Introduction to Data Visualization

How to refine our plots? "themes"

Halil Bisgin, Ph.D.

Changing appearances w/ themes

- Themes can be turned on or off using the theme_set() function
  - **theme_set**(**theme_bw**())
  - **theme_set**(**theme_dark**())
- Once set, a theme applies to all subsequent plots and it remains active until it is replaced by a different theme.
- You can still use the theme() function to fine-tune any aspect of your plot

More theme options

- ggplot comes with several built-in themes:
  - **theme_minimal()**
  - **theme_classic()**
  - **theme_gray()**
  - **theme_grey()** *as the default*

- Install the ggthemes for many more options
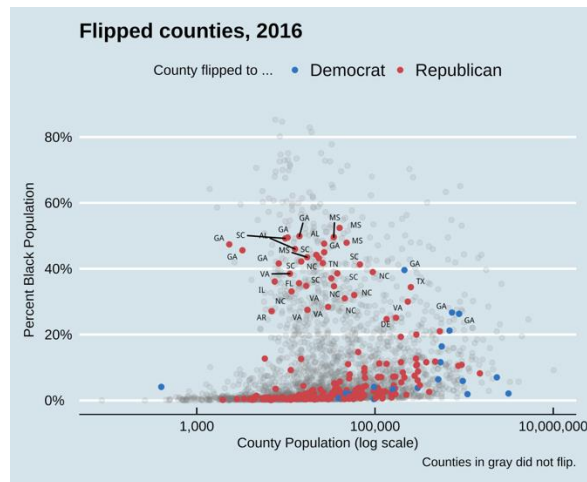
theme_set(theme_economist())
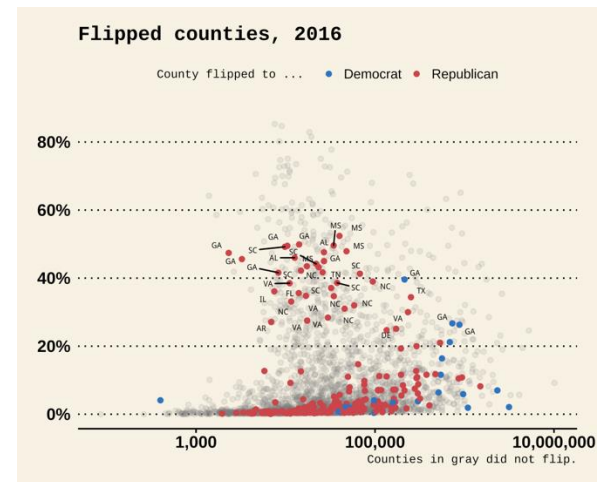
theme_set(theme_wsj())

Themes are like aesthetic

- The choices you make should harmonize with the broader printed or displayed material.

- When starting out it can be wisest to stick to the defaults or consistently use a suitable theme.

Which theme?

- Themes with colored backgrounds customized typefaces are best used
    - *when making one-off graphics or posters,*
    - *when preparing figures to integrate into a slide presentation, or*
    - *when conforming to a house or editorial style for publication*
- Journal publications need a different set of themes:
    - Claus O. Wilke's *cowplot* package
- Alternatively:
    - Bob Rudis's *hrbrthemes*
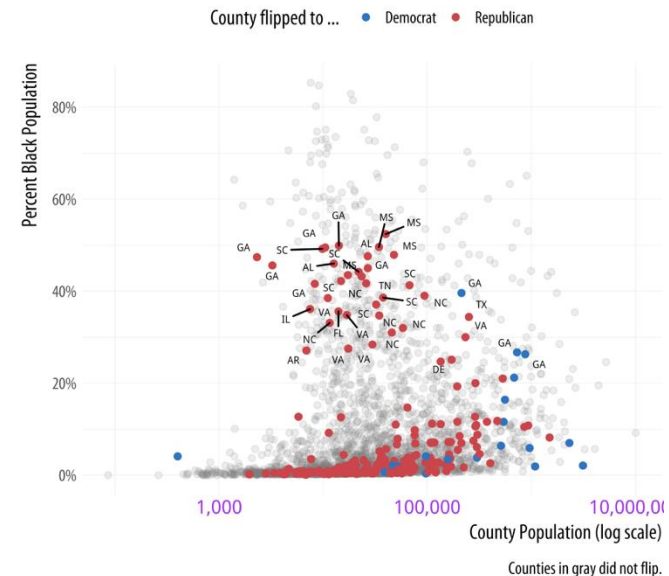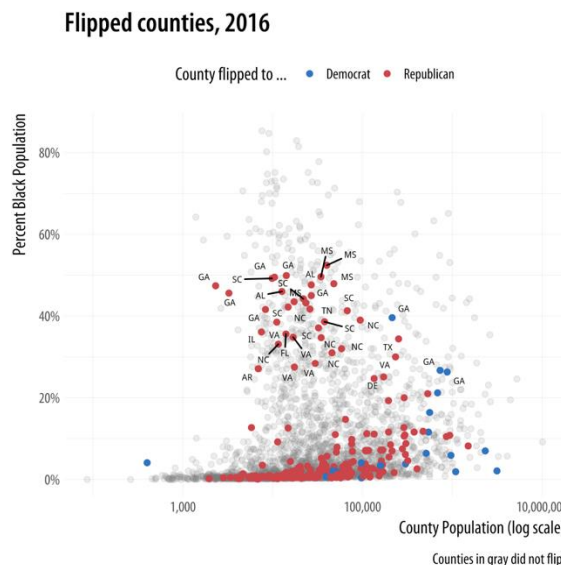
theme() function gives more options

