© 2024. This work is licensed under a Creative Commons "Attribution 4.0 International" license.

Digital Humanities at the Intersection of Three Approaches to Data Visualisation: Statistical Graphics, Data Humanism, and Humanistic Interpretation

Aida Horaniet Ibañez

Center for Contemporary and Digital History (C2DH), University of Luxembourg

Digital humanities has a critical role in the progress of the data visualisation field. First, humanities scholars engage with the concept of knowledge as interpretation. Second, they leverage computational tools and statistical methods for the analysis and visualisation of data and metadata. And finally, Digital Humanities projects are a space for experimentation where different epistemic cultures (which often go beyond the Humanities and Computer Science) negotiate new forms of knowledge. Digital Humanities are therefore in a privileged position to benefit from the full potential of data visualisation, using and improving existing methods and tools, and participating in the development of those that are much needed but do not yet exist. This article presents a classification of current approaches to data visualisation into (I) statistical graphics, (II) data humanism and (III) humanistic interpretation. These three approaches are based on four main aspects: a) the intellectual habits around the definition of data, b) the characteristics of the data visualisation and its objectives, c) the relation between the data and the data visualisation, and d) the expected user interaction. The classification is not intended to narrowly categorise any data visualisation, but rather to help navigate the visualisation continuum across epistemological practices. Moreover, based on a large collection of data visualisation examples compiled in an online gallery, several techniques are identified, that allow to make the transition between the different approaches, potentially facilitating interdisciplinary projects such as the LuxTIME Machine, that leverage multiple types of sources across languages and modalities.

Keywords: digital humanities, data visualisation, statistical graphs, data humanism, interpretation

1 Introduction

This research is developed in the context of the LuxTIME Machine (LuxTIME), an interdisciplinary project about the industrial history of Belval, Luxembourg. We use data visualisation to explore, understand and communicate about data from historical archives, scientific data from the fields of eco-hydrology and cheminformatics, and socio-cultural data. In LuxTIME, we hypothesize that data visualisation can facilitate interdisciplinary work, this research provides one study for understanding the needs of the different disciplines, the common grounds, and the options for visualising data integrating different approaches.

Leveraging data visualisation for research involving multimodal and interdisciplinary data requires understanding the needs and current practices along the epistemological continuum across quantitative, qualitative, rhetorical, and creative practices. In this research, I aim to answer the following three central questions: 1) Is there a data visualisation classification that could help us understand the needs and navigate the different solutions across epistemological practices? 2) What distinguishes the different approaches to data visualisation? And 3) which specific data visualisation techniques are used to transition between those approaches?

In section 3.1, I propose a classification into statistical graphics, data humanism and humanistic interpretation, based on a) the intellectual habits around the definition of data, b) the characteristics of the data visualisation and its objectives, c) the relation between the data and the data visualisation, and d) the expected user interaction. This classification is intended to help us understand the main needs and how they are reflected in the practice of data visualisation in the different fields. In section 3.2, with the aim of understanding which specific data visualisation techniques enable the transition from statistical graphics to data humanism and humanistic interpretation, I curate and analyse a collection of 400 data visualisations.

Finally, in section 4, I discuss the current gaps between the approaches, the possibilities for future development to facilitate the use of a variety of data visualisation techniques across fields of research, and the role of the Digital Humanities.

2 Methods

This part of the research project, which lays the theoretical foundation for further practical experimentation, is conducted in three phases.[1] First, I explored a collection of data visualisations from research articles, mainly published in journals in the related fields of the project: Hydrology (e.g., Hydrology Sciences Journal, Advances in Water Resources), Environmental Cheminformatics (e.g., Exposome Journal), and Digital Humanities (e.g., Digital Humanities Quarterly, Journal of Digital History, Digital Scholarship in the Humanities). This initial exploration resulted in a predominance of visualisations using standard statistical graphs, including bar charts, histograms, line plots, heatmaps, scatterplots, pie charts, contour plots, networks graphs and choropleth maps, among others. This is probably due, among other factors, to the fact that numerous data visualisation tools mostly support this type of

[1] In this article only the theoretical framework is discussed. The practical application to the project has been discussed in a separate article (Aida Horaniet Ibañez and Dagny Aurich, 2023)

standard graphs, accessible to all types of users with minimal technical and statistical knowledge. During this initial phase of the research, some data visualisations, later classified as humanistic interpretation, were also found. This initial collection gathered about 200 data visualisations [2](see Figure 1), extracted from the different volumes of each journal starting with the most recent publications. It is important to mention that the use of data visualisation in the Digital Humanities goes beyond Digital Humanities journals. It is present in discipline (and subdiscipline) -specific journals such as Digital History, Digital Art History, or Literary Studies, among others. Similarly, among the publications in data visualisation journals, there are design studies focusing on use cases in the humanities that have not been included in this first exploration. It is beyond the scope of this research to make an exhaustive collection of all data visualisations in the field.

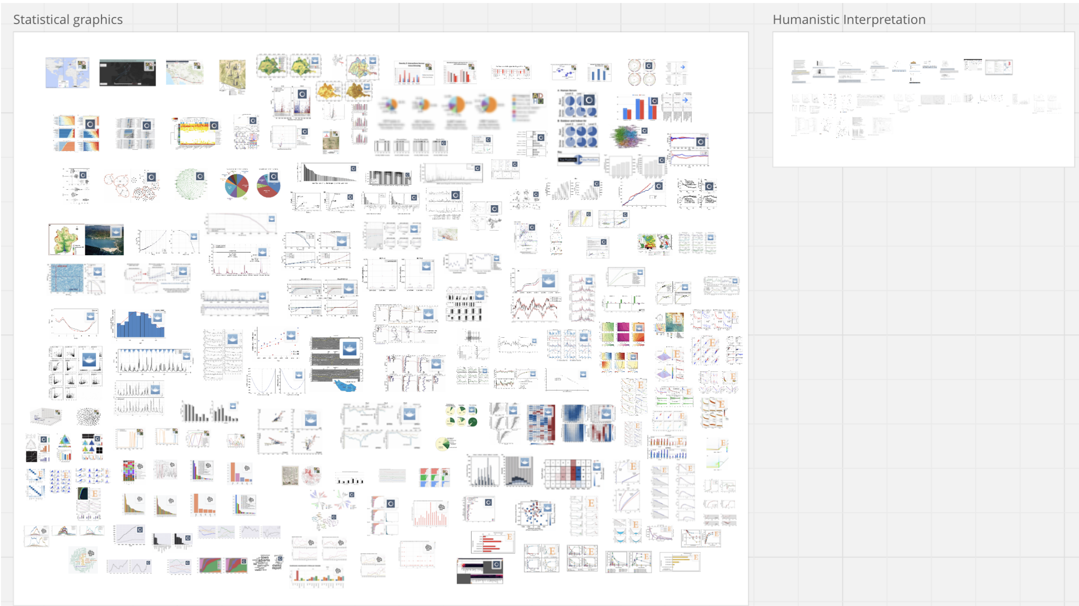

Figure 1: Initial collection

Secondly, I gathered and explored a collection of about 400 data visualisations (see Figure 2) that excluded those using strictly standard charts.[3] As it will be discussed in section 4.2, the number of visualisations currently published in the online gallery is 400 because many links are no longer valid, but the number has been higher throughout the project. In addition to research articles, this second collection also includes non-research data sources such as online newspapers and magazines (e.g., Nightingale, National Geographic, South China Morning Post, Scientific American), social media posts (e.g., Twitter/X, Instagram), information designers' portfolios, data visualisation tool galleries (e.g., Tableau Public Gallery), community projects (e.g., Viz for Social Good, journaldataviz) and data visualisation awards (e.g., Information is Beautiful Awards). Due to the variety of sources, this collection has been compiled

[2] This collection of data visualisations is collected in a Miro board. The separation into statistical graphics and humanistic interpretation was done later to integrate the latter in the second gallery and analysis.

