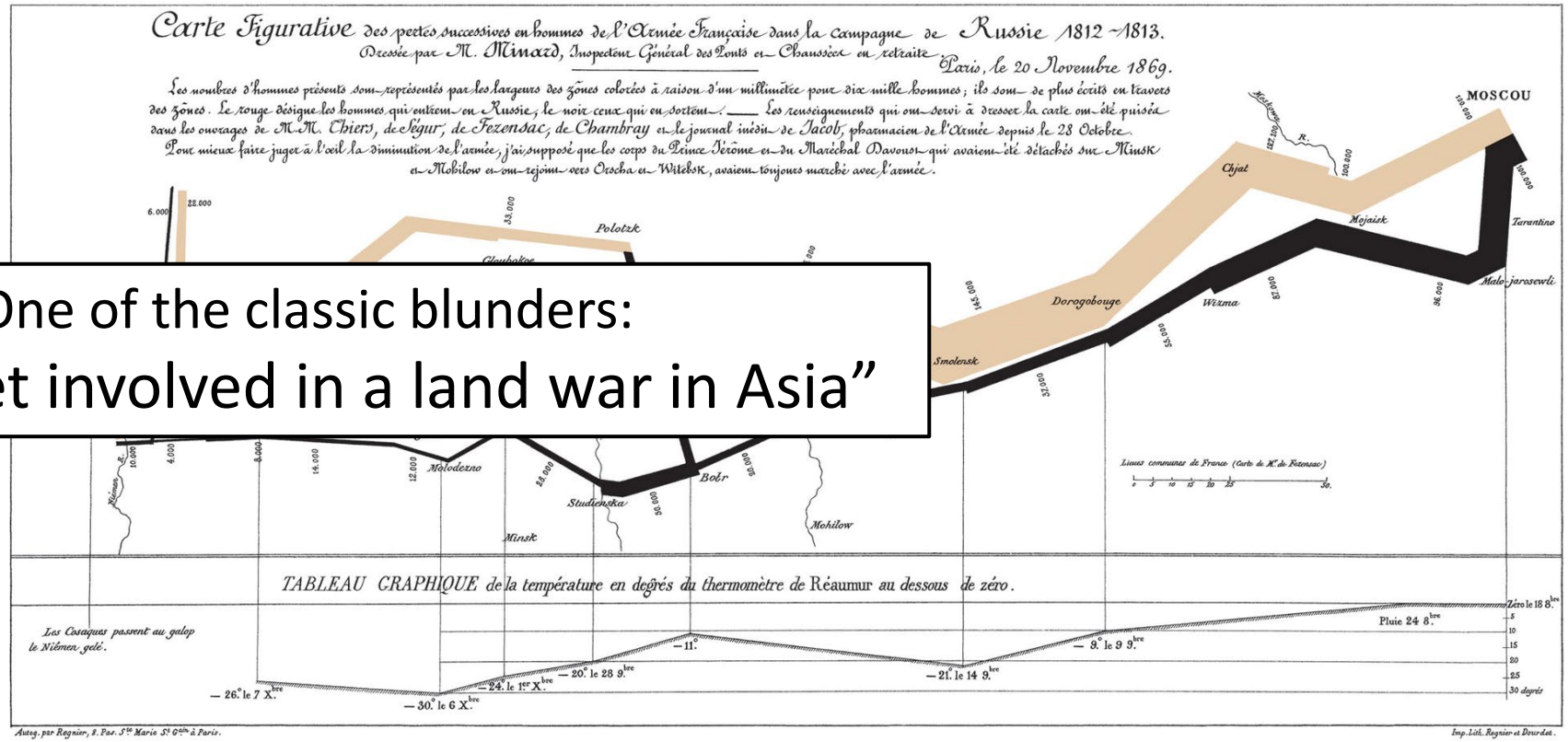
Effective Visualization I

Lab 4



One of the classic blunders:

"Never get involved in a land war in Asia"

Charles Joseph Minard, Paris, 1869

Agenda

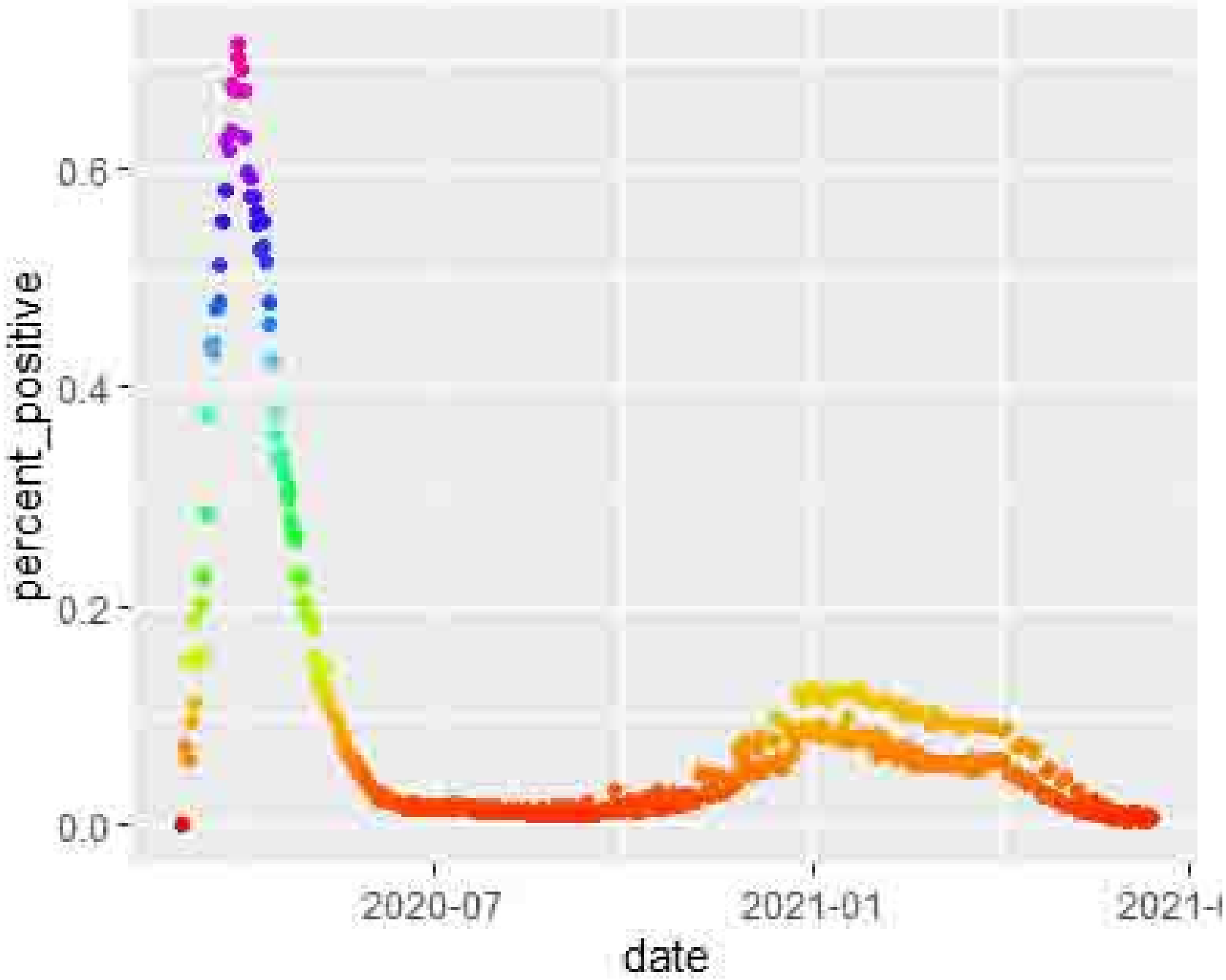
- Principles of effective visualization
- Color and color scales

