

* Recording?

INTERPRETING DATA VISUALIZATIONS & MAPS

Julia Guy, MA, MLIS (She/her)
Digital Projects & GIS Librarian

John Brosz, PhD
Data & Visualization Curator

August, 28, 2024

Slides: <https://brosz.ca/slides/>

Peoples of Treaty 7

Blackfoot Confederacy

Siksika



Piikani



Kainai



Tsuut'ina



Stoney Nakoda

Bearspaw



Chiniki



Goodstoney



Métis Nation of Alberta Districts 5 & 6



The University of Calgary, located in the heart of Southern Alberta, both acknowledges and pays tribute to the traditional territories of the peoples of Treaty 7, which includes the Blackfoot Confederacy (comprised of the Siksika, the Piikani, and the Kainai First Nations), the Tsuut'ina First Nation, and the Stoney Nakoda (including the Chiniki, Bearspaw, and Goodstoney First nations). The City of Calgary is also home to the Métis Nation of Alberta (Districts 5 and 6).



Why Interpreting Visuals Correctly Matters

- Research builds on research
- Anyone can make something that looks legit
- Visualizations are very persuasive
- Visuals and maps are shared without context

Root Causes of Misleading Visuals

When Is a Visualization/Map Misleading?

When a visualization (chart, graph) or map leads you to an incorrect interpretation of the data

Data vs Information

Visuals that Mislead

Data

- Poor Data

Poor Design

- Misuse of size
- Abuse of axes
- Pie charts that don't add up
- Misrepresentative bin sizes
- Correlation vs Causation
- Distorted maps

Intent to Mislead

- Cherry-picking / missing information
- Misleading descriptions
- Bad Colour

Data

Data does not need to be made up to be misleading

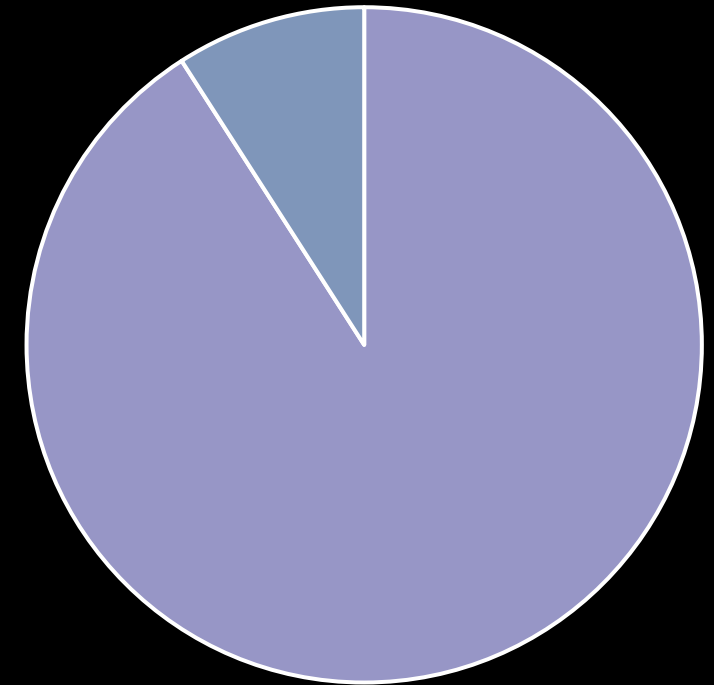
Data can be manipulated or shown in a way that means to deceive

It can be misunderstood

Data can be manipulated

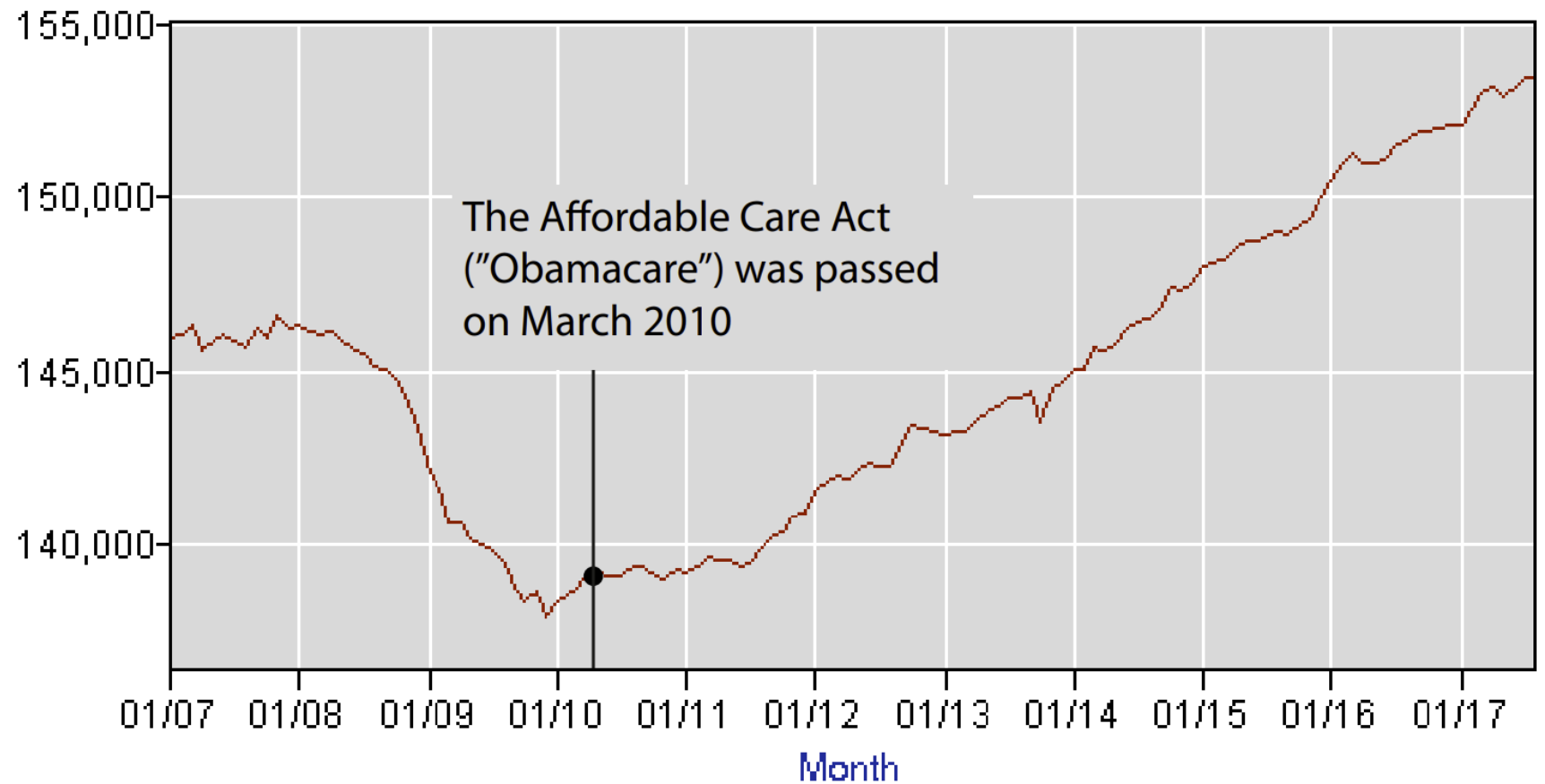
A visualization or map only shows what it shows

What Team is Most Likely to Win the NHL Playoffs

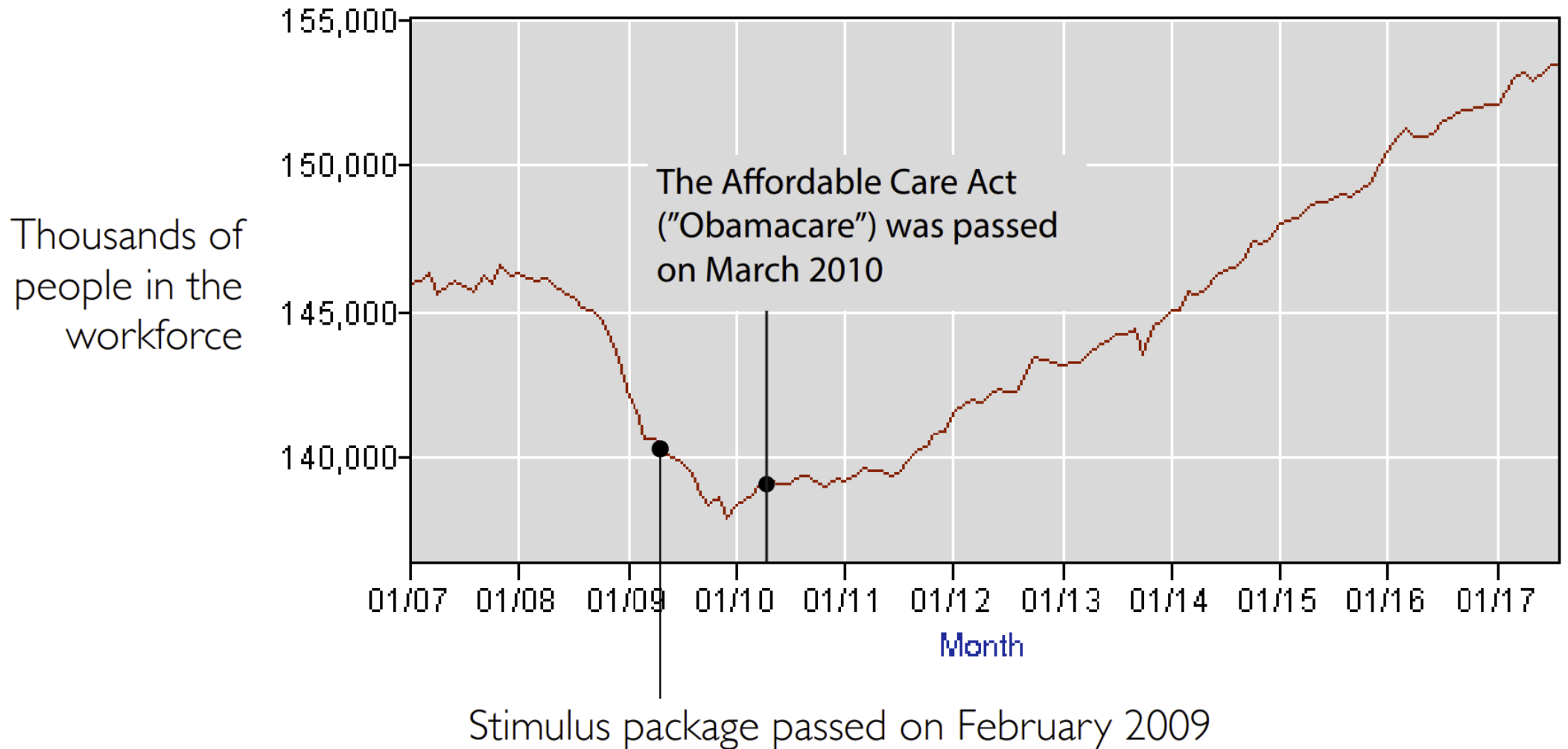


■ Calgary Flames ■ Edmonton Oilers

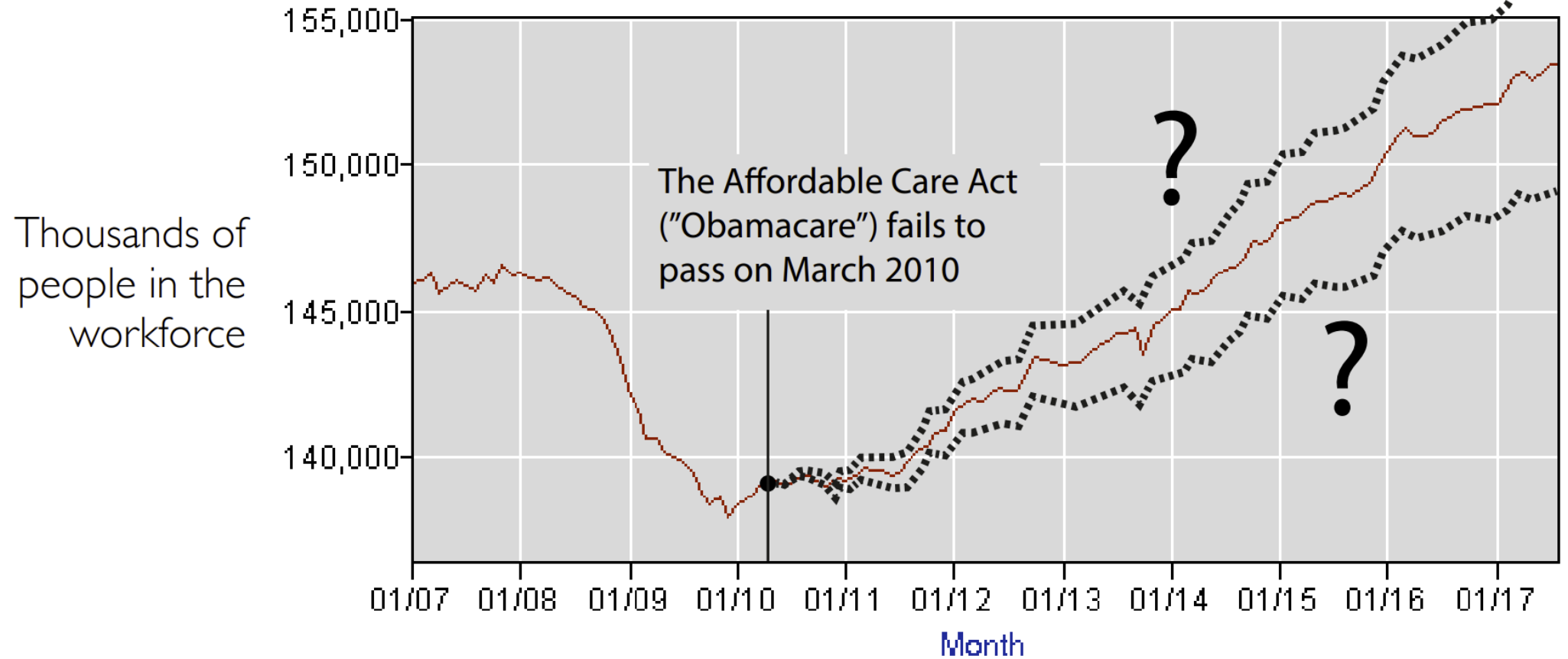
Thousands of
people in the
workforce



Key principle: A chart or infographic shows only what it shows *and nothing else*

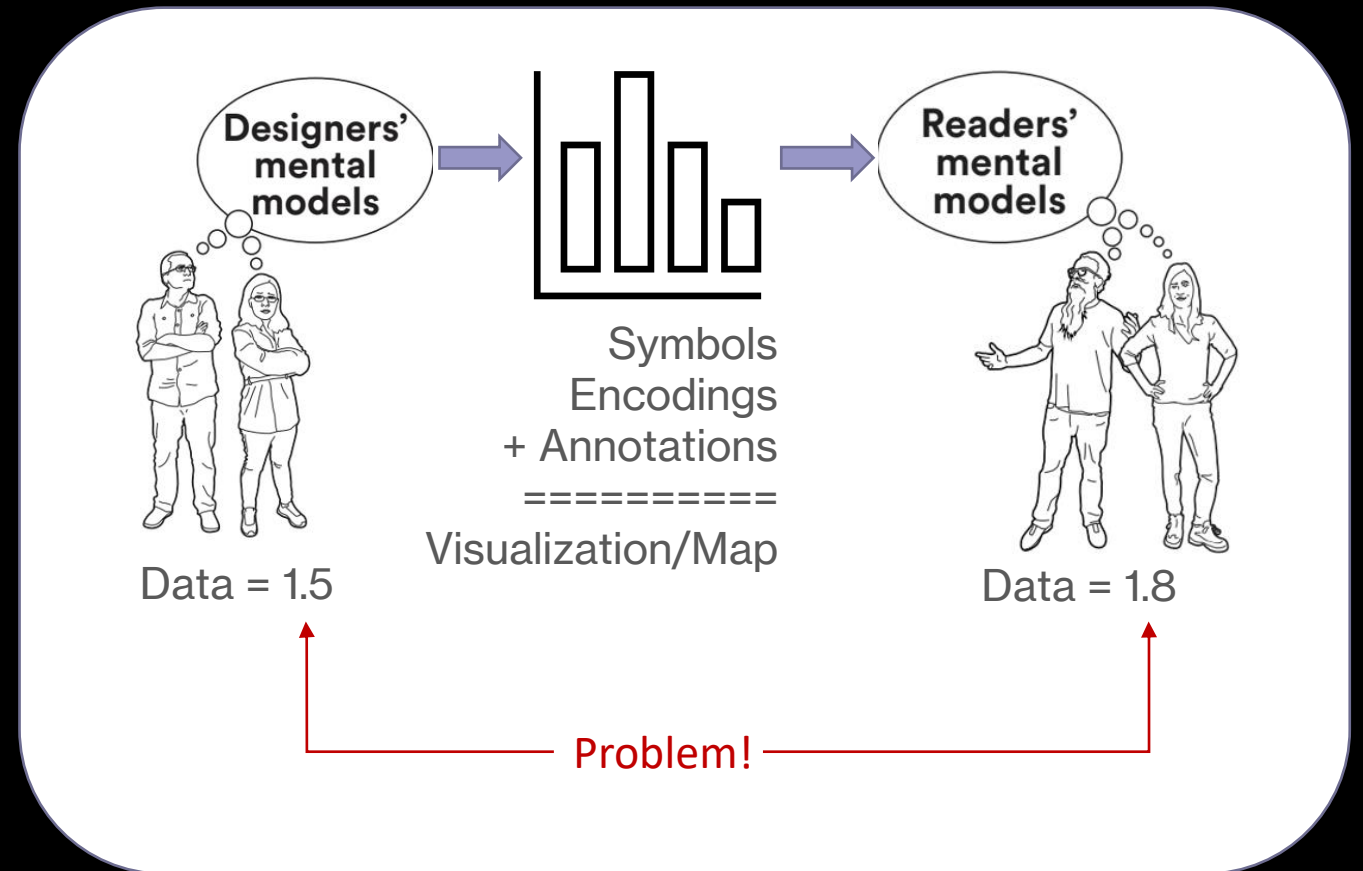


What If Obamacare Hadn't Passed?



