Deep Blue https://deepblue.lib.umich.edu/documents

2016-02-06

Encounters with Data: Thinking Critically about Context and Presentation in Statistics and Visualizations

Steelberg, Tierney; Stuit, Martha

https://hdl.handle.net/2027.42/146733
http://creativecommons.org/licenses/by-nc-sa/4.0/

Downloaded from Deep Blue, University of Michigan's institutional repository

Encounters with Data: Thinking Critically about Context and Presentation in Statistics and Visualizations

Tierney Steelberg & Martha Stuit
University of Michigan School of Information

QuasiCon, February 6, 2016

Image source:
http://libweb5.princeton.edu/visual_materials/maps/websites/thematic-maps/quantitative/sociology-economics/playfair-chronology-1824.jpg

Creating Data Literate Students

http://dataliteracy.si.umich.edu/

What's wrong with this picture?

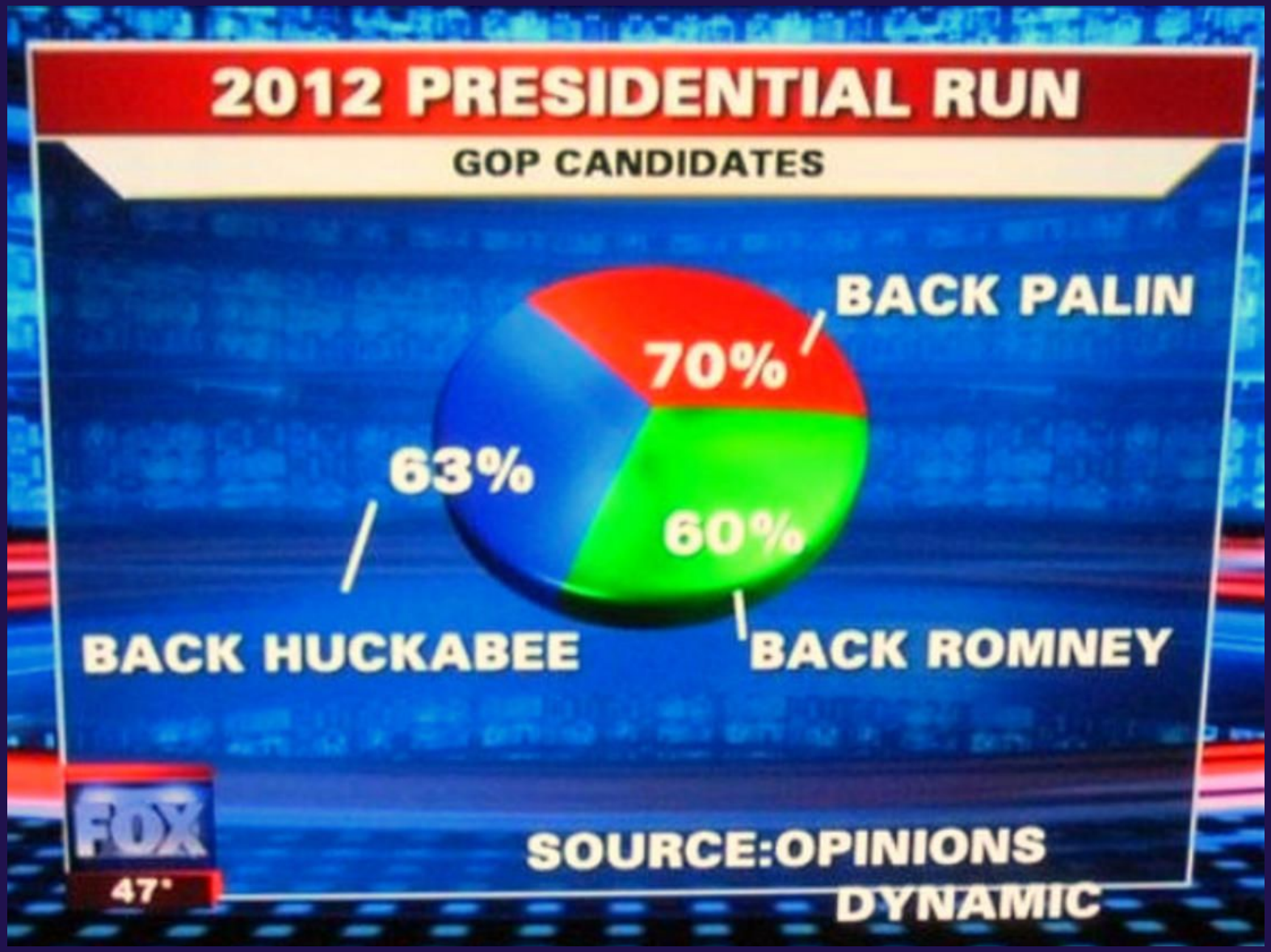


Image source: http://media.nbcchicago.com/images/1200*900/Fox's+Pie+Chart.jpg

Numbers often seem like facts...

but it's okay to question them!

Image source: Canva

What does it mean to be "data literate?"

Data Literacy

Statistical Literacy

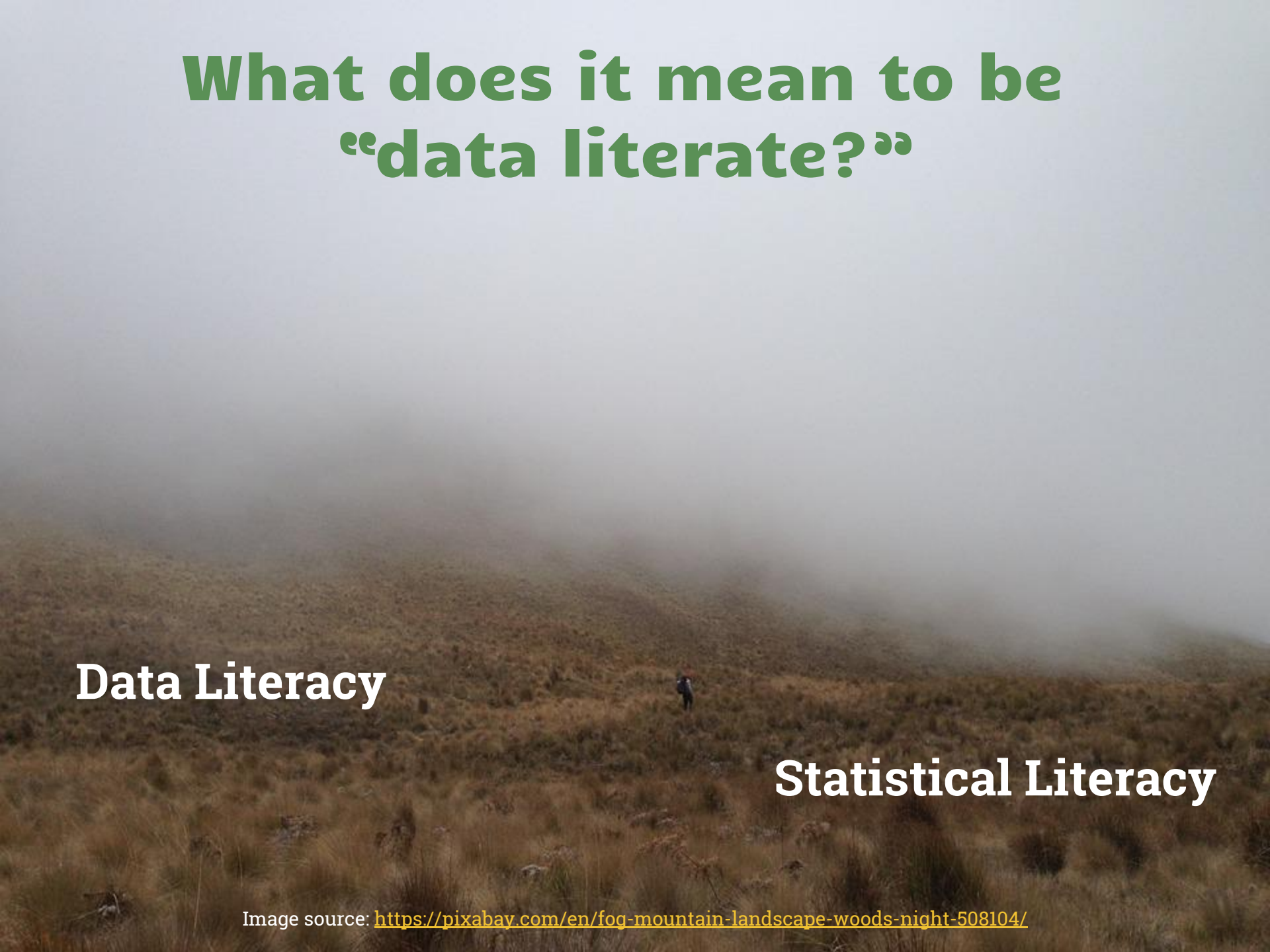


Image source: https://pixabay.com/en/fog-mountain-landscape-woods-night-508104/

The

- Who,
- What,
- When,
- Where,
- Why,
- and How

of statistics

How to Approach Statistics

Image source:
https://en.wikipedia.org/wiki/Iceberg#/media/File:Iceberg.jpg

Who

Who collected the data or made the statistic?

Who presented the statistic?

Who is represented in the statistic?

<u>Sample:</u> the group that was counted, surveyed, polled, or studied (a subset of the population)

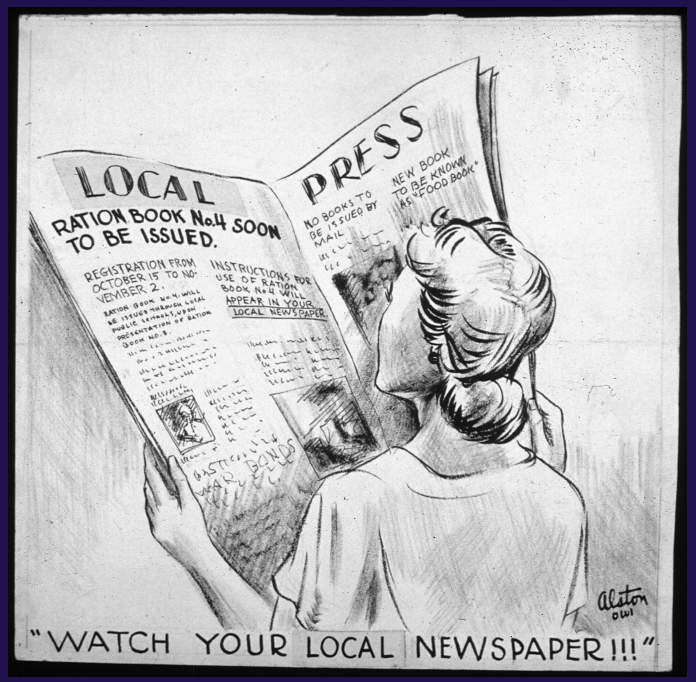


Image source: https://commons.wikimedia.org/wiki/File:%22WATCH_YOUR_LOCAL_NEWSPAPER%5E%5E%22_-_NARA_-_535653.jpg

Who

USA SNAPSHOTS©

Income disclosure

35-44

is the only age group in which a majority has disclosed their income to friends or family.

Note Other four age groups: 18-34, 45-54, 55-64 and 65 and older

Source Ally Bank survey of 1,008 adults

JAE YANG AND JANET LOEHRKE, USA TODAY

Yang, Jae, and Janet Loehrke. "USA SNAPSHOTS: Income Disclosure." *USA Today*, January 29, 2016, Weekend edition, Money.

What

What topic is the statistic illustrating? What problem is the statistic highlighting?

What is represented in the number?

What type of average is the statistic?

What information is missing from the presentation of the statistic?

