



# WEEK 12: DATA VISUALIZATION

Dr. Kai Li

School of Information Sciences  
University of Tennessee, Knoxville  
Spring 2025

# Review/Overview

- Visualization
- No class next week (Spring Recess)!
- Assign 2 submission: please only submit the rmd file with all your finished code and writing.

# Final project presentation

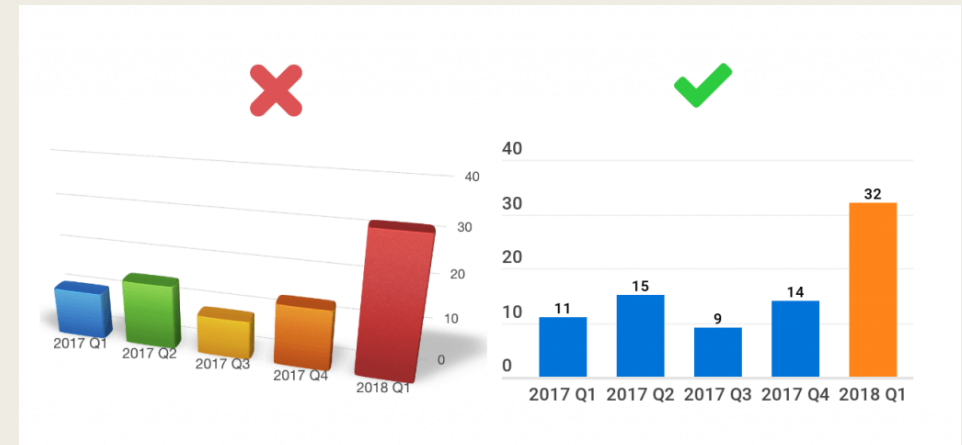
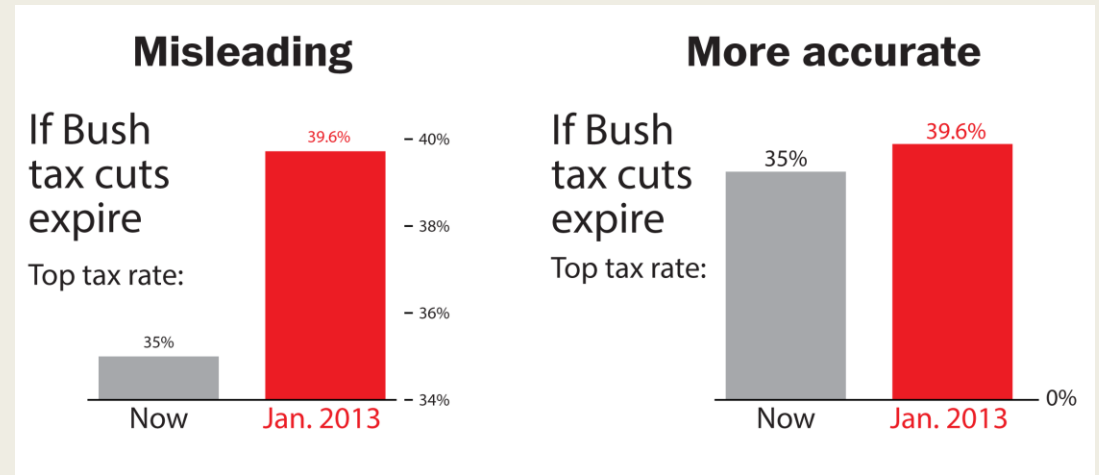
- Please prepare around 10-15 minutes of presentation to tell your story.
- Registration sheet:  
<https://docs.google.com/spreadsheets/d/1bt52DmQuYn8kAEZemCRrhmgHgOIHLd7EUY-PD1RSA7k/edit?usp=sharing>
  - *Please register your time and title (if you prefer) before next Thursday (4/17)*

# Final project presentation

- For the submission:
  - *You can upload a .docx or .html file for your final report.*
  - *Or, you can also consider creating a GitHub page and just send the link to Canvas.*
    - One example would be: [https://github.com/nalsi/dh\\_visualization](https://github.com/nalsi/dh_visualization)

