Statistical Quality in Data Visualization

By

Steve Klement*

Rosemary Byrne*

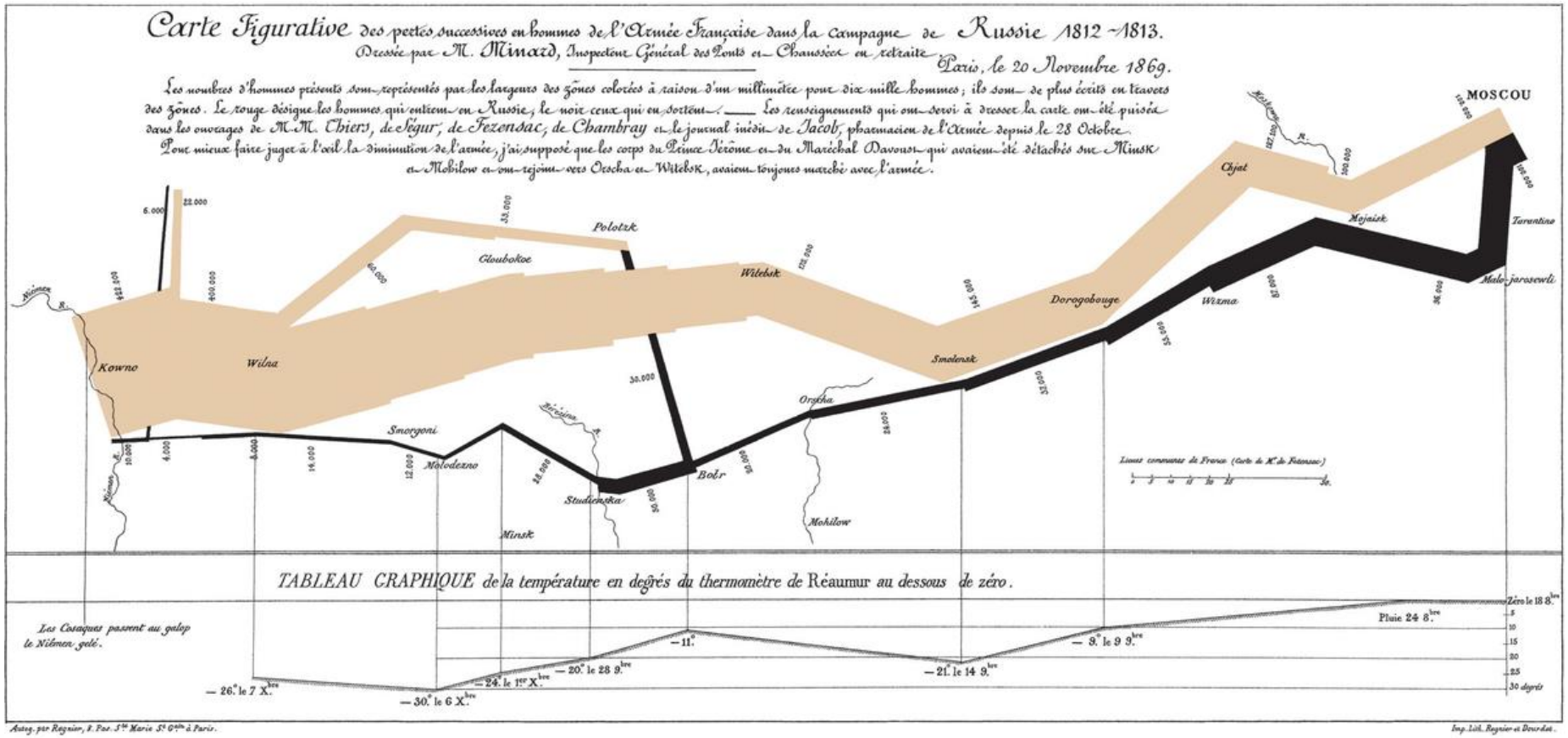
United States® Census Bureau

* Any opinions and expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau.

