GENERAL ASSEMBLY

COMMUNICATING WITH DATA

Joseph Nelson, Data Science Immersive

AGENDA

- Introduction
- Math and stats
- Visual Design
- Coding Viz

“All models are wrong. Some are useful.”

— George Box, 1978

AGENDA

- Data Science Immersive Instructor
- **From:** Des Moines, Iowa
- **Influences:** Marc Andreessen & Ben Horowitz, Zuckerberg, Andrew Ng, Yann LeCun, Jürgen Schmidhuber
- **Likes:** Hockey, SaaS, bad data science puns, running

YOU

- Your background:

- Name
- How do you currently or need to communicate with data? (Brevity counts)

- Rank these three:
- 1. Mathematics and statistics
- 2. Visual design
- 3. Communication strength

PART 1: MATH AND STATS REVIEW

BASIC DESCRIPTIVE STATISTICS

- Mean
- Median
- Mode
- Max
- Min
- Quartile
- Inter-quartile Range
- Variance
- Correlation

MEAN

- What is the mean?
- What is another name for the mean?

MEDIAN

- What is the median?
- How do you find the median?

MEDIAN

- What is the median?
- How do you find the median?
- Bonus: Why might the median be advantageous instead of the mean? When does this condition NOT hold?

MODE

- What is the mode?

- The mode of a set of values is the value that occurs most often.

- A set of values may have more than one mode or no mode.

MEAN VS MEDIAN...

HOW TO LIE WITH STATISTICS: EXERCISE 1

- For each picture:
- 1) What could go wrong
- 2) How to fix it

- Work in pairs!

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

Score Category	Number of Tests
90s	0
80s	0
70s	2
60s	1
50s	1
40s	1
30s	1
20s	1

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

HOW TO LIE WITH STATISTICS

QUARTILES AND THE INTER QUARTILE RANGE

- Quartiles divide a rank-ordered data set into four equal parts.
- The values that divide each part are called the first, second, and third quartiles; and they are denoted by Q1, Q2, and Q3, respectively.
- The interquartile range (IQR) is a measure of variability, based on dividing a data set into quartiles. It is the “middle 50” of your data. Also called the H-spread. IQR = Q3-Q1
- Outliers: Q1 - 1.5(IQR), Q3 + 1.5(IQR)

CRITERIA FOR GOOD VISUALIZATION

PART 2: EFFECTIVE VISUAL DESIGN

CRITERIA FOR GOOD VISUALIZATION

- We'll break this section into parts:

- 1.) Examining best examples
- 2.) Distilling those examples
- 3.) Discussing tools to achieve those examples

CRITERIA FOR GOOD VISUALIZATION

- First, we'll look at examples. We'll discuss, as a class, what makes those examples successful.

- 1: http://fivethirtyeight.com/features/lionel-messi-is-impossible/

CRITERIA FOR GOOD VISUALIZATION

- First, we'll look at examples. We'll discuss, as a class, what makes those examples successful.

- 1: http://fivethirtyeight.com/features/lionel-messi-is-impossible/
- 2: http://www.nytimes.com/interactive/2012/10/15/us/politics/swing-history.html?_r=2&

AN ASIDE ON DATA JOURNALISM

CRITERIA FOR GOOD VISUALIZATION

- 1. Simplified
- 2. Easy to Interpret
- 3. Clearly Labeled

- Bonus: 4. Interactivity

CRITERIA FOR GOOD VISUALIZATION

- 1. Simplified
- 2. Easy to Interpret
- 3. Clearly Labeled

- Ask yourself:
- Who is my target audience?
- What do they already know, and what do they need to know?
- How does my project affect this audience? How might they interpret (or misinterpret) the data?

TOOLS

- So you have developers you can ask to do this for you…

- D3.js: https://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/
- Tableau: https://public.tableau.com/en-us/s/blog/2015/07/analyzing-airbnb-data

- Excel + PPT (for real!)
- Python (Bokeh, Seaborn, matplotlib)
- R (ggplot)

CHARTJUNK

A Pyramid
Sky
Light Side of the Pyramid
Shady Side of the Pyramid
imgflip.com

CHARTJUNK

CHARTJUNK

GRAPHJUNK

90%
of Americans
Support
universal
background
checks
83%
of gun owners
Hillary Clinton
@HillaryClinton
Follow
Dear Congress,
Let's get this done.

GRAPHJUNK

90%
of Americans
Support
universal
background
checks
83%
of gun owners
Hillary Clinton
@HillaryClinton
Follow
Dear Congress,
Let's get this done.

People who know
how to make
Venn Diagrams
Hillary's graphic
design staff
In reply to Hillary Clinton
Michael Deppisch @deppisch · May 20
.@hillaryclinton
3.7K
5.4K

IT'S A BIPARTISAN PROBLEM

NEXT STEPS

- Keep in touch!

- www.josephofiowa.com/entropy

@josephofiowa

josephnelson@generalassemb.ly