Showing Data about People: A Design Space of Anthropographics

Luiz Augusto de Macêdo Morais, Yvonne Jansen, Nazareno Andrade, and Pierre Dragicevic

Abstract—When showing data about people, visualization designers and data journalists often use design strategies that presumably help the audience relate to those people. The term *anthropographics* has been recently coined to refer to this practice and the resulting visualizations. Anthropographics is a rich and growing area, but the work so far has remained scattered. Despite preliminary empirical work and a few web essays written by practitioners, there is a lack of a clear language for thinking about and communicating about anthropographics. We address this gap by introducing a conceptual framework and a design space for anthropographics. Our design space consists of seven elementary design dimensions that can be reasonably hypothesized to have some effect on prosocial feelings or behavior. It extends a previous design space and is informed by an analysis of 105 visualizations collected from newspapers, websites, and research papers. We use our conceptual framework and design space to discuss trade-offs, common design strategies, as well as future opportunities for design and research in the area of anthropographics.

Index Terms—Anthropographics, design space, empathy, compassion, prosocial behavior.

1 Introduction

When communicating data about people, information designers and data journalists regularly create visualizations[1] meant to foster an emotional connection with the persons whose data is represented. Figure 1a shows one example where the reader can see the story and personal information of each person who died in a public mass shooting in the USA from 1966 to 2019. Meanwhile, the visualization in Figure 1b conveys the hardship of the life of refugees by narrating the story of S.W.G., a 26 years old refugee who left Pakistan and spent 651 days before arriving at his final destination. Both visualizations were crafted to bring readers closer to the persons whose data is visualized.

The practice of visualizing data about people in a way that helps the audience relate has been called *anthropographics* [3], [4]. Boy at al. [3] first coined this term as an abbreviation for *anthropomorphized data graphics* — visualizations with human-shaped symbols. The term was then used to refer more generally to *"visual strategies to make the connection between data and the humans behind them more direct and, hopefully, more empathic"* [4]. Similarly, the term *data humanism* was coined to refer to a range of visualization design practices intended to promote humanistic values [5]. In this article, we reconcile these different views by defining anthropographics as:

> *visualizations that represent data about people in a way that is intended to promote prosocial feelings (e.g., compassion or empathy) or prosocial behavior (e.g., donating or helping).*

Like Boy et al. [3], this definition refers to a class of visualizations rather than a set of design strategies, and like Bertini [4], it generalizes beyond the use of human-shaped symbols. Based on findings that empathy is not necessarily conducive to helping behavior [6], the definition considers prosocial behavior as a potential design goal, as well as other prosocial feelings such as compassion, which is presumably more directly connected to prosocial behavior [6].

Visualization designers have explored many strategies for creating anthropographics. A popular strategy is to use human-shaped symbols [3] (e.g., see Figure 1a). Other strategies include the use of text annotations to make each person appear unique [3] (e.g., see the top left of Figure 1b), the use of visual metaphors (e.g., a red bar chart to symbolize blood or death), or the representation of persons as individual marks rather than aggregated data (e.g., the marks in Figure 1a) [7], [8]. So far, only a few of these strategies have been empirically tested, and even though initial studies have been mostly inconclusive [3], there is an enormous potential for information visualization research.

Anthropographics is a rich and growing area, but the work so far has been scattered. As such, the visualization community presently lacks a precise language for discussing anthropographics, as well as an overview of design strategies and opportunities for design. The goal of this paper is to contribute to fill this gap by proposing a design space and conceptual framework that are meant to help researchers and practitioners reason and communicate about anthropographics. We aim at providing conceptual foundations and a language to facilitate the design, critique, comparison, and empirical evaluation of anthropographics.

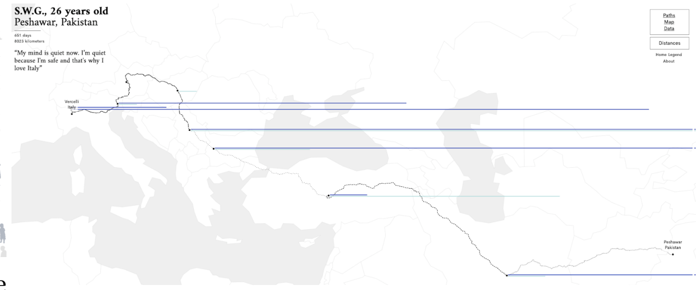
Our framework extends the work of Boy at al. [3], who coined the term "anthropographics" and proposed an initial design space. Their primary focus, however, was not on introducing a comprehensive design space but on reporting empirical studies. In contrast, the present article is fully dedicated to laying out a design space of anthropographics. It extends Boy et al.'s work by introducing useful basic terminology, by refining and extending their design space,

- *Luiz Augusto de Macêdo Morais and Nazareno Andrade are with Universidade Federal de Campina Grande. E-mail: luizaugustomm@lsd.ufcg.edu.br*
- *Yvonne Jansen is with Sorbonne Université, France.*
- *Pierre Dragicevic is with Inria, France.*

Manuscript received December 2, 2019; revised ???.

1. In this article, we employ *visualization* in a broad sense that also includes infographics.

(a) [PERMISSION PENDING] Mass Shooting Statistics in the United States
Source: Washington Post [1]



(b) [PERMISSION PENDING] Stories Behind a Line
Source: storiesbehindaline.com [2]

Fig. 1: Examples of anthropographics: (a) the persons who died due to public mass shootings in the USA from 1966 to 2019; (b) the story of a refugee who fled from his home country to Italy.

and by providing a range of illustrations and examples.

Our design space consists of seven design dimensions that can be reasonably believed to have some effect on prosocial feelings or behavior, namely: *granularity*, *specificity*, *coverage*, *authenticity*, *realism*, *physicality*, and *situatedness*. These dimensions were identified by examining a collection of 105 communicative visualizations that convey data about people, some of which are anthropographics and some of which are not (e.g., some are conventional statistical charts). After describing the design space and the conceptual framework it is based on, we identify recurrent combinations of dimensions (i.e., families of visualization designs), and discuss opportunities for future research and design.

Our conceptual framework is meant to be *descriptive* and *generative* [9]: it has been devised to help designers think more clearly about existing and possible designs. However, it is not meant to be *evaluative* [9]: it cannot help predict what designs will work best, and cannot prescribe what designs to use. More studies are needed before an evaluative framework is viable. Descriptive frameworks such as ours can, however, help researchers reach that goal faster by facilitating the design of informative empirical studies.

2 Methodology

The starting point of our work was the design space of anthropographics by Boy et al. [3]. We extended this design space based on a thorough analysis of a corpus of visualizations, and based on data visualization and psychology research, as well as blog posts from practitioners. This section describes how we collected our corpus of visualizations, and how we used this corpus to inform our design space.

2.1 Scope

We collected a total of 105 visualizations that convey data about people (see Supplementary Material A for a fixed list, or the interactive website for an updated list[2]). The collection process started with visualization repositories (e.g., flowingdata.com and dataphys.org/list) and proceeded through a snowball sampling to include blogs and newspapers. We also included visualizations from an additional material provided by Boy et al. [3]. Finally, the collection was complemented by charts found in social media, and by visualizations from our own practice. Three criteria were considered to include a visualization:

1) *The visualization shows data about people*: examples include statistics about people who died due to a gunshot, the path taken by refugees to escape from war, or the characteristics of women who fought breast cancer.
2) *The data visualized is about existing people*. This condition led us to exclude, for example, simulation-based visualizations where imaginary individuals are created to convey life expectancy data or causes of death[3].
3) *The visualization was published with a communicative intent*. This criterion includes infographics from newspapers, for example, but excludes visualizations produced for purely analytic purposes, for which there is no communicative intent.

The resulting set of 105 visualizations was compiled between October 5, 2018, and July 30, 2019. Among all visualizations, 29% are from blog posts, 27% from newspapers or magazines, 21% from project web pages, 10% from public exhibitions, 10% from books or reports, and 3% from academic papers. The collection includes both interactive and static visualizations. It also includes both anthropographics and non-anthropographics (i.e., visualizations that were likely not designed to promote prosocial feelings or behavior). Doing so allowed us to lay out a more comprehensive design space, and to consider the full continuum between anthropographics and non-anthropographics.

2.2 Development of the Design Space

We started with the set of dimensions proposed by Boy et al. [3], and progressively iterated over them by characterizing each visualization from our collection according to each of the dimensions. This process helped us establish the characteristics and boundaries of each of the dimensions, and identify gaps. In parallel to defining and refining the dimensions of the design space, we developed a conceptual framework and set of elementary definitions (described in Sections 3.1 and 3.2) in order to have a firm conceptual ground on which to rest our final design space.

As our corpus contains a broader range of visualizations than initially considered by Boy et al., the dimensions that could not properly fit the corpus or could not be

2. [*Not available yet*] Interactive list of visualizations: www.examplewebsite.com.

3. See, for example, *Visualizing smoking risk*: https://bit.ly/2I78lix

easily operationalized were removed (see the justifications in subsection 3.7). During the process, we also came up with new dimensions, partly inspired from past literature in psychology (e.g., [10]) and data visualization (e.g., [11]).

After a seemingly stable set of definitions and dimensions was established and the corpus of visualizations had been fully categorized by the first author, we performed a multi-coder evaluation. We collectively wrote a codebook, which the last three authors used to classify a random sample of 17 visualizations. Afterwards, we discussed difficulties and discrepancies in the codings, and iterated one last time on the design space and concept definitions.

The resulting design space consisted of three dimensions from Boy et al.'s design space [3] (see subsection 3.7 for more details), and four new dimensions.

