INDIVIDUAL ASSIGNMENT

TECHNOLOGY PARK MALAYSIA

CT127-3-2-PFDA

PROGRAMMING FOR DATA ANALYSIS

UC2F2008CS

HAND OUT DATE	: **01 OCTOBER 2020**
HAND IN DATE	: **22 NOVEMBER 2020**
WEIGHTAGE	: **50%**

Name:

STEVEN

TP. No:

TP054243

Table of Contents

1. Introduction and Assumptions

1.1. Introduction

Nowadays, data is an essential thing where it can be useful in almost any kind of sector. According to Import.io (2018), data is a collection of facts such as numbers, words, measurements, and observations that have been translated into a form that computers can process. Considering the importance of data has been noticed by the people, the number of data is growing as well. Forbes (2015) found that every second, people produce new data. For example, there are 40,000 search queries every second on Google, which makes it 3.5 billion searches per day and 1.2 trillion searches per year. Hence, the way of processing the data which is data analysis is important as well. ORI (2016) states that data analysis is the systematic process of applying statistical and/or logical techniques to describe, illustrate, and evaluate the data.

In this paper, the hourly meteorological data for LaGuardia Airport (LGA) and John F. Kennedy International Airport (JFK) in the United States is provided and will be analyzed with a proper visualization, techniques, and explanation. Considering the data is quite huge, therefore R Studio will be used for the analysis process. Lastly, useful information or decision will be produced for each analysis.

1.2. Assumptions

- For season column, assumed that:
  - 1: Spring
  - 2: Summer
  - 3: Autumn
  - 4: Winter
- Both LaGuardia Airport (LGA) and John F. Kennedy International Airport (JFK) are in New York, USA which is part of Northern Hemisphere area.
- Assumed that, temperature at both airports represents the average temperature in Queens, New York.
- Relative humidity is in percentage (%).
- In this paper, assumed that 100% humidity is indicating a rain.
- The temperature in the dataset is ambient temperature.

2. Pre-processing

2.1. Add season column based on day and month

```
#-add season column based on day and month
full_data=mutate(full_data, season =
               if_else((month==3&day>=20)|month==4|month==5|(month==6&day<=19),1,
               if_else((month==6&day>=20)|month==7|month==8|(month==9&day<=21),2,
               if_else((month==9&day>=22)|month==10|month==11|(month==12&day<=20),3,4))))
```

Figure 2.1.1 Coding for adding season column

This process is adding `season` column in the raw data (`full_data`) based on the day and month for each season. According to Calendar Date (2020), for the Northern Hemisphere area, the Spring for year 2013 started on March 20th, Summer started on June 20th, Autumn started on September 22nd, and Winter started on December 21st. Considering both LGA and JFK are located in New York which is the Northern Hemisphere area (Britannica, 2020), hence the information can be used to create the `season` column.

Here is the result:

origin	year	month	day	hour	temp	dewp	humid	wind_dir	wind_speed	wind_gust	precip	pressure	visib	time_hour	season
JFK	2013	3	19	21	42.08	30.02	62.04	320	20.71404	27.61872	0.00	1008.4	10.0	19/03/2013 21:00	4
JFK	2013	3	19	22	39.92	24.98	54.81	300	17.26170	21.86482	0.00	1009.2	10.0	19/03/2013 22:00	4
JFK	2013	3	19	23	39.02	26.06	59.37	290	14.96014	19.56326	0.00	1009.7	10.0	19/03/2013 23:00	4
JFK	2013	3	20	0	37.04	24.08	59.09	310	13.80936	NA	0.00	1010.1	10.0	20/03/2013 00:00	1
JFK	2013	3	20	1	37.04	19.94	49.62	290	13.80936	24.16638	0.00	1010.6	10.0	20/03/2013 01:00	1
JFK	2013	3	20	2	35.96	15.08	41.99	280	11.50780	NA	0.00	1011.4	10.0	20/03/2013 02:00	1

Figure 2.2.2 Coding for adding season column

3. Analysis

3.1. Analysis 1: Finding the best season for holiday based on the temperature

```
#-manipulation
full_data%>%group_by(season)%>%summarise(avg_temp=mean(temp),
                                          max_temp=max(temp),
                                          min_temp=min(temp), .groups='drop')
#-visualization
ggplot(full_data, aes(season, temp, fill=season)) + geom_boxplot(aes(group=season))
```

Figure 3.1.1 Coding for Analysis 1

In the Figure 3.1.1 shows that the temperature will be sorted by season using `group_by()`. Then, the average, maximum, and minimum temperature for each season will be summarised using `summarise()`. In addition, `.groups='drop'` is used to remove the group attribute of `summarise()`, hence there will be no group overlapping.

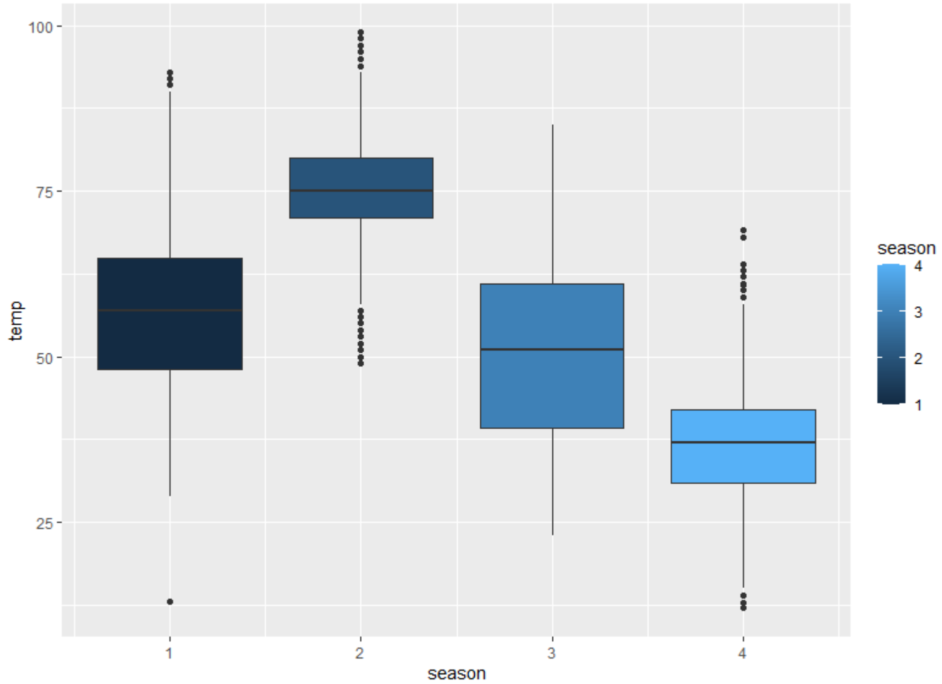

Figure 3.2.2 Boxplot diagram for Analysis 1

```
# A tibble: 4 x 4
  season avg_temp max_temp min_temp
   <dbl>    <dbl>    <dbl>    <dbl>
1      1     57.0     93.0     13.1
2      2     75.3     99.0     48.9
3      3     50.3     84.9     23
4      4     36.6     69.1     12.0
```

Figure 3.3.3 Temperature summary for each season

According to Scientific American (2009), the human body feels comfortable when the ambient temperature is around 70° F. We know that holiday is supposed to make people comfortable and relax. Based on Figure 3.1.2 and Figure 3.1.3, it shows that season 2 (Summer) has the desired temperature. **Therefore, Summer is the best season for a holiday at Queens, New York.**

3.2. Analysis 2: Analyzing average humidity at both airports for each month

```r
#-manipulation
jfk_humid=full_data %>% filter(origin=="JFK") %>% group_by(month) %>% #JFK
  summarise(humid=mean(humid),origin, .groups='drop')
lga_humid=full_data %>% filter(origin=="LGA") %>% group_by(month) %>% #LGA
  summarise(humid=mean(humid),origin, .groups='drop')

#-visualization
ggplot() +
  geom_line(data = jfk_humid, aes(x = month, y = humid, color=origin)) +
  geom_line(data = lga_humid, aes(x = month, y = humid, color=origin)) +
  scale_x_continuous(breaks=1:12,labels=c("jan","feb","mar","apr","may","jun",
                                          "jul","aug","sep","oct","nov","dec")) +
  labs(title="LGA and JFK Humidity in a year", y="Humidity (%)",x="Month")
```

Figure 3.2.1 Coding for Analysis 2

In this analysis, the raw data will be separated in order to get different humidity value for each airport. After separated, the average humidity will be calculated using `mean()`. Furthermore, the visualization will have two lines where each line shows the data for each airport. Then, for the purpose of readability, the month value which is in numeric form will be changed into character form using `scale_x_continuous()`.

