More Practice with Plots

Open RStudio and create a new project under your Module 3 folder and call it **Mod3Assignment3.** For this assignment, you will be creating an R Markdown document that will include the creation of plots and basics around exploratory data analysis (EDA). Once completed, all you need to do is submit the word document that is created.

Create the R Markdown Document

1.) In RStudio, select *File -> New File -> Text File*. This will create a blank text file in the same area that scripts were created in previous assignments (upper left panel). Save this file to your project as **Mod3Assign3Answer.rmd** (it is important to save with the .rmd extension as this saves the text file as an R Markdown file).

2.) Create a Header 1 with the title: **Module 3 - Assignment 3**

3.) Create a Header 2 with the title: **Last Name, First Name** (replace with your name)

4.) Create a Header 3 with the title: **More Practice with Plots**

5.) Click on the dropdown arrow next to the Knit icon at the top of the R Markdown Pane in RStudio and select Knit to Word.

6.) Notice that you now have a document in your files for the project named AssignmentAnswer.docx. This is the file you will be uploading later to Canvas.

7.) Add a chunk of code that will load the tidyverse package. **You will be using ggplot to create all the plots in this assignment.**

8.) In the same chunk of code, you will need to import the file *GOOG.csv* which can be found in Canvas on Assignment 3. Check to make sure the date fields are imported correctly (See the video on Assignment 1 for more details).

Part 1 – Basic Plots

For part 1, you will be creating a variety of basic plots that have been discussed in class. For this assignment, R Graphics Cookbook (https://r-graphics.org/) provides a good reference when creating plots.

1.) Create a Header 4 with the title: **Google Stock Price Plots**

2.) Write the following text to explain what you will be doing in this section:

The following is an analysis of Google's stock price from October of 2019 to October of 2020. This will include a scatter, line, bar, histogram and boxplot. All the plots represent the closing price on the dates listed on the x-axis.
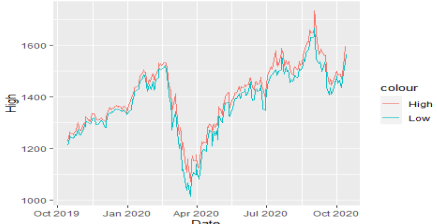
CSBMSBA

3.) For these plots, use the GOOG dataset you just imported in the previous section. For all of the following plots, you will be using the variable *Date* for the x-axis and the variable *Close* for the y-axis.

4.) Create a new chunk of code and name the chunk Basic Plots. Include code to create the following plots in this section:

    1.) Scatterplot
    2.) Line Graph
    3.) Bar Graph (hint: geom_col)
    4.) Histogram (use the variable *Close* for the x-axis, nothing will be needed for the y-axis)
    5.) Box Plot (use the variable *Close* for the x-axis, nothing will be needed for the y-axis)

5.) Finally, before leaving this chunk of code, include a line graph that also displays the points along the line (see https://r-graphics.org/recipe-quick-line for explanation of how to code this).

6.) Immediately following the chunk of code, include the following question and answer it within the markdown document:

You have created many different plots of the same data but some are more helpful than others. Based on the plots you created, which one do you find most useful and why?

Part 2 – Advanced Plots

1.) For the next section, we will be comparing the open/close prices as well as the high/low prices within the GOOG dataset.

2.) Create a Header 4 with the title: **Google Stock Daily Price Comparisons**

3.) In this section, code to show an example of plotting different variables is provided and then, you are to create a similar plot using the different variables stated.

4.) Create a new chunk of code and name it Advanced Plots.

5.) The first plot to include will be a line plot that will include two lines. For this code, you will need to include two geom_line statements. Notice, that to be able to plot two lines we need to include the aes argument within the geom_line statement since the y-axis is different for each line (one is High and the other is Low). Add the following code to your chunk of code.