- The theme() function allows you to change the color, typeface, and font weight of text.

```
... + theme(legend.position = "top",
plot.title = element_text(size=rel(2),                    lineheight=.5,
family="Times",
face="bold.italic",
colour="orange"),
axis.text.x = element_text(size=rel(1.1),              family="Courier",
face="bold",
color="purple"))
```

Flipped counties, 2016

Flipped counties, 2016

County flipped to …   ● Democrat   ● Republican

Percent Black Population

County Population (log scale)

Counties in gray did not flip.

County flipped to …   ● Democrat   ● Republican

Percent Black Population

County Population (log scale)

Counties in gray did not flip.

Using theme elements for design

- It makes good sense use themes as a way to fix design elements:
  - *you can subsequently ignore them, and*
  - *focus instead on the data you are examining.*
- It is also worth remembering that ggplot's theme system is very flexible.
  - *It permits a wide range of design elements to be adjusted in order to create custom figures.*
- element_text() and element_blank() are useful
  - *to tweak the appearance of various text elements such as titles.*
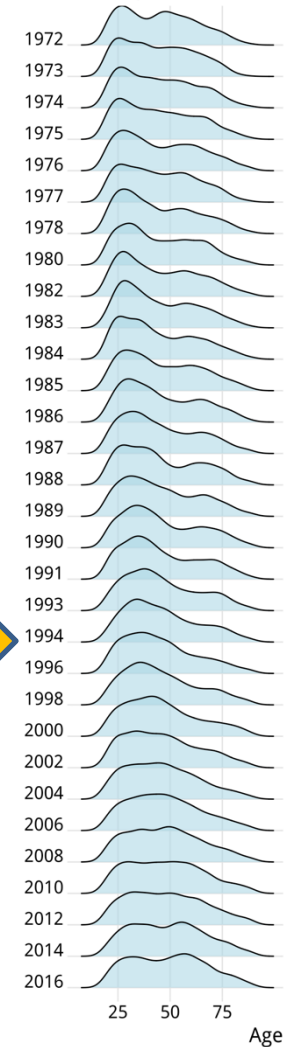 - *to remove several of them altogether.*

Manual vs. ggridges
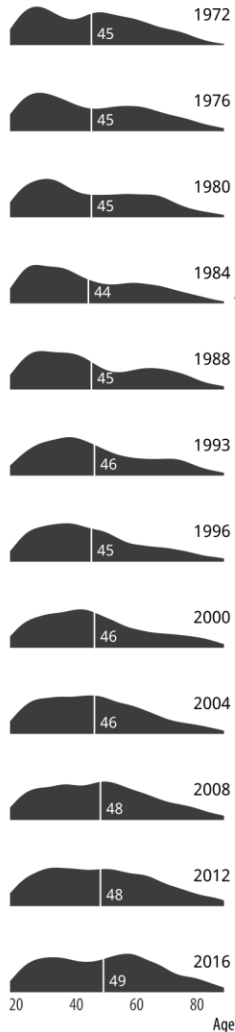
Removing some elements manually

