Airline Output

Purbasha Chatterjee

[pchatterjee10@gmail.com](mailto:pchatterjee10@gmail.com)

Oregon State University, Corvallis

Tue Jul 04 14:48:44 2017

setwd('C:/Users/purba/Desktop/R')  
  
airlines.df <- read.csv(paste("SixAirlines.csv", sep=""))  
summary(airlines.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

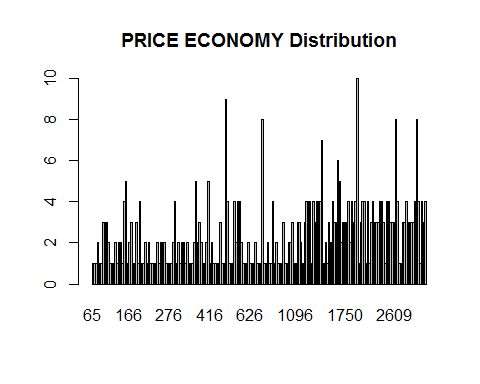
model <- lm(airlines.df$PRICE\_PREMIUM~., data = airlines.df)  
summary(model)

##   
## Call:  
## lm(formula = airlines.df$PRICE\_PREMIUM ~ ., data = airlines.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -801.60 -106.68 -18.84 67.47 1967.27   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -437.50415 3911.60438 -0.112 0.910995   
## AIRLINEBritish 756.74680 144.77762 5.227 2.66e-07 \*\*\*  
## AIRLINEDelta 574.04099 290.92430 1.973 0.049098 \*   
## AIRLINEJet 466.73977 126.26786 3.696 0.000246 \*\*\*  
## AIRLINESingapore 482.28161 169.98519 2.837 0.004760 \*\*   
## AIRLINEVirgin 726.35127 266.28478 2.728 0.006631 \*\*   
## AIRCRAFT 51.11682 38.75110 1.319 0.187815   
## FLIGHT\_DURATION -3.53822 6.49149 -0.545 0.585990   
## MONTH 0.01461 12.71716 0.001 0.999084   
## INTERNATIONAL -102.93062 331.75567 -0.310 0.756509   
## SEATS\_ECONOMY 0.11097 0.42158 0.263 0.792494   
## SEATS\_PREMIUM 1.79395 3.04952 0.588 0.556649   
## PITCH\_ECONOMY 34.81481 106.22255 0.328 0.743253   
## PITCH\_PREMIUM -77.50548 121.51922 -0.638 0.523931   
## WIDTH\_ECONOMY 55.94753 68.14550 0.821 0.412087   
## WIDTH\_PREMIUM 13.49400 181.26476 0.074 0.940691   
## PRICE\_ECONOMY 1.43128 0.02756 51.936 < 2e-16 \*\*\*  
## PRICE\_RELATIVE 1017.17737 40.45870 25.141 < 2e-16 \*\*\*  
## N NA NA NA NA   
## LAMBDA -340.41643 402.63156 -0.845 0.398300   
## QUALITY NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 282.2 on 443 degrees of freedom  
## Multiple R-squared: 0.954, Adjusted R-squared: 0.9521   
## F-statistic: 510.4 on 18 and 443 DF, p-value: < 2.2e-16

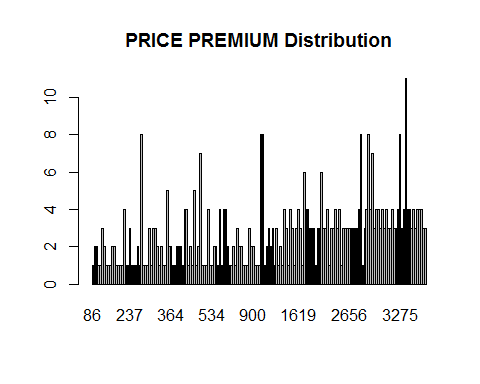
str(airlines.df)

## '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 ...

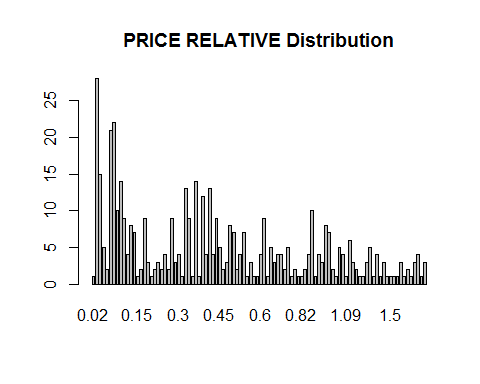
counts <- table(airlines.df$PRICE\_ECONOMY)  
barplot(counts, main = "PRICE ECONOMY Distribution")



