Homework 2: Basic Practice with R

FIRSTNAME LASTNAME

August 30, 2023

Instructions:

The purpose of this homework assignment is to practice using R in RStudio. Make sure you read each question carefully. In each question, I will give you a task to do, and I will tell you want I want you to output. You can write as much code as you want in each code chunk, but make sure you complete the task and only print the output I asked you to print. Don't sort the data unless you are told to sort the data. You should remove the "#" sign in each code chunk before writing your code. **You should knit your RMD file to a PDF after you answer every question.**

For this assignment, we will be using data from the Box Turtle Connection. The Box Turtle Connection is a long-term study anticipating at least 100 years of data collection on box turtles. Their purpose is to learn more about the status and trends in box turtle populations, identify threats, and develop strategies for long-term conservation of the species. Eastern Box Turtle populations are in decline in North Carolina and while they are recognized as a threatened species by the International Union for Conservation of Nature, the turtles have no protection in North Carolina. There are currently more than 30 active research study sites across the state of North Carolina. Turtles are weighed, measured, photographed, and permanently marked. These data, along with voucher photos (photos that document sightings), are then entered into centralized database managed by the NC Wildlife Resources Commission. The *Turtles.csv* dataset contains data collected at The Piedmont Wildlife Center in Durham.

Before answering question 1, remember to run the *setup* code chunk to load all libraries. All of the libraries you load will need to be first installed on your computer.

After you are done, knit the RMarkdown file to PDF and submit the PDF to Gradescope under HW2.

Questions

Q1 (2 Points)

Write code that imports *Turtles.csv* into R and save that data into a data frame named **Turtles**. Print the first 4 rows of **Turtles**.

```
#
```

Q2 (2 Points)

Create a data frame named **Adult_Turtles** that contains only the adult turtles. Print the first 4 rows of **Adult_Turtles**.

```
#
```

Q3 (3 Points)

For the adult turtles sample, I want you to find the three quartiles of the variable **Mass**. I don't care how much code you right, but I only should see three numbers in your output.

```
#
```

Q4 (2 Points)

For the adult turtles, I want you to calculate the ratio of the sample average **Mass** for males to the sample average **Mass** for females. I don't care how much code you right, but I only should see one number in your output.

```
#
```

Q5 (2 Points)

The **Annuli** rings on a turtle represent growth on the scutes of the carapace and plastron. In the past, it was thought that annuli corresponded to age, but findings suggest that this is not the case. However, the annuli are still counted since it may yield important life history information. I want you to calculate the sample correlation between **Annuli** and **Mass** for adult turtles. I don't care how much code you right, but I only should see one number in your output.

```
#
```

Q6 (4 Points)

I want you to go to the website https://r-charts.com/correlation/scatter-plot-group/ and observe the first visual that shows a scatter plot with different colors for the three groups. I want you to modify that code in the first visual for the data in **Adult_Turtles**. I want to see a scatter plot where **Annuli** is the response variable plotted on the y-axis and **Mass** is the predictor variable plotted on the x-axis. I want to see a legend with three different colors so the audience can clearly see which dots are male turtles and which dots are female turtles and which dots are turtles of an unknown sex. I want the legend in the bottom right of the picture instead of the top left. If you do this correctly, you will see that this is a better position for the legend.

You may need to run all the code on the website so you understand how the code works and what you need to do to get this to work for the turtles.

```
#
```

Q7 (5 Points) I want you to perform a two-sample t-test like we did in class (Welch's t-Test) for all the numeric variables in the **Adult_Turtles** dataset. For each t-test, you are testing to see if there is evidence that the population mean for Females is different than the population mean for Males. As you saw in the graphic, there are some adult turtles with an *Unknown* sex; therefore, you must remove these turtles from the data before you perform your t-tests. (Hint: Look up what != means in R and think about how this may be useful)

In your output, I want to see the output of all your t-tests. Then, I want you to write a response in complete sentences where you explain to your audience what you learned about the differences between male and female turtles in the context of the data. Your audience should know where you found statistical significance, but more importantly, the audience should know what this means for the turtles. Be accurate in your discussion of the results. I would recommend reviewing how to interpret the results of a t-test. Since you are discussing 7 variables, you want to be concise so I am not expecting a sentence for every variable. Just make sure your explanation is precise and accurate.

Usually, I would require you to not use the variable names in your explanation since the audience wouldn't know what "PL_AnteriortoHinge" means; however, I didn't quickly fine one place where all these variables are explained so it is okay to reference the variables by their names in the dataset. I do ask you to format your variable names in your paragraph using a pair of single asterisks which *italicize* whatever is written in the middle. This makes the output in the PDF look more professional.

```
#
```

Response in Complete Sentences: REPLACE EVERYTHING IN ALL CAPS HERE WITH YOUR ANSWER IN COMPLETE SENTENCES