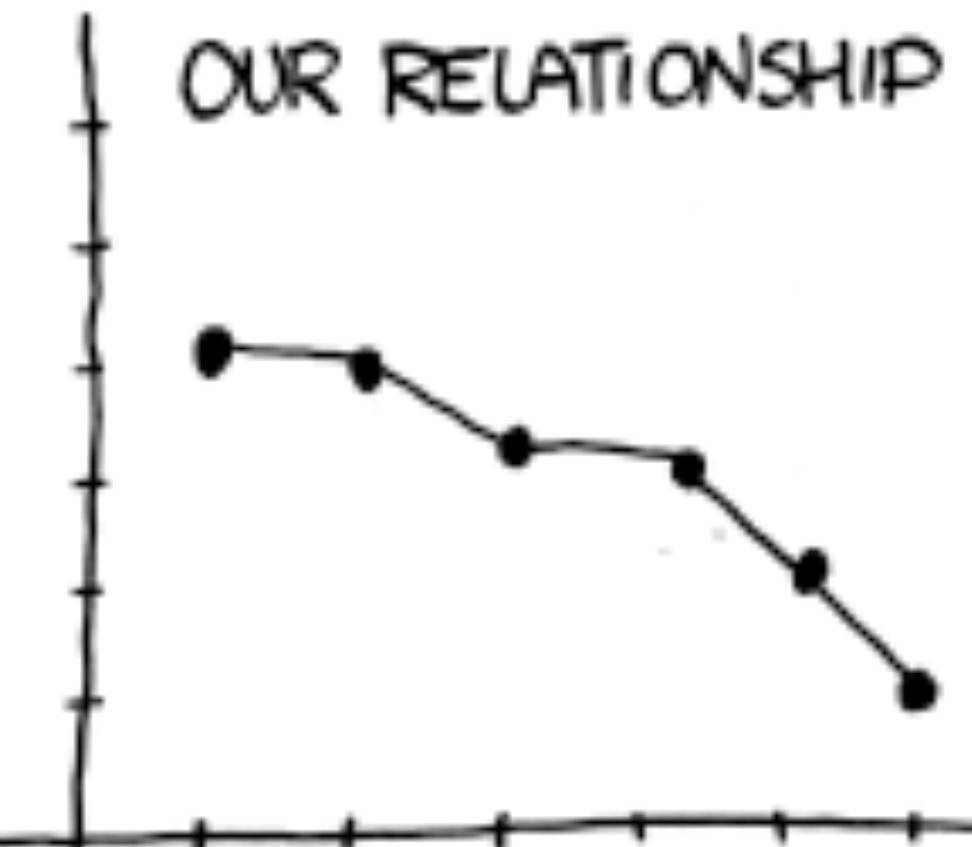


I THINK WE SHOULD
GIVE IT ANOTHER SHOT.

WE SHOULD BREAK
UP, AND I CAN
PROVE IT.



OUR RELATIONSHIP



HUH.



MAYBE YOU'RE RIGHT.

I KNEW DATA WOULD CONVINCE YOU.
NO, I JUST THINK I CAN DO
BETTER THAN SOMEONE WHO
DOESN'T LABEL HER AXES.



Source: xkcd.com/833

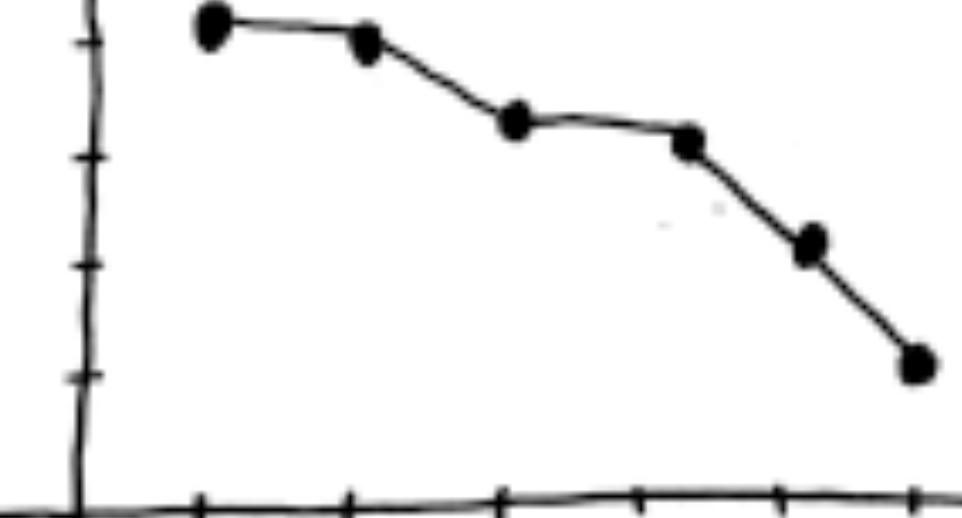
Practice using your Zoom Stamps here

I THINK WE SHOULD
GIVE IT ANOTHER SHOT.

WE SHOULD BREAK
UP, AND I CAN
PROVE IT.



OUR RELATIONSHIP



HUH.



MAYBE YOU'RE RIGHT.

I KNEW DATA WOULD CONVINCE YOU.
NO, I JUST THINK I CAN DO
BETTER THAN SOMEONE WHO
DOESN'T LABEL HER AXES.



Tooltip: ‘And if you labeled your axes, I could tell you exactly how MUCH better!’

Choosing an appropriate visualization & Exploratory Data Analysis

Part 1:

Principles of Effective Visualizations

Principle

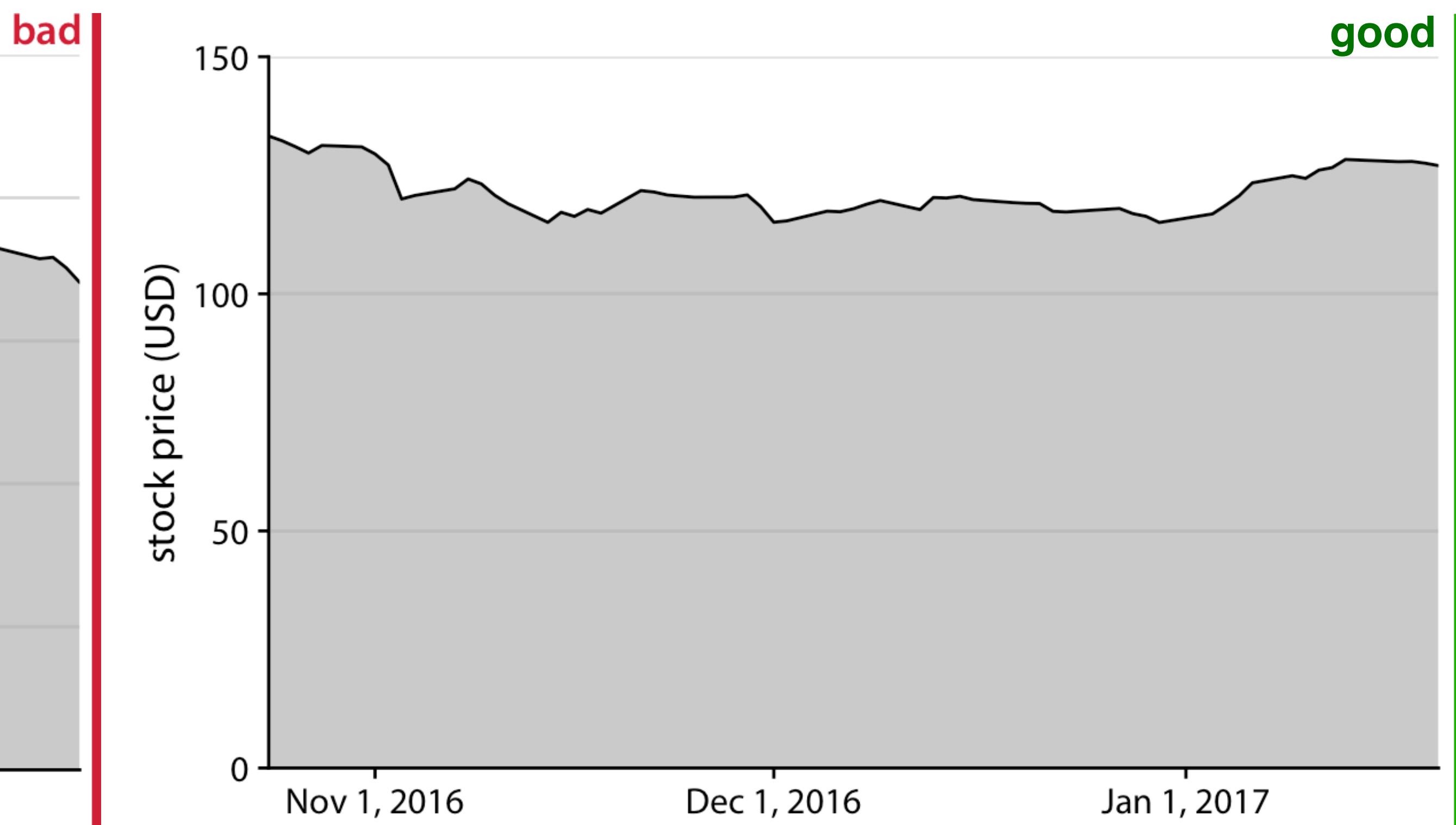
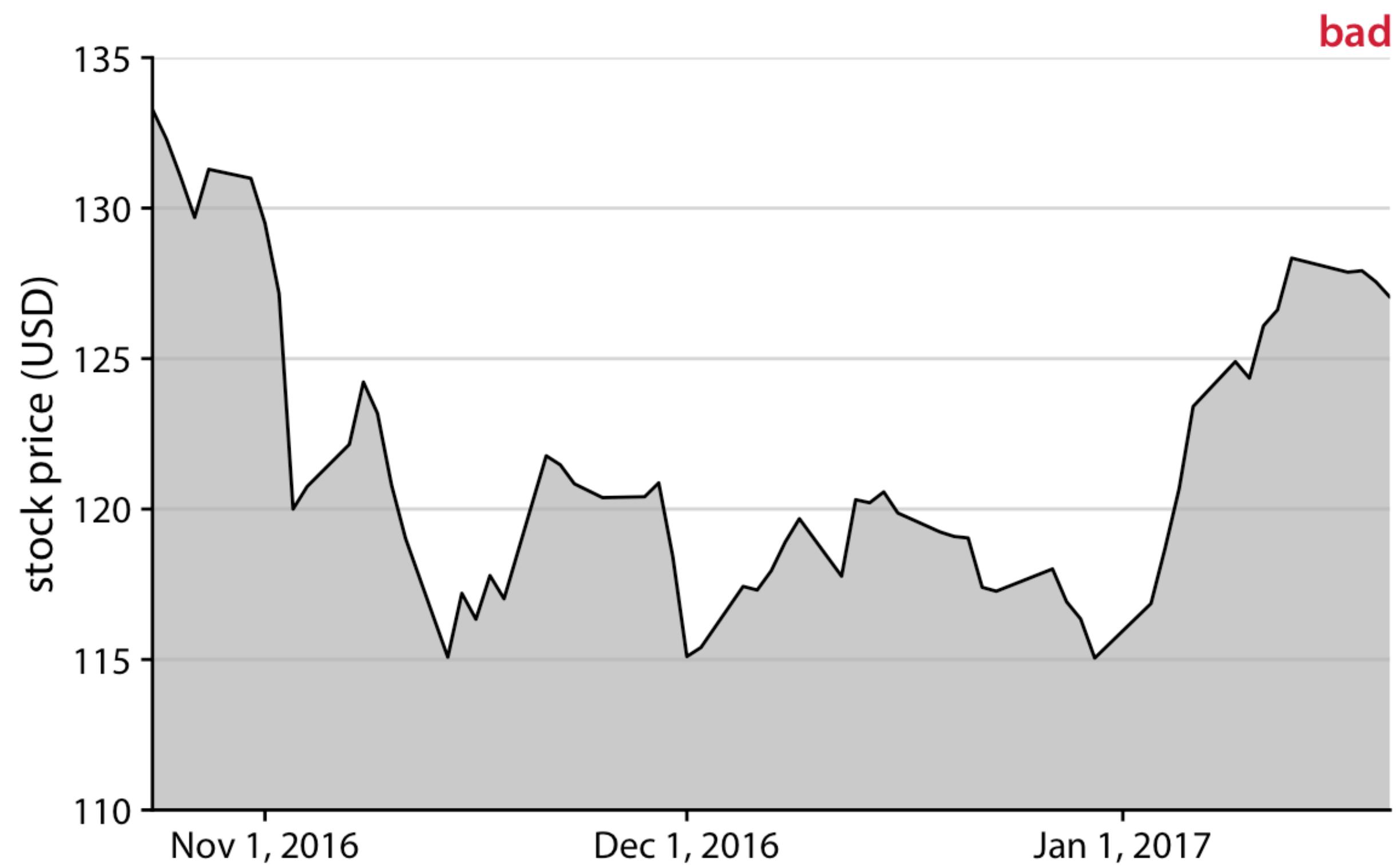
- Proportional Ink

Definition

The amount of ink used to indicate a value should be proportional to the value itself.

Examples

Truncating the y-axis on a bar chart to exaggerate the difference between bars violates the principle of proportional ink



