Data Analysis Project

Student Information

Name: Sai Jaswanth Kumar Kunku
G Number : G01331645

Library and Data Import

```
library(tidyverse)
library(lubridate)
library(ggcorrplot)
library(reshape2)
library(gridExtra)
flights_df <- readRDS(url('https://gmubusinessanalytics.netlify.app/data/dulles_flights.rds'))
```

Raw Data

```
flights_df<-flights_df %>%
   mutate(month = factor(month.name[month], levels = month.name)) %>%
   arrange(month)
head(flights_df,5)
```

```
# A tibble: 5 x 22
  scheduled_fligh~ month_numeric month   day weekday airline tail_num flight_num
  <date>                   <dbl> <fct> <dbl> <fct>   <fct>   <fct>         <dbl>
1 2016-01-01                   1 Janu~     1 Friday  Southw~ N569WN          107
2 2016-01-01                   1 Janu~     1 Friday  Southw~ N466WN          449
3 2016-01-01                   1 Janu~     1 Friday  Southw~ N458WN         1502
4 2016-01-01                   1 Janu~     1 Friday  Southw~ N922WN         2388
5 2016-01-01                   1 Janu~     1 Friday  Southw~ N711HK         1396
# ... with 14 more variables: dest_airport_name <fct>, dest_airport_city <fct>,
#   dest_airport_state <fct>, dest_airport_region <fct>, sch_dep_time <dbl>,
#   dep_time <dbl>, dep_delay <dbl>, taxi_out <dbl>, wheels_on <dbl>,
#   taxi_in <dbl>, arrival_time <dbl>, sch_arrival_time <dbl>,
#   arrival_delay <dbl>, distance <dbl>
```

```
summary(flights_df)
```

```
 scheduled_flight_date month_numeric         month            day
 Min.   :2016-01-01    Min.   : 1.000   July     : 3120   Min.   : 1.00
 1st Qu.:2016-04-16    1st Qu.: 4.000   August   : 3092   1st Qu.: 8.00
 Median :2016-07-12    Median : 7.000   June     : 3071   Median :16.00
 Mean   :2016-07-08    Mean   : 6.756   October  : 3051   Mean   :15.72
 3rd Qu.:2016-10-04    3rd Qu.:10.000   May      : 2887   3rd Qu.:23.00
 Max.   :2016-12-31    Max.   :12.000   November: 2885   Max.   :31.00
                                        (Other) :15327
```

```
     weekday                airline         tail_num       flight_num  
 Sunday   :4710   United         :20653   N63890 :  125   Min.   :  10  
 Monday   :4914   American       : 2597   N69806 :  113   1st Qu.: 365  
 Tuesday  :4917   Delta          : 2565   N66828 :  112   Median : 685  
 Wednesday:4973   Southwest      : 2161   N66841 :  109   Mean   :1050  
 Thursday :4993   JetBlue        : 2013   N68811 :  108   3rd Qu.:1544  
 Friday   :5015   Virgin America: 1613   N62894 :  106   Max.   :6840  
 Saturday :3911   (Other)        : 1831   (Other):32760                 
                  dest_airport_name     dest_airport_city
 San Francisco            : 4034   San Francisco: 4034  
 Los Angeles              : 3846   Los Angeles  : 3846  
 Denver                   : 3628   Denver       : 3628  
 Hartsfield-Jackson Atlanta: 3154   Atlanta      : 3154  
 Logan                    : 2170   Boston       : 2170  
 Orlando                  : 1805   Chicago      : 1900  
 (Other)                  :14796   (Other)      :14701  
    dest_airport_state      dest_airport_region  sch_dep_time  
 California   :9177     West           :15555    Min.   : 5.33  
 Colorado     :3657     South          : 7752    1st Qu.: 8.83  
 Florida      :3511     Northeast      : 4002    Median :14.90  
 Georgia      :3154     Midwest        : 3040    Mean   :14.00  
 Texas        :2641     Southwest      : 3001    3rd Qu.:17.75  
 Massachusetts:2170     Middle Atlantic:   83    Max.   :22.95  
 (Other)      :9123                                             
    dep_time       dep_delay         taxi_out        wheels_on    
 Min.   : 0.02   Min.   : -25.00   Min.   :  1.00   Min.   : 0.02  
 1st Qu.: 8.87   1st Qu.:  -5.00   1st Qu.: 11.00   1st Qu.:10.53  
 Median :14.92   Median :  -2.00   Median : 14.00   Median :15.08  
 Mean   :14.08   Mean   :   9.07   Mean   : 16.95   Mean   :14.87  
 3rd Qu.:17.98   3rd Qu.:   3.00   3rd Qu.: 18.00   3rd Qu.:19.85  
 Max.   :24.00   Max.   :1244.00   Max.   :159.00   Max.   :24.00  

    taxi_in         arrival_time   sch_arrival_time arrival_delay     
 Min.   :  1.000   Min.   : 0.02   Min.   : 0.02    Min.   : -94.0000  
 1st Qu.:  5.000   1st Qu.:10.60   1st Qu.:10.77    1st Qu.: -20.0000  
 Median :  7.000   Median :15.12   Median :15.28    Median : -11.0000  
 Mean   :  8.872   Mean   :14.89   Mean   :14.94    Mean   :  -0.5486  
 3rd Qu.: 10.000   3rd Qu.:19.97   3rd Qu.:20.00    3rd Qu.:   2.0000  
 Max.   :178.000   Max.   :24.00   Max.   :23.98    Max.   :1228.0000  

    distance   
 Min.   : 157  
 1st Qu.: 534  
 Median :1190  
 Mean   :1355  
 3rd Qu.:2288  
 Max.   :4817  
```

