# Data Visualization with ggplot2

Session 1: The Grammar of Graphics

Nathan Barron
nathanbarron@ou.edu

# Graphics are like onions

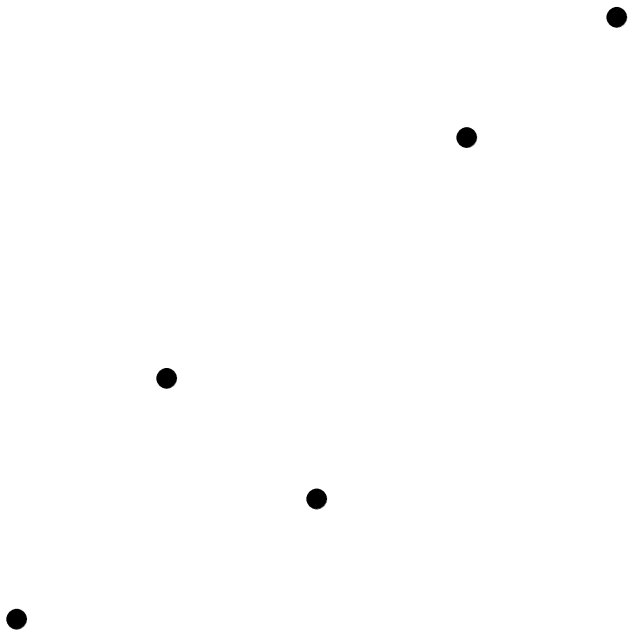
The `ggplot2` package in R is designed with a **grammar**, meaning that there are *core principles* that can be applied to create nearly *any* type of graphic.

- Data: the information you want to visualize
- Mapping: the description of how the data's variables are "mapped" to aesthetic attributes

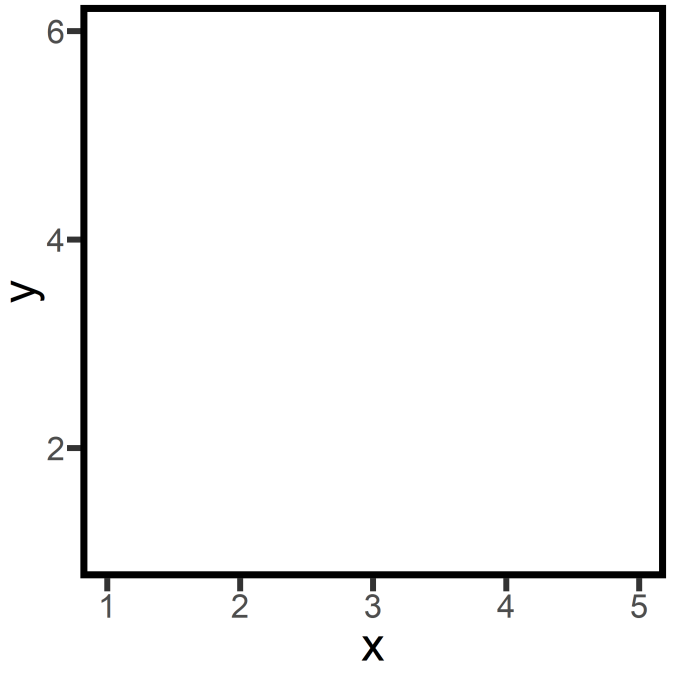

**Essential Mapping Components**

- Layer: collection of geometric elements (or *geom*s)
- Geom: visual components of a plot (e.g. points, lines, bars, etc.)
- Scales: map values in the data space to values in the aesthetic space (e.g. color, shape, size, etc.)
- Coord: coordinate system (e.g., the axes and gridlines)
- Facet: how to display subsets of data
- Theme: display features (e.g. font size, background color, etc.)

