2016-02-06

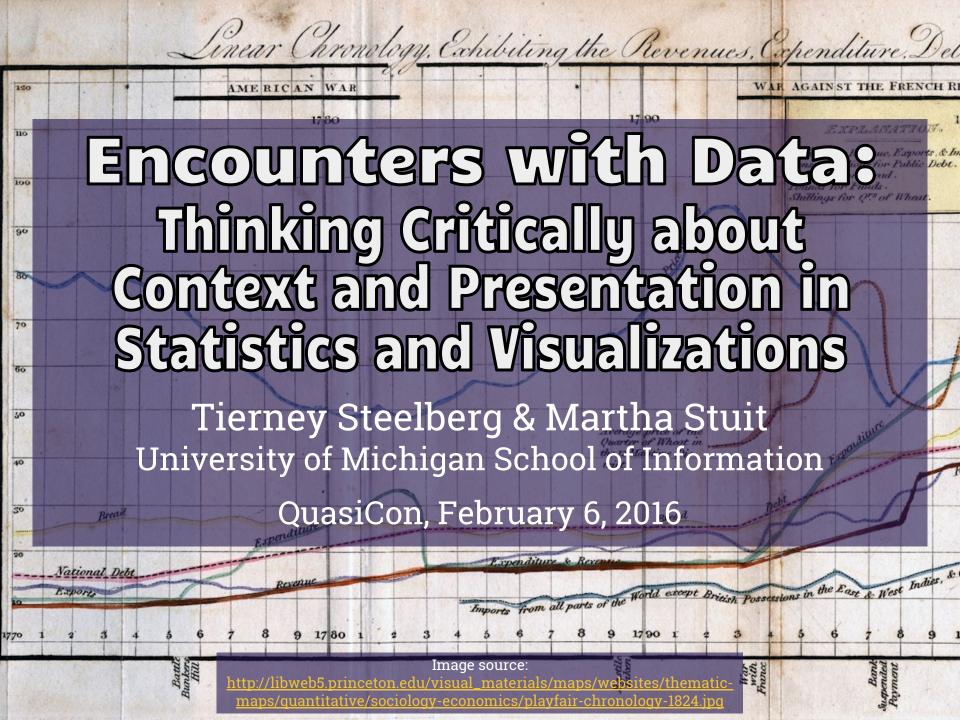
## Encounters with Data: Thinking Critically about Context and Presentation in Statistics and Visualizations

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



#### Data Literacy for High School Librarians

An IMLS-Funded Project of the University of Michigan School of Information and University Library

About Our Project

Blog

Book

Join our Email List

Our Team

Publications and Presentations

Virtual Conference

2

#### **About Our Project**

The Supporting Librarians in Adding Data Literacy Skills to Information Literacy Instruction project is a two-year project running from October 2015 through September 2017 to develop data and statistical literacy skills in high school librarians so they can better support critical comprehension skills in their students.

Over the course of the project, co-PIs Kristin Fontichiaro and Jo Angela Oehrli will partner data experts with curriculum experts to identify mini-lessons, strategies, and "rules of thumb" that librarians can nimbly weave into their instruction. The project recognizes that librarians' time (and that of their classroom colleagues) is limited and looks for the high-leverage practices with the greatest impact on student comprehension.

© 2015 Regents of the University of Michigan and made available under Creative Commons Attribution-NonCommercial-ShareAlike license unless otherwise stated. This project was made possible in part by the Institute of Museum and Library Services RE-05-15-0021-15.

