

CS 571 - Data Visualization & Exploration

Design Guidelines

Instructor: Hamza Elhamdadi

UMassAmherst

Upcoming Dates

Apr 4: Project Milestone

**Homework 3 will be released on Mar 31
(Due Apr 11)**

Design Lessons - Edward Tufte

Tufte describes two main ideas about visualization design:

- Graphical Integrity
- Graphical Excellence

Graphical Integrity

Use clear, detailed, and thorough labeling in your visualizations

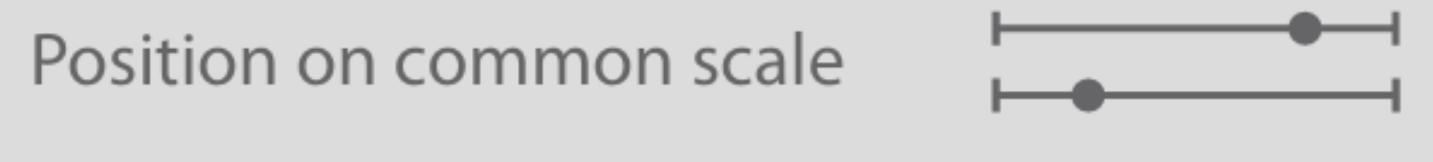
Avoid distortion and ambiguity

How do we avoid distortion/ambiguity?

How do we avoid distortion/ambiguity?

Use Effective Encoding

→ Magnitude Channels: Ordered Attributes



▲ Most Effective ▼ Least Effective

Cake Pie Tea Coffee

Cake Pie Tea Coffee

Cake Pie Tea Coffee

→ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



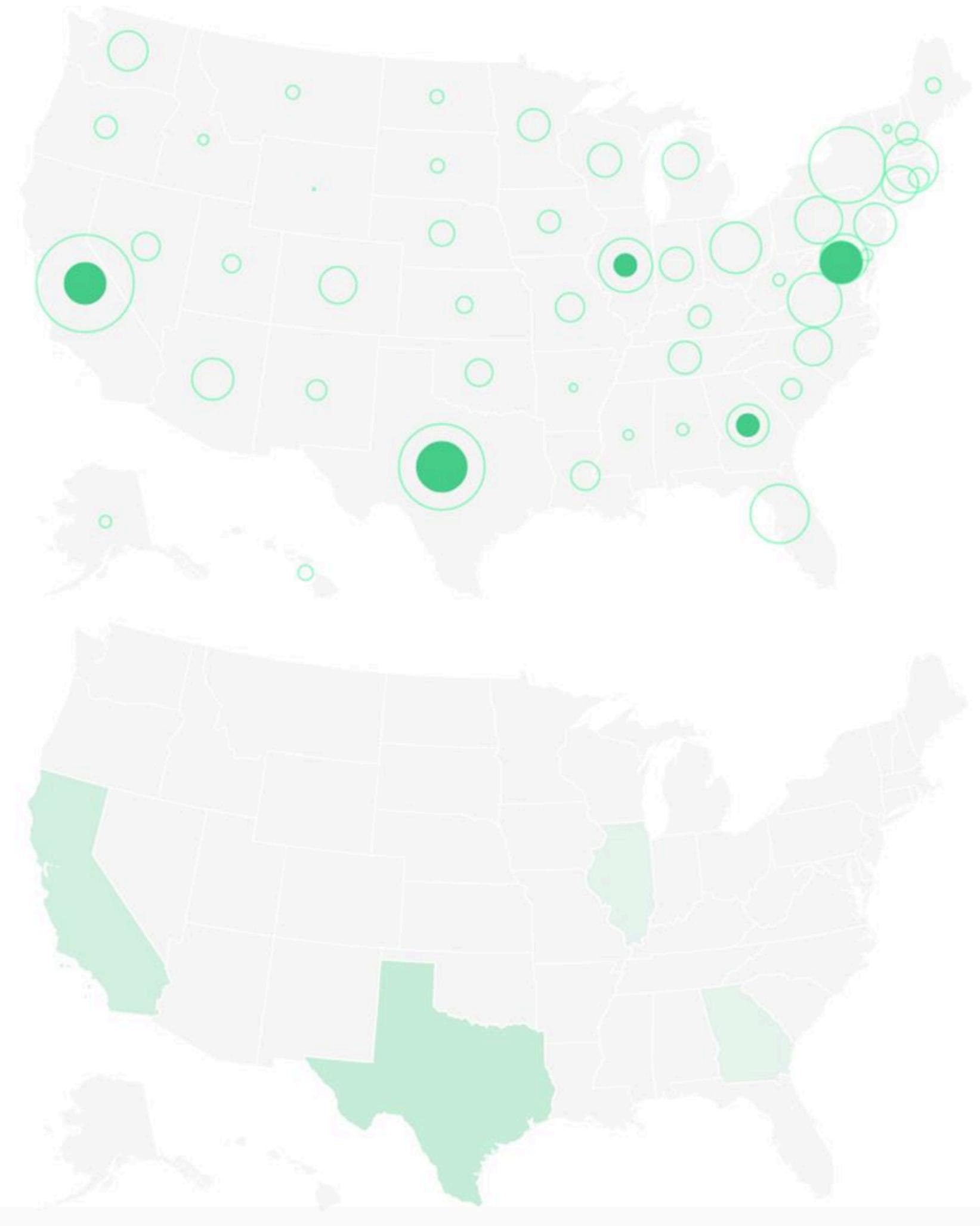
Curvature



Volume (3D size)



▲ Most
Effectiveness
— Same
Least ▼

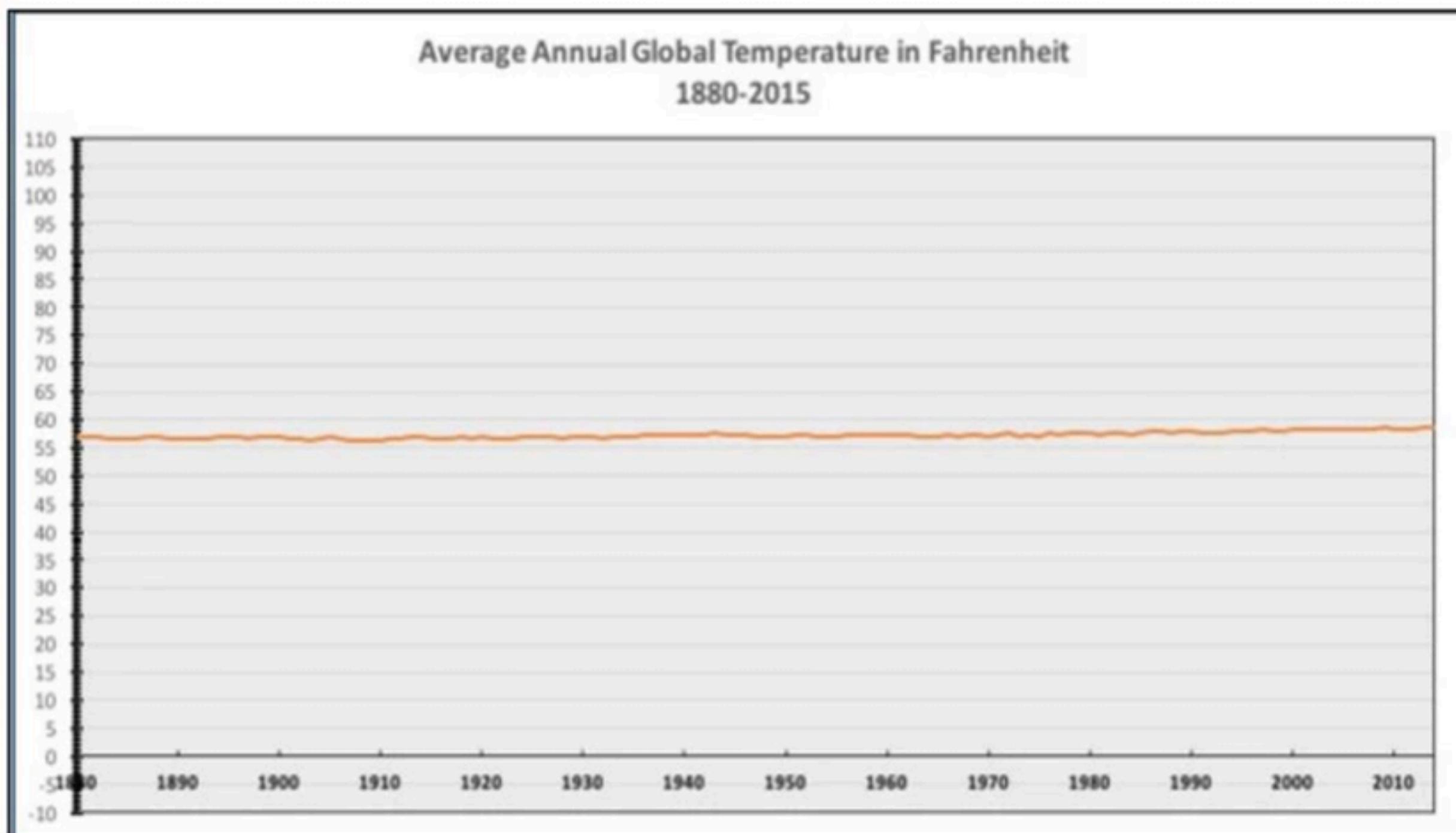


How do we avoid distortion/ambiguity?

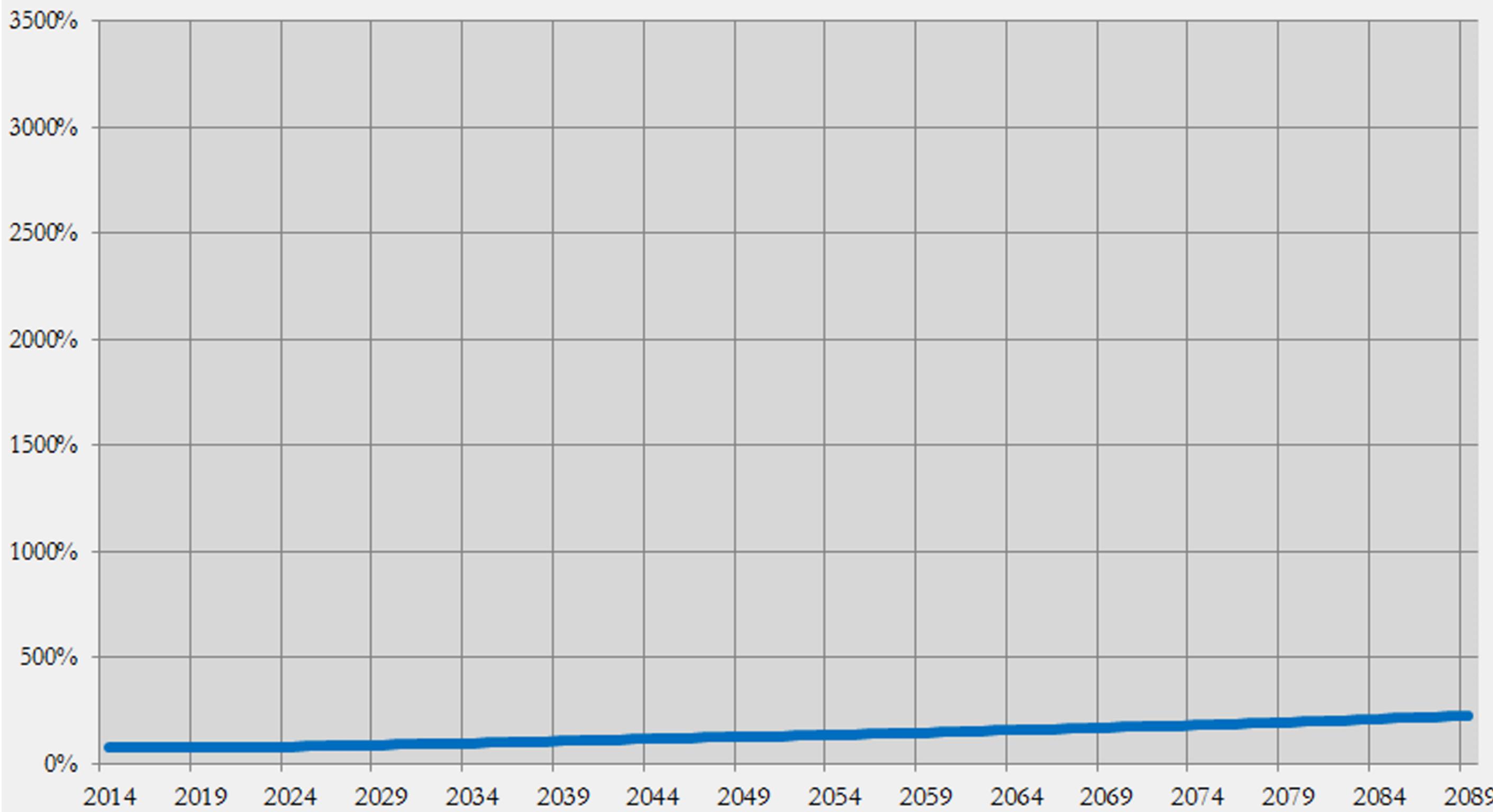
Be Aware of Your Scales

The only #climatechange chart you need to see. natl.re/wPKpro

(h/t [@powerlineUS](#))



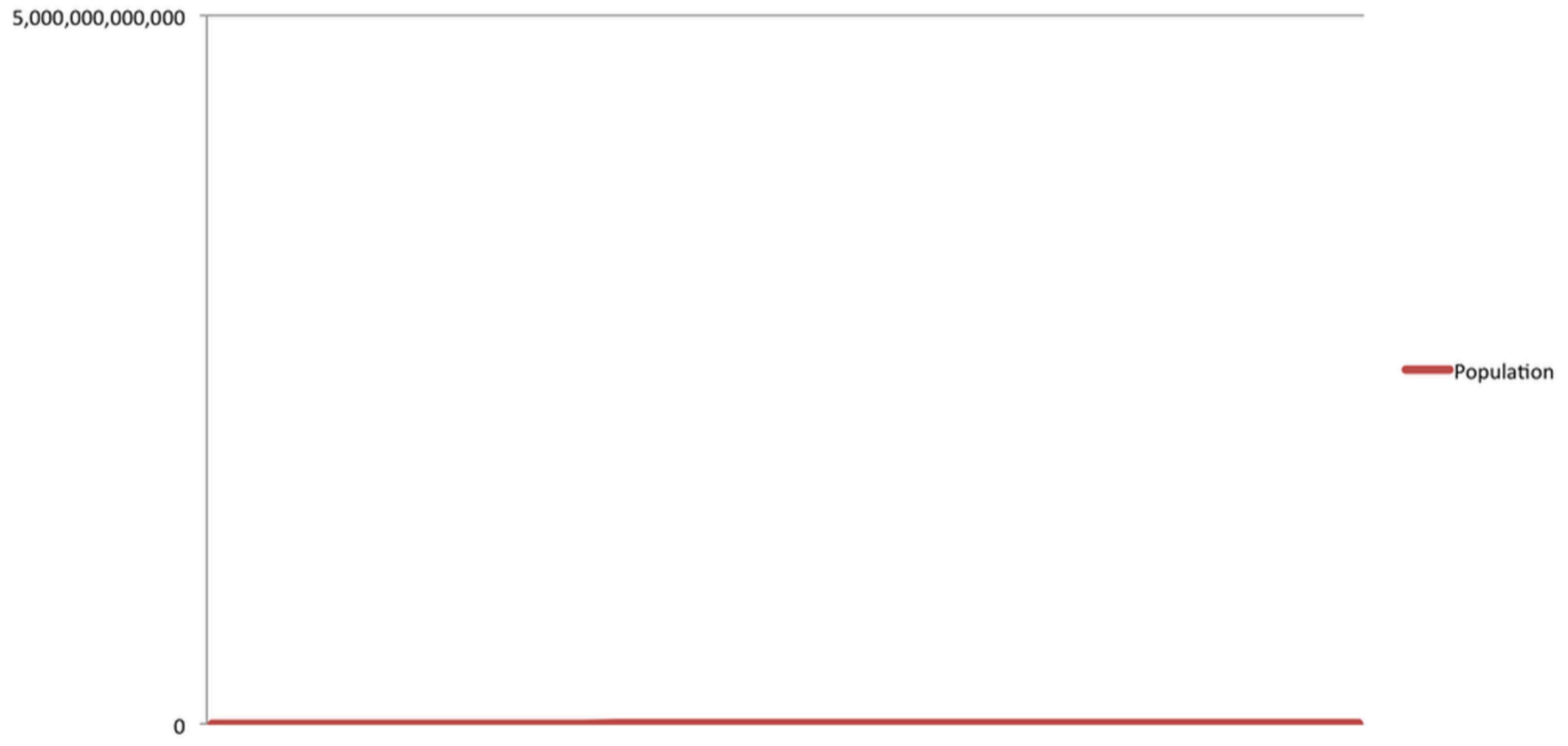
Federal Debt Held by the Public, as a share of GDP



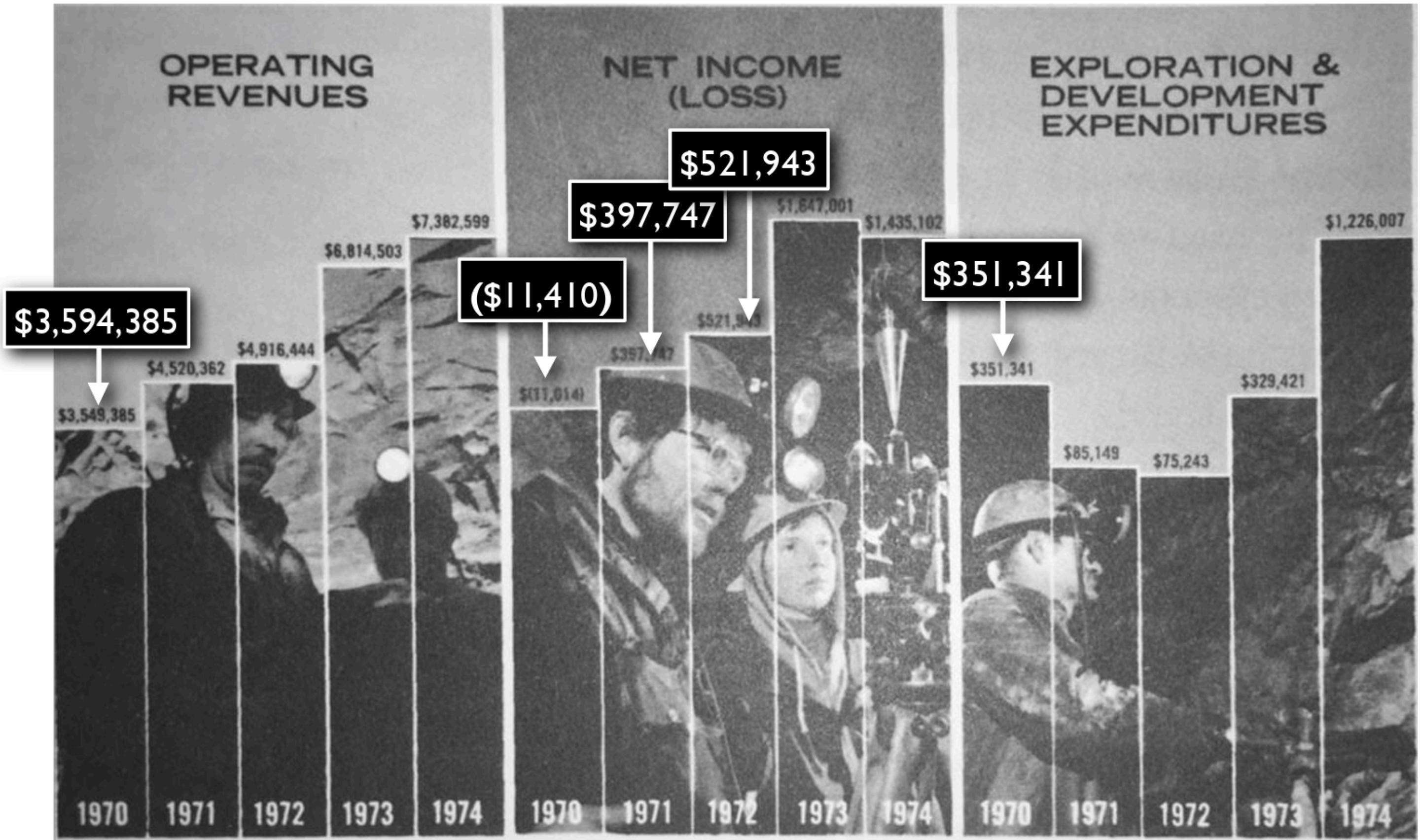
Source: CBO, 2014

@SeanMcElwee

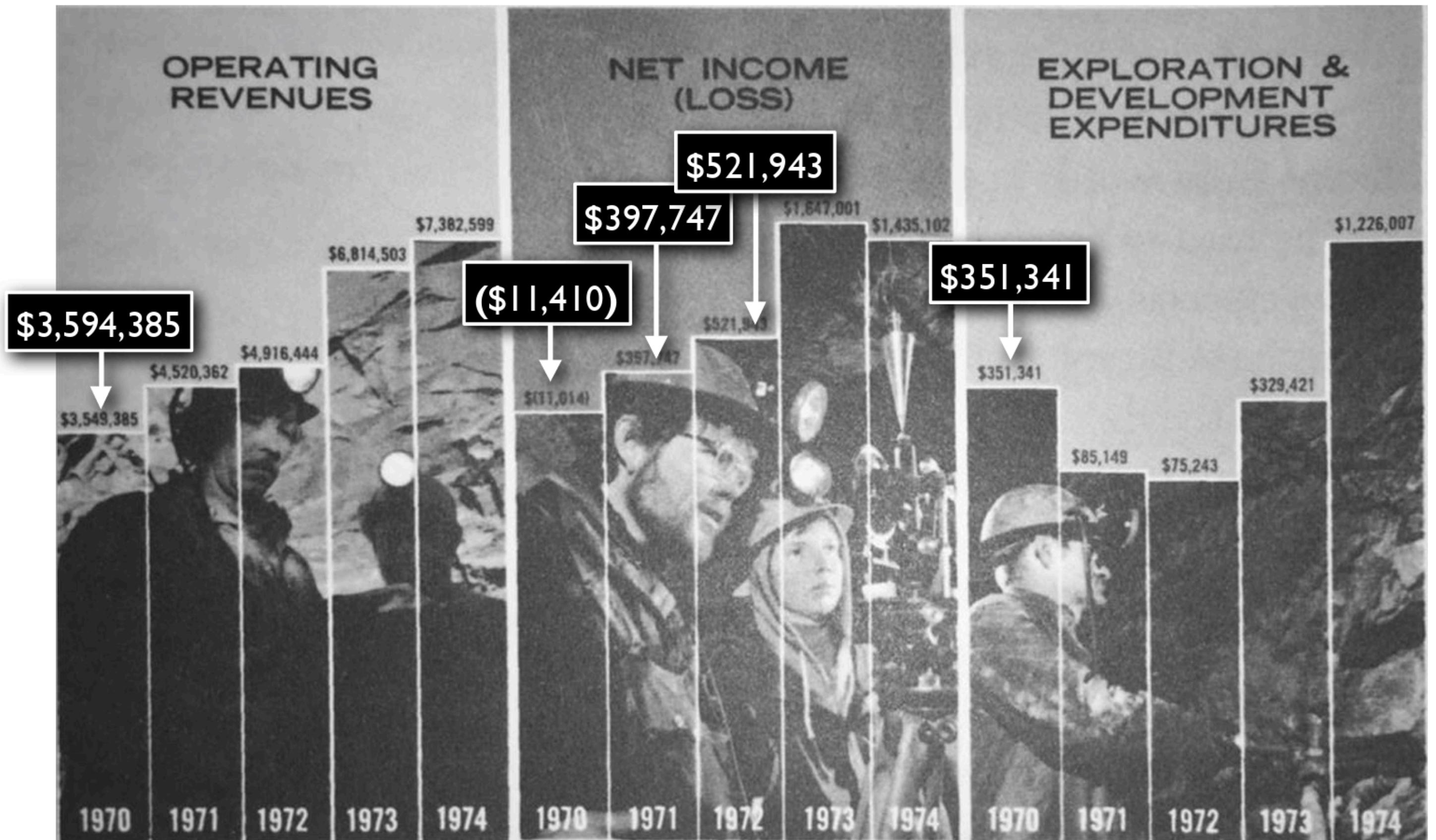
World population from 10,000 BC to present



Missing Scale

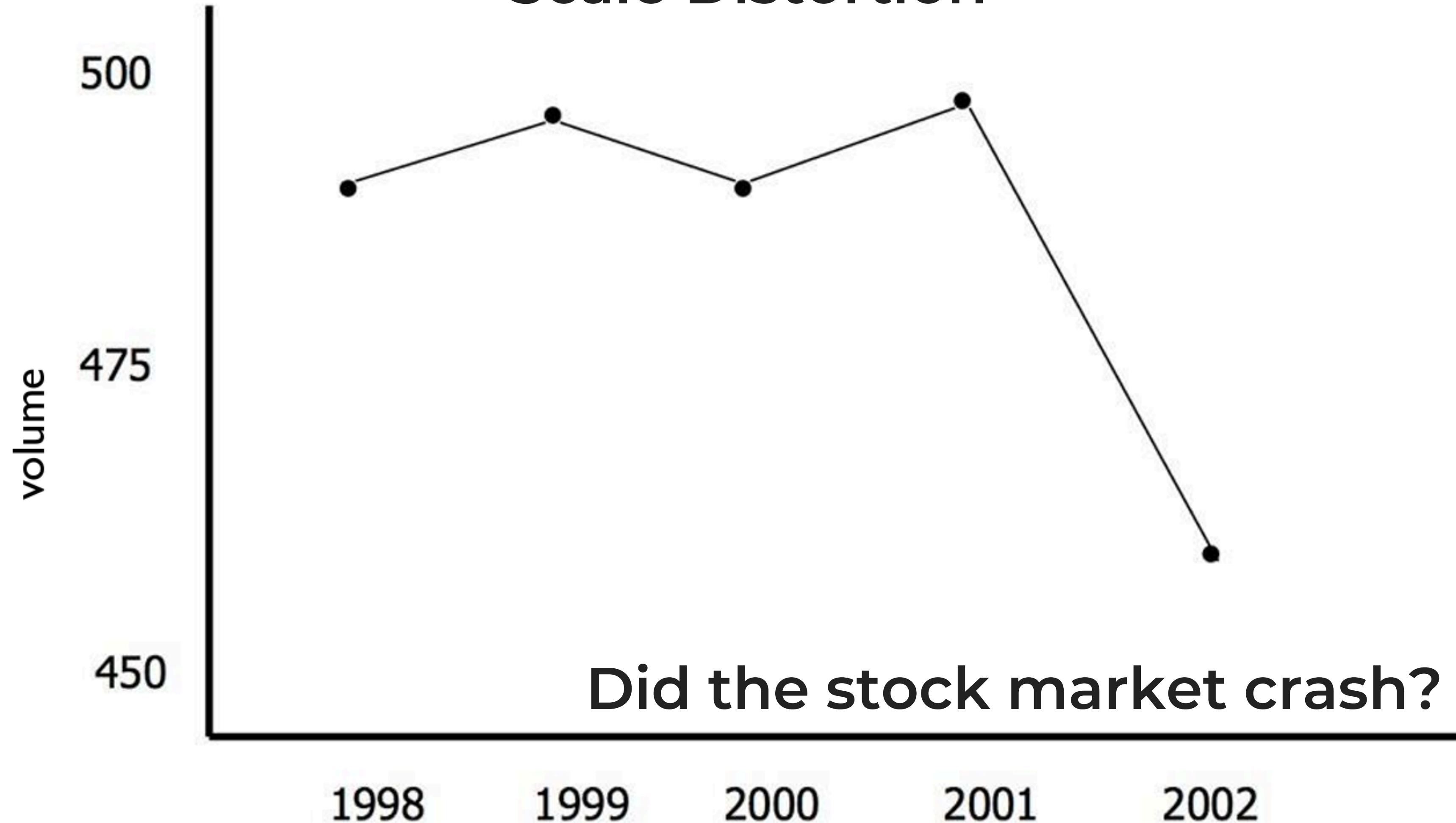


What's the baseline?

