

# CRAFTING STORIES WITH DATA

Joseph Nelson, Represently

---

## AGENDA

---

- ▶ Introduction
- ▶ Math and stats
- ▶ Selecting a Metric
- ▶ Visual Design
- ▶ Communication Tips

 GENERAL ASSEMBLY

---

# INTROS

---

## AGENDA

---

“ ”  
All models are wrong. Some are useful.

— George Box, 1978

## AGENDA

---

- Co-Founder, CEO Represently 
- **From:** Des Moines, Iowa
- **Influences:** a16z, Andrew Ng
- **Likes:** Hockey, bad data science puns, biking, data viz, new tech (especially for good)



FLEISHMANHILLARD



---

# 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

# FROM 30,000 FEET

---

## **WHAT IS CRAFTING STORIES WITH DATA?**

---

- ▶ On your paper... (But not all of it!)
  
- ▶ 1. What is data storytelling?
  
- ▶ 2. What elements are required to tell a good data story?
  
- ▶ 3. What skills are required of a data storyteller?

---

## LET'S HAVE A QUICK LOOK

---

- ▶ 1. [How Americans Spend their Days](#) (Flowing Data)
- ▶ 2. [Swing Voters and Elastic States](#) (FiveThirtyEight)
- ▶ 3. [The Data Behind a Season without Snow Days](#) (Medium, self)
- ▶ 4. [Why Underdogs Do Better in Hockey Than Basketball](#) (Vox)

# MATH AND STATS

# AGGREGATING DATA

*(GROAN)*

---

## 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.



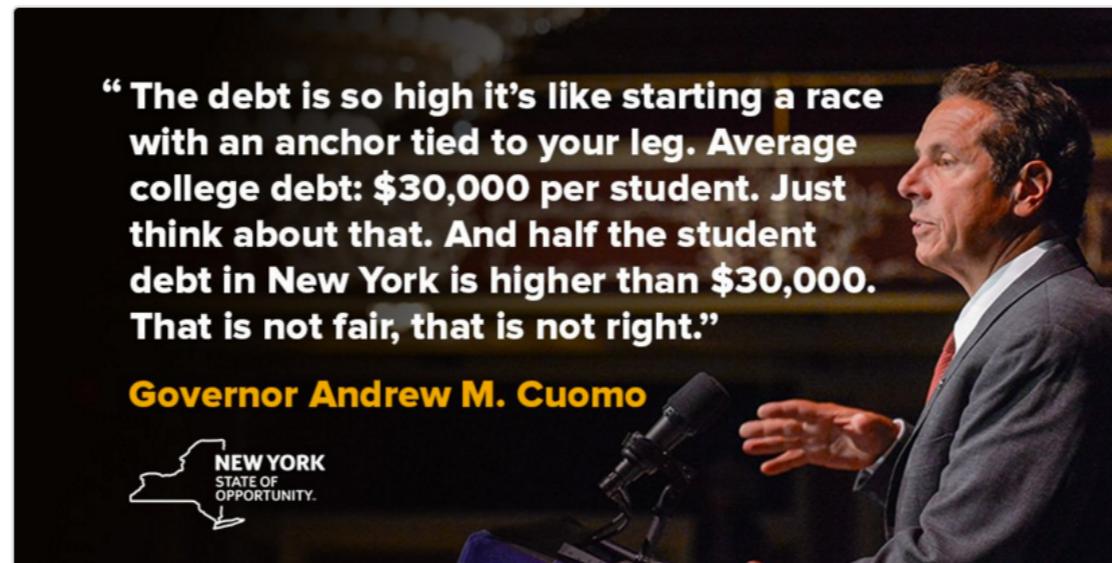
## MEAN VS MEDIAN...



Andrew Cuomo   
@NYGovCuomo

Follow

Student debt is so high it's like starting a race with an anchor tied to your leg. That's why we're going to provide free tuition in NY.



RETWEETS  
**146**

LIKES  
**389**



5:35 PM - 4 Jan 2017



146

389

---

## HOW TO LIE WITH STATISTICS: EXERCISE 1

---

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

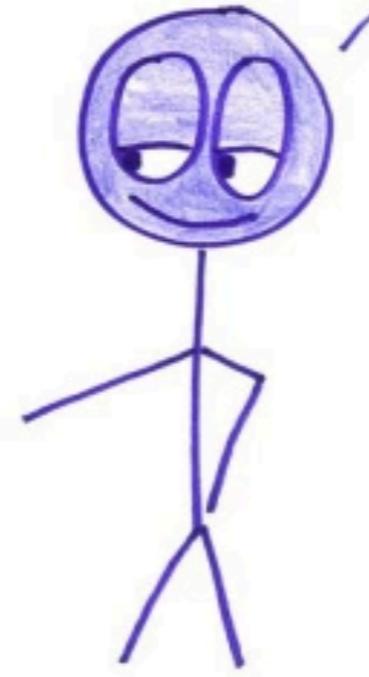
# HOW TO LIE WITH STATISTICS

Mean

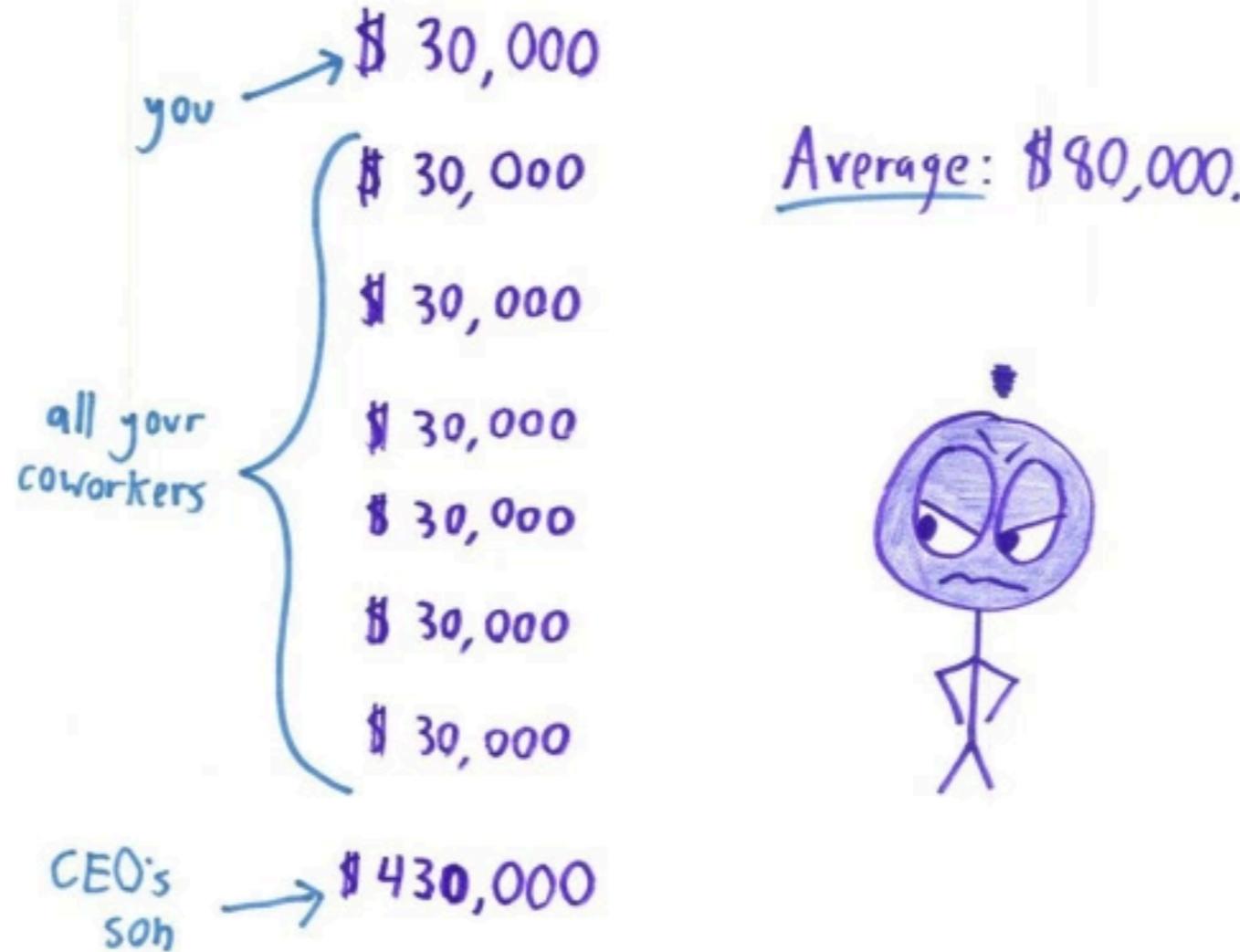
What would my  
starting salary be?



