Black Hat Visualization

Michael Correll
University of Washington

Jeffrey Heer
University of Washington

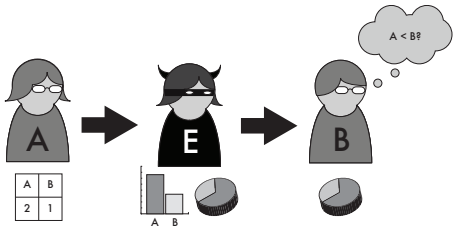


Figure 1: A model of a visualization "attack." Data scientist **A**lex has a dataset they wish to communicate to stakeholder **B**rook. Unfortunately, Alex must go through visualization designer **E**rin, who has malicious intent. Erin has many potential visualization designs at their disposal, but chooses one that is likely to cause Brook to have an incorrect impression of the data.

ABSTRACT

People lie, mislead, and bullshit in a myriad of ways. Visualizations, as a form of communication, are no exception to these tendencies. Yet, the language we use to describe how people can use visualizations to mislead can be relatively sparse. For instance, one can be "lying with vis" or using "deceptive visualizations." In this paper, we use the language of computer security to expand the space of ways that unscrupulous people (black hats) can manipulate visualizations for nefarious ends. In addition to forms of deception well-covered in the visualization literature, we also focus on visualizations which have fidelity to the underlying data (and so may not be considered deceptive in the ordinary use of the term in visualization), but still have negative impact on how data are perceived. We encourage designers to think defensively and comprehensively about how their visual designs can result in data being misinterpreted.

Index Terms: K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 INTRODUCTION

Designers of visualizations have a great deal of power over how data are interpreted. Ideally, designers are interested in accurately presenting data in order to foster responsible and evidence-based decision-making. In the real world, however, designers may have more nefarious goals. In these situations, designers can use visualizations as an *attack vector* to distort how an audience perceives and uses data. We refer to this strategy of altering visualization designs to encourage specific interpretations as "black hat vis." In prior work in the visualization community, many of these black hat techniques are said to generate "deceptive" visualization, or "vis lies" [3]. This language conflates *poor* visualization design and *malicious* visualization design. The binary notion that a visualization is either deceptive or not also elides the subtlety of some of these techniques, which may have an impact on how data are perceived and used without altering the fidelity with which they are presented.

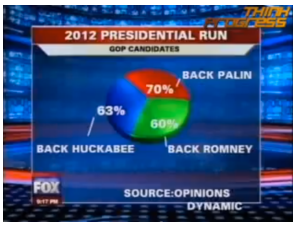
There is extensive prior work on how visualizations can be deceptive, including empirical studies on the efficacy of these deceptions [19]. However, discussions of deceptive visualizations can be ad hoc, built around a limited set of examples. Compared to the rich variety of ways that humans can and have deceived each other (and themselves), visualization has a relatively sparse set of examples of deceptive practices. There are a set of standard examples that are repeated frequently (Fig. 2), giving the impression that deceptive visualizations are the uncommon product of a small set of bad apples. It is our contention that a) the intent to persuade (and, so, potentially, to deceive) is ubiquitous in visualization, and b) the space of known deceptive techniques ought to be expanded to match this ubiquity.

In order to provide an initial framework for expanding the space of deceptive techniques in visualization, we borrow terminology from computer security. In security, "black hats" are attackers with malicious or destructive intent. Visualization designers, as intermediaries between a data set (to which a stakeholder may or may not have direct access) and a wider audience, are well-situated to perform "man in the middle" attacks (Fig. 1). **Black Hat Visualization** is the blanket term we use to describe all of the actions that malicious visualization designers can take to meet their goals. Creating outright "lying" visualizations is just one small category in this larger space of bad behavior. In this paper, we lay out some of the other forms these attacks can take, under the principle that a good defense requires a comprehensive threat analysis. In particular, we focus on four categories of attacks—intentional breaks of convention, data manipulation, obfuscation, and nudging—as examples of more comprehensive ways that visualizations can be used with ill intent. In some cases, these attacks can happen by chance or inexperience. However, if we can defend against the evil, then we can also defend against the incompetent.

