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

Steelberg, Tierney; Stuit, Martha

http://hdl.handle.net/2027.42/146733
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:
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 Oehrl 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.

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?”
How to Approach Statistics

The

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

of statistics

Image source:
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)

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

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

WITH A SNOW BOARD, observers are told to take a reading every six hours and then wipe it clean. Then, they should add together the four individual snowfall measurements at the end of the day to cut down on the effects of compaction.
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.
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.


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 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?

<table>
<thead>
<tr>
<th></th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Get directions,</td>
<td>100</td>
<td>74</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>recommendations,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other info related to</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>your location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listen to an online</td>
<td></td>
<td>53</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>radio/music service,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.g. Pandora or Spotify</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participate in a video</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>call or chat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watch movies or TV</td>
<td></td>
<td>15</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>through a paid subscription service, e.g. Netflix or Hulu Plus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Americans increasingly use smartphones for more than voice calls, texting**

% of U.S. smartphone owners ages 18 and over who have ever used their phone to ...

Note: In 2012, the survey question was asked of cellphone owners who use the internet or email on their cellphone or download apps to their cellphone. In 2013, item wording was “Get directions, recommendations, or other information related to a location where you happen to be.”

Source: Survey conducted June 10-July 12, 2015. Trend data is from previous Pew Research Center surveys.

**PEW RESEARCH CENTER**

**Link:** [Pew Research Center](https://www.pewresearch.org)
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?

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:
Make it easy to read and interpret.
Make it easy to read and interpret.

Poll Results: What Data-related Area Are You Most Interested In?

- Statistics: 21,172
- Design: 16,136
- Business: 16,135
- Cartography: 12,101
- Info. Science: 10,80
- Web Analytics: 8,68
- Programming: 6,50
- Engineering: 3,29
- Mathematics: 2,19
- Other: 5,41

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/](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.

Respect visual and mathematical principles.

Keep things in two dimensions.

Respect visual and mathematical principles.

HE5.1. Nordic and European countries are the tallest
Mean heights for men aged 20 to 49

Do these icons add to the visualization?

Respect visual and mathematical principles.

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: 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?

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.

Image source: http://projects.flowingdata.com/life-expectancy/

Source: The World Bank; Graphic by: Nathan Yau
Movie Genres by Year, 1908-2008

Percentage of films made in each genre

Link: http://blog.undr.com/2012/10/movie-genres-by-year-infographic.png
See also: article on Slate

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

Source: The World Bank; Graphic by: Nathan Yau

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
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?
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: tierneyyc@umich.edu
Martha Stuit: stuitm@umich.edu

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