I'll put it this way:  
our average starting  
salary is \$80,000!



# HOW TO LIE WITH STATISTICS



# HOW TO LIE WITH STATISTICS

Median

So, why should I  
invest with you?

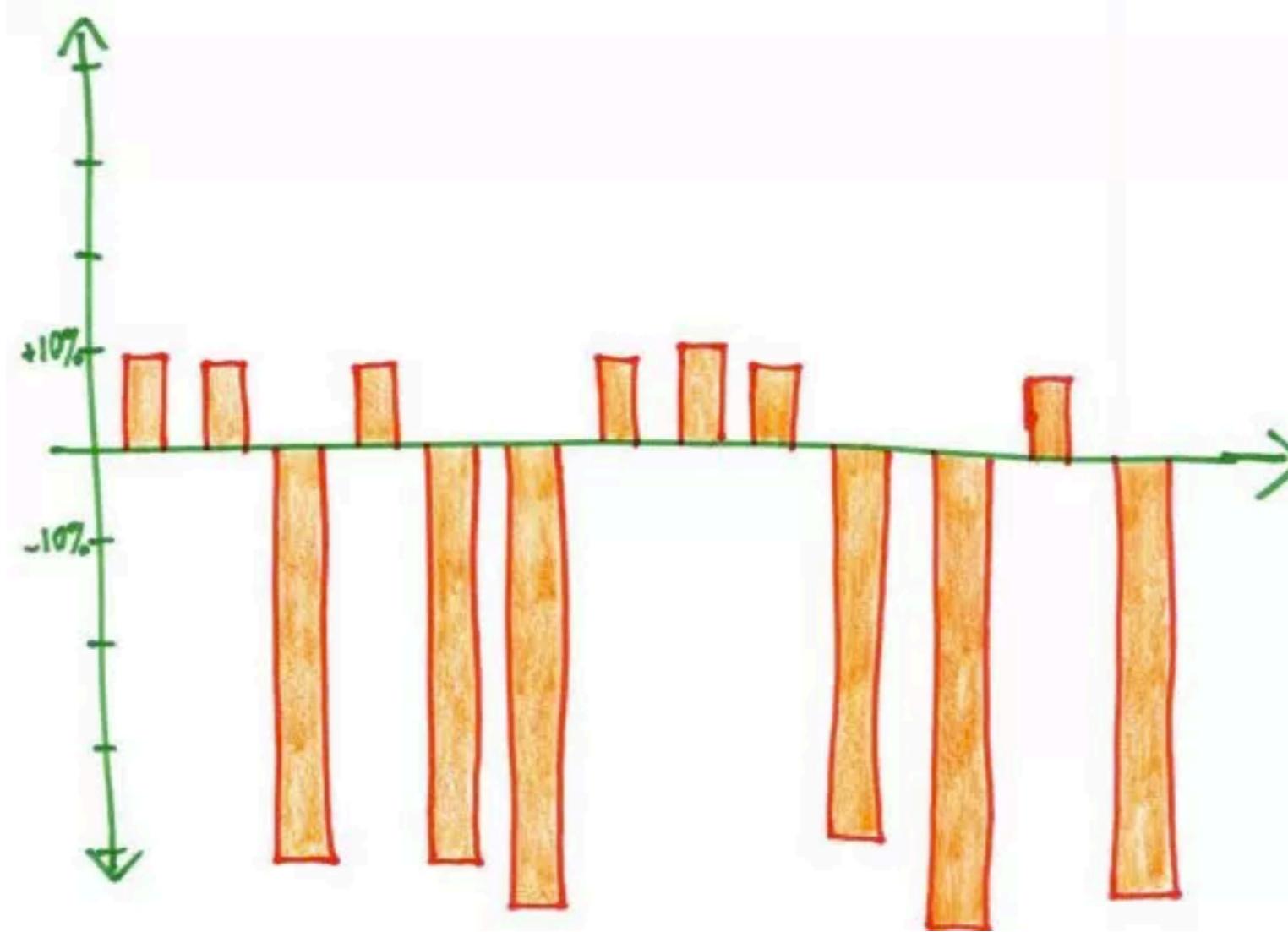


Well, not to brag, but  
my fund has a median  
gain of 8% per year!



# HOW TO LIE WITH STATISTICS

---



# HOW TO LIE WITH STATISTICS

Mode

How are you doing  
on your tests?



My modal category  
is 70-80%!



# HOW TO LIE WITH STATISTICS

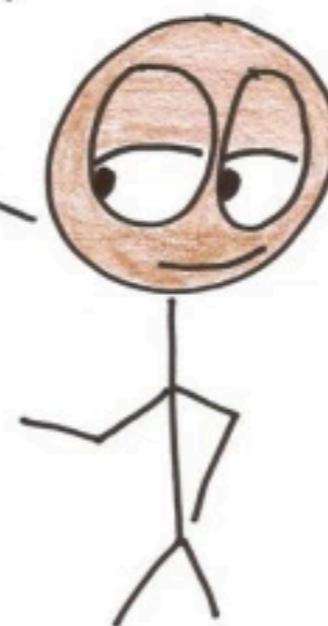


Score Category	Number of Tests
90s	0
80s	0
70s	2
60s	1
50s	1
40s	1
30s	1
20s	1

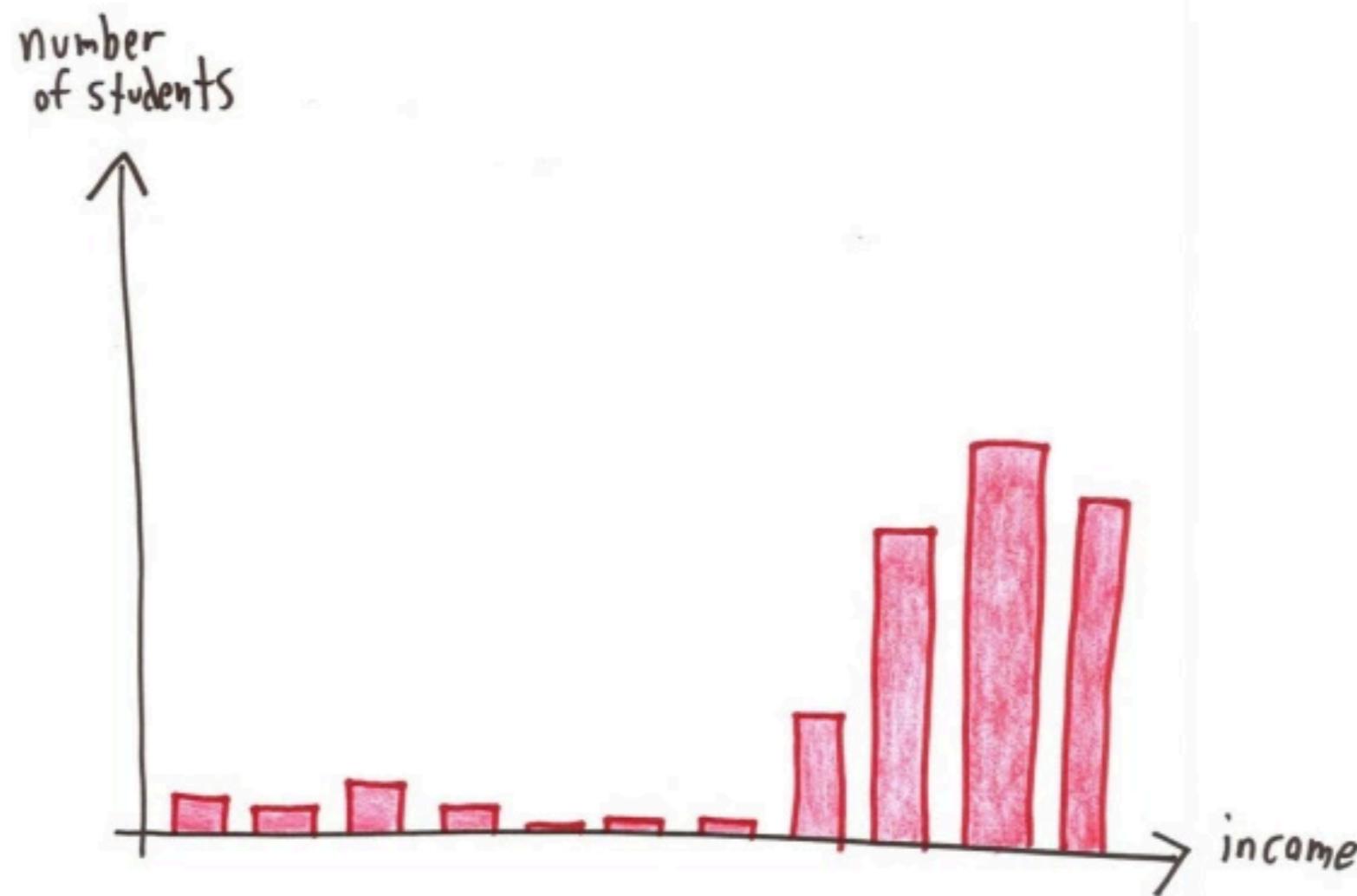
# HOW TO LIE WITH STATISTICS

Range

Our students come from a  
wide range of  
Socioeconomic  
backgrounds...



