Universidade Federal de Campina Grande

Centro de Engenharia Elétrica e Informática

Coordenação de Pós-Graduação em Ciência da Computação

Designing for More than Efficacy: Investigating the Role of Anthropographics on Compassion

Luiz Augusto de Macêdo Morais

Tese submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Visualização de Dados

Nazareno Andrade

(Orientador)

Campina Grande, Paraíba, Brasil

M827d Morais, Luiz Augusto de Macêdo.
Designing for more than efficacy : investigating the role of anthropographics on compassion / Luiz Augusto de Macêdo Morais. - Campina Grande, 2020.
215 f. : il. color.

Tese (Doutorado em Ciência da Computação- Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2020.
"Orientação: Prof. Dr. Nazareno Andrade.
Referências.

1. Antropográficos. 2. Visualização de Dados. 3. Sentimentos - Compaixão. I. Andrade, Nazareno. II. Título.

CDU 004(043)

FICHA CATALOGRÁFICA ELABORADA PELO BIBLIOTECÁRIO GUSTAVO DINIZ DO NASCIMENTO CRB-15/515

DESIGNING FOR MORE THAN EFFICACY: INVESTIGATING THE ROLE OF ANTHROPOGRAPHICS ON COMPASSION

LUIZ AUGUSTO DE MACÊDO MORAIS

TESE APROVADA EM 28/02/2020

NAZARENO FERREIRA DE ANDRADE, Dr., UFCG
Orientador(a)

RAQUEL VIGOLVINO LOPES, Dra., UFCG
Examinador(a)

JOÃO ARTHUR BRUNET MONTEIRO, Dr., UFCG
Examinador(a)

PIERRE DRAGICEVIC, PhD
Examinador(a)

UTA HINRICHS, PhD
Examinador(a)

CAMPINA GRANDE - PB

Resumo

Antropográficos são visualizações de dados sobre pessoas que são projetados com o intuito de provocar sentimentos como compaixão ou empatia. No entanto, não há evidências na literatura para confirmar até que ponto os antropográficos geram compaixão para com as pessoas cujos dados representam. Este trabalho contribui para avançar o entendimento sobre antropográficos, refinando e ampliando o espaço de design de tais visualizações, assim como procurando evidências sobre o quanto elas afetam a compaixão das pessoas. Embora os resultados de um estudo *in-the-wild* tenham sido inconclusivos, um experimento de *crowdsource* encontrou um pequeno efeito ao comparar um antropográfico e um gráfico de barras. Ao verem o primeiro, os participantes se sentiram um pouco piores sobre a situação das pessoas em necessidade e doaram um pouco mais para ajudá-las. Este é o primeiro trabalho que encontra evidências a favor da hipótese de que designs de visualização específicos podem tornar as pessoas mais compassivas. Essa tese abre uma nova perspectiva para estudar quais elementos de design são responsáveis por tal efeito.

Abstract

Anthropographics are data visualizations about people that are designed with the intent of evoking feelings such as compassion or empathy. However, there is no evidence in the literature to confirm to what extent anthropographics make people compassionate with the persons whose data is represented. This work contributes to advance the understanding of anthropographics by refining and extending the design space of such visualizations as well as looking for evidence on how much they affect people's compassion. Although results from an in-the-wild study were inconclusive, a crowdsource experiment found a small effect when comparing an anthropographic with a bar chart. The former leads people to feel a bit worse about the situation of persons in need and donate a little more to help them. This is the first work that finds evidence in favor of the hypothesis that specific visualization designs may make people more compassionate. This thesis opens a new perspective for studying what design elements are responsible for such an effect.

Acknowledgements

I want to show my sincere gratitude to my parents, Luiz and Terezinha, for all the support given during my life. This Ph.D. would have never been possible without you.

I wish to express my indebtedness and gratitude to Naza, my supervisor. The time I spent working with you made me become a better researcher and person. You are a reference for me as a professional and human being.

I would like to recognize the invaluable opportunity that Pierre and Yvonne have provided me. Yvonne, thank you for accepting me for an internship and for helping me out in my studies. Pierre, I am grateful for your advice and support in my academic life. You have contributed significantly to this thesis come true.

I would also like to thank Dandara and Gibran, which helped me immensely with the Harassment Plants study.

I wish to express my gratitude to Lesandro for assisting me in the Our Activity project. Your advice on how to link theory to practice was essential in my early studies.

I want to say thank you for all LSD, AVIZ, and ISIR members that have helped me with my research. I would also like to acknowledge those that went beyond academic life and shared happy moments with me drinking a good beer or wine.

I would like to thank Uta Hinrichs, Pierre Dragicevic, Raquel Lopes, and João Arthur Brunet for being part of the thesis defense's committee and for giving me such relevant advice.

I would like to thank all people who directly or indirectly have contributed to my studies and my research trajectory.

Finally, I would like to thank CNPq for offering me financial support.

Contents

1 Introduction **1**

1.1 Motivation . . . 1

1.2 Objective . . . 3

1.3 Contributions . . . 3

1.4 Thesis History . . . 4

1.4.1 Research Prologue . . . 4

1.4.2 User Experience Study . . . 5

1.4.3 Humanism on Data Visualization . . . 6

1.4.4 Delving into Anthropographics . . . 7

1.4.5 Design of Physical Anthropographics . . . 8

1.5 Thesis Overview . . . 8

2 Background **10**

2.1 Empathy vs. Compassion . . . 10

2.2 Design for Compassion . . . 11

2.2.1 Anthropomorphism . . . 11

2.2.2 Identifiable Victims . . . 12

2.2.3 Near and far . . . 12

2.3 How to Measure Compassion . . . 12

2.4 Studies on Compassion and Charitable Donation . . . 13

3 Design Space of Anthropographics **15**

3.1 Methodology . . . 16

3.1.1 Scope . . . 16

3.1.2 Development of the Design Space 17
3.2 Design Space 18
3.2.1 General Visualization Concepts 18
3.2.2 New Concepts 19
3.2.3 Introducing the Design Space 21
3.2.4 What is Shown 21
3.2.5 How it is shown 31
3.2.6 Interactivity and Animation 37
3.2.7 Differences with Boy et al. 40
3.3 Families of Visualizations 43
3.3.1 Non-anthropomorphic Designs 44
3.3.2 Anthropomorphic Designs 45
3.3.3 Atypical Designs 46
3.4 Discussion 47
3.4.1 Opportunities for Design 47
3.4.2 Opportunities for Research 51
3.4.3 Limitations 53

4 Design of a Situated and Physical Anthropographic 55

4.1 Context 55
4.1.1 Project 56
4.1.2 Motivation 56
4.1.3 Public Space 56
4.2 Method 57
4.2.1 Visualization Sandcastling 58
4.2.2 Team Roles 59
4.3 Challenges 59
4.3.1 Understanding the Audience's Interest 60
4.3.2 Dealing with Data Availability 60
4.3.3 Adapting to Data Changes 62
4.3.4 Considering Environmental & Technological Constraints 64

4.3.5 Fostering Audience Engagement 65
4.4 Conclusion 65

5 In-the-wild Study 67
5.1 Study Overview 67
5.1.1 Dataset 68
5.1.2 Visualization Designs 69
5.1.3 Comparing to Boy et al. 70
5.2 Method 71
5.2.1 Visualization Deployment 71
5.2.2 Participants and Recruitment 72
5.2.3 Data Collection 73
5.2.4 Design Factors 74
5.2.5 Response Variables 74
5.2.6 Analysis 75
5.3 Results 75
5.3.1 Prosocial Behavior 75
5.3.2 Self-reported Compassion 76
5.4 Conclusion 77

6 Charitable Giving Experiments 79
6.1 Studies Overview 79
6.1.1 Dataset 79
6.1.2 Visualization Designs 81
6.1.3 Comparing to Other Studies 83
6.1.4 Ethics 84
6.1.5 Reproducibility 84
6.2 Experiment 1: Is there a clear effect? 84
6.2.1 Method 85
6.2.2 Results and Discussion 87
6.3 Experiment 2: Refined experimental design 89
6.3.1 Pre-study: Topic Selection 89

6.3.2 Method 90
6.3.3 Results and Discussion 92
6.4 Conclusion 95

7 Conclusion 96
7.1 Major findings 96
7.1.1 Prosocial Behavior 96
7.1.2 Self-Reported Compassion and Affect 97
7.2 Research Opportunities 97
7.2.1 Fruther Exploring the Design Space 98
7.2.2 Using More Precise Measurements 98
7.2.3 Investigating Other Response Variables 99
7.3 Concluding Thoughts 99

A Anthropographics List 109

B Data Collection Instruments from the Harassment Project 215

List of Figures

1.1 Mass Shooting Statistics in the United States 2
1.2 Stories Behind a Line 3
1.3 Thesis History 5

3.1 Conceptual datasets 19
3.2 200 Years of Immigration to the US 23
3.3 Granularity Continuum 24
3.4 The Great War 1914-18 25
3.5 Specificity continuum 26
3.6 Gun violence Chart of Deaths in 2019 27
3.7 Data wallpaper 28
3.8 Coverage continuum 29
3.9 Bruises: the Data We Don't See 30
3.10 Cancer is not always the end 31
3.11 100% Montréal 32
3.12 How Many People Have Been Killed by Guns in USA Since Newtown's attack 33
3.13 Authenticity continuum 34
3.14 Poverty in Syria 35
3.15 Realism continuum 36
3.16 The Billionaire Index 37
3.17 San Francisco Gay Men's Chorus representing AIDS deaths until 1993 . . . 38
3.18 Physicality continuum 39
3.19 Activity Clock 40
3.20 Of All the People in All the World: Stats with Rice 41

3.21 Situatedness continuum 42
3.22 U.S. Gun Deaths in 2010 and 2013 43
3.23 Understanding homelessness in USA 44

4.1 Açude Velho 57
4.2 Design Process: initial stage 59
4.3 Focus group 61
4.4 Design Process: intermediate stage 62
4.5 Design Process: final stage 63

5.1 Harassment Plants placed in a public space 68
5.2 Harassment Plants' vase in detail 69
5.3 Harassment Plants' legend 70
5.4 Harassment Information 71
5.5 Comparison between Boy et al. and in-the-wild study 72
5.6 Number of participants in each condition 73
5.7 Prosocial behavior: Harassment Plants vs. Harassment Information 76
5.8 Compassion: Harassment Plants vs. Harassment Information 77

6.1 Bar chart from experiment 1 80
6.2 Anthropographic from experiment 1 82
6.3 Comparison between Boy et al., in-the-wild study, and charitable giving studies 83
6.4 Experiment 1 conditions 86
6.5 Fund allocation question 87
6.6 Distribution of three donation questions 88
6.7 Experiment 2 conditions 91
6.8 Refined donation allocation question 92
6.9 Estimated mean difference in donation allocation between information-rich and information-poor designs with 95% CI. 93
6.10 Estimated mean difference in donation allocation between information-rich and information-poor designs in anonymous and named conditions and its interaction, with 95% CI. 93

6.11 Estimated mean difference in valence and arousal between information-rich and information-poor designs with 95% CI. 94

Chapter 1

Introduction

This chapter motivates the thesis that investigates to what extent certain kinds of visualizations that represent people affect compassion.

1.1 Motivation

When communicating data about people, information designers and data journalists regularly create visualizations meant to foster an emotional connection with the persons whose data is represented. Figure 1.1 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 1.2 conveys the hardship of refugees' life 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.

In the context of charitable giving, visualizations must go beyond the function of just informing or making people feel empathy for the persons in need. Charts that call for donation on NGO websites, for example, must be designed to evoke compassion, which motivates people to donate to the cause. Lambert [49] defends the view that visualizations for social change must affect people's behavior by persuading them to stand for a cause. In this context, it is necessary to understand what design strategies can be made in order to promote people's compassion for the persons the data represents.

The practice of visualizing data about people in a way that helps the audience relate has been called *anthropographics* [14; 8]. Boy and colleagues [14] 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"* [8]. Similarly, the term *data humanism* was coined to refer to a range of visualization design practices intended to promote humanistic values [53]. In this thesis, 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).*

Figure 1.1: Mass Shooting Statistics in the United States. Source: Washington Post (https://www.washingtonpost.com/graphics/2018/national/mass-shootings-in-america/)

Like Boy and colleagues [14], this definition refers to a class of visualizations rather than a set of design strategies, and like Bertini [8], it generalizes beyond the use of human-shaped symbols. Based on findings that empathy is not necessarily conducive to helping behavior [10], 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 [10].

Visualization designers have explored many strategies for creating anthropographics. A popular strategy is to use human-shaped symbols [14] (e.g., see Figure 1.1). Other strategies include the use of text annotations to make each person appear unique [14] (e.g., see the top left of Figure 1.2), 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 1.1) [36; 73]. So far, only a few of these strategies have been empirically tested, and even though initial studies have been mostly inconclusive [14], there is an enormous potential for Information Visualization research.

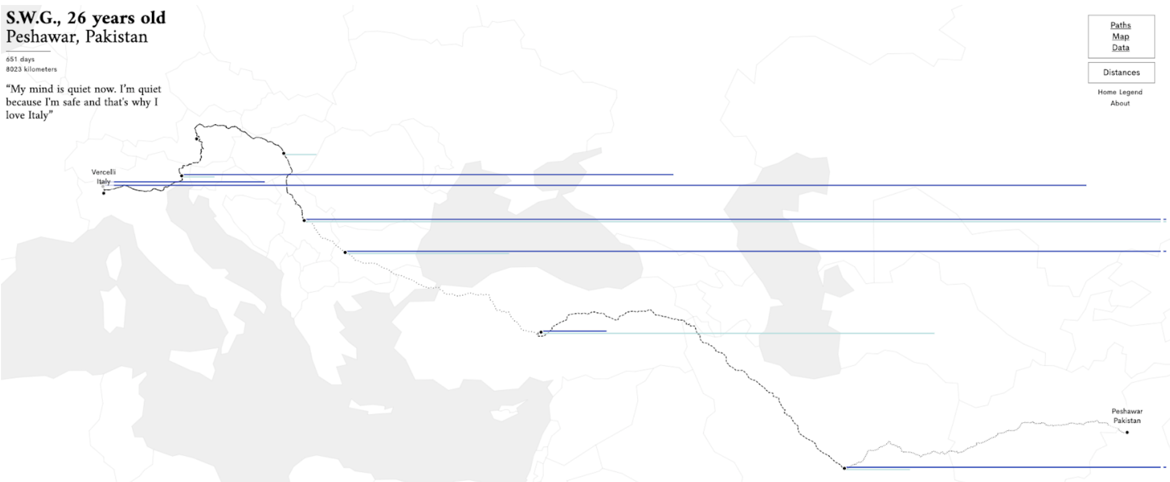


Figure 1.2: Stories Behind a Line. Source: Federica Fragapane and Alex Piacentini (http://www.storiesbehindaline.com)

1.2 Objective

This thesis aimed to advance present knowledge about the design of anthropographics. As more concrete goals, we intended to refine the design space of anthropographics and further investigate the effects of such visualizations on compassion, especially in the context of charitable giving.

1.3 Contributions

This work has contributed to the field of Anthropographics in different aspects. It pushed forward research on whether visualization design may influence people's compassion. It also contributed to giving new perspectives on future work in the field of Anthropographics.

As a first contribution, we have devised a comprehensive design space of anthropographics that advances the one proposed by Boy and colleagues [14]. We did so by establishing a proper language and concepts related to anthropographics and describing an overview of design strategies and opportunities for design.

A second contribution is that we have empirically tested design strategies that have not been extensively explored concerning their effect on compassion. We have investigated anthropographics both in the wild and in a more controlled environment through crowd-sourcing. We have also conducted experiments with higher statistical power than previous work.

Finally, we have also devised a practitioner-oriented contribution. We have presented the whole development of the visualization designed for the in-the-wild study. The lessons learned throughout the process may help designers to reflect on each stage of a physicalization development meant to be exhibited in public space.

1.4 Thesis History

A Ph.D. trajectory is often sinuous, and mine was no different. This section describes the main events that affected the final thesis. This section aims to help readers to understand the context behind the decisions that turned the research trajectory from exploring engagement with situated visualizations to investigating the role of compassion in the context of Anthropographics. The next sections summarise the main events and milestones of this Ph.D.

1.4.1 Research Prologue

I started my Ph.D. inspired by two works: *Data-in-Place: Thinking through the Relations Between Data and Community* [78] and *Embedded data representations* [86]. The former brings the notion of eliciting civic engagement by unveiling the data that exists in a place; the latter provides the terminology and examples of how to represent contextual data. A third influential work is Koeman's thesis [46], which brings a series of studies on urban visualization (situated visualizations exhibited in public space) that shed light on how people engage with such data representations. We initially aimed to create urban visualizations

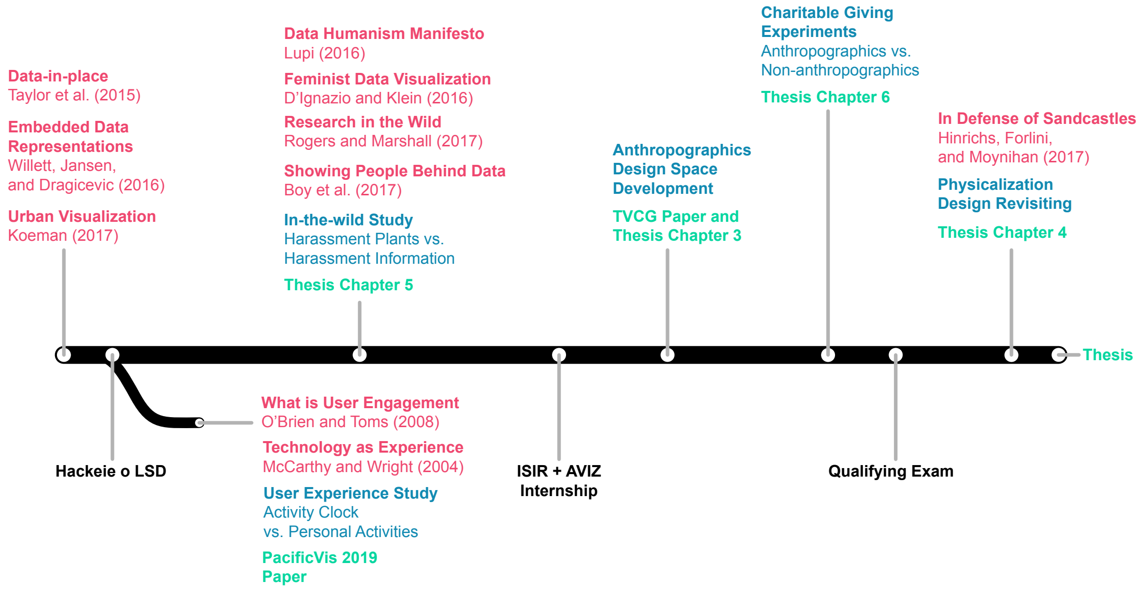


Figure 1.3: *Trajectory towards the final thesis*. Red annotations represent influential papers that affected in some way my research; text in blue corresponds to studies conducted during the Ph.D.; green labels are study reports such as papers or thesis chapters; and text in black corresponds to important events. Milestones (white dots) are in chronological order but the distance between them does not encode periods.

and investigate how users engage and reflect on the social aspects (e.g., people's perceived safety, needs, opinion, etc.) that exist in space. The plan of creating public visualizations was postponed when we decided to create a visualization for a hackathon.

1.4.2 User Experience Study

We attended at a hackathon called *Hackeie o LSD*, ("Let's hack LSD", in English; LSD is the lab I work). During the event, we designed the Activity Clock, a situated visualization that displays the aggregated presence of individuals along the day, which is designed to provoke a sense of defamiliarization for being embedded in a modified but ordinary wall clock. We took advantage of having created a visualization to run a study afterward.

The two main works that influenced the study are *What is user engagement? A conceptual framework for defining user engagement with technology*[64] and *Technology as Experience*[57]. At this point, we progressively changed our focus from engagement to something broader: user experience. By doing so, we expected to capture other perspectives of users' reactions while experiencing the visualization, including their emotional responses.

Before running the study, we created another visualization to contrast some design factors between them. The second visualization is called Personal Activities. It is presented in a conventional printed and framed poster but shows data from each person with an individual glyph as an attempt to lead readers to connect with the personal nature of the data. Both Activity Clock and Personal Activities share the same data and were placed at the same space: the lab's cafeteria.

The study explored the role of defamiliarization (i.e., presenting the visualization in a familiar or defamiliarized physical presentation) and granularity (i.e., representing one or more persons per mark) on the user experience with situated visualizations. At this point, we were interested in exploring how different designs of visualizations about people affect user experience and users' attitudes. The study resulted in a paper at PacificVis 2019 entitled *Defamiliarization, Representation Granularity, and User Experience: A Qualitative Study with Two Situated Visualizations* [21].

Results from the study on user experience are not included in the thesis because despite being part of the Ph.D. trajectory, they do not reflect the topic of compassion. The next section explains how we moved from exploring user experience to compassion.

1.4.3 Humanism on Data Visualization

At the end of the study on user experience, we were more and more interested in investigating how visualizations may affect non-utilitarian factors such as emotion. Before returning to the topic of urban visualization, we dove into topics such as Data Humanism [53] and Feminist Data Visualization [25] to bring a more human-centered and political discourse to the visualizations. This led us to consider creating visualizations to represent more touching data such as data about gender imbalances or other social issues instead of focusing on casual data such as users' presence at a lab.

As an exploratory phase of this research progressed, we focused on something more specific: the effects of data representation design on compassion, based on a paper entitled *Showing People Behind Data* [14]. Instead of hypothesizing that users would care about the data, now we consider that specific visualization designs may influence users to feel more or less concern *about the persons whose data is represented*. As the visualization would be exhibited at a public space, we looked up in the literature which research method best fits such study. We then designed a study based on the concept of research in-the-wild [71], which is an approach where researchers have less control over the factors and are more open to discoveries that happen during the study.

For the study, we created two visualizations: Harassment Plants and Harassment Information. The visualizations aimed to unveil cases of sexual harassment that occurred in a public space in our city. The Harassment Plants is a physicalization that represents individual stories of women harassed in Açude Velho. The Harassment Information represents the same data but in aggregate and more virtual manner.

The study aimed to investigate whether contrasting data representations in regard to representation granularity, physicality, and situatedness could affect users' compassion. During the second study, we initially used the term *empathy* instead of *compassion* because it was used by Boy and colleagues [14] and some visualization practitioners. However, as the work progressed, we found the term *compassion* fits better with the aim of the visualizations we are designing, namely to make users *feel concerned about* the persons represented in the visualizations and become motivated to help them. This is different from "coming to feel as another person feels", the most common definition of empathy [4]. We, therefore, adopt the term *compassion* as "the feeling that arises in witnessing another's suffering and that motivates a subsequent desire to help" [31]. *Compassion* is explored in the rest of the thesis as self-reported scales or through prosocial behavior (e.g., hypothetical donations to charity). Results from this study are reported in Chapter 5.

1.4.4 Delving into Anthropographics

In 2018 I stayed six months in Paris for an internship in collaboration with Yvonne Jansen (Sorbonne/ISIR) and Pierre Dragicevic (INRIA/AVIZ). During that period, I changed my research focus to more controlled experiments. We also decided to adopt the term Anthropographics and focus on visualizations intended to evoke prosocial feelings (e.g., compassion) and prosocial behavior (e.g., donate).

The first contribution of the internship was delving into the design space of anthropographics proposed by Boy and colleagues [14] in order to improve the terminology and advance the understanding of the design space dimensions. We collected a series of visualizations about people, discussed on relevant design dimensions, and wrote a paper summarising our contributions. The paper was conditionally accepted at TVCG, but it is still under review.

The second contribution of the Brazil/France partnership was a series of crowdsourced studies to investigate to what extent anthropographics in the context of charitable giving may increase prosocial behavior and compassion in comparison to a more traditional visualization (a bar chart). The visualizations represent data about the number of migrants who died or survived while trying to cross a border. The study's aim is contrasting the effects found by Boy et al. [14] in dimensions such as granularity and realism, as well as investigating dimensions still unexplored such as information specificity and authenticity. The experiments' results are shown in Chapter 6.

1.4.5 Design of Physical Anthropographics

The last contribution of this thesis was made after the qualyfing exam's feedback. One of the committee members raised the question of how was the Harassment Plants' design process since the visualization is based on the metaphor that harassments were previously hidden and then sprouted from the ground. Leveraged by the Visualization Sandcastling philosophy [37], I gathered all documents from the design process and came up with an overview of the harassment plant's development.

This historical analysis aimed to investigate how the visualization design evolved and what were the challenges during the process. The lessons learned are discussed in Chapter 4.

1.5 Thesis Overview

Chapter 2: Background

This chapter clarifies why this work focuses on compassion and charitable giving and what has been done in the Visualization literature concerning this topic.

Chapter 3: Design Space of Anthropographics

This chapter provides conceptual foundations of Anthropographics, delineates its design space, and illustrates different examples of visualizations designed to represent people.

Chapter 4: Design of a Situated and Physical Anthropographic

This chapter gives an overview of the development of a situated and physical anthropographic as well as discusses the lessons learned throughout the process.

Chapter 5: In-the-wild Study

This chapter presents the results of a study that investigated the effect of a situated and physical anthropographic (vs. a non-situated and virtual one) on people's compassion.

Chapter 6: Charitable Giving Experiments

This chapter reports the results of a series of crowdsourced studies that examined to what extent anthropographics about humanitarian issues affect people's compassion in a charitable giving context.

Chapter 7: Conclusion

This chapter summarises the thesis contributions and sheds light on future research on Anthropographics.

Chapter 2

Background

This chapter discusses why the thesis explores topics of compassion and charitable giving and what has been done in Visualization literature.

2.1 Empathy vs. Compassion

Empathy and *compassion* have been defined loosely in the literature. *Empathy* can be defined as (a) “coming to feel as another person feels”, (b) “feeling distressed at witnessing another person’s suffering”, and (c) “feeling for another person who is suffering” [4], to name a few. The definitions of *compassion* are less diverse than those of empathy, but the use of the concept in research is complicated by the popularity of several synonyms such as *sympathy*, *pity*, and *empathic concern*. Goetz, Keltner, and Simon-Thomas [31] suggest these synonyms belong into a group of “compassion-related states that center upon a concern for relieving the suffering of others”.

In some cases, researchers and practitioners whom we cite in this work use the term *empathy* to refer to *empathic concern*, which is closer to the definition of *compassion*. Based on the work of Jordan, Amir, and Bloom [42], we assume *empathy* and *compassion* are different constructs. In this work, we consider that *empathy* is “the act of experiencing what you believe others are experiencing” [11] and *compassion* is “the feeling that arises in witnessing another’s suffering and that motivates a subsequent desire to help” [31]. Therefore, when citing research that originally uses the term *empathy* to refer to the definition of *empathic concern*, we consider it as *compassion* to assure consistency. This decision reverbs some of the literature on Behavioral Sciences [22].

Another motivation for adopting *compassion* is because this term relates more closely with the objective of the data representations designed in this work. We aim to create visualizations that cause an emotional appeal to users but also motivate them to act towards the causes of the persons represented in the chart. Research suggests that differently from empathy, which leads to actions such as escaping a situation to reduce one's suffering, the experience of compassion results in behaviors that reduce the other's suffering [31]. Thus, we use self-reported responses and prosocial behavior as a proxy to capture compassion, since this emotion is a positive predictor of prosocial behavior [85], whereas empathy "is either not predictive or negatively predictive of prosocial actions" [42].

2.2 Design for Compassion

This section describes some hypotheses of how visualization design may influence users' compassion towards people represented by visualizations. The formula to provoke compassion seems to pass through the use of information-rich, relatable, and identifiable data representations.

2.2.1 Anthropomorphism

One of the most widely used strategies by practitioners to evoke compassion is representing people through anthropomorphic marks (e.g., Figures 3.4 and 3.12), which are also known as wee people, or human silhouettes. Using anthropomorphism in the context of robotics is pointed as a factor to influence humans to emphasize more with robots [70]. Ivanov, Danyluk, and Willett [39] argue that using anthropomorphic marks in the context of Data Visualization is interesting because "they are highly recognizable and immediately suggest that the underlying data corresponds to individual people". However, Harris [36] believes that anthropomorphic visualizations should not stand for aggregates to evoke compassion. Each mark should correspond to a single person. Nevertheless, all claims still need to be tested empirically.

2.2.2 Identifiable Victims

Another design choice used to increase the individuality of people in data representations is presenting more information about each person. Harris [36] suggests that showing individuals through images, personal details, or stories may make users more compassionate with what they see. This design strategy is supported by Psychology literature, which has already found reliable results in favor of the identifiable victim effect [30]. This effect means that users tend to empathize and donate more to vividly identifiable victims than for less identifiable individuals. To the best of our knowledge, this effect was not studied in Data Visualization literature, except for the work of Boy and colleagues [14].

2.2.3 Near and far

Harris [36] also proposes the concept of Near and Far. This notion involves presenting the general and specific levels of information in the same visualization. Displaying a data summary allows users to understand a broader overview of the numbers. Conversely, observing a specific facet of data in detail may lead users to relate with that data [28].

Studies have shown that it is difficult for people to empathize with large numbers of victims. It is easier to feel concerned with a picture of a single starving child in Africa, but it is hard to multiply the feeling for the annual report of children who died from starvation worldwide. The effect known as *compassion fade* [82] proposes that compassion decreases as the number of victims increases. The *near and far* approach may help to overcome the compassion fade [83] by showing not only the aggregated numbers but also specific perspectives.

2.3 How to Measure Compassion

There is a vast collection of measures in Psychology literature devised to capture compassion, ranging from simple scales to long questionnaires. Strauss and colleagues [77] systematically reviewed such measures and came out with a list of nine questionnaires. Gu and colleagues [34] narrowed down the list to four measures and proposed a general questionnaire based on them. The questionnaires have generic questions that capture to what extent a person is compassionate in general. Nevertheless, no existing approach captures the construct in its totality [77; 34], and there is a lack of tools that explore the development of compassion in more specific scenarios such as while experiencing a data visualization. For that reason, this work does not use the previously mentioned measures of compassion.

Compassion is often conceptualized as a construct that contains at least two components: a subjective and a behavioral [82]. This perspective is in line with the definition presented in Section 2.1, which affirms that a person that becomes compassionate is aware of another's distress but also becomes motivated to act to relieve that suffering. Some studies capture the affective and behavioral components of compassion [47; 82] (e.g., self-reported compassion scale and donations) as a form of triangulation, while others (e.g., [56]) measure prosocial actions such as willingness to help or donate as a proxy for the behavioral component of compassion. This work has a primary focus on capturing prosocial behavior through hypothetical donations but also explores compassion using self-reported scales. We chose to use donations as our primary outcome because the responses are less prone to human subjectivity. We also opted to explore self-reported compassion for the sake of construct completeness.

2.4 Studies on Compassion and Charitable Donation

The trend of exploring emotional responses (e.g., empathy or compassion) and prosocial behavior is still new in the Visualization literature. Experts recently called for new ways of assessing the value of visualizations [84], but only a few initiatives have explored factors that go beyond or complement utilitarian purposes.

Under the broad umbrella of Visualization literature, the Virtual Reality field has contributed to exploring how such immersive environments affect people's empathy or prosocial behavior. Ivanov, Danylik, and Willett [39], for example, have argued that using anthropomorphic objects in virtual environments might contribute to increase viewer's level of empathy since "the level of fidelity [of such environments] has the potential to enable a kind of direct face-to-face relationship with visualizations". Other works empirically investigated how VR environments affect people's empathy and prosocial behavior. Calvert, Abadia, and Tauseef [15] have examined through custom questions on whether showing a story using a

VR technology vs. a monitor could affect empathy. Kandaurova and Lee [44] conducted a similar experiment in the context of charitable giving. The authors captured empathy, prosocial behavior (donation of time and money), and other factors such as guilt and responsibility. In the end, all studies suggest that people have higher levels of empathy and are more likely to behave prosocially when exposed to virtual environments.

In the context of Data Visualization, only a few studies have investigated the role of charts on compassion or charitable giving. Most studies do not directly focus on compassion but capture emotional responses similar to the affective component of such construct [27; 38; 45]. The qualitative study conducted by Kennedy and Hill [45], for example, explored emotional engagements with data and visualizations and discovered that participants seem to have become compassionate with migrants after seeing a visualization about the topic. Erlandsson and colleagues [26], on the other hand, conducted a more focused study where they investigated to what extent people allocate money according to the visualization that is presented. The experiment showed different trends concerning cancer death risks, and results suggest that people tend to donate more to projects when they see upward trajectories in the chart. As far as we are aware, only Boy and colleagues [14] investigated both compassion and prosocial behavior in the context of Data Visualization for human rights. The authors introduced the concept of anthropographics and explored the effects of such visualizations on empathic concern and donation allocation. Results were inconclusive but have fostered the opportunity for further investigation. This thesis goes further onto Boy et al.'s work by advancing the design space of Anthropographics and conducting experiments that explore different facets of the design space.

Chapter 3

Design Space of Anthropographics

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 chapter is to contribute to filling 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. [14], 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 work 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, 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* [5]: it has been devised to help designers think more clearly about existing and possible designs. However, it is not meant to be *evaluative* [5]: 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.

3.1 Methodology

The starting point of our work was the design space of anthropographics by Boy et al. [14]. 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.

3.1.1 Scope

We collected a total of 105 visualizations that convey data about people (see Appendix A for a fixed list, or the interactive website for an updated list[1]). 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. [14]. 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.

[1]Interactive list of visualizations: `https://luizaugustomm.github.io/anthropographics/`.

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[2].

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.

3.1.2 Development of the Design Space

We started with the set of dimensions proposed by Boy et al. [14], 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 3.2.1 and 3.2.2) 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 easily operationalized were removed (see the justifications in Section 3.2.7). During the process, we also came up with new dimensions, partly inspired from past literature in psychology (e.g., [82]) and data visualization (e.g., [50]).

[2]See, for example, *Visualizing smoking risk*: https://www.stubbornmule.net/2010/10/visualizing-smoking-risk/

After a seemingly stable set of definitions and dimensions was established, and the corpus of visualizations had been fully categorized, we performed a multi-coder evaluation. We collectively wrote a codebook, which three used to classify a random sample of 17 visualizations. Afterward, 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 [14] (see Section 3.2.7 for more details), and four new dimensions.

3.2 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.

3.2.1 General Visualization Concepts

We assume for simplicity that all **datasets** are flat tables [59]. A **data item** (or simply **item**) is "an individual entity that is discrete", and which corresponds to a row in the table [59]. 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" [59]. 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 [59; 86; 23; 75]. A mark can either consist of a single graphical primitive (e.g., a point, line, or area [7]) or a combination thereof (e.g., a glyph or an icon) [76]. Meanwhile, **perceptual channels**[3] 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 [7]),

[3]This term generalizes the notion of visual channel [59] or visual variable [74] to non-visual data representations [40].

non-graphical attributes (e.g., weight [40]), as well as complex combinations of attributes (e.g., when using icons to convey categorical attribute values).

3.2.2 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. 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.

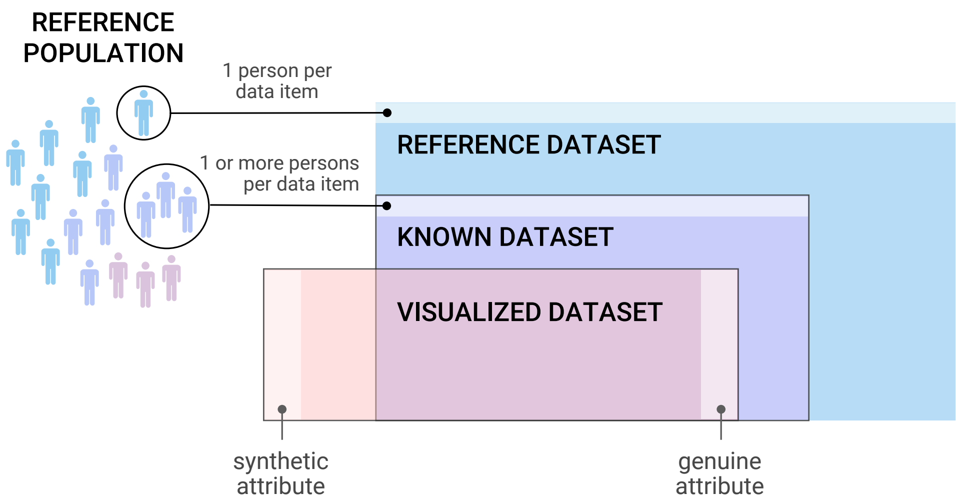


Figure 3.1: The three conceptual datasets we use to capture the message and information conveyed by an anthropographic.

In our conceptual framework, the message and the information conveyed by a visualization involve three datasets (see Figure 3.1). 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[4]. A known dataset is typically a regular dataset (e.g., a csv file).

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 names (e.g., Figure 3.14). 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 of 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 us 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

[4]Missing Migrants website: https://missingmigrants.iom.int/

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 us 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.

3.2.3 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 3.1). The reference population contains the same ten people.

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).

3.2.4 What is Shown

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

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 3.1: Fictional dataset used in Figures 3.3, 3.5, 3.8, 3.13, 3.15, 3.18, and 3.21.

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 3.3-left). An example is depicted in Figure 3.2, 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 3.3-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 3.3-center). Many so-called Isotype visualizations [62; 61] fall into this category. Figure 3.4 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

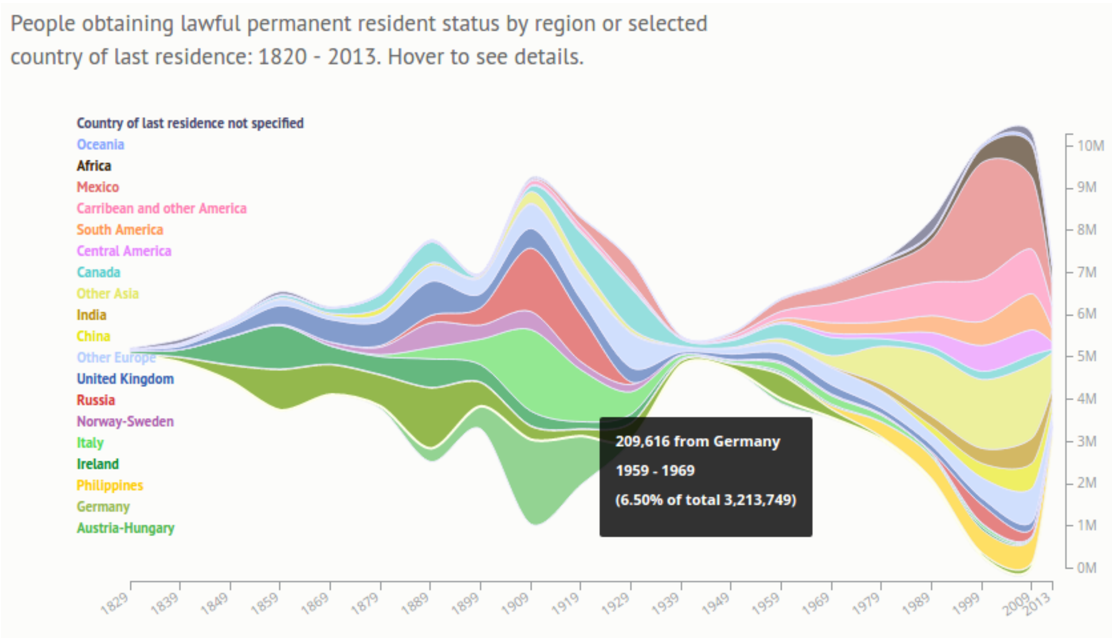


Figure 3.2: *200 Years of Immigration to the US*. This stream graph shows people who obtained lawful permanent resident status by region from 1820 to 2015. Each color corresponds either to the country of last residence or a region. The area represents the number of immigrants who moved to the U.S. over 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 (`http://insightfulinteraction.com/immigration200years.html`).

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 3.3-right). For example, the visualization in Figure 3.6 has maximum 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. On the other hand, 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,

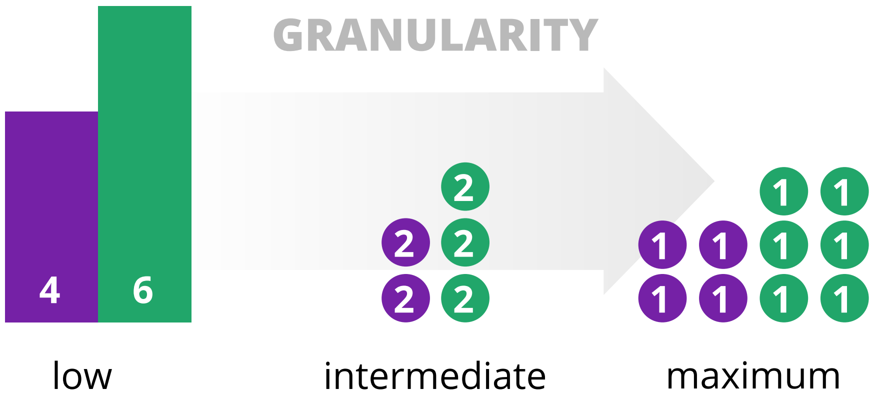


Figure 3.3: *Granularity Continuum*. *Low*: each mark represents a group of people aggregated by one or more attributes (here, their preference for dogs or cats). *Intermediate*: each mark represents a fixed number of persons (here, two). *Maximum*: each mark corresponds to a single person.

or synthetic attributes to humanize the data (e.g., some of Boy et al.'s designs [14]). There is also a belief that representing people using single marks can make readers empathize with the persons represented [73]. However, the 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 [14]. More studies are needed to examine the role of granularity in promoting prosocial feelings or behavior.

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 Section 3.2.2 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

Figure 3.4: *The Great War 1914-18.* An 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 (http://www.laworrell.com/blog/2015/1/14/isotype-international-system-of-typographic-picture-education).

whether they are genuine or synthetic.

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 3.5-left). The Isotype visualization in Figure 3.4 is an example of a low-specificity visualization: the low distinctiveness of the visualized attributes (survival status and side) contributes to making the soldiers look rather deindividualized. The gun violence visualization (Figure 3.6) 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.

Figure 3.5: *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.

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 3.7, the visualization shows six attributes. Although each attribute has low distinctiveness, once combined, the attributes give some sense of individuality to each person.

Finally, in a visualization with **high specificity**, the attributes in the visualized dataset allow the reader to perfectly distinguish all data items (see Figure 3.5-right). One example is shown in Figure 3.11, 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 3.9, for example, the visualization shows the 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 [47; 51]. Although the study from Boy et al [14] 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. Al-

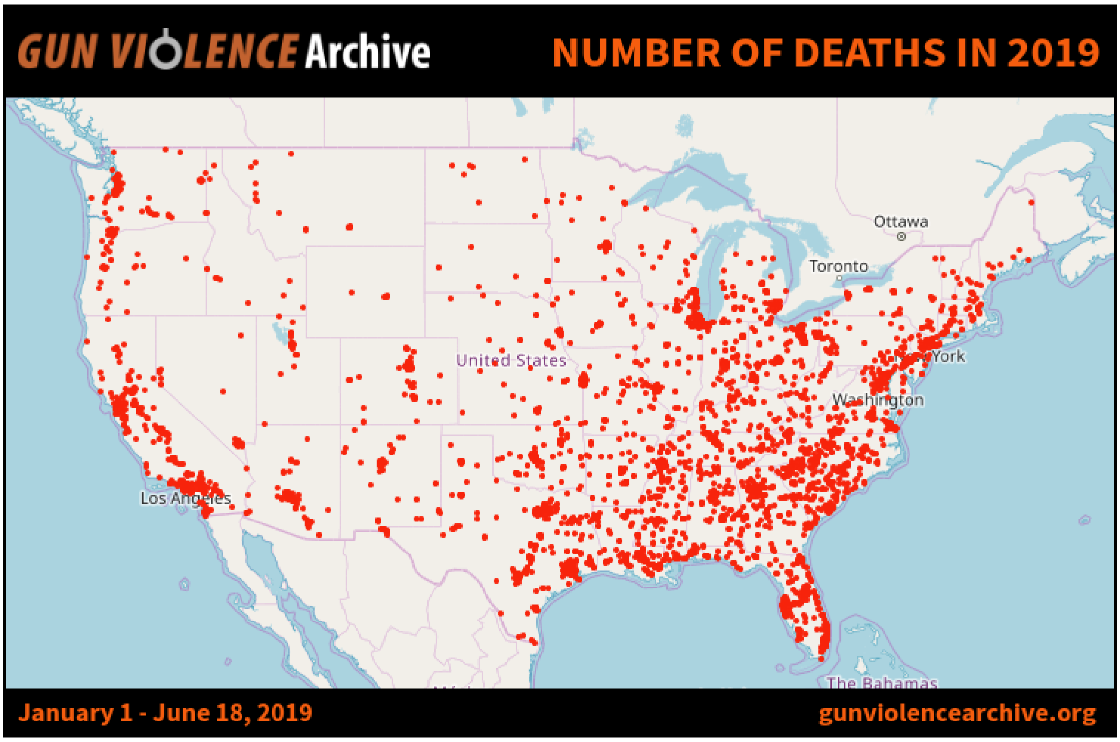


Figure 3.6: *Gun violence Chart of Deaths in 2019*. Depicts all deaths caused by guns in the USA from January to June, 2019. Each red dot is a victim, and its placement indicates where the person was killed. Source: Gun Violence Archive (https://www.gunviolencearchive.org).

though privacy issues often make it hard to reach high specificity in a visualization [14], 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 3.10.

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.

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

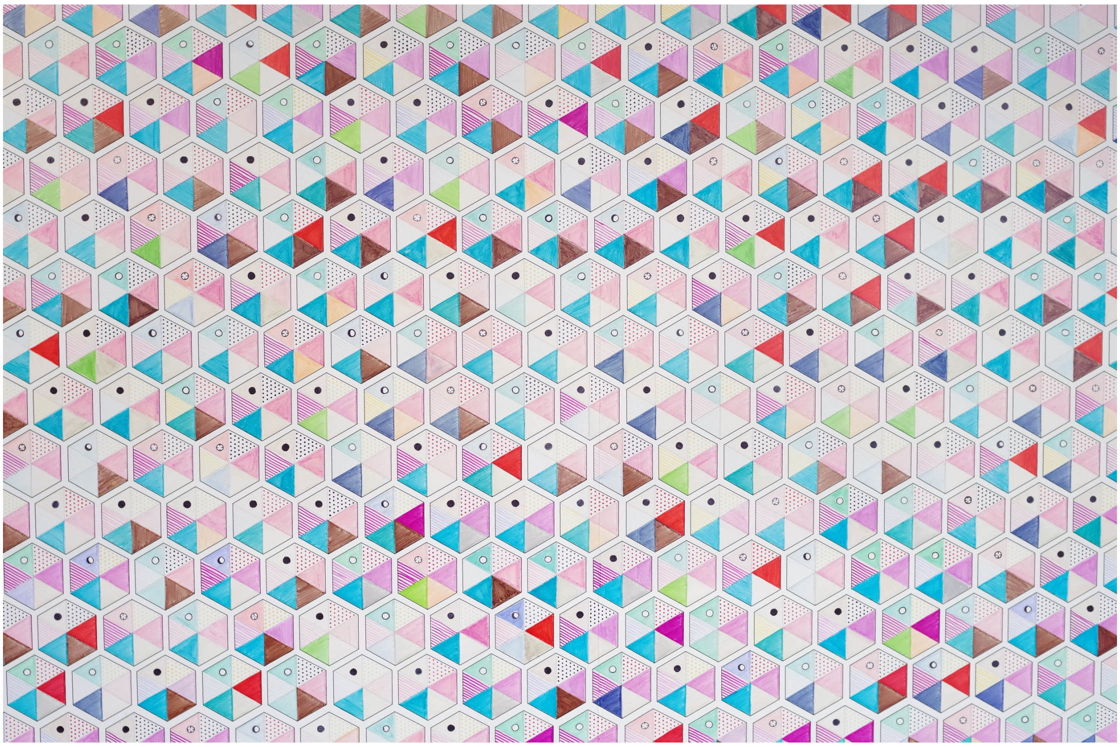


Figure 3.7: *Data wallpaper*. Each glyph represents 6 questions answered by a visitor to a store, ranging from how they see the future of work to how they unlock their creativity. The designer initially installed an empty visualization template as a wallpaper in the store, and the store employees filled the hexagons as they were receiving visitor responses. Source: Giorgia Lupi (http://giorgialupi.com/collaborative-data-wallpaper-for-story/).

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 3.10 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 1.2: “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 to tell a broader story about the arduous life of thousands of refugees around the world.