Early Visualizations

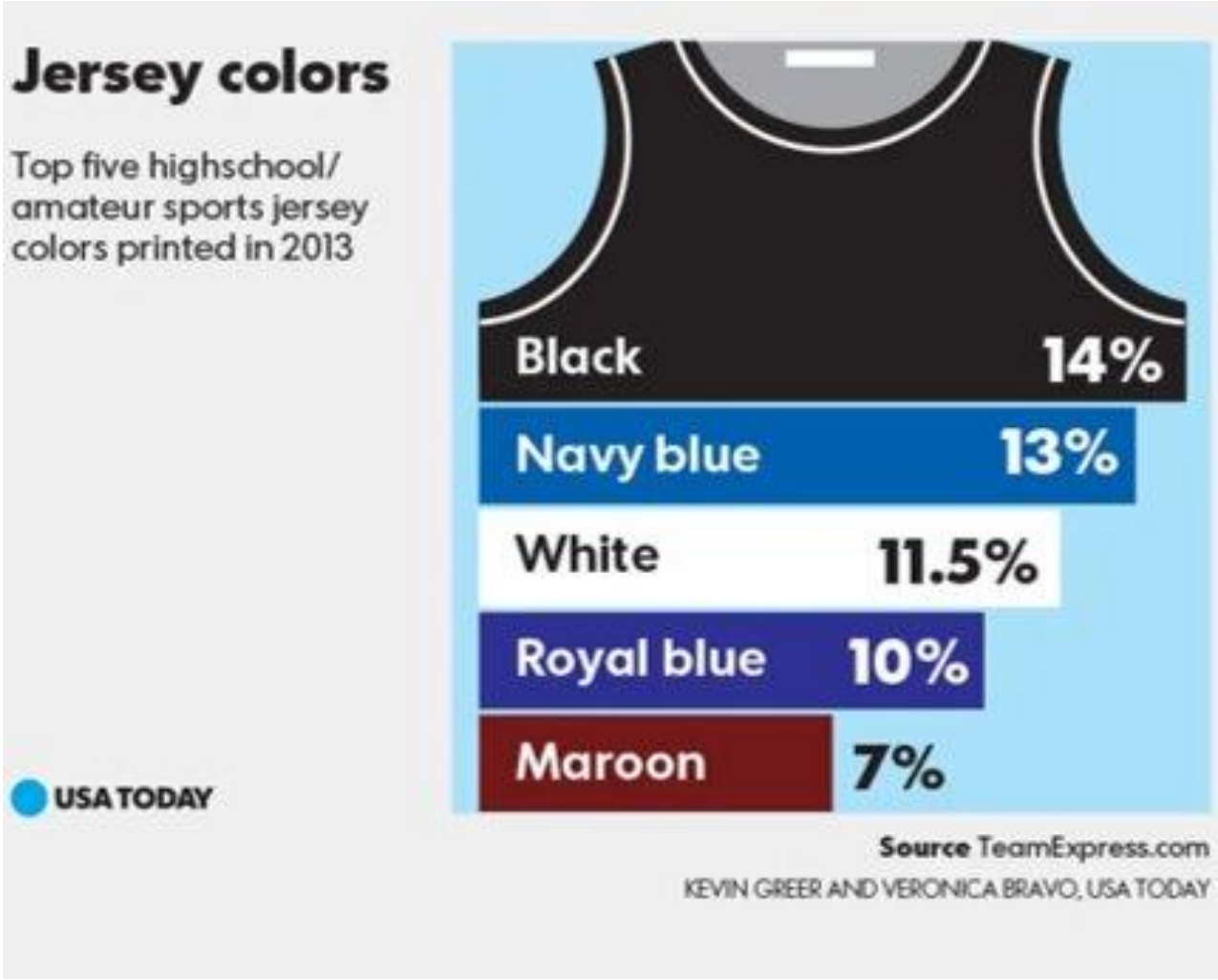
United States® Census Bureau

The Early Age of Data Visualization



And Today We Have This…

- Highly professional
- Easy to understand
- Totally deceiving to the eye



United States® Census Bureau

What is a Standard?

Merriam-Webster dictionary defines “Standard” as:

“...something set up and established by authority as a rule for the measure of quantity, weight, extent, value, or quality”

A standard in brief is either something you ***Shall Do*** or something you ***Shall Not Do***

United States® Census Bureau

Current US Census Bureau Standards

- The Bureau has standards for numerous aspects of our survey work
- We follow all Office of Management & Budget (OMB) statistical policy directives
- Census Bureau Statistical Quality Standards expand upon OMB's Standards and Guidelines for Statistical Surveys
- These cover all aspects of the survey lifecycle
- Census also has standards for disclosure avoidance, paper reviews, etc.

United States® Census Bureau

Standards Pertaining to Data Viz

Edward Tufte in "The Visual Display of Quantitative Information" (1983) states:

- Show the data
- Induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- ***Avoid distorting what the data have to say***
- Present many numbers in a small space
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail, from a broad overview to the fine structure
- Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- Be closely integrated with the statistical and verbal descriptions of a data set.

United States® Census Bureau

Why Data Viz Standards are Needed

- There was an attitude that the more data visualizations we release, the better we serve our users

- There was also a belief that data visualizations that attract attention were good advertising

➡ ***Which could lead to increasing or keeping funding***

United States® Census Bureau

Why Data Viz Standards are Needed (cont.)

