Premium Economy vs Economy Ticket Pricing by Airlines

Aman Gupta
February 16, 2018

This is a mini project based on the study of Premium Economy Vs Economy Ticket Pricing by Airlines. The main research question that we are concerned with is “What factors explain the difference in price between an economy ticket and a premium-economy airline ticket?” Some other questions to be answered are: How Premium and Economy class ticket prices vary in domestic and International Flights respectively? Does the Relative Price increase or decrease with percentage of premium and economy seats? What are the other contributing factors?

Read the data

```
airlines <- read.csv(paste("SixAirlinesDataV2.csv",sep=""))
View(airlines)
```

Attach the dataframe

```
attach(airlines)
```

Summarize the data

```
summary(airlines)
```

```
##      Airline     Aircraft   FlightDuration   TravelMonth
##  AirFrance: 74   AirBus:151   Min.   : 1.250   Aug:127
##  British  :175   Boeing:307   1st Qu.: 4.260   Jul: 75
##  Delta    : 46                Median : 7.790   Oct:127
##  Jet      : 61                Mean   : 7.578   Sep:129
##  Singapore: 40                3rd Qu.:10.620
##  Virgin   : 62                Max.   :14.660
##      IsInternational  SeatsEconomy    SeatsPremium    PitchEconomy
##  Domestic     : 40    Min.   : 78.0   Min.   : 8.00   Min.   :30.00
##  International:418    1st Qu.:133.0   1st Qu.:21.00   1st Qu.:31.00
##                       Median :185.0   Median :36.00   Median :31.00
##                       Mean   :202.3   Mean   :33.65   Mean   :31.22
##                       3rd Qu.:243.0   3rd Qu.:40.00   3rd Qu.:32.00
##                       Max.   :389.0   Max.   :66.00   Max.   :33.00
##   PitchPremium    WidthEconomy    WidthPremium    PriceEconomy
##  Min.   :34.00   Min.   :17.00   Min.   :17.00   Min.   :  65
##  1st Qu.:38.00   1st Qu.:18.00   1st Qu.:19.00   1st Qu.: 413
##  Median :38.00   Median :18.00   Median :19.00   Median :1242
##  Mean   :37.91   Mean   :17.84   Mean   :19.47   Mean   :1327
##  3rd Qu.:38.00   3rd Qu.:18.00   3rd Qu.:21.00   3rd Qu.:1909
```

```
##  Max.   :40.00   Max.   :19.00   Max.   :21.00   Max.   :3593
##   PricePremium    PriceRelative      SeatsTotal  PitchDifference
##  Min.   :  86.0   Min.   :0.0200   Min.   : 98   Min.   : 2.000
##  1st Qu.: 528.8   1st Qu.:0.1000   1st Qu.:166   1st Qu.: 6.000
##  Median :1737.0   Median :0.3650   Median :227   Median : 7.000
##  Mean   :1845.3   Mean   :0.4872   Mean   :236   Mean   : 6.688
##  3rd Qu.:2989.0   3rd Qu.:0.7400   3rd Qu.:279   3rd Qu.: 7.000
##  Max.   :7414.0   Max.   :1.8900   Max.   :441   Max.   :10.000
##  WidthDifference PercentPremiumSeats
##  Min.   :0.000   Min.   : 4.71
##  1st Qu.:1.000   1st Qu.:12.28
##  Median :1.000   Median :13.21
##  Mean   :1.633   Mean   :14.65
##  3rd Qu.:3.000   3rd Qu.:15.36
##  Max.   :4.000   Max.   :24.69

library(psych)
describe(airlines)

##                    vars   n    mean     sd  median trimmed     mad    min
## Airline*              1 458    3.01    1.65    2.00    2.89    1.48  1.00
## Aircraft*             2 458    1.67    0.47    2.00    1.71    0.00  1.00
## FlightDuration        3 458    7.58    3.54    7.79    7.57    4.81  1.25
## TravelMonth*          4 458    2.56    1.17    3.00    2.58    1.48  1.00
## IsInternational*      5 458    1.91    0.28    2.00    2.00    0.00  1.00
## SeatsEconomy          6 458  202.31   76.37  185.00  194.64   85.99 78.00
## SeatsPremium          7 458   33.65   13.26   36.00   33.35   11.86  8.00
## PitchEconomy          8 458   31.22    0.66   31.00   31.26    0.00 30.00
## PitchPremium          9 458   37.91    1.31   38.00   38.05    0.00 34.00
## WidthEconomy         10 458   17.84    0.56   18.00   17.81    0.00 17.00
## WidthPremium         11 458   19.47    1.10   19.00   19.53    0.00 17.00
## PriceEconomy         12 458 1327.08  988.27 1242.00 1244.40 1159.39
65.00
## PricePremium         13 458 1845.26 1288.14 1737.00 1799.05 1845.84
86.00
## PriceRelative        14 458    0.49    0.45    0.36    0.42    0.41  0.02
## SeatsTotal           15 458  235.96   85.29  227.00  228.73   90.44 98.00
## PitchDifference      16 458    6.69    1.76    7.00    6.76    0.00  2.00
## WidthDifference      17 458    1.63    1.19    1.00    1.53    0.00  0.00
## PercentPremiumSeats  18 458   14.65    4.84   13.21   14.31    2.68  4.71
##                       max   range  skew kurtosis    se
## Airline*             6.00    5.00  0.61    -0.95  0.08
## Aircraft*            2.00    1.00 -0.72    -1.48  0.02
## FlightDuration      14.66   13.41 -0.07    -1.12  0.17
## TravelMonth*         4.00    3.00 -0.14    -1.46  0.05
## IsInternational*     2.00    1.00 -2.91     6.50  0.01
## SeatsEconomy       389.00  311.00  0.72    -0.36  3.57
## SeatsPremium        66.00   58.00  0.23    -0.46  0.62
## PitchEconomy        33.00    3.00 -0.03    -0.35  0.03
## PitchPremium        40.00    6.00 -1.51     3.52  0.06
## WidthEconomy        19.00    2.00 -0.04    -0.08  0.03
```

```
## WidthPremium          21.00    4.00 -0.08    -0.31  0.05
## PriceEconomy        3593.00 3528.00  0.51    -0.88 46.18
## PricePremium        7414.00 7328.00  0.50     0.43 60.19
## PriceRelative          1.89    1.87  1.17     0.72  0.02
## SeatsTotal           441.00  343.00  0.70    -0.53  3.99
## PitchDifference       10.00    8.00 -0.54     1.78  0.08
## WidthDifference        4.00    4.00  0.84    -0.53  0.06
## PercentPremiumSeats   24.69   19.98  0.71     0.28  0.23
```

```
#MEAN
apply(airlines[,6:18], FUN=mean, MARGIN=2)
```

```
##       SeatsEconomy       SeatsPremium       PitchEconomy
##        202.3122271         33.6484716         31.2183406
##       PitchPremium       WidthEconomy       WidthPremium
##         37.9061135         17.8384279         19.4716157
##       PriceEconomy       PricePremium      PriceRelative
##       1327.0764192       1845.2576419          0.4872052
##        SeatsTotal     PitchDifference     WidthDifference
##        235.9606987          6.6877729          1.6331878
## PercentPremiumSeats
##         14.6454148
```

```
#STANDARD DEVIATION
apply(airlines[,6:18], FUN=sd, MARGIN=2)
```

```
##       SeatsEconomy       SeatsPremium       PitchEconomy
##         76.3735257         13.2614183          0.6551695
##       PitchPremium       WidthEconomy       WidthPremium
##          1.3139238          0.5575102          1.0971726
##       PriceEconomy       PricePremium      PriceRelative
##        988.2732729       1288.1355206          0.4505873
##        SeatsTotal     PitchDifference     WidthDifference
##         85.2931525          1.7617077          1.1892807
## PercentPremiumSeats
##          4.8424513
```

The average price of an economy seat is USD 1327, while the average price of a premium-economy seat is USD 1845, i.e premium-economy seats are pricier by almost 49%.

Data Types

```
str(airlines)
```

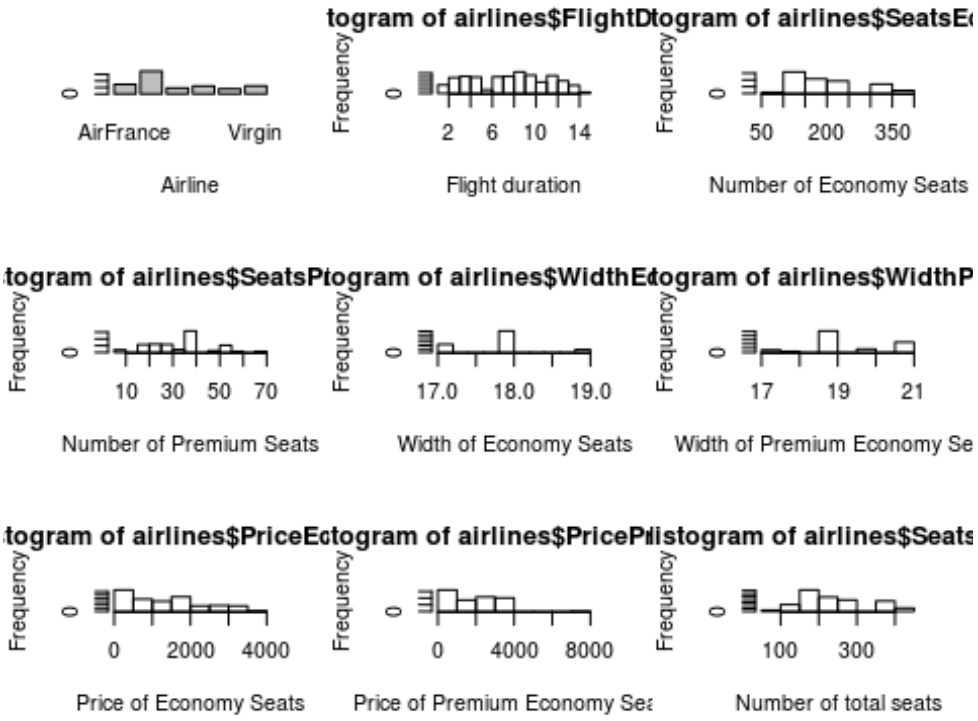
```
## 'data.frame':    458 obs. of  18 variables:
##  $ Airline          : Factor w/ 6 levels "AirFrance","British",..: 2 2 2 2 2 2 2 2
2 2 ...
##  $ Aircraft         : Factor w/ 2 levels "AirBus","Boeing": 2 2 2 2 2 2 2 2 2
2 ...
##  $ FlightDuration   : num  12.25 12.25 12.25 12.25 8.16 ...
##  $ TravelMonth      : Factor w/ 4 levels "Aug","Jul","Oct",..: 2 1 4 3 1 4 3 1
```

```
4 4 ...
##  $ IsInternational    : Factor w/ 2 levels "Domestic","International": 2 2 2 2
2 2 2 2 2 2 ...
##  $ SeatsEconomy      : int  122 122 122 122 122 122 122 122 122 122 ...
##  $ SeatsPremium      : int  40 40 40 40 40 40 40 40 40 40 ...
##  $ PitchEconomy      : int  31 31 31 31 31 31 31 31 31 31 ...
##  $ PitchPremium      : int  38 38 38 38 38 38 38 38 38 38 ...
##  $ WidthEconomy      : int  18 18 18 18 18 18 18 18 18 18 ...
##  $ WidthPremium      : int  19 19 19 19 19 19 19 19 19 19 ...
##  $ PriceEconomy      : int  2707 2707 2707 2707 1793 1793 1793 1476
1476 1705 ...
##  $ PricePremium      : int  3725 3725 3725 3725 2999 2999 2999 2997
2997 2989 ...
##  $ PriceRelative     : num  0.38 0.38 0.38 0.38 0.67 0.67 0.67 1.03 1.03
0.75 ...
##  $ SeatsTotal        : int  162 162 162 162 162 162 162 162 162 162 ...
##  $ PitchDifference   : int  7 7 7 7 7 7 7 7 7 7 ...
##  $ WidthDifference   : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ PercentPremiumSeats: num  24.7 24.7 24.7 24.7 24.7 ...
```

Visualization

Visualization using Histograms

```
par(mfrow=c(3,3))
plot(airlines$Airline, xlab="Airline")
hist(airlines$FlightDuration , xlab="Flight duration")
hist(airlines$SeatsEconomy, xlab="Number of Economy Seats")
hist(airlines$SeatsPremium, xlab="Number of Premium Seats")
hist(airlines$WidthEconomy, xlab="Width of Economy Seats")
hist(airlines$WidthPremium, xlab="Width of Premium Economy Seats")
hist(airlines$PriceEconomy, xlab="Price of Economy Seats")
hist(airlines$PricePremium, xlab="Price of Premium Economy Seats")
hist(airlines$SeatsTotal, xlab = "Number of total seats")
```