#### **Creating Data Literate Students**

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

#### What's wrong with this picture?

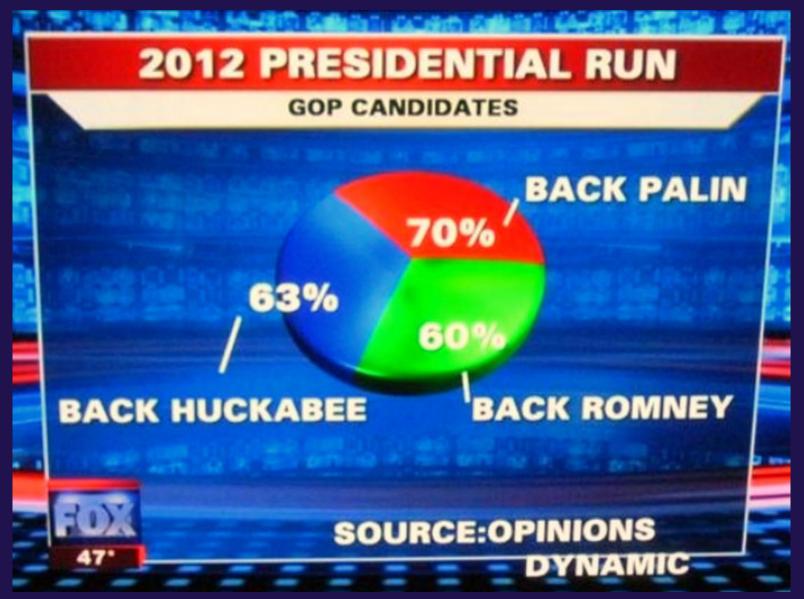


Image source: http://media.nbcchicago.com/images/1200\*900/Fox's+Pie+Chart.jpg

# Numbers often seem like facts...

## but it's okay to question them!





## What does it mean to be "data literate?"

**Data Literacy** 

**Statistical Literacy** 



## 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?

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



**USA SNAPSHOTS®** 

Who

#### 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"

Source: National Weather Service



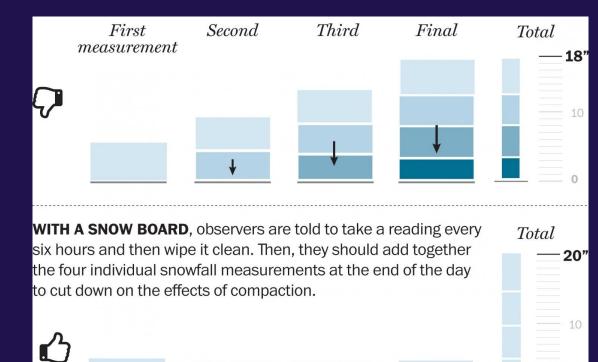
#### <u>Variable:</u> what has been counted

Image source:

http://www.freestockphotos.biz/stockphoto/15142

8

https://img.washingtonpost.com/wp-apps/imrs.php?src=ht tps://img.washingtonpost.com/blogs/capital-weather-gang /files/2016/01/snowdepth0125b.jpg&w=1484



KEVIN UHRMACHER/THE WASHINGTON POST

#### What, continued



- 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.

#### 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.

#### When

When was the data collected?



#### Where

Where was the sample conducted?

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

Rule of thumb: Correlation is not causation.