# Some visualization pitfalls

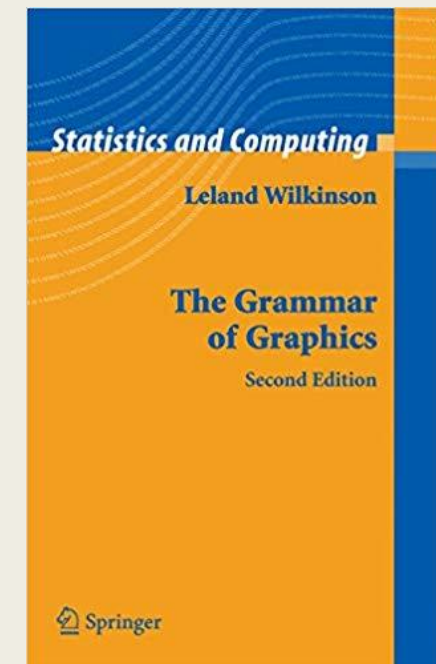
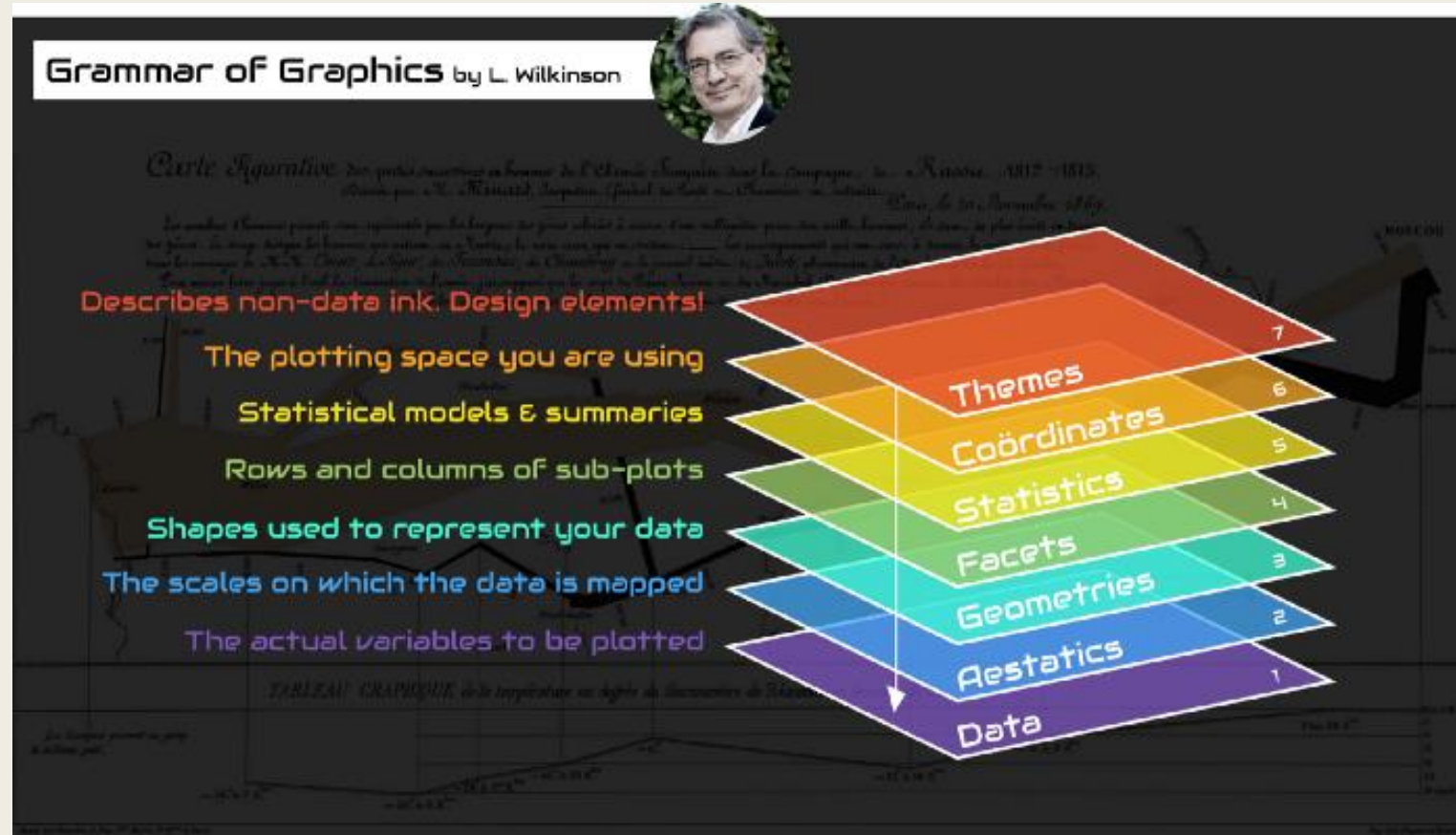
- For bar chart, please:
  - *Do not cut the numeric axis.*
  - *Sort the bars by the number (unless the other axis has another meaning, like time).*
- Do not use fake 3-D presentation!



# Other pitfalls in visualization design

- This is a very good summary of pitfalls in visualization design:  
<https://www.data-to-viz.com/caveats.html>
- It is a good exercise to critical evaluate every visualization work that you come across.

# Grammar of Graphics



# ggplot2

- ggplot2 in R: the very original package implementing GG
  - <https://ggplot2.tidyverse.org/>
  - Extensions: <https://exts.ggplot2.tidyverse.org/gallery/>
  - *It was translated into the plotnine package in Python, with the same grammar but without any extension.*
  - A very useful cheatsheet:  
<https://github.com/rstudio/cheatsheets/blob/main/data-visualization.pdf>



# Why ggplot2?

- By implementing GG model, we can manage all details in one layer without having to mess up with other layers.
  - *For example, we can flip the axes very easily with whatever graph that we have.*
- We have a process to follow: data → aesthetics → geometry...
- Extensions

# Aesthetics

- Aesthetics are the features that we see in the graph, each is mapped to a variable in the data. In general, you should consider using the following aesthetic elements (or “visual patterns”):

