Mayank_Sagar_Airline_Code.R

dell

Wed Jul 05 01:44:39 2017

```
# Analysis of Airline Ticket Pricing
# NAME: Mayank Sagar
# EMAIL: mayanksagar26@gmail.com
# COLLEGE : R.V.COllege Of Engineering, Bengaluru.



#######PREMIUM ECONOMY VS ECONOMY TICKET PRICING BY AIRLINES#######
#####################################################################
#Airline Fares

##setting the directory and assigning a variabel to the data frame
setwd("D:/Data Science and Analytics using R/Data files")

#Reading the dataset and creating a data frame
air.df<-read.csv(paste("SixAirlines.csv",sep = ""))

#Viewing the data frame
View(air.df)

##Analyzing the summary of the data and describing the variables
library(psych)
describe(air.df)
```

```
##                 vars   n    mean      sd  median trimmed     mad   min
## AIRLINE*           1 462    3.02    1.65    2.00    2.90    1.48  1.00
## AIRCRAFT           2 462    0.33    0.47    0.00    0.28    0.00  0.00
## FLIGHT_DURATION    3 462    7.55    3.54    7.75    7.54    4.82  1.25
## MONTH              4 462    1.67    1.05    2.00    1.71    1.48  0.00
## INTERNATIONAL      5 462    0.91    0.28    1.00    1.00    0.00  0.00
## SEATS_ECONOMY      6 462  200.71   77.96  185.00  193.76   85.99 17.00
## SEATS_PREMIUM      7 462   33.54   13.26   36.00   33.20   11.86  8.00
## PITCH_ECONOMY      8 462   31.21    0.66   31.00   31.25    0.00 30.00
## PITCH_PREMIUM      9 462   37.92    1.32   38.00   38.06    0.00 34.00
## WIDTH_ECONOMY     10 462   17.83    0.56   18.00   17.81    0.00 17.00
## WIDTH_PREMIUM     11 462   19.48    1.10   19.00   19.54    0.00 17.00
## PRICE_ECONOMY     12 462 1317.06  989.81 1224.00 1231.30 1163.84 65.00
## PRICE_PREMIUM     13 462 1832.35 1289.97 1710.00 1782.94 1852.51 86.00
## PRICE_RELATIVE    14 462    0.49    0.45    0.38    0.43    0.42  0.02
## N                 15 462  234.25   86.88  227.00  227.69   90.44 38.00
## LAMBDA            16 462    0.15    0.06    0.13    0.14    0.03  0.05
## QUALITY           17 462    6.72    1.78    7.00    6.79    0.00  2.00
##                     max   range  skew kurtosis    se
## AIRLINE*           6.00    5.00  0.59    -0.95  0.08
## AIRCRAFT           1.00    1.00  0.74    -1.46  0.02
## FLIGHT_DURATION   14.66   13.41 -0.05    -1.12  0.16
## MONTH              3.00    3.00 -0.16    -1.20  0.05
## INTERNATIONAL      1.00    1.00 -2.93     6.60  0.01
## SEATS_ECONOMY    389.00  372.00  0.61    -0.26  3.63
## SEATS_PREMIUM     66.00   58.00  0.25    -0.46  0.62
## PITCH_ECONOMY     33.00    3.00 -0.03    -0.38  0.03
## PITCH_PREMIUM     40.00    6.00 -1.48     3.43  0.06
## WIDTH_ECONOMY     19.00    2.00 -0.03    -0.12  0.03
## WIDTH_PREMIUM     21.00    4.00 -0.09    -0.34  0.05
## PRICE_ECONOMY   3593.00 3528.00  0.52    -0.88 46.05
## PRICE_PREMIUM   7414.00 7328.00  0.51     0.41 60.01
## PRICE_RELATIVE     1.89    1.87  1.14     0.61  0.02
## N                441.00  403.00  0.61    -0.44  4.04
## LAMBDA             0.55    0.50  2.70    14.02  0.00
## QUALITY           10.00    8.00 -0.51     1.67  0.08
```

```
summary(air.df)
```

```
##       AIRLINE       AIRCRAFT      FLIGHT_DURATION      MONTH
##  AirFrance: 74   Min.   :0.0000   Min.   : 1.250   Min.   :0.000
##  British  :175   1st Qu.:0.0000   1st Qu.: 4.250   1st Qu.:1.000
##  Delta    : 46   Median :0.0000   Median : 7.750   Median :2.000
##  Jet      : 65   Mean   :0.3268   Mean   : 7.549   Mean   :1.671
##  Singapore: 40   3rd Qu.:1.0000   3rd Qu.:10.500   3rd Qu.:3.000
##  Virgin   : 62   Max.   :1.0000   Max.   :14.660   Max.   :3.000
##  INTERNATIONAL    SEATS_ECONOMY   SEATS_PREMIUM   PITCH_ECONOMY
##  Min.   :0.0000   Min.   : 17.0   Min.   : 8.00   Min.   :30.00
##  1st Qu.:1.0000   1st Qu.:127.0   1st Qu.:21.00   1st Qu.:31.00
##  Median :1.0000   Median :185.0   Median :36.00   Median :31.00
##  Mean   :0.9134   Mean   :200.7   Mean   :33.54   Mean   :31.21
##  3rd Qu.:1.0000   3rd Qu.:243.0   3rd Qu.:40.00   3rd Qu.:32.00
##  Max.   :1.0000   Max.   :389.0   Max.   :66.00   Max.   :33.00
##  PITCH_PREMIUM   WIDTH_ECONOMY   WIDTH_PREMIUM   PRICE_ECONOMY
##  Min.   :34.00   Min.   :17.00   Min.   :17.00   Min.   :  65.0
##  1st Qu.:38.00   1st Qu.:17.00   1st Qu.:19.00   1st Qu.: 404.8
##  Median :38.00   Median :18.00   Median :19.00   Median :1224.0
##  Mean   :37.92   Mean   :17.83   Mean   :19.48   Mean   :1317.1
##  3rd Qu.:38.00   3rd Qu.:18.00   3rd Qu.:21.00   3rd Qu.:1903.0
##  Max.   :40.00   Max.   :19.00   Max.   :21.00   Max.   :3593.0
##  PRICE_PREMIUM  PRICE_RELATIVE         N             LAMBDA
##  Min.   :  86   Min.   :0.0200   Min.   : 38.0   Min.   :0.0500
##  1st Qu.: 524   1st Qu.:0.1000   1st Qu.:162.0   1st Qu.:0.1200
##  Median :1710   Median :0.3800   Median :227.0   Median :0.1300
##  Mean   :1832   Mean   :0.4926   Mean   :234.2   Mean   :0.1503
##  3rd Qu.:2989   3rd Qu.:0.7475   3rd Qu.:279.0   3rd Qu.:0.1500
##  Max.   :7414   Max.   :1.8900   Max.   :441.0   Max.   :0.5500
##     QUALITY
##  Min.   : 2.000
##  1st Qu.: 6.000
##  Median : 7.000
##  Mean   : 6.716
##  3rd Qu.: 7.000
##  Max.   :10.000
```

```
##Analyzing data with respect to every airline

by(air.df$PRICE_ECONOMY,air.df$AIRLINE,mean)
```

```
## air.df$AIRLINE: AirFrance
## [1] 2769.784
## ------------------------------------------------------------
## air.df$AIRLINE: British
## [1] 1293.48
## ------------------------------------------------------------
## air.df$AIRLINE: Delta
## [1] 560.9348
## ------------------------------------------------------------
## air.df$AIRLINE: Jet
## [1] 269.6769
## ------------------------------------------------------------
## air.df$AIRLINE: Singapore
## [1] 860.25
## ------------------------------------------------------------
## air.df$AIRLINE: Virgin
## [1] 1603.532
```

```
by(air.df$PRICE_PREMIUM,air.df$AIRLINE,mean)
```

```
## air.df$AIRLINE: AirFrance
## [1] 3065.216
## ------------------------------------------------------------
## air.df$AIRLINE: British
## [1] 1937.029
## ------------------------------------------------------------
## air.df$AIRLINE: Delta
## [1] 684.6739
## ------------------------------------------------------------
## air.df$AIRLINE: Jet
## [1] 475.4
## ------------------------------------------------------------
## air.df$AIRLINE: Singapore
## [1] 1239.925
## ------------------------------------------------------------
## air.df$AIRLINE: Virgin
## [1] 2721.694
```

```
by(air.df$SEATS_ECONOMY,air.df$AIRLINE,mean)
```

```
## air.df$AIRLINE: AirFrance
## [1] 214.4595
## ------------------------------------------------------------
## air.df$AIRLINE: British
## [1] 216.5886
## ------------------------------------------------------------
## air.df$AIRLINE: Delta
## [1] 137.2174
## ------------------------------------------------------------
## air.df$AIRLINE: Jet
## [1] 132.7231
## ------------------------------------------------------------
## air.df$AIRLINE: Singapore
## [1] 243.6
## ------------------------------------------------------------
## air.df$AIRLINE: Virgin
## [1] 230.1774
```

```
by(air.df$SEATS_PREMIUM,air.df$AIRLINE,mean)
```

```
## air.df$AIRLINE: AirFrance
## [1] 26.7027
## ------------------------------------------------------------
## air.df$AIRLINE: British
## [1] 43.18286
## ------------------------------------------------------------
## air.df$AIRLINE: Delta
## [1] 22.56522
## ------------------------------------------------------------
## air.df$AIRLINE: Jet
## [1] 15.98462
## ------------------------------------------------------------
## air.df$AIRLINE: Singapore
## [1] 31.2
## ------------------------------------------------------------
## air.df$AIRLINE: Virgin
## [1] 42.53226
```

```
by(air.df$WIDTH_ECONOMY,air.df$AIRLINE,mean)
```

```
## air.df$AIRLINE: AirFrance
## [1] 17.56757
## ------------------------------------------------------------
## air.df$AIRLINE: British
## [1] 18
## ------------------------------------------------------------
## air.df$AIRLINE: Delta
## [1] 17.3913
## ------------------------------------------------------------
## air.df$AIRLINE: Jet
## [1] 17.10769
## ------------------------------------------------------------
## air.df$AIRLINE: Singapore
## [1] 19
## ------------------------------------------------------------
## air.df$AIRLINE: Virgin
## [1] 18
```

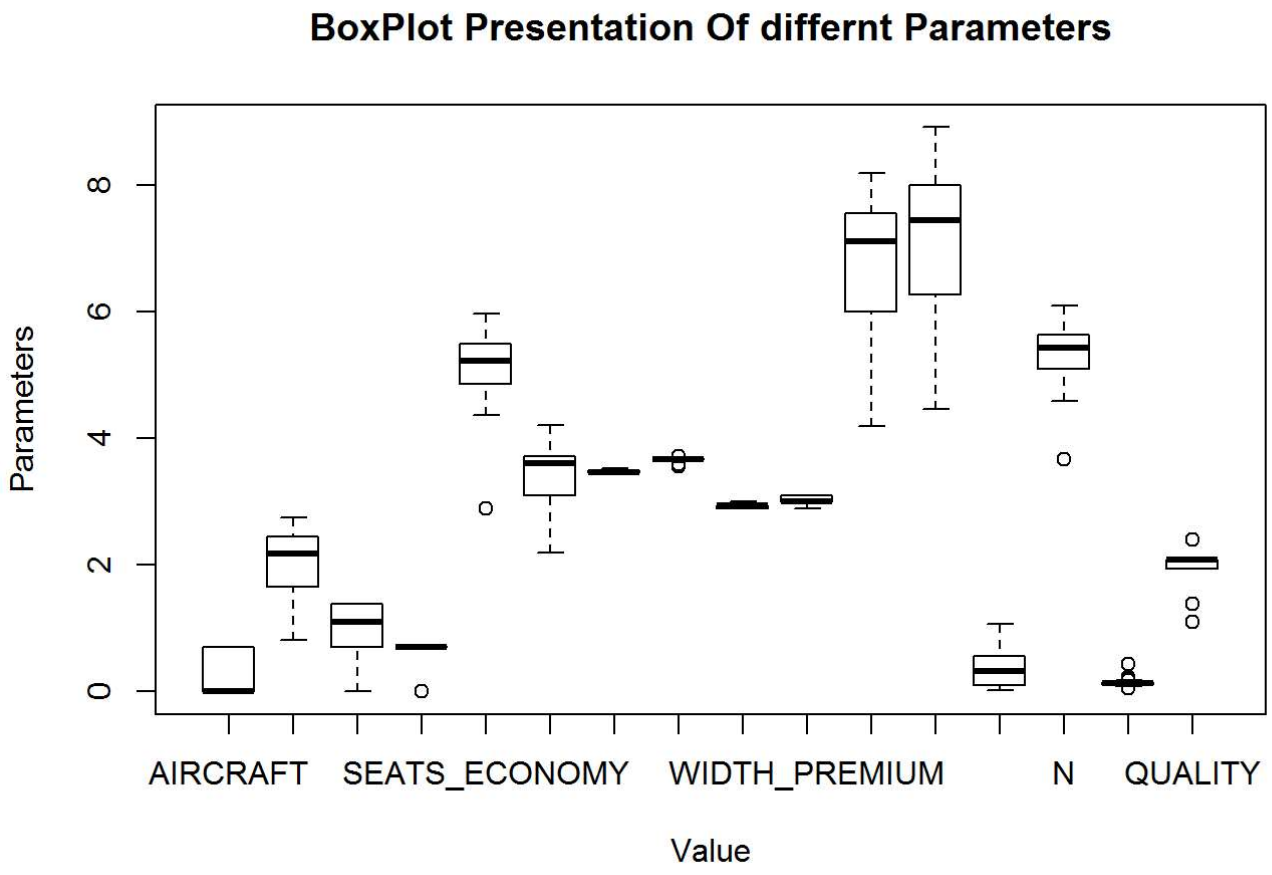
```
by(air.df$WIDTH_PREMIUM,air.df$AIRLINE,mean)
```

```
## air.df$AIRLINE: AirFrance
## [1] 19
## ------------------------------------------------------------
## air.df$AIRLINE: British
## [1] 19
## ------------------------------------------------------------
## air.df$AIRLINE: Delta
## [1] 17.78261
## ------------------------------------------------------------
## air.df$AIRLINE: Jet
## [1] 20.78462
## ------------------------------------------------------------
## air.df$AIRLINE: Singapore
## [1] 20
## ------------------------------------------------------------
## air.df$AIRLINE: Virgin
## [1] 21
```

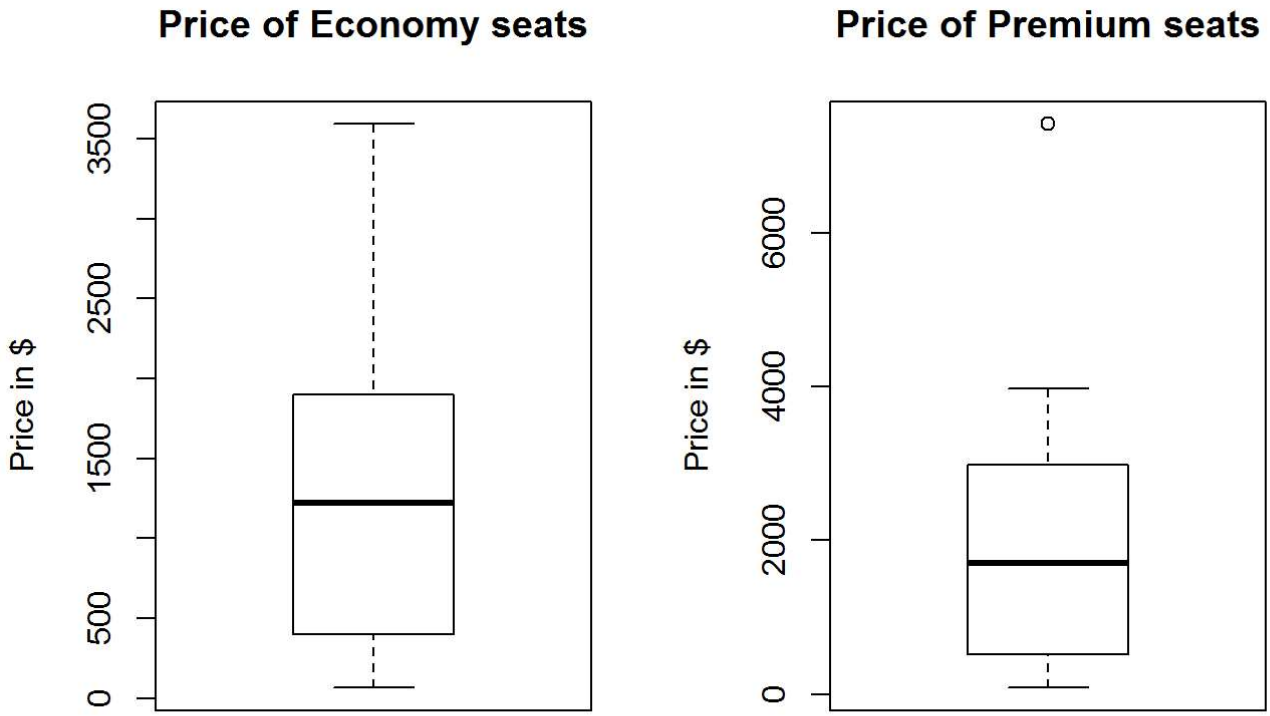
```
#Taking information about the type of flights , whether international(1) or domestic(0) , in the
 data set
table(air.df$INTERNATIONAL)
```

```
##
##   0   1
##  40 422
```

