Project 3: All About Climate

A potato chip company has three different manufacturing methods that they use to produce their chips. They have collected data on each of these manufacturing methods and have hired you to determine which method produces the most consistent chip.

General:

You will work through the project at your own pace. The project will often ask you to do things that you do not know how to do. Do not worry, this is intentional. Any time you feel stuck, ask a TA or the instructor for help. Alternatively, use Google/YouTube/Canvas to find resources to help you.

Instructions:

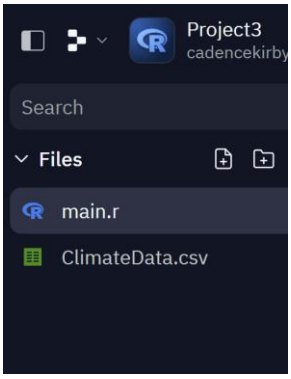
1.) Go to Canvas -> Files -> Project 3 and download a copy of ClimateData.csv. Import this into your Office 365 account by going to

Office 365 -> Home -> Upload (you may have scroll down a little bit to find the "Upload" tab)

2.) In this project we are going to be learning R, which is a programming language. We will be using an online R compiler/interpreter called Replit

3.) If you need to watch or rewatch the brief introduction to R and Replit you can do so by going to Canvas -> Files -> Miscellaneous -> IntroToR.mp4 (this will not be posted until after the 06/29 online session is over). In this project you will definitely need to use Google to look up functions, syntax, and errors. Google is one of the most powerful tools you can use when learning a new programming language.

4. Import the ClimateData.xlsx file into your project files. Your Files tab should look like this after your import.



5.) Read the data by typing this line and pressing the green "Run" button at the top. Replace everything inside the **.

```
# load climate data
climateData <- read.csv(*name of csv file*)
```

6. Print the data and then print a summary of the data. Copy the code you used to do this in your What is the difference between printing the data and printing a summary of the data? (Respond as a comment in main.r). *Note, for all coding projects don't delete any correct code that you write for any of these problems. At the end of the project you will be turning in main.r for your submission. If you don't want all of your previous code to output to the console just comment out your previous code.*

7. Import ClimateDataLegend.txt (which can be found in Canvas -> Files -> Project 3) into your files so you can decipher the column headers. What does the column header FG represent? (Respond as a comment in main.r)

8. How are empty values represented in this data? (Respond as a comment in main.r). Write and execute a statement to print out the number of empty values in this dataset (Respond as a comment in main.r)

9.

```
print(class(*variableName*))
```

What is our data variable stored as in R? Replace "variableName" with the name of your variable and run the line above in your script to get some more information about your variable in the console. Google the output of this line and write down what you learn as a comment in main.r

STOP: Check-in with TA or instructor before continuing

10. It is time to do some data cleaning, although we are only going to worry about duplicate data here. Write some code to identify duplicate years. For example, if there are two rows from 1979, your code should identify that. Finally, write some code to remove the duplicate data. Check to make sure you correctly removed duplicate rows before proceeding.

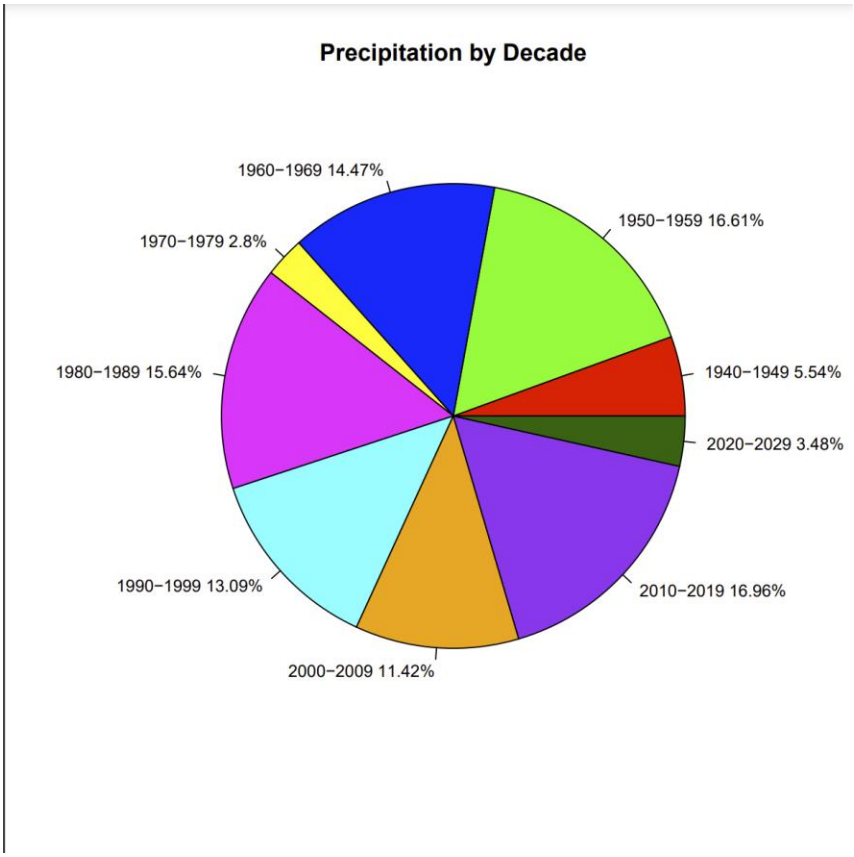
11. Compute the mean and median of each column in the data. Hint: The mean and median of each column can be calculated with a single line of code. Record these results as a comment in main.r.