[3] This collection of 400 data visualisations is available on Github (https://aidahi-unilu.github.io/AlternativeDataViz), including author(s), title, description, source, and the applied techniques, as discussed in the results section. This collection may be enlarged in the future with submissions from other practitioners in the field, through a call for contributions.

manually, with the sole criterion of excluding standard statistical graphs. The reason why I exclude statistical graphs in this second phase is that there is abundant literature on their use, and therefore, they do not introduce new evidence. Only with the initial exploration I can confirm that they potentially represent an approach on their own, how they later differ from the second gallery, and validate the existing literature as to their characteristics.

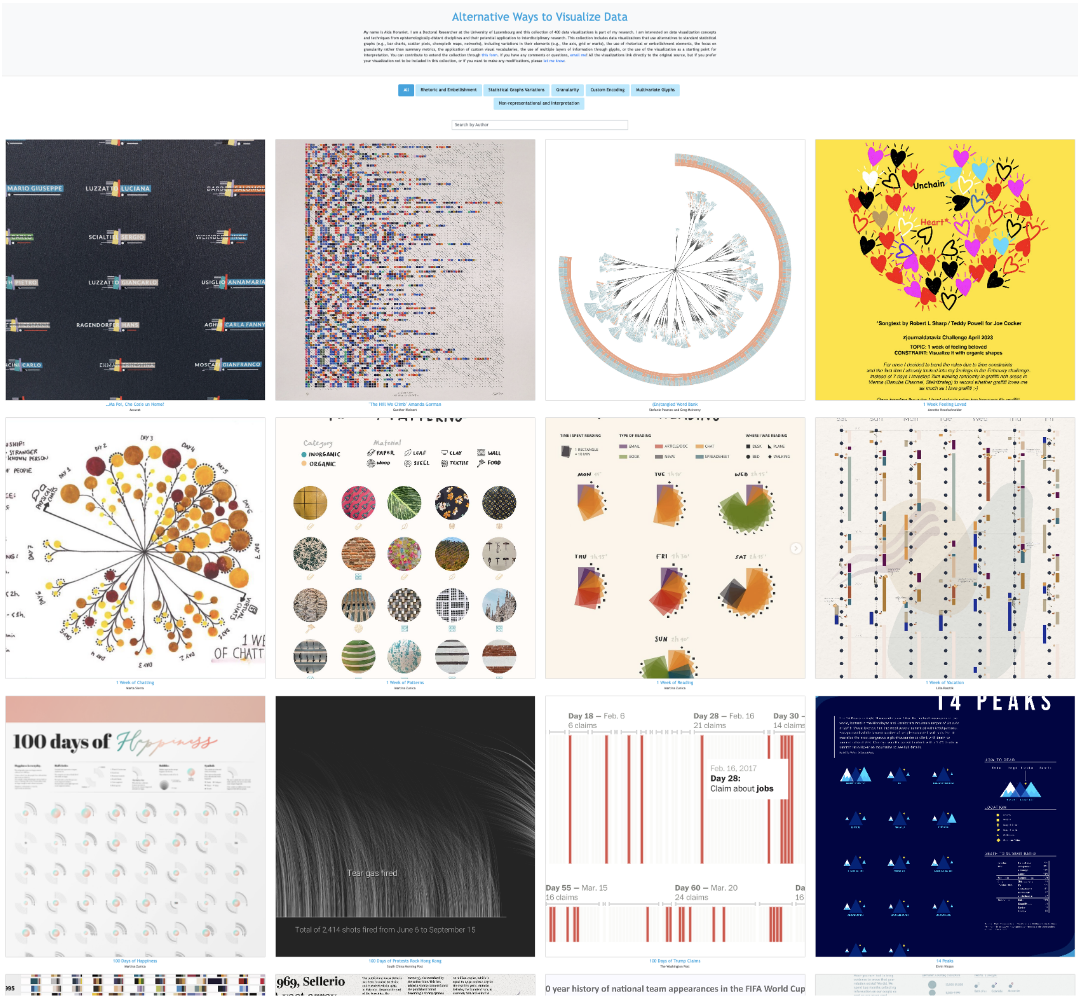

Figure 2: Non-standard charts gallery

Finally, based on these two collections, I propose a classification of approaches to data visualisation, and describe their main characteristics. Moreover, I identify a list of techniques used to move between these approaches. In the second collection, I have labelled the techniques used in the visualisations, and not in which approach they should be classified, because in some cases different combinations of the techniques allow to create visualisations belonging to different approaches. In other cases, the classifications are ambiguous, as the visualisations apply concepts from different approaches.

The classification is intended to help us to read the current landscape in general terms, while the techniques, give us the practical tools to move between the different approaches. These could be potentially used to explore epistemologically distant techniques in interdisciplinary setups and to be standardized and integrated in data visualisation tools.

3 Results

3.1 Classification overview

Table 1 presents a very concise summary of the classification that will be described in detail in the following sections. It includes the suggested name for the approach, the intellectual habits around the definition of data as given i.e., a representation of a pre-existing fact or taken i.e., actively constructed, the main objectives of the data visualisation and its characteristics, the relation data to data visualisation, the expected user interaction, and the fields where it is most frequently used.

Table 1: Summary of Approaches to Data Visualisation

Approach	Data	Objective	Characteristics	Data to Visualisation Relation	User Interaction	Main Fields
Statistical Graphics	Given	Complexity reduction, decision-making	Summary statistics, single clear message	Representational	Known charts, formatted answers	Business, Engineering, Natural Sciences
Data Humanism	Given (context emphasis)	Emotional connection, engagement	Details, context, multiple narratives	Representational	Custom visual vocabularies, open exploration	Journalism, Education, Information Design
Humanistic Interpretation	Taken	Model interpretation	Non-standard metrics	Non representational	Close reading, marking, interpretation	History, Art, Literature

3.1.1 Approach names

To facilitate the subsequent analysis, each approach has been identified with a name that describes it as closely as possible: Approach 1: Statistical graphics, Approach 2: Data humanism and Approach 3: Humanistic interpretation.

Statistical graphics includes all the known forms of graphical representation based on statistical principles used to describe and summarize data, including plots such as scatter plots, histograms, box plots, choropleth maps, bar charts, line charts or heatmaps. The name chosen for the category is the most frequent name for this set of charts. I have included networks in this approach, because although they depict individual entities and relationships; they are standard charts (unless they include variations) based on mathematical structures, often used to gather statistics. In Figure 1, a collection of standard statistical graphs was presented.

