**Output to R code :**

str(airlines)

'data.frame': 462 obs. of 17 variables:

$ AIRLINE : Factor w/ 6 levels "AirFrance","British",..: 1 1 1 1 1 1 1 1 1 1 ...

$ AIRCRAFT : int 0 0 1 0 1 1 1 1 0 0 ...

$ FLIGHT\_DURATION: num 6.91 6.91 9.5 6.91 9.5 13 13 9.5 6.91 8.33 ...

$ MONTH : int 2 3 3 1 2 2 3 1 0 1 ...

$ INTERNATIONAL : int 1 1 1 1 1 1 1 1 1 1 ...

$ SEATS\_ECONOMY : int 216 216 147 216 147 389 389 147 216 200 ...

$ SEATS\_PREMIUM : int 24 24 21 24 21 38 38 21 24 28 ...

$ PITCH\_ECONOMY : int 32 32 32 32 32 32 32 32 32 32 ...

$ PITCH\_PREMIUM : int 38 38 38 38 38 38 38 38 38 38 ...

$ WIDTH\_ECONOMY : int 17 17 18 17 18 18 18 18 17 17 ...

$ WIDTH\_PREMIUM : int 19 19 19 19 19 19 19 19 19 19 ...

$ PRICE\_ECONOMY : int 648 648 630 700 743 1522 1522 990 1094 2918 ...

$ PRICE\_PREMIUM : int 1710 1710 1611 1710 1611 3289 3289 1611 1710 3972 ...

$ PRICE\_RELATIVE : num 1.64 1.64 1.56 1.44 1.17 1.16 1.16 0.63 0.56 0.36 ...

$ N : int 240 240 168 240 168 427 427 168 240 228 ...

$ LAMBDA : num 0.1 0.1 0.13 0.1 0.13 0.09 0.09 0.13 0.1 0.12 ...

$ QUALITY : int 6 6 6 6 6 6 6 6 6 6 ...

summary(airlines)

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 PRICE\_PREMIUM

Min. :34.00 Min. :17.00 Min. :17.00 Min. : 65.0 Min. : 86

1st Qu.:38.00 1st Qu.:17.00 1st Qu.:19.00 1st Qu.: 404.8 1st Qu.: 524

Median :38.00 Median :18.00 Median :19.00 Median :1224.0 Median :1710

Mean :37.92 Mean :17.83 Mean :19.48 Mean :1317.1 Mean :1832

3rd Qu.:38.00 3rd Qu.:18.00 3rd Qu.:21.00 3rd Qu.:1903.0 3rd Qu.:2989

Max. :40.00 Max. :19.00 Max. :21.00 Max. :3593.0 Max. :7414

PRICE\_RELATIVE N LAMBDA QUALITY

Min. :0.0200 Min. : 38.0 Min. :0.0500 Min. : 2.000

1st Qu.:0.1000 1st Qu.:162.0 1st Qu.:0.1200 1st Qu.: 6.000

Median :0.3800 Median :227.0 Median :0.1300 Median : 7.000

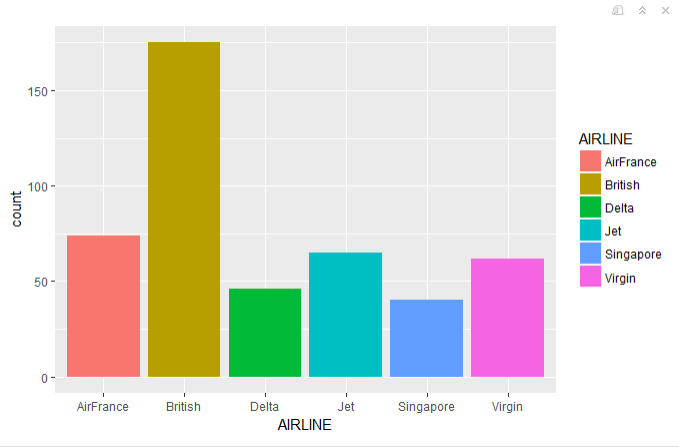
Mean :0.4926 Mean :234.2 Mean :0.1503 Mean : 6.716