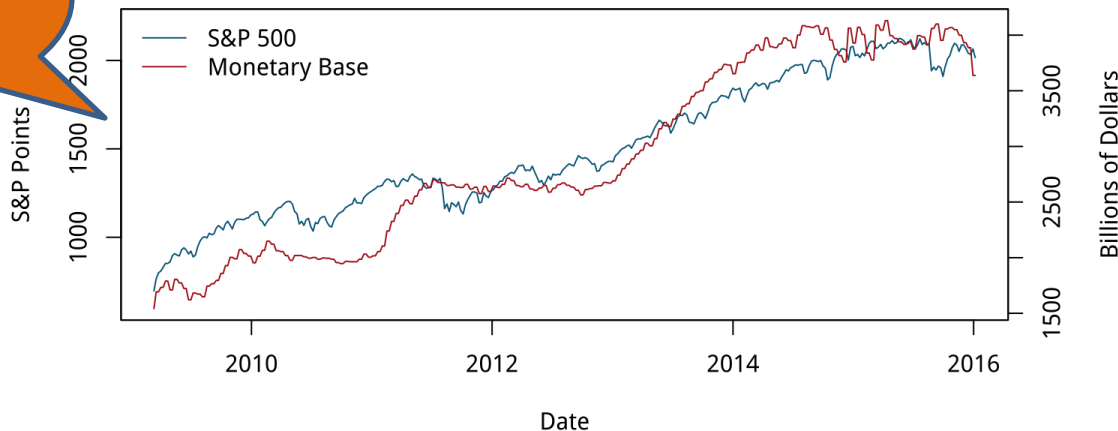Using ggridges

Case study 1-Two y-axes

- In January of 2016, Liz Ann Sonders, Chief Investment Strategist with Charles Schwab, Inc, tweeted about the apparent correlation between two economic time series:
  - *the Standard and Poor's 500 stock market index*
 - *the Monetary Base, a measure of the size of money supply*

Two y-axes are dangerous

- People tend to use two y-axes with a desire to have close lines and a belief of a substantive association.

- However, using two y-axes makes it even easier than usual to mislead about the degree of association.

- It allows to adjust the scaling of the axes to relative to one another in way that moves the data series around more or less however you like!

How were we fooled?

- For the first half of the graph, the red Monetary Base line tracks below the blue S&P 500 and is above it for the second half.
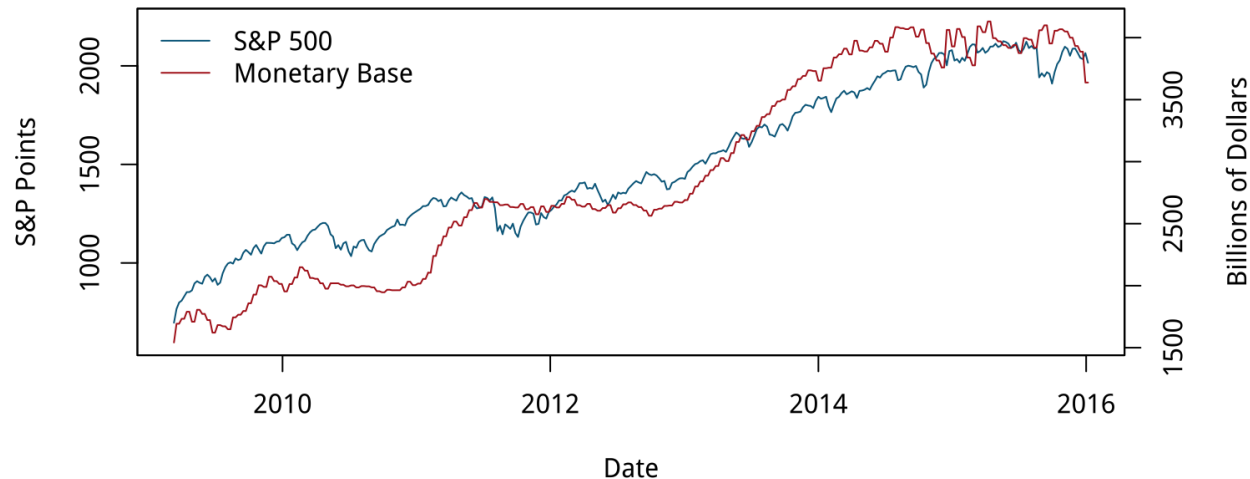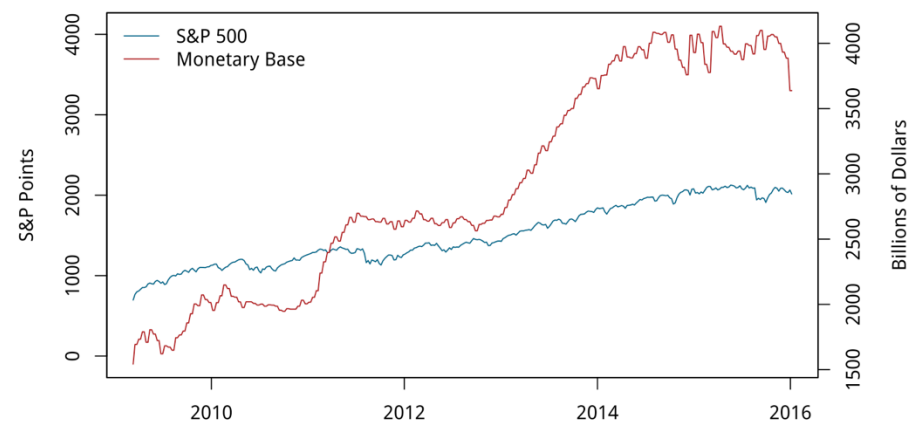
How to "fix" then?

