Call of Data: An Interactive Exploration of the Bias in Visualization-Making

Aoudi, Lama*. Clochard, Axelle*

*Massachusetts Institute of Technology

Abstract

Data exploration is often introduced as a crucial first step to creating a visualization. However, interacting with one's data throughout the design process introduces a huge potential for bias by increasing the likelihood that designers will make data decisions based on whether the resulting graph agrees with their prior assumptions. Our final project, *Call of Data,* is an attempt to simulate the experience of creating a visualization in "data-blind" conditions. Throughout the course of this interactive experience, users will be asked to make a number of data decisions guided only by their *a priori* knowledge of the variables in question. At the conclusion of the experience, they will find out which caricatural personality type the biases of their resulting graph most closely resembled. They will then be able to visualize the impact of each of their decisions in real time, and be prompted to contemplate whether they would have made different decisions had they had access to this knowledge. We hope that this experience will inspire users to consider a new design protocol: one where experimental interaction with data does not drive the creation of a data product, but rather becomes secondary to what we know about the variables at hand.

1 Introduction

A cornerstone of scientific research is hypothesis testing. In short, it is a process that allows scientists to test the validity of a theory by obtaining experimental data and using statistics to evaluate whether this data conforms to the conditions specified under a null hypothesis (usually, the conditions under which the theory would not be valid), or to the conditions specified under the alternative hypothesis (usually, the conditions under which the theory could be valid).[1] The standard protocol is to refrain from looking at the data before formulating these hypotheses and the appropriate tests that go with them. This is because any prior engagement with the data could influence the direction of the testing framework by allowing the scientist to identify conditions that would be more likely to support their theory.[2]

The practice of creating data visualizations is built on different tenets. Crucially, visuals are polished products and data is their raw matter. Therefore, unlike in hypothesis testing, exploring and manipulating one's data is an expected, core part of the process. An important consequence of this approach is that there is no expectation that the designer should refrain from visualizing the impact of their data manipulation decisions on the product. However, this interactivity introduces a huge potential for bias and provides an argument in support of the claim that it is impossible to create an unbiased visual. Despite this, visuals are often perceived by both their makers and consumers as a part of the scientific process: a neutral tool with which to evaluate a theory.[4]

Our final project is an attempt to simulate the experience of creating a data visualization under the "data-blind" conditions of hypothesis testing. Throughout the course of our interactive application, the user will be asked to make a number of data manipulation decisions without having looked at the data and without the ability to gauge the impact of their decision on the resulting graph. As in the context of hypothesis testing, they will be guided only by their *a priori* knowledge (some of which we will provide) of the variables in question.

2 Related Work

2.1 Literature

An important source of inspiration for this project was the White Hat/Black Hat Visualization assignment from MIT's 6.859 Interactive Data Visualization class. This assignment prompted students to use a single dataset to create two visualizations: one honest and the other "intentionally inappropriate" and "misleading".[3] The idea was for students to adopt "two perspectives" when manipulating and shaping their data, and to ultimately become more cognizant of the places where bias and deception could lurk in seemingly straightforward visuals.[3] We were interested in how the standard visualization process facilitates the presence of bias and deception even when the designer does not have any ill intent.

In fact, many of the texts associated with 6.859 sought to expose the ways in which bias is part and parcel of visualization-making. Particularly relevant here are the arguments of Michael Correl "against the neutrality of visualization", where he asserts that "[all] visualizations are rhetorical, and have the potential to persuade" through even "minor choices" that were made "without conscious knowledge".[4] This is in part because, as Correl explains, designers "are often the first and only contact a person might have with an underlying store of data" and therefore "control the curation, presentation, rhetorical content of the" resulting visualizations. [4] Both their choice of what is omitted and what is included will be the result of a chain of judgements that could never be totally bias-free.

Similarly, Lee et al. used an analysis of grassroots "counter-visualizations" to point out that sophisticated and seemingly objective data visualization methods could be used in the service of making non-scientific claims.[5] Their juxtaposition of the scientific community - with its peer review process and "laboratory" produced "knowledge"- and what the authors called the "anti-maskers- who "champion science as a personal practice that prizes rationality and autonomy"- served to emphasize that visualizations exist in value-laden contexts, and that the choice of process itself is driven by individual principles.[5]

Finally, D'Ignazio provided a particularly powerful illustration of how bias can strike at the very root of visualization-making, through the necessarily "reductive" character of data. [6] Indeed, because data "encodes" information through "classification systems", it decides which categories best reflect the information at hand, or quite simply "what counts", and in doing so often reflects deeply-ingrained “values or judgements”. [6] Together, these authors remind us that methods, data, and visualization techniques will always reflect some human bias, even in spite of the designer’s best intentions.