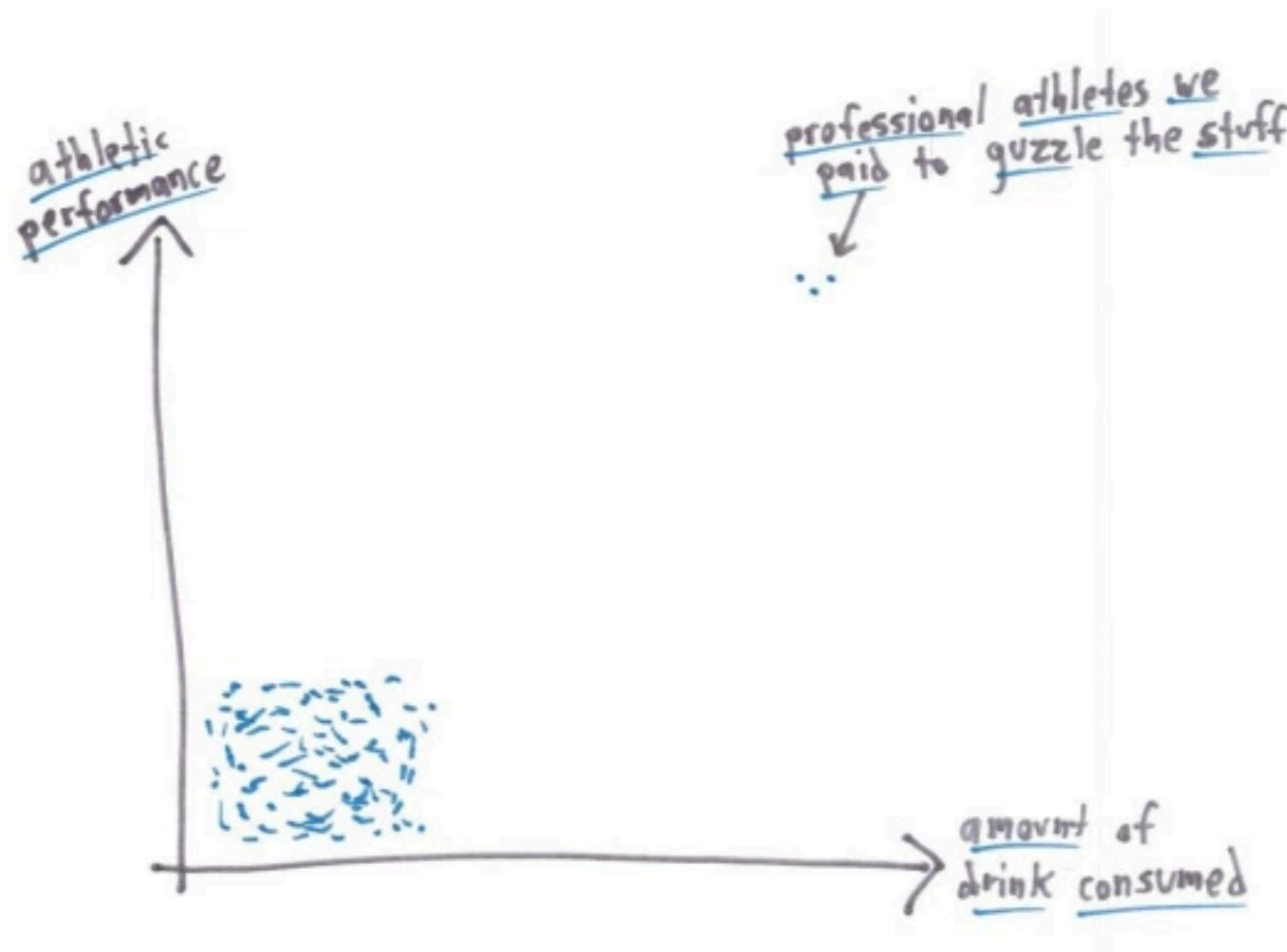
# HOW TO LIE WITH STATISTICS



# HOW TO LIE WITH STATISTICS



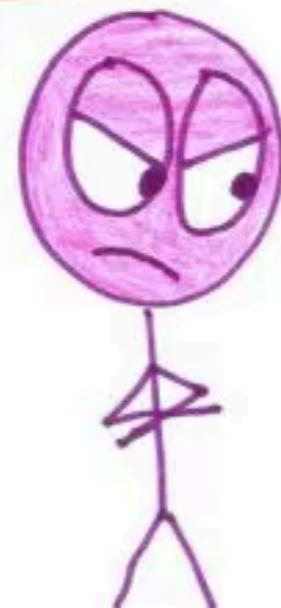
# HOW TO LIE WITH STATISTICS



# HOW TO LIE WITH STATISTICS

## Variance

These results are  
a disaster!

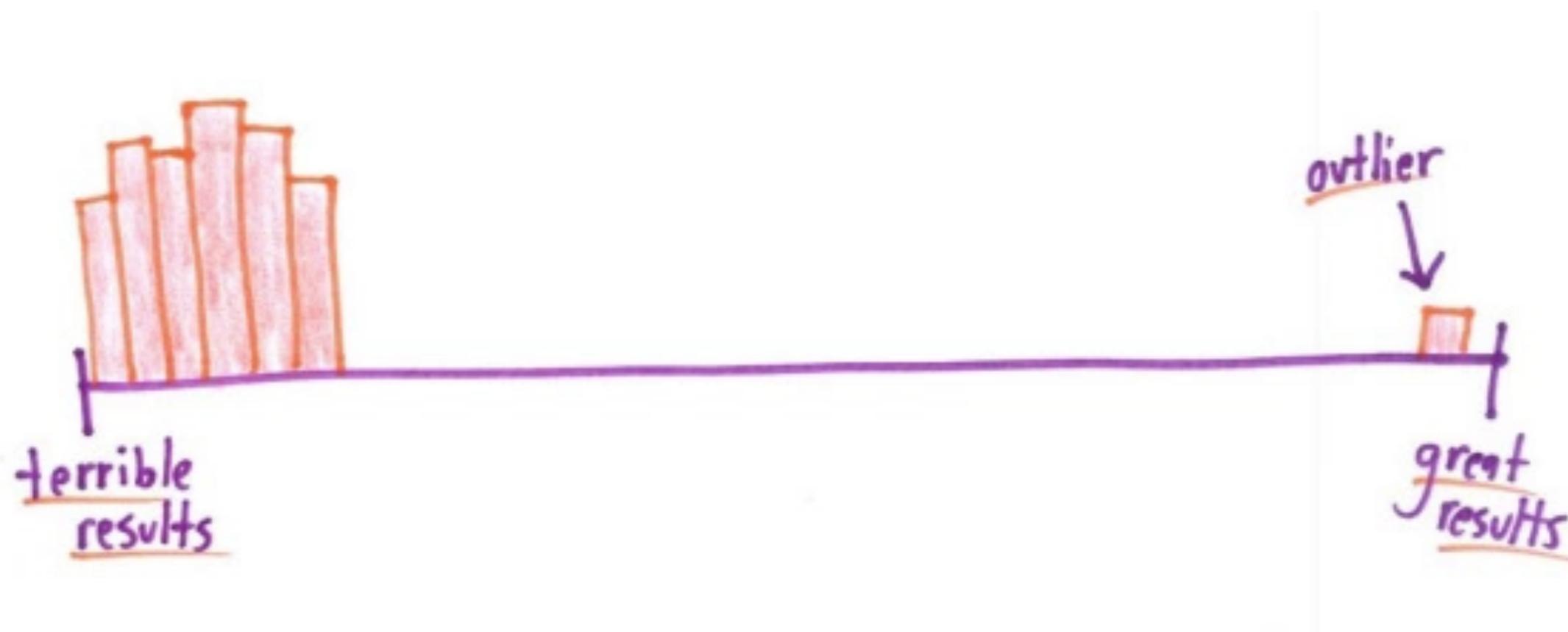


Sure, they look bad,  
but there's a lot of variance!