- Start the second y-axis at zero, which shifts the Monetary Base line above the S&P line for the first half of the series and below it later on.

- Adjust the axes so that the axis tracking the S&P starts at zero.

- Muted the association between the two variables.
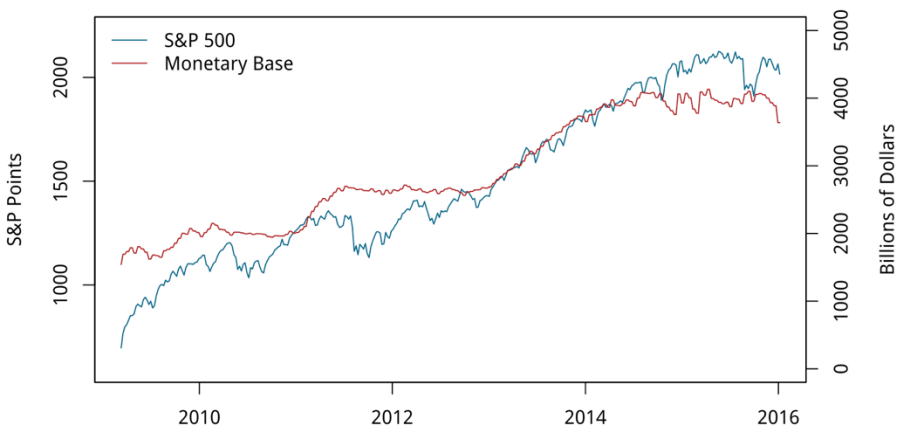
How else might we draw this data?

- We could use a split- or broken-axis plot to show the two series at the same time.
  - better perceptual properties than overlayed charts w/ dual axes
  - useful in cases where the series are of the same kind, but of very different magnitudes. (not the case here)
- If the series are not in the same units (or of widely differing magnitudes), is to rescale one of the series
  - *dividing or multiplying it by a thousand*
  - *to index each of them to 100 at the start of the first period, and then plot them both.*

