

## Topic 5. Visualization and Univariate Graphics

Prof. Josh Clinton

# Today's Agenda

- ▶ Introduction to data visualization!
- ▶ Principles of Data Visualization

## Communicating & Visualizing Data is Essential

- ▶ Data does not exist in a vacuum – it is always interpreted in relationship to something.
- ▶ Humans infer causality from relationships (much too quickly!).
- ▶ All data-science should be question driven! What is the question you are answering with the visualization you are providing?
- ▶ Always think: what is the answer that your visualization is providing?
- ▶ Goal is accuracy, not visual appeal! Science, not Art!
- ▶ Does your visualization communicate the relationship cleanly and accurately?

January 28, 1986: 29 Celcius



## You are NASA (or Thokiol)...

- ▶ What is the question you are asking?
- ▶ What is the information you need to make that determination?

# Do You Launch? Pre-Launch Analysis of Damage

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

SRM No.	Cross Sectional View			Top View			Clocking Location (deg)
	Erosion Depth (in.)	Perimeter Affected (deg)	Nominal Dia. (in.)	Length Of Max Erosion (in.)	Total Heat Affected Length (in.)		
22A	None	None	0.280	None	None	36°--66°	
22A	NONE	NONE	0.280	NONE	NONE	338°-18°	
15A	0.010	154.0	0.280	4.25	5.25	163	
15B	0.038	130.0	0.280	12.50	58.75	354	
15B	None	45.0	0.280	None	29.50	354	
41D	RH Forward Field	0.028	110.0	0.280	3.00	None	275
41C	LH Aft Field*	11A	None	0.280	None	None	--
41B	LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50
STS-2	RH Aft Field	28	0.053	116.0	0.280	--	351
						90	

\*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.

\*\*Soot behind primary O-ring.

\*\*\*Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

# Numbers are Only Meaningful in terms of Relationships

## BLOW BY HISTORY

### SRM-15 WORST BLOW-BY

- 2 CASE JOINTS ( $80^\circ$ ), ( $110^\circ$ ) Arc
- MUCH WORSE VISUALLY THAN SRM-22

### SRM 22 BLOW-BY

- 2 CASE JOINTS ( $30-40^\circ$ )

### SRM-13A, 15, 16A, 18, 23A 24A

- NOZZLE Blow-by

## HISTORY OF O-RING TEMPERATURES (DEGREES - F)

MOTOR	MBT	AMB	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

73 seconds after launch...

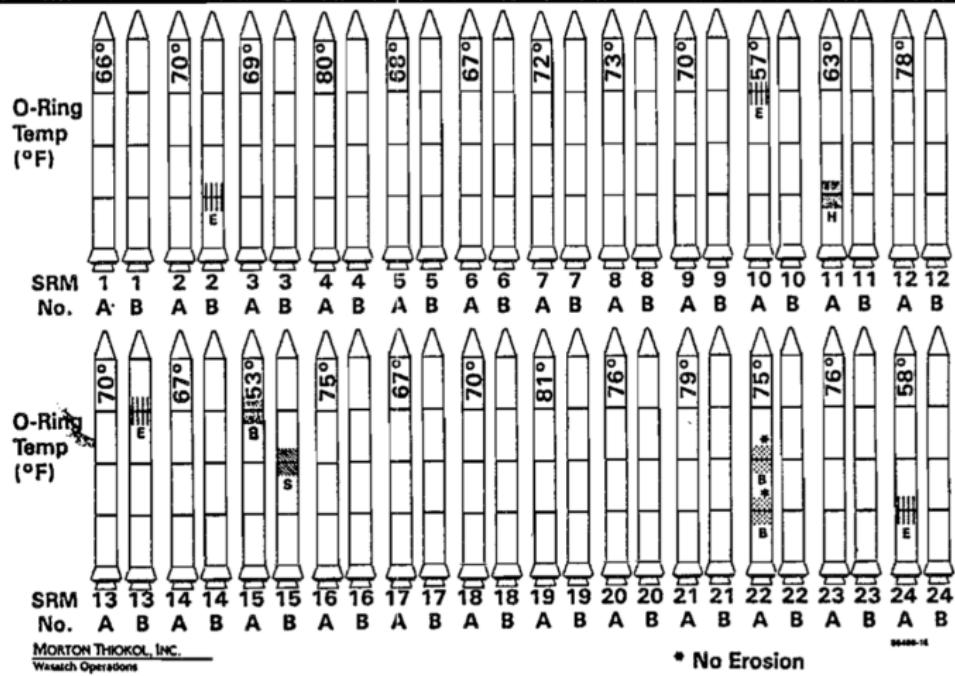


## Aftermath: Feynman



# Post Launch: Official Analysis

## History of O-Ring Damage in Field Joints (Cont)



INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION  
AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

## What We Can Do Instead

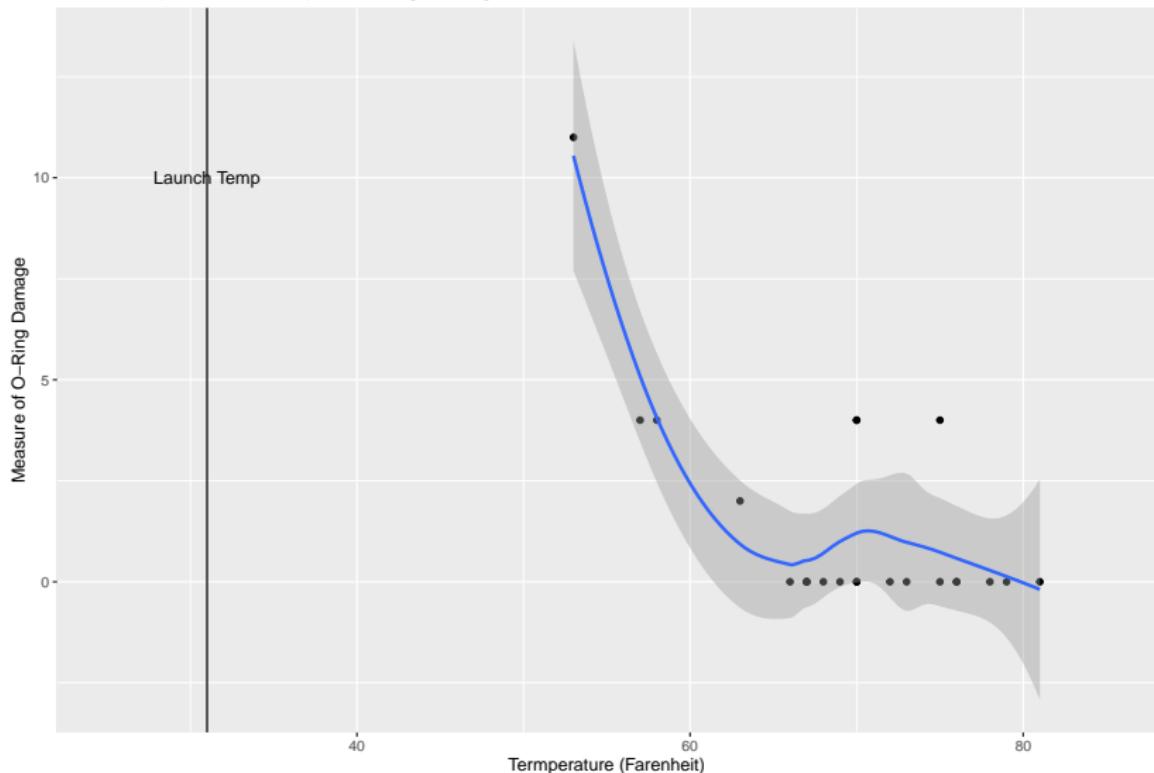
```
library(tidyverse)
library(alr4)
data(Challeng)

Betterplot <- Challeng %>%
  ggplot(aes(x=temp,y=damage)) +
  geom_point() +
  labs(title = "Relationship Between Temp. & O-Ring Damage",
       x="Termperature (Farenheit)",
       y="Measure of O-Ring Damage") +
  scale_x_continuous(limits=c(25,85)) +
  geom_vline(xintercept=31) +
  annotate("text",x = 31, y=10, label = "Launch Temp") +
  geom_smooth(method = "loess")
```

# What We Can Do Instead

## Betterplot

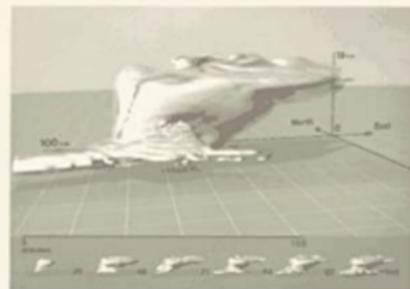
Relationship Between Temp. & O-Ring Damage



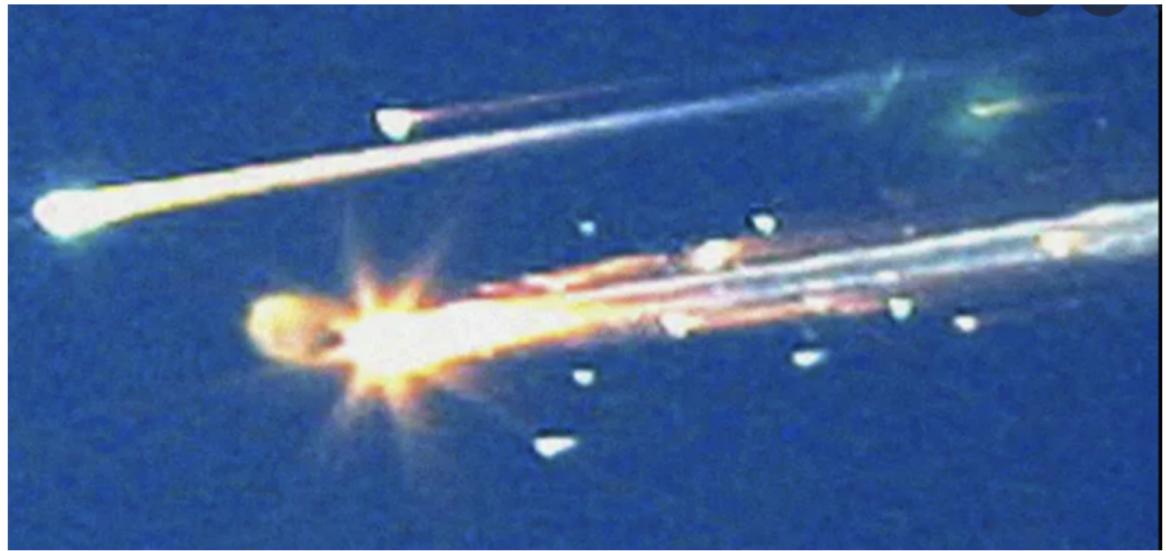
More Info?



