1. Project writeup and code

Flight Delays .

Course-end Project 1

Description

Problem Statement:

Airport arrival performance and other events are affected by weather, which may result in delays or capacity constraints. Almost half of all airport traffic delays are caused by adverse weather conditions. These uncertainties during airport operations can result in significant delays and inconvenience to passengers. Therefore, the airport authority wants to analyze the flights that are delayed and the effect of weather on the delays.

Objective: To visualize the data with the help of histograms, scatter plots, box plots, and pie charts and understand the effect of weather conditions and other factors on flight delays

Data Description

Variable	Description
schedtime	Scheduled time
Carrier	Airline codes
deptime	Time of departure
dest	Destination of flight
distance	Travelling distance
date	Date of travel
flightnum	Flight number
origin	Airport codes
weather	Coded as: 0 – ontime 1 - delayed
dayweek	Coded as: 1 - Sunday and Monday 0 - for all other days
daymonth	Number of days in month
tailnu	Tail number of flight
delay	Delay status

Steps performed

1. Read the dataset
2. Read the dataset description
3. Understand the data
4. Find out the null values
5. Install the required packages
6. Understand the summary of descriptive statistics
7. Plot the histograms to understand the relationships between scheduled time, carrier, destination, origin, weather, and day of the week
8. Plot the scatter plot for flights on time and delayed
9. Plot the box plot to understand how many days in a month flights are delayed by what time
10. Define the hours of departure
11. Create a categorical representation of data using a table
12. Redefine the delay variables
13. Understand the summary of major variables
14. Plot histograms of major variables
15. Plot a pie chart to see how many flights were delayed

2. #code

```
install.packages()

library(readxl)

library(dplyr)

library(ggplot2)

library(tidyr)

library(janitor)

library(lubridate)


#Read the dataset

my_data<-read_excel("C:\\Users\\mahi2\\Desktop\\Data Analytics with
R\\1657873325_flightdelays.xlsx")

my_data


###Read the dataset description

##Reading the names of columns

names(my_data[1:5,])


#Exploratory data analysis


head(my_data)

dim(my_data)

View(my_data)

nrow(my_data)

ncol(my_data)

dim(my_data)

names(my_data)

summary(my_data)
```

```
str(my_data)
head(my_data)
tail(my_data)

### Find out null values
##Printing null values
my_data[is.na(my_data)]
length(my_data[is.na(my_data)])

print(sapply(my_data, function(x) sum(is.na(x))))
##Showing the values of first 3 rows
my_data[1:3,]
set.seed(1)

###Installing the required packages
install.packages("car")
library(car)

###Understanding the summary of descriptive statistics
summary(my_data)
ScheduleTime <- my_data$schedtime


my_data$dest<- as.numeric(factor(my_data$dest))
my_data$carrier<-as.numeric(factor(my_data$carrier))
my_data$origin<-as.numeric(factor(my_data$origin))
```

```
###Plotting the histograms
##Plotting histogram for scheduled time and number of flights
hist(ScheduleTime,col="darkmagenta")
##Plotting histogram for airline codes
plot(my_data$carrier, col="blue", main="Histogram of the Carrier",ylim=c(0, 100))
##Plotting histogram for Destination
plot(my_data$dest, col="blue", main="Histogram of the Destination")
##Plotting histogram for Origins
plot(my_data$origin, col="blue", main="Histogram of the  Origin")
##Plotting histogram for Weather
plot(my_data$weather, col="blue", main="Histogram of the Weather")
library(ggplot2)

###Plotting scatter plot
##Plotting scatter plot for scheduled time and departure time
ggplot(my_data, mapping = aes(x = schedtime, y = deptime)) +
  geom_point()
###Plotting the boxplot
##Plotting boxplot for days of the month and departure time
ggplot(my_data, mapping = aes(x = daymonth, y = deptime)) +
  geom_boxplot()
#Plot the box plot to understand how many days in a month flights are delayed
#by what time
ggplot(my_data, aes(x = daymonth, y = delay)) +
  geom_boxplot() +
  labs(title = "Box Plot to Understand How Many Days in a Month Flights are Delayed by What Time", x =
"Day of the Month", y = "Delay (minutes)")
```

```
my_data <-data.frame(my_data)
names(my_data)
head(my_data)
my_data[1:5,]
dim(my_data)
summary(my_data)
library(car)
##Defining hours of Departure
my_data$sched=factor(floor(my_data$schedtime/100))
###Creating categorical representation using table
##Create a categorical representation of data using table
table(my_data$sched)
table(my_data$carrier)
table(my_data$dest)
table(my_data$origin)
table(my_data$weather)
table(my_data$dayweek)
table(my_data$daymonth)
table(my_data$delay)


###Redefining the delay variable
class(my_data$delay)
my_data$delay<-recode(my_data$delay,"'delayed'=1;else=0")
my_data$delay<-as.numeric(levels(my_data$delay)[my_data$delay])
table(my_data$delay)
```

```
###Understanding the summary of major variables
## Summary of the major variables
summary(my_data$sched)
summary(my_data$carrier)
summary(my_data$dest)
summary(my_data$origin)
summary(my_data$weather)
summary(my_data$dayweek)
summary(my_data$daymonth)
summary(my_data$delay)

my_data$dest<- as.numeric(factor(my_data$dest))
my_data$carrier<-as.numeric(factor(my_data$carrier))
my_data$origin<-as.numeric(factor(my_data$origin))

###Plotting histograms of major variables
## Plots and Histograms of the Major Variables
plot(my_data$sched, col="blue", main="Schedule Departure Time")
plot(my_data$carrier, col="darkblue", main="Flight Carriers")
plot(my_data$dest, col="darkred", main="Destination of Flights")
plot(my_data$origin, col="green", main="Origin of Flights")
plot(my_data$weather, col="darkgreen", main="Weather During Flight Days")
hist(my_data$dayweek, col="darkblue", main="Flights Day of the Week", xlab="Day of Week")
hist(my_data$daymonth, col="yellow", main="Flights Day of the Month")
###Plotting pie chart to see number of delayed flights
##Plotting pie chart to find out delayed flights
x <- c(80,19)
labels <- c("Ontime", "Delayed")
pie(x,labels = piepercent,main = "Flight pie chart",col = rainbow(length(x)))
```

```
legend("topright", c("Ontime","Delayed"), cex = 0.8,
       fill = rainbow(length(x)))
```

