Statistical Quality in Data Visualization

Steven S. Klement[1], Rosemary Byrne[1]

[1]*United States Bureau of the Census*

Abstract

Today's powerful graphics software packages can create intricate static or interactive visualizations of data that were unheard of a short while ago. Many statistical agencies are embracing the newest and most powerful techniques to make intriguing visuals, hoping to entice people to use their data products. Meanwhile, the literature is trying to determine which types of visualizations convey the story of the data to the user and which are confusing, or worse, misleading.

The US Census Bureau has been developing statistical standards for data visualizations for a decade. The current version has been in beta mode for about three years, as development was slowed during the COVID shutdowns. The key issue we are dealing with is that a standard is normally something that you either 'shall do,' or something that you 'shall not' do. For example, John Tukey and others have said that use a pie charts should never be used. Based on this expert advice, this dictate could be a standard. However, what if there is a case where a pie chart would be appropriate? We now have a quandary as to how to author a standard that states that you shouldn't use a pie chart and, at the same time, states you can use one.

This paper will lay out the issues we dealt with in getting our standards to their current state, how we resolved these issues, and what may lie ahead in the future. We hope to promote discussion in the international statistical community as to how others are dealing with statistical quality in data visualization.

Keywords: Data Visualization, Statistical Standards, Quality Standards, Visualization

1. Introduction

Visualizations of statistical data started to appear in the 17th century.[1] The field developed slowly over time. Some early works, such as Charles Minard's map of Napolean's Russian Campaign of 1812,[2] are still held up as excellent examples of the use of data visualization to tell a story using multiple categories in one illustration. With the advent of the computer and graphic design software, the types and intricacies of possible visualizations have become almost limitless. This situation requires statisticians to ask themselves, "Just because we can do something, should we, and would doing so show the actual statistical relationships we are trying to convey, in ways that are easily grasped by a specific audience?"

A way to ensure that statisticians stay focused on data storytelling is to create standards to guide visualization development. For those new to creating data visualizations, having a reference will

[1] Mackinlay, Jock D., Winslow, Kevin; "Designing Great Visualizations". https://www.tableau.com/sites/default/files/2022-09/designing_great_visualizations.pdf
[2] Kraak, Menno-Jan (2014); "Mapping Time: Illustrated by Minard's Map of Napoleon's Russian Campaign of 1812". ESRI Press

assist in ensuring that the story behind the data is not misconstrued. Statistical standards have been used in the survey world for decades in order to guide statistical agencies in the development of estimates. With the increasing use of data visualizations, there is a similar need for visualization standards. There is also a need for providing best practices as reference for data visualization creators which we will discuss in section 2.3. The United States Census Bureau (Census) has been working to develop such standards for well over a decade. The effort has been slow due to the nature of how standards are traditionally written as well as the continual evolution of possibilities and best practices in computer-driven visualizations.

In simple terms, a standard is a rule that tells what a person either 'shall do' or shall not do.' It is supposed to be clearly written so that the reader knows what the boundaries are for the particular action they are about to take. For example, a standard may state that your data product shall compute a response rate with a particular formula and that the response rate must be published along with the data. Similarly, a standard may state how an error estimate must be calculated for a certain type of data collection. These types of standards are specific and leave no doubt as to what the statistician must do in a certain case. Standards can be written more loosely in some instances, but they still set guardrails that shall not be crossed. To take another example, the response rate or error estimate may be required, but the standard may leave it to the statistician to determine in which way it is calculated. However, even in this case, it would be required that the statistician use a method that is statistically valid and accepted in the field. In general, the goal of a standard is to be precise but not overly proscriptive.

