Teaching the Basics of Data Journalism Part 2: Data Visualizations

Mai Nguyen

Spring 2021

QTM and Emory Writing Center

Outline

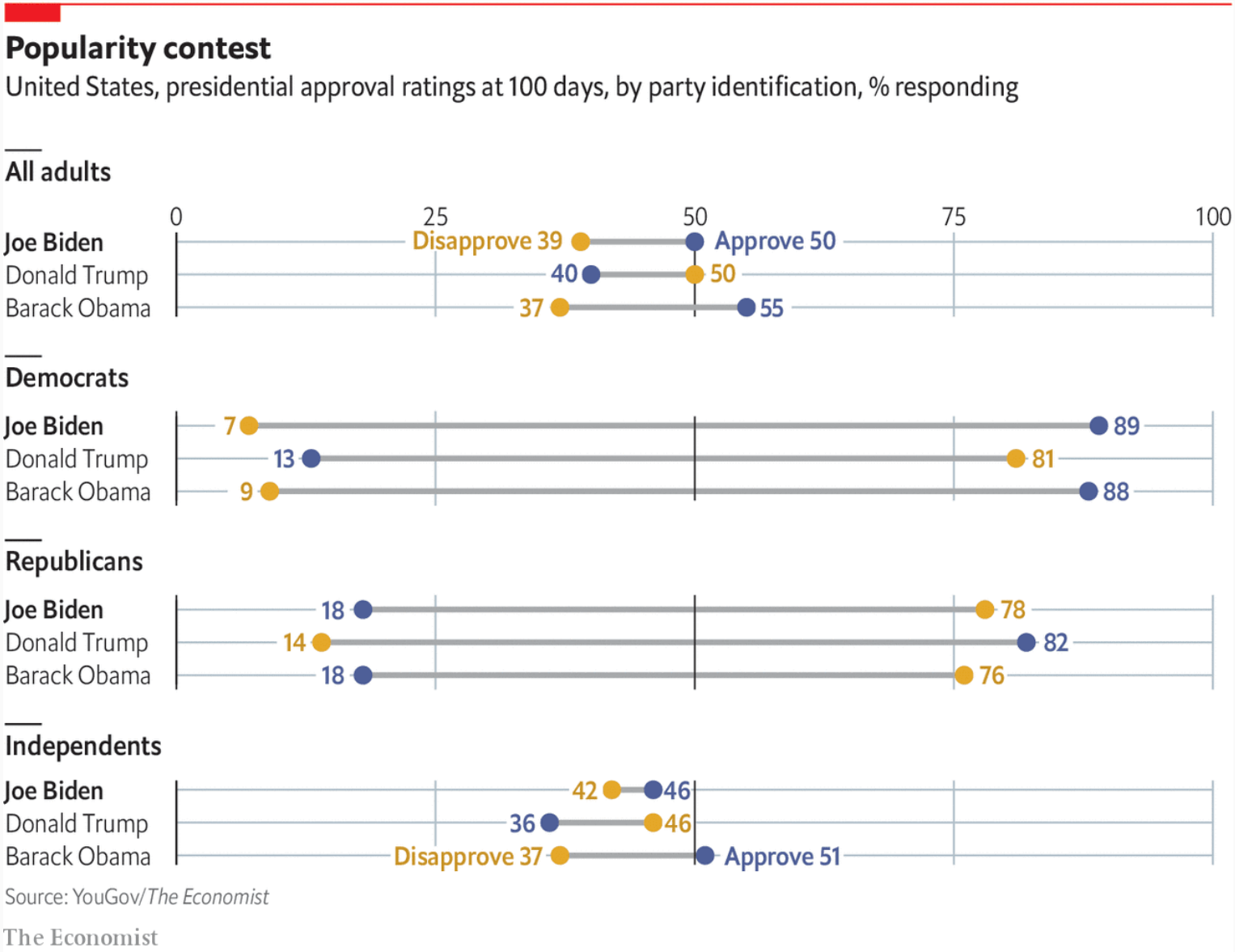
- What makes a good data visualization?
- How to create good data visualizations
  - Some design principles
  - General guidelines
- Additional resources

What Makes a Good Data Visualization?

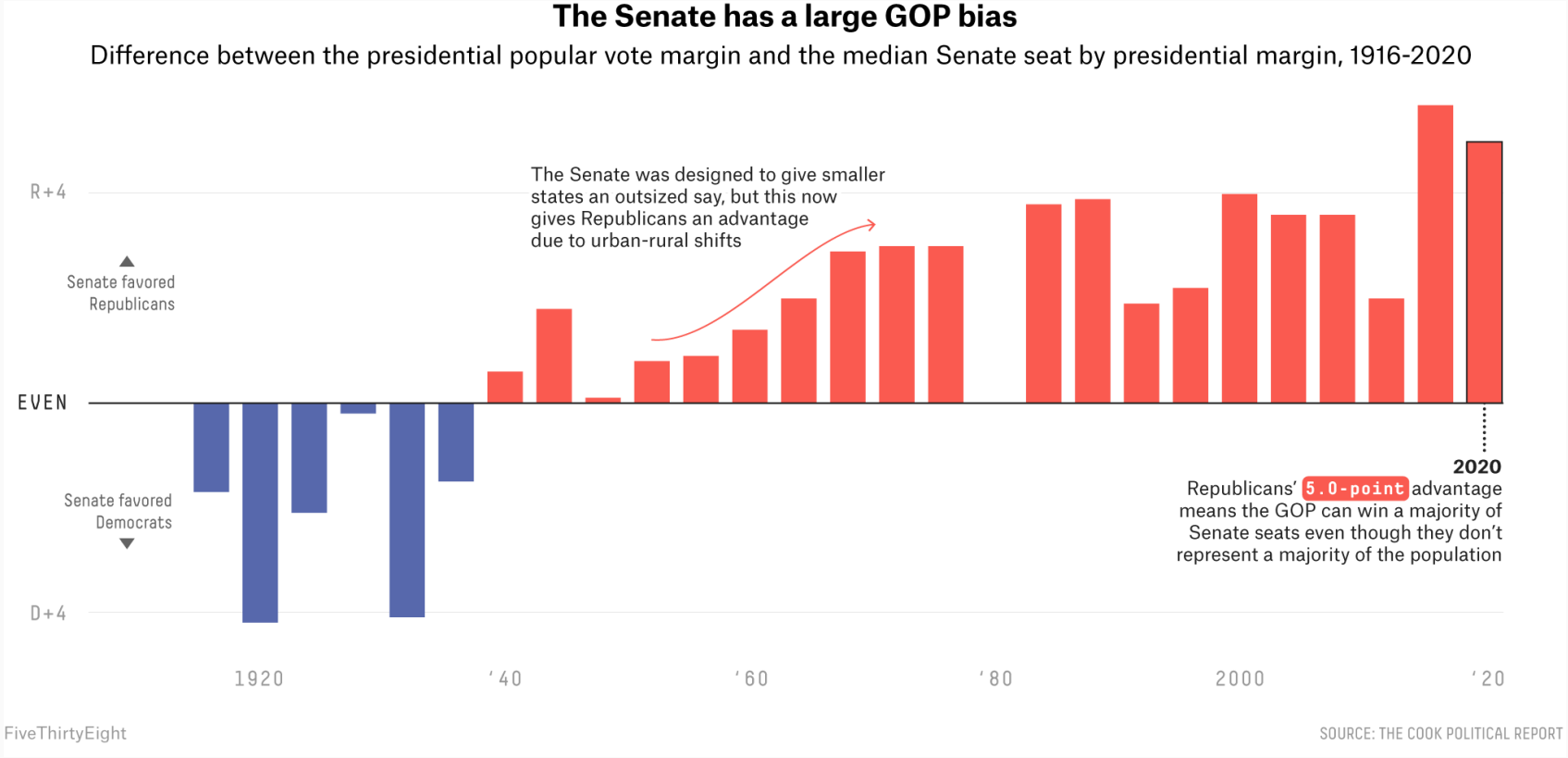
Activity

- I'm going to show you a few data visualizations
- Think about a few things:
  - What is the graphic trying to convey?
  - What works?
  - What doesn't work?

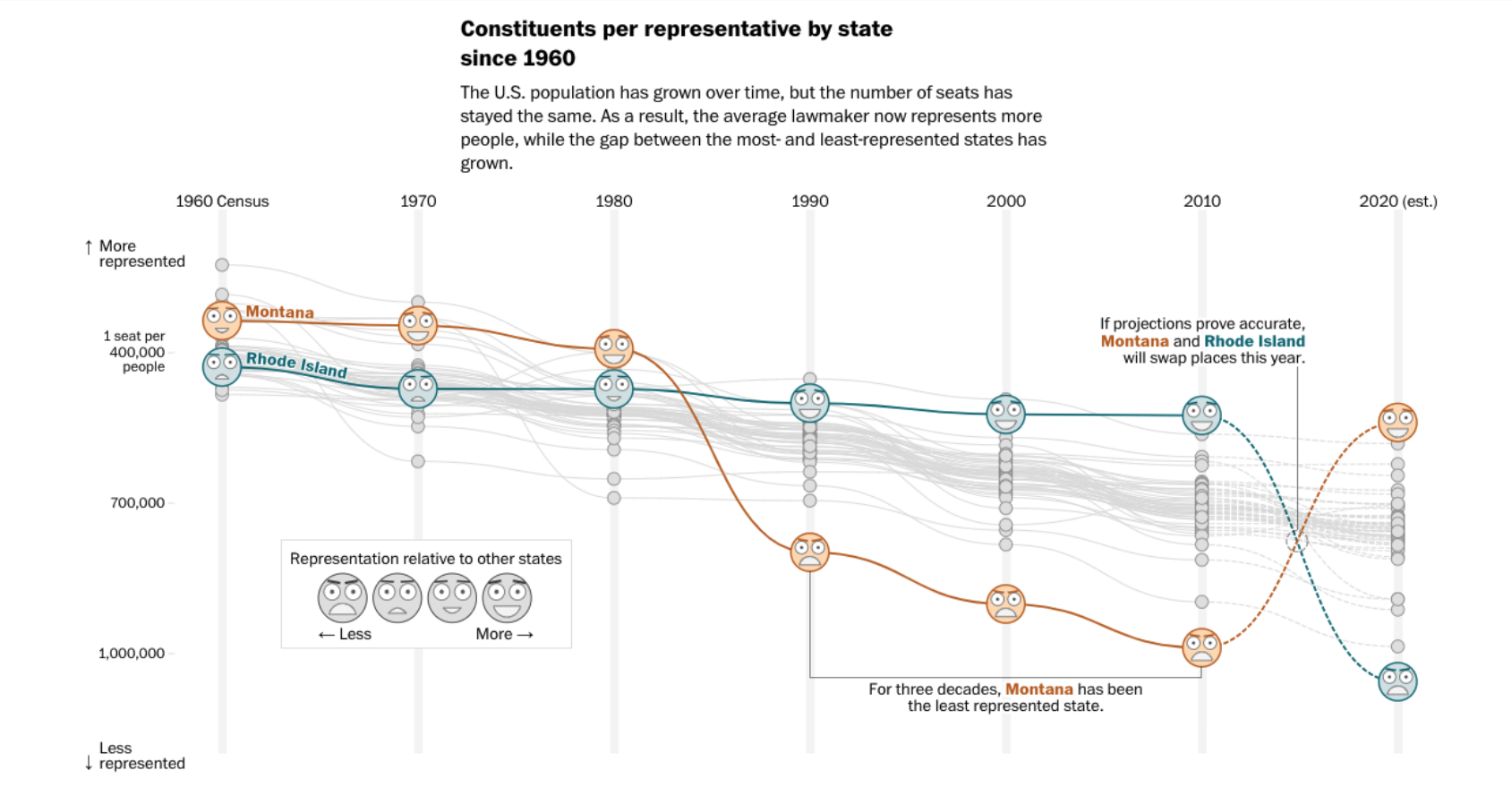
Example 1: The Economist



Example 2: FiveThirtyEight



Example 3: The Washington Post



Graphical Excellence

"...the greatest number of ideas, in the shortest time, using the least amount of ink, in the smallest space."

Edward Tufte

The Visual Display of Quantitative Information (1983)

In other words...

Good data visualization should:

tell a story as efficiently as possible

In other words...

Good data visualization should:

tell a story as efficiently as possible

In other words...

Good data visualization should:

tell a story as **efficiently as possible**

Telling a Story

What's your message?

- Why are you analyzing data?
  - *Exploratory*: identifying patterns and trends in the data
  - *Explanatory*: answering why we see patterns in the data
- What is the key takeaway/recommendation?

Telling a Story

What's your message?

- Why are you analyzing data?
  - *Exploratory*: identifying patterns and trends in the data
  - *Explanatory*: answering why we see patterns in the data
- What is the key takeaway/recommendation?

Telling a Story

What's your message?

- Why are you analyzing data?
  - *Exploratory*: identifying patterns and trends in the data
  - *Explanatory*: answering why we see patterns in the data
- What is the key takeaway/recommendation?

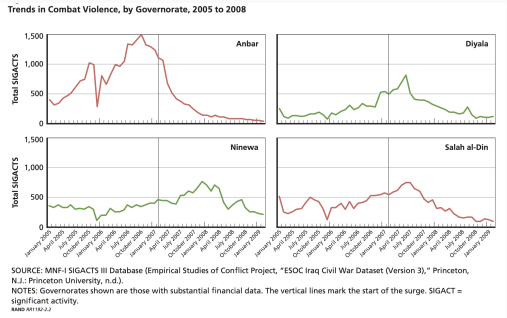
Telling a Story

Who's your audience?