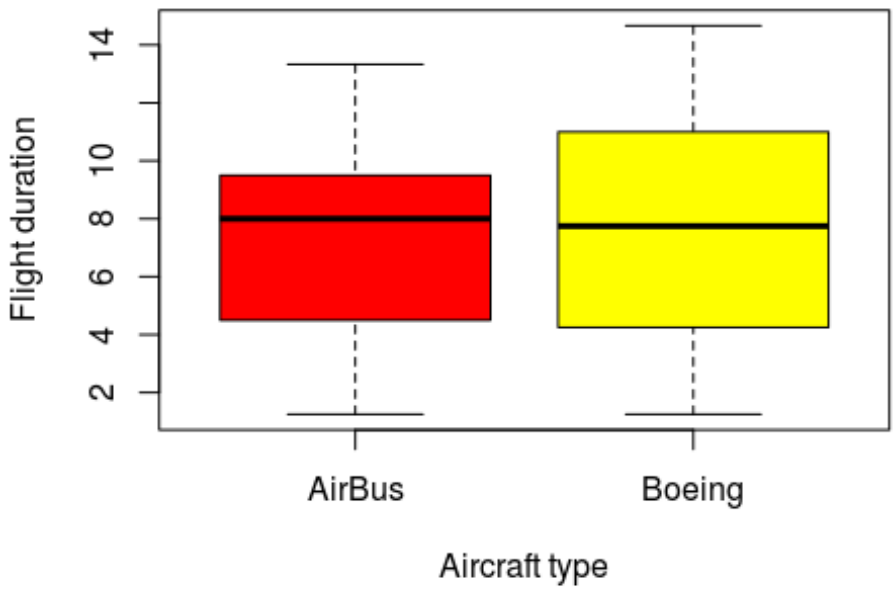



```
par(mfrow=c(1,1))
```

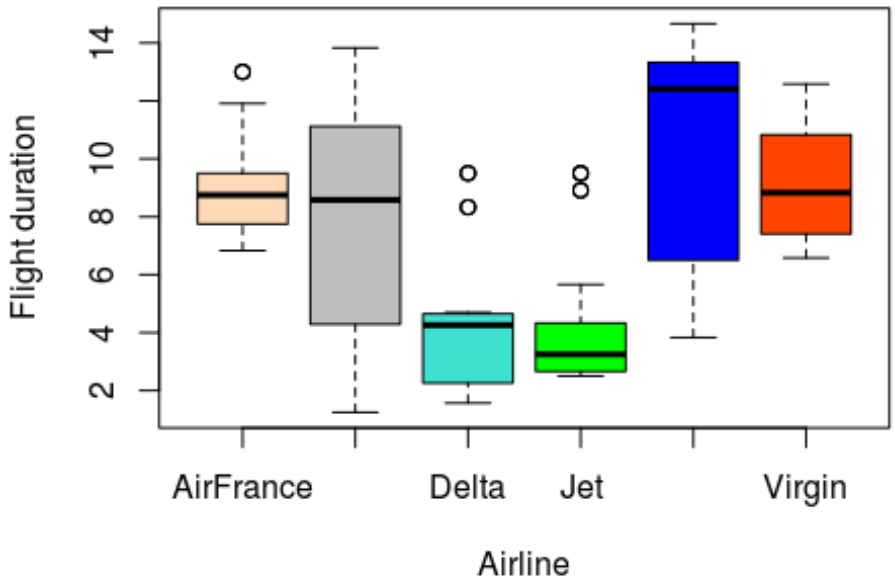
The difference in price can be explained by the enhanced features in premium-economy seats, i.e greater legroom (pitch) and wider seats. On average, premium-economy seats have 6.6 inches more legroom than economy seats, and are wider by around 1.6 inches. There are some other factor affecting the price of the airline ticket, like flight duration, whether the flight is international or domestic or factors like the type of airline it is(Boeing or Airbus).

Visualization using Boxplots

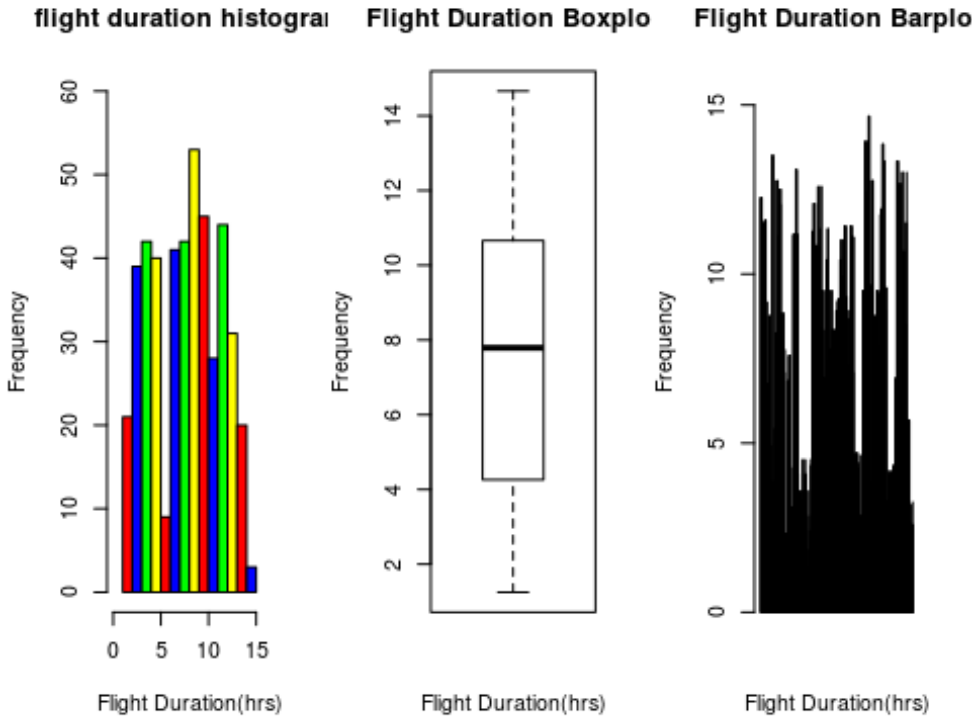
```
boxplot(FlightDuration~Aircraft,data=airlines,xlab="Aircraft type",
ylab="Flight duration",col = c("red","yellow"))
```



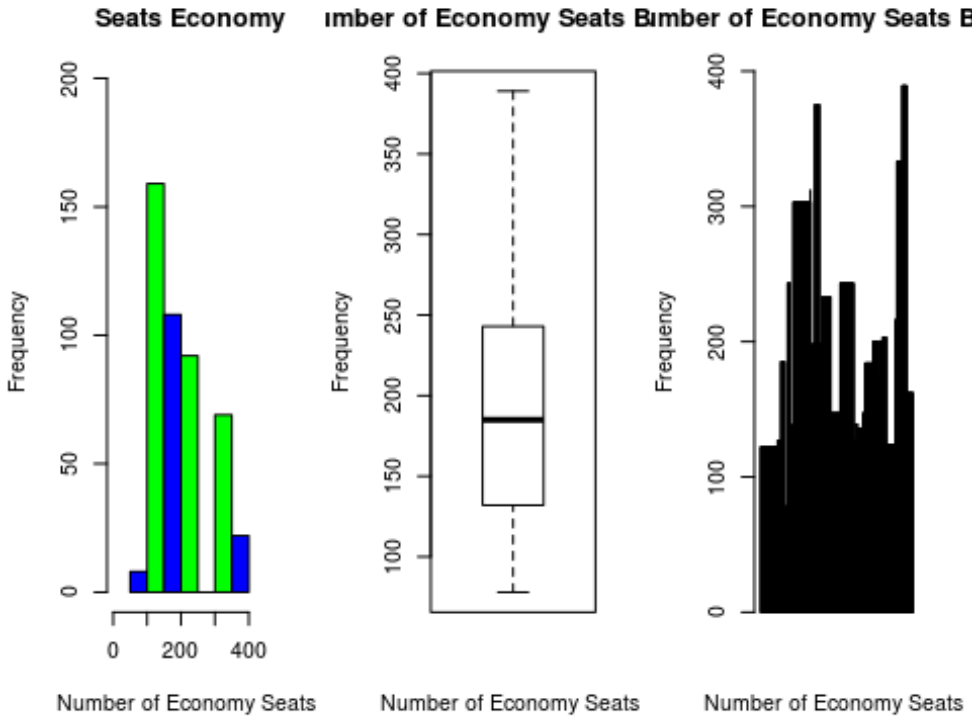

```
boxplot(FlightDuration~Airline,data=airlines,xlab="Airline", ylab="Flight
duration",col = c("peachpuff","gray","turquoise","green","blue","orangered"))
```



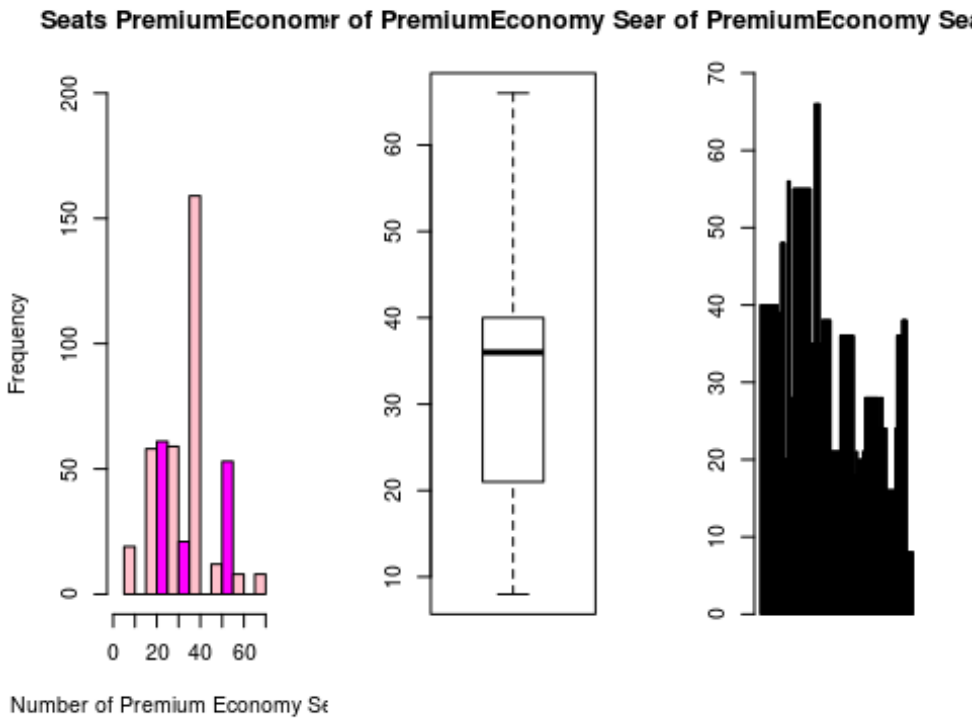
```
par(mfrow=c(1,3))
hist(airlines$FlightDuration,xlab="Flight
Duration(hrs)",ylab="Frequency",main="flight duration histogram",
col=c("red","blue","green","yellow"), breaks=15, xlim=c(0,16), ylim=c(0,60))
boxplot(airlines$FlightDuration,main="Flight Duration Boxplot", xlab="Flight
Duration(hrs)",ylab="Frequency")
barplot(airlines$FlightDuration,main = "Flight Duration Barplot",
ylim=c(0,16), xlab="Flight Duration(hrs)",ylab="Frequency")
```




```
par(mfrow=c(1,3))
hist(airlines$SeatsEconomy,xlab="Number of Economy Seats
",ylab="Frequency",main="Seats Economy", col=c("blue","green"),
xlim=c(0,450),ylim=c(0,195))
boxplot(airlines$SeatsEconomy,main="Number of Economy Seats Boxplot",
xlab="Number of Economy Seats ",ylab="Frequency")
barplot(airlines$SeatsEconomy,main = "Number of Economy Seats Barplot",
ylim=c(0,400), xlab="Number of Economy Seats ",ylab="Frequency")
```

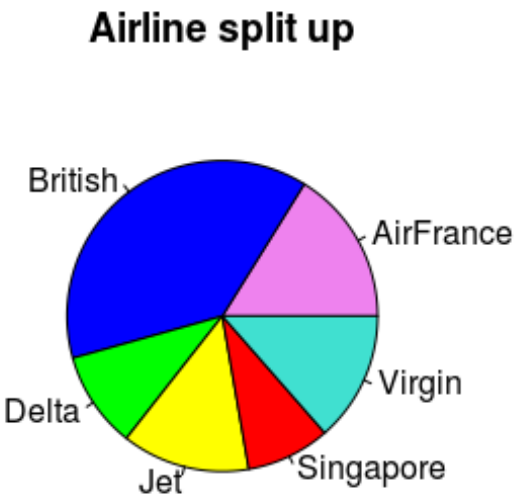



```
par(mfrow=c(1,3))
hist(airlines$SeatsPremium,xlab="Number of Premium Economy Seats
",ylab="Frequency",main="Seats PremiumEconomy",
col=c("pink","magenta"),xlim=c(0,70),ylim=c(0,200))
boxplot(airlines$SeatsPremium,main="Number of PremiumEconomy Seats
Boxplot")
barplot(airlines$SeatsPremium,main = "Number of PremiumEconomy Seats
Barplot",ylim=c(0,70))
```

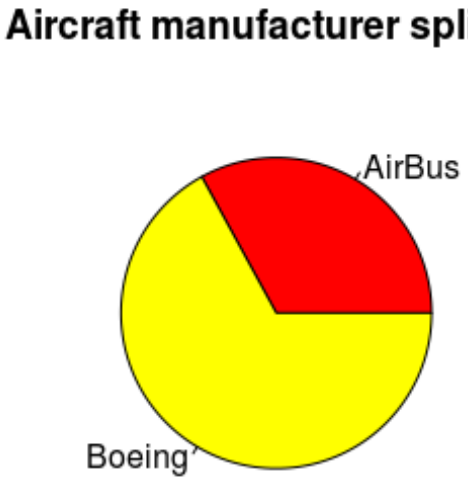


Visualization using Pie-Charts

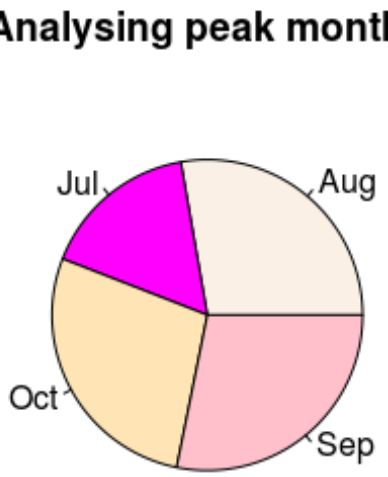
```
par(mfrow=c(1,1))
pie(table(airlines$Airline),col=c("violet","blue","green","yellow","red","turquoise"),main="Airline split up")
```




```
pie(table(airlines$Aircraft),col=c("red","yellow"),main="Aircraft manufacturer split up")
```