Output screenshots

Perform the following tasks on the dataset provided using R

1. Exploratory data analysis:

- Read the dataset
- Read the dataset description

```
> my_data
# A tibble: 2,201 × 13
   schedtime carrier deptime dest  distance date  flightnumber origin weather dayweek
       <dbl> <chr>     <dbl> <chr>    <dbl> <chr>         <dbl> <chr>    <dbl>   <dbl>
 1      1455 OH         1455 JFK        184 37987         5935 BWI          0       4
 2      1640 DH         1640 JFK        213 37987         6155 DCA          0       4
 3      1245 DH         1245 LGA        229 37987         7208 IAD          0       4
 4      1715 DH         1709 LGA        229 37987         7215 IAD          0       4
 5      1039 DH         1035 LGA        229 37987         7792 IAD          0       4
 6       840 DH          839 JFK        228 37987         7800 IAD          0       4
 7      1240 DH         1243 JFK        228 37987         7806 IAD          0       4
 8      1645 DH         1644 JFK        228 37987         7810 IAD          0       4
 9      1715 DH         1710 JFK        228 37987         7812 IAD          0       4
10      2120 DH         2129 JFK        228 37987         7814 IAD          0       4
# i 2,191 more rows
# i 3 more variables: daymonth <dbl>, tailnu <chr>, delay <chr>
# i Use `print(n = ...)` to see more rows
> ###Read the dataset description
> ##Reading the names of columns
> names(my_data[1:5,])
 [1] "schedtime"    "carrier"      "deptime"      "dest"         "distance"
 [6] "date"         "flightnumber" "origin"       "weather"      "dayweek"
[11] "daymonth"     "tailnu"       "delay"
> head(my_data)
# A tibble: 6 × 13
  schedtime carrier deptime dest  distance date  flightnumber origin weather dayweek
      <dbl> <chr>     <dbl> <chr>    <dbl> <chr>         <dbl> <chr>    <dbl>   <dbl>
1      1455 OH         1455 JFK        184 37987         5935 BWI          0       4
2      1640 DH         1640 JFK        213 37987         6155 DCA          0       4
3      1245 DH         1245 LGA        229 37987         7208 IAD          0       4
4      1715 DH         1709 LGA        229 37987         7215 IAD          0       4
5      1039 DH         1035 LGA        229 37987         7792 IAD          0       4
6       840 DH          839 JFK        228 37987         7800 IAD          0       4
# i 3 more variables: daymonth <dbl>, tailnu <chr>, delay <chr>
```

```
6       840 DH          839 JFK        228 3/987         7800 IAD          0       4
# i 3 more variables: daymonth <dbl>, tailnu <chr>, delay <chr>
> dim(my_data)
[1] 2201   13
> View(my_data)
> nrow(my_data)
[1] 2201
> ncol(my_data)
[1] 13
> dim(my_data)
[1] 2201   13
> names(my_data)
 [1] "schedtime"    "carrier"      "deptime"      "dest"         "distance"
 [6] "date"         "flightnumber" "origin"       "weather"      "dayweek"
[11] "daymonth"     "tailnu"       "delay"
> summary(my_data)
   schedtime      carrier            deptime         dest              distance
 Min.   : 600   Length:2201        Min.   :  10   Length:2201        Min.   :169.0
 1st Qu.:1000   Class :character   1st Qu.:1004   Class :character   1st Qu.:213.0
 Median :1455   Mode  :character   Median :1450   Mode  :character   Median :214.0
 Mean   :1372                      Mean   :1369                      Mean   :211.9
 3rd Qu.:1710                      3rd Qu.:1709                      3rd Qu.:214.0
 Max.   :2130                      Max.   :2330                      Max.   :229.0
     date            flightnumber     origin             weather           dayweek
 Length:2201        Min.   : 746   Length:2201        Min.   :0.00000   Min.   :1.000
 Class :character   1st Qu.:2156   Class :character   1st Qu.:0.00000   1st Qu.:2.000
 Mode  :character   Median :2385   Mode  :character   Median :0.00000   Median :4.000
                    Mean   :3815                      Mean   :0.01454   Mean   :3.905
                    3rd Qu.:6155                      3rd Qu.:0.00000   3rd Qu.:5.000
                    Max.   :7924                      Max.   :1.00000   Max.   :7.000
    daymonth        tailnu             delay
 Min.   : 1.00   Length:2201        Length:2201
 1st Qu.: 8.00   Class :character   Class :character
 Median :16.00   Mode  :character   Mode  :character
 Mean   :16.02
 3rd Qu.:23.00
 Max.   :31.00
```

```
> str(my_data)
tibble [2,201 × 13] (S3: tbl_df/tbl/data.frame)
 $ schedtime   : num [1:2201] 1455 1640 1245 1715 1039 ...
 $ carrier     : chr [1:2201] "OH" "DH" "DH" "DH" ...
 $ deptime     : num [1:2201] 1455 1640 1245 1709 1035 ...
 $ dest        : chr [1:2201] "JFK" "JFK" "LGA" "LGA" ...
 $ distance    : num [1:2201] 184 213 229 229 229 228 228 228 228 228 ...
 $ date        : chr [1:2201] "37987" "37987" "37987" "37987" ...
 $ flightnumber: num [1:2201] 5935 6155 7208 7215 7792 ...
 $ origin      : chr [1:2201] "BWI" "DCA" "IAD" "IAD" ...
 $ weather     : num [1:2201] 0 0 0 0 0 0 0 0 0 0 ...
 $ dayweek     : num [1:2201] 4 4 4 4 4 4 4 4 4 4 ...
 $ daymonth    : num [1:2201] 1 1 1 1 1 1 1 1 1 1 ...
 $ tailnu      : chr [1:2201] "N940CA" "N405FJ" "N695BR" "N662BR" ...
 $ delay       : chr [1:2201] "ontime" "ontime" "ontime" "ontime" ...
> head(my_data)
# A tibble: 6 × 13
  schedtime carrier deptime dest  distance date  flightnumber origin weather dayweek
      <dbl> <chr>     <dbl> <chr>    <dbl> <chr>        <dbl> <chr>    <dbl>   <dbl>
1      1455 OH         1455 JFK        184 37987         5935 BWI          0       4
2      1640 DH         1640 JFK        213 37987         6155 DCA          0       4
3      1245 DH         1245 LGA        229 37987         7208 IAD          0       4
4      1715 DH         1709 LGA        229 37987         7215 IAD          0       4
5      1039 DH         1035 LGA        229 37987         7792 IAD          0       4
6       840 DH          839 JFK        228 37987         7800 IAD          0       4
# i 3 more variables: daymonth <dbl>, tailnu <chr>, delay <chr>
> tail(my_data)
# A tibble: 6 × 13
  schedtime carrier deptime dest  distance date      flightnumber origin weather dayweek
      <dbl> <chr>     <dbl> <chr>    <dbl> <chr>            <dbl> <chr>    <dbl>   <dbl>
1       700 RU          650 EWR        213 1/31/2004         2855 IAD          0       6
2       645 RU          644 EWR        199 1/31/2004         2761 DCA          0       6
3      1700 RU         1653 EWR        213 1/31/2004         2497 IAD          0       6
4      1600 RU         1558 EWR        199 1/31/2004         2361 DCA          0       6
5      1359 RU         1403 EWR        199 1/31/2004         2216 DCA          0       6
```

```
> my_data[is.na(my_data)]
<unspecified> [0]
> length(my_data[is.na(my_data)])
[1] 0
> print(sapply(my_data, function(x) sum(is.na(x))))
   schedtime      carrier      deptime         dest     distance         date flightnumber
           0            0            0            0            0            0            0
      origin      weather      dayweek     daymonth       tailnu        delay
           0            0            0            0            0            0
> ##Showing the values of first 3 rows
> my_data[1:3,]
# A tibble: 3 × 13
  schedtime carrier deptime dest  distance date  flightnumber origin weather dayweek
      <dbl> <chr>     <dbl> <chr>    <dbl> <chr>        <dbl> <chr>    <dbl>   <dbl>
1      1455 OH         1455 JFK        184 37987         5935 BWI          0       4
2      1640 DH         1640 JFK        213 37987         6155 DCA          0       4
3      1245 DH         1245 LGA        229 37987         7208 IAD          0       4
# i 3 more variables: daymonth <dbl>, tailnu <chr>, delay <chr>
>
```

