Volume 6 Number 6 August 2018

Digital Journalism

DIGITAL JOURNALISM

ISSN: 2167-0811 (Print) 2167-082X (Online) Journal homepage: https://www.tandfonline.com/loi/rdij20

Patterns in Award Winning Data Storytelling

Story Types, Enabling Tools and Competences

Adegboyega Ojo & Bahareh Heravi

To cite this article: Adegboyega Ojo & Bahareh Heravi (2018) Patterns in Award Winning Data Storytelling, Digital Journalism, 6:6, 693-718, DOI: 10.1080/21670811.2017.1403291

To link to this article: https://doi.org/10.1080/21670811.2017.1403291

Published online: 30 Nov 2017.

PATTERNS IN AWARD WINNING DATA STORYTELLING

Story Types, Enabling Tools and Competences

Adegboyega Ojo and **Bahareh Heravi**

Data storytelling is rapidly gaining prominence as a characteristic activity of digital journalism with significant adoption by small and large media houses. While a handful of previous studies have examined what characterises aspects of data storytelling like narratives and visualisation or analysis based on single cases, we are yet to see a systematic effort to harness these available resources to gain better insight into what characterises good data stories and how these are created. This study analysed 44 cases of outstanding data storytelling practices comprising winning entries of the Global Editors Network's Data Journalism Award from 2013 to 2016 to bridge this knowledge gap. Based on a conceptual model we developed, we uniformly characterised each of the 44 cases and then proceeded to determine types of these stories and the nature of technologies employed in creating them. Our findings refine the traditional typology of data stories from the journalistic perspective and also identify core technologies and tools that appear central to good data journalism practice. We also discuss our findings in relations to the recently published 2017 winning entries. Our results have significant implications for the required competencies for data journalists in contemporary and future newsrooms.

KEYWORDS data-driven journalism; data journalism skills; data journalism tools; data story; data storytelling types; Global Editor Network

Introduction

Data journalism is an aspect of contemporary journalism in which techniques such as data analytics, programming and narrative visualisation are employed in additional to traditional journalistic methods to create data stories (Appelgren and Nygren 2014). Data stories are artefacts for revealing and communicating insights gained from the analysis of data-sets obtained from the public domain, crowdsourcing or big data sources. Data storytelling (i.e. the practice of creating data stories) is a structured approach comprising data, visuals and narratives for communicating insights from data (Dykes 2016). The object of developing data stories is to give voice to the data to inform, explain, persuade or engage the target audience (Slaney 2012).

Routledge Taylor & Francis Group

Digital Journalism, 2018
Vol. 6, No. 6, 693–718, https://doi.org/10.1080/21670811.2017.1403291

Despite the rapidly growing popularity of data journalism (Hewett 2015) and its adoption by large media organisations such as The Times, The Washington Post and The Guardian (Segel and Heer 2010), scholarly publications on data storytelling are limited. For instance, a search on Google Scholar for the term "data storytelling" in early July 2016 returned only 202 documents. The same search returned 316 documents in April 2017, which illustrated the growing interest in the field in the previous year. A similar Google Scholar search on "data journalism" returned 2910 documents in April 2017, 639 of which are produced since the beginning of 2016. Similarly, a search on the Scopus bibliographic database in April 2017 returned only 13 documents listed as containing the phrase "data storytelling" in their titles, abstracts or keywords and 83 documents for those mentioning "data journalism". This compares to more than 68,000 publications returned by Google Scholar using the search terms "data analytics" and 148,000 for "data science" and just over 5000 documents and 1300 documents on Scopus, respectively. Thus, roughly 0.5–4 per cent of the research attention in data analytics and science is devoted to data journalism and data storytelling—arguably one of its most valuable aspects. Despite this paucity, there are a few notable publications on data storytelling (Segel and Heer 2010) and (Lee et al. 2015) which identify core elements and rigorously describe the design space for data stories and narrative visualisations. Other works have attempted to prescribe good data story practices (Alexander and Vetere 2011) and analysed concrete data storytelling practices (Pouchard, Barton, and Zilinski 2014). In addition, few practitioner-directed articles such as (Stikeleather 2013) that have sought to contribute good storytelling practices.

The Global Editors Network (GEN) a cross-platform community for editors and media innovators (GEN 2016) has recognised outstanding practice in Data Journalism since 2012. The winning entries are presented on their community portal (community.globalnetwork.org). In our opinion, this repository of good practice constitutes an invaluable source of information for deconstructing data storytelling to produce more systematised knowledge about options for developing different types of data stories. One of the first steps in undertaking this challenge is to develop a conceptual framework for characterising data storytelling. Such a framework should enable the user to answer basic journalistic questions about the data story cases regarding "who, what, where, why, when and how" (5W-1H) from the resulting knowledgebase. In a recent work Young, Hermida, and Fulda (2017), the authors analysed the nature and quality of the subset of all Canadian finalists and winners in this repository between 2012 and 2015.

This study provides complementary analysis of GEN winning entries with the goal of better understanding the nature or type of data stories in the repository and how different technologies are being combined to create data stories. To this end, we developed a conceptual framework based on extant literature and then applied the framework to describe 44 data storytelling cases recognised as outstanding in Data Journalism (DJA) Award from 2012 to 2016. The resulting repository of cases was analysed using a multi-case approach (Baxter and Jack 2008) and content analysis. Findings from our work refine the traditional "typology of intent" of data stories in particular inform and explain (Slaney 2012) from the journalistic perspective. The findings also provide a "Data Journalism Technology Competency Architecture" for training and development of the future data journalist, configuration of teams working and dynamics in contemporary and future newsrooms.

Conceptual Foundation

Data journalism could be described as the "application of data science to journalism, where data science is defined as the study of the extraction of knowledge from data" (Howard 2014). Howard (2014) further explains that data journalism encompasses "gathering, cleaning, organising, analysing, visualising, and publishing data to support the creation of acts of journalism". A similar definition for the discipline of data journalism is provided in Berret and Phillips (2016) as a "field [that] encompasses a suite of practices for collecting, analysing, visualising, and publishing data for journalistic purposes". Many consider data journalism to have its roots in Computer Assisted Reporting (CAR) which dates back to the 1960s, and involves the application of social science methods in journalism (Knight 2015). Philip Meyer is considered the pioneer of CAR and held to be one of the first American journalists to have used computers for investigative journalism. Shortly after the initial popularity of CAR, and long before the term "data journalism" became popular, Meyer (1973) coined the term of "precision journalism" for the type of journalism involving the use of computers to analyse data. Meyer's specific view on precision journalism was that journalists would be wrong less often if they use a scientific approach to analysing data, employing social science research methods (Meyer 1973, 2002).

Another closely related field to data journalism is Computational Journalism; described as "the application of computing and computational thinking to the activities of journalism including information gathering, organisation and sense-making, communication and presentation, and dissemination and public response to news information" (Gynnild 2013). It combines algorithms, data and knowledge from social sciences to enable journalists to explore the increasingly large amount of structured and unstructured information as they search for stories (Flew et al. 2012). In general, the goals of computational journalism are same as those of data journalism and the older field of computer-assisted reporting. Consequently, for the purpose of this paper, the authors of this study do not distinguish between these terms.

Data are at the heart of data storytelling. Kitchin (2014) describes data as raw elements that are abstracted from phenomena and measured and recorded in different ways. In the most accurate sense, data are actually those elements that have been selected and harvested from all possible data—"as such, data are inherently partial, selective and representative, and the distinguishing criteria used in their capture has consequence" (Kitchin 2014). In common use, data are characterised as unprocessed, abstract, discrete and aggregative. Data can be captured directly through some form of measurement from observations, experiments and records or produced by devices as by-products (exhaust) of their main function. An example of exhaust data is customer data that are captured by an online payment system which could be exploited for other purposes later, for instance in determining buying patterns (Manyika et al. 2011). Exhaust data could be transient in nature if they are never examined or used after they are generated, for instance due to cost of storage or processing. Data could also be derived from other data through analysis and processing of captured data (Kitchin 2014). There are at least two primary ways in which data can be generated. First, data can be captured directly through some form of measurements such as observation, surveys, lab and field experiments, record keeping (e.g. filling out forms or writing a diary), cameras, scanners and sensors. In these cases, data are usually the deliberate product of measurement; that is, the intention was to generate useful data. Exhaust data are inherently produced by a device or system as a by-product of the main function rather than the primary output (Manyika et al. 2011). In addition to the source of data, there are other typologies for data based on form (quantitative versus qualitative), structure (structured versus unstructured) and type (indexical, attribute and metadata) and type (numerical, textual, spatial, networked and temporal) (Kitchin 2014; Manyika et al. 2011).

