Andrew Jop
Noah Greski
Programming Languages

Language Project Progress

Since the last report our main goal up until this point has been to learn the basics of R, but also to begin to manipulate and analyze data sets. Up to this point we have learned the foundations of the R programming language and how to write simple programs. Like any other language R has basic data types, there are numeric, character, logical, vectors and NA which is the equivalent to null and others. Assignment syntax is <- compared to = seen in other languages.You begin counting elements at position one not zero. The data types do not need to be defined explicitly in this language. The most basic data objects are Vectors, and there are six types with four of them being listed above and they are essentially lists of the same type.(W3Schools)Examples of these data types and basic R features can be seen Rexample1.R. This program shows a basic understanding of R and has some testing during our learning process of the language. R supports vectorized calculations with no need for loops to perform a single operation. R is an interpreted language meaning that there is no need for a compilation process to occur. Another benefit is that since R is open source there is a wide variety of packages to help many areas of study from data science to machine learning. R can present data through the use of a data frame. A data frame is an R object that stores tabular data in a table structure made up of rows and columns.(Codeacademy) Data frames present data in a table format, but what highlights the language R is the manipulation and visualization of your data. There are various examples shown in our repositories of manipulation of the data using the package ggplot2.

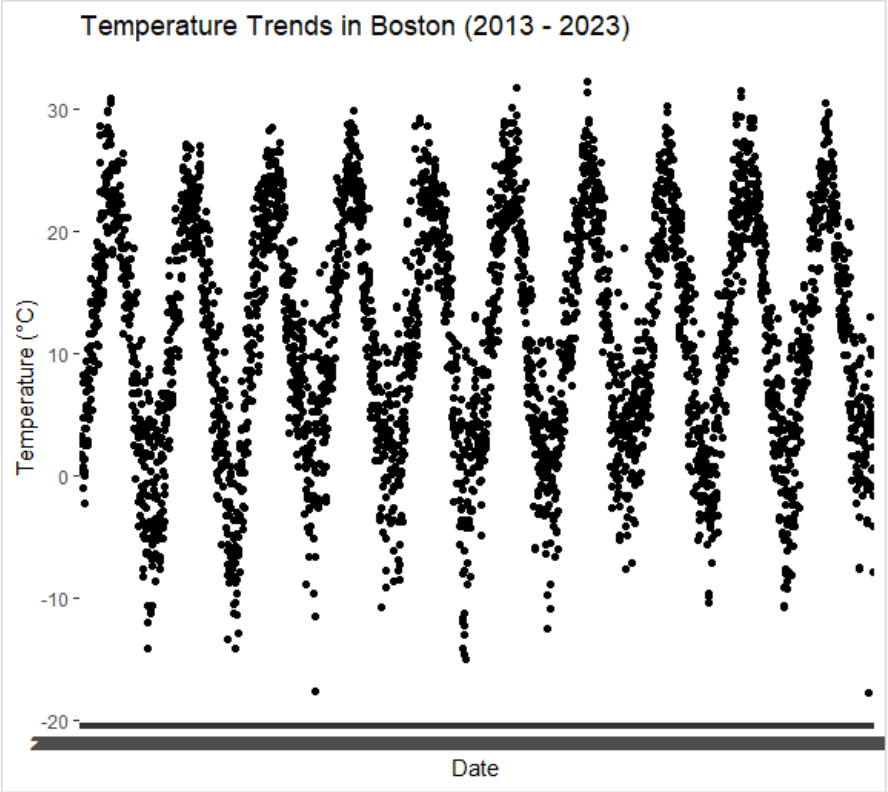
In the example Boston.R, we read a comma separated value(csv) file which has data about weather in Boston, Massachusetts from 2013 - 2023 using the read.csv() built in function. We retrieved this data from Kaggle(Meher 2023). From the data we read in we will then create a scatter plot using the ggplot() function with the first argument as our data, we then use aes() which is a ggplot2 function to set the label of our our x and y axes. We then add geom_point() to add points to the plot and set labels for the plot. As seen just in this initial block to create the scatter plot there are many functions that take care of the underlying work. After this we have created a bar chart based on the precipitation values in the csv file, but before we did this we had

