

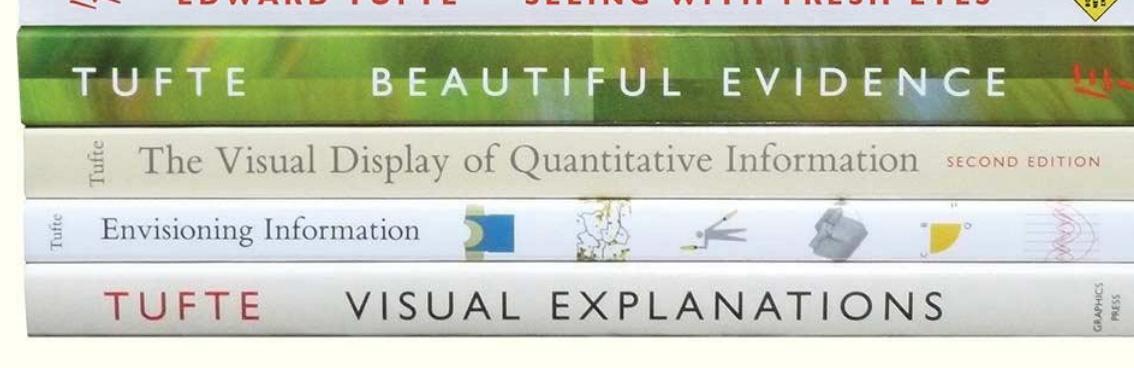
Principles of Effective Visualizations

Principle	Definition	Examples
Proportional Ink	The amount of ink used to indicate a value should be proportional to the value itself	Truncating a position channel on a bar chart to exaggerate the difference between bars violates the principle of proportional ink.
Data:ink ratio	Remove distracting visual elements to focus attention on the data	Lighten line weights, remove backgrounds, never use 3D or special effects, avoid unnecessary/redundant labels.
Labels & legends	Use axes labels and titles to highlight/ communicate data	Never leave your data column names as axes labels! Generally good to add a title.
Colour & Accessibility	Colour can be used to encode information or for aesthetics/style/ design. However, colour can also be distracting if used inappropriately or poorly.	Choose a perceptually uniform colour palette; can be sequential or diverging for quantitative data. Opt for colour-blind friendly palettes. Categorical data can use qualitative colour schemes.

Principles of Effective Visualizations

Principle	Definition	Examples
Task-Encoding Alignment	Start with focus on functionality; aesthetics can be improved later as refinement	Using embellished charts for wow factor and memorability as opposed to using plain bar chart for statisticians. Both valid, depends on function
Eyes Beat Memory	External cognition vs. internal memory - easier to compare by moving eyes between side-by-side views than comparing visible item to memory	Small multiples (showing time with space) vs. animation (showing time with time); fertility rate comparisons across countries
Justified Dimensionality	Use 3D only when the task requires spatial shape perception. For abstract data, 2D is nearly always superior due to depth perception limits, occlusion, and distortion.	Justified: Molecular structures, anatomical models, fluid flow visualization, terrain maps Unjustified: Bar charts, pie charts, abstract networks, non-spatial data
Scale Honesty	Plots that display two related datasets using separate y-axes on the same x-axis; useful for compact comparisons but prone to scale confusion, visual bias, and misleading correlations, so they should be used only when clearly justified	Clear labeling required, consistent scales preferred, purposeful use only

Tufte Principles



1. **Micro/Macro Readability:** Present data clearly at both detailed and overall levels to ensure comprehensive understanding.
2. **Small Multiples:** Employ series of similar graphs with consistent scales for easy comparison and to highlight patterns over time.
3. **Avoiding Chartjunk:** Remove unnecessary decorative elements, complex grid lines, and extraneous labels to enhance clarity.
4. **Narrative Integrity:** Maintain honesty and accuracy in data representation to preserve the story's integrity and prevent misleading viewers.

Advice from Tufte

- **Narrative Integrity:** Prioritize honesty and accuracy in data representation to maintain the story's integrity and avoid misleading viewers.
- **Layering and Separation:** Use design elements like color, shading, and positioning for clarity, helping viewers easily differentiate various layers or aspects of the data.

Visualization for Data Science

Design Missteps



Except when otherwise stated on the slide, the images come from
Alberto Cairo's book *How Charts Lie*

Administrivia

- Course Epilogue on Wednesday
 - Time for Course Feedback (please complete the feedback survey before class so we can have more time for review)
 - Review
 - Open Questions in Viz
 - Final Content & How to Study
- Office Hours This Week
 - Monday after class
 - Tuesday (Dr. K) 2 – 3 pm
 - Tuesday from 3:30 – 6 in the Lab room (with Tas)
- Due on Thursday
 - 6pm, EVERYTHING.

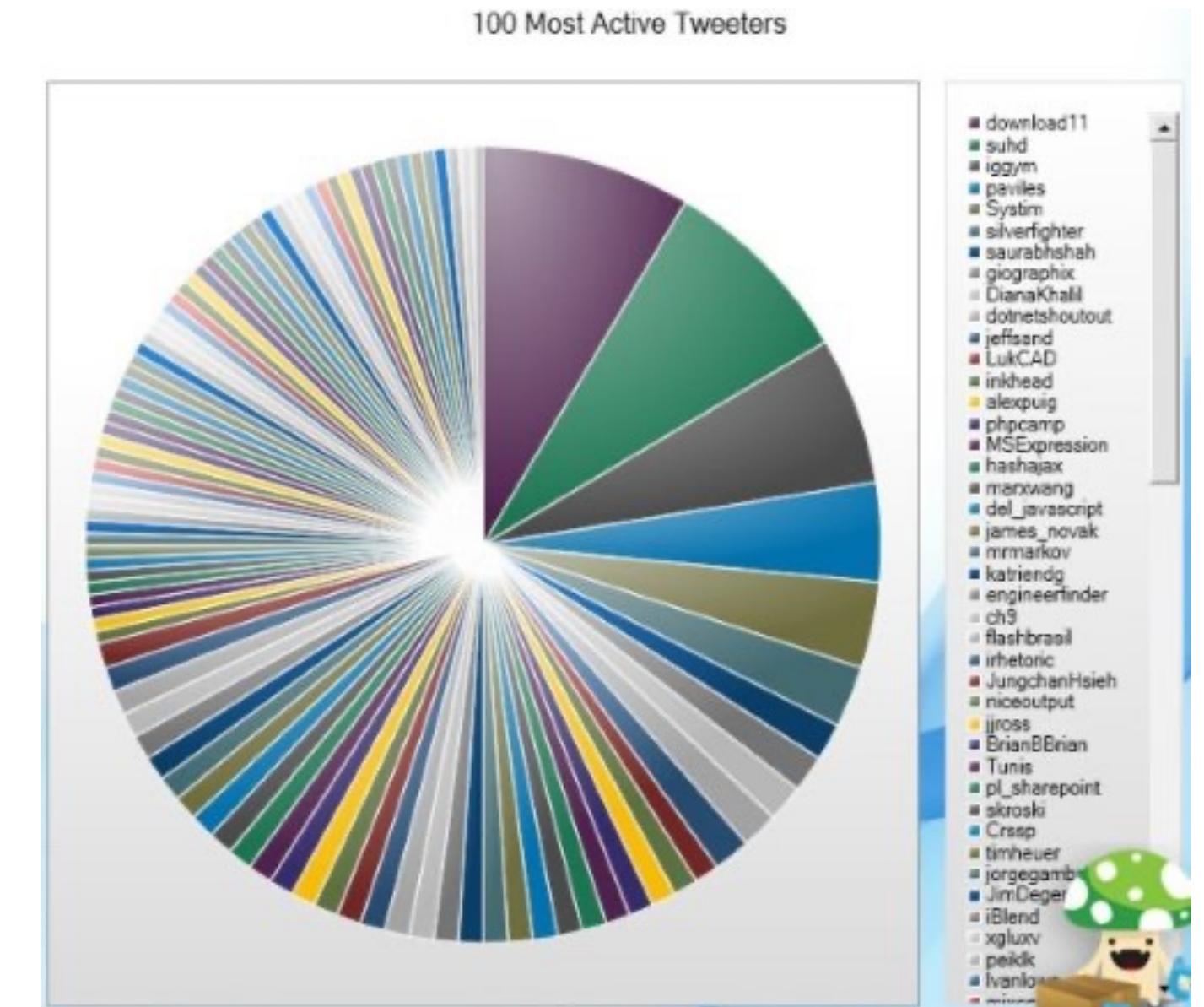
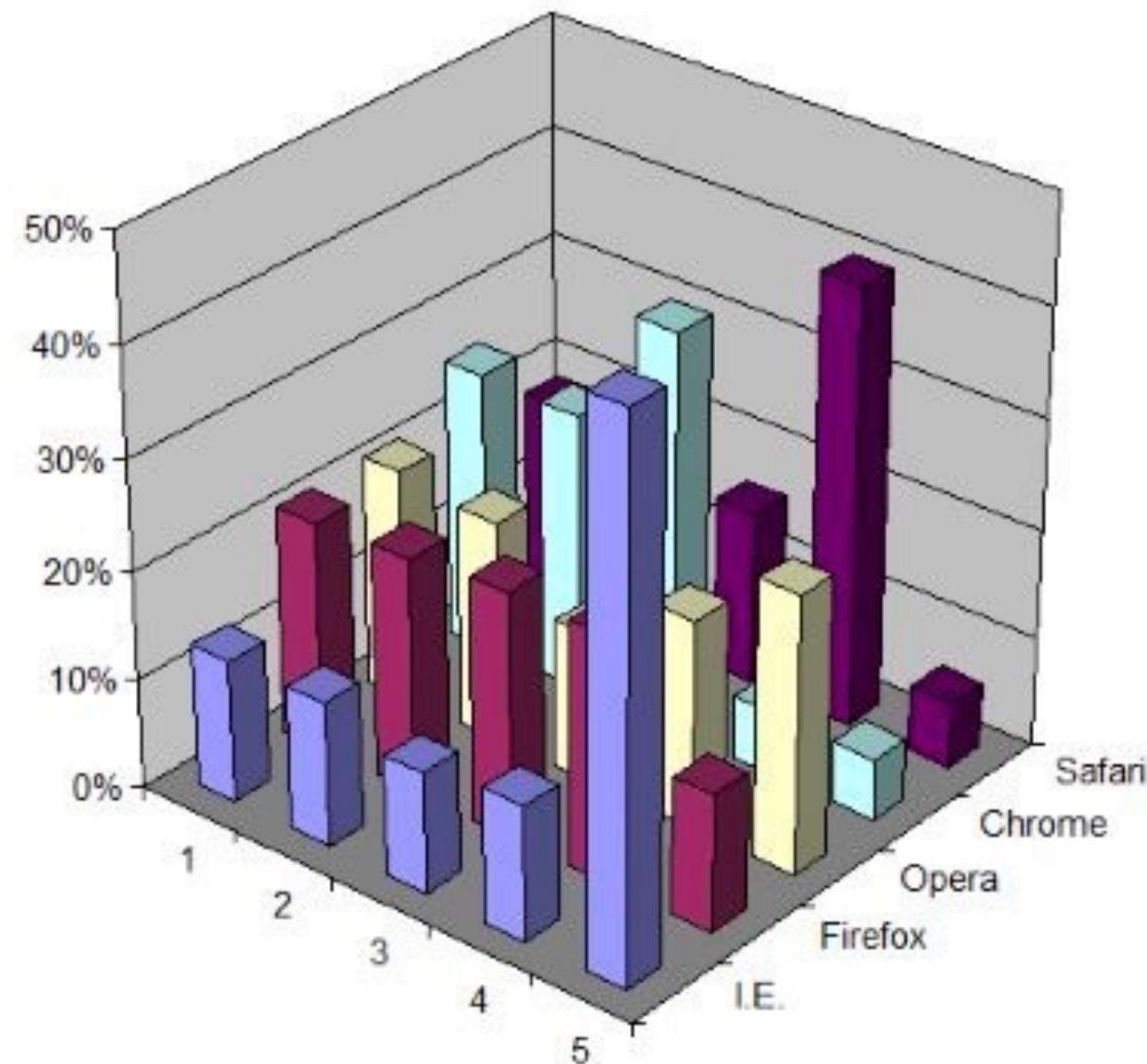
Bad Viz

- Incomprehensible Visualizations
- Scale Manipulation
- Color Misuse
- Data Mapping Errors
- Data and Statistical ~~Deception~~ Issues

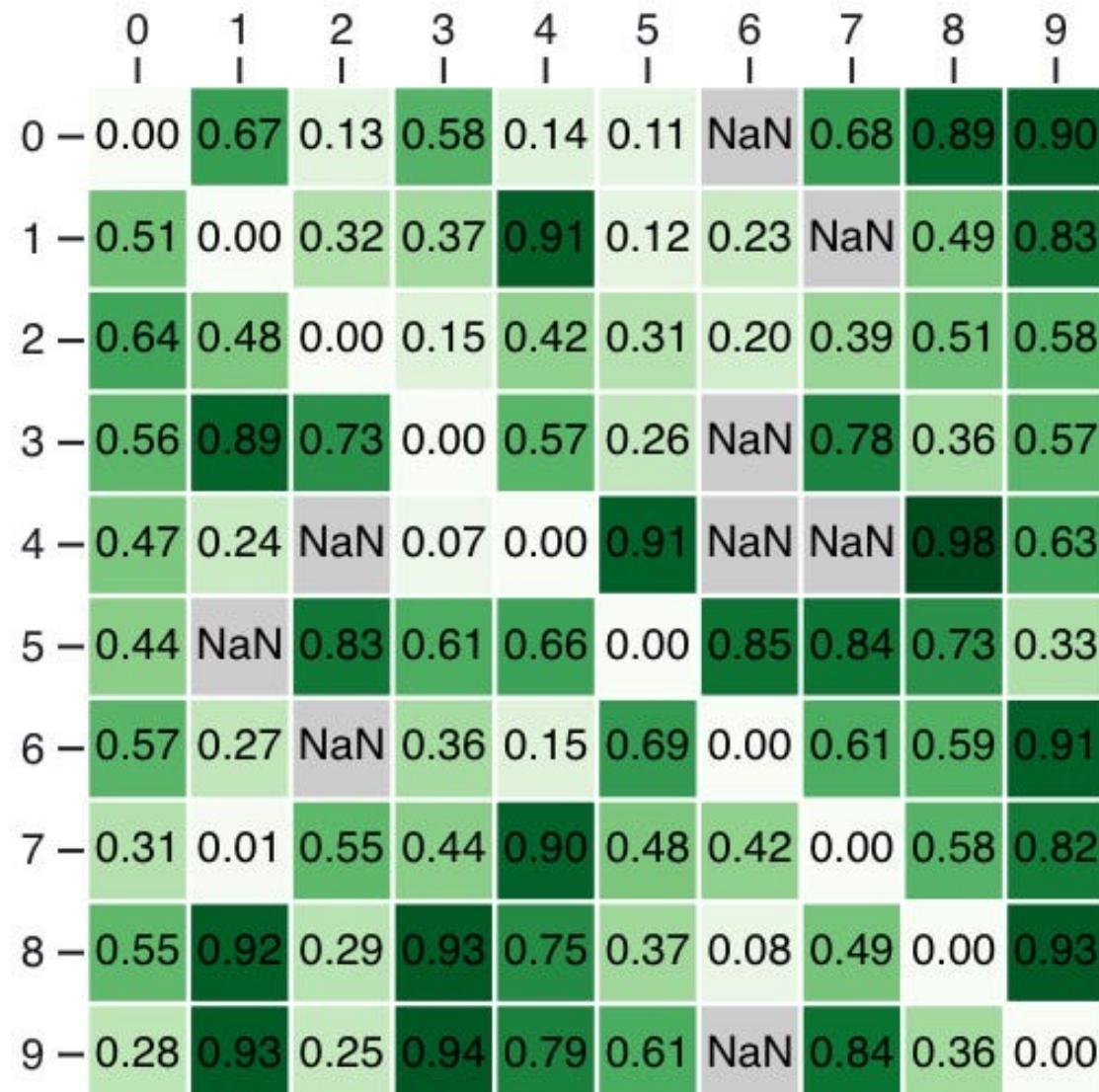
Incomprehensible Visualizations

- **Encoding Overload:** Too many visual channels used simultaneously
- **Chart Junk:** Excessive decoration obscuring data (Tufte's concept)
- **Over-complexity:** Inappropriate chart type for the data/task
- **Poor Layout:** Cluttered arrangement preventing comprehension
- **Illegibility:** Text too small, overlapping labels, poor contrast

