**PROGRAMMING FOR DATA ANALYSIS**

**U.S Weather Report Data Management**

APU2F2102CS(IS)

Abiyyu Taj Mahasin Bagindo

TP058652

Date Assigned: March 17th 2021

Date Completed:

**Introduction**

The objective of this project is to analyze the annual weather data report from the United States using the R programming language to help decision making. This expansive weather data contains multiple nation-wide variables such as temperature, rainfall, air pressure, sunshine, and more. The data records the aforementioned variables of every day in United States during a span of a whole year. Therefore, a wide analysis can be made, and multiple conclusions can be reached.

**First Question – When is the best time to dry your clothes?**

**Analysis 1-1: Which days have zero rainfall?**

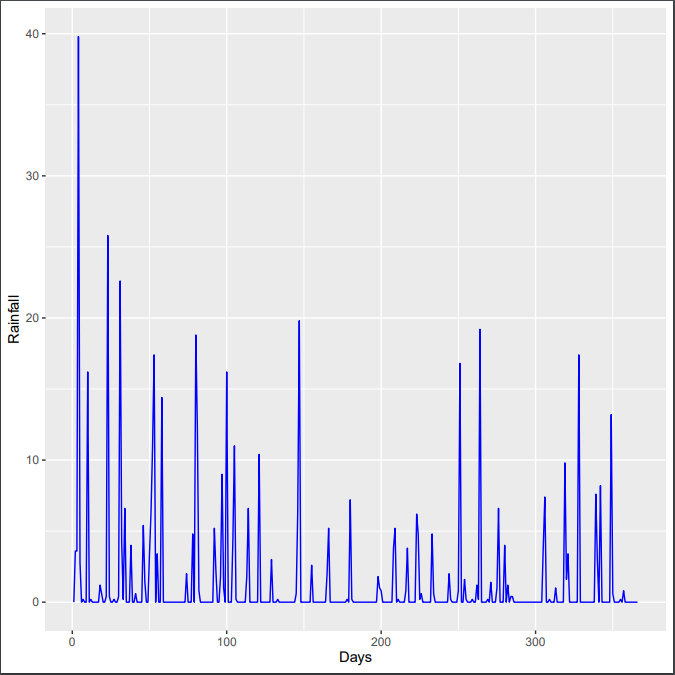


Figure 1 Rainfall graph

The first analysis is to filter out days with zero rainfall, because the last thing you want when drying your clothes is for them to be wet again. The code below starts with the initialization of a NULL vector with the name *rainyday* which will be used to store the days with zero rainfall. A variable called *rainydayindex* is then assigned a value of 1. A for loop that loops 366 times for each day of the year that contains an if statement then starts. The if statement checks whether the contents of the second column of the *rainfallframe* data frame, which is used to store the rainfall data is below 0.2 or not because 0.2 is the lowest value of rainfall. If it is lower than 0.2, that day has zero rainfall and the value of the counter variable *i* is then assigned to the index *rainydayindex* of the vector *rainyday*. The *rainydayindex* variable is then incremented by 1 and the loop repeats. This algorithm is used multiple times throughout the program and we will refer to it as *The Data Frame Crawler Algorithm* in the next parts of the document to reduce repetition. The algorithm shows that there are 264 days with zero rainfall.