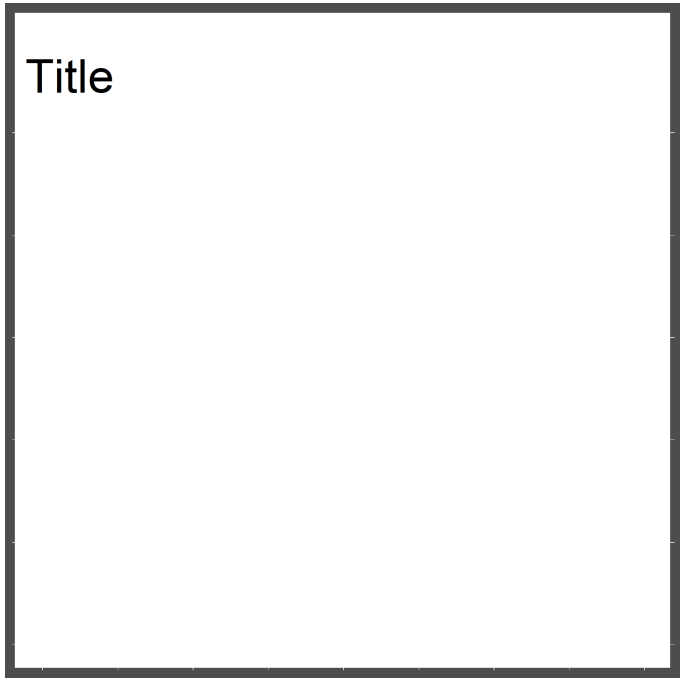
# Plot layers



These are the `geom_point` elements. They are the visual representations of your data.



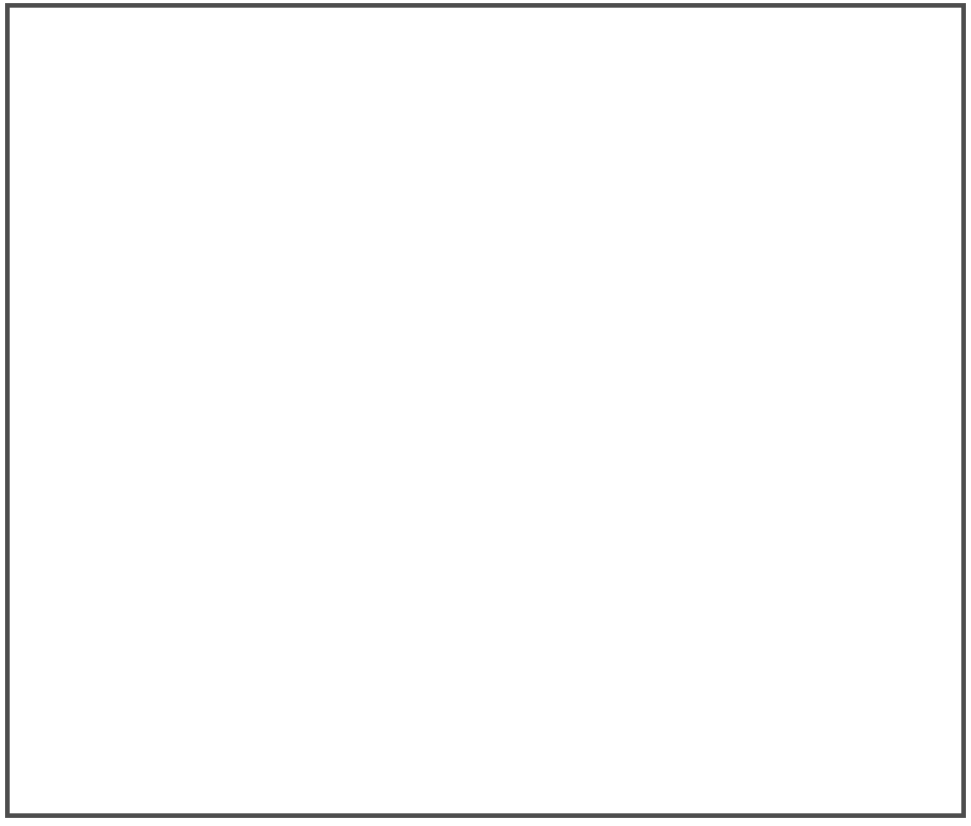

These are the scales and coordinate system. They are the axes and legends so that we can read values from the graph




These are the plot annotations. They provide crucial information for easy interpretation.

