Visualisations of Country-Level Panel Data and Hierarchical Models

By:

Oluwayomi Akinfenwa

Under the supervision of:

Professor Niamh Cahill and Professor Catherine Hurley

A thesis submitted in fulfillment of the requirements for the Ph.D. degree in Statistics

at the

Hamilton Institute
Maynooth University
Maynooth, Co. Kildare, Ireland

August 18, 2026

Blessed is the remembrance of the saint.
In loving memory of my father, ***Chief Jeremiah Olufemi Akinfenwa****.*
I wish you were still here to hold this thesis in your hands and read it,
but God, your maker, called you home to eternal glory.
I know you are proud of your baby girl's achievements, even in death.

Declaration

I, Oluwayomi Akinfenwa, declare that this thesis titled, "**Visualisations of Country-Level Panel Data and Hierarchical Models**" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

This thesis work was conducted from September 2021 to February 2025 under the supervision of Professor Niamh Cahill and Professor Catherine Hurley of Hamilton Institute, Maynooth University, Ireland.

Signed: _____

Date: _____

Acknowledgements

"Thus far, by GOD" - This thesis stands as a testament to grace.

Grace sustained, strengthened, and carried me through every season of this doctoral journey. I give all the glory to God Almighty, the giver of wisdom, strength, and endurance. Through seasons of uncertainty, intellectual stretching, emotional exhaustion, and quiet perseverance, He daily renewed my strength and anchored my hope. Time and again, I was reminded of Ecclesiastes 9:11 — that the race is not to the swift, nor the battle to the strong. This journey has affirmed that progress is neither merely about ability nor speed, but about divine enablement, resilience, and unwavering faith. Thus far, it has truly been by God.

I honour the memory of my beloved father. Even in death, your love and sacrifices continue to speak. Thank you, Daddy, for your unwavering commitment to my education, for the investments you made in my dreams long before they had form, and for believing that learning would open doors far beyond what we could see. Your legacy lives in this achievement.

To my precious and supportive mother, thank you for your immeasurable love, your constant encouragement, and your resilience. You pushed me beyond my comfort zones in ways I did not always understand at the time, but now deeply appreciate. When I felt stretched beyond my limits, your words, prayers, and counsel became my anchor. This milestone carries your fingerprints all over it.

To my extraordinary siblings, my true VVIPs, thank you for your consistent prayers, check-ins, laughter, and emotional reinforcement. You reminded me that I never journeyed alone. I carried you in my heart through every milestone, every challenge, and every victory.

To my darling husband, my companion from the inception of this PhD, your presence in this journey has been one of my greatest gifts. From conversations on

Bayesian hierarchical models to reflections on principles of visualisation, you embraced my world fully and wholeheartedly. Thank you for being my safe haven, for offering your shoulder as a place of rest when the weight felt overwhelming, and for believing in me more fiercely than I sometimes believed in myself. You restored my confidence repeatedly and reminded me of my capability at every stage. Your steady love, intellectual partnership, and emotional strength have been indispensable. I am profoundly grateful to walk this journey with you.

My deepest appreciation goes to my supervisors, whose mentorship shaped this work and refined me as a researcher. To Niamh Cahill, thank you for being more than a supervisor. Your guidance, encouragement, and calm reassurance transformed PhD challenges into endurable steps. You consistently reminded me of my capacity and helped me see possibility even in moments of doubt.

To Catherine Hurley, thank you for your rigour, precision, and high standards. You pushed me intellectually in ways that were demanding but transformative. Your insistence on excellence sharpened my thinking and strengthened my ability. Though the process was sometimes intense, it cultivated resilience, clarity, and depth in my research. I am sincerely grateful for your investment in my growth. Working under your joint supervision has been both an honour and a privilege.

I am also deeply grateful to my parents in Dublin, Pastor and Pastor (Mrs.) Gbadebo, their family, and to the church siblings they graciously gave me. Your prayers, counsel, warmth, and kindness, especially during demanding periods created a home away from home. Your spiritual covering and consistent encouragement strengthened me more than words can fully express.

I gratefully acknowledge Science Ireland for the fully funded scholarship that made this doctoral pursuit possible. This opportunity has been a privilege I do not take for granted. I remain thankful to my study country, Ireland, and look forward to contributing meaningfully in return for the investment made in me.

My sincere thanks to the staff of the Hamilton Institute for providing a supportive, stimulating, and welcoming research environment. The administrative and academic support contributed significantly to the successful completion of this work.

Finally, to my PhD friends and colleagues, thank you for the shared experiences, discussions, and encouragement. The conversations, laughter, and mutual support enriched this journey beyond the academic.

This thesis is more than a collection of projects compiled into chapters, it is a journey of love, sacrifice, pain, days of endless tears, faith, and grace.

Funding

This thesis has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant number 18/CRT/6049.

Abstract

This thesis develops diagnostic and visualisation methods to support the exploration, modelling, and evaluation of country-level panel data in global development research. Chapter 3 introduces `wdiexplorer`, an R package that provides an end-to-end workflow for sourcing World Development Indicators data, computing group-aware diagnostic indices, and visualising temporal dynamics across countries and predefined group structures. The package enables identification of patterns, heterogeneity, and outliers in longitudinal indicators, illustrated through an application to global air pollution data.

Building on this exploratory foundation, Chapter 4 presents principled visualisations for the comparison of Bayesian hierarchical model structures in the data space and the evaluation of multiple models simultaneously. These visual tools are designed to complement conventional numerical summaries by revealing how alternative hierarchical specifications influence parameter distributions and uncertainty. Through a case study modelling cross national mathematics trends from the PISA (Programme for International Student Assessment) database, the proposed approach demonstrates how visual comparison can support more transparent and informed model selection.

The methods from Chapter 3 and 4 are then adapted and applied in Chapter 5 in a substantive analysis of global family planning indicators. Using group-aware diagnostic indices and visual comparisons, Chapter 5 evaluates the alignment between observed survey data and model-based estimates from a Bayesian hierarchical time-series framework used for international monitoring of family planning indicators. By jointly examining observed data, modelled trajectories, and diagnostic measures across countries and sub-regions, the analysis highlights where model estimates capture underlying trends and where discrepancies emerge.

Together, the contributions show how integrating structured diagnostics with

carefully designed visualisations enhance the understanding of country-level panel data, improve evaluation of hierarchical models, and support more transparent interpretation of global development indicators.

Publications

Some of the chapters contained in this thesis are being prepared for peer-reviewed journals. Chapter 3 will be submitted to *The R Journal* and the R package to the Comprehensive R Archive Network (CRAN). Chapter 4 has been submitted to the Journal of Data Science, Statistics, and Visualisation and currently under review. Chapter 5 will be submitted to the Public Library of Science (PLOS ONE).

Submitted article (under review):

- Akinfenwa, O., Cahill, N., & Hurley, C. (2024). Visualization for Exploratory Modeling Analysis of Bayesian Hierarchical Models. Under review in the *Journal of Data Science, Statistics, and Visualisation*. *arXiv* preprint: https://doi.org/10.48550/arXiv.2412.03484.

Articles in preparation

- Akinfenwa, O., Cahill, N., & Hurley, C. (2025). wdiexplorer: An R package Designed for Exploratory Analysis of World Development Indicators (WDI) Data. *arXiv* preprint: https://doi.org/10.48550/arXiv.2511.07027
- Akinfenwa, O., Cahill, N., & Hurley, C. (2026). Using Time Series Measures to Explore Family Planning Survey Data and Model-based Estimates. *arXiv* preprint: https://doi.org/10.48550/arXiv.2602.17034

List of Abbreviations

Abbreviation Definition

APIs ..... Application Programming Interfaces

BHM ..... Bayesian Hierarchical Model

DHS ..... Demographic and Health Surveys

FP2030 ..... Family Planning 2030 Initiative

FPEM ..... Family Planning Estimation Model

FPET ..... Family Planning Estimation Tool

GBD ..... Global Burden of Disease

MICS ..... UNICEF Multiple Indicator Cluster Surveys

OECD ..... Organisation for Economic Co-operation and Development

OWID ..... Our World in Data

PISA ..... Programme for International Student Assessment

PMA ..... Performance Monitoring for Action

PM2.5 ..... Particulate Matter with a diameter of 2.5 micrometers (μm)

UNESCO .... United Nations Educational, Scientific and Cultural Organisation

UNPD ..... United Nations for Population Division

WDI ..... World Development Indicators

WHO ..... World Health Organisation

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Country-level panel data are repeated measurements of the same variables collected over time across multiple countries. These data are used by international organisations, including the World Bank, Our World in Data, and the Organisation for Economic Co-operation and Development, and are extensively used to support research and policy analysis across a wide range of sectors, including public health, education and finance (Yaffee, 2003; Hsiao, 2022). Exploratory analysis of country-level panel data allows researchers to examine both changes within countries over time as well as differences between countries (Hsiao, 2005) and visualisation plays a vital role in supporting such analysis.

Exploratory analysis, supported by visualisation, provides a framework to examine temporal dynamics and cross-country variation in country-level panel data. It reveals the underlying structures and identifies anomalies, heterogeneity, and other interesting characteristics of the data, revealing features of the data that may not be apparent from summary statistics alone (Unwin and Valero-Mora, 2018).

A range of methods and tools have been developed to support the exploratory analysis of panel and time series data. Feature-based approaches, for example, those developed by Hyndman and Athanasopoulos (2021), summarise time series using measures of data characteristics such as trend and seasonality, enabling comparison of series across multiple countries. In addition, there are existing R packages that support the exploration of time series such as `tsibble` Wang et al. (2020), `tsibbletalk` Wang and Cook (2020), `brolgar` Tierney et al. (2022) providing functionalities for summarising, and exploring multiple time series simulta-

neously. These approaches have proven valuable for understanding temporal behaviour across collections of series.

However, many existing methods are not designed to explore temporal patterns in the context of hierarchical structures that are often present in the data. For example, countries can be structured into meaningful groupings such as geographical regions, or income classifications. Incorporating grouping structures into the exploratory analysis of country-level panel data could facilitate the identification of group-level patterns, within-group differences, or countries whose temporal behaviour deviates from their peers. This motivates the need for an exploratory framework that integrates grouping information directly into diagnostic measures and visualisations, which forms a central focus of this thesis.

Beyond exploratory analysis of raw data, visualisation has played a central role in exploratory modelling analysis, that is, the investigation, interpretation, and evaluation of statistical model outputs. There is a growing recognition of the importance of visualisation throughout the entire modelling workflow, particularly for hierarchical models. In this context, Unwin and Valero-Mora (2018) characterise exploratory modelling analysis as the "evaluation and comparison of many models simultaneously", a task for which visualisation is especially well suited. Within Bayesian analysis, the value of visualisation across all stages of the workflow, from model development and diagnostics to the communication of results, has been well established (Gabry et al., 2019; Gelman, 2007; Raudenbush and Bryk, 2002; Correa-Álvarez et al., 2023).

Wickham et al. (2015) identify three complementary strategies for visualising statistical models: (1) displaying models in data space, (2) examining and comparing collections of models, and (3) exploring the process of model-fitting. Building on these principles, this thesis develops methods and tools that advance exploratory modelling analysis specifically for hierarchical models by enabling more structured visualisation of model outputs in relation to their underlying grouping structures. These approaches support the systematic exploration, comparison, and interpretation of both multilevel regression and Bayesian models, strengthening understanding of model behaviour across levels and improving the comprehension of hierarchical modelling.

1.2 Thesis outline

This thesis develops and evaluates methods for analysing and visualising both country-level panel data and outputs of hierarchical models fit to that data. It introduces a group-informed exploratory workflow and presents visualisation principles to enhance exploration, interpretation, and communication of both the data and model outputs (posterior estimates, parameter and hyper-parameter estimates).

Chapter 2: Background

Chapter 2 reviews the literature relevant to the exploration and visualisation of country-level panel data and hierarchical models. It examines established approaches for summarising, diagnosing, and visualising temporal data, as well as methods for analysing hierarchical and Bayesian models, including commonly used frameworks such as JAGS and `brms`. The chapter also reviews principles and best practices for data visualisation.

Chapter 3: wdiexplorer: An R package for Exploratory Analysis of World Development Indicators (WDI) Data.

Chapter 3 presents the `wdiexplorer`, an open source R package and demonstrates its functionality using the $PM_{2.5}$ air pollution dataset. As the acronym implies, "World Development Indicators explorer", `wdiexplorer` is primarily designed for the exploratory analysis of country-level panel data from the World Development Indicators (WDI) database, but it is not limited to this database as it can be used to explore other datasets that are organised as repeated observations of the same variables over time across multiple countries. The WDI database provides a vast array of global development data across hundreds of countries over many decades. However, the complexity of these datasets often pose challenges for interpreting and identifying underlying temporal patterns across countries. To address these challenges, Chapter 3 introduces `wdiexplorer`, a comprehensive workflow that supports the exploratory analysis of country-level panel data.

The development of `wdiexplorer` is inspired by the scagnostic (scatterplot diagnostics) framework (Tukey and Tukey, 1985; Wilkinson and Wills, 2008). Scagnostics quantify structural features of two-dimensional scatterplots, such as outlier, skewness, and shape, to support data exploration across a large collection of scat-

terplot matrices and facilitate the identification of distinctive patterns Dang and Wilkinson (2014) in unstructured two dimensional relationships. In this chapter, we adapt the conceptual framework of scagnostics, without applying the original measures directly, and extend it with some selected time series features from Hyndman and Athanasopoulos (2021) which are already implemented in the `feasts` R package alongside additional diagnostic measures suitable for panel data series.

The `wdiexplorer` package workflow is structured into three main stages: data sourcing and preparation, computation of diagnostic indices, and visualisations to support the interpretation of these indices. While existing tools like `tsibbletalk` and `brolgar` offer time series exploration, `wdiexplorer` specifically addresses the need for comprehensive exploratory analysis of country-level panel data by integrating predefined grouping structures with its collection of ten diagnostic indices and visualisation tools. This integrated approach facilitates the understanding, interpretation and identification of temporal patterns, outliers and other underlying features in country-level panel datasets.

Chapter 4: Visualisations for Exploratory Modelling Analysis of Bayesian Hierarchical Regression Models

Chapter 4 presents visualisation principles for exploratory modelling analysis of Bayesian hierarchical regression models. This chapter is motivated by the importance of visualisation in exploratory data analysis, which has been described as "the evaluation and comparison of many models simultaneously" (Unwin and Valero-Mora, 2018). Visualisation has been widely established as beneficial at all stages, from exploratory data analysis to model development, diagnostics and communication of results (Gabry et al., 2019).

In this chapter, we propose a set of five visualisation principles designed for displaying hierarchical model estimates in data space and for examining collections of multiple Bayesian hierarchical models through their parameter and hyperparameter estimates simultaneously. These principles are not limited to Bayesian model outputs and can be applied to model outputs from other multi-level regression analyses. The proposed visualisation principles support model comparison by highlighting information sharing and variation across hierarchical levels. Our visualisation designs give careful consideration to choices of layout, colour, ordering and scaling to facilitate effective comparisons.

The relevance of our visualisation principles is demonstrated by a subset of the Programme for International Student Assessment (PISA) dataset obtained from the Organisation for Economic Cooperation and Development (OECD) website. Using this dataset, Chapter 4 evaluates model fits, posterior parameter and hyperparameter distributions, facilitates model comparison and supports informed model choice. We also extend our visualisation principles to compare model-based predictions from the selected model with observed PISA 2022 outcomes, illustrating the importance of our visualisation design for interpreting deviations between predicted and observed values.

Chapter 5: Using Time Series Measures to Explore Family Planning Survey Data and Model-based Estimates

Building on the relevance of Chapters 3 and 4 to exploring panel data series, Chapter 5 applies the diagnostic indices of `wdiexplorer` and the outlined visualisation principles for effective visualisation of panel data and model outputs to examine the alignment between model-based estimates and observed data using a case study on modern contraceptive use, a key family planning indicator. The Family Planning Estimation Model (FPEM) is a Bayesian hierarchical time series model that produces annual estimates and projections of key family planning indicators and takes the survey observations as input. The observed data are presented as the proportion of the population with access to modern contraceptive methods.

This chapter assesses consistency and variation between the survey data and model-based estimates of modern contraceptive use using selected diagnostic indices of `wdiexplorer` and the five visualisation principles proposed in the previous chapter which are: (1) one panel per individual-level (country), (2) panel grouping to represent hierarchical structures, (3) the use of colour to distinguish and compare hierarchies, (4) panel ordering to facilitate comparison, and (5) consistent panel scaling for effective comparison and interpretation. The selected indices quantify temporal variation, and trend strength in the proportion of modern contraceptive use across the 85 focus countries of the FP2030 initiative, which represent populations with significant barriers to accessing family planning services. This supports the identification of cases where the model aligns with the observed data and where notable deviations arise. Sub-regional groupings are incorporated as a predefined hierarchical structure, extending the framework proposed in Chapter 3

to support exploratory analysis. The analysis also examines the residuals using suitable diagnostic indices of `wdiexplorer` and its visualisation functionalities are used to present the residuals, data series and the model-based trends for countries with notable deviations.

In general, Chapter 5 demonstrates how group-based exploratory diagnostics and visualisations enhance comprehension, interpretability, and assessment of model-based estimates. It highlights cases where the model diverges from the observed data, particularly in the presence of abrupt changes, temporal noise, or situations where the group-level pooling effects may not fully capture local temporal dynamics. It also emphasises that this type of analysis can be extended to any model-based estimates to assess their alignment with observed data.

Finally, Chapter 6 concludes the thesis with final remarks, summarises the contributions made, and discusses relevant future work. Key achievements include the creation of the `wdiexplorer` package for analysing temporal patterns in country-level panel data, advancements in examining and comparing multiple hierarchical model simultaneously using our proposed visualisation principles, and the enhancement of the diagnostic indices and visualisation principles to assess the alignment between model-based estimates and observed survey data.

1.3 Outline of the code

The computational work of this thesis is implemented across multiple repositories, with each repository corresponding to a specific chapter. All repositories are publicly available on GitHub via `https://github.com/Oluwayomi-Olaitan?tab=repositories`, providing access to the code, documentation, and example datasets to facilitate reproducibility and further use of the proposed methods to new datasets.

The wdiexplorer repository contains the implementation of the `wdiexplorer` package. It includes scripts for all package functions, R Markdown vignettes demonstrating example workflows for both annual and three-year interval data, and detailed documentation for each functionality.

The Visualisation-for-exploratory-analysis-of-BHM repository contains the European mathematics PISA dataset, scripts for data cleaning, data processing, exploratory analysis of model in data space, Bayesian hierarchical modelling, visualisations for examining a collection of multiple models and the investigation of PISA2022 predictions. Codes for each section of Chapter 4 are presented in a

separate script.

The Using-Time-Series-Measures-to-Explore-Family-Planning-Survey-Data-and-Model-based-Estimates repository contains the family planning datasets (survey and model-based), along with scripts for data processing, exploratory analysis of both datasets and residual analysis. Codes for each section of Chapter 5 are presented in a separate script.

All repositories are structured with clear documentation.

1.4 Summary of datasets

This section gives an overview of the datasets used throughout this thesis, describing their sources, structure, and key variables.

1.4.1 The Programme from International Student Assessment (PISA) dataset

The Programme for International Student Assessment (PISA) is an international study developed and coordinated by the Organisation for Economic Co-operation and Development (OECD). It aims to provide policy-oriented, and internationally comparable measures of student performance in reading, mathematics, and science literacy for 15-year old students (Schleicher, 2019). Tests developed and mandated by OECD are administered in three-year cycles since its initiation in 2000. The assessment includes a 120-minute cognitive test and a 35-minute student questionnaire, which collect students' demographic information, learning environment, and attitudes toward learning. PISA scores are reported on a scale with a mean score of 500 points and a standard deviation of 100 points (OECD, 2023a).

In Chapter 4, PISA mathematics scores are obtained directly from the OECD database, and the European subset of participating countries is extracted for the period 2003 to 2022. The 2000 PISA mathematics scores are excluded, as the first full assessment in mathematics was conducted in 2003 (as documented by PISA). The extracted dataset provides country-level mean PISA mathematics scores, alongside other variables which are the grouping variables (region and income), used for the integration of grouping structures in the analysis and visualisation presented in Chapter 4.

1.4.2 The $PM_{2.5}$ air pollution dataset

$PM_{2.5}$ is the concentration of fine particulate matter in the air with diameter of 2.5 micrometers (μm) or less (Harishkumar et al., 2020). The World Health Organisation identifies ambient air pollution, particularly $PM_{2.5}$, as a major contributor to the Global Burden of Disease (GBD), with its impacts felt most strongly by children, older adults, and individuals with pre-existing health conditions (Woodruff et al., 2006). Short-term exposure to $PM_{2.5}$ is associated with respiratory irritation, airway inflammation, and exacerbation of asthma, while long-term exposure increases the risk of cardiovascular disease, chronic respiratory diseases, stroke, lung cancer, and premature mortality (Corrigan et al., 2018). To quantify population exposure, the World Development Indicators(WDI) provide estimates of mean annual exposure to $PM_{2.5}$ air pollution ($\mu g/m^3$), drawing on air pollution exposure data from the GBD Study 2021 produced by the Institute for Health Metrics and Evaluation, with coverage spanning 1990 to 2021 (The World Bank, 2025b).

The WDI (Arel-Bundock, 2025b) R package provides access to this dataset through the indicator code `EN.ATM.PM25.MC.M3`. The dataset contains 13 variables, which is consistent with the standard variable structure of WDI datasets. This dataset contains mean annual exposure to $PM_{2.5}$ measured annually for each country. A distinctive feature of the WDI data is its grouping variables, that can be considered to form a hierarchy which makes it well suited for demonstrating the `wdiexplorer` workflow. In Chapter 3, the dataset is used to illustrate the functionalities of `wdiexplorer`, compute diagnostic indices, visualise and interpret the results in the context of PM2.5 air pollution, mean annual exposure.

1.4.3 The family planning dataset

Progress in family planning are monitored via data collected through household surveys across various key indicators (Cunningham et al., 2015). These indicators include contraceptive prevalence, which measures the proportion of women using any contraceptive method either modern or traditional contraceptive methods. According to Hubacher and Trussell (2015), modern contraceptive methods include medical products or procedures such as pills, intrauterine devices, implants, condoms, and sterilisation, which are designed to reliably prevent pregnancy. Traditional contraceptive methods include behavioural and natural approaches, such as withdrawal, rhythm methods, douching, and folk methods. Contraceptive prevalence is defined

as the proportion of the target population, typically women of reproductive age (15 - 49) using a specific method at a given time relative to the total population of the same group.

Family planning survey data are publicly available through the United Nations Population Division website (United Nations Department of Economic and Social Affairs, Population Division, 2024). In Chapter 5, these data are accessed using the `fpemlocal` R package Guranich et al. (2021), which provides observational data for each country-year pair across key family planning indicators. For the analysis in Chapter 5, we extract observational data on modern and traditional contraceptive use only.

2

Background

The background of the thesis is presented in this chapter, providing a clear overview of country-level panel data, Bayesian hierarchical modelling, and existing approaches for visualising such data and model. It also identifies gaps in the literature that motivate this research.

2.1 Country-level Panel Data

Country-level panel data, also referred to as longitudinal data, are repeated observations for multiple countries over a defined time period. The data have two dimensions: a spatial dimension that represents countries and a temporal dimension that represents periodic observations such as years, quarters, months or days (Yaffee, 2003). Each observation is identified by a country and a time period, which allows the study of both differences across countries and changes within countries over time (Hsiao, 1985). In country-level panel, when all countries have data for every period, the dataset is called a balanced panel, and when some countries have missing observations, it results in an unbalanced panel (Hsiao, 2005). The data often include time-invariant variables, which remain constant for a country over time, such as geographic location or land area, and time-varying variables, which change across both countries and periods, such as economic or social indicators.

Country-level panel data are sourced and maintained by a wide range of international organisations and national statistics offices. They provide comprehensive information across economic, health, education, social, and other sectors. Among the most frequently cited sources is the World Bank, which maintains databases such as the World Development Indicators, the Worldwide Governance Indicators,

Statistical Capacity Indicators, Education Statistics, Gender Statistics, and Health National and Population Statistics (World Bank, 2026). Various United Nations agencies also provide country-level data, including the Food and Agriculture Organisation for agricultural and food self-sufficiency statistics, United Nations Educational, Scientific and Cultural Organisation (UNESCO) also provide literacy and education statistics (UNESCO, 2026). Health-related outcomes can be obtained from the World Health Organisation through the Global Health Observatory and the National Health Accounts databases (World Health Organization, 2026).

In addition to the above databases, these organisations provide Application Programming Interfaces (APIs) and packages for accessing country-level panel data programmatically (Warin, 2021). These APIs follow a design that allows users to specify dimensions such as indicator codes, country codes, and time ranges to retrieve data in formats like JSON or XML directly from their servers. In the R programming environment, there are packages that interface with these organisations APIs to streamline data retrieval. To retrieve data from the World Bank API, the widely used R packages are `WDI` (Arel-Bundock, 2025b) and `wbstats` (Piburn, 2020), a more recent package that also interfaces with the World Bank API which is `wbwdi` (Scheuch, 2025b). These packages enable users to download data for single or multiple indicators across all countries or selected countries within user-defined time periods.

There are organisations that maintain databases of secondary and standardised country-level panel data, such as the Organisation for Economic Co-operation and Development (OECD) (Organisation for Economic Co-operation and Development (OECD), 2026) and Our World in Data (OWID) (Our World in Data, 2026). Both organisations compile from national statistical offices and other global organisations like the World Bank, WHO, and UNESCO. In some domains, the OECD collects data directly through structured questionnaires, such as the PISA surveys. The OECD and OWID data portals provide interactive charts and dashboards for visualising indicators. Programmatic access to these resources is available through R packages: the `owidapi` (Scheuch, 2025a) R package provides programmatic access to the OWID Chart API, which hosts more than 5,000 publicly available datasets covering global challenges such as poverty, disease, hunger, climate change, and inequality. The `OECD` (Persson, 2021) R package also provides comparable functionality for OECD databases, offering access to over 300 indicators across multiple economic and social categories, with tools to explore and retrieve data (Warin,

2021).

Data cleaning and tidying are essential steps after extracting country-level panel data from databases, APIs, or R packages. Raw data often contain country-specific blocks of observations recorded over multiple periods and may not be structured in a form suitable for analysis. Cleaning involves harmonising country codes where necessary and addressing missing values, while tidying involves identifying the underlying variables and observational units and restructuring the dataset into a consistent analytical format. A widely adopted standard is the tidy data format proposed by Wickham et al. (2017), where each variable is stored in a separate column, each observation is represented by a row, and each type of observational unit forms a table.

In R, several packages are used to clean and tidy data. The `countrycode` (Arel-Bundock et al., 2018) package is widely applied to harmonise country identifiers across different International Organisation for Standardisation coding schemes. The `dplyr` (Wickham et al., 2023) package is commonly used for data transformation and restructuring. The `tsibble` (Wang et al., 2020) package extends tidy data principles to temporal data by explicitly defining an index variable for time and a key variable for country, enabling panel datasets to be represented as indexed longitudinal data.

Other packages in R that focus on tidying panel data include `pmdplyr` and `panelr`. The `pmdplyr` (Huntington-Klein and Khor, 2026) package extends the tidy data workflow for panel datasets by requiring an identifier variable (country) and a time variable (year) for panel-specific transformation. The `panelr` (Long, 2020) package provides functions for reshaping panel data between long and wide formats and for preparing data for panel modelling. These packages facilitate data cleaning, restructuring and preparation necessary for further exploratory analysis or modelling of country-level panel data.

2.2 Bayesian Hierarchical Modelling

The hierarchical structure of country-level panel data makes it well suited for Bayesian hierarchical models (BHMs), which can effectively handle repeated observations over time across multiple countries. BHMs account for both variation within and between countries, resulting in more accurate model estimates. In Chapter 4 of this thesis, we analyse PISA mathematics scores for European countries

using multiple BHMs, while in Chapter 5, we use visualisation methods to explore the estimates and projections of the family planning estimation model (FPEM), a country-based BHM used to model family planning indicators.

BHMs provide an analytical framework for integrating information across multiple levels data (Gelman and Pardoe, 2006; Rouder et al., 2014). The models are built on Bayesian analysis methods, which differs from the traditional frequentist approaches that evaluate the probability of the data under fixed parameter assumptions. Bayesian methods, in contrast, combine prior information with observed data to estimate the posterior distribution of parameters (McGlothlin and Viele, 2018). BHMs have two main attributes: (1) the ability to integrate information across multiple levels of the grouping structure of the data into the model; and (2) the use of prior distributions to represent available information about expected values and variability at each level of the model, even when that prior information is limited. These priors are combined with observed data to produce estimates of underlying effects at every level of the hierarchy.

A key feature of BHMs is their partial pooling ability, which provides a compromise between treating each country as entirely independent (no pooling) and assuming all countries are identical (complete pooling). For country-level panel data, partial pooling enables information sharing across countries, especially for those with limited data. The estimates for countries with fewer observations are pulled towards their group, which could be defined, for example, by geographical region or income level, while estimates for countries with larger sample sizes rely more on their own observed data.

Bayesian inferences are performed using computational methods. Many models use Markov Chain Monte Carlo (MCMC) techniques, such as Gibbs sampling or Hamiltonian Monte Carlo (HMC) (Carlo, 2004; Gamerman and Lopes, 2006), implemented in tools including Bayesian inference Using Gibbs Sampling (BUGS) (Spiegelhalter et al., 1996), Just Another Gibbs Sampler (JAGS) (Plummer et al., 2003), and Stan (Gelman et al., 2015). Other approaches, such as Integrated Nested Laplace Approximation (INLA) (Martins et al., 2013), employ deterministic approximations instead of MCMC, particularly latent Gaussian models (Jreich et al., 2022).

In R, there are packages that facilitate the estimation of BHMs using these computational methods. JAGS is commonly accessed through the `rjags` (Plummer, 2025) and `R2jags` (Su and Yajima, 2024) packages which allow users to specify the

distribution of the data at different levels of the hierarchy, parameters that represent unknown quantities such as group-level effects, population means, and variances, run MCMC simulations, and analyse results within the R environment. The `brms` (Bürkner, 2018) package provides a high-level interface to Stan, making it easier to fit hierarchical models using HMC. In Chapter 4 of this thesis, we employ the computational approach of `brms` to analyse PISA mathematics scores for European countries.

The `brms` package adopts an extended `lme4` (Bates et al., 2015) formula syntax to simplify model specification without manually writing Stan code. `lme4` provides a frequentist framework in R for fitting multilevel models to hierarchical data. One of the key strengths of `brms` is its support for many model families, including Gaussian, binomial, Poisson, and non-linear models. In addition, `brms` provides flexible prior specification for model parameters and hyper-parameters, although default priors are available. It also provides tools for model assessment and comparison such as leave-one-out cross-validation (LOO) (Vehtari et al., 2017) and the Watanabe-Akaike information criterion (WAIC) (Watanabe and Opper, 2010), leveraging the `loo` (Vehtari et al., 2024) R package. For users less comfortable with R syntax, the `shinybrms` (Weber et al., 2022) package offers a graphical user interface to simplify model specification and analysis.

A simple example of a BHM can be specified as follows, illustrating how country and group-level parameters, as well as population-level hyper-parameters, are modelled simultaneously:

$$Y_{ct} \sim \text{Normal}(\alpha_c + \beta_c t, \sigma^2) \tag{2.1}$$

where

Y_{ct} is the observed data for individual country c at time t,

α_c is the country-specific intercept,

β_c is the country-specific slope, and

σ^2 is the variation within countries.

With countries as the first stage of the hierarchical structure, the model is written as:

$$
\begin{aligned}
\alpha_c &\sim \text{Normal}\left(\theta_{g[c]},\ \sigma^2_{\alpha_{g[c]}}\right) \\
\beta_c &\sim \text{Normal}(\omega_{g[c]},\ \sigma^2_{\beta_{g[c]}}) \\
\theta_g &\sim \text{Normal}\left(\mu_\theta,\ \sigma^2_\theta\right) \\
\omega_g &\sim \text{Normal}\left(\mu_\omega,\ \sigma^2_\omega\right)
\end{aligned}
\tag{2.2}
$$

where

θ_g is the intercept corresponding to grouping structure g (such as geographical region or income group),

ω_g is the slope corresponding to grouping structure g,

$g[c]$ indicate the group(region or income level) associated with country c,

$\sigma^2_{\alpha_g}$ represents the intercept variation across countries within group g,

σ^2_β represents the slope variation across countries within group g,

μ_θ represents the overall population intercept,

μ_ω represents the overall population slope,

σ^2_θ captures the variation of the group-level intercept θ_g around the population intercept μ_θ,

σ^2_ω captures the variation of the group-level intercept ω_g around the population intercept μ_ω.

An application that we focus on for this thesis is BHMs fit to PISA study. In previous work the OECD applies non-Bayesian multilevel regression models that incorporate socio-economic measures at both student and school-level to examine how performance relates to factors such as persistence or emotional control (Schleicher, 2019). Goldstein (2004) further applied a two-factor multi-level model to analyse PISA mathematics scores for France and England, to demonstrate how complex modelling can reveal detailed and informative differences between countries than standard unidimensional approaches commonly used by the OECD. With the two-factor multi-level model, Goldstein (2004) simultaneously accounts for multiple dimensions of student ability and the hierarchical grouping of students within schools. In Chapter 4 of this thesis, we analysed PISA mathematics scores across European countries using the BHM fitted with `brms`, while accounting for its hierarchical grouping.

BHMs are widely used in other fields with grouped data. In medicine and

public health, they support meta-analyses of clinical trials and studies of rare diseases, where small sample sizes benefit from borrowing information across groups (McGlothlin and Viele, 2018). In psychology and cognitive science, BHMs help eliminate asymptotic bias in non-linear models of recognition memory and signal detection (Rouder and Lu, 2005). In the context of this thesis, in Chapter 5 we focus on visualisations for the model-based estimates of a BHM that has been used in monitoring family planning progress through the Family Planning Estimation Model (FPEM), which provides country-level estimates and projections of key family planning indicators (Alkema et al., 2013; New et al., 2017; Cahill et al., 2018) to support national monitoring, data review, and progress tracking toward commitments such as the Family Planning 2030 initiative (FP2030) (FP2030, 2025).

2.3 Summarising Country-level Panel Data

One of the initial steps in understanding complex datasets is to summarise them into interpretable measures that capture their key characteristics. This is especially important for country-level panel data, which combine observations from multiple countries over time across a range of variables. Data can be summarised numerically through measures that describe trends, variation, correlation and observational patterns. It can also be summarised graphically, using plots and visual diagnostics to reveal patterns, outliers and clusters across countries. Together, these approaches facilitate an overview of the overall behaviour of the data and structural behaviours between variables and over time.

There are various methods for summarising data in different contexts. For instance, Scagnostics (scatterplot diagnostics), originally proposed by Tukey and Tukey (1985) and formalised by Wilkinson and Wills (2008), provide a framework for summarising the geometric properties of two-dimensional scatterplots. These measures are: skewed, striated, clumpy, sparse, convex, skinny, stringy, and monotonic which aim to characterise the density, shape and association of point distributions using proximity graphs (Wilkinson and Wills, 2008; Dang and Wilkinson, 2014). Building on this idea, Dang et al. (2012) proposed TimeSeer, a framework for applying scagnostics in the context of time-varying data series, where multiple variables are observed for many entities across time. It converts scatterplots into numerical summaries, reducing the complexity of comparing large datasets while ensuring the identification and interpretation of patterns, anomalies and other in-

teresting characteristics of the data.

While TimeSeer Dang et al. (2012) focuses on summarising geometric relationships in multivariate scatterplots over time using scagnostic measures, there are other measures that assess the temporal behaviour of time series data such as the Highly Comparative Time-Series Analysis framework, which represents each time series using thousands of automatically extracted features capturing distributional properties, autocorrelation structure, nonlinearity, and stationarity (Fulcher et al., 2013). Other approaches include the set of time series summary measures proposed by Hyndman and Athanasopoulos (2021) to characterise the temporal behaviour of individual series. These measures capture key features of time series dynamics, including trend strength, seasonality strength, autocorrelation structure, nonlinearity, and temporal variability. The set of measures proposed by Hyndman and Athanasopoulos (2021) is implemented in the R package `feasts` (O'Hara-Wild et al., 2024b) and provides a useful basis for summarising and comparing temporal patterns across countries which has been applied in exploratory time series analysis of country-level panel data.

In Chapter 3 of this thesis, we present ten diagnostic indices inspired by the scagnostic framework, incorporating some of the already established time series measures of Hyndman and Athanasopoulos (2021). This collection of indices is designed to summarise country-level panel data by quantifying key temporal characteristics such as variation, trend, shape and temporal differences both within and across countries. Our proposed indices extend existing time series measures by explicitly incorporates the grouping structure in country-level panel data into the diagnostic computation, a feature not considered in the `feasts` measures or other existing time series measures.

2.4 Data Visualisation

This section provides an overview of the principles of data visualisation, including approaches for visualising country-level panel data and Bayesian hierarchical models.

2.4.1 Principles of Data Visualisation

History has proven that humans generally process visual information more efficiently than numbers or texts (Tukey et al., 1977; Friedman, 2008). In this context, data visualisation refers to the graphical representation of data designed to reveal patterns, relationships and structure, thereby supporting understanding, comparison, and communication of datasets (Manovich, 2011). By transforming data into visual elements, visualisation facilitates interpretation and understanding beyond what numerical or textual representations can provide (Tukey et al., 1977; Friedman, 2008).

The objective of graphing is twofold: to present quantitative information visually and to support visual analysis by leveraging the human ability to detect structure and patterns rapidly through perception rather than computation (Cleveland, 1985). This makes visualisation particularly effective for exploratory analysis, as it enables the efficient presentation of large datasets and helps reveal trends, correlations, and relationships that may remain hidden in text or numerical based summaries (Islam and Jin, 2019).

Traditionally, data visualisation has been guided by two principles: reduction and the use of spatial variables (Manovich, 2011). The principle of reduction involves representing complex data and relationships through basic graphical elements like points, lines, bars and other geometric shapes. The second principle prioritises the dimensions of the data by mapping them to visual attributes such as position, size or other encodings. These two principles form the basis of many graphical techniques, including bar charts, line graphs, and scatterplots, which use position and shape to represent key characteristics of the data (Kirk, 2019).

Wilkinson (2011) proposed the grammar of graphics which provides a framework for understanding and constructing statistical graphics which treat visualisation as a structured language rather than a collection of isolated chart types. In this framework, a plot is built by combining data with aesthetic mapping, scales, coordinate systems and layering. The grammar of graphics shifts the focus from using predefined graphics templates to create graphs by mapping the variables to the visual properties such as position, shape, and colour.

The visualisation principles described above can be implemented in practice using modern visualisation tools. In the R environment, the `ggplot2` package (Wickham, 2016) provides a structured way to apply these principles when constructing

statistical graphics. `ggplot2` conceptualises graphics as compositions of independent but interrelated components of the underlying data, aesthetic mapping, scales, colours, shapes, coordinate systems and faceting structures. Its layered approach follows the principles of Wilkinson (2011).

Modern approaches to data visualisation extend traditional principles by incorporating interactivity and direct visualisation into data visualisation framework (Cleveland and McGill, 1988; Buja et al., 1996). Interactivity and linked views complement graphical representation by allowing users to manipulate displays, filter information, and explore multiple perspectives of the same data simultaneously (Buja et al., 1996). Direct visualisation is visualisation without the reduction principle. It constructs graphical representations of data using media objects such as images or video frames, thereby preserving visual properties that may be lost when data is reduced to basic graphical techniques. Together, these advancements improve the usefulness of visualisation for data exploration and summarisation.