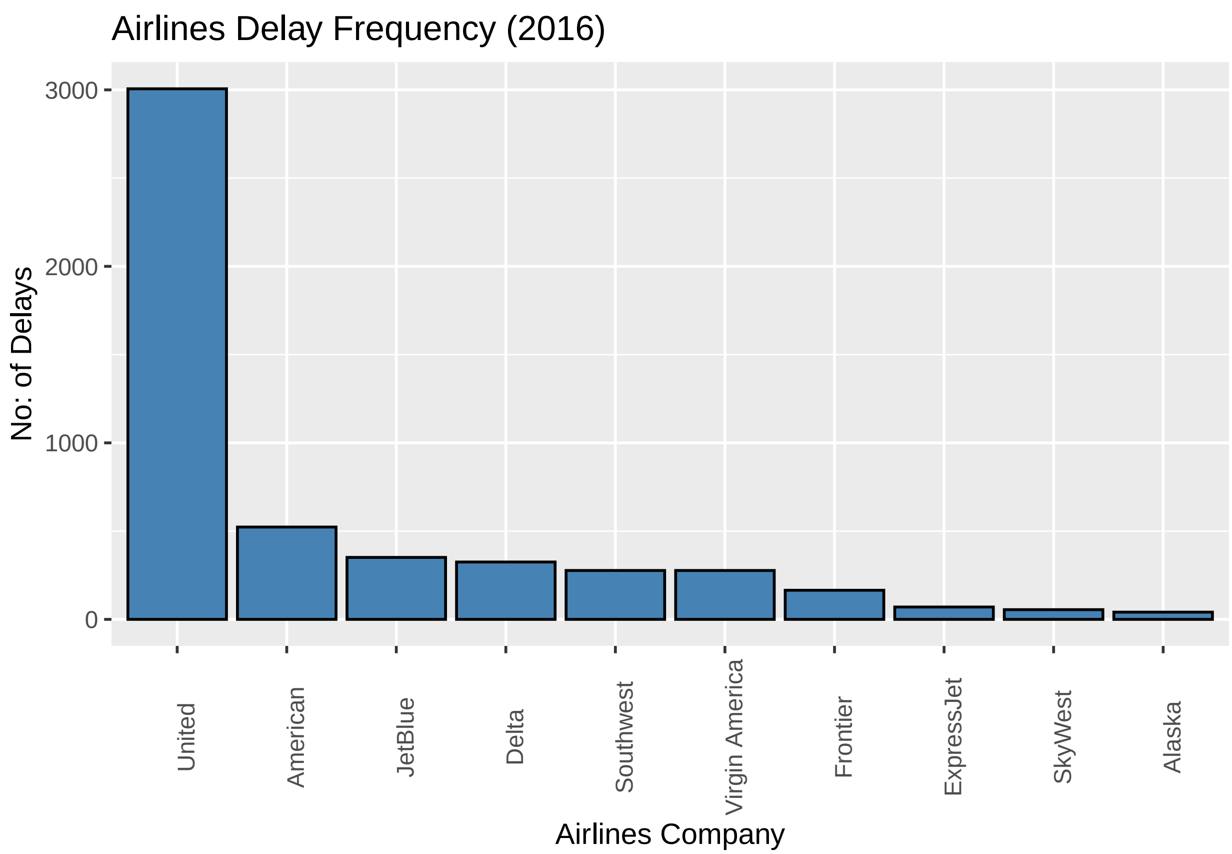
Question 1: Are certain destinations or airlines prone to delays?

Answer:United airlines are highly prone to delays with a delays more than 3000 followed by American and Jetblue with less than 500 delays with majority of them occurring in South and West regions .In addition to that, airports located in west region cities like Los Angeles,Francisco and Denver are experiencing around 700 delays followed by cities in northeast and South regions close to 400 delays.

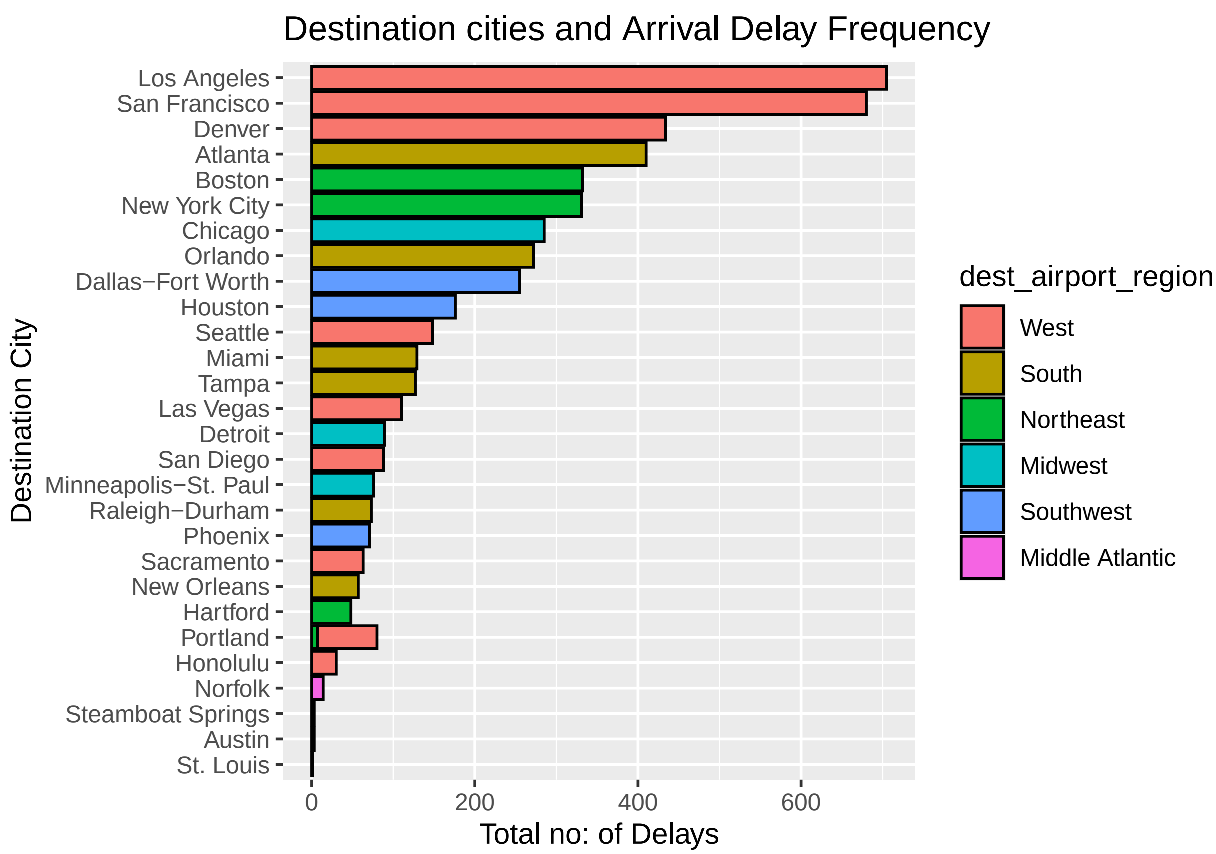
```r
#Total Arrival Delays By Airlines
delay_summary<-flights_df %>% filter(arrival_delay > 15) %>%
  group_by(airline,dest_airport_region) %>% summarise(total_delays=n()) %>% mutate(percent_of_delays=rou

delay_summary
```

```
# A tibble: 24 x 4
# Groups:   airline [10]
   airline  dest_airport_region total_delays percent_of_delays
   <fct>    <fct>                      <int>             <dbl>
 1 United   West                        1639             54.5
 2 United   South                        444             14.8
 3 United   Northeast                    366             12.2
 4 United   Southwest                    282              9.38
 5 United   Midwest                      261              8.68
 6 United   Middle Atlantic               14              0.47
 7 American Southwest                    223             42.6
 8 American West                         172             32.9
 9 American South                        128             24.5
10 Delta    South                        283             87.1
# ... with 14 more rows
```

```r
flights_df %>% filter(arrival_delay > 15) %>%
  count(airline, sort = TRUE, name = 'total_delays') %>% ggplot(aes(x=reorder(airline,-total_delays,),y
  geom_bar(fill="steelblue",stat="identity",color="black") +
  labs(title = "Airlines Delay Frequency (2016)",
        y = "No: of Delays",
        x = "Airlines Company")+theme(axis.text.x = element_text(angle = 90))
```