EDWARD R. TUFTE  
VISUAL EXPLANATIONS



Unfortuately... in 2003



## Unfortuately... in 2003

### Review of Test Data Indicates Conservatism for Tile Penetration

- The existing SOFI on tile test data used to create Crater was reviewed along with STS-87 Southwest Research data
  - Crater overpredicted penetration of tile coating significantly
    - Initial penetration to be described by normal velocity
      - Varies with volume/mass of projectile (e.g. 200ft/sec for 3cu. ln)
    - Significant energy is required for the softer SOFI particle to penetrate the relatively hard tile coating
      - Test results do show that it is possible at sufficient mass and velocity
  - Conversely, once tile is penetrated SOFI can cause significant damage
    - Minor variations in total energy (above penetration level) can cause significant tile damage
- Flight condition is significantly outside of test database
  - Volume of ramp is 1920cu in vs 3 cu in for test

## GOAL: Your visuals must tell an accurate story

- ▶ Humans over-interpret relationships as causal. You must be informative without being misleading.
- ▶ Tables and graphs are essential.
- ▶ Visualizations must be stand-alone as much as possible.

NOTE: Rule of thumb: show it to someone without explanation. If they are confused, re-do!

# Dimensions of Visualization

Essential elements of every good visualization:

- ▶ Well-labeled axes (no variable names!).
- ▶ Title describing the data being plotted.
- ▶ Meaningful scale for x and y axes.
- ▶ Appropriate visualization for the type of data being characterized.
- ▶ ADVANCED: Uncertainty (if relevant)

# Dimensions of Visualization

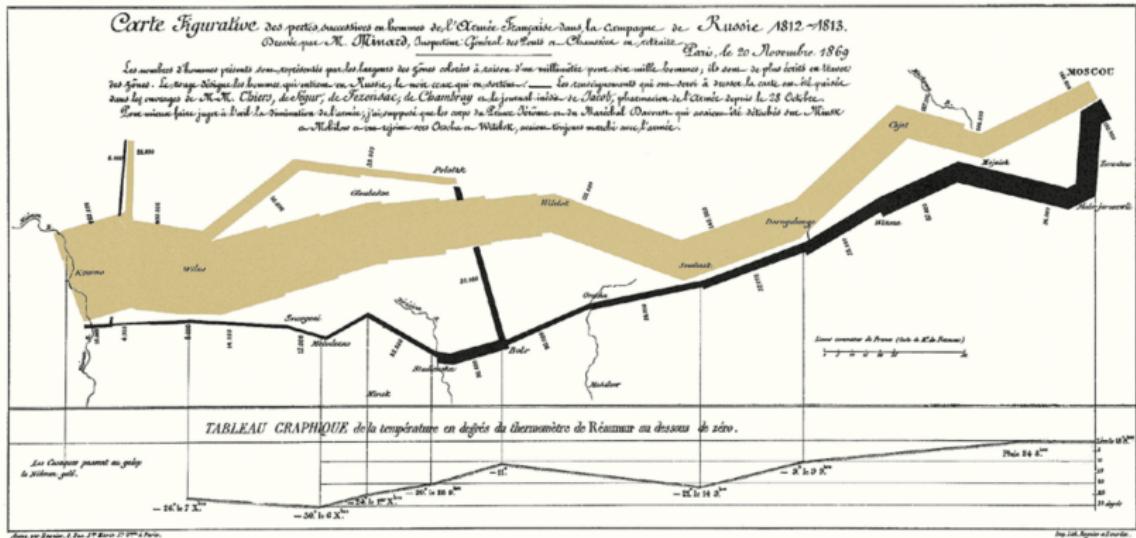
You have several “dimensions” to use when presenting information

- ▶ Horizontal dimension (x-axis)
- ▶ Vertical dimension (y-axis)
- ▶ Shape of data points
- ▶ Color
- ▶ Size of data points

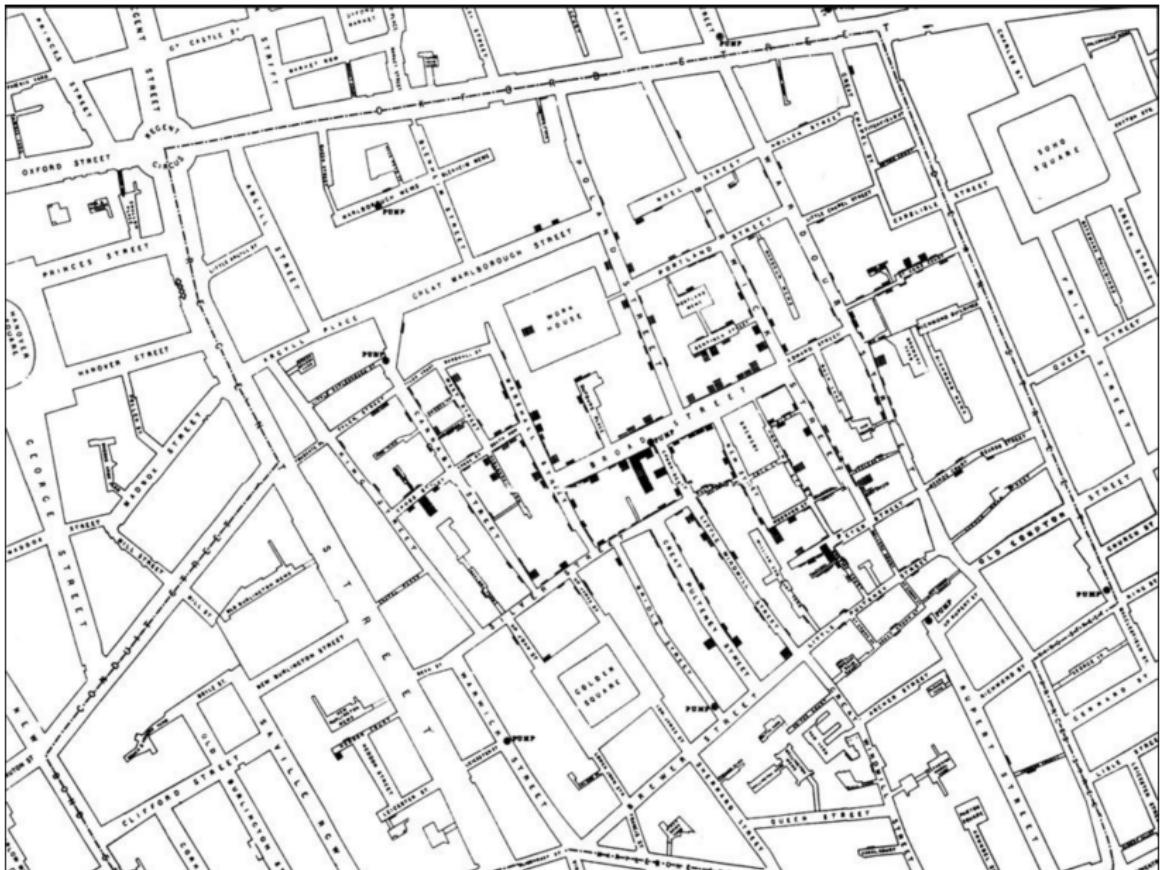
## Don't be confusing

- ▶ Map each variable to at most one dimension.
- ▶ Be intuitive – don't assign small numbers large dots, etc.
- ▶ Don't include extraneous information/visual elements.

# Best Graph Ever? Charles Joseph Minard

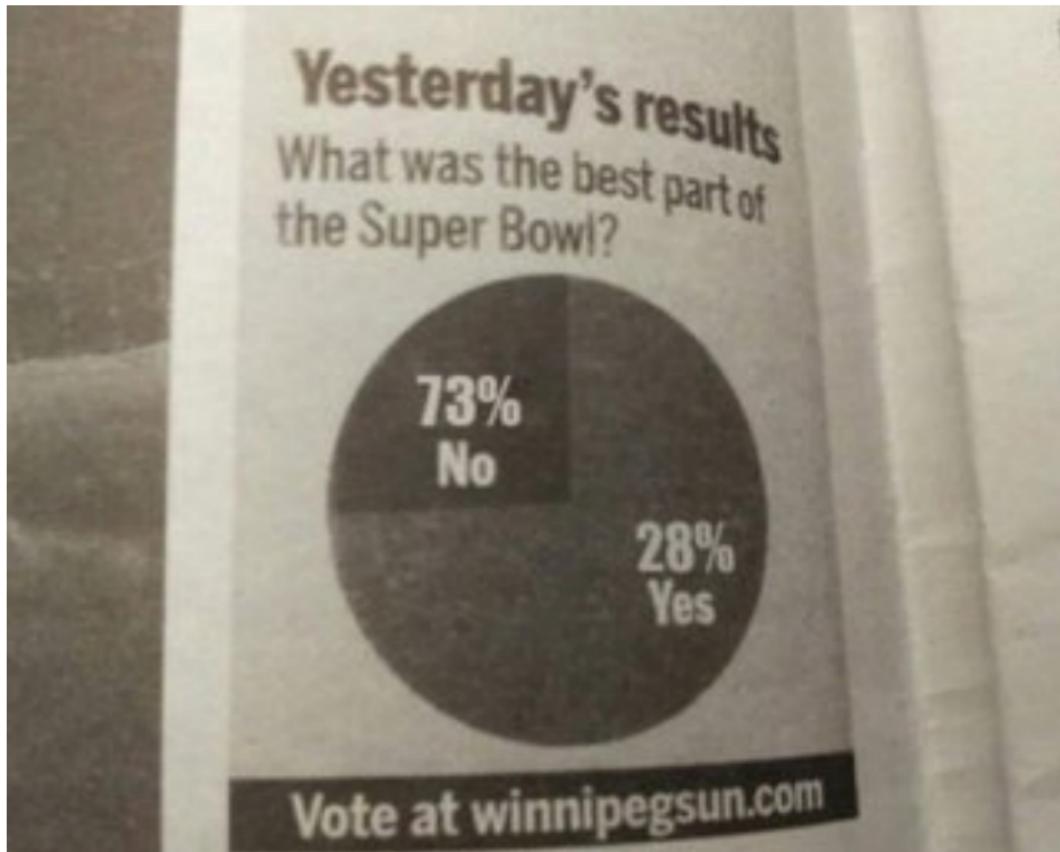


Or maybe John Snow? (Not Jon Snow!)



