R and RStudio Refresher Assignment

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

Before you begin: Ensure that you have installed R and RStudio or updated to the most recent versions of these software tools. The easiest way to update R is to simply install the latest version. The latest version of R is available here: https://cran.r-project.org/bin/windows/base/. The latest version of RStudio is available here: https://www.rstudio.com/products/rstudio/download/#download.

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 example, 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: Each question in the assignment corresponds to a question in the Canvas quiz associated with this assignment.

Before you begin: Install and load the "tidyverse" package using the "install.packages" and "library" commands. I strongly suggest running this function once (by running the R 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. **NOTE: Recall that the "tidyverse" package is a collection of packages. When you install and then load the "tidyverse" package you then DO NOT need to install and load packages that are part of the tidyverse. For example, please do not install and load the "readr" package separately as it is installed as part of the "tidyverse" package.**

If you wish to use the "esquisse" package to assist you in creating plots, you will also need to install and library that package.

Question 1: 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 an appropriate R command to answer the question: How many rows are in this dataset?

Question 2: How many columns are in the "diamonds" dataset?

Question 3: Using ggplot, create a scatterplot of caret (x axis) versus price (y axis). Which statement best describes the relationship between "caret" and "price"?

A. As caret decreases price increases
B. As caret increases price increases
C. As caret increases price does not change

Question 4: Repeat Question 3, but in this plot color the scatterplot points by the "cut" variable. Which statement best describes the relationship between "cut" and "price"?

A. As cut quality increases price increases
B. As cut quality decreases price increases
C. As cut quality increases price does not change

Question 5: Repeat Question 4, but in this plot you should facet by "color". Faceting a plot is described at this link: http://www.cookbook-r.com/Graphs/Facets_(ggplot2)/.

What does "faceting" do to the plot?

A. Creates a separate scatterplot for each "cut"
B. Creates a separate scatterplot for each "carat"

C. Creates a separate scatterplot for each "color"
D. Creates a separate scatterplot for each "price"

Question 6: Use the "read_csv" function to read-in the "InventoryData.csv" file as a data frame called "inventory". NOTE: In this course we will ALWAYS (unless instructed otherwise) use the "read_csv" command. Do NOT use the "read.csv" command.

In the "inventory" data frame that you have just created, what type of variable is "Supplier"?

A. numeric
B. integer
C. factor
D. character

Question 7 In the "inventory" data frame that you have just created, what is the mean of the "On Hand" variable?

Question 8: 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?

Question 9: 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`)
```

A. Creates a new variable called "inventoryA"
B. Creates a new variable called "OnHandRatio"
C. Creates a new data frame called "OnHandRatio"

Question 10: Using the "inventoryA" object that you created in Question 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".

What is the "SKUAvgCost" of items with an "Item SKU" of 011?