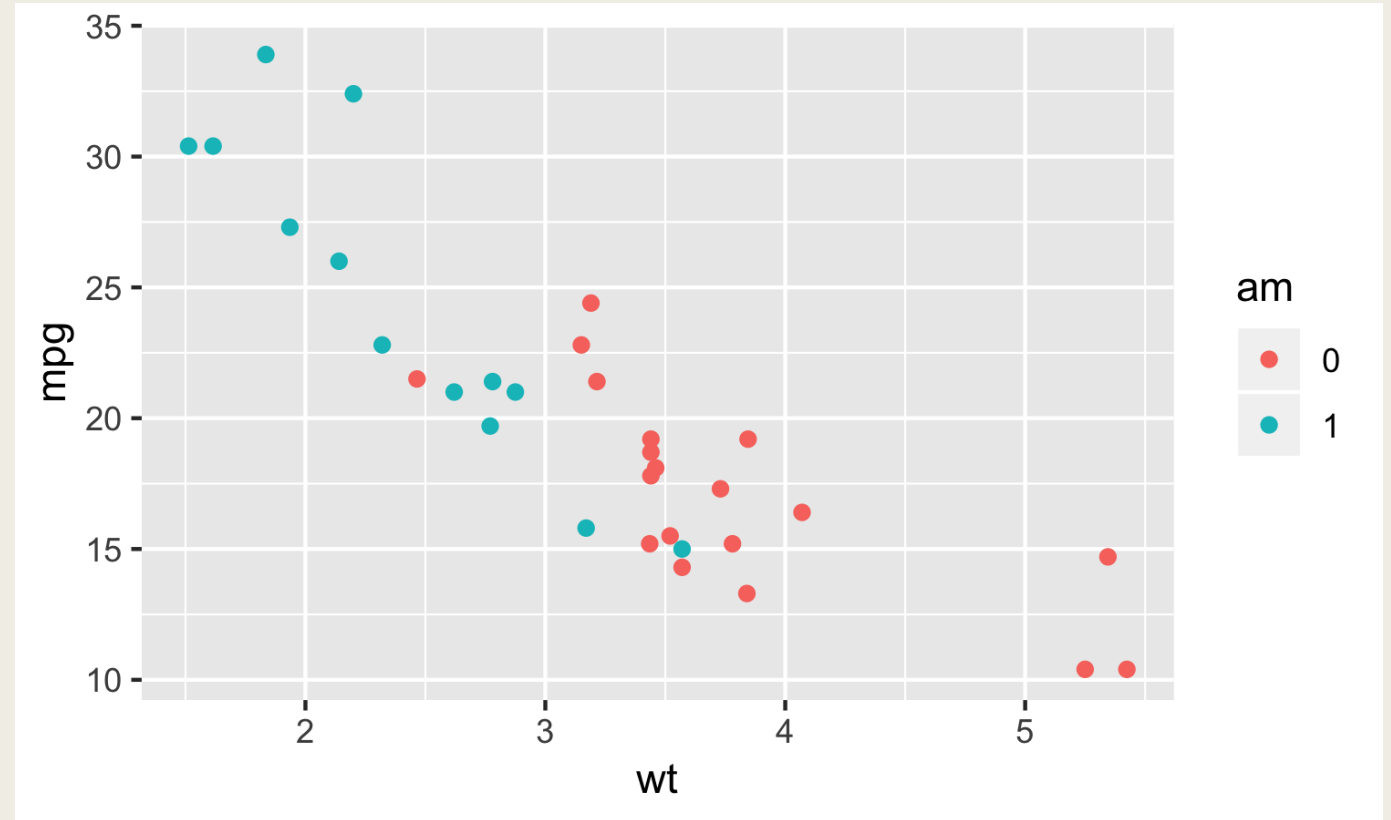
- *Axes*
- *Color*
- *Shape*
- *Size*
- *Transparency*

1. Position along a common scale
2. Position on identical but nonaligned scales (e.g., small multiples)
3. Length
4. Angle, Slope
5. Area
6. Volume, Density, Color Saturation
7. Color hue

Shapes on the top of list are more suitable to be mapped to a numeric variable vs. categorical, and vice versa.

