

Effective Visualization I

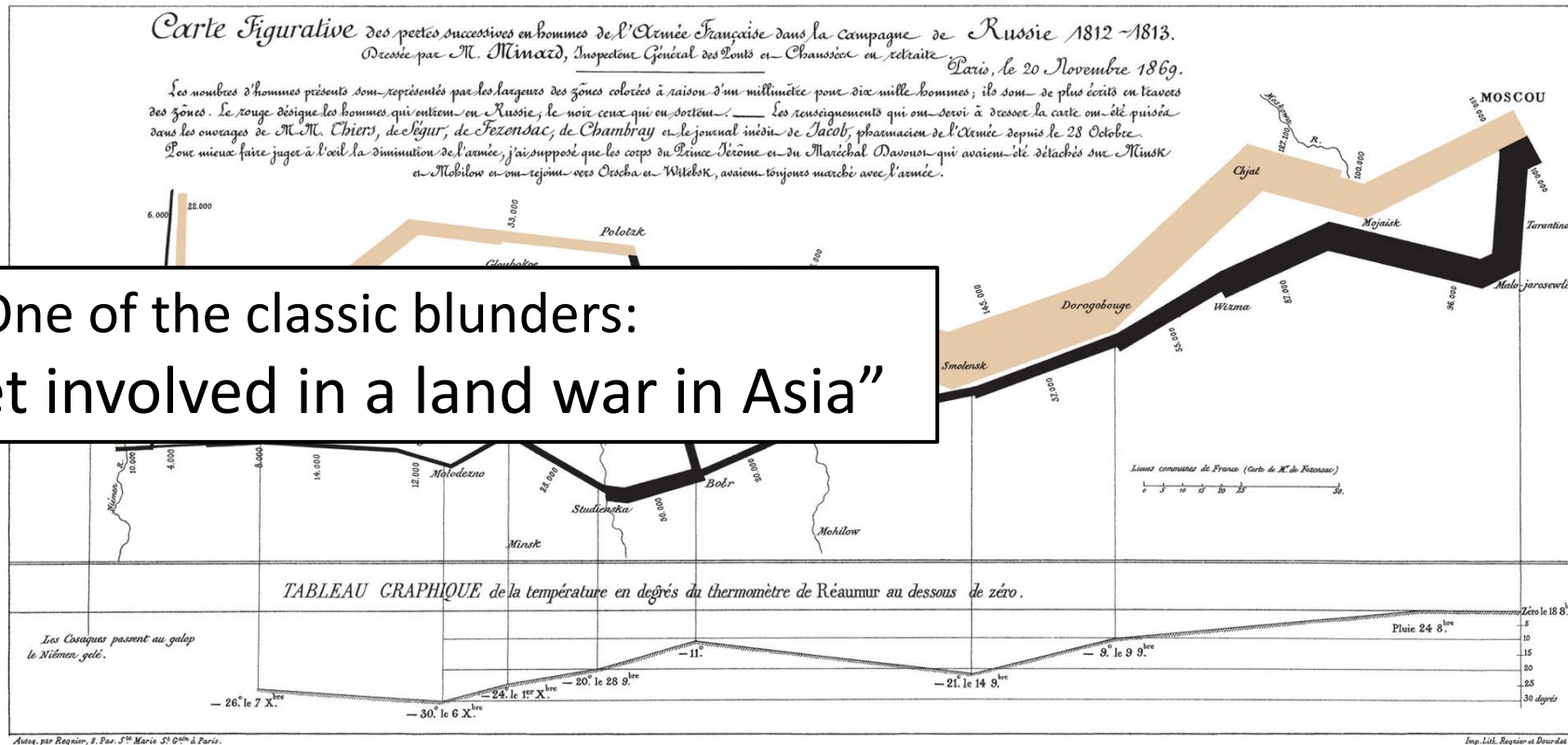
Lab 4

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui ont été en Russie; le noir ceux qui en sont sortis. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Légar, de Fozendac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mohilow et qui rejoignent vers Orscha et Witebsk, avaient toujours marché avec l'armée.

One of the classic blunders:

“Never get involved in a land war in Asia”



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

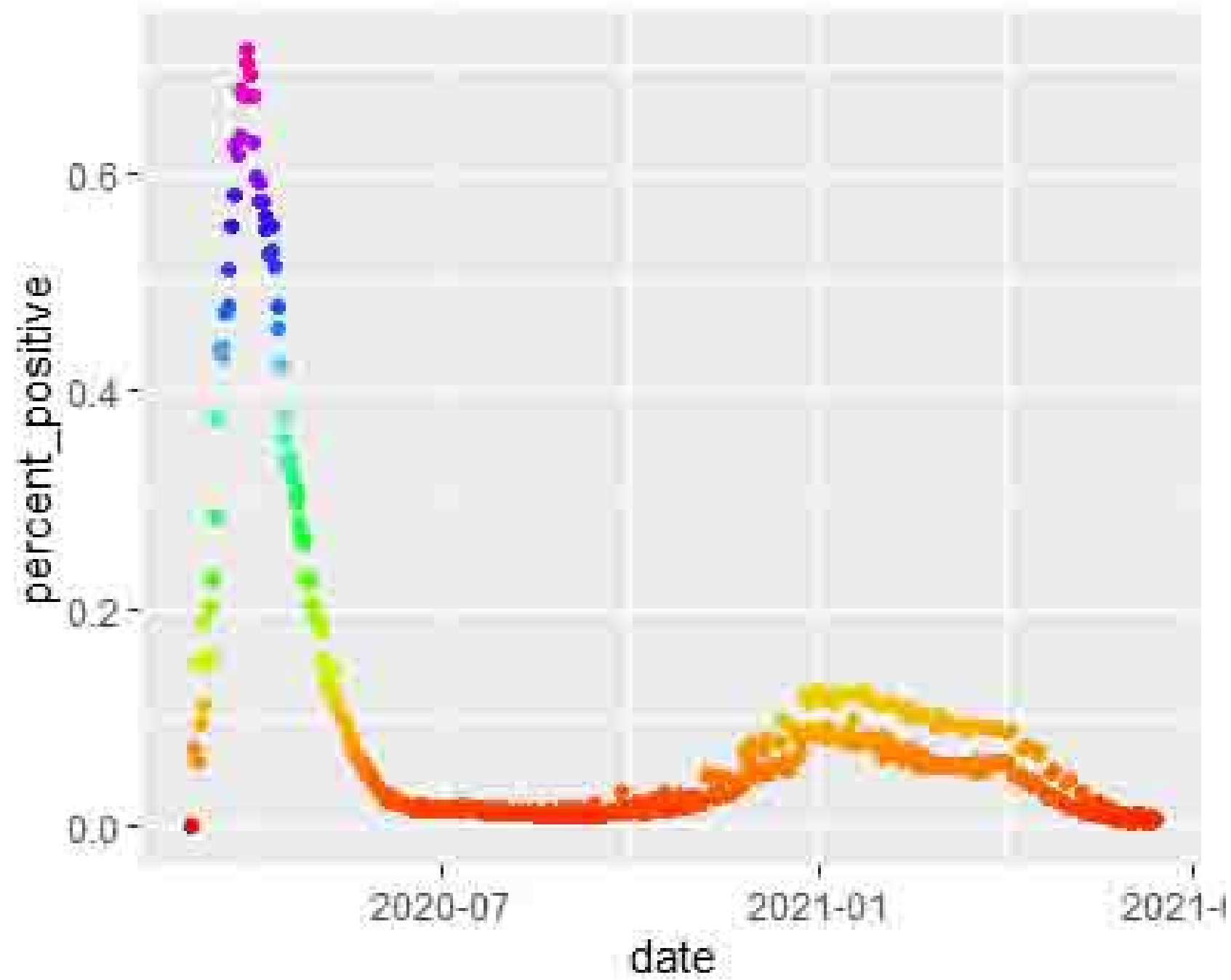
[Hadley Wickham *et al.*'s ggplot2 Book](#)

