https://doi.org/10.2352/EI.2023.35.10.HVEI-251

Am I Safe? A Preliminary Examination of How Everyday People Interpret COVID Data Visualizations

Bernice Rogowitz, Visual Perspectives, Westchester, NY; Paul Borrel, Visual Perspectives, Sète, France.

Abstract

During these past years, international COVID data have been collected by several reputable organizations and made available to the worldwide community. This has resulted in a wellspring of different visualizations. Many different measures can be selected (e.g., cases, deaths, hospitalizations). And for each measure, designers and policy makers can make a myriad of different choices of how to represent the data. Data from individual countries may be presented on linear or log scales, daily, weekly, or cumulative, alone or in the context of other countries, scaled to a common grid, or scaled to their own range, raw or per capita, etc. It is well known that the data representation can influence the interpretation of data. But, what visual features in these different representations affect our judgments? To explore this idea, we conducted an experiment where we asked participants to look at time-series data plots and assess how safe they would feel if they were traveling to one of the countries represented, and how confident they are of their judgment. Observers rated 48 visualizations of the same data, rendered differently along 6 controlled dimensions. Our initial results provide insight into how characteristics of the visual representation affect human judgments of time series data. We also discuss how these results could impact how public policy and news organizations choose to represent data to the public.

Introduction

We have witnessed an enormous growth in the availability of open-source data, provided by reputable scientific and public institutions. These data have been distilled and analyzed by public health organizations, governmental agencies, universities, the press, and by a legion of bloggers and tweeters, producing a wellspring of visualizations. To track and understand the COVID-19 epidemic, for example, worldwide data on a wide range of demographic, geographic, and epidemiological factors have been analyzed and interpreted, and represented in a wide range of data visualizations [1]. Graphs abound, plotting the number of cases, deaths, and vaccines over time, by geography, broken down by age, socio-economic status, etc.

Although it is well-known that data representations can be manipulated to add unwarranted emphasis or even deception, the great majority of data visualizations are created in earnest by serious scientists, journalists, and educators. Moreover, there are many guidelines available to guide how the data are mapped onto the lines, dots, colors and shapes that make up a data visualization, to ensure that the features in the data are accurately represented to the viewer. Whether this intent is faithfully achieved by the data visualization, however, is an open question.

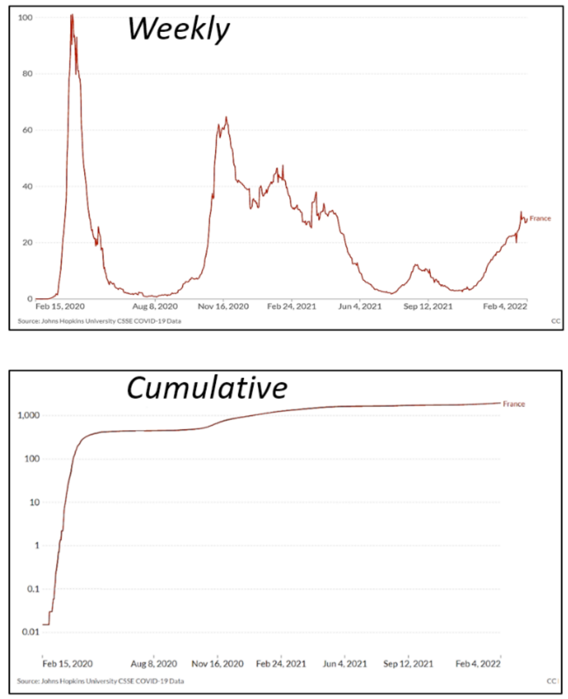

Figure 1. Weekly vs. Cumulative. Identical time series data for COVID-19 fatality is plotted in two standard and legitimate manners. The graph in the top panel plots the number of deaths each week over a two-year span. The graph in the bottom panel plots the cumulative number of deaths, week-by-week, over the same two-year span. Although the data in both representations is identical and both data representations are correct, viewers glean different impressions about the course of the fatalities over time, which influences their judgment about how safe they would feel traveling to this country.

