Exploring the Work Values of Irish Youth: A follow up study of Growing Up in Ireland child cohort

Bharvi Dhall

A thesis presented in fulfilment of the requirements for the degree of Research Masters

Supervisor(s): Dr. Catherine Hurley and Dr. Katarina Domijan

Department of Mathematics and Statistics

Maynooth University

Maynooth Co. Kildare, Ireland

October 2022

Acknowledgements

Throughout my work on this dissertation, I have received an abundance of support and assistance. First, I would like to acknowledge and thank my supervisors, Dr. Catherine Hurley and Dr. Katarina Domijan, whose knowledge and skill assisted me in doing the analysis for this research. These years have been difficult, and I have not been the best student to my supervisors, but they have encouraged me and pushed me through my research. I would like to acknowledge Science Foundation Ireland (Grant number 18/CRT/6049) and Maynooth University for financing this research. The Lockdown of Covid-19 was frustrating, but my Hamilton Institute colleagues were always supportive. I would like to thank Professor Ken Duffy for his guidance and support.

I am grateful to my parents and my friends who stood by my side during the dreary days. Abundance of love to my dog(Ollie) for being my source of joy.

Abstract

Our research investigates social, cognitive, and environmental factors that may be associated with the work values of 17–18-year-olds in Ireland and examines if these findings remain consistent at age 20. It also explores if both genders have diverse work value preferences and scrutinizes the role of school type and career guidance in influencing work values. The data is obtained from the Child Cohort of the National Longitudinal Study of children, the Growing Up in Ireland (GUI) wave 3 and wave 4. We use visualizations to build an understanding of work value choices and model results. Ordered heatmaps have been used for picking interesting features and mapping the association between work values and potential factors influencing these work values. Self-beliefs of adolescents about work, family, and religion, along with cognitive scores, are found to be associated with work value selection, while career guidance and school type have limited importance. The results reveal that the two genders have different choices of work values. Work values can be broadly classified as intrinsic which are bringers of inner peace and psychological satisfaction like an interesting job and extrinsic which are led by external reward-based job aspirations like high income. Our research indicates that girls tend to pick more intrinsic work values compared to boys. Contradictory to the findings from other similar studies carried out in different countries, job security was found to be a more popular choice among Irish adolescent males compared to females, but became a more desirable choice for females as the adolescents transitioned to adulthood.

Table of Contents

List of Figures

List of Tables

List of Code Snippets

CHAPTER 1

Introduction

1.1 Background and Motivation

The adolescent years have a significant influence in preparing and shaping the career of individuals [Chow et al., 2017]. The work values in early adulthood are found to be predictive of career and life satisfaction in later years [Axelsson et al., 2005]. Teenage work aspirations have a substantial impact on post-secondary schooling, which in turn affects future employment [Johnson and Mortimer, 2011]. Improving the transition from education to the workforce has been an Organisation for Economic Co-operation and Development (OECD) policy objective for decades [Zancajo, 2011] as youth unemployment is connected with substantial economic and social costs, and long-term unemployment is a risk factor for poor physical and mental health outcomes [Krug and Eberl, 2018]. Numerous factors, such as an individual's personality, educational attainment, available educational opportunities and advice, socioeconomic class, and parental expectations, might influence career aspirations [Fouad, 2007]. The awareness of factors influencing work values can support students in accomplishing their educational and career ambitions and assist them in their career-related decision-making process. This thesis investigates the career considerations of Irish adolescents through the child cohort of the National Longitudinal Study of Children, the Growing Up in Ireland (GUI) dataset [Murphy et al., 2018] and tracks them at two ages: 17/18 and 20. In addition, it seeks to study the relevant factors associated with these considerations. This information will serve as a starting point for future study that will aid in the improvement of policies and processes pertaining to the Irish school context. Section 1.2 outlines various chapters described in this thesis.

1.2 Outline of Thesis

This thesis comprises of six chapters. Chapter 1 is an introduction to different chapters described in this thesis. Chapter 2 discusses the process for setting up the R workflow and introduces several fundamentals for familiarising oneself with GUI data set. Visualization techniques used for exploration of categorical variables have been discussed in chapter 3. Chapter 4 explores the work values of Irish adolescents and investigates social, cognitive, and environmental factors that may be associated with the work values of 17–18-year-olds in Ireland. Chapter 5 examines if the findings from wave 3 remain consistent in the follow-up wave 4 study of GUI at age 20. Chapter 6 discusses the conclusions from the analysis of GUI data and highlights potential future work.

1.3 Reproducible workflow for GUI data

Despite the fact that the GUI datasets are accessible via the Irish Social Science Data Archive (ISSDA) [ISS], their analysis needs substantial background knowledge, data processing, and linking. Due to the absence of a clear guide on how to begin working with these datasets, everyone must undergo a demanding learning process. This procedure is extremely inefficient because it prevents researchers from utilising the data, requires significant resources and time to complete the learning process, and increases the likelihood of errors occurring when multiple research groups independently replicate the same lengthy procedure. We overcome these challenges by providing code samples and offering thorough explanations of the data processing approaches. We anticipate that many research organisations will use this work, hence increasing the repeatability and validity of the findings. This section of the thesis outlines the project workflow that we have found to be efficient in making reproducible research reports using R statistical software [R Core Team, 2022] with 'markdown' [Allaire et al., 2019] for dynamic document generation, and tidyverse principles [Wickham et al., 2019] for datasets from GUI study [McNamara et al., 2019]

1.4 Visualizing Categorical Data: Exploratory Analysis

Numerous areas, including the behavioural and social sciences, public health, biomedical science, education, and marketing, utilise categorical data [Agresti, 2013]. Moreover, given that datasets frequently include a variety of data types, categorical data can provide crucial context for comprehending continuous variables [Kosara et al., 2006]. Growing up in Ireland data has a large number of categorical variables and it is often difficult to explore categorical variables. This chapter provides an overview of the existing methods for visualising categorical data, which in turn will help in the exploration of GUI data and other similar data sets.

1.5 Work Values of Irish Adolescents: An Exploratory Study

The work values in adolescent years are related to the career outcomes in later years [Chow et al., 2017]. They are also found to be reflective of life and career satisfaction and may lead to a higher quality of life Axelsson et al. [2005]. This chapter investigates social, cognitive, and environmental factors that may be associated with the work values of 17–18-year-olds in Ireland. It also explores if both genders have diverse work value preferences and scrutinizes the role of school type and career guidance in influencing work values. The GUI child cohort wave 3 data has been used for this analysis.

1.6 Work Values Follow-up 20 year olds

Adolescence and early adulthood are the most formative years in terms of professional dedication, as this is when major life choices are made [Super et al., 1963]. However, it is considered that adolescent career intentions are fragile and alter periodically until adulthood [Super, 1980]. This research is a follow-up of research conducted on the work values of Irish adolescents between the ages of 17 and 18 using GUI wave 3 dataset and intends to determine if the selection of preferred work values at the age of 20 using GUI wave 4 data remains consistent with the work values selected in adolescent years.

CHAPTER 2

Reproducible workflow for GUI data

2.1 Introduction

Growing Up in Ireland (GUI) is the nation's first longitudinal study of children and youth. It is a study undertaken by the Economic and Social Research Institute (ESRI) and Trinity College Dublin using government funding. The study began in 2006 and follows two cohorts of children beginning at age 9 (child cohort) and 9 months (infant cohort). The purpose of the study was to acquire an understanding of the experiences of Irish children over time and to discover the fundamental elements that promote or inhibit growth. GUI is a crucial research instrument for everyone interested in child development. To date, the knowledge acquired from GUI has significantly influenced government policies. For instance, evidence addressing the prevalence of obesity [Williams et al., 2013] was employed to develop the Health Service Executive's document titled 'Tackling Childhood Obesity' (HSE). The research presented in this thesis is focused on the child cohort. The child cohort had first been surveyed at age 9 years in 2007 and were followed up at ages 13, 17/18 and 20 years. The total population of all 9-year-olds was 56,497, of which 8000 children were first chosen for the study, representing one in every seven children

(14 percent of 9-year-old residents in the country). Because accurate counts were impossible to obtain as the process entailed obtaining agreement from the children and their families, the final sample of 8,568 youngsters was selected at random. A comprehensive list of all schools in Ireland was compiled with the support of the National Education System, and the schools served as the primary sampling unit from which a sample of 9-year-olds was generated.

While the GUI datasets are accessible through the Irish Social Science Data Archive (ISSDA), their study requires extensive baseline knowledge, data processing, and linkage. Although there is sufficient instruction and training available for analysing GUI data with software like SPSS, there is no equivalent information available for R statistical software. R has the advantage of reproducibility, a rich ecosystem of packages for data manipulation, visualisation and analysis and the ability for researchers to implement their own statistical methodology extensions. The R workflow discussed in this chapter attempts to bridge this gap. We provide code samples and explain the data processing techniques in extensive detail. We anticipate that numerous research organisations will utilise this work, hence boosting the repeatability and validity of the findings. The code provided can be utilized in analysing similar longitudinal studies of other countries, like Growing Up in Scotland and Growing Up in New Zealand.

This section of the thesis outlines the project workflow that we have found to be efficient in making reproducible research reports using R statistical software [R Core Team, 2022] with ‘markdown’ [Allaire et al., 2019] for dynamic document generation, and tidyverse principles [Wickham et al., 2019] for datasets from the GUI study [McNamara et al., 2019]. Section 2.2 explains the process of setting up the data in R and discusses the fundamentals of imported data structures. Section 2.3 describes the GUI naming convention and provides a collection of R functions that facilitate the construction of subsets in a quick and efficient manner.

2.1.1 Data: Description and Naming Convention

Given that GUI is a longitudinal study, the same individuals are followed over time. This information has been published in multiple waves and two distinct formats: Anonymised Microdata Files (AMFs) and Research Microdata Files (RMF). The analysis carried out in our research utilises Growing Up In Ireland Cohort '98 (Child Cohort) Wave 3 - 17/18 years. Through the Irish Social Sciences Data Archive, these datasets are made available to academics in a secure and non-identifying manner (ISSDA) [1]. The AMF is a publicly accessible anonymized dataset designed to safeguard the confidentiality of participants, but the RMF is a more detailed dataset including raw data that can only be accessed by appointment of the researcher as an Officer of Statistics by the Central Statistics Office (CSO). [2]

The GUI study is comparable to other longitudinal studies like Growing Up in Australia, Growing Up in New Zealand, and Growing Up in Scotland. These studies utilise a different naming convention than social science datasets and have a distinctive design. We have used the anonymized (AMF) dataset to conduct the research presented in this thesis. The anonymised version of the GUI dataset is accessible in two distinct naming schemes. Convention A is based on a questionnaire, while Convention B is theme-based. Convention A is used for this paper. In accordance with this convention, variable names consist of three components: questionnaire (alpha), section (alpha), and question indication (alpha/numeric). Figure 2.1 provides an illustration of this [3]. The first two characters in figure 2.1 represent the questionnaire in which the question was included, the third character represents the section within the questionnaire, and the last two characters represent the question number. MML12 denotes the twelfth question in Section L of the Primary Caregiver Main Questionnaire in the Child Cohort Wave 1 data collection.

2.1.2 Composition of the Sample and Response rates

8,568 9-year-olds and their families were questioned for the GUI study. Since the GUI study is longitudinal, these families were contacted again to re-interview the

[1] Growing Up In Ireland Cohort '98 (Child Cohort) Wave 3 - 17/18 years, 2016 data set can be accessed through the Irish Social Science Data Archive(ISSDA) `https://www.ucd.ie/issda/data/guichild/guichildwave3/`

[2] Growing Up In Ireland Cohort '98 (Child Cohort) Wave 3 - 17/18 years RMF dataset can be requested through CSA `https://www.cso.ie/en/aboutus/lgdp/csodatapolicies/dataforresearchers/rmfregister/`

[3] This figure has been obtained from the document Variable naming and longitudinal data dictionary - Child - Wave 1 and 2 which can be obtained from the Irish Social Science Data Archive website`https://www.ucd.ie/issda/data/guichild/guichildwave1/`

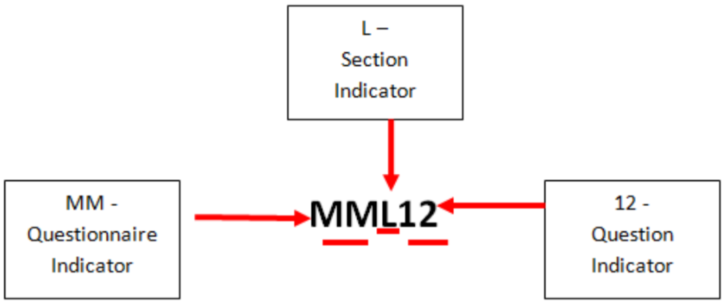


Figure 2.1: Naming Convention A for GUI dataset

Study Child at 13 years old in Wave 2. Nevertheless, 665 families declined to participate, and they were excluded from the target group since they no longer resided in Ireland. Only 7525 of the remainder 7903 families took part in the survey, yielding an 89 percent response rate.

A total of 8,277 households were given to field interviewers for the third round of interviews, which took place while the research participants were between the ages of 17 and 18. Families were invited to participate in the 17/18-year wave of fieldwork if they had been interviewed during the study's first wave (when the Study Child was 9) but not in the wave after that. Families for whom there was no current address or where the family had specifically requested not to be contacted for additional waves of the study were not attempted to be reinterviewed by the study team.

At Wave 3, there were 6,216 finished households in total, corresponding to a 76 percent response rate. The Study Team contacted all former participants for Wave 4 (at age 20) unless the family had previously categorically declined to be contacted in subsequent waves of the study or was ineligible (for example, the entire family had moved abroad or the young adult had sadly passed away).

In total, 7,925 20-year-olds' contact information was provided to interviewers in Wave 4 out of which 5,190 20-year-olds completed questionnaires IRELAND'S [2021].

2.2 Workflow in R

2.2.1 Setting Up File Directory Structure

To promote reproducible workflows, we use RStudio [RStudio Team, 2022] as an integrated development environment (IDE) for the R programming language. All files involved with data analysis are organised as RStudio Projects. RStudio additionally gives simpler access to version control technologies such as 'Git'[Chacon and Straub, 2014]. Git maintains a series of snapshots (commits) of all tracked files, while a git repository keeps track of the complete project. Git facilitates examining the history of all commits, visually comparing differences across versions, and reverting the repository to a specified former state. We exchange and collaborate on projects using GitHub [4].

The raw GUI data and documentation are stored in a subdirectory within the project directory. These data sub-directories are not tracked by Git. The R scripts within the project are created with the assumption that they will be executed by a new R process with the project directory as the working directory. This is how knitr/rmarkdown [Xie et al., 2013] documents function by default. When R is launched from the project's root directory, the working directory is set to the project directory.

When reading or writing files, the 'here' function from the 'here' package [Müller, 2020] is used to construct paths relative to the top-level directory. This package is simple to use and efficient for file reference when numerous users utilise the same piece of code.

2.2.2 Labelled data in R

The GUI datasets are available in SPSS, SAS, and Stata formats. These packages have created infrastructures for labelled data, which consists of variable and value labels as metadata. Variable labels are descriptions of the variables, whereas value labels describe the values a variable can take. Several R libraries can be used to

[4] GitHub is a code hosting platform for version control and collaboration; website `https://github.com/`. GitHub repository containing the code for the analysis; website `https://github.com/Bharvi-Dhall/Exploring-work-values-using-GUI-data`

import SPSS, SAS, and Stata data sets into R, but since we intend to utilise the labelled data features, we use 'haven :read_spss' to import the SPSS dataset and save it as a tibble [Müller and Wickham, 2022]. Code snippet 2.1 gives an example of reading a dataset using here().

Listing 2.1: Reading the Wave 3 Convention A dataset using here()

```
here::i_am("Repo_name/Folder_name/File.Rmd")
#This line of code directs to the R script where code is located.

wave_3 <- haven::read_spss(here("Repo_name","Folder_name","Sub_folder","
    Dataset_wave3.sav"), user_na = TRUE)
#here guides the path to the dataset
```

To reduce file size and improve loading times, we often save the objects as an RDS file, which generates a serialised version of the dataset (refer 2.2).

Listing 2.2: Saving and Reading data as RDS

```
#Saving Data as .RDS
saveRDS(wave_3, here("Repo_name","Folder_name","Sub_folder", "Dataset_
    wave3.RDS"))

#Reading data as .RDS file and storing it wave_3
wave_3 <- readRDS(here("Repo_name","Folder_name","Sub_folder", "Dataset_
    wave3.RDS"))
```

Variable Labels Variable labels are stored in the 'label' attribute of each variable. Code snippet 2.3 gives sample R code for extracting variable labels with relevant output.

Listing 2.3: Extracting variable labels

```
wave_3$CognitiveVocabularyTotal %>% attr("label")
wave_3$w3cq_workbelief %>% attr("label")

#Output
[1] "Cognitive test-Vocabulary test"
[1] "YP Belief in the value of work Wave 3"
```

Value Labels A labelled vector in SPSS is a data structure that allows assigning text labels to specific values. SPSS file variables with value labels imported using 'haven::read_sav' are given a class called 'haven_labelled_spss' and 'haven_labelled'.

Not all variables in the GUI datasets are labelled. Code snippet 2.4 demonstrates the function class() which returns the labels for the variables.

Listing 2.4: Extracting value labels

```
wave_3$CognitiveVocabularyTotal %>% class()
wave_3$w3cq_workbelief %>% class()


#Output
> wave_3$CognitiveVocabularyTotal %>% class()
[1] "haven_labelled_spss" "haven_labelled"
[3] "vctrs_vctr" "double"
> wave_3$w3cq_workbelief %>% class()
[1] "numeric"
```

In R, the value labels are stored in a 'labels' attribute. For non-labelled variables, this attribute will be empty (refer code listing 2.5).

Listing 2.5: Attribute information for different value labels

```
wave_3$CognitiveVocabularyTotal %>% attr("labels")
wave_3$w3cq_workbelief %>% attr("labels")


#Output
> wave_3$CognitiveVocabularyTotal %>% attr("labels")
 2 or less 17 or more Refusal Dont Know
2 17 98 99
> wave_3$w3cq_workbelief %>% attr("labels")
NULL
```

The function 'is.labelled' from the haven package can be used to determine if a variable is haven_labelled. The code fragment 2.6 demonstrates that the variable w3cq_workbelief is not haven_labelled, although CognitiveVocabularyTotal is.