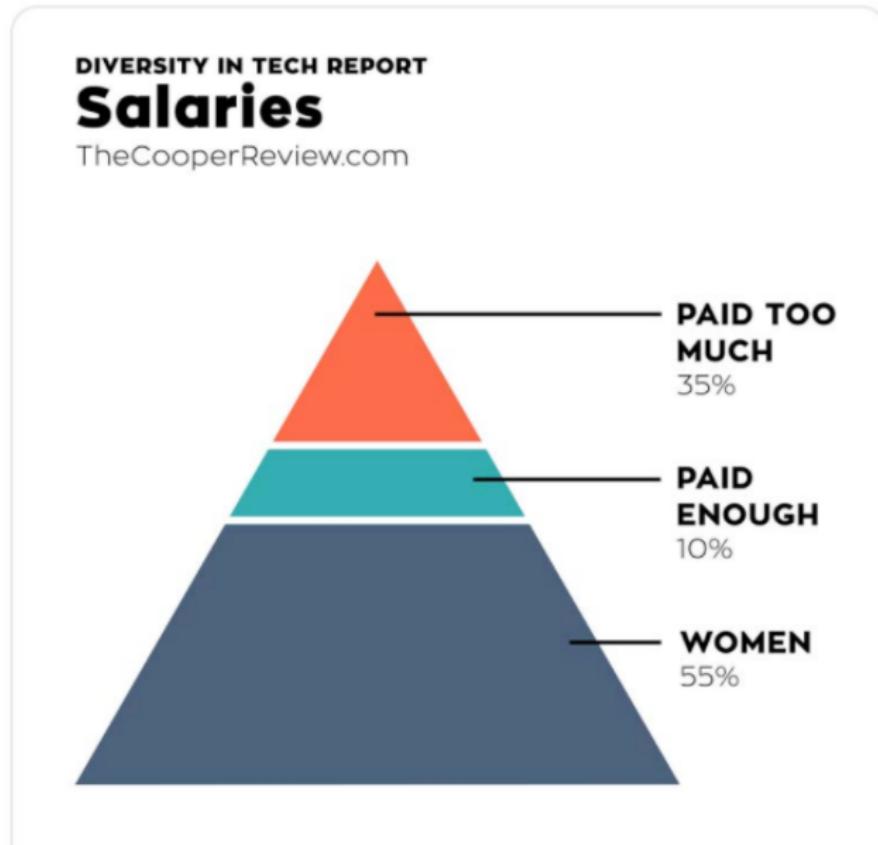
## Your graphs should make sense!

- The visualization should clearly answer the question.



## Your graphs should make sense!

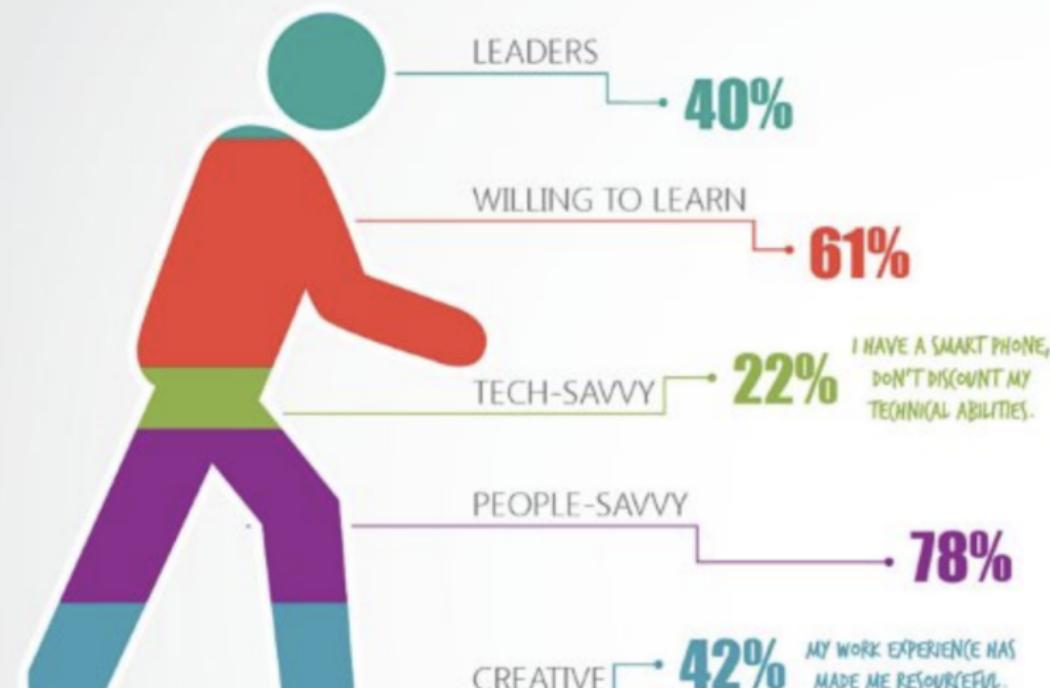
- The visualization should clearly answer the question.



## Your graphs should make sense!

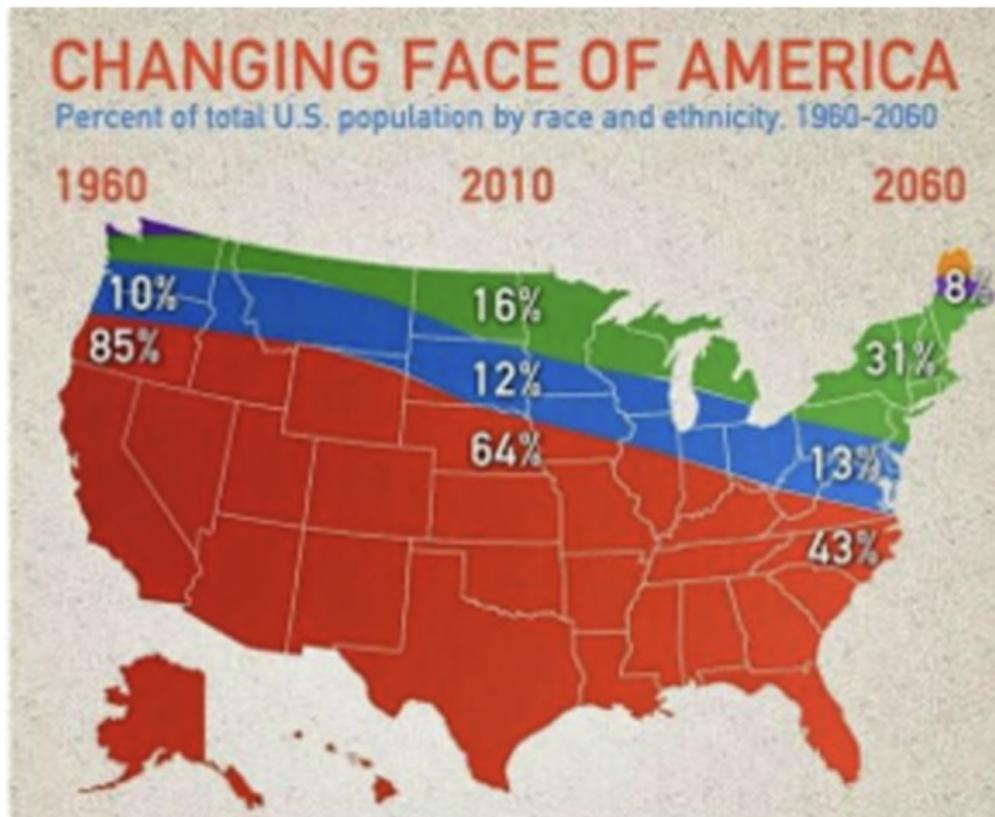
- The visualization itself should not be the focus of the visualization.  
The focus should be on the relationship being visualized.

### HOW BABY BOOMERS DESCRIBE THEMSELVES



## Your graphs should make sense!

- The visualization should be complete (present all the information) but not include extraneous/distracting information/visual-elements.



# Your graphs should be accurate and not mislead

- ▶ Numbers (and patterns in numbers) should match the visualization



Arieih Kovler

@ariehkovler

Follow

"Is 34 bigger or smaller than 14?"

"Smaller. Definitely smaller"

"What about zero?"

"Zero's a bit less than 34 but it's much more than 22"



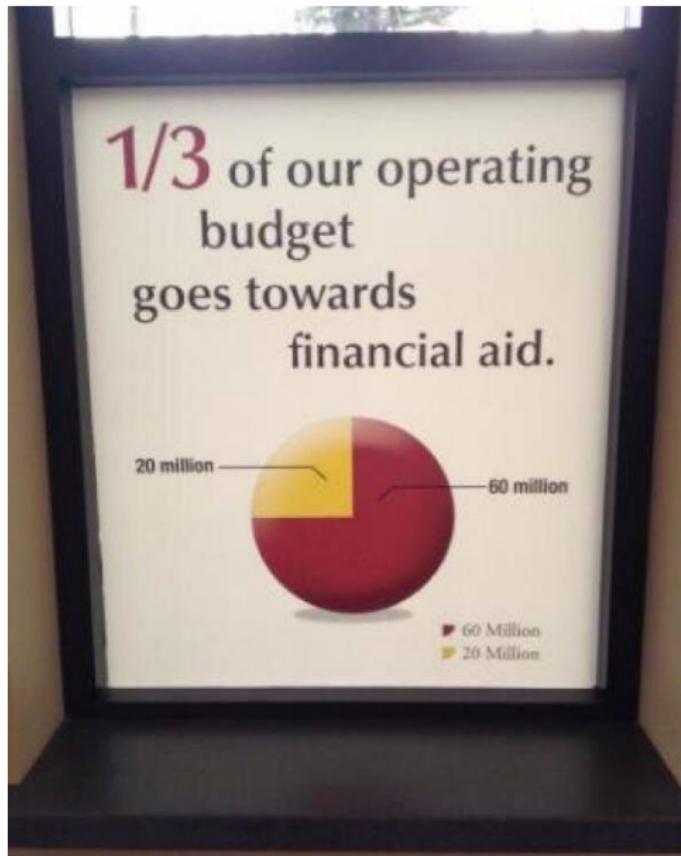
Your graphs should be accurate and not mislead