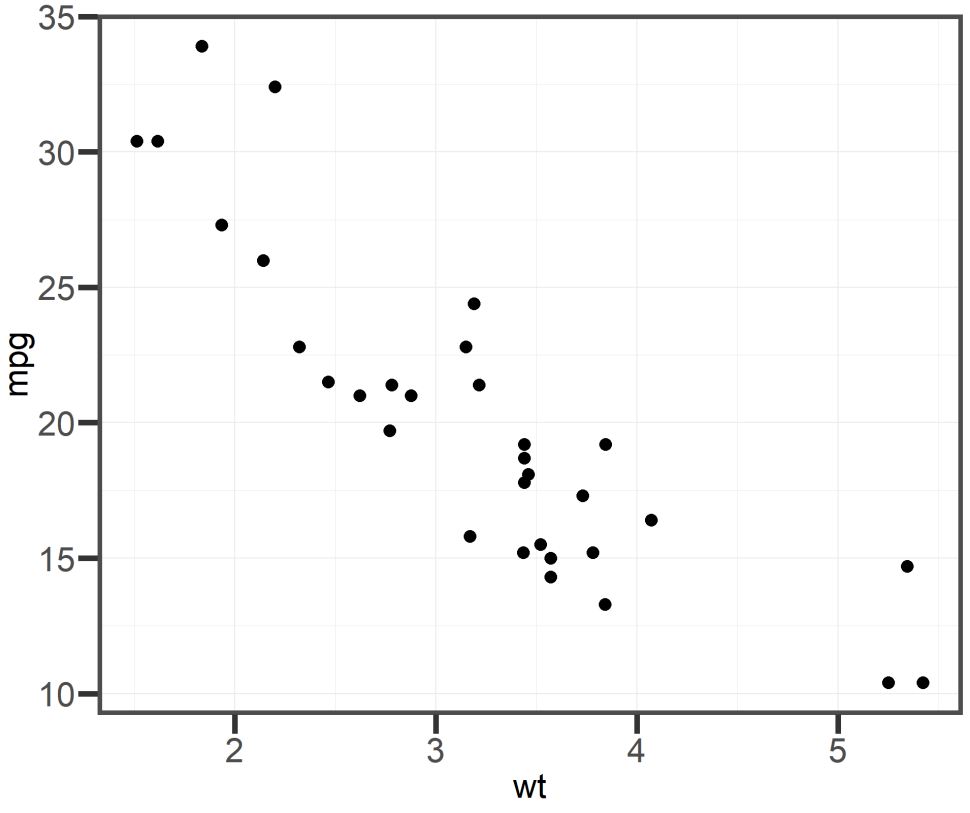
# The foundation

```
1  ggplot(mtcars)
```



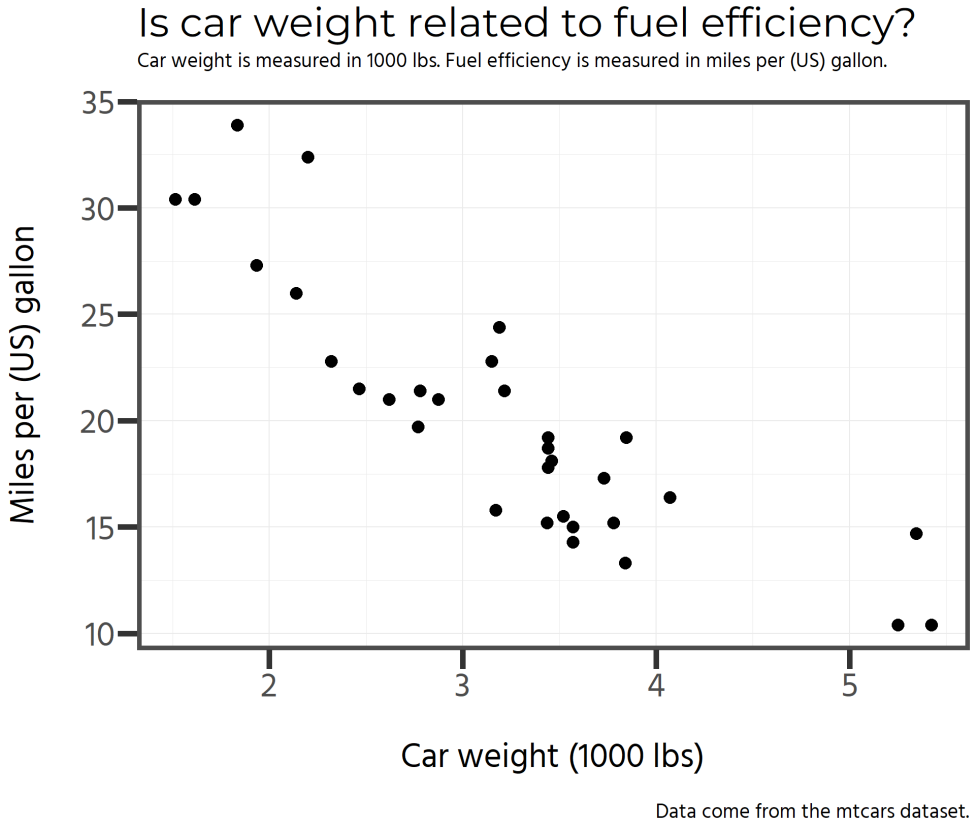
# Plotting the data

```
1  ggplot(mtcars) +
2    geom_point(aes(x=wt, y=mpg))
```