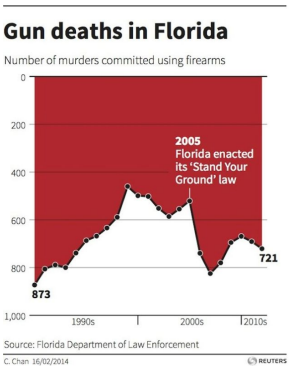
2 BROKEN CONVENTIONS

The interpretation of visualizations is an acquired ability, and this ability is neither monolithic nor universal [7]. For instance, a Pew Research poll on scientific knowledge found that only 63% of Americans could correctly interpret the trends in a scatterplot [12]. For those with graphical literacy, a great deal of the information in a chart is conveyed *implicitly*, through the conventions of the medium [15]. For instance, we expect a pie chart to convey a part/whole relationships, and thus for its components to add up to 100%, and are surprised when the components do not (as in Fig. 2a). We expect that the heights of bars in a bar chart are proportional to their value; truncated y-axes defy this expectation (as in Fig. 2c). People who predominantly write left-to-right expect time to mapped from left to right on the x-axis [5, 24], and so on. By defying these implicit conventions, an attacker can create a visualization which is technically "correct" (in that it is a valid and straightforward mapping from data to mark), but still misleads.

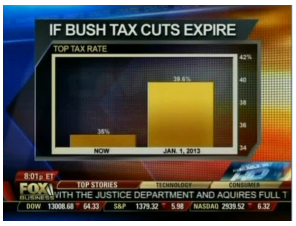
Another form of attack that piggybacks on established visual conventions is to have no relation between data and mark (as in



(a) A "lying" pie chart



(b) A "lying" line graph



(c) A "lying" bar chart

Figure 2: Examples of lying visualizations that are nearly ubiquitous in the discussion of deceptive visualizations. A reverse Google image search reveals millions of hits for each visualization, used in sites with titles ranging from "How to Lie with Data Visualization" to "6 Data Visualizations That Failed at Life."

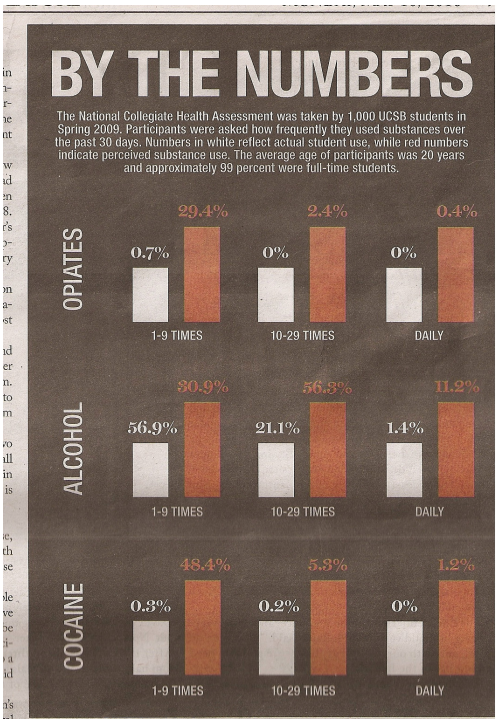
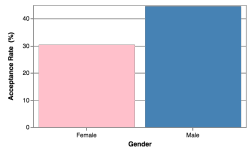
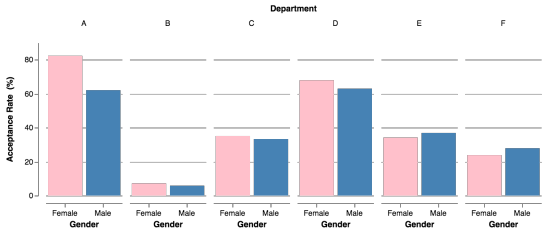


Figure 3: A set of non sequitur bar charts [2]. The visualization follows the visual conventions of bar charts, but there is no connection between the data values and the height of the bars.



