

Statistical Quality in Data Visualization

By

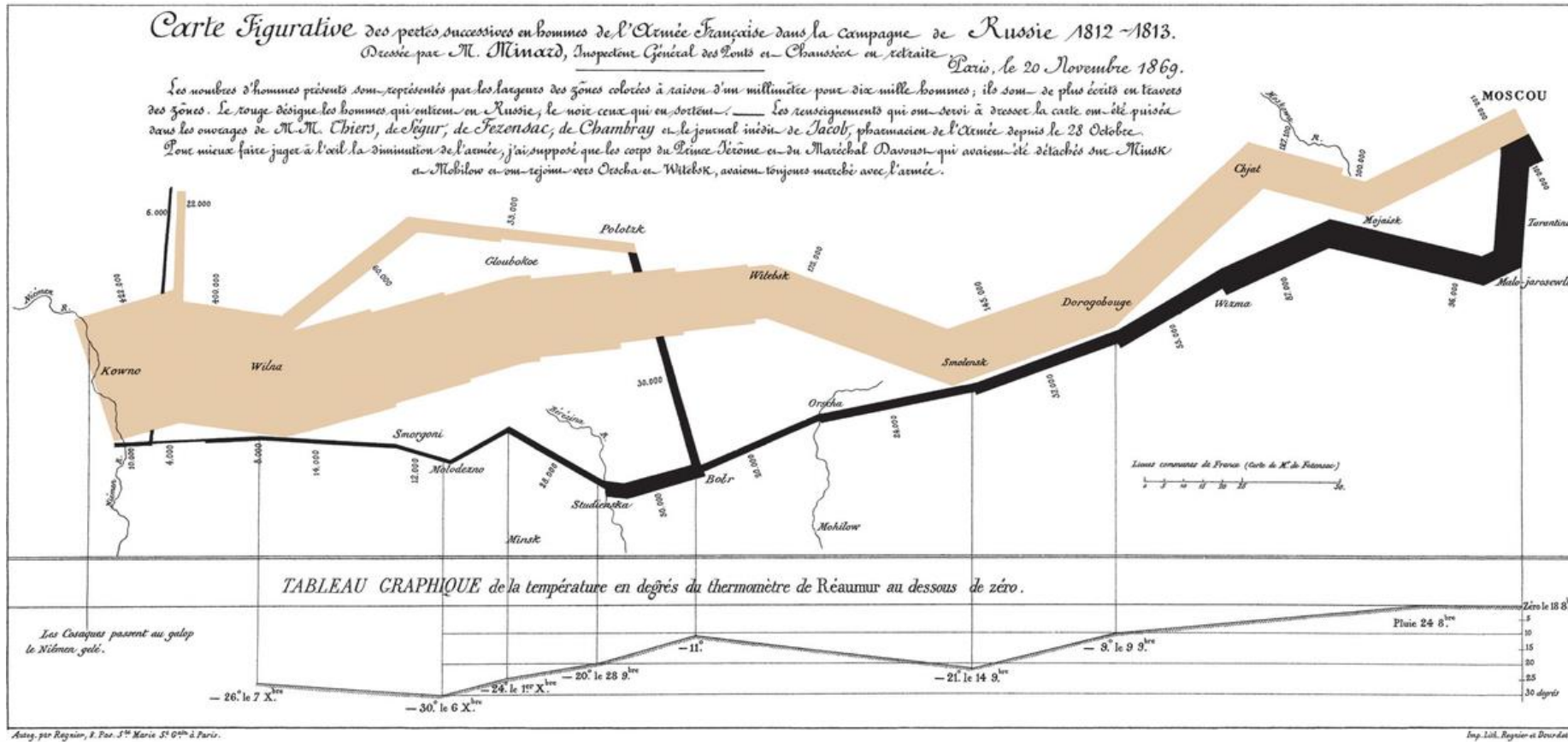
Steve Klement*

Rosemary Byrne*

Early Visualizations



The Early Age of Data Visualization

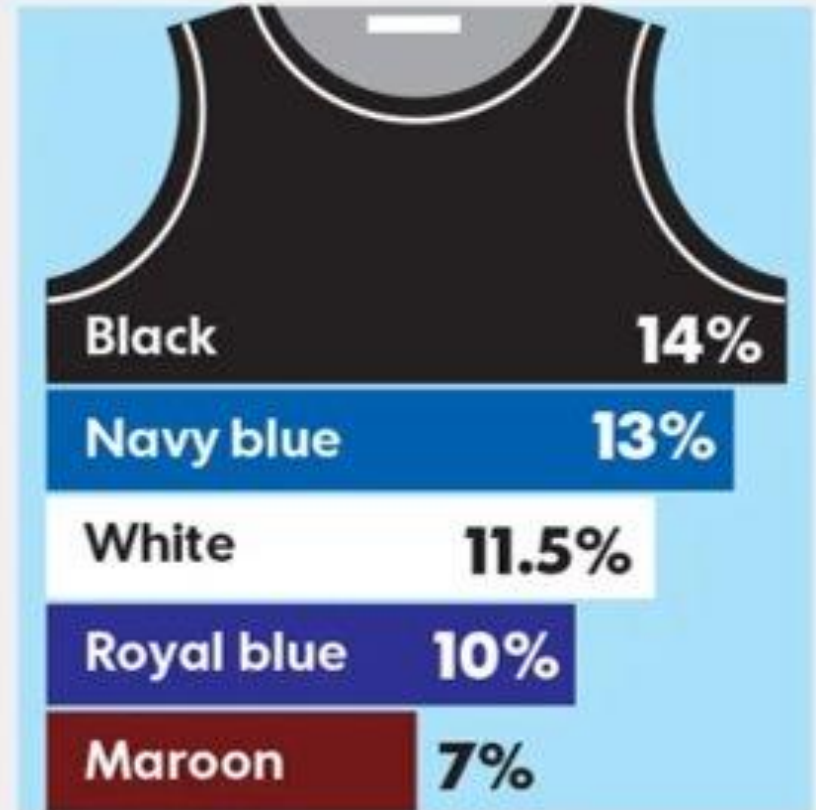


And Today We Have This...

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

Jersey colors

Top five highschool/
amateur sports jersey
colors printed in 2013



USA TODAY

Source TeamExpress.com

KEVIN GREER AND VERONICA BRAVO, USA TODAY

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***

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.

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.

