INDIVIDUAL ASSIGNMENT

TECHNOLOGY PARK MALAYSIA

CT127-3-2-PFDA

PROGRAMMING FOR DATA ANALYSIS

Name: Leong Khai Lok

TP Number:TP051968

Intake Code: UC2F2008IT(NC)

HAND OUT DATE: 6 OCTOBER 2020

HAND IN DATE: 23 NOVEMBER 2020

Lecturer Name: MARY TING

TABLE OF CONTENT

1.0 Introduction and assumptions

- In this research, I explored the application of techniques of data analytics and applied to the dataset “Hourly weather data” that was provided. The study is about particular data problems that is related to the dataset, and giving consideration to the properties of the problem domain, then testing the data analytics techniques to perform. In this assignment, I have learnt skills such as data import, visualization, exploration, manipulation concepts. It is to guide and help me progress on the assignment. It is also justified of the technique I used for each analysis on the code, as well as the analysis explanation. In general, the weather condition in airports are critical because of considering of the takeoff and landing of each airplane. Therefore, it is advisable for airports to meet example weather conditions such as:
  - High visibility
  - Dry or less water or snow on the landing ground
  - Weak wind or headwinds
  - No heavy rain or thunderstorms rains.

  In countless events, airport usually suffers from events such as thunderstorms, strong winds, low visibility, heavy rains, winter snows and many more. Therefore, it is recommended to understand the importance of law of weather changes because it will heavily effect the flights departure time or arrival of flights on airports. In this project, I am assigned and analyzed the hourly weather data that was given to me to understand the weather changes at LaGuardia airport (LGA) and John F. Kennedy International Airport(JFK) in the U.S. it will greatly help the airport employees to arrange arrival and departure flights in different season to avoid any damages from the weather events.

2.0 Preprocessing data

Install Packages

```
install.packages("ggplot2")
install.packages("crayon")
install.packages("readxl")
install.packages("missForest")
install.packages("VIM")
library(ggplot2)
library(crayon)
library(readxl)
library(dplyr)
library(missForest)
library(VIM)
```

- On the diagram above, it shows that my code on the Rstudio program to be installing and use 5 packages, which are called "ggplots2","crayon","readxl","missForest","VIM" by using install.packages("ggplot2"), install.packages("crayon"), install.packages("readxl"), install.packages("missForest"), install.packages("VIM"), then use library to activate all of the packages.

Reading excel file

```
setwd("C:/Users/crazy/Desktop/Asia Pacific University/Degree level 2/UC2F2008IT(NC)/Semester 1/PFDA/Individual")
```

- I will use the code "setwd("C:/Users/crazy/Desktop/Asia Pacific University/Degree level 2/UC2F2008IT(NC)/Semester 1/PFDA/Individual")." to set the work directory to the file location of the excel data file.

```
Dataassignment <- read.csv(file = "4.0 Hourly weather data.csv")
```

- As now, I will read the excel file by using the code "Dataassignment <- read.csv(file = "4.0 Hourly weather data.csv")".

```
> head(Dataassignment)
  origin year month day hour  temp  dewp humid wind_dir wind_speed wind_gust precip pressure visib        time_hour
1    JFK 2013     1   1    1 39.02 26.06 59.37      260   12.65858        NA      0   1012.6    10 01/01/2013 01:00
2    JFK 2013     1   1    2 39.02 26.06 59.37      270   11.50780        NA      0   1012.4    10 01/01/2013 02:00
3    JFK 2013     1   1    3 39.92 26.96 59.50      260   14.96014        NA      0   1012.7    10 01/01/2013 03:00
4    JFK 2013     1   1    4 39.92 28.04 62.21      250   17.26170        NA      0   1012.6    10 01/01/2013 04:00
5    JFK 2013     1   1    5 39.02 26.96 61.63      260   14.96014        NA      0   1012.1    10 01/01/2013 05:00
6    JFK 2013     1   1    6 37.94 26.96 64.29      260   13.80936        NA      0   1012.6    10 01/01/2013 06:00
```

- By using the code "head(Dataassignment)", I will obtain the first 6 data of inside of the excel file.

2.1 Filtering origin (data manipulation)

```
#Data 1 filtering JFK data only
JFK = filter(Dataassignment, origin =="JFK")
#Data 2 filtering LGA data only
LGA = filter(Dataassignment, origin =="LGA")
```

- On this code, I separated both of the airport origin, which are JFK and LGA from the entire csv so I can select my data easily on most of my analysis.

3.0 Analysis

3.1 Analysis 1: Comparing daily Temperature recorded of JFK and LGA in 2013

```
options(warn=-1) #SUPRESS Warning

JFK %>% ggplot(aes(x= day,y = temp))+ #ggploting JFK data

  geom_jitter(aes(color= temp, na.rm = TRUE))+ #setting geom_jitter in color for temp and remove NA values

  facet_wrap(~month)+ #seperate graph by month

  labs(title="Temperature for each day

       in 2013 in JFK", x="Months"

       , y="Temperature in Fahrenheit") #setting title, x axis and y axis names for graph.


LGA %>% ggplot(aes(x= day,y =temp, na.rm = TRUE))+ #ggploting LGA data and removing NA value

  geom_jitter(aes(color= temp))+facet_wrap(~month)+ #Setting as geom_jitter in color for temp and seperate graph by month

  labs(title="Temperature for each day

       in 2013 in LGA", x="Months"

       , y="Temperature in Fahrenheit") #setting title, x axis and y axis names for graph .
```

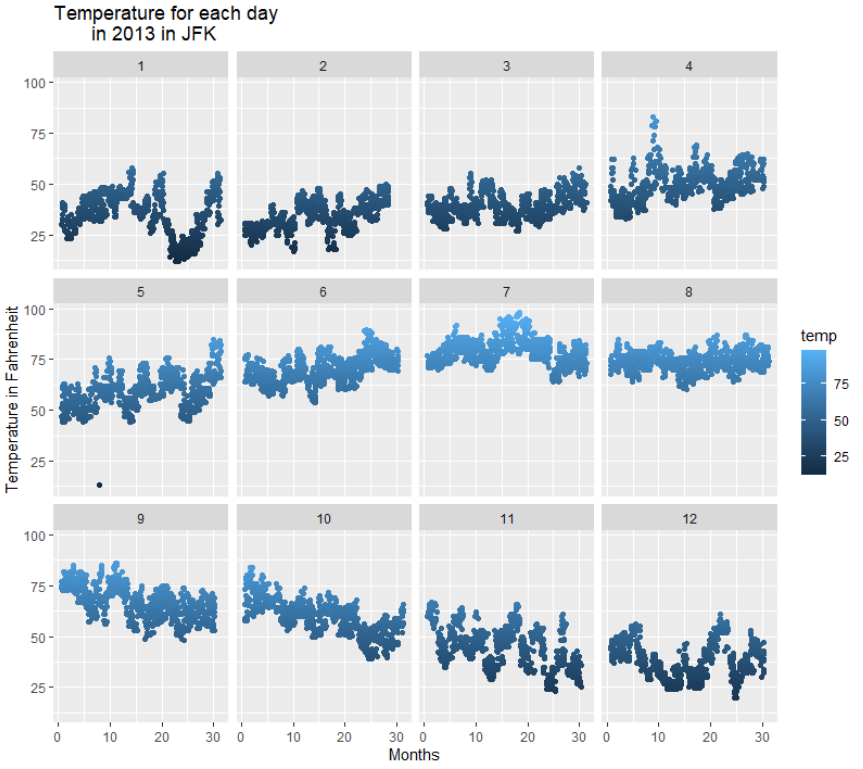
- The analysis is about daily temperature of JFK and LGA in monthly graphs and separated by the two origin in 2013. The analysis is conducted to review the overall of the temperature monthly and it is based on the hourly weather data on the csv file with the code at preprocessing data.
- On this code, I have suppressed the warning with options(warn=-1) for both JFK and LGA origin function that I have separated on preprocessing data. Then, I have displayed daily temperature of JFK and LGA in 2013 with geom_point type of graph and separated them by month.

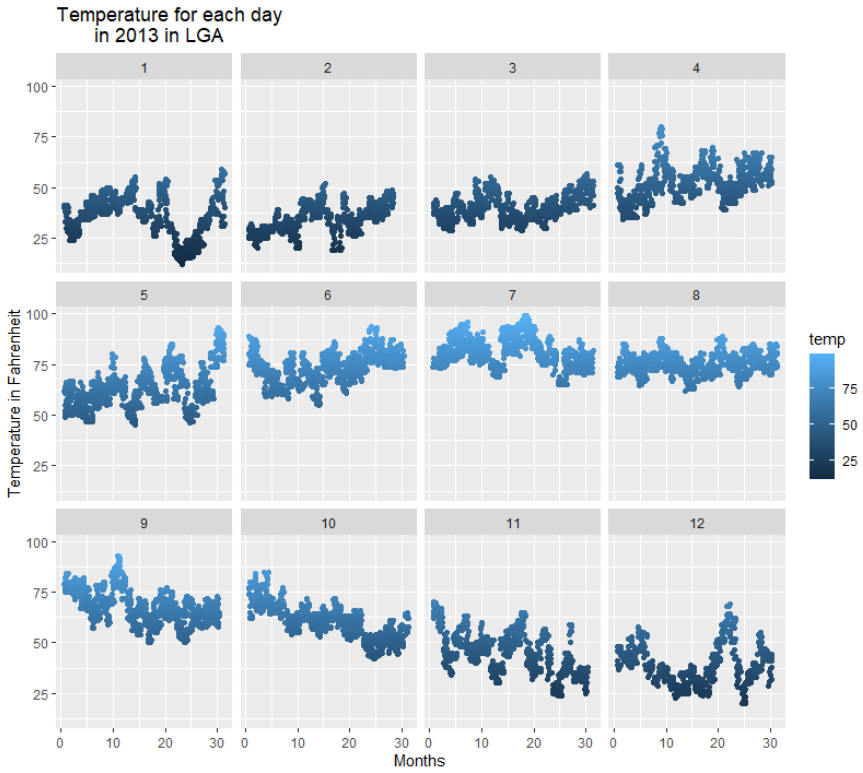
```
> summary(JFK$temp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.02   39.92   53.96   54.47   69.08   98.06
```

- On this summary of the analysis, the minimum of the temperature for JFK is 12.02°F, with the median of 53.96°F and the maximum temperature of 98.06°F.

```
> summary(LGA$temp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.02   39.92   55.94   55.76   71.06   98.96
```

- On this summary of the analysis, the minimum of the temperature for LGA is 12.02°F, with the median of 55.94°F and the maximum temperature of 98.96°F.





- Based on the diagram, the temperature for both origin starts on low temperature on the first four months and slowly increase temperature and drops again at the end of the year. It is advisable for the airport to give discounts or having promotions during the cold temperature under 30°F, as there will be many people prefer to travel during normal temperature rates rather than cold temperature months.