- ▶ Numbers (and patterns in numbers) should match the visualization



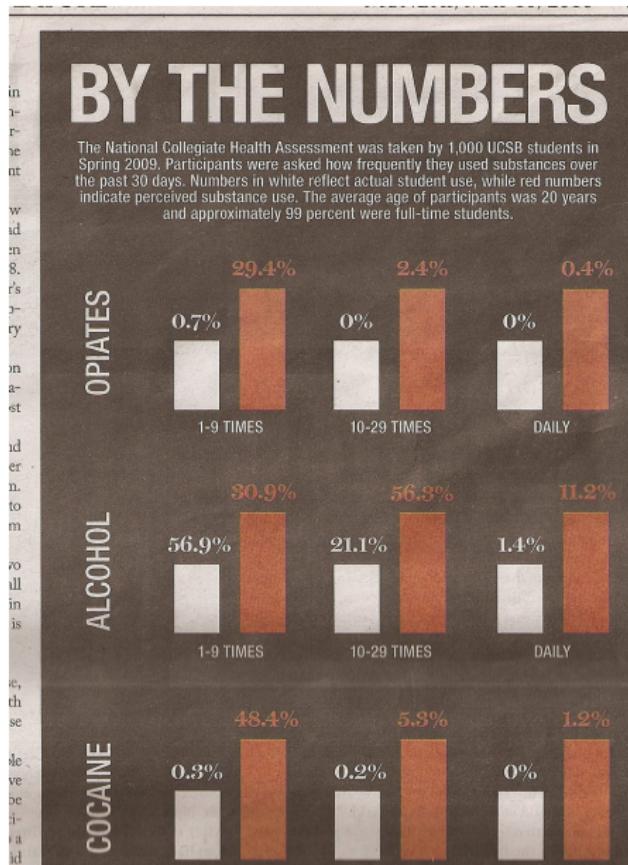
Your graphs should be accurate and not mislead

- ▶ Numbers (and patterns in numbers) should match the visualization



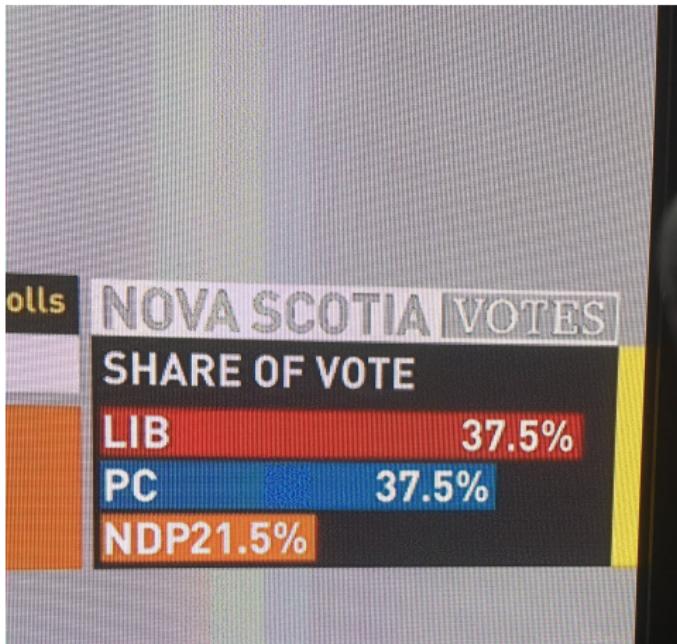
# Your graphs should be accurate and not mislead

- ▶ Numbers (and patterns in numbers) should match the visualization



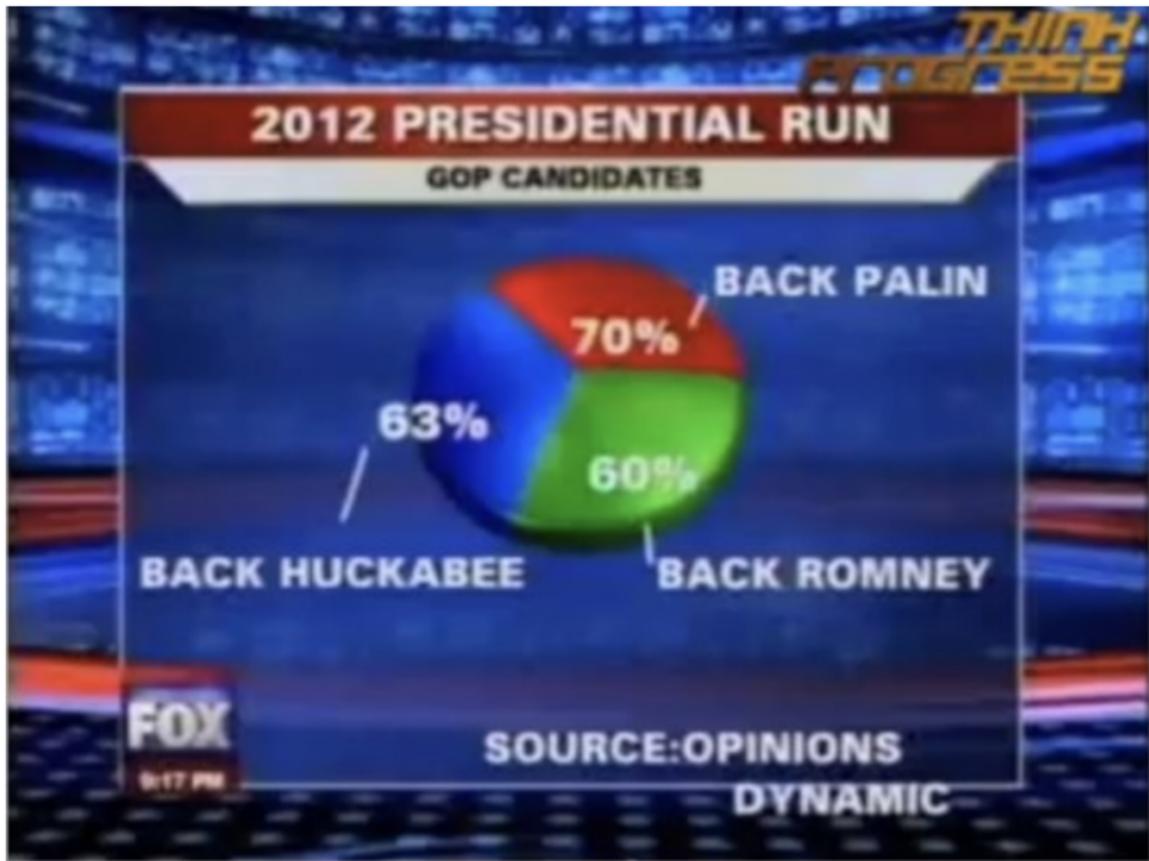
## Your graphs should be accurate and not mislead

- ▶ Numbers (and patterns in numbers) should match the visualization



Your graphs should be accurate and not mislead

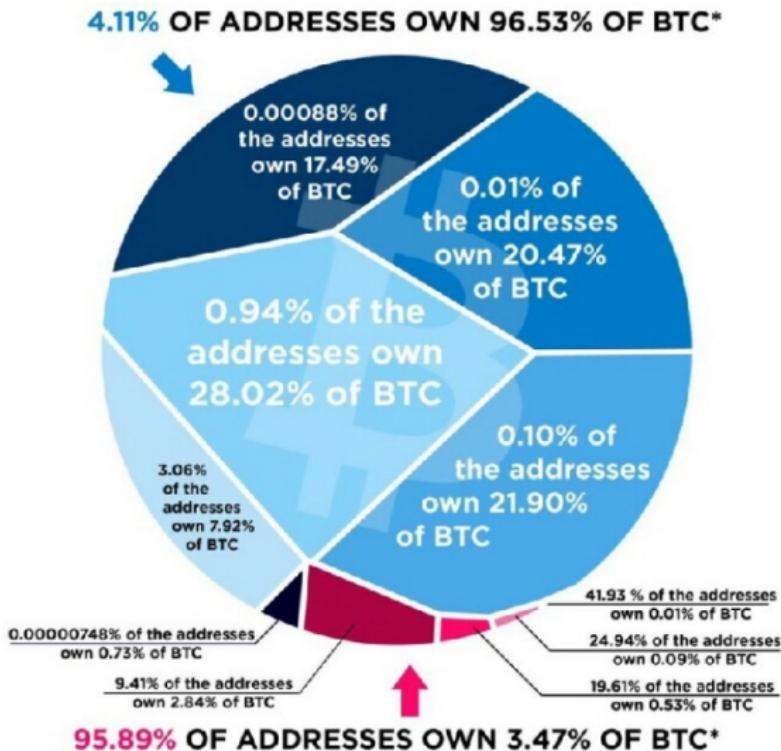
- Avoid piecharts, but if you use them make sure they sum to 100.



Your graphs should be accurate and not mislead

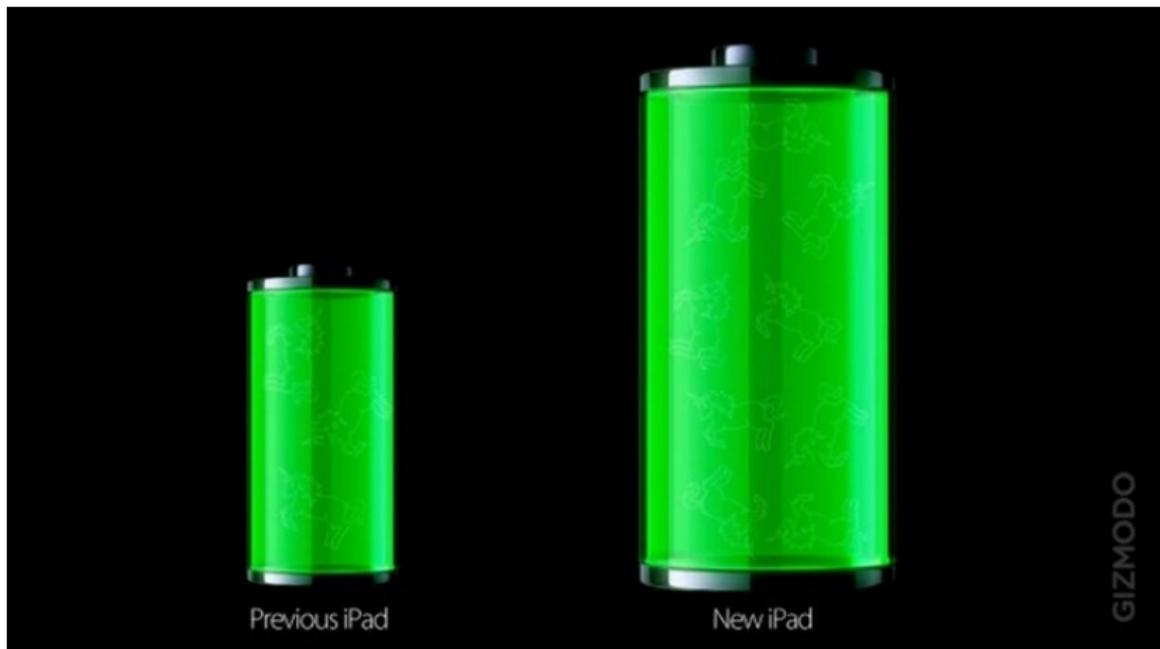
- Avoid piecharts, but if you use them make sure they sum to 100.

