R Homework Week 4

*Data Visualisation*

Topics:

⇒ Creating scatterplots using ggplot2 and the mpg dataframe.

⇒ How to use multiple geoms.

⇒ How to use facets and aesthetic mappings.

⇒ Creating your own scatterplot using our data.

Introduction:

For this homework, you’ll be working with the mpg dataframe that’s included with ggplot2. This dataframe contains real observations collected by the Environmental Protection Agency on 38 different car models. Some variables you’ll be working with are displ (a car’s engine size in liters) and hwy (a car’s fuel efficiency on the highway in miles per gallon). You can learn about all the variables in mpg by typing ?mpg into your console. Come to me (Michael) with any questions.

Here is a reusable template for making graphs with ggplot2:

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

1. Run **ggplot**(data = mpg). What happens? Why can’t you see any data?
2. What does the drv variable describe? Remember, you can use ?mpg to access more information about the dataset.
3. Make a scatterplot of hwy vs. cyl.
4. What happens if you make a scatterplot of class vs. drv? Why isn’t this useful?
5. Which variables in mpg are categorical? Which are continuous?
6. What’s wrong with this code? Why aren’t the points blue?

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

1. Map a continuous variable to color, size, and shape. How do these aesthetics behave differently for categorical vs. continuous variables?
2. What happens if you map the same variable to multiple aesthetics?
3. What does the stroke aesthetic do? What shape does it work with? (Hint: use ?geom\_point).
4. What happens if you facet on a continuous variable?
5. What plots does the following code make? What does . do?

**ggplot**(data = mpg) +   
 **geom\_point**(mapping = **aes**(x = displ, y = hwy)) +  
 **facet\_grid**(drv ~ .)  
  
**ggplot**(data = mpg) +   
 **geom\_point**(mapping = **aes**(x = displ, y = hwy)) +  
 **facet\_grid**(. ~ cyl)

1. Read ?facet\_wrap. What does nrow do? What does ncol do? What other options control the layout of the individual panels? Why doesn’t facet\_grid() use these arguments?