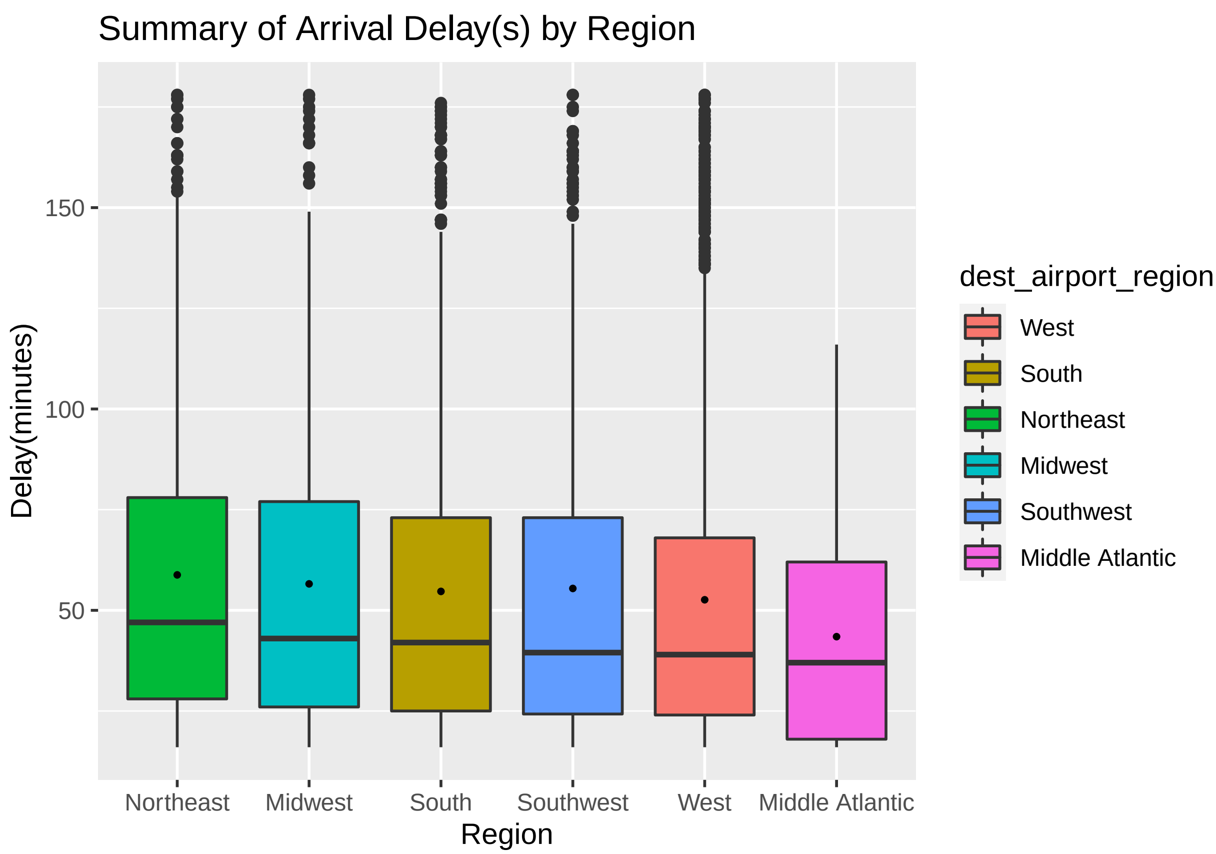

```
#Total Delays by Destination city
tot_dest_delays<-flights_df %>%
  select(dest_airport_region,arrival_delay,dest_airport_city) %>%
  filter(arrival_delay > 15) %>% group_by(dest_airport_region) %>%
  count(dest_airport_city, sort = TRUE, name='total_delays') %>%
  arrange(desc(total_delays))

ggplot(tot_dest_delays,aes(x=reorder(dest_airport_city, total_delays),y=total_delays,fill=dest_airport_r
  geom_bar(stat="identity",color="black") +coord_flip() +
  labs(title = "Destination cities and Arrival Delay Frequency",
             x = "Destination City",
             y = "Total no: of Delays")
```



Question 2 : How the delay times are varying in different regions and which airports are experiencing more arrival delay times in each region?

Answer: The median arrival delay time is highest in Northeast region(~48minutes) and lowest in Middle Atlantic region(~37minutes).The top 3 airports exhibiting higher delay times with respect to each region are Daniel K Inouye,McCarran ,Seattle-Tacoma ,Orlando ,Tampa,Raleigh-Durham,Bradley ,Logan ,Newark Liberty ,Chicago OHare,Minneapolis-St Paul,Detroit Metro Wayne County,Dallas-Fort Worth ,Phoenix Sky Harbor, George Bush Intercontinental,Norfolk.

```
#Removal of Outliers
  x<-flights_df[flights_df$arrival_delay>15,]
  outliers <- boxplot(x$arrival_delay, plot=FALSE)$out
  x<- as.data.frame(x[-which(x$arrival_delay %in% outliers),])

#Plot
  ggplot(x,aes(x=reorder(dest_airport_region, -arrival_delay, FUN = median),
                         y=arrival_delay,
                         fill=dest_airport_region)) +
geom_boxplot()+ stat_summary(fun=mean, colour="black", geom="point",shape=16,
                           size=1, show.legend=FALSE) +
    labs(title = "Summary of Arrival Delay(s) by Region", y = "Delay(minutes)",
         x = "Region")
```




```r
#Top 3 airports with higher delay times in each region
x %>%
select(dest_airport_region,arrival_delay,dest_airport_name) %>%
filter(arrival_delay > 15) %>%
group_by(dest_airport_region,dest_airport_name) %>%
summarise(arrival_delay_time=median(arrival_delay),total_arrival_delays=n()) %>%
filter(total_arrival_delays >10) %>% top_n(3, `arrival_delay_time`) %>%
arrange(dest_airport_region,desc(arrival_delay_time))
```

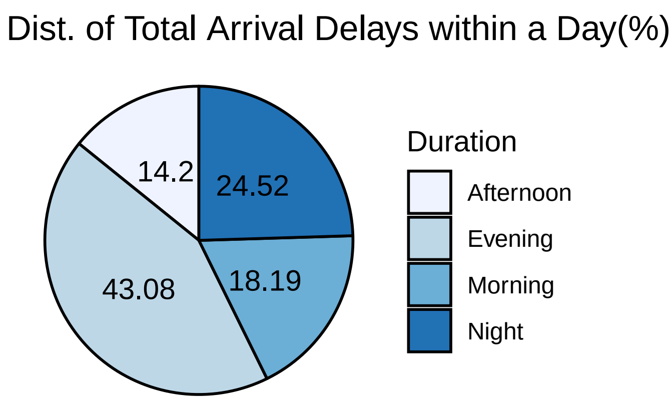
```
# A tibble: 16 x 4
# Groups:   dest_airport_region [6]
   dest_airport_region dest_airport_name     arrival_delay_ti~ total_arrival_de~
   <fct>               <fct>                             <dbl>             <int>
 1 West                Daniel K Inouye                    51                  25
 2 West                McCarran                           47                 101
 3 West                Seattle-Tacoma                     45.5               138
 4 South               Orlando                            49                 247
 5 South               Tampa                              47.5               118
 6 South               Raleigh-Durham                     45                  69
 7 Northeast           Bradley                            55                  45
 8 Northeast           Logan                              49                 302
 9 Northeast           Newark Liberty                     48                 131
10 Midwest             Chicago OHare                      46                 237
11 Midwest             Minneapolis-St Paul                44                  66
12 Midwest             Detroit Metro Wayne ~              40.5                76
13 Southwest           Dallas-Fort Worth                  41                 229
```

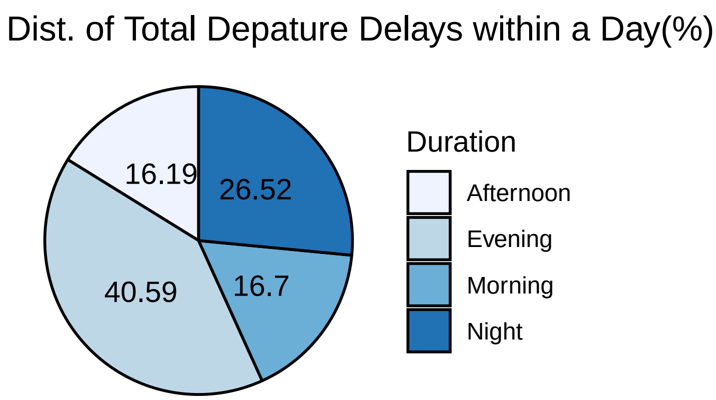
```
14 Southwest            Phoenix Sky Harbor                 38              65
15 Southwest            George Bush Intercon~              37             153
16 Middle Atlantic      Norfolk                            37              13
```

Question 3 : Are certain times of the day or year problematic?