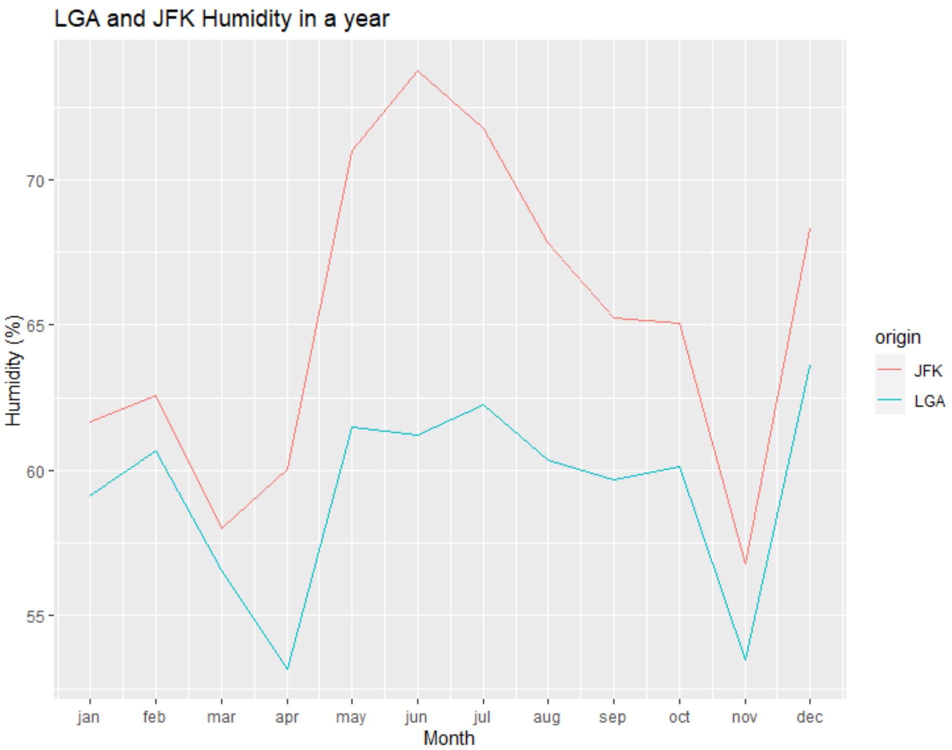

Figure 3.2.2 Line diagram for Analysis 2

According to PhysLink (n.d), humidity is one of the factors that decreases the performance of most aircraft, not only because of its effect on the wings, but also the effect on the engines. As can be seen in Figure 3.2.1, JFK airport has higher humidity percentage than LGA. **Therefore, extra treatments and maintenance will be needed at JFK airport to make sure the aircraft that stay at JFK airport have a good performance**.

3.3. Analysis 3: Analyzing rainy days based on humidity

Smart Fog (2016) found that when it rains, the humidity percentage is 100, which is why the clouds are unable to hold any more water.

```
#-manipulation
#--assumed that humid = 100 is indicating a rain
jfk_rain=full_data %>% filter(humid==100, origin=="JFK") %>% group_by(month) %>% #JFK
  summarise(rain=n_distinct(day),origin,.groups='drop') #multiple rain in a day will be counted as once
lga_rain=full_data %>% filter(humid==100, origin=="LGA") %>% group_by(month) %>% #LGA
  summarise(rain=n_distinct(day),origin,.groups='drop') #multiple rain in a day will be counted as once

#-visualization
ggplot() +
  geom_point(data=jfk_rain, aes(x = month, y = rain, color=origin),size=3) + #JFK plot
  geom_point(data=lga_rain, aes(x = month, y = rain, color=origin),size=3) + #LGA plot
  annotate("rect", xmin=3.6, xmax=6.6, ymin=1 , ymax=6, alpha=0.1, fill="green")+ #Summer time
  scale_x_continuous(breaks=1:12,labels=c("jan","feb","mar","apr","may","jun",
                                          "jul","aug","sep","oct","nov","dec"))+
  scale_y_continuous(breaks=1:7) + labs(title="Number of Rainny Days in A Year",
                                        x="Month", y="Rainny Days")
```

Figure 3.3.1 Coding for Analysis 3

The code above explains that there will be separated data again for each airport, then it will be filtered where only humidity equals to 100 is taken. In the visualization, there are two plots for each airport and there will be a rectangle area that is indicating Summer. In addition, the rectangle area will be created using `annotate()`.

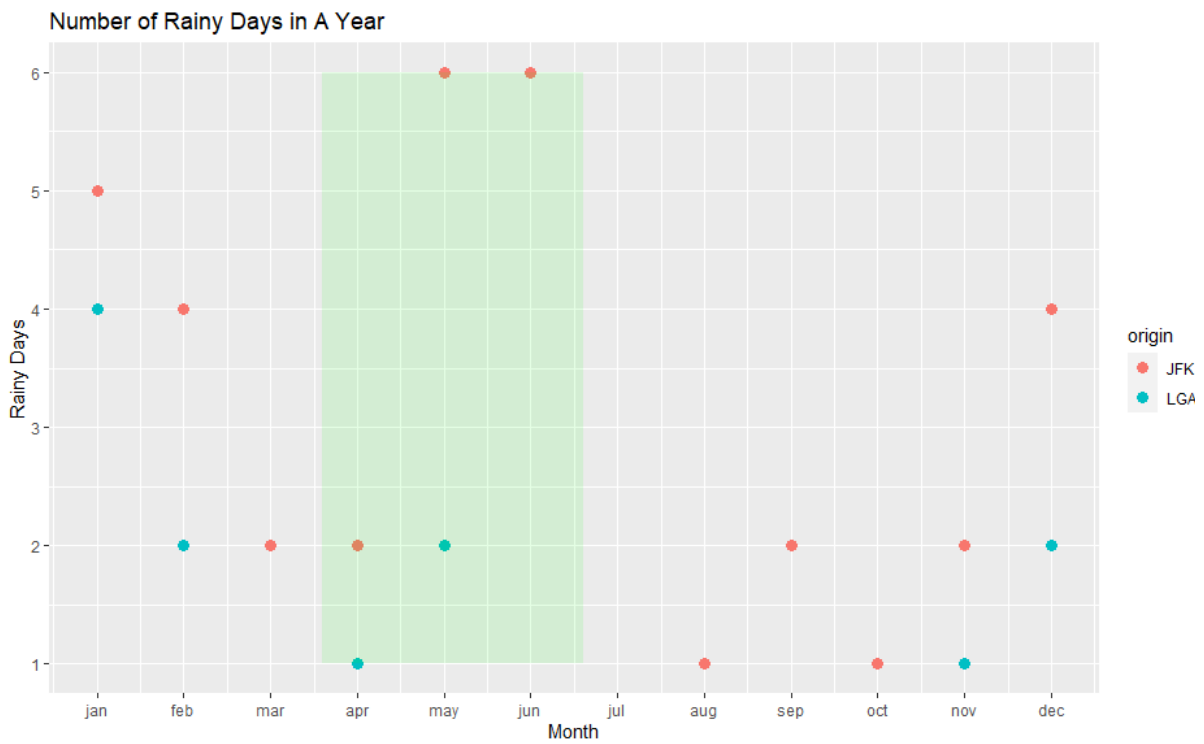

Figure 3.3.2 Plot diagram for Analysis 3

Based on the plot diagram above, it depicts that during Summer, JFK airport has a higher number of rainy days than LGA airport. **Thus, for people who are planning to travel in Summer, it is better to have a flight from LGA airport than JFK airport in order to avoid delays because of rain.**

3.4. Analysis 4: Analyzing Queens temperature in Summer

```
#-manipulation
summer_temp=full_data %>% filter(season==2) %>% group_by(day) %>%
  summarise(temp=max(temp),.groups='drop')

#-visualization
ggplot(summer_temp, aes(day, temp))+geom_line(aes(color=temp))+geom_point(aes(color=temp))+
  scale_x_continuous(breaks=1:31)+geom_hline(yintercept=90,color="orange")+ #show 31 days and make a warning
  labs(title="Queens Temperature in Summer 2013",subtitle="Average of maximum temperature per day",
       y="Temperature(F)",x="Day",color="Temperature")
```

Figure 3.4.1 Coding for Analysis 4

In this analysis, only summer data will be taken. Then, the data will only show the maximum temperature for each day. Moreover, a horizontal warning line will be drawn by using `geom_hline()` that indicates temperature that may harm human body. Furthermore, there will be 2 diagrams which are plot diagram and line diagram which proposed to make the result clearer.