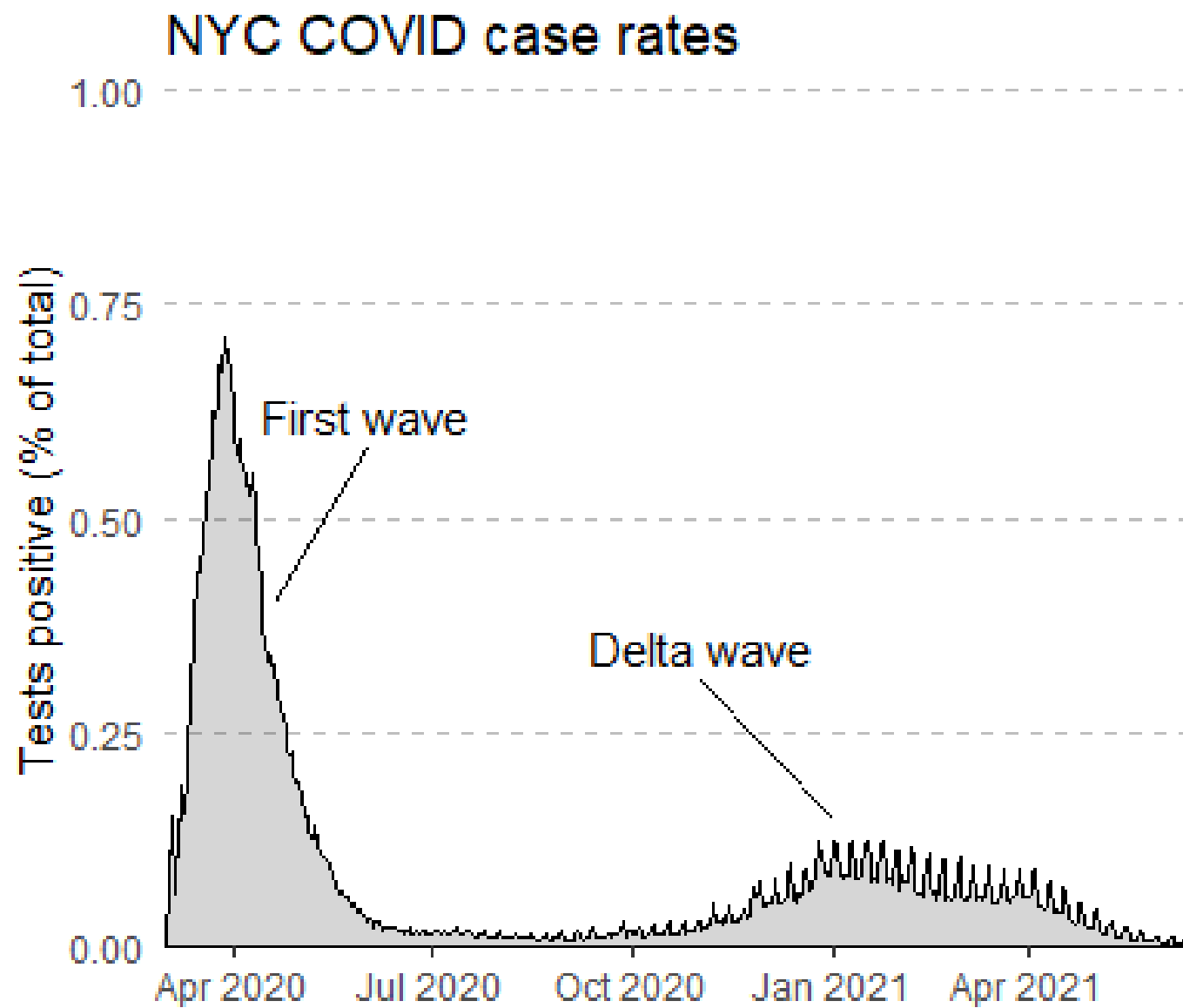
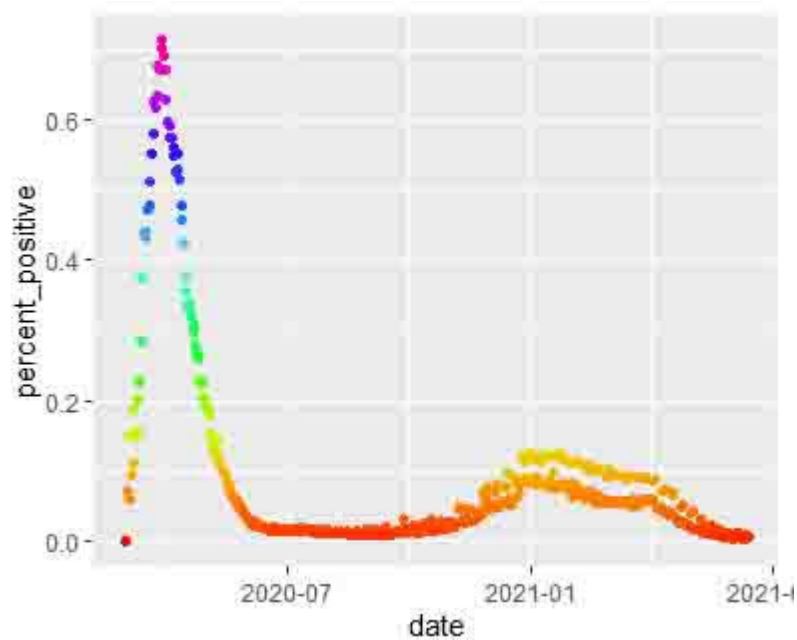
R graphing philosophy

[Winston Chang's R Graphics Cookbook](#)