Figure 1 shows two graphs of the same data. The graph on the top shows a time-series of COVID-19 fatality data for a single country, France. The graph on the bottom shows these same data; however, instead of plotting weekly values, the graph plots the cumulative number of fatalities. Although these graphs are both perfectly correct representations of the same data, the shape of the curve is different, and may give rise to different interpretations. In the weekly graph, the recent rise in the number of fatalities is very salient, and much more difficult to spot in the cumulative graph. On the other hand, the cumulative plot shows the enormous contribution of the initial spike in fatalities to the overall death rate, which is more difficult to spot in the weekly curve.

The question, then, is what do viewers infer from these different representations? One way to address this has been to test the visualization literacy of the viewers. The Visual Literacy

Assessment Test [2] for example, measures the degree to which data values can be extracted from a graph, in the same way a reading literacy test measures the ability to decipher text. VLAT tests, for example, whether the viewer can read off the y value at a certain x value, or compare y values at different points in time, or report on whether the curve is ascending or descending. But, even if the viewer is perfectly capable of extracting and comparing data values, they may still arrive at different conclusions depending on rendering choices in the visualization. That is, the overall impression of the meaning in the data may depend on rendering choices made by the designer.

Perceptual differences between different data renderings of the same data have been pointed out in the literature. Rogowitz and Goodman [3], for example, showed how different renderings of three financial risk variables could reveal very different features in the data. Ziemkiewicz and Kosara [4] showed that a tree-map and the equivalent node-link representation lead to different inferences about the underlying structure, even though they rendered exactly the same data. Recently Padilla, et al. [5] have shown that different choices in representing time-series forecasts can strongly impact higher-level interpretations, such as trust. The observation that different equivalent plots can provide different insights has been cited as a rationale for providing multiple linked views [6] , and motivates this research.

In this paper, we visualize a very simple 3-column spreadsheet of numbers, showing COVID-19 fatality over a two-year period for three countries. These data are visualized under 6 different conditions, in which we vary whether the data are rendered a weekly or cumulative, over a 2-year span or the most recent 6 months, whether the y axis is linear or log, whether data for one country is shown alone or in the context of the data from the other countries, and if alone, whether the y axis is scaled to the one country's own range, or to the range of the whole contextual group.

The goal of this research is to create valid visualizations that manipulate different visualization parameters, and to use these variations to explore which visual cues everyday citizens use to make judgments about a topic that concerns them personally, their safety.

Methods and Procedures

Data

Covid-19 data were downloaded from the Our World in Data website [7], which collects, cleans, curates, and hosts data from multiple reputable sources. Our data set included weekly fatality data for three countries: France, the United States and Peru, measured weekly. We chose fatality data because, unlike data on cases or hospitalizations, it is less likely to be influenced by variations in data collection, and less influenced by biases in reporting. The original source of these data is the COVID-19 Data Repository, Center for Systems Science and Engineering (CSSE), Johns Hopkins University (JHU).

Data Visualizations

Many types of COVID-19 visualizations appear on the internet and in the press, but the most prevalent, by far is the time-series representation, where the magnitude of a dependent variable is plotted over time. To decide which parameters to select for our experiment, we studied many exemplars, and categorized their parameters. For each graph, there were certain set-up parameters, such as which dependent variable was being represented, the number of entities shown (e.g., multiple countries, age groups, etc.), and the type of visualization (e.g., line drawing, mountain). There were choices for the X-axis, such as the time interval (e.g., 6 months, 2 years) or normalization (e.g., days since 100th case), and aggregation strategy (e.g., daily, weekly, 3-week moving average). For the Y axis, selections were made for scale (e.g., linear, logarithmic), the range (and how it was set), and any normalization (none (e.g., total cases) vs. per capita, e.g., rate/million).

Based on this survey, we selected the most frequently used representations, and used the Our World in Data end-user tool to create 48 exemplars. All graphs used the same dependent variable, COVID fatalities. We visualized data from three countries: France, the United States, and Peru. Our stimuli varied along 6 parameters. Examples are shown in Figures 1, 2, 3, and 4.