Poor Design

- Decoding values
- Misuse of size
- Abuse of axes
- Pie charts that don't add up
- Misapplication of colour
- Correlation vs Causation
- Misleading maps



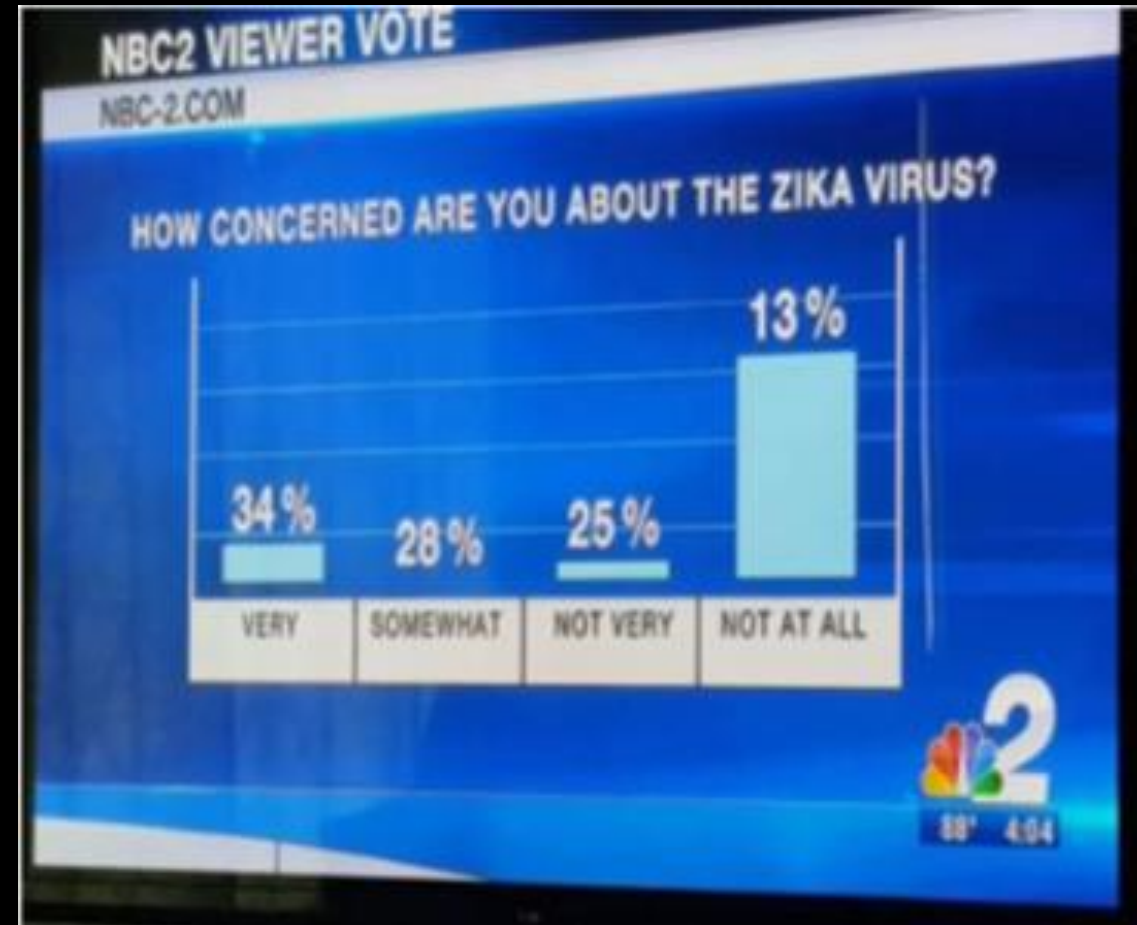
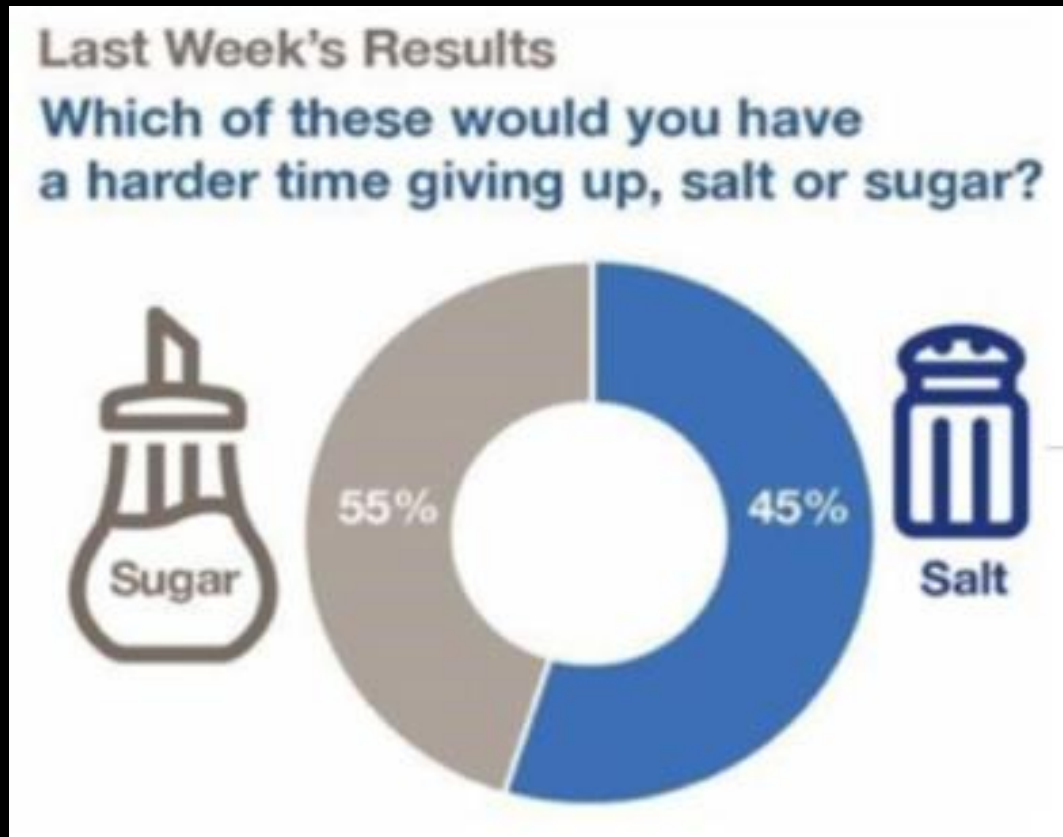
Lie Factor

Less noticed but no less important, Edward Tufte came up with a formula called the “Lie Factor” to calculate the level of accuracy of a graphic. The closer the Lie Factor is to 1.0, the more accurate the graphic is.

$$\text{LIE FACTOR} = \frac{\text{SIZE OF EFFECT SHOWN IN THE GRAPHIC}}{\text{SIZE OF EFFECT IN DATA}}$$

The “Lie Factor” — Edward Tufte (1983)

Decoding Values



Correlation & Causation

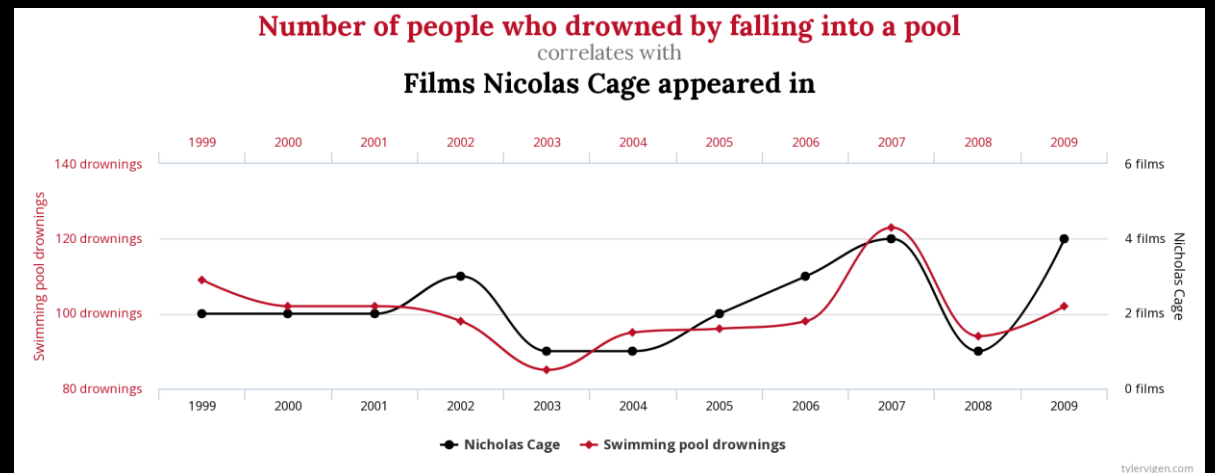
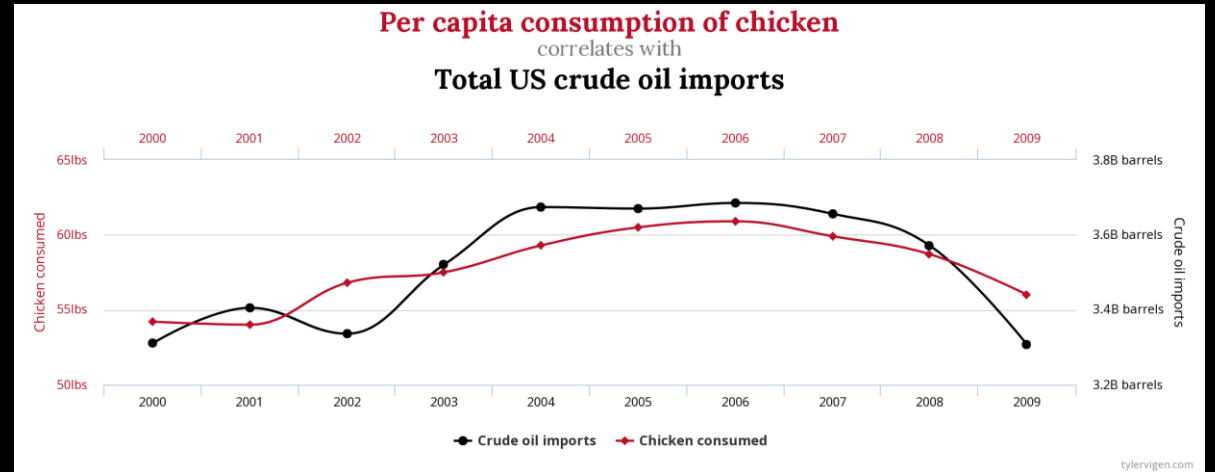


Oded Rechavi 🐦 @OdedRechavi · 14h
Correlation Vs. Causation



53 876 5,774

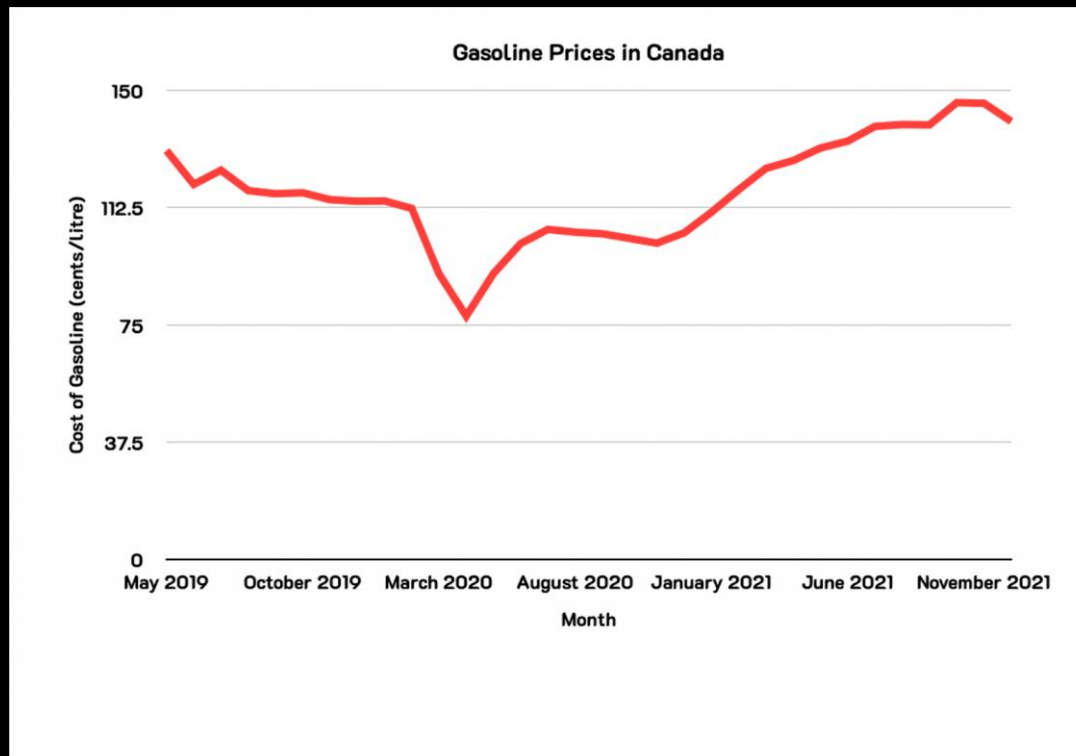
<http://tylervigen.com/spurious-correlations>



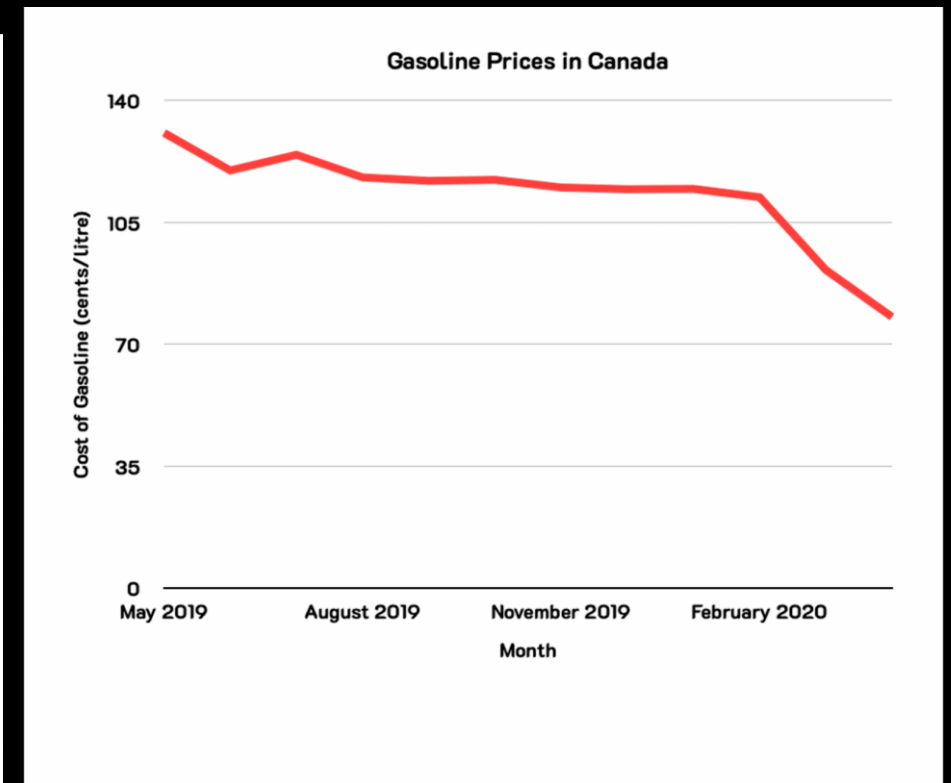
Intent to mislead

- Cherry-picking
- Misleading descriptions
- Bad Colour

Cherry Picking Data Aggregation

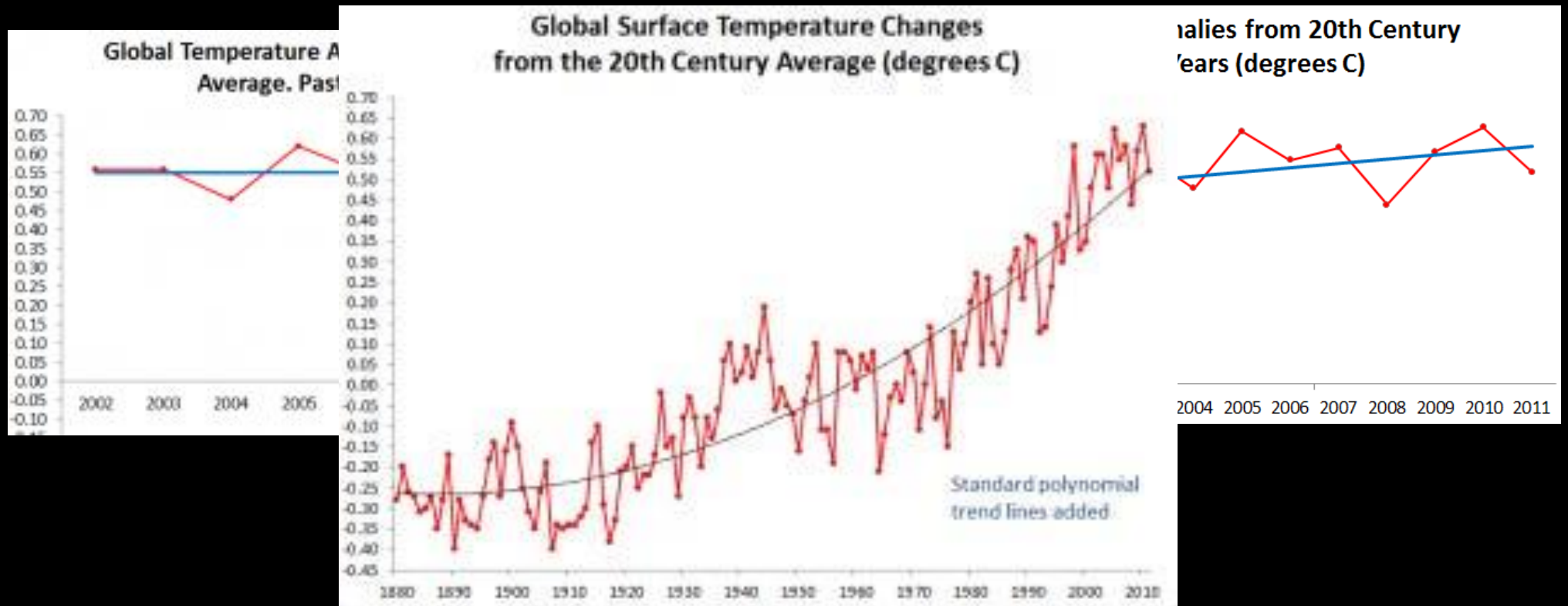


Weekly Average



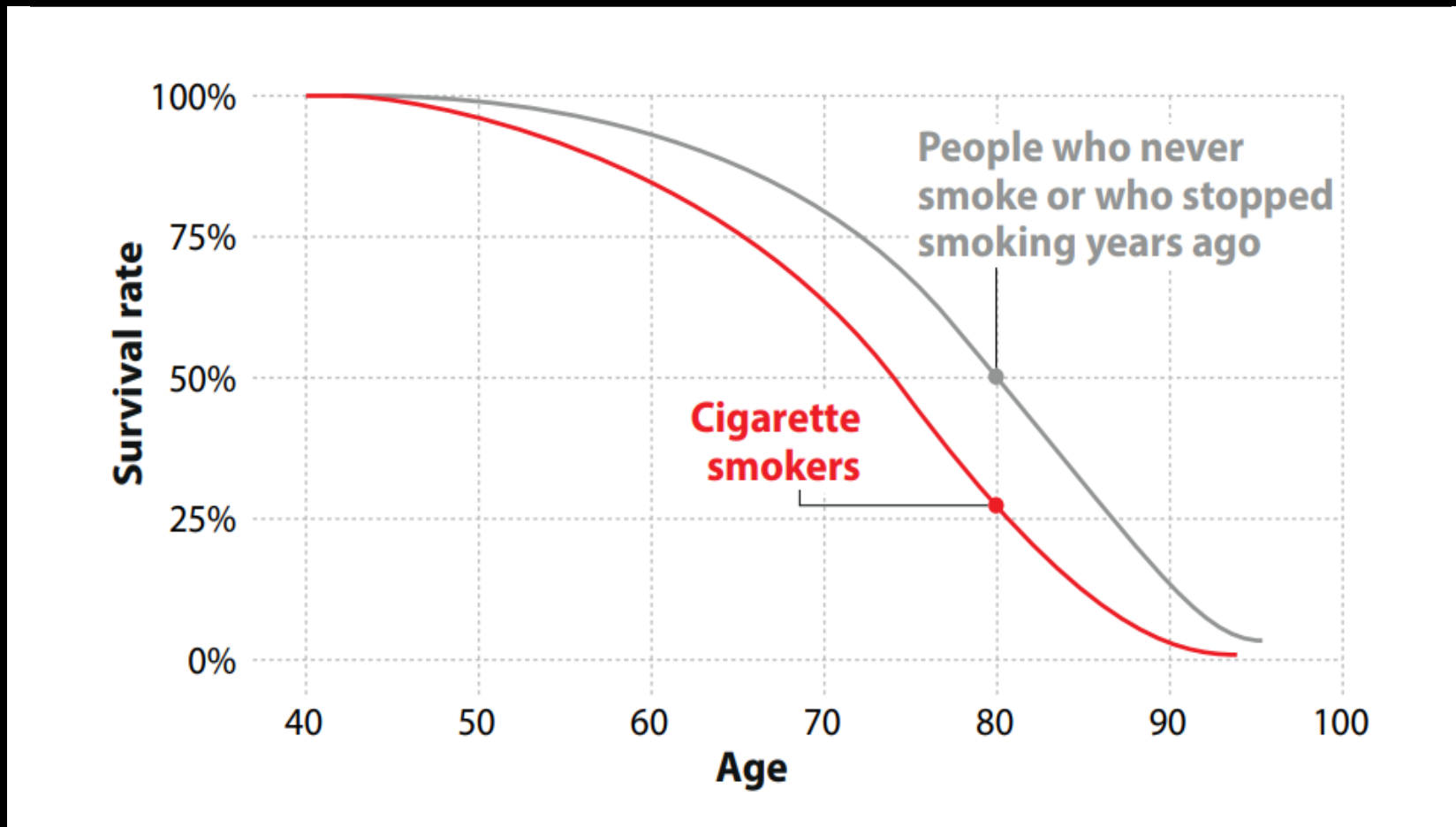
Monthly Average

Cherry Picking Data Selecting the “Right” Range



DESCRIPTIONS

The way we describe the content of a chart to ourselves may bias our understanding of that chart.