# For example

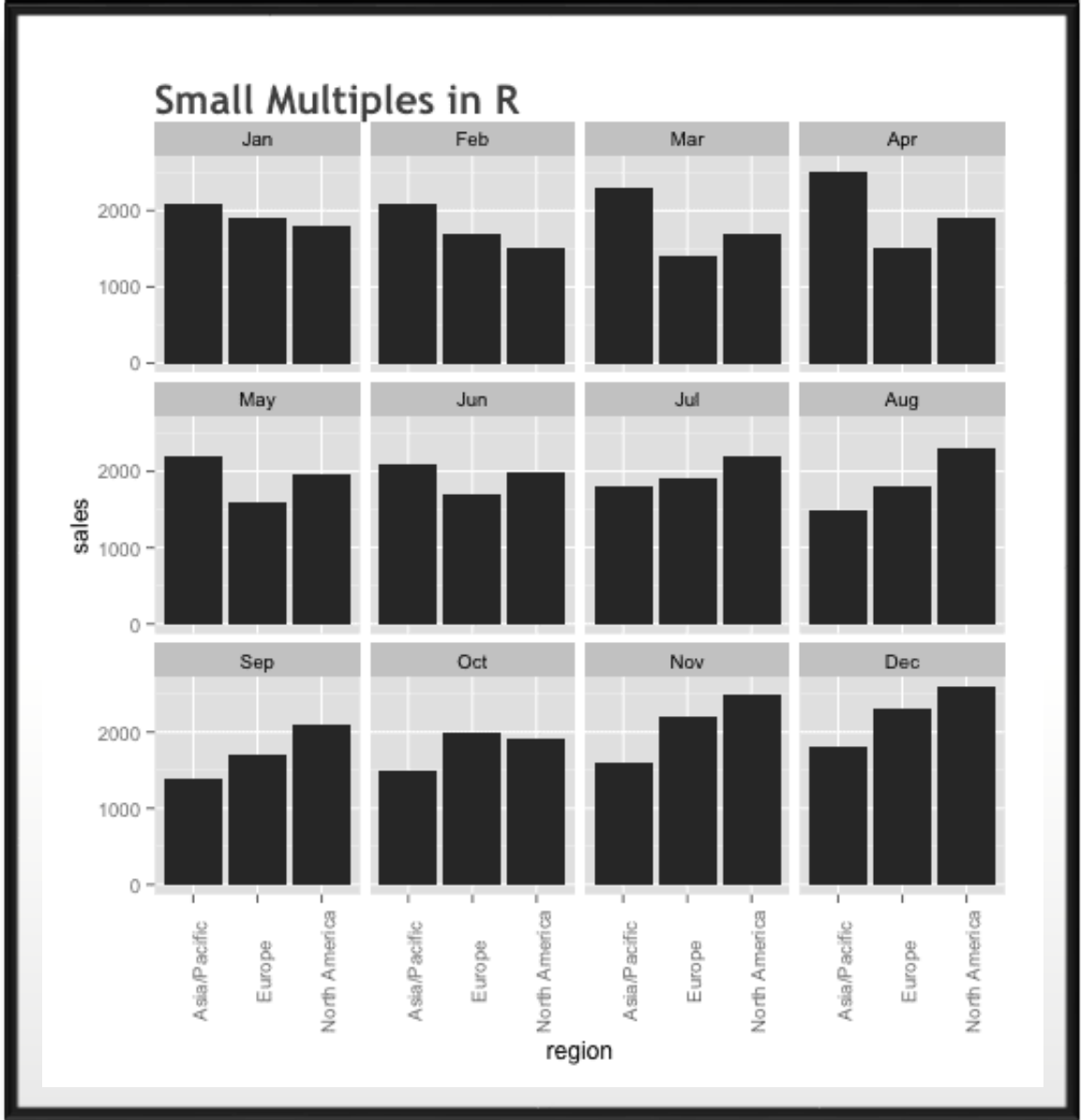
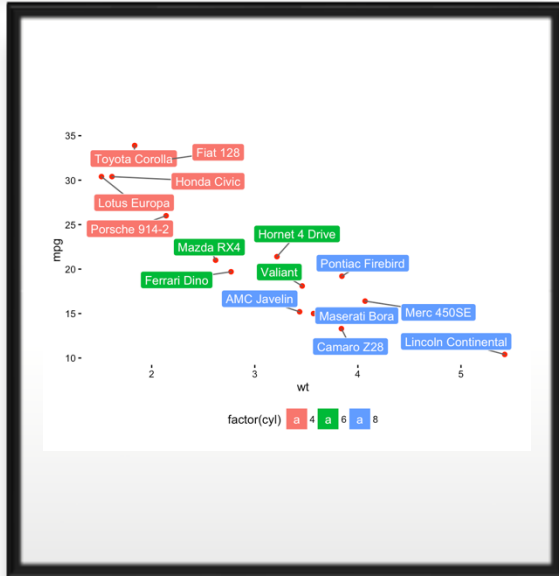
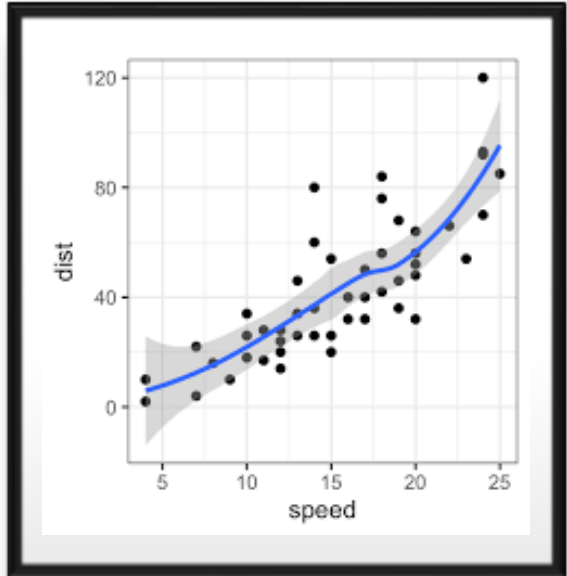
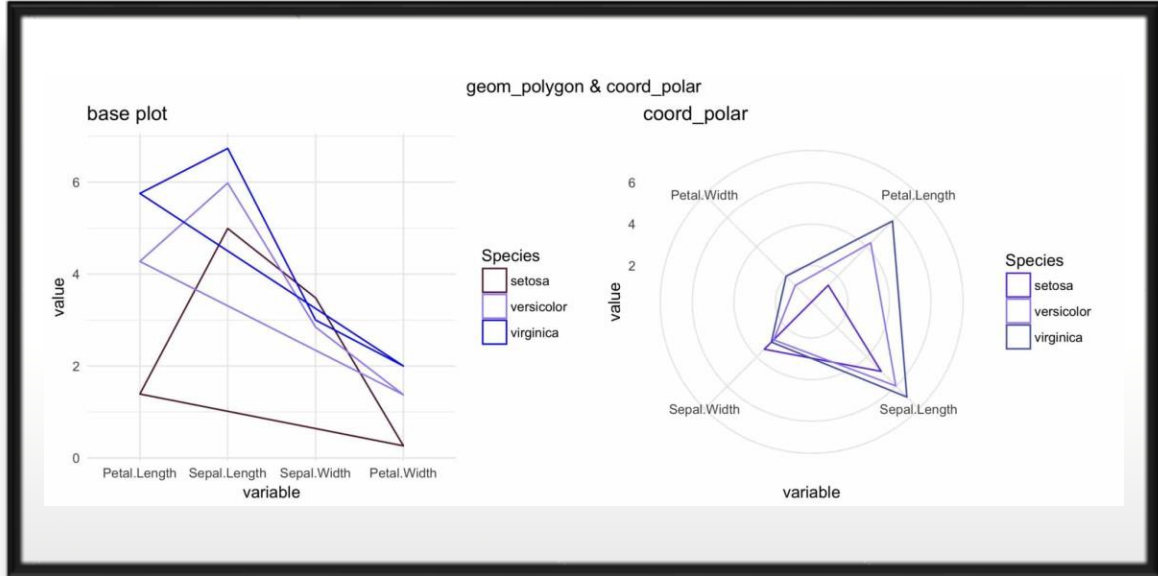
- `x <- wt`
- `y <- mpg`
- `color <- am`
- `geom_point()` shows that we are creating a scatterplot.



```
ggplot(mtcars) +  
  geom_point(aes(x = wt, y = mpg, color = am))
```