- X axis accumulation: Weekly *vs.* Cumulative
- X axis intervals: 6 months *vs.* 2 years
- Y axis scaling: Linear *vs.* Logarithmic
- Y axis normalization: Total *vs.* Per Capita (fatalities per million)
- Data from one country alone (France) *vs.* that country within the context of two others (the US and Peru)
- Y axis scale: Scaled to highest value for an individual country *vs.* Scaled to highest value for the Group

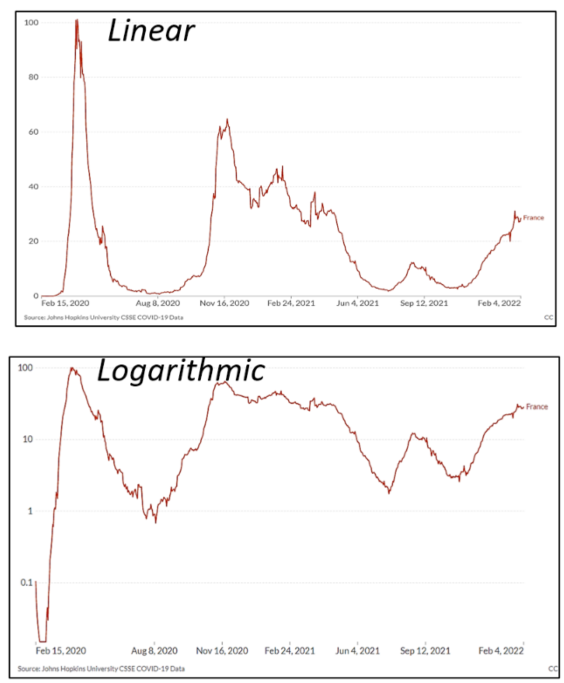

Figure 2. Linear vs. Logarithmic. *Magnitude of the data is commonly represented linearly, that is, with equal data values corresponding to equal distances along the y axis. Data are also commonly represented logarithmically, where equal ratios correspond to equal distances along the y axis. Although both representations correctly convey the magnitude of the data, the impression gleaned by the observer may be different.*

Experimental Design

We used Google Forms to create an online survey that contained 48 randomized visualizations based on these 6 parameters. For each visualization, the observer responded to two questions.

1. How safe would you feel travelling to France?
2. How confident are you about your judgment?

Each question was responded to on a 5-point Likert Scale. The scale ranges were denoted in quintiles (e.g., 0-20%; 20-40%, 40-60%, 60-80%, and 80-100%). This scaling was used so that we could easily separate out negative responses (0-40%) from neutral responses (40-60 %) and positive responses (80-100%).

After the survey, which took 20-30 minutes, participants provided information about their experience with data analysis and visualization and answered the 5 line-drawing data literacy questions from the VLAT test. This preliminary study reports on the results of 24 observers, drawn from our professional circle.

Hypotheses

The graphs judged in these experiments were all drawn from three columns in a spreadsheet. For three countries, The United States, France, and Peru, the rows contained mortality data over a two - year period. If the design choices used to create visualizations of these data had no effect on the impressions gleaned by these subjects, then they should feel equally safe, independent of which rendering was presented. This is the null hypothesis for these.