2.2 Implementation

In terms of the actual implementation of exposing the bias in visualization by way of a game, we were not able to find anything similar on the internet. Of course, the format, name, and characters were tongue-in-cheek callbacks to existing games, universes, and other online interactive experiences. The “personality quiz” format was inspired by Buzzfeed’s infamous quizzes, the name “Call of Data” was a wink to the immensely famous video game “Call of Duty”, the avatars were existing pixel art characters found online (sources in presentation), and the overall style was meant to be reminiscent of 80s-style, retro console games. These choices were made not only out of convenience, but also to emphasize the playful aspect of the experience and to reduce the cognitive load of the user who has to make a number of intellectually-demanding decisions throughout the game.

3 Methods

Call of Data is a game-like experience that can be divided into three parts: a mission, data selection, and results dashboards. The game is set-up with the user as a data scientist called upon to help a fictitious president create an ‘honest’ visual about climate change. The visual should help convince world leaders to ‘save the planet’. Once accepting the mission, users are probed to select a player. The player has no bearing on the outcome of the game and is only intended to increase user engagement, and keep with the video-game narrative. The game will then prompt the player to make a series of decisions on which data they would like to visualize, and how it should be manipulated. A crucial element to our game is the user’s inability to view how their choices are being encoded as they make them. In lieu of real-time visualization, the significance of the decision is specified at the top of each section, and each choice is accompanied by a compelling argument for why it might be effective at accomplishing their mission. See Figure 1 for an example of a choice. There are a total of five choices to be made. The game records the users’ choices to create the final visualization, as well as to sort them into one of the three character archetypes: (1) the Oil Tycoon, (2) the Honest Scientist, and (3) the Tech Enthusiast. Once complete, the game explains how the user’s visualization is biased, and encourages them to engage with a final dashboard that will allow them to toggle between their previous choices and observe how the final chart, and its message, shifts.

3.1 Data Exploration

An important consideration when designing the game was choosing the dataset. The subject matter needed to be widely understood to guarantee users’ had enough *a priori* knowledge to engage with the mission. The dataset (and the subject) had to include multiple independent variables users could choose between and equally accomplish their task with. The data on climate change exactly satisfies these requirements and is why we chose to center the president’s mission around ‘saving the planet’.

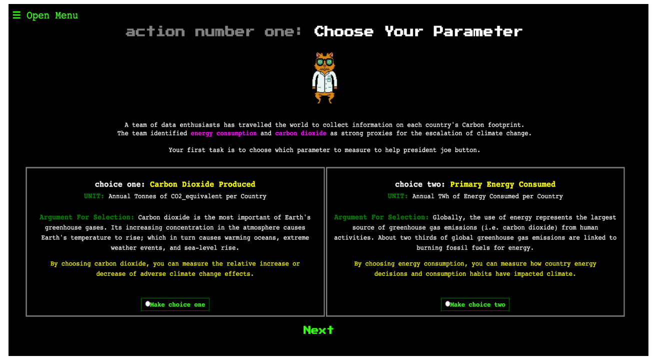

Figure 1- An example of how choices are presented to users.

Due to climate change’s multifacetedness, most of our early data exploration involved identifying the key metrics to narrow in on. Driving our decision-making was the requirement that the metric values were available for all countries, at the same time range (i.e. no missing data). Accordingly, we decided to measure two densely-recorded metrics: annual carbon dioxide emissions and annual primary energy consumption. We used the open-access dataset, *Our World in Data*, which aggregated CO_2 emission data (measured in tonnes of CO_2 produced) from the Global Carbon Project, and collected national energy information (measured in TWh) from both the BP Statistical Review of World Energy and the World Bank’s World Development Indicators. Country statistics, like total population and total gross-domestic-product (adjusted for inflation), are also provided and were aggregated by several sources. The statistics are incorporated into our game as potential confounding variables the users can normalize with in their visualization.

