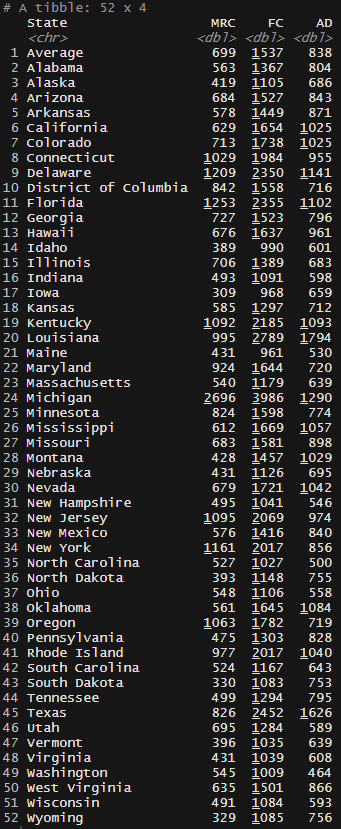
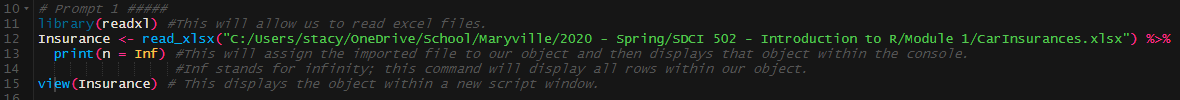
Matthew Stacy

**Project 1**

1. **Read the dataset in** [**CarInsurances.xlsx**](https://maryville.instructure.com/courses/42430/files/6930657/download?wrap=1)**[Preview the document](https://maryville.instructure.com/courses/42430/files/6930657/download?wrap=1) into R. Call the loaded data Insurance. Make sure that you have the directory set to the correct location for the data.**

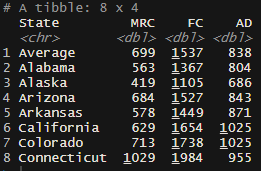
The top center picture shows the script that will import the excel file and assign it to the object “Insurance”. The left picture shows the tibble through the print() command. The reason why we put n = Inf in the print() command is to show ever row of the tibble. Since there are not a lot of rows within the dataset, I did not need to specify the number the rows to display and felt confident that infinity would display the correct number of rows. The right picture shows the imported excel file in a new R Script window through the view() command.

1. **How many rows in the data set? What do the rows represent?**

The top picture shows the script that will display the number of rows in the console and the bottom pictures shows the output that will appear in the console after the command is run. There are 52 rows found within the dataset. Each row represents a state and displays 4 separate results, one for each column.

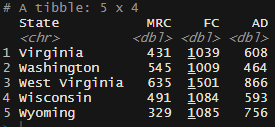
1. **How many columns in the data set? What do the columns represent?**

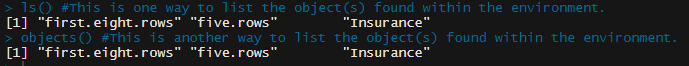
The top picture shows the script that will display the number of columns in the console and the bottom picture shows the output that will appear in the console after the command is run. There are 4 columns found within the data set. The first column represents the different states in alphabetical order and the other three columns represent the rate for either Minimum Required Coverage, Full Coverage, or the Annual Difference between the two coverages.

1. **Assign the first eight rows of the data set to a variable: first.eight.rows and print it out using print() function.**

The top picture shows the script that will assign the first 8 rows of the dataset to a new variable called “first.eight.rows”. We then pipe that command to the print command which displays the result of the command in the console. This is pictured to the left in the tibble. The convenient thing about printing results in the console is that it prints the number of rows just to the left of our tibble telling the user exactly how many different variables there are.

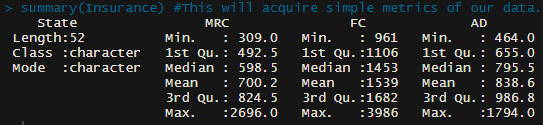
1. **Assign the last five rows of the data set to a variable: five.rows and print it out using print() function.**

The top picture shows the script that will assign the last 5 rows of the dataset to a new variable called “five.rows”. We then pipe that command to the print command which displays the result of the command in the console. This is pictured to the left in the tibble. As mentioned in Question 4, tibbles are extremely convenient for displaying data and ensuring you are performing your commands correctly.

1. **List all objects in the memory using two methods.**

The top picture shows the script that will display (or list) all the objects that are found within the environment. There are two different ways to perform this actions: either ls() which stands for list objects or objects(). They both result in the same output, which can be seen in the bottom picture. There are three objects listed in alphabetical order.

1. **We want to summarize the data. To do it, we may use the summary function. Before asking others for help, it’s generally a good idea for you to try to help yourself either using help() function or google it. Please help yourself and summarize the data first. Then answer the following questions:**

****

The top picture shows the script that will display the summary of each of our columns that contains values and will also display how many variables we have to the left. It states we have 52 variables (or states); we have 50 states, Washington D.C. and an Average.

* 1. **What is the mean of MRC (annual premium of Minimum Required Coverage)?**

As indicated by the bottom picture, the mean for the MRC is 700.2.

* 1. **What is the mean of FC (annual premium of Full Coverage)?**

As indicated by the bottom picture, the mean for the FC is 1539.

* 1. **What is the mean of AD (annual premium differences between MRC and FC)?**

As indicated by the bottom picture, the mean for the AD is 838.6.

We could further improve our data analysis by eliminated any possible outliers and reconducting the summary command. One way we could identify outliers is to perform a boxplot() command. This way we could also determine how large our confidence interval is by providing the command notch = TRUE.