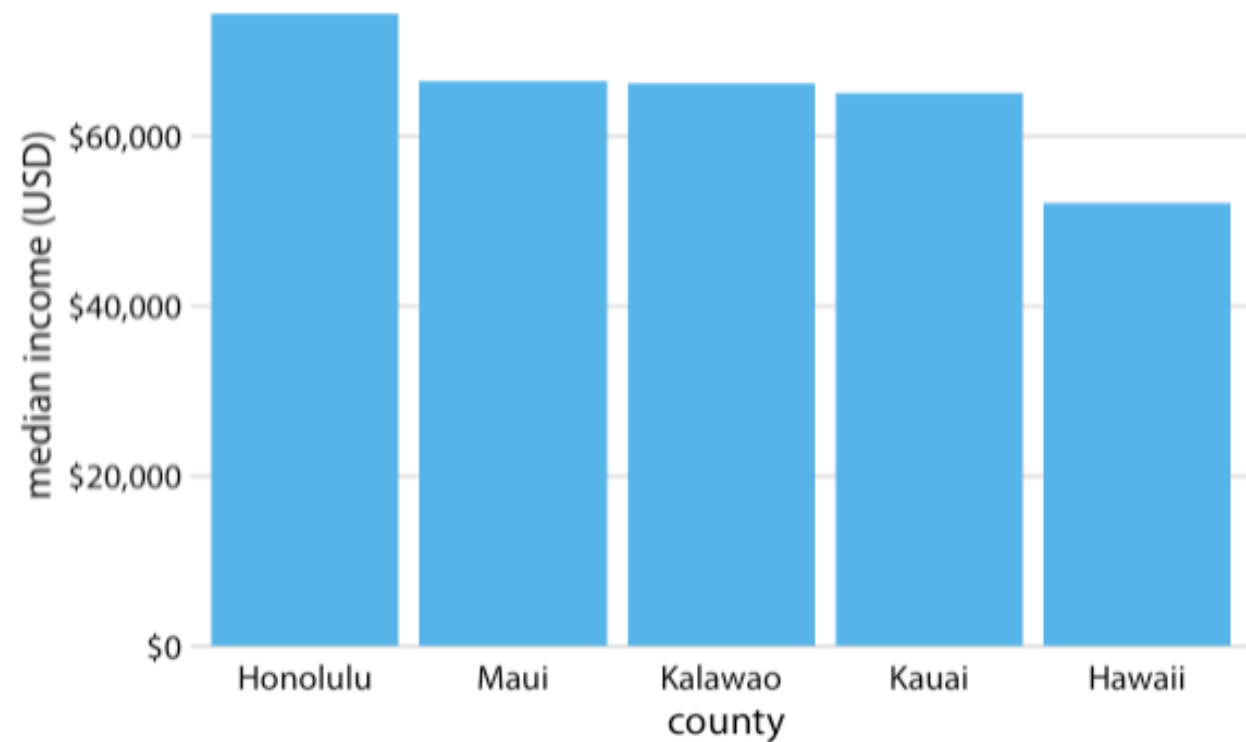
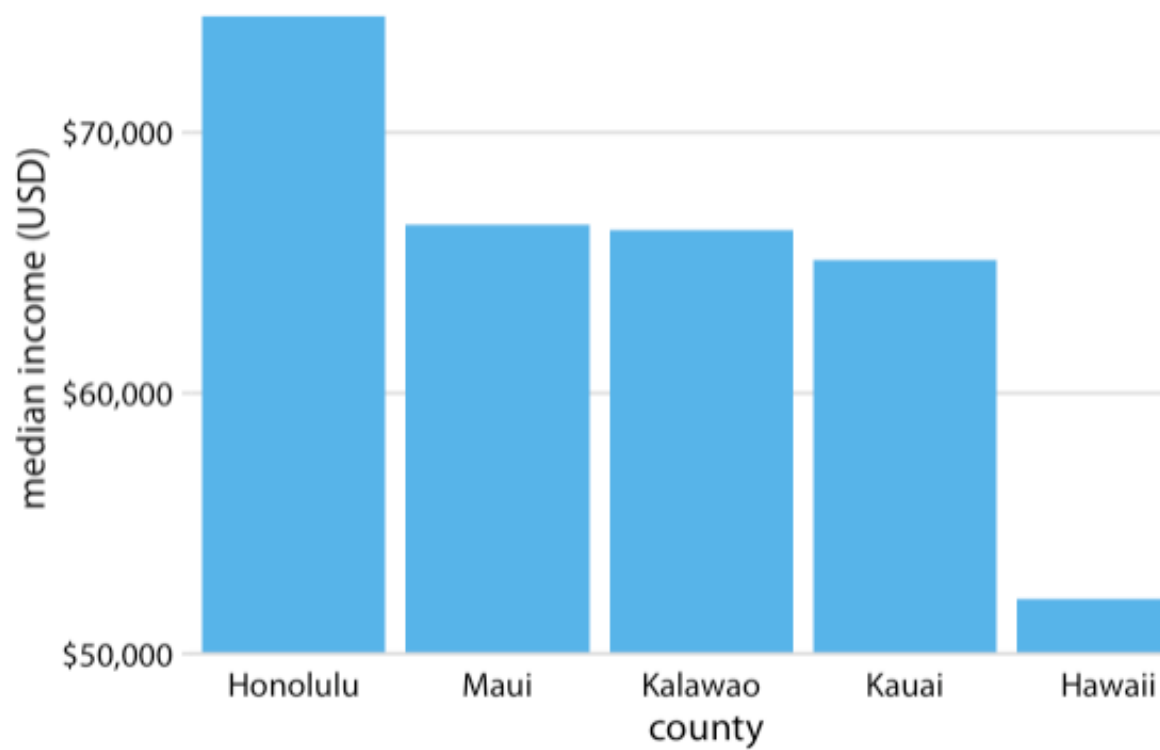
“how do I make this plot?”

[ggplot2 Cheat Sheet](#)