BUT this introduced:

- Inconsistencies
  - Over time
  - How different areas treat similar data
- Inaccuracies
- Unintentional missingness: context, variable relationships, other important dimensions
- Dense and confusing content

United States® Census Bureau

Current Data Visualization Standards

- ***Don't exist at the U.S. Federal Level***
- However, many agencies are working to develop them
- Census has been working on the issue for the last decade
- A main issue with developing standards is the Shall/Shall Not aspect

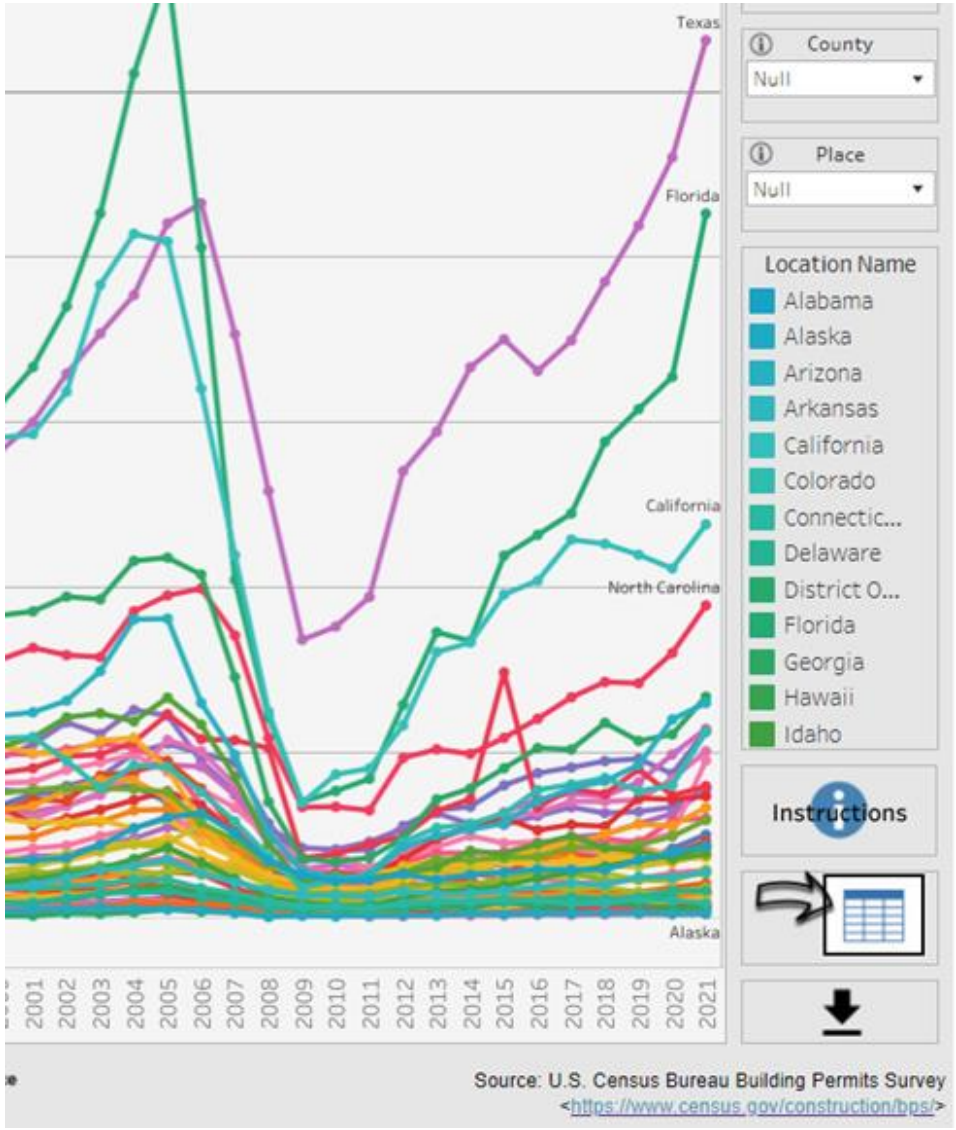
Never use a Pie Chart

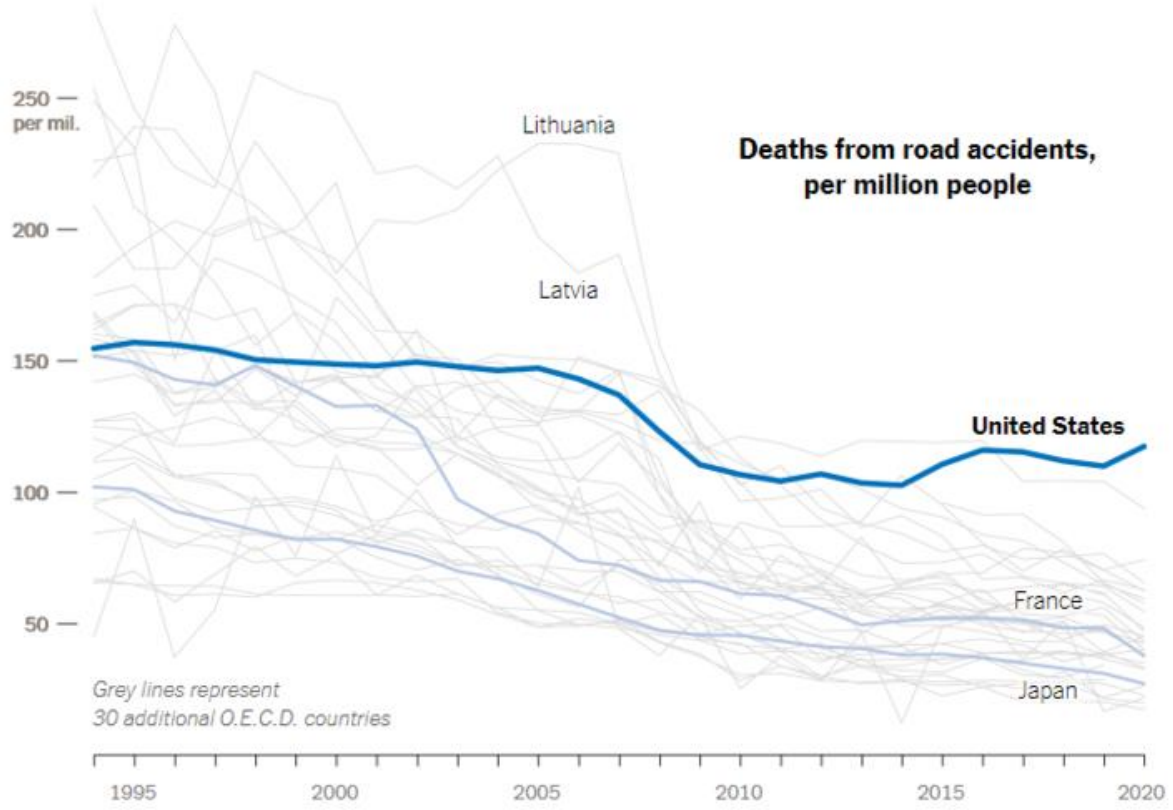
- *Well, except for those couple of cases you can use a pie chart*

Always include a legend

- *But what if direct labeling is a better option?*

United States® Census Bureau





United States® Census Bureau

So, Can There Be Standards?

- Yes, but actually we've found that there are not many true standards that can be applied universally
- Which is why in developing standards, we've realized that what is even more important is giving guidance
  - Give what best practices are out there for different types of visualizations
  - Explain why some ways to do things are better in some situations, but not in other situations
- Show that while there can be standards, it is up to the visual creator to realize that complying with standards does not necessarily guarantee a good visualization