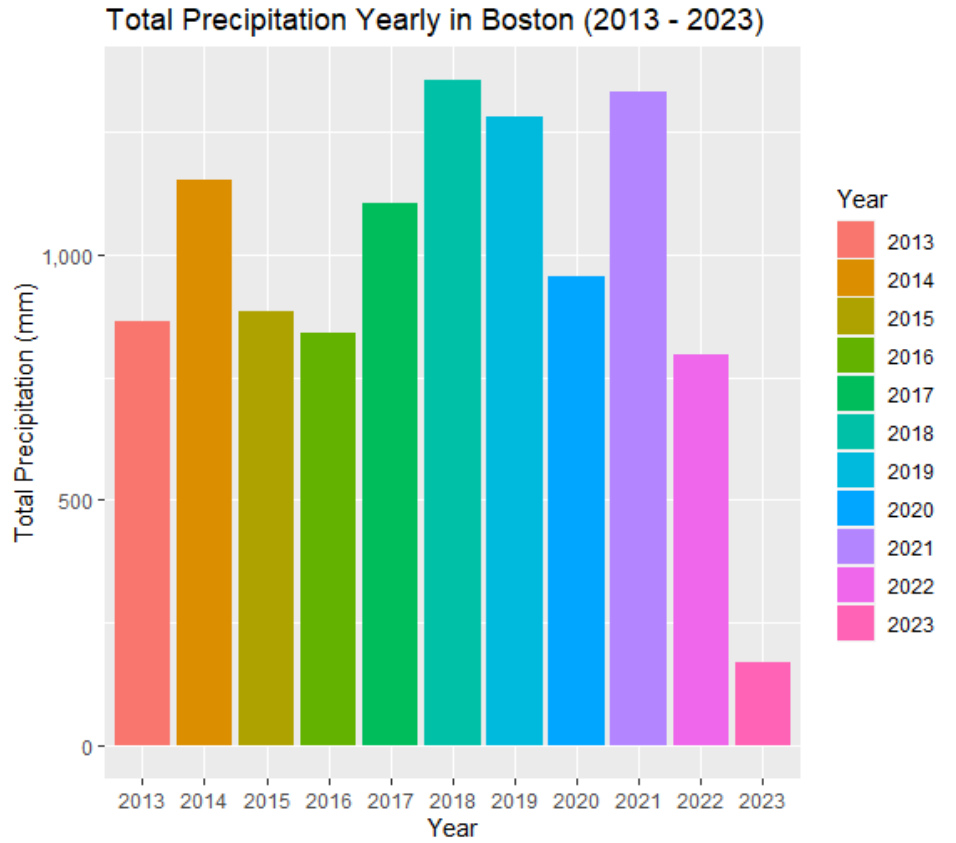
to create a date object to then create a year variable which takes the year value. This is done using the format() function from the time column in the csv specifying "%Y" to extract a four digit format. We can use the ggplot() as before and specify the parameters for a bar chart.

We created another example for data visualization in Chipotle.R. This example uses several packages to use a map to display locational data. The dataset used for this example is called "chipotle_stores.csv." This dataset contains coordinates with every Chipotle restaurant in the US. First as other examples did, the csv file is read into the program using the read.csv() function. Next objects are created for each location on the map using the longitude and latitude from the dataset. Then the US map is imported using the maps package and the map_data() function. Lastly the map is created with all the locations. Several of the map's features are edited including colors, sizes and titles. This was done by using the ggplot(), geom_polygon() and goem_sf() function. The final map is displayed in the "Plots" section in RStudio. All examples done with these smaller datasets have a similar pattern. Install and load the necessary packages, read in the dataset and then edit the features and display the plot. We got this data from Kaggle(Braun 2020).