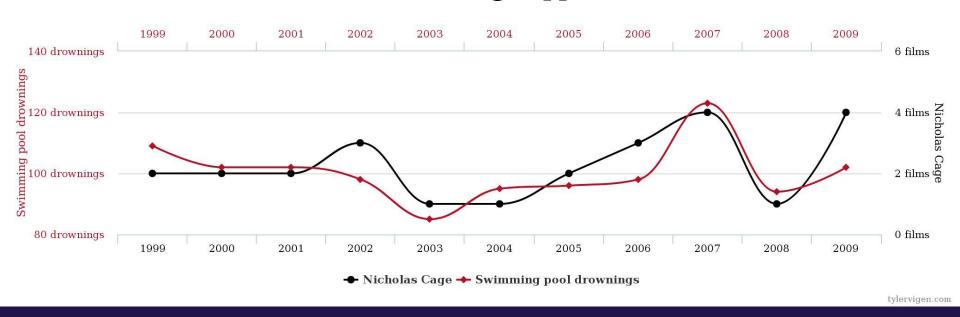
#### Where

#### Rule of thumb: Correlation is not causation.

#### Number of people who drowned by falling into a pool

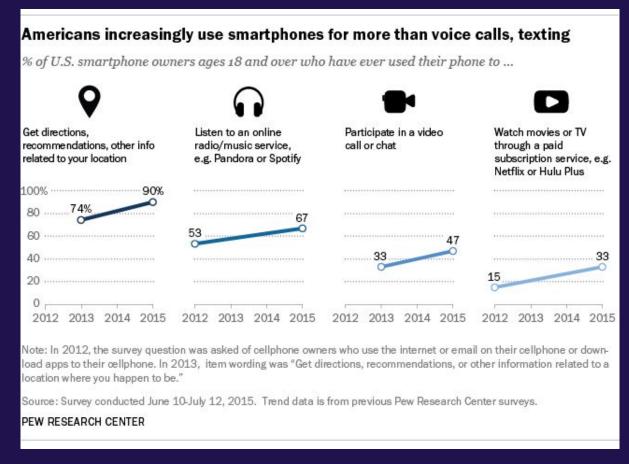
correlates with

#### Films Nicolas Cage appeared in



#### Why

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

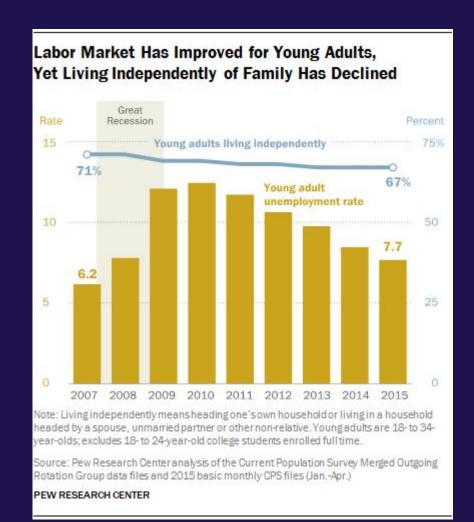


Link: Pew Research Center

#### 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?)



## 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?

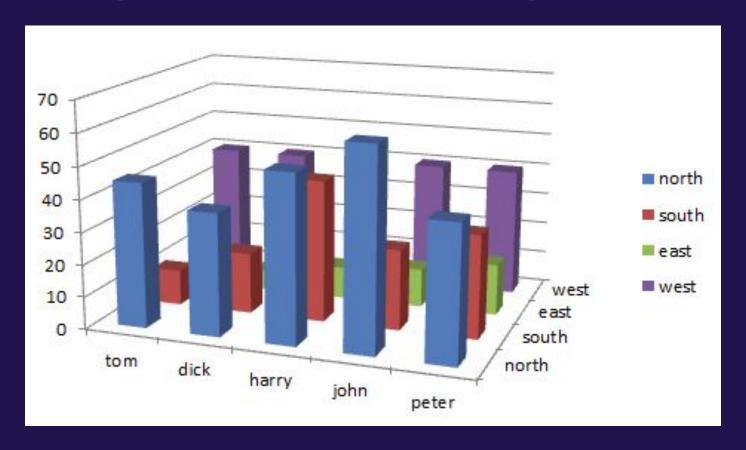


#### General Rules of Thumb for Presenting Your Data



## Clarity and simplicity are key.

#### Clarity and simplicity are key.



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

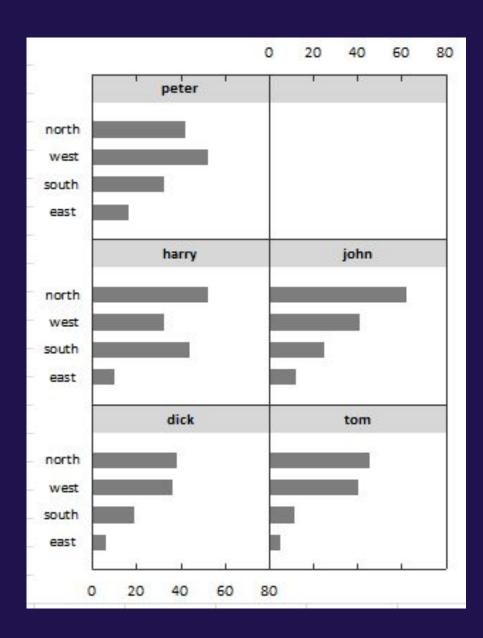
Image source: <a href="http://www.forbes.com/sites/naomirobbins/2012/06/07/trellis-plot-alternative-to-three-dimensional-bar-charts/">http://www.forbes.com/sites/naomirobbins/2012/06/07/trellis-plot-alternative-to-three-dimensional-bar-charts/</a>