## The Bitcoin Wealth Distribution



## Your graphs should be accurate and not mislead

- ▶ Variation in a variable should vary in one-visual dimension, not two!  
(Piecharts bad)

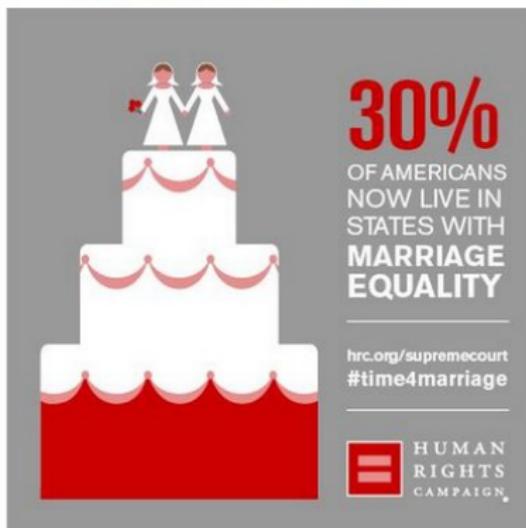


## Your graphs should be accurate and not mislead

- Variation in a variable should vary in one-visual dimension, not two!  
(Piecharts bad)



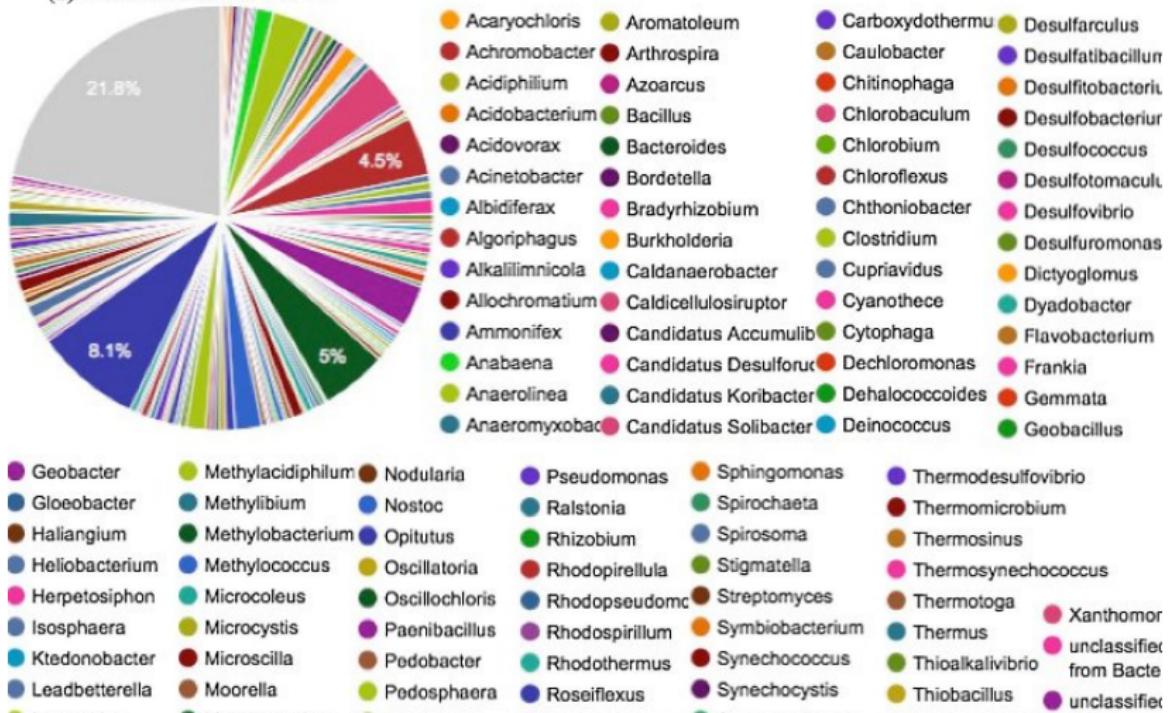
Reply Retweet Favorite More



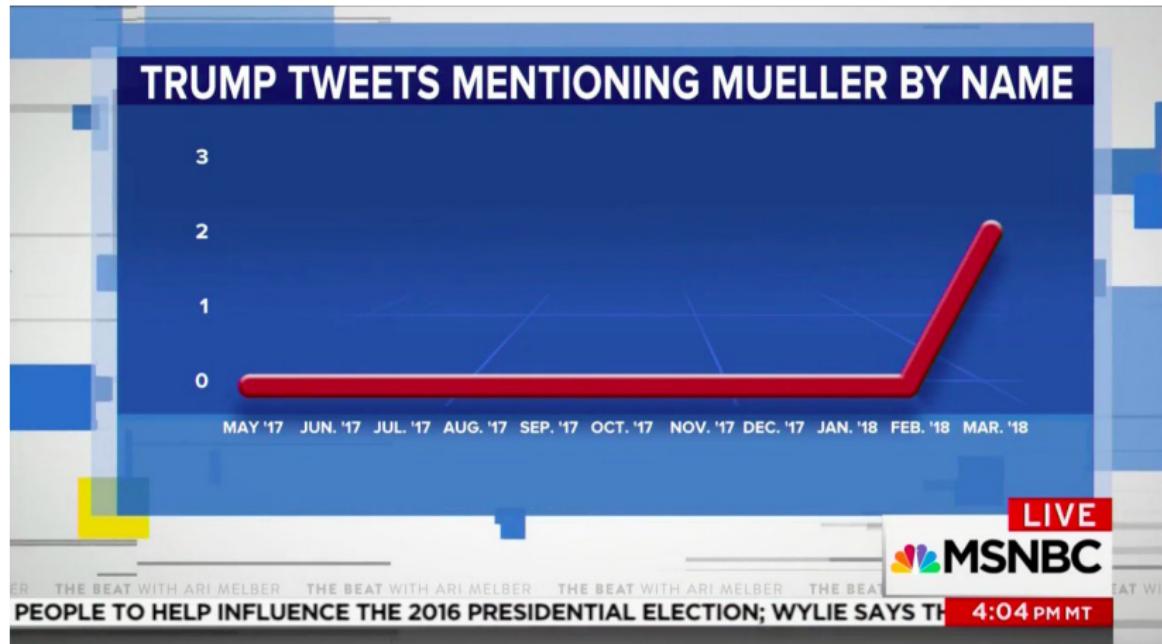
# But, don't put too much info on a graph!

- Think about the central point you are trying to convey!
- A visualization with too much data (and not enough variation in that data?) is meaningless. . .

(f) Distribution of Genus



Don't use a graph if you don't need a graph!



# Don't use a graph if you don't need a graph!

---

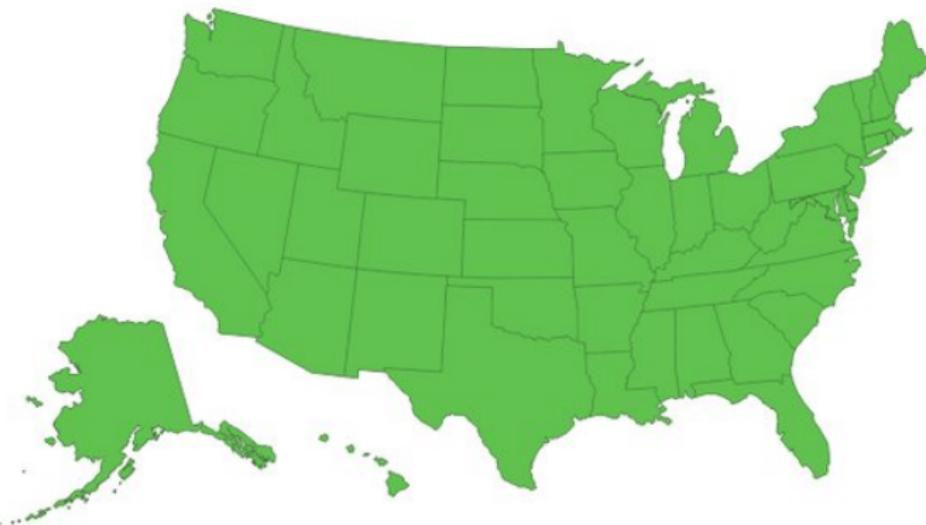
LGBT, 2012

---

State:

% LGBT: 3.5

---



---

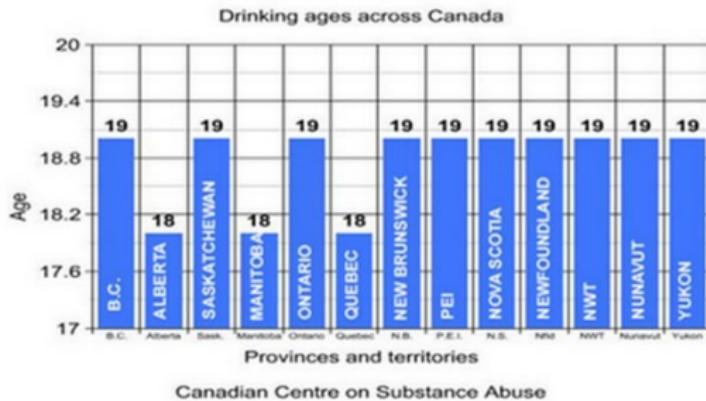
Share

GALLUP®

# Don't use a graph if you don't need a graph!

## Drinking age will remain 19 in Saskatchewan

CBC News Posted: Mar 4, 2013 11:59 AM CST | Last Updated: Mar 4, 2013 11:55 AM CST □ 25