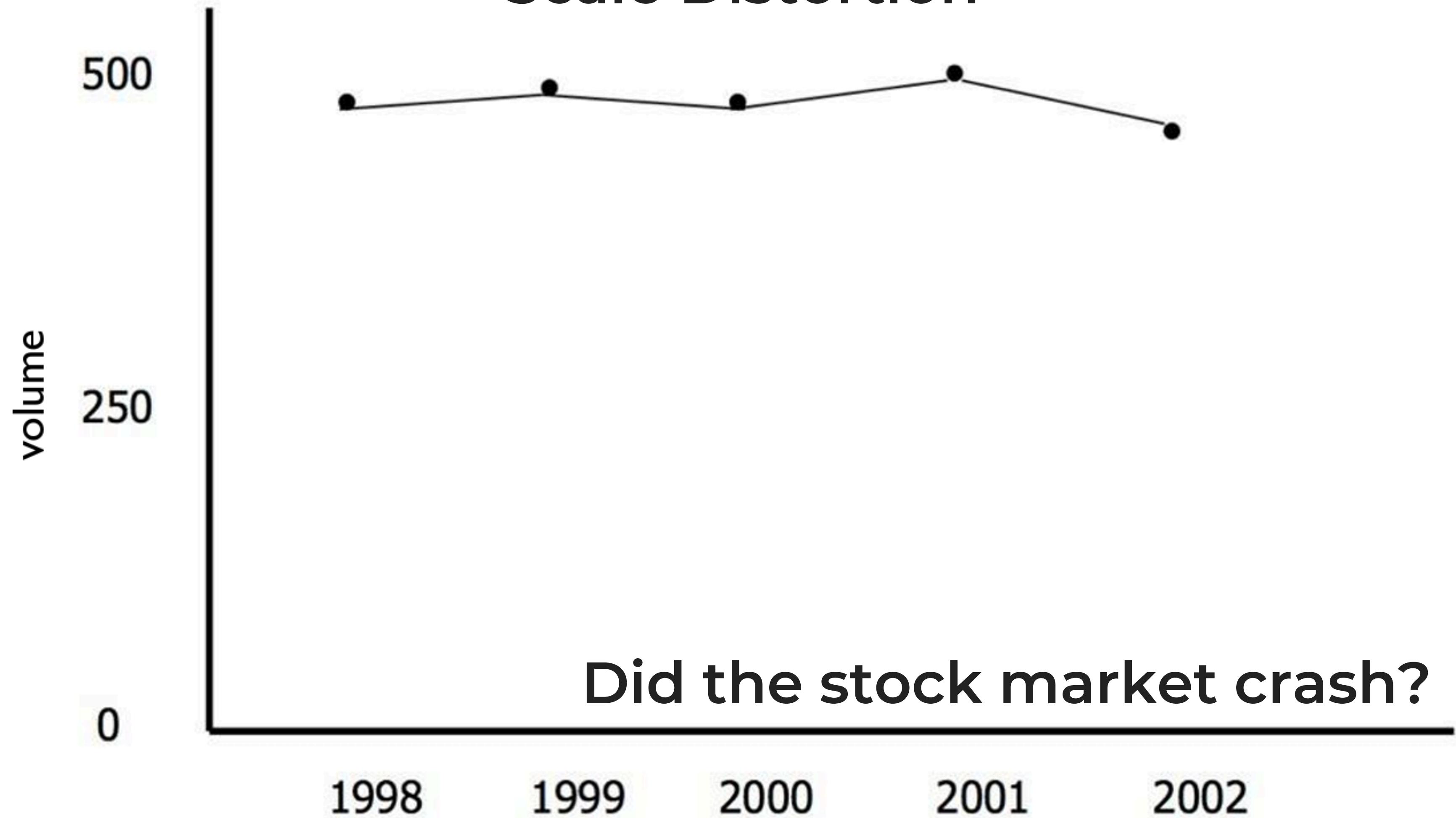


- **\$4.2 million**

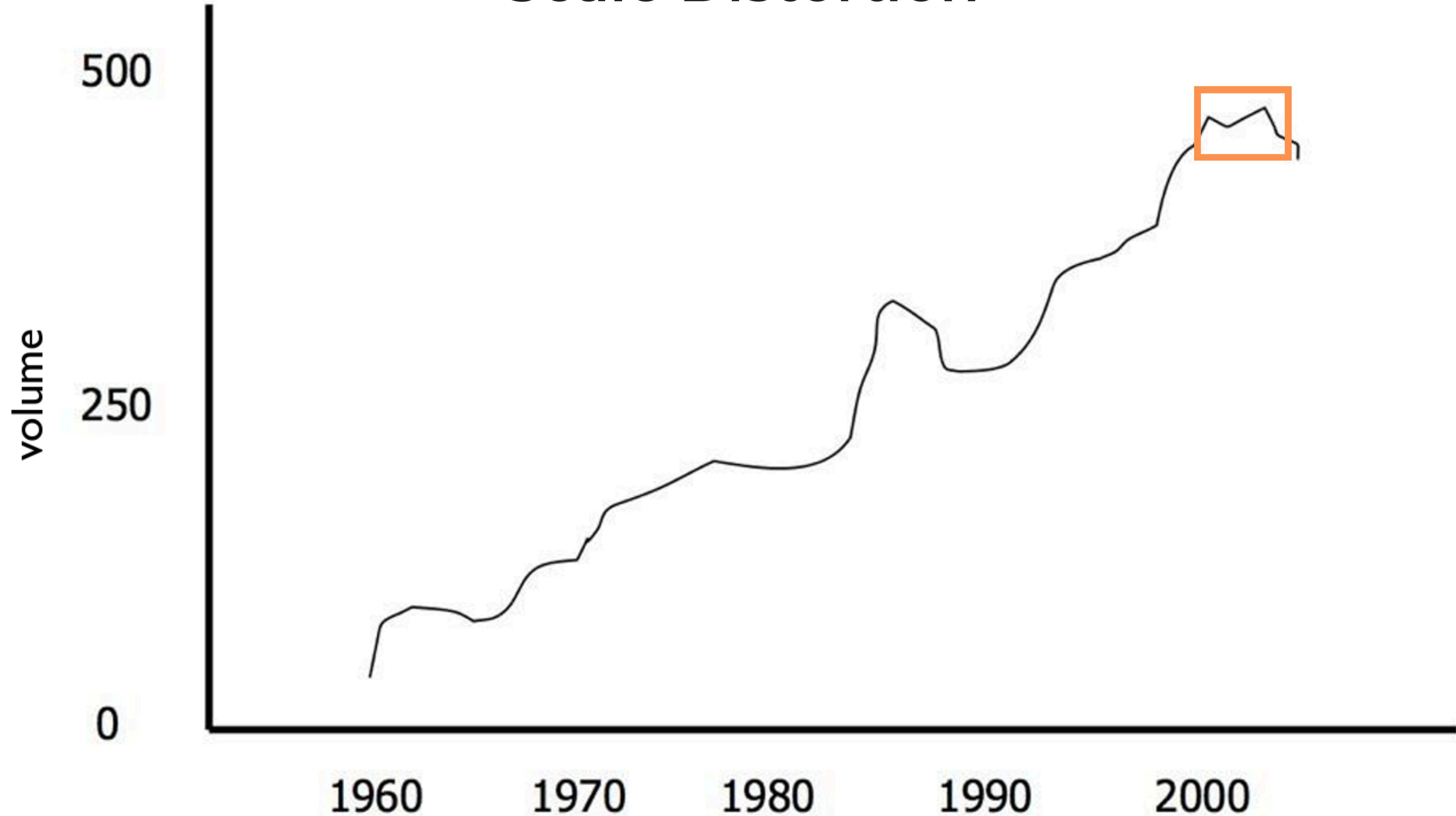
Scale Distortion



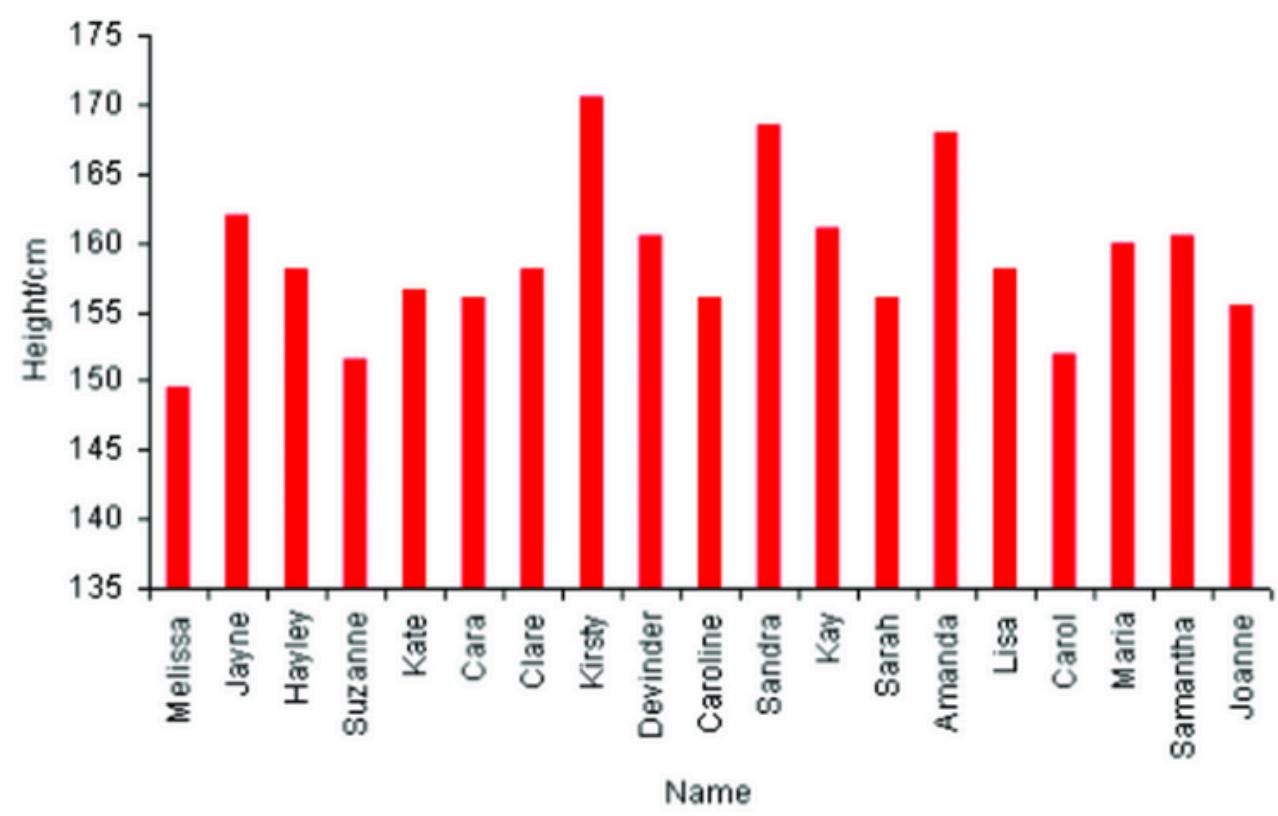
Scale Distortion



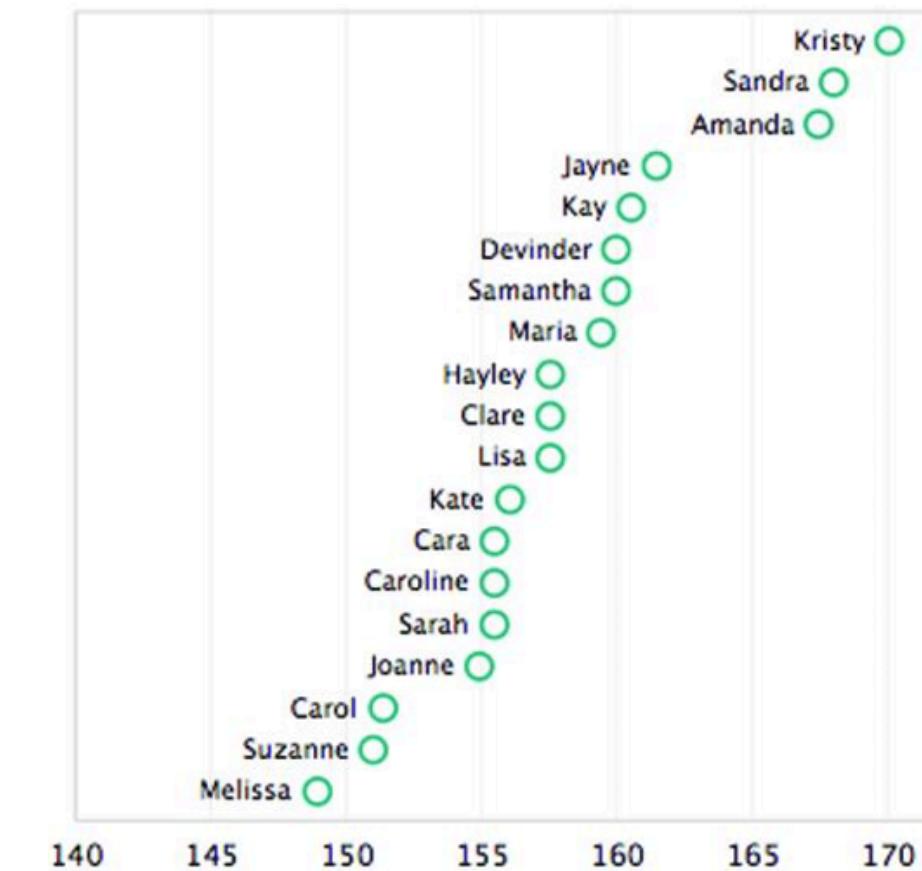
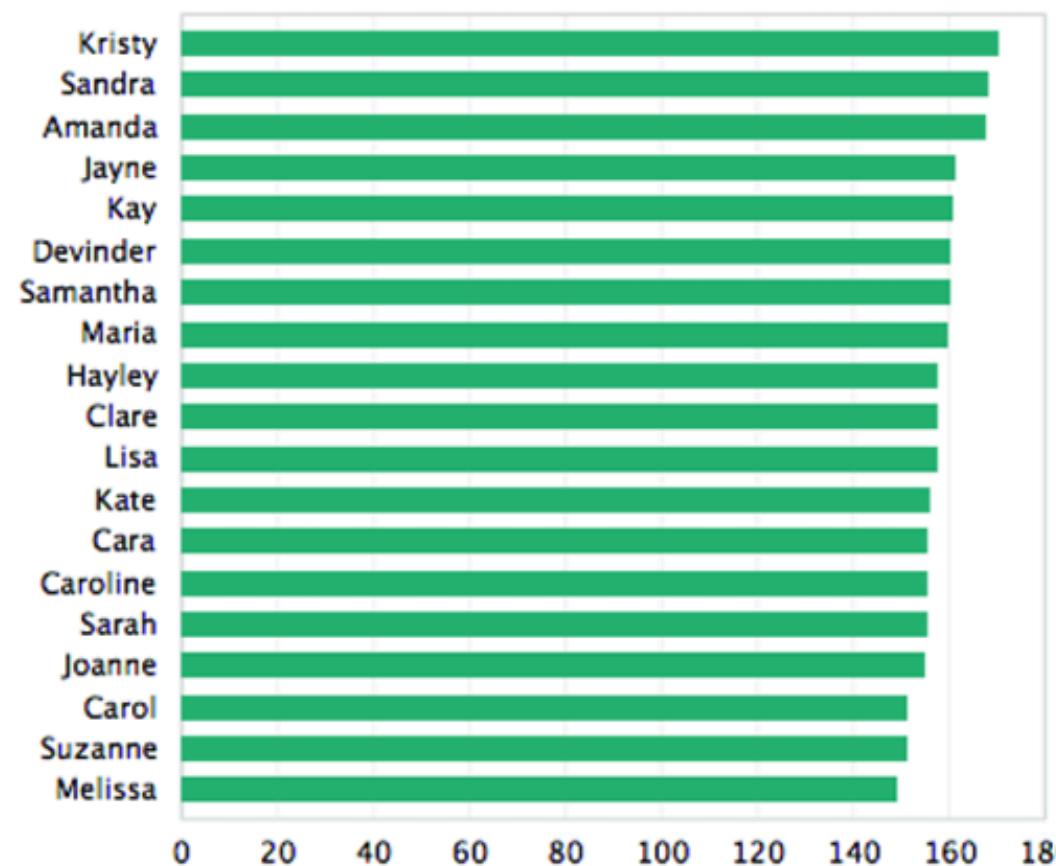
Scale Distortion



Truncated Axes



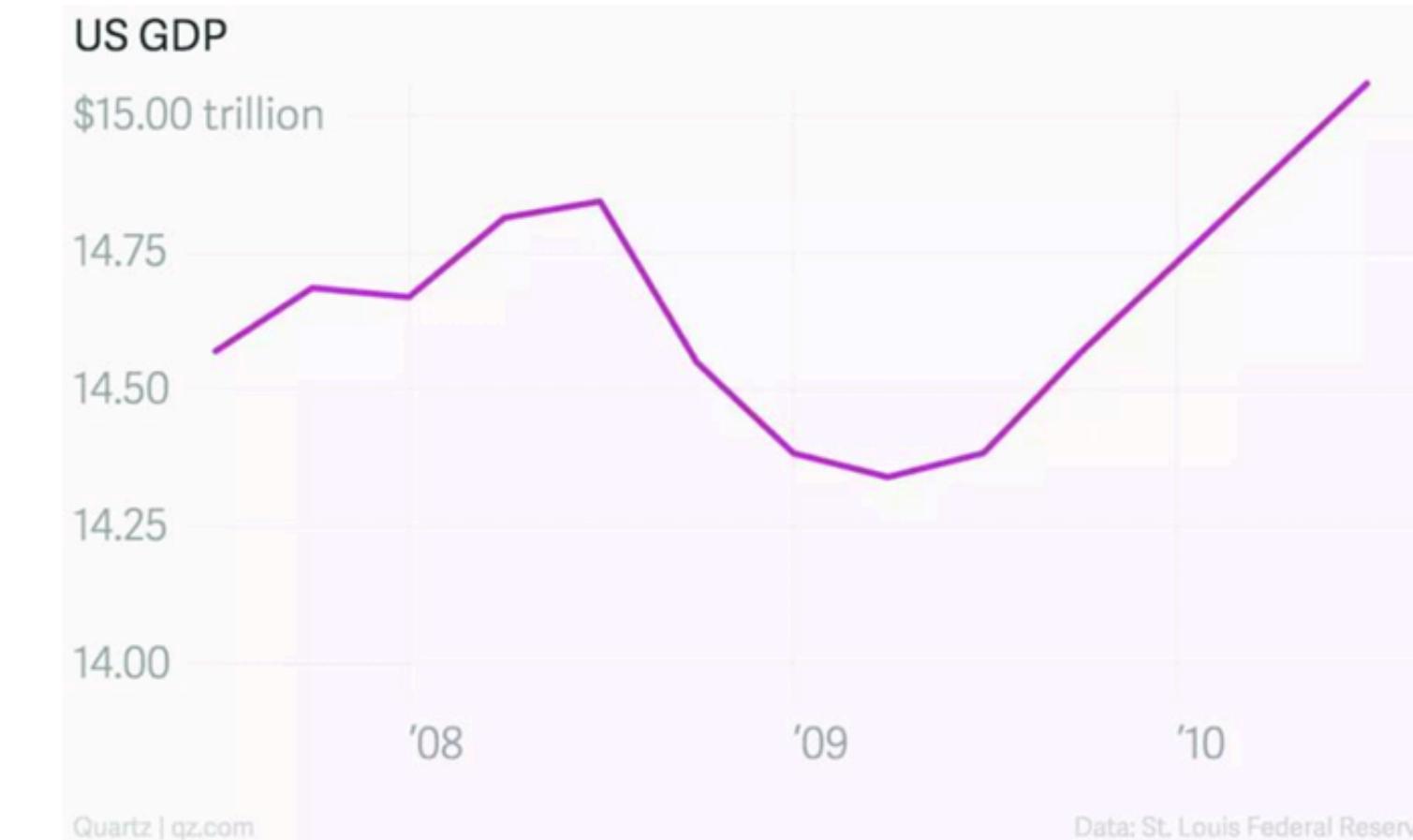
Compare Proportions →
↓
Compare Relative Position



Truncated Axes

You might want to truncate your axes:

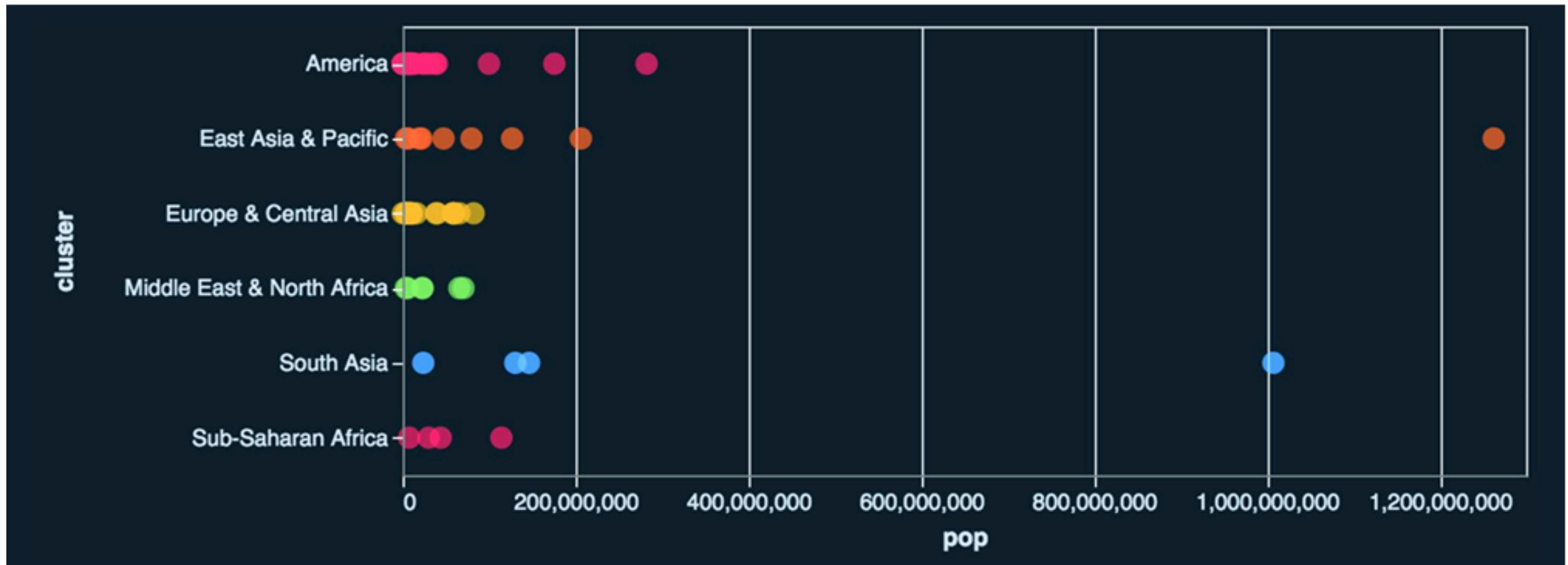
- if the zero baseline doesn't make sense for your data
- to emphasize the relative position comparisons
- if it is the norm (e.g., stock charts)



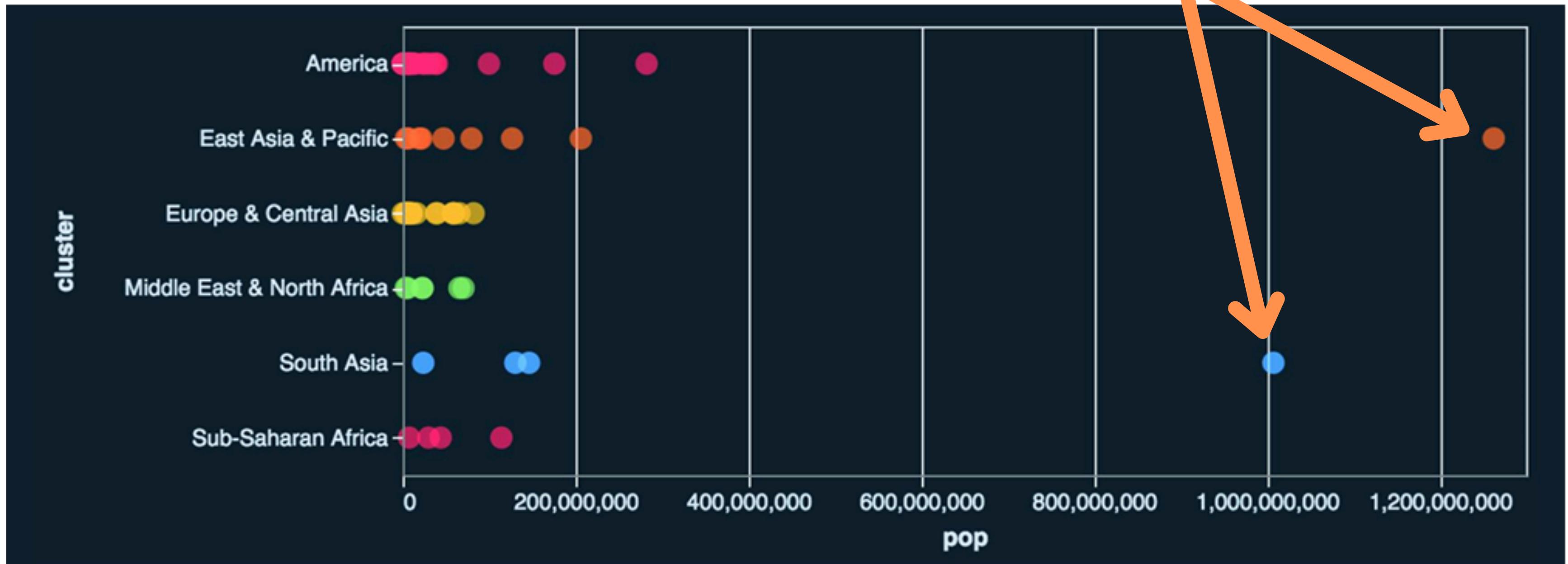
How do we avoid distortion/ambiguity?

Visualize Outliers Wisely

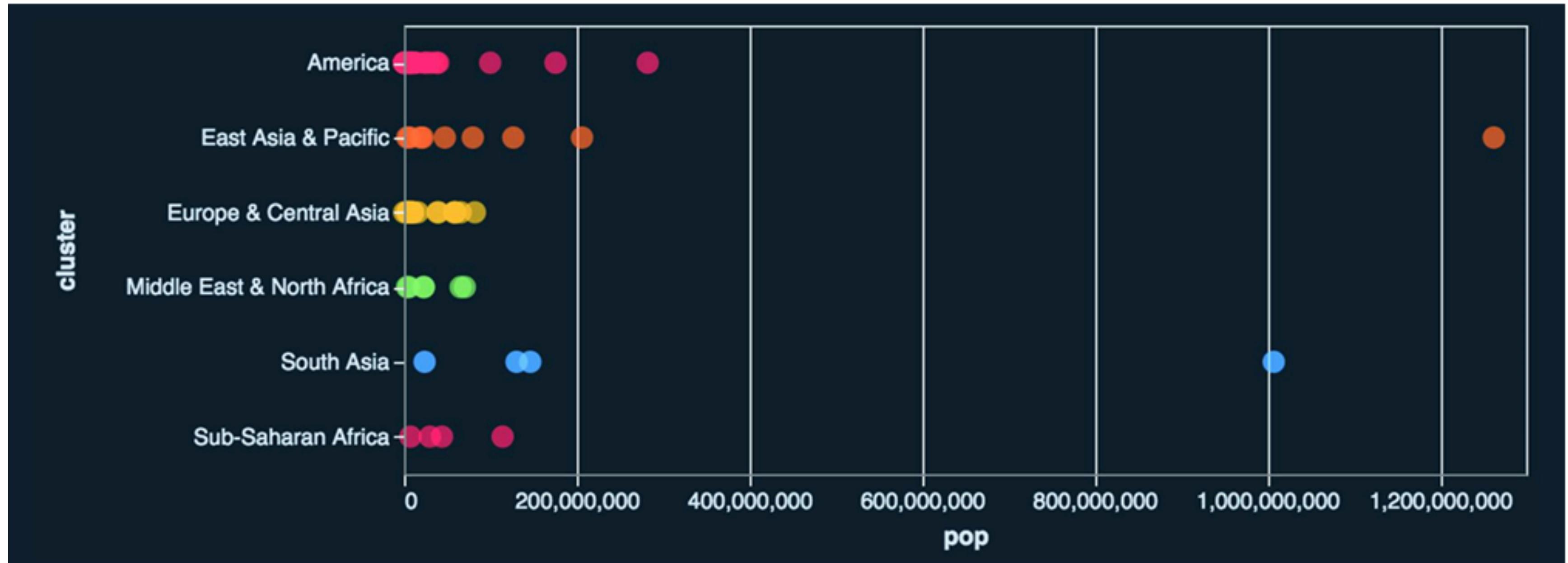
Most of our data is here



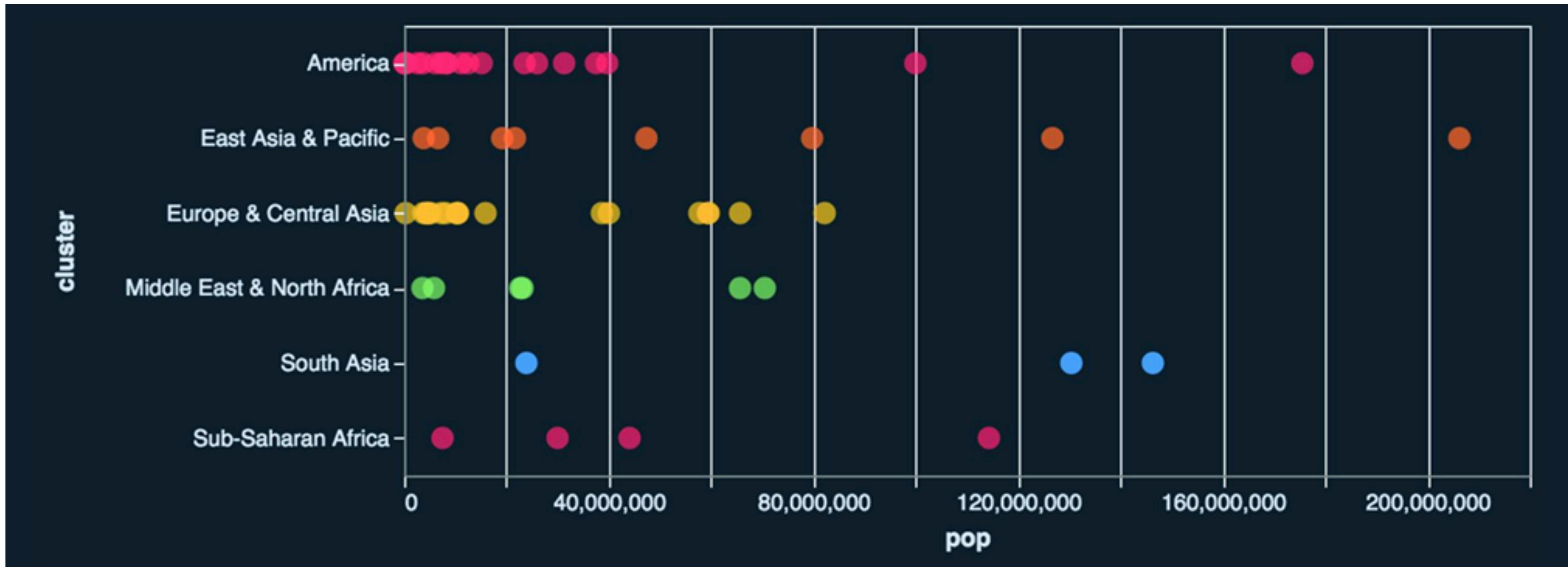
How do we handle these outliers?



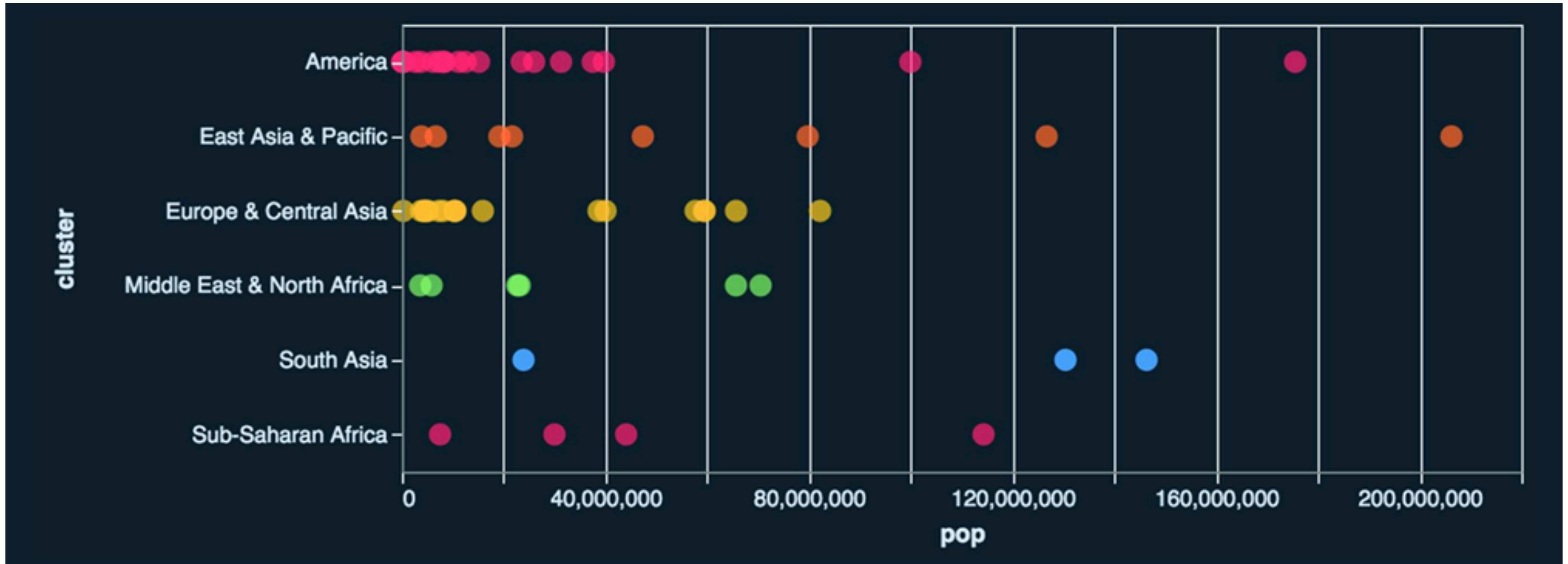
Option 1: Clip them out



Option 1: Clip them out

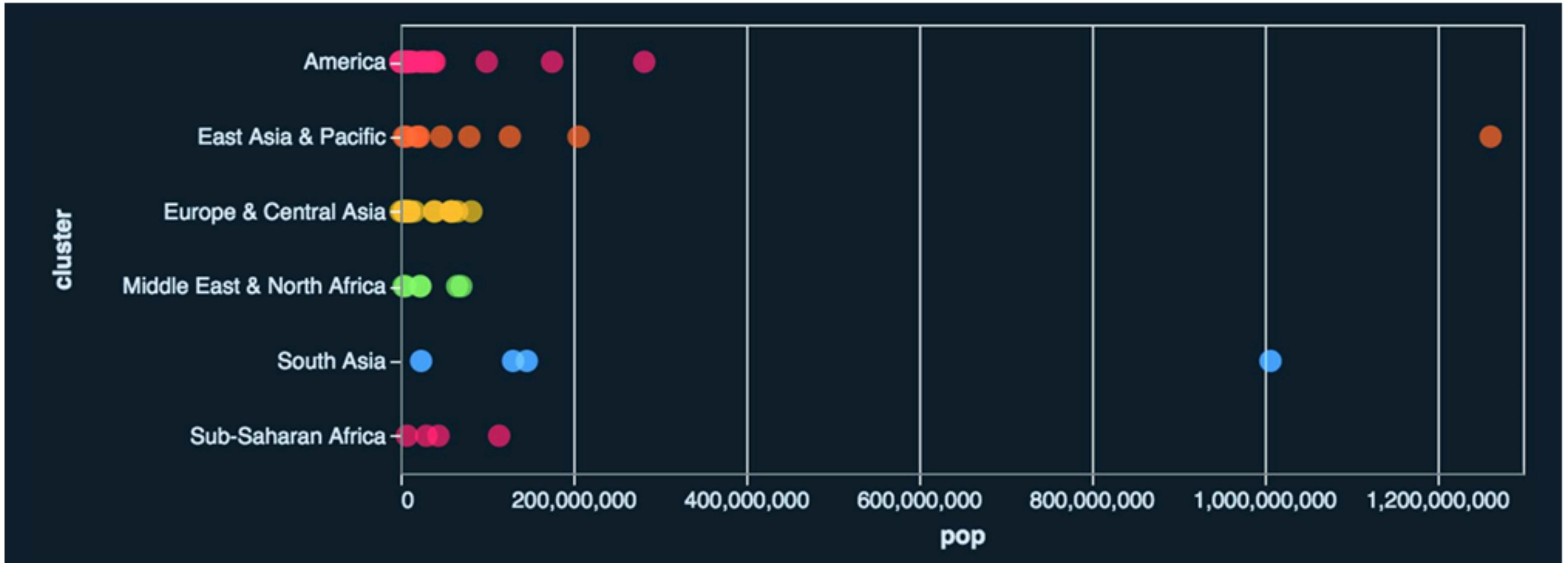


Option 1: Clip them out

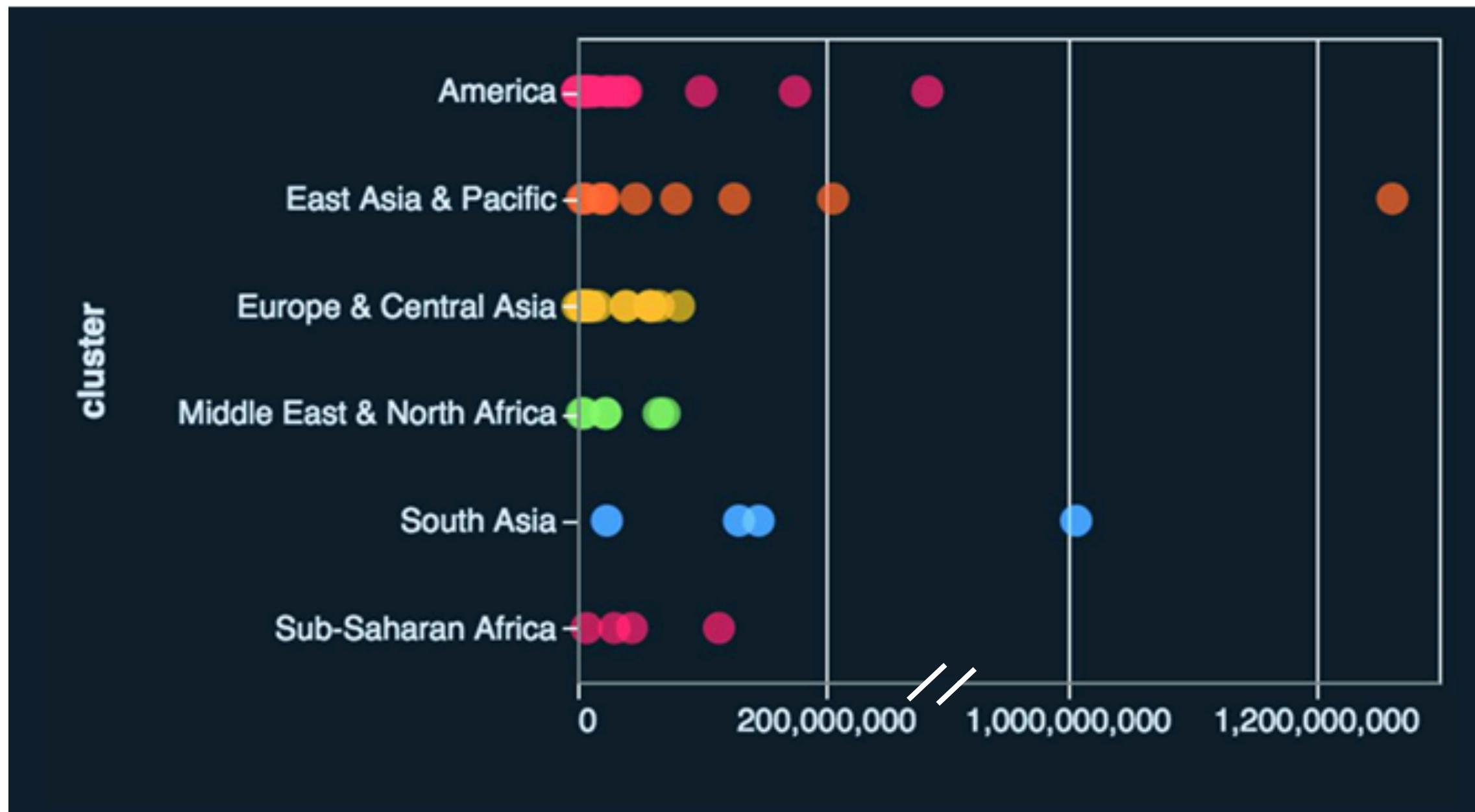


But, this ignores real data points

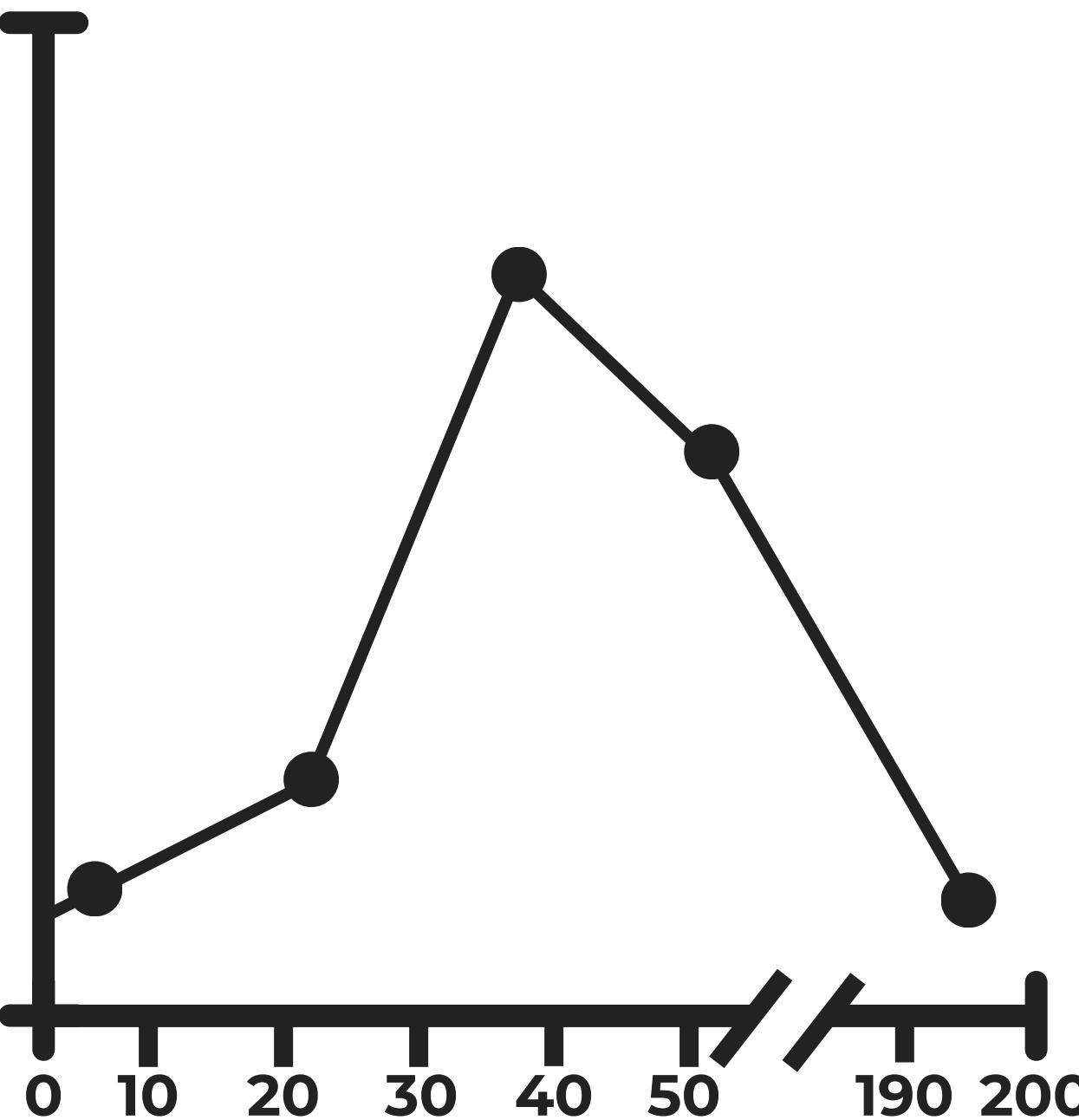
Option 2: Use scale breaks



Option 2: Use scale breaks

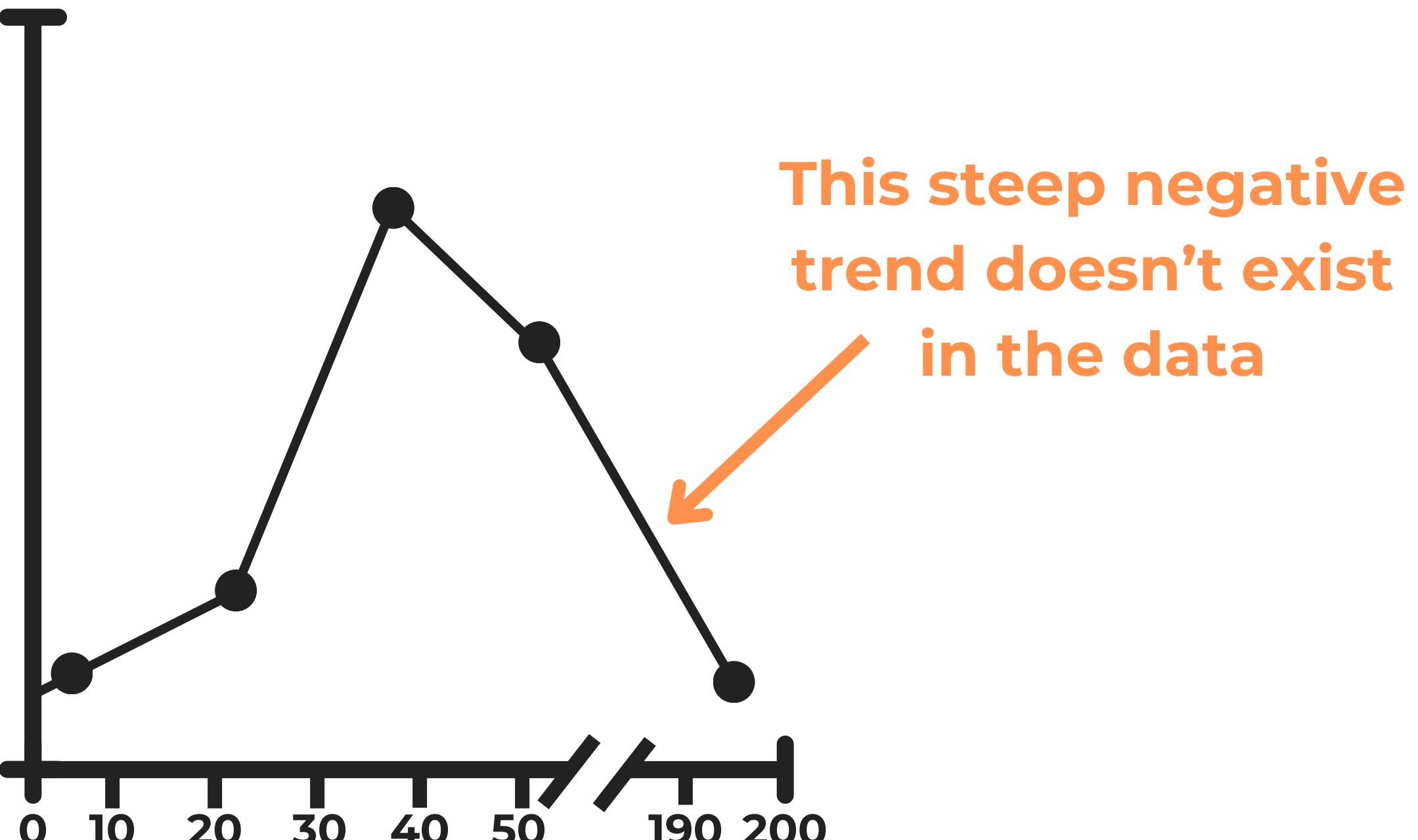


Option 2: Use scale breaks



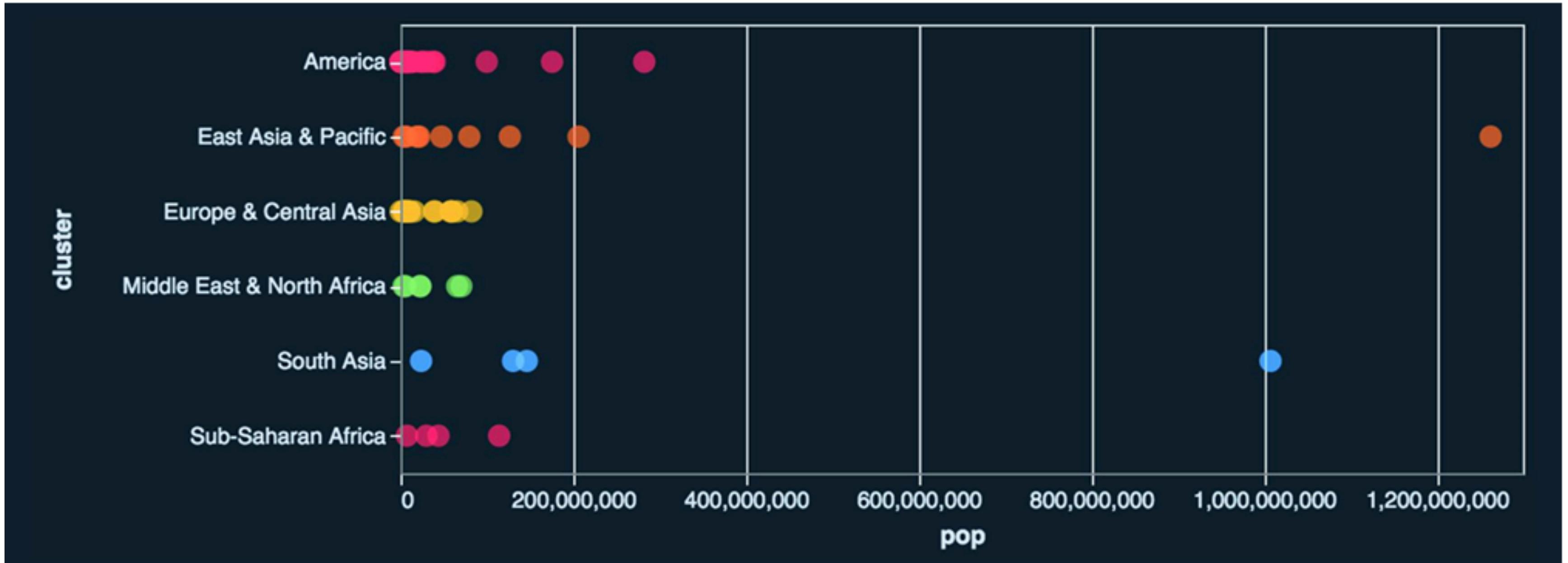
Might violate the expressiveness principle