Telling stories from data or "data storytelling" is central to the practice of data journalism. Stories are fundamental components of human experience (Slaney 2012). They are mechanisms for communicating information in a psychologically efficient format (Segel and Heer 2010). Like stories, a data story comprises a set of story pieces which are backed up with specific facts and are often visualised to support one or more intended messages (Lee et al. 2015). According to the same authors, the story pieces making up a data story are presented in a meaningful order (i.e. creating the plot) to achieve the intended author's high-level goals which include educating or entertaining the viewers, convincing and persuading the audience with thought-provoking opinions. According to Slaney (2012), authors of data stories may also be interested in comforting, entertain, terrorise or inform the intended audience.

The notion of story pieces and required ordering of these pieces into a (data) story is consistent with the fundamental nature of stories. Stories or those stories that are considered meaningful have specific narrative structures that are recognisable by an audience (Rayfield 1972). In fact, stories have certain structure could be used to legitimately recognise it as such (Rayfield 1972). Thus, stories have some implicit complexity comprising structure, elements and concepts (Lee et al. 2015). Specifically, Rayfield (1972) argued that: (1) a listener of a story will accept an item as story only if it has certain structure characterised by some minimal and maximal complexity, (2) the degree of the complexity and nature of such minimal and maximal strictures are largely the same across cultures. This rigorous characterisation could be useful in delineating what a data story is and what it is not. For instance, Lee et al. (2015) in describing what they considered as a data story in their work indicated that web-based interactive visualisations that support completely free exploration without guidance is not considered as a data story nor are charts posted on the web with no written explanations and annotations that help the reader to capture the intended message.

Data storytelling could be described as the act of creating and communicating data stories. Lee et al. (2015) characterised the visual data storytelling process as comprising three major phases—exploring data, making a story and telling a story. Exploration of data comprises activities centred on exploratory analysis to collect data excerpts containing derived insights from data, variations, or quick externalisation of data. For instance the Google Sheet Explore feature provides this type of analysis to summarise and identify variations or outliers in data. The story making is achieved over a number of iterations and involves constructing a storyline or plot to fit the data excerpts generated in the exploration phase (Lee et al. 2015). The core activities involved in story making includes ordering, connecting story pieces and formulating main and closing messages. The third step according to Lee and co-authors (2015); include building a representation by choosing materials and medium, sharing the story and receiving feedback from the audience. Segel and Heer (2010) closely link the goal of storytelling to that of visualisation which includes communicating information in a psychologically efficient format. A similar viewpoint is expressed in Stikeleather (2013) where the essence of a good data visualisation or narrative visualisation is considered to be good story telling.

The narrative style of narrative visualisations lies in the spectrum of author-driven and reader-driven approaches (Segel and Heer 2010). The author-driven style has a linear path through the visualisation with strong messaging and no interactivity. On the other extreme, the reader-driven approach according to Segel and Heer (2010) has no prescribed ordering of images, no messaging and has a high degree of interactivity. While the author-driven approach is effective for efficient communication, the reader-driven approach is ideal for activities such as data diagnostics, pattern discovery and hypothesis formation (Segel and Heer 2010). The same authors identified three hybrid approaches including (1) Martini Glass structure where the narrative begins with an author-driven style and later opens up to a reader-exploration; (2) Interactive Slideshow in which the typical slideshow format allows some interactivity in some of the slides before progressing to subsequent slides; (3) Drill-down stories in which the visualisation structure provides a general theme and then allows the users to indicate particular instance to explore in details.

A myriad of analytical techniques is employed in generating insights from a large amount of data for storytelling. Manyika et al. (2011) identified several of these techniques including association rule mining, classification, cluster analysis, data fusion and integration, machine learning, genetic algorithms, natural language processing, network analysis, pattern recognition, predictive modelling, regression analysis, sentiments analysis, spatial analysis and time series analysis. Some of the tools implementing these analytical techniques include R, Python, Microsoft Excel, IBM-SPSS, Tableau, NodeXL, Google Fussion Tables, RapidMiner and Knime Analytics (Berret and Phillips 2016; Fink and Anderson 2014; Gynnild 2013; Hewett 2016; Müller et al. 2016; Parasie and Dagiral 2013; Segel and Heer 2010; Uskali and Kuutti 2015).

On the research aspect, a number of studies in recent years have examined the practice of data journalism in various countries or specific cities/states, such as Sweden (Appelgren and Nygren 2014), Norway (Karlsen and Stavelin 2014), Belgium (De Maeyer et al. 2015), United Kingdom (Knight 2015), United States (Fink and Anderson 2014; Parasie 2015; Parasie and Dagiral 2013). These studies have largely examined a number of news organisations, or data journalists in respective countries with regards to educational background and skills of data journalists and the nature of tools used for data journalism. Specifically, the reviewed works examined the evolving skill-set and competences required for data journalism and data storytelling and adoption pattern by newsrooms. Fink and Anderson (2015) in particular found that larger news organisations in general have higher level of technically skilled data journalists, and thus able to develop more interesting data stories. Parasie and Dagiral (2013) also reported that integrating programmers into newsrooms opens up the technology-driven innovation potentials of news organisations. This category of work is complemented with the body of research on visualisation, and narrative visualisations such as those reported in Segel and Heer (2010) and Lee et al. (2015). Table 1 highlights key data storytelling concepts described above.

While past studies have looked into aspects of the skills required or currently held by data journalists in newsrooms, barriers to good data storytelling in the studies newsrooms as well as insight into the practice of data journalism they are yet to specifically

TABLE 1
Design space and options for data storytelling from literature

Aspect	Options		Reference
Purpose	Inform, Persuade, Entertain, Comfort, Explain, Terrorise		Lee et al. (2015), Slaney (2012)
Audience	General Public, Specific Group		Lee et al. (2015)
Story Elements	Character, Context, Plot		Rayfield (1972), Lee et al. (2015)
Medium	Browser, Mobile, Print		Lee et al. (2015)
Data	Source	Captured, Exhaust, Derived	Agarwal and Dhar (2014), Kitchin (2014)
	Form	Qualitative, Quantitative	
	Structure	Structured, Unstructured	
	Type	Numerical, Textual, Spatial, Networked and Temporal	
Narrative Style	Author-driven, Reader-driven, Hybrid (Martini-glass, Interactive Slideshow and Drill-down)		Segel and Heer (2010)
Interactivity	Static, Interactive, Limited Interactivity, Search/Filtering/ Selection		Segel and Heer (2010), Young, Hermida, and Fulda (2017)
Representation	Magazine Style, Annotated Graph/Map, Partitioned Poster, Flowchart, Comic Strip, Slideshow, Film/Video/Animation, Others		Segel and Heer (2010), Young, Hermida, and Fulda (2017)
Analytical Techniques	Statistical Analysis, Data Mining, Natural Language Processing, Machine Learning, Spatial Analysis		Manyika et al. (2011)
Technological Tools	Excel, Tableau Public, Open Refine, Google Fusion, R, Python, SPSS, Rubby, PHP. JavaScript, HTML, ESRI Mapping Software, Many Eyes		Parasie and Dagiral (2013), Gynnild (2013), Berret and Phillips (2016), Fink and Anderson (2014), Uskali and Kuutti (2015), Müller et al. (2016), Segel and Heer (2010), Hewett (2016), Bakker (2014)

study the characteristics of "good data storytelling/journalism", required competences for developing high-quality data stories, or specific tools or techniques used in data journalism work. We believe the only study that attempted to examine characteristics of good data storytelling, with a similar cohort of data, is Young, Hermida, and Fulda (2017). Unlike our study that looks into winning entries globally, Young, Hermida, and Fulda (2017) study the shortlisted entries that originated from Canada in three different journalism awards, including the GEN Data Journalism Awards, between 2012 and 2015. In their study they looked into 26 cases and examined the quality of the Canadian shortlisted entries in terms of the quality of tools used, the diversity of the team involved in the projects, interactivity of the stories and the visual elements used. While Young, Hermida, and Fulda (2017) study has a narrower scope than the study in hand in terms of diversity of cases and their quality; given that they were shortlisted cases and not winning cases, certain aspects of their findings are not far from our findings as shown in the discussion section.