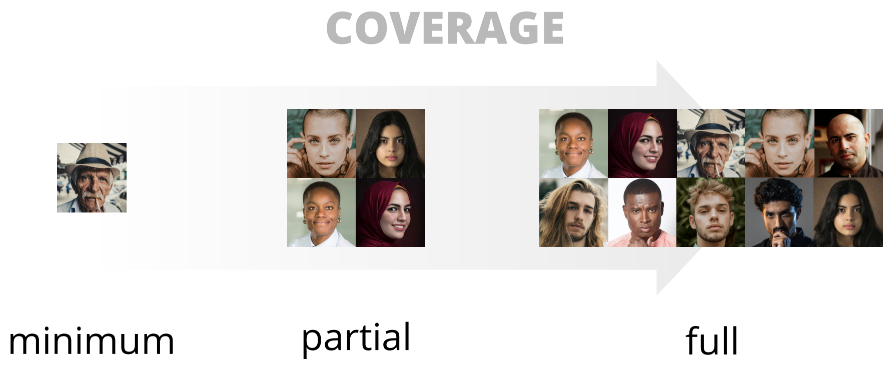


Figure 3.8: *Coverage continuum*. *Minimum*: a single person from the reference dataset is visualized. *Partial*: a subset of people from the reference dataset is visualized. *Full*: all people from the reference dataset are visualized.

In a **full coverage** visualization, the visualized dataset contains *all* the persons from the reference dataset. See Figures 1.1, 3.6, and 3.12 for example.

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 3.10). 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 [47]. Similarly, studies have suggested that as the number of suffering people increases, people feel less empathy for them and donate less [82]. Therefore, it is possible that visualizations with partial or minimum coverage could help observers be more compassionate about suffering populations. However, it remains necessary to test this hypothesis empirically.

Authenticity

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

A visualization with **partial authenticity** contains both genuine and synthetic attributes

Figure 3.9: *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 (http://giorgialupi.com/bruises-the-data-we-dont-see).

(terms defined in Section 3.2.2). The more visualized attributes are synthetic, the less authentic the visualization is. In Figure 3.13-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 3.1) which makes these attributes *synthetic*. A real example of a partially authentic visualization is shown in Figure 3.14. 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.

In a visualization with **full authenticity**, all visualized attributes are genuine. In Figure 3.13-right, the second visualization is fully authentic because all the information presented comes from the known dataset. Although it uses realistic silhouettes like the visual-

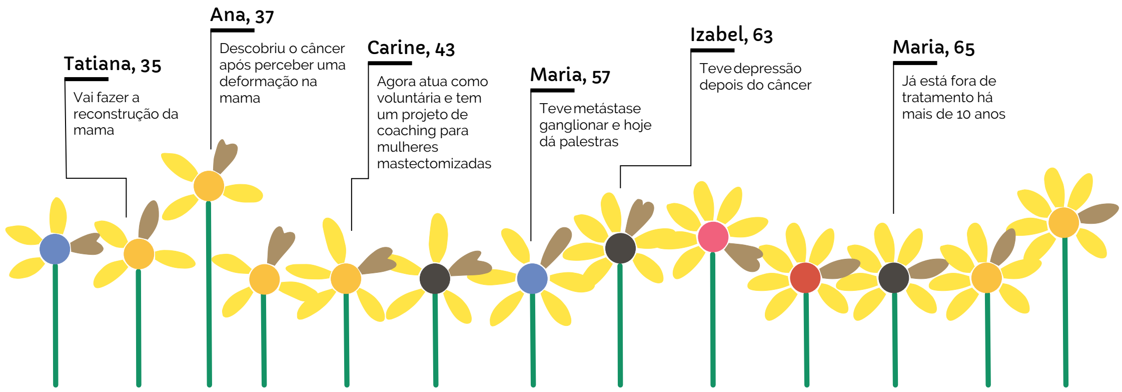


Figure 3.10: *Cancer is not always the end* shows data about 13 women with breast cancer who has been in remission for years. Each flower is a woman and each part of a flower shows a characteristic of the woman such as her age or how she dealt with the treatment. Information about some women is also shown by annotations. Source: Oncoguia (http://www.oncoguia.org.br/conteudo/small-data-e-o-cancer-de-mama-a-jornada-da-paciente/12899/1195/).

ization 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 3.14 or in Visualization 67 in Appendix 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.

3.2.5 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

Figure 3.11: *100% Montréal* is composed of 100 persons from Montréal who have been selected to represent the city 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 (`https://www.rimini-protokoll.de/website/en/project/100-montreal`)

marks are represented.

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 3.15-left). Such marks include dots, bars, abstract glyphs, or shapes that evoke inanimate objects. Figures 3.2, 3.7, and 3.10 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 3.15-center. The visualizations from figures 3.12 and 3.14 are also visualizations with intermediate realism since they are composed of human-shaped icons or silhouettes.

A visualization with **high realism** is made of *realistic anthropomorphic marks*, which closely resemble an actual person. Figure 3.15-right shows two cases of realistic marks: a

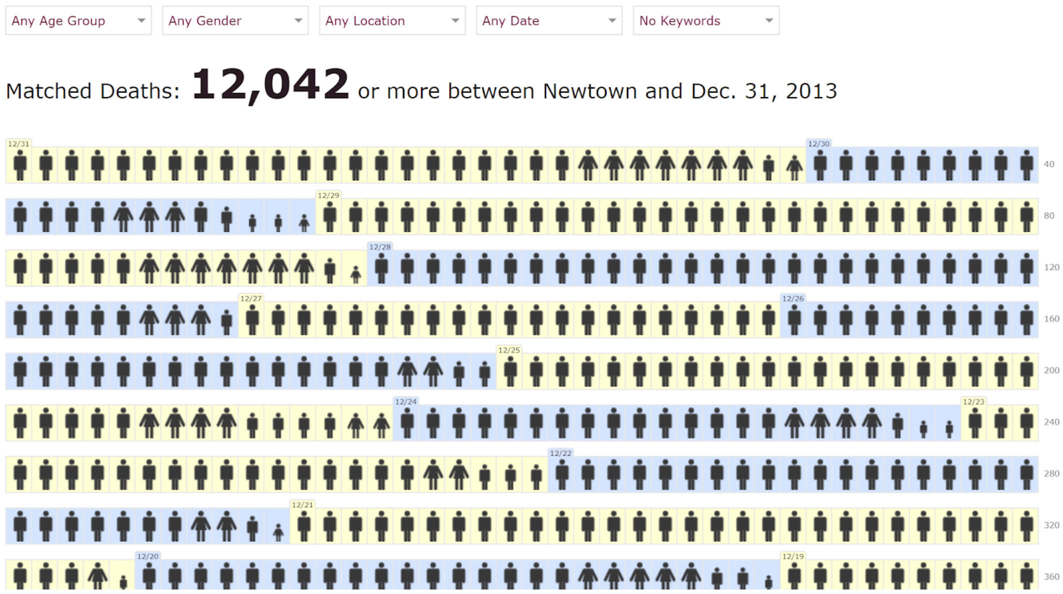


Figure 3.12: *How Many People Have Been Killed by Guns in USA Since Newtown's attack*. This visualization shows people who died due to a gun shot between the Newtown attack in 2012, and December, 2013. Each icon is a person. Sex and age are shown through the icon's shape. Additional information about each person can be obtained by clicking on their icon. Source: Slate (http://www.slate.com/articles/news_and_politics/crime/2012/12/gun_death_tally_every_american_gun_death_since_newtown_sandy_hook_shooting.html).

detailed drawing of a person and a photograph. A realistic anthropomorphic visualization is shown in Figure 3.16, where the marks are detailed drawings of the richest 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 3.11 and 3.17.

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 [14; 36; 16], and that the more realistic the marks are, the more effective they could be at promoting empathy [14]. However, none of these hypotheses has been experimentally confirmed.

Figure 3.13: *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.

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 [40]. Figure 3.18 illustrates the physicality continuum.

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 3.4, 3.14, and 3.16) fall in this category. The visualization from figure 3.7 also have low physicality because it is drawn on a flat wall.

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 [50] 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 3.20, where the marks are grains of rice, and in Figures 3.11 and 3.17, where the marks are real persons.

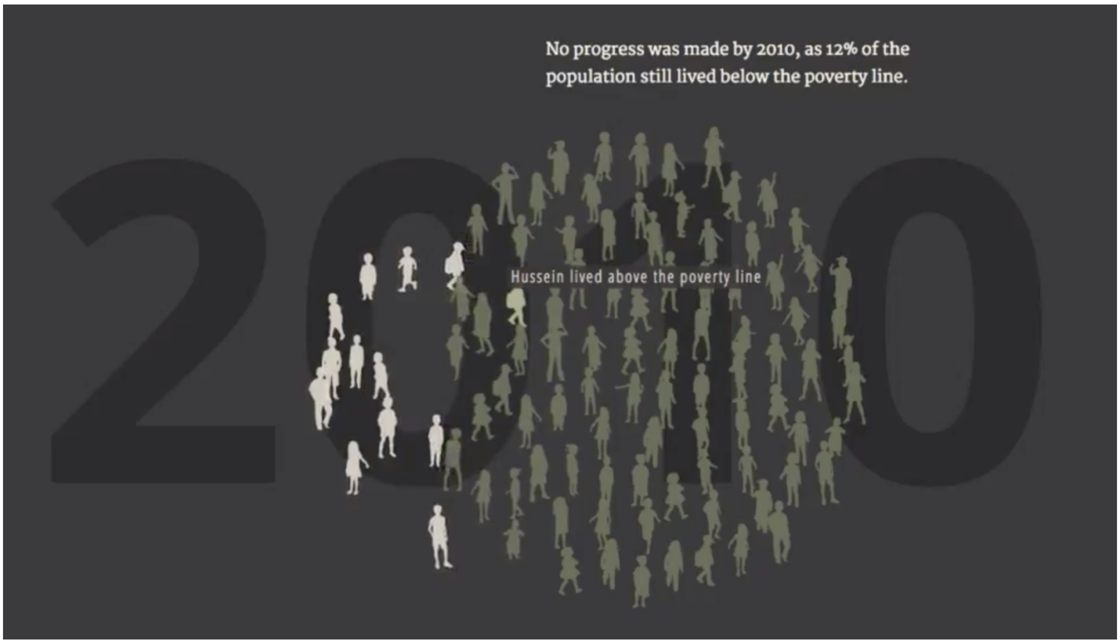


Figure 3.14: *Poverty in Syria* shows the proportion of children below (darker color) or above the poverty line (lighter color) in Syria, in 2010. Each mark is a fixed number of children. Hovering over a mark shows synthetic information about a person. Source: Boy et al. [14].

Le Goc et al. [50] hypothesize that it is easier to empathize with people when they are represented by physical objects than when virtual objects represent them. Nevertheless, it is so far unknown whether using physical marks may indeed increase prosocial feelings such as empathy or compassion.

Situatedness

Situatedness refers to how spatially close the mark's physical presentation [79] is — or was[5] — to its physical referent [86]. In the context of anthropographics, the physical referents are the persons the data describes. Figure 3.21 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 3.12 and 3.6 are examples of non-situated visualizations because they are displayed on computer screens, which are far from the represented victims in most cases.

[5] As we will see, we generalize Willet et al.'s [86] notion of situatedness by also considering spatial relationships in the past.

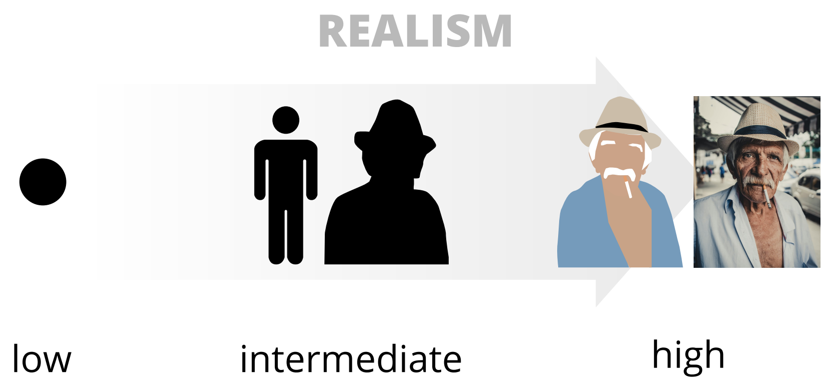


Figure 3.15: *Realism continuum. Low*: the marks do not evoke a person. *Intermediate*: the marks are simplified depictions of a person. *High*: the marks are realistic depictions of a person.

In a visualization with **intermediate situatedness**, the marks are either presented at a location where the persons used to be in the past or the marks used to be in proximity to the persons they represent. The Data Wallpaper (see Figure 3.7) 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 3.19), 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 3.11) 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

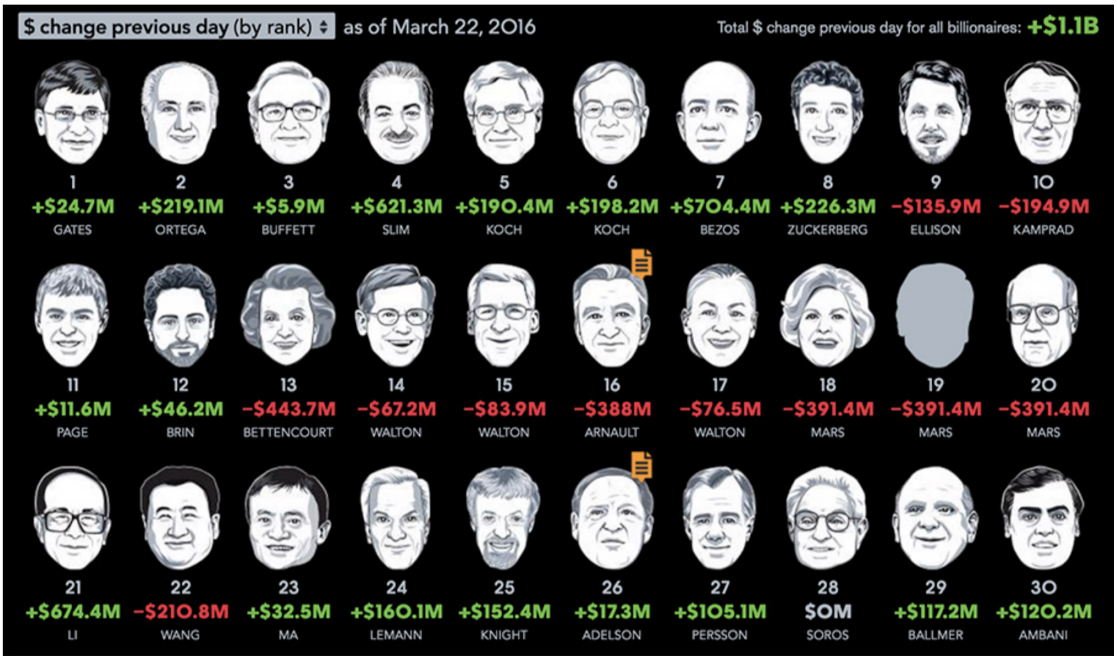


Figure 3.16: *The Billionaire Index* presents data about the 30 richest persons in the world. Below each mark is an index (corresponding to the position of the person in the ranking of billionaires), the last change (in dollars) and the person's name. Color encodes whether the last change was positive or negative. Source: Bloomberg (https://www.bloomberg.com/billionaires/).

significant event has affected a person or a group of persons, and memorabilia can acquire emotional power by having been touched or worn by a person [63] (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 [86].

3.2.6 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) [1]. 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

Figure 3.17: *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: Courtesy of San Francisco Gay Men's Chorus.

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 3.22, 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 3.14 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 3.22, 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. Also, in interactive visualizations, users can focus on people's attributes they care about the most.

Finally, dynamic visualizations can let users explore the same data — items and attributes — through *different representations*. In Figure 3.23, 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

Figure 3.18: *Physicality continuum*. *Low*: the marks are shown on a flat medium. *Maximum*: the marks are physical entities.

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 1.1, 3.6, 3.14, and 3.22, 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 3.23 lets users change its granularity from intermediate (every dot represents five persons) to low (all the dots are combined to form a bar). 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 3.22, 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 3.14 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 3.22, 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. Also, in

Figure 3.19: *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.

interactive visualizations, users can focus on people's attributes they care about the most.

3.2.7 Differences with Boy et al.

As we mentioned previously, our design space of anthropographics extends an earlier proposal by Boy et al. [14]. Our extension both broadens the original design space (that is, it captures a larger variety of designs) and sharpens it (that is, it makes a 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 represents a single person, and those where each mark represents a fixed number of persons.

Figure 3.20: *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: emydot user on Flickr (https://www.flickr.com/photos/24463781@N02/2316420045/).

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 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 visualization expressiveness often arises largely from the meaning of the dataset and it can be manipulated by a variety of visual design strategies like the use of metaphors, which 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 specif-

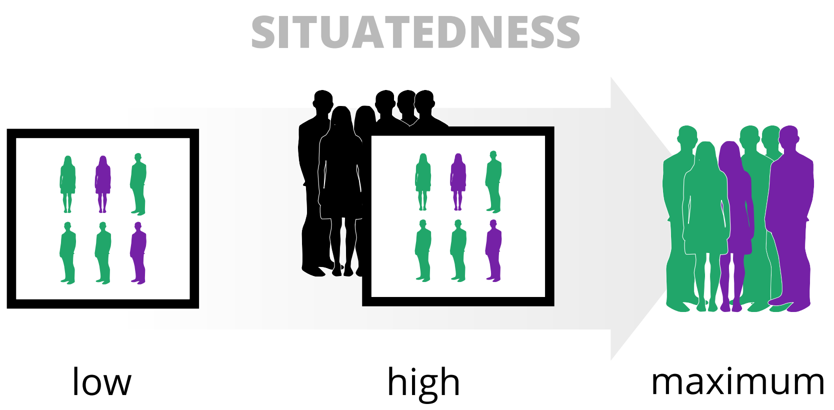


Figure 3.21: Situatedness continuum. *Low*: the marks are presented far from the persons they represent, or their physical location is not under the control of the designer. *High*: the marks are presented close to the persons they represent (the persons in black). *Maximum*: the marks are the persons they represent.

ically 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 many 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

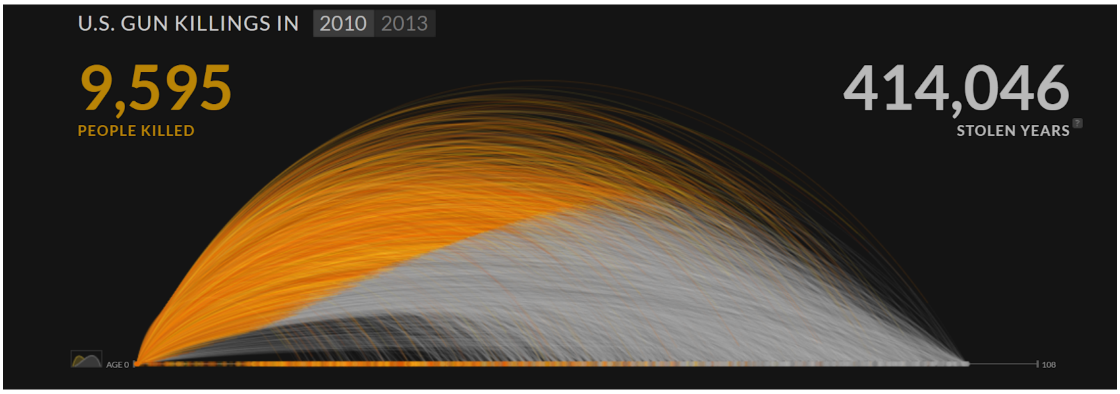


Figure 3.22: *U.S. Gun Deaths in 2010 and 2013*. Each line in this interactive unit visualization represents a person. The orange segment corresponds to the period lived and the gray segment represents the estimated years stolen from the person. Readers can select the dataset by choosing the year and also filter by sex, age, region, and time. Source: Periscopic (https://guns.periscopic.com/?year=2013).

"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. However, 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 to any empirical finding.

3.3 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 specific type of design that we did not particularly emphasize during our search and yet appears a lot in our collection, then this indicates 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.

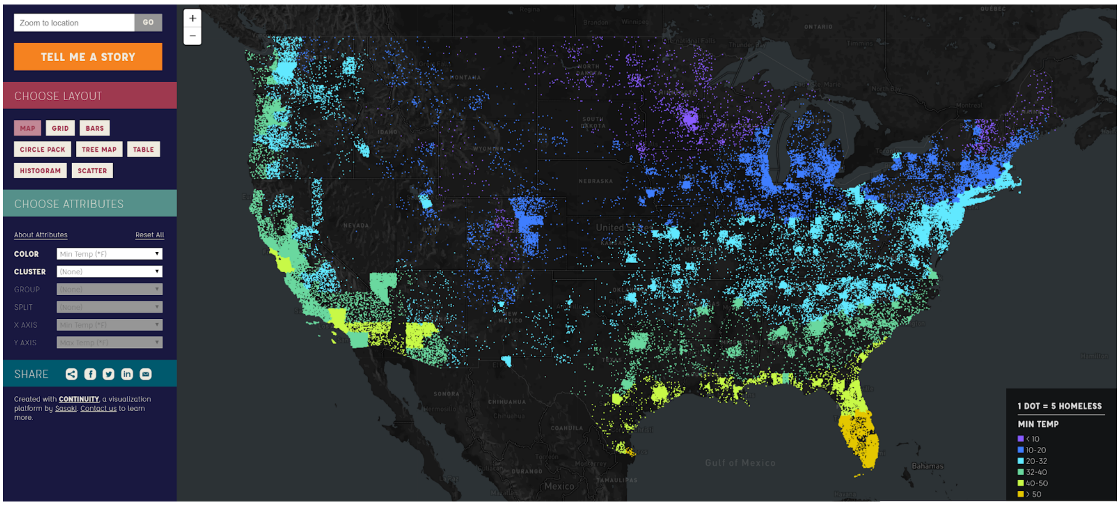


Figure 3.23: *Understanding homelessness in USA*. This interactive unit visualization represents the homeless population in the USA. Each dot corresponds to five homeless people. The representations can be chosen among maps, grids, bars, etc. The viewer can also explore different factors from categories such as geographic, economic, or social. Source: Understanding Homelessness (http://www.understandhomelessness.com/explore/).

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 broad 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.

3.3.1 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 3.19, 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 3.2, 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 3.6 and 3.20 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 3.10 and 3.22.

3.3.2 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 [14], 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 3.4 and 3.14.

Individual wee-people charts are anthropomorphic unit visualizations with intermediate realism and maximum granularity. They have varying degrees of information specificity: visualizations 10 and 24 (see Appendix A) have low information specificity, while the one in Figure 1.1 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 3.16, 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.

3.3.3 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 3.14, where neither the name nor the gender of the silhouettes originates from the known dataset. Another example is presented in Figure 3.17, 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 3.9 provides a striking example. Many other examples exist outside of anthropographics, such as biographical and autobiographical visualizations [2; 66; 67]. 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 3.9 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" [54].

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 3.10 and 1.2 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 3.11) 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 Appendix 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 [40] — 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 3.20) while others are anthropomorphic (Figure 3.17).

3.4 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.

3.4.1 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.

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 that argues that charity should be driven by reason and facts to have a real impact [55]. Many of the anthropographic designs discussed in this paper may not satisfy such thirst for factual information. They 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 [43]. 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* [36]. 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.

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 [10]. 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 Section 3.2.6, we discussed an example of a visualization that lets users select the attributes to visualize and filter people by their demographic characteristics (see Figure 3.22). This kind of chart offers users the opportunity to focus on the type of person or the characteristics they care about the most. Visualization 78 in Appendix 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 [32] 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.

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 [82]. For example, in Figure 1.2, 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 Appendix A or Figure 3.10 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 3.9) comes closest to a minimum-coverage visualization but it remains ambiguous (see *single-person charts* in Section 3.3 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.

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 Section 3.2.5, 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 Appendix 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 Appendix 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 [72].

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 3.7), 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 [63]. 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[6] are fully situated visualizations when they convey data about the people who are wearing them [86]. 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

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

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.

3.4.2 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.

Testing Basic Design Dimensions

Boy et al. [14] 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 mostly 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 [19]. Furthermore, many other designs remain to be tested, some of which could prove more effective. For example, the anthropographics used by Boy et al. [14] 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.

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 [82]. 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.

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 3.11) and wearable physicalizations would be particularly interesting to test since they are maximum in both the situatedness and physicality dimensions.

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 3.4, where each mark explicitly represents 1 million soldiers — coverage and granularity become ambiguous. In Figure 3.11, 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?

3.4.3 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 Section 3.2.1). Although tables are likely the most common data model, other types of datasets exist such as networks [59]. 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 3.1). 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 3.1), 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 [69; 80]. 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 to 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 3.9). 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 [65]. 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.

Chapter 4

Design of a Situated and Physical Anthropographic

This chapter examines the conception, development, and exhibition of a situated physicalization that represents stories of harassment experienced by women in a public lakeside in a 400,000-inhabitants Brazilian city. We analyzed the physicalization's design from the perspective of historical documentation (e.g., sketches, photographs, and documents) and of the authors' experience to explain the detours that took place throughout the process. This work chronicles the design process and leverages the theoretical lens of sandcastling [37] to interpret it.

We created the physicalization as part of a research project where researchers investigated the effect of visualization's situatedness and physicality on compassion. During the project, our plans for data collection and physicalization design have changed drastically due to a multitude of forces. The analysis presented in this chapter contributes to the practitioner-oriented literature of Data Physicalization and Anthropographics by examining the challenges faced at the different stages of a physicalization's development.

4.1 Context

This section describes how the physicalization was conceived and where it was exhibited.

4.1.1 Project

The physicalization described in this chapter is part of a research project where researchers were interested in investigating the effects of situated and physical visualizations on compassion. The plan was to design and examine two different visualizations that represent the same data. The first would contain a series of bar charts in a printed poster and would be shown out of context. The second would be physical and situated — i.e., presented in the place the data refers to. This chapter focuses solely on depicting the design process of the latter.

4.1.2 Motivation

Our initial motivation was to visualize data that leads to discussions about the power imbalance that occurs in public space concerning gender. In line with the concept of uncounted, undercounted, or silenced bodies [24], we were interested in exploring the often unnoticed differences in the use of space between women and men. According to the geography of fear [81], women tend to avoid certain places at specific times to stay safe, which limits their use of space. We initially decided to create a physicalization to make passers-by reflect on women's negative experiences in the public space. We explain in Section 4.3 that the initial topic has changed because of a series of challenges.

4.1.3 Public Space

The place we chose to exhibit the physicalization is called Açude Velho — a public lakeside at the downtown of Campina Grande, a 400,000-inhabitants city in Brazil. Açude Velho is considered one of the main destinations for people who want to leisure or practice sports in Campina Grande. Despite the gender-friendly appearance of the space, a study [48] about the presence of people in Açude Velho points out that the percentage of women is lower than men's at most periods of the day. Therefore, the place that we chose justifies our initial motivation since there might be a power imbalance related to gender.

Figure 4.1: Açude Velho. One of the most popular places for leisure in Campina Grande, Brazil. The lake is surrounded by cycle paths and sidewalks.

4.2 Method

The challenges and lessons described in this chapter come from the team's experience during the physicalization development and historical documents. The historical documents were created as part of the design process and comprised sandcastles (i.e., intermediate sketches, prototypes, etc.), design justification documents, focus group recordings, notes, among other resources. Besides historical records, two team members (including the thesis author) wrote down the development process as they recall it separately and listed challenges faced throughout the process. These members compared and discussed their reports, and created together an overview of the design process, which is presented in Figures 4.2, 4.4, and 4.5.

4.2.1 Visualization Sandcastling

We use the metaphor of visualization sandcastling [37] to depict an overview of the physicalization's design process. We use the concept of detours and consider visualization development as a speculative process. Each new design decision has produced a sandcastle, which contains lessons learned by itself. We explain the sandcastling metaphor and its concepts in the following paragraphs.

Sandcastling is a philosophy for designing visualizations. According to the metaphor, visualizations are sandcastles because designers can shape them according to the **detours** (e.g., decisions, limitations, new requisites) that occur in the course of the design process. Instead of focusing on predetermined goals, sandcastling encourages designers to rebuild the visualization based on criticism, questions, and ideas that emerge throughout the development of sandcastles. Sandcastles can be considered as three different things.

Sandcastles are **aesthetic provocations**. Creating visual representations (either based on data or not) allows people to reflect on the visual components and, consequently, to refine the visualization concept. The aesthetic concerns may evolve as data collection proceeds, but data can also be reshaped according to the design decisions made throughout the speculative process. Aesthetic provocations thus contribute to producing knowledge for researchers and designers during each step in the design process.

Sandcastles are a **speculative process**. Designing a visualization is never a straight path. There are always preliminary sketches or prototypes that are then refined through a speculative process. The sandcastles are transient, unstable, and unfinished, but they do help designers to get insights and speculate about design decisions that would not be noticeable without a visual representation [37].

Finally, sandcastles are **dynamic mediators** because they are visual manifestations of individual ideas [37]. As such, visualizations serve as mediators between people from different domains (e.g., researchers, end-users, designers), allowing them to reflect and discuss how information is obtained and represented critically.

4.2.2 Team Roles

The project was composed of four members: three visualization design researchers and one product designer. The visualization researchers were responsible for selecting and validating the physicalization's topic, collecting data, devising visualization designs, constructing prototypes and the final physicalization, and transporting the visualization to the public space. The product designer provided advice regarding the design process and materials to use. Besides the primary team, women that are used to visit Açude Velho also contributed to the processes of topic validation and data collection.

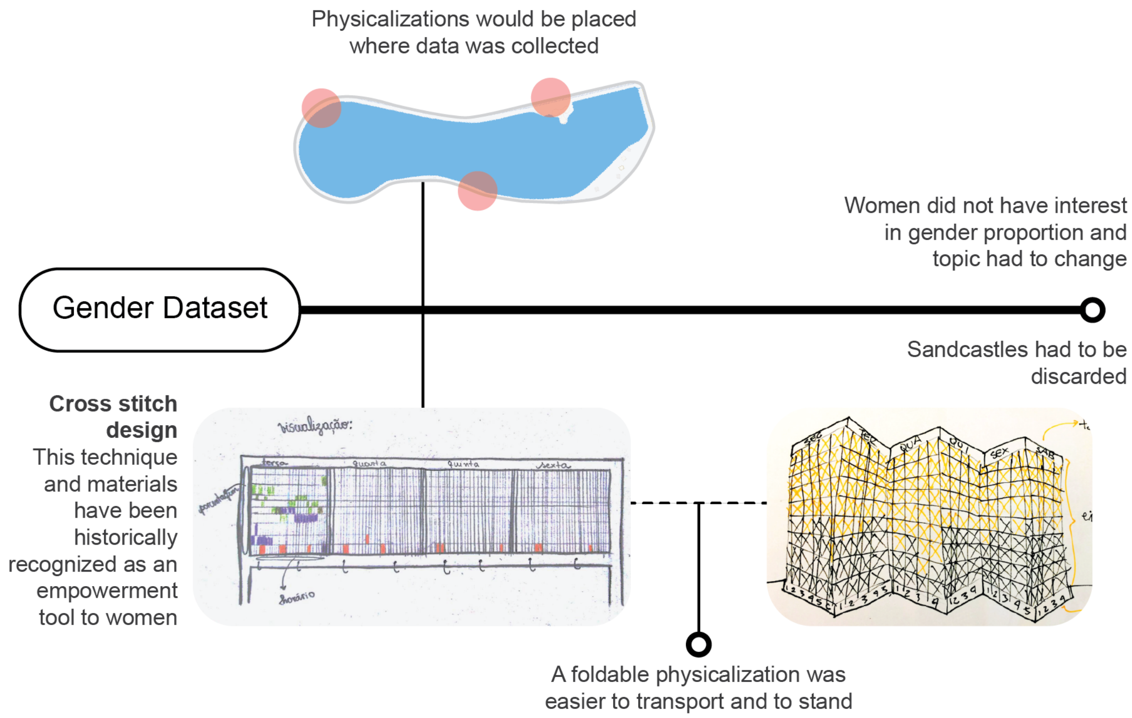


Figure 4.2: Design Process: initial stage

4.3 Challenges

This section unfolds challenges that came out during the design process of the situated and public physicalization.

4.3.1 Understanding the Audience's Interest

One step that is usually neglected while creating a visualization for public space is understanding what the audience is interested in, especially in the case of situated physicalizations, where the data is related to the place. In hindsight, our experience initiated with a similar approach: our first sketches started from a dataset made available by a partner research group (LabRua, www.labrua.org) with counts of the proportion of men and women along the day at Açude Velho.

As a first development, we conducted a focus group with six women that are used to go to Açude Velho to validate that our sketches raised relevant views on this data. Interestingly, the focus groups suggested that our audience was interested in something else. The participants in our focus group suggested that women who visit Açude Velho are not chiefly concerned with the fact that they are a minority in the place; they are more concerned with the situations of danger that they experience. The interviewees described situations of harassment that they went though in Açude Velho, which motivated us to change the topic to something more specific: sexual harassment. Figure 4.2 illustrates the initial design process and how it was affected by the focus group.

Changing the physicalization topic usually has a cascading effect. New data often needs to be collected, and previous sandcastles might be discarded because they do not fit the new data. In our case, for example, all sandcastles designed with gender proportions in mind had to be discarded since the new data corresponds to a completely different topic.

A central decision stemming from this first experience was that we would then collect data about harassment directly from those using the public space.

4.3.2 Dealing with Data Availability

Collecting data from people who visit public and uncontrolled spaces is hard for a series of reasons. This section describes the challenges we faced when we changed the physicalization topic and had to collect new data. Challenges have varied between getting people's interest to contribute, data collection constraints, and issues with the data format.

One of the main difficulties of creating a physicalization with personal data in a public space is motivating people to share their data. We initially asked women through online

Figure 4.3: Focus group to evaluate the topic relevance. Participants represent the target audience, which are women who are used to visit Açude Velho.

and offline advertisements to add their data to an online platform (the intended dataset is the *Situated Harassments*, shown in Figure 4.4). In two weeks of advertisement, no one contributed. Some women who were invited to report their experience affirmed that they did not contribute to the platform because it was cumbersome. This led the team to decide to collect women's in person at Açude Velho.

Another challenge for data collection in the context of situated and public physicalizations about people is formulating the right questions. In our case, for example, we started to approach women and raise specific questions about sexual harassment, which is a sensitive issue. That approach was not useful to start a discussion and produce relevant information because women sometimes did not feel comfortable using the term harassment to events that occurred in their lives (we were trying to collect data to fill the *structuded reports*, shown in Figure 4.4). When we changed the questions to something less explicit such as "have you ever was bothered by men here at Açude Velho?", more relevant answers started to show up (the open-ended answers are available in the *semi-structuded reports* dataset, shown in Figure 4.4).

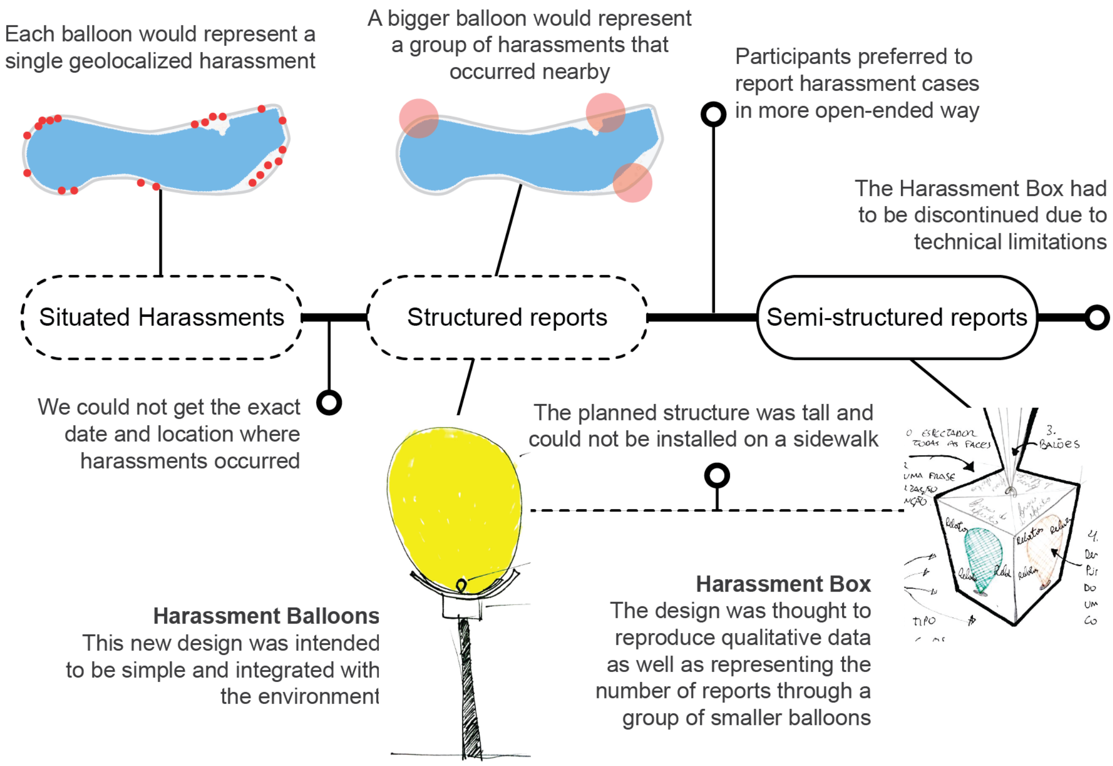


Figure 4.4: Design Process: intermediate stage

Data is not always available, and sometimes people are reluctant to share their personal information. In a physicalization project where personal information is planned to be exhibited in public space, one must think in advance of strategies to get the right information. In such scenarios, empathy with the audience and understanding the context where data lies is paramount to the data collection process.

4.3.3 Adapting to Data Changes

During the initial stages of a visualization project, data may not be wholly available or even need to be collected yet. In those circumstances, any design decision risks being discarded if the data changes drastically. That happened after we changed the physicalization's topic and had to collect new data, for instance. We designed a series of sketches representing the proportion of men and women at Açude Velho but had to drop the visualization sandcastles when we changed the topic to represent reports of sexual harassment (see some sandcastles in Figure 4.2).

In the case of situated and physical visualizations, the lack of precise information about the available data can also affect decisions regarding data materialization in space. We were initially interested in presenting the physicalization at the exact location where the harassment occurred. However, as we started to interview women, we found that they did not remember the precise date and location of the harassment. It was thus necessary to adapt the physicalization's positioning in order to deal with such uncertainty.

Since the physicalization development heavily relies on data, it is crucial to establish the vision and objectives of the physicalization as well as to understand the data before diving into the physicalization's design.

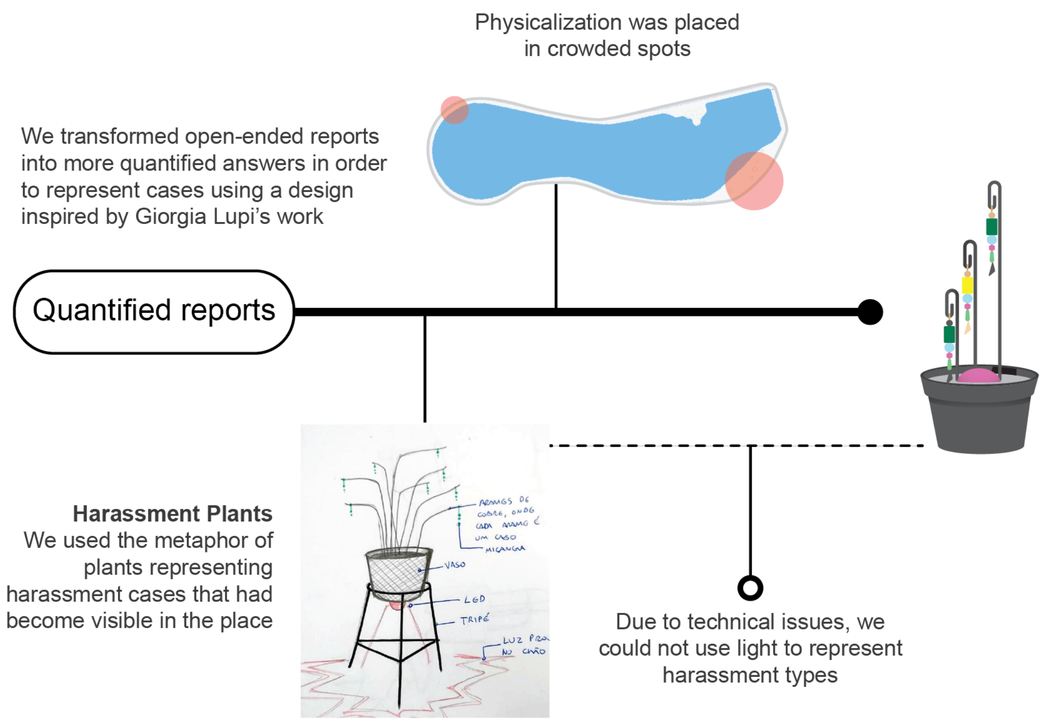


Figure 4.5: Design Process: final stage

4.3.4 Considering Environmental & Technological Constraints

Presenting a physicalization in public space requires thorough consideration regarding environmental and technological constraints. One must specify where, when, and how to exhibit the physicalization as well as what materials and technologies to use to ensure feasibility, visibility, mobility, safety, among other aspects.

Exhibiting a visualization in a public and open place adds a layer of complexity to the design process. One needs to consider issues such as weather, vandalism, theft, and transport. In our case, for example, as we planned to present the physicalization in July, which has occasional rains in Brazil, we needed to make sure all components would be waterproof. Other design decisions, such as using cheap and lightweight materials, were made to facilitate transport and mitigate issues related to safety.

Installing a physicalization in a public place also requires consideration regarding space limitations. We had to install the physicalization on a public sidewalk, which cannot be altered because it may be considered vandalism. This limited, for example, the use of tall structures that need an attachment to the ground. Any structure should also not obstruct people's way. At the same time, we needed to create something that could stand out in the crowd. Therefore, it is essential to understand space limitations and considering alternatives to overcome environmental constraints.

Another common constraint concerning physicalization development is the materials and technologies available to compose the data representation. In the context of this work, we did not possess sophisticated equipment such as 3D printers or laser cutters. We were, therefore, limited to lower-cost materials and technologies available at regular stores. This design constraint limited our space of possibilities but also allowed us to develop physicalizations that could be reproducible at a low cost.

One last issue regarding technology is staff experience. We faced several issues regarding implementation because project members did not have experience with the technologies we chose. One example was the Harassment Box's sound quality (see the sandcastle in Figure 4.4). We spent a considerable amount of time working on the box to finding out in the end that the Arduino sound modules were not loud enough to the environment. Another issue was related to the LED system in the Harassment Plants (see the sandcastle in Figure 4.5). As project members are not experts in electronics, they did not expect that by connecting LEDs together would interfere in the blinking patterns we intended to generate. These experiences highlight the need to ponder materials and technologies according to project members' expertise.

4.3.5 Fostering Audience Engagement

Physicalizations may elicit public engagement [41], but our experience emphasizes that there are issues that make this more difficult in the context of public space. This section presents challenges of creating a physicalization for a public and open space where the audience visits to leisure or to practice exercises.

In our experience of deploying the Harassment Plants, we realized that people did not usually break their routine (e.g., walking or doing exercise) to stop and look at the visualization. For that reason, we tested strategies to get people's attention, such as putting the physicalization close to places where people rest or calling them to explore the visualization, for example. Even though those strategies seem to have increased engagement, a considerable number of people did not get interested in the physicalization.

Besides the fact that people are reluctant to break their routines to explore a physicalization in a public space, another reason seems to have negatively affected engagement: people did not understand what Harassment Plants were at first sight. As the physicalization does not make a clear reference to the topic of harassment, people who passed by Açude Velho thought that those objects had a completely different purpose. Some participants who explored the visualization said that they initially thought the objects were some decoration or crafts for sale. The persons who said so only stopped to look at the visualization after being informed that the plants are data representations about sexual harassment that happened in that place. The physicalization design has failed in making people aware of the topic. Therefore, in the context of public spaces, it seems to be important, creating an easily recognizable physicalization.

4.4 Conclusion

This chapter illustrated a physicalization's development and discussed challenges that happened throughout the process. Every aesthetic provocation and detour that took place during the sandcastles' development helped designers to learn how to create a situated physicalization in a public space. The main lesson we take from this experience is that designers must get to know the context (i.e., people and place) where the data lies in order to deal with the particularities of creating a public and situated data representation.

Getting to know the persons behind the data facilitates the process of data collection. Collecting personal data from public spaces is not an easy task since the data is often unknown. Talking to people to which the data refer helps designers to understand the underlying aspects of data. In our context, only after directly contacting women, we could understand data limitations and how to approach people in order to get the correct data.

Designers that intend to create a public and situated visualization must know the place to which data belongs in order to understand the environmental limitations and how the audience uses space. In doing so, one can choose the best technologies and materials that can be used and where to install the physicalization. In our case, getting to know the space helped us to identify limitations about installing the physicalization on the sidewalk and also allowed us to discover possible spots where people stop.

This chapter contributed to increasing the understanding of the design process of data physicalizations in the context of anthropographics. It is essential to make clear that the lessons learned in this work may not be not generalizable to other contexts. Nevertheless, this work is a first attempt to understand the possible challenges in developing a situated, physical, and public anthropographic.

Chapter 5

In-the-wild Study

This chapter presents results from an in-the-wild study that compares to what extent two different visualizations affect the compassion of passers-by towards the topic of sexual harassment in a public space. Studies that follow research in the wild agenda allow researchers to explore "how a range of factors can influence user behavior *in situ*" [71] and provide a more naturalistic way of understanding how people use visualizations in practice. In-the-wild experiments are more informative because they uncover unexpected behaviors and situations that may not be emulated in a lab setting.

5.1 Study Overview

We have conducted an in-the-wild study at two public spaces: Açude Velho, which is described in Chapter 4, and Parque da Criança, which is a park that is situated close to Açude Velho. The study explored situatedness, physicality, and granularity on people's compassion through donation and self-reported scales. We aim to answer the following research questions:

1. To what extent does the experience with a situated, physical, and fine-grained anthropographic affects people's donation in comparison with seeing non-situated, virtual, and coarse-grained visualizations in a poster?

2. To what extent does the experience with a situated, physical, and fine-grained anthropographic affects people's self-reported compassion with seeing non-situated, virtual,

and coarse-grained visualizations in a poster?

Figure 5.1: Harassment Plants placed in a public space. Each vase contains reports of a different type of harassment.

5.1.1 Dataset

The dataset consists of sexual harassment reports made by women who were passing-by Açude Velho, which is a public space described in Chapter 4. We collected reports by taking notes of verbal answers to a fixed set of open questions. Such questions are related to the harassment itself, the reaction of the victim, and the characteristics of the harasser. Examples of questions are the description of the harassment, the immediate reaction of the woman, the type of harassment (e.g., a catcalling, stalking, undesired physical contact, etc.), the perceived age of the harasser, etc. All categories were defined after analyzing the open answers provided. No question was used to identify the victims, and participants authorized all collected data. Data comprises 28 harassment stories.

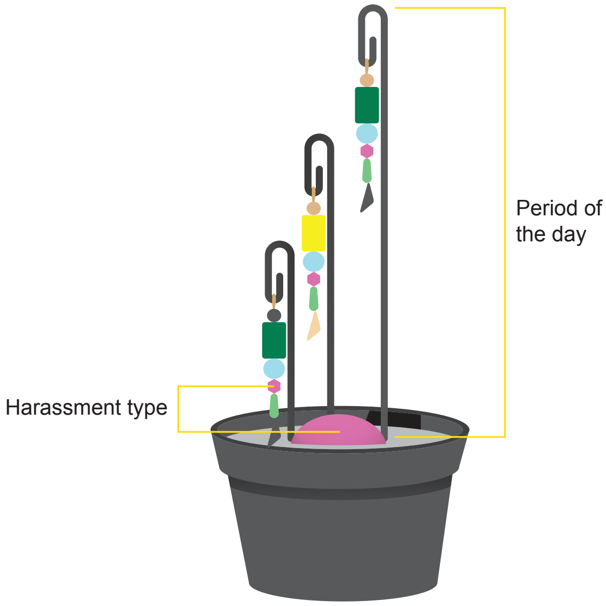


Figure 5.2: Harassment Plants' vase in detail. The rod's size represents a period of the day: the smaller is morning, the taller is night. The color in the middle of the vase represents the harassment type.

5.1.2 Visualization Designs

Two visualizations are explored in this study. The Harassment Plants (see Figures 5.1, 5.2, and 5.3) is a physicalization that represents individual stories of women who were harassed in Açude Velho. The Harassment Information (see Figures 5.4) represents the same data but in aggregate and more virtual manner. The two data representations were exhibited in public spaces of Campina Grande, Brazil.