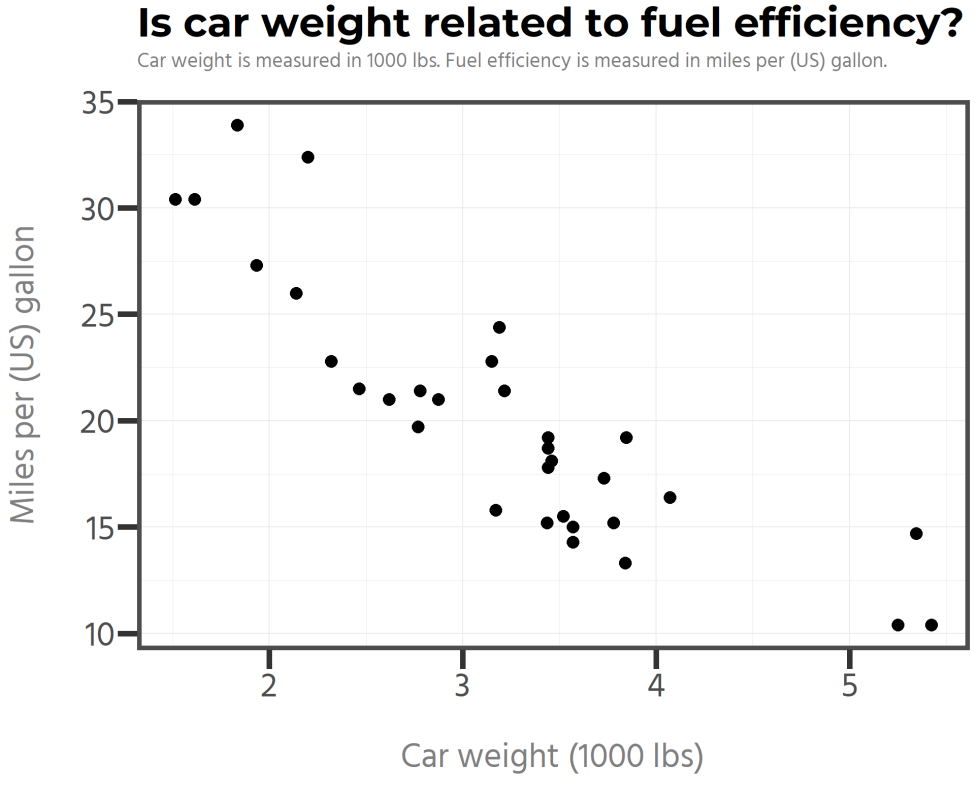
# Annotating your plot

```
ggplot(mtcars) +
  geom_point(aes(x=wt, y=mpg)) +
  labs(x='Car weight (1000 lbs)',
       y='Miles per (US) gallon',
       title = 'Is car weight related to fuel efficiency?
       subtitle = 'Car weight is measured in 1000 lbs. Fu
       caption = 'Data come from the mtcars dataset.')
```