3.2 Analysis 2: Comparing daily Wind Direction recorded of JFK and LGA in 2013(EXTRA FEATURE 1: GEOM_SMOOTH)

```
options(warn =-1) #Suppress warning

JFK %>% ggplot(aes(wind_dir))+ #ggploting JFK data

  geom_bar(aes(color= wind_dir, na.rm = TRUE))+facet_wrap(~month)+ #setting as geom_bar in color for wind_direction, remove NA values and seperate graph by month

  labs(title="Wind direction for each day

       in 2013 of JFK", x="Wind_Direction in degree") #setting title and x axis name for graph.


LGA %>% ggplot(aes(wind_dir))+ #ggploting LGA data

  geom_bar(aes(color= wind_dir, na.rm = TRUE))+facet_wrap(~month)+ #setting geom_bar in color for wind_direction, remove NA values and seperate graph by months

  labs(title="Wind direction for each day

       in 2013 of LGA", x="Wind_Direction in degree") #setting title and x axis name for graph


ggplot(Dataassignment, aes(x = month, y = wind_dir, na.rm = TRUE, suppressWarnings(expr))) + #ggploting wind_direction from csv folder with Dataassignment, remove NA value and suppress warnings

  geom_smooth(size = 1, se = FALSE, aes(color = origin)) + #setting geom_smooth in color and seperate points by origin (Extra feature 1, Geom_Smooth)

  labs(x = "Month", y = "Wind Direction", title = "Wind direction in both JFK and LGA") +

  scale_x_discrete(limits = c(1:12)) #setting title, x and y axis name and scale limit for x of 1 to 12 on graph
```

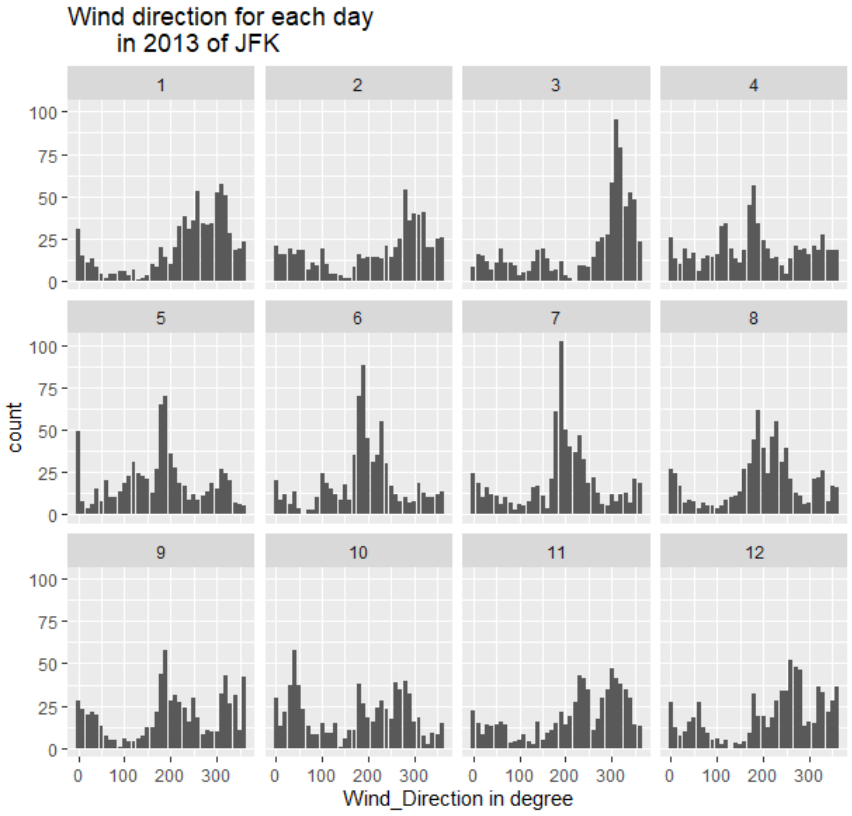
- This analysis is about finding of wind direction on both origin JFK and LGA in 2013 and compare the differences daily in a separated monthly graph. The analysis is conducted to research by using windsocks on the airport to indicate the direction and strength of the wind.
- On the code, I have suppressed the warnings with options(warn=-1). Then, for both JFK and LGA origin, I have separated them by functions to do some manipulation with the help from my preprocessing data to have display the data daily wind_direction of JFK and LGA in 2013 easily with geom_bar and separated them by month with facet_wrap and scaled the x axis limit of 1 to 12. By using extra feature which is geom_smooth type of graph, I make the visualization more smooth and display the data of the finding of analysis.

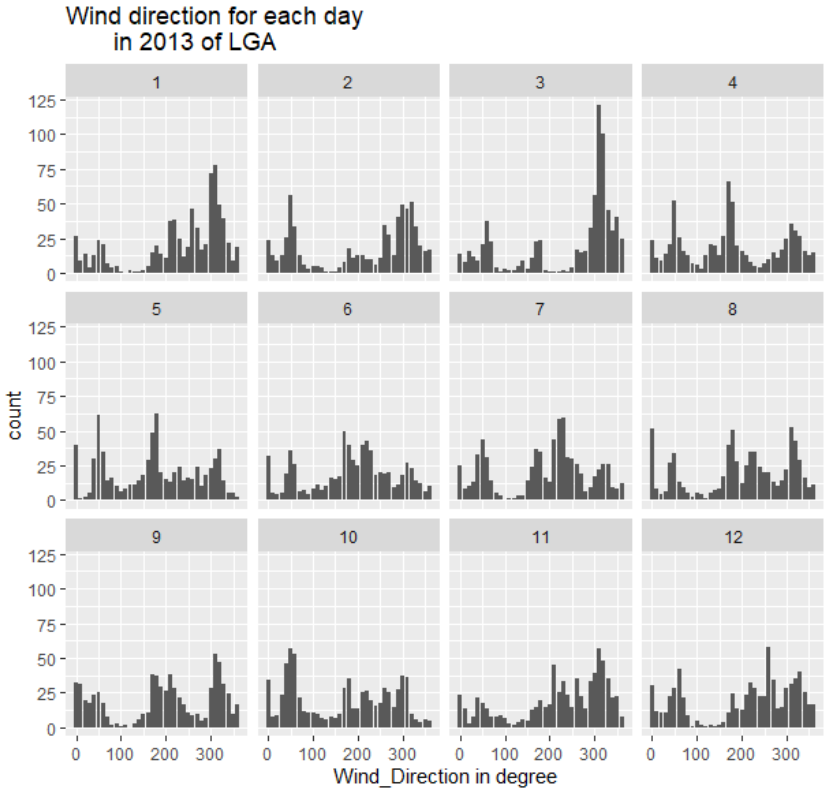
```
> summary(JFK$wind_dir)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    0.0   140.0   220.0   204.2   290.0   360.0      51
```

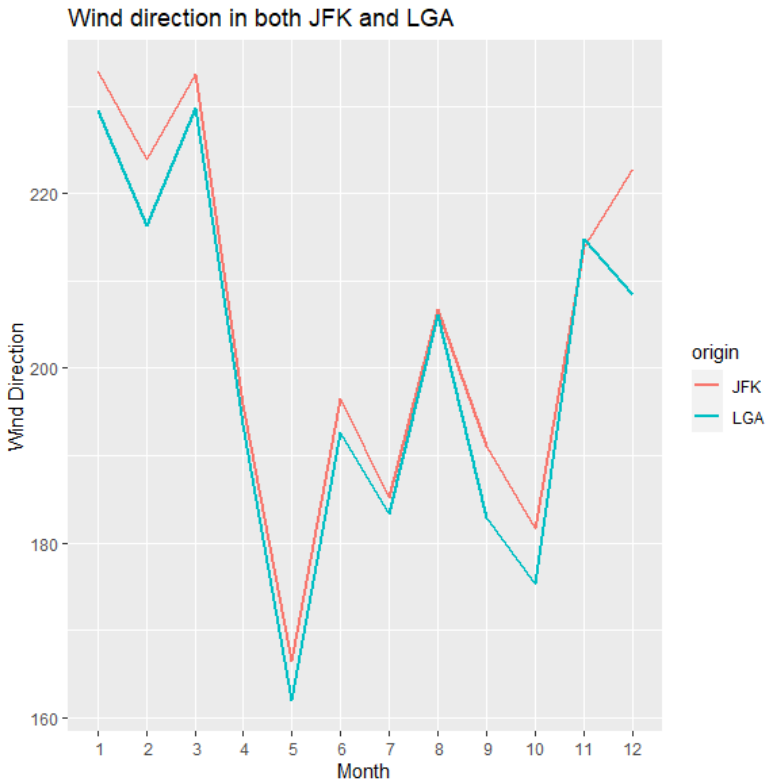
- Summary of wind direction on JFK. Minimum is 0.0, maximum is 360.0, with 51 NA value data.

```
> summary(LGA$wind_dir)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    0.0   100.0   220.0   199.5   300.0   360.0     153
```

- Summary of wind direction on LGA. Minimum is 0.0, maximum is 360, with 153 NA value data.







- Based on the diagram, the result shows the wind direction for both JFK and LGA airport mostly in similar statistic in bar graph but the wind direction in JFK airport are slightly higher in wind direction compared to LGA monthly in the geom_smooth graph.

3.3 Analysis 3: Comparing daily Wind Speed recorded of JFK and LGA in 2013

```
ggplot(JFK, aes(x=day,y=wind_speed, color= wind_speed, na.rm = TRUE)) + #ggploting JFK data for wind speed, set color for wind speed and removing NA values
  geom_point() + facet_wrap(~month) + #setting geom_point and seperating data in months
  labs(title="Daily wind speed of JFK", x="Day", y="Wind speed in KM/H") #setting title, x and y axis name for graph.
```

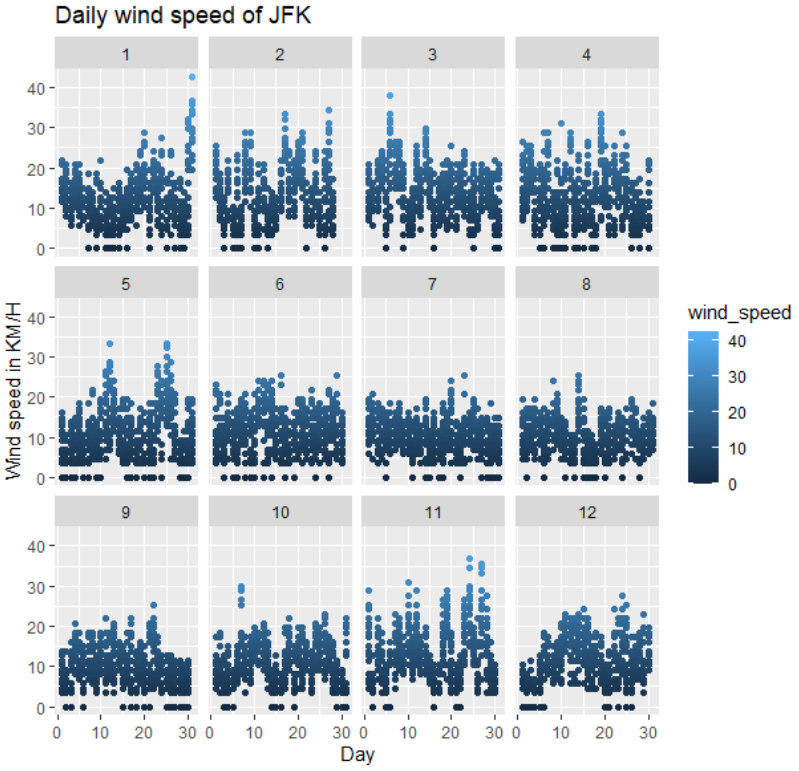
```
ggplot(LGA, aes(x=day,y=wind_speed, color= wind_speed, na.rm = TRUE)) + #ggploting LGA data for wind speed, set color for wind speed and removing NA values
  geom_point() + facet_wrap(~month) + #setting geom_point and seperate graph by months
  labs(title="Daily wind speed of LGA", x="Day", y="Wind speed in KM/H") #set title, x and y axis name for graph.
```