Who & What

The Washington Post

"Lower-than-expected D.C. snowfall total raises questions about its measurement"

Variable: what has been counted

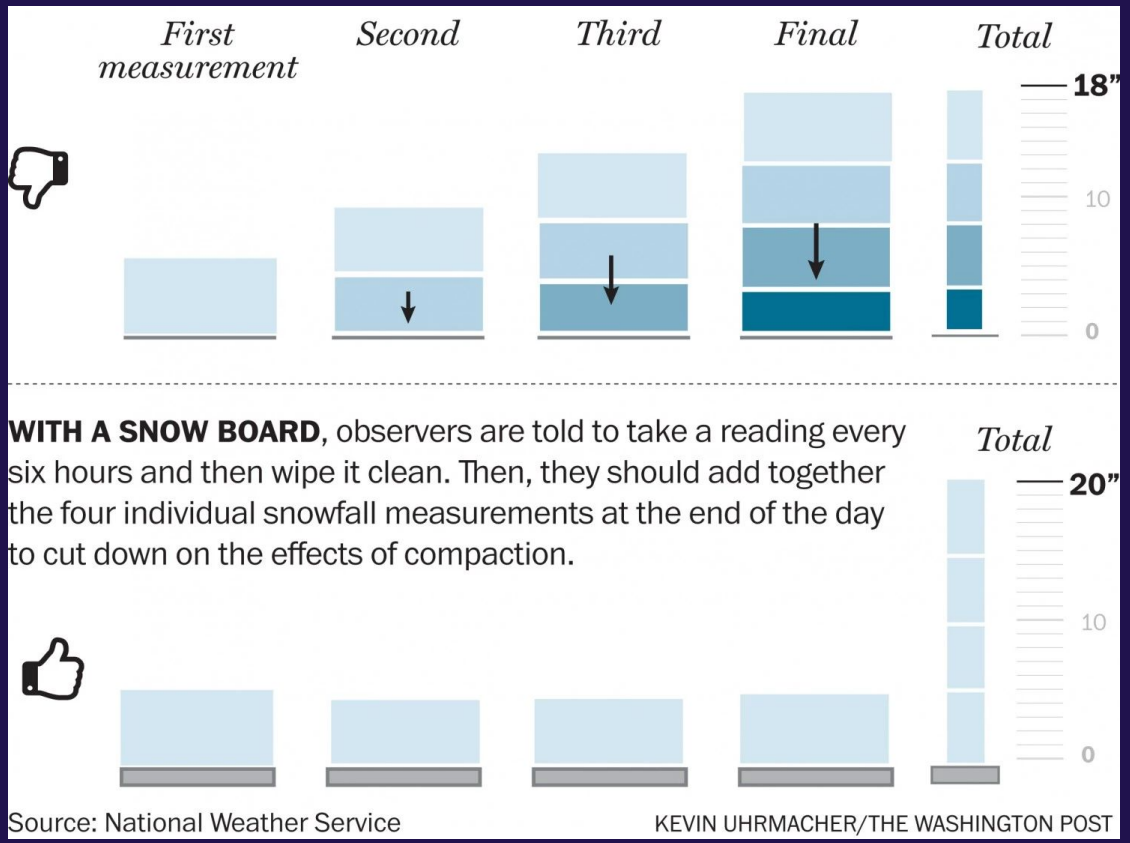
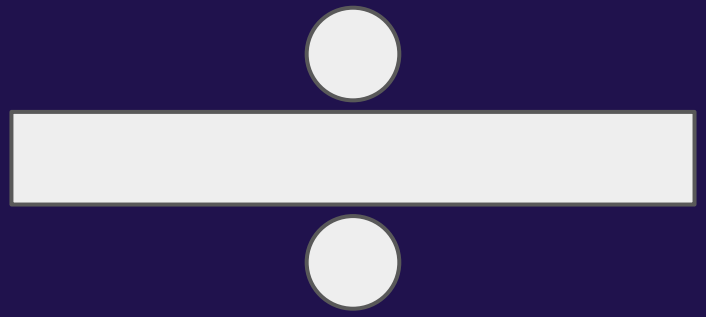


Image source:
http://www.freestockphotos.biz/stockphoto/15142
&
https://img.washingtonpost.com/wp-apps/imrs.php?src=https://img.washingtonpost.com/blogs/capital-weather-gang/files/2016/01/snowdepth0125b.jpg&w=1484

What, continued



What type of average is it?

- Mean: all values added together and divided by the number of values
    - sensitive to extremes
- Median: middle value when values are ordered from highest to lowest
- Mode: most frequently appearing value

Rule of thumb: Present all three types of average.

What, continued

Ten people who make $40,000 a year are in a restaurant. University of Michigan President Mark Schlissel walks in. What is the average income when he joins?

Hint: President Schlissel makes $772,500 a year.

Mark Schlissel salary (source):
https://www.michigandaily.com/section/news/regents-unanimously-raise-schlissels-salary-thursday-meeting
Image source: https://usatftw.files.wordpress.com/2014/11/ap-michigan-brandon-football.jpg?w=1200

What, continued

Ten people who make $40,000 a year are in a restaurant. President Mark Schlissel walks in. What is the average income when he joins?

Mean: **$106,590.91**

Median: **$40,000**

Mode: **$40,000**

Rule of thumb: Present all three types of average.

Image source: https://usatftw.files.wordpress.com/2014/11/ap-michigan-brandon-football.jpg?w=1200

When

When was the data collected?

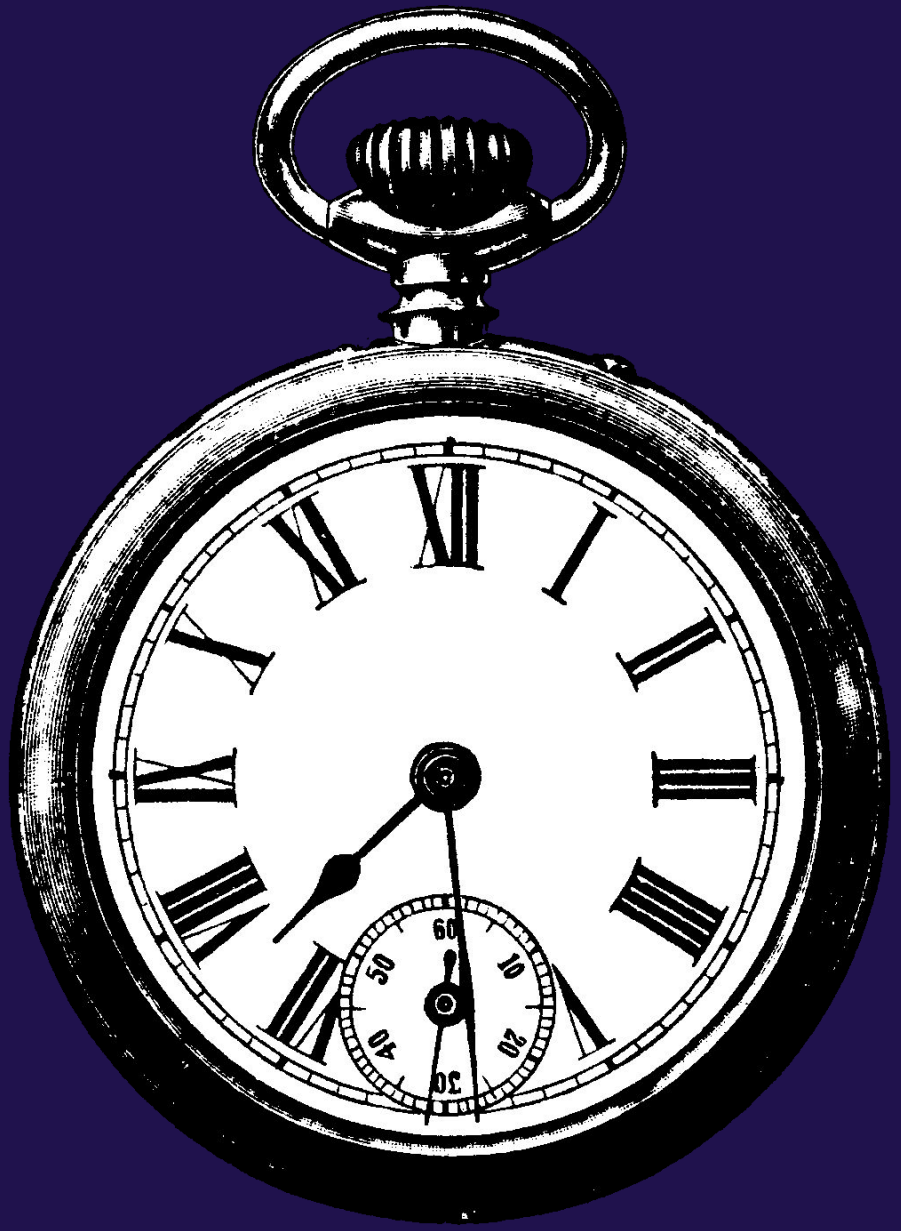


Image source: http://www.clipartbest.com/cliparts/dir/7px/dir7pxM4T.png

Where

Where was the sample conducted?

Where does the data apply? (a particular population? or location?)

Rule of thumb: Correlation is not causation.

Where

Rule of thumb: Correlation is not causation.