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***

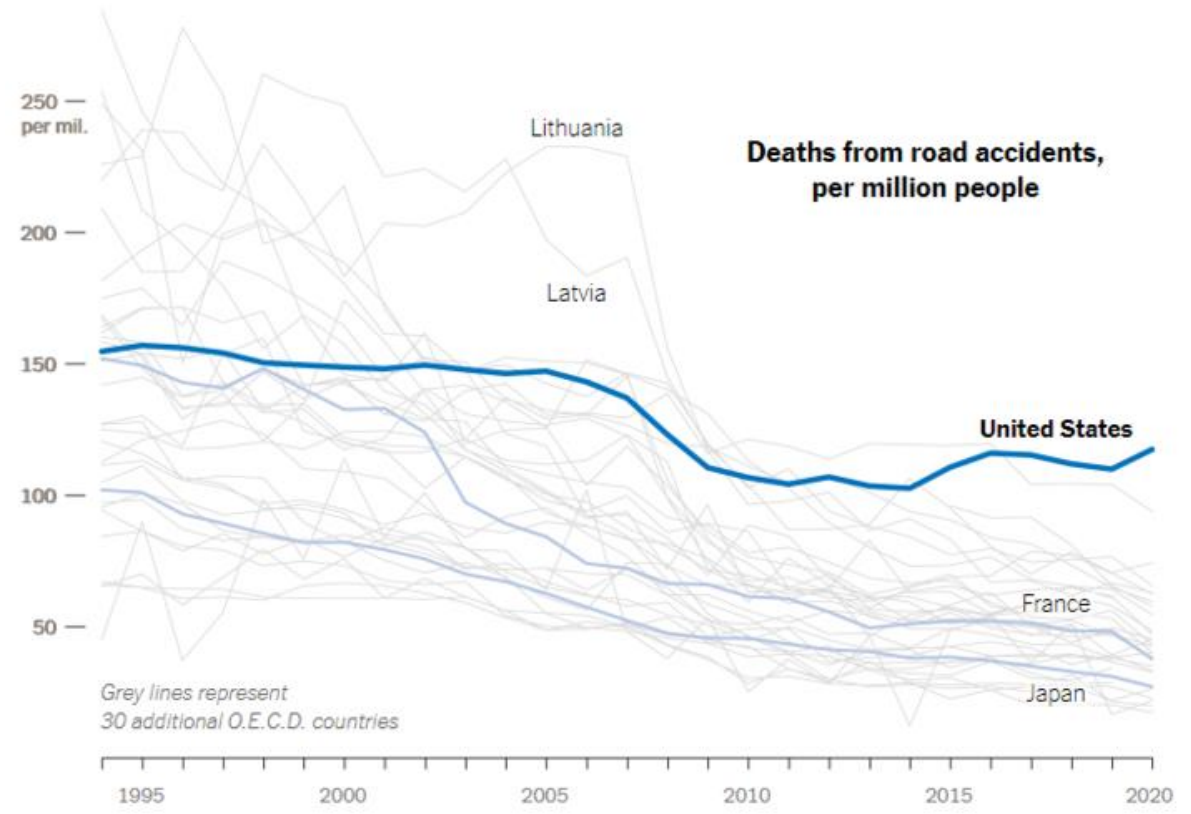