Given the increasing availability of good data stories and data storytelling exemplars in data journalism practices, there is an opportunity to ask questions about the nature of good data stories. Specifically, we could inquire about the specific types of data stories or narratives that appear to better engage the audience and what it takes to create such stories. The GEN, a cross-platform community of editors-in-chief and media innovators[1] launched the Data Journalism Award in 2012 as an annual series of events to recognise outstanding projects in data journalism. Since then, awards have been given to about 50 projects from large newsrooms, small newsrooms and individuals on theme related to investigative journalism, data visualisation, data journalism website, open data award, best use of data in breaking news and news data apps. The winning entries over the years collectively provide examples of how the use of data stories is shifting the frontiers of the practice of journalism. Specifically, the GEN community portal at https://community.globaleditorsnetwork.org/projects provides anecdotes and short descriptions of these projects. In our opinion, a systematic capture of these descriptions could provide a rich stock of best practice exemplars that could be analysed to provide concrete answers to following research questions:

RQ1: What data story types are characteristic of the winning entries in data Journalism Awards?

RQ2: How are the different technologies combined (or used together) to create these compelling data stories?

We explain below our approach to answering these questions. Specifically, we describe how we curated a data-set describing core elements of the data story cases, and how we carried out both exploratory and explanatory analyses of the stories to answer the above questions.

Curating the GEN Data Journalism Award Data-set

The first step in our work involved curating a data-set describing core elements of the GEN Data Story cases selected as winning entries from 2013 to 2016. According to Simon Rogers, Data Journalism Awards Director,[2] the following five criteria were considered in identifying or characterising data story entries as exemplary and a winning entry. The first criterion assesses the innovation character of the data story. It seeks to know if the piece of work is doing something new, for instance in how data is reported. The second criterion examines the "content and story" element of the entry. This criterion looks beyond producing attractive visualisation to whether if there is some compelling story being told. It examines the journalistic aspects (including edit quality) of the data story. The third criterion is centred on how technology is used in telling the data story. It examines if the adopted technology in the case works across different media (mobile in addition to browser), and how it takes advantage of the peculiarities of these media in telling the stories. The fourth criterion is about the context of the story (or plot). This criterion rewards ambitious work from war zones, developing nations or other difficult situations. It looks at the conditions that the data journalists worked under. The fifth and last criterion re-examines the data stories to see if it exemplifies "best journalism around data", going beyond reproducing statistics but endowing these analyses with greater meaning. An implication of these set of criteria is that at least in the GEN community, data storytelling or data journalism is first and foremost about the journalism. About 50 cases were selected by the jurors based on these criteria from 2012 to 2016.

However, we omitted entries for the year 2012 due to the sparsity of information about the cases. In total, 44 cases were selected from the list of winning stories from 2013 to 2016 based on the availability of information on different aspects of the story. The concepts described in Table 1 were employed in describing each of the cases. The authors carefully read the different descriptions of the winning cases on the GEN Data Journalism Award website and created an entry for each of the cases in a Google Sheet.

The selected cases originated from 14 different countries as indicated in Figure 1. However, United States dominates (46 per cent of the 44) as the home of GEN Data Journalism awardees. There were five cases from the United Kingdom, four from France and Argentina and two from Peru. Other countries with at least 1 winning entry include Switzerland, Spain, Italy, India, Hong Kong, Germany, Denmark, Costa Rica and Canada. As Figure 1 shows, North and South America, Europe and Asia are represented in the GEN award map, while the Africa, Australia, Russia are unrepresented. The case of Africa and conflict regions is interesting as these regions are story-rich and in fact, some of the settings of the winning data stories are centred on these regions.

The curated data-set comprises 44 rows and the following 16 columns—Title of project or entry, the of the project description on GEN Community website, the address of the project website, category of the entry, description of the entry, tags used in describing the entry, technology employed in developing the entry or data story, country of journalist, organisation, year of submission, nature of the interactivity in the data story, targeted audience of the data story, nature of data underpinning the story, sector targeted by the story, source of the data for the story and type of the story.

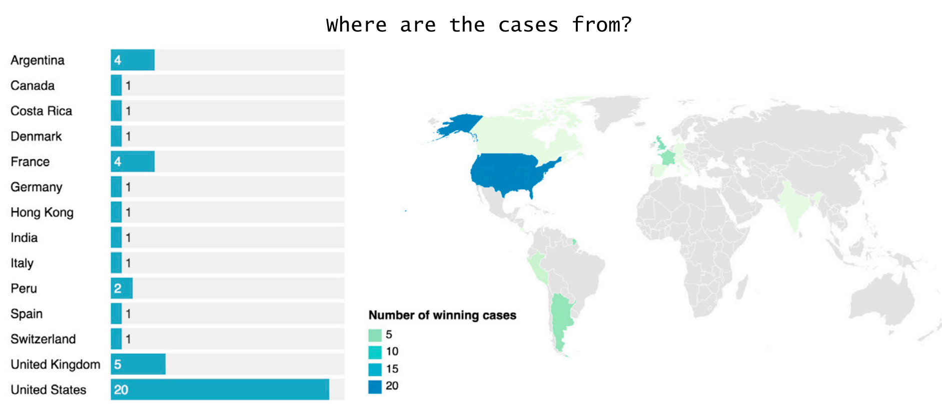

FIGURE 1
Source of the cases

Exploring the Data-set

After coding each of the 44 cases, we explored the data-set to determine emerging patterns in terms of: purpose of the stories, the media used for telling the stories, story types, how the stories are represented, degree of interactivity of the stories and the technology tools employed in the narrating and presenting the data stories. These patterns are presented in Table 1. We show for each element, the most common items across the 44 cases. For instance, most of the data stories are largely aimed at "informing" the audience about a specific topic or phenomenon with the "web browser" as the medium of the choice. Figures 2–5 provide additional information on the observed patterns with respect to the data story elements in Table 2.

In general, many of the cases had more than one goal, for instance, a story may aim at informing the public and simultaneously persuading or entertaining. Specifically, Figure 2 shows that about 73 per cent of the cases had as part of their goals to "inform" the target audience (e.g. that linking metadata information about citizens' call records to email, bank data, etc. is sufficient to reveal the thought and living patterns of subjects) while about 41 per cent of the stories were also interested in "persuading" the audience towards adopting some positions (e.g. persuading parents who are doubtful about vaccination that there vaccination programmes works. About 39 per cent of cases tried "explaining" some phenomena to the public, for instance how millions of voters that are disproportionately minorities could inadvertently be prevented from voting based on the rules used by a computer programme designed to identify irregularities during the election. Some of the reviewed cases were collections of different works (18 per cent) and consequently had a combination of different goals. The overall picture shows that even when the agenda of a data story is to persuade or explain; good practice may require informing the audience about the context and background of the subject matter.

The level of interactivity employed in telling a data story directly affects the story experience. Figure 3 shows 59 per cent of the reviewed data story cases were "interactive" while 27 per cent of the stories provided features for searching, filtering and selection. Only 7 per cent of the cases employed map-based interactivity. Static images and graphics were used 14 per cent of the cases. Most (about 77 per cent) of the interactive features were rendered through annotated graphics and maps according to Figure 4. Between 10 per cent and 18 per cent of the cases used videos, web application, games as media for interactive data storytelling.

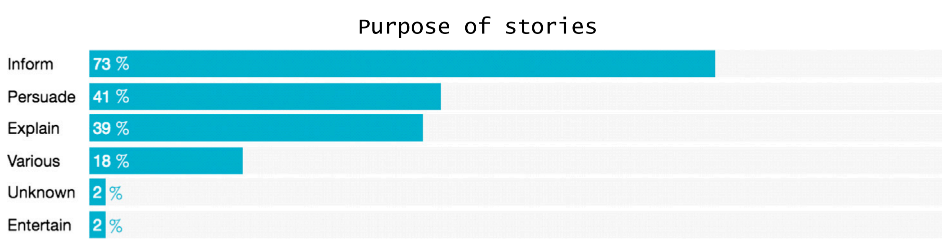

FIGURE 2
What do the data stories want to achieve?