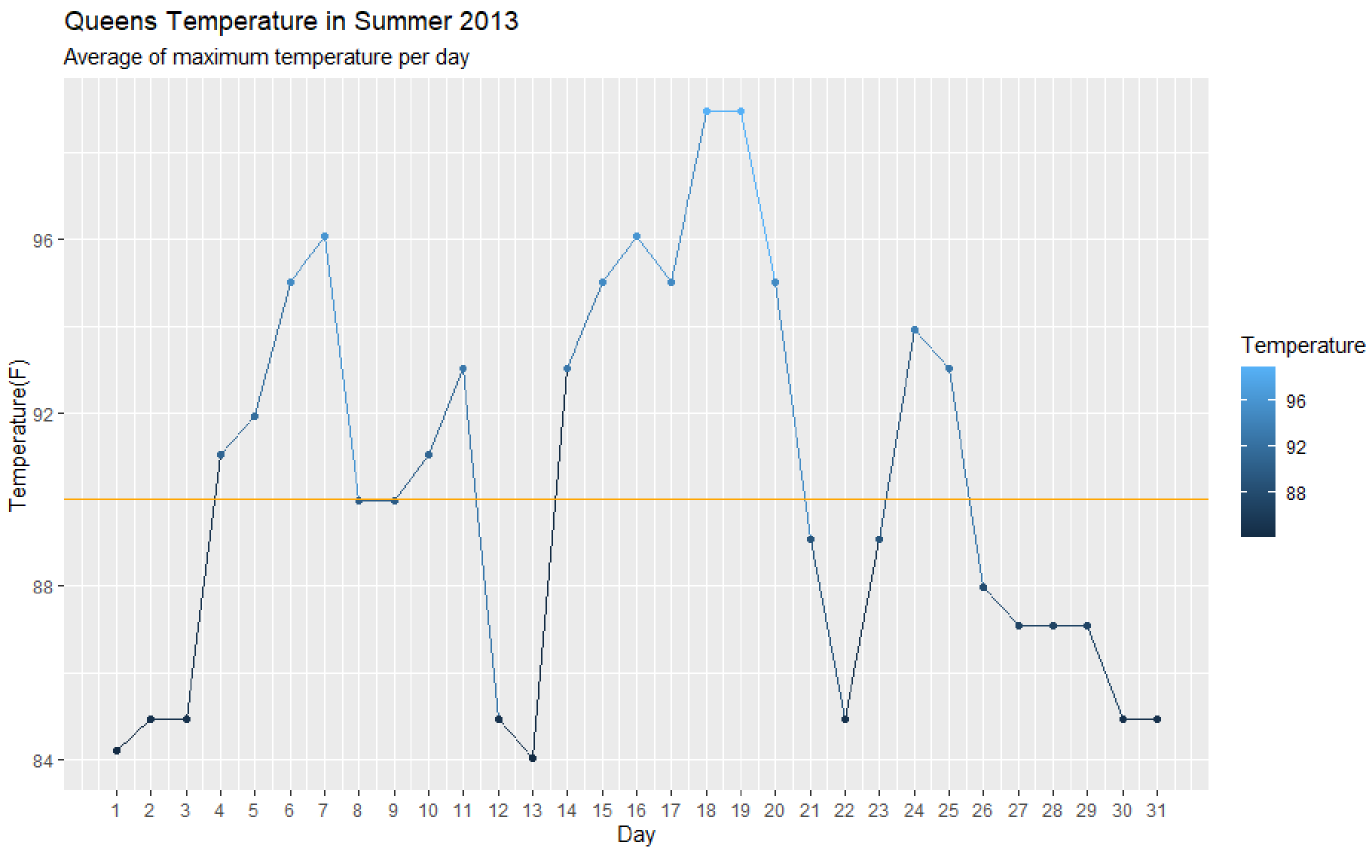

Figure 3.4.2 Plot diagram for Analysis 4

As can be seen in diagram above, there are quite a lot of days where the temperature is possibly harm human body. According to Elaine (2018), high environmental temperatures above 90˚F can caused heat cramps, heat exhaustion, or even heatstroke. **Hence, it is better to keep aware of the temperature and drink more water to stay hydrated.**

3.5. Analysis 5: Relationship between Temperature and Wind Speed in Winter

```
#-manipulation
winter_wind=full_data%>%filter(season==4)%>%select(temp,wind_speed)
summary(winter_wind) #show summary for temperature and wind speed in winter

#-visualization
ggplot(winter_wind, aes(temp,wind_speed)) + geom_point(na.rm=TRUE)+ #remove any NA value
  geom_vline(xintercept=10,color="blue",alpha=0.3)+ #frostbite potential temperature
  geom_hline(yintercept=55,color="blue",alpha=0.3)+ #frostbite potential wind speed
  annotate("pointrange", x=10, y=55, ymin=55, ymax=55,colour ="blue", size=0.77, alpha=0.6)+
  labs(title="Relationship between Temperature and Wind Speed",subtitle="Winter 2013",
       x="Temperature(F)",y="Wind Speed(mph)")+
  geom_text(label="Frostbite potential", x=16,y=56.5,color="blue")
```

Figure 3.5.1 Coding for Analysis 5

Based on the code, the analysis will use winter temperature and wind speed. Then, a vertical line and a horizontal line will be drawn to show the frostbite point. The lines will be assisted with a point that is created by using `annotate()`.

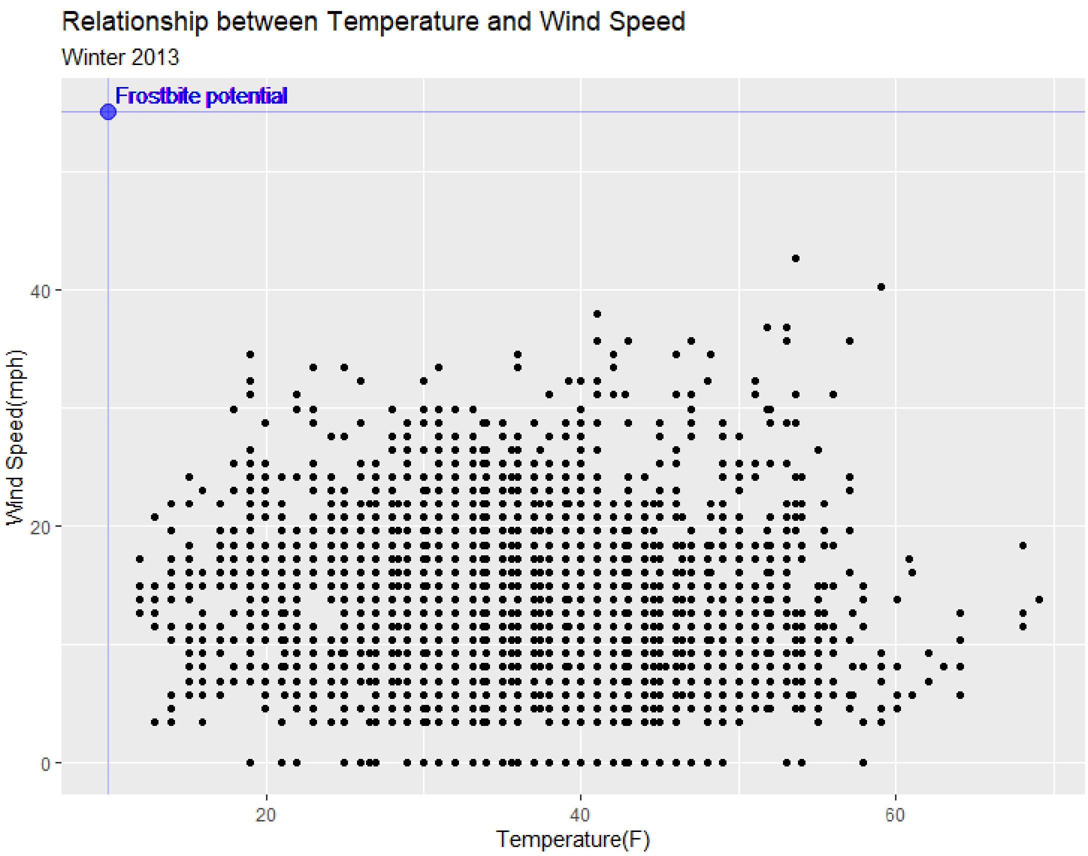

Figure 3.5.2 Plot diagram for Analysis 5

NOAA (2015) states that the combination of air temperature with the speed of the wind can affect an index called wind chill index. It determines how cold our bodies feel when we are outside. The stronger the wind is in the winter, the colder it feels and the lower is the wind chill index. According to Wenz (2015), based on the wind chill chart, 10˚F temperature with 60 mph wind speed will create a frostbite potential. **Thus, from Figure 3.5.2, it shows that there is no potential of frostbite during winter 2013.**

3.6. Analysis 6: Analyzing Queens annual temperature per season

```
#-manipulation
full_data%>%select(season,temp)%>%group_by(season)%>%summarise(avg_temp=mean(temp),.groups='drop')
#-visualization
ggplot(full_data, aes(x = temp, y=frequency(temp), fill = factor(season))) +
  geom_bar(width=0.18,stat = "identity")+ theme_bw() + labs(fill="Season",
                                                            x="Temperature(F)",
                                                            y="count",
                                                            title="Queens Temperature per-season",
                                                            subtitle = "2013")
```

Figure 3.6.1 Coding for Analysis 6

From the code, it indicates that this analysis will focus on temperature per season. It will be summarized in manipulation part and it will be visualize using `geom_bar()`.