Option 2: Use scale breaks

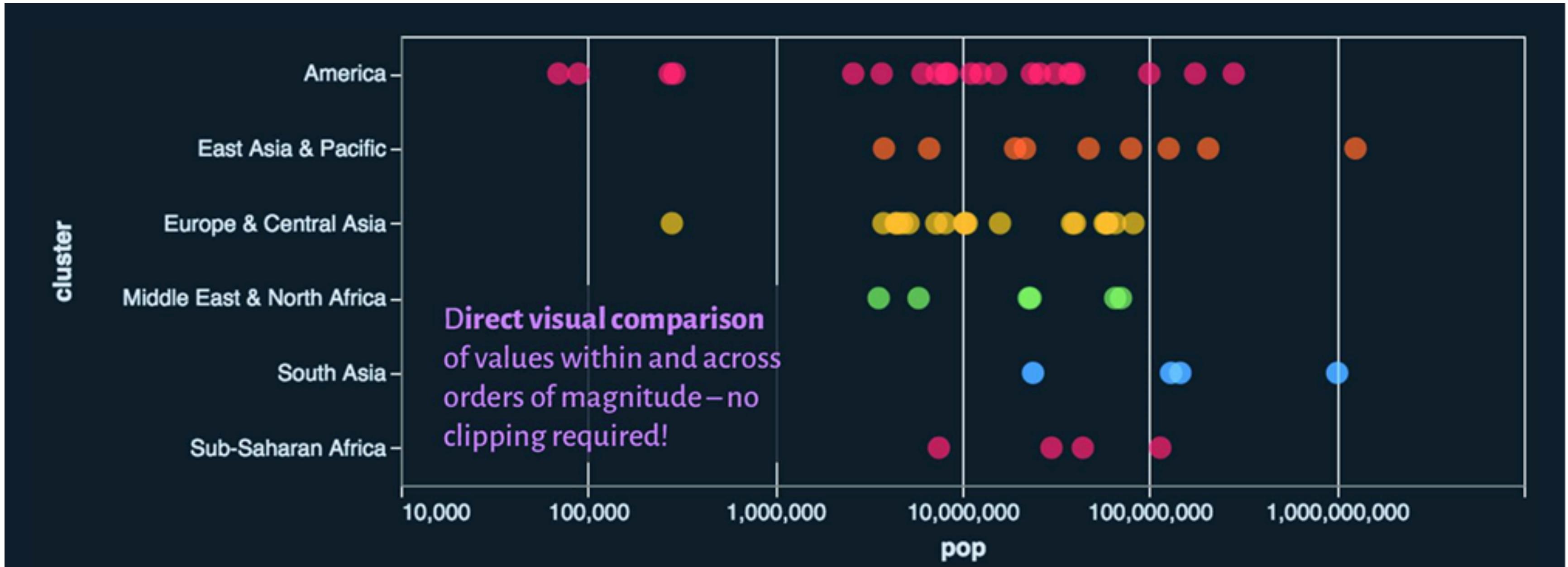


Might violate the expressiveness principle

Option 3: Use log scales



Option 3: Use log scales



Linear Scale



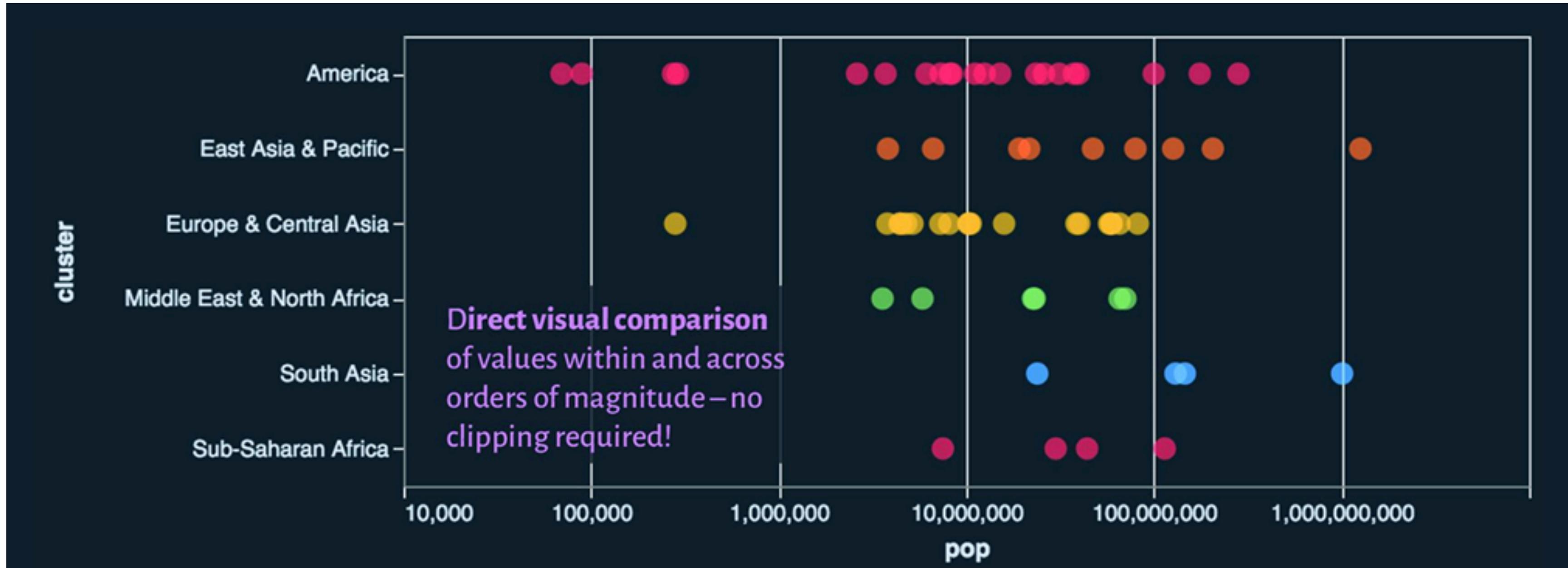
Log Scale



- Absolute change
- 10 visual units (pixels) represent 10 additional data units

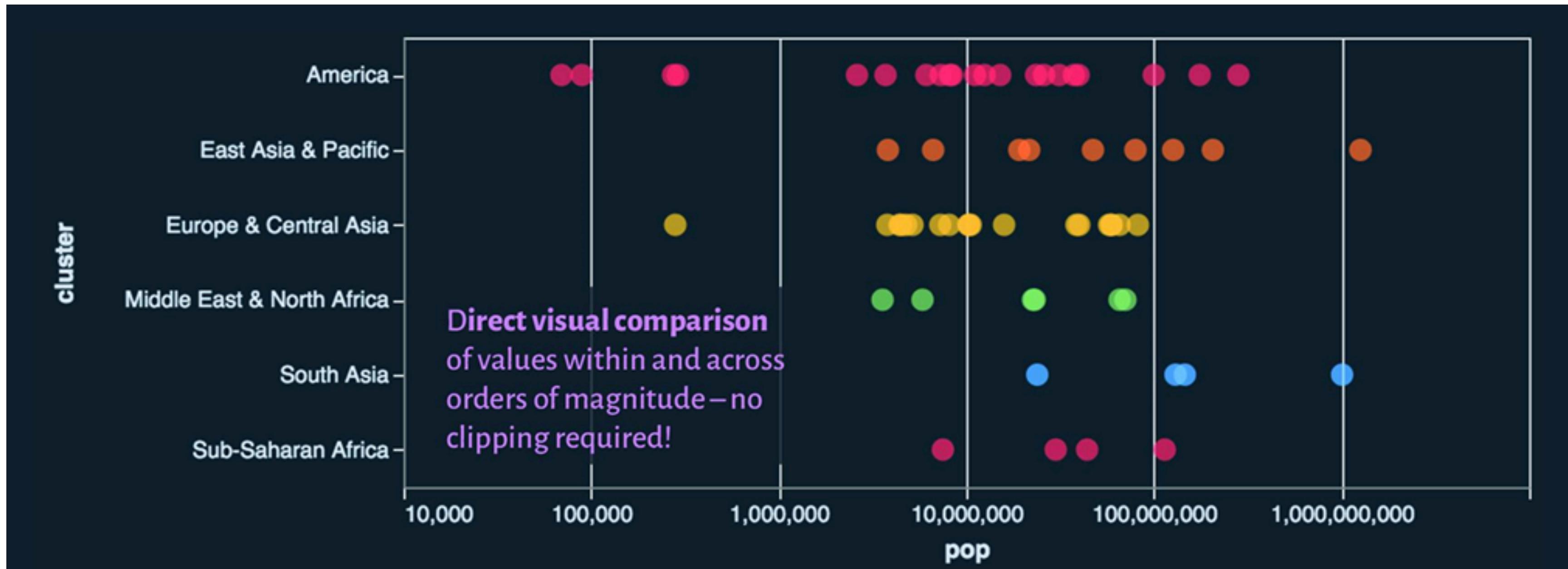
- Percentage change
- 10 visual units (pixels) represent a multiplication of 10 data units

Option 3: Use log scales



Only works for positive, non-zero values

Option 3: Use log scales

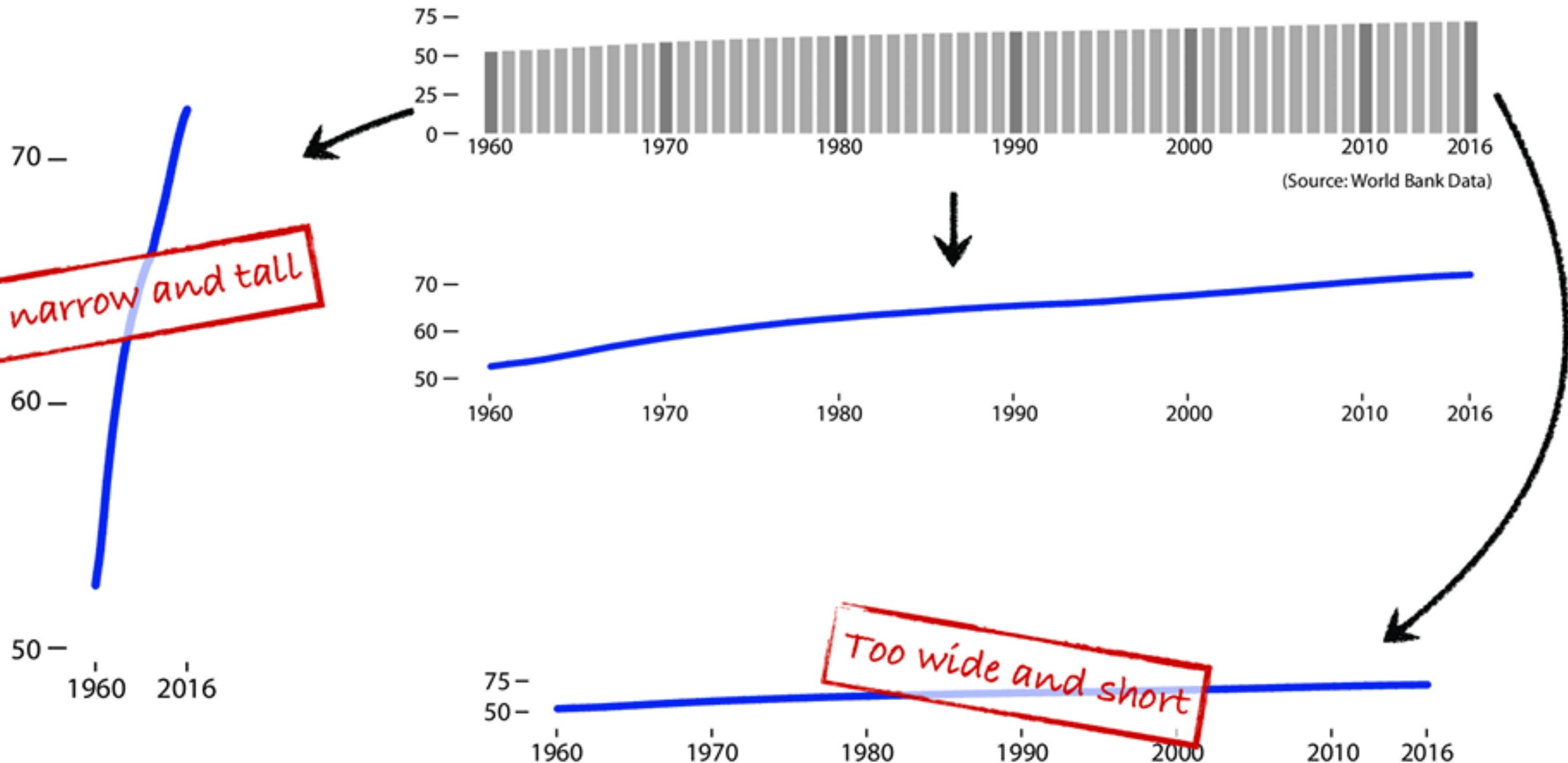


Viewers might be less familiar with log scales

How do we avoid distortion/ambiguity?

Use the Right Aspect Ratio

Average world life expectancy at birth (years)



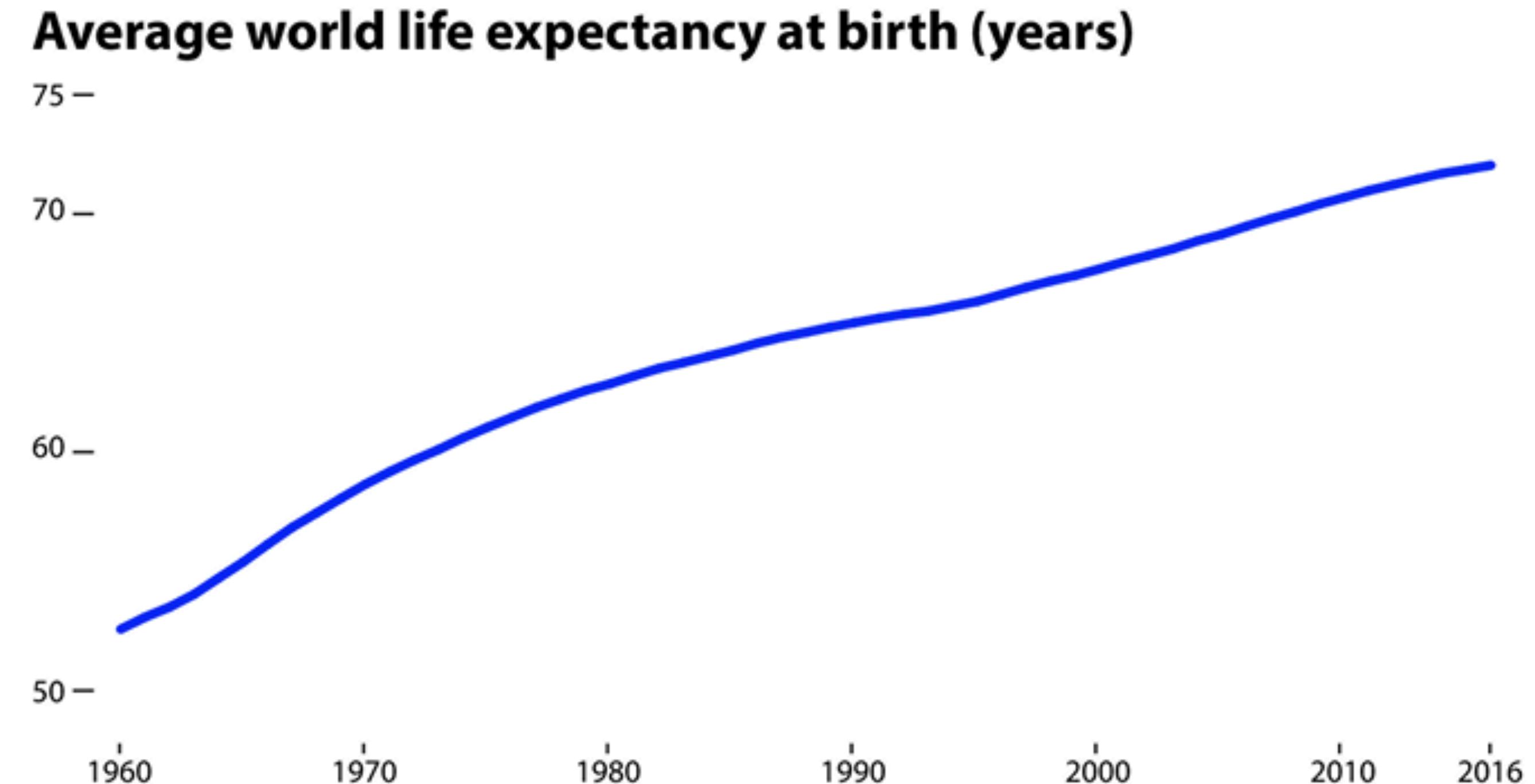
[Alberto Cairo. *How Charts Lie*, 2019]

How do we pick the correct aspect ratio?

Approximate the aspect ratio of the chart to match the trend in the data

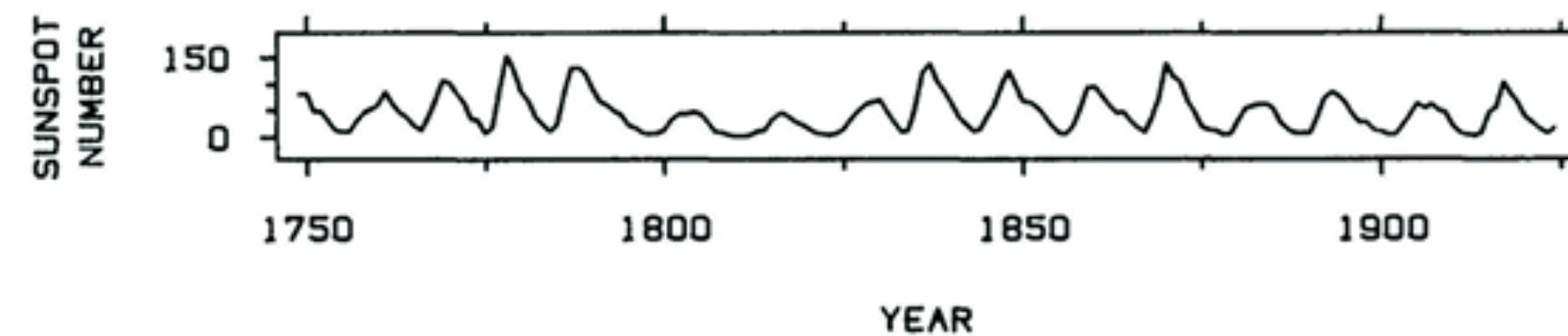
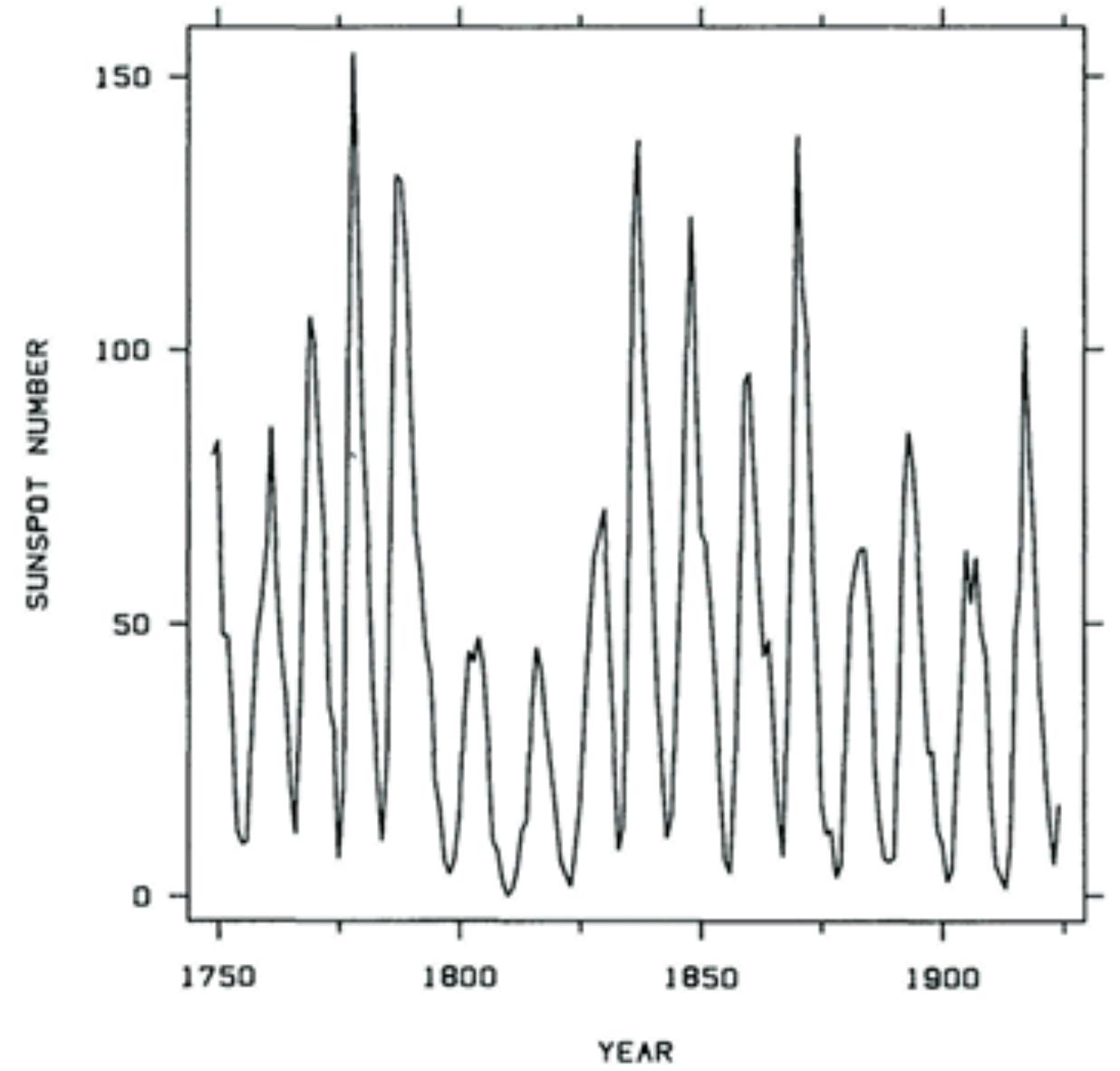
35% increase
in data $\sim 1/3$

use 3:1 aspect
ratio



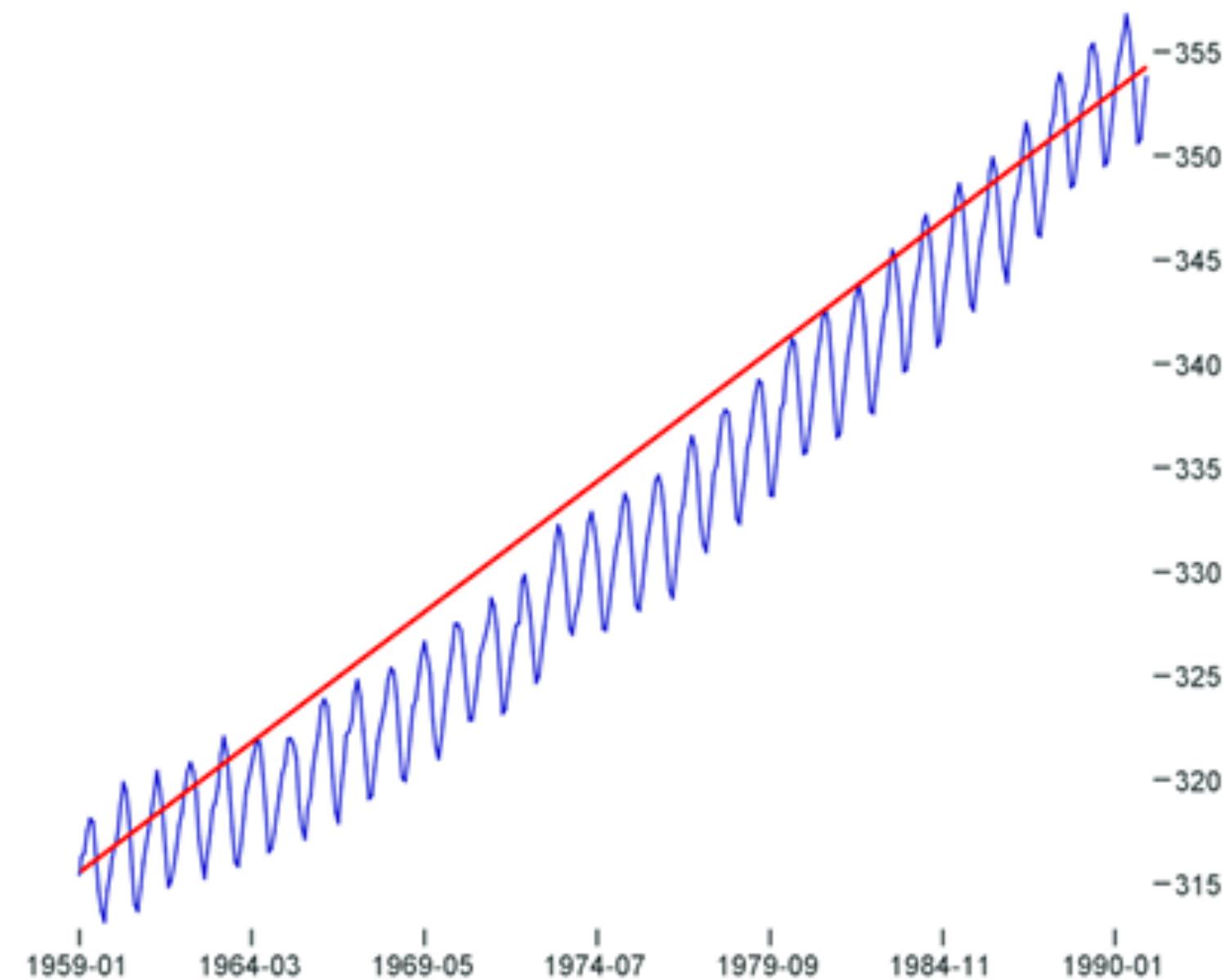
What if there isn't a single trend?

Choose an aspect ratio that makes the average line segment orientation 45°



What if there isn't a single trend?

Or choose an aspect ratio that makes the orientation of the fitted trend line 45°



How do we avoid distortion/ambiguity?

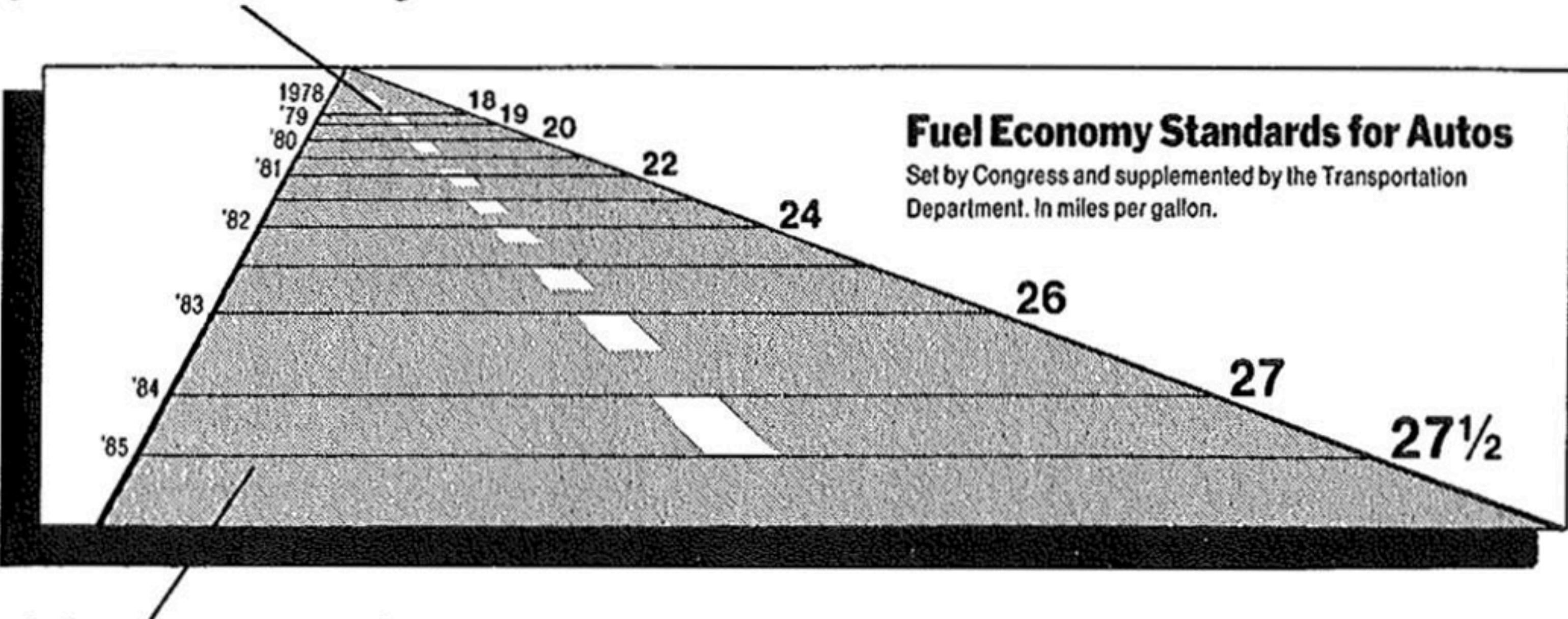
Consider the Lie Factor

What is the Lie Factor?

The representation of numbers (as physically measured on the visualization itself) should be directly proportional to the numerical quantities represented

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



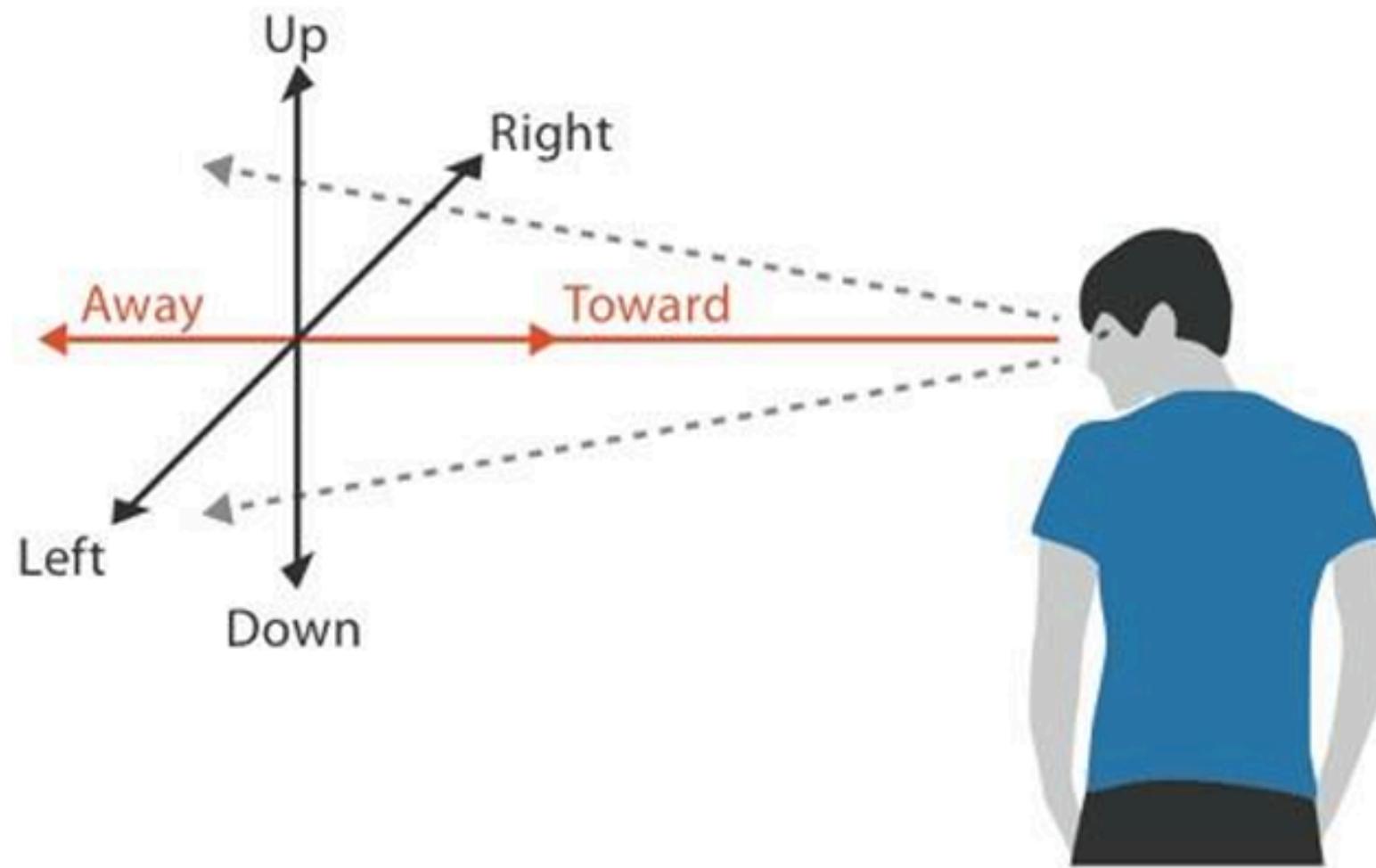
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

How do we avoid distortion/ambiguity?