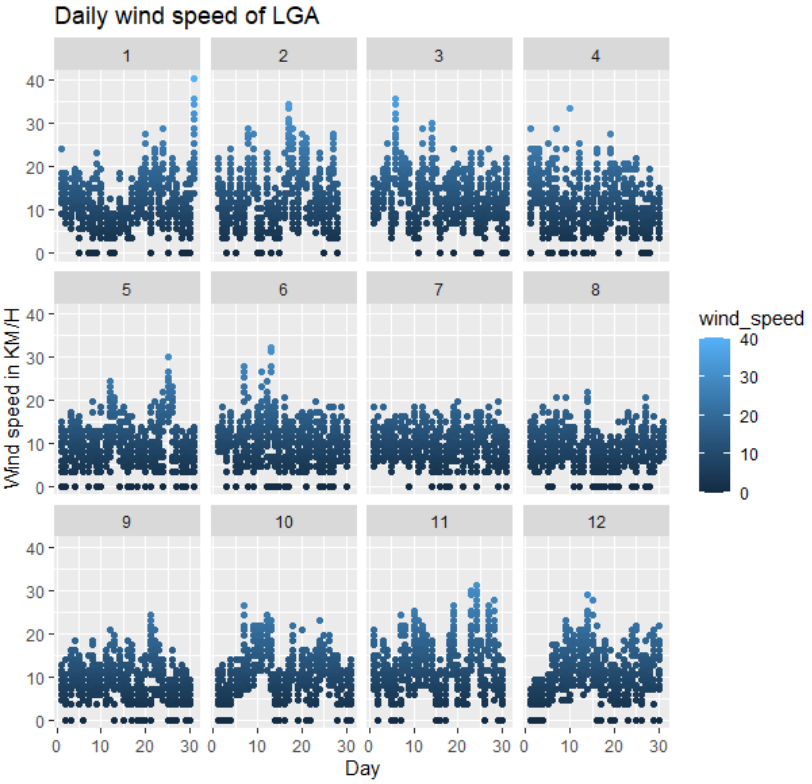
- This analysis finding is to compare daily wind speed of JFK and LGA airport in monthly graph of 2013. The analysis is conducted to measure wind speed to increase the safety for both airports by ensuring safety taking off from the airport and landing of the plane with no issue.
- On the code, I have separated the origin with the help from my preprocessing data and displayed daily wind_speed of JFK and LGA in 2013 with geom_point type of ggplot graph and separated them by month with facet_wrap.

```
> summary(JFK$wind_speed)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  0.000   6.905  10.357  11.468  14.960  42.579       3
```

- Summary for wind_speed on JFK data

```
> summary(LGA$wind_speed)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   6.905  10.357  10.623  13.809  40.277
```



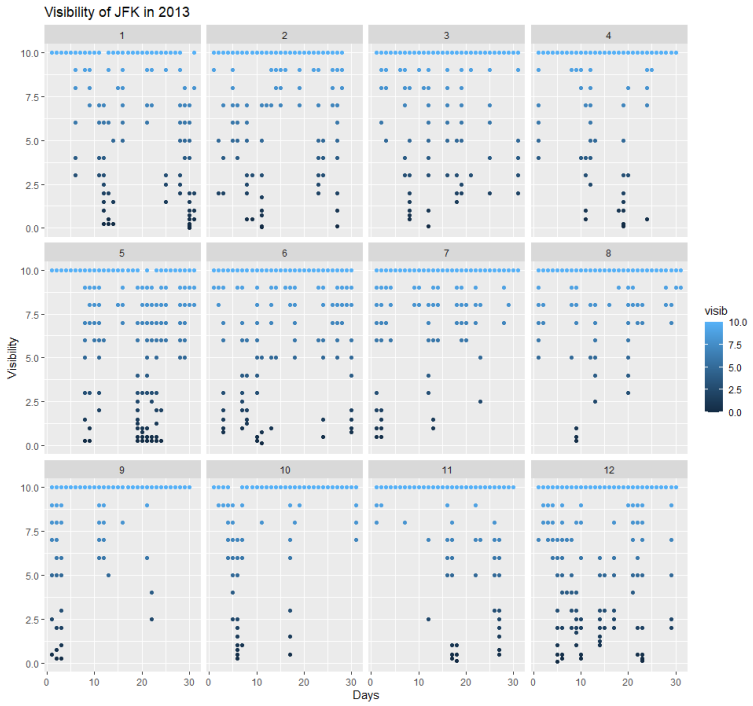


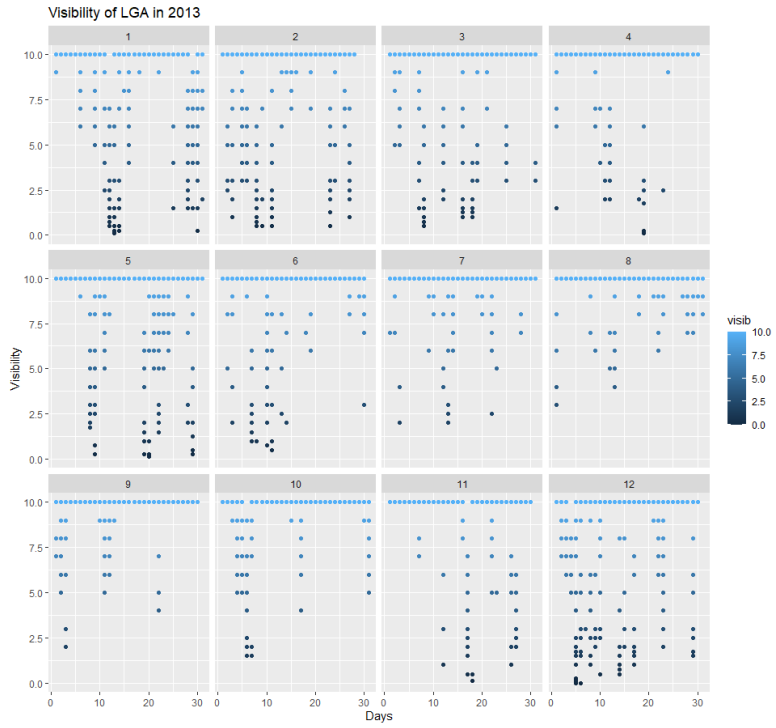
- Based on the diagram, the finding of wind speed for JFK and LGA airport in 2013 are mostly similar. However, there are few days of risk as the wind speed has reached above 30 kilometer per hour.

3.4 Analysis 4: Comparing daily visibility recorded of JFK and LGA in 2013

```
ggplot(JFK, aes(day,visib, color= visib, na.rm = TRUE)) + #ggploting JFK data for visibility, set color to visib and removing NA values
  geom_point() + facet_wrap(~month)+ #setting geom_point and seperating by month
  labs(title="Visibility of JFK in 2013",x="Days", y="Visibility") #setting title,x and y axis name to graph

ggplot(LGA, aes(day,visib, color= visib, na.rm = TRUE)) +  #ggploting LGA data for visibility, set color to visib and removing NA values
  geom_point() + facet_wrap(~month)+ #setting geom_point and seperate graph by month
  labs(title="Visibility of LGA in 2013",x="Days", y="Visibility") #setting title, x and y axis name to graph.
```

- This analysis finding is about finding and comparing the daily visibility of JFK and LGA airport in 2013. The analysis is conducted due to the safety on the plane on each flights. Visibility is important when it comes to the airplane close to the ground as it needs to battle against low visibility such as fogs, strong winds, big storms, frosts and many more. For instance, airplanes when during landing or take off.
- On the code, I have separated the JFK and LGA by using a function on the preprocessing to select data from JFK and LGA filtered data. By using geom_point type of graph, I displayed the graph for visibility of JFK and LGA, then separated it by monthly in 2013.



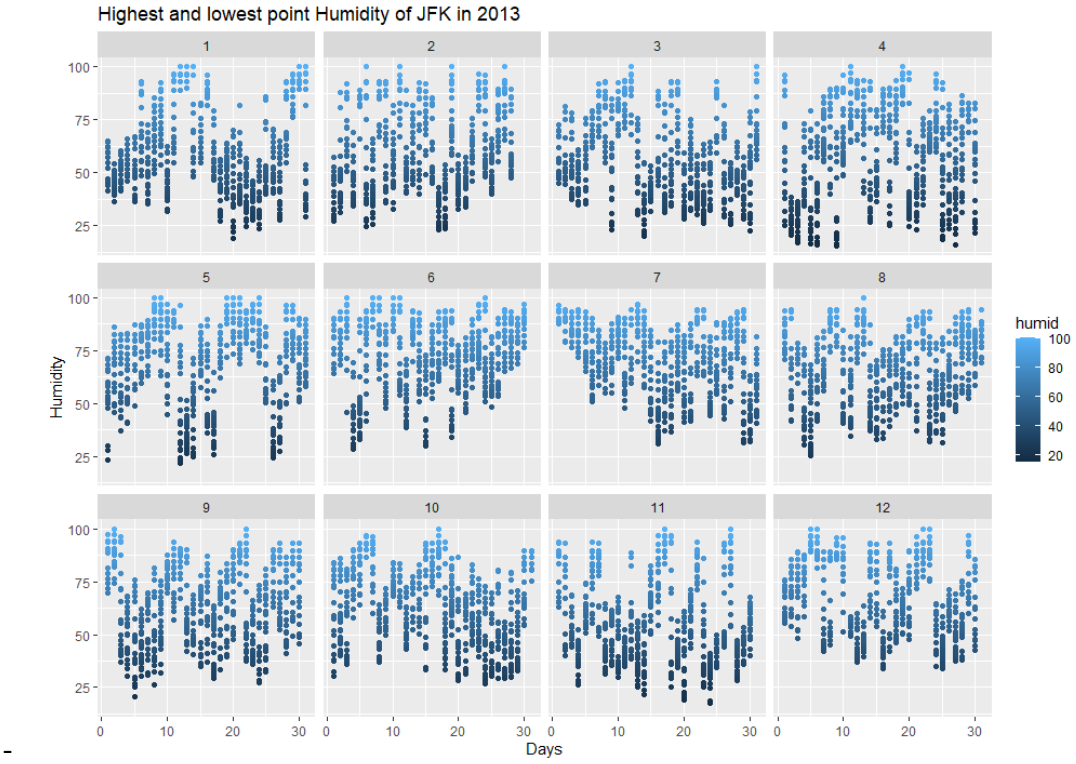


- Based on findings of the graph, the visibility on JFK and LGA is mostly high every month but there are high numerous of low visibility between the month of January, February, March, December.