(a) Overall university acceptance rates, with a bias in favor of men.



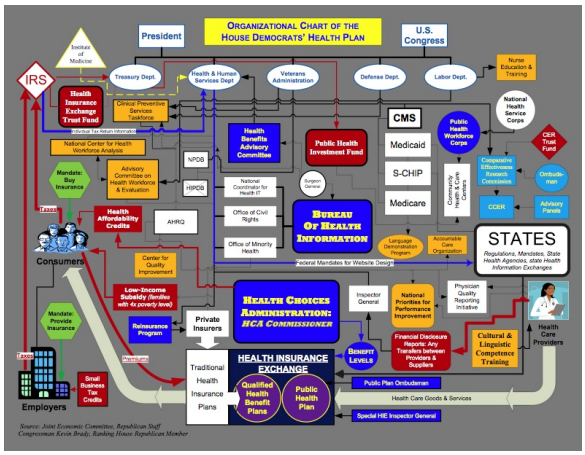
(b) Per-department acceptance rates, with a bias in favor of women in all but two departments.

Figure 4: Choices in how data are aggregated can completely alter the messages conveyed in visualizations. Here, the same data can produce bar charts with two seemingly opposite conclusions.

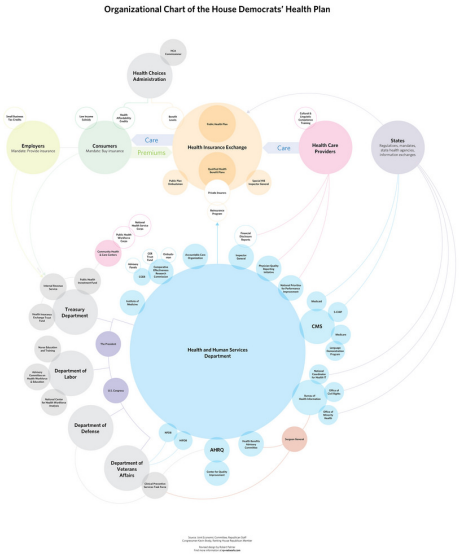
Fig. 3). In the wild, many charts and graphs are simply conceived of as "number decoration" [22]. Charts are frequently associated with scientific evidence and other trustworthy sources of information. Charts can therefore be persuasive even if they are superfluous or uninformative [23]. A black hat designer can take advantage of this fact by choosing not to directly visualize data that may be contrary to their intended goals, but instead representing it with "non sequitur" visualizations. A non sequitur visualization could consist of "decorative" bar charts or pie charts with arbitrary values, and the actual (unfavorable) data values represented textually elsewhere.

3 Data Manipulation

If the attacker has full control over the data, then the potential for harm quite high. Similar to the non sequitur attacks above, the attacker can arbitrarily change the data and then make visualizations that faithfully present these false values. However, even if we presume that the individual data values are unaltered, there are still a number of ways that a designer can create harmful visualizations, by making malicious choices in how the data are aggregated, summarized, or filtered. Visualization designers are often given "raw" data, and are free to these choices themselves, and thus have a great deal



(a) The original Republican flowchart of the ACA [4].



(b) Robert Palmer's cleaner version [17].

Figure 5: An example of intentional obfuscation in visualization design. Republicans in the U.S. House of Representatives produced a flowchart purporting to show the structure of the plan that would become the Affordable Care Act (ACA). Clashing colors, edge crossings, and busy fonts contribute to the impression that the plan is overly complex and impossible to understand. Robert Palmer, using the same topology information, produced a competing graphic titled "Do not fuck with graphic designers" as a passionate response to this intended obfuscation.

of power over what messages a viewer can take away from the data.