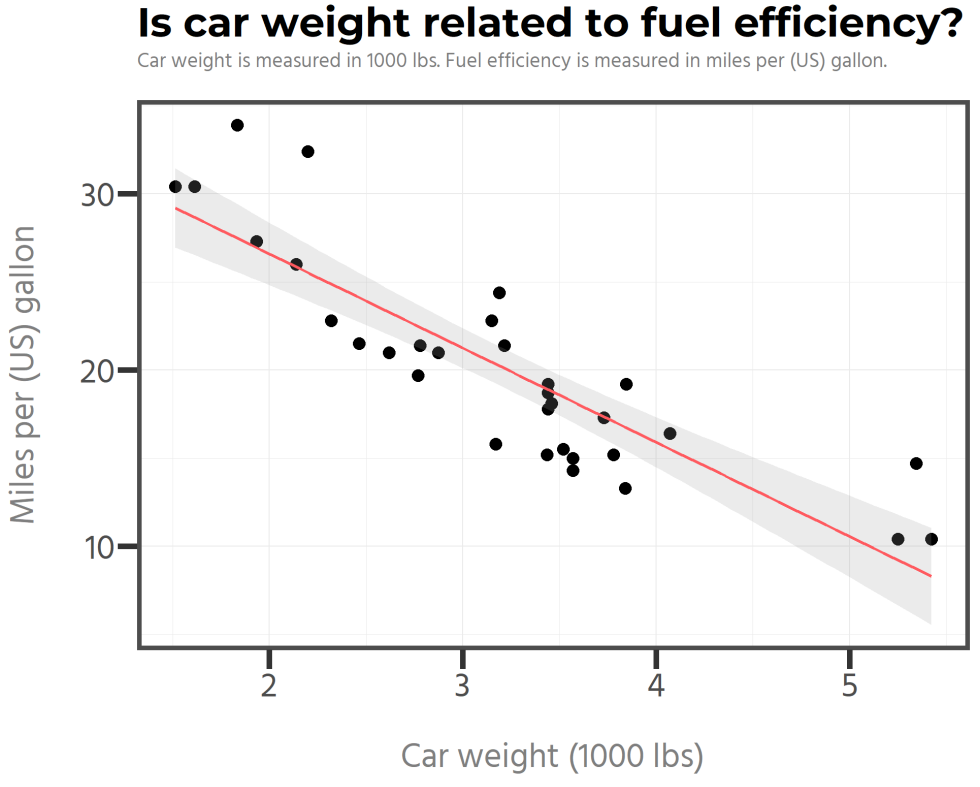
# Changing the theme

```
ggplot(mtcars) +
  geom_point(aes(x=wt, y=mpg)) +
  labs(x='Car weight (1000 lbs)',
       y='Miles per (US) gallon',
       title = 'Is car weight related to fuel efficiency?
       subtitle = 'Car weight is measured in 1000 lbs. Fu
       caption = 'Data come from the mtcars dataset.') +
  theme(axis.title = element_text(color = 'grey50'),
        plot.title = element_text(face = 'bold'),
        plot.subtitle = element_text(color = 'grey50'),
        plot.caption = element_text(face = 'italic'))
```



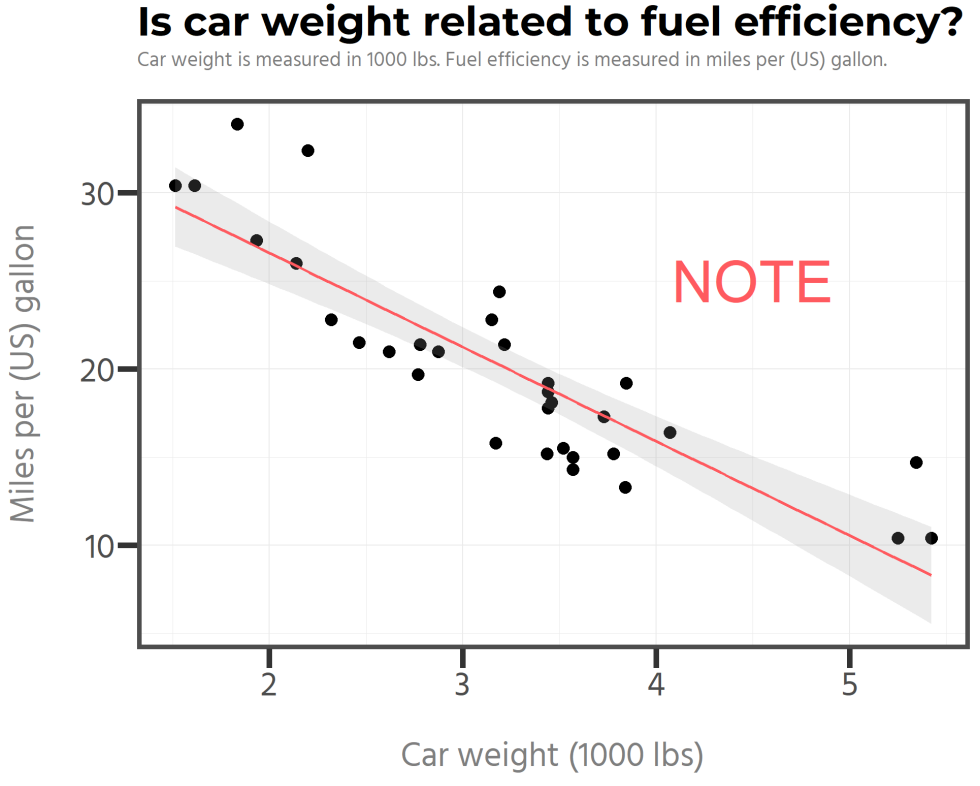
# Adding new layers (fitted line)

```
ggplot(mtcars) +
  geom_point(aes(x=wt, y=mpg)) +
  geom_smooth(aes(x=wt, y=mpg),
              method = 'lm',
              alpha = .20,
              color = '#FF5A5F')+
  labs(x='Car weight (1000 lbs)',
       y='Miles per (US) gallon',
       title = 'Is car weight related to fuel efficiency?
       subtitle = 'Car weight is measured in 1000 lbs. Fu
       caption = 'Data come from the mtcars dataset.') +
  theme(axis.title = element_text(color = 'grey50'),
        plot.title = element_text(face = 'bold'),
        plot.subtitle = element_text(color = 'grey50'),
        plot.caption = element_text(face = 'italic'))
```



