**Table of Contents**

[**1.** **Introduction and Assumptions** 2](#_Toc99293605)

[**2.** **Aims and Objectives** 4](#_Toc99293606)

[**3.** **Background** 4](#_Toc99293607)

[**4.** **Data Import and Pre-Processing** 5](#_Toc99293608)

[**4.1.** **Install Packages** 5](#_Toc99293609)

[**4.2.** **Load Packages** 6](#_Toc99293610)

[**4.3.** **Read CSV File** 6](#_Toc99293611)

[**4.4.** **Pre-processing (Filtering origin)** 6](#_Toc99293612)

[**5.** **Analysis** 7](#_Toc99293613)

[**Analysis 1:** 8](#_Toc99293614)

[**Result:** 9](#_Toc99293615)

[**Analysis 2:** 10](#_Toc99293616)

[**Result** 12](#_Toc99293617)

[**Analysis 3:** 13](#_Toc99293618)

[**Result:** 14](#_Toc99293619)

[**Analysis 4** 15](#_Toc99293620)

[**Result** 16](#_Toc99293621)

[**Analysis 5** 17](#_Toc99293622)

[**Result** 18](#_Toc99293623)

[**Analysis 6:** 19](#_Toc99293624)

[**Result** 20](#_Toc99293625)

[**Analysis 7:** 21](#_Toc99293626)

[**Result** 22](#_Toc99293627)

[**Analysis 8:** 23](#_Toc99293628)

[**Result** 24](#_Toc99293629)

[**Analysis 9:** 25](#_Toc99293630)

[**Result** 27](#_Toc99293631)

[**Analysis 10:** 28](#_Toc99293632)

[**Result** 29](#_Toc99293633)

[**Analysis 11:** 30](#_Toc99293634)

[**Result** 32](#_Toc99293635)

[**Analysis 12:** 33](#_Toc99293636)

[**Result** 35](#_Toc99293637)

[**Analysis 13** 36](#_Toc99293638)

[**Result** 38](#_Toc99293639)

[**Analysis 14:** 39](#_Toc99293640)

[**Result** 40](#_Toc99293641)

[**6.** **Future Recommendation** 41](#_Toc99293642)

[**7.** **Conclusion** 42](#_Toc99293643)

[**8.** **References** 43](#_Toc99293644)

# **Introduction and Assumptions**

In this report, I had used different data analyzing and manipulation techniques to explore the given weather data set to extract the information which needs to conclude. Different techniques like data pre-processing, data visualization, data exploration, data manipulation were involved during this project which was already discussed in our course. Using all these techniques helps me to improve my visualization, exploration, manipulation skills in R programming. Using which at first, I import the given hourly data set in R Studio, then necessary pre-processing was done to apply appropriate commands so that the data set can be converted into the desired format. In this project, relevant graphics (i.e., bar graphs, scatterplots, histograms, and so on) were used to support the findings through R Studio.

The provided data set for this project is about hourly weather data set which contains two origins i.e., John F Kennedy International Airport (JFK) and LaGuardia Airport (LGA) of US. In the airport, every weather factor is seriously considered when taking the flight. It includes temperature, visibility, wind direction, pressure, and many more. Similarly, all those factors are kept inside a single hourly weather data set which consists of 15 columns and 17, 412 rows. All the descriptions of columns inside the dataset are listed below:

|  |  |
| --- | --- |
| **Columns** | **Description** |
| Origin | It describes the weather station |
| Hour, day, month, year | Time of recording the data |
| Temp, dewp | Temperature(in degree Fahrenheit) and dew point |
| humid | humidity |
| Wind\_dir,wind\_speed,wind\_gust | The wind direction in degrees, wind speed, and gust in mph |
| precip | Precipitation(in inches) |
| pressure | Pressure in millibars |
| visib | Visibility in miles |

Table 1: Description of columns

# **Aims and Objectives**

The major aim of this assignment is to use the R programming language to study the different factors of weather and extract necessary information from hourly weather data set for the year 2013 which cause climate change in two airports.

**Objective**

1. The primary objective is to import the given data set in R studio.
2. The secondary objective is to explore, analyze, manipulate and visualize the data set.

# **Background**

R programming language is used for different statistical computing and graphics. According to Johnson (2019), R is one of the most used programming languages in data mining. It was developed by Ross lhaka and Robert Gentleman in 1993. In the R programming package, ggplot2 has become the most popular package due to its aesthetic and interactivity. R programming is used in different fields and its purposes vary according to the sectors like finance, banking, healthcare, social media, e-commerce, and different manufacturing companies.

R programming language is used in most large companies like Uber, Facebook, Google, and so on. Similarly, Facebook uses R to update its status, social network graph and used to forecast how their colleagues would interact with R. Ford motors rely on R for statistical analysis and for assisting data-driven decisions. In google, R is used to calculate ROI on advertising campaigns, forecast economic activities, and increase the effectiveness of internet advertising. Microsoft uses R programming for their Xbox matchmaking services as well as a statistical engine with Azure ML architecture. Similarly, In Mozilla Firefox R programming is served as the backbone as it is used to display web activities (Burns, 2019).

If such big companies are using the R programming language, we can understand the different applications of the R programming language

