Lab 3: Introduction to R.

Daniel Ocampo

October 15, 2017

1 Objectives

- To enhance the understanding of the basic R functionality, including the use of R Studio and producing data visualizations.
- Reading Assignment: Chapters 3, 4, 6 and 8

2 Questions

The questions to answer are the following. 3.2.4, 3.3.1, 3.5.1, 3.6.1, 3.7.1, 3.8.1, 3.9.1, 4.4, 6.3 Note: For the question 6.3.1, where a tip is adopted from Twitter, please give the tip and then explain how it is useful to you.

3 Chapter 3 from textbook

```
    install.packages("tidyverse")
library(tidyverse)
package::function()
ggplot2::ggplot()
ggplot2::mpg
?mpg
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
#ploting
ggplot(data = mpg)
#making the data equal to mpg

ggplot(data = mpg) +
  geom_point(mapping = aes(x = cyl, y = hwy))
# cylinder to miles in highway
ggplot(data = mpg) +
```

```
    geom_point(mapping = aes(x = class, y = drv))
    #the type to  wheel drive
    ggplot(data = mpg) +
geom_point(mapping = aes(x = displ, y = hwy, color = class))
    #changing the color by type E.g
    ggplot(data = mpg) +
      geom_point(mapping = aes(x = displ, y = hwy, size = class))
    #changing the size of the display
    ggplot(data = mpg) +
      geom_point(mapping = aes(x = displ, y = hwy, alpha = class))
    #changing by shading
    ggplot(data = mpg) +
      geom_point(mapping = aes(x = displ, y = hwy, shape = class))
    #difftent shape to type of car
    ggplot(data = mpg) +
      geom_point(mapping = aes(x = displ, y = hwy), color = "blue")
    #change data to all blue

    ggplot(data = mpg) +
      geom_point(mapping = aes(x = displ, y = hwy, color = "blue"))
    mpg##
    View(mpg)
    ## displays all of the data
    #next three will be about 3.3.1 question 3
    ggplot(data = mpg) +
      geom_point(mapping = aes(x = displ, y = hwy, color = cty))
    ### making some R code with continous values for color
    ggplot(data = mpg)+
      geom_point(mapping = aes(x = displ , y = hwy , shape = class))
    ### making shapes based off class
    ggplot(data = mpg)+
        geom_point(mapping = aes(x = displ , y = hwy , size = trans))
    ### the size of the plots based on the transmission
    #### 3.3.1 Question 4
    ggplot(data = mpg)+
      geom_point(mapping = aes(x = displ , y = hwy , size = trans, color = tra
    ## what if we add mutliple arguments we can get a mixture of both
    ?geom_point
    ## for help Use the stroke aesthetic to modify the width of the
    # border
    ggplot(mtcars, aes(wt, mpg)) +
      geom_point(shape = 21, colour = "black", fill = "white", size = 5, strok
```

```
#exa  mple from ?geom_point

ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ class, nrow = 2)
#### seprating data based on class
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_grid(drv ~ cyl)
### plot a grid
```

```
# A tibble: 234 x 11
   manufacturer      model displ  year   cyl      trans   drv   cty   hwy    fl
          <chr>      <chr> <dbl> <int> <int>      <chr> <chr> <int> <int> <chr>
 1         audi         a4   1.8  1999     4   auto(l5)     f    18    29     p
 2         audi         a4   1.8  1999     4 manual(m5)     f    21    29     p
 3         audi         a4   2.0  2008     4 manual(m6)     f    20    31     p
 4         audi         a4   2.0  2008     4   auto(av)     f    21    30     p
 5         audi         a4   2.8  1999     6   auto(l5)     f    16    26     p
 6         audi         a4   2.8  1999     6 manual(m5)     f    18    26     p
 7         audi         a4   3.1  2008     6   auto(av)     f    18    27     p
 8         audi a4 quattro   1.8  1999     4 manual(m5)     4    18    26     p
 9         audi a4 quattro   1.8  1999     4   auto(l5)     4    16    25     p
10         audi a4 quattro   2.0  2008     4 manual(m6)     4    20    28     p
# ... with 224 more rows, and 1 more variables: class <chr>
```

3.2.4 Exercises

1. Run ggplot(data = mpg). What do you see?
   I did not see anything when the code ran.

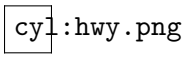
2. How many rows are in mpg? How many columns?
   I would say that there is 0 rows for mpg and two columns for MPG, both hwy and cty. (The view of the data is down below).