Principle

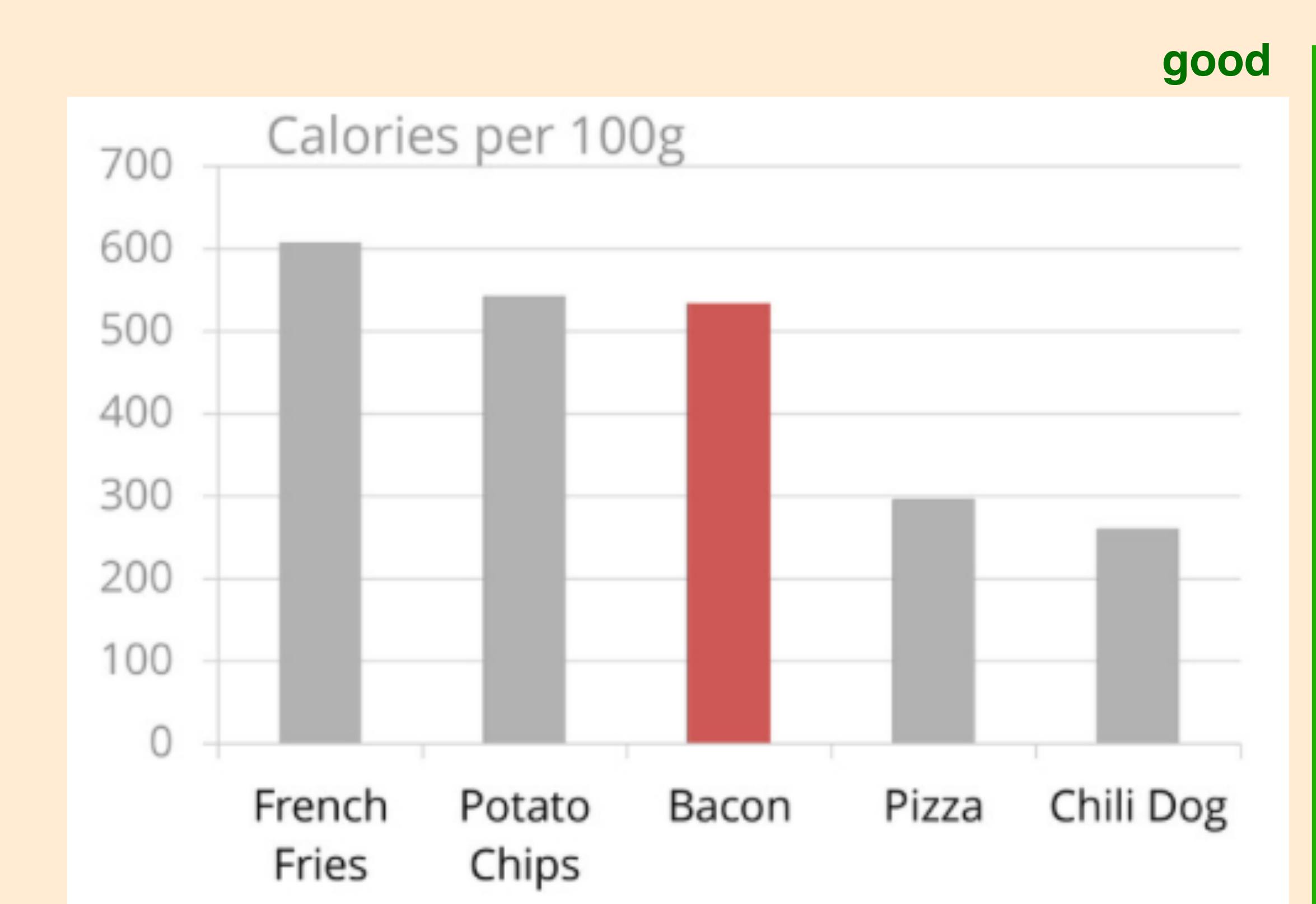
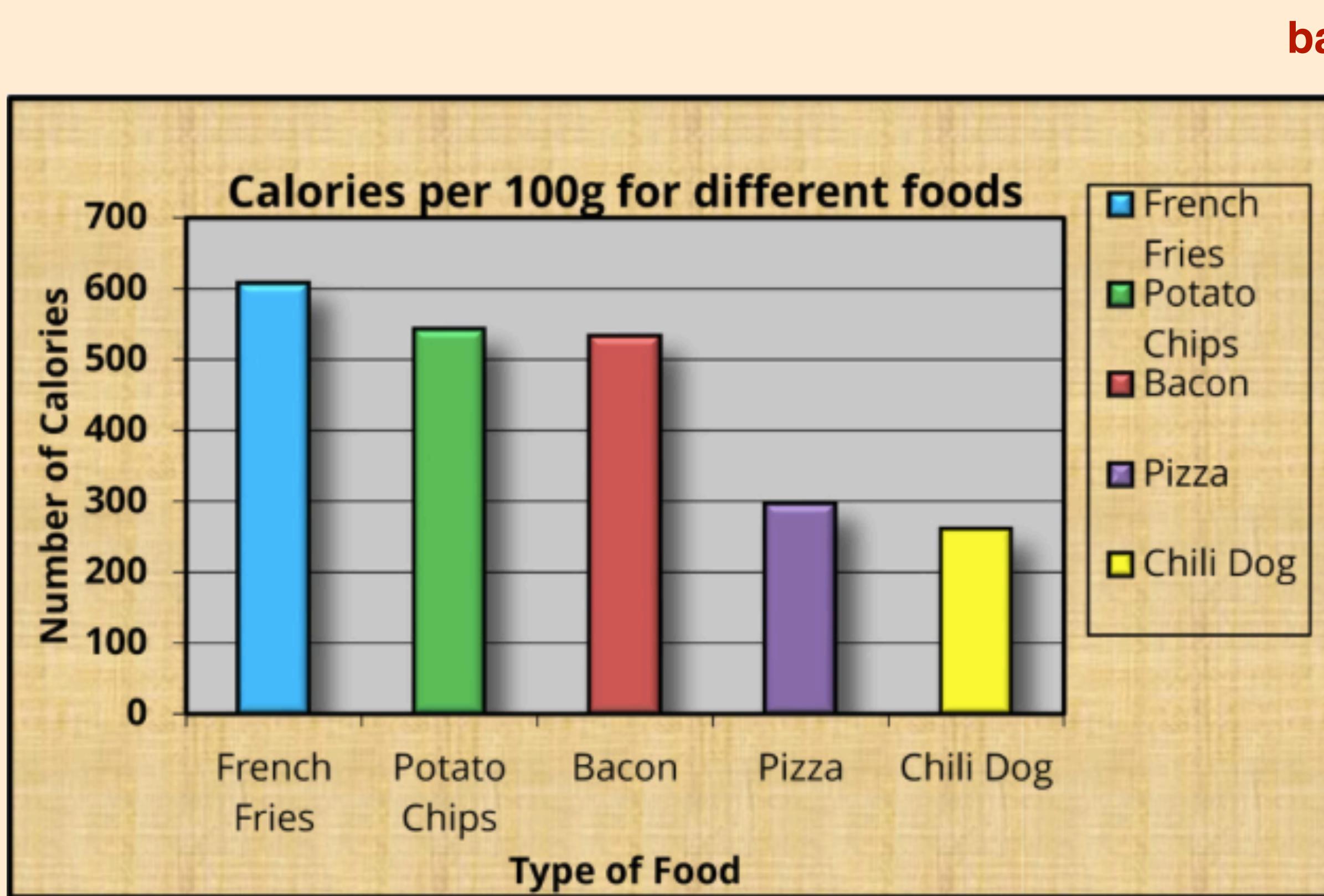
- Data:ink ratio

Definition

Remove distracting visual elements to focus attention on the data

Examples

Lighten line weights, remove backgrounds, never use 3D or special effects, remove unnecessary/redundant labels, etc



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 the y-axis 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, ~~remove~~ 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.

- **Overplotting**

With large datasets, points overlap, resulting in large clouds of data

To fix overplotting, could plot just a sample subset of the data, use alpha, and use smaller points. Or, jitter - but check if appropriate!

- **Visualization choice**

Must be informed by the **data** you have, the **research question** being asked and the **audience** that cares.

Pick the simplest plot that best shows most/all of the data needed to answer the research question. If you only have summary statistics, cannot show distributions. Tailor the visualization to your audience (within reason) but don't dumb it down.

- **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.

Part 2:

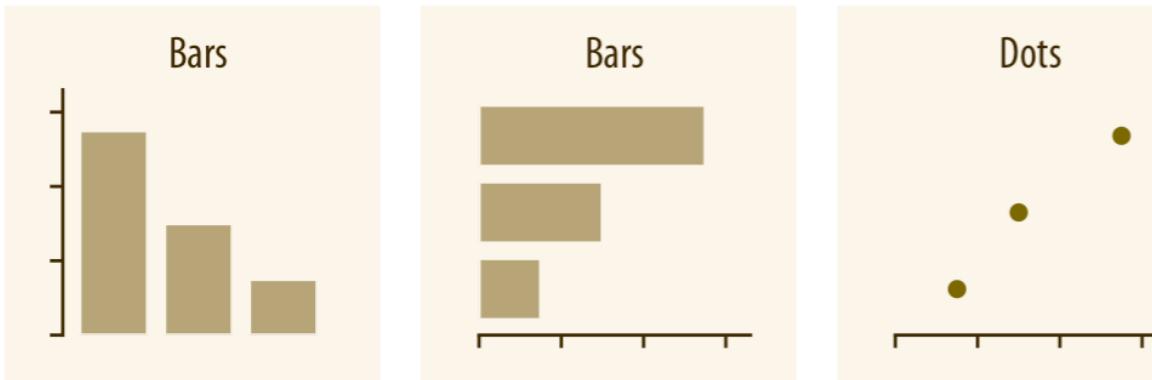
Choosing an appropriate data visualization

Directory of Visualizations

Fundamentals of Data Visualization

visualize data. It is meant both to serve as a table of contents, in case you are looking for a particular visualization whose name you may not know, and as a source of inspiration, if you need to find alternatives to the figures you routinely make.

5.1 Amounts



The most common approach to visualizing amounts (i.e., numerical values shown for some set of categories) is using bars, either vertically or horizontally arranged (Chapter 6). However, instead of using bars, we can also place dots at the location where the corresponding bar would end (Chapter 6).



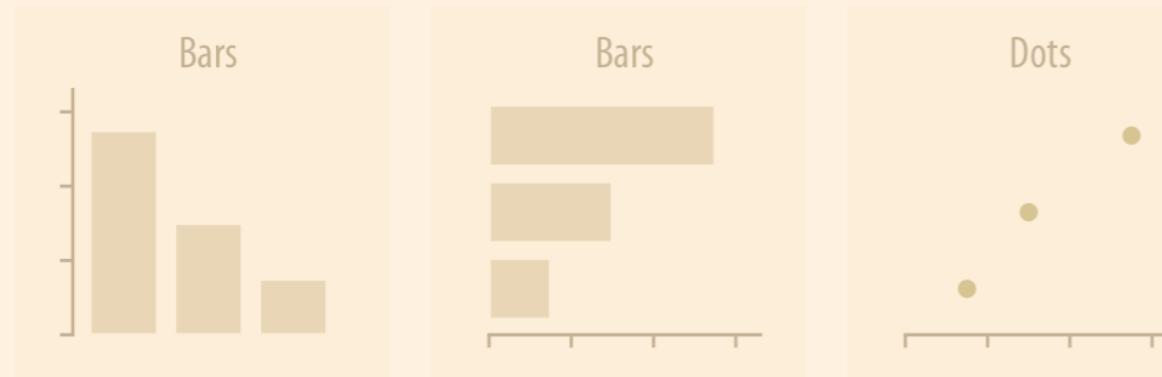
If there are two or more sets of categories for which we want to show amounts, we can group or stack the bars (Chapter 6). We can also map the categories onto the x and y axis and show amounts by color, via a heatmap (Chapter 6).

Directory of Visualizations

Fundamentals of Data Visualization

visualize data. It is meant both to serve as a table of contents, in case you are looking for a particular visualization whose name you may not know, and as a source of inspiration, if you need to find alternatives to the figures you routinely make.

5.1 Amounts



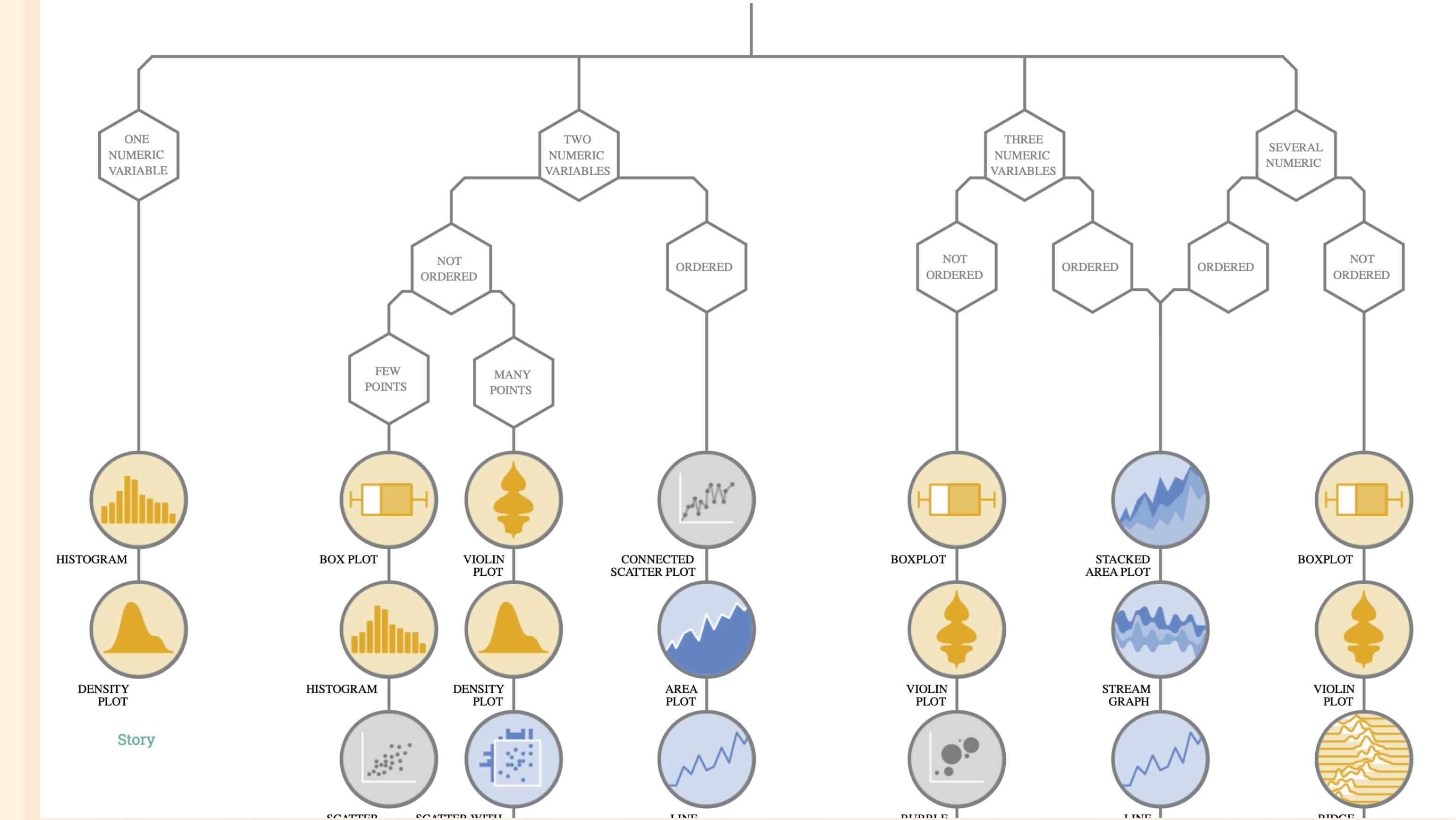
The most common approach to visualizing amounts (i.e., numerical values shown for some set of categories) is using bars, either vertically or horizontally arranged (Chapter 6). However, instead of using bars, we can also place dots at the location where the corresponding bar would end (Chapter 6).