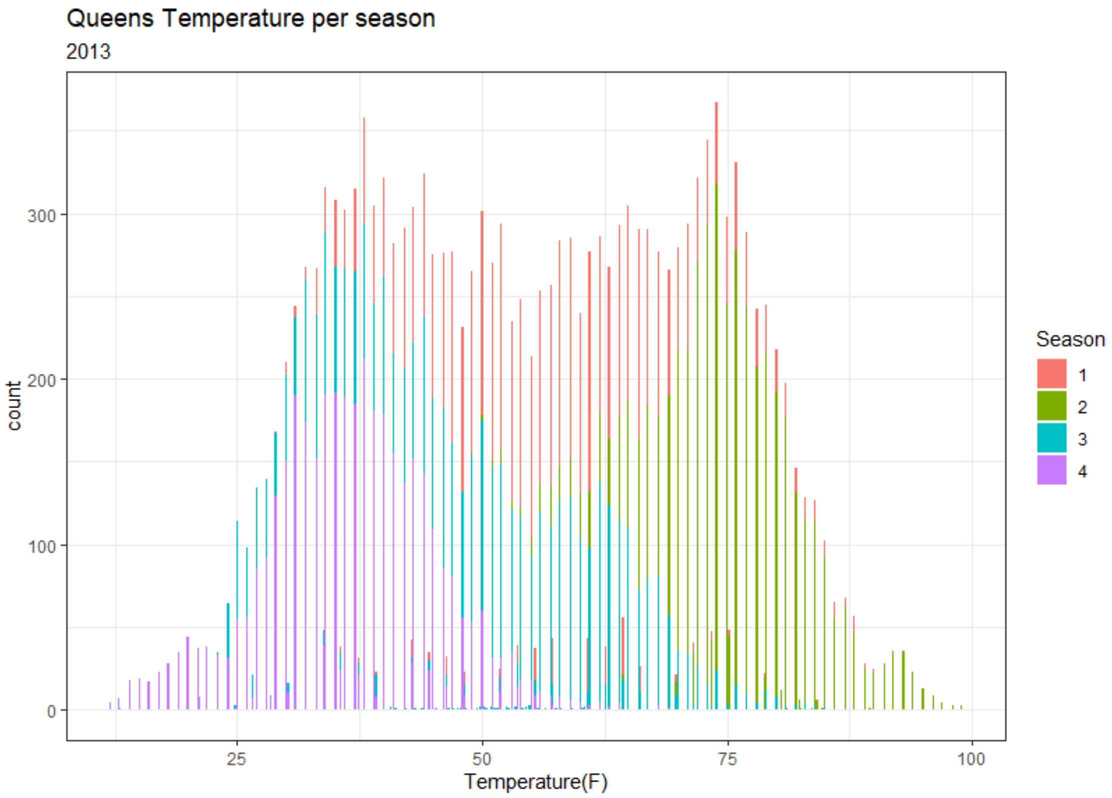

Figure 3.6.2 Diagram for Analysis 6

The diagram above shows nicely the temperature per season at Queens. As can be seen, **Spring has a moderate temperature where it is not too hot and not too cold**, hence working in Spring will be joyful and we can wear normal clothes as well. In addition, here is the table that shows the average temperature per season:

```
# A tibble: 4 x 2
  season avg_temp
   <dbl>    <dbl>
1      1     57.0
2      2     75.3
3      3     50.3
4      4     36.6
```

Figure 3.6.3 Average temperature per season

3.7. Analysis 7: Average temperature comparison between LGA and JFK in 2013

```
#-manipulation
jfk_temp=full_data %>% filter(origin=="JFK") %>% group_by(month) %>%
  summarise(temp=mean(temp),origin, .groups='drop') #find average temp for JFK
lga_temp=full_data %>% filter(origin=="LGA") %>% group_by(month) %>%
  summarise(temp=mean(temp),origin, .groups='drop') #find average temp for LGA
#-visualization
ggplot() +
  geom_line(data = jfk_temp, aes(x = month, y = temp, color=origin)) +
  geom_line(data = lga_temp, aes(x = month, y = temp, color=origin)) +
  scale_x_continuous(breaks=1:12,labels=c("jan","feb","mar","apr","may","jun",
                                          "jul","aug","sep","oct","nov","dec")) +
  labs(title="LGA and JFK Average Temperature in a year", y="Temperature(F)",x="Month")
```

Figure 3.7.1 Coding for Analysis 7

Figure 3.7.1 shows the code for analysis 7. The data will be separated again based on the origin airport. Instead of using overall temperature, this analysis will use the average temperature for each airport.

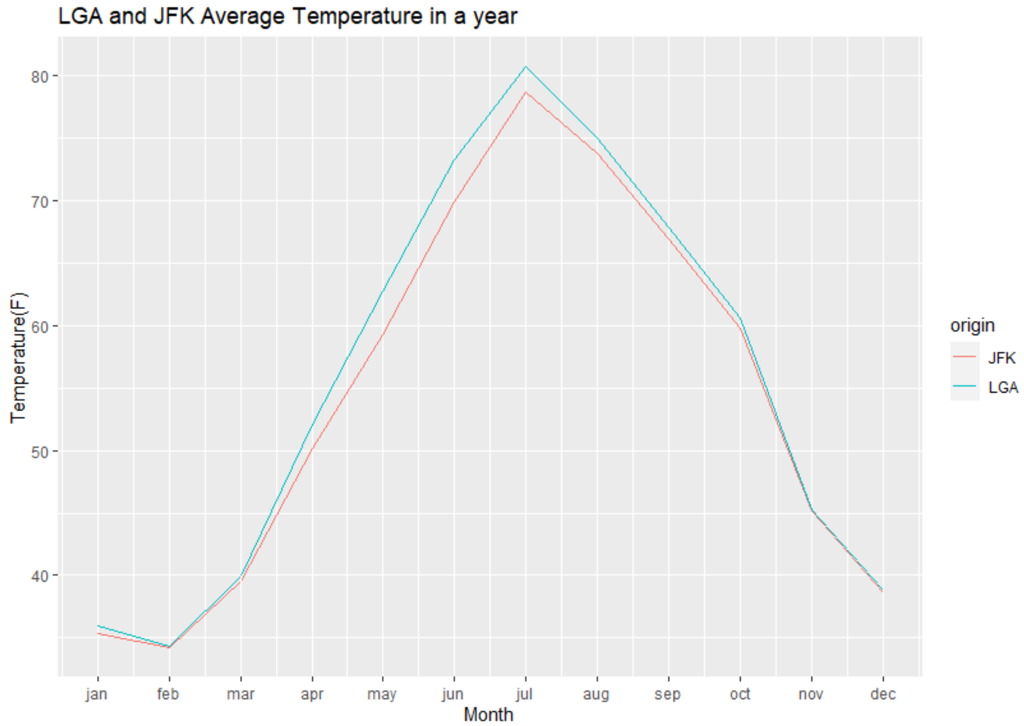

Figure 3.7.2 Diagram for Analysis 7

Line chart above describes that both airports have almost the same average temperature in 2013. In fact, LGA is located on the north of Queens and JFK is on the south. Considering the opposite location and the result above, it shows that **location of the airport does not give a massive impact towards temperature**.

Figure 3.7.3 Airport location
(Weill Cornell Medicine, n.d)

3.8. Analysis 8: Analyzing temperature for energy saving

```
#-manipulation
#--show temperature summary for each month
full_data%>%select(month,temp)%>%group_by(month)%>%summarise(avg_temp=mean(temp),
                                                            max_temp=max(temp),
                                                            min_temp=min(temp),.groups='drop')
#--for labeling purpose
monlab=c("1"="jan","2"="feb","3"="mar","4"="apr","5"="may","6"="jun","7"="jul",
         "8"="aug","9"="sep","10"="oct","11"="nov","12"="dec")

#-visualization
ggplot(full_data, aes(x=temp))+geom_histogram(bins=30,aes(fill = factor(season)))+
  facet_wrap(~month,labeller = labeller(month = monlab)) + #changing label name for each table
  geom_vline(xintercept=64,color="blue",alpha=0.5) + #safe and well-balanced indoor temperature line
  labs(title="LGA and JFK Temperature", y="Count",x="Temperature(F)",fill="Season",
       subtitle="2013")
```

Figure 3.8.1 Coding for Analysis 8

In this analysis, it will show the temperature in different tables for each month using `facet_wrap()`. Then, `geom_histogram()` will be used to show the result. In addition, a vertical line that indicates the safe and well-balanced indoor temperature will be created.