Maps

- Maps are already inherently misleading
 - The earth is round (ish) and maps are not
 - Western bias is common in maps
 - Generalization and simplification are necessary
 - You can't fit everything on a map, something will always be missing

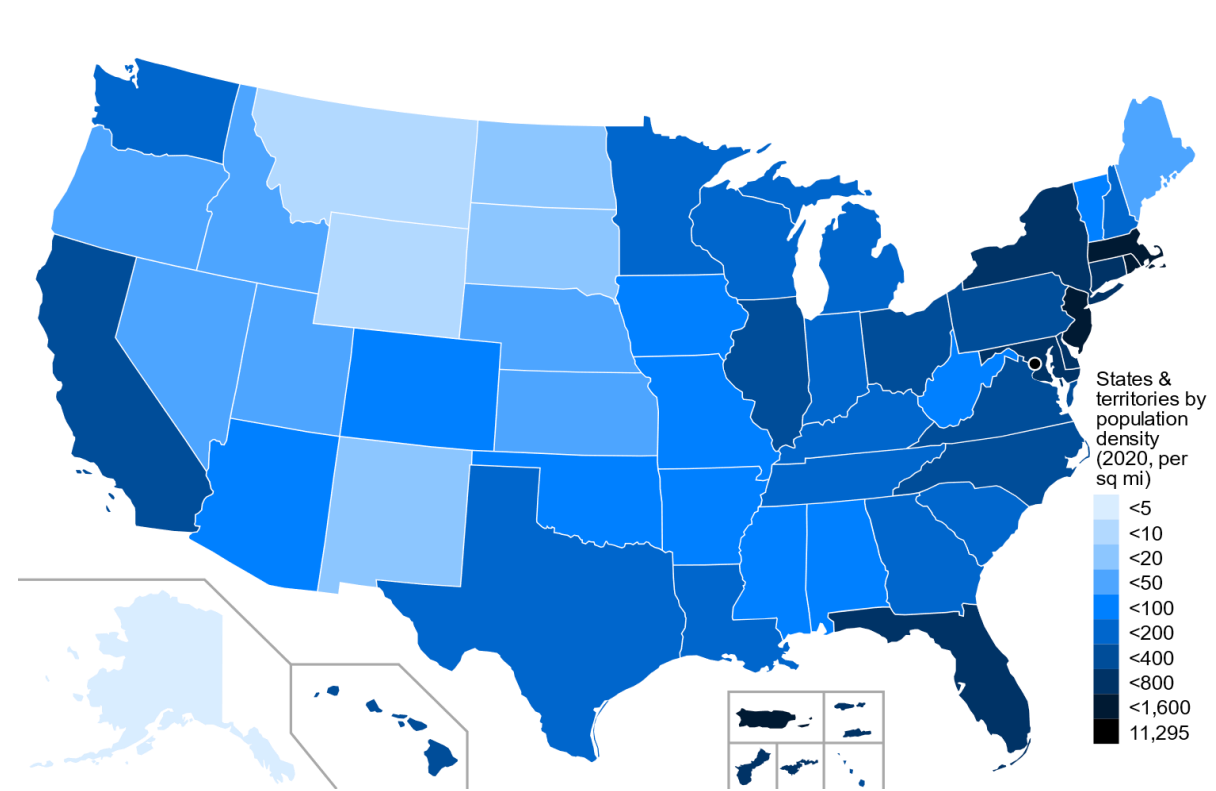
Examples of Types of Maps

Choropleth: Uses established boundaries like government/administrative boundaries or borders and fills these shapes with a colour to show the data

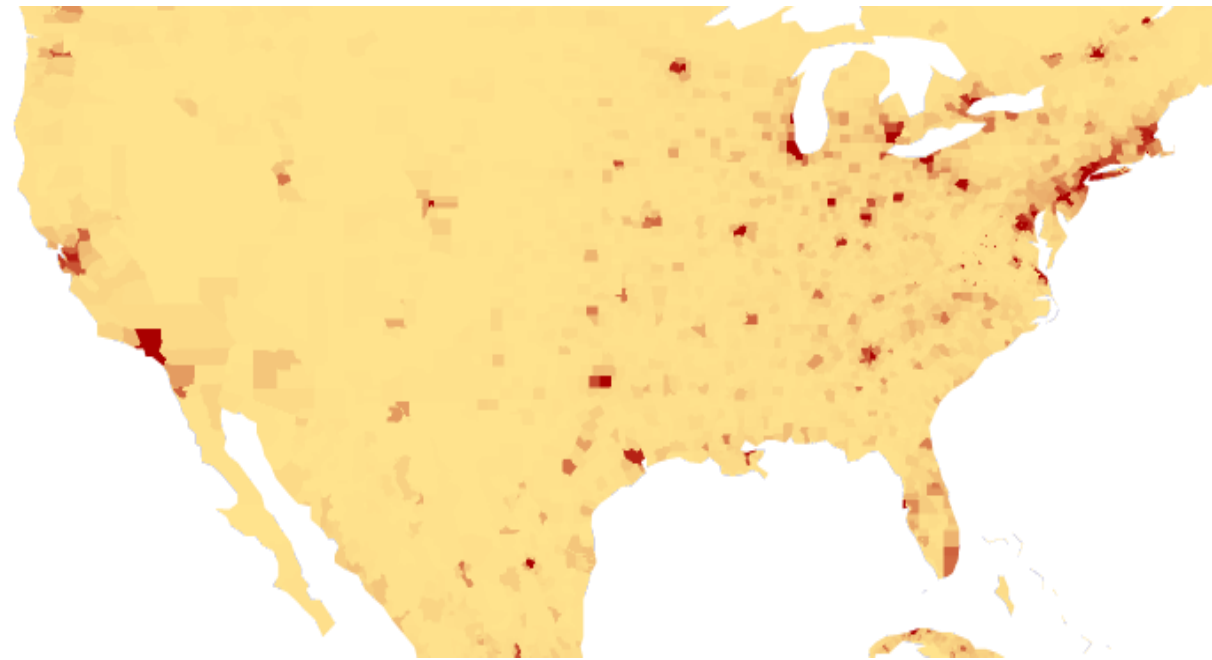
Heat Map: Shows distribution of data independent of established boundaries (e.g. a grid map where each cell is the same size, or a point for each data point)

Cartogram: A map-like image where the geographic size is modified/warped to reflect something else (like population or another variable)

Choropleth Map vs. Heat Map



U.S. states and territories shaded by population density as of the 2020 census

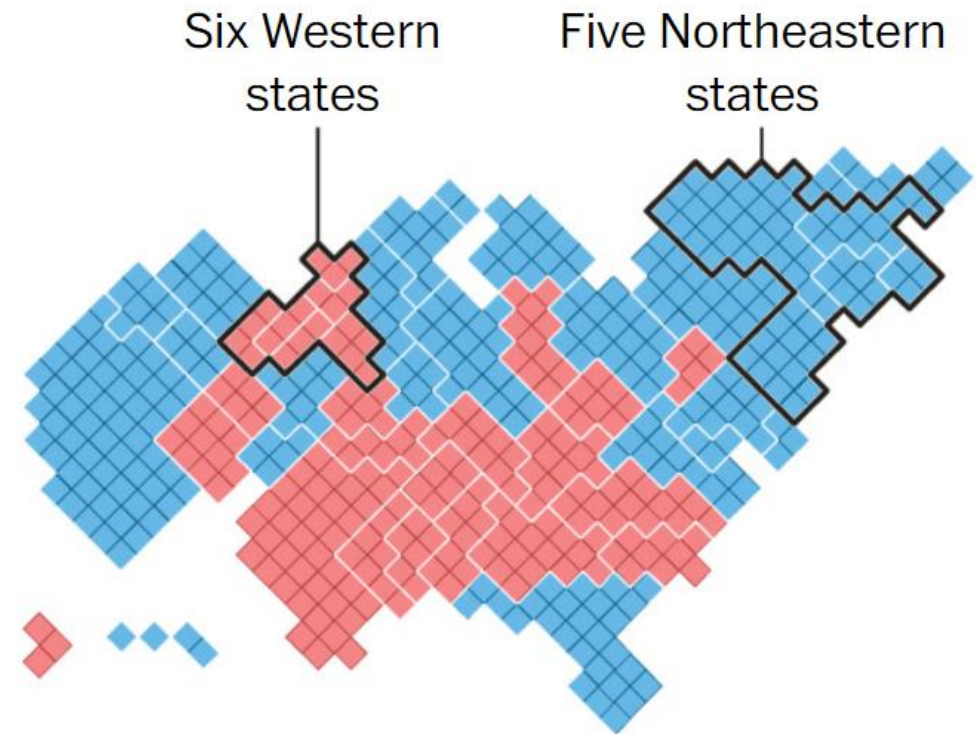
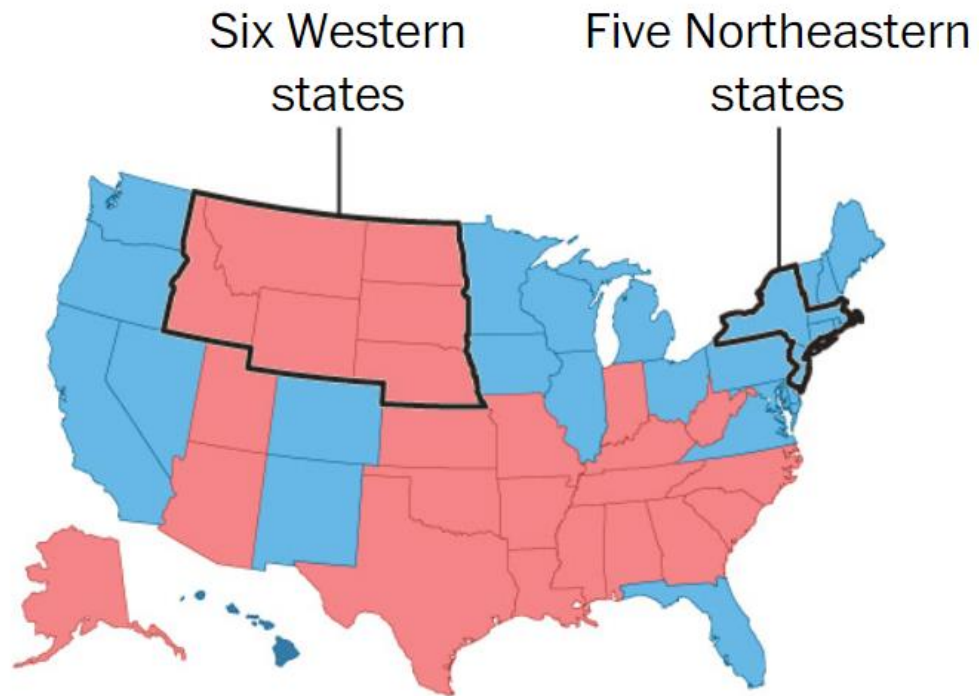


The number of people per square kilometer around the world in 1994 (NASA)

Choropleth vs. Cartogram

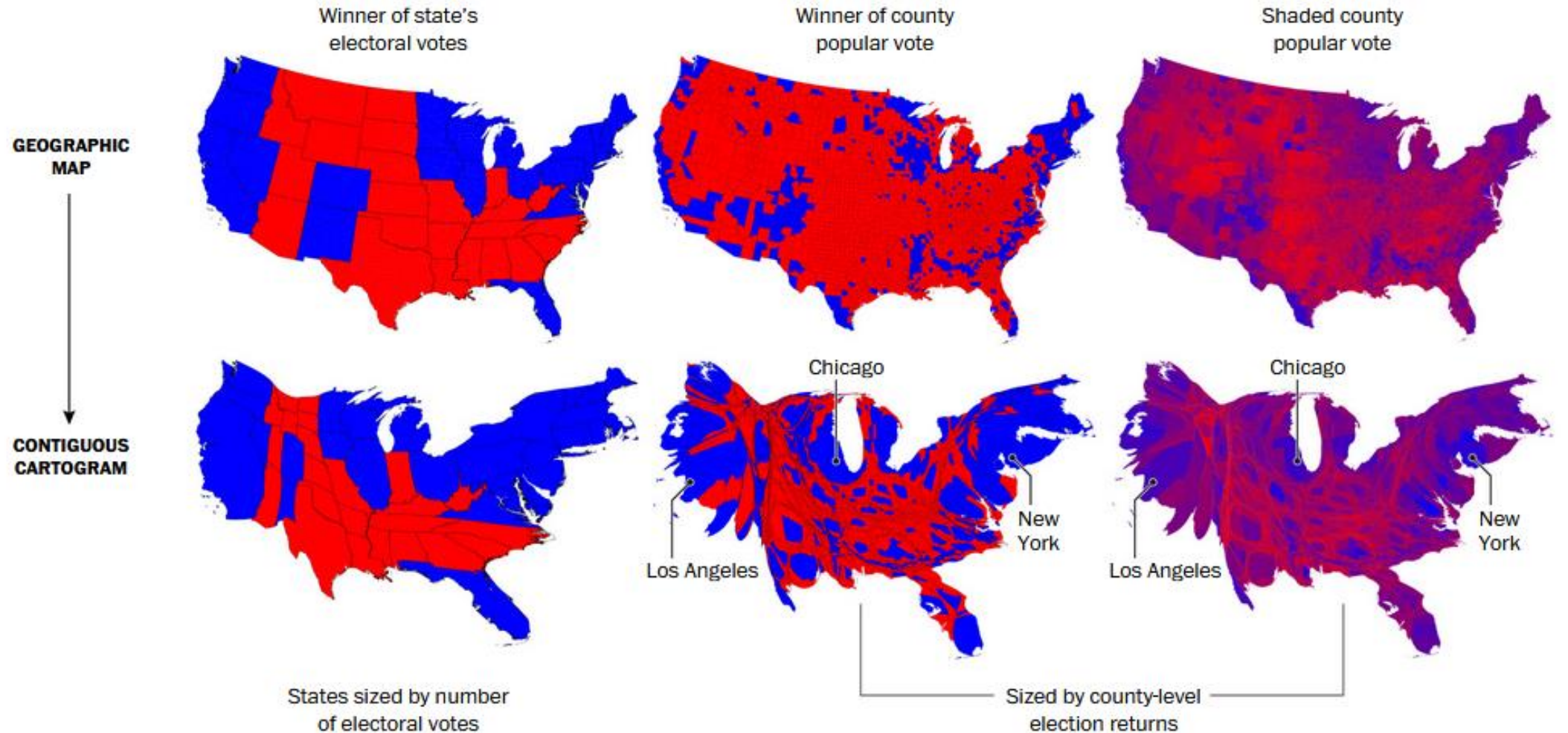
GEOGRAPHIC MAP

CARTOGRAM OF ELECTORAL VOTES



Washington Post created this cartogram for poll data
<https://www.washingtonpost.com/graphics/politics/2016-election/how-election-maps-lie/>

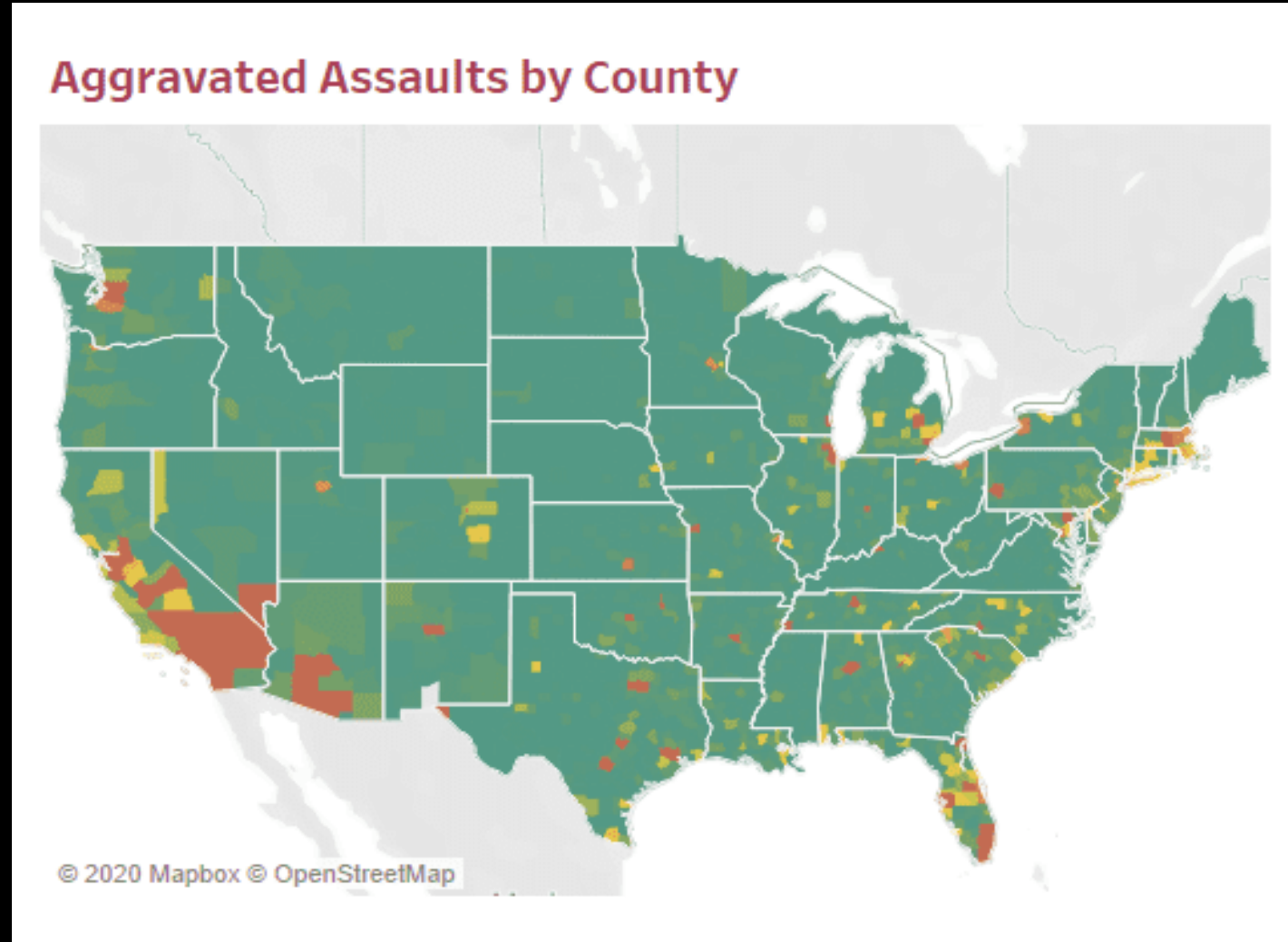
Choropleth vs. Cartogram cont.