# Other layers matter too!

- Facet: the subplots (small multiples)
- Statistics: if to add any statistical summary to the data, such as smooth line?
- Coordinate: do we want to use the original coordinate, or use some transformed version?
- Theme: the overall aesthetic theme used in the graph.
- And always remember that titles and labels matter!



# Demo of ggplot2

- 1. Basic structure of ggplot2 code
- 2. Adding different layers
- 3. Exporting ggplot2 graphs
- 4. Color: we will talk about it in the second half of the lecture

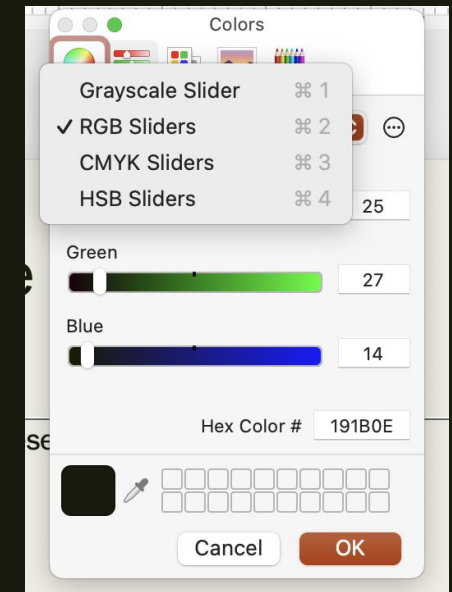
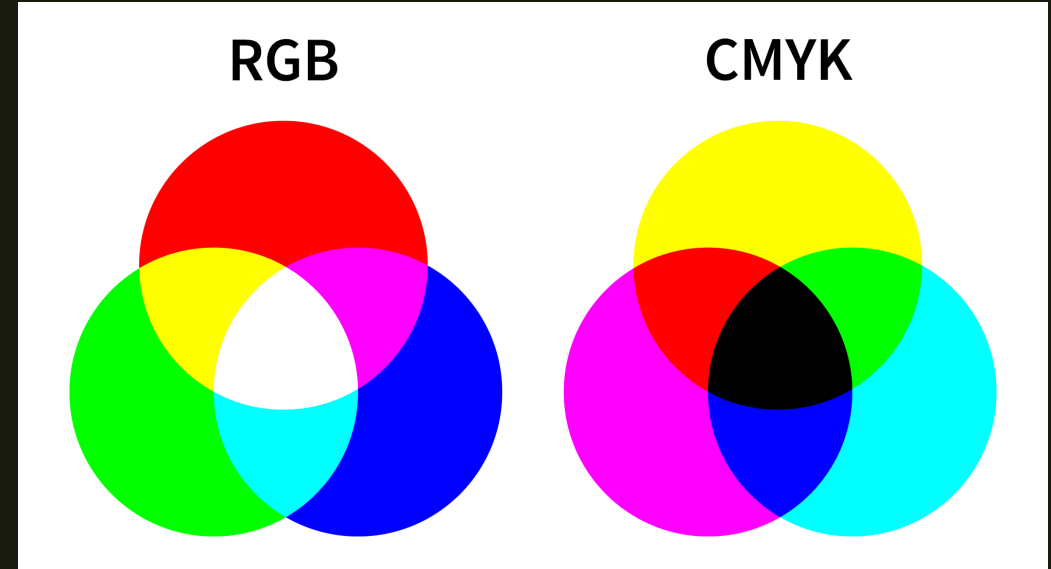
# Colors in visualization