3.2 Visual Encodings

For simplicity, we chose to limit the final visual to an ordinal line-graph, to emphasize how even just one encoding can lead to many possible outcomes. The x-axis is fixed as time (i.e. the year) and color is used to denote the country. The variable on the y-axis is set to change with the user’s manipulation of the data. Figure 2 provides a schematic of the visualization choices and further detail on each is described below. Recall that each choice is equally reasonable and users are supplied a valid argument why either should be chosen. The language of each argument is carefully written to avoid any implicit favoring of one alternative over the other.

I. Parameter

Users are given the choice between two parameters: CO_2 emissions or energy consumption. CO_2 is used as a proxy for greenhouse gas emissions and is the canonical indicator of worsening climate effects. Energy consumption signifies the growth or decline of carbon-emitting and renewable fuels. Both parameters suffer from omitted-variable bias when studied alone; and are thus powerful examples of “objective data visualization methods” used to make incomplete scientific claims.

II. Time Frame

Though the data spans 1965-2016, we created a separate subset for the years 1990-2016. The decision was made after a preliminary data exploration in Tableau showed trends differed between 1965-1990 and 1990-today. The difference was large enough that choosing between either could lead to over- or under-estimation of the effects of climate change, and thus make a compelling argument for how seemingly benign visualization choices can lead to bias.

III. Country Types

Users are then given the choice between different types of countries to study their parameters throughout. To ensure the line graph remained interpretable, only fifteen countries could be visualized at once. Four groups of fifteen countries were pre-selected by us. To create the groups we leveraged existing preconceptions around what characteristics of a country would make them a key nation to study in a visualization on climate change. Furthermore, each group was designed to ostensibly align with one of the biased character archetypes introduced earlier. The groups are as follows:

- The largest oil producing nations
- Countries with the largest populations
- Most technologically advanced countries
- Countries with the most diverse energy mix (ostensibly countries with the highest capacity of renewable energy)

IV. Data Manipulation

Once the data is adequately subsetted, users are able to decide how to manipulate the raw data to convey their message. The first decision is whether to plot absolute values of their parameters or the relative growth rate. Growth rates are pre-calculated for each time frame with t = 0 set at 1965 and 1990 respectively. The second data decision is whether to scale the parameters by country statistics, e.g. gross development product (GDP) or population, or neither. Since GDP is typically used as a signal of a country's development and wealth, one can assume that countries with higher GDPs will rely on higher amounts of energy-intensive products. Similarly, countries with larger populations will necessarily produce and consume more, not necessarily because their lifestyles demand it, but because of numbers alone. Plotting per-capita or per-GDP values removes the external effects of total people or wealth from comparisons between countries in the same group, and facilitates apples-to-apples comparison of CO_2 or energy consumption over time.

3.3 Implementation of Interactivity

The objective with Call of Data is to *show*, not tell, users that all visualizations can be biased. As such, interactivity was crucial to the success of our game. The primary interactive elements are the data features selection, final dashboard, and character selection.

I. Storing Data Selection & Dashboard

The user's data choices were stored as variables along the way for two purposes. The first was to feed them through a personality-quiz algorithm which is described in the next section. The second was to generate the parameters for a graph. The graph, labels, legends, axes, and line elements were all reactive to these choices, using event listeners that picked up on new user input and appropriately subset the underlying dataset and updated the graph in response.

After the user's final graph is revealed to them, they are also able to assess, this time in an exploratory manner, how each of their selections affected the final outcome. They are also prompted to think about whether visualizing the impact of their actions would have affected their data decisions, and what that means for the "truthfulness" of the graph.

II. Characters: Buzzfeed Personality Quiz

As described above the data choices were stored as variables and fed through a personality-quiz algorithm; which resembles a Buzzfeed Personality Quiz by trying to identify what type of character the user's choices most resemble. In our case, we have three characters and each is an embodiment of one of the many polarizing opinions on climate change:

- *The honest scientist* believes climate change is real, CO_2 emissions are rising, and energy consumption is increasing globally. The honest scientist will always create an upward ticking graph.
- *The oil tycoon* believes climate change is a hoax, CO_2 emissions are declining, energy consumption is good and tapering off. The oil tycoon will produce a downward trending graph.
- *The tech enthusiast* resents the claim that server farms and the tech industry are harming the planet, and believes that a technology revolution is not associated with negative effects on the climate. The tech enthusiast will produce a flat graph, with small perturbations.