In the process of developing standards for visualizations at Census, we have found that while standards are necessary, they are often very hard to put into the 'shall do' or 'shall not do' category. For instance, it's popular to eschew pie charts for very sound reasons, but the reality is that people are familiar with pie charts and tend to like them. Also, in a complex visualization with multiple bar charts, simple pie charts can add a visually distinctive and pleasing diversity to the look and feel of the visualization. Ultimately, we chose to provide guidance steering data visualization creators away from pie charts, but we do not forbid them. As a team, we quickly realized that the field of data visualization is not guided simply by statistical principles. As with pie charts, we use the instrumental work of pioneers in measuring visual perception such as Cleveland and McGill[3] to provide justifications for our guidance. While the visualization must

[3] *Cleveland, William; McGill, Robert (1984). "Graphical Perception and Graphical Methods for Analyzing Scientific Data". Journal of the American Statistical Association.* ***79*** *(387): 531–544. doi:10.1080/01621459.1984.10478080.*

convey the meaning behind the data and respect the underlying statistical distribution, we must also consider how to depict the data in ways to maximize pre-frontal processing over very intensive mental calculations. The statistician must consider the aesthetics of the visualization and whether it helps or hinders understanding of the data. In this paper, we'll discuss Census's effort to develop data visualization standards, the current state of the standards, and where we hope to go in the future. We hope this paper will spur a conversation with other statistical agencies and elicit their thoughts and efforts in this area to share knowledge and best practices.

2. Development of Standards

2.1 Early development

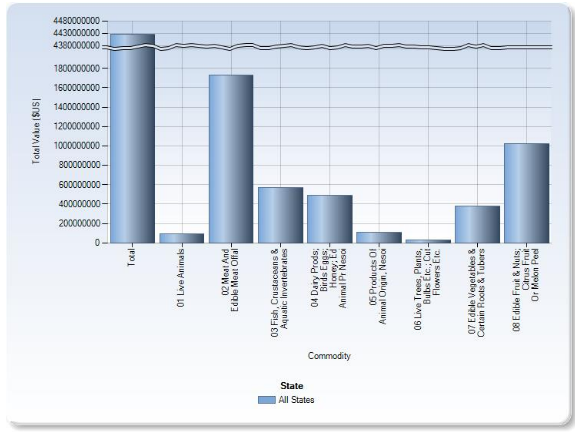
We started our effort by classifying the types of data visualizations that the Census currently uses. We began with the standard bar chart, histogram, and line chart before moving on to data visualizations such as population pyramids and choropleths. The complete list of thirteen visualizations can be found at: https://xdgov.github.io/data-design-standards/. We recognize that there are other possible data visualization types, but these are the most used at the Census and thus the focus of our research and development.

Our team started with what was perceived to be the easiest and most agreed upon standards. The first "shall do" was to mandate including zero on the y-axis of bar charts (or x-axis for column charts). In addition, we mandated the preservation of the area within bar chart comparison rectangles.[4] We can be confident in this standard because it is consistently stated across every authoritative source in addition to being backed by the science of visual perception pioneered by Tufte and others.

An example of the issue that arises when one does not preserve the actual size of the rectangles is illustrated in Figure 1. The data show that the first value is almost 2.5 times larger than the third, but visually the columns representing the data values appear similar in height. Because the human eye is most adept at processing relative rectangular areas pre-frontally, the viewer is easily misled. However, the list of easy to justify standards was quickly exhausted, and the team found they lacked the expertise needed to go further.

[4] Tufte, Edward (2007), *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press. pp.53-77

Figure 1: Illustration of Issue with Comparison Rectangle Size



2.2 Additional Fields of Expertise Needed

Geographers were asked to join the team to lend their expertise. Theirs is a discipline with centuries of experience in applying rules and standards to the practice of depicting three-dimensional reality on a two-dimensional surface. This field has established agreed upon conventions to preserve the integrity of this abstraction. The geographers brought new rigor to the guidance we provide on everything from map legends to the standard that does not allow for raw counts, such as population totals, to be depicted using choropleth maps.