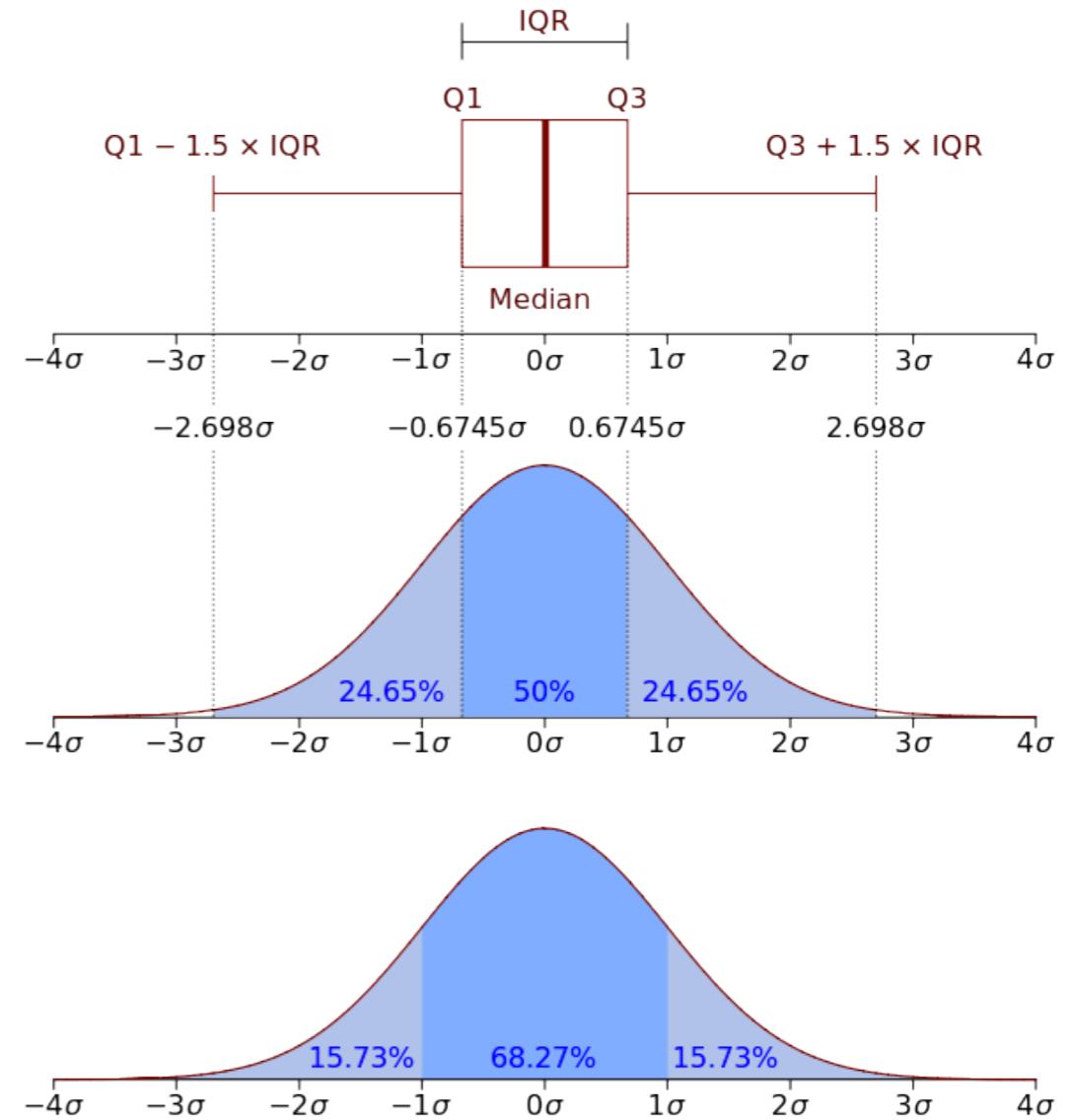
Don't rush  
to judgment.

# HOW TO LIE WITH STATISTICS



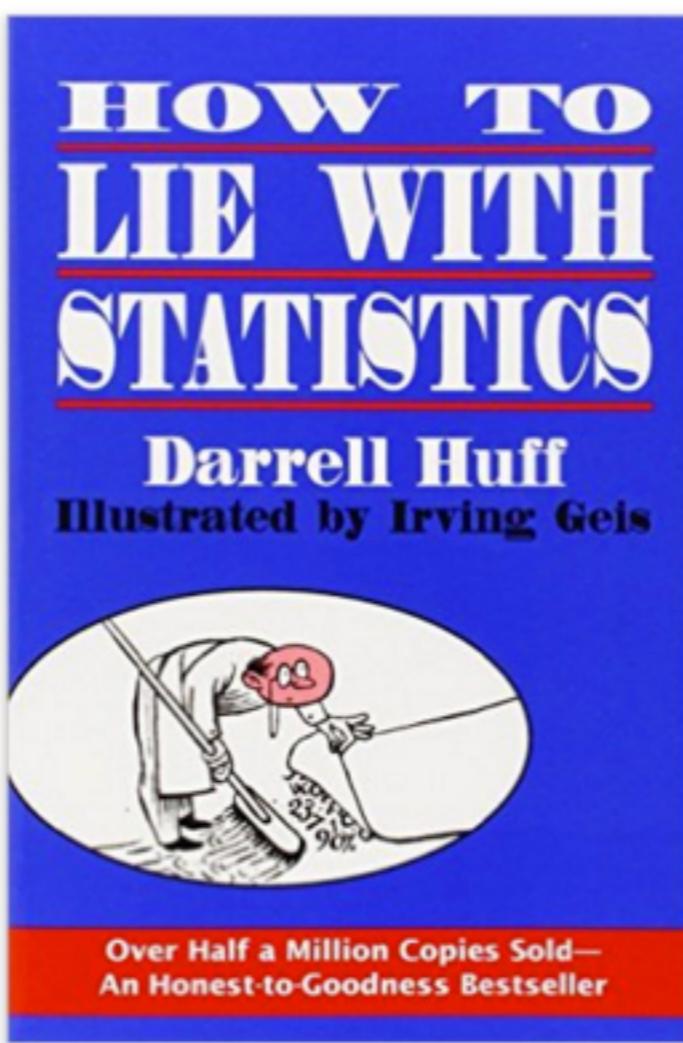
## QUARTILES AND THE INTER QUARTILE RANGE

- 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 Q<sub>1</sub>, Q<sub>2</sub>, and Q<sub>3</sub>, 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 = Q<sub>3</sub>-Q<sub>1</sub>
- Outliers: Q<sub>1</sub> - 1.5(IQR), Q<sub>3</sub> + 1.5(IQR)



## CRITERIA FOR GOOD VISUALIZATION

---



# METRICS (BUSINESS STORIES)

---

## A CONVERSATION ON METRICS...

---

- Choosing a metric is (almost) everything.
- What are you optimizing?
- Why are you optimizing for that?
- Does that capture the full picture? (No)

---

## **YOU RUN A MOBILE APPLICATION COMPANY CALLED SMACEBOOK**

---

- What metric do you choose to prove your success?

---

You run a mobile application company called smacebook

---



- Site visit
- App download
- Profile creation
- Usage
- Referral

---

## YOU RUN A MOBILE APPLICATION COMPANY CALLED SMACEBOOK

---

- What metric do you choose to prove your success?
- Monthly active users (MAU) is a top choice.
- Facebook waited for Harvard University to demonstrate 50% of the campus to have DAILY active usage before launching to Stanford. This was controversial.
- Source: <http://a16z.com/2016/07/16/network-effects-event/>

NPS

---



UBER

Your Opinion Is Powerful.