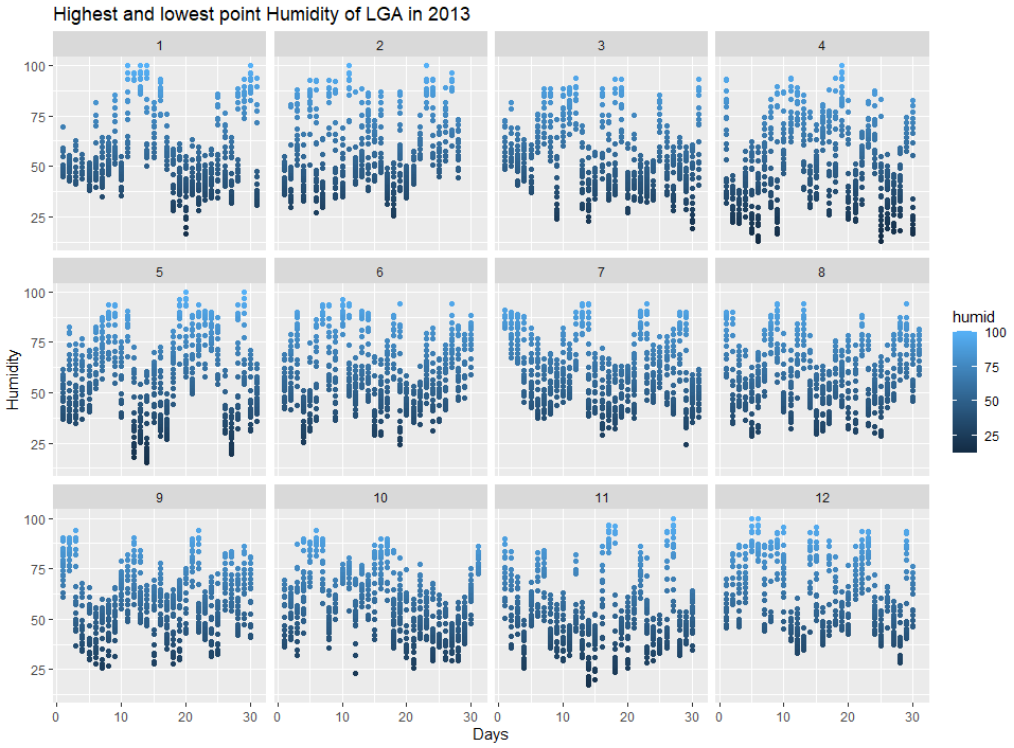
3.5 Analysis 5: Comparing daily humidity of JFK and LGA in 2013

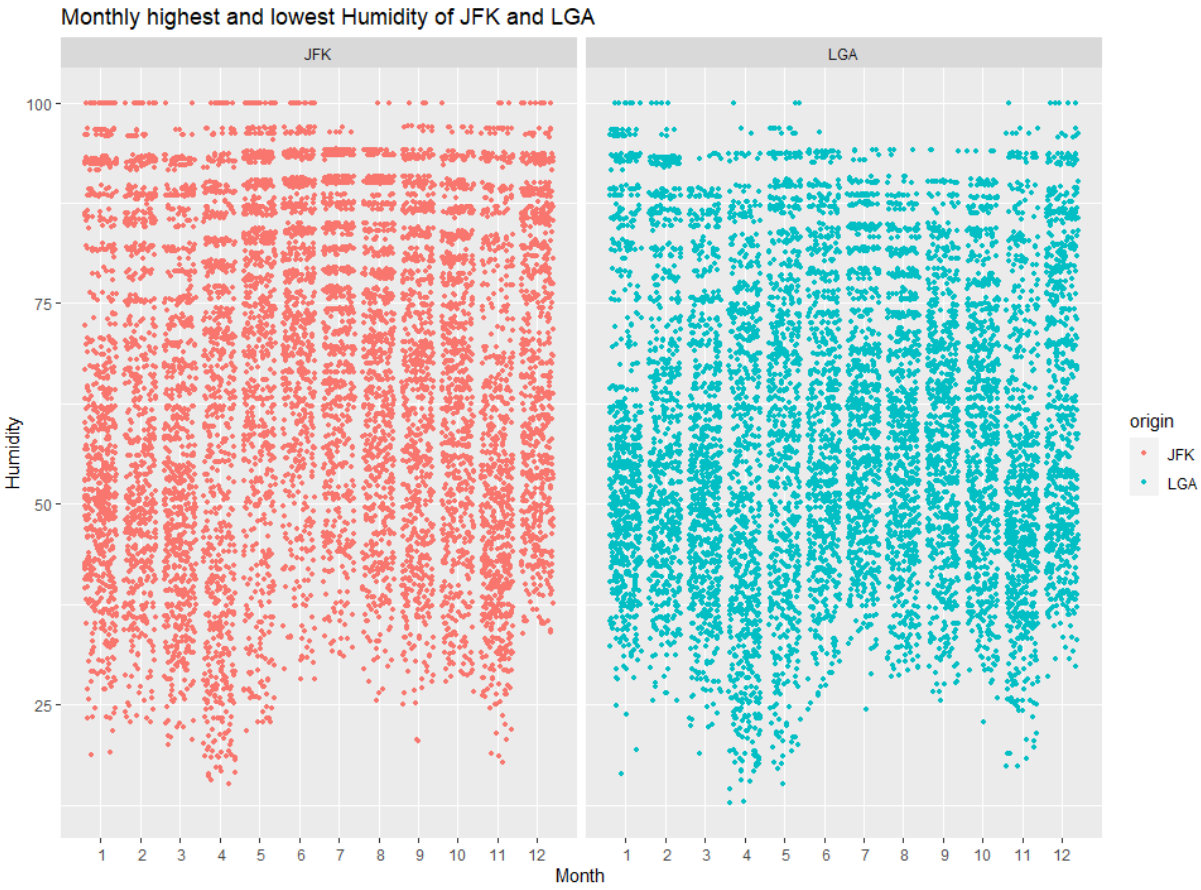
```
ggplot(JFK, aes(day,humid, color= humid, na.rm = TRUE)) + #ggploting JFK data for humidity and removing NA values

  geom_point() + facet_wrap(~month)+ #setting geom_point and seperating by month

  labs(title="Highest and lowest point Humidity of JFK in 2013",x="Days", y="Humidity") #set title, x and y axis name to graph


ggplot(LGA, aes(day,humid, color= humid, na.rm = TRUE)) + #ggploting LGA data for humidity and removing NA values

  geom_point() + facet_wrap(~month)+ #setting geom_point and seperating by month

  labs(title="Highest and lowest point Humidity of LGA in 2013",x="Days", y="Humidity") #set title, x and y axis name to graph


ggplot(Dataassignment, aes(x= month, y=humid, na.rm = TRUE)) + #ggploting the csv file and removing NA data

  geom_jitter(method = 'gam', size=1, aes(color=origin)) + #setting geom_hitter with using method "gam", and set and seperate by color to origin

  labs(title = "Monthly highest and lowest Humidity of JFK and LGA", x ="Month", y ="Humidity") + #setting title,x and y axis name to graph

  scale_x_discrete(limits = c(1:12)) + facet_wrap(~origin) #set scale x axis limit to 1 to 12 and seperate graph by origin
```

- The analysis is about finding and comparing humidity of both origin, JFK and LGA in 2013. This analysis of high heat and humidity is conducted that combat against unsafe conditions on the airport that might cause fatigue and delays reaction times due to improper hydration of all of the airport workers.
- On the code, I separated JFK and LGA filtered data and plotted a graph with geom_point type of graph and separating both filtered data graphs by month. To make the data more accurate, I removed NA values for both filtered data.


Highest and lowest point Humidity of JFK in 2013
1
2
3
4
5
6
7
8
9
10
11
12
Humidity
Days
humid
100
80
60
40
20

-


Highest and lowest point Humidity of LGA in 2013
1
2
3
4
5
6
7
8
9
10
11
12
Humidity
Days
humid
100
75
50
25

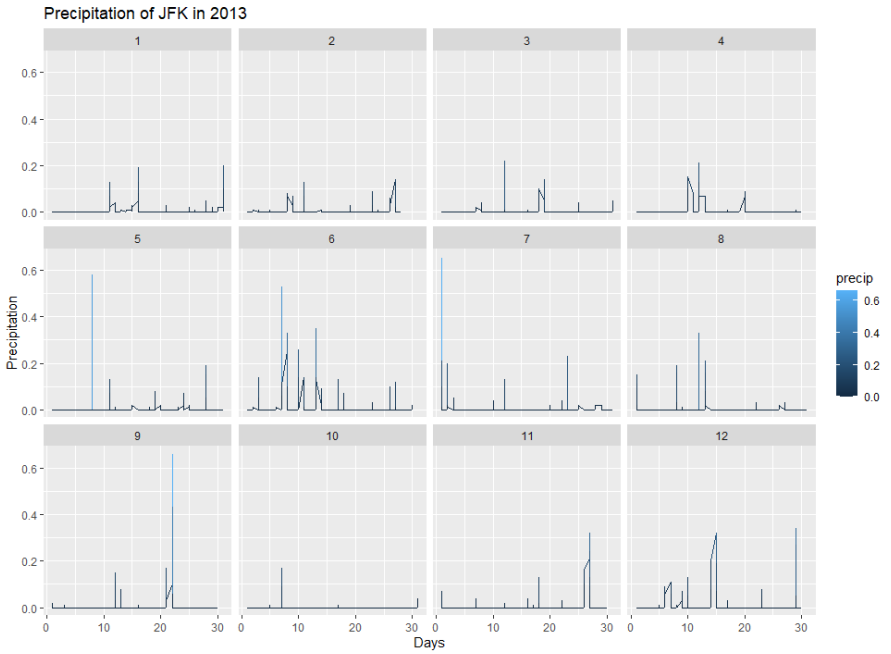


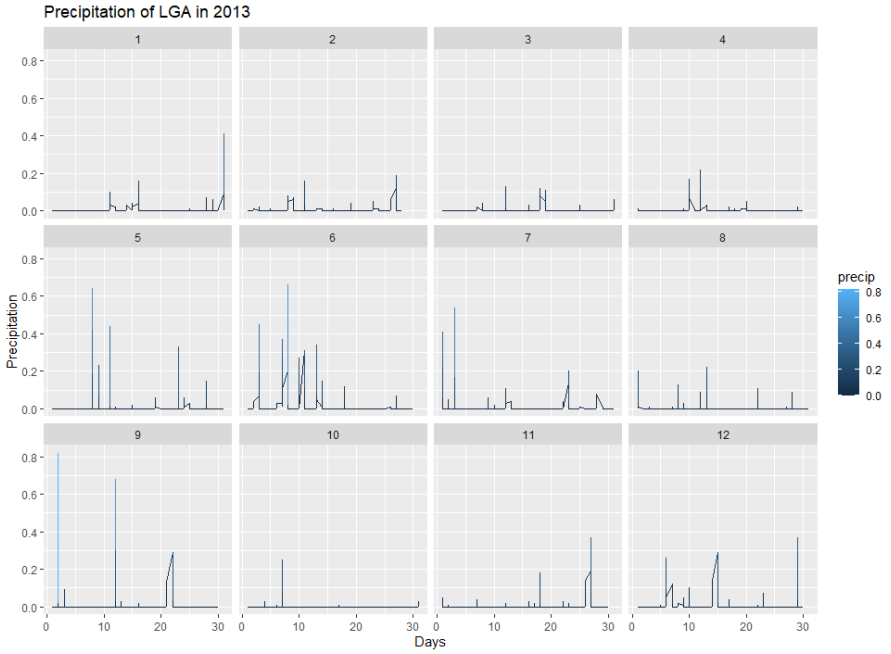
- According to the graphs, the analysis finding of JFK and LGA shows that, there are some days of both of the airport contains data for mostly days with normal humidity.