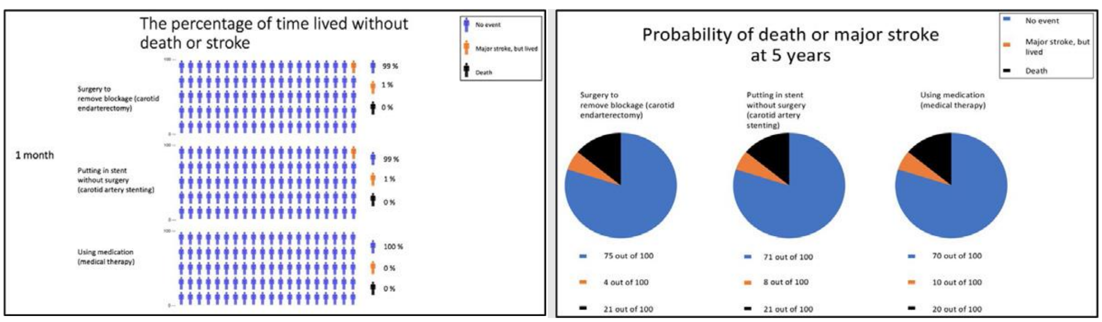
We also added public affairs and communications experts to share their perspective. This group helped balance the desires of statisticians to preserve the statistical validity of the visualization with the knowledge of experts in consistent use of language, Census labeling conventions, and user-engagement metrics. Previously these two sides had opposing viewpoints about the relative importance of these traits. By working together, we forged a path through and gained appreciation for the concerns of all parties.

Our colleagues in the Section 508 Program Office helped advocate for our standards to reflect the best practices mandated to ensure those with visual disabilities can access graphical information through software.[5] They contributed extensively to our discussions around colors and captions. Their expertise had the added benefit that our standards would have compliance incorporated at inception.

[5] https://www.section508.gov/manage/laws-and-policies/

The final data visualization standards team had these specialties along with survey statisticians (survey stats), mathematical statisticians (math stats), and data scientists. Disagreements occurred of course, but we approached this by trying to have open conversations and researching what we could find to support our differing opinions. We strived to back up our decisions with the best science we could find. We knew that we'd far prefer to hear differing opinions at this point in the process than to go live prescribing standards that were flawed. Data visualization is a fast-moving field, and there are many opinions and disagreements about what constitutes the "best science". The prior reference to pie charts is a perfect example. There are those leading the science who state unequivocally that a pie chart should never be used, and while we're inclined to discourage their use, there is also an emerging dialog in the data science field about exceptions to this rule. Peter Scalia, et. al. wrote a compelling study showing an interesting case where patients far preferred seeing cancer outcomes in a pie chart rather than a more complex display.[6] As seen in figure 2, the two charts are telling basically the same story, but the average cancer

Figure 2: Illustration from "Presenting time-based risks of stroke and death for Patients facing carotid stenosis treatment options: Patients prefer pie charts over icon arrays"[6]



patient preferred being shown the pie chart as it was in their view easier to understand. This was the type of information the public affairs team members were able to convey to the team.

2.3 Combining Standards with Best Practices

As we ran out of easily identifiable standards to address, the team quickly realized that there were times we agreed that in most circumstances a developer 'shall not' do something, but in others, it could be a possibility. The Official Census Bureau Statistical Quality Standards are written in such a way as to only include standards and not recommendations or best practices. By limiting the document to this type of guidance there is no confusion as to what is required. In turn, these standards are used by the math stats at Census to ensure that any published Census content

[6] https://www.sciencedirect.com/science/article/pii/S0738399119301715

adheres to these standards. This decision was held up by recent Federal lawsuits challenging the Census' method of conducting various work.[7] Because the Statistical Standards were created by leading professionals in the statistical field it was hard to challenge the validity of Census work that complied with them. This method worked in developing the data visualization standards in some cases (e.g. bar charts and choropleth maps), but not most. There are too many variables to balance and the field is nascent, we felt more confident broadening the scope to more than shall's and shall not's to govern them. Therefore we decided to incorporate best practices, not just standards, into the data visualization standards document.

