Volume 13 Number 9 November 2025

DIGITAL JOURNALISM

Digital Journalism

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

Predictive Data Stories: Characterizing a Nascent Data-Journalistic Genre

Denise Fechner J. & Christian Pentzold

To cite this article: Denise Fechner J. & Christian Pentzold (05 May 2025): Predictive Data Stories: Characterizing a Nascent Data-Journalistic Genre, Digital Journalism, DOI: 10.1080/21670811.2025.2498975

To link to this article: https://doi.org/10.1080/21670811.2025.2498975

Published online: 05 May 2025.

DIGITAL JOURNALISM
https://doi.org/10.1080/21670811.2025.2498975

Routledge
Taylor & Francis Group

ORIGINAL ARTICLE

Predictive Data Stories: Characterizing a Nascent Data-Journalistic Genre

Denise Fechner J.[a] and Christian Pentzold[b]

[a]University of Bremen, Bremen, Germany; [b]Leipzig University, Leipzig, Germany

ABSTRACT

Making sense of the future is a precarious endeavor because there is no such thing as a definitive forecast. For that reason, data-driven news has been welcomed as a source of more reliable outlooks that strengthen journalism's contribution to anticipating future events and setting the public agenda. However, despite large-scale investment into computational and data-rich previsions, little is known about the way these prognoses are told. Based on a sample of cases, we venture into this area of sensemaking and examine the diagrammatic patterns along which data-based projections have been crafted into a journalistic story. In our analysis of this nascent genre of data journalism, we characterize three forms of predictive storytelling: concentration on a single scenario, contrasting different scenarios, and the conjunction of several future scenarios into a prognostic tendency. In all these forms of predictive data stories, more or less conclusive information is arranged into a sequence of events and trends. In most of the resulting stories, issues of probability and uncertainty are not submitted for interactive exploration but become integrated into a directed explanation offered by the news pieces.

KEYWORDS

Storytelling; data journalism; data visualization; data story; predictive analytics; diagrammatic pattern; temporality

Predictive data journalism is a nascent genre of newsmaking. It is part of the far-reaching transformation of journalistic products and practices in the context of digitization and datafication (Anderson 2018; Nguyen 2019). Predictive data journalism emerges at the nexus of, on the one hand, the future-making capacity of the news, that is, to anticipate upcoming events while being an agent of change itself, and, on the other, the ambitions around predictive data analysis that drive the exploitation of large troves of digital information. Predictive data journalism is, Diakopoulos (2022) defined, "the incorporation of predictive information such as forecasts, nowcasts, hindcasts, or other estimates into news production processes … by using or relying on computational modeling techniques".

Yet while predictive data journalism responds to the imminent interest in reducing uncertainty—not only but especially in dealing with the lack of factual information about upcoming developments—it is at odds with journalistic norms of accuracy, verification, and transparency (Zelizer 2021). The uncertainty of predictions, particularly

CONTACT Christian Pentzold christian.pentzold@uni-leipzig.de

data-driven prognostications, renders them a problematic journalistic endeavor since they cannot be proved and their probabilistic calculations can only rely on data of the past from which unknown futures are extrapolated. Still, future-oriented predictive data journalism exists. Think of the dynamically changing infographics being used in the run-up to an election, projections of player performance in sports, or hurricane forecasts. Journalists' reliance on large amounts of data, not only to cover the daily news but to predict future developments, has become particularly evident in the use of predictive models during the coronavirus pandemic (Pentzold, Fechner, and Zuber 2021). Nevertheless, how data-driven newsmaking communicates its future predictions and renders them knowable is little understood.

In general, data journalism has been characterized as a combination of data visualization, programming, and interface design (Gray, Bounegru, and Chambers 2012). Special emphasis is given to storytelling and narrative visualizations so that data and visuals become means to fulfil journalism's key mission of telling stories (Riche et al. 2018). On these terms, number-based stories, or "data stories" as Weber, Engebretsen, and Kennedy (2018) have called them, are "multimodal hybrid artifacts that weave together numbers, words, images and design into a coherent whole" (p. 192). They can be characterized according to their narrative style and level of interactivity, their purpose, forms of representation, data sources, and underlying analytical techniques (Hao et al. 2024; Ojo and Heravi 2018; Stalph and Heravi 2023). Moreover, data stories may, in principle, be separated into an author-driven approach that is more explanatory and aimed at directing readers, and a more reader-driven approach that is exploratory in character and fosters discovery (Segel and Heer 2010).

What these characterizations ignore is the fundamentally temporal orientation of news narratives. Besides tracking a certain sequence of events and meaningful relationships over a period of time, data stories themselves exist on a wider temporal grid between the past, the present, and the future. In principle, they can be either retrospective and use data to look back in time, or they can be prospective and employ quantification to anticipate future events. A multi-dimensional typology of data stories indeed found that future-oriented data stories were one relevant form among many (Davenport 2014).

Zooming on in the genre of predictive data stories, the question is how they address issues of probability and uncertainty with their repertoire of narrative visualization. It is also unclear where forward-looking stories can be located on the continuum between author-driven and reader-driven approaches. In response to these questions, we investigate how calculated futures are told through arrangements of verbal, pictorial, graphic, and other elements. In our definition, predictive data stories are diagrammatical artefacts that represent future-oriented insights generated from an analysis of datasets. We ask: How can predictive data stories be usefully characterized and grouped as forms of a nascent data-journalistic genre? (*RQ1*) That is, what are the diagrammatic patterns through which predictive data stories communicate outlooks and render them meaningful? We further ask: With which means do predictive data stories seek truthfulness? (*RQ2*) Put differently, how do the stories reflect upon and make transparent the uncertainty of outlooks? And we ask: How do predictive data stories balance story control with the freedom of readers? (*RQ3*) In other words, are they following an author-driven approach, a reader-driven approach, or a form of interactivity and reader engagement in-between? The overall aim of this study is to gain a better understanding of data journalism's capacity to make sense of the future by interlacing numerical information, chronology, and causal relations into narrative accounts.

Literature Review: Predictive Data Stories

Journalism does not merely present the latest news and discuss what has happened or is about to happen—it also envisages the future and what will happen next (Neiger and Tenenboim-Weinblatt 2016). Predictive data journalism is fueled by this general interest in forward-looking newsmaking. It promises to complement traditional forms of forecasting based on expert opinion, past experience, or speculation with probabilistic models (Pentzold and Fechner 2021). It is on this note that Maycotte (2015) has stated that "by using available data, journalists will be able to orchestrate predictions and write tomorrow's headlines and stories accordingly." Hence, predictive data journalism is ascribed the potential to forecast and understand the development of complex processes, to inspire counterfactual thinking, and foresee upcoming trends (Diakopoulos 2022). In this capacity, it has for instance contributed to combat COVID-19 and explain the necessity of countermeasures (Pentzold, Fechner, and Zuber 2021; Fu and Stasko 2024).

Aspirations to use vast amounts of data to make more reliable forecasts have deep historical roots in fields like political microtargeting, retail, and preemptive policing (Pietruska 2017). They culminate in predictive analytics, where data are used "to see and then manipulate possible futures" (Beer 2019, p. 29). What predictive analytics generate, however, is not one particular and more certain outlook but a plethora of probable futures. They "articulate a range of possible outcomes," as Silver (2015, p. 61) has asserted. Although these futures are perceived to offer a more future-proof form of anticipation than the human imagination, they are probabilistic in character and cannot determine the future. Capturing this probabilistic nature of data-driven models and the variety of more or less likely previsions is both a key challenge and a major asset of predictive data journalism. On the one hand, it has to reckon with the difficulties of not being able to supply one definitive prognosis and therefore lack the accuracy traditionally sought by journalists (Diakopoulos 2022; Zelizer 2021). On the other hand, it is exactly this sort of information that sets it apart from other kinds of prognostication.

Using Data to Tell Stories

Data journalism is commonly defined as a form of journalistic storytelling that predicates on the analysis of datasets (Anderson and Borges-Rey 2019; Morini 2025). Thus, data journalism is about telling data stories—"artefacts for revealing and communicating insights gained from the analysis of data-sets" (Ojo and Heravi 2018, p. 693).

Not all data journalistic products need to be narratives, tough. Moreover, what constitutes a data story is open to debate. Leaning on Genette's narrative theory, Weber (2020) stressed that a story "is a sequence of events or happenings that are temporally structured and coherently related to each other, involving one or more characters or anthropomorphic agents or objects" (p. 297). A narrative, in turn, is the "textual, visual, or multimodal representation that presents a story" (p. 297). In a similar vein, Weber, Engebretsen, and Kennedy (2018) used the notion of a "narrative" and cited Bell (1991), who defined a narrative as "a sequence of events that are temporally structured and coherently related to each other with bonds of (strong or weak) causality" (Weber, Engebretsen, and Kennedy 2018, p. 194). Irrespective of such definitions, they have likewise noted the less strict usage of the term among journalists, who employ the terms "stories" and "news" interchangeably and associate practices of describing, explaining, and arguing with both. In response to the loose journalistic understanding of what a story is, Riche et al. (2018) have maintained that data stories "are based on a set of events from a narrative, but they may adjust the presentation by changing the order, shortening the length, adding extra context, etc." (p. 8). However, the data story is not bound to a strict sequential format or plot template (Heravi 2022). The fact that data stories are not necessarily adhering to a strict narrative formula has been demonstrated by Kosara (2017), who showed that many of them do not provide a clear-cut story arc. Only a portion of data stories are linear, narrative-based stories (Heravi 2022).