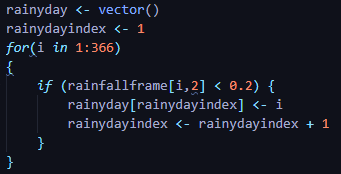


Figure 2 Zero rainfall filter code

**Analysis 1-2: Which days have high evaporation rate?**

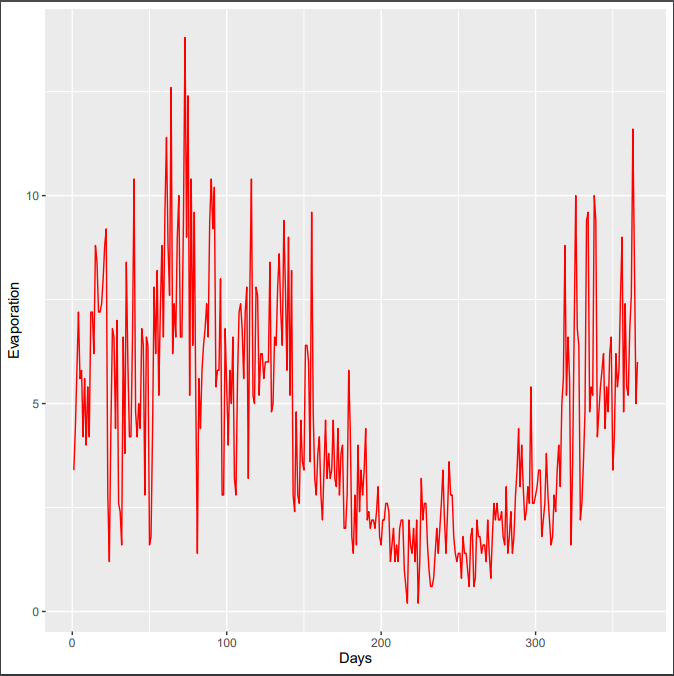


Figure 3 Evaporation graph

The second analysis is to filter out the days that have above average evaporation rate. The higher the evaporation rate, the faster the water on your clothes will evaporate, therefore the faster it will dry. The algorithm used in the code below is the data frame crawler algorithm which is the same one used in the previous analysis. With the difference being the data that is fed to the algorithm and the value that is used to filter out days. In this analysis, we need to find the days with high evaporation, as a bottom line for the filter, the average evaporation rate has been used. The average evaporation rate is obtained by using the *mean()* function which returns the average of the argument used in the form of a float. To use it as an integer, the *round()* function is then used which returns a rounded version of the argument. This rounded value is then assigned to the *evameanr* variable as an integer using the *as.integer()* function. The *evadayindex* is then printed at the end to display the days where the evaporation rate is above average. The algorithm shows that there are 147 days with above average evaporation.

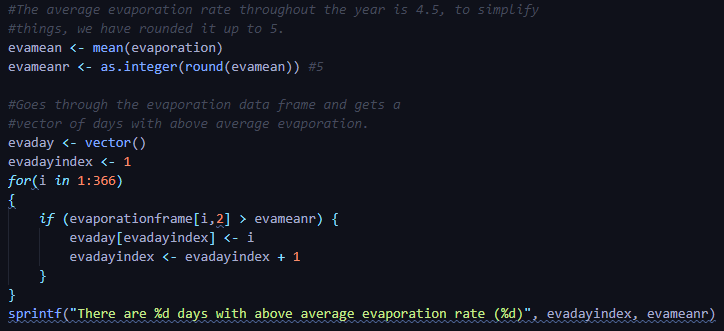


Figure Evaporation mean and filter code.

**Analysis 1-3: Which days have above average sunshine?**

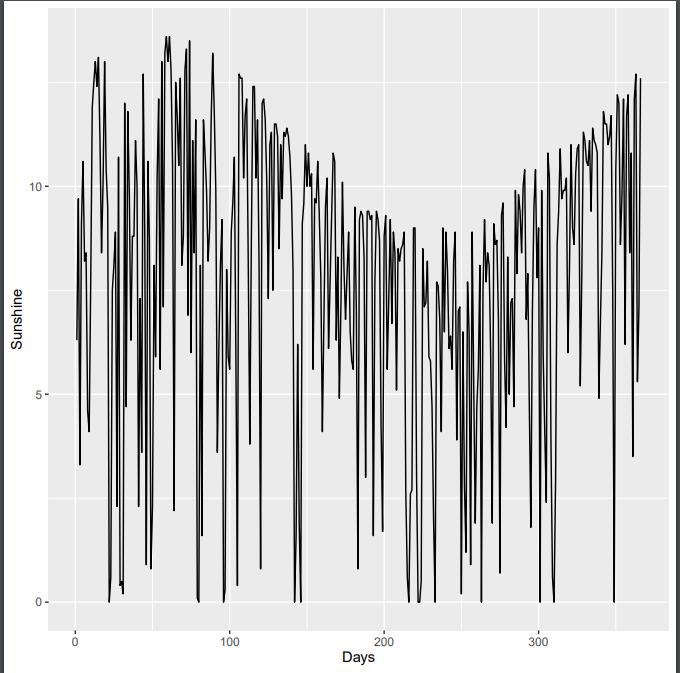


Figure 5 Sunshine graph

The whole activity of drying clothes naturally, is only possible due to the energy carried by sunlight. Therefore, the third analysis is to filter out days with above average sunshine. Once again, we used the same data frame crawler algorithm to find the days with above average sunshine. The code that finds the days with above average sunshine is the same as the one used in Analysis 1-2 with only the variable changes. The algorithm shows that there are 214 days with above average sunshine.