# **Data Import and Pre-Processing**

## **Install Packages**

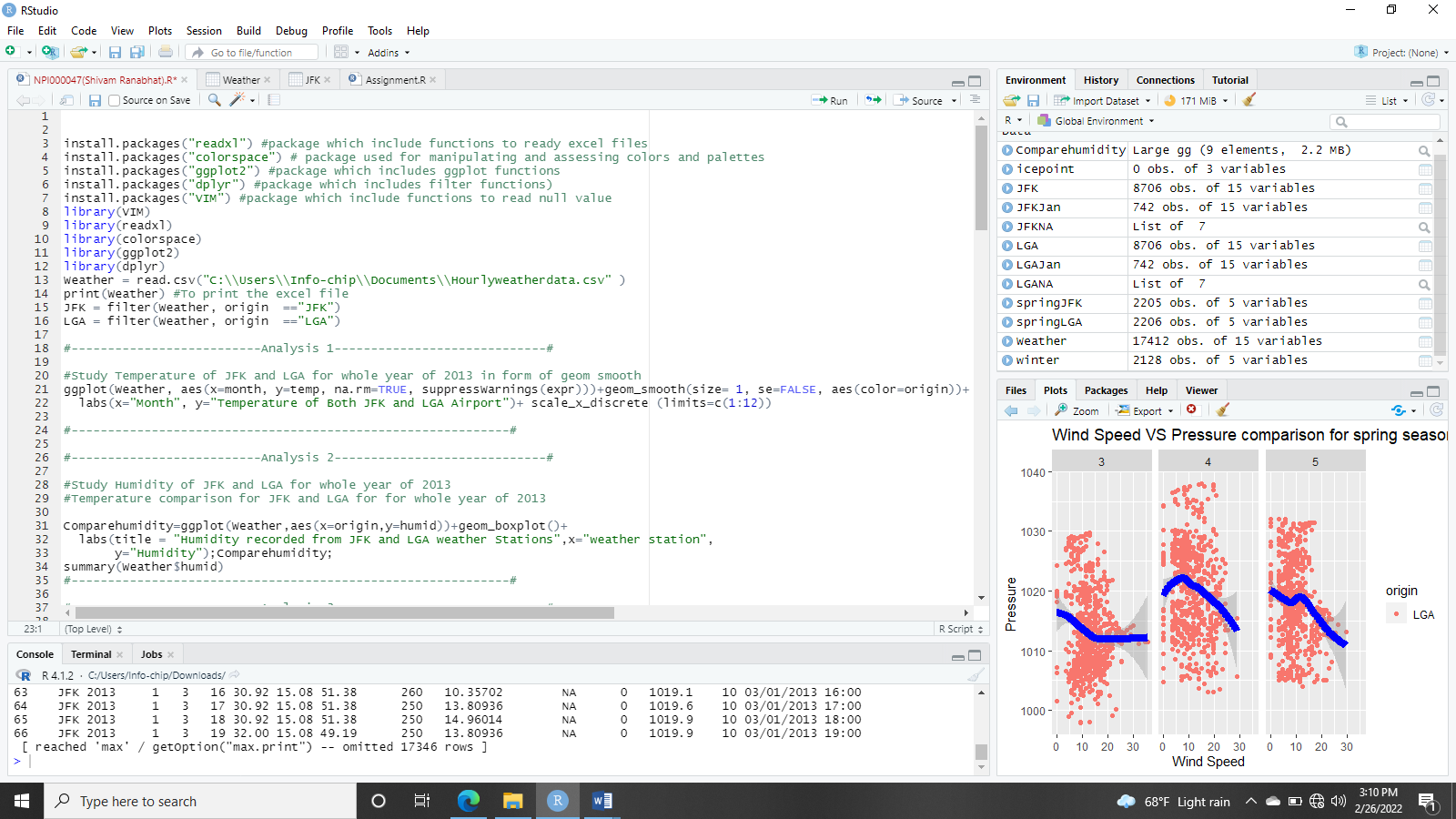


Fig 1: Installing packages

## **Load Packages**

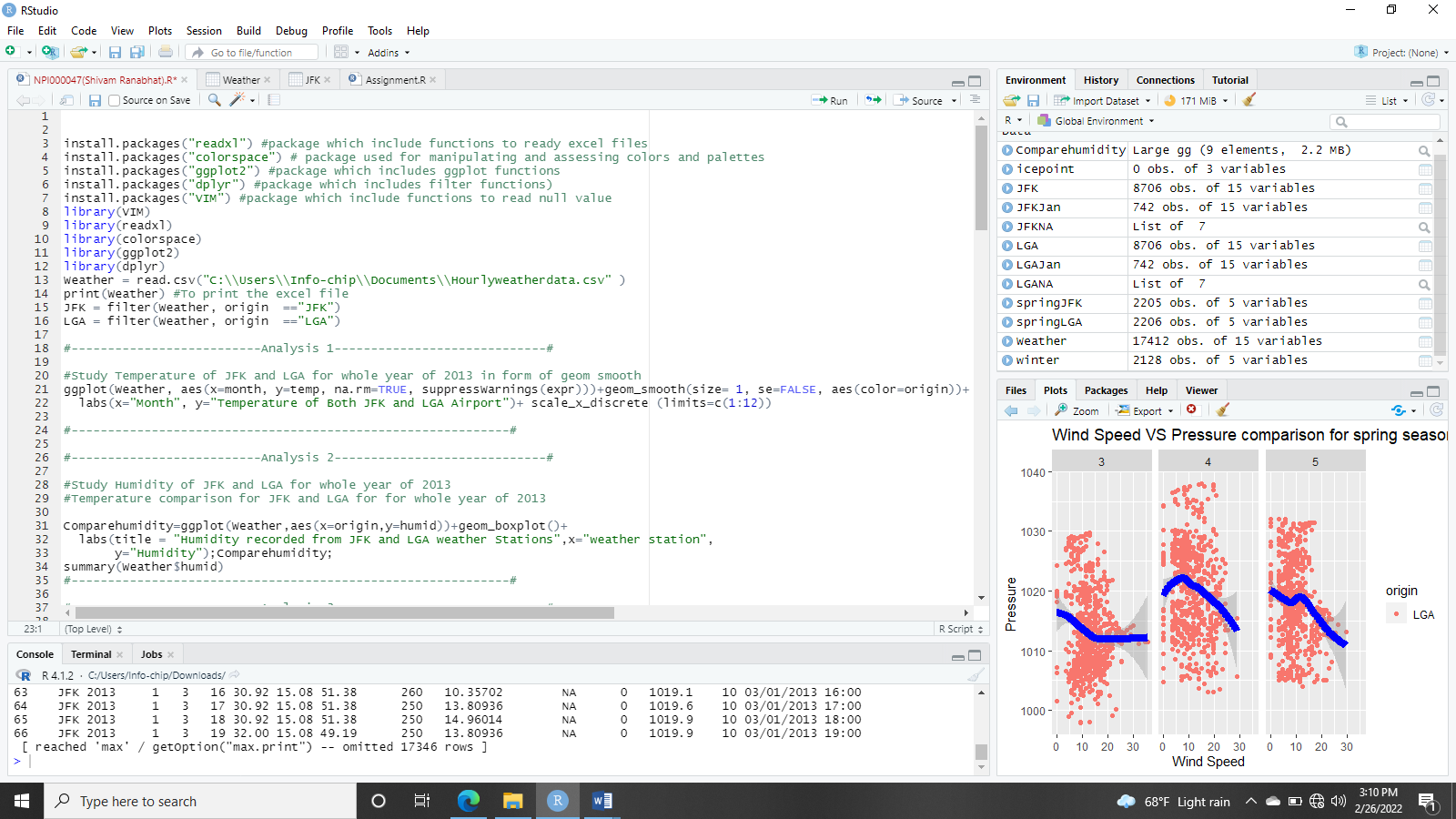
****

Fig 2: Load packages

## **Read CSV File**

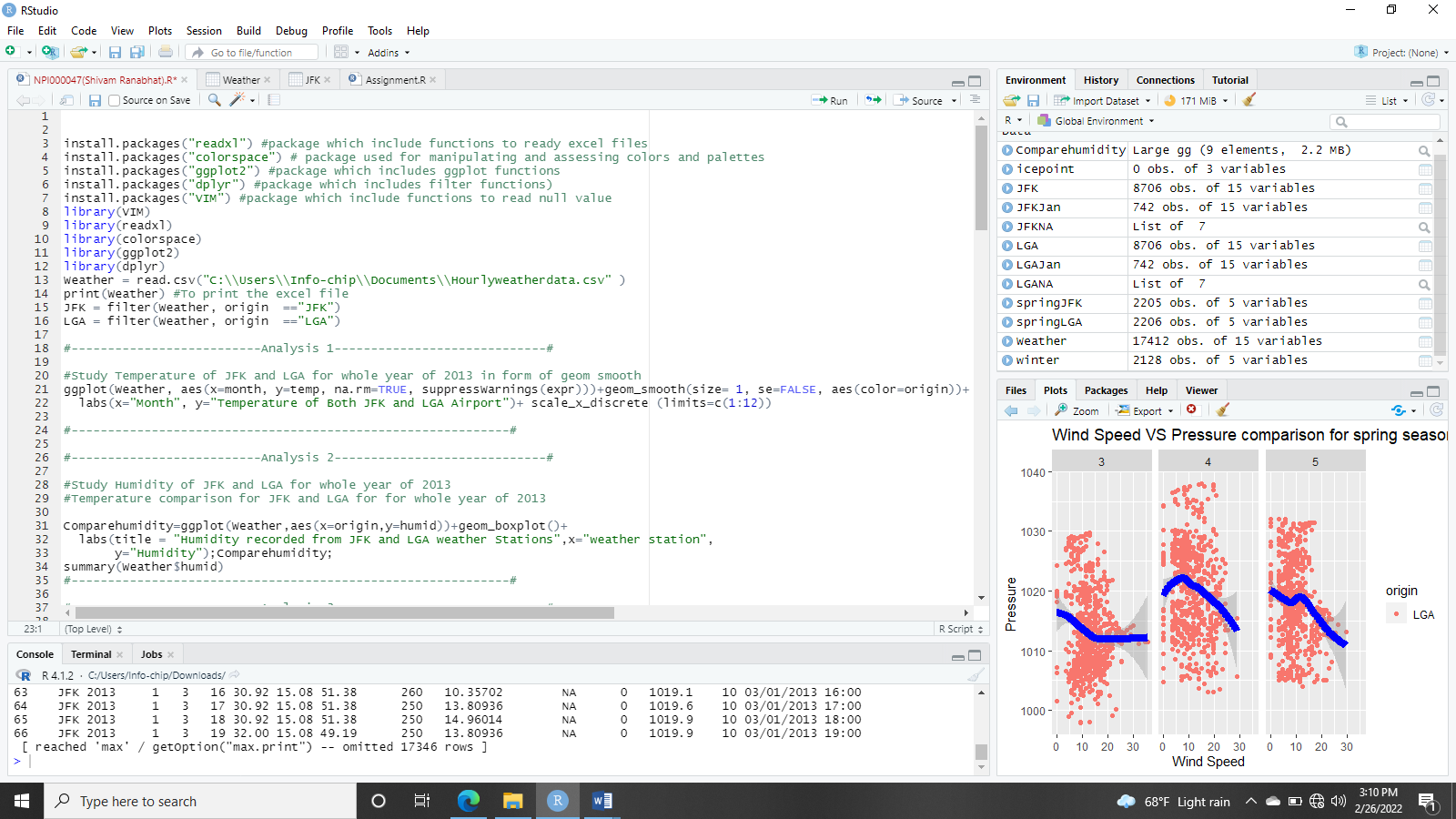
****

Fig 3: Read Csv file

At first, all the necessary packages should be installed using the install.packages() function, then after installing using library() function it is loaded in the R environment. Then the given file is imported into the Rafter viewing the extension of file format. The given file is in Excel CSV format, after installing the readxl package read.csv() function is used.

Before performing other techniques, necessary packages must be loaded to the R environment. Similarly, “ggplot2”, “VIM”, “colorspace”, “dplyr” packages are first installed using install.packages() function and then loaded using library() function.