Answer: Operations happening in the Evening(after 7pm) and Night(before 5am) seems to be problematic as they are accounting for 68.3% of total arrival delays as well as 67.11% of departure delays respectively.While the departure delays tend to be constant over the months,the arrival delay is high in the months from June-August and December.

```
flights_df <- flights_df %>%
  mutate(time_period = case_when(
    sch_arrival_time >= 5 & sch_arrival_time <12 ~ 'Morning',
    sch_arrival_time >= 12  & sch_arrival_time <17 ~ 'Afternoon',
    sch_arrival_time >= 17 & sch_arrival_time <21 ~ 'Evening',
    TRUE ~ 'Night'))

dt1<-flights_df %>% group_by(time_period) %>% filter(arrival_delay > 15) %>%
  summarize(total_delays=n()) %>%  ungroup() %>%
  mutate(percent_of_delays =  round(100*(total_delays/sum(total_delays)),2))

dt2<-flights_df %>% group_by(time_period) %>% filter(dep_delay > 15) %>%
  summarize(total_delays=n()) %>%  ungroup() %>%
  mutate(percent_of_delays = round(100*(total_delays/sum(total_delays)),2))

fig1<- ggplot(dt1, aes(x = "", y = percent_of_delays, fill = time_period)) +
  geom_col(color = "black") +guides(fill=guide_legend(title="Duration"))+
  geom_text(aes(label = percent_of_delays),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y")+theme_void() +  scale_fill_brewer() +
  labs(title = "Dist. of Total Arrival Delays within a Day(%)")

fig2<- ggplot(dt2, aes(x = 2, y = percent_of_delays, fill = time_period)) +
  geom_col(color = "black") +guides(fill=guide_legend(title="Duration"))+
  geom_text(aes(label = percent_of_delays),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y")+theme_void() +  scale_fill_brewer() +
  labs(title = "Dist. of Total Depature Delays within a Day(%)")

grid.arrange(fig1, fig2, nrow = 2)
```






```
monthly_dep_delay_percentage<-flights_df %>% group_by(month) %>%
summarize(total_dep_delays=n()) %>%  ungroup() %>%
mutate(percent_of_delays = round(100*(total_dep_delays/sum(total_dep_delays)),2)) %>% arrange(desc(perce

monthly_dep_delay_percentage
```

```
# A tibble: 12 x 3
   month     total_dep_delays percent_of_delays
   <fct>                <int>             <dbl>
 1 July                  3120              9.33
 2 August                3092              9.25
 3 June                  3071              9.19
 4 October               3051              9.13
 5 May                   2887              8.64
 6 November              2885              8.63
 7 September             2877              8.61
 8 December              2768              8.28
 9 April                 2700              8.08
10 March                 2554              7.64
11 January               2245              6.71
12 February              2183              6.53
```

```
monthly_arr_delay_percentage<-flights_df %>% group_by(month) %>%
filter(arrival_delay > 15) %>% summarize(total_arr_delays=n()) %>%
ungroup() %>%
mutate(percent_of_delays = round(100*(total_arr_delays/sum(total_arr_delays)),2)) %>% arrange(desc(perce
```

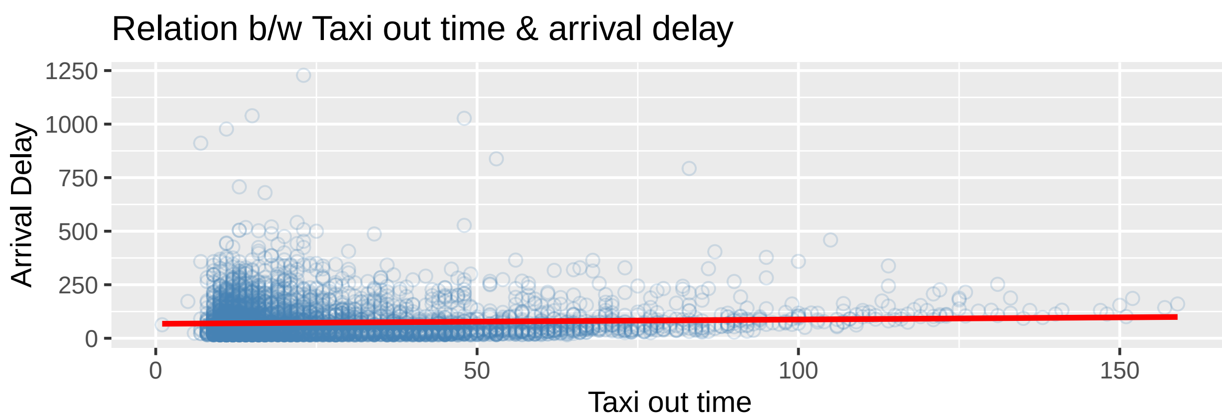
```
monthly_arr_delay_percentage
```

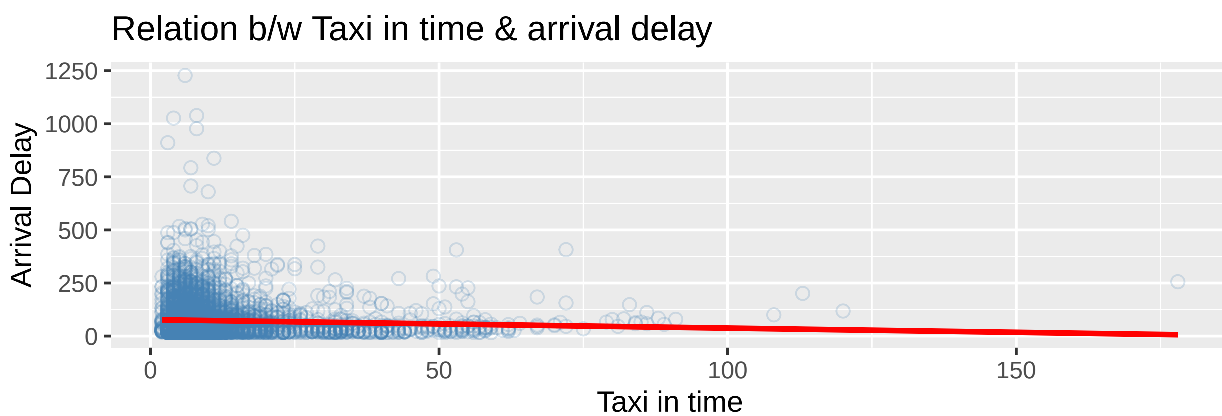
```
# A tibble: 12 x 3
   month     total_arr_delays percent_of_delays
   <fct>                <int>             <dbl>
 1 July                   729             14.3
 2 June                   665             13.1
 3 December               640             12.6
 4 August                 529             10.4
 5 October                393              7.72
 6 May                    385              7.56
 7 September              373              7.33
 8 March                  316              6.21
 9 April                  315              6.19
10 February               264              5.19
11 November               256              5.03
12 January                225              4.42
```

Question 4:Are flight delays affected by taxi-out and taxi-in time?