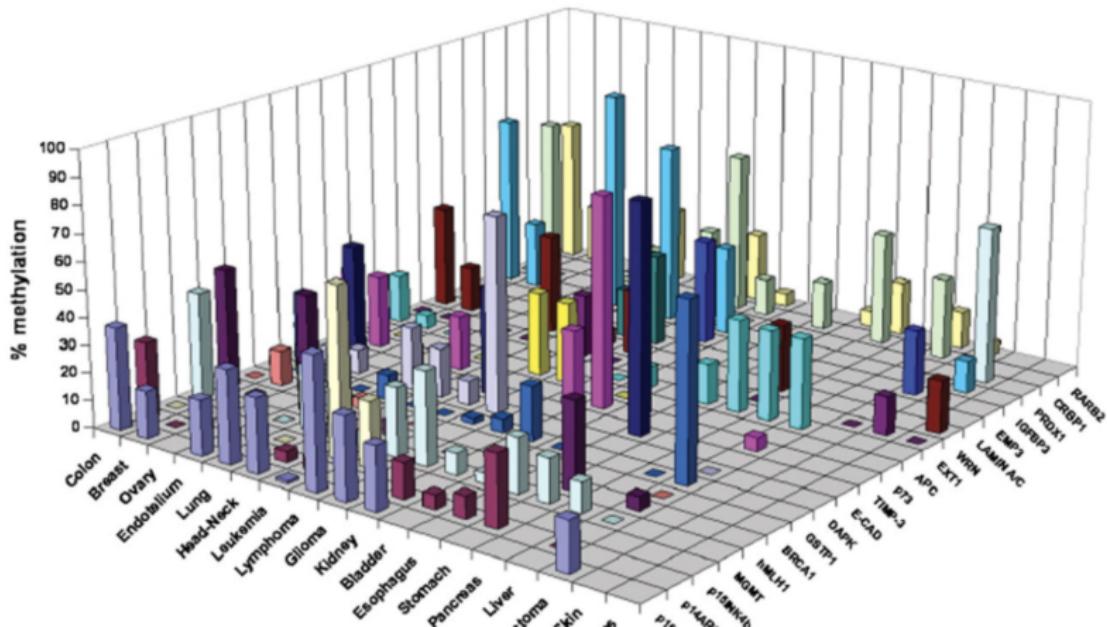
You have to be 19 in Saskatchewan to have a drink, while in Alberta and Manitoba, the drinking age 18. (CBC)

The Saskatchewan Party government has ruled out lowering the drinking age, four months after party members put the issue in the public eye.

Barplots are good, but not 3d barplots!

- ▶ Does the visualization reveal all of the variation of interest?
  - ▶ Depth is a dimension that can show variation, but it can be hard to interpret!

# A CpG Island Hypermethylation Profile of Human Cancer

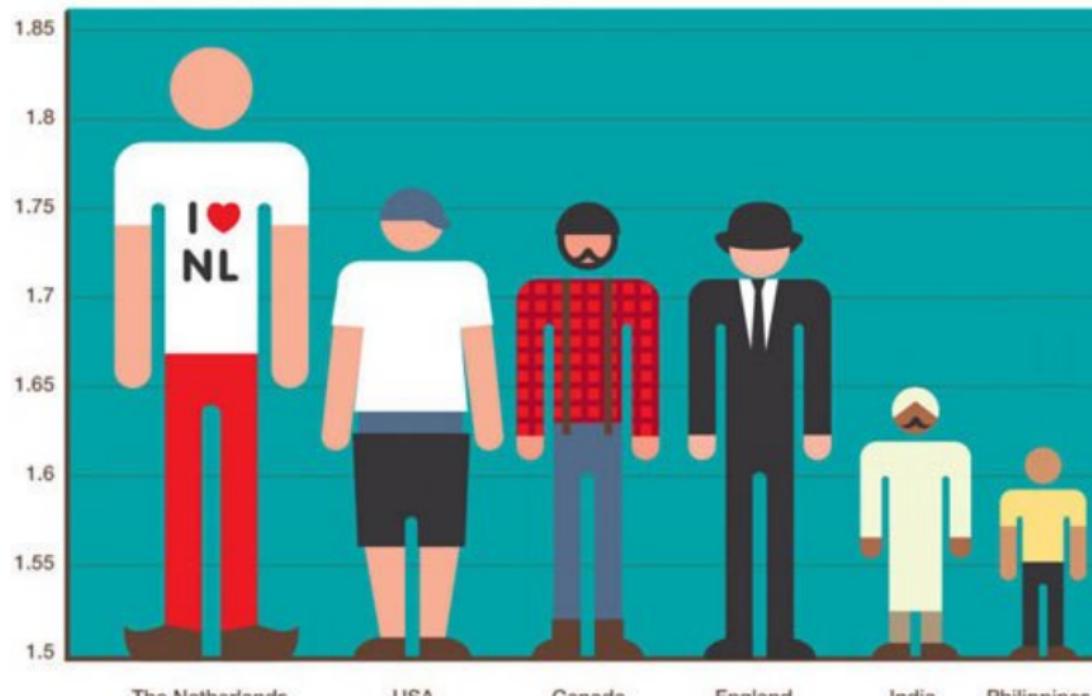


# Don't use deceptive y-axes...

- We don't always need to include 0, unless not doing so is confusing.

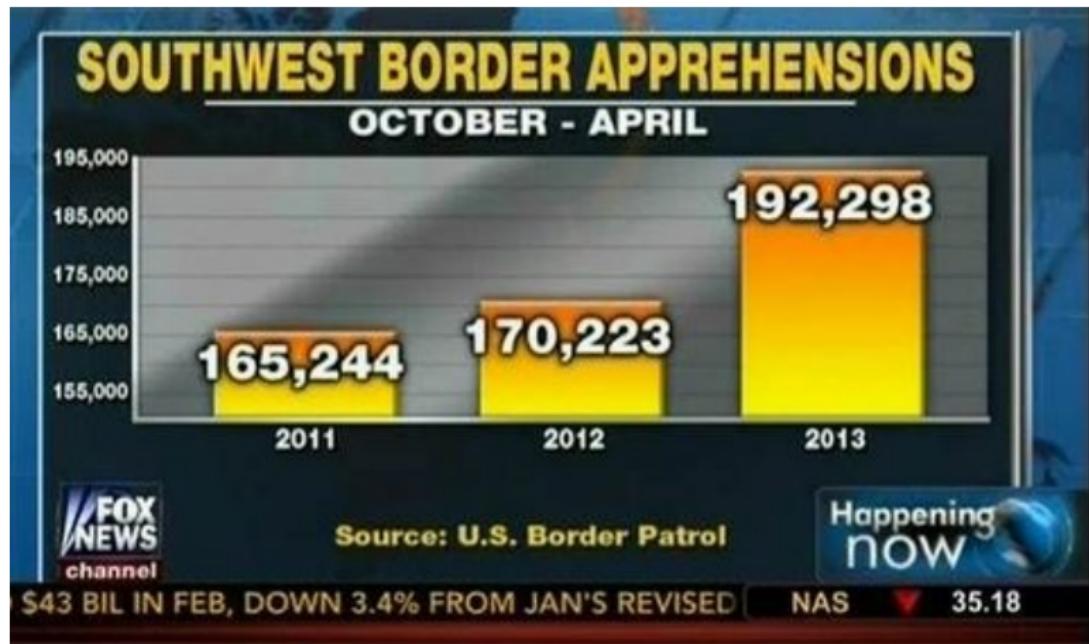
## LOOKING DOWN ON THE REST OF THE WORLD

(Average male height in m)



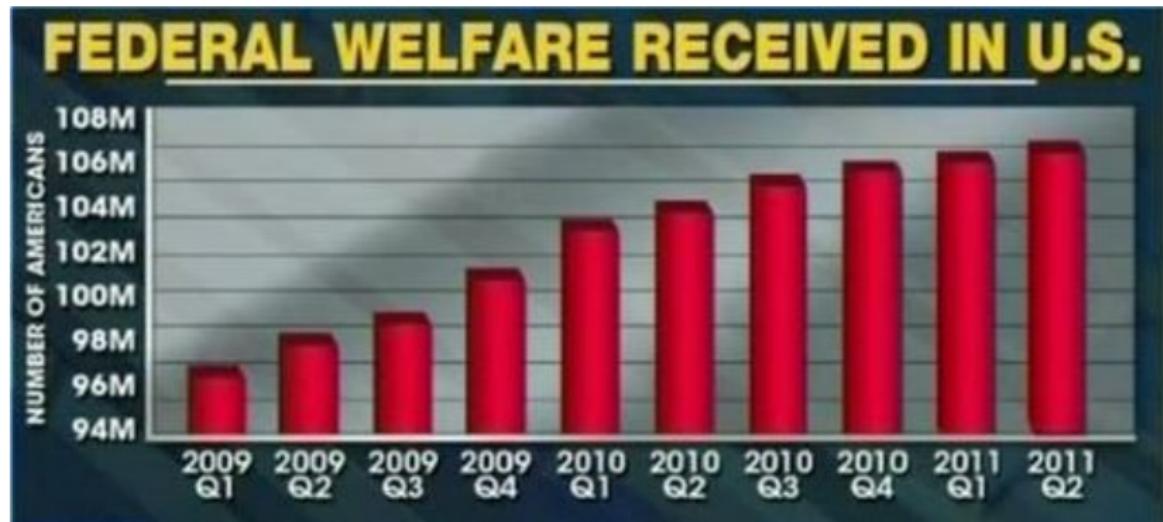
Don't use deceptive y-axes. . .

- ▶ Axes can make small changes look big and big changes look small.

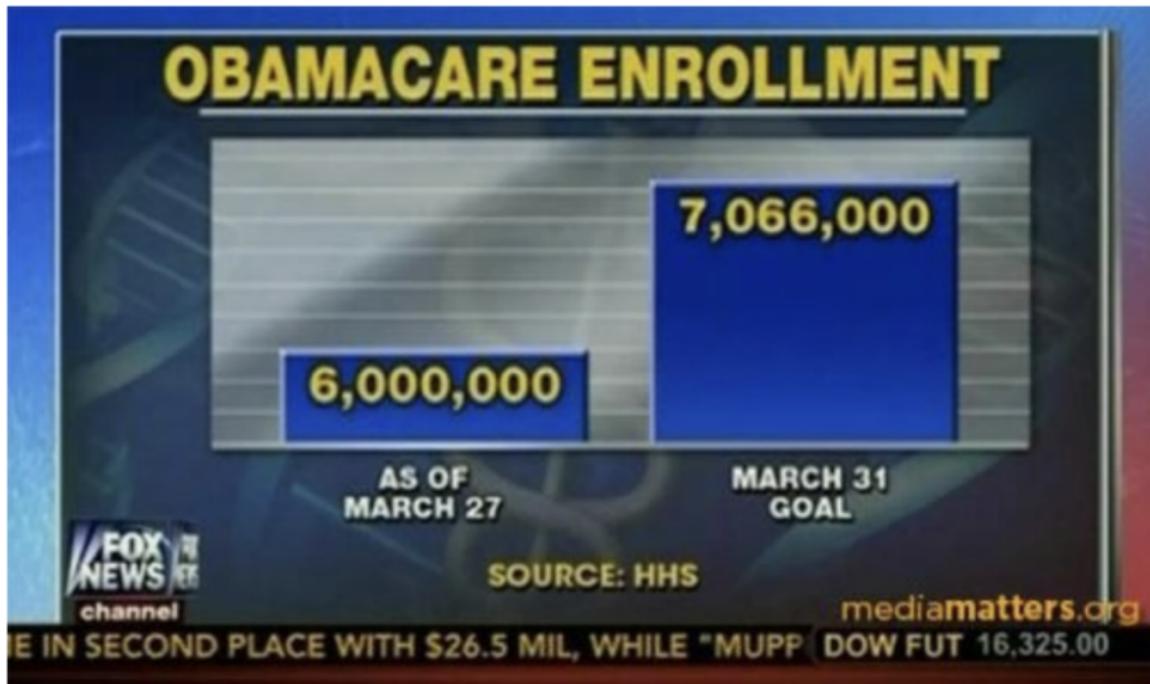


## Don't use deceptive y-axes...

- ▶ The range can also change how you interpret the data in relationship to suggested relationships.