- Understand the data
- Find out the null values

2. Install the required packages

```
> install.packages()
Warning in install.packages :
  package 'car' is in use and will not be installed
> library(readxl)
> library(dplyr)
> library(ggplot2)
> library(tidyr)
> library(lubridate)
> tdelays.xlsx")
+ my_data
+ #Read the dataset
+ my_data<-read_excel("C:\\Users\\mahi2\\Desktop\\Data Analytics with R\\1657873325_flightdela
ys.xlsx")
```

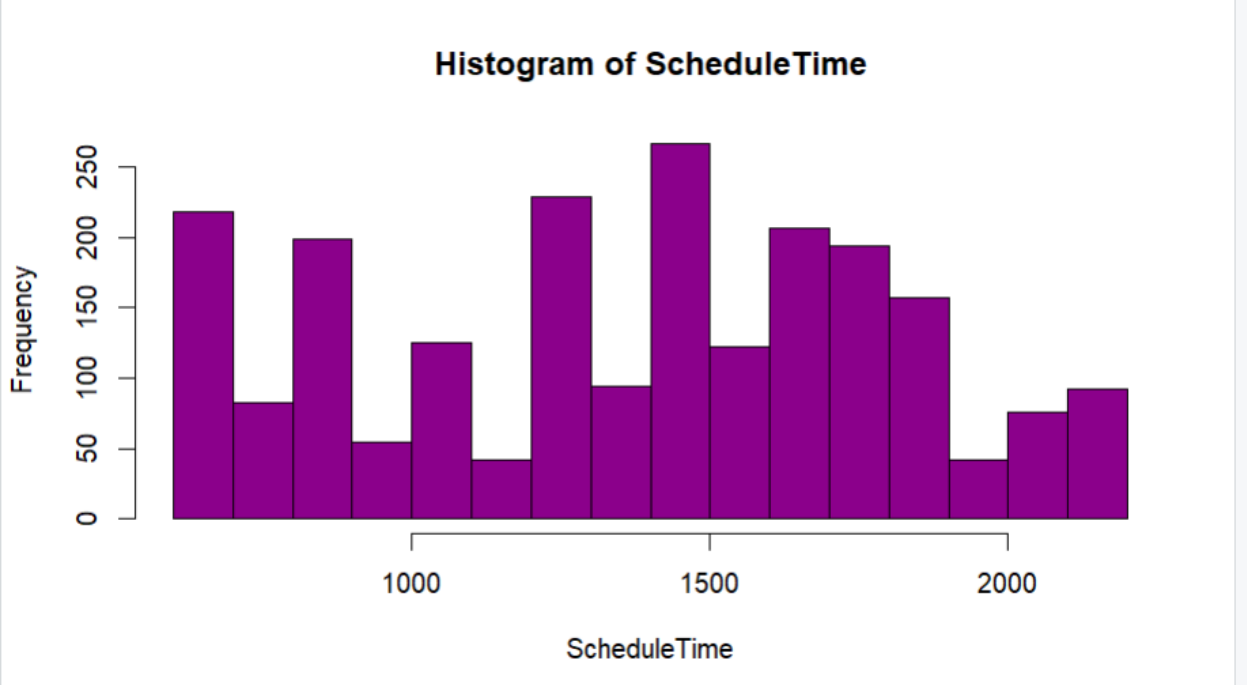
3. Understand the summary of descriptive statistics

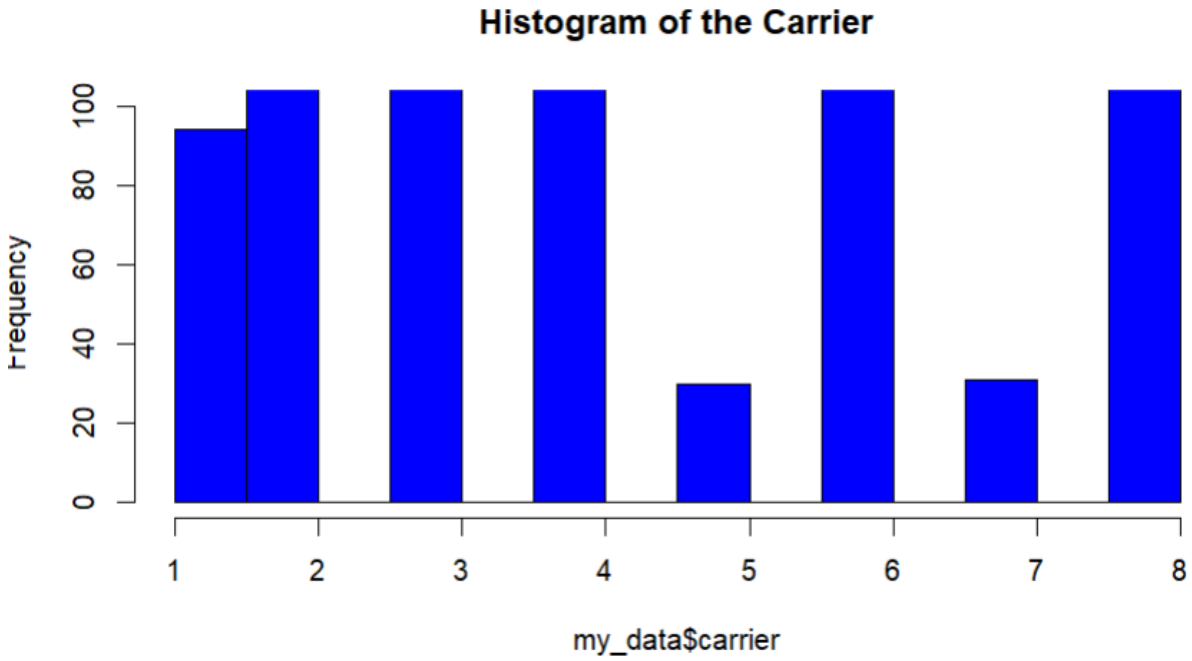
```
> my_data[is.na(my_data)]
<unspecified> [0]
> length(my_data[is.na(my_data)])
[1] 0
> print(sapply(my_data, function(x) sum(is.na(x))))
   schedtime      carrier      deptime         dest     distance         date flightnumber
           0            0            0            0            0            0            0
      origin      weather      dayweek     daymonth       tailnu        delay
           0            0            0            0            0            0
> ##Showing the values of first 3 rows
> my_data[1:3,]
# A tibble: 3 × 13
  schedtime carrier deptime dest  distance date  flightnumber origin weather dayweek
      <dbl> <chr>     <dbl> <chr>    <dbl> <chr>         <dbl> <chr>    <dbl>   <dbl>
1      1455 OH         1455 JFK        184 37987          5935 BWI          0       4
2      1640 DH         1640 JFK        213 37987          6155 DCA          0       4
3      1245 DH         1245 LGA        229 37987          7208 IAD          0       4
# i 3 more variables: daymonth <dbl>, tailnu <chr>, delay <chr>
> |
```