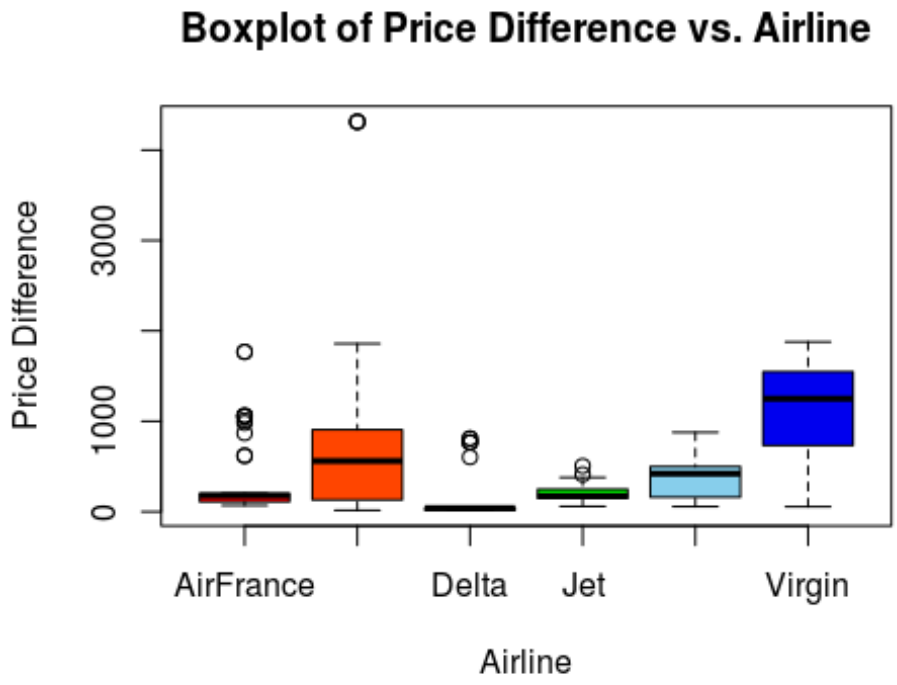
```
pie(table(airlines$TravelMonth),main="Analysing peak
months",col=c("linen","magenta","moccasin","pink"))
```



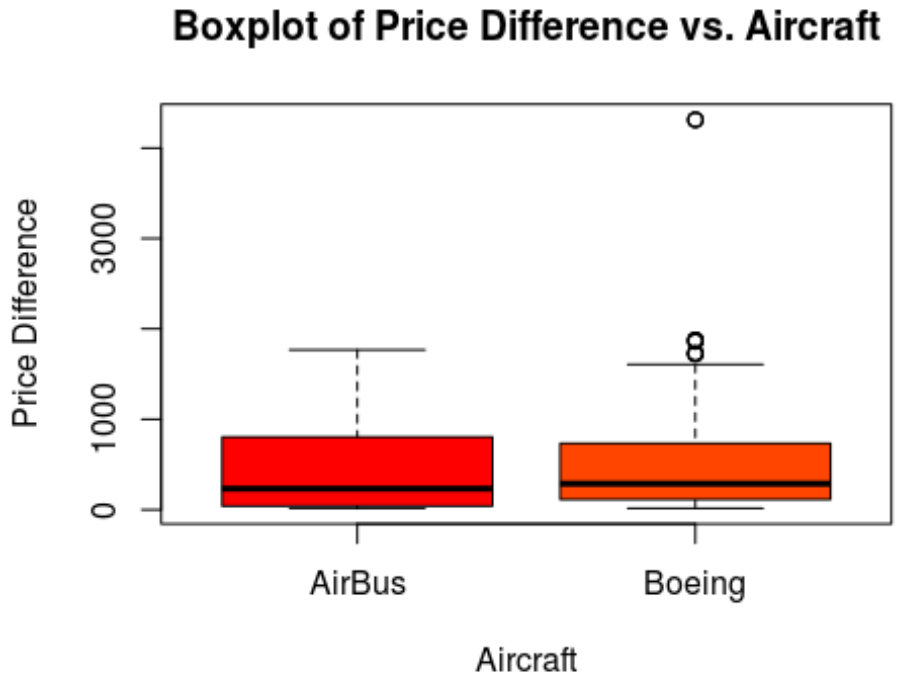

```
# Creating a new column PriceDifference
airlines$PriceDifference <- airlines$PricePremium - airlines$PriceEconomy
```

Visualisation of PriceDifference

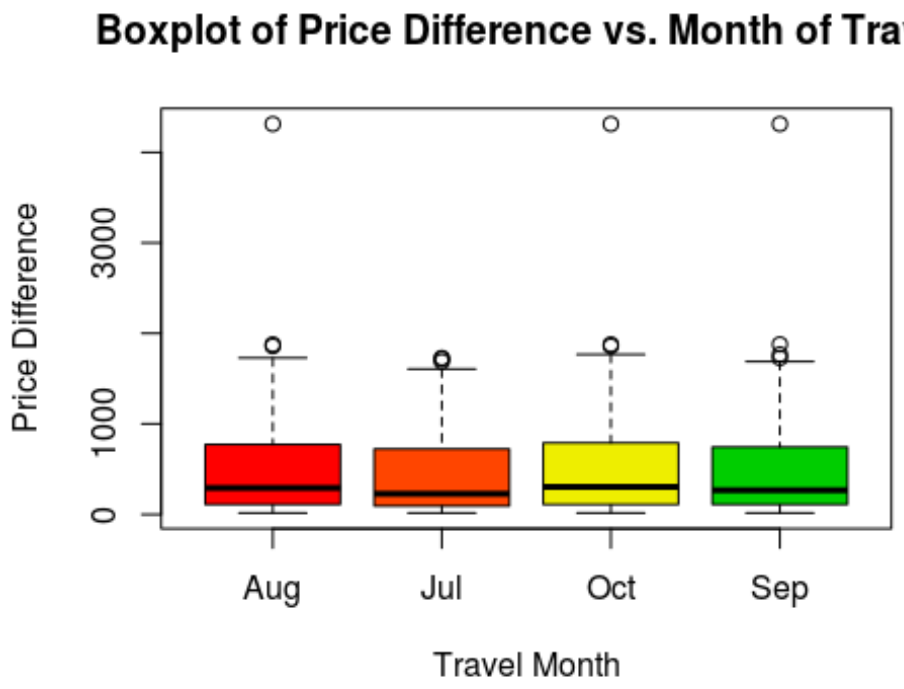
```
boxplot(airlines$PriceDifference~airlines$Airline, ylab="Price Difference",
xlab="Airline", main="Boxplot of Price Difference vs. Airline",
col=c("red","orangered","yellow2","green3","skyblue","blue2"))
```



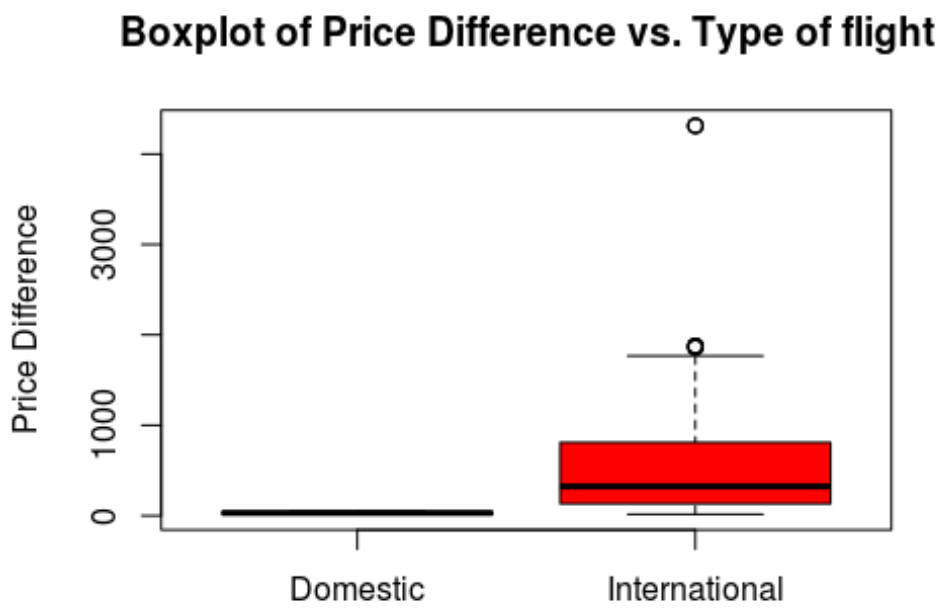

```
boxplot(airlines$PriceDifference~airlines$Aircraft, ylab="Price Difference",
xlab="Aircraft", main="Boxplot of Price Difference vs. Aircraft",
col=c("red","orangered","yellow2","green3","skyblue","blue2"))
```



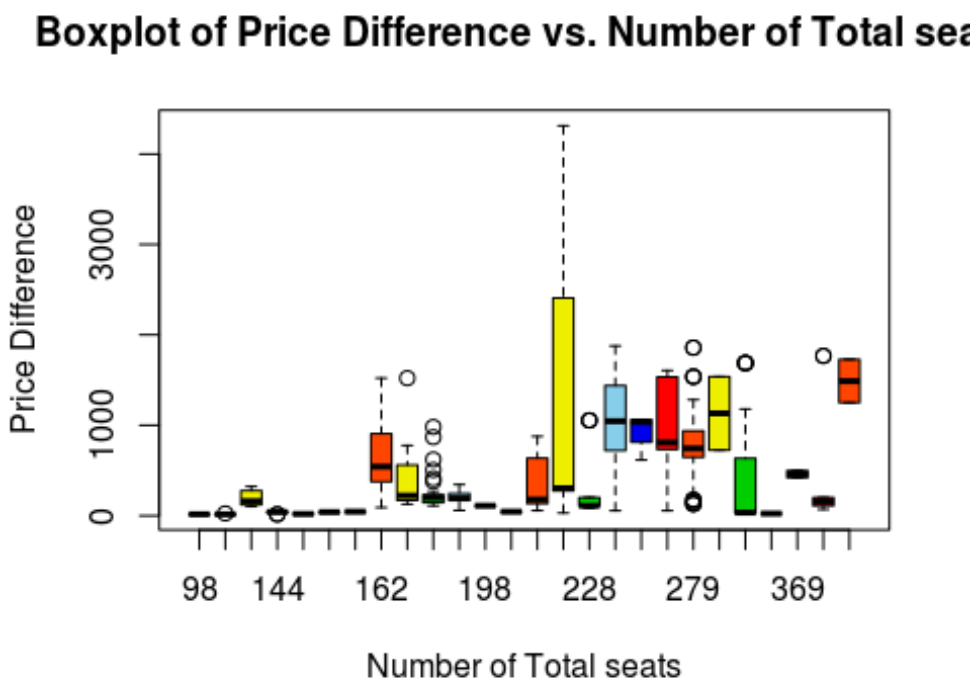
```
boxplot(airlines$PriceDifference~airlines$TravelMonth, ylab="Price
Difference", xlab="Travel Month", main="Boxplot of Price Difference vs.
Month of Travel",
col=c("red","orangered","yellow2","green3","skyblue","blue2"))
```



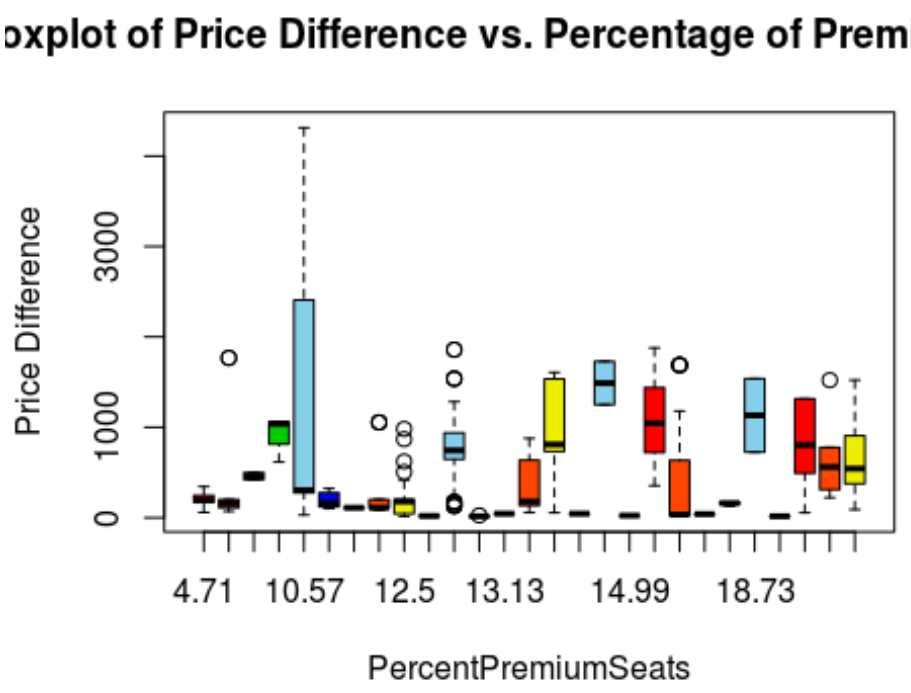

```
boxplot(airlines$PriceDifference~airlines$IsInternational, ylab="Price
Difference", main="Boxplot of Price Difference vs. Type of flight",
col=c("orangered","red","yellow2","green3","skyblue","blue2"))
```