3.6 Analysis 6: Comparing daily precipitation of JFK and LGA in 2013

```
ggplot(JFK, aes(x=day,y=precip, color= precip, na.rm = TRUE)) + #ggploting JFK data, set color to precip and removing NA value

  geom_line() + facet_wrap(~month)+ #setting geom_point and seperating by month

  labs(title="Precipitation of JFK in 2013",x="Days", y="Precipitation") #setting title, x and y axis name to graph


ggplot(LGA, aes(x=day,y=precip, color= precip, na.rm = TRUE)) + #ggploting LGA data, set color to precip and removing NA value

  geom_line() + facet_wrap(~month)+ #setting geom_point and seperating by month

  labs(title="Precipitation of LGA in 2013",x="Days", y="Precipitation") #setting title, x and y axis name to graph
```

- The analysis is about the finding and comparing the daily precipitation of JFK and LGA in 2013. This analysis is conducted to reduce impact on airport performances that cause delays on flights due to weathers events.
- On the code, I separated JFK and LGA filtered data and plotted a graph with geom_line type of graph and separating both filtered data graphs by month. To make the data more accurate, I removed NA values for both filtered data.





- Based on the graph, the analysis finding shows that for both airport, most of the months sparkles less to no precipitation. However, the precipitation sparkled the most during the 6th month in 2013.

3.7 Analysis 7: Monthly Mean temperature of JFK and LGA in 2013

```
ggplot(JFK %>% group_by(month) %>% #ggploting JFK data and grouping data by month
         summarise(temp = mean(temp)), #summarize mean temperature monthly
       aes(x = month, y = temp)) + geom_point() + #set geom_point
  geom_line() + scale_x_discrete(limits = c(1:12)) + #add geom_line and scale x axis limit to 1 to 12 for graph.
  labs(title="temperature mean of JFK in 2013",x="Month", y="Mean") #setting title,x and y axis to graph
```

```
ggplot(LGA %>% group_by(month) %>% #ggploting JFK data and grouping data by month
         summarise(temp = mean(temp)), #summarize mean temperature monthly
       aes(x = month, y = temp)) + geom_point() + #set geom_point
  geom_line() + scale_x_discrete(limits = c(1:12)) + #add geom_line and scale x axis limit to 1 to 12 for graph.
  labs(title="temperature mean of LGA in 2013",x="Month", y="Mean") #setting title,x and y axis to graph
```

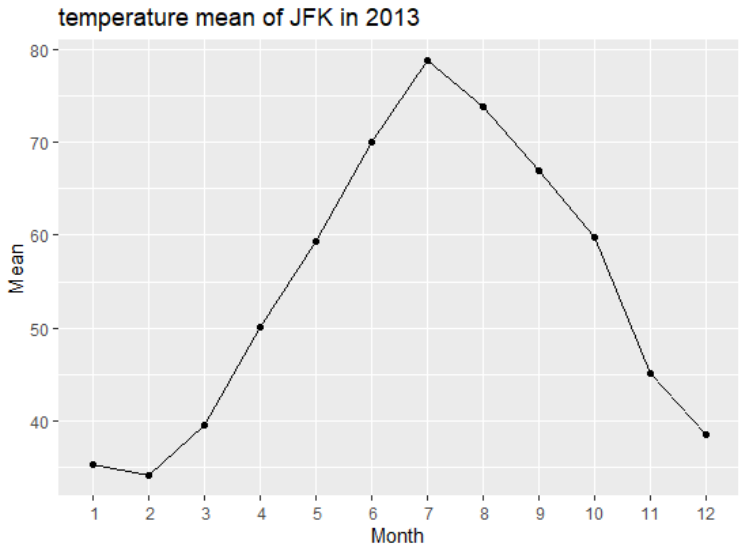
- This analysis is about the mean temperature of JFK and LGA airports on 2013. The analysis is conducted to find the mean point for each month of the temperature on both airports in 2013.
- On the code, I have seperated JFK and LGA filtered data and plotted different graph for both with airports, and seperated them by origin. Then, I summarized the temperature mean monthly and displayed it in geom_point and geom_line type of graph with scaling x axis for an year on 2013.

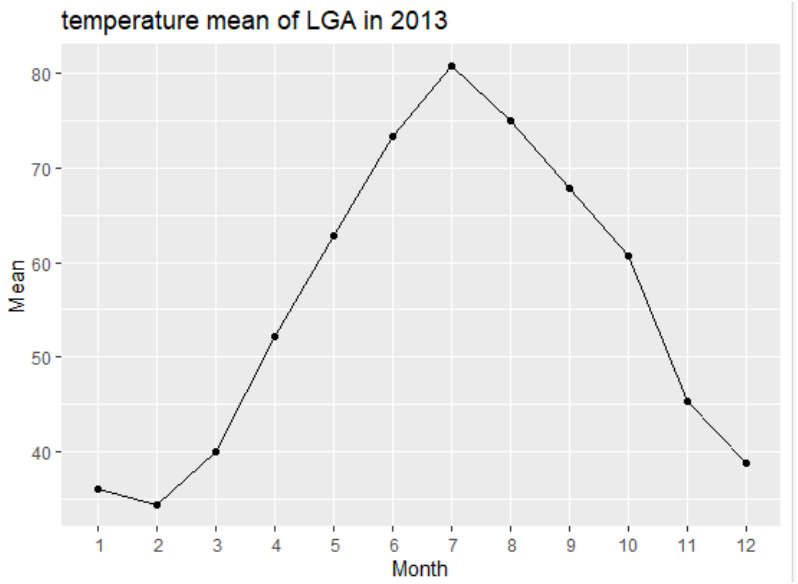
```
> summary(JFK$temp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.02   39.92   53.96   54.47   69.08   98.06
```

- Based on this finding, The overall mean temperature for 2013 in JFK is 54.47.

```
> summary(LGA$temp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.02   39.92   55.94   55.76   71.06   98.96
```

- Based on this finding, the overall mean temperature for 2013 in LGA is 55.76



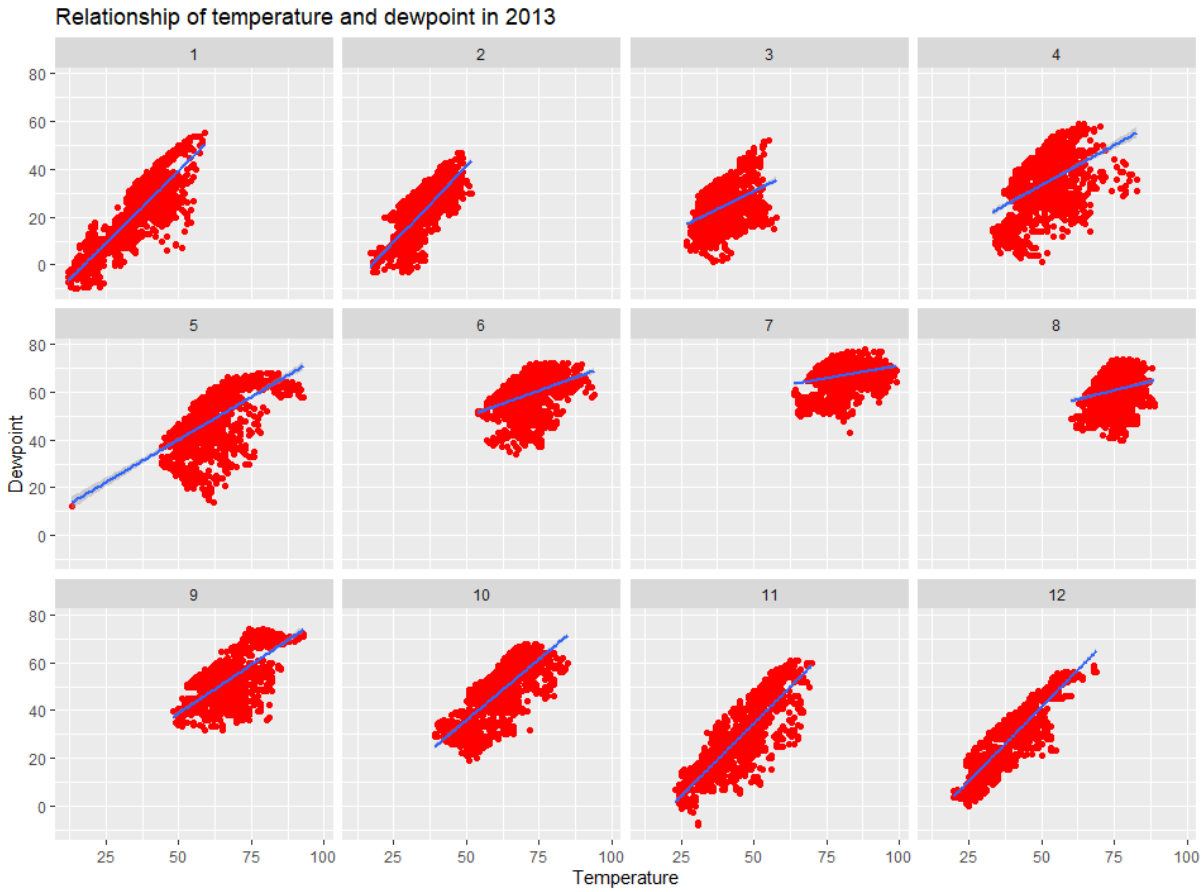


- Based on the analysis finding, the overall mean temperature from LGA airport is higher than JFK airport as LGA mean is recorded at 55.76 while mean of JFK airport is recorded at 54.47.

3.8 Analysis 8: Relationship between temp and dewp of JFK and LGA in 2013

```
options(warn=-1) #suppressing warning messages

ggplot(Dataassignment, aes(x=temp,y=dewp, na.rm = TRUE)) + #ggploting csv folder and removing NA values

  geom_point(color ='red') + geom_smooth(method = lm) + #setting geom_point with color red and add geom_smooth with method =lm

  facet_wrap(~month) + #seperating graph by month

  labs(title="Relationship of temperature and dewpoint in 2013",x="Temperature", y=”Dewpoint") #setting title,x and y axis to graph
```

- This analysis is about the co-variation of between temperature and dewpoint in overall data. The analysis is conducted because to show the co-variation nrelationship changes among the different month in 2013.
- On the code, firstly, I suppressed the warning messages, and start ggploting the csv folder and removing the NA values for accurate data plots. I set the plot with geom_point and color the graph with red and add smooth smooth with method = lm and separate the graphs it by months. Finally, set title, x and y axis to the graph.



- Based on the analysis, it shows that the dewpoint has a strong relationship with temperature in 2013. However, the co-variation of temperature and dewpoint varies among different month. In overall, the dot plots gradually decrease from January to June, but the line increases from July to December. It is resulted in indicating of the dewpoint increases more faster than the temperature in the seasons with lower temperature.