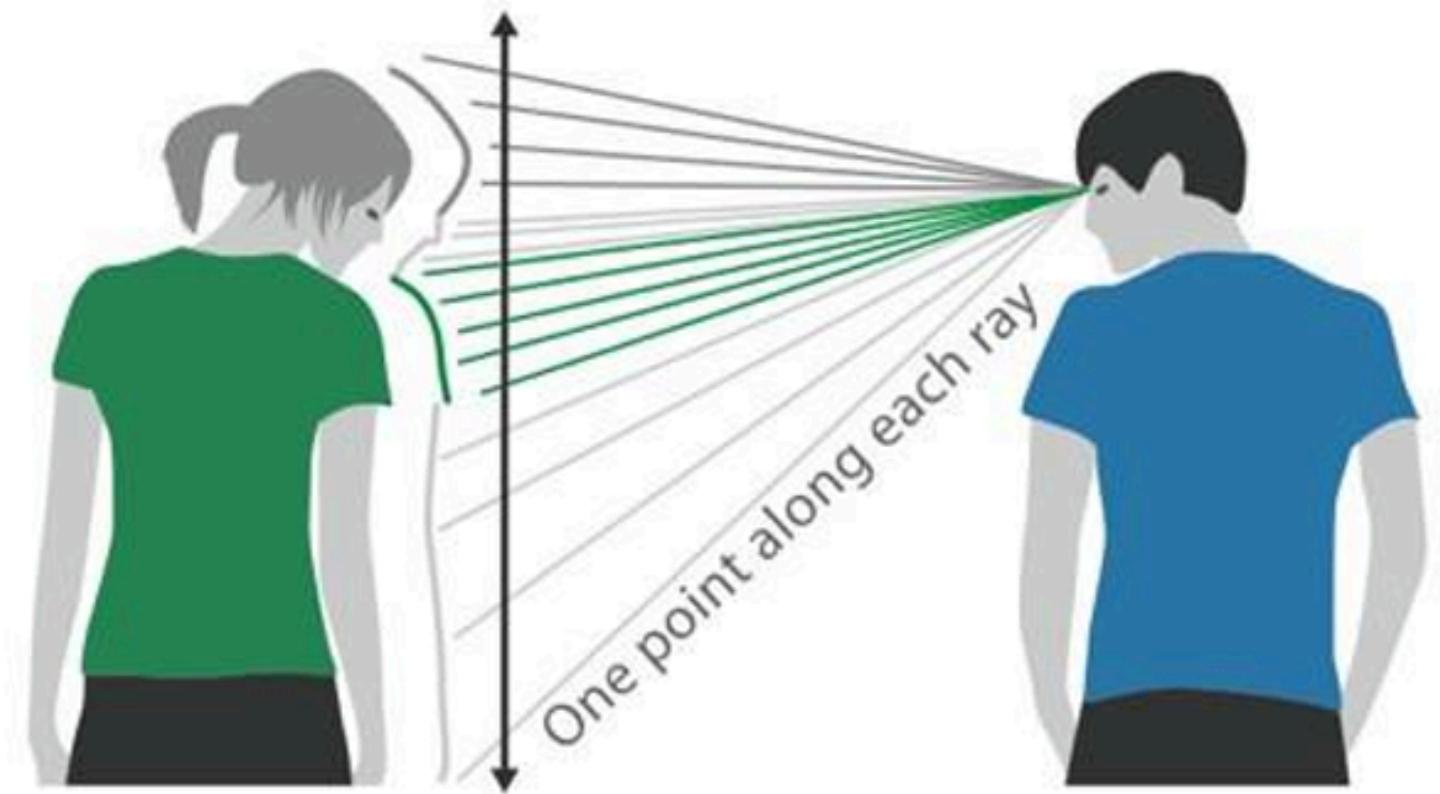
Avoid 3D Visualizations

Why avoid 3D?

We see the world as a ~~2.5D~~ 2.05D space



Thousands of points up/down and left/right

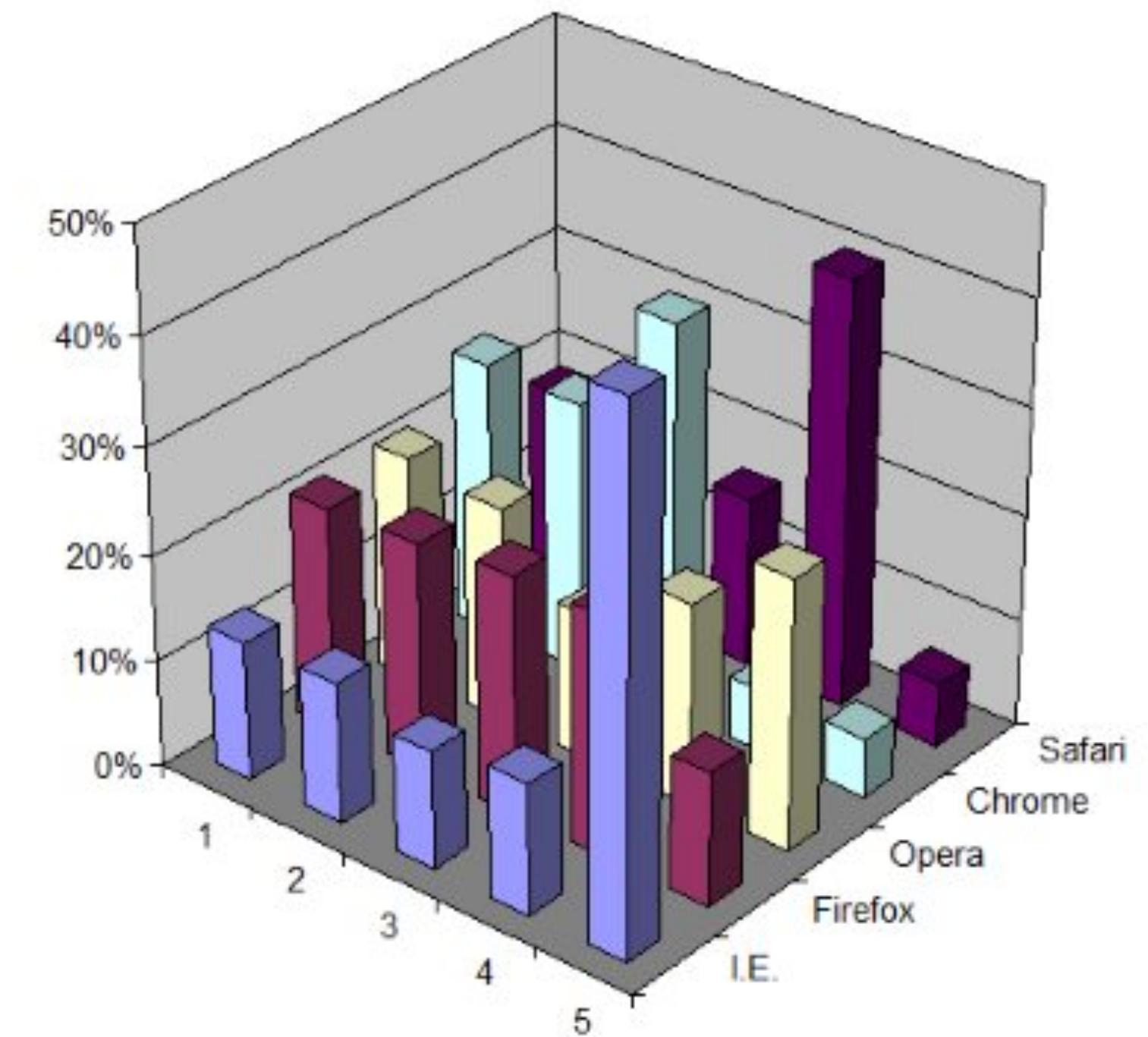
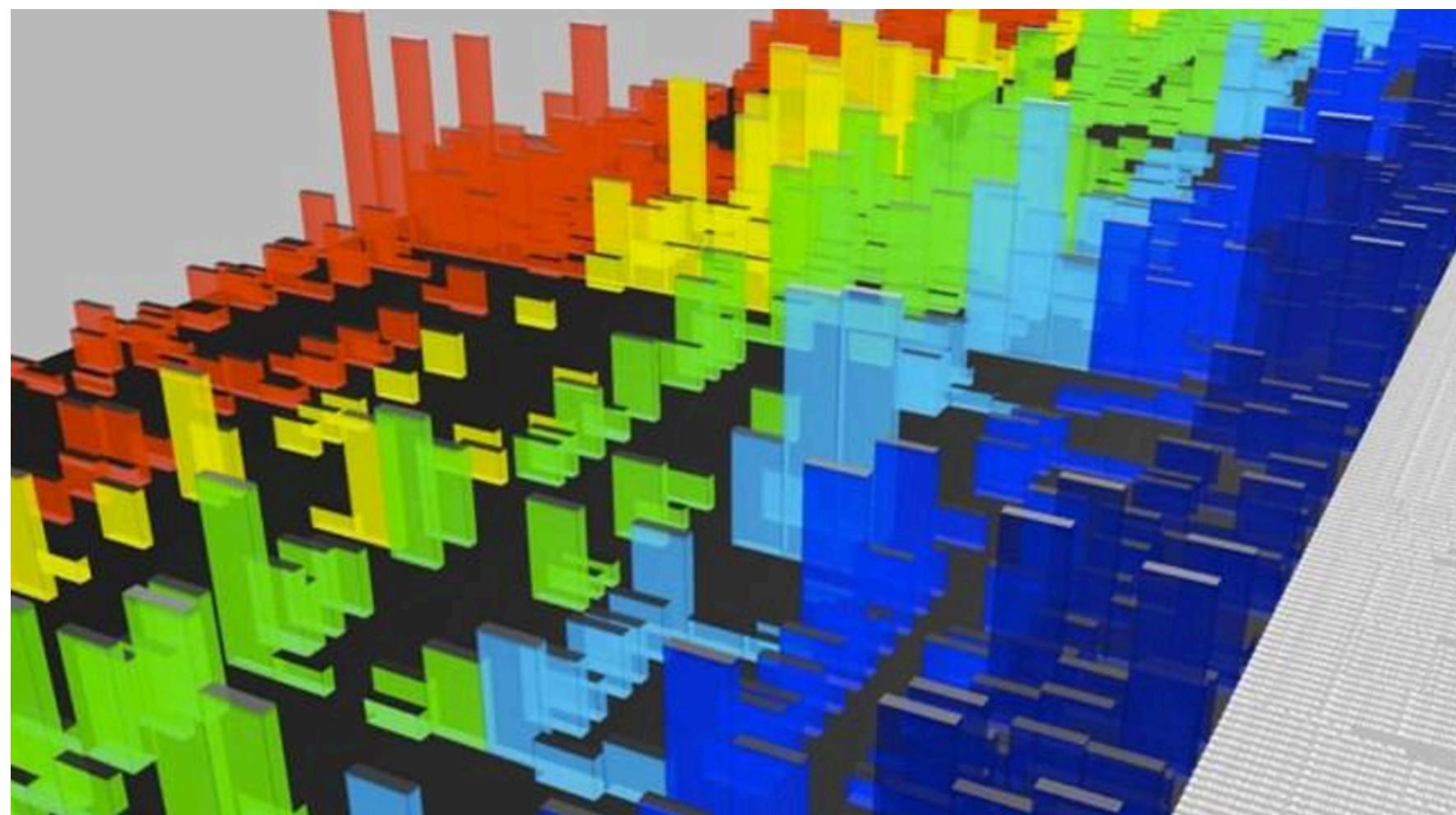


We can only see the outside shell of the world

Why avoid 3D?

Perspective interferes with color and size channels

Data is often occluded



How do we avoid distortion/ambiguity?

Visualize Time-Series Data Wisely

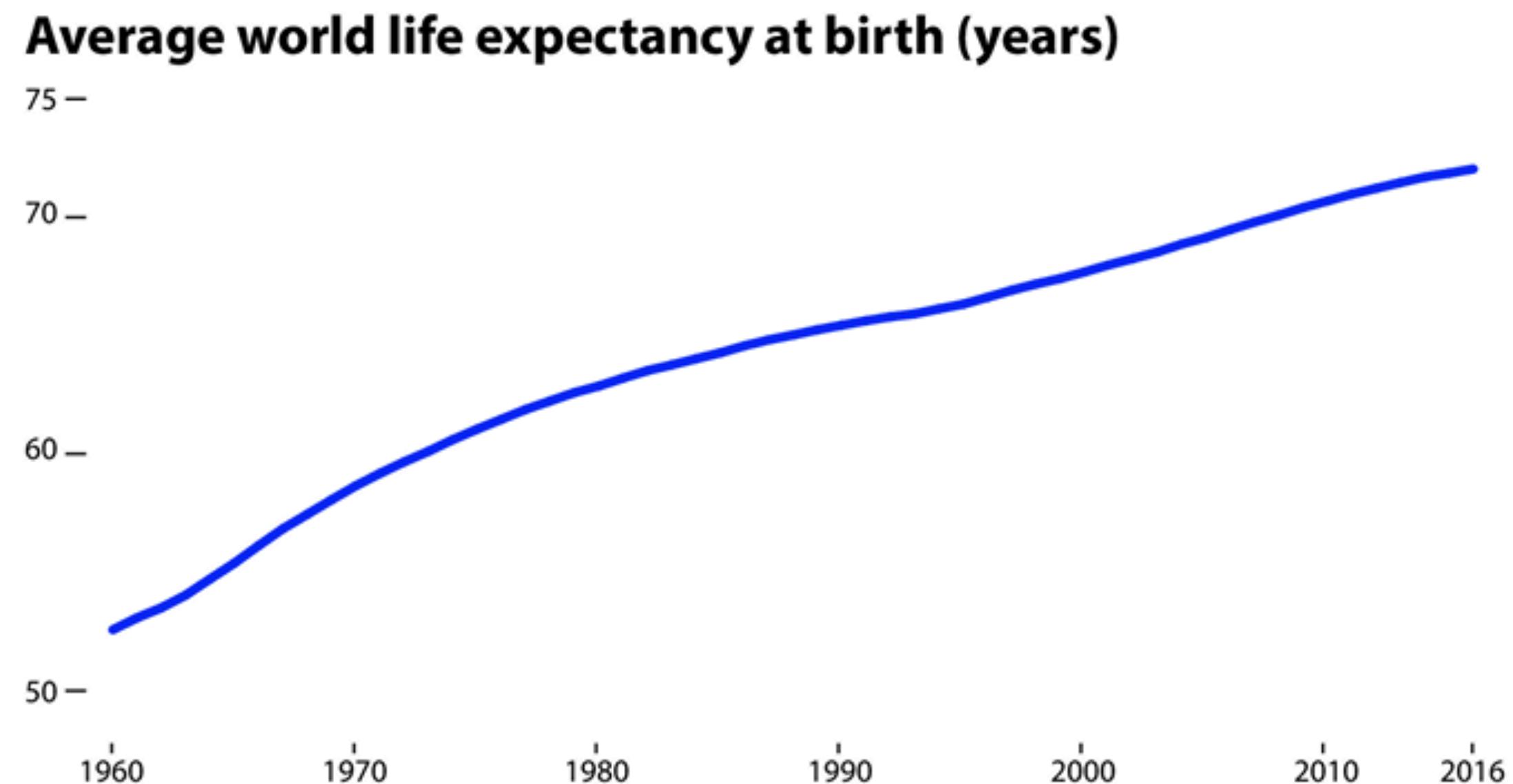
What is Time-Series Data?

Time-series data refers to data where each item represents a measurement at a specific point in time

- e.g., weather data, brain monitoring data, stock prices

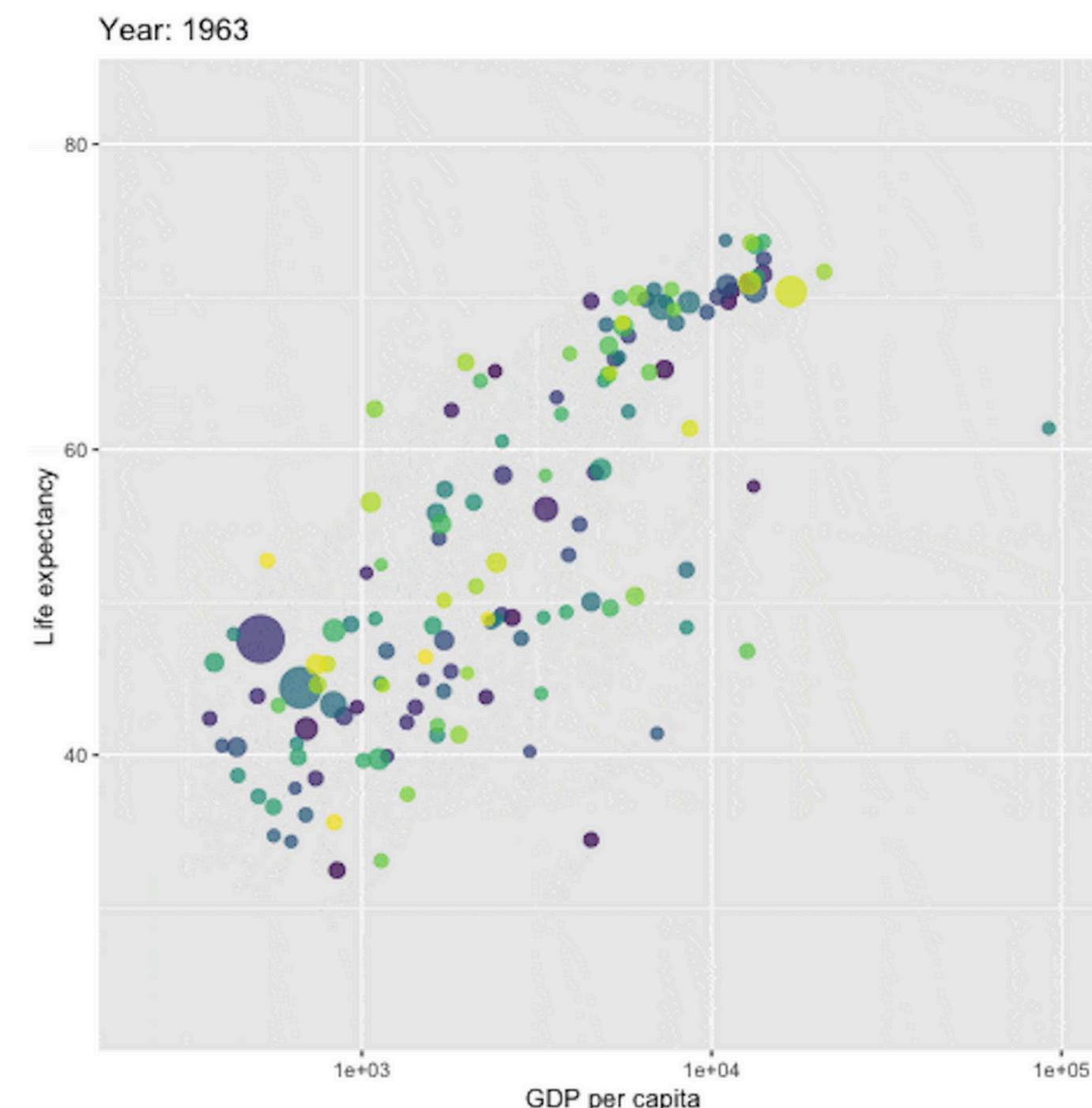
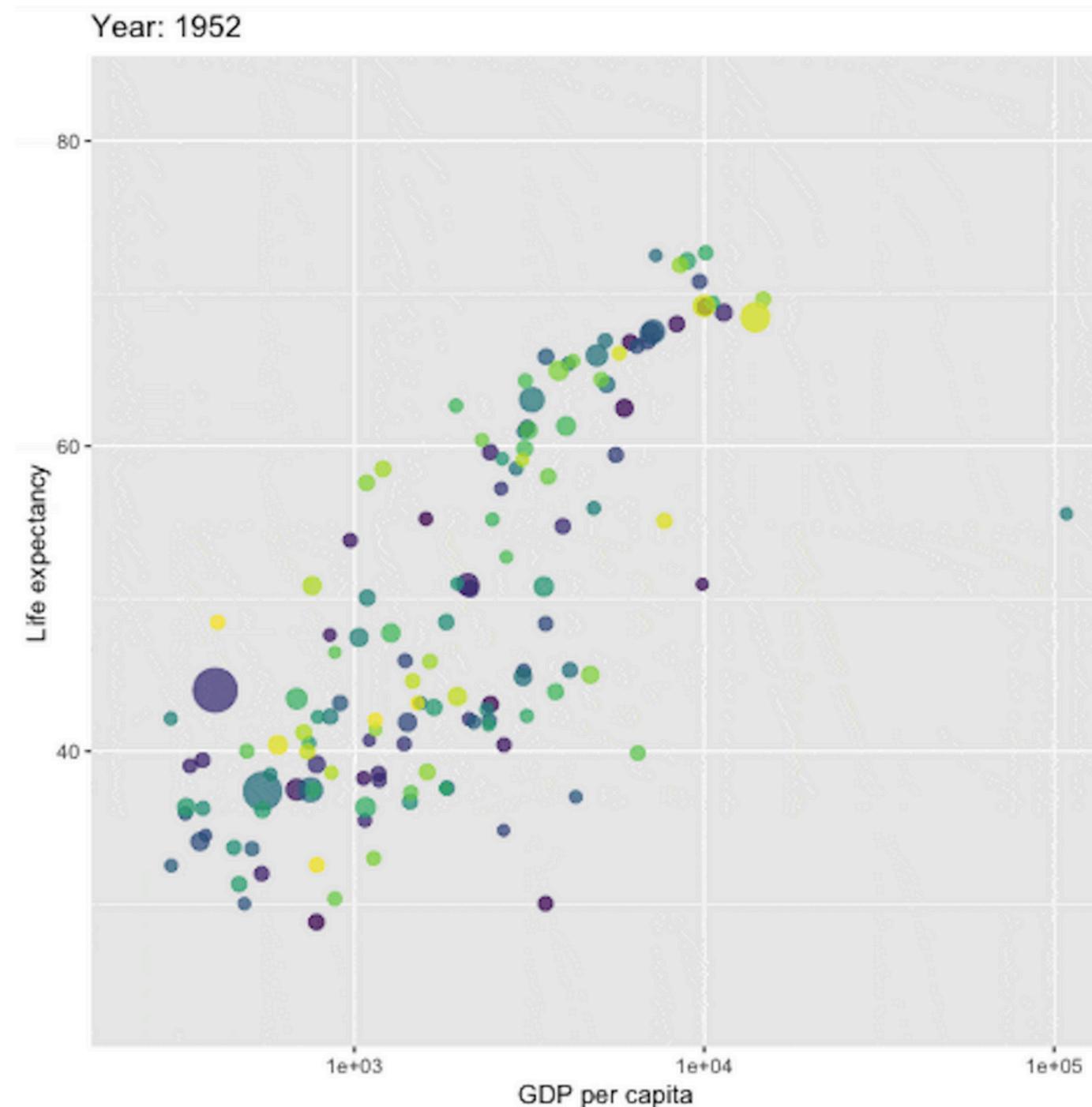
What is Time-Series Data?

We often visualize time series data by using one axis to encode time



What is Time-Series Data?

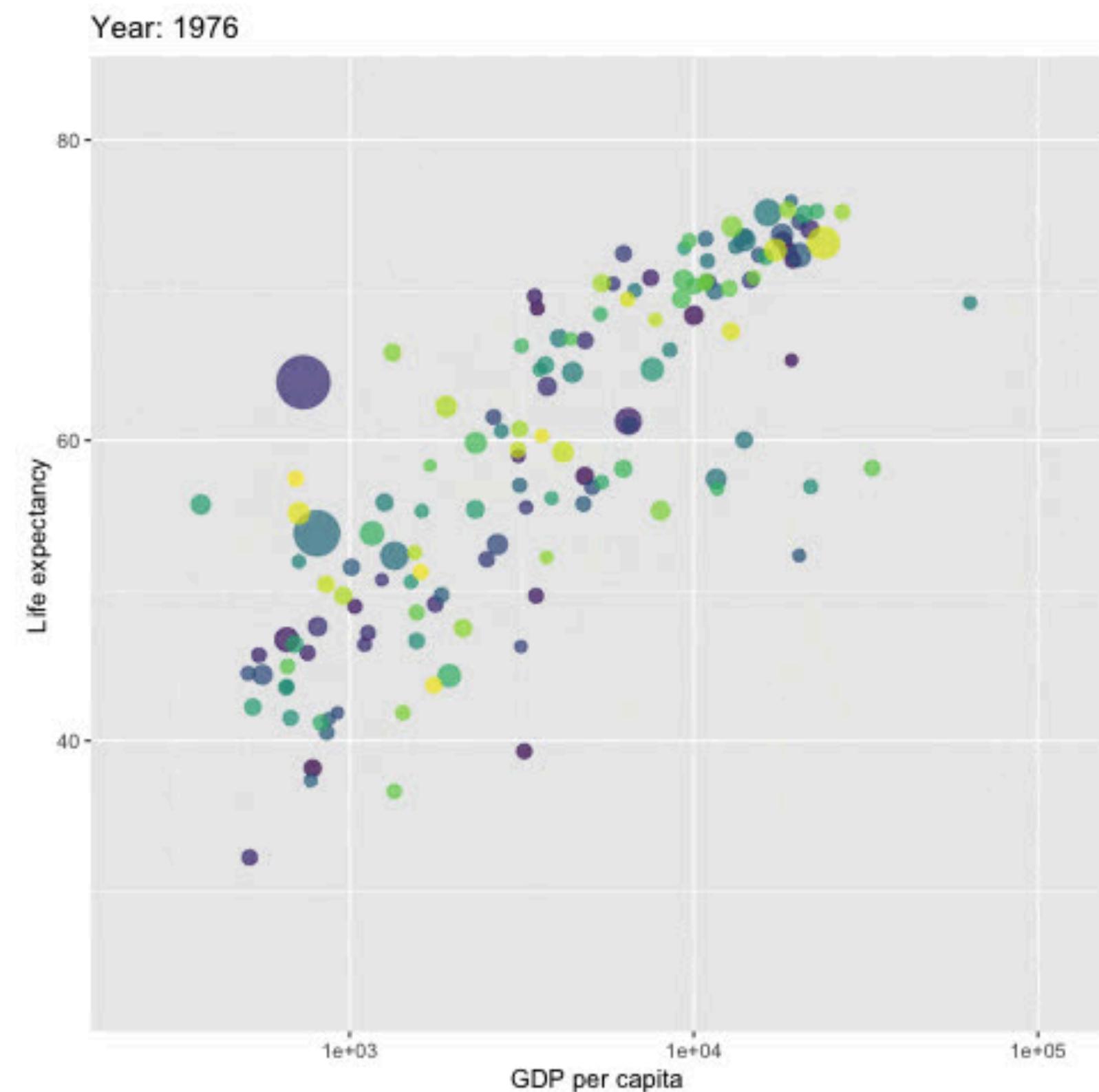
But you could also encode time by creating multiple static visualizations



What is Time-Series Data?

But you could also encode time by creating multiple static visualizations

and animating them?



But Beware!



But Beware!

Animation is useful for storytelling and transitions

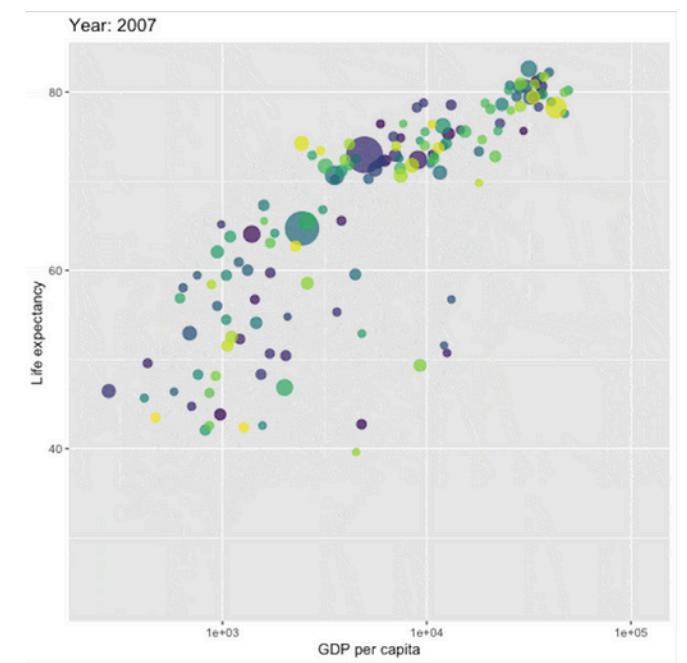
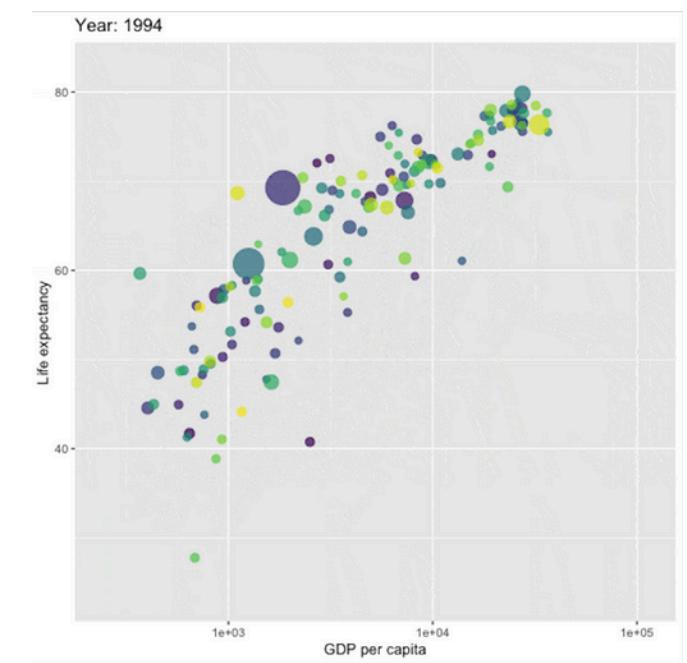
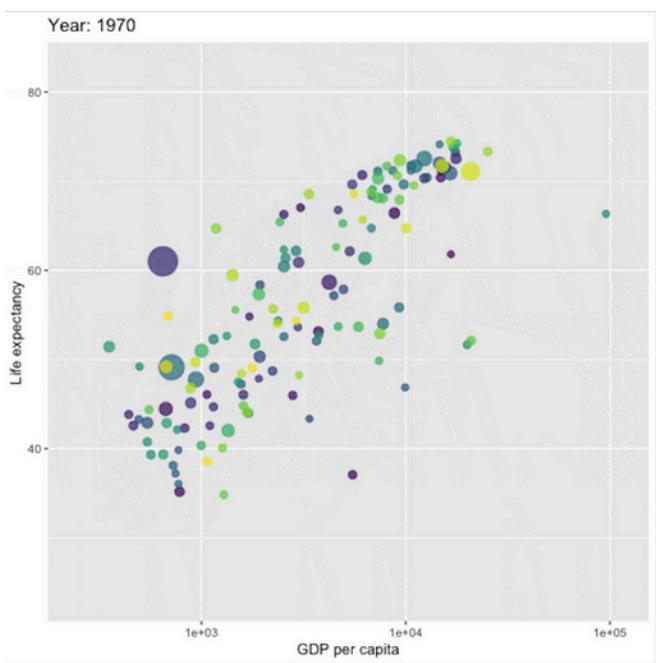
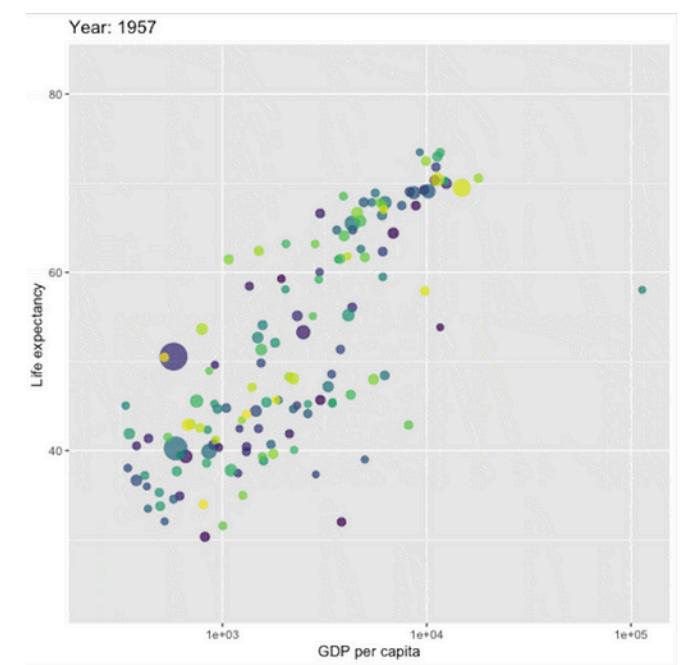
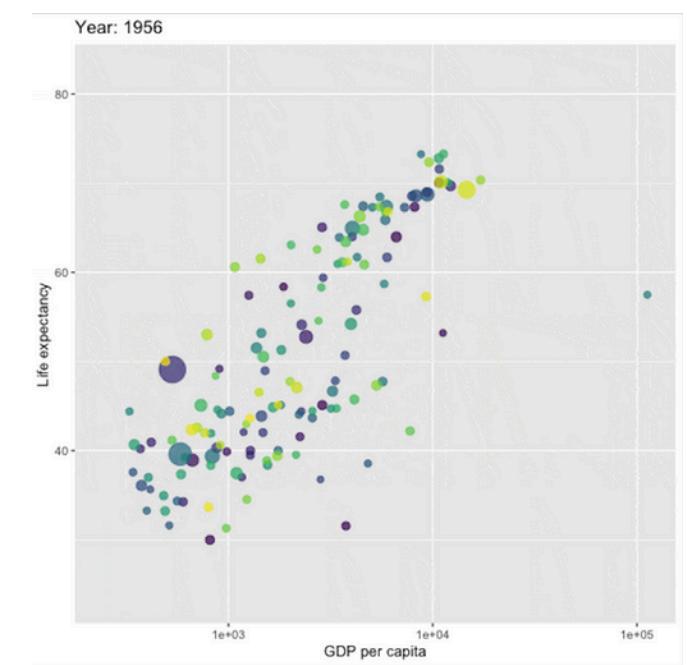
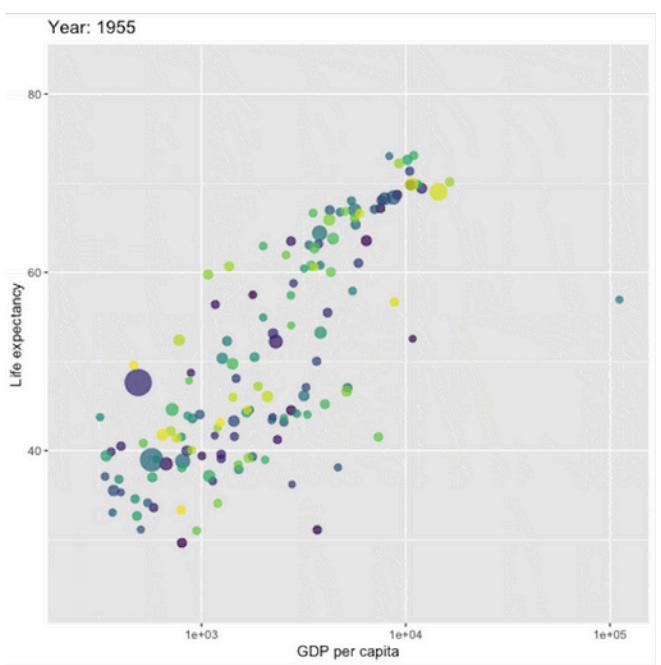
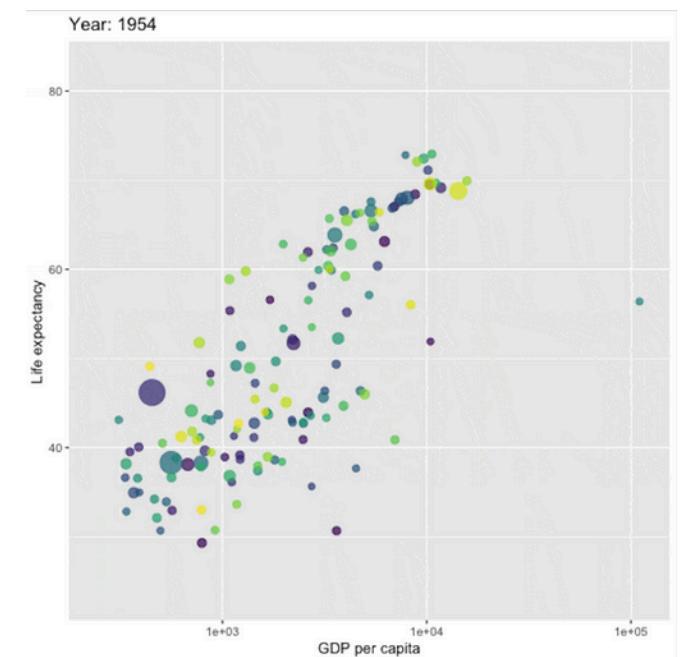
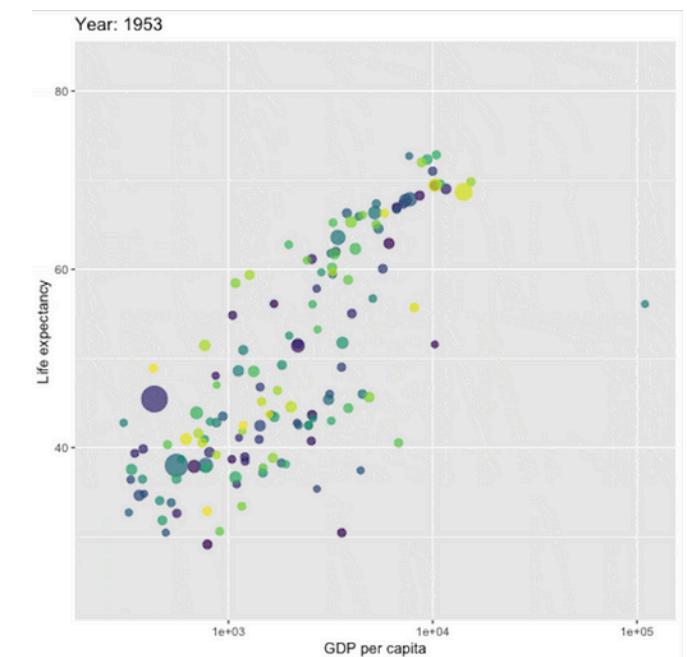
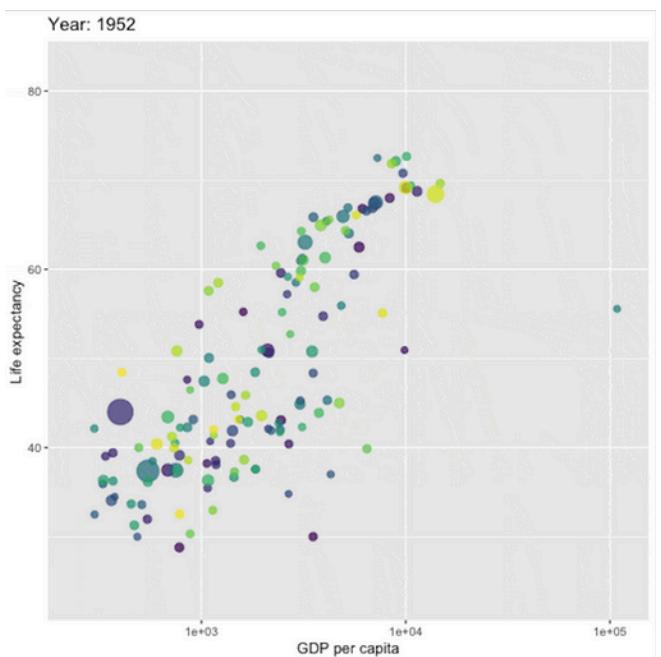
But it's **not** a great idea when you want **to compare multiple complex state changes over time**

For comparison of multiple states, use **small multiples**

Small Multiples

Multiple small, aligned visualizations

Useful for comparing state changes



Graphical Excellence

Graphical Excellence

Design a visualization that gives the viewer:

- the greatest number of ideas
- in the shortest time
- with the least ink
- in the smallest space

“An explanation should be as simple as possible, but not simpler.”