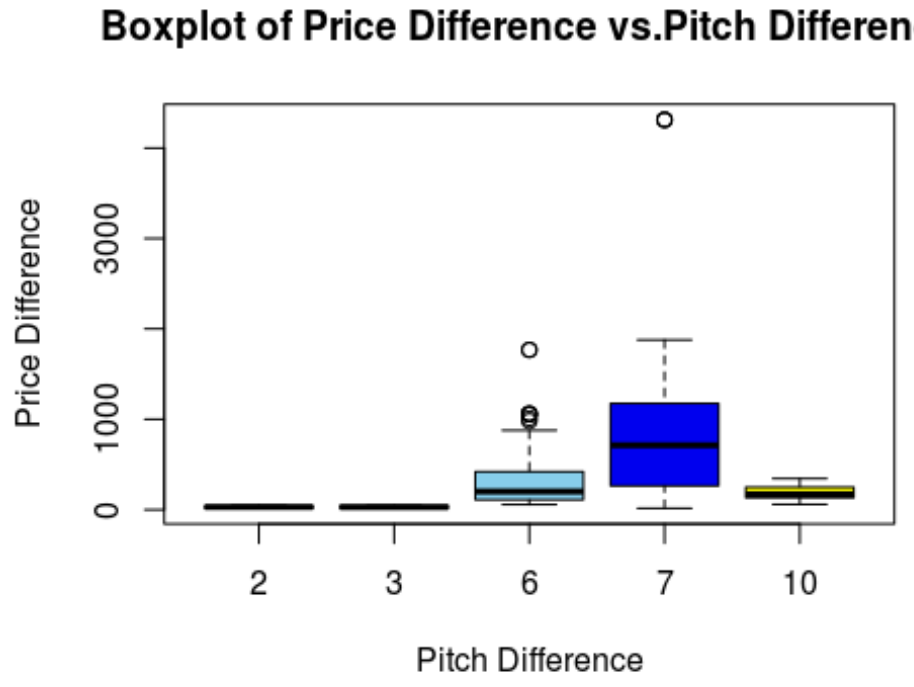

```
boxplot(airlines$PriceDifference~airlines$SeatsTotal, ylab="Price Difference",
xlab="Number of Total seats", main="Boxplot of Price Difference vs. Number
of Total seats",
col=c("red","orangered","yellow2","green3","skyblue","blue2"))
```



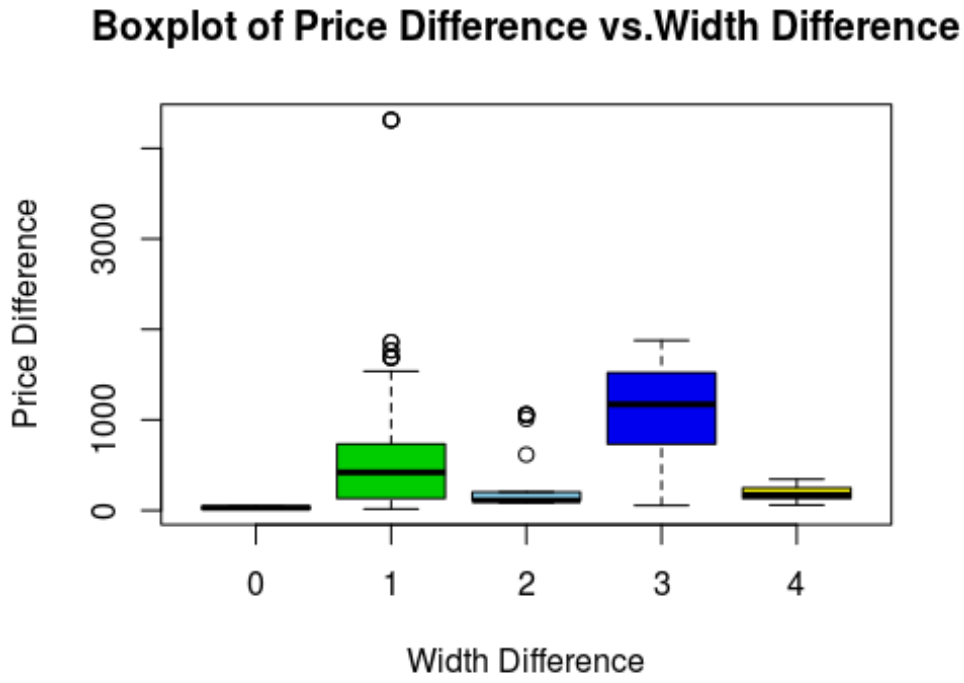

```
boxplot(airlines$PriceDifference~airlines$PercentPremiumSeats, ylab="Price
Difference", xlab="PercentPremiumSeats", main="Boxplot of Price Difference
vs. Percentage of Premium seats",
col=c("red","orangered","yellow2","green3","skyblue","blue2"))
```



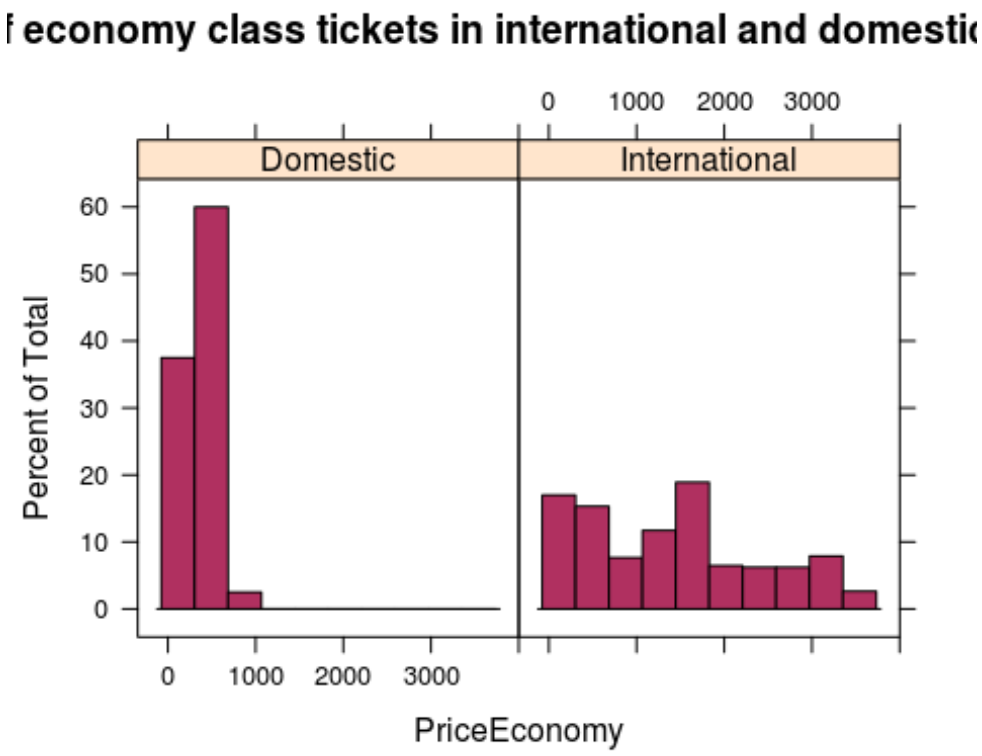

```
boxplot(airlines$PriceDifference~airlines$PitchDifference, ylab="Price
Difference", xlab="Pitch Difference", main="Boxplot of Price Difference
vs.Pitch Difference", col=c("yellow2","green3","skyblue","blue2"))
```



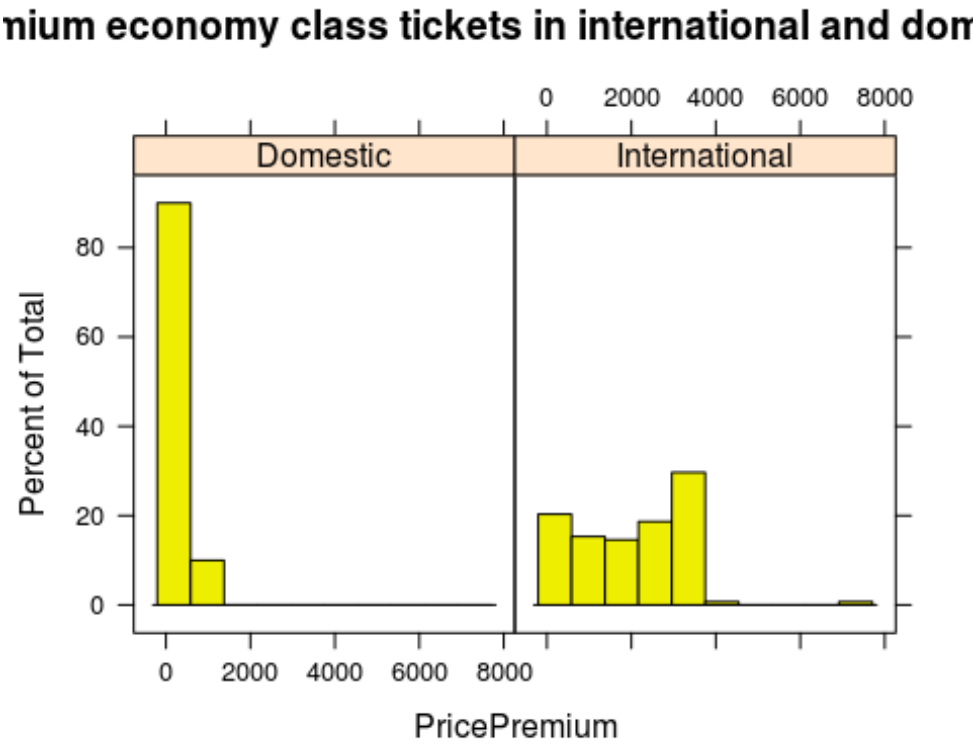
```
boxplot(airlines$PriceDifference~airlines$WidthDifference, ylab="Price
Difference", xlab="Width Difference", main="Boxplot of Price Difference
vs.Width Difference", col=c("yellow2","green3","skyblue","blue2"))
```



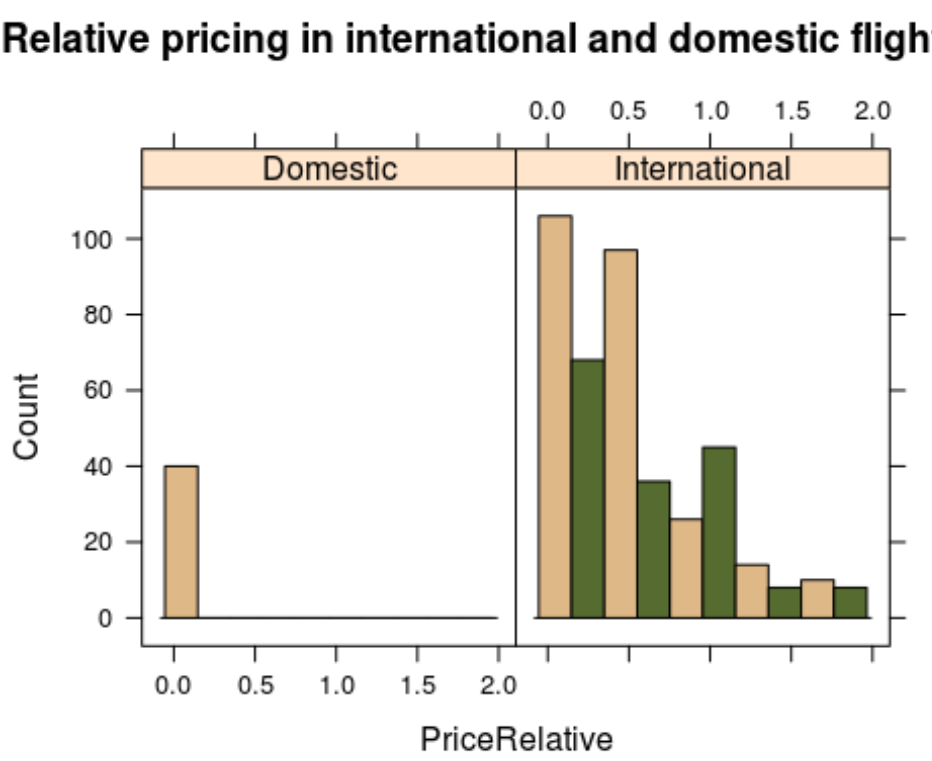

```
library(lattice)
par(mfrow=c(2,1))
histogram(~PriceEconomy | IsInternational,data=airlines,
col="maroon",main="Price of economy class tickets in international and
domestic flights")
```