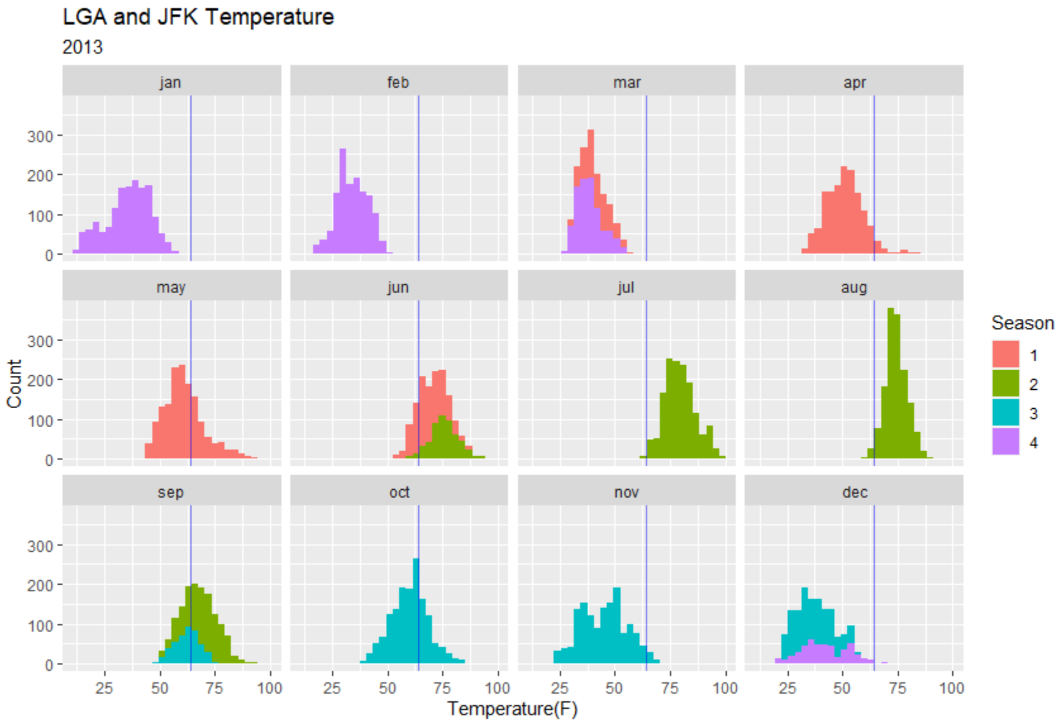

Figure 3.8.2 Histogram for Analysis 8

In an airport, it is an important thing to maintain the temperature in order to make the people inside feel comfortable. According to WHO (2018), 18˚C or 64˚F has been proposed as a safe and well-balanced indoor temperature to protect the health of general populations especially during cold seasons. As can be seen in the histogram above, the vertical line indicates the safe and well-balanced indoor temperature. September and October temperatures show the ideal temperature. Hence, the **airport management can turn off their Air Conditioner (AC) in September and October to save energy**. Referring to Analysis 7 result, this decision can be implemented at both airports.

3.9. Analysis 9: Effect of Wind Speed and Wind Direction towards Flight at LGA

```
#-manipulation
lga_hwind=full_data%>%filter(origin=="LGA")%>%
  #convert mph into Knot & calculate headwind
  mutate(knot_speed=wind_speed/1.151,headwind=knot_speed*cos(wind_dir-45))%>%
  select(month,knot_speed,wind_dir,headwind)%>%filter(headwind>=0)
#--show average and maximum headwind per month
lga_hwind%>%group_by(month)%>%summarise(hwind_avg=mean(headwind),
                                        hwind_max=max(headwind),.groups='drop')
#-visualization
ggplot(lga_hwind, aes(month,headwind))+geom_boxplot(aes(group=month))+
  scale_x_continuous(breaks=1:12,labels=c("jan","feb","mar","apr","may","jun",
                                          "jul","aug","sep","oct","nov","dec"))+
  labs(title="LGA Airport Headwind in 2013", y="Headwind",x="Month",subtitle="Runway 4")
```

Figure 3.9.1 Coding for Analysis 9

In this analysis, LGA data is chosen. However, the wind speed will be converted into Knot because we want to find the headwind which is using Knot in the calculation. Then, only headwind value that above 0 is selected because this analysis focused on the positive impact of headwind towards flight. Lastly, the result will be represented in the range of month using `geom_boxplot()`.

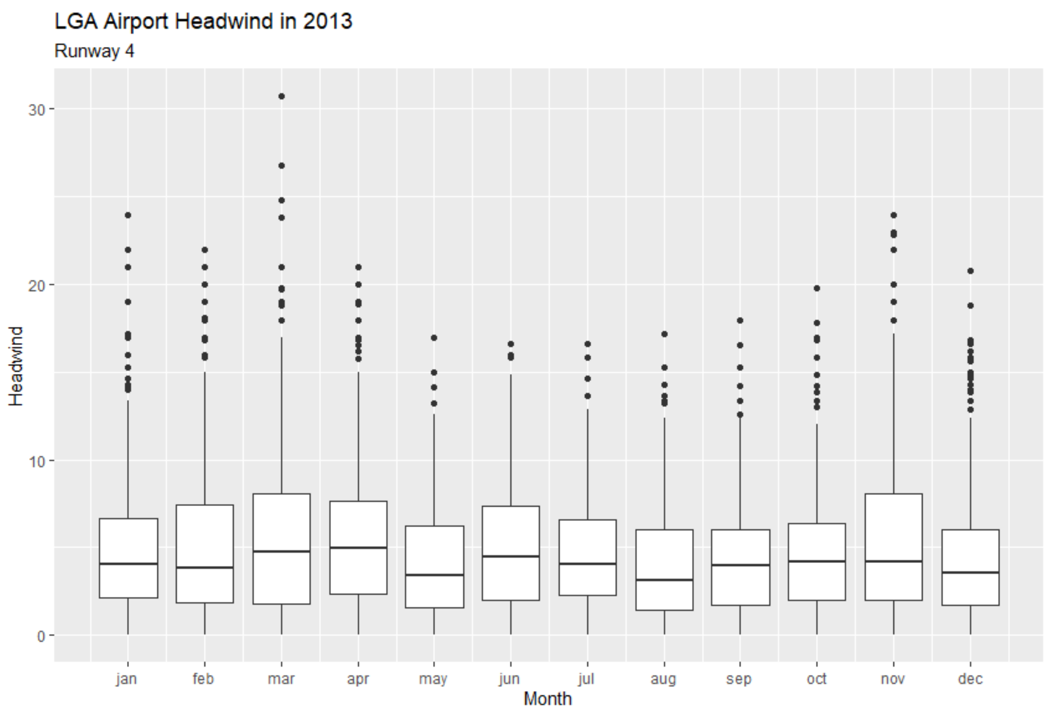

Figure 3.9.2 Boxplot for Analysis 9

According to Finavia (2017), headwind is wind blowing towards the plane. Moreover, pilots prefer to land and take off in headwind because it increases the lift. Then in this analysis, the headwind calculation is conducted by using formula that is based on IVAO (2020). Furthermore, we will analyze the flight that use Runway 4 (45°) of LGA airport only.

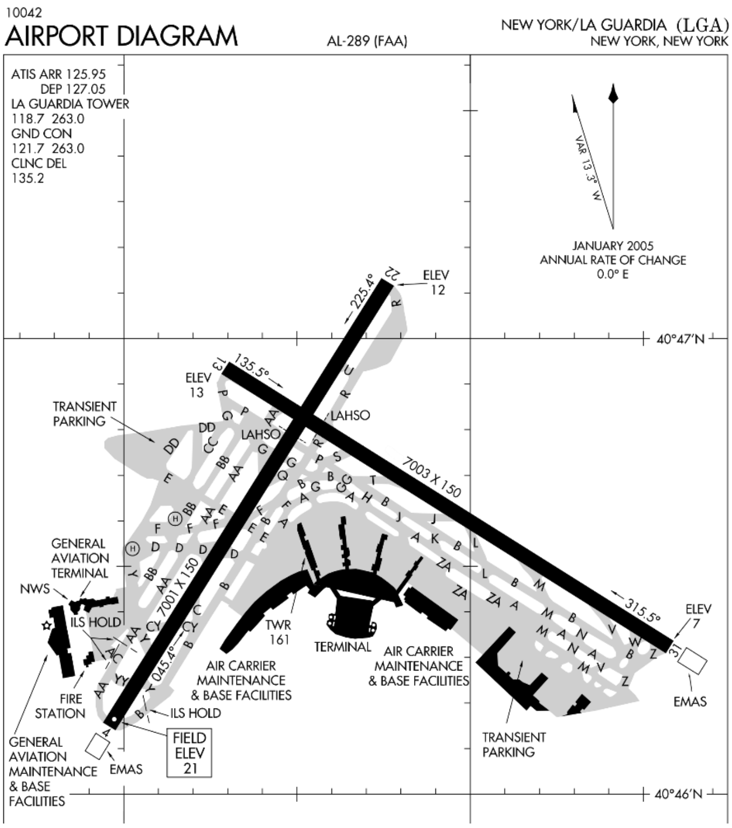

Figure 3.9.3 LGA Airport Runway (Wikimedia, n.d)

```
# A tibble: 12 x 3
   month hwind_avg hwind_max
   <int>     <dbl>     <dbl>
 1     1      4.85      24.0
 2     2      5.40      22.0
 3     3      5.92      30.7
 4     4      5.73      21.0
 5     5      4.18      17.0
 6     6      5.00      16.6
 7     7      4.55      16.6
 8     8      3.87      17.1
 9     9      4.28      18.0
10    10      4.63      19.8
11    11      5.65      24.0
12    12      4.52      20.8
```

Figure 3.9.4 Headwind Summary per month

As can be seen in Figure 3.9.2 and Figure 3.9.4, it shows that March is the month where the strongest headwind occurred and the highest headwind average. **Thus, the flight on March will use less energy than other months and use less runway to take off** (Finavia, 2017).