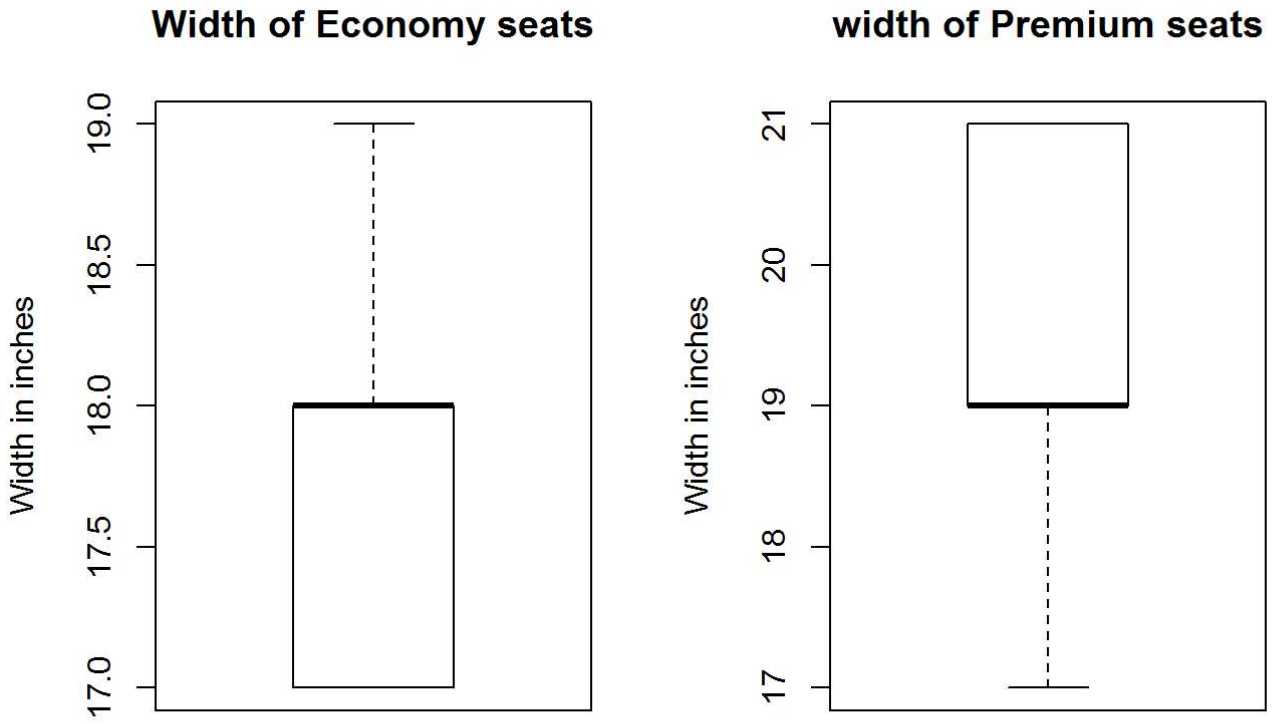
```
##Drawing boxplots among comparable parameters
#Reading variables except airline beacause  it's a string
air1 = air.df[,2:17]
View(air1)
##Taking log
air2=log(air1+1)

 boxplot(air2, xlab="Value", ylab="Parameters", main="BoxPlot Presentation Of differnt Parameter
s")
```



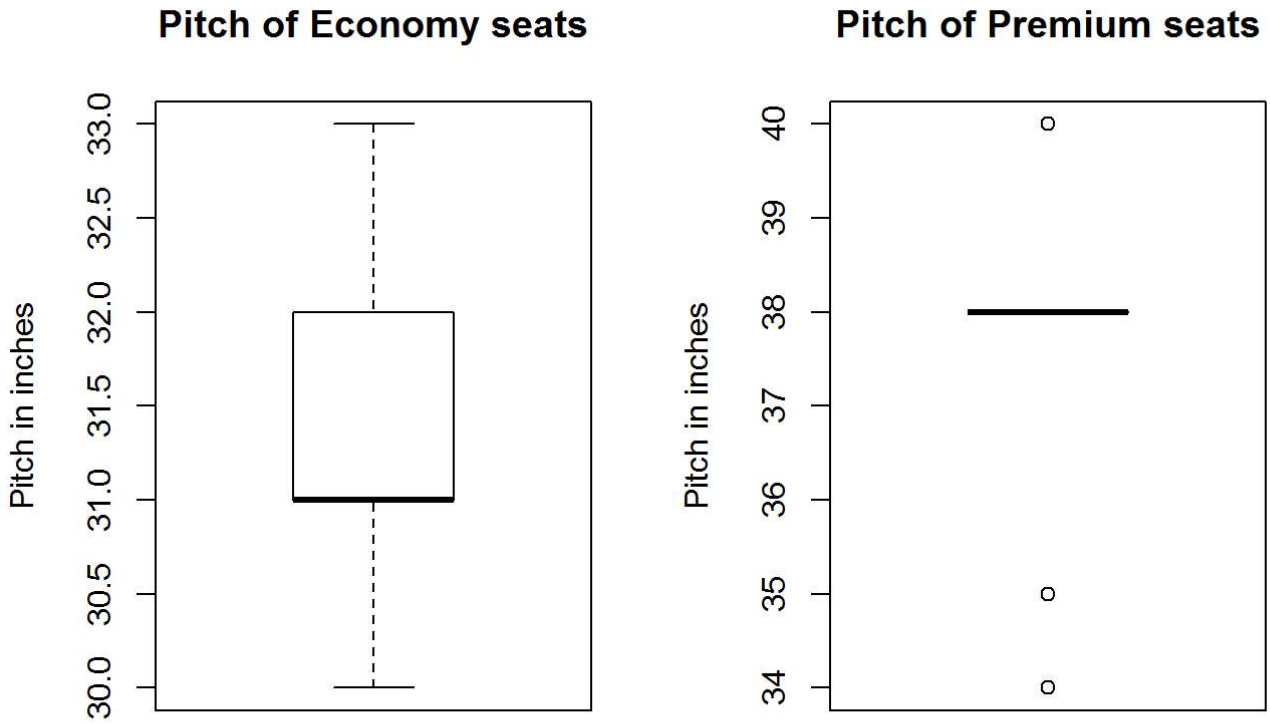

```
##Individual boxplots for comparable parameters

 par(mfrow=c(1,2))
 with(air.df, boxplot(air.df$PRICE_ECONOMY,main="Price of Economy seats",ylab="Price in $"))
 with(air.df, boxplot(air.df$PRICE_PREMIUM,main="Price of Premium seats",ylab="Price in $"))
```



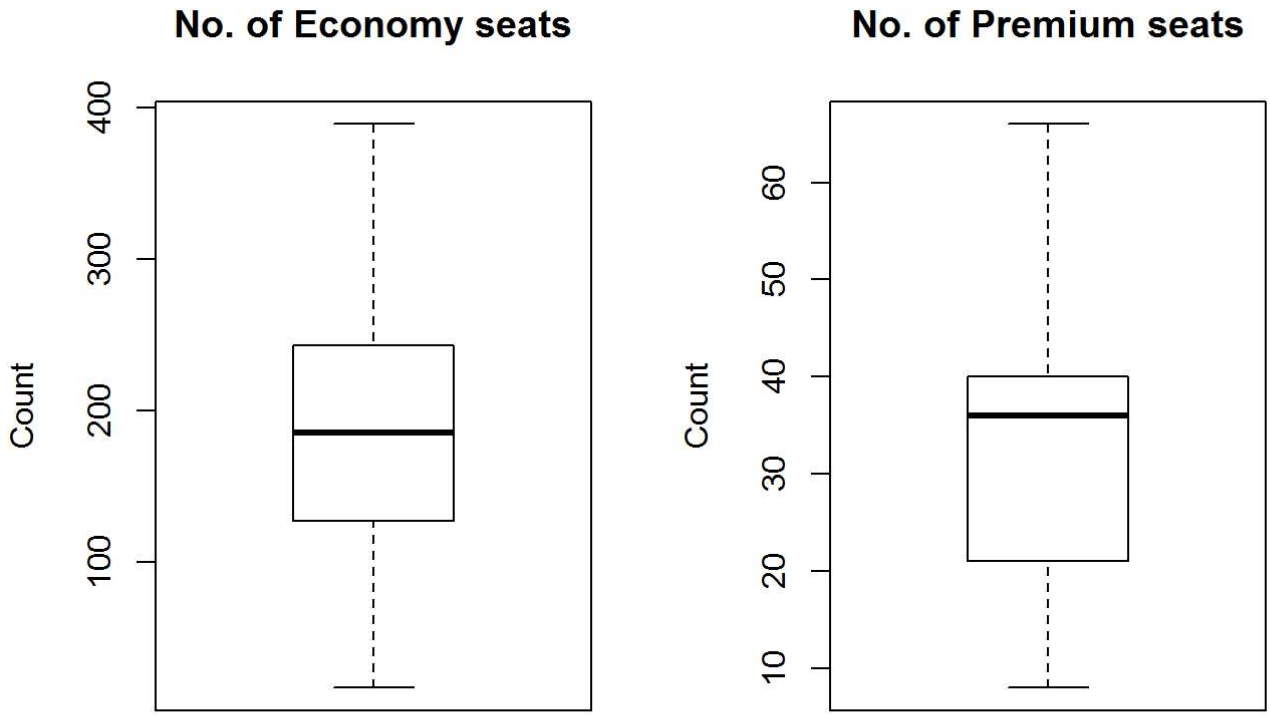

```
 par(mfrow=c(1,1))

 par(mfrow=c(1,2))
 with(air.df, boxplot(air.df$WIDTH_ECONOMY,main="Width of Economy seats",ylab="Width in
inches"))
 with(air.df, boxplot(air.df$WIDTH_PREMIUM,main="width of Premium seats",ylab="Width in
inches"))
```



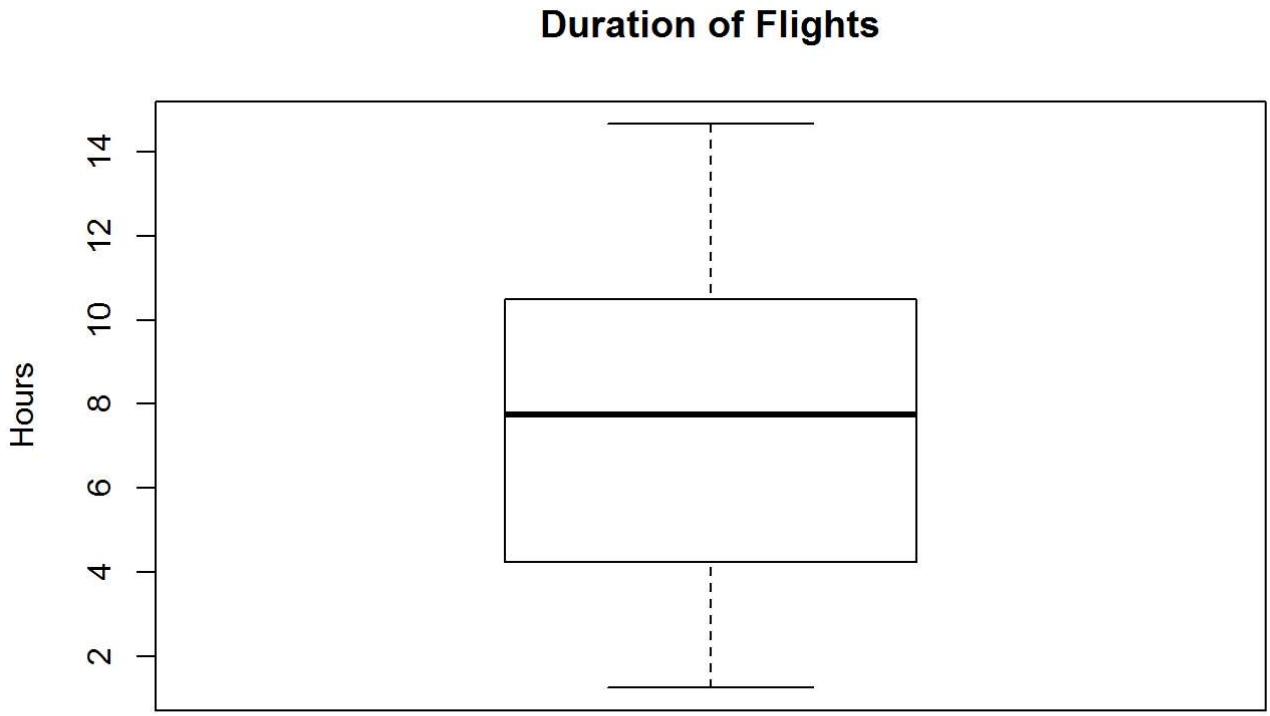

```
 par(mfrow=c(1,1))

 par(mfrow=c(1,2))
 with(air.df, boxplot(air.df$PITCH_ECONOMY,main="Pitch of Economy seats",ylab="Pitch in
inches"))
 with(air.df, boxplot(air.df$PITCH_PREMIUM,main="Pitch of Premium seats",ylab="Pitch in
inches"))
```



