R and RStudio Refresher Assignment

In this assignment you will complete a variety of basic R/RStudio tasks. This assignment is designed as a refresher on R basics and is drawn primarily from the R for Data Science textbook (link to the textbook in the course syllabus).

Before you begin: Ensure that you have installed R (version 3.4 or newer) and RStudio (version 1.0.153 or newer). Be sure to install R before installing R Studio.

Good habits: I strongly recommend creating a new RStudio project for every assignment and for each lecture as you follow-along. Using a good directory structure will make it much easier for you to find your work later. For this assignment, you might create a directory for all Module 1 work. Within this directory, create a folder called "Refresher Assignment" (or similar). Create an R Project (with an appropriate name) in this folder.

Deliverable: All of your work for this assignment should be done in an R Markdown document. Knit your document into a Word file and submit the Word file as the deliverable for this assignment.

Task 1: Install and load the "tidyverse" package using the "install.packages" and "library" commands. Note: It is strongly advised that you do not actually run the "install.packages" command when knitting your Markdown document. I strongly suggest running this function (by running the R code code chunk) to allow the "tidyverse" package to be installed. After doing this, comment out the "install.packages" command before knitting your Markdown document. Recall that placing a # before a line of R code comments it out and prevents it from being executed.

Task 2: The "ggplot2" package (part of the tidyverse set of packages) includes a dataset containing data on diamonds. Use the line of code below to read in this dataset into a data frame called "diamonddata".

```
diamonddata = diamonds
```

Use appropriate R commands to determine how many rows and columns are in this dataset.

Task 3: Using ggplot, create a scatterplot of caret (x axis) versus price (y axis). Briefly describe the relationship between these two variables.

Task 4: Repeat Task 3, but in this plot color the scatterplot points by the "cut" variable. Briefly describe the relationship between these three variables (carat, price, and cut).

Task 5: Repeat Task 4, but in this plot facet by "color". Briefly describe the relationship between the four variables (carat, price, cut, and color).

Task 6: Use the "readr" package (part of the tidyverse) to read-in the "InventoryData.csv" file as a data frame called "inventory". Examine the structure and summary of the data frame.

Task 7: Use a filter to create a new data frame called "inventoryA" containing only inventory from Supplier A. How many rows are in this new data frame?

Task 8: What does the line of code shown below do? Note the use of the backtick character (on the tilde key on your computer's keyboard) to delineate the variable names with spaces in them.

```
inventoryA = mutate(inventoryA, OnHandRatio = `On Hand` / `Annual Demand`)
```

Task 9: Create a new data frame called "avg_cost" that contains the average "Cost per Unit ($)" by each "Item SKU" (let this quantity be in a variable called "SKUAvgCost. Hint: Recall the summarize and group_by functions and the use of the backtick character from Task 8. Your data frame should have only two columns:"Item SKU" and "SKUAvgCost".

Task 10: Given your previous course experience with R/RStudio, what topics/concepts did you find to be most challenging?