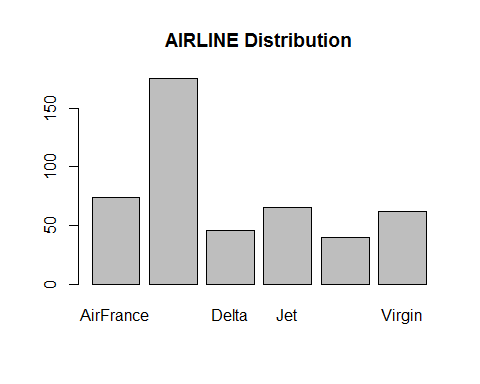
counts <- table(airlines.df$PRICE\_PREMIUM)  
barplot(counts, main = "PRICE PREMIUM Distribution")



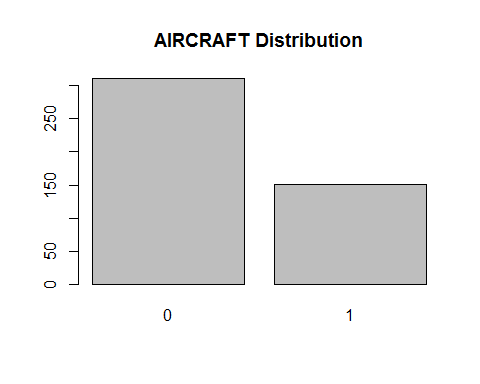
counts <- table(airlines.df$PRICE\_RELATIVE)  
barplot(counts, main = "PRICE RELATIVE Distribution")



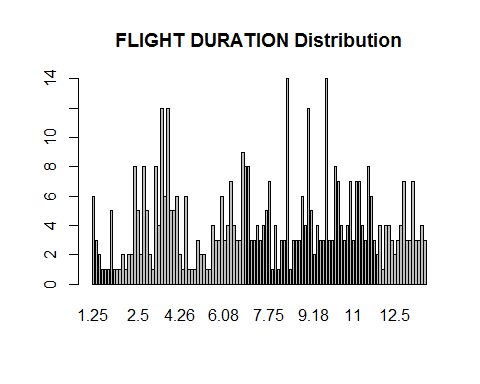
counts <- table(airlines.df$AIRLINE)  
barplot(counts, main = "AIRLINE Distribution")



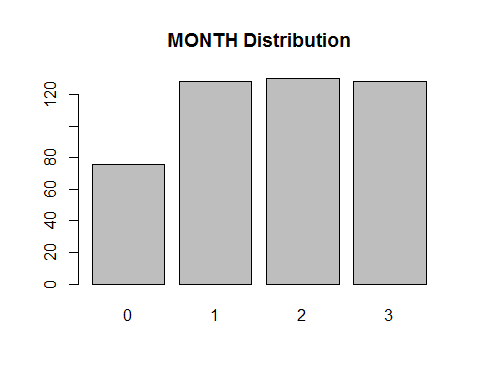
counts <- table(airlines.df$AIRCRAFT)  
barplot(counts, main = "AIRCRAFT Distribution")



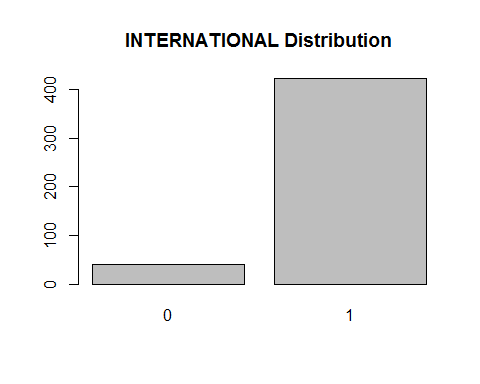
counts <- table(airlines.df$FLIGHT\_DURATION)  
barplot(counts, main = "FLIGHT DURATION Distribution")