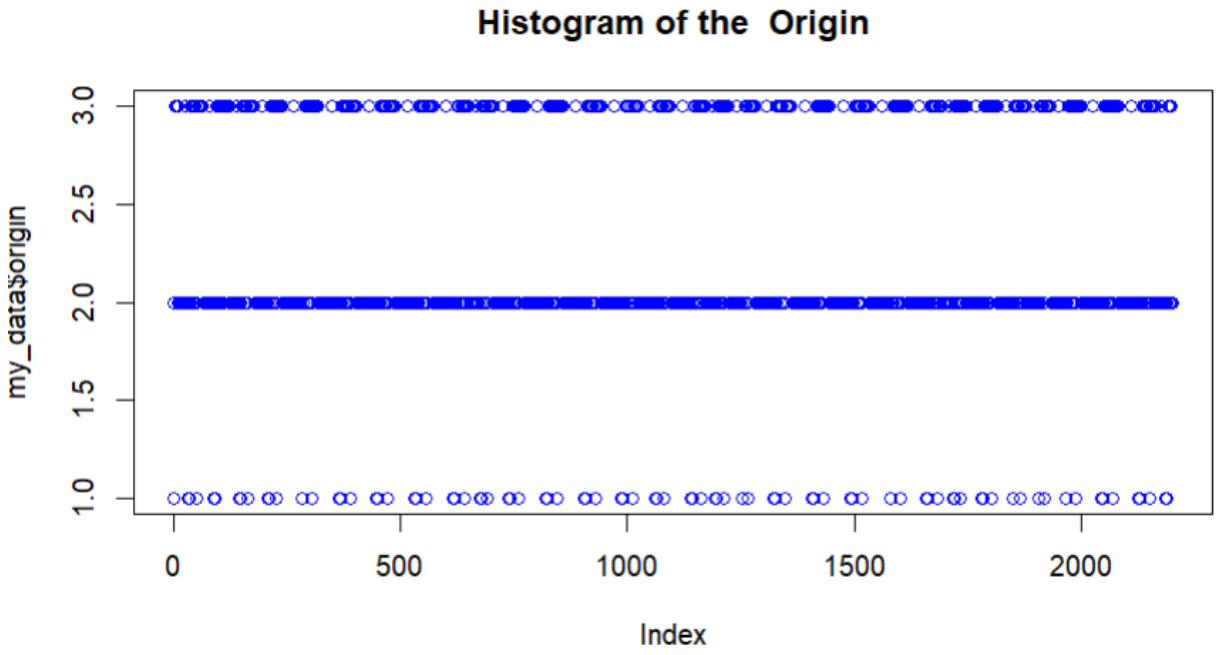
```
> str(my_data)
tibble [2,201 × 13] (S3: tbl_df/tbl/data.frame)
 $ schedtime   : num [1:2201] 1455 1640 1245 1715 1039 ...
 $ carrier     : chr [1:2201] "OH" "DH" "DH" "DH" ...
 $ deptime     : num [1:2201] 1455 1640 1245 1709 1035 ...
 $ dest        : chr [1:2201] "JFK" "JFK" "LGA" "LGA" ...
 $ distance    : num [1:2201] 184 213 229 229 229 228 228 228 228 228 ...
 $ date        : chr [1:2201] "37987" "37987" "37987" "37987" ...
 $ flightnumber: num [1:2201] 5935 6155 7208 7215 7792 ...
 $ origin      : chr [1:2201] "BWI" "DCA" "IAD" "IAD" ...
 $ weather     : num [1:2201] 0 0 0 0 0 0 0 0 0 0 ...
 $ dayweek     : num [1:2201] 4 4 4 4 4 4 4 4 4 4 ...
 $ daymonth    : num [1:2201] 1 1 1 1 1 1 1 1 1 1 ...
 $ tailnu      : chr [1:2201] "N940CA" "N405FJ" "N695BR" "N662BR" ...
 $ delay       : chr [1:2201] "ontime" "ontime" "ontime" "ontime" ...
> head(my_data)
# A tibble: 6 × 13
  schedtime carrier deptime dest  distance date  flightnumber origin weather dayweek
      <dbl> <chr>     <dbl> <chr>    <dbl> <chr>         <dbl> <chr>    <dbl>   <dbl>
1      1455 OH         1455 JFK        184 37987          5935 BWI          0       4
2      1640 DH         1640 JFK        213 37987          6155 DCA          0       4
3      1245 DH         1245 LGA        229 37987          7208 IAD          0       4
4      1715 DH         1709 LGA        229 37987          7215 IAD          0       4
5      1039 DH         1035 LGA        229 37987          7792 IAD          0       4
6       840 DH          839 JFK        228 37987          7800 IAD          0       4
# i 3 more variables: daymonth <dbl>, tailnu <chr>, delay <chr>
> tail(my_data)
# A tibble: 6 × 13
  schedtime carrier deptime dest  distance date      flightnumber origin weather dayweek
      <dbl> <chr>     <dbl> <chr>    <dbl> <chr>            <dbl> <chr>    <dbl>   <dbl>
1       700 RU          650 EWR        213 1/31/2004         2855 IAD          0       6
2       645 RU          644 EWR        199 1/31/2004         2761 DCA          0       6
3      1700 RU         1653 EWR        213 1/31/2004         2497 IAD          0       6
4      1600 RU         1558 EWR        199 1/31/2004         2361 DCA          0       6
5      1359 RU         1403 EWR        199 1/31/2004         2216 DCA          0       6
```

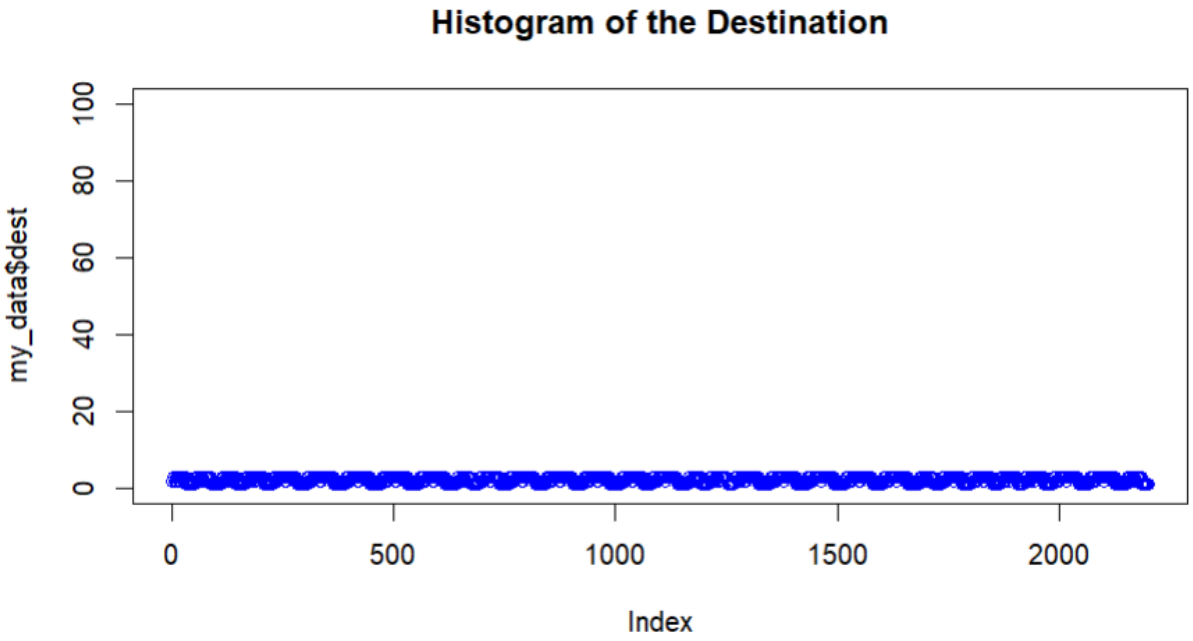
```
package 'car' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\mahi2\AppData\Local\Temp\RtmpmeCs0c\downloaded_packages
> ###Understanding the summary of descriptive statistics
> summary(my_data)
   schedtime      carrier            deptime        dest              distance
 Min.   : 600   Length:2201        Min.   :  10   Length:2201        Min.   :169.0
 1st Qu.:1000   Class :character   1st Qu.:1004   Class :character   1st Qu.:213.0
 Median :1455   Mode  :character   Median :1450   Mode  :character   Median :214.0
 Mean   :1372                      Mean   :1369                      Mean   :211.9
 3rd Qu.:1710                      3rd Qu.:1709                      3rd Qu.:214.0
 Max.   :2130                      Max.   :2330                      Max.   :229.0
     date            flightnumber     origin            weather           dayweek
 Length:2201        Min.   : 746   Length:2201        Min.   :0.00000   Min.   :1.000
 Class :character   1st Qu.:2156   Class :character   1st Qu.:0.00000   1st Qu.:2.000
 Mode  :character   Median :2385   Mode  :character   Median :0.00000   Median :4.000
                    Mean   :3815                      Mean   :0.01454   Mean   :3.905
                    3rd Qu.:6155                      3rd Qu.:0.00000   3rd Qu.:5.000
                    Max.   :7924                      Max.   :1.00000   Max.   :7.000
    daymonth        tailnu             delay
 Min.   : 1.00   Length:2201        Length:2201
 1st Qu.: 8.00   Class :character   Class :character
 Median :16.00   Mode  :character   Mode  :character
 Mean   :16.02
 3rd Qu.:23.00
 Max.   :31.00
>
```