We'd like to hear it.

Joseph, how likely are you to recommend riding with Uber to friends and family?



Not at all likely

Extremely Likely

---

## A CONVERSATION ON METRICS...

---

- Can metrics mislead us?

Date	Users
1/1/16	3
1/2/16	5
1/3/16	9
1/4/16	10
1/5/16	17
1/6/16	19
1/7/16	22
1/8/16	26
1/9/16	30

## A CONVERSATION ON METRICS...

---

- Percentage growth is sensitive to small base rate changes.

Date	Users	Percent Growth
1/1/16	3	0
1/2/16	5	0.6666666667
1/3/16	9	0.8
1/4/16	10	0.1111111111
1/5/16	17	0.7
1/6/16	19	0.117647059
1/7/16	22	0.157894737
1/8/16	26	0.181818182
1/9/16	30	0.153846154

## A CONVERSATION ON METRICS...

---

- Percent growth of percent growth tells an even different story

Date	Users	Percent Growth	“Growth Rate”
1/1/16	3	0	0
1/2/16	5	0.666666667	0
1/3/16	9	0.8	0.2
1/4/16	10	0.111111111	-0.861111111
1/5/16	17	0.7	5.3
1/6/16	19	0.117647059	-0.831932773
1/7/16	22	0.157894737	0.342105263
1/8/16	26	0.181818182	0.151515152
1/9/16	30	0.153846154	-0.153846154

---

## YOU RUN A MOBILE APPLICATION COMPANY CALLED SMACEBOOK

---

- To summarize:
- Metrics are situational specific.
- We can tell different (and misleading) stories with each.
- Growth is new and retained. Be mindful.

---

## SO, GENERAL METRICS PRINCIPLES?

---

- ▶ Let's generalize:
- ▶ 1. Metrics should be easy to understand and use.
- ▶ 2. Metrics should be easily replicated.
- ▶ 3. Metrics must be actionable
- ▶ Read more: <https://hbr.org/2015/07/identify-the-marketing-metrics-that-actually-matter>

---

## YOUR TURN

---

- You came to tonight's workshop to communicate your data more effectively. Applied learning should be the key takeaway. And, fortunately, we can all learn from one another. Let's take advantage of that.
- **5MIN:**
- Exercise: in your workplace, what determines you've been successful? What *should* the metric that determines your success be? What does this successfully capture?
- You will be asked to justify this for your classmates 😊

# EFFECTIVE VISUAL DESIGN

# MAKE IT COMPELLING

---

## 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

---

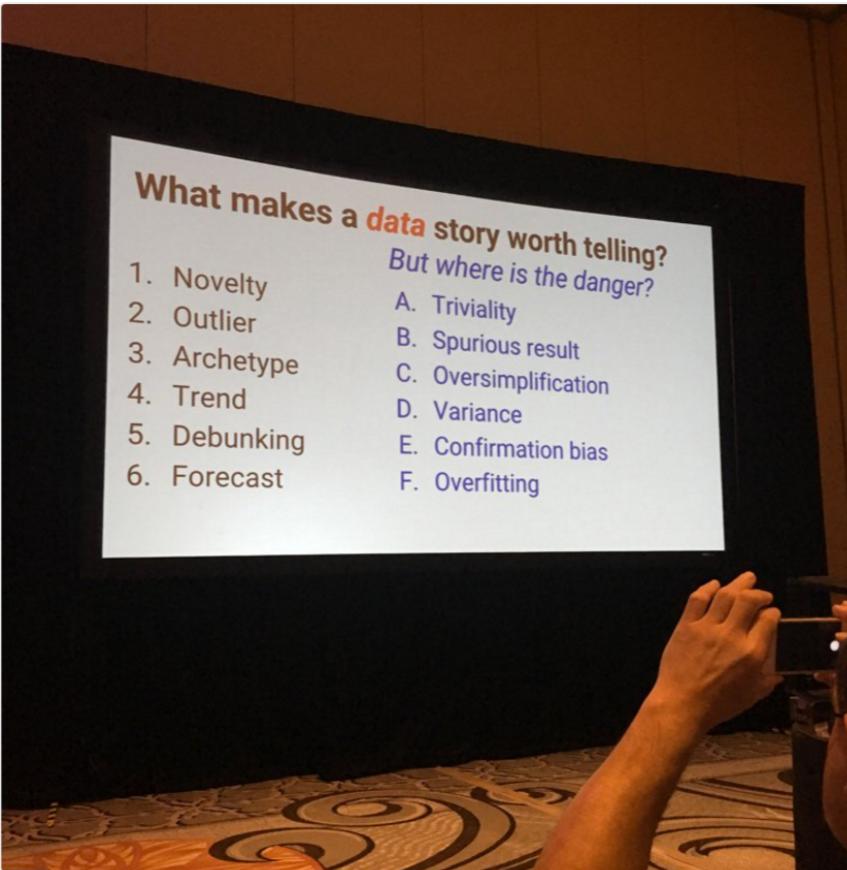
- First, we'll look at examples. We'll discuss, as a class, what makes those examples successful.
- 1: <http://fivethirtyeight.com/features/lionel-messi-is-impossible/>
- 2: [http://www.nytimes.com/interactive/2012/10/15/us/politics/swing-history.html?\\_r=2&](http://www.nytimes.com/interactive/2012/10/15/us/politics/swing-history.html?_r=2&)

## AN ASIDE ON DATA JOURNALISM

Hadley Wickham  [@hadleywickham](#)

[Follow](#) ▾

Six types of data story + their dangers from  
[@andrewflowers #rstudioconf](#)



What makes a **data** story worth telling?

But where is the danger?

- 1. Novelty
- 2. Outlier
- 3. Archetype
- 4. Trend
- 5. Debunking
- 6. Forecast

- A. Triviality
- B. Spurious result
- C. Oversimplification
- D. Variance
- E. Confirmation bias
- F. Overfitting

RETWEETS 594 LIKES 1,144



---

## CRITERIA FOR GOOD VISUALIZATION

---

- ▶ 1. Simplified
  - ▶ 2. Easy to Interpret
  - ▶ 3. Clearly Labeled
- 
- ▶ Bonus: 4. Interactivity