Similarly, for large datasets, Wills (2011) adopts the Goal-Question-Metric approach, originally from software engineering. In his words, visualisations should be designed with a clear goal, whether exploratory (discovering new features) or presentational (communicating known features) using the following methodology: (1) define the goal by identifying the object of study, (2) formulate questions that the visualisation will answer (3) apply mapping by choosing the most effective visual encodings. These ensures that visualisation designs capture the dataset of interest and the characteristics of its variables.

2.4.2 Visualising Country-level Panel Data

In country-level panel data, visualisation provides a systematic way to assess structure by revealing similarities and differences between countries, and by examining temporal behaviour across countries. Patterns such as trends, fluctuations and outliers in country trajectories can be readily identified. Castellacci and Natera (2011) describes visualisation as a quality check to verify data reliability, while revealing missing values. As such, visualisation serves as a preliminary step of data assessment and exploration prior to any statistical modelling.

Line plots offer an effective means of visualising country-level panel data. Line plots display the data trajectory of each country over time, highlighting trends, fluctuations, and potential anomalies (Friendly and Wainer, 2021). Box plots, the

box and whisker plot introduced by Tukey et al. (1977), summarise distributions across countries or years, emphasising variation and outliers, while heat maps use colour to represent magnitudes, allowing large datasets to be inspected quickly for patterns or unusual observations (Wilkinson, 2011; Gehlenborg and Wong, 2012).

Visualisation of country-level panel data can also be approached by representing the data trajectory of each country over time as a continuous function rather than as discrete observations (Hyndman and Shang, 2010). Exploratory Functional Data Analysis adopts this perspective to examine temporal dynamics, identify trends, and detect outliers. Common tools include line plots to display individual trajectories, rainbow plots to highlight progression with an ordering metric, and heatmaps for summarising magnitudes across time and countries (Qu et al., 2025). Interactive plots further enable dynamic exploration, allowing users to navigate and compare curves efficiently.

To manage high-dimensional functional data, robust functional principal component analysis reduces curves into a lower-dimensional score space (Sawant et al., 2012). Visualisations such as the functional bagplot and functional boxplot map these scores back to the original functional space, highlighting curves, and outlying patterns (Wang et al., 2016). Outliers can be classified as magnitude outliers, which fall outside the typical range, or shape outliers, which deviate structurally despite remaining within the general range (Hyndman and Shang, 2010). Additional tools, such as outliergrams and shape-based boxplots, provide further detection of unusual functional patterns. These methods complement classical visualisation techniques by revealing temporal and structural dynamics that might otherwise remain hidden in country-level panel data (Hyndman and Shang, 2010; Qu et al., 2025). Other existing approaches for examining country-level temporal behaviour include time path plots, which are used to visualise dynamic processes over extended periods by tracking how variables evolve across countries and over time (Castellacci and Natera, 2011).

A variety of R packages, beyond `ggplot2`, facilitate the visualisation of country-level panel data. `tsibbletalk` (Wang and Cook, 2020) provides a structured framework for interactive exploration of country trajectories. `Crosstalk` (Cheng and Sievert, 2023) provides linked interactions across multiple plots, enabling the exploration of patterns in one plot while simultaneously observing their correspondence in other plots. `Brolgar` (Tierney et al., 2022) is another package designed for longitudinal data, with functionalities that offer both summarisation and visualisation

tools that emphasise deviations, trends, and temporal dynamics across countries. Additionally, packages such as `plotly` (Sievert, 2020) and `ggiraph` (Gohel and Skintzos, 2025) render conventional plots interactive, enabling hovering and dynamic filtering of observations by country and time period.

2.4.3 Visualising Bayesian Hierarchical Models

Visualisation plays an important role in the development, assessment, and communication of hierarchical models. Graphical displays complement numerical model summaries by facilitating the understanding of model structure, diagnosing model fit, and interpreting complex parameter estimates (Wickham et al., 2015).

Prior work has highlighted the effectiveness of visualisation in model building, validation, evaluation and communication, particularly for hierarchical and multilevel models (Gelman, 2007; Raudenbush and Bryk, 2002; Gabry et al., 2019). Correa-Álvarez et al. (2023) demonstrates how visualisation is used to compare predictive performance between single-level and multilevel Bayesian logistic regression models. Liu et al. (2021) also describe visualisation as a primary methodological means for conveying the uncertainty inherent in predictive modelling.

According to Gelman et al. (1995), graphical inspection is central to Bayesian model evaluation, particularly for models estimated using Markov Chain Monte Carlo (MCMC) methods. Gelman et al. (2013), introduces trace plots to visualise sampled parameter values across iterations, to assess convergence behaviour. To complement trace plot, they also proposed quantitative convergence diagnostics and posterior predictive checks to assess whether Bayesian models adequately reproduce observed data patterns (Gelman et al., 1996).

Gabry et al. (2019) developed `bayesplot` to formalise Bayesian visualisation as a core component of the Bayesian workflow. In the R environment, `bayesplot` (Gabry and Mahr, 2025) provides a systematic framework for visualising parameter estimates, uncertainty, convergence behaviour, and predictive performance. It includes trace plots, posterior predictive checks, posterior distributions and predictive summaries, residual and quantile-quantile plots.

Another relevant R package for visualising Bayesian workflow is `tidybayes` (Kay, 2023), which facilitates the extraction, summarisation, and visualisation of Bayesian model outputs in a tidy data format. The package allows posterior estimates, predictions, and uncertainty intervals to be readily extracted and formatted

for graphical representation using `ggplot2`.

2.5 What's Next?

Collectively, the preceding Sections 2.1, 2.2, 2.3, 2.4 provide the background for the main contribution of this thesis.

Chapter 3 introduces diagnostic indices and visualisation tools for summarising and exploring country-level panel data in the WDI context. By extending time series measures to incorporate the grouping structure of panel data, it demonstrates how visual diagnostics can enhance the identification and interpretation of patterns, outliers, and temporal behaviours. The chapter also extends the silhouette concept for cluster validation to develop visual tools for assessing group structure in both the data and the proposed diagnostics.

Existing approaches for visualising Bayesian workflow discussed in Subsection 2.4.3 do not explicitly represent the hierarchical structure of Bayesian models in their plots. As a result, relationships within hierarchical structures are often not directly conveyed. This gap will be addressed in Chapter 4, where we will propose visualisation principles that incorporate the hierarchical structure of BHMs into the display of model outputs. Chapter 4 will demonstrate how model can be visualised in data space and how multiple models can be examined collectively while incorporating their hierarchical structure, by the display of model parameters and hyper-parameters estimates.

Chapter 5 will demonstrate the application of the diagnostic indices from Chapter 3 and the visualisation principles from 4 to explore and assess the survey data and model-based estimates of family planning.

3

wdiexplorer: An R package for Exploratory Analysis of World Development Indicators (WDI) Data.

3.1 Introduction

The World Development Indicators (WDI, "the World Bank collection of development indicators") (The World Bank, 2025d), is a database maintained and published by the World Bank that provides a wide range of global development data. It includes economic, social, demographic and environmental indicators, sourced from 60 international databases. The five largest sources are Education Statistics, the Atlas of Social Protection: Indicators of Resilience and Equity, Quarterly External Debt Statistics, the Disability Data Hub, and the World Development Indicators, with Education Statistics contributing the highest number of indicators (7,182). These data are standardised by the World Bank to ensure consistency across countries and over time. While most WDI series began in 1960, coverage varies by indicator and country, with many indicators, particularly those related to the sustainable development goals, and renewable energy adoption, starting in the 1990s and later years. Data are predominantly collected on an annual basis. However, the reporting frequency varies by indicator, with some recorded biennially or triennially, and some at quarterly intervals.

The WDI data are structured as country-level panel data, consisting of repeated observations of the same variables collected over time from multiple countries. This structured and consistent format captures both changes within individual coun-

tries and variation across countries over time. However, the underlying temporal patterns across countries are not easily identified. To facilitate the exploration of WDI country-level panel data, the WDI database website provides an online interactive tool that allows users to visualise the temporal behaviour of data series across countries. Despite this functionality, overlapping visuals often create complexity highlighting the need for enhanced exploratory visualisation tools. Such tools should allow users to explore data in ways that uncover hidden patterns, identify outliers and other interesting characteristics of the data, and also enhance better interpretation of temporal behaviours.

This chapter introduces the `wdiexplorer` R package, a set of tools designed to enhance the exploratory analysis of country-level panel data of the WDI. In addition to the country-level information of the WDI dataset, it includes predefined grouping variables such as region, income and lending category. The `wdiexplorer` explicitly incorporate these groupings into the exploration of country-level panel data. It computes group-based measures that characterise panel data behaviour and produces interactive visualisations to identify patterns, temporal dynamics and shape-based features within and across countries. This package also helps to identify outliers and structural variations that might otherwise remain hidden when examining countries individually or as a global aggregate. It facilitates these by enabling comparisons of each country with others in the same predefined group, `wdiexplorer` supports a more context-aware exploration of development trends. The package is currently available on GitHub at https://github.com/Oluwayomi-Olaitan/wdiexplorer/.

The remaining sections of this chapter are organised as follows. Section 3.2 provides an overview of the background to this study, outlining the importance of incorporating grouping structures in country-level panel data exploration, reviewing existing exploratory tools, and examining summaries of time series data. Section 3.3 details our methodology, which extends scagnostics concepts and some established time series measures to exploratory analysis workflow through a collection of diagnostic indices that characterise the key behaviours of the data series. In Section 3.4, we describe the functions of the `wdiexplorer` package. The utility of the package is illustrated using the mean annual exposure levels to ambient $PM_{2.5}$ air pollution indicator data of the WDI in Section 3.5. Finally, in Section 3.6, we conclude by evaluating the benefits of `wdiexplorer` package, as well as other related future work.

3.2 Background

In this section, we discuss WDI data. We highlight grouping variables present in the data set, alongside the main indicator of interest, that can enhance meaningful interpretation and comparison of country-level behavioural patterns. We also review existing exploratory tools for analysing country-level data, with a focus on identifying patterns, outliers, and other characteristics within the data set.

3.2.1 World Development Indicators (WDI) Data

In R, World Bank-hosted databases are accessible through several packages, including `wbstats` (Piburn, 2020), `WDI` (Arel-Bundock, 2025b), and `worldbank` (Mücke, 2025), each offering slightly different approaches to sourcing and downloading the World Bank data. This chapter focuses on the `WDI` package to source and download data, which provides access to 22,806 indicators across 296 reporting entities. These entities include not only countries, but also territories, regional aggregates, and income-based aggregates. According to The World Bank (2025a), there are currently 189 World Bank member countries along with 28 other territories with populations over 30,000. After excluding non-country aggregates from the full list of WDI reporting entities, 217 distinct country and territory entities remain with individual records.

The WDI data contain additional variables that capture geographic and economic factors of each country and territory entities. They are classified by the World Bank into 7 geographic regions: East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia, and Sub-Saharan Africa. The World Bank also divides their economies into four income groups: low income, lower-middle income, upper-middle income, and high income. This classification is updated annually on the 1st of July, based on the Gross National Income (GNI) per capita from the previous calendar year. Once assigned, a country's income group remains fixed for the entire fiscal year until the next update. Countries are classified into three lending categories: International Development Association (IDA), International Bank for Reconstruction and Development (IBRD), and Blend based on the operational policies of the World Bank. IDA countries are low income and qualify for financial loans with low to no interest. IBRD countries are middle-income and have access to standard market-based loans

while Blend countries are eligible for both IDA loans and standard IBRD loans (The World Bank, 2025c).

Given the country-level panel structure of the WDI data, the predefined grouping information available in the data set (region, income, and lending) groupings provides essential information that can be utilised to enhance more meaningful comparisons of country behavioural patterns. Gelman (2007) highlight that grouping structures in country-level panel data are widely recognised in both exploratory and modelling analyses, and emphasise that overlooking grouping structures, such as countries nested within geographic and socioeconomic groupings can lead to oversimplified or inaccurate interpretation of the data. Explicitly accounting for inherent groupings facilitates the identification of meaningful patterns that might be hidden when such structures are ignored. It also provides a clearer understanding of group-wise patterns, highlights sources of variation, and supports the detection of outliers. For example, interpreting Ireland's birth rate over time is more informative when explored in relation to other European countries, where shared policy may shape similar demographic trends, rather than comparing it to countries in geographically and socio-economically distinct regions, like Sub-Saharan Africa or South Asia.

Grouping structures offer two crucial comparative perspectives: understanding countries relative to other countries in their geographical or socioeconomic groups, and identifying group-wise patterns across levels, facilitating comparison between groups. Guided by these perspectives, our approach to explore country-level panel data is organised around two complementary strategies: (1) a collection of diagnostic indices that characterise panel data behaviour, (2) group-informed exploration of country-level panel data that leverage the predefined groupings of the data through interactive visuals to capture behavioural patterns and highlight group-based features.

3.2.2 Country-level Panel Data Exploration

Exploration of country-level panel data is the process of revealing patterns, identifying outliers and other structural behaviours among observational units across countries and over time. Traditional approach of exploring country-level panel data often rely on line plots of series trajectories organised by grouping variables. However, these can become overwhelming and insufficient to capture complex tem-

poral and group structures in large data sets. While ensemble graphics provide an effective way to explore multiple faceted temporal data, Wang and Cook (2021) discuss the challenges of managing such complexity during exploratory analysis, particularly the difficulty in visualising grouping structures and temporal dynamics simultaneously. To address this challenge, Wang and Cook (2021) developed the `tsibbletalk` (Wang and Cook, 2020) R package, which leverage the `tsibble` data structure (Wang et al., 2020) that produces a tidy format of time series and panel data as index and key for series identification. The `tsibbletalk` package is integrated with the `crosstalk` R package (Cheng and Sievert, 2023) that enables filtering of data views and highlighting in html-widgets without relying on a Shiny app, to provide linked brushing by hovering across multiple coordinated views to make grouping structures interactive and visually explorable.

Several other `R` packages also provide valuable tools for exploring time series and panel data. For example, `brolgar`, which also uses the `tsibble` data structure to produce a tidy data series, enhances the exploration of longitudinal data by providing tools to visualise individual trajectories in large datasets (Tierney et al., 2022). It supports sampling, stratification, and feature-based summaries of the data series, which facilitates the identification of unusual patterns within large data sets of individual series.

Also, there are existing measures specifically designed to capture the temporal characteristics of time series data (e.g., Hyndman and Athanasopoulos (2021), Henderson and Fulcher (2025)). The `theft` ecosystem proposed by Henderson and Fulcher (2025) provides tools for deriving time-series features and supporting exploratory visual analysis of time series data. Also, the time series features proposed by Hyndman and Athanasopoulos (2021) include summary statistics, autocorrelation measures, trend and seasonal strength measures, as well as tests for consistent behaviour over time. Such measures aim to uncover meaningful temporal behaviours and are implemented in the `feasts` package. `feasts` (O'Hara-Wild et al., 2024b) focuses on a concise set of features proven effective in time series analysis. `tsibbletalk` and `brolgar` leverage `feasts` and `fabletools` (O'Hara-Wild et al., 2024a) respectively to summarise the temporal features of the data series but does not specifically address the integration of grouping structures into its exploratory workflow, limiting its ability to explore and compare patterns across and between groups with their temporal features.

Although the WDI data do not strictly qualify as functional data, they exhibit

comparable characteristics, consisting of observations from multiple countries over time. Qu et al. (2024) introduced visualisation tools for functional data (which they call "exploratory functional data analysis" or EFDA) such as rainbow plots, functional boxplots and functional bagplots and provided techniques for outlier detection and functional data clustering to detect dominant behaviours and anomalies across curves. Functional clustering, in particular, is used to partition data based on shared shape or magnitude features, supporting data-driven identification of subgroups. The rainbow plot incorporates data ordering feature that colours observations based on their order, which can reflect various indices such as time or other relevant characteristics to enhance the exploration process. However, while the Qu et al. (2024) work supports identifying groupings and patterns based on shared data features, it does not explicitly incorporate predefined grouping structures into its framework. The `wdiexplorer` package builds on its idea of rainbow plot to order series trajectories based on summary measures of the proposed collection of diagnostic indices.

If we consider WDI data as scatter plots (plots of observational units versus time) for individual countries, the idea of scagnostics (scatterplot diagnostic) (Tukey and Tukey, 1985; Wilkinson and Wills, 2008) can be employed to highlight interesting relationships within the data. Scagnostics quantify structural features of two-dimensional scatterplots, such as outliers, skewness, and shape, which support data exploration across large scatterplot matrices by identifying distinctive patterns (Dang and Wilkinson, 2014). However, while scagnostics provide valuable insights into the distributional and geometric characteristics of scatterplots, they are designed for unstructured 2D relationships and do not account for structured observational units or grouping variables that are key attributes of country-level panel data.

The `wdiexplorer` package adapts the conceptual framework of scagnostics measures (whilst we do not directly apply the original scagnostics measures), incorporate and extend their underlying principles alongside selected time series features from Hyndman and Athanasopoulos (2021) (as implemented in the `feasts` package), and additional measures tailored to panel data series. The package construct a collection of diagnostic indices that quantify variation, trend and shape, and sequential temporal structure across time series. Measures are calculated with a consideration of grouping structures to enhance the exploratory analysis of country-level panel data. Additionally, it refines the interactive visualisation approach used by `tsibbletalk`

and incorporates the concept of colouring individual units based on their ordering, as seen in the rainbow plot of EFDA, to enhance interactive visuals that summarise behavioural patterns and highlight group-based features of country-level panel data based on their metric values.

3.3 Methodology

In this section, we present our methodology, inspired by the scagnostics (scatterplot diagnostic) framework originated by Tukey and Tukey (1985) and formalised by Wilkinson and Wills (2008), which provides a set of quantitative measures to detect anomalies in density, shape, association, and other features of 2D scatterplots. We extend scagnostics concepts with some time series features of the `feasts` package to the exploratory analysis of country-level panel data through a collection of diagnostic indices that characterise key behaviours such as trends, variability and similarities of the data series. In addition, considering the heterogeneity typically present in country-level panel data, we explicitly incorporate grouping information (e.g., geographic or socioeconomic) present in the data into the exploratory process through static and interactive visuals.

3.3.1 A Collection of Diagnostic Indices that Characterise Panel Data Behaviour

Our collection of diagnostic indices are organised into three categories: (1) variation features, (2) trend and shape features, and (3) sequential temporal features. Together, these indices characterise country-level panel data and support the identification of trends, patterns, outliers, and other potentially interesting aspects of variation over time.

(1) **Variation Features**

The variation features capture the overall differences in the variable of interest within and between groups of countries. These include an overall dissimilarity measure for each country, a dissimilarity measure for each country relative to its predefined group, and an assessment of how well each country fits its group. These variation measures are formally defined below.

- **Country dissimilarity** quantifies the overall dissimilarity for a country. For each country, it is calculated as the average of the Euclidean distances between that country and every other country.

  A high dissimilarity value indicates that the series for a country differs substantially from other countries, signalling potential outlier or unique behaviour, while a low dissimilarity value suggests the series closely resembles the behaviour of other countries in the dataset.

- **Group-wise dissimilarity** measures how dissimilar each country is from others in the same group. For each country, it is calculated as the average of the Euclidean distances between that country and every other country in its predefined group.

- **Silhouette width** assess how well the series for a country fits its group relative to other groups (Rousseeuw, 1987). For each country, it is calculated by comparing the average distance to countries within its group with the distance to the nearest neighbouring group. The silhouette width ranges from -1 to $+1$.

  A high silhouette width (close to $+1$) indicates that the country is well aligned within its predefined group with clear separation from other groups. A silhouette width close to 0 indicates a country lies near the boundary between two groups, and a silhouette width close to -1 denotes a country closer to a group other that its predefined group.

(2) **Trend and Shape Features**

In our exploratory analysis, trend and shape features describe how country-level data patterns evolve and whether they deviate from smooth, or consistent trajectories. These measures help us understand whether data changes follow steady patterns or more abrupt shifts. These are measured using trend strength, linearity, curvature and smoothness.

Trend strength, linearity, and curvature are part of the set of time series features proposed by Hyndman and Athanasopoulos (2021) and implemented in the `feasts` package (O'Hara-Wild et al., 2024b). These features are estimated using a time series decomposition Cleveland et al. (1990), which takes the form

$$y_t = T_t + S_t + R_t,$$

for seasonal data, and

$$y_t = T_t + R_t,$$

for non-seasonal data. T_t is the smoothed trend component computed using Friedman's SuperSmoother (Luedicke, 2015), implemented through the `stats`::*supsmu* function (R Core Team, 2024), and R_t is a remainder component.

- **Trend strength** measures the extent to which data follows a consistent pattern over time, which could be linear or curved relative to random fluctuations (Hyndman and Athanasopoulos, 2021). This measure, which ranges from 0 to 1, captures how pronounced and stable the overall trend is, distinguishing steady patterns from irregular noise. As described by Hyndman and Athanasopoulos (2021), a data series is considered strongly trended if $\text{Var}(R_t) < \text{Var}(T_t + R_t)$, and **weakly trended** if $\text{Var}(R_t) > \text{Var}(T_t + R_t)$. Therefore, the strength of trend F_T is defined as

$$F_T = \max\left(0,\ 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)}\right).$$

- **Linearity** measures how closely a series follows a straight-line pattern over time. It assesses whether changes occur consistently without sharp curves or bends. Following Hyndman and Athanasopoulos (2021), here it is defined as the coefficient of the linear term from a regression fitted to the trend component (T) of the decomposition, where the predictors are the first two orthogonal polynomials of the time index.

$$\text{T} = \beta_0 + \beta_1 \text{P}_1(t) + \beta_2 \text{P}_2(t)$$

where;
t is the time index.
β_1 is the linear term which quantifies linearity.
$\text{P}_1(t)$ is the first-degree orthogonal polynomial of t.

β_2 is the quadratic term which quantifies curvature.

$P_2(t)$ is the second-degree orthogonal polynomial of t.

A high positive linearity value (β_1) indicates a strong and pronounced increasing linear trend while a high negative linearity value indicates a pronounced decreasing linear trend.

A low positive or negative linearity value indicates a weak trend without a consistent directional pattern, while a β_1 value close to 0 suggests the absence of any linear trend.

- **Curvature** measures the extent to which a series deviates from a straight-line trend over time, capturing the presence of non-linear patterns; upward or downward bends. Curvature is quantified as β_2, the coefficient of the quadratic term in a polynomial regression fitted to the trend component of the decomposed series, where the predictors are the first two orthogonal polynomials of the time index (Hyndman and Athanasopoulos, 2021).

  A high positive curvature value indicates that the trend of the data series decreases and increases over time characterised by a U-shaped pattern while a high negative curvature value indicates trend increases to a peak and then declines over time characterised by an inverted U-shaped pattern.

  A low positive or negative curvature value suggests a weak U-shaped or inverted U-shaped pattern respectively, that is, the trend of the data series is not pronounced enough to be considered the dominant shape of the non-linear trend. When the curvature value is close 0, whether slightly positive or negative, it indicates that the trend follows a nearly linear trajectory, with minimal or no curvature.

- **Smoothness** measures how steadily the data for each country evolves over time. Here we quantify smoothness by the standard deviation of the lagged differences. It captures the degree of fluctuation in the trajectory of the data.

  Lower smoothness values indicate a smoother, more continuous shape, suggesting that the series evolves smoothly over time. Higher values suggest more irregular changes.

(3) **Sequential Temporal Features**

In our analysis, sequential temporal features capture the order and timing of changes in the data series, focusing on patterns that unfold with time rather than the overall trend. These measures reveal persistence, oscillations, and periods of stability over time within individual series. These are: number of crossing points, longest flat spot, and autocorrelation.

Number of crossing points and longest flat spot are features that have been proven useful in time series analysis (Hyndman and Athanasopoulos, 2021) and implemented in the `feasts` package O'Hara-Wild et al. (2024b).

- **Number of crossing points** count how many times a data series crosses below or above its median value, capturing the extent of fluctuation around the central tendency (Hyndman and Athanasopoulos, 2021).

  A higher number of crossing points indicates frequent median crossings, signifying fluctuations above and below the median over time and reduced stability in the data series of a country. A lower number suggests more sustained behaviour in one direction, where the series tends to remain on the same side of the median for extended periods.

- **Longest flat spot** measures the longest consecutive sequence in a series where data points lie within a single interval. This is calculated by dividing the data range into ten equal-sized intervals and identifying the maximum run length of consecutive observations that fall within the same interval (Hyndman and Athanasopoulos, 2021).

  High values indicate that the series, for a given country, remains relatively unchanged, reflecting strong temporal stability, while low values suggest that the series frequently moves between different intervals.

- **Autocorrelation** measures the correlation between each country data series and its one-period lagged values, capturing the strength and direction of association between lagged observations. The value of autocorrelation coefficient ranges from -1 to $+1$.

  A high positive autocorrelation (close to $+1$) indicates strong persistence, where increases (or decreases) tend to be followed by similar increases (or decreases), reflecting a smooth, consistent temporal pattern in the data series of a country. A high negative autocorrelation (close to -1)

indicates a strong inverse relationship, where increases tend to be followed by decreases (and vice versa), suggesting oscillating or alternating behaviour over time. Autocorrelation values near 0 indicate little or no linear dependence between consecutive observations.

The collection of diagnostic indices offers a structured summary of country-level panel data, capturing its key characteristics. While each measure provides insight on its own, their combined use with visualisation allows for a more complete view of how countries behave over time. We developed interactive visualisations to support the interpretation of individual metrics, combinations of metrics, or the full collection of the diagnostic indices.

3.3.2 Static and Interactive Visuals that Summarise Behavioural Patterns and Group-Based Features

We develop a set of visual tools that connect the collection of diagnostic indices to the underlying data series in different ways. These facilitate clearer interpretation of behavioural patterns of each series, within and across groups, while helping to identify unusual or distinctive cases.

(1) **The distribution plot** displays the distribution of metric values, either for all diagnostic indices or for a selected metric (one or multiple). It summarises the variation of metric values by showing the spread and shape of their distributions. There are two versions of the distribution plot: an ungrouped version, which shows the values of one or more metrics across all countries, and a grouped version, which displays the distribution within each level of a specified grouping variable.

(2) **The partition plot** presents metric values for individual countries grouped by a specified grouping variable. The metric value of each country is represented by a coloured bar ordered in descending order, while a lighter-shaded rectangular bar beneath indicates the average value of the metric at each group-level. This plot design is inspired by the silhouette plot (Rousseeuw, 1987), which quantifies and visualises how well each observation fits its assigned group. It also adapts the visual idea of presenting group averages as a shadow beneath individual values, as implemented in the `flexclust` package (Dolnicar et al., 2018).

(3) **The interactive data trajectory plot** displays the trajectory of the data series for each country. It supports two modes and each mode can be rendered in two versions: ungrouped and grouped versions. Both modes display all series as uniform line plots, while the second mode highlights countries that fall within a specified percentile of any chosen diagnostic metric values. In the ungrouped version, all countries are plotted together without any grouping structure, and when highlighting is enabled, series representing countries are highlighted based on whether their metric value exceed an overall threshold calculated by the specified percentile. In the grouped version, data are faceted by a chosen grouping variable (e.g., geographic or economic groupings); when highlighting is applied, countries are highlighted based on group-specific thresholds. In both cases, the highlighted series are emphasised using a colour gradient mapped to the metric values and hovering over the lines will display the correspondence country name and metric values so that the patterns identified by the metrics can be viewed directly within the temporal behaviour of each country for a clearer understanding of what each metric captures.

(4) **The interactive parallel coordinate plot** simultaneously displays all diagnostic metrics, with each metric represented as a vertical axis. Each country is shown as an interactive line that intersects all axes, with the position along the x-axis corresponding to the diagnostic indices. To ensure comparability across metrics, all values are normalised to a scale of 0 to 1. Exact metric values for each country and diagnostic indices are displayed when hovering over the lines at each axis. Similar to other plots, the interactive parallel coordinate plot supports both ungrouped and grouped versions. The ungrouped version displays all countries together as parallel lines, where lines are coloured by a grouping variable to enhance visual distinction. In the grouped version, parallel lines representing countries are faceted by a grouping variable, where countries in each group level are grouped together, and each group level are normalised to a scale of 0 to 1 to facilitate within-group comparison. A consistent colour scheme is applied to both views, with countries in the same group distinguished by unique colours.

(5) **The interactive link-view of metrics and series plot** displays an interactive visualisation that connects diagnostic indices values with their corre-

sponding series trajectories. One panel shows a scatterplot of two selected diagnostic indices (e.g., linearity and curvature), while the other shows the line plot of the data series for each country. Like the other plots, it also supports both ungrouped and grouped versions. In the ungrouped version, all countries appear in a single scatterplot and line plot panel. In the grouped version, countries are separated into facets based on a specified grouping variable (e.g., geographic or economic groupings), allowing both the scatterplot and time series panels to reflect group-specific structures. Hovering over or selecting a data point in the scatterplot highlights the corresponding trajectory with the country name, and vice versa.

(6) **The missingness plot** displays a visual overview of missing data patterns across countries and years. Countries are grouped by the levels of a specified grouping variable, and the presence or absence of data is indicated by colour: missing entries are shown in black, while available data points are displayed in a light grey colour. This plot offers a structured summary of data availability, making it easy to spot countries with complete data gaps and years with widespread missingness.

All visualisations are built using the `ggplot2` (Wickham, 2016) framework, with additional support from the `ggiraph` (Gohel and Skintzos, 2025) and `ggdist` (Kay, 2024) packages. The ggiraph package enables interactivity, while the ggdist package is used to present the distributions of the diagnostic metrics.

3.4 Package Structure

Our `wdiexplorer` package provides a set of functions designed to facilitate the efficient exploration and visualisation of one indicator at a time of the World Development Indicators (WDI) country-level panel data. It focuses on identifying patterns, including temporal, shape-based features, outliers, and structural variations within the data, while enabling comparisons across countries and within predefined groups. The package functions, listed in Table 3.1, are categorised into three groups that reflect the stages of the exploration workflow, namely data sourcing and assessing missing data, computing diagnostic indices, and generating both static and interactive visualisations. Examples of the usage of all of these functions are given in Section 3.5.

Table 3.1: An overview of the wdiexplorer function names and their functionalities.

Function name	Functionality
`get_wdi_data`	Sources and locally stores data from the WDI database.
`plot_missing`	Visualises missingness in the data.
`get_valid_data`	Extracts valid observations and reports gaps in country and year.
`compute_dissimilarity`	Computes Euclidean distance between pairwise countries.
`compute_variation`	Computes average country dissimilarities, group-wise dissimilarities, and silhouette widths.
`compute_trend_shape_features`	Computes trend strength, linearity, curvature, and smoothness.
`compute_temporal_features`	Computes number of crossing points, longest flat spot, and autocorrelation.
`compute_diagnostic_indices`	Computes all ten diagnostic indices collectively.
`add_group_info`	Adds grouping information from the WDI data to the computed diagnostic indices.
`plot_metric_distribution`	Visualises grouped or ungrouped distribution plots for all or selected diagnostic indices.
`plot_metric_partition`	Visualises metric values for each country, partitioned by the specified grouping variable.
`plot_data_trajectories`	Grouped or ungrouped interactive line plots of the series with or without metric-based highlighting.
`plot_parallel_coords`	Parallel coordinate plot of all the diagnostic indices.
`plot_metric_linkview`	Interactive linked view of the relationship between two metrics and their associated series trajectories by country.

3.4.1 Data functions

The `get_wdi_data` function sources and locally stores WDI indicators datasets one at a time, leveraging the `WDI` package. During this process, the indicator variable name is stored as a dataset attribute, which subsequent package functions use by default as the `index` argument ensuring consistency while avoiding the need to

specify the variable with the country-year data points in any WDI data repeatedly.

Country-level panel data often include missing observations and irregular time intervals; to address this, we extend the functionality of the `vis_miss` function from the `naniar` (Tierney and Cook, 2023) package by introducing the `plot_missing` function that provides a visual summary of data availability and missingness across countries and years. Complementing this, the `get_valid_data` function inspects the dataset to identify countries and years with no valid data points. A data point is considered invalid if its indicator value is missing, that is NA, NaN. The function returns a valid dataset with a printed summary of excluded countries and years.

3.4.2 Diagnostic indices functions

The functions that compute the collection of diagnostic indices, introduced in Section 3.3.1, are organised according to their respective categories: variation features, trend and shape features and sequential temporal features. Specifically, the `compute_dissimilarity` function returns the dissimilarity matrix quantifying the Euclidean distance between countries computed with the `daisy()` function which handles missing values. This dissimilarity matrix is used as an input to `compute_variation`, which computes the average dissimilarity for each country, the average dissimilarities between a country and every other countries in its predefined group, and the silhouette width.

The function `compute_trend_shape_features` quantifies the directional patterns and shape of each series with trend strength, linearity, curvature and smoothness measures. The `compute_temporal_features` captures sequential temporal features including the number of crossing points, longest flat spot and autocorrelation.

`compute_diagnostic_indices` computes all diagnostic indices collectively.

`add_group_info` function adds the grouping information of the WDI data to the output of any function that computes diagnostic indices. This collection of diagnostic indices offers statistical measures that characterise the behavioural patterns as well as group-based patterns of country-level panel data.

3.4.3 Visualisation functions

The remaining functions focus on visualising the computed diagnostic indices, providing both static and interactive visualisation functions. The `plot_metric_distribution`

function generates a plot that displays the distribution of diagnostic indices. The function is flexible: it can create either an ungrouped or grouped plot version of the distribution for the values of a single metric, a selected subset of metric, or all the diagnostic indices collectively. When visualising a single metric, the function produces a straightforward distribution plot, while multiple metrics are displayed using faceting, each metric appears in its own panel, enabling clear visual comparison of their shapes and spreads. Within each panel, individual country-level metric values are represented by dots. See Figure 3.4 for an example.

The `plot_metric_partition` function creates a layered bar plot that displays country-level metric values, grouped by a specified variable, with individual bars overlaid on lighter bars representing group-level averages. Each group level is distinguished by a unique colour, with the associated countries displayed accordingly, thereby facilitating comparison both within and across groups. Figures 3.6 and 3.7 are examples of a partition plot with the silhouette width metric.

Both `plot_metric_distribution` and `plot_metric_partition` functions generate static plots because the dots in the distribution plot are not rendered interactively, and the bars in the partition plot are sufficient to convey the metric values of each country, as tracing them toward the x-axis indicates their respective values.

Additionally, the package includes three interactive visualisation functions. The `plot_data_trajectories` function creates interactive line plots of the data series over time. It supports two modes: one that displays all series uniformly, and another that also displays the series uniformly while highlighting countries that fall within a specified threshold of the metric values. In both cases, the function uses interactive tooltips from the `ggiraph` package to display the corresponding metric values. Each mode can be rendered in two versions: ungrouped and grouped. In the ungrouped version, all countries are plotted together without any grouping structure, and when highlighting is enabled, it is based on a global percentile threshold derived from the overall metric distribution. In the grouped version, countries are grouped by a specified variable; when highlighting is applied, the threshold is calculated within each group separately, allowing the plot to reflect group-level variation in the metric rather than the overall distribution. Figure 3.8 is an example of ungrouped data trajectories plot.

The `plot_parallel_coords` function generates two versions of interactive parallel coordinates plot, where each line corresponds to a country and spans the ten diagnostic indices. Tooltips appear when hovering over a line across the x-axis,

displaying the corresponding country name, metric and its value for that country, facilitating a detailed comparison across all indices. The ungrouped version colour the lines by the specified grouping variable, while the grouped version also colour the lines by the specified grouping variable and facet the parallel coordinates plots by the specified grouping variable. Figure 3.12 is an example of a parallel coordinate plot.

The `plot_metric_linkview` function offers an interactive dual-view display that links the relationship between two metric values to their corresponding series line plots, enabling exploration of how they relate to the overall data trajectories. See Figure 3.14 for an example.

The majority of these functions accept an `index` argument that, by default, uses the first variable saved as the WDI dataset attribute. A consistent colour scheme is used across all visualisation functions to distinguish countries within the same group, facilitating comparison within and between groups. Additionally, we adopt a clear and descriptive naming format across all functions, aligned with their intended purpose and the overall structure of the package. These choices improve usability and support an intuitive workflow.

3.5 Package Workflow

In this section, we demonstrate the `wdiexplorer` R package workflow categorised into three main stages, using the $PM_{2.5}$ air pollution data as an example. This data is derived from the GBD Study 2021, coordinated by the Institute for Health Metrics and Evaluation providing air pollution exposure estimates for nitrogen dioxide pollution, ozone pollution, ambient particulate matter pollution, and household air pollution. The $PM_{2.5}$ air pollution dataset contains mean annual exposure levels to ambient $PM_{2.5}$, measured in micrograms per cubic meter ($\mu g/m^3$).

3.5.1 Stage 1: Data Sourcing and Preparation

This initial stage of the workflow uses three core functions:
`get_wdi_data`, `plot_missing` and `get_valid_data`.

The `get_wdi_data` function sources the WDI indicator data of interest, taking a single argument `indicator`, which should be a valid WDI indicator code (e.g.,

$PM_{2.5}$ air pollution data with WDI indicator code: "EN.ATM.PM25.MC.M3"). To find any indicator code, users can employ the `WDI::WDISearch()` function with a relevant search keyword as its argument as illustrated below.

```
WDI::WDIsearch("air pollution")
```

```
pm_data <- get_wdi_data(indicator = "EN.ATM.PM25.MC.M3")
```

The dataset for any specified indicator code is a data frame with 13 variables and is stored locally in a "wdi_data" folder for easy future access. The 13 variables of any WDI indicator dataset are described in Table 3.2 below.

Table 3.2: WDI variable names and descriptions.

Variable name	Description
Country	Country name.
iso2c	2-letter ISO country code.
iso3c	3-letter ISO country code.
year	Calendar year representing the time index of the observation.
indicator code	The variable containing values for the specified indicator code. "In this package, this serves as the index variable".
status	Describes the nature of the reported value, whether it is actual, estimated, or forecasted.
lastupdated	Timestamp that indicates the most recent update of the indicator values.
region	Geographical region.
capital	Name of the capital city.
longitude	Geographic coordinate that measures how far east or west a country is from the Prime Meridian.
latitude	Geographic coordinate that measures how far north or south a country is from the Equator.
income	World Bank income classification.
lending	World Bank lending classification.

```
tibble::glimpse(pm_data)

    Downloading WDI indicator:
    EN.ATM.PM25.MC.M3 data using the WDI R package.
Rows: 13,910
Columns: 13
$ country           <chr> "Afghanistan", "Afghanistan", "Afghanistan"...
$ iso2c             <chr> "AF", "AF", "AF", "AF", "AF", "AF", "AF", "...
$ iso3c             <chr> "AFG", "AFG", "AFG", "AFG", "AFG", "AFG", "...
$ year              <int> 1964, 1965, 2007, 1963, 1970, 1969, 1968, 1...
$ EN.ATM.PM25.MC.M3 <dbl> NA, NA, 58.08379, NA, NA, NA, NA, NA, NA, 5...
$ status            <chr> "", "", "", "", "", "", "", "", "", "",...
$ lastupdated       <chr> "2025-07-01", "2025-07-01", "2025-07-01", "...
$ region            <chr> "South Asia", "South Asia", "South Asia", "...
$ capital           <chr> "Kabul", "Kabul", "Kabul", "Kabul", "Kabul"...
$ longitude         <chr> "69.1761", "69.1761", "69.1761", "69.1761",...
$ latitude          <chr> "34.5228", "34.5228", "34.5228", "34.5228",...
$ income            <chr> "Low income", "Low income", "Low income", "...
$ lending           <chr> "IDA", "IDA", "IDA", "IDA", "IDA", "IDA", "...
```

Once the data is sourced, an important initial step is to highlight missing entries across countries and time periods. To facilitate this, the `plot_missing` function takes two arguments: a dataset containing data for a selected WDI indicator, and a predefined grouping variable within the dataset. It also includes an additional argument `index`, which defaults to `NULL` and uses the first variable saved as the WDI dataset attribute. The function produces a missingness plot that provides a structured overview of missing data and shows its distribution over time, across countries, and by the specified grouping variable.