Harassment Plants is a situated, physical, and fine-grained — maximum granularity — anthropographic. Each vase represents a different category of harassment and contains glyphs of the same type (see details in Figure 5.2). The color in the middle of the vase corresponds to the type of harassment, which is also represented by the same color in one of the beads. Each glyph is composed of a rod — which represents the period of the day the harassment occurred — and a pendant with multiple beads — which corresponds to the reaction of the victim, the perceived age of the harasser, etc. The persons interested in reading

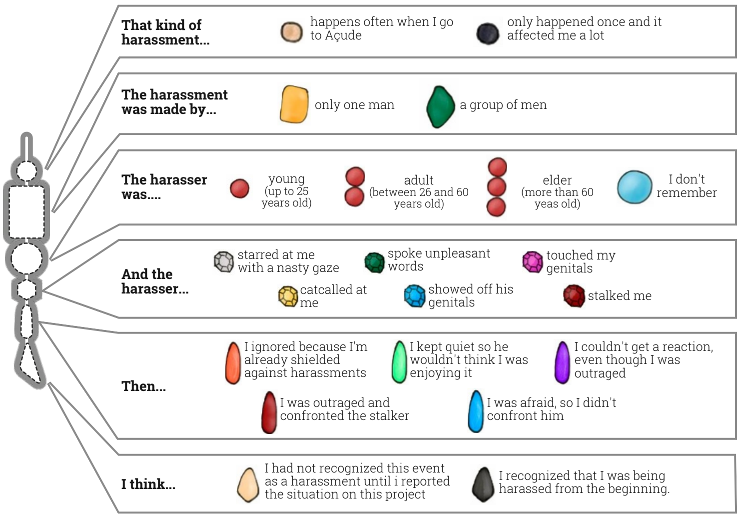


Figure 5.3: Harassment Plants' legend. It was available as a card for participants to read while experiencing the visualization.

the physicalization could understand the meaning of each symbol through a legend that was placed in front of the vases or through legends in cards that were available over the table (the legend can be seen in Figure 5.3).

Harassment Information is a non-situated and coarse-grained — low granularity — visualization. The charts represent statistics about the characteristics of harassment that occurred in Açude Velho. The visualization is presented in a conventional poster format and printed in A2 size. The visualization has the title "Harassments in Açude Velho" and a short introductory text that describes the topic and the period the data was collected. The same title and text also exist in the main legend of the Harassment Plants to promote consistency between the data representations.

5.1.3 Comparing to Boy et al.

This section compares visualization designs from Boy and colleagues [14] with visualizations from this study. Figure 5.5 presents differences regarding each design space dimension.

The Harassment Plants proposed for this study (latest column in Figure 5.5) has a higher level of granularity, physicality, and situatedness than any of Boy et al.'s visualizations as

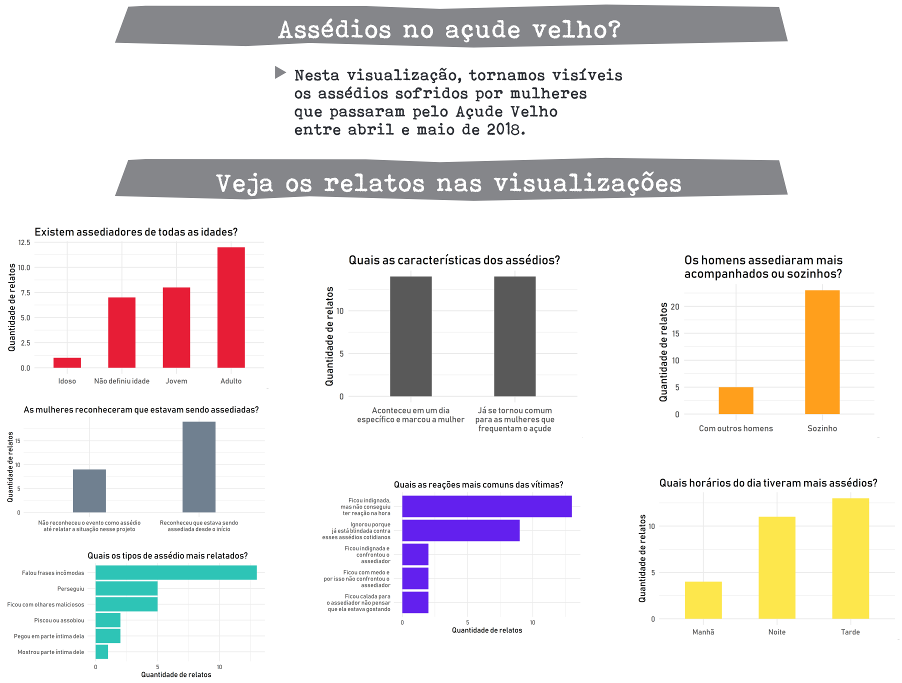


Figure 5.4: Harassment Information. Each chart represents the statistics about one of the variables collected to compose the harassment report. The top-left chart, for example, represents the perceived age of the harasser.

well as the Harassment Information. In contrast, Boy et al.'s visualizations have a higher level of realism and a lower level of authenticity than the visualizations we designed.

5.2 Method

5.2.1 Visualization Deployment

We deployed the Harassment Plants at Açude Velho and the Harassment Information at Parque da Criança. Açude Velho is a lake located at Campina Grande's downtown. Parque da Criança is the main park in Campina Grande, which is located close to Açude Velho. Harassment Plants were presented in three popular areas around Açude Velho. On the other hand, the Harassment Information was displayed where groups of people were gathered at Parque da Criança. The visualizations were presented during different days and times of the week,

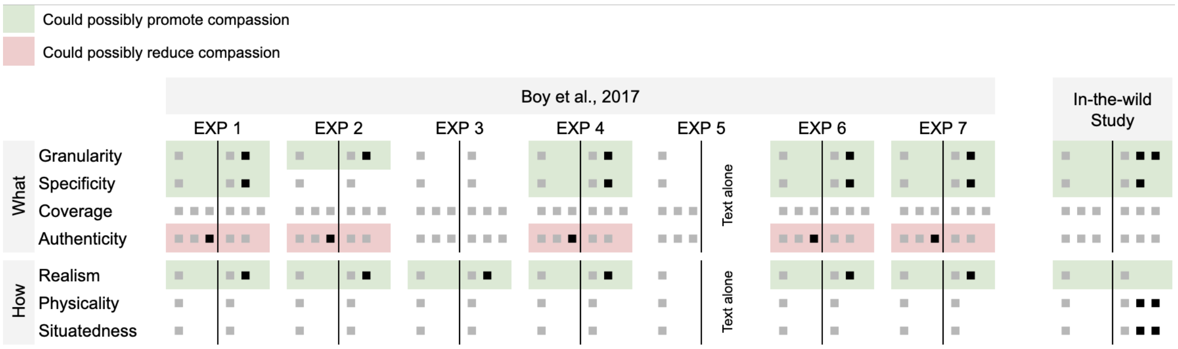


Figure 5.5: Visual summary comparing visualization designs from Boy et al. and the in-the-wild study. The black or gray squares show in which position the visualizations are in the design space according to each dimension. The black squares represent how much a visualization design has a higher level from a specific dimension in comparison to the other visualization. The columns on the left correspond to the visualization from the control condition; the columns on the right are anthropographics.

and they were carried to the places during every new exhibition.

5.2.2 Participants and Recruitment

We recruited individuals who visited one of the places the visualizations were presented. Participants were split into four conditions (see Figure 5.6). Two conditions comprise people from Açude Velho or Parque da Criança who did not see any visualization. The other two conditions contain people who saw the Harassment Plants at Açude Velho or the Harassment Information at Parque da Criança. Overall, 81 persons saw a visualization and were interviewed: 44 persons saw the Harassment Plants, and 37 viewed the Harassment Information. The other 47 persons did not see any visualization but were also interviewed: 21 were at Açude Velho and 26 in Parque da Criança. Participants were walking, having a seat, or eating close to the places the visualizations were exhibited. The researcher stated the artifact represented data about sexual harassment that happened in Açude Velho but did not disclose the research purpose.

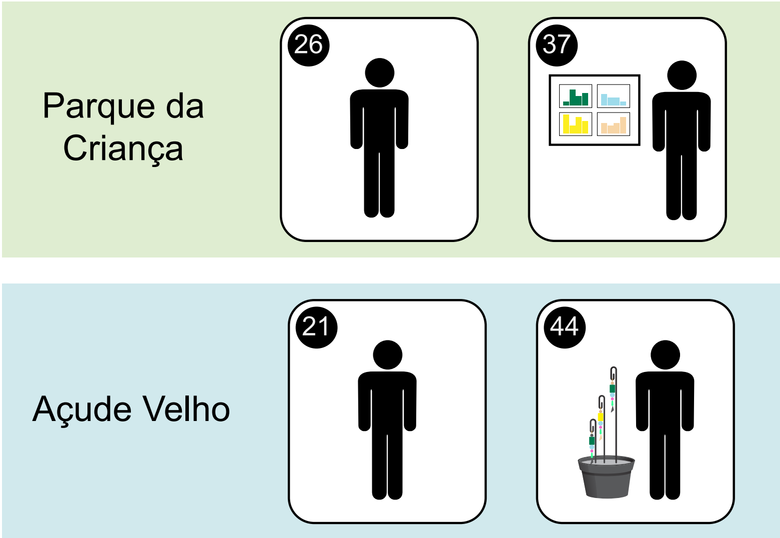


Figure 5.6: Number of participants in each condition. The first column represents participants that did not see any visualization. The top-right cell represents people who saw the Harassment Information. The bottom-right image represents people who saw the harassment plants. Participants only experienced a single experimental condition, which characterizes a between-subjects design.

5.2.3 Data Collection

After interacting with the visualizations, participants were asked to fill a printed questionnaire that contains questions about prosocial behavior (the behavioral facet of compassion), and a self-reported scale of empathic concern (the affective perspective of compassion). Participants who did not see the data representations answered a similar questionnaire but without the self-reported scale of compassion.

Two questions addressed prosocial behavior. In the first question, participants were presented to the following text (translated to English):

> **Picture the following scenario**
>
> Suppose I give you R$ 100 for participating in this survey.
>
> If you wish, I can split this money between you and a donation to an institution that will promote a campaign to combat harassment in the Açude Velho.

Participants had to answer "How much of the R$100 would you donate to the institution?". The other question related to prosocial behavior asked to justify the donation.

Finally, questions related to empathic concern are based on the ones used by Boy et al. [14]. They are 5-point Likert scales that vary between "Not at all" and "Very". They were translated to Portuguese and revised an expert in English/Portuguese translations.

5.2.4 Design Factors

We explored the interaction of three design factors: situatedness, granularity, and physicality. We consider the Harassment Plants as more anthropographic because they have a maximum level of granularity — every glyph represents the story of a single woman, intermediate/high situatedness — marks are presented close to where the data belongs to, and maximum physicality — marks are physical. On the other hand, the Harassment Information lies in a position that is far from the concept of anthropographics: it has low granularity — each mark represents a group of people aggregated by at least one attribute, low situatedness — marks are presented far from where data belongs to, and low physicality — marks are shown in a flat medium.

5.2.5 Response Variables

We aimed to investigate whether compassion — captured using self-reported scales and prosocial behavior as a proxy — would be affected by data representations about persons in another context than the humanitarian. Compassion is measured through two variables. The first is focused on prosocial behavior: participants were asked how much of a fictitious amount of R$100 that they would donate to an institution that fights sexual harassment at Açude Velho. The value answered by each participant is a measure of their compassion. The second variable is related to self-reported compassion: our questionnaire contained a series of 5-point Likert scale questions in which they should answer "how much (. . .) they were feeling about the data they just saw", where the (. . .) corresponds to feeling *sympathetic*, *compassionate*, or *moved*. Self-reported compassion is measured as the median of a participant's answer to these three questions.

5.2.6 Analysis

We use Cliff's delta [18] for statistical inference to compare the data in the different groups from the first experiment. We report Cliff's Delta effect sizes d, with both point and interval estimates. For interval estimates, we consider a 95% confidence level. Consider a comparison of donation values g and h from two groups, and let $P(g > h)$ be the probability that a randomly sampled donation from the first group is higher than the latter. In this situation, a Cliff's delta $d = 0.4$ means that $P(g > h) - P(h < g) = 0.4$. Cliff's delta was chosen because it is a robust method for data that is not normally distributed or for ordinal data such as Likert scales [58].

Some of the terms that compose the scale of empathic concern were removed from the analysis because several participants reported they could not understand what they mean, and this led to clearly inconsistent answers. During data collection, a considerable number of participants reported they could not understand the meaning of the terms *tender*, *warm*, and *softhearted* — translated to Portuguese as *carinhoso*, *caloroso*, and *generoso*. This happened even though these terms had their translation revised by three different persons, including a professional translator. The empathic concern scale is thus left with 3 out of 6 original elements.

5.3 Results

This section presents to what extent the introduction of data representations in a public space affects the compassion of passers-by and how large is the difference in the compassion of participants who interacted with the different visualizations.

5.3.1 Prosocial Behavior

We first compare donations of participants who interacted with one of the visualizations and participants that saw no visualization. Evidence ($d = 0.28$; [0.006, 0.52] 95%CI) suggests that people who saw the Harassment Information might have donated more than those who did not see any visualization at Parque da Criança. On the other hand, the same effect is unlikely for the Harassment Plants ($d = 0.088$; [-0.15, 0.32] 95%CI).

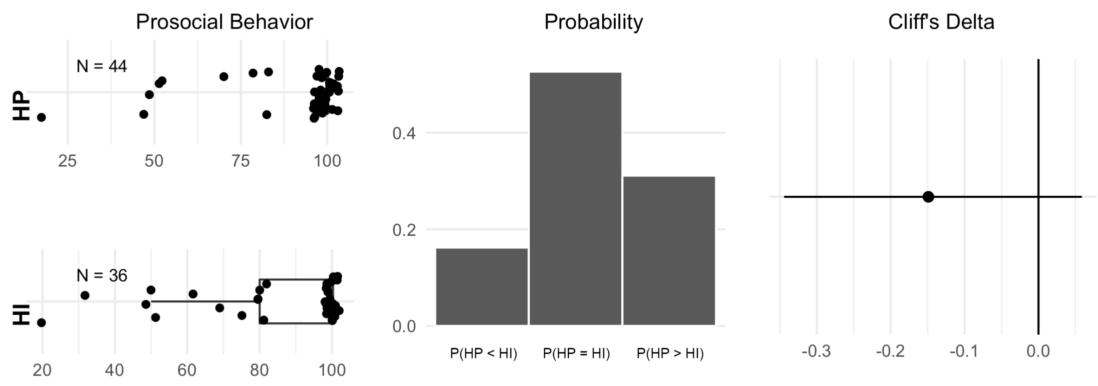


Figure 5.7: Prosocial behavior: Harassment Plants (HP) vs. Harassment Information (HI). Left: responses to "How much of the R$100 would you donate to the institution?". Right: probabilities of randomly choosing a member that scores higher, equal, and lower than the member of the other group; and Cliff's delta. Error bars are 95% CIs.

We also compare donations of the group that interacted with the Harassment Plants and the persons who experienced the Harassment Information. The probability of donating more after seeing the Harassment Plant is likely to be higher than when participants see the Harassment Information ($d = -0.15$; [-0.34, 0.06] 95% CI). However, findings deserve further scrutinization since the evidence we have found is very weak. Nevertheless, a possible effect might have been attenuated since the distribution of donations (see Figure 5.7) suggests that the majority of people chose to donate values close to R$100, which may have caused a ceiling effect in the results.

5.3.2 Self-reported Compassion

We have also investigated whether capturing the compassion of participants in a more direct manner would produce similar results to the ones of prosocial behavior, since the literature strongly supports the relation between the two variables.

We contrast the self-reported compassion between participants who saw the Harassment Plants and the persons who interacted with the Harassment Information (Figure 5.8). Results point to weak evidence in favor of the Harassment Plants ($d = -0.22$; [-0.43, 0.01] 95% CI), though the effect size is unclear. Therefore, more fine-tuned experiments need to be

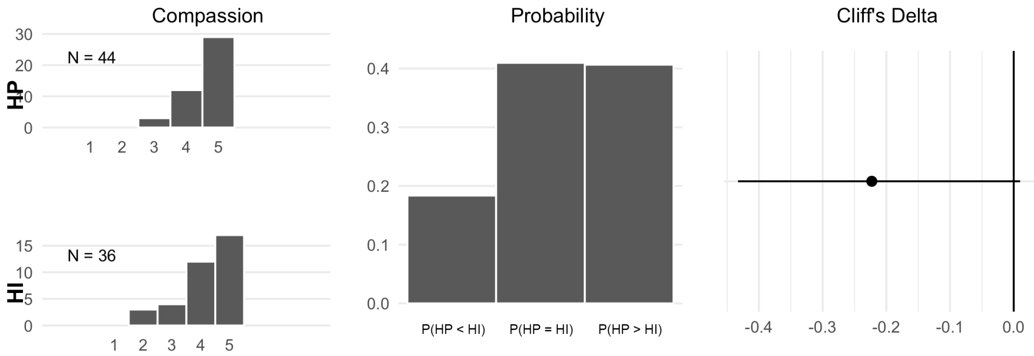


Figure 5.8: Compassion: Harassment Plants (HP) vs. Harassment Information (HI). Left: self-reported compassion calculated by the median of 3 Likert scale questions. Right: probabilities of randomly choosing a member that scores higher, equal, and lower than the member of the other group; and Cliff's delta. Error bars are 95% CIs.

conducted to determine whether and to what extent the anthropographic design might affect people's compassion.

5.4 Conclusion

This chapter presents results from an in-the-wild study that explored the effect of data representations about sexual harassment on compassion through prosocial behavior and self-reported measurements. Findings suggest very weak evidence in favor of the anthropographic design. However, results are mostly inconclusive, and if an effect exists, it is likely to be small. Therefore, more controlled studies are necessary to understand the effect of anthropographics on compassion.

We have some validity threats that might have affected the study. The main threat to external validity is our sample size. It is also crucial to consider that this work was conducted in public spaces, where many contextual variables actuate without our control. Finally, the data contains signs of ceiling effects, which may make it more difficult to draw conclusions from the data.

Despite the limitations, this work has relevant contributions that spark future directions. Future work should explore more reliable ways to capture compassion to avoid ceiling effect, conduct more controlled experiments to isolate possible confounding variables, and qualitatively investigate the reasons why participants demonstrated different behaviors in the presence of the visualizations.

Chapter 6

Charitable Giving Experiments

This chapter presents results from two experimental studies that investigated the role of a specific anthropographic design on compassion in the context of charitable giving. We decided to conduct more controlled experiments after finding inconclusive results from the in-the-wild study presented in Chapter 5. Also, as we had already devised the design space, the experiments were based on the exact positions of each dimension.

6.1 Studies Overview

Both of the studies presented in this chapter were crowdsourced using the Prolific[1] platform. The first experiment is our first tentative to gauge the effect of anthropographics on prosocial behavior through donation questions, building on our experiences in Chapter 5 and Boy's work [14]. The second study is a refined version of the first one: it contains a reformulated donation question, it explores whether data anonymity interferes with results, it has greater statistical power, and it also explores to what extent the anthropographic influences affect. Sections 6.2 and 6.3 describe the experiments and their results.

6.1.1 Dataset

Our main artifact in data collection is a set of visualizations – anthropographic and conventional – showing data about migrants. We chose the topic of migration because it is related

[1] Prolific website: https://www.prolific.co

to charitable giving projects such as UNHCR[2] or IOM[3], which assures ecological validity. Our dataset comes from the Missing Migrants Project[4] and contains data about migrants who tried to cross a border in 2018. The visualizations represent the number of persons who died and the ones who survived while trying to cross a border. Other variables, such as whether the person is a man, a woman or a child, the incident's location, the cause of death, and the estimated date of the incident are also represented in the anthropographic visualizations.

As we initially intended to make participants see two visualizations in the experiment, we selected from the dataset one region for each visualization. We chose data from the Middle East and Southeast Asia because they have a similar distribution of incidents. Our rationale for choosing regions with a similar number of dead and survivors is that, otherwise, participants would tend to be more compassionate with the visualization with a more significant number of deaths.

Figure 6.1: *Bar chart from experiment 1.* The red bar representes the number of migrants who died in 2018 trying to cross a border in the Middle East. The lighter bar represents the persons who survived to the incidents.

[2]United Nations High Commissioner for Refugees: www.unhcr.org
[3]International Organization for Migration: www.iom.int
[4]Missing Migrants website: https://missingmigrants.iom.int

6.1.2 Visualization Designs

The studies from this chapter explore charts with different positions in the design space of anthropographics compared to the visualizations created by Boy and colleagues [14]. The experiments show two different visualizations: an anthropographic and a bar chart. The second experiment also showed the same visualizations but without information about the incident's location in the anthropographic. The next sections detail the visualization designs.

Experiment 1: Anthropographic vs. Bar Chart

Experiment 1 uses two visualizations. The first visualization is an anthropographic designed to show more information (we thus sometimes call it an information-rich visualization) with the intent to connect participants to victims of the incidents. The second visualization is a bar chart, which contains less information about the persons being represented while representing the same core data from the dataset. We manipulated three design space[5] dimensions: granularity, specificity, and realism. We aimed to create visualizations that lie in opposite positions of the design space.

The anthropographic presented in Figure 6.2 represents real migrants that died or have survived to try to cross a border in the Middle East. Each human silhouette corresponds to a single victim. The gender and the approximate age are also indicated by the characteristics of the mark. The visualization was designed to have maximum granularity — each person corresponds to a single mark; an intermediate level specificity — groups of people can be distinguished from each other thanks to a combination of information such as gender, location, date of incident, etc.; and an intermediate degree of realism — marks are human silhouettes.

The bar chart from Figure 6.1 also represents victims from incidents in Middle East borders, but it shows less information. The bars represent the number of people that were involved in the incidents. The color encodes whether the persons died or survived. We intentionally did not represent information about people's gender, cause of death, among other factors to decrease the level of specificity in the chart. Differently from the anthropographic, the bar chart has low levels of granularity, specificity, and realism in order to hinder people

[5] The dimensions are detailed in Chapter 3

from distinguishing one victim from the other and, consequently, avoid participants to relate to individual persons.

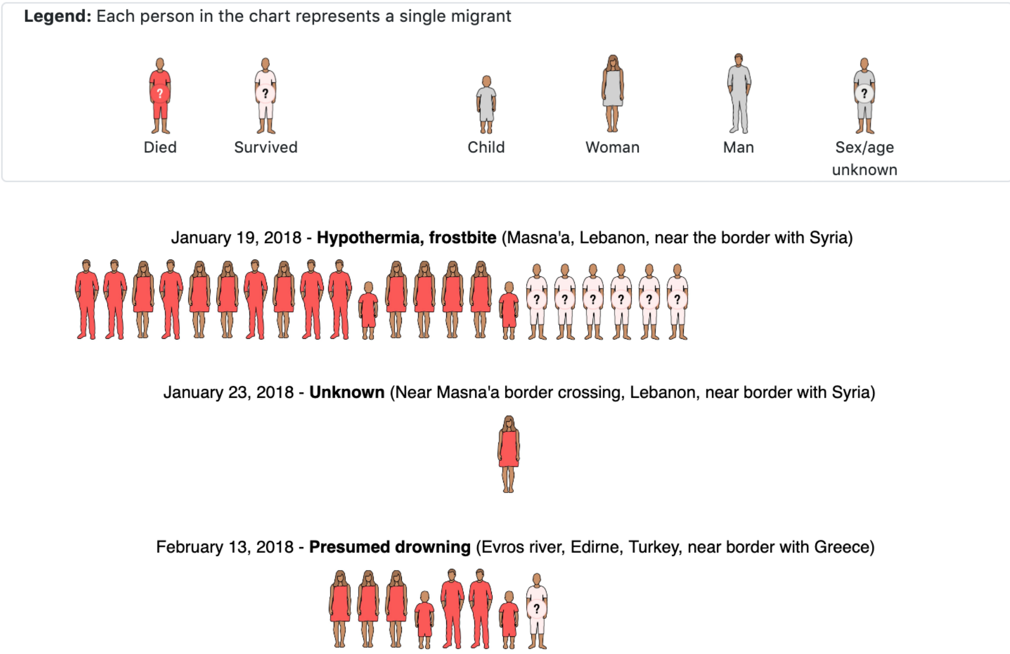


Figure 6.2: *Part of anthropographic from experiment 1*. Participants could scroll the page to see victims of incidents up to December 2018. Each human silhouette represents a real victim that tried to cross a border in the Middle East. The information-rich visualization shows additional information of gender, age (i.e., adult or child), estimated date of the incident, cause of death, and place where the incident occurred.

Experiment 2: Anonymized Anthropographic

Feedback from the first experiment suggests that the location where the incidents happened might have affected participants' responses. For that reason, we decided to remove from visualizations the information about incidents' location for experiment 2. As the bar charts did not have such information already, they are kept the same. On the other hand, we removed the text in parentheses from the anthropographic (e.g., removing "(Masna'a, Lebanon, near the border with Syria)") that described where incidents occurred. Besides the information mentioned above, the remaining visualization designs are identical.

6.1.3 Comparing to Other Studies

This section compares visualization designs from Boy and colleagues [14] with visualizations from the study shown in Chapter 5 and the ones from this chapter. Figure 6.3 presents differences regarding each design space dimension.

The main difference between Boy et al.'s work, and ours is that they designed anthropographics with intermediate or low granularity and partial authenticity. At the same time, we investigated visualizations with maximum granularity and full authenticity. The design choice of testing visualizations with maximum granularity was based on Boy and colleagues' [14] advice:

> *In contexts where disaggregated data may be more readily usable, lower data-granularity may contribute to further reducing the semantic incongruence, which may positively impact viewers' perception of the uniqueness of units. Similarly, increasing the number of units to show absolute values (instead of normalized values) may provide more concrete scales [19], i.e., scales that are easier to relate to, which may also help viewers better grasp the magnitude of certain HR tragedies*

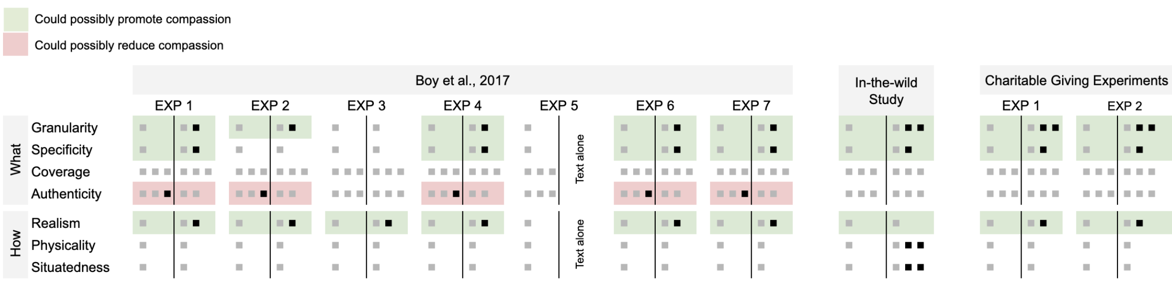


Figure 6.3: Visual summary comparing visualization designs from Boy et al., the in-the-wild study, and the charitable giving studies. The black or gray squares show in which position the visualizations are in the design space according to each dimension. The black squares represent how much a visualization design has a higher level from a specific dimension in comparison to the other visualization. The columns on the left correspond to the visualization from the control condition; the columns on the right are anthropographics.

Boy and colleagues have used synthetic attributes (e.g., fictional names, gender, and age of people) to increase specificity. However, such design decision made their visualizations less authentic and possibly have caused what they call "semantic incongruence" in the quote above. In our case, we decided to maintain a fully authentic design and increased the specificity only by representing genuine attributes — attributes originated from data.

Levels of realism, physicality, and situatedness are the same between Boy et al.'s visualizations and the ones from our last experiment. On the other hand, we have explored different levels of such dimensions in the study from Chapter 5. As was discussed in the previous chapter, we could not find strong evidence that changing those dimensions may significantly affect compassion.

6.1.4 Ethics

Experiments were approved by the Comité d'Ethique pour la Recherche de l'Université Paris-Saclay, reference CER-Paris-Saclay-2019-006. Participants were compensated if they completed the whole experiment without reloading the page or failing any attention check. All participants agreed to participate by clicking on a button labeled "I agree" after reading the informed consent.

6.1.5 Reproducibility

Both experiments were preregistered on OSF. The code, analysis, and registration description can be accessed at `https://osf.io/xqae2/`.

6.2 Experiment 1: Is there a clear effect?

This study was designed to overcome some issues pointed out by Boy and colleagues [14] in their work. As we have mentioned in Section 6.1.3, we chose to explore different visualization designs concerning granularity and authenticity compared to previous work to make sure that those modifications are relevant in affecting compassion. We also created different donation allocation questions with the intent of having effects with lower variability. Finally, we opted for conducting a hybrid experimental design (instead of a simple within-subjects design) since Boy and colleagues pointed out that their design might have been affected by a carryover effect [14] and a between-subjects design could reduce possible confounds.

This study aimed to handle issues found in previous work in order to find a clear effect of anthropographics on compassion. For achieving such an aim, we compared two visualizations that lie in opposite directions regarding granularity, specificity, and realism, which are dimensions that compose the concept of information-richness. Our research question was the following:

> Does an information-rich anthropomorphic visualization of humanitarian data increase donation allocations compared to a simple bar chart?

6.2.1 Method

Participants

One hundred twenty-six workers from Prolific platform participated in the experiment (63 per condition). We only accepted workers with at least 95% acceptance rate on the platform, fluent in English, and that did not participate in any pilot study. Participants are mostly from Europe or the United States, aged 30 on average (age varied from 18 to 68 years), and male (66%). The sample size assures a .8 power to detect a medium Cohen d's effect size of .5, as computed by the G*Power software. It is also more than twice the sample size of a Boy et al.'s [14] study.

Procedure

Participants saw two humanitarian scenarios in the same order: migration incidents in South-East Asia, then in the Middle East. During each scenario, participants were presented to a visualization that represents data about the corresponding region. One of the scenarios is conveyed with an information-rich visualization (anthropographic), while the other shows a simple bar chart. The order that visualizations are exhibited depends on the condition that participants are assigned to.

Participants were assigned to one of two conditions (see conditions in Figure 6.4). In the rich-first condition, the anthropographic (see Figure 6.2) was presented before the bar chart (see Figure 6.1), while in the poor-first condition occurred the opposite. After seeing each visualization, participants were asked to split funds from an organization between the region that was mentioned in the scenario and the rest of the world (see the fund allocation question in Figure 6.5). At the end, participants were also asked to allocate $100 between two organizations that would help migrants in Southeast Asia and in the Middle East. We asked participants to justify their donations.

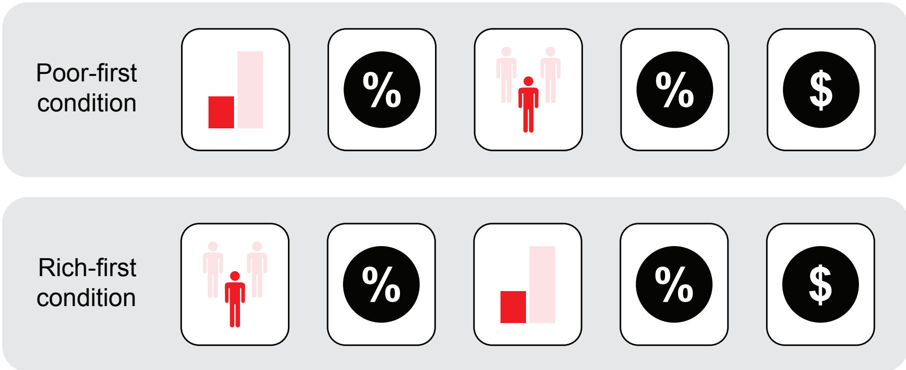


Figure 6.4: Experiment 1 conditions. Visualizations from the first scenario represent data about migrants in Southeast Asia, while the ones in the second scenario show incidents in the Middle East. The % symbols represent fund allocation questions, while the $ signs correspond to a donation allocation question. Questions are the same in both conditions.

Response Variables

We captured prosociality indirectly through donation allocation questions (e.g., splitting money between two charities), as these are less likely to suffer from social desirability bias than donation intention questions (e.g., how much participants report they would be willing to donate to a single charity) [29].

Donation allocation was measured through three different dependent variables:

- DV1 = Additional funds allocated to Southeast Asia / UNHCR compared to the Middle East / IOM
- DV2 = Percentage of global UNHCR funds allocated to Southeast Asia
- DV3 = Percentage of personal money donated to SouthEast Asia / SEARAC

For each of these dependent variables, we estimate the difference on average between the experimental group who sees the information-rich visualization first and the experimental group who sees the information-rich visualization second, yielding three different measures of (simple) effect sizes. Our primary outcome is the difference in DV1. Differences in DV2 and DV3 are treated as secondary outcomes.

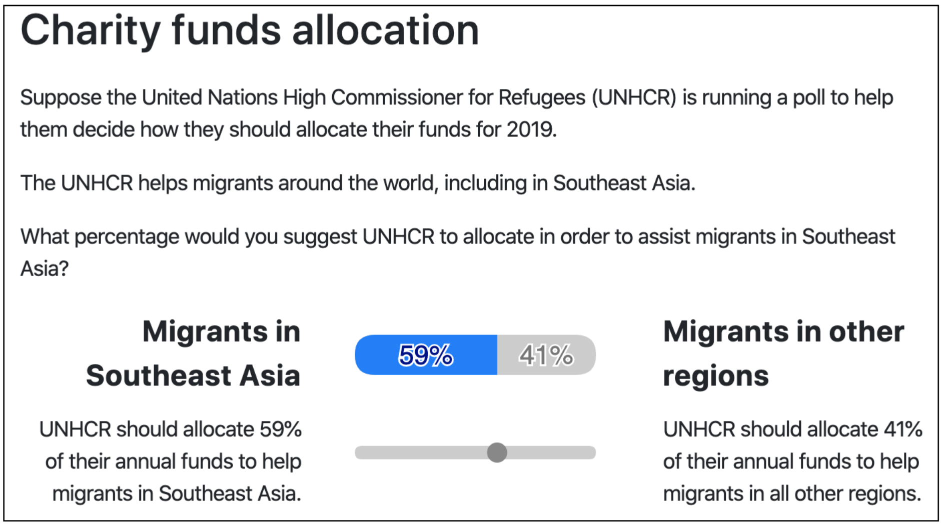


Figure 6.5: Fund allocation question

Analysis

We use interval estimation to report and interpret results. We draw inferences from the graphically-reported point estimates and interval estimates [20]. Since we identify a single primary outcome in this analysis, we do not adjust for multiplicity [6]. We interpret results for the secondary outcomes as tentative and exploratory.

6.2.2 Results and Discussion

The primary effect is the difference in mean DV1 between the rich-first and the poor-first groups. In other words, it is the interaction between the scenario condition (Southeast Asia vs. Middle East, within-subjects) and the visualization order condition (richFirst vs. poorFirst, between-subjects). Results from the primary measurement are inconclusive (3.8, 95% CI [-1.7, 9.3]), since there is no evidence that the information-rich design does increase donations in comparison to the bar chart. At the same time, if there is an effect, it is not plausible that it is of more than 10% difference in donation allocation.

The secondary effects were captured in order to reinforce results from the primary effect. The effects are the difference in mean DV2 and DV3 between rich-first and poor-first groups, respectively. Both effect 2 (2.7, 95% CI [-2.3, 7.9]) and effect 3 (-1.8, 95% CI [-7.1, 3.1]) are also inconclusive. We decided to conduct further analysis to explore the responses in order to understand such results.

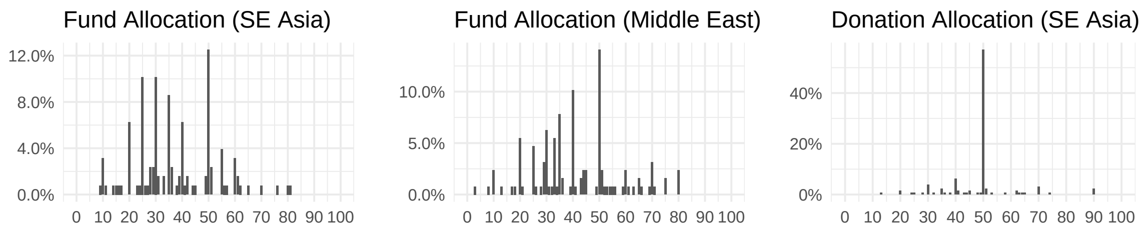


Figure 6.6: Distribution of three donation questions

Looking into participants' responses, we realized that there are patterns of donations (see Figure 6.6). In the first two fund allocation questions, where participants should decide the percentage of funds they would allocate to the scenario region (Southeast Asia or the Middle East, depending on the question) or the rest of the world, they tended to distribute the funds as 50/50 or 30/70. In the last question, where participants had to allocate $100 between only two regions (Southeast Asia vs. the Middle East), it is clear that they tried to split the money evenly. The pattern we have found in participants' donation seems to reflect the diversification bias [68], where people prefer to spread limited resources evenly across a set of possibilities. Therefore, responses might have been affected by this confound.

Another issue that we found when we analyzed participants' donation justifications is that their prior attitudes might also have affected results. One participant that donated $80 to help Southeast Asia migrants, for example, might have been affected because he lived in the region: *"I lived in Thailand for four years. The migrant crisis in that part of the world is important to me as I witness it a few times"*. Another participant has more preference towards an ethnic group: *"I feel more sympathy towards Asian migrants"*. A third participant donated based on her previous donation history: *"I took into consideration my personal experiences in terms of donation to companies/associations helping migrants from specific regions of the globe"*. Therefore, perhaps showing geographic regions (and ethnic groups) in donation descriptions may affect results.

6.3 Experiment 2: Refined experimental design

This study was conceived as an attempt to improve issues from the first experiment. For that, we removed the fund allocation questions to make the experiment shorter, refined the donation allocation question to avoid the diversification bias, explored whether showing region names interferes in results, increased the sample size considerably in order to be able to detect small effects, and explored the effect of anthropographics in participant's affective response.

The main research questions are:

1. To what extent does an information-rich visualization design affect donation allocations compared to an information-poor design?
2. Does this effect depend on whether regions are anonymized?
3. To what extent does an information-rich design influence reported affect compared to an information-poor design?

6.3.1 Pre-study: Topic Selection

Before conducting the second experiment, we ran a pre-study to select a topic to attenuate the effect of the diversification bias we have found in the donation allocation responses of experiment 1. We aimed to choose a topic for cause A — the baseline — that did not make a data bump in 50/50 donations.

We accepted 154 participants located mostly in Europe or in the United States who were required to have a 95% acceptance rate on Prolific and be fluent in English. Participants were randomly assigned to one of four conditions, which correspond to different causes.

We chose four causes to compete with the cause of helping migrants in Southeast Asia: helping people after earthquakes (48 participants), saving forests in the Amazon (31), removing plastic from the oceans (39), and fighting the zika virus (36). Each cause corresponded to a condition where participants had to read a text about two causes — baseline and migration — and allocate $100 between them.

The condition that had the least expressive bump in 50/50 donations was the cause of saving Amazon forests (13% of donations). The remaining causes had 18% (plastic), 22%

(zika), and 27% (earthquake) of 50/50 donations. Therefore, we chose the topic of forests to represent cause A and the migration incidents in Southeast Asia to be cause B.

6.3.2 Method

Participants

Seven hundred eighty-eight workers from the Prolific platform participated in the experiment (197 per condition). We only accepted workers with at least 95% acceptance rate on the platform, fluent in English, and that did not participate in the previous study. Participants are mostly from Europe or the United States and aged 33 on average (age varied from 18 to 74 years). Gender was balanced (51% of males). The sample size assures a .8 power to detect a *small* Cohen d's effect size of .2 for our first research question (where the four conditions are collapsed into 2), as computed by the G*Power software for differences between independent means

Procedure

Every participant sees information about two causes. Cause A — the baseline — contains information about the Amazon forest, and cause B has information about migrants in Southeast Asia. Cause A is described with text only, while cause B is accompanied by a visualization which varies depending on the condition (information-rich vs. information-poor design). Another independent variable determines whether the geographical regions are displayed or anonymized in the descriptions of Cause A and Cause B (anonymized vs. named). The two independent variables are fully crossed, leading to four conditions in total (see Figure 6.7):

Anonymized+Rich condition : the Amazon region and Southeast Asia are anonymized and called Region Blue and Region Orange, respectively. The visualization shown for cause B has an information-rich anthropographic design.

Anonymized+Poor condition : the Amazon region and Southeast Asia are anonymized and called Region Blue and Region Orange, respectively. The visualization shown for cause B has an information-poor design.

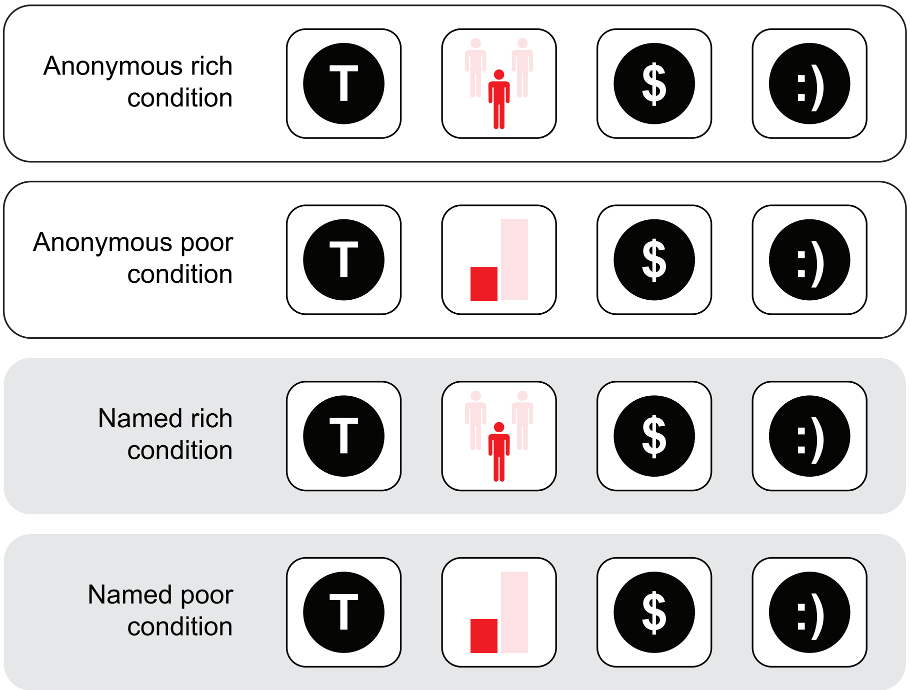


Figure 6.7: Experiment 2 conditions. The **T** corresponds to a text description of two causes: (A) saving a forest and (B) helping migrants. In anonymized conditions, the causes are from Blue and Orange regions, respectively. In named conditions, we make clear that forests are in the Amazon and migrants in Southeast Asia. The **$** signs correspond to a donation allocation question, which has anonymized regions in the corresponding conditions. The **:)** represents the affect question.

Named+Rich condition : the Amazon and Southeast Asia regions are not anonymized. The visualization shown for cause B has an information-rich anthropographic design.

Named+Poor condition : the Amazon and Southeast Asia regions are not anonymized. The visualization shown for cause B has an information-poor design.

After seeing the visualization, participants are asked to allocate a $100 donation between the two causes (see the donation allocation question in Figure 6.8). Finally, Prolific workers needed to choose their levels of valence and arousal while seeing the visualization using the Affective Slider [9].

Response Variables

The study has three dependent variables:

- DV1 (primary) = mean donation allocation to cause B (migrants in need helped by UNHCR) (range: [0, 100])
- DV2v (secondary) = self-reported valence (range: [0, 1])
- DV2a (secondary) = self-reported arousal (range: [0, 1])

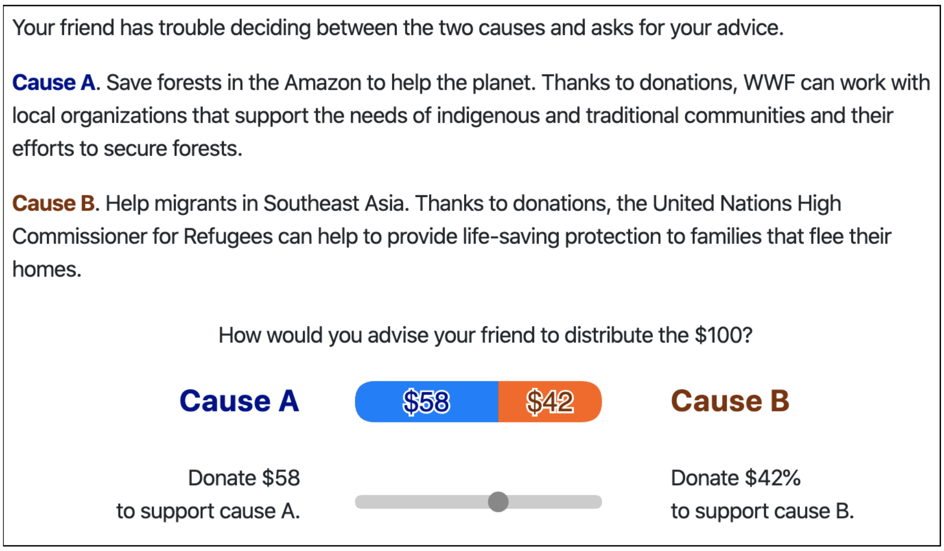


Figure 6.8: Refined donation allocation question

Analysis

Similar to the first experiment, we also use confidence interval estimation to report and interpret our results in this experiment. We interpret results for the other outcomes as tentative and exploratory, especially concerning the auxiliary outcomes. The role of the secondary outcomes is to investigate whether the information-rich anthropographic design may also influence people's affect.

6.3.3 Results and Discussion

The primary outcome is the difference between DV1 in information-rich conditions and information-poor conditions, independent whether scenarios are anonymous or named (see

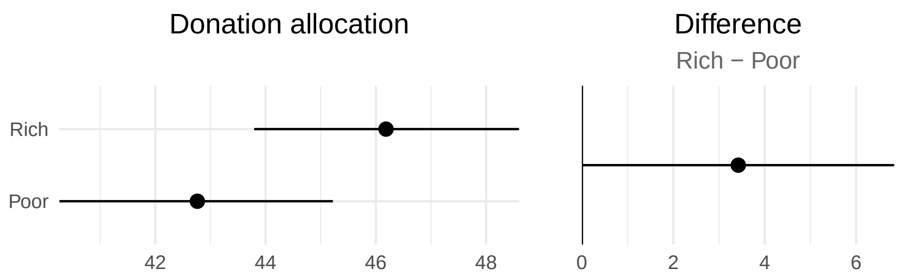


Figure 6.9: Estimated mean difference in donation allocation between information-rich and information-poor designs with 95% CI.

Figure 6.9). We captured this effect in order to answer the question of to what extent an information-rich anthropographic design affects donation allocation compared to an information-poor design. There seems to be a small effect in favor of the information-rich design: 3.4, 95% CI [0.021, 6.8]. This means that participants who saw the anthropographic might have donated a little more than the persons who viewed the bar chart – it is plausible that this difference is anywhere between 0.02 and 6.8% in the population from which our sample is drawn.

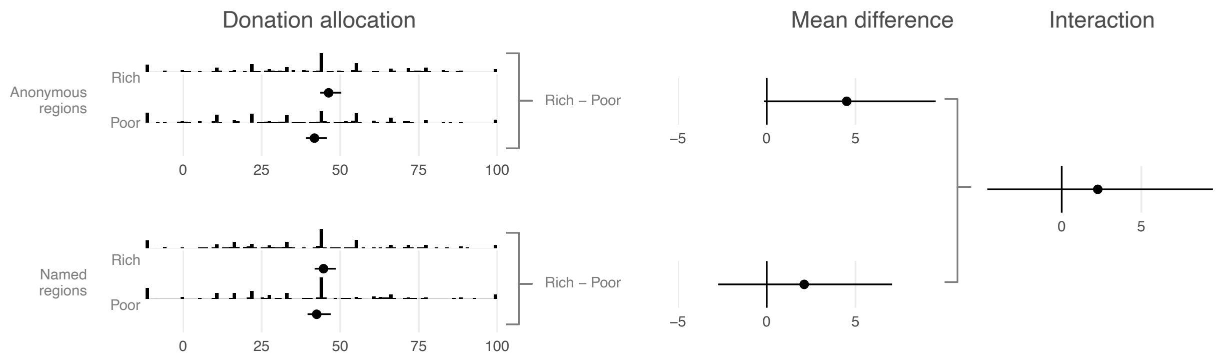


Figure 6.10: Estimated mean difference in donation allocation between information-rich and information-poor designs in anonymous and named conditions and its interaction, with 95% CI.