| Code	Output
```ggplot(GOOG) +` ` geom_line(aes(x=Date,y=High,color="High")) +` ` geom_line(aes(x=Date,y=Low,color="Low"))```	

6.) Notice, that we had to include color="High" and color="Low" for the lines to have a specific color on the plot. These are the default colors which can be changed but for now, let's work with the default colors.

7.) Now, using similar code to the code above, create another line chart that will plot both the *Open* and *Close* variables instead of the High/Low.

8.) The next plot will be using a goem_area command. This will create a plot that will have a line displayed and everything below the line will be filled with a specific color. The code include both a color (which is the color of the line) and a fill (which is the color that will appear below the line). There is also another argument included named alpha (which will change the transparency of the fill color). Add the following code to your chunk of code.

Code	Output
`ggplot(GOOG, aes(x=Date,y=Close)) +` `  geom_area(color="black",fill="blue", alpha=.2)`	
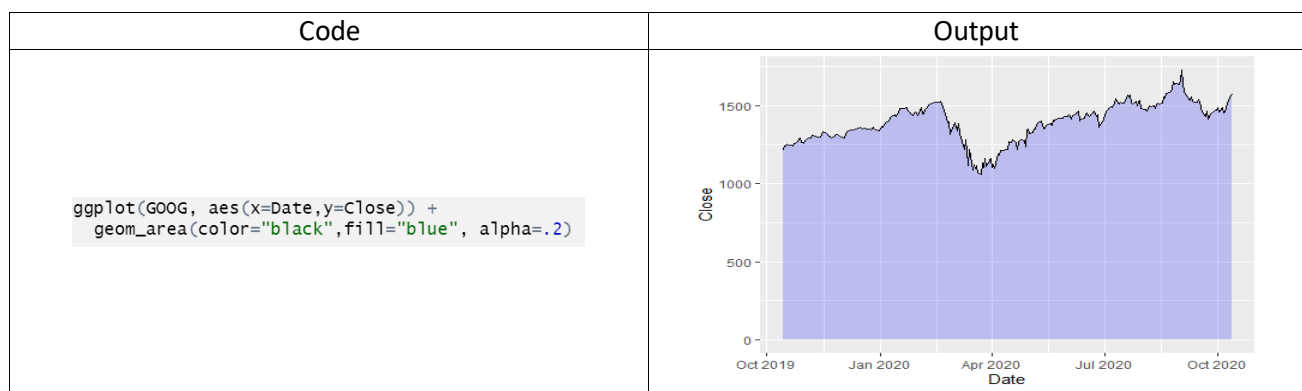
9.) Now, using similar code to the code above, create another line chart that will plot the *Date* and *Open* variables. The line color should be blue and the color below the line should be green. Set the transparency of the color below the line to 0.10

10.) For the last plot, we will be comparing the prices for a specific month by creating a subset of the data. The code to create a subset using the following logic:
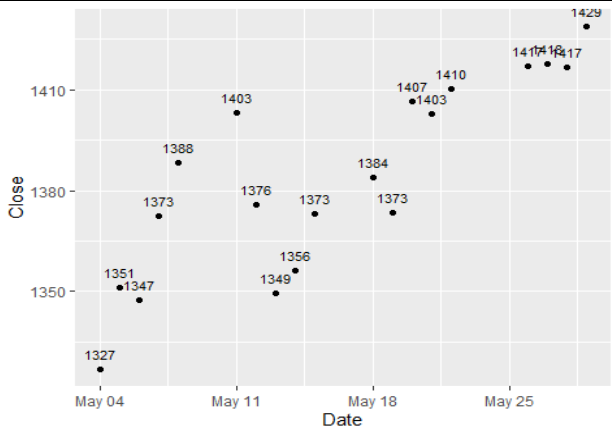
NewDataFile <- subset(*ExistingData, Argument*)

Following the logic above, the code below was used to create a new subset of data called *MayPrice*. Notice that once you run the code, the dataset called *MayPrice* will now be in the Global Environment panel in the upper right corner of RStudio. Add the following code to your chunk of code:

Code	Output
`MayPrice <- subset(GOOG, Date >"2020-04-30" & Date < "2020-06-01" )`	
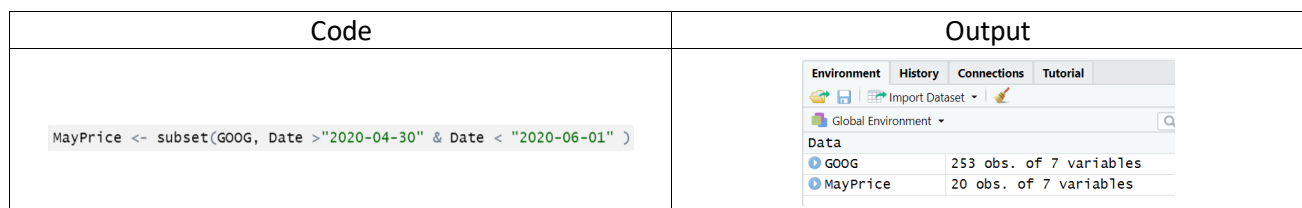
11.) A few things to point out in this code and output. We are using the *Date* variable and selecting the dates that fall between May 1st to May 31st (which is why we use the last day of April and the first day of June). Also, notice that the number of observations is now 20 instead of 253.

12.) Next, we will write the code to plot the points. In this case, we want to also include the text next to the points in our scatterplot. To do this, we include an additional command geom_text. The arguments include a label (what we want to show up next to the point), a size (the font size for the text) and vjust (adjusting the vertical alignment of the text). More information on changing the appearance of text can be found at https://r-graphics.org/recipe-appearance-text-appearance#RECIPE-APPEARANCE-TEXT-APPEARANCE. We have also used the round command for Close. Many times numbers will appear with multiple digits after the decimal point so this command will round all the number up to a whole number. Copy the following code to your chunk of code (I would also encourage you to make adjustments to the arguments to see what happens):

Code	Output
``ggplot(MayPrice, aes(x=Date,y=Close)) +
geom_point() +
geom_text(aes(label = round(Close)), size = 3, vjust=-1)```	

13.) Now, using similar code to the code above, create a new subset called AprilPrice to include only data from the month of April. Next, create a scatterplot of the *Date* and *Open* price which will label the points with the *Open* price, font size of 2 and a vertical adjustment of 1.5

14.) Knit the markdown file to a Word document. Before submitting the Word document, open the document and make sure the plots look good (you may need to adjust the size of the plots manually within the document).

CSBMSBA