The missingness plots in Figures 3.1, 3.2, 3.3 show that no data are present from 1960 to 1989 and from 2021 to 2024. During the period 1990 to 2020, data are available across all countries in the Middle East & North Africa, South Asia, and Sub-Saharan Africa. In contrast, several countries in other regions have no valid data recorded. Specifically, four countries in East Asia & Pacific, six countries in Europe & Central Asia, and seven countries in Latin America & the Caribbean. In total, 17 countries across these three regions have no valid data between 1990 and

2020. These patterns of missingness are clearly captured by the plot, which provides a structured overview across time, countries, and predefined groups (geographic or economic groupings).

```
plot_missing(pm_data, group_var = "region")
```

The World Bank grouped countries into seven regions. To improve readability and legibility, the missingness plots for the 185 countries were displayed across three separate plots Figures 3.1, 3.2, 3.3: the first two plots each contain two regions, while the third plot contains the remaining three regions. Note: the `plot_missing` function presents the missingness patterns for all countries in a single plot; however, for presentation purposes, it was divided into three separate plots.

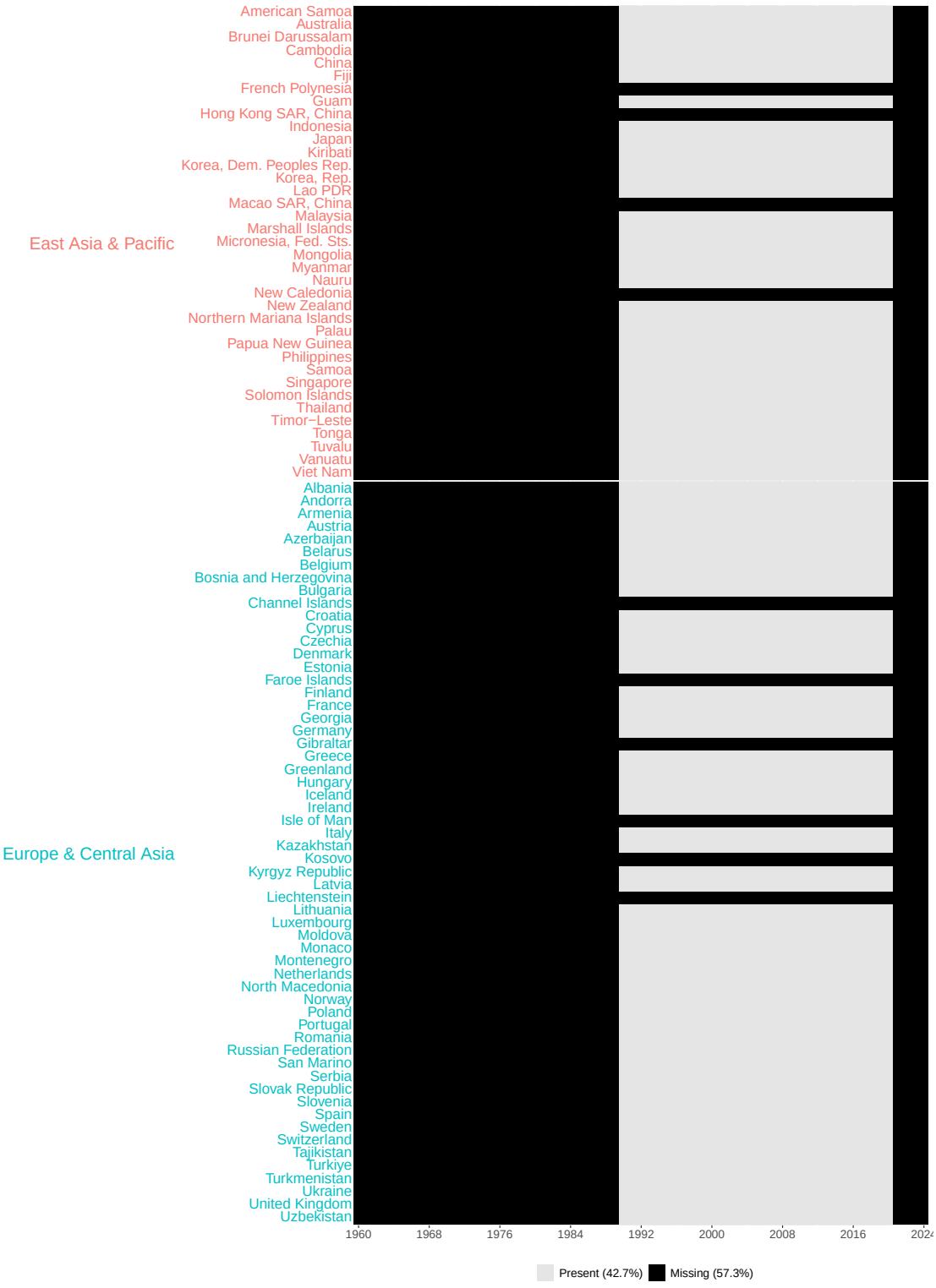

Figure 3.1: Missingness plot of the first two regions, providing information about the years and countries with missing entries and the overall percentages of missing and present data. It also shows that no data points are available during years 1960 to 1989 and 2021 to 2024.

Figure 3.2: Missingness plot of Latin America & Caribbean, and Middle East & North Africa regions, providing information about the years and countries with missing entries and the overall percentages of missing and present data. It also shows that no data points are available during years 1960 to 1989 and 2021 to 2024.

Figure 3.3: Missingness plot of North America, South Asia, and Sub-Saharan Africa, providing information about the years and countries with missing entries and the overall percentages of missing and present data. It also shows that no data points are available during years 1960 to 1989 and 2021 to 2024.

To complement this visual summary, we introduce a second step: calculating the total number of missing entries per country. This quantitative assessment mirrors the information shown in the missingness plot and helps identify countries with particularly high levels of missing data within their group. The following code demonstrates this using our illustrative dataset.

```
index <- "EN.ATM.PM25.MC.M3"

pm_data |>
  dplyr::select(country, region, year, tidyselect::all_of(index)) |>
  dplyr::group_by(region, country) |>
  naniar::miss_var_summary() |>
  dplyr::filter(variable == index) |>
  dplyr::arrange(desc(n_miss))

# A tibble: 215 × 5
# Groups:   region, country [215]
   country                region         variable n_miss pct_miss
   <chr>                  <chr>          <chr>     <int>    <num>
 1 Aruba                  Latin Americ... EN.ATM....     65      100
 2 British Virgin Islands Latin Americ... EN.ATM....     65      100
 3 Cayman Islands         Latin Americ... EN.ATM....     65      100
 4 Channel Islands        Europe & Cen... EN.ATM....     65      100
 5 Curacao                Latin Americ... EN.ATM....     65      100
 6 Faroe Islands          Europe & Cen... EN.ATM....     65      100
 7 French Polynesia       East Asia & ... EN.ATM....     65      100
 8 Gibraltar              Europe & Cen... EN.ATM....     65      100
 9 Hong Kong SAR, China   East Asia & ... EN.ATM....     65      100
10 Isle of Man            Europe & Cen... EN.ATM....     65      100
# 205 more rows
```

In addition, the `wdiexplorer` package provides the `get_valid_data` function, which reports countries with no data points as well as years for which no data are available, and returns a tibble with the valid data for the provided WDI indicator dataset. The function takes one argument, a dataset containing data for a selected WDI indicator.

```
get_valid_data(pm_data)

The 17 countries listed below had no available data and were excluded:
 Aruba
- British Virgin Islands
- Cayman Islands
- Channel Islands
- Curacao
- Faroe Islands
- French Polynesia
- Gibraltar
- Hong Kong SAR, China
- Isle of Man
- Kosovo
- Liechtenstein
- Macao SAR, China
- New Caledonia
- Sint Maarten (Dutch part)
- St. Martin (French part)
- Turks and Caicos Islands

The 34 year(s) listed below had no available data and were excluded:
 1960,  1961,  1962,  1963,  1964,  1965,
 1966,  1967,  1968,  1969,  1970,  1971,
 1972,  1973,  1974,  1975,  1976,  1977,
 1978,  1979,  1980,  1981,  1982,  1983,
 1984,  1985,  1986,  1987,  1988,  1989,
 2021,  2022,  2023,  2024
```

The `get_valid_data` function is primarily used to investigate missing entries and identify countries excluded from the indicator data. This function is executed internally by all other functions in the `wdiexplorer` workflow to filter valid data from the specified indicator data.

3.5.2 Stage 2: Diagnostic Indices

This second stage of the workflow focuses on calculating the diagnostic indices introduced in Section 3.3.1. They measure variation, trend and shape features, as well as sequential temporal characteristics. There are three functions in the `wdiexplorer` package which support the computation of these indices.

Variation Features

To measure variation, the function `compute_dissimilarity` takes one main argument, a dataset of any WDI indicator, and returns a matrix of dissimilarity values between country pairs. The function also includes two additional arguments with default values: `index`, which defaults to `NULL` and uses the first variable saved as the WDI dataset attribute; and `metric`, which specifies the dissimilarity calculation method and defaults to Euclidean distance.

The function `compute_variation` accepts two main arguments: a dataset of any WDI indicator and a grouping variable `group_var`. It also includes an optional dissimilarity matrix argument, `diss_matrix` (defaulting to the output of `compute_dissimilarity`. The `compute_variation` function returns a data frame with `country`, `country_avg_dist`, average distance between each country and all other countries in the data set, `within_group_avg_dist`, average distance between each country and all other countries within its predefined group, and `sil_width`, silhouette widths which measure how well each country fits within its group compared to other groups. The code below demonstrates the use of these functions:

Users can compute the dissimilarity matrix separately and pass it directly as the `diss_matrix` argument into the `compute_variation` function as demonstrated below or allow the function to compute it internally by specifying only the two main arguments:

```
pm_diss_mat <- compute_dissimilarity(pm_data)

pm_variation <- compute_variation(
                    pm_data,
                    diss_matrix = pm_diss_mat,
                    group_var = "region"
        )
```

The output `pm_variation` enables the exploration of computed variation features. It facilitates the identification of the most distinctive countries, the evaluation of within-group differences, and the analysis of how closely aligned countries within a group are compared to those in other groups. However, these measures are not always intuitive to interpret on their own; they are best understood in conjunction with the accompanying line plot, which displays the behavioural patterns over time, offering meaningful context for comparing the trajectories of standout countries to the rest.

```
# country dissimilarity average
pm_variation |>
        dplyr::arrange(desc(country_avg_dist)) |>
        dplyr::slice_head(n = 3)
```

```
# A tibble: 3 × 5
  country    group       country_avg_dist within_group_avg_dist sil_width
  <chr>      <chr>                  <dbl>               <dbl>      <dbl>
1 Qatar      Middle East ...         498.                410.     -0.168
2 Niger      Sub-Saharan ...         479.                408.     -0.207
3 Mauritania Sub-Saharan ...         415.                345.     -0.243
```

```
# within group dissimilarity average
 pm_variation |>
        dplyr::arrange(desc(within_group_avg_dist)) |>
        dplyr::slice_head(n = 3)
```

```
# A tibble: 3 × 5
  country    group       country_avg_dist within_group_avg_dist sil_width
  <chr>      <chr>                  <dbl>               <dbl>      <dbl>
1 Qatar      Middle East ...         498.                410.     -0.168
2 Niger      Sub-Saharan ...         479.                408.     -0.207
3 Mauritania Sub-Saharan ...         415.                345.     -0.243
```

The results above show that Qatar has the highest overall average dissimilarity, followed by Niger and Mauritania. These countries, Qatar, Niger, and Mauritania,

not only rank highest in overall average dissimilarity but also exhibit the highest dissimilarity relative to countries in their respective region groups.

```
pm_variation |>
    dplyr::arrange(desc(sil_width)) |>
    dplyr::slice_head(n = 3)
```

```
# A tibble: 3 × 5
country       group    country_avg_dist  within_group_avg_dist  sil_width
  <chr>        <chr>           <dbl>            <dbl>               <dbl>
1 Canada        North Ame...   151.              13.7               0.836
2 Bermuda       North Ame...   155.              17.2               0.798
3 United States North Ame...   138.              23.0               0.695
```

As shown in the `pm_variation` result, Canada, Bermuda, and the United States (the only countries from the North America region group with available $PM_{2.5}$ air pollution data) exhibit the highest silhouette widths, indicating that these countries have similar $PM_{2.5}$ trajectories, which are dissimilar to those in other region groups.

Trend and Shape Features

To examine trend and shape features, we implement `compute_trend_shape_features`, which takes as the main argument a dataset of any WDI indicator. An additional argument `index` specifies the indicator name. If `index` is `NULL` then the first variable saved as the WDI dataset attribute is used. The function returns a data frame with one row per country and columns for the measures `trend strength`, `linearity`, `curvature`, and `smoothness`.

```
pm_trend_shape <- compute_trend_shape_features(pm_data)
```

The `pm_trend_shape` output enables the exploration of the computed trend and shape features.

```
pm_trend_shape |>
        dplyr::arrange(desc(trend_strength)) |>
    dplyr::slice_head(n = 3)
```

```
# A tibble: 3 × 5
  country        trend_strength linearity curvature smoothness
  <chr>                   <dbl>     <dbl>     <dbl>      <dbl>
1 Ukraine                 0.996     -32.4     2.69       0.547
2 Moldova                 0.995     -34.4     2.65       0.666
3 United Kingdom          0.995     -16.3    -0.217      0.329
```

The output above shows that Ukraine, Moldova, and the United Kingdom are the three countries with the strongest trend strength. In this context, trend strength measures the extent to which data follows a consistent pattern over time, whether linear or curved.

```
pm_trend_shape |>
     dplyr::arrange(desc(linearity)) |>
     dplyr::slice(c(1, 2, dplyr::n(), dplyr::n() - 1))
```

```
# A tibble: 4 × 5
  country      trend_strength linearity curvature smoothness
  <chr>                 <dbl>     <dbl>     <dbl>      <dbl>
1 Mongolia              0.944      27.2      4.47       1.95
2 Saudi Arabia          0.810      25.5    -11.7        4.34
3 Bolivia               0.995    -116.       2.46       2.08
4 Peru                  0.991    -104.     -10.2        2.49
```

The output above lists countries with the strongest positive and negative linear trend as represented by the `linearity` column, Mongolia and Saudi Arabia exhibit the highest positive linearity values, and Bolivia and Peru are countries with the highest negative linearity. A more detailed interpretation of these results is presented in Subsection 3.5.3, where the metric values are explored in conjunction with the trajectories of each country's data series.

The next code segment identifies countries with extreme curvature values.

```
pm_trend_shape |>
    dplyr::arrange(desc(curvature)) |>
     dplyr::slice(c(1, 2, dplyr::n(), dplyr::n() - 1))
```

```
# A tibble: 4 × 5
  country   trend_strength linearity curvature smoothness
  <chr>              <dbl>     <dbl>     <dbl>      <dbl>
1 Singapore          0.970    -21.1      14.4       1.41
2 Senegal            0.575    -13.7      10.1       6.19
3 Kuwait             0.762     13.7     -17.2       4.92
4 Libya              0.713      6.78    -15.2       3.94
```

From the above output, Singapore and Senegal exhibit the most positive curvature values, indicating that their series follow a U-shaped pattern. Conversely, Kuwait and Libya exhibit the most negative curvature values, indicating a pronounced inverted U-shaped pattern.

We proceeded to identify countries with the highest and lowest smoothness values.

```
pm_trend_shape |>
    dplyr::arrange(desc(smoothness)) |>
    dplyr::slice(c(1, dplyr::n()))
```

```
# A tibble: 2 × 5
  country trend_strength linearity curvature smoothness
  <chr>            <dbl>     <dbl>     <dbl>      <dbl>
1 Niger            0.242    13.9      -9.88       8.54
2 Tuvalu           0.906     0.631    -0.596      0.144
```

This output shows that Niger, a country in Middle East & North Africa region, emerges as the country with the highest smoothness value. Tuvalu, located in the East Asia & Pacific region, has the lowest smoothness value.

Sequential Temporal Features

Lastly, to measure the sequential temporal features of the data series, we implement a function named `compute_temporal_features`, which takes the same arguments

as the other functions that calculate diagnostic indices. The function returns a data frame where each row represents a country, and columns for the `country` name, and measures of `number of crossing points`, `longest flat spot`, and `autocorrelation`.

```
pm_temporal <- compute_temporal_features(pm_data)
```

We can use `pm_temporal` to explore the computed sequential temporal features.

```
pm_temporal |>
   dplyr::arrange(desc(crossing_points)) |>
   dplyr::slice(c(1, 2, dplyr::n(), dplyr::n() - 1))
```

```
# A tibble: 4 × 4
  country       crossing_points flat_spot   acf
  <chr>                   <int>     <int> <dbl>
1 Gabon                      10         4 0.338
2 Tunisia                    10         5 0.760
3 Venezuela, RB               1        13 0.942
4 Uruguay                     1        11 0.977
```

The code segment above shows that Gabon and Tunisia have the highest number of crossing points. In contrast, 54 countries have only one crossing point, indicating that these countries cross above or below their median value just once.

We proceeded to identify countries with the highest and the lowest number of flat spot:

```
pm_temporal |>
   dplyr::arrange(desc(flat_spot)) |>
   dplyr::slice(c(1:3, (dplyr::n() - 2):dplyr::n()))
```

```
# A tibble: 6 × 4
  country              crossing_points flat_spot      acf
  <chr>                          <int>     <int>    <dbl>
1 Kyrgyz Republic                    3        18  0.909
2 United Arab Emirates               9        18 -0.00320
3 Armenia                            1        17  0.934
4 Jordan                             3         3  0.846
5 Rwanda                             7         3  0.539
6 West Bank and Gaza                 8         3  0.608
```

We see from this output that Kyrgyz Republic and United Arab Emirates have the longest flat spots, characterised by long consecutive periods during which their data series remain within an interval. In contrast, Jordan, Rwanda, and West Bank and Gaza exhibit the shortest consecutive periods where their series remain within a specified interval.

The next code segment investigates the acf measures in `pm_temporal`:

```
pm_temporal |>
     dplyr::arrange(desc(acf)) |>
     dplyr::slice(c(1, dplyr::n()))
```

```
# A tibble: 2 × 4
  country              crossing_points flat_spot      acf
  <chr>                          <int>     <int>    <dbl>
1 Ukraine                            1         5  0.996
2 United Arab Emirates               9        18 -0.00320
```

The above output shows that Ukraine has the highest autocorrelation value. United Arab Emirates has zero autocorrelation value. United Arab Emirates has the zero autocorrelation value and is also one of the countries with the longest flat spot. This suggests that although the data series in United Arab Emirates contains extended periods of minimal change reflected in its high number of longest flat spot, the

zero autocorrelation indicates an absence of dependence between successive time points, even when its values remain within the same interval for the longest period. This shows that the diagnostic measures are often not informative enough when interpreted in isolation; they become more insightful when combined with their corresponding series trajectories as shown in Subsection 3.5.3.

The `compute_diagnostic_indices` function returns all diagnostic indices collectively in a single data frame. The function takes two arguments: a dataset of any WDI indicator data and a grouping variable `group_var` and returns a data frame with: `country`; `country_avg_dist`, `within_group_avg_dist`, `sil_width`, `trend strength`, `linearity`, `curvature`, `smoothness`, `number of crossing points`, `longest flat spot`, and `autocorrelation` measures.

```
pm_diagnostic_metrics <- compute_diagnostic_indices(
                pm_data, group_var = "region"
            )
```

This `pm_diagnostic_metrics` output can be passed directly to the plot functions to generate both static and interactive visuals of the `wdiexplorer` package.

The `add_group_info` function appends the predefined grouping information from the WDI dataset to the data frame of any computed diagnostic function output. The function takes two arguments: a data frame with the calculated diagnostic indices `metric_summary`; and a dataset of any WDI indicator data. This facilitates the direct use of the resulting data frame with the plot functions to create group versions of the visualisations.

```
pm_diagnostic_metrics_group <- add_group_info(
                    metric_summary = pm_diagnostic_metrics,
                    pm_data
            )
```

3.5.3 Stage 3: Static and Interactive Visualisations

The third stage of the workflow utilises visual summaries to detect potentially interesting features within panel data. Our package offers five core visualisation functions, two static plot functions: `plot_metric_distribution`, and `plot_metric_partition` and three interactive plot functions: `plot_data_trajectories`, `plot_parallel_coords`, and `plot_metric_linkview`, all built on the `ggplot2` (Wickham, 2016) framework

and using `ggiraph` (Gohel and Skintzos, 2025) for interactivity, `ggdist` (Kay, 2024) for the distribution plots and `scales` (Wickham et al., 2025) for managing predefined group colour palettes. Each of these functions is demonstrated below.

Distribution Plot

The `plot_metric_distribution` function takes two main arguments: a data frame containing the computed diagnostic indices and the predefined grouping information `metric_summary`; and a variable, `colour_var` whose levels are mapped to distinct colours in the resulting dot plot. An optional argument, `group_var`, controls whether the distribution is grouped or ungrouped. To generate the distribution plot for specific metrics, users can pass the variable names to the `metric_var` argument.

The code below demonstrates how to use the `plot_metric_distribution` function to generate ungrouped distribution plots for all diagnostic indices.

```
plot_metric_distribution(
      metric_summary = pm_diagnostic_metrics_group,
      colour_var = "region"
      )
```

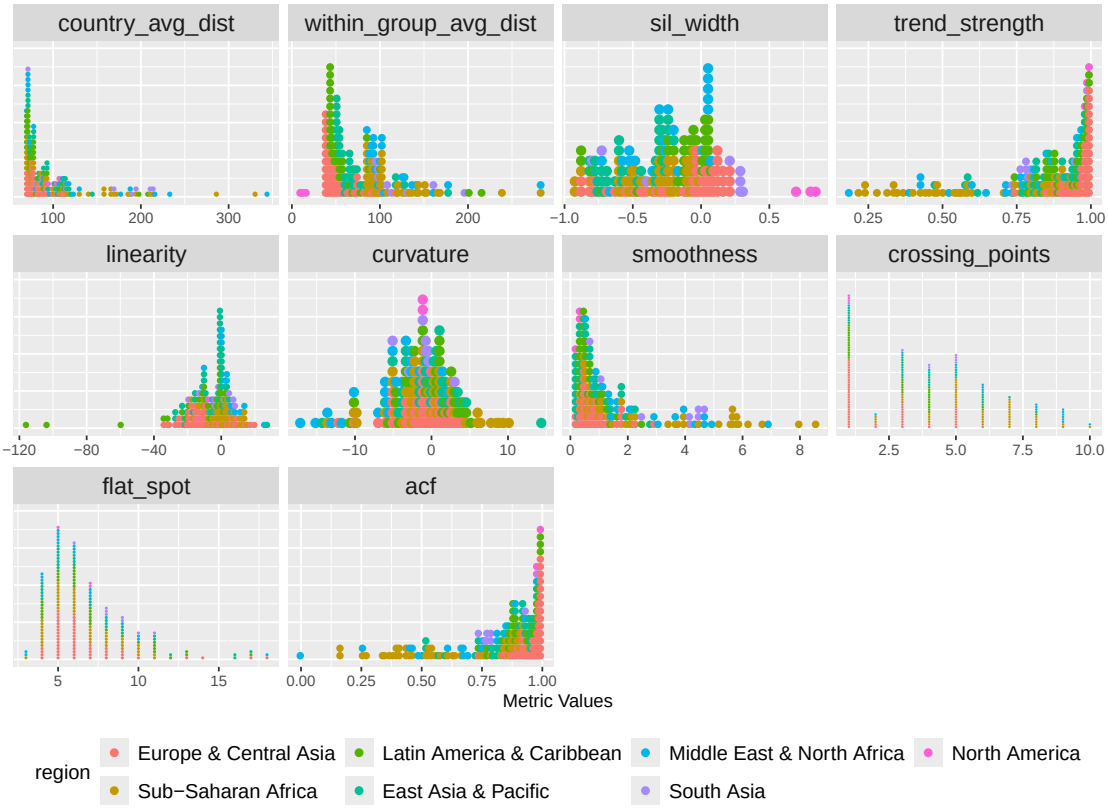

Figure 3.4: Distribution of diagnostic indices where each panel represents a different metric. It shows the spread of the metric values across countries, with each dot representing a country and coloured by region. Countries in the North America region stand out with the lowest within-group average dissimilarity and the highest silhouette width values.

Figure 3.4 shows the distribution of all diagnostic indices using dotplots. Each metric is presented in a separate panel, with each dots showing the per country metric value with dots coloured by region. The figure reveals distinct distributional patterns across the indices. For instance, country average dissimilarity and smoothness measures are right skewed with most countries having low values. These indicate that the majority of countries differ only minimally from one another and tend to follow smooth, gradual changes over time (based on the smoothness measures). In contrast, trend strength and autocorrelation (acf) are left skewed, with most countries showing high metrics values. These reveal that, in majority of the countries, their data series exhibit persistence trends which could be linear or curved and high correlation between successive time points. However, linearity and curvature are centred around zero. Linearity (β_1) value close to 0 suggests the absence of

any linear trend and curvature (β_2) value close to 0 indicates that the trend follows a nearly linear trajectory, with minimal or no curvature. Although the plot is not interactive and the countries corresponding to the dots centred around zero cannot be identified, a literal interpretation suggests that the trends in some countries are neither predominantly linear nor strongly curved.

The Figure 3.4 also highlights that the three countries in the North America region shown in magenta stand out with the lowest within-group average distance and the highest silhouette width values. The low within-group average distance indicates minimal variation among the temporal patterns of these countries, suggesting that their $PM_{2.5}$ air pollution trends are highly similar. The high silhouette width further confirms this, as it measures how well each country fits within its group relative to other groups, higher values indicate that these countries are closely aligned to their predefined group and distinct from others.

The code below demonstrates how to use the `plot_metric_distribution` function to generate grouped distribution plots for all diagnostic indices.

```
plot_metric_distribution(
      metric_summary = pm_diagnostic_metrics_group,
      colour_var = "region",
      group_var = "region"
      )
```

Figure 3.5: Distribution of diagnostic indices grouped by region. Each panel displays a metric, with countries organised by region to facilitate within and between group comparisons. The plot reveals region-specific patterns and outliers. Sub-Saharan Africa and East Asia & Pacific regions show wider spread while North America region are closely aligned.

Figure 3.5 shows the distribution of diagnostic indices grouped by region. As in the ungrouped distribution plot in Figure 3.4, each panel represents a different metric. However, while the ungrouped version presents all countries together as individual dots in a single distribution per metric, the grouped version organises countries by region, making it easier to compare both within and between group metric values across regions. Across all regions, the country average dissimilarity metric is consistently right-skewed, though Sub-Saharan Africa and Middle East & North Africa contain notable outliers that deviate substantially from other countries in their region. A similar pattern is observed in their within-group average dissimilarity metric. In the silhouette width panel, All countries in Latin America

& Caribbean and Sub-Saharan Africa have negative values, suggesting that their temporal patterns may be more aligned with countries outside their assigned groups while the three countries in the North America region have the highest silhouette width values. The North America region exhibit closely aligned metric values across all diagnostic indices with minimal differences. This indicates a highly knitted region with similar temporal behaviour.

Figure 3.5 also highlights outlier countries in South Asia region where some of the countries exhibit negative silhouette width far off from the behaviour in other countries. Outliers are easily identified in most of the regions across metrics in the grouped distribution plot. For example, East Asia & Pacific has the country with the highest positive curvature value while Middle East & North Africa has countries with the most extreme negative curvature values. In smoothness metric, Sub-Saharan Africa has the highest smoothness values. Sub-Saharan Africa and Middle East & North Africa exhibit broader spreads across metrics particularly in country average dissimilarity, trend strength, curvature, smoothness and autocorrelation (acf), reflecting substantial heterogeneity within the data series in these regions.

Although linearity and curvature remain centred around zero in most regions as previously seen in Figure 3.4, the ungrouped version and the grouped version in Figure 3.5 reveals region-specific behaviours. For instance, Latin America & Caribbean appears more left-skewed in linearity, indicating a pronounced downward trend in most of the countries.

Partition Plot

The `plot_metric_partition` function takes three arguments: `metric_summary`, a data frame with the calculated diagnostic indices and the grouping information, `metric_var`, a variable within the data frame that contains the metric values, and `group_var` naming the grouping variable. The `metric_summary` is the output of any diagnostic indices functions merged with the grouping information from the WDI data. This integration is performed using the `add_group_info` function, which appends the relevant group information of each country to the metric summary. The following code demonstrates `plot_metric_partition` using the silhouette width measures of the $PM_{2.5}$ air pollution data.

```
plot_metric_partition(
          metric_summary = pm_diagnostic_metrics_group,
          metric_var = "sil_width",
          group_var = "region"
 )
```

To improve legibility, the silhouette width partition plot is presented across Figures 3.6 and 3.7.

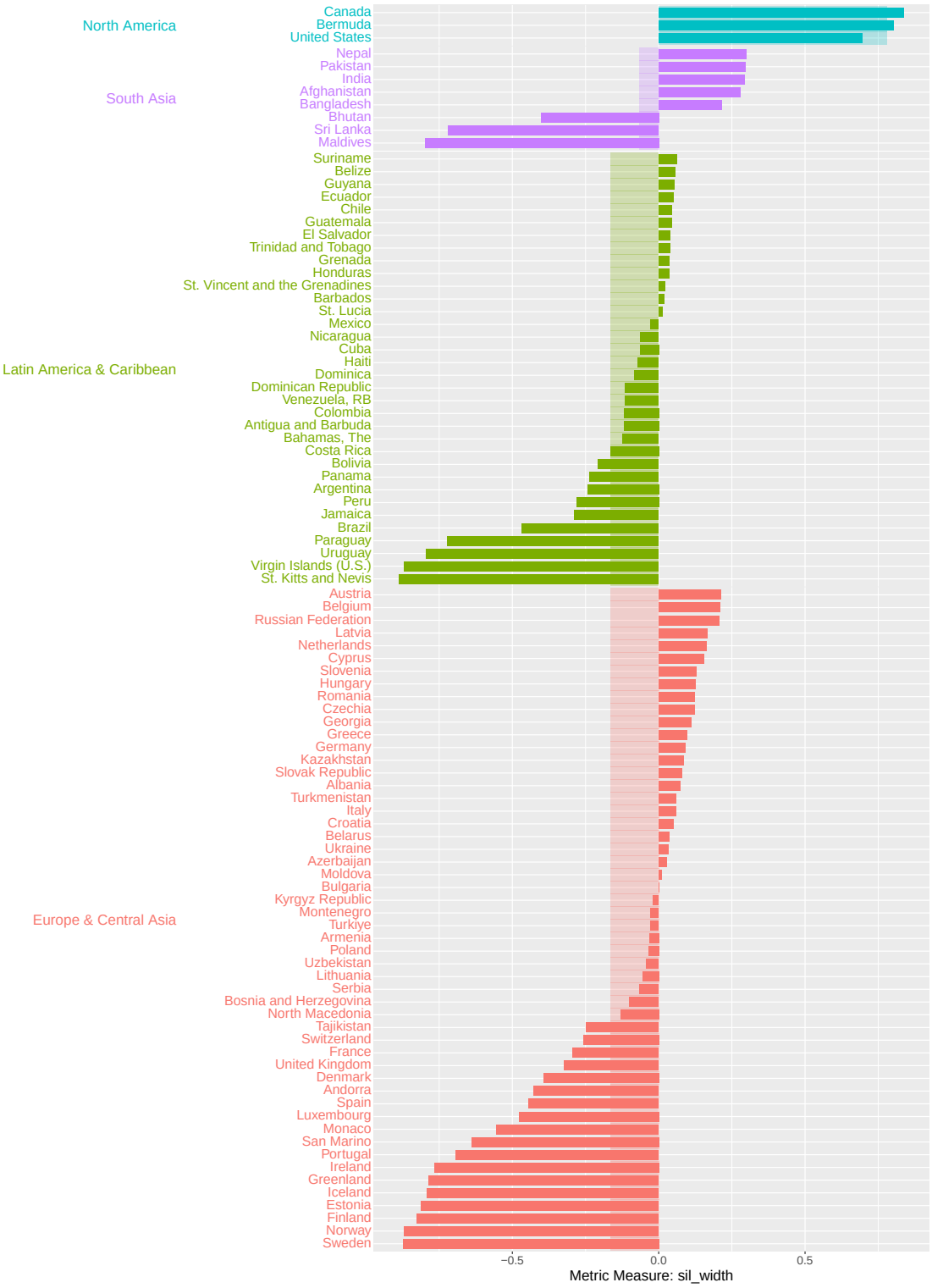

Figure 3.6: Country silhouette widths, grouped by region, with the average silhouette width for each region underlaid beneath the country bars.

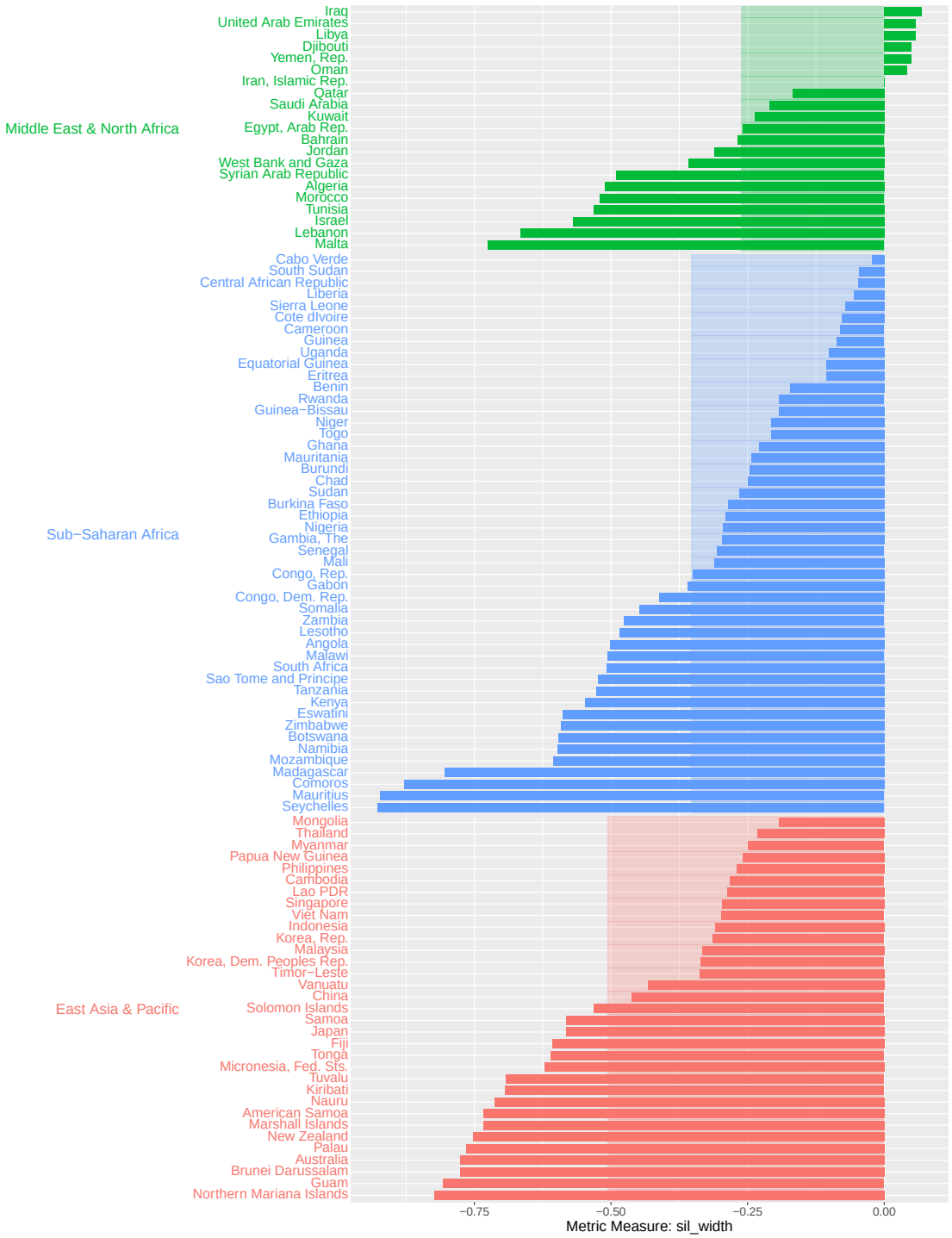

Figure 3.7: Country silhouette widths, grouped by region, with the average silhouette width for each region underlaid beneath the country bars. Countries in Sub-Saharan Africa and East Asia & Pacific regions all exhibit negative silhouette widths, suggesting that they do not fit well within their predefined regional groupings based on their data series, or that their behaviour may be more similar to countries in other regions.

Figures 3.6 and 3.7 present the silhouette width value for each country, partitioned by their respective region groupings. Each bar represent the silhouette width of a country, while the group-level average is underlaid beneath the country bars. A distinct colour palette is used to differentiate region levels, with countries within the same group sharing the same colour.

North America stands out as the region with the most closely aligned data series, with Bermuda, Canada and the United States all exhibiting high positive close to +1 silhouette widths. This indicates that these countries behave consistently in terms of their $PM_{2.5}$ air pollution trends and differ clearly from countries in other regions. In contrast, South Asia displays a mix of both positive and negative silhouette values. Notably, three countries in the region exhibit highly negative values, indicating a weaker fit within the South Asia group. A similar variation is observed in Latin America & Caribbean, Europe & Central Asia and Middle East & North Africa, where a few countries in these regions have slightly positive values, the majority exhibit extreme negative values close to -1 and some close to 0 values reflecting weak within-group similarity among countries within these regions.

Sub-Saharan Africa and East Asia & Pacific regions are characterised by uniformly negative silhouette widths, though with varying degree as some countries exhibit close to -1, while some are close to 0. This reflects noticeable differences in behaviour among countries within the same region. Figures 3.6 and 3.7 reveal that, only countries in North America exhibit consistently similar temporal patterns, while other regions exhibit weaker similarity with more group-wise variability.

The remaining visualisation functions are designed to produce interactive plots with tooltips displaying the country name and their respective metric value. For the purposes of this documentation, we provide static versions instead, while the interactive versions are available in an online archive via [a]. The code examples below demonstrate how each of these functions can be used to interpret the diagnostic measures discussed in Section 3.5.2 in the data trajectories space.

Data Trajectories Plot

The `plot_data_trajectories` function generates interactive line plots. It takes one main argument: a dataset containing data for a selected WDI indicator. It also accepts two optional arguments, `index`, which defaults to `NULL` and uses the first

[a] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/

variable saved as the WDI dataset attribute, and `interactive`, which defaults to `TRUE`. By default, it generates an ungrouped interactive line plot showing country-level trajectories. If an additional optional argument `group_var` is specified, the function instead produces a grouped version of the interactive line plots of the series, faceted by the specified grouping variable. Additional arguments `metric_summary` and `metric_var` can be supplied to highlight the trajectories based on the specified percentile of the metric values. `metric_summary` is a data frame containing the calculated diagnostic indices alongside the grouping information, while `metric_var` specifies which variable in that data frame contains the values of the metric for highlighting.