- Human beings have inaccurate and subjective perception of colors.
  - *TED-ED: how human perceive color:*  
[https://www.youtube.com/watch?v=l8\\_fZPHasdo](https://www.youtube.com/watch?v=l8_fZPHasdo)
- However, color is extremely important to visualization because it makes the graph **more appealing**, and it also **bears meanings**.



# Color systems

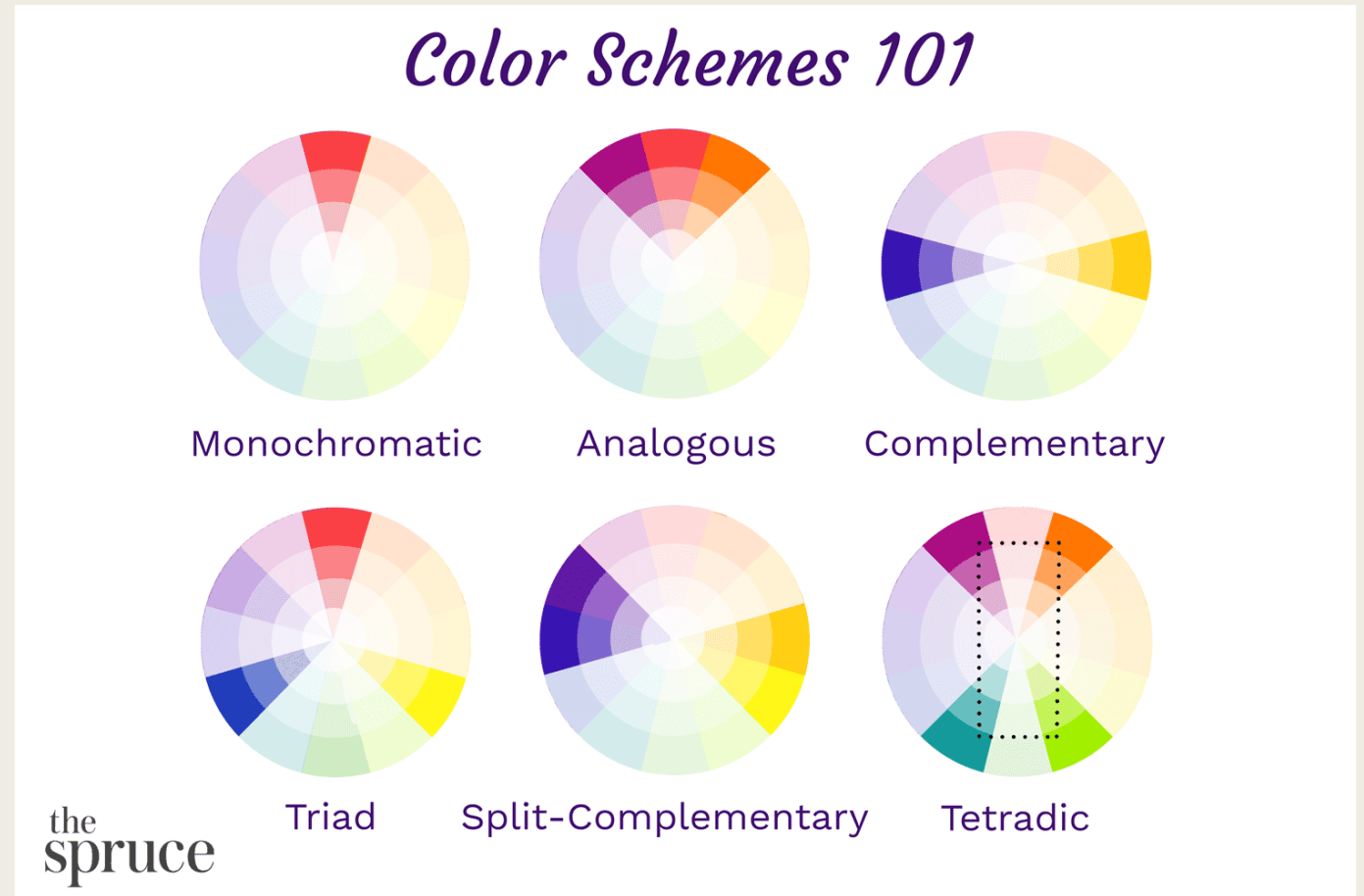
- Each color can be represented by its position in one of the color systems, such as **RGB** and **CMYK**.
- The **color wheel** contains all potential colors (using whichever color system) and can guide the choice of colors in the same work.