United States® Census Bureau

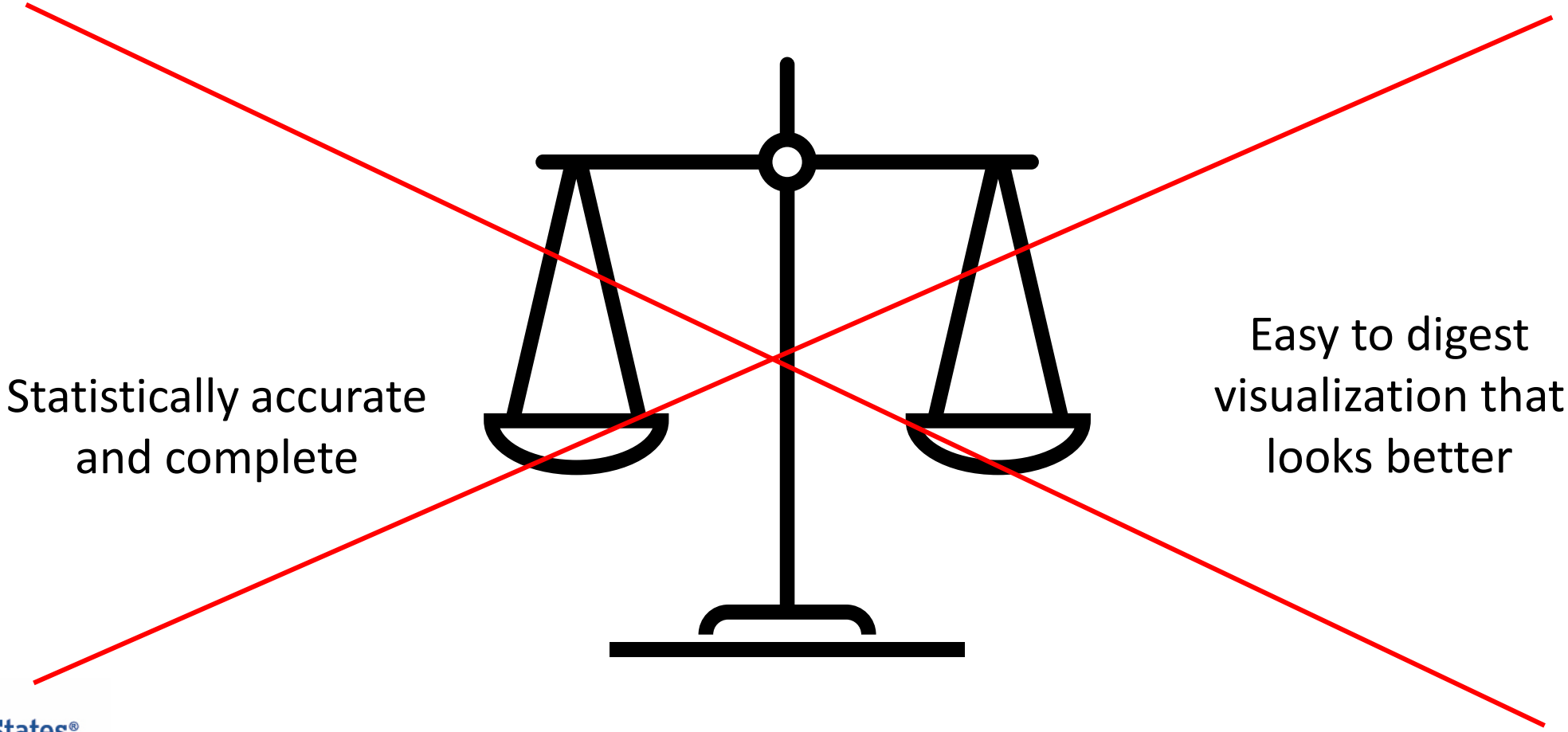
Standards *Do Not* Mean Rigidity

- Allow for creativity and innovation within the bounds of the standards

- Standards and guidance will need to evolve as the science of data visualization advances