The function also accepts a `percentile` argument, which defines the threshold for highlighting countries based on their metric values. By default, `percentile =` 0.95, meaning the top 5% of countries based on the selected metric are interactively emphasised with distinct colour palette. When `group_var` is NULL, the function highlights countries using a global threshold across all countries. When `group_var` is specified, the function produces a faceted plot grouped by the specified variable, and countries are highlighted based on group-specific percentile thresholds rather than a global distribution threshold.

To demonstrate the functionality of the `plot_data_trajectories` for highlighting countries based on the computed metrics, we select two diagnostic indices: dissimilarity and linearity. We use absolute values of the linearity index to ensure that both strong positive and strong negative are treated equally when highlighting the top percentile.

The ungrouped and grouped data series trajectories with the top 5% countries highlighted based on their dissimilarities. The `plot_data_trajectories` function allows users to modify the axis titles and add a plot title to the data trajectories.

```
# ungrouped version
plot_data_trajectories(
  pm_data,
  metric_summary = pm_diagnostic_metrics,
  metric_var = "country_avg_dist",
  interactive = FALSE
) +
  ggplot2::labs(
  x = "Year", y = "Annual mean PM2.5 exposure"
  )
```

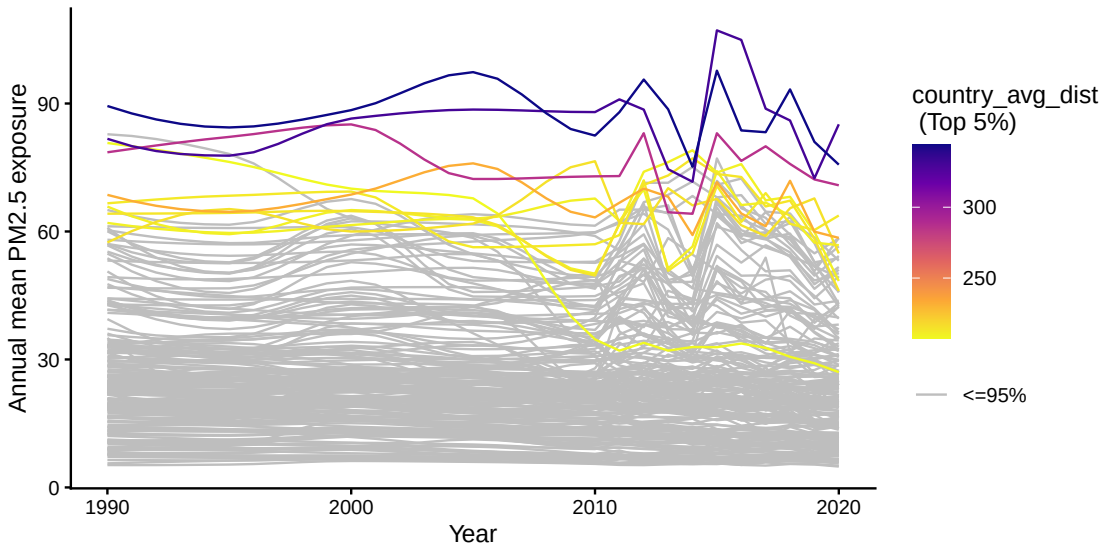

Figure 3.8: Country line plots of $PM_{2.5}$ air pollution dataset. Countries with average dissimilarity distance values below or at the 95th percentile are displayed in grey, while countries with the top 5% average dissimilarity between itself and other countries are highlighted using a colour gradient. Qatar and Niger, countries displayed in purple-blue exhibit the highest dissimilarity values. Hovering any of the highlighted lines in the interactive version reveals the corresponding country name and metric value.

In the interactive version of Figure 3.8 available via [b], hovering over each highlighted line displays the country name and the average dissimilarity distance value. This plot visually complements and reinforces the earlier findings from the

[b] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/interactive-plots/ungrouped_dissimilarity_plot.html

`pm_variation` output generated by the `compute_variation` function in Subsection 3.5.2. Qatar, Niger displayed in purple-blue and Mauritania displayed in reddish-purple are the top three countries with the highest average dissimilarity distance. As seen in Figure 3.8, these three countries also record the highest annual $PM_{2.5}$ air pollution exposure levels, with Qatar leading. Qatar's extensive oil and gas production and large-scale construction activities likely contribute to its elevated pollution levels and its distinctiveness relative to other countries. Ali Ahmadi et al. (2018) identified Qatar arid climate, along with rapid industrialisation, urbanisation, and traffic emissions, as major contributing factors to its elevated $PM_{2.5}$ and PM_{10} pollution levels.

```
# grouped version
plot_data_trajectories(
  pm_data,
  metric_summary = pm_diagnostic_metrics_group,
  metric_var = "within_group_avg_dist",
  group_var = "region",
  interactive = FALSE
) +
  ggplot2::labs(
    x = "Year", y = "Mean annual exposure of PM2.5 air pollution"
    )
```

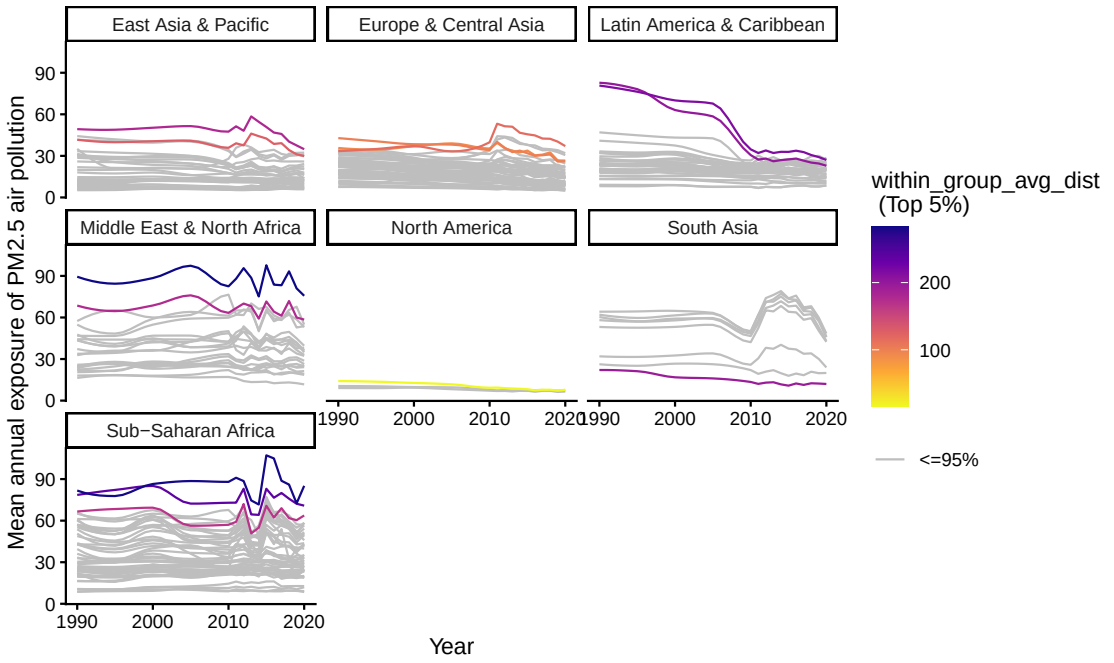

Figure 3.9: $PM_{2.5}$ air pollution data trajectories faceted by region groupings with group-based threshold rather than a uniform global threshold for highlighting countries with the top percentile. Qatar stood out with the highest dissimilarity values across other countries in Middle East & North Africa while Niger, Mauritania and Senegal are identified as countries with the highest dissimilarity within the Sub-Saharan Africa region. Hovering any of the highlighted lines in the interactive version reveals the corresponding country name and metric value.

The interactive version of Figure 3.9 is available online via [c]. Figure 3.9 shows that within each region, countries are highlighted based on group-specific thresholds. Qatar stands out among all other countries in Middle East & North Africa while Niger and Mauritania stand out in Sub-Saharan Africa. This suggests that their pollution trajectory over time are not only unusual at the global level but also relative to other countries in their region. In the study conducted by Bauer et al. (2019), air pollution across the African continent was found to be dominated by Saharan dust, with biomass burning identified as a major contributor, particularly in Central and West Africa. These natural environmental factors largely explain the high $PM_{2.5}$ air pollution levels and structural distinctiveness observed in Niger and Mauritania.

[c] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/interactive-plots/grouped_dissimilarity_plot.html

In Figure 3.8, countries highlighted based on the global threshold include Afghanistan, Bahrain, Egypt, Arab Rep., India, Mali, Mauritania, Niger, Peru, Qatar, and Senegal. However, in Figure 3.9 where countries are highlighted based on the group-specific threshold, some of these countries are no longer highlighted, while new ones appear within their respective regions. Notably, no countries from East Asia & Pacific or North America region were highlighted under the global threshold, yet region-specific highlighting reveals outliers that would otherwise be overlooked. This shows the value of group-specific thresholds in revealing regionally unusual temporal behaviours.

The results from the country average and within-group average dissimilarity align with the patterns observed in the silhouette width partition plot. Countries such as Qatar, Niger, and Mauritania were identified as outliers not only due to their high country average and within-group average dissimilarity, they also belong to regions that were characterised by weak similarity and within group variability. In contrast, Bermuda, Canada and the United States, all in the North America region, exhibit high positive silhouette widths, indicating relatively strong similarity among the countries in the group. However, the group-specific dissimilarity threshold highlights United States as a country with the highest dissimilarity within the region. Although the countries appear relatively similar, the data series for the United States differs slightly from the data series of Bermuda and Canada. This is also reflected in its silhouette width, which is positive but lower than the other countries in the North America region. The alignment across these measures underscores the value of using multiple metrics to identify structural patterns and detect outliers in country-level panel data.

Having introduced the chosen diagnostic indices, the code below demonstrate how the `plot_data_trajectories` function can be used to highlight countries with the strongest linear trends (positive and negative), based on the absolute value of the linearity measure.

```
pm_trend_shape <- pm_trend_shape |>
      dplyr::mutate(abs_linearity = abs(linearity))

plot_data_trajectories(
        pm_data,
        metric_summary = pm_trend_shape,
        metric_var = "abs_linearity",
        percentile = 0.96,
  interactive = FALSE
 ) +
  ggplot2::labs(
  x = "Year", y = "Annual mean PM2.5 exposure"
    )
```

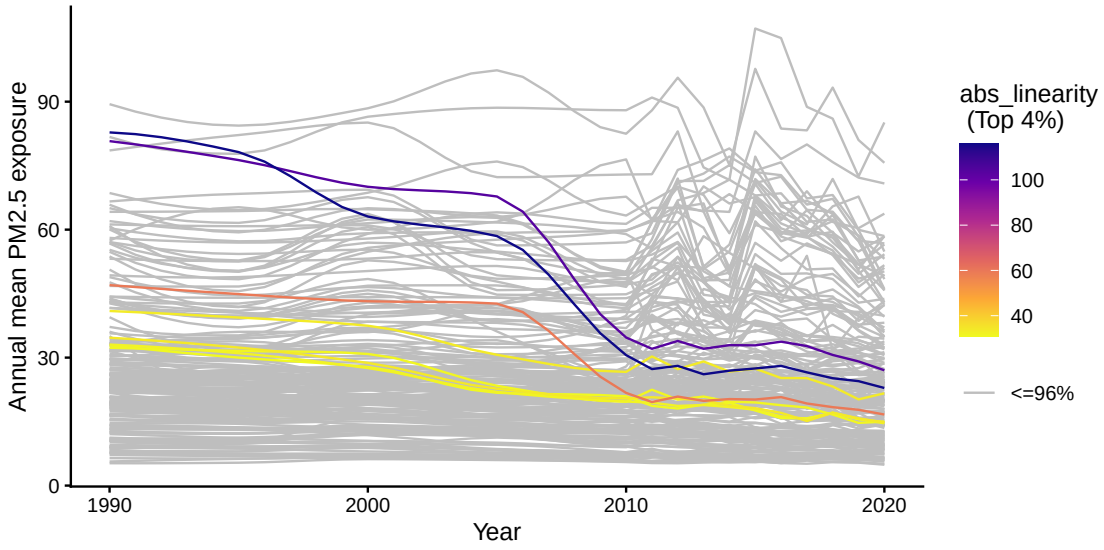

Figure 3.10: Ungrouped $PM_{2.5}$ air pollution data trajectories with highlighted countries based on the linearity metric values. Countries with absolute linearity values below or at the 96th percentile are displayed in grey, while countries within the top 4% absolute linearity values are displayed using a colour gradient. All highlighted countries have negative linear trend, indicating that the most pronounced linear trends in the data reflect strong decline in the decomposed series. Hovering any of the highlighted lines in the interactive version reveals the corresponding country name and metric value.

The interactive version of Figure 3.10 is available online at [d]. Bolivia, Peru were already identified as top 2 countries with highest negative linearity value in Subsection 3.5.2, along with other countries falling within the top 4% are highlighted using a colour gradient. In Figure 3.10, the highlighted countries exhibit negative linear trend. This shows that, the most extreme linearity values (β_1) of the decomposed trend component have sustained downward trajectories. Figure 3.4 supports this as the distribution plot of the linearity metric tends to be left skewed with majority of the countries exhibiting negative linearity values. Sicard et al. (2023) identify that the dominating decrease in air pollution is primarily attributed to significant reductions in emissions and the establishment and implementation of environmental policies and legislation. According to Xu et al. (2023), The Gross Domestic Product (GDP) per capita has influenced the shift of $PM_{2.5}$ air pollution from positive to negative in most countries by 2010, reflecting increased investment in environmental protection and the introduction of air pollution prevention and control bills. Figure 3.10 confirms these reductions in most countries.

```
pm_trend_shape_group <- add_group_info(
      metric_summary = pm_trend_shape,
      pm_data
  )

plot_data_trajectories(
        pm_data,
        metric_summary = pm_trend_shape_group,
        metric_var = "abs_linearity",
        group_var = "region",
        percentile = 0.96,
  interactive = FALSE
 ) +
  ggplot2::labs(
    x = "Year", y = "Mean annual exposure of PM2.5 air pollution"
 )
```

[d] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/interactive-plots/ungrouped_linearity_plot.html

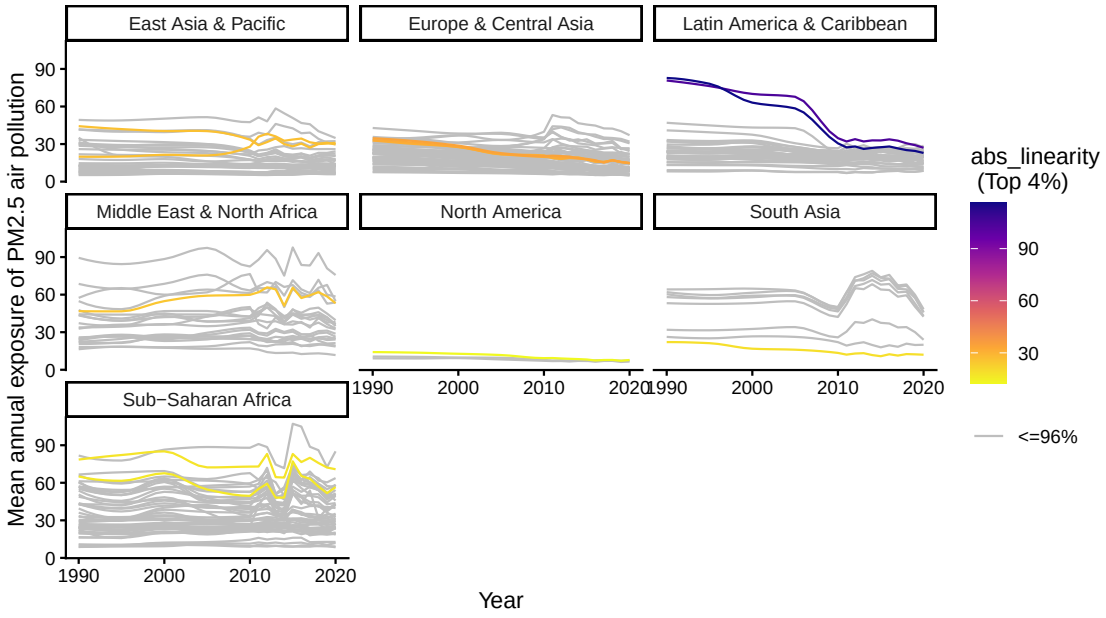

Figure 3.11: $PM_{2.5}$ air pollution data trajectories faceted by region groupings with group-based threshold with highlighted countries based on their linearity metric values. Countries with absolute linearity values below or at the 96th percentile are displayed in grey, while countries within the top 4% absolute linearity values are displayed using a colour gradient. Hovering any of the highlighted lines in the interactive version reveals the corresponding country name and metric value.

In Figure 3.11, countries are highlighted based on a group-specific thresholds, which reveal those exhibiting pronounced linear trends within each region. The interactive version of Figure 3.11 available via [e] shows that in the East Asia & Pacific region, Mongolia and Thailand are the highlighted countries. Mongolia demonstrates a positive linear trend while Thailand exhibits a negative trend. This indicates that, though the ungrouped version shows that highlighted countries within the top 4% absolute linearity values all demonstrates negative trends, within the East Asia & Pacific region shows otherwise. Countries falling within the 96th percentile exhibit positive and negative trends. In Europe & Central Asia, both Moldova and Ukraine show clear decreasing linear trends in their $PM_{2.5}$ data series. Peru and Bolivia in Latin America & Caribbean, have the most pronounced decreasing trend. The United States in North America; Maldives in South Asia; Mauritania and Nigeria in Sub-Saharan Africa also exhibit decreasing trends. In

[e] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/interactive-plots/grouped_linearity_plot.html

contrast, Saudi Arabia in Middle East & North Africa displays an increasing linear trend.

These observed decreasing trends in $PM_{2.5}$ air pollution levels align with the findings of Guerreiro et al. (2014), who reported substantial reductions in particulate matter emissions with significant improvements between 2000 and 2019 in regions such as Eastern United States, Europe, Southeast China, and Japan. These reductions are linked to strong regulatory frameworks and air quality initiatives implemented in these areas, contributing to declining particulate matter concentrations. Precipitation also plays a role through wet deposition, acting as an efficient mechanism for pollutant removal Xu et al. (2023). However, while significant progress is noted in developed areas, challenges remain in developing regions, such as India and the Middle East, where $PM_{2.5}$ concentrations often continue to rise due to the prioritisation of economic development over environmental concerns.

Parallel Coordinates Plot

The `plot_parallel_coords` function generates parallel coordinate plots with normalised metric values to a scale of 0 to 1. It takes two main arguments, a data frame `diagnostic_summary` containing all diagnostic indices values alongside the predefined grouping information, and a variable specified by `colour_var` in the data frame used to assign colours to the parallel lines. If an additional optional argument `group_var` is specified, the function instead produces a grouped parallel coordinate plot, where metric values are normalised within each group before plotting, and the resulting plot is faceted by the specified grouping variable. The parallel coordinate plot is rendered interactively using `ggiraph` with tooltips to display the corresponding country name, metric and metric value of each parallel line.

```
# ungrouped version
plot_parallel_coords(
      diagnostic_summary = pm_diagnostic_metrics_group,
      colour_var = "region"
)
```

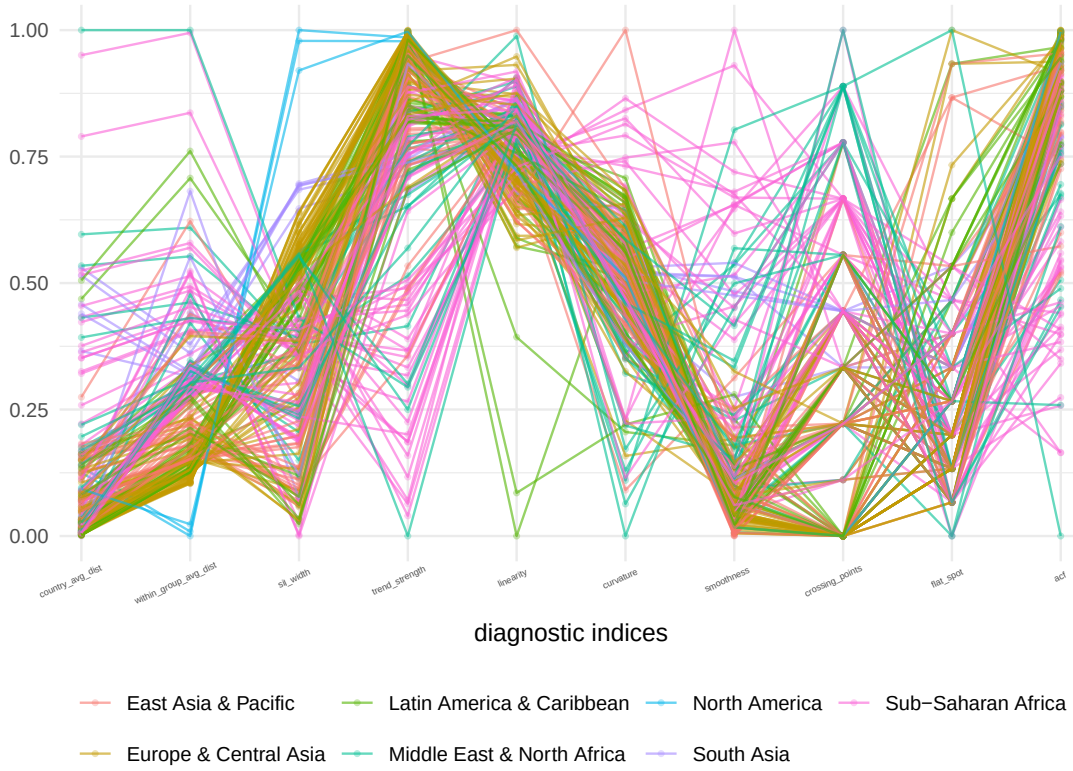

Figure 3.12: Parallel coordinate plot displaying the metric values across all the diagnostic indices. The metric values are normalised to a scale of 0 to 1. Countries in Sub-Saharan Africa region, shown in magenta, display a wide spread across most diagnostic indices. Hovering any parallel line in the interactive version reveals the corresponding country name and hovering the points across the x-axis reveals the metric name, metric value and the country name.

The interactive version of Figure 3.12 available via [f] displays the parallel coordinates across all 10 diagnostic indices. Hovering over the x-axis, the tooltips show the country name of each parallel line, the correspondence metric, and its metric value. This collectively reinforce the findings presented in Subsection 3.5.2.

[f] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/interactive-plots/ungrouped_parallel_plot.html

```
# grouped version
plot_parallel_coords(
      diagnostic_summary = pm_diagnostic_metrics_group,
      colour_var = "region",
      group_var = "region"
)
```

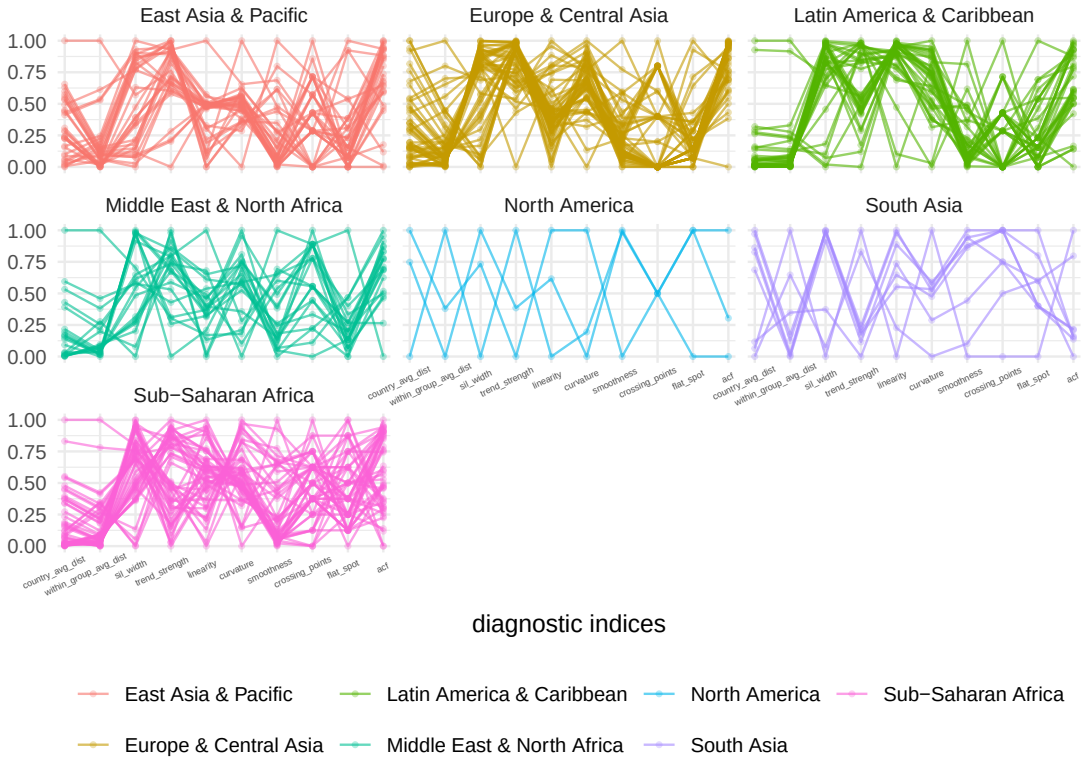

Figure 3.13: Parallel coordinate plot displaying the metric values across all diagnostic indices grouped by region. The metric values are normalised to a scale of 0 to 1 within each group. Countries in Sub-Saharan Africa region, shown in magenta, displays the widest spread across most diagnostic indices. Hovering any parallel line in the interactive version reveals the corresponding country name and hovering the points across the x-axis reveals the metric name, metric value and the country name.

Figure 3.12 reveals that countries in North America, shown in skyblue consistently recorded almost similar values, with slight difference in United States reinforcing earlier findings about the behavioural patterns of $PM_{2.5}$ air pollution levels

within this region, see 3.13 with its interactive version available via [g]. In North America region their silhouette widths are all positive and high as seen in Figure 3.6, further supporting these findings. These countries exhibit high trend strength, coupled with negative linearity and close to 0 curvature, indicating a clear, consistent, and dominant linear decline in $PM_{2.5}$ levels over time, with the United States showing the most pronounced downward trend. Their low smoothness scores further imply that the time series evolve gradually without abrupt changes. The data series of each of these countries crosses its median only once, pointing to a single shift in their series followed by sustained stability. Both Canada and the United States exhibit 7 long flat spots, while Bermuda exhibits 5, indicating periods where their data series remained within an interval. This is consistent with the overall smoothness and linearity of their trends. Additionally, they all show high autocorrelation, with the United States having the highest value, reflecting strong temporal correlation where each series is closely related to its preceding and succeeding time points.

The countries in Sub-Saharan Africa region, shown in magenta, display a widespread index values. In contrast, countries in Europe & Central Asia region, represented in darkgoldenrod, exhibit relatively uniform behaviour in most of the diagnostic indices.

United Arab Emirates in the Middle East & North Africa region, shown in tealgreen is another country that stands out in Figure 3.13, primarily due to its distinctive values across several diagnostic metrics. Notably, the data series records low values for both trend strength and autocorrelation, indicating that it exhibit minimal to no consistent temporal pattern and successive observations are largely independent of one another. However, it simultaneously records some of the highest values for the number of crossing points and the longest flat spot metrics. These suggest extended periods of stability with shifts around its median values.

Bolivia, a country in the Latin America & Caribbean region, shown in leafgreen also emerges as a notable country that stands out in both Figures 3.12 and 3.13, characterised by a unique combination of high trend strength, low linearity, low crossing points, and high autocorrelation. The high trend strength indicates a pronounced directional movement in its $PM_{2.5}$ air pollution data series, while the low linearity value confirms that this movement is a consistent linear decline. The

[g] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/interactive-plots/grouped_parallel_plot.html

low number of crossing points (to be precise 1) implies that the data series crosses its median value once, possibly due to a sustained downward linear trend with no subsequent directional changes. The high autocorrelation reflects strong correlation between each series and its preceding time point. This shows that, the $PM_{2.5}$ air pollution levels over time in Bolivia follow a stable and strongly consistent downward linear trend.

Link-view of Metrics and Series Plot

The `plot_metric_linkview` function creates an interactive linked dual-based visualisation that connects the diagnostic space with the corresponding data trajectories. It takes three main arguments: a dataset containing data for a selected WDI indicator, `metric_summary`, a data frame containing the computed diagnostic indices with the predefined grouping information ; and `metric_var` a pair of metric variables within the `metric_summary` data frame which are used to create a scatterplot. The function also accepts an optional `index` argument which defaults to NULL and the function uses the first variable stored as an attribute in the WDI dataset. By default, the function generates an ungrouped interactive link-based visualisation. However, if an additional optional argument `group_var` is specified, the function instead produces a grouped version of the interactive dual-based visualisation of a scatterplot of two selected diagnostic metrics alongside their series trajectories, faceted by the specified grouping variable.

```
# ungrouped version
plot_metric_linkview(
          pm_data,
          metric_summary = pm_diagnostic_metrics,
          metric_var = c("linearity", "curvature")
        )
```

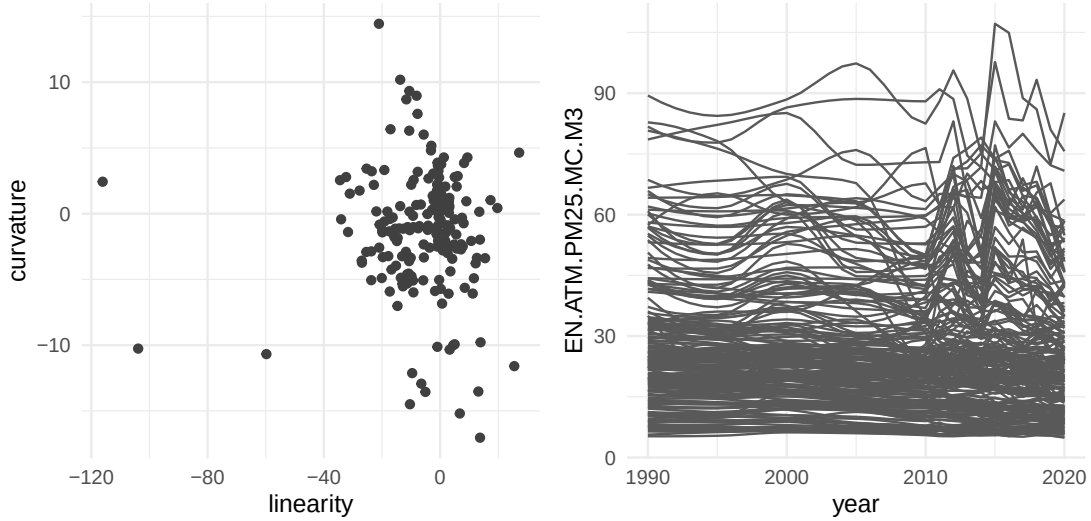

Figure 3.14: Link-based plot showing the relationship between linearity and curvature metrics across all countries. Each point in the scatterplot represents a country, and hovering a point in the interactive version reveals the corresponding trajectory of the data series.

The interactive version of Figure 3.14 available via [h] simultaneously displays a link view of the metric values and their correspondence data trajectories: one panel displays data series line plots for all countries, while the other shows a scatterplot of the relationship between the linearity and curvature metrics. This enables an interactive exploration of viewing the metrics values in the data space. Hovering over a line in the time series plot highlights the corresponding country and its scatter point value in the other plot, and vice versa. This dynamic interaction enhances an intuitive understanding of how temporal behaviours relate to pair of structural features of the data, captured by the collection of diagnostic indices.

```
# grouped version
plot_metric_linkview(
          pm_data,
          metric_summary = pm_diagnostic_metrics_group,
          metric_var = c("acf", "flat_spot"),
          group_var = "region"
      )
```

[h] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/interactive-plots/ungrouped_linkview_plot.html

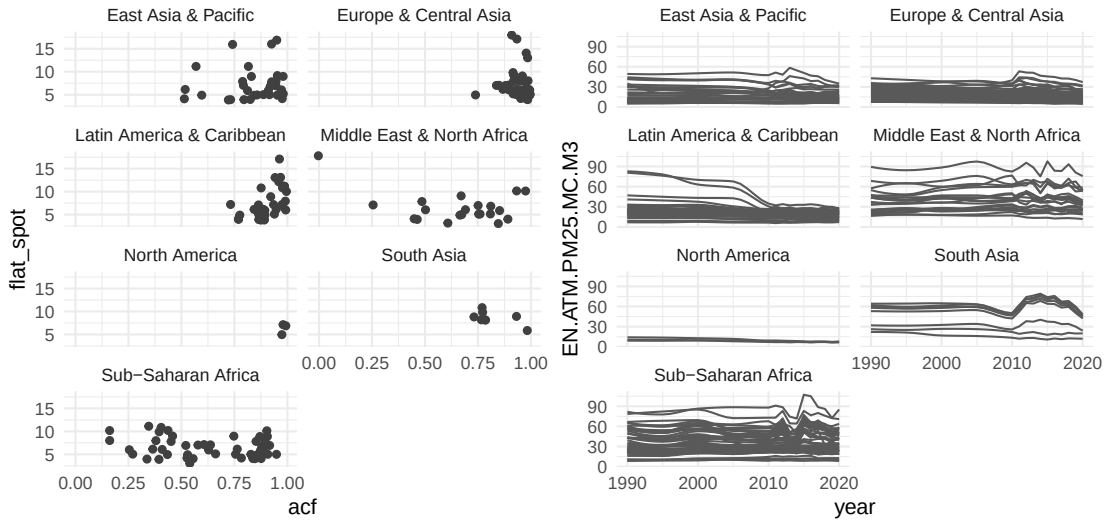

Figure 3.15: Link-based plot showing the relationship between autocorrelation and the longest flat spot metrics across all countries faceted by region. Each point in the scatterplot represents a country, and hovering a point in the interactive version reveals the corresponding trajectory of the data series in its panel.

Figure 3.15 with its interactive version available via [i] extends this view by incorporating group-level structure through faceting to the scatterplot of autocorrelation and the longest flat spot metrics and data trajectories plot. In this version, the scatterplot and data trajectories panels are faceted by region, allowing within-group comparisons of metrics relationships and temporal behaviours. This facilitates the identification of countries that behave differently from others within the same group, based on both their diagnostic indices and data trajectories.

The link-view plot is highly flexible, allowing users to visualise the scatterplot of any two selected diagnostic metrics and the data trajectories plot. Extending this approach, a linked version of the parallel coordinates plot connected to the trajectory plot would further enhance interpretability by enabling simultaneous exploration of multiple metrics for each country and how they reflect in the temporal series.

[i] https://oluwayomi-olaitan.github.io/rjournal-wdiexplorer-interactive-plots/interactive-plots/grouped_linkview_plot.html

3.6 Discussion

The primary contribution of `wdiexplorer` lies in enabling enhanced exploratory analysis of country-level panel data within the R programming environment. Its functionality specifically supports the need to analyse the temporal behaviour of WDI data, both at the level of individual countries and across countries of the same peer group. Furthermore, the incorporation of data grouping structures in visually presenting the analytical results facilitates the identification of unique patterns, outliers, and other interesting characteristics in temporal behaviours. This integration reinforces a comprehensive workflow where numerical assessment and visual representation enhance better understanding and interpretation of the data.

`wdiexplorer`, as the acronym implies "World Development Indicators explorer", is primarily designed for the exploration of WDI country-level panel data, but it is not limited to this database. The package can also accommodate other databases with data characterised by repeated observations of the same variables collected over time from multiple countries passed directly into the package functions, provided the `index` variable is specified across all functions that require the variable with the country-year numerical values. In addition, the computational functions of `wdiexplorer` that compute the diagnostic metrics can be integrated with a wide range of visualisation tools. This flexibility enables users to explore the metrics values within custom visualisations, while enabling visual exploration of the outputs to the underlying quantitative diagnostic measures.

Several functions have been implemented in the `wdiexplorer` package to support (1) data sourcing and preparation, (2) computation of the set of diagnostic indices (3) both static and interactive visualisations. The workflow organises the package functions into these three stages to provide a seamless and user-friendly process. As illustrated in Section 3.5, we demonstrated how to source and prepare the data, use the available 14 functions and interpret the results at every stage of the workflow. To ensure comparability, all functions were applied to the $PM_{2.5}$ air pollution indicator data and interpreted accordingly. These features establish `wdiexplorer` as a self-contained R package for the exploration of country-level panel data.

We provide users with practical guidance to facilitate reproducibility through comprehensive vignettes available via https://oluwayomi-olaitan.github.io/wdiexplorer/articles/pm_analysis.html. This vignette demonstrates how to apply each of the 14 functions of the `wdiexplorer` package to explore World

Development Indicators data. It guides users through example workflows for two indicators: one with a complete annual dataset ($PM_{2.5}$ air pollution data) and another with missing entries and data collected triennially (PISA data). It contains interpretations of the diagnostic indices output and demonstrates the visualisation tools, highlighting their analytical and visualisation capabilities.

A potential direction for future work is to extend the static visualisation functions to support interactivity. The current implementation of `plot_metric_distribution` generates a static representation of the distribution of metric values, summarising the variation by displaying the spread and shape of their distributions, with individual countries represented by dots. Enhancing this function with interactive capabilities will enable users to hover over a point to reveal the corresponding country name in the distribution plot. This would align the `plot_metric_distribution` function with other interactive visual functions available in the `wdiexplorer` package.

In conclusion, the `wdiexplorer` package contribute to advancing a comprehensive workflow for exploratory analysis of country-level panel data. By following our step-by-step workflow, researchers and other R users can better understand their data, identify patterns and outliers, and communicate results in the context of the data structure more effectively.

4

Visualisations for Exploratory Modelling Analysis of Bayesian Hierarchical Regression Models

4.1 Introduction

Visualisation has been a longstanding tool within statistical workflows for simplifying complex concepts into comprehensible insights and plays an important role in investigating, understanding and criticising regression models. It also plays a key role in exploratory modelling analysis, which Unwin and Valero-Mora (2018), described as "the evaluation and comparison of many models simultaneously".

The importance of visualisation throughout the Bayesian analysis workflow has been well-established by Gabry et al. (2019). They emphasise the benefits of visualisation at all stages, from exploratory data analysis to model development, diagnostics and communication of results. They further explain that visualisation helps sift through networks of increasingly complex models that can capture the features and heterogeneity in the data and they state that visualisation supports model comparison, evaluation and decision-making.

In the current chapter, we are not addressing the entire Bayesian workflow, our goal instead is the design of visualisations for exploring parameters and hyperparameters across different BHMs. To date, little work has been done on visualisation in this context.

The BHM relies on a statistical modelling procedure that integrates information across many levels, enabling simultaneous estimation of multiple model parameters

(Gelman et al., 2013). The model possesses two essential features. Firstly, it has a hierarchical or multi-level structure. Secondly, it uses prior distributions to reflect available information, even if it is vague about the probable values and variability at each hierarchical level (McGlothlin and Viele, 2018). The model uncertainties are propagated and quantified by combining the BHM prior structure with the model for the observed data.