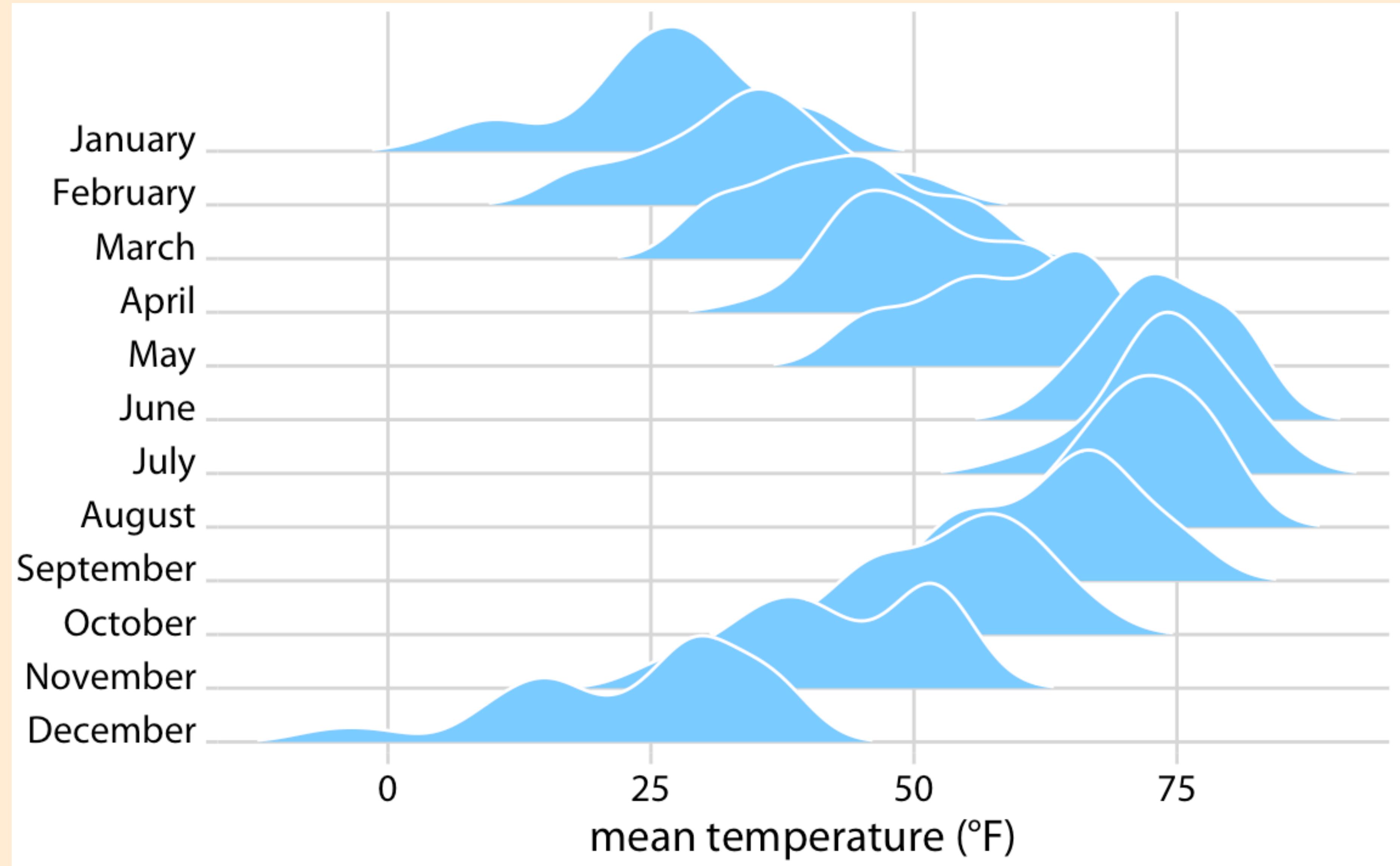
If there are two or more sets of categories for which we want to show amounts, we can group or stack the bars (Chapter 6). We can also map the categories onto the x and y axis and show amounts by color, via a heatmap (Chapter 6).

What kind of data do you have? Pick the main type using the buttons below. Then let the decision tree guide you toward your graphic possibilities.

Numeric Categoric Num & Cat Maps Network Time series



My favourite: Ridgeline plot



Source: [Fig 9.9 of Fundamentals of Data Visualization](#)

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 the y-axis 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, ~~remove~~ 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.

- **Overplotting**

With large datasets, points overlap, resulting in large clouds of data

To fix overplotting, could plot just a sample subset of the data, use alpha, and use smaller points. Or, jitter - but check if appropriate!

- **Visualization choice**

Must be informed by the **data** you have, the **research question** being asked and the **audience** that cares.

Pick the simplest plot that best shows most/all of the data needed to answer the research question. If you only have summary statistics, cannot show distributions. Tailor the visualization to your audience (within reason) but don't dumb it down.

- **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.

Revisiting a principle...

Principle

- Visualization choice

Definition

Must be informed by:

- 1) the **data** you have,
- 2) the **research question** being asked and
- 3) the **audience** that cares

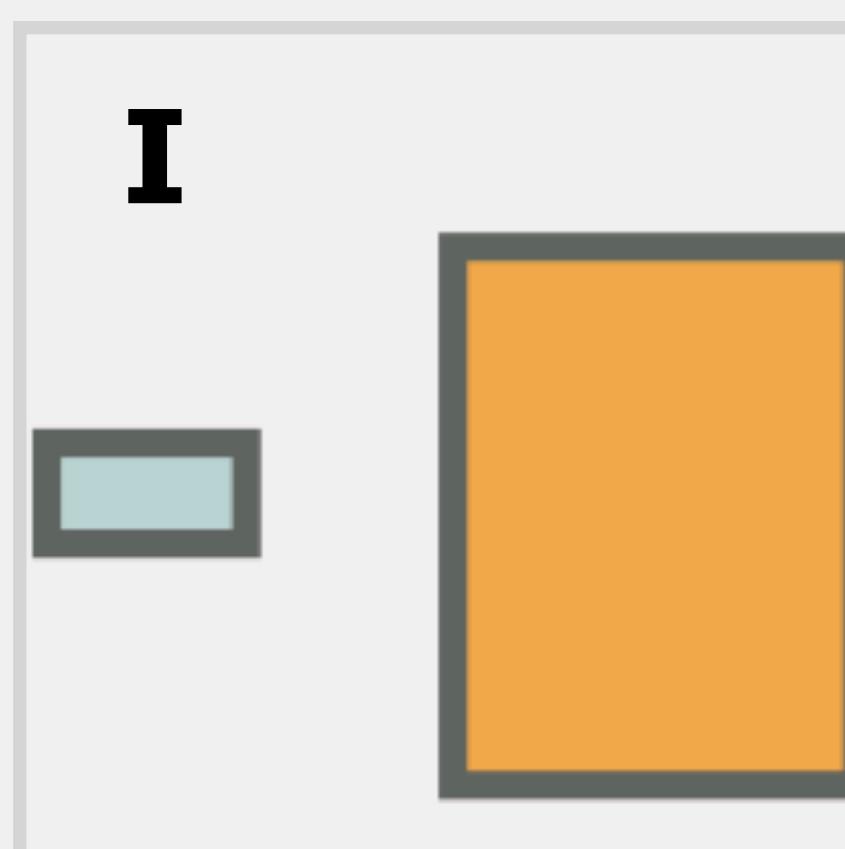
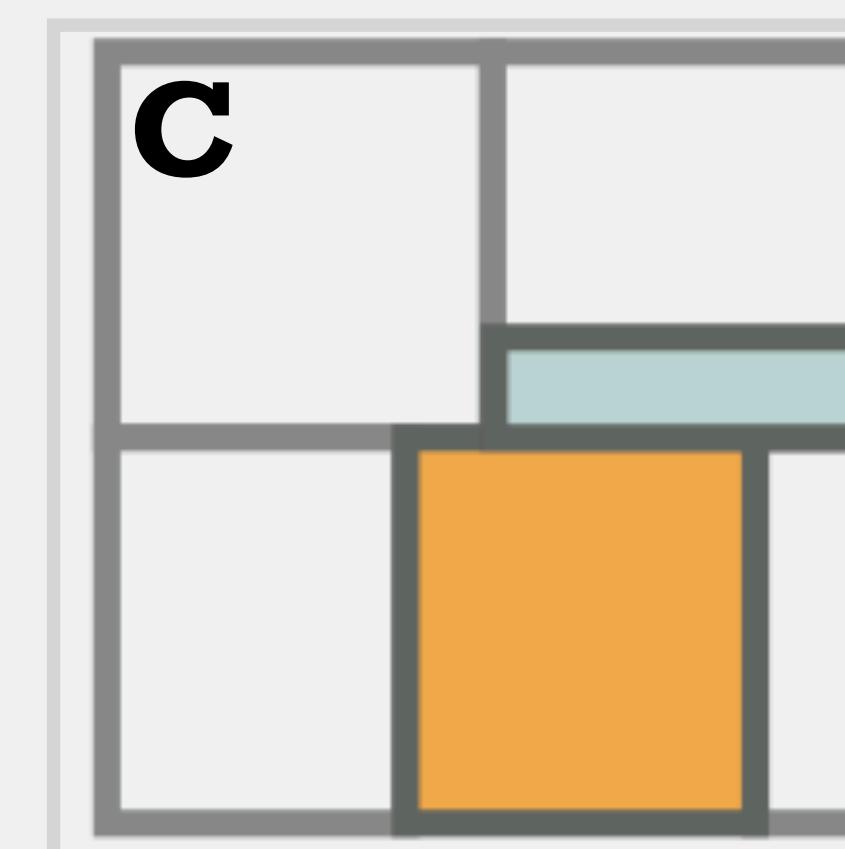
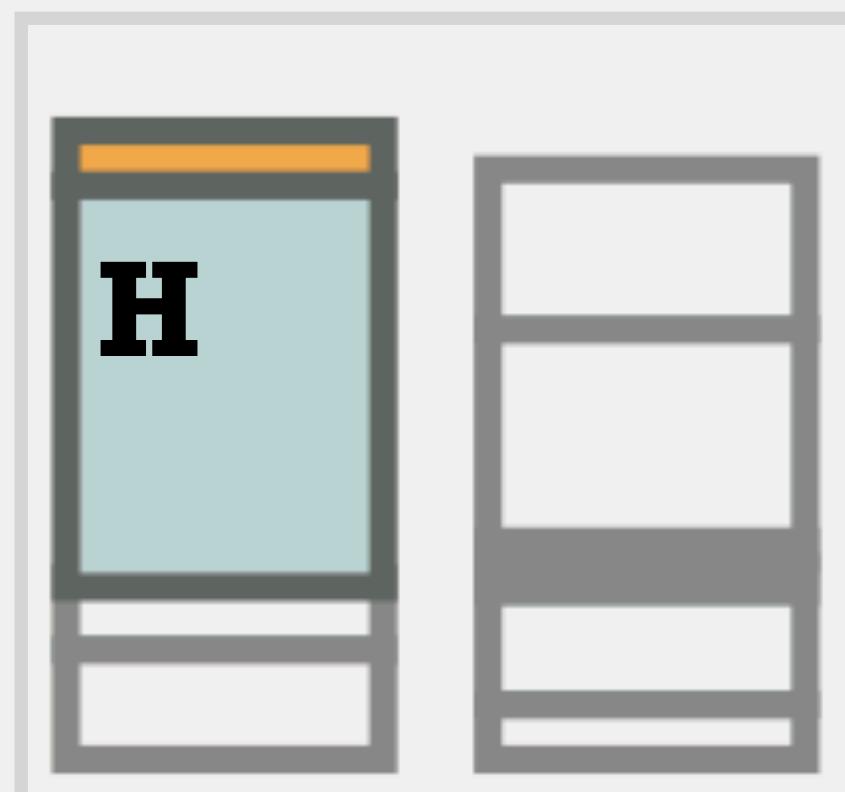
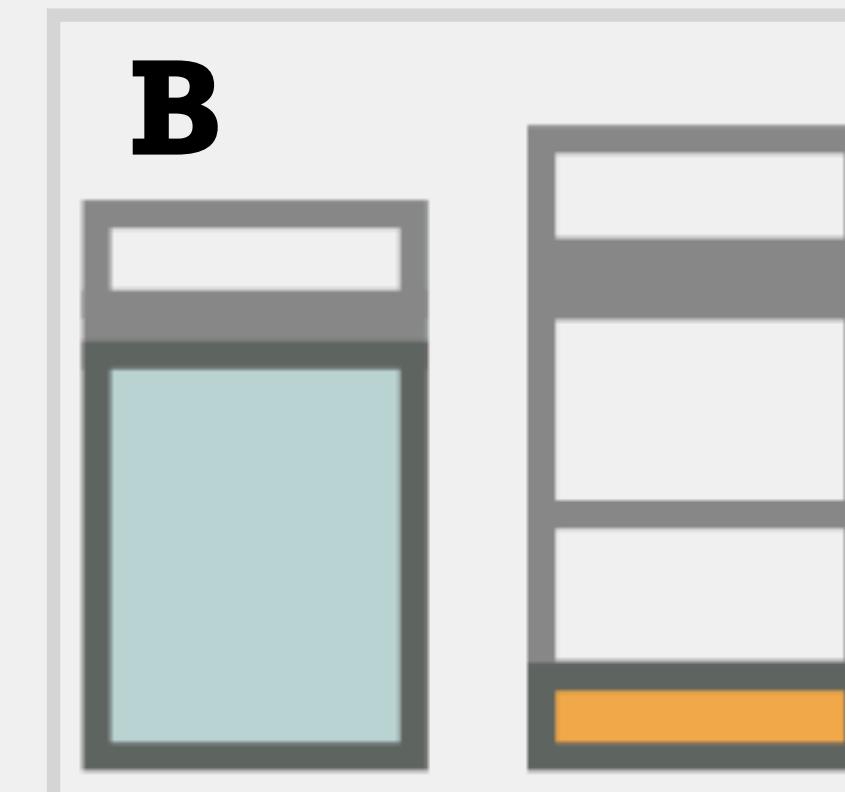
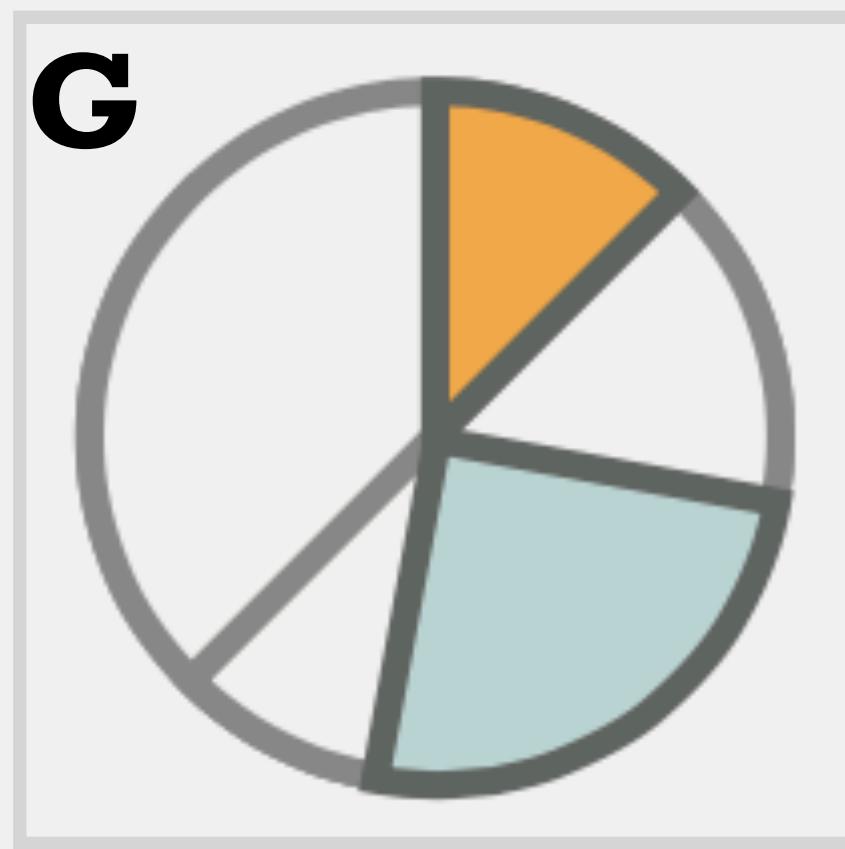
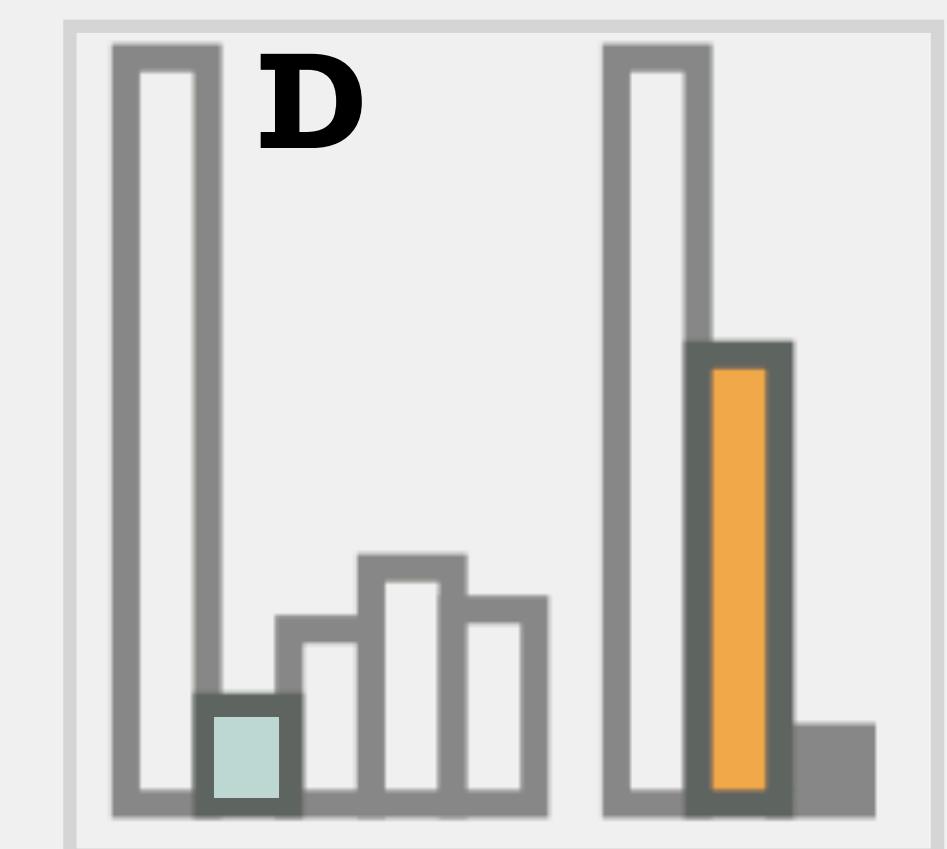
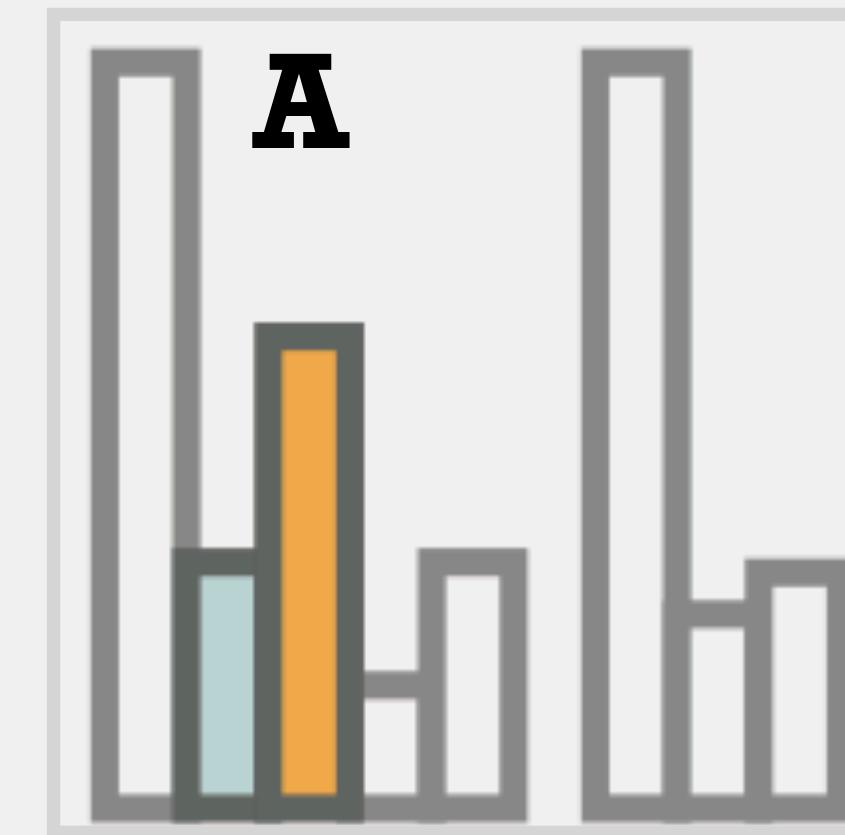
Examples