3 Design Space

In the next sections, we first introduce basic terminology and a conceptual framework that will serve as a foundation for our design space, and then describe and motivate each design space dimension.

3.1 General Visualization Concepts

We assume for simplicity that all **datasets** are flat tables [12]. A **data item** (or simply **item**) is "an individual entity that is discrete", and which corresponds to a row in the table [12]. In our framework, a data item always corresponds either to *a person* or *a group of persons*. Meanwhile, a **data attribute** (or simply **attribute**) is "some specific property that can be measured, observed, or logged" [12]. Examples include the salary, height, or name of a person, or the average salary or height of a group of persons.

Visualizations are human-readable representations of data items and data attributes. Here too, our terminology largely follows previous literature. The main building block of a visualization is the **mark**, which is an element that represents a data item [12], [13], [14], [15]. A mark can either consist of a single graphical primitive (e.g., a point, line, or area [16]) or a combination thereof (e.g., a glyph or an icon) [17]. Meanwhile, **perceptual channels**[4] are properties of a mark that can be varied to convey the value of data attributes. They include elementary graphical attributes (e.g., size, value, color, texture, orientation, and shape [16]), non-graphical attributes (e.g., weight [19]), as well as complex combinations of attributes (e.g., when using icons to convey categorical attribute values).

3.2 New Concepts

We assume that a visualization is created by a designer with an intent, i.e., a message to convey. For any such visualization, we define the **reference population** as the set of all people who are the subject of the visualization's message, i.e., all the persons the visualization designer chose to tell a story about. For example, if a visualization tells a story about WWII casualties, we can assume that the reference population is the set of all people who died due to WWII, even if not all of these people are represented in the visualization, and even if not all of these people are known. Thus, the reference population is a conceptual set, which is not necessarily fully visible in the visualization, and about which we typically can only speculate.

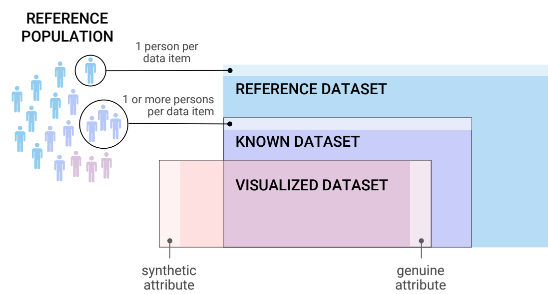


Fig. 2: The three conceptual datasets we use to capture the message and information conveyed by an anthropographic.

In our conceptual framework, the message and the information conveyed by a visualization involve three datasets (see Figure 2). The first one, the **reference dataset**, is purely conceptual as it consists of all the people from the reference population – each mapped to an item, plus all of their characteristics, each mapped to an attribute. An example of a reference dataset is all migrants who died in the Middle East between 2015 and 2019, together with all the information possibly imaginable about them.

The **known dataset** is the subset of the reference dataset available to the designer, possibly aggregated in such a way that each data item corresponds to a group of people instead of a single person. It consists of a finite number of items and attributes, for example, the data about migrants who are known to have died in the Middle East between 2015 and 2019, as reported in the Missing Migrants dataset[5]. A known dataset is typically a regular dataset (e.g., a csv file), although it might partly reside in analog storage artefacts (e.g., a piece of paper with handwritten numerals) or in the designer's mind (e.g., if the designer is working from memory or logging information directly into the visualization).

Finally, the **visualized dataset** consists of all the information — items and attributes — that is represented in the visualization, or in other words, all the information that can potentially be extracted from the visualization. It is a subset of the known dataset with potentially extraneous information added by the designer, for example, for aesthetic reasons or for storytelling purposes. We formalize this by introducing the notions of genuine and synthetic attributes.

Genuine attributes are attributes from the visualized dataset that originate from the known dataset. For example, a genuine attribute could be the age of the migrants who died in the Middle East between 2015 and 2019, as reported in the Missing Migrants dataset. On the other hand, **synthetic attributes** are attributes in the visualized dataset that do not originate from the known dataset. For example, a designer may assign an arbitrary gender to each person (through the use of a male or female silhouette) or may give them fictitious

4. This term generalizes the notion of visual channel [12] or visual variable [18] to non-visual data representations [19].

5. Missing Migrants website: https://missingmigrants.iom.int/

names (e.g., Figure 15). Synthetic attributes may occasionally match the attributes from the reference dataset, either by accident or because the designer made an informed guess. However, we generally cannot assume synthetic attributes to reflect reality, since they are by definition not known. Note that it is not necessarily visible in a visualization whether an attribute is genuine or synthetic.

Another way to characterize attributes is by their distinctiveness, which captures how much information they convey. The **distinctiveness** of an attribute or a set of attributes is the extent to which it allows to distinguish people or groups of people from each other. When data items are individuals, attributes with low distinctiveness are ones whose values are generally shared by many people (e.g., sex, age, or country of origin). Meanwhile, attributes with high distinctiveness are ones that few people have in common (e.g., full names or photographs). Many attributes lie somewhere in the middle of the distinctiveness continuum. When data items are groups of people (e.g., demographic groups or countries), distinctiveness refers to the extent to which the attributes allow to distinguish those groups from each other.

Finally, attributes can be represented in a visualization in two major ways. **Encoded attributes** are attributes from the visualized dataset that are mapped to perceptual channels. One example is mapping people's age to the height of bars. In contrast, **literal attributes** are attributes from the visualized dataset that are presented either in a literal form or using written prose or numerals. One example is fully spelling out somebody's age, e.g., with a label stating "22 years old". Some attributes such as names or photos do not easily lend themselves to visual encoding, and thus are almost always represented as literal attributes.

3.3 Introducing the Design Space

The rest of this section describes the dimensions that make up our design space. We illustrate the dimensions with a series of minimalistic visualizations, all of which assume the same known dataset consisting of ten persons and four attributes: whether each person prefers cats or dogs, their gender, their name, and a photo of their face (see Table 1). The reference population contains the same ten people.

Name	Gender	Dogs or cats?	Photograph
Yousef	Male	Dogs	(image)
Amir	Male	Dogs	(image)
Alana	Female	Dogs	(image)
Olaf	Male	Dogs	(image)
Jeremy	Male	Dogs	(image)
João	Male	Dogs	(image)
Fatima	Female	Cats	(image)
Nadia	Female	Cats	(image)
Asha	Female	Cats	(image)
Michal	Male	Cats	(image)

TABLE 1: Fictional dataset used in Figures 4, 6, 9, 14, 16, 19, and 22.

The dimensions of our design space fall into two broad groups. The *what is shown* group corresponds to dimensions concerned with the information represented in the visualization (e.g., whether all people from the reference population are represented). Meanwhile, dimensions from the *how it is shown* group describe the way information is represented on the visualization (e.g., whether the marks look like people).

3.4 What is Shown

There are four dimensions related to what information and how much information is presented in a visualization.

3.4.1 Granularity

Granularity refers to the extent to which the persons in the visualized dataset are mapped to separate data items – and equivalently, to separate marks on the visualization.

In a visualization with **low granularity**, each mark corresponds to a *group of persons* who have one or more attribute values in common (see Figure 4-left). An example is depicted in Figure 3, where each segment of a different color represents the sum of immigrants in a period of time. Statistical charts where each mark represents a large number of people (thousands or millions) are typical examples of low-granularity visualizations. In such visualizations, length or area are commonly used to encode cardinality (e.g., bar charts as in Figure 4-left, bubble charts, or area charts). Position encoding is also occasionally used (e.g., in line charts of population growth).

In a visualization with **intermediate granularity**, each mark maps to a *fixed number of persons*, this number being greater than one (see Figure 4-center). Many so-called Isotype visualizations [22], [23] fall into this category. Figure 5 shows an example of an Isotype visualization, where each mark represents one million soldiers. Since one million is a large number, the granularity is relatively low in absolute terms, but it is higher than if people were fully aggregated according to the three categories "killed", "wounded" or "others returning home". While most Isotype-like designs show absolute counts, a variation over this design uses marks to show rounded percentages (i.e., the counts from all marks sum up to 100 people).

In a visualization with **maximum granularity**, each data item corresponds to *a single person* (as in Figure 4-right). For example, the visualization in Figure 7 has maximum

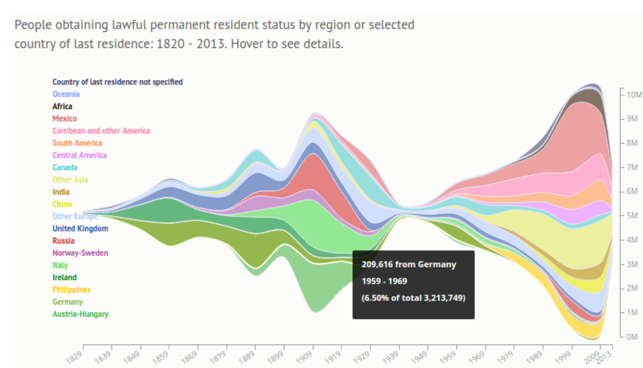


Fig. 3: [PERMISSION PENDING] *200 Years of Immigration to the US*. This stream graph shows people who obtained lawful permanent resident status by region from 1820 to 2015. Each color corresponds either to the country of last residence or a region. The area represents the number of immigrants who moved to the U.S. over the years. When hovering over an area, the reader gets information about the total number of immigrants within a period of 10 years until the selected date. Source: Insightful Interaction [20].