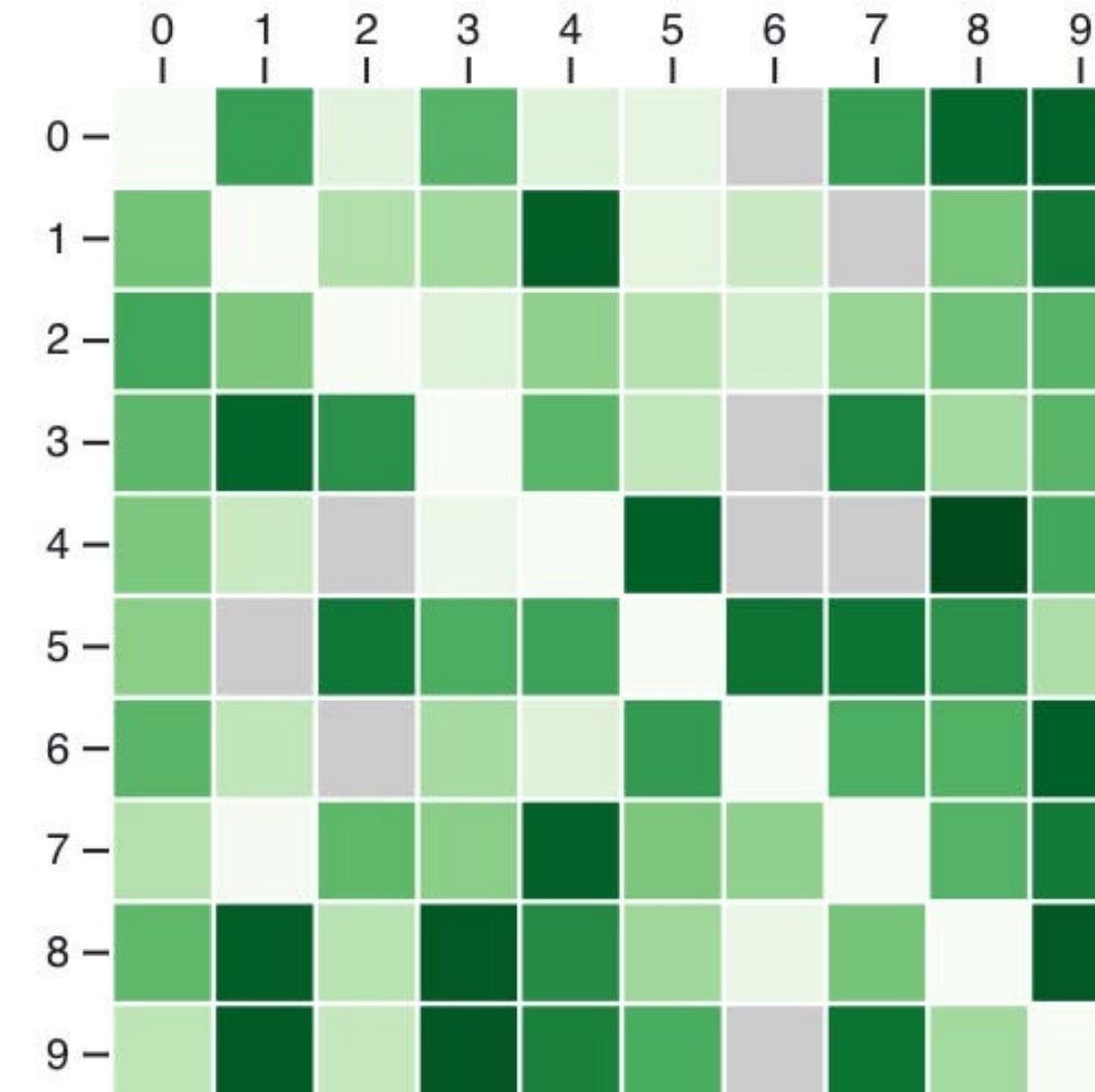
Encoding Overload – Too much data, too many encoded items



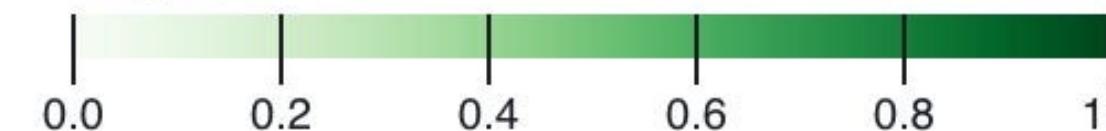
Visual Clutter



Weight



Weight

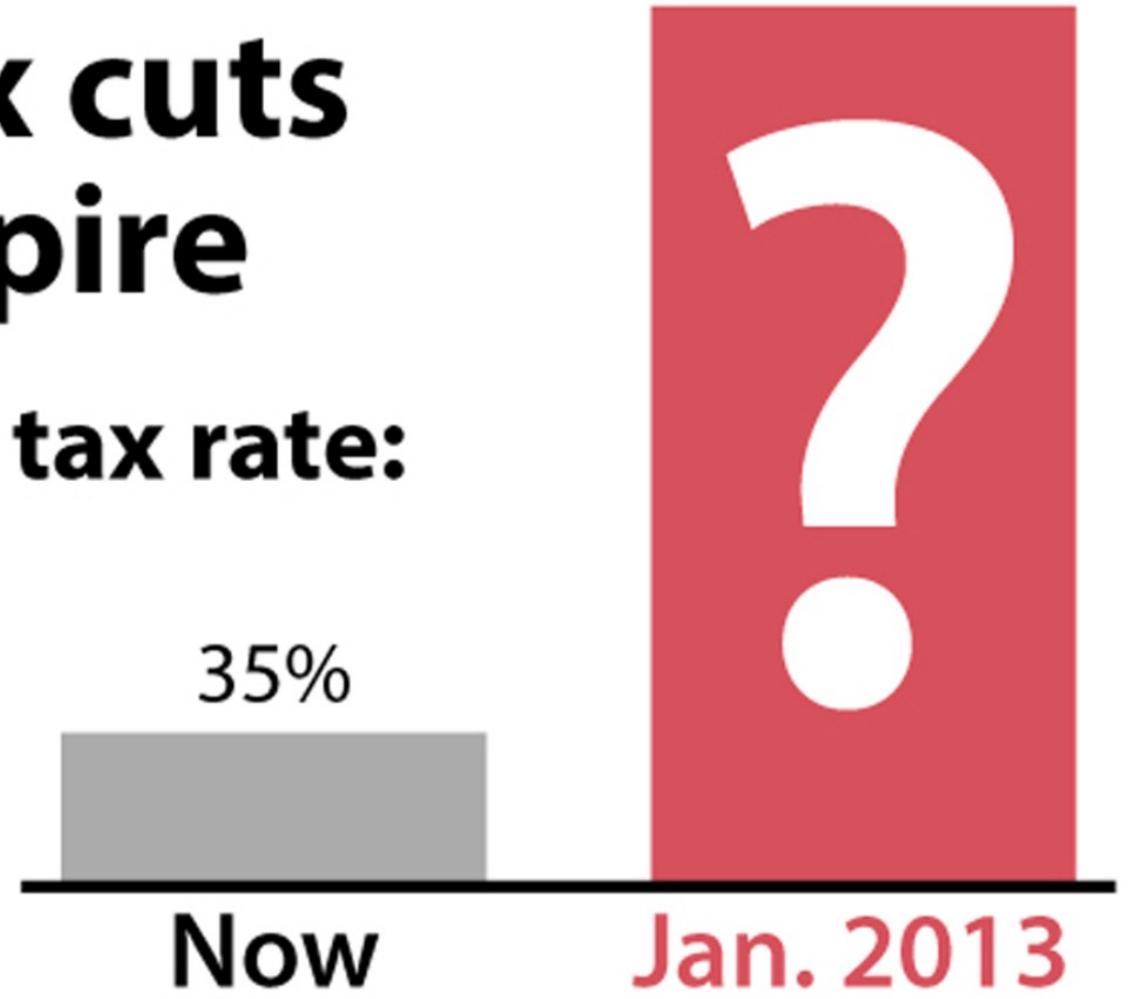


Scale Manipulation

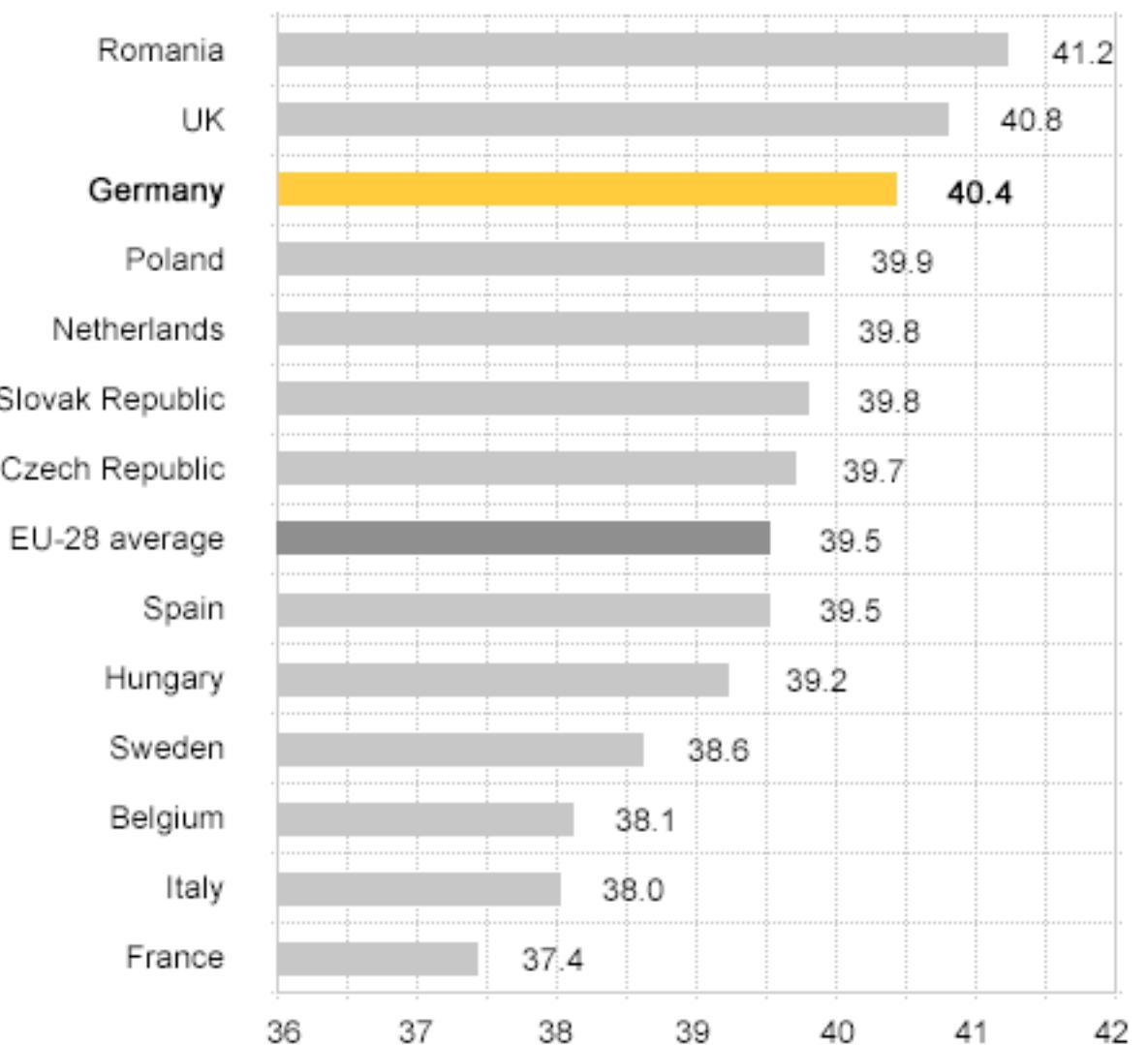
- Truncated Y-Axis: Not starting at zero to exaggerate differences
- Inconsistent Scales: Different scales across comparable charts
- Dual-Axis Deception: Manipulating scales to manufacture correlation
- Non-Linear Scales: Using log/exponential scales without disclosure
- Broken Axes: Hiding data gaps or manipulating visual spacing

**If Bush
tax cuts
expire**

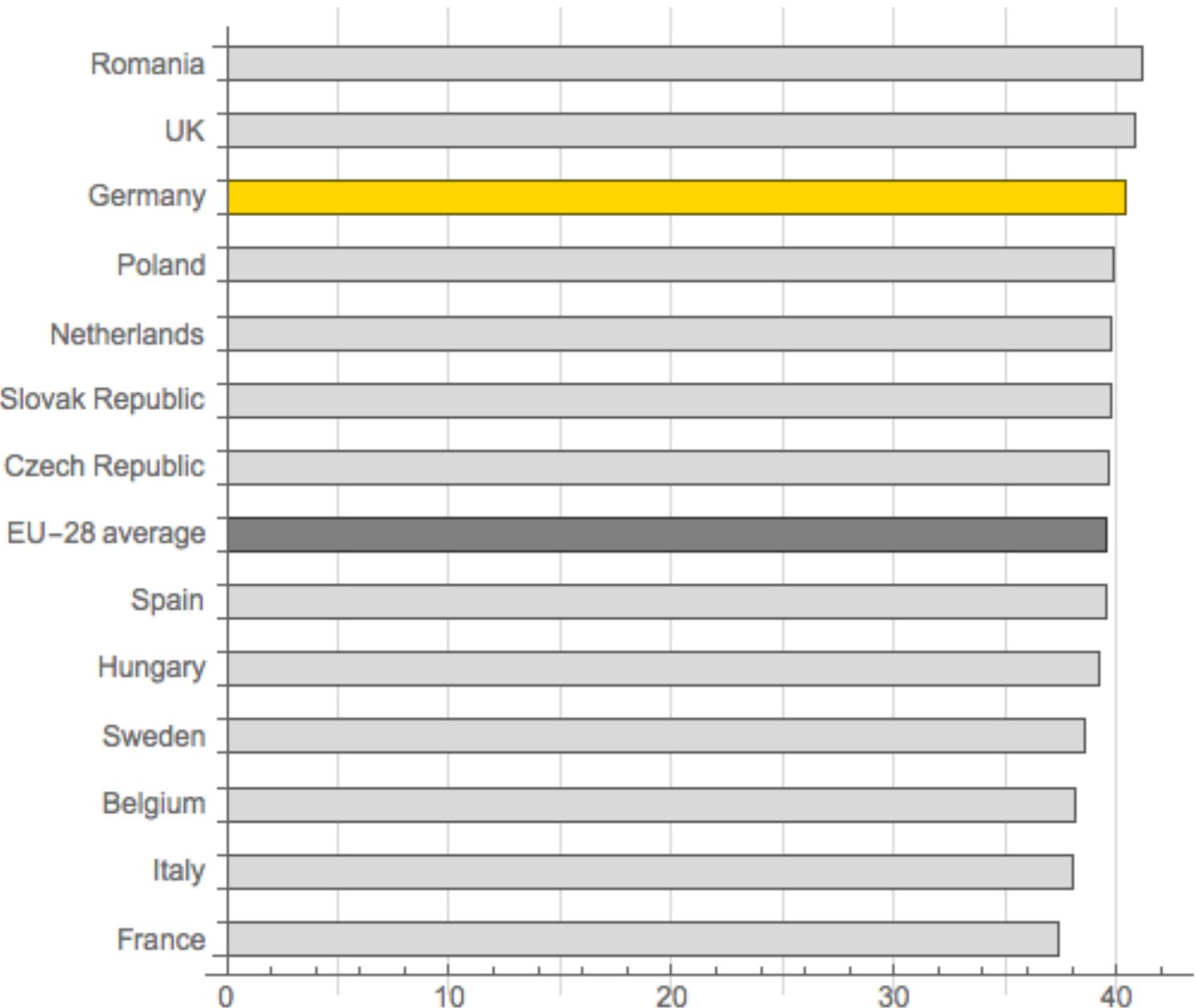
Top tax rate:



Average number of actual weekly hours of work in main job, full-time employees, 2013



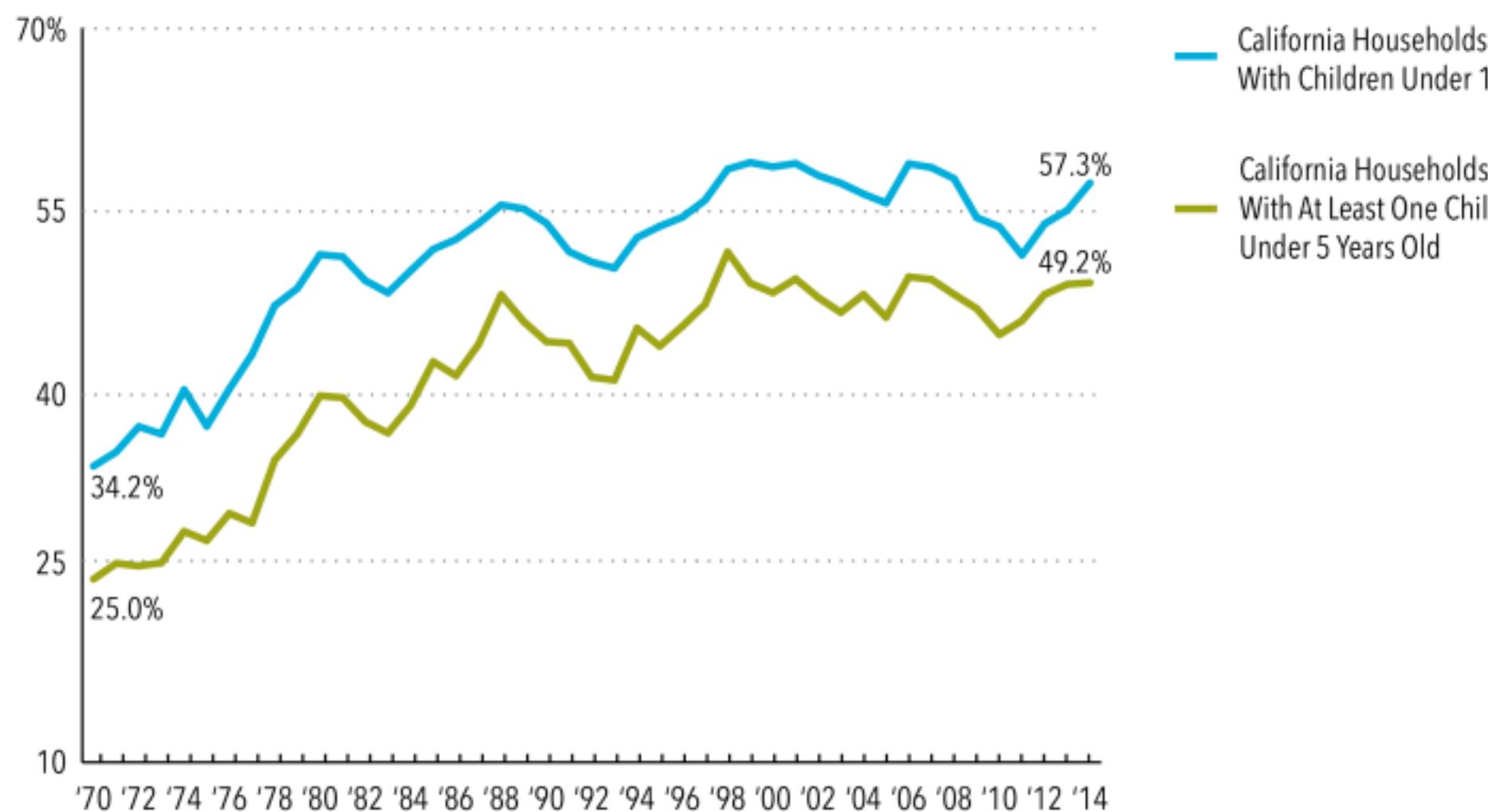
Source: Eurofound 2014



So the moral of the story
is that we should always
start the y-axis at 0

More California Households Have All Parents Working, Making Access to Child Care an Important Priority

Percentage of California Households Where All Parents Work, 1970 to 2014



Note: A "household where all parents work" includes single-parent households and dual-earner households. Parents include stepparents and adoptive parents.
Source: Budget Center analysis of US Census Bureau data

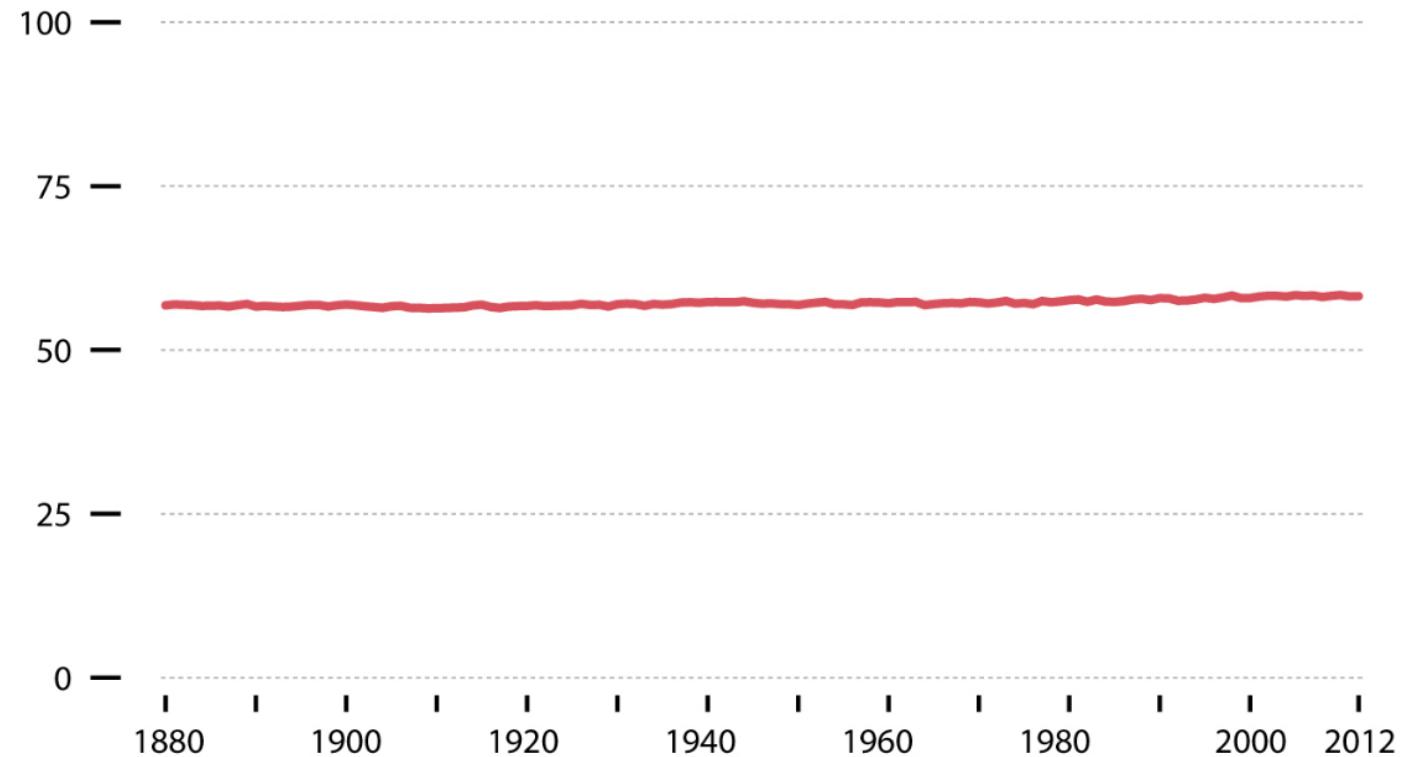


California Budget
& Policy Center
Independent Analysis. Shared Prosperity.

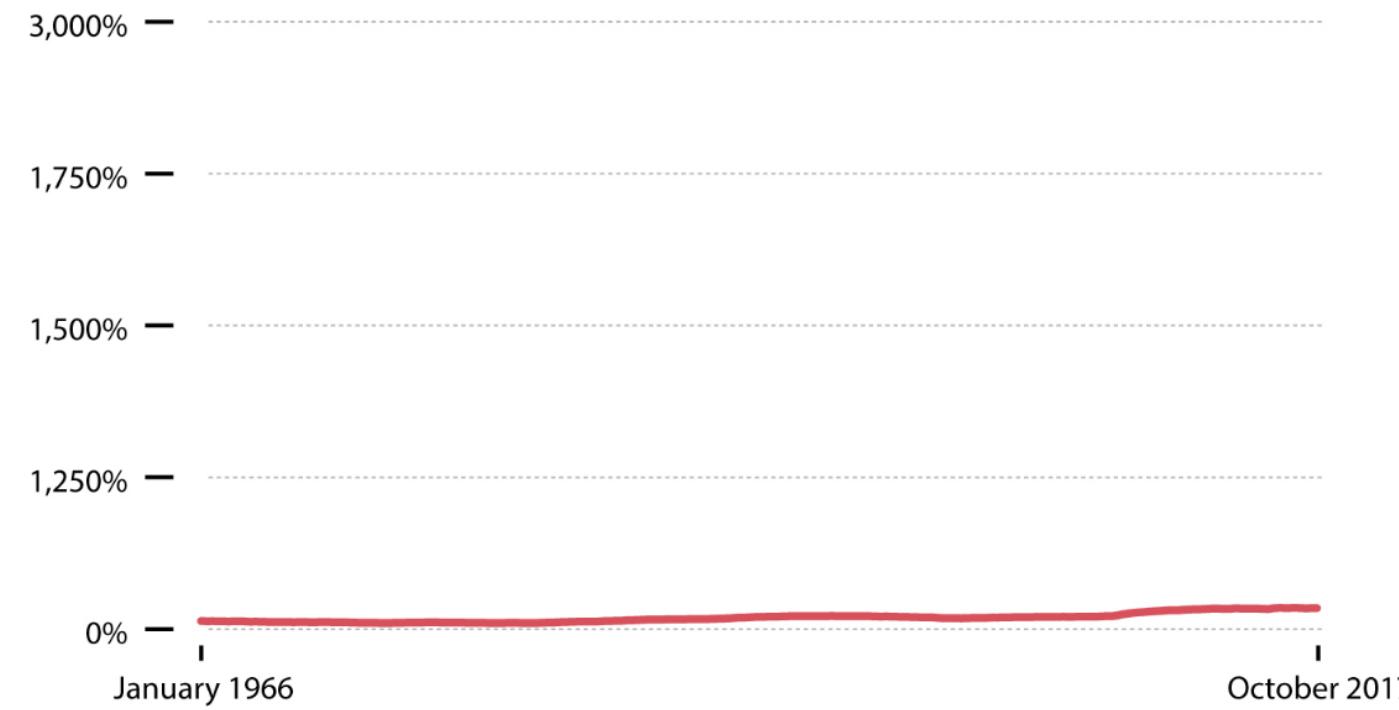
Zero-baseline Scale?



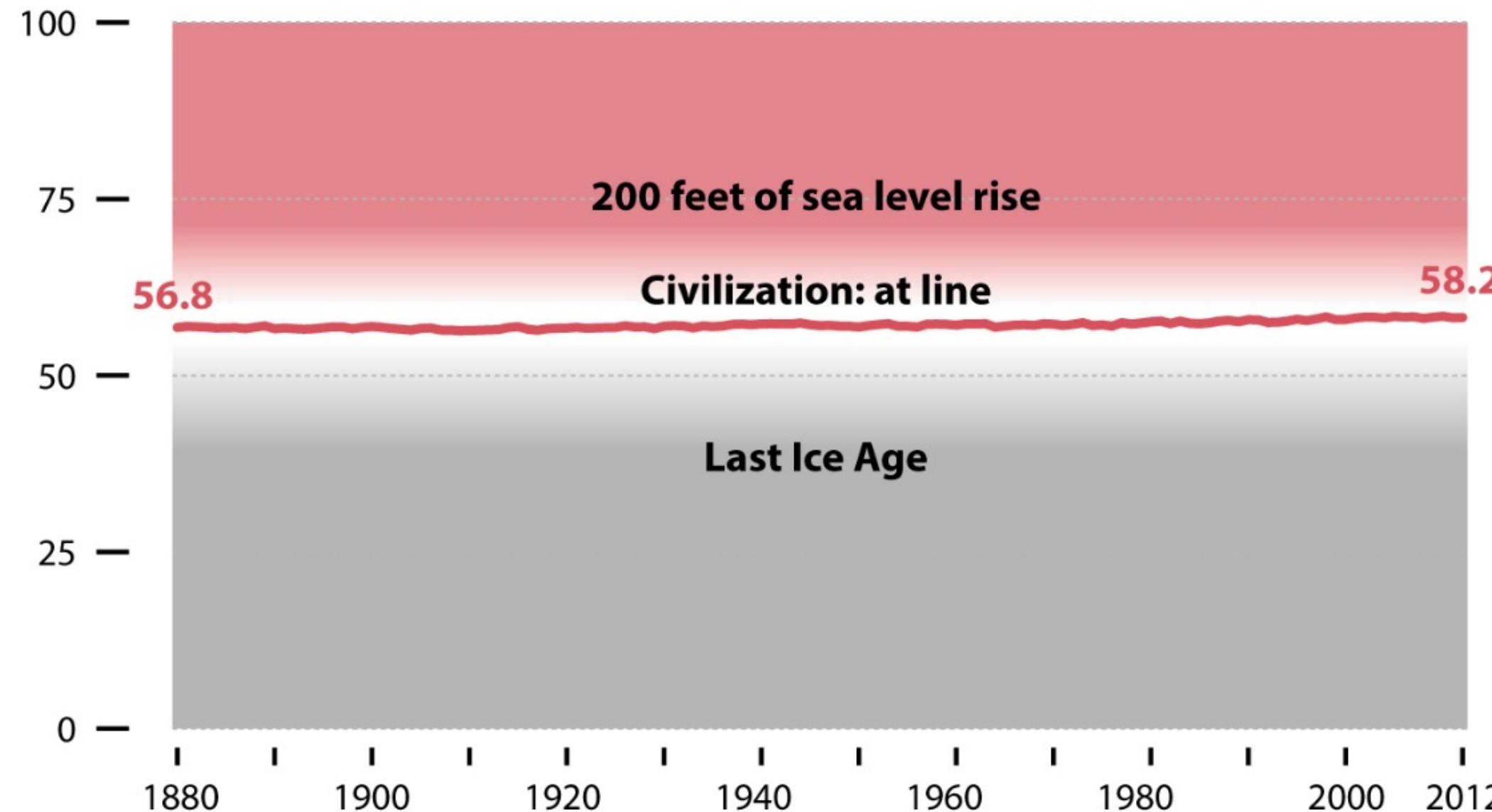
Average annual global temperature in degrees Fahrenheit