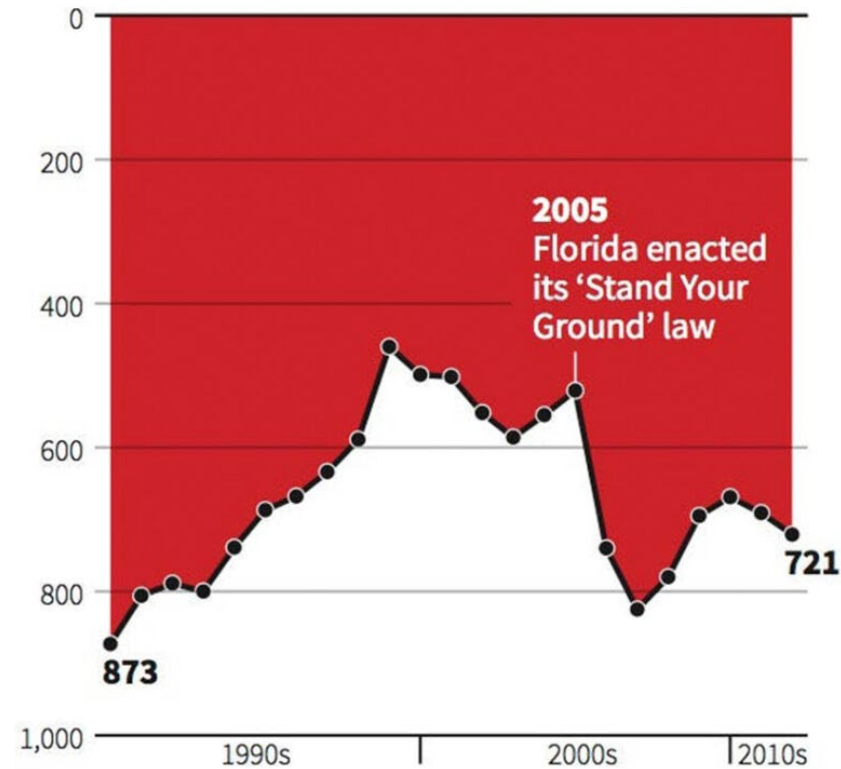
Tell the truth



Tell the truth

Gun deaths in Florida

Number of murders committed using firearms



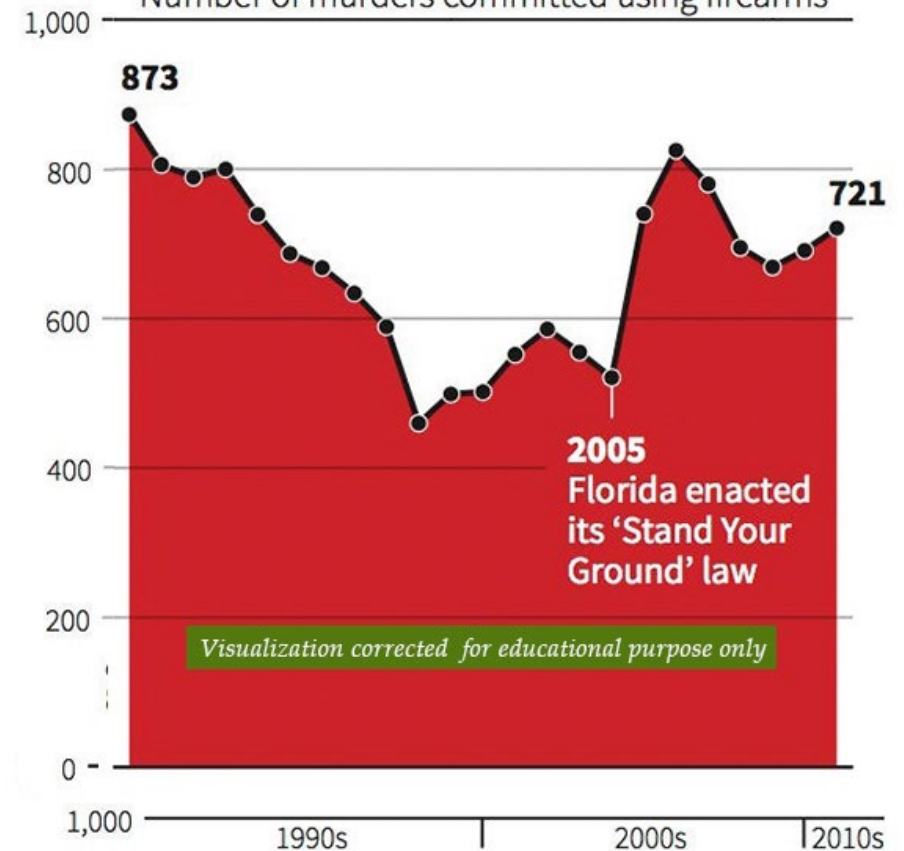
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

Tell the truth

The Economist owns up to visualization mistakes

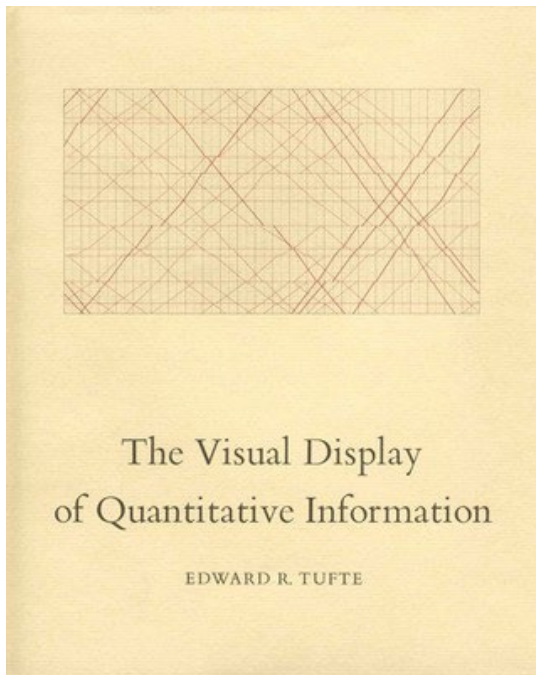
- Poor carpenters blame their tools

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.