Answer: Based on the results from the scatter plot it is evident that flight delays are not affected by taxi in/out times.

```
dt2<-flights_df%>% select(taxi_out,arrival_delay) %>% filter(arrival_delay>15)
fig1<-ggplot(dt2,aes(x=taxi_out,y=arrival_delay)) +
  geom_point(size=2, shape=21,color="steelblue",alpha = 0.2) +
  labs(title = "Relation b/w Taxi out time & arrival delay ",
           y = "Arrival Delay",
           x = "Taxi out time") +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE,color="red")

dt1<-flights_df%>% select(taxi_in,arrival_delay) %>% filter(arrival_delay>15)
fig2<-ggplot(dt1,aes(x=taxi_in,y=arrival_delay)) +
  geom_point(size=2, shape=21,color="steelblue",alpha = 0.2) +
  labs(title = "Relation b/w Taxi in time & arrival delay ",
           y = "Arrival Delay",
           x = "Taxi in time") +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE,color="red")


grid.arrange(fig1, fig2, nrow = 2)
```

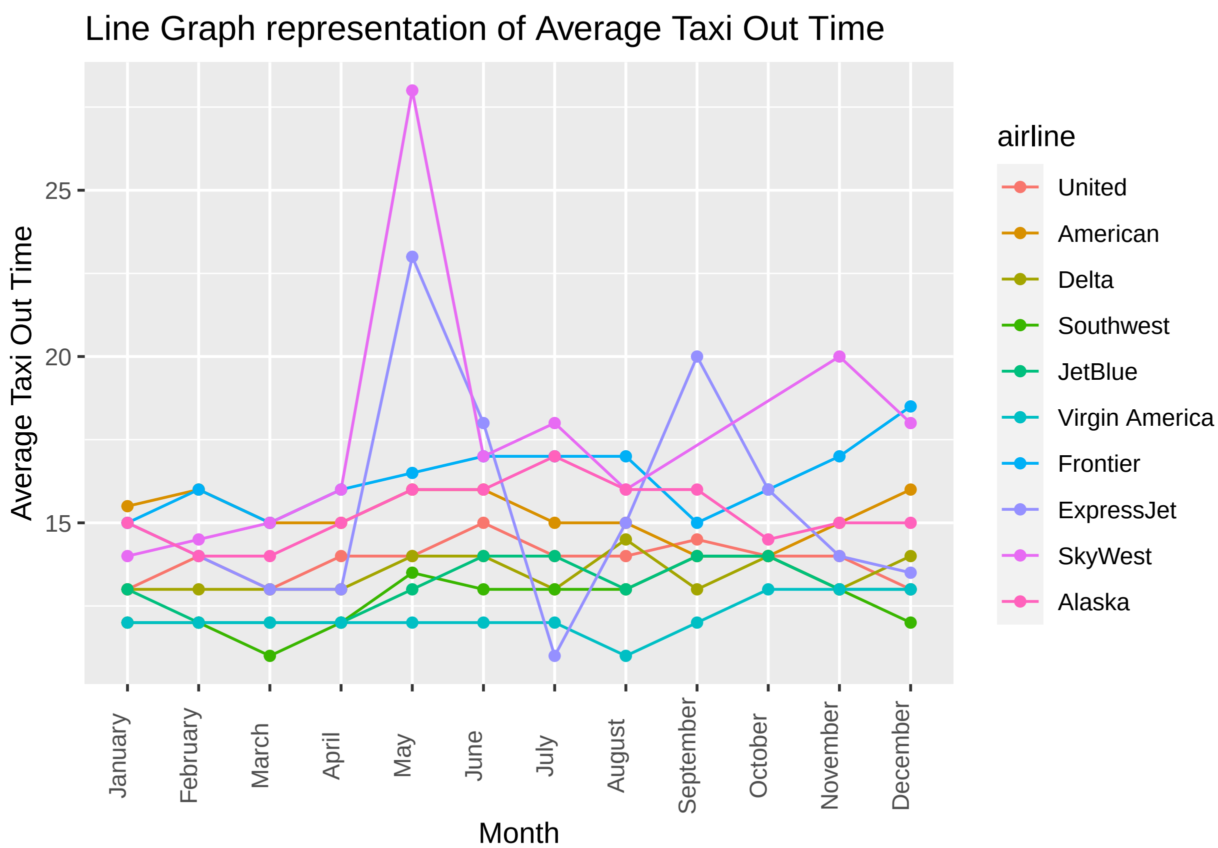




Question 5: Do certain airlines or time of year lead to greater taxi out times (i.e. traffic jams on the runways)?

Answer: Alaska airlines is showing higher taxi out between January and July whereas Skywest is experiencing increased taxi out time from January before reaching a peak time of 30mins in May.While,Frontier is showing greater taxi out times from September; Express Jet is showing higher taxi out times in the month of May with a peak time of 25mins before increasing again in September.

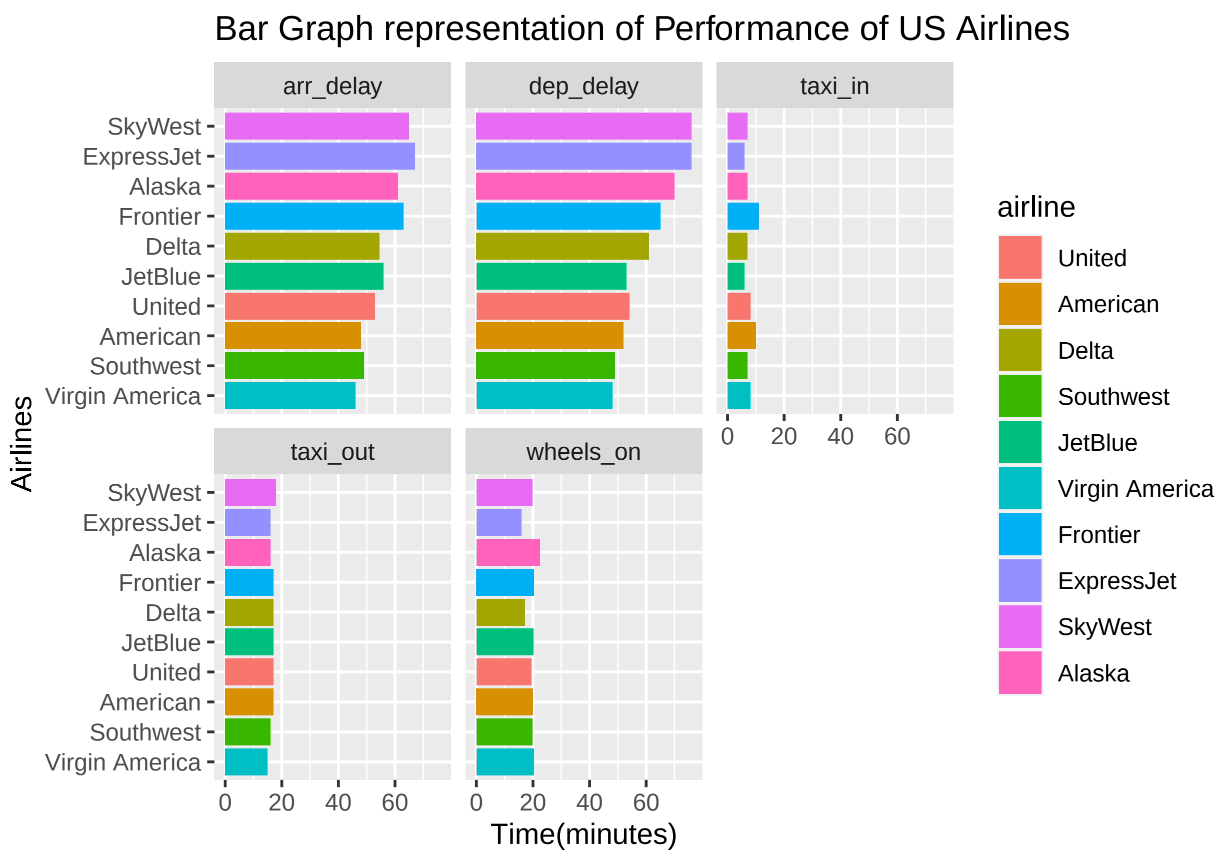
```
flights_df %>% group_by(airline,month) %>%
  summarize(avg_taxi_out=median(taxi_out)) %>%
  ggplot(aes(x=month,y=avg_taxi_out,color =airline,group=airline)) +
  geom_line() +geom_point()+
  labs(title = "Line Graph representation of Average Taxi Out Time",
      y = "Average Taxi Out Time",
      x = "Month") +theme(axis.text.x = element_text(angle = 90, vjust = -0.1))
```