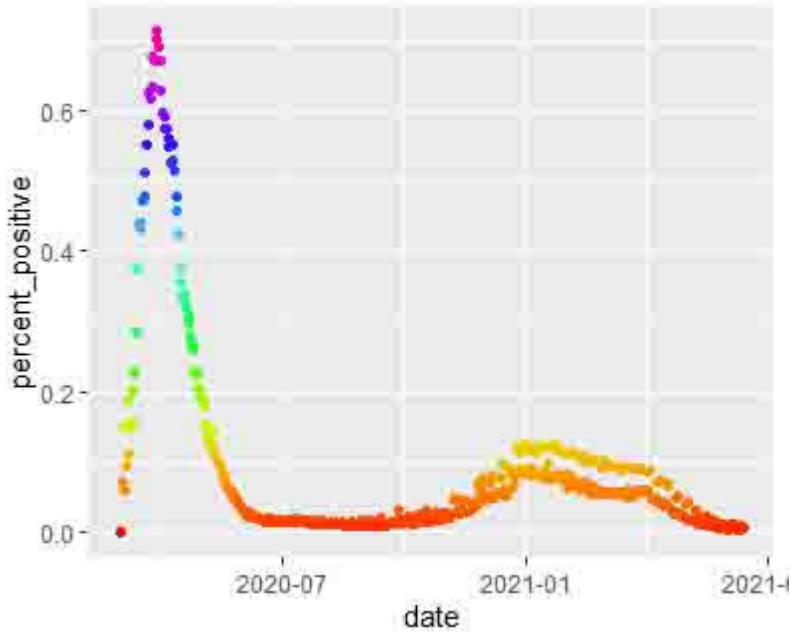
- ggplot intro

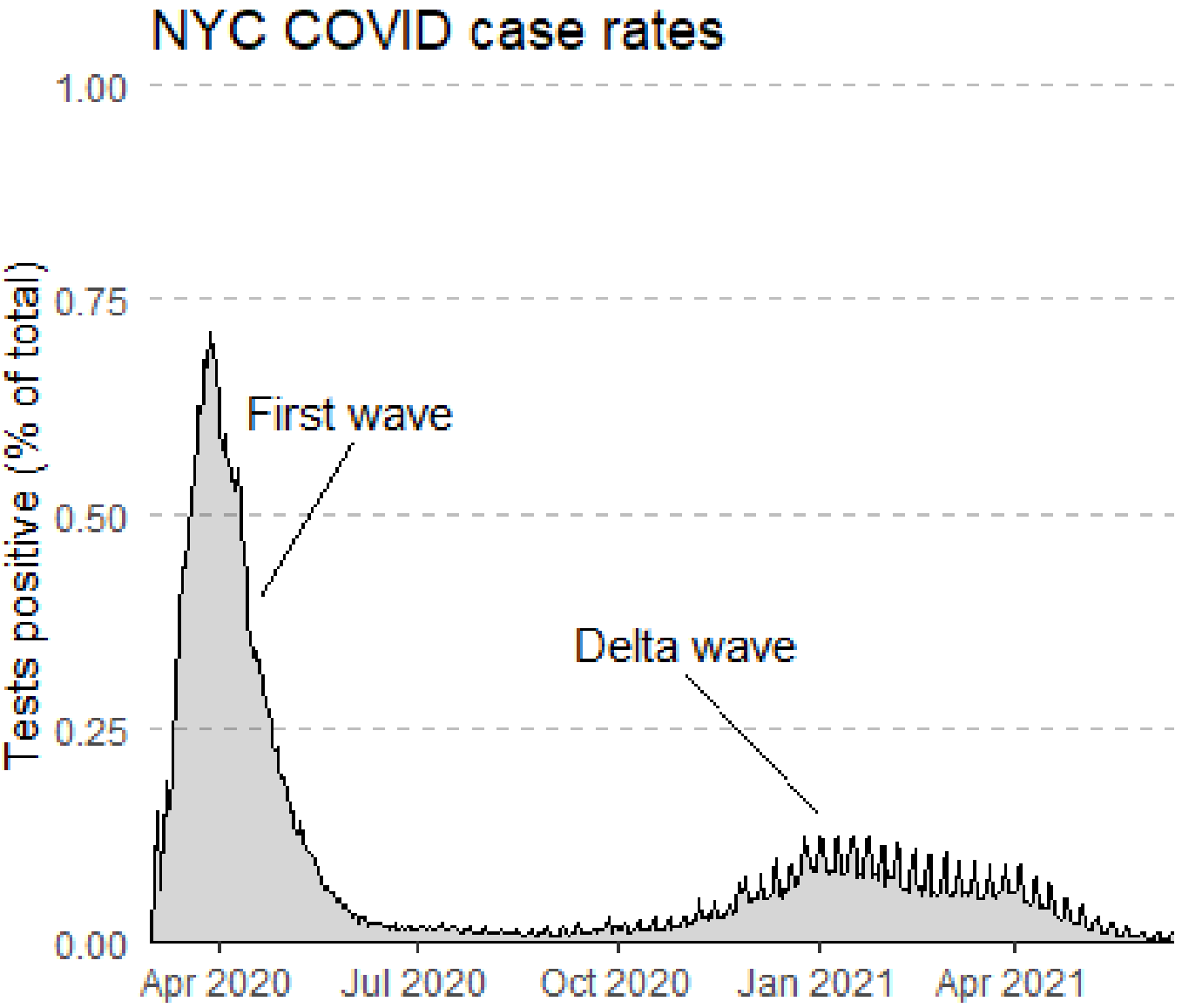
- Next time:
  - Image files (raster and vector, resolution)
  - Putting data on maps

Recommended resources

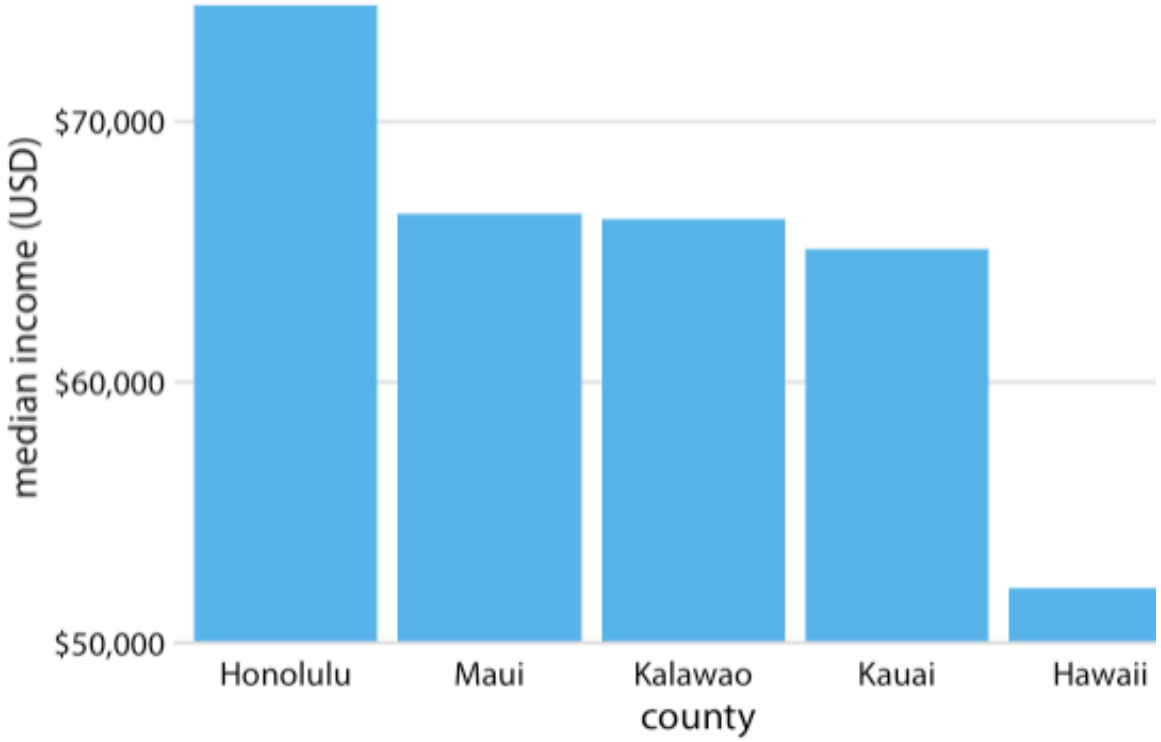
Claus Wilke's Fundamentals of Data Visualization	design concepts
Hadley Wickham *et al.*'s ggplot2 Book	R graphing philosophy
Winston Chang's R Graphics Cookbook	"how do I make this plot?"
ggplot2 Cheat Sheet	