3rd Qu.:0.7475 3rd Qu.:279.0 3rd Qu.:0.1500 3rd Qu.: 7.000

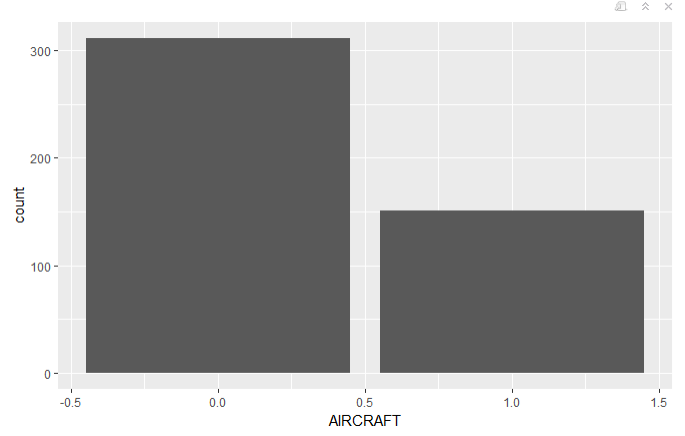
Max. :1.8900 Max. :441.0 Max. :0.5500 Max. :10.000

#Plotting barchart for the airline frequency distribution

ggplot(airlines, aes(x = AIRLINE, fill = AIRLINE)) + geom\_bar()



ggplot(airlines, aes(x = AIRCRAFT, fill = AIRCRAFT)) + geom\_bar()



#Distribution showing aircrafts used by the airlines

table(airlines$AIRLINE, airlines$AIRCRAFT)

0 1

AirFrance 38 36

British 128 47

Delta 34 12

Jet 58 7

Singapore 24 16

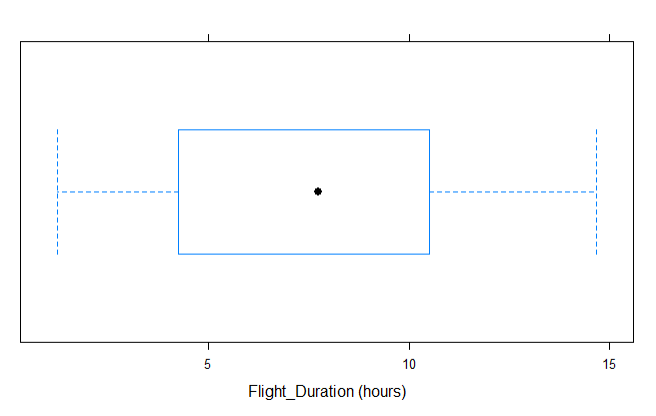
Virgin 29 33

summary(airlines$FLIGHT\_DURATION)

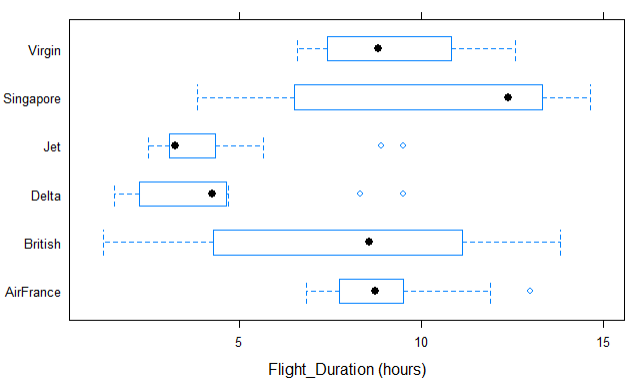
Min. 1st Qu. Median Mean 3rd Qu. Max.

1.250 4.250 7.750 7.549 10.500 14.660

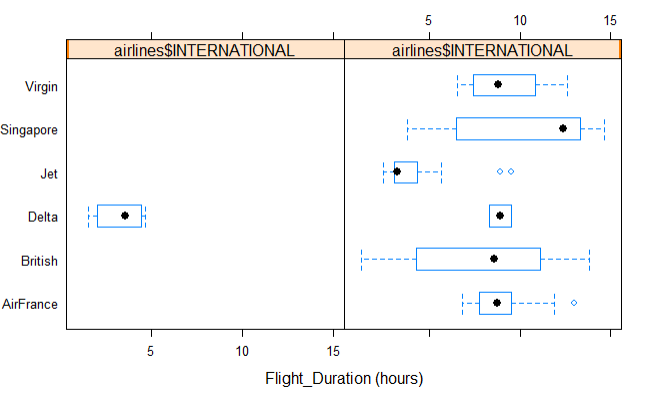
#Box plot for flight duration to understand its distribution