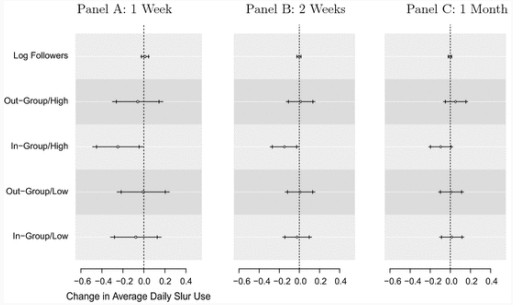
- Different audiences/contexts = different requirements

Media

Industry



Academic



Keep it Simple

- Simplicity $\neq$ Bare
- Think about **data-ink** (Tufte 1983)
  - Data-ink refers to the “ink” on the graph that represents data
  - We want to maximize the ratio of data-ink to non-data-ink
- The *data* should be the star of the show, everything else is support and should be minimized as much as possible

Keep it Simple

- Simplicity $\neq$ Bare
- Think about **data-ink** (Tufte 1983)
  - Data-ink refers to the “ink” on the graph that represents data
  - We want to maximize the ratio of data-ink to non-data-ink
- The *data* should be the star of the show, everything else is support and should be minimized as much as possible

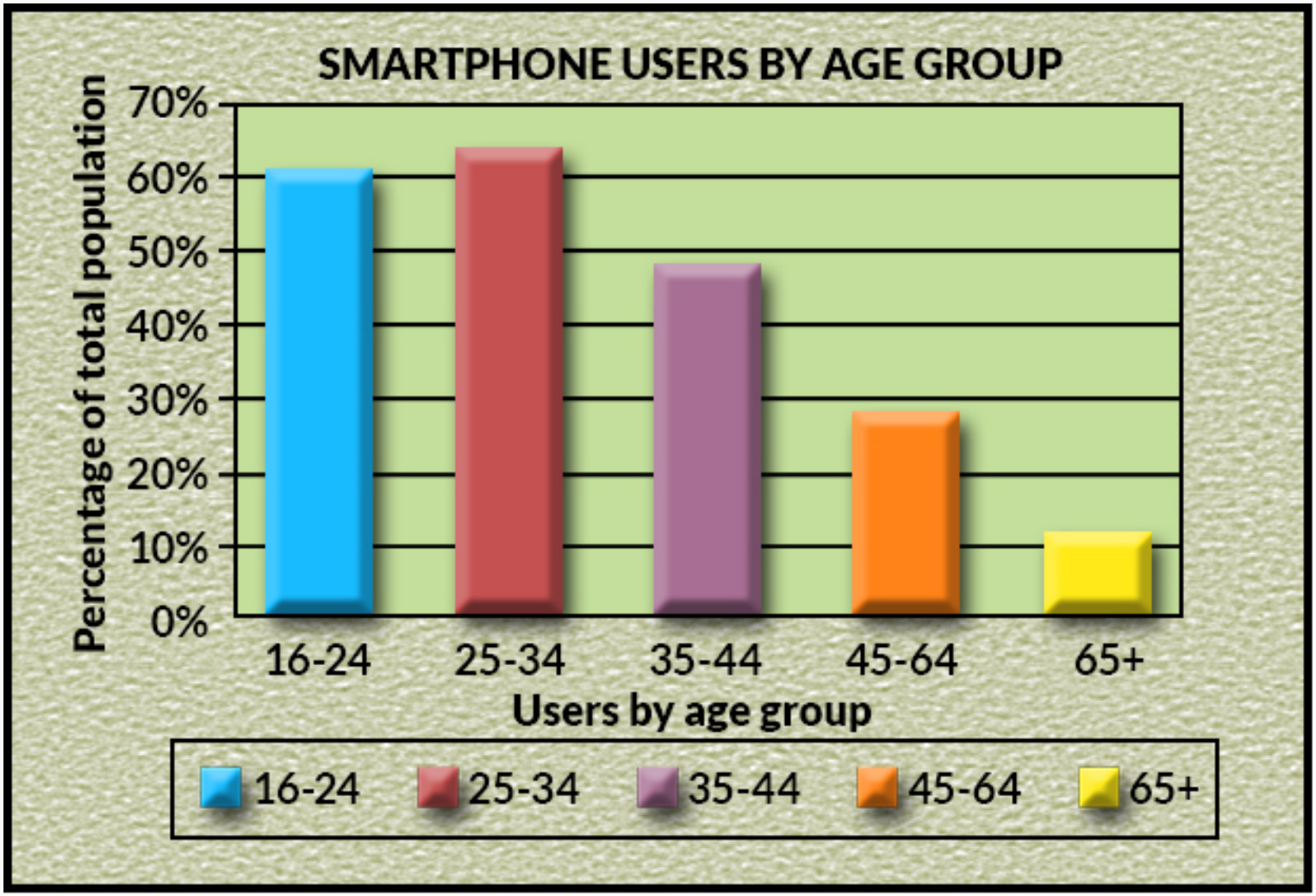
Keep it Simple

- Simplicity $\neq$ Bare
- Think about **data-ink** (Tufte 1983)
  - Data-ink refers to the “ink” on the graph that represents data
  - We want to maximize the ratio of data-ink to non-data-ink
- The *data* should be the star of the show, everything else is support and should be minimized as much as possible

Keep it Simple

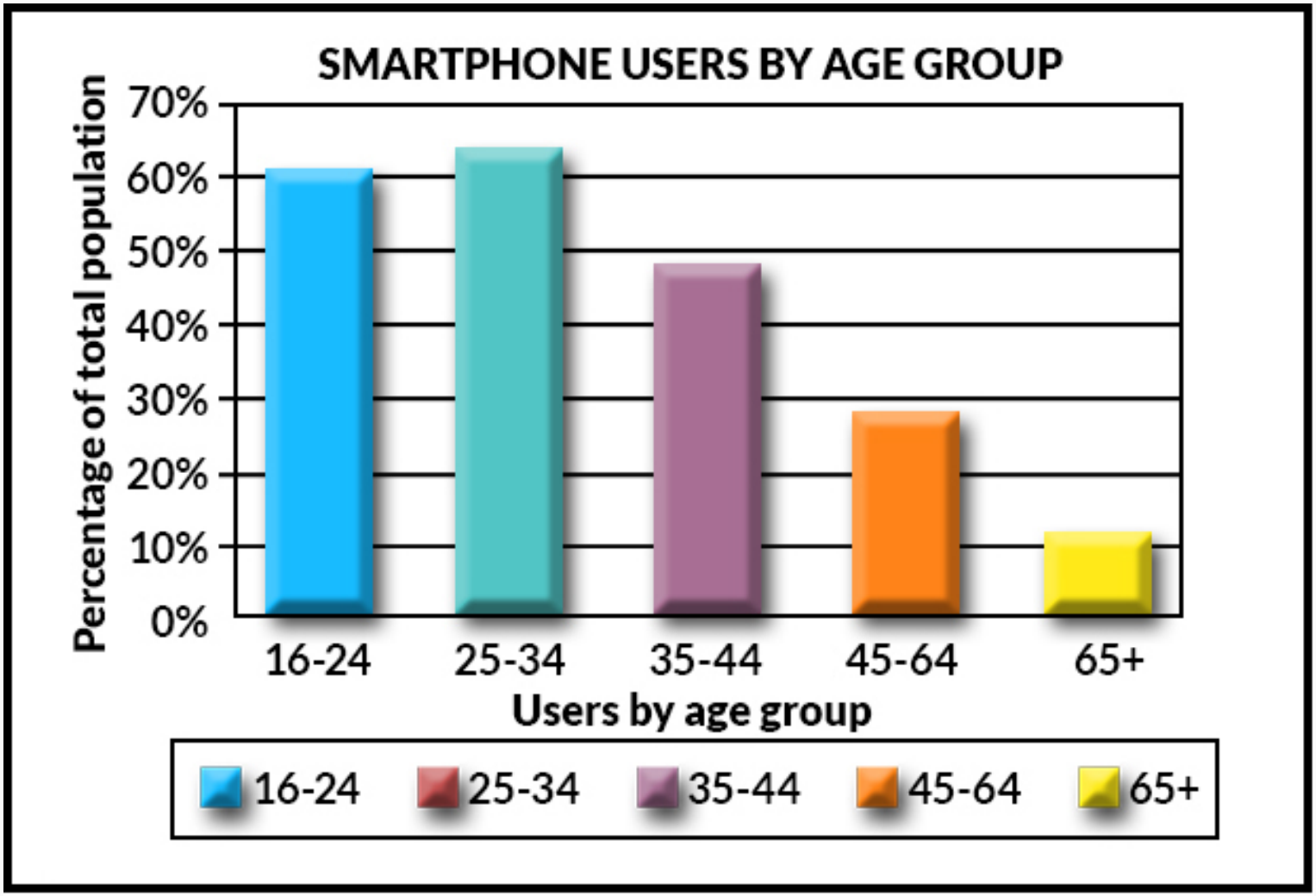
- Simplicity $\neq$ Bare
- Think about **data-ink** (Tufte 1983)
  - Data-ink refers to the “ink” on the graph that represents data
  - We want to maximize the ratio of data-ink to non-data-ink
- The *data* should be the star of the show, everything else is support and should be minimized as much as possible