Supposedly, the flexibility and selectivity of narrative elements and causal connections in data stories become all the more pertinent with regard to the multiple possible outcomes of probabilistic predictions, of which only few can be covered by a story. Probabilistic predictions generate a great number of possible events that make it difficult to develop one coherent sequence that could form the backbone of a data story without unduly curtailing many other possible future paths. Engaging with this challenge, we assume to find two strategies of storying the data: a hypothesis-driven path that starts out with a question and then processes data, and a data-driven path that sets out by exploring data without a guiding assumption (Parasie 2022). Similarly, Borges-Rey (2020) has differentiated between data as story input used to test hypotheses and data as story output resulting from free exploration.

Data stories can, we assume, rest on both procedures as long as they gather, analyze, and visualize data in ways that create a comprehensible, though not necessarily comprehensive, story that enables readers to make sense of one possible future or more than one.

Thus, the first research question is: How can predictive data stories be usefully characterized and grouped as forms of a nascent data-journalistic genre? (*RQ1*)

Truthful Narrative Visualizations

Data stories conventionally employ visualizations. Thus, while images and illustrations have long been key to journalism writ-large, in data journalism visualizations have in particular transformed storytelling "from writing as the main semiotic mode to coding and visualizing as pivotal elements of digital storytelling" (Weber, Engebretsen, and Kennedy 2018, p. 191). Consequently, visualizations have come to be seen as the defining element of data journalism proper (Loosen, Reimer, and De Silva-Schmidt 2020; Young, Hermida, and Fulda 2018). Considering predictive data journalism only, Diakopoulos (2022) even determined that it almost always includes data visualizations. When forming part of data stories, narrative visualizations are deemed a means to enhance user engagement and comprehension (Borges, Correa, and Silveira 2022;

Hullman and Diakopoulos 2011; Fu and Stasko 2024). Visualizations seem to be especially helpful for communicating probabilistic predictions and convey uncertainty (Pentzold and Fechner 2020; Burgio 2019; Diakopoulos 2022; Spiegelhalter, Pearson, & Scott, 2011).

Narrative visualizations can take different shapes which Segel and Heer (2010) have classified as magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and film/video/animation. Regardless of their more or less sophisticated design, they fulfil not only an aesthetic but an epistemological function. They are, Drucker (2014) has explained, diagrammatic in character because they rest on a meaning-producing graphical organization of images, static charts, tables, maps, or infographics that interlace visuals, numeric information, and texts into a multimodal news offering. "Diagrammatic images spatialize relations in a meaningful way," (p. 66) she wrote. They do so according to semiotic conventions that guide the encoding of information into patterns of hierarchy, proximity, or juxtaposition (Stalph and Heravi 2023).

Whilst visualizations can be treated as diagrammatic arguments, can they also transport a narrative? Stressing their narrative capacity, Bounegru et al. (2017) have assumed that visualizations in fact encapsulate the "potential to possess narrativity" (p. 701). Telling a predictive data story in diagrammatic form would therefore imply the use of visual resources of all kinds next to written accounts, as well as their composition into a meaningful narrative (Aiello 2020). Employed as storytelling devices, data visualizations "integrate diagrammatic potentials of certain components" (Stalph and Heravi 2023, p. 1645).

Nevertheless, while visual tools such as graphs, charts, or illustrations make it easier to represent uncertainty, their primarily associative display can assumedly interfere with the linearity of a story. In response to this challenge, journalists have explored different possibilities of data-driven visual storytelling regardless of a news piece's prime temporal orientation. They have experimented with narrative constituents such as the presence of a narrator, sequentiality and change, as well as tellability, that is, the point of a story being made explicit in a visualization (Weber 2020). We assume to find evidence of such self-reflexivity in predictive data stories. Because of the lack of verifiable information and the indeterminacy of the future, emphasis is placed on communicating truthfully (Cairo 2016).

For Tufte (2001), seeking diagrammatical means in order to be honest about the lack of certainty constitutes "graphical excellence" (p. 53). In a survey of predictive data journalism, Diakopoulos (2022) found that many pieces included transparency information so to address the uncertainty about both the predictions and the models and data used to calculate them. What is unclear, however, is how predictive data stories achieve transparency and how they seek to communicate truthfulness through narrative visualizations. In other words, how they "explore the ways in which data that is not meaningful in statistical terms can become meaningful in semiotic terms" (Burgio 2019, p. 230).

A means to do so is the "meta-story" (Weber, Engebretsen, and Kennedy 2018, p. 201). With that, stories are not only enriched by data used for verification or that stem from a data-based investigation—they can also provide meaning behind data (Rogers, Schwabish, and Bowers 2017). A meta-story may only state the data source briefly, or it might expand on the methodology and the decisions made throughout the analysis. It could use visuals, numbers, text, or combinations thereof. The meta-story can explore how the methodology and decisions made come to impact on the predictive outcome. That way, including the meta-story in a data journalistic piece supposedly serves to enhance transparency, thus strengthening the credibility of a prediction, creating awareness of the inherent uncertainty of the data analysis, and demonstrating the influence of analytical decisions on a result (Weber, Engebretsen, and Kennedy 2018, p. 202).

So the second research question is: With which means do predictive data stories seek truthfulness? (*RQ2*)

Engaging Data Stories

Data story typologies usually base on the purpose and intent of a story. In a simple form, Slaney (2012) defined four types of data stories that either inform, explain, persuade, or entertain. By the same token, Ojo and Heravi (2018) derived seven types that either 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, or reveal information about an entity in increasing levels of details. With a hint towards the complexity of the underlying analytical procedure, Gray, Bounegru, and Chambers (2012) named eight types of data stories that rested on measurement, counting, proportion, internal comparison, external comparison, change over time, league table, analysis by category and associations. The question is to what extent predictive data stories fit with these typologies and if the typologies satisfactorily capture their form, analytical base, and purpose.

More generally, Segel and Heer (2010) have identified two core strategies for mapping data stories. Data stories can either follow a more author-driven approach or a more reader-driven approach. While the author-driven approach refers to a linear structure that above all enables authors to tell a story with the data, the reader-driven approach forms a nonlinear structure containing more interactive features that allows readers to explore the data story more freely. In this case, the data story that affords the exploration of the data represents "a tool for readers to analyze what's being presented to them" (Cairo 2013, p. 73). In practitioners' terms, the options therefore either tilt toward an exploratory—namely reader-driven—or an explanatory—namely author-driven—stance (González Veira 2017).

The author-driven approach and the reader-driven approach are not binary choices but represent two ends of a continuum. Journalists who strive to provide data stories are hence not pressed to choose one of the two options; they can also seek to combine elements of both, depending on the presumed purpose of the data story. Thus, Segel and Heer (2010) delineated three hybrid approaches. The Martini-Glass structure starts with an author-driven approach so to propose a narrative and then moves to a reader-driven stage where readers can explore the data. Interactive slideshows enable interactivity within the boundaries of a unit and then move on to the next stage of the story. With Drilldown, a general theme is presented from which readers can follow more specific themes. There are more storytelling structures in newsmaking beyond these three. In an analysis of a broad set of data journalistic pieces, Heravi (2022) found that the most common form was a combination of inverted pyramid and stack of blocks, the 'Water tower': the data stories began with key information and moved on to parts that were each equally important.

As for predictive data stories, we assume that the unavoidable issues of uncertainty and the multiplicity of possible futures suggest a reader-driven, exploratory, and more interactive approach. "Understanding and retention are promoted by interactive graphics because the user is encouraged to engage with the content actively rather than passively, which can also help to counteract differences in numeracy," as Spiegelhalter, Pearson, and Short (2011, p. 1399) thus noted. Letting readers explore different possibilities can eventually foster a higher engagement with the news (Boy, Detienne, and Fekete 2015; Lewis and Westlund 2015). Ideally, "the satisfying story is a means of discovery," Matei and Hunter (2021, p. 10) have postulated. They added: "A compelling story sparks public imagination and draws people to think. … Further, data stories, at their best, push the audience from simply believing, to knowing with a degree of confidence" (p. 10). Moreover, a reader-driven interactive approach might also help to communicate the uncertainty involved in forecasting by presenting possible scenarios that have a different likelihood of occurring in the future (Pentzold and Fechner 2020).