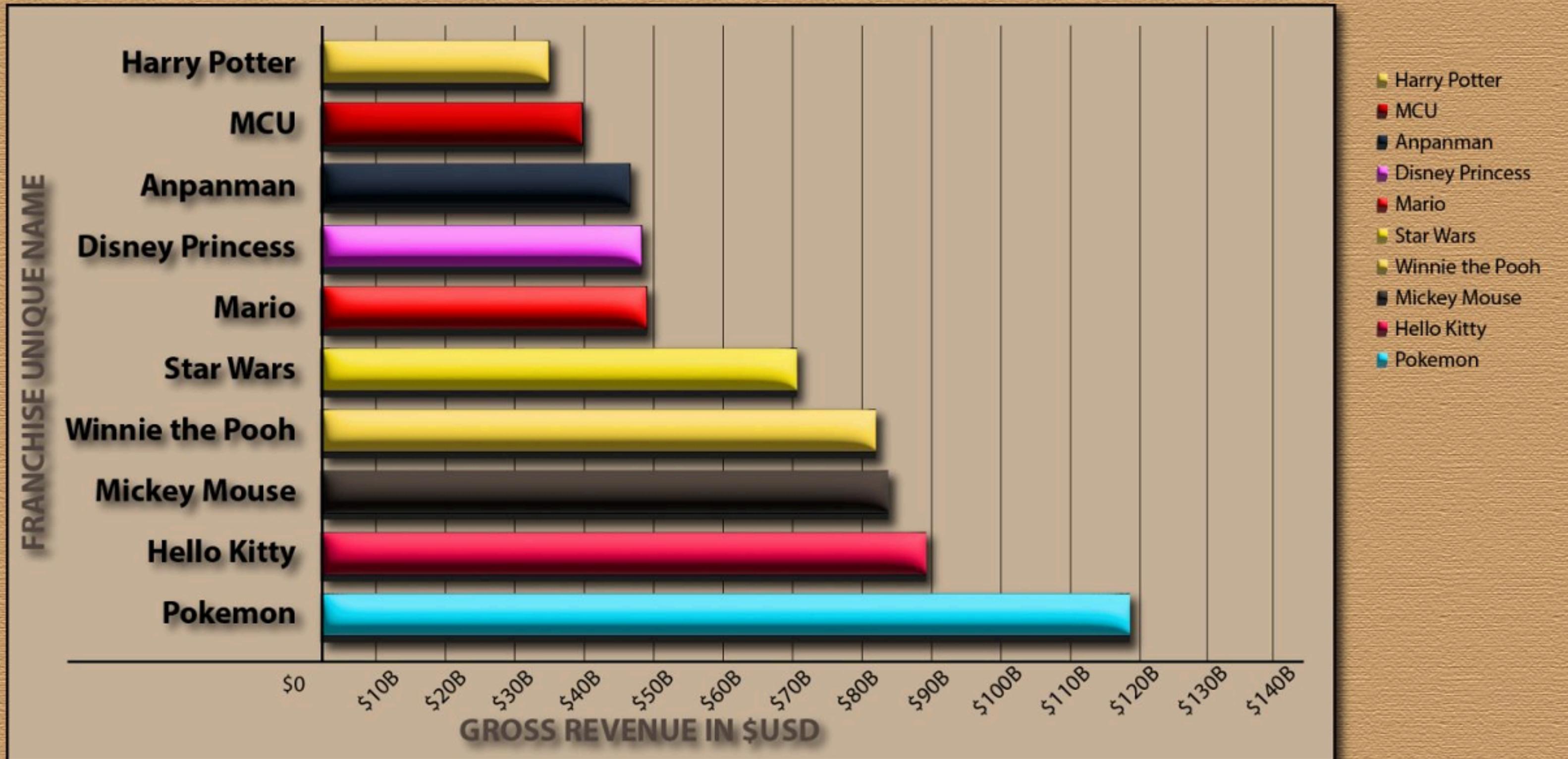
- Albert Einstein

How do we achieve graphical excellence?

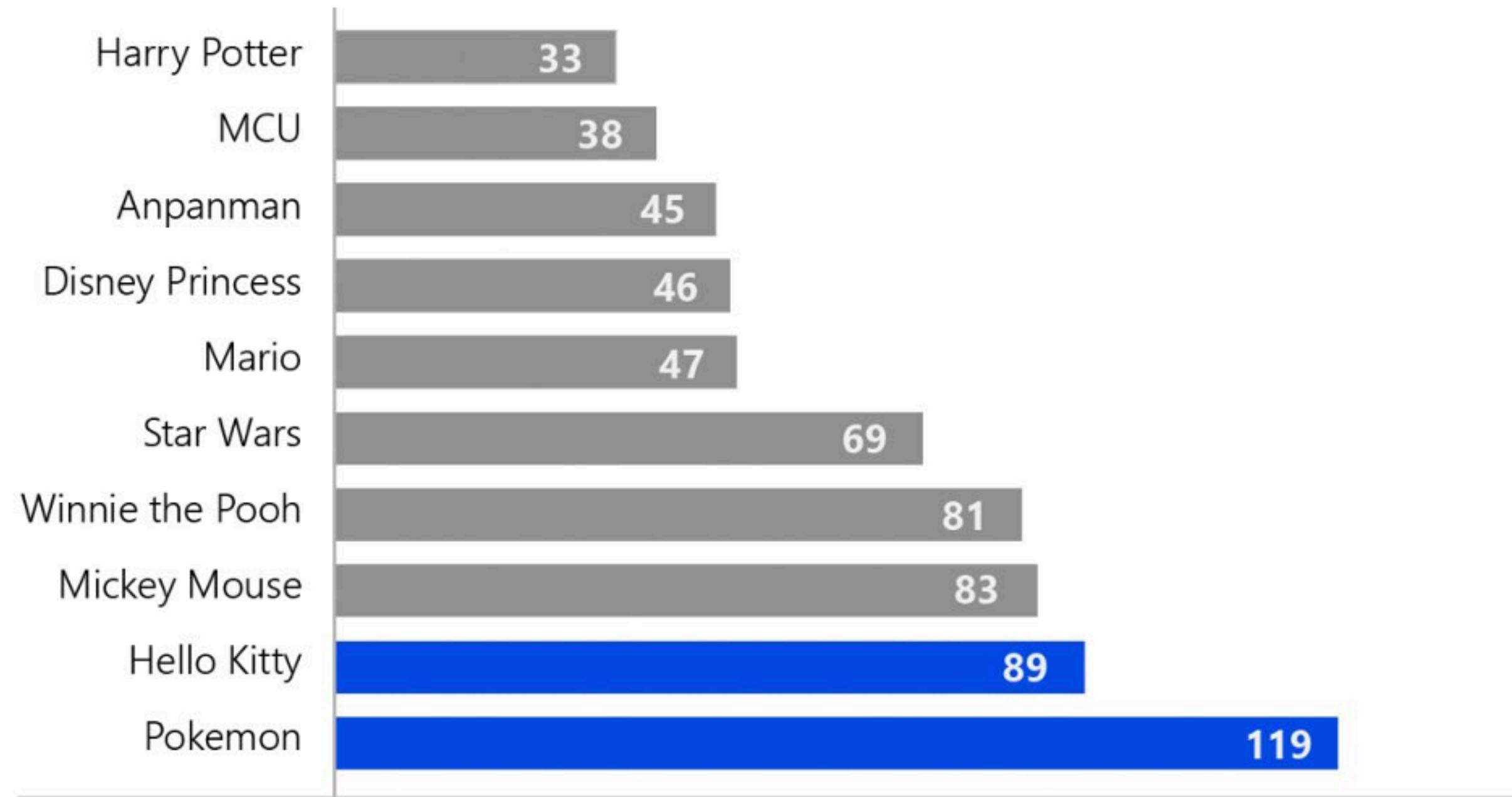
Maximize the Data-Ink Ratio

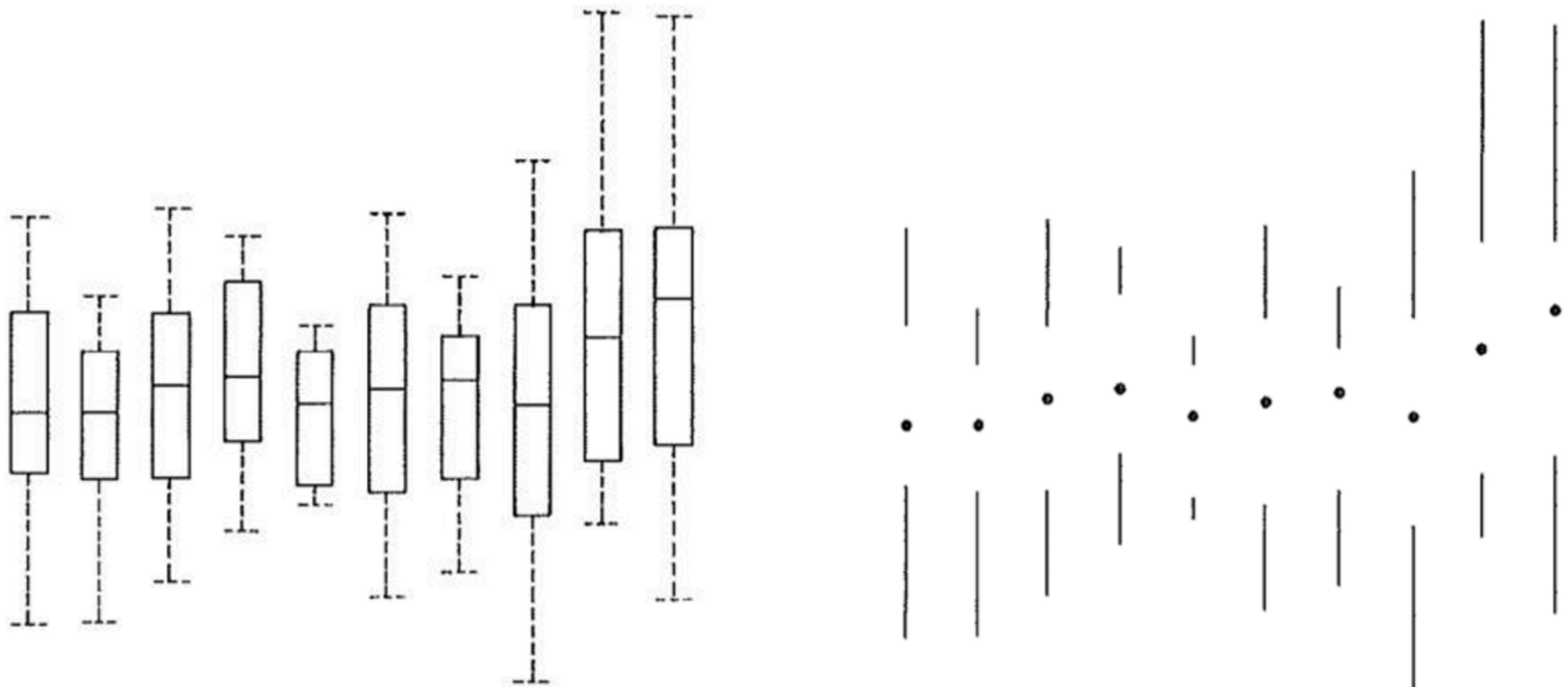
$$\text{Data-Ink Ratio} = \frac{\text{amount of ink used to encode data}}{\text{total ink used in graphic}}$$

Japanese Franchises Top the highest grossing media franchises of all time



Japanese Franchises Top the highest grossing (in Billions \$USD) media franchises of all time





A User Study of Visualization Effectiveness Using EEG and Cognitive Load

E. W. Anderson¹, K. C. Potter¹, L. E. Matzen², J. F. Shepherd², G. A. Preston³, and C. T. Silva¹

¹SCI Institute, University of Utah, USA

²Sandia National Laboratories, USA

³Utah State Hospital, USA

But...

Abstract

Effectively evaluating visualization techniques is a difficult task often assessed through feedback from user studies and expert evaluations. This work presents an alternative approach to visualization evaluation in which brain activity is passively recorded using electroencephalography (EEG). These measurements are used to compare different visualization techniques in terms of the burden they place on a viewer's cognitive resources. In this paper, EEG signals and response times are recorded while users interpret different representations of data distributions. This information is processed to provide insight into the cognitive load imposed on the viewer. This paper describes the design of the user study performed, the extraction of cognitive load measures from EEG data, and how those measures are used to quantitatively evaluate the effectiveness of visualizations.

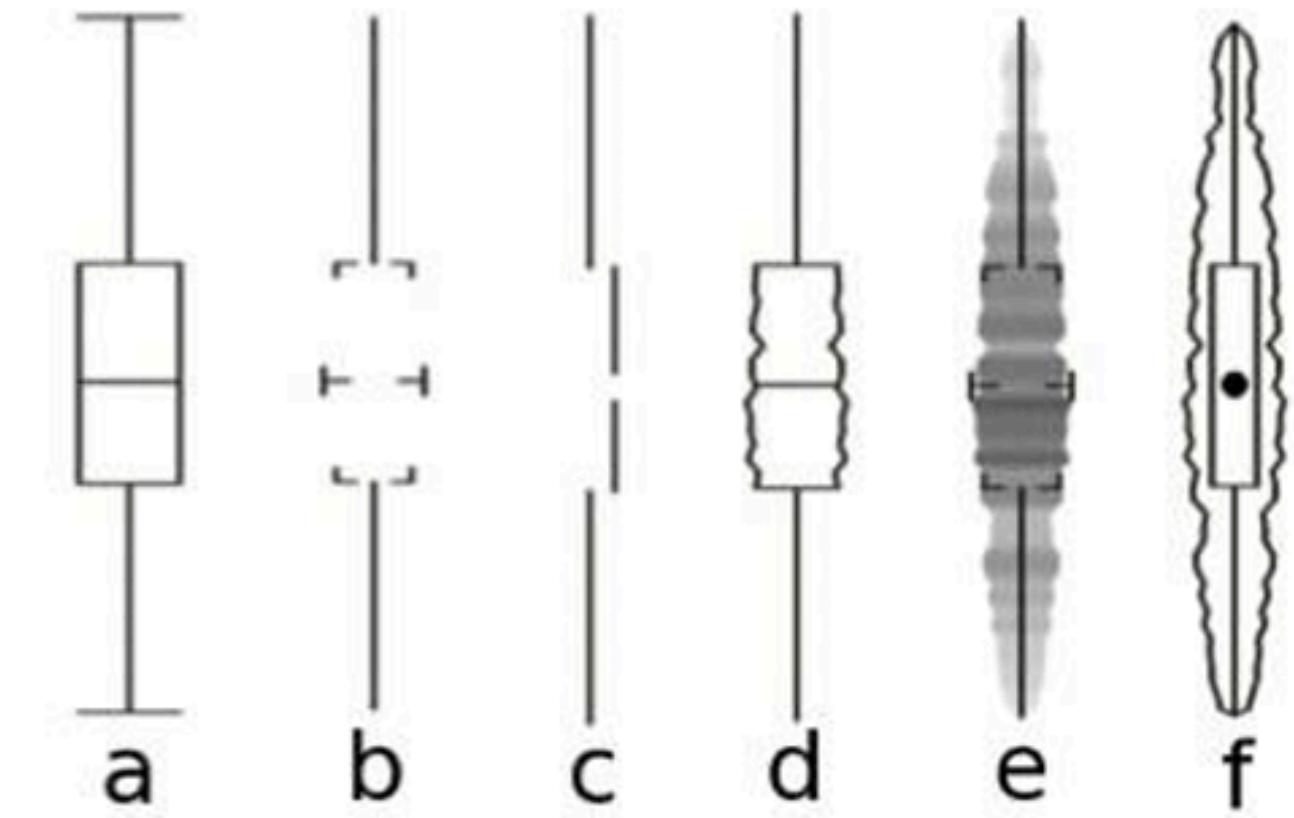
Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: General—Human Factors, Evaluation, Electroencephalography

Experimental Design

Participants were asked to choose with the largest **range**

How the boxplot was displayed varied

Participants' **cognitive load** was measured using EEG

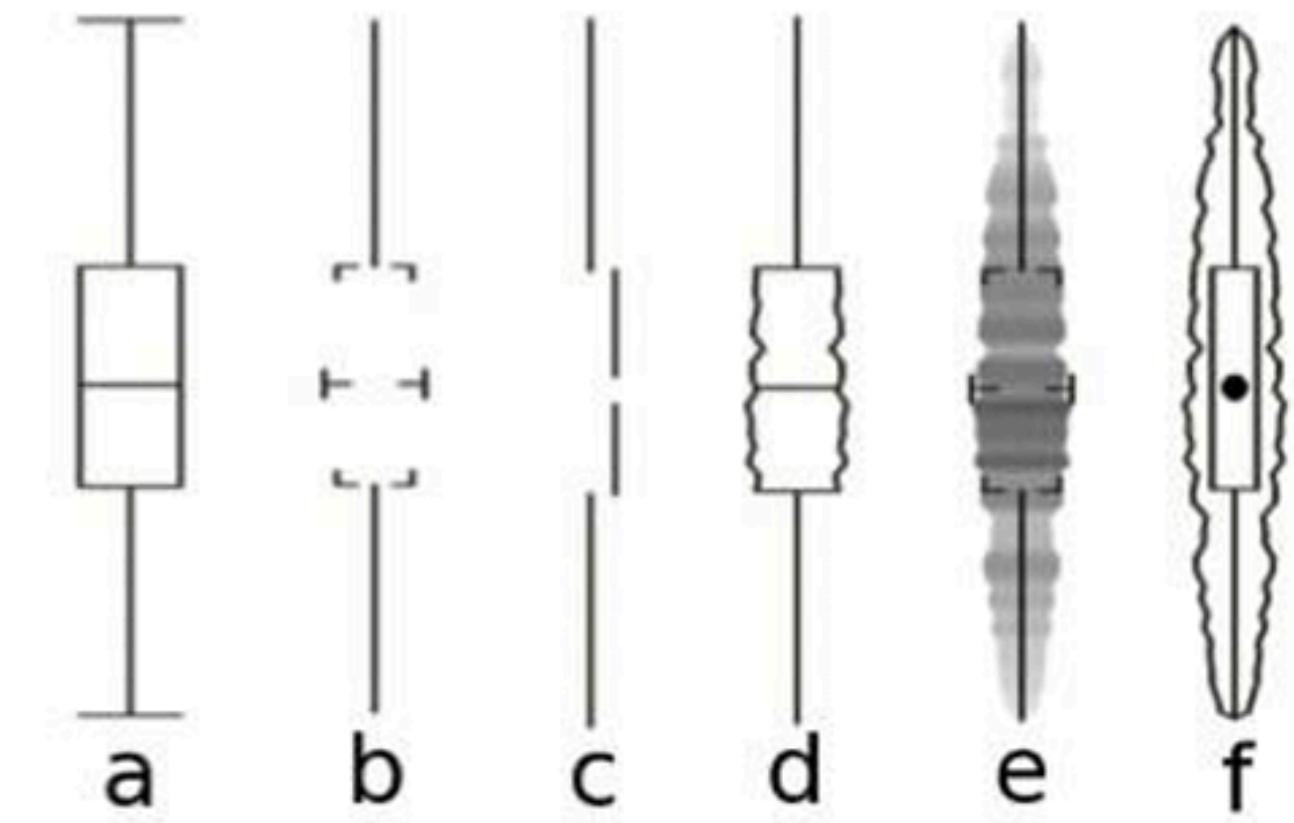


Experimental Results

Cognitive Load was **lowest** when interpreting the **canonical bar chart** (a) and **density plot** (e)

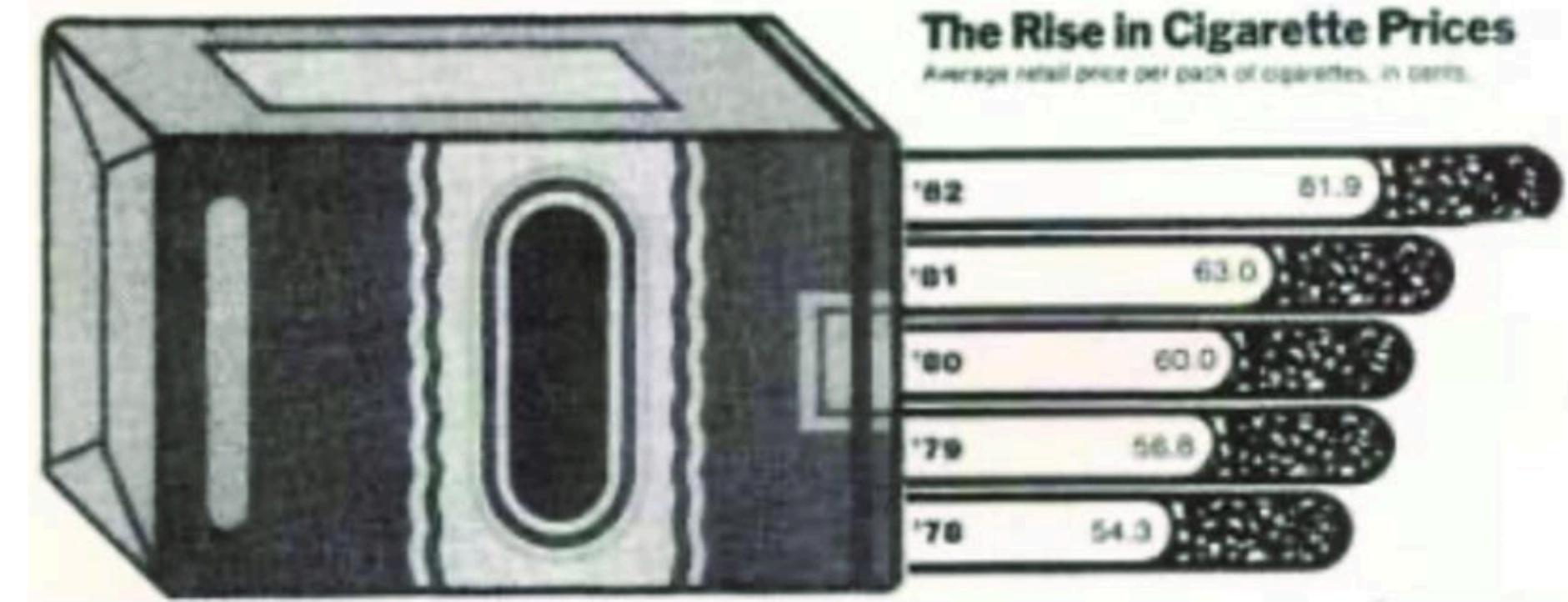
Cognitive Load was **highest** when interpreting the **interquartile plot** (c)

Fully maximizing the data-ink ratio may make it **harder** for people to use your visualizations



How do we achieve graphical excellence?

Avoid Chart Junk



MONSTROUS COSTS

Total House and Senate campaign expenditures,
in millions



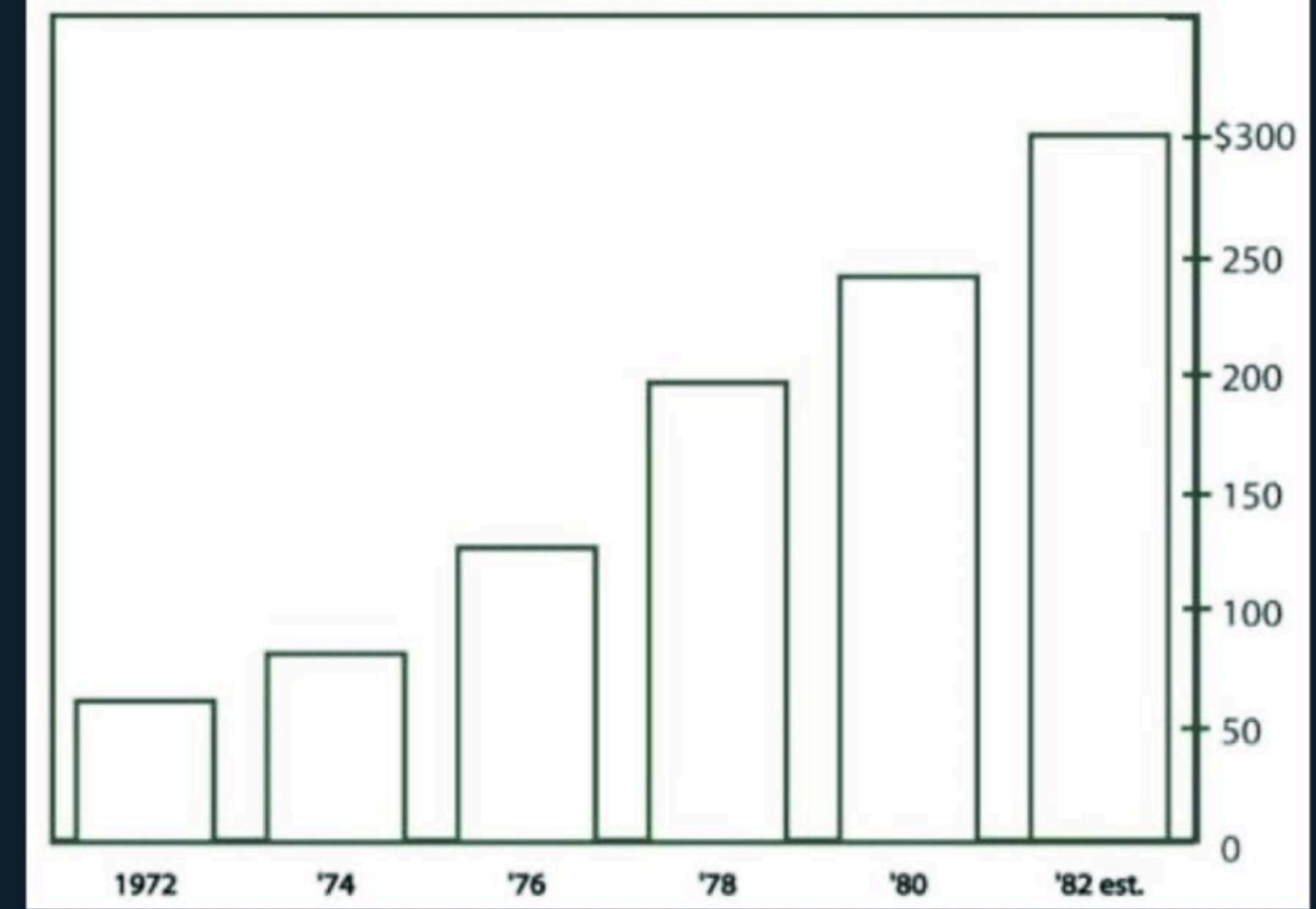
MONSTROUS COSTS

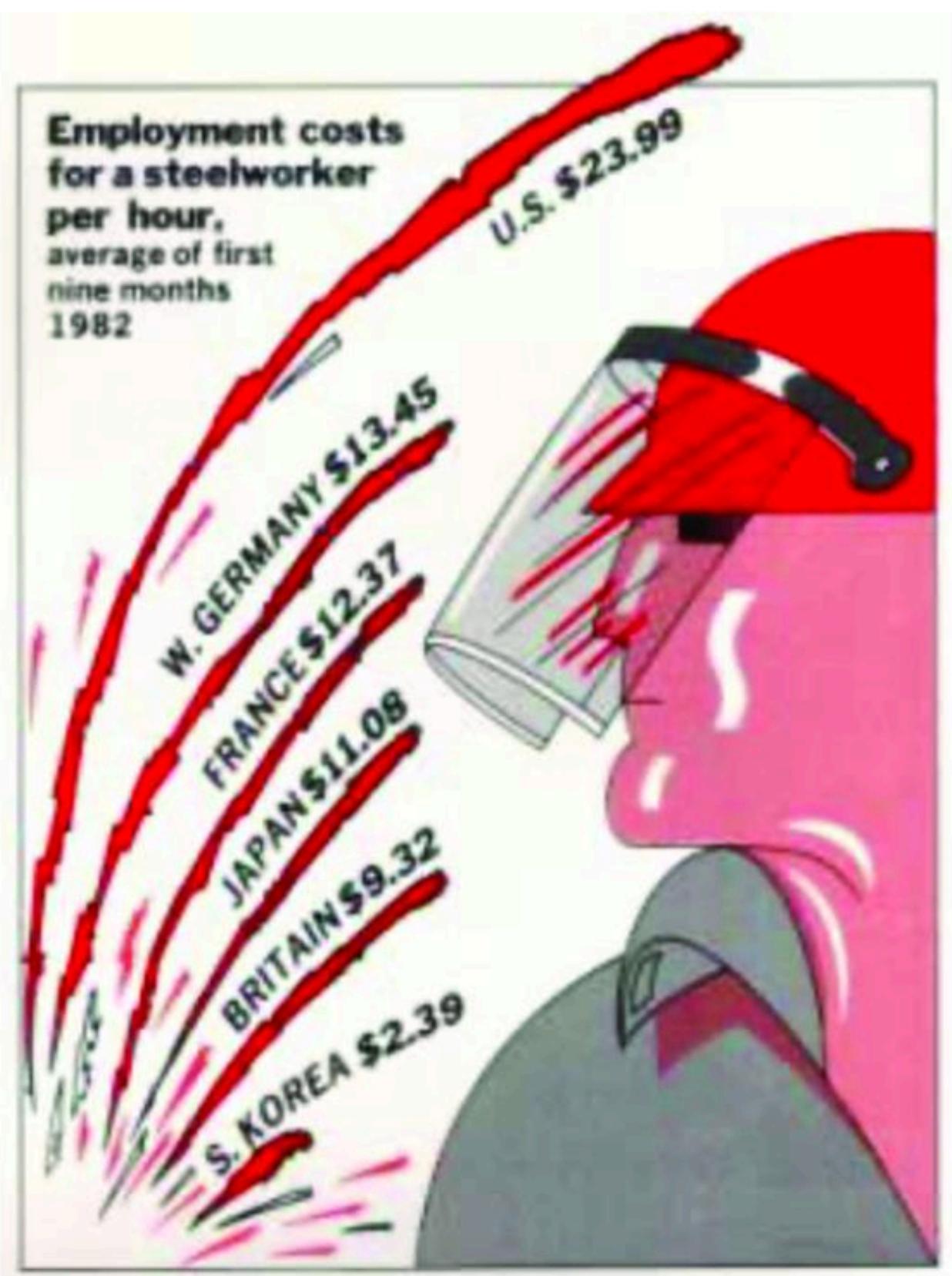
Total House and Senate campaign expenditures, in millions



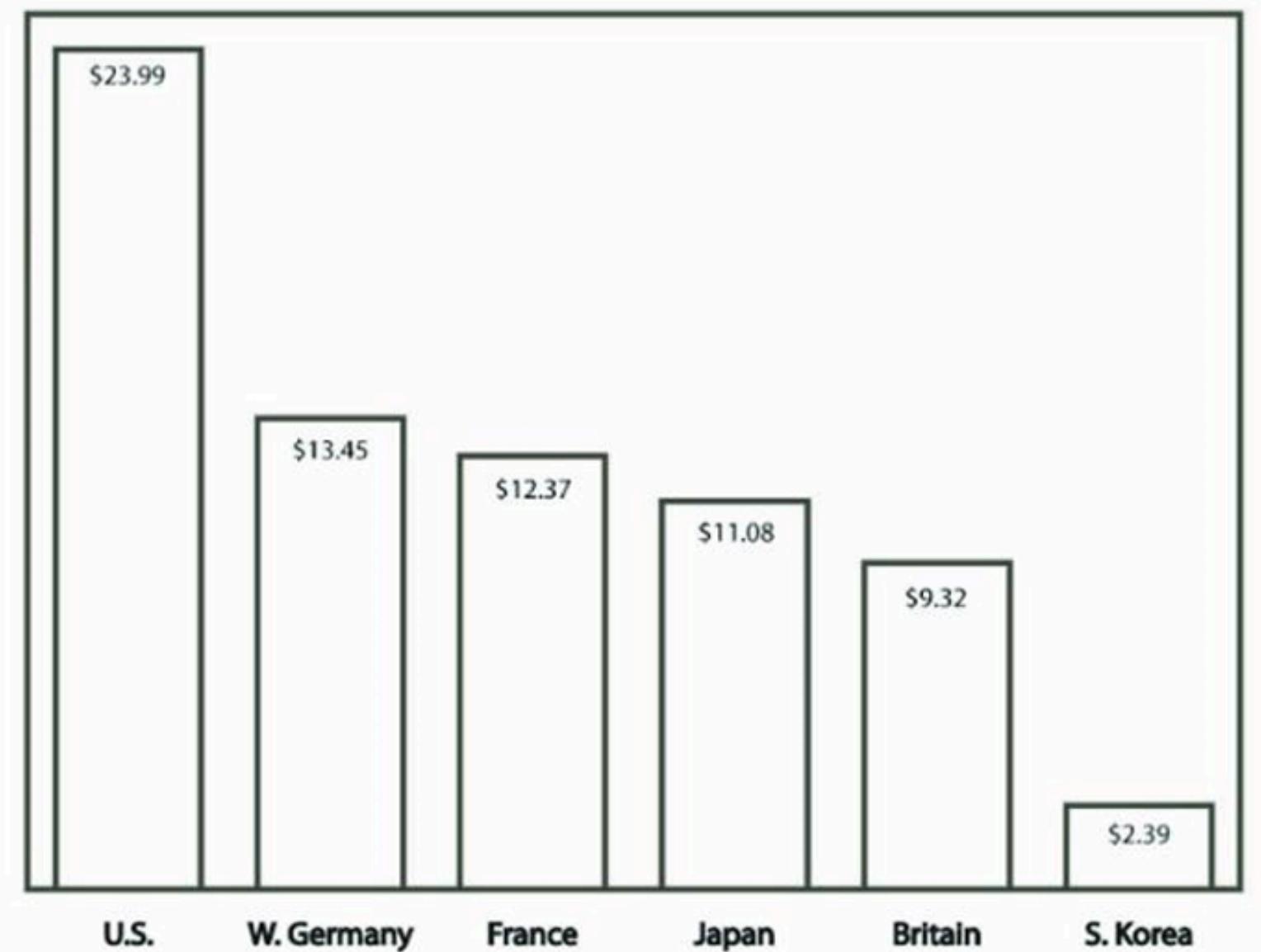
MONSTROUS COSTS

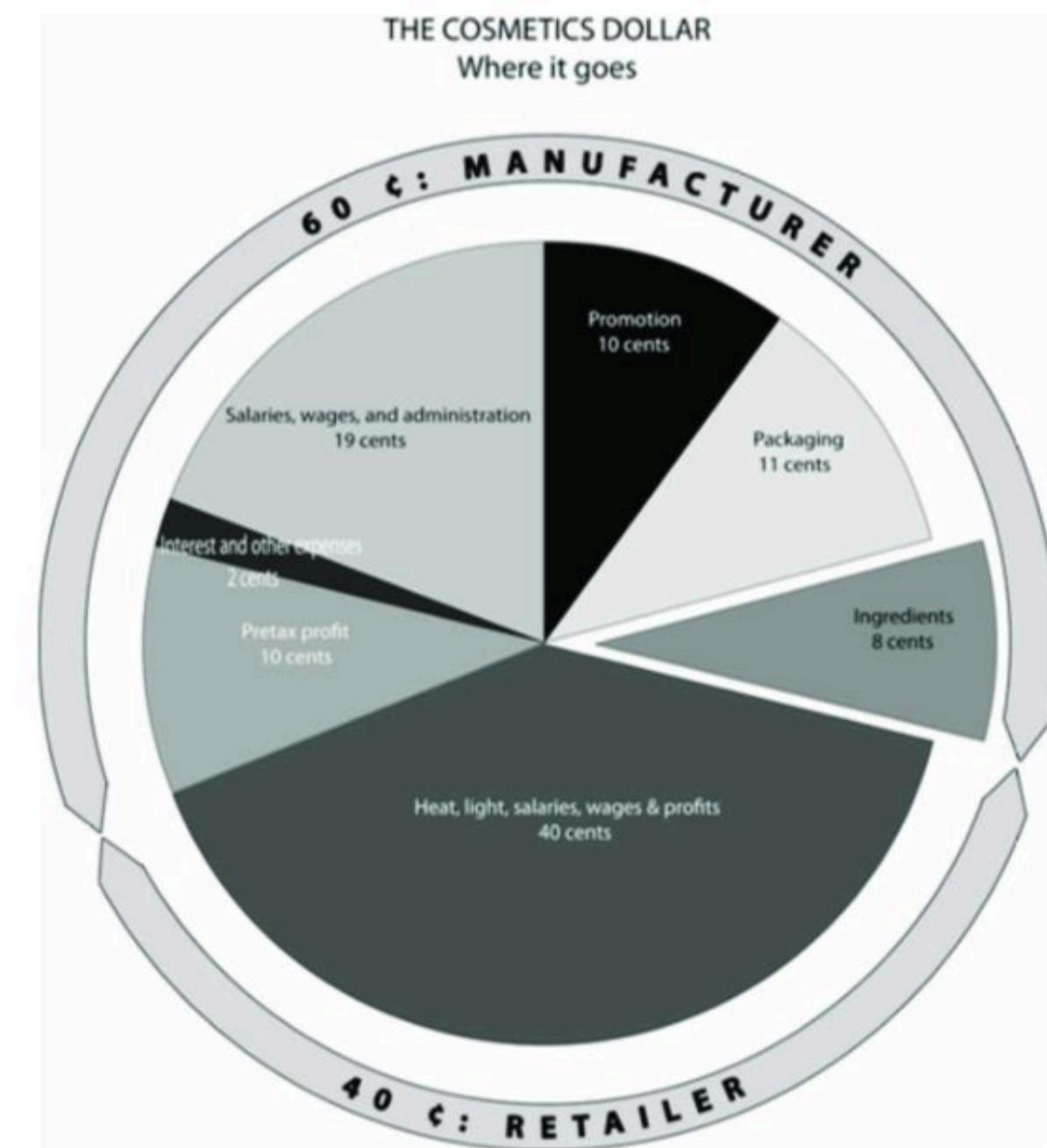
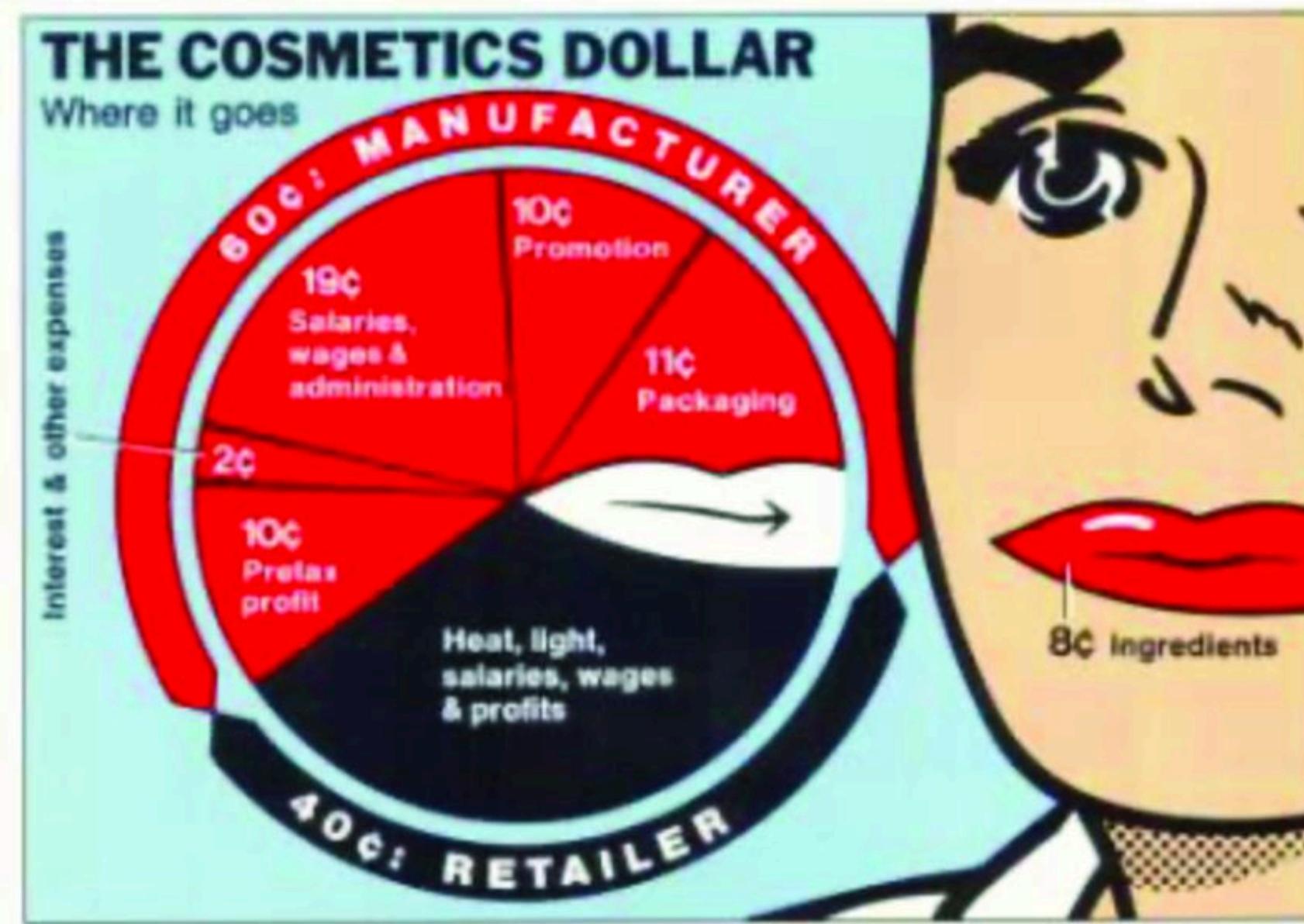
Total House and Senate campaign expenditures, in millions