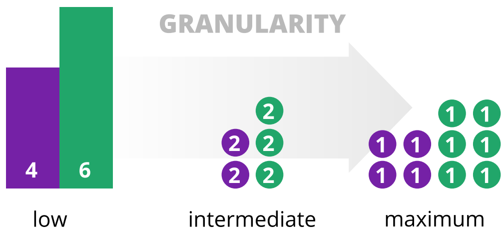


Fig. 4: *Granularity Continuum*. *Low*: each mark represents a group of people aggregated by one or more attributes (here, their preference for dogs or cats). *Intermediate*: each mark represents a fixed number of persons (here, two). *Maximum*: each mark corresponds to a single person.

Fig. 6: *Specificity continuum*. *Low*: items tend to be very similar to each other and most of them cannot be distinguished. *High*: the reader can perfectly distinguish all data items, thanks to two highly distinctive attributes: the name and a photograph of the person.

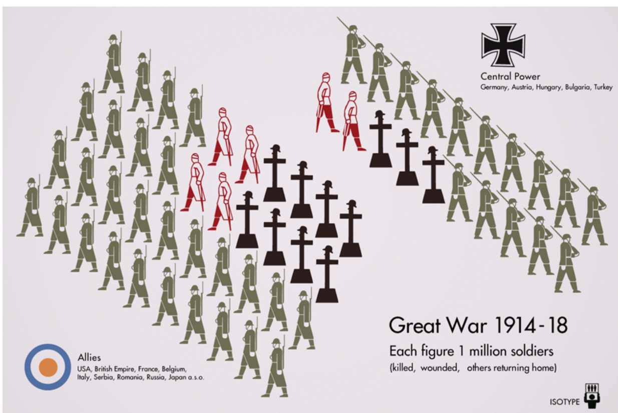


Fig. 5: [PERMISSION PENDING] *The Great War 1914-18*. An Isotype visualization where each mark represents 1 million soldiers. Black crosses represent soldiers who died during the war, while soldiers in red came back wounded, and soldiers in green returned without major physical injuries. The soldiers on the left are the Allies and the ones on the right belong to the Central Power. Source: LA Worrell [21].

granularity because each data item stands for a different person who died by a gunshot in 2019.

Note that in the case of low granularity, the number of persons is typically encoded using perceptual channels, whereas in intermediate and maximum granularity, it is mapped to the number of marks.

Using visualizations that represent people as individual marks allows designers to convey specific information about each person, be them genuine attributes detailing those portrayed, or synthetic attributes to humanize the data (e.g., some of Boy et al.'s designs [3]). There is also a belief that representing people using single marks can make readers empathize with the persons represented [8]. However, maximum granularity may have drawbacks with large datasets: representing thousands of people as different marks demands a large space or forces the designer to reduce each mark to a speck with little details.

It is still necessary to investigate in which scenarios there are relevant benefits in using visualizations with higher granularity. The only study that investigated visualizations with intermediate granularity in the context of human rights was not able to confirm the hypothesis that higher granularity leads to more empathy [3]. More studies are needed to examine the role of granularity in promoting prosocial feelings or behavior.

3.4.2 Specificity

The **information specificity** of a visualization (or simply **specificity**) corresponds to *how distinctive* is the *entire set of attributes* in the visualized dataset (see subsection 3.2 for the definition of distinctiveness). The more the attributes allow to distinguish between data items (either individuals or groups of people) the higher the visualization's specificity. All attributes contribute to a visualization's specificity, whether they are encoded or literal, and whether they are genuine or synthetic.

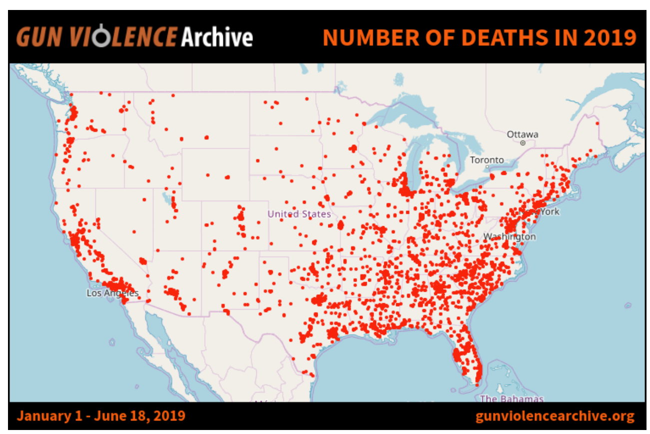


Fig. 7: [PERMISSION PENDING] *Gun violence Chart of Deaths in 2019*. Depicts all deaths caused by guns in the USA from January to June, 2019. Each red dot is a victim, and its placement indicates where the person was killed. Source: Gun Violence Archive [24].

In a visualization with **low specificity**, items tend to be visually very similar to each other and most of them cannot be distinguished (see Figure 6-left). The Isotype visualization in Figure 5 is an example of a low-specificity visualization, where the low distinctiveness of the visualized attributes (survival status and side) contributes to making the soldiers look rather deindividualized. The gun violence visualization (Figure 7) also has somehow low specificity, because the only visualized attribute is the location where each person died. In this case, location can be thought to have relatively low distinctiveness because there are spots on the map where more than one person might have died, which makes it hard to distinguish the victims. Had the map been higher-resolution or zoomable, specificity would have been higher.

In a visualization with **intermediate specificity**, attributes in the visualized dataset allow the reader to easily distinguish many data items from each other. A typical approach is to use glyphs to represent multidimensional data about individual people. For example, in Figure 8, the visualization shows six attributes. Although each attribute has low distinctiveness, once combined together, the attributes give some sense of individuality to each person.

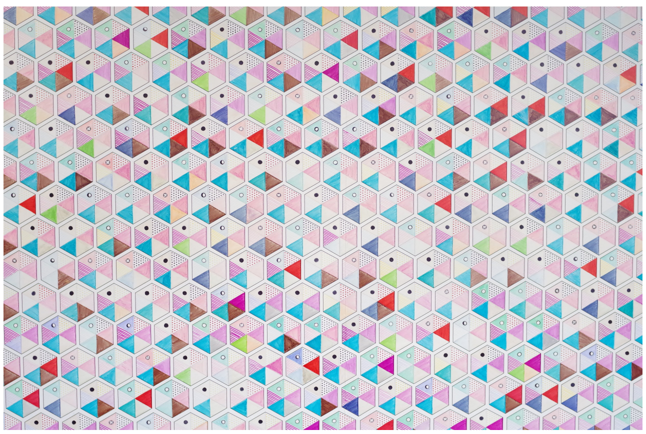


Fig. 8: [PERMISSION PENDING] *Data wallpaper*. Each glyph represents 6 questions answered by a visitor to a store, ranging from how they see the future of work to how they unlock their creativity. The designer initially installed an empty visualization template as a wallpaper in the store, and the store employees filled the hexagons as they were receiving visitor responses. Source: Giorgia Lupi [25].

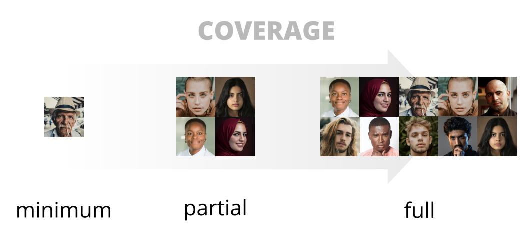


Fig. 9: *Coverage continuum*. *Minimum*: a single person from the reference dataset is visualized. *Partial*: a subset of people from the reference dataset is visualized. *Full*: all people from the reference dataset are visualized.

Finally, in a visualization with **high specificity**, the attributes in the visualized dataset allow the reader to perfectly distinguish all data items (see Figure 6-right). One example is shown in Figure 12, where people's faces and other physical characteristics make each person unique. A visualization can also have high specificity when it shows a set of attributes that is so large that the data is necessarily unique to each person. In Figure 10, for example, the visualization shows the

Fig. 10: [PERMISSION PENDING] *Bruises: the Data We Don't See* shows the progress of a child in coping with an auto-immune disease. Each petal is a day. Red dots are platelet counts (the disease destroys them). Colors in the petals represent various events such as bleeding, using medications, or positive feelings. The texts around the petals are the mother's notes about the day. Source: Giorgia Lupi [26].

story of a child coping with an auto-immune disease. Since in our framework data items are always people or groups of people, the visualized dataset is comprised of a single data item, and thus the visualization consists of a single mark that can be thought of as an extremely complex glyph.

Evidence suggests that showing pictures and details of victims is linked to a higher likelihood of donating to charity [27], [28]. Although the study from Boy et al [3] has failed to find an effect for intermediate-specificity visualizations with intermediate granularity, it remains possible that other designs could promote prosocial feelings and behaviors. Although privacy issues often make it hard to reach high specificity in a visualization [3], it is sometimes possible to reach intermediate levels of specificity without compromising privacy, for example by showing people's first names, age, and other non-identifiable data, as in Figure 11.

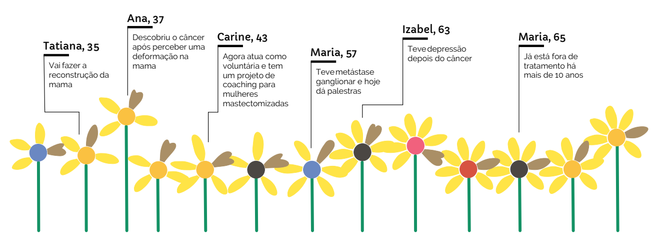


Fig. 11: *Cancer is not always the end* shows data about 13 women with breast cancer who has been in remission for years. Each flower is a woman and each part of a flower shows a characteristic of the woman such as her age or how she dealt with the treatment. Information about some women is also shown by annotations. Source: Oncoguia [29].