The BHM is particularly valuable in situations where sparse data poses a challenge for parameter estimation (Rue and Martino, 2007). The structure enables the leveraging of strength from group level parameters to enhance estimates at the individual-level, resulting in more reliable and robust estimates. In addition, the inherent shrinkage of parameter estimates, pulling estimates toward the overall population distribution, can reduce the influence of outliers in the data.

In this chapter, via a set of visualisation principles, we specifically explore the importance of visualisation for BHMs for gaining insights into model choice and interpretation of model parameters. We follow the strategies for visualising statistical models presented by Wickham et al. (2015), which are the display of models in data space and examining a collection of multiple models. Our visualisation designs give careful consideration to choices of layout, colour, ordering and scaling to facilitate effective comparisons.

To demonstrate the relevance of our visualisation designs to a real-life dataset, we analyse a subset of the PISA dataset, obtained from the OECD website (OECD, 2024). PISA, initiated by the OECD in 1997, provides regular and timely assessments of student achievement with policy-oriented comparable indicators. The surveys, conducted among 15-year-old students, assess performance in reading, mathematics, and scientific literacy (Harlen, 2001). Tests developed and mandated by the OECD are administered in three-year cycles since its initiation in 2000. The resulting comprehensive dataset offers valuable information for comparing subject performance at continental, regional, national, and sub-national levels over time. The multi-level nature of this dataset makes it suitable for illustrating our proposed visualisation principles for exploratory modelling analysis of BHMs. For the purpose of this analysis, models are fit to a subset of the PISA data using the mathematics average scores for forty European countries from 2003 to 2018. PISA 2000 results in mathematics are not considered, since the first full assessment in mathematics took place in 2003. The 2021 results were delayed to 2022 and were not published until recently, due to the global pandemic.

Our objective is to create visualisations that evaluate model fits, posterior parameter and hyper-parameter distributions and facilitate model comparison. We present five principles for effective visualisation of models in data space and for examining a collection of multiple models simultaneously. We extend the conventional way of displaying model fits on the data points to arrange the individual levels by their respective hierarchical structures, using distinct formatting, layouts, ordering, scaling and colouring. The goal is to enhance the interpretation of model results with respect to information sharing and to explore variations in parameter estimates across different models and hierarchical levels.

The outline of this chapter is as follows: Section 4.2 provides an overview of the background to this study, focusing on the governing essential strategies for visualising statistical models as proposed by Wickham et al. (2015). Section 4.3 outlines the principles upon which our new visualisation designs are based, using the PISA dataset for illustration, and we discuss insights gained from the PISA analysis, facilitated by the new visualisations. In Section 4.4, we explore the most recent year (2022) of PISA data and compare the observed estimates with model-based predicted estimates. Finally, in Section 4.5, we conclude with an evaluation of the potential benefits and drawbacks of our methods, as well as other related future research.

4.2 Background

Visual representations of models offer significant value as complementary tools to numerical summaries. In the Bayesian context, several researchers have investigated the effectiveness of visualisation in developing, computing, validating and evaluating statistical models (Gabry et al., 2019; Gelman, 2007; Raudenbush and Bryk, 2002; Correa-Álvarez et al., 2023). Wickham et al. (2015) highlighted three important strategies for visualising statistical models: (1) displaying the model in data space, (2) examining a collection of multiple models, and (3) exploring the process of model fitting. In this work we consider the first two strategies only, which underscore the importance of visualisation in comprehending how a model summarises data, reacts to parameter changes, and fits the data.

There are various conventional approaches for displaying the model in data space. For illustration, Figure 4.1 presents a display of estimates from a model fit to mathematics scores from the PISA dataset for 40 European countries using

`brms` (Bürkner, 2018). Here, regression fits and 95% credible intervals are displayed on the observed data, faceted by country, indicating that the linear model broadly captures the patterns in the data and provides information on the direction of trends for each country.

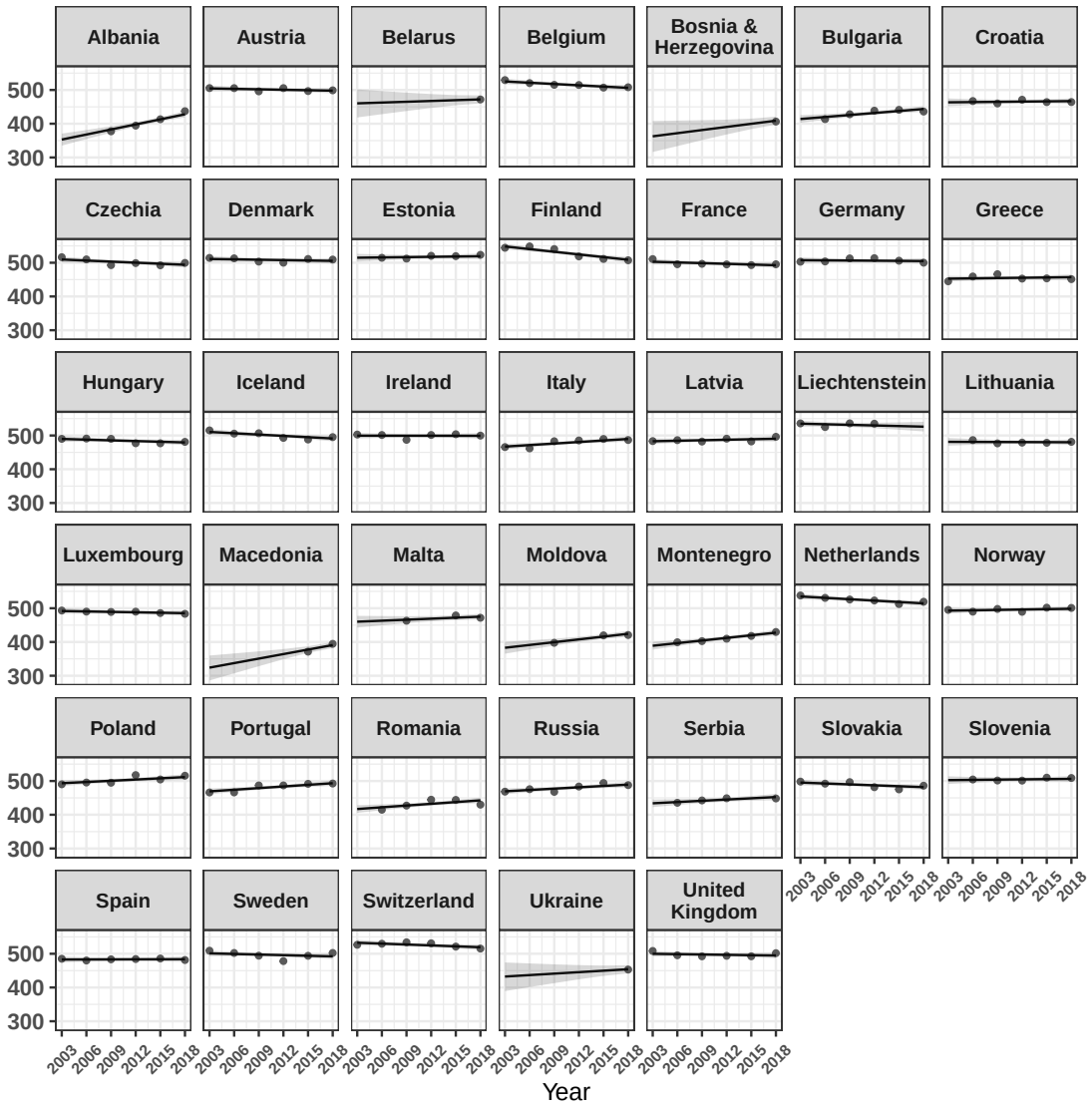

Figure 4.1: Visualisation of PISA European mathematics scores with model fit and 95% credible intervals, showing overall trends across each country and how well the fitted model captures the pattern in each country over time. Belarus, Bosnia & Herzegovina and Ukraine, each with one data point and wider credible intervals, reflecting increased posterior uncertainty inherit the common global slope, due to the hierarchical model.

The model equation for the displayed fit is specified in `brms` as:

$$\text{math} \sim \text{year} + (1 + \text{year}|\text{country}) \qquad \text{country model}$$

The Bayesian model specification is presented as:

$$\begin{aligned} Y_{c,t} &\sim \text{Normal}\left(\alpha_c + \beta_c t, \sigma^2\right) \\ \alpha_c &\sim \text{Normal}\ \left(\mu_\theta,\ \sigma_\theta^2\right) \\ \beta_c &\sim \text{Normal}\ \left(\mu_\omega,\ \sigma_\omega^2\right) \end{aligned} \qquad (4.1)$$

where,
α_c is a country specific intercept for the 40 countries in Europe,
β_c is a country specific slope for the 40 countries in Europe,
σ^2 is the population variation,
μ_θ represents the overall population intercept,
μ_ω represents the overall population slope,
σ_θ^2 captures the variation of the country intercept α_c around the population intercept μ_θ,
σ_ω^2 captures the variation of the country intercept β_c around the population intercept μ_ω.

The prior specification for the *country model* is as follows:

$$\begin{aligned} \mu_\theta &\sim \text{Normal}\ \left(500, 100^2\right) \\ \mu_\omega &\sim \text{Normal}\ \left(0, 50^2\right) \\ \sigma_\theta^2 &\sim \text{Truncated-t}\ (0, 10^{-2}, 1)T(0,) \\ \sigma_\omega^2 &\sim \text{Truncated-t}\ (0, 2^{-2}, 1)T(0,) \\ \sigma^2 &\sim \text{Truncated-t}\ \left(30, 10^{-2}, 1)T(0,)\right) \end{aligned}$$

This model leverages global parameter distributions to enhance parameter estimation at the country level. Therefore, linear fits for countries with one data point (Belarus, Bosnia & Herzegovina and Ukraine) are obtained from the global parameter estimates, as can be seen in Figure 4.1. With this exception, the display provides no insight into the hierarchical structure of the model.

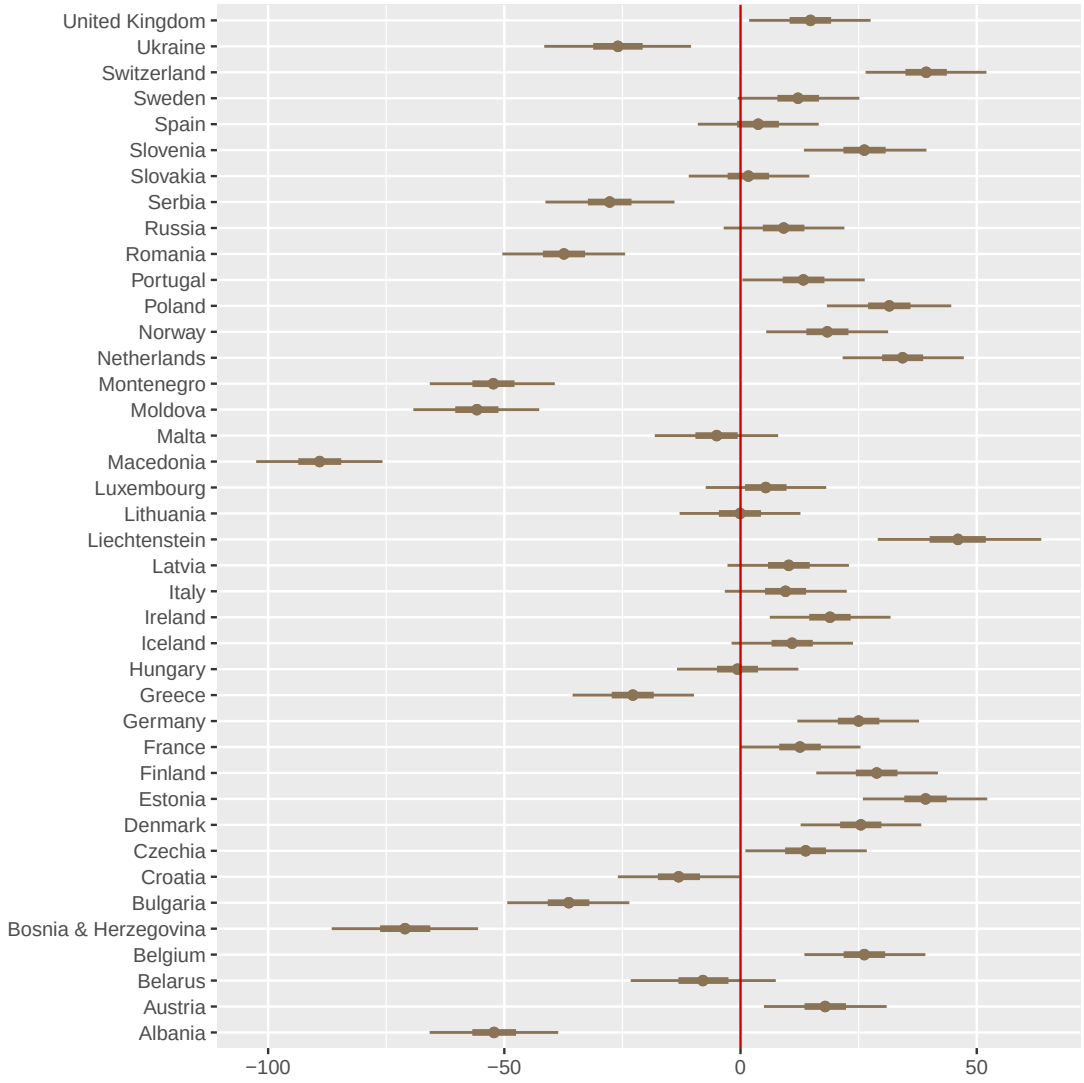

Figure 4.2: The intercept offset median and 95% credible intervals for the 40 European countries, showing the deviation from the global estimate. Hungary, Lithuania, Slovakia, and Spain have little deviation from the global estimate, while 2018 PISA mathematics performance for Macedonia and Bosnia & Herzegovina is comparatively low.

In hierarchical modelling, observations are structured into nested levels allowing models to capture heterogeneity and shared characteristics across different levels of the hierarchy. This can result in more stable and robust estimates (Gelman, 2007). For example, countries in the PISA data could be grouped by geographical region. Hierarchical modelling also automatically incorporates the shrinkage of parameter

estimates (i.e., pulling parameter estimates toward the overall hierarchical distribution) and this shrinkage can reduce the influence of outliers. These characteristics of hierarchical modelling prove particularly valuable in situations where sparse data poses a challenge (Rue and Martino, 2007). Hence, displaying the estimates from these types of models require showing the information sharing between the individual level and the hierarchical structures.

Parameters at a higher level of hierarchical models typically represent global or group-level quantities, and the deviation of the individual level from the global estimates or the group-level (hyper-parameter) estimates are the "offsets". Figure 4.2 shows parameter offsets of the hierarchical model presented in Figure 4.1 illustrating the offsets of each country from the global mean. (Note that our models are parameterised so that the intercept represents mathematics scores in 2018, the last year used in the model fits). This conventional method of displaying offset parameters provides insight into how each individual level moves away from the global estimate but provides no information on the global estimate itself.

For the first strategy outlined in Wickham et al. (2015) which emphasises the display of models in data space, Hurley et al. (2022) develops an approach aligned with this idea by presenting a method for displaying collections of statistical models in data space to facilitate model comparison. Related to this perspective of model to data space visualisation, slice-based visualisation techniques provide an additional way of improving interpretability of machine learning models by displaying conditional model fits directly in the predictor space. In our context, these ideas motivate the development of visualisation that support the interpretation of hierarchical statistical models through their behaviour in the data space.

For the second strategy outlined in Wickham et al. (2015), it is advisable to consider and visually compare multiple models. A useful tool for comparing model coefficients is `gg_coef_model()` from the `ggstats` package (Larmarange, 2026), which provides support for `brms` model fit and enables direct comparison of coefficients across multiple fitted models. one of the objectives of this work is to visually compare a collection of multiple models. Consequently, in our context, there is a need for visualisation tools specifically designed to present parameter estimates from multiple hierarchical models. A simple representation in this context might be a display in the style of Figure 4.2 but which shows intercept or slope estimates across multiple models side-by-side.

Recent work has emphasised the importance of interpretable visualisation for

complex statistical models. In particular, Arel-Bundock (2025a) highlight the use of predictions, marginal effects, and parameter summaries for model interpretation, while Heiss (2021) and Heiss (2022) provide general frameworks for communicating hierarchical, Bayesian, and mixed-effects models using similar quantities. Although these approaches primarily focus on models incorporating covariates, they underscore the role of visualisation in improving the interpretability of hierarchical structure. Horan et al. (2024) also demonstrate the use of spatial layouts to present multilevel models of voting behaviour in recent UK general elections. Building on this idea, more informative displays can be achieved by incorporating hierarchical structure and hyper-parameter estimates directly into the visualisation of model comparisons.

In this work, we design new visualisations for BHMs that improve on Figures 4.1 and 4.2 by incorporating hierarchical information from various levels in the model hierarchy. In the design of these visualisations, there are various principles to consider, for example, those described by Munzner (2014) and Wilke (2019). Both authors highlight the need for consistency in design choices, such as uniformity of scaling and consistency in the use of colour and formatting. Wilke (2019) emphasises the use of multiple panels for large and complex datasets, such that the data can be segmented based on the one or more dimensions of the data, visualised separately, then arranged into a grid for an easy interpretation. We have adapted these design principles to explore model estimates in data space and examine a collection of multiple models, highlighting prominent disparities and similarities between and within the hierarchical structures.

4.3 Visualisation Principles

In this section, we present the key principles of our visualisation approach. In Section 4.3.1 we apply these principles to the design of a display for BHMs in data space and in Section 4.3.2, we further apply these to the design of visualisations for the comparison of multiple BHMs.

To illustrate our visualisation approaches, we employ the previously used subset of the PISA data. We decided to consider country groupings according to geographical regions, as defined by the World Health Organisation (WHO, 2023), and level of income, using the World Development Indicators, provided by the WDI (Arel-Bundock, 2025b) R package. These indicators categorised countries

into lower-middle, upper-middle and high-income by their population, environment, economy, global linkages, states and markets. Within the dataset under study, only Ukraine belongs to the lower-middle income category. Consequently, we merged both lower-middle income and upper-middle income classifications into a single "middle-income" category, resulting in two distinct income categories: middle income and high income.

We fit various BHMs to the PISA data, using the `brms` package, which is a front-end for `Stan` (Carpenter et al., 2017).

4.3.1 Model in data space

Visualising the model in the data space requires the presentation of model estimates alongside observed data. Ideally, this should be done in accordance with the model structure to enhance understanding of the model fit in relation to the model specification. In hierarchical modelling, individual-level observations are nested within higher-level groupings to account for the variability between groups and within groups and to capture the average characteristics or influences of the groups on the individual level. An example for the PISA data is to fit a model using a geographical (country within region) hierarchical structure. The model equation is specified in `brms` as:

$$\text{math} \sim \text{year} + (1 + \text{year}|\text{region}) + (1 + \text{year}|\text{country}) \qquad \text{region model}$$

with the following Bayesian model specification:

$$\begin{aligned} \alpha_c &\sim \text{Normal}\left(\theta_{g[c]},\ \sigma^2_{\alpha_{g[c]}}\right) \\ \beta_c &\sim \text{Normal}\left(\omega_{g[c]},\ \sigma^2_{\beta}\right) \\ \theta_g &\sim \text{Normal}\left(\mu_\theta,\ \sigma^2_{\theta}\right) \\ \omega_g &\sim \text{Normal}\left(\mu_\omega,\ \sigma^2_{\omega}\right) \end{aligned} \qquad (4.2)$$

At the top level, the region model specifies a normal likelihood for the response variable $Y_{c,t}$ parameterised by $\alpha_c|\theta_{g[c]}, \mu_\theta, \sigma^2_{\alpha_{g[c]}}$. The model uses regional-level distributions to enhance parameter estimation at the country level. Each observation in the country c is assumed to vary randomly about a mean consisting of both population-level mean θ_g and a random-effect component μ_θ. At the next level,

the model expresses a prior belief that each group-level mean arises from a normal distribution with a mean given by the overall mean μ_θ, and variance σ^2_θ.

The following prior assumptions used to analyse this model are informed by prior information from the PISA data.

$$
\begin{aligned}
\mu_\theta &\sim \text{Normal } \left(500, 100^2\right) \\
\mu_\omega &\sim \text{Normal } \left(0, 50^2\right) \\
\sigma^2 &\sim \text{Truncated-t } \left(30, 10^{-2}, 1)T(0,)\right) \\
\sigma^2_\theta &\sim \text{Truncated-t } (30, 10^{-2}, 1)T(0,) \\
\sigma^2_\omega &\sim \text{Truncated-t } (0, 2^{-2}, 1)T(0,) \\
\sigma^2_{\alpha_g} &\sim \text{Truncated-t } (30, 10^{-2}, 1)T(0,) \\
\sigma^2_\beta &\sim \text{Truncated-t } (0, 2^{-2}, 1)T(0,)
\end{aligned}
$$

Figure 4.3 illustrates the regional hierarchical model in data space. This figure improves Figure 4.1 (where countries are ordered alphabetically) to order countries according to the regional structures used in the model fitting. Here, panel layouts and colours tailored to the regional structure of the data facilitate comparisons within each region. This approach also displays the regional fit (as determined by the hierarchical model hyper-parameters) superimposed on regional data, in the first column of each row.

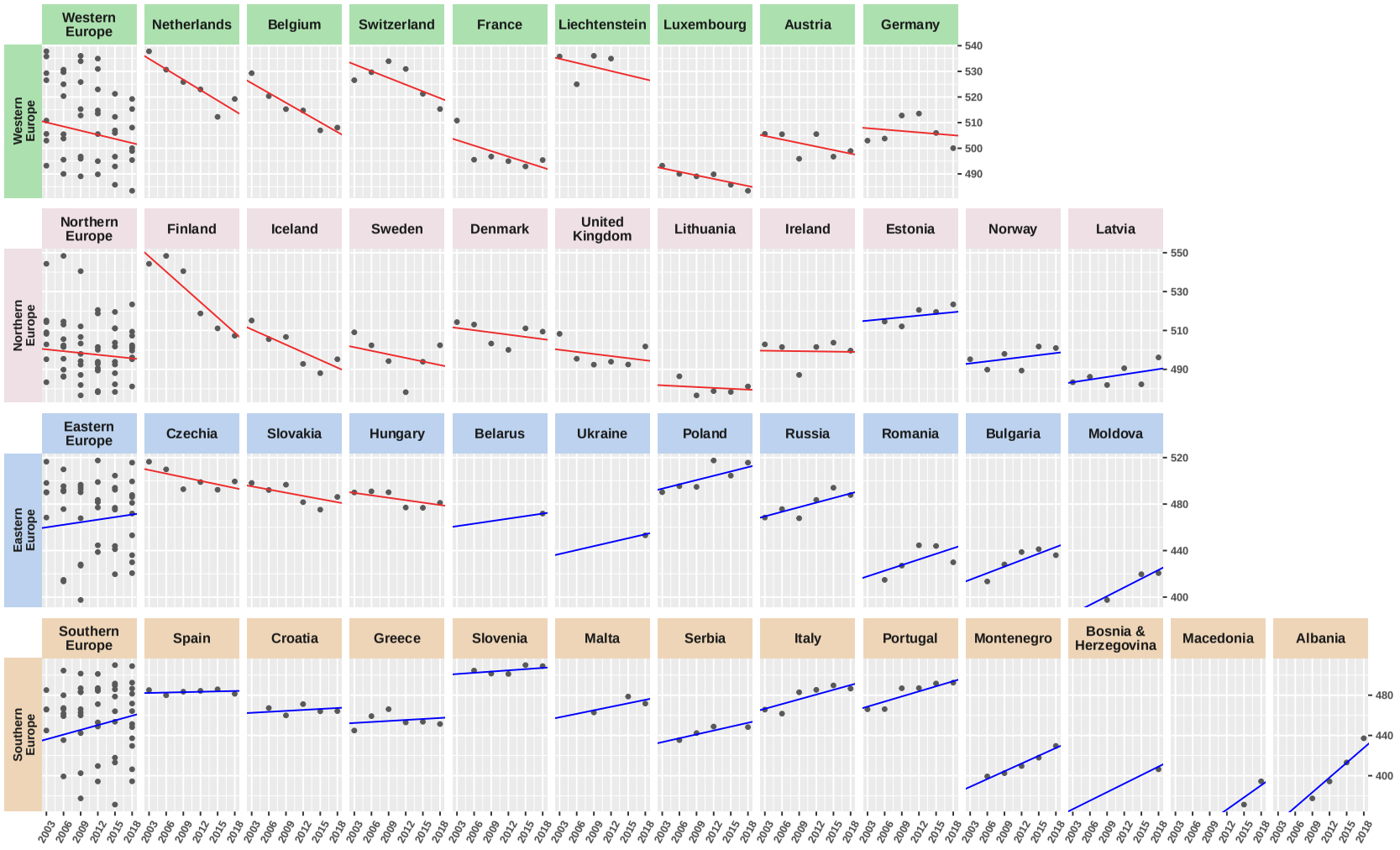

Figure 4.3: Regional hierarchical model fits to mathematics scores of PISA European data. Western and Southern European countries have a consistent negative and positive trend respectively, but trends are mixed for Northern and Eastern groupings.

We used the following principles in the design of Figure 4.3:

P1 - One panel per individual-level (unit): In Figure 4.3, like Figure 4.1, we use one panel to display the model result in data space per country, as the country level is the lowest level. This clearly reveals the country trends, allowing anomalies to stand out, such as the observation for Ireland in 2009, Liechtenstein in 2006 and Sweden in 2012. In the case of fewer countries or fewer countries per region, visualisation designs that overlay countries could be considered.

P2 - Panel grouping to show the model's hierarchical structure: In the realm of hierarchical modelling, our objective is to organise all individual level units according to the hierarchical structure of the model to facilitate comparison between and within groups. In Figure 4.3, each region appears in a separate row. Each panel, representing a country, is organised based on its respective geographical region. We see that mathematics scores in Western Europe are decreasing, while those in Southern Europe are increasing. In addition, this principle facilitates the identification of the mixed trends across countries in Eastern and Northern Europe, despite overall positive and negative trends in the Eastern and Northern European regions respectively.

P3 - Colour to differentiate groups and to enable comparison: In Figure 4.3, colours are used in two ways. Within each panel, to distinctly differentiate the direction of the model fit on observed data, negative trends are shown in red and positive trends in blue. Colour is also used alongside position to group countries, with labels for Western Europe in the light-green colour, Northern Europe with lavender, Eastern Europe in steel-blue, and Southern Europe in the bisque colour. The same region colours will be used consistently in other displays.

P4 - Panel ordering to enable model comparisons: In Figure 4.3, regions (rows) are ordered by increasing regional slope parameter estimates. Furthermore, within each region, countries are arranged by increasing slope. This arrangement places Finland first in the Northern Europe group because, while it has high mathematics scores, these scores dropped rapidly over time. Belarus and Ukraine are adjacent because, with one observation apiece, the slope displayed in their panels is the overall Eastern Europe slope (slope hyper-parameter estimate for the region).

P5 - Panel scaling for effective comparisons: Ideally, in displays such as Figure 4.3, panels would use the same scaling to facilitate comparison across countries. However, as the range of differences in Western and Northern Europe is small, this choice obscures trends within each country as most slopes appear

flat in these groups. Rather than free scaling across all panels, in this case we have opted to use the same scaling for the panels in each row and different scaling across rows. This choice enables easy comparison of countries within each region. However, comparison of mathematics scores across countries in different regions requires careful inspection of the row axes.

4.3.2 Examining a collection of multiple models

Displaying the model fit by region, as seen in Figure 4.3, noticeable patterns (positive and negative trends) emerge in Eastern and Northern Europe. To identify additional sources of variation in the dataset, we further classified countries based on their respective geographical region and income level and introduced another grouping structure termed "income-region", which combines region and income as a new variable, leading to three identified sources of heterogeneity and shared characteristics (region, income and income-region) to construct hierarchical models.

For comparison purposes, we consider five linear models fit to the PISA data: The first three models are: a non-pooled model, a country-specific hierarchical model where country grouping provides an additional source of variation, and a two-level hierarchical model where countries are nested in region and both region and country are sources of variation. The non-pooled model assumes all observations are independent, while the hierarchical models leverage group-level distributions to enhance parameter estimation at the country level. These five models are specified in `brms` as:

- **Model 1**

$$\text{math} \sim \text{year} * \text{country} \qquad \text{non-pooled}$$

with Bayesian model specifications:

$$\begin{aligned} Y_{c,t} &\sim \text{Normal}\left(\alpha_c + \beta_c t, \sigma^2\right) \\ \alpha_c &\sim \text{Normal}\ \left(500, 100^2\right) \\ \beta_c &\sim \text{Normal}\ \left(0, 50^2\right) \\ \sigma^2 &\sim \text{Truncated-t}\ \left(30, 10^{-2}, 1\right) T(0,) \end{aligned} \qquad (4.3)$$

- **Model 2**

$$\text{math} \sim \text{year} + (1 + \text{year}|\text{country}) \qquad \text{country model}$$

The country model specification is presented in Section 4.2 (Equation 4.1) and the model appeared previously in Figures 4.1 and 4.2.

- **Model 3**

$$\text{math} \sim \text{year} + (1 + \text{year}|\text{region}) + (1 + \text{year}|\text{country}) \qquad \text{region model}$$

The region hierarchical model specification is presented in Sub-section 4.3.1 (Equation 4.2) and the model was previously displayed in Figure 4.3.

The equations and model specifications for two further models (4) and (5) are the same as (3), except with the grouping variable region replaced by income and income-region respectively.

We chose the intercept and slope parameters to summarise and interpret the model results while examining a collection of multiple models simultaneously. Recall, our models are parameterised so that the intercept represents mathematics scores in 2018. In Figure 4.4, we display the median and 80%, 95% credible intervals of the intercept for each country alongside their respective hyper-parameter distributions from models (1) to (5). Similarly, Figure 4.5 presents the slope estimates. Comparing the estimates from various models highlights the shrinkage effects.

In Figure 4.4 and Figure 4.5, each model is represented by a vertical axis, with country-level estimates shown in darker shades and hierarchical (hyper-parameter) distributions in lighter shades alongside. Countries are represented by separate panels, carefully arranged by the map layout following the regional structure, with labels coloured to differentiate income categories. The panels are individually scaled due to the variation in estimates across countries, facilitating comparison of models within each country.

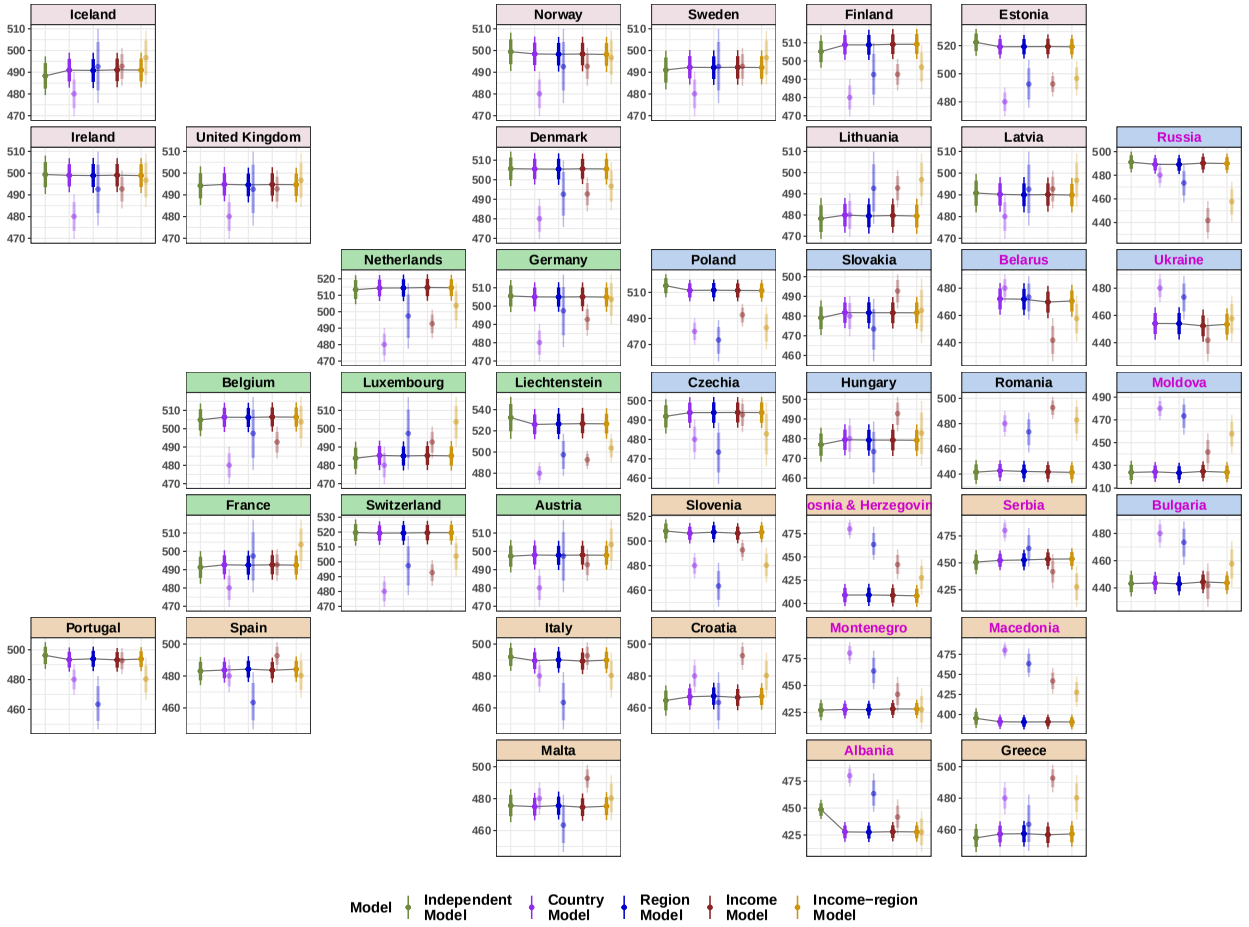

Figure 4.4: Intercept distributions (2018 median estimates with 80% and 95% credible intervals) from five model fits to mathematics scores from European PISA data. Hyper-parameter estimates are presented in lighter shades beside the parameter estimates. Countries are arranged according to the map layouts and labels are coloured to differentiate the income categories. Panels are individually scaled, because of the variation in estimates across countries.

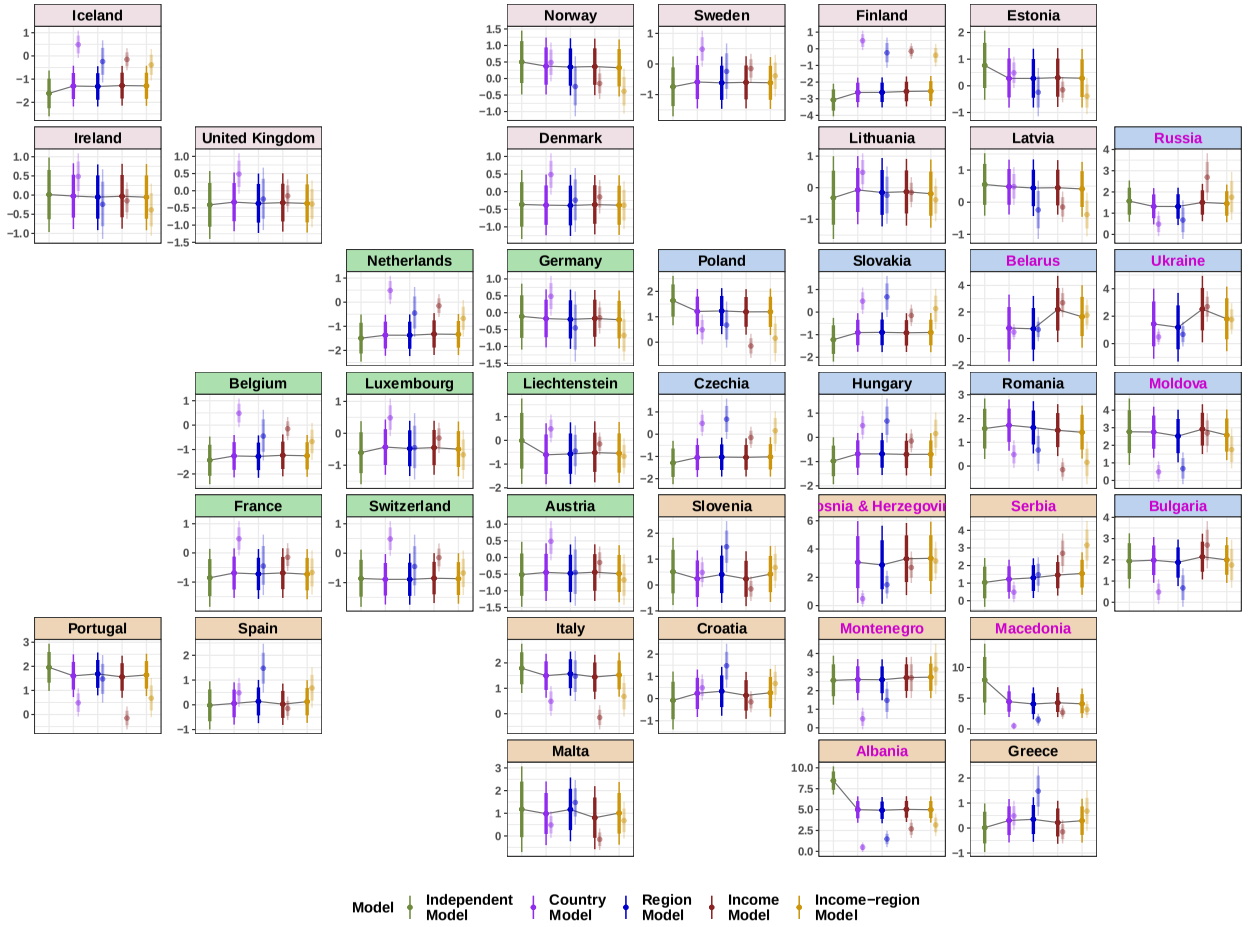

Figure 4.5: Slope distributions (median estimates with 80% and 95% credible intervals) from five model fits to mathematics scores from European PISA data. Hyper-parameter estimates are presented in lighter shades beside the parameter estimates. Countries are arranged according to the map layouts and labels are coloured to differentiate the income categories. Panels are individually scaled, because of the variation in estimates across countries.

The principles P1–P5 were also used in the design of our Figures 4.4 and 4.5:

P1 - One panel per individual-level (unit): As in Figure 4.3, one panel per country is used in Figures 4.4 and 4.5. Hence, each panel represents the country intercept/slope parameter distribution respectively using median, 80% and 95% credible intervals as distribution summaries. Models are arranged along the horizontal axis, in order of model (1) to (5). The country estimates are connected using a horizontal dashed line to differentiate them from the estimates from the higher-level groups.

P2 - Panel layout to show the model's hierarchical structure: Rather than arranging the countries by region as in Figure 4.3, we use a rough geographical map layout to arrange country panels, so that countries in the same region group are positioned nearby.

P3 - Colour to differentiate groups and to enable comparison: Figures 4.4 and 4.5 use colours in a number of ways: (1) to differentiate region groups and income groups, (2) to distinguish each model fit. The same region colour scheme used in Figure 4.3 was employed and colour is further used for the country label to indicate the income level, high income in black and middle income in magenta. Note that the middle income countries are mostly grouped together on the right hand side. In addition, we use one colour per model fit. Within each panel of Figures 4.4 and 4.5, the hierarchical structure parameter estimates (hyper-parameter) for the hierarchical models (2) to (5) are shown in a lighter shade positioned beside each model's parameter estimates. For model (2), the global intercept and slope distributions are in lightpurple, while lightblue, brown and darkgoldenrod are similarly used for models 3, 4 and 5.