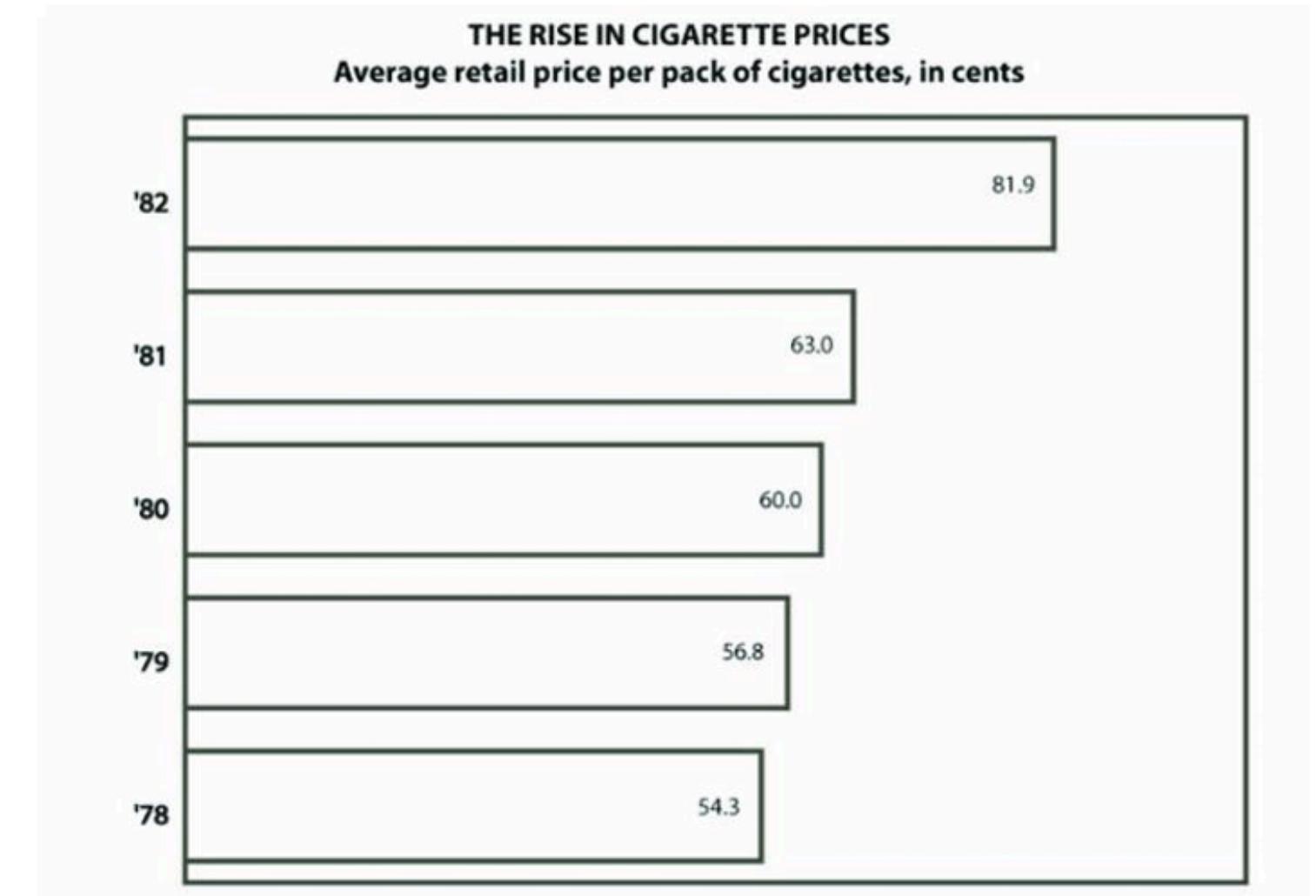
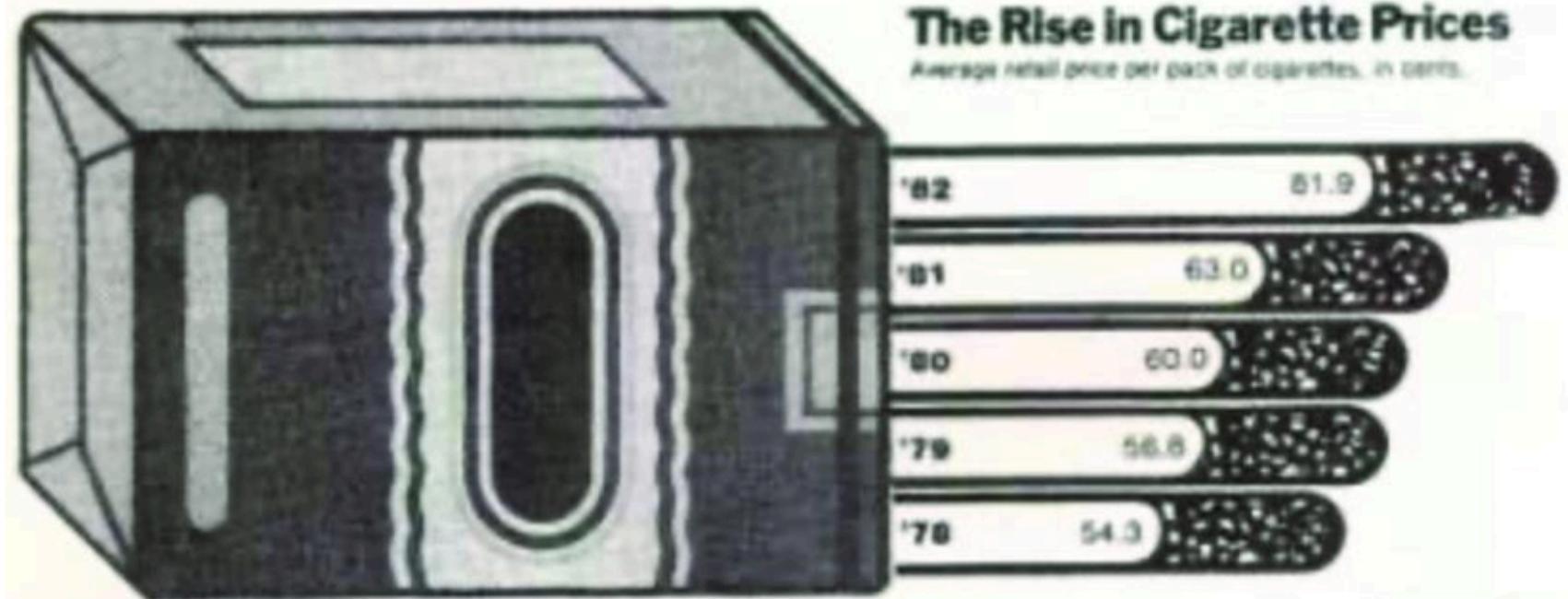


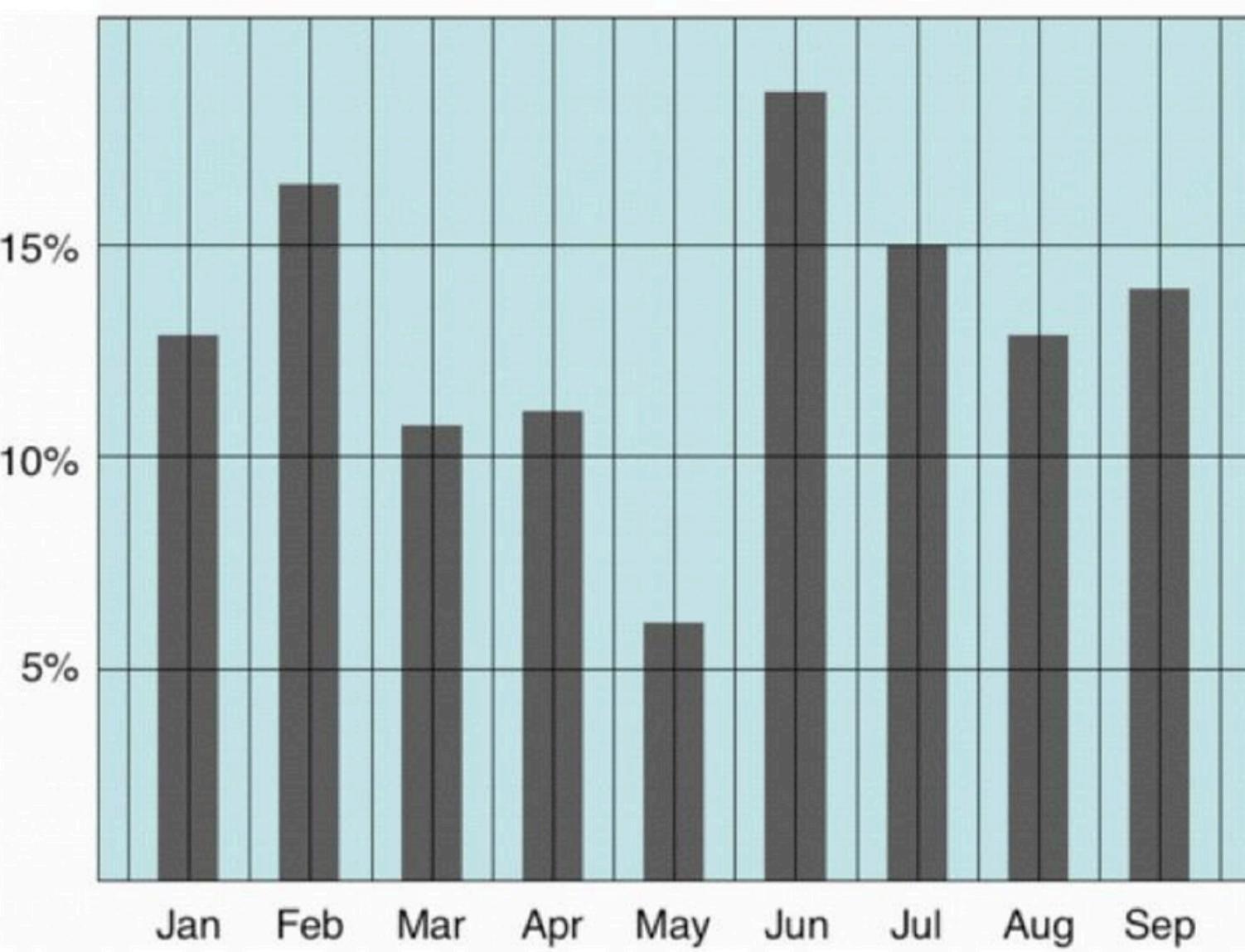


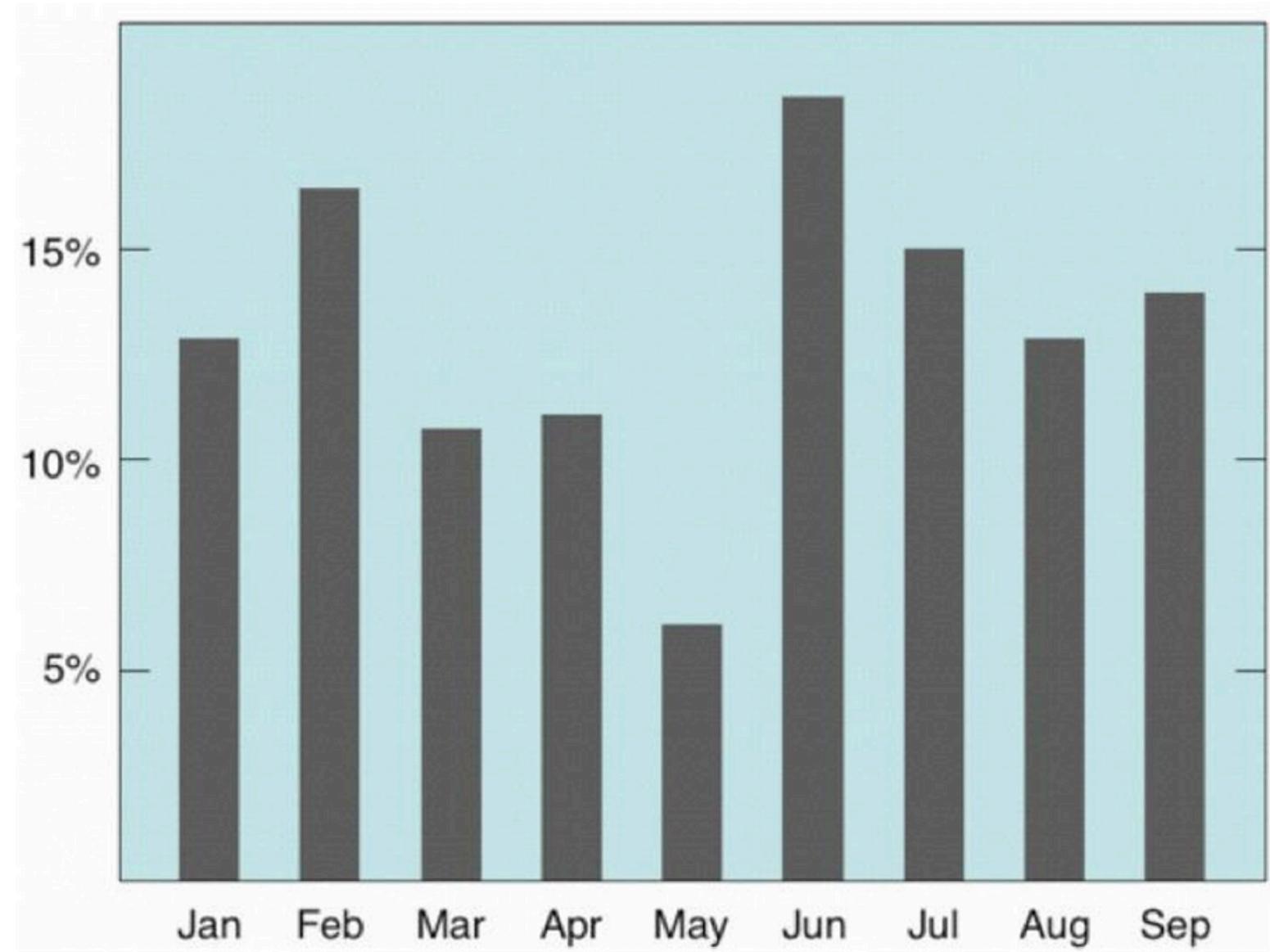
EMPLOYMENT COSTS FOR A STEELWORKER PER HOUR
Average of first nine months, 1982

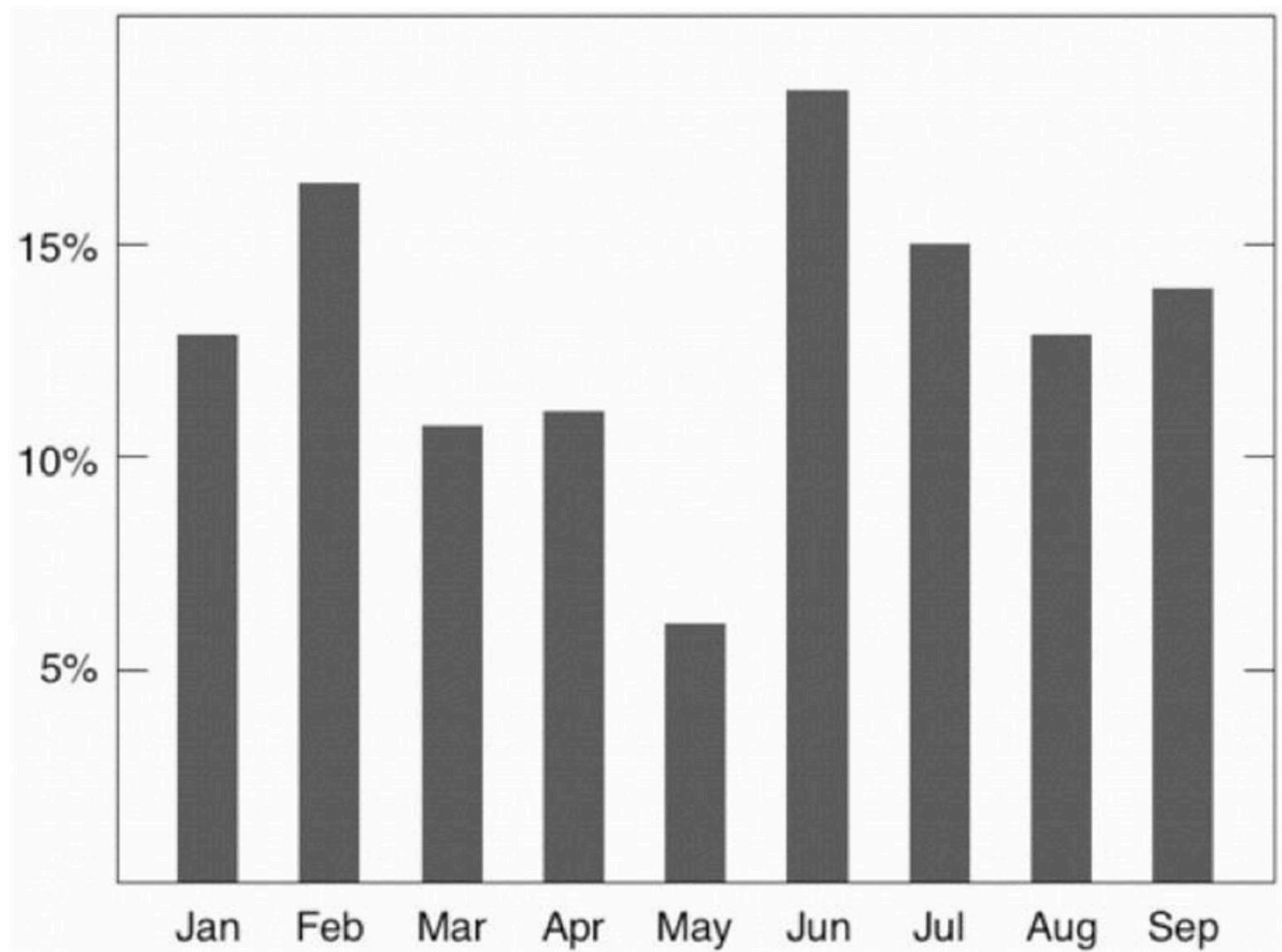


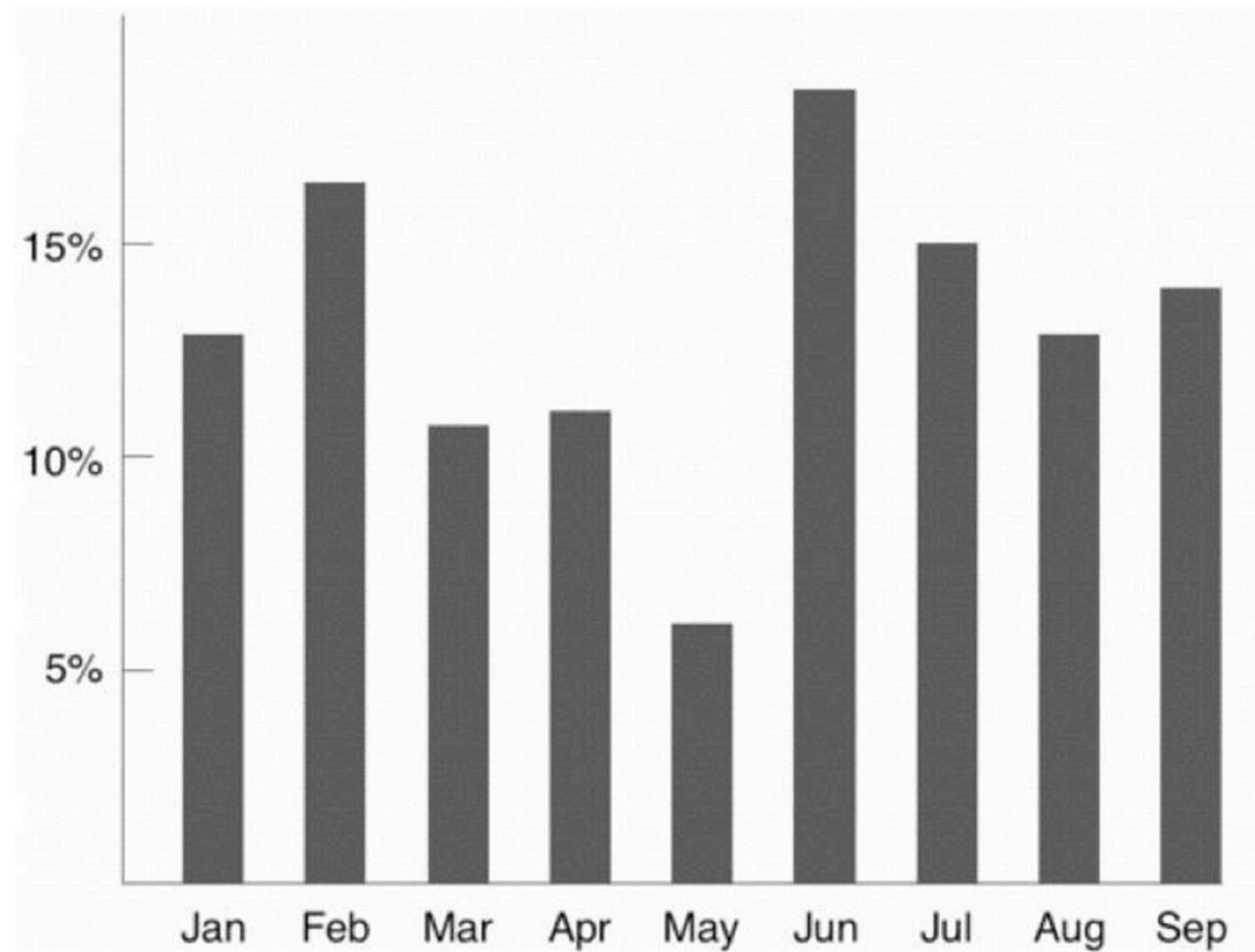


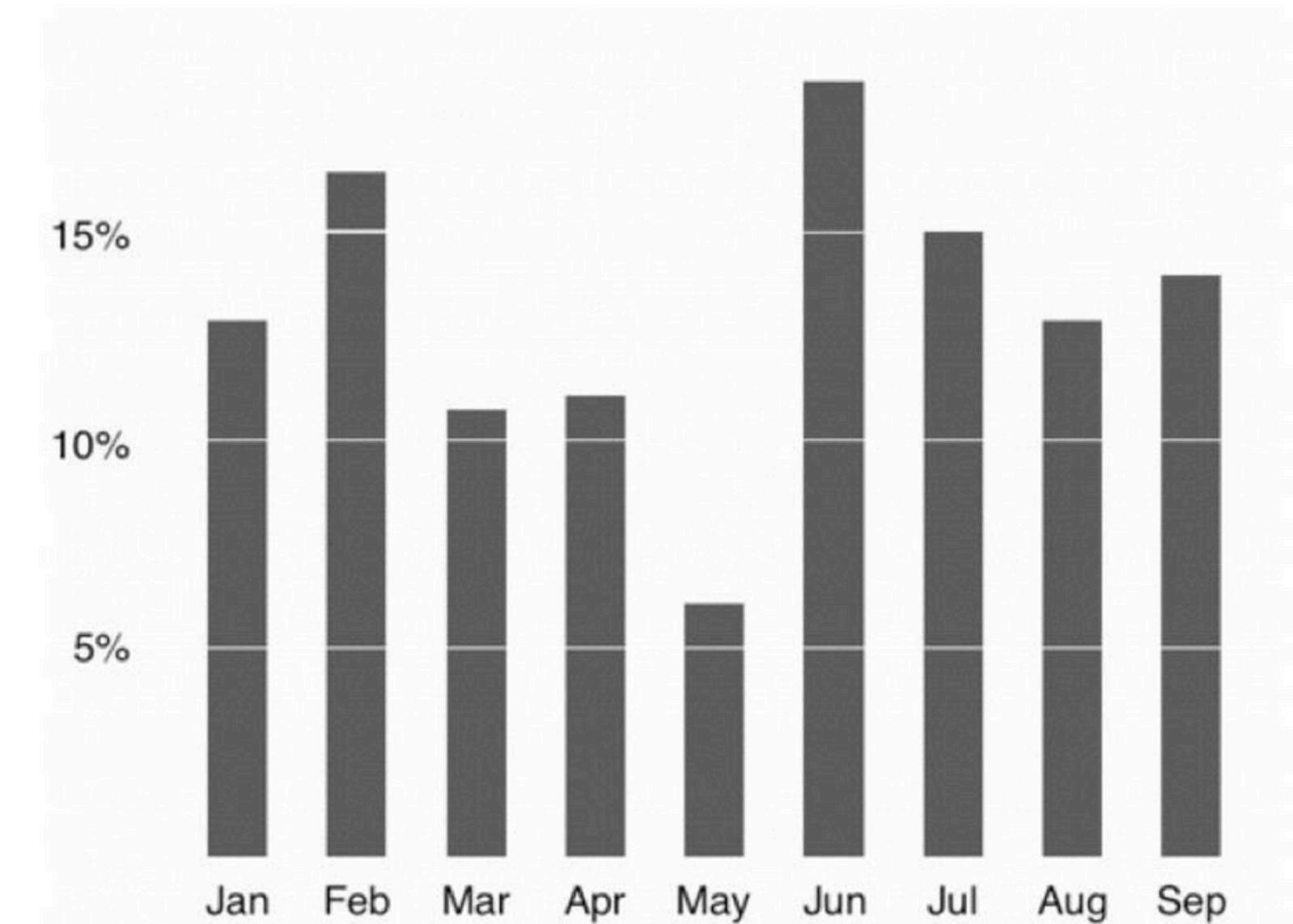












Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks

Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT

Guidelines for designing information charts often state that the presentation should reduce ‘chart junk’ – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people’s accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords

Charts, information visualization, imagery, memorability.

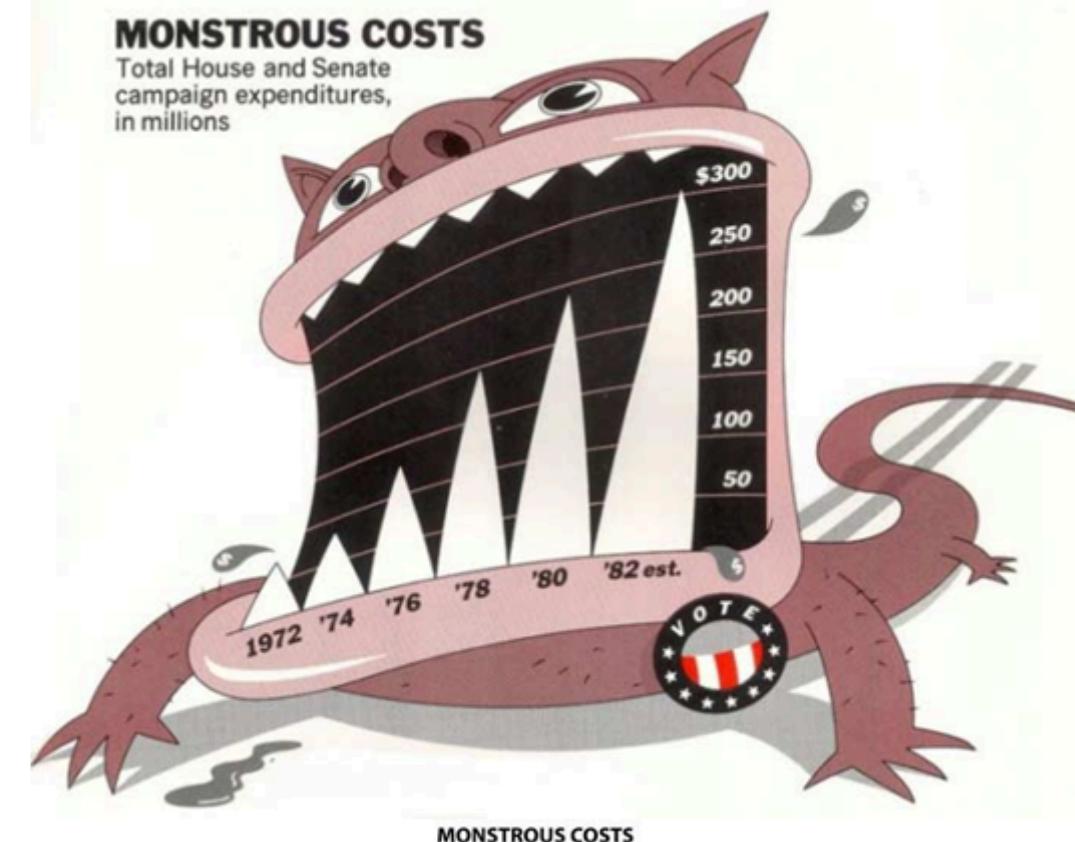
ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI):

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).

MONSTROUS COSTS

Total House and Senate campaign expenditures, in millions



Experimental Questions

Do visual embellishments cause comprehension problems?

Do embellishments provide additional valuable information to viewers?

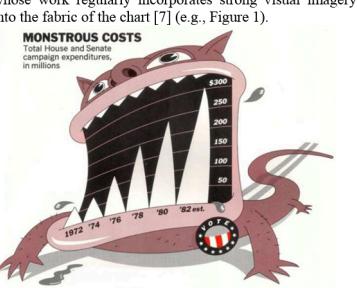
Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks

Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT
Guidelines for designing information charts often state that the presentation should reduce ‘chart junk’ – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people’s accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords
Charts, information visualization, imagery, memorability.
ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI).



Experimental Results

Do visual embellishments cause comprehension problems?

- No significant difference in interpretation accuracy between plain and embellished charts
- No significant difference in recall accuracy

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks

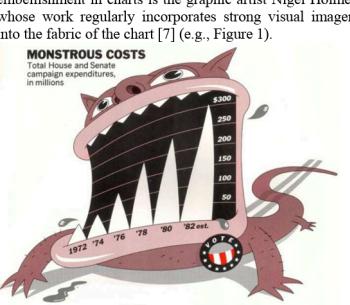
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT

Guidelines for designing information charts often state that the presentation should reduce ‘chart junk’ – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people’s accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords
Charts, information visualization, imagery, memorability.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI)



Experimental Results

Do embellishments provide additional valuable information to viewers?

- Better recall after a long-term gap (2-3 weeks) for embellished charts
- “Value messages” were more noticeable from embellished charts
- Embellished charts were more attractive and enjoyable

CHI 2010: Graphs

April 10–15, 2010, Atlanta, GA, USA

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

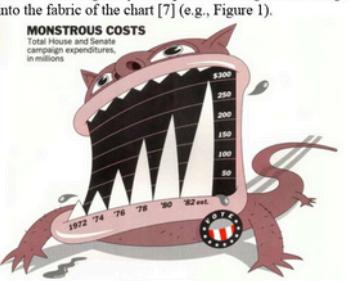
Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks

Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT
Guidelines for designing information charts often state that the presentation should reduce ‘chart junk’ – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people’s accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords
Charts, information visualization, imagery, memorability.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI).



What Makes a Visualization Memorable?

Michelle A. Borkin, *Student Member, IEEE*, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, *Student Member, IEEE*, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister, *Senior Member, IEEE*

And...

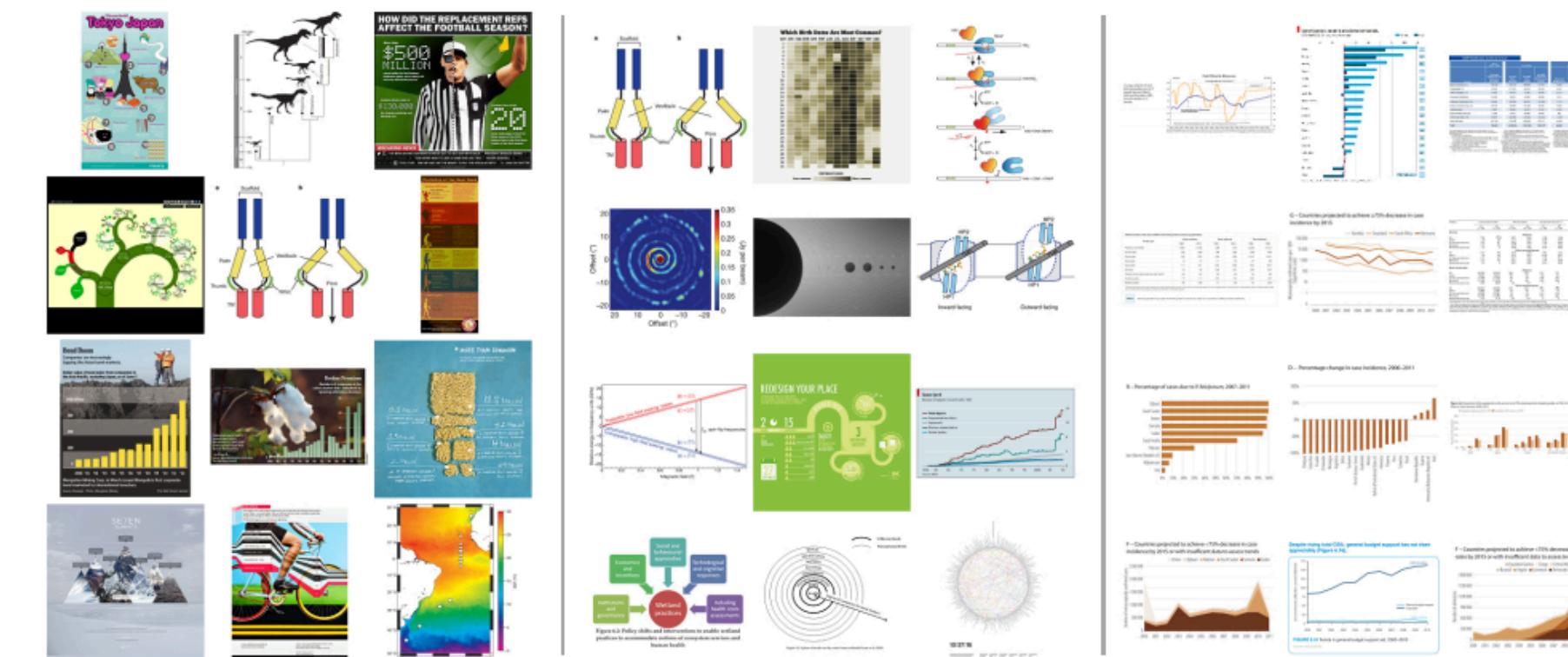


Fig. 1. Left: The top twelve overall most memorable visualizations from our experiment (most to least memorable from top left to bottom right). **Middle:** The top twelve most memorable visualizations from our experiment when visualizations containing human recognizable cartoons or images are removed (most to least memorable from top left to bottom right). **Right:** The twelve least memorable visualizations from our experiment (most to least memorable from top left to bottom right).