# Adding new layers (annotations)

```
ggplot(mtcars) +
  geom_point(aes(x=wt, y=mpg)) +
  geom_smooth(aes(x=wt, y=mpg),
              method = 'lm',
              alpha = .20,
              color = '#FF5A5F') +
  annotate("text", x=4.5, y=25, label = 'NOTE', color = '
  labs(x='Car weight (1000 lbs)',
       y='Miles per (US) gallon',
       title = 'Is car weight related to fuel efficiency?
       subtitle = 'Car weight is measured in 1000 lbs. Fu
       caption = 'Data come from the mtcars dataset.') +
  theme(axis.title = element_text(color = 'grey50'),
        plot.title = element_text(face = 'bold'),
        plot.subtitle = element_text(color = 'grey50'),
        plot.caption = element_text(face = 'italic'))
```

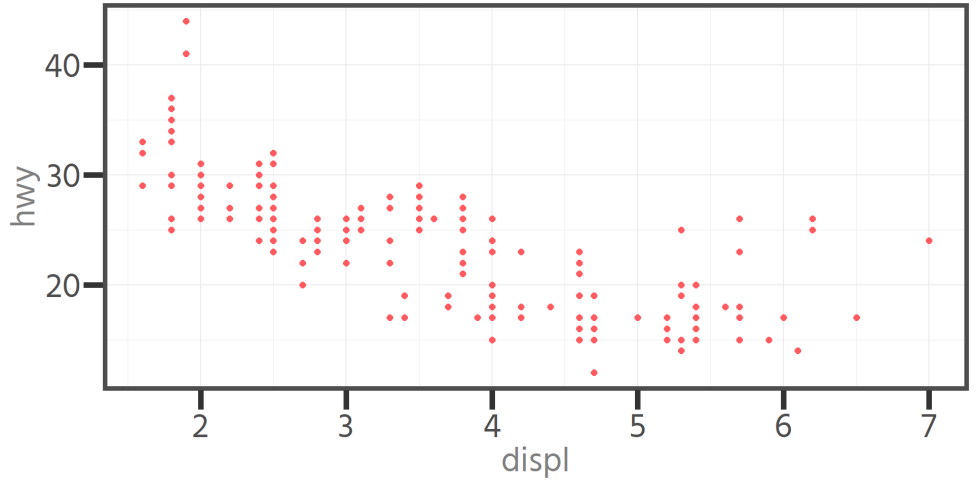


# Aesthetic attributes

*Aesthetics* are parameters that can change the appearance of a plot.

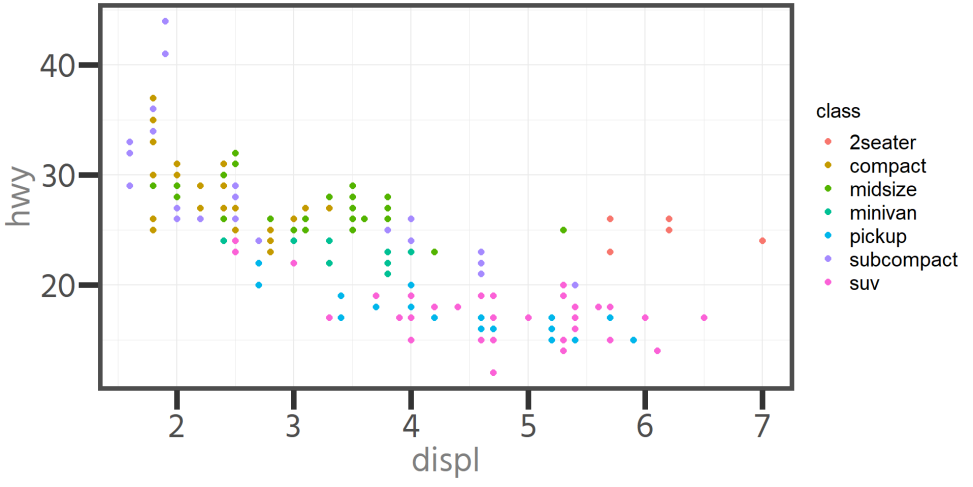
**Global aesthetics** affect all data at the same time.

```
ggplot(mpg) +
  geom_point(aes(displ, hwy),
             color = 'red')
```