Data humanism is based on the concept defined by (Lupi, 2017) where she claims that “whenever the main purpose of data visualisation is to open people’s eyes to fresh knowledge, it is impractical to avoid a certain level of visual complexity”. This approach refers to a practice that embraces complexity in data visualisation, slowness, engagement looking at details – the grain, understanding the individual cases behind the numbers, avoiding standards when they do not fit the purpose, and adding context to understand the uniqueness of the representation. In Figure 3, we can see a collection of data visualisations that introduce some of these ideas.

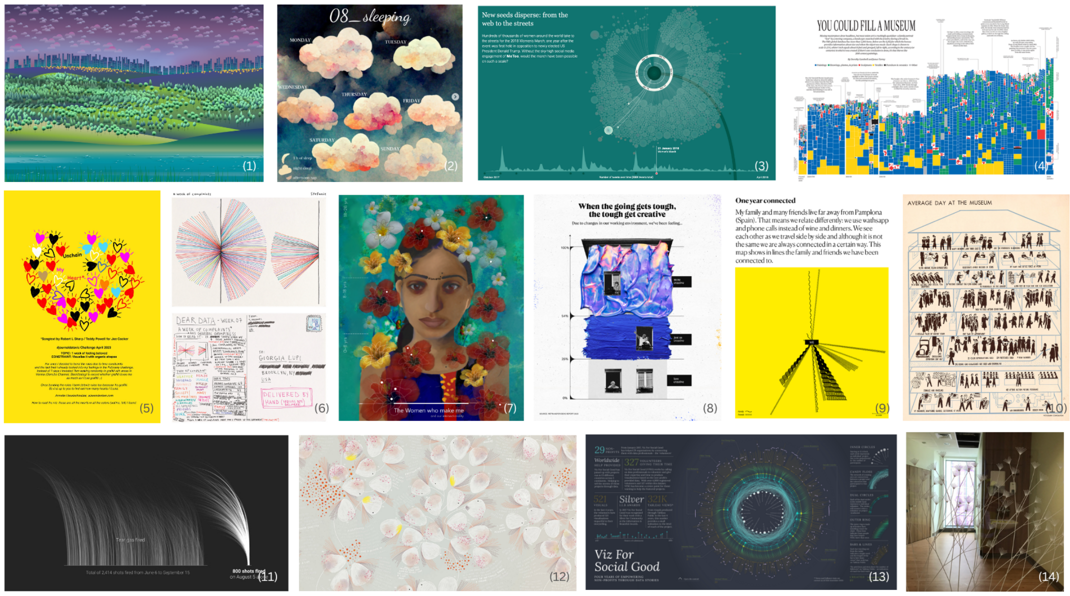

Figure 3: (1)(Sonja Kuijpers, 2019), (2)(Alenka Gucek, 2023), (3)(Valentina D'Efilippo and Lucia Kocincova, 2017), (4)(Gambrell, 2018), (5)(Annette Hexelschneider, 2023), (6)(Lupi and Posavec, 2016), (7)(Kadambari Komandur, 2022), (8)(Gabrielle Merite, 2020), (9)(Alberto Molina Arce, 2022), (10)(MoMA Archives, 1946),(11)(Pablo Robles, Darren Long, Dennis Wong, 2019), (12)(Giorgia Lupi, 2017), (13)(Samuel Parsons, 2020), (14)(Stefanie Posavec, 2019)

Humanistic interpretation refers to the approach that uses data visualisation to model interpretation and produce knowledge. A non-representational approach, in which data visualisation is not a surrogate of the dataset but the primary act of knowledge production (Drucker, 2020). Besides the basic graphical features used in other approaches, it uses activators and inflectors and dimensions of interpretation, such as point of view or relative scales. This concept is based on the work of Drucker.[4] In Figure 4, we can see a less homogeneous collection of data visualisations representing humanistic interpretation. As we will discuss later, the examples are fewer, and thus we observe several attempts at interpretation through visualisation, from a mixture of proposed sketches from articles, physicalisations, and personal projects, among others.

3.1.2 Data

The intellectual habits around the definition of data are the first element of differentiation among approaches: given or taken.[5] I refer to the intellectual habits, and not to the definition of data itself, because all data is taken (as opposed to found or given). However, in the three approaches presented here, we observe different levels of engagement with the data construction process, or how that data is taken or produced, before being reused.

In statistical graphics, the intellectual habit is to use data as the starting point, a pre-existing fact, used practically as if it was given. In science and business, a

4 (Drucker, 2011, 2017, 2020; Drucker and Nowviskie)

5 Based on Drucker's concept of data as capta being taken instead of given (Drucker, 2011)

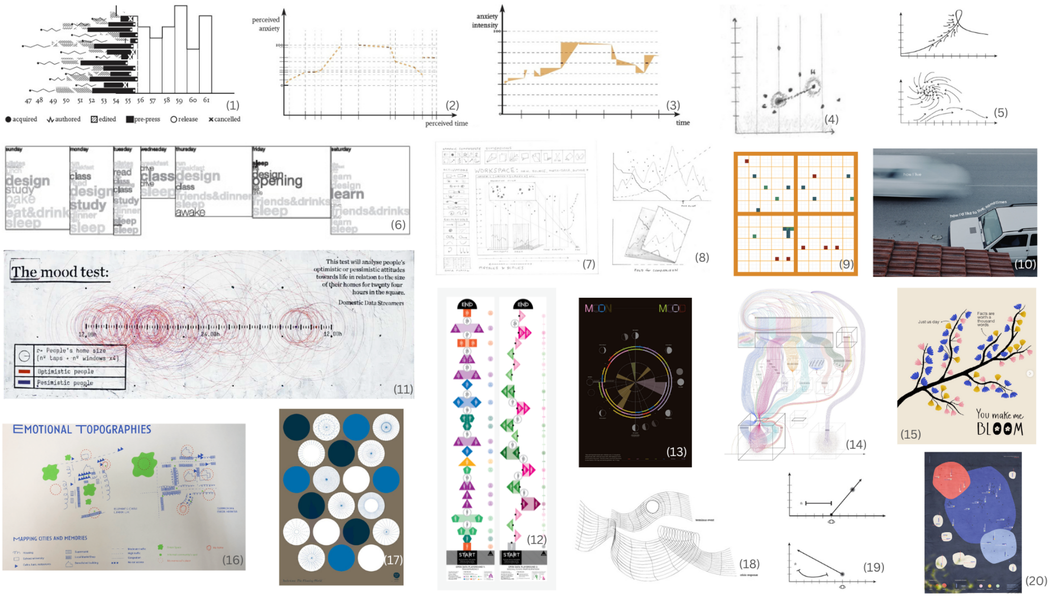

Figure 4: (1,2,3,5,6,18,19)(Drucker, 2011), (4,7,8)(Drucker, 2017), (9)(Lauren Klein et al., 2016), (10)(Alessia Musìo, 2022), (11)(Streamers, 2013), (12)(Stefanie Posavec, 2014), (13)(Sarah Stern, 2022), (14)(Stefanie Posavec, 2023), 15)(Martina Zunica, 2023),(16)(Syennie Valeria, 2022),(17)(Stefanie Posavec, 2011),(20)(Camilla De Amicis, 2022)

data-driven culture suggests to always "start with the data" (Patil and Mason, 2015), what somehow implies that the data is "the truth", a value-neutral object. The visualisations cite the data source, and if that source is considered reliable in the context in which it is presented, that is sufficient. Information about uncertainty is collected statistically in the form of confidence intervals and mathematical errors. Data is observer independent.

