

# Data Visualization with ggplot2

@erikaris

July 14, 2019

# Why do we visualize?

- ▶ to understand relationship between variables.

# How to visualize in R?

- ▶ basic R plot
- ▶ ggplot2

# What is ggplot2?



- ▶ ggplot2 is a system for declaratively creating graphics
- ▶ it is based on The Grammar of Graphics.
  - ▶ Grammar of Graphics is a concept that defines a plot as a set of component layers: **aesthetic and geometry**.
  - ▶ You **provide the data**, tell ggplot2 how to map variables to **aesthetics**, what graphical **geometries** to use, and it takes care of the details.

## Drawing your first plot

- ▶ Make sure you have the library installed on your machine:

```
install.packages('ggplot2')
```

- ▶ load the ggplot library:

```
r library(ggplot2)
```

- ▶ use the dataset mtcars, available in R Studio.

- ▶ explore its structure:

```
## 'data.frame':  32 obs. of  11 variables:
## $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22
## $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
## $ disp: num  160 160 108 258 360 ...
## $ hp  : num  110 110 93 110 175 105 245 62 95 123 .
## $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69
## $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
## $ qsec: num  16.5 17 18.6 19.4 17 ...
## $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
```

You can convert the temperature unit from Kelvin to Celsius with the formula

$$celsius = kelvin - 273.15$$

And you can convert the result to Fahrenheit with the formula

$$fahrenheit = celsius \times \frac{9}{5} + 32$$

## Cleaning

For the remainder of the report, we will look only at data from the year 2000. We aggregate our data by location, using the *R* code below.

```
means <- atmos %>%  
  filter(year == year) %>%  
  group_by(long, lat) %>%  
  summarize(temp = mean(temp, na.rm = TRUE),  
            pressure = mean(pressure, na.rm = TRUE),  
            ozone = mean(ozone, na.rm = TRUE),  
            cloudlow = mean(cloudlow, na.rm = TRUE),  
            cloudmid = mean(cloudmid, na.rm = TRUE),  
            cloudhigh = mean(cloudhigh, na.rm = TRUE)) %>%  
  ungroup()  
  
clouds <- means %>%  
  select(-(temp:ozone)) %>%  
  gather("altitude", "coverage", 3:5)
```