# Color palette

- A color palette is a small subset of colors that can be used together.
- There are some general principles to choose colors into a palette.

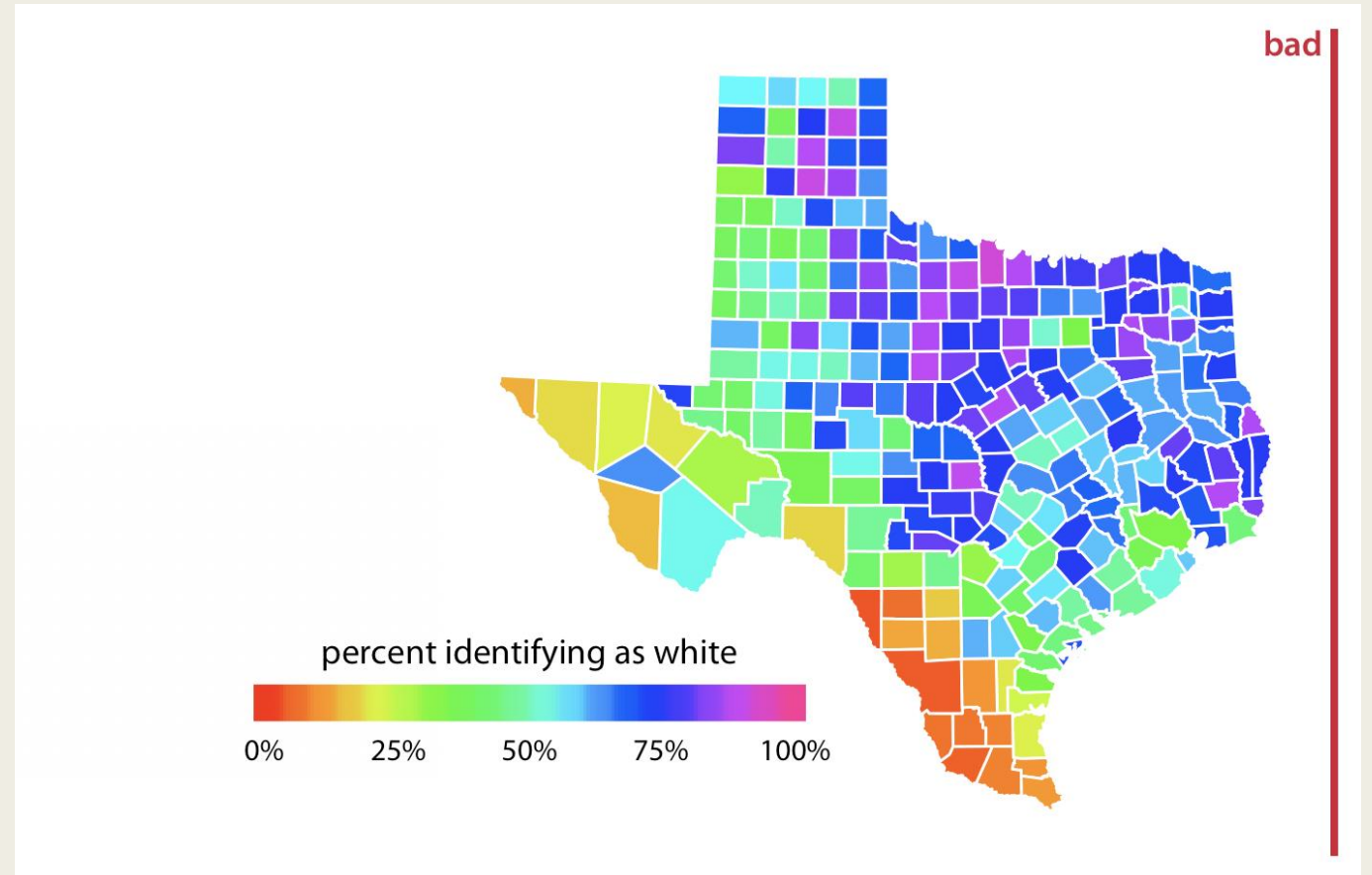


# Color palette

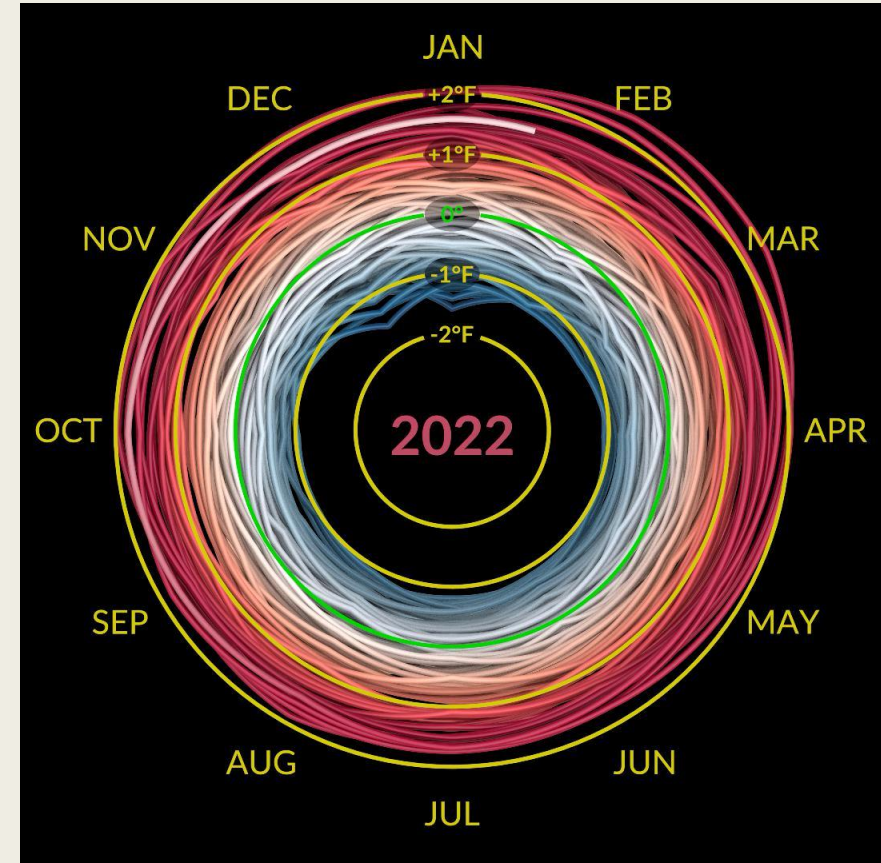
- But we can also use pre-defined color palettes in visualization tools!
  - Example of RColorBrewer: <https://www.r-graph-gallery.com/38-rcolorbrewers-palettes>
  - Another popular project in R is viridis: <https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html>.