Data visualisation, as we predominantly know it today is the result of two main transformations: first, the data collection that is the mapping of a phenomenon to data, and second, the data visualisation or generation of a visual display from that data. This process involves a series of decisions from the occurrence of the phenomenon to the data visualisation, and beyond to the understanding of it: what elements do we collect depending on the physical or emotional context, the purpose, the person's previous knowledge or the configuration of the machine; what is grouped together; where the data is stored; which elements we choose to visualise, using which representations, what level of granularity or aggregation we choose and why. Even before adding the complexity of more advanced statistical models, the process involves many decisions, and each combination of them leads to a different result, a different dataset, and a different data visualisation. "Rethinking histories of data requires not only better answers to existing questions, but also better questions" (Strasser and Edwards, 2017). In Big Data is the Answer…But What is the Question? Strasser and Edwards suggest several questions such as "What counts as data?", "How are objects related to data?", or "Why do we keep data, and how do we decide which data to lose or forget?".

In data humanism, the intellectual habit is still to use data mostly as if it was given. However, a greater effort is made to situate it and to understand the reductive process that occurs during its generation. This will be visible later in the exploration of the data visualisation that becomes co-dependent with the user. In this approach, citing the data source, regardless of how trustworthy it is, is not enough. It requires a problematisation of the provenance of the data, an identification of the stakeholders, and other more human elements necessary to understand "the where, why, and how of data" (D'Ignazio, 2017). It interrogates the context, limitations, and validity of the same. It could be argued that such problematisation allows to classify data humanism in the intellectual habit of using data as taken and not given. Nevertheless, in many of the examples analyzed, we see that the discussion occurs mostly at the level of the visualisation, and to a lesser extent at the data level. It remains true that some visualisations classified within this approach could indeed fit into the category of taken.

In humanistic interpretation, the intellectual habit is to always approach data as taken. Loukissas states that "things become data within interpretative acts" (Loukissas, 2019). Rosenberg emphasizes on the rhetorical meaning of data: "Data has no truth (...). It may be that the data collected has no relation to truth or reality whatsoever beyond the reality that data help us to construct" (Gitelman, 2013). Drucker refers to data as capta, 'taken' actively, while data is assumed to be a 'given' able to be recorded and observed" (Drucker, 2011). In this approach to data visualisation, special attention is paid to the process itself, to the sources and their cultures because "when data are assembled from heterogeneous sources, each with their own local conditions, multiple settings are juxtaposed, creating a clash between discordant originating data cultures (Loukissas, 2019); where it is stored since "making data means bringing a subject into a pre-existing system, defined by durable conditions of data collection as well as storage, analysis, and dissemination" (Loukissas, 2019). Manovich refers to datasets as not just any collection of information but "objects structured in ways that allow them to exist within a computational medium and be analyzable by particular methods" (Manovich, 2020). It looks at the limitations of the data itself, what exactly it represents, and the definition of social identities created through the definition of categories (Suchman, 1993).

While the intellectual habits around the definition of data in statistical graphics are clearly different from the other two approaches, the difference between data humanism and humanistic interpretation is subtle. Perhaps because a large part of the visualisations classified as data humanism comes from statistical graphics practitioners who have started to experiment with new visual vocabularies and high levels of granularity to avoid the limitations of the graphics they know, but without the critique of the process more rooted in the humanities. Furthermore, data humanism constitutes an attempt to make data more human, while humanistic interpretation, born out of the humanities disciplines, studies artefacts created by humans, objects of culture.

3.1.3 Objectives and characteristics

The second factor of differentiation between approaches concerns the objectives of the data visualisation and how they characterize the graphical representation.

Statistical graphics aim at simplifying complex topics, they follow a reductionist approach that allows to convey high amounts of information into digestible charts very quickly. The objective is to support effective decision-making and to communicate a clear message to a given audience. To achieve this goal, the visualisations are characterized by their abstraction, reduction, standardization, representation, and legibility. Using predefined charts and standard practices facilitates the fast processing of the content. These charts are selected based on the number of dimensions and measures, the type of data to be represented (e.g., continuous, discrete, categorical) and the data relationships that need to be displayed (e.g., deviation, ranking, flow, part-to-whole). The extensive research about graphic semiology and use of visual elements (e.g., color, shapes, size), the mechanics of the sight (e.g., Gestalt principles) and user experience, is applied to optimize the process, that is to ensure that the user receives the message as effectively as possible. It adheres to the concept of data-ink ratio introduced by Tufte (Tufte, 1999), according to which any visual elements that interfere with the attention to the data are clutter, and therefore they should be avoided.

Data humanism focuses on promoting human values to foster an emotional connection with the user. Data visualisations show complex topics in depth. The objective is to engage the user in the interpretation of the visualisation, introducing multiple narratives when necessary. Granularity, specificity, full coverage, high realism, physicality, i.e., how data is perceived through human senses, and situatedness, i.e., the relation between the data and the context (e.g., space, time), are the main characteristics of this approach. Special importance is given to showing the gaps in the data and the complexity of the visualisation process itself. Unique visual vocabularies adapted to the visualisation of a particular theme are introduced, using organic shapes and original elements. These are defined from any combination of visual elements such as color, shape, angle, or size; and they are often explained in a "how to read" section. The use of non-essential or embellishment visuals is common. Communication and storytelling theory is applied to facilitate the navigation through the visualisation of the multiple narratives (e.g., audience analysis, story structure). It embraces pluralism and includes multiple perspectives to increase reflexivity (D'Ignazio and Klein, 2020).

In humanistic interpretation, the goal is the interpretative process itself. A co-dependent constructivism allows to directly construct arguments through visual means. The visualisations are characterized by the use of non-standard visual elements such as non-discrete categories, unequal scale divisions, metrics as a factor of a point of view, of assumptions, presumptions or convention; non-continuous, non-homogeneous and multidirectional temporalities (Drucker, 2017). Figure 5 revisits Snow's cholera chart, where Drucker suggests looking at the individual profiles of the people behind the dots (what would be considered as data humanism), but she goes further and suggests to "take the rate deaths, their frequency, and chart them on a temporal axis inflected by increasing panic". Redrawing the urban streetscape to express the emotional landscape is humanistic interpretation.

3.1.4 Relation between data and visualisation

The third differentiator explores the relation between the data and the visualisation. In statistical graphics and data humanism, data is the starting point and from the data we generate the data visualisation. They are both based on a representational approach in which "the relation between data and display is uni-directional, the data precede the display, and the data are presumed to have some reliable representational relation to the phenomena from which they have been abstracted" (Drucker, 2017). Humanistic interpretation focuses on a non-representational approach, that uses the "graphical

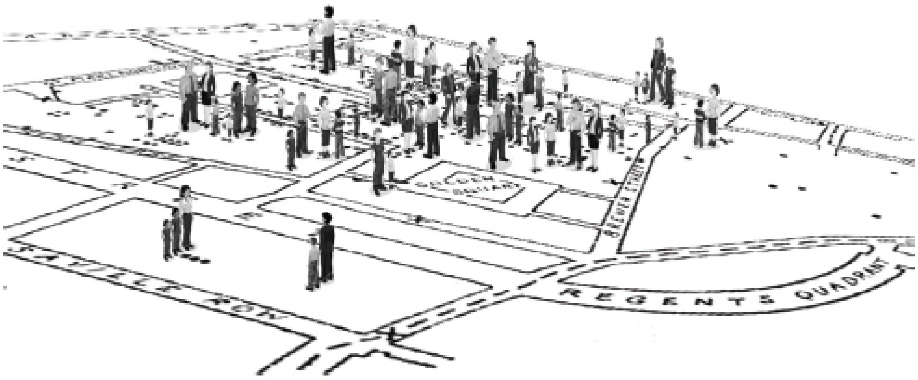

Figure 5: Snow's chart revisited to show the individuals behind the dots, emphasizing that "each dot represents a life, and none of them are identical" (Drucker, 2011)

input as primary means of interpretative work" (Drucker, 2017). The interaction with the data visualisation changes the data or creates new versions of it.