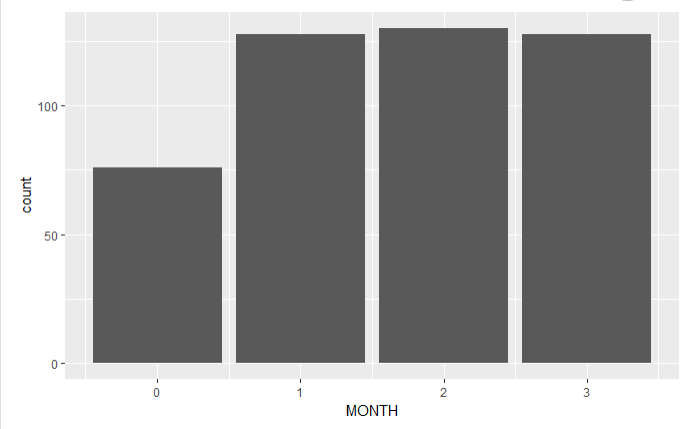
#Box plot showing the distribution for flight duration of different airlines side by side



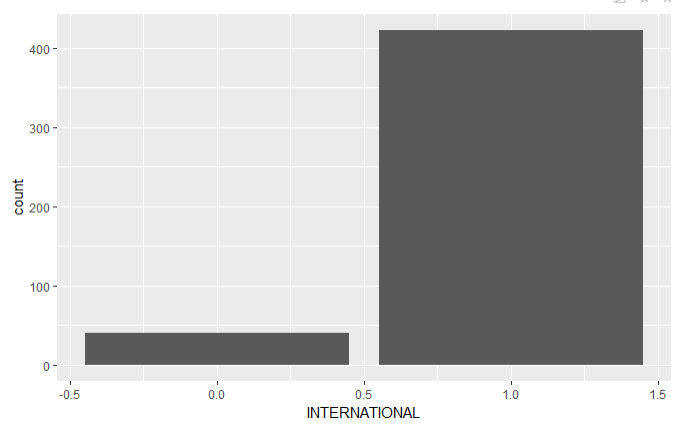
#Box plot showing the distribution for flight duration of different airlines based on whether the flight is domestic or international side by side



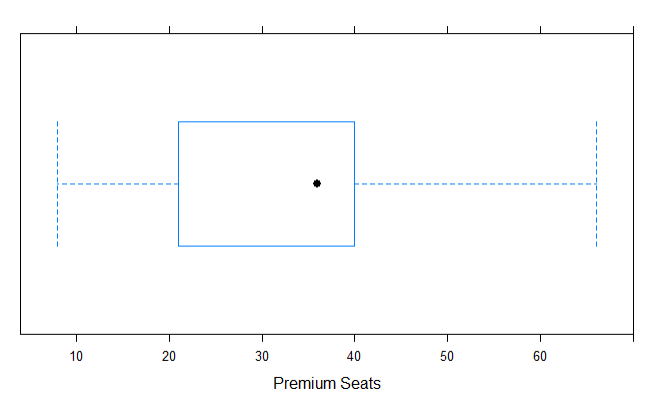
#Plot to see in which months are the flights most frequent i.e. from July, Aug, Sept and Oct.