P4 - Panel ordering to enable model comparisons: Within the panels of Figures 4.4 and 4.5, models are ordered from (1) to (5), going from the simplest non-pooled model (1), followed by the country-specific one-level hierarchical model (2), then the region hierarchical model (3) with two levels (country in region), and models (4) and (5) also with two levels (country in income and country in income-region) respectively.

P5 - Panel scaling for effective comparisons: In Figure 4.4, our main goal is to compare the intercept distributions from the five models within each country. Similarly, Figure 4.5 for the slopes. Therefore, we use free scaling across all panels, so that differences across the distributions for the models are visible in each of the countries.

The principles 1–5 underlying our visualisations are straightforward and aim to visualise model estimates that align closely with the model specification. Ideally, this means being able to compare estimates both within and across various hierarchical structures and to interpret these estimates while considering the effect of information sharing.

4.3.3 Insights from PISA model visualisations

Here, we will highlight key insights from the visualisations of PISA models in Figures 4.3, 4.4 and 4.5.

The structured arrangement of Figure 4.3 (showing model (3)), and the use of colours to indicate the direction of trends, promotes pattern recognition across countries, such that similar trends and disparities can be identified within the regional groups that govern the hierarchical structure of the model. For instance, the Eastern European countries Czechia, Slovakia and Hungary exhibit negative trends despite the overall positive trend in the region. Likewise in Northern Europe, Estonia, Norway and Latvia demonstrate positive trends over time notwithstanding the negative regional trend.

Within each row of Figure 4.3, countries are ordered by increasing slope, so we can easily identify countries with the most and the least decline in the region. In Western Europe, Netherlands and Belgium have a high rate of decline. Finland had the highest PISA score in 2003, but has the steepest decline across Northern European countries and indeed overall, though this last assessment is made more difficult by the different vertical axis scaling across the rows. Similarly, Albania is notable for its low initial PISA scores and by far the highest rate of improvement across the years.

Next we move to Figures 4.4 and 4.5, which compare five different PISA models. For all countries, there is little variation in the country intercept estimates (corresponding to PISA performance in 2018; Figure 4.4) across the four hierarchical models (2–5), even though the estimates for the groups depend on the choice of hierarchical model (see for example the group estimates for the Southern Europe middle income countries, whose labels have magenta text and bisque background). However, comparing the non-pooled model (1), we see that for Albania pooling produces a notable reduction in the intercept estimate, while for Finland and Iceland, pooling produces a small increase in the intercept estimate.

Looking at the slopes in Figure 4.5, pooling produces a notable reduction in the slope estimate for Albania and Macedonia, and small increases for Finland and Iceland. The non-pooled fit produces a practically zero slope estimate for Liechtenstein, while the hierarchical models estimate a reduction of about 0.5 per year. For most countries, slope estimates and credible intervals are not affected much by the choice of hierarchical model. There are some small differences in slope estimates evident for the middle income Eastern European countries Belarus and Ukraine, with the income model in brown giving higher slope estimates. This discrepancy is not surprising as these countries have just one observation apiece. For the other country with just one observation Bosnia & Herzegovina, the slope estimates are remarkably stable across models, though with a high-level of uncertainty across all models.

4.4 Investigating PISA 2022 Predictions

Recall that all model results displayed previously were based on PISA results up to 2018. It will be interesting to compare model-based predictions for 2022 with the the observed data available from the most recent PISA survey in 2022 [a]. A comparison of models 2–5 via leave-one-out cross-validation (LOO-CV Vehtari et al., 2024) shows little difference between models, agreeing with our findings from the visualisations of Figures 4.4 and 4.5. Therefore, we chose the richest hierarchical model for this investigation, which is the income-region model (5). In ideal circumstances this comparison could be used to assess the model's predictive accuracy. However, as we anticipate PISA results have been negatively impacted by COVID-19, this comparison will instead allow us to compare the level of impact across countries.

[a]The PISA test was originally due to be conducted in the year 2021 but was delayed by one year because of the COVID-19 pandemic. The exceptional circumstances throughout this period, including lock-downs and school closures in many countries, which led to the difficulties in collecting some data OECD (2023b).

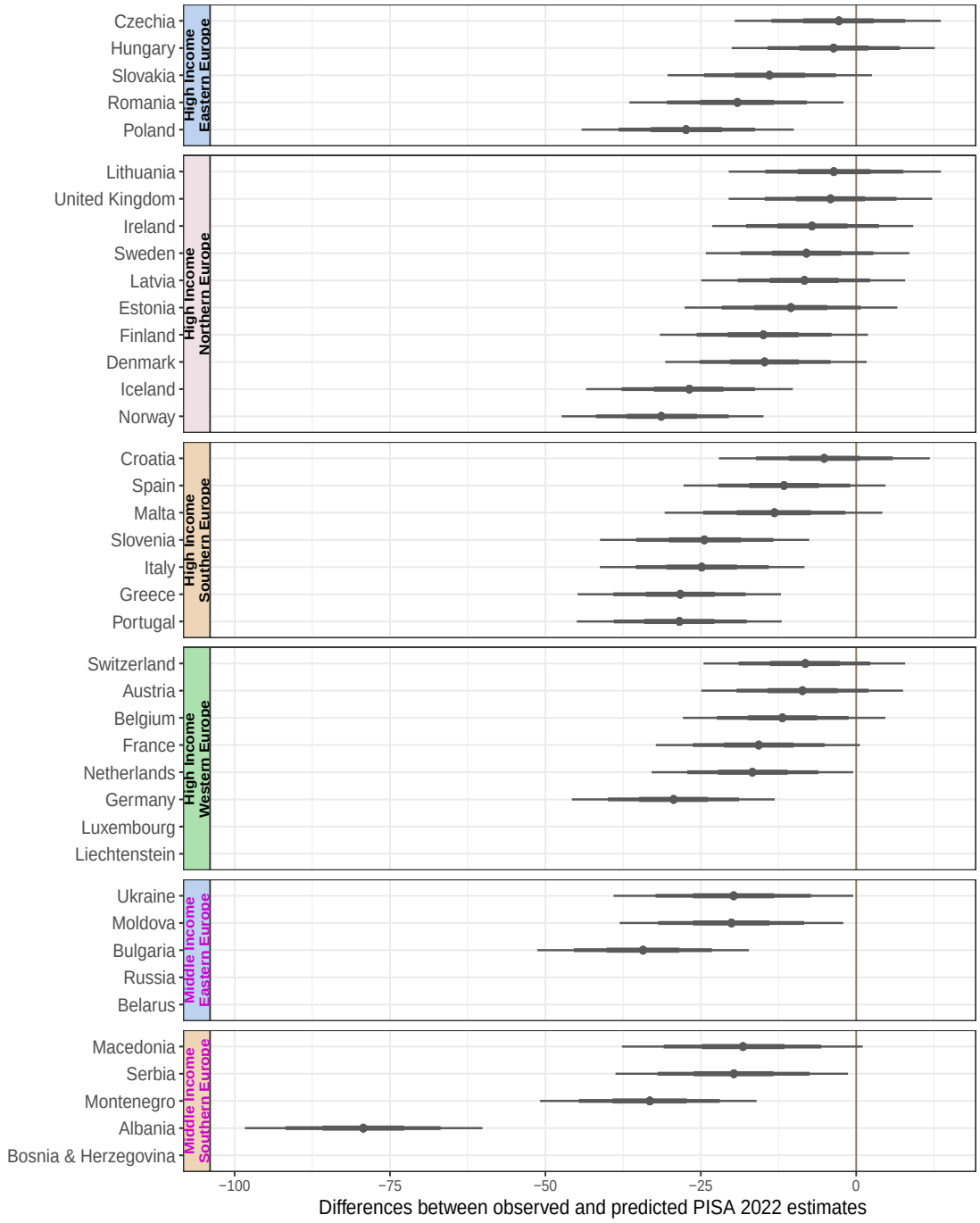

Figure 4.6: Differences between observed PISA maths data for 2022 and predictions from the income-region model (5), alongside 80%, and 95% credible intervals. Note that Luxembourg, Liechtenstein, Belarus, Russia, and Bosnia & Herzegovina did not participate in the PISA 2022 survey.

In Figure 4.6, we present the prediction error (observed - predicted) with the

50%, 80%, and 95% prediction intervals for the 35 out of 40 European countries that participated in the PISA 2022 survey. We arrange countries by their respective income-region group, order countries within each group according to their increasing prediction error and utilise the same region and income colour schemes as proposed in Section 4.3 to distinctly label the income-region groups.

In all countries the PISA 2022 values are lower than predicted. The country with by far the biggest drop is Albania, and in fact this 2022 score has wiped out all of the gains achieved since 2009. A possible reason for this decline, in addition to COVID-19, is that Albania experienced a strong earthquake in 2019 which caused damage to infrastructure, including schools and houses. It is also evident that PISA mathematics scores for middle income countries showed higher decline than those in high-income countries, suggesting that high-income countries were in a better position to mitigate the effect of COVID-19 on the school system.

4.5 Discussion

We have presented new visualisations and proposed informative principles for displaying hierarchical model estimates according to the strategies outlined by Wickham et al. (2015). The principles in Section 4.3.1 focused on the display of model results in the data space to provide an in-depth examination of the model fit in relation to the hierarchical structure of the model. We utilised layout, scales, colours, point sizes, and line types, to differentiate and emphasise patterns and to facilitate comparisons. The same principles (Section 4.3.2) offer a systematic approach for examining a collection of multiple models, facilitating comparisons of parameter and hyper-parameter distributions across models with different hierarchical structures.

In Section 4.3.3, we demonstrated that the new visualisations led us to identify countries with similar performance and those with unusual performance in PISA mathematics results. In this case study, the choice of hierarchical model is not crucial. For most countries, there is little to choose between models 2–5, except for a few countries with very little data. Applying the same visualisation principles in Section 4.4 to compare PISA 2022 mathematics results with model predictions established that across all countries, 2022 results were lower than expected (likely due to COVID-19), drastically so in the case of Albania, though overall, high income countries were less impacted.

The model fits and summaries presented in this chapter utilised the following

R packages, `brms` (Bürkner, 2018), `dplyr` (Wickham et al., 2023), and `tidybayes` (Kay, 2023). The visualisations were obtained using the following R packages, `ggplot2` (Wickham, 2016), `ggragged` (Marttila, 2023), `ggdist` (Kay, 2024), and `geofacet` (Hafen, 2024), all available from the Comprehensive R Archive Network (CRAN) at `https://cran.r-project.org/`. The `ggragged` R package is designed for creating arrays where rows can have different lengths and can be manipulated to suit desired arrangements, as in Figure 4.3. The `ggdist` R package is designed to present distributions and uncertainty, as seen in Figures 4.4 and 4.5. The package displays the median intercept and slope respectively along with their credible intervals using point intervals. The `geofacet` R package is designed for creating spatial visuals in a clear and comprehensible manner by arranging panels to mimic the geographical map layout, as in Figures 4.4 and 4.5.

The new visualisation designs are simple, flexible, adaptable and can be applied to non-Bayesian models, to models that fit relationships beyond linear regression, and to models with a richer hierarchical structure. In our illustrative example there are 40 countries under study, having a larger number of individual levels might result in clutter, which could be mitigated by selecting a smaller number of individual units for detailed examination. The visualisations presented used layouts based on spatial positioning of countries in Figures 4.4 and 4.5. However, in other applications these model comparison visualisations could use layouts based on groups such as those in Figure 4.3. We acknowledge that displaying hierarchical structures richer and more complex than those illustrated here could be challenging, but the principles described in 4.3.2 should provide helpful guidance.

For future work, these visualisation principles could be extended to richer hierarchical PISA models with covariates, most of the existing model visualisation literature (e.g., Heiss (2021); Heiss (2022)) focuses on models that include covariates, where interpretation is based on predictions, marginal effects, and parameter summaries. An additional extension would be to visualise the joint distribution of intercept and slope estimates in the context of hierarchical model structures, which would provide further insight into parameter uncertainty and correlation structure. Providing an R package that allows users to select and visualise relevant model parameters would also be a useful addition.

In conclusion, our principles for visualising hierarchical model estimates provide a comprehensive and insightful approach to understanding complex model structures. By following our step-by-step principles, researchers can achieve a clearer

understanding of hierarchical structures present in the data, and communicate results more effectively. Despite potential challenges, such as managing large numbers of individual levels and hierarchical structures, our approach remains highly valuable for examining and comparing hierarchical models. Through these visualisation principles, we contribute to advancing the field of hierarchical modelling, offering tools and insights that can be adapted to a wide range of research contexts.

5

Using Time Series Measures to Explore Family Planning Survey Data and Model-based Estimates

5.1 Introduction

The right of any individual to decide freely and responsibly on the number and spacing of their children has long been recognised as a fundamental human right (Bracke, 2022) and is central to improving lives and building equitable societies (United Nations, Department of Economic and Social Affairs, Population Division, 2019). This positions improvements to family planning as a cornerstone of progress as reflected in Sustainable Development Goal indicator 3.7.1: "Proportion of women who have their need for family planning satisfied with modern methods". Family planning is therefore a key global development priority, integral to reproductive health services and broader socio-economic development (Bracke, 2022). A range of family planning programmes operate at both global and national levels to expand access to reproductive health services, strengthen health systems, and promote informed contraceptive choice. Among these initiatives is Family Planning 2030 (FP2030) FP2030 (2025), which supports the global commitment to achieving universal access to family planning by 2030.

FP2030 evolved from the FP2020 initiative, launched at the 2012 London Summit on family planning with a vision tagged "120 by 20" numeric vision, which sought to enable 120 million additional women and girls access to modern contraception by 2020 (Hardee and Jordan, 2021). As the FP2020 initiative concluded

in 2020, it gave way to FP2030 with a broader vision to promote the right of women and girls to lead healthy lives and make informed choices about contraception (FP2030, 2025). This is an initiative that is closely aligned with the global commitment of universal access to sexual and reproductive health under SDG 3.7.

Despite the global commitment to family planning, progress varies widely across countries and regions, partly due to differences in access, program implementation, and socio-demographic factors (Kuang and Brodsky, 2016). To monitor family planning progress, data are collected through household surveys such as, the Demographic and Health Survey(DHS), Performance Monitoring for Actions(PMA), UNICEF Multiple Indicator Cluster Surveys (MICS), national surveys, and other surveys (United Nations, Department of Economic and Social Affairs, Population Division, 2019). These household surveys are typically conducted every 3 to 5 years, resulting in temporal gaps of up to 6 years for certain countries. In some rare cases, national surveys are conducted annually or semi-annually (Cahill et al., 2021).

Temporal gaps in large-scale survey data make consistent monitoring of family planning progress challenging. To address this limitation, the Family Planning Estimation Model (FPEM) was developed (Alkema et al., 2013; Cahill et al., 2018). FPEM produces annual survey-informed estimates and projections for key family planning indicators using Bayesian hierarchical time series inference. Where survey data are available, the model interpolates estimates and it projects estimates for data gaps between infrequent survey observations. FPEM employs Bayesian hierarchical inference to share information across hierarchical structures, countries nested in subregions, subregions in regions, and regions within the global level. The model estimates are smoothed by design, because they tend to reflect the hierarchical overall behaviour, which makes it important to explore how these modelled trends align with the underlying survey observations.

This chapter examines the extent to which modelled estimates from FPEM align with the underlying survey data, using selected diagnostic indices from the `wdiexplorer` package of 3. These indices are time series measures designed to quantify variation, trend, shape and sequential temporal features of country-level panel data. By using `wdiexplorer` to jointly analyse the survey observations and the corresponding modelled trajectories, as well as evaluating the residuals, the analysis identifies where the model aligns with the observed data and where notable differences arise. This facilitates the identification and comparison of temporal behaviours of the data and the associated model trends, enhancing understanding of

patterns in modern contraceptive use across countries. All analyses and visualisations presented in this chapter are fully reproducible and available on GitHub at https://github.com/Oluwayomi-Olaitan/Family-Planning-Exploratory-Analysis.

The remaining sections of this chapter are organised as follows. Section 5.2 provides background to the study, outlining the family planning survey data, the model-based estimates, and the `wdiexplorer` methodology. Section 5.3 describes the data processing steps for the survey dataset and the dataset of estimates and presents an initial visualisation of the survey data. Section 5.4 introduces the exploratory analysis of the survey data and model-based estimates using selected `wdiexplorer` diagnostic indices. Section 5.5 presents the residual-based exploratory analysis. Finally, Section 5.6 concludes with the findings, benefits of assessing how well the model-based trends align with the underlying survey observations, the limitations of the diagnostic indices in the context of family planning data, and directions for future work.

5.2 Background

In this section, we discuss the family planning survey data, the model-based estimates, and our `wdiexplorer` methodology. The survey and model-based data were sourced from the UNPD, while `wdiexplorer` offers a workflow with time-series diagnostic measures and visualisation tools for exploring these measures to identify patterns, outliers, and other notable features.

5.2.1 Family Planning Survey Data

According to Hubacher and Trussell (2015), modern contraceptive methods include medical products or procedures such as pills, intrauterine devices, implants, condoms, and sterilisation, which are designed to reliably prevent pregnancy. Traditional contraceptive methods include behavioural and natural approaches, such as withdrawal, rhythm methods, douching, and folk methods. The prevalence of either modern or traditional contraceptive methods is defined as the proportion of the target population, typically women of reproductive age (15 - 49) using that method at a given time relative to the total population of the same group. These proportions (prevalence of modern and traditional contraceptive use) are the primary variables

of interest for our analysis.

These proportions alongside other demographic variables are captured through family planning surveys. A collation of family planning survey data is organised and maintained by the UNPD in the World Contraceptive Use database (United Nations Department of Economic and Social Affairs, Population Division, 2024) and updated annually, providing country-level observations of family planning indicators to support the monitoring of family planning progress. The dataset provides survey observations of contraceptive use (modern and traditional) and unmet need, obtained from nationally representative household surveys, including Demographic and Health Surveys (DHS), UNICEF Multiple Indicator Cluster Surveys (MICS), Performance Monitoring for Action (PMA), Reproductive Health Surveys, Contraceptive Prevalence Surveys (CPS), national surveys, and other sources (Cahill et al., 2018).

The dataset spans 197 countries, including Taiwan and Kosovo, which are not universally recognised as sovereign nations, and covers 1950 to 2021, although coverage and completeness vary across countries, time, and indicators. Survey observations are disaggregated by data series type, capturing the survey source, union status, indicating whether individuals are married/in a union or not, alongside other demographic characteristics and measures of observation uncertainty in the form of standard errors.

5.2.2 Family Planning Model-based Estimates

The Family Planning Estimation Model (FPEM) is a Bayesian hierarchical time series model used by the UNPD to generate annual estimates and projections of key family planning indicators (Alkema et al., 2013; Cahill et al., 2018) using survey data as input. The model leverages the Bayesian inference framework, where country-level estimates are informed by subregional, regional, and global trends. This structure allows reliable projections even in countries with sparse data, as information is effectively borrowed from nested hierarchical levels. Its structure consists of a process model that captures long-term changes in contraceptive prevalence through logistic growth curves reflecting the realistic assumption that growth accelerates at intermediate levels of use and slows as prevalence reaches saturation (Alkema et al., 2013; Cahill et al., 2018).

FPEM serves as the core family planning estimation framework within the

UNPD, providing projections for family planning indicators and is implemented in the family planning estimation tool for country-level use, available in the R package `fpemlocal` (Guranich et al., 2021). While FPEM is central, there are other methodologies employed to complement it in assessing program inputs and impact and setting strategic targets.

The model estimates of family planning indicators are available on the UNPD website (United Nations Department of Economic and Social Affairs, Population Division, 2024) and the dataset contains estimates from 1970 to 2030. There are estimates for 14 indicator levels, with both numerical counts and percentages for contraceptive prevalence, demand for family planning, and unmet need, classified by method type (any, modern, traditional). Each record is associated with a *LocationID*, corresponding to the division numeric code and *Location*, country name, with *Time* representing the year. The dataset also includes *Variant* that captures the estimates and their credible intervals, by 95% lower bound, 80% lower bound, median estimate, 80% upper bound, and 95% upper bound. It contains the *Age range* information, *Categories*, with 3 population groups ("All women", "Married or in a union women", and "Unmarried women"), and *Estimate method* which indicates whether the estimate is interpolated for observed data periods or projected for periods without data.

These model-based annual estimates serve as a dataset to monitor progress toward global family planning goals, such as the FP2030 initiative, and other key family planning targets. The dataset provides consistent and comparable estimates across countries, sub-regions, and regions over time.

5.2.3 wdiexplorer: Time Series Diagnostic Measures

To assess and explore the family planning survey data related to contraceptive use and corresponding model-based annual estimates generated by FPEM, we apply our `wdiexplorer` R package to these two datasets. `wdiexplorer` is a time series exploratory tool originally designed for exploratory analysis of country-level panel data of the World Development Indicators (WDI), as the name implies. It can also be applied to any dataset formatted as repeated observations over time and across countries. The package provides a structured workflow for data exploration through diagnostic indices and visualisation tools. By explicitly incorporating predefined grouping structures, such as geographical region, income, or other

categories, `wdiexplorer` facilitates the identification of patterns, outliers, and variations within and across group levels, as well as temporal and shape-based features.

In `wdiexplorer`, we use a three-stage process designed for the exploratory analysis of country-level panel data. The first stage involves obtaining data from the WDI, or another source, and assessing its completeness and validity. This stage uses visual summaries to assess data availability and missingness across countries and years, grouping countries by a predefined variable to compare patterns within and across groups. The second stage computes a set of ten diagnostic indices that quantify key characteristics of the data series. Inspired by the scagnostics framework (Tukey and Tukey, 1985; Wilkinson and Wills, 2008) and some established time series features (Hyndman and Athanasopoulos, 2021), these indices capture variation, shape and trend, and sequential temporal behaviours of the data. The final stage utilises visual tools to identify outliers, patterns, potentially interesting features and to support interpretation of the diagnostic indices.

The `wdiexplorer` package was originally designed for exploratory analysis of WDI data, one indicator at a time. We adapt some of the diagnostic indices and visualisation tools of `wdiexplorer` to demonstrate how it can also be used to assess consistency, variation, and comparison between observed data and model-based estimates of any country-level panel data. Using family planning survey observations and the corresponding modelled estimates as an example, we show how different functionalities of the package are used depending on whether the focus is on exploring the data and model outputs or the residuals.

The selected diagnostic indices for exploring the family planning datasets are silhouette width and trend strength. Silhouette width (Rousseeuw, 1987) is typically used in cluster analysis to assess the performance of groups obtained from a clustering procedure. Here, we use it to assess a predefined grouping of countries. It evaluates how well a country's data aligns with its predefined group compared to other groups. It is computed by comparing the average distance to countries within the same group against the distance to the closest neighbouring group. The silhouette width ranges from -1 to $+1$. A value near $+1$ indicates that the country is well aligned within its assigned group, while a value close to -1 suggests the country is closer to an alternative group. Silhouette width in our context quantifies the degree to which the contraceptive use pattern of each country align with its sub-region in both the survey data and the modelled output.

Trend strength quantifies the extent to which a data series follows a consistent

pattern over time, which could be linear or curved, relative to random fluctuations. The trend strength is defined as the ratio of the variance of the remainder component (R_t) to the variance of the combined sum of the trend and remainder components ($T_t + R_t$) from a time series decomposition (Hyndman and Athanasopoulos, 2021). The value ranges from 0 to 1 expressed in proportions, distinguishing smooth patterns from irregular noise.

To assess model fit, we use the diagnostic indices linearity and curvature to analyse the model residuals. Linearity measures how closely a series follows a straight-line pattern over time. It assesses whether changes occur consistently without sharp curves or bends while curvature measures the extent to which a series deviates from a straight-line trend over time, capturing the presence of non-linear patterns; upward or downward bends. Linearity is quantified as β_1, the coefficient of the linear term and curvature is quantified as β_2, the coefficient of the quadratic term in a polynomial regression fitted to the trend component of the decomposed series, where the predictors are the first two orthogonal polynomials of the time index (Hyndman and Athanasopoulos, 2021). If the model perfectly captures the patterns in the data, the residuals for each country should behave randomly, and the corresponding linearity and curvature metrics should be zero or close to zero.

Other diagnostic measures of `wdiexplorer` are: country average dissimilarity, within group average dissimilarity, smoothness, number of crossing points, longest flat spot and autocorrelation. These measures are less suitable for the family planning datasets for several reasons. The infrequent nature of the survey data with intervals of 3–5 years limits the comparability of temporal measures such as the number of crossing points or longest flat spot with the model-based annual estimates. Silhouette width and trend strength diagnostic measures are considered most suitable for capturing the behaviour of family planning data. We proceed to explore these two datasets, compare their diagnostic measures, and examine patterns in the residuals using linearity and curvature indices.

5.3 Preliminary Data Exploration

5.3.1 Data Processing

The family planning datasets were processed to align with the requirements of the `wdiexplorer` workflow, which expects a single observation per country-year pair.

Survey data on contraceptive use and other family planning indicators were collected from multiple survey types. In 41 countries, there are multiple survey observations per country-year pairs. To address this, a yearly observation is obtained from the surveys prioritised in the following order: DHS first, followed by MICS, PMA surveys, national surveys, and other sources.

We prioritise survey types in this order because DHS is generally considered the most reliable and widely comparable, serving as a global survey for health data. Analyses using FPEM have demonstrated that the random errors associated with non-DHS data are greater than those associated with DHS data (Alkema et al., 2013). MICS and PMA follow because they are also nationally representative and rigorous, while national surveys and other sources are used only when higher-priority data are unavailable, acknowledging their potentially greater measurement error and variability. Despite prioritising survey types to obtain one observation per country–year pair, 16 country-year pairs across 12 countries have two observations of the same survey type. For each country-year pair with two observations for the same survey type, the mean of the proportions was computed and used as the representative value for such pair.

For the analysis, we select the 85 focus countries of the FP2030 initiative, previously identified as the 69 priority countries of FP2020. These countries were characterised as focus countries because they represent populations with significant barriers to accessing family planning services and have the largest gaps in meeting demand for modern contraception (United Nations, Department of Economic and Social Affairs, Population Division, 2019). Many are located in Western, Eastern Africa and South-Central Asia, predominantly identified by their socioeconomic status, consisting of low and lower-middle income countries. Table A.1 (see Appendix) presents the 85 focus countries along with the total number of surveys conducted and the year of the most recent survey. The dataset spans five world regions, with Africa having the largest number of countries (48) across five sub-regions, and Europe having the fewest, with only one country in Eastern Europe.

We further limited the dataset for this analysis to survey observations from 1990 through the most recent available survey year. Prior to 1990, family planning programs were primarily focused on reducing fertility and population growth, and the corresponding survey data reflect these early program goals rather than the rights-based perspective on reproductive health and well-being (RamaRao and Jain, 2015). We also limited the analysis to the population of women who are married

or in a union. This group is often referred to as Married/In-Union Women of Reproductive Age (Cahill et al., 2020).

5.3.2 Visualisation of Family Planning Survey Data

Chapter 4 outlined five principles for effective visualisation of panel data: (1) one panel per individual-level (country), (2) panel grouping to represent hierarchical structures, (3) the use of colour to distinguish and compare hierarchies, (4) panel ordering to facilitate comparison, and (5) consistent panel scaling for effective comparison and interpretation. These principles guided our preliminary exploration of the family planning survey data. We applied them by constructing stacked plots of modern and traditional contraceptive use, with each panel representing a country. The panels were grouped according to the UNPD sub-regional classifications and ordered by their average total contraceptive use (modern + traditional) for their most recent survey year. Countries within each sub-region were ordered by their total contraceptive use for the most recent survey year. We maintained consistent panel scaling to ensure that the differences between countries were meaningfully represented.

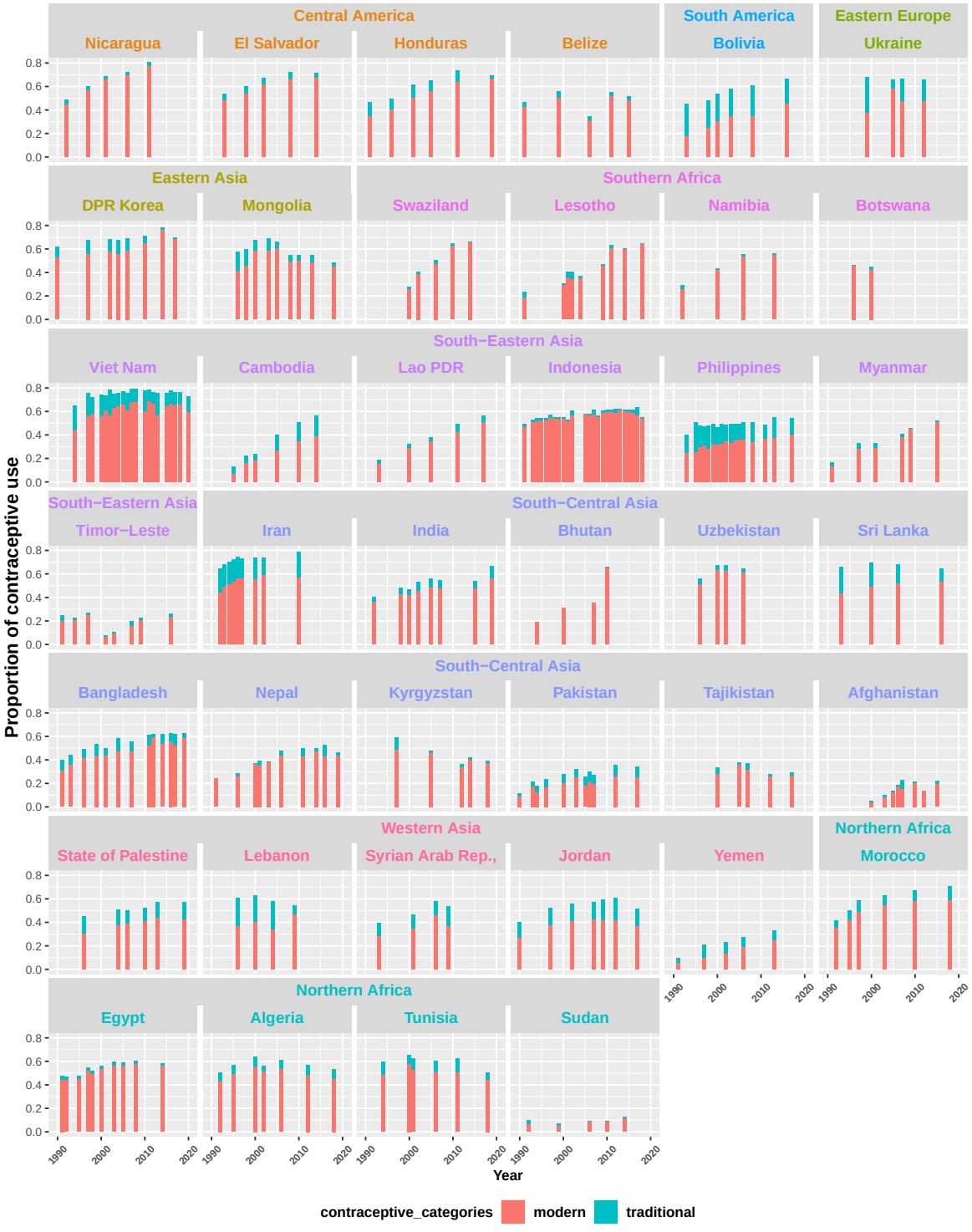

Figure 5.1: Stacked bar plot of modern and traditional contraceptive use across countries in the top nine sub-regions. Countries are grouped by sub-region, with sub-regions ordered by their average total (modern + traditional) contraceptive use in the most recent year and countries within sub-regions are also ordered accordingly. Central America sub-region has the highest total contraceptive use.

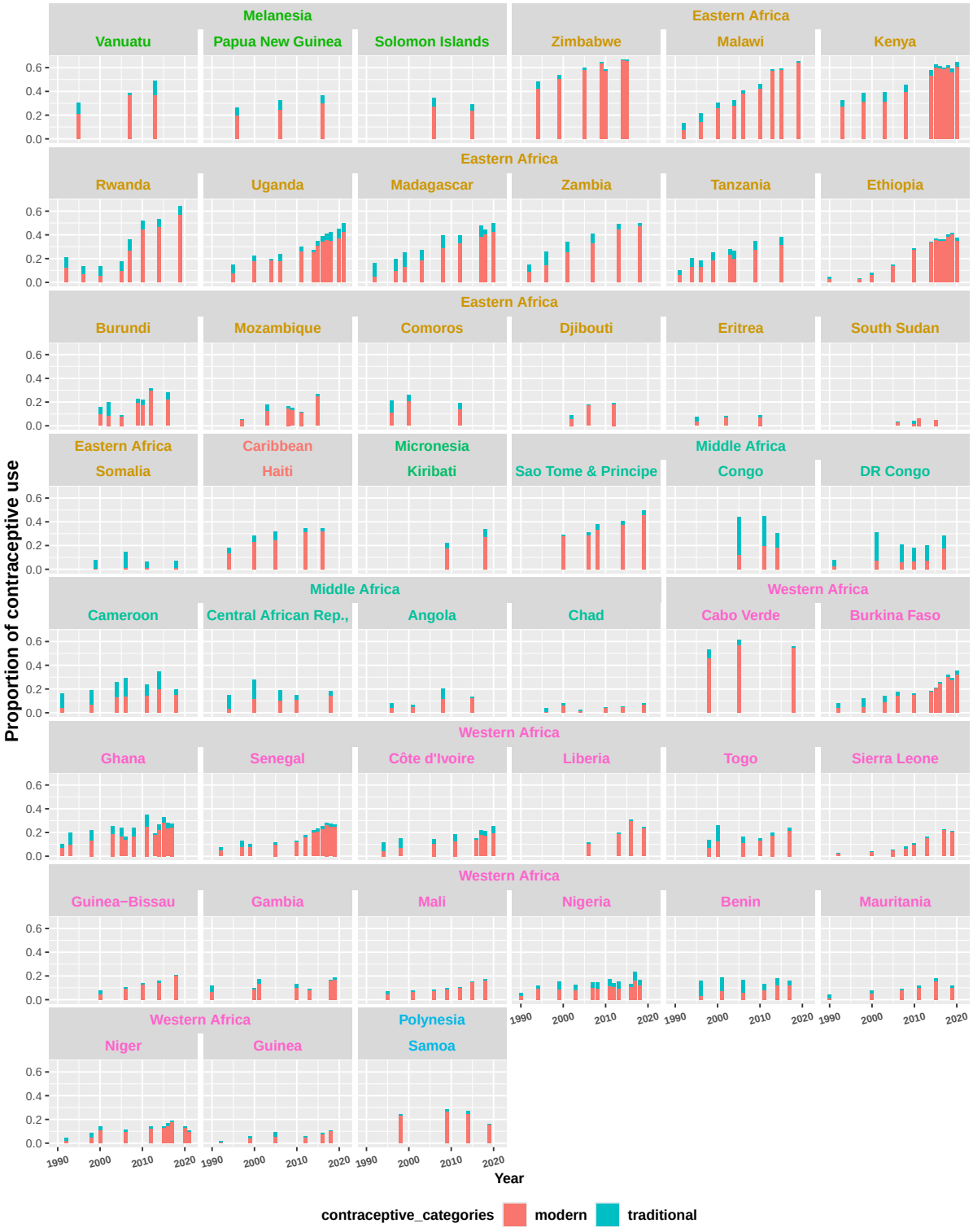

Figure 5.2: Stacked bar plot of modern and traditional contraceptive use across countries in the remaining seven sub-regions. Sub-regions and countries are ordered by average total contraceptive use in the most recent year. Polynesia has the lowest total contraceptive use.

Figures 5.1 and 5.2 present the proportions of modern and traditional contraceptive use among FP2030 focus countries. Figure 5.1 displays the top nine sub-regions with forty countries ordering sub-regions by their average total (modern + traditional) contraceptive use in the most recent year and countries within sub-regions are also ordered according to their average total (modern + traditional) contraceptive use in the most recent year. Figure 5.2 displays the remaining seven sub-regions with forty-five countries ordering sub-regions by their average total (modern + traditional) contraceptive use in the most recent year and countries within sub-regions are also ordered accordingly.

At the sub-regional level, Central America records the highest average contraceptive use, followed by South America and Eastern Europe. South America and Eastern Europe are each represented by a single FP2030 focus country, Bolivia and Ukraine, respectively. There are noticeable differences in total contraceptive use within and across sub-regions. In the Eastern Asia subregion there are just two FP2030 focus countries, the Democratic People's Republic of Korea and Mongolia, both of which have high total contraceptive use. In South-Eastern Asia, a few countries exhibit consistently high total contraceptive use, with Vietnam being the highest. In South-Central Asia, Iran records the highest total contraceptive use in its most recent survey year. Likewise, Morocco in Northern Africa, Zimbabwe in Eastern Africa record the highest total contraceptive use in their most recent survey year.

Figure 5.1 also reveals variation in survey frequency across countries, as recorded in Table A.1 (see Appendix). Vietnam, Indonesia, and Philippines, countries in South-Eastern Asia, display the longest periods of annually recorded survey data. This indicates that the total number of surveys conducted in these countries between 1990 to 2021 exceeds that of any other FP2030 focus country. In contrast, some countries exhibit only short periods of annual data collection. For example, Lesotho has annual survey data between 2000 and 2002 only. Other countries, particularly countries in Central America and Melanesia sub-regions, show consistent gaps of 3 to 5 years between surveys. Beyond variation in survey frequency, Figures 5.1 and 5.2 show that some sub-regions include only a single FP2030 focus country, as the sole population in the sub-region facing significant barriers to accessing family planning services: Bolivia in South America, Ukraine in Eastern Europe, Haiti in Caribbean, Kiribati in Micronesia and Samoa in Polynesia.

Generally, in Figures 5.1 and 5.2, the bar plot reveal that modern contraceptive

use accounts for a larger share of total contraceptive use than traditional methods. However, exceptions are observed in some countries where the proportion of traditional contraceptive use is higher. For example, the proportion of traditional contraceptive use was higher in Bolivia during its first survey year (1993), likewise all survey years in Somalia, a country in Eastern Africa. Similar pattern occurs in initial survey years across countries in Middle Africa sub-region, as well as some survey years in countries of the Western Africa sub-region. This pattern is consistent with findings reported by United Nations, Department of Economic and Social Affairs, Population Division (2019) which indicate that some countries report higher proportion of traditional contraceptive use over modern revealing barriers to accessing family planning services.

According to the results of United Nations, Department of Economic and Social Affairs, Population Division (2019), the median prevalence of modern contraceptive use among women of reproductive age (15–49 years) in 2019 was 44.3% (95% uncertainty interval: 42.1 – 47.0%). Traditional methods accounted for approximately 4.2% of use, calculated as the difference between overall contraceptive prevalence (48.5%) and modern methods. Marquez et al. (2018) also emphasised that modern contraceptive methods are more reliable than traditional methods and are positioned higher in the hierarchy of contraceptive effectiveness. Backed by these findings, the subsequent analysis focuses on modern contraceptive use.