Keeping it Simple



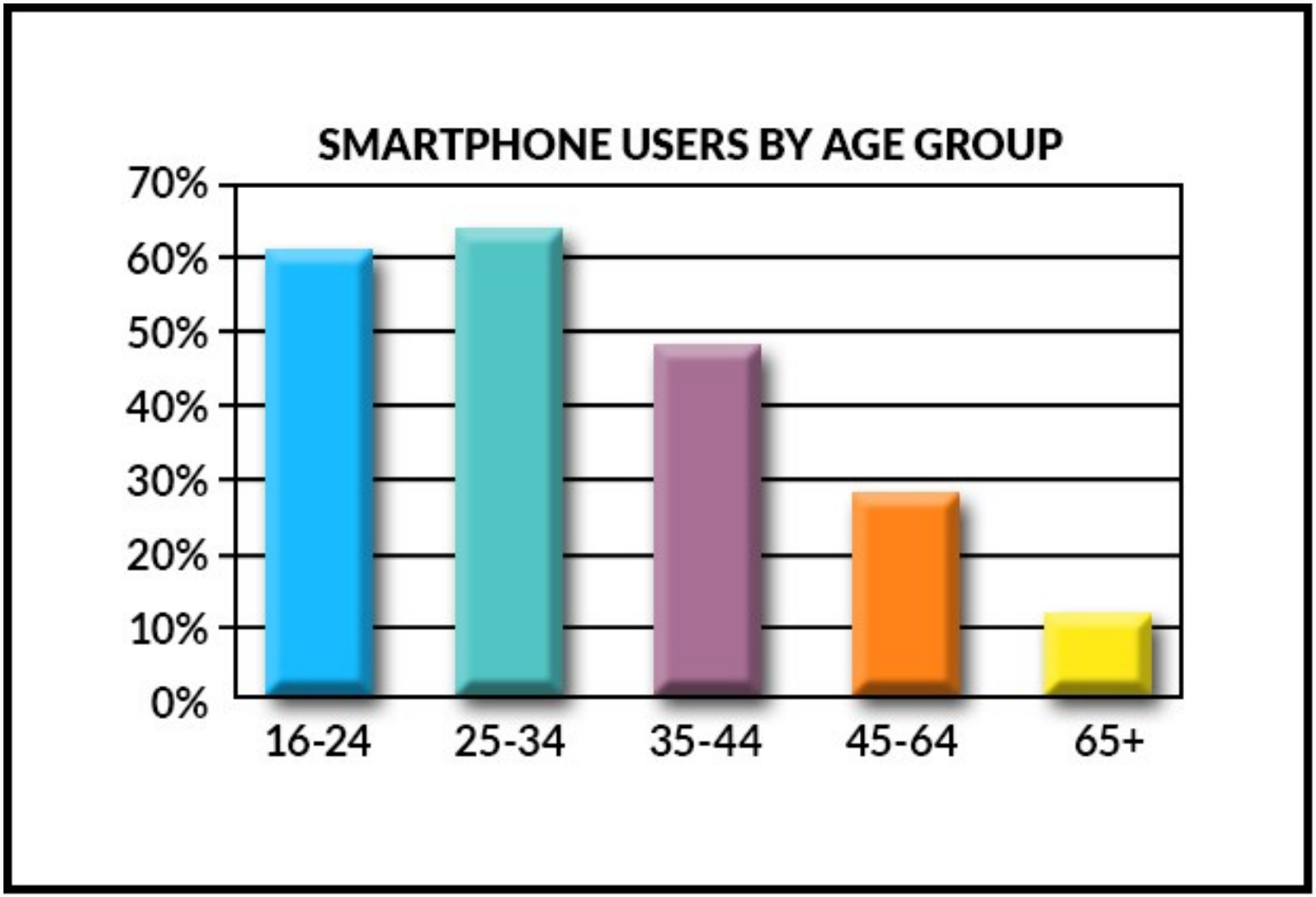
Source: GiveGoodUX

Keeping it Simple



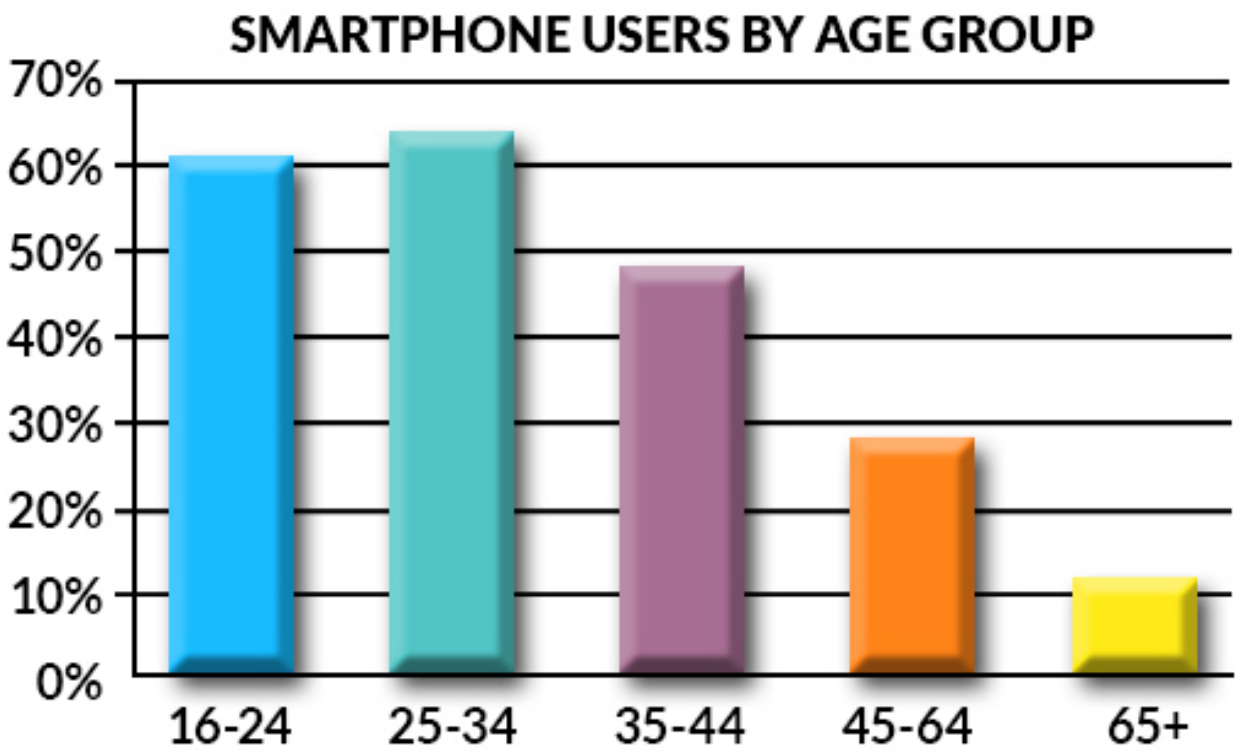
Source: GiveGoodUX

Keeping it Simple



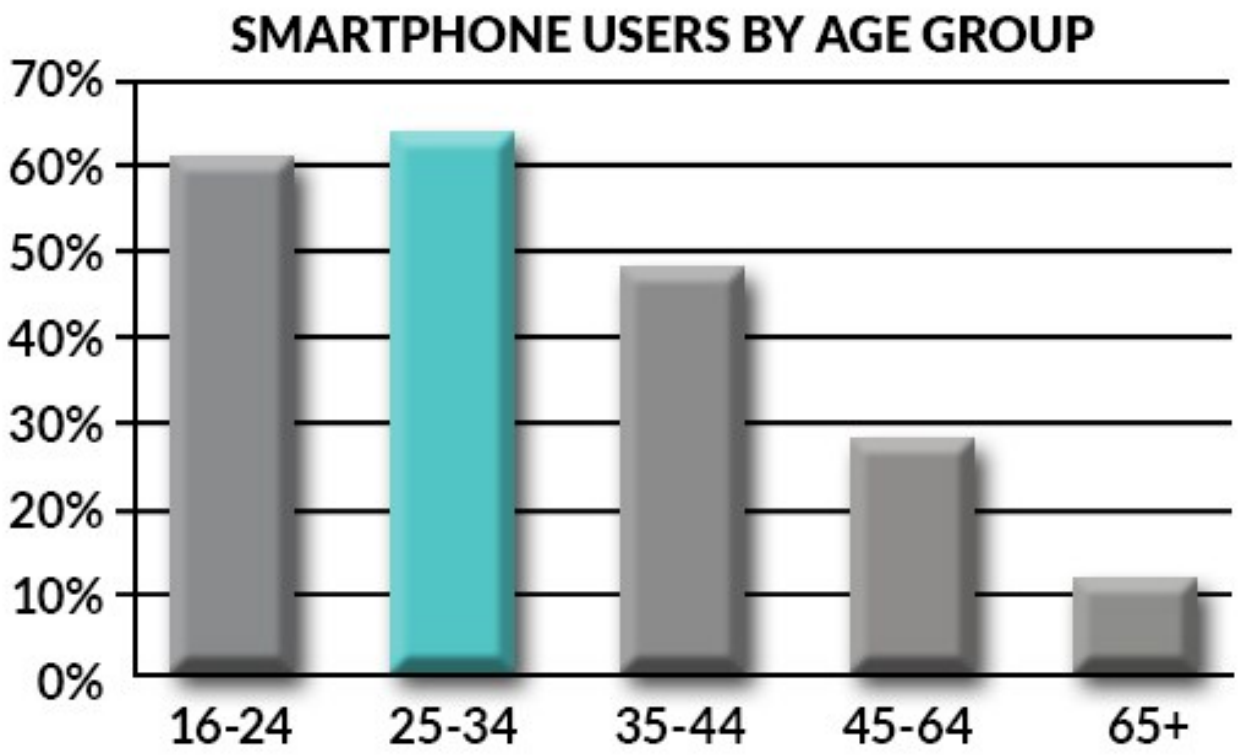
Source: GiveGoodUX

Keeping it Simple



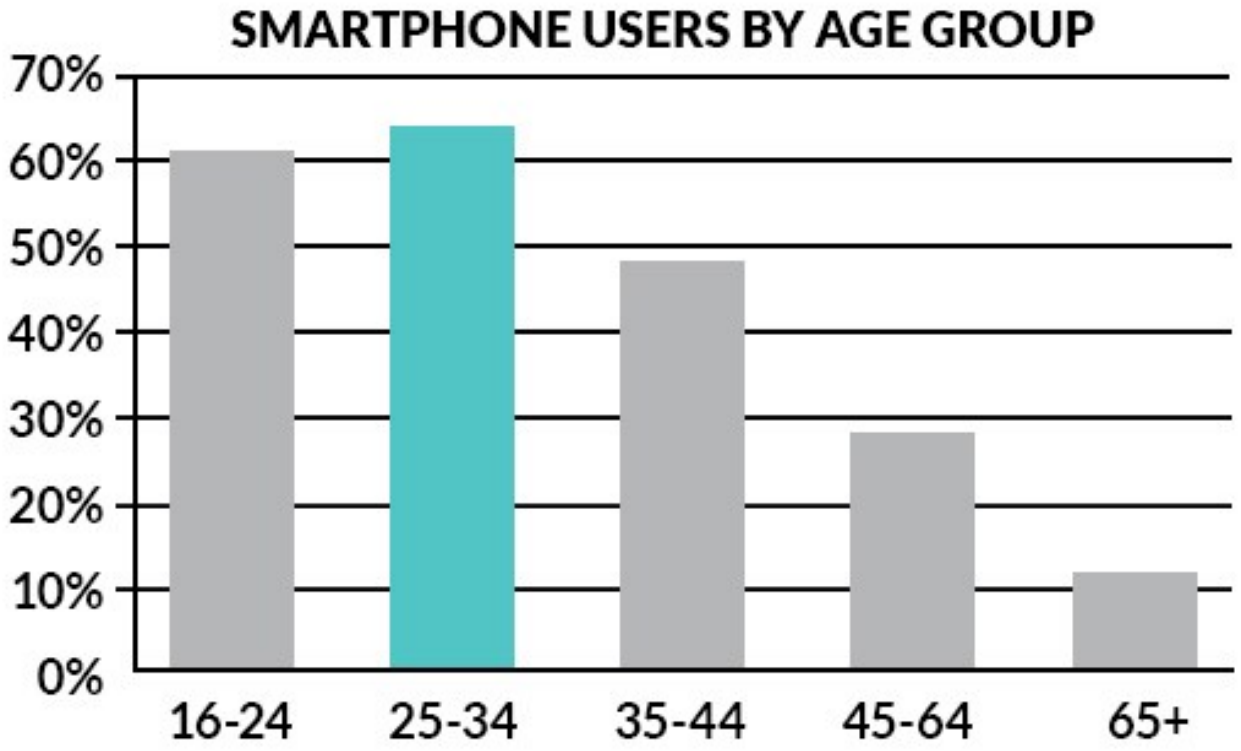
Source: GiveGoodUX

Keeping it Simple



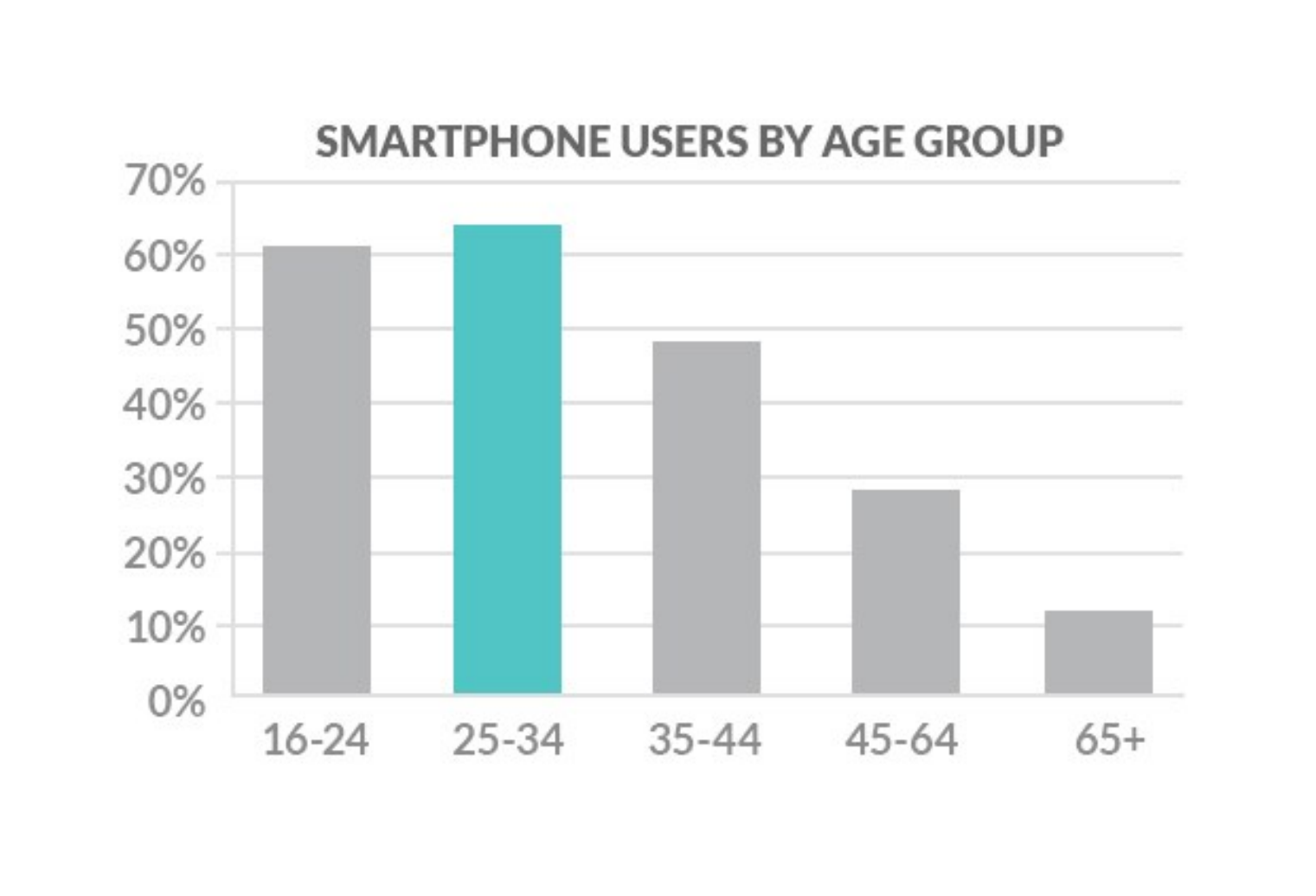
Source: GiveGoodUX

Keeping it Simple



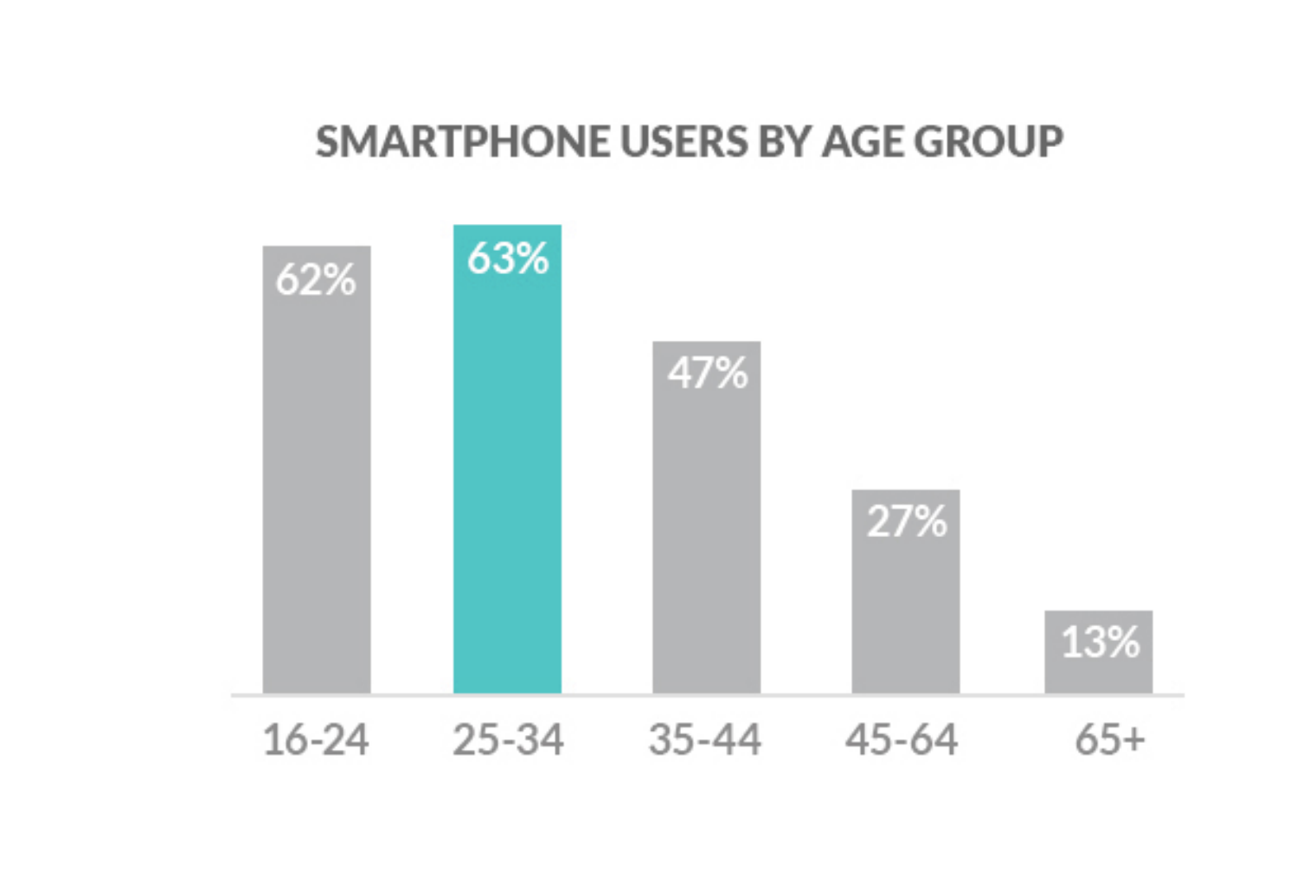
Source: GiveGoodUX

Keeping it Simple



Source: GiveGoodUX

Keeping it Simple



Source: GiveGoodUX

How to Create Good Data Visualizations

Choose the Right Chart Type for Your Data

- **What variables are you working with?**
  - How many variables are you using?
  - What type of variables are they? Continuous? Categorical?