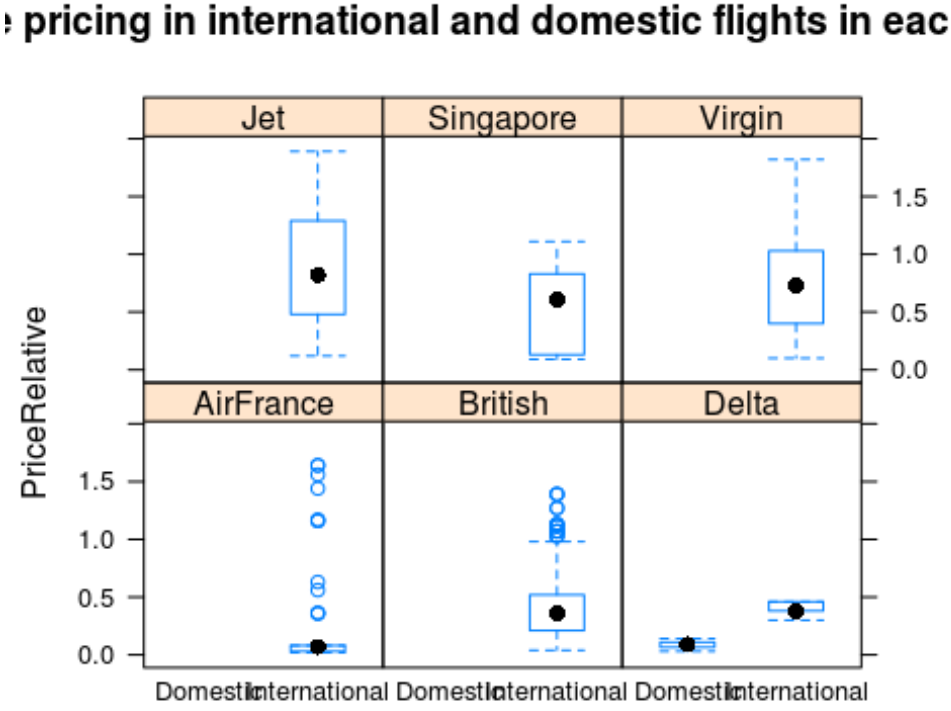

```
par(mfrow=c(2,1))
histogram(~PricePremium | IsInternational,data=airlines,col="yellow2",
main="Price of premium economy class tickets in international and domestic
flights")
```



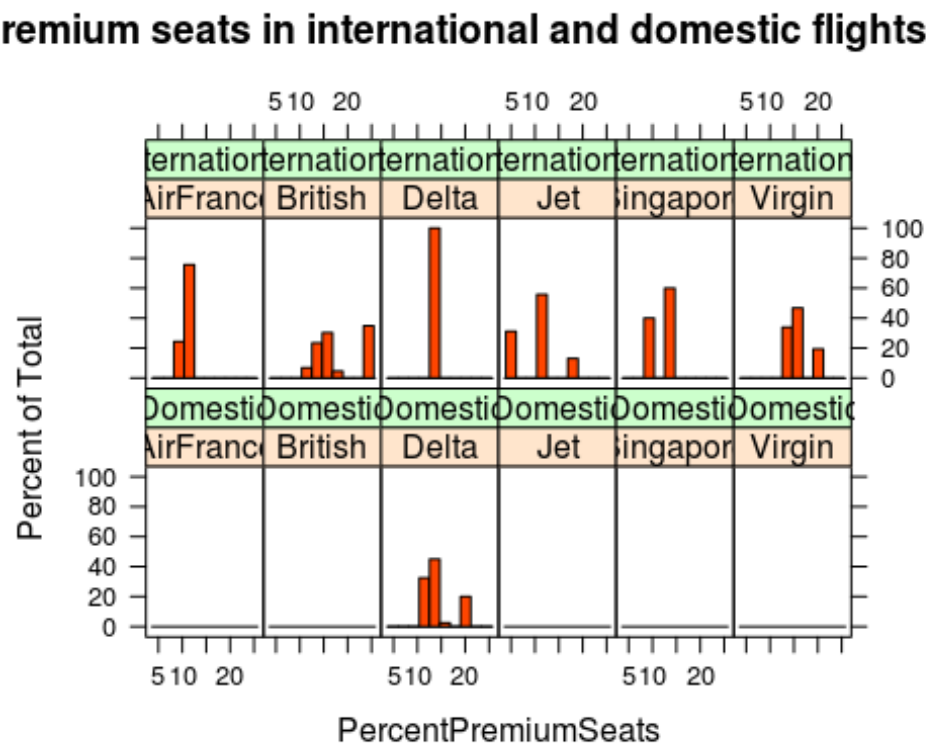

```
par(mfrow=c(1,2))
histogram(~PriceRelative | IsInternational, data=airlines,
        type="count",
        layout=c(2,1),
        col=c("burlywood", "darkolivegreen"),main="Relative pricing in
international and domestic flights")
```



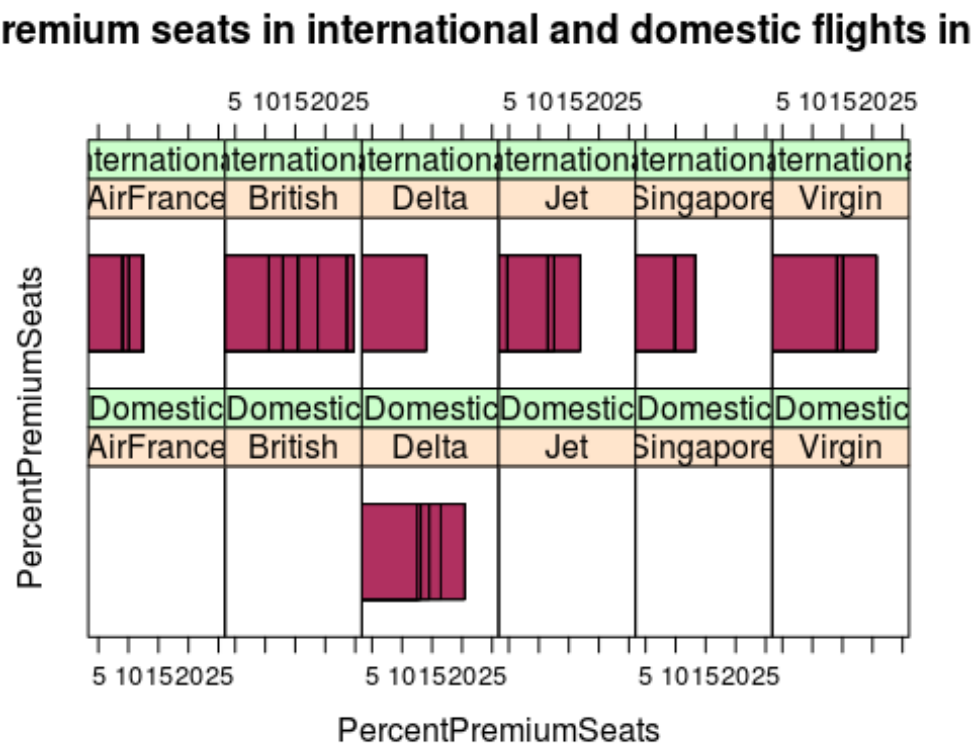

```
bwplot(airlines$PriceRelative~ airlines$IsInternational |
airlines$Airline,ylab="PriceRelative",main="Relative pricing in international
and domestic flights in each arilines")
```



```
par(mfrow=c(2,1))
histogram(~PercentPremiumSeats | Airline + IsInternational, data=airlines,
       layout=c(6,2),
         col="orangered", main="Percent of premium seats in international and
domestic flights in each airline")
```

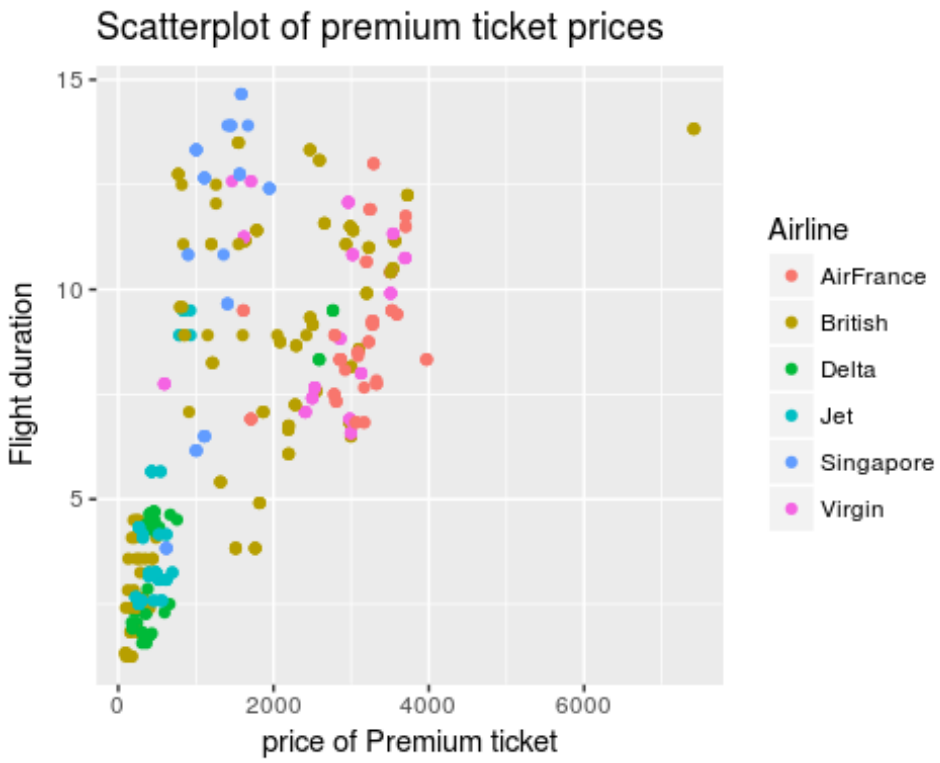



```
barchart(~PercentPremiumSeats | Airline +
IsInternational,ylab="PercentPremiumSeats",data=airlines,col="maroon",
                layout=c(6,2), main="Percent of premium seats in international
and domestic flights in each airline")
```

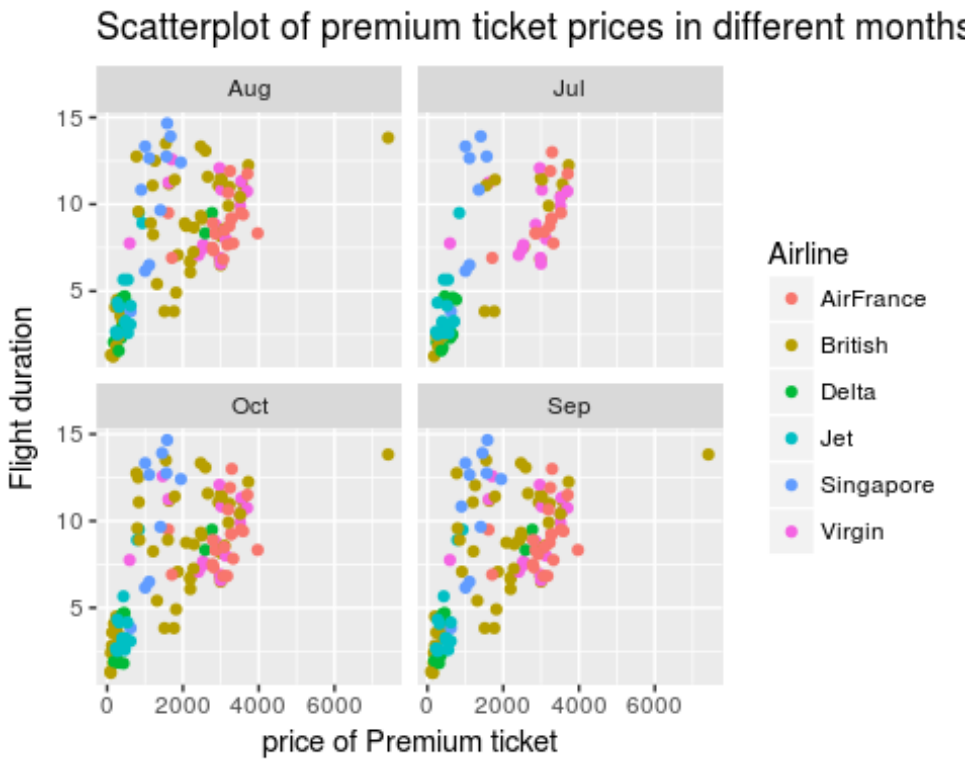


Scatterplots

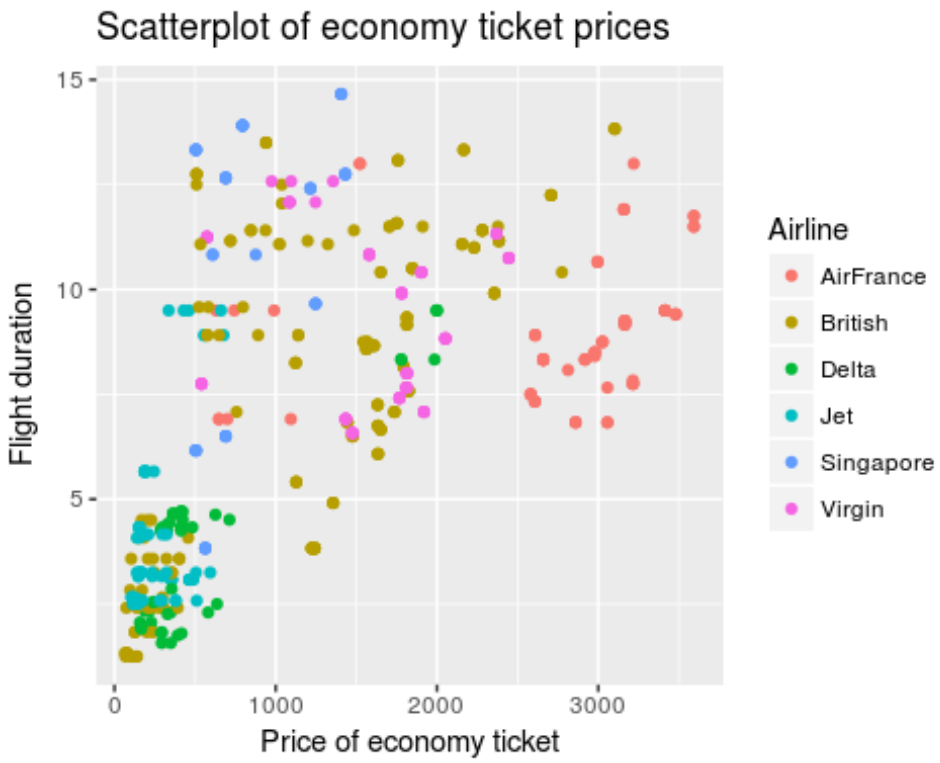
```
library(car)

##
## Attaching package: 'car'

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

library(ggplot2)

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha

ggplot(airlines, aes(PricePremium, FlightDuration)) + geom_point(aes(color =
Airline)) +  scale_x_continuous("price of Premium ticket") +
scale_y_continuous("Flight duration")+ labs(title="Scatterplot of premium
ticket prices")
```




```
library(ggplot2)
ggplot(airlines, aes(PricePremium, FlightDuration)) + geom_point(aes(color =
Airline)) +  scale_x_continuous("price of Premium ticket") +
scale_y_continuous("Flight duration")+ labs(title="Scatterplot of premium
ticket prices in different months") + facet_wrap( ~ TravelMonth)
```