Redrawing transformed data

- Two plots: i) two series in one, ii) their difference
- The S&P index runs above the Monetary Base for almost the whole series as opposed to the original
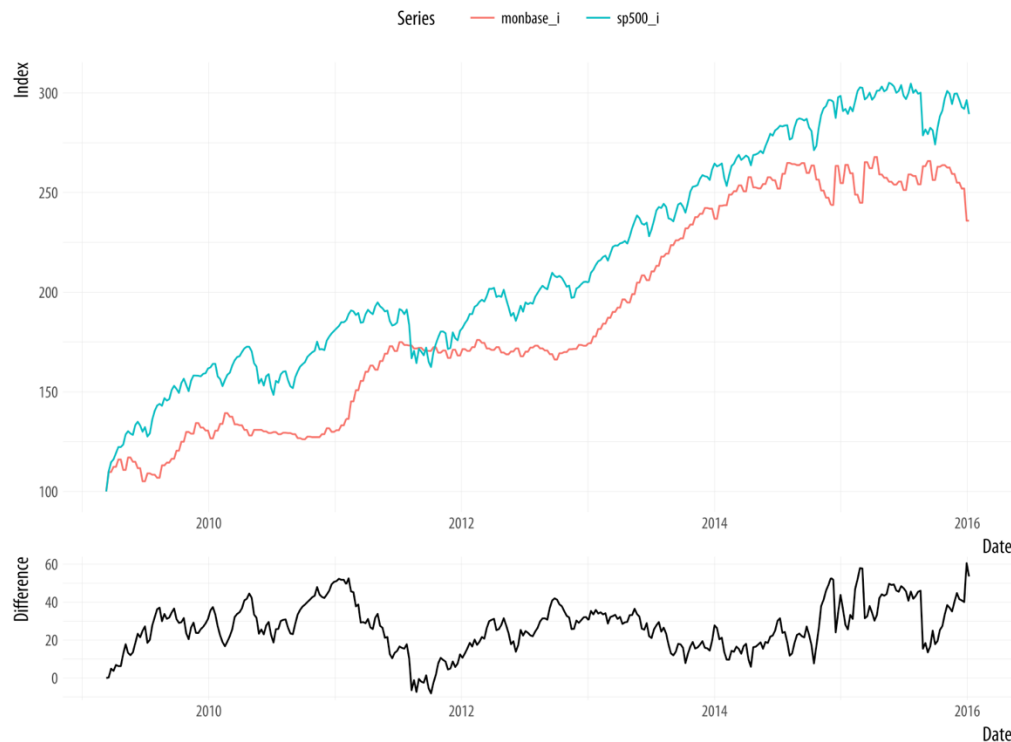
Case study 2-A bad slide