Maps courtesy of Mark Newman

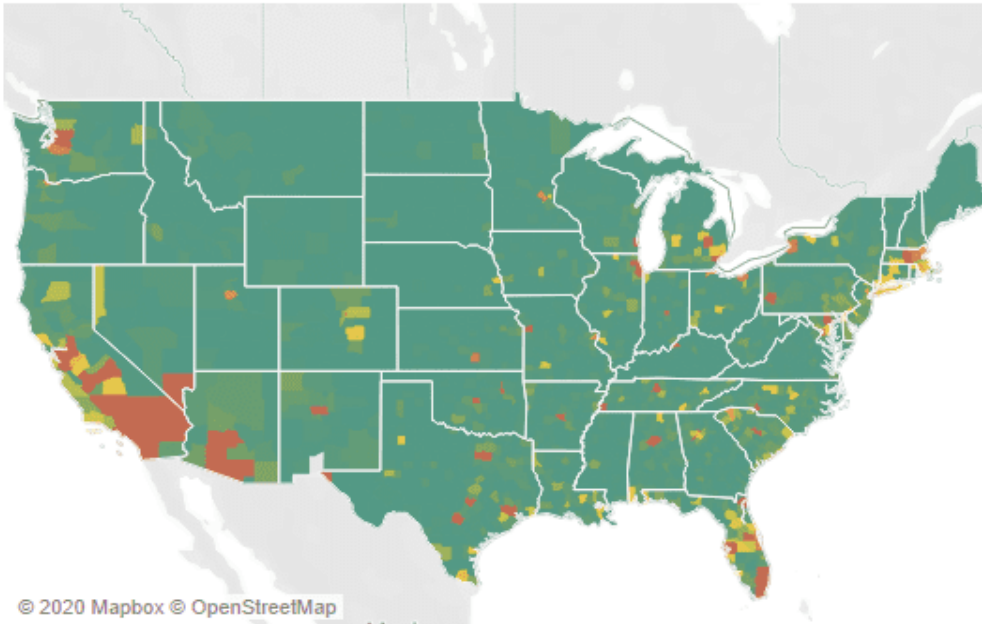
Mark Newman created these maps to show 2004 Presidential Election Results
<https://www.washingtonpost.com/graphics/politics/2016-election/how-election-maps-lie/>

What could explain this data?

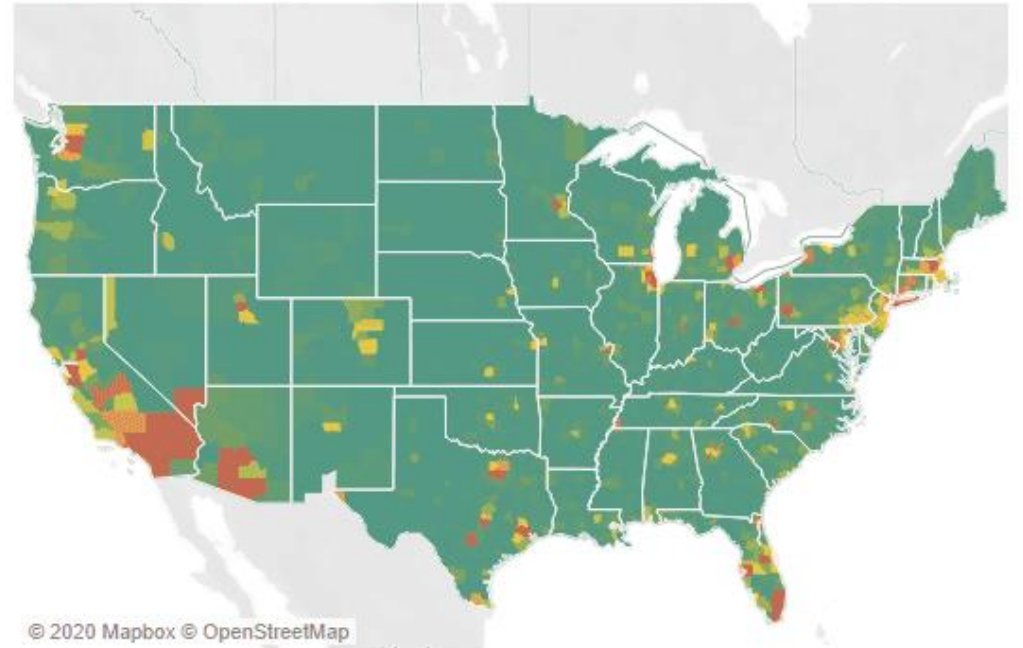


Distorted Maps

Aggravated Assaults by County

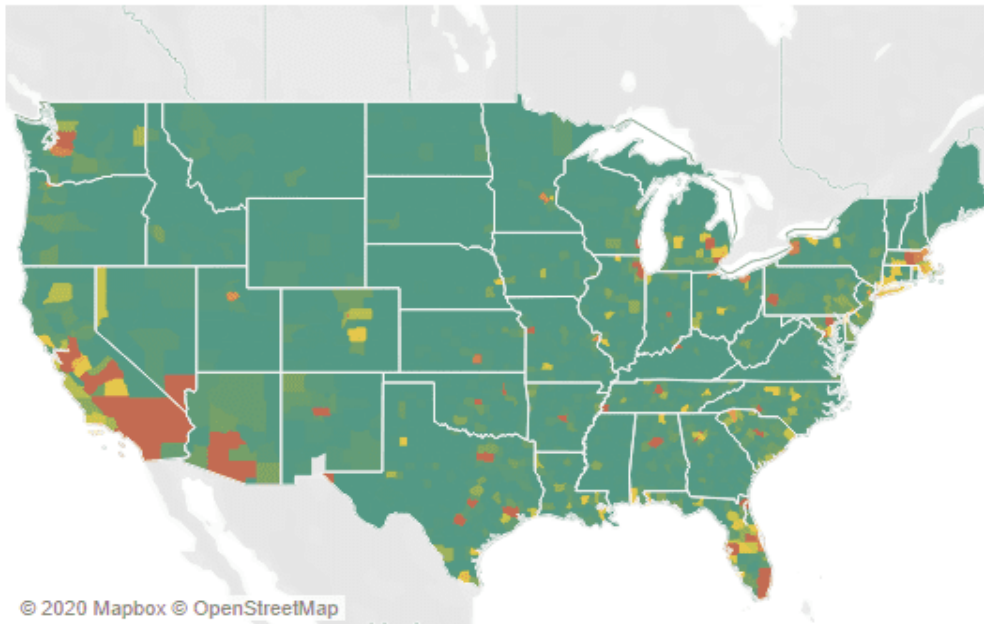


Population by County

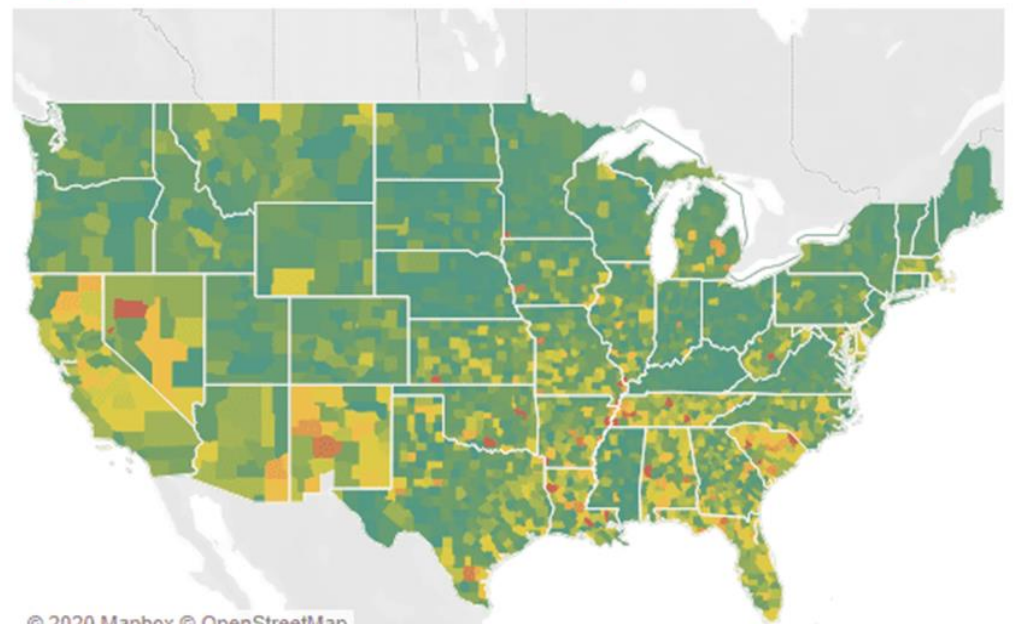


Distorted Maps

Aggravated Assaults by County



Aggravated Assaults Per Capita by County



Data Classification

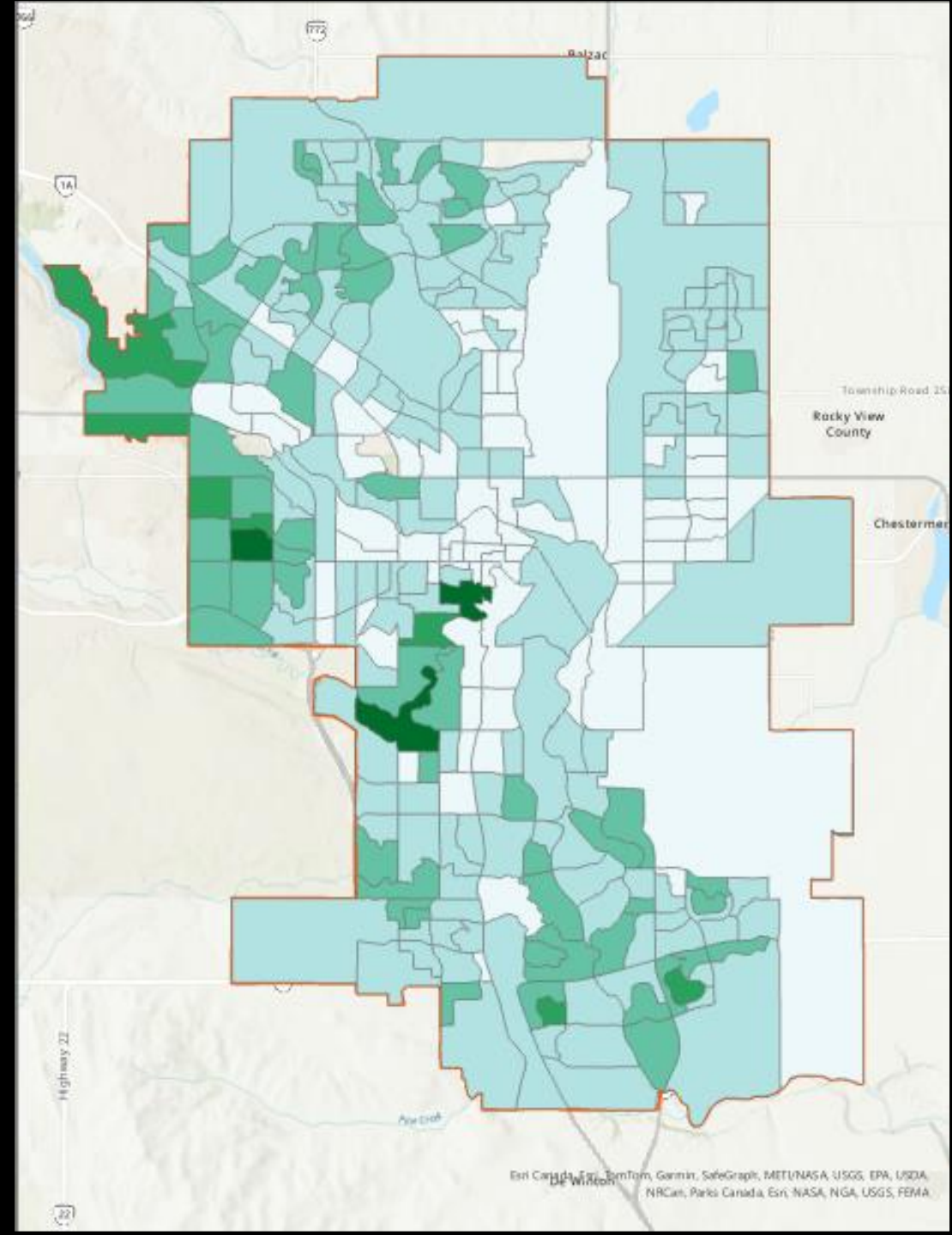
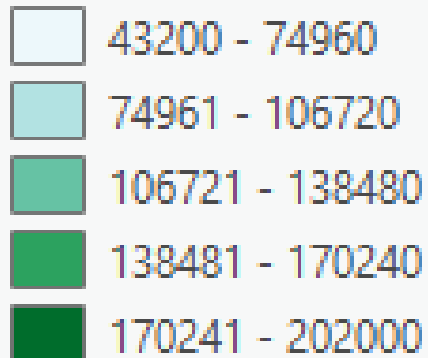
Methods Examples

- Equal interval – Each class represents an equal ranges
- Quantile – Each class contains an equal number of features
- Manual interval – Define classes in any way you want
- Standard interval – Classes created using mean values and the standard deviations from the mean to show how the values vary from the mean
- Natural breaks (Jenks) (ESRI) – Classes grouped by “natural” patterns in the data

Example: Equal Interval

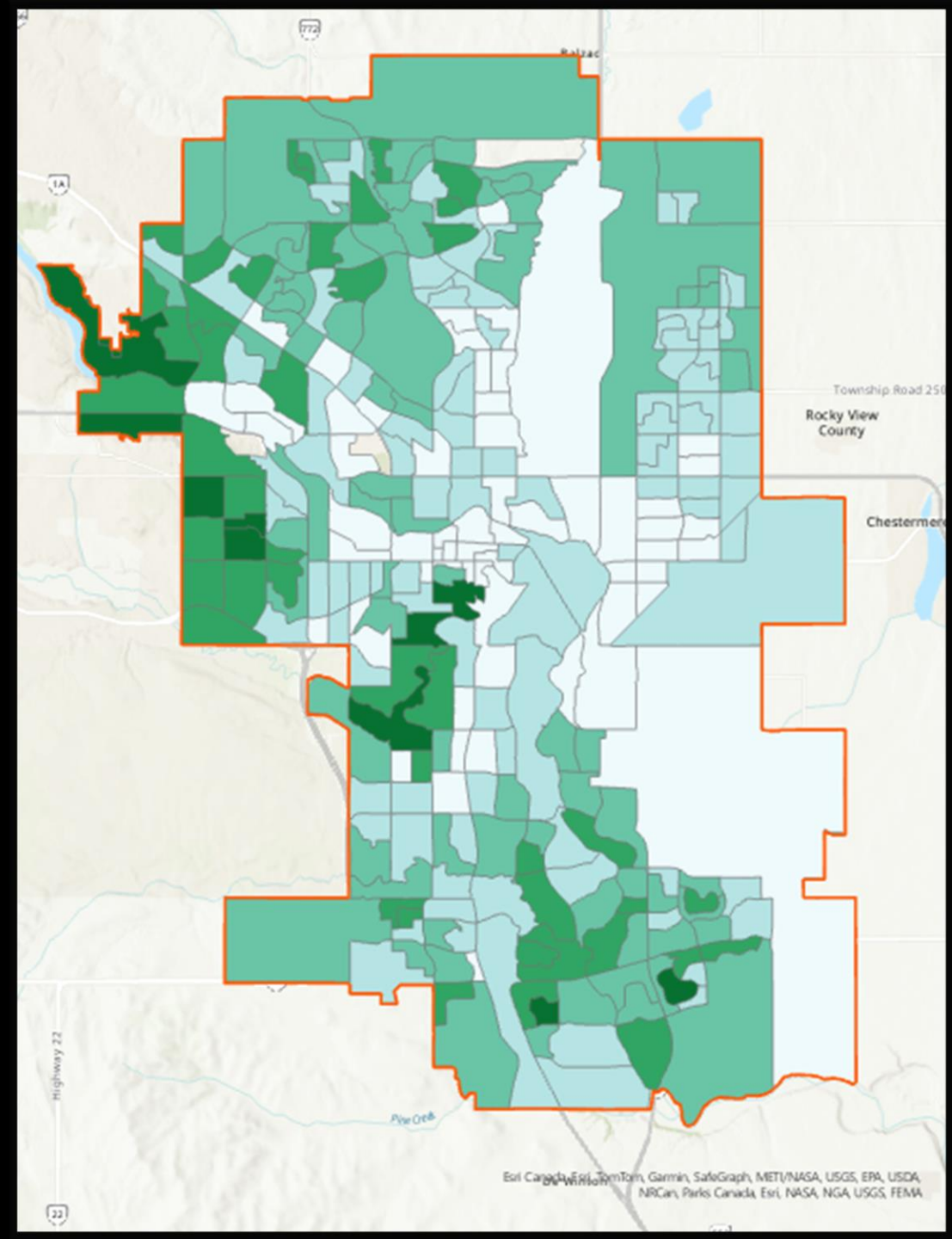
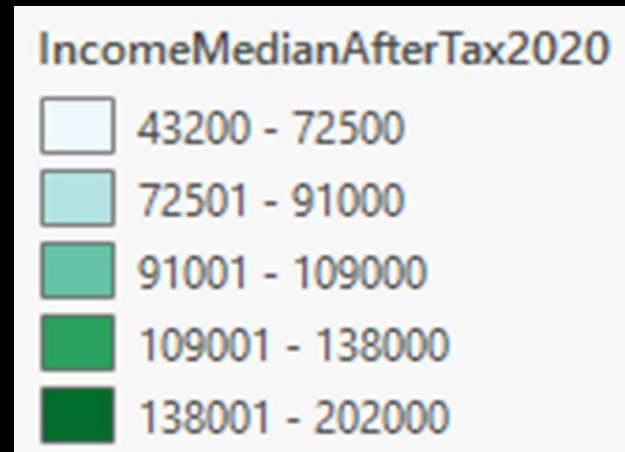
2021 Census Data by Census Tract
2020 Median Household Income After Tax