Listing 2.6: Investigating is the variable is haven_labelled

```
wave_3$CognitiveVocabularyTotal %>% haven::is.labelled()
wave_3$w3cq_workbelief %>% haven::is.labelled()

#Output
> wave_3$CognitiveVocabularyTotal %>% haven::is.labelled()
[1] TRUE
> wave_3$w3cq_workbelief %>% haven::is.labelled()
[1] FALSE
```

2.2.3 Data dictionary

A data dictionary provides metadata about the data and facilitates rapid exploration of variables of interest in R. They are utilised to offer comprehensive information on the contents of a dataset or database, including the names of measured variables, their data types or formats, and textual descriptions. A data dictionary offers a succinct reference to comprehending and utilising data. Package 'labelled' [Larmarange, 2018] provides a number of useful tools to manipulate variable and value labels using the haven_labelled and haven_labelled_spss classes introduced in haven.

Function 'generate_dictionary' generates a data dictionary from the variable names and the labels. It is useful to store this in an object before coercing the labelled variables to a standard R class. The code snippet 2.7 demonstrates steps to create a data dictionary and uses the head() function to view the first few rows of it.

Listing 2.7: Generating the data dictionary

```
wave_3_dict <- labelled::generate_dictionary(wave_3)

# to look at the first few rows
wave_3_dict%>% head()

#Output
> wave_3_dict%>% head()
 pos variable label col_type values
```

```
1 ID Household ID dbl
2 WGT_17YRa Weighting Factor - 17yr~ dbl
3 GROSS_17YRa Grossing Factor - 17yr ~ dbl
4 WGT_17YRb Weighting Factor - 17yr~ dbl
5 GROSS_17YRb Grossing Factor - 17yr ~ dbl
6 xxwave1 Family participated in ~ dbl
```

The R libraries 'dplyr' and 'stringr' encloses numerous functions to lookup subsets of variables from the data dictionary. For instance, the str_detect function in the stringr package picks all variables containing the string 'cq3g4' (see code snippet 2.8).

Listing 2.8: Illustration of the dplyr and stringr functions to extract subsets of data

```
#Subsetting the data dictionary for variable CognitiveVocabularyTotal and
   w3cq_workbelief
wave_3_dict%>% filter(variable=="CognitiveVocabularyTotal" |variable=="
   w3cq_workbelief")

#Display position, variable name, label and column type of all variables
   containing sstring "cq3g4"
wave_3_dict %>% filter(str_detect(variable, "cq3g4"))%>% select(pos,
   variable, label, col_type)
```

2.2.4 Coercing to a standard R class

We coerce 'haven_labelled' vectors to a conventional R class after importing, as many of the packages we interact with for analysis may lack methods for haven-labelled objects. It should be noted that variable labels can be used to clean and recode data before conversion. For example, if the answers "Refusal" and "Don't know" were to be removed from the analysis, they might be recoded as missing values using the value labels at this point in the procedure. However, each wave of GUI datasets has hundreds of variables, necessitating caution, as illustrated in the code snippet 2.9.

Listing 2.9: Investigating value labels before recoding

```
# The below code generates a data dictionary and then filters all the
   value labels which are not NA and groups them to find out the number
   of observations for each value labels. The labels and their counts are
    then arranged in ascending orders.
wave_3_dict <- labelled::generate_dictionary(wave_3) %>%
  labelled::lookfor_to_long_format()
wave_3_dict %>%
  filter(!is.na(value_labels)) %>%
  select(value_labels) %>%
  group_by(value_labels) %>%
  tally() %>%
  arrange(desc(n))

#Output
# A tibble: 1,159  2
   value_labels n
   <chr> <int>
 1 [8] Refusal 887
 2 [9] Dont Know 820
 3 [1] Yes 363
 4 [2] No 225
 5 [98] Refusal 88
 6 [9] Dont know 67
 7 [1] yes 59
 8 [99] Dont Know 56
 9 [2] Ordinary 40
10 [999] Dont Know 37
#  with 1,149 more rows
```

At this stage, we skip performing data cleaning and recoding on the full dataset, opting instead to perform these operations on smaller groups of variables for subsequent analysis. For example, for choose all that apply (CATA) variables, non-selection is coded as a missing value as opposed to a separate value, and the user would need to recode this as a new factor level beforehand in order to keep the distinction between non-selection and non-answer. See an example in the code snippet 2.10 illustrates that Yes, Refusal and Don't Know have been coded as 1, 8, and 9 respectively for the variables starting with 'cq3g2'.

Listing 2.10: Extracting attribute labels for CATA variables

```
wave_3 %>%
  select(starts_with("cq3g2")) %>%
     sapply(attr, "labels")


# Ouput
          cq3g2a cq3g2b cq3g2c cq3g2d cq3g2e cq3g2f
Yes 1 1 1 1 1 1
Refusal 8 8 8 8 8 8
Dont Know 9 9 9 9 9 9
          cq3g2g cq3g2h cq3g2i cq3g2j cq3g2k
Yes 1 1 1 1 1
Refusal 8 8 8 8 8
Dont Know 9 9 9 9 9
```

It is possible to convert 'haven_labelled' vectors into numeric and factor classes. As there are hundreds of variables, the vast majority of which are categorical, we have chosen to convert them into factors. Moreover, converting to numeric eliminates value labels, while converting to factors replicates value labels to factor levels. The factor levels are arranged according to value codes. A lot of labelled variables in the GUI are ordinal or likert scales with labelled first and last points.

We work with factors rather than strings as they are useful in modelling (e.g. one can control the ordering of factor levels) and in visualisation as they allow categorical variables to be mapped to aesthetic attributes [McNamara and Horton, 2018].

Packages 'haven', 'labelled', and 'sjlabelled' provide useful functions for the labelled data features. They overlap in functionality and, unfortunately, use different naming schemes for the same operations. For example 'haven::as_factor' converts the labelled variables to factors, as does 'sjlabelled::as_label'. However 'sjlabelled::as_factor' performs a different operation. See examples in code snippet 2.11.

Listing 2.11: Exploring useful functions for labelled data

```
haven::as_factor(wave_3$CognitiveVocabularyTotal)%>% levels()
forcats::as_factor(wave_3$CognitiveVocabularyTotal)%>% levels()
sjlabelled::as_label(wave_3$CognitiveVocabularyTotal, add.non.labelled =
    TRUE)%>% levels()
sjlabelled::as_factor(wave_3$CognitiveVocabularyTotal)%>% levels()
```

Because the naming conventions can mask functions from other libraries, e.g., loading 'sjlabelled' library after 'forcats' masks 'forcats::as_factor' we access functions using double colon operators as opposed to loading the libraries.

We use 'dplyr::mutate_if' and 'haven::is.labelled' to take variables that are labelled and convert them to factors (refer code 2.12). The numerical variables are unchanged.

Listing 2.12: Conversion of labelled variables to factors

```
wave_3 <- wave_3 %>%
  mutate_if(
    haven::is.labelled,
    ~haven::as_factor(.x)
  )
```

The attribute information is removed from the labelled variables once they have been turned into factors (see code snippet 2.13).

Listing 2.13: Checking for attribute Information

```
wave_3$CognitiveVocabularyTotal %>% class()
wave_3$CognitiveVocabularyTotal %>% attr("labels")


#Output
> wave_3$CognitiveVocabularyTotal %>% class()
[1] "factor"
> wave_3$CognitiveVocabularyTotal %>% attr("labels")
NULL
```

2.3 Variable Naming

The GUI survey was carried out by recording the responses to various questionaires as illustrated in section. Every individual was requested to respond to a main and a sensitive questionaire. Main Questionaire consisted of more broad and less sensitive questions whereas the sensitive questionnaire dealt with more sensitive questions like one's behaviours and activities. Except for derived variables, variable names begin with a prefix and correspond to a question number on the questionnaire from which they originated. Variables for Wave 3 of the child cohort are prefixed with 'pc3' for Parent One, 'sc3' for the Parent Two and 'cq3' for the Young Person. The '3' indicates that the data came from the third wave of the project [Murphy et al., 2018]. The variable name pc3f1educ, for instance, pc3 indicates that it is derived from the primary care givers questionnaire from wave 3 and section F (Respondent's Lifestyle), with the question 'What is the greatest degree of schooling you have completed?'. If the question was from the sensitive questionnaire, a 's' is added to the variable name; for instance, the variable name for question A1 from the Young Person Sensitive Questionnaire was cq3sa1.

The only exceptions to this convention are the household grid variables, which are prefixed with the person number. For example, the variable for the sex of the person on line 1 of the grid is 'P1sexW3' where 'W3' indicates Wave 3 data [Murphy et al., 2018].

The GUI wave 3 dataset is a collection of variables obtained in the order listed:

- Household Grid (p1xxW3, p2xxW3)
- Parent One Main Questionnaire (pc3)
- Parent One Sensitive Questionnaire (pc3s)
- Parent Two Main Questionnaire (sc3)
- Parent Two Sensitive Questionnaire (sc3s)
- Young Person Main Questionnaire (cq3)
- Young Person Sensitive Questionnaire (cq3s)
- Standardised Scale Scores (w3)

Table 2.1: Top 10 counts of Variables corresponding to different themes

Section	Count
cq3	618
pc3	242
p3q	144
sc3	76
w3c	37
w3p	16
w3s	14
w3i	12
w3f	6
w3T	5

- Physical Measurements (w3)
- Derived Variables (w3)

Table 2.1 lists the top 10 counts of the variables listed under different sections.

The GUI Research Team suggests using the data in conjunction with the Questionnaire Documentation to gain the most efficient overview of the data. The code snippet 2.15 provides a table indicating the counts of variables corresponding to each theme.

Listing 2.14: Checking for attribute Information

```
wave_3_dict %>%
  select(variable) %>%
  unique() %>%
  mutate(name_type = substr(variable, start = 1, stop = 3)) %>%
  group_by(name_type) %>%
  count() %>%
  arrange(desc(n))
```

Keeping the original codes for variable names facilitates the selection of columns containing answers to the same questions or the identification of questions from a subset of the survey. For example the code in listing 2.15 selects answers to the question: "Consideration when choosing a job".

The code 2.16 explains how to subset variables using the select_with() function and creates a correlation plot of all subset variables using the corrplot() function obtained from the package 'corrplot' [Wei et al., 2017].

Listing 2.15: Using codes to subset variables

```
df <- wave_3 %>% select(starts_with("cq3g2"))


#identify questions from a subset of the survey
wave_3 %>% select(starts_with("p1")) %>% sjlabelled::get_label() %>%
    enframe()
```

One could use R's text analysis tools to select variables using their names or labels from data dictionary and to summarise/visualise information about the structure/themes of the survey. Thus it bridges the gap between Naming Convention A and B.

The function 'sjlabelled::label_to_colnames' converts variable label to variable name. However, the variable labels for GUI data are excessively lengthy. To develop custom labels for subsets of variables, the labels are altered using the function 'Hmisc::upData' (see code snippet 2.16).

Listing 2.16: Creating Custom Labels

```
var.labels <- c(cq3g2a = "high_inc",
               cq3g2b = "training_opp",
               cq3g2c = "interesting_job",
               cq3g2d = "flex_hours",
               cq3g2e = "time_off",
               cq3g2f = "good_for_career",
               cq3g2g = "being_boss",
               cq3g2h = "help_society",
               cq3g2i = "job_security",
               cq3g2j = "travel_abroad",
               cq3g2k = "other"
 )
```

```
df <- Hmisc::upData(df,labels = var.labels)
df <- sjlabelled::label_to_colnames(df) # uses sjlabelled
```

2.4 Data wrangling using tidyverse

Data from large surveys can introduce unique issues in data wrangling. Tidyverse principles when dealing with factor variables produce simpler and less fragile code than base R [McNamara and Horton, 2018]. They introduce users to a collection of data structures, functions, and operations that make dealing with data more intuitive compared to base R. This section lists a few examples to illustrate the use of functions in tidyverse for manipulating and processing the GUI variables.

Missing Values The survey datasets like GUI are abundant with categorical variables which have factor levels as NA, the 'forcats::fct_explicit_na' is used to recode all the missing values as 'not selected'.

Listing 2.17: Recoding missing values as 'No'

```
wave_3 <- wave_3 %>%
  mutate(across(starts_with("cq3g2"), ~fct_explicit_na(.x, "No")))
```

Creating Derive Variables To facilitate analysis, many variables from the GUI dataset were utilized to generate new variables known as derived variables. The forcats package has been used to collapse factor levels. For instance, the functions 'forcats::fct_collapse' and 'dplyr::case_when' were used to create type of school variable using the variables the number of girls and number of boys in the school from the principal questionnaire.

Listing 2.18: Creating derived variable: school_type

```
wave_3 %>% mutate(p3q4b = fct_collapse(p3q4b, no = "0", other_level = "yes
    ")) -> wave_3
wave_3 %>% mutate(p3q4a = fct_collapse(p3q4a, no = "0", other_level = "yes
    ")) -> wave_3
```

```
wave_3 <- wave_3 %>% mutate(school_type = case_when(
   p3q4b == "yes" & p3q4a == "yes" ~ "mixed",
   p3q4b == "yes" | p3q4a == "yes" ~ "single"))
```

Changing data type Numerous variables in the GUI dataset are ordinal categorical variables with multiple categories and lengthy names. If necessary, these variables could be converted into numeric variables for the analysis. Code snippet 2.19 lists the steps involved in selecting multiple columns at once (using across()) and recoding the variables as numeric.

Listing 2.19: Converting the variables to numeric data type

```
wave_3_dict%>% filter(str_detect(variable, "cq3g4a"))


wave_3 <- wave_3 %>% mutate(across(starts_with("cq3g4"),
~ fct_recode(.x, '1' = "not important at all", '6' = "very important",
    NULL = "Don't Know",NULL = "Refusal"))) %>%
mutate(across(starts_with("cq3g4"), ~ as.numeric(.x)))
```

CHAPTER 3

Visualizing Categorical Data: Exploratory Analysis

3.1 Introduction

Growing up in Ireland is a survey of Irish children's views and opinions on their lives in Ireland. The main objectives of this survey include identifying key factors for children's development, examining the progress and well-being of children at important time intervals from birth to adulthood and describing the lives of children in various age categories. The dataset includes responses to different questions answered by the study child, child's caregivers, principal, and teachers. The dataset mainly contained answers to the tick-box questions where respondents were asked to select one (or potentially more) of a fixed number of possible options. As a result, there is a large number of categorical variables present in the Growing up in Ireland dataset.

Categorical variables are classified in two types : ordinal and nominal. For instance, in the GUI data the variable the individuals were asked 'How often have/did your parents ask how you are/were getting on with your teachers/lecturers?' and the young person was asked to pick one response that applies out of: 'Don't Know', 'Refusal', 'Never or hardly ever', 'A few times a year', 'About once a month', 'Several times a month', and 'Several times a week'. This is an example of ordinal variable. Examples of the nominal type of categorical variables include gender with factor levels 'male' and 'female'. Exploring association among ordinal variables is relatively easier compared to the nominal variables as there is no natural ordering among the categories of a nominal variable. The data analysis for categorical variables can be performed using contingency tables (two or higher dimensions) of nominal and/or ordinal data. It mainly includes looking for independence structures or patterns using methods such as independence tests. Visualization techniques are an alternative way to perform analysis on these variables in order to detect structures present in the data which may not be evident from the numeric output such as test statistics.

Graphical displays for categorical data are useful in exploring association among variables and finding patterns. In addition to exploring patterns, they are also useful in model diagnosis and as graphical summaries for representing the modelling results [Friendly, 1992]. The display techniques such as barplot, mosaic, sieve, and association plots offer excellent visual summaries for either exploratory or model based tasks.

In section 3.2 we explore the GUI survey data using different visualization techniques such as bar plots, stacked bar plots, heatmap displays, network plots, alluvial plots, mosaic plots and UpSet plots.

3.2 Visualizations for categorical data

3.2.1 Bar Plot

Bar charts are basic visual representations of a categorical variable or a one-way frequency table, in which the categories of the variable are often displayed along the horizontal axis and the height of each bar symbolises the frequency of each category. Incorporating an additional categorical variable and providing a quick visual overview of the association between two categorical variables, bar plot variants such as grouped bar plots and stacked bar plots are becoming increasingly popular. Concerning survey data, stacked bar displays are commonly used to represent Likert-type responses. Figure 4.2 illustrates the frequency of boys and girls who selected a particular work value using basic bar plots.

3.2.2 Heatmap Plot

A heatmap is a two-dimensional colour representation of the magnitude of a value within a data matrix. One such example is the correlation matrix plot, in which the colour and its intensity denote the direction and value of correlation, respectively. Figure 3.1 shows a heatmap between the two ordinal variables from the wave 3 of the GUI dataset. The variable on the Y-axis reflects rating given by teenagers for how often their parents have discussed their future plans with five factor levels (1 = 'Never or hardly ever', 2 = 'A few times a year', 3 = 'About once a month', 4 = 'Several times a month', and 5 = 'Several times a week'), while the variable on the X-axis represents the importance of the mother in determining the respondents' future plans after high school (three factor levels). Apparently, there are a large number of adolescents who discuss their futures at home and who believe that their mothers play a crucial role in helping them determine their futures. This suggests that mothers have been engaging with the teenagers regarding their future during their teen years.