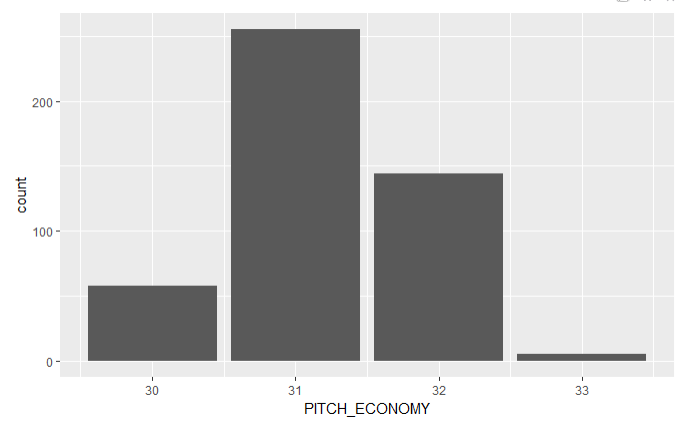


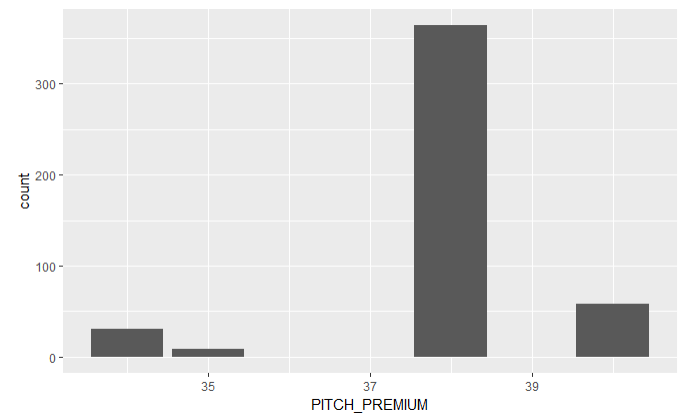
Plot to see how many flights are international and how many aren't.

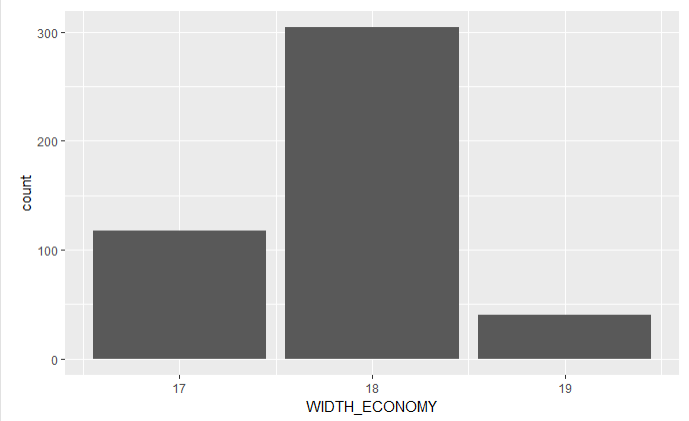


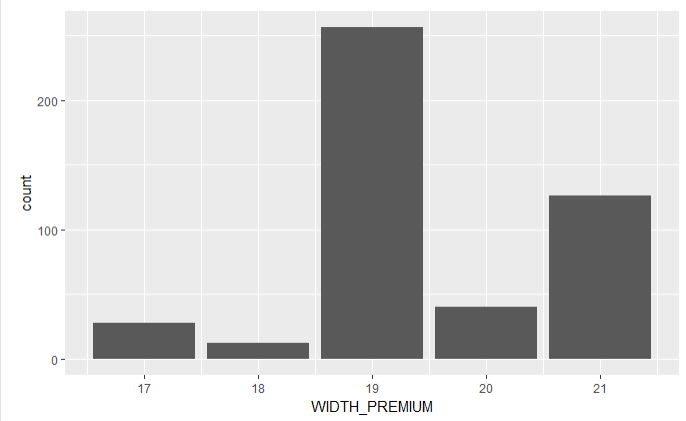
Distribution of economy and premium seats

