STAT 209
A Taxonomy of Graphics

June 8, 2021

Colin Reimer Dawson

Topics for This Video

- Plot (geom) types
- Color palettes
- ggplot2 themes

Coming Up

- Thursday: Version Control (`git` and GitHub)
- Next Tuesday: Integrative lab to "reverse engineer" a published visualization
- Next Thursday: Project 1 workshop

Outline

Quick Summary: ggplot2

Step 1: load library

```
> library(ggplot2)
```

Result

(no change)

Figure: Slide by Jordan Crouser at Smith College

Quick Summary: ggplot2

Step 2: make `ggplot()` object

```
> library(ggplot2)
> ggplot()
```

Result

Figure: Slide by Jordan Crouser at Smith College

Quick Summary: ggplot2

Step 3: tell ggplot about data

```
> library(ggplot2)
> ggplot(arbuthnot)
```

Result

(no change)

Figure: Slide by Jordan Crouser at Smith College

Quick Summary: ggplot2

Step 4: map variables to dimensions

```
> library(ggplot2)
> ggplot(arbuthnot,
+        aes(x=boys,
+            y=girls))
```

No closing parentheses tells R to wait for more commands

Result

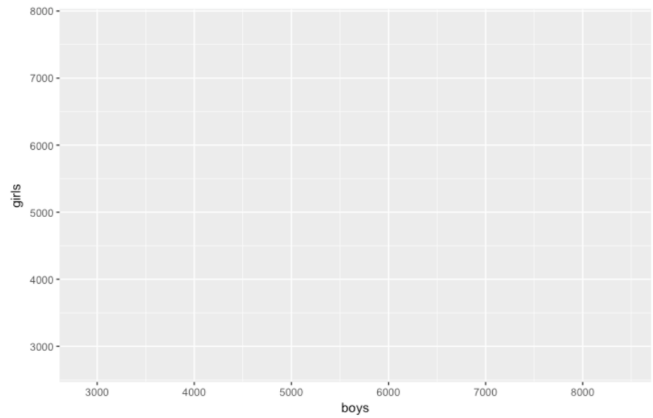


Figure: Slide by Jordan Crouser at Smith College

Quick Summary: ggplot2

Step 5: add appropriate geoms

```
> library(ggplot2)
> ggplot(arbuthnot,
+        aes(x=boys,
+            y=girls)) +
+ geom_point()
```

Result

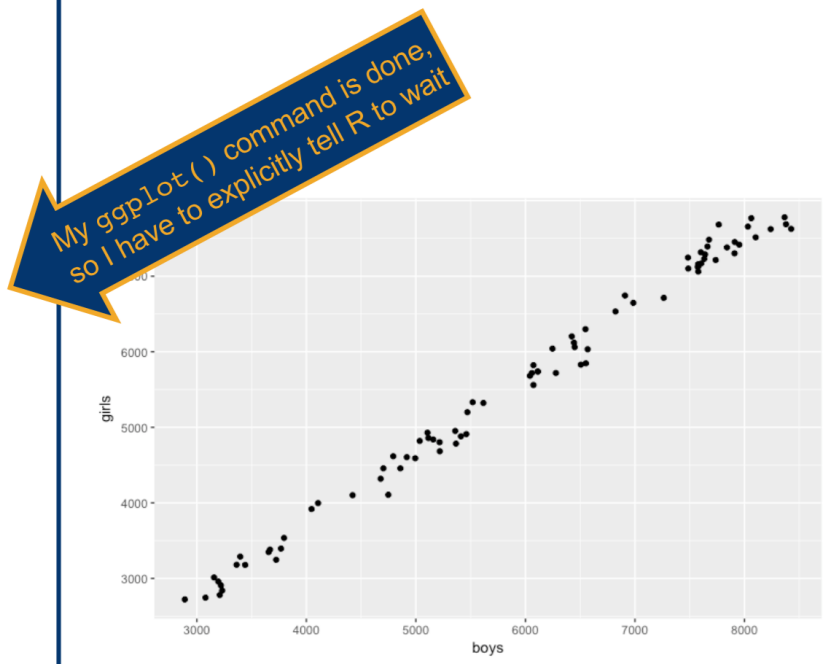


Figure: Slide by Jordan Crouser at Smith College

Quick Summary: ggplot2

Step 6: style the geoms

```
> library(ggplot2)
> ggplot(arbuthnot,
+        aes(x=boys,
+            y=girls)) +
+ geom_point(aes(
+    color = boy_ratio))
```

Result

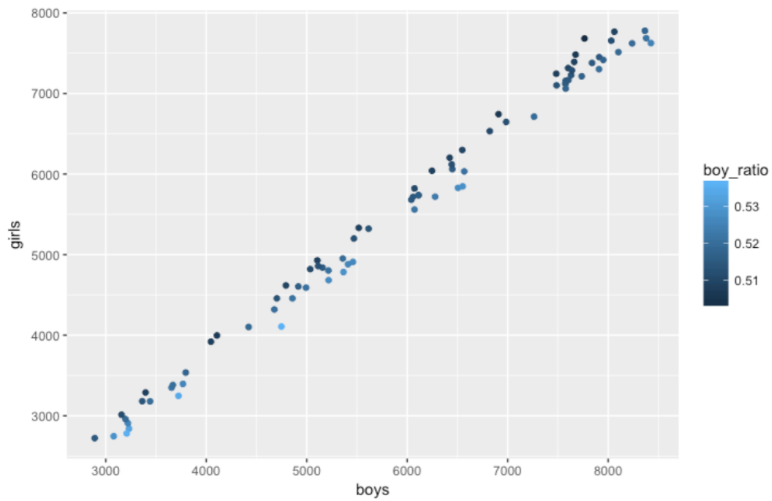


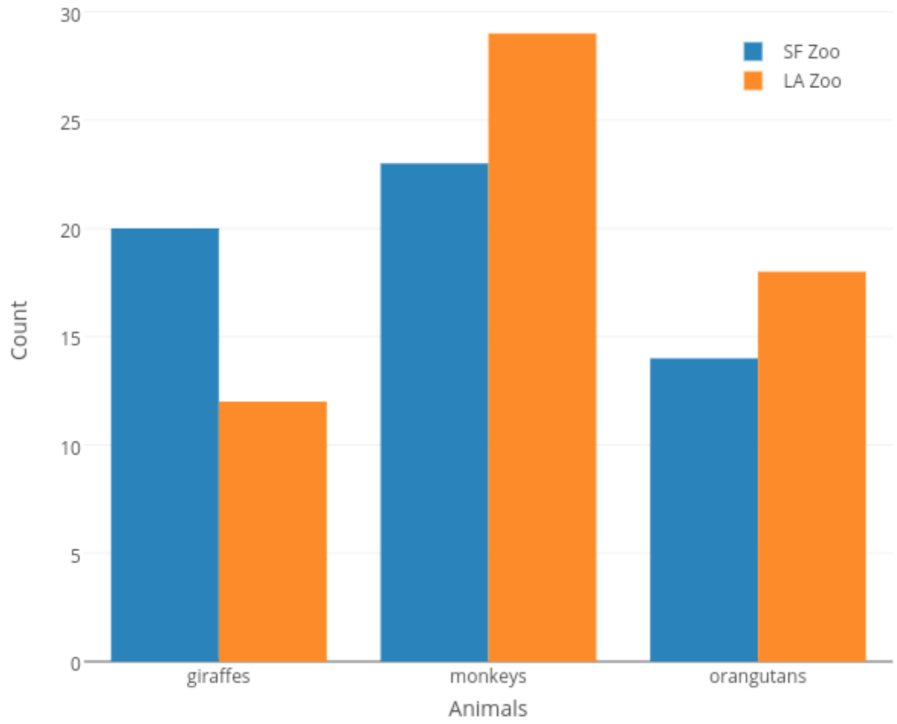
Figure: Slide by Jordan Crouser at Smith College

Outline

Common Graph Types

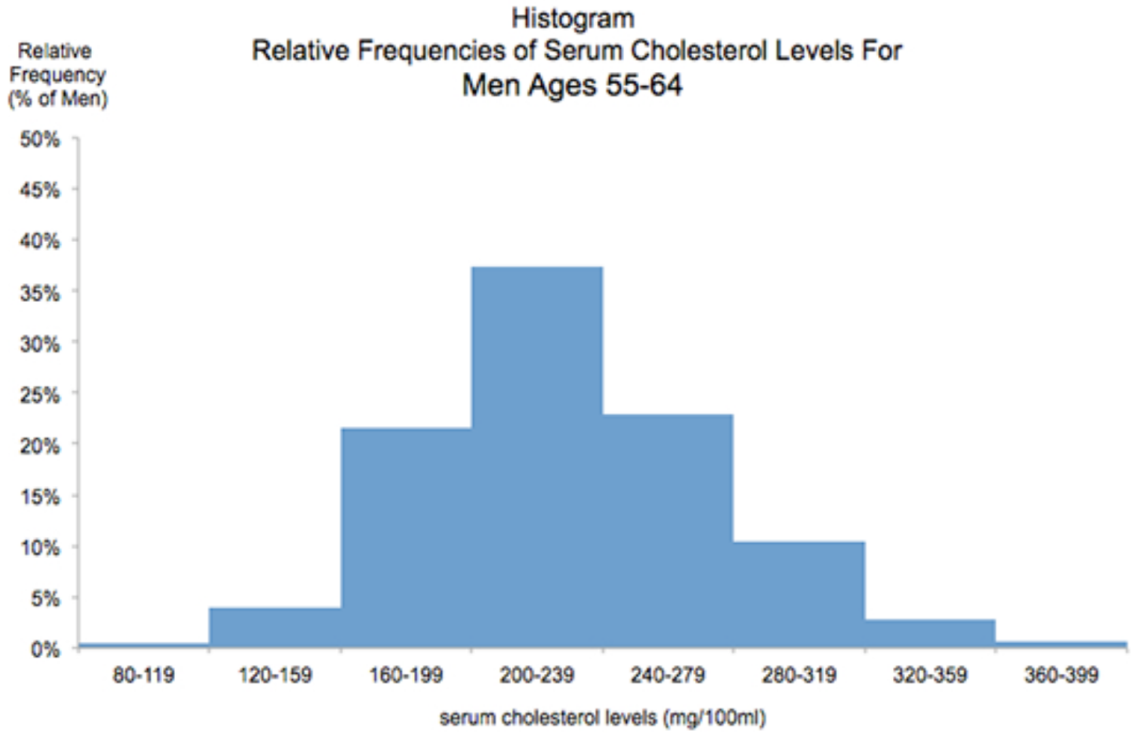
- Bar Graph
- Histogram
- Box Plot
- Scatterplot
- Line Chart
- Map

Bar Graph



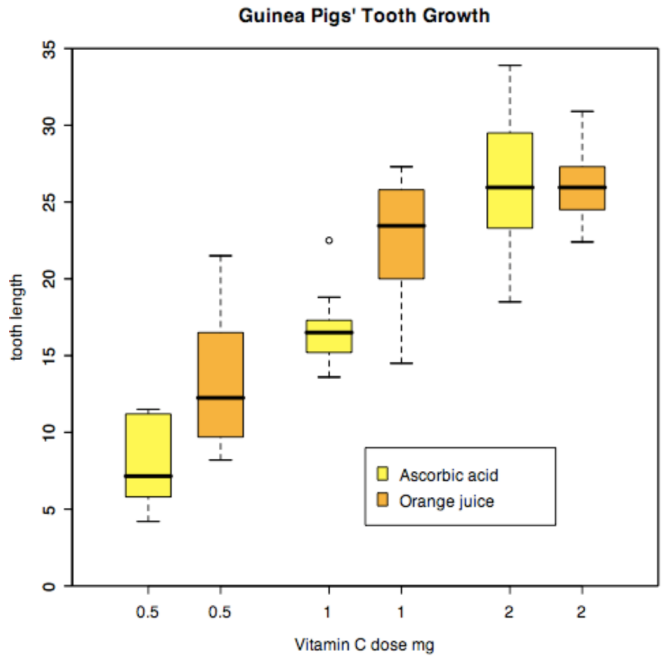
- Allows comparison of a statistic (often just "amount of data") across categories
- Can use grouping or stacking to bring in a second categorical variable and depict combinations

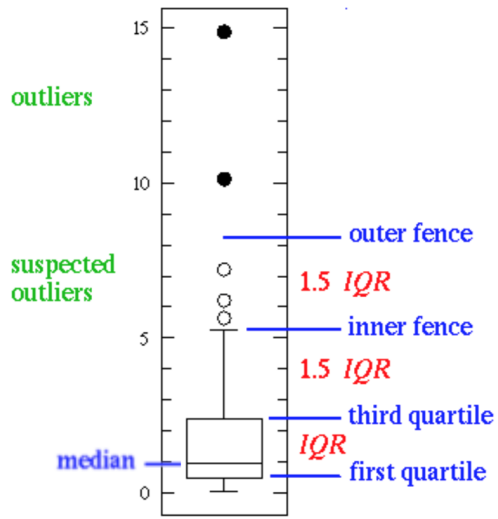
Histogram



- Show count or proportion within bins of a *quantitative* variable
- Choice of bin width/cutoffs can affect impression dramatically

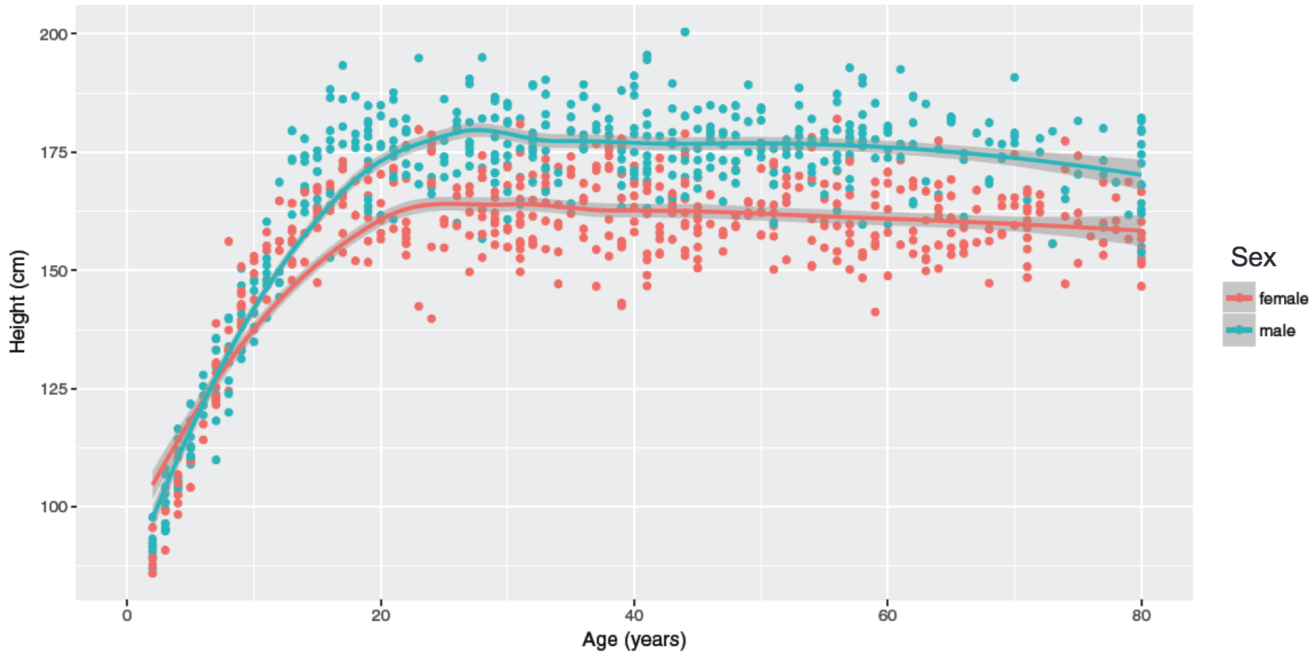
Box Plot





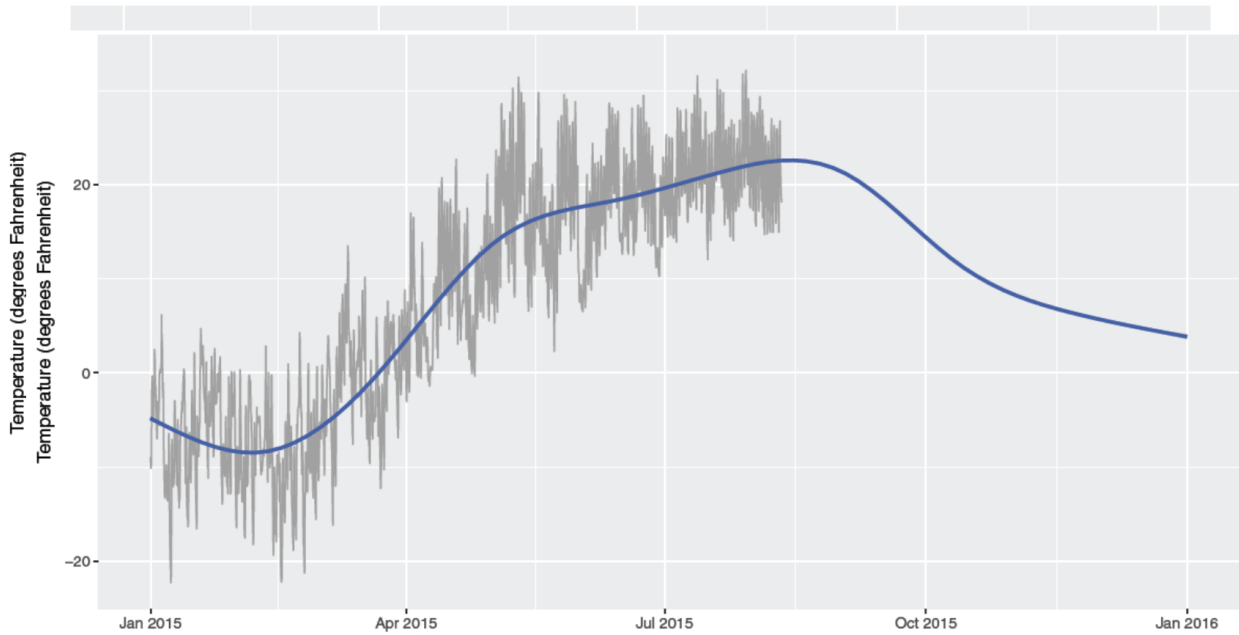
- Alternative to histogram to show distribution of a quantitative variable
- Can group by a categorical variable to compare distributions

Scatterplot



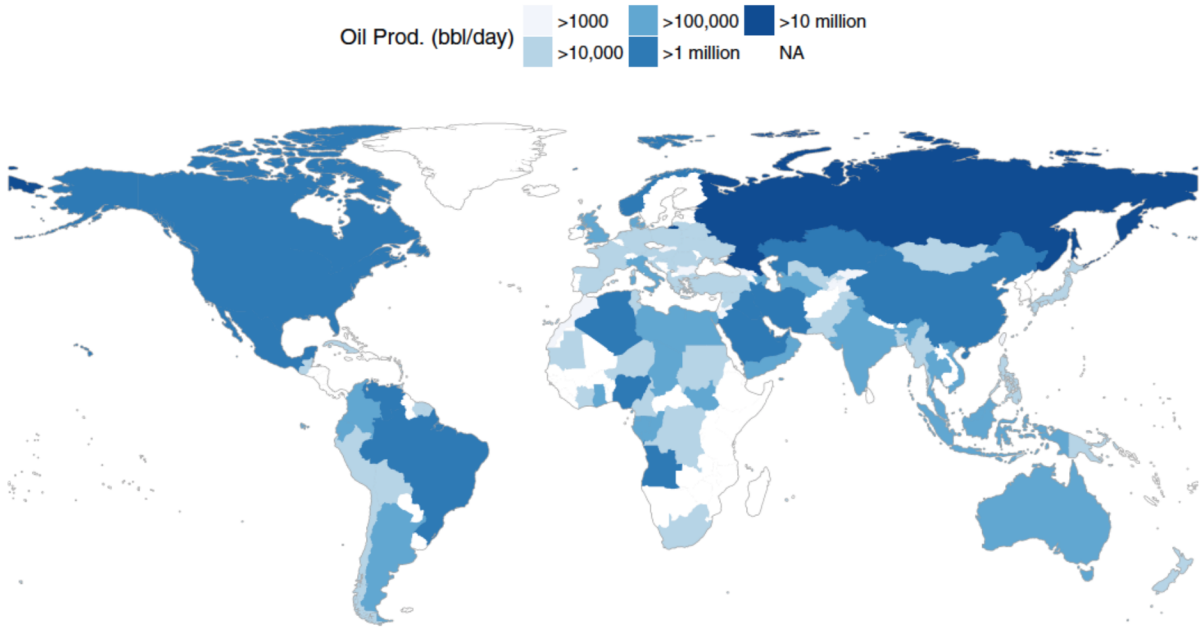
- Depicts the relationship between two quantitative variables
- Can bring in additional variables using hue, saturation, symbol
- Can show trend line based on a model (e.g., regression)

Line Chart



- Useful to depict trends in *sequential* data
- Essentially a scatterplot with dots connected, but this must make sense
- Can depict groups, or overlay multiple variables to see when variables move together/apart/switch ordering

Map



- Useful for geographic data (duh)
- A filled map (like this) is called a **chloropleth**
- May alternatively plot data point in specific locations

Pitfalls and Distortions of Maps

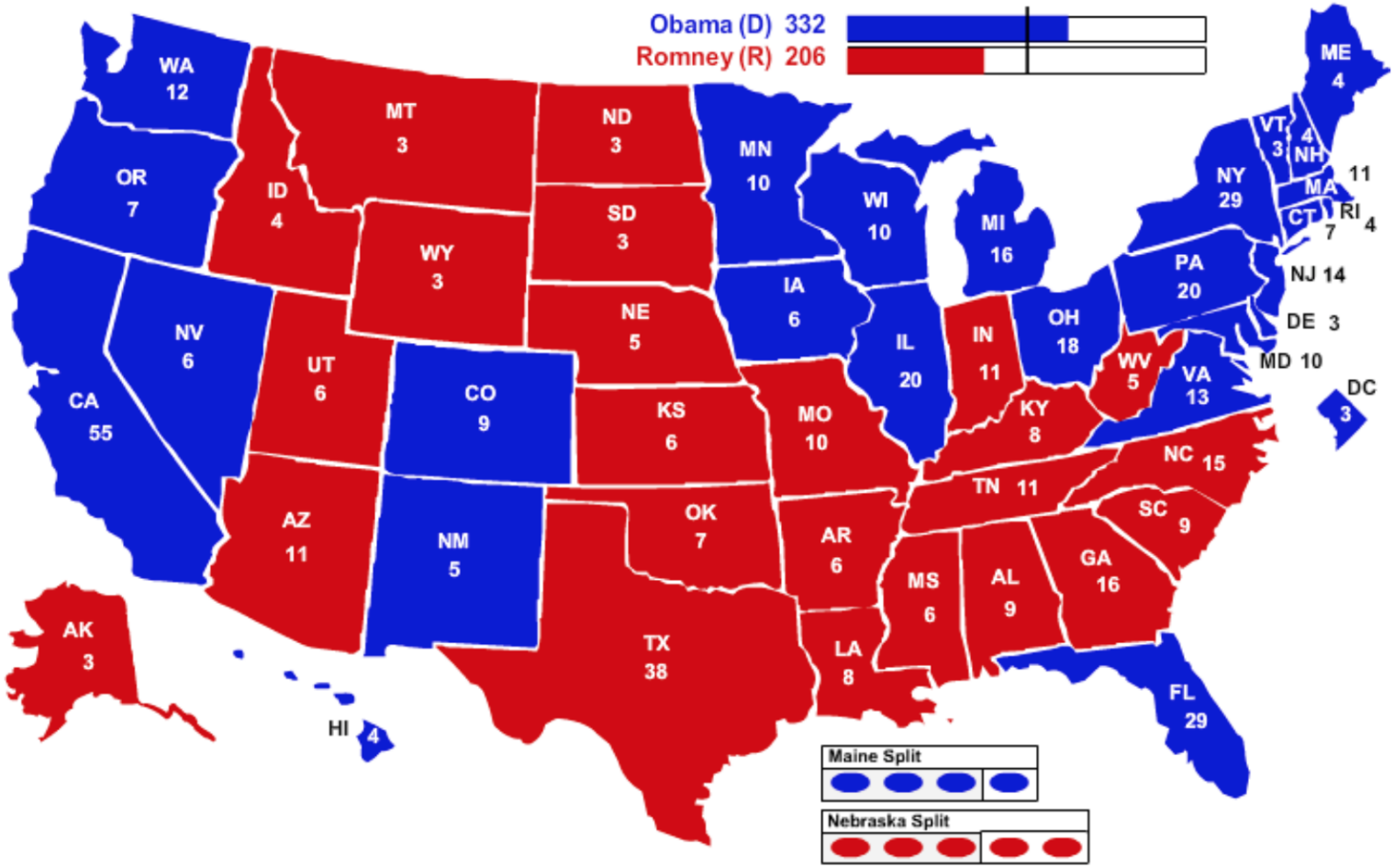


Image courtesy of the 270toWin.com

Pitfalls and Distortions of Maps

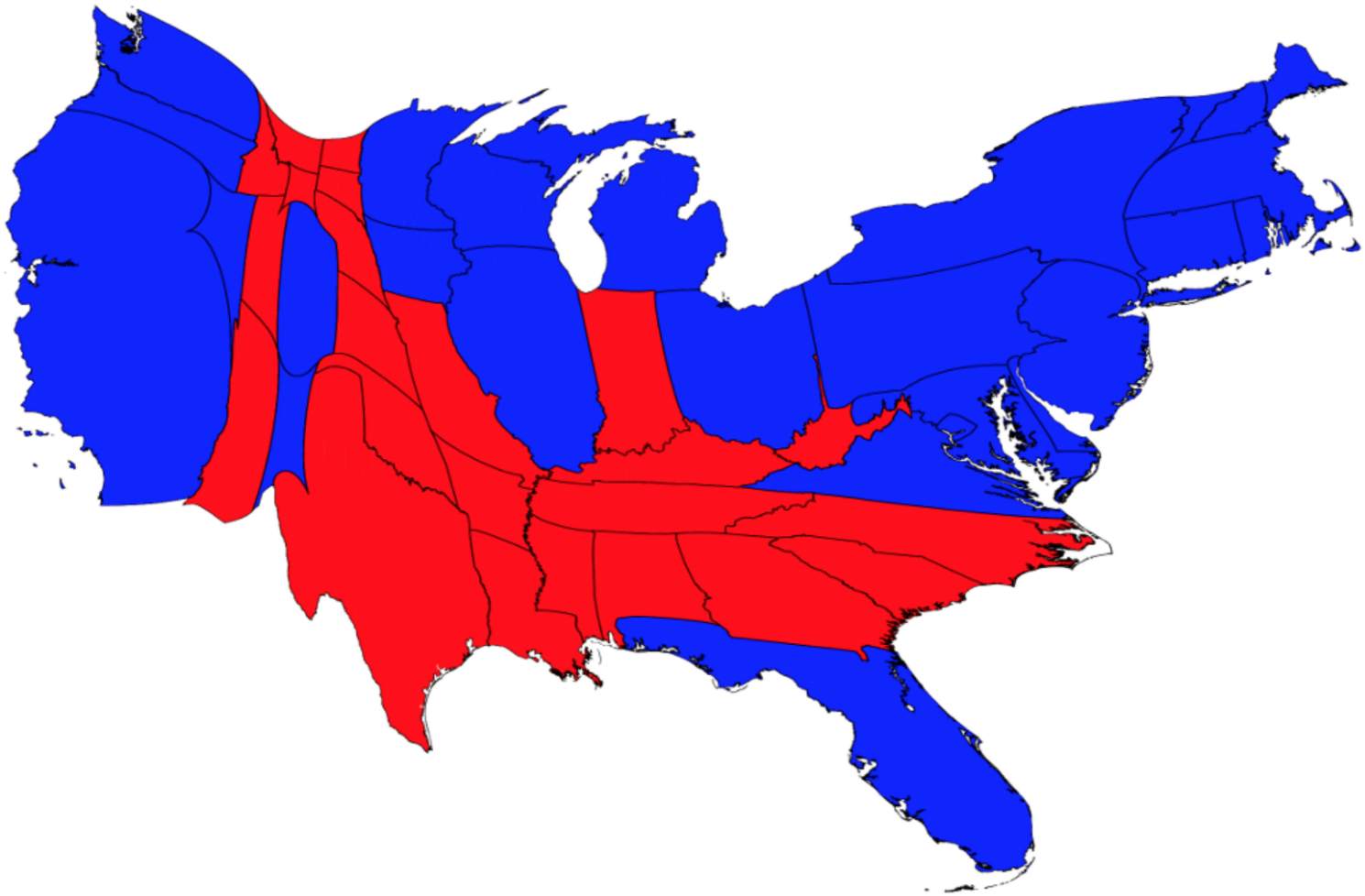


Image courtesy of Mark Newman, University of Michigan

Pitfalls and Distortions of Maps

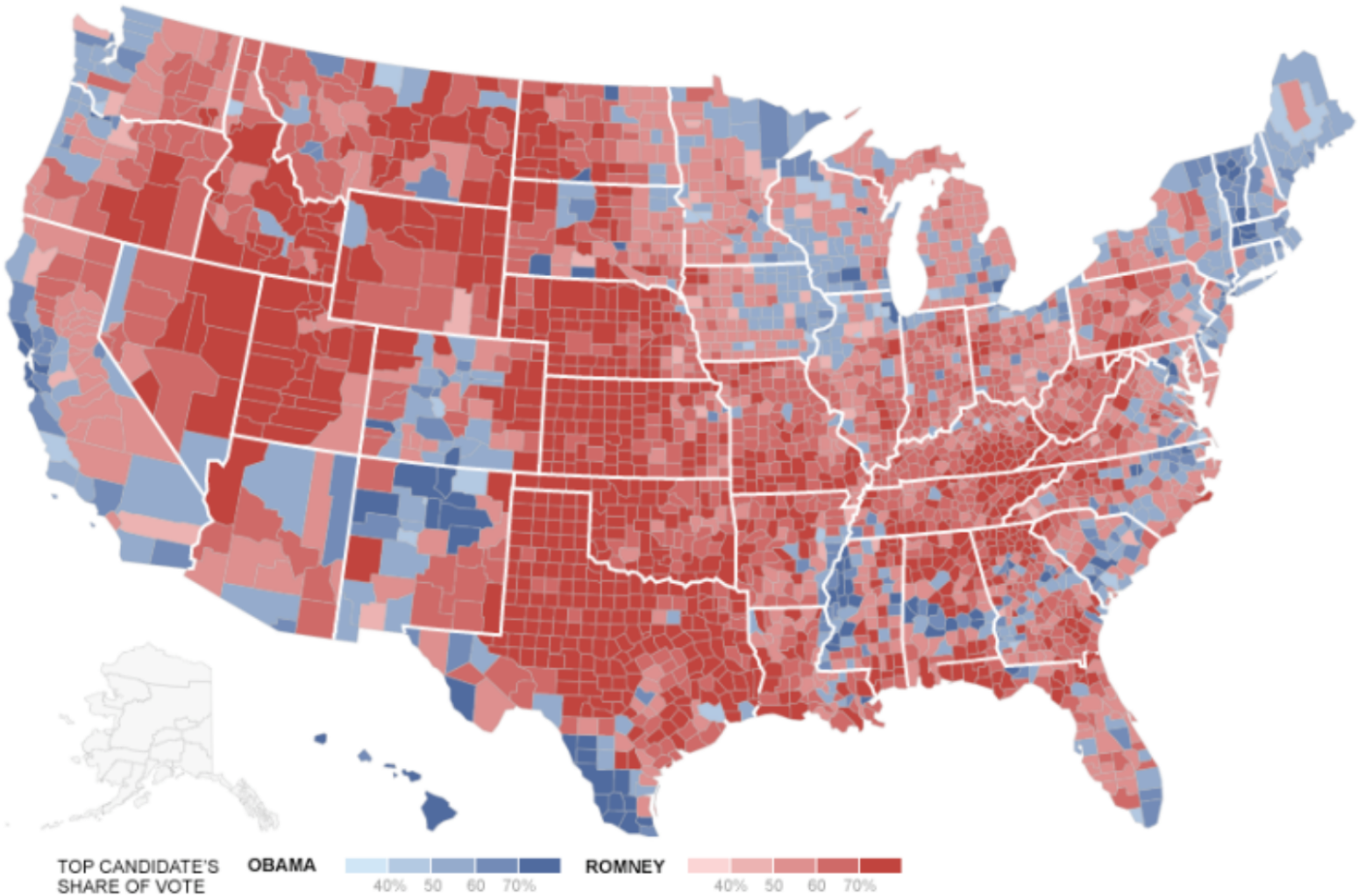


Image courtesy of the New York Times

Pitfalls and Distortions of Maps

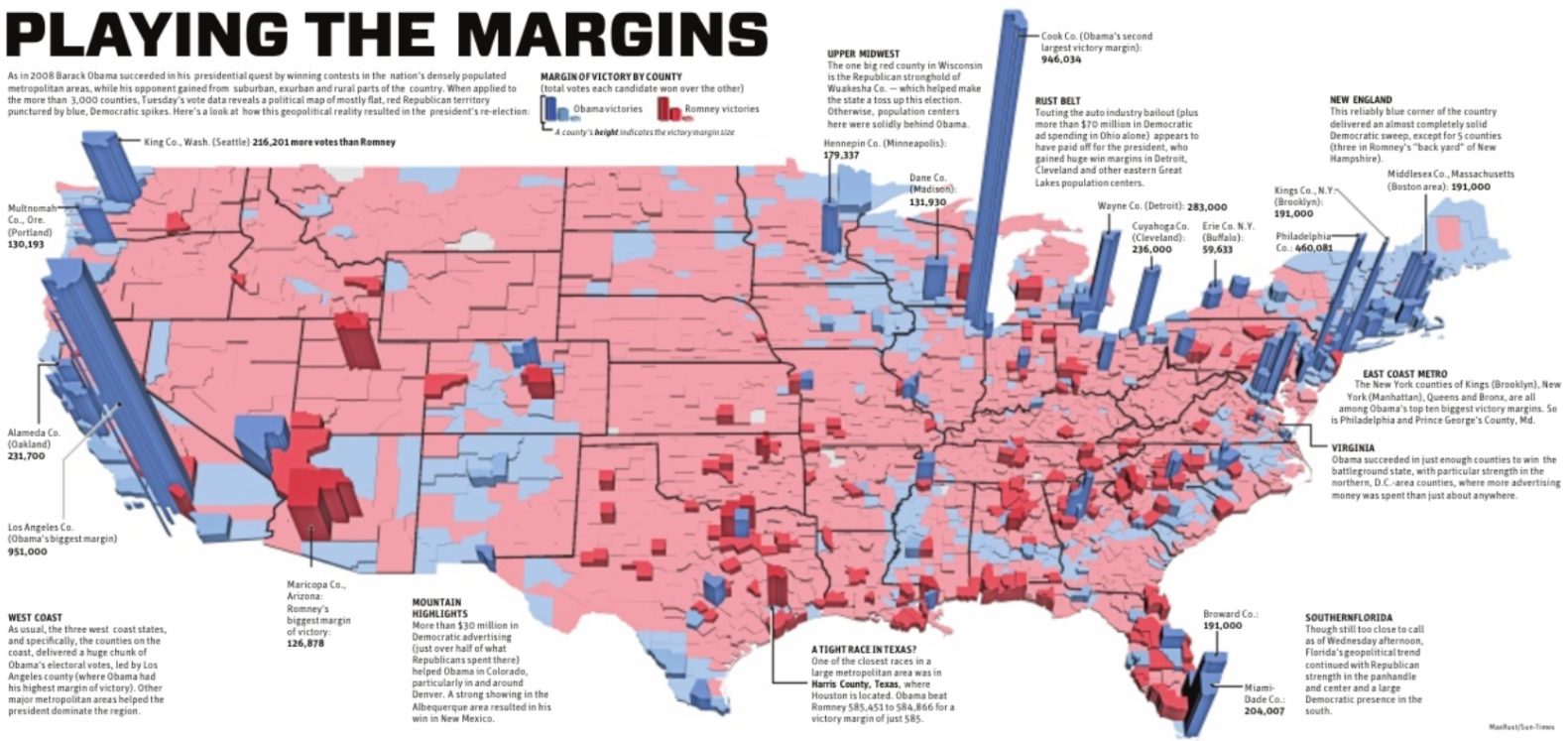


Image courtesy of the Chicago Sun Times

Pitfalls and Distortions of Maps

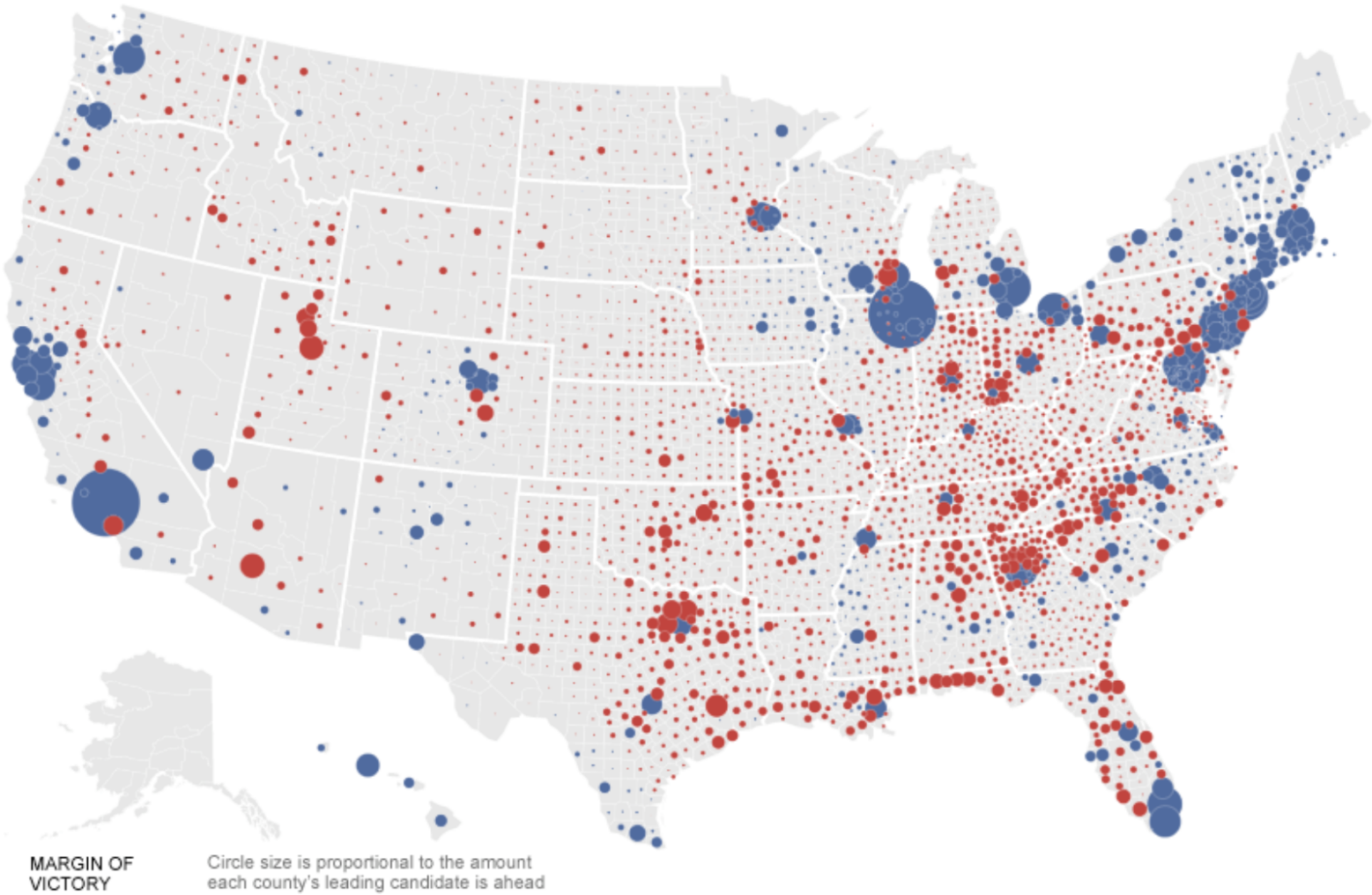


Image courtesy of the New York Times

My Personal Fave

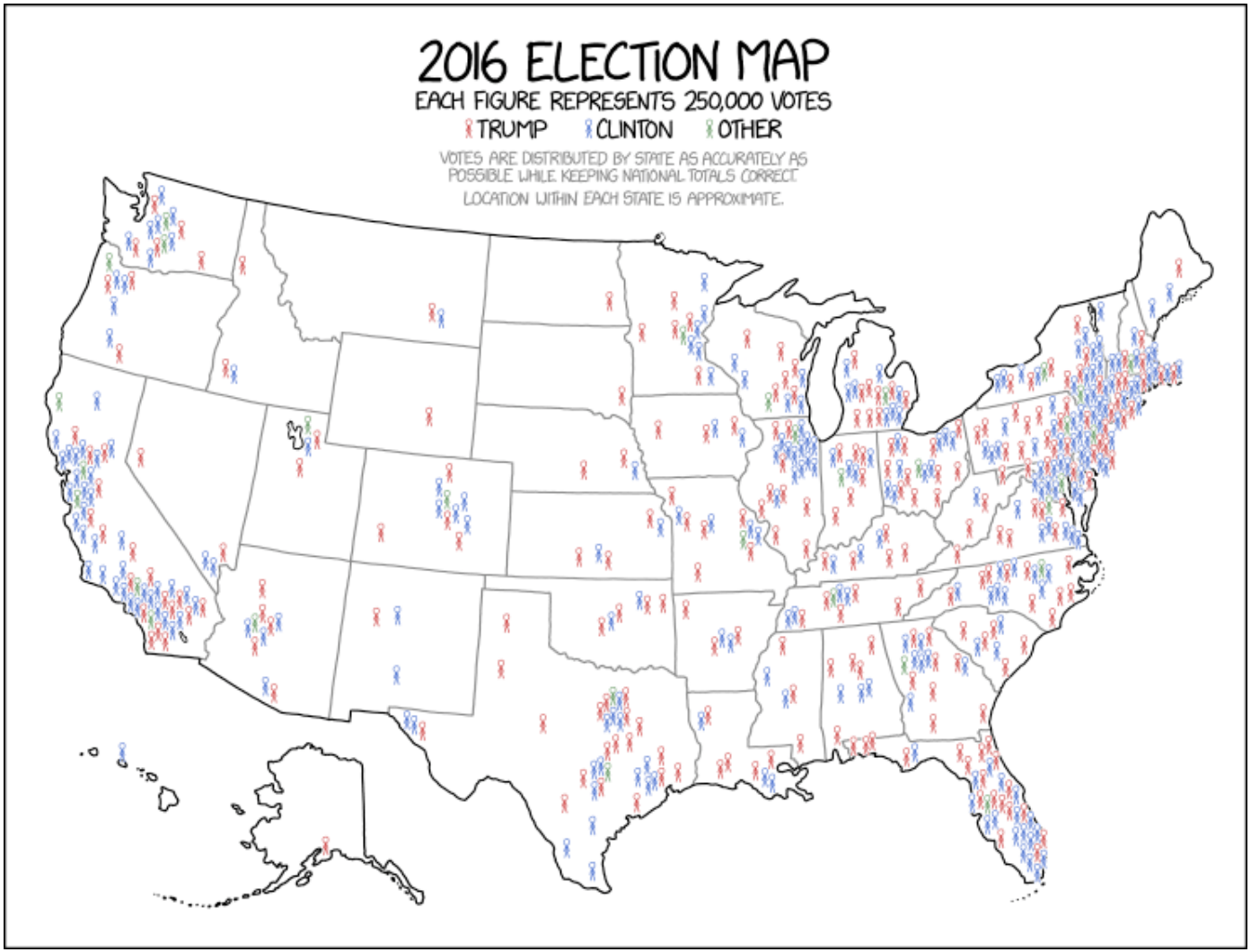


Figure: http://xkcd.com/1939/

Outline

Choose Colors Intentionally

- Color is not just about aesthetic preference! Colors communicate information
- Universal design: Keep in mind not everyone sees color the same way

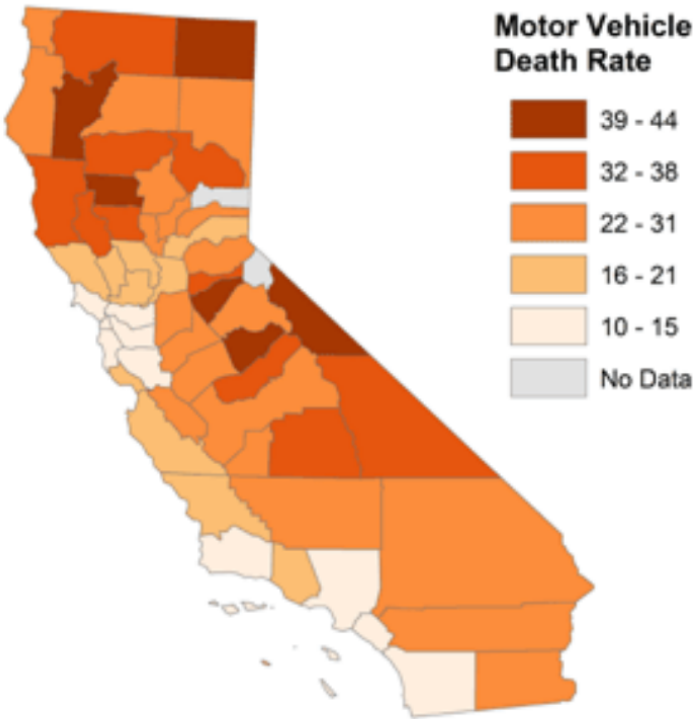
Color Coordinate Systems

- RGB (Red/Green/Blue)
  - denotes intensity of each primary color in the color
  - useful for color generation, not particularly intuitive for color perception
- HSL (Hue/Saturation/Lightness)
  - More naturally describe perceptual properties
  - Hue: the dominant "color word" in the rainbow
  - Saturation: how "pure" vs. "muted" is the color
  - Lightness: how much white or black is in the color

Taxonomy of Color Palettes

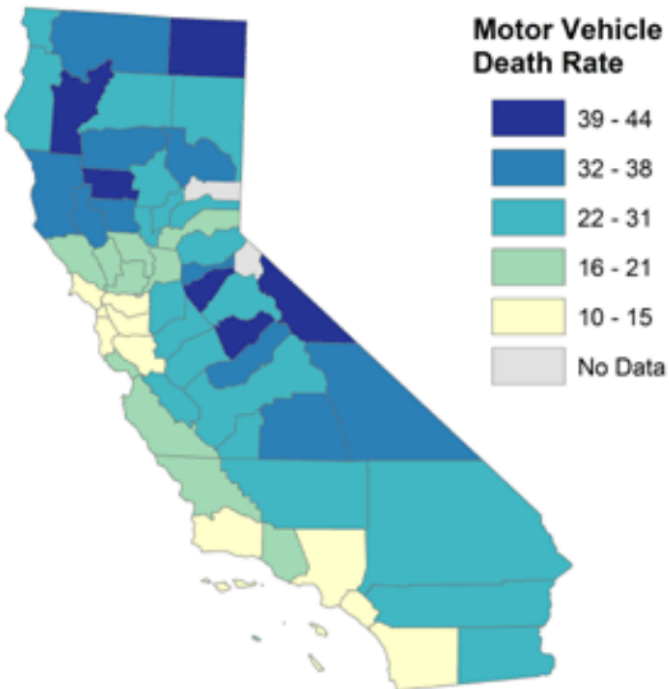
- Sequential: Colors fall on an ordered scale
  - Single hue: Uses only saturation/lightness to distinguish
  - Multi hue: Varies colors on a hue spectrum
- Diverging: Colors move two directions from a neutral point
- Categorical: No sense of a quantitative scale; mapping for categorical variables only

Sequential Palette: Single Hue



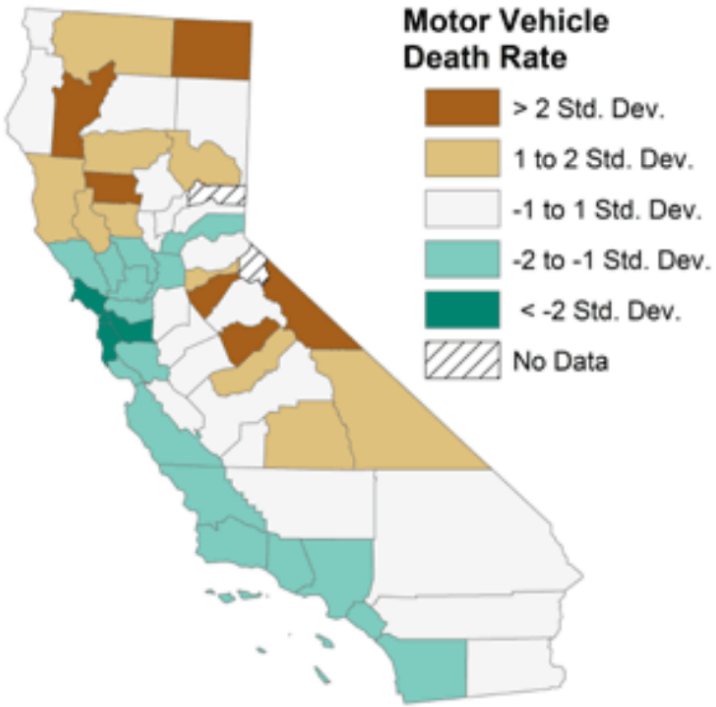
- Maps a numeric variable to the *saturation* dimension (sometimes combination of saturation and lightness)
- Higher saturation = More of the thing

Sequential Palette: Multi Hue



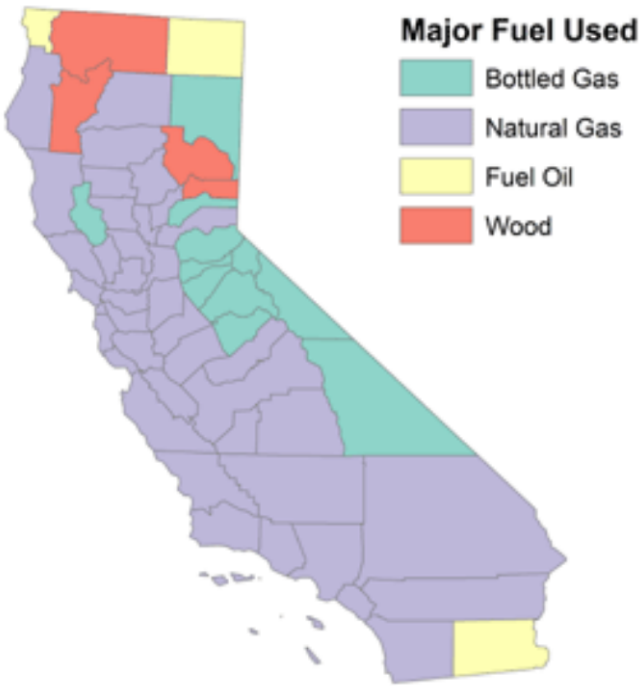
- Maps a numeric variable to an arc in HSB color space, varying hue
- Typically, darker = more

Diverging Palette



- Two sequential (single hue) scales that "meet in the middle" at a neutral (low saturation) color
- Useful for signed quantitative variable with anchor at zero
- Careful to distinguish missing data!

Categorical Palette



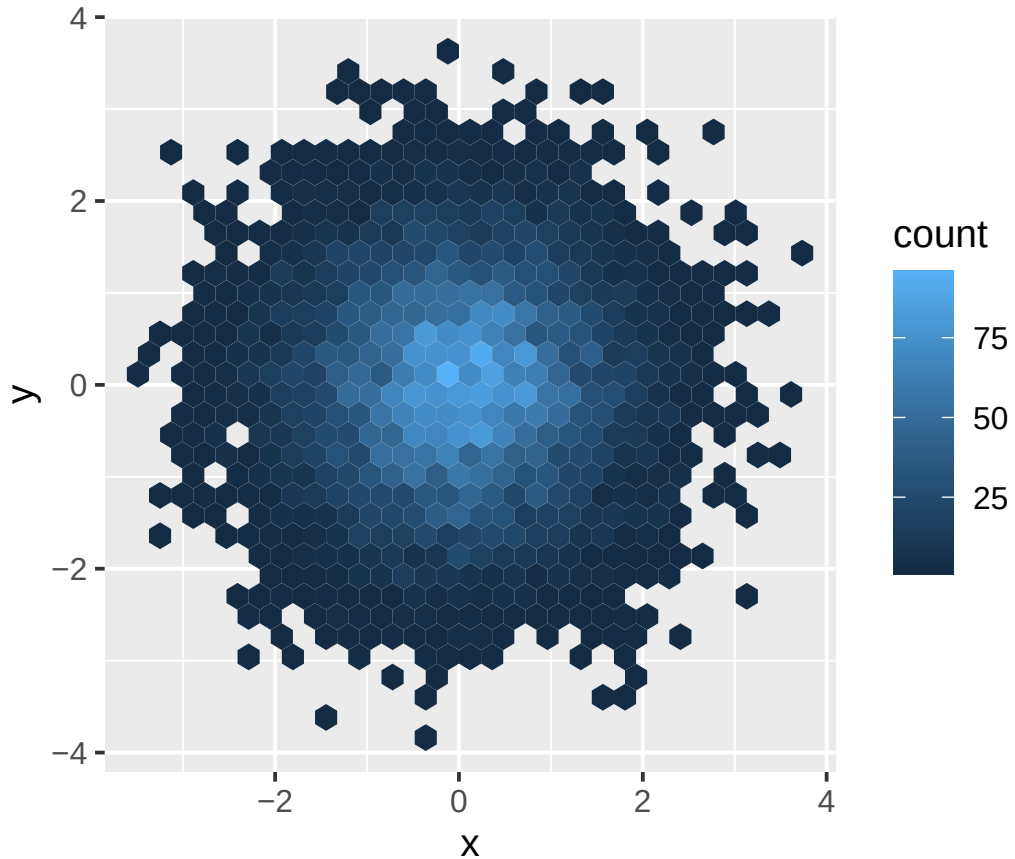
- A set of (roughly) equally spaced points in a circle in HSB space (i.e., only hue varies)
- Colors indicate levels of a *categorical* variable

Some convenient built-in options

- `RColorBrewer`
  - by Cynthia Brewer (`http://www.colorbrewer2.org`)
  - has sequential, diverging, categorical palettes
- `viridis`
  - Rudis, Ross and Garnier (link to CRAN documentation)
  - Based on color schemes in Python's `matplotlib`
  - Sequential (multi-hue) only

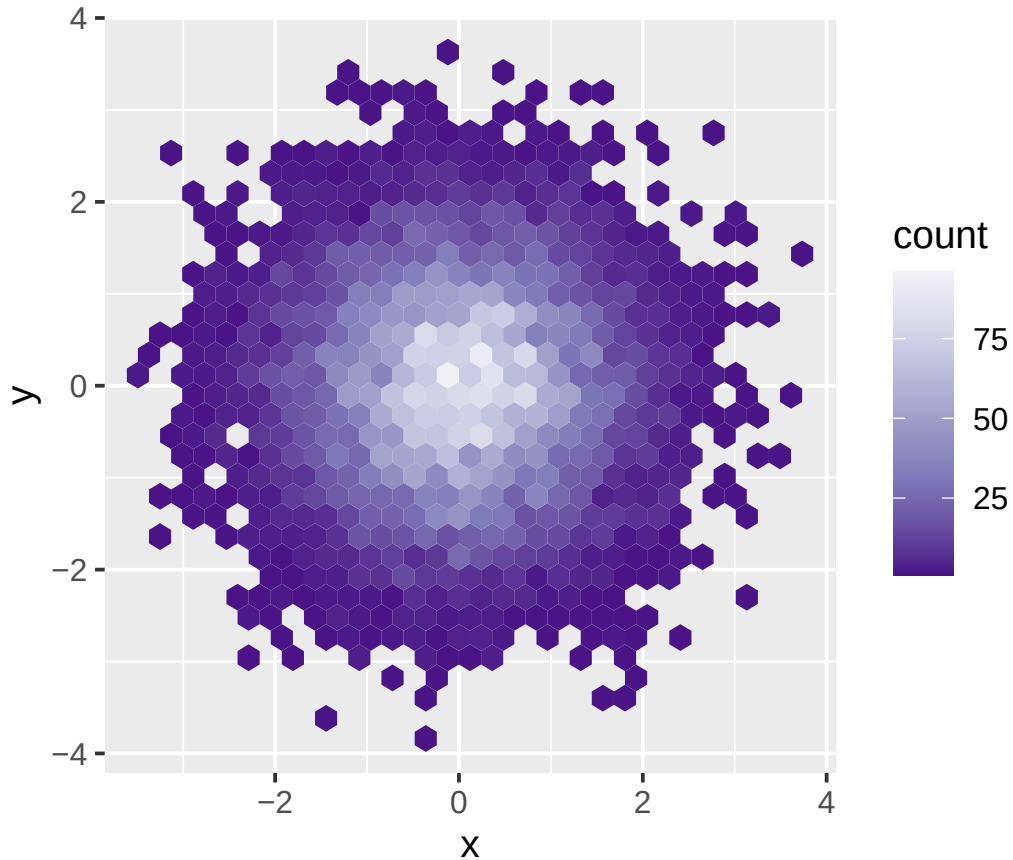
ggplot2 default

```
library(tidyverse)
rand_data <- data.frame(x = rnorm(10000), y = rnorm(10000))
norm_plot <- ggplot(rand_data, aes(x = x, y = y)) +
  geom_hex() + coord_fixed()
norm_plot
```



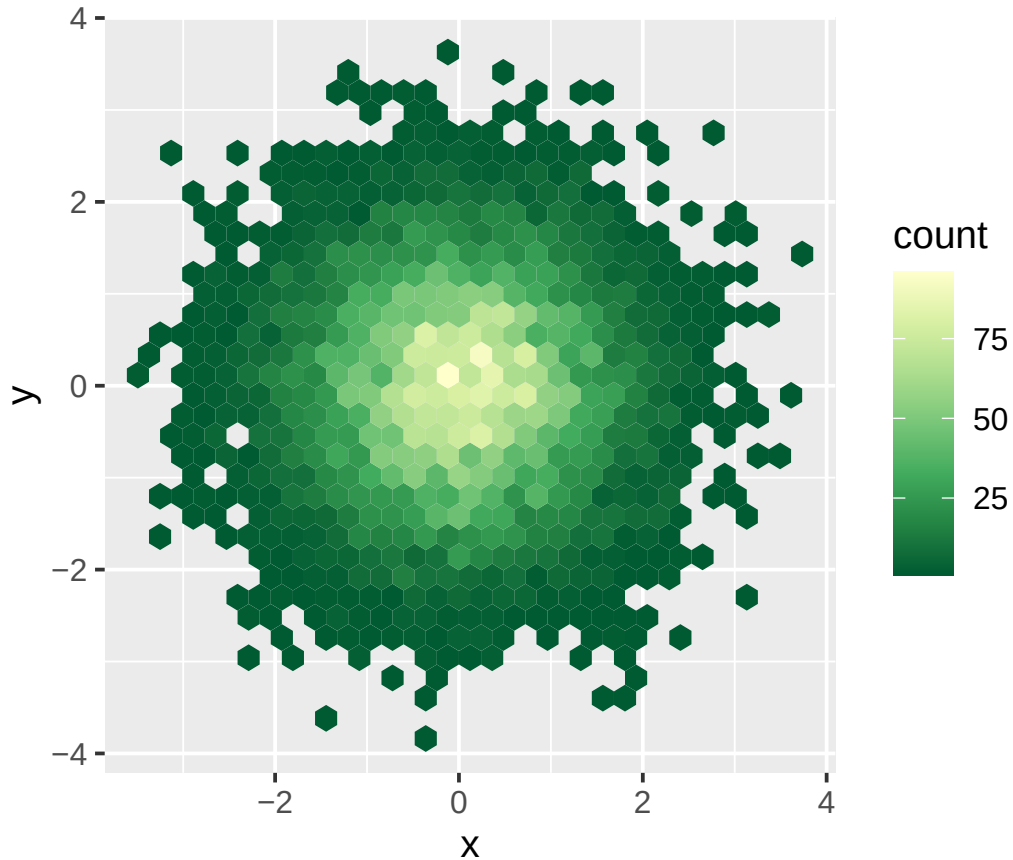
ColorBrewer Purples palette

```
library(RColorBrewer)
norm_plot + scale_fill_distiller(palette = "Purples")
```



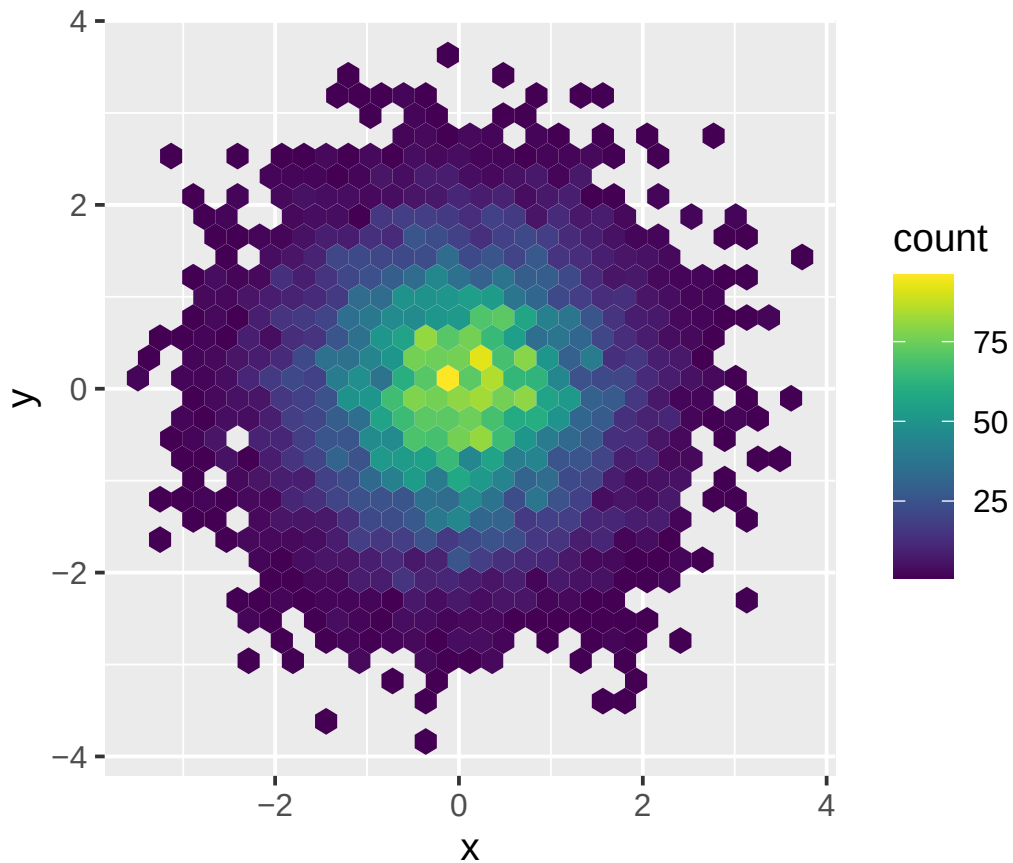
ColorBrewer YlGn palette

```
library(RColorBrewer)
norm_plot + scale_fill_distiller(palette = "YlGn")
```



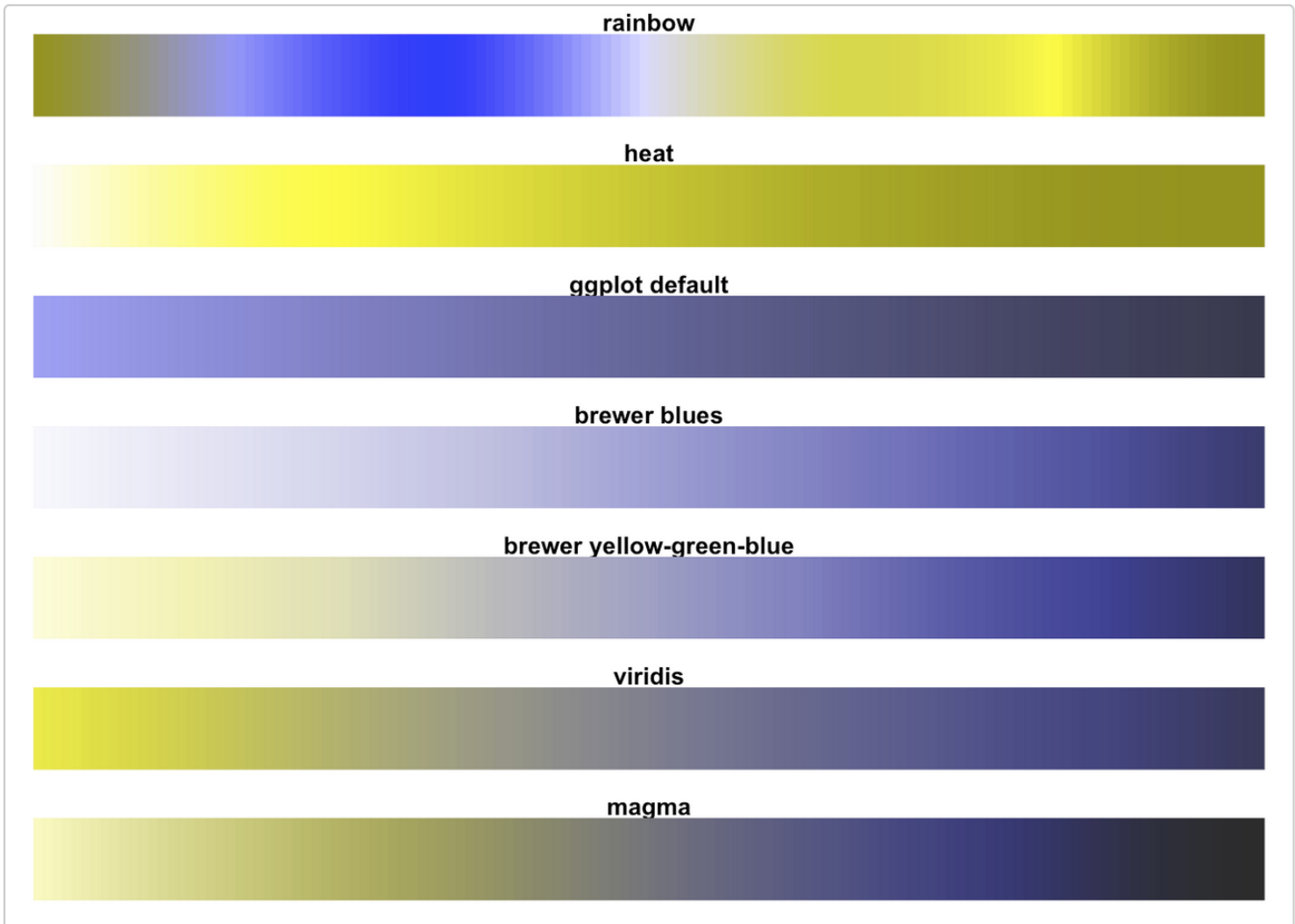
viridis palette

```
library(viridis)
norm_plot + scale_fill_viridis()
```



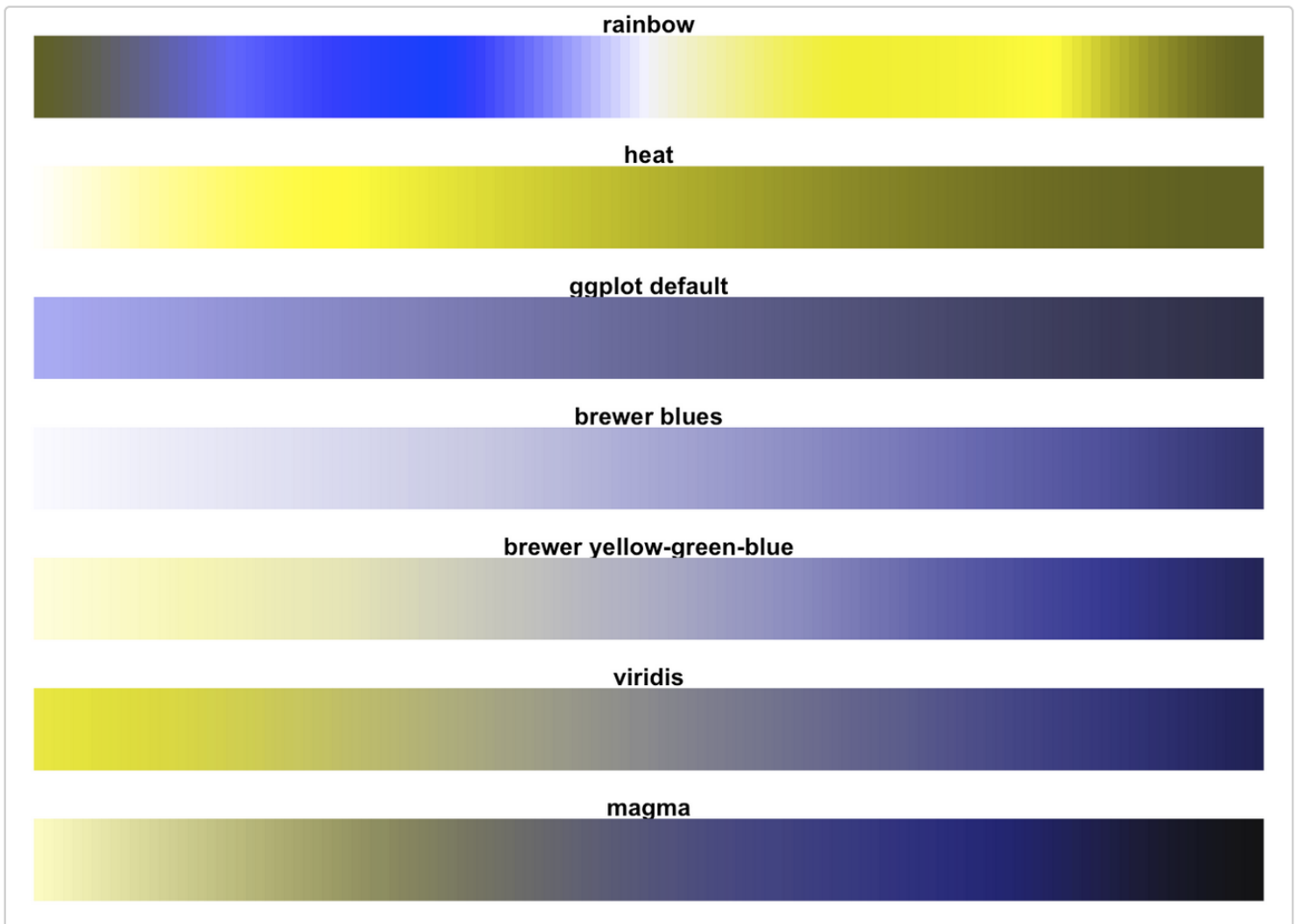
Universal Design: Color Blindness

Green-Blind (Deuteranopia)

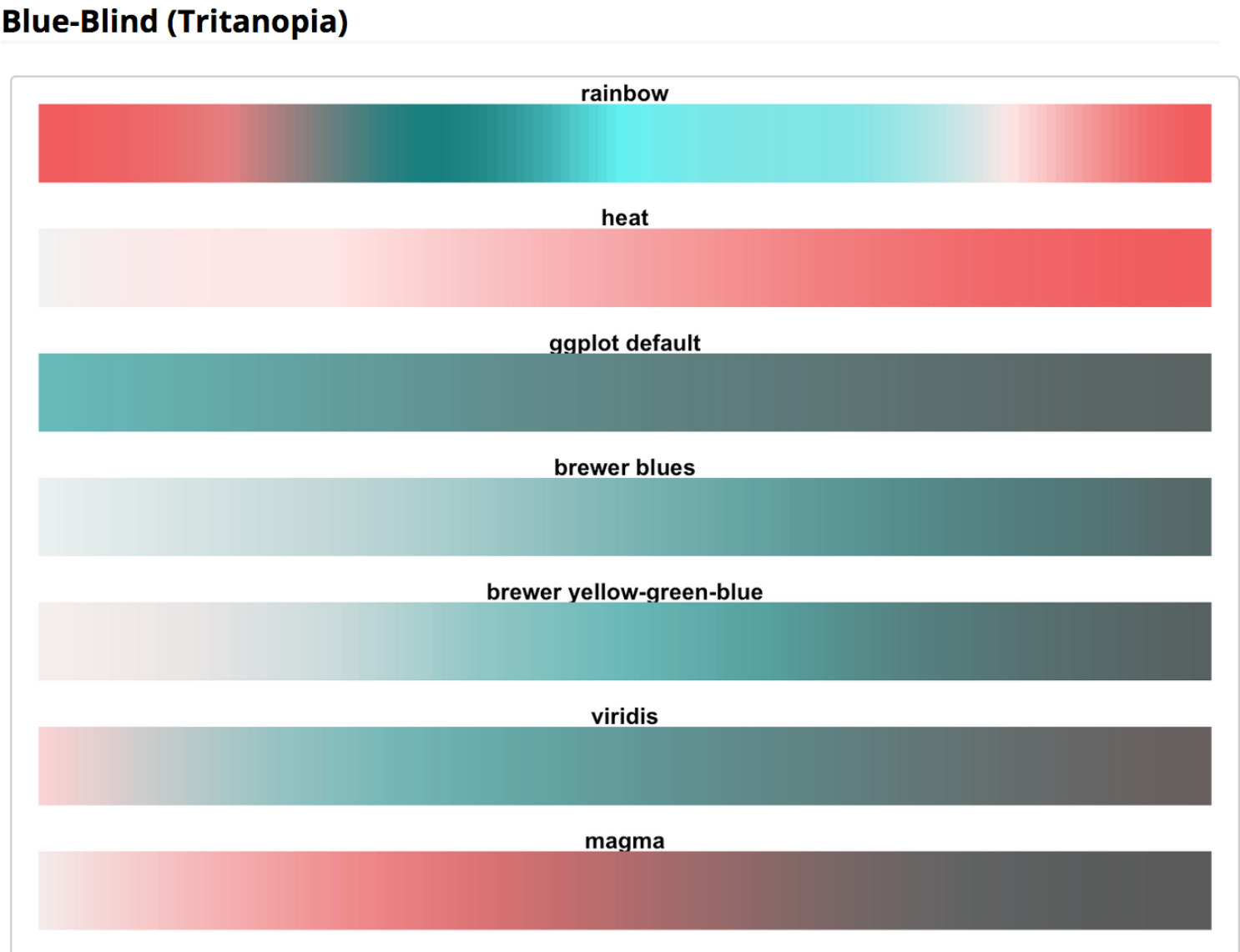


Universal Design: Color Blindness

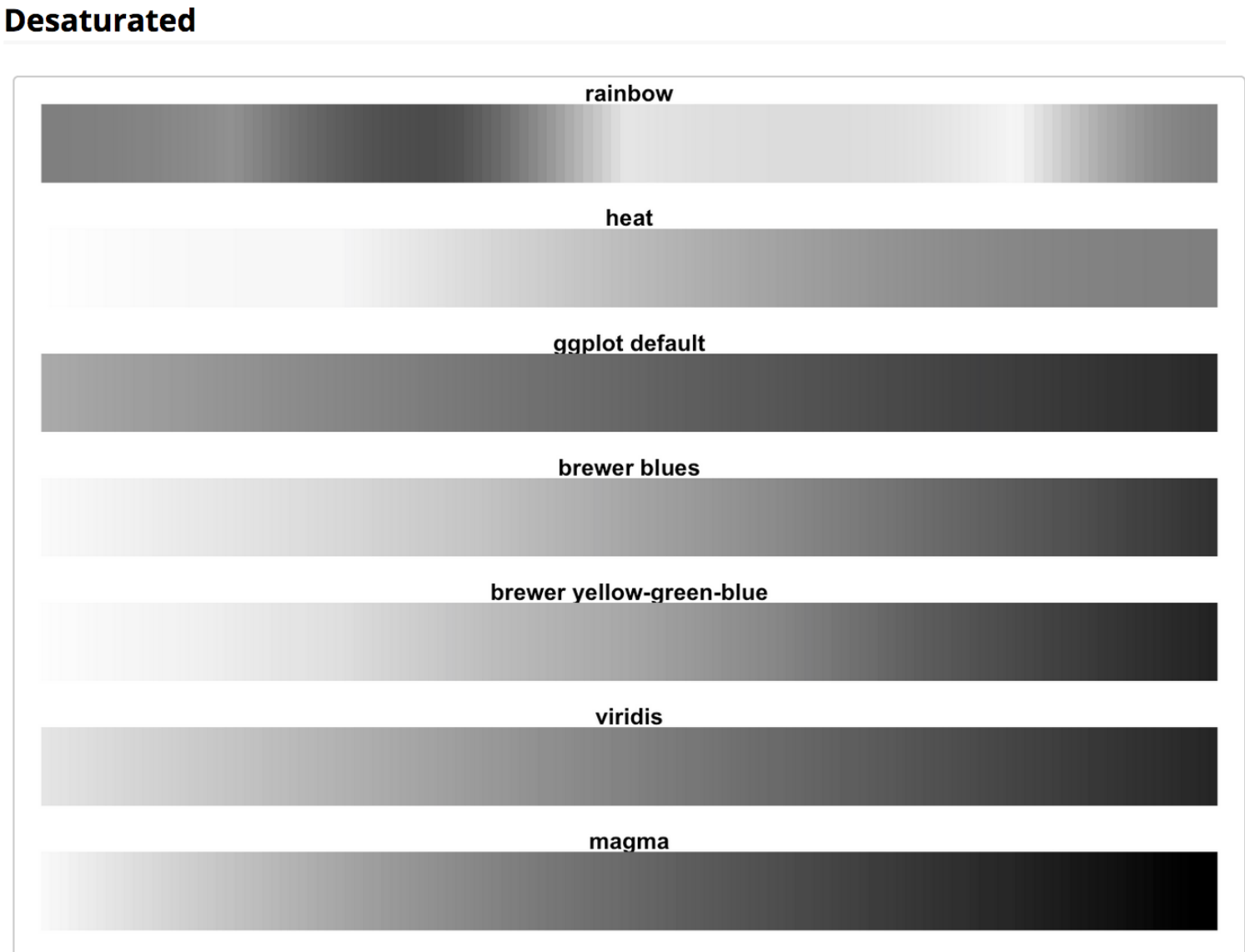
Red-Blind (Protanopia)



Universal Design: Color Blindness



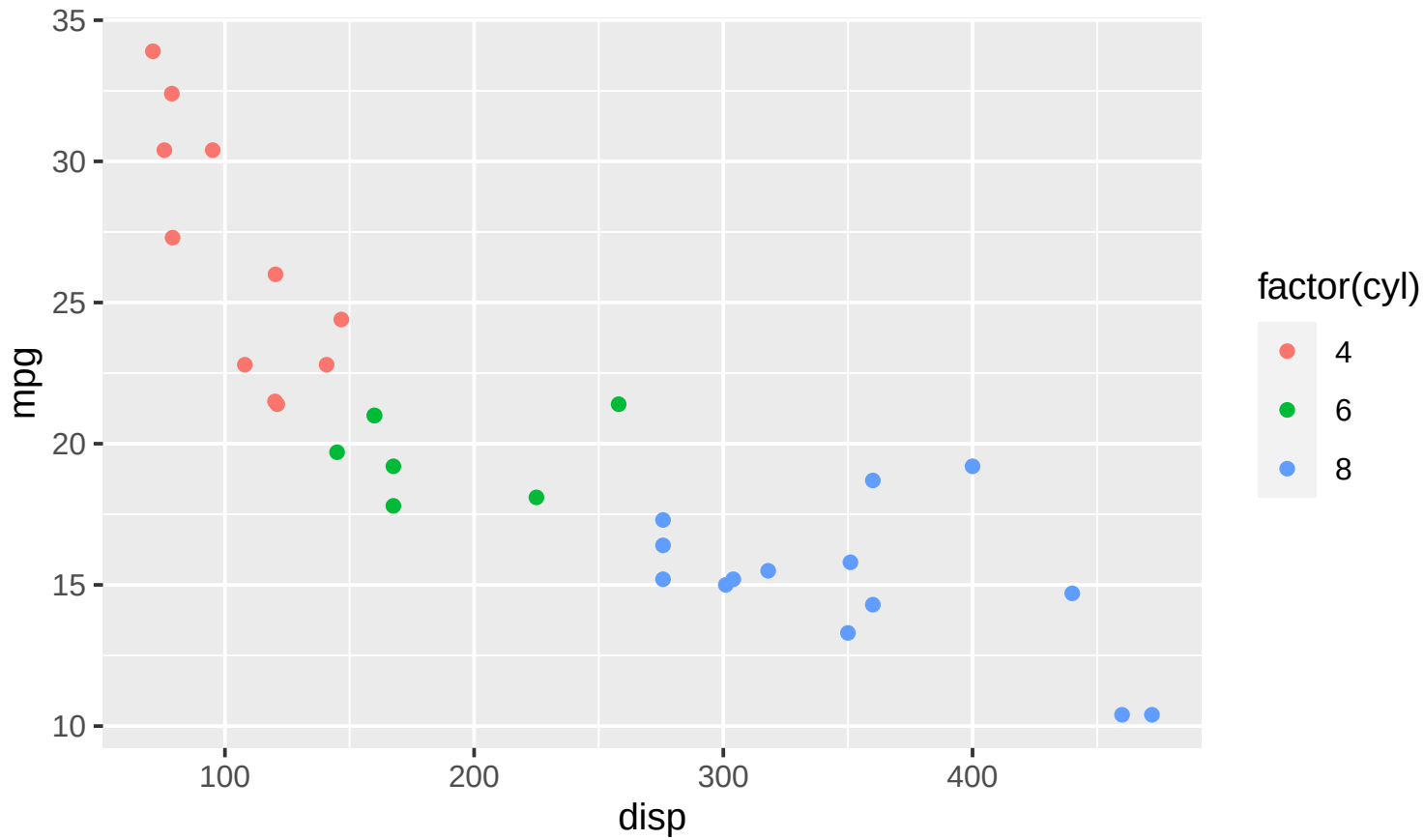
Universal Design: Color Blindness



Outline

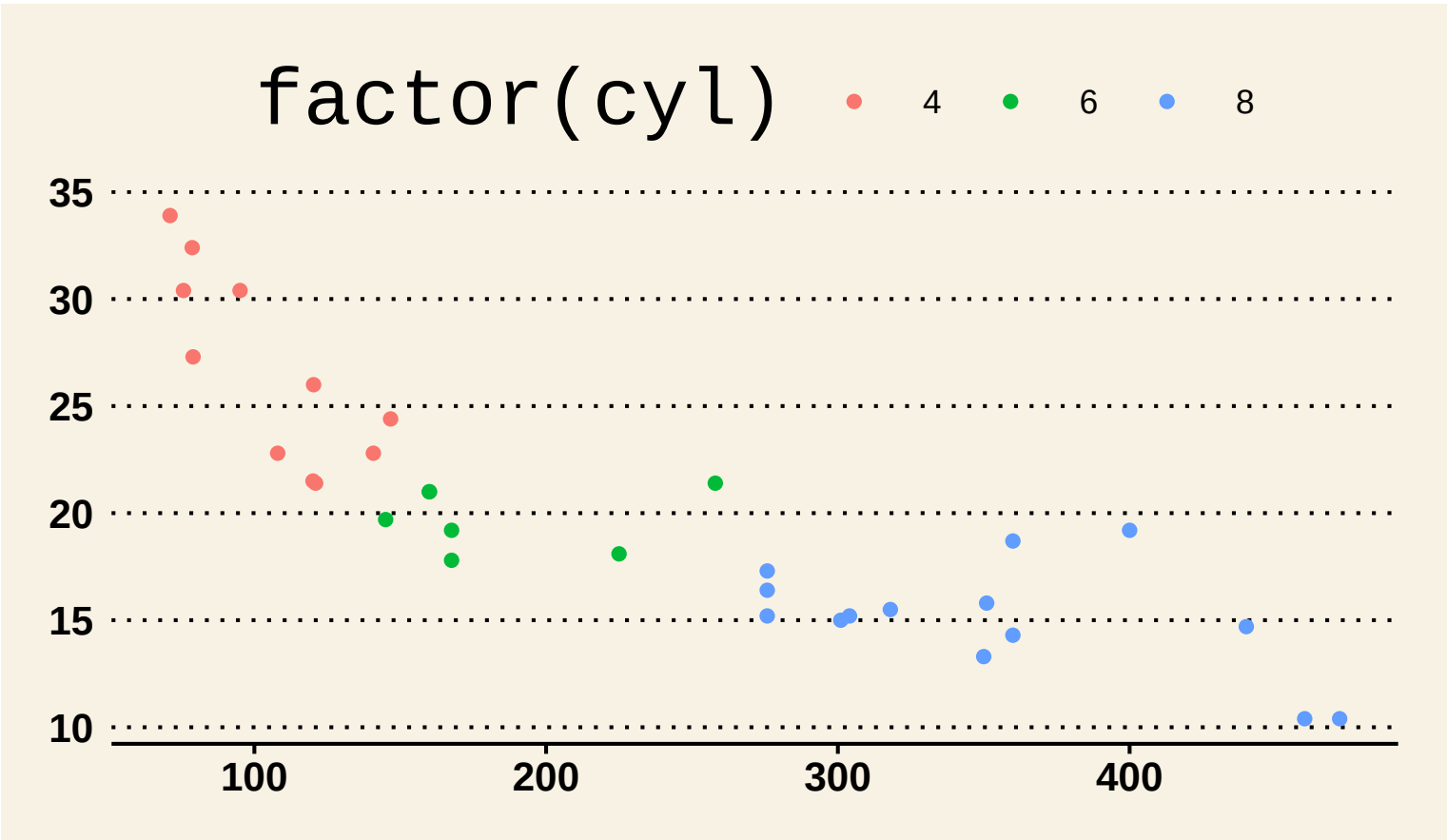
ggplot2 themes

```
library(tidyverse)
p1 <- ggplot(mtcars, aes(x = disp, y = mpg, color = factor(cyl))) +
  geom_point()
p1
```



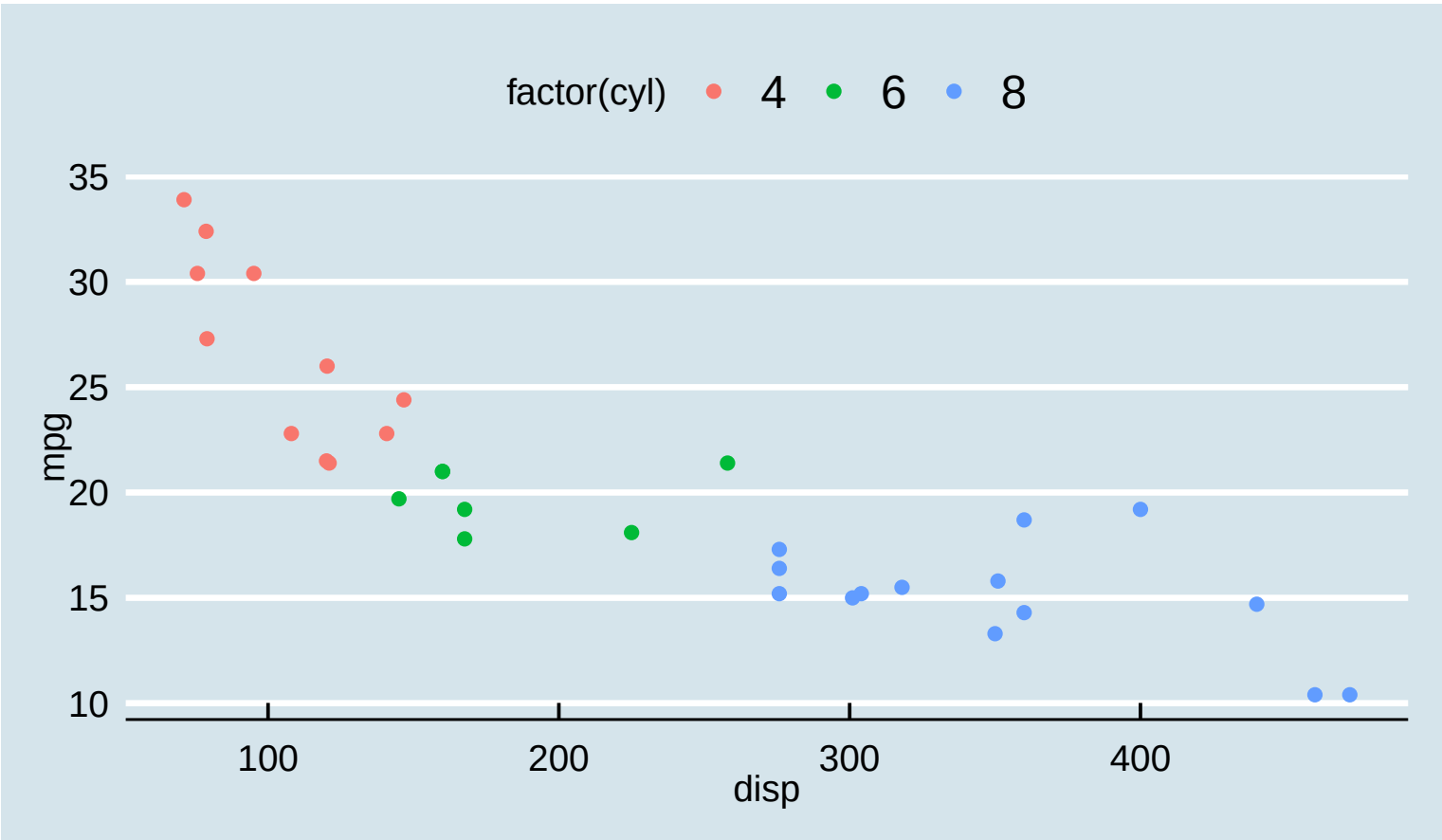
Wall Street Journal Theme

```
library(ggthemes)
p1 + theme_wsj()
```



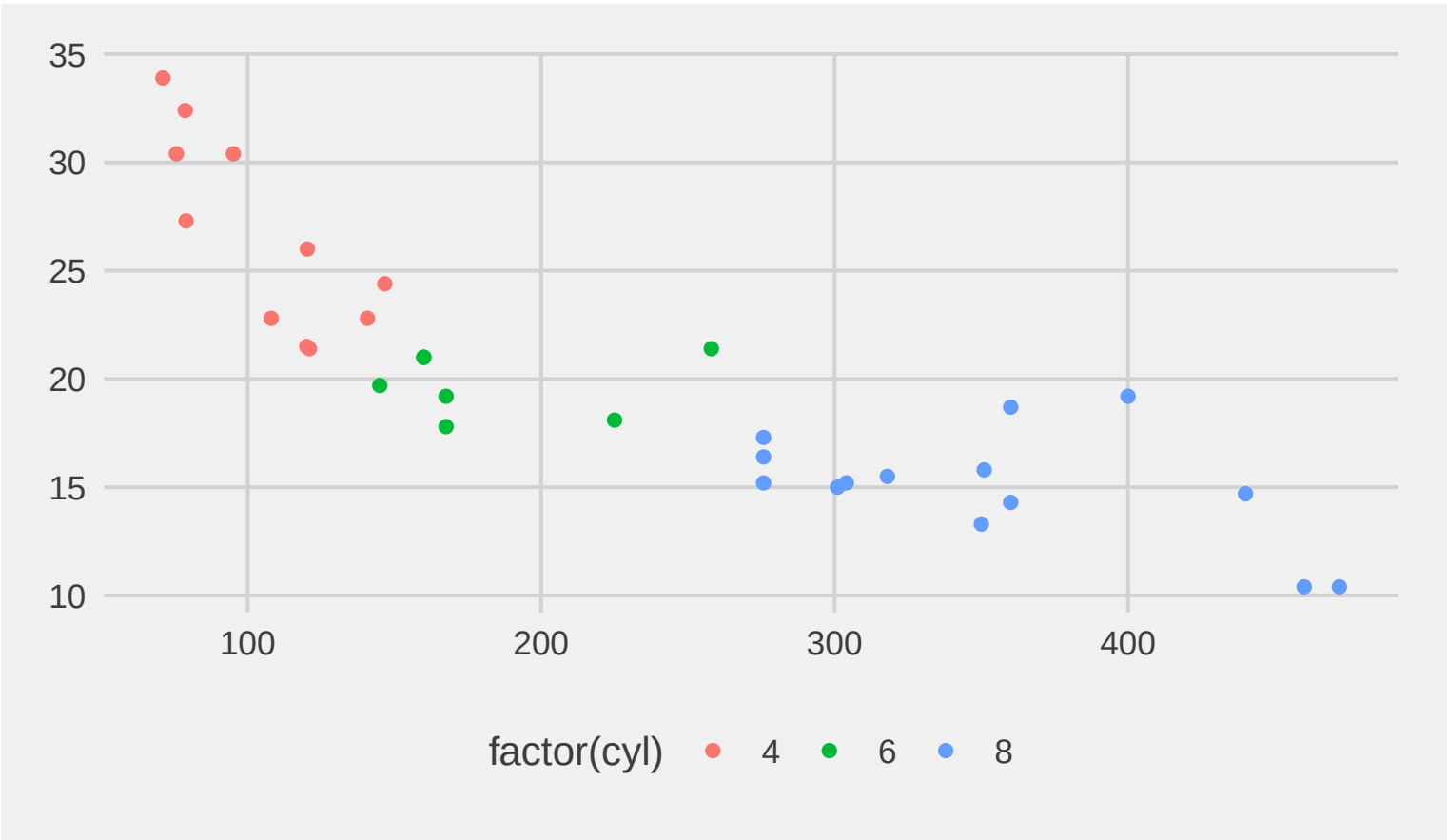
Economist Theme

```
library(ggthemes)
p1 + theme_economist()
```



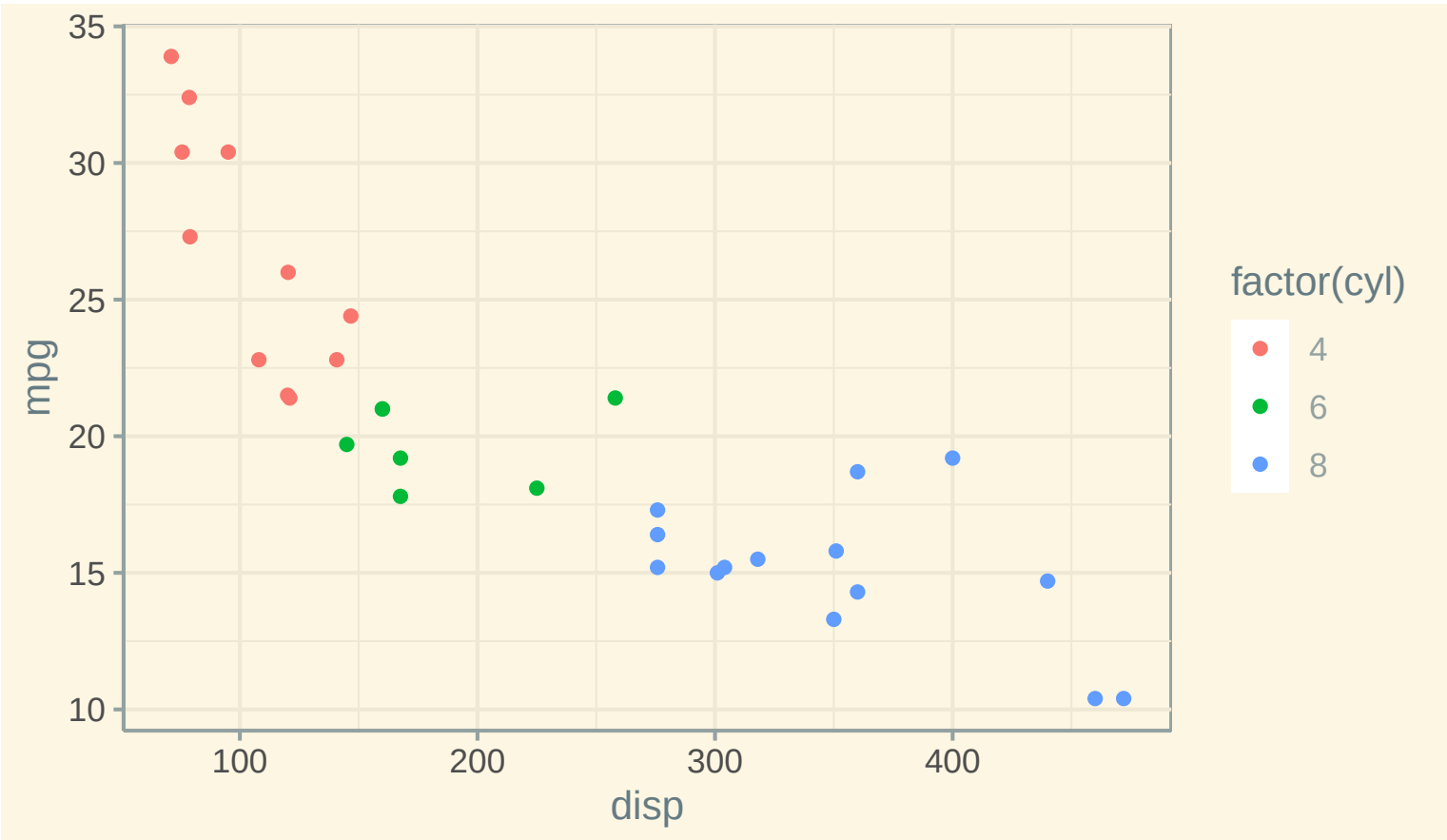
FiveThirtyEight Theme

```
library(ggthemes)
p1 + theme_fivethirtyeight()
```



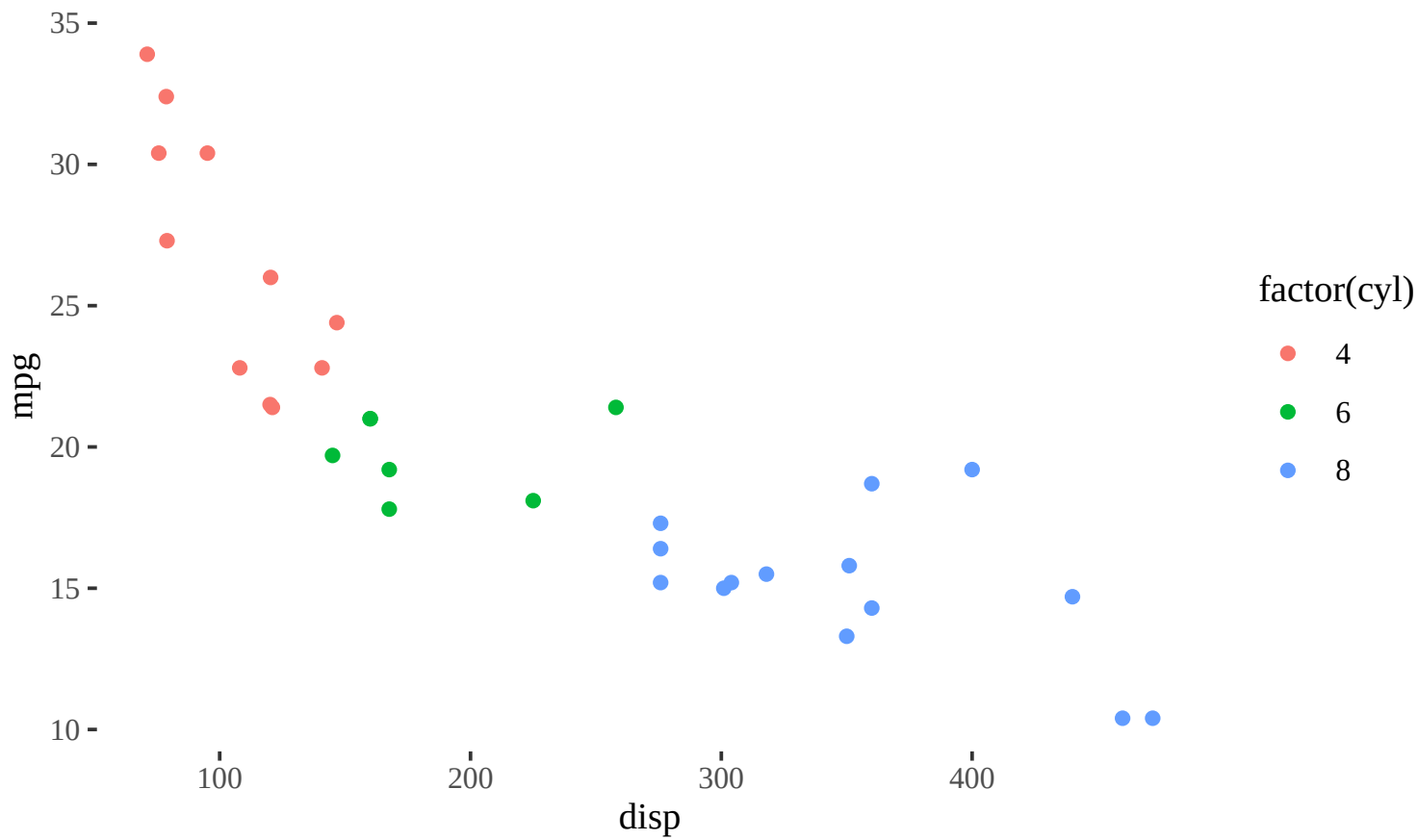
Solarized Theme

```
library(ggthemes)
p1 + theme_solarized()
```



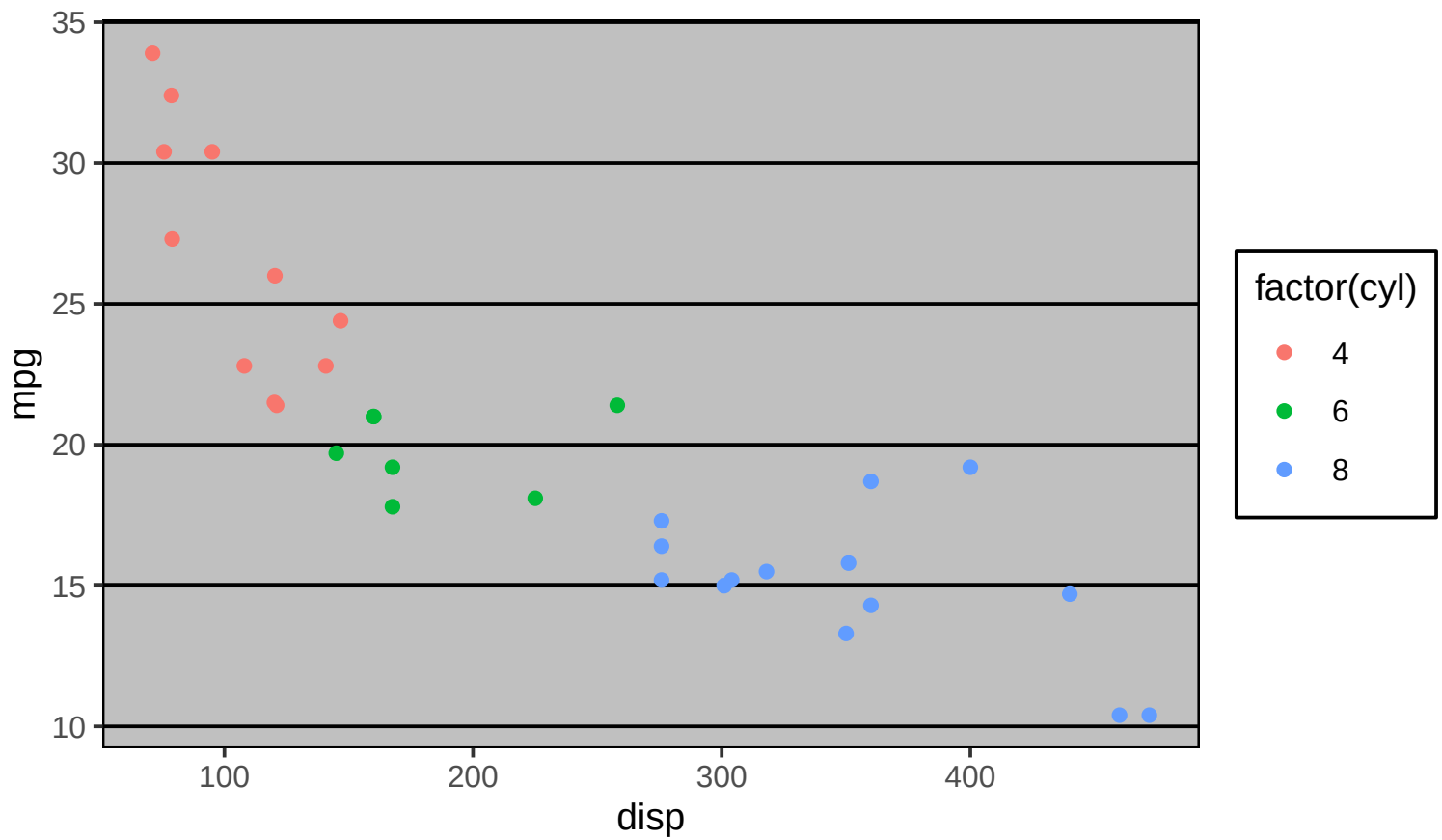
Edward Tufte's Minimalist Theme

```
library(ggthemes)
p1 + theme_tufte()
```



Why Does This Exist?

```
library(ggthemes)
p1 + theme_excel()
```



Many More...

- ggthemes package documentation link