3.2.3 Network plot

Network plots can be used to visualize the relationship between two binary variables. The 'igraph' [Csardi, 2013] R package is useful in constructing network visualizations. For instance, to display career choices which were picked together such plots were created (see figures 3.2, 3.3 and 3.4). The node size in the display represents the frequency of picking a particular work value and the edges joining the nodes represent the count of combination picked together. Plot 3.2 displays individuals who selected high income, and visualised additional choices of work values which are frequently selected alongside high income. The counts of each work value being selected ere used to modify the node size in order to gain a better understanding of which options are more popular than others. Likewise, plot 2 depicts all students who have chosen interesting job. Plot 3.2 reveals interesting job and high income were often picked together. Figure 3.3 individuals who pick high income but do not

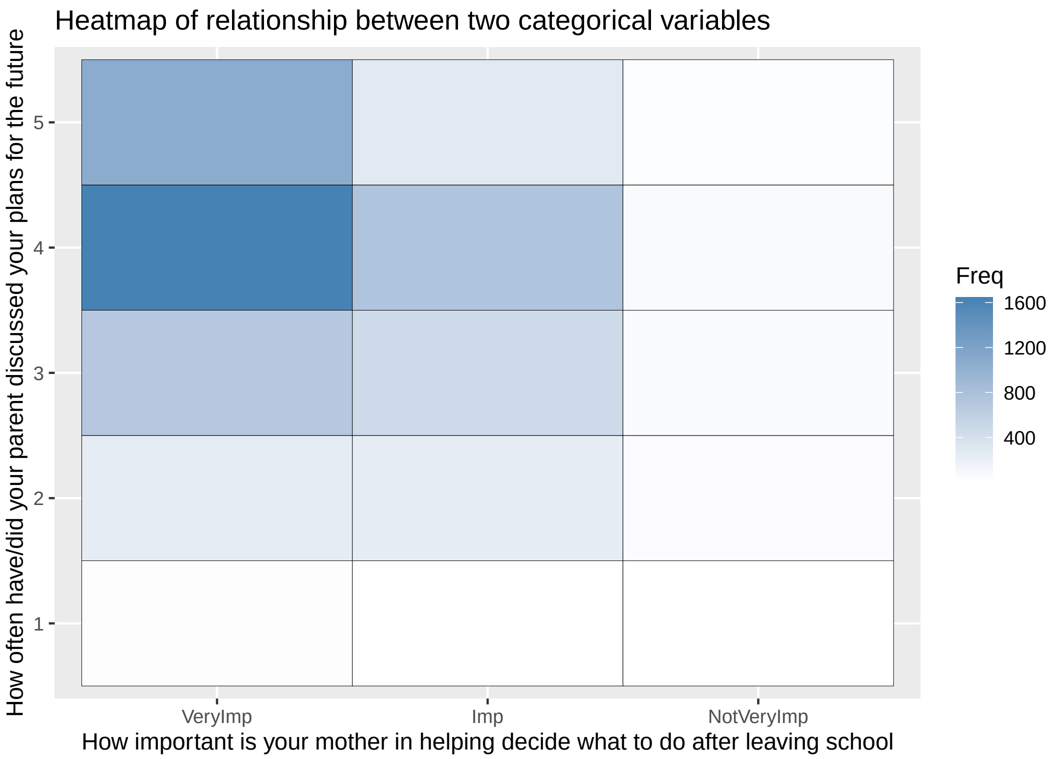


Figure 3.1: On the Y-axis of the heatmap is the frequency of parental interaction regarding future plans, and on the X-axis is the importance score for mother in determining what to do after high school. The plot demonstrates that those who view their mother as an essential decision-making resource discuss future plans with their parents more frequently.

select interesting job. The darkest and the thickest edge highlights that among students who do not interesting job, the work values high_income and interesting_job are often selected together. Similar plot has been constructed for individuals who pick interesting job but not high income (see figure 3.4). This plot suggests that the top two combinations for individuals who do not pick high_income include interesting_job with helping_society, and travel_abroad.

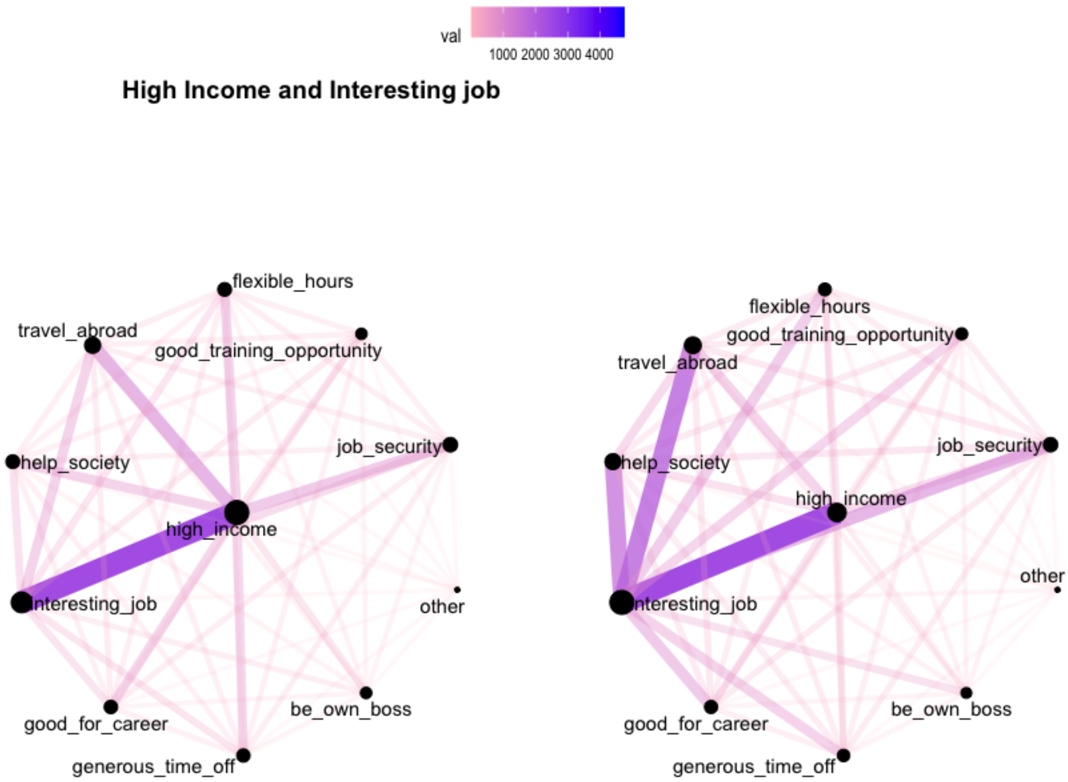


Figure 3.2: The first network diagram on the left displays individuals who selected high income and connects it to other options that were also selected. Similarly, the plot on the right depicts which work values were selected together with interesting job. The two plots are similar but travel_abroad and helping_society are far more popular for those individuals picking interesting_job than for those selecting high_income. The node size in the first plot illustrate that high_income and interesting_job are the most popular choices. The darker and thicker edge between these two nodes represent that this combination was more often picked together than others.

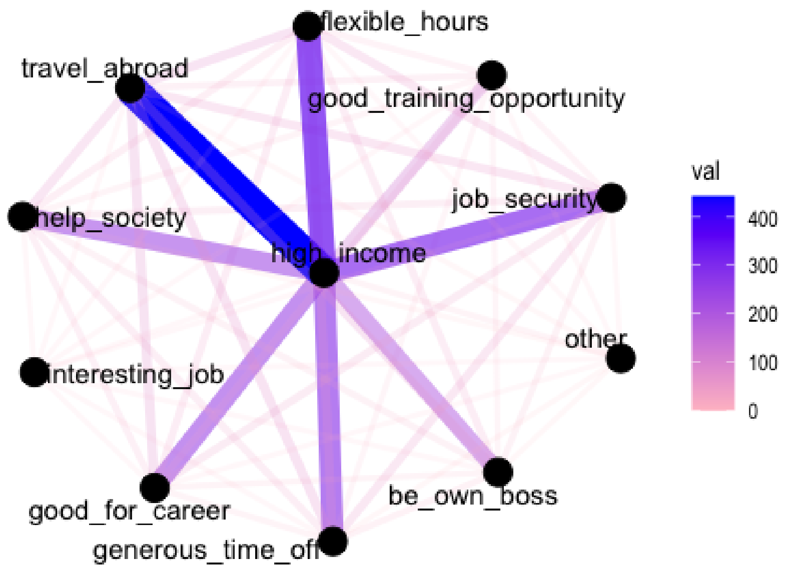


Figure 3.3: The network plot representing individuals who pick high income but do not select interesting job. In addition to an interesting job, students who choose a high income also pick travel abroad, flexible hours, and work stability

3.2.4 Alluvial Plot

Alluvial diagram is a form of flow diagram that represents changes and is comparable to Parallel Coordinates Plot (PCP), but for categorical data. The parallel axes reflect the categorical variables under consideration, with blocks on each axis representing the values for a given category (of another variable). The vertical dimensions of the blocks are proportional to the category's frequency. Each combination of the values of the categorical variables considered represents alluvia, and their breadth is proportional to their frequency. These graphs are useful for investigating various conditional and unconditional patterns in multivariate data sets.

Figure 3.5 shows an alluvial diagram of the categorical variables: whether dislike school, whether school helps and future education plan for the teenagers conditional on whether the respondent is male or female. The plots shows that a high frequency of male and female respondents in GUI data do not dislike school, and think that the school is helpful. These individuals aspire to go for a degree in future suggesting that the school might play a role in motivating students to go for a degree in their

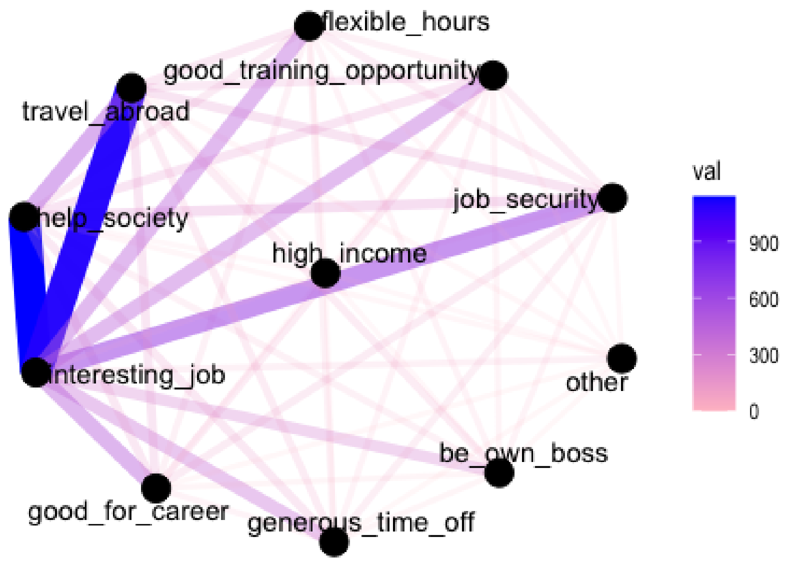


Figure 3.4: The network plot representing individuals who pick interesting job but do not select high income. In addition to an high income, students who choose a interesting job also pick travel abroad and helping society.

future.

3.2.5 Mosaic Plot

The mosaic display, proposed by Hartigan and Kleiner [1984] and extended by Friendly [1994] represents the counts in a contingency table directly by tiles whose area is proportional to the observed cell frequency. The mosaic plots are easily extendable to three-way and higher-way tables. These displays are useful during exploratory as well as model building step.

Figure 3.6 shows a mosaic plot for 3x3 contingency table for primary caregiver's education, young person's aspiration regarding education and young person's gender. The plot shows that the highest level of education for primary caregiver for a high proportion of teenagers (both male and female) is Leaving Certificate in GUI data. The plot also shows that a high proportion of the respondents with Leaving Certificate as highest education for primary caregiver would like to go for a degree in their future.

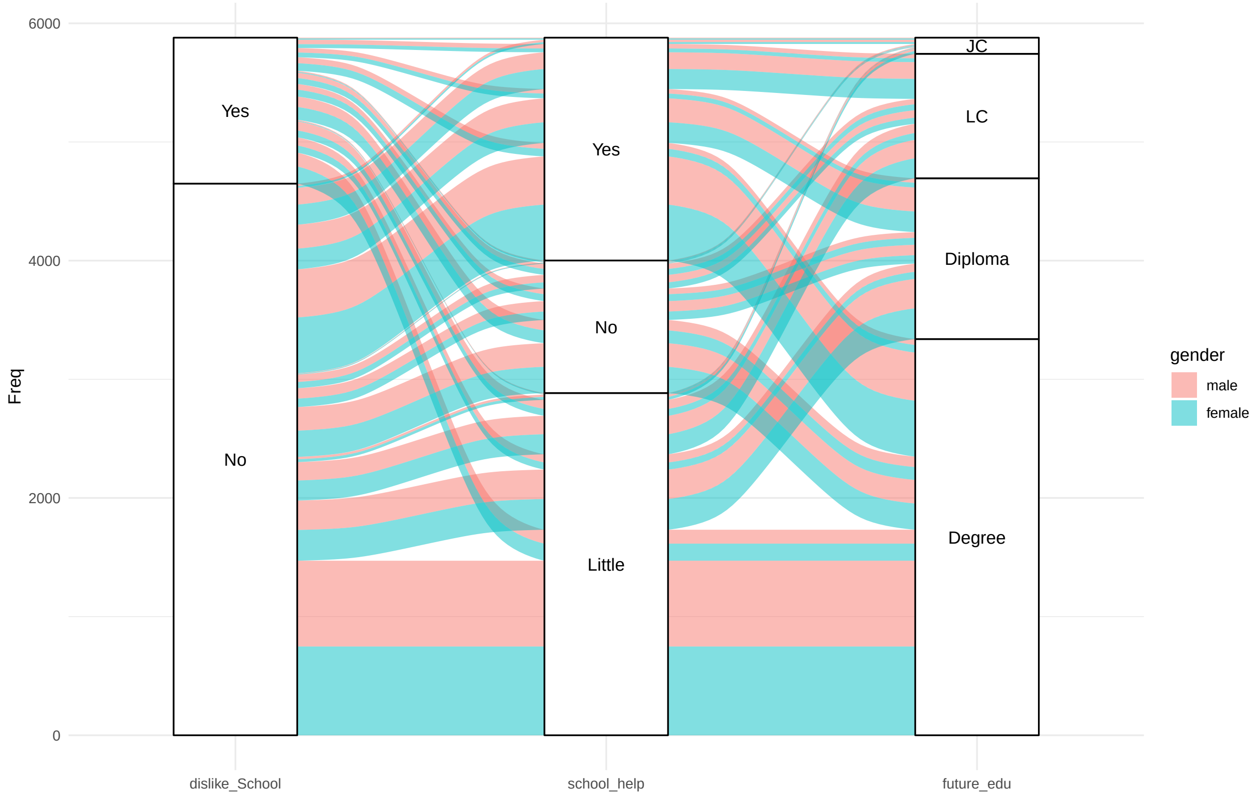


Figure 3.5: Alluvial diagram for variables: whether the respondent dislikes school, whether school is helpful for respondent and future education plan for the teenagers. The breadth of each alluvia represents the proportion of male and female respondents present in each combination of values of the variables.

3.2.6 UpSet Plot

UpSet plot Lex et al. [2014a] is a visualisation technique for analysis of sets and their intersections. The UpSet plots display a large number of set intersections in a matrix layout which enables effective presentation of the data such as number of elements in the set intersections. We use these plots for exploring Choose All that Apply (CATA) data in GUI dataset.

Figure 3.7 displays an UpSet plot for the work values selected by only females in the GUI data. Each respondent was asked to pick three most important choices out of the 11 possible work values. The plot shows that around 8 female students picked the combination of travelling abroad, flexible hours and time off which was not picked by any male student.

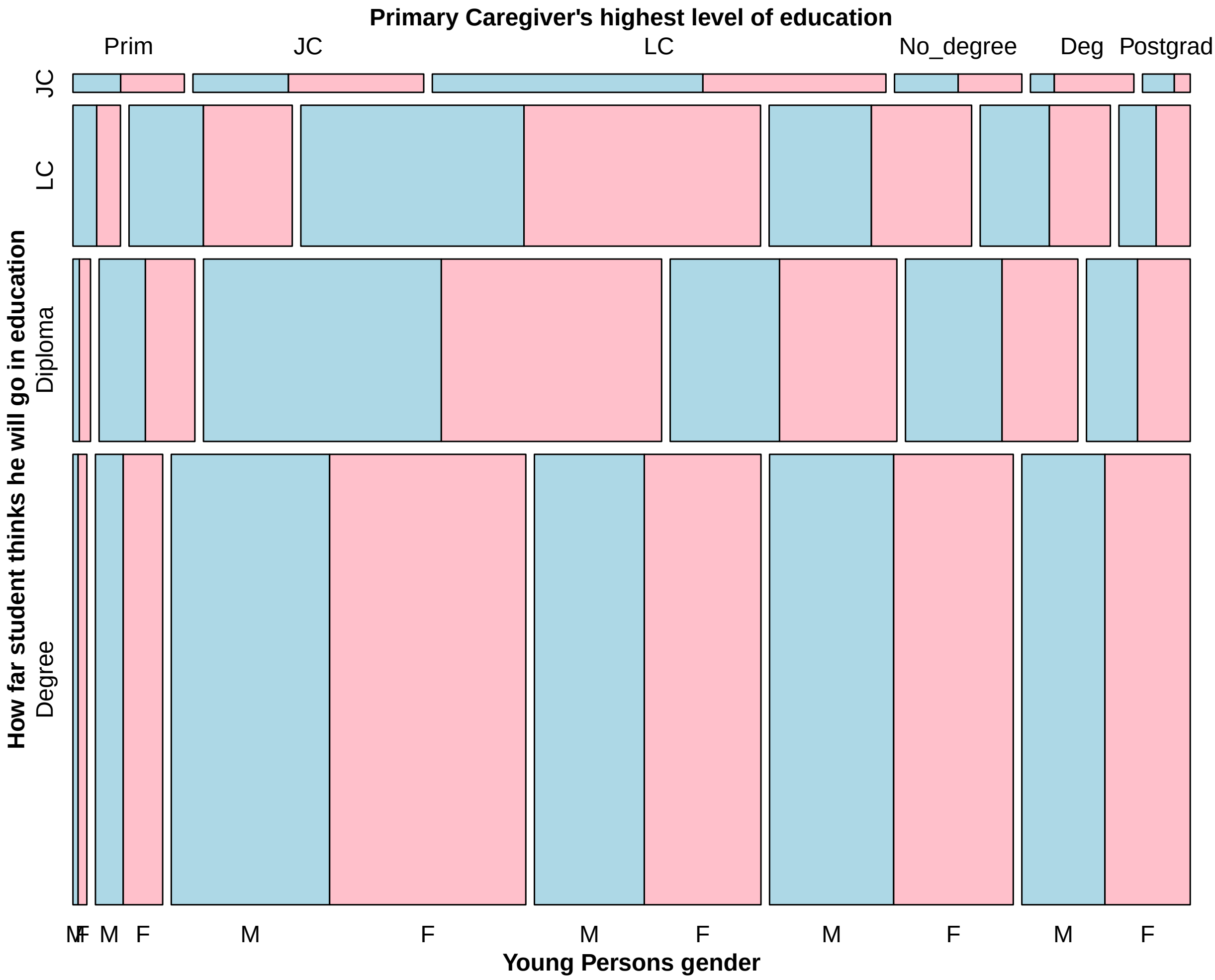


Figure 3.6: The mosaic plot showing relationship between the gender of the respondent, respondent's primary caregiver's education and respondent's plan for education in future. A higher tile area suggests a higher observed frequency for a particular category.