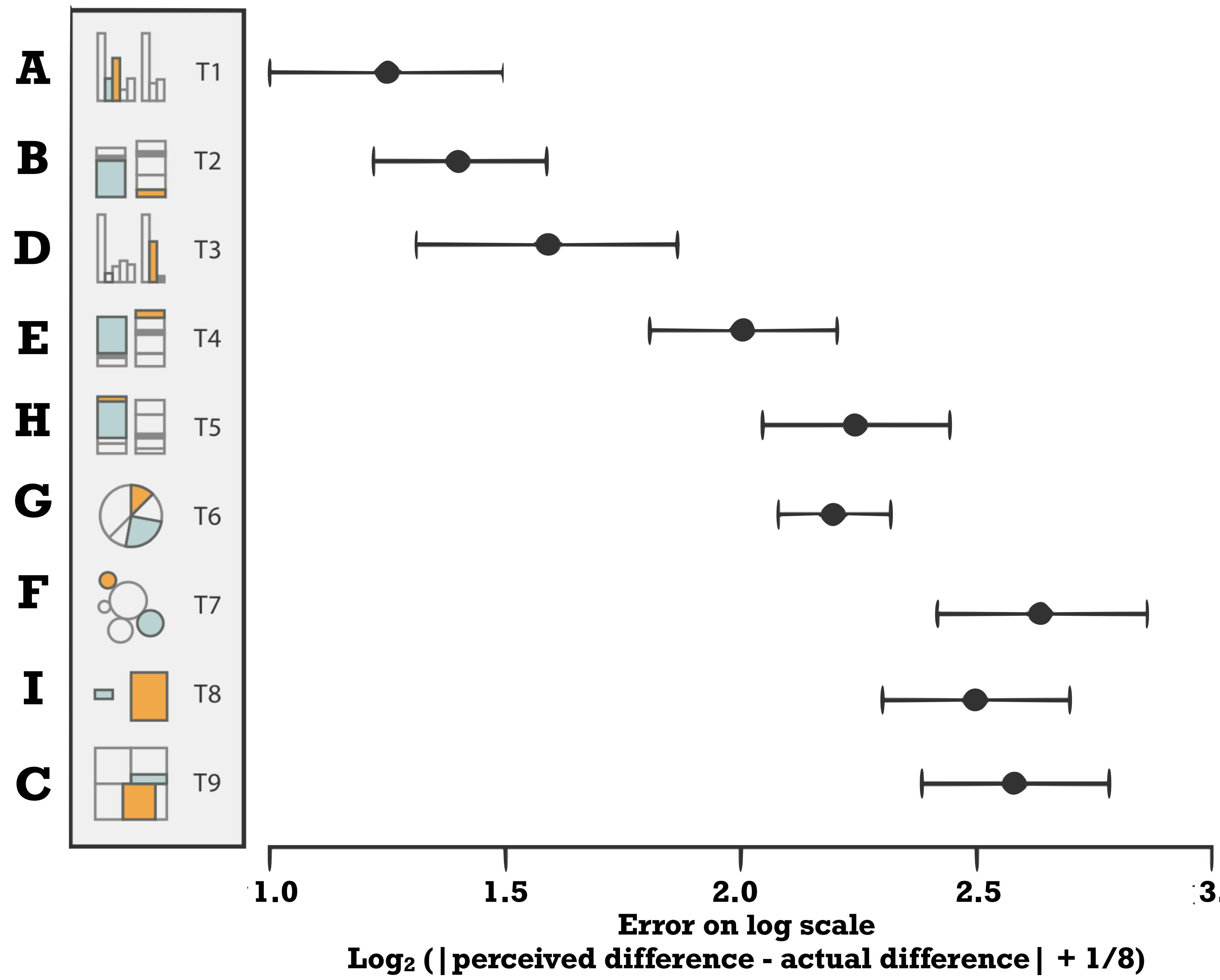
- Summary statistics >> do not show distributions
- Pick the simplest plot that best shows most/all of the data needed to answer the research question
- Tailor the visualization to your audience (within reason)

Select a Plot

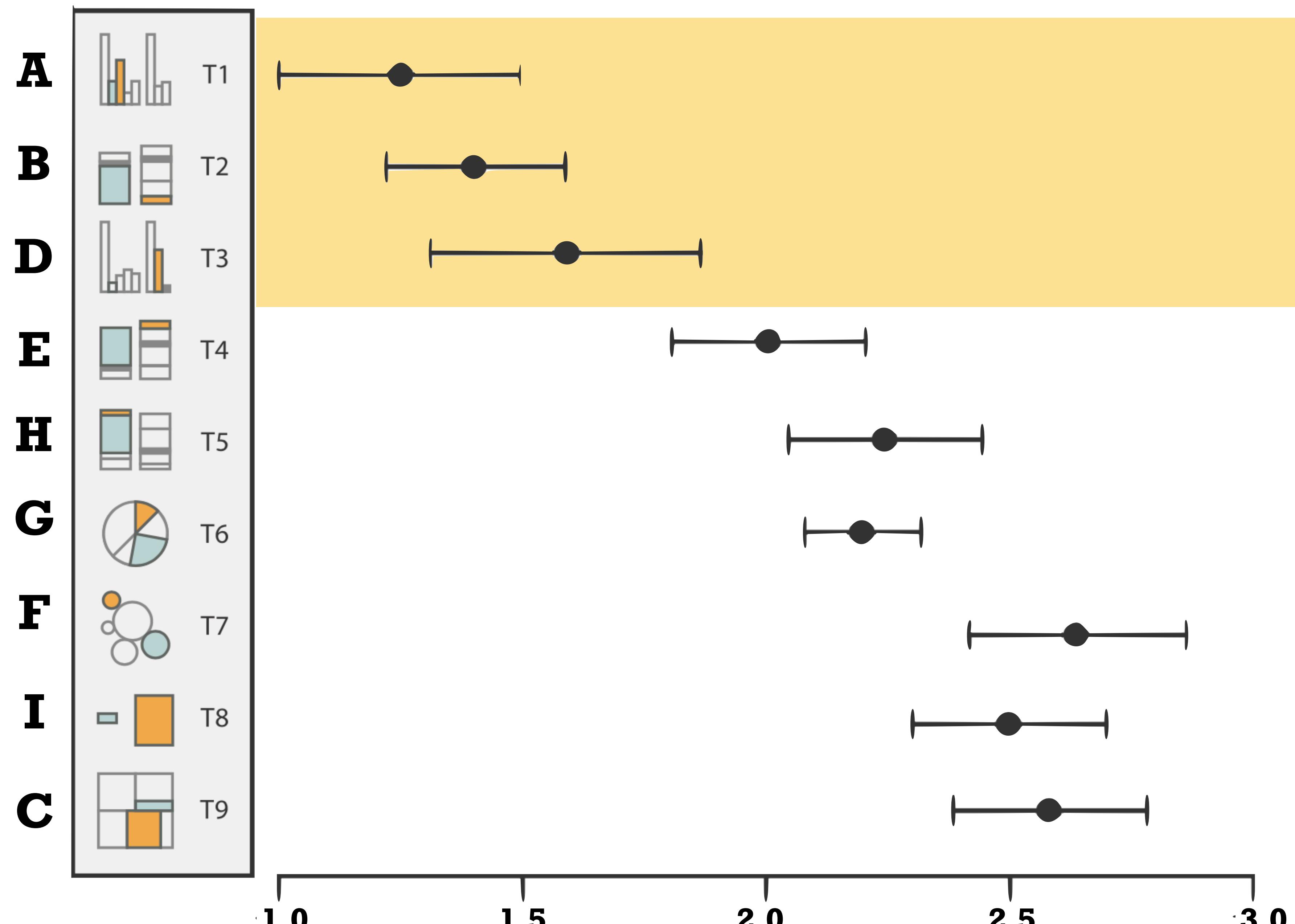
Use Zoom Stamps to
select the “best” 
and “worst” 
plot that:

Quantifies the
difference between
orange and green
regions.





Sources: [Heer & Bostock](#) and
[Tamara Munzner's textbook](#)



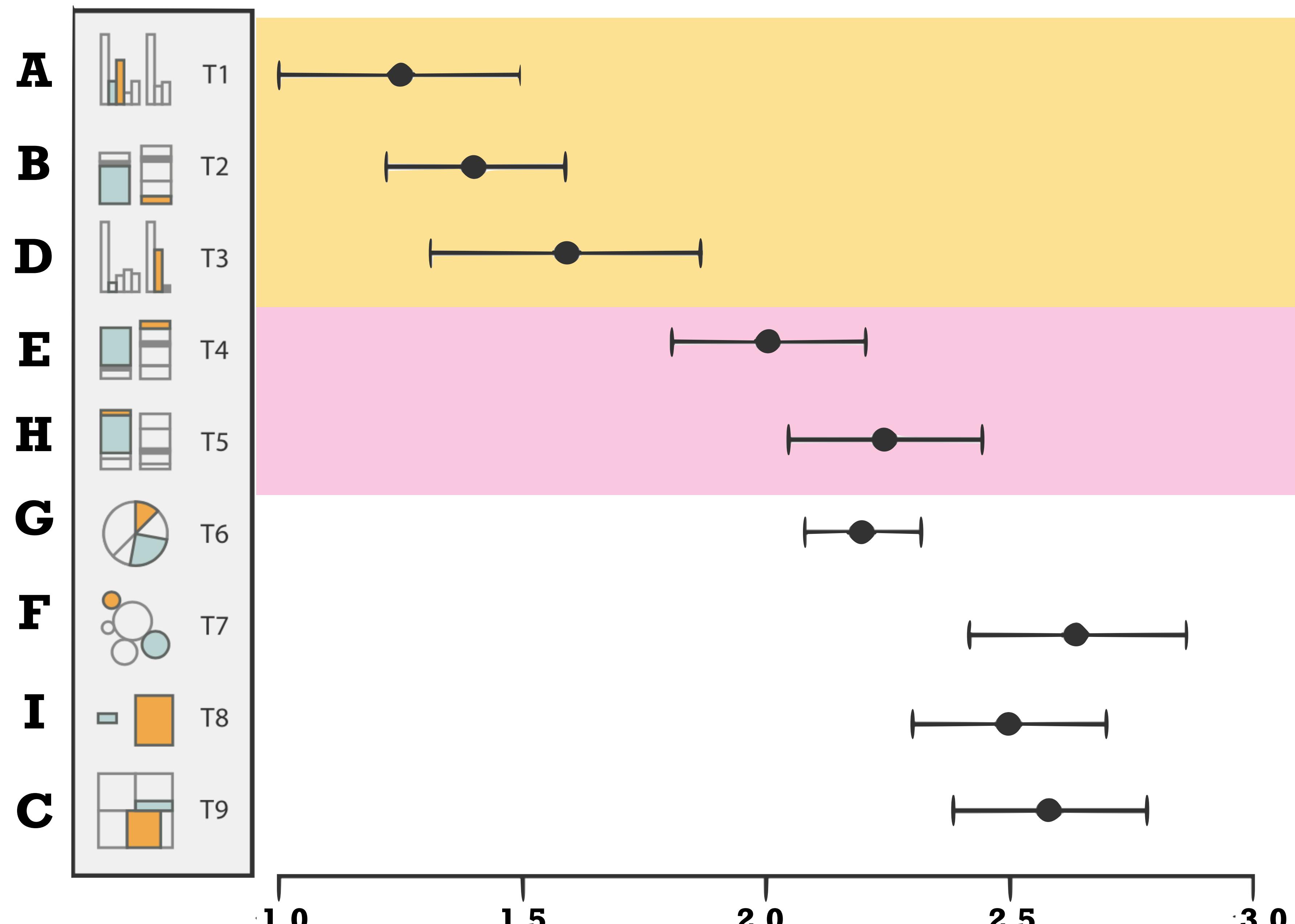
**Position on a scale
(y-axis)**

**Length/Height
comparison**

Angles/curves

Areas

Sources: [Heer & Bostock](#) and
[Tamara Munzner's textbook](#)