United States® Census Bureau

A False Tradeoff



United States® Census Bureau

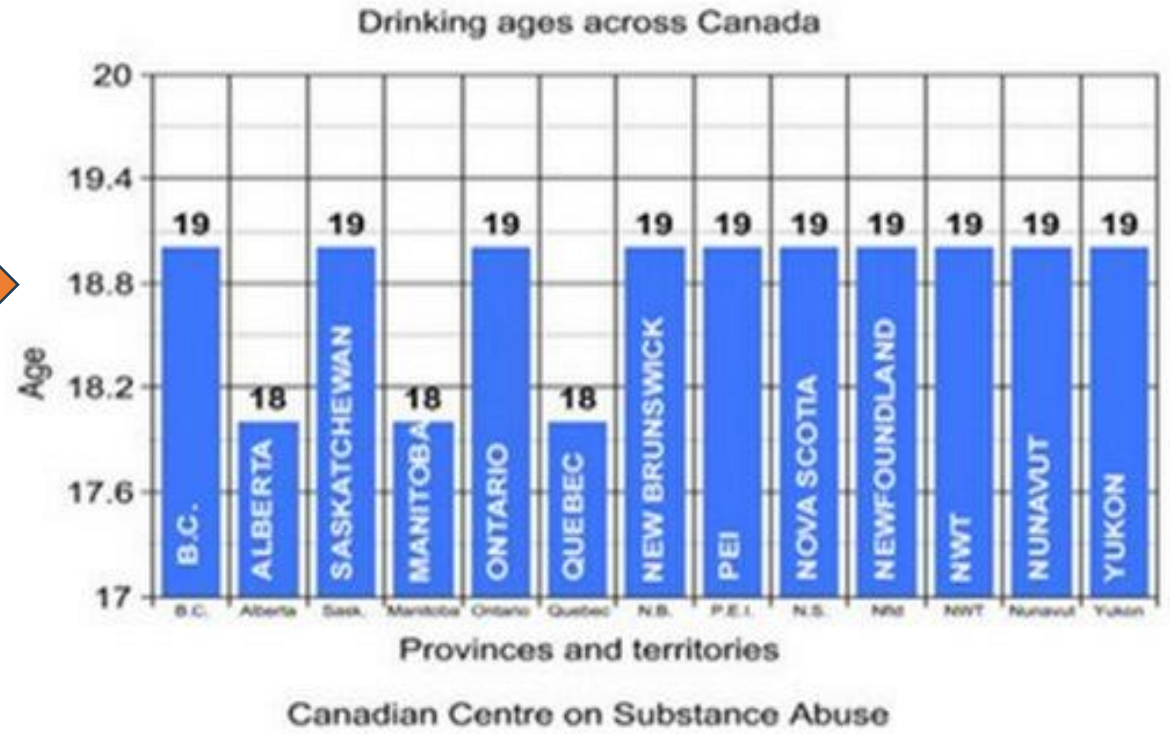
Strengthening the Process is Key

- We all have different strengths! (Take the subject matter expert, geographer, and communication expert)
- Just because you can, does not mean you should
- Just because it looks better doesn't mean it is still meaningful
- Many data viz aspects are not cut and dried yes/no items

United States® Census Bureau

Drinking age will remain 19 in Saskatchewan

CBC News Posted: Mar 4, 2013 11:59 AM CST | Last Updated: Mar 4, 2013 11:55 AM CST 25



You have to be 19 in Saskatchewan to have a drink, while in Alberta and Manitoba, the drinking age 18. (CBC)

The Saskatchewan Party government has ruled out lowering the drinking age, four months after party members put the issue in the public eye.