5.4 Diagnostic Indices of Survey Data and Model-based Estimates

FPEM relies on hierarchical Bayesian inference, which allows the model to share information across countries, sub-regions, regions and globally. The model produces smoothed estimates, and the hierarchical structure is especially useful where sparse data poses a challenge. We assess not only how well the Bayesian model captures the temporal and variational patterns observed in the survey data, but also where hierarchical pooling influences the estimates. Using diagnostic measures such as trend strength and silhouette width from the collection of diagnostic indices of the `wdiexplorer` package, as described in Subsection 5.2.3, we compare survey data and model-based estimates, and identify countries where deviations are most pronounced, highlighting contexts in which the model hierarchy has a substantive

impact.

5.4.1 Trend Strength Measure

Trend strength is one of the diagnostic indices of the `wdiexplorer`, under the category of trend and shape features. It quantifies the degree to which a series follows a consistent temporal pattern, whether linear or non-linear, relative to random fluctuations. Trend strength is the proportion of variance in the data explained by the variance of the remainder component of a decomposed time series relative to the variance of the sum of the smoothed trend component and the remainder component (Hyndman and Athanasopoulos, 2021). A series is considered to have a strong trend when the variance of the remainder component is small relative to the total variance. A weakly trended series has a large remainder variance, reflecting greater irregular fluctuations.

To evaluate the strength of trends in the family planning survey data and model-based estimates, we calculate trend strength of each country in both datasets using the `wdiexplorer` package. These metrics were compared by calculating the ratio of model to survey trend strength.

We present a scatterplot of the model to survey trend strength ratio versus the survey data trend strength for each of the 85 FP2030 focus countries in Figure 5.3.

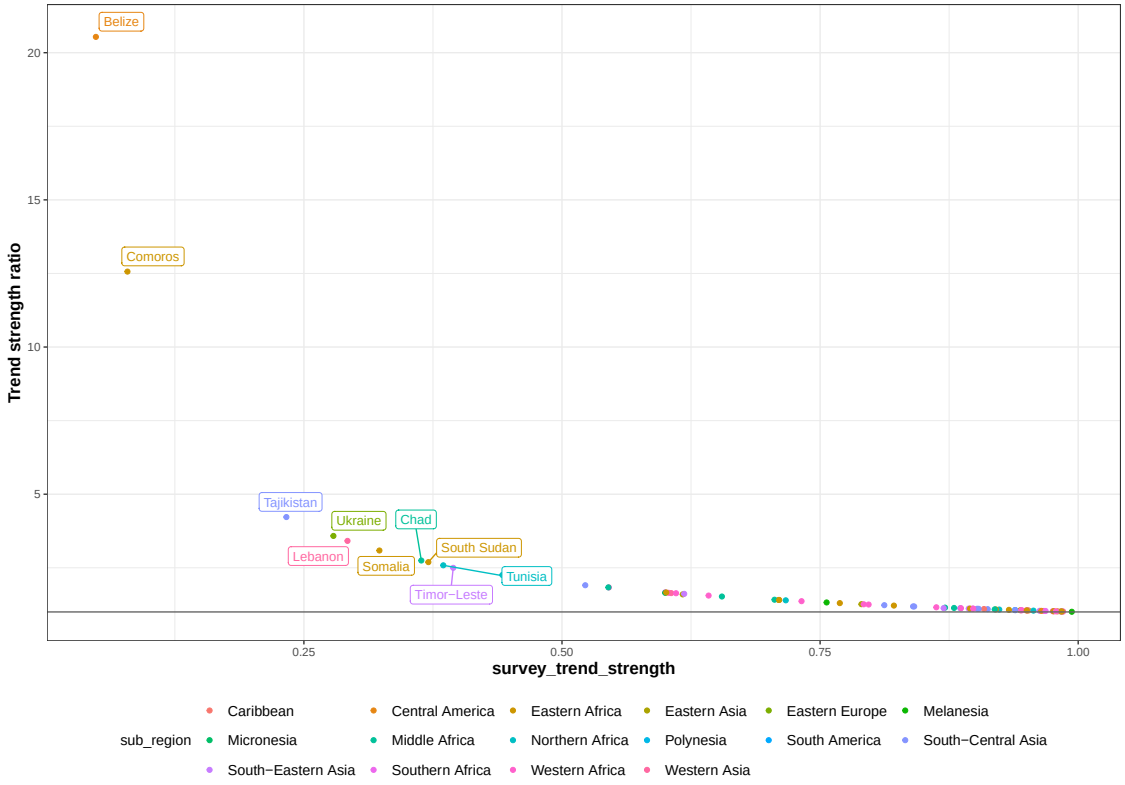

Figure 5.3: Comparison of trend strengths in survey data and model-based estimates across countries. Each point represents the ratio of model-based to survey trend strength plotted against the survey trend strength for a country. All countries have ratios above 1, indicating that the model-based trends are stronger than the survey trends. Labelled points correspond to countries with extreme ratios of model-based to survey trends strength. Hovering any point in the interactive version reveals the corresponding country name, its survey and model trend strength.

Figure 5.3 shows that across all 85 focus countries, the trend strength ratio exceeds 1, indicating that the model-based trend strength is stronger than that derived directly from survey data. This pattern is consistent with the smoothing effect of FPEM. Smoothing increases apparent trend strength because it removes short-term noise and measurement error, borrows information from neighbouring time points to reinforce consistency, assumes gradual continuous change rather than abrupt shifts, and down-weights outliers; together these effects suppress variability in the raw data and make the underlying long-term pattern appear clearer, more consistent, and stronger than it does in the observed data points. Labelled points correspond to the top ten countries with the largest ratios, pointing to cases where the model estimates differ most from the observed survey trends. We further exam-

ine these countries by comparing their survey data points with their corresponding model-based trends.

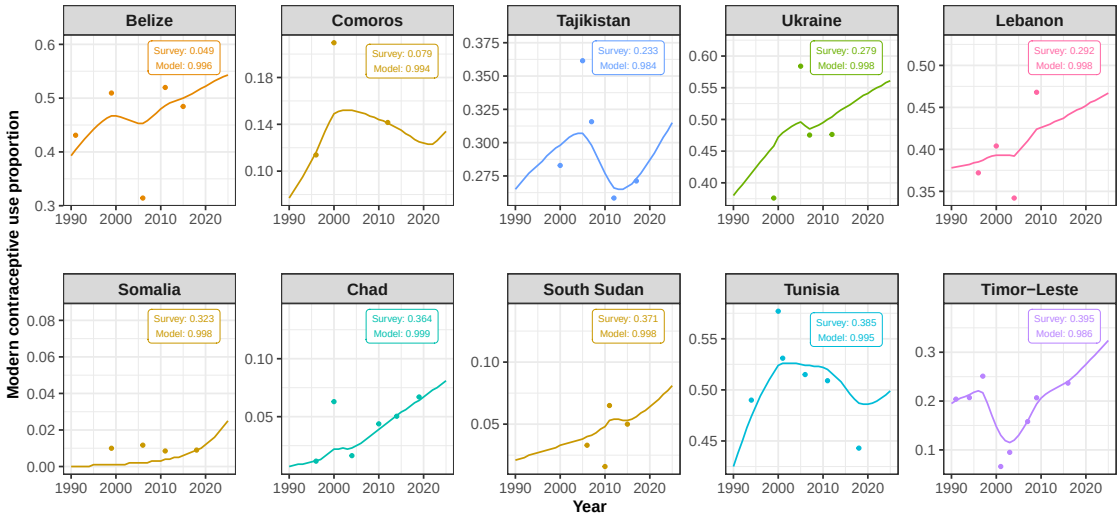

Figure 5.4: Survey data trajectories and model-based trends for the top 10 countries with the largest ratios of model-based to survey trend strength. Each panel presents the survey observations and corresponding model trajectory of a country, ordered by their ratios. Belize shows the highest ratio, reflecting a dip in its 2006 smoothed out by the model. Similarly, Comoros exhibits a sharp spike in 2000, with an observed proportion of modern contraceptive use far higher than other years.

For the 10 countries with the largest deviations, Figure 5.4 displays their survey data points alongside the corresponding model-based trends over time. Each country is represented by a separate panel, and countries are ordered in descending order of the ratio value. We ensure panel-specific scaling to accurately present the fluctuations in data points and trends for each country. Figure 5.4 illustrates how FPEM smooths survey fluctuations in the top 10 countries with the highest ratios. The smoothed curves are designed to highlight the underlying signal by filtering out some of the noise in the data. That makes the model based long-term trends look stronger and clearer than they appear in the raw observed data.

All 10 countries have trend strength values calculated from the modelled estimates that are very close to 1, as one might expect. The differing ratios reflect variation in the trend strength of the observed data. Belize has a very low trend measurement for the survey observations, and its panel shows that the proportion of modern contraceptive use decreased substantially in 2006 compared to previous years, followed by higher values in subsequent surveys (2011 and 2015). This dip

in the survey data is weakly captured in the model-based trend, as the model assumes gradual, continuous change over time rather than abrupt shifts. Comoros has a similarly low trend measurement for the survey observations, but based on just three observations, where the middle one is unusually large.

Other countries with high ratios, namely Tajikistan, Ukraine, Lebanon, and South Sudan, have sparse data, which does not follow a consistent upward or downward trend. In Somalia, the fitted trend exhibits bias and lies consistently below the observed survey points. In Figure 5.2, Somalia exhibits the lowest proportion of modern contraceptive use across all survey years, both within its sub-region and relative to other sub-regions. Given this persistent deviation, one might expect the model-based trend to be partially pulled toward the sub-regional pattern, but the trend remains low, due to the influence of the low survey values.

The high trend scores for the models relative to the surveys across the 10 countries are attributed to FPEM including temporal smoothing and information sharing across countries within its hierarchical structure. Note also, while we rank survey types and select the data point from the highest-ranked type for each country-year during our survey data processing, FPEM uses all available survey types to generate model estimates. Together, these modelling choices reduce the influence of survey noise, resulting in smoother trajectories and stronger apparent trends in the model-based estimates compared with the survey data.

5.4.2 Silhouette Width Measure

Silhouette width is an output of the silhouette analysis and is one of the variation features offered by `wdiexplorer`. It quantifies how well a data series align with its assigned group relative to other groups (Rousseeuw, 1987). The silhouette value ranges from -1 to $+1$, where a value approaching $+1$ indicates that a country is well-aligned within its group and clearly separated from other groups, a value near 0 suggests a country is not clearly separated from other groups, and a value near -1 indicates that the data trajectory of a country is closer to those of a different group.

Silhouette analyses for both the survey and model-based estimates were conducted using the complete family planning dataset for all participating countries. We calculated dissimilarities for all 185 countries, capturing the distance of each country to all others as well as its distance to countries within its sub-region. These

dissimilarities formed the basis for computing the silhouette width. Although the analysis draws on the complete family planning dataset available from the UNPD website, the partition plots display only FP2030 focus countries with more than one country in their sub-region. Sub-regions with only a single FP2030 country were excluded in the partition plot.

Figures 5.5 and 5.6 are the partition plots of survey data and model-based estimates silhouette widths respectively. The partition plot is produced by one of the visualisation functions of the `wdiexplorer` and displays diagnostic metrics for each country while accounting for predefined groupings, here sub-regions used by FPEM. Each country metric value is represented by a coloured bar over a lighter-shaded rectangular bar indicating the group average, allowing easy comparison of silhouette widths of countries within their group.

Silhouette widths highlight how closely countries follow the patterns of their assigned groups and reveal meaningful within-group deviations. However, when a country exhibits behaviour that diverges from that of other countries in the group, average dissimilarity for the group increases, which can influence the relative alignment of countries within the group. This pattern illustrates how outlier countries can pull the group average dissimilarity, affecting the alignment of the majority of countries within the group. A silhouette analysis based on median rather than average dissimilarity may be more appropriate for our context, but we have not investigated this.

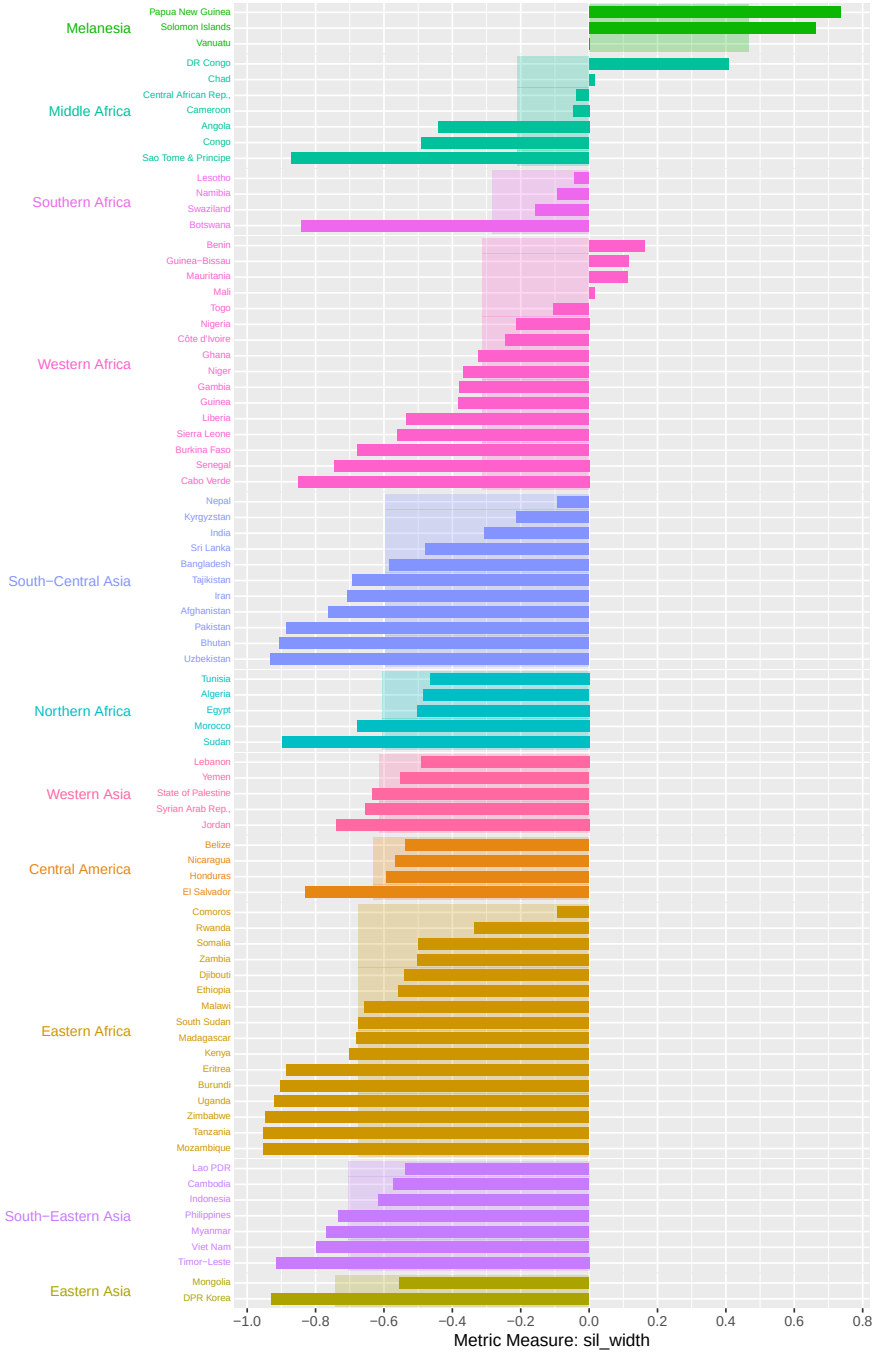

Figure 5.5: Silhouette widths of the survey data across the FP2030 focus countries, grouped by sub-region excluding sub-regions with only one country. Each bar represents the silhouette width of its corresponding country, while the light-shaded rectangular bars indicate the average silhouette width for each sub-region. Sub-regions are ordered by their average silhouette width, and countries within the same group are ordered accordingly with the same colour to facilitate visual distinction within groups. All sub-regions except Melanesia exhibit negative group average silhouette width.

Figure 5.5 displays the silhouette widths for the survey data of the 85 FP2030 focus countries, grouped by sub-region. Haiti, Ukraine, Kiribati, Bolivia, and Samoa are excluded from the partition plot because they are the only FP2030 country in their sub-region, leaving no other focus countries for comparison. Note, other countries do exist in their sub-regions in the complete family planning dataset used to compute the silhouette analysis, as reported in Table A.2 (see Appendix A). Sub-regions are arranged in descending order of their average silhouette widths, and within sub-regions, countries are ordered by decreasing silhouette widths.

Melanesia, a sub-region in the Oceania region, is the only sub-region with a positive average silhouette width. In this sub-region, Figure 5.2 shows that the three FP2030 focus countries consistently exhibit proportions of modern contraceptive use within a similar range across all survey years, although Vanuatu displays slightly higher levels than the other two countries. This overall similarity is reflected in their comparable and positive silhouette widths. Papua New Guinea with only three data points as reported in Table A.1 (see Appendix A) has a relatively high silhouette width of 0.74 and Solomon Islands, with two data points, has a silhouette width of 0.66. Vanuatu has a silhouette width of 0, because none of its surveys were conducted in the same years as those of other 5 countries in the Melanesia sub-region. (Recall that the complete dataset is used for the silhouette analysis). The group average silhouette width in this sub-region also reflects the observed low within group dissimilarity with positive average group silhouette width.

In Central America sub-region, the focus countries consistently exhibit proportions of modern contraceptive use within a similar pattern across survey years, see Figure 5.1. However, the silhouette widths shown in Figure 5.5 do not reflect this observation. This is explained by differences in survey years across countries. Survey observations were conducted in different years, and in this sub-region, only a few surveys were conducted in the same year. As a result, dissimilarities for some country pairs were computed using only one or, at most, two common survey years. Silhouette widths in this sub-region are estimated solely on the years where both countries have recorded observations, which may not adequately capture the observed overall similarity.

Eastern Asia has the lowest average group silhouette width, with two of our focus countries exhibiting high negative silhouette widths. In Middle Africa, the Democratic Republic of Congo and Chad show positive silhouette widths, whereas other countries in the sub-region have negative values, with São Tomé and Príncipe

exhibiting the lowest. In Southern Africa, Botswana has an extreme negative silhouette width, indicating that its trajectory of modern contraceptive use is distinct from all other countries in the sub region.

The influence of an outlier country is seen across most sub-regions. For example, in South-Eastern Asia, Timor-Leste exhibits the lowest proportion of modern contraceptive use, as shown in Figure 5.1. Its silhouette width, which quantifies how closely a country follows the patterns in its assigned group, indicates that Timor-Leste has the lowest value, reflecting its minimal alignment with the South-Eastern Asia sub-region (Figure 5.5). This extreme deviation contributes to an increased group average dissimilarity, reflecting the influence of an outlying country. A similar effect is observed in Sudan within the Northern Africa sub-region, where its contraceptive use is the least compared to the neighbouring countries, resulting in low silhouette width and higher group average dissimilarity.

In Western Africa, Cabo Verde has the highest proportion of modern contraceptive use across its survey years relative to other countries in the sub-region. Its low silhouette width indicates that it is the least aligned country within the Western Africa sub-region, thereby contributing to a negative group average silhouette width and thus higher overall group dissimilarity. By contrast, several countries within a similar range of proportions of modern contraceptive use in Western Africa sub-region, namely Benin, Guinea-Bissau, and Mauritania exhibit positive silhouette widths, demonstrating better alignment.

Overall, most countries exhibit strongly negative silhouette widths. In the literal interpretation of silhouette width, these values indicate that the data series trajectories of these countries are more similar to countries outside their assigned sub-regions than to those within. Several factors may contribute to these deviations. Firstly, we observed in Figures 5.1 and 5.2 that there is high variation in contraceptive use within each subregion. Furthermore, as discussed above, the presence of outlier countries can affect group alignment. In addition, family planning survey data were collected at irregular intervals, and because dissimilarities are computed only between observations from the same year, silhouette widths are estimated using dissimilarities of only those years where both countries in a pair have recorded observations. These inconsistencies could have amplified negative silhouette widths and weaken the alignment of countries within their sub-region.

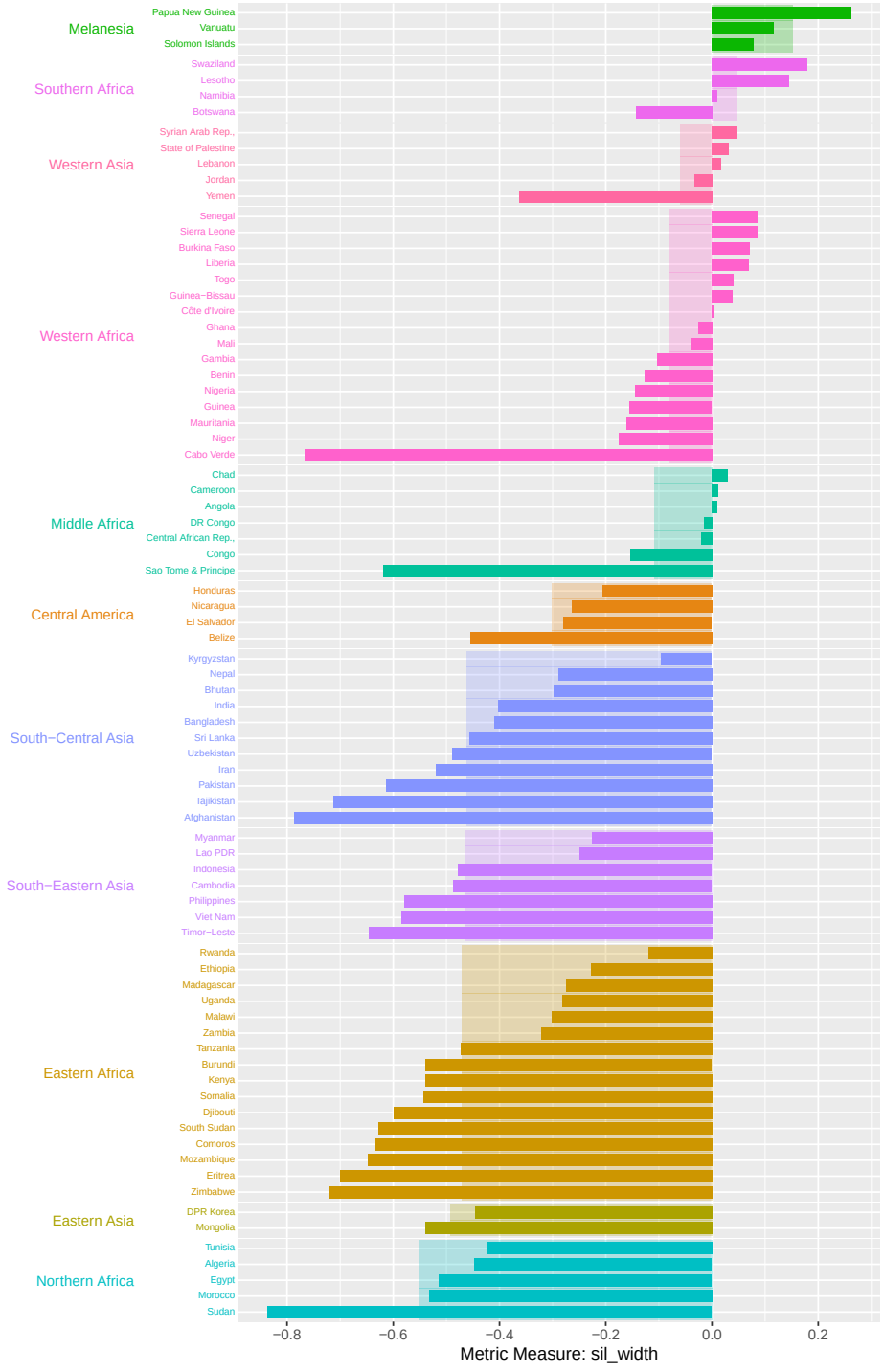

Figure 5.6: Silhouette widths of the model-based estimates across the 85 FP2030 focus countries, grouped by sub-region excluding sub-regions with only a single country. Sub-regions are ordered by their average silhouette width. Countries within the same sub-region share the same bar and axis text colour to facilitate visual distinction within groups. All sub-regions have negative average silhouette widths, except Melanesia and Southern Africa with positive average silhouette widths.

Figure 5.6 is the partition plot of the model-based estimates dataset for the 85 focus countries, excluding the 5 countries that are the only country in their sub-regions. Figure 5.6 shows that all countries in the Melanesia sub-region exhibit positive silhouette widths as seen in Figure 5.5. Notably, Vanuatu now has a silhouette value; in the survey partition plot of Figure 5.5, it lacked a value because the survey years of its observations did not match the survey years of the other countries in the sub-region. With FPEM generating annual estimates, this is no longer an issue.

Southern Africa shows a relatively small positive average silhouette width, with Botswana as the only country exhibiting a negative value, whereas in Figure 5.5, all countries exhibit negative silhouette widths. This reflects how the use of model-based annual estimates from FPEM, which enables consistent year-to-year comparisons across countries, can result in more robust silhouette widths that better represent alignment within this sub-region. In Western Africa, Cabo Verde stands out with an extremely low silhouette width. This pattern was also evident in Figure 5.5, as Cabo Verde is the only focus country in the Western Africa sub-region with a high uptake on modern contraception. Similarly, Yemen stands out among Western Asia countries with an extremely low silhouette width, though this was not apparent in Figure 5.5. Looking back at Figure 5.2, we see that Yemen has by far the lowest contraceptive use in its sub region.

Other sub-regions including Central America, South-Central Asia, South-Eastern Asia, Eastern Africa, Eastern Asia, and Northern Africa are dominated by countries with moderate to strongly negative silhouette widths. This same pattern was observed in the survey-based silhouette width as seen in Figure 5.5.

In Figure 5.5, countries in the Western Asia sub-region show negative silhouette widths, with an average group width of -0.61. The model-based estimates in Figure 5.6 present a slightly different picture: while Jordan and Yemen remain negative, other countries exhibit positive (though close to zero) silhouette widths, with an average group width of -0.06. This comparison highlights how the annual model-based estimates of FPEM adjust the alignment of countries within the sub-region relative to the survey-based data. Building on this, we compared the silhouette widths from the model-based estimates with those from the survey data to assess how well the model reproduces the patterns observed in the survey and to highlight cases with high differences in their values.

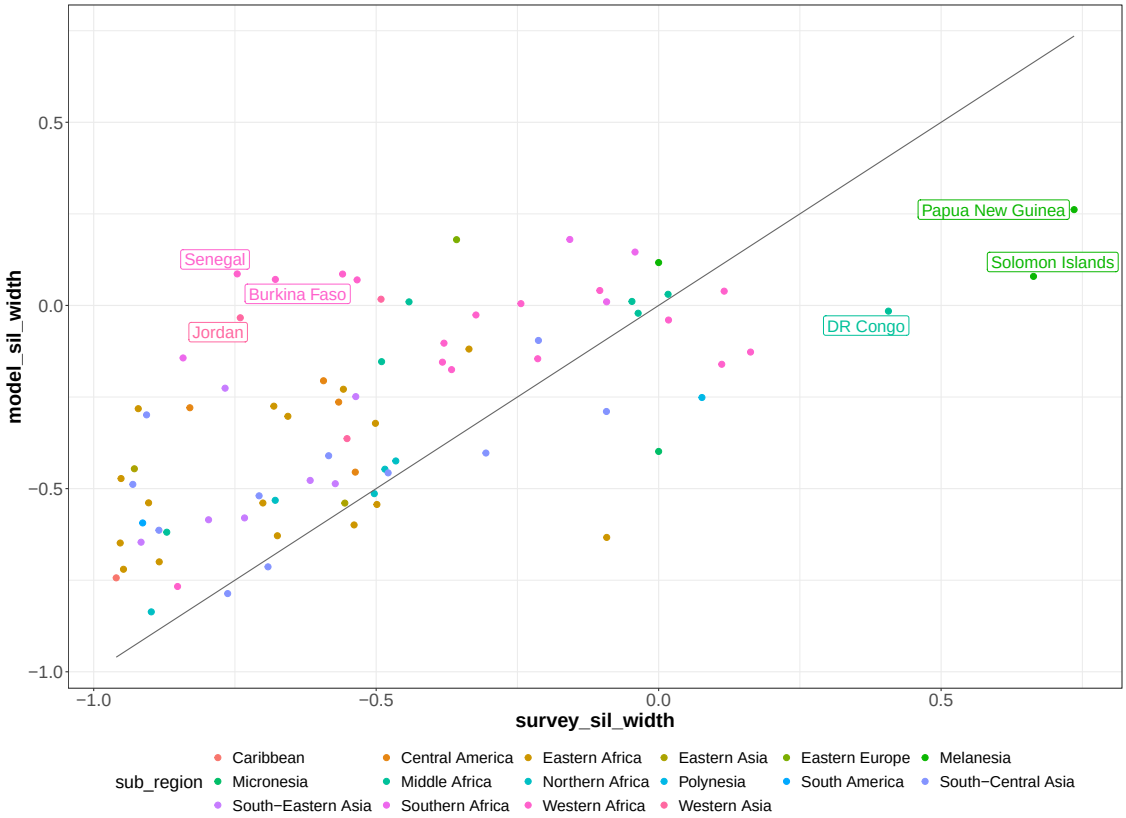

Figure 5.7: Comparison of silhouette widths between survey data and model-based estimates across countries. Each point represents the model-based silhouette width plotted against the survey silhouette width of a country. The diagonal line represents the identity line. Countries are coloured by sub-region to facilitate the identification of within group patterns. The labelled points correspond to the top three countries with the largest absolute differences between the two metrics and three other countries with the highest positive survey silhouette width. Countries represented by points close to the diagonal have similar silhouette widths in both the survey and model-based estimates. Hovering any point in the interactive version reveals the corresponding country name, its survey and model-based silhouette widths.

Figure 5.7 presents the scatterplot of model-based silhouette widths against the survey silhouette width for each country. The labelled countries are positioned far away from the diagonal line and their observations and model-based trends are shown in more detail in Figure 5.8.

Generally, we would expect model-based silhouette values to be bigger than those calculated from the survey data. This is due to the effect of smoothing, which flattens out local discrepancies and also the influence of the Bayesian hierarchical

framework which borrows strength in modelling trajectories within sub-regions. Consequently, most of the points in Figure 5.7 are above the diagonal line. The three labelled countries in Figure 5.7 (Burkina Faso and Senegal in Western Africa, and Jordan in Western Asia) with model-based silhouette widths considerably higher than values from the survey data would consequently appear to benefit a lot from the effects of modelling, even though their model-based silhouette values are still near zero. To further examine the large differences in their silhouette widths, we compare the survey data trajectories to the corresponding model-based trends of the labelled countries.

Figure 5.8: Survey data trajectories and model-based trends for the labelled countries with high absolute differences between model-based and survey silhouette widths in Figure 5.7. The labelled countries are presented along the other countries in their sub-region to compare their data series and model trends. Sub-regions are ordered by their silhouette width. Labelled countries are uniquely highlighted by distinct colours while the model trajectory of other countries in their sub-region are presented in grey colour. Countries in Western Africa have higher model-based silhouette width, while countries in Melanesia have higher survey silhouette width. Survey and model-based silhouette widths for each labelled country are presented in their panel. Hovering any line in the interactive version reveals the corresponding country name, and its survey and model silhouette widths.

Figure 5.8 also labels three countries with unusually high survey-based silhouette width, relative to those from the model trajectories. These are the three countries with the highest silhouette width in Figure 5.5, and their observations and trends are shown in Figure 5.8. As noted previously and as shown in Figure 5.7, the

two Melanesia countries (Papua New Guinea and Solomon Islands) have just a few survey observations and so the survey-based distances and thus silhouette widths are unreliable. The third country, DR Congo based on its survey silhouette value is the most middling country in Middle Africa, which is consistent with Figures 5.1 and 5.2. The drop in silhouette value for its model trajectory is likely explained by the fact that the smooth remains flat between 2000 and 2014, unlike other Middle Africa countries.

The comparison of survey and model-based trend strengths and silhouette widths highlights the influence of the Bayesian hierarchical framework and the time series modelling in FPEM, which smooths country trajectories. The low silhouette values for the observed data reflects the fact that most of the sub-regions have quite a lot of variation across their observed data trajectories. The silhouette values calculated from the model are somewhat higher. This analysis suggests that a grouping structure other than sub-region might be more informative for family planning modelling. To further strengthen our exploratory analysis, we analysed the residuals between the model estimates and survey observations.

5.5 Exploratory Residual Analysis of the Survey and Model-based Family Planning Datasets

To further compare how the model-based estimates of modern contraceptive use differs from the survey data, we analyse the residuals using two selected diagnostic indices of the `wdiexplorer` as discussed in Subsection 5.2.3. We compute the linearity and curvature of the residuals for each country. If the model captures the patterns in the data trajectories, the residuals should appear as random fluctuations around the zero line, with linearity and curvature values close to zero. When the residuals instead show some patterns, such as consistent upward or downward movements or fluctuations over time, it indicates that the fitted model has not fully captured important structure in the data.

Linearity is one of the diagnostic indices implemented in the `wdiexplorer`. It measures the overall direction of a data series over time. When applied to the residuals of the family planning datasets, this measure evaluates how closely the residuals follow a straight-line pattern. The linearity value is obtained as the coefficient of the linear term (β_1) in an orthogonal polynomial regression fitted to the

smoothed trend component of the decomposed series. Positive values indicate an increasing linear pattern in the residuals, while negative values indicate a decreasing pattern. Because residuals from a well-fitting model should resemble random noise, we expect minimal to no linear structure, with linearity values near zero.

The curvature feature implemented in the `wdiexplorer` measures how much a series bends away from a straight-line pattern over time, capturing non-linear movements such as gradual upward or downward curves (U- shaped and the inverted U-shaped). It is computed as the coefficient of the quadratic term of an orthogonal polynomial regression fitted to the trend component of the decomposed residuals (Hyndman and Athanasopoulos, 2021). As with linearity, residuals that behave like random noise should have curvature values close to zero.

To examine the linearity and curvature of the residuals in the family planning datasets, we present a scatterplot of these two metrics in Figure 5.9. Most countries cluster near zero for both metrics. The labelled points indicate countries where either linearity or curvature deviates much from the expected random behaviour.

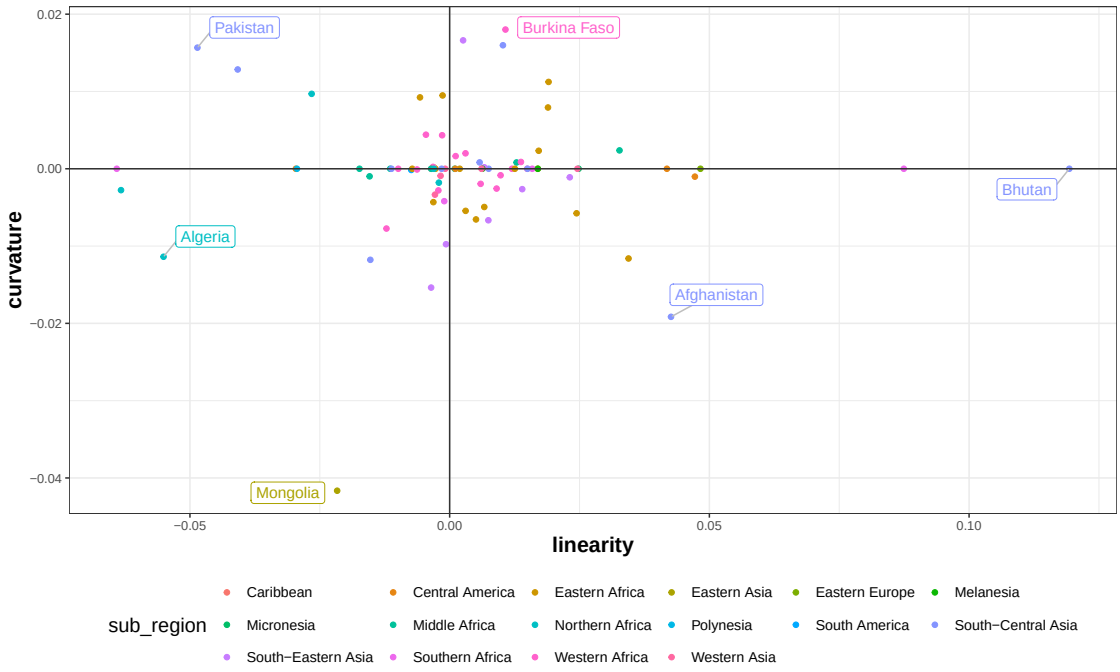

Figure 5.9: A scatterplot of residual curvature against linearity from the FP survey and model-based datasets, with each point representing a country and coloured by sub-region. It assesses whether linear or non-linear patterns remain in the residuals of each country. Most countries are around the vertical and horizontal lines indicating zero and near-zero curvature or linearity values. The labelled points correspond to countries with slight unexplained variation between the survey and the model-based estimates. Each panel presents the curvature and linearity values of its country. Hovering any point in the interactive version reveals the corresponding country name, the curvature and linearity of its residuals.

Figure 5.9 shows that the model performs well for most countries. The residuals cluster close to zero on both the linearity and curvature axes, indicating little evidence of systematic linear or non-linear structure in the unexplained variation. A small number of countries, which are labelled in the plot, fall somewhat farther from zero, although these deviations are not large, they show a higher absolute linearity or curvature relative to other countries. For example, Mongolia has a curvature value of -0.042 and a linearity value of -0.022, illustrating modest fluctuations in its residual pattern. In contrast, Bhutan has a curvature value of zero, while its linearity deviates noticeably from zero, suggesting a linear residual structure. Figure 5.10 further examines these residual patterns.

Figure 5.10 presents residual trajectories, a superimposed quadratic fit, the data

series and model trends for the labelled countries from Figure 5.9. Burkina Faso is also included: although the residuals analysis indicates good model fit, it stood out in earlier analysis of silhouette widths for the survey and model-based trends. This inclusion illustrates a situation where strong diagnostic from residuals coexist with atypical trend-strength measures.

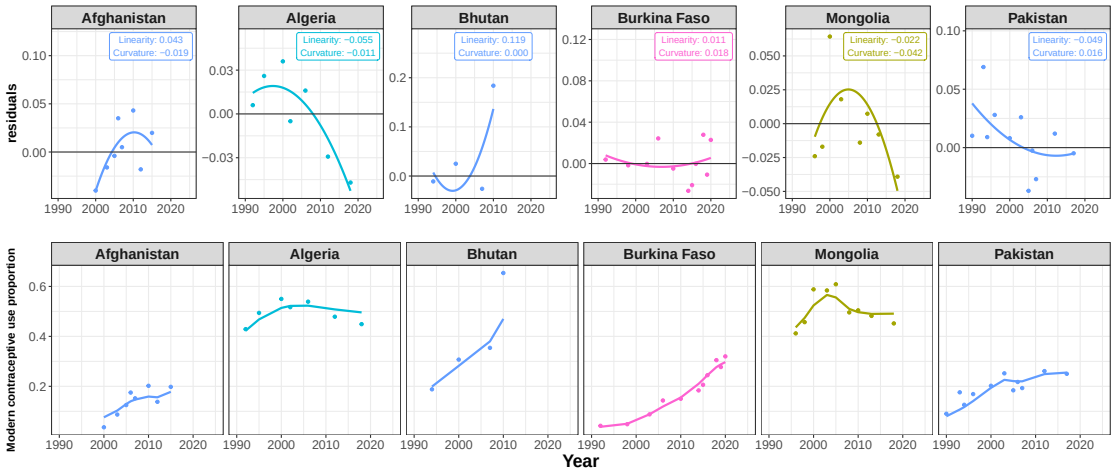

Figure 5.10: Labelled countries with non-zero curvature and linearity residual scores. The top panels show the residual trajectories and a superimposed quadratic fit of each country while the bottom panels display their survey data and corresponding model trends, illustrating the extent to which the model captures the patterns in each labelled country.

