412/612 PRACTICE TEST

PART 1 GGPLOT PRACTICE (Use tidyverse coding for all problems)

1) Using the mpg data frame, create a scatter plot that shows a relationship between the variables **displ** and **cty.** (**displ** =x and **cty**=y)

2 Using the mpg data frame, create a scatter plot that shows a relationship between the variables **displ** and **cty.** (**displ** =x and **cty**=y), and also map colors of your scatter plot to the variable **fl.**

3. Using the mpg data frame, create a scatter plot that shows a relationship between the variables **displ** and **cty.** (**displ** =x and **cty**=y), and also include code so that all of your scatter plot points are green.

4. Using the mpg data frame, create a scatter plot that shows a relationship between the variables **displ** and **cty.** (**displ** =x and **cty**=y), and also include code so that the scatter plot has different shapes or characters according to class. Code so that your scatter plot has the title *City vs Highway* , the y axis is labled *City* and the x axis is labeld *Highway.* (Check out the ggplot graphing example towards the end of chapter one in your book to get data points with different shapes)

5. Using the mpg data frame, create a smooth line fitted to the data **displ** and **cty.** (**displ** =x and **cty**=y).

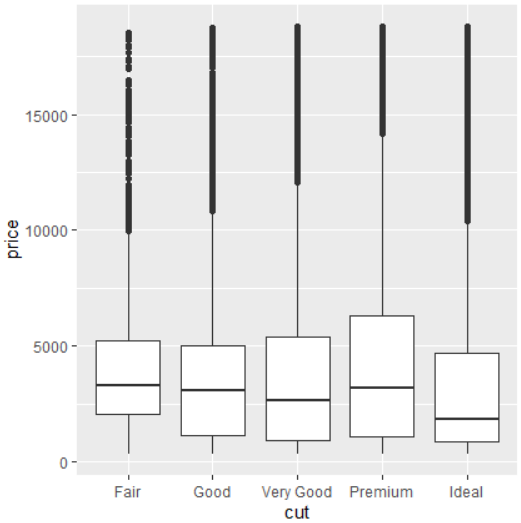
6. Using the mpg data frame, create smooth lines over the scatter plot for the data **displ** and **cty.** (**displ** =x and **cty**=y) based on **drv** levels.

7. Using the mpg data frame, create a scatter plot that shows a relationship between the variables **displ** and **cty.** (**displ** =x and **cty**=y), and also include code that produces overlayed regression lines based on **drv** levels. The data points should also be color coded according to levels of **drv.**

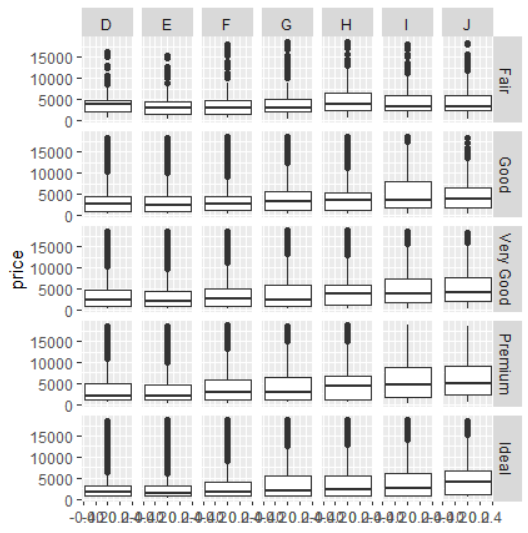
8. Using the mpg data frame, create a bar graph for the variable **drv.** include code so that each bar has a different color. Title your bar graph *DRV Bar Graph*

9. Using the mpg data frame, for the variable **drv,** create a bar graph that shows **color stacked bars** over the variable **trans** for the variable **hwy**

10. Using the stacked bar graph from number 10, create a pie chart.

11. Use and show R code that will produce the following side by side box plots from the **diamonds** data frame.

12. Use and show R code that will produce the following **faceted** display of boxplots from the **diamonds** data frame.



13. Use and show R code that will produce a histogram that is colored blue for the **hwy** variable from the **mpg** data frame. Give the histogram the title *Blue Histogram*.

PART II DPLYR PRACTICE (You may use the piping method or the assignment method)

1. Using the DPLYR **filter** function and the mpg data frame, produce a data frame that only has output for a Dodge Durango 4wd.

2. Using the DPLYR **filter** function and the mpg data frame, produce a data frame that only has output for vehicles whose city mileage is less than 10 miles per gallon and whose highway mileage is less than 16 miles per gallon.

3. Using the DPLYR **arrange** function and the mpg data frame, produce a data frame that displays **displ** in descending order

4. Using the DPLYR **arrange** function and the mpg data frame, produce a data frame of 30 observations that display city miles per gallon in ascending order. Which vehicle has the lowest city miles per gallon ?

5. Using the DPLYR **filter** and **select**  functions and the mpg data frame, produce a data frame that displays all ford vehicles for 1999 whose city miles per gallon is less than 16 and whose highway miles per gallon is also less than 16.

6. Using the DPLYR filter, select and mutate functions and the mpg data frame, produce a data frame that displays the difference between highway mileage and city mileage for the ford mustang.

A partial data frame is given below;

manufacturer model cty hwy difference

*<chr>* *<chr>* *<int>* *<int>* *<int>*

1 ford mustang 18 26 8

2 ford mustang 18 25 7

3 ford mustang 17 26 9

7. Install the New York City flights data package;

**install.packages(“nycflights13”)**

Now code and execute the following library

**library(nycflights.13)**

And then code and call the following data frame

**flights**

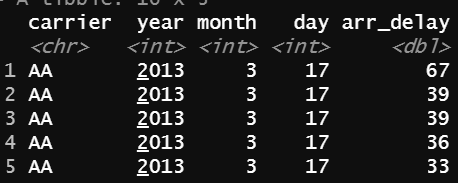
Use the following commands to further explore the data frame **flights**

**?flights**

**View(flights)**

Use DPLYR functions and the piping operator to produce a data frame that shows arrival delay times in descending order for American Airlines on March 17, 2013.

A partial table is given below.

****

8. Using the **mpg** data frame , dplyr functions and the pipe operator, produce a data frame that displays the mean mpg for city driving for manufacturers in the year 1999 only in descending order. Which manufacturer got the best average gas mileage in 1999?