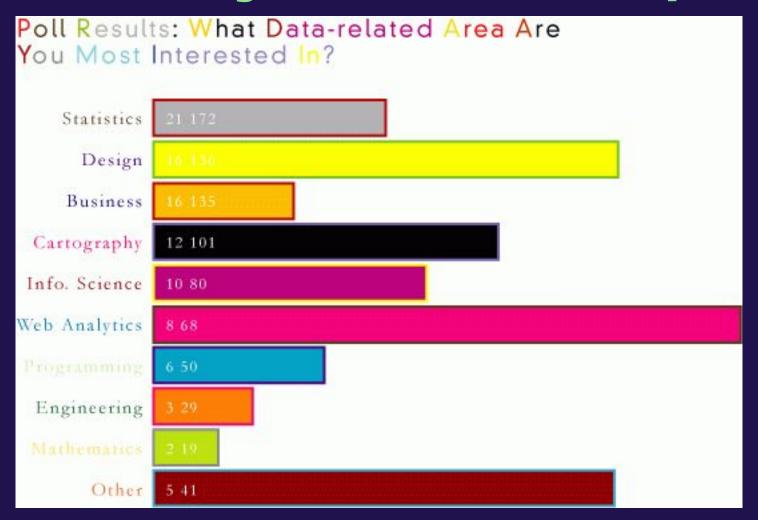
## Clarity and simplicity are key.

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

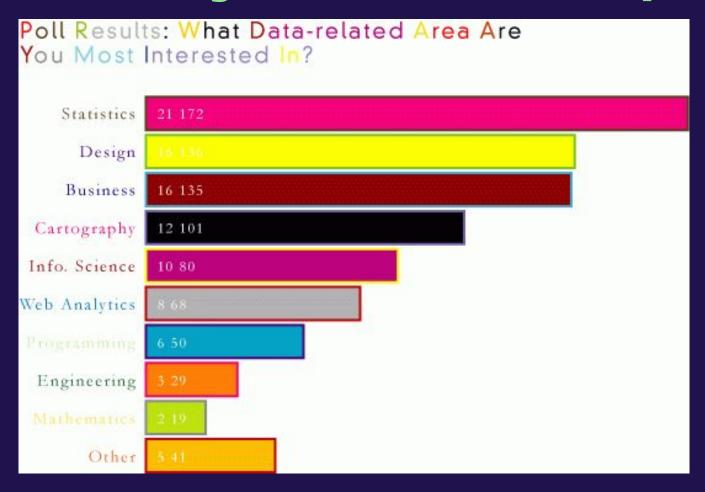
Image source:

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

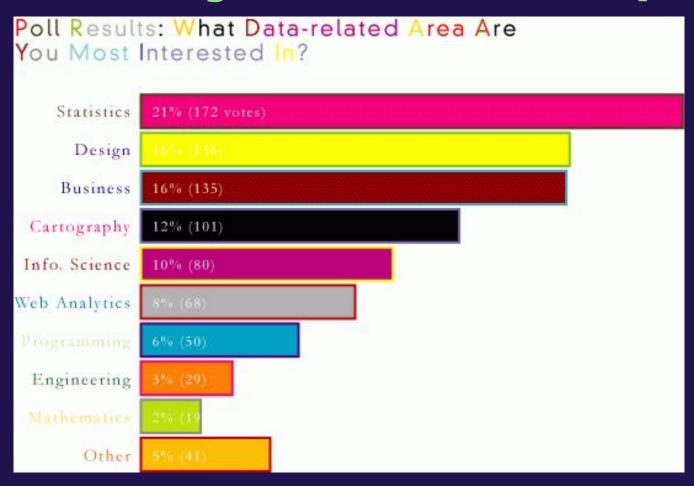




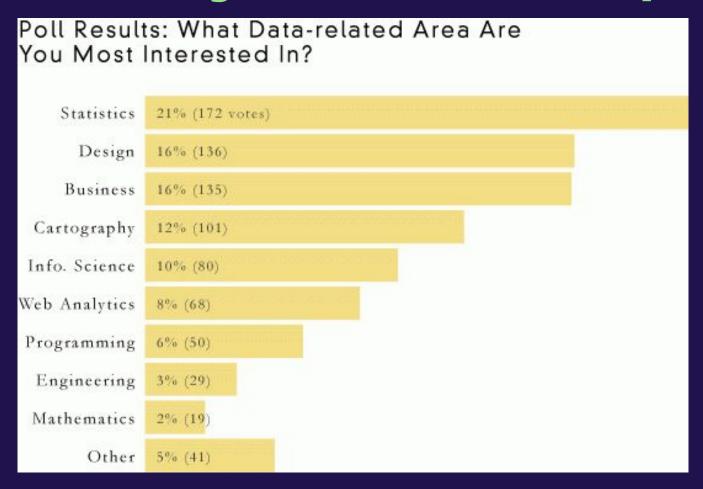
#### What's wrong with this picture?



Organize values in a meaningful order.

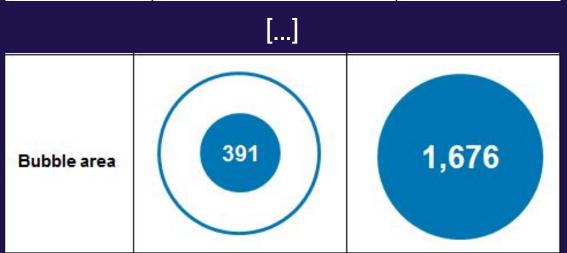


Provide a legend and labels.
Clarify units.



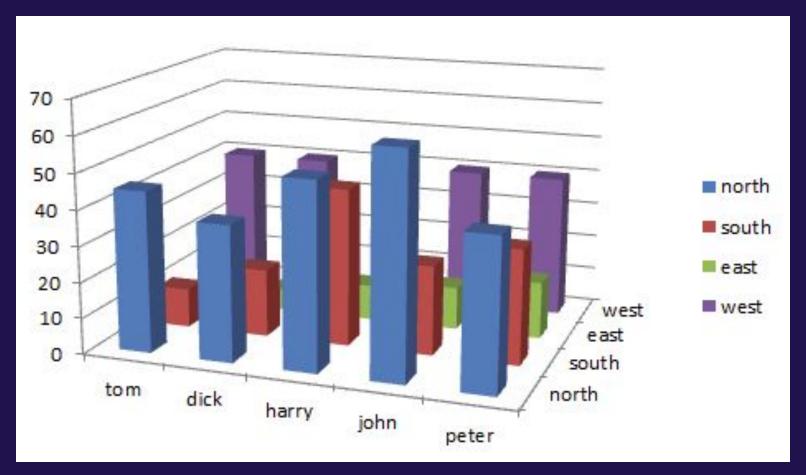
Use a simple color scheme. Avoid using color combinations that are difficult to distinguish.





Size twodimensional shapes proportionally according to their area.