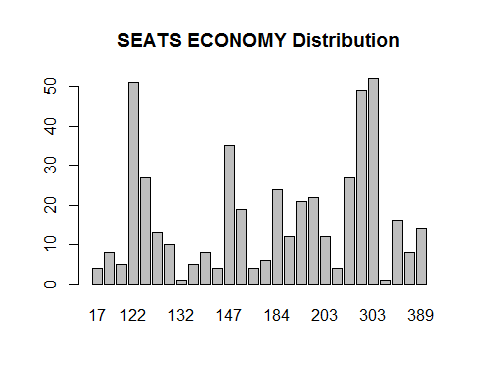
counts <- table(airlines.df$MONTH)  
barplot(counts, main = "MONTH Distribution")



counts <- table(airlines.df$INTERNATIONAL)  
barplot(counts, main = "INTERNATIONAL Distribution")



counts <- table(airlines.df$SEATS\_ECONOMY)  
barplot(counts, main = "SEATS ECONOMY Distribution")  
  
library(car)



par(mfrow=c(2, 2))  
with(airlines.df, scatterplot(airlines.df$AIRLINE , airlines.df$FLIGHT\_DURATION))

## [1] "6" "7" "74" "250" "251" "252" "253" "254" "255" "306" "327"  
## [12] "329" "350" "351" "352" "353"

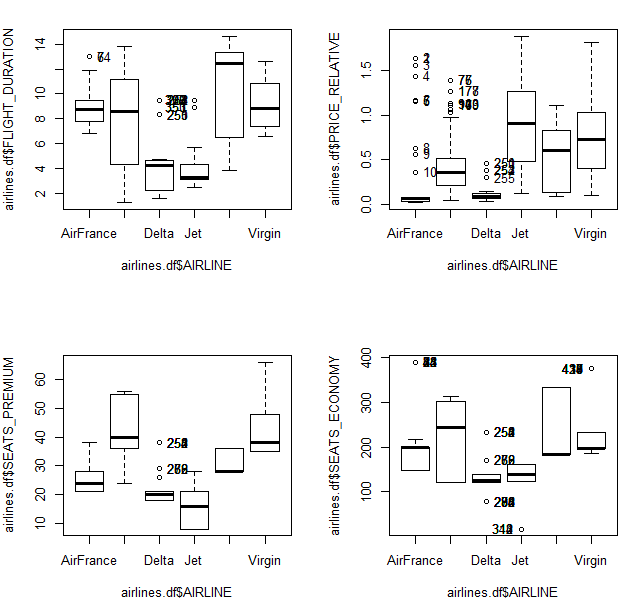
with(airlines.df, scatterplot(airlines.df$AIRLINE , airlines.df$PRICE\_RELATIVE))

## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "75"   
## [12] "76" "77" "177" "178" "90" "91" "143" "169" "170" "250" "251"  
## [23] "252" "253" "254" "255"

with(airlines.df, scatterplot(airlines.df$AIRLINE , airlines.df$SEATS\_PREMIUM))

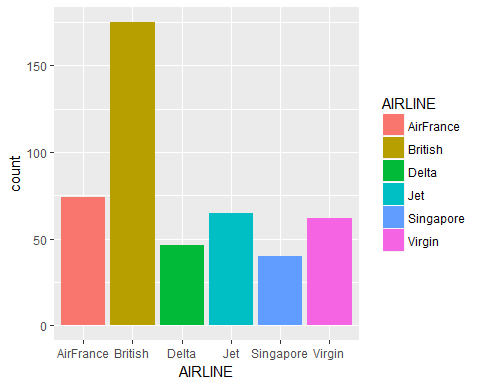
## [1] "250" "251" "252" "253" "254" "255" "258" "269" "272" "286"

with(airlines.df, scatterplot(airlines.df$AIRLINE , airlines.df$SEATS\_ECONOMY))

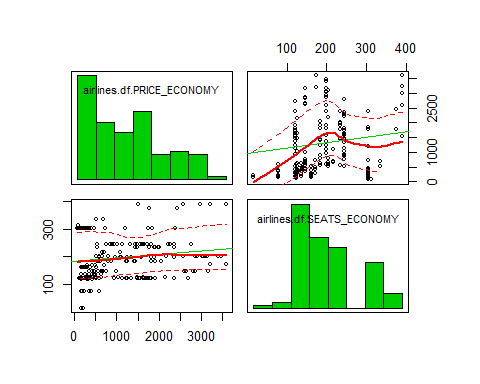


## [1] "6" "7" "24" "25" "33" "34" "35" "42" "43" "44" "276"  
## [12] "280" "283" "284" "285" "292" "294" "295" "250" "251" "252" "253"  
## [23] "254" "255" "258" "269" "272" "286" "312" "313" "314" "340" "417"  
## [34] "418" "419" "420" "434" "435" "436" "437"

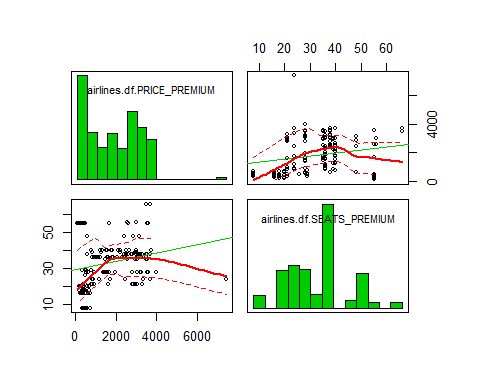
airlines.df$airname <- ifelse(airlines.df$AIRLINE=="AirFrance",1,  
 (ifelse(airlines.df$AIRLINE=="British",2,  
 (ifelse(airlines.df$AIRLINE=="Delta",3,  
 (ifelse(airlines.df$AIRLINE=="Jet",4,  
 (ifelse(airlines.df$AIRLINE=="Singapore",5,  
 (ifelse(airlines.df$AIRLINE=="Virgin",6, 0  
 ))  
 ))))  
 )))))  
  
library(ggplot2)  
ggplot(airlines.df, aes(x = AIRLINE, fill = AIRLINE)) + geom\_bar()



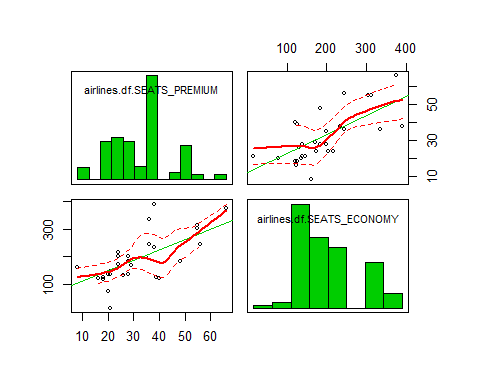
library(car)  
scatterplotMatrix(formula=~airlines.df$PRICE\_ECONOMY+airlines.df$SEATS\_ECONOMY,cex=0.6,diagonal="histogram")



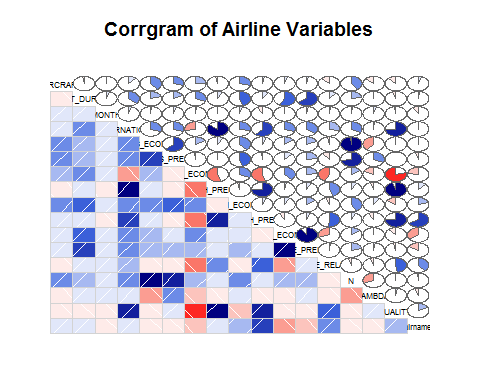
scatterplotMatrix(formula=~airlines.df$PRICE\_PREMIUM+airlines.df$SEATS\_PREMIUM,cex=0.6,diagonal="histogram")



scatterplotMatrix(formula=~airlines.df$SEATS\_PREMIUM+airlines.df$SEATS\_ECONOMY,cex=0.6,diagonal="histogram")



library(corrgram)  
corrgram(airlines.df, main="Corrgram of Airline Variables", lower.panel=panel.shade,  
 upper.panel=panel.pie,  
 text.panel=panel.txt)



cor.test(airlines.df$PITCH\_ECONOMY,airlines.df$PRICE\_ECONOMY)

##   
## Pearson's product-moment correlation  
##   
## data: airlines.df$PITCH\_ECONOMY and airlines.df$PRICE\_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

cor.test(airlines.df$PITCH\_PREMIUM,airlines.df$PRICE\_PREMIUM)

##   
## Pearson's product-moment correlation  
##   
## data: airlines.df$PITCH\_PREMIUM and airlines.df$PRICE\_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

t.test(airlines.df$airname, airlines.df$QUALITY)