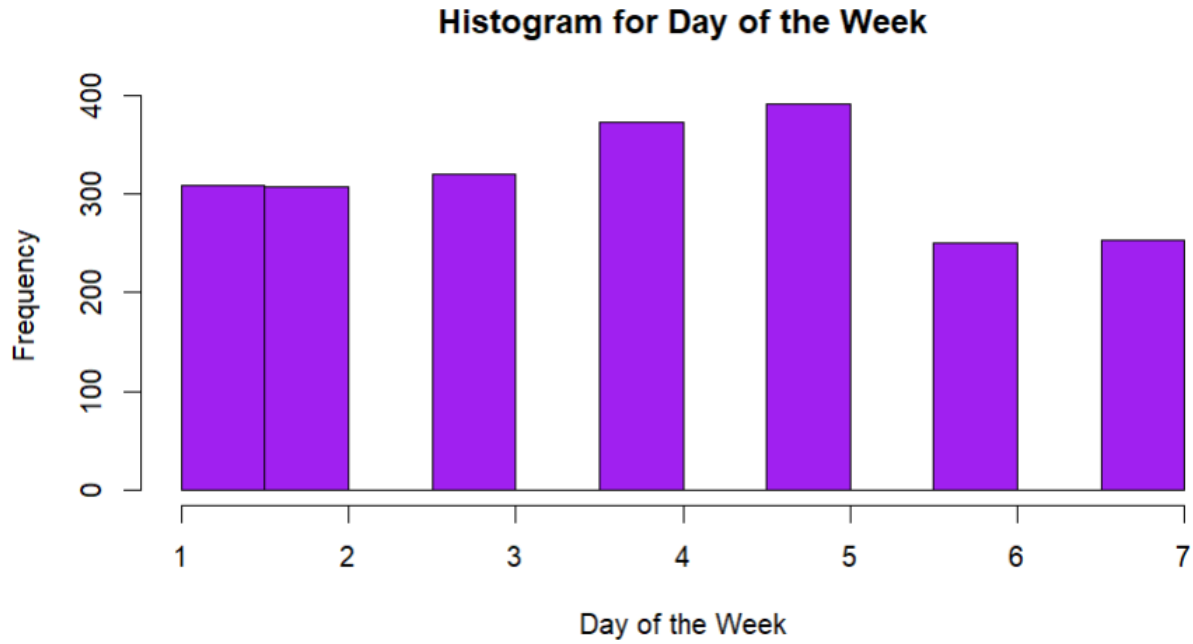
4. Plot the histograms to understand the relationships between scheduled time, carrier, destination, origin, weather, and day of the week

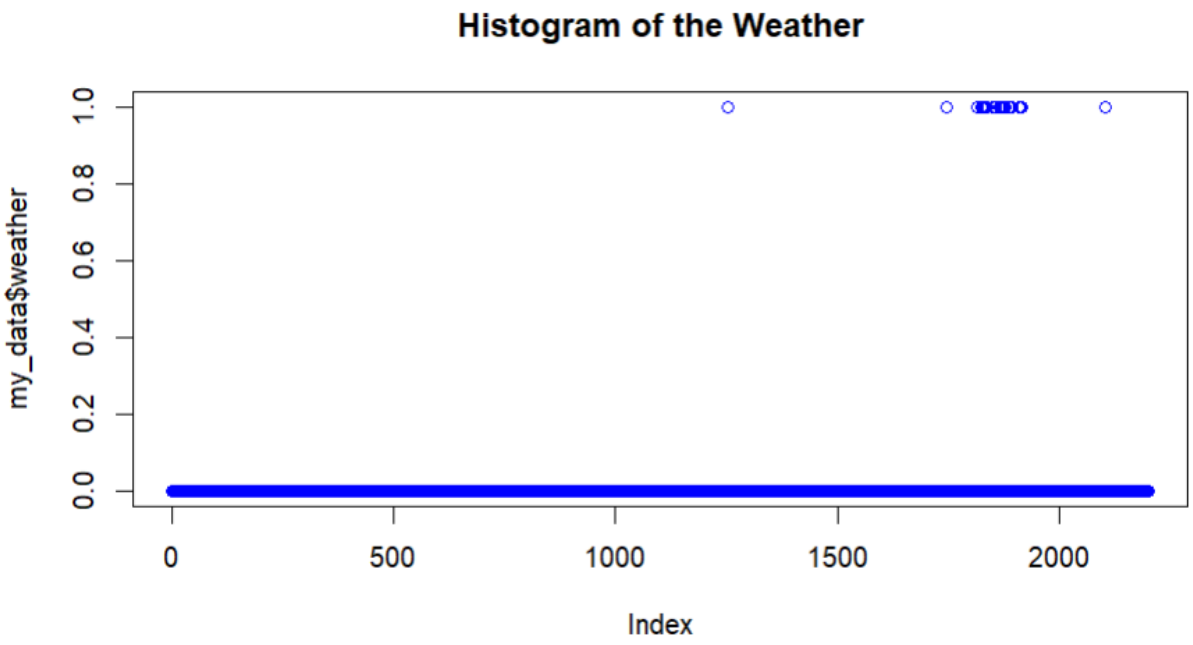



Histogram of the Carrier
Frequency
0
20
40
60
80
100
1
2
3
4
5
6
7
8
my_data$carrier