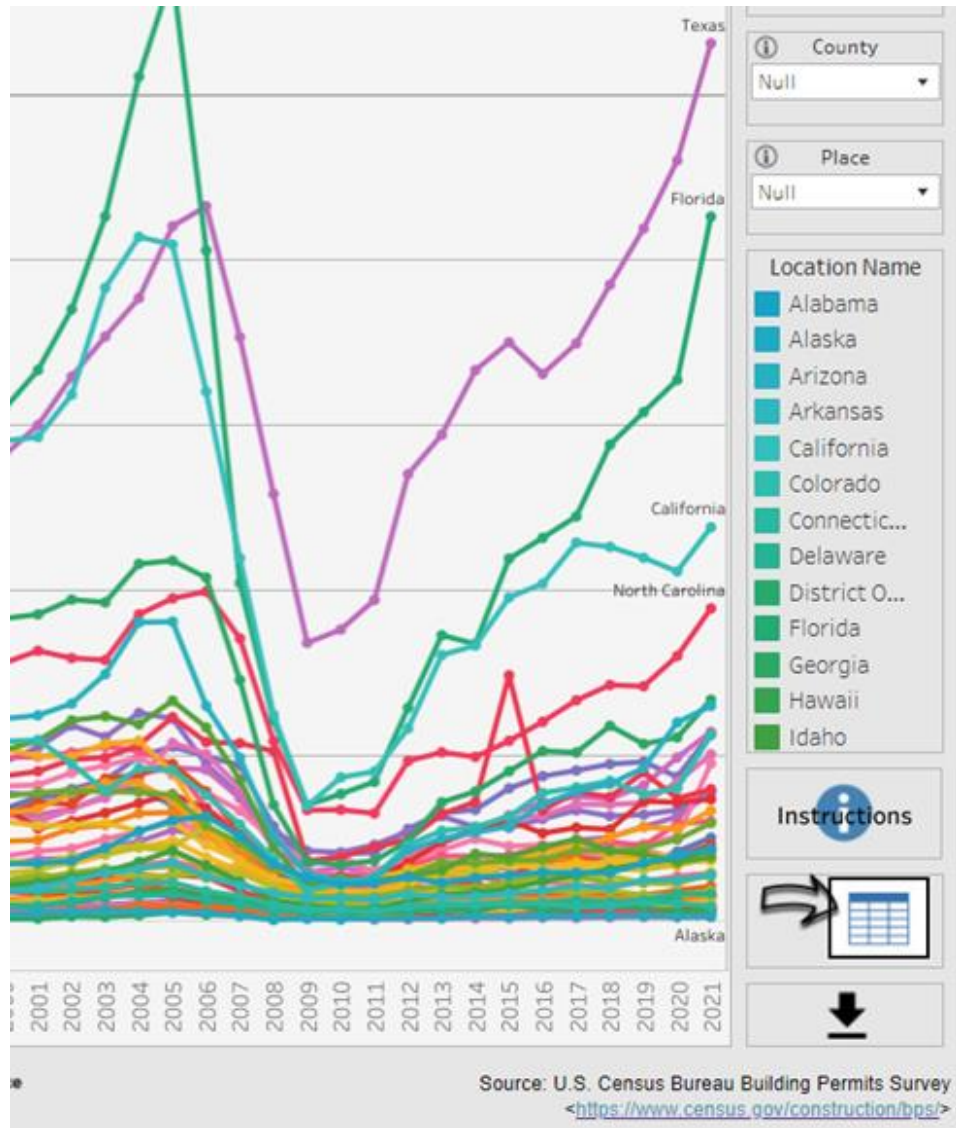
Why Data Viz Standards are Needed (cont.)