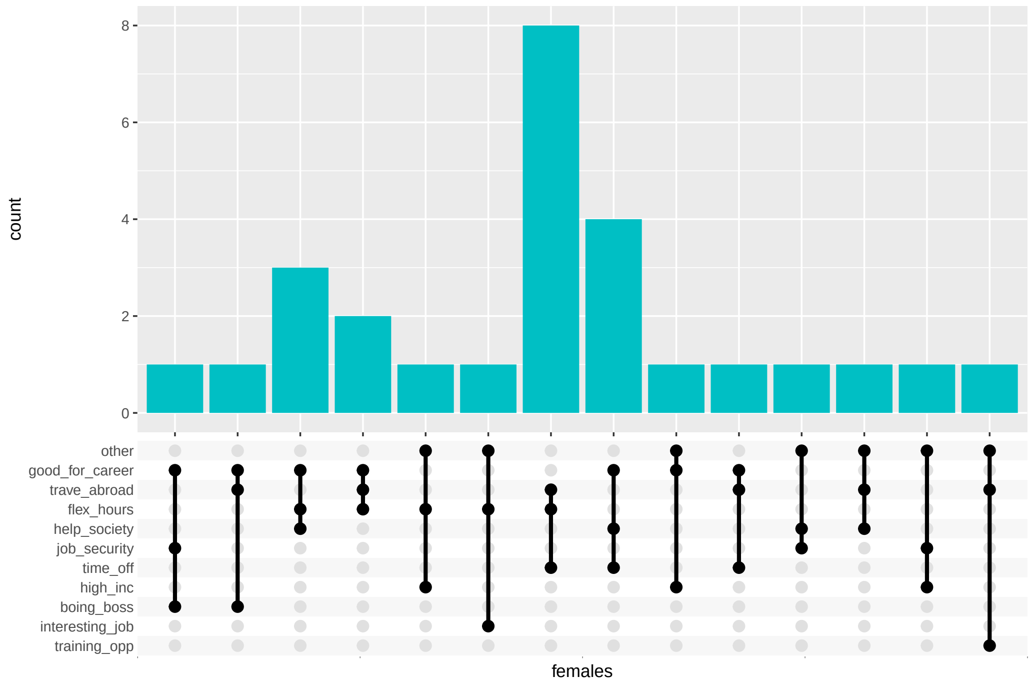


Figure 3.7: The UpSet plot showing the unique work value combinations selected by females but not males. The bar chart shows the number of respondents in the data who selected particular combination of work values. Each bar represents a different combination.

CHAPTER 4

Work Values of Irish Adolescents: An Exploratory Study

4.1 Introduction

Work values are defined as the desirable characteristics of work which one considers important in their life [Lyons et al., 2010]. In the literature, these work values or attitudes to work have been classified and defined by Schwartz [1999], Lyons et al. [2010], Marini et al. [1996], and Ros et al. [1999] broadly into 4 categories which are intrinsic, extrinsic, social or altruistic and prestige. The intrinsic work values such as an interesting job or taking a good step on the career ladder are bringers of inner peace and psychological satisfaction in an individual. The extrinsic work values are led by external reward-based aspirations of work such as getting a job or career which provides high income, job security and flexible hours. The altruistic work values pertain to maintaining good relations within the office environment or engaging in a career which contributes to uplifting society. Lastly, prestige corresponds to the values which are related to an individual's status in a society like being powerful, being your own boss or having a respectful job. Studying these work values is vital for numerous reasons. The work values reflect an individual's perception of work which has considerable psychological meaning and holds predictive value in identifying future occupational choices [Holland et al., 1990]. The work values in adolescent years are related to the career outcomes in later years [Chow et al., 2017]. They are also found to be reflective of life and career satisfaction and may lead to a higher quality of life Axelsson et al. [2005].

According to Super's [1990] developmental theory of career development, high school students are at the exploration stage of career development and many influences can stimulate career development like one's ethics, capabilities, social learning experiences and necessities [Super, 1990]. The awareness of these factors influencing work values can support students in accomplishing their educational and career ambitions and assist them in their career-related decision-making process. Various studies have been conducted in the literature across different countries and cohorts to learn about the work values of adolescents and to find indicators influencing them. Ferry [2006] looks at the factors influencing career choices of adolescents in rural Pennsylvania and reveals the cultural and social context of family and community as instrumental factors. A cross-sectional follow-up study of three cohorts of adolescents (15-16 years) in Finland [Sinisalo, 2004] found gender to be associated with work values, boys considered extrinsic values more important than girls and intrinsic values were valued more highly among girls who expressed plans to seek higher educational routes. Hirschi [2010] conducted a study on 268 Swiss adolescents to investigate the relationship of intrinsic and extrinsic work values to positive career development and found gender to be significantly related to work values with girls endorsing more intrinsic and less extrinsic work values compared to boys.

We use Growing Up in Ireland (GUI), child cohort wave 3 [Murray et al., 2010] to explore the work values of Irish teenagers. The GUI is the first national longitudinal study of children conducted to study the well-being of children in Ireland. Our study investigates the factors which influence the work values of adolescents in Ireland and evaluates if these work values differ by gender, school type and career guidance. In our research, we use visualizations to communicate the findings of the research. GUI is a survey dataset which records responses to diverse questions answered by the study child, child's caregivers, principal, and teachers. The data set was comprised of tick-box questions where respondents were asked to select one (or potentially more) of a fixed number of possible options resulting in categorical data (ordinal or nominal). For this study, the question related to career considerations of the study child was selected from the child questionnaire wave 3. The response to this question included 11 different choices. These choices were coded as different columns (11 in count) in the dataset and have been captured as binary responses (1 refers to a particular choice being picked and 0 refers to the choice being dropped). We use visualizations to build an understanding of survey data and model results. To capture interesting features and map the association between work values and potential factors which influence these work values we use seriated heatmaps. Seriation pulls the most interesting patterns in the heatmap together. Other studies use tables to present the findings from modelling, but our research communicates the results obtained from modelling the work values through meaningful visualizations.

Section 4.2.1 describes the data set, the question of interest from the GUI study and introduces the different variables which were considered for the analysis. Different steps that were carried out to manipulate and process the data are elaborated in section 4.2.2. The exploratory analysis and their findings are presented in section 4.2.3. The novel visualization techniques which assisted in answering the research questions are introduced in section 4.3. Section 4.3 also presents results from modelling. The findings along with potential future work are discussed in section 4.4.

4.2 Exploring work values in GUI data

4.2.1 GUI : Data

The data set used for this research is Growing Up in Ireland (GUI). The GUI is the National Longitudinal Study of children in Ireland. The survey is piloted to study the well-being of children in Ireland. The study was done on two age cohorts - the child and infant cohorts. For this research, the study of interest is the child cohort study. The wave 1 child cohort data surveyed 8,568 nine-year-olds living in Ireland in 2009. Since GUI is a longitudinal study, the same participants are followed over time and this information has been published in different waves. The study participants and the families of Cohort'98 have been interviewed when the study child was 9 (wave 1), 13 (wave 2) and 17/18 (wave 3) and 20 (wave 4) years of age. This chapter is based on the Growing Up in Ireland wave 3 data set from the GUI child cohort which can be accessed through the Irish Social Science Data Archive (ISSDA) [1]. This wave records data from 6,216 17-18-year-olds. The work values are obtained from the question, 'What would you consider when choosing a job'. The students were asked to pick the three most important considerations out of the 11 possibilities (see figure 4.1). A subset of wave3 data has been constructed by selecting the columns corresponding to the work values and potential variables of interest which might influence these work values to perform the analysis presented in this chapter. The variables of interest picked following the work done in the literature [Johnson, 2002], [Super, 1990] and exploratory analysis which is discussed later in this chapter.

G2. [CARD G2] Here are some factors a person might consider when choosing a job. Please choose the three most important things for you personally. **[TICK THREE ONLY]**

a. High income.. ☐1
b. A job that offered good training opportunities .... ☐2
c. An interesting job.. ☐3
d. Flexible working hours.. ☐4
e. Generous holidays/time off.. ☐5
f. A good step on the career ladder.. ☐6
g. Be your own boss... ☐7
h. A job that is useful to society or helps other people................ ☐8
i. Job security .... ☐9
j. Opportunity to travel/work abroad .... ☐10
k. Other (please specify)____________________ .... ☐11

Figure 4.1: This question is obtained from the child questionnaire wave 3. The 17-18-year-old's were asked to pick the three most important considerations when choosing a career.

Since the respondents (study children) were asked to pick the three most important considerations out of the 11 possibilities in the question of interest, the pupils picking only three choices were picked for the analysis. A subset of the dataset was obtained following literature [Johnson, 2002], [Super, 1990] to include the variables indicating primary caregiver's education, employment, family income, gender, career guidance, school type, cognitive scores, self-beliefs about work and religion.[2].

[1] Growing up In Ireland Cohort'98 (child cohort) wave 3 - 17/18 years, 2016 data set can be accessed through the Irish Social Science Data Archive(ISSDA) `https://www.ucd.ie/issda/data/guichild/guichildwave3/`

[2] The list of all the selected variables which were considered for the analysis and their short names have been listed in table 6.1

These variables were obtained from the parents, child (young person), teacher and principal's questionnaires. The data was collected by conducting interviews with the study child (young person was age 17 at the time of wave 3) and their families. The interviews were based on their respective questionnaires. The principal questionnaire was filled by the principal of every school that participated in the survey once for each wave and involved generic questions about the school.

4.2.2 GUI : Data Preparation

The originally selected subset of the GUI wave 3 data had 6216 observations and 45 variables including the 11 work value choices (see 6.1 in the Appendix). The variables picked for the analysis (see section 4.2.1) consisted of primarily factor (categorical) variables with multiple levels and some numerical variables. Numerous steps taken to prepare data for the analysis are detailed in this section. The variables with over 20% of missing values were not selected.

The subset of data used for analysis was obtained by replacing the don't know and refusal factor levels in categorical variables with NA values. The spellings of the factor levels have been rectified to eliminate redundancy. The factor levels of some categorical variables were renamed to obtain meaningful short names to make the text in the plots aesthetically pleasant. There are 11 work values used as response variables selected in the initial subset, out of which the work value 'other' has been used only for the exploratory analysis but has been omitted for the modelling as it had very low count for the choice selection (N = 99, 0.5% of the total responses). The work values in the dataset have been coded as categorical variables with levels 'Yes' and 'NA'. For each participant in the survey, the choices they selected have been coded as 'Yes', while the choices not picked were left as 'NA'. For this analysis, all the NA's have been coded as factor level 'No' as R software treats 'NA' as missing values and regression models only use complete cases. The numerical variables in the data set have been scaled to obtain values that lie between 0 and 1 to bring numerical variables on the same scale and for removing units, thus making variables comparable for easier interpretation. A correlation plot (refer figure 4.4) was constructed for examining the association between work values and the factors influencing them. The results of this plot have been used to find predictors that were correlated so that they can be dropped prior to modelling as high correlations may lead to unreliable regression estimates.

The variables which were sub-questions under the same theme have been utilized to create new derived variables. The subset used for analysis has four derived variables namely parental_interactions, guidance_score, school_type and cognitive_scores.

The derived variable parental_interactions was developed by concatenating six question-based variables that indicated the interaction of 17/18-year-olds with their parents. These variables correspond to questions like 'How often have/did your parents have discussed how you are getting on with different subjects at school/college?', 'How often have/did your parents ask how you are/were getting on with your teachers/lecturers?' into one variable. These questions had multiple choices and the young person was asked to pick one response that applies out of: 'Don't Know', 'Refusal', 'Never or hardly ever', 'A few times a year', 'About once a month', 'Several times a month', and 'Several times a week'. Since the responses to these questions are ordinal categorical, for modelling and visualizations these questions have been transformed into numerical scores. The responses recorded as 'Never or hardly ever' or 'A few times a year' have been assigned the lowest scores and the highest scores have been given to the response 'Several times a week'. Since these six question variables were informing the interactions of individuals with their parents and variables share same factor levels, these variables were then coalesced by taking the mean values of the scores assigned and a new variable named parental_interactions was created. A high parental_interactions score indicates frequent interaction between the Young Person (study child) and their parents.

Similarly, a set of questions were asked from the Young Person. For example, In thinking about what you will (would) do after you leave school: 'Have you had career talks at your school?', 'Used a specialist guidance website (such as Qualifax)?', 'Looked at university/institute of technology/college websites?'. Responses to these questions were recorded as binary with levels 0 (refers to 'No') and 1(refers to 'Yes'). The derived variable career_guidance was generated by pooling these question variables and calculating the mean value of the recorded binary responses. The higher value of the mean refers to a higher guidance score which indicates utilization of external resources and extra help to make informed decisions about their career choices.

The variable school_type was created using the number of boys and girls in a school obtained from wave 3 principal's questionnaire to indicate if the study child belongs to a single-sex or mixed school. For the variables number of boys and number of girls, the principal's questionnaire revealed a large amount of missing values, with roughly 20% missing. To prevent information loss, wave 1 and wave 2 figures for the number of boys and girls are used to derive wave 3 numbers where these values were missing.

Three variables cognitive maths, cognitive naming and cognitive vocabulary indicating an individual's cognitive development or academic performance were pooled to create the derived variable: cognitive_scores. These variables had multiple categories with '0' indicating the lowest score and larger numbers indicating higher scores. The mean of these three variables was calculated and stored in the variable cognitive_score.

The final dataset obtained after manipulating factor levels and collapsing columns to create derived variables had 27 variables selected for the analysis (16 predictors and 11 responses). The final list of variables used for modelling is displayed in table 4.1. The variables used for creating derived variables were dropped from the model.

4.2.3 GUI : Exploratory Data Analysis

The prepared data had 6216 observations and 27 variables given in table 4.1(including the 11 response variables corresponding to the work values listed in table 4.2). Exploratory data analysis was carried out on this data set to explore different work values. Bar plots were created to visualize the responses by 17-18-year old's. Figure 4.2 illustrates the frequency of boys and girls who selected a particular work value. In the GUI data set, the number of boys (N= 3024) and girls (N=3192) are equivalents. The plot reveals interesting jobs, high income, and travel opportunities to be the most popular work values in both genders. Boys are found to pick high_income and be_own_boss more frequently compared to girls. More girls picked intrinsic work values like considering a job which is interesting or is a good step on the career ladder.

The question of interest used in our study involves picking the three most impor-

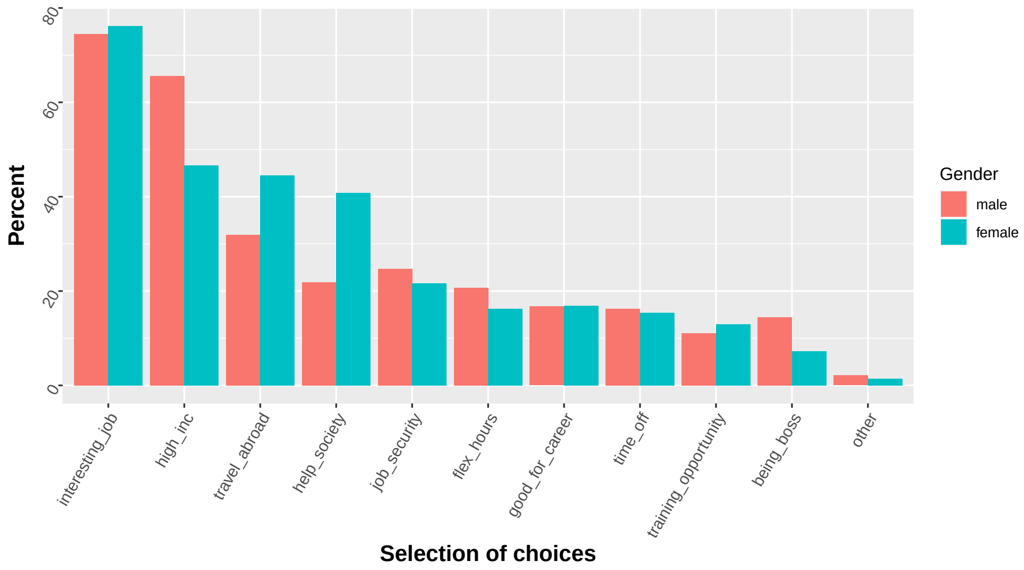


Figure 4.2: The bar plot of most popular work values selected by the two genders. Interesting job and high income are the most popular choices in the two genders.

tant work values out of the 11 options. This results in $\binom{11}{3}$ combinations to represent the top 3 choices. Upset plots [Lex et al., 2014b] have been used to visualize the three top choice combinations in both genders. They are used for representing the intersection of sets visually like Venn diagrams. These plots first decompose the sets into all possible set intersections and then allow the user to analyze these intersections individually or as aggregates. Figure 4.3 is an upset plot that shows the 10 most popular combinations of three work values in males and females. For instance, more than 300 females in the data set picked the combination of an interesting job, travelling opportunity and helping society. For boys, the combination of an interesting job, high income and the opportunity to travel is the top choice. It is notable that the top choice for boys and girls includes interesting jobs and travel, but these are chosen alongside helping society for girls and high income for boys. It is noteworthy that the top 10 combinations of choices for boys are more diverse than for girls. The work values being_boss and training_opp are less popular among girls and do not qualify in the top 10 combinations.

Figures 4.2 and 4.3 demonstrate the differences in choice selection among the two genders. These plots indicate an effect of gender in picking work values.