3.9 Analysis 9: Relationship between wind gust and wind speed of JFK and LGA in 2013

```
options(warn=-1) #suppressing warning messages

ggplot(Dataassignment, aes(x=wind_gust,y=wind_speed, na.rm = TRUE)) + #ggploting csv folder and removing NA values

  geom_point(color ='red') + geom_smooth(method = lm) + #setting geom_point with color red and add geom_smooth with method =lm

  facet_wrap(~month) + #seperating graph by month

  labs(title="Relationship of wind_gust and wind_speed in 2013",x="Temperature", y="Dewpoint") #setting title,x and y axis to graph
```

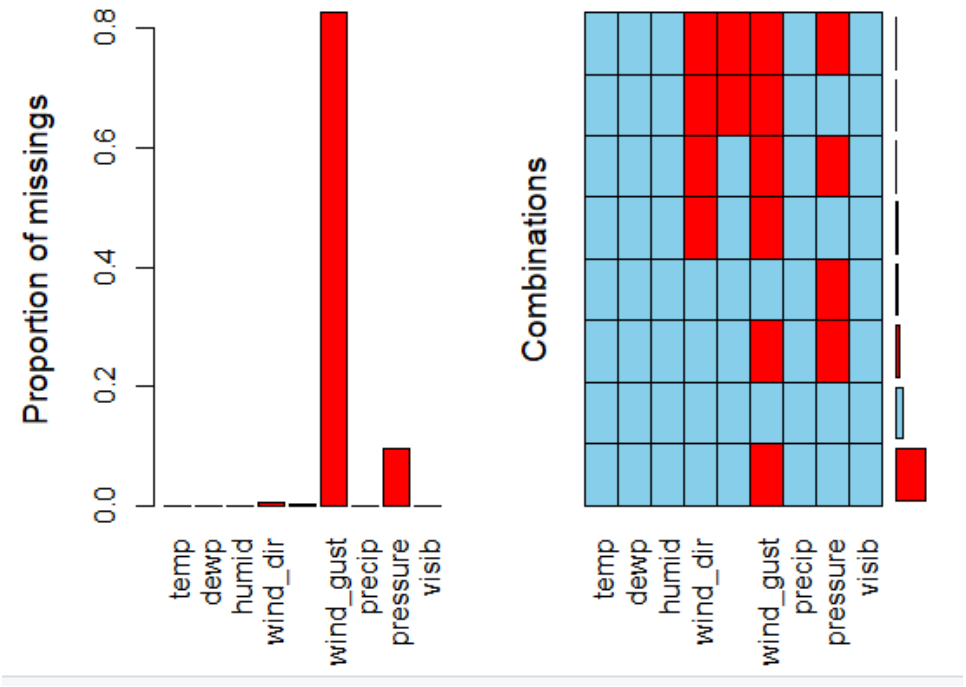
- This analysis is about the co-variation of between wind_gust and wind_Speed in overall data. The analysis is conducted because to show the co-variation nrelationship changes among the different month in 2013.
- On the code, I suppressed the warning and read the csv file, and removing NA values for more accurate plotting. Then I use geom plot type of graph and add color “red” to the graph and add geom_smooth graph with method = lm. after that, I seperated the graphs in 2013 by monthly and set the title, x and y axis to the graph.

- Based on the analysis, it shows that the dewpoint has a strong relationship with temperature in 2013. However, the co-variation of wind gust and wind speed varies among different month. The plot started maintaining the plots from January to April, then started to decrease afterwards on the months, meanwhile the line decreased from January until June and suddenly the line sparkled to a long distance on July, and decreased again afterwards from august to December in 2013.

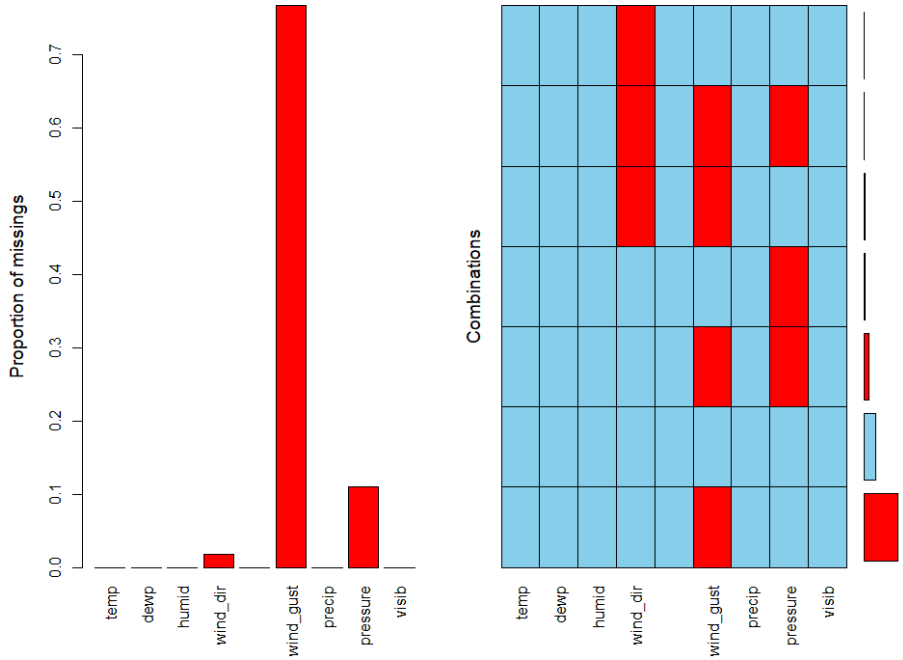
3.10 Analysis 10: summarization of NA values of JFK and LGA in 2013

```
JFKNAvalues = JFK %>% select(temp:visib) %>% #select temp to visib from JFK data and set to JFKNAvalues

  aggr(prop=T, numbers=T) #use aggr to return an array of values and turn prop and numbers to TRUE

summary(JFKNAvalues)


LGANAvalues = LGA %>% select(temp:visib) %>% #select temp to visib from JFK data and set to LGANAvalues

  aggr(prop= T, numbers = T) #use aggregation to return an array of values and turn prop and numbers to TRUE

summary(LGANAvalues)
```

- This analysis is about showing the statistic of missing NA values of the JFK and LGA in 2013. This analysis is conducted to compare the statistic of the missing NA values between both airport.
- On the code, I use the JFK and LGA filtered data and select the data from temperature to visibility on the csv file.



- Statistic of missing value of JFK on 2013.

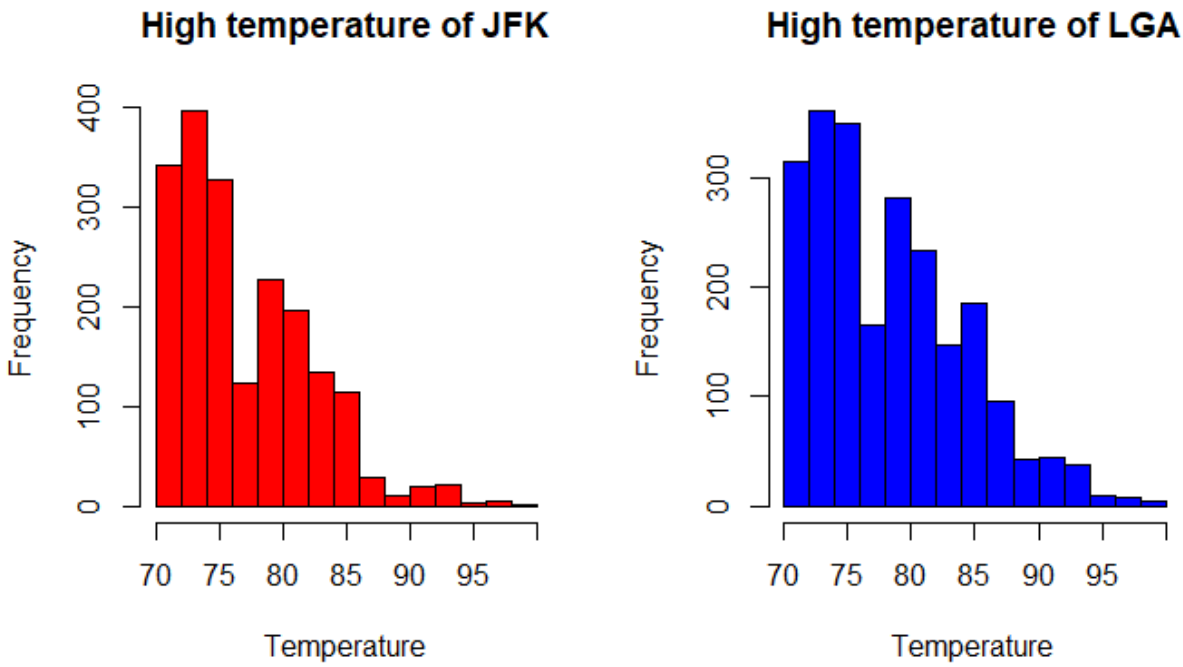


- Statistic of missing value of LGA on 2013.

3.11 Analysis 11: Temperature of 70°F or above on JFK and LGA in 2013

```
JFKTemp = filter(JFK, temp >70) #set JFKTemp to filter temp above 70°F only in JFK data

LGATemp = filter(LGA, temp >70) #set LGATemp to filter temp above 70°F only in LGA data

par(mfrow = c(1,2)) #create a simple multi-paneled plot 1x2 mfrow

hist(JFKTemp$temp, main ="High temperature of JFK in 2013",xlab ="Temperature", col ="red") #plot histogram JFKTemp$temp with setting title, x axis name and color histogram to red

hist(LGATemp$temp, main ="High temperature of LGA in 2013",xlab ="Temperature", col ="red") #plot histogram LGATemp$temp with setting title, x axis name and color histogram to red
```

- This analysis is about showing the high temperature of JFK and LGA in 2013. This analysis is conducted to show and compare of the high temperature of both airport ,JFK and LGA in 2013.
- On the code, I filtered both JFK and LGA to show only 70°F or above of temperature. I made a multi-panel plot of 1x2 mfrow and plot a histogram for both JFK and LGA data to compare.

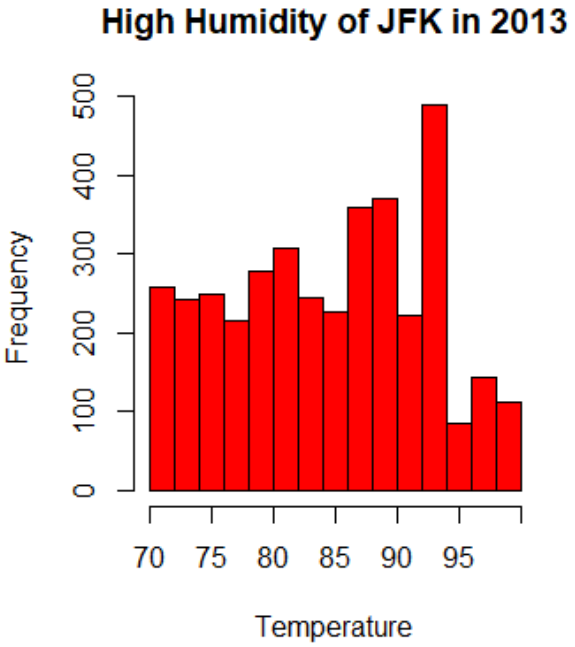


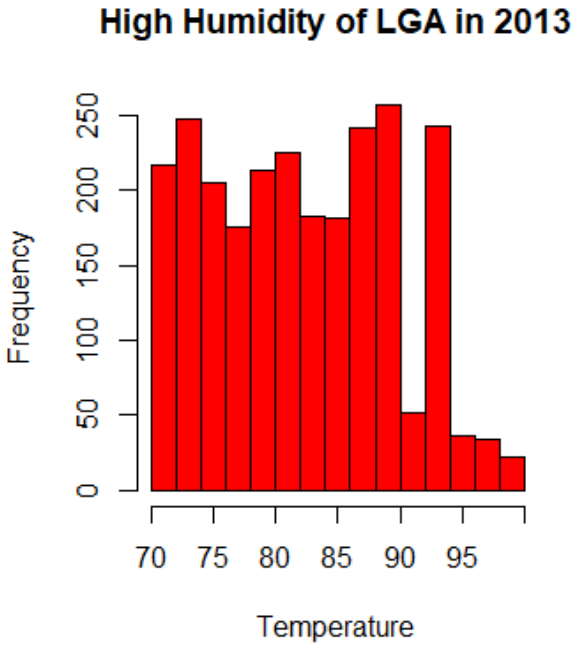
- Based on the graph, it shows the high temperature of both JFK and LGA airport had few similar of data frequency in 2013.

