Visualize Data

## Your Turn 0

Add a setup chunk that loads the tidyverse packages.

mpg

## # A tibble: 234 × 11  
## manufacturer model displ year cyl trans drv cty hwy fl class  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>  
## 1 audi a4 1.8 1999 4 auto… f 18 29 p comp…  
## 2 audi a4 1.8 1999 4 manu… f 21 29 p comp…  
## 3 audi a4 2 2008 4 manu… f 20 31 p comp…  
## 4 audi a4 2 2008 4 auto… f 21 30 p comp…  
## 5 audi a4 2.8 1999 6 auto… f 16 26 p comp…  
## 6 audi a4 2.8 1999 6 manu… f 18 26 p comp…  
## 7 audi a4 3.1 2008 6 auto… f 18 27 p comp…  
## 8 audi a4 quattro 1.8 1999 4 manu… 4 18 26 p comp…  
## 9 audi a4 quattro 1.8 1999 4 auto… 4 16 25 p comp…  
## 10 audi a4 quattro 2 2008 4 manu… 4 20 28 p comp…  
## # ℹ 224 more rows

## Your Turn 1

Run the code on the slide to make a graph. Pay strict attention to spelling, capitalization, and parentheses!

ggplot(data = mpg) +  
 geom\_point(mapping = aes(x = displ, y = hwy))



## Your Turn 2

Add color, size, alpha, and shape aesthetics to your graph. Experiment.

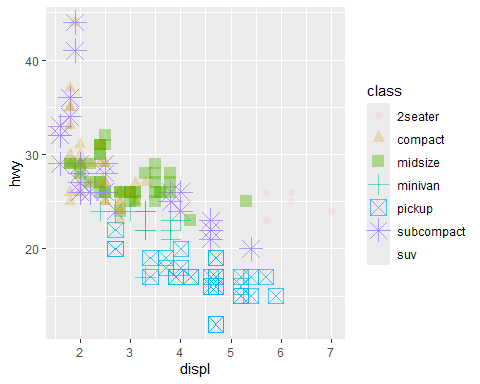
ggplot(data = mpg) +  
 geom\_point(mapping = aes(displ, hwy, color = class, size = class, alpha = class, shape = class))

## Warning: Using size for a discrete variable is not advised.

## Warning: Using alpha for a discrete variable is not advised.

## Warning: The shape palette can deal with a maximum of 6 discrete values because more  
## than 6 becomes difficult to discriminate  
## ℹ you have requested 7 values. Consider specifying shapes manually if you need  
## that many have them.

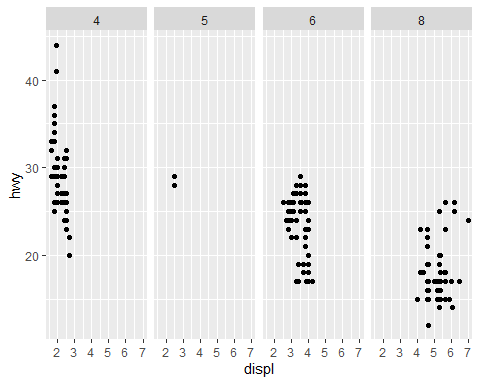
## Warning: Removed 62 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



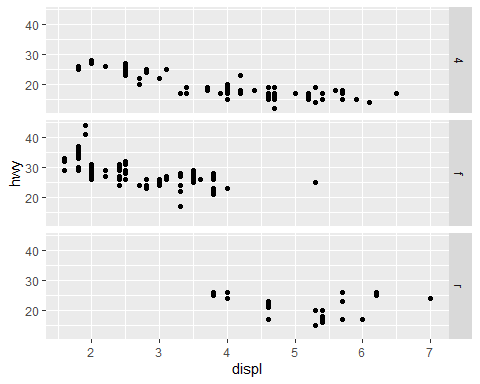
## Help Me

What do facet\_grid() and facet\_wrap() do? (run the code, interpret, convince your group)

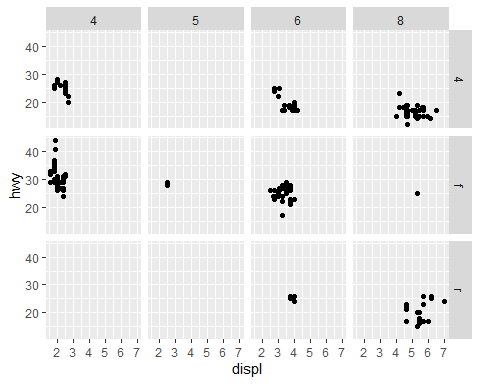
# Makes a plot that the commands below will modify  
q <- ggplot(mpg) + geom\_point(aes(x = displ, y = hwy))  
  
q + facet\_grid(. ~ cyl)