3.4.3 Coverage

Coverage corresponds to the extent to which the visualized dataset includes the persons from the reference dataset. Although in many visualizations it may not be possible to identify with certainty the reference population considered by the designer and consequently the coverage, it remains possible to speculate, and more generally, to reason about coverage on a theoretical level.

Fig. 12: [PERMISSION PENDING] *100% Paris* is composed of 100 Parisians who have been selected to represent the Paris population. Depending on personal questions asked during the show, the persons moved to the left or the right of the green circle, depending on whether their answer was "me" (on the left) or "not me" (on the right). Source: Rimini Protokoll (Vimeo) [30].

In a **minimum coverage** visualization, the reference population consists of more than one person but the visualized dataset only contains data about *a single* person from that population.

In a visualization with **partial coverage**, the visualized dataset contains data about a *subset* of people from the reference dataset. The people may be selected at random or chosen according to some specific attributes. For example, the visualization in Figure 11 delivers a message about women who have breast cancer in Brazil and are in remission, but only includes data about 13 of these women. Another example of a partial coverage visualization is shown in Figure 1b: "The Stories Behind a Line" starts by showing six lines, each of which contains initials of a refugee. The user can click on a line to see the story of a person who left their home seeking a better life. In this example, the six individuals were presumably chosen in an arbitrary manner, and used as illustrations in order to tell a broader story about the arduous life of thousands of refugees around the world.

In a **full coverage** visualization, the visualized dataset contains *all* the persons from the reference dataset. See Figures 1a, 7, and 13 for examples.

While full coverage is the most straightforward design choice, partial coverage is necessary when datasets are incomplete (e.g., not all bodies have been found in a disaster), or may be a more convenient choice when visualizing rich personal data that needs to be explicitly gathered from the people (e.g., Figure 11). A related advantage of partial and minimum coverage is that they reduce the number of marks, and thus leave space to show richer data about individuals. Furthermore, there is evidence suggesting that telling the story of a single suffering individual can better promote compassion than telling the story of an entire group [27]. Similarly, studies have suggested that as the number of suffering people increases, people feel less empathy for them and donate less [10]. Therefore, it is possible that visualizations with partial or minimum coverage could help observers be

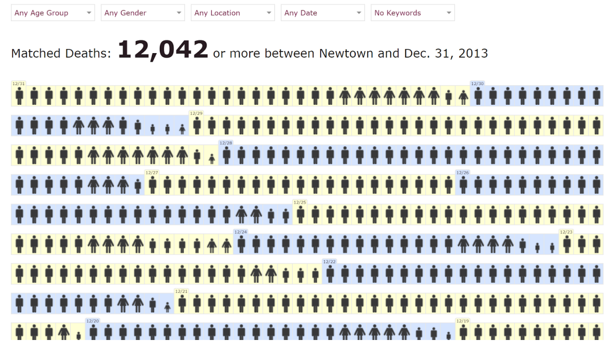


Fig. 13: [PERMISSION PENDING] *How Many People Have Been Killed by Guns in USA Since Newtown's attack*. This visualization shows people who died due to a gun shot between the Newtown attack in 2012, and December, 2013. Each icon is a person. Sex and age are shown through the icon's shape. Additional information about each person can be obtained by clicking on their icon. Source: Slate [31].

more compassionate about suffering populations. However, it remains necessary to test this hypothesis empirically.

3.4.4 Authenticity

Authenticity refers to the proportion of genuine attributes in the visualized dataset.

Fig. 14: *Authenticity continuum*. The two visualizations have the same encoded attributes: whether people prefer cats or dogs and their gender. *Partial*: the visualization contains synthetic attributes such as a person in a wheelchair or differences in body height, information not in the known dataset. *Full*: two types of silhouettes are used depending on the gender indicated in the known dataset.

A visualization with **partial authenticity** contains both genuine and synthetic attributes (terms defined in subsection 3.2). The more visualized attributes are synthetic, the less authentic the visualization is. In Figure 14-left, the designer gave the marks different silhouettes to make them look unique: we can see for example a child, and a person in a wheelchair. The designer also annotated an individual with personal information. However, none of this information is in the known dataset (see Table 1) which makes these attributes *synthetic*. A real example of a partially authentic visualization is shown in Figure 15. The only genuine attribute is whether a person is above or below the poverty line in 2010, whereas the name and gender of the persons are synthetic attributes that were probably used to increase information specificity and make the persons look more unique, or to increase the visual appeal of the visualization.

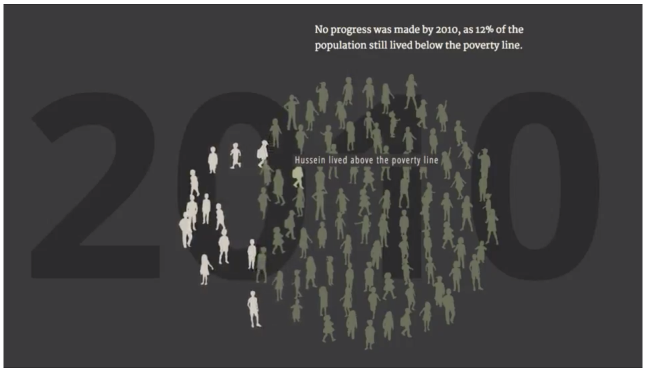


Fig. 15: [PERMISSION PENDING] *Poverty in Syria* shows the proportion of children below (darker color) or above the poverty line (lighter color) in Syria, in 2010. Each mark is a fixed number of children. Hovering over a mark shows synthetic information about a person. Source: Boy et al. [3].

In a visualization with **full authenticity**, all visualized attributes are genuine. In Figure 14-right, the second visualization is fully authentic because all the information presented comes from the known dataset. Although it uses realistic silhouettes like the visualization on the left, only two types of silhouettes are used and they serve to encode gender, an attribute that is in the known dataset.

While full authenticity is the most obvious design choice, designers sometimes use synthetic attributes in visualizations about people such as in Figure 15 or in Visualization 67 in Supplementary Material A. In those examples, the designers used anthropomorphic marks with different genders and ages that likely do not originate from the known dataset, possibly as an attempt to promote compassion. However, it is possible that this technique can backfire as a result of readers feeling manipulated, possibly causing them to doubt even the genuine information and the visualization as a whole. As far as we know, this technique and its possible trade-offs have never been explicitly discussed in the information visualization literature.

3.5 How it is shown

This section describes design dimensions that capture how information is represented. All of the dimensions here are properties of *marks*. The properties of marks are often coherent across a visualization, and thus a visualization can be characterized with respect to how its marks are represented.

3.5.1 Realism

Realism refers to the degree of resemblance of the visualization's marks to actual persons.

A visualization with **low realism** represents people or groups of people using *symbolic marks* that are non-anthropomorphic, i.e., they bear no resemblance with a human (see Figure 16-left). Such marks include dots, bars, abstract glyphs, or shapes that evoke inanimate objects. Figures 3, 8, and 11 are examples of low-realism visualizations.

A visualization with **intermediate realism** is made of *pictorial anthropomorphic marks*. Examples of such marks are simple icons or human silhouettes, as shown in Figure 16-center. The visualizations from figures 13 and 15 are also visualizations with intermediate realism, since they are composed of human-shaped icons or silhouettes.

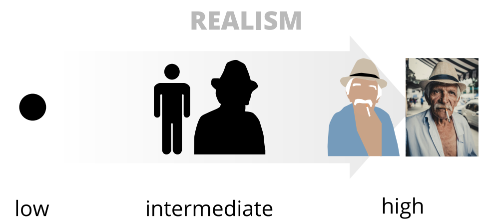


Fig. 16: *Realism continuum*. *Low*: the marks do not evoke a person. *Intermediate*: the marks are simplified depictions of a person. *High*: the marks are realistic depictions of a person.

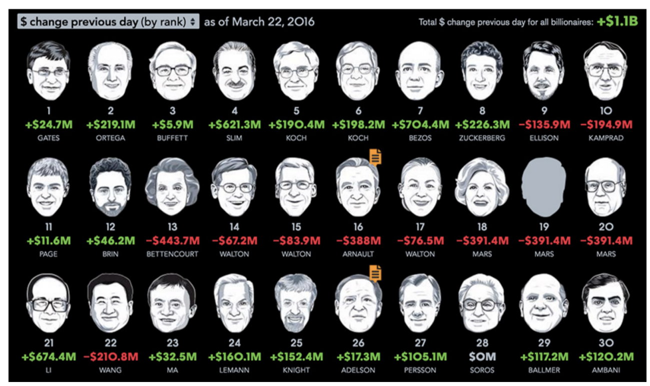


Fig. 17: [PERMISSION PENDING] *The Billionaire Index* presents data about the 30 richest persons in the world. Below each mark is an index (corresponding to the position of the person in the ranking of billionaires), the last change (in dollars) and the person's name. Color encodes whether the last change was positive or negative. Source: Bloomberg [32].

A visualization with **high realism** is made of *realistic anthropomorphic marks*, which closely resemble an actual person. Figure 16-right shows two cases of realistic marks: a detailed drawing of a person and a photograph. A realistic anthropomorphic visualization is shown in Figure 17, where the marks are detailed drawings of the most rich people in the world. Other examples of realistic marks include 3D avatars, physical sculptures, and even real persons, such as the data physicalizations from Figures 12 and 18.

Naturally, the realism of anthropomorphic marks is best thought of as a continuum, as real persons are higher on the realism spectrum than photographs, which are themselves higher than drawings, which are in turn higher than simple icons. Designers typically use anthropomorphic marks to reinforce the fact that the data is about real persons. It has been hypothesized that doing so could promote empathy [3], [7], [34], and that the more realistic the marks are, the more effective they could be at promoting empathy [3]. However, none of these hypotheses has been experimentally confirmed.