3.12 Analysis 12: High and Low humidity of JFK and LGA in 2013

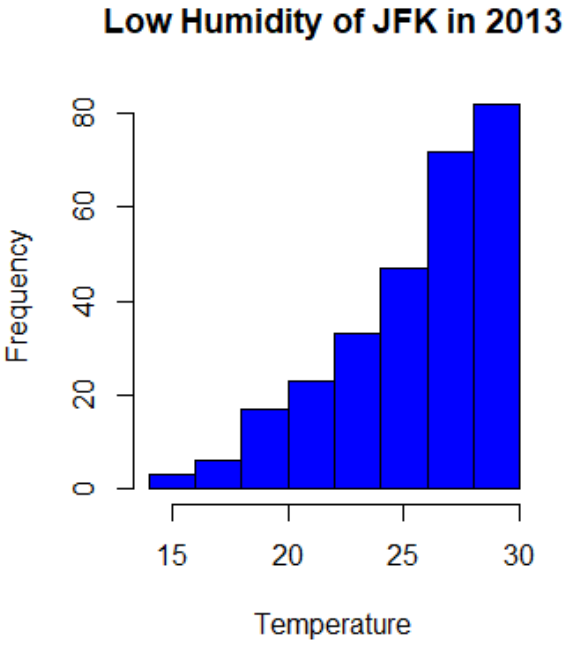
```
JFKHighHumid = filter(JFK,humid >70) #set JFKHighTemp to filter humidity above 70 only in JFK data

LGAHighHumid = filter(LGA,humid >70) #set LGAHighTemp to filter humidity above 70 only in LGA data

par(mfrow = c(1,2)) #create a simple multi-paneled plot 1x2 mfrow

hist(JFKHighHumid$humid, main ="High Humidity of JFK in 2013",xlab ="Temperature", col ="red") #plot histogram JFKHighHumid$humid with setting title, x axis name and color histogram to red

hist(LGAHighHumid$humid, main ="High Humidity of LGA in 2013",xlab ="Temperature", col ="red") #plot histogram LGAHighHumid$humid with setting title, x axis name and color histogram to red


JFKLowHumid = filter(JFK,humid <30) #set JFKLowTemp to filter humidity below 30 only in JFK data

LGALowHumid = filter(LGA,humid <30) #set LGALowTemp to filter humidity below 30 only in LGA data

par(mfrow = c(1,2)) #create a simple multi-paneled plot 1x2 mfrow

hist(JFKLowHumid$humid, main ="Low Humidity of JFK in 2013",xlab ="Temperature", col ="blue") #plot histogram JFKLowHumid$humid with setting title, x axis name and color histogram to blue

hist(LGALowHumid$humid, main ="Low Humidity of LGA in 2013",xlab ="Temperature", col ="blue") #plot histogram LGALowHumid$humid with setting title, x axis name and color histogram to blue
```

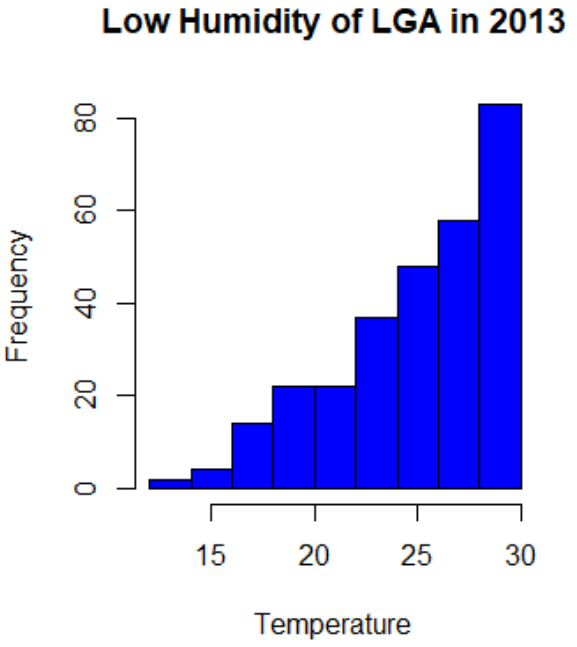
- This analysis shows the temperature between 30 or below or 70 or above that appears on JFK and LGA in 2013. This analysis is conducted to show low humidity and high humidity of 2013 and purpose is to reduce hazards and risks that effects the pilot and airplanes due to extreme heat.
- On the code, I filtered out the data from JFK and LGA data to show only humidity with 70 or above and 30 or below. Then I created multi paneled plot of 1x2 mfrow and plot the data each in every histogram and set colors.





- Based on the graph of high humidity of JFK and LGA in 2013, it shows that JFK has slightly more data from 90 and onwards when comparing to LGA airport.



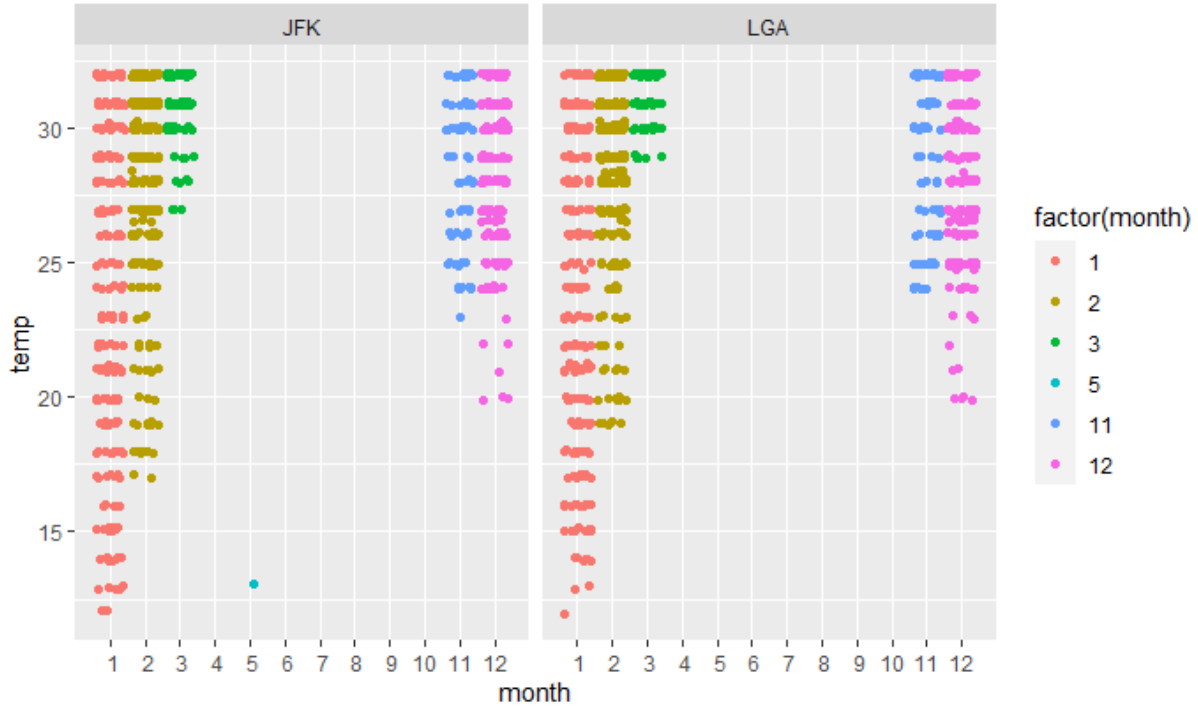


- Based on the graph of low humidity of JFK and LGA in 2013, it shows that JFK has more data starting from 25 to 30 humidity when comparing to LGA airport.

3.13 Analysis 13: Temperature below ice point (32°F) of JFK and LGA in 2013

```
options(warn = -1) #suppressing warning messages

Dataassignment %>% filter(temp <= 32) %>% #read csv file and filter temp under 32°F.

  ggplot(aes(x=month, y=temp, group =month, color = factor(month), na.omit()))+ #ggplot graph, set color to month and remove NA values

  geom_jitter()+ facet_wrap(~origin) + scale_x_discrete(limits = c(1:12)) #set geom_jitter, seperate graph by origin and scale x limit from 1 to 12
```

- This analysis is about finding the data of temperature below ice point of JFK and LGA in 2013. This analysis is conducted to find out the temperature below ice point which is under 32°F to reduce
- On the code, I selected and filtered temperature below ice point which is 32°F from the csv file, and create a extra feature jitter plot type graph because it helps to visualize the separation of the data better.



- Based on this graph findings, most of the ice point temperature appears on the month of January, February, March, November and December for both JFK and LGA airport in 2013.

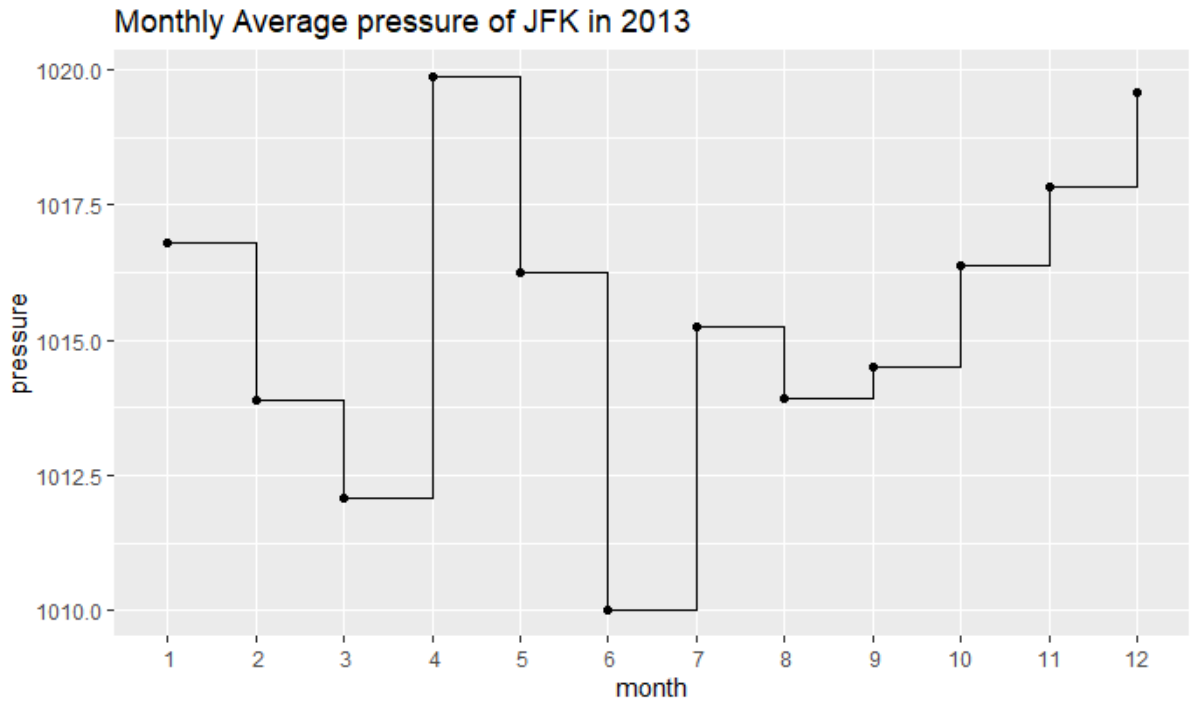
3.14 Analysis 14: Average pressure by month of JFK and LGA in 2013 (Extra Feature 2: Geom_Step)