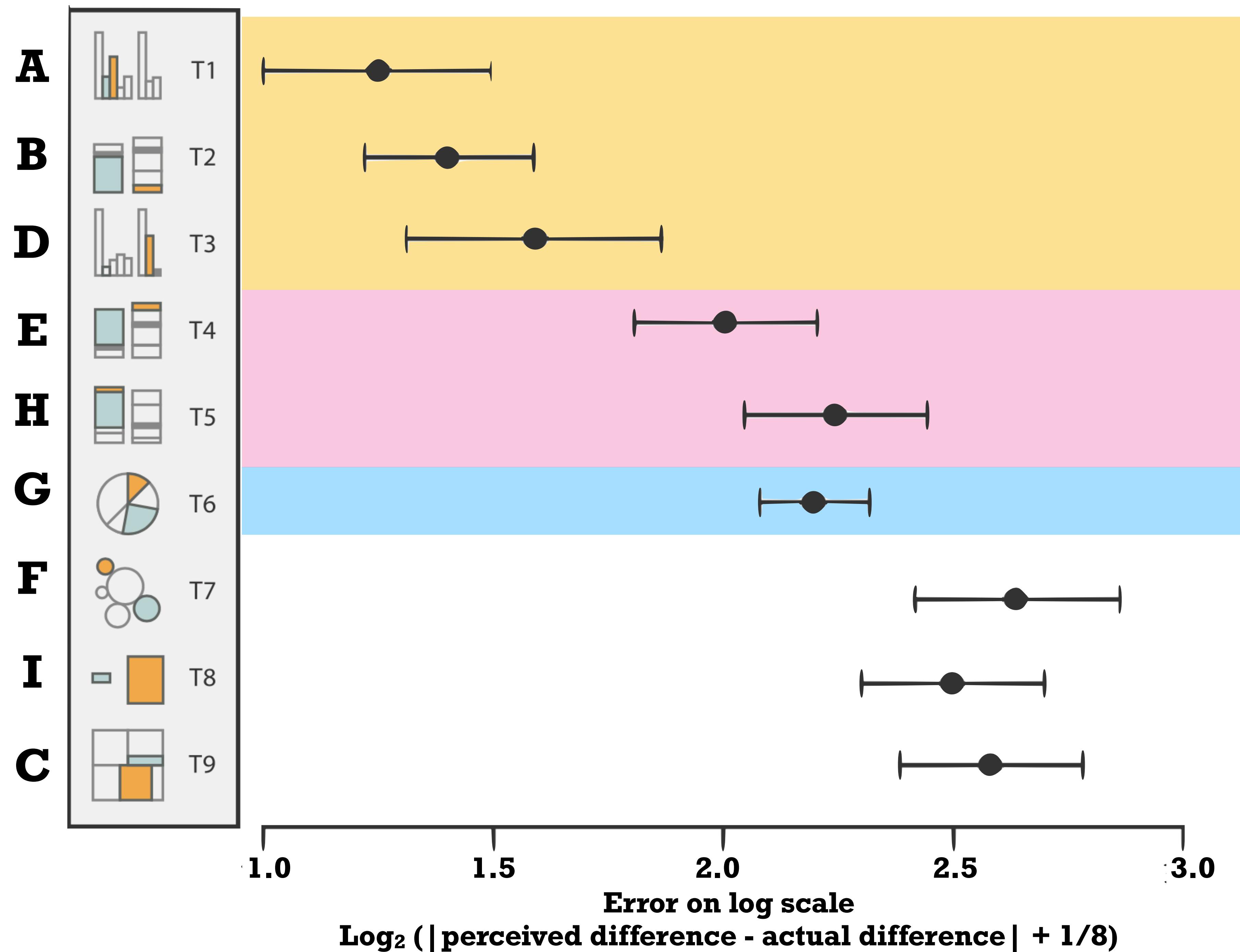
**Position on a scale
(y-axis)**

**Length/Height
comparison**

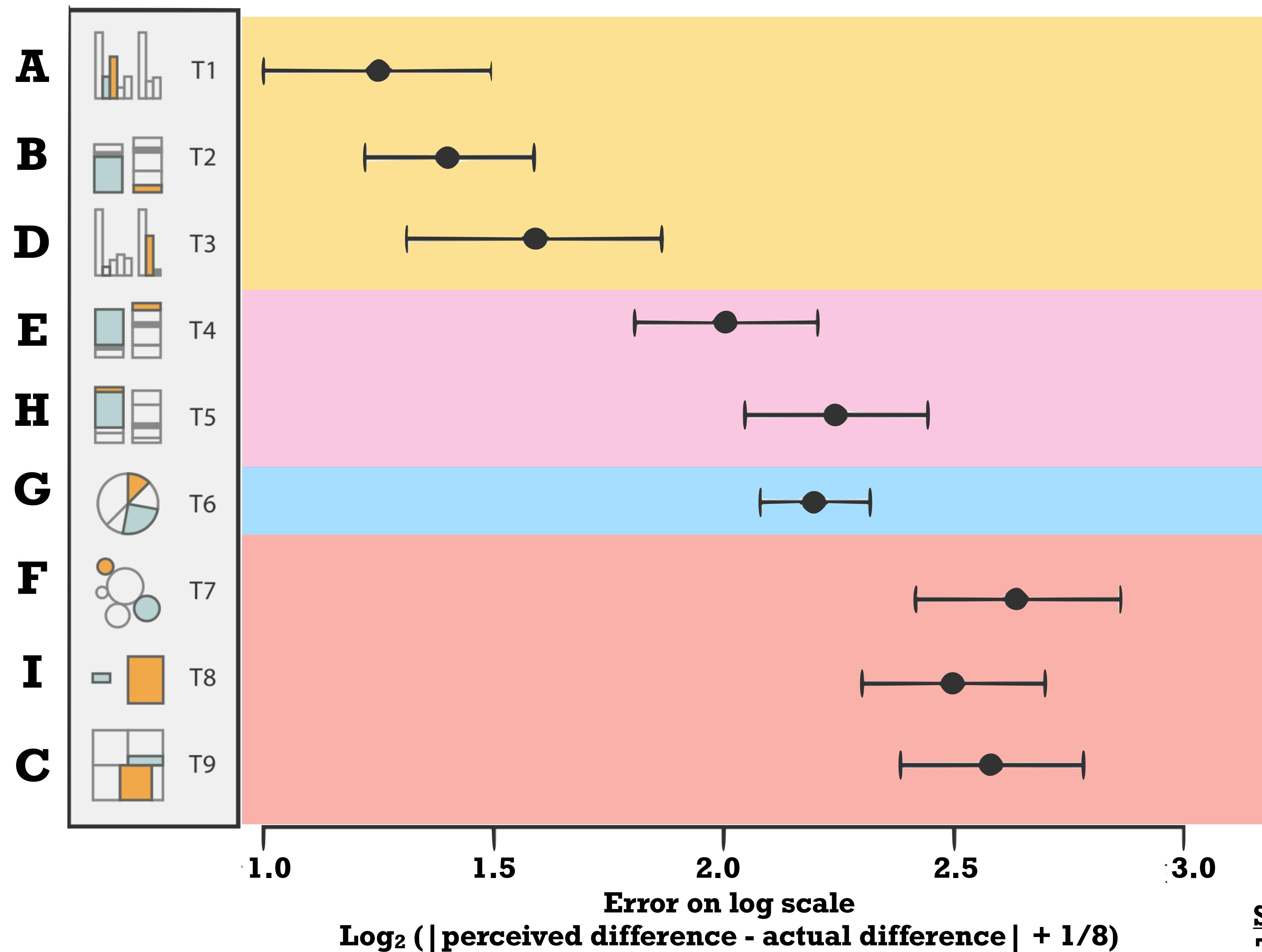
Angles/curves

Areas

Sources: [Heer & Bostock](#) and
[Tamara Munzner's textbook](#)



Sources: [Heer & Bostock](#) and
[Tamara Munzner's textbook](#)

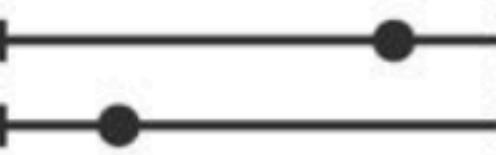


Sources: [Heer & Bostock](#) and
[Tamara Munzner's textbook](#)

Channels: Expressiveness Types and Effectiveness Ranks

→ **Magnitude** Channels: Ordered Attributes

Position on common scale



Effectiveness
↑ Most

Position on unaligned scale

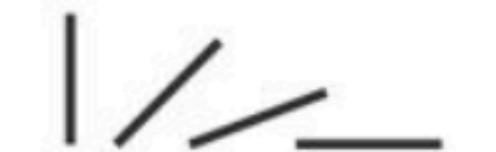


Length (1D size)



Effectiveness
↑ Most

Tilt/angle



Effectiveness
↑ Most

Area (2D size)



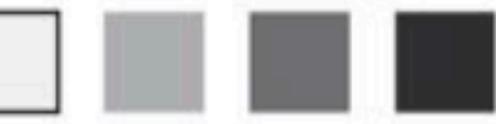
Effectiveness
↑ Most

Depth (3D position)



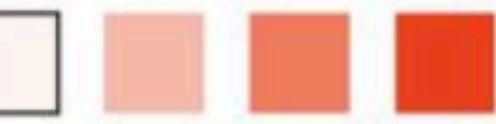
Effectiveness
↑ Most

Color luminance



Effectiveness
↑ Most

Color saturation



Effectiveness
↑ Most

Curvature



Effectiveness
↑ Most

Volume (3D size)



Sources: Chapter 5 of Tamara Munzner's textbook

If you're interested...

Sources: Chapter 5 of Tamara Munzner's textbook and originally:
“On the Psychophysical Law.”,
Psychological Review 64:3 (1957)