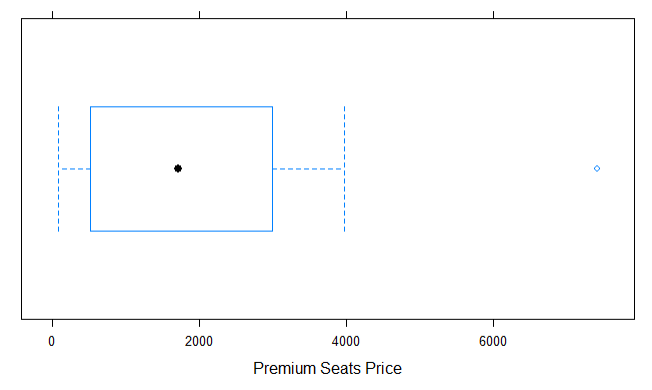


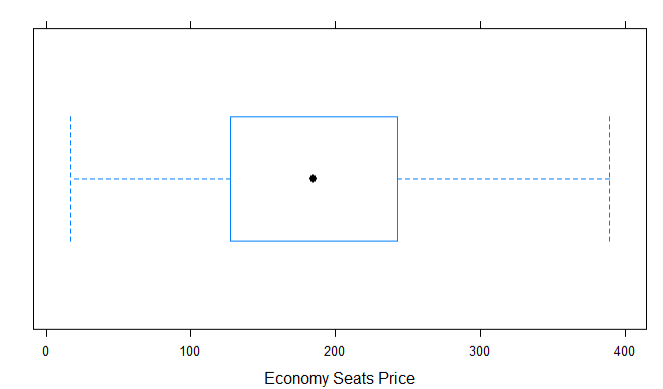




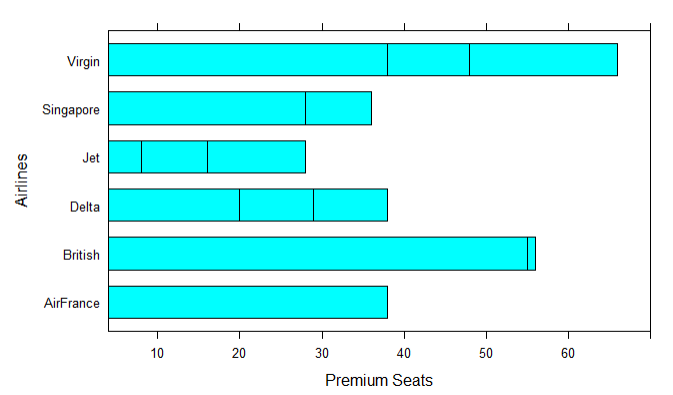




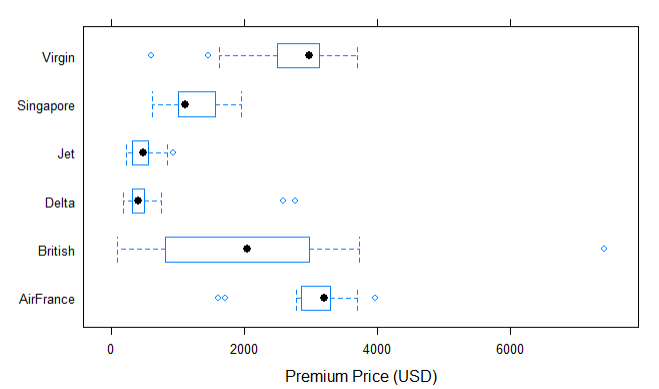




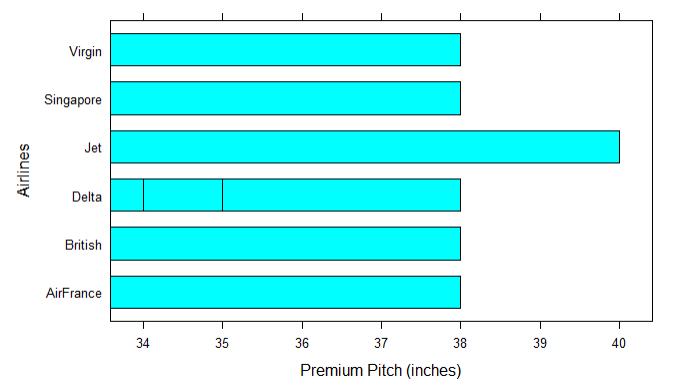
#Distribution showing airlines and premium seats distribution

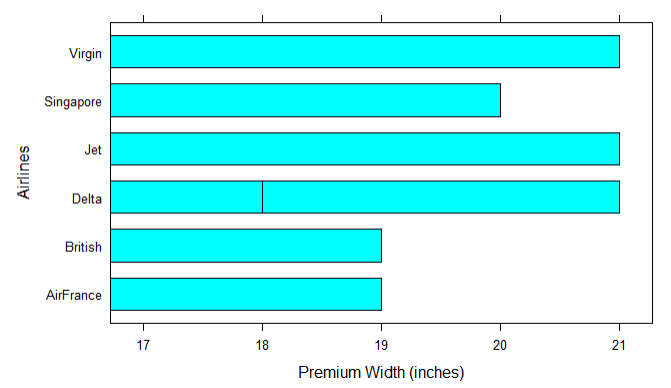


#Let's do a box plot for the premium seat price based on the airlines.