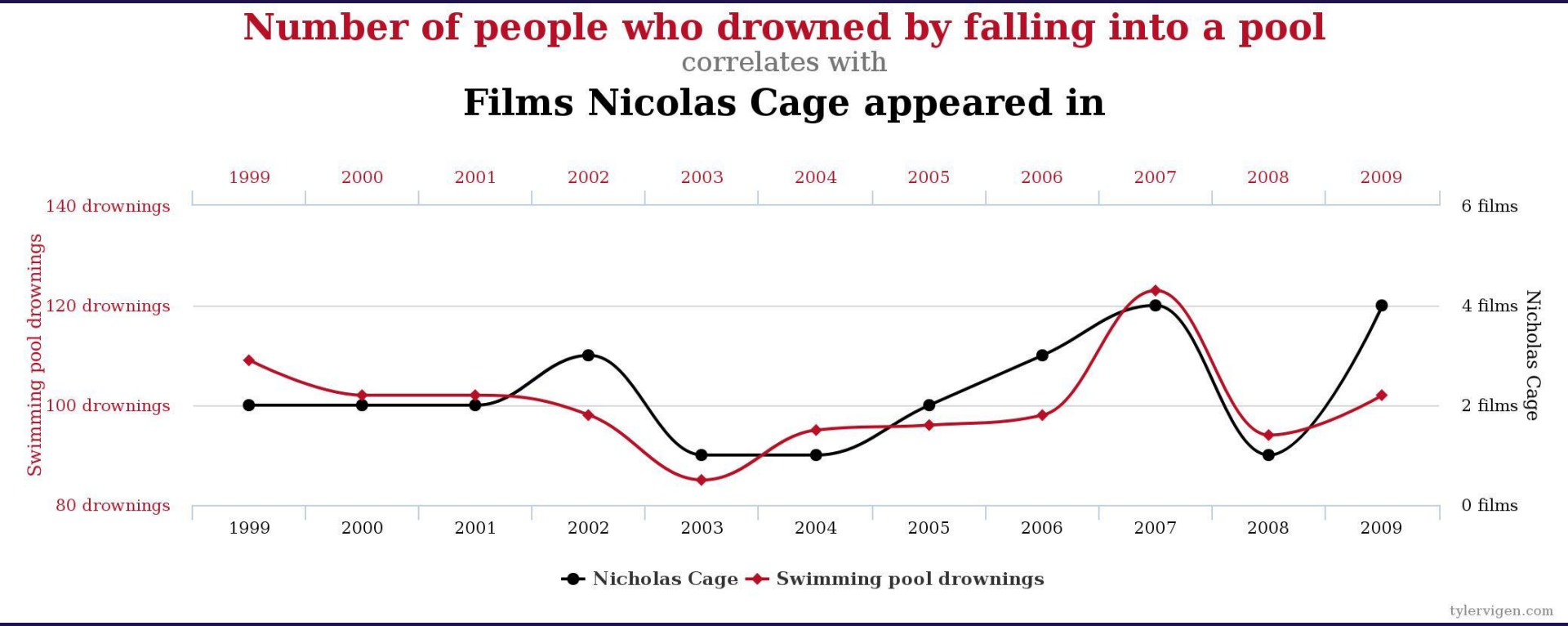
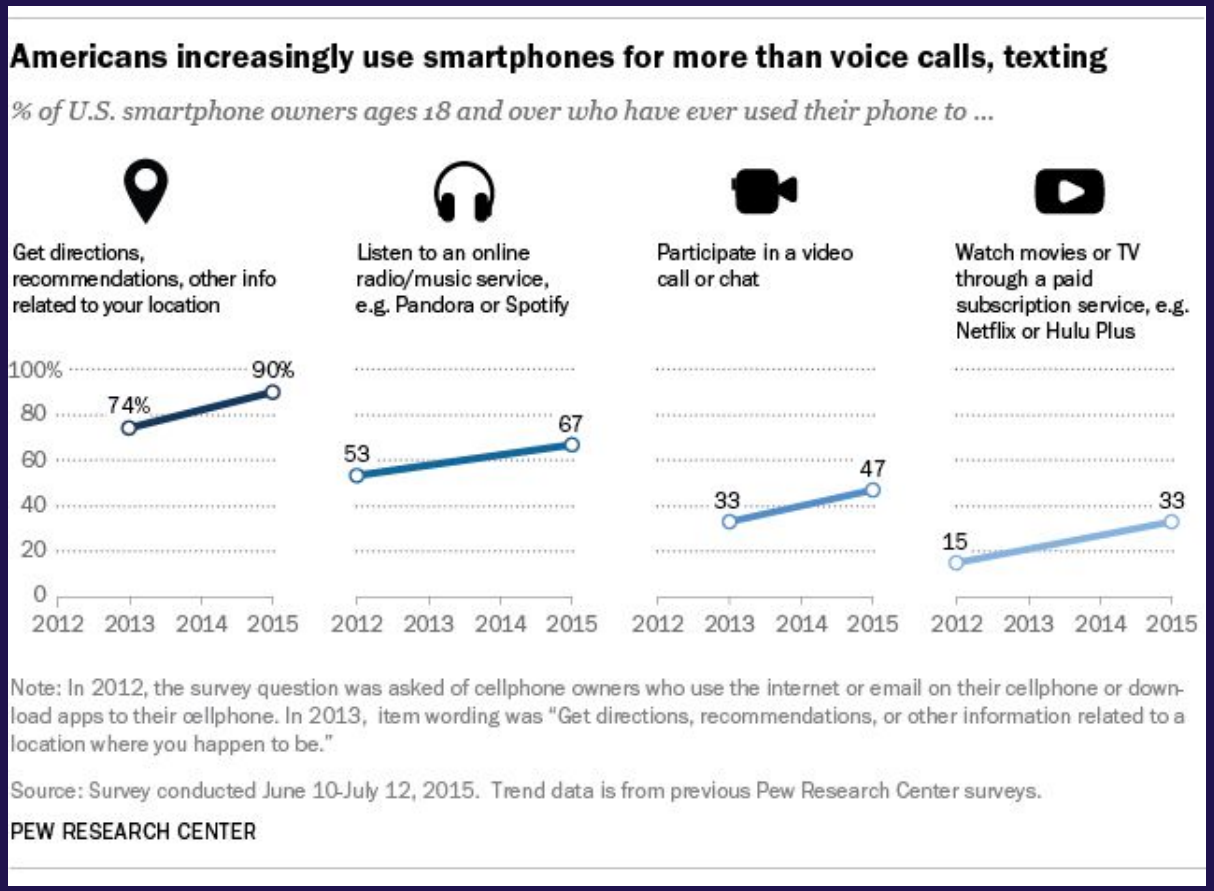


Image source: http://www.tylervigen.com/spurious-correlations

Why

Why is this number significant?
Why did someone calculate this statistic?

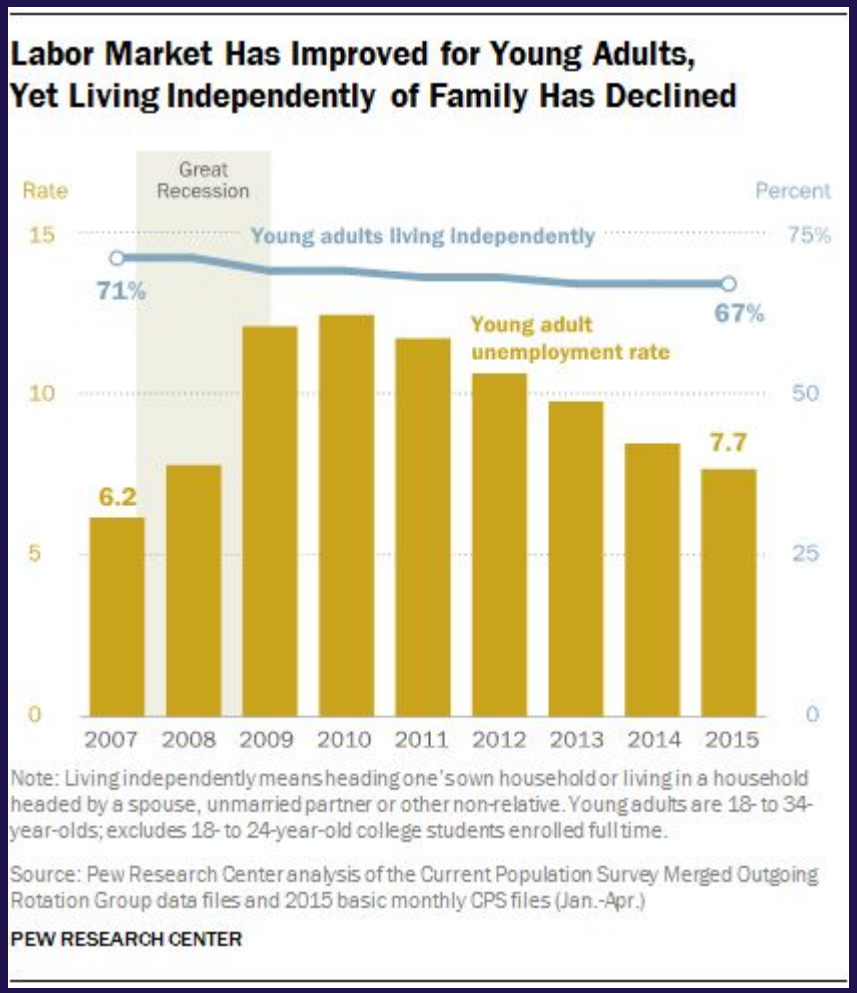


Link: Pew Research Center

How

How was this statistic calculated? Is it a prediction, or is it an accurate count?

How big or small is the number? (Does it seem extreme?)



Source: http://www.pewsocialtrends.org/2015/07/29/more-millennials-living-with-family-despite-improved-job-market/

Context of Statistics: Rules of Thumb

1. Correlation does not equal causation.
2. Interrogate the statistics that you encounter by asking:
    - who?
    - what?
    - when?
    - where?
    - why?
    - and how?

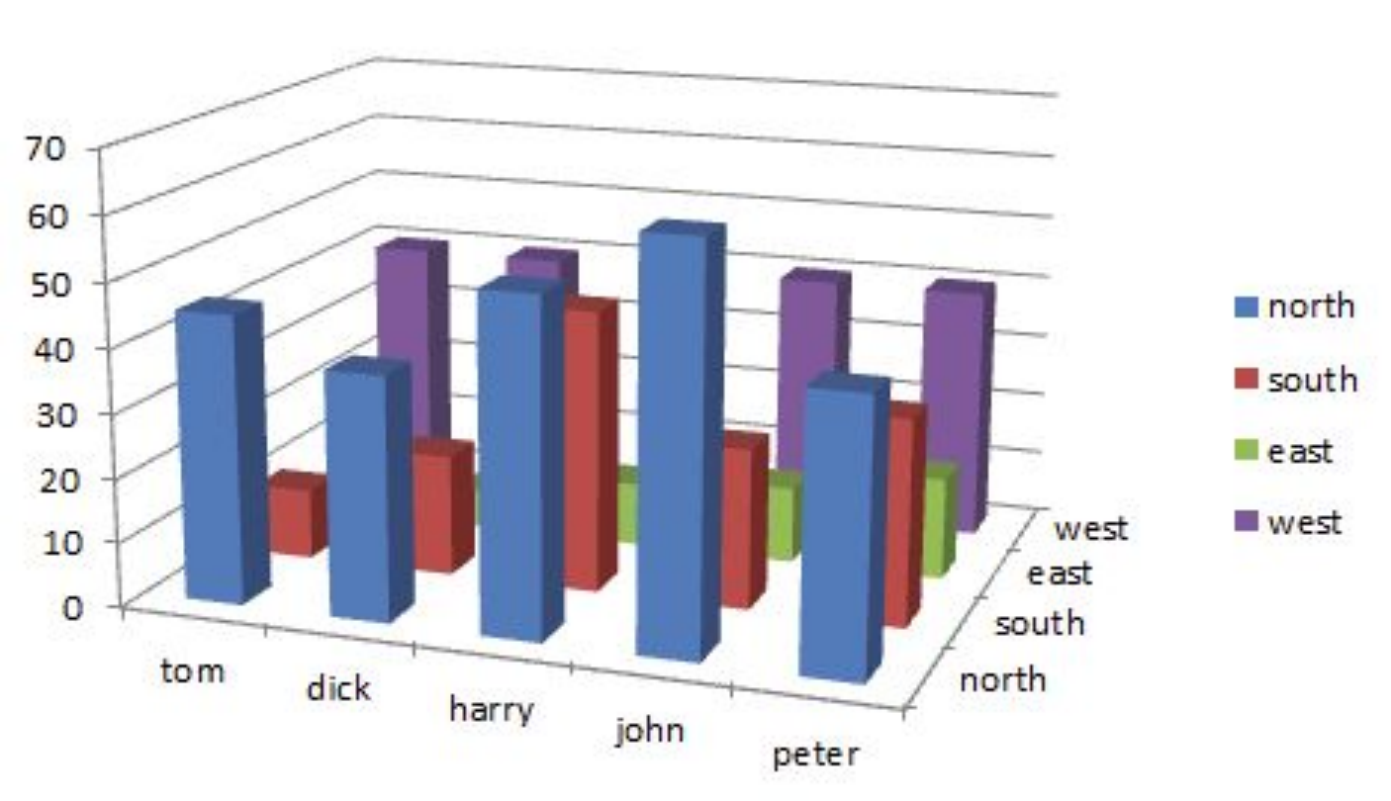
Image source: https://commons.wikimedia.org/wiki/File:Symbol_thumbs_upzel.svg (edited for color)

General Rules of Thumb for Presenting Your Data

Image source: http://chicagouic.estaging.intrafinity.com/common/resources/images/Cliparts/Math/Bar%20Chart%20Glassy.png

Clarity and simplicity are key.

Clarity and simplicity are key.



What do you think of the way this data is presented?

Image source: http://www.forbes.com/sites/naomirobbins/2012/06/07/trellis-plot-alternative-to-three-dimensional-bar-charts/

Clarity and simplicity are key.