To perform an exploratory check for association among variables we visualize the correlation of the work values as well as the predictors selected for the analysis. In

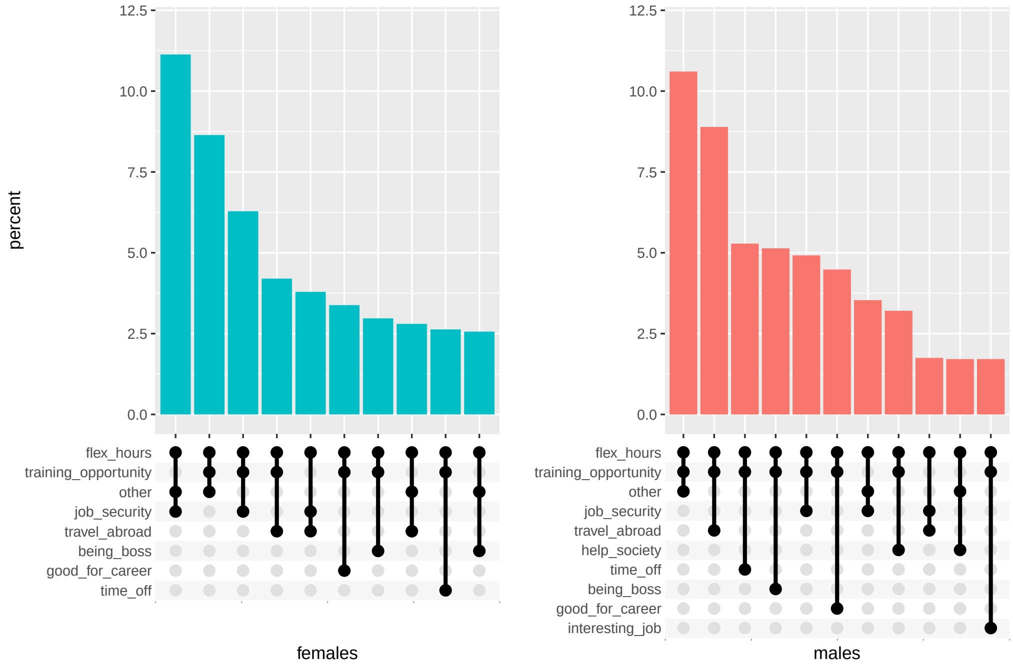


Figure 4.3: Top 10 combinations of most preferred three work values across the two genders. An interesting job, travelling opportunity and helping society is the top choice combination for girls whereas boys prefer a job which is interesting job and offers high income and travel opportunities.

Figure 4.4 the work values of 17-18-year-olds are the first 10 variables of the plot. It is evident from Figure 4.4 that most individuals who opted for transition year are still in school. The gender being female shows a positive association with career guidance and parental interactions. Figure 4.4 also demonstrates the correlation of work values on the top left. High income and helping society are uncovered as negatively associated whereas training opportunities and good for career reveal positive association.

The dark coloured cells in Figure 4.4 highlight the variable pairs which have higher magnitude correlation. It is evident that the variables left_school (indicates if the adolescents have left school) is negatively associated with TY-Y (indicates transition year) and PC1_emp (the employment if primary caregiver) and -equalised_inc (equalised family income) are found to be positively associated.

Exploratory data analysis carried out in our research indicates that the selection of work values differ by gender. The correlation plot of the variables in Figure 4.4 highlights positive association of career guidance and family involvement with the

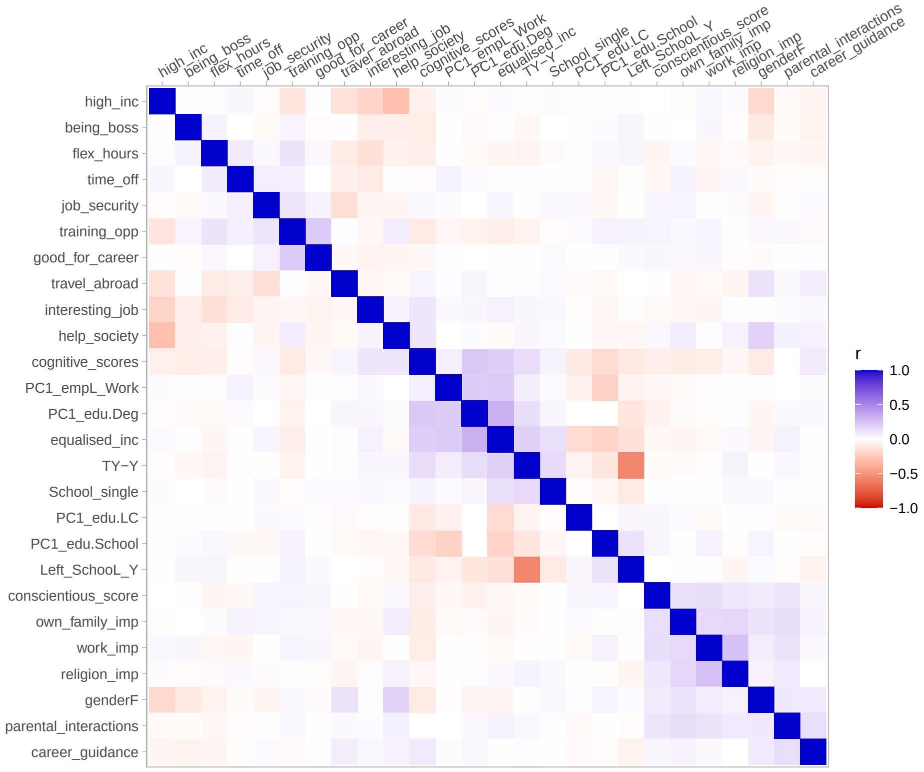


Figure 4.4: Correlation matrix plot constructed for the Work Values as well as the drivers of these work values. In this plot, blue represents positive correlation and red indicates negative correlations. Darker blue highlights strong positive correlation and darker red corresponds to strong negative correlation. We can see from the plot that cognitive scores of Irish teens are positively associated with Primary Caregiver's education and family income.

choice selection for females. The variables indicating primary caregiver's education and employment along with the household region were dropped as they were associated with the variable equalised income. Following exploratory data analysis 13 variables variables were left to perform modelling.

It will be interesting to investigate if these findings are consistent with the results obtained after modelling. To investigate the findings further and to get the predictors associated with these work values, models were fit and results obtained from modelling have been communicated using visualizations in section 4.3.

4.3 Models for work values

The dataset used for modelling had 13 predictors and 6216 cases. Logistic regression was performed on the resulting variables to model work values. Ten logistic regressions were fit, one for every work value with the same set of predictors and Type II Anova was used to obtain p-values from chi-squared tests to quantify the additional effect of predictors on the dependent variable (work values). 862 observations were deleted due to missingness when modelling work values. The variables left_school and TY (transition year) were dropped in the final model as they were not significant when controlling for other predictors (*p-value* > 0.5 for all 10 fits). Logistic regression models with two-way interactions with gender and three-way interactions with gender, career guidance and school type were also considered. The final models incorporated variables gender, cognitive scores, beliefs about work, having their own family and religion, conscientious score, equalised family income, school type, career guidance, parental interactions, and interactions of gender with conscientious score, school type and beliefs about having their own family. The final models are presented in figure 4.5 which is a seriated heatmap that includes significant main effects and significant interactions with gender. Seriation was achieved using ordering of p-values for the variables of interest. The p-values of the variables of interest were classified as positive significant, positive insignificant, negative significant and negative insignificant. Then, for every work value the p-values were arranged in descending order, by keeping the variables with most significant one's towards left. The work values with maximum significant p-values were displayed first. The work values are represented on the y-axis and the variables which influence work values are on the x-axis. The colours in the cells represent p-values from the chi-squared tests for logistic regressions. The dark blue values of the cell indicate a highly significant association where *p-value* is less than *0.01*. The cyan blue colour is used for highlighting significant associations where *p-value* is between *0.01* and *0.05*, the cells coloured grey indicate the associations which did not come out significant in presence of other predictors.

The more important associations, which had a higher number of significant p-values across all work values have been pulled towards the top left of the plot. Gender is found to have significant interactions with (a) conscientious score for work value travel_abroad and interesting_job. (b) one's beliefs about having own family for work values job_security, training_opportunity and interesting job, and (c) school type for work value job_security. The same can be seen from figure 4.5. Cognitive scores, gender, and beliefs about work and family are found to have a strong association with work values overall. The figure 4.5 indicates beliefs about the importance of work do not have a significant association with a number of extrinsic work values like travel_abroad and job_security. Cognitive scores do not show a significant association with work values job_security and time_off.

Analyzing the GUI study, we discovered limited importance of the variables school type and career guidance in influencing work values as they are only associated with fewer work values in presence of other predictors. Career_guidance is found to be highly associated with work values travel_abroad and being_boss and moderately associated with helping_society. School type has one significant interaction with gender for work value job_security. It is apparent from the model that variables relating to self-efficacy and cognitive performance are stronger influences in presence of other factors like school type and family income which affect the work value selection among Irish adolescents.

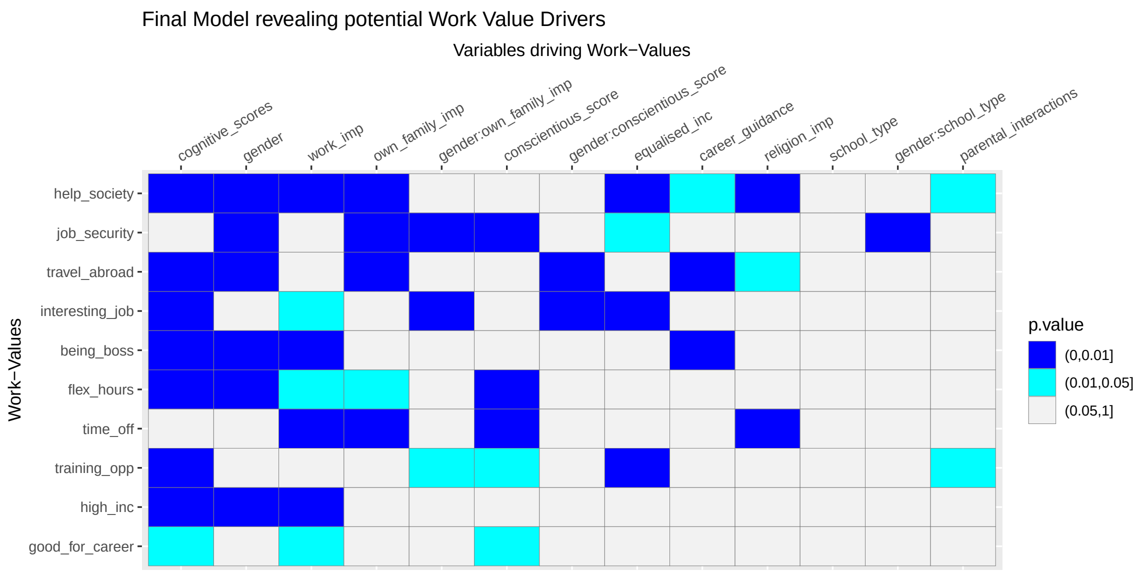


Figure 4.5: The final model considering the interactions with gender represented using a seriated heatmap plot. Seriation changes the ordering of predictors and response variables to pull significant associations to top-left.

The heatmap in figure 4.5 only visualizes if the selected variables have any association with the work values, it does not give any information about the direction of the association. The coefficient plots (refer figure 4.6) were created for the main effects for each work value using the GGally package [Schloerke et al., 2021] in R to investigate the direction of the interesting association. The main effects only include the terms or variables which are not involved in interactions with gender, that is, conscientious score, school type and importance of family have not been considered for the creation of coefficient plots.

Figure 4.6 represents different coefficient plots constructed using the values of coefficients for different main effects obtained from logistic regression. Importance of Work for 17-18-year-olds and cognitive scores have significant p-values for high_income (refer figure 4.6(f)). It also reveals that females with lower cognitive scores compared to males opted for high income [3]. Figure 4.6(b) highlights family income to be a significant predictor in picking job_security as a preferable work choice. Figure 4.6(d) reveals that individuals with good family background give more importance to intrinsic rewards like aiming for a job which is interesting compared to those with low family income. Individuals who consider religion as important, pick helping society and time off more often compared to others (refer 4.6(a)).

Since the heatmap plot in figure 4.5 reveals an interactive effect of gender and importance of having one's own family on job_security. To explore the effect of gender on job_security for variables school_type, conscientious_score and own_family_imp the interaction with gender was plotted for these variables (4.7). Figure 4.8 reveals job security to be a popular choice among males compared to females. The females who prioritise having their own family are more inclined to consider job security. Figure 4.7 uncovers that attending mixed schools for females reduces the chance of selecting job security.

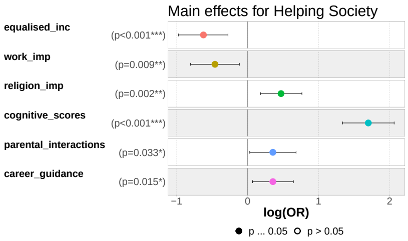
The interaction effects have been plotted using the 'effects plot' function obtained from the ggeffects package [Lüdecke, 2018] (refer figure 4.8) which were created using ggpredict() function to look at the marginal effects between different variables influencing work values. It returns the predicted values for the response of a model as effects conditioned on a factor level (usually the reference level) to visualize association. Effects plot were constructed only for the predictors which had significant

[3] Males is taken as the reference category for the factor variable gender

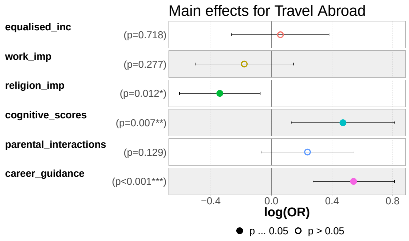
interactions with gender. Effect plots in figure 4.8 were created to look at own family importance and gender for different work values. Figure 4.8(a) and 4.8(b) highlight the work values which were selected by females more often compared to males. Females who consider having their own family as important prefer a job which helps society, while those who give less importance to having their own family prefer a job which offers opportunity to travel abroad. The figures 4.8(c), 4.8(d) and 4.8(e) represent work values preferred by boys compared to girls and the remaining plots highlight the crossover between the two genders. Plot 4.8(a) and 4.8(f) reveal that girls pick helping society and travelling abroad more that boys. Females who consider having their own family as important do not prefer a job which offers them opportunity to travel abroad . The individuals who consider family as important look for flexible hours. The remaining plots demonstrates the choices where the two genders have different preferences for a career choice given one considers having own's family as important. Boys who consider having family as important prefer interesting job whereas girls who with lower importance of having own family look for an interesting job.

4.4 Conclusion and Discussion

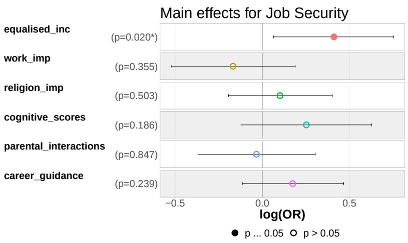
The research conducted on the GUI study reveals that cognitive scores, one's beliefs about work and family plans in addition to gender are the most important influencers of work values. The analysis reveals that Irish adolescent boys have a higher inclination toward extrinsic work values (like high-paying job, flexible hours and job security) than females. The females value helping society and travel more than males when adjusting for other predictors, whereas high income and being one's own boss are more important to males. Surprisingly, being_boss is one of the least picked work values by females. We found that the 'having one's own family' predictor has some interesting interactions with gender. Family income and background are found to be key influences in picking intrinsic work values. Job security is more important for males, but interestingly, only for females keen to have a family which is contradictory to some studies which investigated work values among adolescents [Tan, 1996]. This group of females also place less importance on having an interesting job, whereas, for males, the choice of an interesting job increases with keenness to have a family. School type has low relevance overall, aside from the tendency of females attending a mixed school to place reduced value on job security. An interesting job has been a popular choice of selection for the two genders. It also highlights that these choices are different across the two genders as females place greater importance than males on intrinsic and altruistic career choices and that males place greater importance than females on extrinsic work-related rewards like high income and job security which is in line with the literature Marini et al. [1996], [Bridges, 1989], [Herzog, 1982], [Johnson, 2002]. Career guidance and parental interactions seem to have a higher positive influence on females than males. However, school and career guidance seem to have limited importance at this stage for individuals. The paper [Johnson, 2002] talks about the region (urban/rural) to have an influence on picking work values. It describes that the families of individuals living in rural communities are more inclined toward intrinsic rewards rather than extrinsic, but we were unable to find any strong associations between the region and career choice selection in presence of other important predictors which are present in the final model. It would be interesting to see if these work value preferences remain consistent in the next wave of GUI child cohort study.



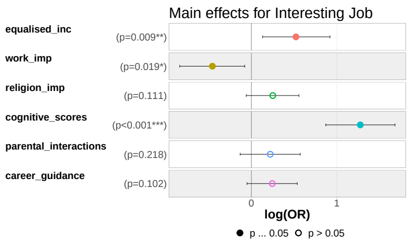
(a) Helping Society



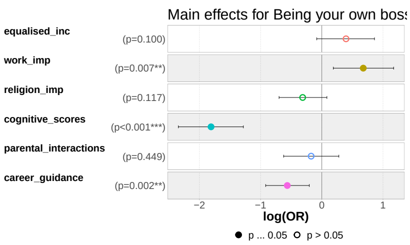
(b) Travel Abroad



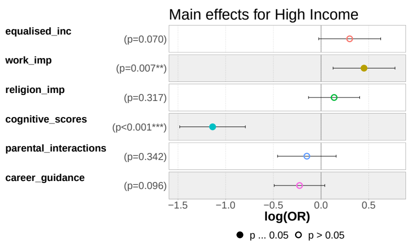
(c) Job Security



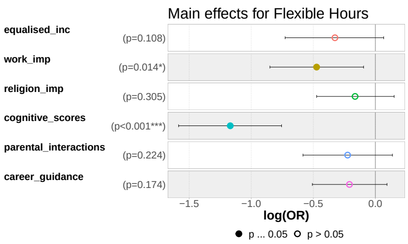
(d) Interesting Job



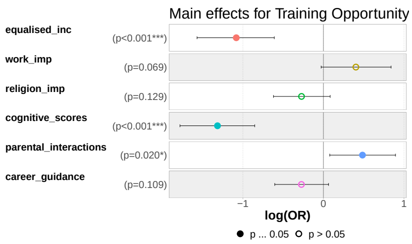
(e) Being Own Boss



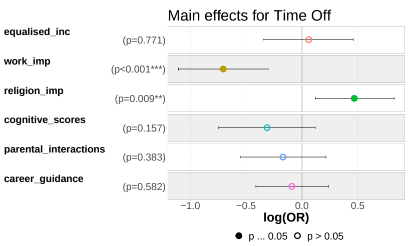
(f) High Income



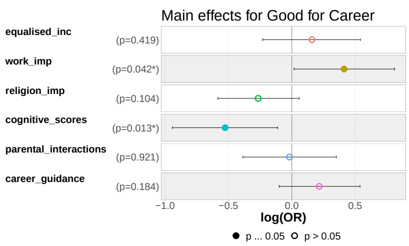
(g) Flexible Hours



(h) Training Opportunities

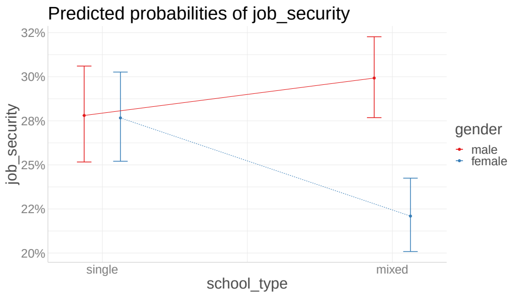


(i) Time Off

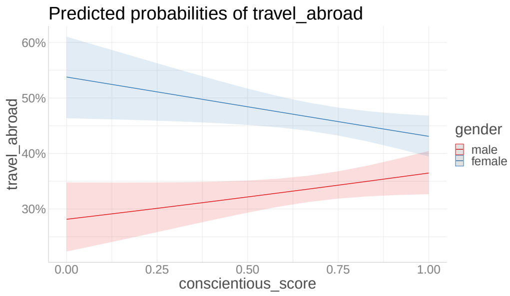


(j) Good For Career

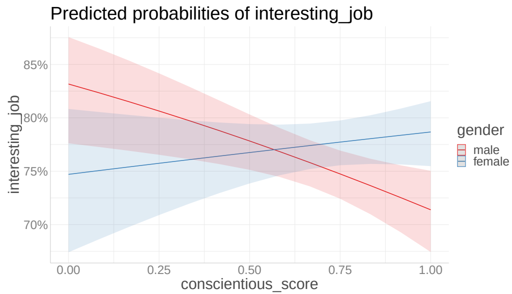
Figure 4.6: Coefficient plots for different work values constructed using the values of coefficients for different main effects obtained from logistic regression. Figure 4.6(b) highlights family income to be a significant predictor in picking job security as a preferable work choice.