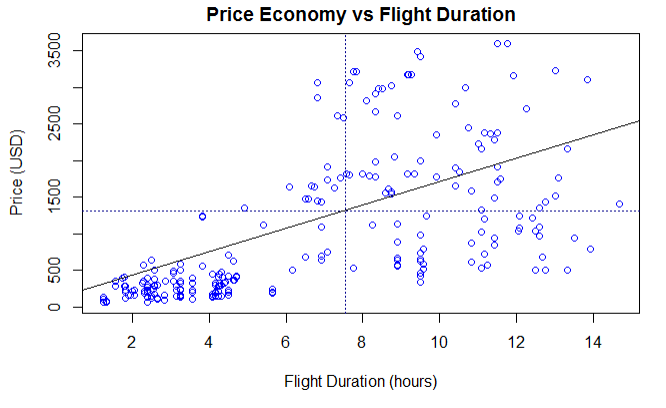


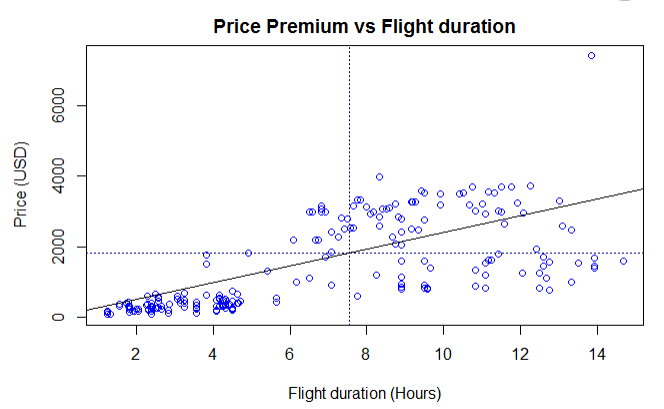
Let's see the distribution of premium width and pitch based on the diff. airlines.



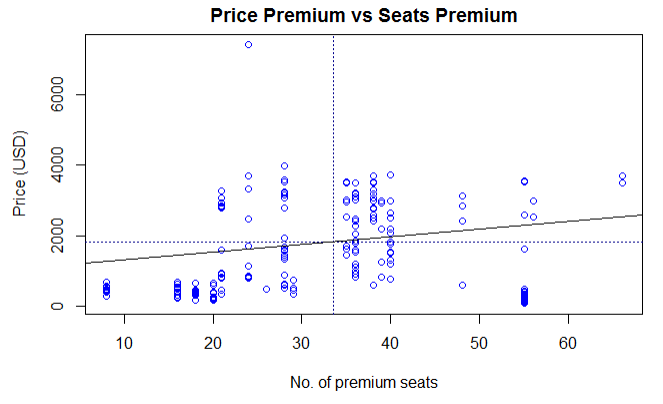


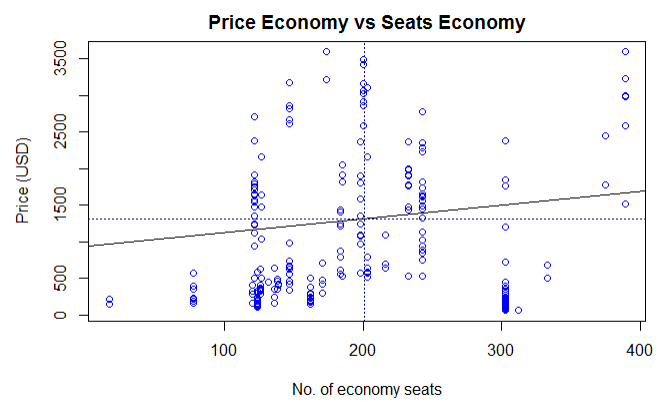
Let's do a scatterplot of price economy vs flight duration –