- Keep it simple: avoid unnecessary ornamentation.
- Split things up into multiple charts if the display starts getting crowded.

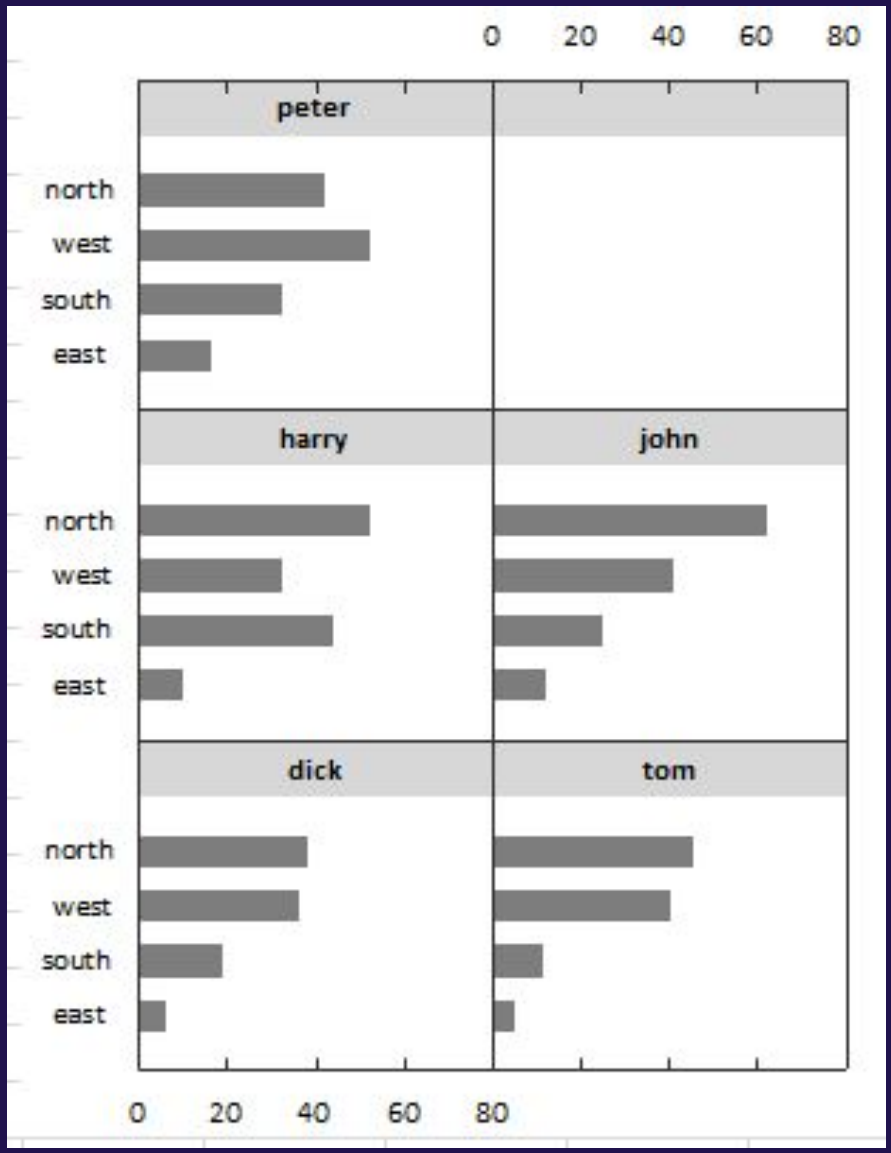
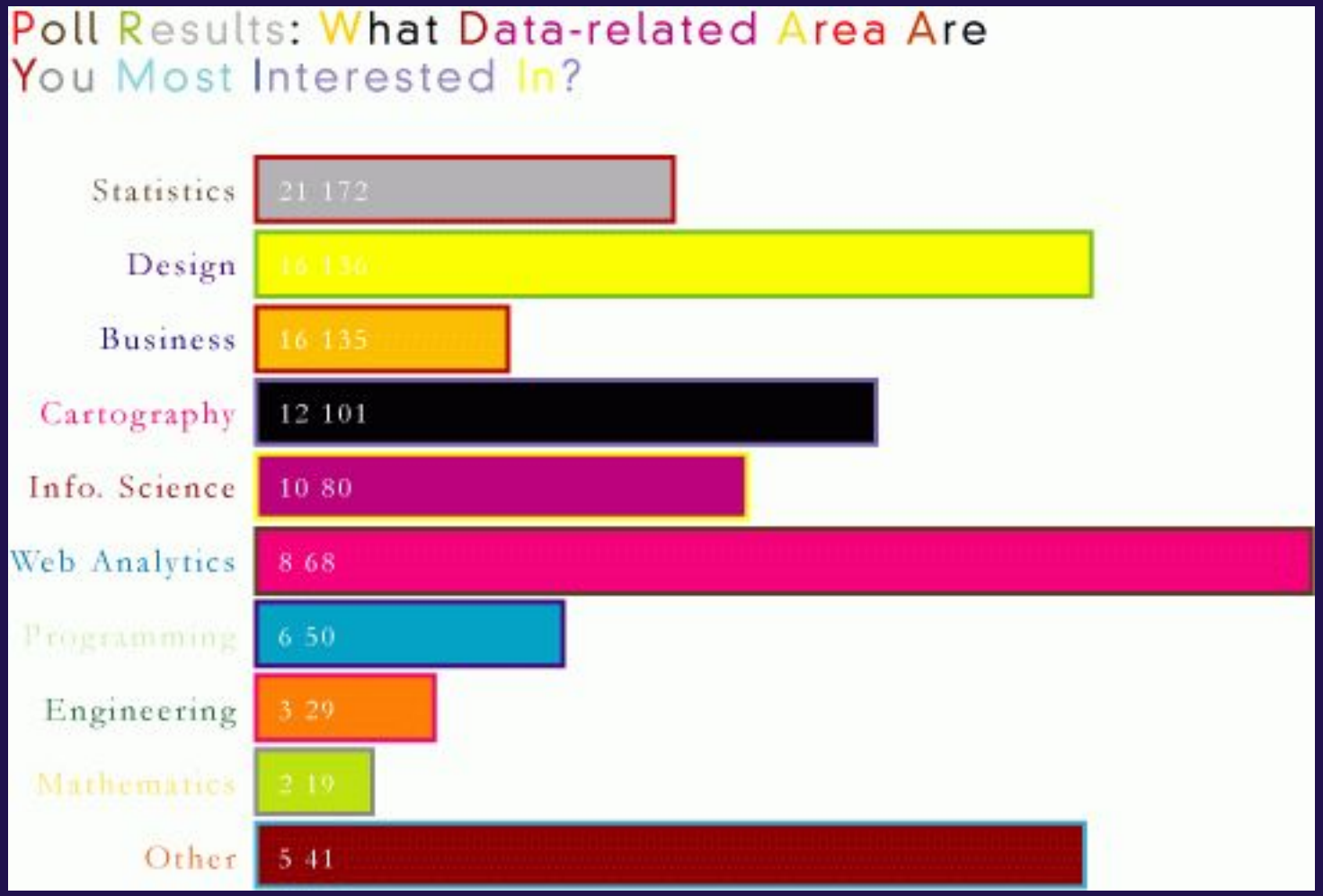


Image source:
http://www.forbes.com/sites/naomirobbins/2012/06/07/trellis-plot-alternative-to-three-dimensional-bar-charts/

Make it easy to read and interpret.

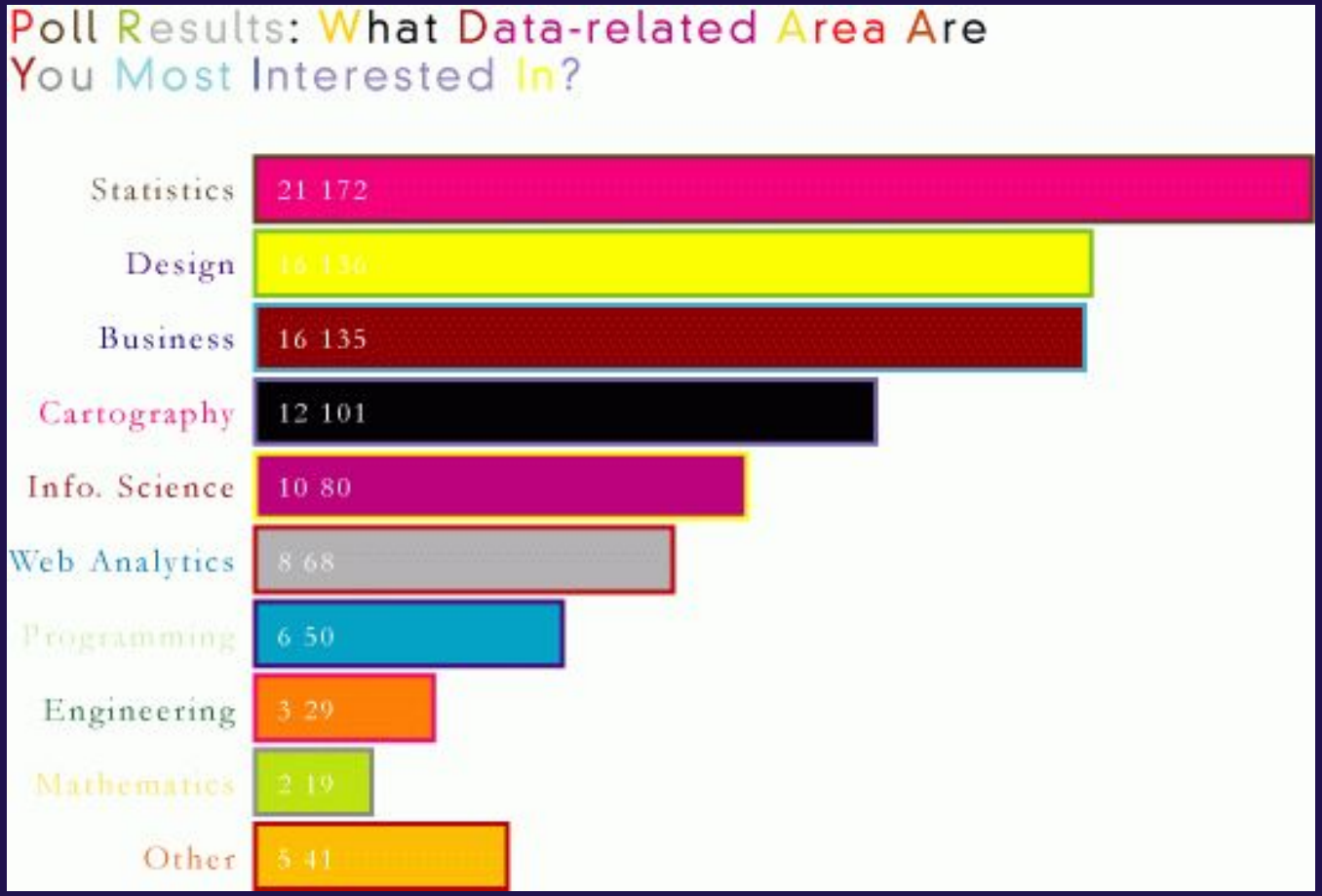
Make it easy to read and interpret.



What's wrong with this picture?