Question 6: How various airlines are performing on the runway and with respect to delay times?

Answer: During taxi-in American and Frontier are experiencing higher delay times.In terms of delay in arrivals and departures Alaska, Skywest,Express Jet and Frontier are highest when compared to other airlines.

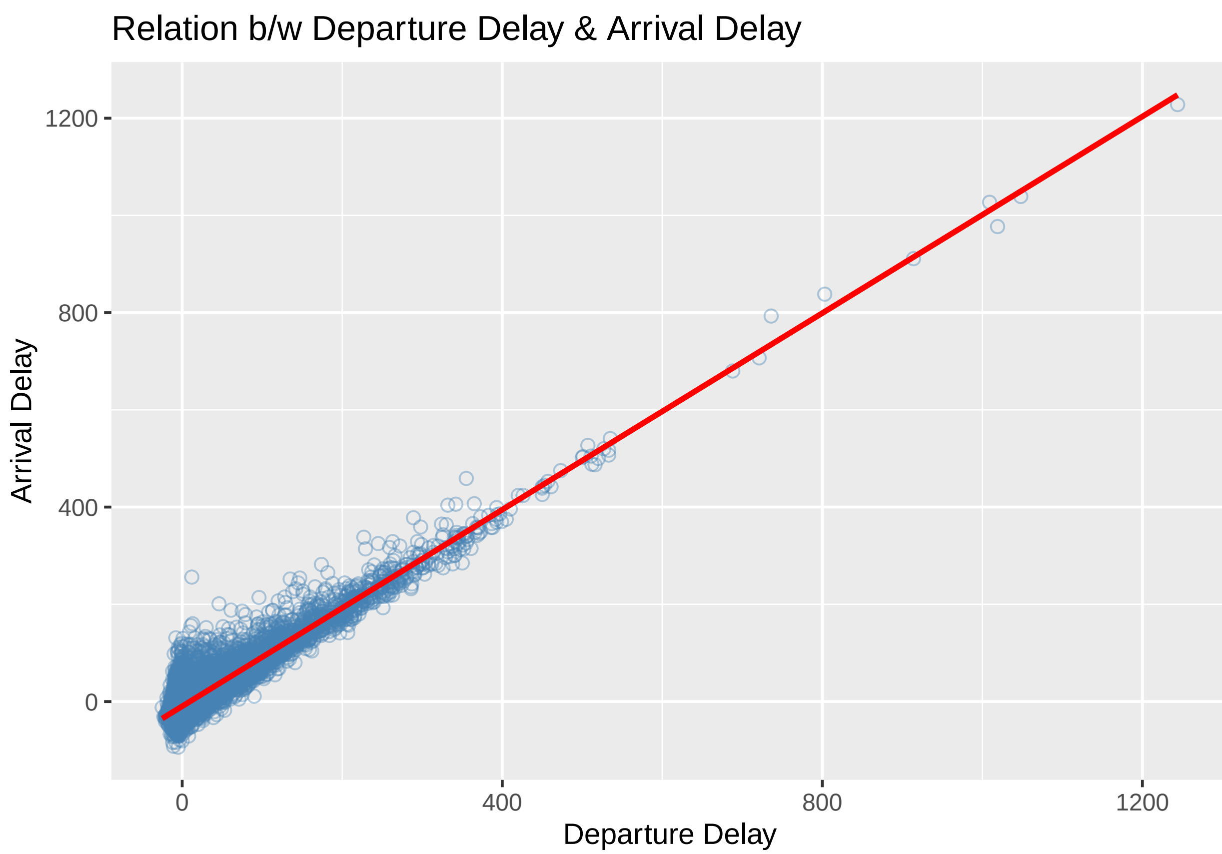
```
flights_df %>% group_by(airline) %>% filter(dep_delay >0 & arrival_delay>15) %>%
  summarize(arr_delay=median(arrival_delay),
                      dep_delay=median(dep_delay),
                      taxi_in=median(taxi_in),
                    wheels_on=median(wheels_on),
                      taxi_out=median(taxi_out)) %>%
                      arrange(desc(taxi_out)) %>%
  pivot_longer(cols = c(arr_delay, dep_delay,taxi_in,taxi_out,wheels_on),
            names_to = 'value_type',
            values_to = 'total_count') %>%
  ggplot(aes(x=reorder(airline,total_count),y=total_count,fill =airline)) +
  geom_bar(stat="identity",position = "dodge") + facet_wrap(~value_type) +
  coord_flip() +
  labs(title = "Bar Graph representation of Performance of US Airlines",
            x = "Airlines",
            y = "Time(minutes)")
```



Question 7: Does delay in departures cause arrival delay in destination?

Answer: Yes,from the scatter plot it can be observed that the arrival delay is increasing linearly with respect to the increase in departure delay.