IncomeMedianAfterTax2020



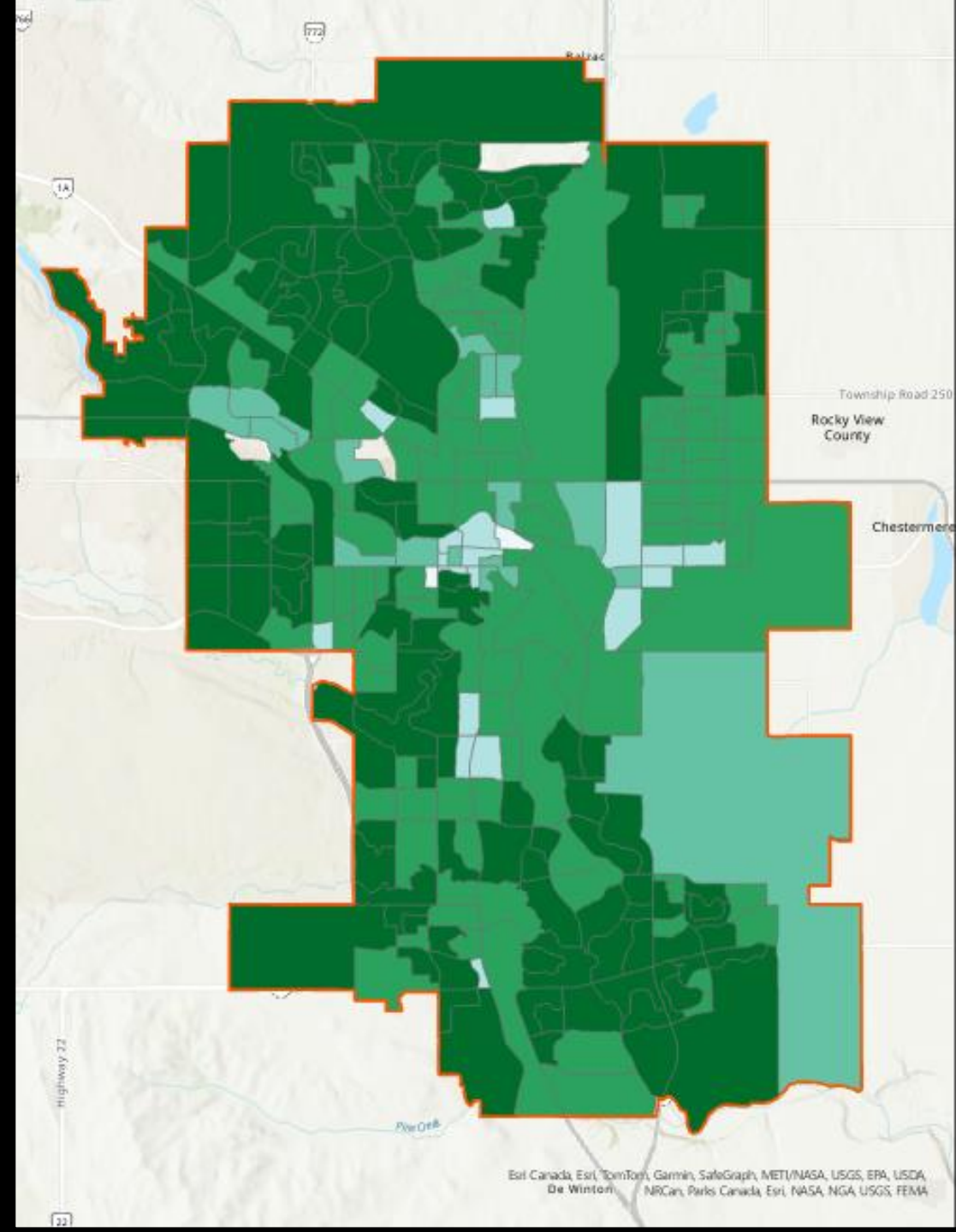
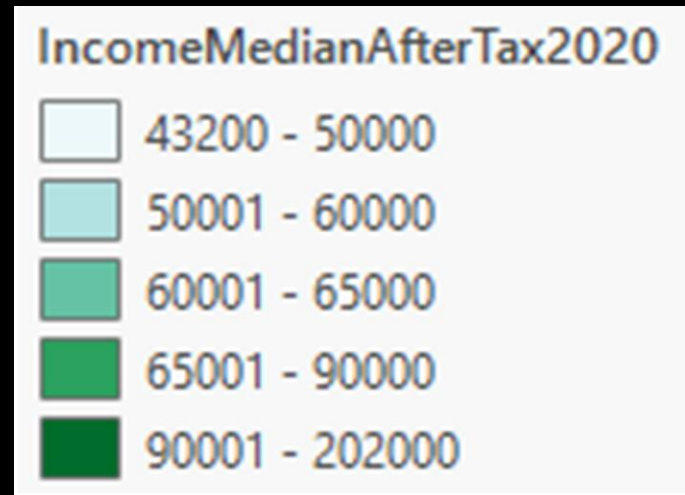
Example: Natural Breaks (Jenks)

2021 Census Data by Census Tract
2020 Median Household Income After Tax



Example: Manual Interval

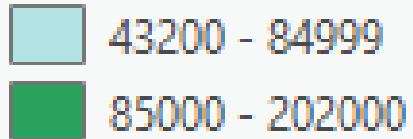
2021 Census Data by Census Tract
2020 Median Household Income After Tax



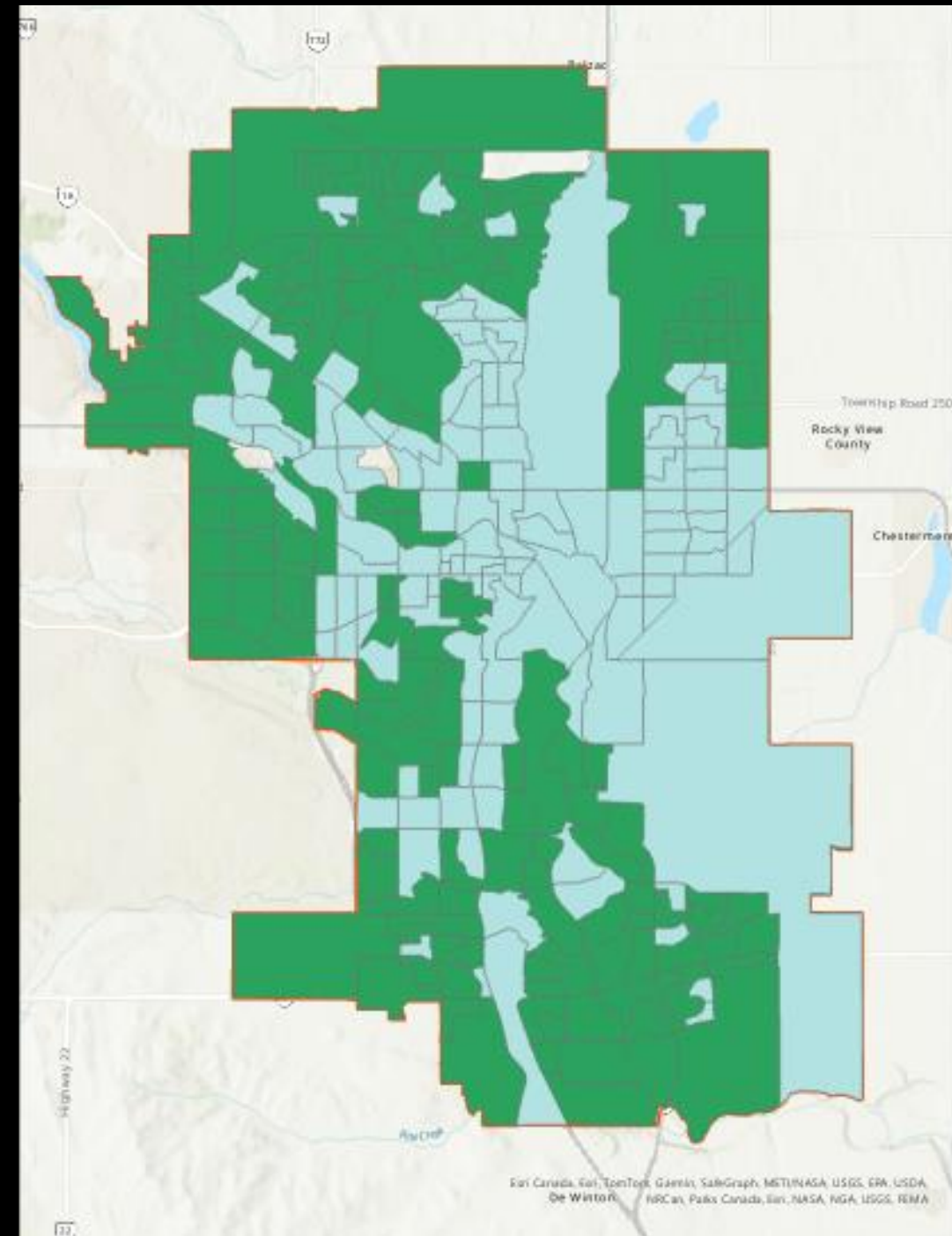
Example: Manual Interval 2

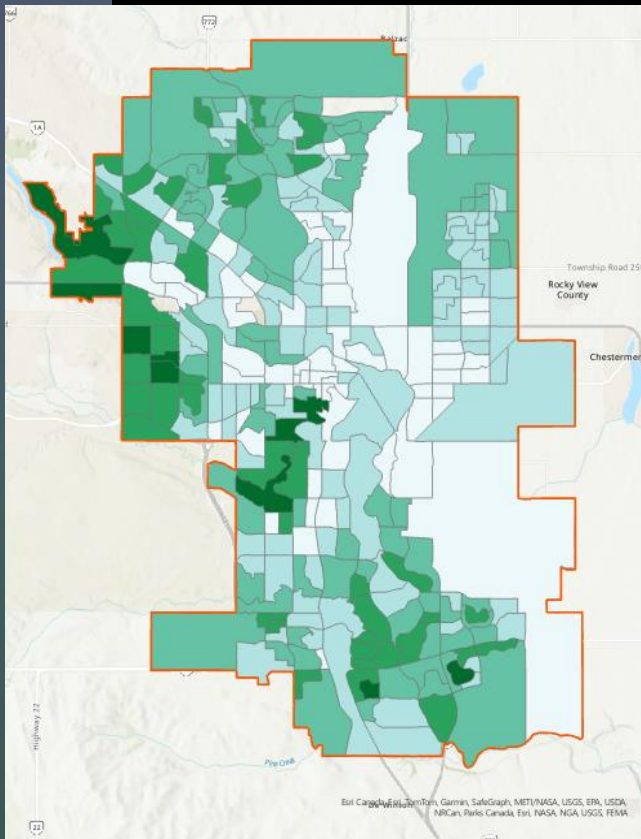
2021 Census Data by Census Tract
2020 Median Household Income After Tax

IncomeMedianAfterTax2020

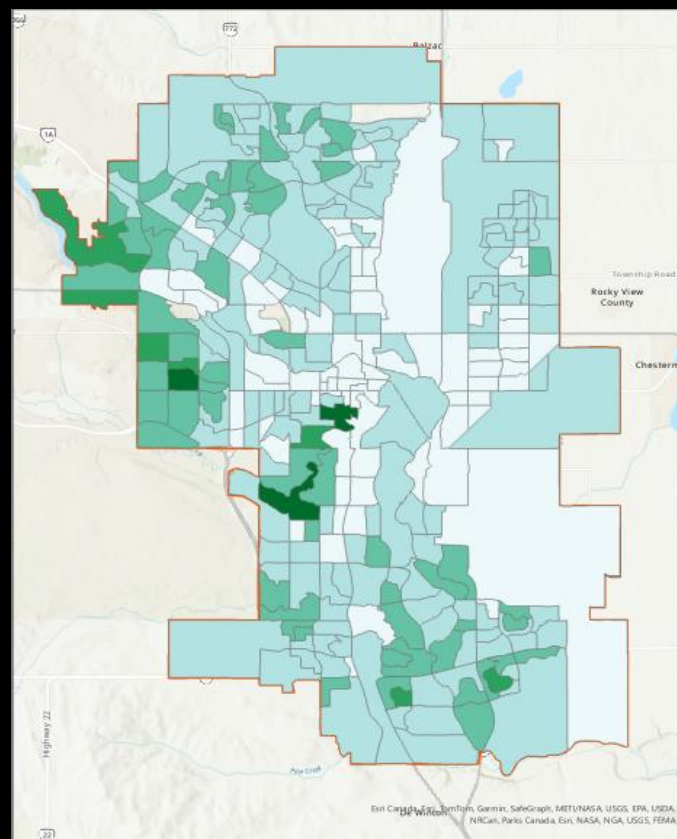


2020 Median Household income
after tax for Calgary: 85,000

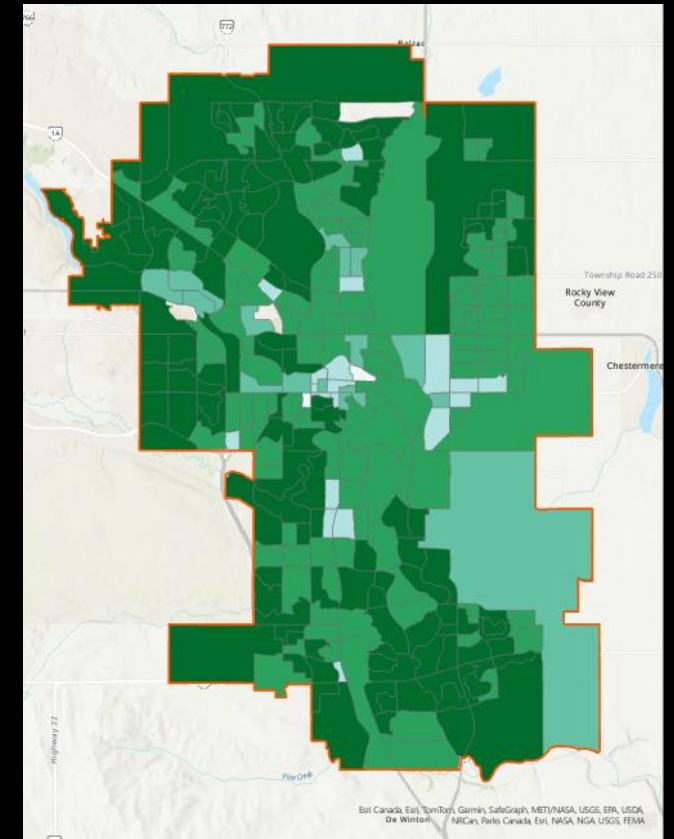




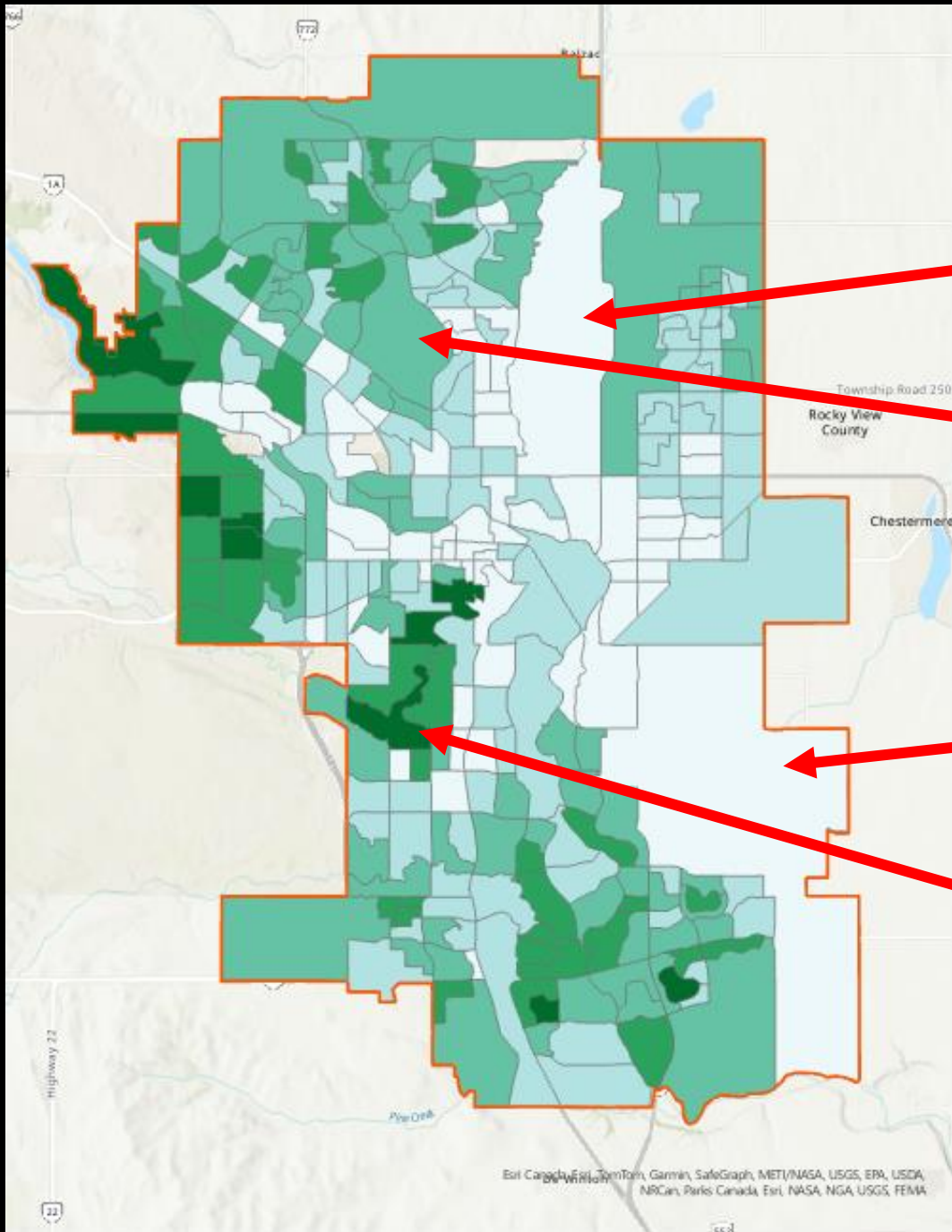
Natural Breaks



Equal Interval



Manual Interval (1)



Airport area

Park area

Largely commercial area

Reservoir

Reminder: Choropleth maps can give false impressions

Symbology

- Bias in symbology:
 - Colour – We associate darker colours with higher quantity, red with heat, pink and blue with gender etc.
 - Size – How large the symbol is will impact our interpretation
 - Icons – What icon has been chosen to represent that thing and why

Understand the Context

Look at the source, creator/organization/author, metadata, full article the map is from, or any other accompanying information.

Who created this map? Why?

What assumptions have they made about you, the viewer?

Who funded this research?

How and when was data collected?

What/who is included and what/who is missing?

AI?

What is Data Visualization?

Charts & Graphs

Visual representation of data

“Transformation of the symbolic into the geometric”
[McCormick et al, 1987]

“... artificial memory that best supports our natural means of perception” [Bertin, 1967]

“Use of computer-generated, interactive, visual representations of data to amplify cognition”
[Card, Mackinlay, & Shneidermann, 1999]

Visualizations aren't just pretty pictures!!!!