... But please do use axes!



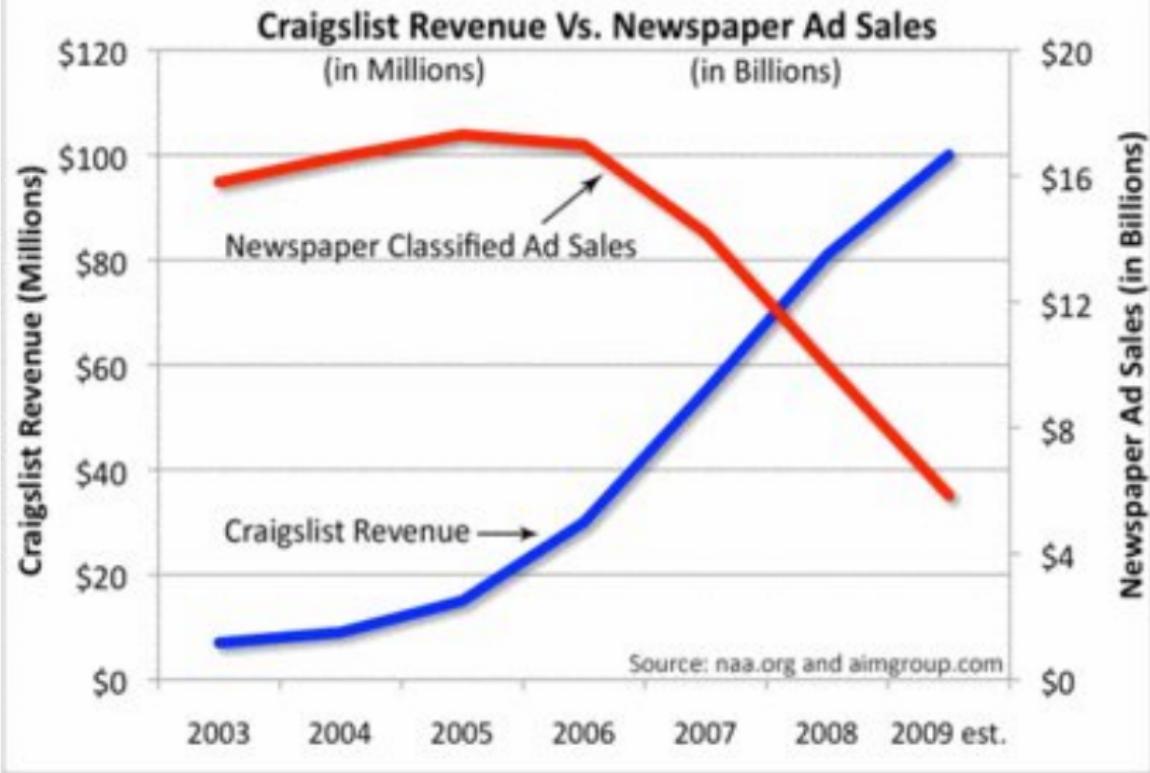
... But please do use axes!



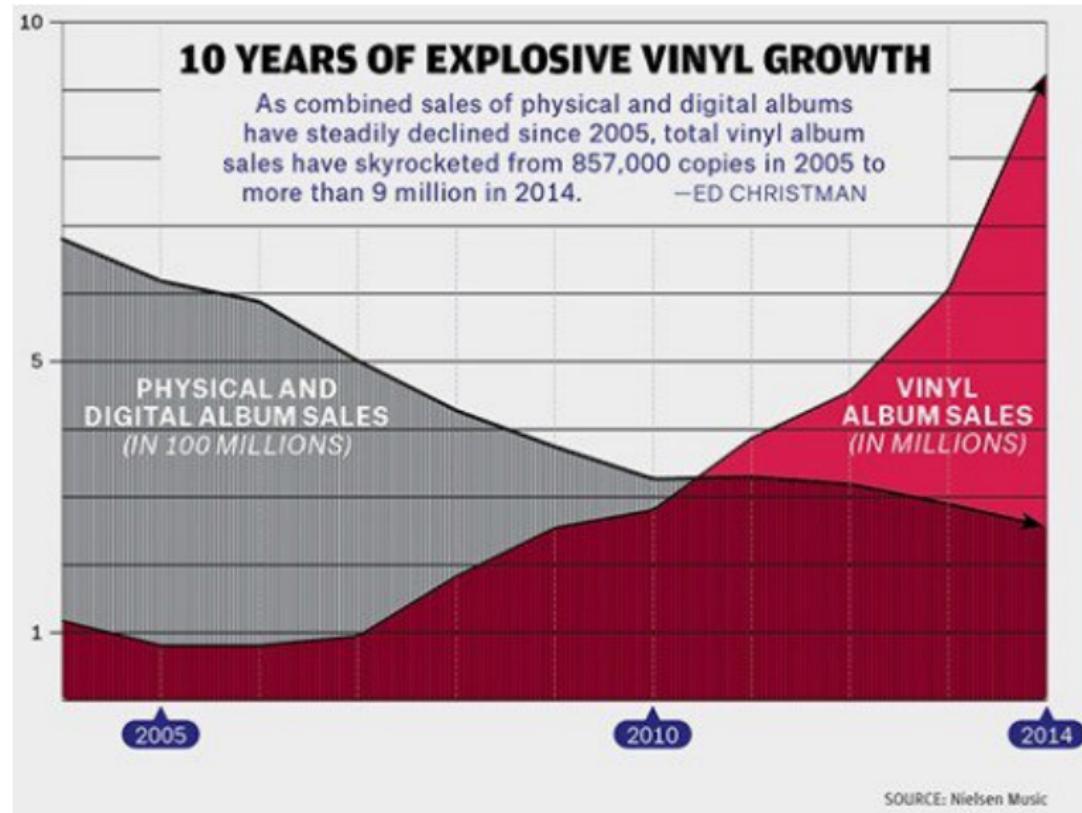
But not multiple y-axes!

Silicon Alley Insider

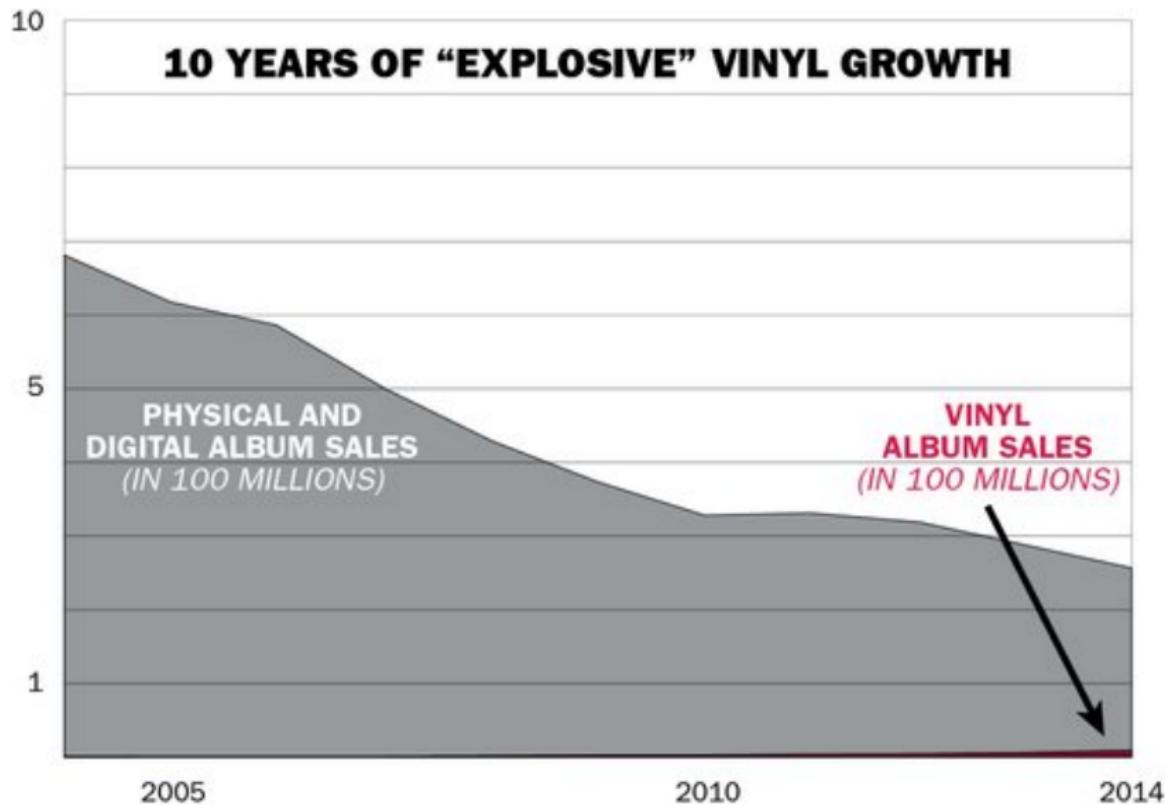
Chart of the Day



But not multiple y-axes!



But not multiple y-axes!



## But not multiple y-axes!



But not multiple y-axes!