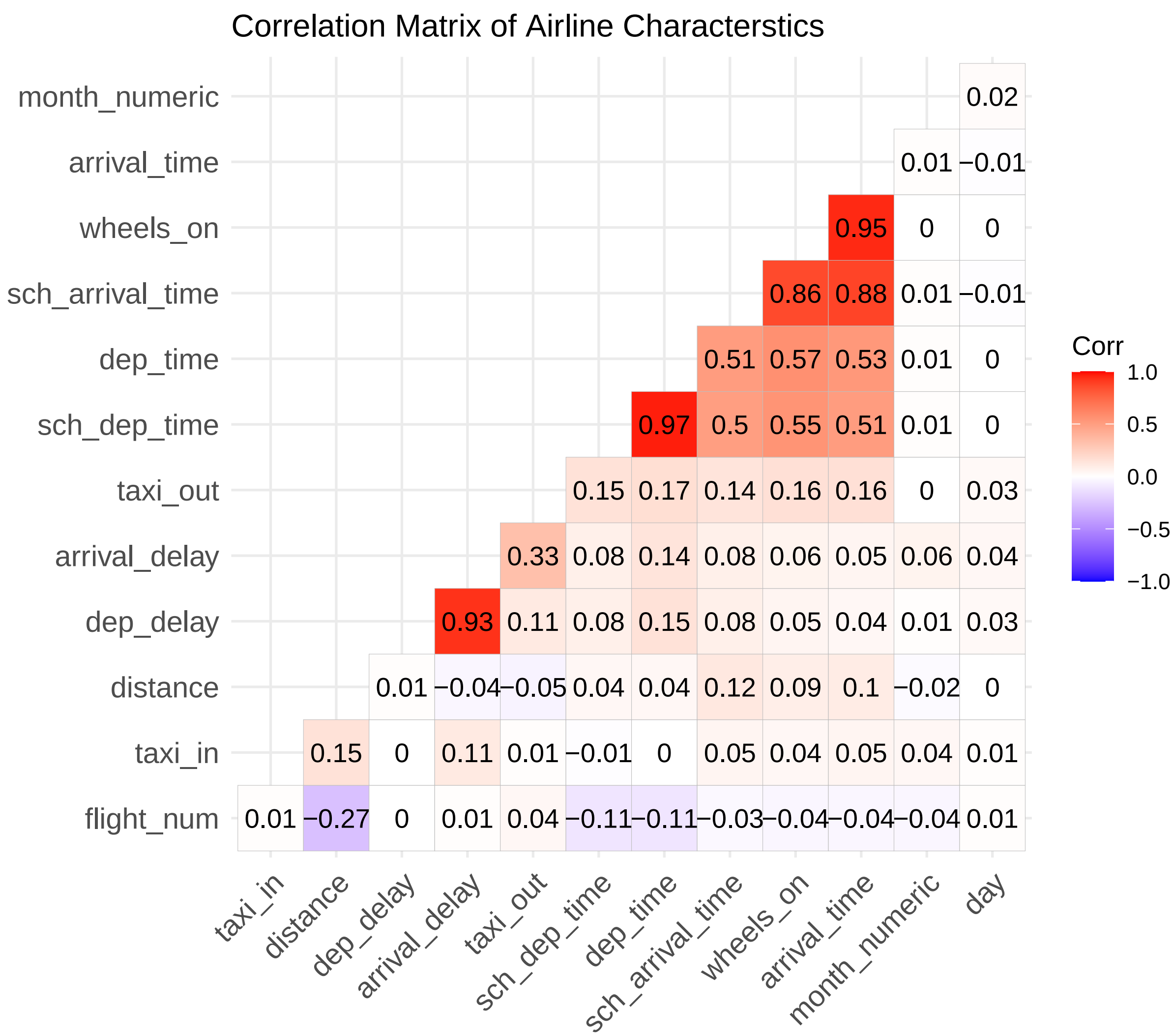
```
ggplot(flights_df,aes(x=dep_delay,y=arrival_delay)) +
  geom_point(size=2, shape=21,color="steelblue",alpha = 0.4) +
  labs(title = "Relation b/w Departure Delay & Arrival Delay ",
            x = "Departure Delay",
            y = "Arrival Delay") +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE,color="red")
```



Question 8:How are the airline characterstics related to each other**

Answer:The heat map explains that the wheels on time is highly correlated with the arrival time and moderately correlated with the departure time.It also explains that departure delay is positively correlated with arrival delay.It is interesting to note that delays are not affected by distance variable.

```
numeric_df<-Filter(is.numeric, flights_df)
ggcorrplot(cor(numeric_df), hc.order = TRUE, type = "lower",
           lab = TRUE,digits = 2)+
  labs(title = "Correlation Matrix of Airline Characterstics")
```



Question 9 :What are the taxi in/taxi-out times for the flights that arrived early over the flights that are delayed? **Answer**:Airlines that are delayed are experiencing additional taxi out,wheels on time of 5 to 10minutes when compared to the flights that arrived early.For taxi in time there has been any major difference expect for Jet and Frontier airlines which have taxi in time greater than 10minutes.

```
flights_df %>%
  mutate(arrival_state = case_when(
    arrival_delay <=15 ~ 'Early/Ontime',
    TRUE ~ 'Delayed')) %>% group_by(airline,arrival_state) %>%
  summarise(taxi_out_time=median(taxi_out),taxi_in_time=median(taxi_in),wheels_on=median(wheels_on)) %>
  arrange(arrival_state)
```

```
# A tibble: 20 x 5
# Groups:   airline [10]
   airline        arrival_state taxi_out_time taxi_in_time wheels_on
   <fct>          <chr>                 <dbl>        <dbl>     <dbl>
 1 United         Delayed                  18            8      19.3
 2 American       Delayed                  19           11      20.0
 3 Delta          Delayed                  19            7      16.8
 4 Southwest      Delayed                  20            7      19.8
 5 JetBlue        Delayed                  19            6      20.2
```

```
 6 Virgin America Delayed                    15          8       20.3
 7 Frontier       Delayed                    18         12       20.8
 8 ExpressJet     Delayed                    17          6       16.4
 9 SkyWest        Delayed                    23          7       19.7
10 Alaska         Delayed                    17          7       22.4
11 United         Early/Ontime               14          7       14.9
12 American       Early/Ontime               14         10       13.5
13 Delta          Early/Ontime               13          6       13.0
14 Southwest      Early/Ontime               12          7       14.6
15 JetBlue        Early/Ontime               13          6       12.1
16 Virgin America Early/Ontime               12          8       12.6
17 Frontier       Early/Ontime               16         12       18.9
18 ExpressJet     Early/Ontime               15          6       14.1
19 SkyWest        Early/Ontime               16          7       13.6
20 Alaska         Early/Ontime               15          7       21.3
```

Bonus Question:Does weather have any impact on delay of airlines**

Answer:Yes ,weather has a great impact on the airline delays and it is evident from the increasing delay,arrival and taxi out between July and November when temperatures are falling down.On the other hand from January,when the temperature are slowly rising, the delay rates are decreasing gradually.

```r
temp_df<-read.csv("temperature.csv")
#Parsing Dates
flights_df$scheduled_flight_date<- ymd(flights_df$scheduled_flight_date)
temp_df$Date<- ymd(temp_df$Date)
temp_df$Average<-(temp_df$Average-32)*5/9
#Joining Temperature dataset by Date column
final_df<-left_join(flights_df, temp_df,
                    by = c("scheduled_flight_date" = "Date"))
#Checking for NA rows after joining
sum(is.na(final_df))
```