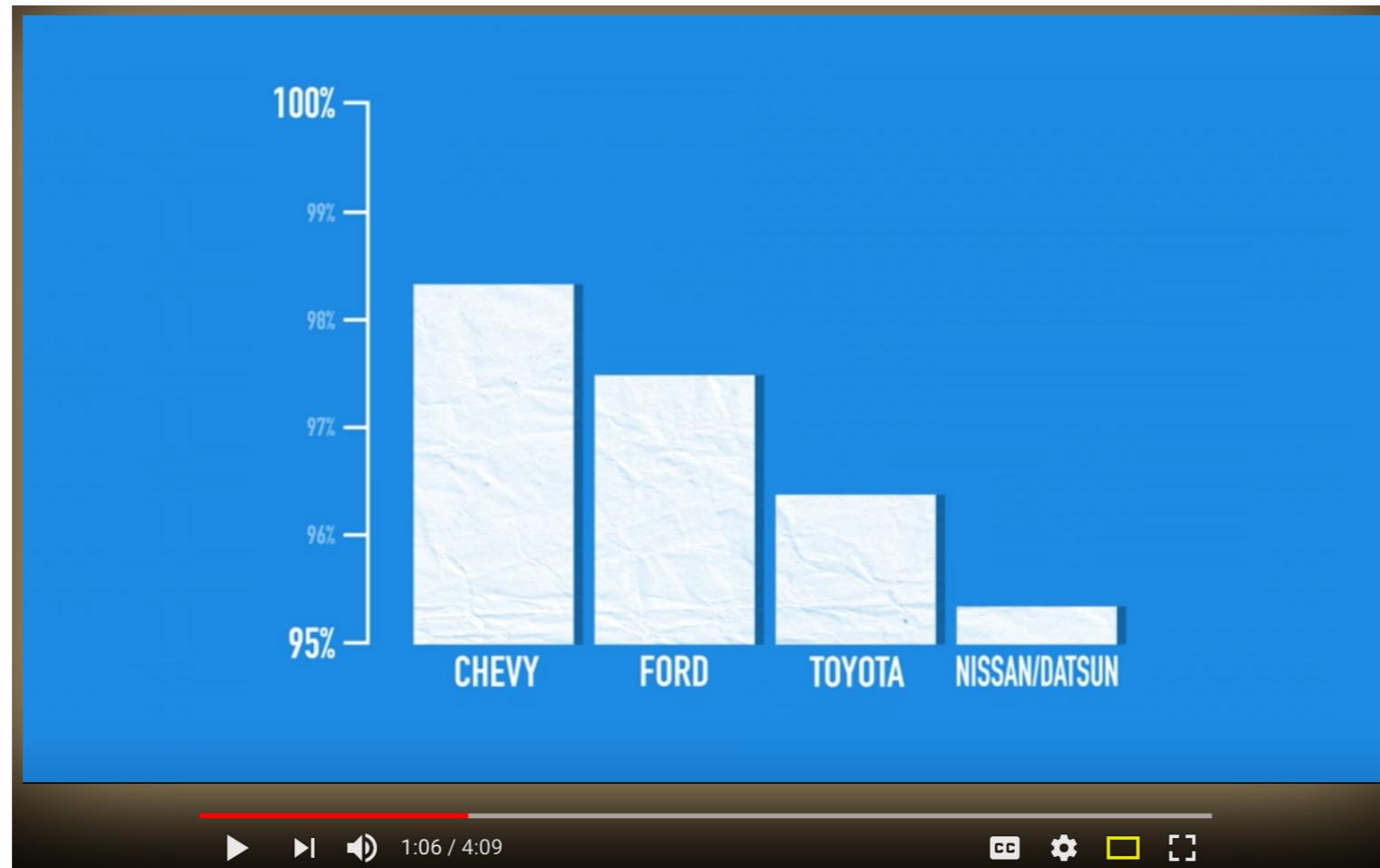
---

## 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?

## YES, AXES MATTER. LOTS.



<https://www.youtube.com/watch?v=E91bGT9BjYk>

---

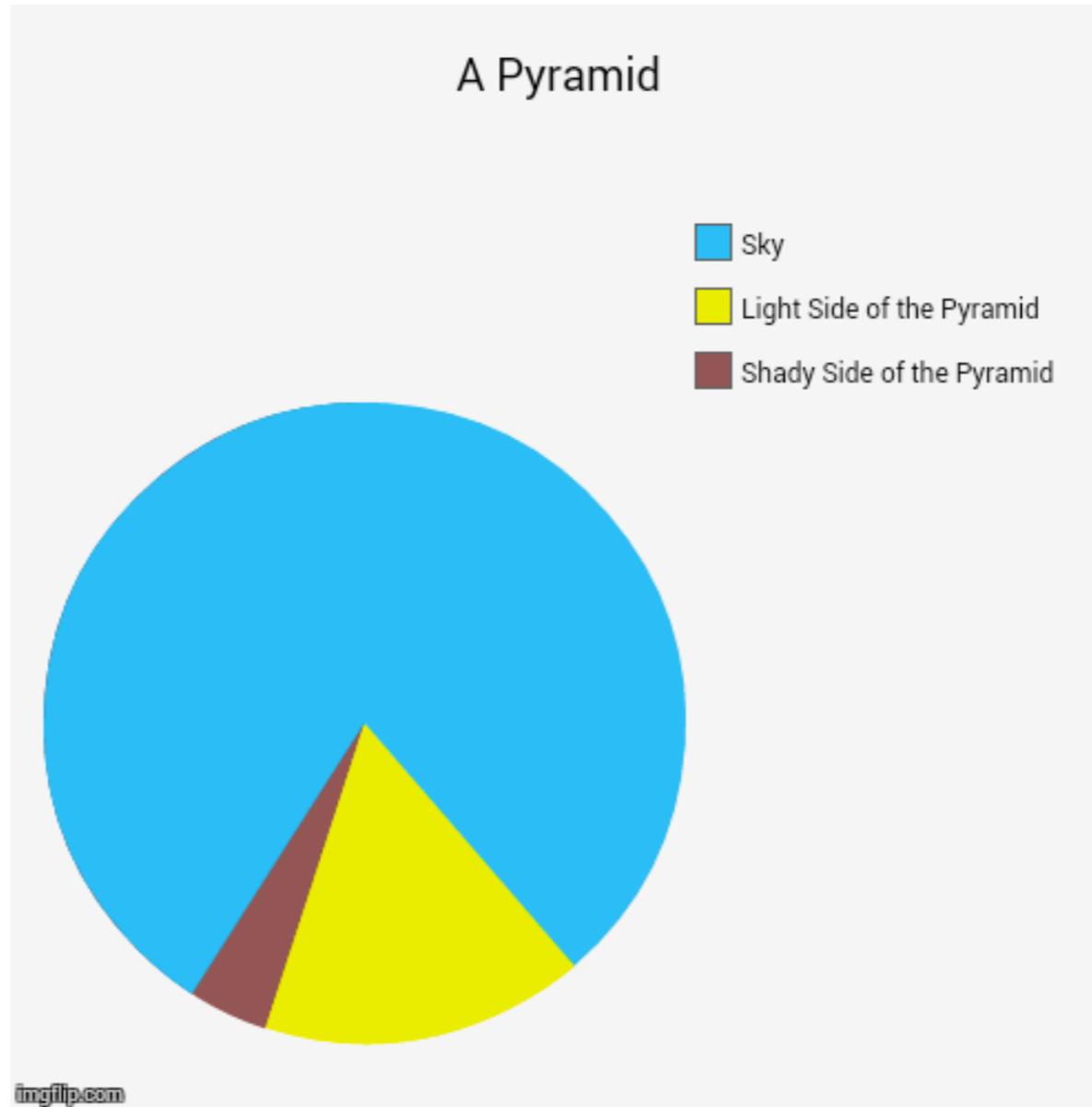
## TOOLS

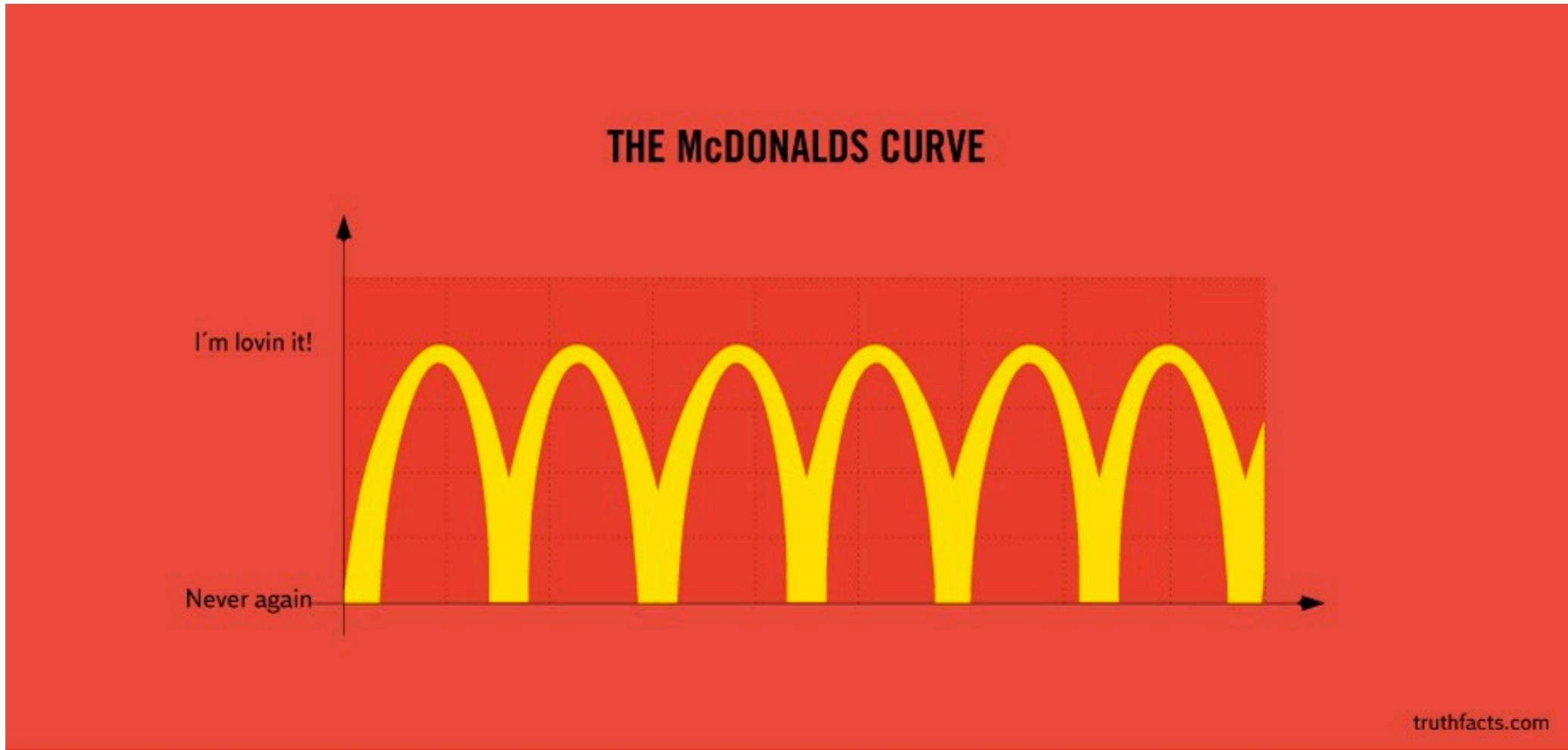
---

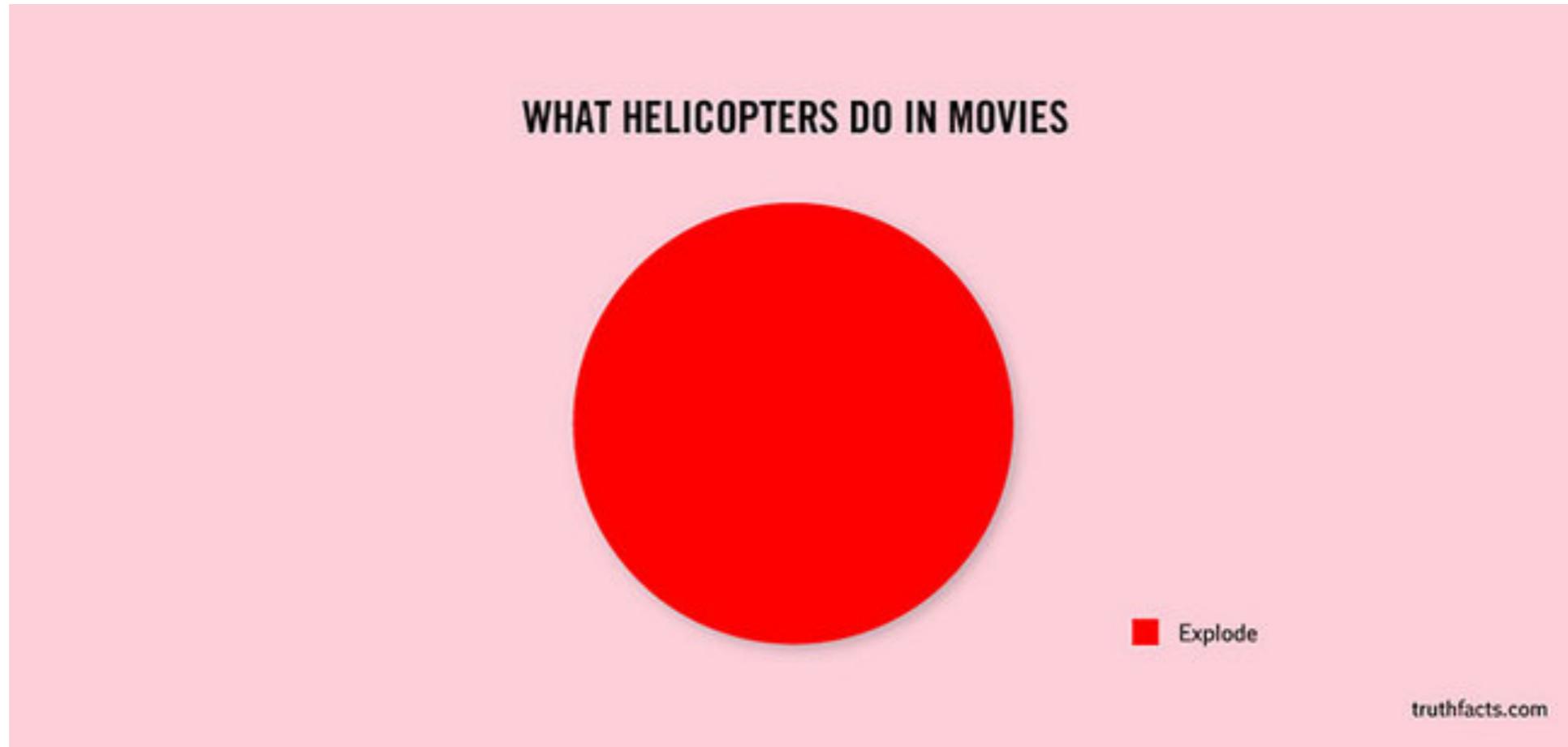
- So you have developers you can ask to do this for you...
- D3.js: <https://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/>
- Tableau: <https://public.tableau.com/en-us/s/blog/2015/07/analyzing-airbnb-data>
- Excel + PPT (for real!)
- Python (Bokeh, Seaborn, matplotlib)
- R (ggplot)