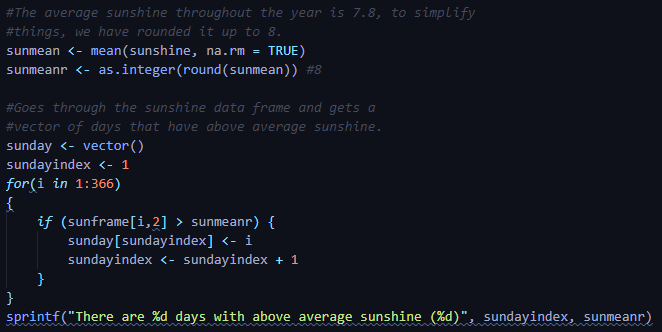


Figure 6 Sunshine mean and filter code

**Question 1: Results**

By using the three aforementioned analysis methods, it has been found that there are 264 days that do not rain, 147 days with above average evaporation rate (5), and 214 days with above average sunshine (8). The vectors containing the days will then be compared with each other to find days that exist in all three vectors. Those days will have all three properties for a good day to hang clothes. First, a null vector that will contain all days with zero rainfall and above average evaporation rate called *rainevaday* will be initialized. A variable called rainevaindex will also be initialized and assigned the value of 1. The following process works similarly to the data frame crawler algorithm but with an added layer and it compares values of two vectors instead. When the compared values of the two vectors matched, it will then be assigned as a value to the *rainevaday* vector. The process is then repeated with another vector which contains the days with above average sunshine and the days where all three properties are met will be printed. We will call this algorithm the *Vector Comparison Algorithm* for future references.

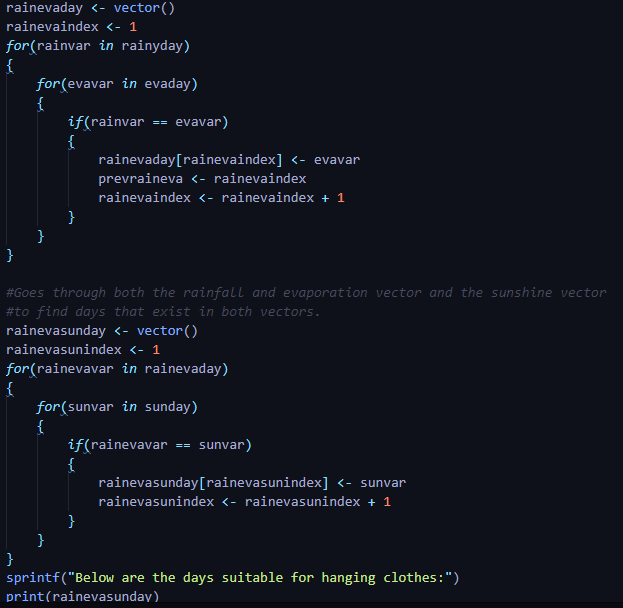


Figure Vector comparison code

**Second Question – When is the best time to exercise in the evening?**

**Analysis 2-1: Which days do not rain?**

**Analysis 2-2: Which days have comfortable levels of humidity in the evening?**

**30-50**

**Analysis 2-3: Which days have non-windy evenings?**

**Third Question – When is the best time to fly from the east to the west?**

**Analysis 3-1: Which days have winds travelling to the west?**

**Analysis 3-2: Which days have non-cloudy mornings?**

**Analysis 3-3: Which days have low rainfall?**

**Fourth Question – When is the best time to ski?**

**Analysis 4-1: When is winter?**

**Analysis 4-2: Which days have non-windy mornings?**

**Analysis 4-3: Which days lack of gust?**

**Fifth Question – When will a storm occur?**

**Ana 5-1: Which days have low air pressure?**

**Ana -52: Which days have high rainfall?**

**Ana 5-3: Which days have strong gusts of wind?**

**Conclusion**