Image source: http://flowingdata.com/2009/06/15/6-easy-steps-to-make-your-graph-really-ugly/

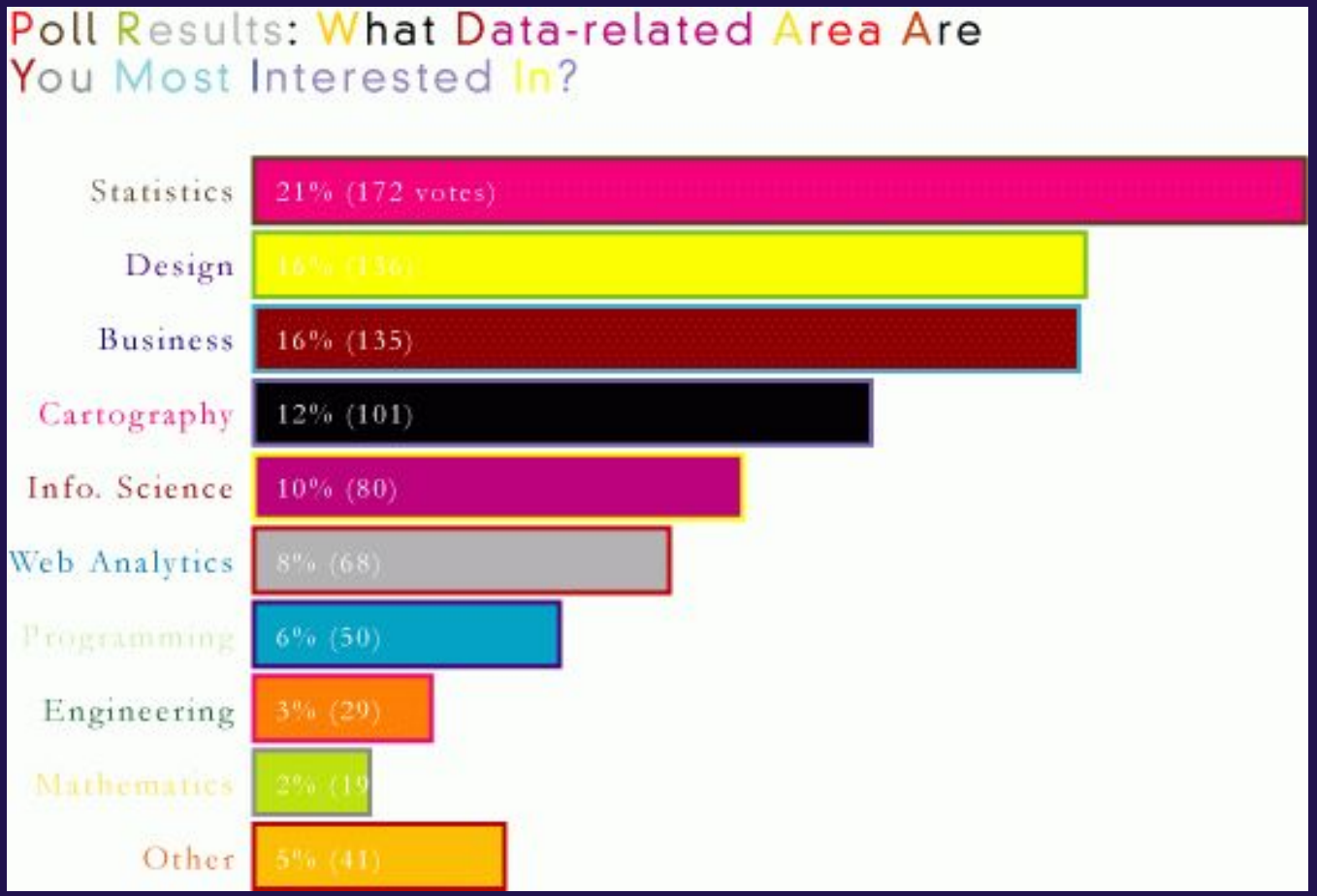
Make it easy to read and interpret.



Organize values in a meaningful order.

Image source: http://flowingdata.com/2009/06/15/6-easy-steps-to-make-your-graph-really-ugly/

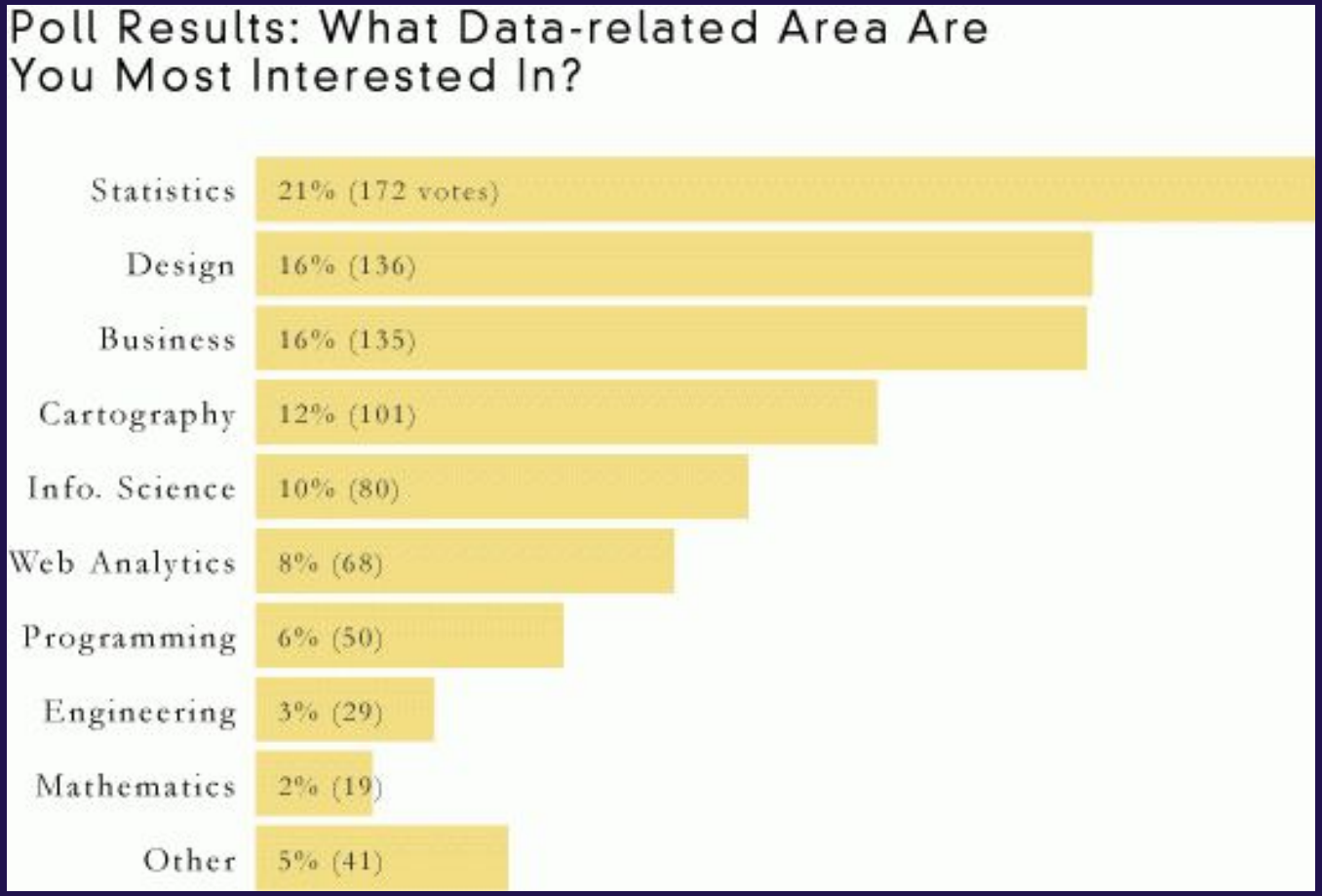
Make it easy to read and interpret.



Provide a legend and labels.
Clarify units.

Image source: http://flowingdata.com/2009/06/15/6-easy-steps-to-make-your-graph-really-ugly/

Make it easy to read and interpret.

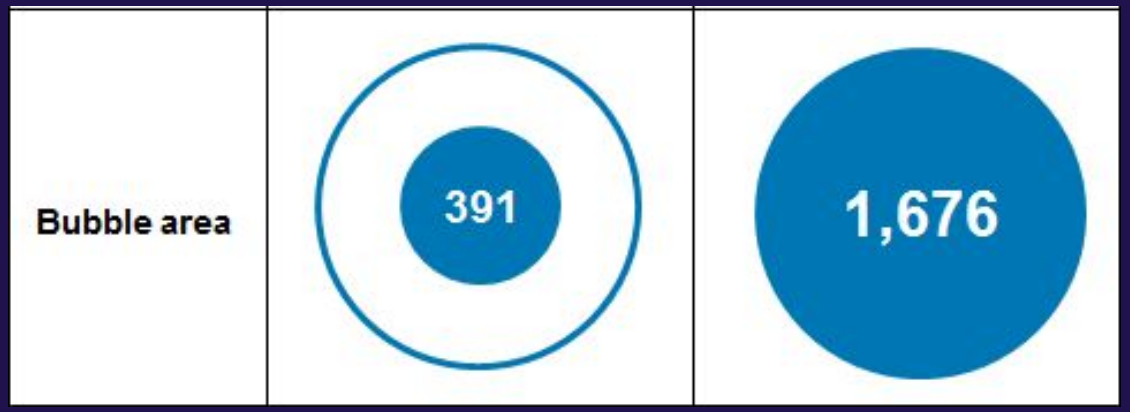


Use a simple color scheme. Avoid using color combinations that are difficult to distinguish.

Image source: http://flowingdata.com/2009/06/15/6-easy-steps-to-make-your-graph-really-ugly/

Respect visual and mathematical principles.

[...]

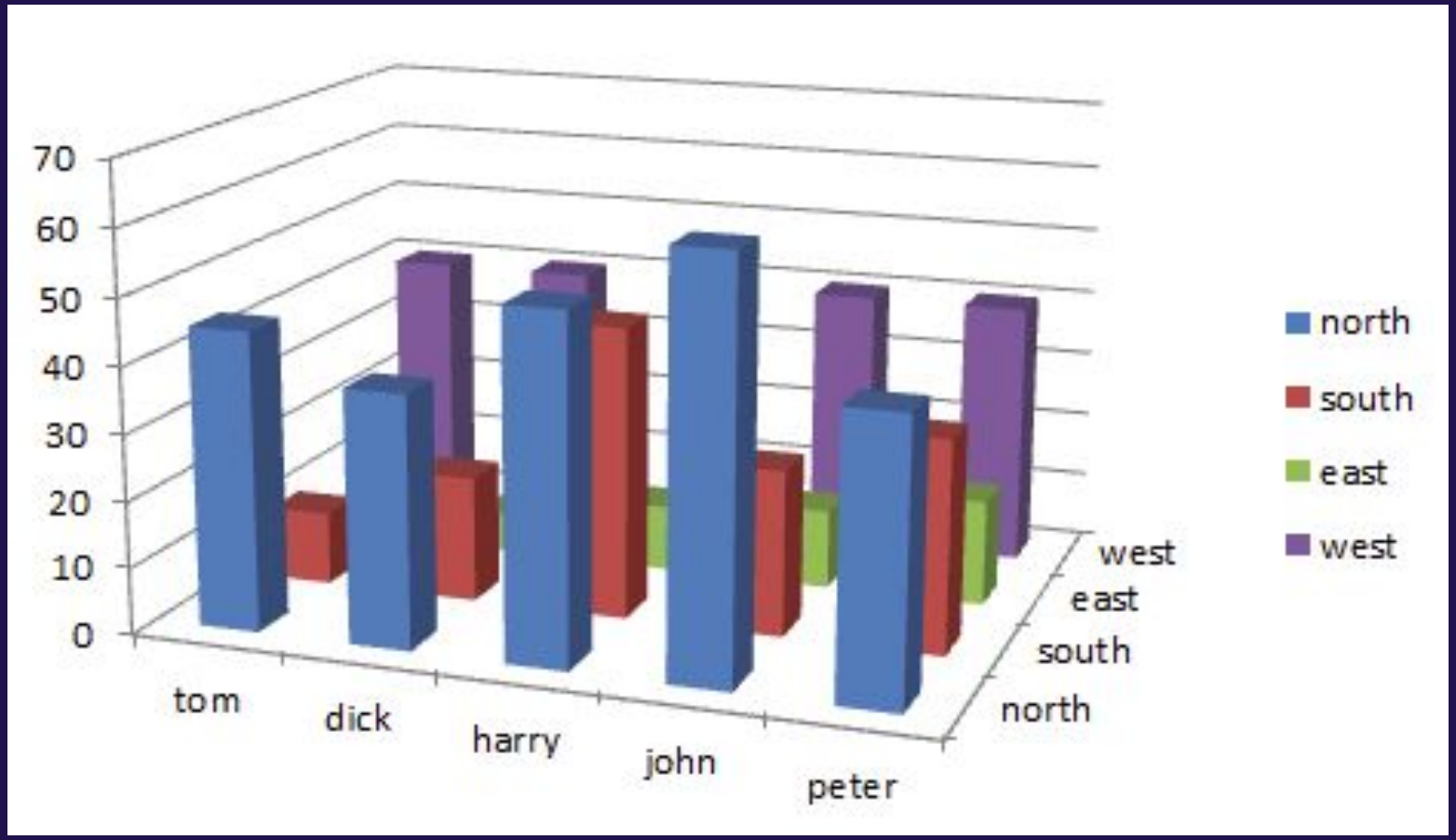


Respect visual and mathematical principles.