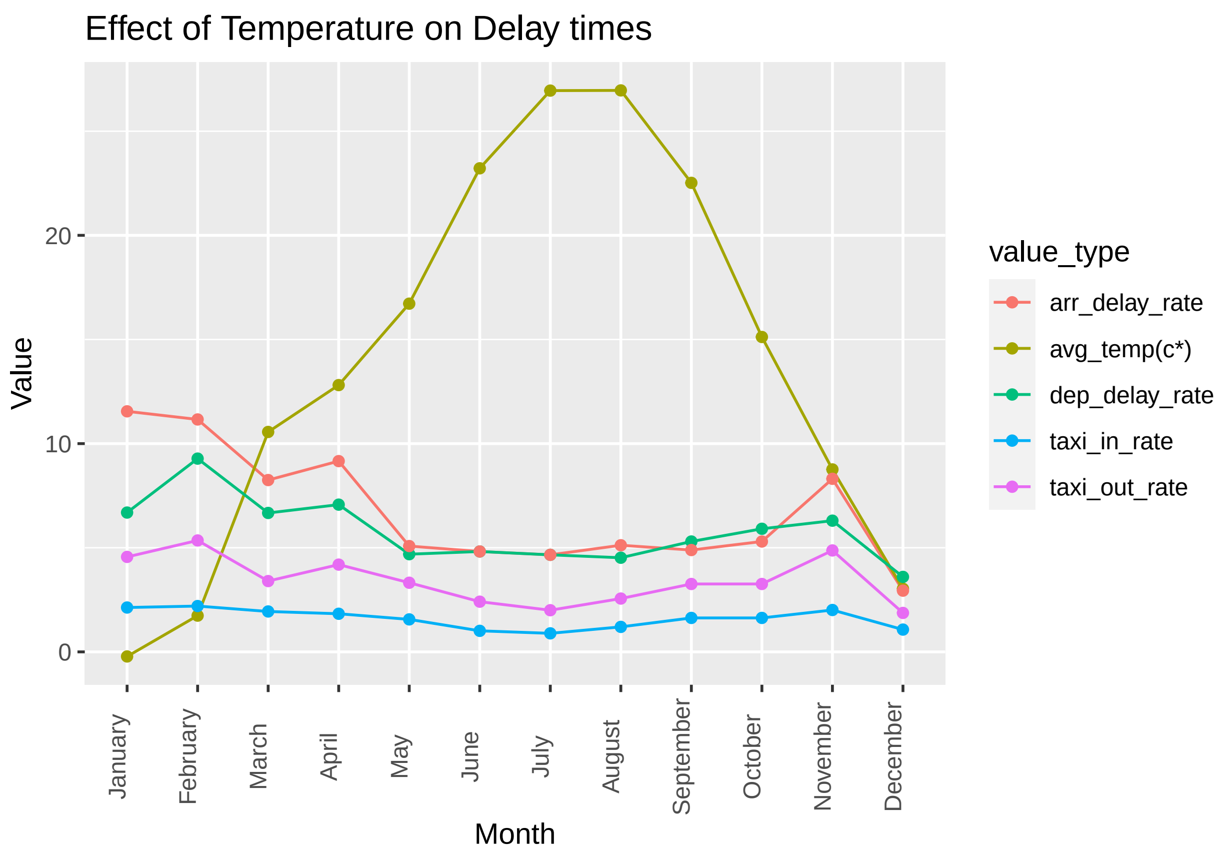
```
[1] 0
```

```r
weather_df<-final_df %>% group_by(month) %>% filter(dep_delay >0,arrival_delay>0) %>%
  summarize(`avg_temp(c*)`=round(mean(Average),2),
            total_flights=n(),
            dep_delay_rate=round((median(arrival_delay)/total_flights)*100,2),
            arr_delay_rate=round((median(dep_delay)/total_flights)*100,2),
            taxi_in_rate=round(median(taxi_in)*100/n(),2),
            taxi_out_rate=round(median(taxi_out)*100/n(),2)) %>%
pivot_longer(cols = c(`avg_temp(c*)`,dep_delay_rate,arr_delay_rate,taxi_in_rate,taxi_out_rate),
               names_to = 'value_type',
               values_to = 'values')

weather_df
```

```
# A tibble: 60 x 4
   month    total_flights value_type     values
   <fct>            <int> <chr>           <dbl>
 1 January            329 avg_temp(c*)    -0.22
 2 January            329 dep_delay_rate   6.69
 3 January            329 arr_delay_rate  11.6
 4 January            329 taxi_in_rate     2.13
 5 January            329 taxi_out_rate    4.56
```

```
 6 February           318 avg_temp(c*)     1.74
 7 February           318 dep_delay_rate   9.28
 8 February           318 arr_delay_rate  11.2
 9 February           318 taxi_in_rate     2.2
10 February           318 taxi_out_rate    5.35
# ... with 50 more rows
```

```
ggplot(weather_df,aes(x=month,y=values,color =value_type,group=value_type)) +
  geom_line() +geom_point()+
  labs(title = "Effect of Temperature on Delay times",
       y = "Value",
       x = "Month") +theme(axis.text.x = element_text(angle = 90, vjust = -0.1))
```



Executive Summary

1.Introduction

Delays is one of the important factors that needs to be considered by Airline authority as they cause great inconvenience to the passengers forcing them to spend more spend more money and also to the airports thereby affecting overall annual profits and reputation.The goal of this analysis is to identity the airports/airlines that are more prone to delays and are experiencing higher delay times in the runways.It also tries to identify the percentage of delays caused within a day and also over the year with the help of weather data.Identifying above problems and providing solutions through in-depth analysis will help the airports to streamline operations effectively and airline companies to hold their positions.

2.Key Findings

Some of the key findings from my analysis reveal that United Airlines are more prone to arrival delays than followed by American and Jetblue and it is due to fact that it has more domestic operations than any other airlines.Flights that are delayed are experiencing additional taxi out,wheels on time of 5 to 10minutes when compared to the flights that arrived early/on time.Airports located in West,Northeast,south region cities such as Los Angeles,San Francisco,Denver,Boston,Network are experiencing delays in several hundreds annually with an average delay time greater than 35 minutes.It should be noted that around 67% of these delays are occurring during Evenings and Night times and they show up to increase between June-August.When it comes to traffic on runways,the taxi-out time doesn't tend to show any impact on the flight arrival delay but airlines such as Skywest,Alaska are facing increased taxi out times between January and April whereas Frontier and ExpressJet are showing higher taxi out times during September along with median delay time greater than 60minutes.Based on summary from the operations,it is observed that Alaska, Skywest,Express Jet and Frontier are exhibiting more delay and arrival times while the taxi in and taxi out times among airways are almost identical.Some of the features such as arrival delay,departure delay and wheels on time are highly correlated with each other and delay in one of these variables affects other significantly.It should be noted that delays and taxi out times are increasing as the temperature starts to drop between July and November and decrease from January as it starts to raise.

3.Recommendations

Based on my analysis,I would recommend FAA to initially reschedule some of the flights operating between Evenings and Nights to Morning and Afternoon as these account for about majority of delays as well as increase flights operations in Summer as this period is more susceptible to delays.It has been observed that majority of delays are happening in cities located in West,South and Northeast regions with a median arrival delay time greater than 37 minutes.While the tax-out and taxi-in times are considerably similar for all the airlines,some of the less popular airlines are experiencing overall delays and taxi-out times especially in Spring and Fall seasons.To overcome above mentioned issues,it is necessary to expand terminals and runways for the corresponding airports/construct additional domestic airports if necessary.Since, arrival delay,departure delay and wheels on time are dependent on one other,addressing one of the issues would eliminate existing problems.It is also important to hire more airport staff to meet the increasing demand in airport operations.Considering delays due to climatic conditions, airlines should take necessary actions such as halting operations during extreme weather conditions,providing reschedules and refunds without any extra charges.