TABLE 2
Patterns in data story aspects across cases

No.	Element	Most common values for story aspects
1	Purpose	Inform (32), Persuade (18), Explain (17)
2	Medium	Browser (43), Mobile (1)
3	Story-type	Hybrid-Drill down story (14), Hybrid-Slideshow (8) Interactive (8), Author-driven (6), Reader-driven (5), Hybrid-Martini-Glass (3)
4	Representation	Annotated Graph/Map (34), Video (8), Magazine Style (6), Image (5), Web App (6), Question Game (5), Graphics (3)
5	Interactivity	Interactive (26), Filtering (12), Selection (12), Search (12), Static graphics (4), Map (3)
6	Technology	Javascript (16), Excel (13), D3 (10), HTML5/HTML (17), Python (9), Adobe Illustrator (5), CSS3 (5), MySQL (6), Google Maps (5), R(4), jQuery (3), Tableau (3)

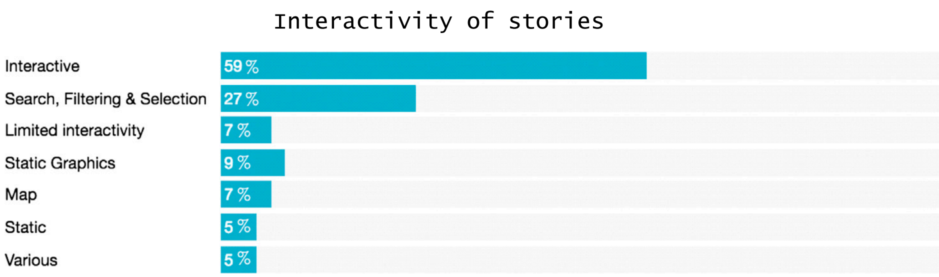

FIGURE 3
What kinds of interactivity are employed in telling data stories?

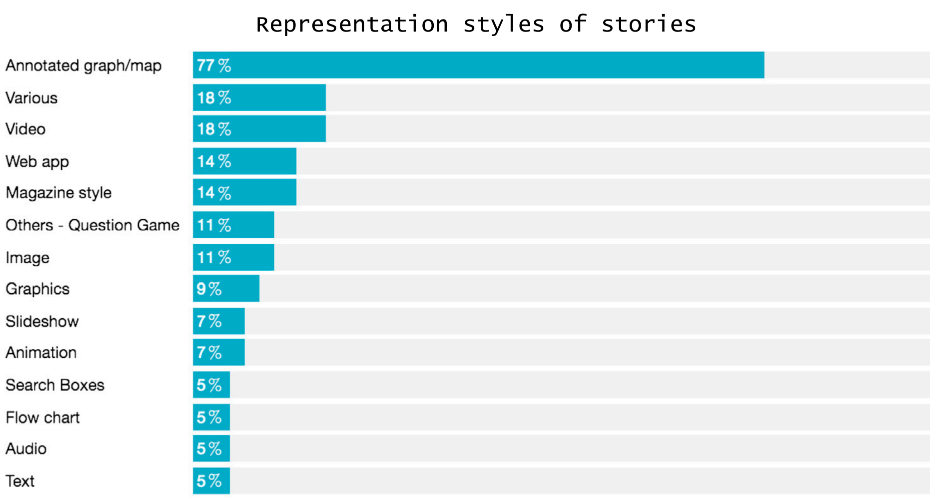

FIGURE 4
How are these data stories represented?

FIGURE 5
What kinds of technologies are employed in developing data stories?

Lastly, a bird's eye view of the tools employed in constructing the data stories according to Figure 5 reveals that Web development tools, Data Analysis/Analytics tools, Data Visualisation frameworks and database are most prominent. JavaScript and HTML appear most frequently as Web development tools, Microsoft Excel and Python appear are the most frequently used tools in the data analysis and analytics category, D3 is the most common data visualisation framework employed across the cases while MySQL standout as the most popular database tools. Other notably popular tools include jQuery (Scripting library), Google Refine (data preparation and refinement), Google Maps (mapping) and Adobe Illustrator (Graphics publishing). Together over 130 different tools and frameworks were employed across the 44 cases.

Analysing of the Data-set

The exploration of our data-set was followed by two types of analyses. The first analysis involved the use of the conventional approach to content analysis (Hsieh and Shannon 2005) to establish the nature of the data stories and the tools used in creating the data stories. The use of conventional content analysis is suitable when existing theory or research literature in a domain is limited. In this approach, categories and names (codes) for categories flow from the data. One advantage of this approach is that the knowledge generated is directly drawn from the data as opposed to when categories are determined a priori based on literature or some theoretical framework. Using the conventional content analysis approach, categories were developed to *represent the types of data stories (cases) and the different types of tools associated with the creation of these stories*. The coding process is characterised by two kinds of operations; a reductionist operation to establish the initial set of fine-grained categories based on story intent followed by a clustering operation to merge similar granular categories into more compact categories. The final set of categories were described over a number of cycles to improve coverage and obtain consensus. Considering the great importance attributed to the "content and story" element of data stories by the jurors and experts, the reductionist operation that produced our data story *types* was based on the "specific purpose or intent" of each of the 44 data stories. Consequently, our *final set* of data story categories or types are directly related and indeed, *refine* the more abstract story purposes shown in Table 2 which include inform, persuade and explain. These more specific types originating directly from journalistic practices potentially provides models that could be emulated to develop compelling data stories. In the case of our technology categories, the resulting categories provide logical groupings for the specific technologies associated with the 44 cases to enable further analysis, such as technology usage pattern and dependencies (the second type analysis discussed below). Frequency counts of our data story types and technology categories are presented as summaries in Figures 7 and 8.

The second type of analysis includes the determining the patterns of use of the different technologies in data storytelling. More specifically, based on our data-set, we are interested in establishing how different categories of technologies tend to be used together in the same project and dependencies among these tools. For this purpose, we mine association rules (Zhao 2003) and implication rules (Godin and Missaoui 1994) in our data-set using the "ConExp" tool.[3] The ConExp tool is a research tool developed for the analysis of simple attribute-object tables and exploration of the different dependencies that exist among attributes. In our case, the objects were the different data story projects and the attributes comprise the different categories of technologies employed in these projects (see Figure 6). Our goal is to analyse the dependencies in the use of the different categories of technologies when carrying out data story projects. This information will provide us with the set of technologies that are often used together in data story projects. It also allows us to determine the consequence of using a technology or set of technologies.

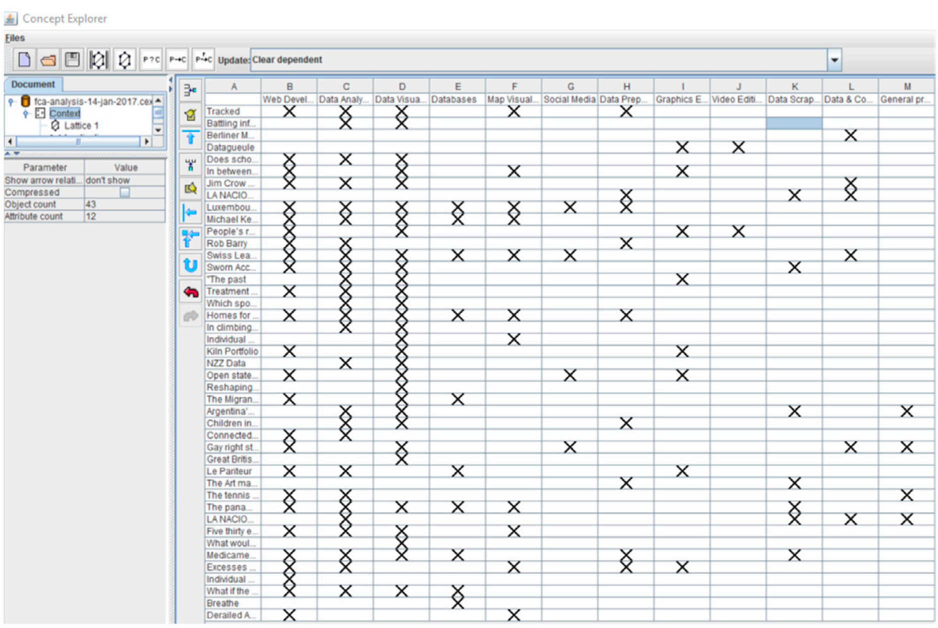

FIGURE 6
Screenshot of our data-set displayed in the ConExp Tool

Findings

This section reveals and describes the different story types discovered across the case, the prevalence and use of different technologies across projects and the patterns of usage of these technologies.