3.10. Analysis 10: Finding possibility of heat stroke in Summer

```
#-manipulation
#--temp >= 86 because heat stroke potential occurs from temperature 86F
summer_dewp=full_data%>%filter(season==2, temp>=86)%>%select(time_hour,dewp,temp)
#--data frame for heat index which indicates heat stroke potential (>104)
heatindex=data.frame(temp=c(90:100),dewp=c(78,77,76,75,76,73,72,71,70,68,68))

#-visualization
ggplot() +
  geom_point(data=summer_dewp,aes(x=dewp,y=temp))+
  geom_line(data=heatindex, aes(x=dewp,y=temp),color="red",size=2,alpha=0.2)+ #heatstroke potential line
  scale_x_continuous(breaks=53:80)+scale_y_continuous(breaks=86:100)+
  labs(title="Heat Stroke Potential in Summer", y="Temperature(F)",x="Dew Point(F)")
```

Figure 3.10.1 Coding for Analysis 10

This analysis will be conducted by using Summer data, dew point, and temperature that is more than or equal to 86˚F because the heatstroke potential occurs from that temperature (iWEATHERNET, 2016). In the code, a data frame of heatstroke potential temperatures and dew points will be created and will be displayed in the diagram. This analysis is using `scale_y_continuous()` to change the Y axis values into desired values.

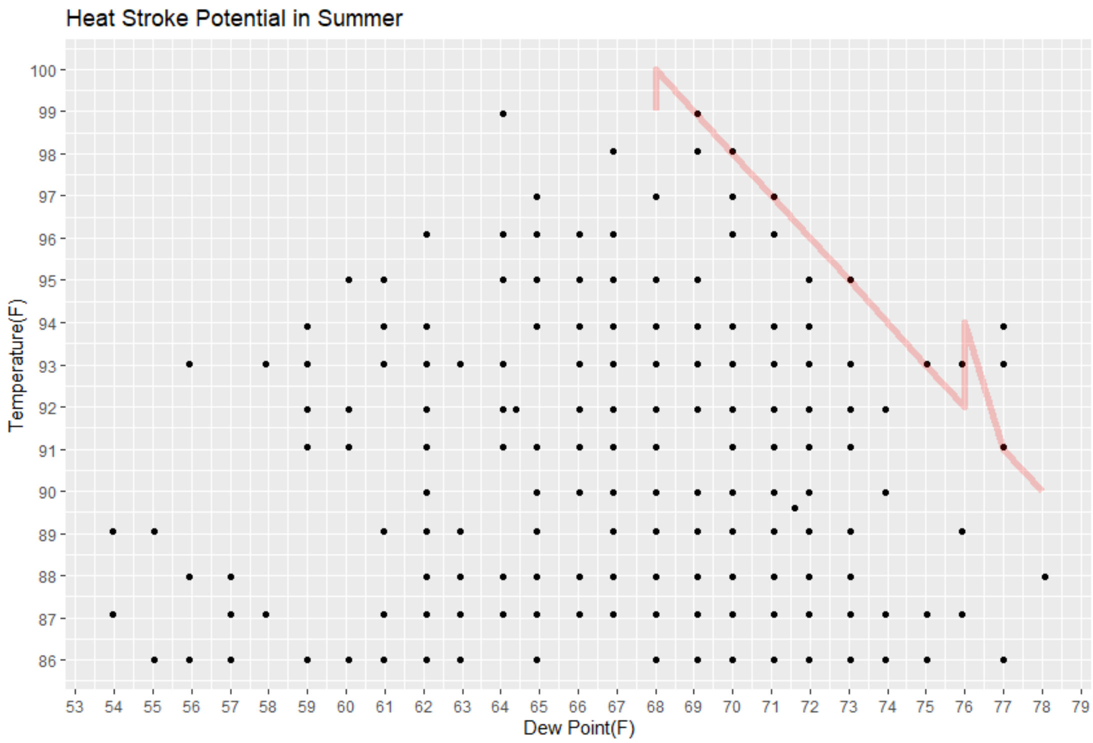

Figure 3.10.2 Plot diagram for Analysis 10

Chandler (2018) states that the higher the dew point, the muggier it will be and give uncomfortable feeling. Then, the combination of dew point and temperature is known as heat index. From this index, we can indicate the potential heatstroke. Based on the graph above, the red line shows the heatstroke potential and there are some points which are categorized as heatstroke potential. Therefore, in Summer 2013, **there are possibilities of heatstroke, hence it is better to drink plenty of cold drinks, take cool baths, wear light-colored clothing, and avoid the sun between 11am and 3pm** (NHS, 2018).

3.11. Analysis 11: Relationship between temperature and dew point

```
#-manipulation
#--check the correlation between temperature and dew point
cor(full_data$temp,full_data$dewp)
#-visualization
ggplot(full_data,aes(dewp,temp))+geom_jitter()+geom_smooth(method="lm")+
  labs(title="Correlation Between Temperature and Dew Point", y="Temperature(F)",x="Dew Point(F)")
```

Figure 3.11.1 Coding for Analysis 11

In the code above, it uses `cor()` to check whether both data have correlation or not. Cor = 1 indicates correlation. Then, the data will be represented using `geom_jitter()` with the help of `geom_smooth()`.

```
> cor(full_data$temp,full_data$dewp)
[1] 0.8959831
```

Figure 3.11.2 Correlation result

Figure 3.11.2 shows that 0.89 is not a small number, which means both data have correlation.

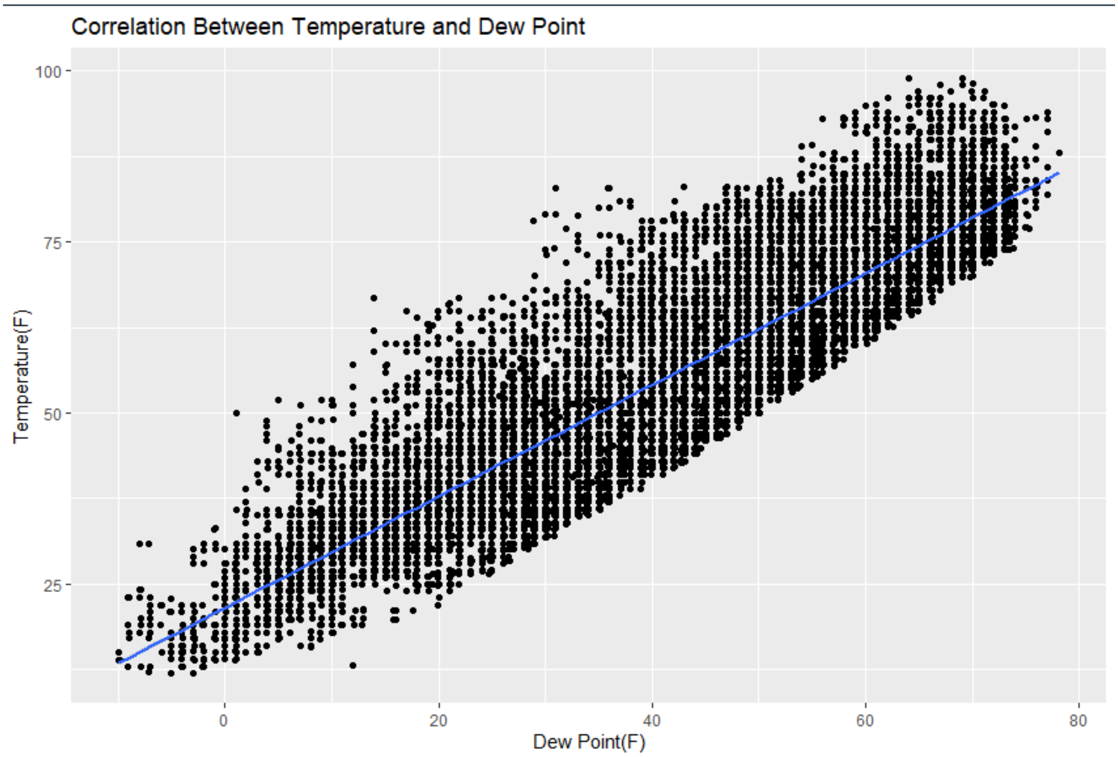

Figure 3.11.3 Diagram for Analysis 11

Diagram above shows that, the higher the dew point, the higher the temperature. NOAA (2020) defines that at constant pressure, the dew point is the temperature the air needs to be cooled to in order to achieve a 100% of relative humidity (RH). **Therefore, it is proven that dew point and temperature have correlation where the higher the dew point, the higher the temperature.**