## **Pre-processing (Filtering origin)**

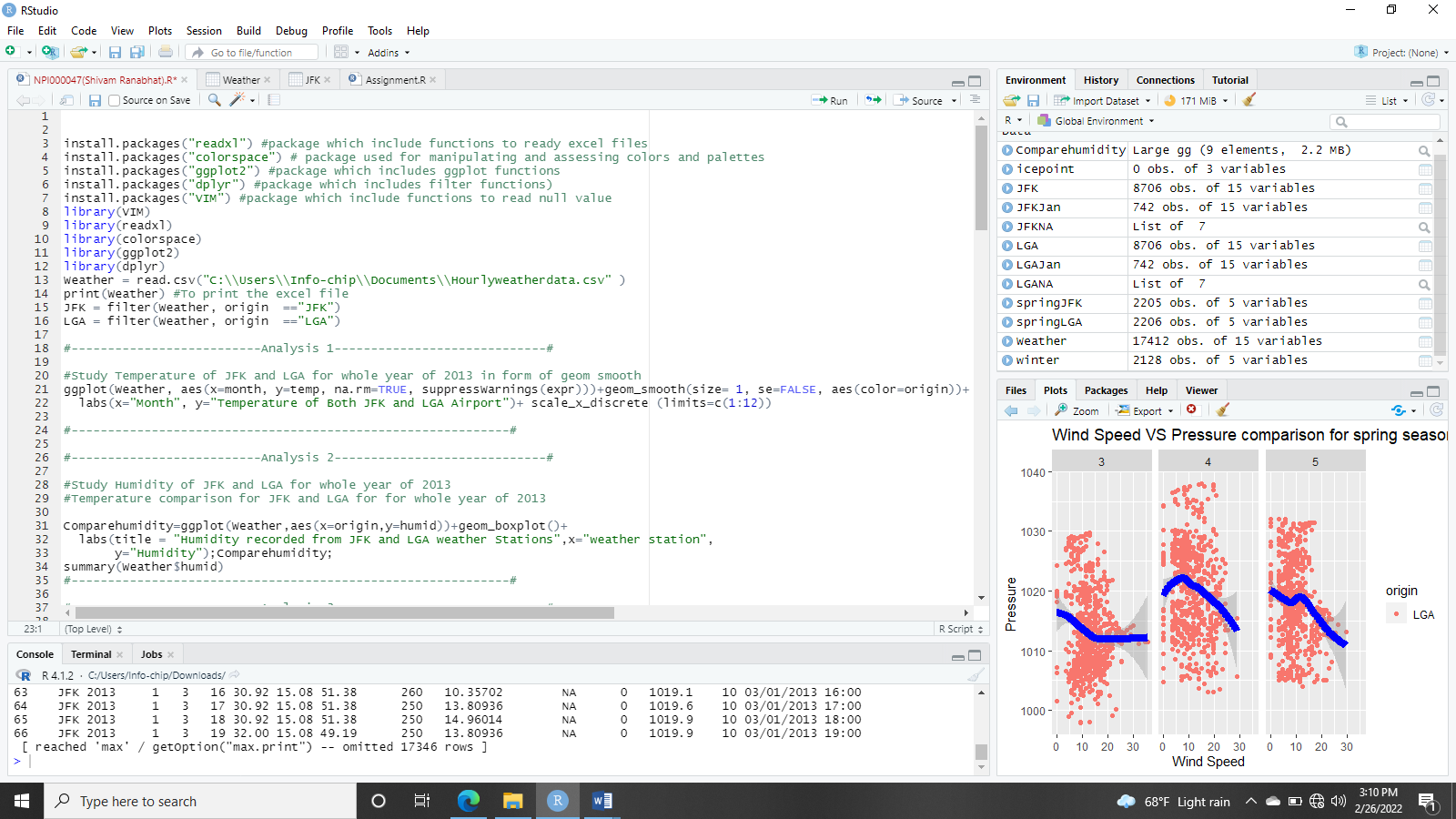
****

Fig 4: Data Filtering

As there are two origins in the data set i.e., JKF and LGA airport so, filter all the data based on origin and store it into two variables (JKF for JKF airport and LGA for LGA airport). Using this we can easily extract the necessary information of a particular airport.

# **Analysis**

Overview Topics of the analysis

Analysis 1: Comparing overall temperature of JKF and LGA

Analysis 2: Comparing humidity of overall year of JKF and LGA

Analysis 3: Comparing wind direction of the whole year 2013 of both origins

Analysis 4: Comparing wind speed of JKF and LGA for the Year 2013

Analysis 5: Comparing maximum precipitation of both origins for the whole year 2013

Analysis 6: Comparing the dew point of both origins for the whole year 2013

Analysis 7: Comparing wind gust of JFK and LGA for the Year 2013

Analysis 8: Finding the null value of January 2013

Analysis 9: Comparing January temperature below the ice point of both origins

Analysis 10: Comparing wind speed and pressure of spring season

Analysis 11: Comparing the temperature of Dec 22 of JFK and LGA

Analysis 12: Comparing the hourly dew point of 1st January of JFK and LGA

Analysis 13: Comparing the visibility of December of JFK and LGA

Analysis 14: Comparing the maximum pressure of 2013 in both origins

## **Analysis 1:**

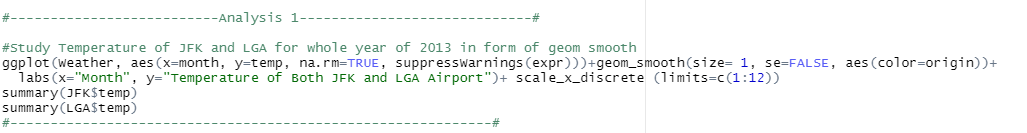


Fig 5: Temperature of JKF and LGA for the overall year 2013

* In this analysis the overall temperature of both origins is compared throughout the year 2013. The main purpose of this analysis is to study the monthly rise and fall in temperature in the year 2013 for both airports.
* In the above code, I have plotted two continuous position variables in a graph using the geom\_smooth() function and along the x-axis, I have filled month scaling it from 1 to 12 using scale\_x\_discrete() function as the data consist from January to December. The warning function is removed using the suppresswarning () function.
* Using summary function(), the minimum temperature, 1st quartile, median, mean, 3rd quartile, and maximum temperature are found as follows:



Fig 6: summary of overall temperature

## **Result:**

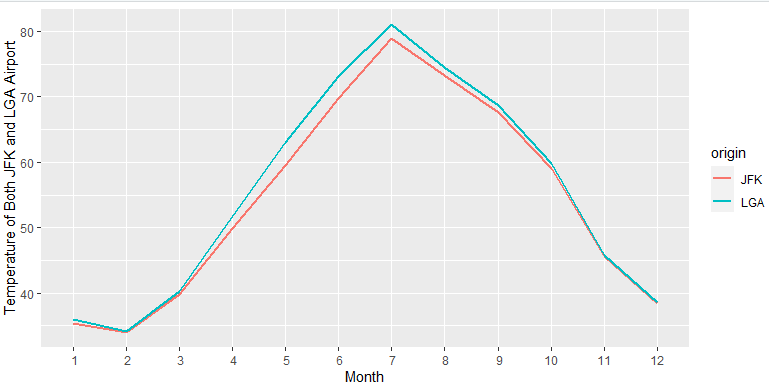
****

Fig 7: Temperature of JKF and LGA

In the above diagram, the red line represents the temperature of JKF and blue represents the temperature of LGA airport, and 1 to 12 represents the number of months. The temperature of both airports is found to be a minimum in February. The maximum temperature is found in July in both airports. It starts increasing from February and to July and then slowly decreasing from July to December. While comparing both the airports the maximum temperature is found in LGA airport and the minimum temperature is almost the same in both airports.

## **Analysis 2:**

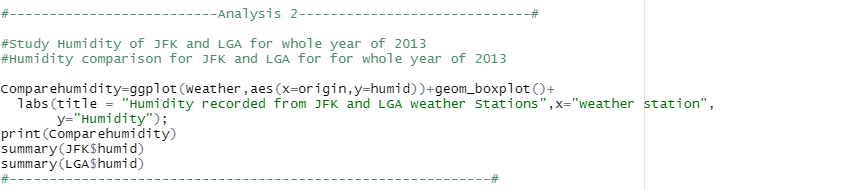
****

Fig 8: Comparing humidity of the overall year

* This analysis is about comparing the humidity of both JKF and LGA airports throughout the year 2013. Using this analysis, we can find the difference in humidity found in both airports along with the airport which has maximum and minimum humidity in the whole year 2013.
* In the code of the above figure, I have stored the annual humidity in the “compare humidity” variable. The data is further displayed in the form of the boxplot using the geom\_boxplot() function.
* Using the summary() function, the minimum temperature, 1st quartile, median, mean, 3rd quartile, and maximum humidity are found as follows:



Fig 9: summary of humidity of JKF and LGA

## **Result**

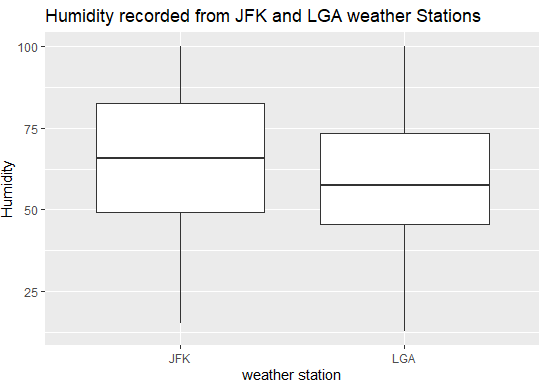
****

Fig 10: Boxplot of JFK and LGA for the humidity of the 2013 year

From the above figure, the maximum humidity is the same throughout the year 2013, the humidity is minimum in LGA than JFK, the median of humidity is higher in JFK than LGA. Similarly, the 1st and 3rd quartile for humidity in JFK is found to be greater than the LGA.