The Types

After a careful analysis of the themes and purpose or intent of the data stories projects, we identified seven Types comprising: Refute Claims, Reveal unintended consequences, Reveal information of personal interest, Enable deeper understanding of a phenomenon, Reveal anomalies and deficiencies in Systems Track changes in systems, and Reveal information about an entity in increasing levels of details. These categories are defined with examples in Table 3. While types like "revealing anomalies and deficiencies in systems", "revealing unintended consequences" and "refuting claims" are investigative in nature, others types easily find applications in areas of public education and advocacy (e.g. enabling a deeper understanding of a phenomenon). The "track changes" in systems" is particularly generic theme for monitoring and has wide applicability beyond the transparency contexts such as in the case study involving the tracking of the wealth of politicians serving in public offices in Argentina to performance monitoring of services and infrastructures in cities.

In terms of the frequency of these types in our data-set, 25 per cent (or 11 cases) of the stories were about the "Explaining a Phenomenon for deeper understanding" type. 20.4 per cent (9 cases) and 18 per cent (8 cases) of the cases were related to *revealing anomalies in systems* and revealing information of personal interest, respectively. About 6.8 per cent (3 cases) of the cases were about refuting claims and tracking changes in the system. Lastly, 5.4 per cent or 2 cases each were *about revealing unintended consequences* and *revealing information in increasing level of details.* We were unable to classify 6 out of the 4 cases due to lack of adequate information. These results which are also illustrated in Figure 7 will be discussed later in relations to good story telling practices.

Categories of Technologies

After describing the types of data stories in our data-set, we examine here the types of technologies used in creating these winning and compelling stories. Following the iterative content analysis process described in the methodology section, we identified 12 categories of tools employed across the 44 cases. These categories include Web Development and Publishing; Data Analysis; Data Visualisation; Map Visualisation; Databases; Social Media; Data Preparation and Wrangling; Graph Editor and Publishing; Video Editing and Processing; Data Scrapping; Data and Content Management platforms and General Programming. We briefly describe each of these categories below and provide concrete examples of these categories from our data-set.

- *Web Development and Publishing*—These are tools and technologies that are used for web-based application programming, development and publishing. Tools in this category also support the management of the code for such programmes. Examples are HTML, CSS, JavaScript, Browserify and Python.

TABLE 3
Types of data stories discovered from cases

No.	Case name	Description of type	Related data stories in sample
1	Refute claims	Reveal evidence that proves that a (widely) held assertion, view or opinion may not be correct or that it is at least contestable. *For example, the claim that harvesting metadata of messages alone does not compromise privacy of citizens may not be true*	1, 11, 37
2	Reveal unintended consequences	Provide insights that show that an action taken by an organisation or state may be having serious unintended consequences. *For example, a government policy to grant citizens the right to refuse some vaccination may inadvertently endanger a large segment of its population in the near future*	2, 7
3	Reveal information of personal interest	Provide concrete information about an event or issue of high personal and public interest. *For instance revealing the true cause of an accident by analysing the data related to the event or showing the wage gap between men and women in a countries workforce*	5, 8, 17, 20, 29, 30. 43, 44, 31
4	Enable deeper understanding of a phenomenon	Employs data to provide the public with detailed facts about phenomena. *This includes providing new facts on wars, consequences of policies, poverty or issues in specific regions*	4, 6, 19, 21, 25, 27, 28, 32, 38, 42, 43
5	Reveal anomalies and deficiencies in systems	Reveals how a group or category of actors may be dubiously taking advantage of a system. *For instance, revealing how tennis players may be involved in match fixes, or how older demographic group are being exploited by rogue health insurance agents*	9, 12, 13, 16, 18, 10, 33, 34, 39
6	Track changes in systems	Enables the tracking of changes in system or entity over time. *For instance tracking the changes in the assets of politicians in public office before, during and after taking a public office or how a city infrastructure or service has evolved over time*	14, 15, 24
7	Reveal information about an entity in increasing levels of details	This involves zooming into facts about a specific entity or phenomenon. *For instance, providing details about the assets of politicians at a point in a time in way that enables the reader from a high-level or cursory level to details of the different aspects of the entity*	23, 26

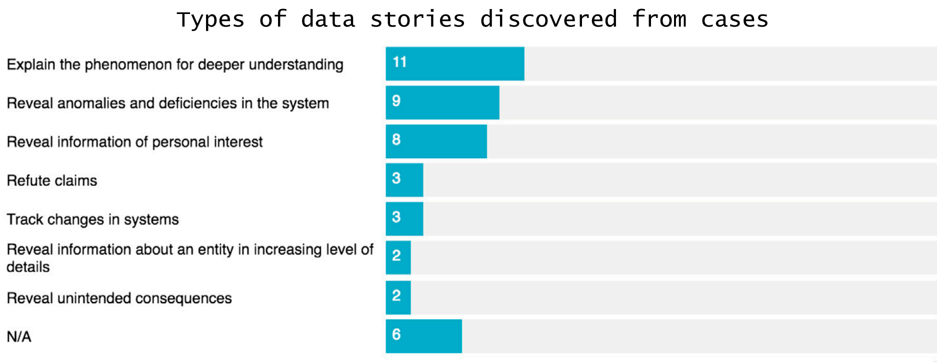

FIGURE 7
What types of data stories were found from cases?

- *Data Analysis*—These are tools that are used for exploring and analysing data. Examples in this category include Spreadsheet software, mentions of data analysis and data analysis techniques, statistics tools and technologies such as SPSS, SAS, Stata, Python data analysis libraries such as Pandas, R and object graph.
- *Data Visualisation*—These are tools and techniques used for user-facing data and information visualisation, in both static and interactive forms. Examples of tools in this category are Tableau Public, R (when used to data visualisation), Google fusion tables, Brackets, Highcharts, Linkurious, D3.js and certain other JavaScript libraries such as Raphaël.
- *Map Visualisation*—This category is related to visualisations specifically concerned with geospatial data, and visualisation on maps. Example of tools here include ArcGIS, MapBox, Mapper, Open Street Maps API and Google maps.
- *Databases*—This category includes database management tools, which are tools that are used for storage and retrieval of data of different types. Database can support structured data, unstructured data or graph data. Examples in this category are MySQL, SQL Server, SQL Base, PostgreSQL, MongoDB and Neo4j.
- *Social Media*—This category includes tools that are used for communicating on the social media, or downloading social data. Examples are Oxwall, Trello and Facebook Connect API.
- *Data Preparation and Wrangling*—This category includes tools that are used for data preparation, data cleaning and data transformation or wrangling. These tools are used for pre-processing data for suitable for data analysis. Examples are OpenRefine, Mr Dataconverter, Wigle, Nitro PDF and Tabula.
- *Graphics Editor and Publishing*—These are Graphic design tools that are used by graphic designers and graphics team for presenting and publishing data on both different media. Examples are Adobe Illustrator, Adobe Photoshop and Adobe After Effect.
- *Video Editing and Processing*—These are video editing and processing tools. They enable the presenting of data stories as videos. Examples are Avid, Adobe Premiere Pro and Processing.
- *Data Scraping*—These are tools that are used for scraping or collecting data from webpages, PDFs, or scanned documents. Examples in this category are Imacros, HTTrack, Omnipage and Nokogiri.

- *Data and Content Management Platforms*—These are tools that are used for document management, indexing and retrieval of documents and as document management platforms. Examples are Apache Solr, Backlight, Drupal 7, Google drive, Wordpress and Junar platform.
- *General programming*—These are tools and techniques that are used for coding/programming, which not specifically for any of the categories above, or tools that are used for producing, maintaining and sharing the code. Examples are Visual Basic, .NET Framework, Github and Jupyter Notebooks.

Our analysis shows that Data Visualisation (65 per cent), Web Development and Publishing (63 per cent) and Data Analysis tools (56 per cent) are the most common set of technologies for across the 44 cases. These three categories clearly standout as the core technologies for data storytelling in our data-set. Other notable categories of tools (we may call supporting tools) include Map Visualisation (26 per cent), Graphics Editing and Publishing (23 per cent), Databases (23 per cent), Data and Content Management Platforms (21 per cent) as well as Data Preparation and Wrangling (19 per cent). Figure 8 provides more information.