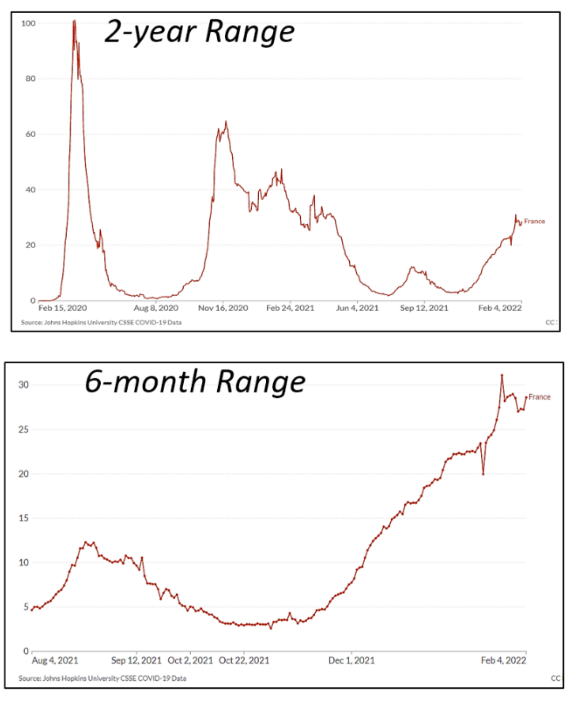

Figure 3. X-axis range. *Different visualizations will represent the data over different time frames. For our experiments, we selected a 2-year time frame, and compared it a recent 6-month range. The rise in cases in France is perceived quite differently when seen within a 2-year context.*

We hypothesized that different renderings would, indeed, generate different responses. First, we expected that *visual characteristics* experiments: independent of the rendering, all 48 charts should feel equally safe, since they plotted the same data, using the same plotting conventions for line drawings would play an important role. If the curve rising steeply at the end of the range, the impression is that fatalities are growing, thereby feeling less safe, and if the line appeared to drop, or remain constant, the apparent risk would feel weaker. Second, we expected that people would have difficulty interpreting logarithmic plots, even though all graph axes were labelled. A constant increase in mortality looks like an accelerating line in linear space, but like a flat line in log coordinates. We also predicted that plotting the normalized fatality rate would produce more veridical judgments than the raw data, especially when data from multiple countries was plotted in the same graph. Additionally, we hypothesized that the judgments people made would be independent of their performance on a visualization literacy test, because making an emotional judgment about a trend is fundamentally different from simply being able to read its values. We expected people sophisticated in data analysis and visualization to be better at making these judgments.

Preliminary Results

Our first observation is that the participants' estimation of perceived safety varied from graph to graph, showing that their judgment was, indeed impacted by the way it was rendered.

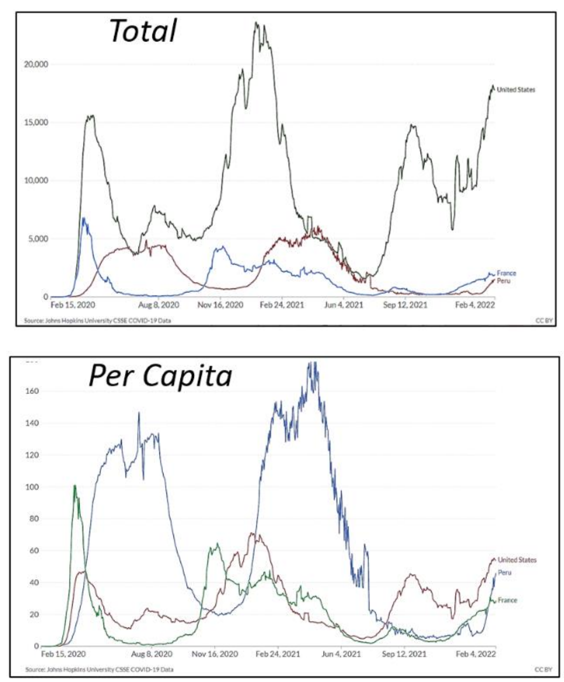

Figure 4. Total vs. Per Capita.. *Normalization is a common method used for comparing variables. The total fatality count is the number of COVID fatalities and the per capital value is the number of fatalities per million residents, that is, the proportion of the population affected. Note that the total number of Fatalities in France was lower than in the US, the per capita rate far exceeded the two comparison countries during the two peaks, and appears to be more similar by the end of the recording interval than when raw numbers are plotted.*

We next looked at the main dimensions along which these visualizations varied. For each graph, we counted the proportion of observers who felt highly safe (rating in the 80-100% range) or highly unsafe (rating in the 0-40% range), and looked at how these judgments varied by condition. For our small set of observers, people felt safer with linear Plots than logarithmic, and safer with raw data than with per capita plots. They felt safer looking at the most recent range, and safer with other countries' data plotted for comparison.