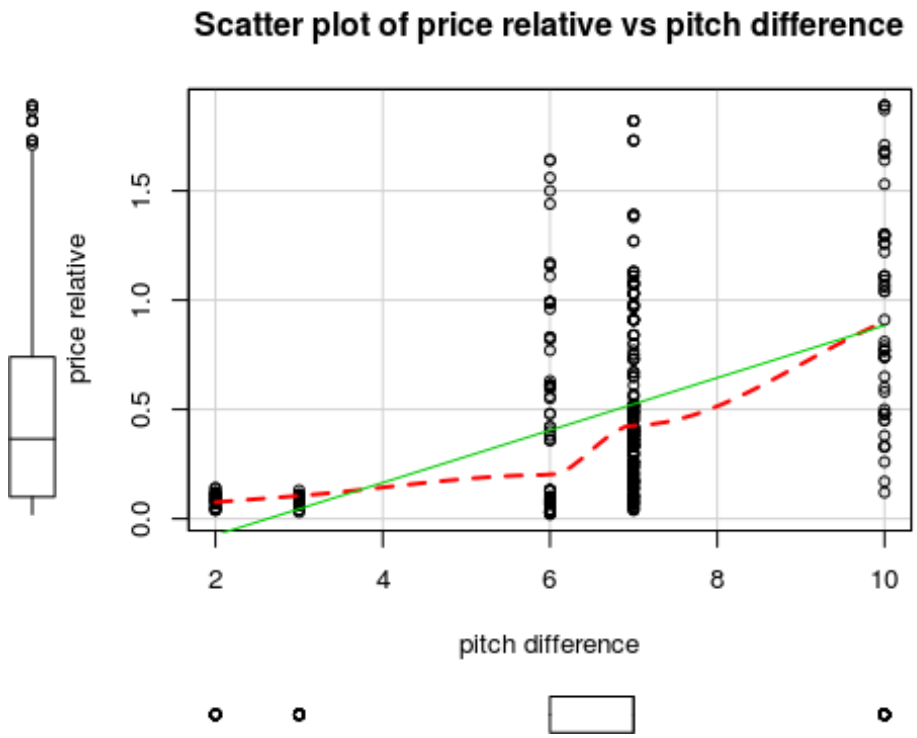



```
ggplot(airlines, aes(PriceEconomy, FlightDuration)) + geom_point(aes(color = Airline)) +  scale_x_continuous("Price of economy ticket") + scale_y_continuous("Flight duration") + labs(title="Scatterplot of economy ticket prices")
```

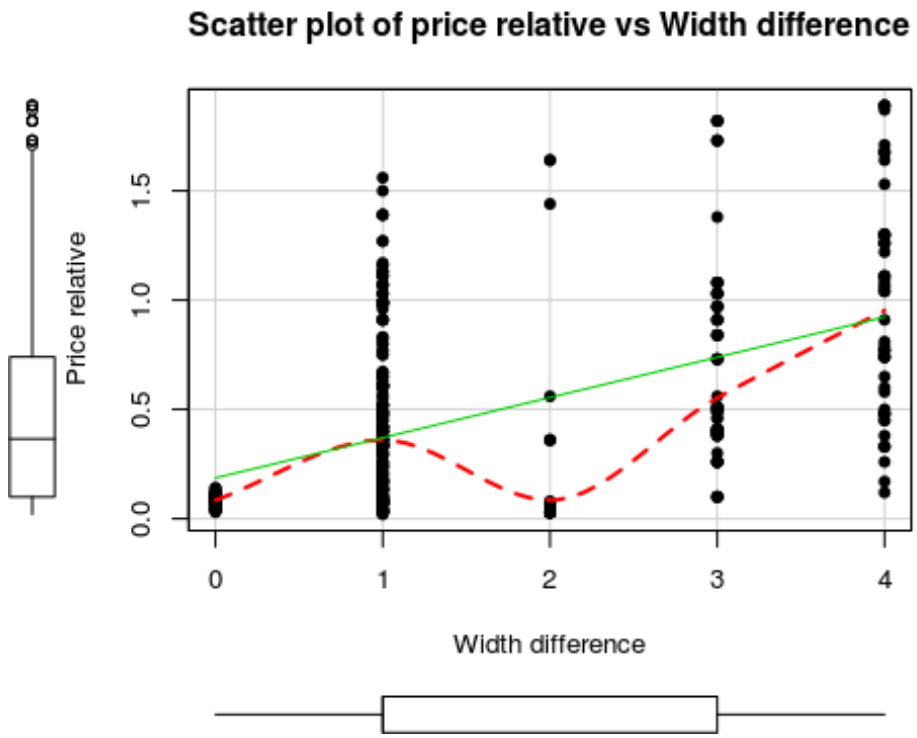



```
ggplot(airlines, aes(PriceEconomy, FlightDuration)) + geom_point(aes(color = Airline)) +  scale_x_continuous("price of economy ticket") + scale_y_continuous("Flight duration") + labs(title="Scatterplot of ticket prices in different months") + facet_wrap( ~ TravelMonth)
```

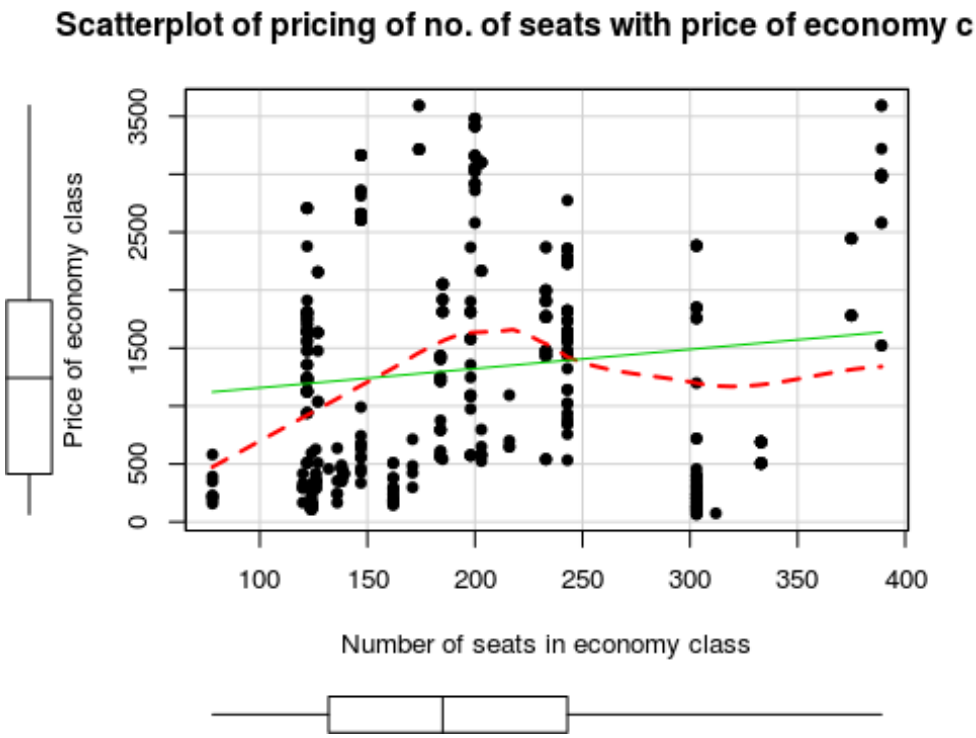

```
scatterplot(PriceRelative ~PitchDifference,     data=airlines,
          spread=FALSE, smoother.args=list(lty=2),
          main="Scatter plot of price relative vs pitch difference",
          xlab="pitch difference",
          ylab="price relative")
```



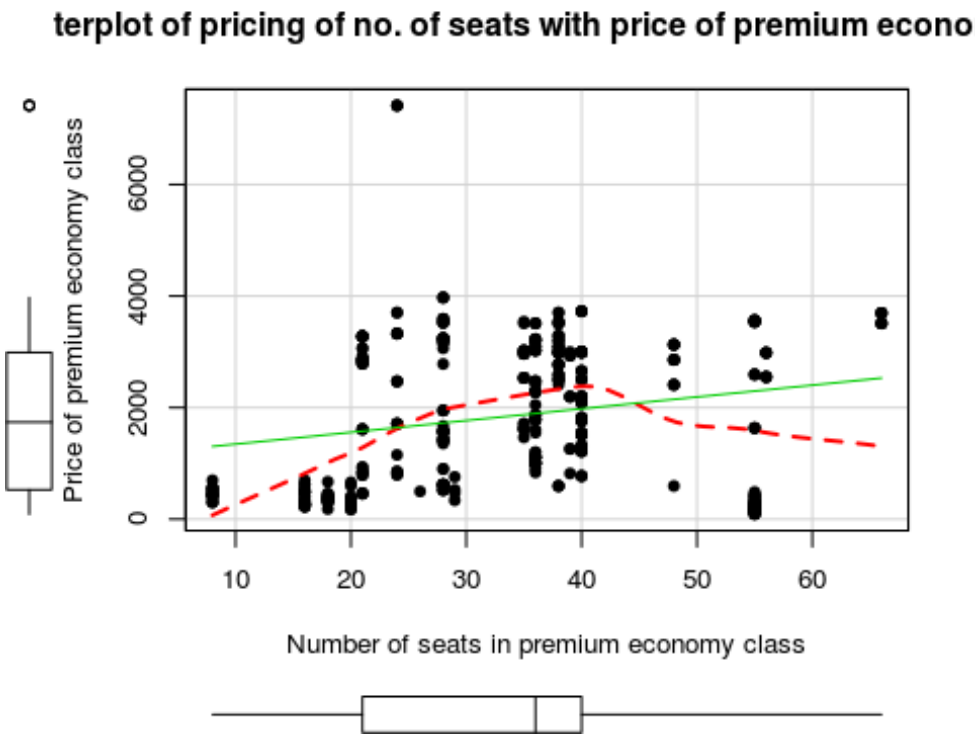

```
scatterplot(PriceRelative ~ WidthDifference, data= airlines,
          spread=FALSE, smoother.args=list(lty=2), pch=19,
          main="Scatter plot of price relative vs Width difference",
          xlab="Width difference",
          ylab="Price relative")
```




```
scatterplot(PriceEconomy~SeatsEconomy, data= airlines,
          spread=FALSE, smoother.args=list(lty=2), pch=19,
          main="Scatterplot of pricing of no. of seats with price of economy
class",
          xlab = "Number of seats in economy class",
          ylab= "Price of economy class")
```

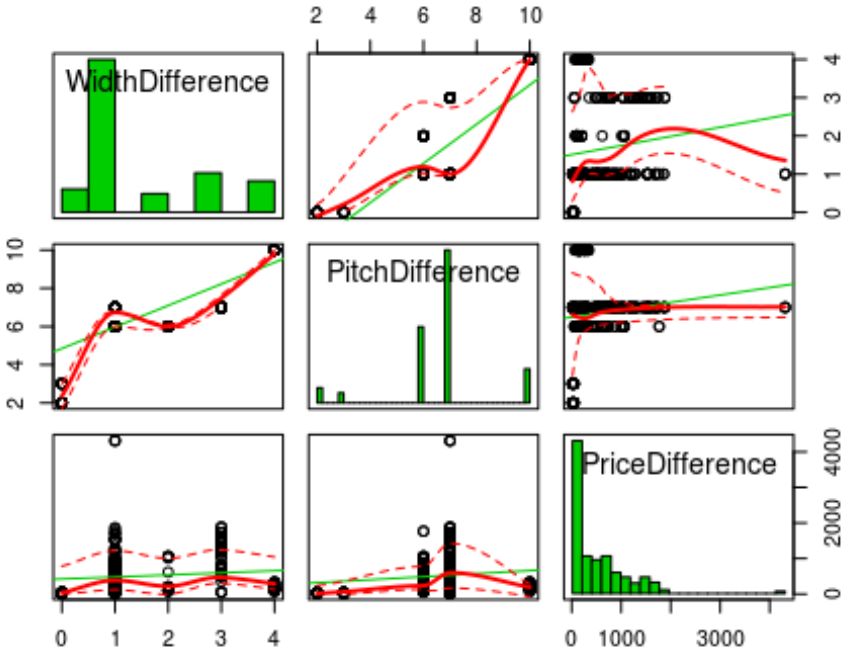



```
scatterplot(PricePremium~SeatsPremium, data= airlines,
          spread=FALSE, smoother.args=list(lty=2), pch=19,
          main="Scatterplot of pricing of no. of seats with price of premium
economy class",
           xlab = "Number of seats in premium economy class",
          ylab= "Price of premium economy class")
```