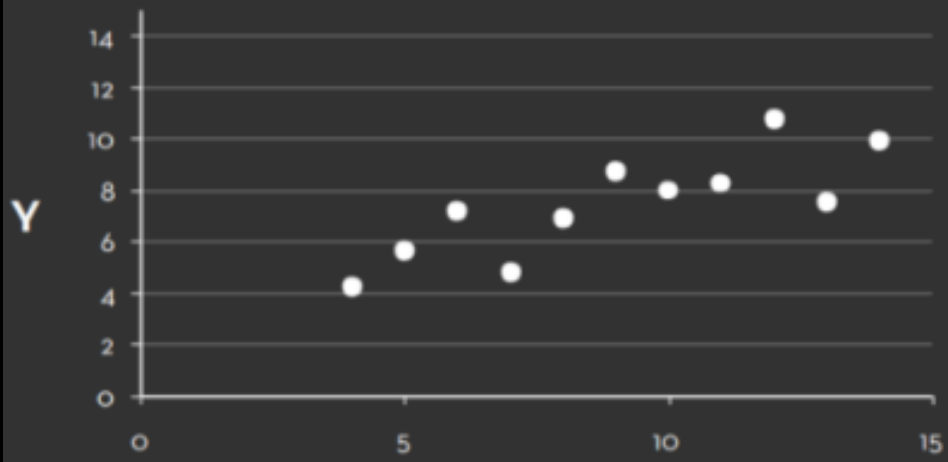


Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.08	10	9.14	10	7.47	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

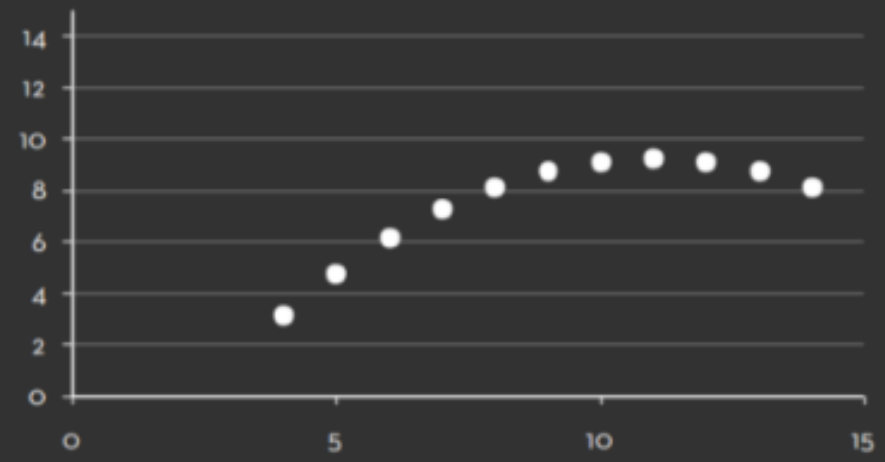
Summary Statistics
 $\mu_X = 9.0$ $\sigma_X = 3.317$
 $\mu_Y = 7.5$ $\sigma_Y = 2.03$

Anscombe's Quartet (Anscombe, Francis J., 1973)

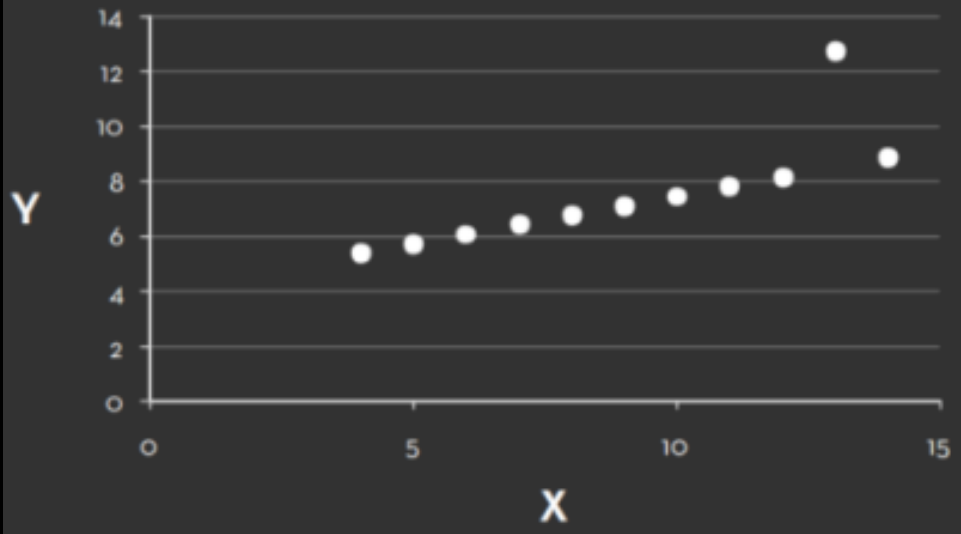
Set A



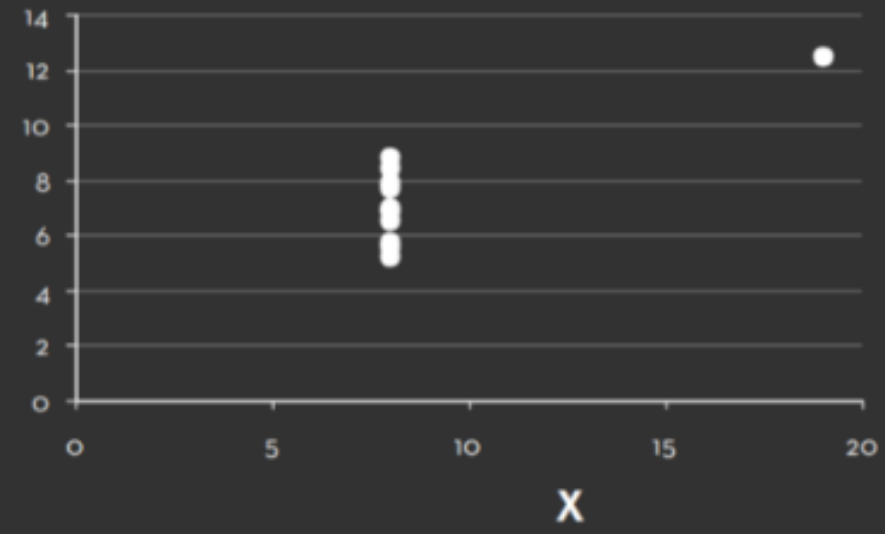
Set B



Set C



Set D



Every time Ford and Kavanaugh answered the question — and didn't answer the question

Click on any part of the transcript to expand

■ Answers question ■ Does not answer question

Ford



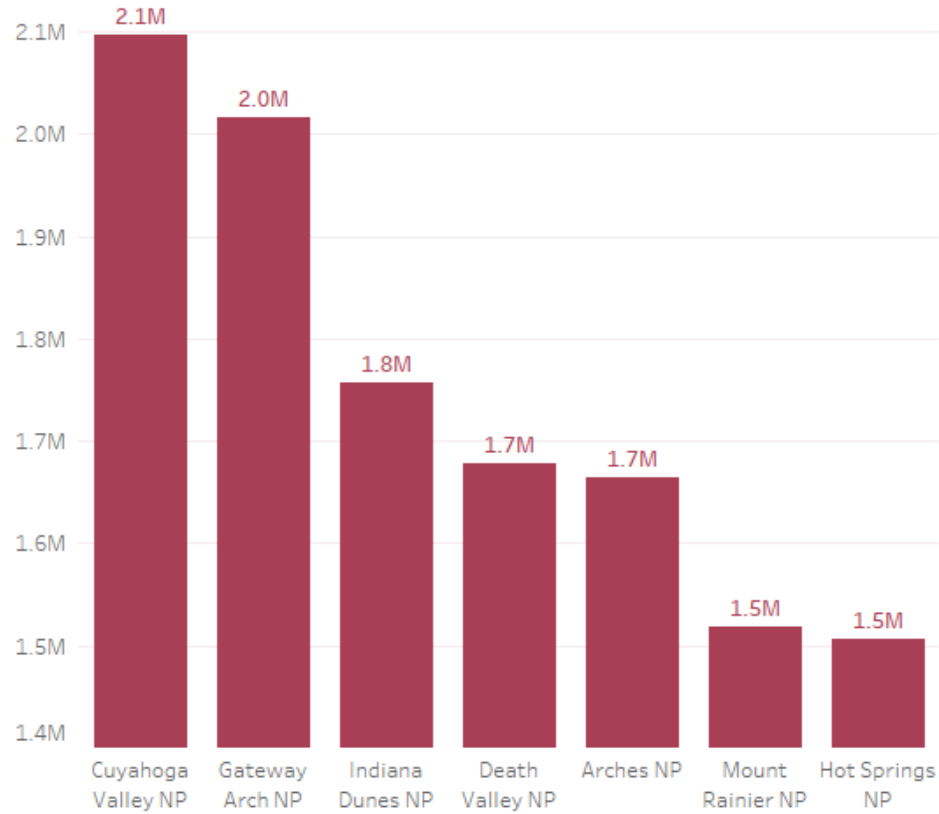
Kavanaugh



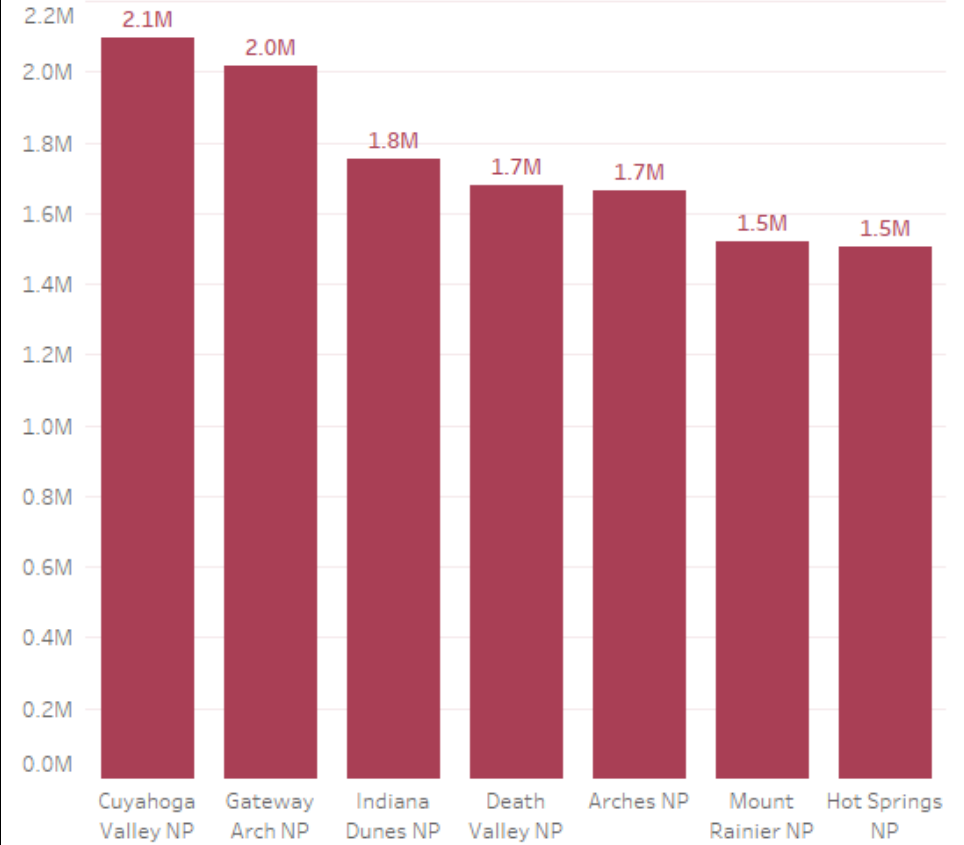
<https://www.vox.com/policy-and-politics/2018/9/28/17914308/kavanaugh-ford-question-dodge-hearing-chart>

Misuse of Size

Selected National Parks Visitation | 2018



Selected National Parks Visitation | 2018





18" (diameter)



12" (X2)

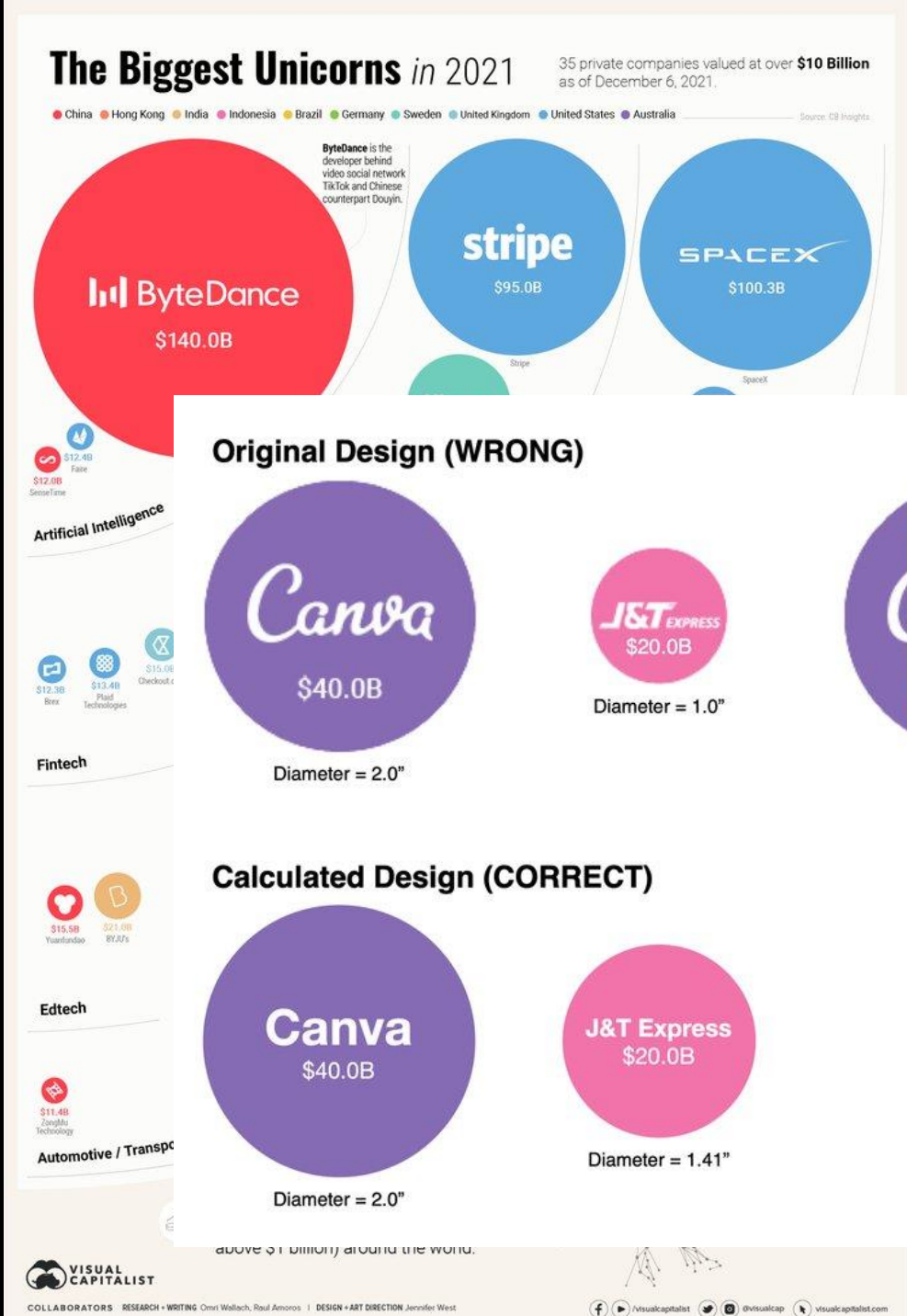
“Just think about it: if you tell a software tool to scale something 200 percent, it will make it twice as tall and twice as wide. Therefore, you aren’t doubling the size of your original circle. You’re making it four time larger.”

- The Truthful Art, Alberto Cairo (@albertocairo)

Pizza images from
<https://twitter.com/fermatslibrary/status/1082273172114862083/>

Misuse of Size (2D)

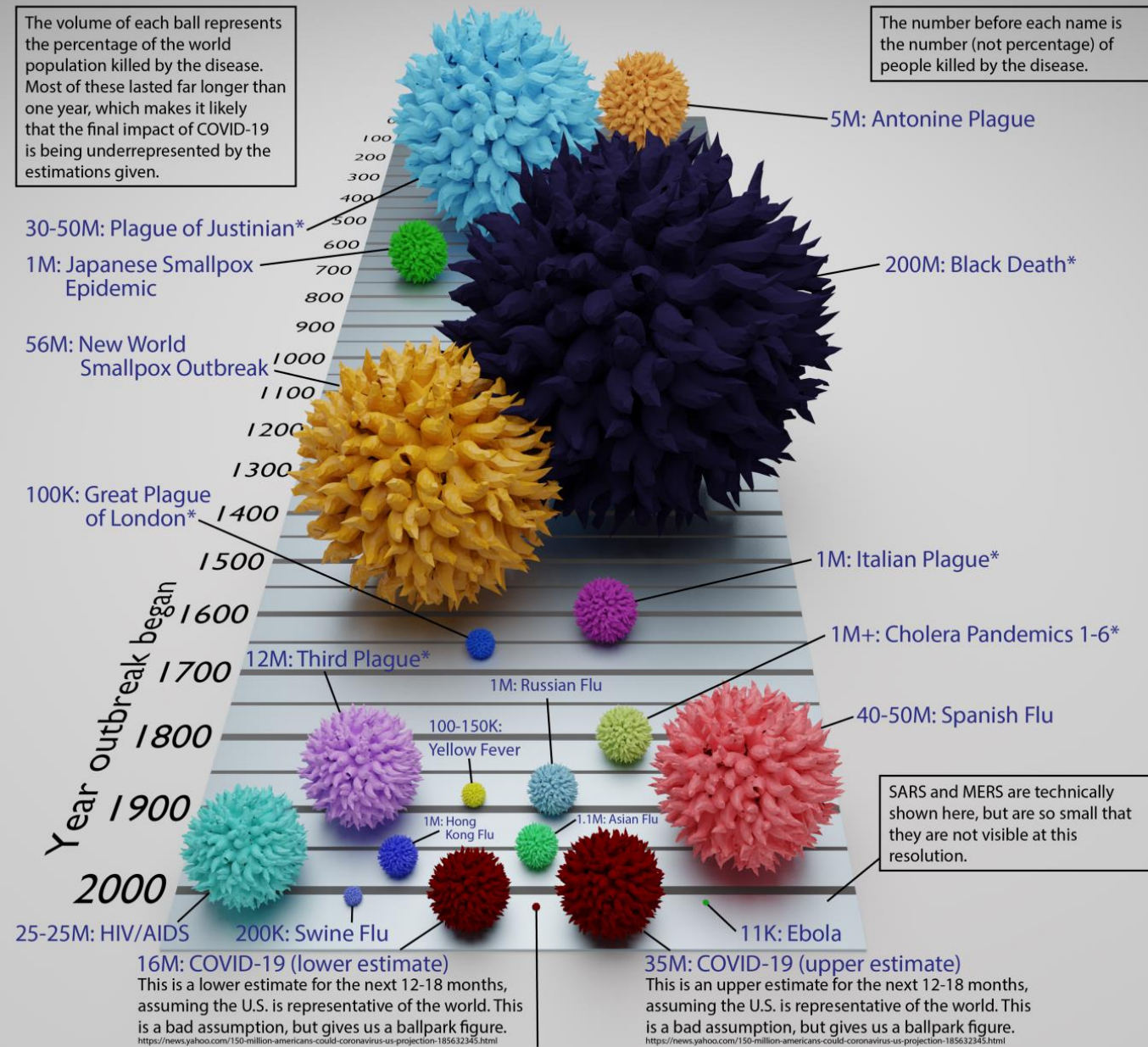
<https://coolinfographics.com/blog/2021/12/15/false-visualizations-getting-circles-sizes-wrong>



Misuse of Size (3D)

The volume of each ball represents the percentage of the world population killed by the disease. Most of these lasted far longer than one year, which makes it likely that the final impact of COVID-19 is being underrepresented by the estimations given.

The number before each name is the number (not percentage) of people killed by the disease.