We also investigated whether the effect depends on making the regions anonymized in the experiment. Although it seems that anonymizing the scenarios might have contributed to a greater difference between information-rich and information-poor visualizations (4.6, 95% CI [-0.16, 9.5]) in comparison to the named scenarios (2.2, 95% CI [-2.7, 7.1]), results are inconclusive. Our data does not support the claim that anonymizing regions to where the money goes increases or decreases donations.

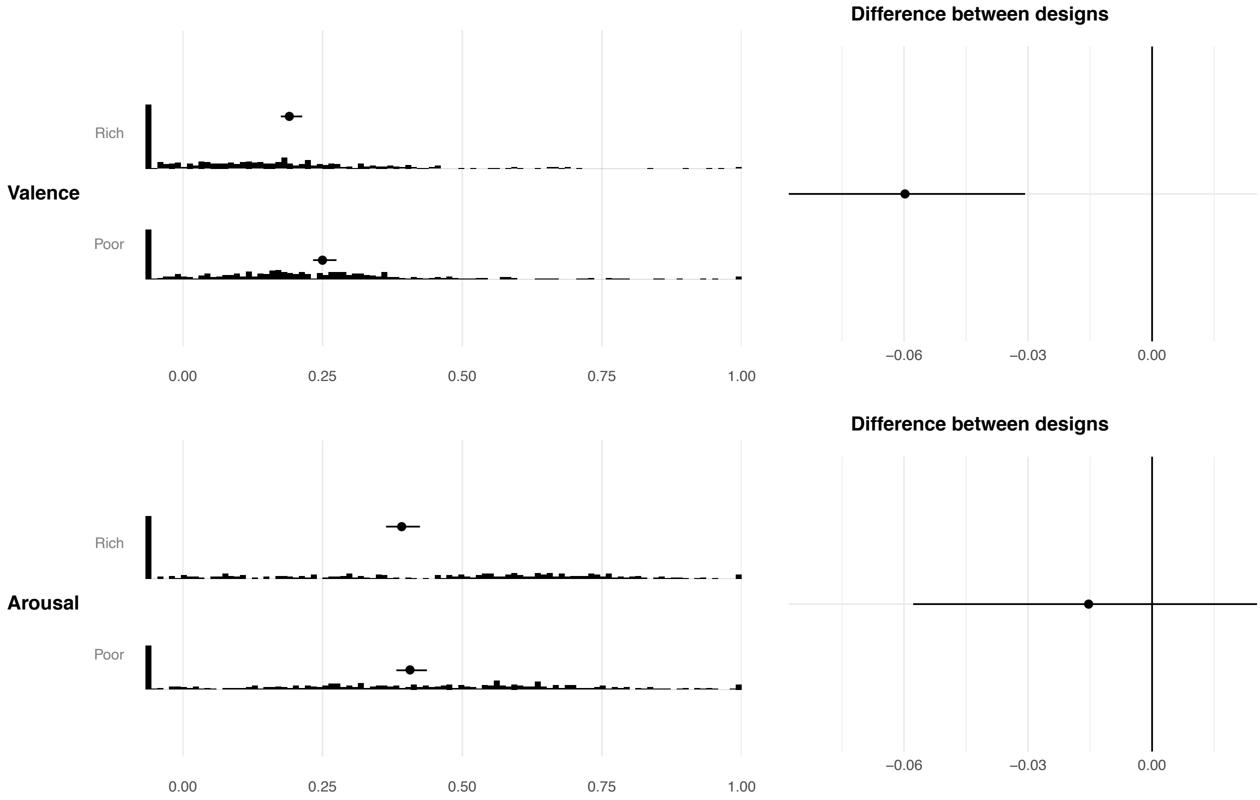


Figure 6.11: Estimated mean difference in valence and arousal between information-rich and information-poor designs with 95% CI.

Finally, we investigated to what extent information-rich design influences reported affect compared to an information-poor design. Affect is composed of valence and arousal scales, which may vary from negative (less than 0.5) to positive (greater than 0.5). According to our data, the evidence that the anthropographic affected arousal is inconclusive (-0.015, 95% CI [-0.058, 0.025]). On the other hand, we found a clear difference in levels of valence (-0.059, 95% CI [-0.088, -0.031]) between visualization designs. Results suggest that participants who saw the anthropographic reported more negative feelings but with a probably similar intensity than the ones who experienced the bar chart.

6.4 Conclusion

This chapter presents results from crowdsourced studies that investigated the role of anthropographics on prosocial behavior and affect, as proxies for compassion. We found a small effect on donations in favor of the information-rich design. Data also suggests that participants felt more negative feelings when they experienced the anthropographic condition. The findings open new research opportunities for exploring different facets of compassion.

The work has some limitations. First, we only used a single anthropographic design in the experiments, which limits results to that region in the design space. Second, we did not measure prosocial behavior using real donation, which might have affected results because of the windfall effect [17]. Therefore, new studies need to be conducted to explore different spots in the anthropographics design space and capture compassion with more accurate measurements.

Chapter 7

Conclusion

This thesis has contributed to advance the field of Anthropographics from different perspectives. From a theoretical view, we have developed concepts and terms to facilitate the design, critique, comparison, and empirical evaluation of anthropographics as well as extending their design space. From an empirical perspective, we have made progress in understanding the role of anthropographics on compassion. Finally, from a design-oriented angle, we have discussed the challenges and lessons learned during the development of a situated and physical anthropographic. This chapter summarises the results found throughout the thesis and points out directions towards future work.

7.1 Major findings

We have investigated the role of anthropographics on compassion through in-the-wild and crowdsource studies. As a consequence, we have advanced the understanding of to what extent visualizations designed with the intent to make people compassionate with the persons represented may affect different facets of compassion.

7.1.1 Prosocial Behavior

We have tried to capture prosocial behavior through different donation tasks. The study from Chapter 5 brings very weak evidence that showing a situated, physical, and fine-grained anthropographic in public space instead of a non-situated, virtual, and coarse-grained bar chart might affect how much people donate to a cause. In Chapter 6, we have found stronger evidence suggesting a small effect in favor of anthropographics after adapting the measurement to avoid some cognitive biases such as social desirability bias [33] and diversification bias [68]. Results suggest that people tend to donate about $3.4 more when they see an anthropographic in comparison to a bar chart. Even though the amount of additional money is small since participants had $100 to allocate, this result brings up the discussion on whether it is worth for an NGO to change the visualization design in order to gain a few more dollars. It is also essential to keep in mind that we have only tested a limited part of the design space. Other design decisions might produce different gains.

7.1.2 Self-Reported Compassion and Affect

We also have captured the affective perspective of compassion through self-reported empathic concern and a more general scale of affective response. Results from the self-reported compassion in Chapter 5 show weak evidence in favor of the anthropographic design but with inconclusive information about the effect size. Besides the sample size, another aspect that seems to have affected results is because participants got confused by the terms that were translated from English to Portuguese. We decided to use a more straightforward and more intuitive scale in the crowdsourcing experiment.

We have found clear evidence that people present lower levels of valence when they see an anthropographic design instead of a bar chart. Although we cannot say the valence scale directly translates to compassion, this evidence points out that a specific kind of visualization design can affect people's emotional response. This result opens new opportunities to explore the role of anthropographics on compassion and affect.

7.2 Research Opportunities

This thesis opens opportunities for future research on Anthropographics. This section discusses possible research directions.

7.2.1 Fruther Exploring the Design Space

This thesis has explored positions in the design space of anthropographics that have never been investigated before. Until now, we could only find evidence in favor of anthropographics when combining maximum granularity, intermediate specificity, and intermediate realism. Dimensions such as coverage, authenticity, and situatedness are still unexplored. Other positions in the design space that belong to already explored dimensions are still lacking further investigation. Therefore, further studies need to be conducted in order to understand better which design decisions influence people's compassion. It is also important to make clear that, until now, no work has investigated anthropographics' design space dimensions separately. This step can also be an opportunity for future studies.

7.2.2 Using More Precise Measurements

Self-reported scales or hypothetical questions are easy to be used because they do not need additional costs (e.g., buying instruments or spending money with real donations). As a downside, such measurements are more prone to errors and subjective responses. Further studies on Anthropographics could capture compassion and prosocial behavior through more precise measurements.

Compassion and empathy could be measured through physiological measurements. Västfjäll and colleagues [82], for example, have used an instrument that captured compassion through the mapping of facial expressions. Such measures are less prone to cognitive biases and subjectivity [60]. However, using instruments that measure brain activity or even facial expressions is costly.

Future studies can also use real donations to measure prosocial behavior. We only used hypothetical questions because of IRB's constraints. However, further studies could capture prosocial behavior more precisely by asking real donation questions where participants would have to spend part of their experiment's payment. An intermediary approach is making participants perform real-effort tasks in order to donate (e.g., pressing a button 100 times to donate \$1) since people tend to donate less when they earn money by effort [52], which avoids the social desirability bias [33].

7.2.3 Investigating Other Response Variables

Recent studies find that the inclusion of pictures can help make charts more memorable [3; 12; 13], though it can also make them more challenging to process [12; 35]. One study found that the use of pictorial symbols to represent units of data can help people remember information and can encourage them to inspect visualizations more closely, apparently without any negative impact on legibility [35]. This suggests that when showing data about people, the use of anthropomorphic symbols may have benefits on memorability and engagement without clear downsides. However, except for the study from Boy et al. [14], we are not aware of any study that examines the benefits of this design strategy on prosocial feelings or behavior.

The focus of this thesis is testing to what extent anthropographics may evoke compassion for people. As far as we know, the effect on people's emotional response and prosocial behavior is small. Future studies may go farther and explore whether such visualization designs might also affect other outcomes such as memorability, comprehension, or engagement.

7.3 Concluding Thoughts

This thesis is one more step in understanding the role of anthropographics on compassion. We have conducted studies that point to evidence that anthropographics may affect how people donate to and feel about persons in need. There is a vast design space that can be explored in order to find whether the effect of creating visualizations with the intent to evoke compassion is minimal, or there are certain decisions that produce a higher emotional response or more significant donations. Therefore, our contributions create new research opportunities as well as motivate practitioners to explore new possibilities in the design space of anthropographics.

Bibliography

[1] Benjamin Bach, Pierre Dragicevic, Daniel Archambault, Christophe Hurter, and Sheelagh Carpendale. A review of temporal data visualizations based on space-time cube operations. In *Eurographics conference on visualization*, 2014.

[2] Benjamin Bach, Pierre Dragicevic, Samuel Huron, Petra Isenberg, Yvonne Jansen, Charles Perin, André Spritzer, Romain Vuillemot, Wesley Willett, and Tobias Isenberg. Illustrative data graphics in 18th-19th century style: A case study. In *IEEE Conference on Visualization-IEEE VIS 2013*, 2013.

[3] Scott Bateman, Regan L Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2573–2582. ACM, 2010.

[4] C Daniel Batson. These things called empathy: eight related but distinct phenomena. 2009.

[5] Michel Beaudouin-Lafon. Designing interaction, not interfaces. In *Proceedings of the working conference on Advanced visual interfaces*, pages 15–22. ACM, 2004.

[6] Ralf Bender and Stefan Lange. Adjusting for multiple testing—when and how? *Journal of clinical epidemiology*, 54(4):343–349, 2001.

[7] Jacques Bertin. *Sémiologie graphique: Les diagrammes-Les réseaux-Les cartes*. Gauthier-VillarsMouton & Cie, Paris, 1973.

[8] Enrico Bertini. Can visualization elicit empathy? our experiments with "anthropographics". https://medium.com/@FILWD/

can-visualization-elicit-empathy-our-experiments-with-anthropograph
2017. [Online; accessed 8-July-2019].

[9] Alberto Betella and Paul FMJ Verschure. The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS one*, 11(2), 2016.

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

[11] Paul Bloom. Empathy and its discontents. *Trends in cognitive sciences*, 21(1):24–31, 2017.

[12] Rita Borgo, Alfie Abdul-Rahman, Farhan Mohamed, Philip W Grant, Irene Reppa, Luciano Floridi, and Min Chen. An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2759–2768, 2012.

[13] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013.

[14] Jeremy Boy, Anshul Vikram Pandey, John Emerson, Margaret Satterthwaite, Oded Nov, and Enrico 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*, pages 5462–5474. ACM, 2017.

[15] James Calvert, Rhodora Abadla, and Syed Mohammad Tauseef. Design and testing of a virtual reality enabled experience that enhances engagement and simulates empathy for historical events and characters. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 868–869. IEEE, 2019.

[16] Sarah Campbell. Feeling numbers the rhetoric of pathos in visualization. Master's thesis, Northeastern University, 2018.

[17] Fredrik Carlsson, Haoran He, and Peter Martinsson. Easy come, easy go-the role of windfall money in lab and field experiments. *rapport nr.: Working Papers in Economics 374*, 2009.

[18] Norman Cliff. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological bulletin*, 114(3):494, 1993.

[19] Jacob Cohen. The earth is round (p<. 05). In *What if there were no significance tests?*, pages 69–82. Routledge, 2016.

[20] Geoff Cumming and Sue Finch. Inference by eye: confidence intervals and how to read pictures of data. *American psychologist*, 60(2):170, 2005.

[21] Luiz Augusto de Macêdo Morais, Nazareno Andrade, Dandara Maria Costa de Sousa, and Lesandro Ponciano. Defamiliarization, representation granularity, and user experience: A qualitative study with two situated visualizations. In *2019 IEEE Pacific Visualization Symposium (PacificVis)*, pages 92–101. IEEE, 2019.

[22] David DeSteno. Compassion and altruism: How our minds determine who is worthy of help. *Current opinion in behavioral Sciences*, 3:80–83, 2015.

[23] Selan Rodrigues Dos Santos. *A framework for the visualization of multidimensional and multivariate data*. PhD thesis, University of Leeds, 2004.

[24] Catherine D'Ignazio and Lauren Klein. *Data Feminism*. 2019. [Online; accessed 16-December-2019].

[25] Catherine D'Ignazio and Lauren F Klein. Feminist data visualization. In *Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore. IEEE*, 2016.

[26] Arvid Erlandsson, Sigrid Møyner Hohle, Erik Løhre, and Daniel Västfjäll. The rise and fall of scary numbers: The effect of perceived trends on future estimates, severity ratings, and help-allocations in a cancer context. *Journal of Applied Social Psychology*, 48(11):618–633, 2018.

[27] Ana Figueiras. How to tell stories using visualization. In *Information Visualisation (IV), 2014 18th International Conference on*, pages 18–18. IEEE, 2014.

[28] Ana Figueiras. Narrative visualization: A case study of how to incorporate narrative elements in existing visualizations. In *Information Visualisation (IV), 2014 18th International Conference on*, pages 46–52. IEEE, 2014.

[29] Robert J Fisher. Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20(2):303–315, 1993.

[30] Alexander Genevsky, Daniel Västfjäll, Paul Slovic, and Brian Knutson. Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *Journal of Neuroscience*, 33(43):17188–17196, 2013.

[31] Jennifer L Goetz, Dacher Keltner, and Emiliana Simon-Thomas. Compassion: an evolutionary analysis and empirical review. *Psychological bulletin*, 136(3):351, 2010.

[32] Colin M Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L Toombs. The dark (patterns) side of ux design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 534. ACM, 2018.

[33] Pamela Grimm. Social desirability bias. *Wiley international encyclopedia of marketing*, 2010.

[34] Jenny Gu, Kate Cavanagh, Ruth Baer, and Clara Strauss. An empirical examination of the factor structure of compassion. *PloS one*, 12(2):e0172471, 2017.

[35] Steve Haroz, Robert Kosara, and Steven L Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 1191–1200. ACM, 2015.

[36] Jacob Harris. Connecting with the dots. https://source.opennews.org/articles/connecting-dots/, 2015. [Online; accessed 29-January-2019].

[37] Uta Hinrichs and Stefania Forlini. In defense of sandcastles: Research thinking through visualization in dh. In *Proceedings of the conference on Digital Humanities*. International Alliance of Digital Humanities Organizations (ADHO), 2017.

[38] Trevor Hogan, Uta Hinrichs, and Eva Hornecker. The elicitation interview technique: Capturing people's experiences of data representations. *IEEE transactions on visualization and computer graphics*, 22(12):2579–2593, 2016.

[39] Alexander Ivanov, Kurtis Thorvald Danyluk, and Wesley Willett. Exploration & anthropomorphism in immersive unit visualizations. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, page LBW008. ACM, 2018.

[40] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper 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, April 2015. ACM.

[41] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. Opportunities and challenges for data physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3227–3236. ACM, 2015.

[42] Matthew R Jordan, Dorsa Amir, and Paul Bloom. Are empathy and concern psychologically distinct? *Emotion*, 16(8):1107, 2016.

[43] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[44] Maria Kandaurova and Seung Hwan Mark Lee. The effects of virtual reality (vr) on charitable giving: The role of empathy, guilt, responsibility, and social exclusion. *Journal of Business Research*, 100:571–580, 2019.

[45] Helen Kennedy and Rosemary Lucy Hill. The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology*, 52(4):830–848, 2018.

[46] Lisa Koeman. *Urban visualisation: the role of situated technology interventions in facilitating engagement with local topics*. PhD thesis, UCL (University College London), 2017.

[47] Tehila Kogut and Ilana Ritov. The "identified victim" effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, 18(3):157–167, 2005.

[48] LabRua. Espaços Públicos do Açude Velho, Campina Grande-PB. http://www.labrua.org/pesquisas/acudevelho/. [Online; accessed 16-December-2019].

[49] Steve Lambert. And what do i do now? using data visualization for social change. https://c4aa.org/2016/01/data-visualization-for-what/, 2016. [Online; accessed 29-January-2019].

[50] Mathieu Le Goc, Lawrence H Kim, Ali Parsaei, Jean-Daniel Fekete, Pierre Dragicevic, and Sean Follmer. Zooids: Building blocks for swarm user interfaces. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 97–109. ACM, 2016.

[51] Seyoung Lee and Thomas Hugh Feeley. The identifiable victim effect: A meta-analytic review. *Social Influence*, 11(3):199–215, 2016.

[52] Huafang Li, Jie Liang, Hui Xu, and Yingxi Liu. Does windfall money encourage charitable giving? an experimental study. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 30(4):841–848, 2019.

[53] Giorgia 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.

[54] Giorgia Lupi. Bruises: The data we don't see. http://giorgialupi.com/bruises-the-data-we-dont-see, 2019. [Online; accessed 17-July-2019].

[55] William MacAskill. *Doing good better: Effective altruism and a radical new way to make a difference*. Guardian Faber Publishing, 2015.

[56] Ezra Markowitz, Paul Slovic, Daniel Vastfjall, and Sara Hodges. Compassion fade and the challenge of environmental conservation. 2013.

[57] John McCarthy and Peter Wright. *Technology as Experience*, volume 1. MIT Press, London, 2004.

[58] K Meissel. A practical guide to using cliff's delta as a measure of effect size where parametric equivalents are inappropriate. In *ACSPRI Social Science Methodology Conference*, 2010.

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

[60] David L Neumann, Raymond CK Chan, Gregory J Boyle, Yi Wang, and H Rae Westbury. Measures of empathy: Self-report, behavioral, and neuroscientific approaches. In *Measures of personality and social psychological constructs*, pages 257–289. Elsevier, 2015.

[61] Marie Neurath. Isotype. *Instructional science*, 3(2):127–150, 1974.

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

[63] George E Newman and Paul Bloom. Physical contact influences how much people pay at celebrity auctions. *Proceedings of the National Academy of Sciences*, 111(10):3705–3708, 2014.

[64] Heather L. O'Brien and Elaine G. Toms. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6):938–955, apr 2008.

[65] Evan M Peck, Sofia E Ayuso, and Omar El-Etr. Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. *arXiv preprint arXiv:1901.01920*, 2019.

[66] Charles Perin. The symmetry of my life: An autobiographical visualization. In *Poster at IEEE VIS*, 2017.

[67] Charles Perin. The symmetry of my life ii. In *Poster at IEEE VIS*, 2018.

[68] Daniel Read and George Loewenstein. Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, 1(1):34, 1995.

[69] Tom Regan and Peter Singer. *Animal rights and human obligations*. Prentice-Hall, Englewood Cliffs, NJ, 2 edition, 1989.

[70] Laurel D Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. How anthropomorphism affects empathy toward robots. In *Proceedings of the*

4th ACM/IEEE international conference on Human robot interaction, pages 245–246. ACM, 2009.

[71] Yvonne Rogers and Paul Marshall. Research in the wild. *Synthesis Lectures on Human-Centered Informatics*, 10(3):i–97, 2017.

[72] Joe Rosenberg. The known unknown. https://99percentinvisible.org/episode/the-known-unknown/, 2019. The 99% Invisible podcast.

[73] Lisa Charlotte Rost. Data point moves into a bar. https://lab.dsst.io/slides/33c3/7999.html, 2016. [Online; accessed 15-March-2019].

[74] Robert E Roth. Visual variables. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, pages 1–11, 2016.

[75] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, 23(1):341–350, 2017.

[76] Hikmet Senay and Eve Ignatius. *Rules and principles of scientific data visualization*. Institute for Information Science and Technology, Department of Electrical Engineering, 1990.

[77] Clara Strauss, Billie Lever Taylor, Jenny Gu, Willem Kuyken, Ruth Baer, Fergal Jones, and Kate Cavanagh. What is compassion and how can we measure it? a review of definitions and measures. *Clinical psychology review*, 47:15–27, 2016.

[78] Alex S. Taylor, Siân Lindley, Tim Regan, and David Sweeney. Data-in-Place: Thinking through the Relations Between Data and Community. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, pages 2863–2872, 2015.

[79] Bruce H Thomas, Gregory F Welch, Pierre Dragicevic, Niklas Elmqvist, Pourang Irani, Yvonne Jansen, Dieter Schmalstieg, Aurélien Tabard, Neven AM ElSayed, Ross T Smith, et al. Situated analytics. In *Immersive Analytics*, pages 185–220. Springer, 2018.

[80] Brian Tomasik. The importance of wild-animal suffering. *Rel.: Beyond Anthropocentrism*, 3:133, 2015.

[81] Gill Valentine. The geography of women's fear. *Area*, pages 385–390, 1989.

[82] Daniel Västfjäll, Paul Slovic, Marcus Mayorga, and Ellen Peters. Compassion fade: Affect and charity are greatest for a single child in need. *PloS one*, 9(6):e100115, 2014.

[83] Data Visualization and Human Rights Website. Near and far. http://visualizingrights.org/kit/improvements/near-and-far.html, 2018. [Online; accessed 11-February-2019].

[84] Yun Wang, Adrien Segal, Roberta Klatzky, Daniel F Keefe, Petra Isenberg, Jörn Hurtienne, Eva Hornecker, Tim Dwyer, and Stephen Barrass. An emotional response to the value of visualization. *IEEE computer graphics and applications*, 39(5):8–17, 2019.

[85] Mark Ottoni Wilhelm and René Bekkers. Helping behavior, dispositional empathic concern, and the principle of care. *Social Psychology Quarterly*, 73(1):11–32, 2010.

[86] Wesley Willett, Yvonne Jansen, and Pierre Dragicevic. Embedded data representations. *IEEE transactions on visualization and computer graphics*, 23(1):461–470, 2017.

Appendix A

Anthropographics List

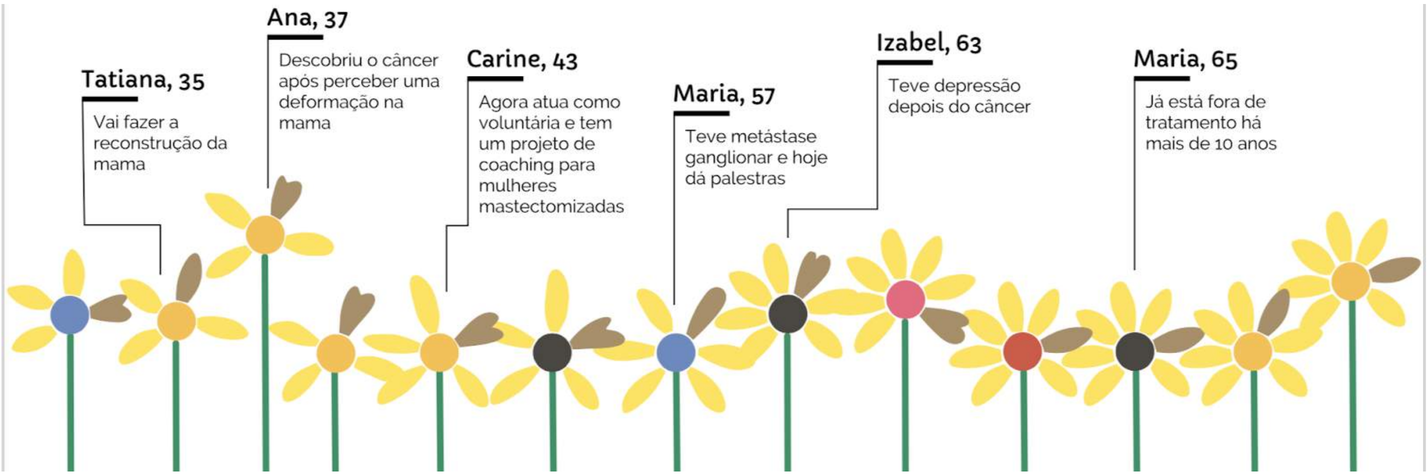
This appendix shows a list of images that lie in different positions of the anthropographics design space presented in Chapter 3.

1. Personal activities [Morais and Andrade]

http://www.luizaugustomm.me/papers/2019_pacificvis_defamiliarization.pdf

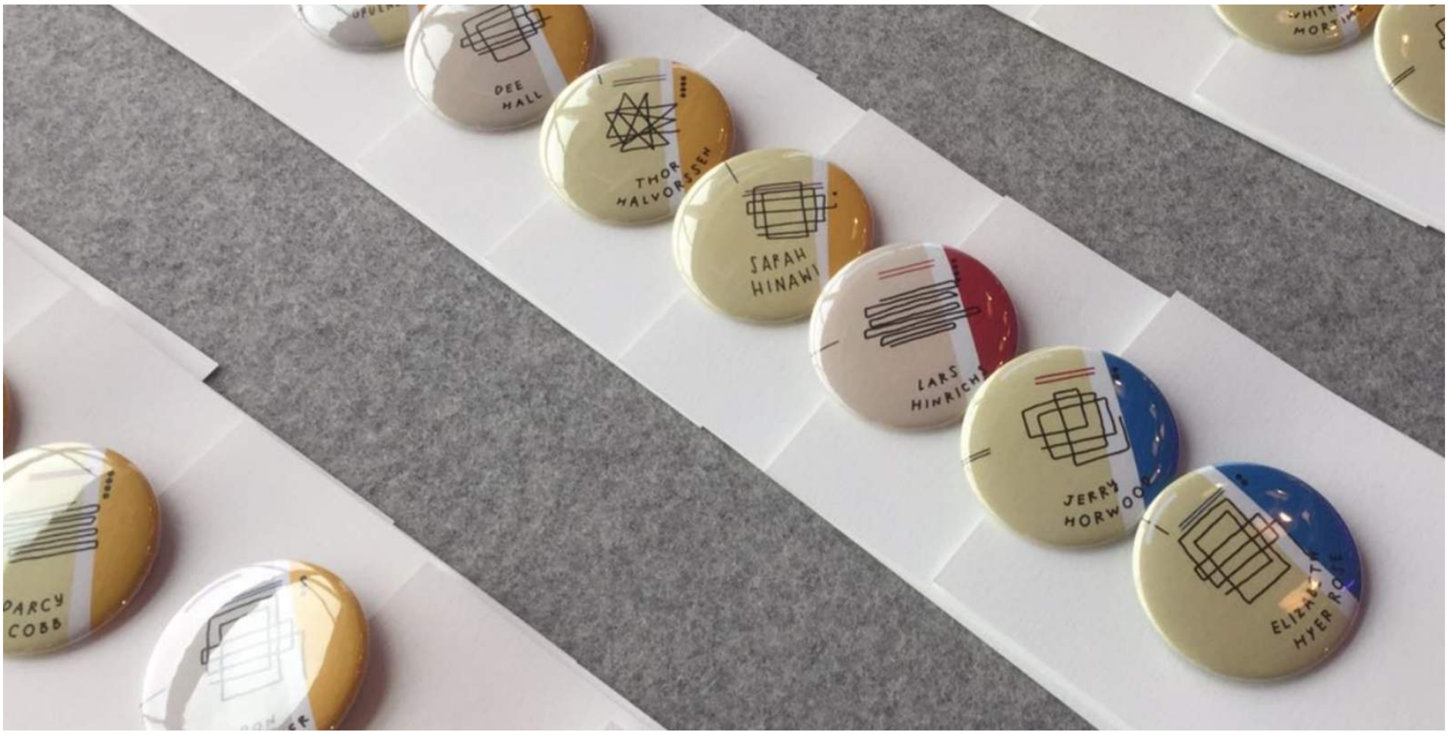
2. Harassment plants [Morais, Yasser, Sousa, and Andrade]

3. Cancer is not always the end [Morais]



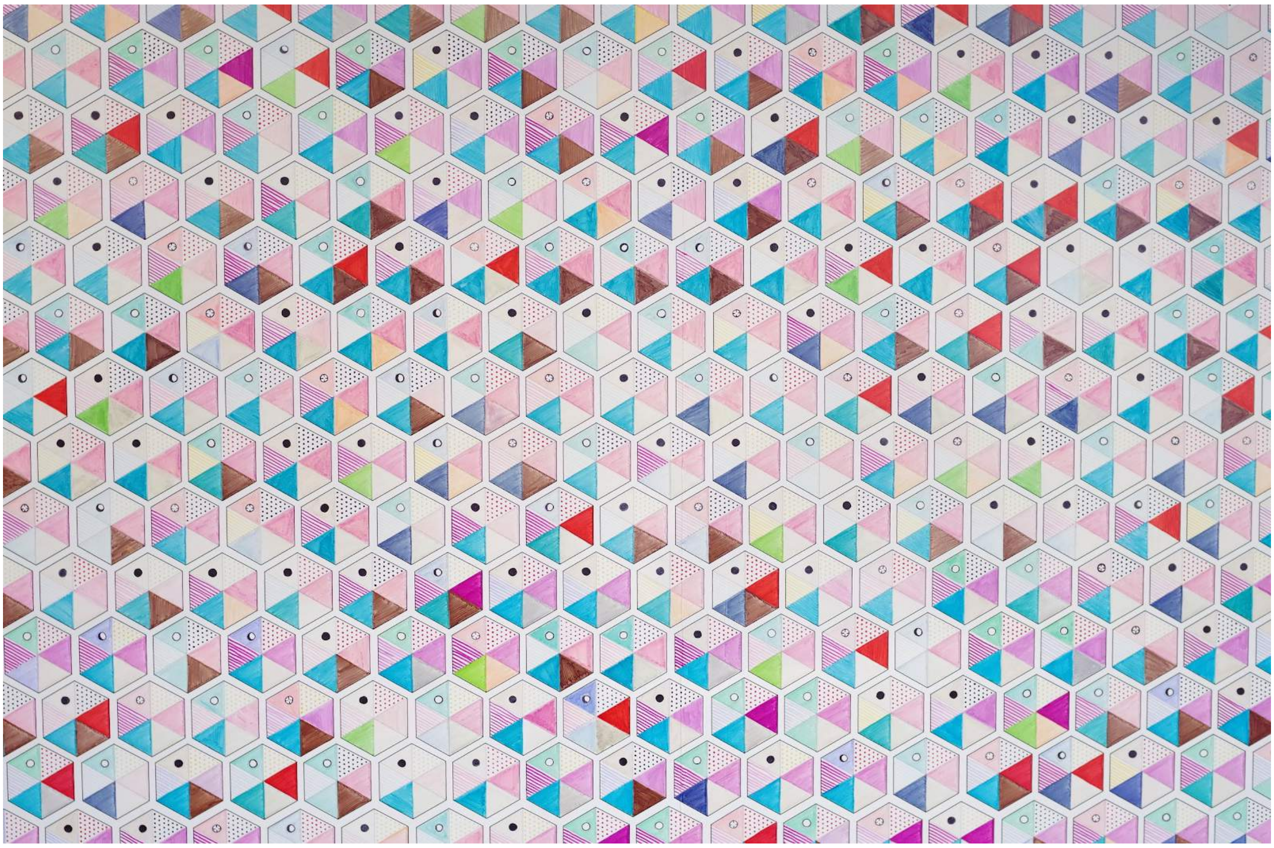
https://drive.google.com/open?id=10XfdsE-fvyEhV_v_2L9LFNjOoDcNd-Ve

4. Data portraits [Lupi]



http://giorgialupi.com/data-portraits-at-ted2017/

5. Data wallpaper [Lupi]

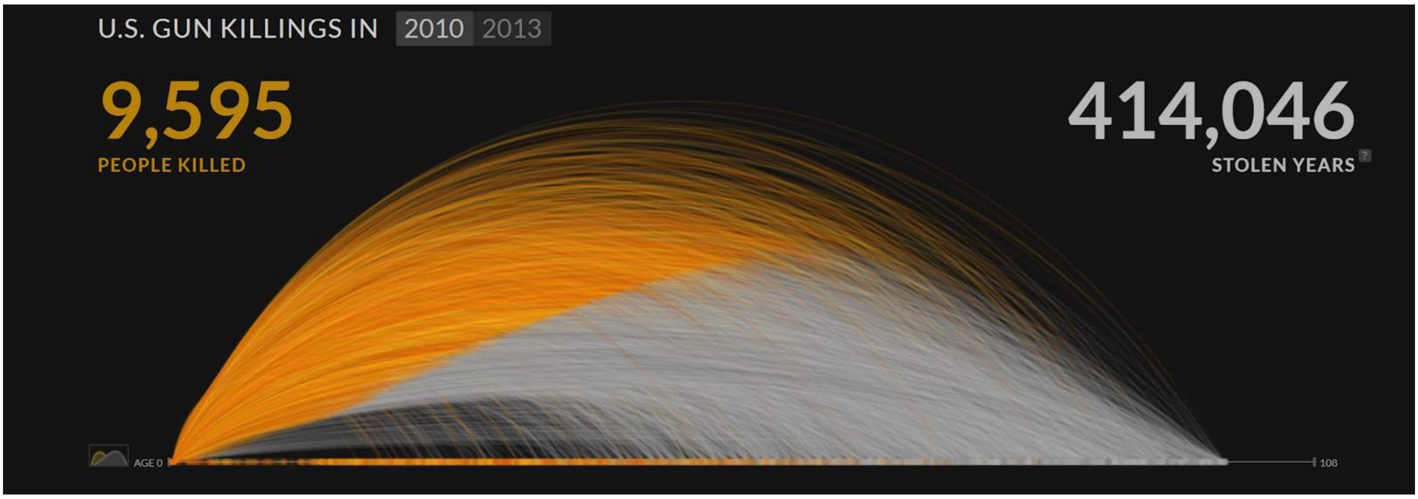


http://giorgialupi.com/collaborative-data-wallpaper-for-story/

6. Titanic dots [Unknown]

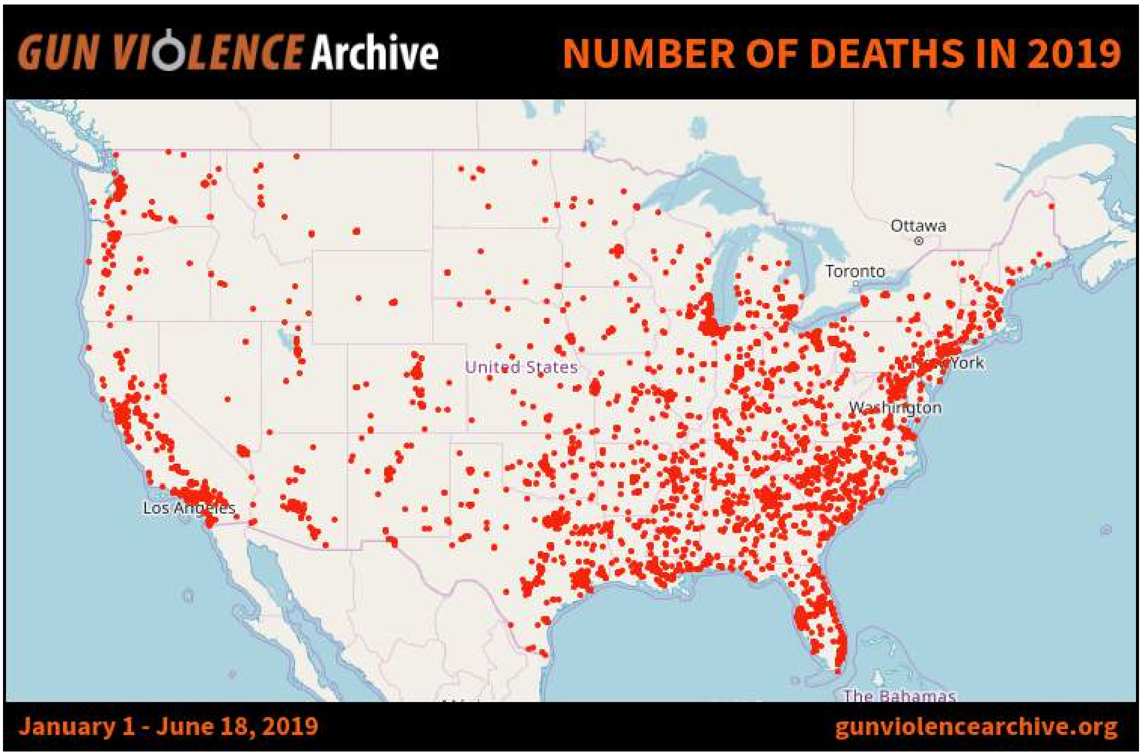
http://bl.ocks.org/aragorn87/raw/03e01fd46d488015ba072fda0476690b/

7. U.S. gun deaths [Periscopic]



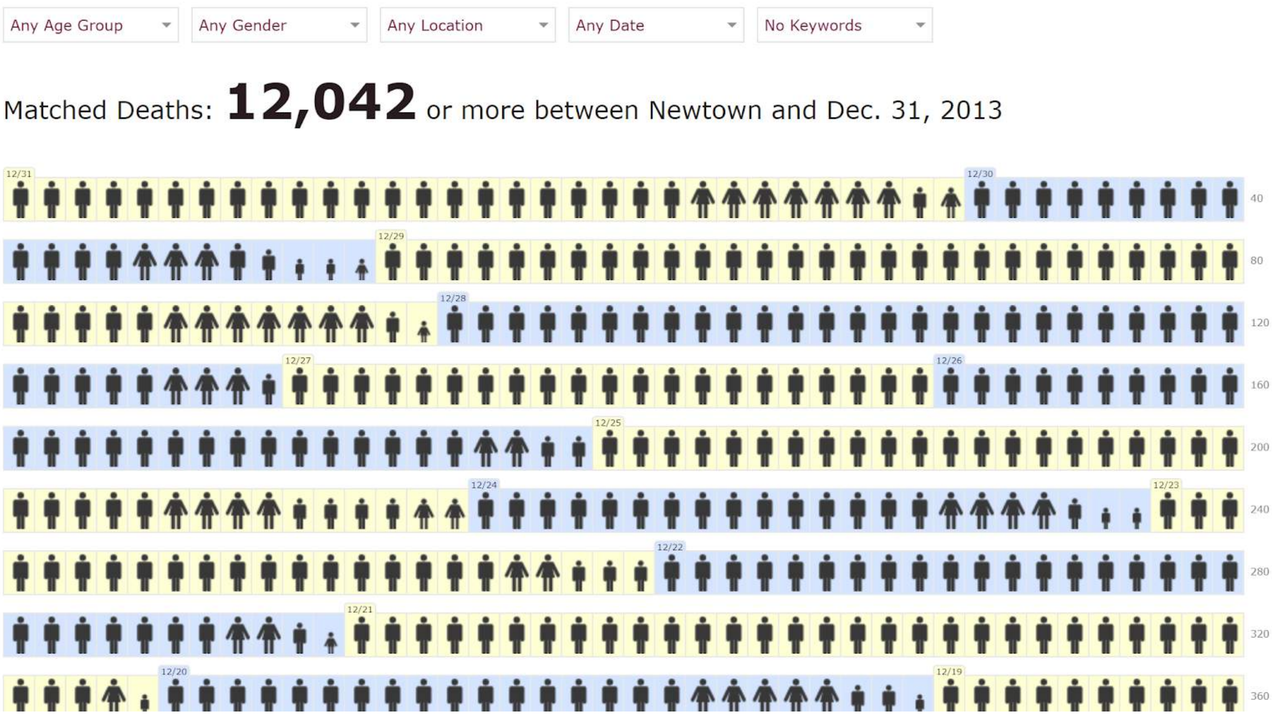
https://guns.periscopic.com/?year=2013

8. Gun violence chart [Gun violence archive]



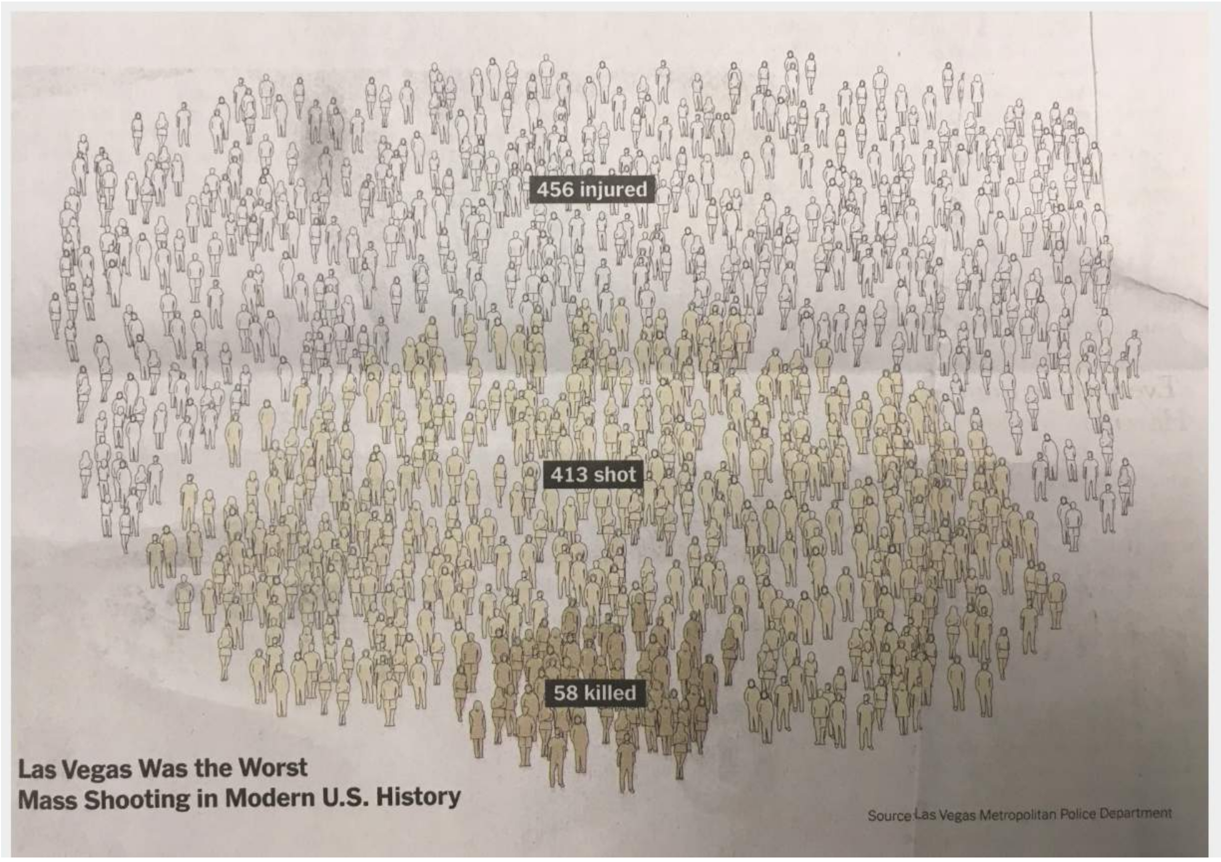
https://www.gunviolencearchive.org/

9. How Many People Have Been Killed by Guns Since Newtown? [Slate]



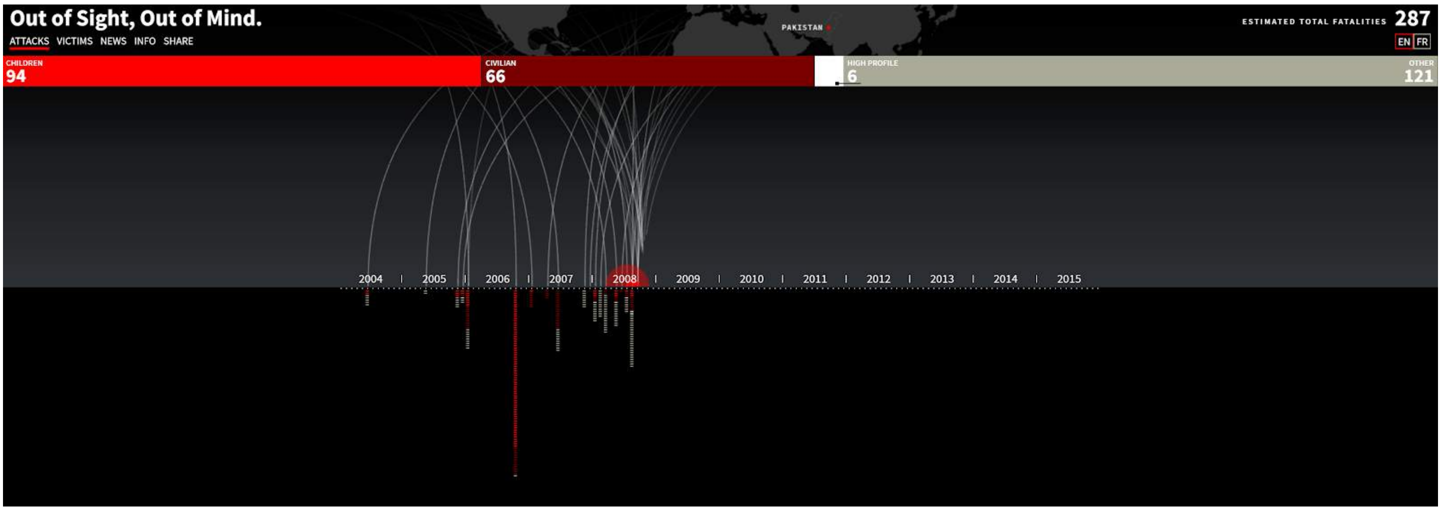
http://www.slate.com/articles/news_and_politics/crime/2012/12/gun_death_tally_every_american_gun_death_since_newtown_sandy_hook_shooting.html

10. Las Vegas shooting [NYT]



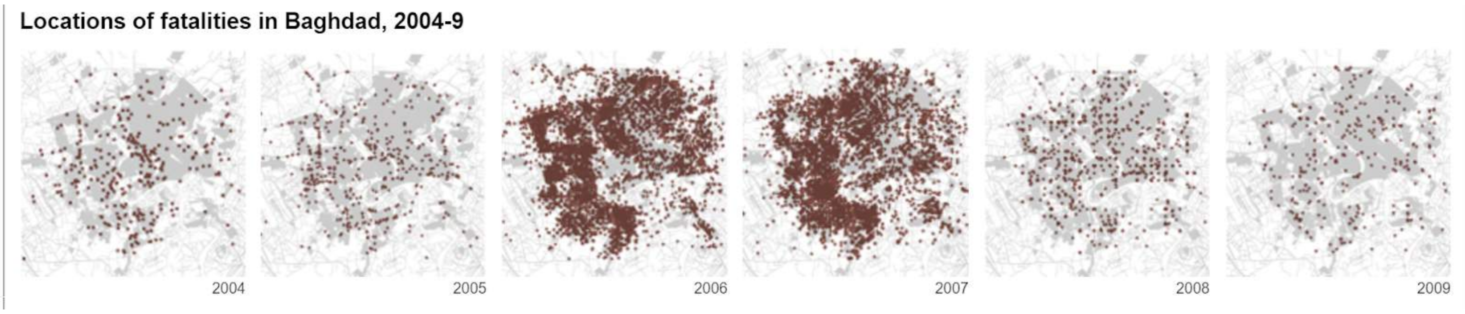
https://dataanddragons.wordpress.com/2018/10/08/time-and-space/