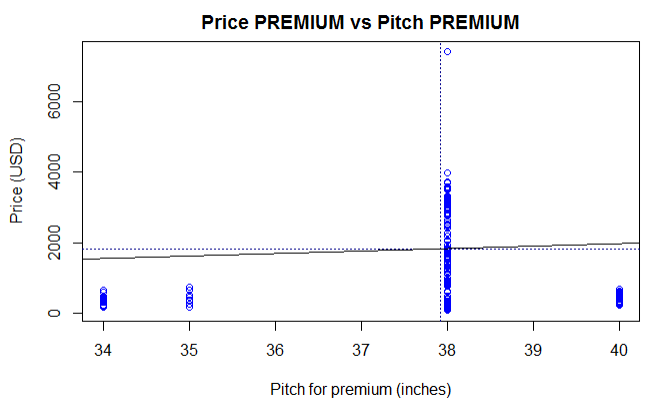


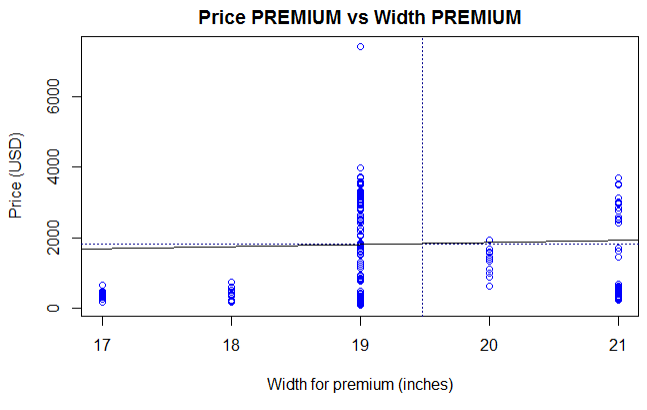


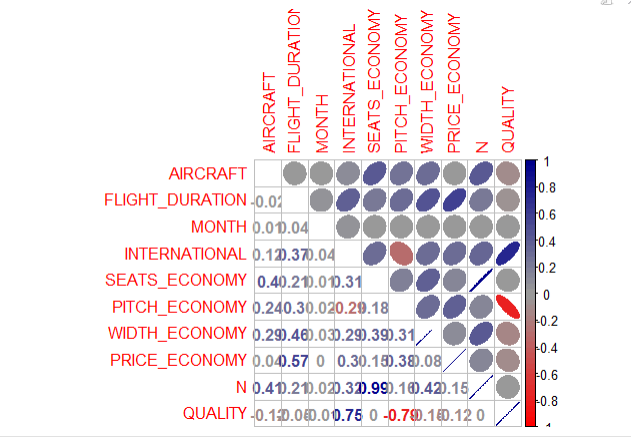
Let's do a similar exercise for seats\_premium and price\_premium.