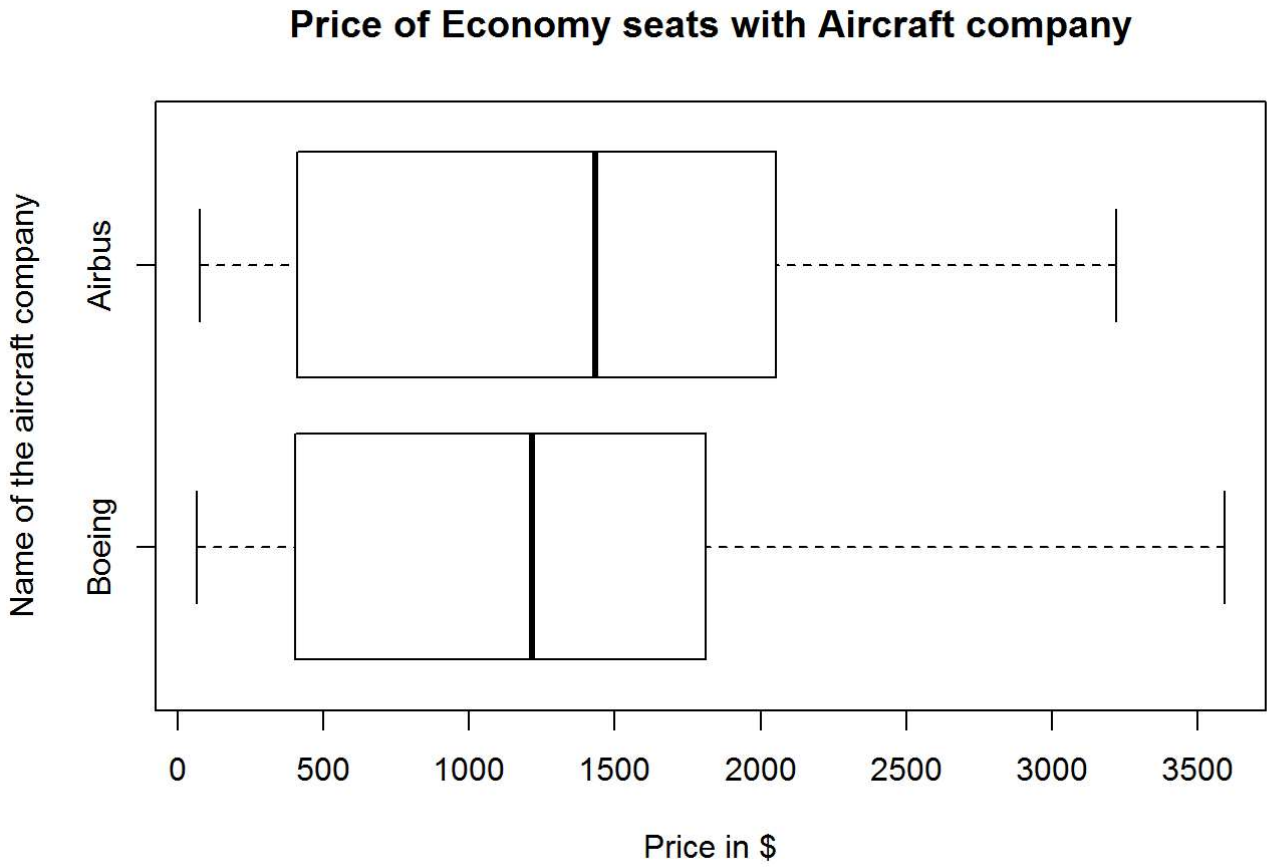

```
par(mfrow=c(1,1))

par(mfrow=c(1,2))
with(air.df, boxplot(air.df$SEATS_ECONOMY,main="No. of Economy seats",ylab="Count"))
with(air.df, boxplot(air.df$SEATS_PREMIUM,main="No. of Premium seats",ylab="Count"))
```



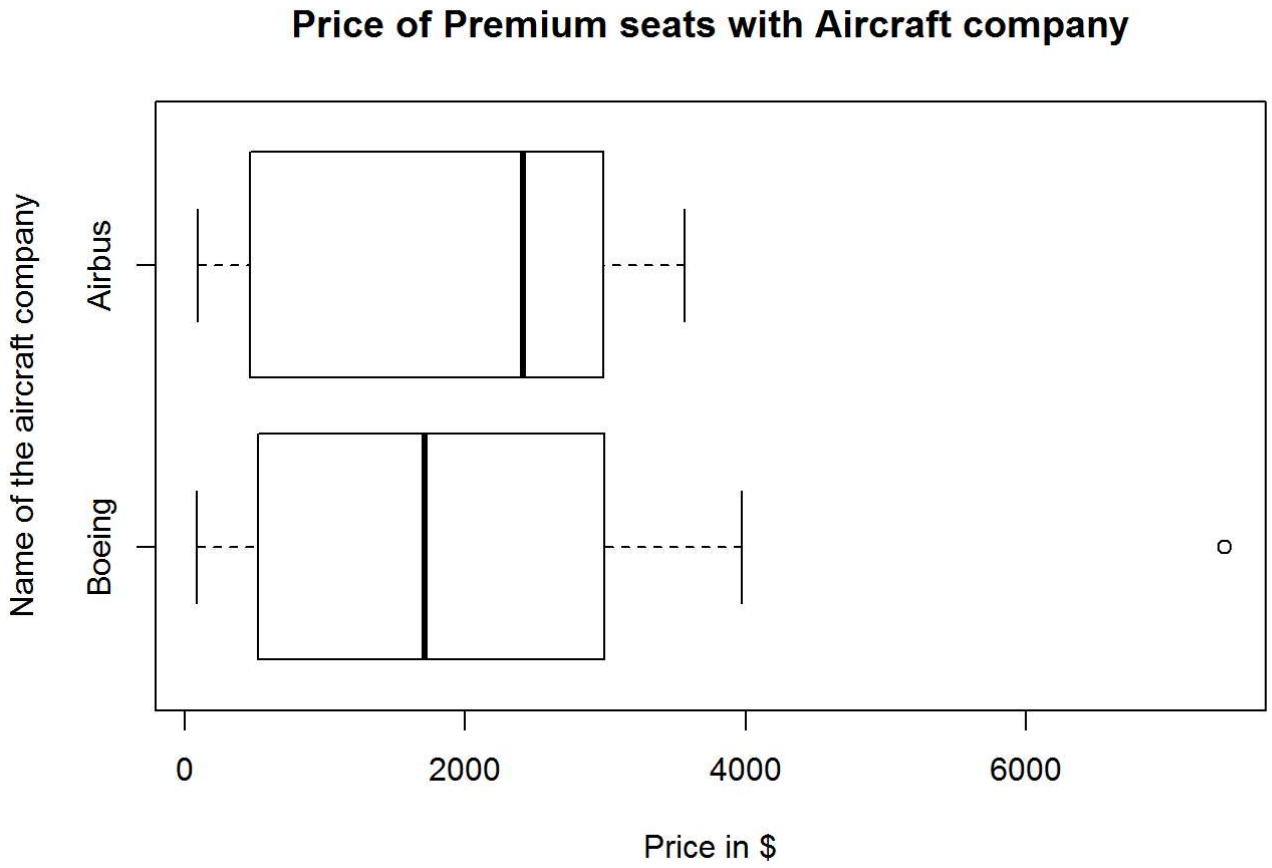

```
par(mfrow=c(1,1))

boxplot(air.df$FLIGHT_DURATION, main="Duration of Flights",ylab="Hours")
```



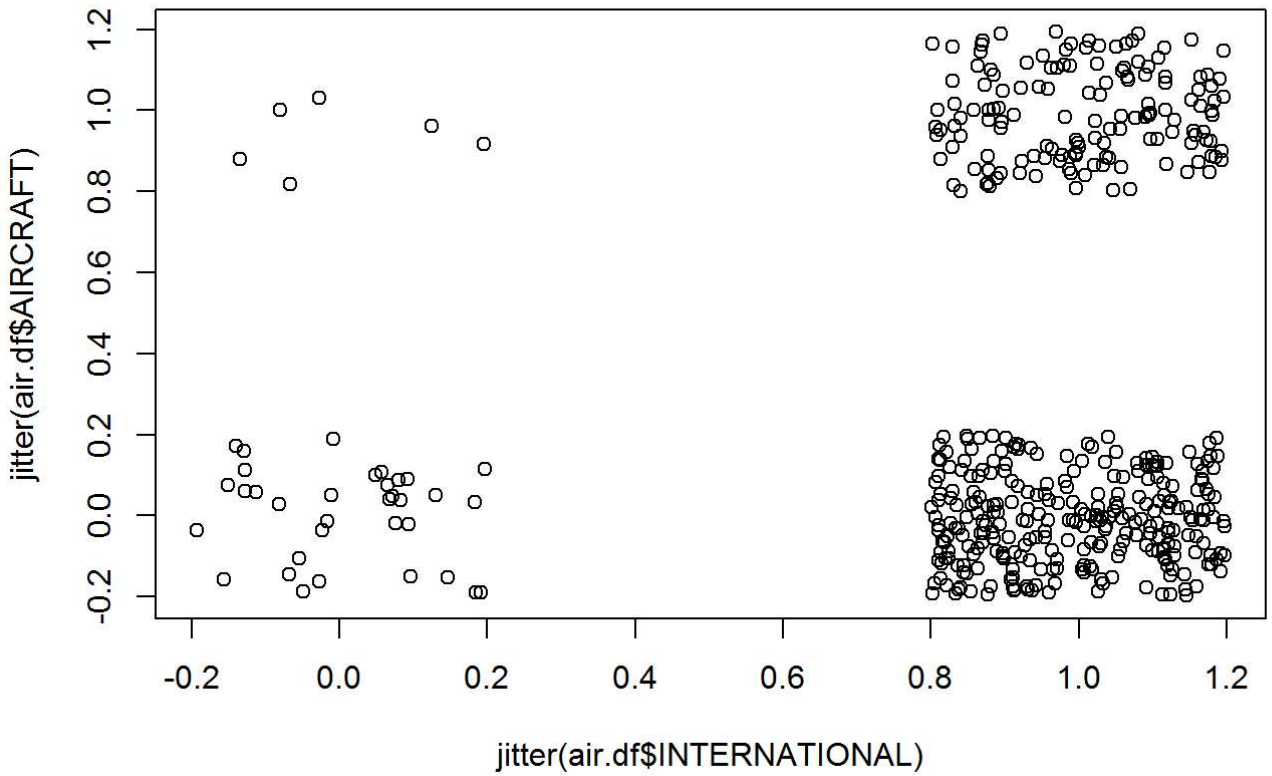

```
 ##Price of seats varying with aircart company
 boxplot(air.df$PRICE_ECONOMY~air.df$AIRCRAFT,yaxt="n", horizontal=TRUE,main="Price of Economy s
eats with Aircraft company",xlab="Price in $", ylab="Name of the aircraft company")
 axis(side=2, at=c(1,2), labels = c("Boeing", "Airbus"))
```



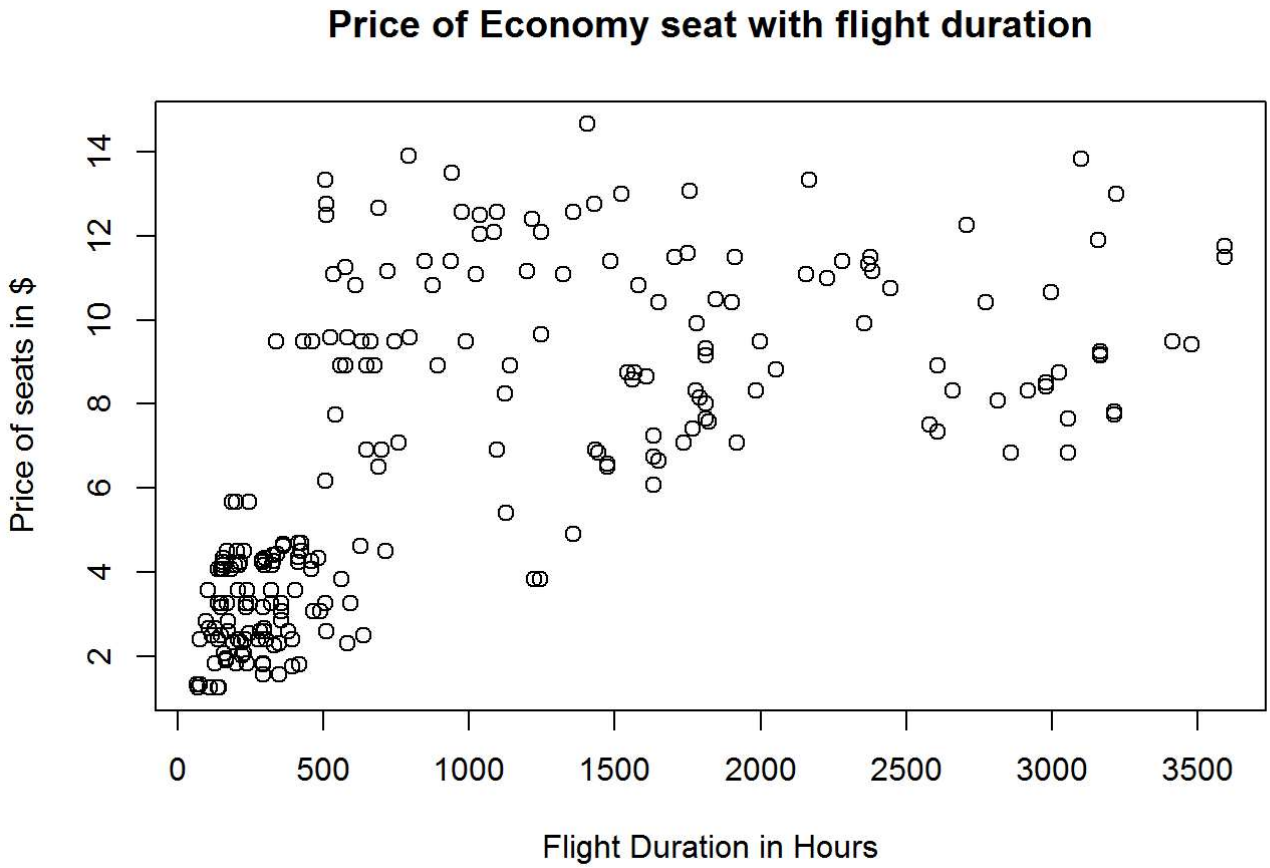

```
 boxplot(air.df$PRICE_PREMIUM~air.df$AIRCRAFT,yaxt="n", horizontal=TRUE,main="Price of Premium s
eats with Aircraft company",xlab="Price in $", ylab="Name of the aircraft company")
 axis(side=2, at=c(1,2), labels = c("Boeing", "Airbus"))
```




```
##Plots

 plot(jitter(air.df$INTERNATIONAL), jitter(air.df$AIRCRAFT))
```




```
 ##PRICE OF SEAT VARYING WITH FLIGHT DURATION
 plot(air.df$FLIGHT_DURATION~air.df$PRICE_ECONOMY,main="Price of Economy seat with flight durati
on",xlab = "Flight Duration in Hours", ylab="Price of seats in $", cex=1.1)
```