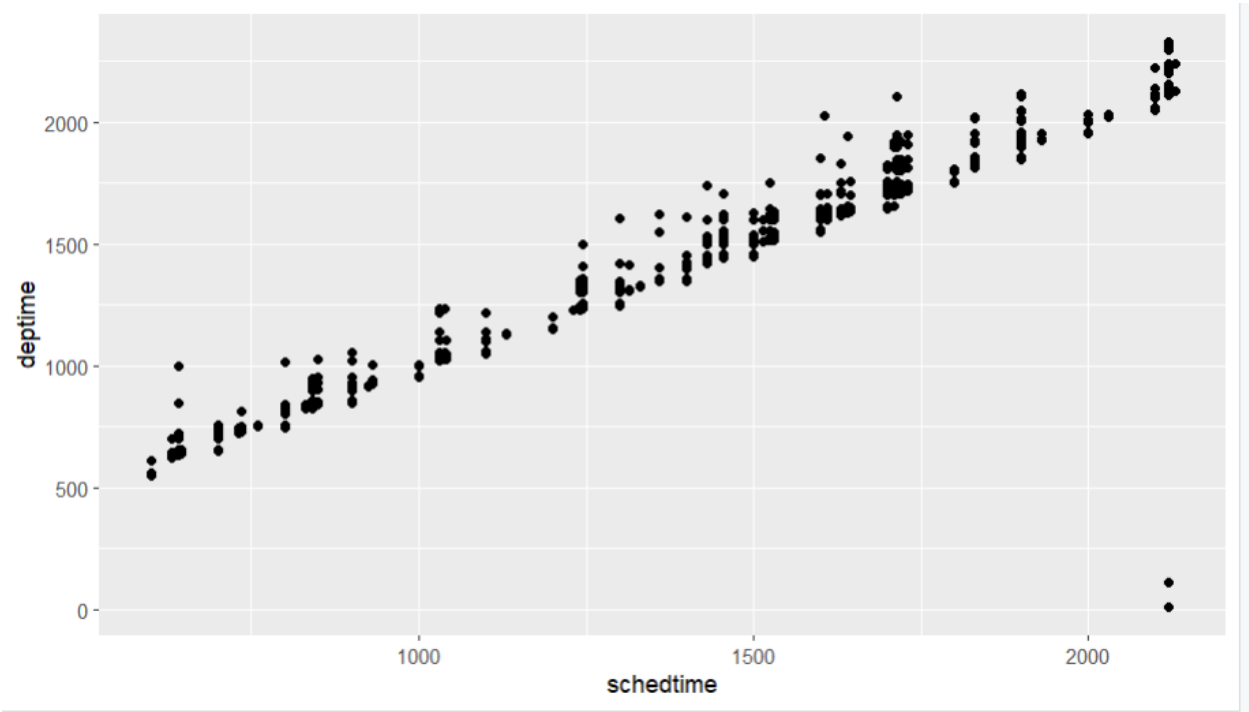

Histogram of the Origin
my_data$origin
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
2000
Index


Histogram of the Destination
my_data$dest
0
20
40
60
80
100
0
500
1000
1500
2000
Index


Histogram for Day of the Week
Frequency
0
100
200
300
400
1
2
3
4
5
6
7
Day of the Week

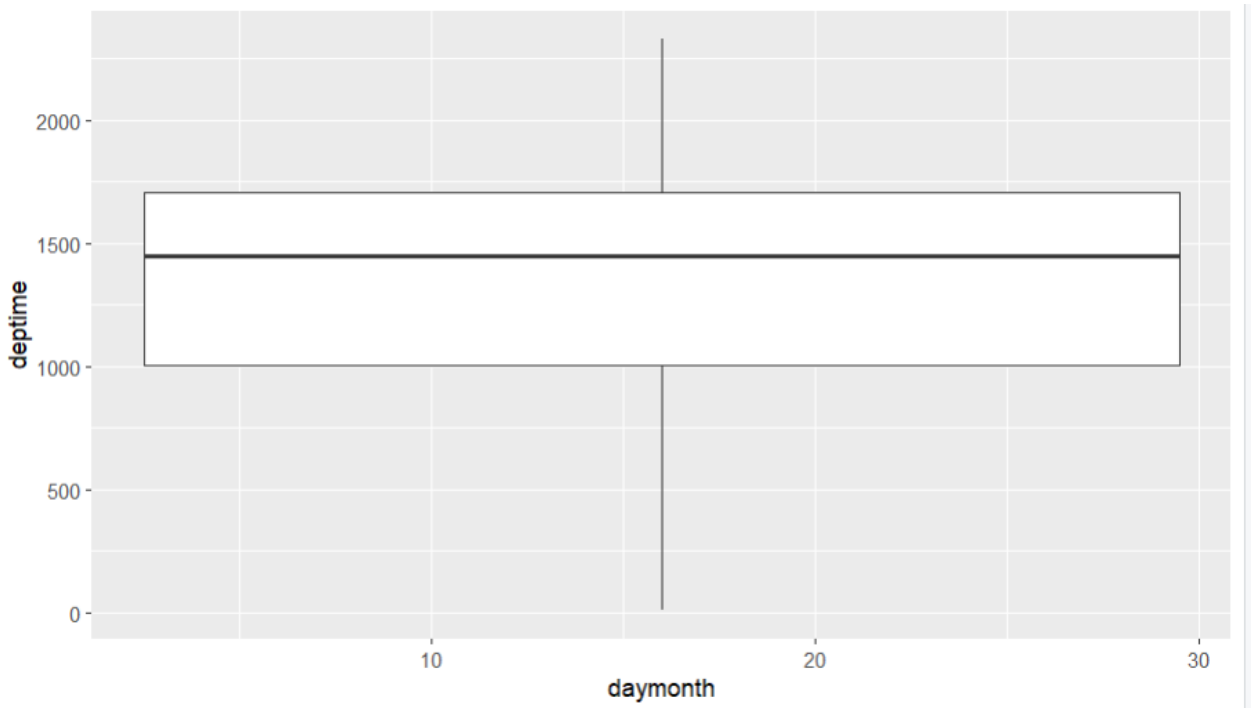


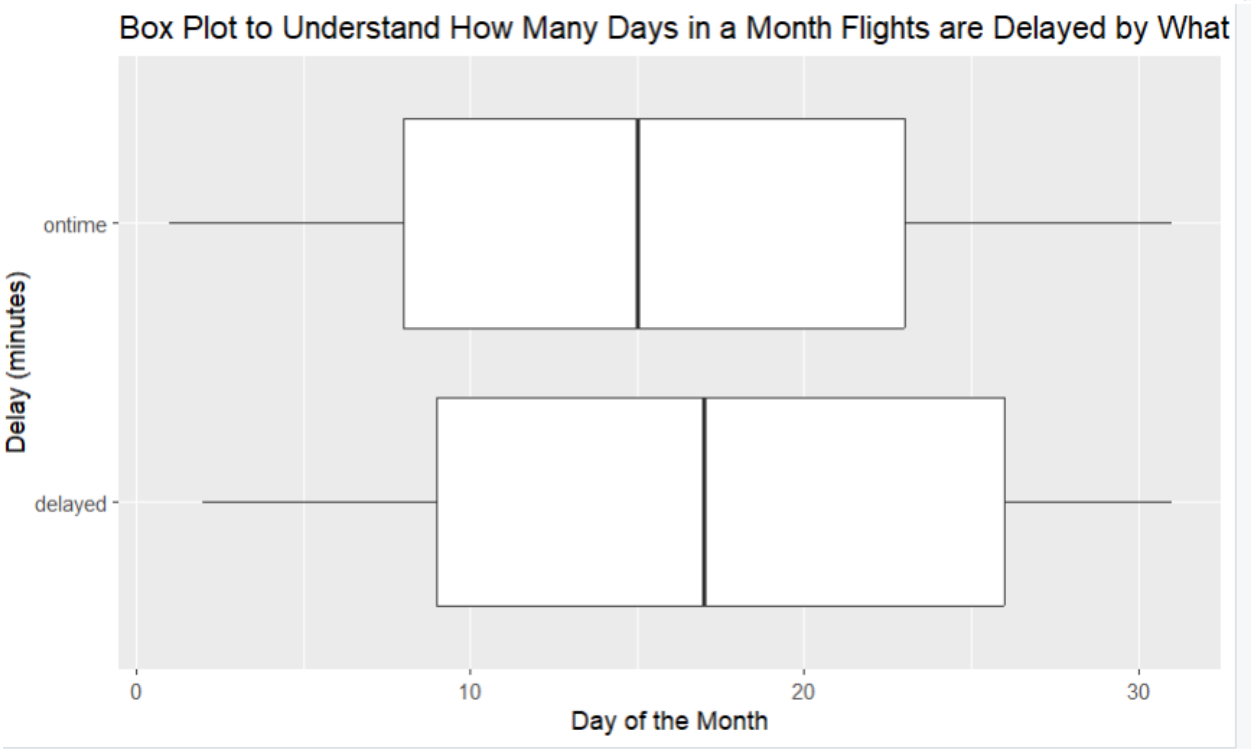
5. Plot the scatter plot for flights on time and delayed Tasks to Perform