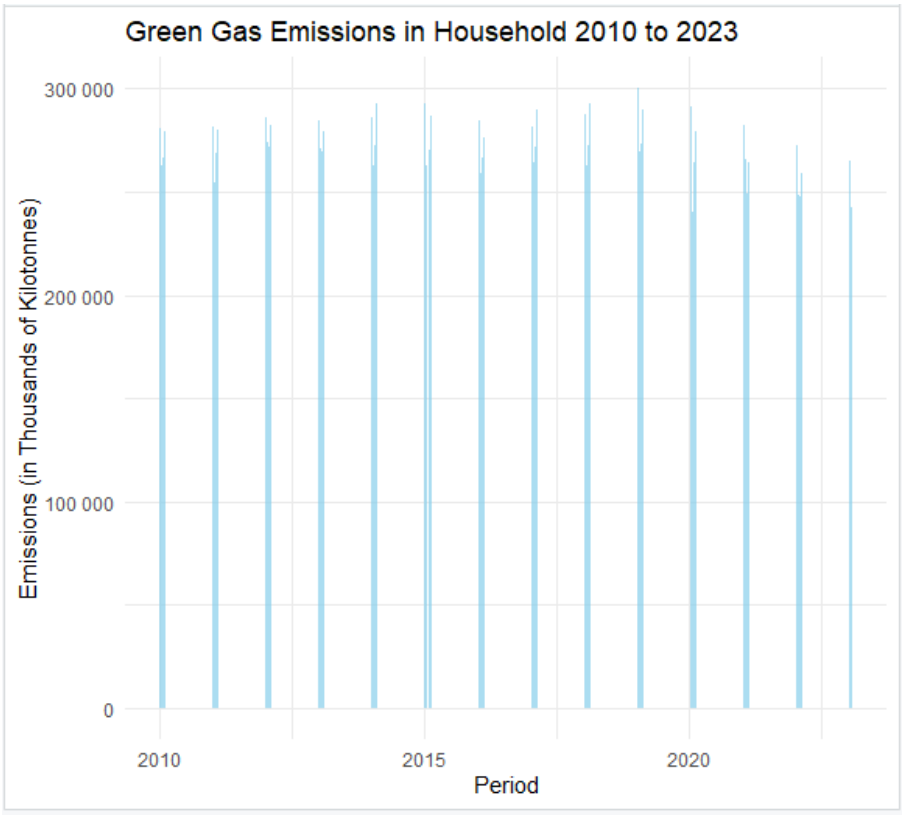
We plan to design an R program which uses a map that uses locations and based on a user input show the zoomed in location and present specific data based on the updated location. Disclaimer we may not do this. This plan may change slightly due to what is possible in this language. From what we have learned so far in R is that creating maps off datasets is possible as are several other data visualizations. However, the concept of this project may change based on what else we learn in the next few weeks. For now, we plan for an interactive map in R, with some sort of locational data included. We plan to have this larger programming portion completed by December 4th.

Boston Example :

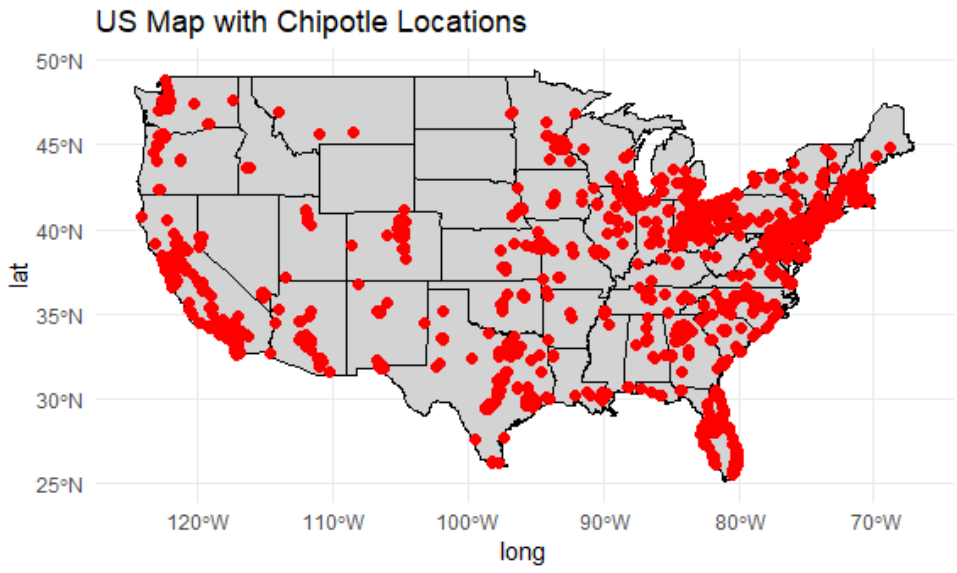




Emissions Example:



Chipotle Example:



Citations

Braun, Jeffrey. 2020, “Chipotle Locations”, *Kaggle*, https://www.kaggle.com/datasets/jeffreybraun/chipotle-locations.

Codecademy ,2023,www.codecademy.com/courses/learn-r/lessons.

Meher, Swaroop. 2023, “Boston Weather 2013-2023.” *Kaggle*, https://www.kaggle.com/datasets/swaroopmeher/boston-weather-2013-2023/

R vectors. (n.d.). https://www.w3schools.com/r/r_vectors.asp

Greenhouse gas emissions (industry and household): March and June 2023 quarters | Stats NZ. https://www.stats.govt.nz/information-releases/greenhouse-gas-emissions-industry-and-household-march-and-june-2023-quarters/