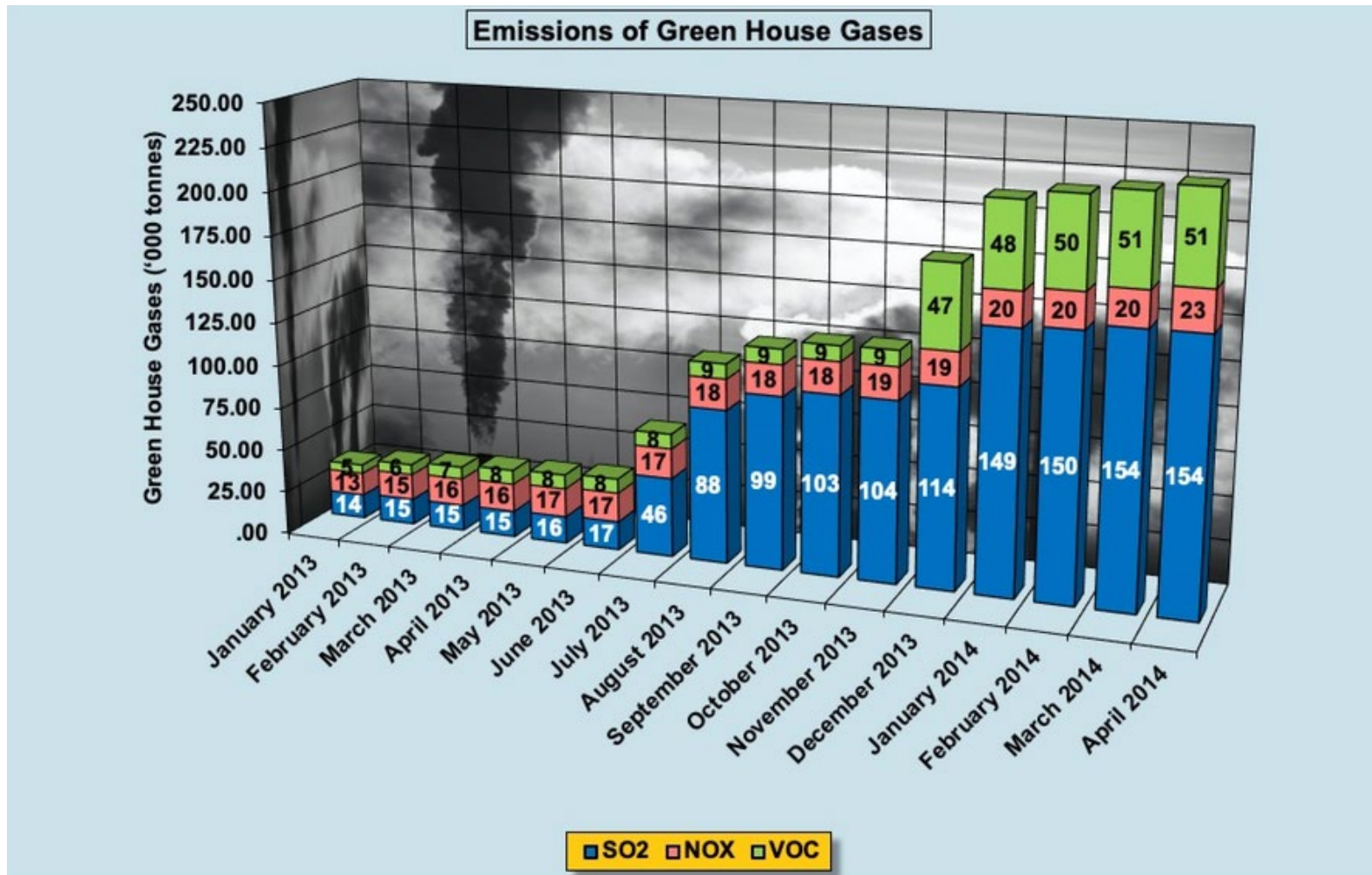
Data-ink ratio and chartjunk

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

$$= 1 - \text{proportion of a graphic that can be erased}$$



Data-ink ratio and chartjunk



Data-ink ratio and chartjunk

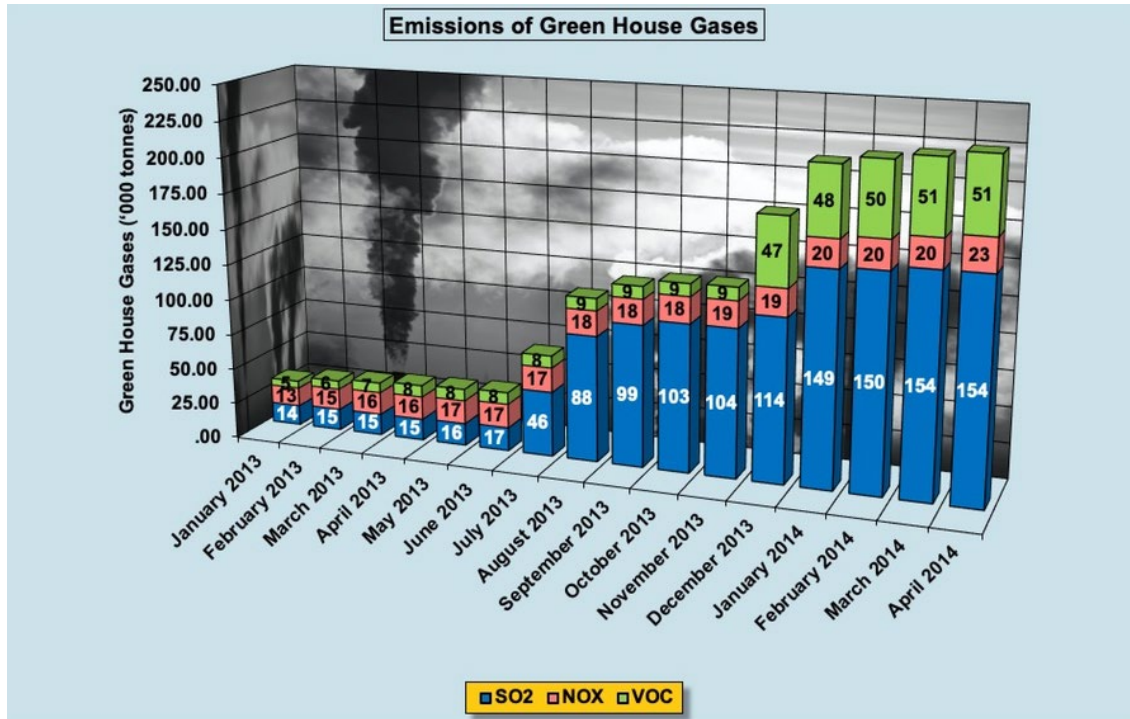
The Data-ink-ratio

Tricks to create great looking charts