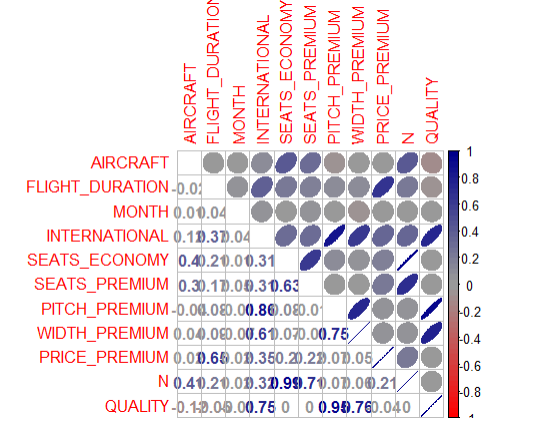


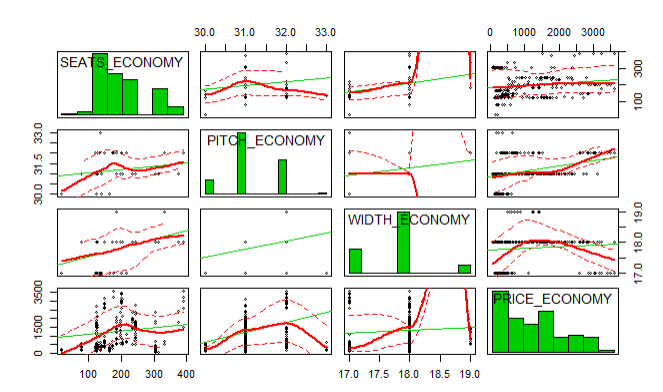


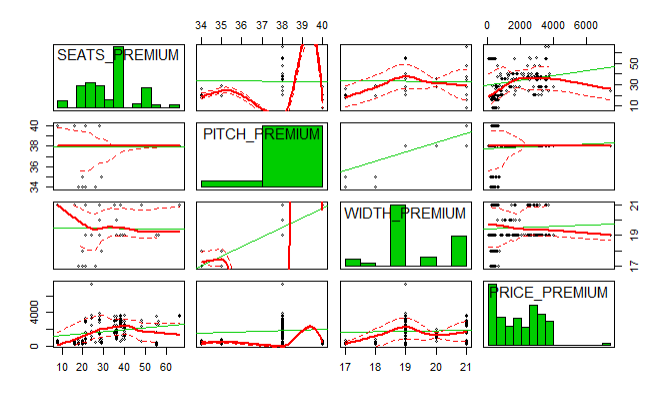












Pearson's product-moment correlation

data: airlines$PRICE\_ECONOMY and airlines$PITCH\_ECONOMY

t = 8.8003, df = 460, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.2987210 0.4550742

sample estimates:

cor

0.379605

Pearson's product-moment correlation

data: airlines$PRICE\_ECONOMY and airlines$WIDTH\_ECONOMY

t = 1.764, df = 460, p-value = 0.0784

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.009330795 0.171911298

sample estimates:

cor

0.0819679

Pearson's product-moment correlation

data: airlines$PRICE\_PREMIUM and airlines$PITCH\_PREMIUM

t = 1.5338, df = 460, p-value = 0.1258

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.02002801 0.16150915

sample estimates:

cor

0.07133125

Pearson's product-moment correlation

data: airlines$PRICE\_PREMIUM and airlines$WIDTH\_PREMIUM

t = 1.0592, df = 460, p-value = 0.2901

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.04209336 0.13992426

sample estimates:

cor

0.04932498

t.test(airlines$PRICE\_PREMIUM~airlines$AIRCRAFT,var.equal = TRUE)

Two Sample t-test

data: airlines$PRICE\_PREMIUM by airlines$AIRCRAFT

t = -0.43103, df = 460, p-value = 0.6666

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-306.8544 196.4588

sample estimates:

mean in group 0 mean in group 1

1814.305 1869.503

Since the p-value is > 0.05, we fail to reject the null hypothesis. Thus, we can say that type of aircraft does not affect the premium prices.

# Formulating a regression model -

Hypothesis - The premium seat prices are affected by flight duration and no. of premium seats offered in a flight as well as width and pitch.

y = premium prices

x1 = flight duration

x2 = width premium

x3 = pitch premium

x4 = seats\_premium

Call:

lm(formula = PRICE\_PREMIUM ~ FLIGHT\_DURATION + WIDTH\_PREMIUM +

PITCH\_PREMIUM + SEATS\_PREMIUM, data = airlines)

Residuals:

Min 1Q Median 3Q Max

-2364.7 -615.1 31.3 733.2 4212.7

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1311.690 1329.358 -0.987 0.32431

FLIGHT\_DURATION 230.206 12.997 17.713 < 2e-16 \*\*\*

WIDTH\_PREMIUM -59.732 62.558 -0.955 0.34017

PITCH\_PREMIUM 57.544 52.054 1.105 0.26954

SEATS\_PREMIUM 11.562 3.456 3.345 0.00089 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 969.3 on 457 degrees of freedom

Multiple R-squared: 0.4403, Adjusted R-squared: 0.4354

F-statistic: 89.87 on 4 and 457 DF, p-value: < 2.2e-16

Call:

lm(formula = PRICE\_PREMIUM ~ FLIGHT\_DURATION + SEATS\_PREMIUM +

QUALITY + PITCH\_PREMIUM + WIDTH\_PREMIUM, data = airlines)

Residuals:

Min 1Q Median 3Q Max

-2268.9 -657.6 42.9 717.4 4376.0

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -10050.497 4134.568 -2.431 0.015448 \*

FLIGHT\_DURATION 215.311 14.561 14.787 < 2e-16 \*\*\*

SEATS\_PREMIUM 12.502 3.467 3.606 0.000345 \*\*\*

QUALITY -211.814 94.941 -2.231 0.026166 \*

PITCH\_PREMIUM 306.858 123.184 2.491 0.013091 \*

WIDTH\_PREMIUM -19.326 64.867 -0.298 0.765892

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 965.1 on 456 degrees of freedom

Multiple R-squared: 0.4463, Adjusted R-squared: 0.4402

F-statistic: 73.51 on 5 and 456 DF, p-value: < 2.2e-16