3.1.5 User Interaction

Last, in statistical graphics, the user is expected to know the charts used in the visualisation. It requires a certain data literacy, including knowledge about statistical graphs (e.g., selecting the right charts or knowing how to read them). A user who knows how to read a type of graph can quickly read a graph related to any data set, ask predefined questions (e.g., in the case of a histogram: where most values fall and how much the variation is among values), and generate formatted answers. A critical approach is expected in terms of assessing whether the rules of statistical graphs are respected such as using the wrong graph, manipulating the axes, partially selecting data, or going against conventions to mislead the user. The user can interact with the visualisation by using filters or drilling up and down to different levels of aggregation (e.g., by date – monthly, yearly; by category). The visualisation is designed to effectively answer these predefined questions.

In data humanism, the user does not know beforehand what to expect from the visualisation. The design and the topic stimulate the user's interest in exploring it further. No standard graphics are used, and if they are, they have been significantly redesigned in some way. Based on the user's own understanding of the data representation, guided by the frequently used 'how to read' section, the visualisation allows to ask open questions and explore the answers. The user is expected to take ample time to explore it, as effectiveness is not an important factor. We do not know if special skills or some level of data literacy is required to read the visualisations, as they are not standardized and are usually accompanied by instructions. Lupi, in her manifesto, "Data Humanism, The Revolution will be Visualized" states that "the first wave [referring to cheap marketing infographics] was successful in making others more familiar with new terms and visual languages", implying that some kind of learning process is required to read these data visualisations (Lupi, 2017). We "might want to move beyond literacy with datasets and towards literacies with infrastructure" since “datasets do not simply neutrally designate aspects of the world, they also render the world in accordance with different visions, values and cultures, making it navigable through data” (Gray et al., 2018). To gain an in-depth understanding of the implications, data literacy research needs to be extended to data humanism and humanistic interpretation.

In humanistic interpretation, the user interacts with the visualisation to build knowledge, the process happens through a series of interactions in which the user can mark and annotate the data. A set of tools allow the interaction with the visualisation, such as activators to introduce attraction and repulsion, weight, force, or sequence; or the use of graphical elements of interpretation such as point of view, layers, slices, tilt, fold, parallax or split (Drucker, 2017). Finally, the visualisation produced is also subject to interpretation. The user of this type of visualisations usually has a humanistic background, accustomed to the principles of ambiguity, complexity, and interpretation.

3.2 Data visualisation techniques: from statistical graphics to data humanism and humanistic interpretation

Based on the second collection described in section 2. I identified several techniques to move between statistical graphics and data humanism, and humanistic interpretation. They range from subtle variations of standard charts (e.g., eliminating the grid or the tick marks in a bar chart), decomposing data showing the granularity behind the aggregated values, using custom visual vocabularies such as organic shapes, overlaying multiple layers of information using data glyphs, to completely challenging the principles of standard charts such as homogeneously spaced time units. All of them allow us to walk that continuum across epistemological practices. In the following sections, they are discussed separately in detail. Furthermore, these techniques are not mutually exclusive, but can be applied simultaneously, as can be seen in the online gallery through the multifilter selection.

3.2.1 Variations of statistical graphs

When using standard statistical charts, a series of rules are applied to convey the information efficiently. These rules include, among others, the selection of a graph adapted to the data and the analytical purpose, the right level of aggregation, the precise display of axes, labels, gridlines, legends, or annotations; or a low data-ink ratio avoiding distracting elements. In this analysis, I consider variations of statistical graphs, all those data visualisations including graphs where the standard graph is still recognizable, and knowing the rules that apply to such a graph, one or more rules are not applied for a specific purpose.

The options are multiple, but from the collection studied, it is worth noting the total elimination of grids, axes, or tick marks, to draw attention to the changes and not the specific values (Kim Albrecht, 2016);(Lopez et al., 2019); the turning, superposition and use of curved axes (Pentagram, 2021); and the over-encoding, using multiple visual attributes to encode the same information (Lazzaroni, 2020). It is not considered a variation of a statistical graph when the graphs are misleading (e.g., extending the y-axis to minimize a variance), or when, for no specific purpose, the rules of standard charts do not apply (e.g., the legend is missing and there is no way to understand the content). These visualisations must be based on decisions made deliberately by the designer, to move between different approaches. Although the techniques presented below could be generally described in some cases as other types of variations, in the online gallery, I have only identified in this category those where the reference graph is fully recognizable.

3.2.2 Custom encoding

The use of custom visual vocabularies implies leaving aside the standard visual attributes associated with statistical graphs such as lines, rectangles and circles and creating an alternative way of encoding data variables, from replacing some of the elements of a standard statistical graph to completely redesigning all the elements of the visualisation.

Frequently observed encodings include organic forms (Clark, 2022; Kadambari Komandur, 2022; Martina Zunica, 2023; Sonja Kuijpers, 2018), but also new combinations of geometric forms (Nwosu, 2022). Among others, the use of custom visual vocabularies allows us to disaggregate information (by not having the constraint of standard graphs based on aggregate metrics), to display simultaneously multiple layers of information by integrating multiple variables in the visual vocabulary (see multivariate data glyphs below), to use rhetoric to attract the reader's attention (see mapping rhetoric below), and to potentially remove barriers to data literacy, since the reader does not need prior knowledge in statistics to read the visualisation[6]. This type of visualisation requires a "how to read" section, so that learning to read it is part of the process.

3.2.3 Rhetoric and embellishment

Contrary to data visualisations with a low data-ink ratio, where "extra" elements are avoided, there are also data visualisations where these "embellishing" elements are used with a rhetorical purpose to inform, persuade or motivate a given audience in a particular situation. These elements can be external to the graphics and therefore not linked to the data, in the form of images or drawings. Furthermore, rhetoric and embellishment elements can be part of the graph, not directly to encode the data but for example by using colour to create a specific effect (Gabrielle Merite, 2020; Scientific American, 2015), customizing the layout (Valentina D'Efilippo and Lucia Kocincova, 2017), rotating and superimposing graphs or through annotations (Bravo, 2020). Last, rhetoric can also be applied directly in connection to the data, often in combination with the use of custom encodings through rhetorical mapping, where the visual attributes used to encode the data, apply visual metaphors to present the data in terms of the topic, the context, or the implications (Ervin Vinzon, 2021; Kate Wong et al., 2021). Besides "mapping rhetoric", or "the information presentation via the data-to-visual transfer function" (Hullman and Diakopoulos, 2011), Hullman and Diakopoulus also describe other rhetoric techniques (e.g., information access rhetoric, provenance rhetoric, linguistic-based rhetoric, procedural rhetoric).