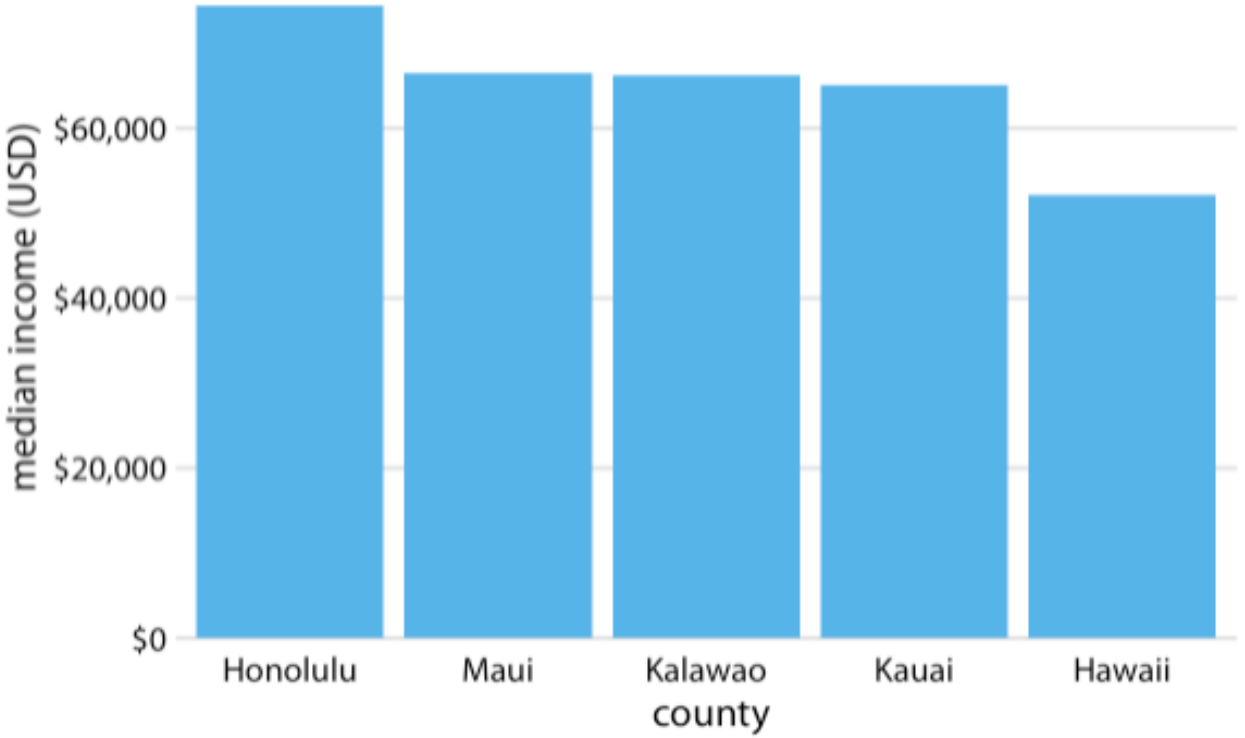

percent_positive
0.6
0.4
0.2
0.0
2020-07
2021-01
date


percent_positive
0.6
0.4
0.2
0.0
2020-07
2021-01
2021-
date

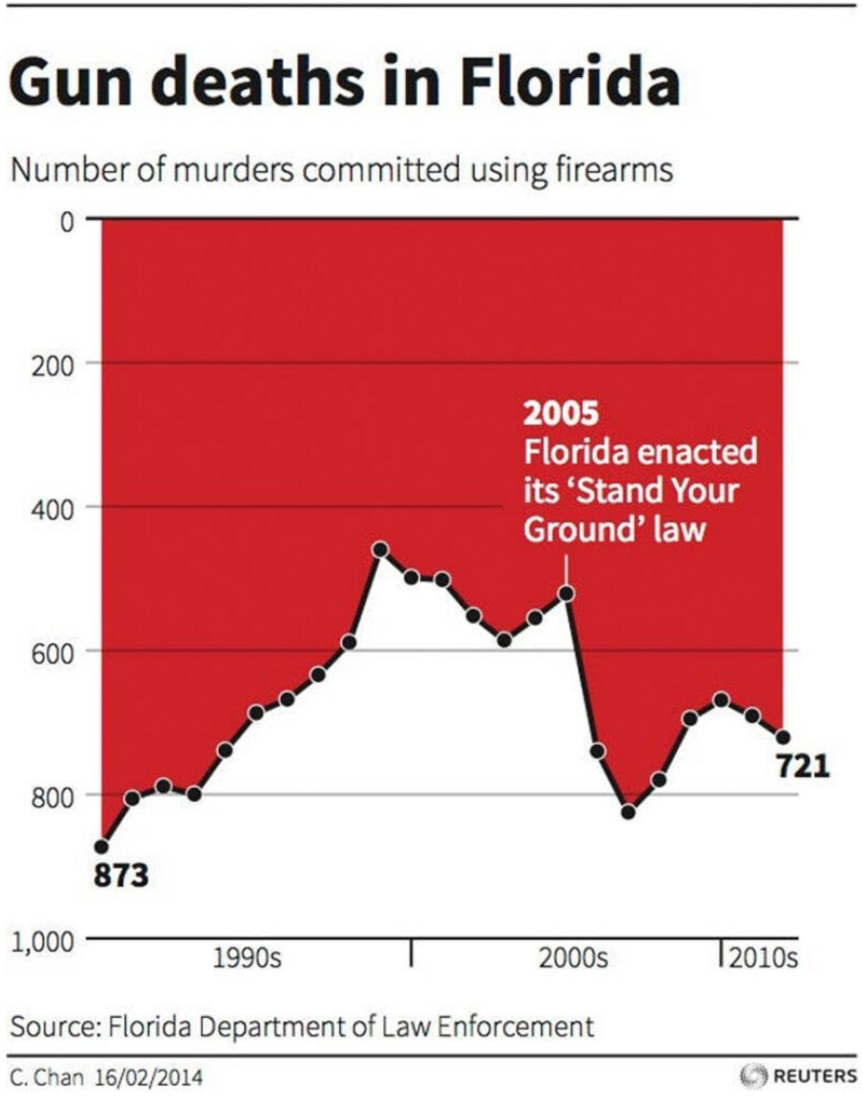

NYC COVID case rates
1.00
0.75
0.50
0.25
0.00
Tests positive (% of total)
First wave
Delta wave
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Apr 2021