“Death count of various pandemics as a ratio of world population” from r/dataisbeautiful
https://www.reddit.com/r/dataisbeautiful/comments/fp76db/death_count_of_various_pandemics_as_a_ratio_of/

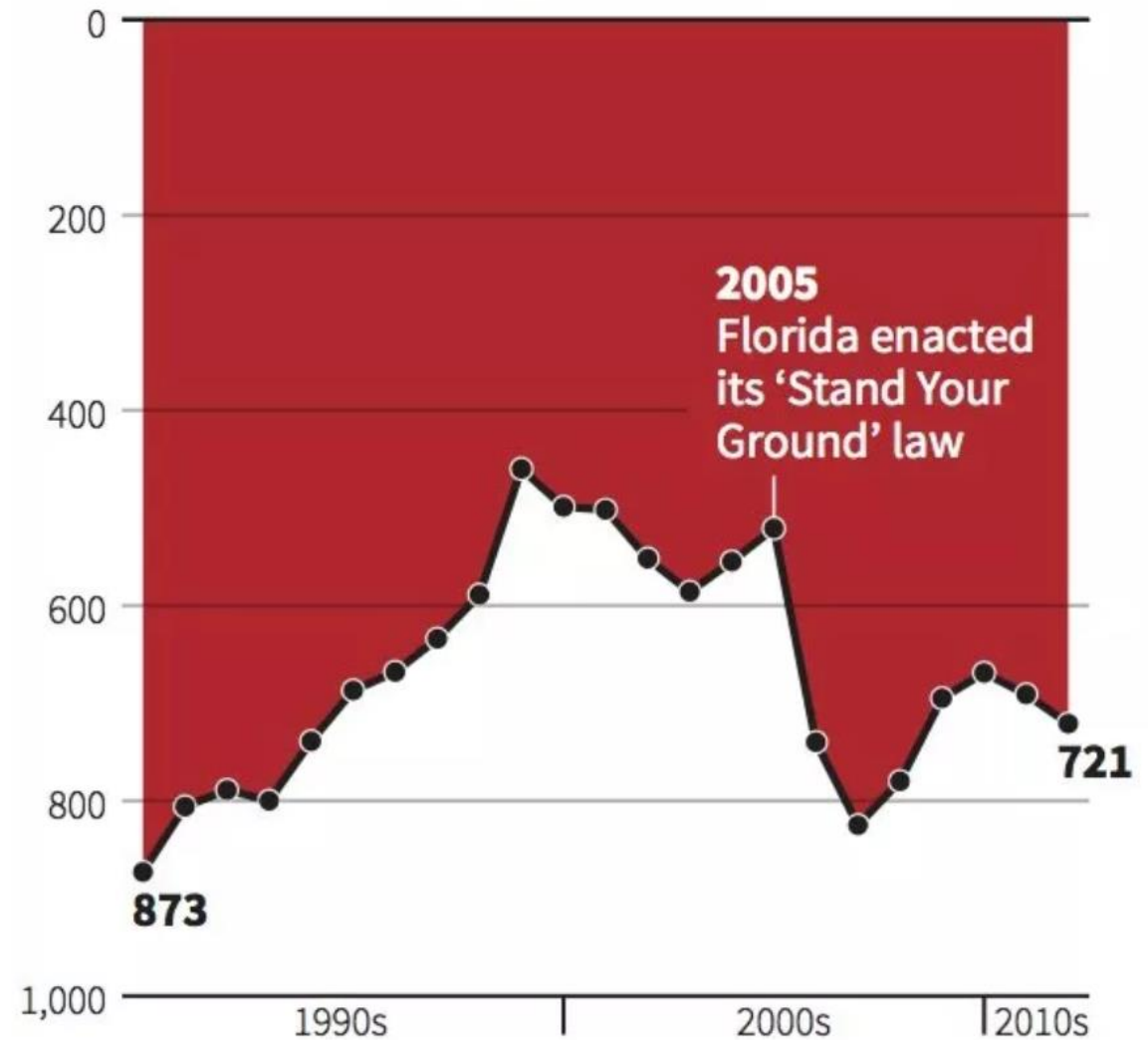
This visual is HEAVILY inspired by <https://www.visualcapitalist.com/history-of-pandemics-deadliest/>

*Caused by bacteria, not a virus

Abuse of Axes

Gun deaths in Florida

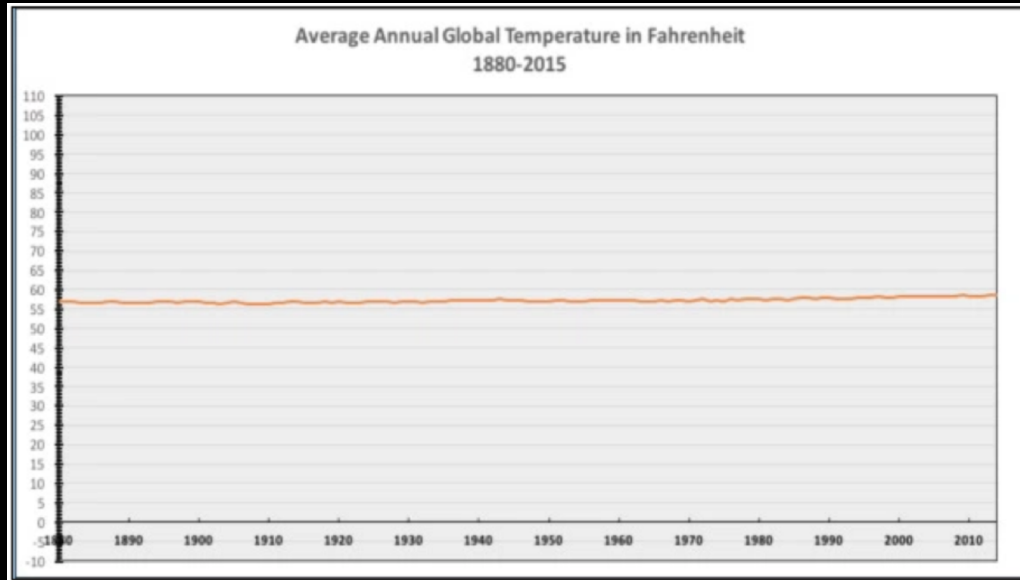
Number of murders committed using firearms



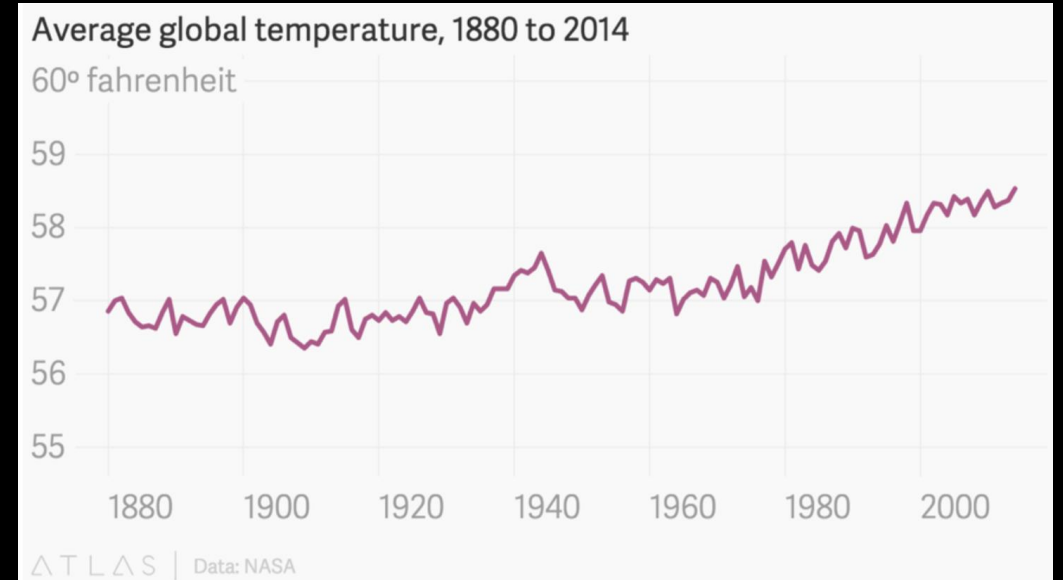
Source: Florida Department of Law Enforcement

Abuse of Axes

Scale Games



Source: National Review

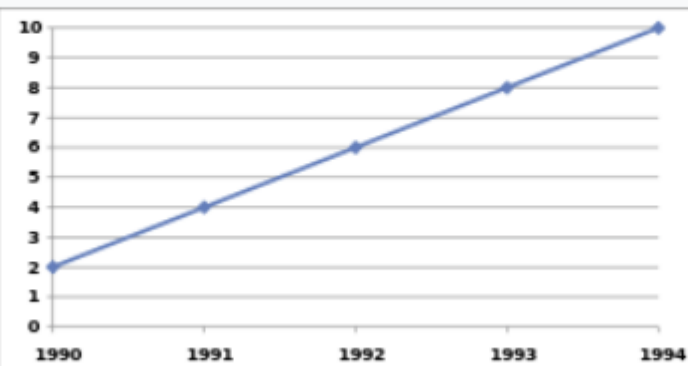


Source: Quartz

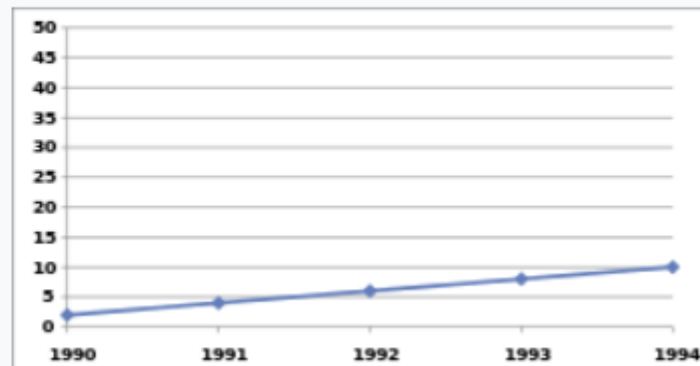
Original graph



Smaller maximum



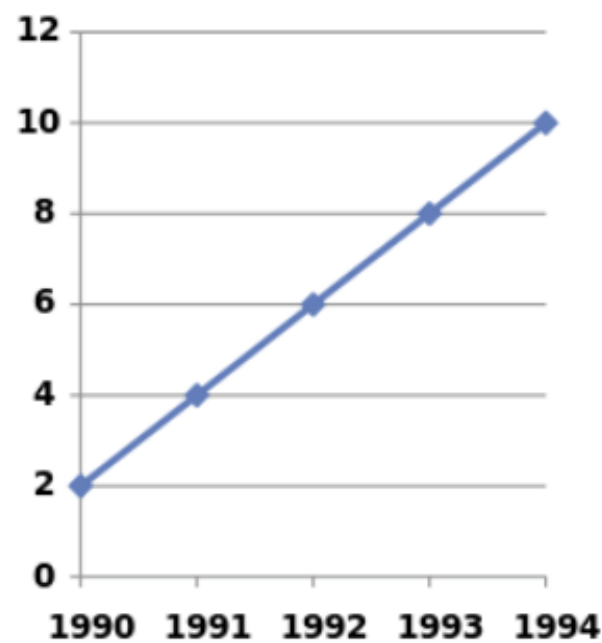
Larger maximum



Original graph



Half width, twice height



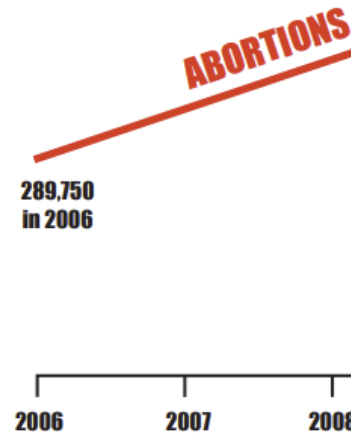
Twice width, half height



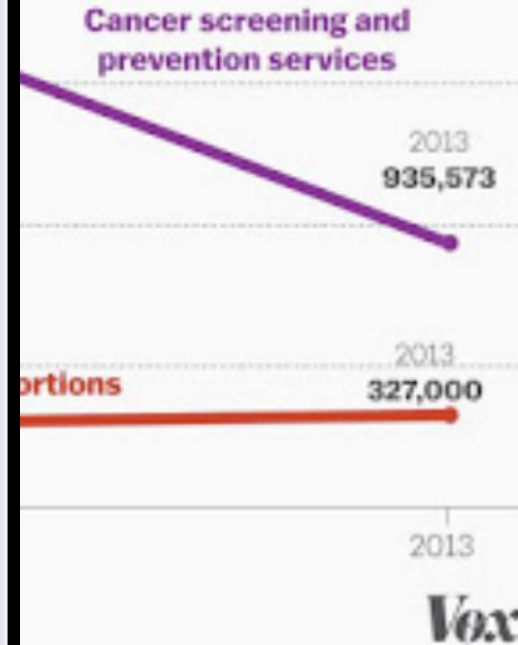
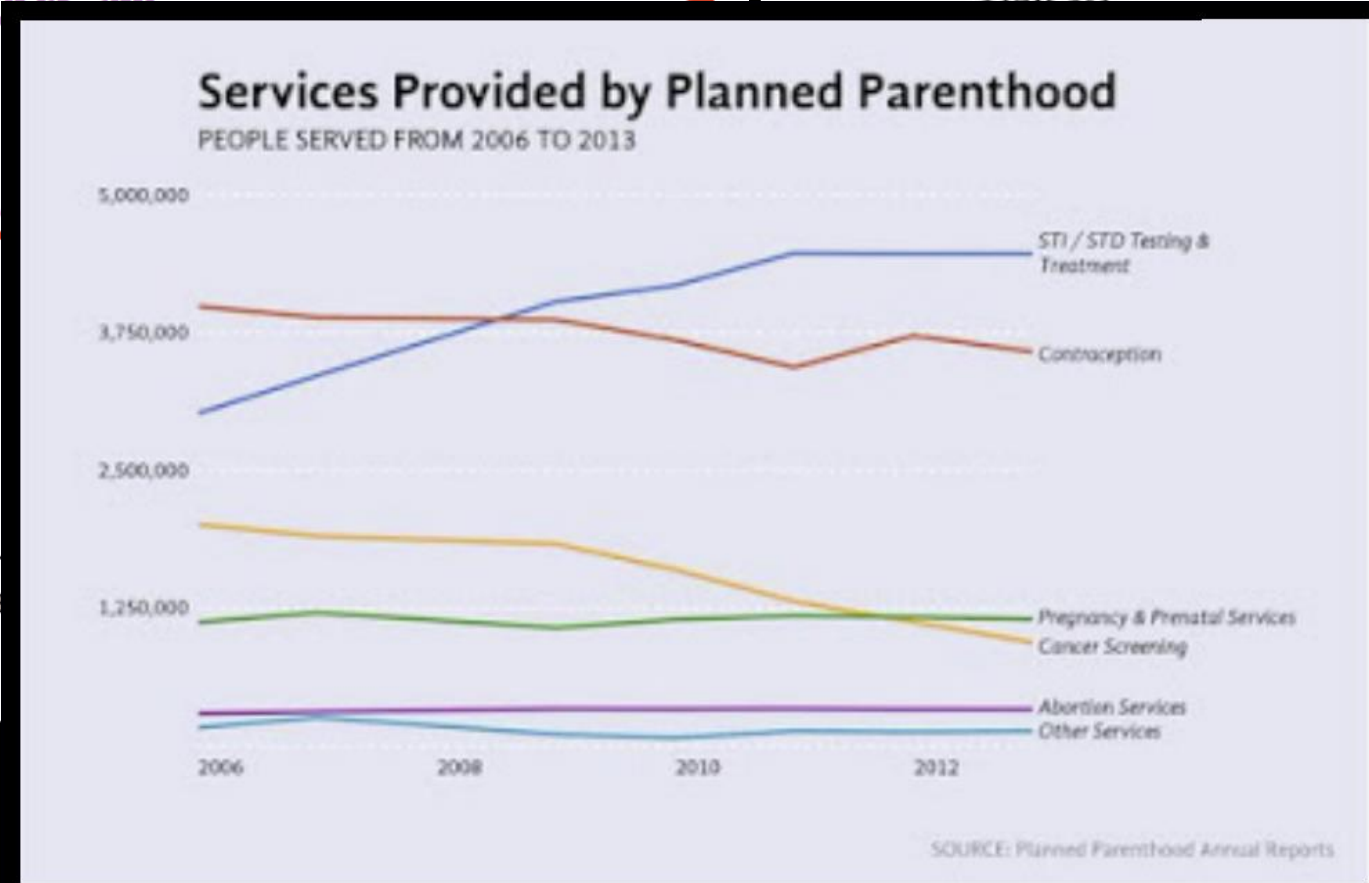
Vox - Shut up about the y-axis. It shouldn't always start at zero.