3.12. Analysis 12: Analyzing precipitation and temperature in Winter

```
#-manipulation
winter=full_data%>%filter(season==4,precip>0)
#--to show the days
full_data%>%filter(season==4,precip>0,temp<=32)
#-visualization
ggplot(winter, aes(temp,precip))+geom_count(alpha=0.2,color="blue") + facet_wrap(~hour)+
  #--this rectangle area indicates snowfalls
  annotate("rect", xmin=15, xmax=32, ymin=0.0 , ymax=0.6, alpha=0.1, fill="cyan")+
  labs(title="Hourly Temperature and Precipitation",
       y="Precipitation(in)",x="Temperature(F)",subtitle = "Winter, 2013")
```

Figure 3.12.1 Coding for Analysis 12

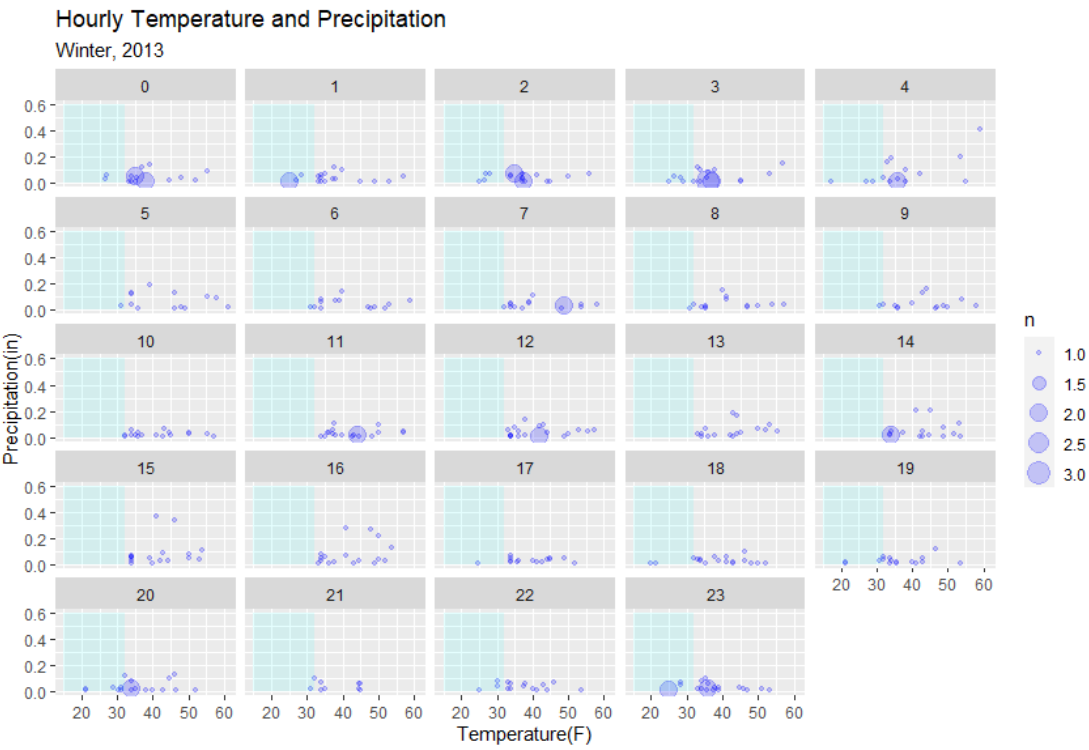

Figure 3.12.2 Diagram for Analysis 12

NSIDC (n.d) states that Snow forms when the atmospheric temperature is at or below freezing (32 degrees Fahrenheit) and precipitation is less than 0.6. Therefore, the diagram shows the rectangle that indicates the area of snowfall potential. **As can be seen, mostly, the snowfall at 12pm**.

3.13. Analysis 13: Analyzing Pressure

```
#-manipulation
#--show pressure summary for both airports
full_data%>%filter(origin=="JFK")%>%select(jfk_pressure=pressure)%>%summary()
full_data%>%filter(origin=="LGA")%>%select(lga_pressure=pressure)%>%summary()
#-visualization
ggplot(full_data,aes(pressure))+geom_histogram(bins=20,aes(group=origin,fill=origin),na.rm=T)+
  labs(title="Histogram for Pressure",
       y="count",x="Pressure",subtitle = "2013")
```

Figure 3.13.1 Coding for Analysis 13

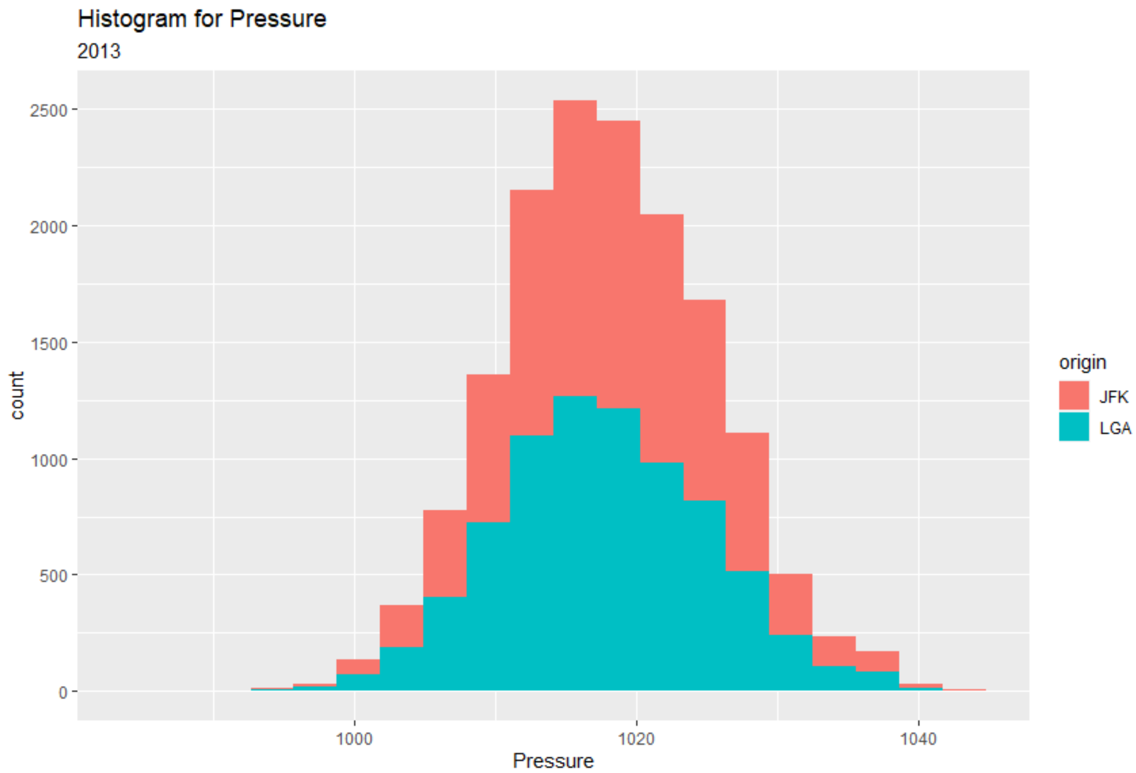

Figure 3.12.3 Diagram for Analysis 13

As can be seen in the diagram above, it shows that JFK has higher pressure than LGA. **This may affected by the elevation** where JFK elevation is 4m and LGA elevation is 6m.

3.14. Analysis 14: Analyzing Crosswind

```r
#-manipulation
#--LGA
lga_cwind=full_data%>%filter(origin=="LGA")%>%
  #convert mph into Knot & calculate crosswind for Runway 4
  mutate(knot_speed=wind_speed/1.151,crosswind=knot_speed*sin(wind_dir-45))%>%
  filter(crosswind>0)%>%select(origin,month,knot_speed,wind_dir,crosswind)
#--show average and maximum headwind at LGA
lga_cwind%>%summarise(cwind_avg=mean(crosswind,na.rm=T),cwind_max=max(crosswind,na.rm=T),.groups='drop')
#--JFK
jfk_cwind=full_data%>%filter(origin=="JFK")%>%
  #convert mph into Knot & calculate crosswind for Runway 13R
  mutate(knot_speed=wind_speed/1.151,crosswind=knot_speed*sin(wind_dir-133))%>%
  filter(crosswind>0)%>%select(origin,month,knot_speed,wind_dir,crosswind)
#--show average and maximum headwind at JFK
jfk_cwind%>%summarise(cwind_avg=mean(crosswind,na.rm=T),cwind_max=max(crosswind,na.rm=T),.groups='drop')

#-visualization
ggplot()+
  geom_jitter(data = jfk_cwind, aes(y=crosswind,x=month,color=origin)) +
  geom_jitter(data = lga_cwind, aes(y=crosswind,x=month,color=origin)) +
  scale_x_continuous(breaks=1:12,labels=c("jan","feb","mar","apr","may","jun",
                                          "jul","aug","sep","oct","nov","dec"))+
  scale_y_continuous(breaks=c(0,5,10,15,20,25,30,35)) +
  geom_hline(yintercept=30,color="darkred",size=2,alpha=0.2)+
  labs(title = "Jitterplot for Crosswind in 2013", subtitle="JFK: Runway 13R | LGA: Runway 4")
```