Abstract—An ongoing debate in the Visualization community concerns the role that visualization types play in data understanding. In human cognition, understanding and memorability are intertwined. As a first step towards being able to ask questions about impact and effectiveness, here we ask: “What makes a visualization memorable?” We ran the largest scale visualization study to date using 2,070 single-panel visualizations, categorized with visualization type (e.g., bar chart, line graph, etc.), collected from news media sites, government reports, scientific journals, and infographic sources. Each visualization was annotated with additional attributes, including ratings for data-ink ratios and visual densities. Using Amazon’s Mechanical Turk, we collected memorability scores for hundreds of these visualizations, and discovered that observers are consistent in which visualizations they find memorable and forgettable. We find intuitive results (e.g., attributes like color and the inclusion of a human recognizable object enhance memorability) and less intuitive results (e.g., common graphs are less memorable than unique visualization types). Altogether our findings suggest that quantifying memorability is a general metric of the utility of information, an essential step towards determining how to design effective visualizations.

Experimental Findings

Color and human-recognizable objects enhance memorability

Common chart types are less memorable than unique chart types

What Makes a Visualization Memorable?

Michelle A. Borkin, *Student Member, IEEE*, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, *Student Member, IEEE*, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister, *Senior Member, IEEE*

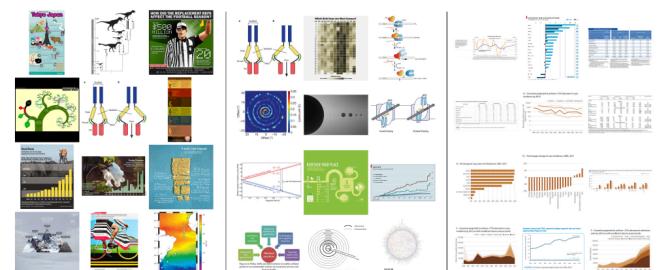


Fig. 1. **Left:** The top twelve overall most memorable visualizations from our experiment (most to least memorable from top left to bottom right). **Middle:** The top twelve most memorable visualizations from our experiment when visualizations containing human recognizable cartoons or images are removed (most to least memorable from top left to bottom right). **Right:** The twelve least memorable visualizations from our experiment (most to least memorable from top left to bottom right).

Abstract—An ongoing debate in the Visualization community concerns the role that visualization types play in data understanding. In human cognition, understanding and memorability are intertwined. As a first step towards being able to ask questions about impact and effectiveness, here we ask: “What makes a visualization memorable?” We ran the largest scale visualization study to date using 2,070 single-panel visualizations, categorized with visualization types (e.g., bar chart, line graph, etc.), collected from news media sites, government reports, scientific journals, and infographics sources. Each visualization was annotated with additional attributes, including ratings for data-ink ratios and visual densities. Using Amazon’s Mechanical Turk, we collected memorability scores for hundreds of these visualizations, and discovered that observers are consistent in which visualizations they find memorable and forgettable. We find intuitive results (e.g., attributes like color and the inclusion of a human recognizable object enhance memorability) and less intuitive results (e.g., common graphs are less memorable than unique visualization types). Altogether our findings suggest that quantifying memorability is a general metric of the utility of information, an essential step towards determining how to design effective visualizations.

So, should you avoid chart junk?

It depends...

Pros of using chart junk:

- more persuasive
- more memorable
- more engaging

Cons of using chart junk:

- may be less trustworthy
- may be harder to interpret
- may be an inefficient use of space

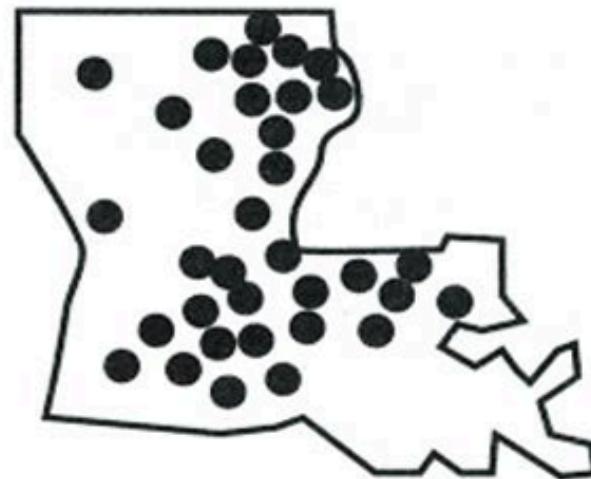
How do we achieve graphical excellence?

Maximize the Data Density

$$\text{Data Density} = \frac{\text{number of data items visualized}}{\text{area of graphic}}$$

Shrink the Graphics (with small multiples)

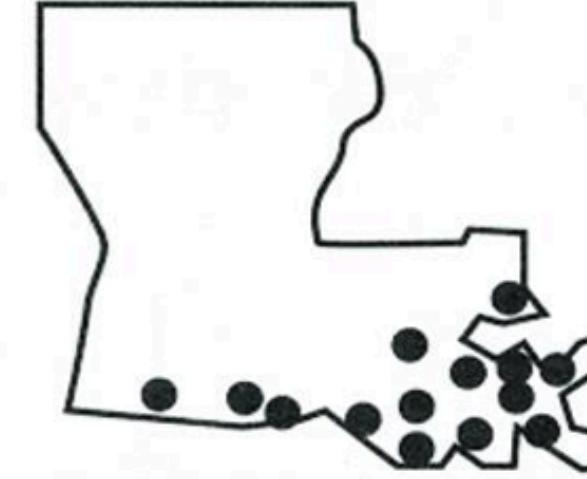
alfisol



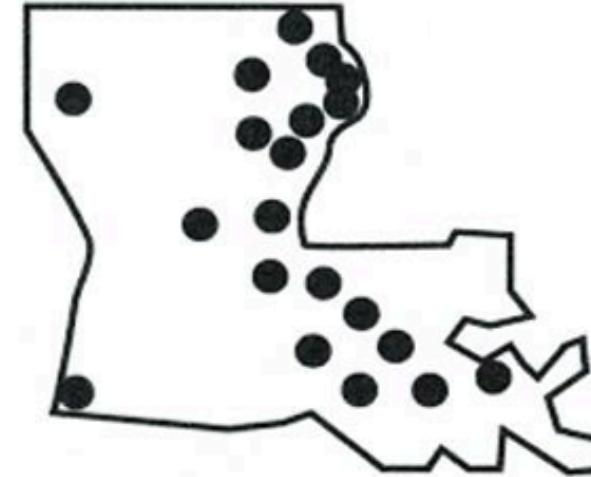
entisol



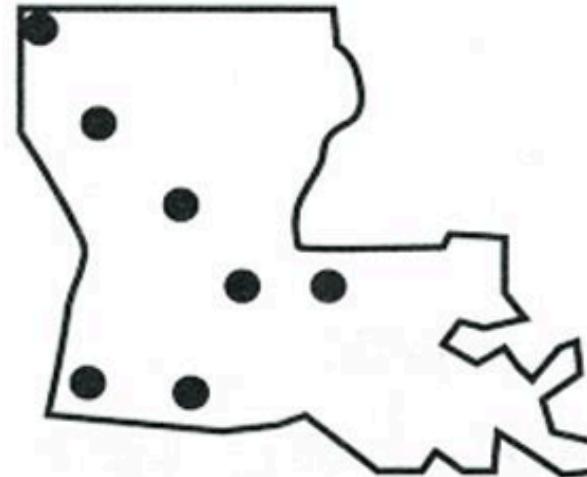
histosol



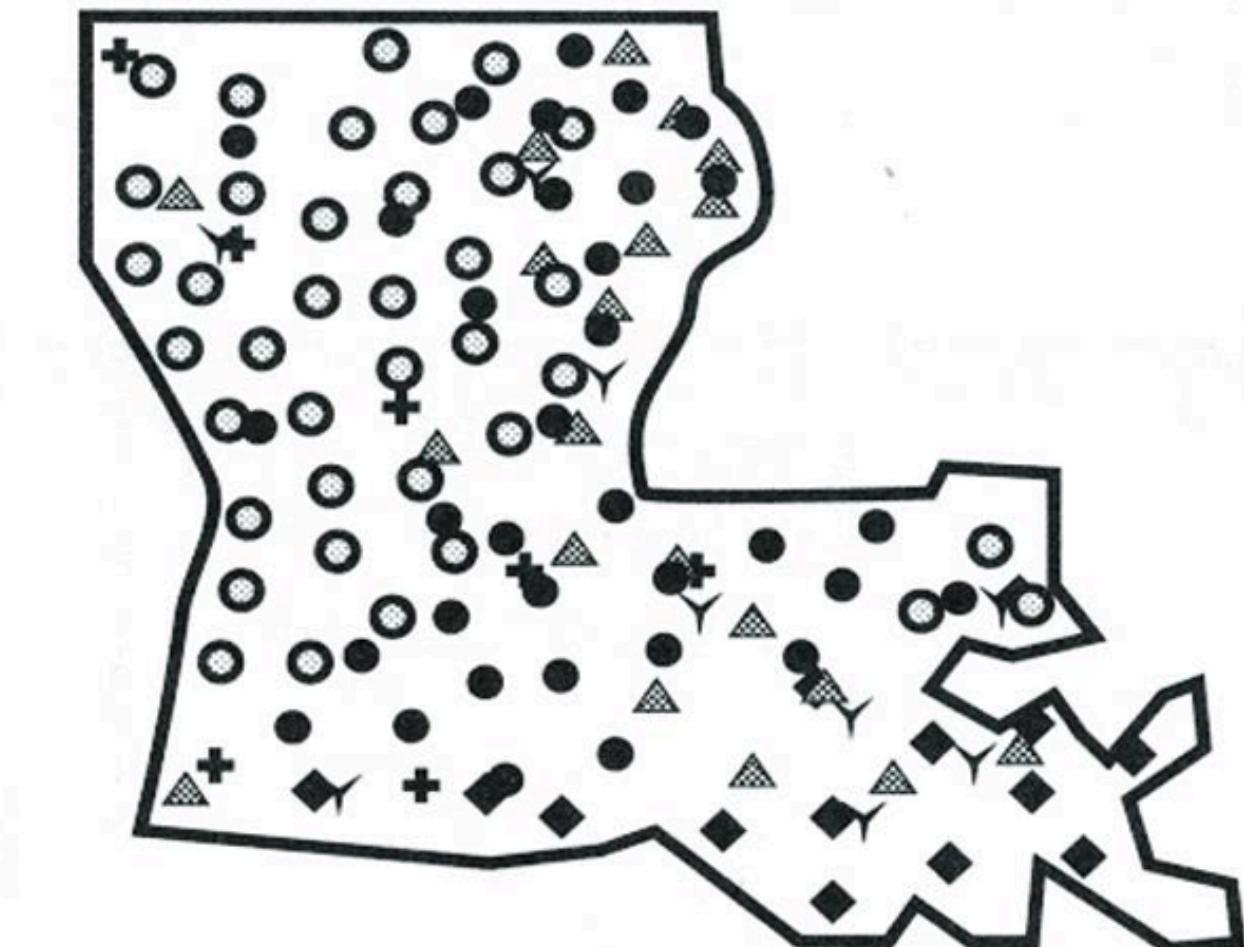
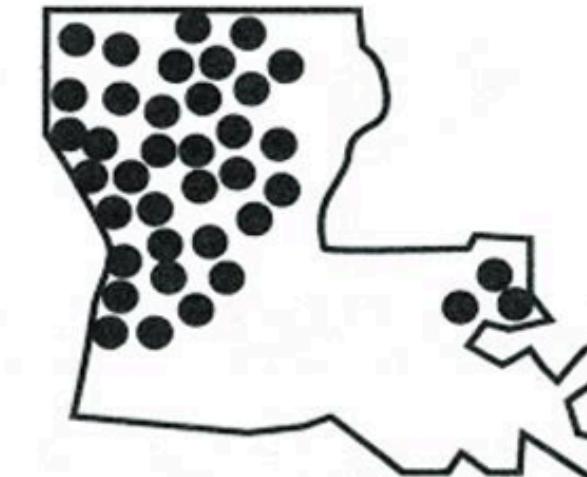
inceptisol



mollisol



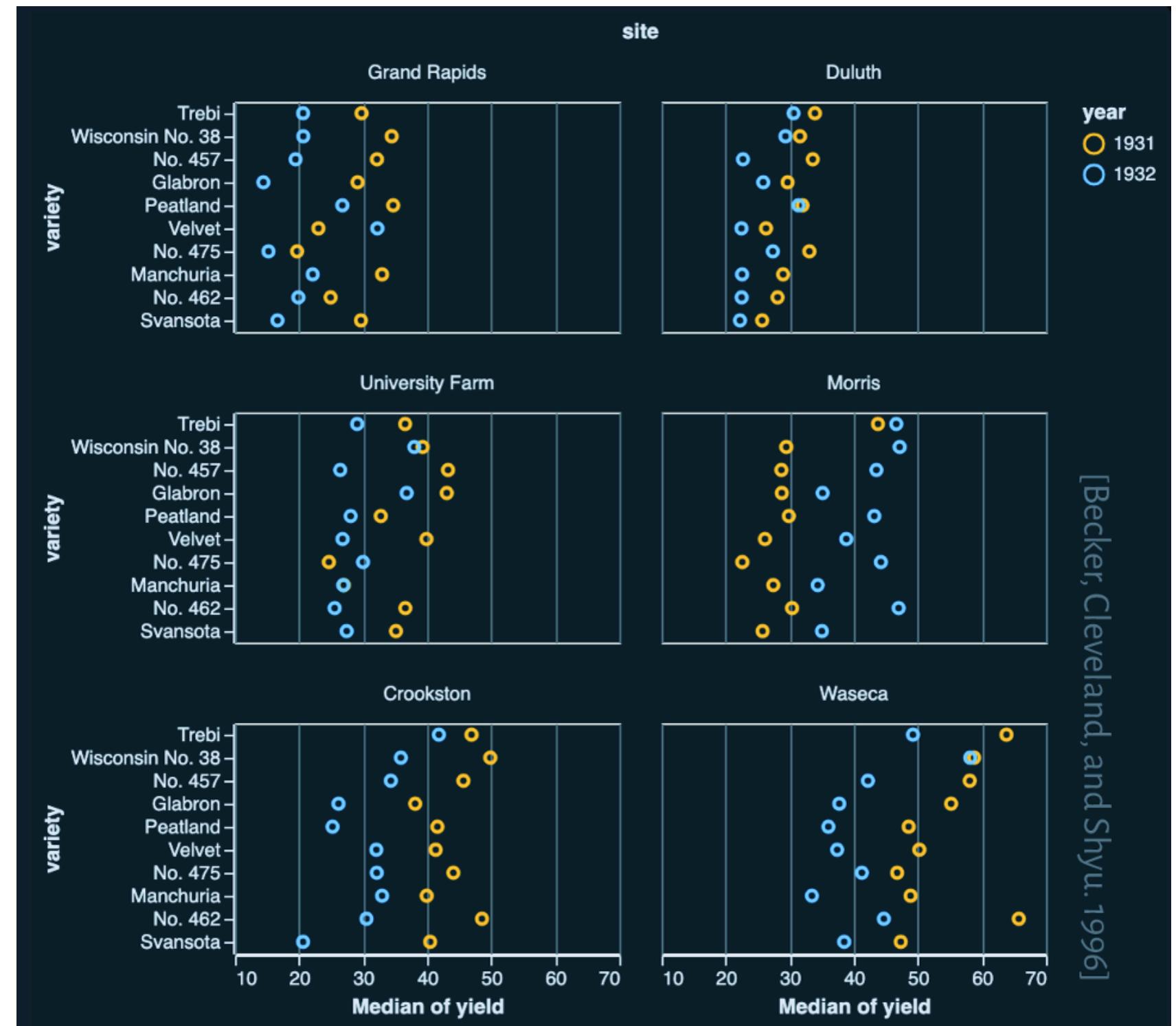
ultisol



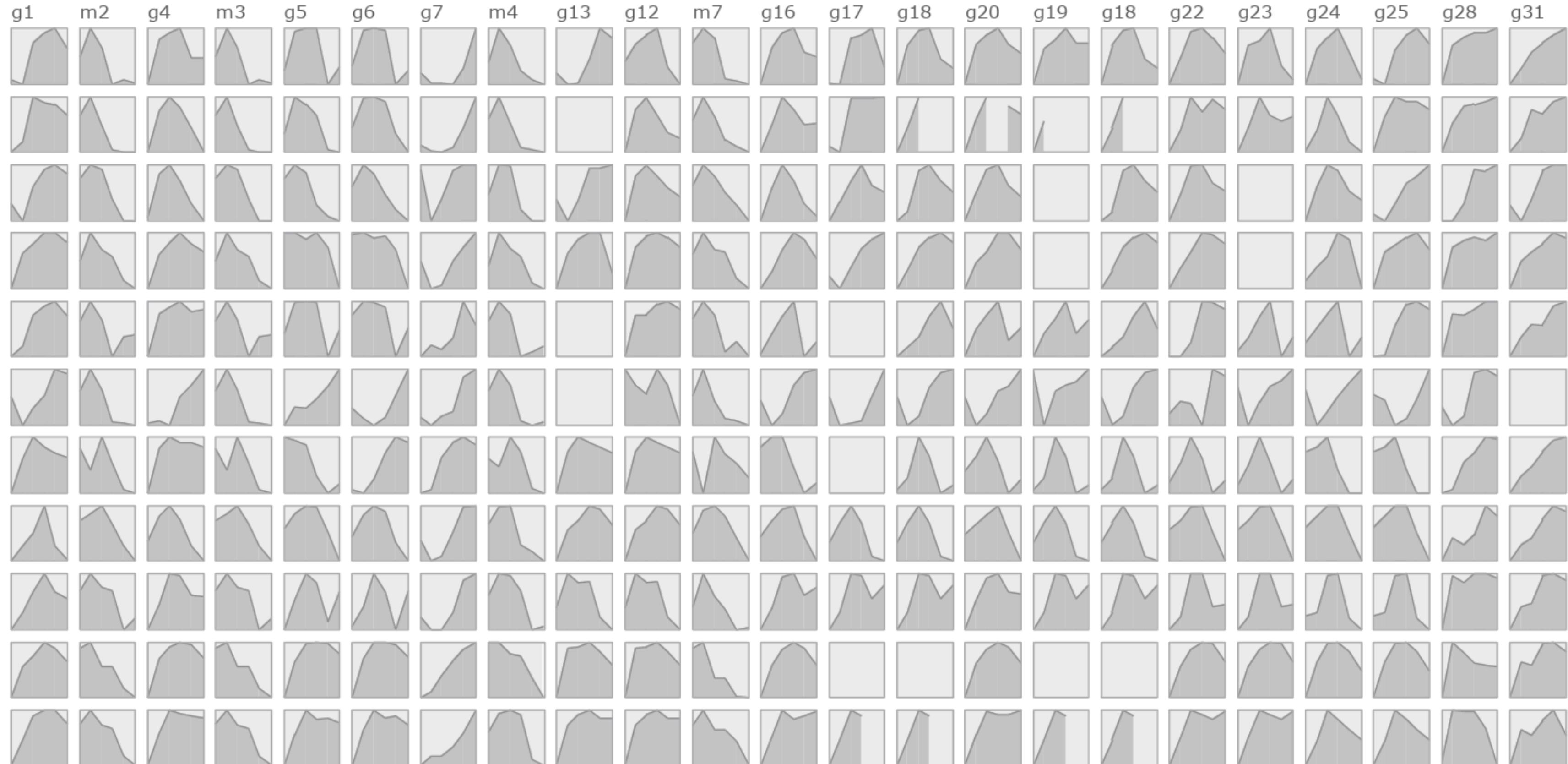
Shrink the Graphics (with small multiples)

Sometimes called Trellis plots

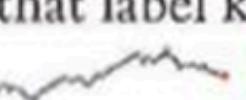
Categorical or Ordinal variables
are typically used as dimensions
for subdivision

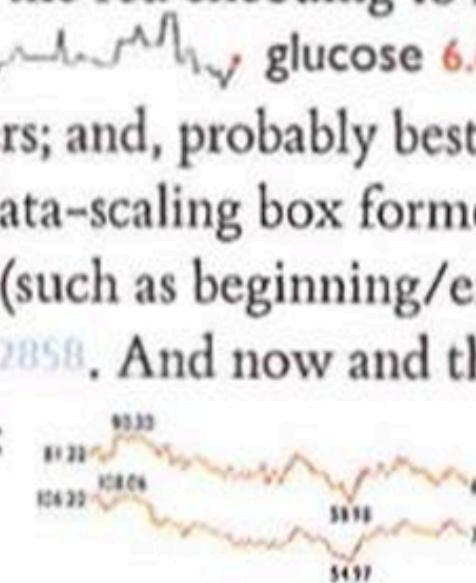


Shrink the Graphics (with small multiples)



Shrink the Graphics (with sparklines)

Dequantification In exchange for an enormous increase in graphical resolving power, the wordlike size of sparklines precludes the overt labels and scaling of conventional statistical displays. Most of our examples have, however, depicted *contextual methods* for quantifying sparklines: the gray bar for normal limits and the red encoding to link data points in sparklines to exact numbers  glucose 6.6; global scale bars and labels for sparkline clusters; and, probably best of all, surrounding a sparkline with an implicit data-scaling box formed by nearby numbers that label key data points (such as beginning/end, high/low) 1.1025  1.1907 1.0783 1.2858. And now and then sparklines might be scaled by very small type: 



Production methods Data lines produced by conventional statistical graphics programs must be gathered together, rescaled, and resized into sparklines. Sometimes this can be quickly done by cutting and pasting data lines, then resizing the printed output to sparkline resolutions. To produce and display really elegant sparklines, however, currently requires elaborate software: (1) a *page layout* program, (2) a *graphic design* program that gives complete control over type, tables, linework, and (3) a *statistical analysis* program to generate hundreds of chartjunk-free sparklines for export into design and layout operations. Once the basic templates for sparklines are worked out, then ongoing production and

Unseen and Unaware: Implications of Recent Research on Failures of Visual Awareness for Human–Computer Interface Design

D. Alexander Varakin and Daniel T. Levin

Vanderbilt University

Roger Fidler
Kent State University

But...

ABSTRACT

Because computers often rely on visual displays as a way to convey information to a user, recent research suggesting that people have detailed awareness of only a small subset of the visual environment has important implications for human–computer interface design. Equally important to basic limits of awareness is the fact that people often over-predict what they will see and become aware of. Together, basic failures of awareness and people’s failure to intuitively understand

Experimental Findings

People over-predict what they are actually aware of

Too many visual features makes it hard to find specific bits of information

Each visual object is not interpreted/understood deeply

**Unseen and Unaware:
Implications of Recent Research
on Failures of Visual Awareness for
Human–Computer Interface Design**

D. Alexander Varakin and Daniel T. Levin
Vanderbilt University

Roger Fidler
Kent State University

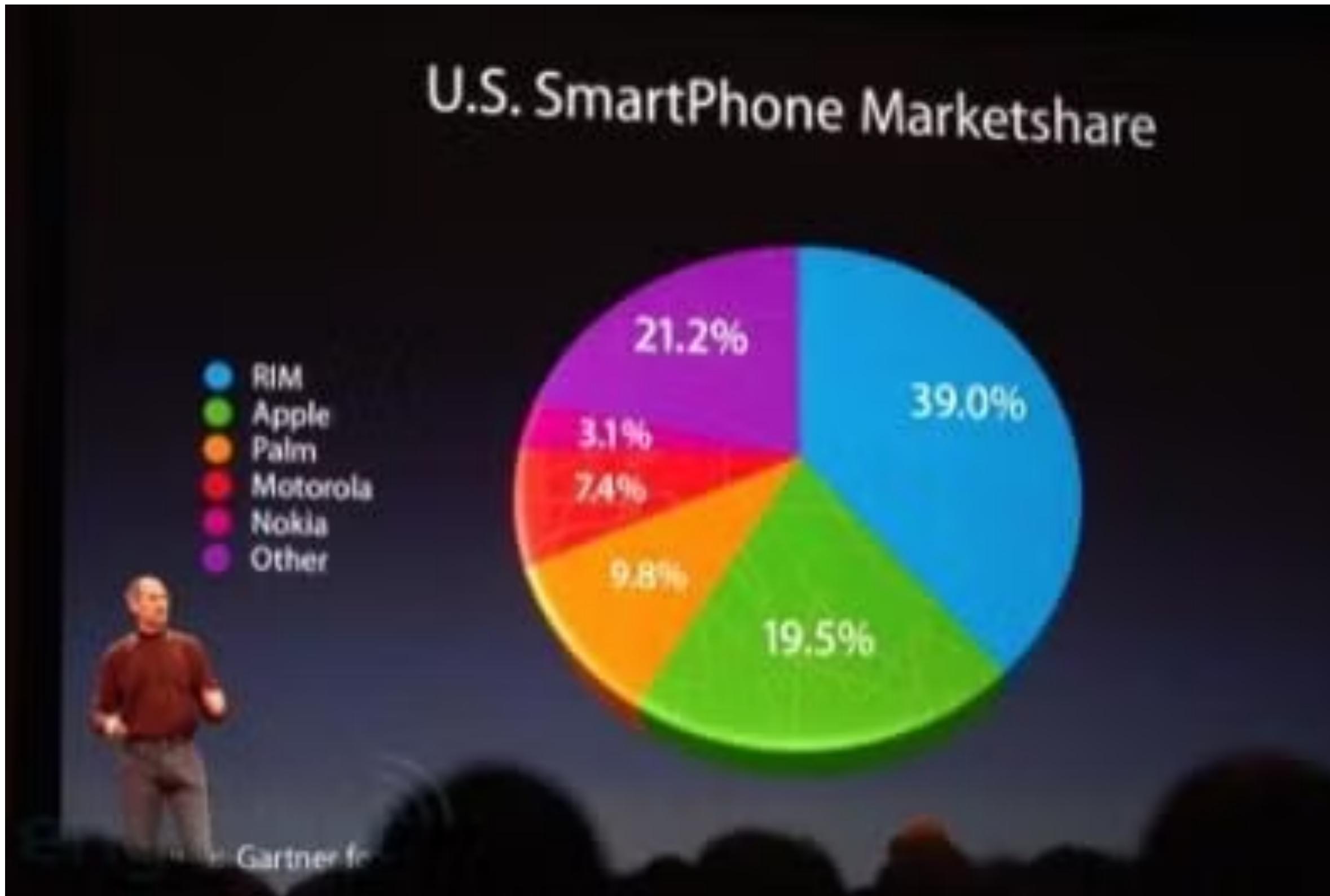
ABSTRACT

Because computers often rely on visual displays as a way to convey information to a user, recent research suggesting that people have detailed awareness of only a small subset of the visual environment has important implications for human–computer interface design. Equally important to basic limits of awareness is the fact that people often over-predict what they will see and become aware of. Together, basic failures of awareness and people’s failure to intuitively understand

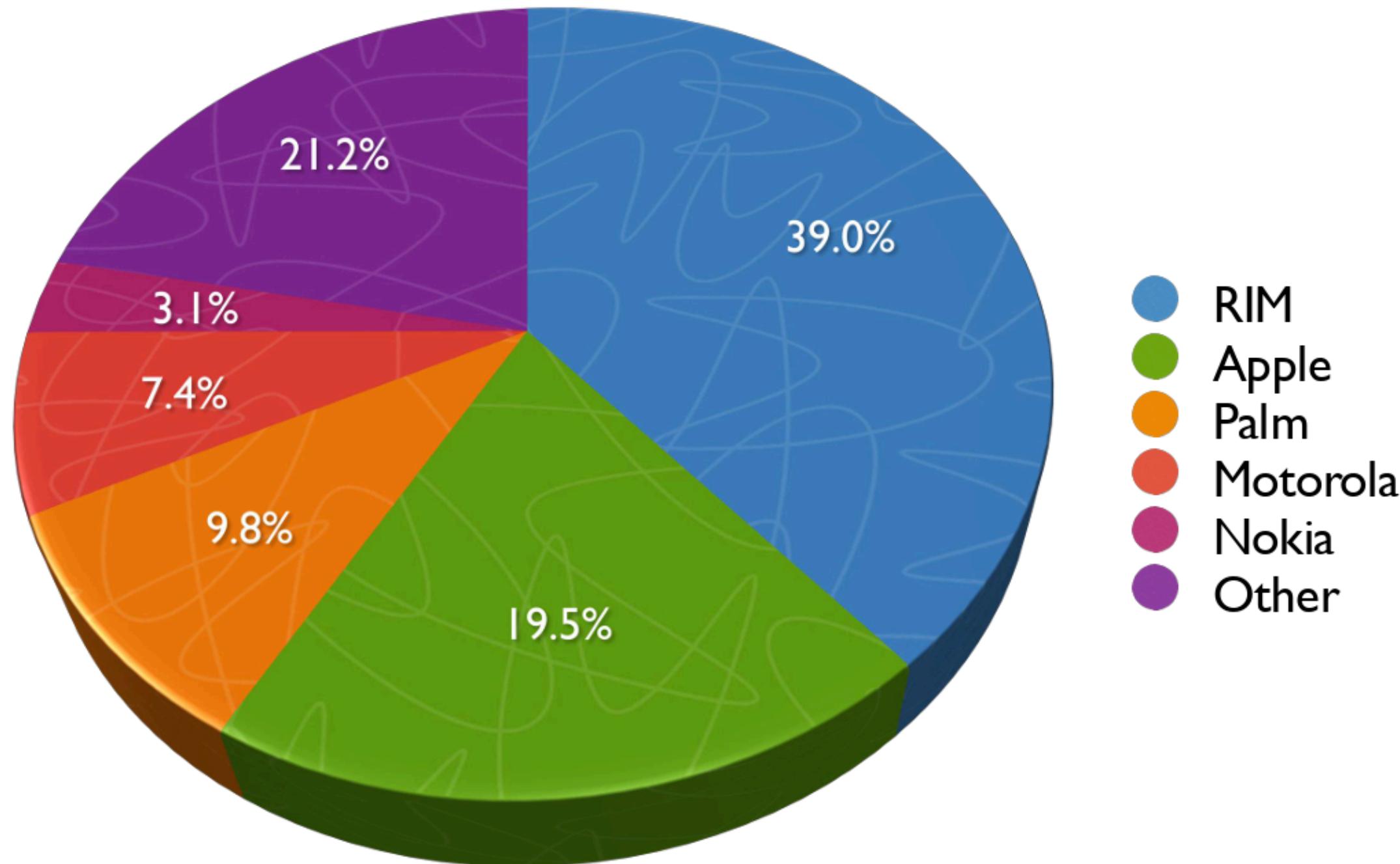
How do we achieve graphical excellence?

Avoid Misleading Encodings

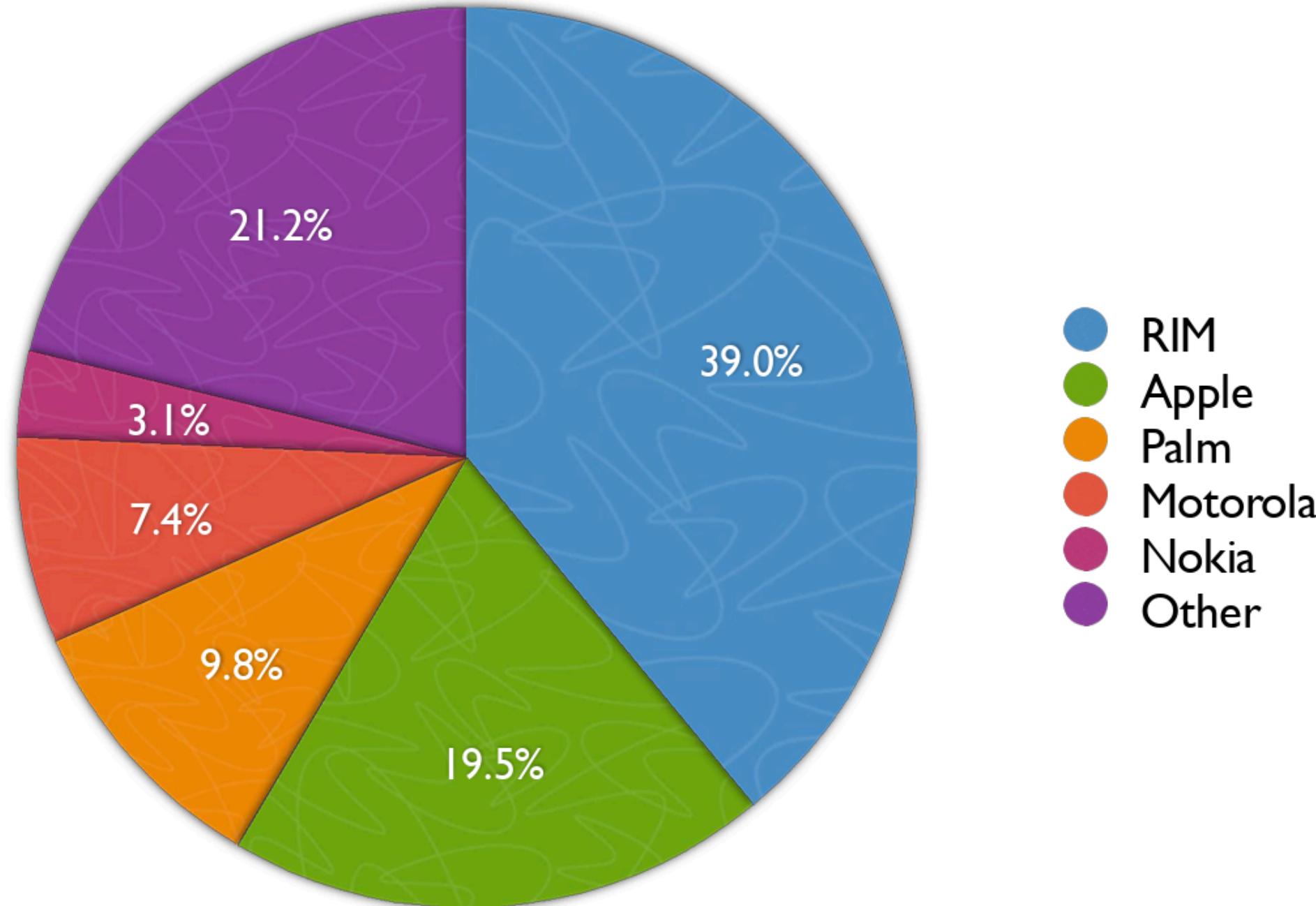
Misleading “U.S. Smartphone Marketshare” Vis



Misleading “U.S. Smartphone Marketshare” Vis



Better “U.S. Smartphone Marketshare” Vis



Better “U.S. Smartphone Marketshare” Vis

④ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)