[6] From The Historical Archive To The Citizens: Visualizing Census Data From Brill Street in 1922 (Aida Horaniet Ibañez et al., 2023) presents an example of the use of custom encodings to make historical data accessible to citizens.

3.2.4 Granularity

Granularity is defined in this research as the intent to disaggregate summary data typically presented in standard statistical graphs. The "grain" in a dataset is the smallest unit collected. In a data visualisation focused on granularity, importance is given to individual and unique cases behind the aggregated numbers. There might be different levels of aggregation/disaggregation, and they might be combined in the same data visualisation (drill-up/down approach). For example, in the data visualisation about the history of national team appearances in a world cup (Sonja Kuijpers, 2018), the grain would be the data related to each national team's appearance (e.g., year, result), the semi-aggregated data would present summary metrics by country (e.g., number of appearances, total number of points), and the aggregated data would describe the worldwide data (e.g., total number of tournaments, average number of points, total number of regions). These levels of aggregation are relative to the objectives of the visualisation, since the data can be broken down from many different perspectives. The granularity presented in a data visualisation gives us an important clue as to where it is situated in the epistemological practices. On the one hand, the Natural and Applied Sciences use predominantly quantitative methods to measure an "objective" reality conceptualised as cause-effect relationships that can be generalized, that requires to summarize to explain and predict. On the other hand, the Humanities use predominantly critical and rhetorical methods to interrogate particular symbolic actions in order to understand a "subjective reality". The closer to the "grain" the closer to the Humanities, the rhetorical analysis and the interpretation; the more aggregated, the closer to the Natural and Applied Sciences, standardization, and generalization.

Based on the collection studied, one of the most commonly used techniques is displaying individual points, icons or drawings to humanize data, such as arrests during a protest (Post, 2020), Covid-19 deaths (Pentagram, 2020), personal experiences (Lupi and Posavec, 2016) or cells in a Jupyter notebook to create a unique article fingerprint (C2DH, 2021). Another approach consists of replacing (or overlaying) traditional shapes, such as the bars in a bar chart, with the data points that constitute them (Gambrell, 2018), showing both the aggregated and the disaggregated view at the same time; or using direct visualisation (Manovich, 2011), preserving to some extent the original form of the grain (Flavio Gortana et al., 2018); (Ferry, 2020).

3.2.5 Multivariate data glyphs

A way to display multiple layers of information simultaneously, while maintaining a certain degree of granularity, is the use of multivariate data glyphs, a visual representation of data where the attributes of the graphical entity are defined by the attributes of the data record. To form the data glyphs, the encoding might rely on standard graphs and different levels of variations (Silvia Romanelli, 2022), customized visual vocabularies (Kimly Scott, 2023), or a combination of both. Every data glyph defines the lowest level of detail and they usually appear as small multiples, where each representation is derived from the same design structure, to leverage visual consistency, economy of perception and uninterrupted visual reasoning (Chuah and Eick, 1998).

3.2.6 Non-representational and interpretation

In contrast to representational approaches, in a non-representational approach, the existence of data or other representations is not assumed prior to the interpretative work. Examples include data visualisations that use visual attributes to collect data not based on measurements or observations, such as personal experience or perception (Panagiotidou et al., 2020; Pentagram, 2021; Streamers, 2013). Another example is the use of visual vocabularies for interpretation, such as the uncertainty about a time duration (Nadieh Bremer, 2016) or more complex concepts such as point of view, layers, scales or folds (Drucker, 2017). Finally, it includes data visualisations where not only the graph elements can be aesthetically modified, but also the underlying principles, using for example ambiguous categories, multidirectional or perceived timeliness (Drucker, 2011). Although these elements could be evaluated separately, I have decided to group them together because there are very few visualisations that implement them.

4 Discussion

4.1 Approaches and gaps

The classification proposed in this article helps us to understand the current landscape in data visualisation beyond disciplinary boundaries. The collection of data visualisations belonging to data humanism and humanistic interpretation allows us to identify, through numerous examples, a series of techniques linking these three classifications along the continuum of epistemological practices.

Data humanism and humanistic interpretation, both present options for shifting beyond statistical graphics. However, it is also interesting to discuss the gaps between the two as possibilities for future research. First, both promote the use of new visual vocabularies, however the "interpretative" visual language described in humanistic interpretation is perhaps more along the lines of a new standardization than the freedom to use any visual language, i.e., a specific symbol would mean attraction, as opposed to the designer deciding which representation of attraction is the most appropriate to a specific case, as it would be the case in data humanism. Second, both use variations of statistical graphs: In data humanism the aesthetics of the graph elements are modified, while in humanistic interpretation, the basic principles of their construction are challenged. Even so, aesthetically, the latter seems closer to the statistical graphs. In the literature, data humanism and humanistic interpretation are closer to each other than to statistical graphics (e.g., relevance of context to situate the data, transparency of the design process). However, based on the examples analyzed in the gallery, visually, it is data humanism that is more distant. In addition, we find numerous examples that symbolize the gradual transition between statistical graphics and data humanism, as well as between statistical graphics and humanistic interpretation (in proportion to the lower number of examples available in this category). We do not find so many examples situated between data humanism and humanistic interpretation, though. This leads to the question of what case studies would require such a combination, and how to bridge this gap between the two approaches, which are not the same.

4.2 Future applications and development

The techniques discussed above, among others, have been applied in the continuation of the LuxTIME project to facilitate interdisciplinary work (Aida Horaniet Ibañez and Dagny Aurich, 2023). In addition, the identified techniques could be used as a reference to standardize certain functionalities in the existing data visualisation tools, since as we can observe, most of them have been created either with design programs (e.g. Adobe Illustrator), sacrificing the connection with the data and the interactivity; or with programming libraries such as (d3.js), limiting the accessibility to users without development knowledge. With some basic statistical and digital knowledge, any user can create standard statistical charts using tools like Excel or Tableau. However, moving away from statistical graphics requires specialized tools. This implies that the natural and applied sciences are not exposed to other approaches, and that the Humanities, which need these techniques for interpretative work, are limited to quantitative analysis and standard visualisations. Without any doubt, some of the elements discussed above, especially regarding rhetorical elements and customized visual vocabularies, have a degree of variability that makes their development challenging, but offering some options would already be a step forward. Despite the existence of many successful design studies specific to the Digital Humanities e.g., (McCurdy et al., 2016) and tools that address one or more humanities challenges (e.g., Voyant), these are rarely integrated into other existing tools, so often their impact remains limited. Such integration would support the cross-fertilisation between the different approaches and facilitate the use of different methods and techniques in disciplinary and interdisciplinary contexts.