The relationships among these technologies in terms of usage pattern is shown in Table 4, the results of association and implication analysis carried out using the formal concept analysis tool—ConExp. Our results show that for between 8 and 11 cases, *Data Analysis and Database tools* were used along with *Web development tools*. Similarly, *Map Visualisation Tools* are used with both *Data Analysis* and *Web Development tools*. To confirm the core or base technologies, we consult the implication results in Table 4. For instance, the use of *Data Analysis and Database tools* implies the use of *Web development and publishing tools*. Similarly, we also see that the use of Databases and Map Visualisation tools implies the use of *Web development and publishing* as well as Data Analysis tools. In all the cases where the implication rules were significant, the three Web development and publishing, Data Analysis and Data Visualisation are clearly identified as base technologies for data storytelling (since they are the only ones in the Y column). From the implication rules, we also note that Social media, Data preparation

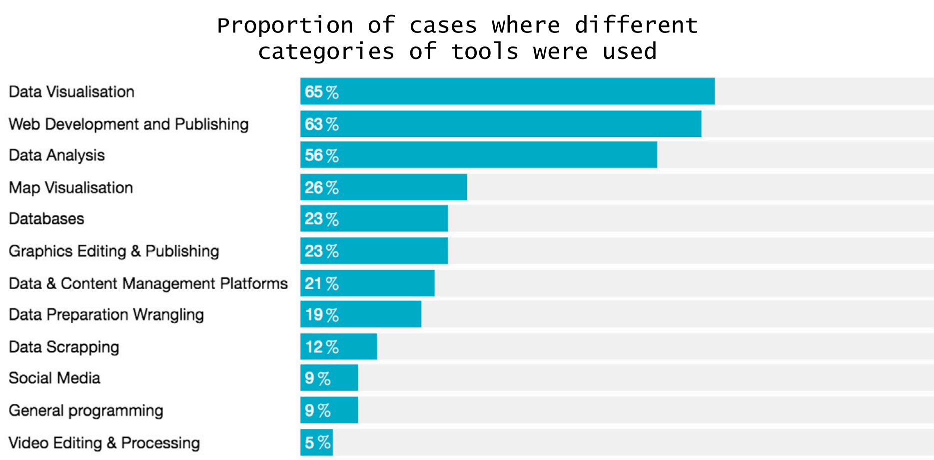

FIGURE 8
How prevalent is the use of specific categories of tools across cases?

TABLE 4
Technology use patterns across cases

	Associations in use of tools (Technology *X* and *Y* are used together)[a]			Implications founds in the use of tools (Using Technology *X* implies the use of *Y*)		
No.	*X*	*Y*	Confidence	*X*	*Y*	Cases
1	Map Visualisation	Web Development and Publishing	91%	Data Analysis and Databases	Web Development and Publishing	8
2	Databases	Web Development and Publishing	90%	Data Visualisation and Databases	Web Development and Publishing	8
3	Map Visualisation	Data Visualisation	82%	Data Analysis and Map Visualisation	Web Development and Publishing	8
4	Data Analysis and Databases	Web Development and Publishing;	100%	Databases and Map Visualisation	Web Development and Publishing and Data Analysis and Data Visualisation	5
5	Data Analysis and Map Visualisation	Web Development and Publishing	100%	Social Media	Web Development and Publishing and Data Visualisation	4
6	Data Visualisation and Databases	Web Development and Publishing	100%	Web Development and Publishing and Data Preparation Wrangling	Data Analysis	6
7	Data Visualisation and Map Visualisation	Web Development and Publishing	89%	Data Visualisation Data Preparation Wrangling	Data Analysis	5
8	Web Development and Publishing and Databases	Data Visualisation	89%	Map Visualisation and Data Preparation Wrangling	Web Development and Publishing and Data Analysis	4
9	Web Development and Publishing and Data Visualisation and Map Visualisation	Data Analysis	88%	Data Visualisation and Data Scrapping	Data Analysis	4
10	Web Development and Publishing	Data Visualisation	88%	Web Development and Publishing	Data Visualisation	3

(*Continued*)

TABLE 4. (*Continued*)

	Associations in use of tools (Technology *X* and *Y* are used together)[a]			Implications founds in the use of tools (Using Technology *X* implies the use of *Y*)		
No.	*X*	*Y*	Confidence	*X*	*Y*	Cases
	Data Analysis Map Visualisation			and Data "&" Content Management Platforms		

[a]BETWEEN 8 CASES AND 11 WERE ASSOCIATED WITH THE ASSOCIATION RULES.

and Wrangling and Data scrapping are needed as supporting technologies (albeit with 4–6 cases as support) in addition to Databases and Map Visualisation tools.

From our analysis and results, we arrive at the technology use architecture shown in Figure 9 for data storytelling. The figure identifies the three core technologies and two sets of supporting technologies: *web development and publishing*, *data visualisation* and *data analysis*. These are considered to be the core skills needed in newsrooms for data journalism projects, or in other words the most crucial skills to be expected from data journalists to hold. These core technologies are supported by two sets of technologies: (1) *databases* and *map visualisation*, and (2) *social media*, *data preparation and wrangling* and *data scraping*. The first set of supporting technologies were more often used with core technologies than the second set. We may consider the two sets of support technologies as complementary. We shall discuss below the implications of technology-use pattern when considering the competency framework for data journalists in future newsrooms.

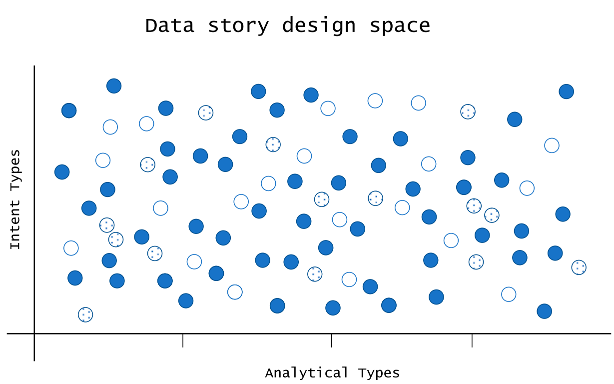

FIGURE 9
Example of a data story design space

Discussion

We believe the only study that has attempted to conduct a similar investigation to this work is Young, Hermida, and Fulda (2017), which studies the shortlisted entries from the Data Journalism Awards originated in Canada, along with Canadian entries from two other journalism awards. The study examined the quality of the shortlisted projects that originated from Canada from 2012 to 2015 in terms of the quality of tools used, the diversity of the team involved in the projects, interactivity of the stories and the visual elements used. Although Young, Hermida, and Fulda (2017) examined shortlisted entries only from Canada, and while there was only one common case between the two studies (Vancouver Sun one of the winners of the GEN Data Journalism Awards in 2014), the information in Table 1 largely agrees with conclusions in their work regarding the use of annotated graph/maps and videos as the primary visual elements or representation in reviewed data stories.

Regarding data story types, several authors have attempted to provide one or more typologies for data stories. The closest to our typology is the typology presented in Slaney (2012) comprising four types inform, explain, persuade and entertain, subset of which is shown as common purposes of data stories in Table 2. Our typology which comprises 7 types and presented in Table 3 are also based the intent, purpose or essence of a story. As explained in the method section, the basis for our decision to characterise the stories based on their intent is based on the importance attributed to this facet of data storytelling by expert jurors of the Data Journalism Award. *In our opinion, a typology based on a core facet of compelling or winning stories offers concrete models for emulation in practice and for further investigation.* When compared with the 4 types in Slaney (2012), our 7 types are sub-types of the "inform" and "explain" types. This relationship is shown in first column of Table 5.

In addition, we observe that the typologies presented in Kang (2015) and Gray, Chambers, and Bounegru (2012) are based on the nature of analyses that underpin data stories, hence we refer to both typologies as providing *Analytical types*. Specifically, seven types of stories are described in Kang (2015) including Narrating change over time, Start Big and Drill Down, Start small and zoom out, Highlight Contrast, Explore the intersection, Dissect the factors and Profile the outliers. The Handbook of Data Journalism (Gray, Chambers, and Bounegru 2012) identified 8 types of data stories comprising: measurement or counting, proportion, internal comparison, external comparison, change over time, league table, analysis by category and associations. This typology which is also analysis-centric appears to present some form of order in terms of complexity of the underpinning analysis types; starting with *counting* at the low end and stories based on *league tables and associations* at the high end.