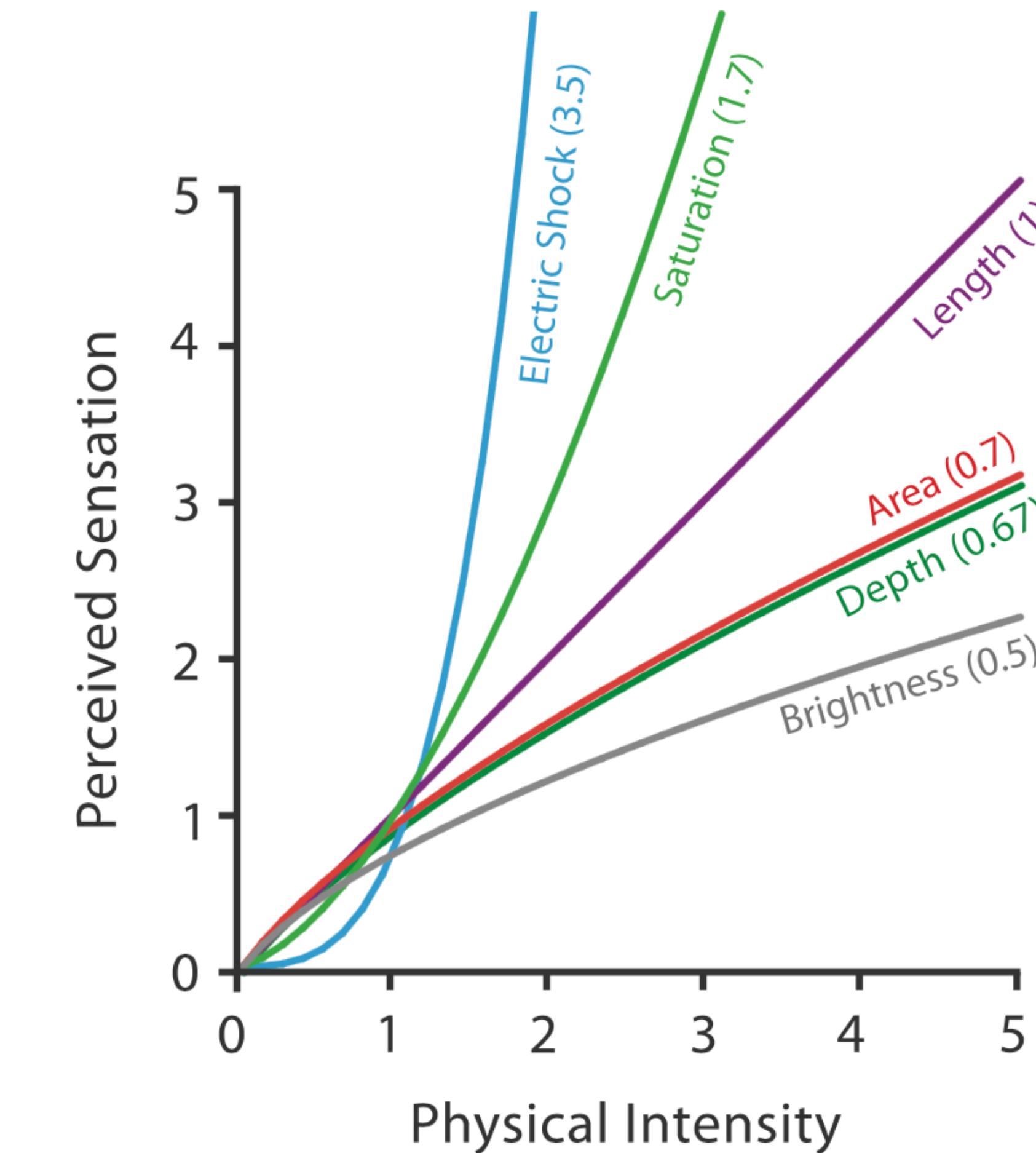
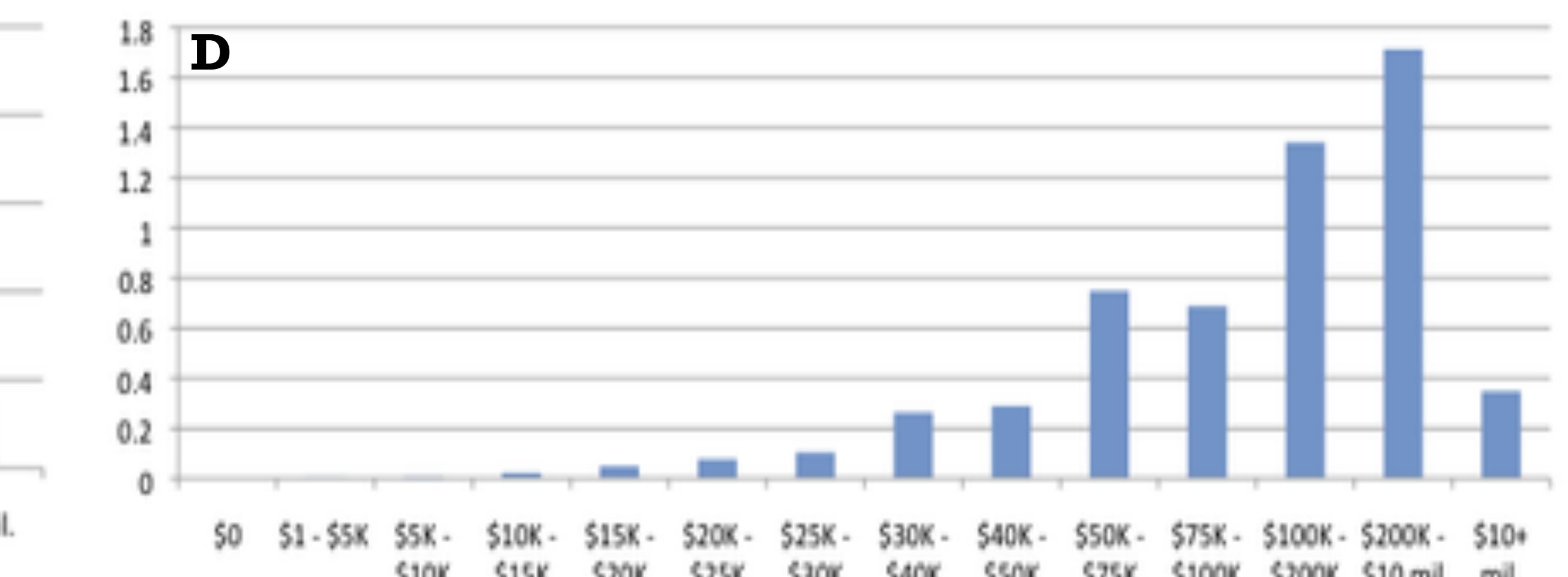
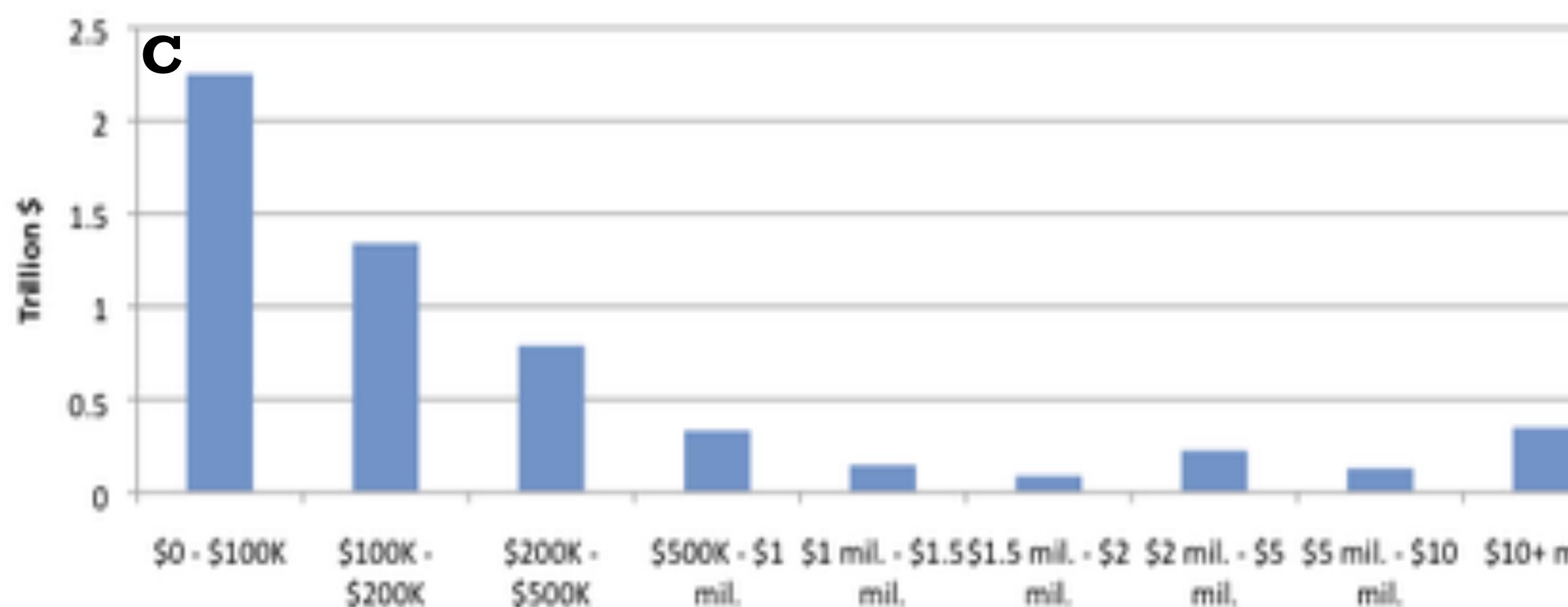
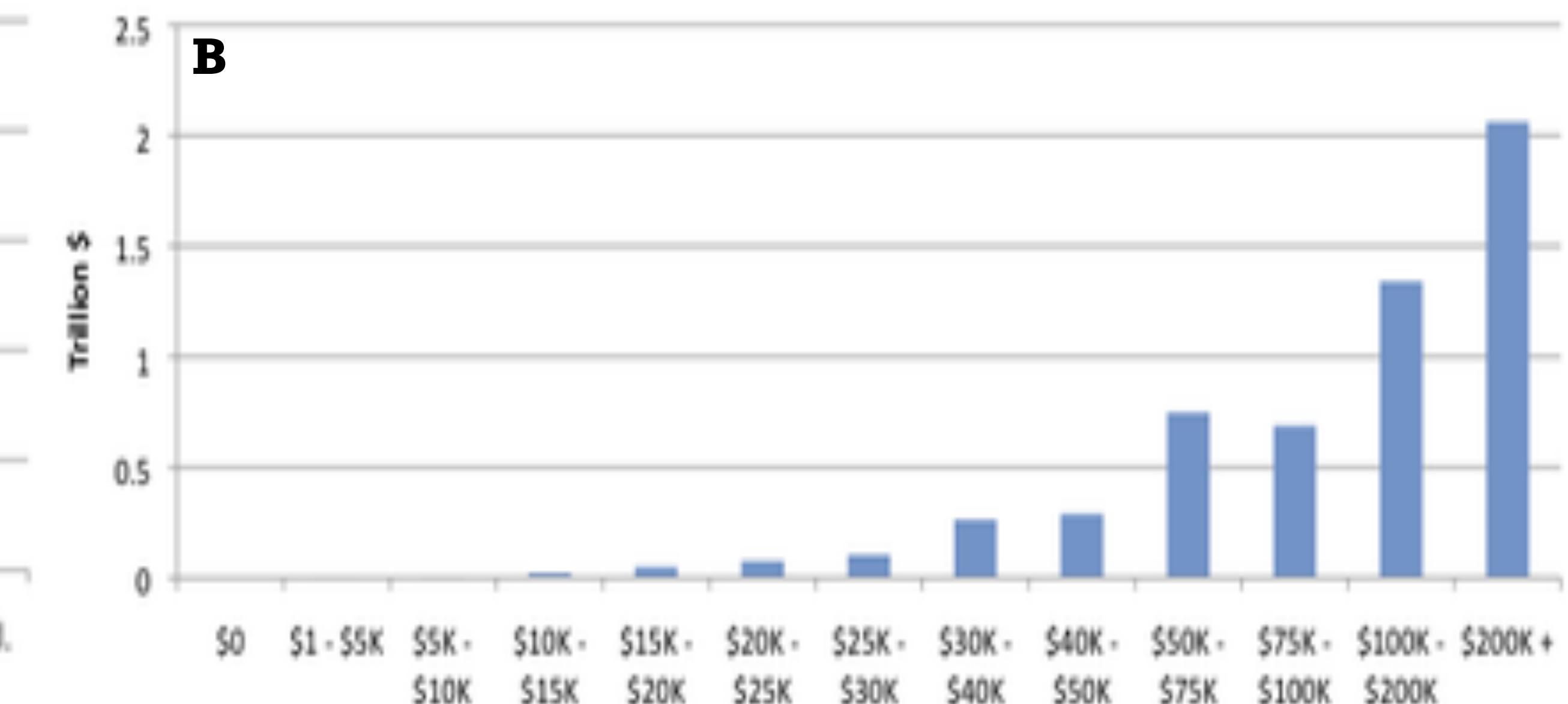
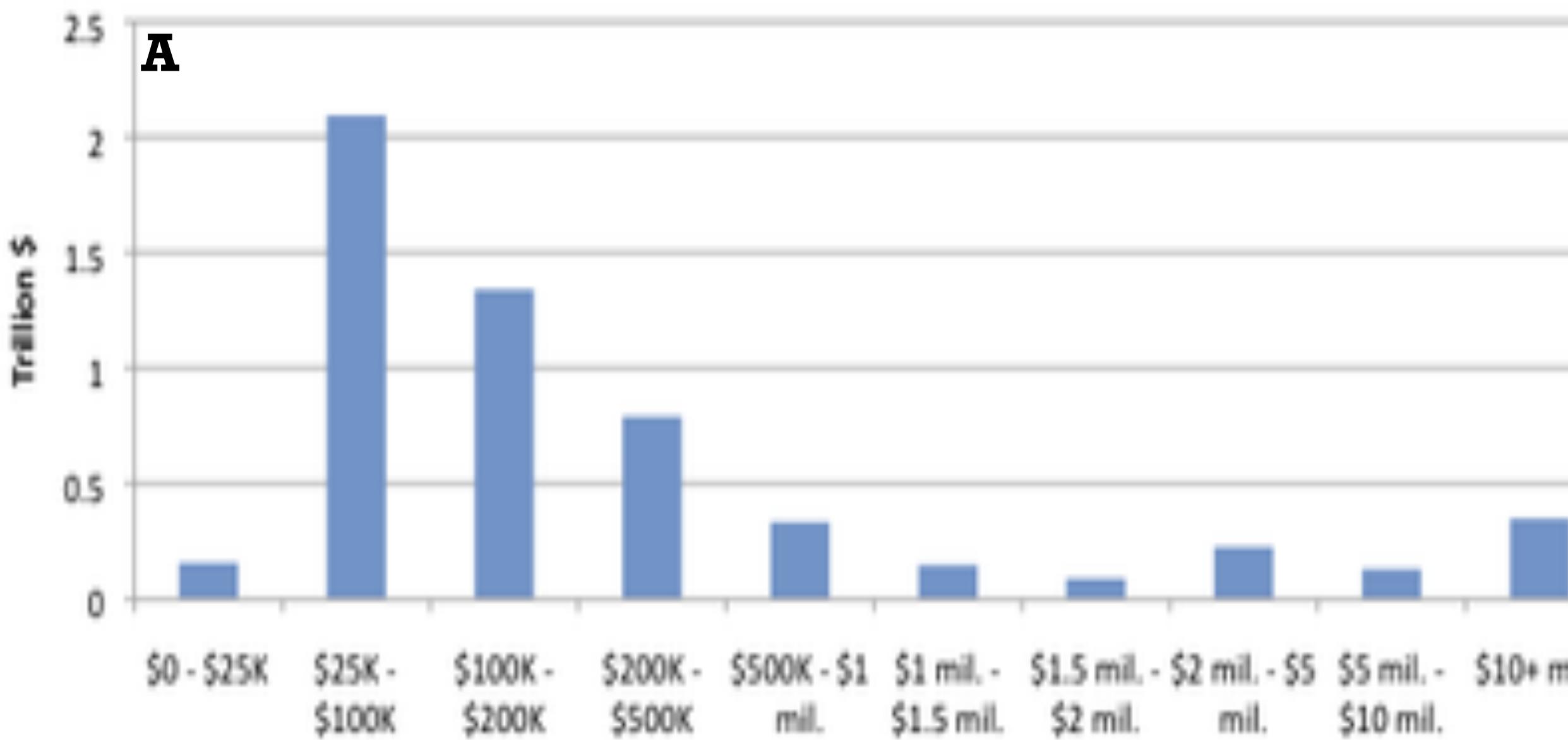


Figure 5.7. Stevens showed that the apparent magnitude of all sensory channels follows a power law $S = I^n$, where some sensations are perceptually magnified compared with their objective intensity (when $n > 1$) and some compressed (when $n < 1$). Length perception is completely accurate, whereas area is compressed and saturation is magnified. Data from Stevens [Stevens 75, p. 15].

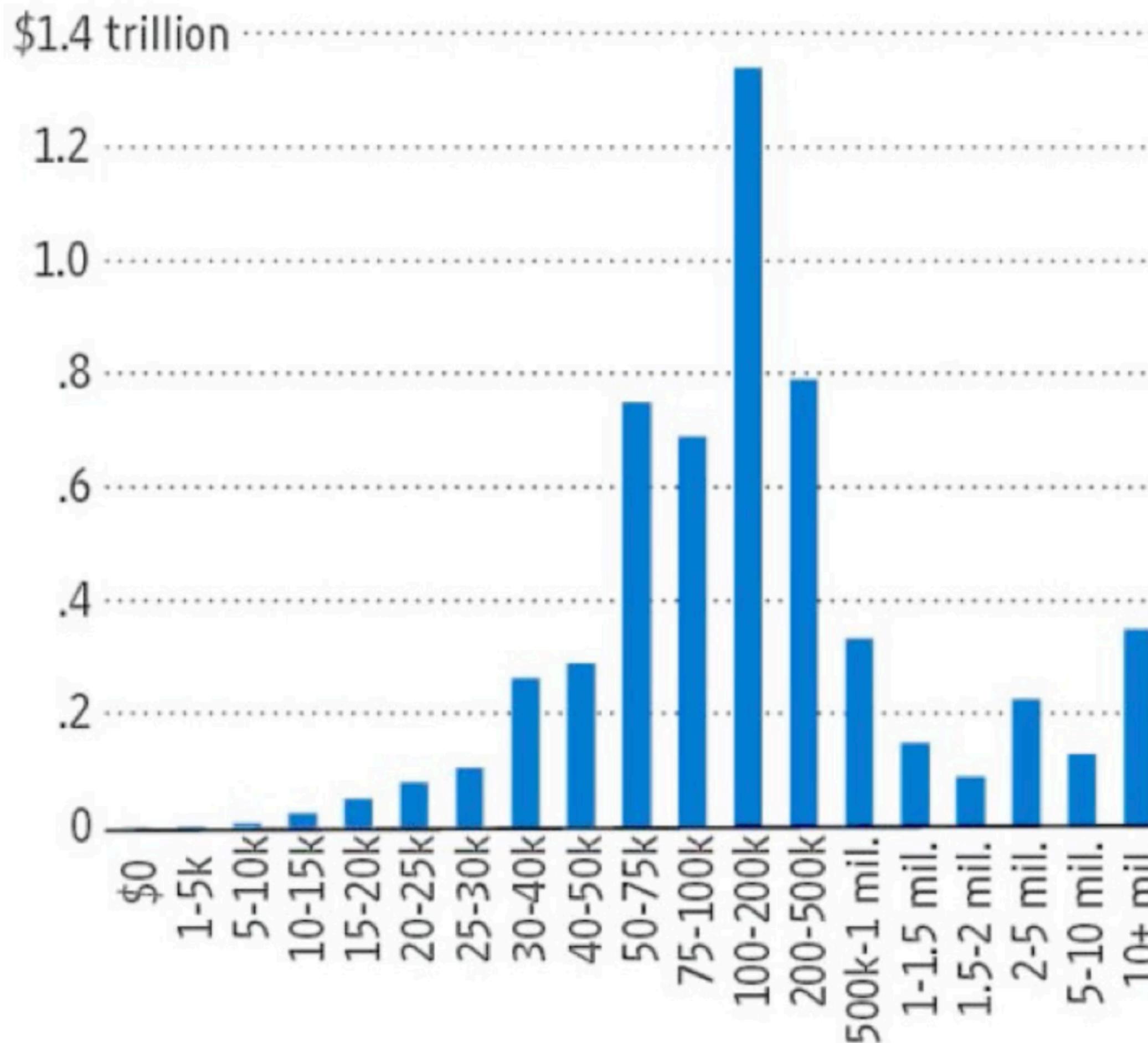
Which of these four plots do you think best represents the “true” wealth distribution in the United States today?