3. What does the drv variable describe? Read the help for ?mpg to find out.
   F = Front-wheel drive
   4 = Four wheel drive
   R = Rear wheel drive

4. Make a scatterplot of hwy vs cyl.

```
ggplot(data = mpg) +
    geom_point(mapping = aes(x = cyl, y = hwy))
```

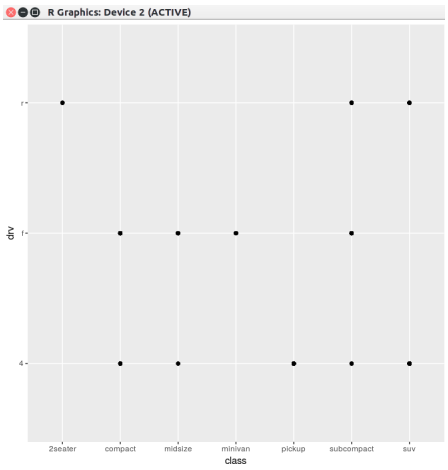
   The output is posted in the figure below.



5. What happens if you make a scatterplot of class vs drv? Why is the plot not useful?

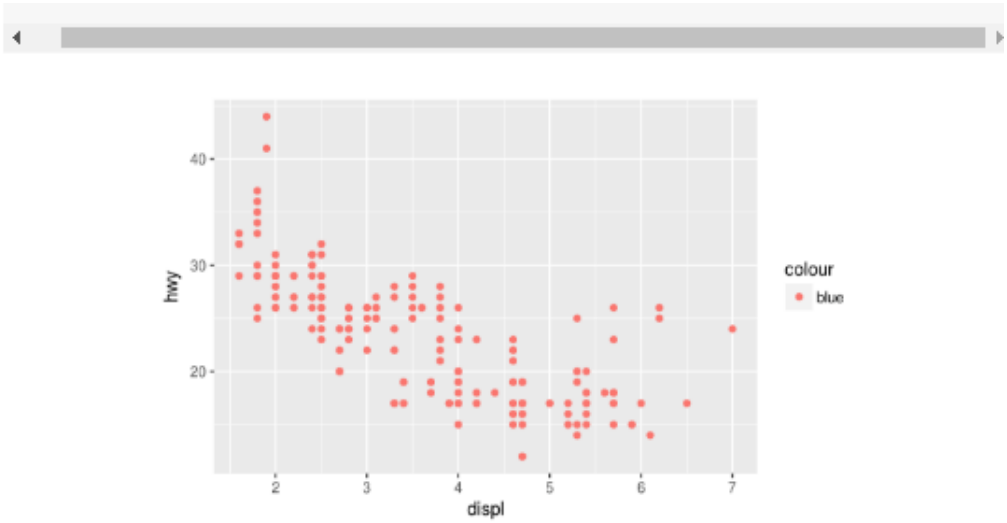
```
ggplot(data = mpg) +
    geom_point(mapping = aes(x = class, y = drv))
```

   It has no useful information, what point of knowing what brand they are and if it is rear or 4 wheel drive, not important information. (The image below shows the output of the code).



3.3.1

1. Whats gone wrong with this code? Why are the points not blue?
   The reason why it is not blue is because it needs to go out the aesthetic



2. Which variables in mpg are categorical? Which variables are continuous? (Hint: type ?mpg to read the documentation for the dataset). How can you see this information when you run mpg?

   I would say that model, manufacturer, trans, drv, class would be considered categorical. Variables that are considered continuous would be city, hwy ,cyl, and displ. when I type in mpg, I get to see what are int and strings, so it can give me a better guess of weather it is categorical or continuous.