Simpson's paradox is one example (Fig. 4) of this power. The same college admissions data can be used to show a bias either for or against women. This apparent contradiction is due to the fact that women were more likely to apply to departments with lower overall acceptance rates, and vice versa. By choosing how to aggregate and normalize the same data, the designer of the visualization can control the message of the chart. Presenting the full range of counts, percentages, and groups to circumvent these sorts of paradoxes requires visualizations more complex than simple histograms [20]. Data preparation choices related to aggregation and normalization, such as whether or not to discard outliers, or whether or not to smooth data, can also dramatically affect how data are visualized and subsequently interpreted.

4 Obfuscation

Another form of attack is to intentionally obscure the data values. That is, to make the process of visually extracting particular data values as arduous as possible. Even if the obfuscation does not make data illegible, a successful attack can make people less confident in the data, or their own understanding of the data. In some cases, presenting certain data as too complex to parse is a form of attack in and of itself (as in Fig 5).

Another attack which relies on obfuscation is to hide relevant information amongst a sea of less relevant or less damaging information. Dashboards and infographics often contain a dozen or more separate visualizations. If certain views or facets contain information contrary to the intended message of the attacker, then they can be placed in locations that require extra effort from the viewer to read (for instance, they require zooming, panning, or scrolling to encounter). Alternatively, black hat designers could attempt to counter solitary charts with damning information by including a large number of charts with supportive information. Even if the data in these supporting charts provides only weak or spurious evidence in support of the attacker's goal, their sheer number may mentally outweigh or exhaust the viewer (in argumentation, this technique is sometimes called a "Gish Gallop" [1]).

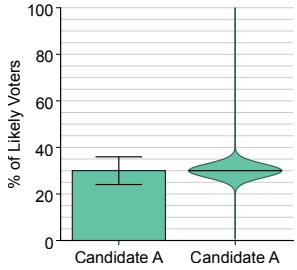


Figure 6: An example of a nudge, on polling data. The bars in bar charts create a visual metaphor of containment. This causes outcomes that would fall inside of the bar to perceived as liklier than values outside of the bar. Other visualizations lack this bias. In this sample polling data, viewers were more confident that Candidate A would underperform when presented with a bar chart (as on the left) than with a violin plot (as on the right) [10].

5 Nudges

The last form of attack we discuss is where the design of the visualization is altered to emphasize or de-emphasize certain patterns in the data. That is, the viewer is nudged toward or away from a particular interpretation of the data. These attacks are common in marketing; for instance, consumers are more likely to underestimate the price of goods when the price ends in 99 rather than a round number (i.e., a product costing $4.99 is more likely to be misremembered as costing less than a product costing $5.00) [21]. However, these nudges are also present in visualizations. For instance, points in scatterplots are perceived as being more highly correlated when they are compressed in scale [9]. Red glyphs in choropleth maps are perceived as having more area than green glyphs [8]. Points within the visual area of bars in bar charts are perceived as likelier

than points outside of the bars [16]. Estimates of y-intercepts in area charts are lower than in scatterplots [11]. Attackers can take advantage of one or more of these nudges to systematically influence how data are perceived (as in Fig. 6).

Although the impact of these individual nudges can be minor, these nudges could be compounded together for larger potential effects. Unlike the prior attack strategies, which can be easy for visualization experts to spot and correct, nudging can be subtle to identify, and (in some cases) even unavoidable. The interplay between visualization and decision-making is complex; minor tweaks in visual design can be sufficient for major changes in decision-making [14].

6 Discussion

As the collection and presentation of data becomes a larger part of peoples' lives, visualization becomes more and more important. This importance has a moral component: all powerful technologies have the potential for both use and abuse. We as a field should consider the full space of how visualizations can be misused. Performing this sort of analysis requires putting ourselves in the mindset of bad actors, and to consider the persuasive and rhetorical content of the visualizations that we make, even with the best of intentions [13]. Enumerating these attacks offers the promise that we can detect them, and correct them where they are encountered.