D. None of the above

The Middle Class Tax Target

The amount of total taxable income (left scale) for all filers by adjusted gross income level for 2008



Source: IRS

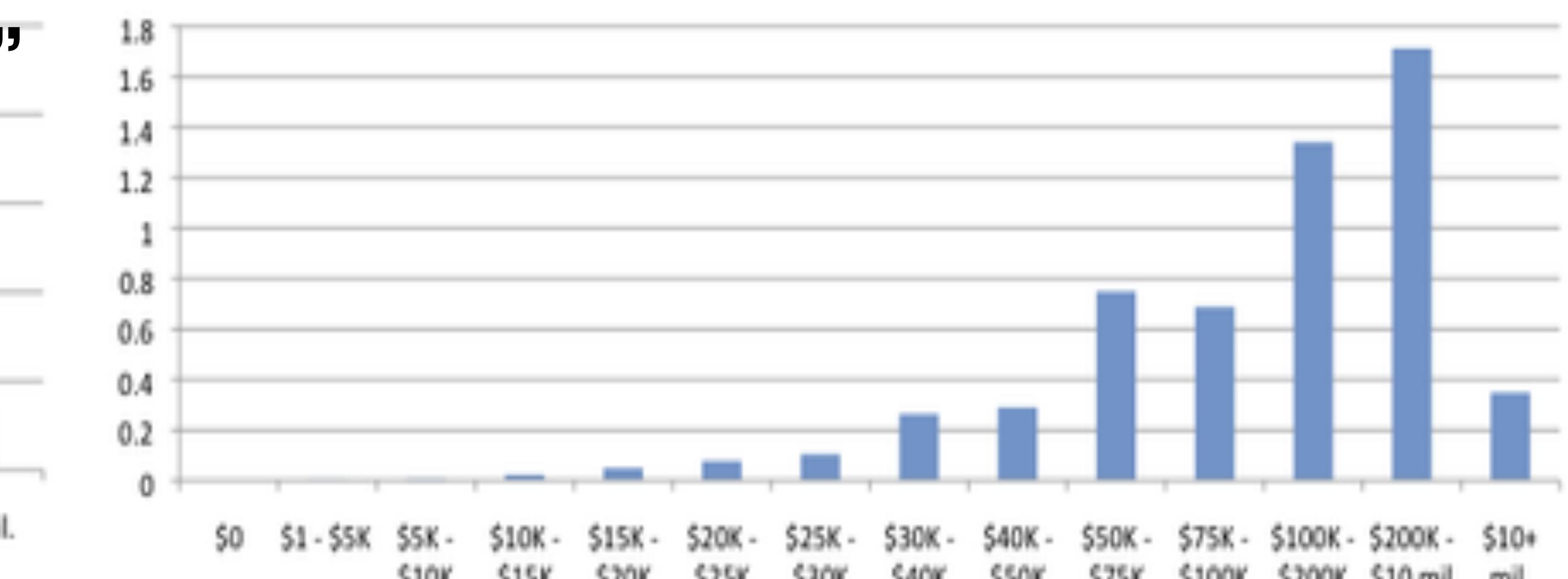
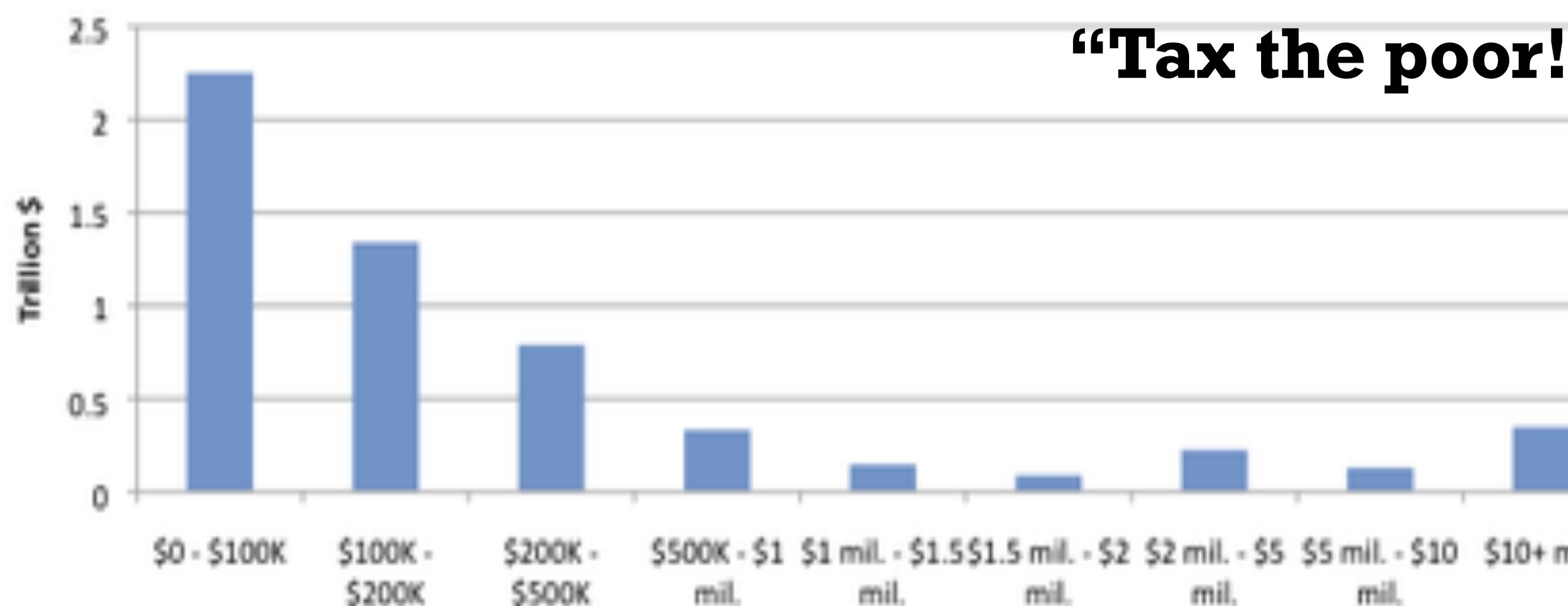
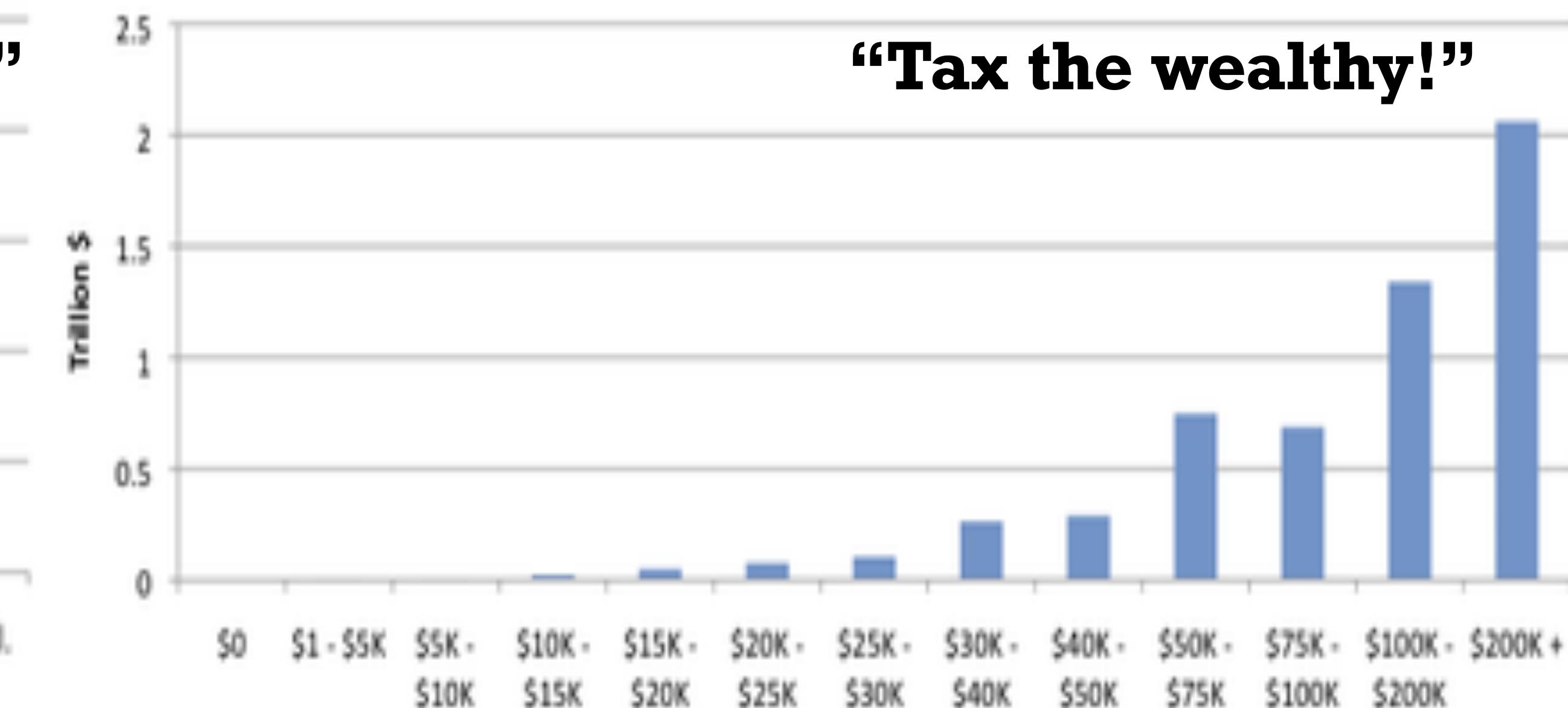
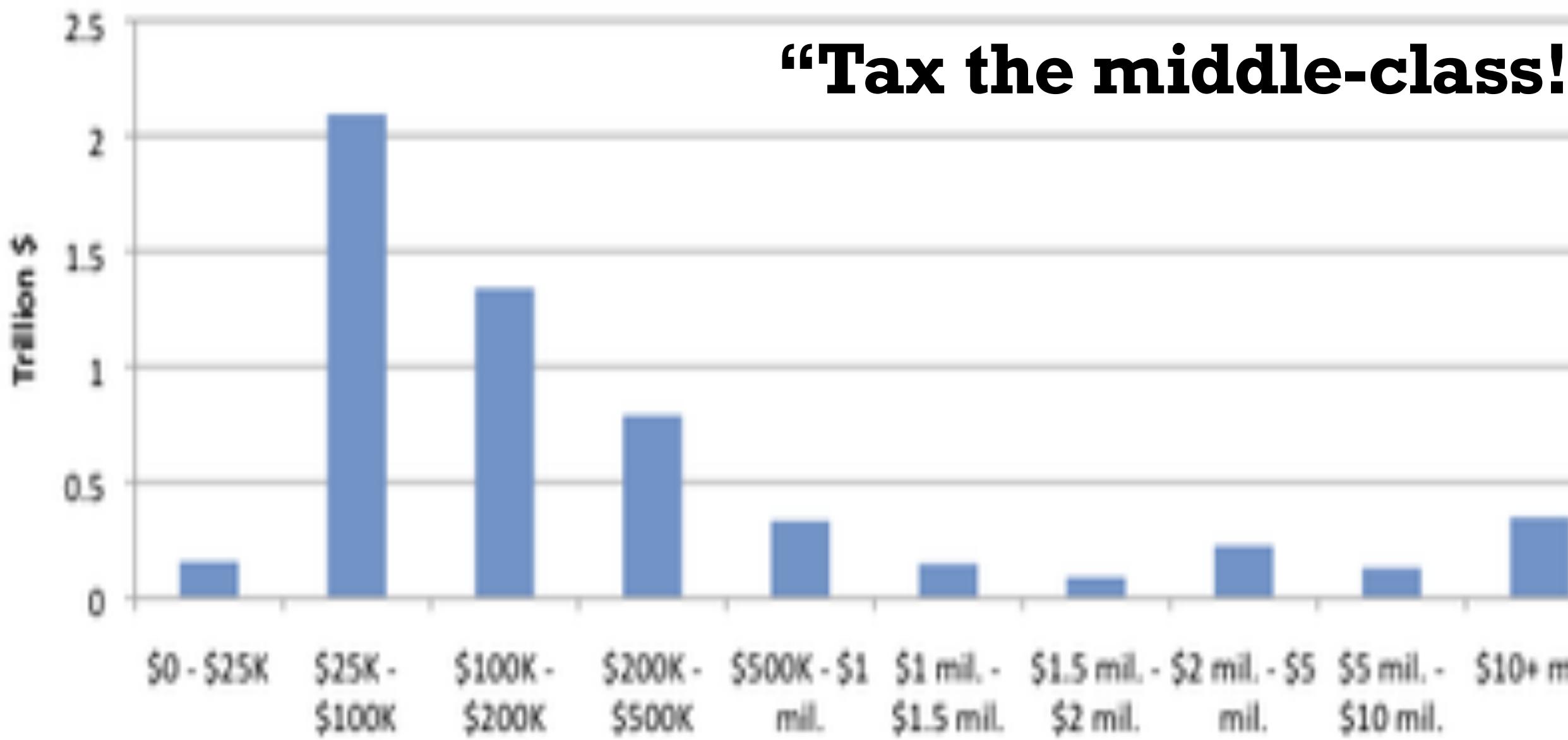
"The rich, in short, aren't nearly rich enough to finance Mr. Obama's entitlement state ambitions—even before his health-care plan kicks in."

So who else is there to tax? Well, in 2008, there was about \$5.65 trillion in total taxable income from all individual taxpayers, and most of that came from middle income earners. The nearby chart shows the distribution, and the big hump in the center is where Democrats are inevitably headed for the same reason that Willie Sutton robbed banks."

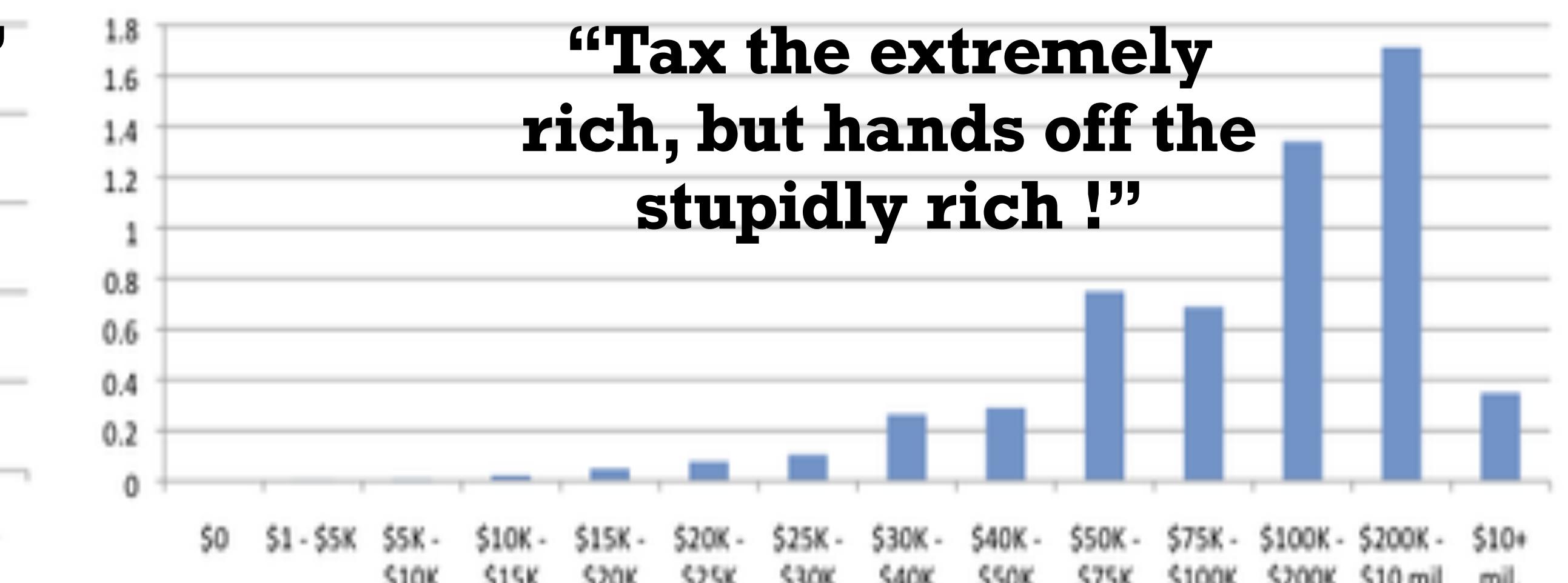
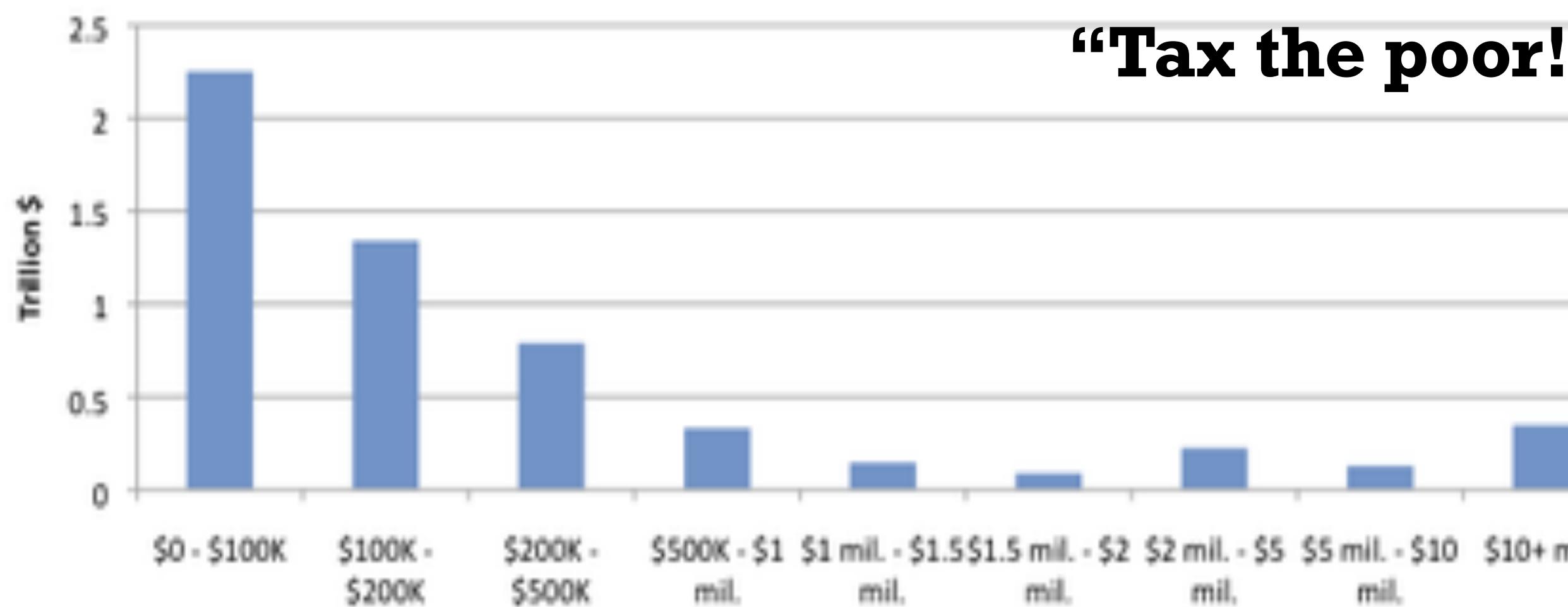
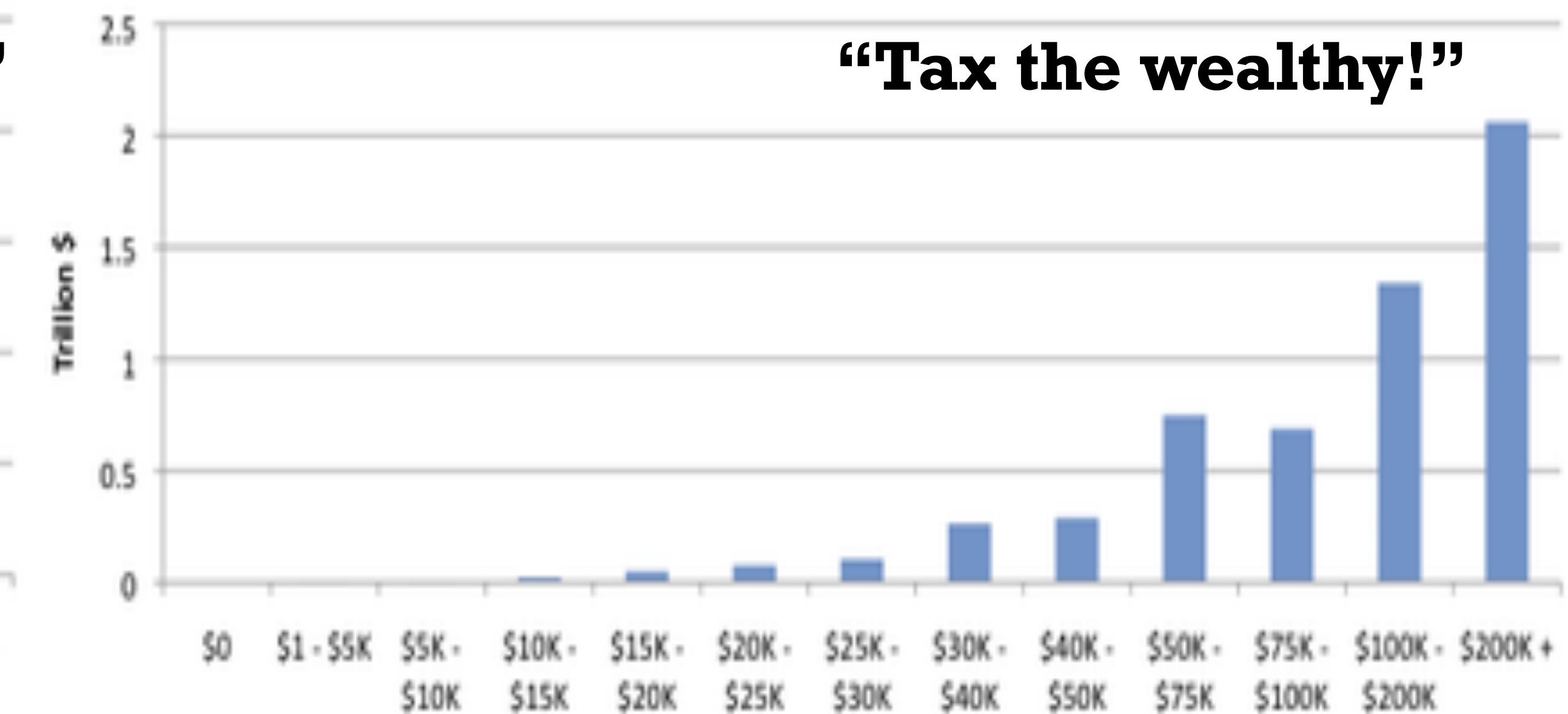
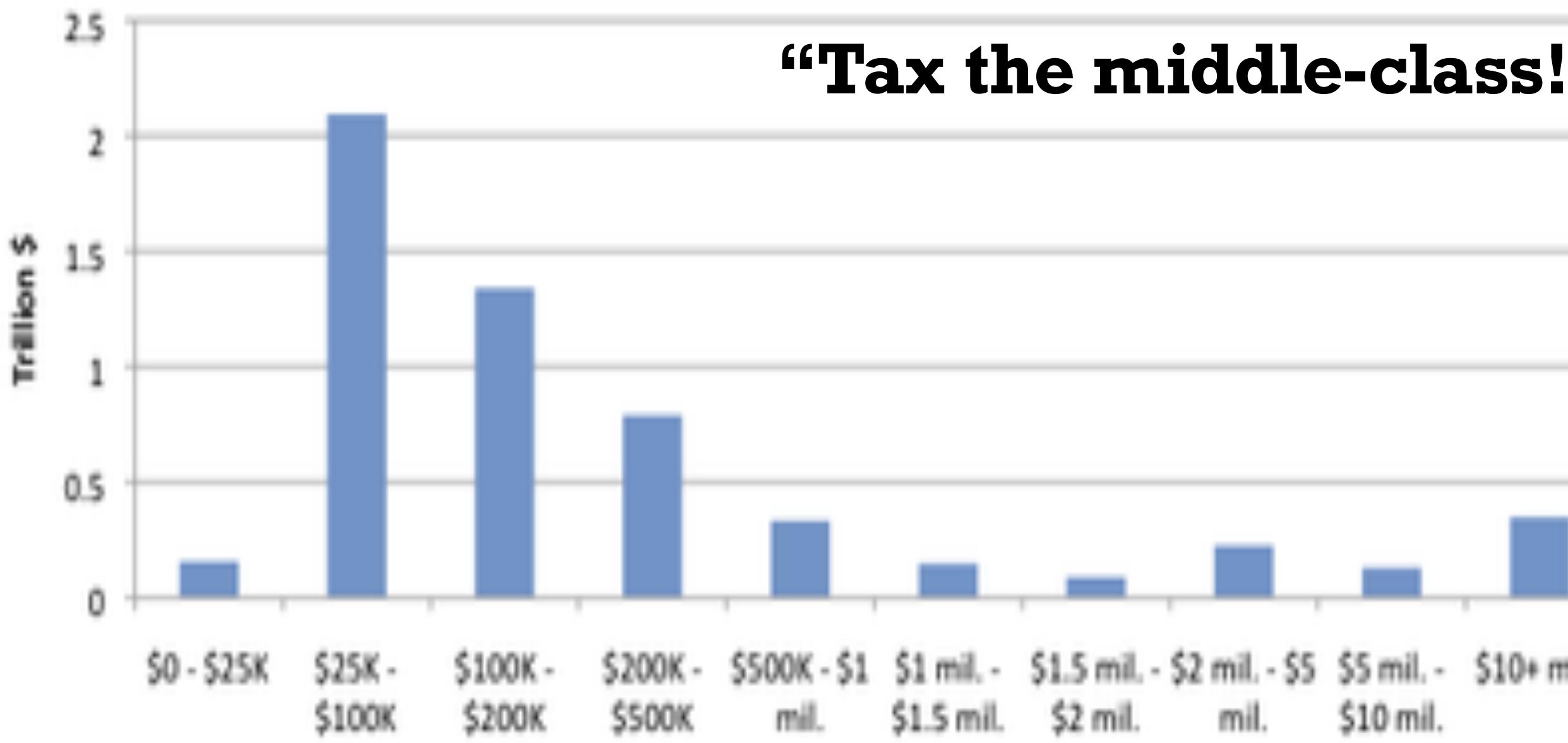
-The Wall Street Journal
April 17, 2011

Sources:
callingbull.org and
[this blog](#)

Which of these four plots do you think best represents the “true” wealth distribution in the United States today?



Which of these four plots do you think best represents the “true” wealth distribution in the United States today?



10 minutes

Part 3:

Judicious use of Colours

Colour

- Use of colours and shading is essential in most (if not all) visualizations
- However, **appropriate** colours and schemes must be used to retain plot effectiveness
- Accessibility: colour vision deficiency affects 1 in 12 men and 1 in 200 women¹

Colour blindness

Drag and drop or paste your file in the area below or: [Choose File](#) no file selected

Trichromatic view: Anomalous Trichromacy: *Dichromatic view:* *Monochromatic view:*

Normal Red-Weak/Protanomaly Red-Blind/Protanopia Monochromacy/Achromatopsia
 Green-Weak/Deuteranomaly Green-Blind/Deuteranopia Blue Cone Monochromacy
 Blue-Weak/Tritanomaly Blue-Blind/Tritanopia

Use lens to compare with normal view: No Lens Normal Lens Inverse Lens

[Reset View](#)



Zoom, move and lens functionality only with your own images available.

Check figures through an online [colour blindness simulator](#)

Colour blindness

Drag and drop or paste your file in the area below or: [Choose File](#) no file selected

Trichromatic view: Anomalous Trichromacy: *Dichromatic view:* *Monochromatic view:*

Normal Red-Weak/Protanomaly Red-Blind/Protanopia Monochromacy/Achromatopsia
 Green-Weak/Deuteranomaly Green-Blind/Deuteranopia Blue Cone Monochromacy
 Blue-Weak/Tritanomaly Blue-Blind/Tritanopia

Use lens to compare with normal view: No Lens Normal Lens Inverse Lens

[Reset View](#)



Zoom, move and lens functionality only with your own images available.

Drag and drop or paste your file in the area below or: [Choose File](#) no file selected

Trichromatic view: Anomalous Trichromacy: *Dichromatic view:* *Monochromatic view:*

Normal Red-Weak/Protanomaly Red-Blind/Protanopia Monochromacy/Achromatopsia
 Green-Weak/Deuteranomaly Green-Blind/Deuteranopia Blue Cone Monochromacy
 Blue-Weak/Tritanomaly Blue-Blind/Tritanopia

Use lens to compare with normal view: No Lens Normal Lens Inverse Lens

[Reset View](#)



Run your figures through an online [colour blindness simulator](#)

4.1 Color as a tool to distinguish

We frequently use color as a means to distinguish discrete items or groups that do not have an intrinsic order, such as different countries on a map or different manufacturers of a certain product. In this case, we use a *qualitative* color scale. Such a scale contains a finite set of specific colors that are chosen to look clearly distinct from each other while also being equivalent to each other. The second condition requires that no one color should stand out relative to the others. And, the colors should not create the impression of an order, as would be the case with a sequence of colors that get successively lighter. Such colors would create an apparent order among the items being colored, which by definition have no order.

Many appropriate qualitative color scales are readily available. Figure 4.1 shows three representative examples. In particular, the ColorBrewer project provides a nice selection of qualitative color scales, including both fairly light and fairly dark colors (Brewer 2017).

Okabe Ito



ColorBrewer Dark2



ggplot2 hue