In another article published in the Harvard Business Review (Davenport 2014), a multi-dimensional typology of stories were identified based on four key dimensions including time—is the story about past, present or future; focus—type of question to being answered by the story, what, why or how?; depth—the level of investigation involved in the story; and method—the nature of the statistical analysis. Based on this four dimension, Davenport (2014) identified the following stories types: three time-based types which are *reporting*, *explanatory survey* and *predictions* stories; three focus-based types including *what*, *why* and *how to address this issue* stories; two depth-based stories which include "CSI" and "Eureka" stories and finally two Methods-based stories

TABLE 5
Relating data stories typologies

Type 1—Intent (Slaney 2012)		Type 2—Analytical (Kang 2015)	Type 3—Analytical (Gray, Chambers, and Bounegru 2012)	Type—4 Multidimensional (Davenport 2014)	
Inform	• Refute Claim • Reveal unintended consequences • Reveal anomalies/deficiencies • Track changes in system • Reveal information of interest	• Narrating change over time • Start big and drill down • Start small and Zoom in • Highlight Contrast • Explore intersection • Dissect Factors • Profile outliers	• Measurement or counting • Proportion • Internal comparison • External Comparison • Change over time • League table • Analysis by category • Association	Time	Past—Reporting Present—Explanatory Future—Predictive
Explain	• Enable deeper understanding • Reveal information in increasing details			Focus	What? Why? How?
Persuade				Depth	CSI Eureka
Entertain				Methods	Correlation Causation

including correlation and causation stories. *Given that a number of portfolios studied were about investigative journalism, almost all the story types we identified in Table 2 are in Davenport's terms "depth"-based stories.*

If we consider in particular our intent typology and the two analytical typologies described above as some form of *data story design patterns*, the three typologies could be used complementarily to construct a data story design space as shown in Figure 9. Such design space provides concrete mechanisms and strategies (analytical types) to realise compelling data stories characterised by our intent types. For instance two of analytical types in Kang (2015) can be directly associated with the mechanisms for two of our intent types in Table 3—Narrating change over time (*Track changes in systems*) and Start Big and Drill Down (*Reveal information about an entity in increasing level of details*). Similarly, Gray, Chambers, and Bounegru (2012) analytical types and Davenport (2014) methods could be employed as mechanisms to realise our intent types.

Another major implication of our results is the pre-eminence of the web technologies along with data visualisation and data analysis tools for data storytelling. Based on empirical analysis, we have provided in Figure 10 the technology use pattern across the data journalism projects reviewed. These results show that contemporary and future data journalism practices require a team with computation and programming skills, data analysis skills and data visualisation skills, in addition to traditional journalism skills. Specifically, the availability of database, map visualisation, social media, data preparation and wrangling and data scrapping skills are very important in addition to web development, data visualisation and data analysis for mature data journalism practices. This reinforces the argument expressed by the authors of Berret and Phillips (2016)—"The authors of this report believe that all journalism schools must broaden their curricula to emphasise data and computational practices as foundational skills". It contributes to the "programmer-journalist" discussion in Fink and Anderson (2014), as well as a statement made by Mar Cabra, the head of Data and Research at the International Consortium for Investigative Journalism—the organisation behind the Panama Papers, in a podcast interview with one of the authors—"Universities have to start adding [data journalism] to their curriculum, because if not, they are failing their students".[4]

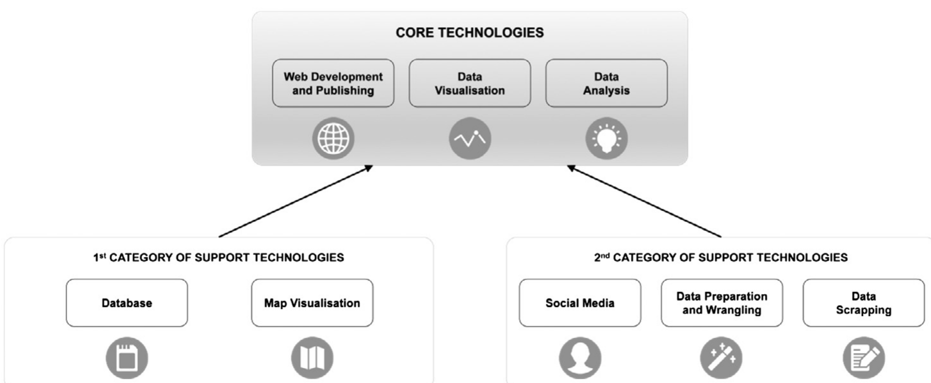

FIGURE 10
What are the core and supporting technologies for Data Storytelling?

We now examine our findings in light of the recently published information on the winning entries for the 2017 GEN awards.[5] By describing the 12 winning entries for 2017 using the same schema employed for our initial 44 cases (2013–2016), we were able to determine if: (1) our data story typology could adequately and consistently classify the 2017 stories, (2) the established technology and tool use pattern from the initial 44 cases is consistent with the technology use patterns in 2017 validation cases and (3) if the interactivity pattern observed in the 2017 validation set is consistent with the pattern observed across the original set of 44 cases. We observed that our seven types adequately characterised the 12 new stories in terms of their purpose. In terms of interactivity, 75 per cent of the 2017 entries were interactive while 33 per cent provided search, filtering and selection features. This is largely consistent with the patterns we obtained from the analysis of the original 44 cases (59 per cent interactive and 27 per cent search, filter and selection). In the area of technology use, web development tools, data analysis and visualisation tools remain the core technologies for 2017. In addition, we notice the increasing adoption of video editing and processing tools in the 2017 winning stories.

We end our discussion with some thoughts on how our findings contribute to the evolving understanding of what constitutes high-quality data journalism. To this end, we consider how our findings may collectively characterise a compelling or high-quality data story. While the DJA director specified a number of criteria for selecting the winning stories (see method section), our finding provide more specific features to characterise what we may call "data story or data journalism ideal types" (Doty and Glick 1994) in terms of intent, interactivity, representation style and use of technology. From our findings, the ideal data story and journalism type will tend to (1) have one or a more of our seven intents, (2) provide a high degree of interactivity or search, filtering and selection features, (3) employ annotated graphs, charts and maps with videos and (4) implemented using web development tools in addition to data and visual analytics tools suitable for presentation over a variety of channels. We nevertheless caution that this is just one of the emerging perspectives on what constitutes good data storytelling or journalism. This perspective is centred on the primacy of journalistic features in data journalism and sees the data, digital or computational aspects as only providing a supporting role in data journalism practice. Other viewpoints are emerging. One such alternative viewpoint s discussed in a recent blog which considered the depth of data analysis involved in the data journalism to be central to quality data journalism practice (Dickinson 2017). Dickinson (2017) appears to view data journalism as a domain in itself though with roots in journalism with data and computational analysis as pivotal elements. We are sure that this debate will continue as scholarly work increase in this area.

Conclusion

Newsrooms across the world and the journalism community have seen a tremendous shift in the ways in which data and algorithms are used in journalistic practices. From the simple representation of information, to complicated data-driven investigations and newsroom tool development, we have seen an ever-growing use of data, algorithm and computational tools in newsrooms in recent years.

While data journalism is maturing, attention is shifting to the qualitative aspects of data stories; specifically on how it impacts on journalism, how it sits within newsrooms norms and best practices, what new skills and toolsets should be employed and in general how the data can improve journalistic work. In search for best practices for data storytelling and as a result high-quality data journalism work, this study analysed 44 cases of award-winning data journalism work, comprising winning entries of the GEN's Data Journalism Award from 2013 to 2016. It further presents a conceptual model using which we uniformly characterised each of the 44 cases to determine types of these stories, the nature of technologies, and skills employed in creating these stories. Regarding our two research questions stated in the method section, our findings show that:

1. Seven types of data stories characterise the 44 winning projects in our study based on their theme or intent. A quarter of our winning stories were categorised as *explaining a Phenomenon for deeper understanding* type, followed by cases related to *revealing anomalies in systems* and *revealing information of personal interest* types.
2. W*eb development and publishing*, *data analysis* and *data visualisation* are the core technologies needed in creating successful data stories. These three categories of technologies appear to be the most important skills for data journalism and newsrooms aiming at successful data journalism work. Database, map visualisation, social media, data wrangling and data scrapping tools are also important supporting tools in these projects.
3. In our opinion, our findings contribute to the growing discussion of what constitutes good data storytelling and journalism. However, we also note that improving our understanding of what constitutes quality data storytelling and journalism requires going beyond the analysis of the secondary data on the data storytelling and journalism projects as we have done in this study. We believe that having direct access to some of the winning teams and interviewing them should offer deeper and significantly richer insight into beliefs, processes and dynamics of creating high quality and compelling data stories. Part of our future work will be in this direction.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

NOTES

1. Global Editors Network—About Us, https://www.globaleditorsnetwork.org/about-us/.
2. The authors exchanged messages in early April 2016 with one of the jurors of the GEN Data Story Awards to obtain the complete list of criteria employed in selecting the winning entries from 2012 to 2016.
3. https://conexp.sourceforge.net/index.html.