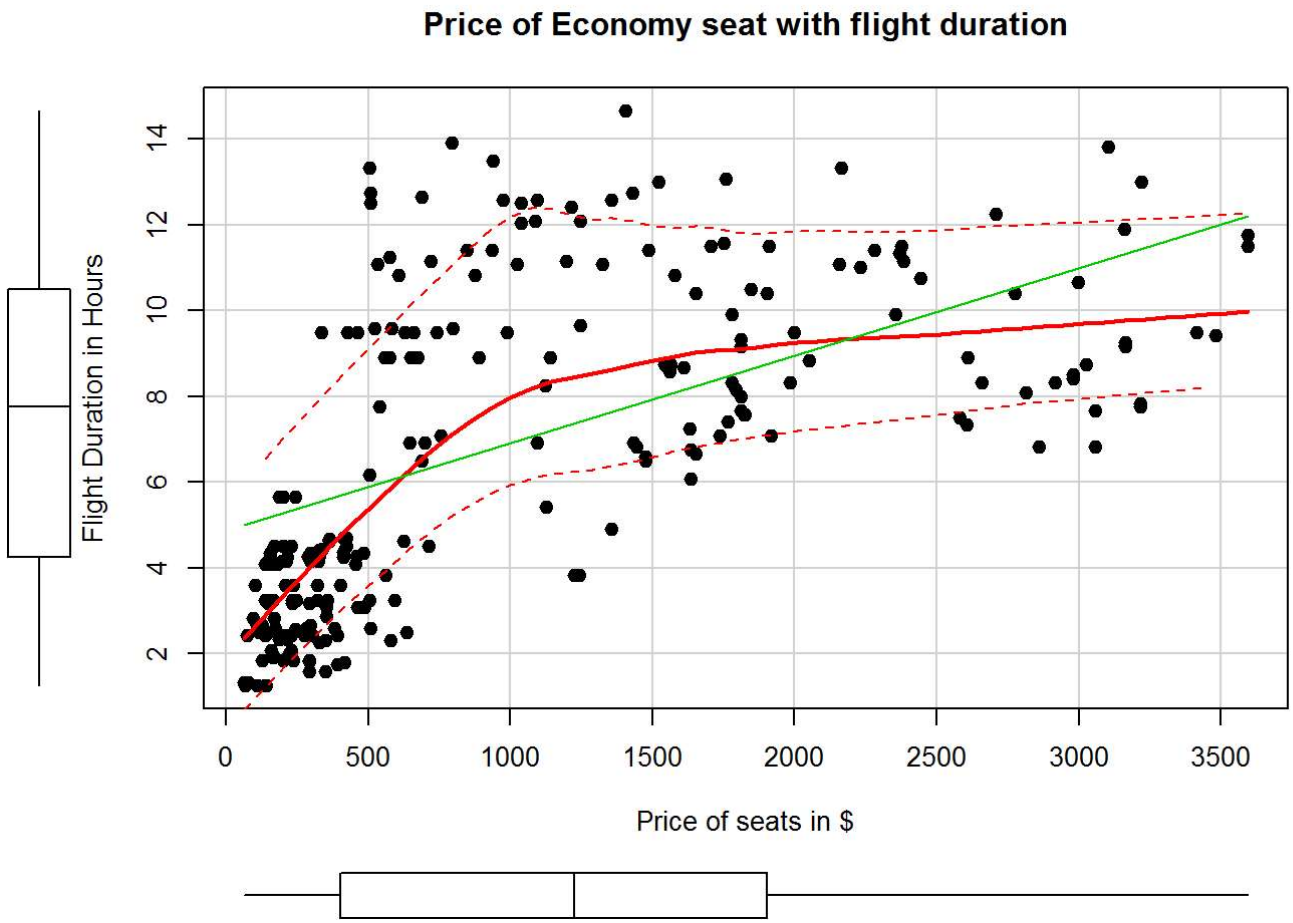



```
library(car)
```

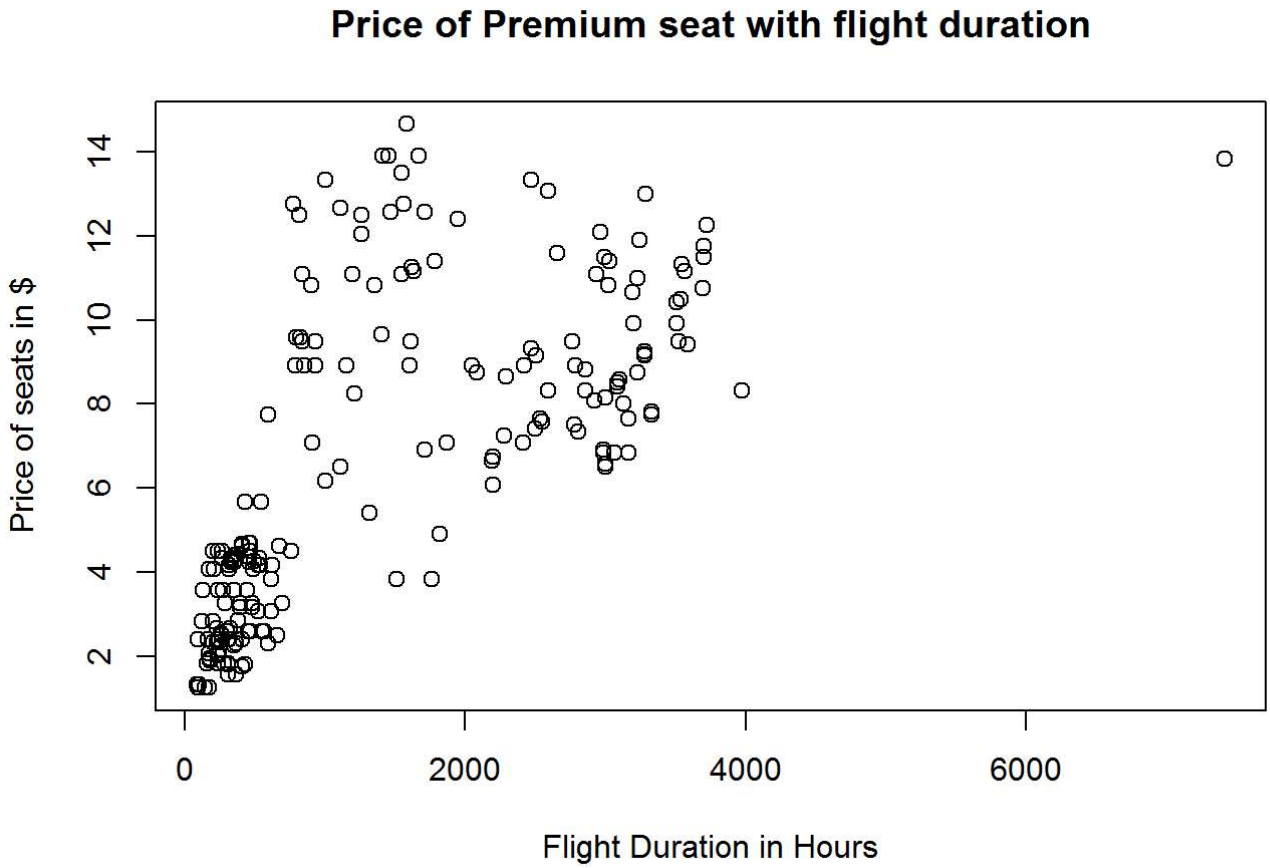
```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

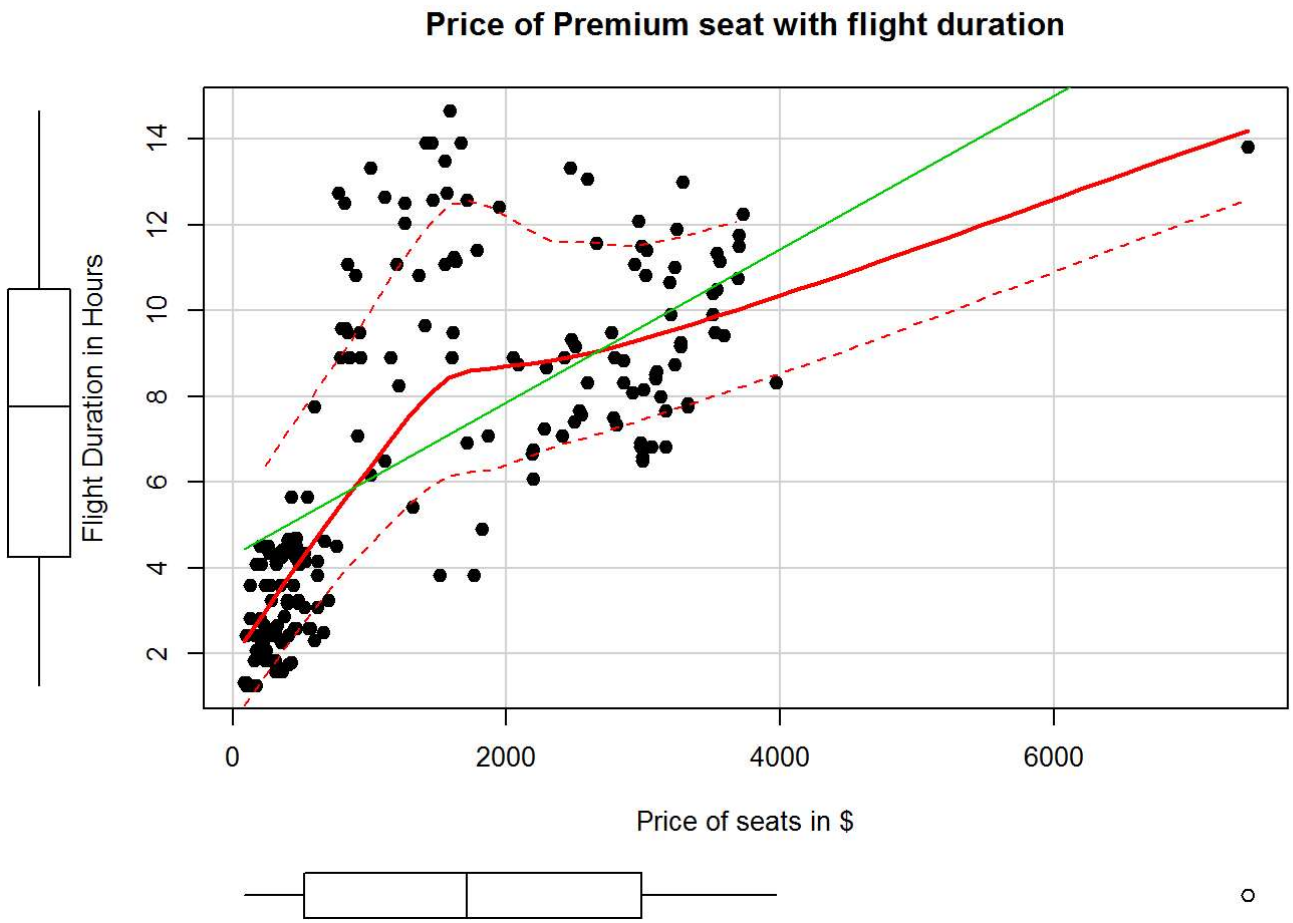
```
scatterplot(air.df$PRICE_ECONOMY,air.df$FLIGHT_DURATION,main="Price of Economy seat with flight
duration",ylab = "Flight Duration in Hours", xlab="Price of seats in $",cex=1.1,pch=19)
```



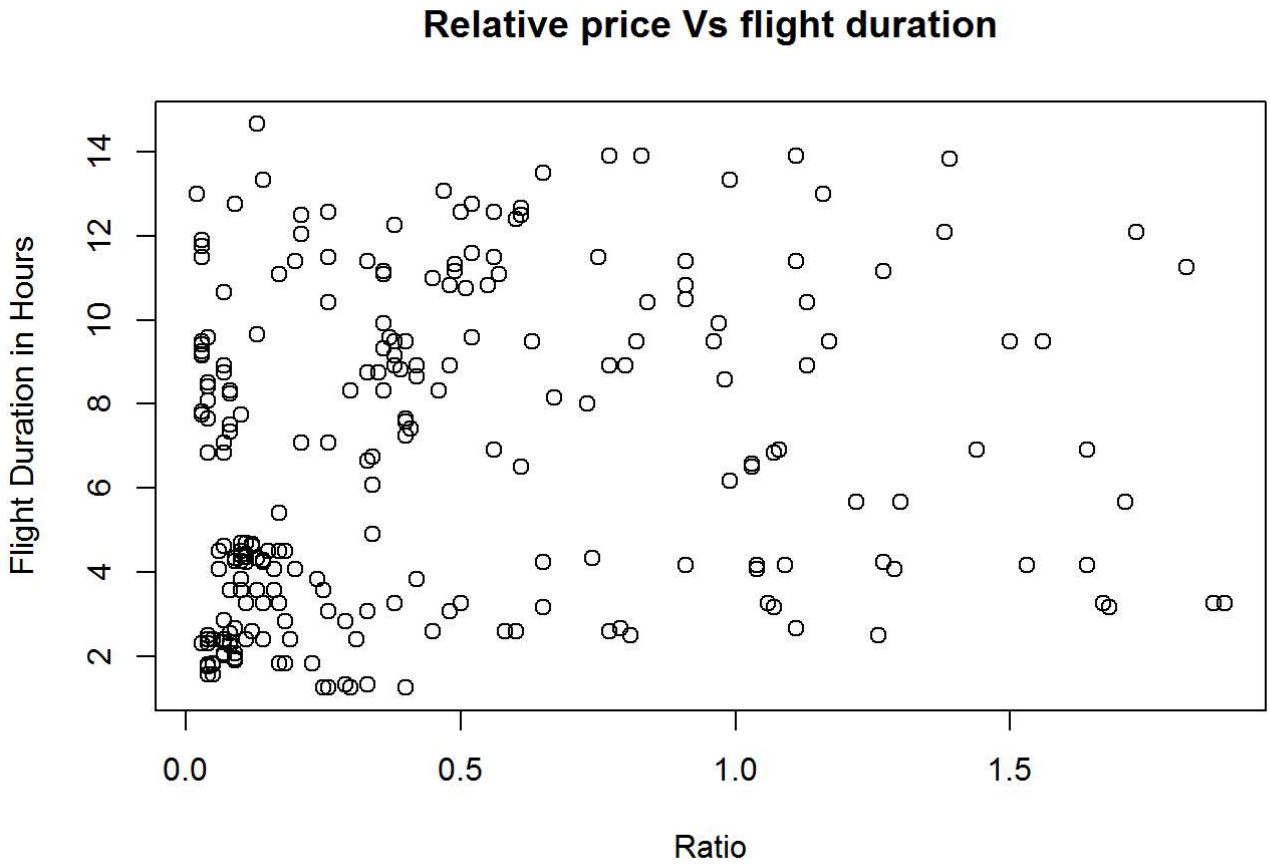

```
 plot(air.df$FLIGHT_DURATION~air.df$PRICE_PREMIUM,main="Price of Premium seat with flight durati
on",xlab = "Flight Duration in Hours", ylab="Price of seats in $", cex=1.1)
```



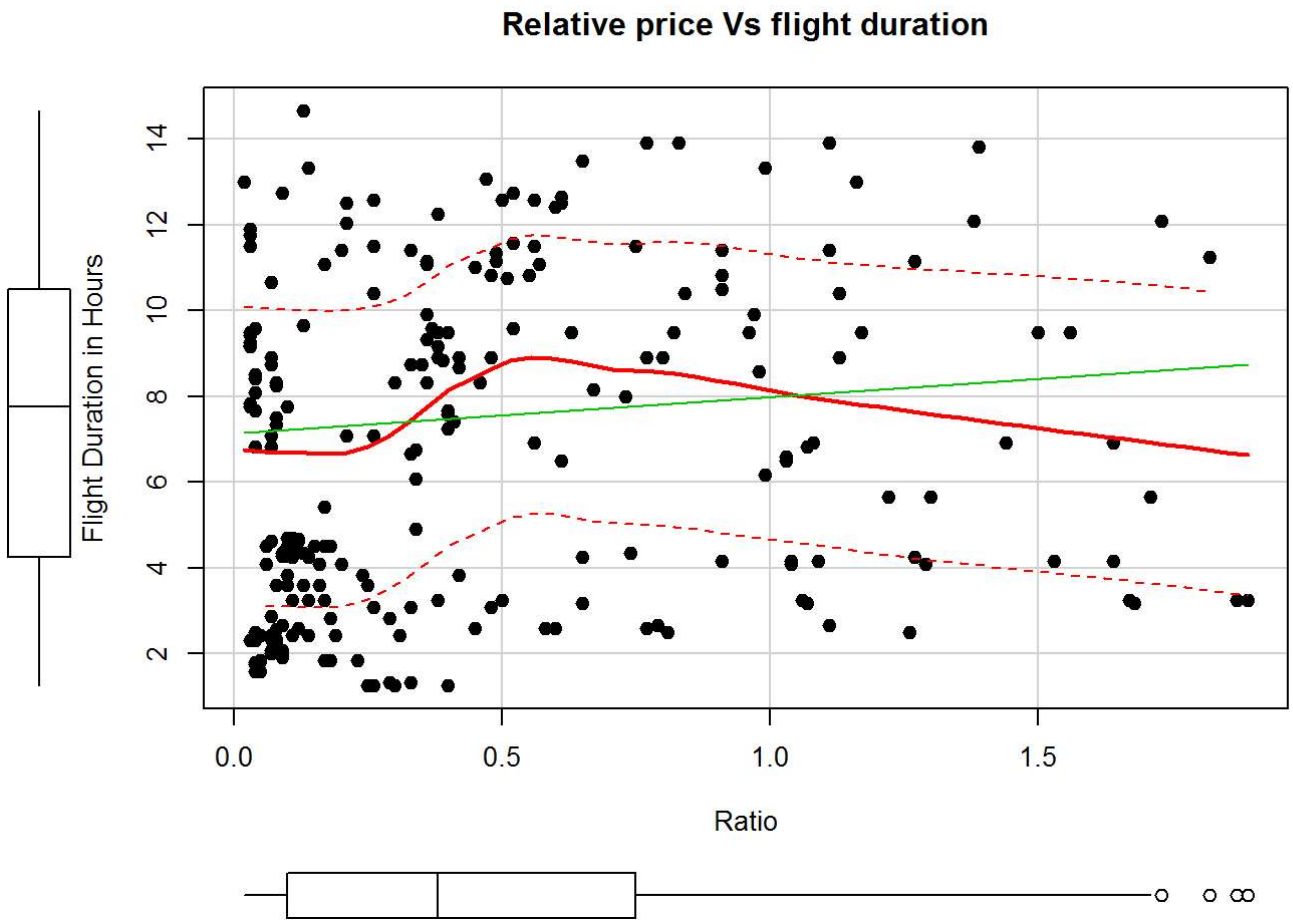