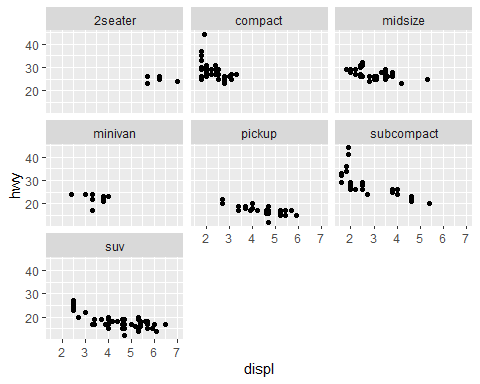
q + facet\_grid(drv ~ .)



q + facet\_grid(drv ~ cyl)

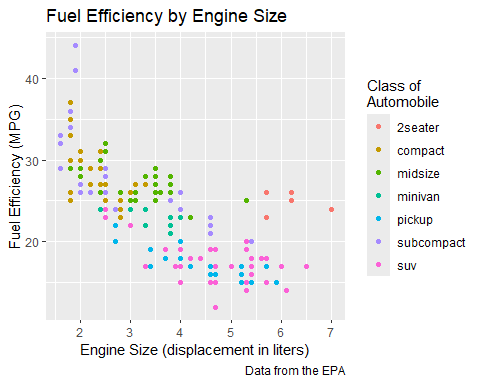


q + facet\_wrap(~ class)

 It separates each type of car into a different scatter plot. ## Your Turn 3

Add the black code to your graph. What does it do?

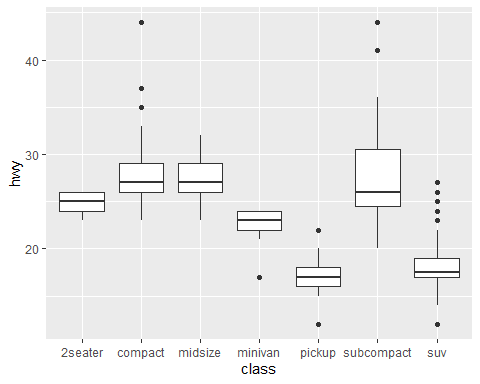
ggplot(data = mpg) +  
 geom\_point(mapping = aes(displ, hwy, color = class)) +  
 labs(title = "Fuel Efficiency by Engine Size",  
 x = "Engine Size (displacement in liters)",  
 y = "Fuel Efficiency (MPG)",  
 color = "Class of \nAutomobile",  
 caption = "Data from the EPA")

 Adds a title, subtitle, color title, x and y text, and a caption

## Your Turn 4

Replace this scatterplot with one that draws boxplots. Use the cheatsheet. Try your best guess.

ggplot(mpg) + geom\_boxplot(mapping = aes(x = class, y = hwy))

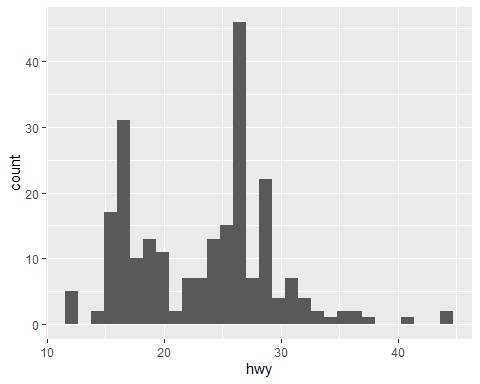


## Your Turn 5

Make a histogram of the hwy variable from mpg. Hint: do not supply a y variable.

ggplot(data = mpg) + geom\_histogram(mapping = aes(x = hwy))

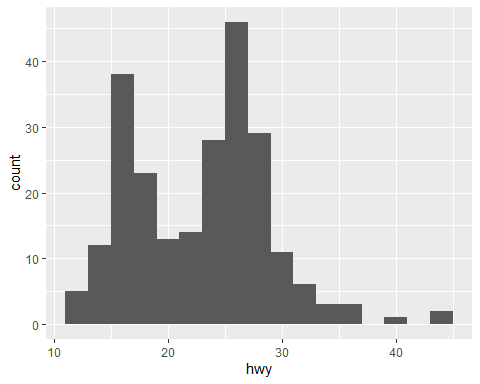
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## Your Turn 6

Use the help page for geom\_histogram to make the bins 2 units wide.