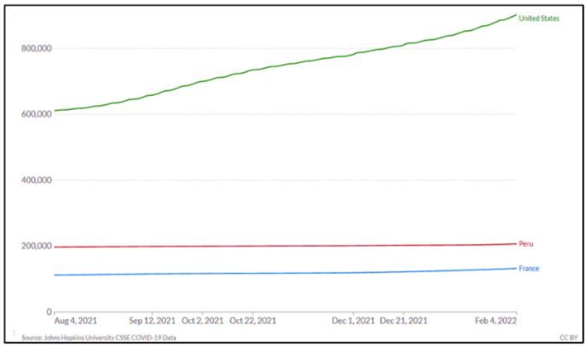

Figure 5. "Safest" graph. *These data show linear, cumulative raw fatality data over the most recent data range. In this period, fatalities were growing in the US, but were flat in France. Participants felt safest in this condition.*

Figure 5 shows the graph that made people feel safest about travelling to France. The graph plots raw fatality data for France in the context of the plots for the US and Peru. There are several reasons why this graph may have felt the most safe. Perhaps having a frame of reference was important, since France is clearly lower than the other two countries over this timeframe. However, this may be a false impression, since these data are not normalized. Yes, the US has 4x more fatalities, but it also has 7x the population. Or, perhaps it was the fact that the number of additional deaths over this timeframe appears to be flat, so although there are fatalities, the accumulation appears to be slow, especially compared with the US. Our ongoing research is aimed at understanding these judgments, how they interact, and which visual cues drive these decisions.

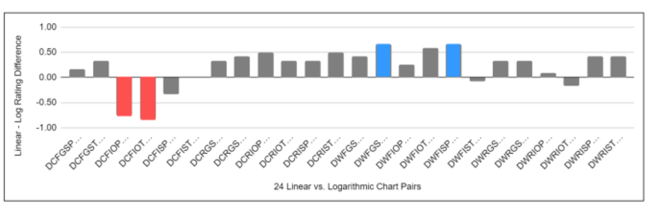

Figure 5 Linear vs. Logarithmic. *This graph shows the difference in safety ratings for linear (above 0) and logarithmic (below 0) for the 24 linear vs. log pairs of visualizations. S's felt safer in 17 of these comparisons, with most showing at least a 0.5 point rating difference. Blue and red colors highlight the plots greatest differential ratings for log and linear, respectively.*

As a first step toward understanding these interactions, we looked more closely at the Linear vs. Log comparison. In a recent study, Sevi, et al.[10] presented a single time-series of COVID-19 data, plotted with either a linear or logarithmic y axis, and found no differences in participants' judgments about the need for a lockdown. In our experiment, we considered many more conditions, including various aggregation strategies and context. To assess the impact of log vs. linear, we computed, for each observer, the difference between the rating they provided for the 24 linear graphs relative to the 24 matched logarithmic graphs, which were identical in all other parameters. Mean ratings are shown in Figure 5. In 17 of 24 combinations, the observers felt safer when the data were plotted on a linear scale than when plotted on a logarithmic scale. The dark blue bars indicate the two conditions that showed the greatest advantage of linear over logarithmic. In these cases, France's data were plotted weekly, in the context of the other two countries, and scaled to the range of the country with the greatest maximum. This advantage is not ubiquitous, however, The red bars show to cases where the logarithmic transformation afforded a stronger feeling of safety. In these plots, France's data were plotted cumulatively, and alone, and scaled to their own range. And, consistent with [10], there were some cases where no difference was observed between log and linear representations.

Conclusion

This paper reports preliminary results based on 24 observers, and our analysis is also very preliminary. Still, we are encouraged by the apparent strength of our observations. Legitimate visualizations of the same data produced very different impressions, despite the fact that all the graphs were created using the same style, and conformed to best visualization practices [8,9]. Even people with strong data analysis and visualization backgrounds, who performed perfectly on the visualization literacy test, were swayed by specific visualization choices. Continuing work will be directed at characterizing the visual features of the graphs, to better understand which visual features drive the overall impression of the data, and a deeper analysis of how different transformations, normalizations, comparisons, and selections influence judgments.