Federal debt: total public debt as percent of gross domestic product



Average annual global temperature in degrees Fahrenheit



Average annual global temperature in degrees Fahrenheit

59 —

58 —

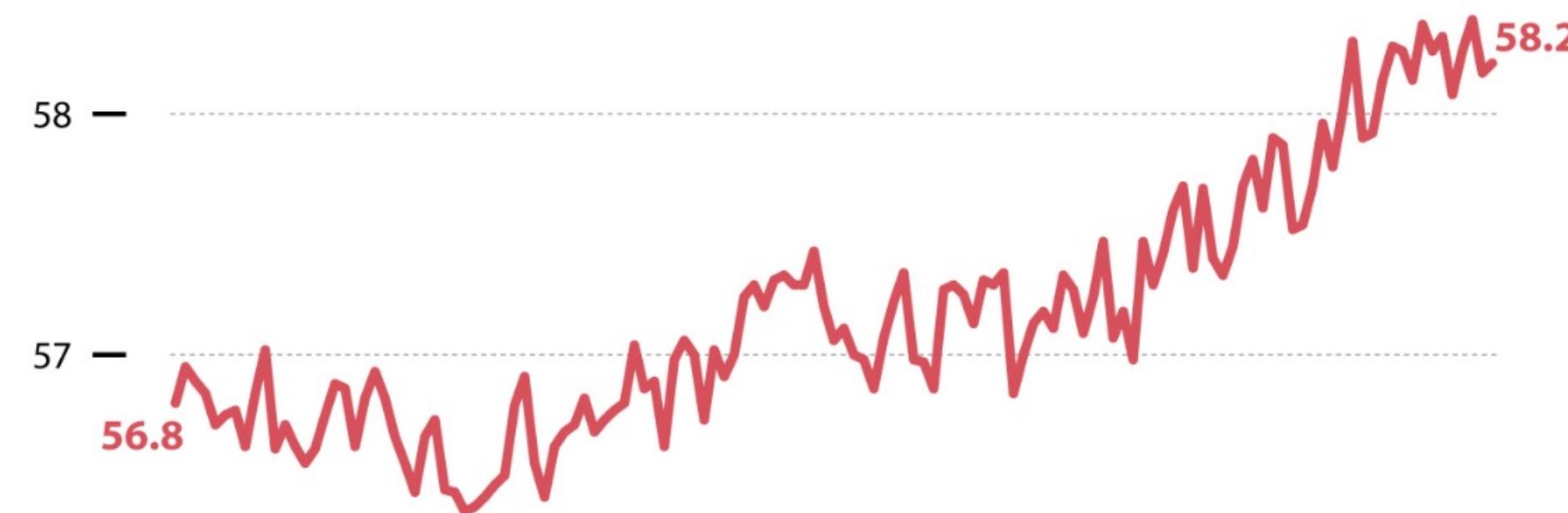
57 —

56.8

56 —

1880 1900 1920 1940 1960 1980 2000 2012

58.2



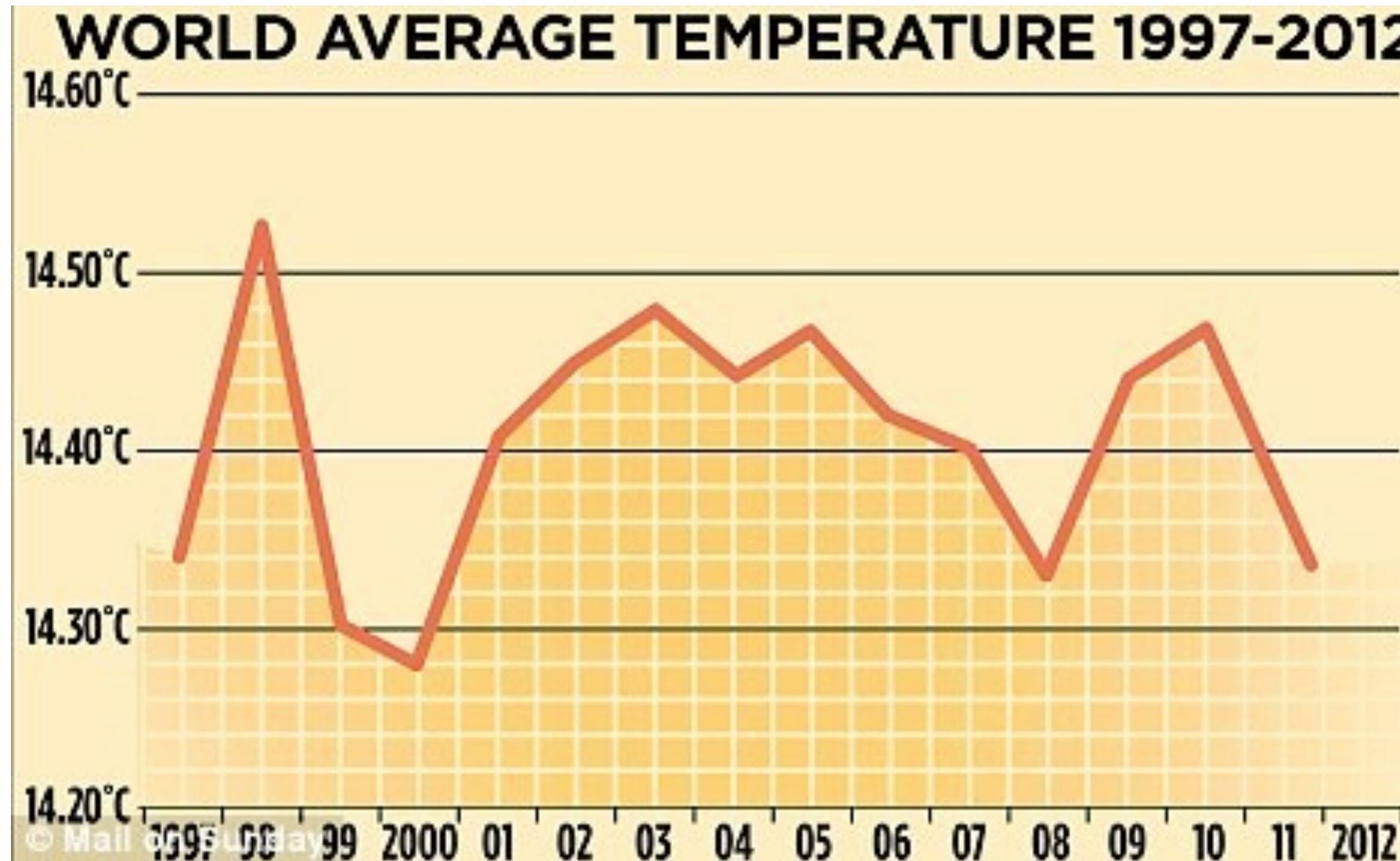
Zero-baseline Scale Advice for Alberto Cairo

- **Zero Baseline for Height/Length Encodings:** recommends a baseline of zero for charts where data is encoded through height or length (e.g., bar charts), to prevent misinterpretation of differences between values.
- **Non-Zero Baseline for Position/Angle Encodings:** For line charts, where the encoding is through position and angle, a non-zero baseline might be appropriate, as it doesn't distort the data representation and can offer clearer trend insights.
- Remember that CONTEXT MATTERS

Advice from Alberto Cairo

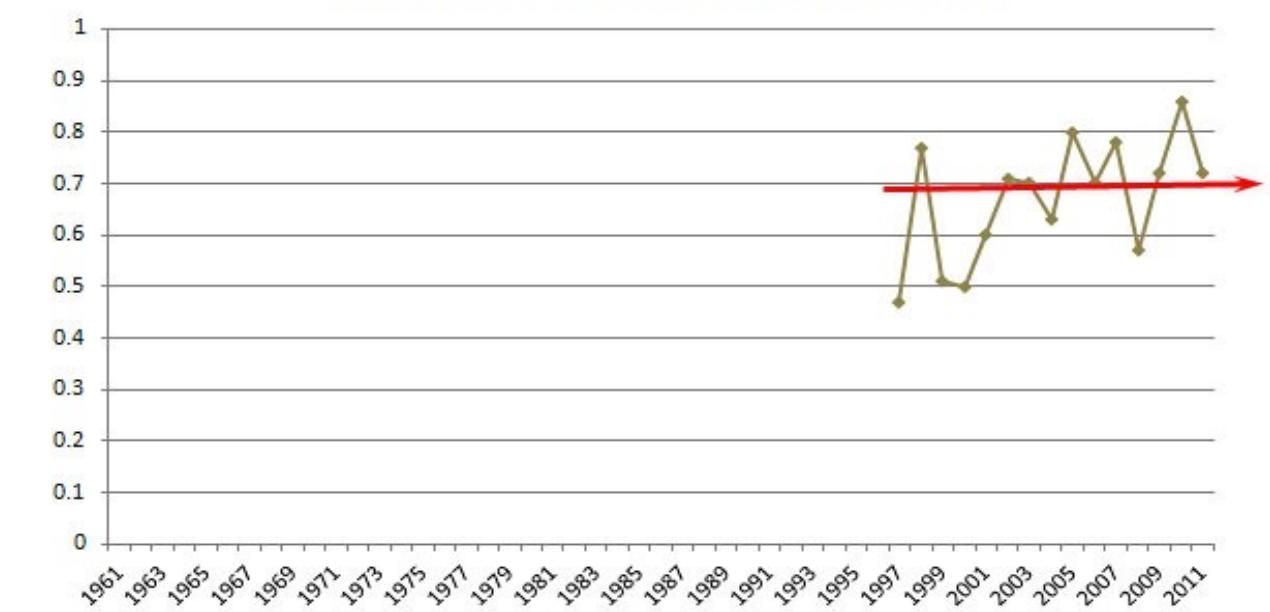
- **Clarify Over Simplify:** Aim to clarify data rather than oversimplify it, ensuring that important details and distributions are not lost.
- **Contextual Detailing:** Adjust the level of detail based on the context, audience, and nature of the data to effectively convey the story.
- **Balance in Simplicity:** Strive for simplicity by removing irrelevant details while adding meaningful information to enhance understanding and attractiveness. (Similar to Tufte)

Global Warming?

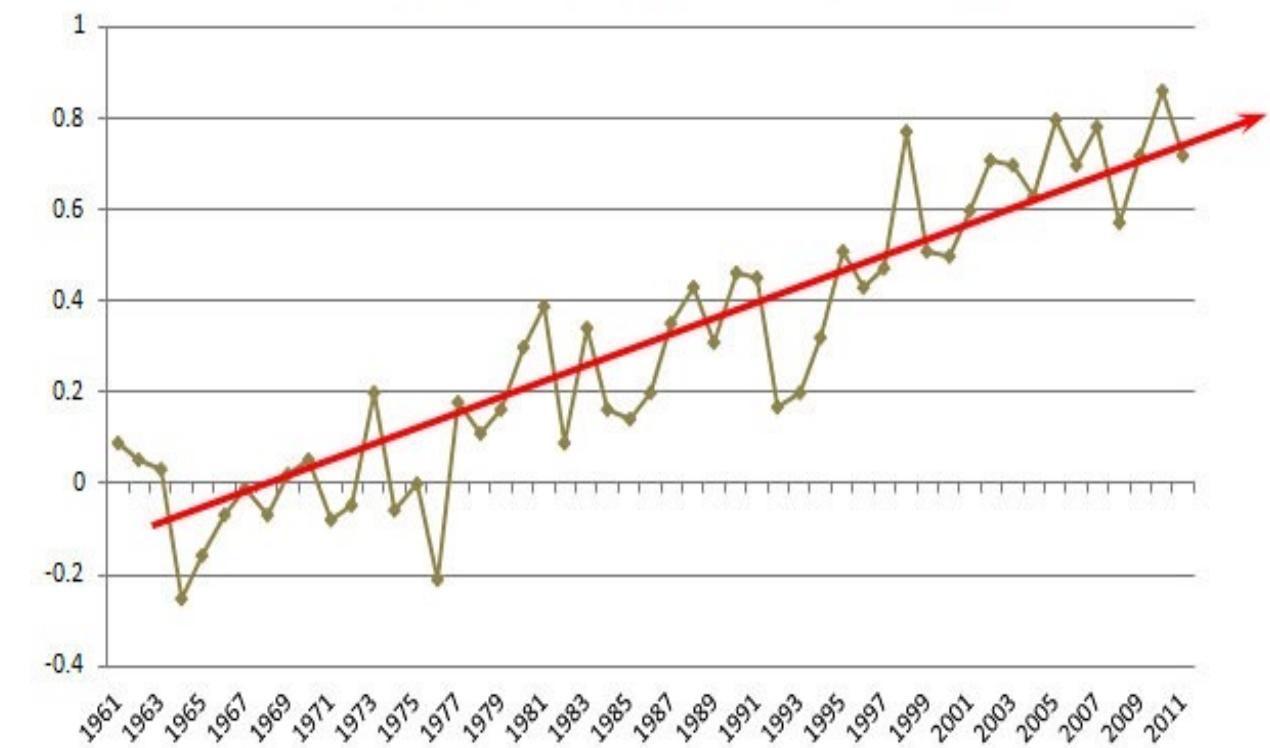


<https://www.dailymail.co.uk/sciencetech/article-2093264/Forget-global-warming--Cycle-25-need-worry-NASA-scientists-right-Thames-freezing-again.html>

Temperature Anomaly -- Annual Mean (°C)



Temperature Anomaly -- Annual Mean (°C)



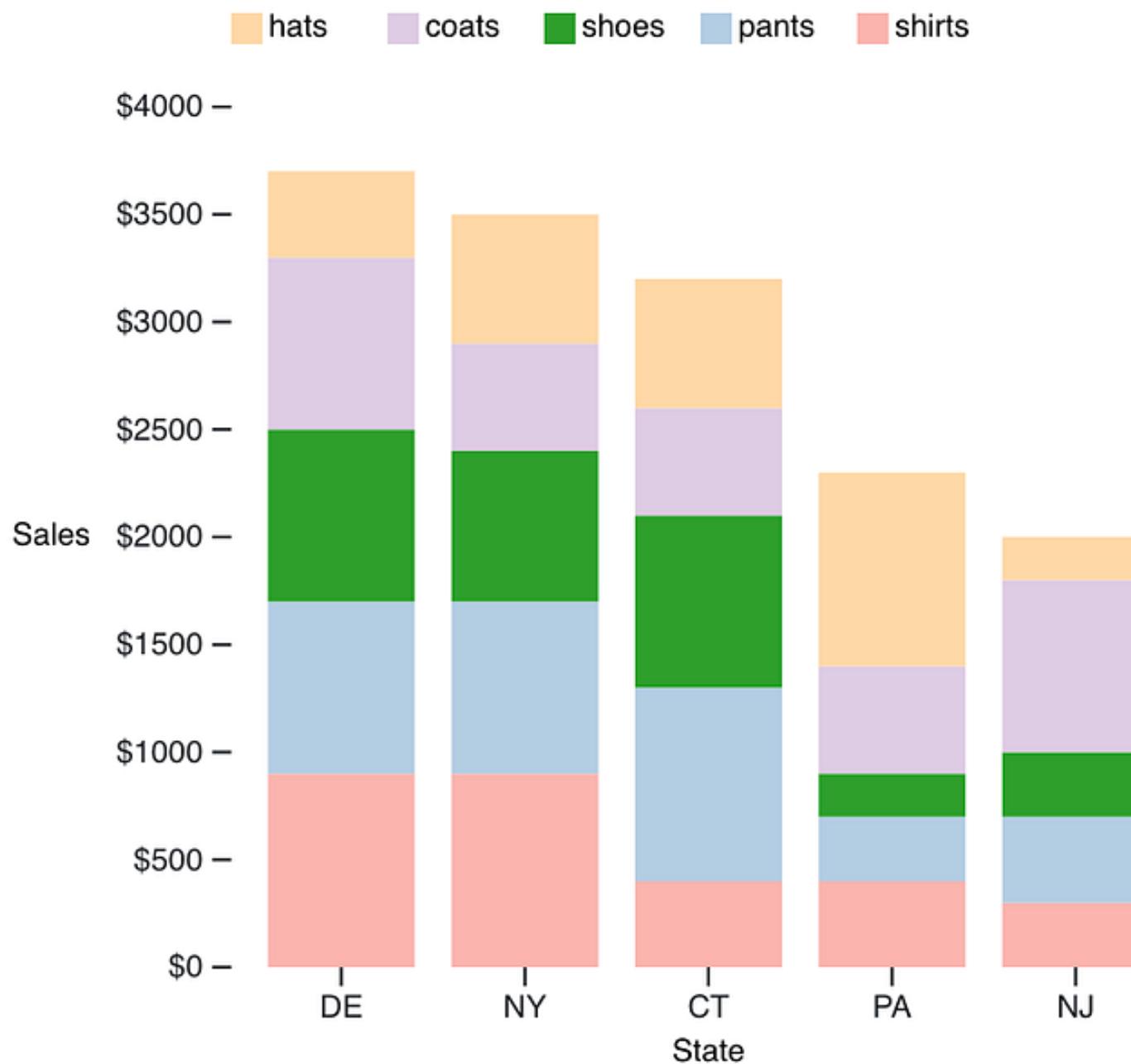
<https://www.motherjones.com/kevin-drum/2012/01/lying-charts-global-warming-edition/>

What is the issue here?