11. Out of sight, out of mind [Pitch Interactive]



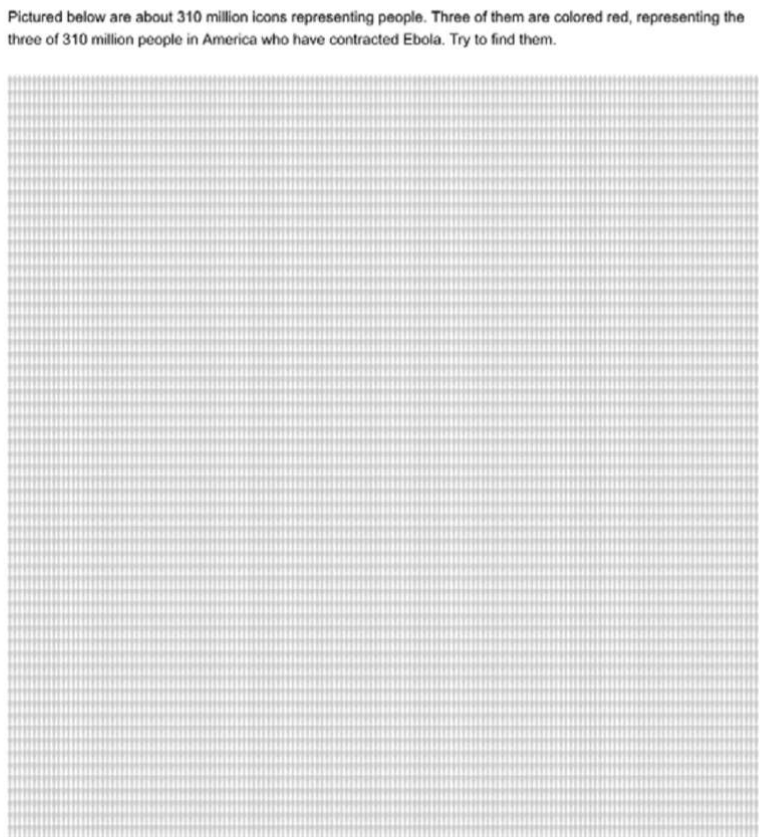
http://drones.pitchinteractive.com/

12. Local fatalities in Baghdad [NYT]



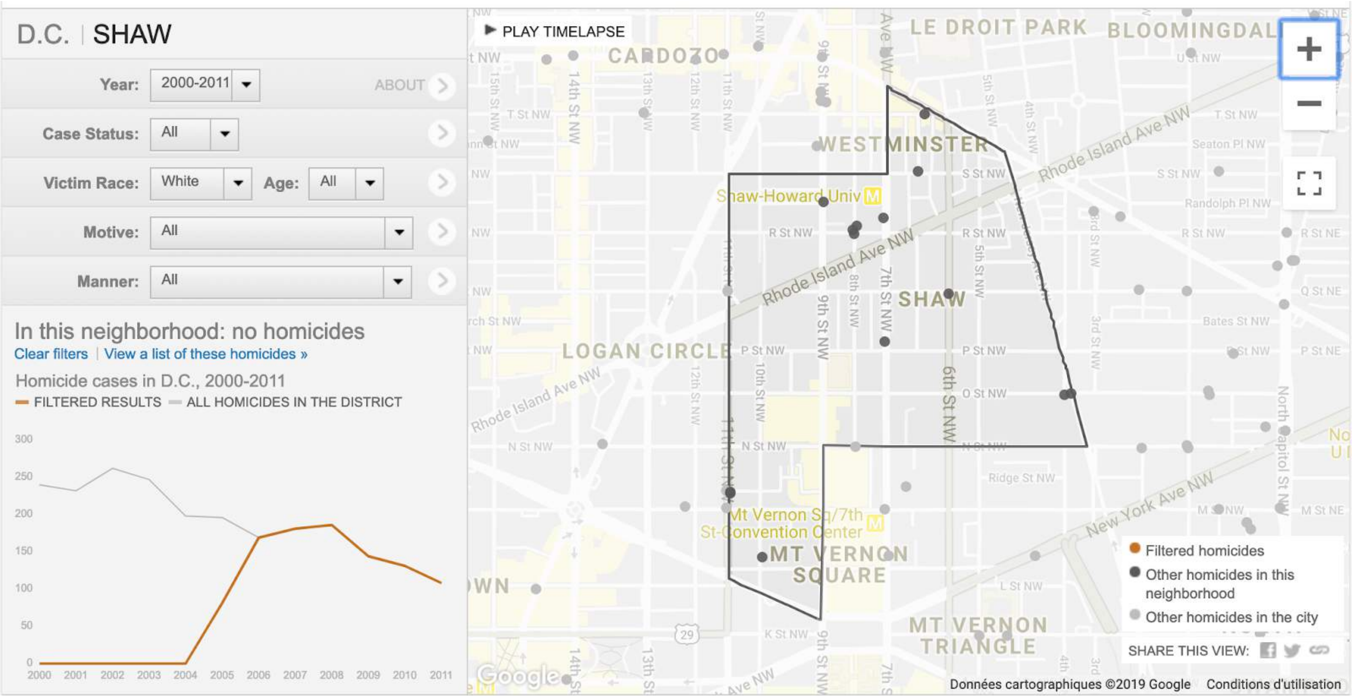
https://archive.nytimes.com/www.nytimes.com/interactive/2010/10/24/world/1024-surge-graphic.html

13. Want to feel better about Ebola? This (massive) chart should do the trick. [WP]



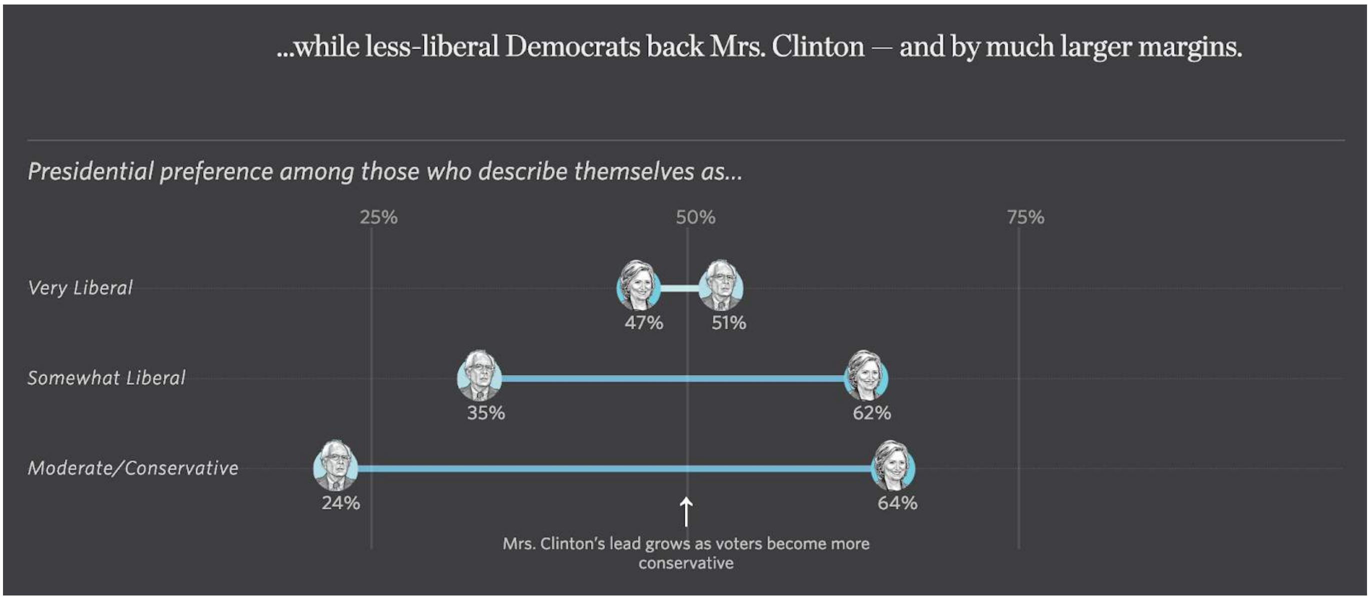
https://www.washingtonpost.com/news/the-fix/wp/2014/10/16/want-to-feel-better-about-ebola-this-massive-chart-should-do-the-trick/

14. Homicides in the District [WP]



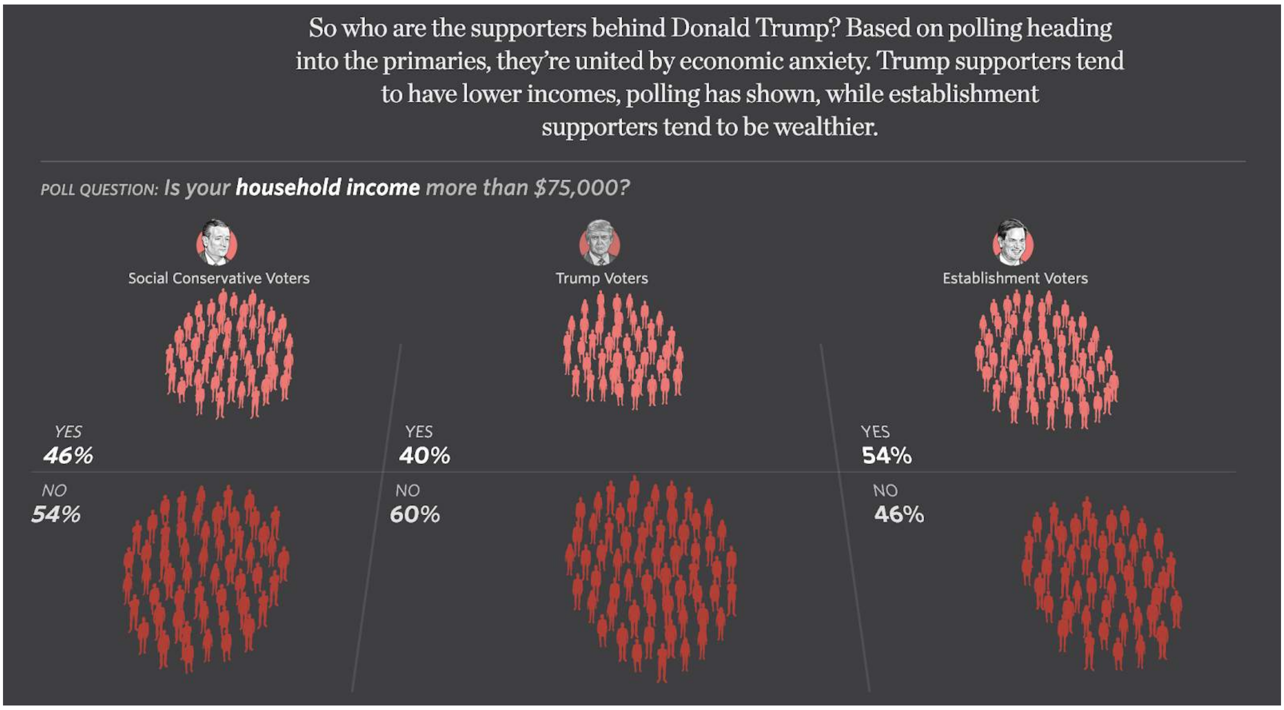
http://apps.washingtonpost.com/investigative/homicides/#211:0:all:all:white:all:all:all

15. How Sanders Happened [WSJ]



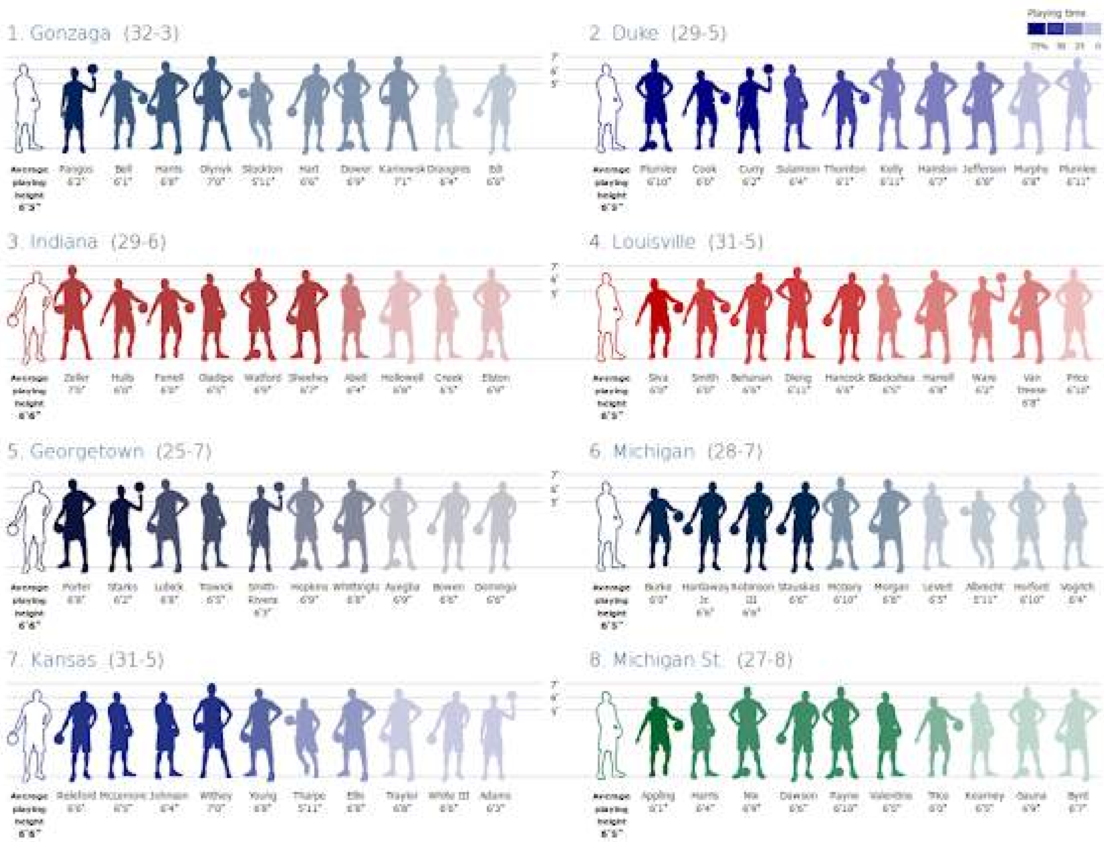
http://graphics.wsj.com/elections/2016/how-sanders-happened/

16. How Trump Happened [WSJ]



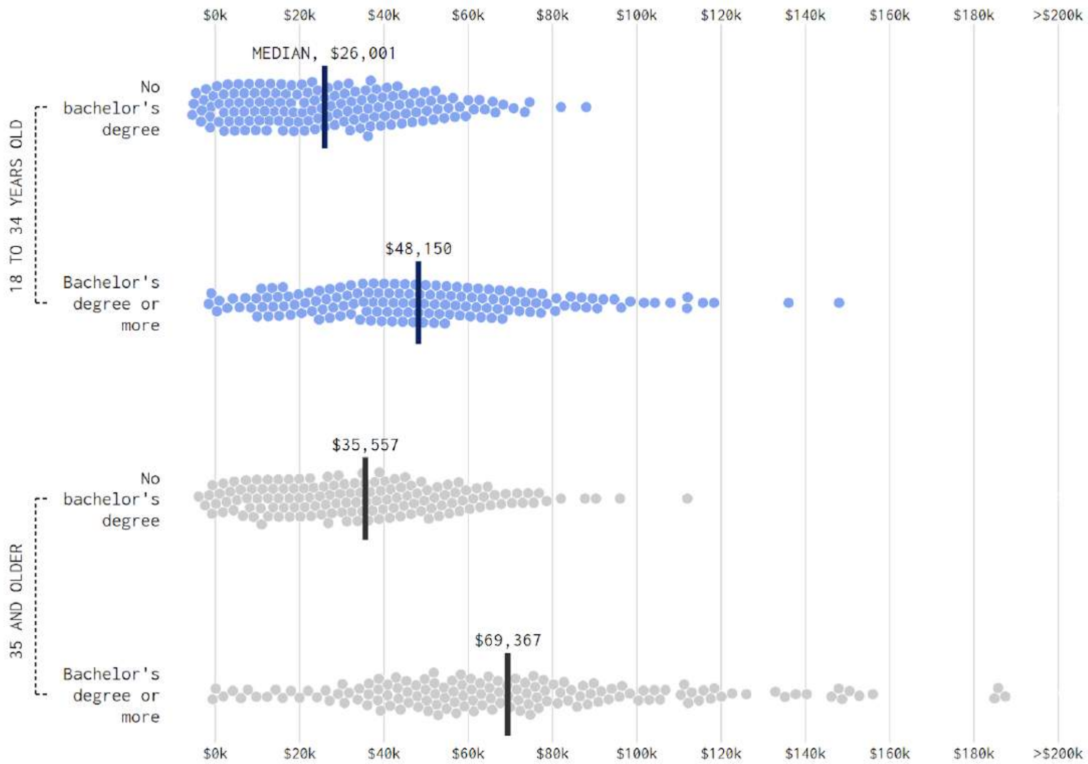
http://graphics.wsj.com/elections/2016/how-trump-happened/

17. March Madness: do the tallest teams always win the NCAA championship? [The Guardian]



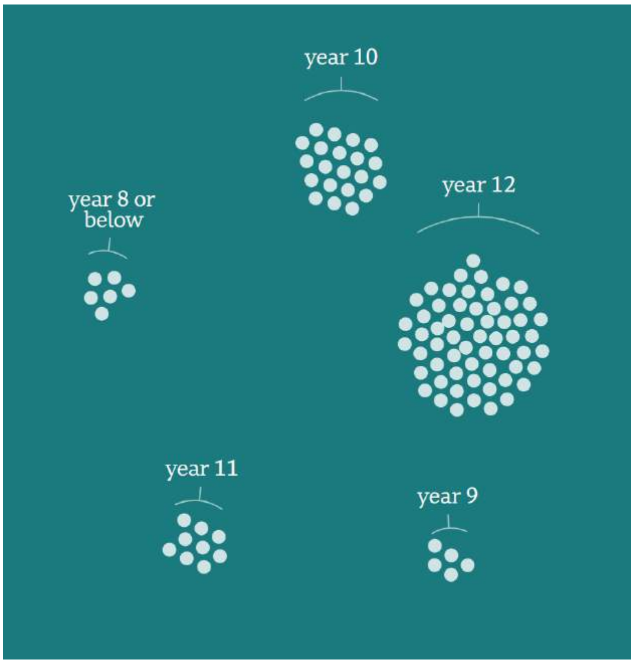
https://www.theguardian.com/sport/interactive/2013/mar/18/ncaa-tournament-team-matchups-height#g=Top%2025

18. Shifting incomes for young people [Yau]



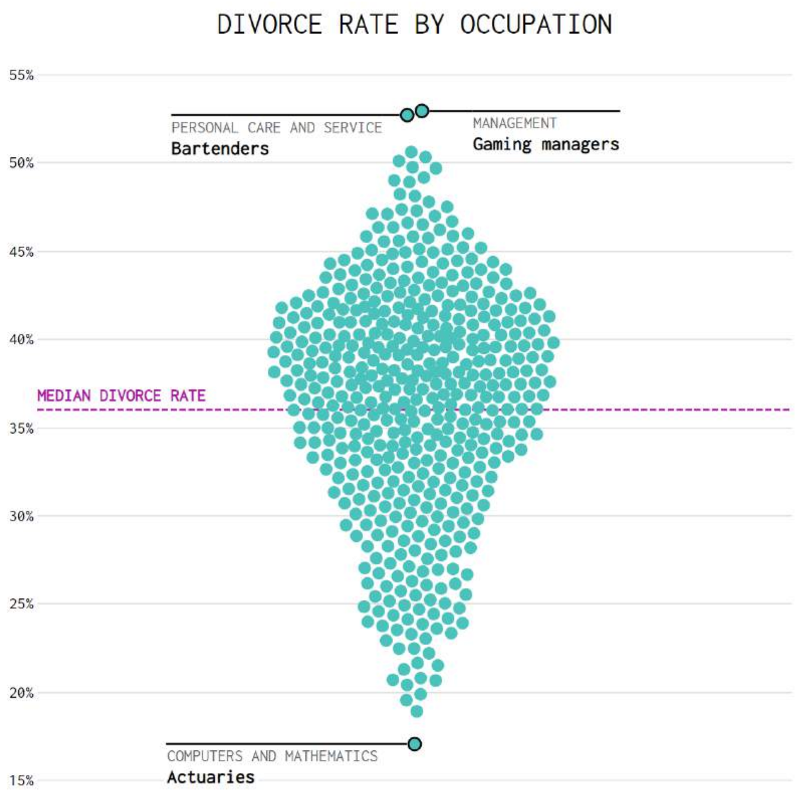
https://flowingdata.com/2017/05/02/shifting-incomes-for-young-people/

19. Australia as 100 people [Hanrahan and Elvery]



http://www.abc.net.au/news/2017-06-27/census-australia-as-100-people/8634318

20. Divorce and occupation [Yau]



https://flowingdata.com/2017/07/25/divorce-and-occupation/

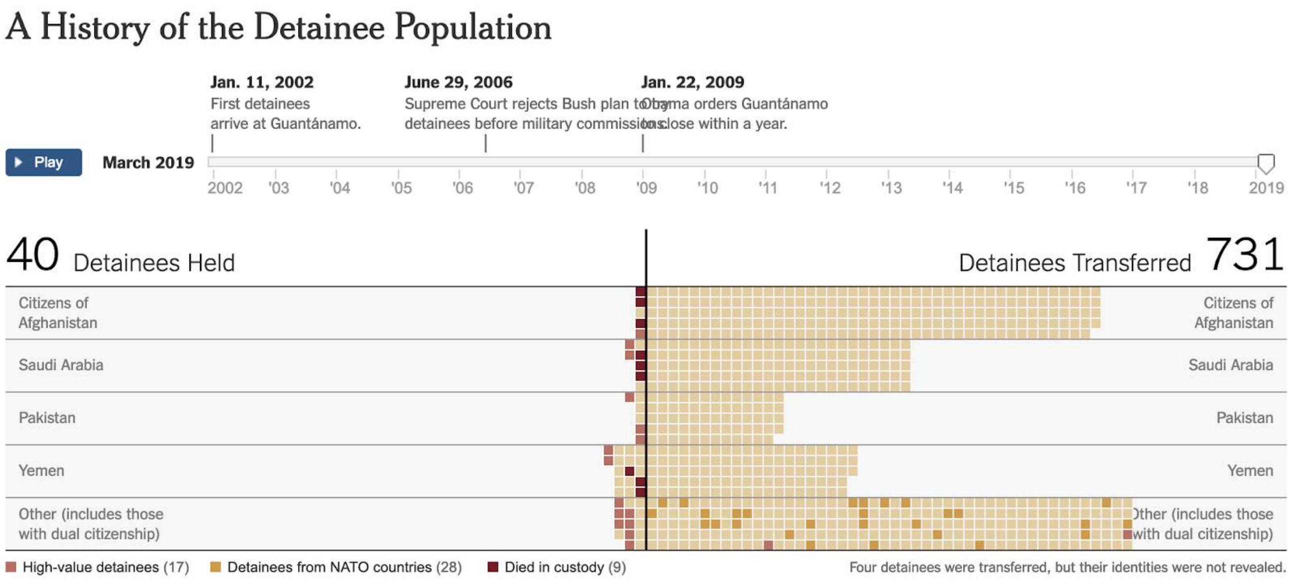
21. If only 100 people were in twitter [McCandless]

https://gizmodo.com/5330049/if-only-100-people-were-in-twitter

22. A message for the candidates [NYT]

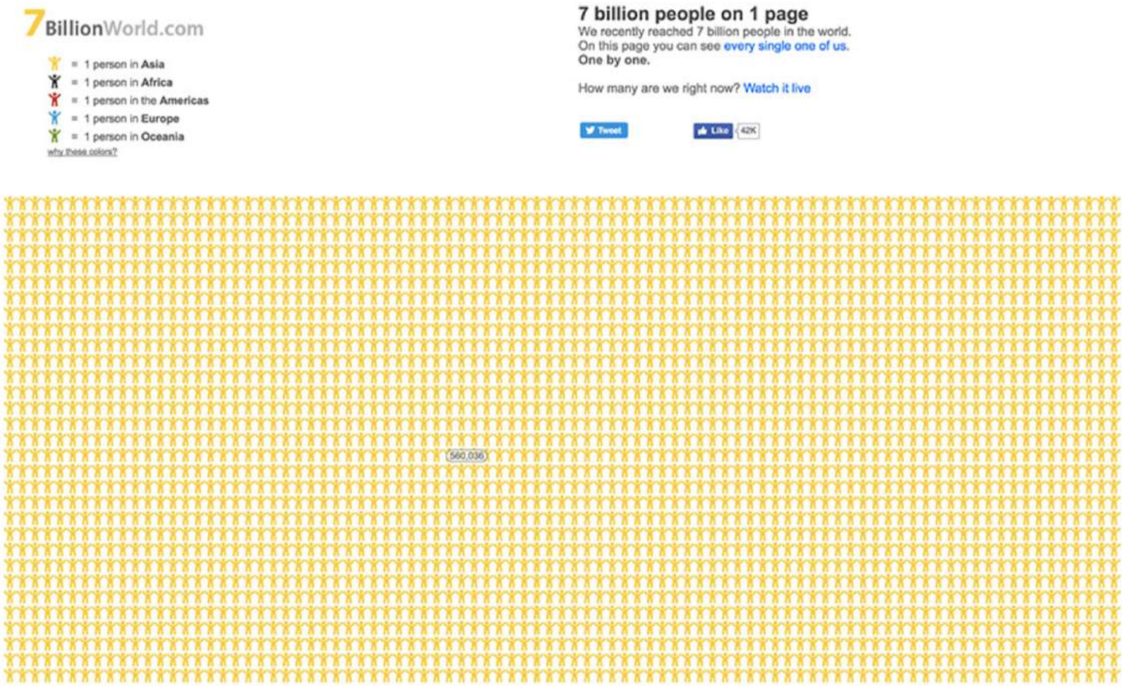
https://www.nytimes.com/2012/06/21/education/857-desks-call-attention-to-dropout-problem.html

23. A History of the Detainee Population [NYT]



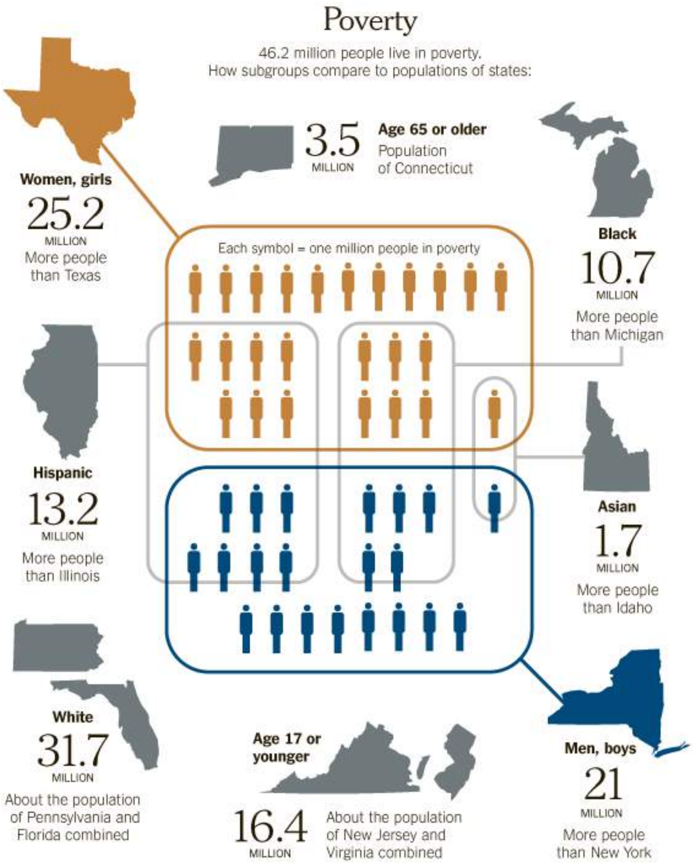
http://projects.nytimes.com/quantanamo

24. Seven Billion World [Unknown]



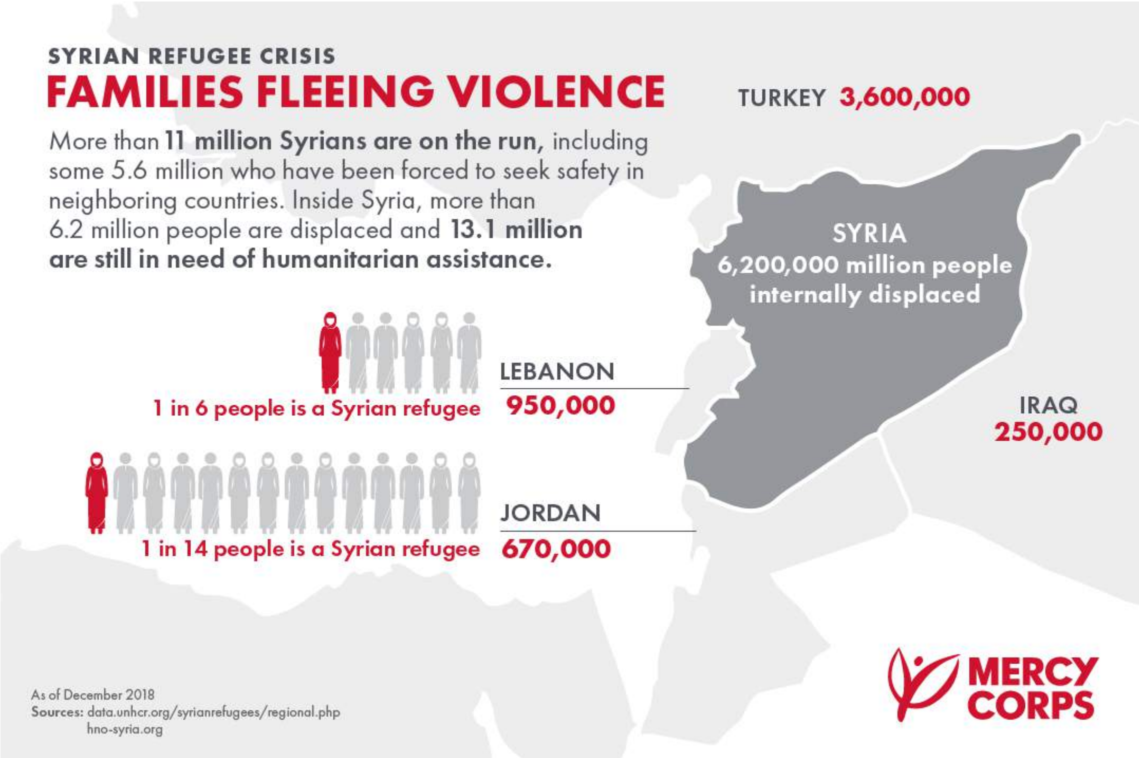
http://www.7billionworld.com/

25. The Impoverished States of America [Kuntz and Marsh]



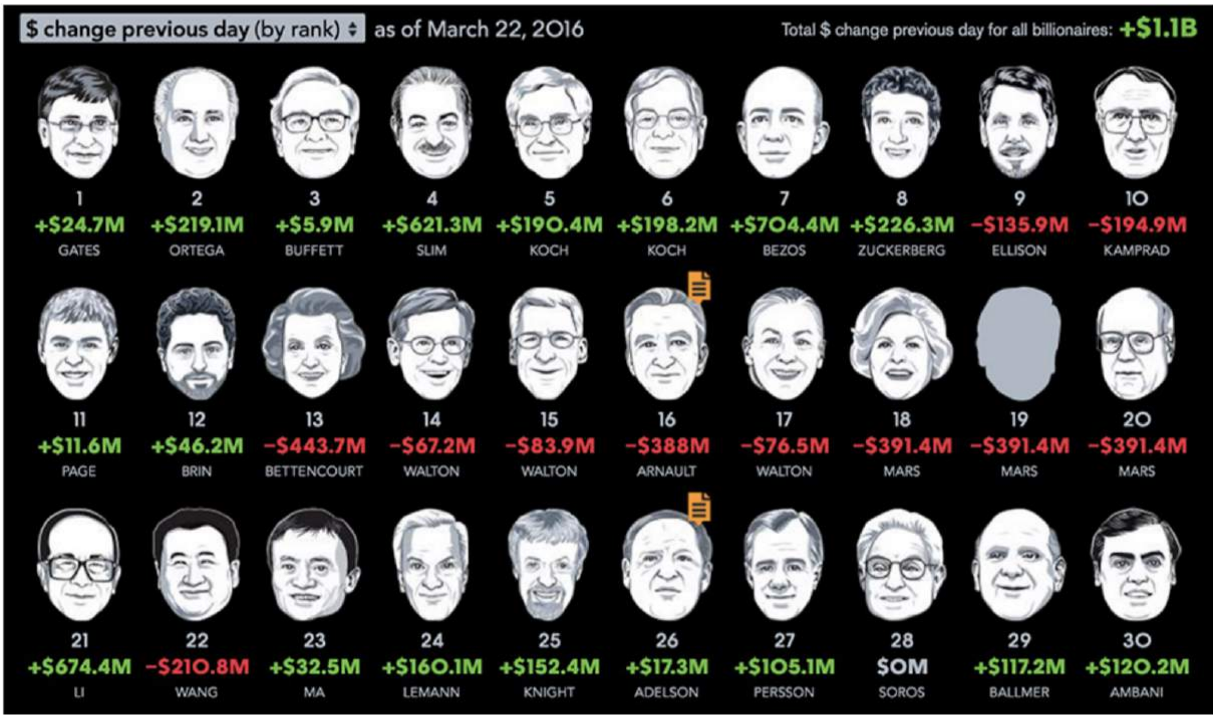
https://archive.nytimes.com/www.nytimes.com/interactive/2011/09/18/sunday-review/20110918_Poverty.html?_r=0

26. Quick facts: What you need to know about the Syria crisis [Mercy Corps]



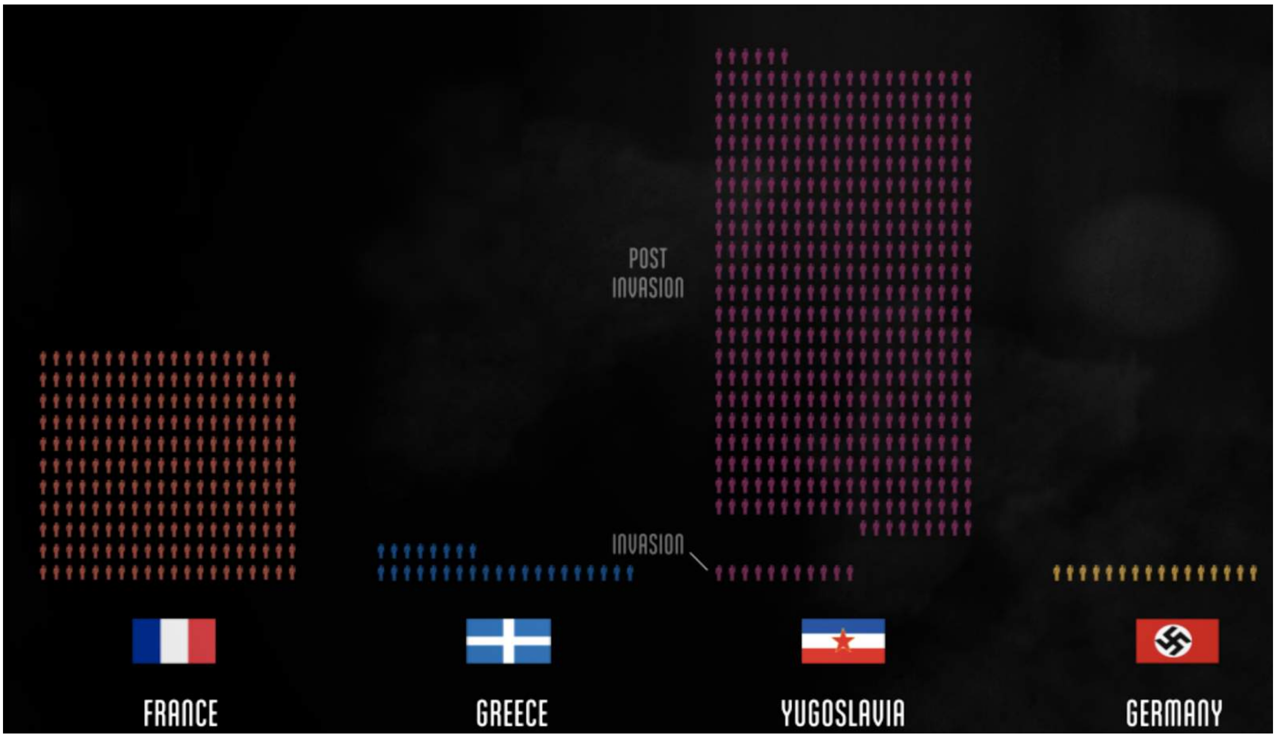
https://www.mercycorps.org/articles/iraq-jordan-lebanon-syria-turkey/quick-facts-what-you-need-know-about-syria-crisis

27. The Billionaire Index [Bloomberg]



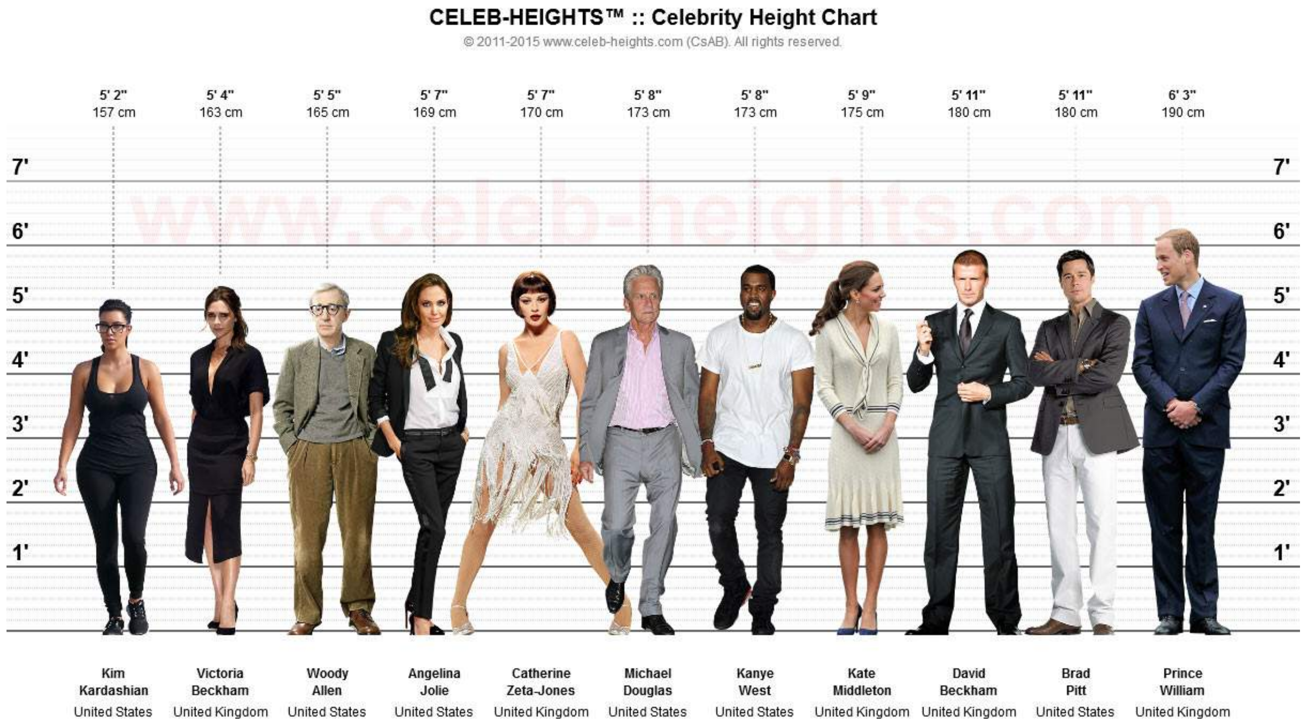
http://www.bloomberg.com/billionaires/2016-03-22/cya

28. The fallen of World War II [Halloran]



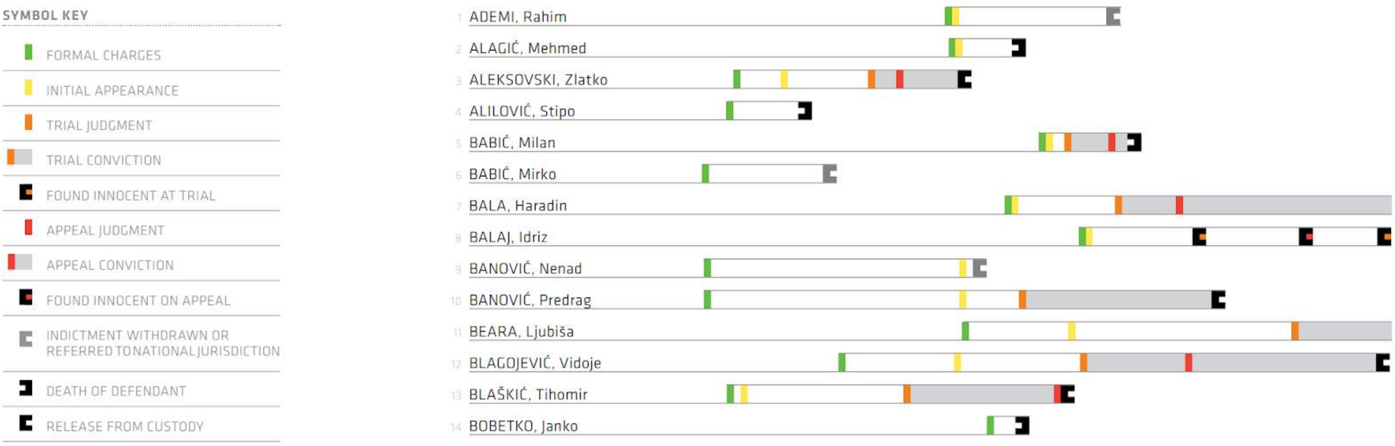
http://www.fallen.io/ww2/

29. Celebrity Height Chart [Unknown]



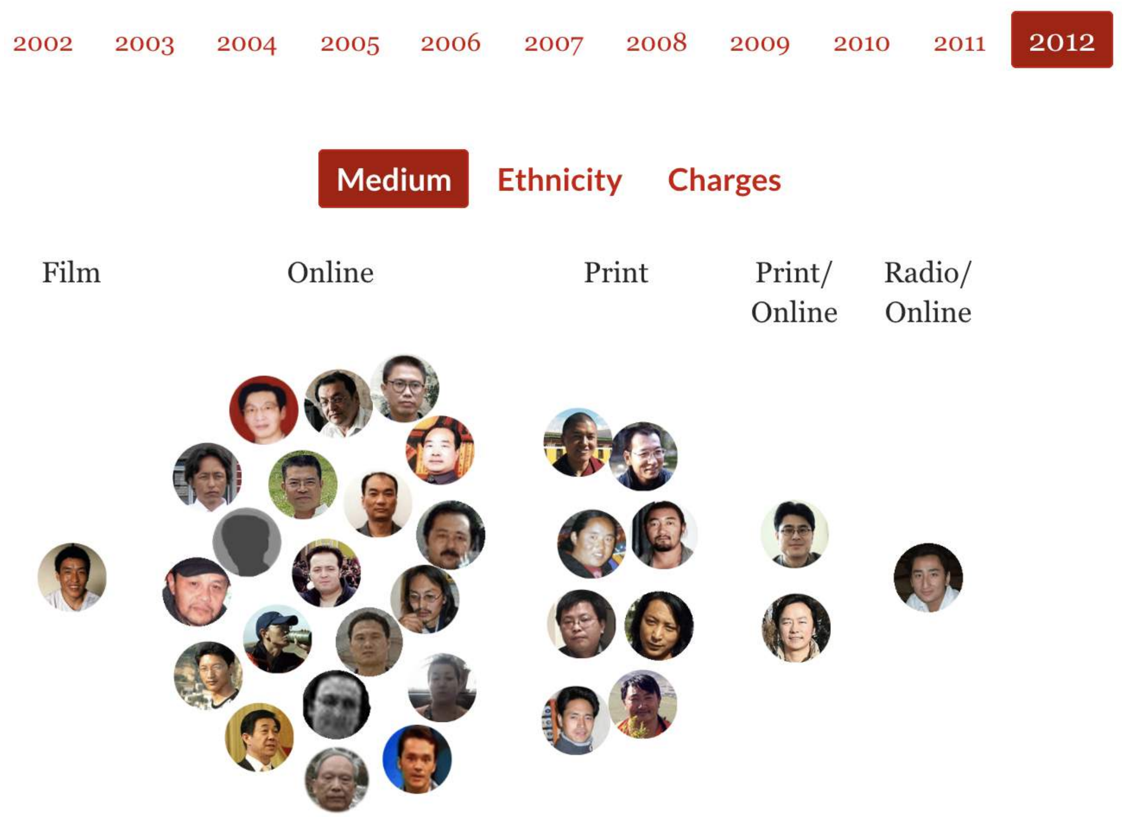
https://celeb-heights.com/celebrity.php?name=Angelina+Jolie

30. International Criminal Tribunals [Leitner Center]



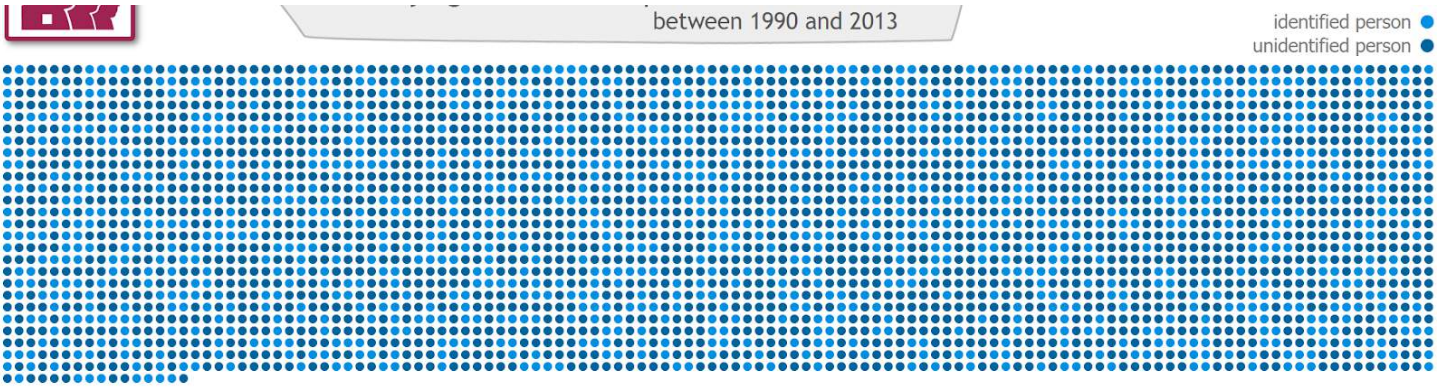
http://www.leitnercenter.org/files/News/International%20Criminal%20Tribunals.pdf

31. Journalists imprisoned in China [CPj]



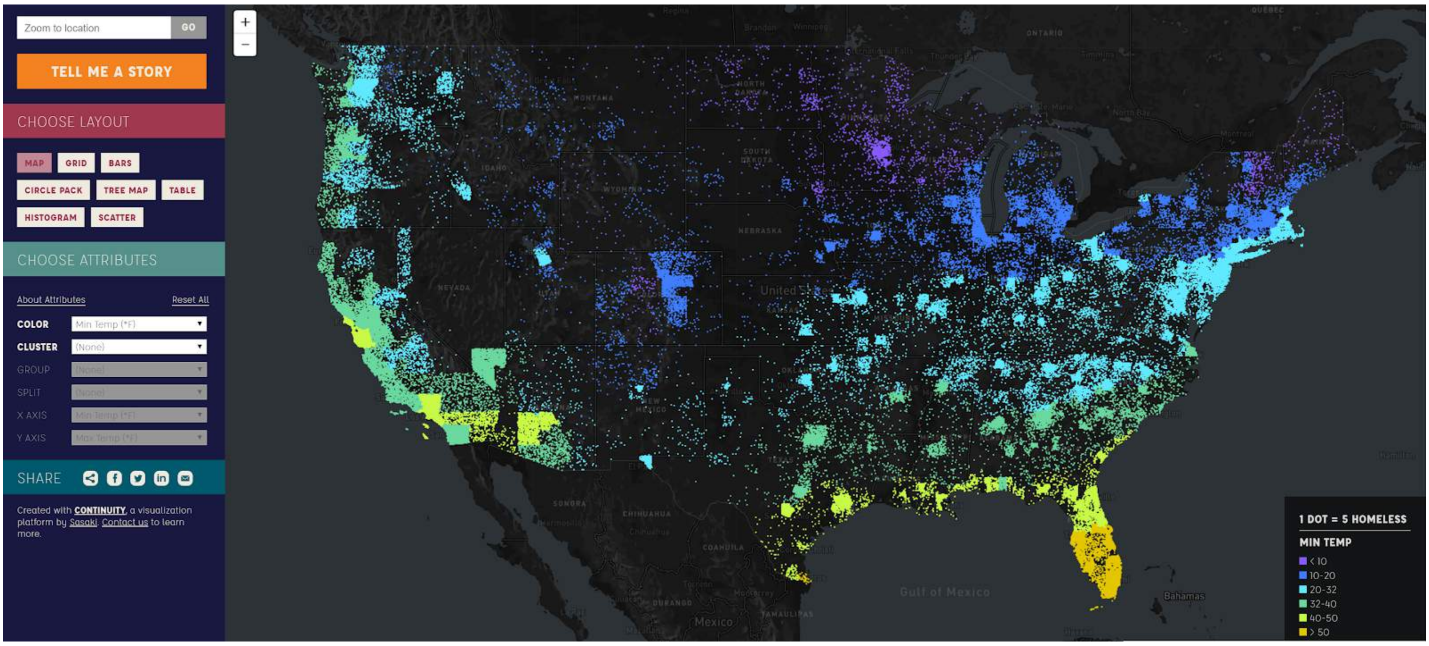
https://cpj.org/reports/2013/03/challenged-china-media-censorship-graphic-imprisoned.php