We incorporated the shall's and shall nots into the existing Statistical Quality Standards and are now creating an accompanying standards and guidelines on our data.census.gov platform, which distributes all Census data products. Here, users can see requirements and recommendations for creating data visualizations. An example of a completed standard that the team developed is the standard for proportional symbol maps, available on our GitHub site:

Requirements

Always

- Include a legend.
- Choose an appropriate symbol for your map.
- Consider the relative size of all symbols. If the symbols overlap, is the map still legible?
- Consider the orientation of the symbol. If the symbol is not a circle, ensure the symbol is oriented properly.
- Choose a method for grouping your data carefully when using a graduated symbol map.
- The Natural Breaks classification method groups data based on gaps in the data.
- The Equal Interval method divides the data range into equal intervals. This method works best when data are evenly distributed. If the distribution is uneven, the predominant values will dominate the map.
- The Quantile method partitions the data so that every group has an equal number or nearly equal number of values. This can give a choropleth map an even, aesthetically pleasing color distribution.
- Subject matter expertise can be used to place data into groups.

Recommendations

Recommended

- Limit the number of classes to five or fewer categories.
- Fewer symbols on the map are more easily distinguished by the eye and will make trends in the data easier to see.

---

[7] https://www.brennancenter.org/issues/gerrymandering-fair-representation/fair-accurate-census/2020-census-litigation

- Use an Albers equal-area projection, including insets of Alaska, Hawaii, and Puerto Rico, if applicable.

Not Recommended

- Don’t have symbols on the map overlap. This can sometimes be appropriate depending on the type of symbol and scale of the map.

2.4 Special Cases

Data Visualization Standards led us to a broader discussion about how we as the standards and guidelines creators can nudge those creating Census visualizations toward earlier consideration of the story they are attempting to tell. While visualization concepts do go through a vetting process as stories are created, the closest scrutiny of the visualizations comes very late in the process. As this review is often very close to a desired publication date, visualization creators are loathe to rethink fundamental approaches to depicting their content. We’ve identified a clear need for an earlier review by visualization experts on proposed visualizations.

Equally as important as careful upfront consideration of how best to tell a given story is the process of ensuring that the visualizations we do create provide value to the viewer. In particular, we’ve identified some traits of our tabular data that present genuine challenges to visualization creators. In many cases, not only do these traits add little value, but we also found them to be mentally taxing and too dense to truly convey what the data show. As part of our guidelines (as opposed to standards), we’re incorporating recommendations for how to display tabular data that is nested and how to handle data in non-mutually exclusive categories.

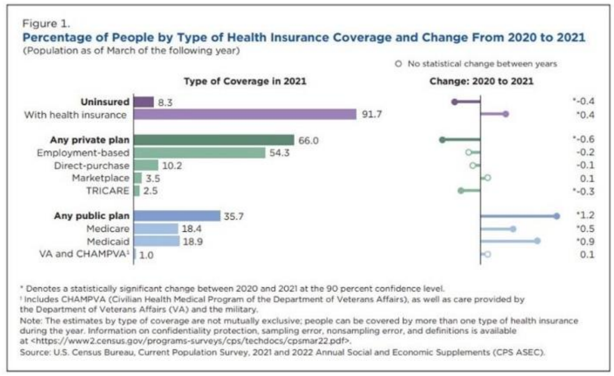
An example of this type of confusing visual can be shown in the case of nested data. In Figure 3, the universe of insurance data is split into those with insurance and those without. Then the category of those with insurance is broken down into various types. In the table on the left, it is relatively easy to understand the nested nature of the data presented. However, there are several miscues in the visualization on the right. First, all three charts use a basic color and font scheme with shading to convey the groupings. The first bar is bolder, and the following bars are lighter. For the green and blue charts, the lighter text and shading imply these are subsets of the first category. But this implication does not hold true for the data in the first chart, which causes cognitive dissonance. This type of issue shows the difficulties the team must address when developing standards. There are so many combinations and permutations of possible types of data that it is nearly impossible to address every situation in a ‘shall’ or ‘shall not’ manner.