3. Map a continuous variable to color, size, and shape. How do these aesthetics behave differently for categorical vs. continuous variables?

```
#next three will be about 3.3.1 question
ggplot(data = mpg) +
    geom_point(mapping = aes(x = displ, y = hwy, color = cty))
### making some R code with continous values for color
 ggplot(data = mpg)+
   geom_point(mapping = aes(x = displ , y = hwy , shape = class))
 ### making shapes based off class
 ggplot(data = mpg)+
     geom_point(mapping = aes(x = displ , y = hwy , size = trans))
```

4. What happens if you map the same variable to multiple aesthetics?

```
ggplot(data = mpg)+
geom_point(mapping = aes(x = displ , y = hwy , size = trans ,
color = trans))
```

The thing that you get is a mixture of both of them, so you would get the size and the color at the same time.

5. What does the stroke aesthetic do? What shapes does it work with? (Hint: use ?geom_point)

According R studio, it Uses the stroke aesthetic is to modify the width of the border. o

```
    ggplot(mtcars, aes(wt, mpg)) +
geom_point(shape = 21, colour = "black", fill = "white", size = 5,
stroke = 5)
```

This was the example that was given, so apparently it supports size.

6. What happens if you map an aesthetic to something other than a variable name, like aes(colour = displ ¡ 5)?

Well I think what is going to happen is that it going to give a true or false, and change the colors based on part of the spectrum the engine size is located.

3.5.1

1. What happens if you facet on a continuous variable?
   well continuous data may be to much for R studio, it would be hard to plot continuous data. In my opnion it would be point less because what if you have a data point by itself in a graph

2. What do the empty cells in plot with facet_grid(drv ~ cyl) mean? How do they relate to this plot?

```
ggplot(data = mpg) +
geom_point(mapping = aes(x = drv, y = cyl))
```

   The following code just has few points showing that there are cars with front wheel drives, have 4 cylinders and so on so fourth, Don't see how this data can be useful.

3. What plots does the following code make? What does . do?
   The . function is similar to ~ in that it is used to capture the name of variables, not their current value. This is used throughout plyr to specify the names of variables (or more complicated expressions).

```
ggplot(data = mpg) +
geom_point(mapping = aes(x = displ, y = hwy)) +
facet_grid(drv ~ .)
```

   This part of code make a longer horizontal grid which, just puts all of the types on grid and plot miless on hwy based on how big the size of the engine while also separating them by the wheel drive.

```
ggplot(data = mpg) +
geom_point(mapping = aes(x = displ, y = hwy)) +
facet_grid(cyl ~ .)
```

   The same thing but just based on the cylinders.

4. Take the first faceted plot in this section:

```
ggplot(data = mpg) +
geom_point(mapping = aes(x = displ, y = hwy)) +
facet_grid(drv ~ .)
```

   What are the advantages to using faceting instead of the colour aesthetic? What are the disadvantages? How might the balance change if you had a larger dataset?

   Well this is a hard question to answer because, it all depends on the dataset

3.6.1

1. What geom would you use to draw a line chart? A boxplot? A histogram? An area chart?
   I can figure this problem out by simply doing the ?ggplot.
   geom_line() = a line chart
   geom_boxplot()= box plot chart
   geom_histogram() = a historgram box
   geom_area()= this equalts a area chart.

2. Run this code in your head and predict what the output will look like. Then, run the code in R and check your predictions.

```
ggplot(data = mpg, mapping = aes(x =
displ, y = hwy, color = drv)) +
 geom_point() +
 geom_smooth(se = FALSE)
```

   I think the graph is going to produce the engine size compared to amount of miles, while changing the color by the wheel drive.

   What happened when I ran this is exactly what I thought would happen the only difference is that I forgot to mention the smooth part, which would draw a line. around the areas where wheel drive had switched.

3. What does show.legend = FALSE do? What happens if you remove it? Why do you think I used it earlier in the chapter?

   It removes the legend, I think he used to help show how to make line of a scatter plot.

4. What does the se argument to geom_smooth() do.

   I think what it does is a condtion so if it true draw the line else if do not draw it.