- There are three types of color scales:
  - Sequential scales (numerical and categorical)
  - Diverging scale (numerical and categorical)
  - Qualitative scales (categorical)
- Don't forget that color is still best used for categorical values.

## Some general suggestions/pitfalls of using colors

- Don't use the rainbow palette.
- Don't use too many colors.
- Offering contrast to categorical values is more powerful!

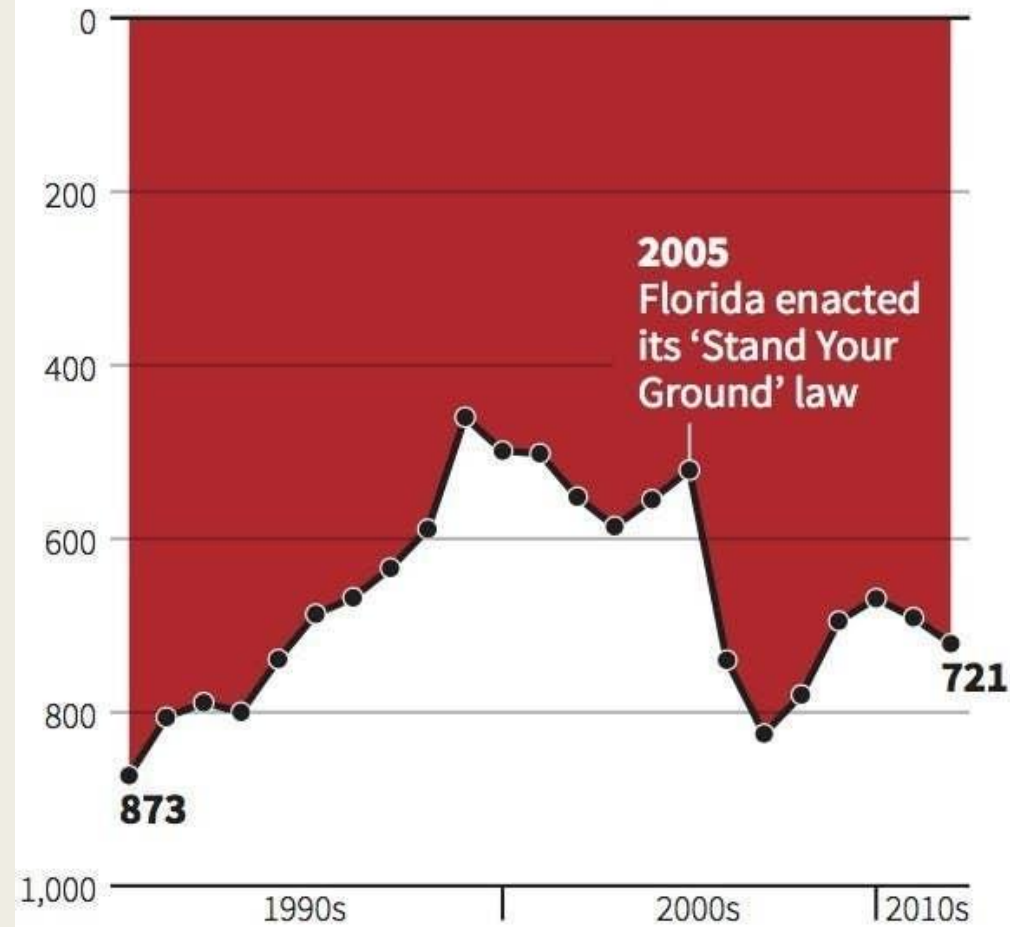


# Metaphors of color



# Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement