# Analysis of Airline Ticket Pricing

# NAME: <ASHUTOSH SHAW>

# EMAIL: <ashu.shaw09@gmail.com>

# COLLEGE: <DTU>

**OUTPUTS**

summary(flight.df)

#output for the summary generated

Airline Aircraft FlightDuration TravelMonth IsInternational SeatsEconomy SeatsPremium PitchEconomy PitchPremium

AirFrance: 74 AirBus:151 Min. : 1.250 Aug:127 Domestic : 40 Min. : 78.0 Min. : 8.00 Min. :30.00 Min. :34.00

British :175 Boeing:307 1st Qu.: 4.260 Jul: 75 International:418 1st Qu.:133.0 1st Qu.:21.00 1st Qu.:31.00 1st Qu.:38.00

Delta : 46 Median : 7.790 Oct:127 Median :185.0 Median :36.00 Median :31.00 Median :38.00

Jet : 61 Mean : 7.578 Sep:129 Mean :202.3 Mean :33.65 Mean :31.22 Mean :37.91

Singapore: 40 3rd Qu.:10.620 3rd Qu.:243.0 3rd Qu.:40.00 3rd Qu.:32.00 3rd Qu.:38.00

Virgin : 62 Max. :14.660 Max. :389.0 Max. :66.00 Max. :33.00 Max. :40.00

WidthEconomy WidthPremium PriceEconomy PricePremium PriceRelative SeatsTotal PitchDifference WidthDifference PercentPremiumSeats

Min. :17.00 Min. :17.00 Min. : 65 Min. : 86.0 Min. :0.0200 Min. : 98 Min. : 2.000 Min. :0.000 Min. : 4.71

1st Qu.:18.00 1st Qu.:19.00 1st Qu.: 413 1st Qu.: 528.8 1st Qu.:0.1000 1st Qu.:166 1st Qu.: 6.000 1st Qu.:1.000 1st Qu.:12.28

Median :18.00 Median :19.00 Median :1242 Median :1737.0 Median :0.3650 Median :227 Median : 7.000 Median :1.000 Median :13.21

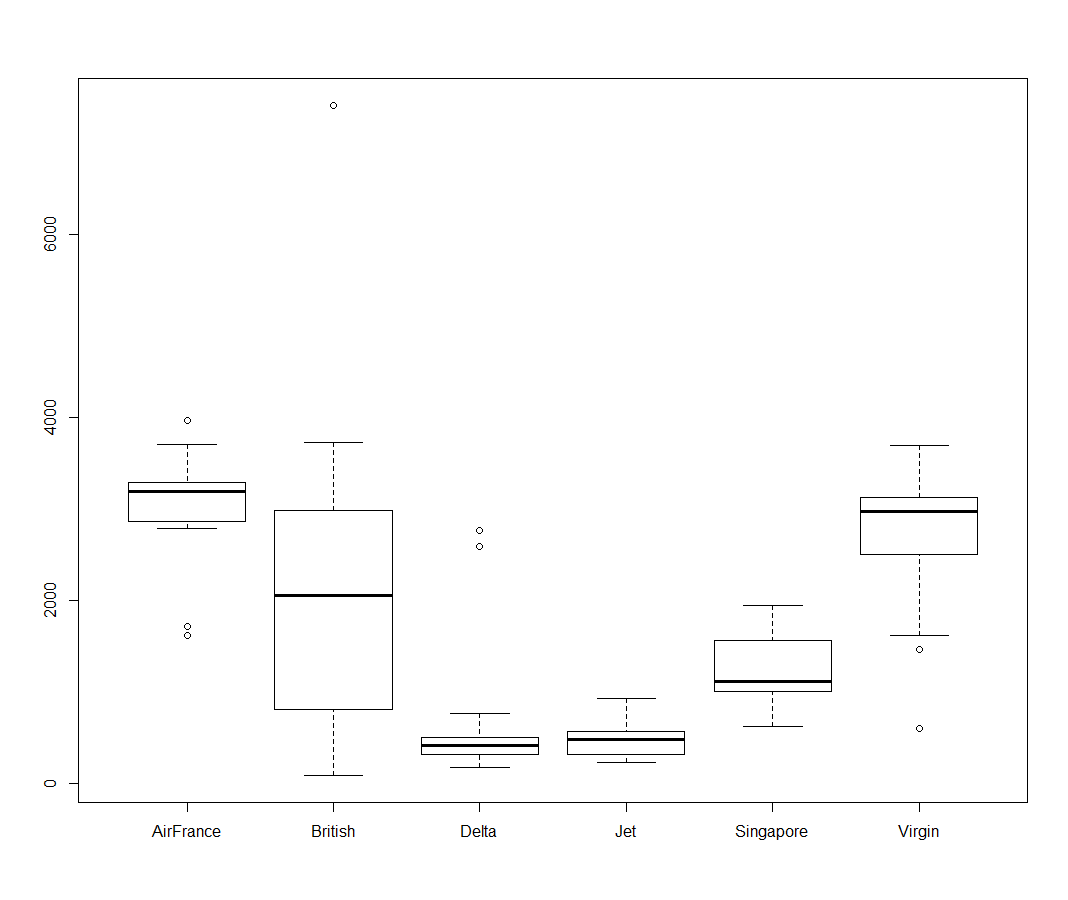
Mean :17.84 Mean :19.47 Mean :1327 Mean :1845.3 Mean :0.4872 Mean :236 Mean : 6.688 Mean :1.633 Mean :14.65

3rd Qu.:18.00 3rd Qu.:21.00 3rd Qu.:1909 3rd Qu.:2989.0 3rd Qu.:0.7400 3rd Qu.:279 3rd Qu.: 7.000 3rd Qu.:3.000 3rd Qu.:15.36

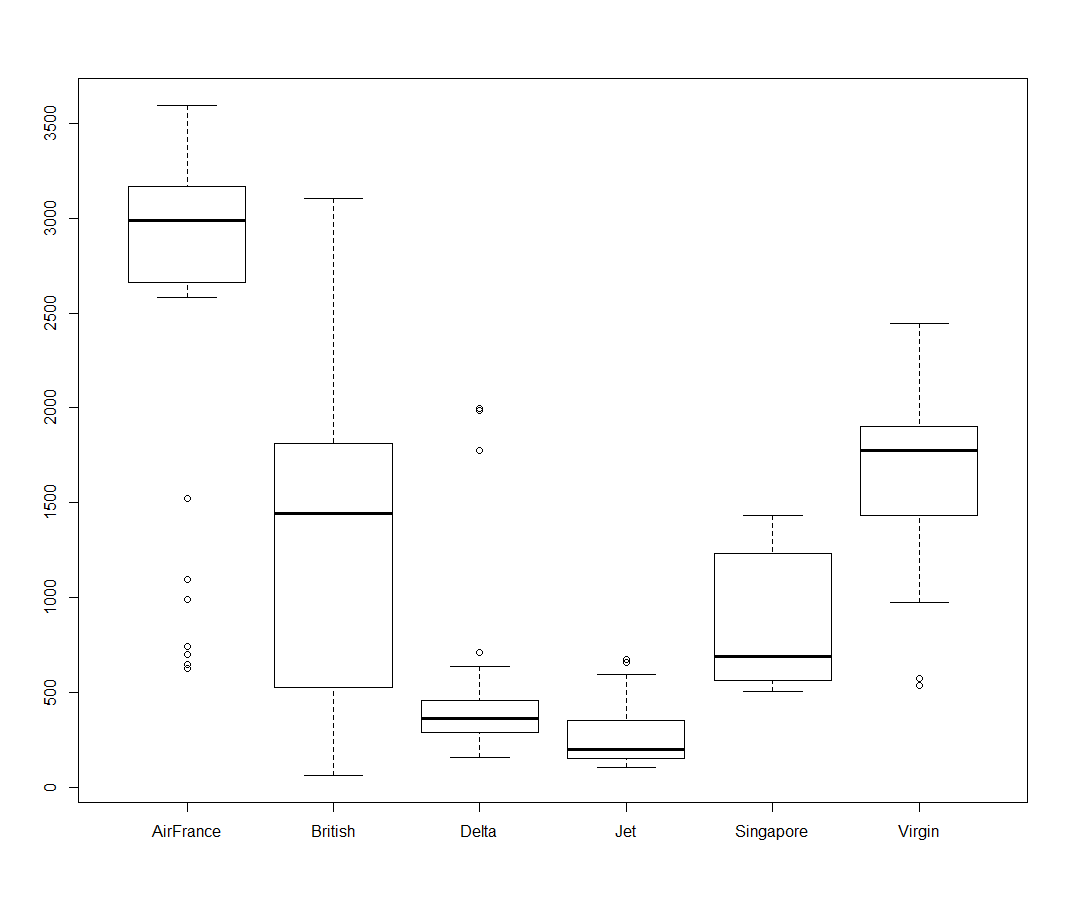
Max. :19.00 Max. :21.00 Max. :3593 Max. :7414.0 Max. :1.8900 Max. :441 Max. :10.000 Max. :4.000 Max. :24.69

# Visualizing using plots

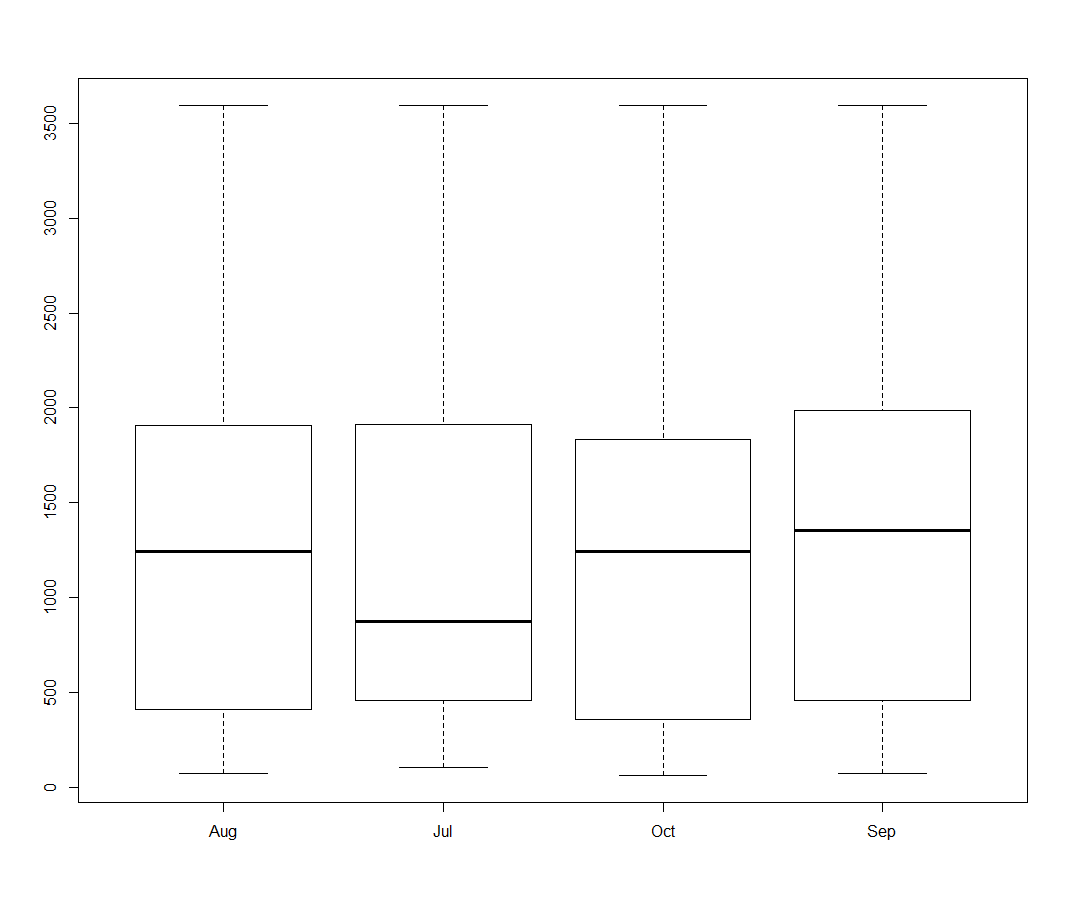
> boxplot(flight.df$PricePremium~flight.df$Airline)



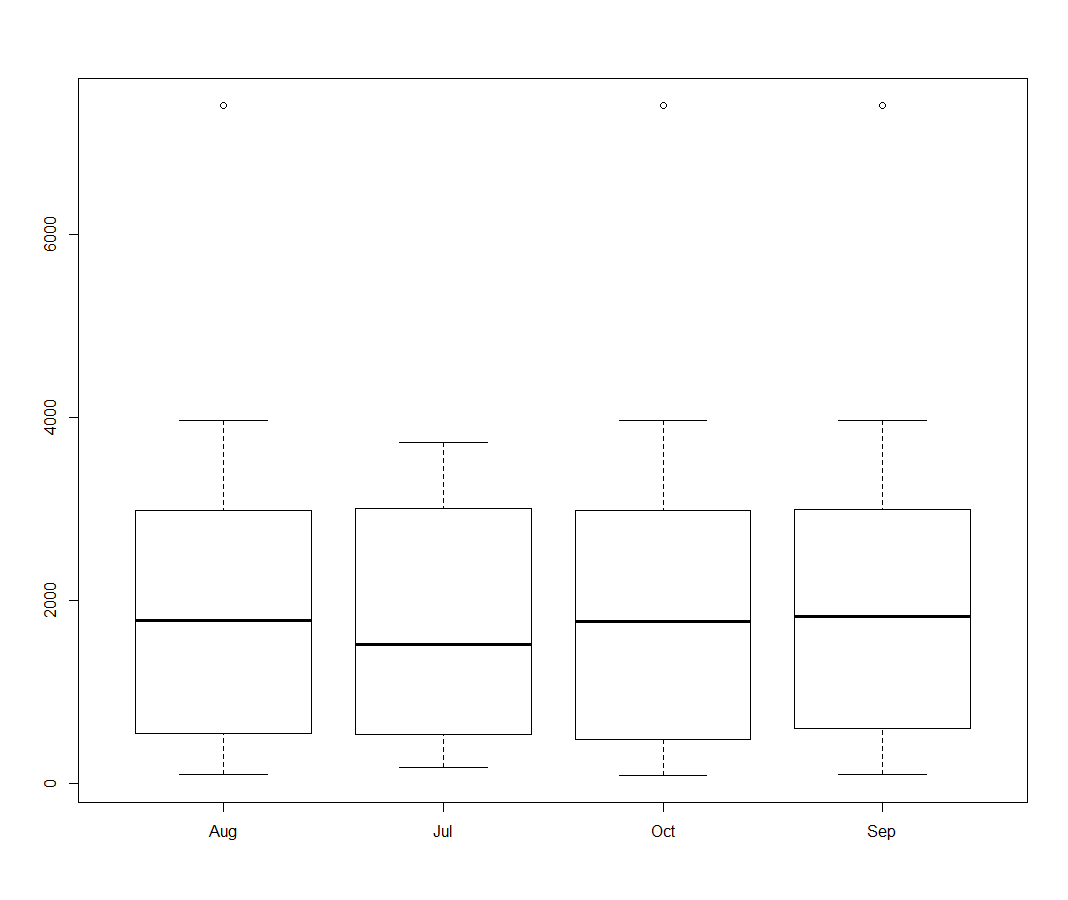
boxplot(flight.df$PriceEconomy~flight.df$Airline)



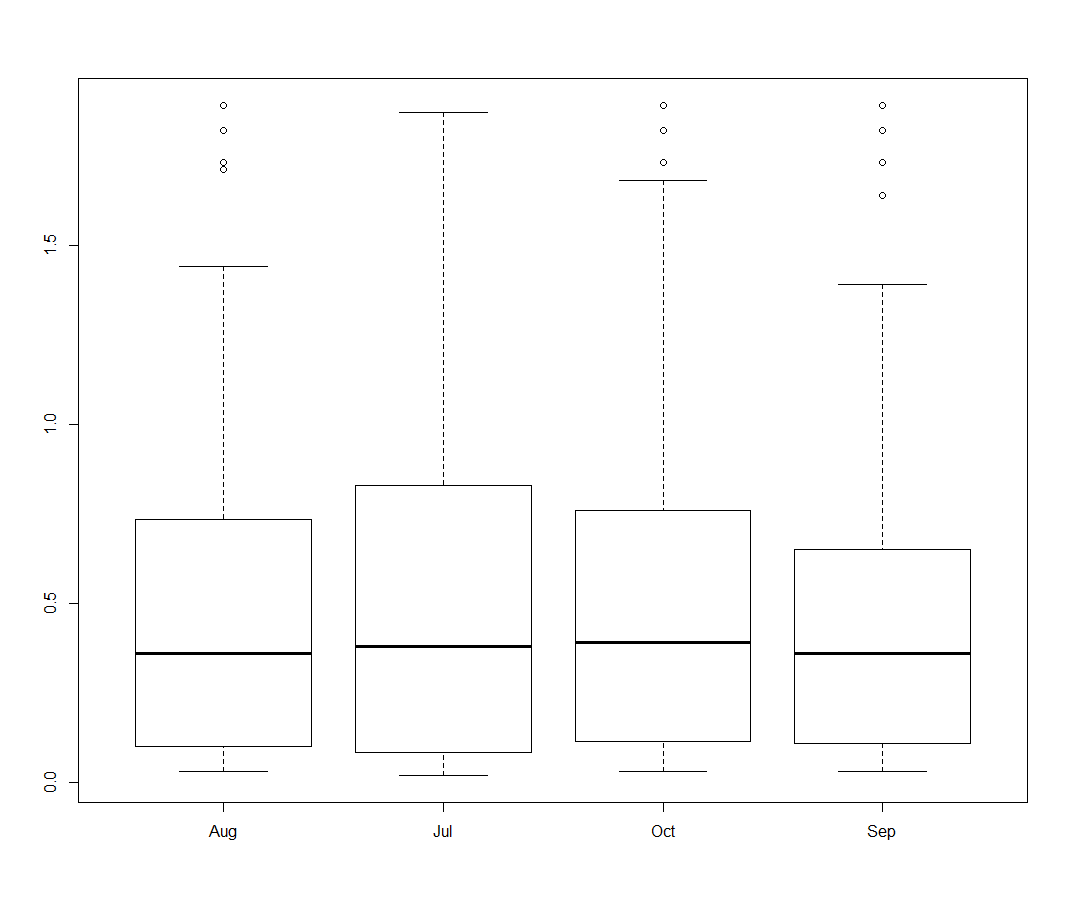
boxplot(flight.df$PriceEconomy~flight.df$TravelMonth)



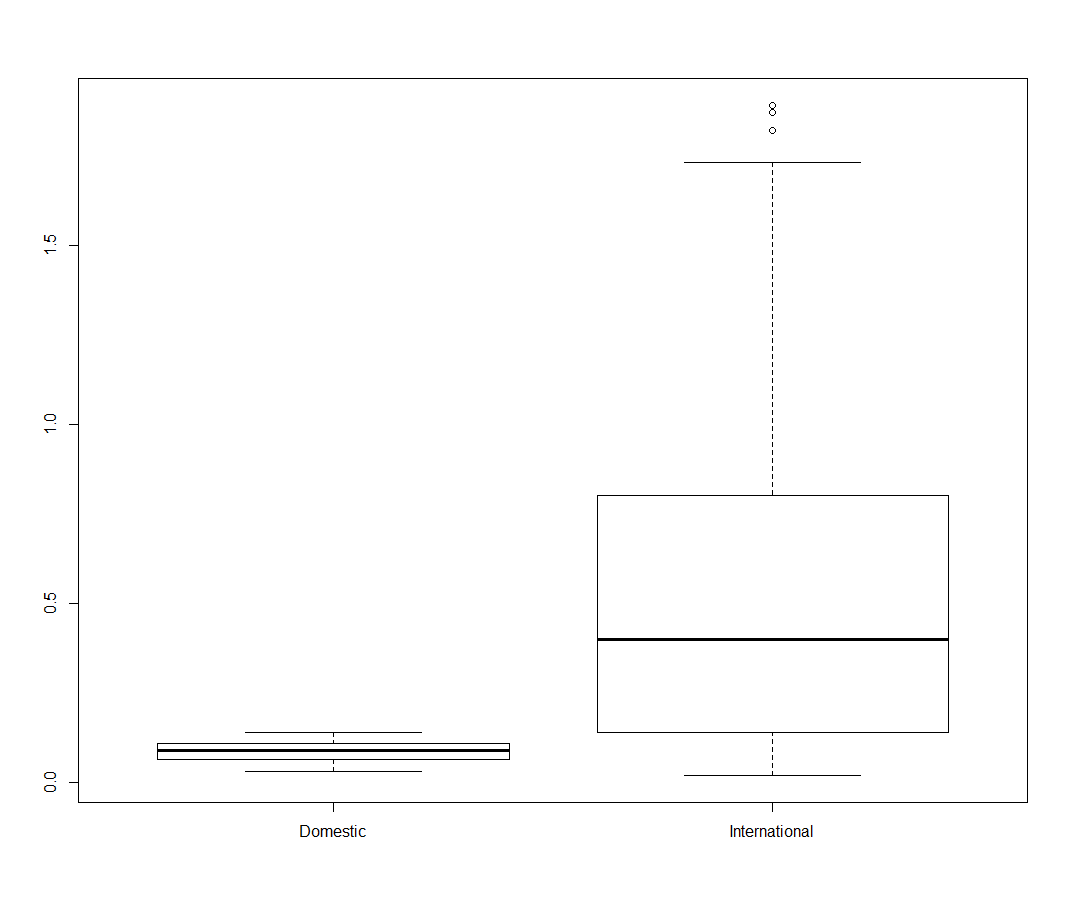
boxplot(flight.df$PricePremium~flight.df$TravelMonth)



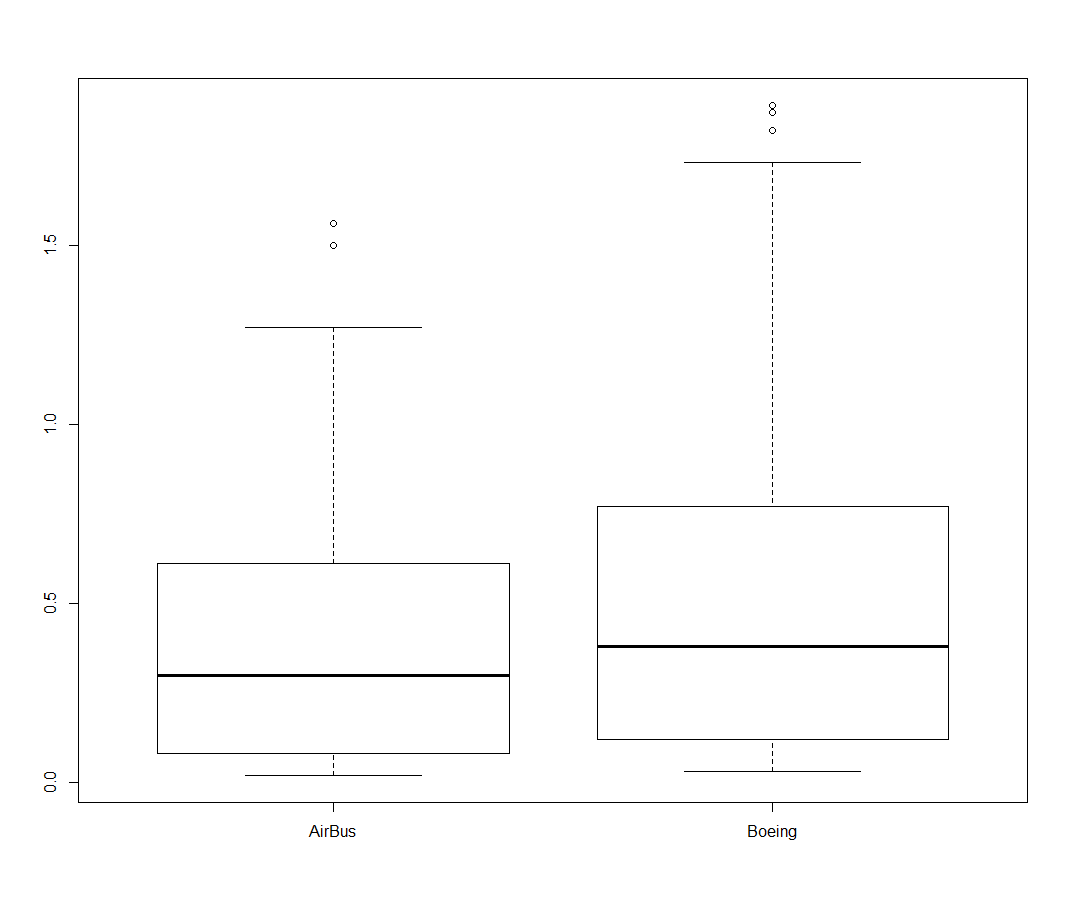
boxplot(flight.df$PriceRelative~flight.df$TravelMonth)



boxplot(flight.df$PriceRelative~flight.df$IsInternational)

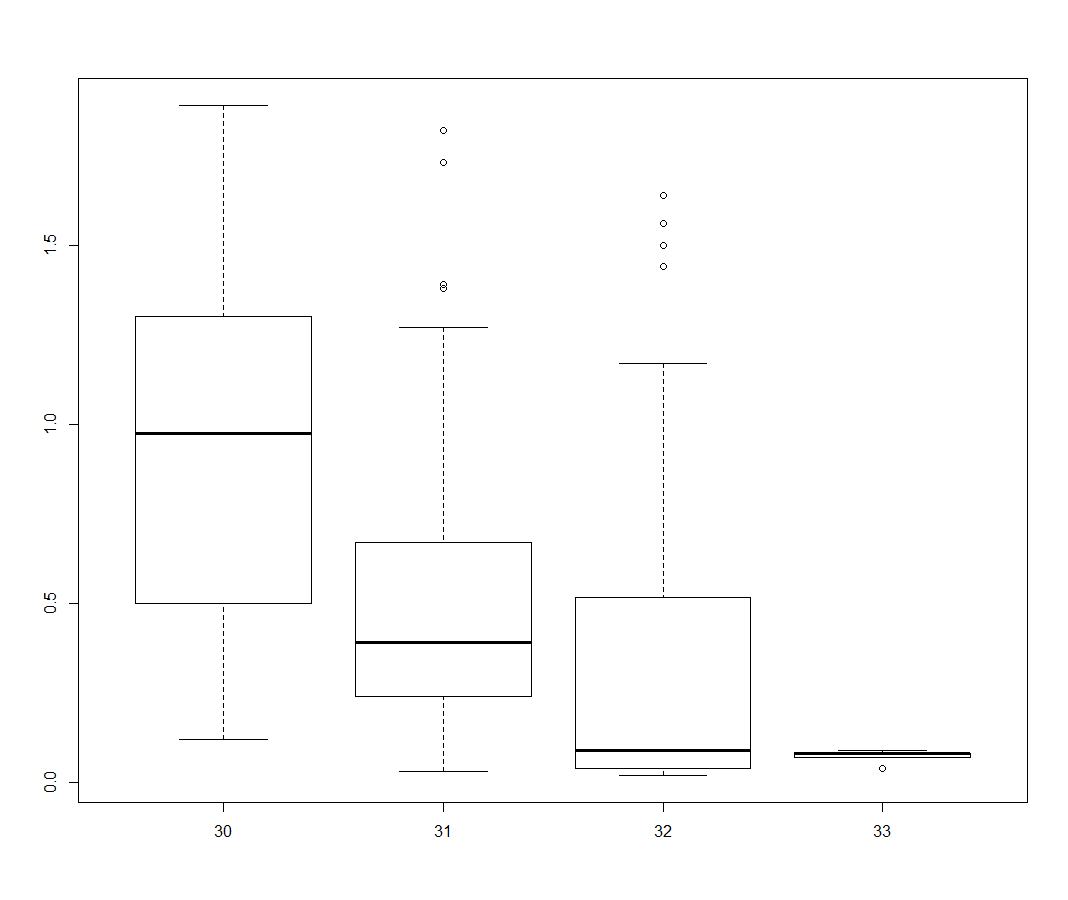


boxplot(flight.df$PriceRelative~flight.df$Aircraft)



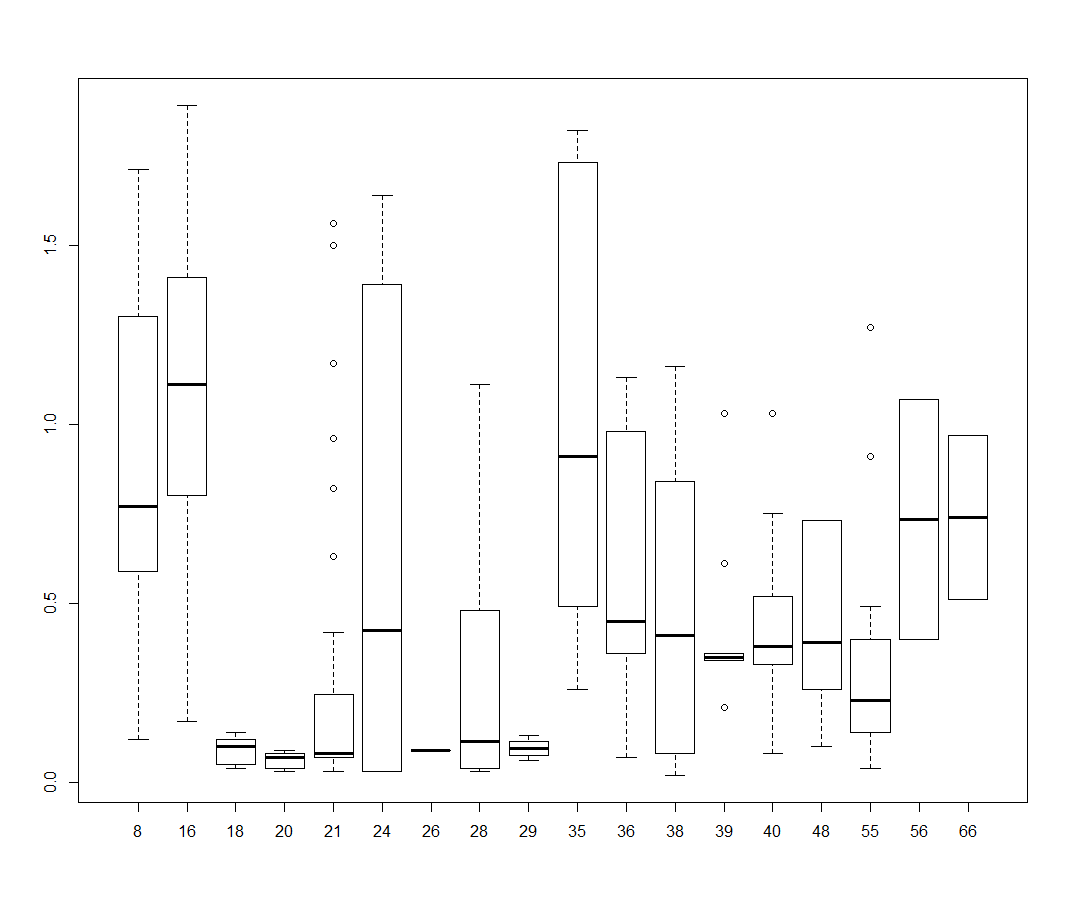
> x1<-factor(flight.df$SeatsPremium)

> boxplot(flight.df$PriceRelative~x1)



x2<-factor(flight.df$PitchEconomy)

> boxplot(flight.df$PriceRelative~x2)

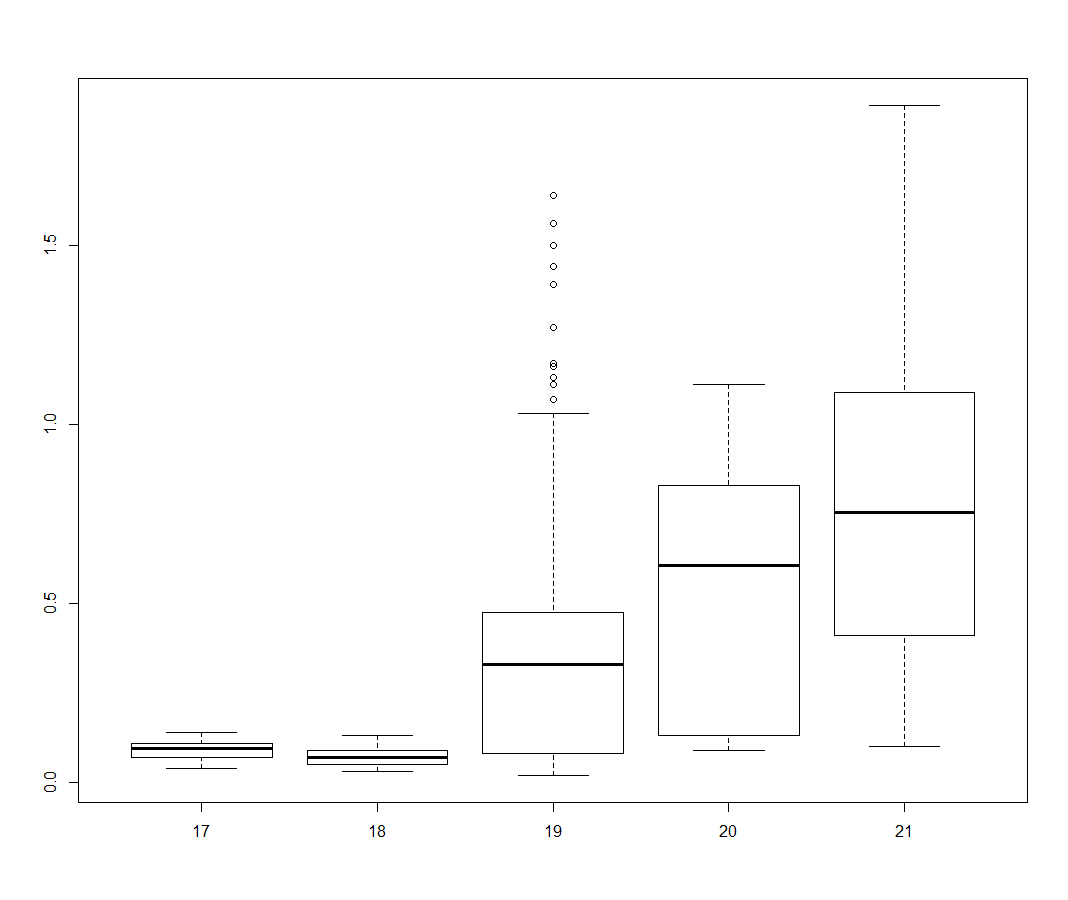


x3<-factor(flight.df$PitchPremium)

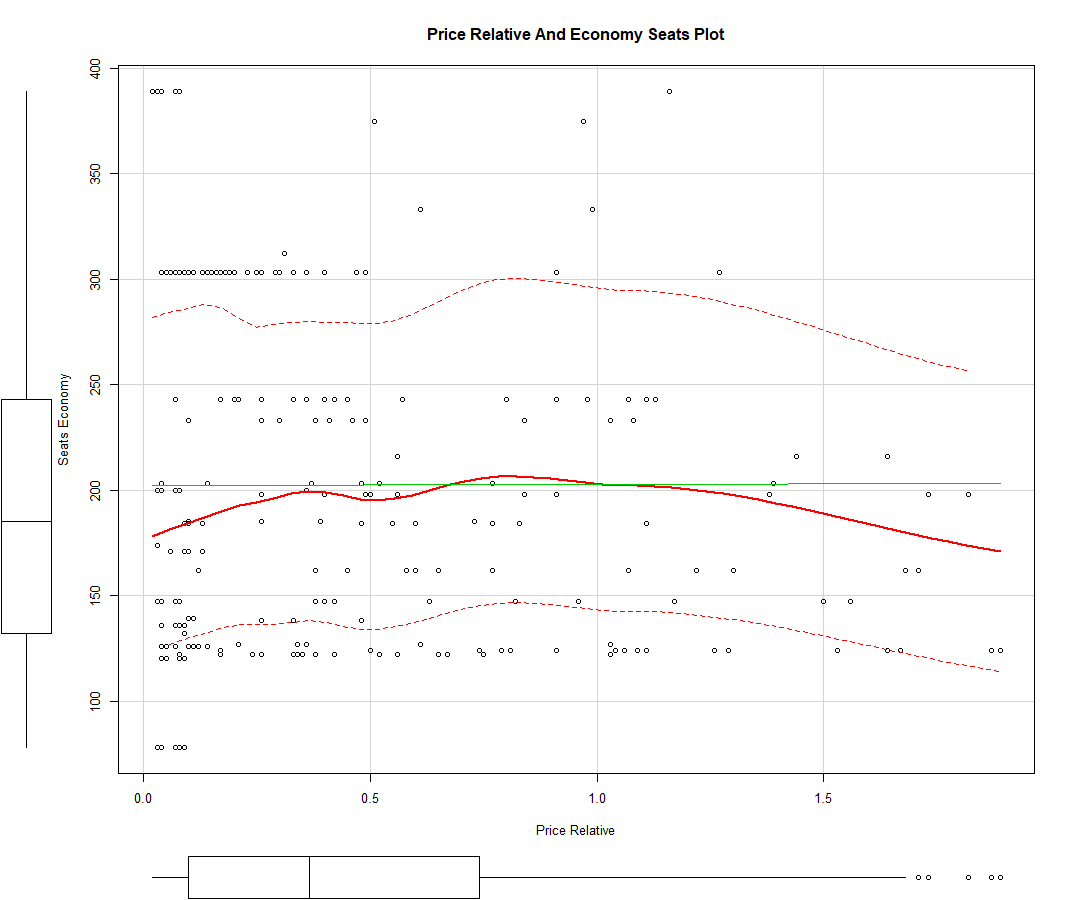
> boxplot(flight.df$PriceRelative~x3)

x4<-factor(flight.df$WidthPremium)

> boxplot(flight.df$PriceRelative~x4)



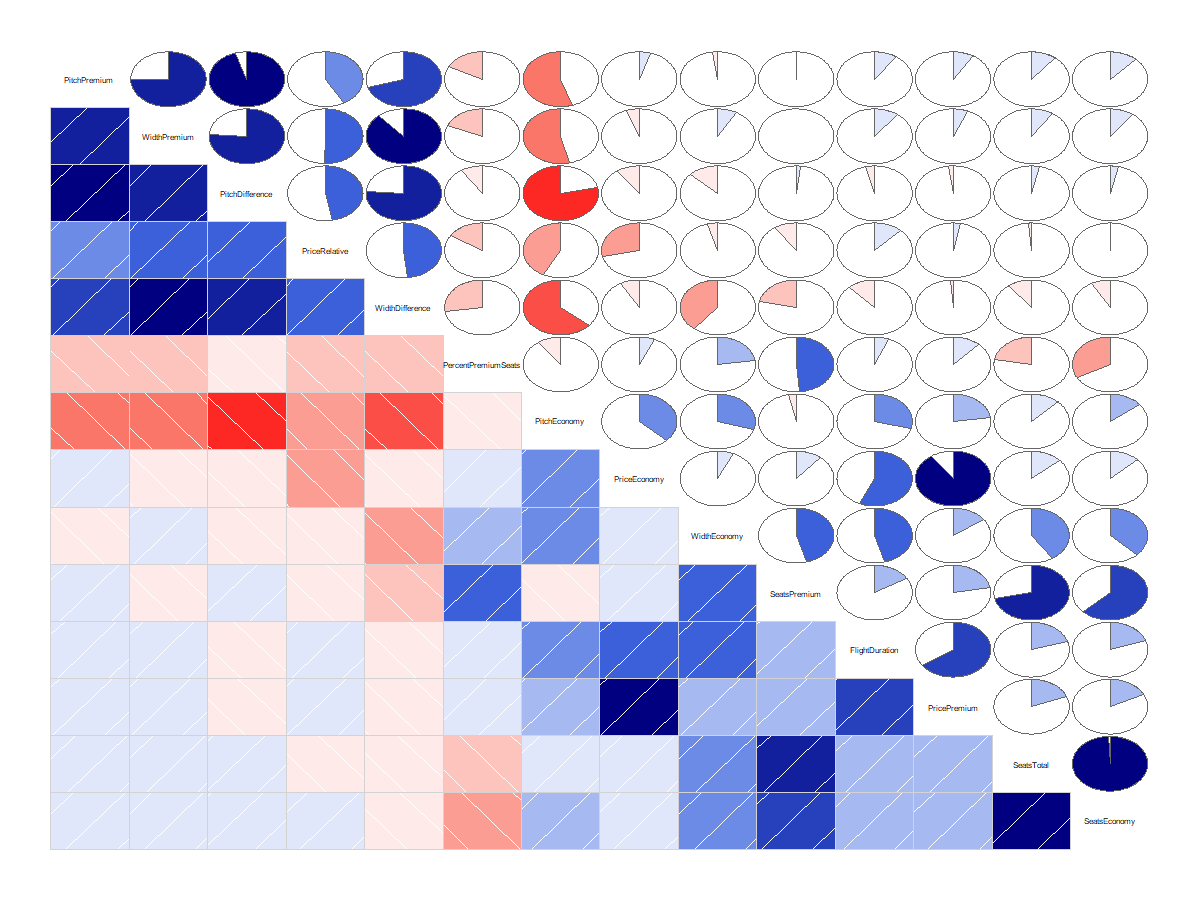
scatterplot(flight.df$PriceRelative,flight.df$SeatsEconomy , main="Price Relative And Economy Seats Plot" , xlab="Price Relative" , ylab="Seats Economy")



#CORRGRAM

> library(corrgram)

> corrgram(flight.df, order=TRUE, lower.panel=panel.shade,upper.panel=panel.pie, text.panel=panel.txt)



my\_data <- air.df[, c(3,6:18)]

res<-cor(my\_data)

round(res,2)

FlightDuration SeatsEconomy SeatsPremium PitchEconomy PitchPremium WidthEconomy WidthPremium PriceEconomy PricePremium PriceRelative SeatsTotal

FlightDuration 1.00 0.20 0.16 0.29 0.10 0.46 0.10 0.57 0.65 0.12 0.20

SeatsEconomy 0.20 1.00 0.63 0.14 0.12 0.37 0.10 0.13 0.18 0.00 0.99

SeatsPremium 0.16 0.63 1.00 -0.03 0.00 0.46 0.00 0.11 0.22 -0.10 0.72

PitchEconomy 0.29 0.14 -0.03 1.00 -0.55 0.29 -0.54 0.37 0.23 -0.42 0.12

PitchPremium 0.10 0.12 0.00 -0.55 1.00 -0.02 0.75 0.05 0.09 0.42 0.11

WidthEconomy 0.46 0.37 0.46 0.29 -0.02 1.00 0.08 0.07 0.15 -0.04 0.41

WidthPremium 0.10 0.10 0.00 -0.54 0.75 0.08 1.00 -0.06 0.06 0.50 0.09

PriceEconomy 0.57 0.13 0.11 0.37 0.05 0.07 -0.06 1.00 0.90 -0.29 0.13

PricePremium 0.65 0.18 0.22 0.23 0.09 0.15 0.06 0.90 1.00 0.03 0.19

PriceRelative 0.12 0.00 -0.10 -0.42 0.42 -0.04 0.50 -0.29 0.03 1.00 -0.01

SeatsTotal 0.20 0.99 0.72 0.12 0.11 0.41 0.09 0.13 0.19 -0.01 1.00

PitchDifference -0.04 0.04 0.02 -0.78 0.95 -0.13 0.76 -0.10 -0.02 0.47 0.03

WidthDifference -0.12 -0.08 -0.22 -0.64 0.70 -0.39 0.88 -0.08 -0.01 0.49 -0.11

PercentPremiumSeats 0.06 -0.33 0.49 -0.10 -0.18 0.23 -0.18 0.07 0.12 -0.16 -0.22

PitchDifference WidthDifference PercentPremiumSeats

FlightDuration -0.04 -0.12 0.06

SeatsEconomy 0.04 -0.08 -0.33

SeatsPremium 0.02 -0.22 0.49

PitchEconomy -0.78 -0.64 -0.10

PitchPremium 0.95 0.70 -0.18

WidthEconomy -0.13 -0.39 0.23

WidthPremium 0.76 0.88 -0.18

PriceEconomy -0.10 -0.08 0.07

PricePremium -0.02 -0.01 0.12

PriceRelative 0.47 0.49 -0.16

SeatsTotal 0.03 -0.11 -0.22

PitchDifference 1.00 0.76 -0.09

WidthDifference 0.76 1.00 -0.28

PercentPremiumSeats -0.09 -0.28 1.00

> cor(flight.df$PriceRelative,flight.df$PitchDifference)

[1] 0.4687302

cor.test(flight.df$PriceRelative,flight.df$PitchDifference)

p-value < 2.2e-16

cor(flight.df$PriceRelative,flight.df$WidthDifference)

[1] 0.4858024

> cor.test(flight.df$PriceRelative,flight.df$WidthDifference)

p-value < 2.2e-16

t.test(air.df$PitchDifference,air.df$PriceRelative )

Welch Two Sample t-test

data: air.df$PitchDifference and air.df$PriceRelative

t = 72.974, df = 516.54, p-value < 2.2e-16

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

95 percent confidence interval:

6.033640 6.367495

sample estimates:

mean of x mean of y

6.6877729 0.4872052

lm1<-lm(PriceRelative~PitchDifference+WidthDifference,data=air.df)

summary(lm1)

Call:

lm(formula = PriceRelative ~ PitchDifference + WidthDifference,

data = air.df)

Residuals:

Min 1Q Median 3Q Max

-0.84163 -0.28484 -0.07241 0.17698 1.18778

Coefficients:

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

(Intercept) -0.10514 0.08304 -1.266 0.206077

PitchDifference 0.06019 0.01590 3.785 0.000174 \*\*\*

WidthDifference 0.11621 0.02356 4.933 1.14e-06 \*\*\*

---

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

Residual standard error: 0.3886 on 455 degrees of freedom

Multiple R-squared: 0.2593, Adjusted R-squared: 0.2561

F-statistic: 79.65 on 2 and 455 DF, p-value: < 2.2e-16

coefficients(lm1)

(Intercept) PitchDifference WidthDifference

-0.10514235 0.06019158 0.11621441

t.test(air.df$PriceRelative,air.df$PercentPremiumSeats, paired=TRUE)

Paired t-test

data: air.df$PriceRelative and air.df$PercentPremiumSeats

t = -61.394, df = 457, p-value < 2.2e-16

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

95 percent confidence interval:

-14.61140 -13.70502

sample estimates:

mean of the differences

-14.15821

fit <- lm(PriceRelative ~ PercentPremiumSeats, data = air.df)

> summary(fit)

Call:

lm(formula = PriceRelative ~ PercentPremiumSeats, data = air.df)

Residuals:

Min 1Q Median 3Q Max

-0.55358 -0.38866 -0.08661 0.28655 1.35446

Coefficients:

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

(Intercept) 0.70738 0.06632 10.665 < 2e-16 \*\*\*

PercentPremiumSeats -0.01503 0.00430 -3.496 0.000518 \*\*\*

---

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

Residual standard error: 0.4452 on 456 degrees of freedom

Multiple R-squared: 0.0261, Adjusted R-squared: 0.02397

F-statistic: 12.22 on 1 and 456 DF, p-value: 0.0005185

> fit$coefficients

(Intercept) PercentPremiumSeats

0.70737829 -0.01503358

confint(fit)

2.5 % 97.5 %

(Intercept) 0.57703918 0.837717404

PercentPremiumSeats -0.02348423 -0.006582934