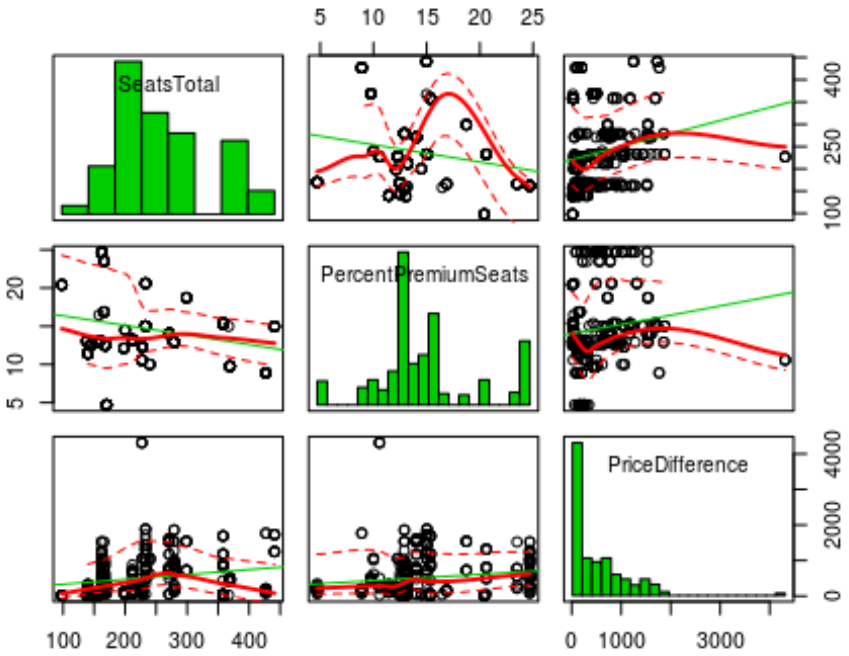


Scatterplot Matrix

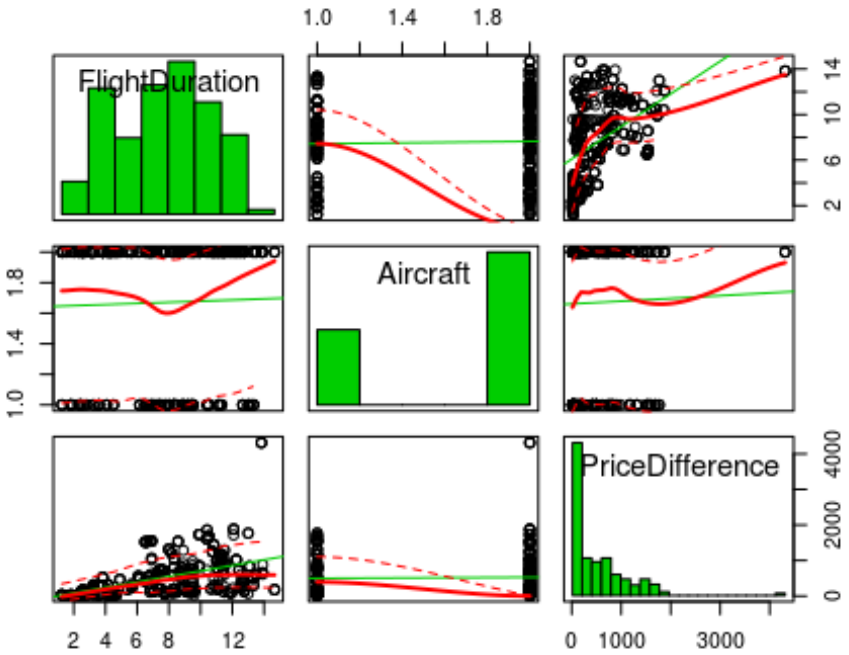
```
scatterplotMatrix( ~  WidthDifference + PitchDifference + PriceDifference ,
                   data = airlines , diagonal = "histogram")
```



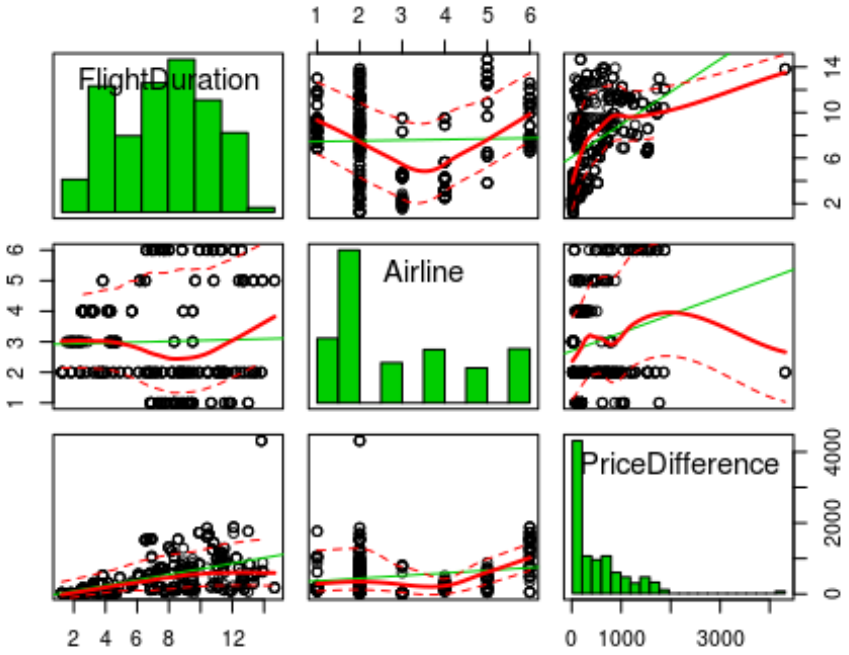

```
scatterplotMatrix( ~  SeatsTotal + PercentPremiumSeats + PriceDifference,
data = airlines , diagonal = "histogram")
```



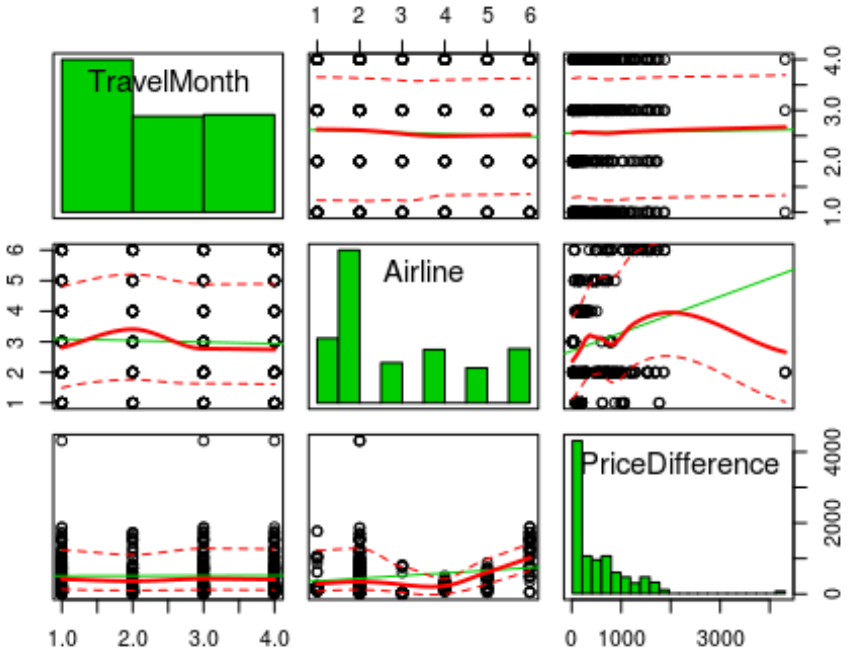
```
scatterplotMatrix( ~  FlightDuration + Aircraft + PriceDifference, data =
airlines , diagonal = "histogram")
```




```
scatterplotMatrix( ~  FlightDuration+ Airline + PriceDifference, data =
airlines , diagonal = "histogram")
```

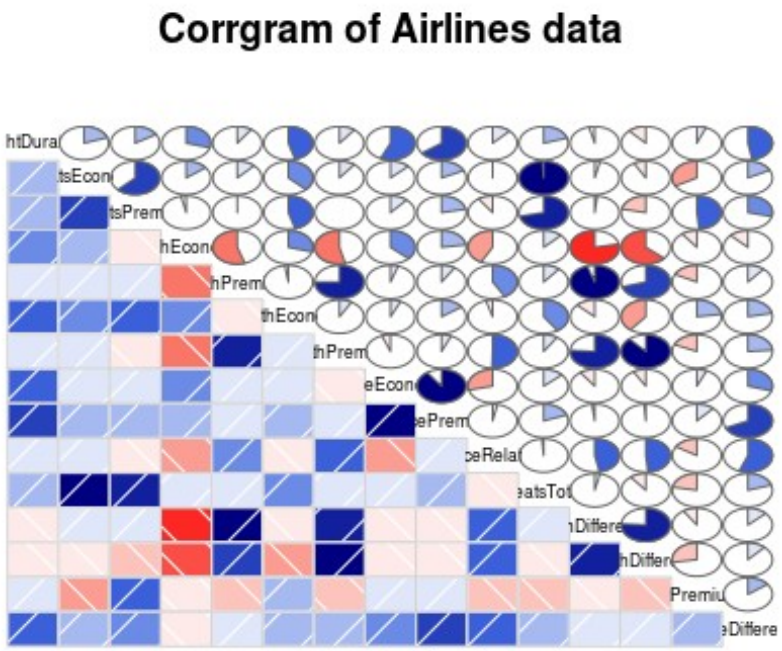



```
scatterplotMatrix( ~  TravelMonth + Airline + PriceDifference, data = airlines ,
diagonal = "histogram")
```



Corrgram

```
library(corrgram)
par(mfrow=c(1,1))
corrgram(airlines, order=FALSE, lower.panel=panel.shade,
        upper.panel=panel.pie, text.panel=panel.txt,
        main="Corrgram of Airlines data")
```




```
correlationmatrix <- cor(airlines[,6:19])
round(correlationmatrix,digits = 2)
```

```
##                  SeatsEconomy SeatsPremium PitchEconomy PitchPremium
## SeatsEconomy             1.00         0.63         0.14         0.12
## SeatsPremium             0.63         1.00        -0.03         0.00
## PitchEconomy             0.14        -0.03         1.00        -0.55
## PitchPremium             0.12         0.00        -0.55         1.00
## WidthEconomy             0.37         0.46         0.29        -0.02
## WidthPremium             0.10         0.00        -0.54         0.75
## PriceEconomy             0.13         0.11         0.37         0.05
## PricePremium             0.18         0.22         0.23         0.09
## PriceRelative            0.00        -0.10        -0.42         0.42
## SeatsTotal               0.99         0.72         0.12         0.11
## PitchDifference          0.04         0.02        -0.78         0.95
## WidthDifference         -0.08        -0.22        -0.64         0.70
## PercentPremiumSeats     -0.33         0.49        -0.10        -0.18
## PriceDifference          0.17         0.29        -0.13         0.11
##                  WidthEconomy WidthPremium PriceEconomy PricePremium
## SeatsEconomy             0.37         0.10         0.13         0.18
```

```
## SeatsPremium               0.46        0.00        0.11        0.22
## PitchEconomy               0.29       -0.54        0.37        0.23
## PitchPremium              -0.02        0.75        0.05        0.09
## WidthEconomy               1.00        0.08        0.07        0.15
## WidthPremium               0.08        1.00       -0.06        0.06
## PriceEconomy               0.07       -0.06        1.00        0.90
## PricePremium               0.15        0.06        0.90        1.00
## PriceRelative             -0.04        0.50       -0.29        0.03
## SeatsTotal                 0.41        0.09        0.13        0.19
## PitchDifference           -0.13        0.76       -0.10       -0.02
## WidthDifference           -0.39        0.88       -0.08       -0.01
## PercentPremiumSeats        0.23       -0.18        0.07        0.12
## PriceDifference            0.22        0.24        0.30        0.68
##                    PriceRelative SeatsTotal PitchDifference
## SeatsEconomy                0.00       0.99           0.04
## SeatsPremium               -0.10       0.72           0.02
## PitchEconomy               -0.42       0.12          -0.78
## PitchPremium                0.42       0.11           0.95
## WidthEconomy               -0.04       0.41          -0.13
## WidthPremium                0.50       0.09           0.76
## PriceEconomy               -0.29       0.13          -0.10
## PricePremium                0.03       0.19          -0.02
## PriceRelative               1.00      -0.01           0.47
## SeatsTotal                 -0.01       1.00           0.03
## PitchDifference             0.47       0.03           1.00
## WidthDifference             0.49      -0.11           0.76
## PercentPremiumSeats        -0.16      -0.22          -0.09
## PriceDifference             0.56       0.20           0.13
##                    WidthDifference PercentPremiumSeats PriceDifference
## SeatsEconomy                 -0.08             -0.33           0.17
## SeatsPremium                 -0.22              0.49           0.29
## PitchEconomy                 -0.64             -0.10          -0.13
## PitchPremium                  0.70             -0.18           0.11
## WidthEconomy                 -0.39              0.23           0.22
## WidthPremium                  0.88             -0.18           0.24
## PriceEconomy                 -0.08              0.07           0.30
## PricePremium                 -0.01              0.12           0.68
## PriceRelative                 0.49             -0.16           0.56
## SeatsTotal                   -0.11             -0.22           0.20
## PitchDifference               0.76             -0.09           0.13
## WidthDifference               1.00             -0.28           0.12
## PercentPremiumSeats          -0.28              1.00           0.15
## PriceDifference               0.12              0.15           1.00
```