PLANNED PARENTHOOD FEDERATION OF AMERICA: ABORTIONS UP — LIFE-SAVING PROCEDURES DOWN

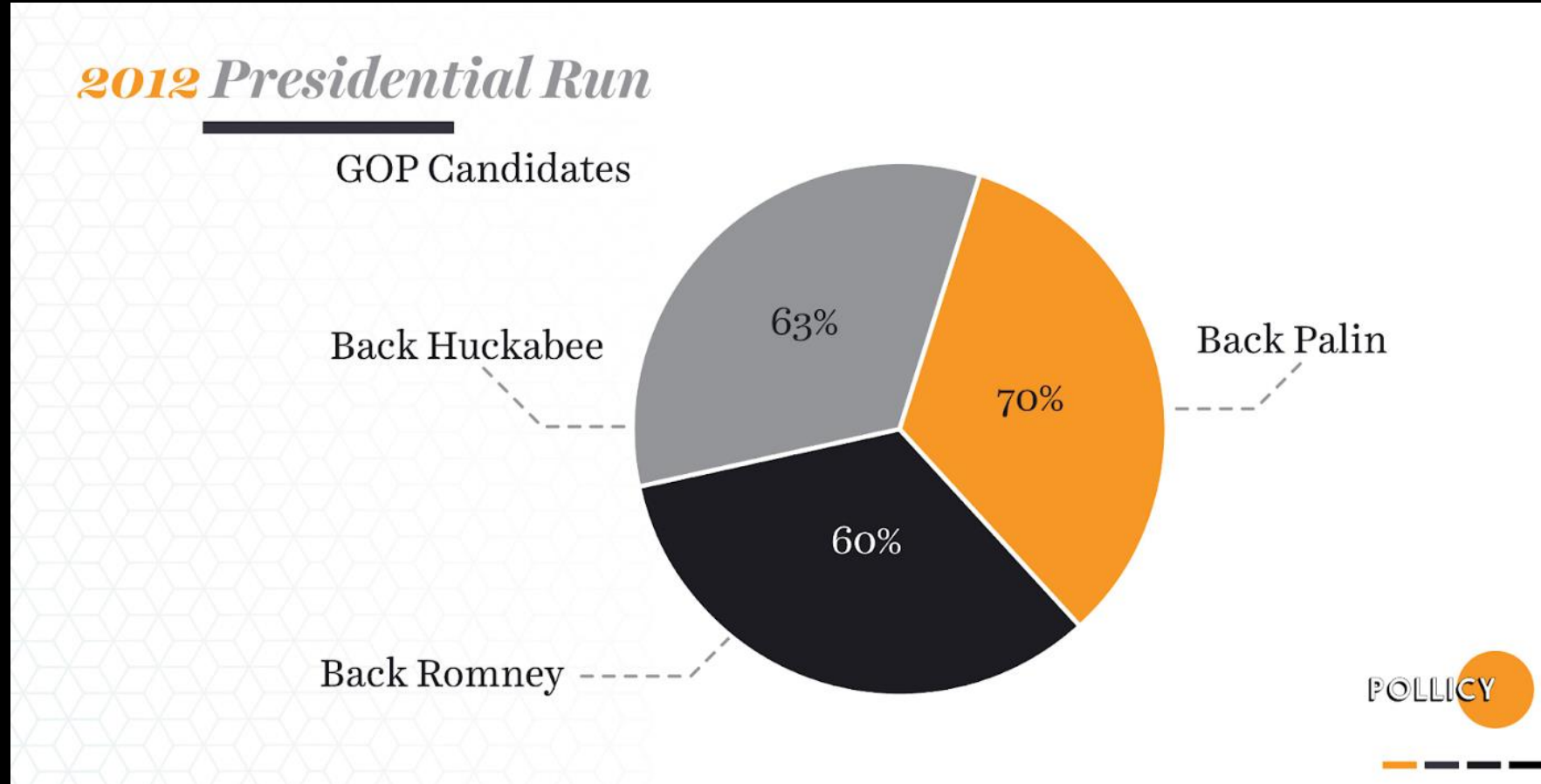


Services provided by Planned Parenthood



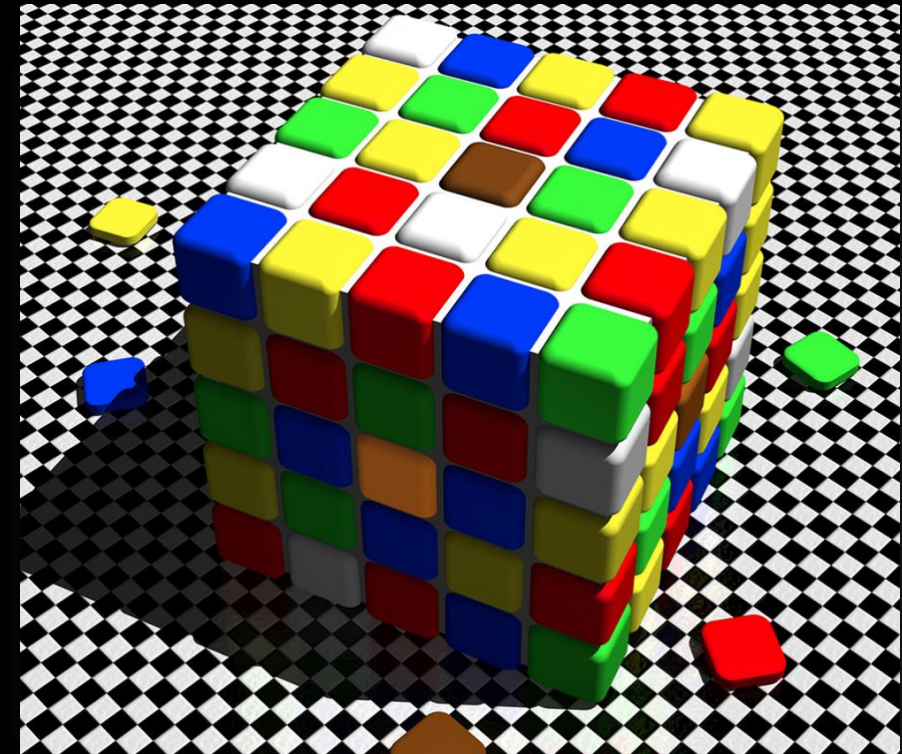
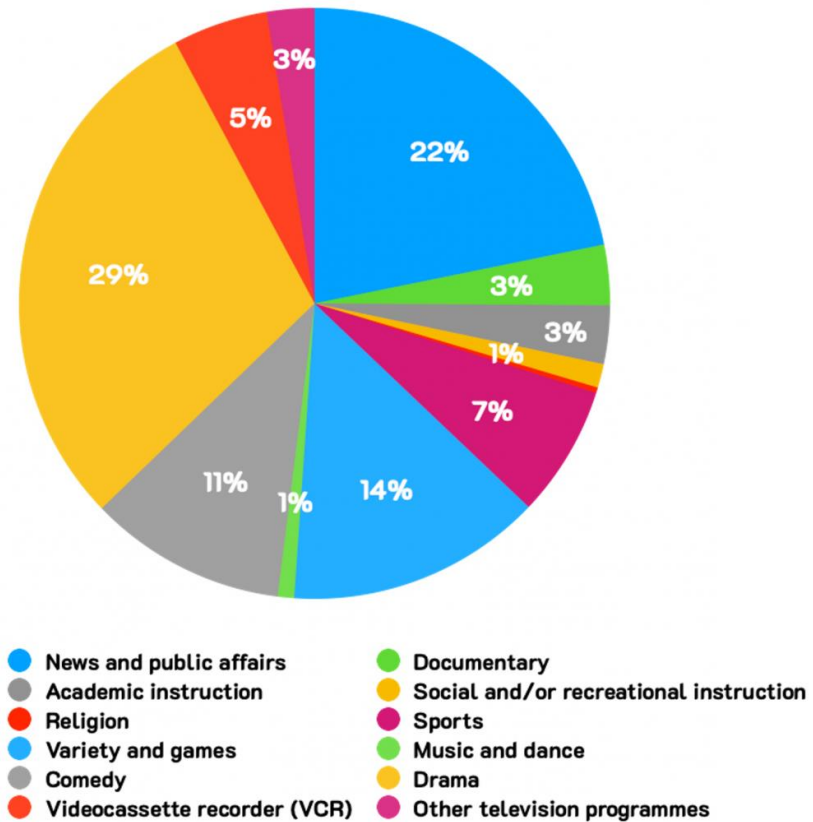
Graph by: Emily Schuch

Pie Charts That Don't Add Up

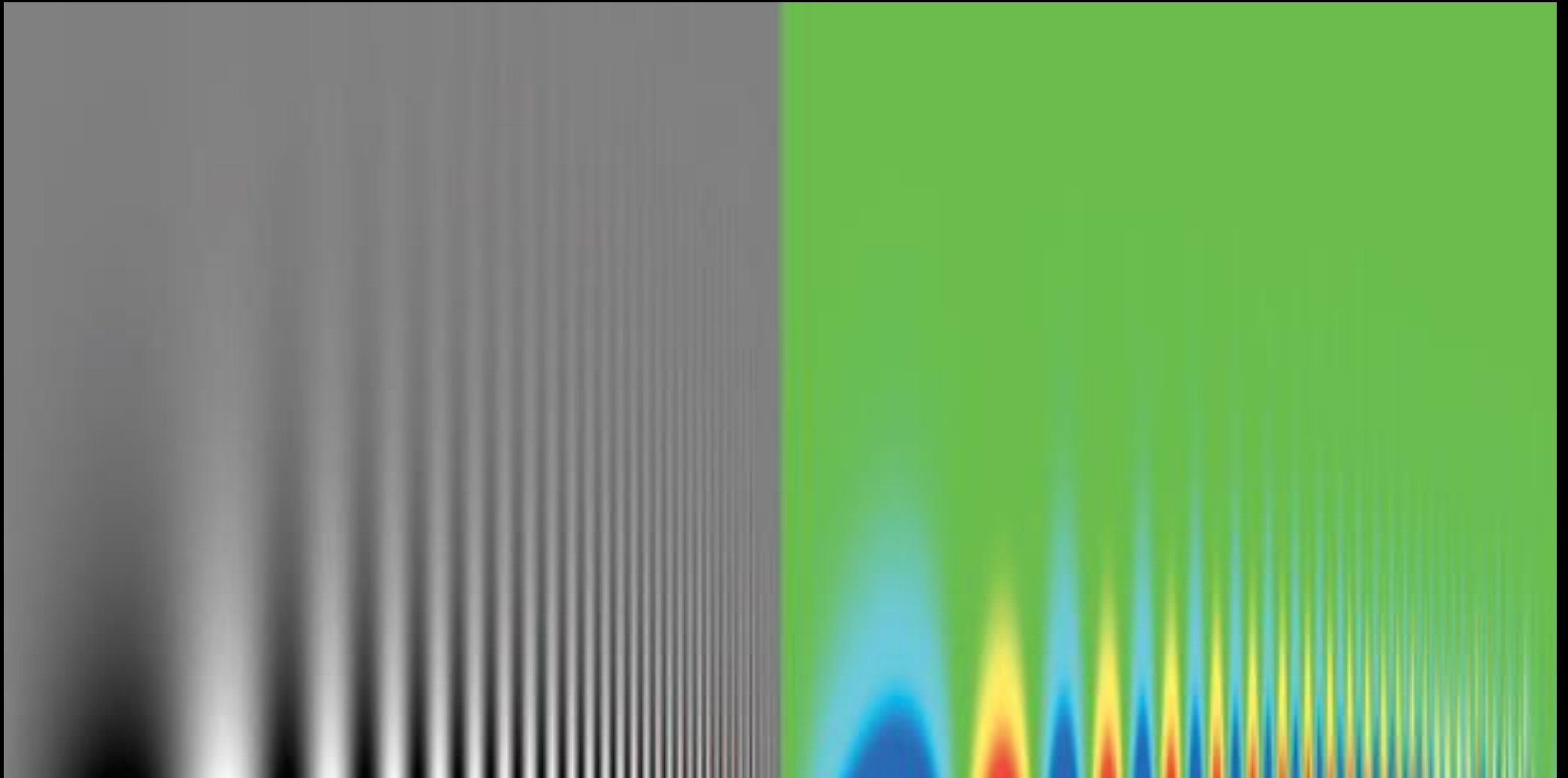


Misapplication of Colour

Ontario Television Viewing in 2004

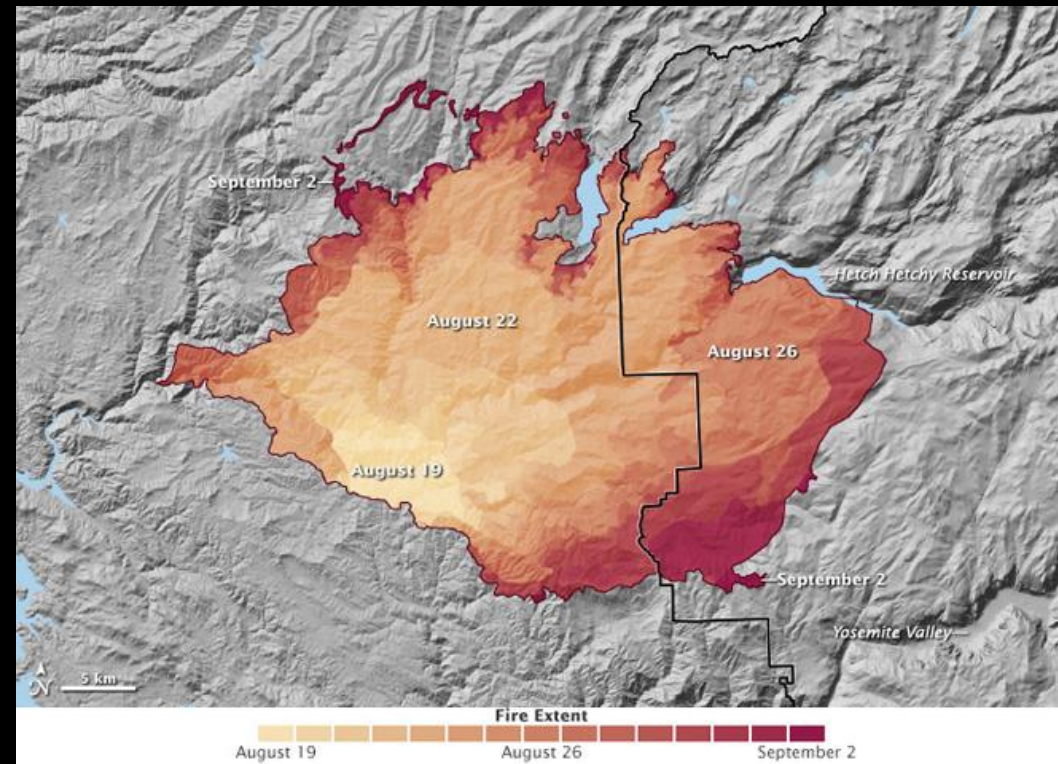
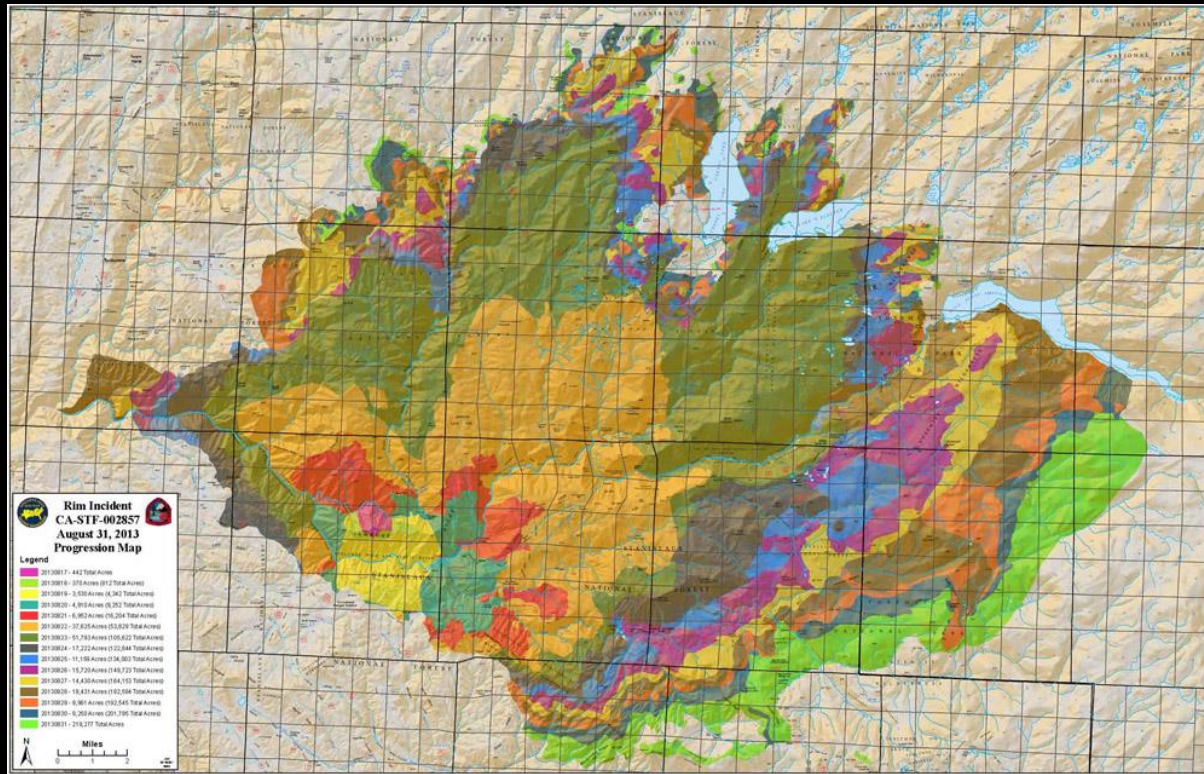


Misapplication of Colour

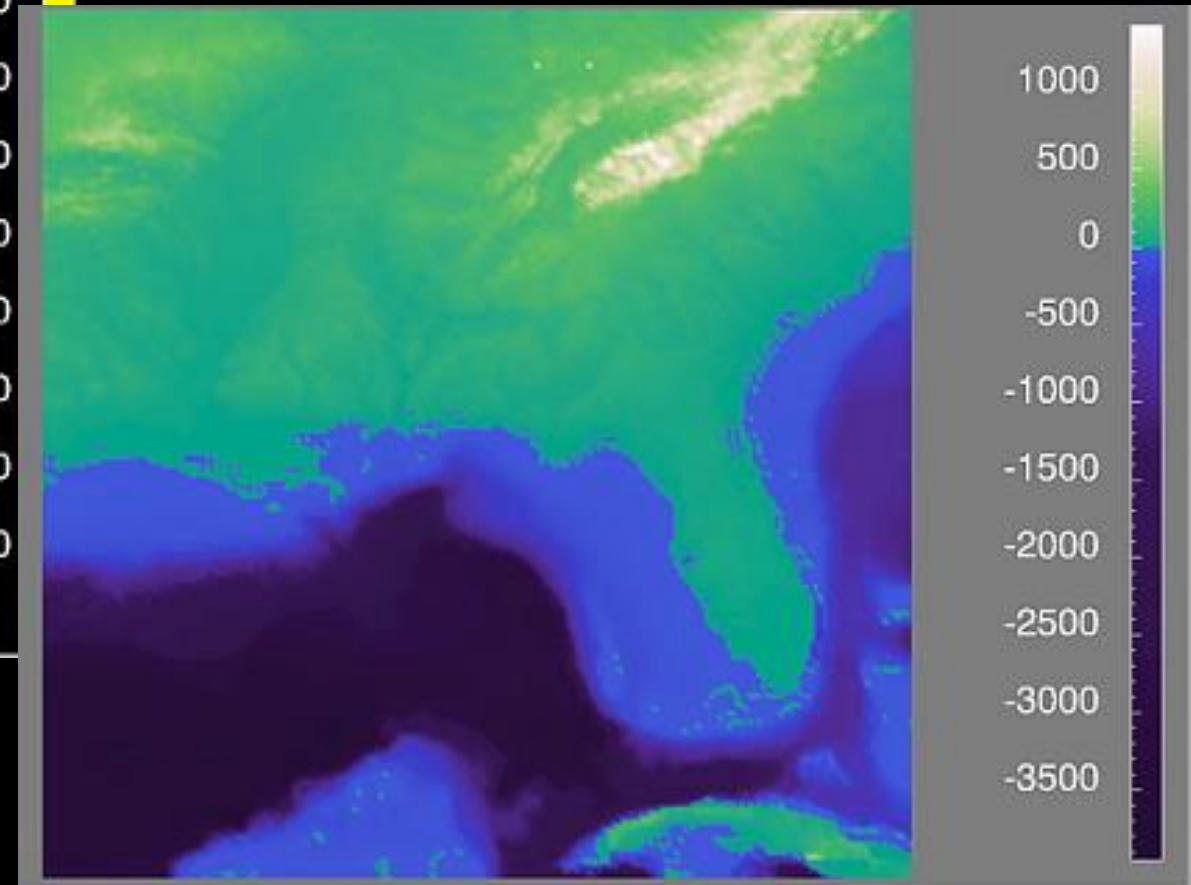
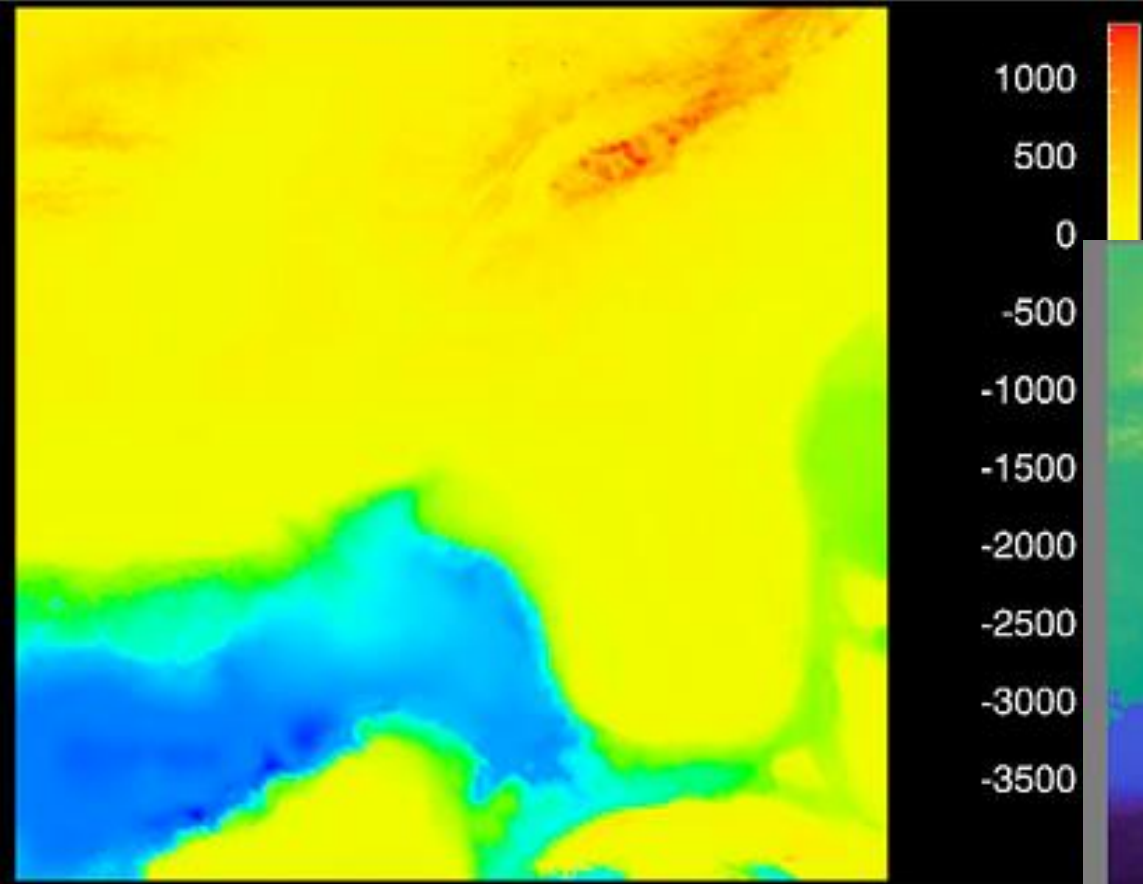


From *Borland & Taylor II (2007), Rainbow Color Map (Still) Considered Harmful. IEEE Visualization Viewpoints, pp 14-17.*

Misapplication of Colour



Misapplication of Colour





How to avoid being mislead

How to spot misinformation in visualization?



Is there a listed data source?

How was the data collected?
Have the authors limited the data's range?
What was not measured?



Consider the motivation of the person creating the visual



Are there signs of poor design?



Read the text

captions
labels
axes



Go back to the data and create your own visual



Slides: <https://brosz.ca/slides/>

Questions

Julia Guy

JULIA.GUY@UCALGARY.CA

<https://sands.ucalgary.ca/>

John Brosz

JDLBROSZ@UCALGARY.CA

<https://library.ucalgary.ca/visualization>