## **Analysis 3:**

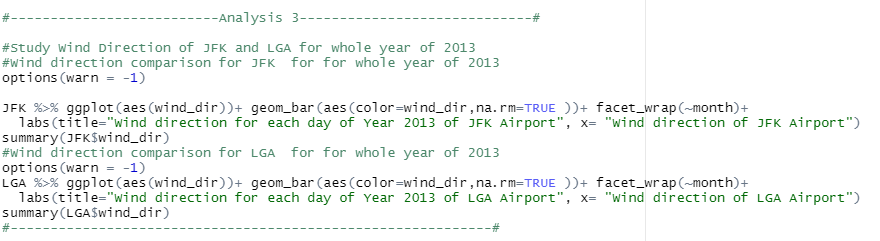
****

Fig 11: Wind Direction of the whole year 2013 of both origins

* In this analysis, the wind direction of the overall year 2013 is compared for both JFK and LGA airports. From this analysis, we can compare the wind direction in the degree of both airports.
* In the code shown above, I have used the option(warn = -1) is used to avoid the warning. Similarly, the data is displayed in the form of a bar graph using the geom\_bar() function, and passing the parameter na.rm = TRUE the null value present in the wind\_dir column will be removed.
* The overall summary is displayed using the summary() function which contains minimum temperature, 1st quartile, median, mean, 3rd quartile, and maximum temperature as follows:

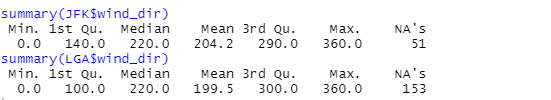
****

Fig 12: Summary of humidity for JFK and LGA

## **Result:**

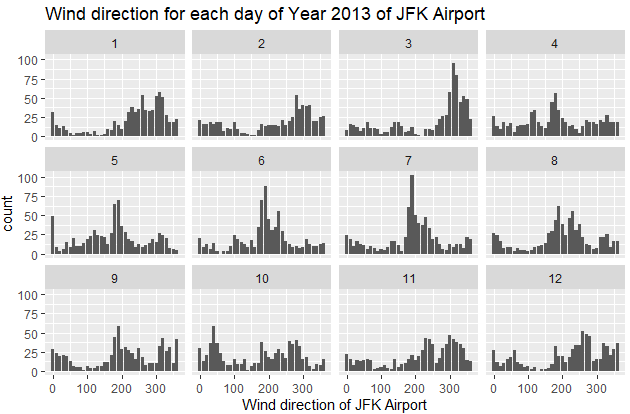
****

Fig 13: Wind direction of JFK for the year 2013

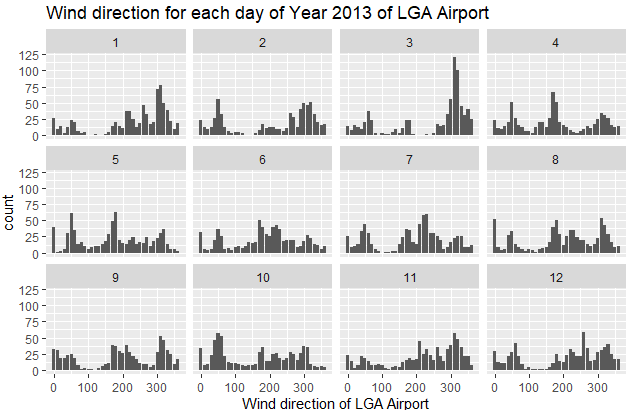
****

Fig 14: Wind direction of LGA for the year 2013

From the above two figures, the JFK airport has a higher wind direction in comparison to LGA airport. In March, June, and July the wind direction is found to be higher than other months in JFK airport. Similarly, In LGA airport, the wind direction value is found higher only in March. It concluded that the wind direction is found higher in spring seasons rather than others.

## **Analysis 4**

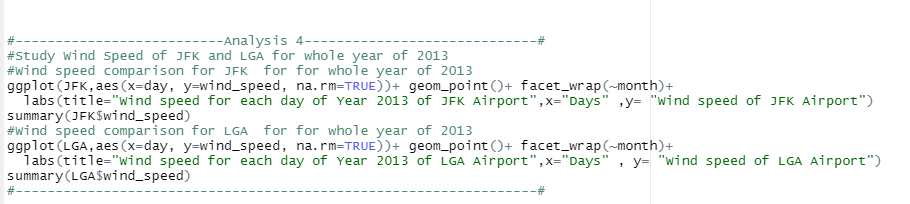
****

Fig 15: wind speed of JFK and LGA of 2013

* In this analysis, the overall wind speed of the year 2013 is compared between JFK and LGA. A scatterplot is filled individually for both JFK and LGA so that it would be easy for comparing the wind speed.
* In the above code, along the y-axis wind\_speed, is filled in relation with days along the x-axis, and the geom\_point() function is used to plot the scatterplot.
* Similarly, a summary() function is used to find out min, max, mean, number of null values, and so on.

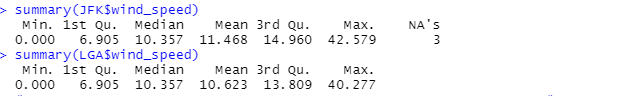
****

Fig 16: Summary of wind speed of JFK and LGA

## **Result**

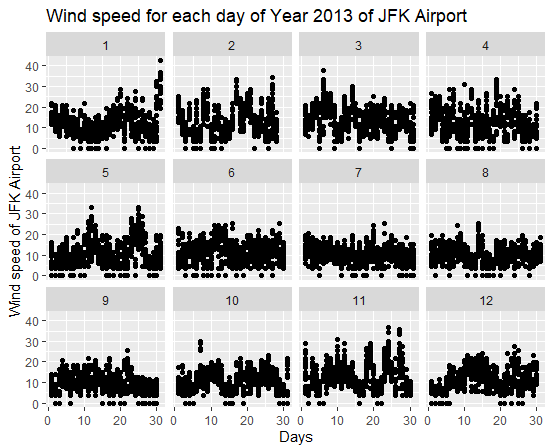
****

Fig 17: Scatterplot for the wind speed of JFK Airport

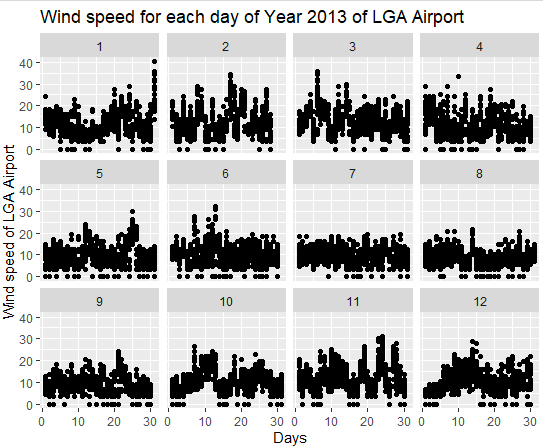
****

Fig 18: Scatterplot for the wind speed of LGA Airport

By comparing the above plots of two airports it is found that the wind speed of both JFK and LGA is maximum of January but the wind speed is maximum in JFK than LGA airport which is found to be 42.579 mph. But the minimum value is the same for both JFK and LGA which is 0 mph. After January, the wind speed is higher in November month of both airports.

## **Analysis 5**

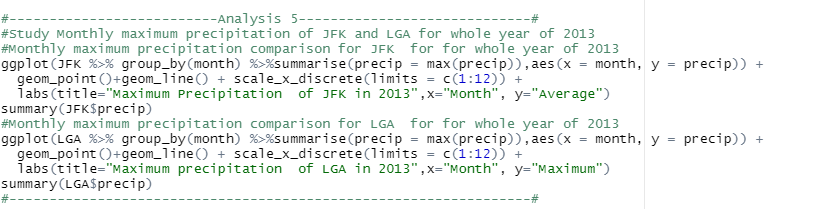
****

Fig 19: Maximum precipitation of both origins for the whole year 2013

* In this analysis, the monthly maximum precipitation of both JFK and LGA is compared for the whole year of 2013. For this, a line is plotted using the geom\_line() function where precipitation along the y-axis and month along the x-axis is plotted. And using summarise () function the precipitation column is selected with maximum values. Scale\_x\_discrete() function is used to set the number of months that are placed in the vector.
* summary() function is used to retrieve min, max, median, mean, and so on.

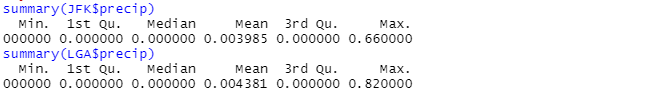


Fig 20: Summary of maximum precipitation for the year 2013

## **Result**

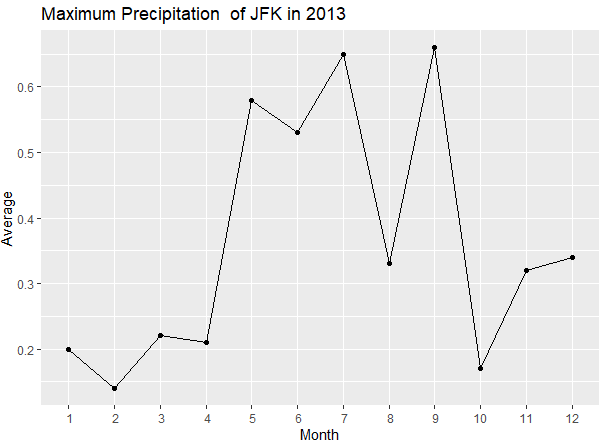


Fig 21: Line plot for maximum precipitation of JFK for 2013

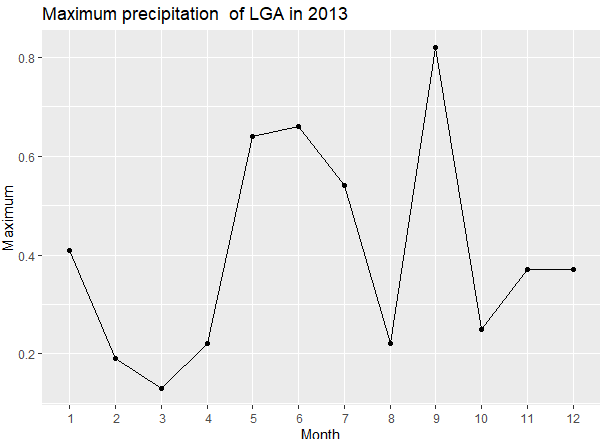


Fig 22: Line plot for maximum precipitation of LGA for 2013

From the above figure, maximum precipitation is found in LGA airport. In LGA airport the precipitation is found maximum in July and September. Similarly, in LGA the maximum precipitation is found in September month which is 0.82 inches. In JFK the maximum precipitation is found to be 0.66 inches. And both airports have the same minimum values. In the rest of the months, the maximum precipitation is normal in both JFK and LGA airports.