- Marissa Mayer's performance as CEO of Yahoo was being criticized by many observers.
  - *One of them, Eric Jackson, an investment fund manager, sent a 99-slide presentation to Yahoo's board outlining his best case against M...*
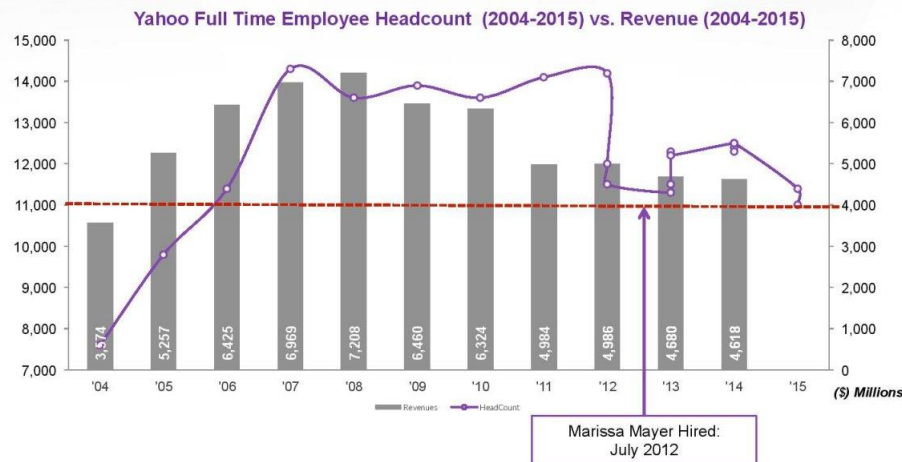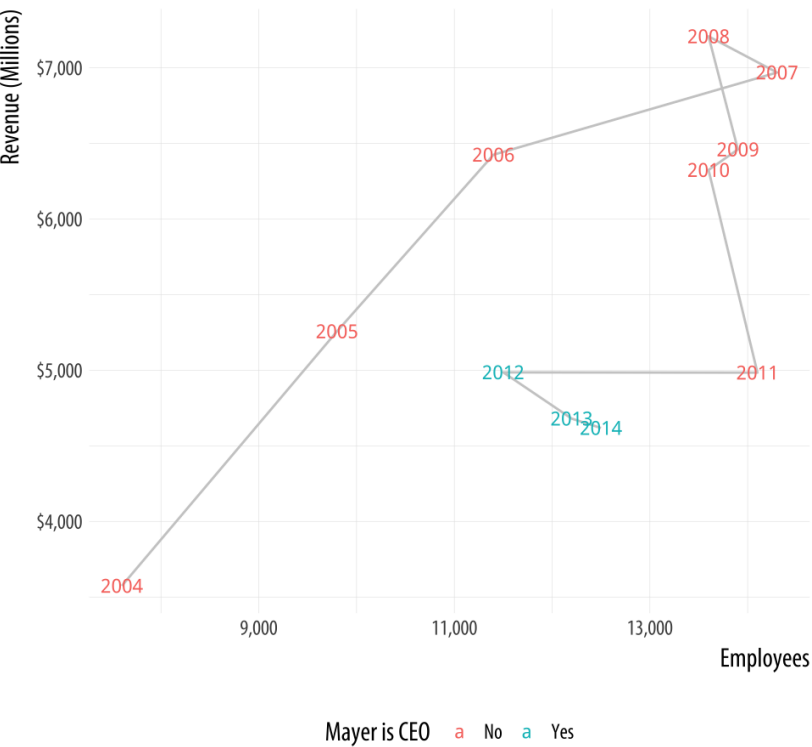
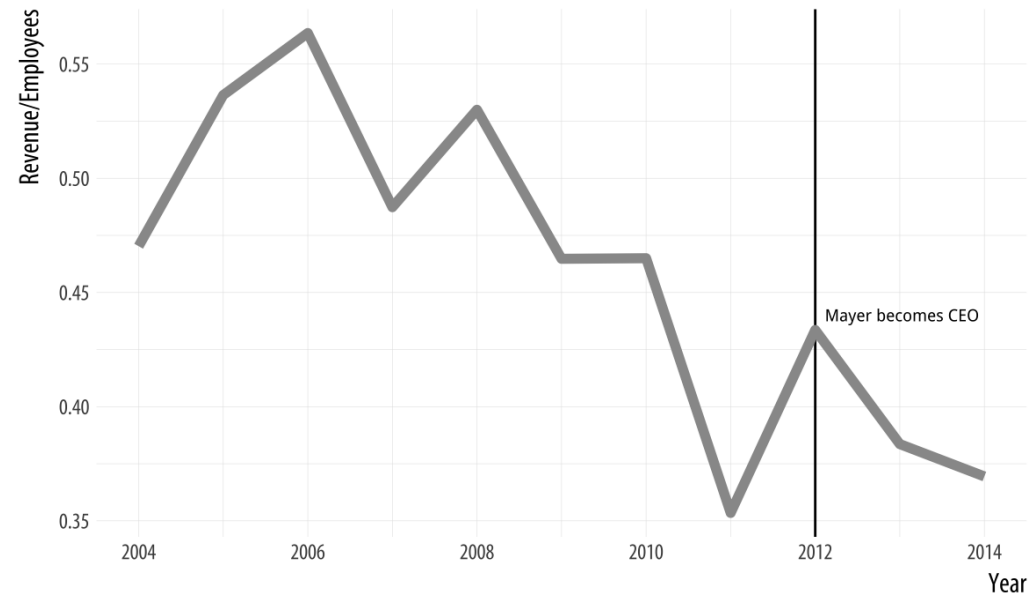# So Mayer performed poorly?

- Connected plot

Yahoo Employees vs Revenues, 2004-2014

erformance: revenue/employee

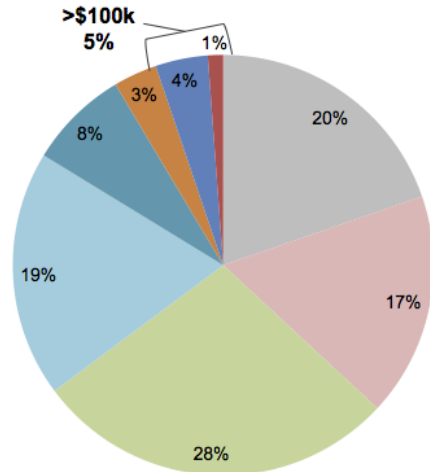Yahoo Revenue to Employee Ratio, 2004-2014