Therefore, the third research question is: How do predictive data stories balance story control with the freedom of readers? (*RQ3*)

Data and Method: Looking Beyond the Pinnacle of Data Journalism

Because many studies limit their analysis to award-winning data journalistic projects, they single out a minority of available stories that are not representative of the average data-journalistic output (Diakopoulos 2022; Loosen, Reimer, and De Silva-Schmidt 2020; Ojo and Heravi 2018; Stalph 2018; Stalph and Heravi 2023; Young, Hermida, and Fulda 2018). Yet even some of the studies that looked at examples of journalistic excellence only discovered limited interactive features (Appelgren 2018; Kennedy, Weber, and Engebretsen 2020; Weber, Engebretsen, and Kennedy 2018), a finding that was confirmed by studies of data journalism as part of daily news reporting whose diagrammatic displays are also usually little interactive and make do with tables, photographs, and static charts (Borges-Rey 2016; Hao et al. 2024; Heravi 2022; Zamith 2019).

In order to enlarge our scope and include material beyond the pinnacle of data journalism awards, our data collection was based on a comprehensive sample that was taken from 70 different sources and completed in 2020. Besides the longlists of award schemes found on *datajournalismawards.org*, these included directories of data journalism and data visualization projects like *The Pudding* and Nate Silver's *FiveThirtyEight*, websites and blogs listing data journalism projects (*Datasketch.es, Datenjournalist.de, The Functional Art*, *Flowing Data*) and data journalism sections of online daily newspapers from the United States (including, for example, the New York Times's *Upshot*), the United Kingdom (encompassing, among others, the Guardian's *datablog*), and Germany (containing, for instance, Der Spiegel with its *SPIEGEL data* section). We included only news pieces that dealt with the future in their title, sub-title, or the first paragraph. (see the full list here: https://osf.io/5hk8f/?view_only=1e42c47fadae429f97c0c5c9e0304f21).

The procedure yielded a small set of news items ($N = 150$) compared to the vast number of data journalistic pieces published in the same time frame where we found 2,644 items in total. Some of them appeared in several places and were only included once in their original version if there was one. With the study's focus on predictive data stories, a restriction to award-winning contributions would assumedly have yielded an even more minuscule sample. The analysis was not comparative in character because previous research has noted strong cross-country influences and a project-based collaboration of data journalistic teams that involves fluctuation and freelance work (Pentzold and Fechner 2020; Heft 2021). Our selection of material from the United States, United Kingdom, and Germany hence did not aim to identify national patterns but followed practical concerns of linguistic and cultural access to understand and contextualize the news pieces. This selection also included prominent journalistic powerhouses.

According to our definition, predictive data journalism projects seek to envision future developments, use predictive models, and delineate possible events in projective scenarios. We identified their forward-looking perspective based on their headline or subtitle as well as their inclusion of grammatical markers of a future tense. Following Hansen (2016), we looked for a future orientation marked by time auxiliaries ("will," "won't," "shall," "is going to") and time adverbials (e.g., "tomorrow," "later," "then," "when," "next," "soon"). Because temporal deixis can also be indexical, like with "already," "after," or "before," no definite dictionary could be employed. Consequently, we also had to consider the general temporal direction of a piece, that is, whether it included references to future developments or events, even if this was not indicated by the grammatical tense (Neiger and Tenenboim-Weinblatt 2016).

The final selection of cases was downloaded and stored using the free Firefox browser extension SavePageWE, which saves files in PDF. In order to capture interactive features, we also used a screen recorder addon. The files were imported to the qualitative data analysis software MAXQDA that helped to organize the codings. Since many cases used animations and dynamic objects, the analytical process moved back and forth between the files and the websites.

A total of 5.6% of all cases in our sample were future-oriented; this was striking and considerably lower than the ratio found in legacy newspapers. For instance, in U.S. and Israeli outlets, 43.4% of the headlines had a projective outlook, as found by Tenenboim-Weinblatt and Neiger (2015). Similar to their approach, we identified the temporal orientation of a news item with grammatical markers and references to potentially forthcoming events. Yet whilst selection criteria cannot help to explain the disparity, there may be more than one reason why predictive data journalism is still a niche pursuit: Newsmakers may be reluctant to adopt the kind of proactive, interventionist position that is required for making data-based forecasts (Hanitzsch 2007; Zelizer 2017). Other factors may involve the complex models and calculations that are necessary for generating data-based predictions. Their translation into diagrammatic displays is far from trivial and often requires bespoke solutions which are both costly and time-consuming to produce (Pentzold and Fechner 2021).

To identify the diagrammatic patterns of predictive data stories, we carried out an inductive conceptualization following a Grounded Theory-based coding methodology. The coding considered predictive data stories holistically, analyzing their diagrammatic composition of both textual and visual elements in combination with narrative and temporal aspects.

The initial set of codes consisted of heuristic codes taken from the available literature (see Table 1). In this first phase, each predictive data story was coded according to the following aspects. The code *data set or data source* was used to register the number of data sets, the sources and sorts of data found in a news piece, and the patterns of referring to the data sets (Hao et al. 2024; Stalph and Heravi 2023). *Text-image relations* were identified including lead and patterns of visualization captions (Hao et al. 2024; Weber, Engebretsen, and Kennedy 2018). *Visualization types* were grouped into categorical (bar charts, bubble chart, etc.), hierarchical (pie chart, Venn, Dendrogram, etc.), correlations (scatter plot chart, heat map, etc.), temporal (line chart, Gantt, etc.), spatial (Choropleth chart, Dot map, Flow map, etc.), and mixed types (Stalph and Heravi 2023). *Story types*

Table 1. Coding categories.

Categories	Description
Data set and data source (Hao et al. 2024; Stalph and Heravi 2023)	Number of data sets Sources of data sets (government; NGO/NPO; enterprise/company; non-university research institute/center; university/academia; other media outlet; own data collection; n.a.) Sorts of data (quantitative; qualitative; mixed) Data source patterns (direct link; data source mentioned)
Text-image relation (Hao et al. 2024; Weber, Engebretsen, and Kennedy 2018)	Relation between text and visualizations (stand-alone graphic; text alternating with graphics) Lead (visualization(s) as the leading element(s) accompanied by text or not; text as the leading element accompanied by visualization(s)) Visualization caption patterns (mentioning the data source; visualization attribution; scope of data collection; describing the content of the visualization; additional information)
Visualization types (Stalph and Heravi 2023)	Visualization types (categorical; hierarchical; correlations; temporal; spatial, other)
Story type (Hao et al. 2024; Heravi 2022; Ojo and Heravi 2018; Segel and Heer 2010; Weber, Engebretsen, and Kennedy 2018)	Story structure (linear predefined structure; open and explorative structure; linear-nonlinear structure) Narrative visualization (Martini-Glass; Interactive slideshow; Drilldown) Story convention (inverted pyramid; dramatic arc; kabob; stack of blocks; water tower; hybrid, other) Data story type (quick update; briefing; chart description; investigation; in-depth investigation) Story purpose (inform; explain; persuade; comfort; terrorize; entertain)
Interactivity (Diakopoulos 2022; Hao et al. 2024; Anderson and Borges-Rey 2019; Stalph and Heravi 2023)	Interactivity level (no interactive features; multiple predefined variations of multimedia components; user manipulation through scaling, rotating, navigating or choosing perspectives; graphic generated through user input/query; comprehensive toolset offered to build component; feedback based on intelligent programming adding a meaningful layer to symbolic content) Visualization interaction patterns (inspecting; zooming; exploring; scrollytelling; other)
Future scenarios or trends (Grewal, Goodwin, and Dwyer 2021)	Discreteness continuum (high degree of discreteness (pinpoint the most likely outcome), low degree of discreteness (multiple outcomes))
Meta-story (Diakopoulos 2022; Rogers, Schwabish, and Bowers 2017; Weber, Engebretsen, and Kennedy 2018)	Story–data relation (stories enriched by data; stories using data to investigate; stories explaining data) Transparency level (presentation of the editorial processes behind the data story; explanation of the editorial processes behind the data story) Uncertainty dimension (about prediction; about data; about model, none) Uncertainty information (quantifiable information; no quantifiable information)

were classified with respect to story structure, narrative visualization, story convention, data story type, and story purpose (Hao et al. 2024; Heravi 2022; Ojo and Heravi 2018; Segel and Heer 2010; Weber, Engebretsen, and Kennedy 2018). We also coded for *interactivity* levels and visualization interaction patterns like inspecting, zooming, exploring, and scrollytelling (Anderson and Borges-Rey 2019; Diakopoulos 2022; Hao et al. 2024; Stalph and Heravi 2023). The *future scenarios or trends* offered were located on a discreteness continuum (Grewal, Goodwin, and Dwyer 2021). The *meta-story* was characterized according to the story–data relation (stories enriched by data; stories using data to investigate; stories explaining data), the transparency level, the uncertainty dimension (about prediction, about data, or about model), and the existence of quantifiable or not quantifiable uncertainty information (Diakopoulos 2022; Rogers, Schwabish, and Bowers 2017; Weber, Engebretsen, and Kennedy 2018).

In this first phase, coding was done by both authors working separately. In this process, the initial set of heuristic codes deducted from the literature was inductively elaborated and refined in accord with the empirical material. Regular team meetings were held to discuss all developed and adapted codes, a process that served to confirm congruent codings and clear out divergent readings. The phase was completed once all material was taken into consideration, no new codes could be formulated, and all codings were harmonized. As this was a piecemeal and reflexive process of interpretation, no inter-coder reliability measures could be calculated.