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

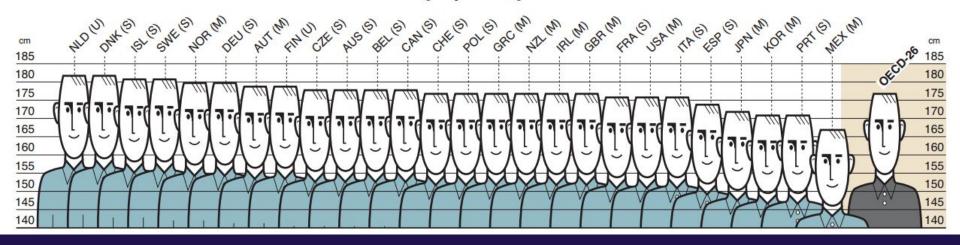


Keep things in two dimensions.

Image source:

HE5.1. Nordic and European countries are the tallest

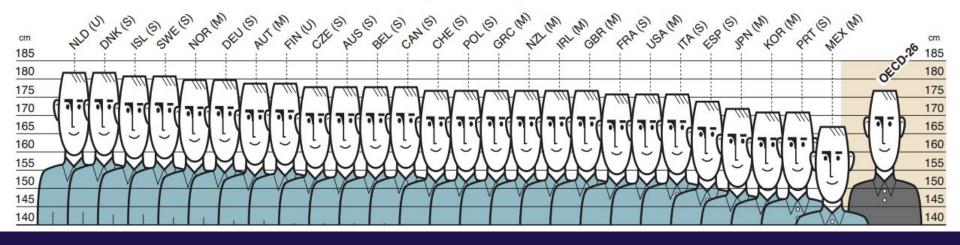
Mean heights for men aged 20 to 49



Do these icons add to the visualization?

HE5.1. Nordic and European countries are the tallest

Mean heights for men aged 20 to 49



In general, forego icons in the data visualization itself.

### 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:
<a href="https://commons.wikimedia.org/wiki/File:">https://commons.wikimedia.org/wiki/File:</a>
<a href="mailto:Microsoft\_Excel\_2013\_logo.svg">Microsoft\_Excel\_2013\_logo.svg</a>

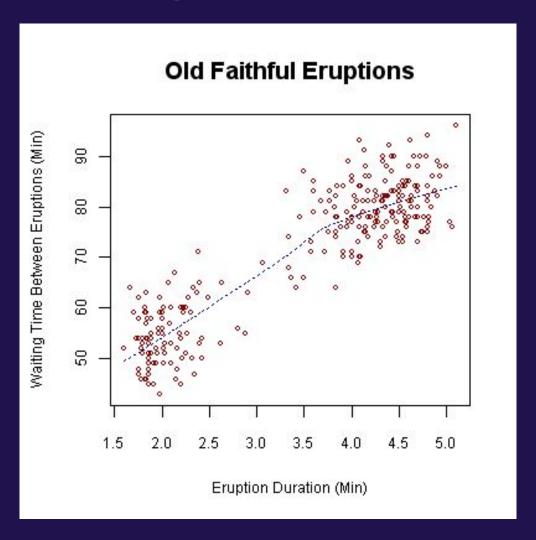
#### **Google Sheets**



Image source: <a href="http://eci511-emarsh-blog.blogspot.com/">http://eci511-emarsh-blog.blogspot.com/</a>