```
scatterplot(air.df$PRICE_PREMIUM,air.df$FLIGHT_DURATION,main="Price of Premium seat with flight
duration",ylab = "Flight Duration in Hours", xlab="Price of seats in $",cex=1.1,pch=19)
```



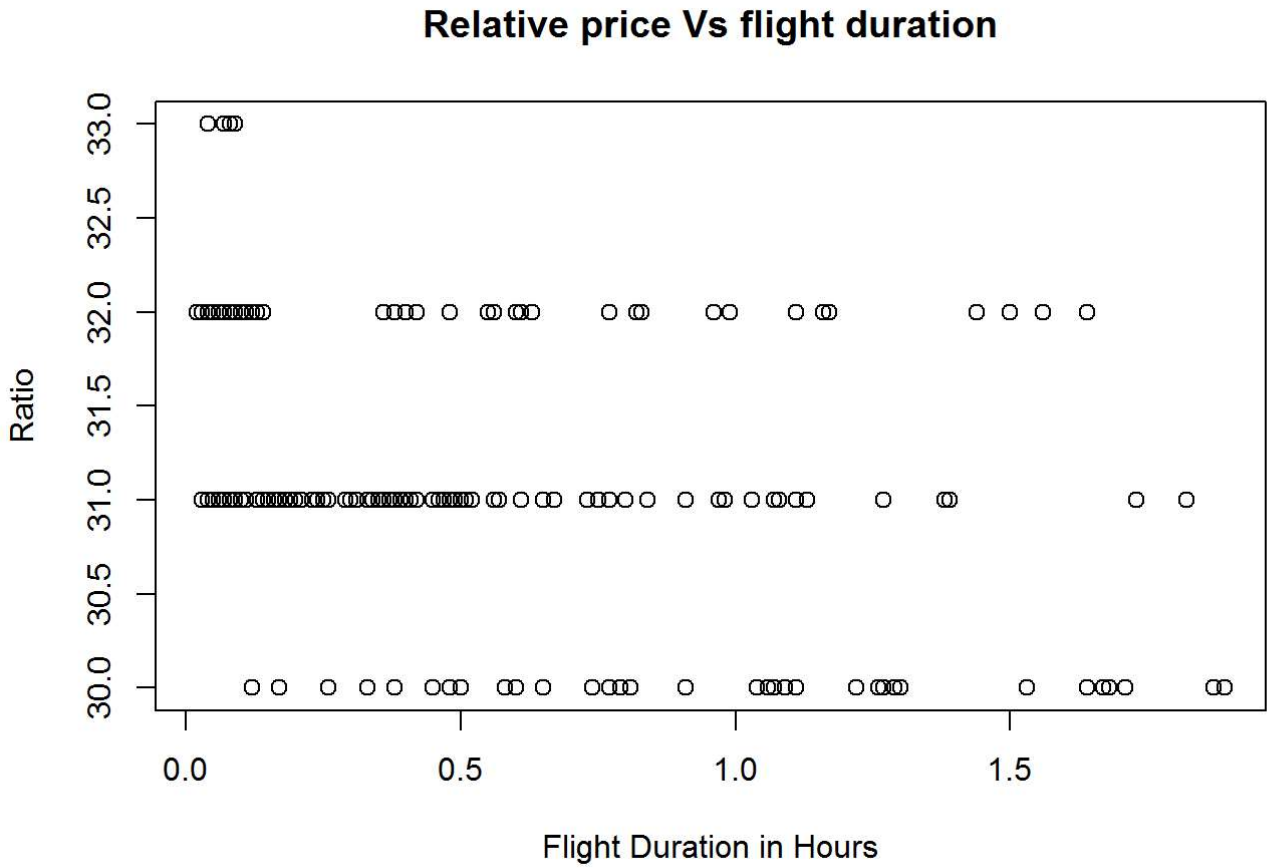

```
plot(air.df$FLIGHT_DURATION~air.df$PRICE_RELATIVE,main="Relative price Vs flight duration",ylab
= "Flight Duration in Hours", xlab="Ratio", cex=1.1)
```



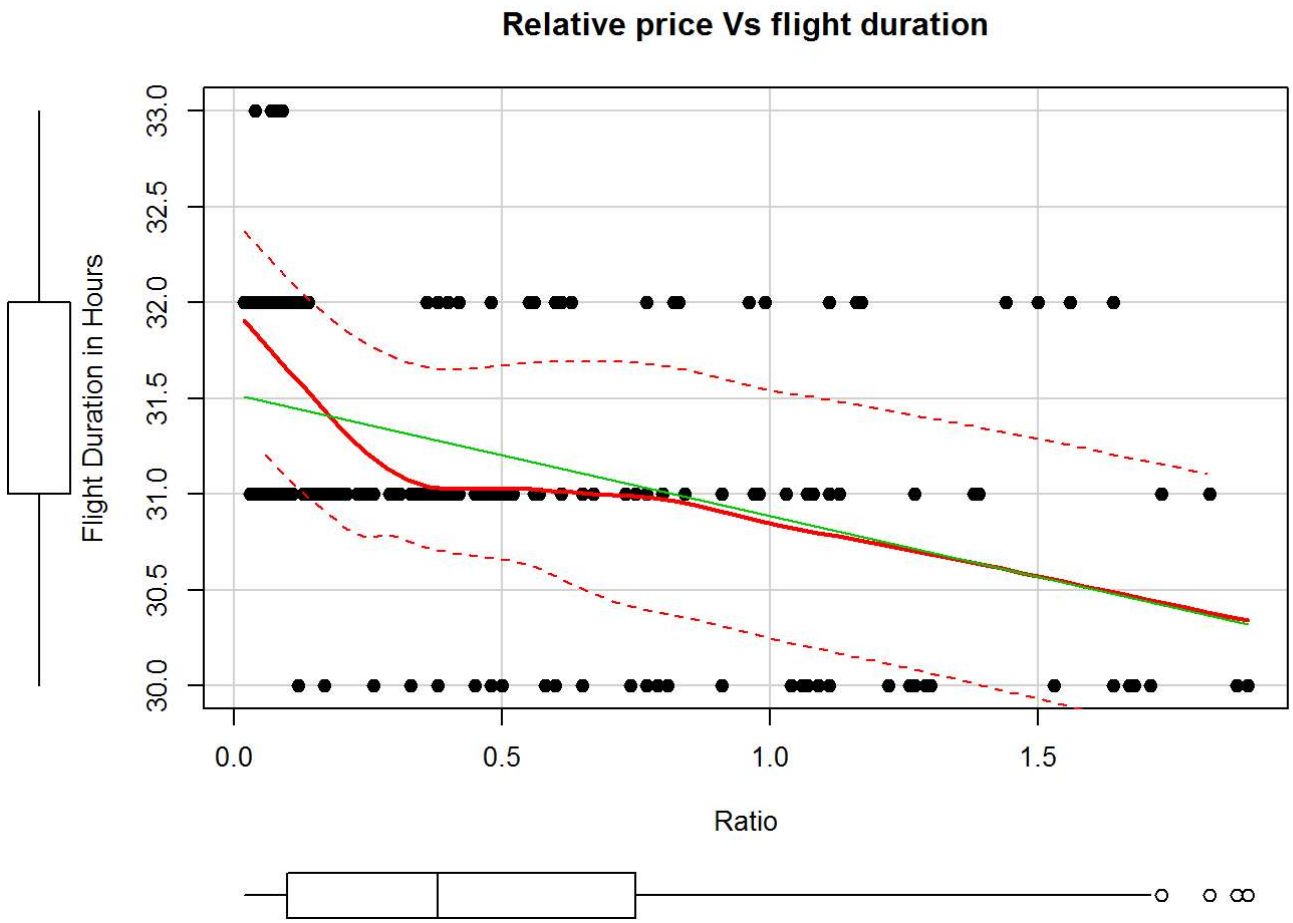

```
 scatterplot(air.df$PRICE_RELATIVE,air.df$FLIGHT_DURATION,main="Relative price Vs flight duratio
n",ylab = "Flight Duration in Hours", xlab="Ratio",cex=1.1,pch=19)
```



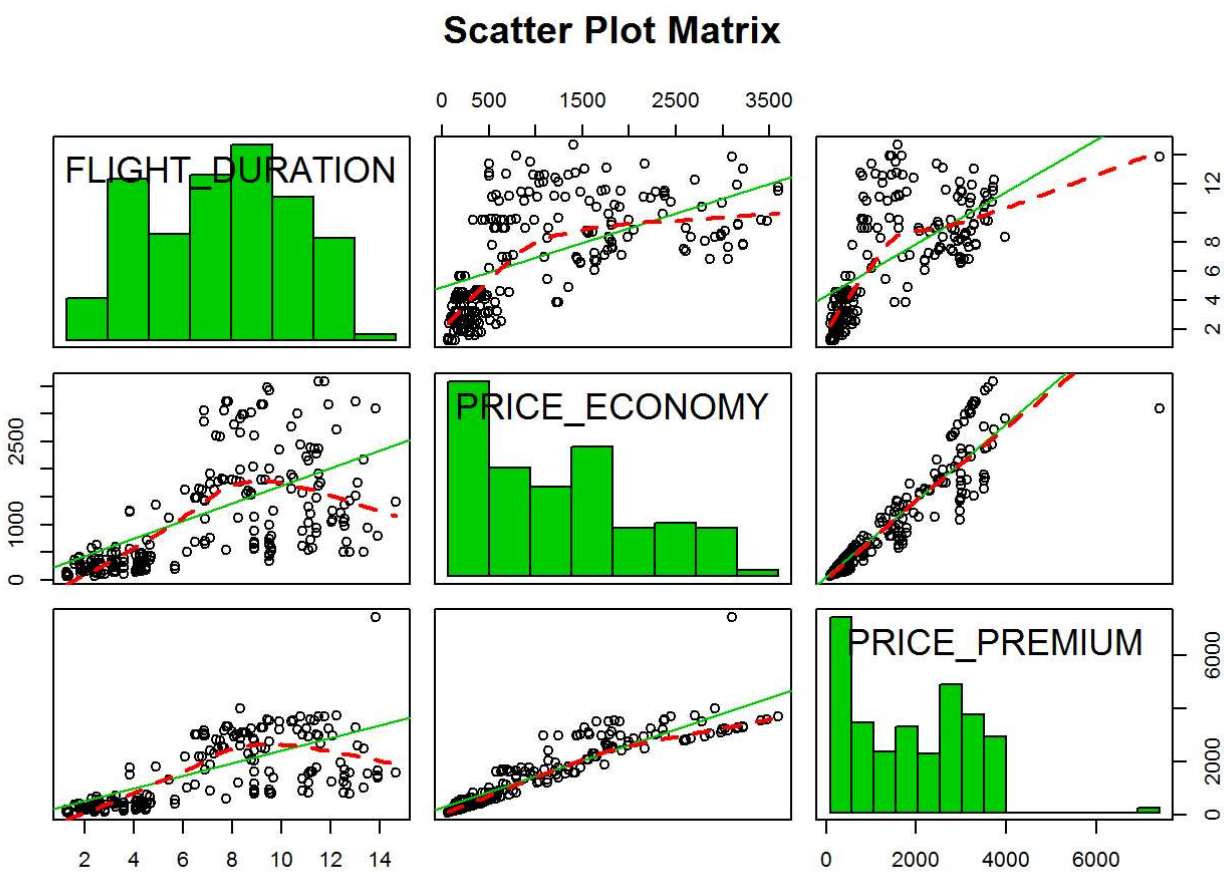

```
plot(air.df$PITCH_ECONOMY~air.df$PRICE_RELATIVE,main="Relative price Vs flight duration",xlab =
"Flight Duration in Hours", ylab="Ratio", cex=1.1)
```




```
 scatterplot(air.df$PRICE_RELATIVE,air.df$PITCH_ECONOMY,main="Relative price Vs flight
duration",ylab = "Flight Duration in Hours", xlab="Ratio",cex=1.1,pch=19)
```




```
library(car)
scatterplotMatrix(
  air.df[
    ,c("FLIGHT_DURATION","PRICE_ECONOMY","PRICE_PREMIUM")],
  spread=FALSE, smoother.args=list(lty=2),
  main="Scatter Plot Matrix", diagonal = "histogram")
```

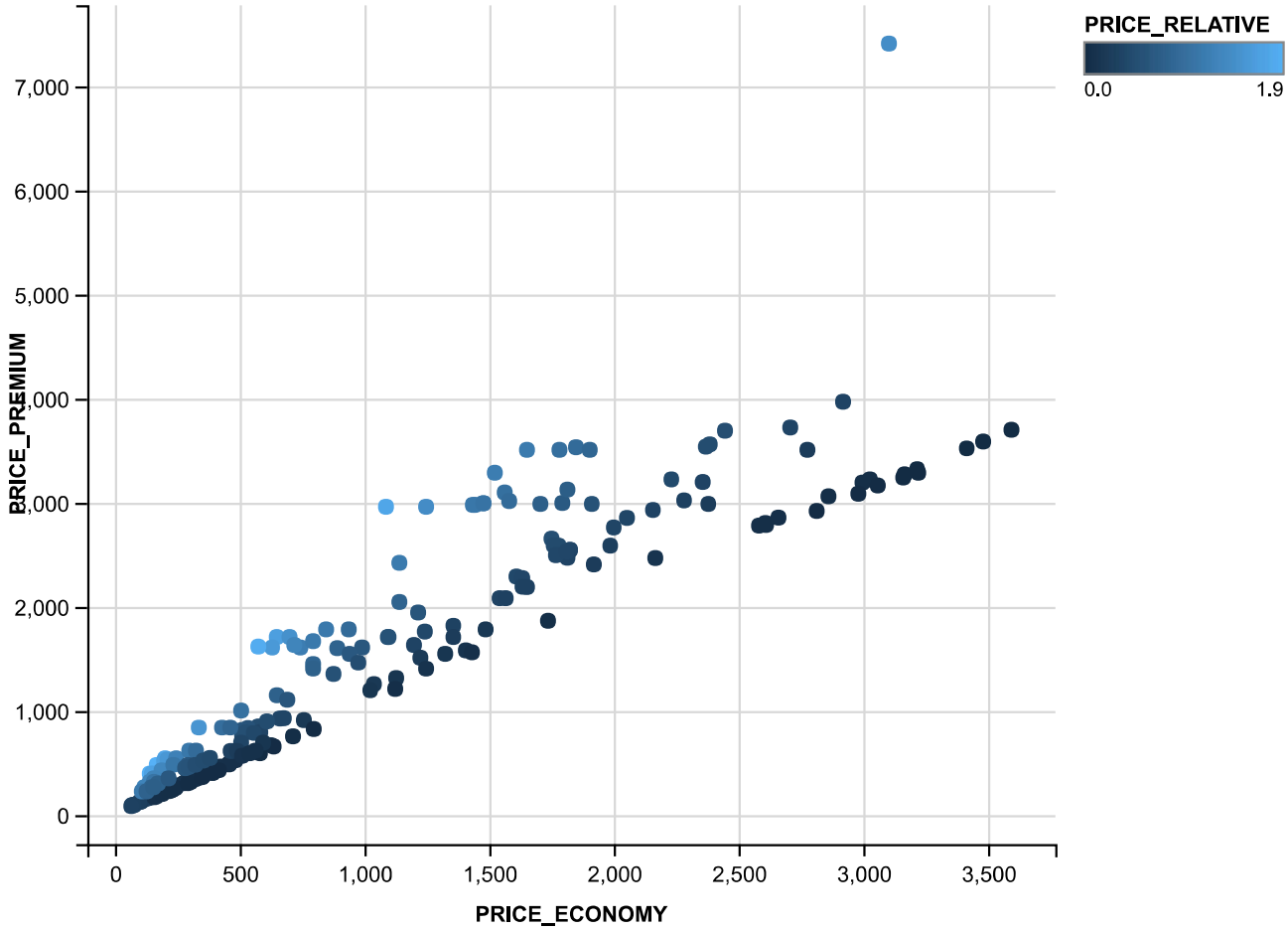