12. Compute the standard deviation of the "T" column. Why does it return "NA"? Respond as a comment in r.main?

13. To fix the problem in 12, you need to tell the sd function in R to ignore NA values. Determine out how to do this and then calculate the standard deviation of each column in the data. Record your results as a comment in main.r.

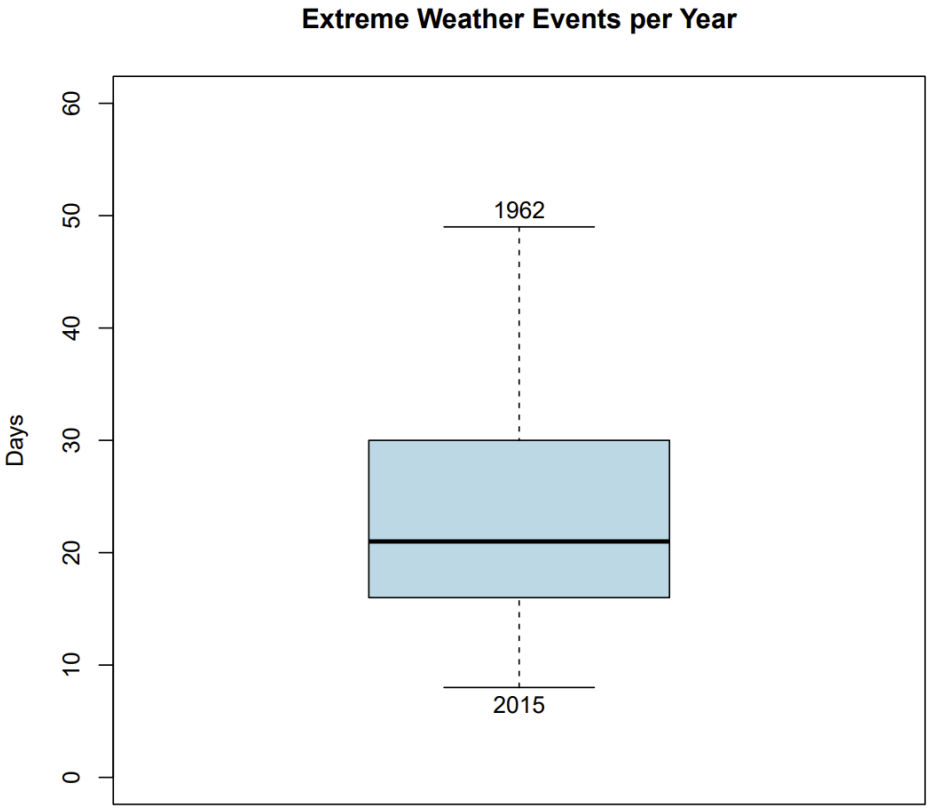
STOP: Check-in with TA or instructor before continuing

14. Using the data, create a pie chart that looks like this. Make sure to customize the graph title, colors, labels, and font.



Know that this problem will require several consecutive steps. All graphs made in Replit using R functions are uploaded to a file to the left called RPlots.pdf. Why do you think that the 70's had such little precipitation? Feel free to do some research, but respond as a comment in main.r.

15. Make a box-and-whiskers plot of all extreme weather events in the entire dataset. Make sure to include a title, axis labels, and label the min and max on the plot with the corresponding year. Extreme weather days are described as days with either snow, storm, tornado, or hail. For example, if the year 1944 had 30 snow days, 25 storm days, 5 tornado days, and 10 hail days, we would say that 1944 had 70 extreme weather events (30+25+5+10). Your plot should look something like below:



16. Make a histogram of all of the average annual wind speeds. Configure your bins to be 10-11, 11-12, 12-13,…, all the way up to 19-20. Make sure to include a title, axis labels, and proper scale on your axes. Which bin has the greatest frequency (respond as a comment in main.r).

17. Make a bar graph of number of rainy days vs. Year. Include a descriptive title as well as a label for minimum and maximum year. Introduce a color gradient for this graph so that the bars go from light blue to dark blue as you look left to right. Do you think that a bar graph is the best way to represent this data? If yes, why? If no, what other plot would you make instead? Respond as a comment in main.r.

18. Make a scatterplot of average annual maximum temperature vs. Year. From the plot, does it look like average annual maximum temperature is trending upwards? After you respond, watch this video about Correlation vs. Causation. Finally, record whether or not you can determine from this graph whether the relationship between average annual maximum temperature and year is causational. Respond as a comment in main.r.

STOP: Check-in with a TA or instructor before continuing.

19. Now it's time to get some more comfort with the logical side of R. Complete challenges 1-10 as mentioned in this article. You may have to Google some terms you are unfamiliar with.

STOP: Check-in with a TA or instructor before continuing.

20. Make a new script called RockPaperScissors.r. Your job is to code a game of rock-paper-scissors in which a user plays against the computer in the console. The computer can randomly generate either rock, paper, or scissors, or if you are feeling brave give the computer some basic strategy.

Congrats! You have finished Project 3! Make sure to submit your R scripts and your pdf of your graphs to Canvas.