Variance- Covariance Matrix

```
VarianceCovariancematrix <- var(airlines[,6:19])
round(VarianceCovariancematrix, 2)
```

```
##                   SeatsEconomy SeatsPremium PitchEconomy PitchPremium
## SeatsEconomy           5832.92      633.07        7.21        11.96
## SeatsPremium            633.07      175.87       -0.30         0.09
## PitchEconomy              7.21       -0.30        0.43        -0.47
## PitchPremium             11.96        0.09       -0.47         1.73
## WidthEconomy             15.91        3.37        0.11        -0.02
## WidthPremium              8.58       -0.04       -0.39         1.08
## PriceEconomy           9673.79     1489.38      238.70        65.43
## PricePremium          17413.25     3717.36      190.85       149.85
## PriceRelative             0.14       -0.58       -0.12         0.25
## SeatsTotal             6465.99      808.94        6.91        12.05
## PitchDifference           4.75        0.38       -0.90         2.20
## WidthDifference          -7.33       -3.41       -0.50         1.10
## PercentPremiumSeats    -122.39       31.15       -0.33        -1.12
## PriceDifference        7739.46     2227.98      -47.85        84.43
##                   WidthEconomy WidthPremium PriceEconomy PricePremium
## SeatsEconomy             15.91        8.58     9673.79     17413.25
## SeatsPremium              3.37       -0.04     1489.38     3717.36
## PitchEconomy              0.11       -0.39      238.70      190.85
## PitchPremium             -0.02        1.08       65.43      149.85
## WidthEconomy              0.31        0.05       37.46      108.12
## WidthPremium              0.05        1.20      -61.85       90.48
## PriceEconomy             37.46      -61.85   976684.06   1147494.77
## PricePremium            108.12       90.48  1147494.77   1659293.12
## PriceRelative            -0.01        0.25     -128.50       18.48
## SeatsTotal               19.28        8.54    11163.18    21130.62
## PitchDifference          -0.12        1.47     -173.28      -41.00
## WidthDifference          -0.26        1.15      -99.32      -17.64
## PercentPremiumSeats       0.61       -0.97      312.61      726.02
## PriceDifference          70.66      152.33   170810.71   511798.35
##                   PriceRelative SeatsTotal PitchDifference
## SeatsEconomy               0.14    6465.99           4.75
## SeatsPremium              -0.58     808.94           0.38
## PitchEconomy              -0.12       6.91          -0.90
## PitchPremium               0.25      12.05           2.20
## WidthEconomy              -0.01      19.28          -0.12
## WidthPremium               0.25       8.54           1.47
## PriceEconomy            -128.50   11163.18        -173.28
## PricePremium              18.48   21130.62         -41.00
## PriceRelative              0.20      -0.44           0.37
## SeatsTotal                -0.44    7274.92           5.13
## PitchDifference            0.37       5.13           3.10
## WidthDifference            0.26     -10.74           1.59
## PercentPremiumSeats       -0.35     -91.24          -0.79
## PriceDifference          146.98    9967.44         132.28
##                   WidthDifference PercentPremiumSeats PriceDifference
## SeatsEconomy                -7.33            -122.39        7739.46
## SeatsPremium                -3.41              31.15        2227.98
## PitchEconomy                -0.50              -0.33         -47.85
## PitchPremium                 1.10              -1.12          84.43
```

```
## WidthEconomy              -0.26          0.61       70.66
## WidthPremium               1.15         -0.97      152.33
## PriceEconomy             -99.32        312.61   170810.71
## PricePremium             -17.64        726.02   511798.35
## PriceRelative              0.26         -0.35      146.98
## SeatsTotal               -10.74        -91.24     9967.44
## PitchDifference            1.59         -0.79      132.28
## WidthDifference            1.41         -1.59       81.68
## PercentPremiumSeats       -1.59         23.45      413.41
## PriceDifference           81.68        413.41   340987.65
```

Hypothesis

H0 : There is no correlation between “Price Difference of Premium Economy and Economy airline seat tickets” and “The variables present in the data provided”. H1 : Alternate Hypothesis i.e. Yes, there is a correlation between the above mentioned variables.

```
newairlines <- airlines
newairlines$Airline <- as.numeric(airlines$Airline)
newairlines$Aircraft <- as.numeric(airlines$Aircraft)
newairlines$IsInternational <- as.numeric(airlines$IsInternational)
newairlines$TravelMonth <- as.numeric(airlines$TravelMonth)
newairlines <- newairlines[order(newairlines$Airline),]
```

T-test

```
t.test(newairlines$PriceEconomy, newairlines$PricePremium)

##
##  Welch Two Sample t-test
##
## data:  newairlines$PriceEconomy and newairlines$PricePremium
## t = -6.8304, df = 856.56, p-value = 1.605e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -667.0831 -369.2793
## sample estimates:
## mean of x mean of y
##  1327.076  1845.258
```

There is a significant difference between pricing of economy class and premium economy class tickets.

Pearson’s Correlation Test

```
cor.test(newairlines$PriceDifference, newairlines$FlightDuration)

##
##  Pearson's product-moment correlation
##
## data:  newairlines$PriceDifference and newairlines$FlightDuration
## t = 11.435, df = 456, p-value < 2.2e-16
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3976578 0.5403379
## sample estimates:
##       cor
## 0.4720837
```

```
cor.test(newairlines$PriceDifference, newairlines$PitchDifference)
```

```
##
##  Pearson's product-moment correlation
##
## data:  newairlines$PriceDifference and newairlines$PitchDifference
## t = 2.7688, df = 456, p-value = 0.005855
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03739893 0.21764764
## sample estimates:
##       cor
## 0.1285851
```

```
cor.test(newairlines$PriceDifference, newairlines$WidthDifference)
```

```
##
##  Pearson's product-moment correlation
##
## data:  newairlines$PriceDifference and newairlines$WidthDifference
## t = 2.5291, df = 456, p-value = 0.01177
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.02627012 0.20700978
## sample estimates:
##       cor
## 0.1176138
```

These three correlations tests suggest that the difference in pricing of the two class of tickets depends strongly on the flightduration and also on the pitch and width difference(p-value<0.05).

Regression Analysis

Consider the following Regression equation

$$PriceDifference=\beta_0+\beta_1 PitchDifference+\beta_2 WidthDifference+\beta_3 FlightDuration+\beta$$

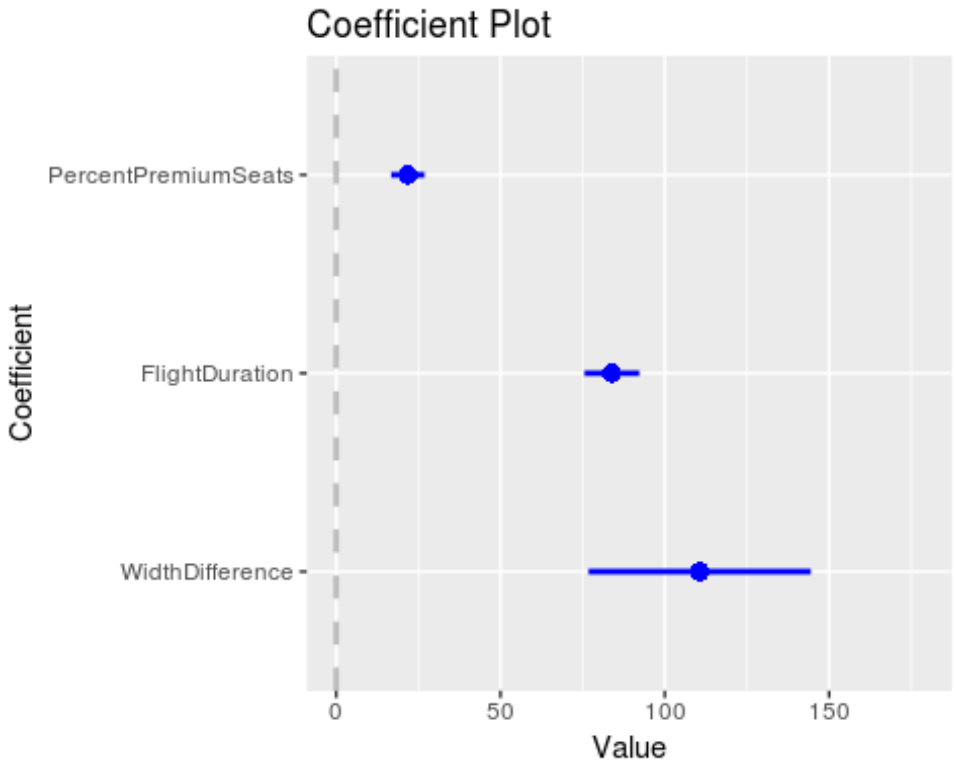
```
fit=lm(PriceDifference ~ PitchDifference + WidthDifference + FlightDuration
+ IsInternational + PercentPremiumSeats, data = airlines)
summary(fit)
```

```
##
## Call:
## lm(formula = PriceDifference ~ PitchDifference + WidthDifference +
##     FlightDuration + IsInternational + PercentPremiumSeats, data =
airlines)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -831.0 -273.9  -57.9  145.6 3434.3
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -590.753    139.161  -4.245 2.65e-05 ***
## PitchDifference              10.907     33.505   0.326  0.74493
## WidthDifference             110.598     33.820   3.270  0.00116 **
## FlightDuration               83.894      8.356  10.040  < 2e-16 ***
## IsInternationalInternational -109.303    168.162  -0.650  0.51603
## PercentPremiumSeats           21.808      5.085   4.289 2.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 496.9 on 452 degrees of freedom
## Multiple R-squared:  0.2837, Adjusted R-squared:  0.2758
## F-statistic:  35.8 on 5 and 452 DF,  p-value: < 2.2e-16

# beta coefficients
fit$coefficients

##                (Intercept)             PitchDifference
##                  -590.75337                    10.90682
##             WidthDifference             FlightDuration
##                   110.59821                    83.89419
## IsInternationalInternational         PercentPremiumSeats
##                  -109.30292                    21.80784

# confidence intervals
confint(fit)

##                                2.5 %     97.5 %
## (Intercept)                -864.23662 -317.27013
## PitchDifference             -54.93797   76.75160
## WidthDifference              44.13500  177.06141
## FlightDuration               67.47306  100.31532
## IsInternationalInternational -439.77839  221.17255
## PercentPremiumSeats           11.81514   31.80054
```

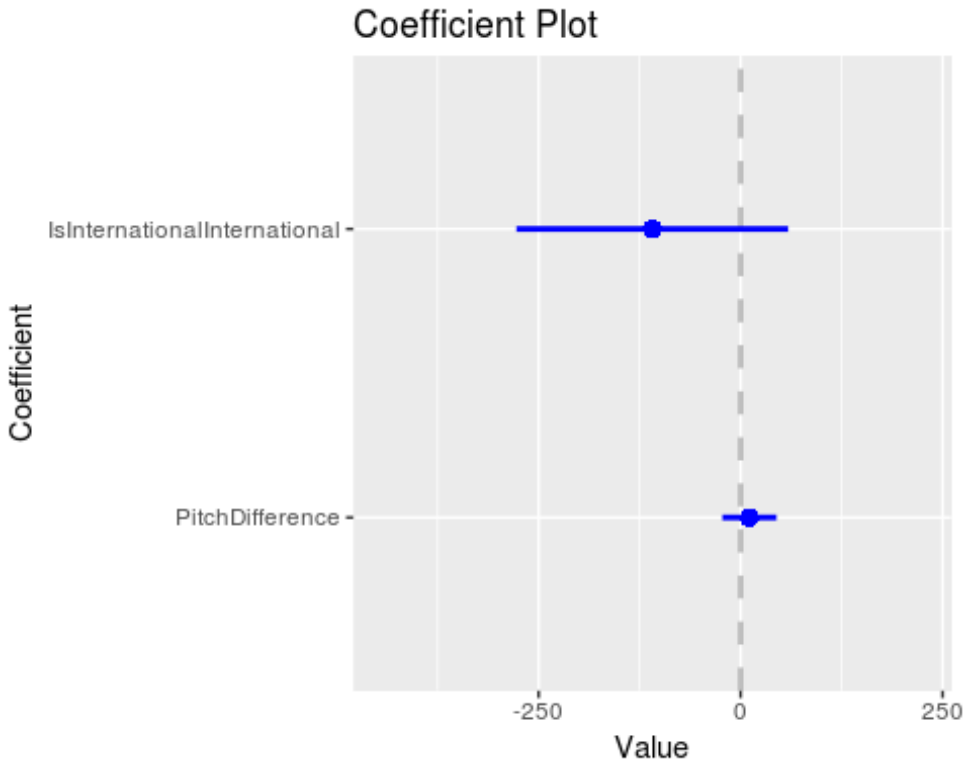
Visualizing the beta coefficients

```
library(coefplot)
# 1. WidthDifference, FlightDuration and PercentPremiumSeats are
statistically significant
```

```
coefplot(fit, predictors=c("WidthDifference", "FlightDuration",
"PercentPremiumSeats"))
```




```
# 2. PitchDifference and IsInternational are NOT statistically significant.
# We infer this since its confidence interval includes zero within it.
coefplot(fit, predictors=c("PitchDifference","IsInternational"))
```



- Hence, the p-values and the coefficients suggest that the model is a good fit and the regression is good and we can reject the null hypothesis.
- Prices of premium economy seats is more with increasing width, pitch and flight duration.
- WidthDifference and FlightDuration are statistically significant for PriceDifference.
- PitchDifference and IsInternational are not statistically significant, since their Confidence Inteval includes 0.