Perform the following tasks on the dataset provided using R:

6. Plot the box plot to understand how many days in a month flights are delayed by what time





7. Define the hours of departure

8. Create a categorical representation of data using a table

```
> library(car)
> ##Defining hours of Departure
> my_data$sched=factor(floor(my_data$schedtime/100))
> ###Creating categorical representation using table
> ##Create a categorical representation of data using table
> table(my_data$sched)

  6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21
126 135 162 108 100  68 142 157 236 154 178 241  85 119  53 137
> table(my_data$carrier)

 CO  DH  DL  MQ  OH  RU  UA  US
 94 551 388 295  30 408  31 404
> table(my_data$dest)

 EWR  JFK  LGA
 665  386 1150
> table(my_data$origin)

 BWI  DCA  IAD
 145 1370  686
> table(my_data$weather)

   0    1
2169   32
> table(my_data$dayweek)

  1   2   3   4   5   6   7
308 307 320 372 391 250 253
> table(my_data$daymonth)

 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
50 62 51 63 78 85 83 85 84 50 68 84 85 84 69 80 49 55 81 85 85 86 81 50 67 65 52 68 82 84
31
50
> table(my_data$delay)

delayed  ontime
    428    1773
>
```

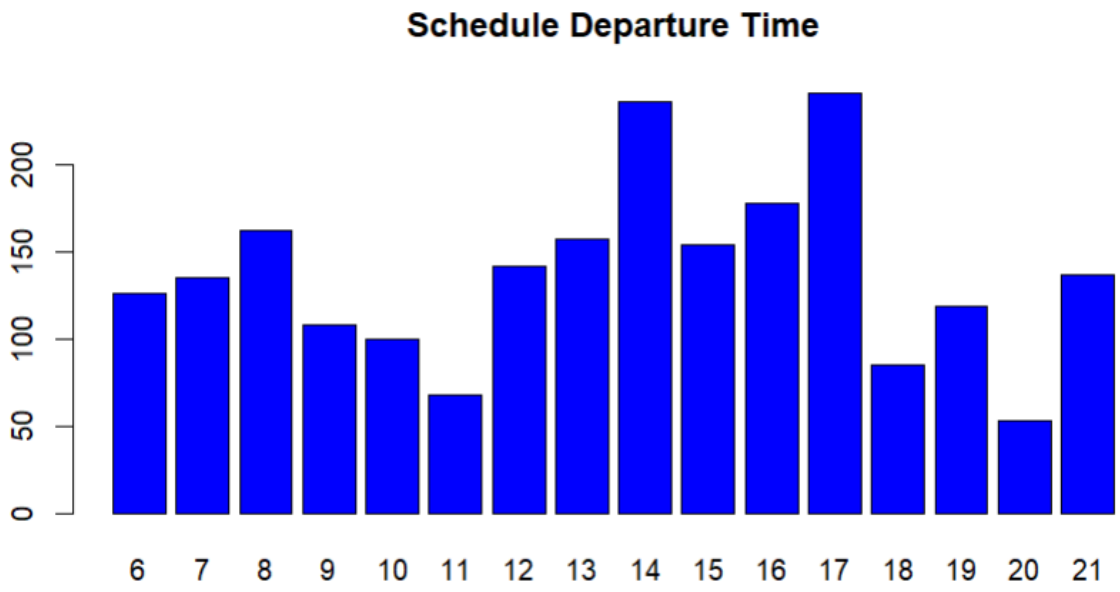
9. Redefine the delay variables

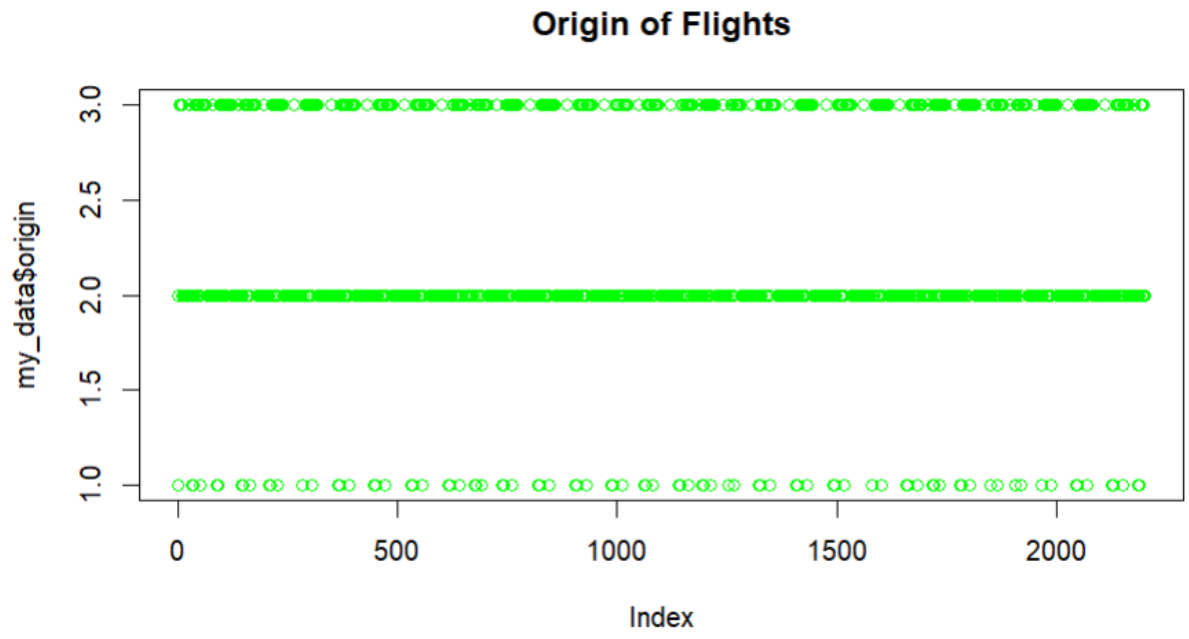
```
> names(my_data)
 [1] "schedtime"    "carrier"      "deptime"      "dest"         "distance"
 [6] "date"         "flightnumber" "origin"       "weather"      "dayweek"
[11] "daymonth"     "tailnu"       "delay"
> head(my_data)
  schedtime carrier deptime dest distance  date flightnumber origin weather dayweek
1      1455      OH    1455  JFK      184 37987          5935    BWI       0       4
2      1640      DH    1640  JFK      213 37987          6155    DCA       0       4
3      1245      DH    1245  LGA      229 37987          7208    IAD       0       4
4      1715      DH    1709  LGA      229 37987          7215    IAD       0       4
5      1039      DH    1035  LGA      229 37987          7792    IAD       0       4
6       840      DH     839  JFK      228 37987          7800    IAD       0       4
  daymonth tailnu  delay
1        1 N940CA ontime
2        1 N405FJ ontime
3        1 N695BR ontime
4        1 N662BR ontime
5        1 N698BR ontime
6        1 N687BR ontime
> dim(my_data)
[1] 2201   13
> summary(my_data)
   schedtime      carrier            deptime         dest              distance
 Min.   : 600   Length:2201        Min.   :  10   Length:2201        Min.   :169.0
 1st Qu.:1000   Class :character   1st Qu.:1004   Class :character   1st Qu.:213.0
 Median :1455   Mode  :character   Median :1450   Mode  :character   Median :214.0
 Mean   :1372                      Mean   :1369                      Mean   :211.9
 3rd Qu.:1710                      3rd Qu.:1709                      3rd Qu.:214.0
 Max.   :2130                      Max.   :2330                      Max.   :229.0
     date            flightnumber     origin             weather           dayweek
 Length:2201        Min.   : 746   Length:2201        Min.   :0.00000   Min.   :1.000
 Class :character   1st Qu.:2156   Class :character   1st Qu.:0.00000   1st Qu.:2.000
```