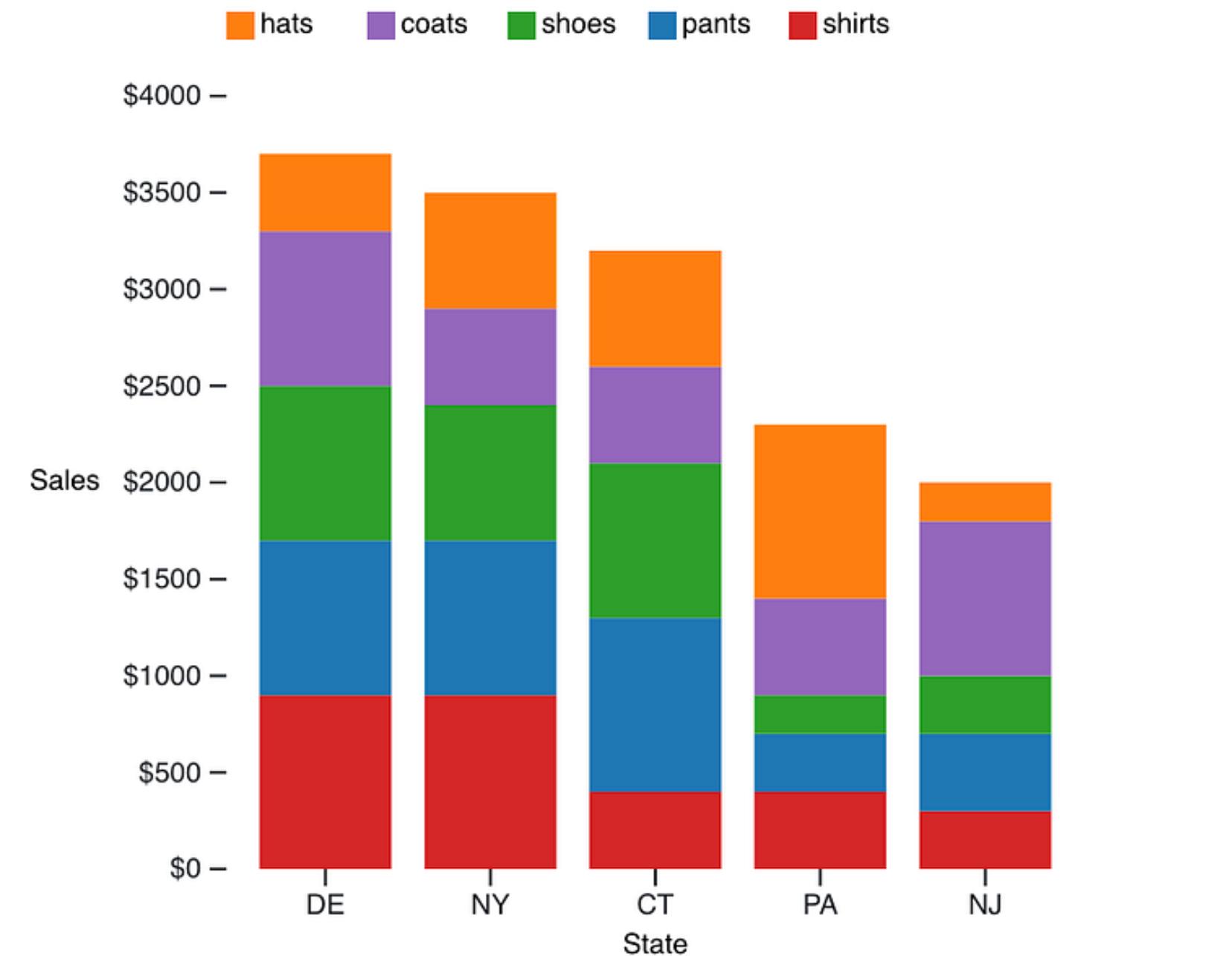




Product Sales by State



Product Sales by State

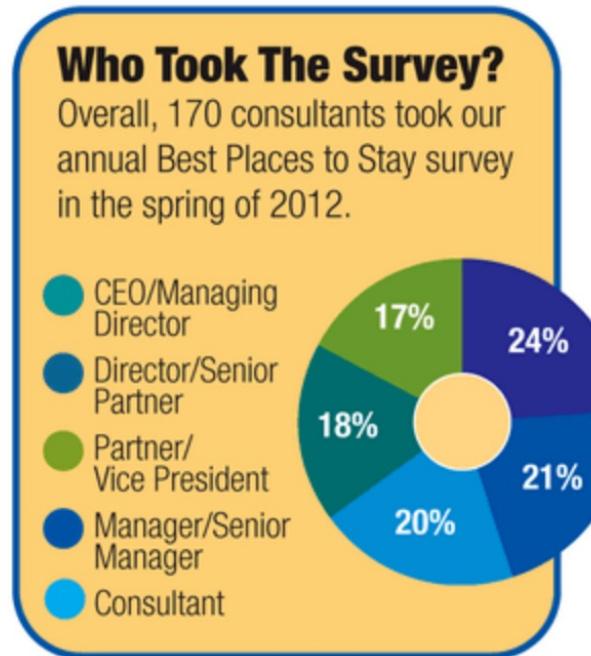
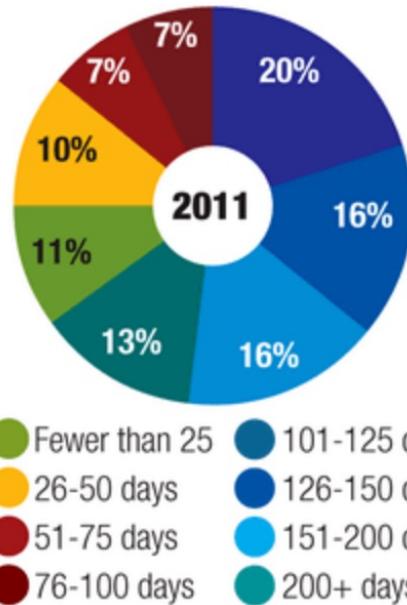
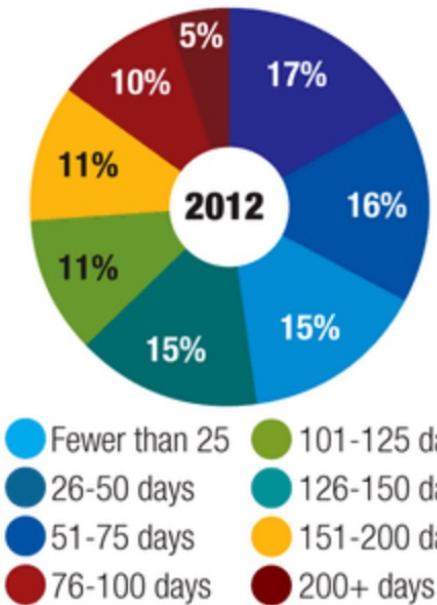


Color Misuse

- Semantic Incongruence: Red for positive, green for negative
- Rainbow Colormaps: For continuous data (creates false boundaries)
- Encoding Redundancy: Multiple channels encoding same variable
- Color Profusion: Too many colors to distinguish
- Poor Accessibility: Not colorblind-safe palettes

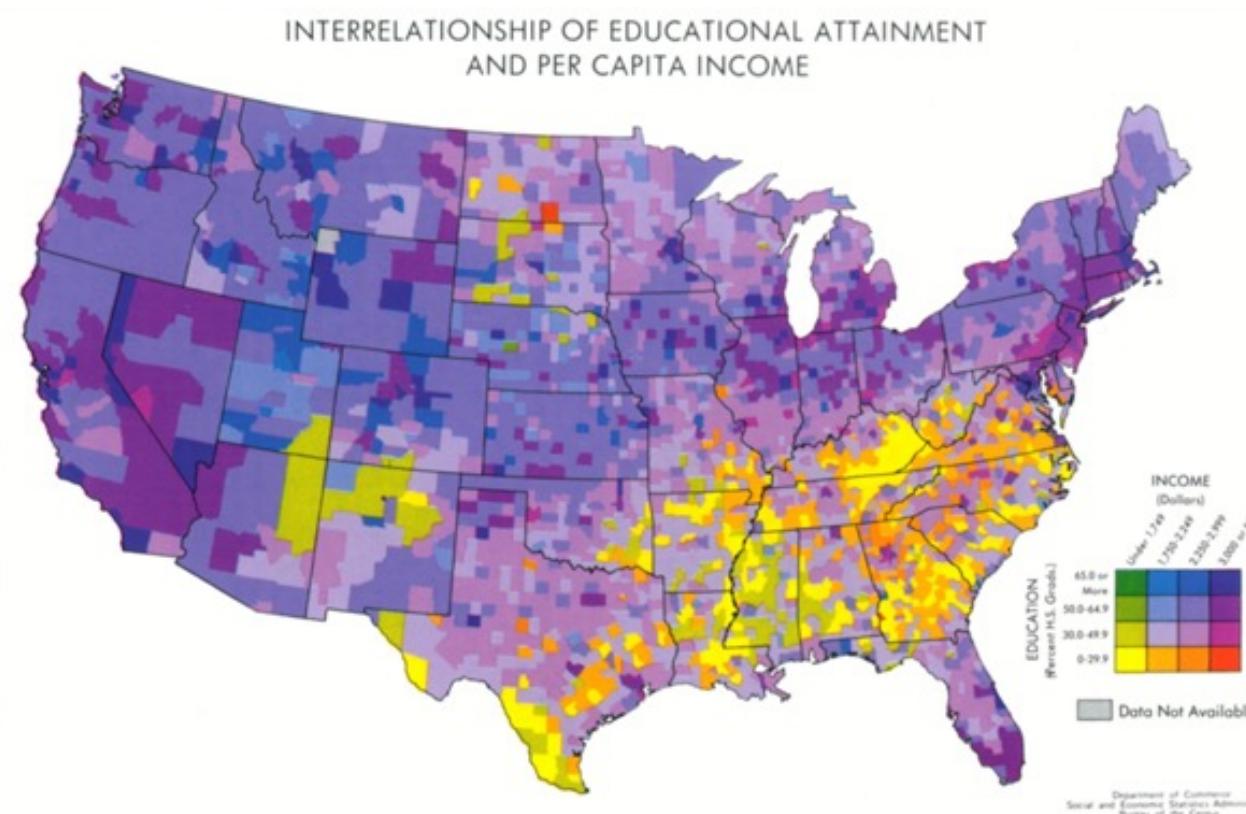
Color Abuse: Hues, Attributes, Cardinality

How many nights do you typically stay away from home due to work?