Size two-dimensional shapes proportionally according to their area.

Image source:
http://viz.wtf/post/131758708391/does-the-big-circle-show-391-connections-656-or
(cropped for clarity)

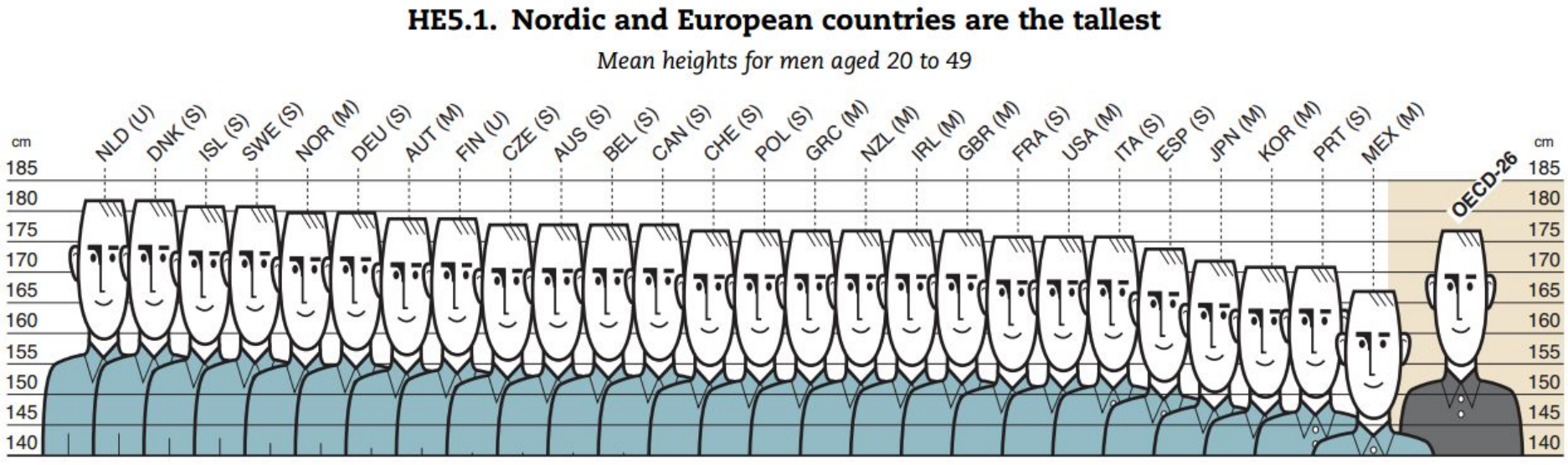
Respect visual and mathematical principles.



Keep things in two dimensions.

Image source:
http://www.forbes.com/sites/naomirobbins/2012/06/07/trellis-plot-alternative-to-three-dimensional-bar-charts/

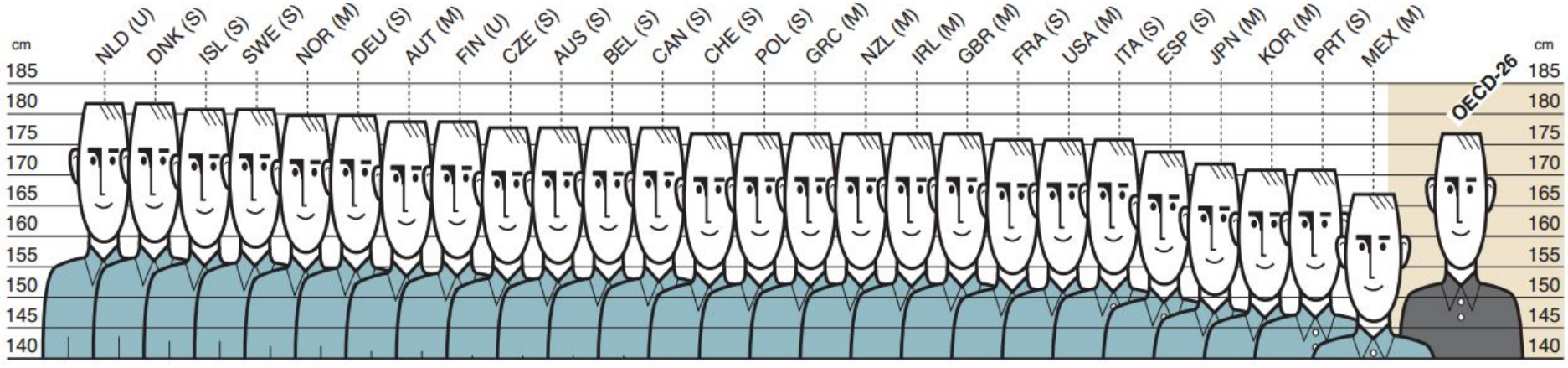
Respect visual and mathematical principles.



Do these icons add to the visualization?

Image source: OECD via WTF Visualizations (http://viz.wtf/image/131032061255).

Respect visual and mathematical principles.



In general, forego icons in the data visualization itself.

Image source: OECD via WTF Visualizations (http://viz.wtf/image/131032061255).

Play around with your data!

Play around with your data!

Try out different charts and graphs, using software readily at your disposal: it's as easy as the click of a button.

Microsoft Excel

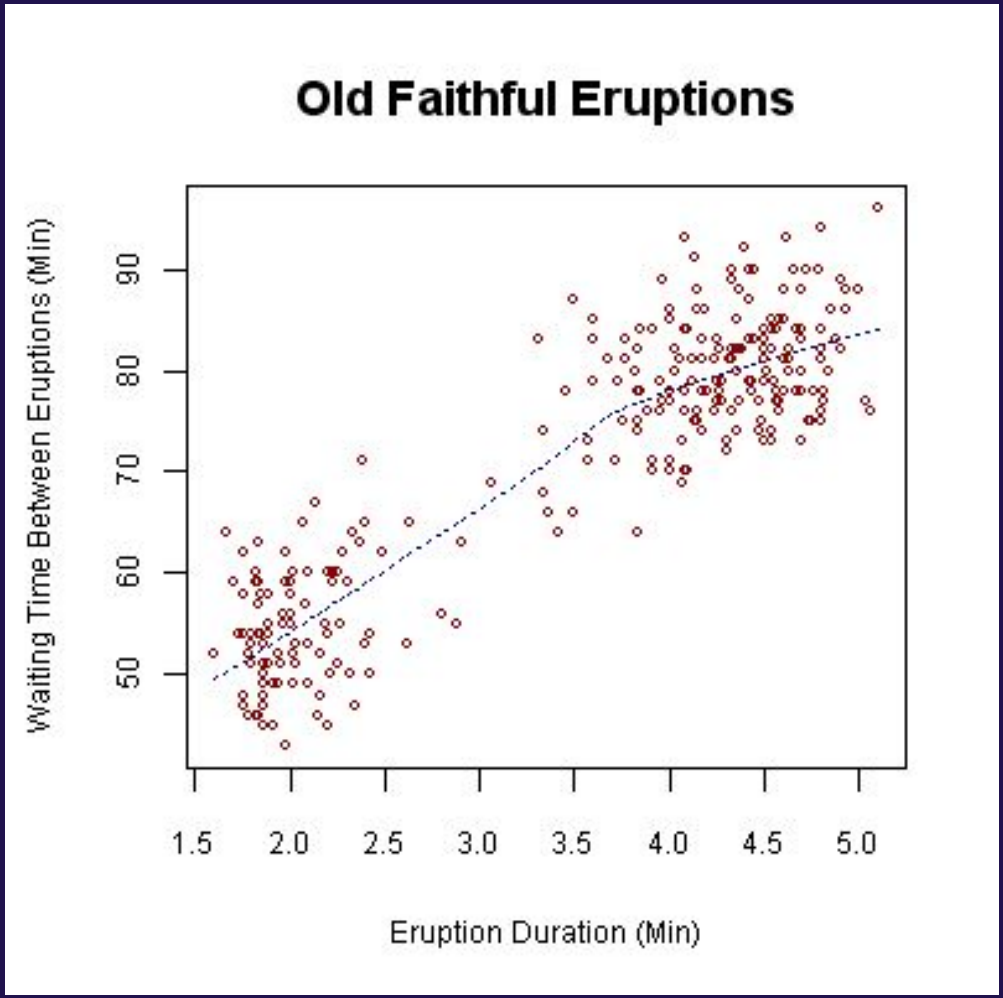
Image source:
https://commons.wikimedia.org/wiki/File:Microsoft_Excel_2013_logo.svg

Google Sheets

Image source:
http://eci511-emarsh-blog.blogspot.com/

Cite your sources (or even provide your dataset).

Cite your sources.