5. Will these two graphs look different? Why/why not?

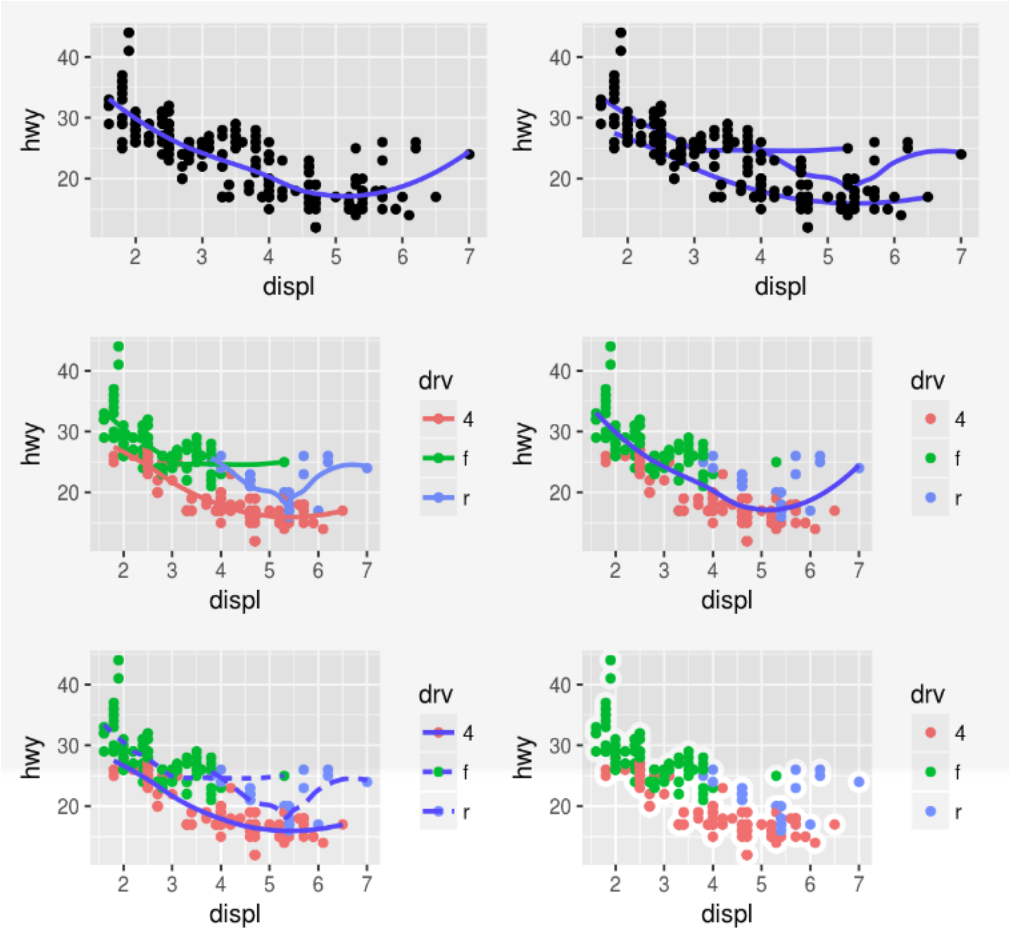
```
ggplot(data = mpg, mapping =
aes(x = displ, y = hwy)) +
 geom_point() +
 geom_smooth()

ggplot() +
  geom_point(data = mpg, mapping =
```

```
aes(x = displ, y = hwy)) +
geom_smooth(data = mpg, mapping =
aes(x = displ, y = hwy))
```

I am not sure because they use the same data, It should produce the same graph, but there is a lot details that might be missed in the format of the code

6. Recreate the R code necessary to generate the following graphs.




```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
geom_point() +
geom_smooth(se = FALSE)
##### number 1
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
geom_smooth(aes(group = drv), se = FALSE) +
```

```
geom_point()
#### number 2
ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +
geom_point() +
geom_smooth(se = FALSE)

### number 3

ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
geom_point(aes(color = drv)) +
geom_smooth(se = FALSE)

### 4
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
geom_point(aes(color = drv)) +
geom_smooth(aes(linetype = drv), se = FALSE)

### number 5
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
geom_point(size = 4, colour = "white") +
geom_point(aes(colour = drv))

### number 6
```

3.7.1

1. What is the default geom associated with stat_summary()? How could you rewrite the previous plot to use that geom function instead of the stat function?