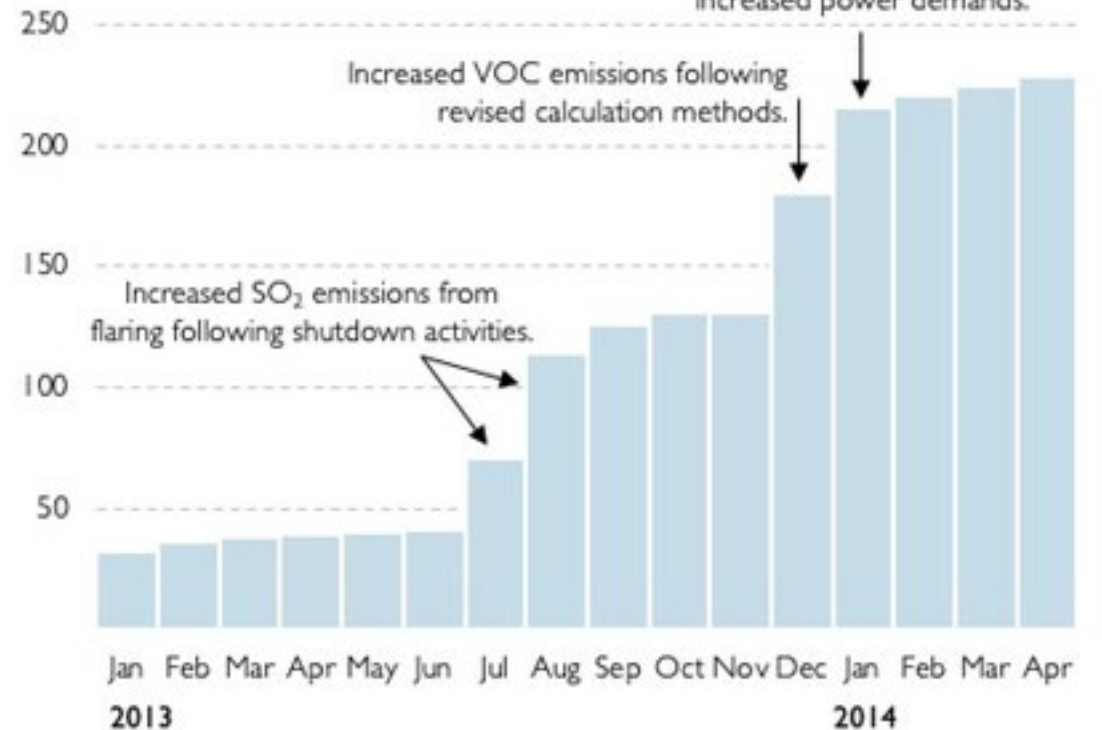
simplexCT
Consulting and Training

Data-ink ratio and chartjunk



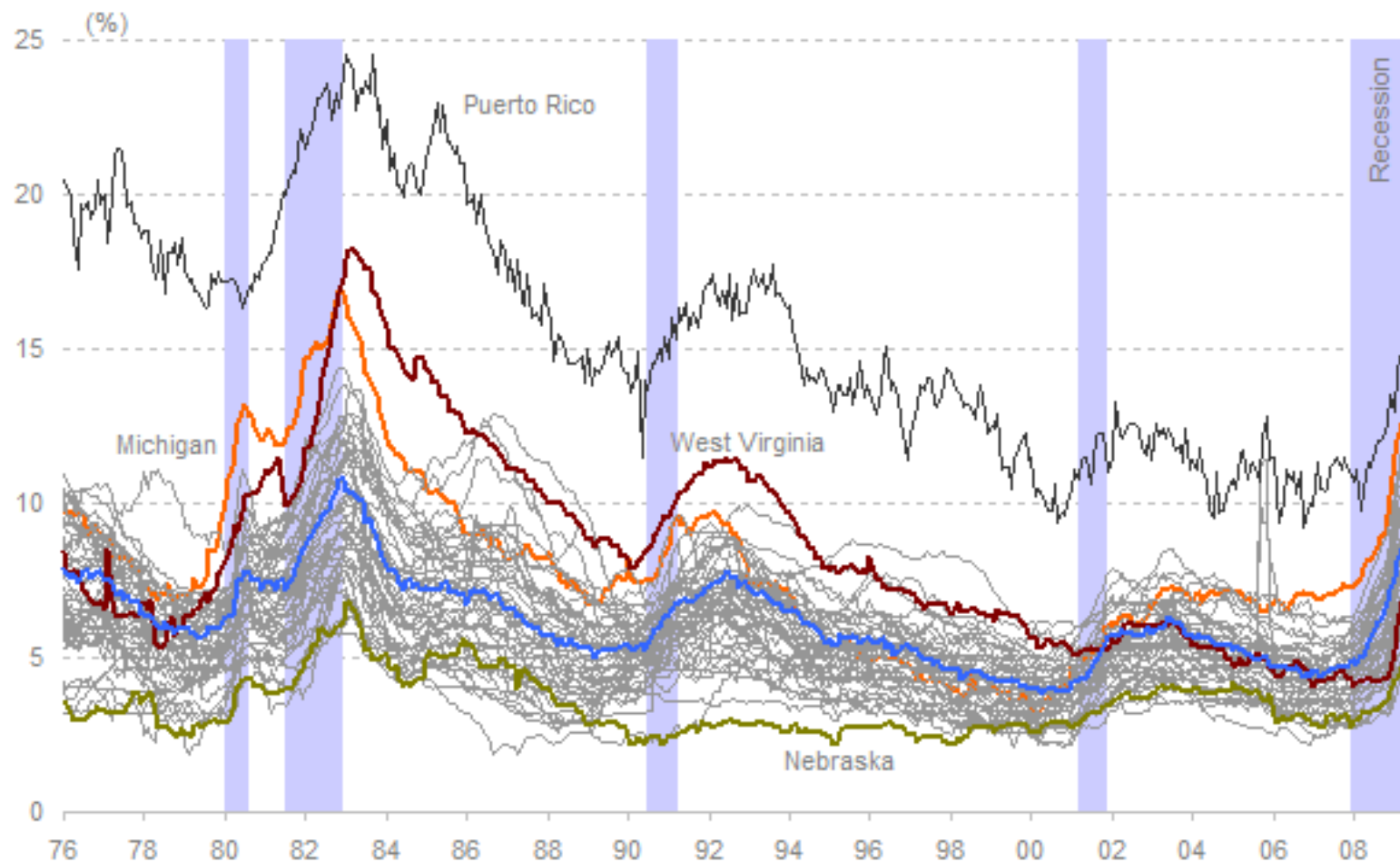
Emissions of Green House Gases

'000 Tonnes of SO₂/NO_x/VOC



Multivariate data – small multiples

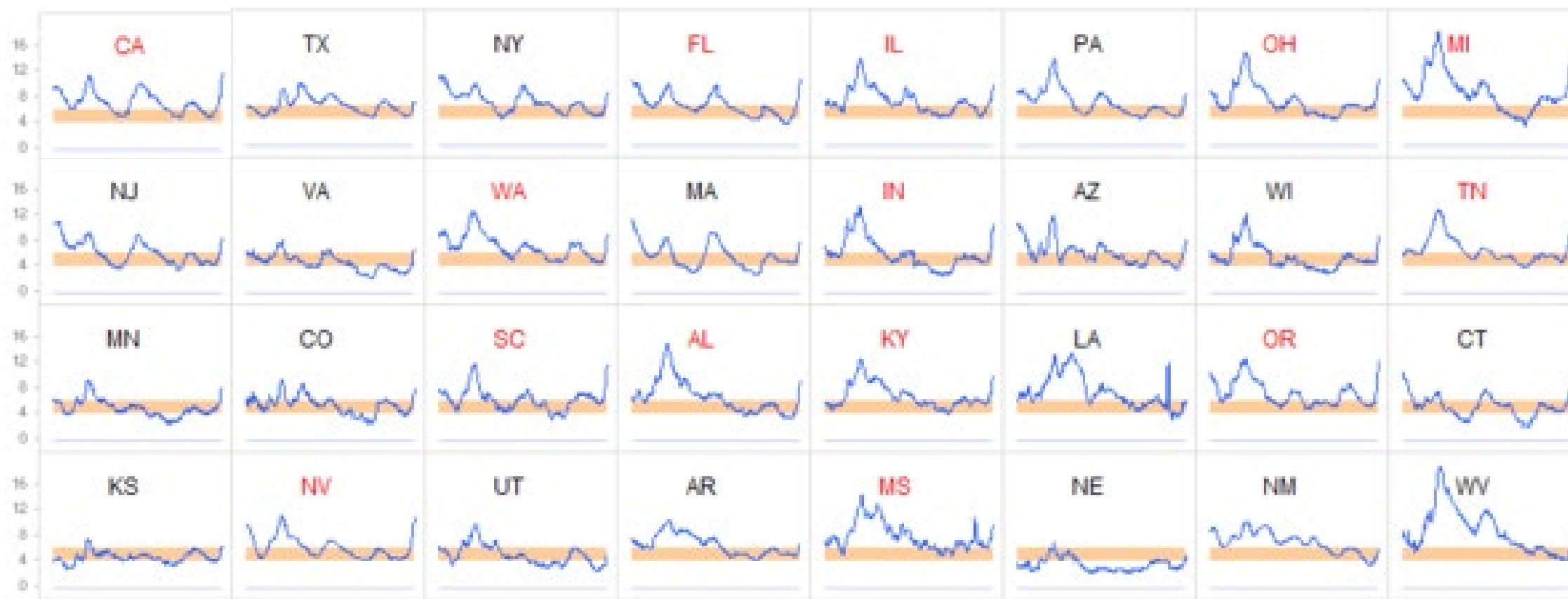
Monthly Unemployment Rate by State Jan 1976 - Apr 2009