The algorithm assigns users a character using a point system. All characters are initialized with 0 points. Each possible choice is then given works by initializing each character to have 0 points, then mapping which mapseach path of choices with a number of points that will skew the results in favor of the associate. For example, each possible choice assigns a fixed number of points to each character. If a choice does not skew the chart in favor of any archetype then all characters receive an identical number of points. Paths that lead to upward, flat, or downward trending charts were pre-identified and the point system was defined accordingly. Conditional statements were used to incorporate dependencies between choices, i.e. a user must select the timeframe "1990-2016" and "Tech Advanced countries" to receive x number of points. Figure 2 attempts to highlight a few examples of inter-choice dependencies. The schematic describes the necessary sequence of choices to be an *honest scientist,* beginning at "CO_2" and ending at the choice "No Division". Any alternative choice along the way may have resulted in a completely different bias to chart, i.e. character.

4 Results

The following section will illustrate how toggling between two variables can make the difference between an oil-tycoon and an honest scientist. Abiding by the sequence in the schematic Figure 2 the resulting chart for an oil tycoon is displayed in Figure 3. The chart shows all countries in the group "Top Oil Producers"are showing negative growth trends in energy consumption. The non-intuitive finding is the result of normalizing the values by population, which speaks nothing to the nations' contribution to global fuel consumption or their related emissions. To demonstrate the bias, we adjust our final choices to "No Division" and the chart is re-classified as "Honest Scientist", see Figure 4. The honest scientist's chart looks as we expect, and energy consumption increases for all countries.

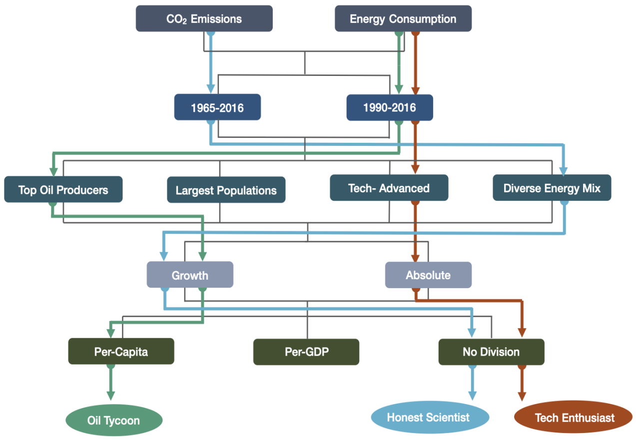

Figure 2 - A schematic of the visualization encoding decision tree. Example decisions needed to obtain a specific character are shown in the three color node-to-arrow diagram. Green signifies the *Oil Tycoon*, blue is the *Honest Scientist* and Red is the *Tech Enthusiast*.

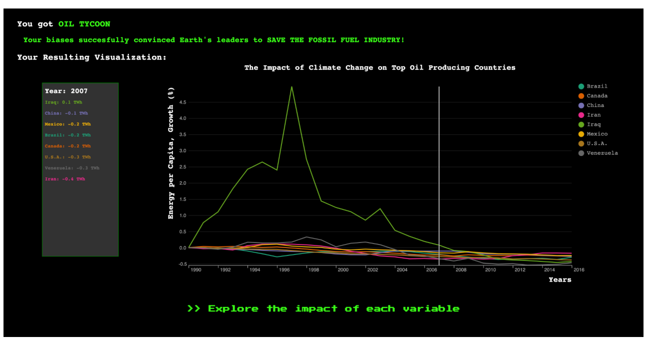

Figure 3- The resulting chart following the path of the oil tycoon.

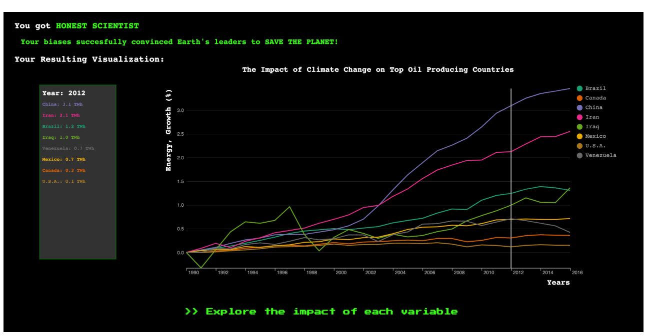

Figure 4- The resulting chart after modifying the final choice to "no division". The result is an honest scientist biased graph.