## **Analysis 6:**

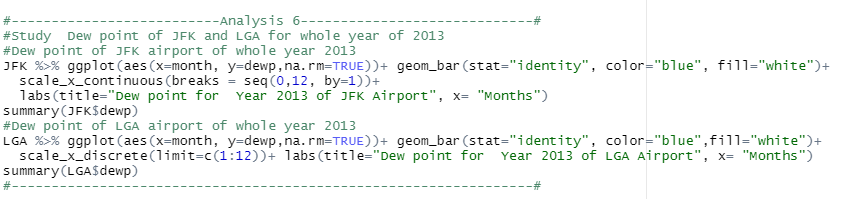


Fig 23: Dew point of both origins for the year 2013

* In this analysis, the dew point of both JFK and LGA is compared for the whole year of 2013. For this, a bar is plotted using the geom\_bar() function where the dew point along the y-axis is plotted. The background color is provided white using fill and blue to the bars. Scale\_x\_discrete() function is used to set the number of months that are placed in the vector.
* Similarly, a summary() function is used to find the min, max, mean, median, and so on as follows:

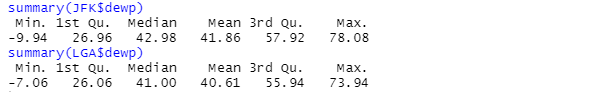


Fig 24: Summary of dew point for the year 2013

## **Result**

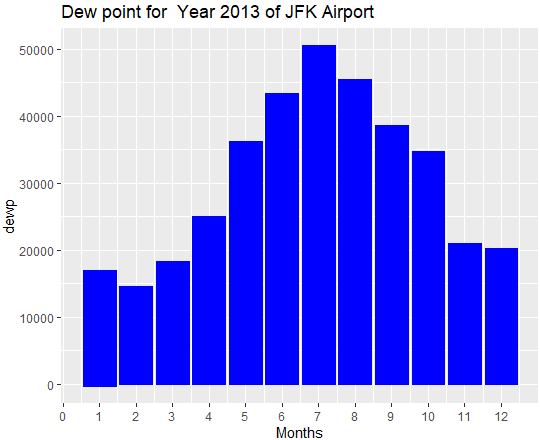
****

Fig 25: Bar chart for the monthly dew point of JFK

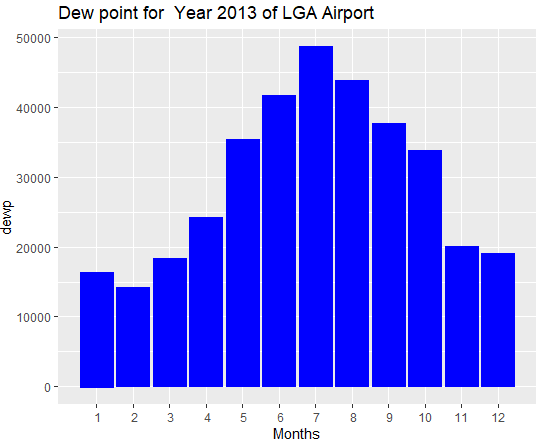


Fig 26: Bar chart for the monthly dew point of LGA

From figures 24 and 25, we can compare the overall monthly dew point of JFK and LGA. The maximum value of dew point is found in July of both airports where JFK has a higher dew point with a value of 78.78 degrees Fahrenheit than LGA whose maximum value of dew point is 73.94 degrees Fahrenheit. Similarly, the minimum dew point is found in February where the value is -9.94 degrees Fahrenheit in JFK is greater than LGA airport having a dew point of -7.06 degrees Fahrenheit.

## **Analysis 7:**

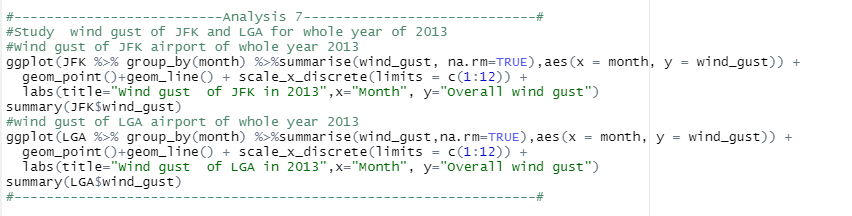
****

Fig 27: Wind gust comparison for JFK and LGA

* In this analysis, the wind gust of the whole year 2013 for both JFK and LGA airports is compared in the form of a line graph. In the above analysis ggplot() function is used to plot the graph which is grouped by month using group () function, summarise() function is used to select the desired column, geom\_line() and geom\_point() is used to fill the graph in the form of the line with points and scale\_x\_discrete is used to provide the numbers for the month along the x-axis.
* summary () function is used to find the overall summary i.e., min, max, median, number of null values, and so on.

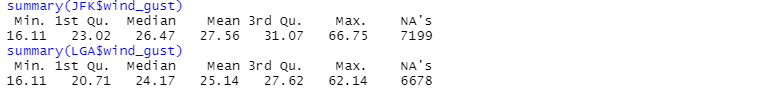


Fig 28: Summary of Wind gust for both Airports

## **Result**

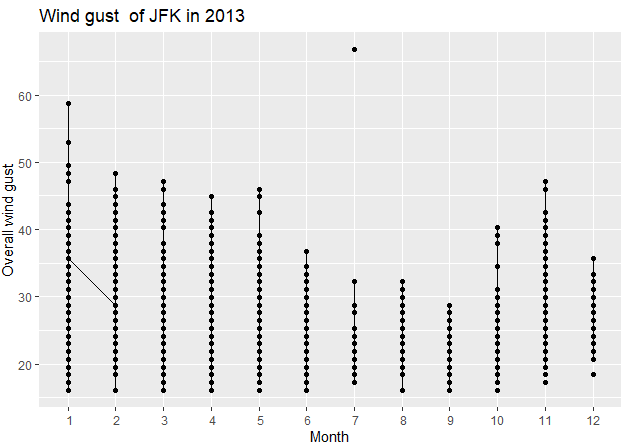


Fig 29: Line plot with points of wind gust in JFK

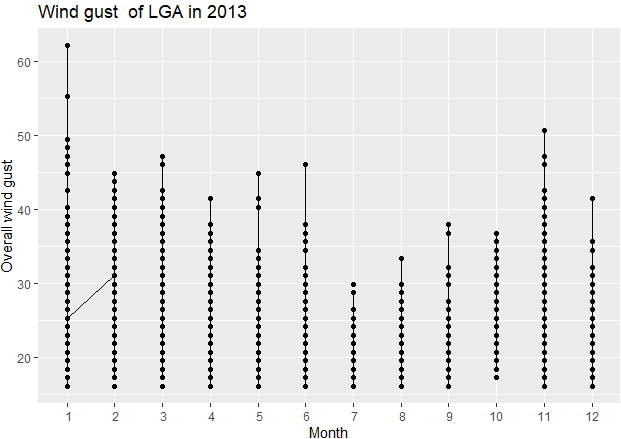


Fig 30: Line plot with points of wind gust in LGA

After comparing both figures 28 and 29 it is found that the wind gust is higher in JFK in July which is 66.75 mph. Similarly, in LGA airport, the maximum value of wind gust is found in the month of is 62. 14 mph. In December both wind gusts are found slightly higher after the overall maximum value. Similarly, the minimum value for wind gust is the same for both JFK and LGA whose value is 16.11 mph.

## **Analysis 8:**

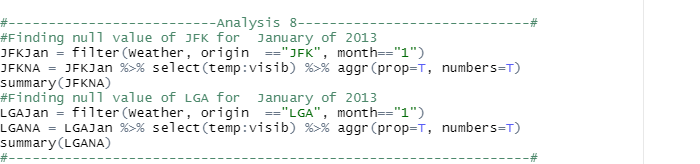


Fig 31: Finding the null value of January 2013 of JFK and LGA

* This analysis is about finding the null values and the columns where the maximum and a minimum number of values are missing. After this, both JFK and LGA are compared about the number of null values.
* In this analysis, pre-processing is performed to filter out the origin and their data using the filter () function from the weather data.
* To plot the graph of proportion aggr () function is used which is enabled due to installing and loading the VIM package. After providing TRUE to both proportion and number the graph is generated which shows the data both in proportion and number of missing values.
* Similarly, the summary () function is used to find the missing numbers along with their columns as follows:

|  |  |
| --- | --- |
| C:\Users\Info-chip\Pictures\Screenshots\Screenshot (238).png | C:\Users\Info-chip\Pictures\Screenshots\Screenshot (245).png |