  - Do you have time-series data?

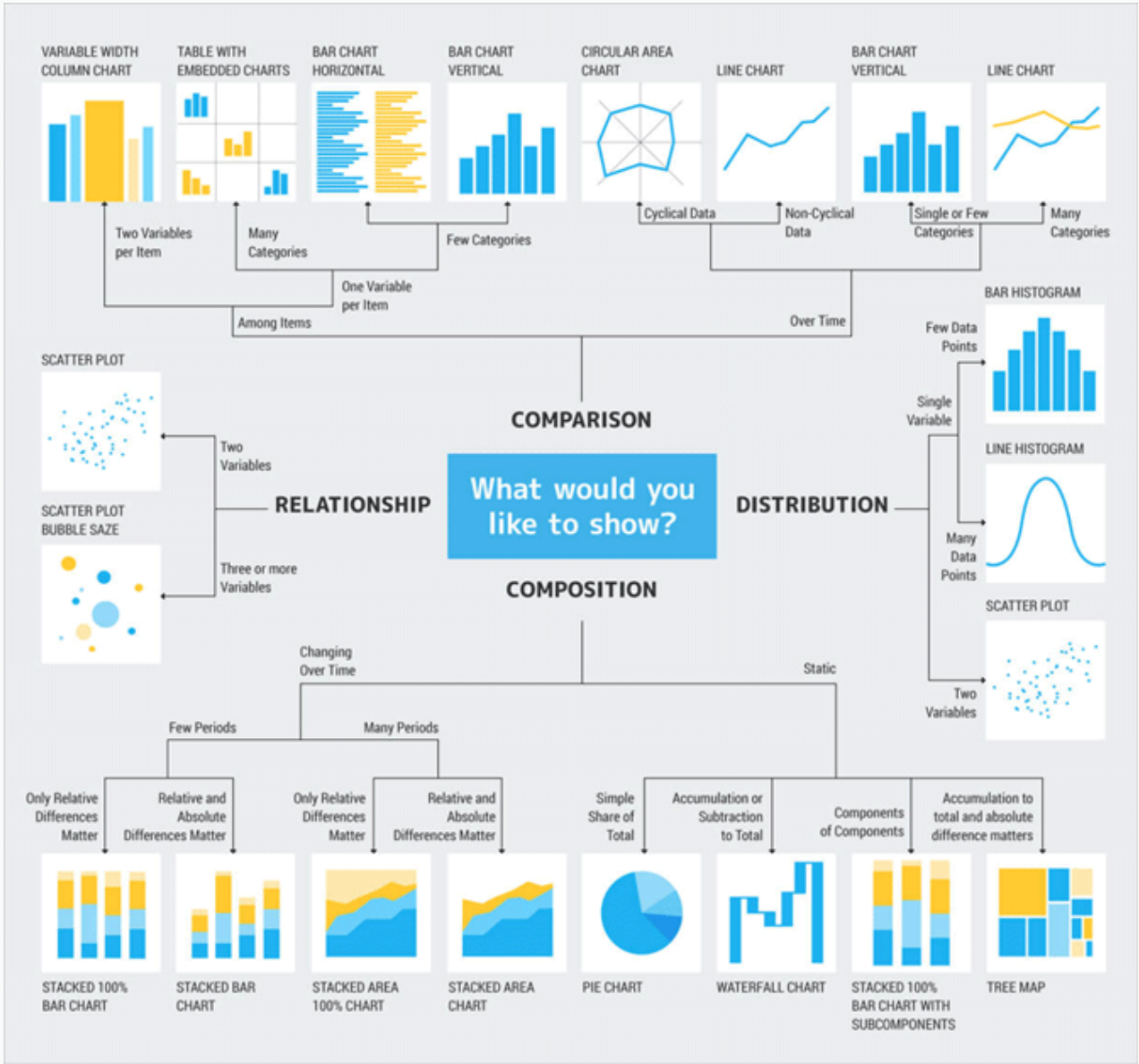
- **What are you trying to show with these variables?**
  - Are you trying to show the relationship between two variables?
  - Or do you want to show the breakdown and distributions of specific variables?

Choose the Right Chart Type for Your Data

- **What variables are you working with?**
  - How many variables are you using?
  - What type of variables are they? Continuous? Categorical?
  - Do you have time-series data?

- **What are you trying to show with these variables?**
  - Are you trying to show the relationship between two variables?
  - Or do you want to show the breakdown and distributions of specific variables?

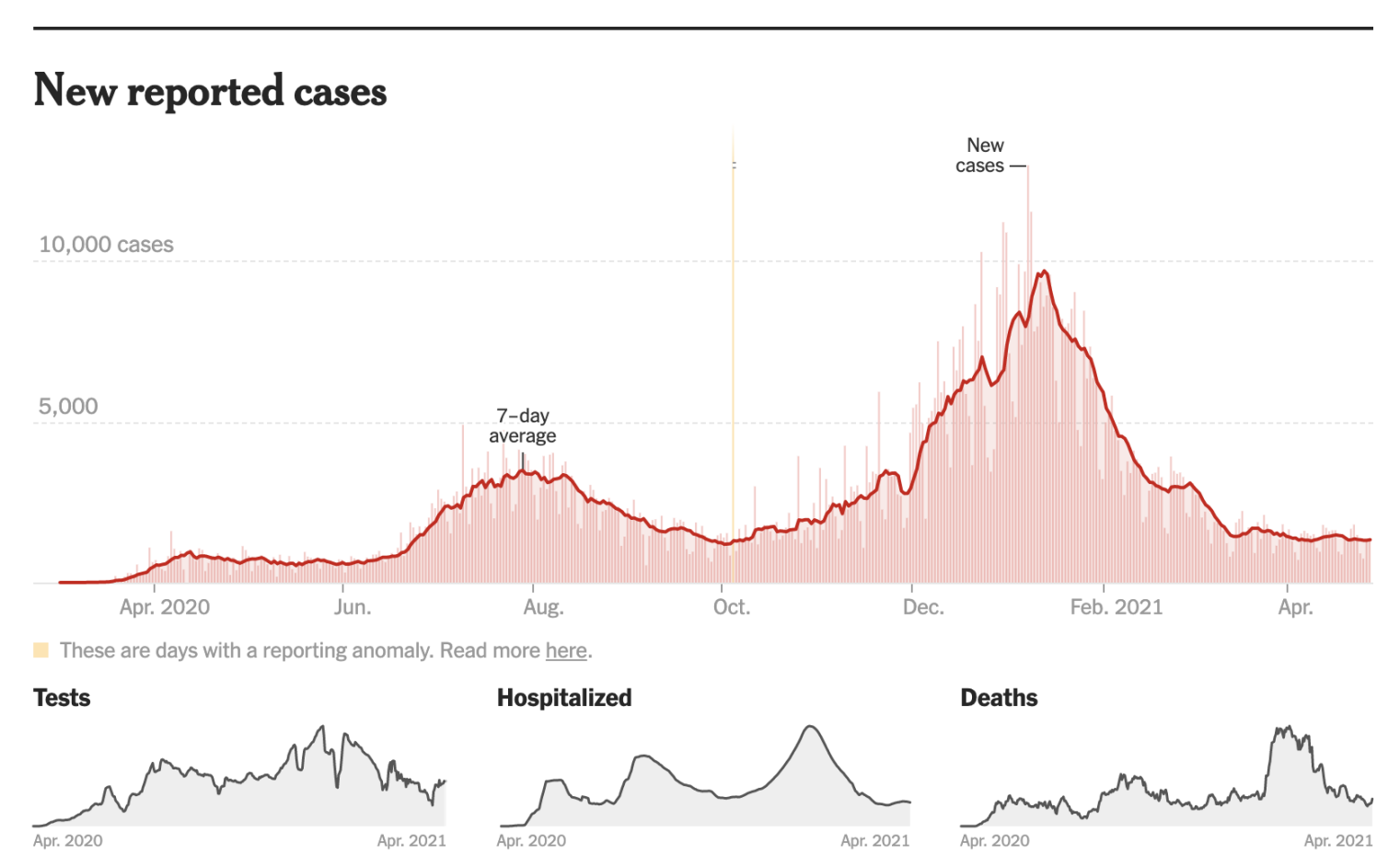
Choose the Right Chart Type for Your Data



Source: Andrew Abela

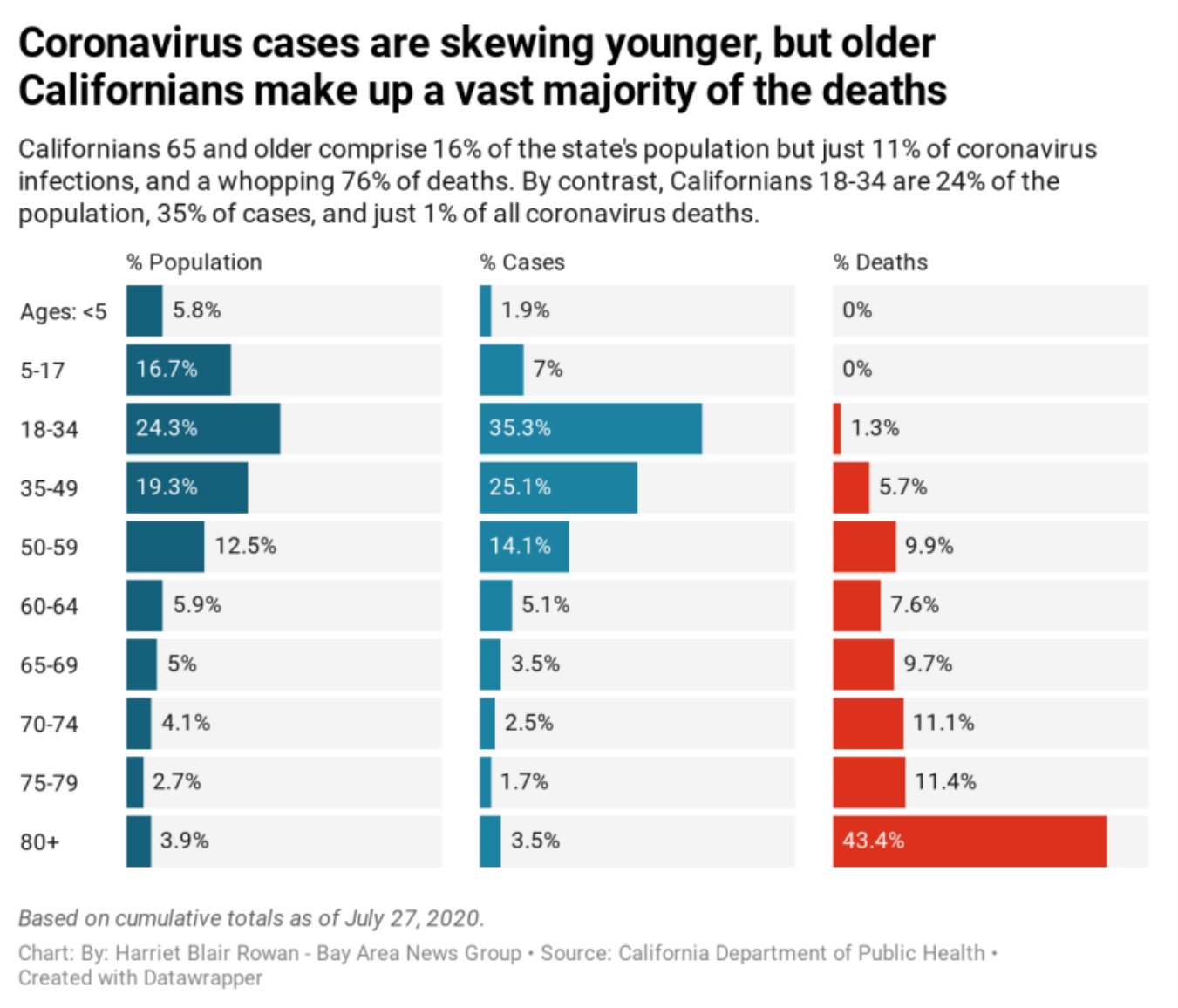
Line Chart

Uses: showing changes over time



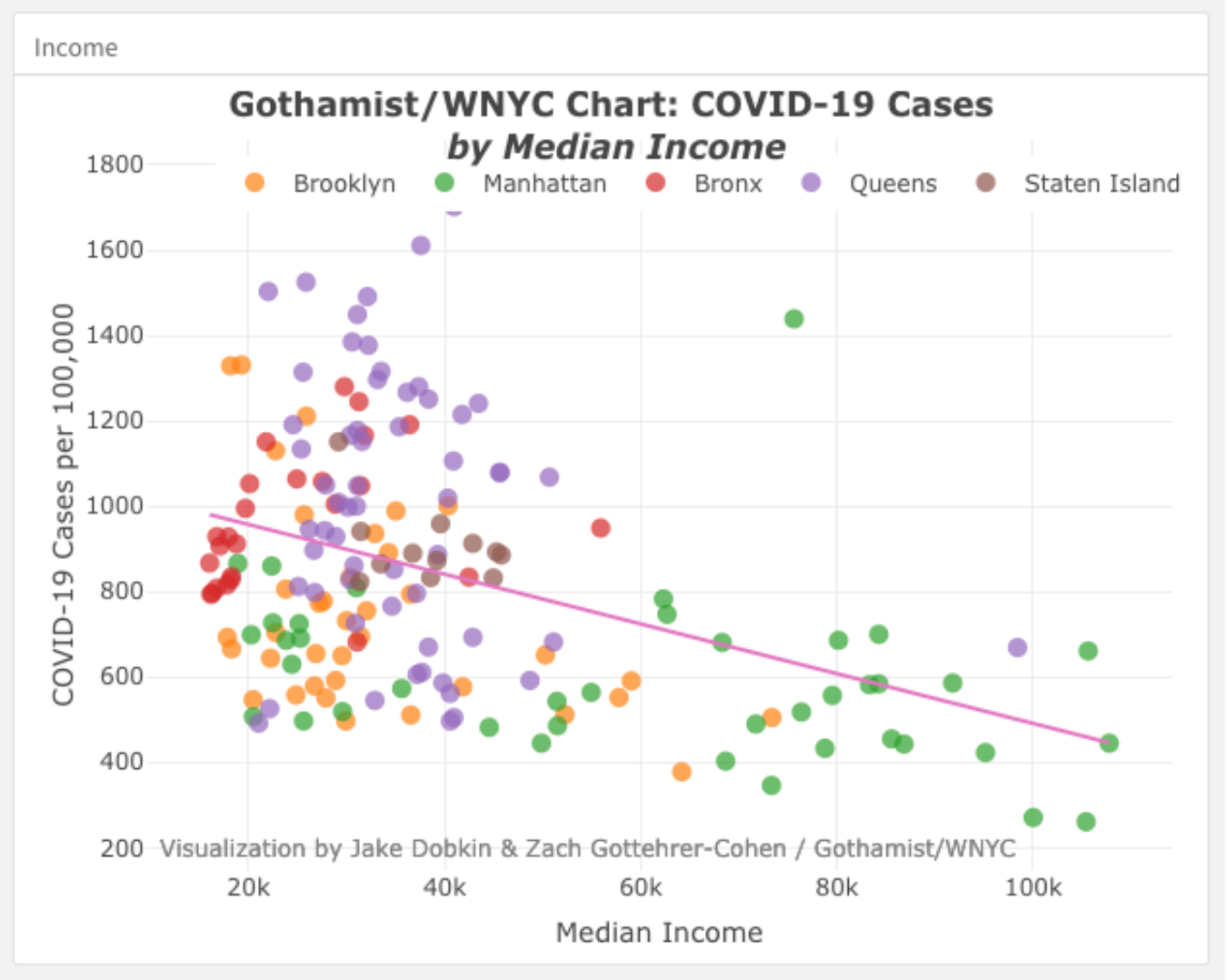
Bar Chart

Uses: showing similarities or differences between groups



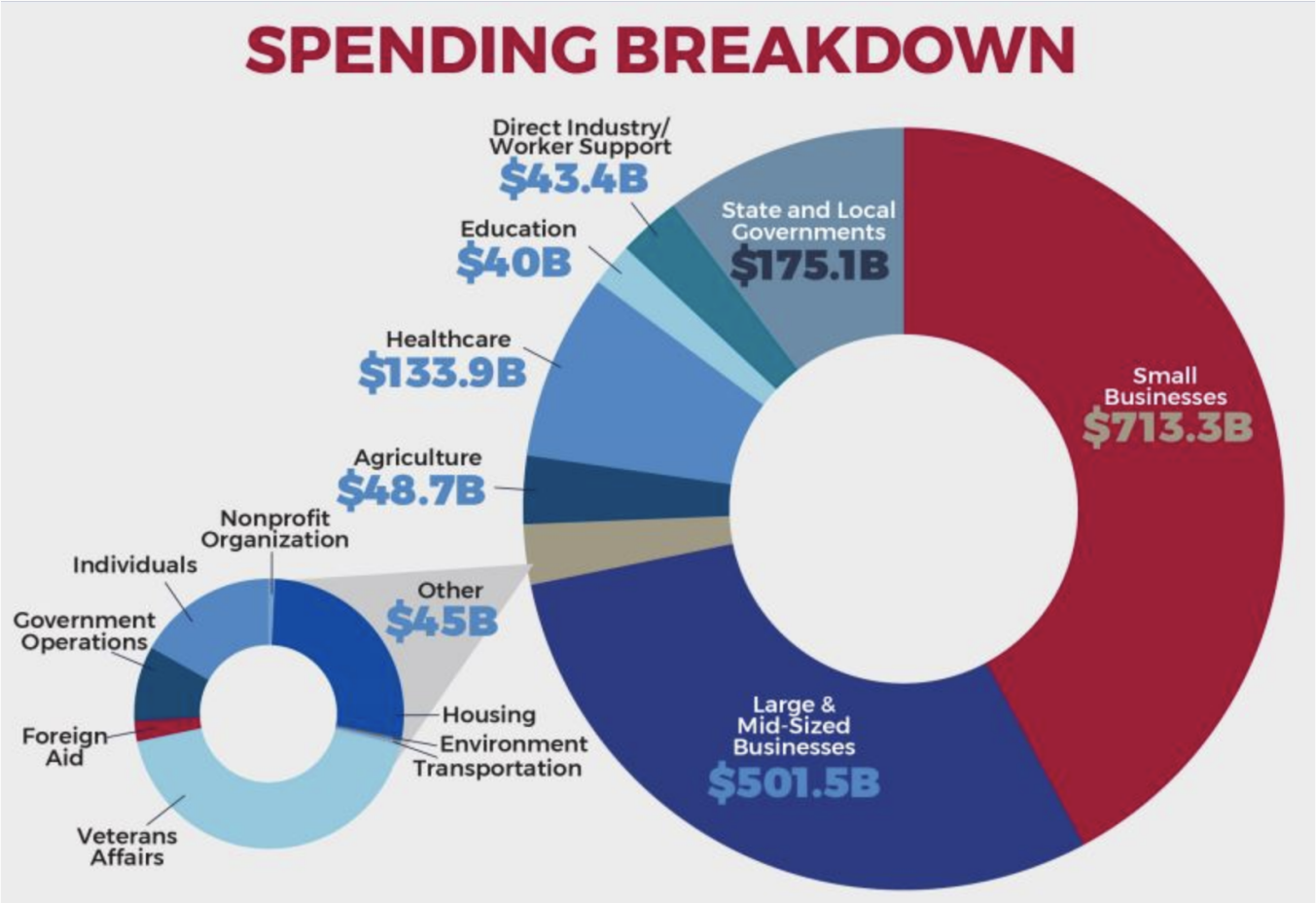
Scatter Plot

Uses: showing relationships



Pie Chart

Uses: showing compositions



The Humble Table

Uses: when exact values are important

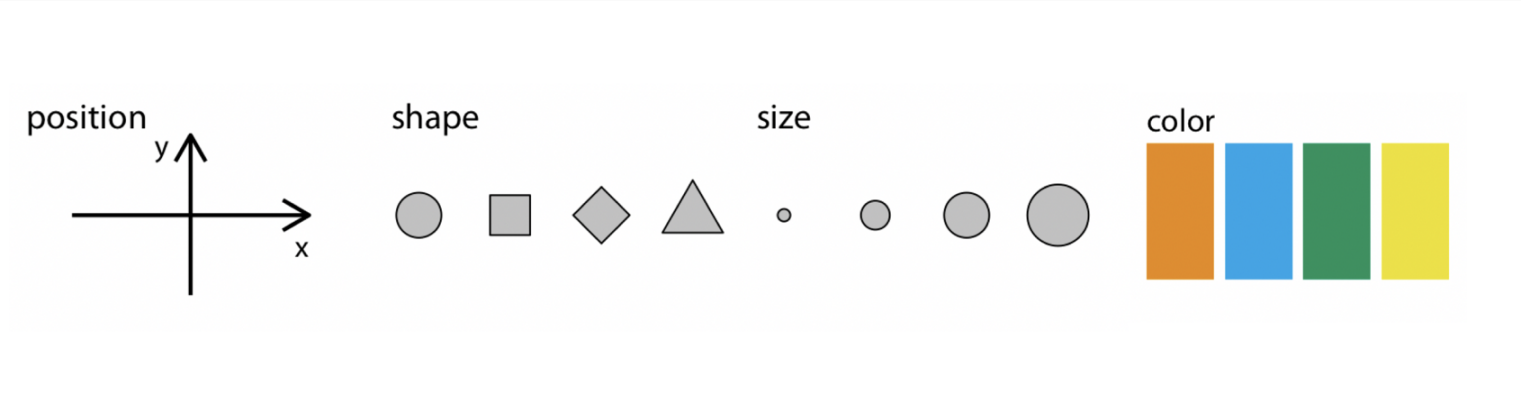
Ten counties with highest rates of reported cases

Deaths | Cases | Adj. for population | Totals

County	Total reported cases per 100k	New cases in last 7 days per 100k
Willacy County, Tex.	13,469	1,508
Chattahoochee County, Ga.	36,213	817
Motley County, Tex.	9,948	779
Huron County, Mich.	12,462	745
Grant County, Ore.	6,682	696
Muskegon County, Mich.	8,196	648
Briscoe County, Tex.	10,867	647
Hemphill County, Tex.	14,430	616
Coos County, N.H.	5,827	612
Klamath County, Ore.	5,889	504

Using the Right Tool for the Job

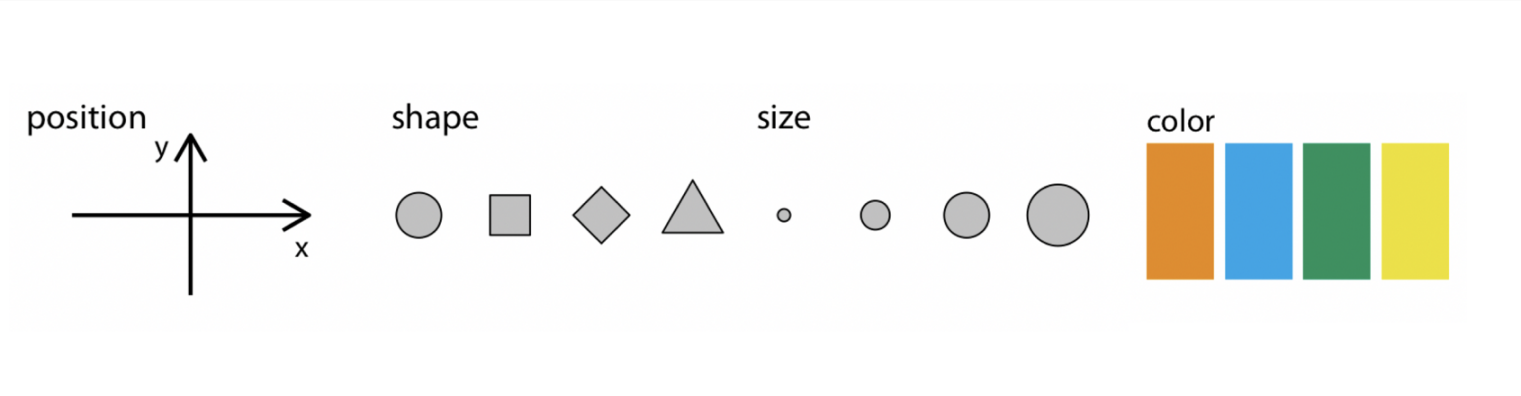
- Tools we can use to encode information include:



- **Perceptual topology should match data topology**
  - Understand how people generally perceive things
  - Goal is to be **emphathetic** to the audience

Using the Right Tool for the Job

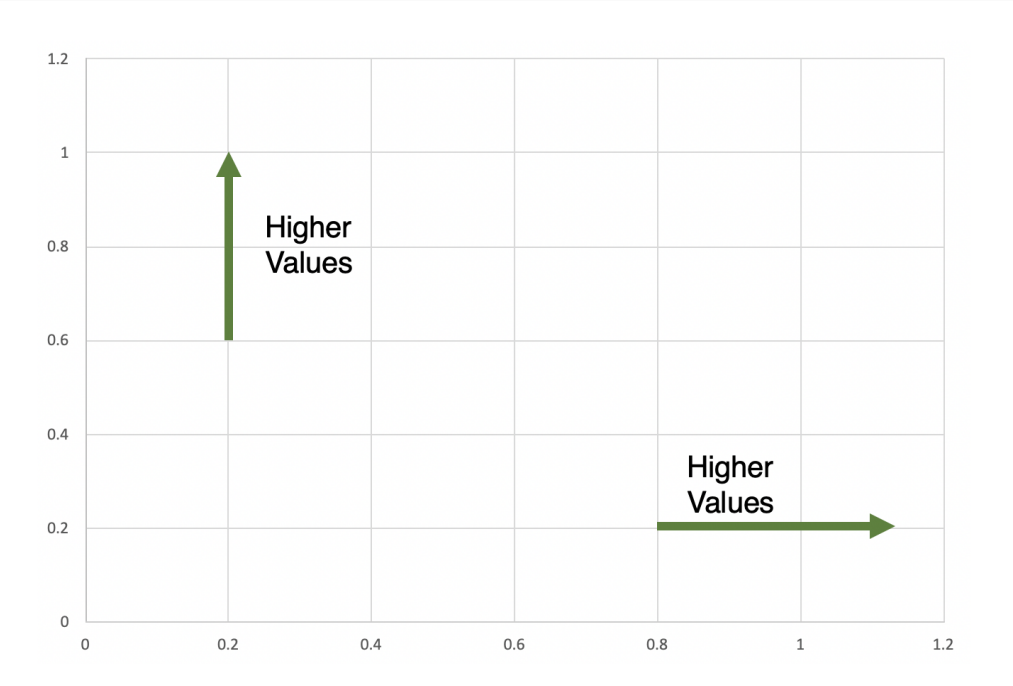
- Tools we can use to encode information include:



- **Perceptual topology should match data topology**
  - Understand how people generally perceive things
  - Goal is to be **emphathetic** to the audience

Position

- Arguably the most important tool in your toolbox
- Placement of things matter for how people will interpret the information



- Position from left to right can also indicate time
- Additionally, placement of objects together signal similarity

Size

- Size is inherently ordered
  - Objects that are larger imply larger values
  - Use this to your advantage
- Be consistent when using size and avoid distortions
  - An object that is a third of the size of another should also be a third of the value
- When using size, make sure there is enough variation so that different values can be easily interpreted

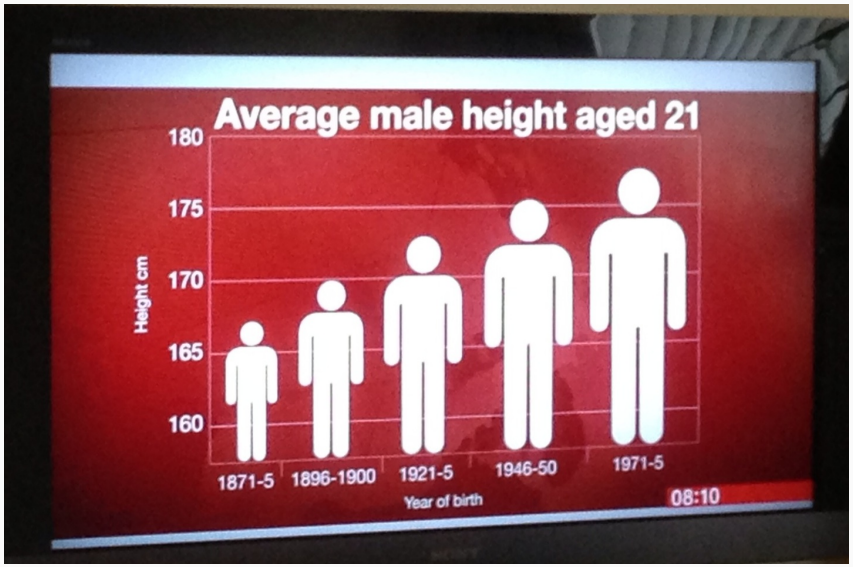
Size

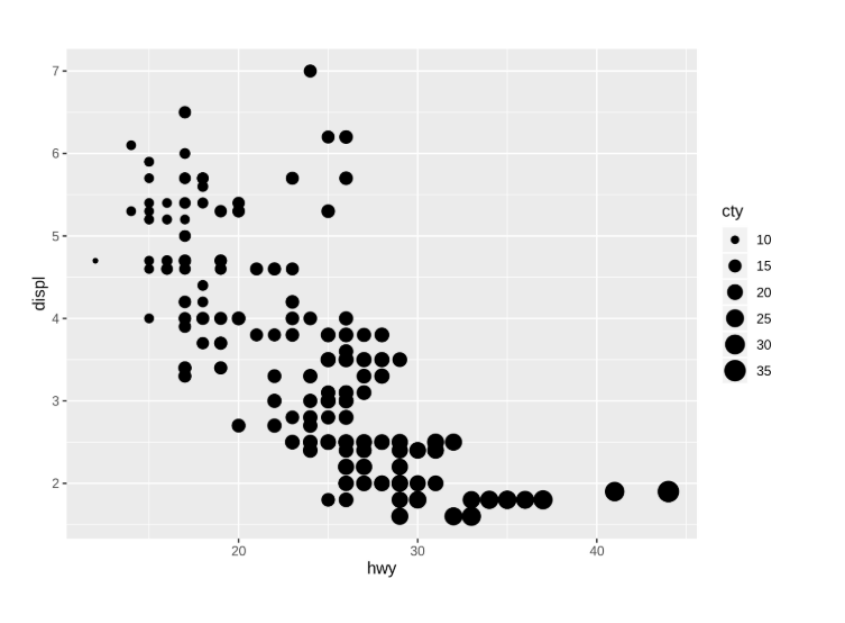
- Size is inherently ordered
  - Objects that are larger imply larger values
  - Use this to your advantage
- Be consistent when using size and avoid distortions
  - An object that is a third of the size of another should also be a third of the value
- When using size, make sure there is enough variation so that different values can be easily interpreted

Size

- Size is inherently ordered
  - Objects that are larger imply larger values
  - Use this to your advantage
- Be consistent when using size and avoid distortions
  - An object that is a third of the size of another should also be a third of the value
- When using size, make sure there is enough variation so that different values can be easily interpreted

Size


Average male height aged 21
180
175
170
165
160
Height cm
1871-5
1896-1900
1921-5
1946-50
1971-5
Year of birth
08:10

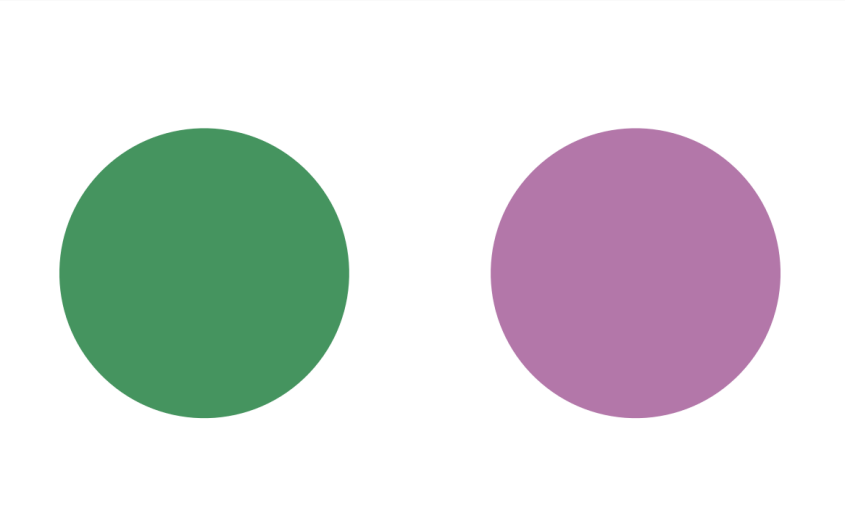

7
6
5
4
3
2
displ
20
30
40
hwy
cty
10
15
20
25
30
35

Color

- Color can be potentially very powerful, but must be used meaningfully
- Be cognizant of:
  - Hue or color choice
  - Intensity or transparency
- Colors can represent different categories of things as well as values

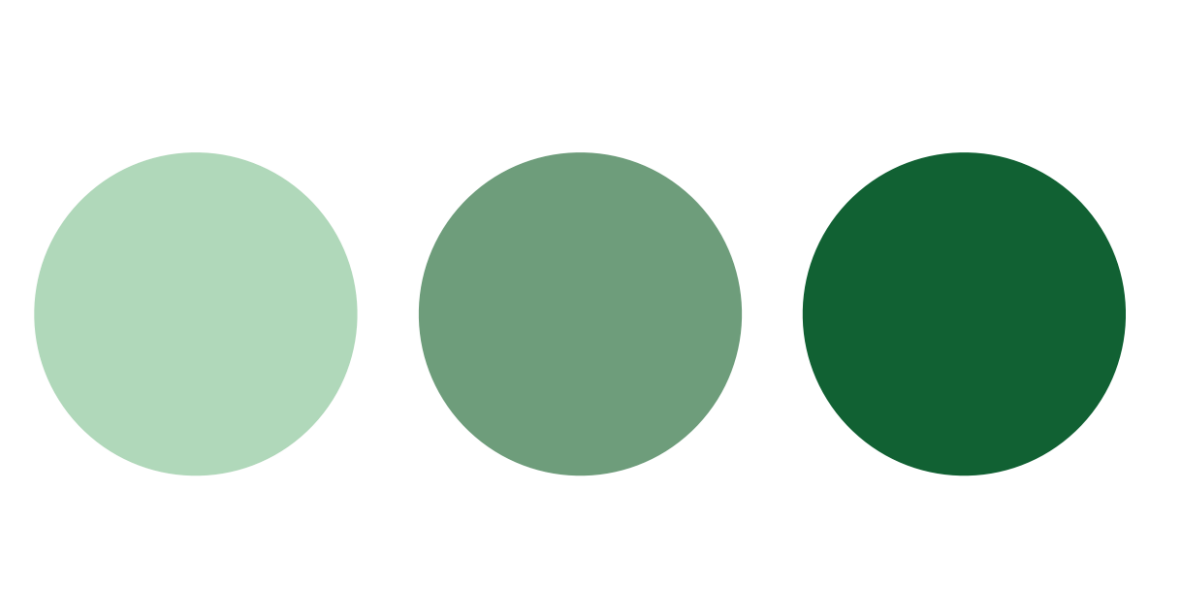
Color

Which represents the larger value?



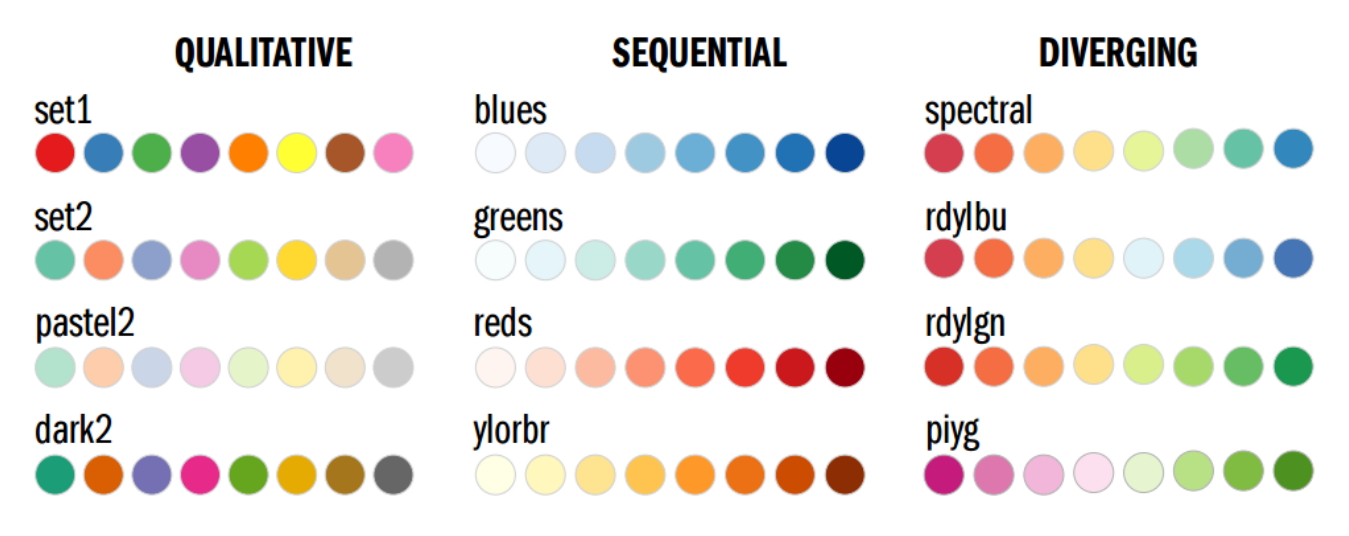
Color

Which represents the larger value?



Color

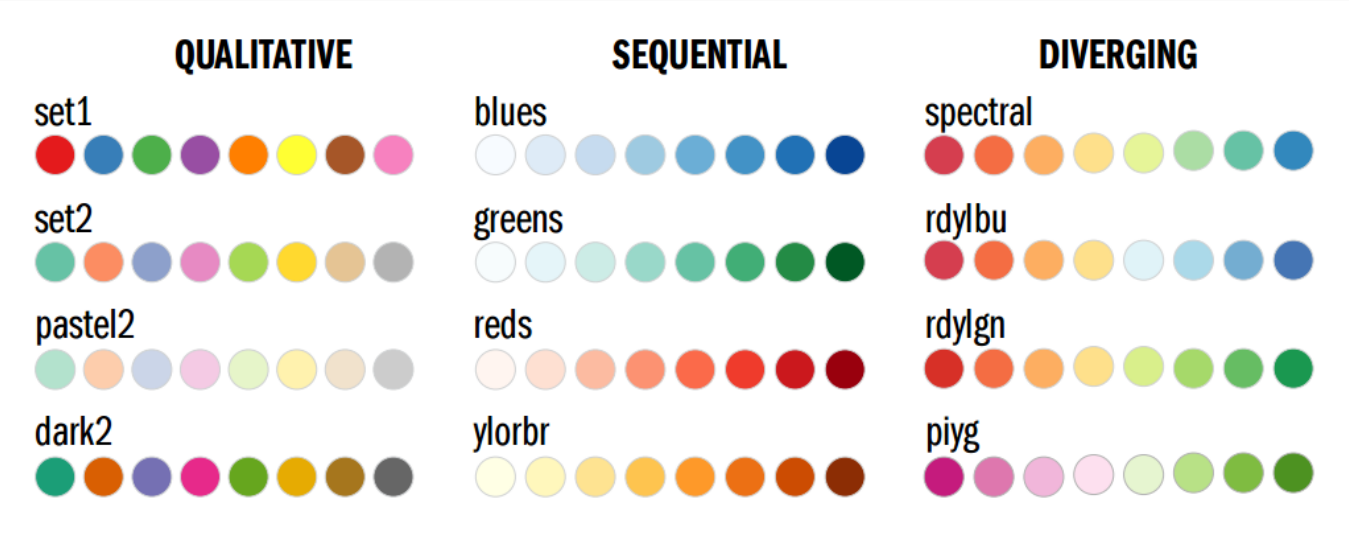
Choose how you use color carefully!



- Other color tips:
  - Use color-blind friendly palettes when possible
  - Don't use too many colors (ideally, no more than 6)

Color

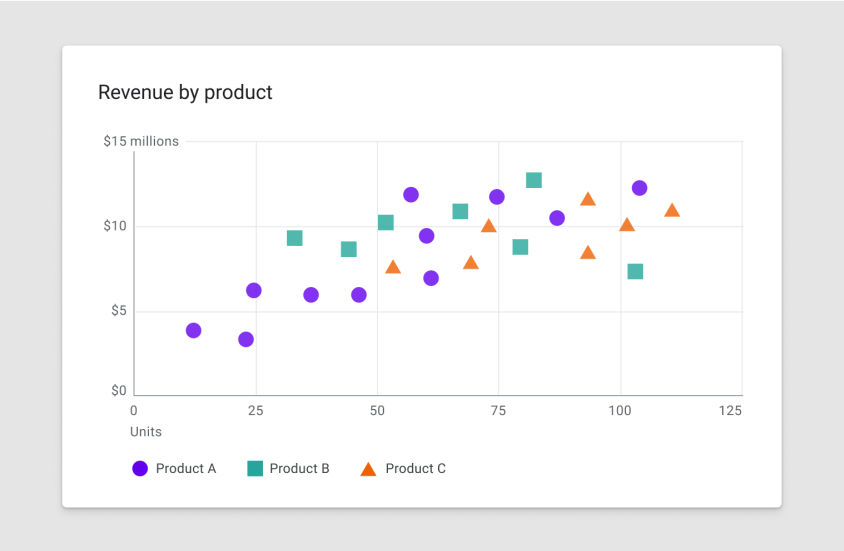
Choose how you use color carefully!

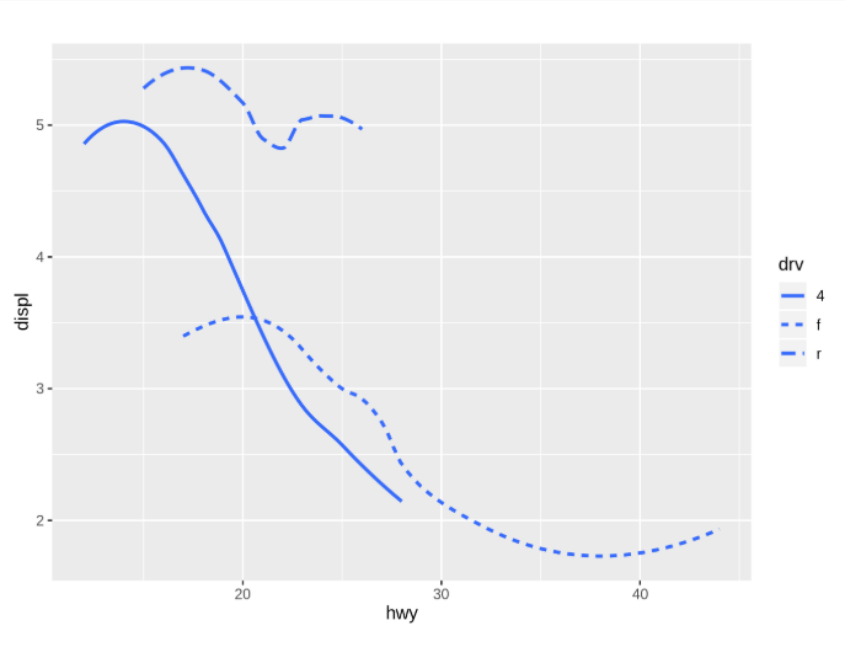


- Other color tips:
  - Use color-blind friendly palettes when possible
  - Don’t use too many colors (ideally, no more than 6)

Shape

- Shapes don't inherently indicate values
  - Therefore, better used to distinguish between categories

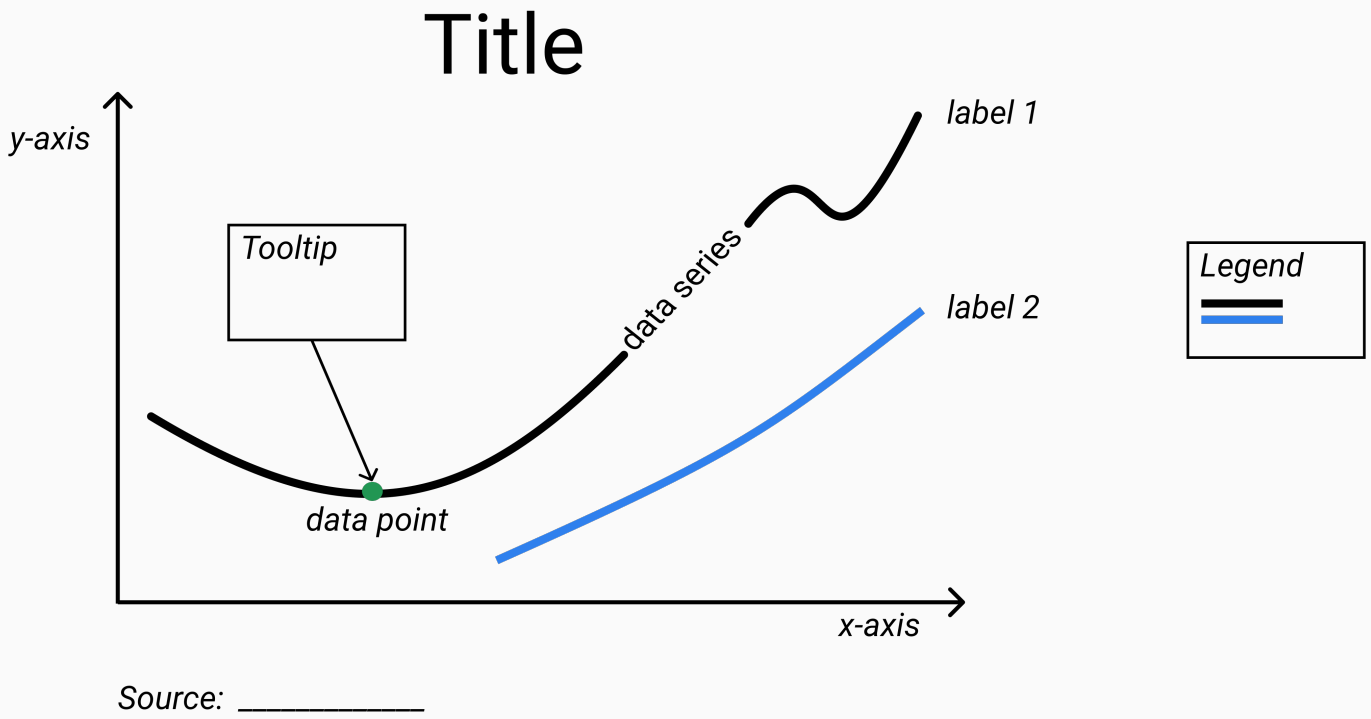




General Guidelines When Creating Data Visualizations

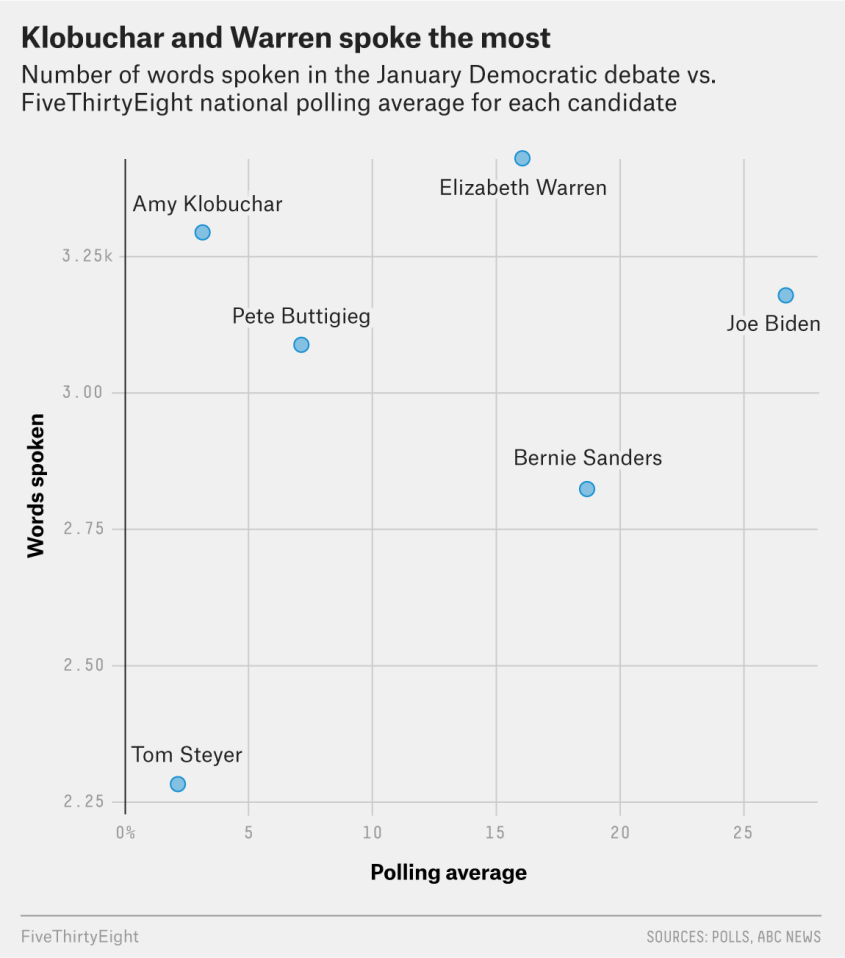
Labels, Titles and Text

There are many ways to use text in a data visualization:



Labels, Titles and Text

- You should always include a title
  - Be descriptive without going too overboard
  - Use the title to say something
  - Use the subtitle to tell people how to read the chart



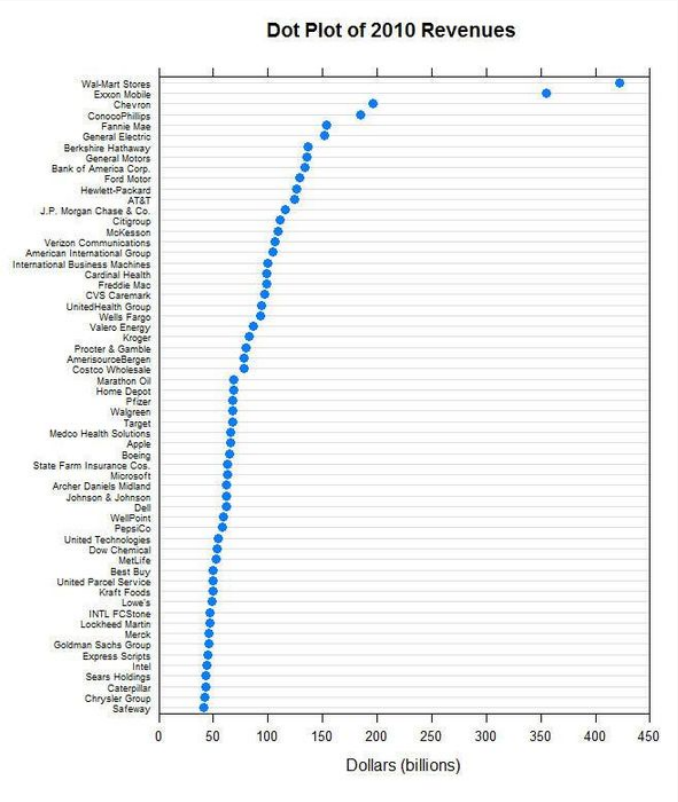
Labels, Titles and Text

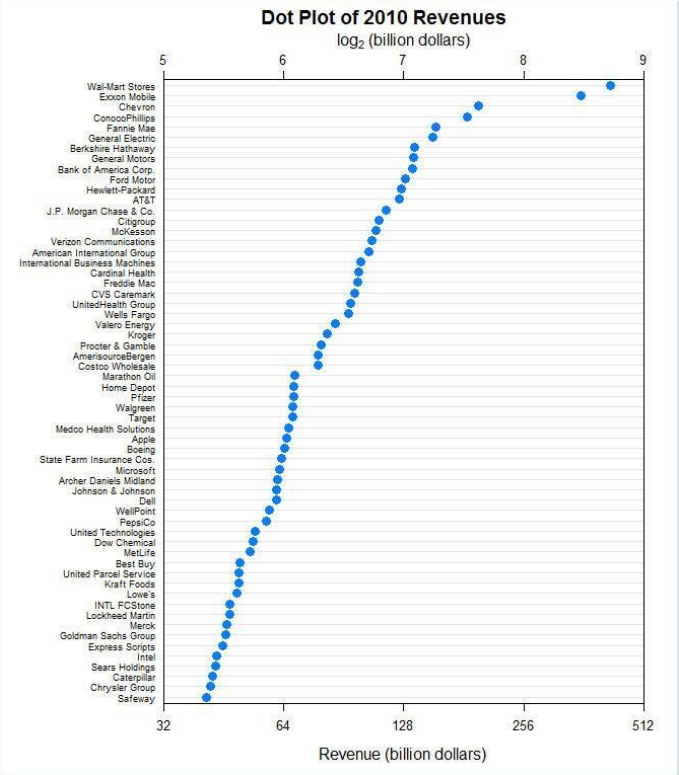
Other notes:

- Label the axes!
- Callouts can be helpful when you want to highlight one data point
- Legends may not be necessary if you use labels wisely
- Footnotes or chartnotes should generally be avoided in presentations, unless absolutely necessary

Data Transformations

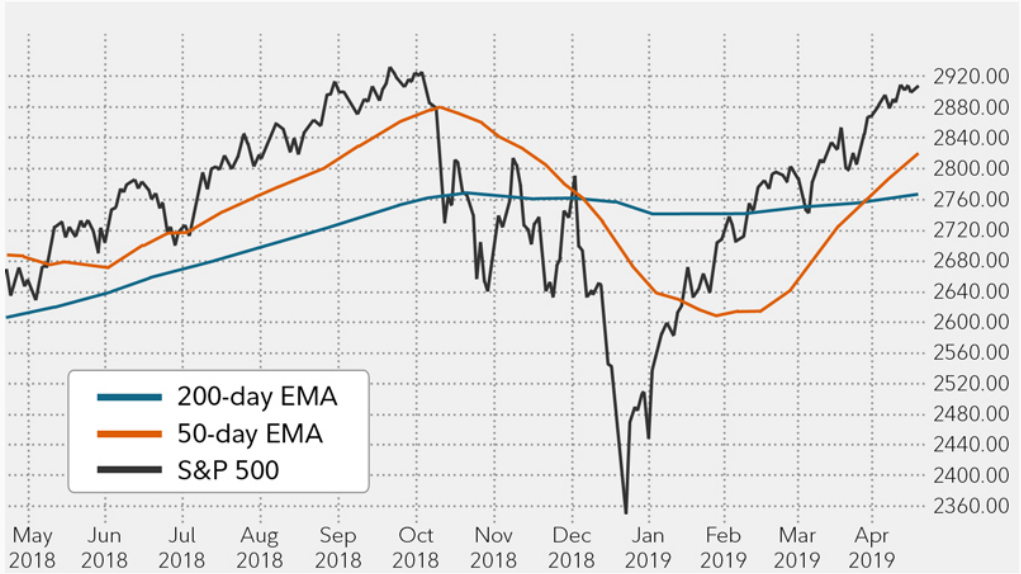
- Sometimes it's necessary to transform data before adding to a chart
  - Outliers or differences can skew the data
  - Example: Common to apply a log transformation for dollar amounts





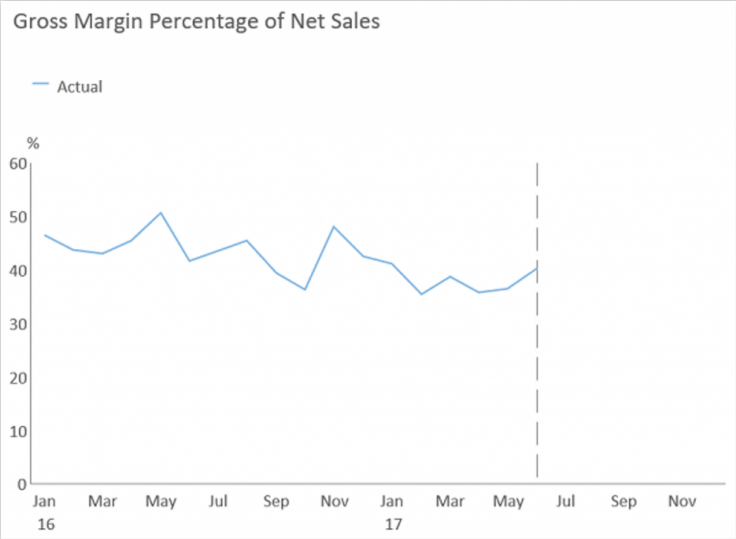
Data Transformations

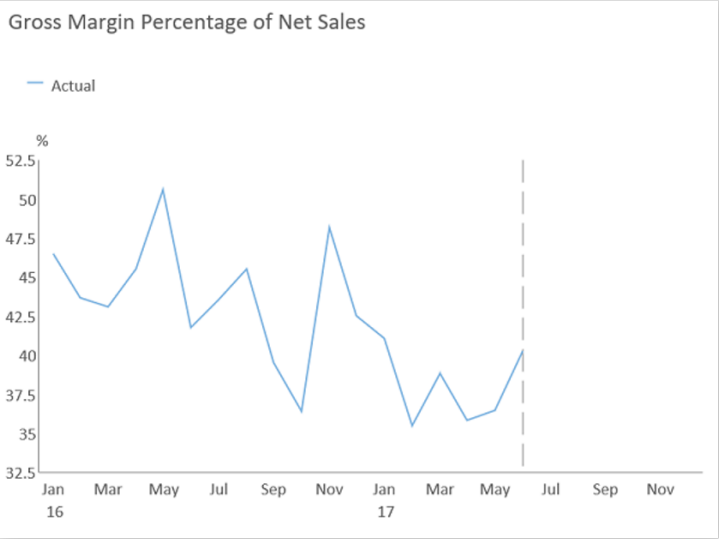
- Sometimes it's necessary to transform data before adding to a chart
  - Too many observations
  - Aggregate in different ways



Axes and Scales

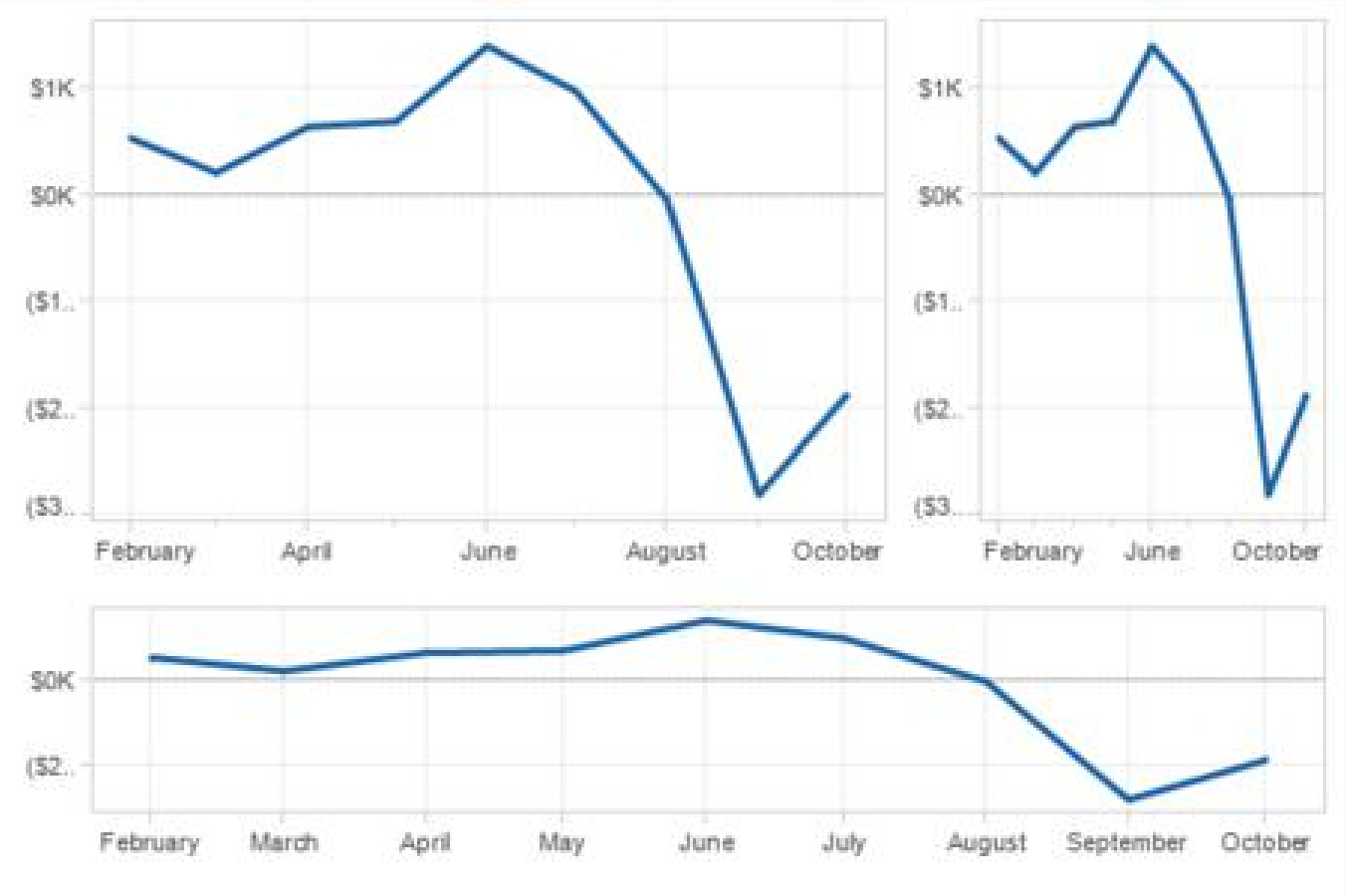
- Use meaningful scales and axis ranges
- Avoid distortions that obscure the message





Axes and Scales

- Similarly, be cognizant of how aspect ratios may obscure patterns



Summary

- What point am I making?
  - Graphics should convey a limited number of things as efficiently as possible
- Is this the appropriate way to show my point?
  - Many different types of visualizations, choose wisely
- Understand how people generally perceive things
  - Shapes, colors, lengths, sizes
- Label things!
  - Use legends or callouts when needed
- Avoid distortions
  - What are the axes and scales?

Summary

- What point am I making?
  - Graphics should convey a limited number of things as efficiently as possible
- Is this the appropriate way to show my point?
  - Many different types of visualizations, choose wisely
- Understand how people generally perceive things
  - Shapes, colors, lengths, sizes
- Label things!
  - Use legends or callouts when needed
- Avoid distortions
  - What are the axes and scales?

Summary

- What point am I making?
  - Graphics should convey a limited number of things as efficiently as possible
- Is this the appropriate way to show my point?
  - Many different types of visualizations, choose wisely
- Understand how people generally perceive things
  - Shapes, colors, lengths, sizes
- Label things!
  - Use legends or callouts when needed
- Avoid distortions
  - What are the axes and scales?

Summary

- What point am I making?
  - Graphics should convey a limited number of things as efficiently as possible
- Is this the appropriate way to show my point?
  - Many different types of visualizations, choose wisely
- Understand how people generally perceive things
  - Shapes, colors, lengths, sizes
- Label things!
  - Use legends or callouts when needed
- Avoid distortions
  - What are the axes and scales?

Summary

- What point am I making?
  - Graphics should convey a limited number of things as efficiently as possible
- Is this the appropriate way to show my point?
  - Many different types of visualizations, choose wisely
- Understand how people generally perceive things
  - Shapes, colors, lengths, sizes
- Label things!
  - Use legends or callouts when needed
- Avoid distortions
  - What are the axes and scales?

Resources

- Tools for data visualizations:
  - Tableau Data Visualization beginner's guide
  - Visualizations using Python
  - Guide to ggplot2
- More on guidelines and do's/don'ts:
  - https://datajournalism.com/read/handbook/one/
  - https://datavizcatalogue.com/
  - *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics* By: Nathan Yau
  - Any book by Edward Tufte
- Inspiration:
  - http://www.visualcomplexity.com/vc/
  - https://flowingdata.com/
  - https://informationisbeautiful.net/

Final Thoughts

- Make something that is memorable...
- ...BUT be honest and credible
  - Your data analysis is not going to be perfect.
  - That's okay - be honest about it
  - Talk about caveats and explain uncertainty

Thanks!