```
#Visualizing the data through ggvis

library(ggvis)
```

```
## Warning: package 'ggvis' was built under R version 3.4.1
```

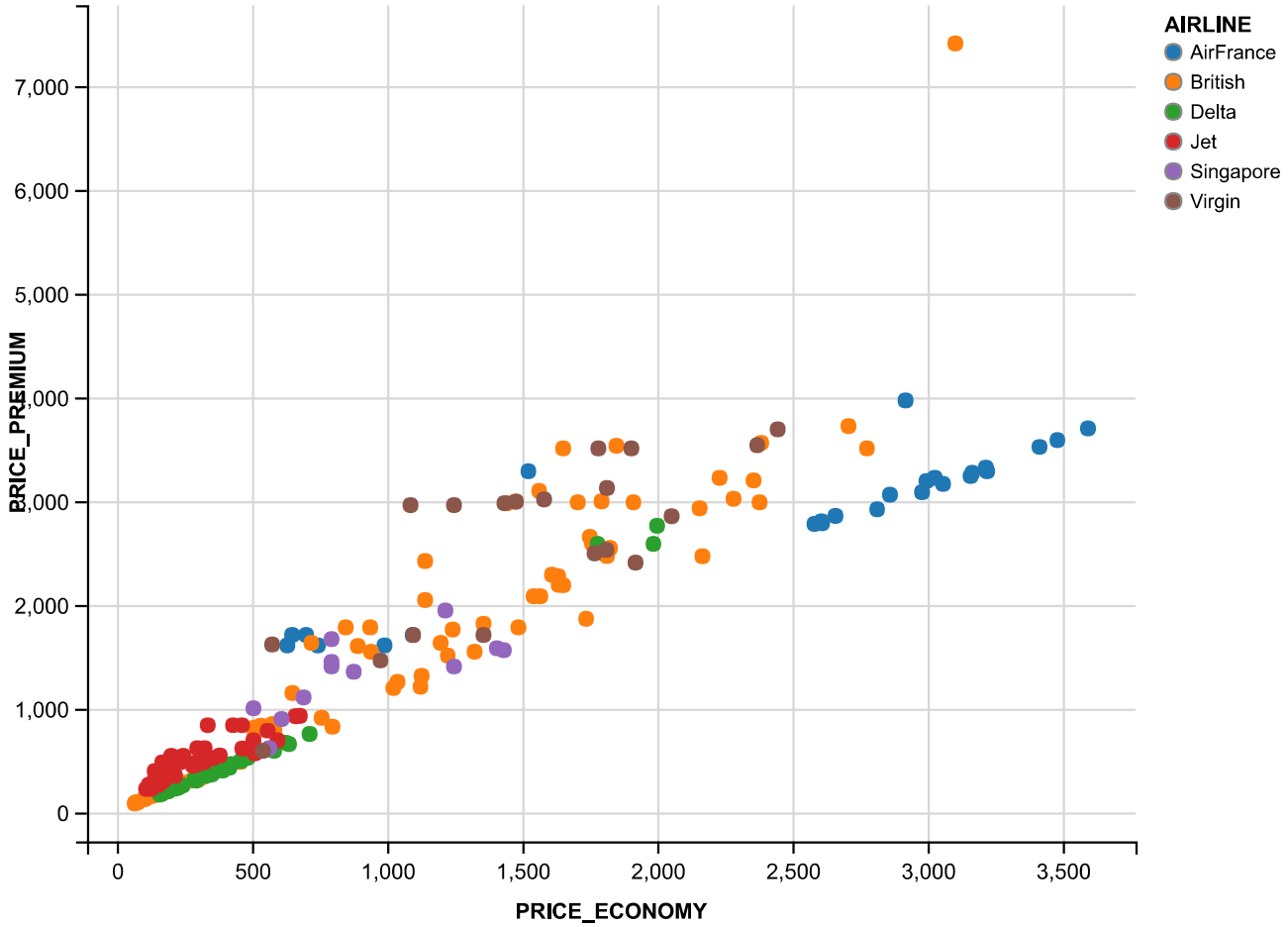
```
ggvis(~PRICE_ECONOMY,~PRICE_PREMIUM,fill=~PRICE_RELATIVE,data=air.df)
```