# 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

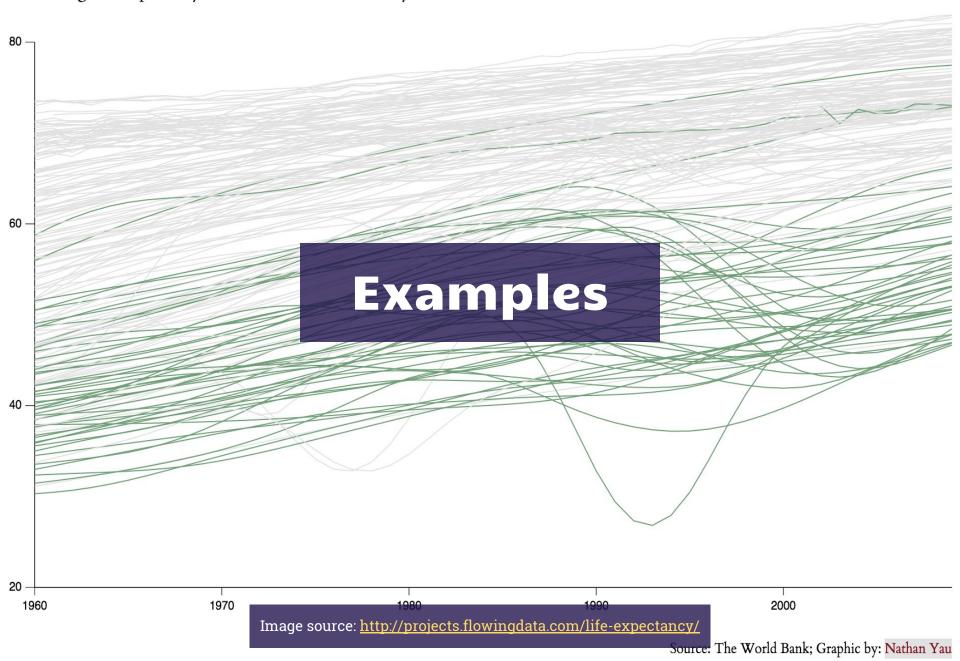
## Data Presentation: Rules of Thumb

- 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.

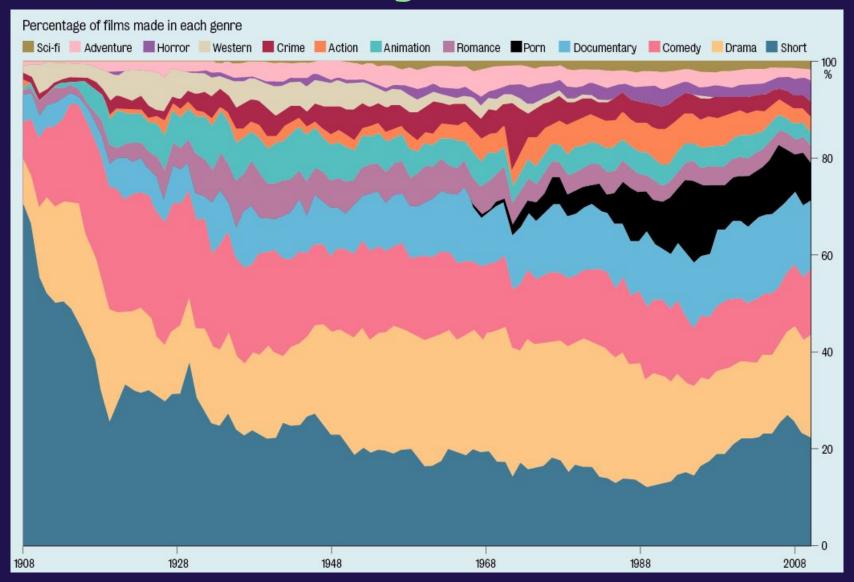


#### WORLD

The average life expectancy in the world in 2009 was 69 years.