In a second phase, we looked for characteristic combinations of the codes within the data stories. This process of comparison and conceptualization was aimed at characterizing and grouping predictive data stories. It accommodated the differences and similarities in terms of diagrammatic display and its meaningful relationship with story elements and temporal references. In this interpretative step, we worked in a team of two with regular meetings that served to validate tentative proposals about the most salient features and cross-case connections which would justify formulating a form vis-à-vis another.

In this stepwise procedure, three different forms of predictive data stories emerged. All pieces we sampled were structured along a future-oriented timeline of events and were thus stories. Most notably, these stories differed in the number of futures they projected and how they were arranged into one ore multiple lines of events. We call these three forms (1) *concentration*, (2) *contrast*, and (3) *conjunction*. Of the 150 pieces studied, 59 belong to the first category, 34 to the second, and 57 to the third. In what follows, we introduce each of them and illustrate them with examples from our corpus.

Overview: Forms of Predictive Data Stories

Predictive data stories combine visual elements and textual elements into a coherent narrative. In all of the examined news items, journalists unfolded their stories and exposed causal relations and possible future outcomes by the use of narrative elements such as actors, motives, expectations, actions, histories, and proclivities set on a future-oriented path. Guided by a forward-looking perspective, predictions offered a hypothesis about the future that placed the narrative elements into a logical sequence so to bring temporal order and progress to the diagrammatic argument. The data-based assumptions about the future hence rested on cause-and-effect relationships that were told through textual and visual elements composed into a meaningful narrative (Aiello 2020).

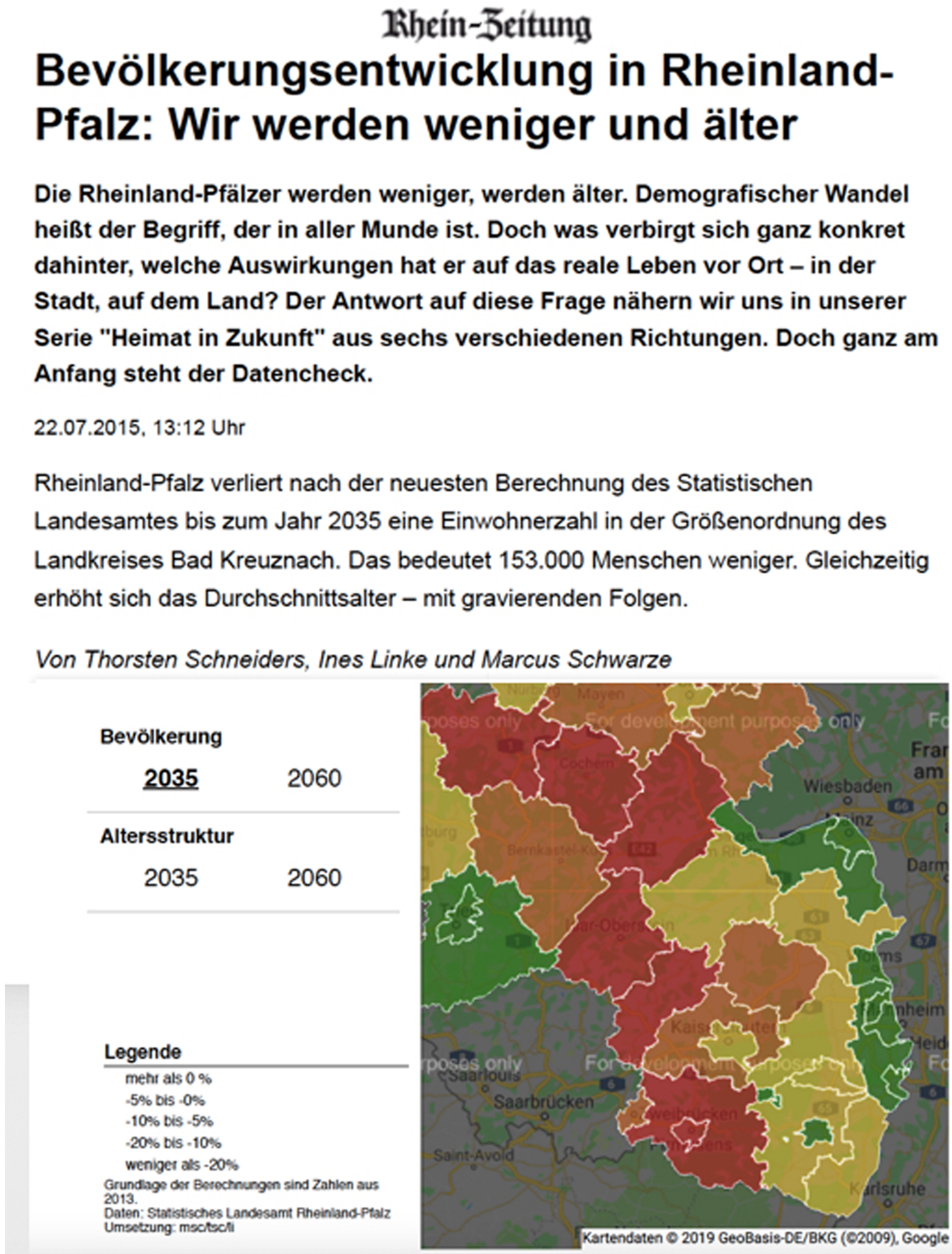
In response to *RQ1*, the three forms of predictive data stories which we derived from the material and grouped together are characterized. In the following, we start with concentration-style stories before discussing stories told by way of contrast and conjunction, respectively. *RQ2* and *RQ3* will be considered in the next section.

Form 1—Concentration

The first form of predictive data story focuses on a single future forecast or scenario in its narrative and multimodal presentation. It employs a linear structure, following one prognostic version and omitting alternative trajectories, maintaining an author-driven approach. While readers can adjust certain details, this format offers little space for exploring or comparing different developments. The story focuses on one trend, sidelining alternative predictions and other possible future paths influenced by changing factors. Visualization techniques, such as "scrollytelling," give an interactive feel but limit the readers' freedom to explore, keeping control with the authors. The meta-story, presented at the end, offers additional information but does not play a dominant role.

A subset of these stories allows readers to adjust visualizations, but the authors' interpretation remains dominant. The linear narrative starts with a strong authorial position and then opens slightly for exploration, resembling the Martini-Glass structure described by Segel and Heer (2010), where a broad opening sits on a narrow stem. Weber, Engebretsen, and Kennedy (2018) describe this as a "hybridization process" (p. 195) between author-driven and reader-driven approaches, though these stories remain primarily author-driven, with limited interactive features bound to the main storyline.

This form is exemplified by a news piece published in the German newspaper Rhein-Zeitung (Figure 1). The predictive data story there analyzed demographic trends in one German federal state and predicted that the population in the region would shrink and be older on average. The aim of the predictive data story was to explain the impact of a smaller and predominantly geriatric population. The piece zeroed in on that particular scenario and specified its ramifications for the people living in the area. Its data visualization showed one concrete trend in different cities and locations, while the accompanying text encouraged readers to see for themselves what the future would look like in their home region. While the map shown in the piece was interactive, it was not responsive in the sense that it allowed readers to explore the data, change the future outlook, and engage with different possible future scenarios. Instead, the interactive element simply served the purpose of readability, as it allowed readers to zoom in on details. The readers could not otherwise manipulate the visualization or trajectory of the story. The visualization summed up the essence of the predictive data story and expanded a general trend projected for a single federal state into a detailed view of the future presumably valid for other regions.



Rhein-Zeitung

Bevölkerungsentwicklung in Rheinland-Pfalz: Wir werden weniger und älter

Die Rheinland-Pfälzer werden weniger, werden älter. Demografischer Wandel heißt der Begriff, der in aller Munde ist. Doch was verbirgt sich ganz konkret dahinter, welche Auswirkungen hat er auf das reale Leben vor Ort – in der Stadt, auf dem Land? Der Antwort auf diese Frage nähern wir uns in unserer Serie "Heimat in Zukunft" aus sechs verschiedenen Richtungen. Doch ganz am Anfang steht der Datencheck.

22.07.2015, 13:12 Uhr

Rheinland-Pfalz verliert nach der neuesten Berechnung des Statistischen Landesamtes bis zum Jahr 2035 eine Einwohnerzahl in der Größenordnung des Landkreises Bad Kreuznach. Das bedeutet 153.000 Menschen weniger. Gleichzeitig erhöht sich das Durchschnittsalter – mit gravierenden Folgen.

Von Thorsten Schneiders, Ines Linke und Marcus Schwarze

Figure 1. Example of predictive data story for form 1–*concentration*. Source: rhein-zeitung.de.

In simpler form, a story in the New York Times (Figure 2) provided a simulation of the effects if working-age men without advanced education would own a college degree, and would start earning wages typical for college graduates. The scenario was fixed and readers had no opportunity to tinker with different routes. The same pattern is at the base of a Guardian story (Figure 3) that offers one estimate on how the world's population might grow until 2100.