```
## Guessing layer_points()
```



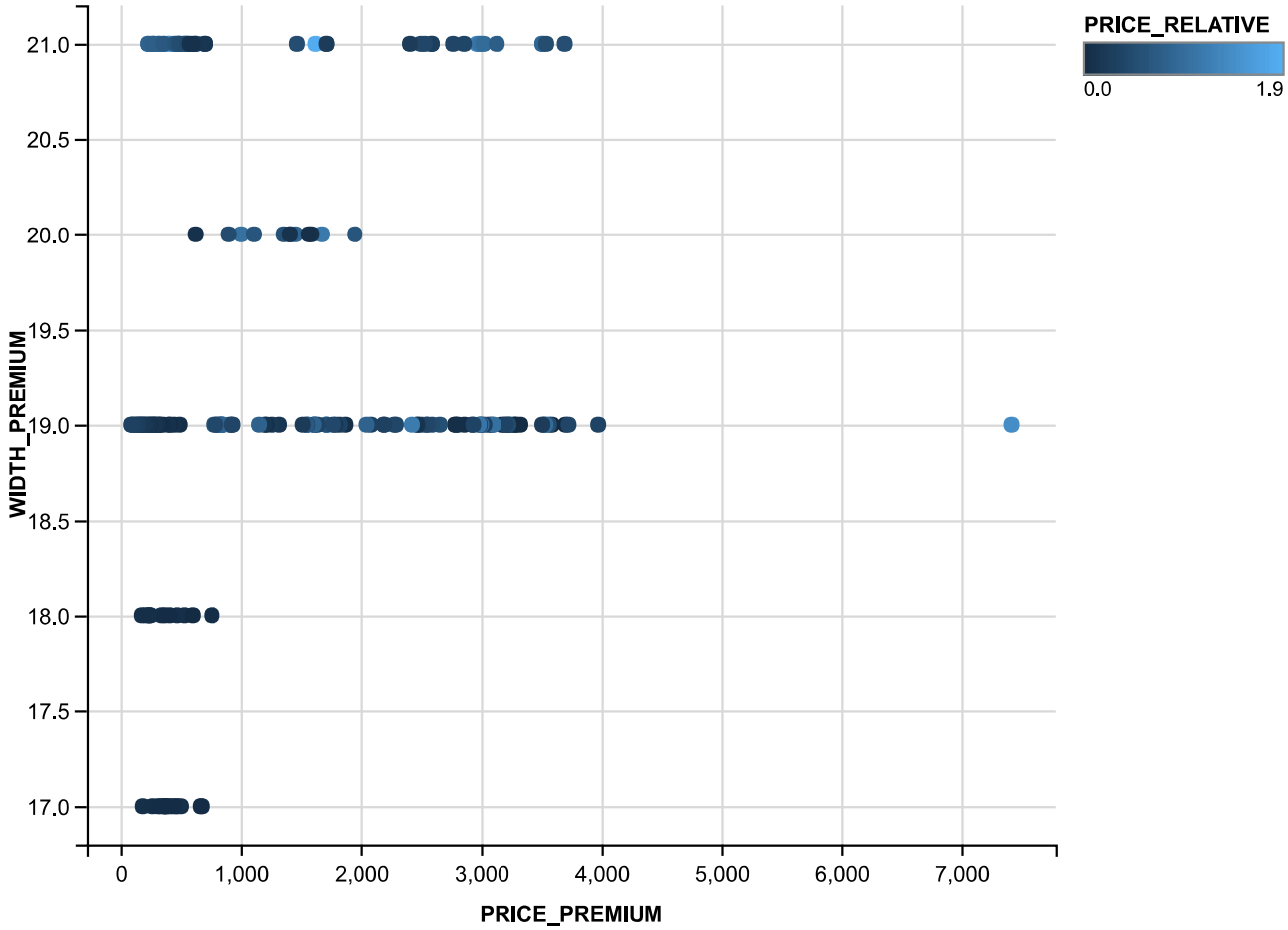

```
ggvis(~PRICE_ECONOMY,~PRICE_PREMIUM,fill=~AIRLINE,data=air.df)
```

```
## Guessing layer_points()
```



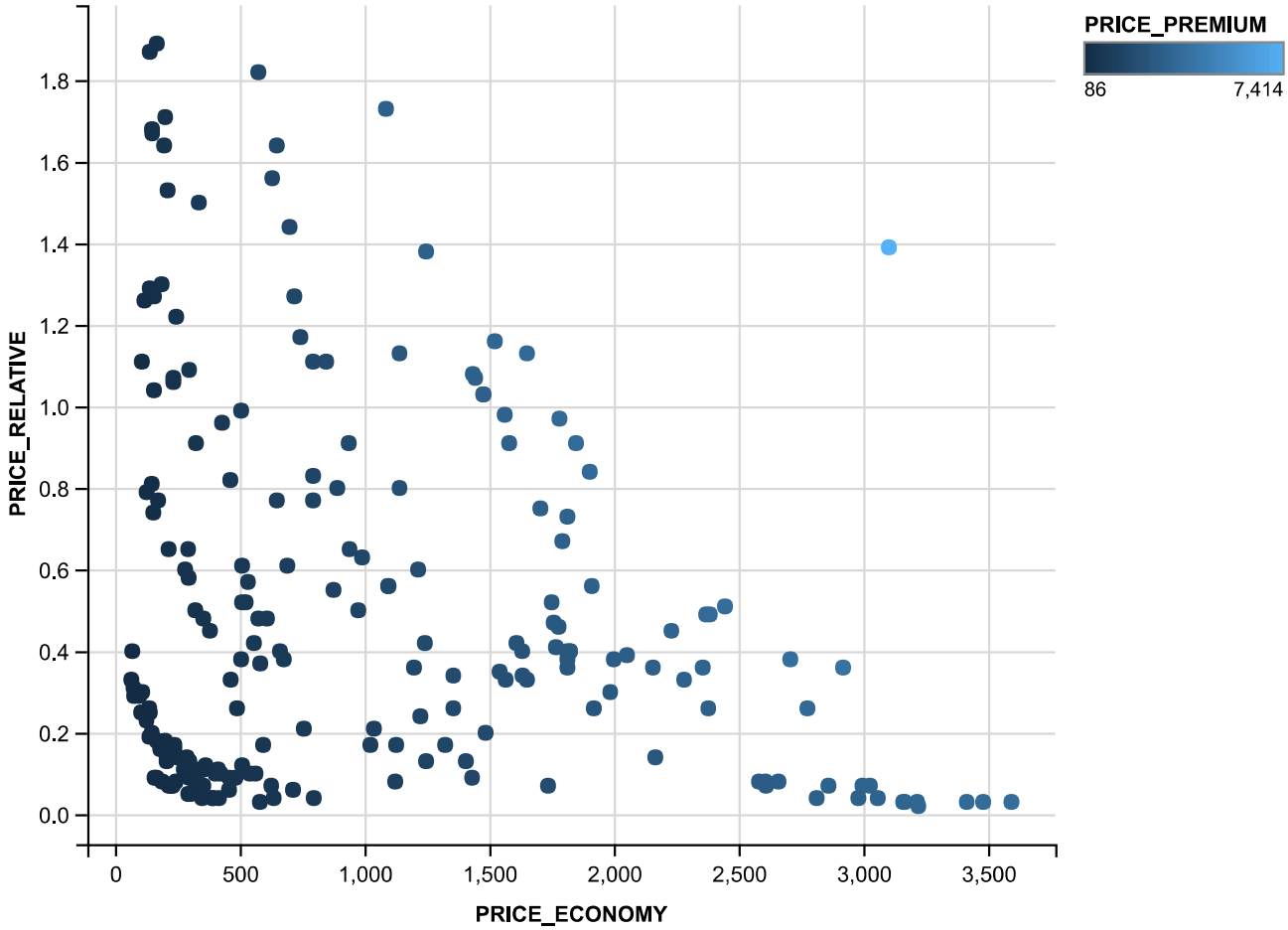

```
ggvis(~PRICE_PREMIUM,~WIDTH_PREMIUM,fill=~PRICE_RELATIVE,data=air.df)
```

```
## Guessing layer_points()
```



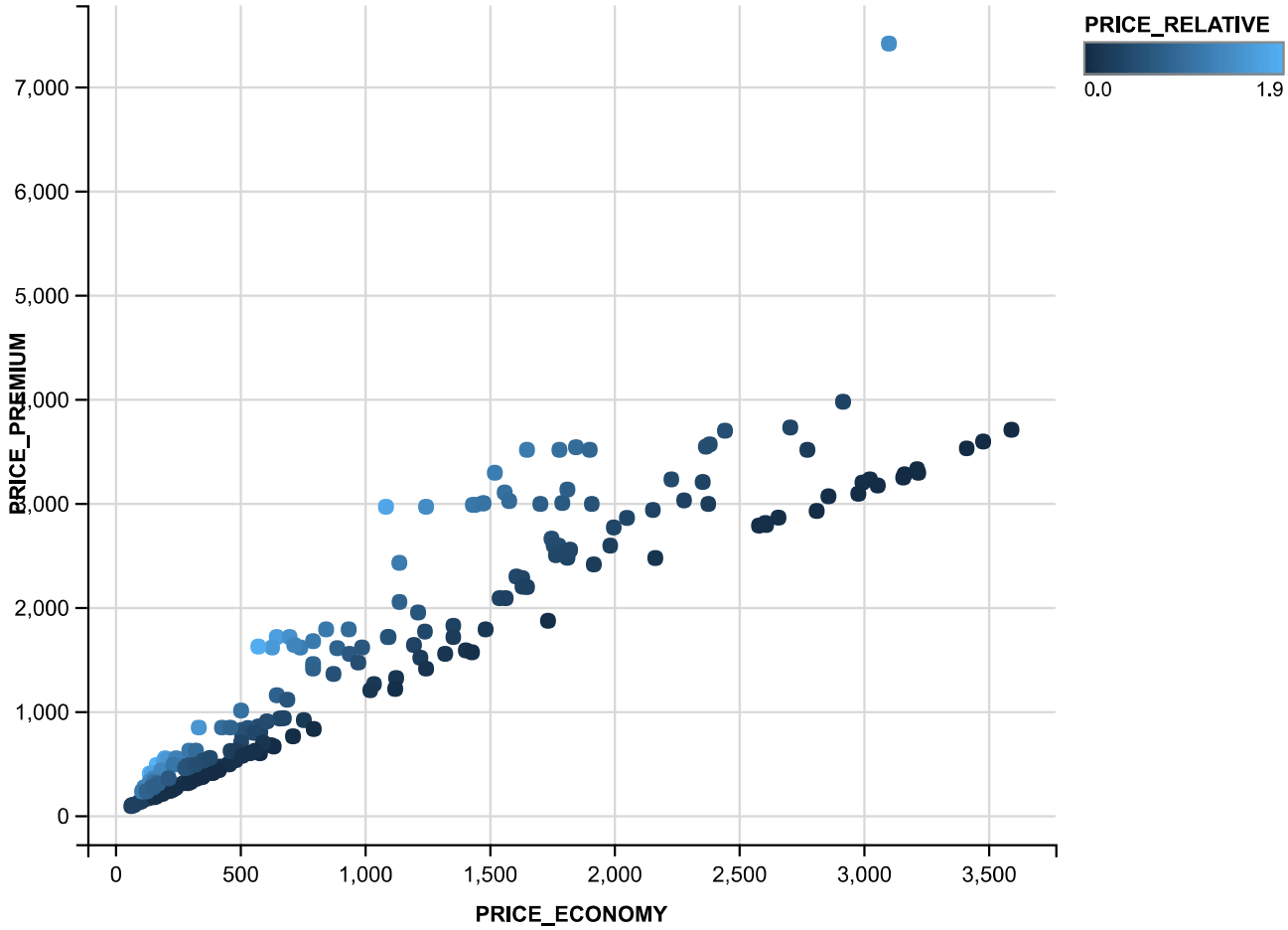

```
#Interaction between the price quantities

ggvis(~PRICE_ECONOMY,~PRICE_RELATIVE,fill=~PRICE_PREMIUM,data=air.df)
```

```
## Guessing layer_points()
```



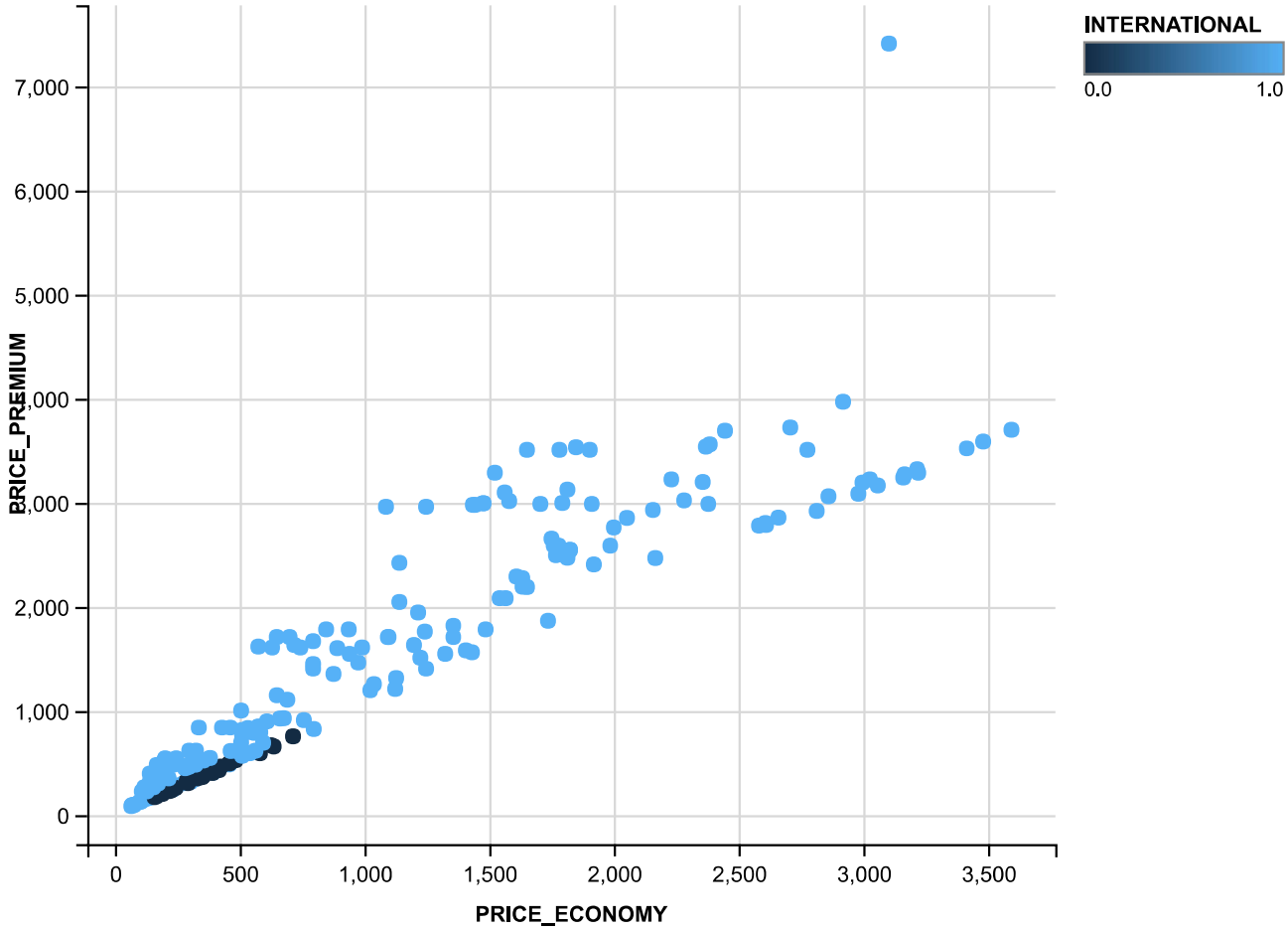

```
ggvis(~PRICE_ECONOMY,~PRICE_PREMIUM,fill=~PRICE_RELATIVE,data=air.df)
```

```
## Guessing layer_points()
```




```
ggvis(~PRICE_ECONOMY,~PRICE_PREMIUM,fill=~INTERNATIONAL,data=air.df)
```

```
## Guessing layer_points()
```




```
##Correlation tests to find relationship between different parameters
# Correlation matrix,covariance matrix, Corrgram

cor(air1)
```

```
##                     AIRCRAFT FLIGHT_DURATION        MONTH INTERNATIONAL
## AIRCRAFT         1.000000000    -0.022265599  0.007377185    0.11607192
## FLIGHT_DURATION -0.022265599     1.000000000  0.042726734    0.37027250
## MONTH            0.007377185     0.042726734  1.000000000    0.04278124
## INTERNATIONAL    0.116071917     0.370272495  0.042781237    1.00000000
## SEATS_ECONOMY    0.401466912     0.209301905  0.013665959    0.30780599
## SEATS_PREMIUM    0.295390514     0.167707636  0.054286479    0.30898003
## PITCH_ECONOMY    0.241596266     0.303239275  0.020511825   -0.28733904
## PITCH_PREMIUM   -0.040380485     0.082007777 -0.008600321    0.86207721
## WIDTH_ECONOMY    0.292529863     0.462436824  0.034357832    0.29207630
## WIDTH_PREMIUM    0.036853395     0.090968063 -0.066095472    0.61130395
## PRICE_ECONOMY    0.037272760     0.570604184  0.002979302    0.29906411
## PRICE_PREMIUM    0.020092754     0.651891797  0.018930560    0.34559444
## PRICE_RELATIVE  -0.119890981     0.108224710  0.019989106    0.27746729
## N                0.405295313     0.213389202  0.020544571    0.32332858
## LAMBDA          -0.128565543     0.003494719  0.038527062    0.01812136
## QUALITY         -0.119809583    -0.051829983 -0.014013167    0.74708385
##                 SEATS_ECONOMY SEATS_PREMIUM PITCH_ECONOMY PITCH_PREMIUM
## AIRCRAFT          0.401466912  0.2953905141    0.24159627  -0.040380485
## FLIGHT_DURATION   0.209301905  0.1677076365    0.30323927   0.082007777
## MONTH             0.013665959  0.0542864791    0.02051183  -0.008600321
## INTERNATIONAL     0.307805993  0.3089800311   -0.28733904   0.862077210
## SEATS_ECONOMY     1.000000000  0.6267956084    0.17614558   0.082656260
## SEATS_PREMIUM     0.626795608  1.0000000000   -0.01847429  -0.008183207
## PITCH_ECONOMY     0.176145581 -0.0184742879    1.00000000  -0.561708319
## PITCH_PREMIUM     0.082656260 -0.0081832073   -0.56170832   1.000000000
## WIDTH_ECONOMY     0.391536716  0.4618817295    0.31103746  -0.043627342
## WIDTH_PREMIUM     0.070712269 -0.0140721606   -0.54892493   0.754851871
## PRICE_ECONOMY     0.148163761  0.1221179841    0.37960497   0.033635434
## PRICE_PREMIUM     0.195287760  0.2249974040    0.23983878   0.071331248
## PRICE_RELATIVE   -0.024509143 -0.1072356234   -0.43465185   0.427823608
## N                 0.992908806  0.7149790849    0.15523329   0.072917300
## LAMBDA           -0.399068794  0.3163713970   -0.17266314  -0.053841073
## QUALITY          -0.004098189  0.0007906561   -0.78895106   0.951521570
##                 WIDTH_ECONOMY WIDTH_PREMIUM PRICE_ECONOMY PRICE_PREMIUM
## AIRCRAFT           0.29252986    0.03685339   0.037272760    0.02009275
## FLIGHT_DURATION    0.46243682    0.09096806   0.570604184    0.65189180
## MONTH              0.03435783   -0.06609547   0.002979302    0.01893056
## INTERNATIONAL      0.29207630    0.61130395   0.299064115    0.34559444
## SEATS_ECONOMY      0.39153672    0.07071227   0.148163761    0.19528776
## SEATS_PREMIUM      0.46188173   -0.01407216   0.122117984    0.22499740
## PITCH_ECONOMY      0.31103746   -0.54892493   0.379604972    0.23983878
## PITCH_PREMIUM     -0.04362734    0.75485187   0.033635434    0.07133125
## WIDTH_ECONOMY      1.00000000    0.06259988   0.081967901    0.16310757
## WIDTH_PREMIUM      0.06259988    1.00000000  -0.070180135    0.04932498
## PRICE_ECONOMY      0.08196790   -0.07018013   1.000000000    0.90253105
## PRICE_PREMIUM      0.16310757    0.04932498   0.902531053    1.00000000
## PRICE_RELATIVE    -0.06093642    0.51167702  -0.298104314    0.01757342
## N                  0.42178638    0.06130173   0.151575655    0.20955518
## LAMBDA             0.09724877   -0.07203714  -0.005672849    0.03550282
## QUALITY           -0.14803745    0.76469927  -0.116146580   -0.03618869
##                  PRICE_RELATIVE            N       LAMBDA       QUALITY
## AIRCRAFT           -0.11989098  0.405295313 -0.128565543 -0.1198095828
```

```
## FLIGHT_DURATION     0.10822471  0.213389202  0.003494719 -0.0518299826
## MONTH               0.01998911  0.020544571  0.038527062 -0.0140131666
## INTERNATIONAL       0.27746729  0.323328584  0.018121358  0.7470838510
## SEATS_ECONOMY      -0.02450914  0.992908806 -0.399068794 -0.0040981889
## SEATS_PREMIUM      -0.10723562  0.714979085  0.316371397  0.0007906561
## PITCH_ECONOMY      -0.43465185  0.155233293 -0.172663141 -0.7889510603
## PITCH_PREMIUM       0.42782361  0.072917300 -0.053841073  0.9515215700
## WIDTH_ECONOMY      -0.06093642  0.421786384  0.097248771 -0.1480374529
## WIDTH_PREMIUM       0.51167702  0.061301732 -0.072037142  0.7646992712
## PRICE_ECONOMY      -0.29810431  0.151575655 -0.005672849 -0.1161465802
## PRICE_PREMIUM       0.01757342  0.209555179  0.035502822 -0.0361886891
## PRICE_RELATIVE      1.00000000 -0.038352287 -0.050057215  0.4793339231
## N                  -0.03835229  1.000000000 -0.309808226 -0.0035565934
## LAMBDA             -0.05005721 -0.309808226  1.000000000  0.0242043400
## QUALITY             0.47933392 -0.003556593  0.024204340  1.0000000000
```

```
x<-air.df[,c("FLIGHT_DURATION","LAMBDA","INTERNATIONAL","N", "MONTH","QUALITY","WIDTH_ECONOMY",
"WIDTH_PREMIUM","SEATS_ECONOMY", "SEATS_PREMIUM", "PITCH_ECONOMY", "PITCH_PREMIUM")]
y<-air.df[,c("PRICE_ECONOMY","PRICE_PREMIUM","PRICE_RELATIVE")]
cor(x,y)
```

```
##                 PRICE_ECONOMY PRICE_PREMIUM PRICE_RELATIVE
## FLIGHT_DURATION   0.570604184    0.65189180     0.10822471
## LAMBDA           -0.005672849    0.03550282    -0.05005721
## INTERNATIONAL     0.299064115    0.34559444     0.27746729
## N                 0.151575655    0.20955518    -0.03835229
## MONTH             0.002979302    0.01893056     0.01998911
## QUALITY          -0.116146580   -0.03618869     0.47933392
## WIDTH_ECONOMY     0.081967901    0.16310757    -0.06093642
## WIDTH_PREMIUM    -0.070180135    0.04932498     0.51167702
## SEATS_ECONOMY     0.148163761    0.19528776    -0.02450914
## SEATS_PREMIUM     0.122117984    0.22499740    -0.10723562
## PITCH_ECONOMY     0.379604972    0.23983878    -0.43465185
## PITCH_PREMIUM     0.033635434    0.07133125     0.42782361
```

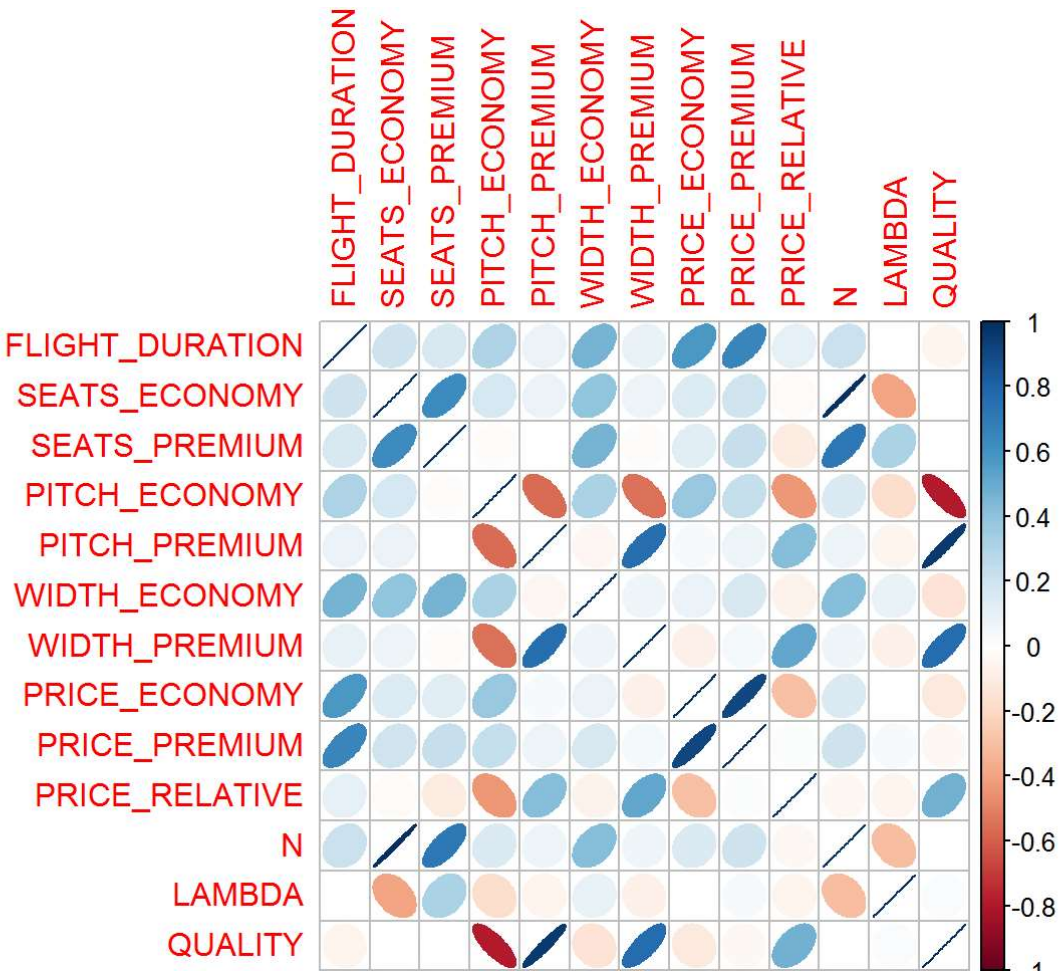
```
cov(x,y)
```