In this analysis, I have decided to group together certain techniques that could be analyzed extensively on their own. However, the lack of a gallery of data visualisations that would allow us to do so, and in cases, such as in humanistic interpretation, the lack of applied examples, make the disaggregation difficult at this stage. This gallery has been created with the objective of having a reference point for the analysis of data visualisation across disciplines and beyond the standard statistical graphs. There are other galleries (Maarten Lambrechts, 2023; Manuel Lima, 2014; Yan Holtz, 2022), but to the best of my knowledge, none focused on the analysis of the approaches described in this research. Future research can leverage this collection and extend the metadata (e.g., time period, sources, interactivity) to perform further analyses, or the identified techniques used could be studied independently in more detail. The collection could be enlarged collaboratively, or automatically through the use of image similarity algorithms. It is worth noting that from the beginning of the project, when I started collecting data visualisations about two and a half years ago, almost one-fifth of the links have been lost. Many of these projects come from social networks, personal blogs, company websites, newspapers and magazines, and other media with very variable lifespans. Consequently, it would be necessary to archive all the visualisation websites in a lasting way (e.g., using web archives), that allows us to maintain and do further research taking into account their specific context, and not only the images.

4.3 The role of the digital humanities

In recent years, numerous research projects in the field of Digital Humanities have focused on other ways to visualise data (Brüggemann et al., 2020; Hinrichs et al., 2019;

Lauren Klein et al., 2016). However, in research articles, as discussed above, the use of standard statistical graphics presenting summary metrics, and network diagrams to depict relationships, continues to predominate. It is possible that by gradually integrating techniques from different disciplines that allow us to walk along this line between objectivity and subjectivity, first in a "manual" way, and then standardizing them in the tools, the possibilities available to all the disciplines will increase. The development of these tools requires researchers accustomed to working in an interdisciplinary way encompassing quantitative and qualitative methods, computational analysis and development, but also familiar with rhetorical analysis and interpretation, as well as the use of aesthetic methods and the creation of knowledge through experience. All these factors are present in the Digital Humanities, which places them in a strong position to advance the field of data visualisation.

References

Aida Horaniet Ibañez and Dagny Aurich. How can data visualization support interdisciplinary research? LuxTIME: studying historical exposomics in Belval. *Frontiers in Big Data*, 2023. doi: 10.3389/fdata.2023.1164885.

Aida Horaniet Ibañez, Daniel Richter, and Joëlla Van Donkersgoed. From The Historical Archive To The Citizens: Visualizing Census Data From Brill Street in 1922, 2023. URL https://nightingaledvs.com/from-the-historical-archive-to-the-citizens-visualizing-census-data-from-brill-street-in-1922/.

Alberto Molina Arce. One Year Connected, 2022. URL https://www.informationisbeautifulawards.com/showcase/5130-one-year-connected.

Alenka Gucek. Sleeping, 2023. URL https://www.instagram.com/p/CnKFvAvI-cJ/?img_index=1.

Alessia Musìo. Slowness, 2022. URL https://www.behance.net/gallery/148388993/Slowness.

Annette Hexelschneider. 1 Week feeling loved, 2023. URL https://twitter.com/ainvienna/status/1651825405815291904.

Liz Bravo. Week 43: Apparel Exports to US, 2020. URL https://www.makeovermonday.co.uk/week-43-2020/.

Viktoria Brüggemann, Mark-Jan Bludau, and Marian Dörk. The Fold: Rethinking Interactivity in Data Visualization. 2020.

University of Luxembourg C2DH. Journal of Digital History - Article Fingerprint, 2021. URL https://journalofdigitalhistory.org/en/issue/jdh001.

Camilla De Amicis. 5 Years, 2022. URL https://www.informationisbeautifulawards.com/showcase/5808-5years.

M.C. Chuah and S.G. Eick. Information rich glyphs for software management data. *IEEE Computer Graphics and Applications*, 18(4):24–29, July 1998. ISSN 0272-1716. doi: 10.1109/38.689658. URL http://ieeexplore.ieee.org/document/689658/.

Sarah Emery Clark. What's cookin, 2022. URL https://www.sarahemeryclark.com/work/whats-cookin.

Catherine D'Ignazio. What would feminist data visualization look like? *Feminist Studies*, 14(3):575, 2017. ISSN 00463663. doi: 10.2307/3178066. URL https://www.jstor.org/stable/3178066?origin=crossref.

Catherine D'Ignazio and Lauren F. Klein. *Data feminism*. Strong ideas series. The MIT Press, Cambridge, Massachusetts, 2020. ISBN 978-0-262-04400-4.

Johanna Drucker. Humanities Approaches to Graphical Display. *DHQ*, 5(1), 2011.

Johanna Drucker. Non-representational approaches to modelling interpretation in a graphical environment. *Digital Scholarship in the Humanities*, 33(2), July 2017. doi: 10.1093/llc/fqx034.

Johanna Drucker. *Visualization and interpretation: humanistic approaches to display*. The MIT Press, Cambridge, Massachusetts, 2020. ISBN 978-0-262-04473-8.

Johanna Drucker and Bethany Nowviskie. Temporal Modelling: Conceptualization and Visualization of Temporal Relations for Humanities Scholarship.

Ervin Vinzon. 14 Peaks, 2021. URL https://public.tableau.com/app/profile/ervin.vinzon/viz/Eight-Thousanders_16386540239820/dash.

Frederic Ferry. 1000 runs, 2020. URL https://public.tableau.com/app/profile/fredfery/viz/My1000Runs/1000Runs.

Flavio Gortana, Franziska von Tenspolde, Daniela Guhlman, and Marian Dörk. Coins, 2018. URL https://uclab.fh-potsdam.de/coins/.

Gabrielle Merite. Data Illustrations for WePresent, 2020. URL https://www.behance.net/gallery/109802031/Data-Illustrations-WePresent-Ideas-Report-2020.

Dorothy Gambrell. You could fill a museum, 2018. URL https://www.informationisbeautifulawards.com/showcase/3266-you-could-fill-a-museum.

Giorgia Lupi. Bruises. The data we don't see., 2017. URL http://giorgialupi.com/bruises-the-data-we-dont-see.

Lisa Gitelman, editor. *"Raw data" is an oxymoron*. Infrastructures series. MIT Press, Cambridge, Mass., 2013. ISBN 978-0-262-31232-5 978-0-262-51828-4.

Jonathan Gray, Carolin Gerlitz, and Liliana Bounegru. Data infrastructure literacy. *Big Data & Society*, 5(2):205395171878631, July 2018. ISSN 2053-9517, 2053-9517. doi: 10.1177/2053951718786316. URL http://journals.sagepub.com/doi/10.1177/2053951718786316.

Uta Hinrichs, Stefania Forlini, and Bridget Moynihan. In defense of sandcastles: Research thinking through visualization in digital humanities. *Digital Scholarship in the Humanities*, 34(Supplement_1):i80–i99, December 2019. ISSN 2055-7671, 2055-768X. doi: 10.1093/llc/fqy051. URL https://academic.oup.com/dsh/article/34/Supplement_1/i80/5146726.

J. Hullman and N. Diakopoulos. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12): 2231–2240, December 2011. ISSN 1077-2626. doi: 10.1109/TVCG.2011.255. URL http://ieeexplore.ieee.org/document/6064988/.

Kadambari Komandur. Data Selfi, 2022. URL https://kadambarik.myportfolio.com/data-selfi.

Kate Wong, Liz Wahid, and Jan Willem Tulp. What Scientists Have Learned from 100 Years of Bird Banding, 2021. URL https://www.scientificamerican.com/article/what-scientists-have-learned-from-100-years-of-bird-banding/.