The sensitivity of the chart to a single decision emphasizes how effective biases are at changing, or re-writing any data's narrative. The abrupt change in the chart is also why we made users choose their visualization parameters blindly. Any foresight into the chart might have contradicted their preconceptions of what they intended their "honest" visual to look like and immediately alter their decision making, diminishing Call of Data's effectiveness at addressing bias in visualization.

5 Discussion

Our intention was that the user would come out of this experience more aware of the extent to which they usually interact with data when building a data visualization, and to question their motivation for doing so as well as the potential pitfalls of this practice. Most designers would probably consider it bad practice not to check the impact of their data decisions on their interim data products. And yet, when we do so, we introduce the potential for bias by creating opportunities for ourselves to make data decisions based on whether the result agrees with our *a priori* assumptions. Our hope was that the process of building a graphic in a "data-blind" manner and receiving a "personality type" based on this process would highlight this tension between experimental knowledge of data and bias.

In an ideal case, our users (most of whom probably consider themselves to be "honest scientists") would be assigned a personality type that they did not expect. This experience would highlight the fact that a graph produced with only *a priori* knowledge of the data would not agree with a graph they might have built with experimental knowledge of the data.

Beyond this specific scenario however, it should be clear to the user that the choices they were asked to make did not necessarily have right answers. The point of this exercise was simply to point out that a good faith decision might lead to a graph that the societal perspective might consider a "counter-visualization".

Finally, the concluding dashboard and its myriad of potential iterations should emphasize that a single question contains a plurality of potential graphs. Our simple example did not ask the user to vary encodings, and only asked the users to make five decisions. Yet, the range of graphs the user could have expected to see was vast.

Of course, one possible objection is that this experience merely replaces one type of bias (judgements based on the appearance of the data) with another (judgements based on prior knowledge of the data). But we are not advocating for an approach where the user completely neglects to interact with their experimental data. Rather, we'd hope that this experience would prompt users to consider a new, more balanced protocol: one where experimental interaction with data does not drive the creation of a data product, but instead takes a back seat to what we know to be true about the variables at hand.

6 Future Work

Several features can be expanded on in future iterations of our work. Most notably, users can be given additional flexibility to choose between a broader list of climate-related independent variables (e.g. global temperature or fossil fuel consumption). In a similar vein, additional characters (or archetypes) could have been incorporated. For example, the "American Patriot" who would like to show that climate metrics are standing still or decreasing in the United States in comparison to other nations. The benefit of additional character and variables is increased nuance in how we define or create a biased visualization. It could also be beneficial to introduce qualitative engagements with users. For instance, upon showing the final graph users might be probed to answer: - *why could this visual be biased?* - prior to being given a character personality. Of course building out the game would require rewriting the quiz algorithm to be more scalable and robust, i.e. not require the pre-identification of paths. This might require meta-software that can detect bias as opposed to pinpointing precise junctions where choices skew the story in any particular direction.

Furthermore, dealing with an issue where there was a clear scientific consensus was intentional on our part, but perhaps this exercise could have been even more effective on an issue where the scientific community is split. It is also not lost on us that a lot of our own biases were baked in this work.

References

[1] Biau, David Jean, Brigitte M. Jolles, and Raphaël Porcher. "P Value and the Theory of Hypothesis Testing: An Explanation for New Researchers." Clinical Orthopaedics and Related Research 468, no. 3 (March 2010): 885–92. https://doi.org/10.1007/s11999-009-1164-4.

[2] Gotesman, Ryan. "Why You Should Pick Your Hypotheses BEFORE Looking at Your Data." Medium, April 27, 2019. https://towardsdatascience.com/why-you-should-pick-your-hypotheses-before-looking-at-your-data-6d087751336c.

[3] "6.859: Interactive Data Visualization (Spring 2021)." Accessed May 16, 2021. http://vis.mit.edu/classes/6.859/A3/.

[4] Michael Correll. 2019. Ethical Dimensions of Visualization Research. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300418

[5] Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. Crystal Lee, Tanya Yang, Gabrielle Inchoco, Graham M. Jones, Arvind Satyanarayan. ACM CHI. 2021.

[6] Chapter 4: What Gets Counted Counts. Catherine D'Ignazio. Data Feminism. 2020.

[7] Data Source, *Our World in Data*: https://github.com/owid/co2-data