Figure 3.14.1 Coding for Analysis 14

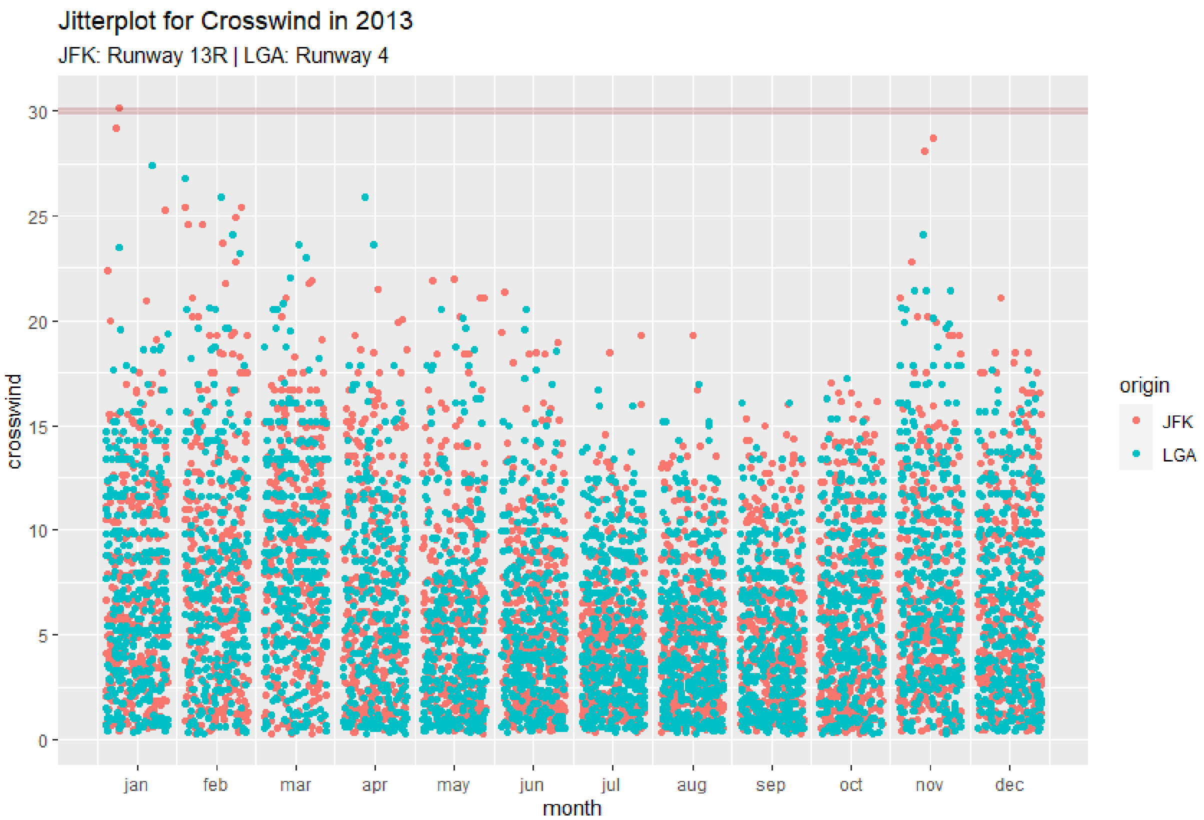

Figure 3.14.2 Diagram for Analysis 14

Referring to the explanation from Analysis 8, crosswind is a wind that blows from the right/left side of the aircraft. Hence, based on the analysis towards Runway 13R at JFK and Runway 4 at LGA, **it can be concluded that there is a point in January where JFK airport will have a delay due to strong crosswind.**

4. Extra Features

- scale_x_continuous() : change x values
- scale_y_continuous() : change y values
- geom_hline() and geom_vline(): make vertical and horizontal line
- annotate(): to make rectangle and point that can help making decisions

5. Conclusion

In conclusion, this analysis shows how important the temperature, dew point, wind speed, wind direction, and precipitation are. It can affect to many sectors and aspect. Therefore, by analyzing this hourly weather data, it can increase the awareness of weather situation. This project tells that 1 data set can bring to many decisions and analysis. It shows how important the data is. Lastly, by doing this project, I can understand more about data analytics, how to manipulate and how to visualize it properly.

6. References

Britannica (2020). *The North*. [Online]. Available from: https://www.britannica.com/place/the-North [Accessed: 18 November 2020]

Calendar Date (2020).*Spring 2013*. [Online]. Available from: https://www.calendardate.com/spring_2013.htm [Accessed: 18 November 2020]

Chandler, N. (2018). *What is Relative Humidity and How Does it Affect How I Feel Outside?*. [Online]. Available from: https://science.howstuffworks.com/nature/climate-weather/atmospheric/question651.htm [Accessed: 21 November 2020]

Elaine, K. (2018). *Hot and Cold: Extreme Temperature Safety*. [Online]. Available from: https://www.healthline.com/health/extreme-temperature-safety#extreme-heat-temperatures [Accessed: 19 November 2020]

Finavia (2017). *How do wind conditions affect flight?*. [Online]. Available from: https://www.finavia.fi/en/newsroom/2017/how-do-wind-conditions-affect-flight [Accessed: 20 November 2020]

Forbes (2015). *Big Data: 20 Mind-Boggling Facts Everyone Must Read*. [Online]. Available from: https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/?sh=177dc07e17b1 [Accessed: 18 November 2020]

Import.io (2018). *Why is data, and why is it important.* [Online]. Available from: https://www.import.io/post/what-is-data-and-why-is-it-important/ [Accessed: 17 November 2020]

IVAO (2020). *Crosswind and Headwind Calculation*. [Online]. Available from: https://mediawiki.ivao.aero/index.php?title=Crosswind_and_Headwind_calculation [Accessed: 20 November 2020] iWEATHERNET (2016). *Heat Index Calculator & Charts*. [Online]. Available from: https://www.iweathernet.com/educational/heat-index-calculator-and-conversion-table [Accessed: 21 November 2020]

NHS (2018). *Heat exhaustion and Heatstroke*. [Online]. Available from: https://www.nhs.uk/conditions/heat-exhaustion-heatstroke/ [Accessed: 21 November 2020]

NOAA (2015). *A Discussion of Water Vapor, Humidity, and Dewpoint, and Relationship to Precipitation*. [Online]. Available from https://www.weather.gov/lmk/humidity [Accessed: 19 November 2020]

NOAA (2020). *Heat Index*. [Online]. Available from: https://www.weather.gov/arx/heat_index [Accessed: 21 November 2020]

NSIDC (n.d). *How Snow Forms*. [Online]. Available from: https://nsidc.org/cryosphere/snow/science/formation.html [Accessed: 22 November 2020]

ORI (2016). *Data Analysis.* [Online]. Available from: https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/datopic.html [Accessed: 18 November 2020]

PhysLink (n.d).*How does humidity effect the way that an airplane flies?*. [Online]. Available from: https://www.physlink.com/education/askexperts/ae652.cfm [Accessed: 19 November 2020]

Schrader, R. (2019). *What Wind Sped Delays Flights?*. [Online]. Available from: https://www.skyscanner.com/tips-and-inspiration/what-windspeed-delays-flights [Accessed: 22 November 2020]

Scientific American (2009). *When the air is the same temperature as our body, why do we feel hot?*. [Online]. Available from: https://www.scientificamerican.com/article/why-people-feel-hot/ [Accessed: 19 November 2020]

Smart Fog (2016).*How Rain and Humidity Connected?*. [Online]. Available from: https://www.smartfog.com/how-rain-and-humidity-connected.html [Accessed: 19 November 2020]

Weill Cornell Medicine (n.d). *Transportation in New York City*. Available from: https://postdocs.weill.cornell.edu/navigating-wcm-and-nyc/upon-arrival/transportation [Accessed: 19 November 2020]

Wenz, J. (2019). *What is Wind Chill, and How Does it Affect the Human Body*. [Online]. Available from: https://www.smithsonianmag.com/science-nature/what-wind-chill-and-how-does-it-affect-human-body-180971376/ [Accessed: 19 November 2020]

WHO (2018).*WHO Housing and Health Guidelines*. [Online]. Available from: https://apps.who.int/iris/bitstream/handle/10665/277465/WHO-CED-PHE-18.10-eng.pdf [Accessed: 20 November 2020]

Wikimedia (n.d). *LGA airport map.* [Online]. Available from: https://commons.wikimedia.org/wiki/File:LGA_airport_map.svg [Accessed: 20 November 2020]