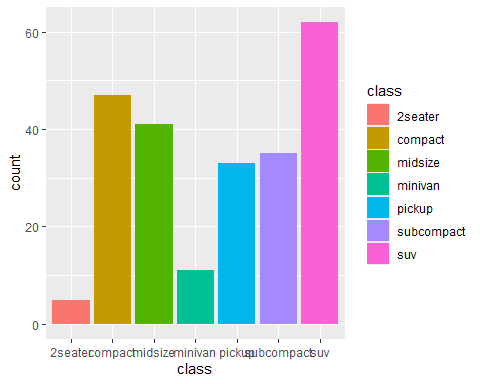
ggplot(data = mpg) +   
 geom\_histogram(mapping = aes(x = hwy), binwidth = 2)



## Your Turn 7

Make a bar chart class colored by class. Use the help page for geom\_bar to choose a “color” aesthetic for class.

ggplot(data = mpg) +  
 geom\_bar(mapping = aes(x = class, fill = class))

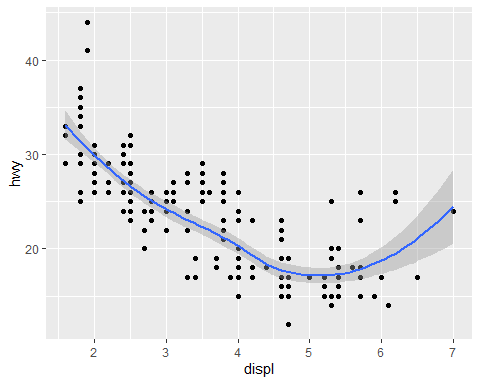


## Quiz

What will this code do?

ggplot(mpg) +   
 geom\_point(aes(displ, hwy)) +  
 geom\_smooth(aes(displ, hwy))

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

 This code will add a scatterplot to the graph and then a smoothing line.

## Quiz

What is different about this plot? Run the code! Note: had to comment out this code in order to knit it.

#p <- ggplot(mpg) +   
# geom\_point(aes(displ, hwy)) +  
# geom\_smooth(aes(displ, hwy))  
  
#library(plotly)  
#ggplotly(p)

This graph is interactive. You are able to see what the x and y values are for the data points or a specific part of the smooth line.

# Take aways

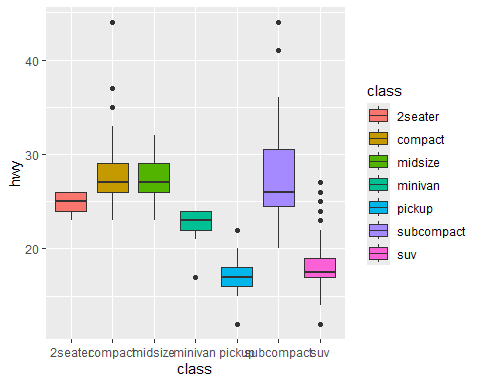
You can use this code template to make thousands of graphs with **ggplot2**.

ggplot(data = <DATA>) +  
 <GEOM\_FUNCTION>(mapping = aes(<MAPPINGS>))

# Other Exercises

1. What happens if you facet on a continuous variable? It’ll create a new panel for each unique value for that variable.
2. What are the advantages to using faceting instead of the color aesthetic? What are the disadvantages? How might the balance change if you had a larger dataset? Some of the advantages include each panel showing a distinct set of the data and easy comparison between all the facets. Some of the disadvantages include a complexity in the vision when there are a large number of facets and not as much details in each one. One will need to consider whether it is better to use faceting (like trying to find a pattern between different variables) or color aesthetic (distinguishing between groups and how they perform compared to each other).
3. What does the se argument to geom\_smooth() do? It tells the argument whether or not to display the confidence interval.
4. What is the default position adjustment for geom\_boxplot()? Create a visualization from the mpg dataset that demonstrates different position adjustments. The default position adjustment for the function is dodge2.

ggplot(data = mpg) +  
 geom\_boxplot(aes(x = class, y = hwy, fill = class))



#vs  
  
ggplot(data = mpg) +  
 geom\_boxplot(aes(x = class, y = hwy, fill = factor(cyl)), position = "identity", alpha =0.5)