```
##                 PRICE_ECONOMY PRICE_PREMIUM PRICE_RELATIVE
## FLIGHT_DURATION  1999.4291957   2976.982377    0.173589562
## LAMBDA             -0.3434499      2.801268   -0.001387251
## INTERNATIONAL      83.3353523    125.505104    0.035391817
## N               13035.3006451  23486.573795   -1.509763642
## MONTH               3.1016706     25.684687    0.009525782
## QUALITY          -204.7190749    -83.129354    0.386737377
## WIDTH_ECONOMY      45.4751951    117.932783   -0.015475111
## WIDTH_PREMIUM     -76.5196214     70.089857    0.255376323
## SEATS_ECONOMY   11433.0342001  19639.218554   -0.865712595
## SEATS_PREMIUM    1602.2664451   3847.355241   -0.644051047
## PITCH_ECONOMY     248.7500352    204.823760   -0.130376370
## PITCH_PREMIUM      44.0309604    121.694406    0.256361007
```

```
var(x,y)
```

```
##                 PRICE_ECONOMY PRICE_PREMIUM PRICE_RELATIVE
## FLIGHT_DURATION  1999.4291957   2976.982377    0.173589562
## LAMBDA             -0.3434499      2.801268   -0.001387251
## INTERNATIONAL      83.3353523    125.505104    0.035391817
## N               13035.3006451  23486.573795   -1.509763642
## MONTH               3.1016706     25.684687    0.009525782
## QUALITY          -204.7190749    -83.129354    0.386737377
## WIDTH_ECONOMY      45.4751951    117.932783   -0.015475111
## WIDTH_PREMIUM     -76.5196214     70.089857    0.255376323
## SEATS_ECONOMY   11433.0342001  19639.218554   -0.865712595
## SEATS_PREMIUM    1602.2664451   3847.355241   -0.644051047
## PITCH_ECONOMY     248.7500352    204.823760   -0.130376370
## PITCH_PREMIUM      44.0309604    121.694406    0.256361007
```

```
 #Visualizing relation through corrplots

library(corrplot)
corrplot(corr=cor(air1[,c(2,5:16)],use = "complete.obs"), method = "ellipse")
```

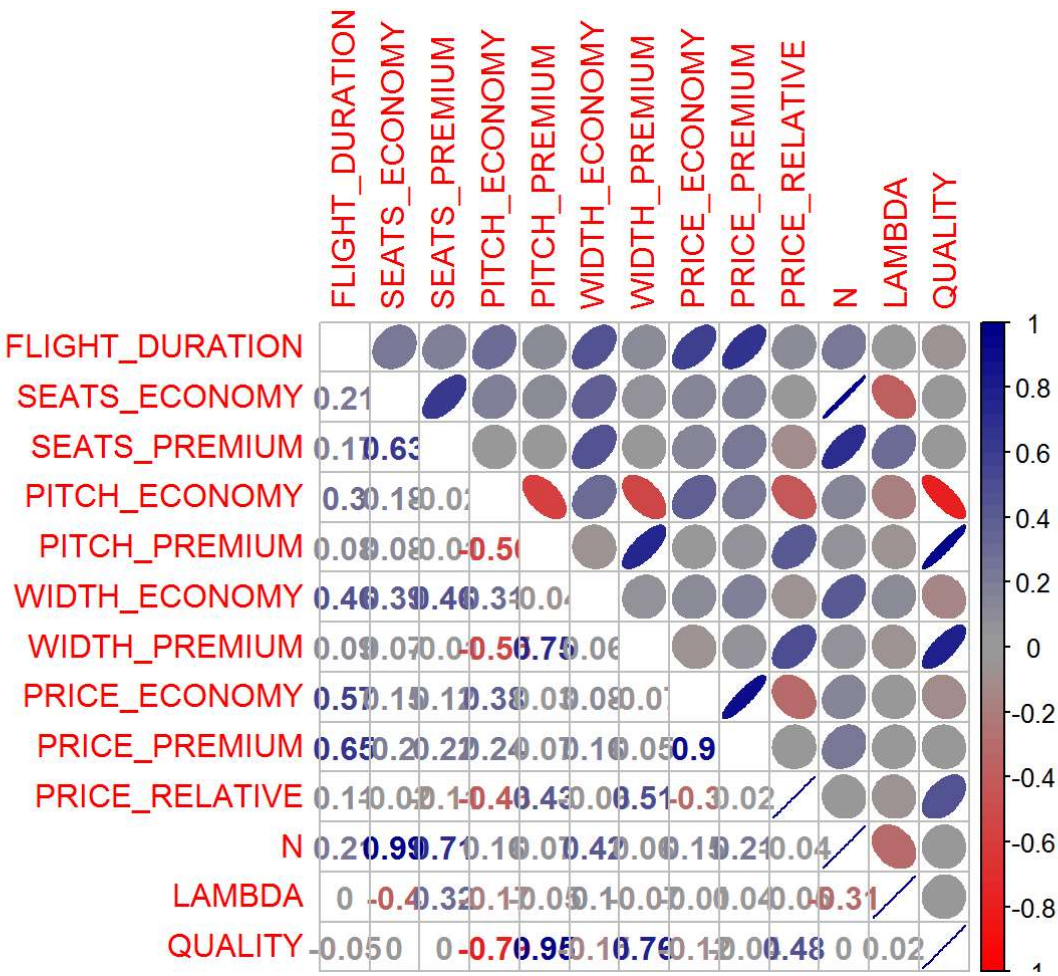



```
library(gplots)
```

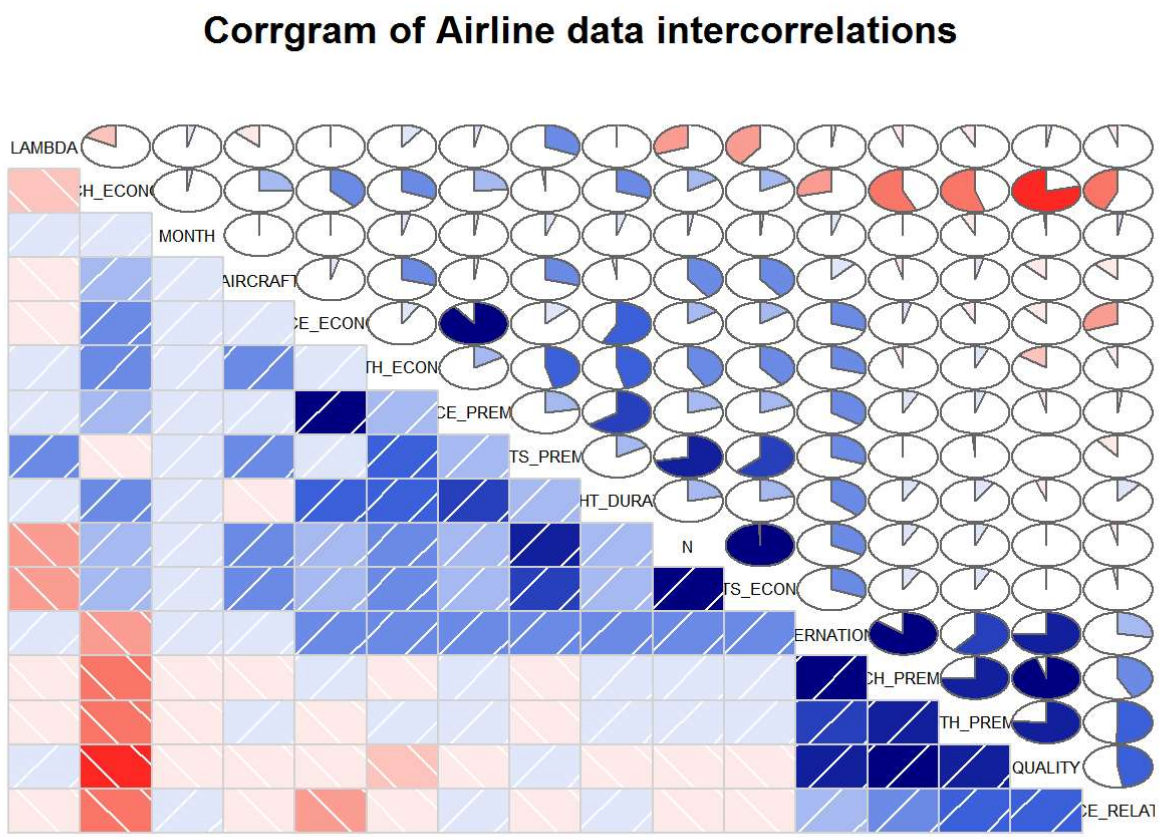
```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
 corrplot.mixed(corr=cor(air1[,c(2,5:16)],use = "complete.obs"), upper = "ellipse", tl.pos = "l
t", col = colorpanel(50, "red", "gray60", "blue4"))
```




```
#VIsualizing by corrgram

library(corrgram)

corrgram(air1, order=TRUE, lower.panel=panel.shade,
         upper.panel=panel.pie, text.panel=panel.txt,
         main="Corrgram of Airline data intercorrelations")
```




```
 ##From the above cor(),corrplot and corrgram command we get to know the correlation , and can c
onsider
 ## FLIGHT_DURATION as a factor for pricing in ECONOMY and PREMIUM seats and for
 ## PRICE_RELATIVE factors such as QUALITY,PITCH_ECONOMY, PITCH_PREMIUM and WIDTH_PREMIUM are we
ll correlated

 #Hypothesis
 #1. Average COST of Premium seats in Boing aircraft is less than Airbus aircraft
 t.test(PRICE_PREMIUM~AIRCRAFT,alternative="less",data=air.df)
```

```
##
##  Welch Two Sample t-test
##
## data:  PRICE_PREMIUM by AIRCRAFT
## t = -0.43785, df = 309.53, p-value = 0.3309
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 152.7841
## sample estimates:
## mean in group 0 mean in group 1
##        1814.305        1869.503
```

```
 #2. Average COST of ECONOMY seats in Boing aircraft is less than Airbus aircraft
 t.test(PRICE_ECONOMY~AIRCRAFT,alternative="less", data=air.df)
```

```
##
##  Welch Two Sample t-test
##
## data:  PRICE_ECONOMY by AIRCRAFT
## t = -0.79106, df = 288.69, p-value = 0.2148
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 85.32482
## sample estimates:
## mean in group 0 mean in group 1
##        1291.386        1369.954
```

```
#3. Average COST of ECONOMY seats in Domestic is less than International
t.test(PRICE_ECONOMY~INTERNATIONAL, alternative="less", data = air.df)
```

```
##
##  Welch Two Sample t-test
##
## data:  PRICE_ECONOMY by INTERNATIONAL
## t = -20.174, df = 436.68, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -965.5663
## sample estimates:
## mean in group 0 mean in group 1
##         356.625        1408.102
```

```
#4.Average COST of PREMIUM ECONOMY seats in Domestic is less than International
t.test(PRICE_PREMIUM~INTERNATIONAL, alternative="less", data = air.df)
```

```
##
##  Welch Two Sample t-test
##
## data:  PRICE_PREMIUM by INTERNATIONAL
## t = -24.295, df = 457.34, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -1476.119
## sample estimates:
## mean in group 0 mean in group 1
##          385.90         1969.45
```

```
#Multiple Linear Regressional Model for Price of Economy Flights as dependent variable
fit1<- lm(PRICE_ECONOMY~ PITCH_ECONOMY + WIDTH_ECONOMY + FLIGHT_DURATION + QUALITY +PRICE_RELAT
IVE+ INTERNATIONAL, data = air.df)
summary(fit1)
```

```
##
## Call:
## lm(formula = PRICE_ECONOMY ~ PITCH_ECONOMY + WIDTH_ECONOMY +
##     FLIGHT_DURATION + QUALITY + PRICE_RELATIVE + INTERNATIONAL,
##     data = air.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1537.13  -436.22    68.69   419.27  1494.89
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     18409.89    3877.90   4.747 2.76e-06 ***
## PITCH_ECONOMY     -70.81     103.48  -0.684    0.494
## WIDTH_ECONOMY    -884.54      65.40 -13.526  < 2e-16 ***
## FLIGHT_DURATION   141.15      10.39  13.588  < 2e-16 ***
## QUALITY          -377.06      65.56  -5.752 1.62e-08 ***
## PRICE_RELATIVE   -679.84      73.37  -9.266  < 2e-16 ***
## INTERNATIONAL    2946.19     298.22   9.879  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 588.4 on 455 degrees of freedom
## Multiple R-squared:  0.6512, Adjusted R-squared:  0.6466
## F-statistic: 141.6 on 6 and 455 DF,  p-value: < 2.2e-16
```

```
 #Multiple Linear Regressional Model for Price of Premium Flights as dependent variable
 fit2<- lm(PRICE_PREMIUM~ PITCH_PREMIUM + WIDTH_PREMIUM + FLIGHT_DURATION + QUALITY +PRICE_RELAT
IVE+ INTERNATIONAL , data = air.df)
 summary(fit2)
```

```
##
## Call:
## lm(formula = PRICE_PREMIUM ~ PITCH_PREMIUM + WIDTH_PREMIUM +
##     FLIGHT_DURATION + QUALITY + PRICE_RELATIVE + INTERNATIONAL,
##     data = air.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2152.2  -752.8   136.9   699.7  4295.9
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     13439.721   5325.779   2.524   0.0120 *
## PITCH_PREMIUM    -401.935    161.602  -2.487   0.0132 *
## WIDTH_PREMIUM       9.289     65.400   0.142   0.8871
## FLIGHT_DURATION   177.065     16.686  10.611  < 2e-16 ***
## QUALITY           -17.521     97.489  -0.180   0.8575
## PRICE_RELATIVE      1.889    120.271   0.016   0.9875
## INTERNATIONAL    2446.630    412.479   5.932 5.95e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 942.4 on 455 degrees of freedom
## Multiple R-squared:  0.4732, Adjusted R-squared:  0.4662
## F-statistic: 68.11 on 6 and 455 DF,  p-value: < 2.2e-16
```

```
 #Multiple Linear Regressional Model for Price relative as dependent variable
 fit<-lm(PRICE_RELATIVE~.-AIRCRAFT-MONTH-SEATS_ECONOMY-SEATS_PREMIUM-LAMBDA-N-WIDTH_ECONOMY, dat
a = air1)
 summary(fit)
```

```
##
## Call:
## lm(formula = PRICE_RELATIVE ~ . - AIRCRAFT - MONTH - SEATS_ECONOMY -
##     SEATS_PREMIUM - LAMBDA - N - WIDTH_ECONOMY, data = air1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74081 -0.11639 -0.01705  0.09721  0.86056
##
## Coefficients: (1 not defined because of singularities)
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1.463e+01  1.437e+00 -10.184  < 2e-16 ***
## FLIGHT_DURATION  1.992e-02  4.600e-03   4.330 1.84e-05 ***
## INTERNATIONAL   -9.237e-01  1.056e-01  -8.752  < 2e-16 ***
## PITCH_ECONOMY    1.153e-01  2.645e-02   4.359 1.62e-05 ***
## PITCH_PREMIUM    3.086e-01  2.616e-02  11.799  < 2e-16 ***
## WIDTH_PREMIUM    2.816e-02  1.646e-02   1.711   0.0878 .
## PRICE_ECONOMY   -7.665e-04  3.031e-05 -25.291  < 2e-16 ***
## PRICE_PREMIUM    5.331e-04  2.414e-05  22.083  < 2e-16 ***
## QUALITY                 NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.237 on 454 degrees of freedom
## Multiple R-squared:  0.7306, Adjusted R-squared:  0.7265
## F-statistic: 175.9 on 7 and 454 DF,  p-value: < 2.2e-16
```