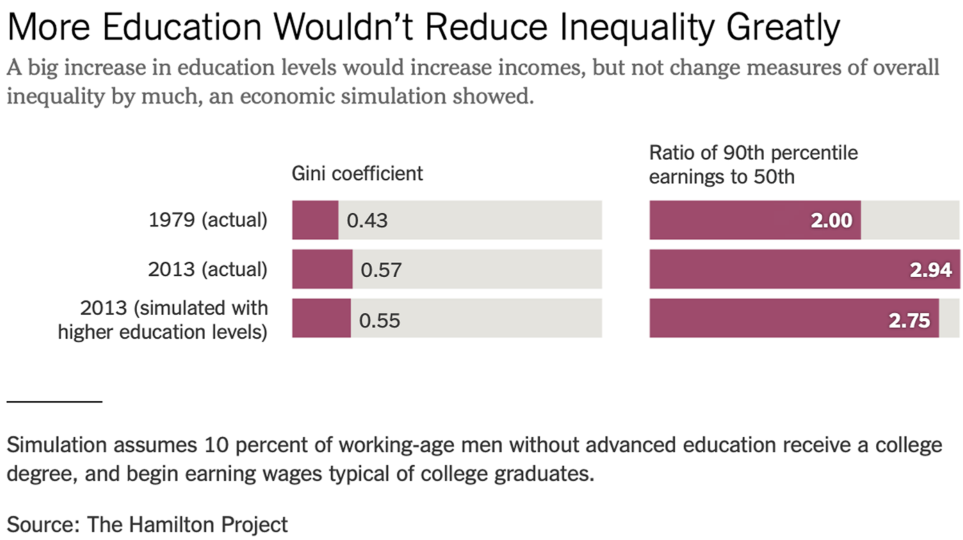


Figure 2. Example of predictive data story for form 1–*concentration*. Source: nytimes.com.

Form 2—Contrast

The second form features a predictive data story presenting two future scenarios. These projections highlight different or diverging trends for a topic, aiming to foster deeper understanding by contrasting outlooks. This form is typically used to explain underlying factors and possible correlations. By juxtaposing futures, it illustrates the uncertainty of predictions and the range of possibilities. It often relies on multiple visualizations to clarify background assumptions, outline scenarios, and suggest lessons. Scrollytelling features are the main exception, with stories using graphical and textual elements chronologically to detail alternative scenarios and address uncertainties. This combination allows for a clear narrative structure, where authors still control the storytelling, guiding readers through a narrow set of options. The low interactivity of scrollytelling limits the readers’ ability to alter the course of events, reinforcing the contrast-style stories’ bi-linear character.

This author-driven approach lets journalists balance different scenarios comprehensively, focusing on causal relations that determine predicted outcomes, shifting attention from a single future to a comparative view of two plausible predictions. The form employs a linear structure and visualizations to communicate scenario selection but does not present all possible outcomes, instead emphasizing key conditions and their implications. The meta-story documents data sources, analytical decisions, and the predictive model, integrating these into the narrative to explain causalities and factors impacting the prognoses.

To illustrate this form, a notable set of examples are news pieces that contrasted different COVID-19 scenarios (Author 2021). Take, for instance, the story “Ausbreitung von Covid-19. Jeder Tag zählt” released by Der Spiegel (Figure 4) that juxtaposed two different waves, one fully colored visualizing the potential spread without any countermeasures, and another version with dashed lines that showed a development when distancing measures would be implemented. The kind of contrastive waves representing two scenarios has not been pioneered by Der Spiegel but was indeed widespread

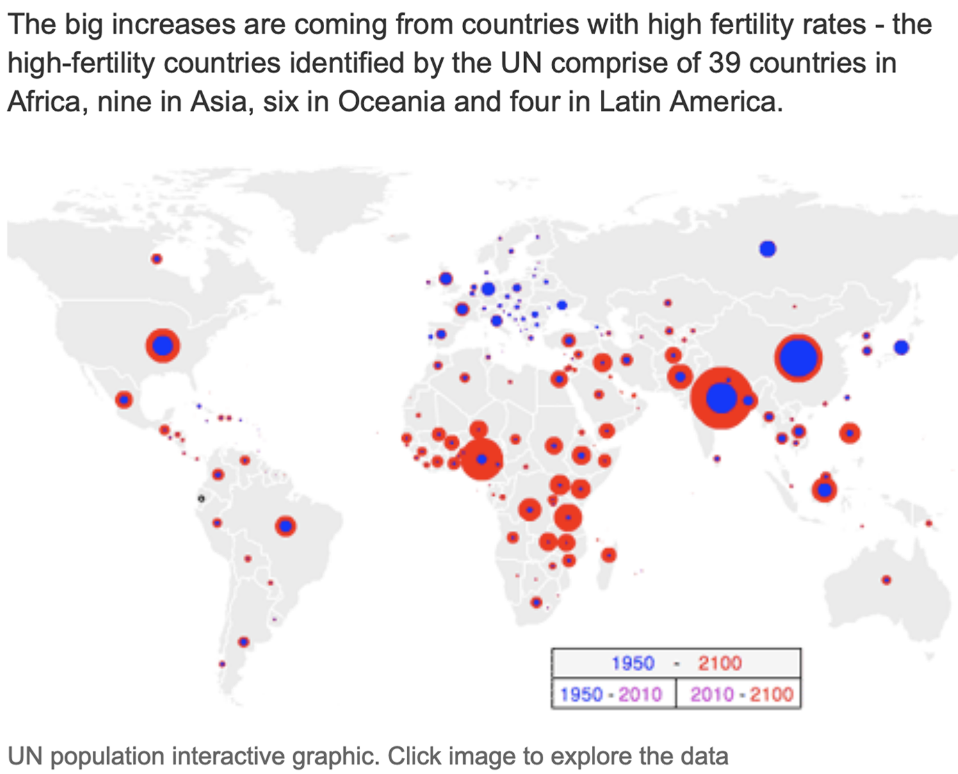


Figure 3. Example of predictive data story for form 1–*concentration*. Source: theguardian.com.

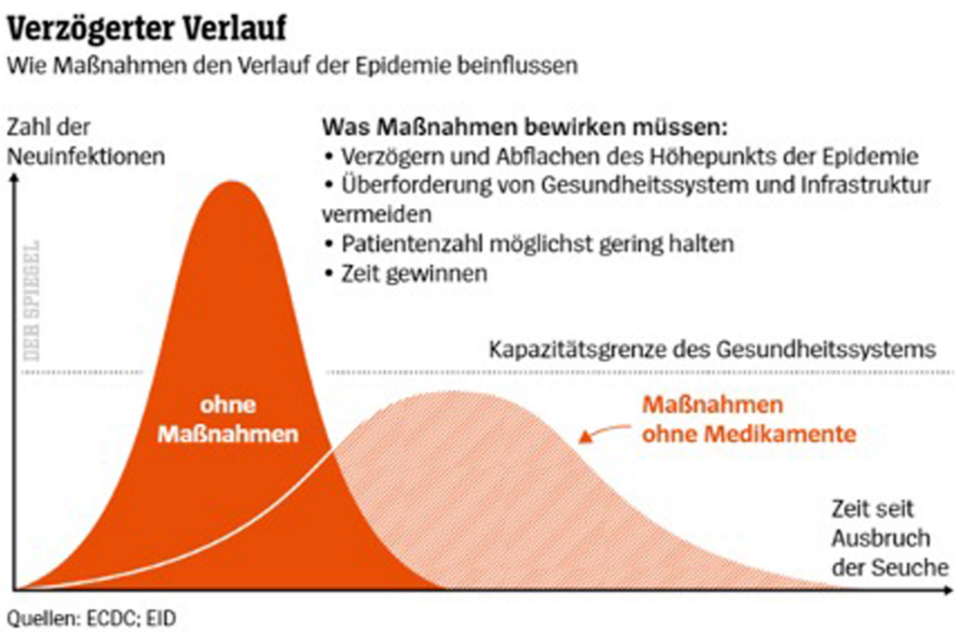


Figure 4. Example of predictive data story for form 2–*contrast*. Source: spiegel.de.

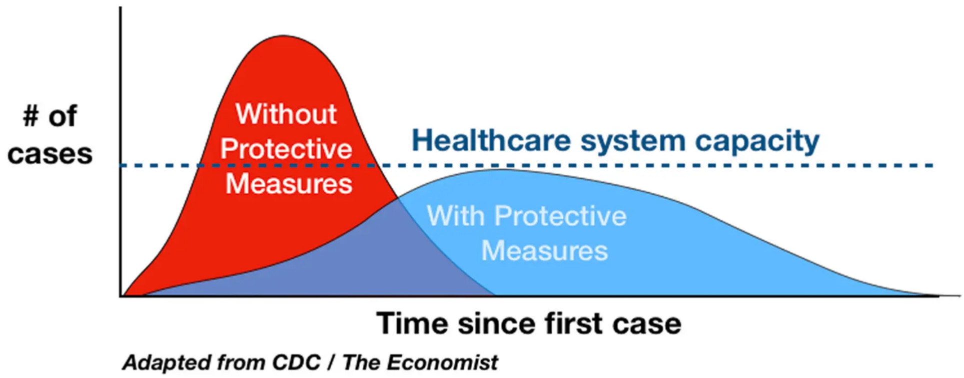


Figure 5. Example of predictive data story for form 2–*contrast*. Source: nytimes.com.

and became almost iconic in the first months of the Coronavirus pandemic. The template spawned many renditions, for instance, in the New York Times (Figure 5).