##   
## Welch Two Sample t-test  
##   
## data: airlines.df$airname and airlines.df$QUALITY  
## t = -32.759, df = 916.62, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3.920746 -3.477522  
## sample estimates:  
## mean of x mean of y   
## 3.017316 6.716450

t.test(airlines.df$airname, airlines.df$PRICE\_ECONOMY)

##   
## Welch Two Sample t-test  
##   
## data: airlines.df$airname and airlines.df$PRICE\_ECONOMY  
## t = -28.535, df = 461, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1404.542 -1223.554  
## sample estimates:  
## mean of x mean of y   
## 3.017316 1317.064935

t.test(airlines.df$airname, airlines.df$PRICE\_PREMIUM)

##   
## Welch Two Sample t-test  
##   
## data: airlines.df$airname and airlines.df$PRICE\_PREMIUM  
## t = -30.481, df = 461, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1947.266 -1711.392  
## sample estimates:  
## mean of x mean of y   
## 3.017316 1832.346320

t.test(airlines.df$SEATS\_ECONOMY, airlines.df$PRICE\_ECONOMY)

##   
## Welch Two Sample t-test  
##   
## data: airlines.df$SEATS\_ECONOMY and airlines.df$PRICE\_ECONOMY  
## t = -24.167, df = 466.72, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1207.128 -1025.586  
## sample estimates:  
## mean of x mean of y   
## 200.7078 1317.0649

t.test(airlines.df$SEATS\_PREMIUM, airlines.df$PRICE\_PREMIUM)

##   
## Welch Two Sample t-test  
##   
## data: airlines.df$SEATS\_PREMIUM and airlines.df$PRICE\_PREMIUM  
## t = -29.971, df = 461.1, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1916.750 -1680.865  
## sample estimates:  
## mean of x mean of y   
## 33.53896 1832.34632

model1 <- lm(airlines.df$SEATS\_ECONOMY~airlines.df$WIDTH\_ECONOMY+airlines.df$PRICE\_ECONOMY+airlines.df$PITCH\_ECONOMY, data = airlines.df)  
summary(model1)

##   
## Call:  
## lm(formula = airlines.df$SEATS\_ECONOMY ~ airlines.df$WIDTH\_ECONOMY +   
## airlines.df$PRICE\_ECONOMY + airlines.df$PITCH\_ECONOMY, data = airlines.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -128.093 -74.623 -8.307 31.292 176.085   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -8.078e+02 1.773e+02 -4.556 6.68e-06 \*\*\*  
## airlines.df$WIDTH\_ECONOMY 5.247e+01 6.246e+00 8.400 5.64e-16 \*\*\*  
## airlines.df$PRICE\_ECONOMY 8.735e-03 3.634e-03 2.404 0.0166 \*   
## airlines.df$PITCH\_ECONOMY 1.968e+00 5.697e+00 0.345 0.7300   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 71.38 on 458 degrees of freedom  
## Multiple R-squared: 0.1671, Adjusted R-squared: 0.1616   
## F-statistic: 30.62 on 3 and 458 DF, p-value: < 2.2e-16

model2 <- lm(airlines.df$SEATS\_PREMIUM~airlines.df$WIDTH\_PREMIUM+airlines.df$PRICE\_PREMIUM+airlines.df$PITCH\_PREMIUM, data = airlines.df)  
summary(model2)

##   
## Call:  
## lm(formula = airlines.df$SEATS\_PREMIUM ~ airlines.df$WIDTH\_PREMIUM +   
## airlines.df$PRICE\_PREMIUM + airlines.df$PITCH\_PREMIUM, data = airlines.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22.6235 -10.5886 0.1343 5.5395 28.8563   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 37.6895528 17.6932568 2.130 0.0337 \*   
## airlines.df$WIDTH\_PREMIUM -0.1921873 0.8350233 -0.230 0.8181   
## airlines.df$PRICE\_PREMIUM 0.0023292 0.0004689 4.967 9.59e-07 \*\*\*  
## airlines.df$PITCH\_PREMIUM -0.1232385 0.6964258 -0.177 0.8596   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.95 on 458 degrees of freedom  
## Multiple R-squared: 0.05132, Adjusted R-squared: 0.04511   
## F-statistic: 8.259 on 3 and 458 DF, p-value: 2.316e-05