MATH 208 Assignment 1

The assignment contains three questions, with 80 possible points. Your answers must be submitted in the form of a PDF and include both the answers to the questions, along with your R code and output used to generate your answers.

Question 1 (30 points)

You can access datasets from the R datasets package by using

```
data(NAME_OF_DATASET)
```

For this question, we will use the ToothGrowth data.

```
data(ToothGrowth)
```

(a) Determine the (i) mode and (ii) class of the ToothGrowth data object.

(b) Determine how many rows and columns the object has by using R functions.

(c) Using boxplots, histograms, and density plots to describe the distribution of odontoblast lengths by supplement type. Does one supplement seem to be associated with greater lengths? Explain your answer.

(d) Based on your output from part (c), which plot do you think is most effective for assessing whether there is a difference in distribution of lengths between the two groups? Explain your answer.

(e) Create an appropriate scatterplot to assess the association betweeen the dose of the supplement and the lengths and to determine whether the nature of the association depends on the type of supplement. Does the association between length and dose seem to depend on the type of supplement? Explain your answer.

(f) Generate a summary table that contains the mean, median, and standard deviation of the lengths for each supplement type.

Question 2 (30 points)

One of the most popular datasets on the UCI Machine Learning repository is the Abalone dataset, which contains characteristics of sea abalone. The goal of this analysis is to predict the number of rings of the abalone shell, which indicates the age of the abalone. The dataset contains the following data:

Name	Data Type	Measurement Unit
Sex	nominal	
Length	continuous	mm
Diameter	continuous	mm
Height	continuous	mm
Whole weight	continuous	grams
Shucked weight	continuous	grams
Viscera weight	continuous	grams
Shell weight	continuous	grams
Rings	integer	

(a) Read in the data directly to a tibble object from the URL (https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data) by using the read_csv() function (note: the column names are NOT included in the dataset).

(b) Assign names to the columns of the tibble. The columns are in order of the measurements given in the table above.

(c) Create a new column for the radius of the abalone shell by using the diameter.

(d) Find the maximum and minimum number of rings for each value of the Sex variable by using R functions.

(e) Using only plots, explain whether you think the association between total weight and the number of rings depends on the value for Sex.

Question 3 (20 points)

Assume that Prof. Steele creates the following list in R to help manage his life:

```r
shopping_list <- list(
  Grocery = list(
    Dairy = c("Milk","Cheese"),
    Meat = c("Chicken","Sausage","Bacon"),
    Spices = c("Cinnamon")
  ),
  Pharmacy = c("Soap","Toothpaste","Toilet Paper")
  )
```

(a) What objects (or values) are returned by the following lines of R code?

```r
shopping_list$Pharmacy
shopping_list[1][[2]]
shopping_list[[1]][[3]]
shopping_list$Grocery[2][1]
```

(b) Using R code, show which statement yield the following three results:

Result 1:

```
[1] "Soap"          "Toothpaste"    "Toilet Paper"
```

Result 2:

```
$Pharmacy
[1] "Soap"          "Toothpaste"    "Toilet Paper"
```

Result 3:

```
[1] "Sausage"
```