Fig 32: Summary of null values of JFK and LGA

## **Result**

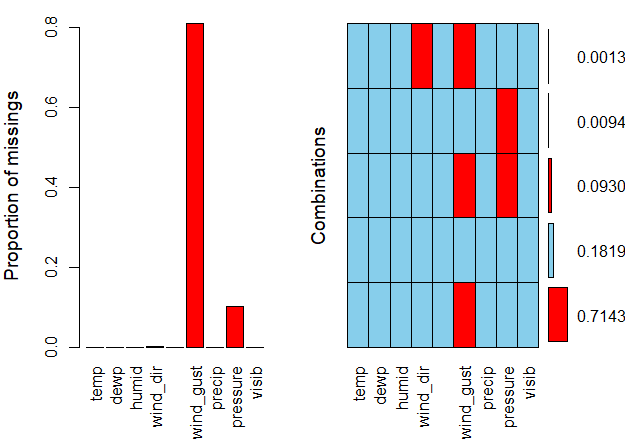
****

Fig 33: Proportion graph of missing values of JFK

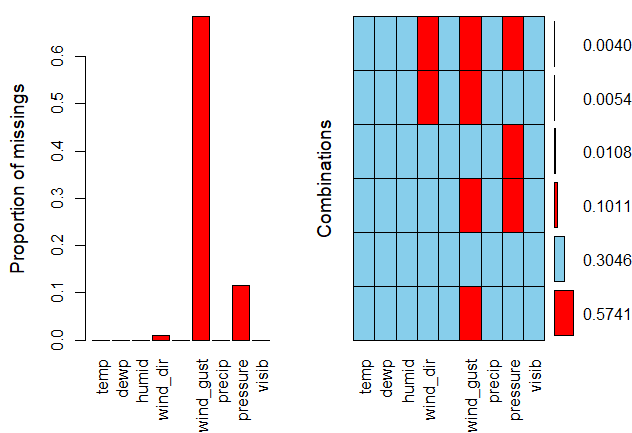


Fig 34: Proportion graph of the missing value of LGA

From figures 32 and 33, it is found that most null values are found in JFK airport. The null values are a little less in LGA compared to JFK. Similarly, the values are mostly missing in wind\_gust and then pressure and wind\_dir in both airports. The missing value of wind\_dir is found more in LGA than in JFK airport.

## **Analysis 9:**

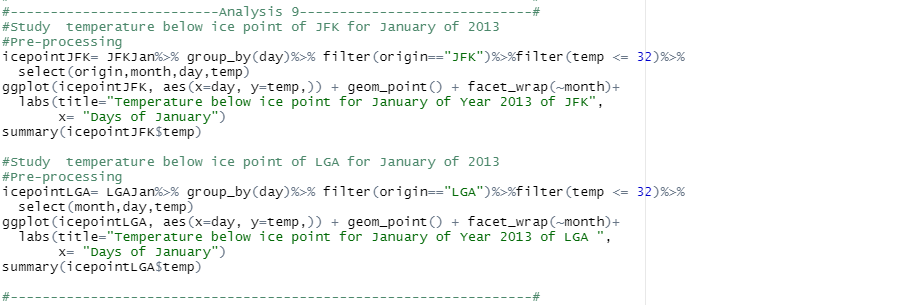
****

Fig 35: Comparing January temperature below ice point of both origins

* In this analysis, the temperature of January below ice point i.e., less than or equals 32 degrees Fahrenheit. At first, I filter the origin including the temperature which is less than or equal to 32, and using the select () function I select origin, month, day, and temp and stored it in a variable i.e., “icepointJFK” for JFK airport and “icepointLGA” for LGA airport.
* Similarly, after completing the pre-processing I use the ggplot () function where the day is kept along the x-axis and filtered temp along the y-axis, and using geom\_point () the graph is plotted in the form of point.
* summary () function is used to find the overall summary i.e., min, max, median, number of null values, and so on.

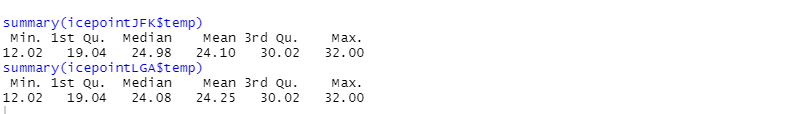


Fig 36: Summary of temperature below ice point

|  |  |
| --- | --- |
|  |  |
|  |  |

Table 2: Table generated from pre-processing

## **Result**

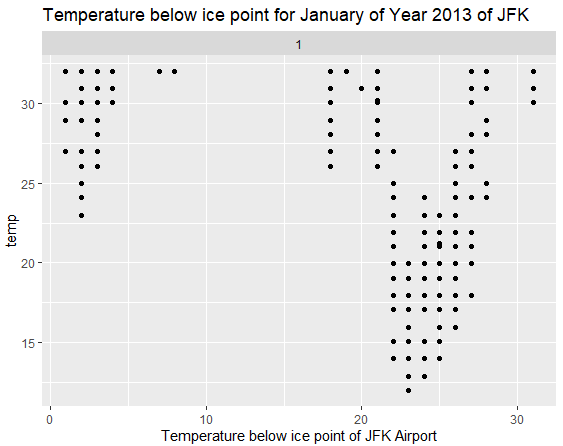
****

Fig 37: Point plot for temperature below ice point of JFK

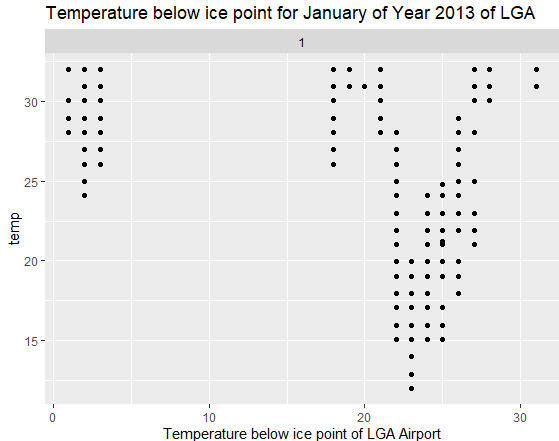
****

Fig 38: Point plot for temperature below ice point of LGA

From above figures 36 and 37, the temperature ice point of both airports is found almost similar. The minimum value of temperature below ice point is also the same for both airports. The temperature below the ice point does not found from day 5 to day 15 in LGA airport from day 5 to 15 in JFK airport. In both airports, the temperature below ice points is mostly found on day 25 to day 27. So, we can say that the temperature is cold at the end of January.

## **Analysis 10:**

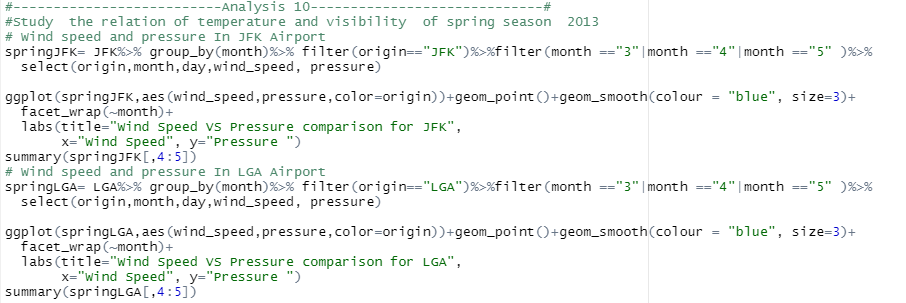
****

Fig 39: Comparing wind speed and pressure of spring season in JFK and LGA

* In this analysis the relation of wind speed and pressure is compared for the spring season i.e., March, April, and May. As the spring season is known as the windy season so, the effect of wind speed and pressure generated is shown in this analysis.
* For pre-processing, the filter () function is used to filter the origin and month then using the select () function, origin, month, day, wind speed, and pressure is selected and stored in a variable (springJFK for JFK and springLGA for LGA). Then, using the ggplot() function it is filled in a graph using geom\_smooth() in the form of the regression line.
* Similarly, the summary () function is used to find the overall summary i.e., min, max, median, number of null values, and so on.

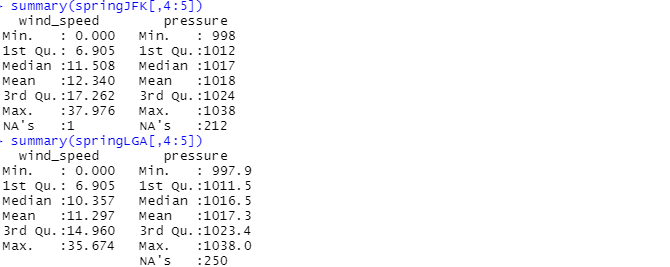


Fig 40: Summary of wind speed and pressure of spring season

## **Result**

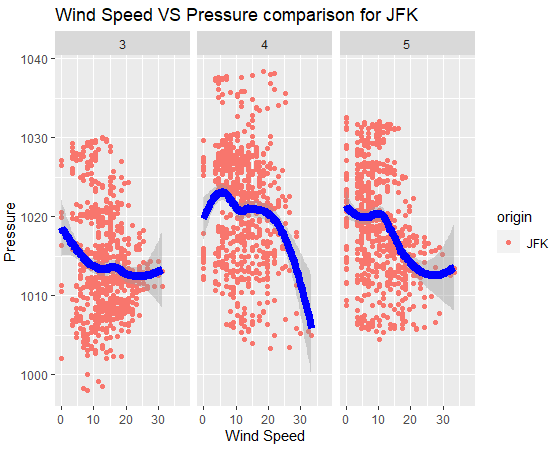


Fig 41: Wind speed vs pressure of spring season of JFK

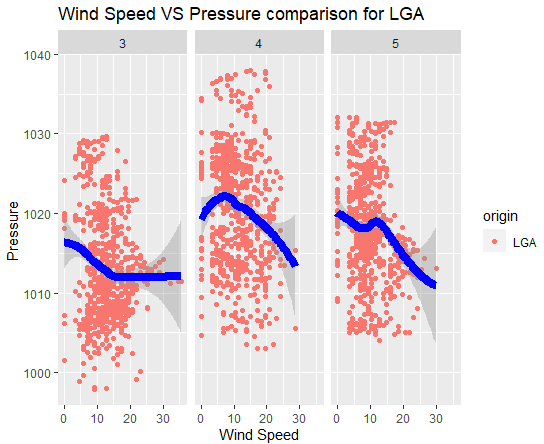


Fig 42: Wind speed vs pressure of spring season of LGA

From above figures 40 and 41, when there is minimum wind speed the pressure increases gradually. In JFK the wind speed and pressure are found maximum than LGA. In April of JFK, the lowest pressure is found where the wind speed the pressure gradually decreases in the comparison of March and May. In May of LGA, the pressure is found to be minimum than others. We can conclude that there is an inversely proportional relationship between the wind speed increases pressure.