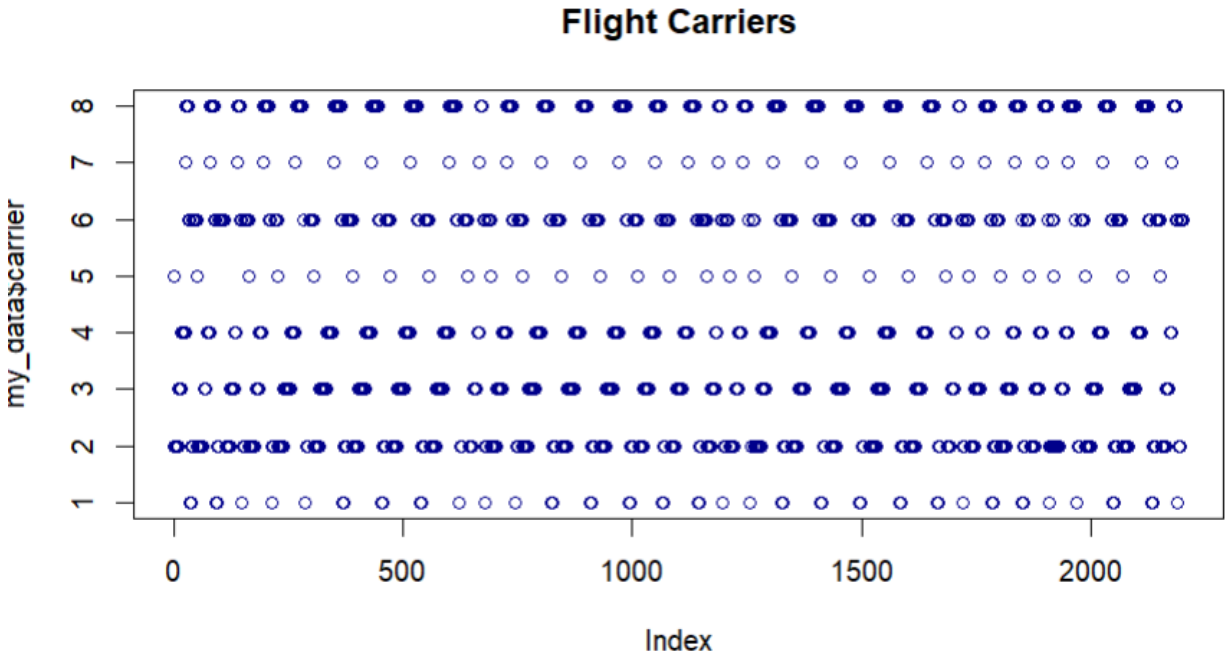
10. Understand the summary of major variables

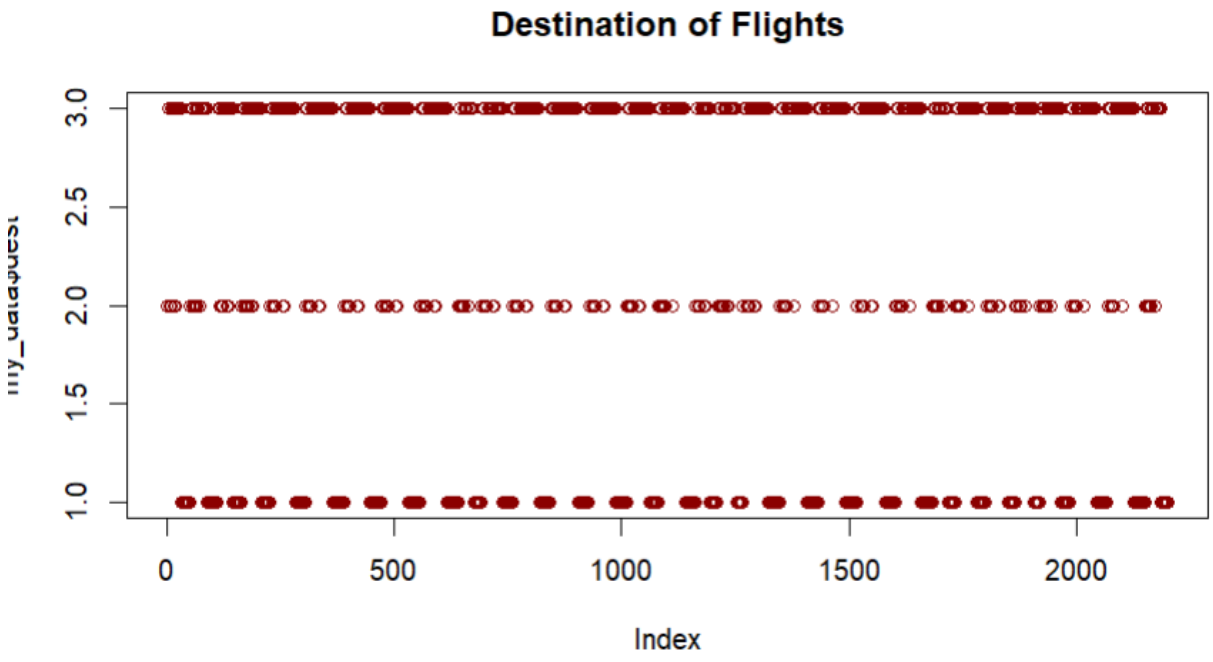
```
> ###Understanding the summary of major variables
> ## Summary of the major variables
> summary(my_data$sched)
  6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21
126 135 162 108 100  68 142 157 236 154 178 241  85 119  53 137
> summary(my_data$carrier)
   Length     Class      Mode
     2201 character character
> summary(my_data$dest)
   Length     Class      Mode
     2201 character character
> summary(my_data$origin)
   Length     Class      Mode
     2201 character character
> summary(my_data$weather)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00000 0.00000 0.00000 0.01454 0.00000 1.00000
> summary(my_data$dayweek)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   4.000   3.905   5.000   7.000
> summary(my_data$daymonth)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    8.00   16.00   16.02   23.00   31.00
> summary(my_data$delay)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0       0       0       0       0       0
>
```

11. Plot histograms of major variables




Origin of Flights
my_data$origin
1.0
1.5
2.0
2.5
3.0
0
500
1000
1500
2000
Index


Flight Carriers
my_data$carrier
1
2
3
4
5
6
7
8
0
500
1000
1500
2000
Index



12. Plot a pie chart to see how many flights were delayed