Source: Bureau of Labor Statistics

Multivariate data – small multiples

Monthly Unemployment Rates by State, Jan 1976 - Apr 2009

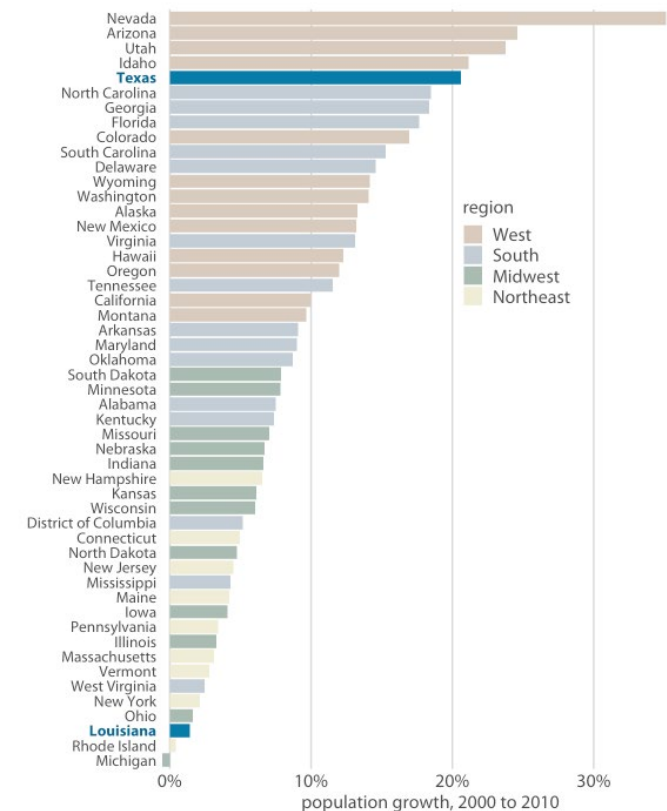
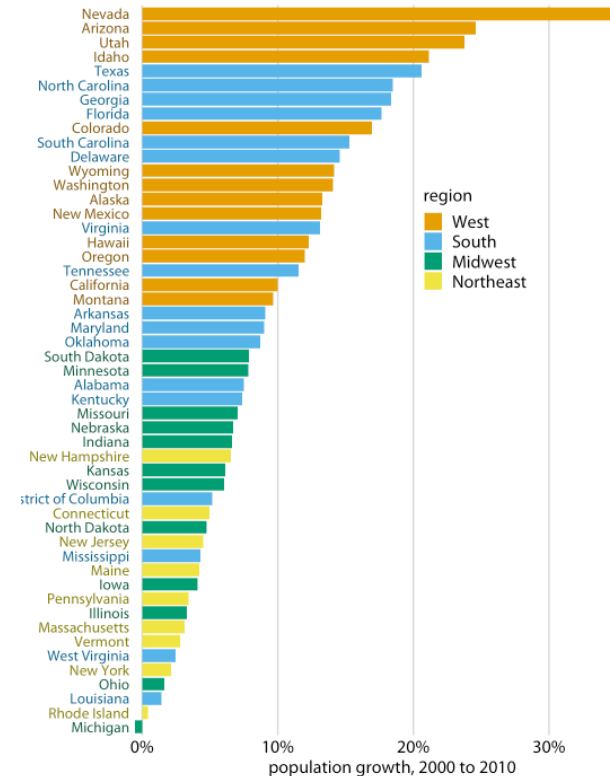
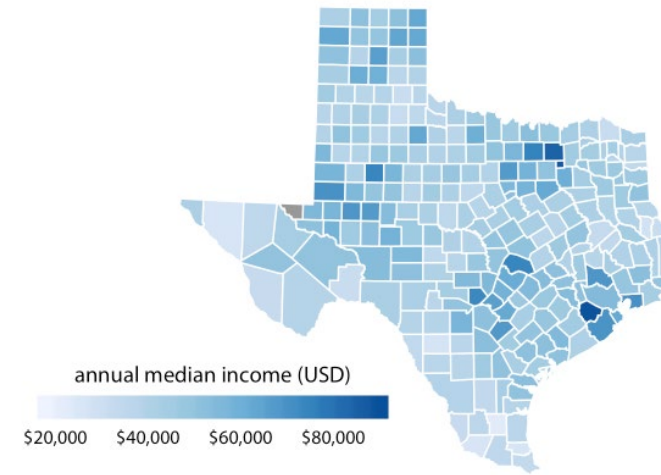


When to use color?

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

Use color to:

- Distinguish groups
- Represent data values
- Emphasize visual elements



Types of color scales

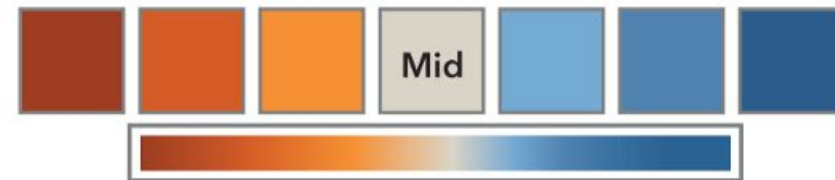
SEQUENTIAL

color is ordered from low to high



DIVERGING

two sequential colors with a neutral midpoint



CATEGORICAL

contrasting colors for individual comparison

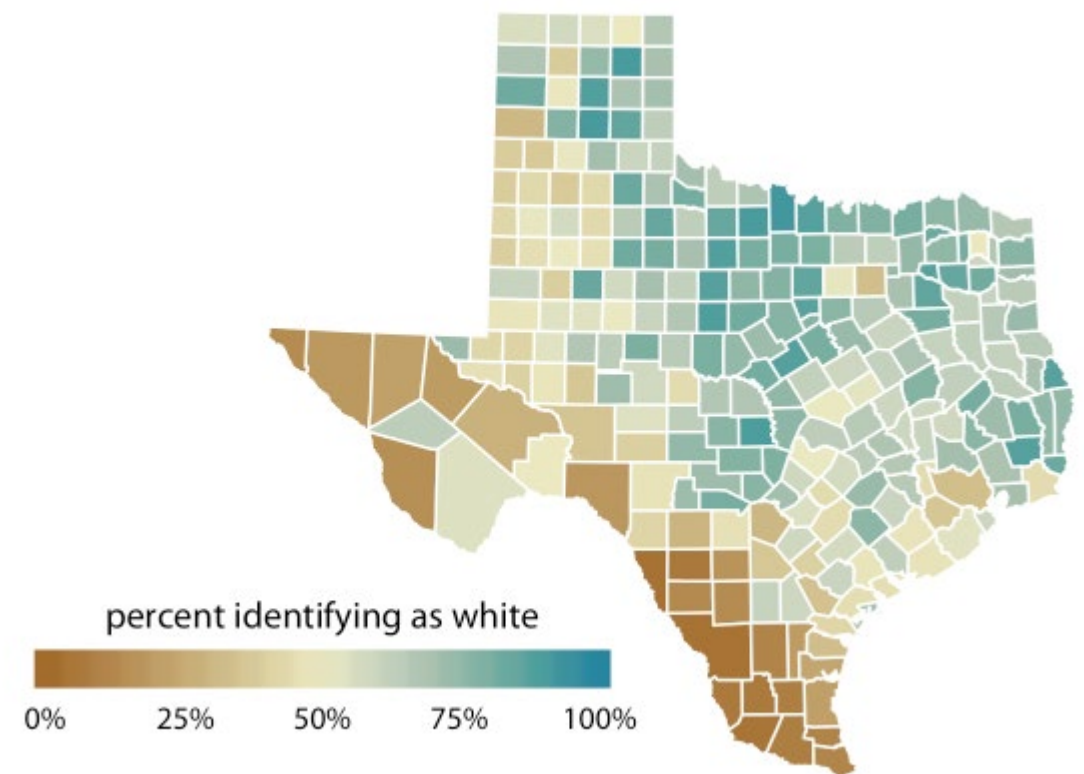
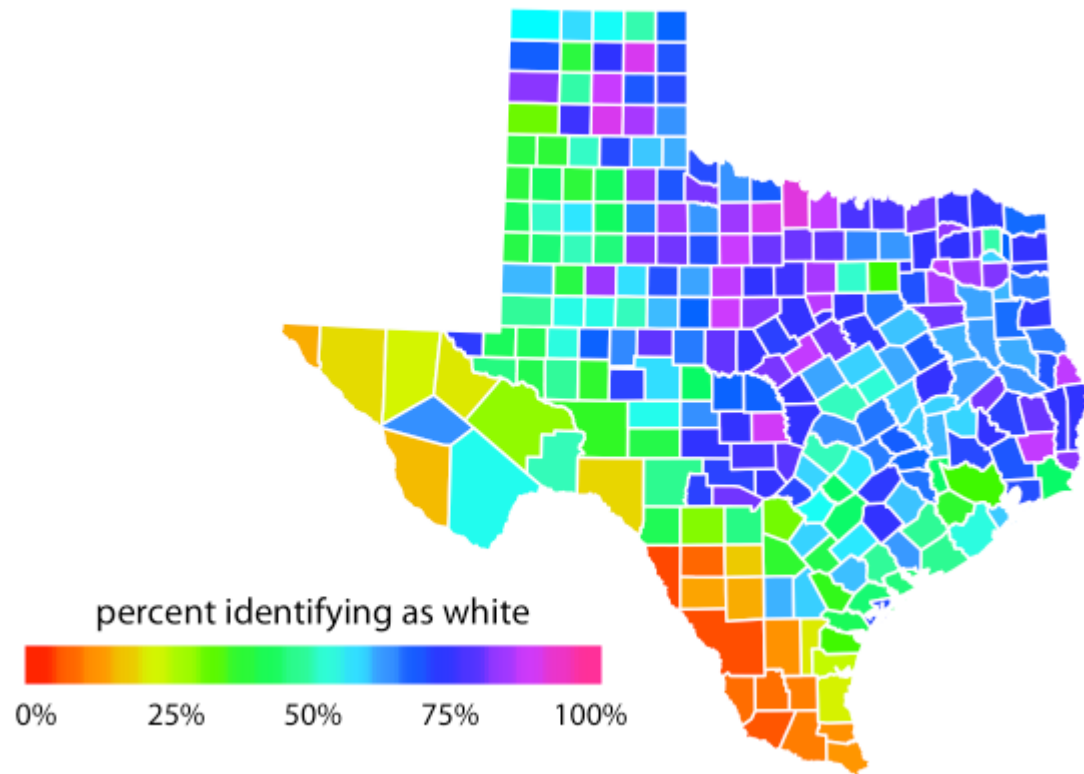


HIGHLIGHT

color used to highlight something

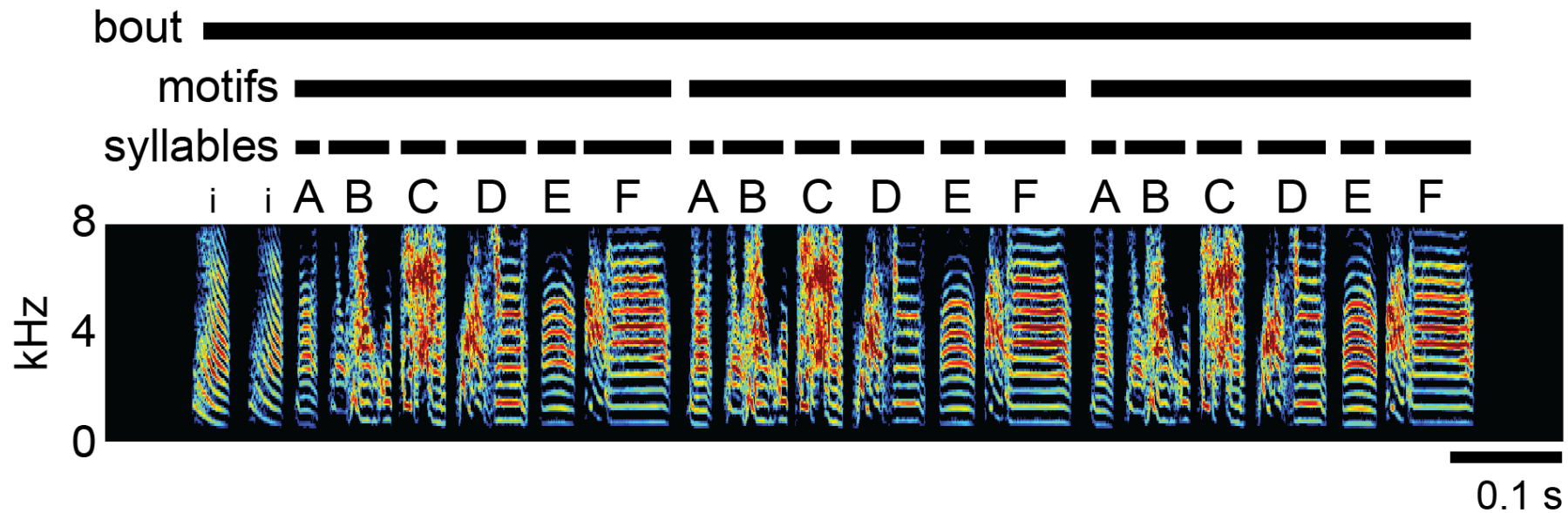


Beware the rainbow

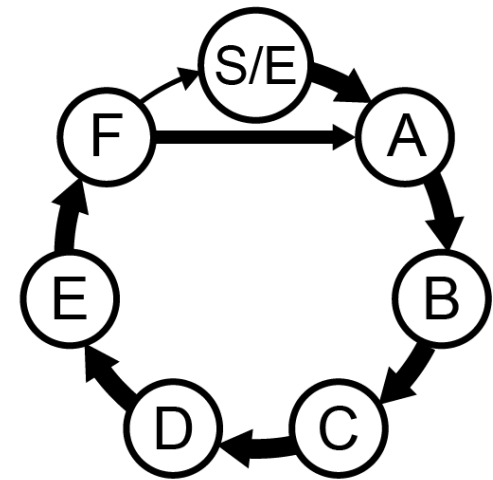


Beware the rainbow (unless your boss wants it)

a



b



Color and palette resources

[Top R Color Palettes to Know](#)

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