Fig. 18: [PERMISSION PENDING] *San Francisco Gay Men's Chorus representing AIDS deaths until 1993*. The men in white are the surviving members of the original San Francisco Gay Men's choir, while the persons in black with their back turned represent the members lost to AIDS. Source: San Diego Gay and Lesbians News [33].

3.5.2 Physicality

Physicality refers to the degree to which a visualization's marks are embodied in physical objects as opposed to shown on a flat display [19]. Figure 19 illustrates the physicality continuum.

Fig. 19: *Physicality continuum*. *Low*: the marks are shown on a flat medium. *Maximum*: the marks are physical entities.

In a visualization with **low physicality**, the marks are shown on a flat medium, such as a computer screen or a sheet of paper. All visualizations designed for the web or for magazines (e.g., Figures 5, 15, and 17) fall in this category. The visualizations from figures 8 and 20 also have low physicality because the former is drawn on a flat wall, while the latter is shown on a wall clock with a printed background.

Visualizations with **intermediate physicality** are characterized by marks that have both physical and virtual qualities. While we could not find an example featuring data about people, examples exist for other types of datasets. For example, the Emoto installation [11] shows tweets, of which some attributes are encoded in physical shape while others are video-projected.

In visualizations with **maximum physicality**, the marks are physical objects or actual persons. Examples are shown in Figure 21, where the marks are grains of rice, and in Figures 12 and 18, where the marks are real persons.

Le Goc et al. [11] hypothesize that it is easier to empathize with people when they are represented by physical objects than when they are represented by virtual objects. Yet it is so far unknown whether using physical marks may indeed increase prosocial feelings such as empathy or compassion.

Fig. 20: *Activity Clock* shows the aggregated presence of persons in a lab from 8 AM to 8 PM during a 3-year period. Each bar is a 15-minute bin with the 10th and 90th percentile of the number of persons in that time. White dots and the color encode the median number of people in the corresponding time. Source: Morais et al. [35].

Fig. 21: [PERMISSION PENDING] *Of All the People in All the World: Stats with Rice* is an installation using grains of rice to show various statistics about people (one grain per person). Source: Flickr [36].

3.5.3 Situatedness

Situatedness refers to how spatially close the mark's physical presentation [37] is — or was[6] — to its physical referent [13]. In the context of anthropographics, the physical referents are the persons the data describes. Figure 22 illustrates the situatedness continuum.

In visualizations with **low situatedness**, which we will also refer to as *non-situated*, the marks are either presented far from the persons they represent, or their physical location is not under the control of the designer. This includes all visualizations designed for magazines or the web. Figures 13 and 7 are examples of non-situated visualizations because they are displayed on computer screens which are in most cases far from the represented victims.

In a visualization with **intermediate situatedness**, the marks are either presented at a location where the persons

6. As we will see, we generalize Willet et al.'s [13] notion of situatedness by also considering spatial relationships in the past.

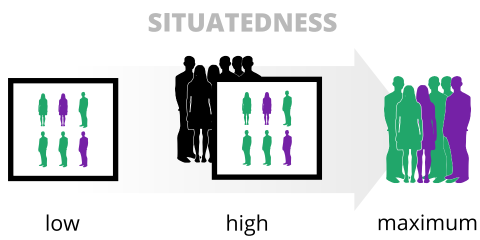


Fig. 22: Situatedness continuum. *Low*: the marks are presented far from the persons they represent, or their physical location is not under the control of the designer. *High*: the marks are presented close to the persons they represent (the persons in black). *Maximum*: the marks are the persons they represent.

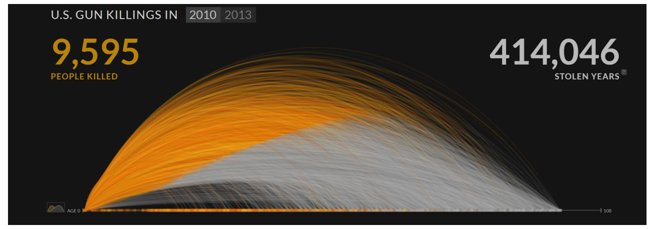


Fig. 23: [PERMISSION PENDING] *U.S. Gun Deaths in 2010 and 2013*. Each line in this interactive unit visualization represents a person. The orange segment corresponds to the period lived and the gray segment represents the estimated years stolen from the person. Readers can select the dataset by choosing the year and also filter by sex, age, region, and time. Source: Periscopic [40].

used to be in the past, or the marks used to be in proximity to the persons they represent. The Data Wallpaper (see Figure 8) is located at this intermediate point of the continuum because the persons who are represented in the visualization provided their personal data next to the visualization, but left the exhibition space afterward.

Visualizations with **high situatedness** are made of marks that are presented close to the persons they represent. An example of this point on the continuum is the Activity Clock (see Figure 20), in which the authors installed a visualization of lab presence data in the cafeteria of the lab itself. The visualization is highly situated because the people it represents are (generally) near the visualization.

Finally, there is **maximum situatedness** where the marks are the persons they represent. This is the most extreme point on the situatedness continuum and corresponds to physical visualizations made of real people, and showing data about those people. The show 100% (see Figure 12) is a maximally-situated visualization where the persons split themselves into groups or hold signs of different colors according to the questions they are asked.

It is possible that achieving at least some degree of situatedness can, in some cases, help observers relate to the people represented. Outside of visualization, situatedness has long been thought to affect people's emotions: memorials are often placed in a location where a significant event has affected a person or a group of persons, and memorabilia can acquire emotional power by virtue of having been touched or worn by a person [38] (both are examples of intermediate situatedness). In contrast, most visualization designs are meant to be easy to replicate and share, which maximizes the number of people who can see them but also means they have low situatedness [13].

3.6 Interactivity and Animation

We have so far assumed static visualizations. Although many visualizations are indeed static, others are *dynamic*, i.e., they can change under the user's influence (interactive charts) or outside the user's influence (animated visualizations such as in educational videos) [39]. Among the 105 visualizations in our collection, 44 are dynamic.

A dynamic visualization can be thought of as a (potentially very large) set of static views, each of them providing a different perspective on the data. There are three major ways in which users can explore data about people by moving from view to view.

First, dynamic visualizations can let users explore different *sets of data items* — people or groups of people — over time. Figure 23, for example, shows a visualization where users can explore the persons who died by a gunshot in different years (which can be chosen by clicking on the corresponding year). Similarly, the animation from Figure 15 changes the dataset while switching from 2010 to 2016.

Second, dynamic visualizations can let users explore different *sets of attributes* for a given set of data items. In Figure 23, users can filter victims by sex, age, or region. This approach is useful when the number of perceptual channels is insufficient to display all attributes simultaneously: users can learn more and more about people over time. In addition, in interactive visualizations, users can focus on people's attributes they care the most about.

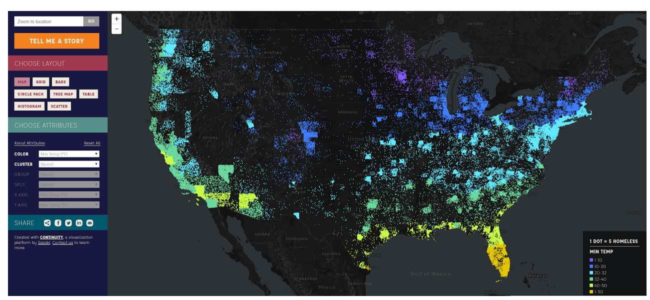


Fig. 24: [PERMISSION PENDING] *Understanding homelessness in USA*. This interactive unit visualization represents the homeless population in the USA. Each dot corresponds to five homeless people. The representations can be chosen among maps, grids, bars, etc. The viewer can also explore different factors from categories such as geographic, economic, or social. Source: Understanding Homelessness [41].

Finally, dynamic visualizations can let users explore the same data — items and attributes — through *different representations*. In Figure 24 for example, homeless people in the USA can be represented through maps, grids, bars, among other types of representations.

Since the different views of a dynamic visualization may have different characteristics according to our design space, interactivity and animation can offer a way for users to dynamically navigate in the anthropographic design space, both in terms of what is shown and how it is shown. The dynamic labels of visualizations from figures 1a, 7, 15, and 23, for example, increase the information specificity by showing detailed data about each individual. Another dimension that can change in dynamic visualizations is granularity. For example, the chart from Figure 24 lets users change its granularity from intermediate (every dot represents 5 persons) to low (all the dots are combined to form a bar).

3.7 Differences with Boy et al.

As we mentioned previously, our design space of anthropographics extends an earlier proposal by Boy et al. [3]. Our extension both broadens the original design space (that is, it captures a larger variety of designs) and sharpens it (that is, it makes finer distinction between related designs).

Boy et al's design space had four main dimensions:

- **Class of visualization**. This dimension distinguishes between *unit* and *aggregate* visualizations. It maps to the *granularity* dimension of our framework. In contrast with the original dimension, granularity sits on a continuum and distinguishes between two types of unit visualizations: those where each mark represent a single person, and those where each mark represents a fixed number of persons.
- **Human shape**. This dimension consists of two sub-dimensions:
  - *Realism* (*abstract—realistic*) directly maps to our *realism* dimension. However, our framework expands the definition of realism to also include more realistic marks such as photographs and real humans.
  - *Expressiveness* (*neutral—expressive*). We initially included expressiveness as a "how it is shown" dimension but removed it after our multi-coder evaluation because we found it hard to define, especially when considering non-anthropomorphic marks. We also realized that the expressiveness of a visualization often arises in large part from the meaning of the dataset and can be manipulated by a variety of visual design strategies like the use of metaphors, that are hard to operationalize.
- **Unit labeling**. This dimension captures three types of text annotations that can be displayed on top of marks: *generic*, *iconic* and *unique*. In our framework, it is incorporated into the more general *information specificity* dimension. Unit labeling is more specifically captured by the concept of attribute *distinctiveness*, which formalizes Boy et al.'s notion of uniqueness and generalizes it to other types of information beyond text annotations. Among other things, distinctiveness captures the use of unique anthropomorphic shapes, which was also discussed by Boy et al., but as part of the realism dimension which we see as orthogonal.
- **Unit grouping**. This dimension captures the spatial layout of the marks, such as *grid-based* or *organic*. We decided not to include this dimension in our framework because it is specific to unit visualizations and cannot be easily generalized to low-granularity visualizations. We also could not find arguments in the past literature in support of spatial layout influencing prosocial feelings or behavior.

Much of these modifications to the original design space were meant to cover a wider range of designs, and thus situate anthropomorphic unit visualizations within a larger design space of visualizations of data about people. Our conceptual framework also extends the original framework in a number of other ways, including by making a useful distinction between two classes of design space dimensions: *what is shown* (which includes two of Boy et al.'s dimensions), and *how it is shown* (which includes another dimension). Our design space also introduces two additional dimensions in each category: *coverage* and *authenticity* in the "what is shown", and *physicality* and *situatedness* in the "how it is shown", all of which we argued are relevant dimensions to consider when designing anthropographics. Finally, like Boy et al.'s work, our design space focuses on design dimensions that could plausibly promote compassion, but it includes more extensive discussions of why this should be the case, and of the underlying trade-offs. At the same time, unlike Boy et al., our work does not contribute any empirical finding.

4 Families of Visualizations

With a corpus of data visualizations and a design space to describe them, we now turn to reflect on the combinations of design choices we observed in our collection of visualizations. Although our collection is not a random sample of all existing visualizations and is likely biased, it can still be informative as a proxy to what exists. For example, if there is a certain type of design we did not particularly emphasize during our search and yet appears a lot in our collection, then this provides an indication that this design is popular. Conversely, if there is a design (i.e., a specific combination of dimensions) that would have caught our eye but of which we found no example, then this should be an indication that this design is at best relatively uncommon.

In our collection, the design dimensions that varied the most are realism, granularity, and specificity. From different combinations of these dimensions we derived several families of visualizations, which we grouped into two large classes: non-anthropomorphic and anthropomorphic. After discussing the designs most frequently used, we describe atypical visualizations that tap into less commonly used design dimensions, namely: coverage, authenticity, physicality, and situatedness.

4.1 Non-anthropomorphic Designs

Non-anthropomorphic visualizations are visualizations whose marks do not resemble a human being. They have varying degrees of granularity and information specificity. We break down this large class into three common families.

Statistical charts have low granularity. Their goal is to show high-level patterns in the data, rather than specific information about individuals. The visualization in Figure 20, for example, shows only two attributes and aggregates observations in a way that does not reveal data about individuals. Statistical charts can be low in information specificity when the goal is to convey simple trends, or they can have intermediate to high specificity, like the visualization in Figure 3 which conveys rich information about each demographic group. However, having low granularity, they cannot show rich information about individuals.

Simple unit charts are non-anthropomorphic visualizations characterized by having maximum granularity but low information specificity. Each individual is visible but little information is conveyed about them: see Figures 7 and 21 for examples. Such visualizations are typically used when the intent is not to show detailed characteristics about each person but to convey numbers of people, or how people are distributed across a few attributes.

Information-rich unit charts are non-anthropomorphic visualizations with both maximum granularity and high information specificity. This family of visualization presents various data attributes — often both encoded or literal, and thus gives readers access to detailed information about each person. Examples are Figures 11 and 23.

4.2 Anthropomorphic Designs

Anthropomorphic visualizations are a broad class of visualizations whose marks resemble human beings. They have varying levels of realism and information specificity. Although creating an aggregated anthropomorphic visualization is possible in principle [3], it does not appear to be a common design choice. In our collection, all anthropomorphic visualizations have intermediate to high levels of granularity, so they can be referred to as unit charts. Two common variations are wee-people designs and more realistic portrait-like designs.

Proportional wee-people charts are anthropomorphic unit visualizations with intermediate realism and intermediate granularity. This kind of visualization usually has low information specificity and uses pictorial marks to represent a fixed group of persons. Examples of such visualizations are Figures 5 and 15.

Individual wee-people charts are anthropomorphic unit visualizations with intermediate realism and maximum granularity. They have varying degrees of information specificity: the visualizations 10 and 24 (see Supplementary Material A) have low information specificity, while the one in Figure 1a achieves a high level of specificity by providing detailed information about each individual, including the name and the circumstances of their shooting.

Face charts are high on the realism dimension. This family consists of visualizations where the marks are either photographs, drawn portraits, or the people themselves. In Figure 17, for example, the chart shows detailed faces of the 30 most rich persons in the world. Since a face is a highly distinctive attribute, all the visualizations in this category are high on the information specificity dimension.

4.3 Atypical Designs

This section discusses less frequent design choices in the design space of anthropographics, which tap into the dimensions of authenticity, coverage, situatedness, and physicality.

Embellished charts are visualizations with partial authenticity. The same way a good storyteller can embellish a true story to make it more poignant, here the designer embellishes real data with synthetic information to enhance its impact. Embellished charts are often anthropomorphic. An example is shown in Figure 15, where neither the name nor the gender of the silhouettes originates from the known dataset. Another example is presented in Figure 18, where the persons who represent the men who died from AIDS are not the actual persons being represented. Embellished charts are higher on information specificity, but not all of the information they convey is genuine.

Single-person charts are visualizations characterized by having a single person in the visualized dataset. These charts are typically non-anthropomorphic and often convey rich information about a particular person, and therefore they have high specificity. Figure 10 provides a striking example. Many other examples exist outside of anthropographics, such as biographical and autobiographical visualizations [42], [43], [44]. Single-person charts can either have full or minimum coverage, depending on whether the message is about one specific person, or about a larger population of which one person has been taken as a representative. The visualization in Figure 10 is ambiguous in that respect, because although the data is about a person the designer knows personally, one of her stated goals was to "empower patients and families dealing with illness or disease" [26].

Example-driven charts are visualizations with minimum to partial coverage, where the designer chooses to focus on data from one or a few individuals in order to convey information or sentiments about a larger population. Focusing on a small number of people increases opportunities for showing richer information since there is more space to show attributes, and it is easier to collect rich data on a few people. Thus, example-driven charts often have intermediate to high information specificity. See Figures 11 and 1b for two examples. As discussed above, example-driven charts can also be *single-person charts*.

Situated visualizations exhibit some form of spatial proximity relationship with the people they represent. The show series "100%" (Figure 12) involve situated visualizations where the marks are the persons themselves (thus they are also *face charts*). Other examples are participatory visualizations where each person crafts their own physical mark (e.g., visualizations 39 and 44 in Supplementary Material A). Situated visualizations usually cover small populations since building them can be costly, especially when the marks are physical objects or people.

Physicalizations — also known as physical visualizations [19] — are visualizations in which the marks are embodied by physical objects as opposed to shown on flat displays. Physicalizations can take on many different forms. For example, some are non-anthropomorphic (e.g., Figure 21) while others are anthropomorphic (Figure 18).

5 Discussion

In this section, we discuss opportunities for design suggested by our design space, as well as opportunities for research and limitations of our work.

5.1 Opportunities for Design

Some relevant design choices and directions are visible in our design space but appear underrepresented in our current collection of examples. We discuss four of them.

5.1.1 Hybrid Designs

One particularly promising direction is to explore designs that share some of the qualities of statistical charts and some of the qualities of anthropographics. Statistical charts are widely used in different domains and for various purposes, including for conveying facts about suffering populations. However, since statistical charts lack properties expected to promote compassion, they may be experienced by many as "cold" and "reducing people to numbers". Meanwhile, there is a growing community who argues that charity should be driven by reason and facts to have a real impact [45]. Many of the anthropographic designs discussed in this paper may not satisfy such thirst for factual information and may be perceived as focusing too much on anecdotes and on emotional appeal to be taken seriously. Most people probably combine those two modes of thinking to some degree [46]. A key challenge, then, is to find designs that can appeal to both types of thinking.

One way the two approaches could be mixed is through the use of interactivity, for example, based on Harris' concept of *Near and Far* [7]. Visualizations based on this approach would allow readers to dive deeper into personal stories (e.g., filtering or querying specific persons and their attributes) while offering them the possibility to explore statistical patterns through an aggregated view. Those detailed views could use anthropographic design principles to help readers relate to specific individuals. The hope is that readers can understand statistical information for effective data-driven decision-making while being able to understand and empathize with the personal experience of individuals.

5.1.2 Customization and Targeting for Compassion

A person's emotional response to a visualization likely depends on the relationship between that person and the population visualized. For example, a person may find it harder to care about the plight of distant populations than about difficulties encountered by people who live nearby and are similar to them [6]. Although this relationship is mostly outside the designer's control, a visualization's emotional strength may be enhanced by playing with the "what is shown", i.e., by showing the appropriate information to the appropriate people.

In subsection 3.6, we discussed an example of a visualization that lets users select the attributes to visualize and filter people by their demographic characteristics (see Figure 23). This kind of chart offers users the opportunity to focus on the type of person or the characteristics they care the most about. Visualization 78 in Supplementary Material A goes further by asking users to enter their personal information and then customizes the visualization by showing populations of people who are like them. More elaborate techniques could be used to emphasize similarities between the user and a suffering population that may otherwise appear very remote and different. For example, if the user has kids and some of the suffering people also have kids, this attribute could be displayed and emphasized. If on the other hand, many of the visualized people are older, and the reader is young, the visualization may hide age information and show another attribute instead.

"Smart" techniques like the above move us away from a *user customization* approach and closer to a more controversial *user targeting* approach. While current systems require explicit interaction (e.g., people need to enter information about themselves), it is possible to go further by using personal information shared by the browser (e.g., country of residence) to optimize visualizations for maximizing prosocial feelings or behavior. While this is a dark pattern [47] that touches on issues of privacy and personal freedom, a lot can still be gained from customization and responsive targeting. Besides humanitarian applications, in a time of intense polarization and tribalism, targeted anthropographics could help individuals become more tolerant by showing them the many characteristics they share with people of a different identity or political leaning.

5.1.3 Exploring Example-Driven Charts Further

When communicating about a particular suffering population or a particular societal concern (e.g., gun fatalities), data visualizations typically show data about the entire concerned population (e.g., all people who fell victim to gunshots in the USA). However, we found examples of visualizations that represent a small subset of the reference population. Those visualizations focus on conveying rich information about a few persons, which may contribute to making readers more compassionate [10]. For example, in Figure 1b, the designer conveyed the difficulties of the life of refugees by creating six narratives of how a refugee traveled from their respective home country to Italy. Other example-driven charts such as Visualization 100 in Supplementary Material A or Figure 11 cover different topics using a similar technique.

Although example-driven charts are occasionally used, it appears that a lot remains to be explored. For example, we have not found a single clear example of a minimum-coverage visualization (i.e., focusing on a single person). *Bruises* (Figure 10) comes closest to a minimum-coverage visualization but it remains ambiguous (see *single-person charts* in section 4 for a discussion). Another interesting direction could be to combine an example-driven approach with the customization/targeting strategy discussed previously. For example, a visualization that uses minimum coverage could choose the representative person wisely, in such a way that the person shares some of the reader's characteristics.

5.1.4 Exploring Situatedness Further

We discussed a few examples of situated visualizations, but much remains to be explored in this area, especially around the notion of intermediate situatedness. As explained in subsubsection 3.5.3, intermediate situatedness can occur when the marks are presented at a location where the persons used to be in the past. This practice has received little attention, yet it can be useful to design visualizations that act as memorials, e.g., that show data about people who perished or experienced hardship at some specific location. Our only example is Visualization 2 in Supplementary Material A, where the stories of 28 women who were harassed in a public space were encoded into a physicalization and presented in different places of this public space during one week.

A non-situated example of a memorial-type design is Visualization 105 in Supplementary Material A, where 7,000 pairs of shoes were placed on the lawn of the U.S. Capitol to symbolize the children killed during the 2012 Newtown shooting. Although the installation conveys an emotional charge, it is located 300 miles away from Newtown. Would it carry more emotion had it been placed exactly where the shooting occurred? Perhaps the practice of anthropographics can learn from memorial design. Many existing memorials are situated, although they rarely visualize rich data about the people. It is possible that doing so would help people relate, although practices exist in memorial design that paradoxically do the exact opposite. For example, the tomb of the unknown soldier at Arlington has been designed to convey as little information as possible, so that anyone could relate and entertain the possibility that the soldier is a lost relative [48].

The second way intermediate situatedness can occur is when the marks used to be in proximity to the persons they represent. This approach has been used in participatory designs such as the Data Wallpaper (see Figure 8), and it also opens up a vast area of untapped possibilities. For example, it could be interesting to explore anthropographic visualizations that tap into the psychological power of memorabilia that were owned, touched, or worn by people [38]. For example, would the installation in Visualization 105 discussed above carry more emotional weight had it used objects that personally belonged to the victims?

Finally, wearable visualizations such as data jewelry and data clothing[7] are fully situated visualizations when they convey data about the people who are wearing them [13]. Perhaps there are ways in which such objects can be used to promote prosocial feelings and behavior towards people in need. For example, if we succeed in developing a visual language for conveying personal histories and people become proficient at reading it, homeless persons in need may be able to share their hardships through wearable visualizations that everyone could understand at a glance.

5.2 Opportunities for Research

None of the design dimensions described in this article has been thoroughly explored in the visualization literature. This section underlines opportunities for empirical research.

5.2.1 Testing Basic Design Dimensions

Boy et al. [3] conducted several experiments to explore the effectiveness of design choices to promote prosocial feelings. Their experiments mostly compared proportional wee-people charts — intermediate granularity and realism — with classical statistical charts — low granularity and realism. Although the results are largely inconclusive, it does not follow that anthropographics are ineffective. Perhaps the effects are small, and the experiments did not have enough statistical power to detect them [49]. Furthermore, many other designs remain to be tested, some of which could prove more effective. For example, the anthropographics used by Boy et al. [3] generally had partial authenticity, and none of them used maximum granularity.

There is a clear need for more empirical studies that investigate the effect of basic anthropographic design dimensions, namely, granularity, specificity, authenticity, and realism. To maximize lessons learned, the isolated effects of these dimensions as well as their interactions must be tested. For example, a strong sense of individuality can only be conveyed if a visualization has both maximum granularity and high information specificity, so it will be interesting to test whether this combination is particularly beneficial. Also, it will be crucial to control for authenticity and assess its effects separately. Although some designers add synthetic attributes to increase information specificity, this comes at the cost of authenticity, the importance of which has been under-discussed in the visualization literature.

5.2.2 Understanding Coverage

Once the effectiveness of basic design dimensions starts to be understood, it will be interesting to study more elaborate strategies, starting with coverage. As we mentioned before, empirical evidence in psychology suggests that as the number of identifiable victims in a narrative increases, prosocial feelings and behavior decrease [10]. Until now, no study has been conducted to test whether this so-called "compassion fade" can also occur in the context of data visualization. Keeping the reference dataset constant and manipulating the coverage dimension can be an effective way of empirically exploring compassion fade in the context of data visualization.

5.2.3 Studying Situatedness and Physicality

Situatedness and physicality may also be relevant to anthropographics and may deserve attention, even though empirically studying them will likely be less practical. One example of an empirical question is: does an intermediate-situatedness visualization that shows victims in-place evokes more compassion than a non-situated visualization that shows the data elsewhere? Can physicality add to the emotional impact or memorability of a design strategy, and in turn promote compassion? Designs like "people-as-mark" (Figure 12) and wearable physicalizations would be particularly interesting to test since they are maximum in both the situatedness and physicality dimensions.

5.2.4 Investigating Ambiguity

Finally, it could be interesting to explore the effect of ambiguity on prosocial feelings and behavior. For example, in cases where the designer does not make explicit how many persons are represented by each mark — as opposed to Figure 5, where each mark explicitly represents 1 million soldiers — coverage and granularity become ambiguous. In Figure 12, for example, each mark can either be interpreted as standing for (A) a proportion or (B) a subset of the Parisian population. If one chooses A as interpretation, the visualization would have intermediate granularity and full coverage. Conversely, if the interpretation is based on B, the visualization would have maximum granularity and partial coverage. Can the same visualization have different effects on prosocial feelings or behavior depending on how it is interpreted?

5.3 Limitations

Our goal was to identify elementary dimensions in the vast design space of anthropographics and propose a conceptual framework and terminology that can help both researchers and practitioners reflect on and communicate about essential aspects of anthropographics design. Nevertheless, this design space is far from complete.

First, our conceptual framework currently only covers flat tables (see subsection 3.1). Although tables are likely the

7. See http://dataphys.org/list/tag/data-jewellery/ and http://dataphys.org/list/tag/data-clothing/ for examples.

most common data model, other types of datasets exist such as networks [12]. Excluding such datasets excludes all social network visualizations, for example.

Second, we restricted our scope to datasets where each item is a person or a group (see section 2). There exist datasets that contain information that can profoundly affect people, despite not containing information about people. For example, datasets on how diseases spread geographically, or data about global warming. But since there is no one in these datasets with which to empathize or feel compassion about, they may be outside the realm of anthropographics.

We also excluded other datasets and visualizations from our investigation in order to keep the scope manageable, although they could be relevant to anthropographics. In particular, we excluded datasets involving simulated people (see section 2), but these may be useful to consider in the future. Also, although our focus was on people, many non-human animals experience suffering and may need compassion as well as help [50], [51]. Our conceptual framework can be easily generalized by considering that data items can refer to other sentient beings than humans.

In choosing the seven dimensions of our design space, we focused on elementary visualization characteristics that are easy to define, to manipulate in studies, and to apply as a designer. Many other factors can contribute into making a visualization effective in promoting prosocial feelings which are much more difficult to operationalize, including aesthetics, the use of visual metaphors, and the use of compelling accompanying stories (see, e.g., Figure 10). Nevertheless, given the current lack of concepts, terms, and empirical knowledge in the domain of anthropographics, we think that our design space is a reasonable starting point.

Finally, it is important to stress that our design space does not capture properties that are inherent to the dataset being visualized, even though they may have a considerable impact on how compassionate readers can feel. Personal connection with the data, for example, has a significant impact on how people experience visualizations [52]. However, such properties do not belong to a design space of visualizations, since the designer has no control over the dataset once it has been chosen.

ACKNOWLEDGMENTS

The first author gratefully acknowledges the financial support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [grant number 88881.188888/2018-01].

REFERENCES

[1] B. Berkowitz, D. Lu, and C. Alcantara, “The terrible numbers that grow with each mass shooting,” https://wapo.st/2RZsYmy, 2019, [Online; accessed 17-July-2019].

[2] F. Fragapane and A. Piacentini, “Stories behind a line,” http://www.storiesbehindaline.com, 2019, [Online; accessed 17-July-2019].

[3] J. Boy, A. V. Pandey, J. Emerson, M. Satterthwaite, O. Nov, and E. Bertini, “Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data?” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, pp. 5462–5474.

[4] E. Bertini, “Can visualization elicit empathy? our experiments with ”anthropographics”,” http://bit.ly/2Gs53FJ, 2017, [Online; accessed 8-July-2019].

[5] G. Lupi, “Data humanism, the revolution will be visualized,” *Published online at https://www.printmag.com/information-design/data-humanism-future-of-data-visualization. Retrieved Feb*, 2017.

[6] P. Bloom, *Against empathy: The case for rational compassion*. Random House, 2017.

[7] J. Harris, “Connecting with the dots,” https://source.opennews.org/articles/connecting-dots/, 2015, [Online; accessed 15-March-2019].

[8] L. C. Rost, “Data point moves into a bar,” https://lab.dsst.io/slides/33c3/7999.html, 2016, [Online; accessed 15-March-2019].

[9] M. Beaudouin-Lafon, “Designing interaction, not interfaces,” in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2004, pp. 15–22.

[10] D. Västfjäll, P. Slovic, M. Mayorga, and E. Peters, “Compassion fade: Affect and charity are greatest for a single child in need,” *PloS one*, vol. 9, no. 6, p. e100115, 2014.

[11] M. Le Goc, L. H. Kim, A. Parsaei, J.-D. Fekete, P. Dragicevic, and S. Follmer, “Zooids: Building blocks for swarm user interfaces,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 2016, pp. 97–109.

[12] T. Munzner, *Visualization analysis and design*. AK Peters/CRC Press, 2014.

[13] W. Willett, Y. Jansen, and P. Dragicevic, “Embedded data representations,” *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 461–470, 2017.

[14] S. R. Dos Santos, “A framework for the visualization of multidimensional and multivariate data,” Ph.D. dissertation, University of Leeds, 2004.

[15] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, “Vega-lite: A grammar of interactive graphics,” *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 341–350, 2017.

[16] J. Bertin, *Sémiologie graphique: Les diagrammes-Les réseaux-Les cartes*. Paris: Gauthier-VillarsMouton & Cie, 1973. [Online]. Available: http://infoscience.epfl.ch/record/51376

[17] H. Senay and E. Ignatius, *Rules and principles of scientific data visualization*. Institute for Information Science and Technology, Department of Electrical Engineering, 1990.

[18] R. E. Roth, “Visual variables,” *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, pp. 1–11, 2016.

[19] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian, and K. Hornbæk, “Opportunities and Challenges for Data Physicalization,” in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. New York, NY, United States: ACM, Apr. 2015. [Online]. Available: https://hal.inria.fr/hal-01120152

[20] T. Bronshtein, “200 years of immigration to the us,” http://bit.ly/2Lrytb0, 2017, [Online; accessed 17-July-2019].

[21] L. A. Worrell, “The great war,” http://bit.ly/2RTjhpS, 2019, [Online; accessed 17-July-2019].

[22] O. Neurath, *International picture language: The first rules of isotype*. Basic English Publ., 1936.

[23] M. Neurath, “Isotype,” *Instructional science*, vol. 3, no. 2, pp. 127–150, 1974.

[24] Gun Violence Archive, “Gun violence chart of deaths in 2019,” https://www.gunviolencearchive.org/, 2019, [Online; accessed 17-July-2019].

[25] G. Lupi, “Collaborative data wallpaper for story,” http://bit.ly/2Sjv6mW, 2019, [Online; accessed 17-July-2019].

[26] ——, “Bruises: The data we don’t see,” http://bit.ly/2LYwev8, 2019, [Online; accessed 17-July-2019].

[27] T. Kogut and I. Ritov, “The “identified victim” effect: An identified group, or just a single individual?” *Journal of Behavioral Decision Making*, vol. 18, no. 3, pp. 157–167, 2005.

[28] S. Lee and T. H. Feeley, “The identifiable victim effect: A meta-analytic review,” *Social Influence*, vol. 11, no. 3, pp. 199–215, 2016.

[29] L. A. d. M. Morais and N. Andrade, “Small data e o câncer de mama: a jornada da paciente,” http://bit.ly/2y6lPp4, 2019, [Online; accessed 24-July-2019].

[30] H. Haug, S. Kaegi, and D. Wetzel, “100% Paris,” http://bit.ly/2y2Qofz, 2015, [Online; accessed 17-July-2019].

[31] C. Kirk and D. Kois, “How many people have been killed by guns since newtown?” http://bit.ly/32Ao8yD, 2013, [Online; accessed 17-July-2019].

[32] Bloomberg, “Bloomberg billionaires index,” https://bloom.bg/2JBoCNs, 2016, [Online; accessed 17-July-2019].
[33] T. Rawles, “Picture from 1993 reminds people of the loss of life due to aids,” http://bit.ly/2Gj6Rk5, 2017, [Online; accessed 17-July-2019].
[34] S. Campbell, “Feeling numbers the rhetoric of pathos in visualization,” Master’s thesis, Northeastern University, 2018.
[35] L. A. d. M. Morais, N. Andrade, D. M. Costa de Sousa, and L. Ponciano, “Defamiliarization, representation granularity, and user experience: a qualitative study with two situated visualizations,” *PacificVis*, 2019.
[36] Emydot, “All the people in all the world at the new york world,” http://bit.ly/2JMqy4S, 2008, [Online; accessed 17-July-2019].
[37] B. H. Thomas, G. F. Welch, P. Dragicevic, N. Elmqvist, P. Irani, Y. Jansen, D. Schmalstieg, A. Tabard, N. A. ElSayed, R. T. Smith *et al.*, “Situated analytics,” in *Immersive Analytics*. Springer, 2018, pp. 185–220.
[38] G. E. Newman and P. Bloom, “Physical contact influences how much people pay at celebrity auctions,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 10, pp. 3705–3708, 2014.
[39] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale, “A review of temporal data visualizations based on space-time cube operations,” in *Eurographics conference on visualization*, 2014.
[40] Periscopic, “U.s. gun deaths in 2010 and 2013,” http://bit.ly/2GePUax, 2013, [Online; accessed 17-July-2019].
[41] Sasaki, “Understanding homelessness,” http://bit.ly/2XLxHMp, 2017, [Online; accessed 17-July-2019].
[42] B. Bach, P. Dragicevic, S. Huron, P. Isenberg, Y. Jansen, C. Perin, A. Spritzer, R. Vuillemot, W. Willett, and T. Isenberg, “Illustrative data graphics in 18th-19th century style: A case study,” in *IEEE Conference on Visualization-IEEE VIS 2013*, 2013.
[43] C. Perin, “The symmetry of my life: An autobiographical visualization,” in *Poster at IEEE VIS*, 2017.
[44] ——, “The symmetry of my life ii,” in *Poster at IEEE VIS*, 2018.
[45] W. MacAskill, *Doing good better: Effective altruism and a radical new way to make a difference*. Guardian Faber Publishing, 2015.
[46] D. Kahneman, *Thinking, fast and slow*. Macmillan, 2011.
[47] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, “The dark (patterns) side of ux design,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 534.
[48] J. Rosenberg, “The known unknown,” https://99percentinvisible.org/episode/the-known-unknown/, 2019, the 99% Invisible podcast.
[49] J. Cohen, “The earth is round (p¡. 05),” in *What if there were no significance tests?* Routledge, 2016, pp. 69–82.
[50] T. Regan and P. Singer, *Animal rights and human obligations*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[51] B. Tomasik, “The importance of wild-animal suffering,” *Rel.: Beyond Anthropocentrism*, vol. 3, p. 133, 2015.
[52] E. M. Peck, S. E. Ayuso, and O. El-Etr, “Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania,” *arXiv preprint arXiv:1901.01920*, 2019.

Luiz Augusto de Macêdo Morais is a Ph.D. candidate in Computer Science at Universidade Federal de Campina Grande, Brazil. His research interests include anthropographics, situated data visualization, data physicalization, and the application of psychological theories to data visualization.

Yvonne Jansen is a research scientist at the Centre National de Recherche Scientifique (CNRS) since 2016. She is affiliated with the Institut des systèmes intelligents et de robotique (ISIR) at Sorbonne Université, Paris, France. She received her doctoral degree from Université Paris Sud in 2014. She previously held a post-doctoral position at the University of Copenhagen in Denmark. Her research interests include data physicalization, situated and embedded data visualization, and methods and techniques to facilitate the transparent reporting of research outcomes.

Nazareno Andrade is a professor at the Systems and Computing Department of Universidade Federal de Campina Grande, Brazil, since 2009. Nazareno received a PhD in Electrical Engineering from the same University, and was a post-doctoral fellow at the Delf University of Technology. His research interests include anthropographics, civic technology, social computing, and music information retrieval.

Pierre Dragicevic is a permanent Research Scientist at Inria (France) since 2007, studying information visualization and human-computer interaction. He received his PhD from the Université de Nantes in 2004 and was a post-doctoral fellow at the University of Toronto from 2006—2007. His research interests include data physicalization, decision making with data visualizations, user study methodology, transparent statistical communication, design spaces and conceptual frameworks.