32. Deaths at the borders of southern Europe [University of Amsterdam]



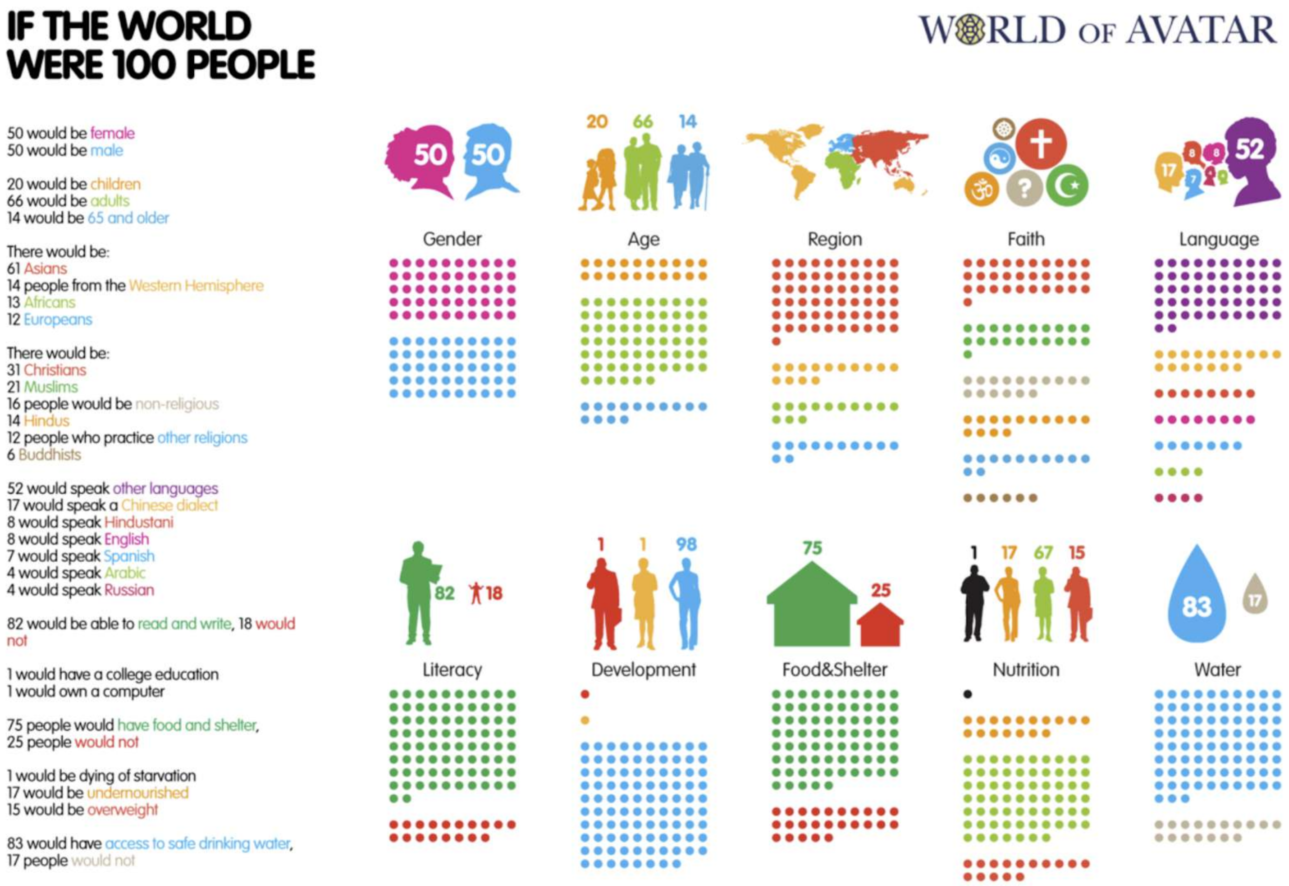
http://www.borderdeaths.org/

33. Understanding homelessness in USA [Continuity]



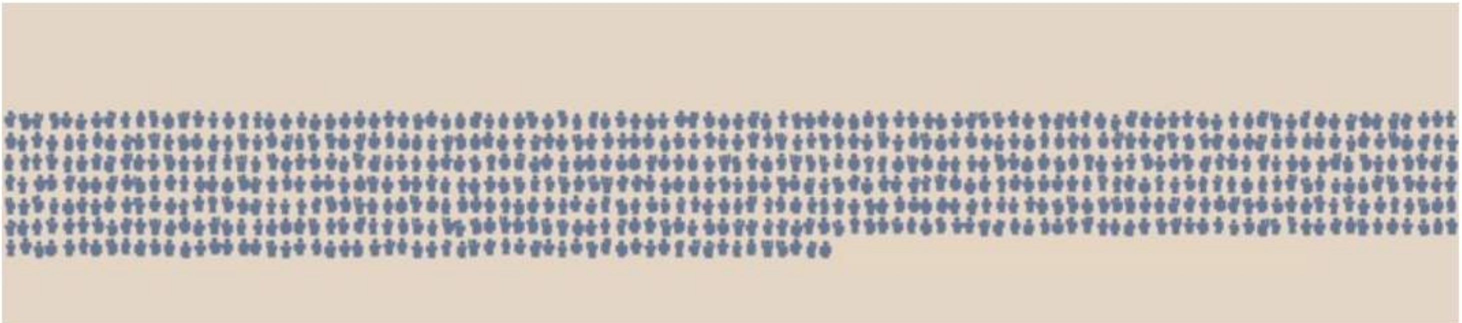
http://www.understandhomelessness.com/explore/

34. If the world were 100 people [Unknown]



http://afrographique.tumblr.com/image/5892814932

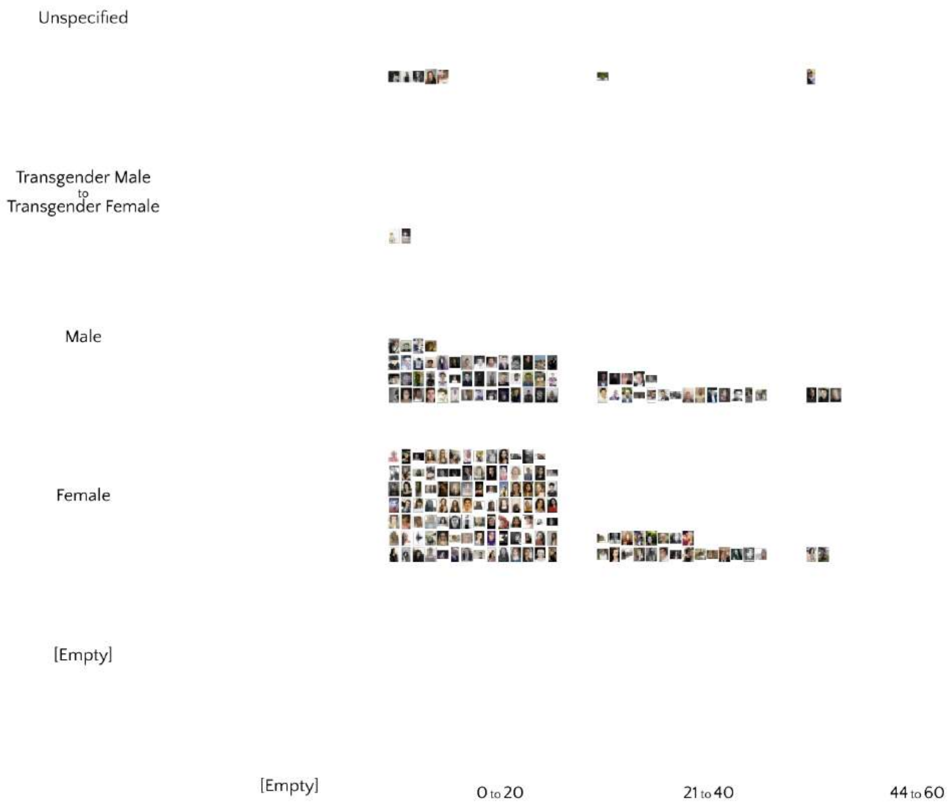
35. Risk, cycling and denominator neglect [giCentre]



657 serious injuries to people riding bicycles in London in 2012

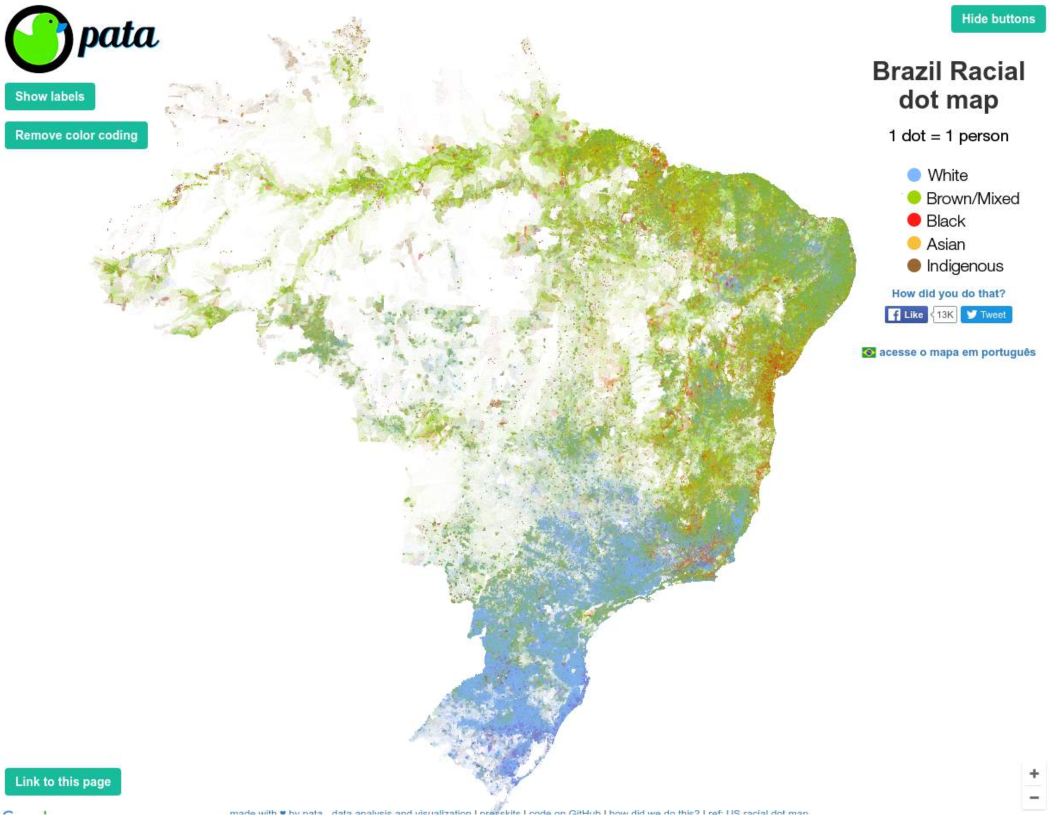
https://www.gicentre.net/blog/2013/11/24/risk-cycling-and-denominator-neglect

36. Acting talent [Zegami]



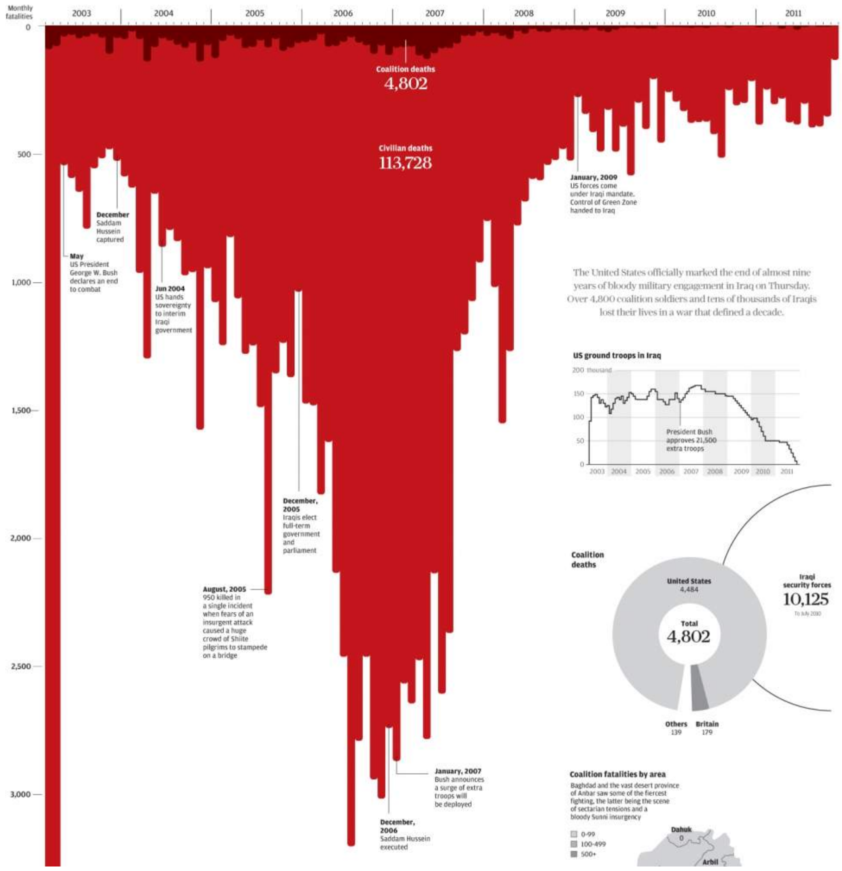
https://zegami.com/collections/public-5b98dc280d316a0001863ccc

37. Brazil Racial dot map [pata]



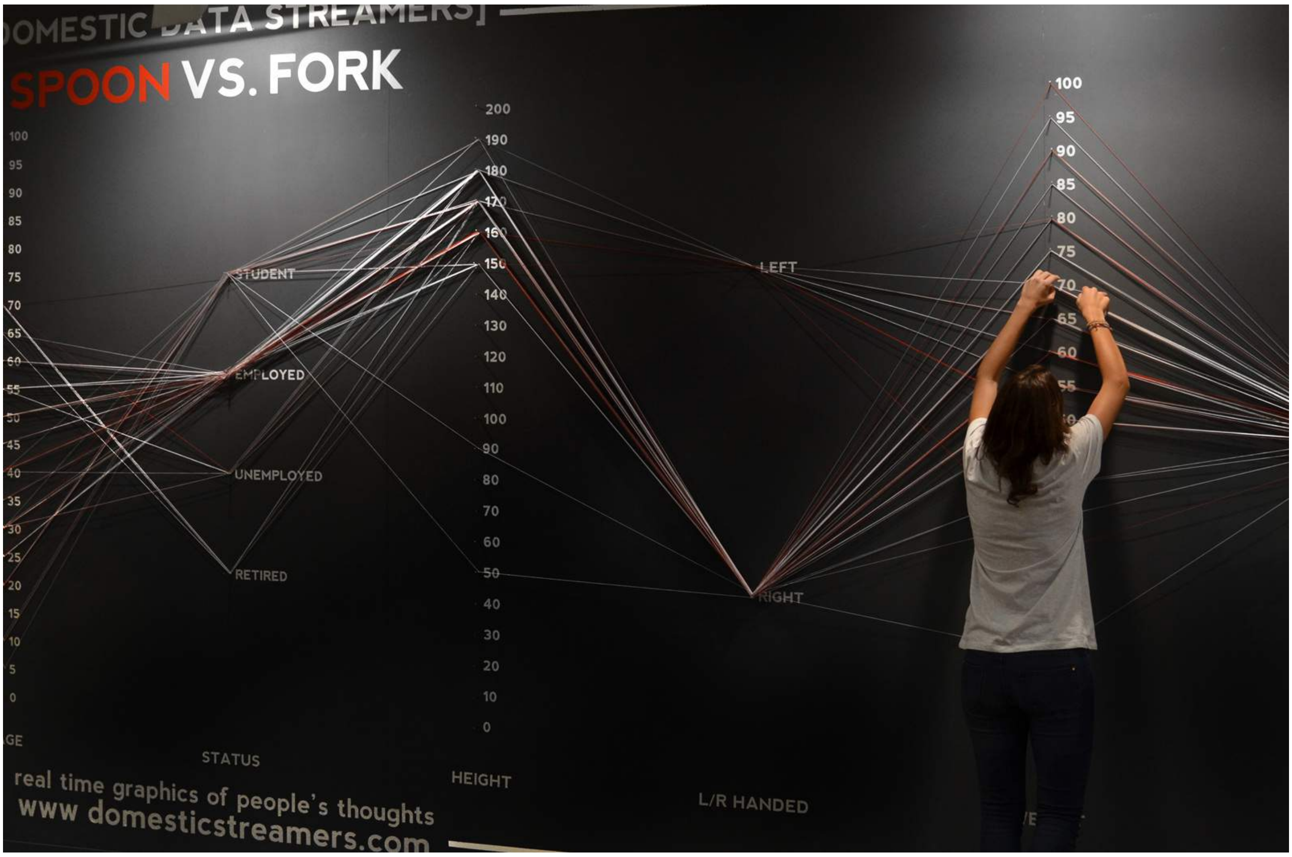
http://patadata.org/maparacial/en.html

38. Iraq's bloody toll [Scarr]



https://www.scmp.com/infographics/article/1284683/iraqs-bloody-toll

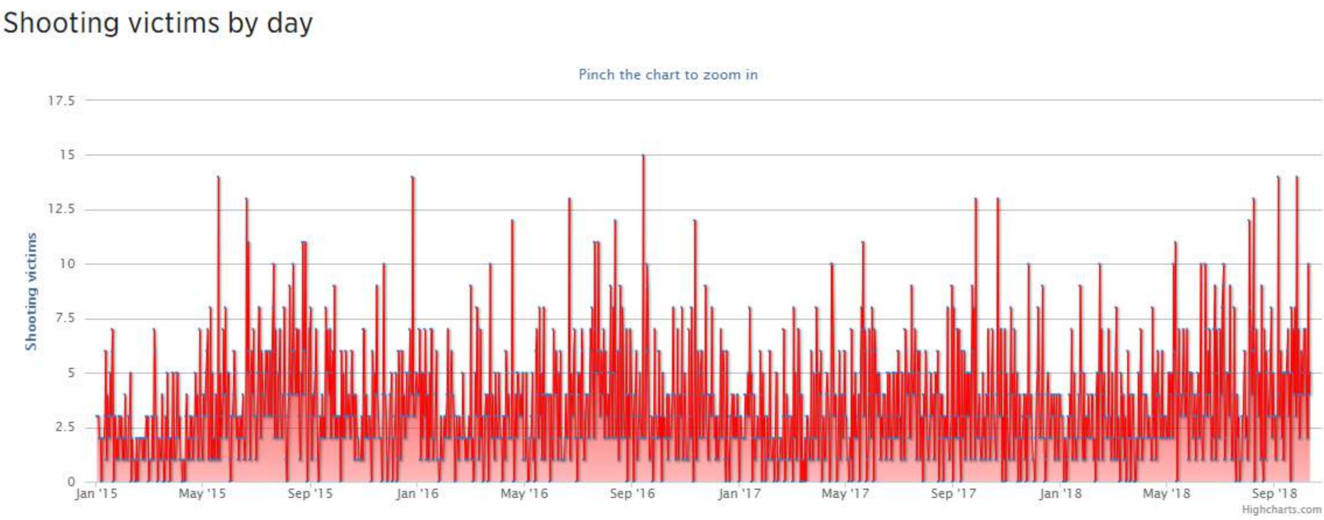
39. Data strings [Domestic Data Streamers]



http://dataphys.org/list/data-strings-physical-parallel-coordinates/

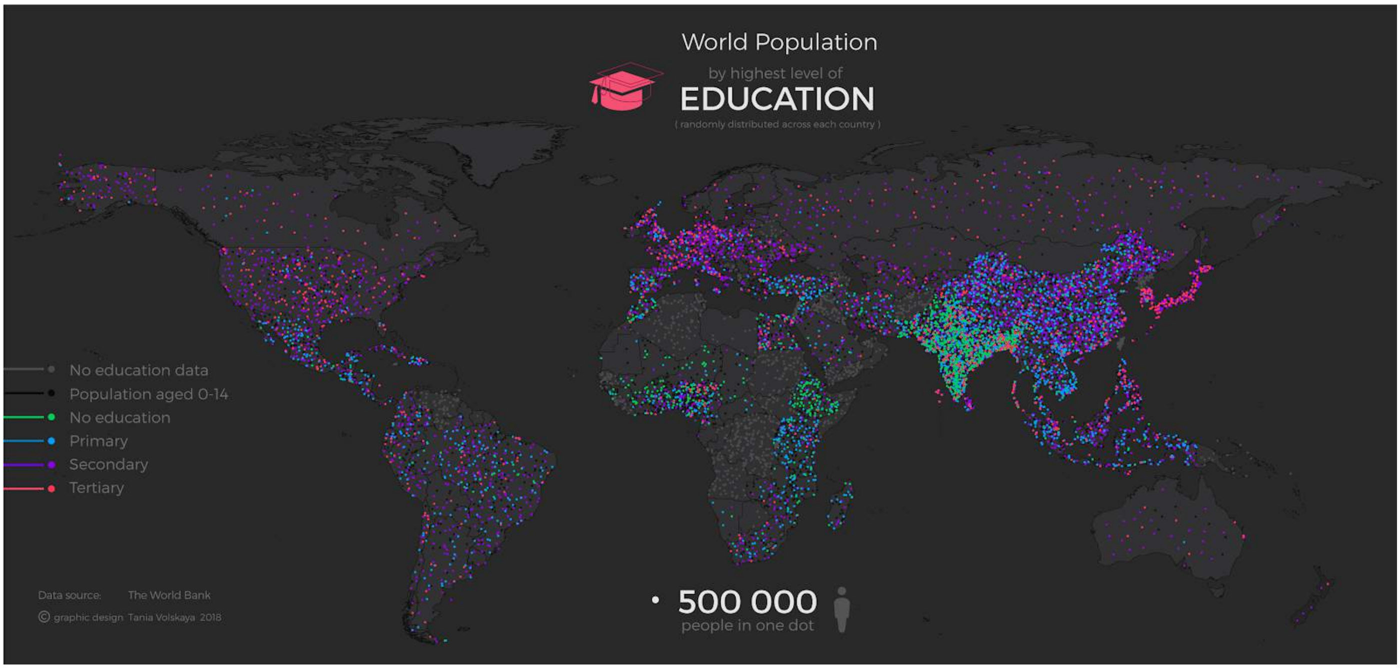
http://domesticstreamers.com/2014/project/data-strings/ (this is the real link)

40. Philadelphia Shooting Victims [Philly.com]



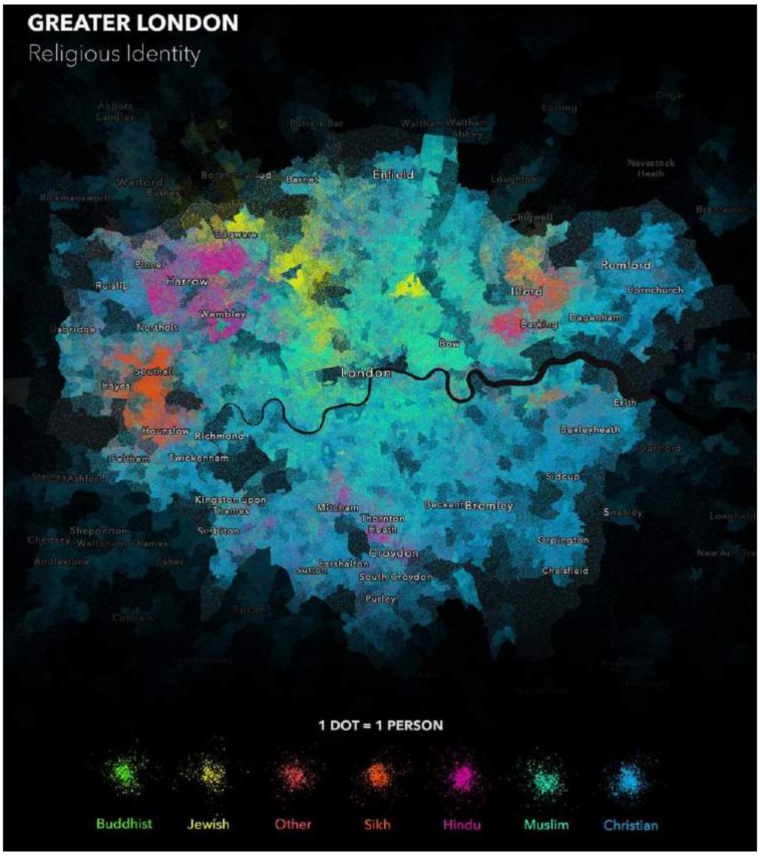
http://data.philly.com/philly/crime/shootings/

41. World Education by highest level of education [Volskaya]



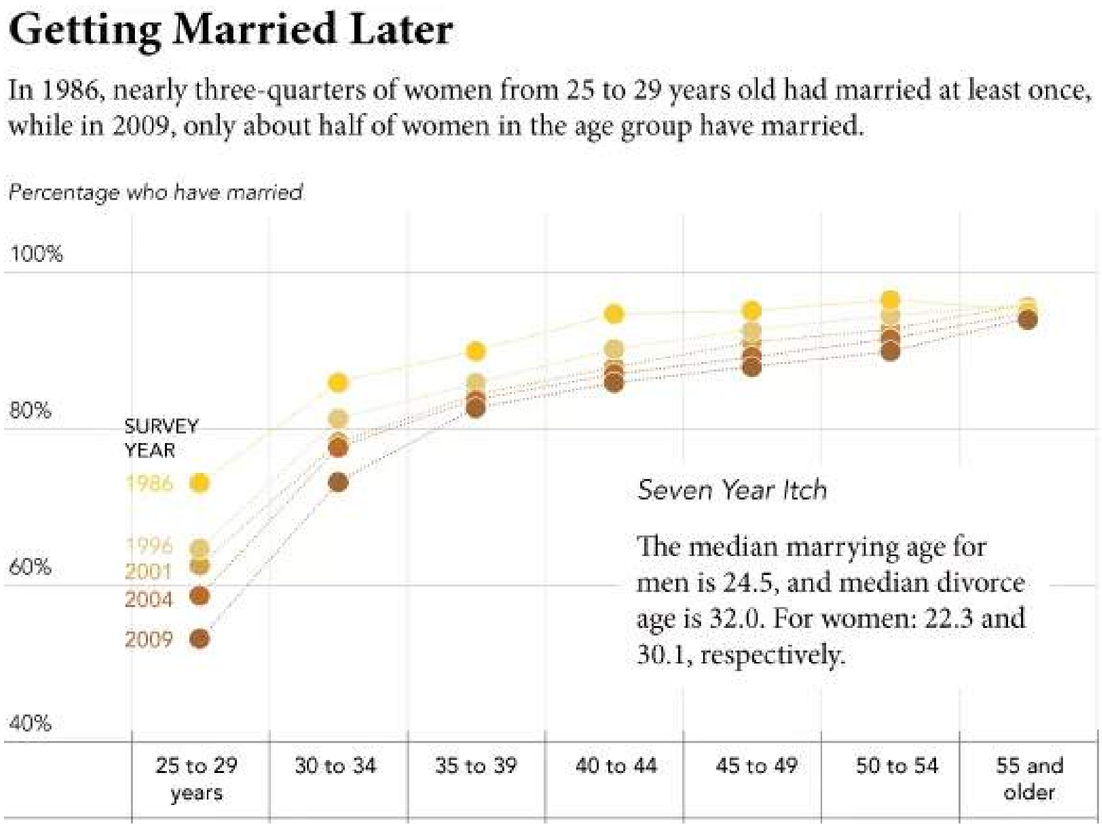
https://www.reddit.com/r/dataisbeautiful/comments/9mdk6w/world_population_by_highest_level_of_education_oc/

42. Religious Identity [Flanagan]



https://www.benflanagan.co.uk/#/religious-identity/

43. Seven year itch: When do people get married and divorced? [Yau]

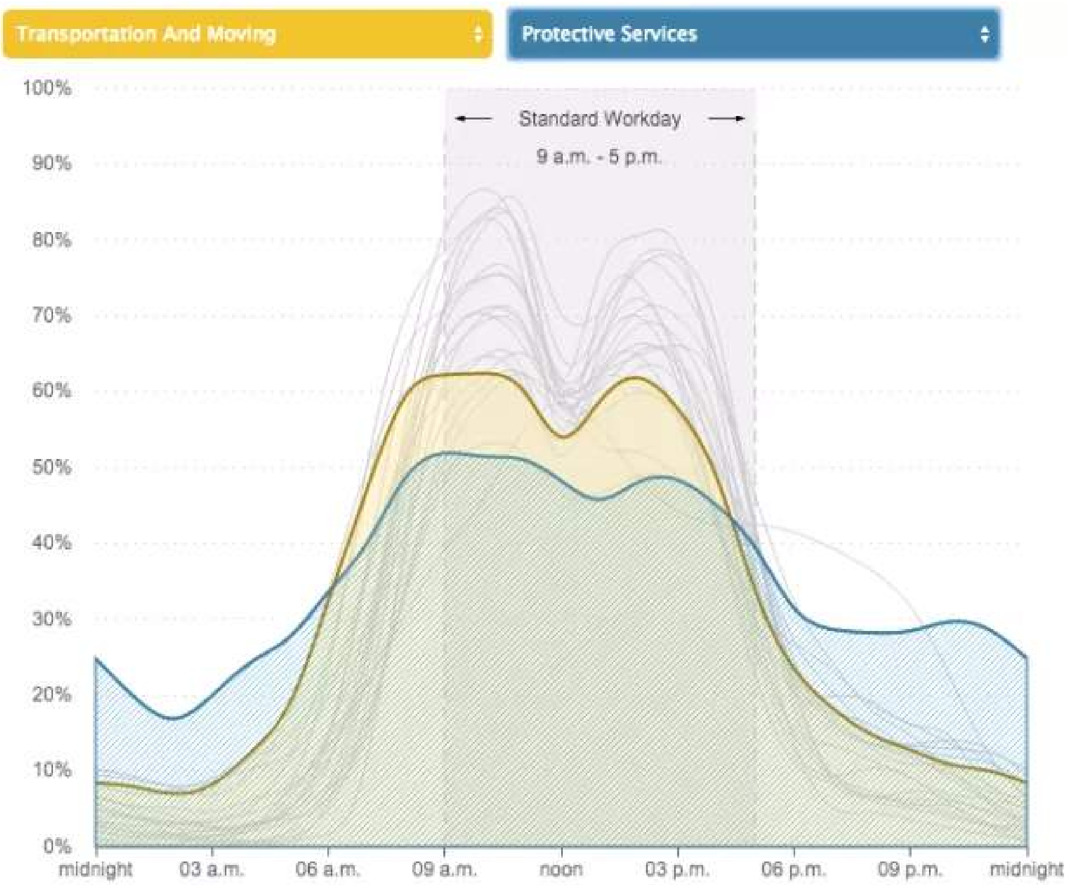


https://flowingdata.com/2011/05/23/when-do-people-get-married-and-divorced/

44. What made me [Grabkowska and Kolec]

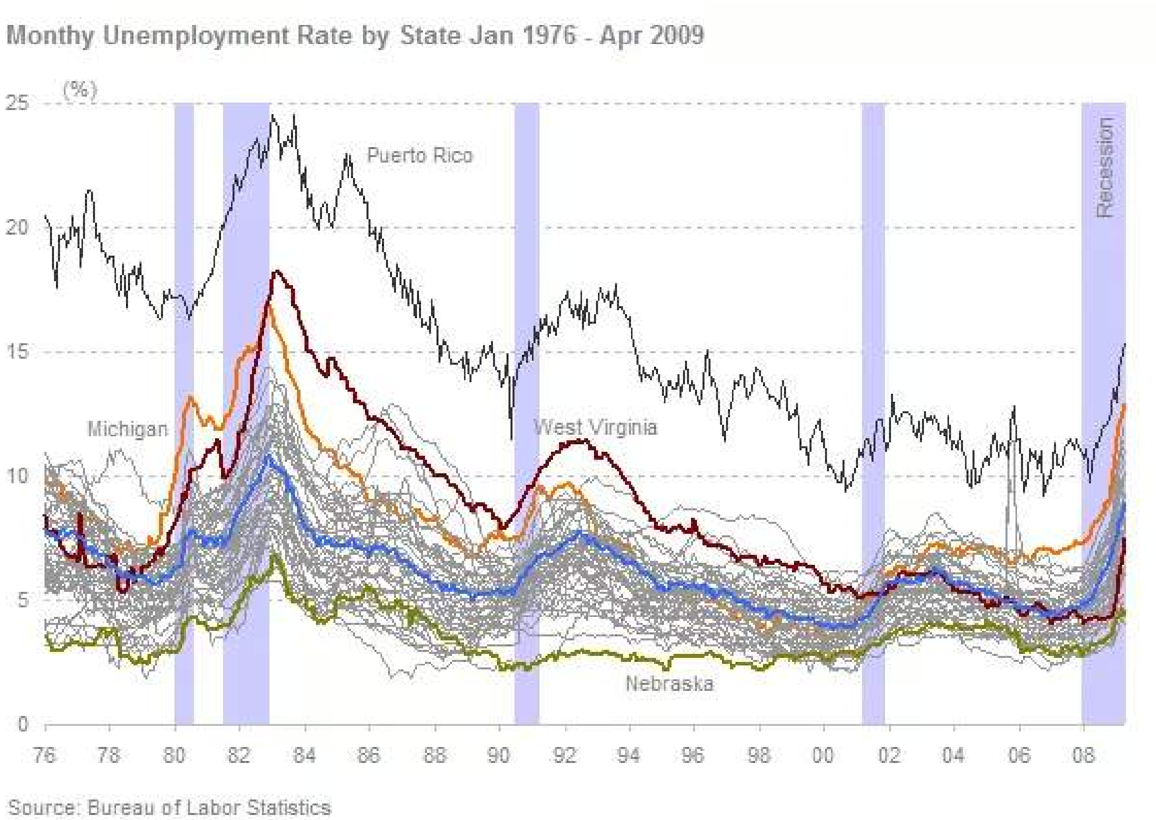
http://dataphys.org/list/what-made-me-interactive-public-installation/

45. When people work? [NPR]



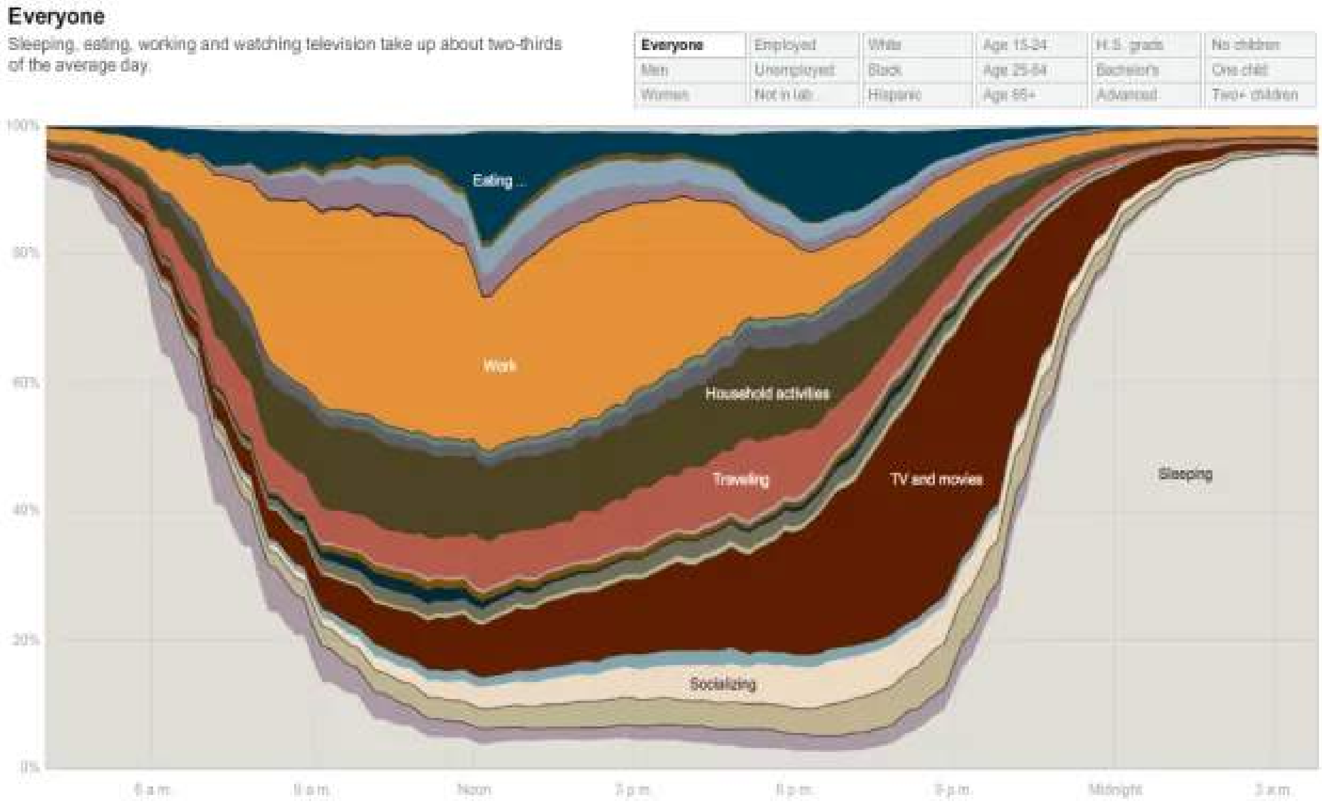
https://www.npr.org/sections/money/2014/08/27/343415569/whos-in-the-office-the-american-workday-in-one-graph

46. Monthly unemployment rates by state [Camoes]



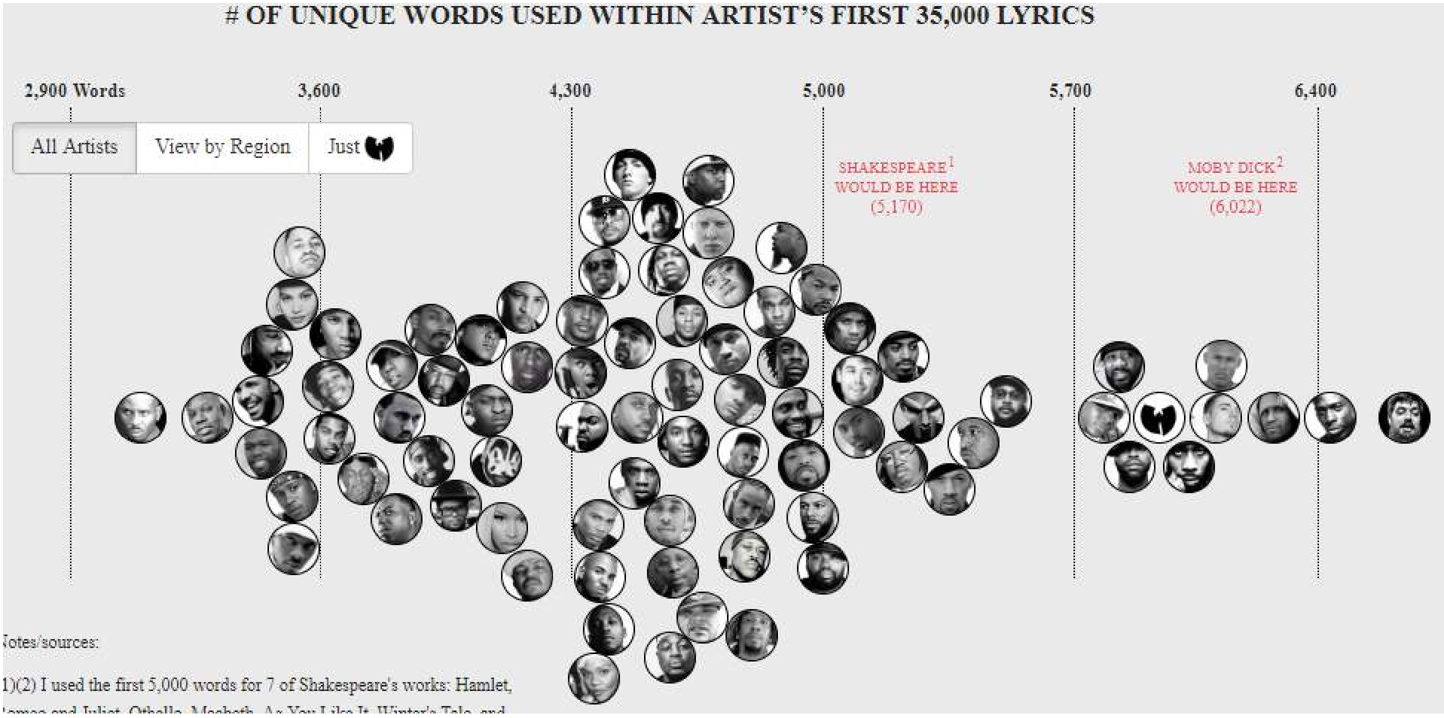
https://excelcharts.com/charts-monthly-unemployment-rates-by-state-1976-2009/

47. How people in America spend their day [Carter, Cox, Quealy, and Schoenfeld]



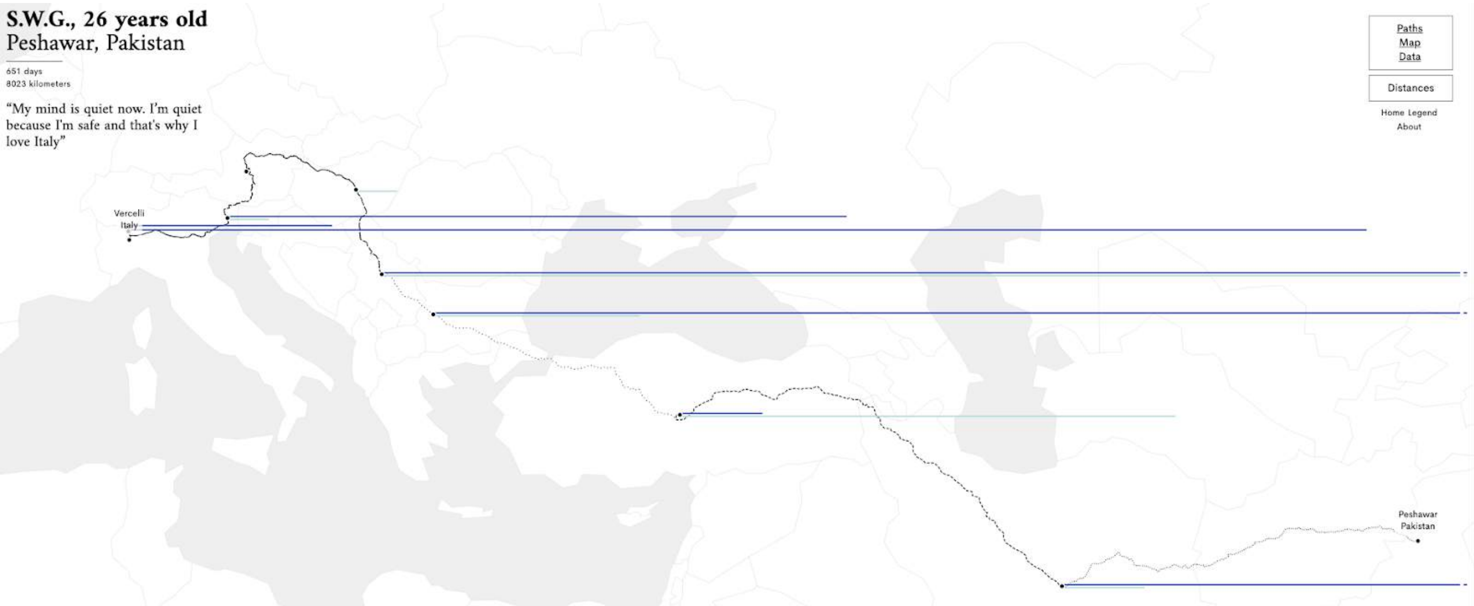
https://flowingdata.com/2009/08/10/how-people-in-america-spend-their-day/

48. Number of unique words used within artist's first 35,000 lyrics [Daniels]



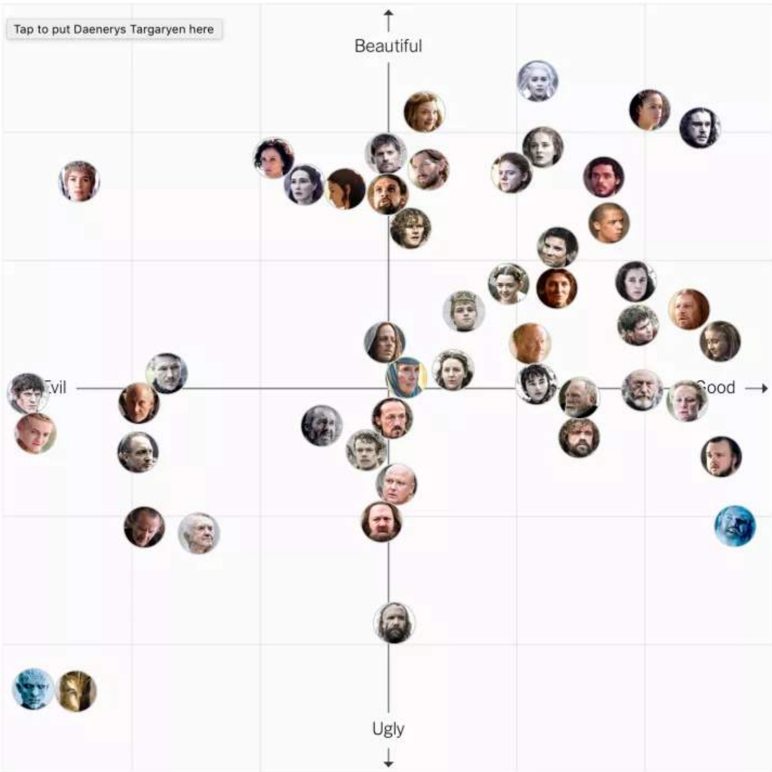
https://pudding.cool/2017/02/vocabulary/

49. The stories behind a line [Fragapane and Piacentini]



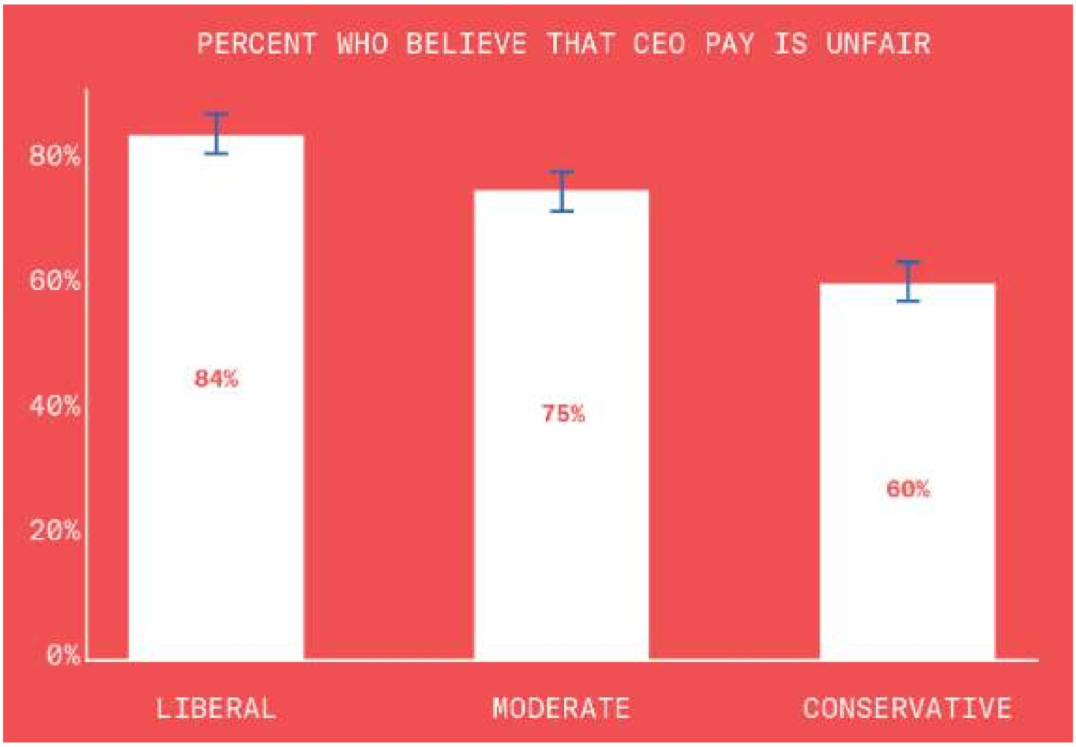
http://www.storiesbehindaline.com/

50. Game of thrones character chart [Flowing data]



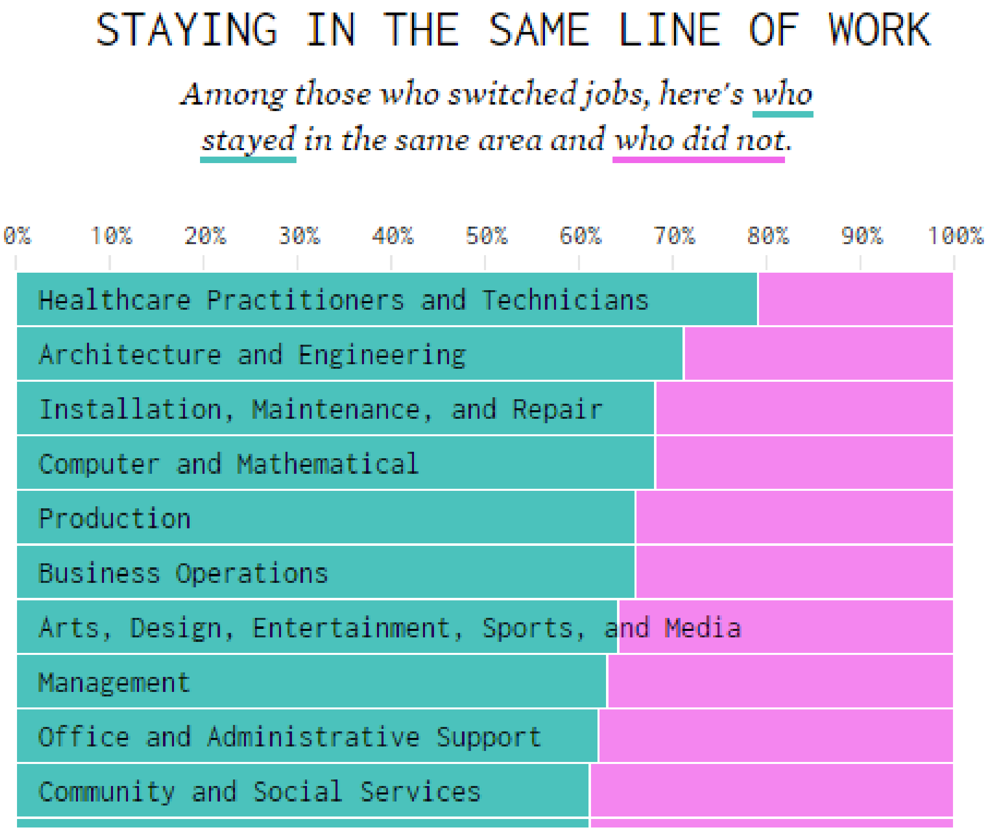
https://flowingdata.com/2017/08/13/game-of-thrones-character-chart-you-decide/

51. Fairness [The Pulse of the Nation]



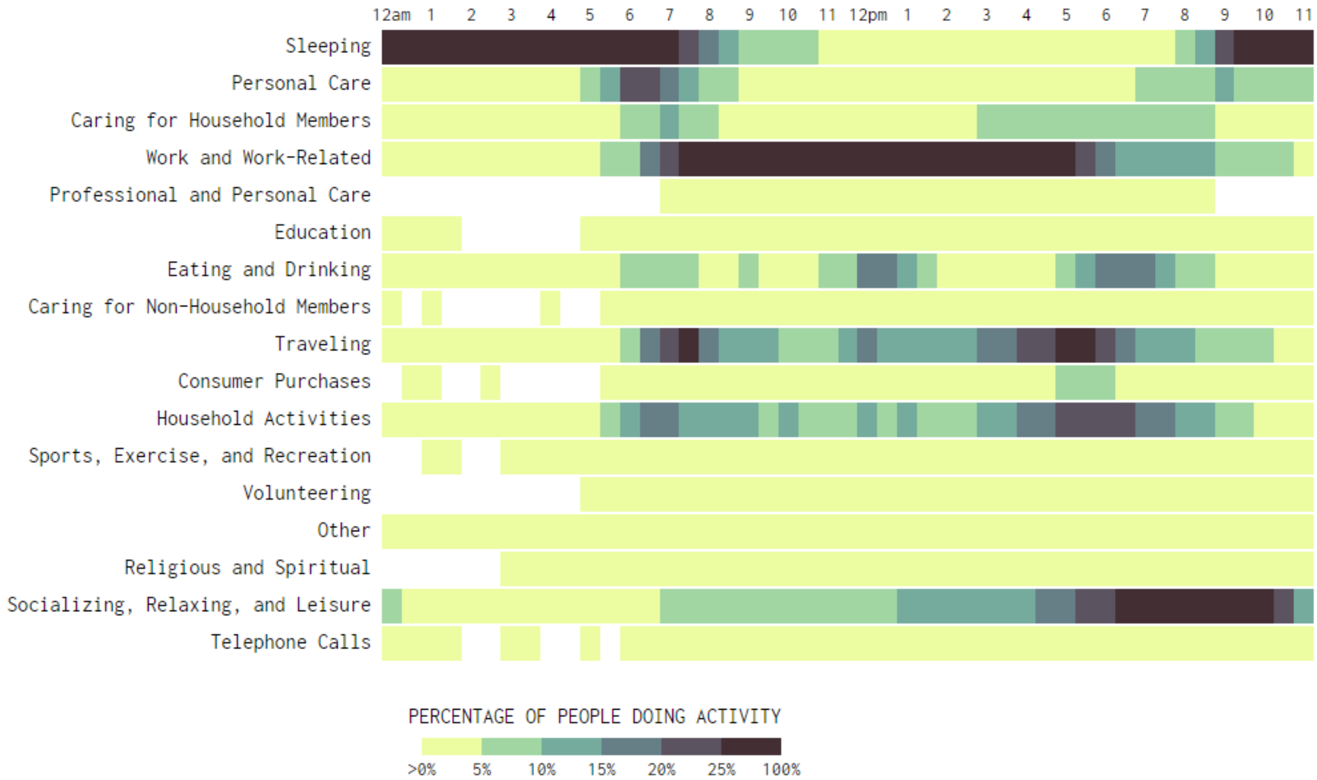
https://thepulseofthenation.com/#fairness

52. Staying in the same line of work [Yau]



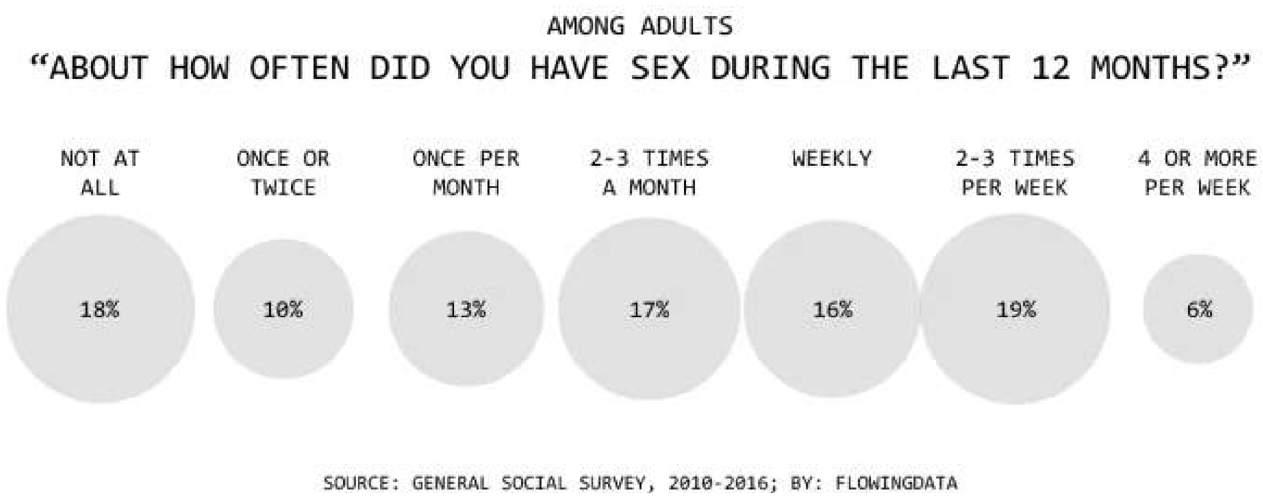
https://flowingdata.com/2017/11/16/switching-jobs/

53. American Daily Routine [Yau]



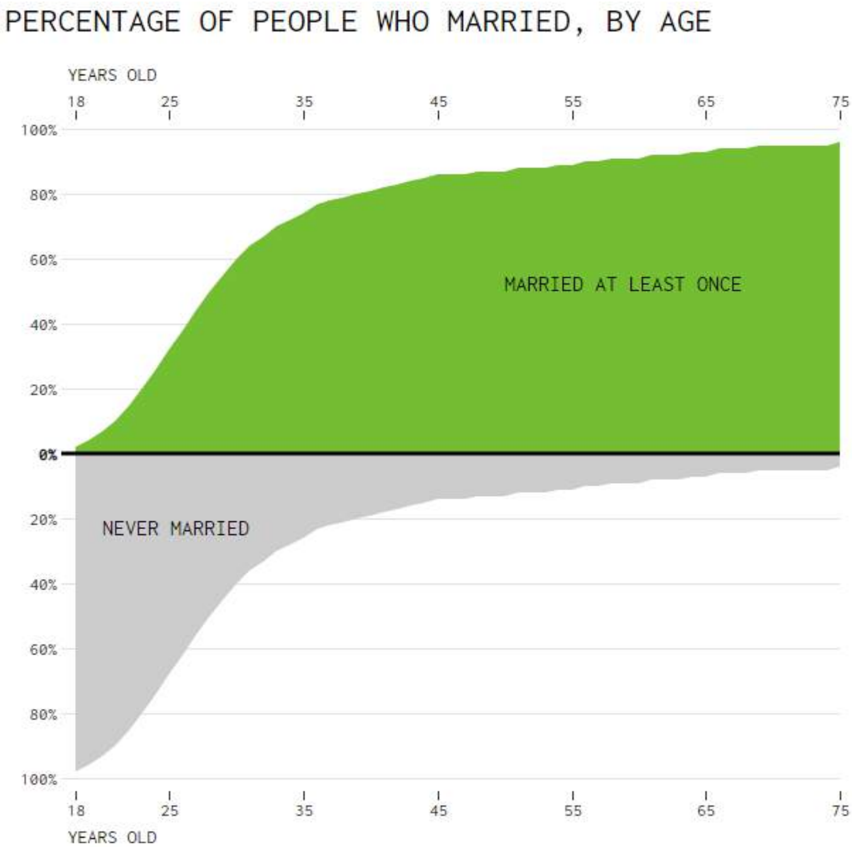
https://flowingdata.com/2017/10/19/american-daily-routine/

54. Married People Have More Sex [Yau]



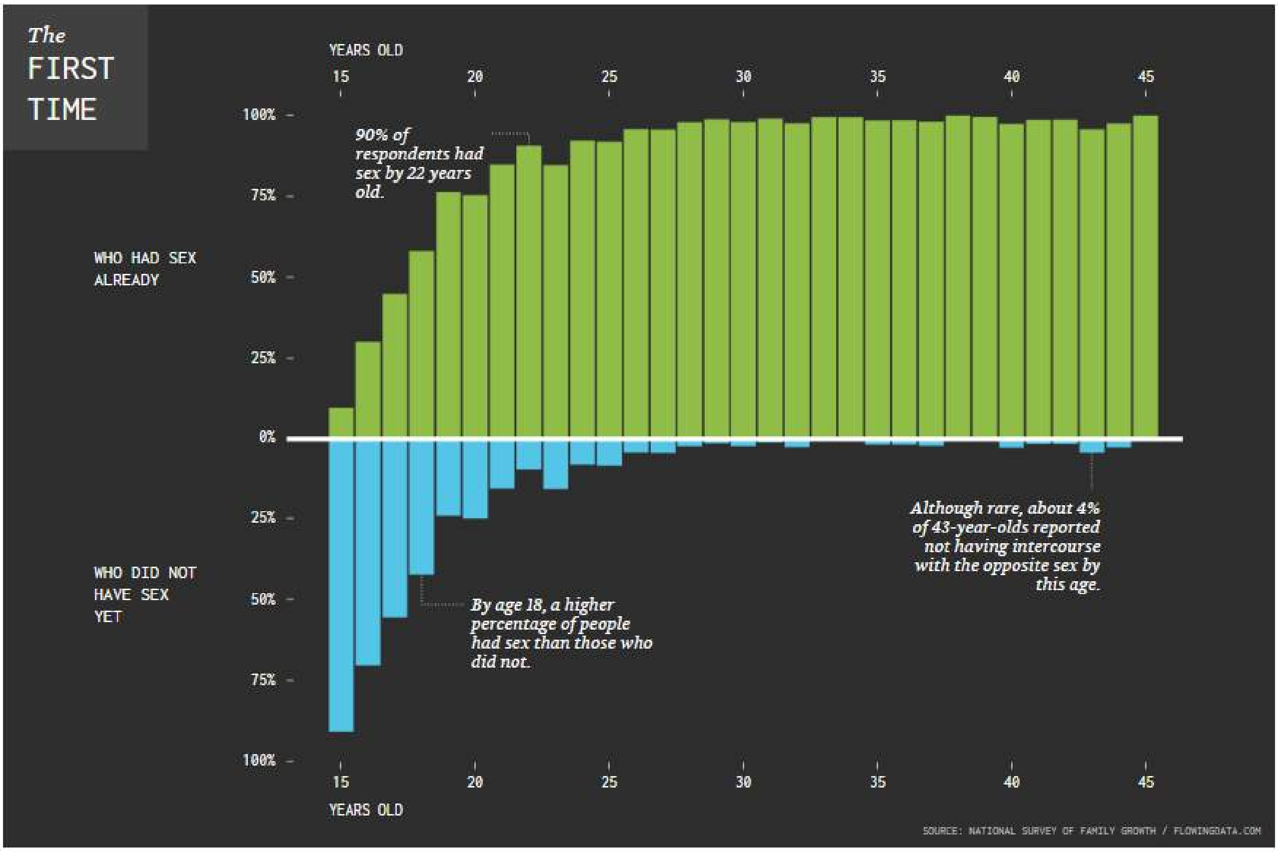
https://flowingdata.com/2017/07/03/married-people-sex/

55. Percentage of people who married by age [Yau]



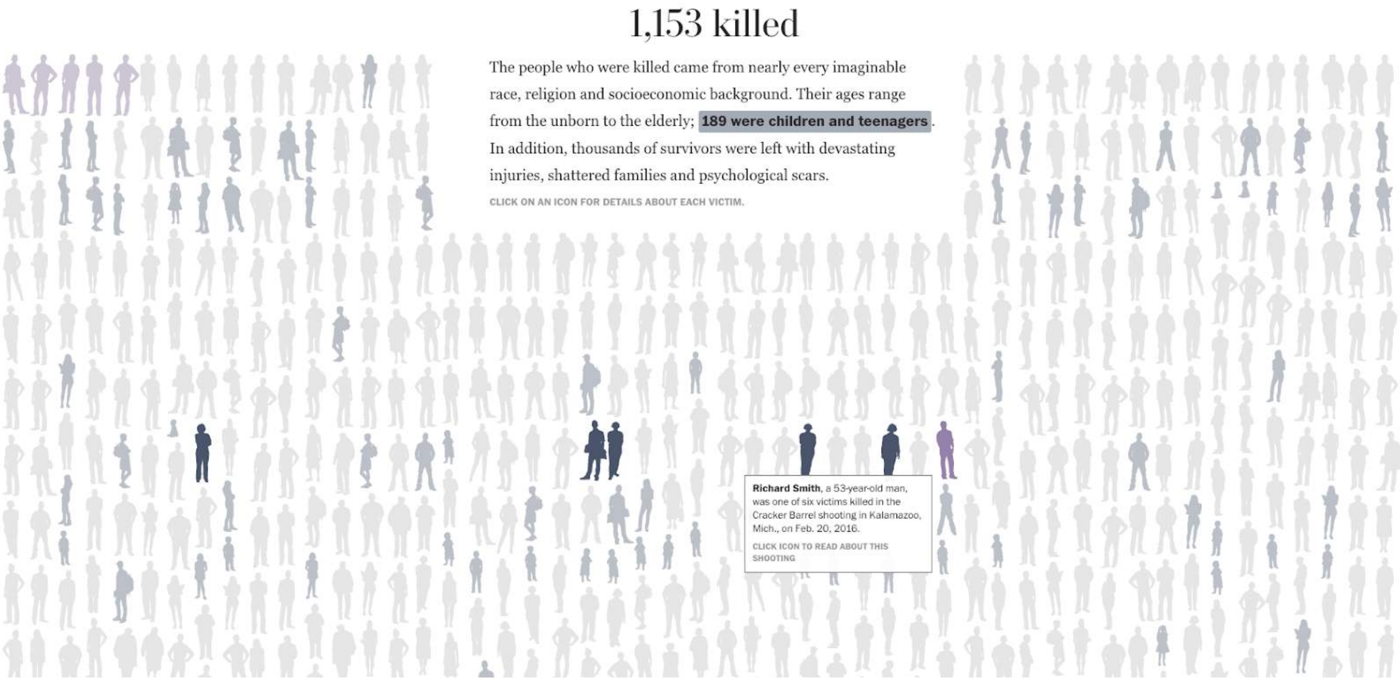
https://flowingdata.com/2017/11/01/who-is-married-by-now/

56. The first time [Yau]



https://flowingdata.com/2017/03/17/when-americans-lost-their-virginity/

57. Mass Shooting Statistics in the United States [WP]

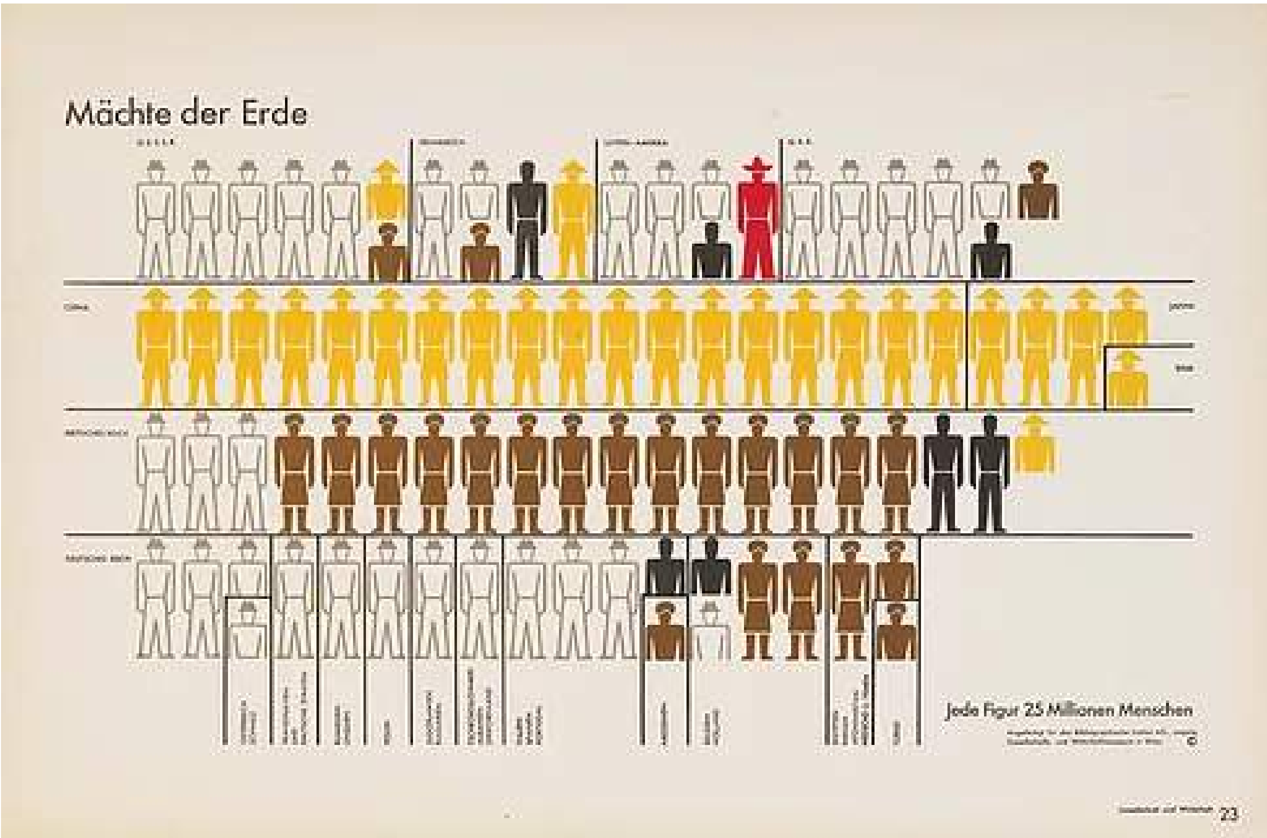


https://www.washingtonpost.com/graphics/2018/national/mass-shootings-in-america/?noredirect=on&utm_term=.de323e91d160

58. Mazamet Ville Morte [DataPhys]

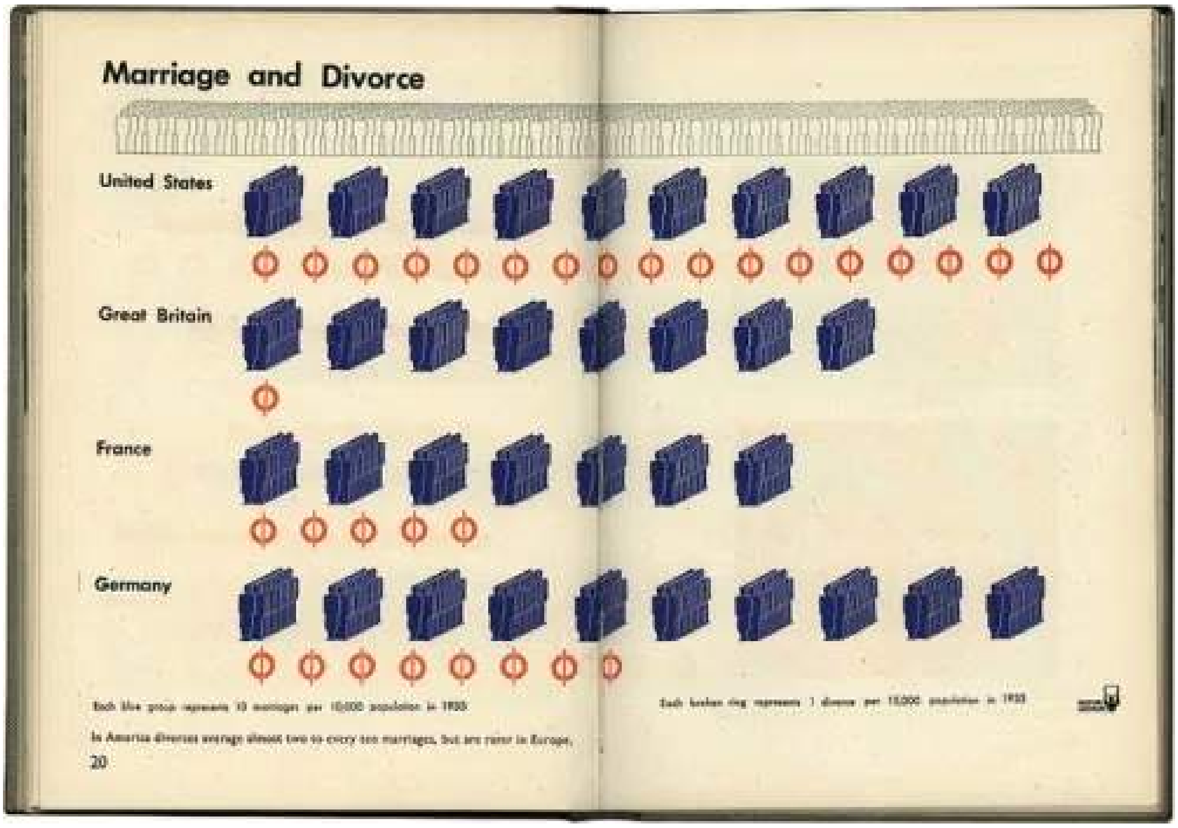
http://dataphys.org/list/mazamet-ville-morte/

59. USA and Great Britain in the world [Nashabolovke Gallery]



https://www.nashabolovke-gallery.com/izosta-eng

60. Mariage and Divorce [Brain Pickings]



https://www.brainpickings.org/2012/11/13/only-an-ocean-between-isotype-infographics/

61. Great war [Worrell]

http://www.laworrell.com/blog/2015/1/14/isotype-international-system-of-typographic-picture-education

62. Ceramic Poppies to Commemorate Fallen Soldiers in WW1 [DataPhys]

http://dataphys.org/list/tower-poppies-888246-ceramic-poppies-to-commemorate-fallen-soldiers-in-ww1/

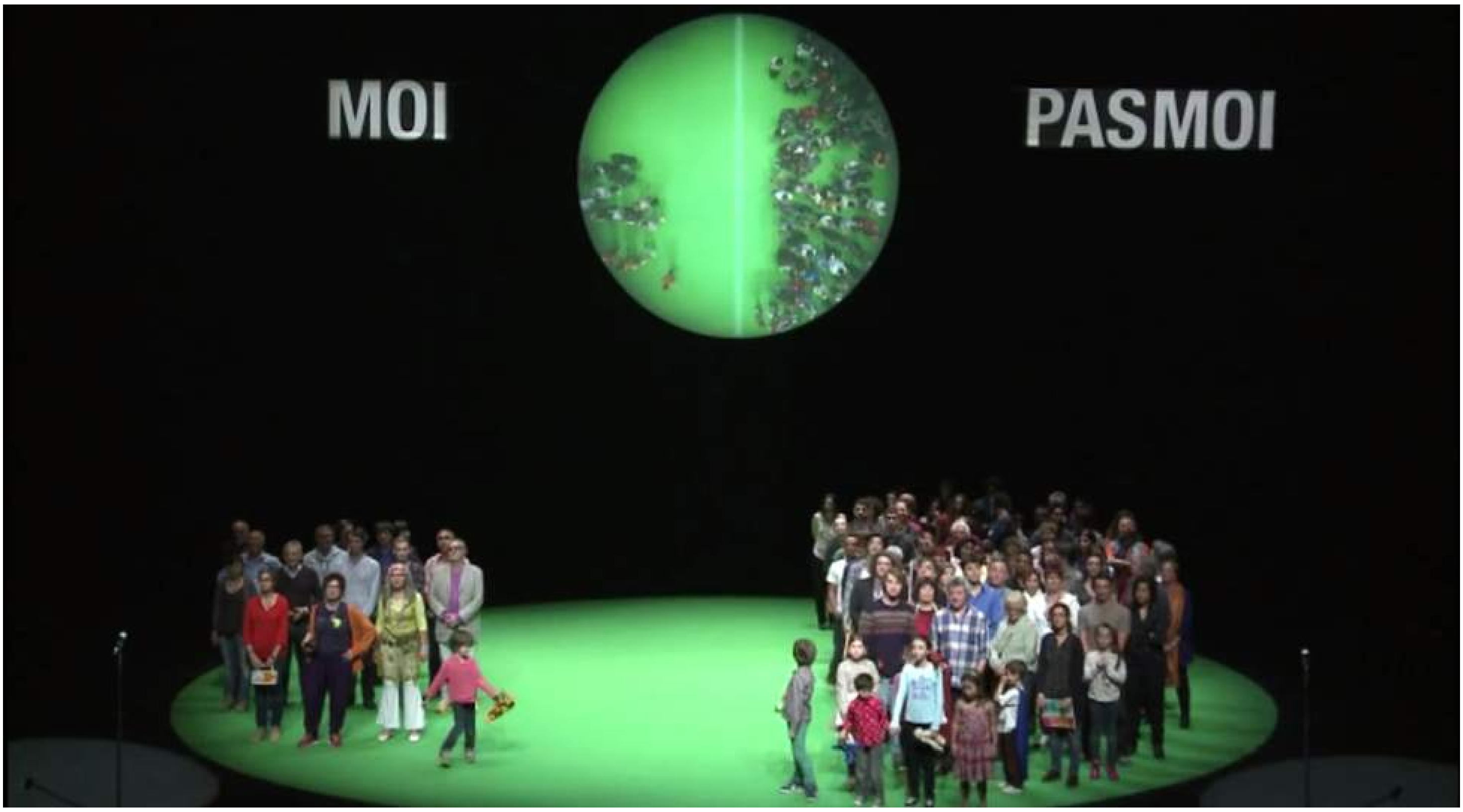
63. Of All the People in All the World: Stats with Rice [DataPhys]

http://dataphys.org/list/of-all-the-people-in-all-the-world-stats-with-rice/

64. Luxembourg American Cemetery [Unknown]

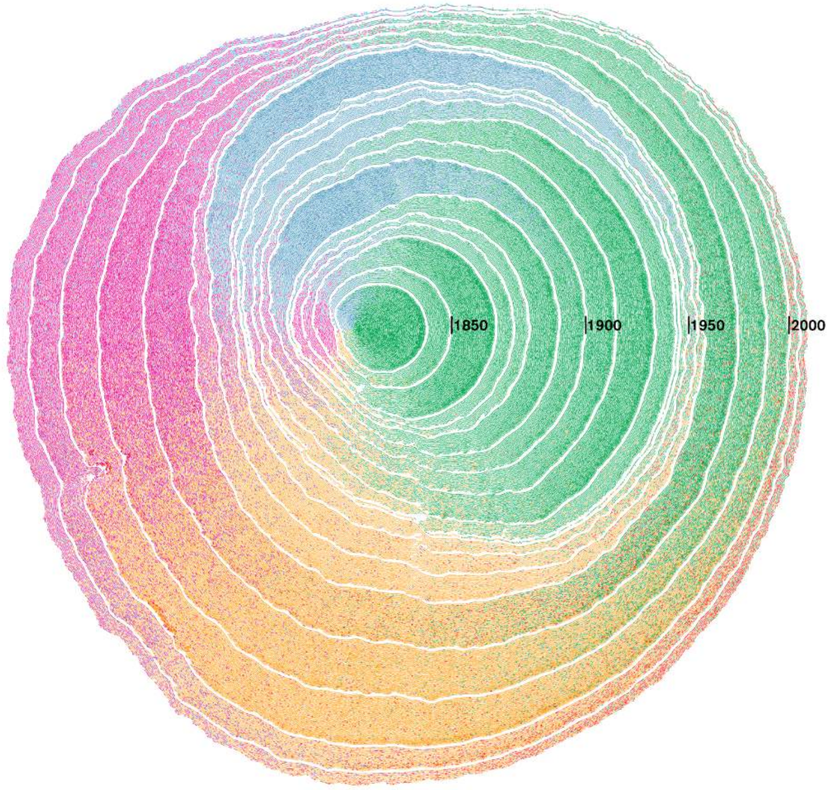
https://www.abmc.gov/cemeteries-memorials/europe/luxembourg-american-cemetery#.W8hMiGiuLIU

65. 100% [Haug, Kaegi, and Wetzel]



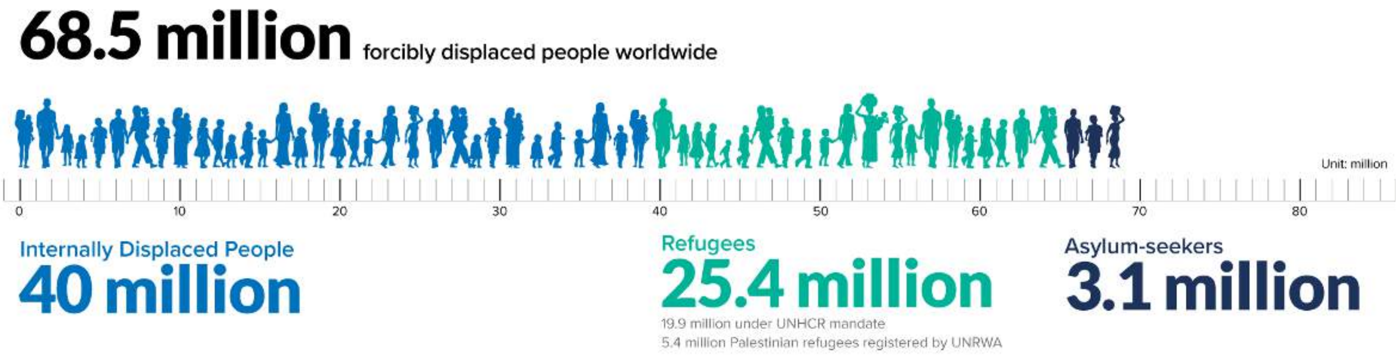
https://vimeo.com/114034410 (first vis starts at 24:00)

66. Simulated Dendrochronology of U.S. immigration [Cruz, Wihbey, Ghael, and Shibuya]



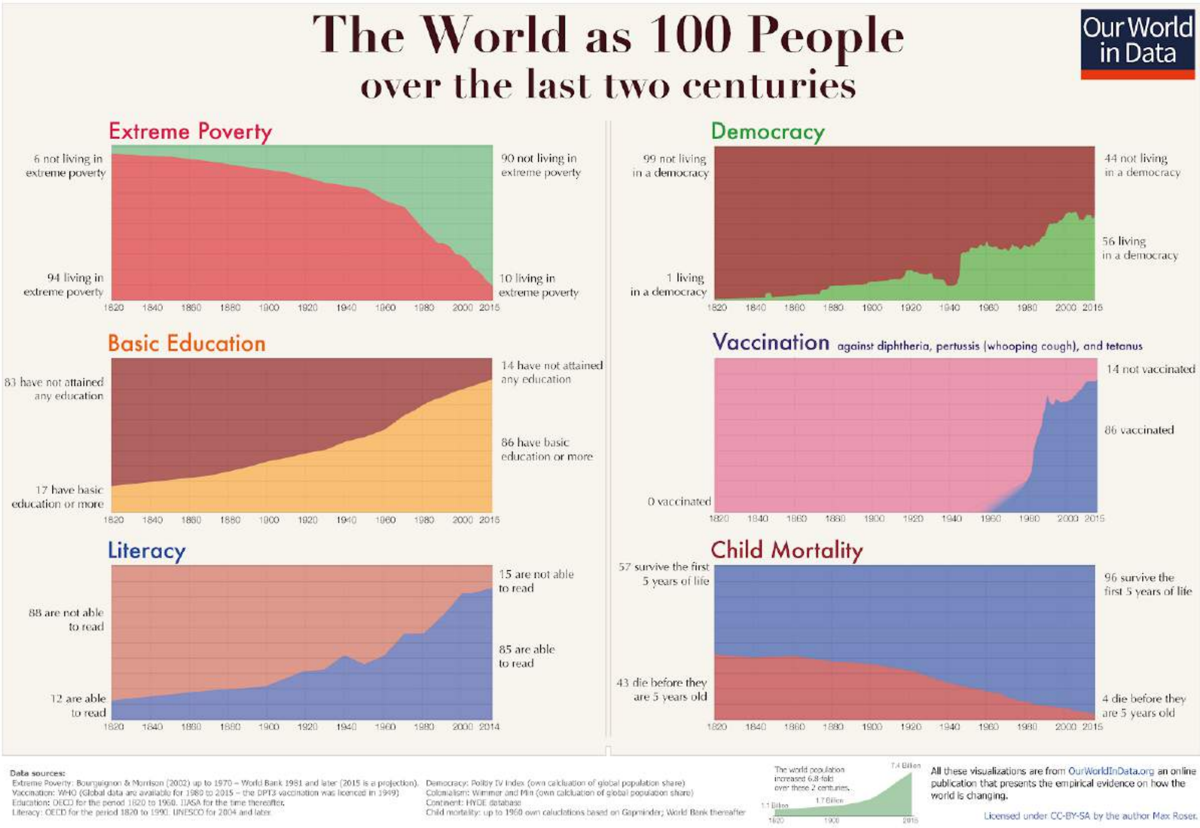
http://pmcruz.com/dendrochronology/

67. Displaced people worldwide [UNHCR]



https://data.humdata.org/organization/unhcr

68. The world as 100 people, in 200 years of history [Our World in Data]

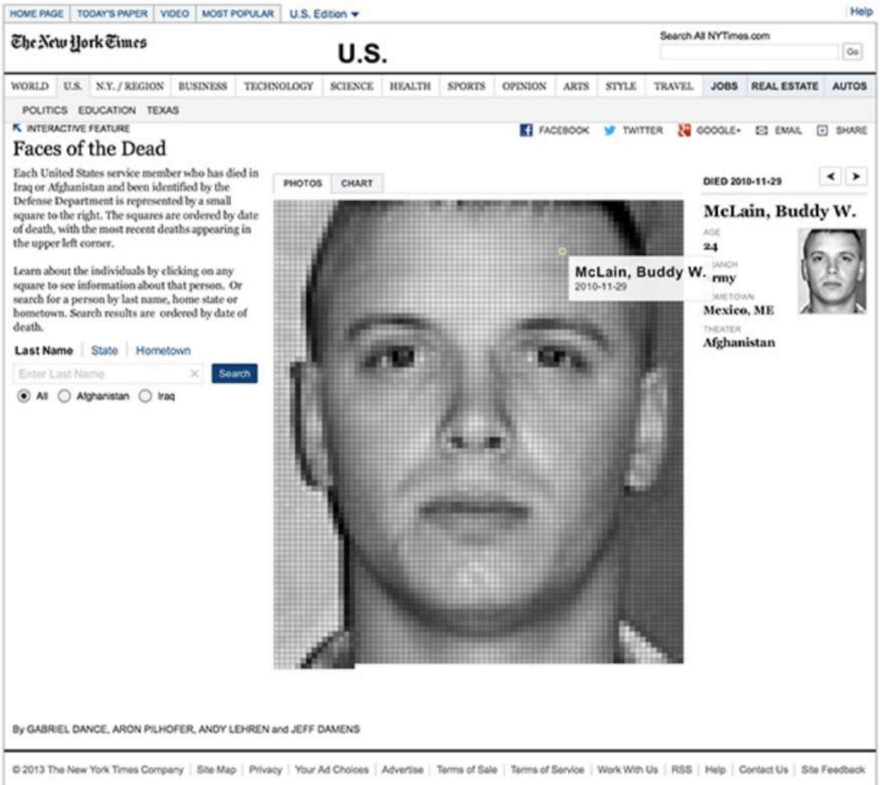


https://www.weforum.org/agenda/2017/01/ricardo-hausmann-why-governments-act-on-their-lies

69. Death in Syria [NYT]

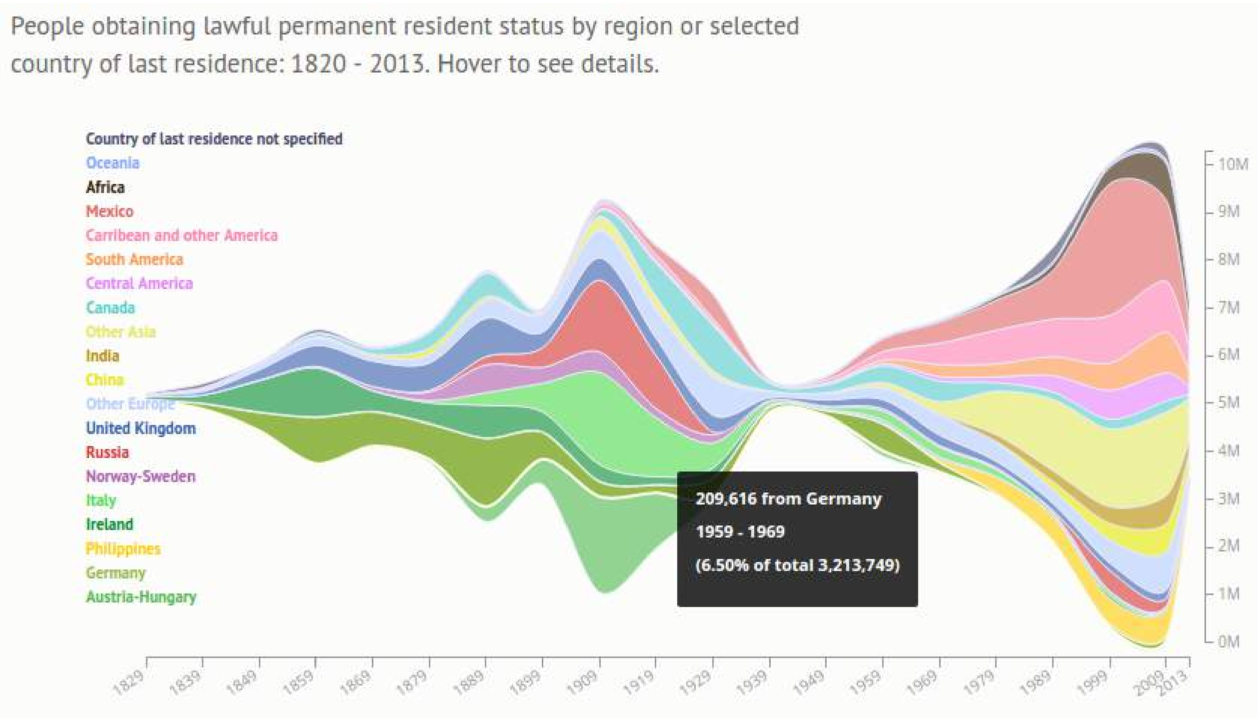
http://www.nytimes.com/interactive/2015/09/14/world/middleeast/syria-war-deaths.html

70. Faces of the Dead [NYT]



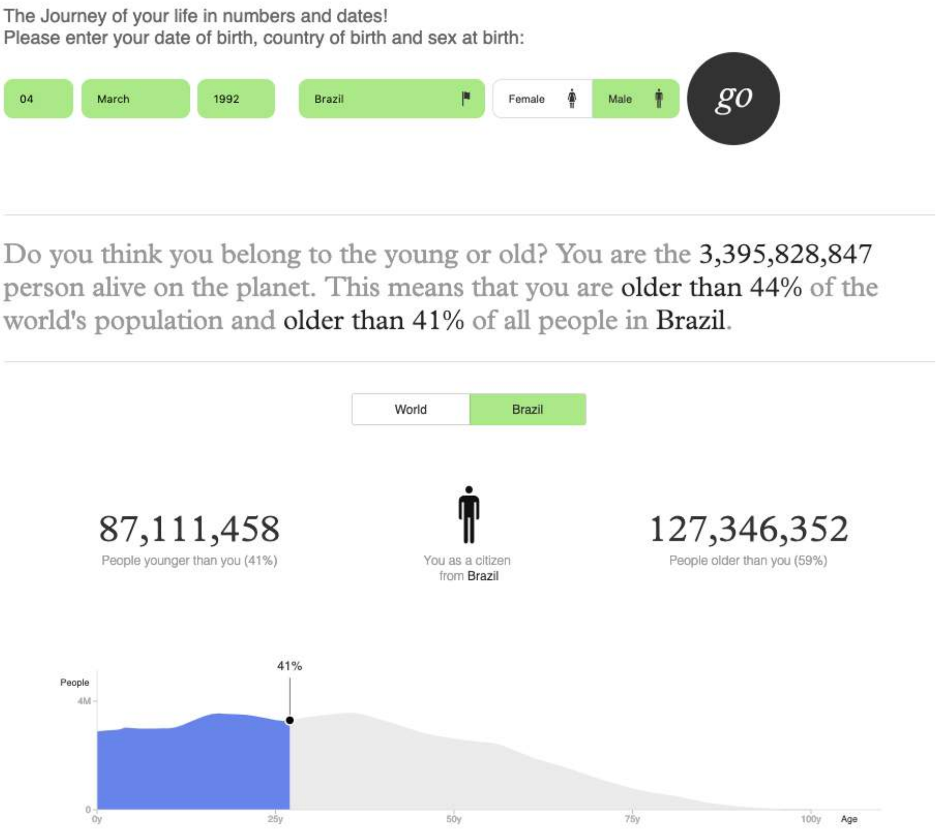
http://www.nytimes.com/interactive/us/faces-of-the-dead.html#/mclain_buddy_w

71. 200 Years of Immigration to the U.S. [Bronshtein]



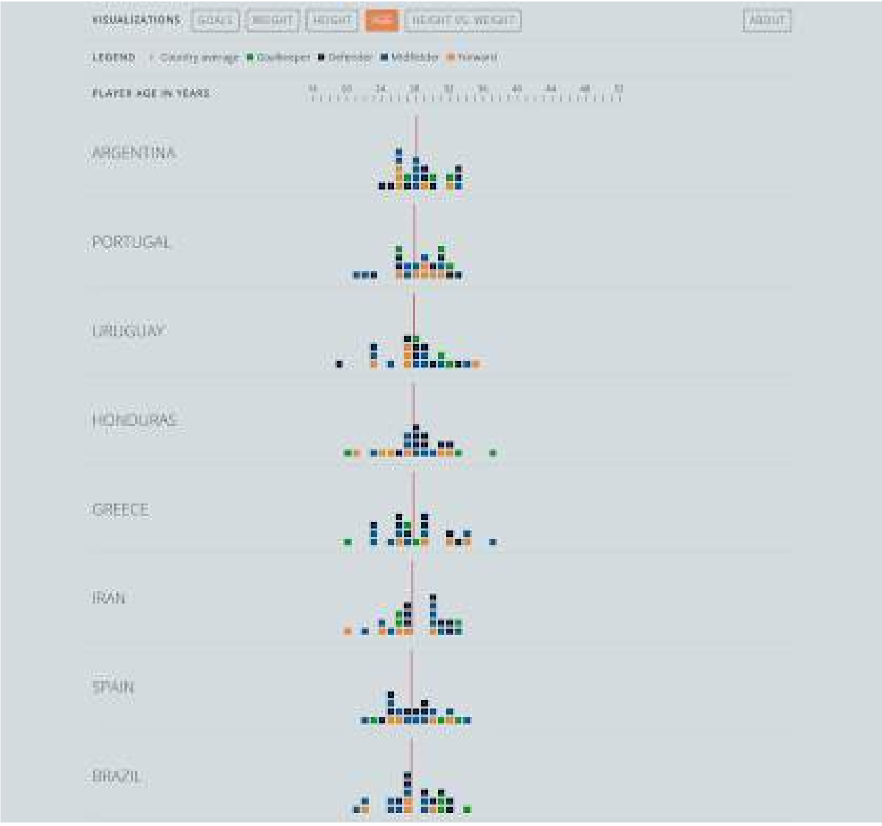
http://insightfulinteraction.com/immigration200years.html