4. http://datadrivenjournalism.net/news_and_analysis/from_zero_to_hero_how_data_journalism_helped_establish_the_icij.
5. https://www.datajournalismawards.org/.

REFERENCES

Agarwal, Ritu, and Vasant Dhar. 2014. "Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research." *Information Systems Research* 25 (3): 443–448. doi:10.1287/isre.2014.0546.

Alexander, Stephanie, and Colleen Vetere. 2011. "Telling the Data Story the Right Way." *Healthcare Financial Management* 65 (October): 104–110.

Appelgren, Ester, and Gunnar Nygren. 2014. "Data Journalism in Sweden." *Digital Journalism* 2 (3): 394–405. doi:10.1080/21670811.2014.884344.

Bakker, Piet. 2014. "Mr. Gates Returns." *Journalism Studies* 15 (5): 596–606. doi:10.1080/1461670X.2014.901783.

Baxter, Pamela, and Susan Jack. 2008. "Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers." *The Qualitative Report* 13 (4): 544–559. doi:10.2174/1874434600802010058.

Berret, Charles, and Cheryl Phillips. 2016. "Teaching Data and Computational Journalism." https://bit.ly/cjsteachdata.

Davenport, Thomas. 2014. "10 Kinds of Stories to Tell with Data." *Harvard Business Review*. https://hbr.org/2014/05/10-kinds-of-stories-to-tell-with-data.

De Maeyer, Juliette, Manon Libert, David Domingo, François Heinderyckx, and Florence Le Cam. 2015. "Waiting for Data Journalism." *Digital Journalism* 3 (3): 432–446. doi:10.1080/21670811.2014.976415.

Dickinson, Andy. 2017. "Two Fundamentals That Define Good Data Journalism." *Medium.com*. https://medium.com/@digidickinson/two-fundamentals-that-define-good-data-journalism-fe79170f8566.

Doty, Harold, and William Glick. 1994. "Typologies as a Unique Form of Theory Building : Toward Improved Understanding and Modeling." *Academy of Management Review* 19 (2): 230–251. doi:10.1177/0891243204273076.

Dykes, Brent. 2016. "Data Storytelling: The Essential Data Science Skills Everyone Needs." *Forbes*, March. https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/#6cf5fcf8f0c8.

Fink, Katherine, and C. W. Anderson. 2014. "Data Journalism in the United States." *Journalism Studies* 16 (4): 467–481. doi:10.1080/1461670X.2014.939852.

Fink, Katherine, and C. W. Anderson. 2015. "Data Journalism in the United States." *Journalism Studies* 16 (4): 467–81. Routledge. doi: 10.1080/1461670X.2014.939852.

Flew, Terry, Christina Spurgeon, Anna Daniel, and Adam Swift. 2012. "The Promise of Computational Journalism." *Journalism Practice* 6 (2): 157–171. doi:10.1080/17512786.2011.616655.

GEN (Global Editors Network) 2016. "Global Editors Network—About US." GEN Homepage. https://www.globaleditorsnetwork.org/about-us/.

Godin, Robert, and Rokia Missaoui. 1994. "An Incremental Concept Formation Approach for Learning from Databases." *Theoretical Computer Science* 133 (2): 387–419. doi:10.1016/0304-3975(94)90195-3.

Gray, Jonathan, Lucy Chambers, and Liliana Bounegru. 2012. *Data Stories. Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. Beijing: O'Reilly.

Gynnild, Astrid. 2013. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15 (6): 713–30. doi:10.1177/1464884913486393.

Hewett, Jonathan. 2015. "Learning to Teach Data Journalism: Innovation, Influence and Constraints." *Journalism*, Special Issue Article: 1–19. doi:10.1177/1464884915612681.

Hewett, Jonathan. 2016. "Learning to Teach Data Journalism: Innovation, Influence and Constraints." *Journalism: Theory, Practice & Criticism* 17 (1): 119–137. doi:10.1177/1464884915612681.

Howard, Alexander. 2014. "The Art and Science of Data-Driven Journalism." Tow Centre for Digital Journalism—a Tow/Knight Report. https://towcenter.org/the-art-and-science-of-data-driven-journalism/.

Hsieh, Hsiu-Fang, and Sarah E. Shannon. 2005. "Three Approaches to Qualitative Content Analysis." *Qualitative Health Research* 15 (9): 1277–1288. doi:10.1177/1049732305276687.

Kang, Martha. 2015. "Exploring the 7 Different Types of Data Stories." *Mediashift*. https://mediashift.org/2015/06/exploring-the-7-different-types-of-data-stories/.

Karlsen, Joakim, and Eirik Stavelin. 2014. "Computational Journalism in Norwegian Newsrooms." *Journalism Practice* 8 (1): 34–48. doi:10.1080/17512786.2013.813190.

Kitchin, Rob. 2014. "Conceptualising Data." In *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, 1–26. Sage Publishing. doi: 10.4135/9781473909472.

Knight, Megan. 2015. "Data Journalism in the UK: A Preliminary Analysis of Form and Content." *Journal of Media Practice* 16 (March): 1–18. doi:10.1080/14682753.2015.1015801.

Lee, Bongshin, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. "More than Telling a Story: Transforming Data into Visually Shared Stories." *IEEE Computer Graphics and Applications* 35 (5): 84–90. doi:10.1109/MCG.2015.99.

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. "Big Data: The next Frontier for Innovation, Competition, and Productivity." McKinsey Global Institute, June, 156. doi:10.1080/01443610903114527.

Meyer, Philip. 1973. *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Bloomington and London: Indiana University Press.

Meyer, Philip. 2002. *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Lanham: Rowman & Littlefield.

Müller, Oliver, Iris Junglas, Jan vom Brocke, and Stefan Debortoli. 2016. "Utilizing Big Data Analytics for Information Systems Research: Challenges, Promises and Guidelines." *European Journal of Information Systems* 25: 289–302. doi:10.1057/ejis.2016.2.

Parasie, Sylvain. 2015. "Data-Driven Revelation?" *Digital Journalism* 3 (3): 364–380. doi:10.1080/21670811.2014.976408.

Parasie, Sylvain, and Eric Dagiral. 2013. "Data-Driven Journalism and the Public Good: 'Computer-Assisted-Reporters' and 'Programmer-Journalists' in Chicago." *New Media & Society* 15 (6): 853–871. doi:10.1177/1461444812463345.

Pouchard, Line, Amy Barton, and Lisa Zilinski. 2014. "Data Narratives: Increasing Scholarly Value." In *Proceedings of the ASIST Annual Meeting*. Vol. 51. Seattle, WA, USA. doi:10.1002/meet.2014.14505101088.

Rayfield, J. R. 1972. "What is a Story?" *American Anthropologiest* 74 (5): 1073–1078.

Segel, Edward, and Jeffrey Heer. 2010. "Narrative Visualization: Telling Stories with Data." *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 1139–1148. doi:10.1109/TVCG.2010.179.

Slaney, Malcolm. 2012. "Tell Me a Story." *IEEE Computer* 4–6.

Stikeleather, Jim. 2013. "The Three Elements of Successful Data Visualizations." *Harvard Business Review*. https://hbr.org/2013/04/the-three-elements-of-successf/.

Uskali, Turo, and Heikki Kuutti. 2015. "Models and Streams of Data Journalism." *The Journal of Media Innovations* 2 (1): 77–88. doi:10.5617/jmi.v2i1.882.

Young, Mary Lynn, Alfred Hermida, and Johanna Fulda. 2017. "What Makes for Great Data Journalism?" *Journalism Practice* 1–21. doi:10.1080/17512786.2016.1270171.

Zhao, Qiankun. 2003. "Association Rule Mining : A Survey." Survey Paper, no. 2003116: 1–20.

Adegboyega Ojo (author to whom correspondence should be addressed), Insight Centre for Data Analytics, National University of Ireland (NUIG), Galway, Ireland. E-mail: adegboyega.ojo@insight-centre.org. ORCID http://orcid.org/0000-0003-0565-2115

Bahareh Heravi, School of Information & Comms Studies, University College Dublin (UCD), Dublin, Ireland. ORCID http://orcid.org/0000-0002-0587-1046