Kim Albrecht. Quantifying the Evolution of Individual Scientic Impact, 2016. URL http://sciencepaths.kimalbrecht.com/.

Kimly Scott. #Moms Who Viz, 2023. URL https://public.tableau.com/app/profile/kimly.scott/viz/MomsWhoVizWomensHistoryMonth/MomsWhoVizWomensHistoryMonth.

Lauren Klein, Caroline Foster, Erica Pramer, Adam Hayward, and Shivani Negi. The Shape of History, 2016. URL http://shapeofhistory.net/.

Michaela Lazzaroni. Space Rocks, 2020. URL https://www.behance.net/gallery/96212805/Space-Rocks-La-Lettura-439-dataviz.

Alberto Lucas Lopez, Ryan Williams, and Kaya Berne. Migration Waves: This is what 50 years of human migration looks like, 2019. URL https://www.nationalgeographic.com/magazine/graphics/graphic-shows-past-50-years-of-global-human-migration?sf215829698=1&sf217104276=1.

Yanni A. Loukissas. *All data are local: thinking critically in a data-driven society*. The MIT Press, Cambridge, Massachusetts, 2019. ISBN 978-0-262-03966-6.

Giorgia Lupi. Data Humanism: The Revolutionary Future of Data Visualization. *PrintMag*, 2017. URL https://www.printmag.com/article/data-humanism-future-of-data-visualization/.

Giorgia Lupi and Stefanie Posavec. *Dear data*. Princeton Architectural Press, New York, first edition edition, 2016. ISBN 978-1-61689-532-7.

Maarten Lambrechts. Xenographics, 2023. URL https://xeno.graphics/.

Lev Manovich. What is visualisation? *Visual Studies*, 26(1):36–49, March 2011. ISSN 1472-586X, 1472-5878. doi: 10.1080/1472586X.2011.548488. URL http://www.tandfonline.com/doi/abs/10.1080/1472586X.2011.548488.

Lev Manovich. *Cultural analytics*. The MIT Press, Cambridge, Massachusetts, 2020. ISBN 978-0-262-03710-5.

Manuel Lima. Visual Complexity, 2014. URL http://www.visualcomplexity.com/vc/.

Martina Zunica. You Make Me Bloom, 2023. URL https://www.instagram.com/p/CrnsBoPt1Mw/?img_index=1.

Nina McCurdy, Julie Lein, Katharine Coles, and Miriah Meyer. Poemage: Visualizing the Sonic Topology of a Poem. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):439–448, January 2016. ISSN 1077-2626. doi: 10.1109/TVCG.2015.2467811. URL http://ieeexplore.ieee.org/document/7192712/.

MoMA Archives. Average Day at the Museum: The Year's Work: Annual Report to the Board of Trustees and Corporation Members of the Museum of Modern Art for the Year June 30, 1939–July 1, 1940., 1946. URL https://www.moma.org/explore/inside_out/2009/11/27/an-average-day-at-the-museum-reinterpreted-by-christoph-niemann/.

Nadieh Bremer. My Life Vacations, 2016. URL https://vacations.visualcinnamon.com/.

Chimdi Nwosu. Build Up Nepal for Viz for Social Good, 2022. URL https://public.tableau.com/app/profile/chimdi.nwosu/viz/VizForSocialGood-BuildUpNepal/Dashboard1.

Pablo Robles, Darren Long, Dennis Wong. 100 days of protests rock Hong Kong, 2019. URL https://multimedia.scmp.com/infographics/news/hong-kong/article/3027462/hong-kong-100-days-of-protests/index.html?src=arcade.

Georgia Panagiotidou, Sinem Gorucu, and Andrew Vande Moere. Data Badges: Making an Academic Profile Through a DIY Wearable Physicalization. *IEEE Computer Graphics and Applications*, 40(6):51–60, November 2020. ISSN 0272-1716, 1558-1756. doi: 10.1109/MCG.2020.3025504. URL https://ieeexplore.ieee.org/document/9201309/.

DJ Patil and Hilary Mason. *Data driven: creating a data culture*. 2015. ISBN 978-1-4919-2119-7. OCLC: 928942563.

Pentagram. Covid-19 Charts, 2020. URL https://www.pentagram.com/work/covid-19-charts/story.

Pentagram. Mindworks: The Science of Thinking, 2021. URL https://www.pentagram.com/work/mindworks-the-science-of-thinking/story.

Pentagram. A Warming Planet, 2021. URL http://giorgialupi.com/bullettin-of-the-atomic-scientists.

South China Morning Post. Arrested Hong Kong protesters, 2020. URL https://multimedia.scmp.com/infographics/news/hong-kong/article/3088009/one-year-protest/index.html.

Samuel Parsons. Viz for Social Good. Self Reflection 2017-2020, 2020. URL https://public.tableau.com/app/profile/sparsonsdataviz/viz/VizForSocialGood-SelfReflection2017to2020/VizForSocialGood-SelfReflection2017to2020.

Sarah Stern. Moon Mood, 2022. URL https://www.informationisbeautifulawards.com/showcase/5525-moon-mood.

Scientific American. Relativity's Influence Is Still Going Strong on Its 100th Birthday, 2015. URL https://www.scientificamerican.com/article/relativity-s-influence-is-still-going-strong-on-its-100th-birthday/.

Silvia Romanelli. Endangered Languages, 2022. URL https://www.behance.net/gallery/128273653/Endangered-Languages.

Sonja Kuijpers. 20 Year History of National Team Appearances in the FIFA World Cup, 2018. URL http://www.studioterp.nl/world-cup-national-team-appearances-datavisualization/.

Sonja Kuijpers. A View on Despair, 2019. URL https://www.studioterp.nl/a-view-on-despair-a-datavisualization-project-by-studio-terp/.

Stefanie Posavec. Inch Time - The Floating World, 2011. URL https://www.stefanieposavec.com/inch-time.

Stefanie Posavec. Open Data Playground, 2014. URL https://www.stefanieposavec.com/open-data-playground.

Stefanie Posavec. Papworth Hospital, Inpatients Ward, 2019. URL https://www.stefanieposavec.com/papworth.

Stefanie Posavec. Data Murmurations: Points in flight (People Like You), 2023. URL https://peoplelikeyou.ac.uk/system/.

Bruno J. Strasser and Paul N. Edwards. Big Data Is the Answer ... But What Is the Question? *Osiris*, 32(1):328–345, September 2017. ISSN 0369-7827, 1933-8287. doi: 10.1086/694223. URL https://www.journals.uchicago.edu/doi/10.1086/694223.

Domestic Data Streamers. The mood test, 2013. URL https://domesticstreamers.com/projects/the-mood-test/.

Lucy Suchman. Do categories have politics? page 14, 1993.

Syennie Valeria. Emotional Topographies. Cities., 2022. URL https://www.behance.net/gallery/140603619/Emotional-Topographies-Cities?tracking_source=search_projects%7Cinformation%20visualization%20data%20viz.

Edward R. Tufte. *The visual display of quantitative information*. Graphics Press, Cheshire, Conn, 17. print edition, 1999. ISBN 978-0-9613921-0-9.

Valentina D'Efilippo and Lucia Kocincova. Me Too, 2017. URL http://metoomentum.com/about.html.

Yan Holtz. Dataviz Inspiration, 2022. URL https://www.dataviz-inspiration.com/.