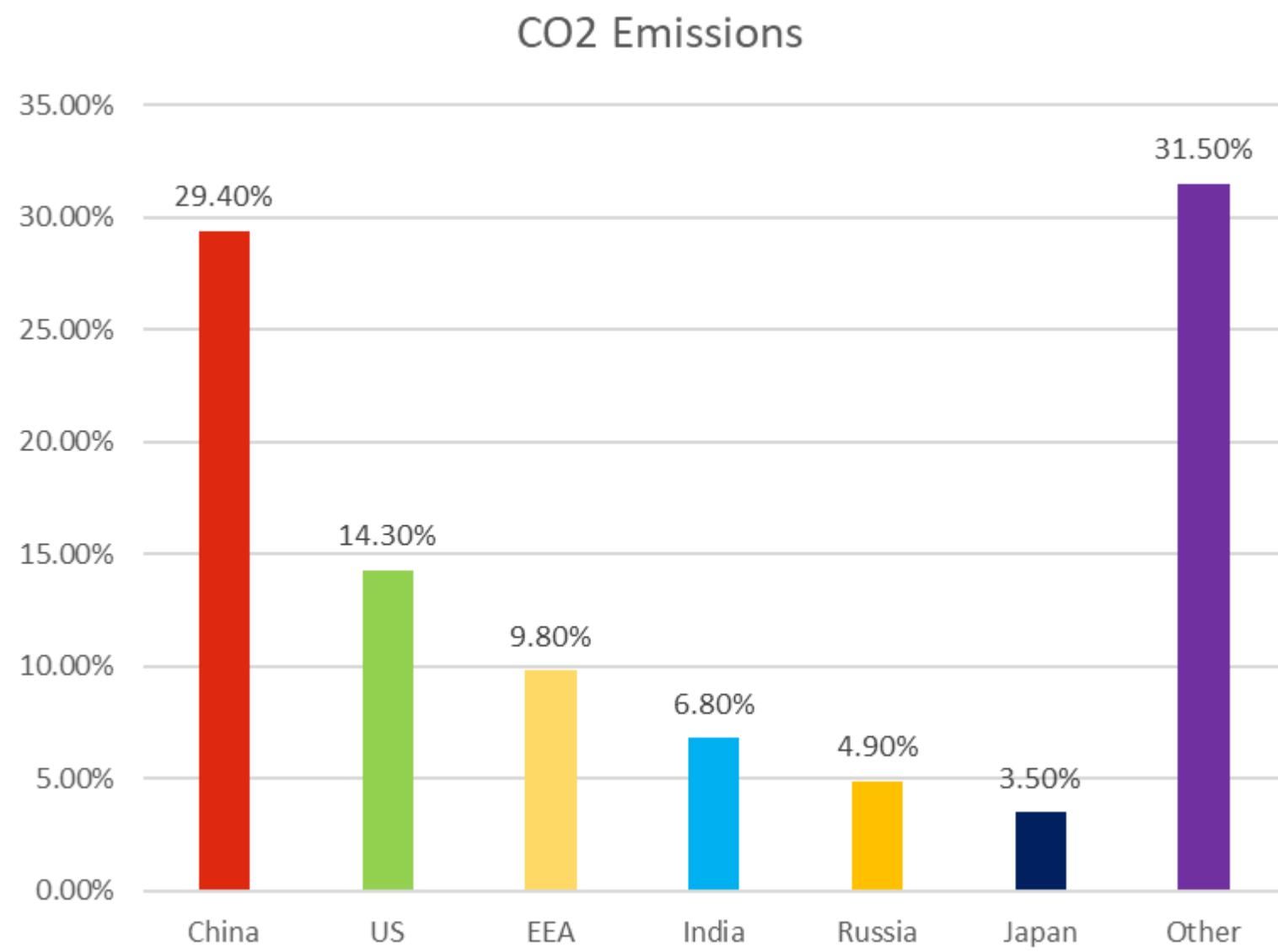


<https://connorrothschild.github.io/v2/post/color-in-data-vis/>

<https://cartographicperspectives.org/index.php/journal/article/view/1538/1819>

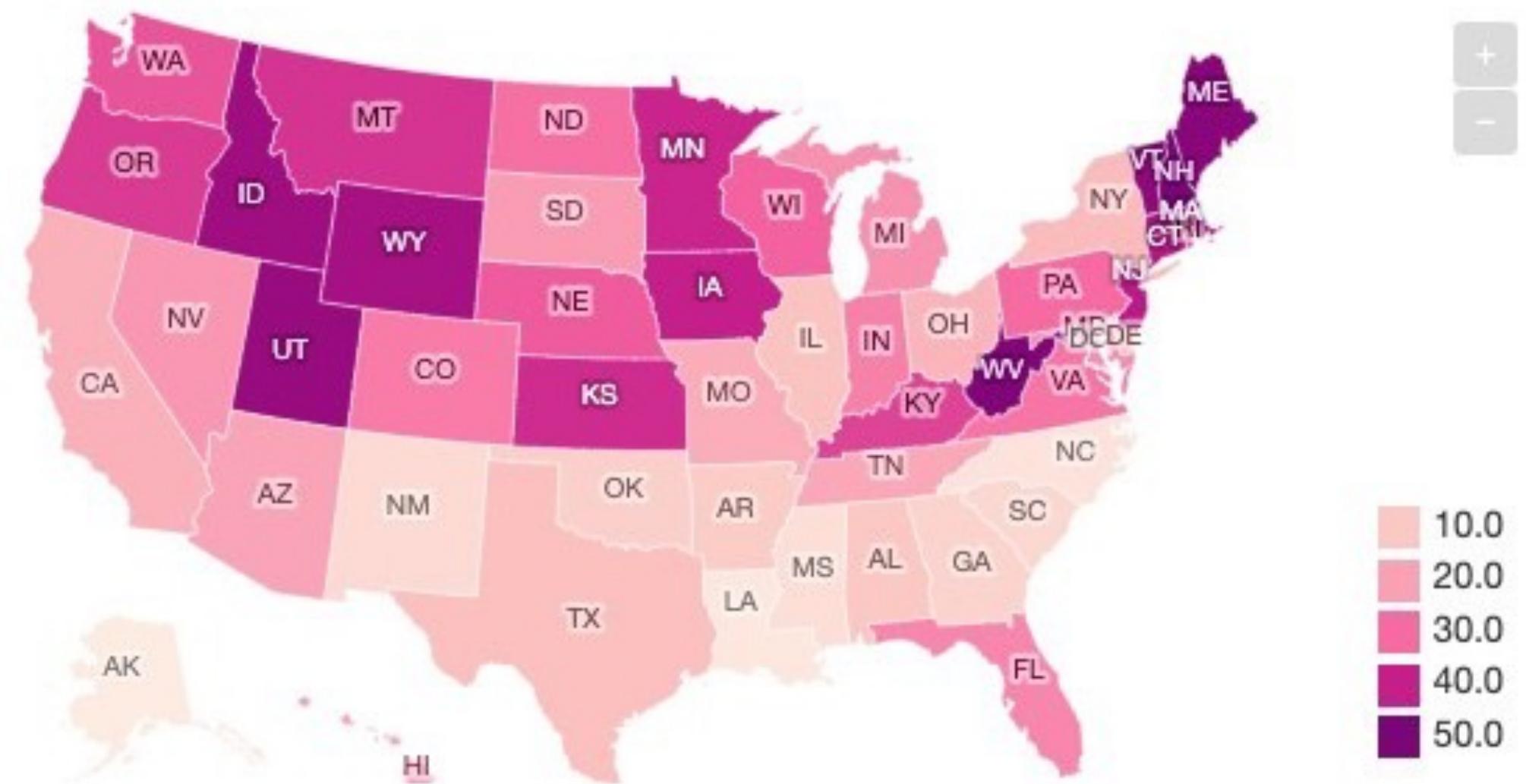


Color Abuse: Encoding Redundancy



Color Abuse: Semantic Incongruence

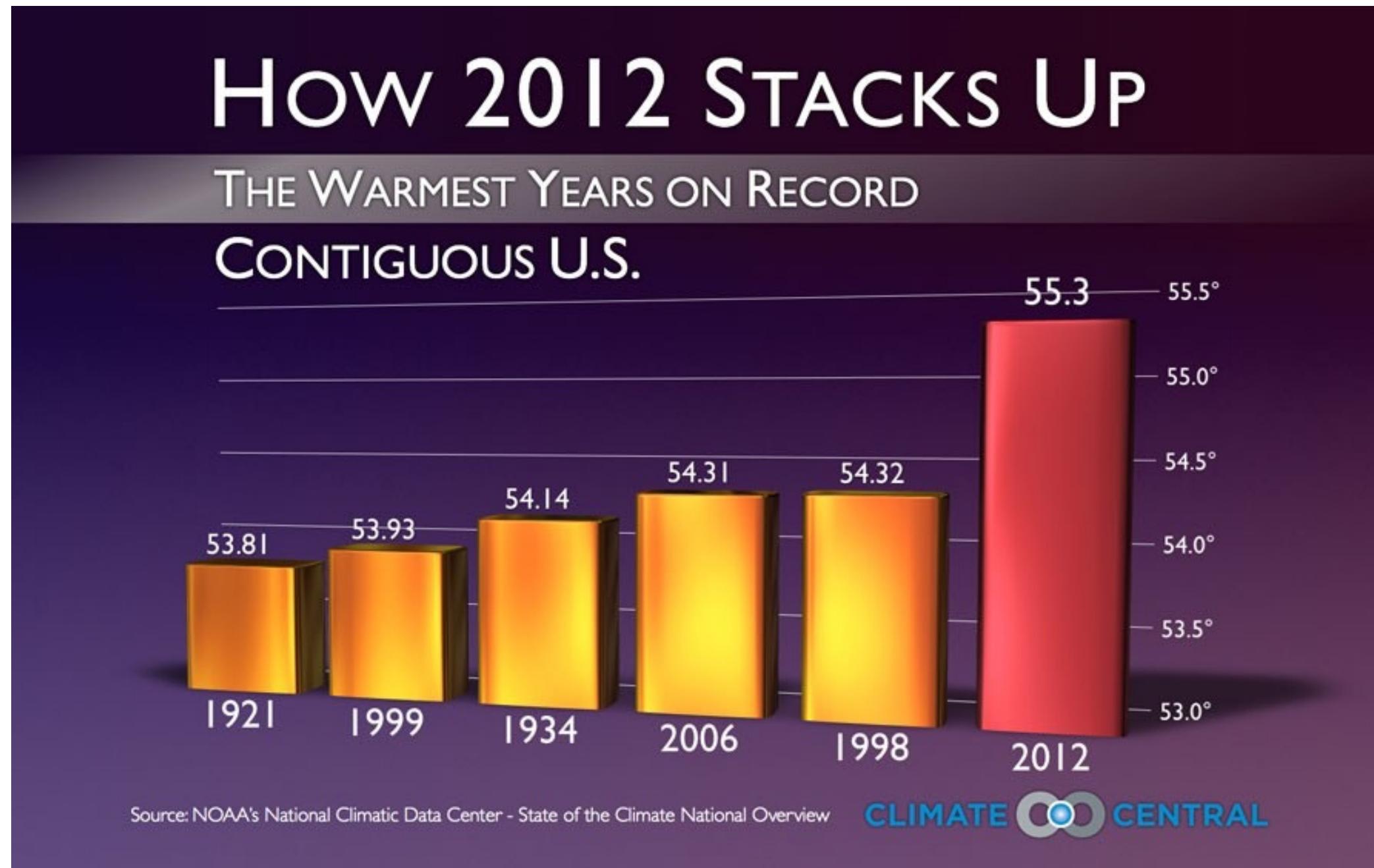
Which states have the most STIs?



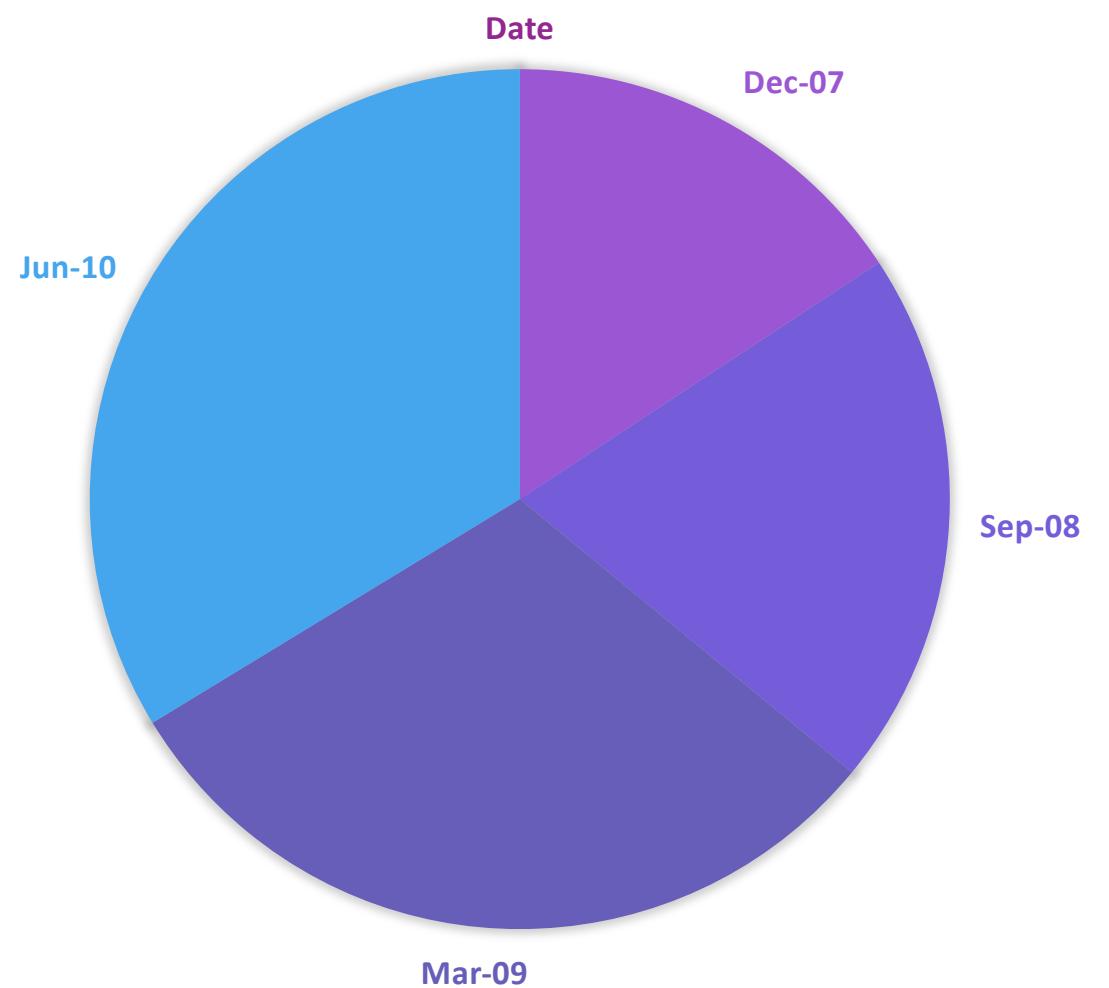
Color Design Tips

- **Limited Palette:** Use about six colors for focus and simplicity. Cardinality matters
- **Color Harmony:** Aim for harmonious color schemes, possibly using natural colors for familiarity.
- **Cultural Conventions and Symbolism:** Incorporate culturally significant colors and understand their symbolism. Recognize that color preferences vary by culture and individual taste.
- **Effective in Black and White:** Ensure the visualization is clear and understandable even without color.
- **Respect for the Color Deficiency:** Choose colors that are distinguishable to those with color vision deficiencies.
- **Avoid Overloading the Color Channel:** Refrain from using colors to represent multiple attributes, to prevent confusion and ensure clear interpretation of each aspect in the visualization.

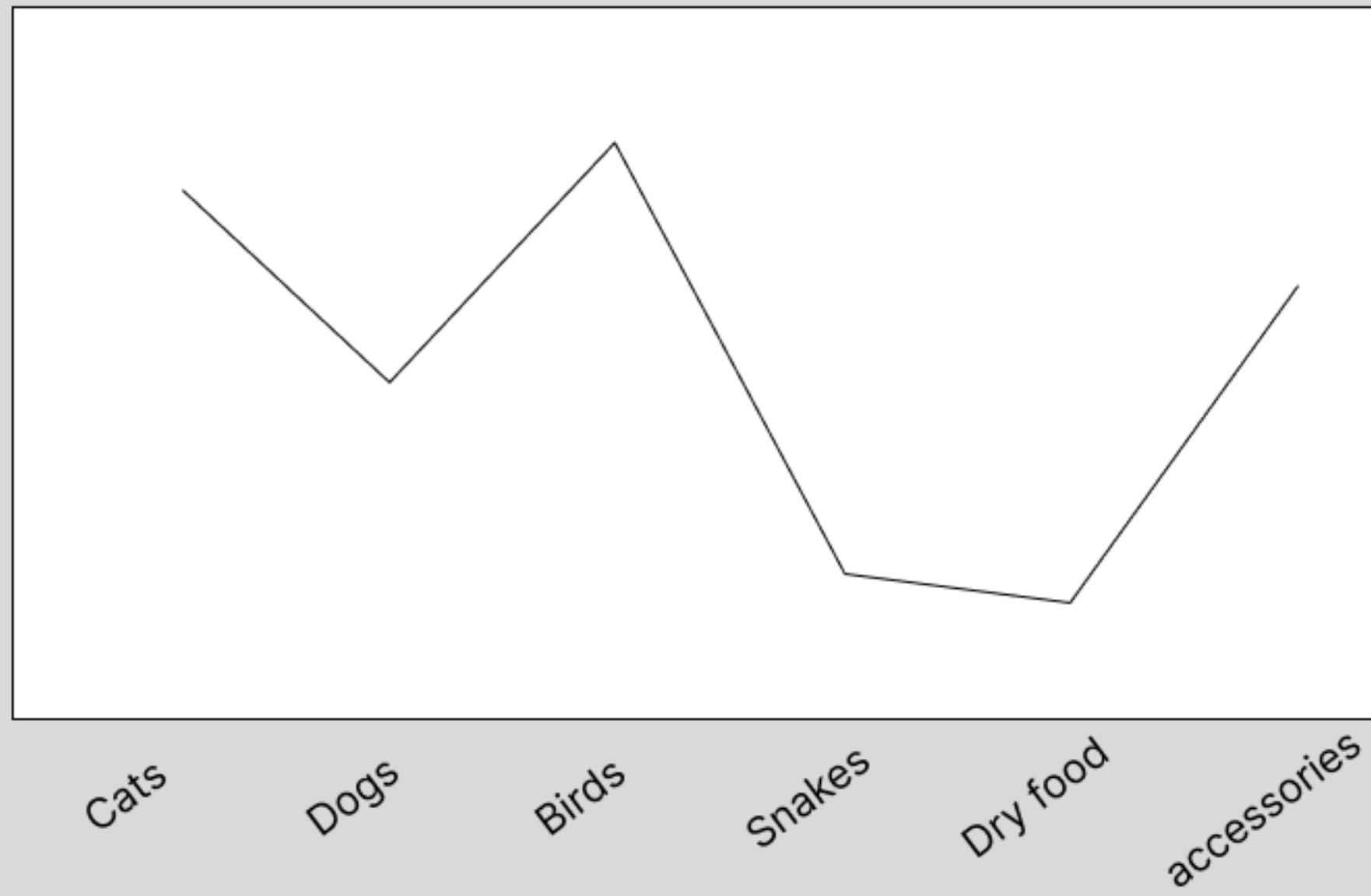
Activity: List the various issues that exist