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

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?*



So, Can There Be Standards?

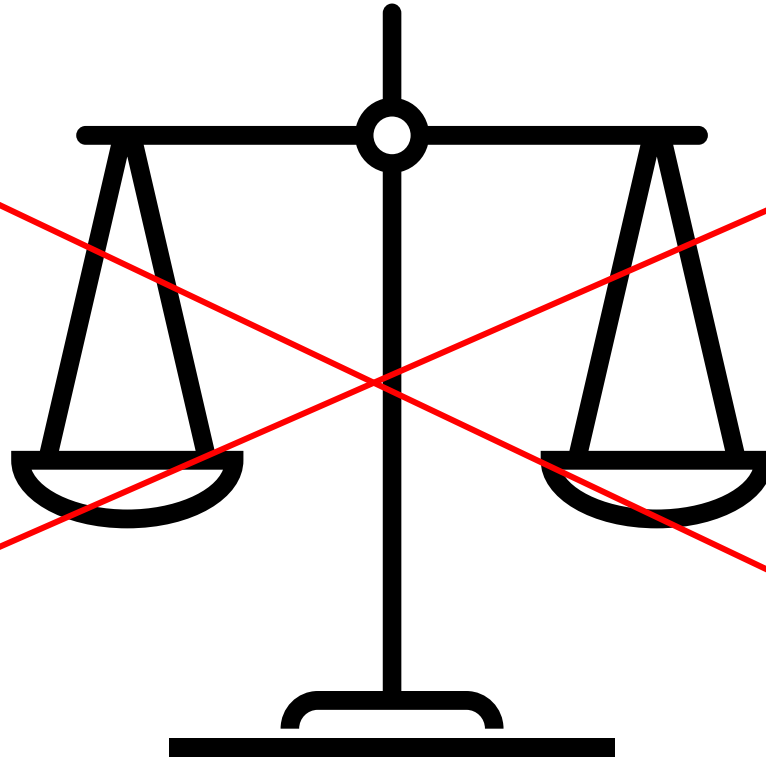
- Yes, but actually we've found that there are not many true standards that can be applied universally
- Which is why in developing standards, we've realized that what is even more important is giving guidance
 - Give what best practices are out there for different types of visualizations
 - Explain 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