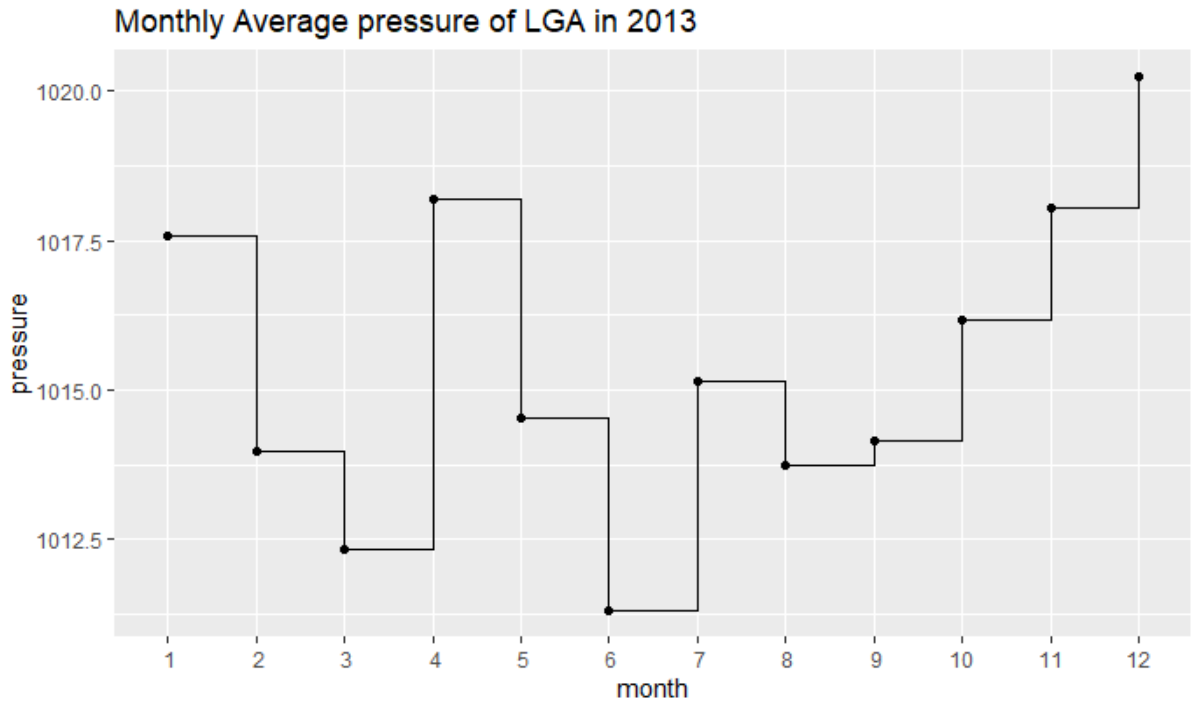
```
ggplot(JFK %>% na.omit() %>% group_by(month) %>% #ggploting JFK data, remove NA values and group data by month
       summarise(pressure = mean(pressure)), #summarize to find mean point of pressure each month
      aes(x = month, y = pressure)) +
  geom_point() + #setting geom_point
  geom_step() + scale_x_discrete(limits = c(1:12)) + #add geom_step and scale x limit from 1 to 12 to graph (Extra feature 2: geom_step)
  labs(title = "Monthly Average pressure of JFK in 2013") #set title to graph
```

```
ggplot(LGA %>% na.omit() %>% group_by(month) %>% #ggploting LGA data, remove NA values and group data by month
       summarise(pressure = mean(pressure)), #summarize to find mean point of pressure each month
      aes(x = month, y = pressure)) +
  geom_point() + #setting geom_point
  geom_step() + scale_x_discrete(limits = c(1:12)) + #add geom_step and scale x limit from 1 to 12 to graph (Extra Feature 2, Geom_step)
  labs(title = "Monthly Average pressure of LGA in 2013") #set title to graph
```

- This analysis finding is about the average mean point of pressure for both airport JFK and LGA in 2013. This analysis is conducted to calculate the monthly mean pressure to describe its changes monthly throughput the time in 2013.
- On the code, I have filtered JFK and LGA data, removed NA values for accurate data and summarized the pressure data by monthly. I used geom_point and geom_step to visualize the graph. For the extra feature, Extra Feature that i used is Geom_step because it is to visualize the graph into showing a better and accurate visualization of the result on the graph.

34.1



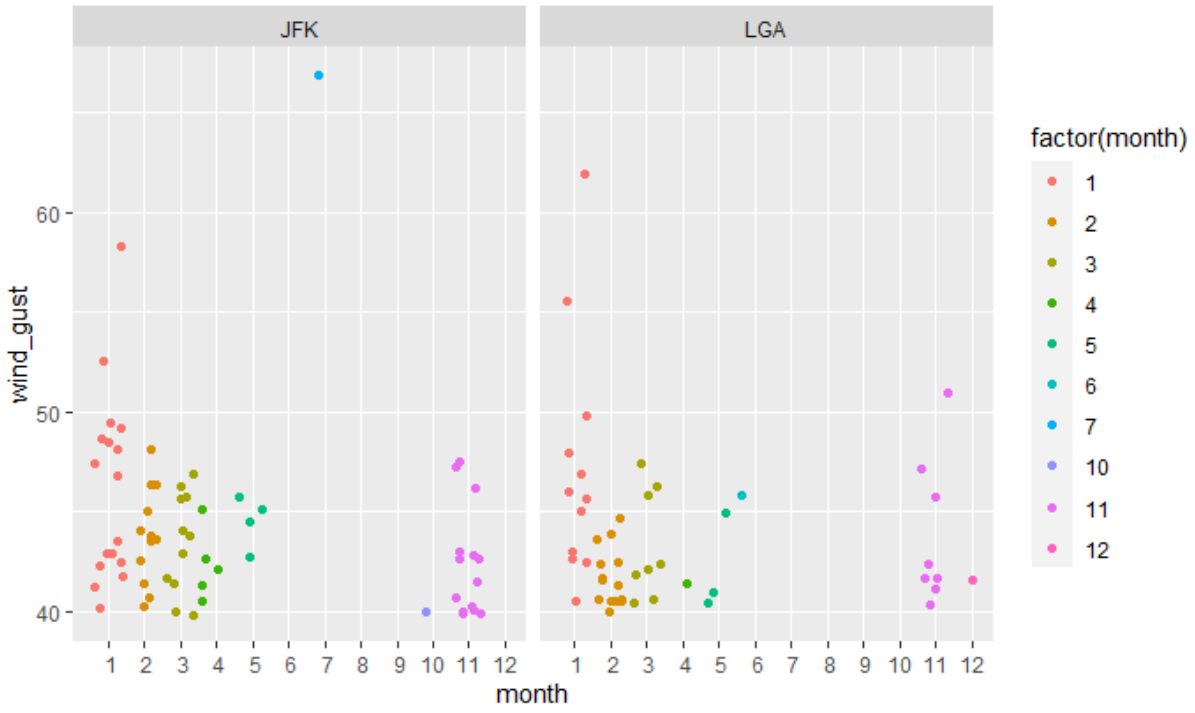


- Based on the graph, the analysis finding shows the monthly air pressure for JFK is slightly higher than LGA from the overall on 2013. It shows both airport of the average pressure the first three months dropped as shown in mean pressure and hit the lowest at June. Then slowly raising back up to the highest pressure mean at December.

3.15 Analysis 15: High wind gust in JFK and LGA in 2013

```
Dataassignment %>% filter(wind_gust >= 40) %>% #filter wind gust data above 40 from csv file

  ggplot(aes(x=month, y=wind_gust, group =month, color = factor(month), na.omit()))+ #ggplot graph, group by both, add color by month and removing NA values

  geom_jitter()+ facet_wrap(~origin) + scale_x_discrete(limits = c(1:12)) #set geom_jitter and seperate graph by origins and scale x limit from 1 to 12.
```

- This analysis finding is about finding and displaying wind gust of 40 or above on both airport JFK and LGA in 2013. This analysis is conducted is to reduce the risk and potential damages to the airport or danger that could delay the airport, injuries or even death due to event such as power outages, transportation disruption or damage to buildings and airplanes or ground vehicles.
- On the code, I have selected the whole data from the csv file and filtered them to show data only wind gusts of 40 or greater. Then I removed the NA values from the data to get more accurate data and seperated the graphs by origins, which are JFK and LGA and scaled the graph to show monthly data.



- Based on the graph, it shows that the high wind gust that extended 40 frequency happens on both airport starting from January to May and November. On June, there is one extremely high wind gust that is soaring pass the 50 frequency of wind gust for both airport and one passed the 60 frequency of wind gust has occurred on JFK airport.

4.0 Conclusion

- based on the 15 analysis and 2 extra features in the analysis that I have listed above, I concluded that I have found some findings in details below:

1. Both of the data of JFK and LGA for weather is very similar. For instance, the temperature and dew points clearly shows the changes during seasonal variation. Example seasonal event would be, strong winds, ice point temperatures and low visibility of both airports. Moreover, those event occurs during spring and winter. Therefore, the result is not very surprising as the both airport is located in New York and being apart of each other lesser than 20 km.

2. the events such as poor visibility, ice and dew point and strong winds happens during the winter season and early spring, which are around November, December, January and February.

3. The heavy rain or storm rain happens during the summer and winter, which are around may and June or November and December. In the findings, it shows that October has the least precipitation than the other month in 2013. Therefore, workers should be more caution to event such as heavy or storm rains during summer and winter.

4. During the early days of January, the humid and air pressure can be used to predict the strong winds as it could be used to explained well.

5.0 Reference List

Finavia.fi. 2017. *How Do Wind Conditions Affect Flight?* | *Finavia*. [online] Available at: <https://www.finavia.fi/en/newsroom/2017/how-do-wind-conditions-affect-flight> [Accessed 17 November 2020].

Ggplot2.tidyverse.org. n.d. *Smoothed Conditional Means — Geom_Smooth*. [online] Available at: <https://ggplot2.tidyverse.org/reference/geom_smooth.html> [Accessed 21 November 2020].

Goglia, J., 2012. *Extreme Heat And Humidity Cause Unsafe Conditions On The Airport Ramp.*. [online] Aviation Pros. Available at: <https://www.aviationpros.com/airports/airports-municipalities/blog/10737557/extreme-heat-and-humidity-cause-unsafe-conditions-on-the-airport-ramp> [Accessed 20 November 2020].

Gyre.umeoce.maine.edu. n.d. *Variable Description*. [online] Available at: <http://gyre.umeoce.maine.edu/data/gomoos/buoy/php/variable_description.php?variable=wind_speed> [Accessed 16 November 2020].

Rstudio.com. 2015. [online] Available at: <https://rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf> [Accessed 19 November 2020].

Sape.inf.usi.ch. n.d. *Ggplot2 Quick Reference: Geom_Step* | *Software And Programmer Efficiency Research Group*. [online] Available at: <http://sape.inf.usi.ch/quick-reference/ggplot2/geom_step> [Accessed 21 November 2020].

Schultz, M., Lorenz, S., Schmitz, R. and Delgado, L., 2018. *(PDF) Weather Impact On Airport Performance*. [online] ResearchGate. Available at: <https://www.researchgate.net/publication/328291850_Weather_Impact_on_Airport_Performance> [Accessed 21 November 2020].

Index of comments