There are three avenues of future work related to black hat vis. The first is taxonometric: we should enumerate the ways that visualizations can deceive or mislead. The examples presented in this paper, and in other works, are a fraction of this enormous space. The second avenue is empirical: rather than taking it on faith that these attacks are or are not effective to our target audiences, we should collect empirical evidence on the size and reliability of these effects (as per [6, 18]). The last avenue is detection and prevention of bad behavior: "white hat vis," using our analogy. There are many automated tools for generating and interacting with visualizations. To encourage the responsible use of these tools, we need to develop methods for detecting and correcting for their misuse. This "defensive design" could take the form of automated assistants (such as systems to detect and correct suboptimal color maps [20]), or could exist as codified design principles and best practices. However, at a minimum, we ought to bake in more supervision and guidance into visual analytics systems, through smart defaults, automated feedback, and explicit warnings to users and consumers of visualizations.

Acknowledgments

This work was supported by a Moore Foundation Data-Driven Discovery Investigator award.

References

[1] Gish gallop — rational wiki. http://rationalwiki.org/wiki/Gish_Gallop.

[2] UCSB by the numbers. https://twitter.com/EagerEyes/status/13821850078. Original source unknown.

[3] Vis lies 2016. http://www.vislies.org/.

[4] Organizational chart of the house democrats' health plan. http://voices.washingtonpost.com/ezra-klein/assets_c/2009/07/jecchart.html, 2009.

[5] B. K. Bergen and T. T. C. Lau. Writing direction affects how people map space onto time. *Frontiers in psychology*, 3, 2012.

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

[7] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics*, 20(12):1963–1972, 2014.

[8] W. S. Cleveland. A color-caused optical illusion on a statistical graph. *The American Statistician*, 37(2):101–105, 1983.

[9] W. S. Cleveland, P. Diaconis, and R. McGill. Variables on scatterplots look more highly correlated when the scales are increased. Technical report, HARVARD UNIV CAMBRIDGE MASS DEPT OF STATISTICS, 1982.

[10] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics*, 20(12):2142–2151, 2014.

[11] M. Correll and J. Heer. Regression by eye: Estimating trends in bivariate visualizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1387–1396. ACM, 2017.

[12] S. K. Goo. The art and science of the scatterplot. http://www.pewresearch.org/fact-tank/2015/09/16/the-art-and-science-of-the-scatterplot/, 2015.

[13] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics*, 17(12):2231–2240, 2011.

[14] O. Inbar. Graphical representation of statistical information in situations of judgment and decision-making. In *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore!*, pp. 265–268. ACM, 2007.

[15] H. Kennedy, R. L. Hill, G. Aiello, and W. Allen. The work that visualisation conventions do. *Information, Communication & Society*, 19(6):715–735, 2016.

[16] G. E. Newman and B. J. Scholl. Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic bulletin & review*, pp. 1–7, 2012.

[17] R. Palmer. Do not fuck with graphic designers. https://www.flickr.com/photos/robertpalmer/3743826461, 2009.

[18] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini. The persuasive power of data visualization. *IEEE transactions on visualization and computer graphics*, 20(12):2211–2220, 2014.

[19] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1469–1478. ACM, 2015.

[20] J. Poco, A. Mayhua, and J. Heer. Extracting and retargeting color mappings from bitmap images of visualizations. *IEEE transactions on visualization and computer graphics*, 2018. "to appear.

[21] R. M. Schindler and A. R. Wiman. Effects of odd pricing on price recall. *Journal of Business Research*, 19(3):165–177, 1989.

[22] M. Stefaner. There be dragons: dataviz in the industry. https://medium.com/visualizing-the-field/there-be-dragons-dataviz-in-the-industry-652e712394a0.

[23] A. Tal and B. Wansink. Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*, 25(1):117–125, 2016.

[24] B. Tversky, S. Kugelmass, and A. Winter. Cross-cultural and developmental trends in graphic productions. *Cognitive psychology*, 23(4):515–557, 1991.