## **Analysis 11:**

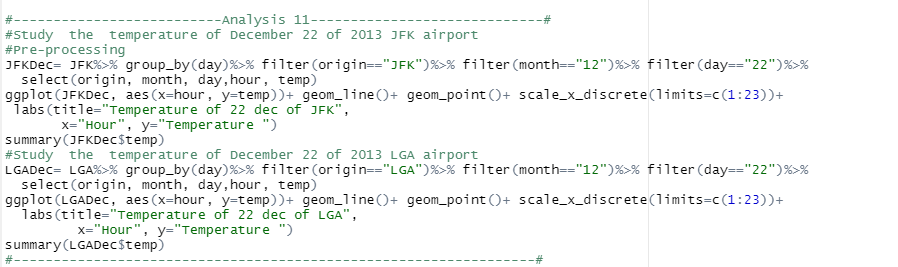
****

Fig 43: Comparing temperature of Dec 22 of JFK and LGA

* In this analysis the hourly rise and fall of temperature of the coldest day of year i.e., December 22 are compared between JFK and LGA.
* For pre-processing, the filter () function is used to filter the origin, month, and day then using the select() function, origin, month, day, hour, and temperature are selected and stored in a variable (JFKDec for JFK and LGADec for LGA). Then, using the ggplot() function it is filled in a graph in the form of a line using geom\_line() and the scale\_x\_discrete () function is used to provide the number for an hour from 1 to 23 as shown in figure 42.
* Similarly, the summary () function is used to find the overall summary i.e., min, max, median, number of null values, and so on.

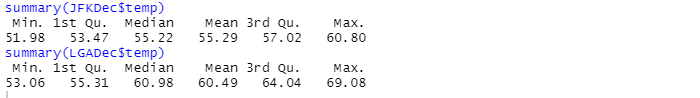


Fig 44: Summary of the temperature of Dec 22

|  |  |
| --- | --- |
|  |  |
|  |  |

Table 3: Table generated after pre-processing

## **Result**

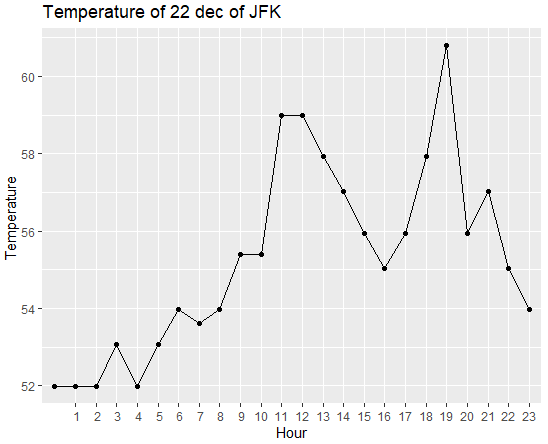
****

Fig 45: Line plot for an hourly temperature of Dec 22 in JFK

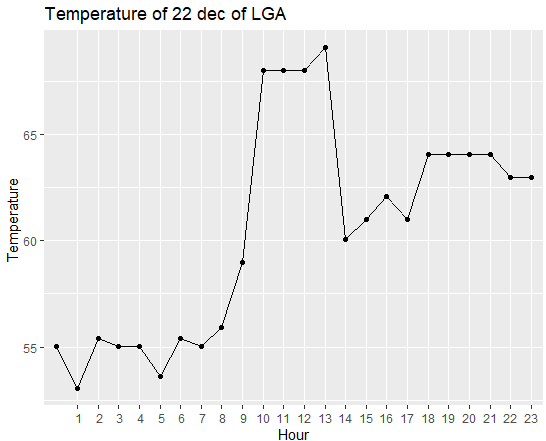


Fig 46: Line plot for an hourly temperature of Dec 22 in LGA

From the above figures 44 and 45, it is found that on the coldest day throughout the year i.e., December 22 the temperature in LGA airport is higher than JFK. At 13 o’clock, the temperature is found higher i.e., 69.08 degrees Fahrenheit in LGA. Similarly, in JFK at 19 o’clock, the temperature is found higher with the value of 60.80 degrees Fahrenheit. From the above diagram, we can also find the rise and fall of temperature on December 22.

## **Analysis 12:**

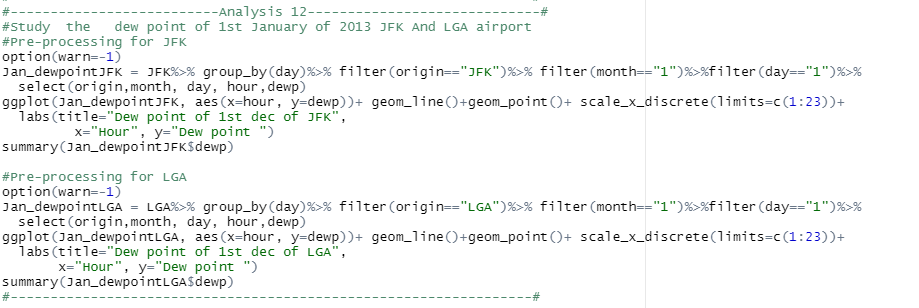
****

Fig 47: Comparing the dew point of 1st January of JFK and LGA

* In this analysis the hourly rise and fall of the dew point of the 1st January are compared between JFK and LGA.
* For pre-processing, the filter () function is used to filter the origin, month, day, hour, and dewp, and are selected using the select () function are selected and stored in a variable (Jan\_dewpointJFK for JFK and Jan\_dewpointLGA for LGA). Then, using the ggplot() function it is filled in a graph in the form of a line with point using geom\_line() and geom\_plot(), and the scale\_x\_discrete () function is used to provide the number for an hour from 1 to 23 as shown in figure 46. Option() function is used to remove the warning message by passing warn = -1 parameter inside it.
* Similarly, the summary () function is used to find the overall summary i.e., min, max, median, number of null values, and so on.

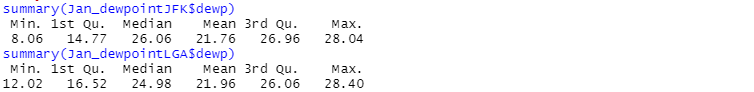
****

Fig 48: Summary of the dew point of 1st January

|  |  |
| --- | --- |
| **C:\Users\Info-chip\Pictures\Screenshots\Screenshot (307).png** | **C:\Users\Info-chip\Pictures\Screenshots\Screenshot (308).png** |

Table 4: Table generated after pre-processing

## **Result**

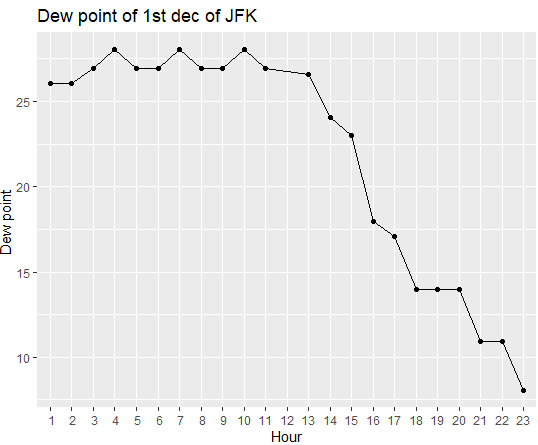
****

Fig 49: Line plot for an hourly dew point of JFK

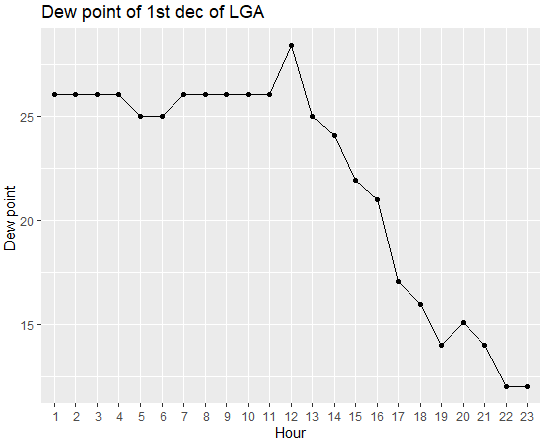


Fig 50: Line plot for an hourly dew point of LGA

From the above figures 48 and 49, it is found that the dew point on 1st January in LGA airport is higher than JFK. At 12 o’clock, the dew point is found higher i.e., 28.40 degrees Fahrenheit in LGA. Similarly, in JFK at 4, 7, 11 o’clock, the dew point is found constant and higher with the value of 20.04 degrees Fahrenheit. From the above diagram, we can also find the rise and fall of dew point on 1st January.