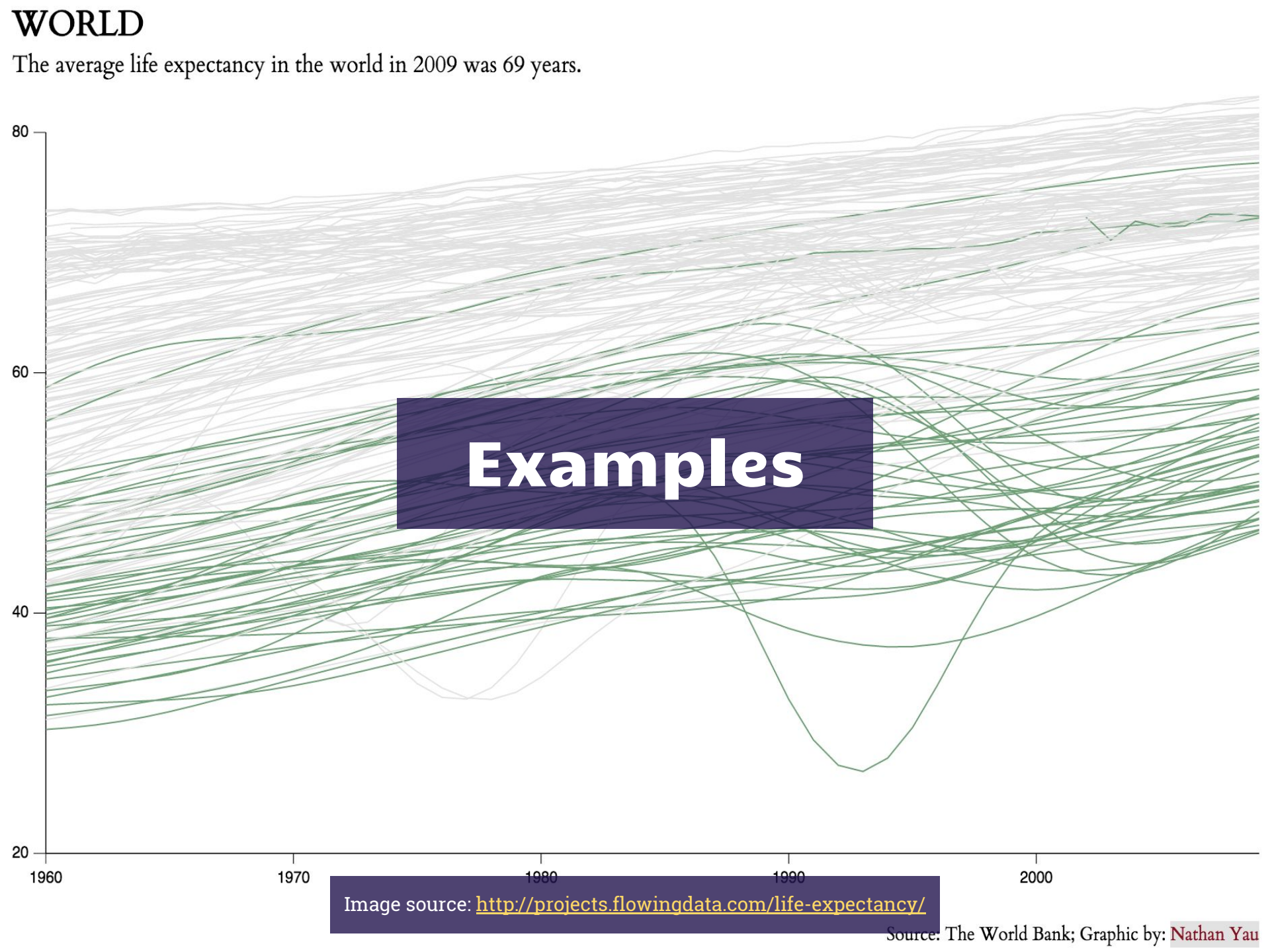
Source is missing: what would context provide?

Image source: https://en.wikipedia.org/wiki/File:Oldfaithful3.png

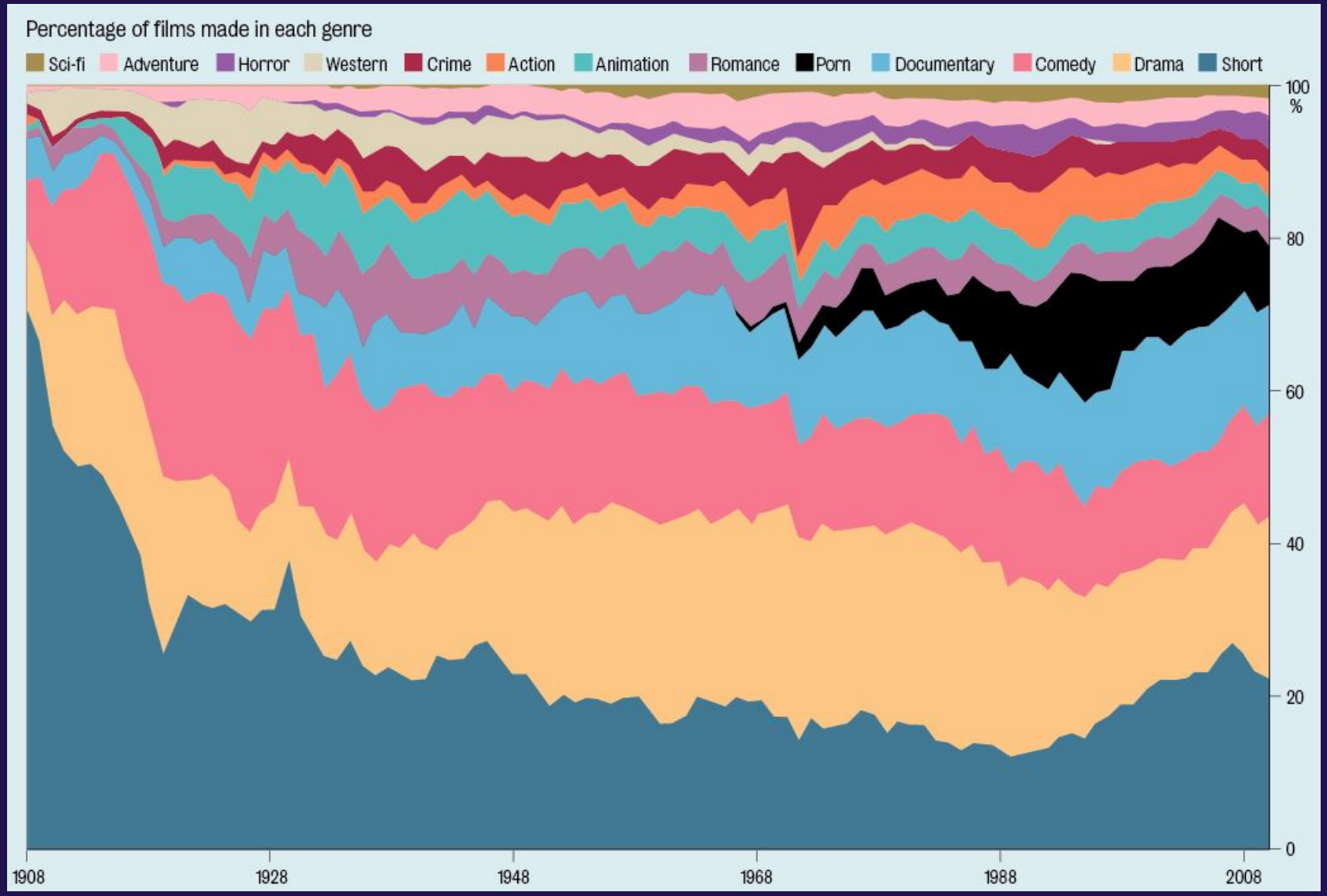
Data Presentation: Rules of Thumb

1. Clarity and simplicity are key.
2. Make it easy to read and interpret.
3. Respect visual and mathematical principles.
4. Play around with your data!
5. Cite your sources or provide your dataset.

Image source: https://commons.wikimedia.org/wiki/File:Symbol_thumbs_upzel.svg (edited for color)


WORLD
The average life expectancy in the world in 2009 was 69 years.
80
60
40
20
1960
1970
1980
1990
2000
Examples
Image source: http://projects.flowingdata.com/life-expectancy/
Source: The World Bank; Graphic by: Nathan Yau