RANKING OF FAVORITE PIZZA TOPPINGS



Pet Store Sales



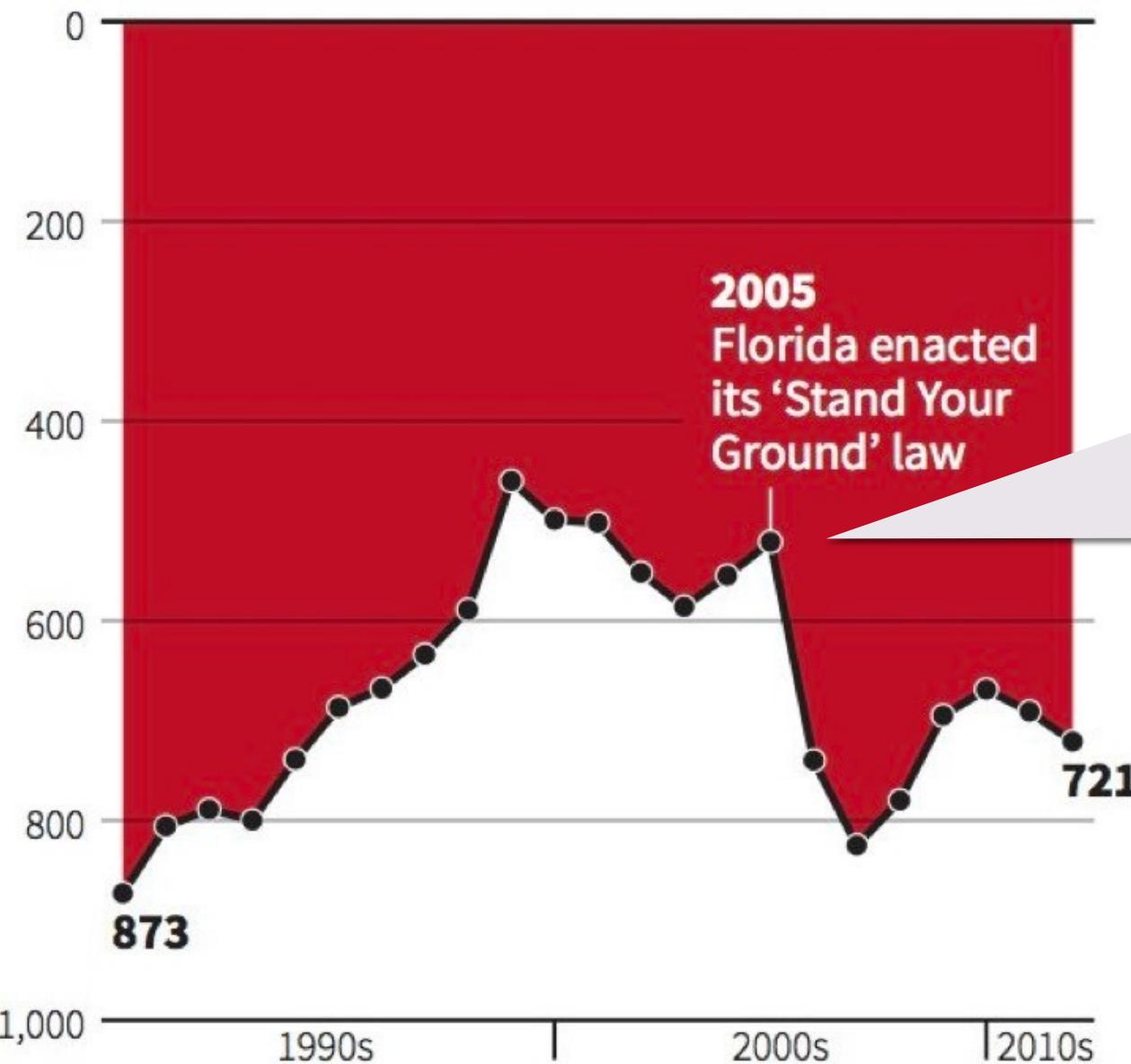
Data mapping Issues

- Incorrect Data to Encoding: Mapping continuous data to discrete categories
- Mismatched Chart Type: Pie chart for rankings, line chart for unordered categories
- Spurious Precision: More decimal places than data accuracy supports

Intentionally Misleading Viz.

Gun deaths in Florida

Number of murders committed using firearms



this creates the immediate visual impression that gun deaths declined sharply after stand-your-ground legislation was enacted in Florida.

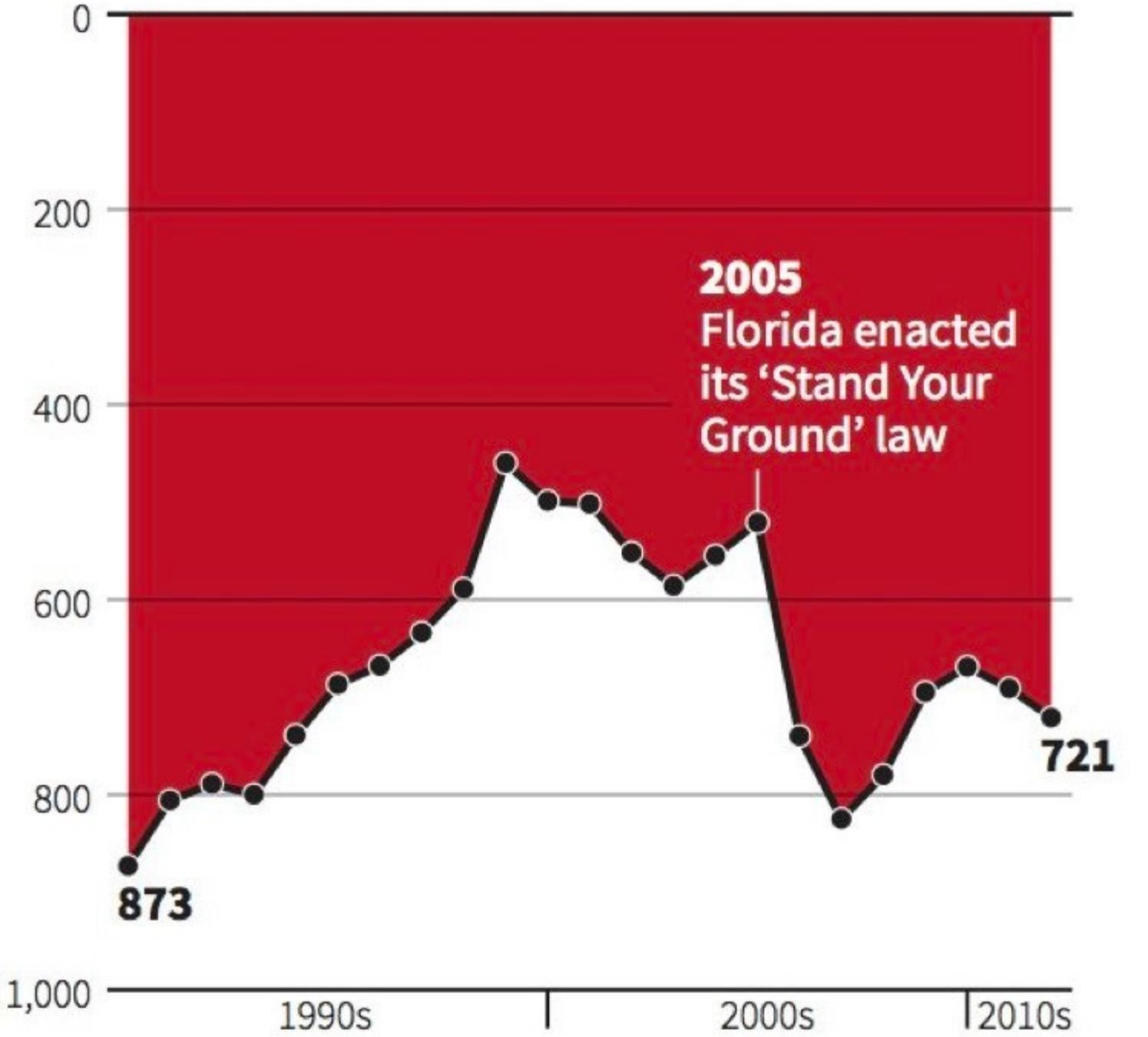
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