# CHARTJUNK

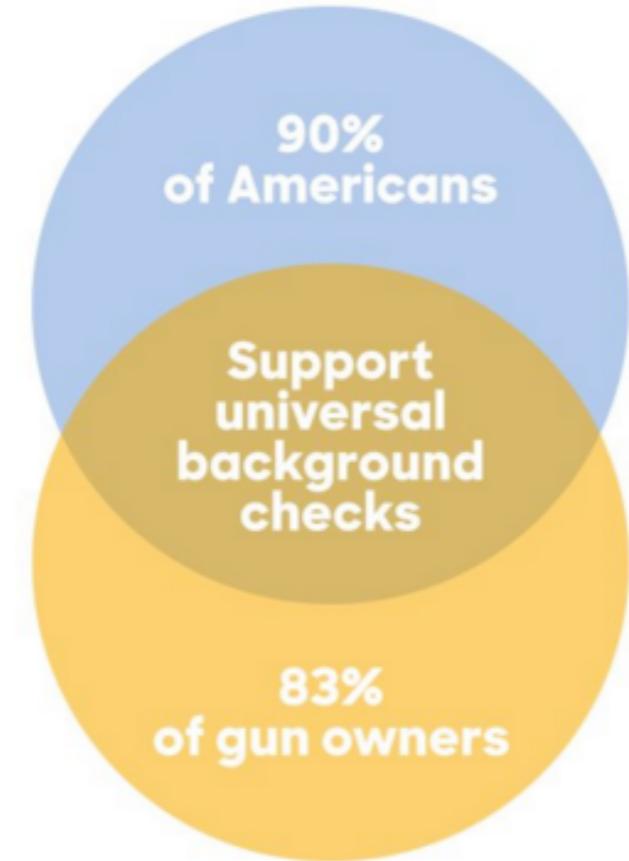
---







# GRAPHJUNK



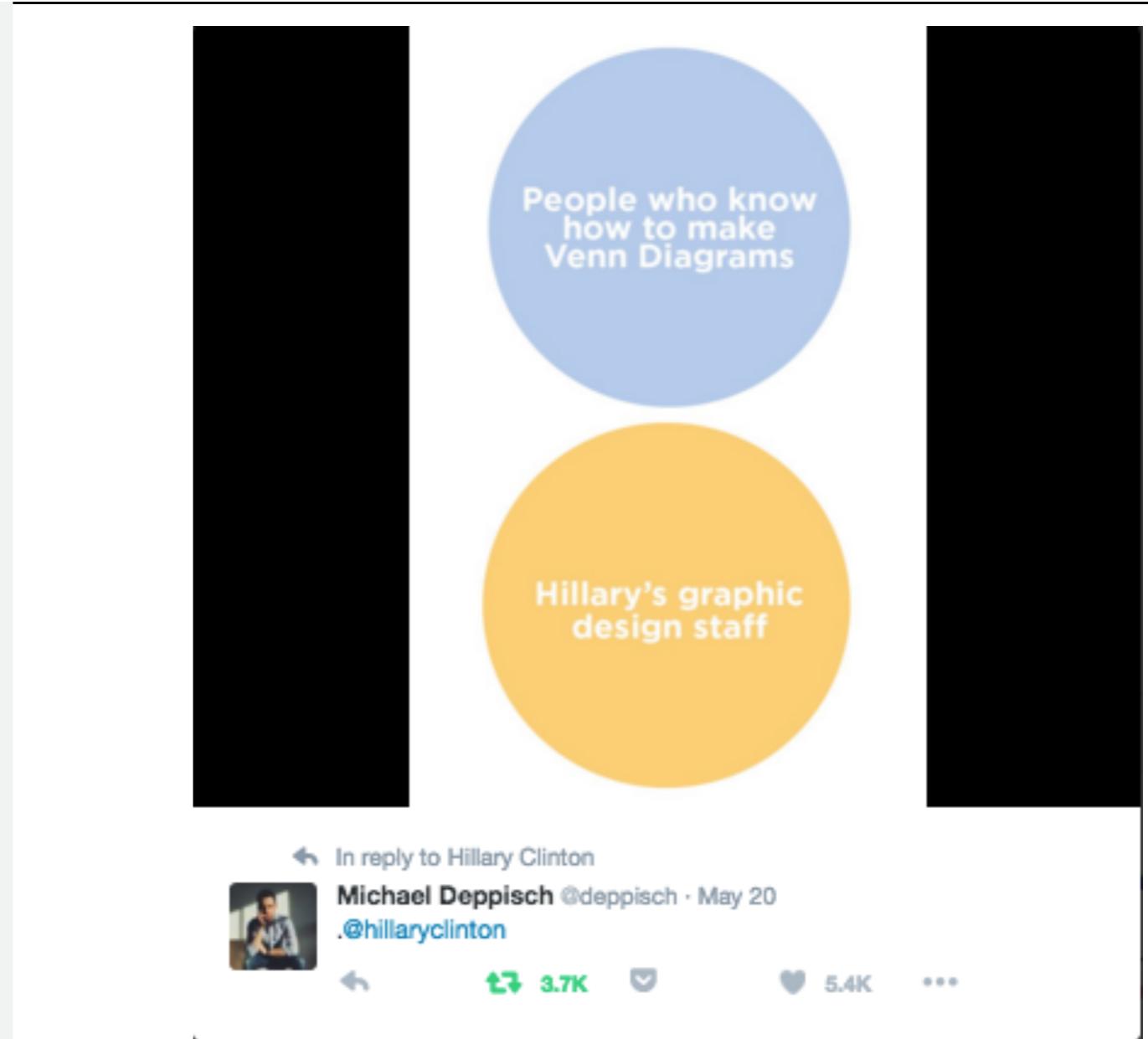
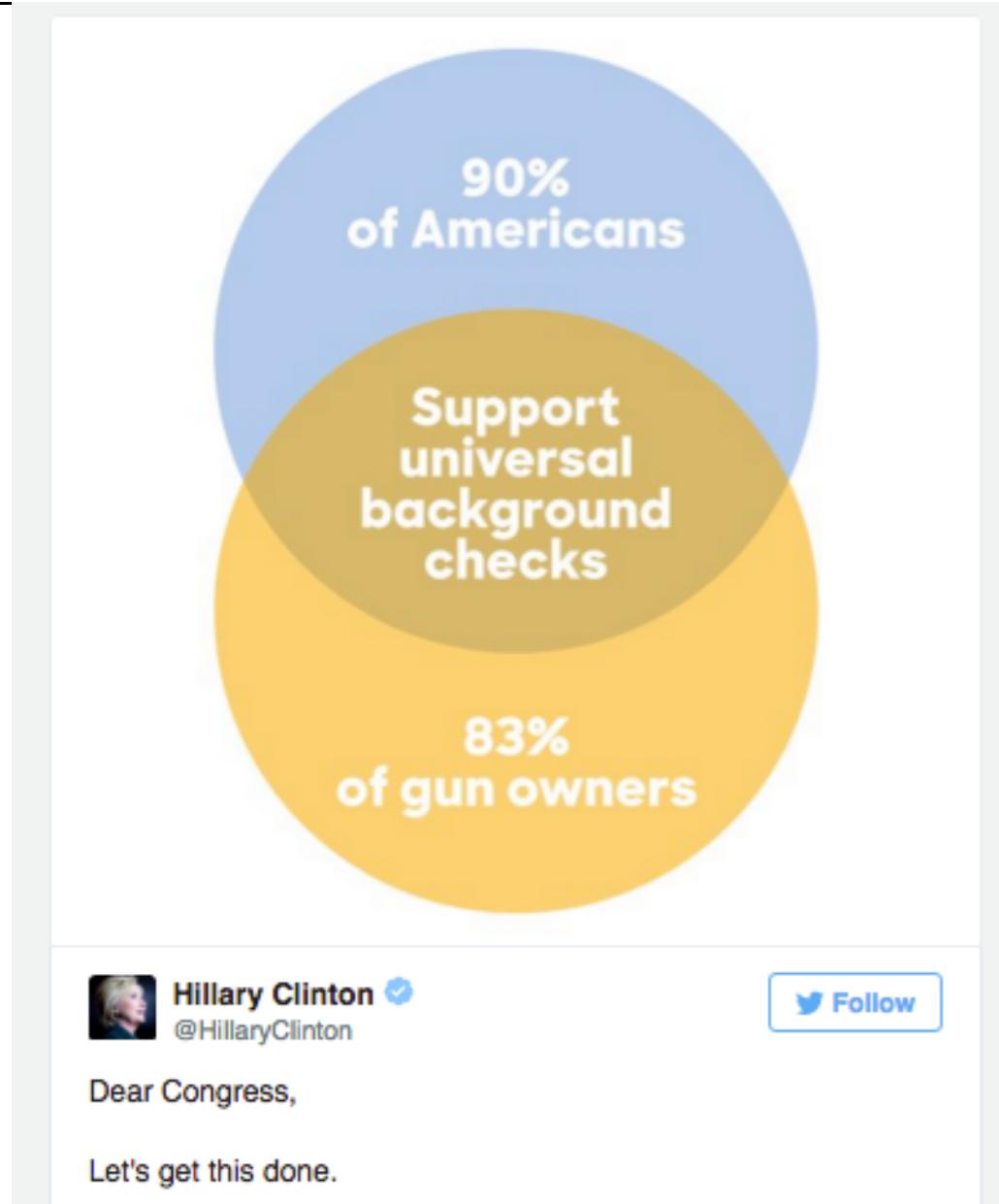
Hillary Clinton   
@HillaryClinton

 Follow

Dear Congress,

Let's get this done.

# GRAPHJUNK



# IT'S A BIPARTISAN PROBLEM

GOP  @GOP

Following

.@HillaryClinton's persistent lies have lost her the trust of the American people 

**VOTERS TRUST TRUMP OVER CLINTON**

— VOTERS SAY TRUMP IS MORE HONEST AND TRUSTWORTHY —

(CNN/ORC, 9/6/16)



**50% TRUMP**

**35% CLINTON**

THE WHITE  
WASHINGTON

RETWEETS LIKES

39 65



3:26 PM - 7 Sep 2016

# EFFECTIVE COMMUNICATION

# GET YOUR POINT ACROSS

---

## COMMUNICATION - CONTENT

---

- ▶ Establish a central thesis
- ▶ Create a narrative arc: problem, addressing, result
- ▶ “Completed X by doing Y as measured by Z”
- ▶ Follow the SAR principles: Situation, Action, Result

---

## COMMUNICATION - DELIVERY

---

- Establish a confident stance
- Engage your audience: fill the room and ensure your audience knows you’re speaking to every one of them
- Speak as though you’re writing: you begin with a thesis, topic sentence, and example
- Offer a “nugget” and reflect humility
- A pause is more powerful than “Um”

---

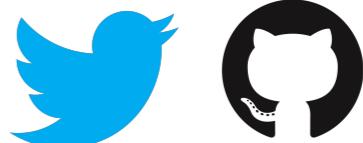
## NEXT STEPS

---

- Keep in touch!
- If you like data, checkout:
- Python for Data Analysis Workshop
- Intermediate Python Workshop
- Part-time Data Science
- Data Science Immersive



@josephofiowa



josephnelson@generalassemb.ly

## A CONVERSATION ON METRICS...

---

- A. We may want to optimize for quality hires.
- B. We may want to optimize for fast hiring.
- What metric may we use to capture A?
- What may that metric miss out on?

Aware company exists  
Visit company website  
Visits “Careers” page  
Begin filling application  
Submit application  
Hear from org to interview  
Interview 1  
Interview 2  
Offer  
Accept + Start  
Refer others to apply