Facebook
Twitter
Share
Email

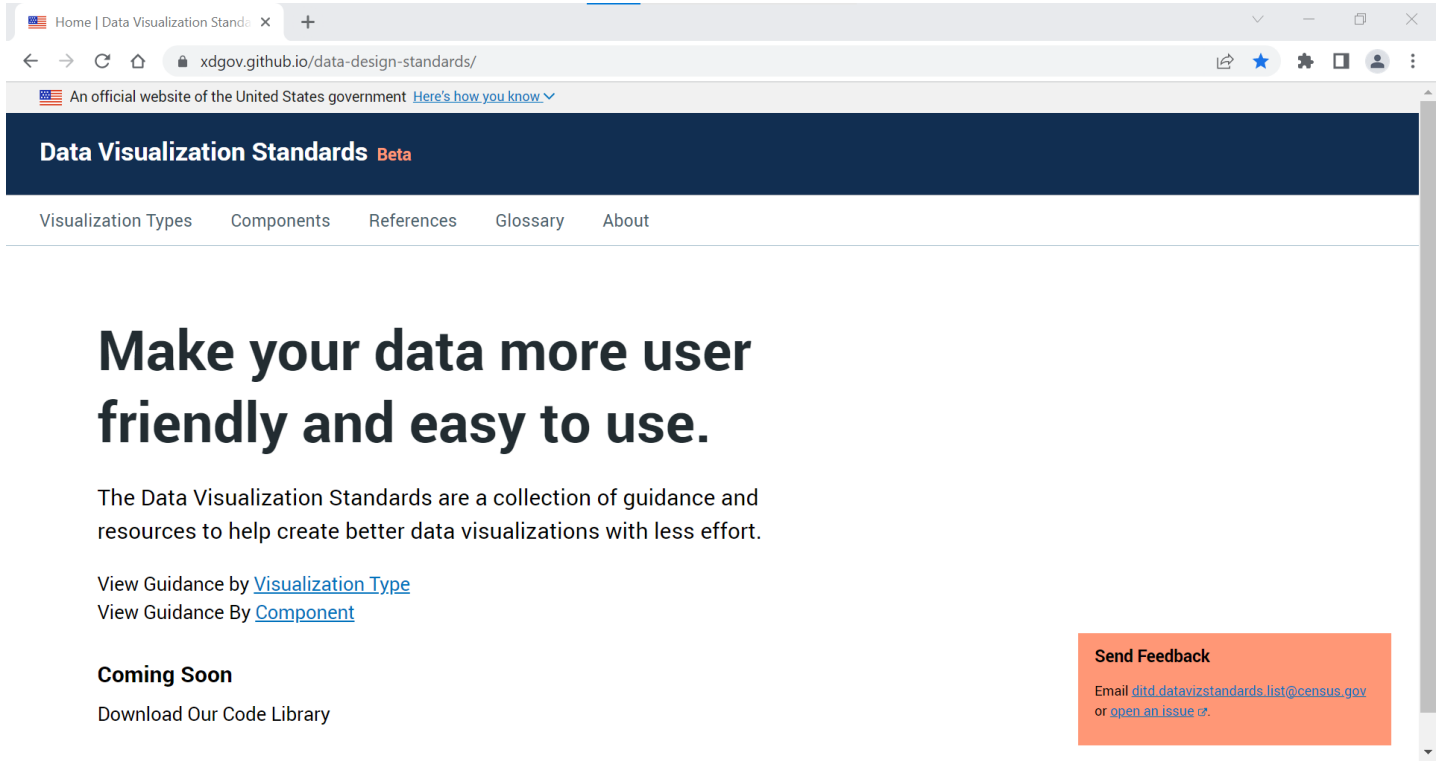
Stay Co
Mobile Fac

United States® Census Bureau

Current Status of Census Standards

- This is our current Beta version of our Data Visualization Standards
- They are laid out by visualization type and include
  - Description
  - Examples
  - Requirements (Shall/Shall Not)
  - Recommendations

https://xdgov.github.io/data-design-standards/



United States® Census Bureau

A Visualization Type Example

Visualization Types | Components | References | Glossary | About

- Bar Chart
- Box and Whisker
- Bubble Chart
- Choropleth Map
- Dot Distribution Map
- Histogram
- Line Graph
- Pie Chart
- Population Pyramid
- Proportional Symbol Map
- Scatter Plot
- Stacked Bar Chart
- Treemap

Bar Chart

Comparative

Also known as: bar graph

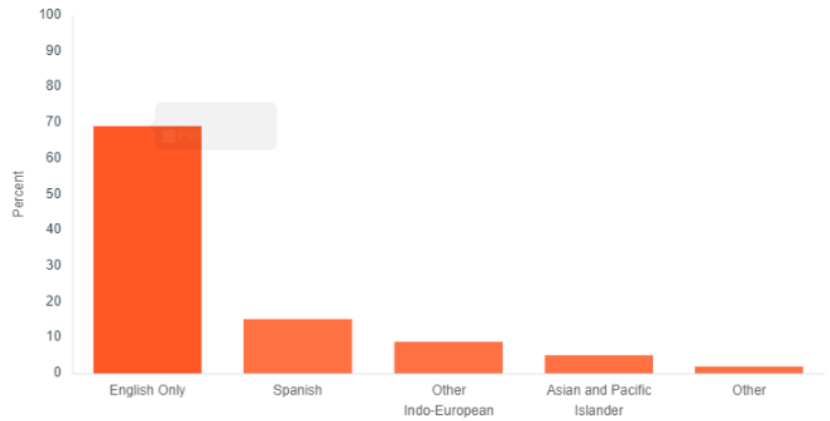
A bar chart displays categorical data with rectangular bars whose length or height corresponds to the value of each data point.

Bar charts can be visualized using vertical or horizontal bars. Bar charts are best used to compare a single category of data or several. When comparing more than one category of data, the bars can be grouped together to created a grouped bar chart.

Bar charts use volume to demonstrate differences between each bar. Because of this, bar charts should always start at zero. When bar charts do not start at zero, it risks users misjudging the difference between data values.

Examples

Types of Languages Spoken at Home in New York



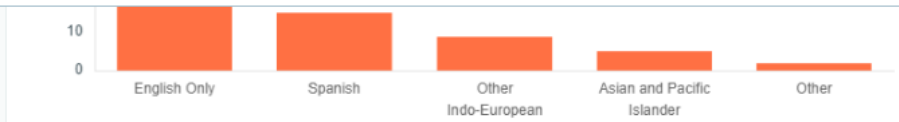
Source: 2016 American Community Survey 1-Year Estimates

United States® Census Bureau

A Visualization Type Example

Visualization Types | Components | References | Glossary | About

- Bar Chart
- Box and Whisker
- Bubble Chart
- Choropleth Map
- Dot Distribution Map
- Histogram
- Line Graph
- Pie Chart
- Population Pyramid
- Proportional Symbol Map
- Scatter Plot
- Stacked Bar Chart
- Treemap



Source: 2016 American Community Survey 1-Year Estimates

Guidance

In addition to the guidance provided in this section, relevant guidance can also be found in the Agency Logo, Axes, Colors, Grids, Labels, Legends, Source, Titles, Typography sections.

Requirements

Always

- Use independent categories of data.
- Label each category of data.
- When visualizing data between 0% and 100%, start tick marks at 0%.

Never

- Never omit the space between bars – otherwise the bar chart will appear to be a histogram.
- Never use three dimensional (3D) graphics as they distort the visual calculation of volume.