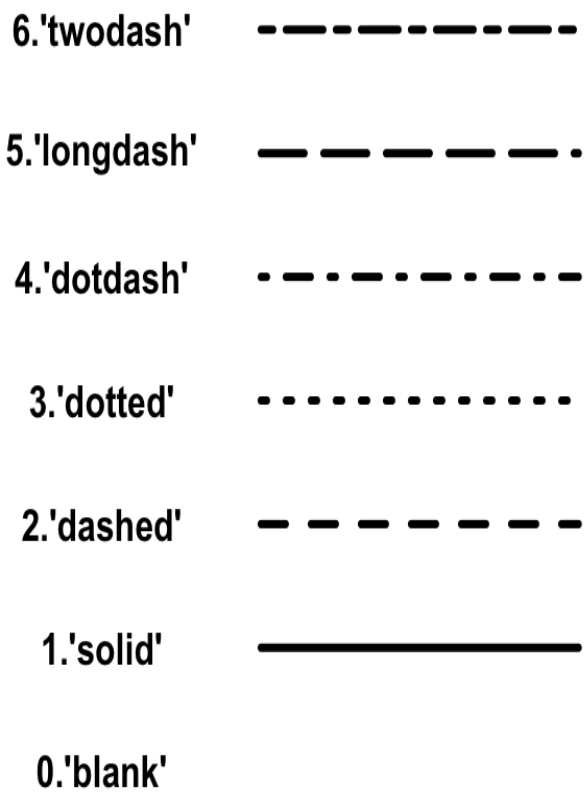


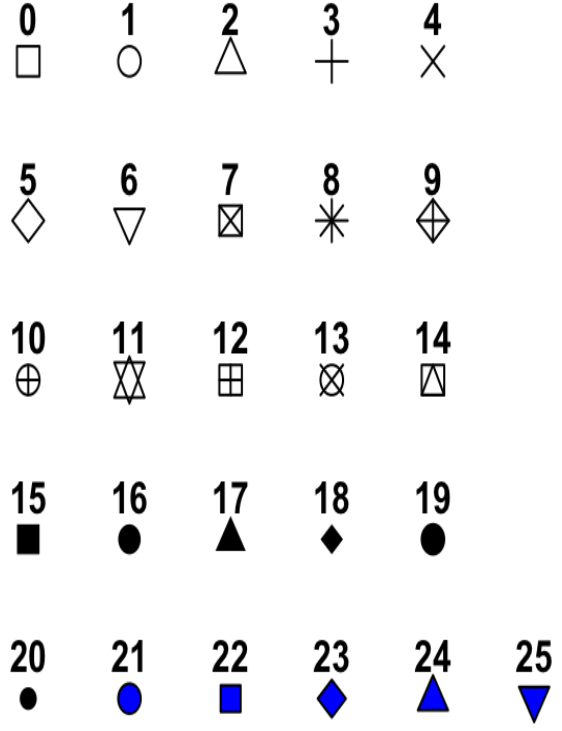

**Mapped aesthetics** affect data based on the values of some specified variable.

```
ggplot(mpg) +
  geom_point(aes(displ, hwy, color = class))
# Notice that 'color = ____' is now inside aes()
```