Use Case 3-Saying no to pie

- Two charts from a New York Federal Reserve Bank briefing on the structure of debt in the United States.
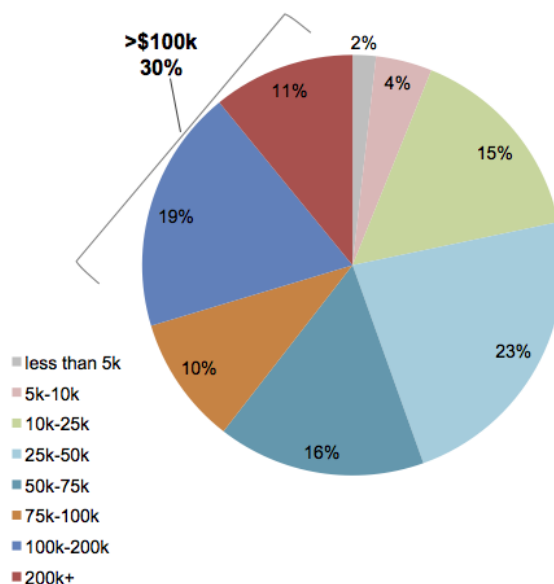
- Harder to compare values.

Borrower Distribution by Outstanding Balance
out of 44 million borrowers in 2016

Debt Distribution by Outstanding Balance
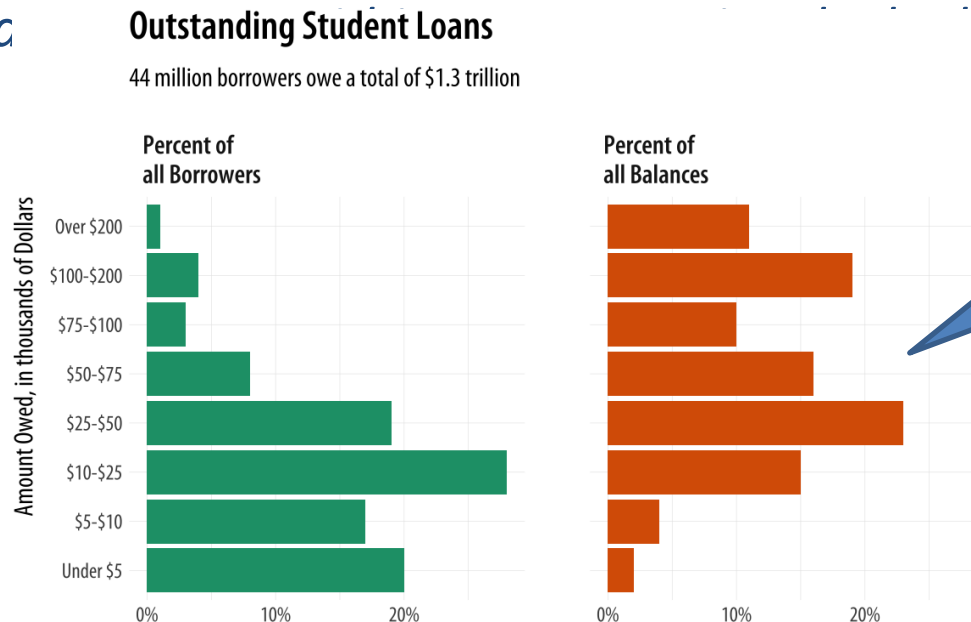out of $1.3 trillion in 2016

- less than 5k
- 5k-10k
- 10k-25k
- 25k-50k
- 50k-75k
- 75k-100k
- 100k-200k
- 200k+

- Clockwise, but hard to follow still
- Ordered, but colors not sequential
- Too much annotation like a table

Alternatively-1

- Split the data into the two categories, and showed the percentage shares as bars.
  - *The percent scores are on the x-axis. Instead of coloring to distinguish the debt categories, put their values on the y-axis instead.*
  - *We ca* ~~~~ *ing down the bars.*

Outstanding Student Loans
44 million borrowers owe a total of $1.3 trillion

fill is redundant, and portion info lost!

Source: FRB NY

Alternatively-2

- Instead of having separate bars distinguished by heights, we can array the percentages for each distribution proportionally within a single bar.

- We will make a stacked bar chart with just two main bars, and lie them on their side for comparison.

Outstanding Student Loans

44 million borrowers owe a total of $1.3 trillion

Amount Owed, in thousands of dollars

| Under $5 | $5-$10 | $10-$25 | $25-$50 | $50-$75 | $75-$100 | $100-$200 | Over $200 |

Source: FRB NY