(a) Job Security

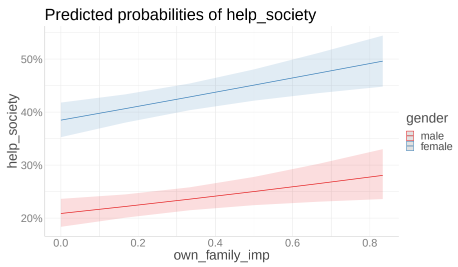


(b) Travel Abroad

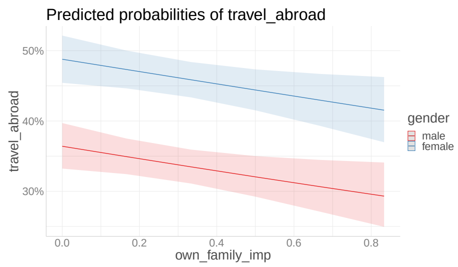


(c) Interesting Job

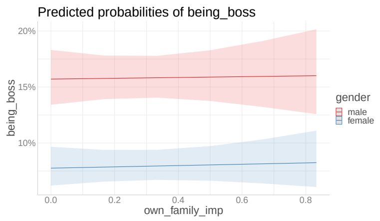
Figure 4.7: Effects plots for different work values with significant interaction terms. Plot a reveals that males pick job security more often compared to females and attending mixed school for females reduces the chance of selecting job security. Plot b highlights that females with higher conscientious score pick travelling abroad less frequently. Plot c demonstrates that boys having lower conscientious score prefer interesting job more that girls.



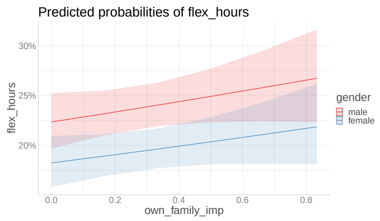
(a) Helping Society



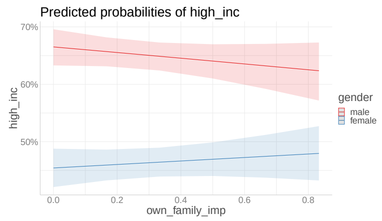
(b) Travel Abroad



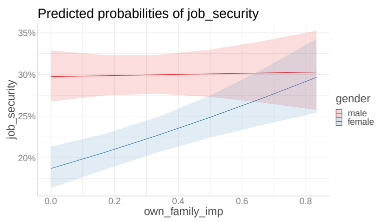
(c) Being Own Boss



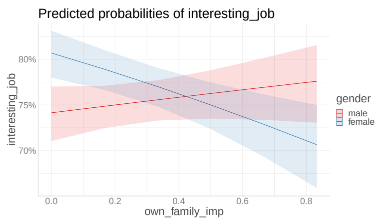
(d) Flexible Hours



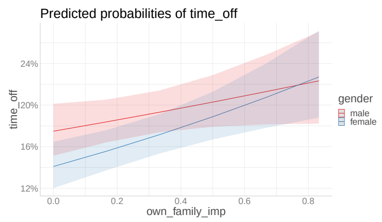
(e) High Income



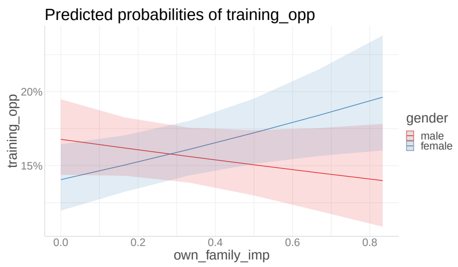
(f) Job Security



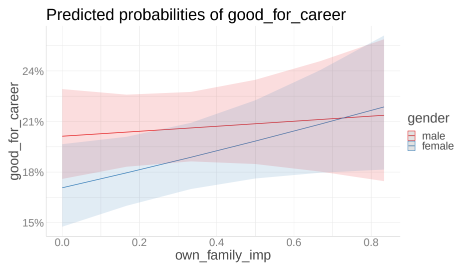
(g) Interesting Job



(h) Time Off



(i) Training Opportunities



(j) Good For Career

Figure 4.8: Effect plots of own family importance and gender for different work values. The first row (plot a and b) reveal that girls pick helping society and travelling abroad more that boys. Females who consider having their own family as important do not prefer a job which offers them opportunity to travel abroad . The second row illustrates the career choices which were preferred by males over females. The individuals who consider family as important look for flexible hours. The third and fourth row demonstrates the choices where the two genders have different preferences for a career choice given one considers having own's family as important.

Questionnaire	Shortname (used for analysis)	Variable Name (GUI dataset question code)	Description/Question	Type	Category
Primary Caregiver	equalised_inc	w3equivinc	Equivalised Household Annual Income	numerical	predictor
Principal	school_type	p3q4a,p3q4b	Derived Variable- to indicate school is mixed or single sex	categorical - 2 levels	predictor
Young Person	gender	p2sexW3	Gender	categorical - 2 levels	predictor
	work_imp	cq3g4f	How important - Profession and Work	categorical- 7 levels	predictor
	religion_imp	cq3g4j	How important - One's own family and children	categorical- 7 levels	predictor
	own_family_imp	cq3g4h	How important - Religion	categorical- 7 levels	predictor
	cognitive_scores	CognitiveMathsTotal,CognitiveNamingTotal,CognitiveVocabularyTotal	Derived Variable-Cognitive Scores	numerical	predictor
	conscientious_score	w3cq_conscientious	Conscientious Subscale wave 3	numerica	predictor
	parental_interactions	cq3e1a, cq3e1b, cq3e1c,cq3e1d, cq3e1e,cq3e1f	Derived Variable- How often did your parents interact with you	numerical	predictor
	career_guidance	cq3c2a, cq3c2b, cq3c2c,cq3c2d, cq3c2e	Derived Variable- In thinking about career did you attend talks and take guidance	numerical	predictor
Identifiers	ID	ID	Unique Identifier-Household ID	numerical	predictor

Table 4.1: Final list of predictor variables obtained after cleaning and preliminary screening

Table 4.2: Data Dictionary of responses in GUI analysis

	Short_name	Variable Name	Description
1	high_inc	cq3g2a	High Income
2	training_opp	cq3g2b	A job that offered good training opportunities
3	interesting_job	cq3g2c	An interesting job
4	flex_hours	cq3g2d	Flexible working hours
5	time_off	cq3g2e	Generous holidays/time off
6	good_for_career	cq3g2f	A good step on career ladder
7	being_boss	cq3g2g	Be your own boss
8	help_society	cq3g2h	A job that is useful to society or helps other people
9	job_security	cq3g2i	job security
10	travel_abroad	cq3g2j	Opportunity to travel/work abroad
11	other	cq3g2k	Other

CHAPTER 5

Work Values Follow-up 20 year olds

5.1 Introduction

Adolescence and early adulthood are the most formative years in terms of professional dedication, as this is when major life choices are made [Super et al., 1963]. However, it is considered that adolescent career intentions are fragile and alter periodically until adulthood [Super, 1980]. This research is a follow-up of research conducted on the work values of Irish adolescents between the ages of 17 and 18 using GUI wave 3 dataset and intends to determine if the selection of preferred work values at the age of 20 using GUI wave 4 data remains consistent with the work values selected in adolescent years.

The research presented in the previous chapter demonstrated that cognitive scores, perceptions about work and family plans, as well as gender, are the most influential factors in determining work values. It also revealed the tendency of boys aged 17 to 18 to pick more extrinsic work values than girls. The importance of job security was found to be greater for men, but, curiously, only for women who consider having their own family as important.

This chapter examines the work values of the same individuals at the age of

20 and examines whether or not these values evolve through time. We examine the relationship between characteristics related with the individuals' work values when they were 17-18 years old and the job values they chose at the age of 20. We investigate whether adult work values vary by gender.

Section 5.2 describes the data set obtained from GUI wave 4 and the question of interest from the GUI study. The steps undertaken to prepare the dataset for the analysis has also been discussed in this section. The exploratory analysis and their findings have been presented in section 5.3. Section 5.4 presents results from modelling and also discusses limitations of the methodology for comparison of work values in wave 3 and wave 4. The findings along with potential future work has been discussed in section 5.5.

5.2 GUI : Data

This research is based on the Growing Up in Ireland wave 4 data set from the GUI Child Cohort which can be accessed through the Irish Social Science Data Archive (ISSDA)[1]. This wave records data from 5,190 20-year-olds. The work values are obtained from the question, 'What would you consider when choosing a job' listed in the young person's questionnaire. The young adults ranked the relevance of thirteen job-related characteristics such as 'high income', 'job security' and 'helpful to society'. They were asked to rank these characteristics from 0 (not at all important) to 10 (very important). The snapshot of the question is presented in figure 5.1. Compared to the question asked concerning career considerations in GUI Wave 3, two additional work values have been added to the list of choices in wave 4 when selecting a job: creative work and promotion opportunities.

In wave 4 of GUI data, there were 5,190 observations. There were seven rows with the value '99' for all choices. Since the young person was asked to rank the career considerations on a scale from 0 to 10, with 10 representing the most important option, these 7 observations were deleted prior to data analysis. The wave 4 data set was merged using inner join with the final subset of wave 3 data (refer section

[1]Growing up In Ireland Cohort '98 (Child Cohort) Wave 4 - 20 years, 2019 data set can be requested through the Irish Social Science Data Archive(ISSDA) `https://www.ucd.ie/issda/data/guichild/guichildwave4/`

J5. [CARD J5] Here are some factors a person might consider when choosing a job. On a scale of 0 to 10 how important would each of these be to you in choosing a job?

0= Not at all important......10= Very important

	Not at all important										Very impo rtant
a. High income	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
b. A job that offered good training opportunities	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
c. A job that offered good promotion opportunities	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
d. An interesting job	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
e. Flexible working hours	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
f. Generous holidays/time off	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
g. A good step on the career ladder	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
h. Be your own boss	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
i. A job that allows you to be creative	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
j. A job that is useful to society or helps other people	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
k. Job security	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
l. Opportunity to travel/work abroad	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10
m.Other (please specify)	☐0	☐1	☐2	☐3	☐4	☐5	☐6	☐7	☐8	☐9	☐10

Figure 5.1: This question is obtained from the young person's questionnaire wave 4. The 20-year-old's were asked to rate the important considerations when choosing a career on a scale of 0 to 10.

4.2.2) using the unique identifier variable: 'Household ID'. Since there were some families in Wave 3 which didnot participated in the GUI Wave 4 survey, thus the resulting subset for analysis contained 4811 observations and 13 predictors. Using the same set of predictors from wave 3, the association between these variables and work values for 20-year-olds have been investigated. The same set of predictors is used to investigate if the association of these predictors with work values still persists when the 17/18 year olds turn 20. In section 5.3, we will explore the choices made by young adults at age 20 using some exploratory plots.

5.3 Exploratory Analysis

An exploratory data analysis was conducted to investigate various work value choices presented to the Irish adults. Since individuals were asked to rank these characteristics from 0 (not at all important) to 10 (very important) we use figure 5.2 and 5.3 to illustrate the scores given by young adults to each work value. There was only one observation which had 'NA' for all work values, thus it has been eliminated from the data set. The x-axis of the plots represent the different career choice considerations and the y-axis highlights the importance score. In Figure 5.2 majority of respondents ranked these values as being of average or higher importance. Work value being_boss has been accorded with a relatively low degree of importance compared to other work values. Interesting job followed by job security have been rated as the most important work values. These visualisations are useful for determining the ranking trend of work values.

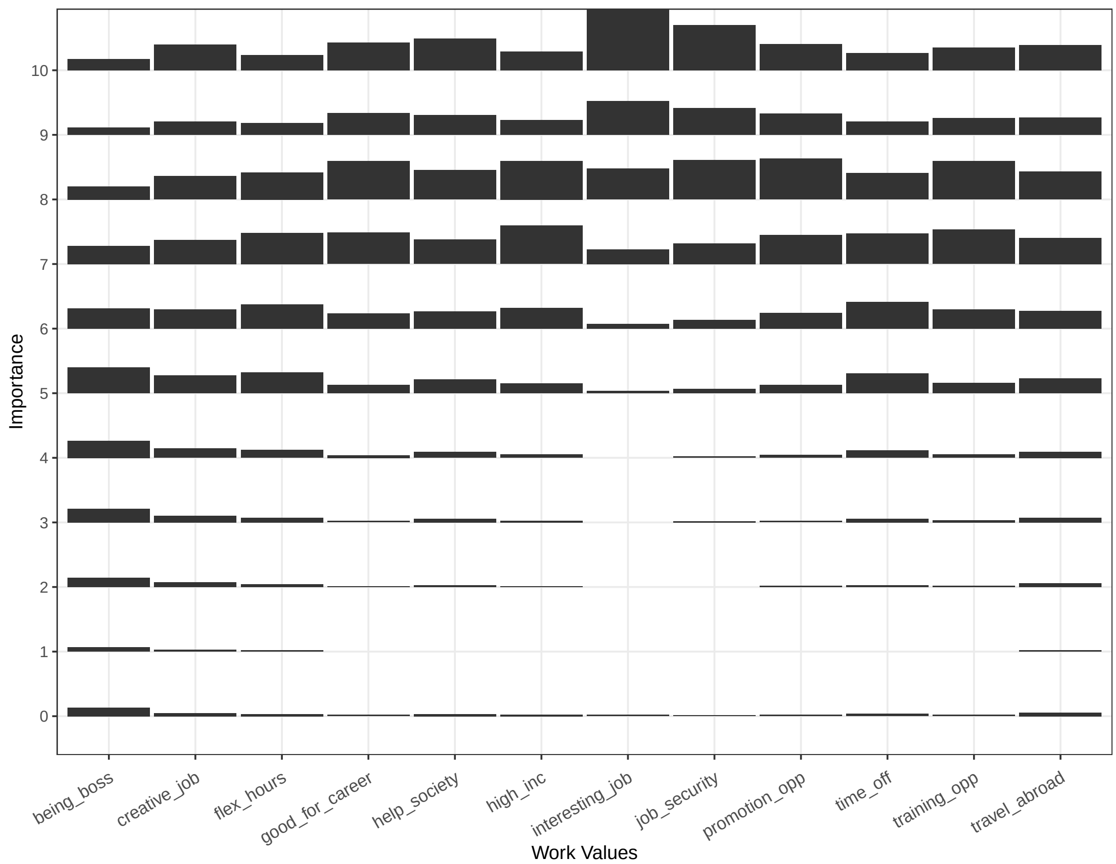


Figure 5.2: The figure highlights lowest scores to be given to work value being_boss followed by creative job and travelling abroad whereas highest scores being assigned to interesting job. The height of the bar represents the occurence of score for a particular work value

To compare the findings obtained from exploring work values in wave 3 with wave 4, similar plots were produced. Figure 5.4 depicts the frequency with which boys and girls selected a particular work value. In the GUI wave 4 data set, there are a comparable number of males (N=2498) and females (N=2692). The plot 5.4 indicates that interesting jobs, job security, and promotion opportunities are the top three work value considerations for both men and women. Boys opt for high pay

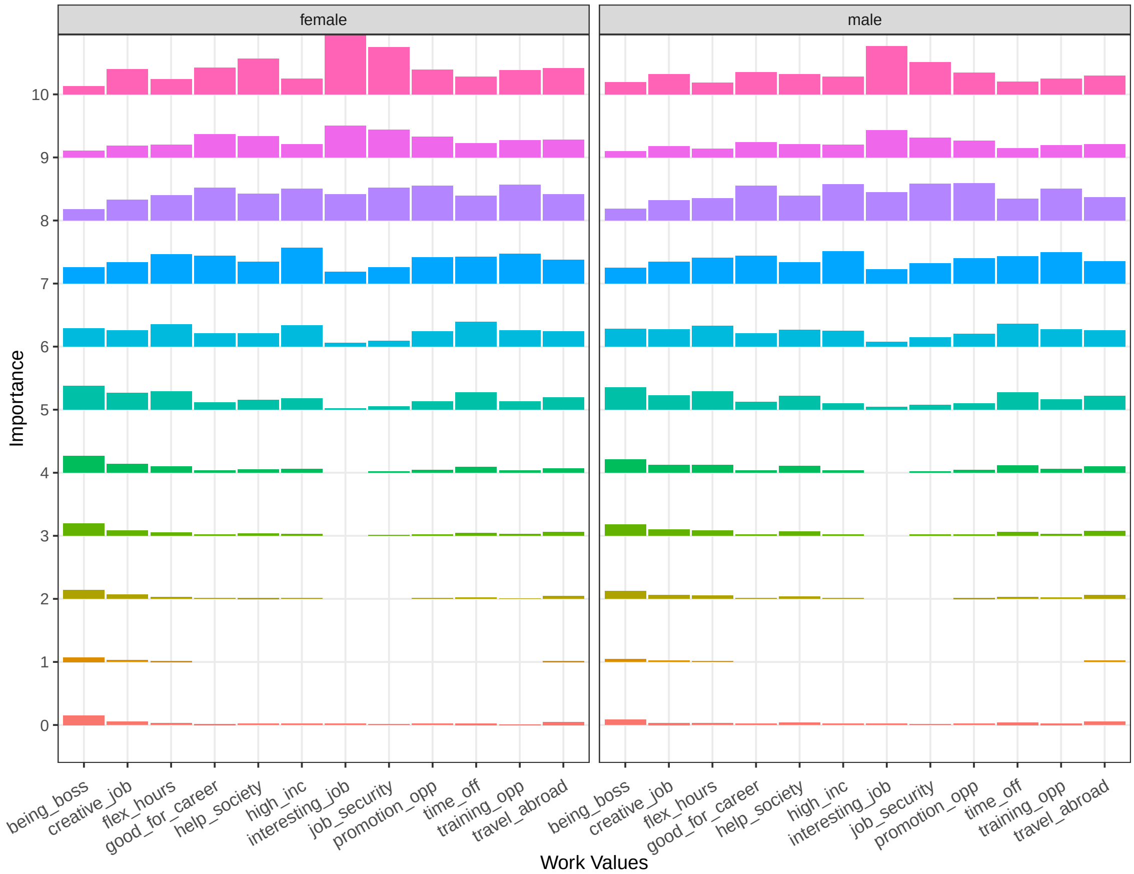


Figure 5.3: The figure illustrates the scores given by young adults to each work value and facets them by gender. The plot reveals that females have given more number of highest scores (score 10) to different work values compared to males. The lowest score awarded to an interesting job, according to both male and female responses, was rate 5.

and being their own boss more frequently than girls. When individuals are 17-18 years old and 20 years old, an interesting job is consistently the most popular option, however high income is no longer a popular choice among individuals. Compared to boys, more girls choose intrinsic work values, such as selecting a job that is interesting or a good step on the career ladder or choosing a creative employment. Being your own boss has been the least popular choice for the individuals at both 17/18 and 20 years. It is interesting to notice that Irish males chose job security more frequently than females between the ages of 17 and 18, but at age 20, girls outnumber boys as they ranked job_security higher compared to males.

Figure 5.5 is a dumble plot that shows the work value choices which where different in males and females. The solid red circle in the plot represents the frequency of male selections for a certain job value, whereas the blue circle represents the total

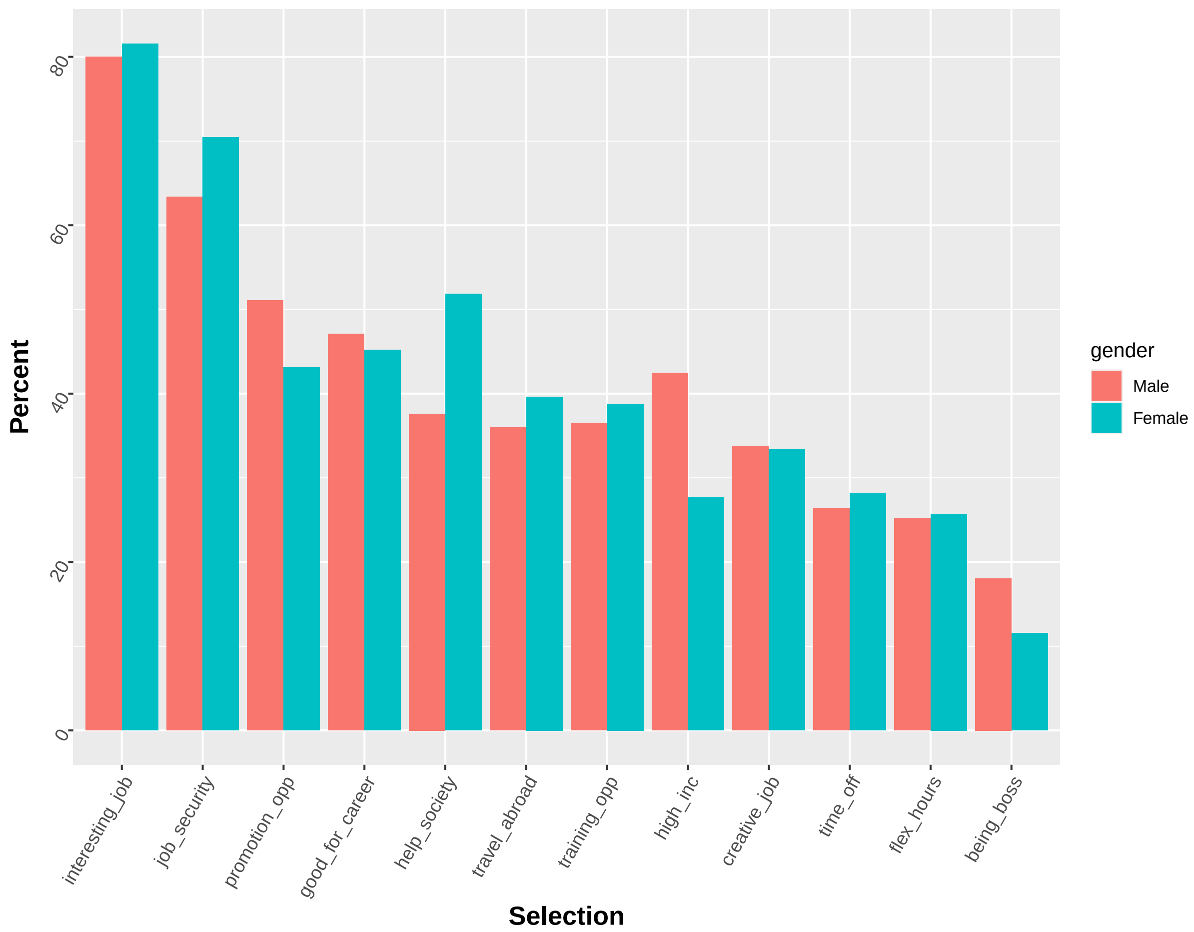


Figure 5.4: The bar plot of most popular work values selected by the two genders. Interesting job and job security are the highly scored choices in the two genders.

number of female selections.. For instance, women are more likely than men to select the work value, helping society. In addition, job security and interesting job appear to be more appealing to women than to men. However, men are more likely than women to choose reward based job ideals such as a high wage and promotion opportunities. Work values such as creative work and flexible hours are desired equally by men and women.

To conduct an exploratory test for association between variables, we display the correlation of work values using a heatmap. Figure 5.6 illustrates the association between choice selection of work values when the individuals attain age 20. The dark blue colour indicates the strength of association (Color intensity and the size of the circle are proportional to the correlation coefficients). The correlation plot is important for determining which work values were selected in tandem. Figure 5.6 demonstrates that picking the work value good_for_career has strong positive

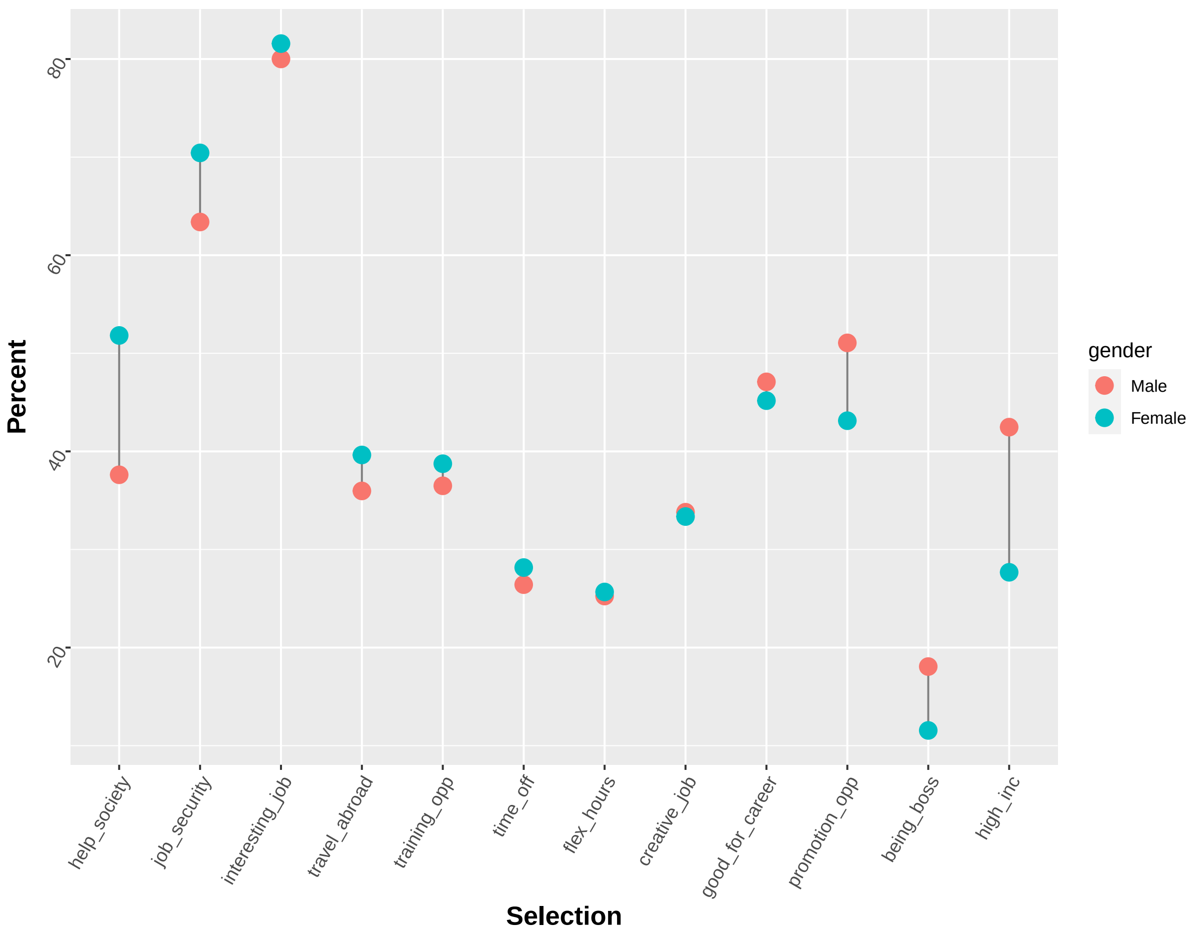


Figure 5.5: The dumble plot emphasises the contrasting choices made by men and women. Compared to women, men value a high income, promotion prospects, and being in charge of their job. Females are highly represented in intrinsic work values such as helping society and interesting job. The work values interesting job and helping society are picked more often by females consistently in both waves, however, in wave 4 more girls compared to boys ranked job security higher.

association with promotion and training opportunities. There is a strong correlation between high income and promotion opportunities. Flexible work hours and time off are also indicative of a positive association. The findings from exploratory analysis indicate the diverse selection of choices in the two genders. Section sec: model will discuss the results from modelling.

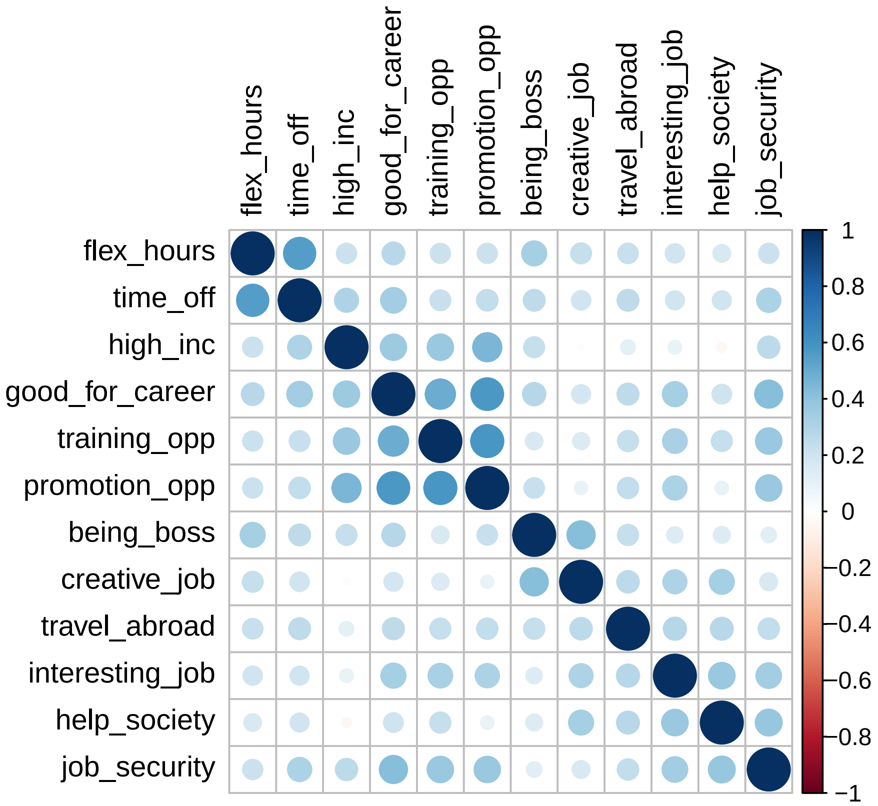


Figure 5.6: Correlation matrix plot is constructed for the work values when the individuals attain age 20. In this plot, blue colour indicates positive correlation and red indicates negative correlations. Darker blue highlights strong positive correlation and darker red corresponds to strong negative correlation. We can see from the plot that good_for_career has strong positive association with promotion and training opportunities

5.4 Models for Work Values

The modelling of work values in GUI wave 4 utilised a data set with 13 predictors and logistic regression was used, similar to the methodology for wave 3. In the subset, thirteen work values were utilised as response variables; however, the work value 'other' was removed from the modelling due to its extremely low selection count (N = 99, 0.5% of the total responses). In GUI wave 3, adolescents were asked to select the three most significant characteristics of a profession, but in wave 4, adults were asked to rank these options from 0 (least important) to 10 (most important). Due to the different underlying structure of the two questions, the responses of adults in wave 4 were filtered to achieve a comparable subset. The scores that adults assigned to these work values were arranged in ascending order, and as there are 12 options for career considerations, scores at position 10 and above were selected (thus picking the top 3 ranks given by the young person) to make data comparable to wave 3. Thus, the subset obtained was comprised of the top three choices selected by individuals; nevertheless, if an adult ranked all options equally, the choices corresponding to ranks at position 10, 11, and 12 from the sorted list will be selected. After getting a comparable subset, a seriated heat map was generated using the same methods as wave 3. The variables left_school and PC1_emp were dropped in the final model as they were not significant when controlling for other predictors (*p-value* > 0.5 for all 12 fits). However, the variables TY and PC1_edu were significant and were therefore retained in the final model. Twelve logistic regressions were fitted, one for every work value with the identical set of predictors. ANOVA was used to get p-values from chi-squared tests to quantify the additional influence of predictors on the dependent variable and these p-values are represeted using a heatmap (see figure 5.7. Logistic regression models with two-way interactions with gender were also considered. The final model incorporated variables gender, cognitive scores, beliefs about work, having their own family and religion, conscientious score, equalised family income, school type, career guidance, parental interactions, transition year, primary caregiver's education and interactions of gender with transition year and beliefs about importance of work. The final model is presented in Figure 5.7 which includes significant main effects and significant interactions with gender. The work values are represented on the y-axis and the variables which influence work values are on the x-axis. The colours in the cells represent p-values from the chi-squared tests for logistic regressions. The dark blue values of the cell indicate a highly significant association where *p-value* is less than *0.01*. The cyan blue colour is used for highlighting significant associations where *p-value* is between *0.01* and *0.05*, the cells coloured grey indicate the associations which did not come out significant in presence of other predictors. The more important associations, which had a higher number of significant p-values across all work values have been pulled towards the top left of the plot.

Figure 5.7 reveals cognitive scores and gender to be strongly associated with most work values. Controlling for other predictors, there is an association between conscientiousness and work values related to career advancement such as picking

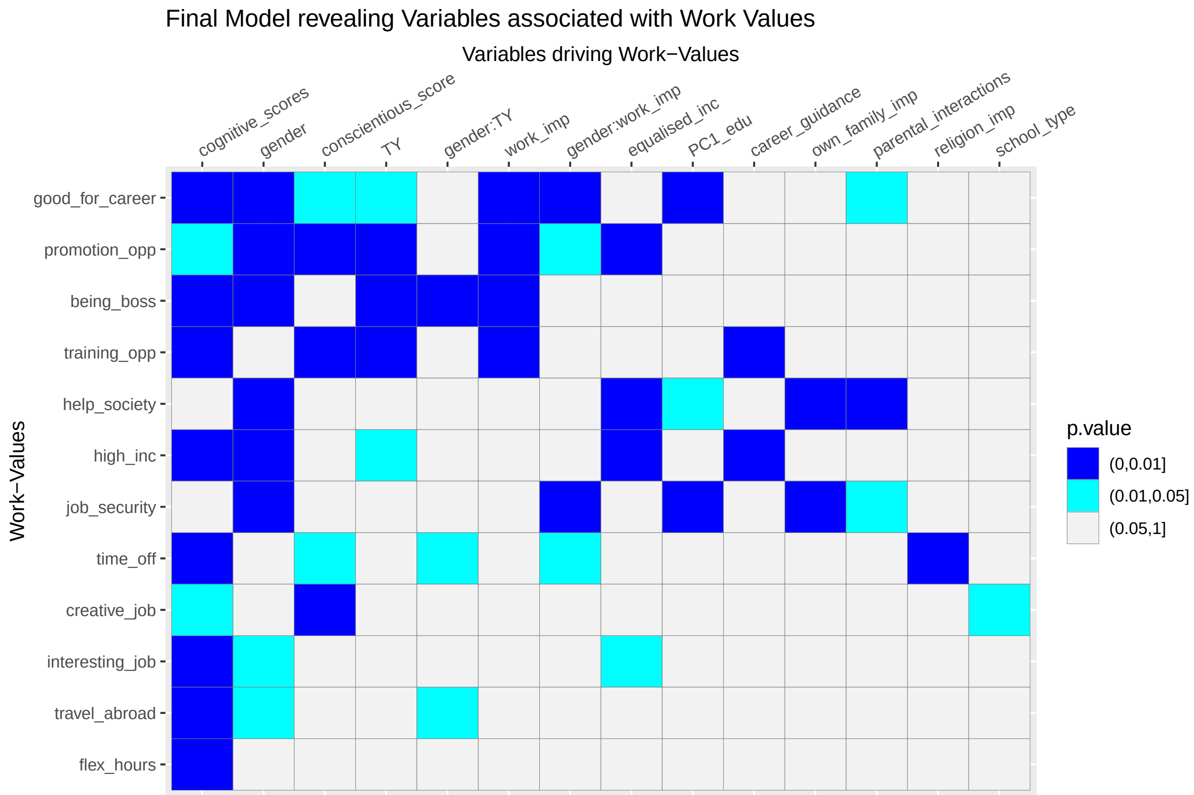
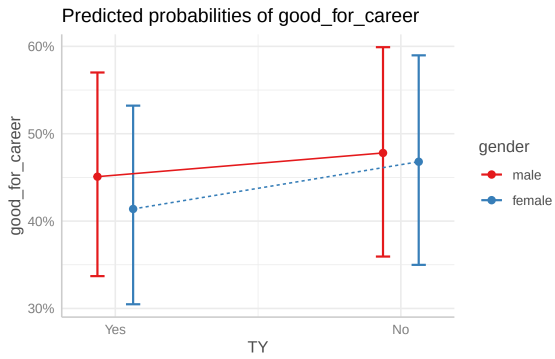


Figure 5.7: The final model considering the main effects and significant interactions with gender represented using a seriated heatmap plot. Seriation changes the ordering of predictors and response variables to pull significant associations to top-left.

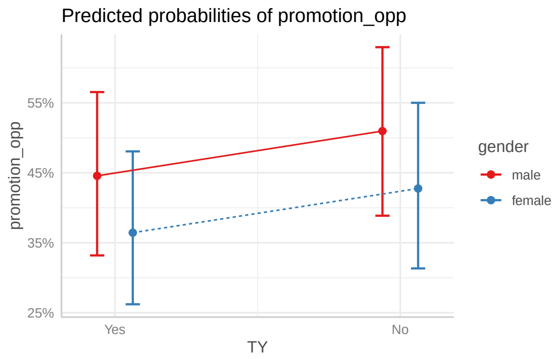
a job which is good for career, offers opportunities to be creative, provides good training and promotion opportunities. The variable TY (indicator of whether the participant attended transition year) was omitted from the final model in wave 3 since it did not indicate any significant relationships; however, in wave 4, it was revealed to be a significant factor in selecting work values related to career growth. The effects plot were created to investigate the effect of gender on work values which had significant association with 'TY'. Figure 5.8 reveals that males who donot attend TY score work values related to career growth higher compared to females. However, for both genders these work values are preferred by those who didnot attend TY.

5.5 Results and Discussions

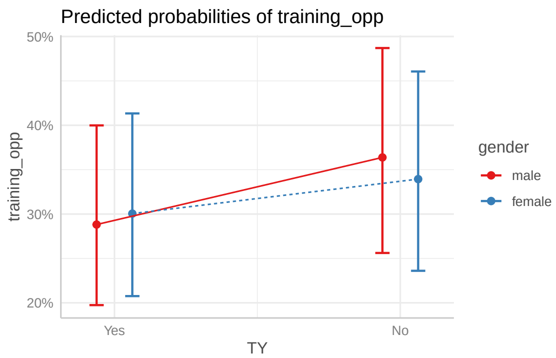
The work value 'an interesting job' was ranked as most important work value by over 60% adults in wave 4 which is consistent with the results obtained from wave 3.



(a) Good_for_career



(b) Promotion_opp



(c) Training_opp

Figure 5.8: Effects plots to investigate the interaction effect of gender and TY for work values related to career growth. The plots (a), (b), and (c) reveal that males who did not attend TY prefer work values: good_for_career, promotion_opp and training_opp

The second-highest response was 'job security' which may reflect the experiences of the group growing up during a period of economic turmoil. As individuals migrate from adolescence to adulthood, a shift in the selection of certain values is noticed between waves 3 and 4. In the third wave, a high wage was viewed as the most picked work value after interesting job; but, in the fourth wave, job stability became more important. School type appears to have limited importance in both waves. The work value: high income was not even in the top 5 picked choices. To determine the direction of association between potential factors and work values, it is possible to examine the significant main effects and interactions with gender. As it did not show any significant relationships, the variable TY (which indicates if the participant attended a transition year) was left out of the final model in wave 3. However, in wave 4, it was found to be a significant factor in choosing work values connected to career advancement. The purpose of the effects plot was to examine how gender affected work values that were strongly associated with TY. Males who did not attend TY, score work values connected to professional progression higher than females do. Yet, individuals who did not attend TY appreciated these labour values for both sexes. Adults' current activity status (whether employed or in higher education) can be examined to see whether they place a high value on 'promotion' and 'career ladder' jobs. But due to insufficient time, this analysis was not conducted.

CHAPTER 6

Conclusions

6.1 Results

According to studies conducted on Irish teens (wave 3) and adults (wave 4), an interesting job is the most desirable work value choice, while being your own boss is the least desired. In both waves, girls preferred helping society more than males. It is interesting to note that between the ages of 17 and 18, Irish males choose job stability more frequently than females, but at age 20, more females selected this work value. The modelling suggests that cognitive scores, work-related beliefs, and gender are the most influential predictors of work values across both waves. In wave 4, it was discovered that transition year has a significant interaction with gender and is highly associated with work values like promotion prospects, training opportunities, and a job that is good for career.

Literature-based research indicates religiousness to be a strong predictor of the work values of adolescents. Those who placed a stronger emphasis on religion valued security and influence, as well as intrinsic, altruistic, and social rewards, more than those who did not [Johnson, 2002]. We discovered a positive relationship between religion and the selection of altruistic/intrinsic job value: helping society among

Irish adolescents (17/18 years old). However, no association was identified between religion and intrinsic or altruistic work values among wave 4 participants (aged 20). As Irish adolescents transitioned from high school to college/work, the relevance of certain work-related values, such as having an engaging career, travelling overseas, and contributing to society, was consistent for both waves.

However, certain work ideals, such as high pay and job stability, witnessed a shift in preferences over time. Initially, a high income was a very popular work value for Irish adolescents between the ages of 17 and 18, but by the time they reached the age of 20, it became less important. The importance of job security increased for Irish youths over time and became the second most popular choice after interesting job.

Some variables, such as the education of primary caregivers (generally mothers) and participation in the transition year, were not shown to have strong relationships with the work value decisions of Irish teenagers after controlling for other variables. Later on, however, these variables indicate substantial relationships with occupational values.

6.2 Discussion and Future Work

A comprehensive understanding of the changes in work values among young adults necessitates a diversity of analytic methodologies and the analysis of several dynamics. Our research takes the first step in exploring these work values for Irish youth at 17/18 years and 20 years. This study outlines changes in work values that occur during the transition to adulthood, as well as explores the association of these work values with various factors like cognitive development and socio-economic background. However, the analysis in chapter 5 is not optimal, as we have longitudinal data (paired) and it would be better to fit models that exploit this structure. It will be interesting to determine whether those in employment are more likely than those in education/training to place a high value on work values such as promotion opportunities, having a job which is a good step on the career ladder' and being your own boss. The work values associated with job and career satisfaction could be explored in future. The next wave of data collection, scheduled to occur at age

25, will provide insights into the employment integration of Young Adults who were still in education at the time of wave 4. It would also provide an opportunity to compare the work value choices of the adults who entered the labour market at the age 20.

Bibliography

Angela Chow, Nancy L Galambos, and Harvey J Krahn. Work values during the transition to adulthood and mid-life satisfaction: Cascading effects across 25 years. *International Journal of Behavioral Development*, 41(1):105–114, 2017.

Lars Axelsson, Ingemar Andersson, Anders Håkansson, and Göran Ejlertsson. Work ethics and general work attitudes in adolescents are related to quality of life, sense of coherence and subjective health–a swedish questionnaire study. *BMC public health*, 5(1):1–10, 2005.

Monica Kirkpatrick Johnson and Jeylan T Mortimer. Origins and outcomes of judgments about work. *Social Forces*, 89(4):1239–1260, 2011.

Adrián Zancajo. Education at a glance 2010: Oecd indicators. *Revista Española de Educación Comparada*, (18):374, 2011.

Gerhard Krug and Andreas Eberl. What explains the negative effect of unemployment on health? an analysis accounting for reverse causality. *Research in Social Stratification and Mobility*, 55:25–39, 2018.

Nadya A Fouad. Work and vocational psychology: Theory, research, and applications. *Annu. Rev. Psychol.*, 58:543–564, 2007.

Daráine Murphy, Amanda Quail, James Williams, Sophie Gallagher, Aisling Murray, Eoin McNamara, and Desmond O'Mahony. A summary guide to wave 3 of growing up in ireland's child cohort (at 17/18 years). *Dublin: Economic and Social Research Institute*, 2018.

Irish social science data archive. URL `https://www.ucd.ie/issda`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL `https://www.R-project.org/`.

JJ Allaire, Jeffrey Horner, Yihui Xie, Vicent Marti, and Natacha Porte. *markdown: Render Markdown with the C Library 'Sundown'*, 2019. URL `https://CRAN.R-project.org/package=markdown`. R package version 1.1.

Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi:10.21105/joss.01686.

Eoin McNamara, Aisling Murray, and James Williams. Growing up in ireland. *Longitudinal and Life Course Studies*, 10, 2019. ISSN 17579597. doi:10.1332/175795919X15699445254797.

Alan Agresti. Categorical data analysis 3rd ed. *John Willey and Sons, USA*, 2013.

Robert Kosara, Fabian Bendix, and Helwig Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE transactions on visualization and computer graphics*, 12(4):558–568, 2006.

Donald E Super, Reuben Starishevsky, Norman Matlin, and Jan Pieere Jordaan. Career development; self-concept theory. 1963.

Donald E Super. A life-span, life-space approach to career development. *Journal of vocational behavior*, 16(3):282–298, 1980.

James Williams, Aisling Murray, Cathal McCrory, and Sinead McNally. *Growing Up in Ireland national longitudinal study of children: development from birth to three years infant cohort.* Department of Children and Youth Affairs, 2013.

OF GROWING UP IN IRELAND'S. A summary guide to wave 4 of growing up in ireland's cohort'98 (child cohort) at 20 years of age. 2021.

RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA, 2022. URL `http://www.rstudio.com/`.

Scott Chacon and Ben Straub. *Pro Git*. 2014. doi:10.1007/978-1-4842-0076-6.

Yihui Xie et al. knitr: A general-purpose package for dynamic report generation in r. *R package version*, 1(7):1, 2013.

Kirill Müller. *here: A Simpler Way to Find Your Files*, 2020. URL `https://CRAN.R-project.org/package=here`. R package version 1.0.1.

Kirill Müller and Hadley Wickham. *tibble: Simple Data Frames*, 2022. URL `https://CRAN.R-project.org/package=tibble`. R package version 3.1.7.

Joseph Larmarange. Labelled: Manipulating labelled data, 2018.

Amelia McNamara and Nicholas J. Horton. Wrangling categorical data in r. *American Statistician*, 72, 2018. ISSN 15372731. doi:10.1080/00031305.2017.1356375.

Taiyun Wei, Viliam Simko, Michael Levy, Yihui Xie, Yan Jin, Jeff Zemla, et al. Package 'corrplot'. *Statistician*, 56(316):e24, 2017.

Michael Friendly. Graphical methods for categorical data. *Proceedings of SAS SUGI*, 17:190–200, 1992.

Maintainer Gabor Csardi. Package 'igraph'. *Last accessed*, 3(09):2013, 2013.

John A Hartigan and Beat Kleiner. A mosaic of television ratings. *The American Statistician*, 38(1):32–35, 1984.

Michael Friendly. A fourfold display for 2 by 2 by k tables. Technical report, Citeseer, 1994.

Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. Upset: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics (InfoVis)*, 20(12):1983–1992, 2014a. doi:10.1109/TVCG.2014.2346248.

Sean T Lyons, Chris A Higgins, and Linda Duxbury. Work values: Development of a new three-dimensional structure based on confirmatory smallest space analysis. *Journal of Organizational Behavior*, 31(7):969–1002, 2010.

Shalom H Schwartz. A theory of cultural values and some implications for work. *Applied psychology*, 48(1):23–47, 1999.

Margaret Mooney Marini, Pi-Ling Fan, Erica Finley, and Ann M Beutel. Gender and job values. *Sociology of Education*, pages 49–65, 1996.

Maria Ros, Shalom H Schwartz, and Shoshana Surkiss. Basic individual values, work values, and the meaning of work. *Applied psychology*, 48(1):49–71, 1999.

John L Holland, Gary D Gottfredson, and Herbert G Baker. Validity of vocational aspirations and interest inventories: Extended, replicated, and reinterpreted. *Journal of Counseling Psychology*, 37(3):337, 1990.

DE Super. A life-span, life-space approach to career development. career choice and development: Applying contemporary theories to practice. *D. Brown, L. Brooks, San Francisco*, 1990.

Natalie M Ferry. Factors influencing career choices of adolescents and young adults in rural pennsylvania. *Journal of Extension*, 44(3):1–6, 2006.

Pentti Sinisalo. Changing work values and expressed educational plans of adolescents: A cross-sectional follow-up of three cohorts in finland. *European Journal of Psychology of Education*, 19(3):227–236, 2004.

Andreas Hirschi. Positive adolescent career development: The role of intrinsic and extrinsic work values. *The career development quarterly*, 58(3):276–287, 2010.

Aisling Murray, Cathal McCrory, Maeve Thornton, James Williams, Amanda Quail, Lorraine Swords, Erika Doyle, and Elaine Harris. Growing up in ireland: Design, instrumentation and procedures for the child cohort (aged 9 years). Technical report, Child Cohort Technical Report, 2010.

Monica Kirkpatrick Johnson. Social origins, adolescent experiences, and work value trajectories during the transition to adulthood. *Social forces*, 80(4):1307–1340, 2002.

Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. Upset: visualization of intersecting sets. *IEEE transactions on visualization and computer graphics*, 20(12):1983–1992, 2014b.

Barret Schloerke, Di Cook, Joseph Larmarange, Francois Briatte, Moritz Marbach, Edwin Thoen, Amos Elberg, and Jason Crowley. *GGally: Extension to 'ggplot2'*, 2021. URL `https://CRAN.R-project.org/package=GGally`. R package version 2.1.2.

Daniel Lüdecke. ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, 3(26):772, 2018. doi:10.21105/joss.00772.

Esther Tan. Comparing the work values of australian and singaporean adolescents. 1996.

Judith S Bridges. Sex differences in occupational values. *Sex roles*, 20(3):205–211, 1989.

A Regula Herzog. High school seniors' occupational plans and values: Trends in sex differences 1976 through 1980. *Sociology of Education*, pages 1–13, 1982.

Appendices

Table 6.1: Appendix I: Preliminary variables selected for the analysis

	Short_name	Description
1	gender	Young Person (YP) Gender Wave 3
2	PC1_edu	What is the highest level of education Primary Caregiver (PC) has completed?
3	PC1_emp	Primary Caregive's employment status Wave 3 Grid
4	equalised_inc	Equivalised Household Annual Income - Quintiles W3
5	work_imp	YP- How important - Profession and work
6	religion_imp	YP-How important - One's own family and children
7	own_family_imp	YP-How important - Religion
8	cog_naming	YP-Cognitive test-Naming Task
9	cog_maths	YP-Cognitive test-Maths Score
10	cog_vocab	YP-Cognitive test-Vocabulary test
11	cognitive_scores	Derived Variable-Cognitive Scores
12	TY	YP-Did you take Transition Year
13	conscientious_score	YP-conscientious Subscale W3
14	ID	Unique Identifier-Household ID
15	left_school	YP-When did you leave school - Year
16	no_boys	Pricipal's Questionnaire- Number of Boys
17	no_girls	Pricipal's Questionnaire- Number of Girls
18	school_type	Derived Variable School Type
19	schoolID	SchoolID-Derived
20	Parental_interaction_subjects	YP-How often have/did your parents discussed how you are getting on with different subjects at school/college

	Short_name	Description
21	Parental_interaction_work	YP-How often have/did your parent asked how you are/were coping with the amount of work (course-work etc) for your courses
22	Parental_interaction_future	How often have/did your parent asked how you are/were getting on with your teachers/lecturers
23	Parental_interaction_future	YP-How often have/did your parent discussed your plans for the future
24	Parental_interaction_friends	YP-How often have/did your parent asked how you are/were getting on with friends
25	Parental_interaction_exams	YP-How often have/did your parent discussed how you did in tests or exams
26	parental_interactions	Derived variable- Parental Interactions
27	career_guidance	Derived Variable- guidance score
28	part_time_job	YP -Ever in a part-time paid job in term-time while attending school or college (don t include jobs during the school holidays)
29	work_for_business	YP- Do you ever do any work for a business owned or run by a member of your family (paid or unpaid work)
30	work_exp	During this/your final school year had a short term work experience placement, as part of your school curriculum
31	household_region	PC- Household Region -Wave 1
32	social_class	Family's Social Class
33	school_ethos	Principal Questionnaire- Describe ethos of school?
34	suitable_accomodation	Do you feel your current accommodation is suitable for your family's needs

	Short_name	Description
35	have_a_car	Would your family like to have a car but you cannot afford it
36	what_would_you_do_post_school	YP-What do you think you are most likely to do when you leave school
37	PC_born_in_Ireland	PC-Were you born in Ireland
38	SC_born_in_Ireland	SC- Were you born in Ireland
39	YP_citizen_of_Ireland	YP- Are you a citizen of Ireland
40	How far does PC think YP will go in Education?	PC_ How far do you expect YP will go in education or training
41	How far does SC think YP will go in Education?	SC-How far do you expect YP will go in education or training
42	region of household	Derived Variable-region
43	grade_lcmaths	YP-Grade in Mathematics
44	level_jcmaths	YP-Level of Mathematics in Junior Cert
45	Level of Education for Secondary Caregiver	What is the highest level of education you have completed