# Linetype and shape aesthetics



# Scales

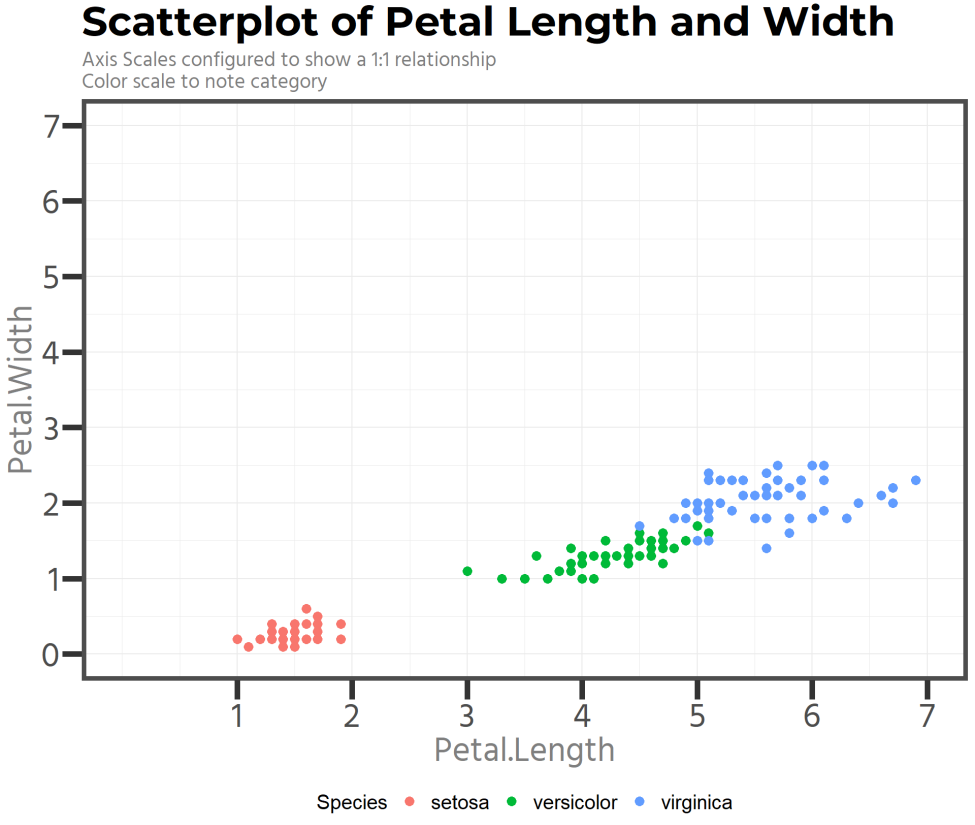
*Scales* refer to how data are translated into visual representations. This includes the x-axis, y-axis, colors, shapes, etc. Scales are essential components of the mapping process.

**Key Elements**

Name: a basic, though often neglected, description of scale

Limits: the range of a scale

Breaks: the subdivisions of range used to mark gradations



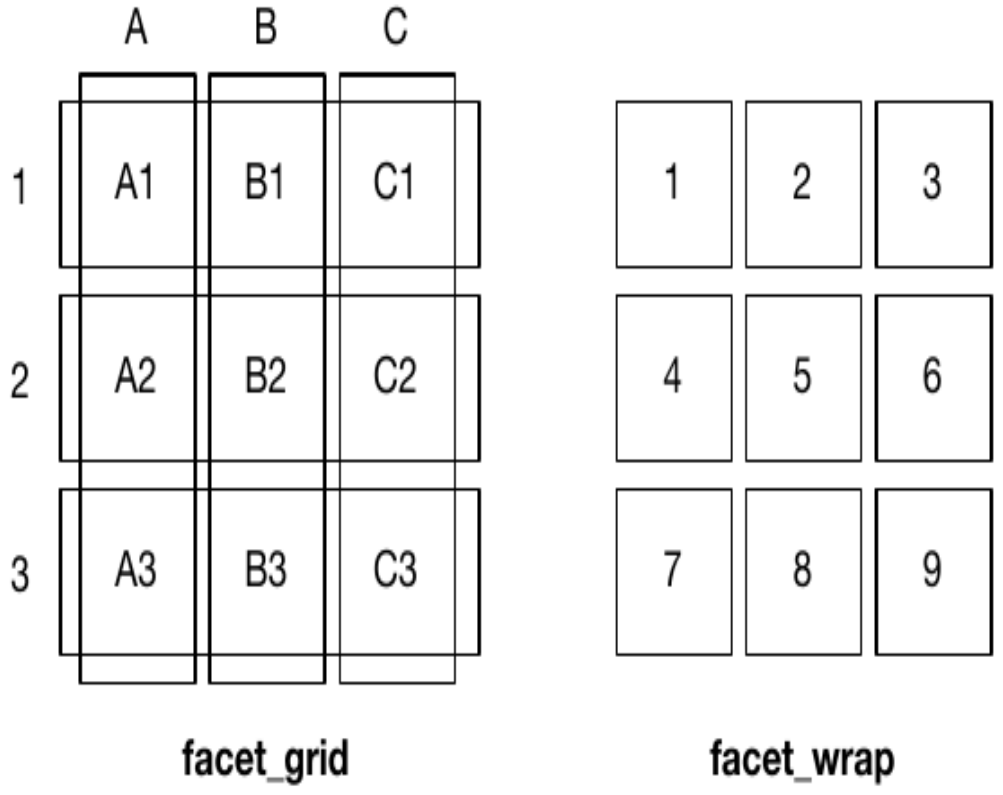
# Facets

*Facets* show subsets of the data in a sequential, side-by-side format.

Two types of facets:

`facet_wrap`:
Subsetting data on one variable

`facet_grid`:
Subsetting data on more than one variable



# Exercise #1