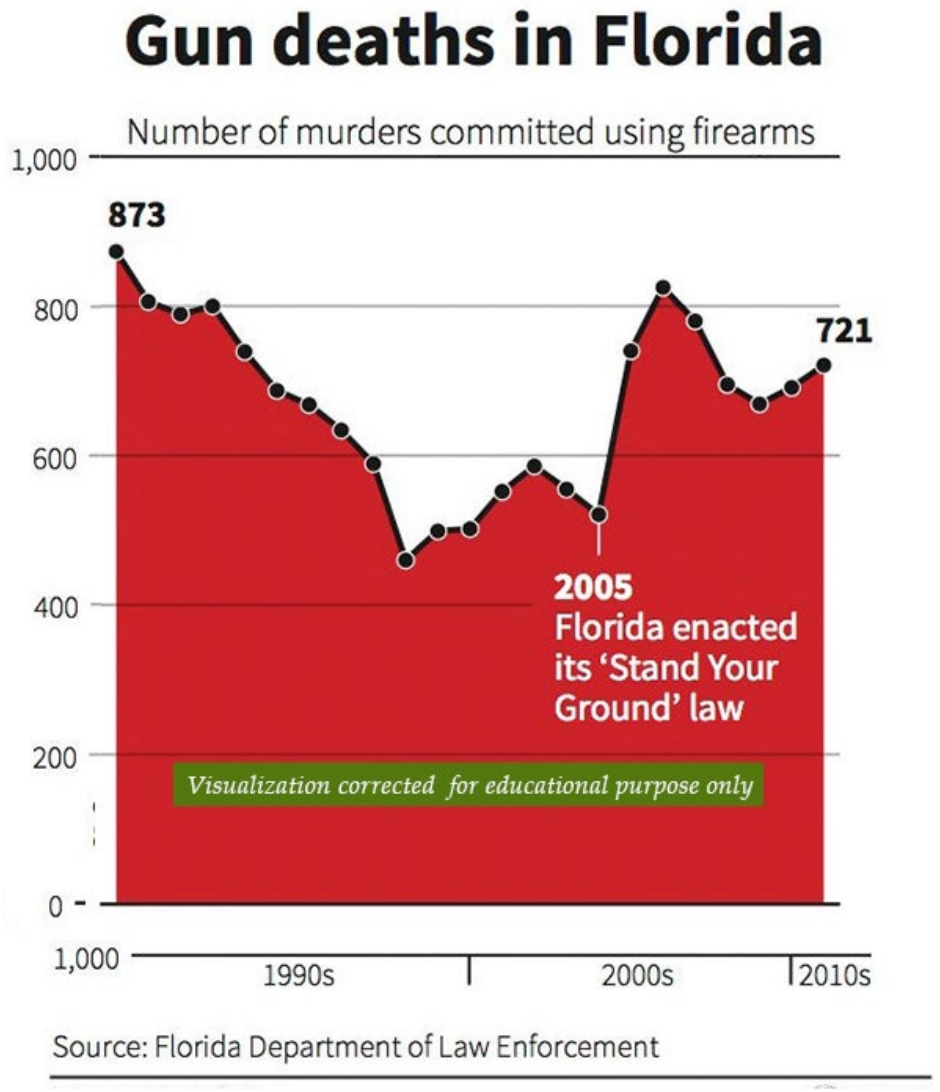
Tell the truth





Tell the truth





Tell the truth

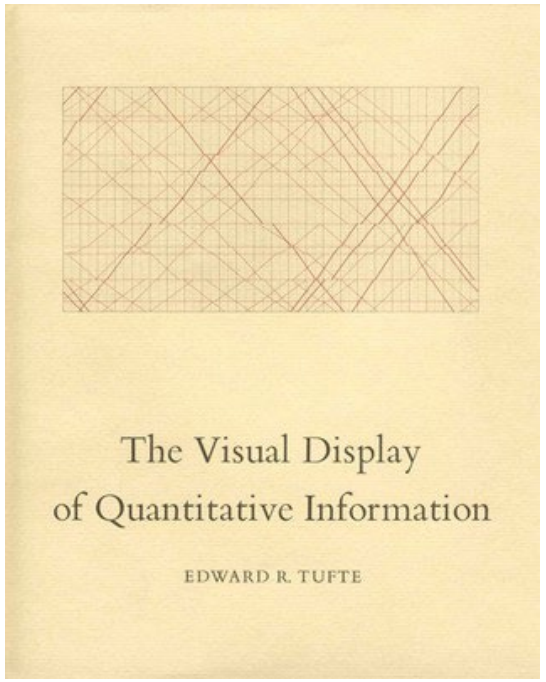
[The Economist owns up to visualization mistakes]

- Poor carpenters blame their tools

> Instead of plotting the individual polls with a smoothed curve to show the trend, we connected the actual values of each individual poll. This happened, primarily, because our in-house charting tool does not plot smoothed lines. Until fairly recently, we were less comfortable with statistical software (like R) that allows more sophisticated visualisations.

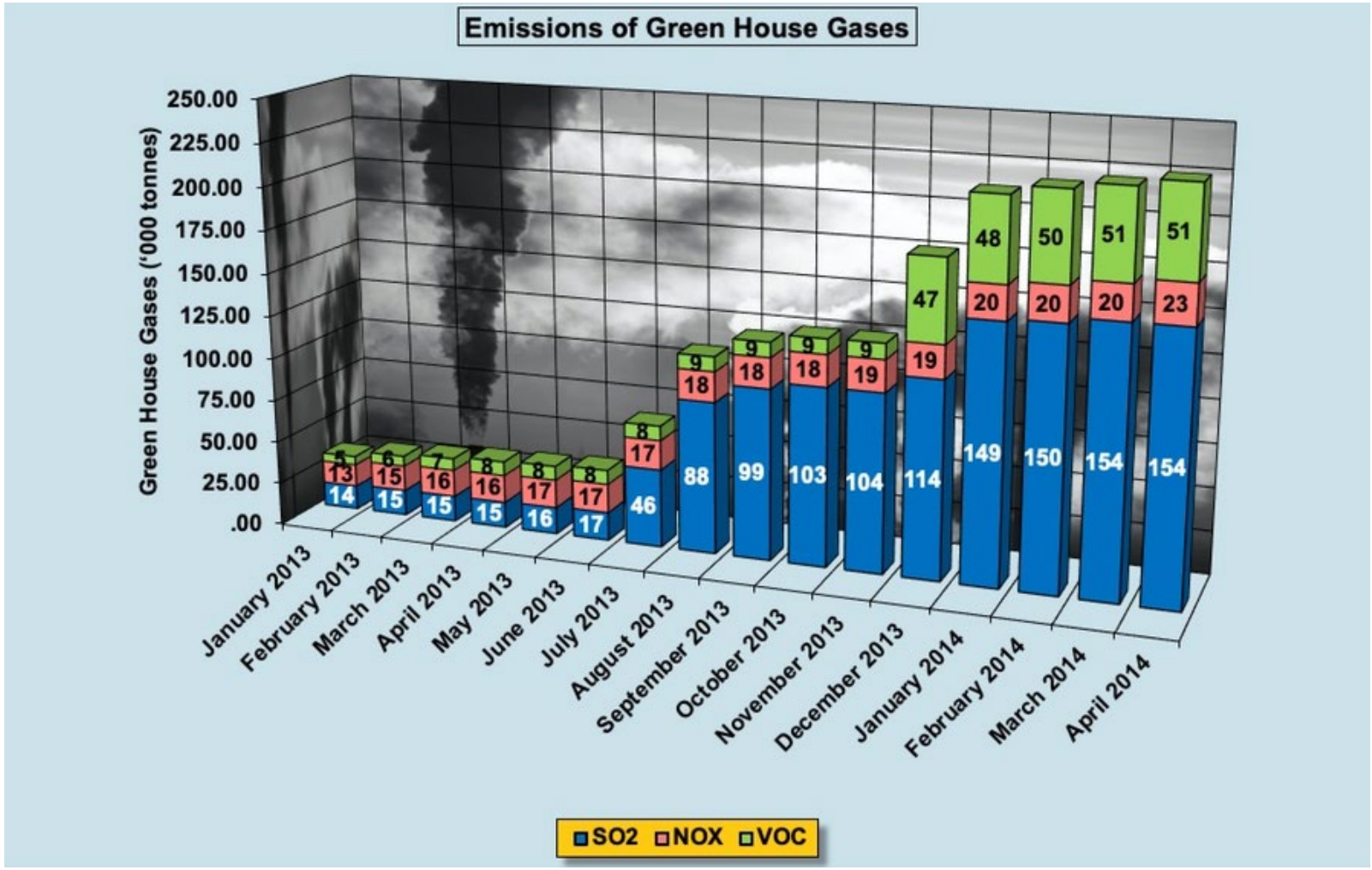
Data-ink ratio and chartjunk

$$\textbf{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

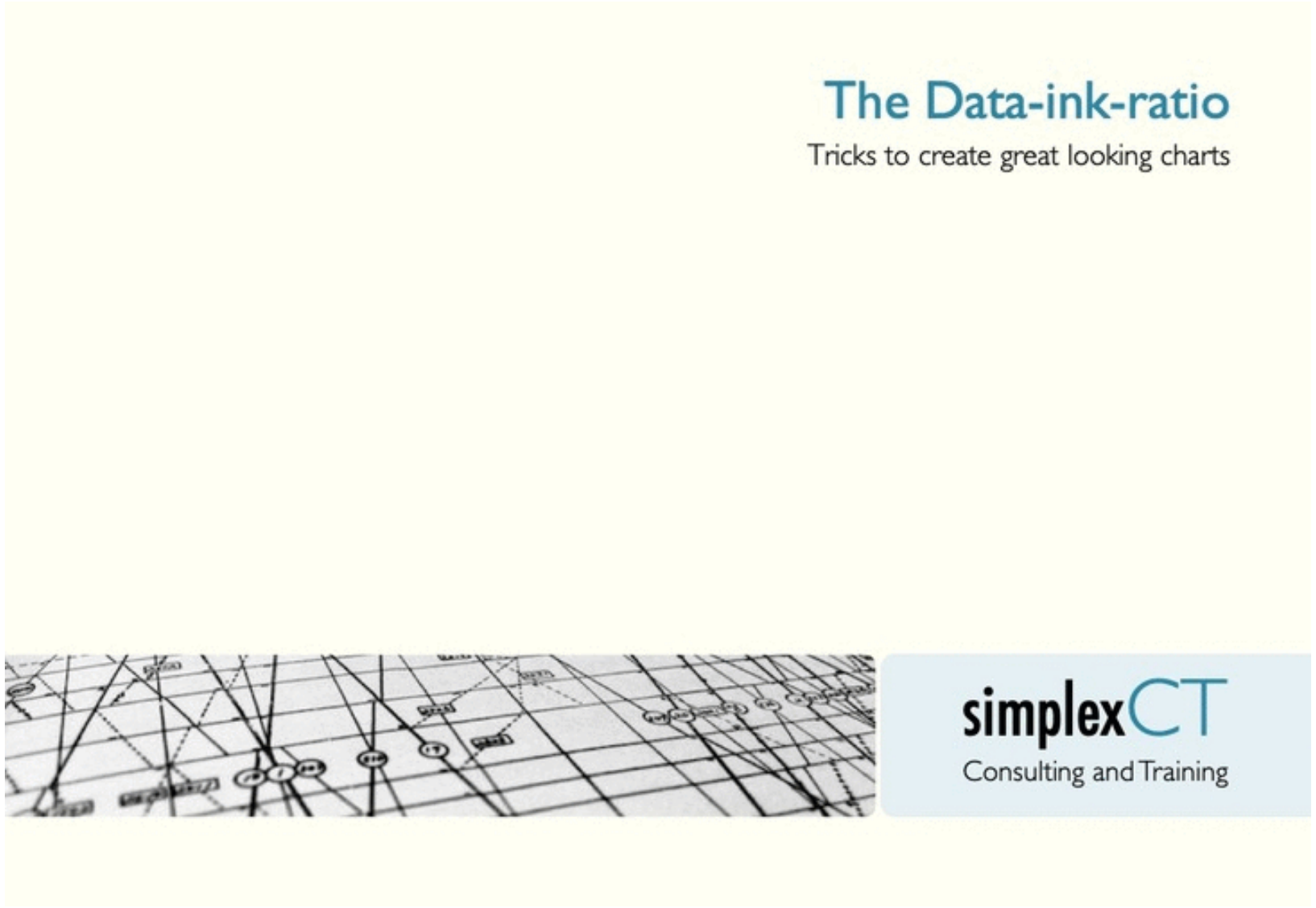


= 1 – proportion of a graphic **that can be erased**

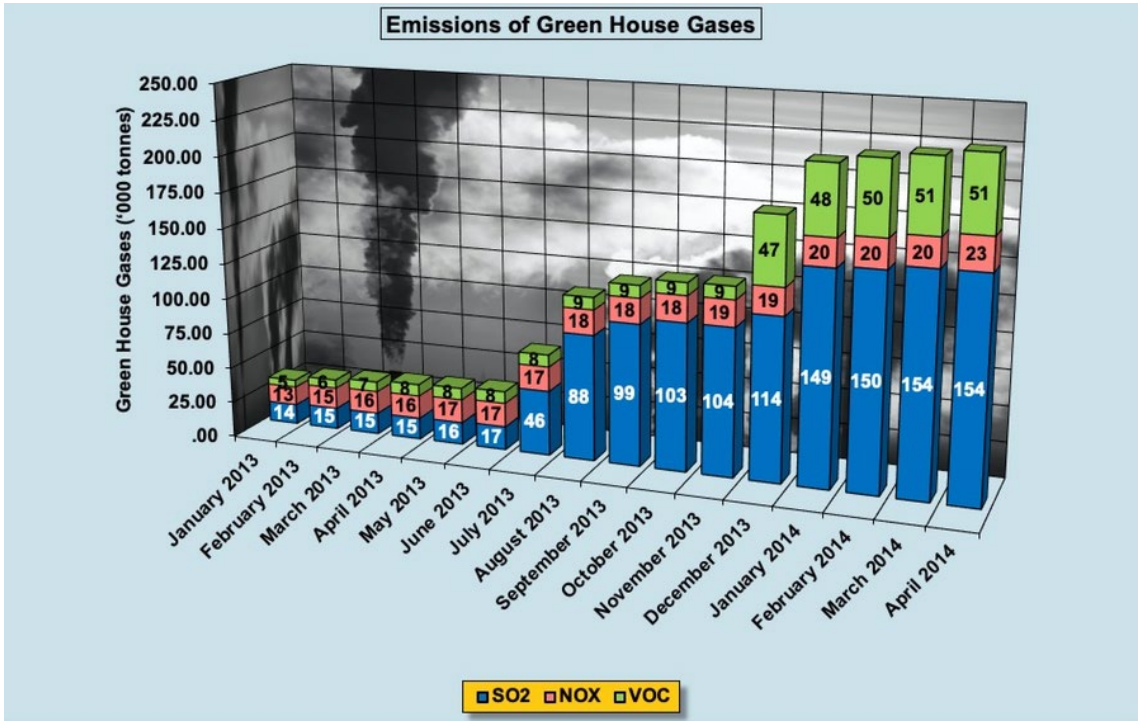
Data-ink ratio and chartjunk

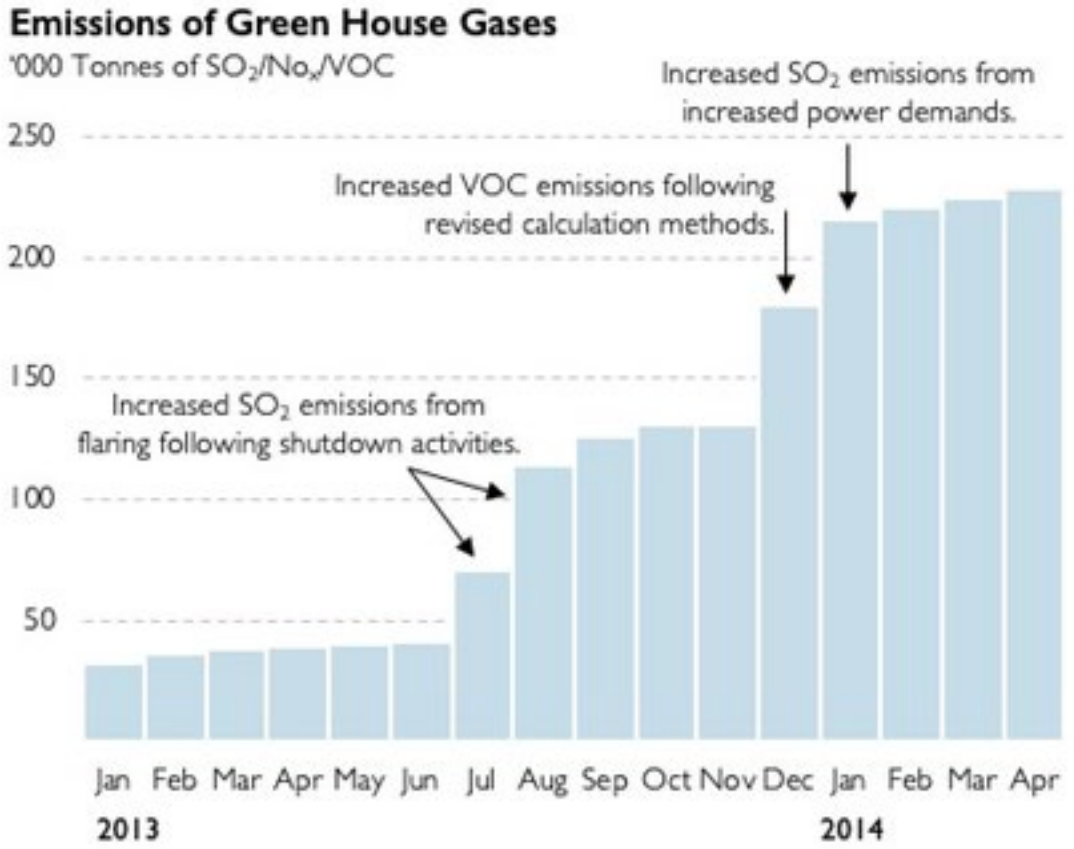


Data-ink ratio and chartjunk

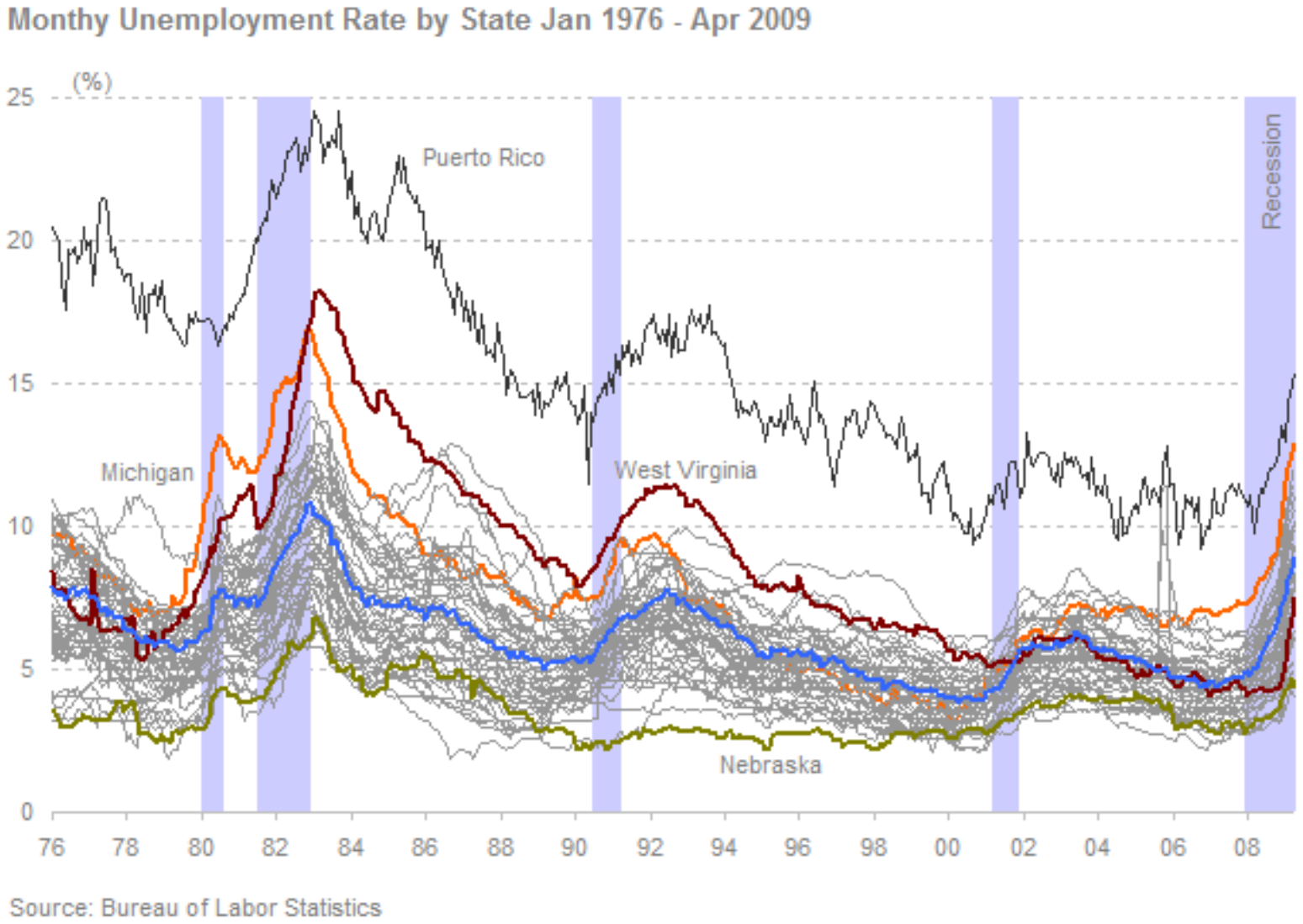


Data-ink ratio and chartjunk

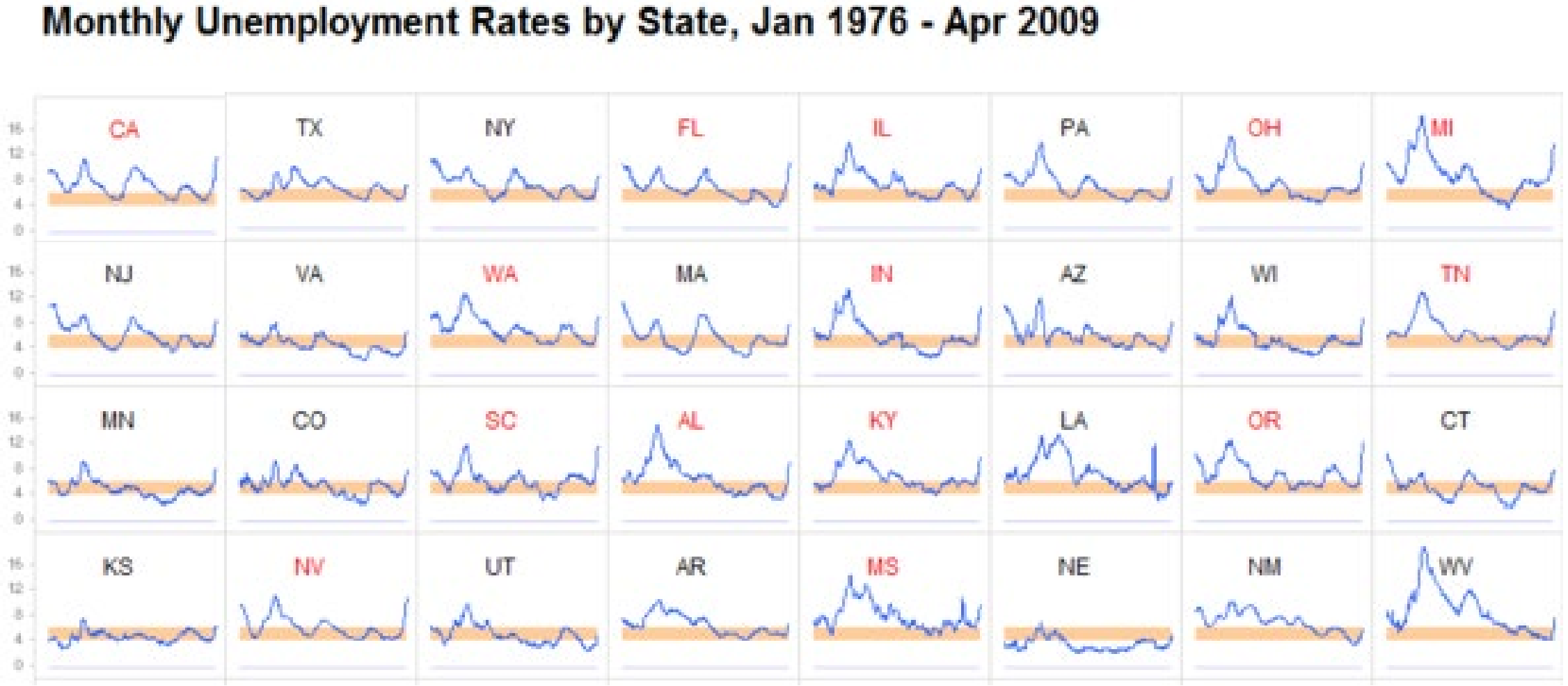




Multivariate data – small multiples

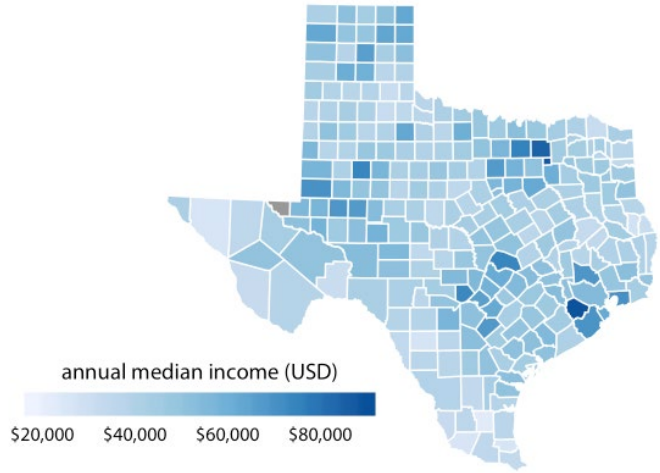


Multivariate data – small multiples



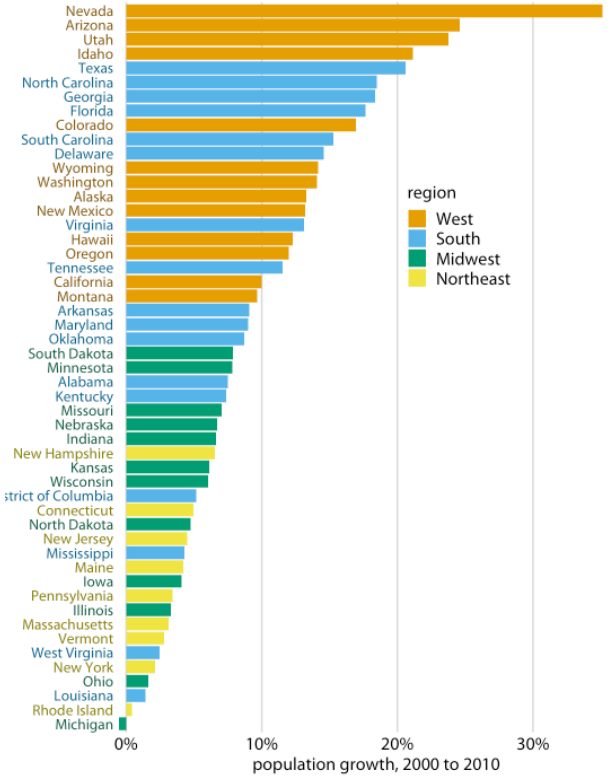
When to use color?

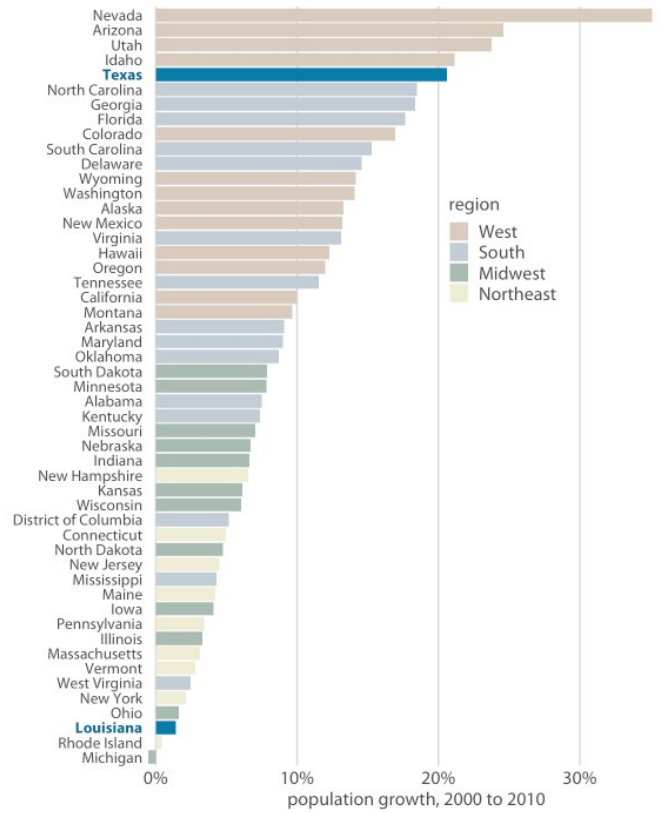
- Will my figure **benefit** from color?