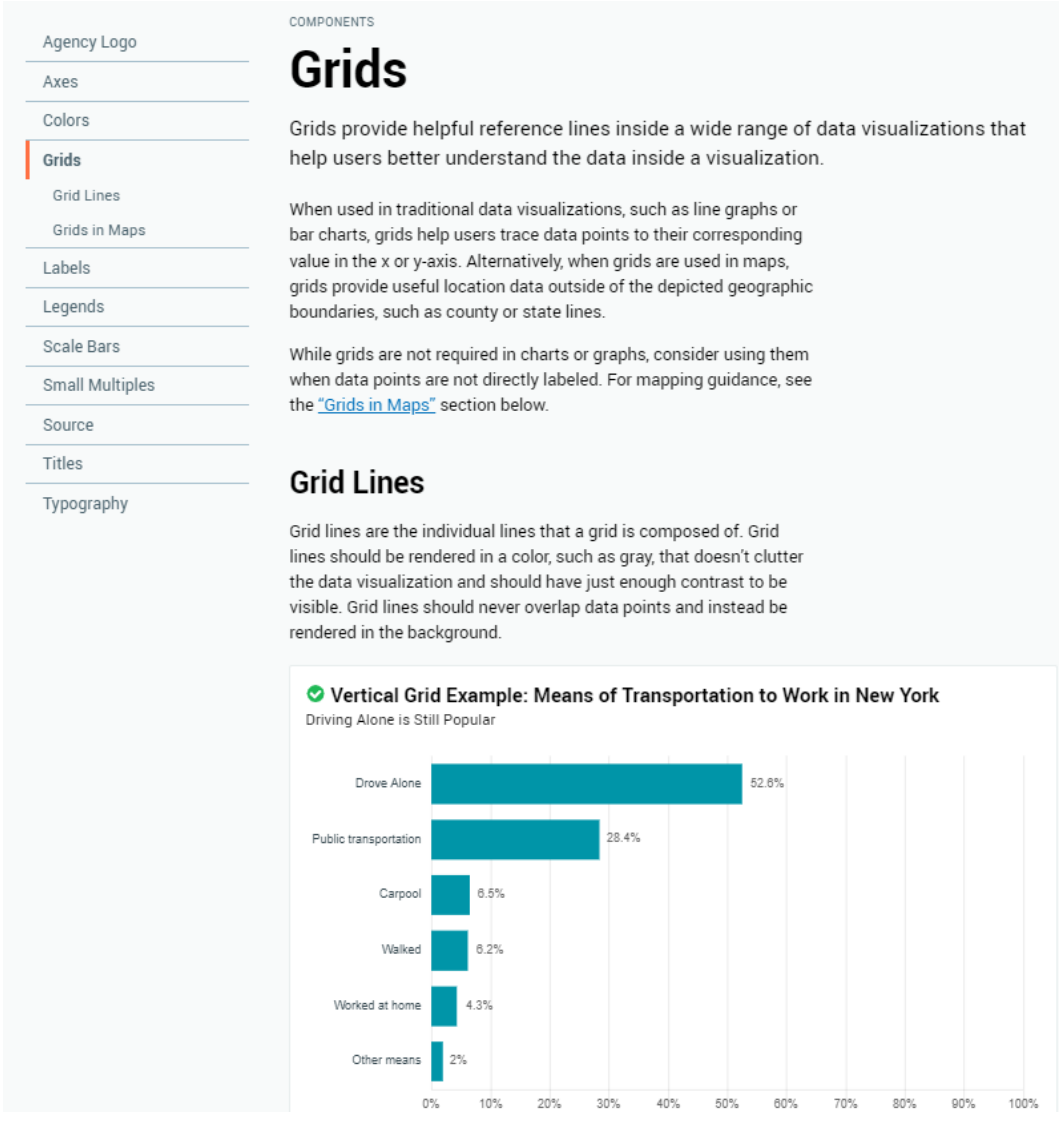
Recommendations

Recommended

- Use bar charts when comparing large changes in data values.
- Limit the number of bars or else a bar chart becomes difficult to understand.

United States® Census Bureau

A Component Example



United States® Census Bureau

Future Plans

- Taking this out of BETA!
- Adding more visualizations
- Adding a section for each with example code in a range of coding languages like R, Python, and SAS

United States® Census Bureau

The Future

- We specifically came to Q2024 to start a conversation with other national and international statistical agencies
- Our goal is to develop international quality standards data visualization
- ***Are you interested?***

United States® Census Bureau

Thank You!

You can reach our data visualization team at:

quality@census.gov

Steve Klement: steven.s.klement@census.gov

Rosemary Byrne: rosemary.byrne@census.gov

United States® Census Bureau