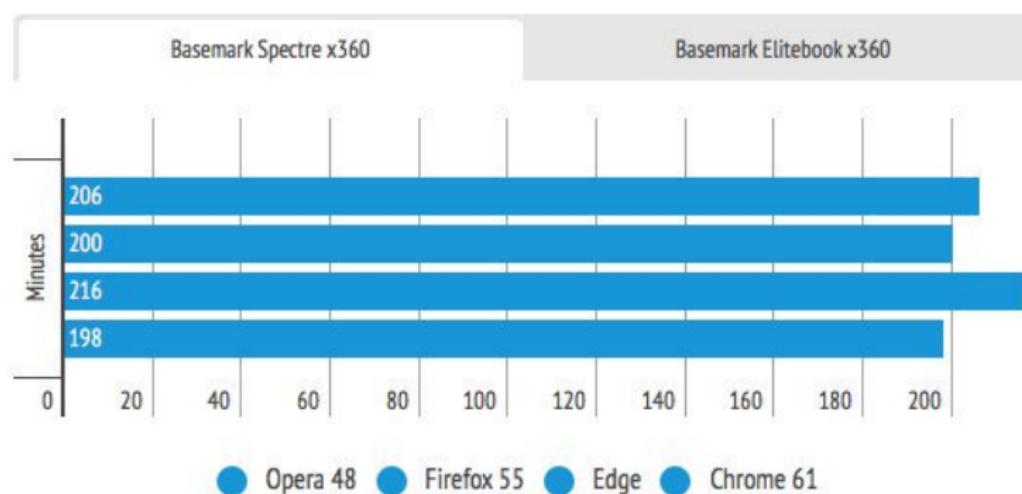
Only use colors if meaningful



Only use colors if meaningful

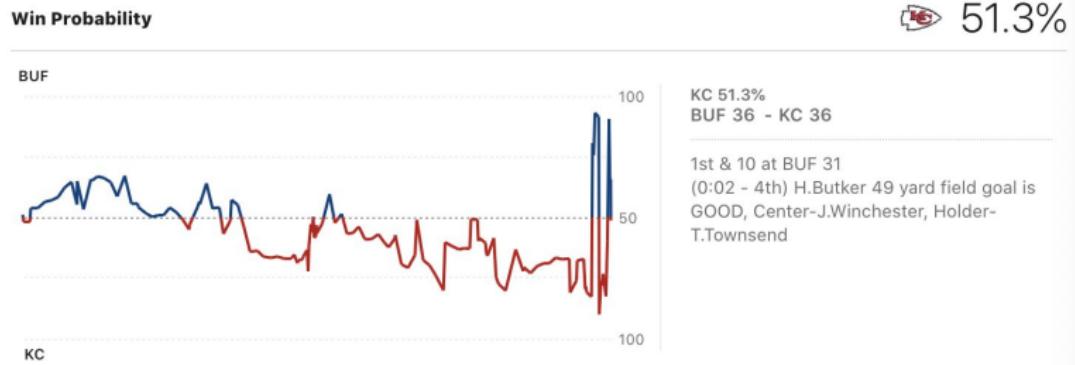
## Battery Performance

Runtime



# Uncertainty?

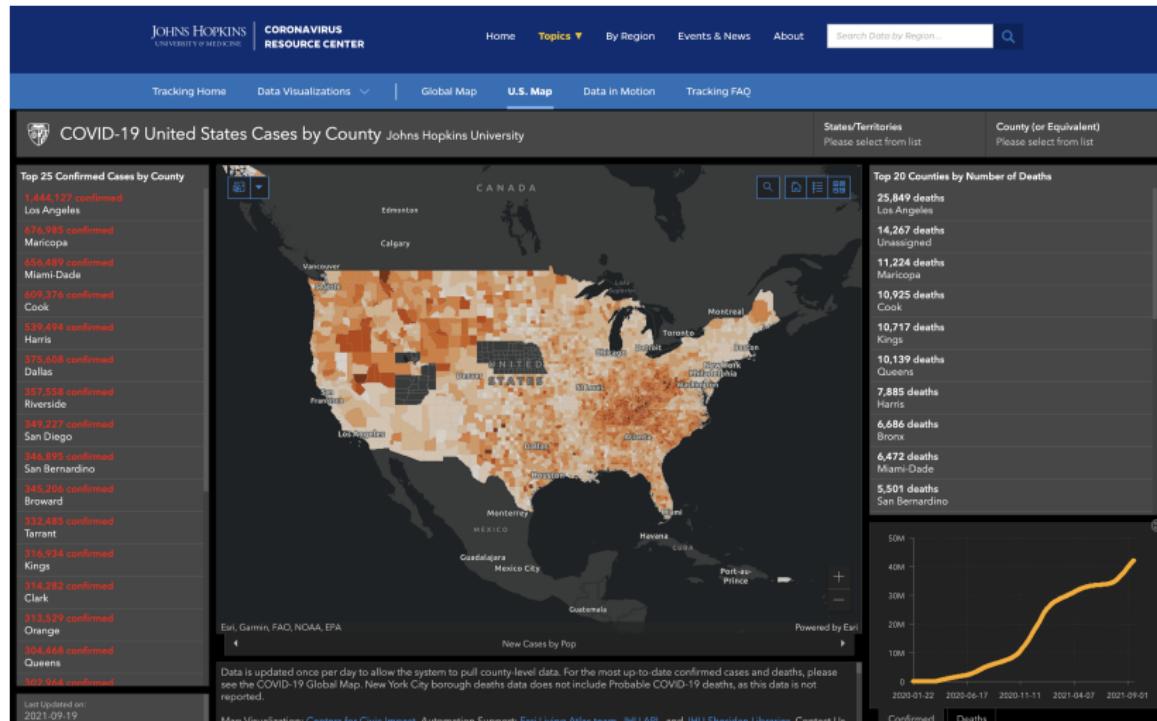
- ▶ How confident are we about our visualized predictions?



## Things to think about

- ▶ What is the goal of the graph – what information do you want to convey and for what purpose?
- ▶ What is your eye drawn to? Is that relevant?
- ▶ What does your brain tell you about what you conclude from that? Is that the correct conclusion?
- ▶ Visualization is reducing complexity (hopefully), but is the reduction informative or confusing?
- ▶ Graphs overwhelm text and tables so think about them!

# Is this a good visualization of the pandemic in the US?



<https://coronavirus.jhu.edu/us-map>

# How about these?

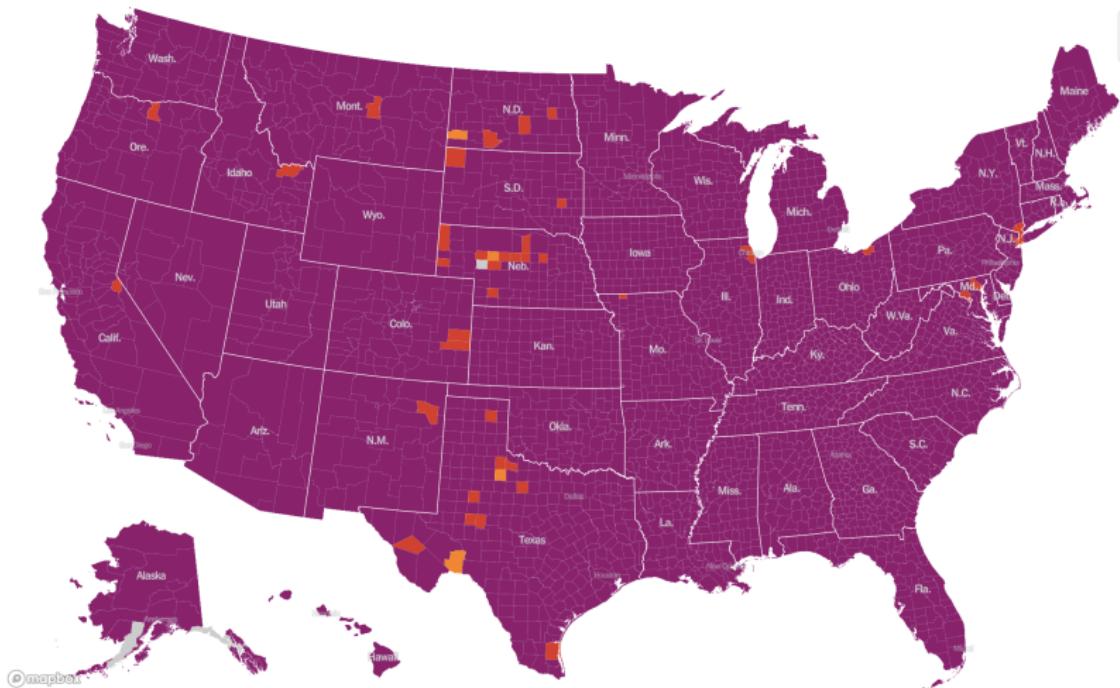
## A look back at Covid-19 risk through time

In the winter of 2020, the overall Covid risk level throughout the country was much worse than earlier in the pandemic. Risk levels before September 2020 are not available because of a lack of widespread testing and data.



# How about these?

Covid-19 risk for unvaccinated people is based on cases and test positivity.  
■ LOW ■ MODERATE ■ HIGH ■ VERY HIGH ■ EXTREMELY HIGH

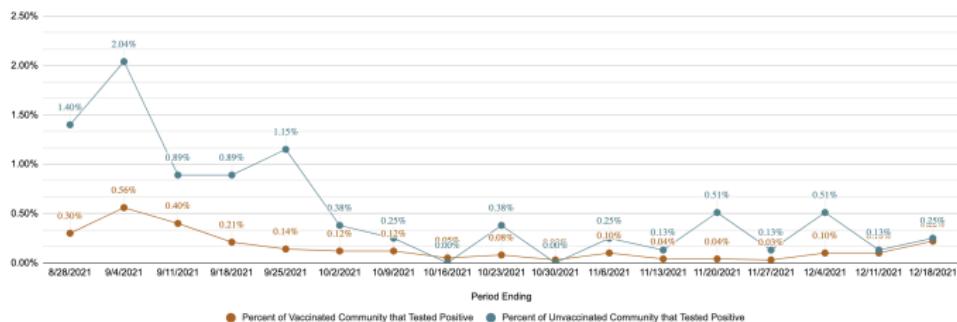


Source: Covid-19 risk assessment by The New York Times and Resolve to Save Lives based on reported cases and test positivity data. [Read more below.](#)

# How about these?

**The Vanderbilt Community is approximately 96 percent fully vaccinated.** This total is subject to change as additional data is available.

Vanderbilt is publishing the final edition of this year's weekly COVID-19 dashboard on Dec. 22, 2021, covering the period Dec. 12-18, 2021, which was the last week of the fall semester. We will not be publishing COVID data over the winter break but will continue to monitor COVID-19 prevalence closely. We will resume publishing data on Jan. 26, 2022, covering the week Jan. 16-22, 2022, the first week of the spring semester.



## Visualization is very powerful

- ▶ Humans infer causality very quickly so make sure that the relationships are not misguided.
- ▶ Make sure people get what you intend from the visualization.