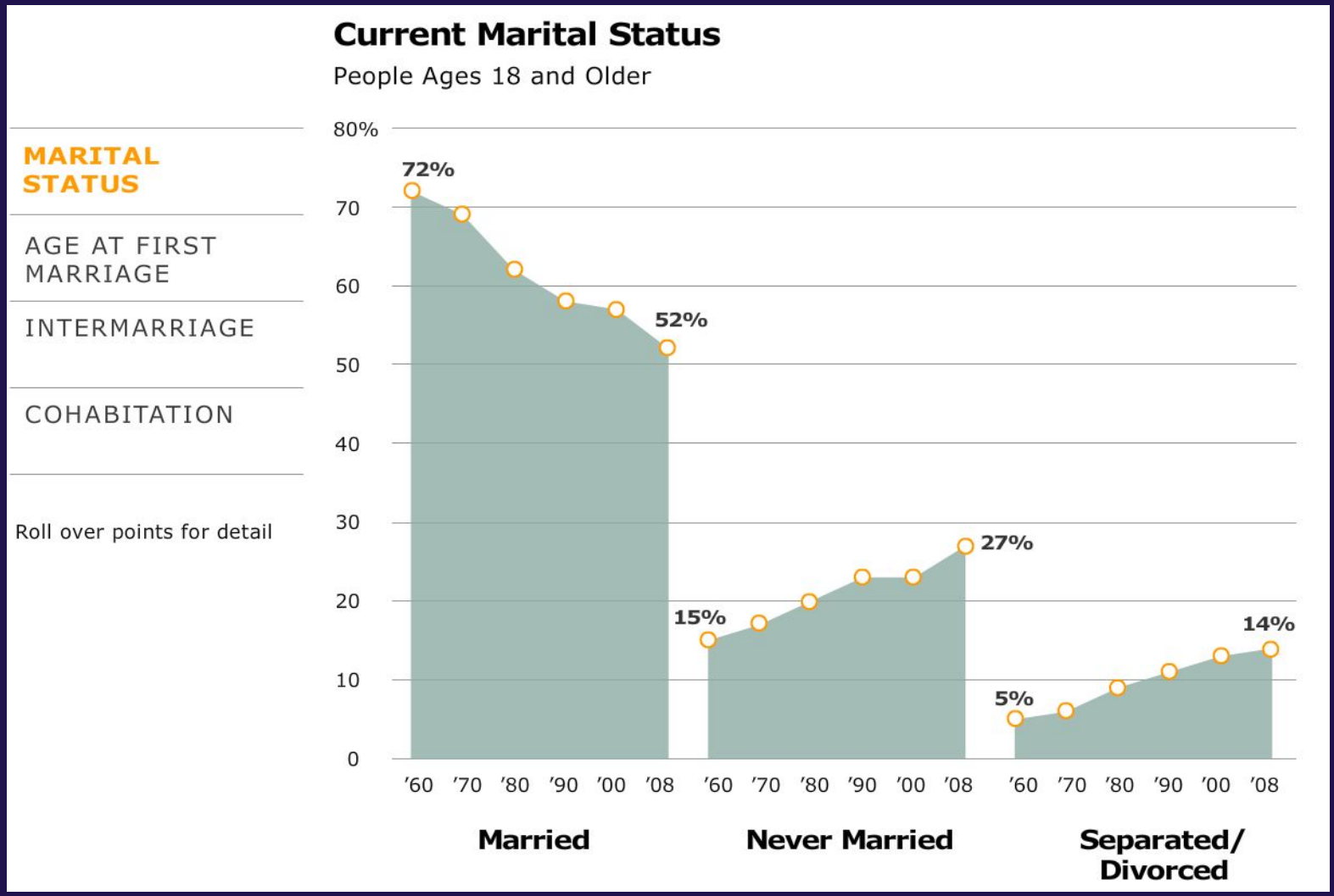
Movie Genres by Year, 1908-2008



Link: http://blog.undr.com/2012/10/movie-genres-by-year-infographic.png

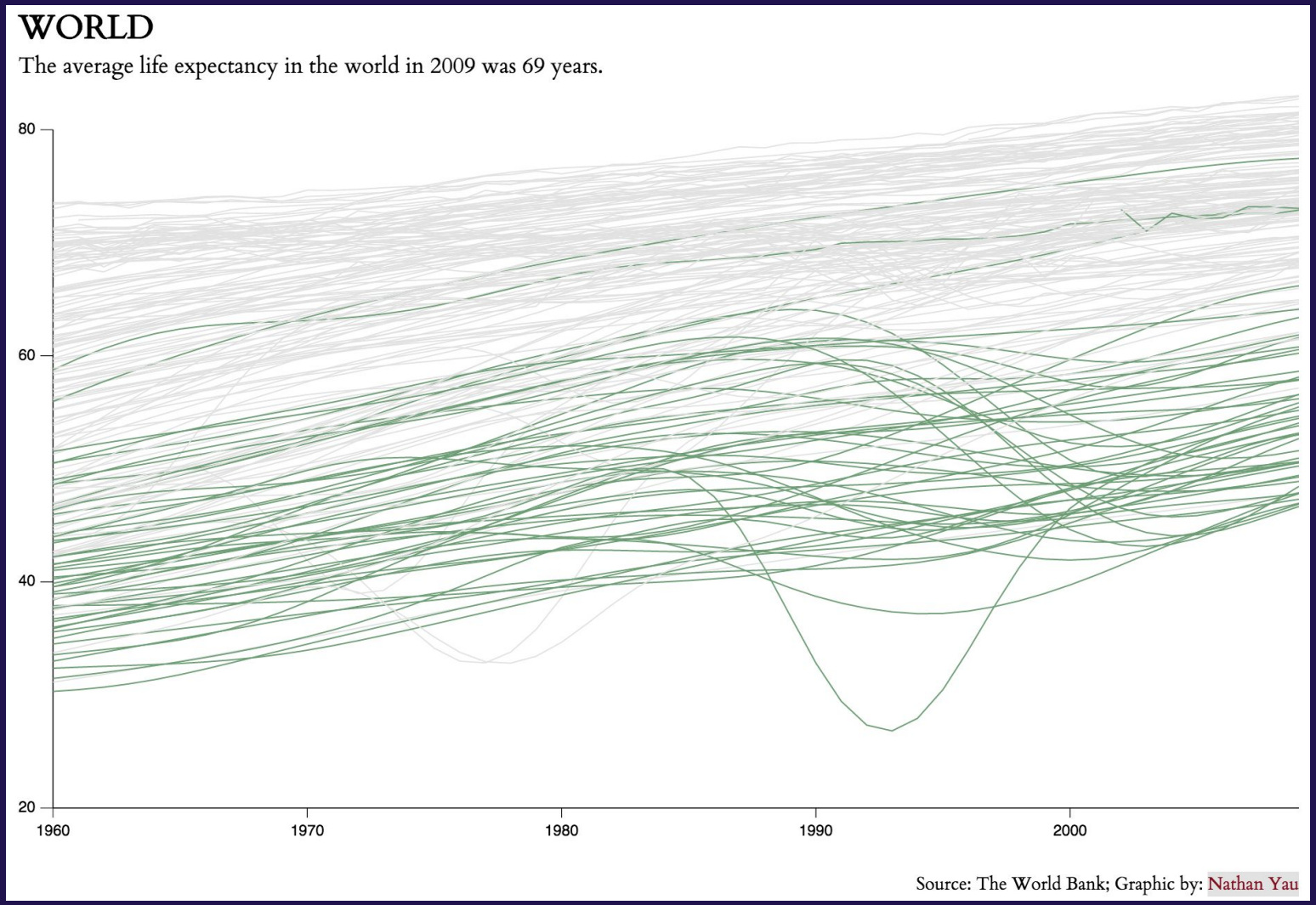
See also: article on Slate

Pew Research Center: The Changing American Family



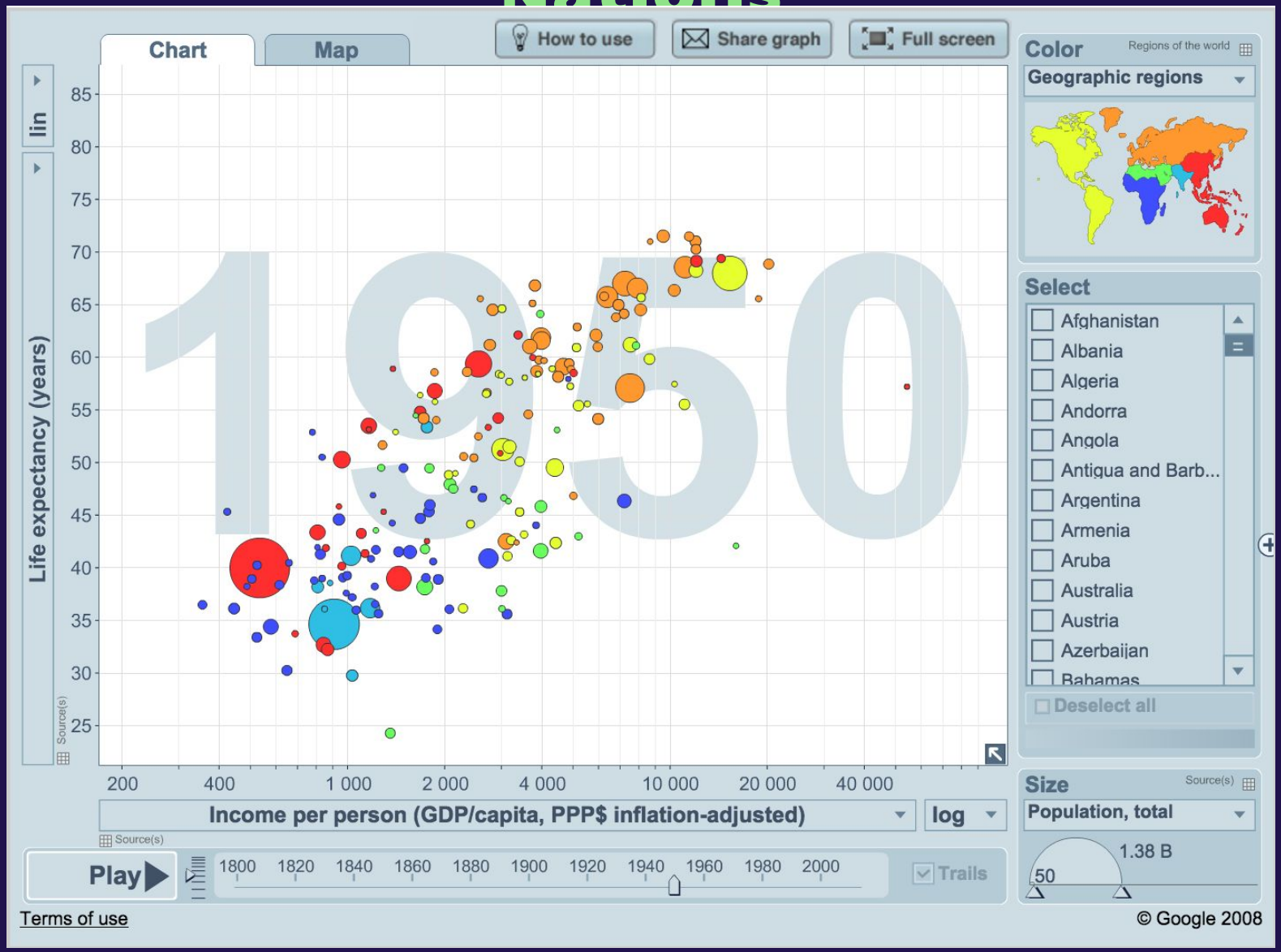
Link: [Pew Research Center]

FlowingData: Life Expectancy Around the World, 1960-2009



Link: FlowingData

GapMinder: Wealth and Health of Nations



Link: [GapMinder]

Play around with Google Public Data Explorer!

https://www.google.com/publicdata/directory

Check out its data from a variety of sources and create some charts of your own! While you work, think about:

- the data itself and its potential context
- the presentation of the data
- things you wish you could see or do with this tool

Image created with Google Public Data Explorer

Google Public Data Explorer: Share Your Findings

- What are your thoughts on context?
- What are your thoughts on presentation?
- What did you wish you could see or do with this tool?

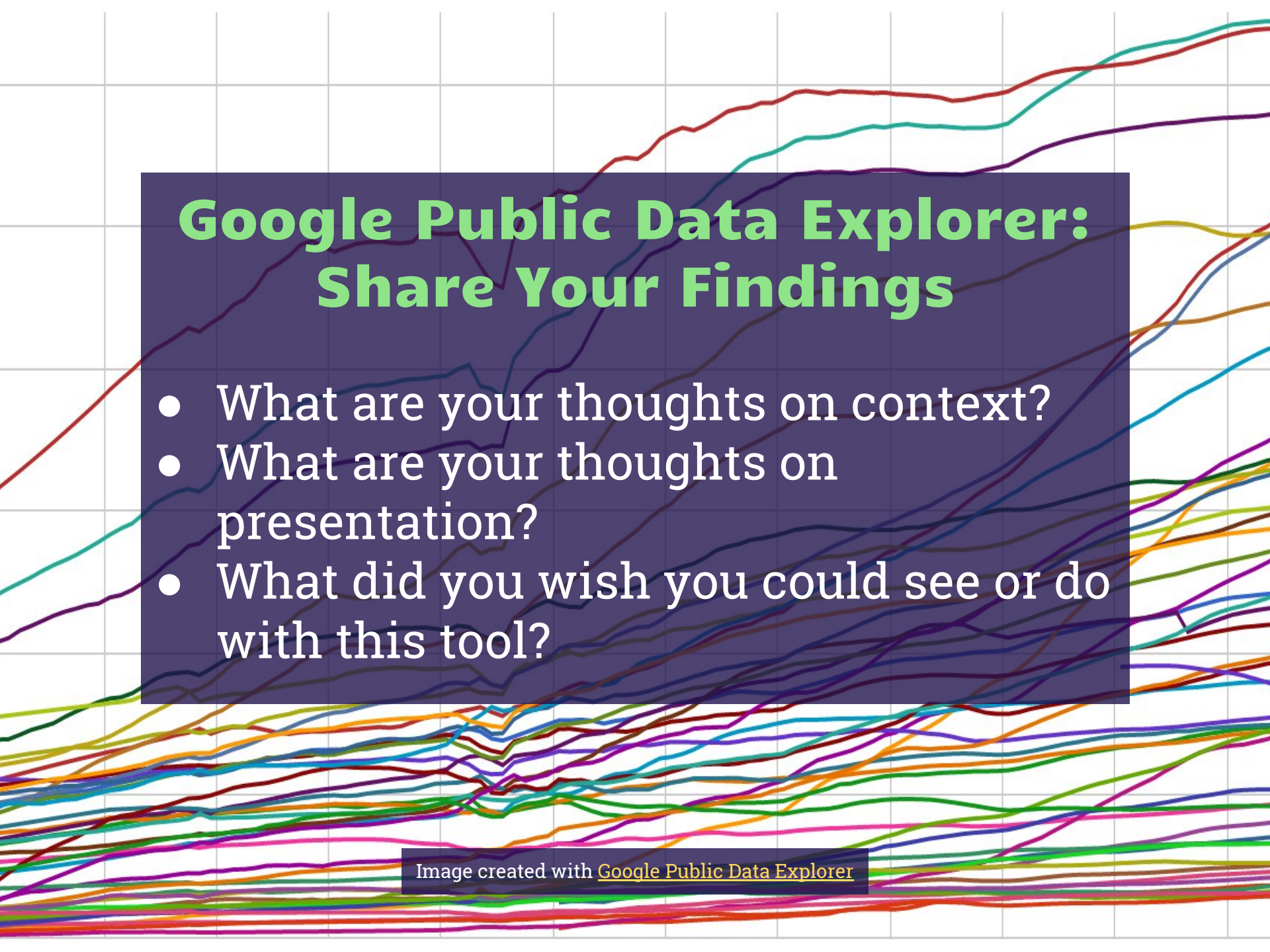


Image created with Google Public Data Explorer

In summary:

Statistical Context

Correlation ≠ causation • Ask yourself: who? • what? • when? • where • why? • how?

Data Presentation

Keep it simple. • Make it readable. • Respect visual / mathematical principles. • Play around! • Cite sources.

Any questions?

Contact us!

Tierney Steelberg: tierneyc@umich.edu
Martha Stuit: stuitm@umich.edu

Image source: https://openclipart.org/image/300px/svg_to_png/192053/remington-typewriter.png (edited for color)