In Figure 5.10, residuals are shown in the top tow and the observed data with the model-based trends in the bottom row, with countries ordered consistently across rows. Bhutan has by far the largest linearity measure, which upon inspection is due to the substantial improvement in contraceptive use in 2010, which is not captured by the model. Upon investigation, earlier surveys in Bhutan, were based on national surveys, whereas the 2010 observation comes from a Multiple Indicator Cluster Survey. While the analysis allows for observations from different surveys and FPEM uses all surveys observations, the sudden spike was not captured by the model. In Afghanistan, the residuals exhibit a linearity of 0.043 and a curvature of -0.019. The observed proportion of modern contraceptive use in 2000 is substantially lower than in other survey years, resulting in a large negative residual for this year. Similarly, in 2010, where the observed value is relatively high, the model exhibits a slight bias, which may be attributed to its hierarchical pooling effect. Together,

these deviations account for the observed upward followed by downward pattern in the residuals, as captured by the linearity and curvature measures.

In Pakistan, the linearity and curvature values capture the discrepancy in the model fit over the period 2003–2012, as reflected in the residuals. Survey observations during this period show sudden increases and decreases in the proportion of modern contraceptive use, which do not follow a consistent temporal pattern. The time-series component of FPEM assumes smooth and continuous change rather than abrupt shifts; as a result, short-term fluctuations are smoothed in the model-based trends, giving rise to U-shaped residual trajectories. This behaviour is reflected in the residual curvature value of 0.016, while the observed dip between 2003 and 2005 is indicated by the negative linearity value of -0.049, corresponding to a downward bias in the model during these years.

Afghanistan, Bhutan, and Pakistan, all belonging to the same sub-region, exhibit notable discrepancies that warrant further investigation of the South-Central Asia sub-region. Across these countries and the other countries examined above, residual analysis highlights discrepancies between model-based trends and survey observations.

Somalia, a country not included in Figure 5.10 shows a downward-biased model prediction, with the fitted trend consistently lying below the observed survey points as shown in Figure 5.4. This downward bias is reflected in the residuals, which though positive are all small in magnitude, resulting in negligible linearity and curvature measures.

These findings establish the importance of exploring the observed data, the model-based trends, and the residuals to better understand the temporal patterns, diagnose where and investigate how well the model captures the observed patterns.

5.6 Conclusion

This study explored survey data and FPEM-based estimates of modern contraceptive prevalence using time series diagnostic indices implemented in the `wdiexplorer` package. Recognising that FPEM estimates are inherently smoothed, we assessed how well the model captures observed temporal patterns of modern contraceptive use across the 85 focus countries of the FP2030 initiative. By evaluating how well model trends reflect the underlying survey observations and the heterogeneity of family planning trajectories across countries and sub-regions, this study provides

insight into the temporal dynamics of the proportions of modern contraceptive use across the 85 FP2030 focus countries and highlight where discrepancies between model-based estimates and survey observations are most pronounced, with visualisation serving as a key tool. Visualisation is central to the analysis, enabling direct comparison of relatively sparse survey observations with annual modelled trajectories and facilitating the interpretation of diagnostic measures at both country and sub-regional levels.

Across the 85 focus countries, the selected diagnostic indices: silhouette width, trend strength, linearity and curvature indicate that FPEM generally captures long-term trends in modern contraceptive use data. Deviations from the survey data are most pronounced in countries and years exhibiting abrupt changes. For example, in Belize and Comoros, the model assumes gradual, continuous change over time rather than the sudden shifts observed in the data. In Somalia, while one would expect the model-based trend to partially align with the sub-regional patterns, it consistently falls below the observed survey points. Residual analysis highlights these discrepancies, providing a detailed view of where and how the model diverges from the observed data. Other countries with notable discrepancies, Afghanistan, Bhutan, and Pakistan, all belonging to the same sub-region, suggest that a grouping structure other than sub-region may improve the modelling of family planning trends.

This analysis is subject to limitations arising from the structure of the family planning survey data. These data are country-level survey observations obtained from multiple nationally representative household surveys, sometimes resulting in multiple observations per survey year and often temporal gaps of up to 6 years. The `wdiexplorer` package is developed to explore country-level panel data with one observation per country-year pair. Also, infrequent temporal gaps make several diagnostic measures of `wdiexplorer` unsuitable for capturing temporal behaviour in the proportion of modern contraceptive use. While silhouette width was considered appropriate, it poses a challenge. Dissimilarity-based measures are sensitive to temporal gaps, and could distort group-level alignment. One possible way to address this limitation is to estimate missing survey data points using linear interpolation prior to calculating dissimilarities. In addition, this analysis uses the mean dissimilarity of country pairs to quantify country-level dissimilarity; however, this measure can be sensitive to extreme distance values, an alternative approach would be the use of median instead of the mean in silhouette analysis, which may provide

a more robust interpretation of group-level patterns in the data.

In conclusion, this study demonstrates that combining time series diagnostic indices with visualisation offers a framework for evaluating how well model-based estimates align with observations. This analysis confirms that FPEM-based estimates generally capture long-term trends in modern contraceptive use data across the 85 FP2030 focus countries, while also highlighting specific countries and periods where the model diverges from observed data, particularly in cases of abrupt changes or where sub-regional grouping may not fully capture local dynamics. By identifying these discrepancies, this approach not only provides insight into the temporal patterns of family planning trajectories but also points to opportunities for refining hierarchical modelling strategies. Despite limitations arising from sparse and unevenly spaced survey data, the integration of diagnostic measures and visualisation enables a nuanced assessment of model performance.

6

Conclusion

This thesis has introduced significant contributions to the exploration and visualisation of country-level panel data and hierarchical model outputs. Through Chapters 3, 4, and 5, we have highlighted the importance of incorporating predefined grouping structures into data and model exploration and visualisation. Our main contributions are diagnostic measures for multiple time series, and carefully designed strategies for constructing rich and informative visualisations.

Chapter 3 introduced `wdiexplorer`, a new R package that supports a workflow for exploring country-level panel data based on time series diagnostic measures and visualisation tools. The primary aim of `wdiexplorer` is to reveal temporal patterns, trends, outliers and other important structural features of panel data. The package extends scagnostics concepts by incorporating existing feature-based time series indices. These diagnostic indices capture variation features, trend dynamics, shape and sequential temporal characteristics of the data series. `wdiexplorer` also uses silhouette analysis for assessing patterns between and within groups, by incorporating grouping structures into diagnostic calculations. The indices are used to draw attention to potentially interesting countries in the visualisations of country-level panel data, and to identify countries for further investigation. Although the name `wdiexplorer` originates from demonstrating the package functionality using the WDI data, the workflow is not restricted to WDI data. The package is applicable to any country-level panel dataset of repeated measurements of the same variables collected over time across multiple countries.

While this development of `wdiexplorer` represents an advancement in time series diagnostic measures by exploring data series within the hierarchical structures present in the data, future research could extend the package functionality

to include an interactive Shiny application, which would allow users with little to no knowledge of the R programming language to explore country-level panel data. Also, the set of diagnostic indices could be extended to capture additional features and more complex temporal behaviours, such as those offered by Henderson and Fulcher (2025). Another possible extension to the diagnostic measures could be the use of classical and robust principal component analysis for anomaly detection in hierarchical time series, building on recent developments in functional data analysis by Hyndman and Shang (2010); Fisch et al. (2024). The use of principal component analysis for dimensionality reduction could be integrated with the diagnostic indices of `wdiexplorer`, which incorporate the hierarchical structure of the data into their computation, to preserve hierarchical heterogeneity while detecting outlying countries.

Chapter 4 presented new visualisations and proposed informative principles for displaying hierarchical model estimates. Our visualisations provide an in-depth examination of model fit in relation to the hierarchical structure of the data. It also support the examination of multiple models simultaneously to understand how individual-level units (countries) respond to hierarchical groupings and how grouping choices influence posterior estimates. This is inspired by the strategies outlined by Wickham et al. (2015) for visualising statistical models. For the visualisation principles, we utilised layout, scales, colours, and line types, to differentiate and emphasise patterns and to facilitate comparisons across model outputs.

The proposed visualisations were illustrated using the outputs of five Bayesian hierarchical models fitted to PISA European mathematics scores, demonstrating their relevance for interpreting complex hierarchical model structures. The visualisation designs are simple, flexible, and adaptable, and can be extended to non-Bayesian models, models capturing non-linear relationships, and models with more complex hierarchical structures. By applying our step-by-step principles, researchers can gain a clearer understanding of hierarchical structures present in the data, how it influence model output, and communicate results more effectively. This contribution advances the visualisation of hierarchical models and offers adaptable tools that can be applied across a wide range of research fields.

Future work could explore the design of visualisation for joint distributions of model parameters in the context of hierarchical models. Additionally, it could be further extended to compare posterior parameter and hyper-parameter estimates across different Bayesian computational methods. For example, hierarchical mod-

els could be fitted using software such as JAGS and BRMS with the same model specifications and likelihoods to compare their estimates and also assess how individual units respond to information sharing under different computational implementations. Beyond its current use, these visualisation could then be extended to compare model outputs across different computational methods. A notable limitation of the work is that it would be difficult to extend the visualisations to accommodate comparisons across more than two hierarchical structures simultaneously. Currently, the strip background colour of each panel represents the regional grouping, while the strip text indicates the income level of each country.

Building on the work in Chapters 3 and 4, Chapter 5 presented a new case study that utilised some of the diagnostic indices of `wdiexplorer`, alongside the visualisation principles proposed in Chapter 4 to explore survey data and Family Planning Model-based estimates of modern contraceptive prevalence. The rationale was to evaluate how well model trends reflect the underlying survey observations and the heterogeneity of family planning trajectories across countries and sub-regions and to highlight where discrepancies between model-based estimates and survey observations are most pronounced.

Future work could extend the application of `wdiexplorer` to develop diagnostic methods that can handle missing or irregularly spaced observations, as was the case with the family planning survey data. For instance, techniques such as linear interpolation or other imputation approaches could be used to fill gaps in the data, enabling the indices to provide more reliable evaluations of model fit and reveal patterns that might otherwise be hidden. Another complementary approach would be to design alternative dissimilarity measures for comparing country-level time series when observations are sparse or unevenly spaced. Currently, dissimilarities can only be computed at time points where both countries have data, which limits the ability to assess similarity across the full temporal range. Developing methods that can accommodate infrequent or incomplete time series would improve the effectiveness and flexibility of country-level panel data diagnostic analyses.

Also, family planning surveys are collected from multiple survey types such as Demographic Health Survey, UNICEF Multiple Indicator Cluster Surveys, Performance Monitoring for Action, Reproductive Health Surveys, national surveys, and other sources, which sometimes yield multiple data points per year. However, `wdiexplorer` is designed to accommodate a single observation per country-year. In Chapter 5, we handled this by ranking the family planning survey types by global

reliability and comparability, and only the observation for the highest-ranked survey type was selected for each country-year. This was a limitation, because model-based estimates generated using all available survey data were compared with survey observations from the highest-ranked survey type. A possible direction for future work would be to extend the diagnostic measures of `wdiexplorer` to handle multiple observations per country-year.

In assessing how well model trends capture underlying survey observations and heterogeneity in the data, alternative hierarchical structures could be considered. In the FPEM, projections for family planning indicators among the married or in-union women are modelled by nesting countries within sub-regions, sub-regions within regions, and regions within continents. For projections of family planning indicators among unmarried populations, which we didn't focus on for this project, models are fitted using combinations of sexual activity and geographical region as grouping variables. Future research could explore additional grouping structures to assess model alignment with survey observations.

Reproducibility has been a key consideration throughout this thesis. All data processing, model fitting, and visualisation workflows have been fully documented, with codes and analyses well structured to facilitate replication. The `wdiexplorer` package provides a systematic step-by-step workflow, with each stage of the workflow carefully described for ease of understanding and adaptation to new datasets. It contains two vignettes, one using an annual dataset and another a triennial dataset, to demonstrate its functionality across different temporal intervals datasets. Each project organised as chapters of this thesis has a separate GitHub repository containing a detailed README file that describe the content and purpose of each script. Scripts are clearly labelled in accordance with the sections of their respective chapters. Additionally, all scripts, data, visualisation codes, and outputs associated with each analysis presented in this thesis have been organised for submission alongside each manuscript, enabling verification of results. The emphasis on reproducibility strengthens adaptation and extension of these methods in future research.

In conclusion, this thesis has made valuable contributions to the exploration and visualisation of country-level panel data and hierarchical model outputs. Through the development of the `wdiexplorer` R package, country-level panel data can be explored using time series diagnostic indices within the context of the predefined hierarchical structures present in the data, enabling examination of patterns, trends,

shape features, outliers, and other temporal characteristics. The five proposed visualisation principles extend beyond standard numerical summaries, providing a clear and structured way to present hierarchical model outputs, which represents a significant step forward in advancing the use of visualisation in the Bayesian workflow. These contributions, `wdiexplorer` R package and visualisation principles developed in this thesis not only facilitate detailed exploration and visualisation of country-level panel data and hierarchical model outputs but also aim to equip others with flexible and reproducible workflows for understanding and interpreting country-level panel data.

A

Appendix A: Using Time Series Measures to Explore Family Planning Survey Data and Model-based Estimates

A.1 Preliminary Data Exploration

Table A.1: Summary of surveys for family planning focus countries

Country	Sub Region	Number of Surveys	Recent Survey
Afghanistan	South-Central Asia	8	2018
Algeria	Northern Africa	7	2018
Angola	Middle Africa	4	2015
Bangladesh	South-Central Asia	13	2019
Belize	Central America	5	2015
Benin	Western Africa	6	2017
Bhutan	South-Central Asia	4	2010
Bolivia	South America	6	2016
Botswana	Southern Africa	2	2017
Burkina Faso	Western Africa	11	2020
Burundi	Eastern Africa	7	2016
Cabo Verde	Western Africa	3	2018

Continued on next page

Table continued from previous page

Country	Sub Region	Number of Surveys	Recent Survey
Cambodia	South-Eastern Asia	6	2014
Cameroon	Middle Africa	7	2018
Central African Republic	Middle Africa	5	2018
Chad	Middle Africa	6	2019
Comoros	Eastern Africa	3	2012
Congo	Middle Africa	3	2014
Côte d'Ivoire	Western Africa	8	2020
DPR Korea	Eastern Asia	8	2017
DR Congo	Middle Africa	6	2017
Djibouti	Eastern Africa	3	2012
Egypt	Northern Africa	10	2014
El Salvador	Central America	5	2014
Eritrea	Eastern Africa	3	2010
Ethiopia	Eastern Africa	12	2020
Gambia	Western Africa	7	2019
Ghana	Western Africa	13	2017
Guinea	Western Africa	6	2018
Guinea-Bissau	Western Africa	5	2018
Haiti	Caribbean	5	2016
Honduras	Central America	6	2019
India	South-Central Asia	8	2019
Indonesia	South-Eastern Asia	25	2018
Iran	South-Central Asia	9	2010
Jordan	Western Asia	7	2017
Kenya	Eastern Africa	11	2020
Kiribati	Micronesia	2	2018
Kyrgyzstan	South-Central Asia	5	2018
Lao PDR	South-Eastern Asia	5	2017
Lebanon	Western Asia	4	2009
Lesotho	Southern Africa	9	2018
Liberia	Western Africa	4	2019

Continued on next page

Table continued from previous page

Country	Sub Region	Number of Surveys	Recent Survey
Madagascar	Eastern Africa	9	2020
Malawi	Eastern Africa	9	2019
Mali	Western Africa	7	2018
Mauritania	Western Africa	6	2019
Mongolia	Eastern Asia	9	2018
Morocco	Northern Africa	6	2018
Mozambique	Eastern Africa	6	2015
Myanmar	South-Eastern Asia	6	2015
Namibia	Southern Africa	4	2013
Nepal	South-Central Asia	10	2019
Nicaragua	Central America	5	2011
Niger	Western Africa	10	2021
Nigeria	Western Africa	12	2018
Pakistan	South-Central Asia	11	2018
Papua New Guinea	Melanesia	3	2016
Philippines	South-Eastern Asia	17	2017
Rwanda	Eastern Africa	8	2019
Samoa	Polynesia	4	2019
Sao Tome and Principe	Middle Africa	5	2019
Senegal	Western Africa	12	2019
Sierra Leone	Western Africa	8	2019
Solomon Islands	Melanesia	2	2015
Somalia	Eastern Africa	4	2018
South Sudan	Eastern Africa	4	2015
Sri Lanka	South-Central Asia	4	2016
State of Palestine	Western Asia	6	2019
Sudan	Northern Africa	5	2014
Swaziland	Southern Africa	5	2014
Syrian Arab Republic	Western Asia	4	2009
Tajikistan	South-Central Asia	5	2017
Tanzania	Eastern Africa	8	2015

Continued on next page

Table continued from previous page

Country	Sub Region	Number of Surveys	Recent Survey
Timor-Leste	South-Eastern Asia	8	2016
Togo	Western Africa	6	2017
Tunisia	Northern Africa	6	2018
Uganda	Eastern Africa	12	2021
Ukraine	Eastern Europe	4	2012
Uzbekistan	South-Central Asia	4	2006
Vanuatu	Melanesia	3	2013
Vietnam	South-Eastern Asia	21	2020
Yemen	Western Asia	5	2013
Zambia	Eastern Africa	6	2018
Zimbabwe	Eastern Africa	7	2015

Table A.2: A tabular representation of the total number of countries in each sub-region, grouped by FP2030 focus country status.

Sub_region	Number of FP2030 countries	Others
Eastern Africa	16	2
Western Africa	16	0
South-Central Asia	11	3
Middle Africa	7	2
South-Eastern Asia	7	3
Northern Africa	5	6
Western Asia	5	11
Central America	4	4
Southern Africa	4	1
Melanesia	3	3
Eastern Asia	2	5
Caribbean	1	11
Eastern Europe	1	9
Micronesia	1	4
Polynesia	1	3
South America	1	11

Bibliography

Ali Ahmadi, A., Balakrishnan, P., Kakosimos, K., and Goktepe, I. (2018), "Determination of the Levels of Particulate Matter 25 and 10 and their Elemental Composition in Qatar," in *Qatar Foundation Annual Research Conference Proceedings*, HBKU Press Qatar, vol. 2018, p. EEPD912.

Alkema, L., Kantorova, V., Menozzi, C., and Biddlecom, A. (2013), "National, Regional, and Global Rates and Trends in Contraceptive Prevalence and Unmet Need for Family Planning Between 1990 and 2015: A Systematic and Comprehensive Analysis," *The Lancet*, 381, 1642–1652.

Arel-Bundock, V. (2025a), *Model to meaning: How to interpret statistical models with r and python*, CRC Press.

— (2025b), *WDI: World Development Indicators and Other World Bank Data*, r package version 2.7.9.

Arel-Bundock, V., Enevoldsen, N., and Yetman, C. (2018), "countrycode: An R package to Convert Country Names and Country Codes," *Journal of Open Source Software*, 3, 848.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015), "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, 67, 1–48.

Bauer, S. E., Im, U., Mezuman, K., and Gao, C. Y. (2019), "Desert Dust, Industrialization, and Agricultural Fires: Health Impacts of Outdoor Air Pollution in Africa," *Journal of Geophysical Research: Atmospheres*, 124, 4104–4120.

Bracke, M. A. (2022), "Women's Rights, Family Planning, and Population Control: The Emergence of Reproductive Rights in the United Nations (1960s–70s)," *The International History Review*, 44, 751–771.

Buja, A., Cook, D., and Swayne, D. F. (1996), "Interactive High-Dimensional Data Visualization," *Journal of Computational and Graphical Statistics*, 5, 78–99.

Bürkner, P.-C. (2018), "Advanced Bayesian Multilevel Modeling with the R Package brms," *The R Journal*, 10, 395–411.

Cahill, N., Sonneveldt, E., Emmart, P., Williamson, J., Mbu, R., Fodjo Yetgang, A. B., Dambula, I., Azambuja, G., Mahumane Govo, A. A., Joshi, B., et al. (2021), "Using Family Planning Service Statistics to Inform Model-Based Estimates of Modern Contraceptive Prevalence," *Plos one*, 16, e0258304.

Cahill, N., Sonneveldt, E., Stover, J., Weinberger, M., Williamson, J., Wei, C., Brown, W., and Alkema, L. (2018), "Modern Contraceptive Use, Unmet Need, and Demand Satisfied Among Women of Reproductive Age who are Married or in a Union in the Focus Countries of the Family Planning 2020 Initiative: A Systematic Analysis Using the Family Planning Estimation Tool," *The Lancet*, 391, 870–882.

Cahill, N., Weinberger, M., and Alkema, L. (2020), "What Increase in Modern Contraceptive Use is Needed in FP2020 Countries to Reach 75% Demand Satisfied by 2030? An Assessment Using the Accelerated Transition Method and Family Planning Estimation Model," *Gates Open Research*, 4, 113.

Carlo, C. M. (2004), "Markov Chain Monte Carlo and Gibbs Sampling," *Lecture notes for EEB*, 581, 3.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017), "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 76.

Castellacci, F. and Natera, J. M. (2011), "A New Panel Dataset for Cross-Country Analyses of National Systems, Growth and Development (CANA)," *Innovation and Development*, 1, 205–226.

Cheng, J. and Sievert, C. (2023), *crosstalk: Inter-Widget Interactivity for HTML Widgets*, r package version 1.2.1.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., Terpenning, I., et al. (1990), "STL: A seasonal-trend decomposition," *J. off. Stat*, 6, 3–73.

Cleveland, W. C. and McGill, M. E. (1988), *Dynamic Graphics for Statistics*, CRC Press, Inc.

Cleveland, W. S. (1985), *The Elements of Graphing Data*, Wadsworth Publ. Co.

Correa-Álvarez, C. D., Salazar-Uribe, J. C., and Pericchi-Guerra, L. R. (2023), "Bayesian Multilevel Logistic Regression Models: A Case Study Applied to the Results of Two Questionnaires Administered to University Students," *Computational Statistics*, 38, 1791–1810.

Corrigan, A. E., Becker, M. M., Neas, L. M., Cascio, W. E., and Rappold, A. G. (2018), "Fine Particulate Matters: The Impact of Air Quality Standards on Cardiovascular Mortality," *Environmental Research*, 161, 364–369.

Cunningham, M., Bock, A., Brown, N., Sacher, S., Hatch, B., Inglis, A., and Aronovich, D. (2015), "Estimating Contraceptive Prevalence Using Logistics Data for Short-Acting Methods: Analysis Across 30 Countries," *Global Health: Science and Practice*, 3, 462–481.

Dang, T. N., Anand, A., and Wilkinson, L. (2012), "Timeseer: Scagnostics for High-Dimensional Time Series," *IEEE Transactions on Visualization and Computer Graphics*, 19, 470–483.

Dang, T. N. and Wilkinson, L. (2014), "Scagexplorer: Exploring Scatterplots by their Scagnostics," in *2014 IEEE Pacific Visualization Symposium*, IEEE, pp. 73–80.

Dolnicar, S., Grün, B., and Leisch, F. (2018), *Market Segmentation Analysis—Understanding It, Doing It, and Making It Useful*, Singapore: Springer.

Fisch, A., Grose, D., Eckley, I. A., Fearnhead, P., and Bardwell, L. (2024), "anomaly: Detection of Anomalous Structure in Time Series Data," *Journal of Statistical Software*, 110, 1–24.

FP2030 (2025), "ABOUT FP2030: Working together to advance family planning around the world." Accessed: 2025-10-24.

Friedman, V. (2008), "Data Visualization and Infographics," *Graphics, Monday Inspiration*, 14, 2008.

Friendly, M. and Wainer, H. (2021), "A History of Data Visualization and Graphic Communication," in *A History of Data Visualization and Graphic Communication*, Harvard University Press.

Fulcher, B. D., Little, M. A., and Jones, N. S. (2013), "Highly Comparative Time-Series Analysis: The Empirical Structure of Time Series and their Methods," *Journal of the Royal Society Interface*, 10.

Gabry, J. and Mahr, T. (2025), "bayesplot: Plotting for Bayesian Models," R package version 1.13.0.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019), "Visualization in Bayesian Workflow," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182, 389–402.

Gamerman, D. and Lopes, H. F. (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall/CRC.

Gehlenborg, N. and Wong, B. (2012), "Heat maps," *Nature Methods*, 9, 213.

Gelman, A. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, Chapman and Hall/CRC.

Gelman, A., Lee, D., and Guo, J. (2015), "Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization," *Journal of Educational and Behavioral Statistics*, 40, 530–543.

Gelman, A., Meng, X.-L., and Stern, H. (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies," *Statistica Sinica*, 733–760.

Gelman, A. and Pardoe, I. (2006), "Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models," *Technometrics*, 48, 241–251.

Gelman, A. et al. (2013), "Bayesian Data Analysis, 3rd: Boca Raton," *Texts in Statistical Science*.

Gohel, D. and Skintzos, P. (2025), *ggiraph: Make 'ggplot2' Graphics Interactive*, r package version 0.8.12.

Goldstein, H. (2004), "International Comparisons of Student Attainment: Some Issues Arising from the PISA Study," *Assessment in Education: Principles, Policy & Practice*, 11, 319–330.

Guerreiro, C. B., Foltescu, V., and De Leeuw, F. (2014), "Air Quality Status and Trends in Europe," *Atmospheric Environment*, 98, 376–384.

Guranich, G., Cahill, N., and Alkema, L. (2021), "Fpemlocal: Estimating Family Planning Indicators in R for a Single Population of Interest," *Gates Open Research*, 5, 24.

Hafen, R. (2024), *geofacet: 'ggplot2' Faceting Utilities for Geographical Data*, https://github.com/hafen/geofacet, https://hafen.github.io/geofacet/.

Hardee, K. and Jordan, S. (2021), "Advancing Rights-Based Family Planning from 2020 to 2030," *Open Access Journal of Contraception*, 157–171.

Harishkumar, K., Yogesh, K., et al. (2020), "Forecasting Air Pollution Particulate Matter (PM2. 5) using Machine Learning Regression Models," *Procedia Computer Science*, 171, 2057–2066.

Harlen, W. (2001), *The Assessment of Scientific Literacy in the OECD/PISA Project*, Dordrecht: Springer Netherlands, pp. 49–60.

Heiss, A. (2021), "A Guide to Correctly Calculating Posterior Predictions and Average Marginal Effects with Multilevel Bayesian Models," .

Heiss, A. (2022), "Marginal and Conditional Effects for GLMMs with Marginaleffects," Accessed: 2026-05-14.

Henderson, T. and Fulcher, B. D. (2025), "Feature-Based Time-Series Analysis in R using the Theft Ecosystem," *The R Journal*, 17, 43–68, https://doi.org/10.32614/RJ-2025-023.

Horan, K., Brunsdon, C., and Domijan, K. (2024), "A multilevel spatial model to investigate voting behaviour in the 2019 UK General Election," *Applied Spatial Analysis and Policy*, 17, 703.

Hsiao, C. (1985), "Benefits and Limitations of Panel Data," *Econometric Reviews*, 4, 121–174.

— (2005), "Why Panel Data?" *The Singapore Economic Review*, 50, 143–154.

— (2022), *Analysis of Panel Data*, no. 64, Cambridge University Press.

Hubacher, D. and Trussell, J. (2015), "A Definition of Modern Contraceptive Methods," *Contraception*, 92, 420–421.

Huntington-Klein, N. and Khor, P. (2026), *pmdplyr: 'dplyr' Extension for Common Panel Data Maneuvers*, r package version 0.3.4, commit a2c309ab8956ef4d2fc7860cbb9b933184219397.

Hurley, C. B., O'Connell, M., and Domijan, K. (2022), "Interactive Slice Visualization for Exploring Machine Learning Models," *Journal of Computational and Graphical Statistics*, 31, 1–13.

Hyndman, R. J. and Athanasopoulos, G. (2021), *Forecasting: Principles and Practice*, Melbourne, Australia: OTexts, 3rd ed. , accessed on 2025-07-05.

Hyndman, R. J. and Shang, H. L. (2010), "Rainbow Plots, Bagplots, and Boxplots for Functional Data," *Journal of Computational and Graphical Statistics*, 19, 29–45.

Islam, M. and Jin, S. (2019), "An Overview of Data Visualization," in *2019 International Conference on Information Science and Communications Technologies (ICISCT)*, IEEE, pp. 1–7.

Jreich, R., Hatte, C., and Parent, E. (2022), "Review of Bayesian Selection Methods for Categorical Predictors using JAGS," *Journal of Applied Statistics*, 49, 2370–2388.

Kay, M. (2023), *tidybayes: Tidy Data and Geoms for Bayesian Models*, r package version 3.0.6.

— (2024), *ggdist: Visualizations of Distributions and Uncertainty*, r package version 3.3.2.

Kirk, A. (2019), "Data Visualisation: A Handbook for Data Driven Design," .

Kuang, B. and Brodsky, I. (2016), "Global Trends in Family Planning Programs, 1999–2014," *International Perspectives on Sexual and Reproductive Health*, 42, 33–44.

Larmarange, J. (2026), *ggstats: Extension to 'ggplot2' for Plotting Stats*, r package version 0.13.0.

Liu, L., Moon, H. R., and Schorfheide, F. (2021), "Panel Forecasts of Country-Level Covid-19 Infections," *Journal of Econometrics*, 220, 2–22.

Long, J. A. (2020), *panelr: Regression Models and Utilities for Repeated Measures and Panel Data*, r package version 0.7.3.

Luedicke, J. (2015), "Friedman's Super Smoother," *Boston College. http://fmwww. bc. edu/RePEc/bocode/s/supsmooth_doc. pdf*.

Manovich, L. (2011), "What is Visualisation?" *Visual Studies*, 26, 36–49.

Marquez, M. P., Kabamalan, M. M., and Laguna, E. (2018), "Traditional and Modern Contraceptive Method Use in the Philippines: Trends and Determinants 2003–2013," *Studies in Family Planning*, 49, 95–113.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013), "Bayesian Computing with INLA: New Features," *Computational Statistics & Data Analysis*, 67, 68–83.

Marttila, M. (2023), *ggragged: Ragged Grids for 'ggplot2'*, r package version 0.1.0.

McGlothlin, A. E. and Viele, K. (2018), "Bayesian Hierarchical Models," *Jama*, 320, 2365–2366.

Munzner, T. (2014), *Visualization Analysis and Design*, AK Peters Visualization Series, CRC Press.

Mücke, M. (2025), *worldbank: Client for World Banks's 'Indicators' and 'Poverty and Inequality Platform (PIP)' APIs*, r package version 0.6.0.

New, J. R., Cahill, N., Stover, J., Gupta, Y. P., and Alkema, L. (2017), "Levels and Trends in Contraceptive Prevalence, Unmet Need, and Demand for Family Planning for 29 States and Union Territories in India: A Modelling Study using the Family Planning Estimation Tool," *The Lancet Global Health*, 5, e350–e358.

OECD (2023a), *PISA 2022 Assessment and Analytical Framework*.

— (2023b), *PISA 2022 Assessment and Analytical Framework*, OECD Publishing.

— (2024), "Programme for International Student Assessment (PISA)," https://www.oecd.org/pisa/, accessed: January 29, 2024.

O'Hara-Wild, M., Hyndman, R., and Wang, E. (2024a), *fabletools: Core Tools for Packages in the 'fable' Framework*, r package version 0.5.0.

— (2024b), *feasts: Feature Extraction and Statistics for Time Series*, r package version 0.4.1.

Organisation for Economic Co-operation and Development (OECD) (2026), "OECD Data," Online data repository, accessed: 12 February 2026.

Our World in Data (2026), "Our World in Data: Research and data to make progress against the world's largest problems," https://ourworldindata.org/, accessed: 12 February 2026.

Persson, E. (2021), *OECD: Search and Extract Data from the OECD*, r package version 0.2.5.

Piburn, J. (2020), *wbstats: Programmatic Access to the World Bank API*, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Plummer, M. (2025), *rjags: Bayesian Graphical Models using MCMC*, r package version 4-17.

Plummer, M. et al. (2003), "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria, vol. 124, pp. 1–10.

Qu, Z., Dai, W., Euan, C., Sun, Y., and Genton, M. G. (2024), "Exploratory Functional Data Analysis," *Test*, 1–24.

— (2025), "Exploratory Functional Data Analysis: Z. Qu et al." *Test*, 34, 459–482.

R Core Team (2024), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

RamaRao, S. and Jain, A. K. (2015), "Aligning Goals, Intents, and Performance Indicators in Family Planning Service Delivery," *Studies in Family Planning*, 46, 97–104.

Raudenbush, S. W. and Bryk, A. S. (2002), *HierarchicalLlinear Models: Applications and Data Analysis Methods*, vol. 1, sage.

Rouder, J. N. and Lu, J. (2005), "An Introduction to Bayesian Hierarchical Models with an Application in the Theory of Signal Detection," *Psychonomic Bulletin & Review*, 12, 573–604.

Rouder, J. N., Morey, R. D., and Pratte, M. S. (2014), "Bayesian Hierarchical Models," *New Handbook of Mathematical Psychology*, 1.

Rousseeuw, P. J. (1987), "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, 20, 53–65.

Rue, H. and Martino, S. (2007), "Approximate Bayesian Inference for Hierarchical Gaussian Markov Random Field Models," *Journal of Statistical Planning and Inference*, 137, 3177–3192.

Sawant, P., Billor, N., and Shin, H. (2012), "Functional Outlier Detection with Robust Functional Principal Component Analysis," *Computational Statistics*, 27, 83–102.

Scheuch, C. (2025a), *owidapi: Access the Our World in Data Chart API*, r package version 0.1.1.

— (2025b), *wbwdi: Seamless Access to World Bank World Development Indicators (WDI)*, r package version 1.0.3.

Schleicher, A. (2019), "PISA 2018: Insights and Interpretations." *OECD Publishing*.

Sicard, P., Agathokleous, E., Anenberg, S. C., De Marco, A., Paoletti, E., and Calatayud, V. (2023), "Trends in Urban Air Pollution Over the Last Two Decades: A Global Perspective," *Science of The Total Environment*, 858, 160064.

Sievert, C. (2020), *Interactive Web-Based Data Visualization with R, plotly, and shiny*, Chapman and Hall/CRC.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996), "BUGS 0.5: Bayesian Inference using Gibbs Sampling Manual (Version ii)," *MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*, 1–59.

Su, Y.-S. and Yajima, M. (2024), *R2jags: Using R to Run 'JAGS'*, r package version 0.8-9.

The World Bank (2025a), "How does the World Bank classify countries?" Accessed: 2025-08-06.

— (2025b), "PM2.5 Air Pollution, Mean Annual Exposure (micrograms per cubic meter)," https://data.worldbank.org/indicator/EN.ATM.PM25.MC.M3, world Development Indicators. Accessed: 2025-01-31.

— (2025c), "World Bank Country and Lending Groups," Accessed: 2025-08-08.

— (2025d), "World Development Indicators (WDI)," https://databank.worldbank.org/source/world-development-indicators, accessed: 2025-08-06.

Tierney, N. and Cook, D. (2023), "Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations," *Journal of Statistical Software*, 105, 1–31.

Tierney, N., Cook, D., and Prvan, T. (2022), "brolgar: An R Package to Browse Over Longitudinal Data Graphically and Analytically in R," *The R Journal*, 14, 6–25, https://doi.org/10.32614/RJ-2022-023.

Tukey, J. W. and Tukey, P. A. (1985), "Computer Graphics and Exploratory Data Analysis: An Introduction," in *Proceedings of the Sixth Annual Conference and Exposition: Computer Graphics*, vol. 85, pp. 773–785.

Tukey, J. W. et al. (1977), *Exploratory Data Analysis*, vol. 2, Springer.

UNESCO (2026), "UNESCO Datahub," Online Data repository, accessed: 2026-02-12.

United Nations, Department of Economic and Social Affairs, Population Division (2019), *Family Planning and the 2030 Agenda for Sustainable Development: Data Booklet*, ST/ESA/SER.A/429, United Nations.

United Nations Department of Economic and Social Affairs, Population Division (2024), "Family Planning Data," https://www.un.org/development/desa/pd/data/family-planning-data, accessed: 2025-11-25.

Unwin, A. and Valero-Mora, P. (2018), "Ensemble Graphics," *Journal of Computational and Graphical Statistics*, 27, 157–165.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., and Gelman, A. (2024), "loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models," R package version 2.8.0.

Vehtari, A., Gelman, A., and Gabry, J. (2017), "Practical Bayesian Model Evaluation using Leave-One-Out Cross-Validation and WAIC," *Statistics and Computing*, 27, 1413–1432.

Wang, E. and Cook, D. (2020), *tsibbletalk: Interactive Graphics for Tsibble Objects*, r package version 0.1.0.

— (2021), "Conversations in Time: Interactive Visualization to Explore Structured Temporal Data," *The R Journal*, 13, 461–469, https://doi.org/10.32614/RJ-2021-050.

Wang, E., Cook, D., and Hyndman, R. J. (2020), "A New Tidy Data Structure to Support Exploration and Modeling of Temporal Data," *Journal of Computational and Graphical Statistics*, 29, 466–478.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016), "Functional Data Analysis," *Annual Review of Statistics and its Application*, 3, 257–295.

Warin, T. (2021), *Data Pipeline with R*.

Watanabe, S. and Opper, M. (2010), "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory." *Journal of Machine Learning Research*, 11.

Weber, F., Ickstadt, K., and Glass, Ä. (2022), "shinybrms: Fitting Bayesian Regression Models Using a Graphical User Interface for the R Package brms." *R J.*, 14, 96–120.

WHO (2023), "UNSD Mthodology," https://unstats.un.org/unsd/methodology/m49/.

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

Wickham, H., Cook, D., and Hofmann, H. (2015), "Visualizing Statistical Models: Removing the Blindfold," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 8, 203–225.

Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023), *dplyr: A Grammar of Data Manipulation*, r package version 1.1.4.

Wickham, H., Grolemund, G., et al. (2017), *R for Data Science*, vol. 2, O'Reilly Sebastopol.

Wickham, H., Pedersen, T. L., and Seidel, D. (2025), *scales: Scale Functions for Visualization*, r package version 1.4.0.

Wilke, C. O. (2019), *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*, O'Reilly Media.

Wilkinson, L. (2011), "The Grammar of Graphics," in *Handbook of Computational Statistics: Concepts and Methods*, Springer, pp. 375–414.

Wilkinson, L. and Wills, G. (2008), "Scagnostics Distributions," *Journal of Computational and Graphical Statistics*, 17, 473–491.

Wills, G. (2011), *Visualizing time: Designing Graphical Representations for Statistical Data*, Springer Science & Business Media.

Woodruff, T. J., Parker, J. D., and Schoendorf, K. C. (2006), "Fine Particulate Matter (PM2. 5) Air Pollution and Selected Causes of Postneonatal Infant Mortality in California," *Environmental health Perspectives*, 114, 786.

World Bank (2026), "World Bank DataBank," https://databank.worldbank.org/, accessed: 12 February 2026.

World Health Organization (2026), "WHO Data Portal," Online data repository, accessed: 2026-02-12.

Xu, X., Shi, K., Huang, Z., and Shen, J. (2023), "What Factors Dominate the Change of PM2. 5 in the World from 2000 to 2019? A Study from Multi-Source Data," *International Journal of Environmental Research and Public Health*, 20, 2282.

Yaffee, R. (2003), "A Primer for Panel Data Analysis," *Connect: Information Technology at NYU*, 8, 1–11.