Although we are just at the beginning of this journey, we feel that there are important implications for policy makers and public health officials. If the visualization designer has many valid ways of representing data faithfully, which of those would best communicate an emerging public danger? We have found that people who score perfectly on data visualization literacy tests, and people with graduate degrees in analytical disciplines, nonetheless judge objectively identical graphs differently, depending on the selected data transformation or normalization. These preliminary results show that more research is needed to understand how chart-drawing choices, and the visual cues they produce, affect our judgment. We think this line of research can lead to a richer model of data understanding, and new methods for assessing what knowledge people infer from data.

References

[1] Zhang, Y., Sun, Y., Padilla, L., Barua, S., Bertini, E., & Parker, A. G. (2021, May). Mapping the landscape of covid-19 crisis visualizations. In Proceedings of the 2021 CHI conference on human factors in computing systems, pp. 1-23, 2021.

[2] Lee, S., Kim, S. H., & Kwon, B. C. VLAT: Development of a visualization literacy assessment test. IEEE transactions on visualization and computer graphics, 23(1), 551-560, 2016.

[3] Rogowitz, B. E., & Goodman, A , Integrating human-and computer-based approaches to feature extraction and analysis. In Proceedings of Human Vision and Electronic Imaging XVII , 8291, 258-268, 2012.

[4] Ziemkiewicz, C., & Kosara, R. The shaping of information by visual metaphors. IEEE transactions on visualization and computer graphics, 14(6), 1269-1276, 2008.

[5] Padilla, L., Fygenson, R., Castro, S. C., & Bertini, E. Multiple Forecast Visualizations (MFVs): Trade-offs in Trust and Performance in Multiple COVID-19 Forecast Visualizations. IEEE Transactions on Visualization and Computer Graphics, 29(1), 12-22, 2022.

[6] Roberts, J. C. On encouraging multiple views for visualization. In Proceedings. IEEE Conference on Information Visualization. An International Conference on Computer Visualization and Graphics (Cat. No. 98TB100246) 8-14, 1998.

[7] Our World in Data web sources : https://ourworldindata.org/coronavirus and https://github.com/owid/covid-19-data/tree/master/public/data

[8] Huff, D. How to lie with statistics. Penguin UK, 1954.

[9] Robbins, N. Creating more Effective Graphs. Wiley Interscience, N.J. 2013.

[10] Sevi, S., Aviña, M.M., Péloquin-Skulski, G., Heisbourg, E., Vegas, P., Coulombe, M., Arel-Bundock, V., Loewen, P.J. and Blais, A., 2020. Logarithmic versus linear visualizations of COVID-19 cases do not affect citizens' support for confinement. Canadian Journal of Political Science/Revue canadienne de science politique, 53(2), 385-390, 2023.

Author Biography

.

Bernice Rogowitz *received her Ph.D. in Vision Science from Columbia University, and was a research fellow in psychophysics at Harvard University. At the IBM T.J. Watson Research Center, she worked on fundamental and applied topics in color, image, pattern, and haptic perception. Her current research includes touch perception, memory in VR and MR, spatial-temporal pattern perception and visualization in the wild. She is a Fellow and Honorary Member of the IS&T, the founding chair of the Conference on Human Vision and Electronic Imaging, and co-Editor in Chief of the IS&T Journal of Perceptual Imaging.*

Paul Borrel *received his Ph.D. from the University of Montpellier, specializing in computer-assisted design (CAD), and worked a CNRS laboratory in France on graphical transformations, projections and coordinate geometries. At the IBM Research, in NY and in Brazil, he pursued research in computer graphics and visualization, haptics, and interactive data integration, and led many projects applying data visualization to real-world problems, including the effect of climate change on water reservoirs.*