Form 3—Conjunction

The third form of predictive data story presents several future scenarios. Unlike contrasting stories that pit two alternatives against each other, conjunctions compile forecasts supporting a single overarching trend. They are multilinear, with the scenarios differing only slightly, all pointing toward the same preferable or probable future. Visualizations emphasize both the uncertainty of the trend and the robustness of the forecast, synthesized from multiple prognoses. These scenarios, while varied, reinforce the main argument of the story, conveying stability and certainty about the future by combining similar projections.

Conjunction-style stories, being less selective than the first two forms, tend to present more predictions and use more visualizations to illustrate convergence. The narrative develops alongside these visualizations, exploring and explaining a trend with strong projective overlap. Projections may stem from different data sets or are calculated from a single set modeling multiple outlooks. The use of diverse data often necessitates corresponding visualizations.

Like other predictive data stories, conjunctions are author-driven, limiting reader interactivity. Story control remains with the authors, and interactive features, when present, typically function through filtering options that aid interpretation rather than exploration of the data or models. These stories also include elaborate meta-stories providing background on the forecasts and the predictive models. By discussing the decisions leading to different predictions, this form connects forecasts with the aim to improve the understanding of causal relationships and the analytical work shaping the projections.

An example of this form of predictive data story comes from *The Pudding* and makes predictions on the time the population of cetaceans in captivity in the U.S. will likely by extinct (Figure 6). Readers were given options to adjust the outlook along an assumed maximum age of the population and the moment breeding would be stopped. Another elaborate forecast was, for example, offered by Nate Silver's project *FiveThirtyEight* on the 2022 U.S. senate elections ("Forecasting the race for the

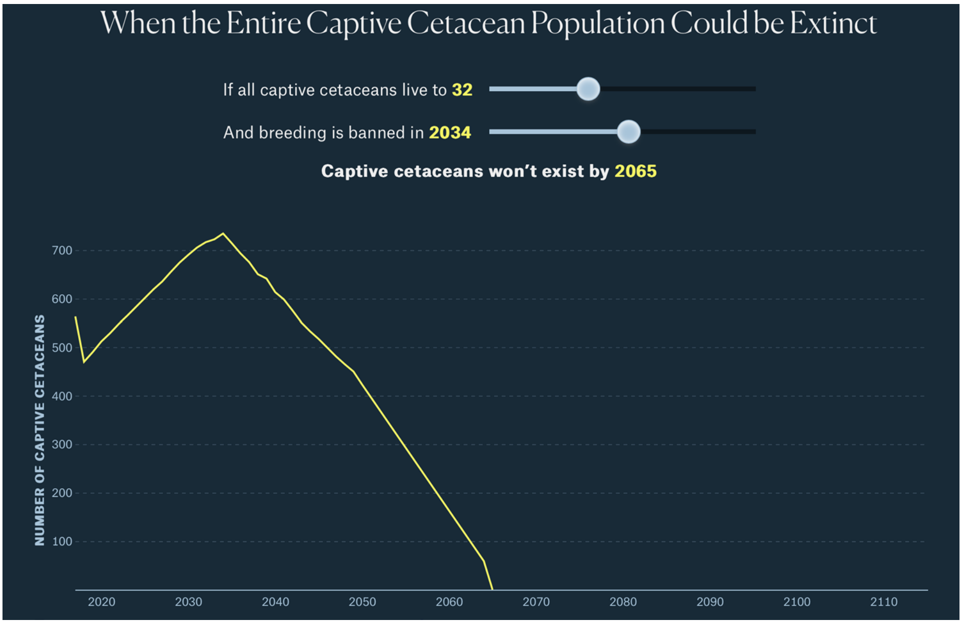


Figure 6. Example of predictive data story for form 3–*conjunction*. Source: pudding.cool.

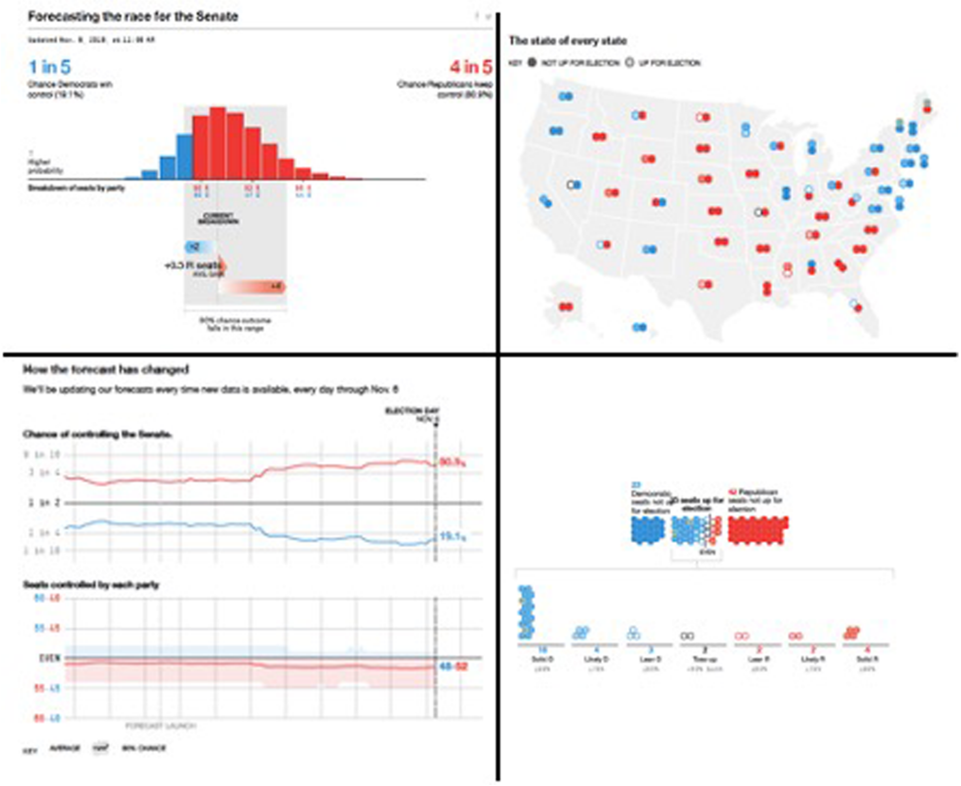


Figure 7. Example of predictive data story for form 3–*conjunction*. Source: projects.fivethirtyeight.com.

Senate") where users were given a range of possibilities to explore the data and possible outcomes (Figure 7).

Discussion: In Line with the Future

Predictive data stories are constructed by means of a limited number of story forms that permit a liaison between predictive models and journalistic analysis in the context of an unfolding narrative sequence. The three forms we were able to distinguish in our analysis following *RQ1* differ in their portrayal of one or multiple futures. At the most general level, they can either emphasize a certain outlook or highlight the multiplicity of possible prospects. No matter if they employ a strategy of emphasizing or de-emphasizing the variance in prognoses, all predictive data stories use a selection of narrative elements like characters, settings, relationships, and events (Riche et al. 2018). While it remains unclear whether the production of the predictive data stories was hypothesis-driven or data-driven, the published stories are told—unlike our assumption—in a hypothesis-driven, explanatory fashion (Parasie 2015). Even where they exhibit elements of exploration, these are conveyed *via* a Martini-Glass structure that provides a strong framework for orienting the story (Segel and Heer 2010).

In all three forms, meta-stories serve as self-reflexive means to communicate truthfully (*RQ2*). Their general purpose is to explain the data and open up the decisions driving the analysis and presentation so to render them meaningful and relevant for readers. In the story form of concentration, meta-stories, usually in text form, are positioned at the conclusion of the accompanying text and provide additional information without assuming a central role. In the contrast stories, the meta-story, again mainly text, acts as a mandatory record of the data sources, analytical decisions, and the predictive model driving the narrative. It is seamlessly embedded within the main story and serves to explain the causal links and factors shaping the potential outcomes. In the story form of conjunction, there are typically detailed text-based meta-stories that offer background information and explanations for the various predictions available. These often clarify a specific aspect of the predictive model that influences the forecast.

Arguably, meta-stories assume an active and knowledgeable audience that is keen to investigate the data (Parasie 2015; Splendore 2016). As with other data-journalistic products, data stories therefore "carry certain epistemological assumptions about how audiences might acquire knowledge" (Lewis and Westlund 2015, p. 7). Yet although some data journalism readers might well engage with the data in this way, others may lack the numeracy and statistical expertise to embrace such a more free-ranging exploratory approach.