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

A False Tradeoff

Statistically accurate
and complete



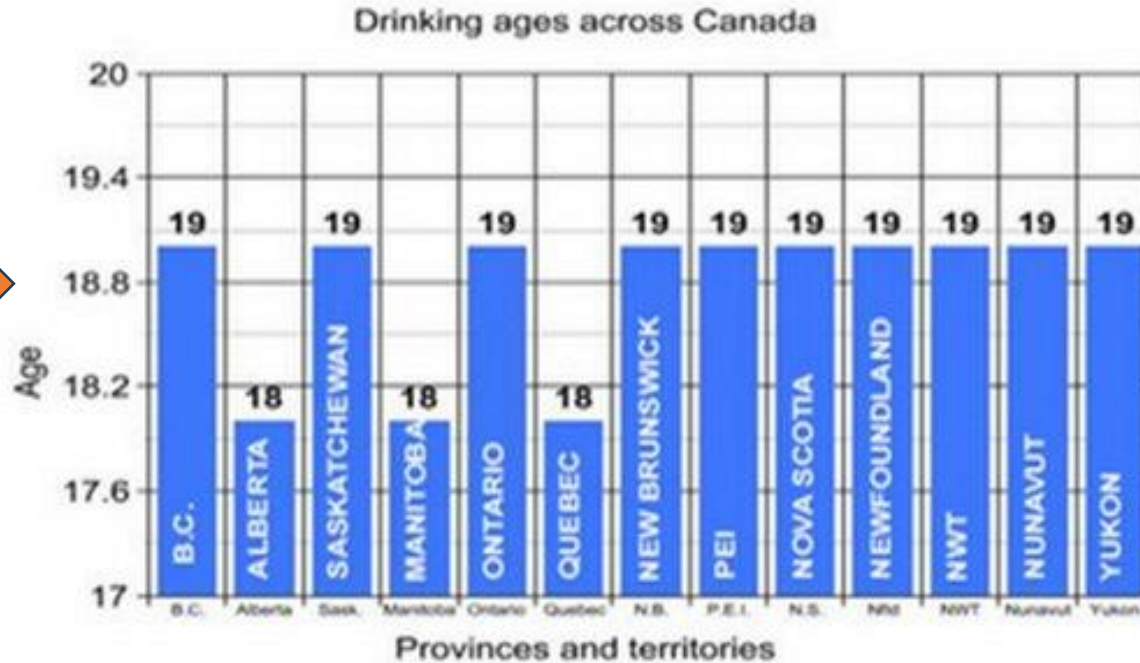
Easy to digest
visualization that
looks better