Figure 3: Table 1 and Figure 1 Excerpt from Heath Insurance Coverage in the United States 2022

Table 1.
Number and Percentage of People by Health Insurance Coverage Status and Type: 2021 to 2022
(Numbers in thousands. Margins of error in thousands or percentage points as appropriate. Population as of March of the following year. Information on confidentiality protection, sampling error, nonsampling error, and definitions is available at <https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar23.pdf>)

Coverage type	2021				2022				Change in percent (2022 less 2021)
	Number	Margin of error[1] (±)	Percent	Margin of error[1] (±)	Number	Margin of error[1] (±)	Percent	Margin of error[1] (±)	
Total	**328,100**	**148**	**X**	**X**	**330,000**	**130**	**X**	**X**	**X**
Any health plan	**300,900**	**748**	**91.7**	**0.2**	**304,000**	**746**	**92.1**	**0.2**	***0.4**
Any private plan[2,3]	216,400	1,077	66.0	0.3	216,500	1,399	65.6	0.4	-0.3
Employment-based[2]	178,300	1,123	54.3	0.3	179,800	1,369	54.5	0.4	0.1
Direct-purchase[2]	33,550	705	10.2	0.2	32,800	661	9.9	0.2	*-0.3
Marketplace coverage[2]	11,390	447	3.5	0.1	11,840	461	3.6	0.1	0.1
TRICARE[2]	8,299	527	2.5	0.2	7,817	485	2.4	0.1	*-0.2
Any public plan[2,4]	117,100	911	35.7	0.3	119,100	1,183	36.1	0.4	0.4
Medicare[2]	60,230	378	18.4	0.1	61,570	392	18.7	0.1	*0.3
Medicaid[2]	61,940	843	18.9	0.3	62,050	1,112	18.8	0.3	-0.1
VA and CHAMPVA[2,5]	3,151	192	1.0	0.1	3,354	214	1.0	0.1	0.1
Uninsured[6]	**27,190**	**727**	**8.3**	**0.2**	**25,940**	**739**	**7.9**	**0.2**	***-0.4**

* An asterisk preceding an estimate indicates change is statistically different from zero at the 90 percent confidence level.
X Not applicable.
[1] A margin of error (MOE) is a measure of an estimate's variability. The larger the MOE in relation to the size of the estimate, the less reliable the estimate. This number, when added to and subtracted from the estimate, forms the 90 percent confidence interval. MOEs shown in this table are based on standard errors calculated using replicate weights.
[2] The estimates by type of coverage are not mutually exclusive; people can be covered by more than one type of health insurance during the year.
[3] Private health insurance includes coverage provided through an employer or union, coverage purchased directly, or TRICARE.
[4] Public health insurance coverage includes Medicaid, Medicare, CHAMPVA (Civilian Health and Medical Program of the Department of Veterans Affairs), and care provided by the Department of Veterans Affairs and the military.
[5] Includes CHAMPVA, as well as care provided by the Department of Veterans Affairs and the military.
[6] In the CPS ASEC, individuals are considered to be uninsured if they did not have health insurance coverage for the entire calendar year.
Note: Estimates may differ from previous publications due to additional rounding implemented to protect respondent privacy.
Source: U.S. Census Bureau, Current Population Survey, 2022 and 2023 Annual Social and Economic Supplements (CPS ASEC).



2.5 Standard Errors

Guidance for how to handle the visualization of standard errors as well as what rules should exist for data that has very large variance remain the team's thorniest areas. Our existing Statistical Quality Standards explicitly forbid written comparison statements that suggest conclusions not supported by statistical tests. We also highly discourage visualizations that promote conclusions that are not warranted given the measure of error. Nevertheless, many times estimates that may not be statistically significant are put in a visualization. Because of the visual power of charts – particularly bar charts - even a statement cautioning a viewer not to draw conclusions requires the user to ignore what their own eyes observe through pre-frontal processing in milliseconds. So far we have not found a good solution to this problem and are interested to hear if others have considered the relative power of a bar chart to a footnote. We also recognize that for static charts requiring error estimates to be graphically represented can make the chart unwieldly to the point that it obfuscates the point of the visualization.

3. Future Work

The Bureau is planning on finalizing the updated data visualization standards this year. Once complete, it will be released for review and eventual adoption by the Census Bureau Methodology and Standards Council. We plan to make these standards a living document and to update it frequently as different types of visualizations are created and more best practices are developed in the future. It is our desire to find experts in other national and international statistical agencies to further the discussion on the nuances of this topic and develop standards, guidance, and best practices that all can use.