Considering *RQ3*, the most evident commonality of all forms of predictive data stories is their strong tendency to accentuate the linearity of future developments. Even though the overview encompasses several ways of narrating data-driven projections, the forms reinstate an author-driven approach that hinges on a linear form of narration. In fact, in order to craft a story around the probabilistic spectrum of more or less likely futures, the complex, at times non-linear models are streamlined into a consecutive temporal structure. Thus, unlike our assumption that issues of uncertainty and the multiplicity of possible futures suggest a reader-driven, exploratory, and more interactive approach, predictive stories use narrative elements to sequence events and trends into a temporally structured and coherently narrated line of reasoning (Weber, Engebretsen, and Kennedy 2018; Weber 2020). Storying the data-driven predictions means also cohering information that is non-coherent but dependent on the intricacies of computational models and their potential incommensurable assumptions or data sources. To craft a narrative account that renders a data-driven prediction meaningful and connects it to political issues or readers' concerns sometimes entails simplifying an account that is inconclusive and unfit to tell a plausible story. Due to this alignment, predictive data stories assumedly have a more pronounced narrative outline than other data stories (Ojo and Heravi 2018; Weber, Engebretsen, and Kennedy 2018). Indeed, standalone interactive visualizations without a story occur only rarely (Kosara 2017). So there is no clear separation between the narrative conveyed by the text and the numeric evidence shown in the visualizations. Instead, both are combined in a multimodal composition that integrates textual and visual elements (Aiello 2020). This interplay is particularly notable in scrollytelling techniques, which are widely used in our sample.

The prevalence of linear stories may be explained by our data collection method that favored routine use, not award schemes although even there not all entries seem to excel in interactivity and audience engagement. Hence, the binary equation of 'normal' data journalism and 'excellent' data journalism along dimensions of openness and responsiveness might be misguided. In fact, Appelgren (2018) and others have looked at projects submitted to an award scheme and found that linearity prevailed there too (Kennedy, Weber, and Engebretsen 2020; Stalph and Heravi 2023; Weber, Engebretsen, and Kennedy 2018).

To explain the lack of interactivity, Klein-Avraham and Reich (2022) explore bifurcations where decisions are made during the editorial process that impede a more resolute use of interactivity. Therefore, the limited interactivity of data stories seems more the outcome of accumulated obstacles. They can be found on the level of the workforce where competences and resources could stand in the way, on the level of the workflow with editorial routines and decisions promoting interactivity or not, and on the technology level during production and consumption. Another factor that might influence the use of interactive features are role models which have been found to change with data journalism from the more traditional notion of gatekeeping for a passive audience to engaging an active audience and giving them exploratory freedom (Borges-Rey 2016; Fu and Stasko 2024). Yet despite that shift in the journalistic mindset, Anderson and Borges-Rey (2019) paradoxically found increased efforts to control and guide readers. This mismatch is corroborated by a number of studies that only discovered limited use of interactive features (Appelgren 2018; Hao et al. 2024; Heravi 2022; Kennedy, Weber, and Engebretsen 2020; Stalph and Heravi 2023; Young, Hermida, and Fulda 2018; Zamith 2019).

Conclusion

Through their capacity to tell anticipatory stories, predictive data stories enrich the repertoire of data journalism. Predictive data stories usually point toward a particular future scenario and recommend a finite list of responses to that scenario. This involves the selection of specific plans for action (or even just one), while discarding others, that are told using narrative templates. As predictive data stories, they are anchored in narrative devices which aim to tell, not show, news forecasts. As such, they can have a considerable impact, for instance on the course of elections or epidemics (Diakopoulos 2022).

In our overview, we have discussed three forms of predictive storytelling: concentration on a single scenario; contrasting different scenarios; and the conjunction of several future scenarios. In all three kinds of predictive data stories, the narrative outlines required journalists to cohere more or less conclusive information into a sequence of events and trends. In the resulting storylines, issues of predictability and uncertainty were usually not opened up for interactive exploration but were integrated into a directed explanation. For sure, journalistic pioneers aspiring to generate highly interactive and self-reflexive data journalism have come up with some striking examples that showcase the power of computationally enhanced newsmaking (Ojo and Heravi 2018). Yet our corpus of data-driven stories points us toward a more common sort. A considerable proportion of the data stories we considered were modest in their use of multimodal displays and interactive features. This lack of interactivity may only be temporary given the evolving nature of data journalistic standards that could make more versatile and mutable stories on the spectrum between author-driven and reader-driven types the norm. Such development is, however, not the only possible route predictive data journalism can take. In fact, the notion of genre suggests a state of consolidation that does merely exist for the time being and may be diffcult to achieve at all. What is more, interactivity might not be the prime characteristic of data stories at all and being little interactive does not necessarily foreclose a stories' capacity to translate numerical information about the future into a narrative account.

In the current study, we did not aim at comparing countries or outlets but were more interested in characterizing the patterns across the sample. A next step could therefore compare the venues and national contexts so to shed light on the potentially varying prominence of the stories. Such comparative analysis can also look for possible similarities and differences in terms of topic areas, temporal scope, and diagrammatic design. Another avenue for exploration is the audience of predictive data stories. This can be the imagined audiences of those producing and communicating the anticipations, that is, the preconceptions data journalists have of the people they hope to cater to. It can further include studying how readers understand and assess the stories based in their numeracy, political opinions, age, or cultural background.

That being said, predictive data stories constitute a meager contingent of the total population of data journalism products. Interrogating this fraction of outputs is still important, we argue, since studying how the future features in data journalism can also point us toward a possible future of data journalism. In its hesitation to take on an interventionist mindset, "journalism's future orientation remains the least likely temporal positioning to develop among journalists," Zelizer (2017) has conceded. Yet she adds that "it is the most directed temporal tool for sustaining journalism's future" (p. 259). In a context of nonstop communication and an unmanageable amount of shared social media content, predictive data stories can help raise people's awareness for matters of future concern that nevertheless require attention and action today.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the Central Research Development Fund, University of Bremen.

ORCID

Christian Pentzold http://orcid.org/0000-0002-6355-3150

Data Availability

Data available from here: https://osf.io/5hk8f/?view_only=1e42c47fadae429f97c0c5c9e0304f21

References

Aiello, G. 2020. "Inventorizing, Situating, Transforming: Social Semiotics and Data Visualization." In *Data Visualization in Society*, edited by Engebretsen, M., & Kennedy, H. Editor, 49–62. Amsterdam: Amsterdam University Press.

Anderson, B., and E. Borges-Rey. 2019. "Encoding the UX: User Interface as a Site of Encounter between Data Journalists and Their Constructed Audiences." *Digital Journalism* 7 (9): 1253–1269. https://doi.org/10.1080/21670811.2019.1607520

Anderson, C. W. 2018. *Apostles of Certainty: Data Journalism and the Politics of Doubt*. Oxford: Oxford University Press.

Appelgren, E. 2018. "An Illusion of Interactivity: The Paternalistic Side of Data Journalism." *Journalism Practice* 12 (3): 308–325. https://doi.org/10.1080/17512786.2017.1299032

Pentzold, C., and Fechner, D.J. 2020. Data Journalism's Many Futures: Diagrammatic Displays and Prospective Probabilities in Data-Driven News Predictions. *Convergence*, 26(4), 732-751. https://doi.org/10.1177/1354856519880790

Pentzold, C., Fechner, D.J., and Zuber, C. 2021. "Flatten the Curve": Data-driven Projections and the Journalistic Brokering of Knowledge During the COVID-19 Crisis. *Digital Journalism*, 9(9), 1370-1393. https://doi.org/10.1080/21670811.2021.1950018

Pentzold, C., and Fechner, D.J. 2021. Probabilistic Storytelling and Temporal Exigencies in Predictive Data Journalism. *Digital Journalism*, 6(9), 715-736. https://doi.org/10.1080/21670811.2021.1878920

Beer, D. 2019. *The Data Gaze*. London: Sage.

Bell, A. 1991. *The Language of News Media*. Oxford: Blackwell.

Borges, M., C. M. Correa, and M. S. Silveira. 2022. "Fundamental Elements and Characteristics for Telling Stories Using Data." *Journal on Interactive Systems* 13 (1): 77–86. https://doi.org/10.5753/jis.2022.2330

Borges-Rey, E. 2016. "Unravelling Data Journalism." *Journalism Practice* 10 (7): 833–843. https://doi.org/10.1080/17512786.2016.1159921

Borges-Rey, E. 2020. "Towards an Epistemology of Data Journalism in the Devolved Nations of the United Kingdom." *Journalism* 21 (7): 915–932. https://doi.org/10.1177/1464884917693864

Bounegru, L., T. Venturini, J. Gray, and M. Jacomy. 2017. "Narrating Networks: Exploring the Affordances of Networks as Storytelling Devices in Journalism." *Digital Journalism* 5 (6): 699–730. https://doi.org/10.1080/21670811.2016.1186497

Boy, J., F. Detienne, and J.-D. Fekete. 2015. "Storytelling in Information Visualizations: Does It Engage Users to Explore Data?." 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM Press, 1449–1458.