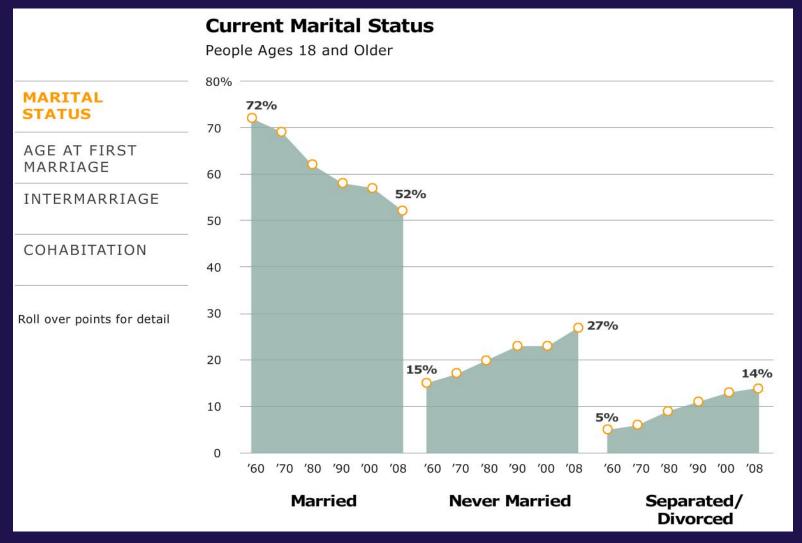


### Movie Genres by Year, 1908-2008



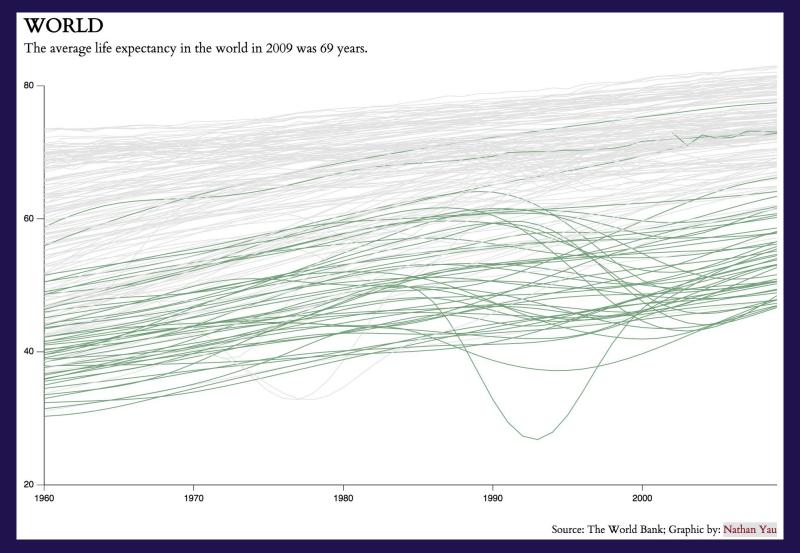
Link: <a href="http://blog.undr.com/2012/10/movie-genres-by-year-infographic.png">http://blog.undr.com/2012/10/movie-genres-by-year-infographic.png</a>
See also: <a href="mailto:article.on.slate">article on Slate</a>

### Pew Research Center: The Changing American Family



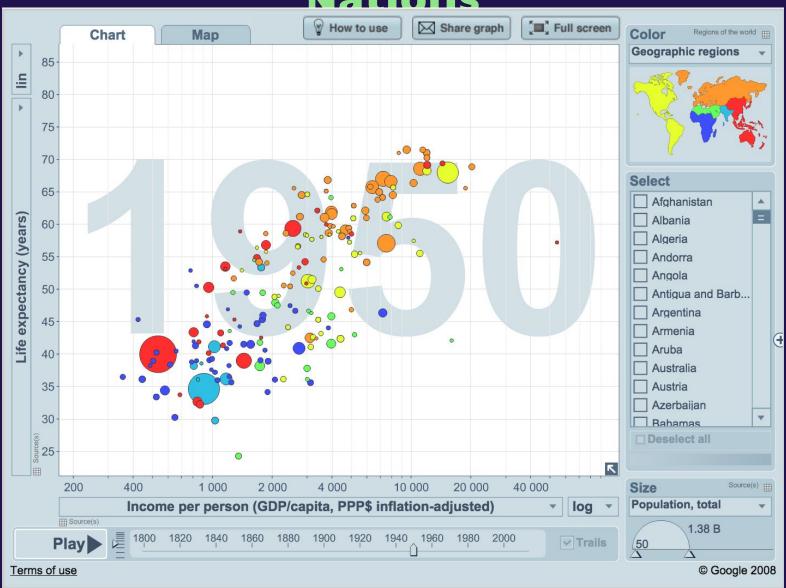
Link: Pew Research Center

### Flowing Data: 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



- 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: <u>tierneyc@umich.edu</u> Martha Stuit: <u>stuitm@umich.edu</u>