## **Analysis 13**

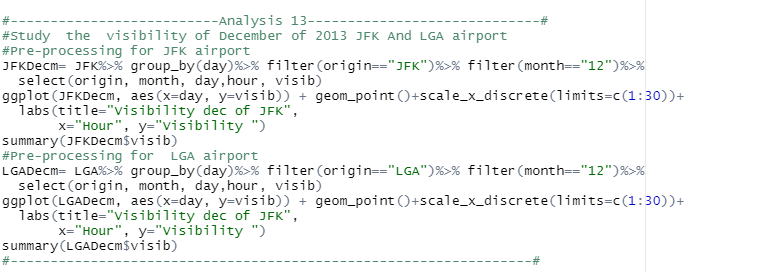
****

Fig 51: Comparing the visibility of December of both origins

* In this analysis the monthly rise and fall of the visibility of December are compared between JFK and LGA.
* For pre-processing, the filter () is used to filter the origin, and month then using the select function, origin, month, day, hour, and visibility are selected and stored in a variable (JFKDecm for JFK and LGADecm for LGA). Then, using the ggplot() function it is filled in a graph in the form of a scatterplot using geom\_point() and the scale\_x\_discrete () function is used to provide the number for an hour from 1 to 30 as shown in figure 50.
* Similarly, the summary () function is used to find the overall summary i.e., min, max, median, number of null values, and so on.

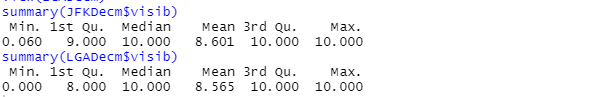
****

Fig 52: Summary of visibility of December

|  |  |
| --- | --- |
| **C:\Users\Info-chip\Pictures\Screenshots\Screenshot (311).png** | **C:\Users\Info-chip\Pictures\Screenshots\Screenshot (312).png** |

Table 5: Table generated after pre-processing

## **Result**

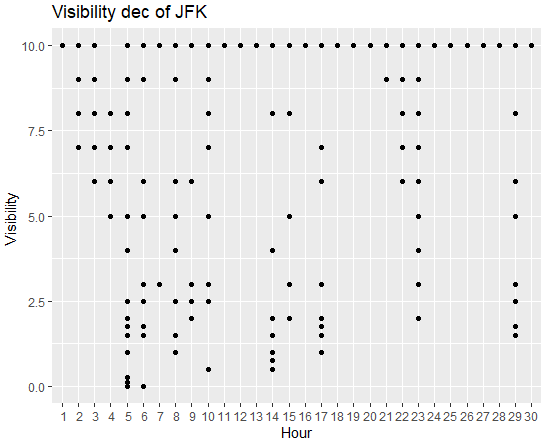
****

Fig 53: Scatterplot for Visibility of December in JFK

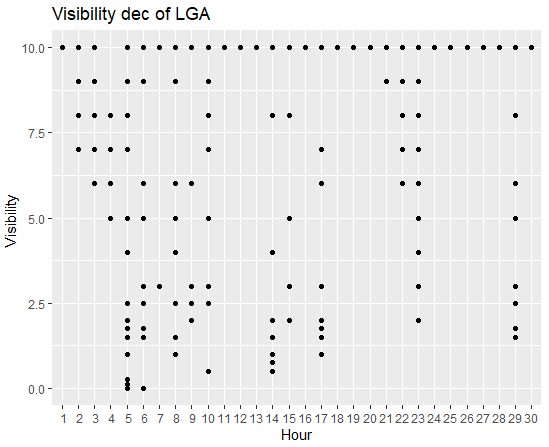
****

Fig 54: Scatterplot for Visibility of December in LGA

From the above figures 52 and 53, it is found that the maximum visibility in LGA and JFK are the same. The maximum value of visibility in both airports is found to be 10 miles which are found on all the days except the 4th of December. But the minimum visibility is found in LGA which is 0.00 miles and 0.060 miles in JFK on the 5th of December in both LGA and JFK.

## **Analysis 14:**

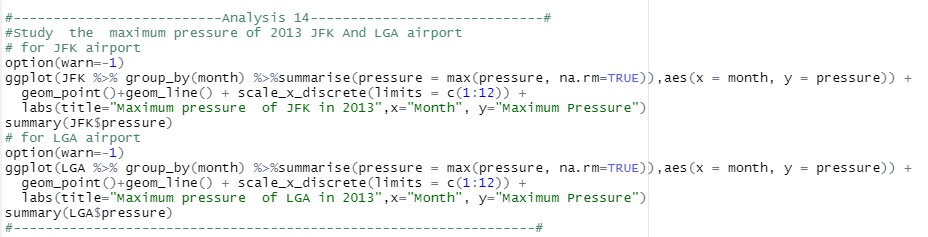
****

Fig 55: Comparing the maximum pressure of 2013 in both origins

* In this analysis the monthly maximum pressure of 2013 is compared between JFK and LGA.
* In the above analysis ggplot() function is used to plot the graph which is grouped by month using group() function, summarise() function is used to select the pressure column with monthly maximum value, geom\_point() and geom\_line() is used to fill the graph in the form of the line with points and scale\_x\_discrete is used to provide the numbers for the month along the x-axis. Similarly, the Option ()function is used to remove the warning message by passing warn = -1 parameter inside it.
* Similarly, the summary () function is used to find the overall summary i.e., min, max, median, number of null values, and so on.

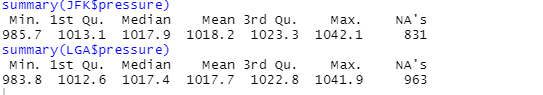
****

Fig 56: Summary of the pressure of JFK and LGA

## **Result**

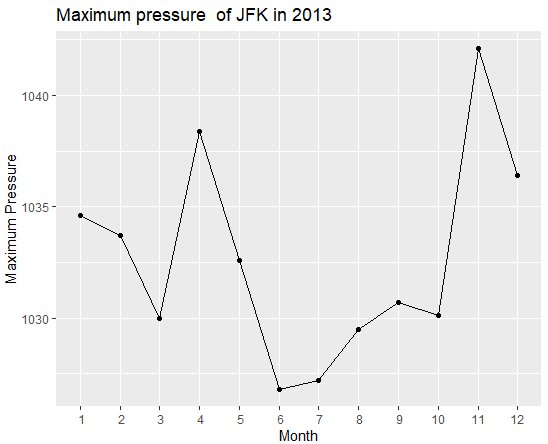


Fig 57: Line plot of monthly maximum pressure of JFK

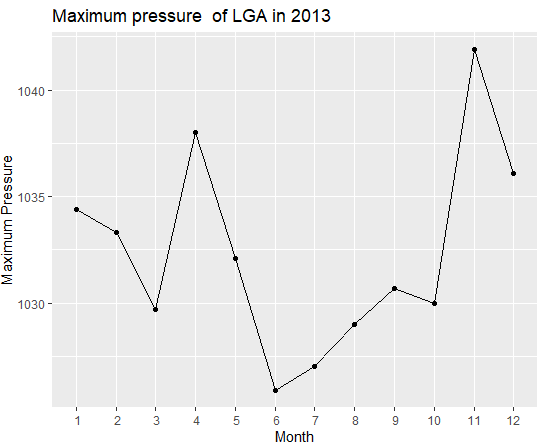
****

Fig 5: Line plot of monthly maximum pressure of LGA

From the above figures 56 and 57, it is found that the maximum pressure in LGA and JFK is found in November. The maximum value of pressure in JFK is found to be 1042.1milli bars which are higher than 1041.9 mill bars of LGA. But the least value of maximum pressure is found in June month of both airports (985.7 in JFK and 983.8 in LGA). Similarly, we can see the rise and fall of maximum pressure in both airports through the whole year of 2013.

# **Future Recommendation**

R is found as an important language for analyzing the data set provided in this assignment and R studio is found as an excellent tool for data analysis which involves visualization and manipulation. R is also found to be a flexible programming language that has no coding rules, it results in some complex pre-processing for extracting a particular data. R studio and the file generated in the R extension are found to be less concerned with memory allocation which occupies less space and helps me to work in R studio without any problem. But the unnecessary errors which don’t affect the results in the R command line reduce the user experience. Similarly, during the analysis process, it was found the late response for the visualization. These two problems were faced during the analysis process in my assignment which needs to upgrade. If these problems will be solved in the future, then R studio and R programming will provide a better experience to the users.

# **Conclusion**

At the end of this assignment, I was able to better comprehend the basics of R programming. It assisted me to understand the concept of data exploration, pre-processing, modification, and visualization which results in developing my analysis techniques and improved the outcome of the data analysis for decision making. Fourteen analyses in this project also assisted me to know the relationship between different weather factors which helps in the field of weather forecasting. Similarly, the techniques employed in R studio have been learned from this assignment which aids to develop a career in the field of data science.

# **References**

1. DataMentor. 2022. *R boxplot() to Create Box Plot (With Numerous Examples)*. [online] Available at: <https://www.datamentor.io/r-programming/box-plot/> [Accessed 7 March 2022].
2. Medium. 2022. *Understanding Boxplots*. [online] Available at: <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51> [Accessed 8 March 2022].
3. Rdocumentation.org. 2022. *hist function - RDocumentation*. [online] Available at: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/hist> [Accessed 8 March 2022].
4. Rdocumentation.org. 2022. *smooth function - RDocumentation*. [online] Available at: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/smooth> [Accessed 7 March 2022].
5. Statmethods.net. 2022. *Quick-R: Scatterplots*. [online] Available at: <https://www.statmethods.net/graphs/scatterplot.html> [Accessed 7 March 2022].
6. Vries, A., 2022. *Interpolation and smoothing functions in base R*. [online] Revolutions. Available at: <https://blog.revolutionanalytics.com/2015/09/interpolation-and-smoothing-functions-in-base-r.html> [Accessed 8 March 2022].