Gun deaths in Florida

Number of murders committed using firearms

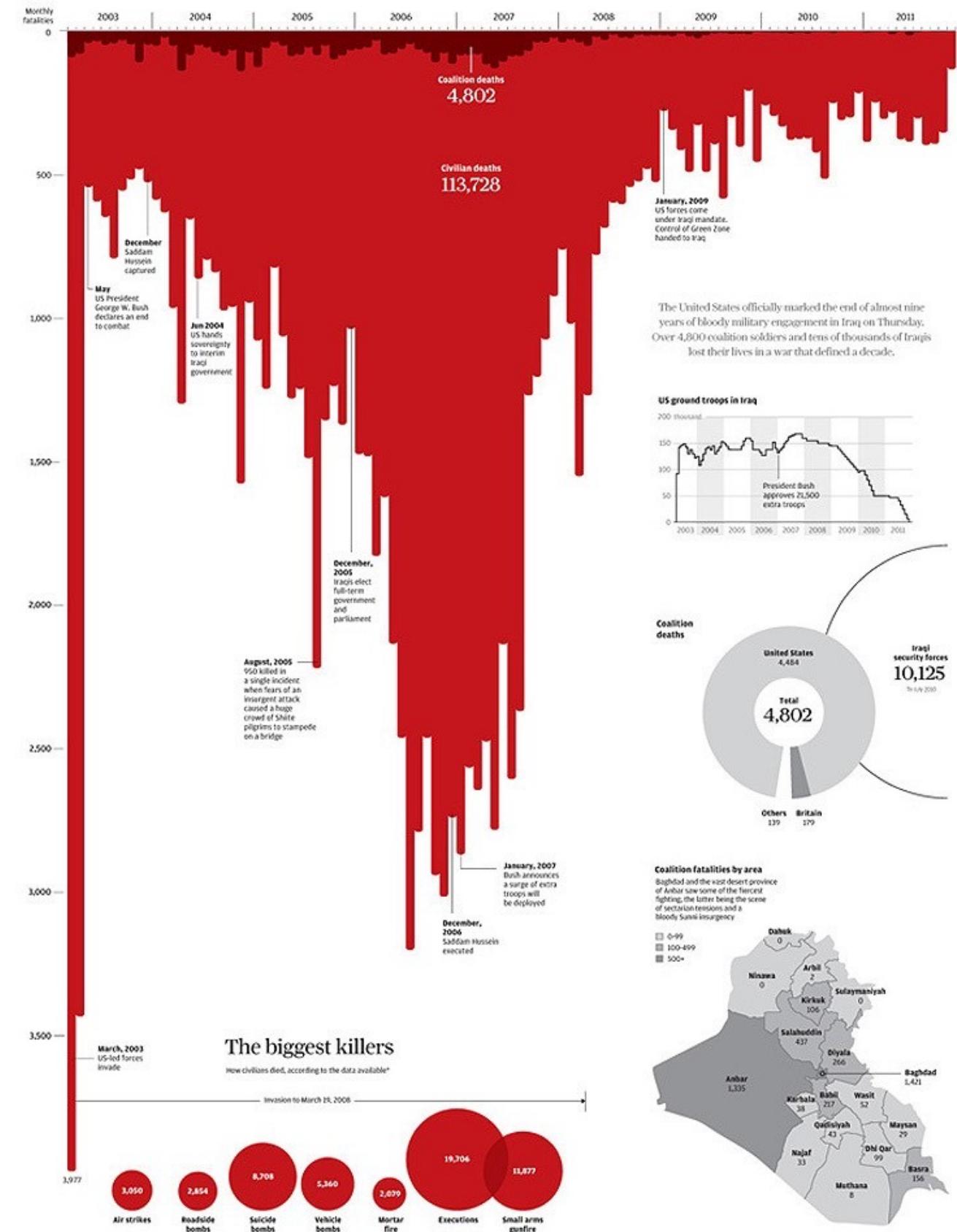


Source: Florida Department of Law Enforcement

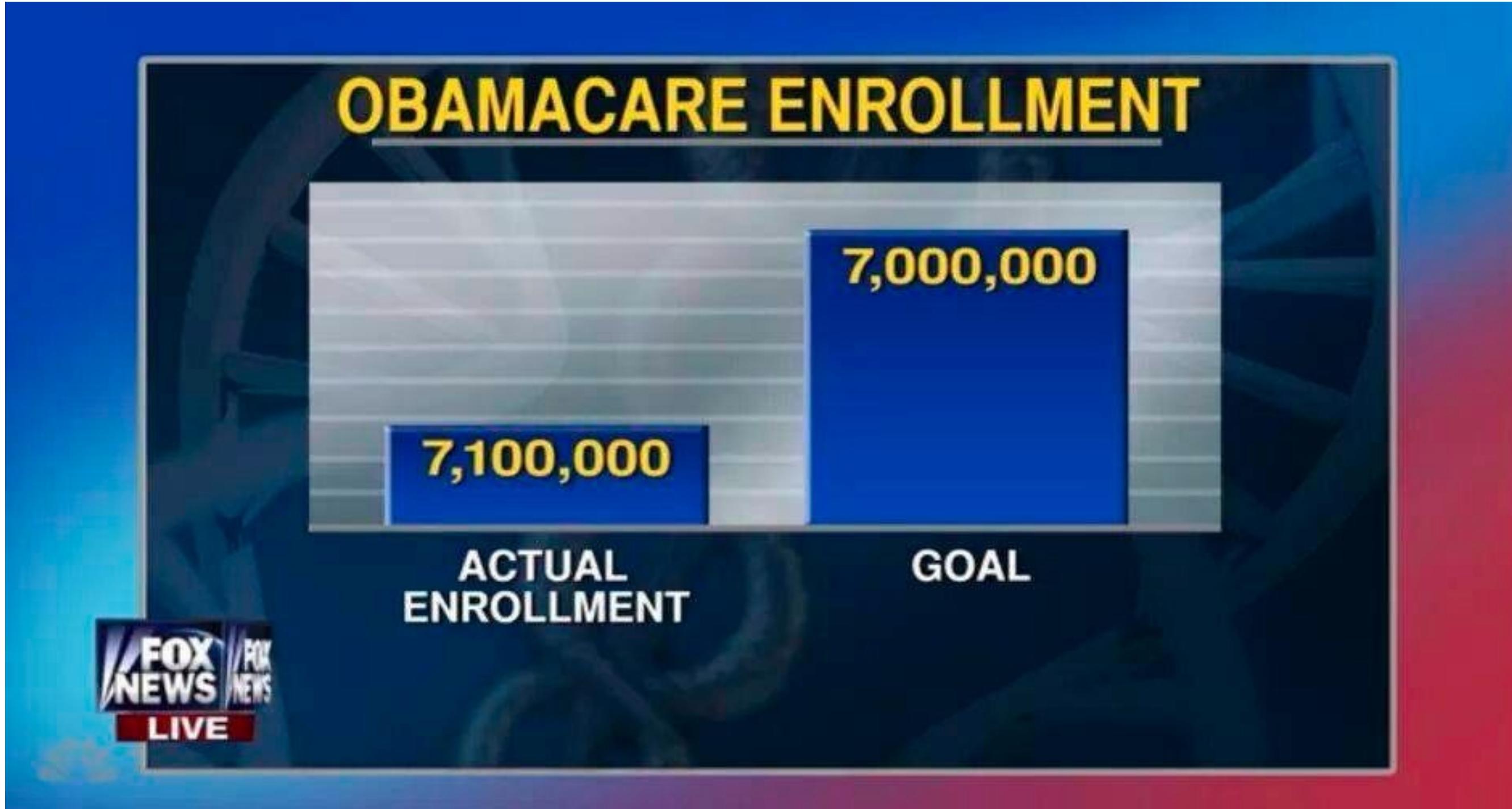
C. Chan 16/02/2014

REUTERS

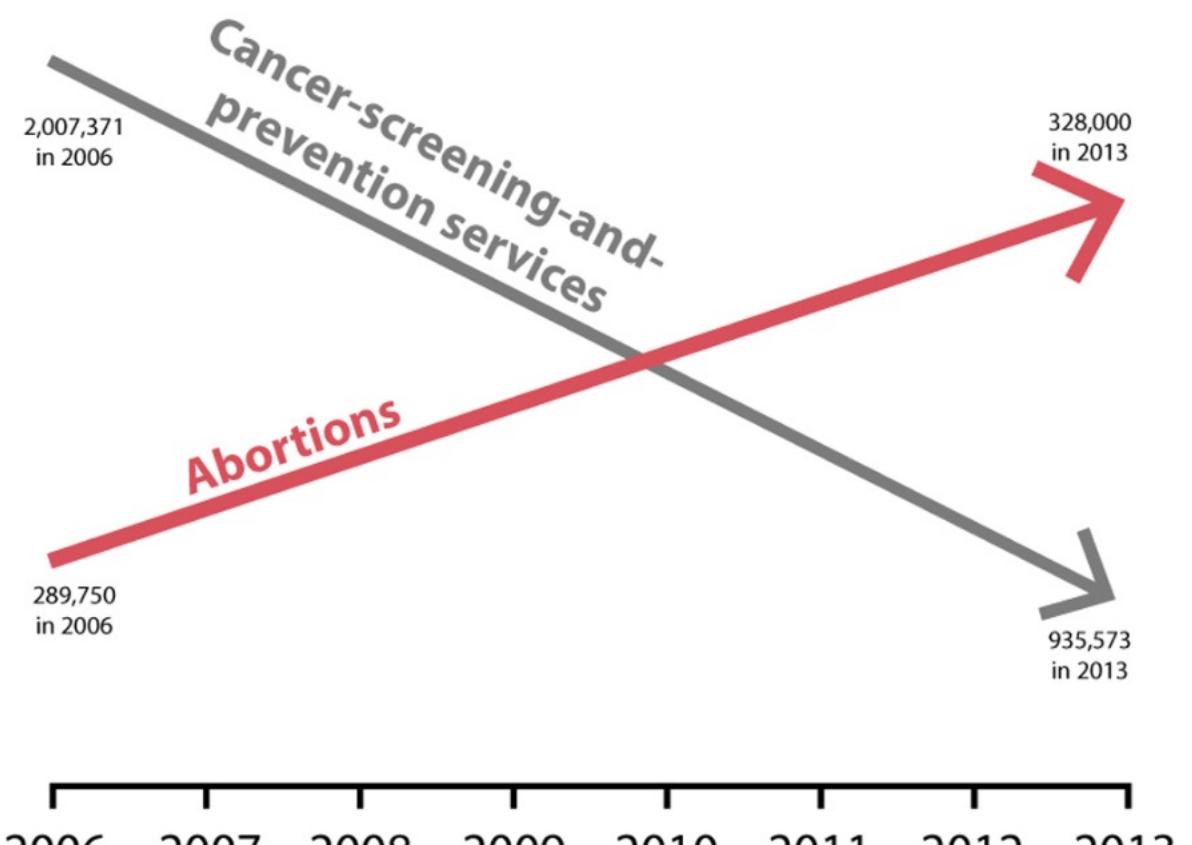
Iraq's bloody toll



What is wrong with this chart?

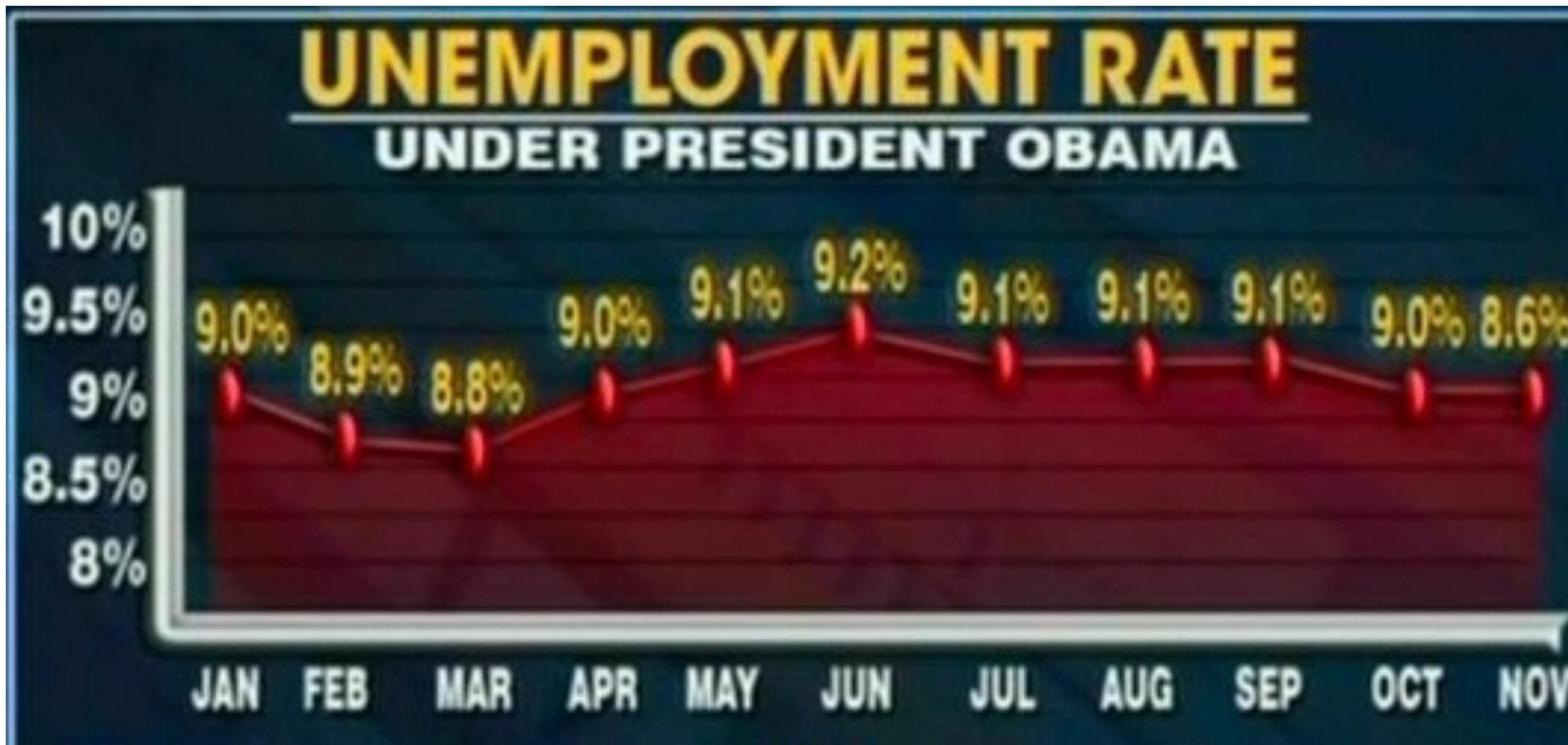


Planned Parenthood Federation of America:
Abortions up—life-saving procedures down



(Source: Americans United for Life)

List the various issues that exist?



Underlying Data Issues

- Incorrect Data: failing to encode the right data
- Cherry-Picking: Selecting only favorable time periods or leaving out contradictory data or categories
- Temporal Manipulation: Strategic start/end dates
- Correlation as Causation: Implying causal relationship
- Ignoring Baseline Effects: Not adjusting for population, inflation, etc.

What is wrong with this chart?



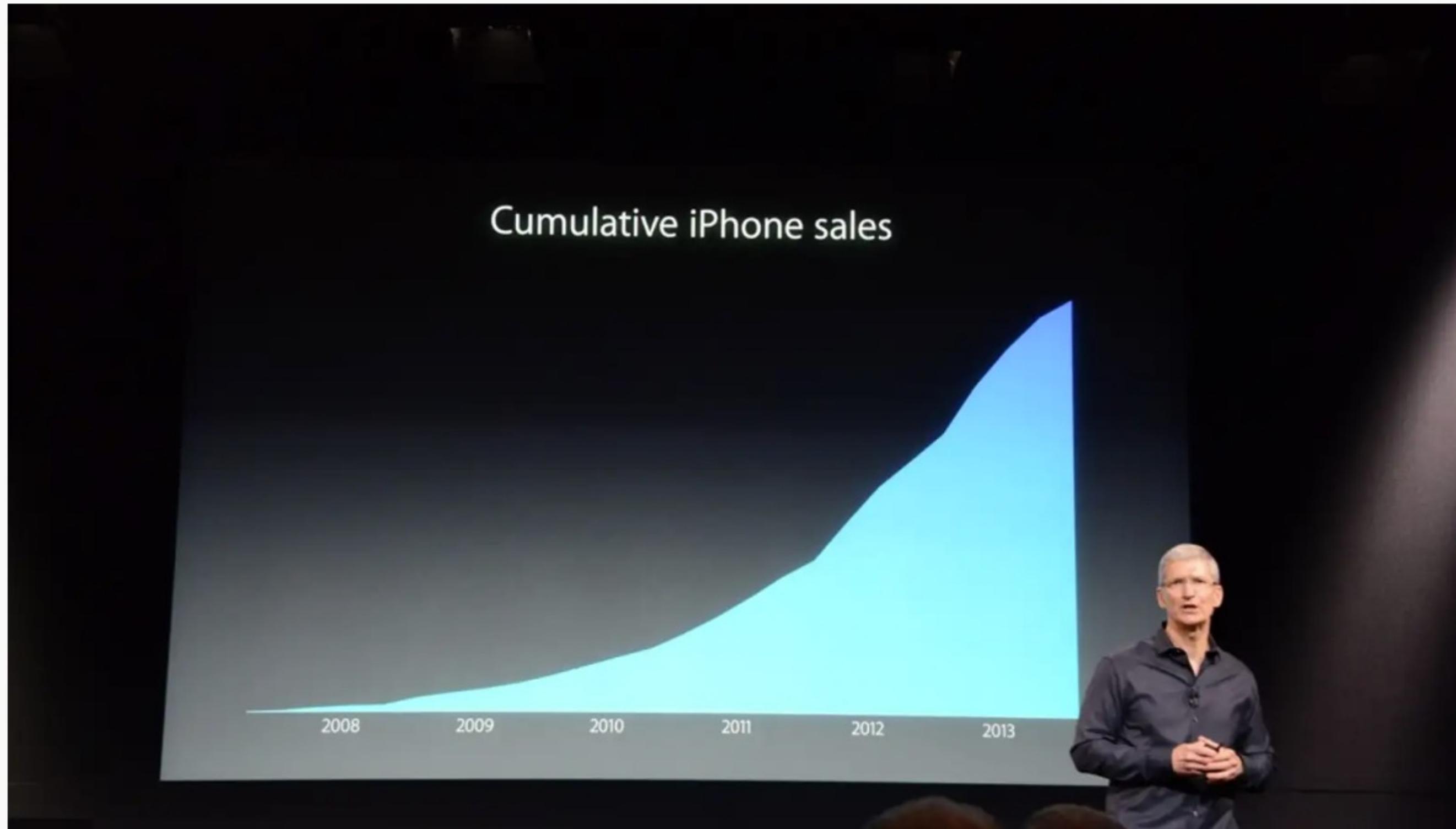
Posted in r/funny by u/veganator



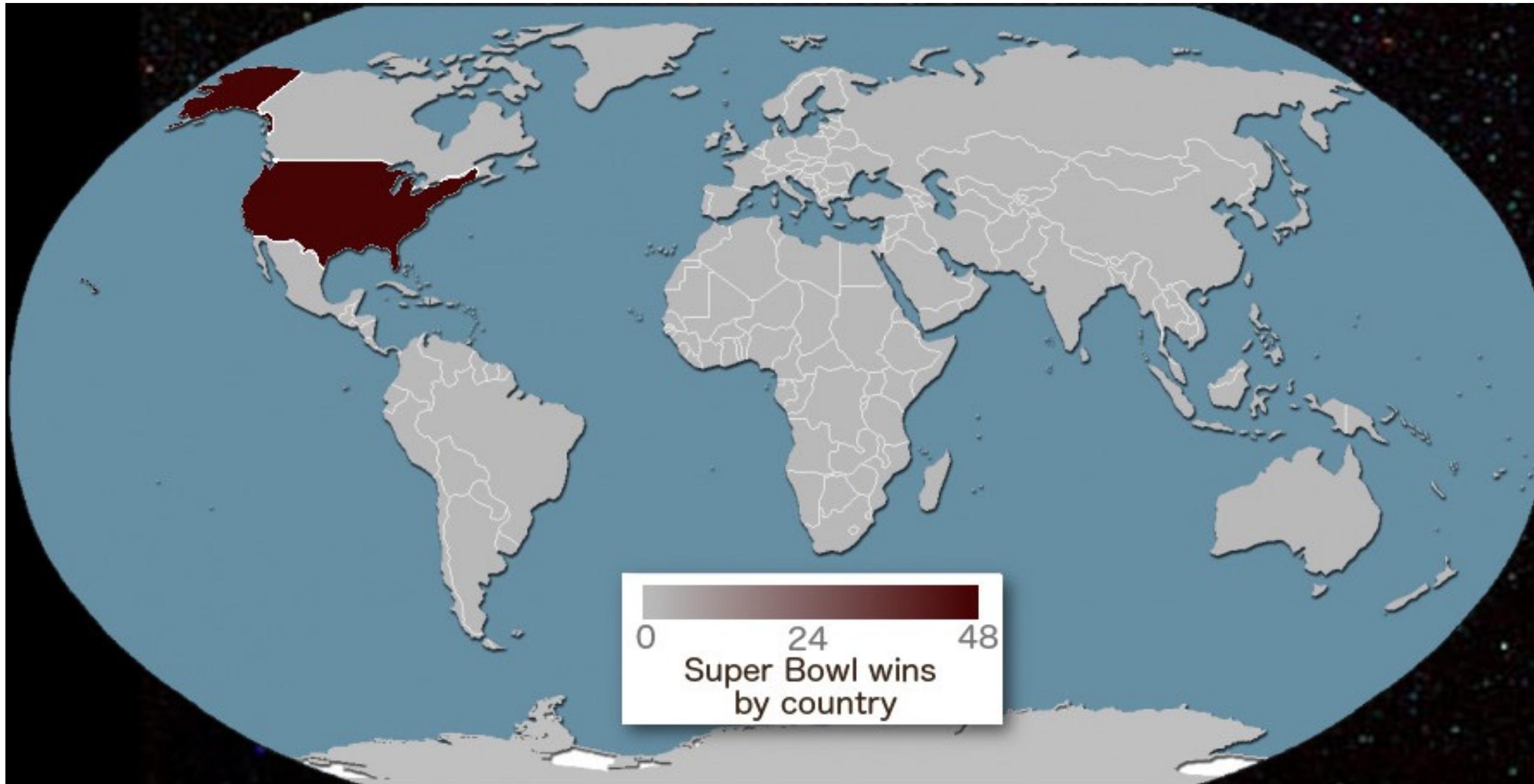
What is wrong with this chart?



What is wrong with this chart?



What is wrong with this chart?



What is wrong with this chart?

What can you conclude from this faceted bar chart?