72. What's my place in the world population? [Groß, Samir and Fengler]



http://population.io/

73. Visualization of age, height and weight of players in the 2014 FIFA World Cup [Johansson]

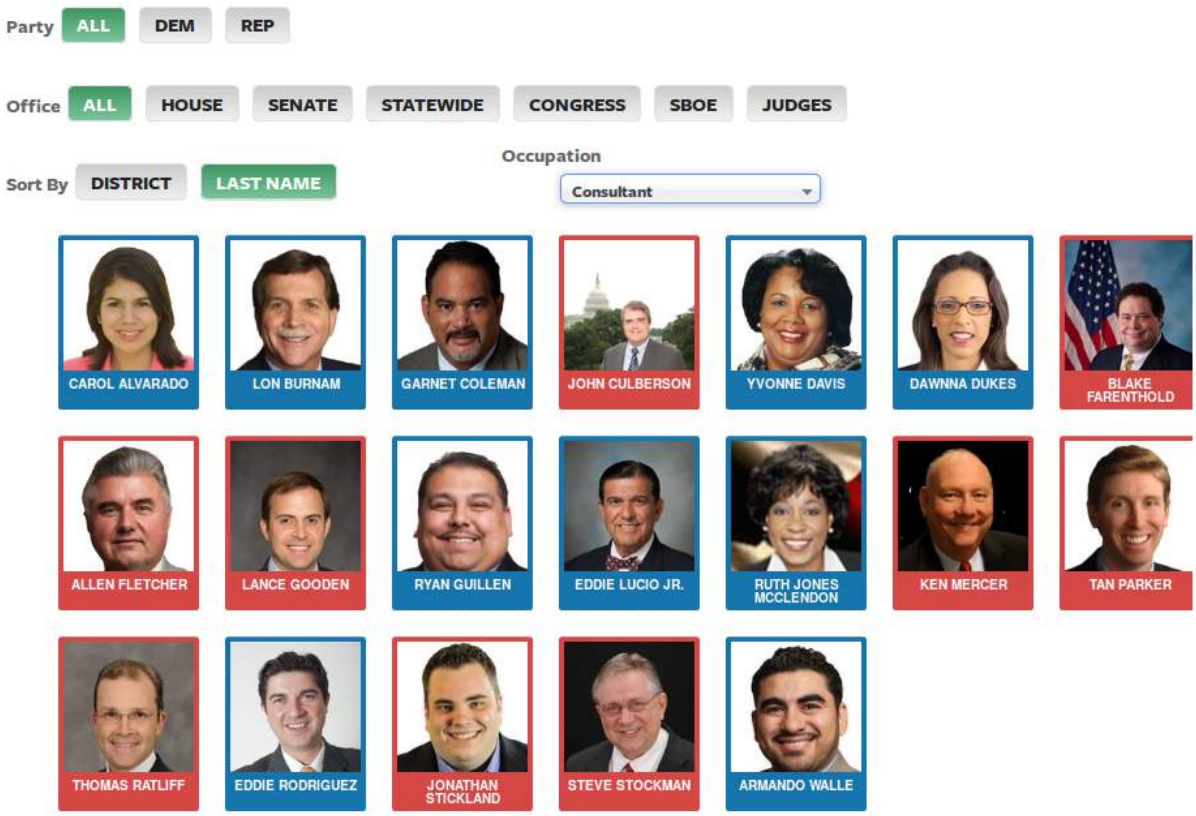


http://ejoh.se/worldcup2014/

74. North Korea vs. South Korea in figures [The Guardian]

http://www.theguardian.com/world/datablog/interactive/2013/apr/09/north-korea-south-korea-interactive

75. Ethics Explorer [Texas Tribune]



https://web.archive.org/web/20130116215215/http://www.texastribune.org/bidness/explore/

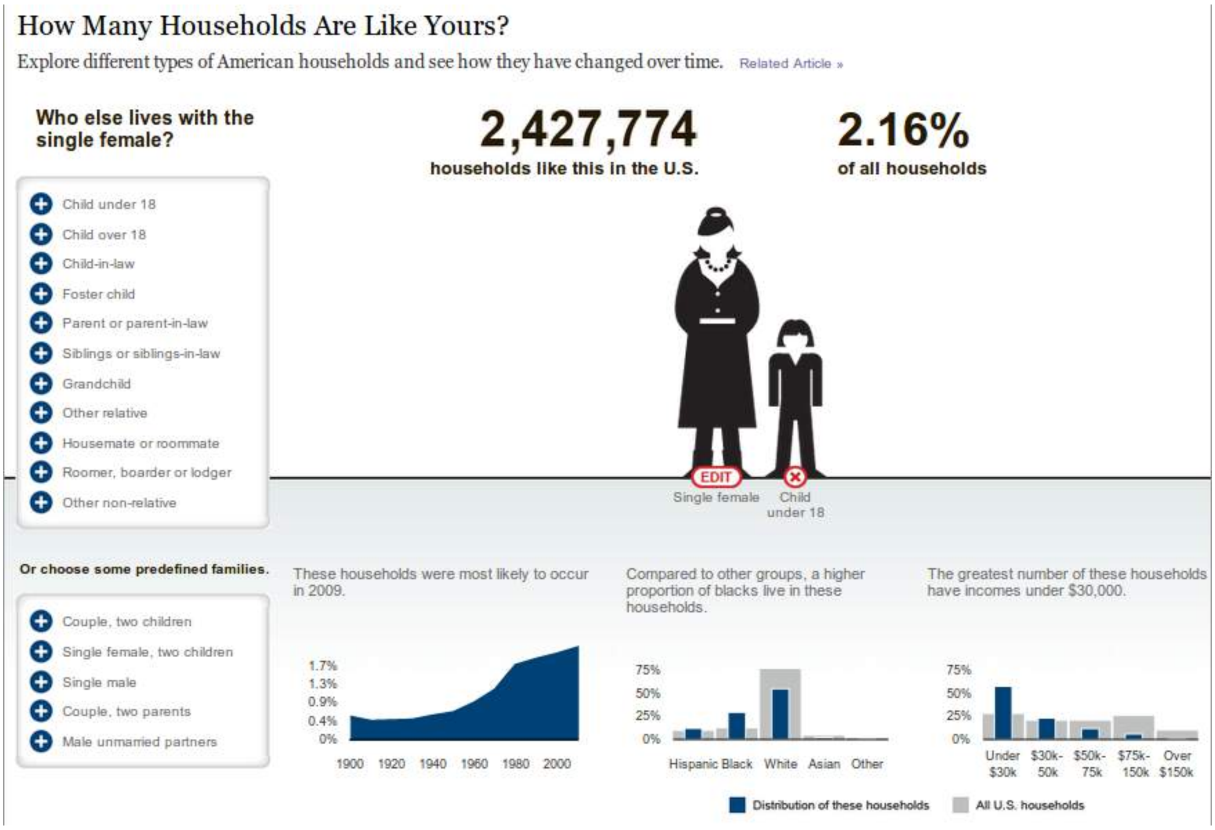
76. British Troops Killed in Afghanistan [Blight]

http://www.theguardian.com/world/interactive/2011/sep/20/british-troops-killed-in-afghanistan-interactive

77. The Workers [Rice, Conrad and Harbaugh]

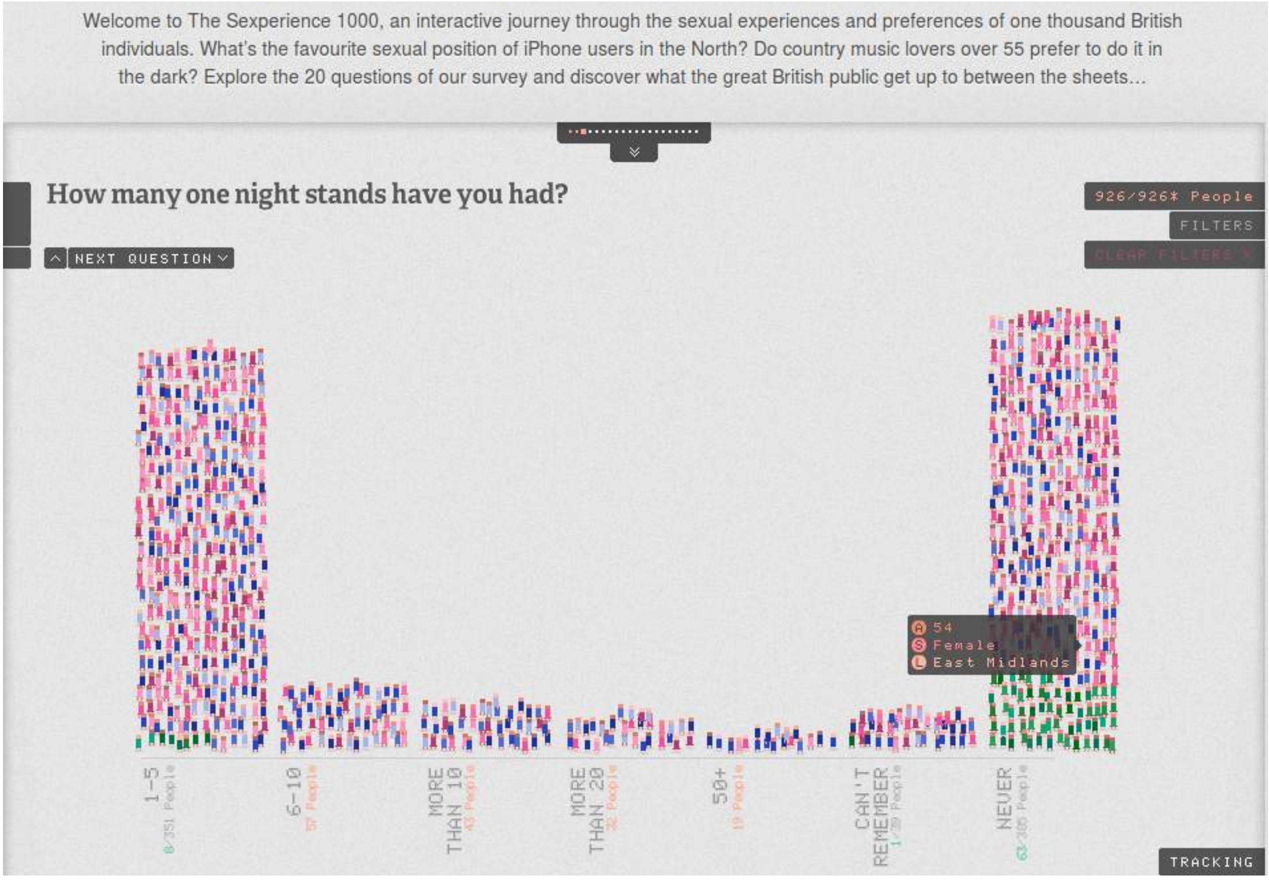
http://www.nytimes.com/interactive/2011/09/08/us/sept-11-reckoning/wtc-workers.html

78. How Many Households Are Like Yours? [Fessenden, White, Ericson, and Pecanha]



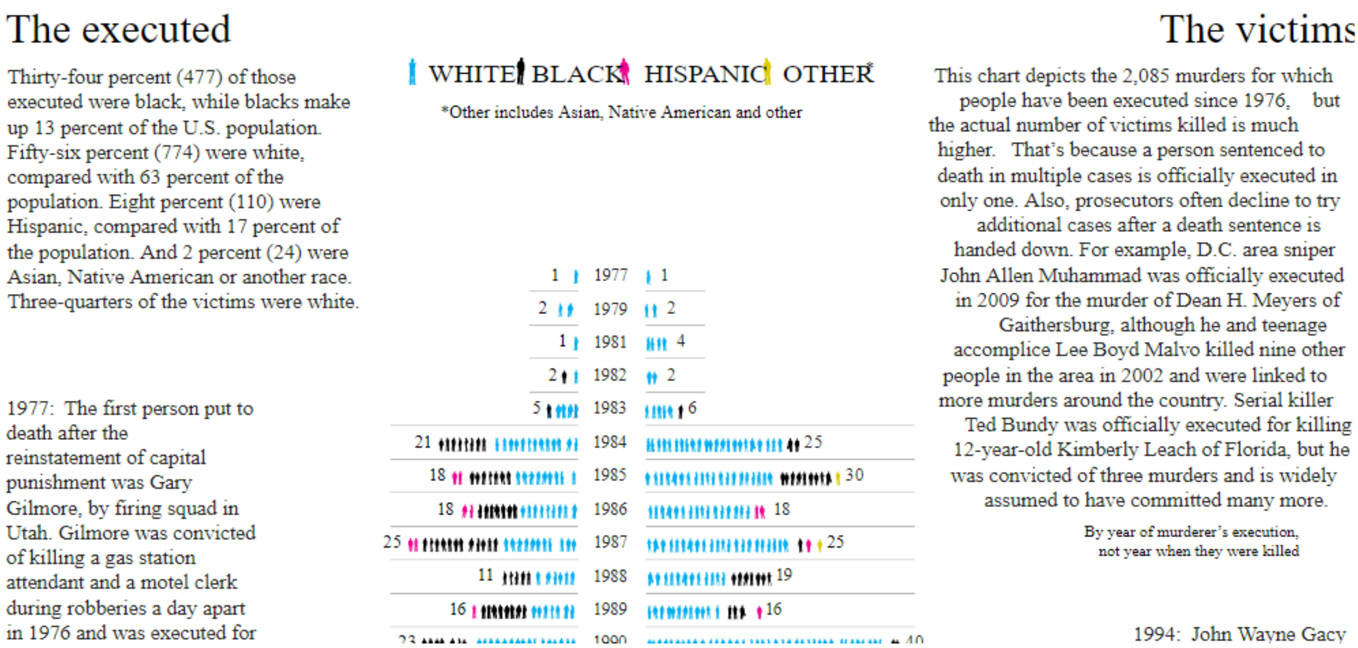
http://www.nytimes.com/interactive/2011/06/19/nyregion/how-many-households-are-like-yours.html

79. The Sexperience 1000 [Unknown]



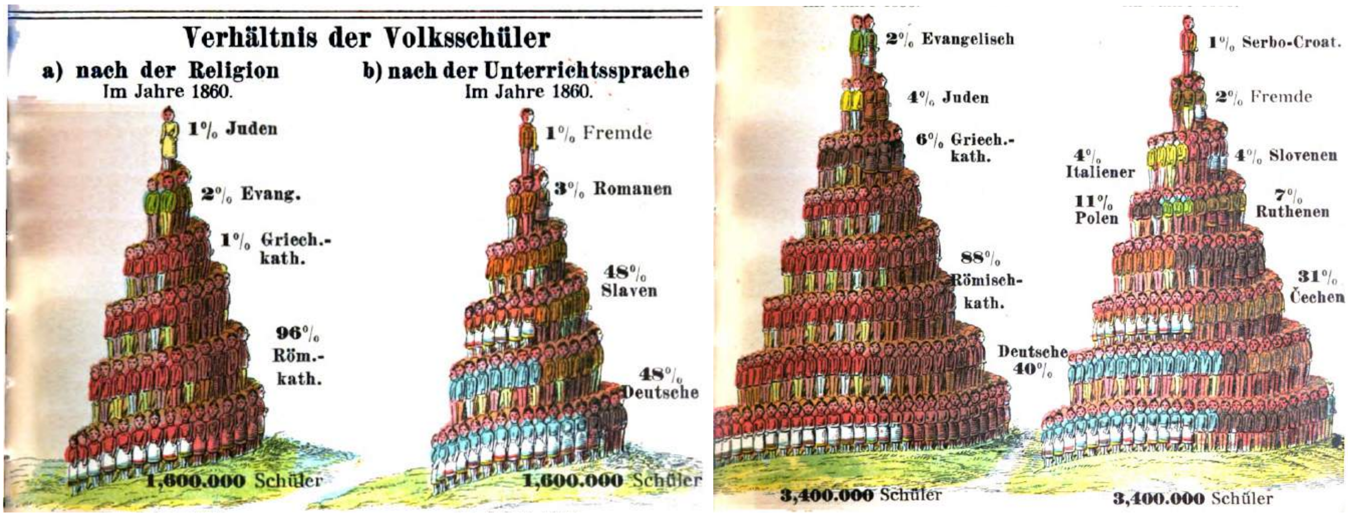
http://sexperienceuk.channel4.com/the-sexperience-1000

80. An eye for an eye? [WP]



https://www.washingtonpost.com/wp-srv/special/outlook/death-penalty/??noredirect=on

81. Proportion of elementary school pupils [Hickmann]



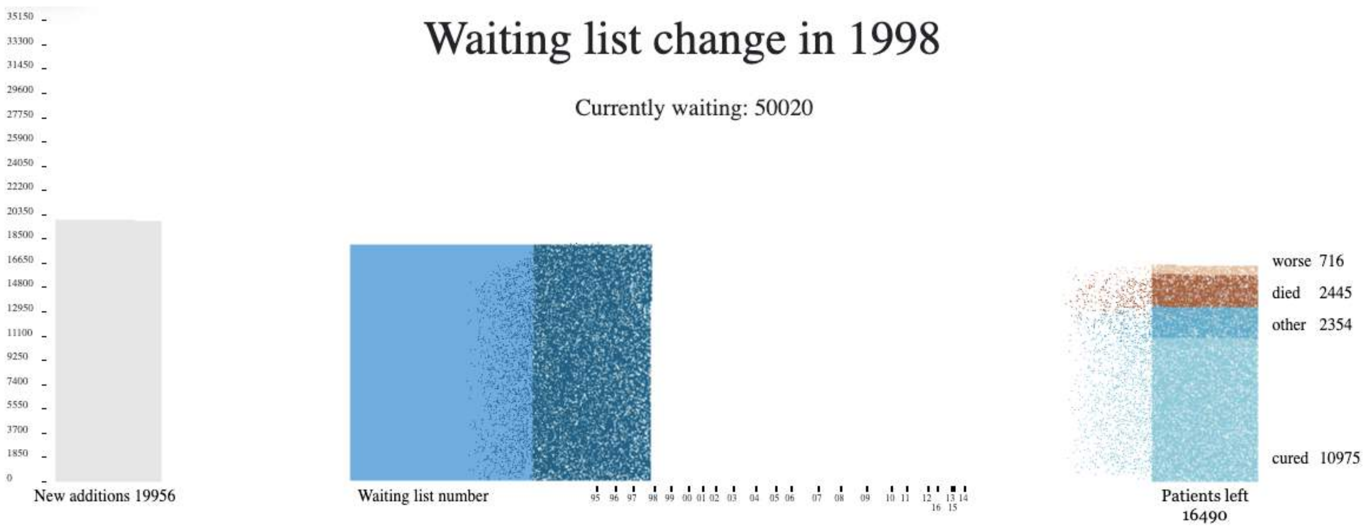
82. We Were Strangers Once, Too [Hoffman, Younse, and Thorp]

https://www.jerthorp.com/wewerestrangers

83. San Francisco Gay Men's Chorus representing AIDS deaths until 1993 [Unknown]

https://sdgln.com/causes/2017/11/28/picture-1993-reminds-people-loss-life-due-aids

84. The Impatient List [Gupta, Xu, Jin, and Sheth]



https://xujunjiejack.github.io/

85. Poverty in Syria [Boy]

https://dl.acm.org/citation.cfm?doid=3025453.3025512 (Source materials)

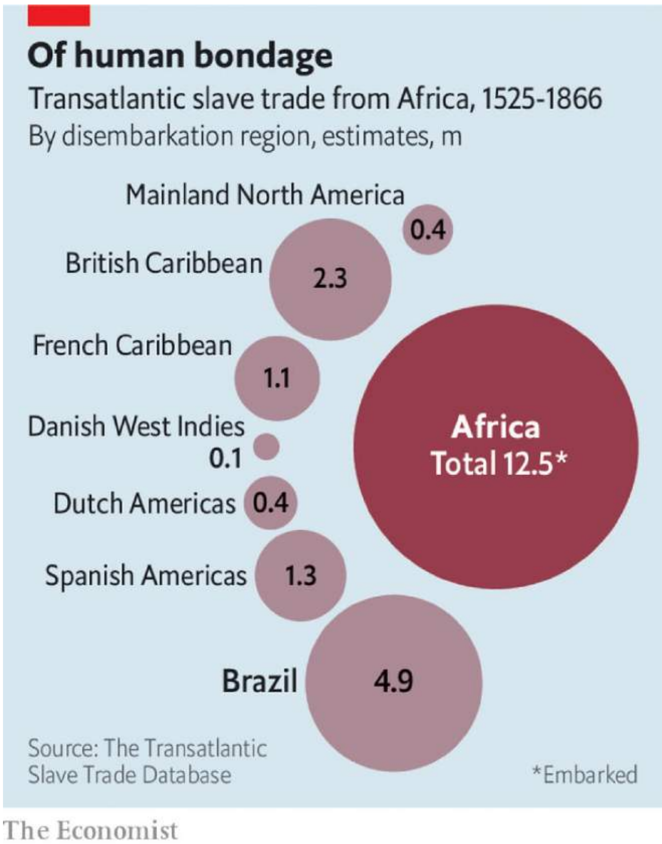
86. Activity Clock [Morais, Andrade]

http://www.luizaugustomm.me/papers/2019_pacificvis_defamiliarization.pdf

87. Hostile Terrain [De León]

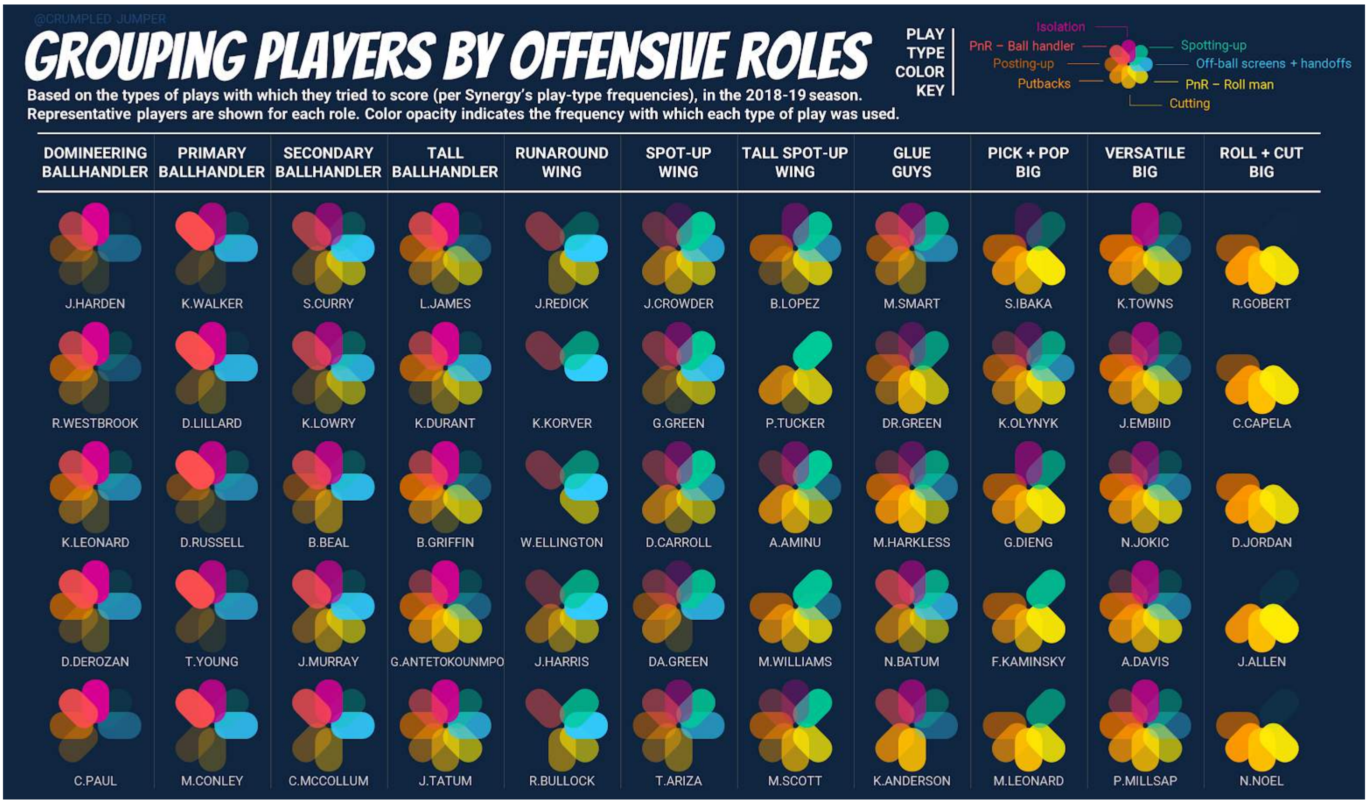
https://hostileterrain94.wordpress.com/about/

88. Transatlantic Slave Trade from Africa [The Economist]



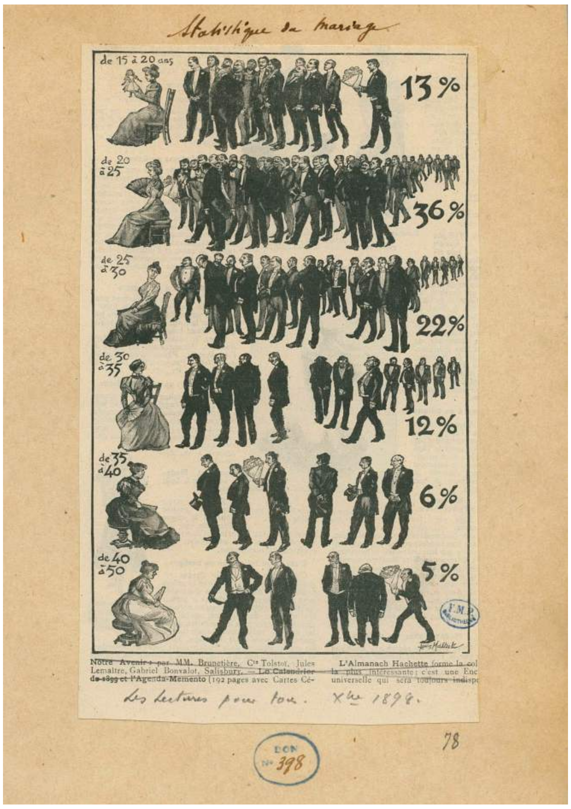
https://www.economist.com/books-and-arts/2019/05/23/for-some-in-brazil-commemorating-slavery-is-vital

89. NBA Players by Offensive Roles [Crumpled Jumper]

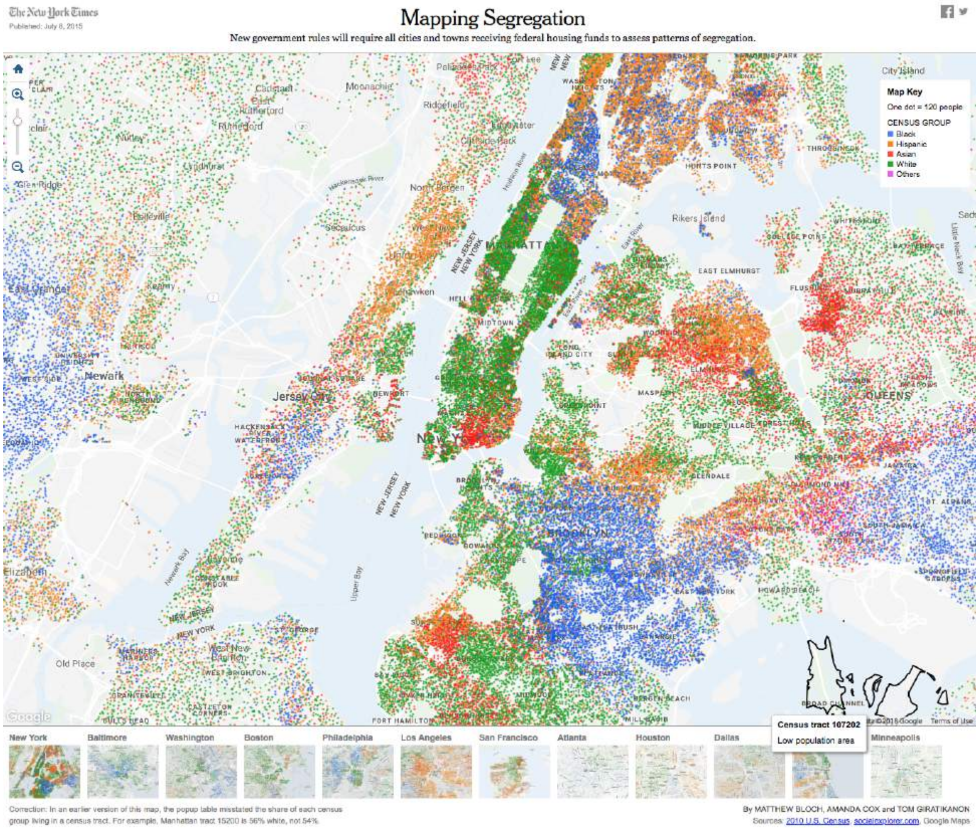


https://fansided.com/2019/05/29/nylon-calculus-grouping-players-offensive-role-again/

90. Statistique du Mariage [Louis Malateste]

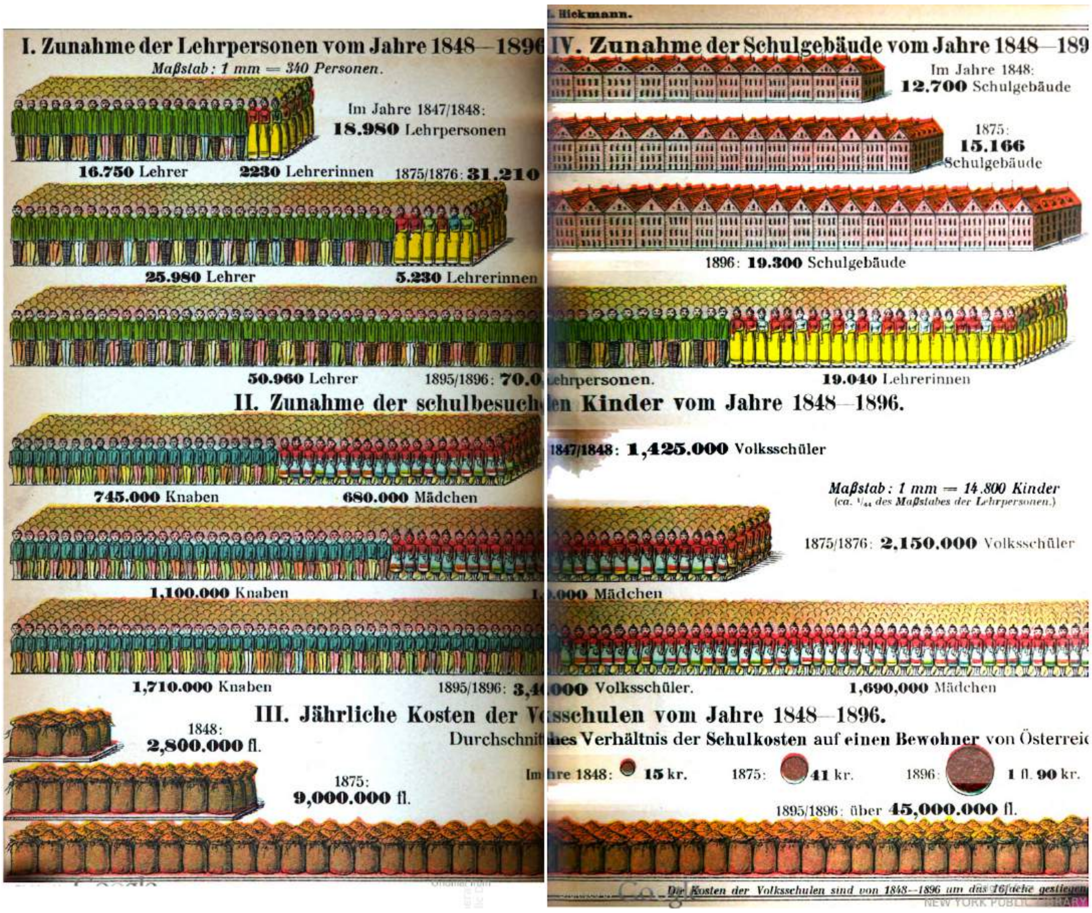


91. Mapping Segregation [NYT]



https://www.nytimes.com/interactive/2015/07/08/us/census-race-map.html?_r=1

92. Series of Old Charts About People [Hickmann]



93. Young People in Maritime Transport in the USSR [Neurath and Arntz]

https://labvis.eba.ufrj.br/infograficos-sovieticos/

94. Young People who Built the First Subway Line [Neurath and Arntz]

https://labvis.eba.ufrj.br/infograficos-sovieticos/

95. Millions of Working Women Have Become Cultured [Nashabolovke Gallery]

https://www.nashabolovke-gallery.com/izosta-eng

96. Bruises [Lupi]

http://giorgialupi.com/bruises-the-data-we-dont-see

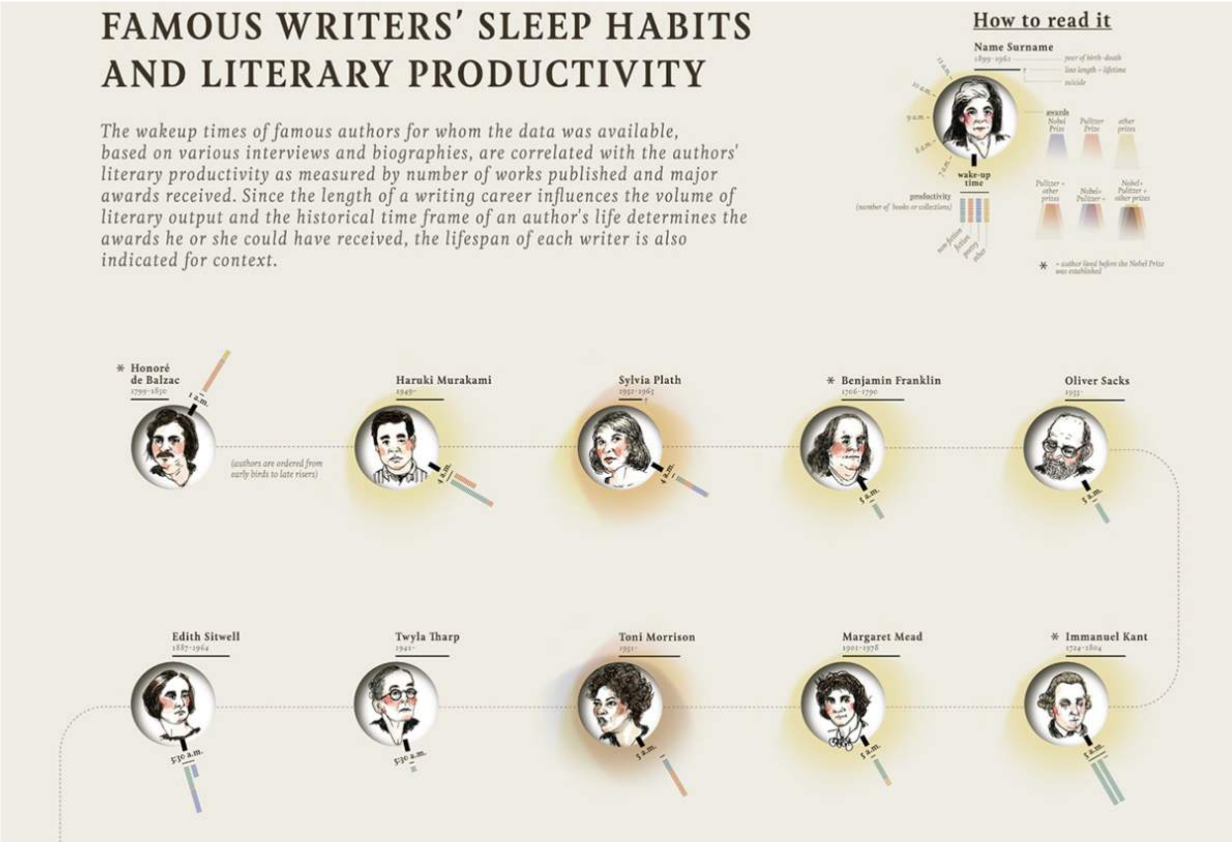
97. Building Hopes [Lupi]

http://giorgialupi.com/building-hopes

98. ..Ma poi, che cos'è un nome? [Lupi]

http://giorgialupi.com/ma-poi-che-cose-un-nome

99. Famous Writers' Sleep Habits and Productivity [Lupi]

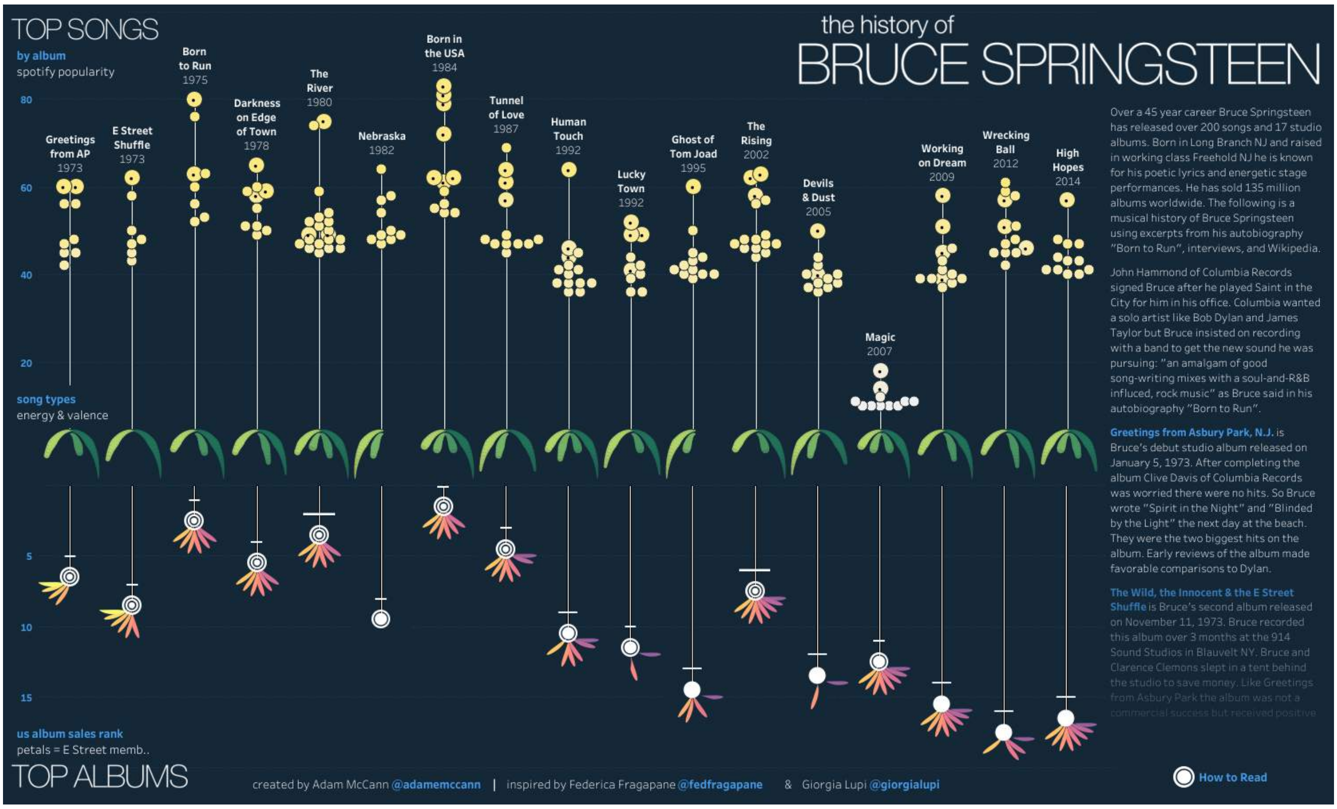


http://giorgialupi.com/famous-writers-sleep-habit-and-productivity

100. The Women of DataViz [DataViz Today]

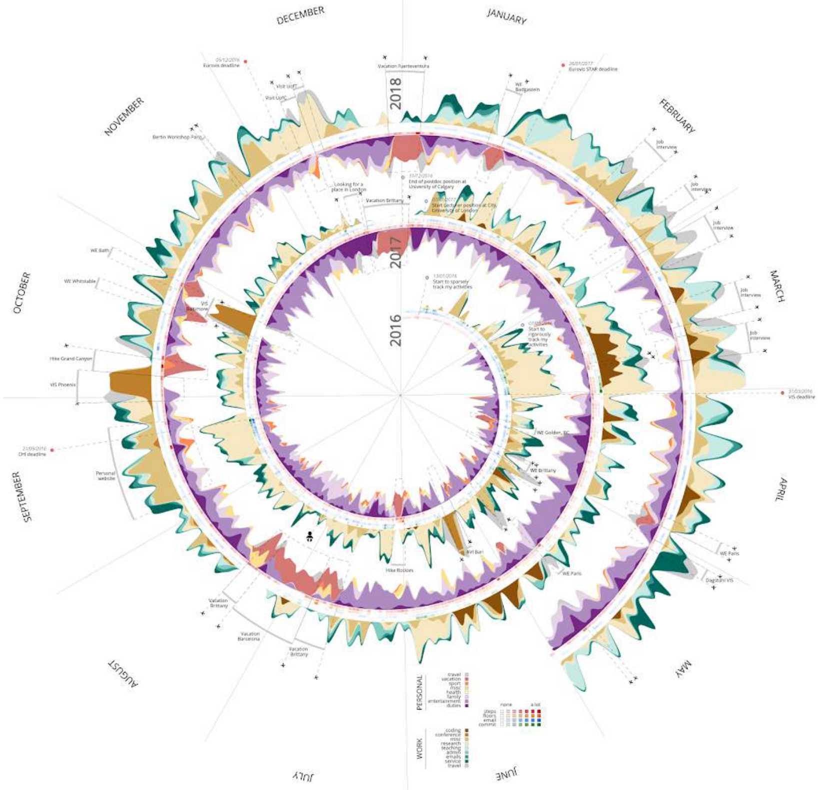
https://dataviztoday.com/shownotes/28

101. History of Bruce Springsteen [Dueling Data]



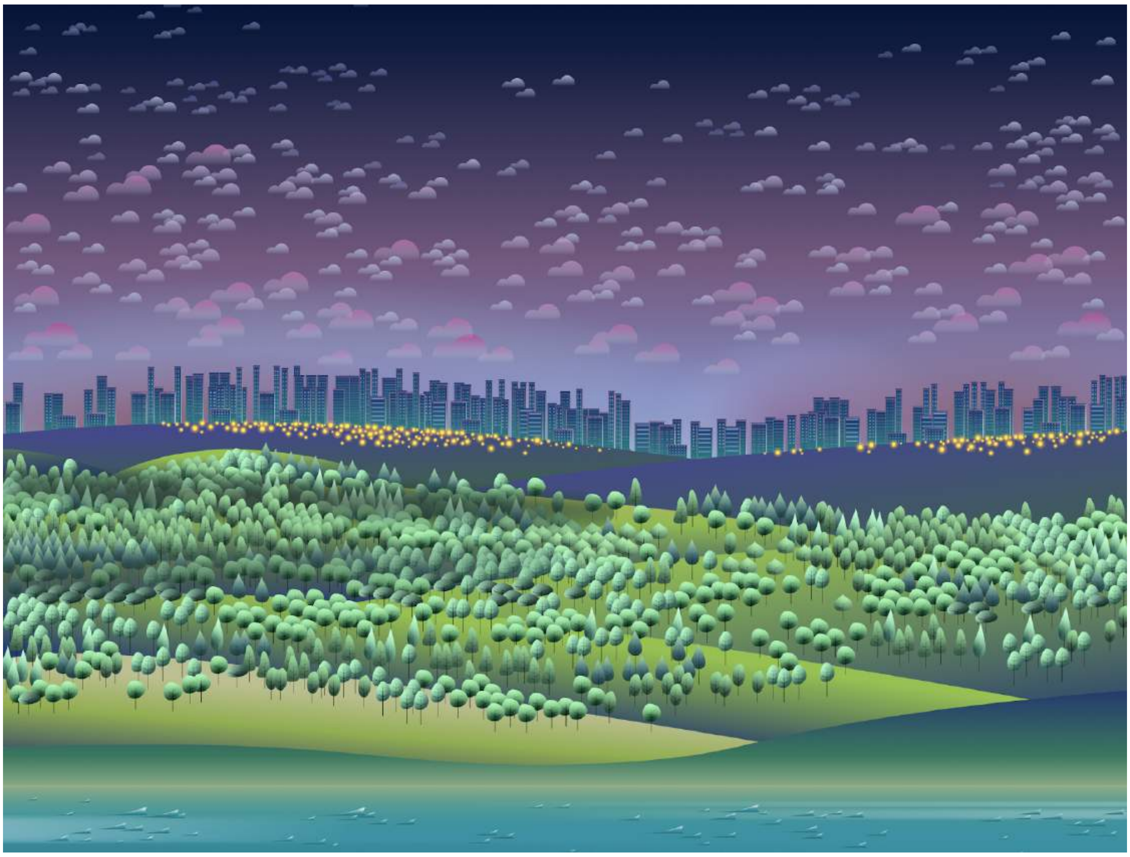
http://duelingdatalarge.blogspot.com/2018/08/history-of-bruce-springsteen.html

102. The Symmetry of My Life [Perin]



http://charlesperin.net/projects/symmetry_of_my_life

103. A View of Despair [Kuijpers]



https://www.studioterp.nl/a-view-on-despair-a-datavisualization-project-by-studio-terp/

104. Deaths and Injuries related to school shootings in the U.S. since 1998 [Daily Emerald]

https://www.dailyemerald.com/news/hundreds-of-lawn-flags-seek-to-visualize-the-number-of/article_39b9fdb6-a1bc-5139-9a3c-311ac56de0f1.html

105. Shoes Representing Gun Violence in the USA [Loeb]

https://www.popsugar.com/news/Gun-Violence-Protestors-Leave-Shoes-Outside-US-Capitol-44663568

Appendix B

Data Collection Instruments from the Harassment Project

This appendix shows questionnaires used to collect data in the Harassment Project presented in chapters 4 and 5. The first questionnaire was used to collect data for the development of the Harassment Plants and Harassment Information. The second form capture passers-by responses regarding compassion. The questionnaires are written in Portuguese.

Coleta feita no dia ____/____/____ das ____:___ às ____:____

Tipo	**Descrição**	**Negou e depois confirmou?**	**Quantidade de assediadores** 0. em geral 1. um 2. vários	**Idade aparente** 0. sem padrão 1. jovem: 0-25 2. adulto: 26-60 3. idoso: 61+

Tipos de assédio
1. Olhares maliciosos | 2. Piscadas ou assobios | 3. Frases incômodas | 4. Ameaças | 5. Mostrando parte íntima dele | 6. Pegando em parte íntima dela | Outro. Qual?

Seu ponto de vista sobre assédios no Açude

*Obrigatório

1. Data *

Exemplo: 7 de janeiro de 2019

2. Local *

Marcar apenas uma oval.

- Açude Velho
- Parque da Criança
- Quiosque
- Marcos/Humanas
- Outro:

3. Sexo *

Marcar apenas uma oval.

- Masculino
- Feminino

Imagine a seguinte situação...

Suponha que eu te darei R$ 100 por participar dessa pesquisa.

Se desejar, eu posso dividir esse dinheiro entre você e uma doação para uma instituição que vai promover uma campanha de combate a assédios no Açude Velho.

4. Quanto dos R$100 você doaria à instituição? *

5. Por que você você decidiu doar esse valor? *

6. Poderia informar seu e-mail ou telefone para a gente te contatar em um segundo momento?

Realizaremos entrevistas com pessoas que interagiram com a visualização. Por favor, deixe seu contato se você tem interesse de contribuir :)

O quanto você concorda com as afirmações a seguir?

Marque numa escala entre “discordo fortemente” a “concordo fortemente”

7. Os assédios que acontecem contra mulheres no Açude Velho são um problema sério e precisam ser combatidos *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente						Concordo fortemente

8. Se uma mulher que está caminhando no Açude Velho leva uma cantada e fica calada, significa que ela gostou *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente						Concordo fortemente

9. Assobiar ou soltar uma piada para uma mulher bonita que está passando pelo Açude Velho é uma forma de elogio *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	◯	◯	◯	◯	◯	Concordo fortemente

10. É inconveniente ficar olhando malicioso para as mulheres que estão passeando no Açude Velho com roupa justa *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	◯	◯	◯	◯	◯	Concordo fortemente

11. É errado caminhar no Açude Velho ao lado de uma mulher desconhecida e soltar cantadas se ela não demonstrar interesse *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo fortemente	◯	◯	◯	◯	◯	Concordo fortemente

O quanto você se sentiu...

Por favor, marque a intensidade das emoções que você sentiu ao explorar a visualização

12. Solidário(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	◯	◯	◯	◯	◯	Muito

13. Comovido(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	◯	◯	◯	◯	◯	Muito

14. Com pena

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	◯	◯	◯	◯	◯	Muito

15. Carinhoso(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	◯	◯	◯	◯	◯	Muito

16. Caloroso(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	◯	◯	◯	◯	◯	Muito

17. Generoso(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	⬭	⬭	⬭	⬭	⬭	Muito

18. Assustado(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	⬭	⬭	⬭	⬭	⬭	Muito

19. Triste

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	⬭	⬭	⬭	⬭	⬭	Muito

20. Preocupado(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	⬭	⬭	⬭	⬭	⬭	Muito

21. Incomodado(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	○	○	○	○	○	Muito

22. Perturbado(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	○	○	○	○	○	Muito

23. Angustiado(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	○	○	○	○	○	Muito

24. Apreensivo(a)

Marcar apenas uma oval.

	1	2	3	4	5	
Nem um pouco	○	○	○	○	○	Muito