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

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



Canadian Centre on Substance Abuse

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

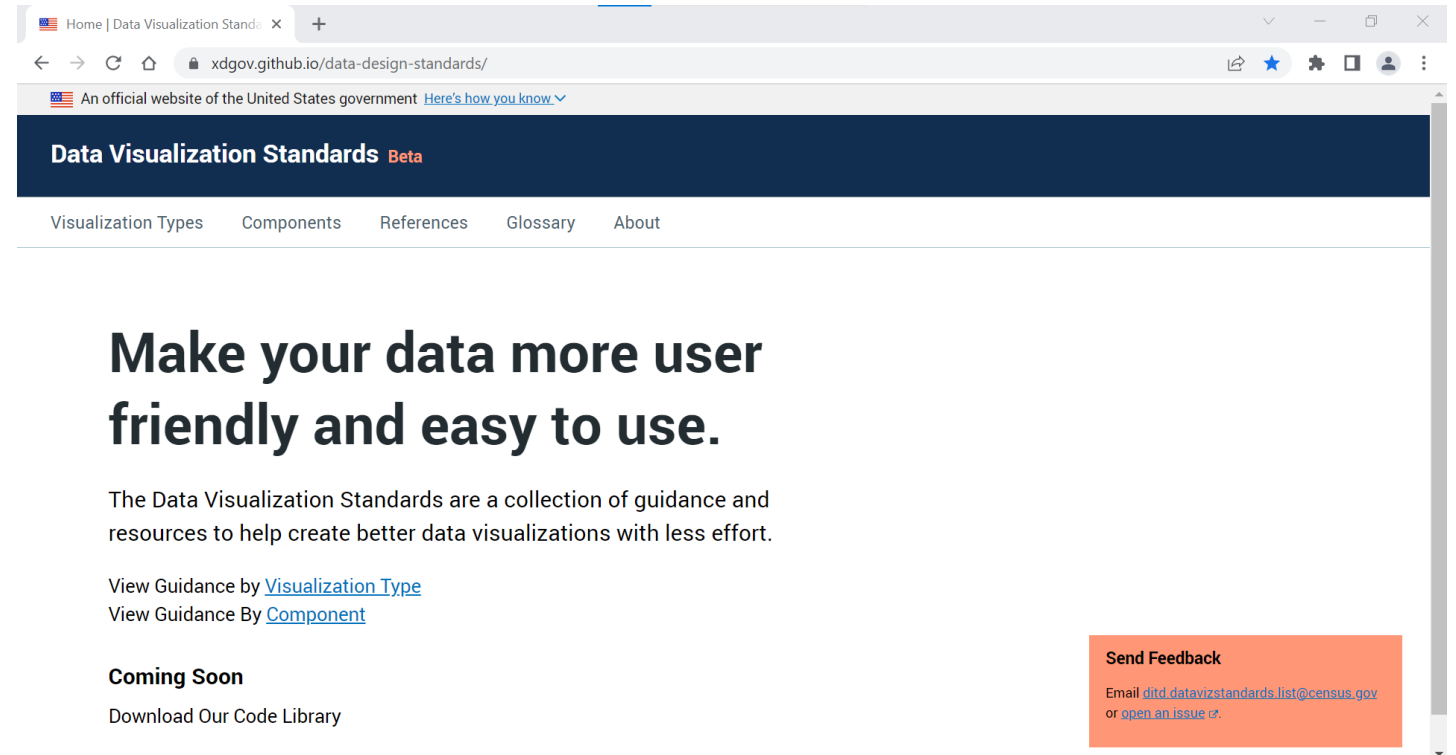


Mobile Fac

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/>



The screenshot shows a web browser window displaying the 'Data Visualization Standards Beta' website. The browser's address bar shows the URL 'xdgov.github.io/data-design-standards/'. The website has a dark blue header with the title 'Data Visualization Standards Beta' and a navigation menu with links for 'Visualization Types', 'Components', 'References', 'Glossary', and 'About'. The main content area features a large heading: 'Make your data more user friendly and easy to use.' Below this, a paragraph states: 'The Data Visualization Standards are a collection of guidance and resources to help create better data visualizations with less effort.' There are two links: 'View Guidance by Visualization Type' and 'View Guidance By Component'. A 'Coming Soon' section includes a link to 'Download Our Code Library'. In the bottom right corner, there is an orange 'Send Feedback' button with the text: 'Email dtd.datavizstandards.list@census.gov or [open an issue](#) on GitHub.'

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 create 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

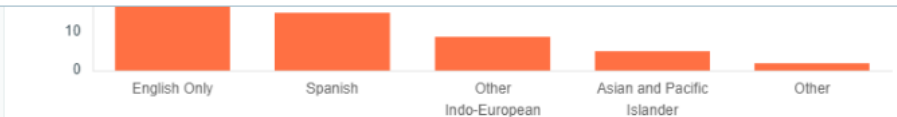
Language	Percent
English Only	68
Spanish	15
Other Indo-European	8
Asian and Pacific Islander	5
Other	2

Source: [2016 American Community Survey 1-Year Estimates](#)

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.

A Component Example

- Agency Logo
- Axes
- Colors
- Grids**
 - Grid Lines
 - Grids in Maps
- Labels
- Legends
- Scale Bars
- Small Multiples
- Source
- Titles
- Typography

COMPONENTS

Grids

Grids provide helpful reference lines inside a wide range of data visualizations that help users better understand the data inside a visualization.

When used in traditional data visualizations, such as line graphs or bar charts, grids help users trace data points to their corresponding value in the x or y-axis. Alternatively, when grids are used in maps, grids provide useful location data outside of the depicted geographic boundaries, such as county or state lines.

While grids are not required in charts or graphs, consider using them when data points are not directly labeled. For mapping guidance, see the [“Grids in Maps”](#) section below.

Grid Lines

Grid lines are the individual lines that a grid is composed of. Grid lines should be rendered in a color, such as gray, that doesn't clutter the data visualization and should have just enough contrast to be visible. Grid lines should never overlap data points and instead be rendered in the background.

✔ **Vertical Grid Example: Means of Transportation to Work in New York**
Driving Alone is Still Popular

A horizontal bar chart with vertical grid lines. The x-axis represents the percentage of workers, ranging from 0% to 100% in 10% increments. The y-axis lists six transportation methods. The bars are teal. The data points are: Drove Alone (52.8%), Public transportation (28.4%), Carpool (6.5%), Walked (6.2%), Worked at home (4.3%), and Other means (2%).

Means of Transportation	Percentage
Drove Alone	52.8%
Public transportation	28.4%
Carpool	6.5%
Walked	6.2%
Worked at home	4.3%
Other means	2%

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

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?***

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