Burgio, V. 2019. “Uncertain Infographics: Expressing Doubt in Data Visualization.” *Semiotica* 2019 (230): 143–166. https://doi.org/10.1515/sem-2018-0114

Cairo, A. 2013. *The Functional Art: An Introduction to Information Graphics and Visualization*. Indianapolis: New Riders.

Cairo, A. 2016. *The Truthful Art. Data, Charts, and Maps for Communication*. Indianapolis: New Riders.

Davenport, T. 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

Diakopoulos, N. 2022. “Predictive Journalism: On the Role of Computational Prospection in News Media.” *Tow Center for Digital Journalism* https://ssrn.com/abstract=4092033

Drucker, J. 2014. *Graphesis*. Cambridge, MA: Harvard University Press.

Fu, Y., and J. Stasko. 2024. “More than Data Stories: Broadening the Role of Visualization in Contemporary Journalism.” *IEEE Transactions on Visualization and Computer Graphics* 30 (8): 5240–5259. https://doi.org/10.1109/TVCG.2023.3287585

González Veira, X. 2017. “I like Data-Driven Storytelling.” *Medium* https://medium.com/xocas/i-like-datadriven-storytelling-i-like-loaded-wordsd066d83ee909

Gray, J., Bounegru, L., & Chambers, L. Eds 2012. *The Data Journalism Handbook*. O’Reilly.

Grewal, Y., S. Goodwin, and T. Dwyer. 2021. “Visualising Temporal Uncertainty.” Proceedings 2021 IEEE 14th Pacific Visualization Symposium, 41–45.

Hanitzsch, T. 2007. “Deconstructing Journalism Culture.” *Communication Theory* 17 (4): 367–385. https://doi.org/10.1111/j.1468-2885.2007.00303.x

Hansen, K. R. 2016. “News from the Future: A Corpus Linguistic Analysis of Future-Oriented, Unreal and Counterfactual News Discourse.” *Discourse & Communication* 10 (2): 115–136. https://doi.org/10.1177/1750481315611240

Hao, S., Z. Wang, B. Bach, and L. Pschetz. 2024. “Design Patterns for Data-Driven News Articles.” Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI ‘24), *231*, 1–16.

Heft, A. 2021. “Transnational Journalism Networks “from below”. Cross-Border Journalistic Collaboration in Individualized Newswork.” *Journalism Studies* 22 (4): 454–474. https://doi.org/10.1080/1461670X.2021.1882876

Heravi, B. 2022. “Storytelling Structures in Data Journalism: Introducing the Water Tower Structure.” *Proceedings Computation + Journalism* https://cj2022.brown.columbia.edu/proceedings/

Hullman, J., and N. Diakopoulos. 2011. “Visualization Rhetoric: Framing Effects in Narrative Visualization.” *IEEE Transactions on Visualization and Computer Graphics* 17 (12): 2231–2240. https://doi.org/10.1109/TVCG.2011.255

Kennedy, H., W. Weber, and M. Engebretsen. 2020. “Data Visualization and Transparency in the News.” In Engebretsen, M. & Kennedy, H. (eds), *Data Visualization in Society* (pp. 169–188). Amsterdam: Amsterdam University Press.

Klein-Avraham, I., and Z. Reich. 2022. “Industrial Production of Post-Industrial Products: How Interactive Visualizations Miss Their Potential as a Journalistic Form of Knowledge.” *Journalism Studies* 23 (15): 1881–1898. https://doi.org/10.1080/1461670X.2022.2121745

Kosara, R. 2017. “An Argument Structure for Data Stories.” Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization. ACM Press, 31–35.

Lewis, S. C., and O. Westlund. 2015. “Big Data and Journalism: Epistemology, Expertise, Economics, and Ethics.” *Digital Journalism* 3 (3): 447–466. https://doi.org/10.1080/21670811.2014.976418

Loosen, W., J. Reimer, and F. De Silva-Schmidt. 2020. “Data-Driven Reporting: An on-Going ®Evolution? An Analysis of Projects Nominated for the Data Journalism Awards 2013–2016.” *Journalism* 21 (9): 1246–1263. https://doi.org/10.1177/1464884917735691

Matei, S. A., and L. Hunter. 2021. “Data Storytelling is Not Storytelling with Data: A Framework for Storytelling in Science Communication and Data Journalism.” *The Information Society* 37 (5): 312–322. https://doi.org/10.1080/01972243.2021.1951415

Maycotte, H. O. 2015. “Big Data Triggers Predictive Journalism.” *Nieman Lab* http://www.niemanlab.org/2015/12/big-data-triggers-predictive-journalism/

Morini, F. 2025. "Different yet Complementary: A Systematic Literature Review on Data Journalism in Visualization Research and Journalism Studies." *Journalism* 26Online first. (2): 425–444. https://doi.org/10.1177/14648849241237897

Neiger, M., and K. Tenenboim-Weinblatt. 2016. "Understanding Journalism through a Nuanced Deconstruction of Temporal Layers in News Narratives." *Journal of Communication* 66 (1): 139–160. https://doi.org/10.1111/jcom.12202

Nguyen, F. Ed. 2019. *News, Numbers and Public. Opinion in a Data-Driven World*. London: Bloomsbury.

Ojo, A., and B. Heravi. 2018. "Patterns in Award Winning Data Storytelling: Story Types, Enabling Tools and Competences." *Digital Journalism* 6 (6): 693–718. https://doi.org/10.1080/21670811.2017.1403291

Parasie, S. 2015. "Data-Driven Revelation? Epistemological Tensions in Investigative Journalism in the Age of "Big Data." *Digital Journalism* 3 (3): 364–380. https://doi.org/10.1080/21670811.2014.976408

Parasie, S. 2022. *Computing the News*. New York: Columbia University Press.

Pietruska, J. 2017. *Looking Forward*. Chicago: Chicago University Press.

Riche, N.H., Hurter, C., Diakopoulos, N., & Carpendale, S. Eds 2018. *Data-Driven Storytelling*. CRC Press.

Rogers, S., J. Schwabish, and D. Bowers. 2017. "Data Journalism in 2017. The Current State and Challenges Facing the Field Today." *Google News Lab, September 2017* Retrieved from https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf

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

Silver, N. 2015. *The Signal and the Noise*. New Yoek: Penguin.

Slaney, M. 2012. "Tell Me a Story." *IEEE Computer* 19 (1): 4–6.

Spiegelhalter, D., M. Pearson, and I. Short. 2011. "Visualizing Uncertainty about the Future." *Science (New York, N.Y.)* 333 (6048): 1393–1400. https://doi.org/10.1126/science.1191181

Splendore, S. 2016. "Quantitatively Oriented Forms of Journalism and Their Epistemology." *Sociology Compass* 10 (5): 343–352. https://doi.org/10.1111/soc4.12366

Stalph, F. 2018. "Classifying Data Journalism: A Content Analysis of Daily Data-Driven Stories." *Journalism Practice* 12 (10): 1332–1350. https://doi.org/10.1080/17512786.2017.1386583

Stalph, F., and B. Heravi. 2023. "Exploring Data Visualisations: An Analytical Framework Based on Dimensional Components of Data Artefacts in Journalism." *Digital Journalism* 11 (9): 1641–1663. https://doi.org/10.1080/21670811.2021.1957965

Tenenboim-Weinblatt, K., and M. Neiger. 2015. "Print is Future, Online is past: Cross-Media Analysis of Temporal Orientations in the News." *Communication Research* 42 (8): 1047–1067. https://doi.org/10.1177/0093650214558260

Tufte, E. R. 2001. *The Visual Display of Quantitative Information*. 2nd. Edition. Cheshire, CN: Graphics Press.

Weber, W. 2020. "Exploring Narrativity in Data Visualization in Journalism." In *Data Visualization in Society*, Edited by M. Engebretsen & H. Kennedy (Eds.), pp. 295–311. Amsterdam: Amsterdam University Press.

Weber, W., M. Engebretsen, and H. Kennedy. 2018. "Data Stories: Rethinking Journalistic Storytelling in the Context of Data Journalism." *Studies in Communication Sciences* 18 (1): 191–206. https://doi.org/10.24434/j.scoms.2018.01.013

Young, M. L., A. Hermida, and J. Fulda. 2018. "What Makes for Great Data Journalism? A Content Analysis of Data Journalism Awards Finalists 2012–2015." *Journalism Practice* 12 (1): 115–135. https://doi.org/10.1080/17512786.2016.1270171

Zamith, R. 2019. "Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism." *Digital Journalism* 7 (4): 470–489. https://doi.org/10.1080/21670811.2018.1554409

Zelizer, B. 2017. *What Journalism Could be*. Cambridge: Polity.

Zelizer, B. 2021. "Why Journalism's Default Neglect of Temporality is a Problem." *Media, Culture & Society* 43 (7): 1213–1229. https://doi.org/10.1177/01634437211015846