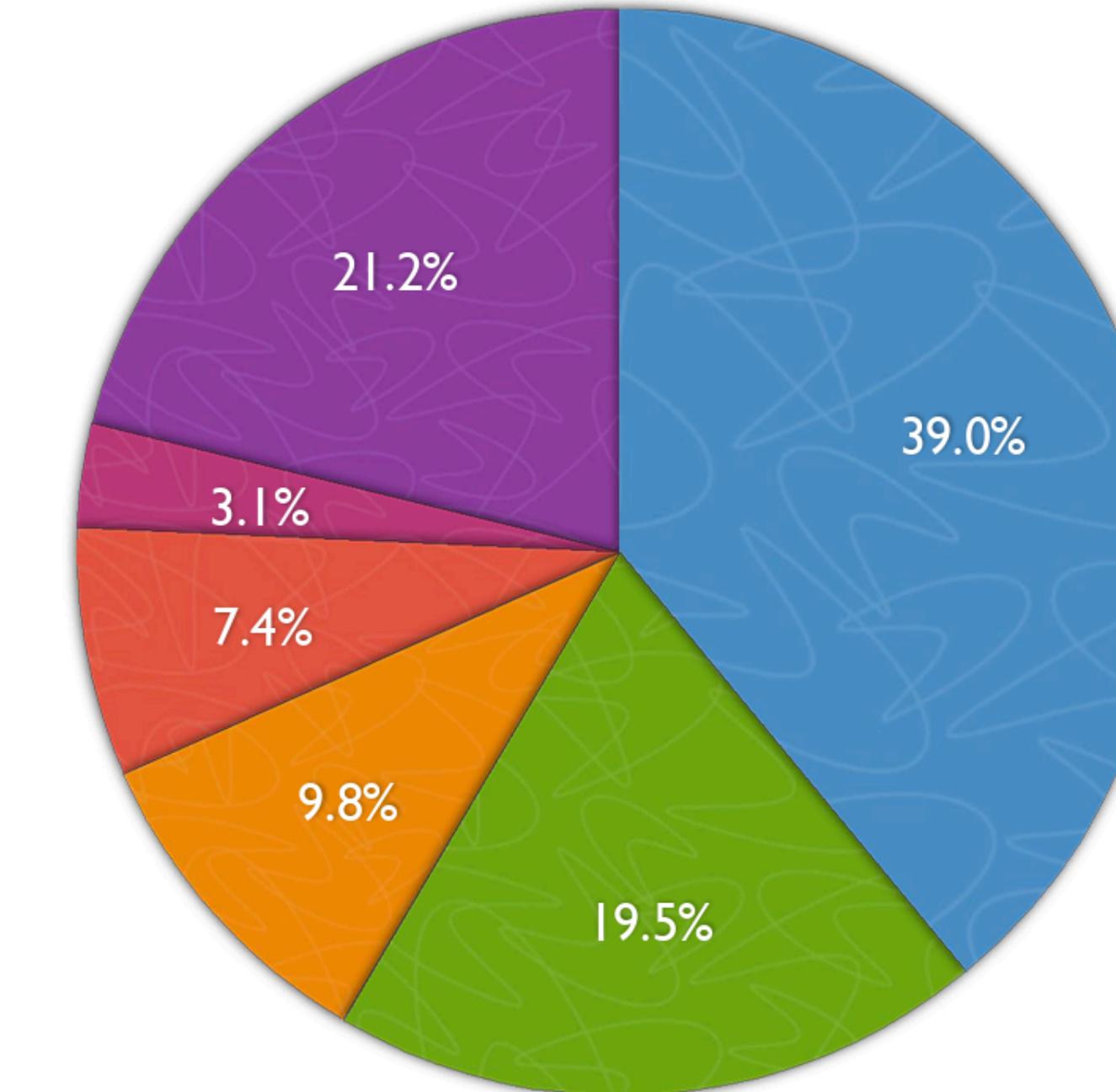


▲ Most

Effectiveness

▼ Least

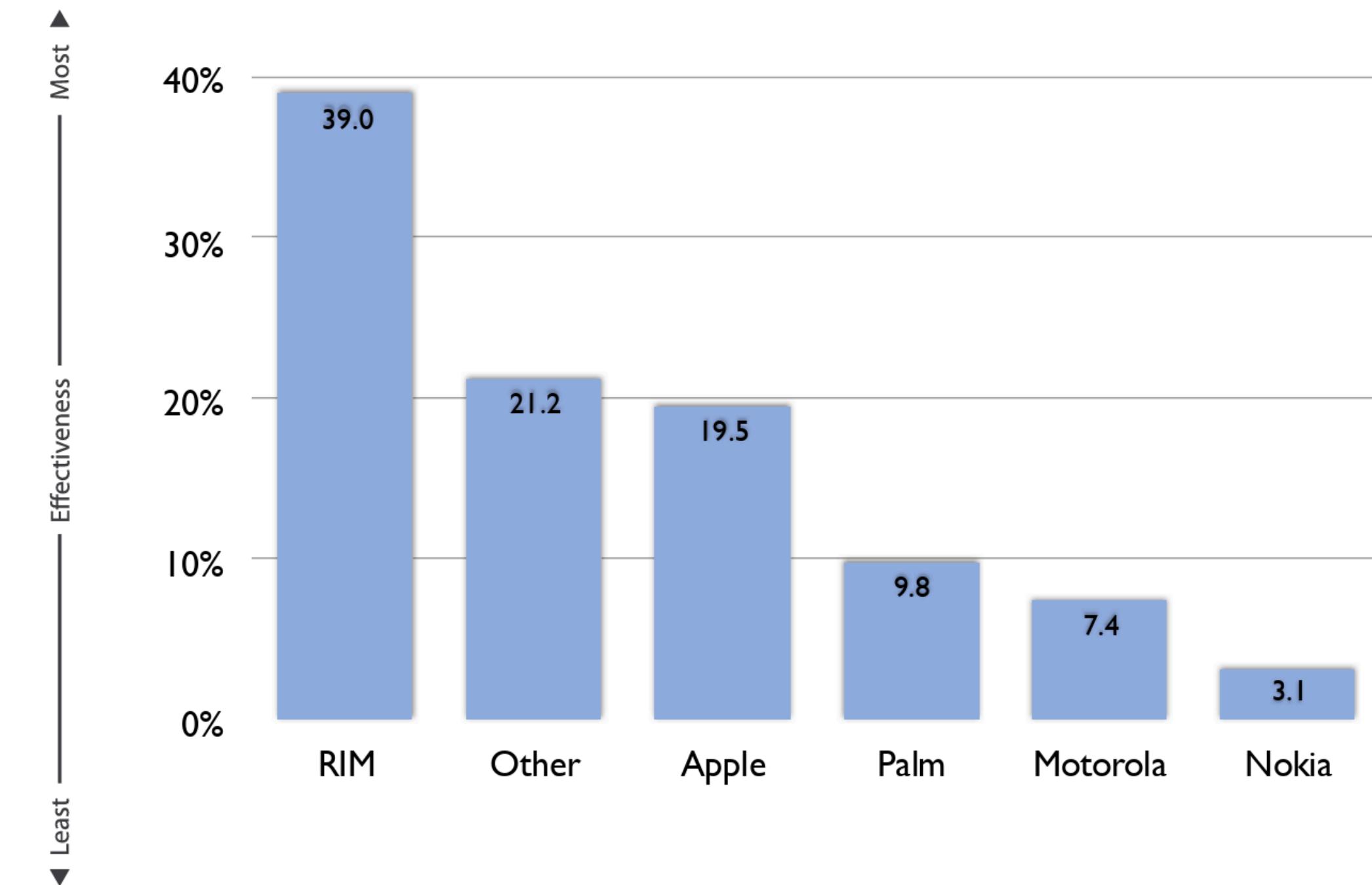
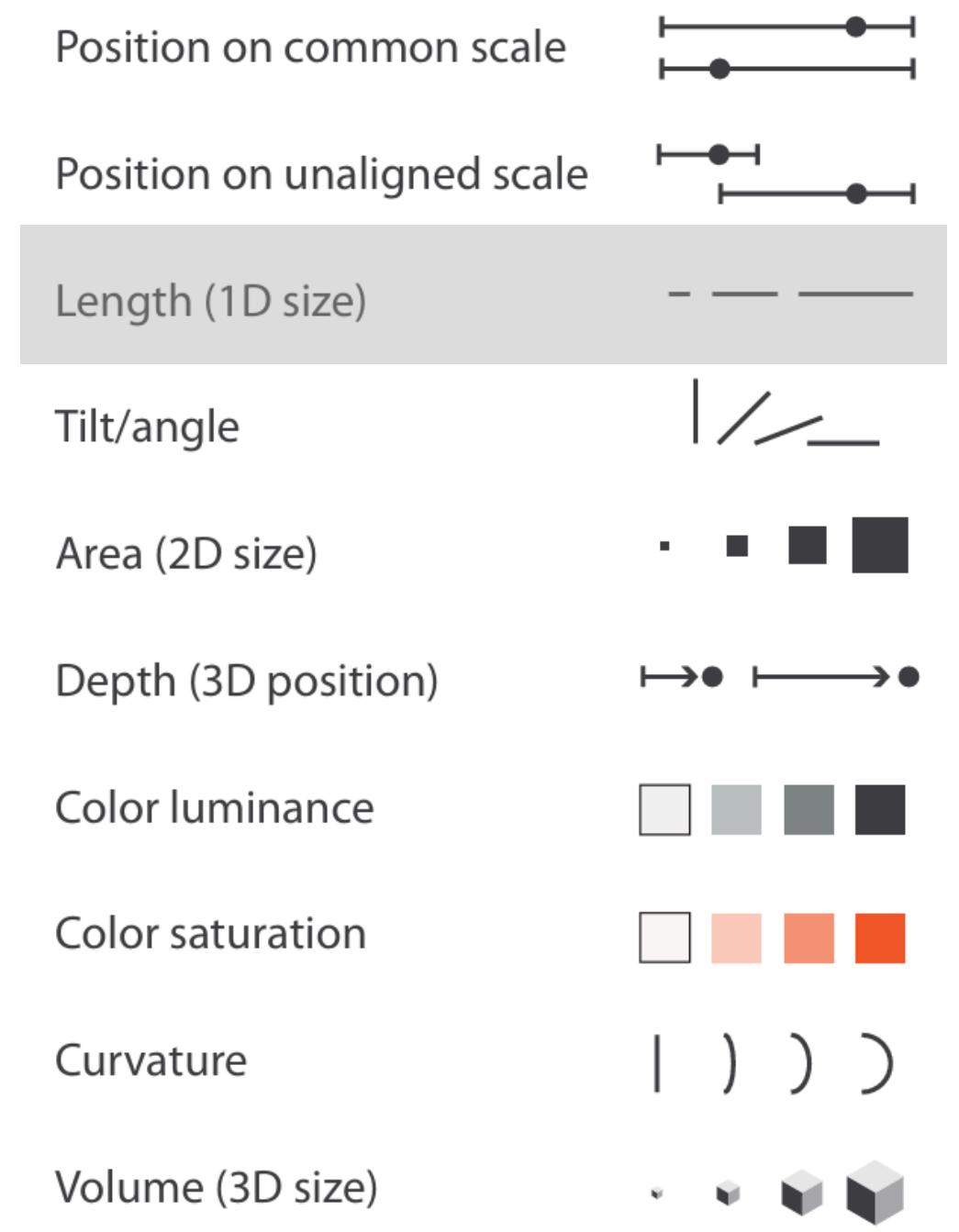
| Same |



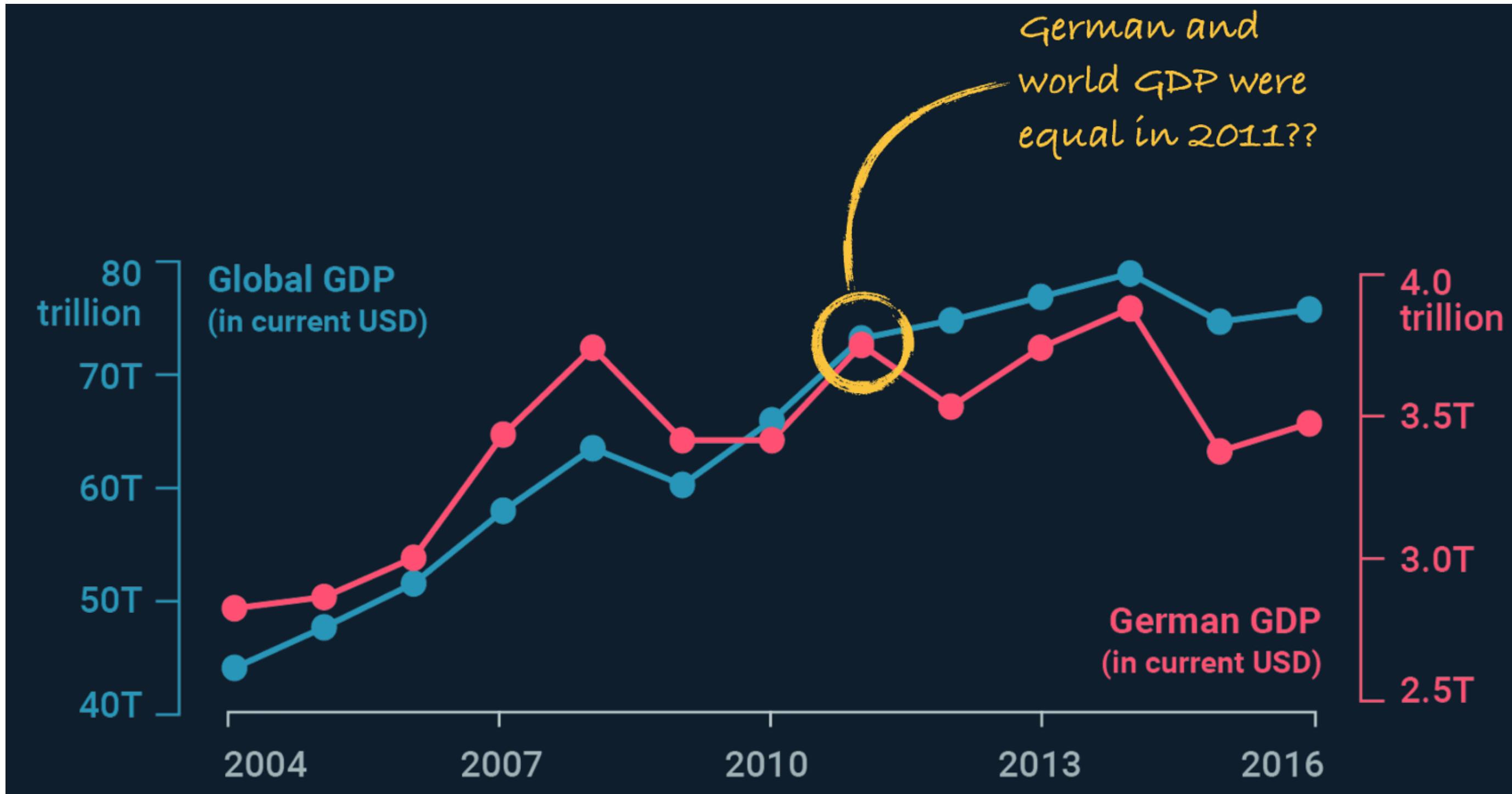
- RIM
- Apple
- Palm
- Motorola
- Nokia
- Other

Even Better “U.S. Smartphone Marketshare” Vis

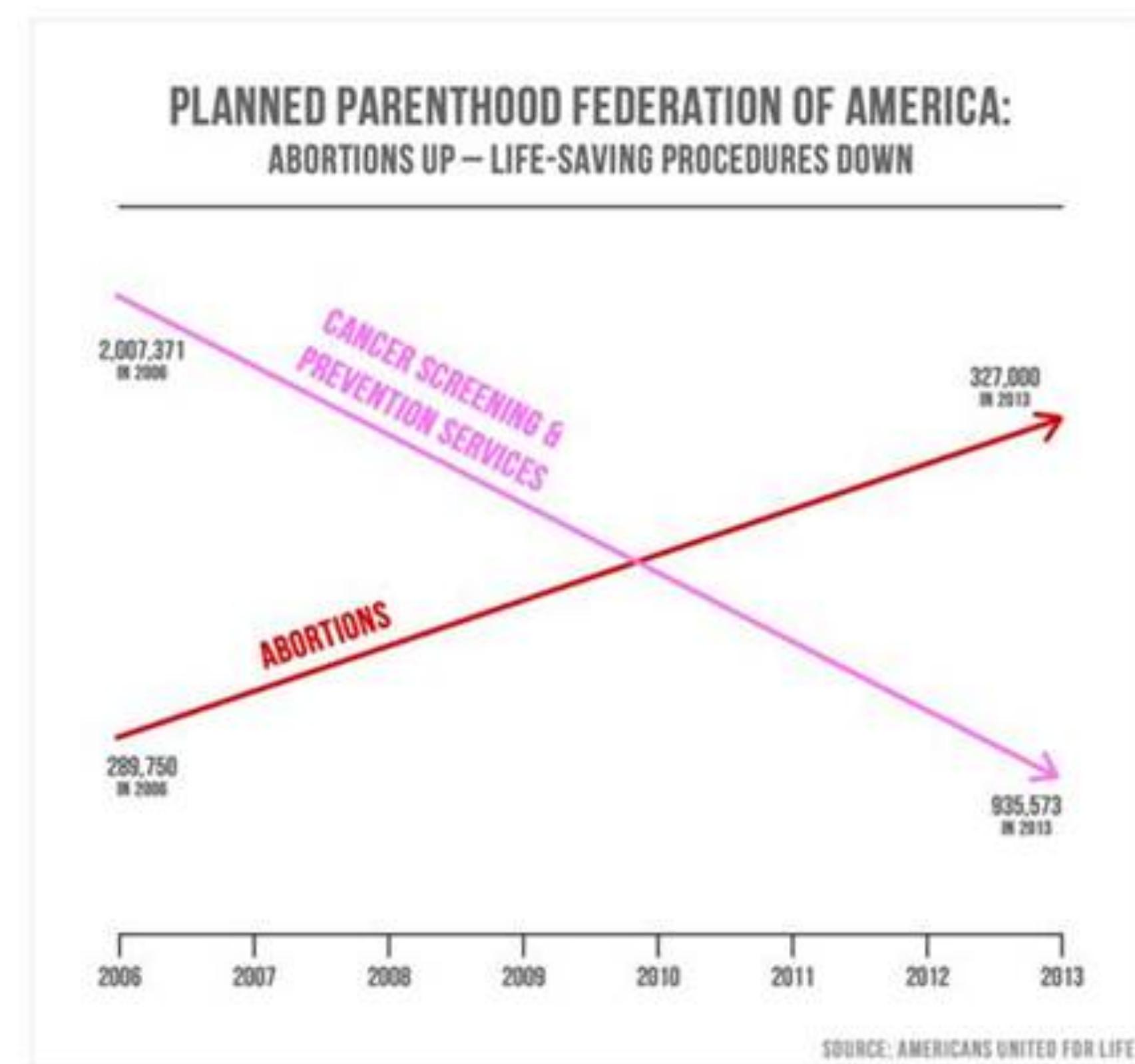
④ Magnitude Channels: Ordered Attributes



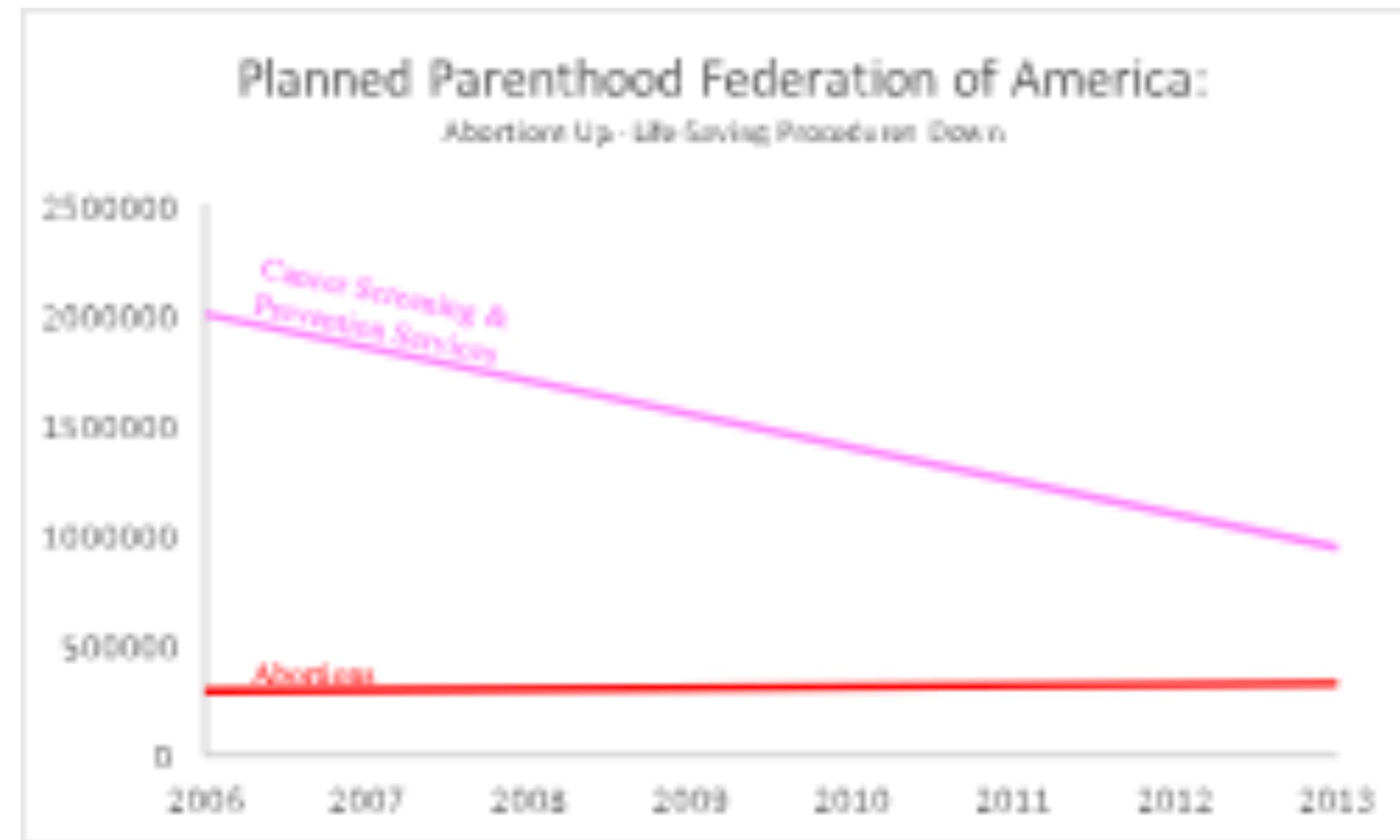
Avoid Dual Axis Charts



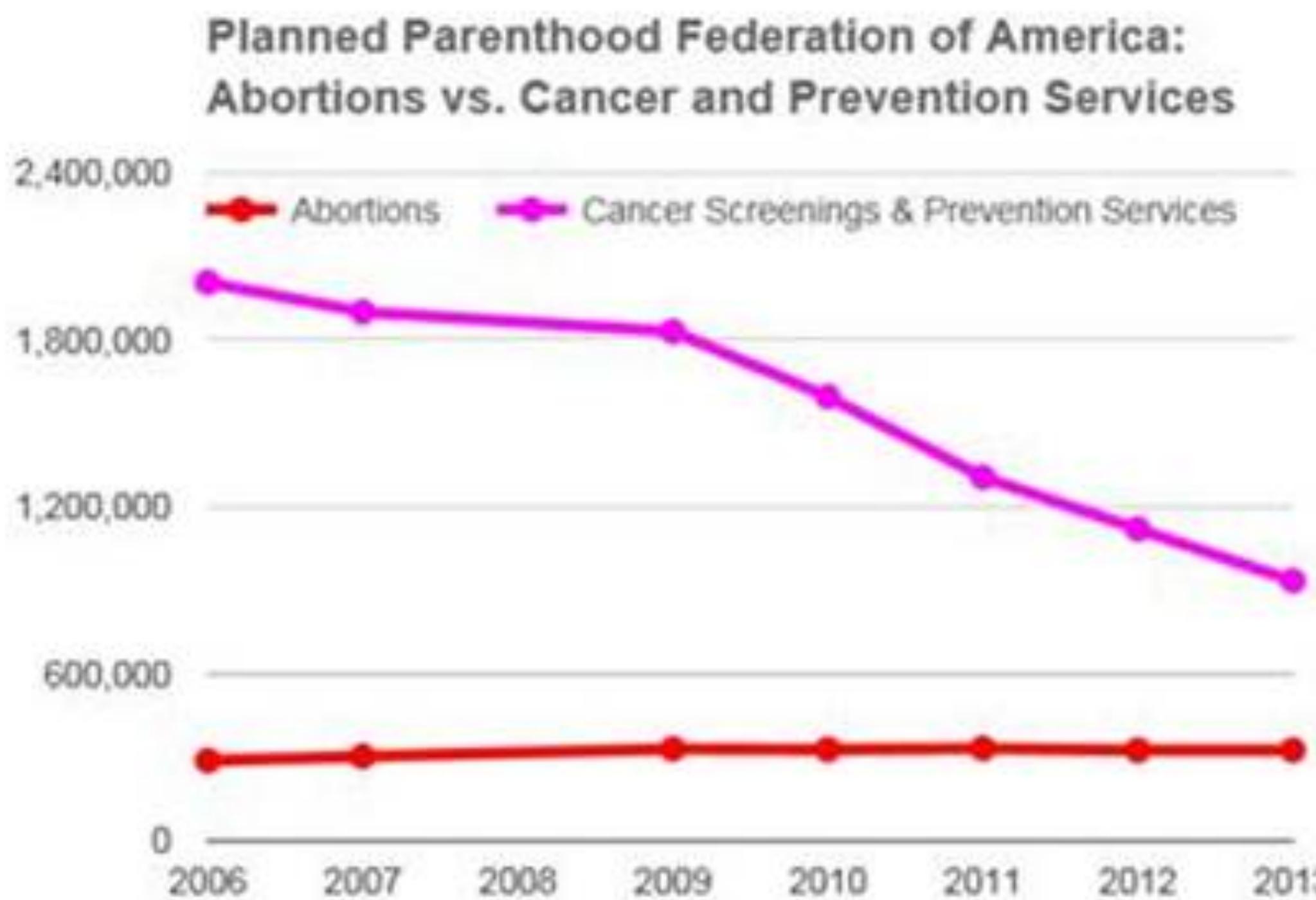
Avoid Dual Axis Charts



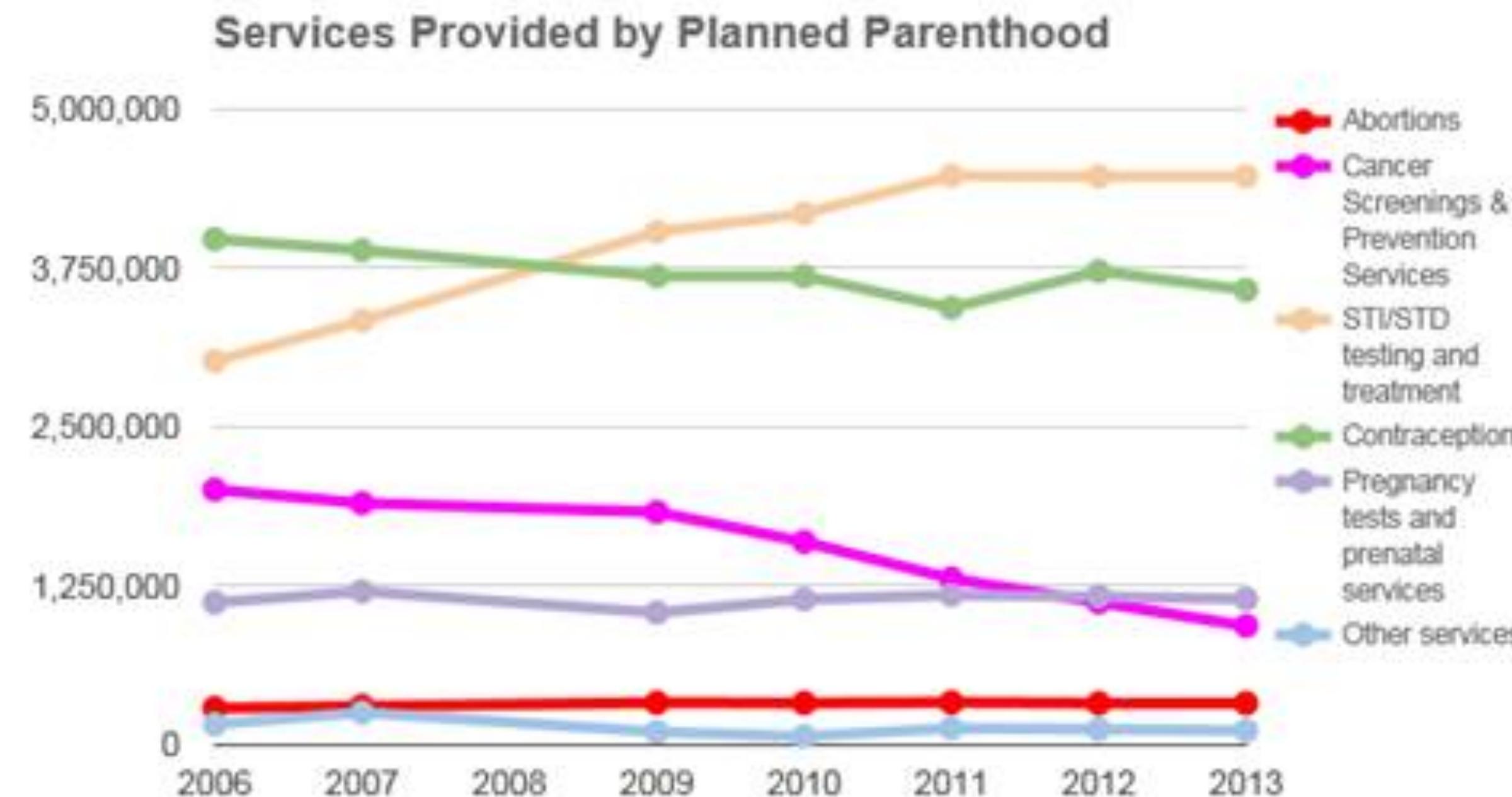
Avoid Dual Axis Charts



Provide Sufficient Context



Provide Sufficient Context

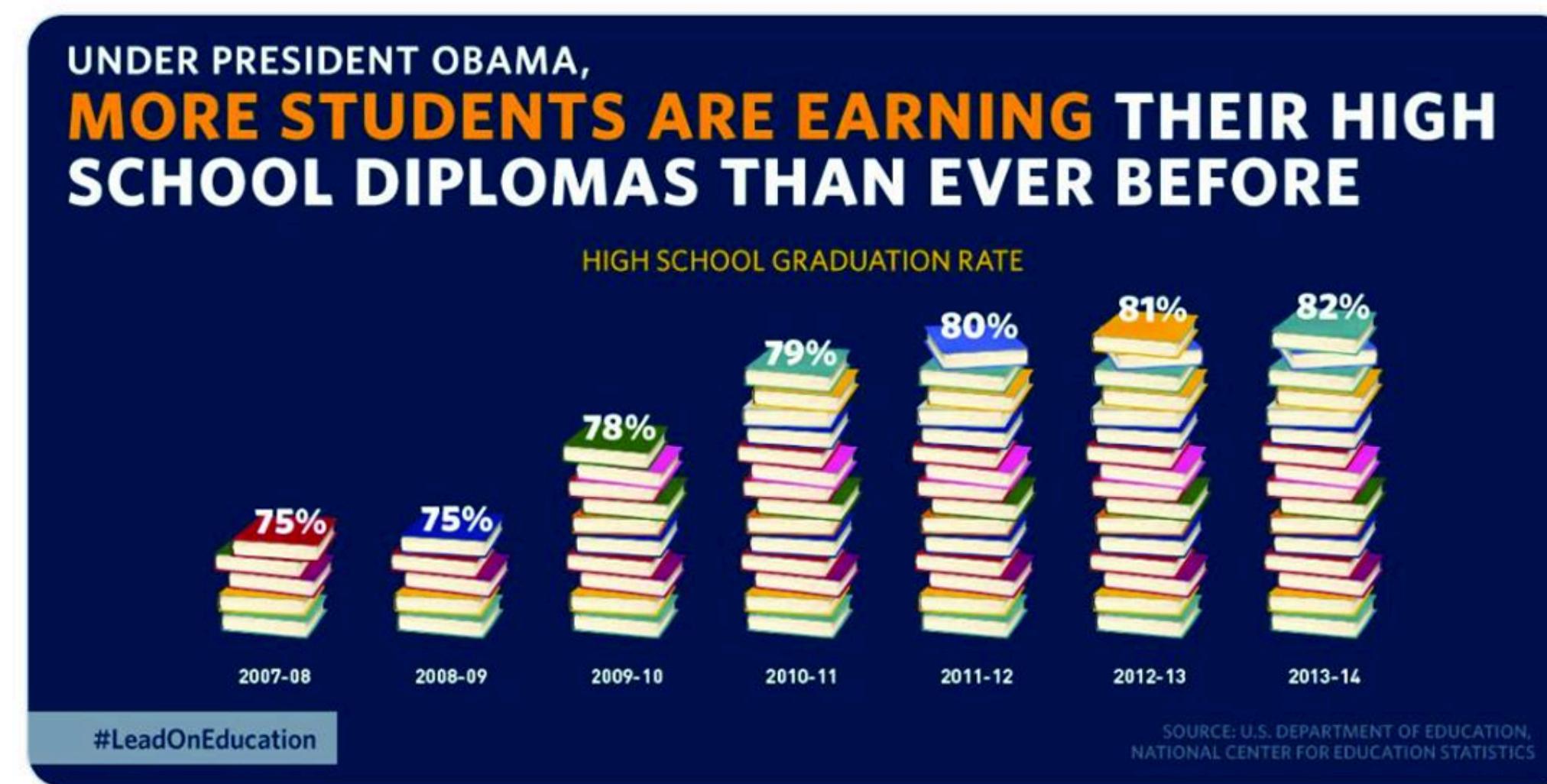


Provide Sufficient Context



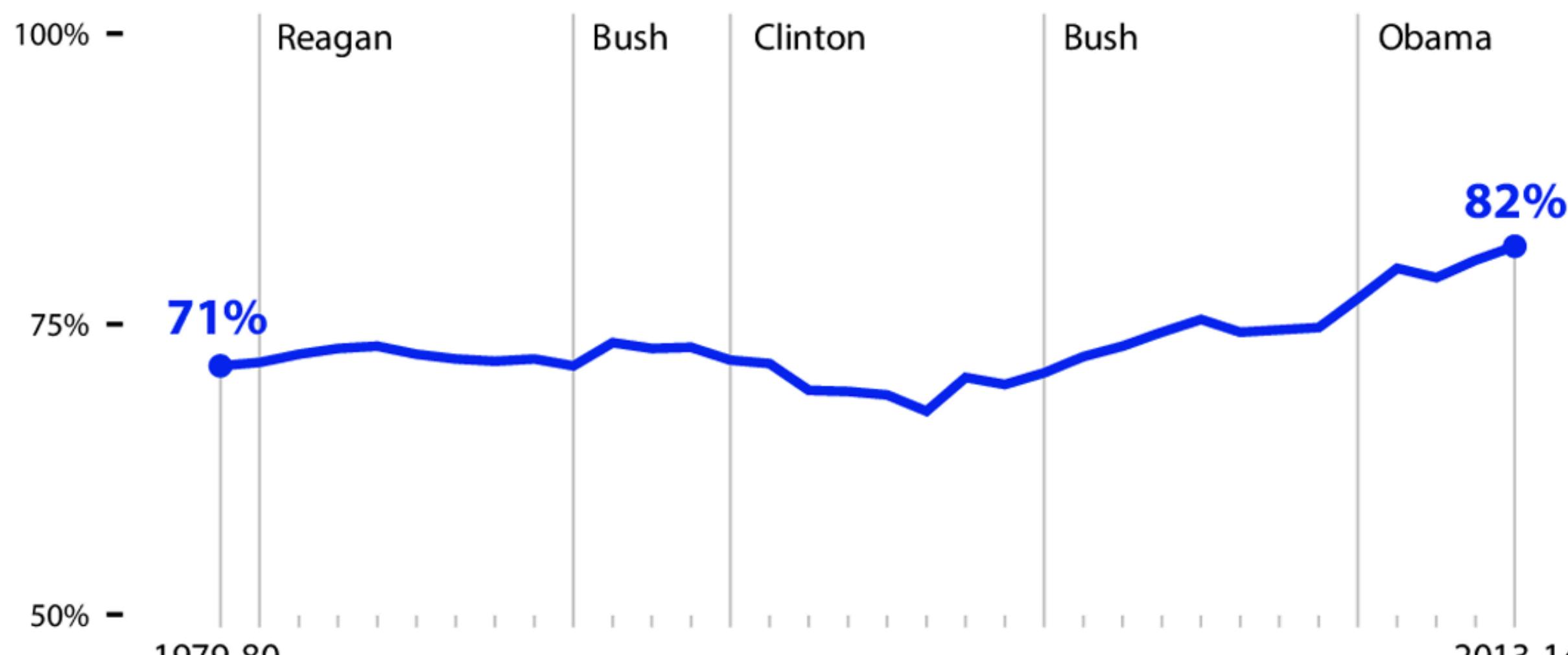
White House Archived @ObamaWhiteHouse

Good news: America's high school graduation rate has increased to an all-time high. wapo.st/1m40Mei



Provide Sufficient Context

High school graduation rates under each president

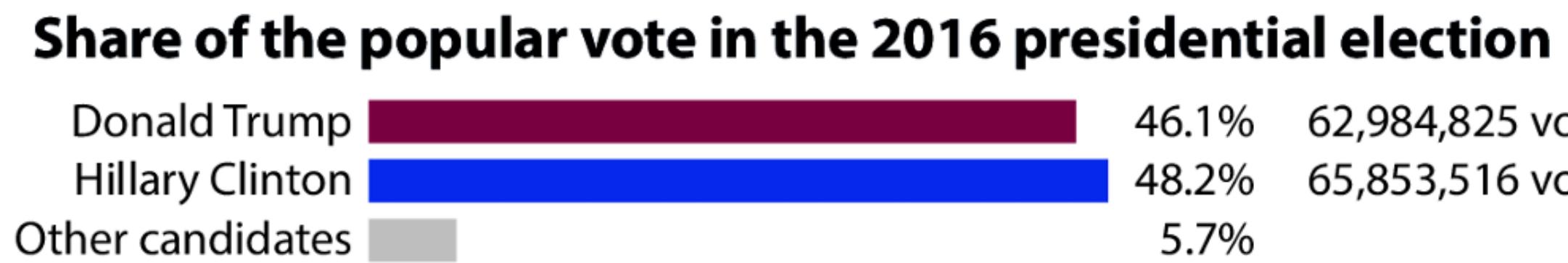


(Source: National Center for Education Statistics)

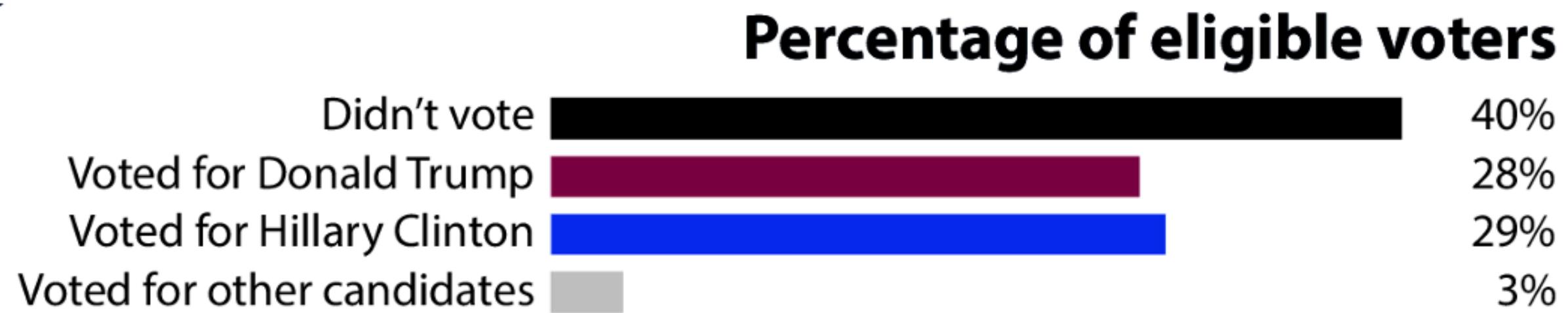
Provide Sufficient Context



Provide Sufficient Context



Provide Sufficient Context



No but for this one...

You should always avoid misleading encodings

FIN

Upcoming Dates

Apr 4: Project Milestone

**Homework 3 will be released on Mar 31
(Due Apr 11)**