Use color to:

- Distinguish groups
- Represent data values
- Emphasize visual elements

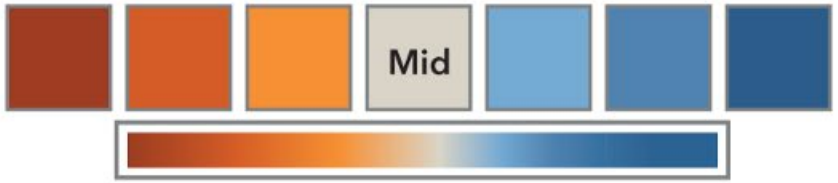




Types of color scales

SEQUENTIAL
color is ordered from low to high

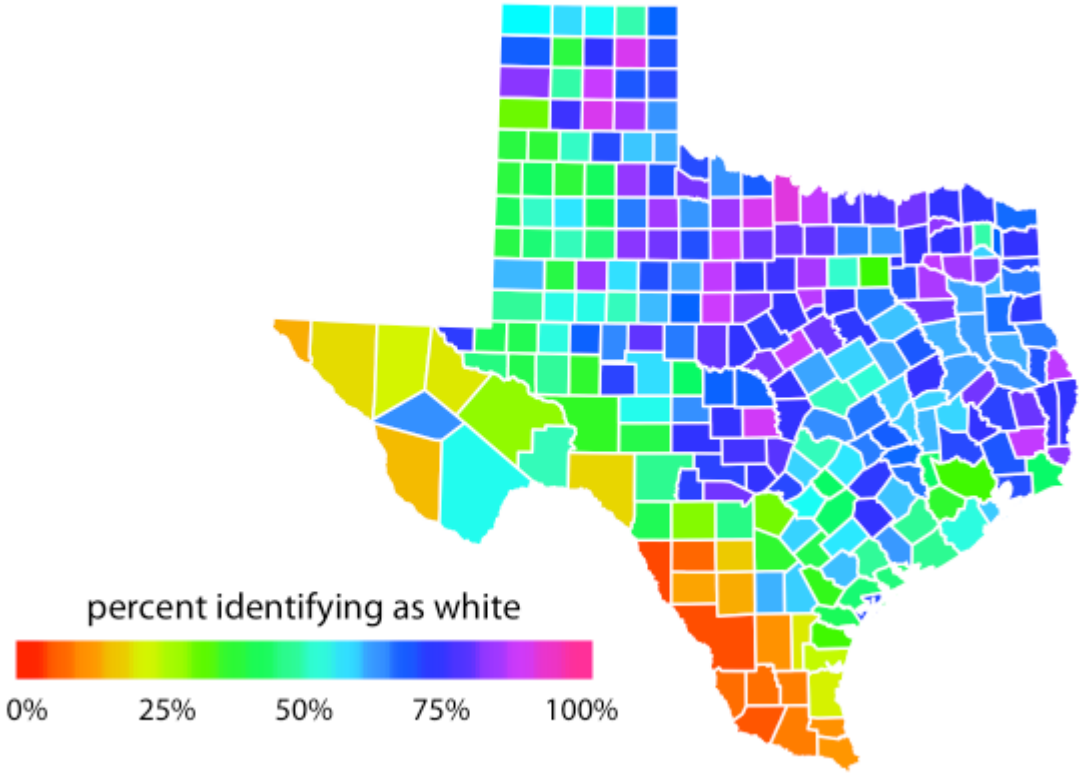
DIVERGING
two sequential colors with a neutral midpoint

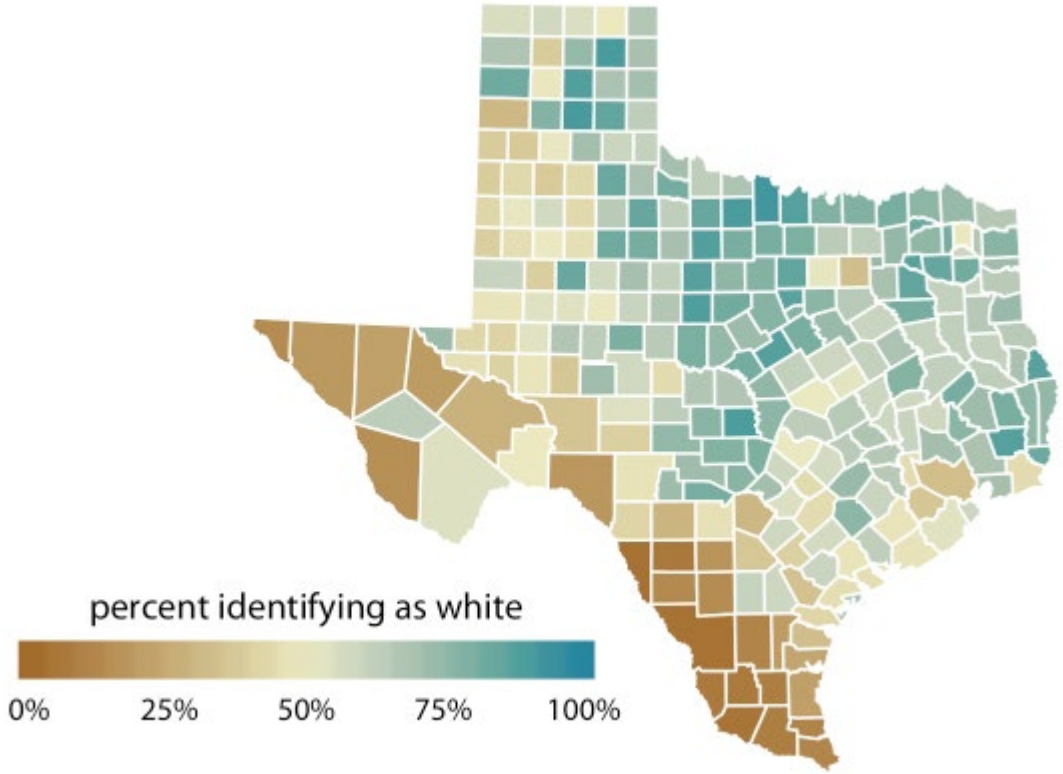


CATEGORICAL
contrasting colors for individual comparison

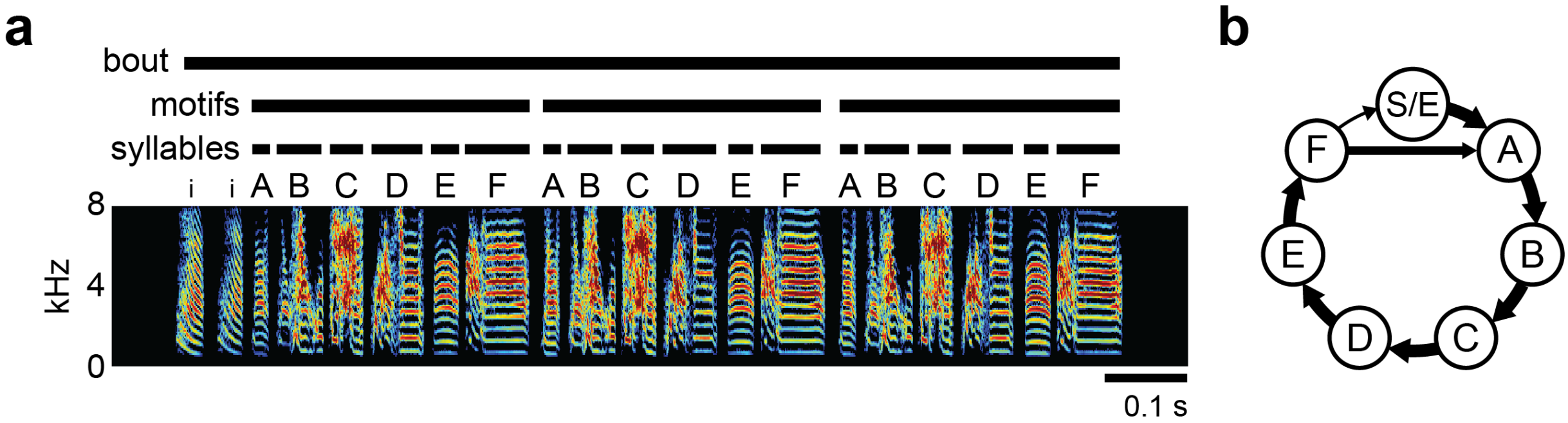
HIGHLIGHT
color used to highlight something

Beware the rainbow





Beware the rainbow (unless your boss wants it)



Color and palette resources

[Top R Color Palettes to Know]

[Kenneth Moreland's Color Advice (highly recommended!)]