HW1

Step 1, 2 : Reading the two data files, checking the structure of the data.

```
library(MASS)
library(readxl)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(schoolmath)
library(tinytex)
#step1

FAA1_1 <- read_excel("FAA1-1.xls")
FAA2_1 <- read_excel("FAA2-1.xls")
dim(FAA1_1)

## [1] 800   8

#Step 2
str(FAA1_1)

## Classes 'tbl_df', 'tbl' and 'data.frame':    800 obs. of  8 variables:
##  $ aircraft    : chr  "boeing" "boeing" "boeing" "boeing" ...
##  $ duration    : num  98.5 125.7 112 196.8 90.1 ...
##  $ no_pasg     : num  53 69 61 56 70 55 54 57 61 56 ...
##  $ speed_ground: num  107.9 101.7 71.1 85.8 59.9 ...
##  $ speed_air   : num  109 103 NA NA NA ...
##  $ height      : num  27.4 27.8 18.6 30.7 32.4 ...
##  $ pitch       : num  4.04 4.12 4.43 3.88 4.03 ...
##  $ distance    : num  3370 2988 1145 1664 1050 ...

str(FAA2_1)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    150 obs. of  7 variables:
##  $ aircraft    : chr  "boeing" "boeing" "boeing" "boeing" ...
##  $ no_pasg     : num  53 69 61 56 70 55 54 57 61 56 ...
##  $ speed_ground: num  107.9 101.7 71.1 85.8 59.9 ...
##  $ speed_air   : num  109 103 NA NA NA ...
##  $ height      : num  27.4 27.8 18.6 30.7 32.4 ...
##  $ pitch       : num  4.04 4.12 4.43 3.88 4.03 ...
##  $ distance    : num  3370 2988 1145 1664 1050 ...
```

Step 3: Merging the two files together

In addition, there are no duplicates to be found in my initial analysis of the data. Had I found any duplicates, I would have removed them from my analysis as duplication might have skewed my results in this scenario. Also, checked for any NA values present in the data.

```
#Step 3
FAA <- merge(FAA1_1,FAA2_1,all.x=TRUE)
dim(FAA)

## [1] 800   8

#Removing Duplicate rows
FAA_uniq <- FAA %>% distinct(distance,.keep_all = TRUE)
dim(FAA_uniq)

## [1] 800   8

#Hence No duplicates

#NA values
colSums(is.na(FAA_uniq))

##     aircraft      no_pasg speed_ground    speed_air       height 
##            0            0            0          600            0 
##        pitch     distance     duration 
##            0            0            0
```

Step 4:Checking the structure and inital exploratory analyis of the dataset.

```
#Step 4
str(FAA_uniq)

## 'data.frame':    800 obs. of  8 variables:
##  $ aircraft    : chr  "airbus" "airbus" "airbus" "airbus" ...
##  $ no_pasg     : num  36 38 40 41 43 44 45 45 45 45 ...
##  $ speed_ground: num  47.5 85.2 80.6 97.6 82.5 ...
##  $ speed_air   : num  NA NA NA 97 NA ...
##  $ height      : num  14 37 28.6 38.4 30.1 ...
##  $ pitch       : num  4.3 4.12 3.62 3.53 4.09 ...
```

```
##  $ distance    : num  251 1257 1021 2168 1321 ...
##  $ duration    : num  172 188 93.5 123.3 109.2 ...
```

```r
summary(FAA_uniq)
```

```
##    aircraft            no_pasg       speed_ground      speed_air     
##  Length:800         Min.   :29.00   Min.   : 27.74   Min.   : 90.00  
##  Class :character   1st Qu.:55.00   1st Qu.: 65.87   1st Qu.: 96.16  
##  Mode  :character   Median :60.00   Median : 79.64   Median :100.99  
##                     Mean   :60.13   Mean   : 79.54   Mean   :103.83  
##                     3rd Qu.:65.00   3rd Qu.: 92.33   3rd Qu.:109.48  
##                     Max.   :87.00   Max.   :141.22   Max.   :141.72  
##                                                      NA's   :600     
##      height           pitch          distance          duration     
##  Min.   :-3.546   Min.   :2.284   Min.   :  34.08   Min.   : 14.76  
##  1st Qu.:23.338   1st Qu.:3.658   1st Qu.: 900.95   1st Qu.:119.49  
##  Median :30.147   Median :4.020   Median :1267.44   Median :153.95  
##  Mean   :30.122   Mean   :4.018   Mean   :1544.52   Mean   :154.01  
##  3rd Qu.:36.981   3rd Qu.:4.388   3rd Qu.:1960.44   3rd Qu.:188.91  
##  Max.   :59.946   Max.   :5.927   Max.   :6533.05   Max.   :305.62  
## 
```

Step 5:Preliminary Observations of the dataset

- Distance has a huge difference in the median and mean values meaning that the data is left skewed.
- Speeed_Air has 642 missing values
- It can be seen from the histogram that speed_air values are truncated (values below ~90 do not exist)
- Airbus and Boeing both consists of 400 records
- Some abnormal values rpesent in some of the columns which woould require to be cleaned.

```r
sum(FAA_uniq$aircraft=="airbus")
```

```
## [1] 400
```

```r
sum(FAA_uniq$aircraft=="boeing")
```

```
## [1] 400
```

Data Cleaning and further exploartion

Step 6

- Removed 19 observations based on abnormal value data definitions

```r
#Step 6
FAA_uniq <- FAA_uniq[!(FAA_uniq$duration <= 40), ]
FAA_uniq <- FAA_uniq[!(FAA_uniq$speed_ground < 30 | FAA_uniq$speed_ground > 
140), ]
```

```r
FAA_uniq <- FAA_uniq[!(FAA_uniq$height < 6), ]
FAA_uniq <- FAA_uniq[!(FAA_uniq$distance > 6000), ]
```

Step 7

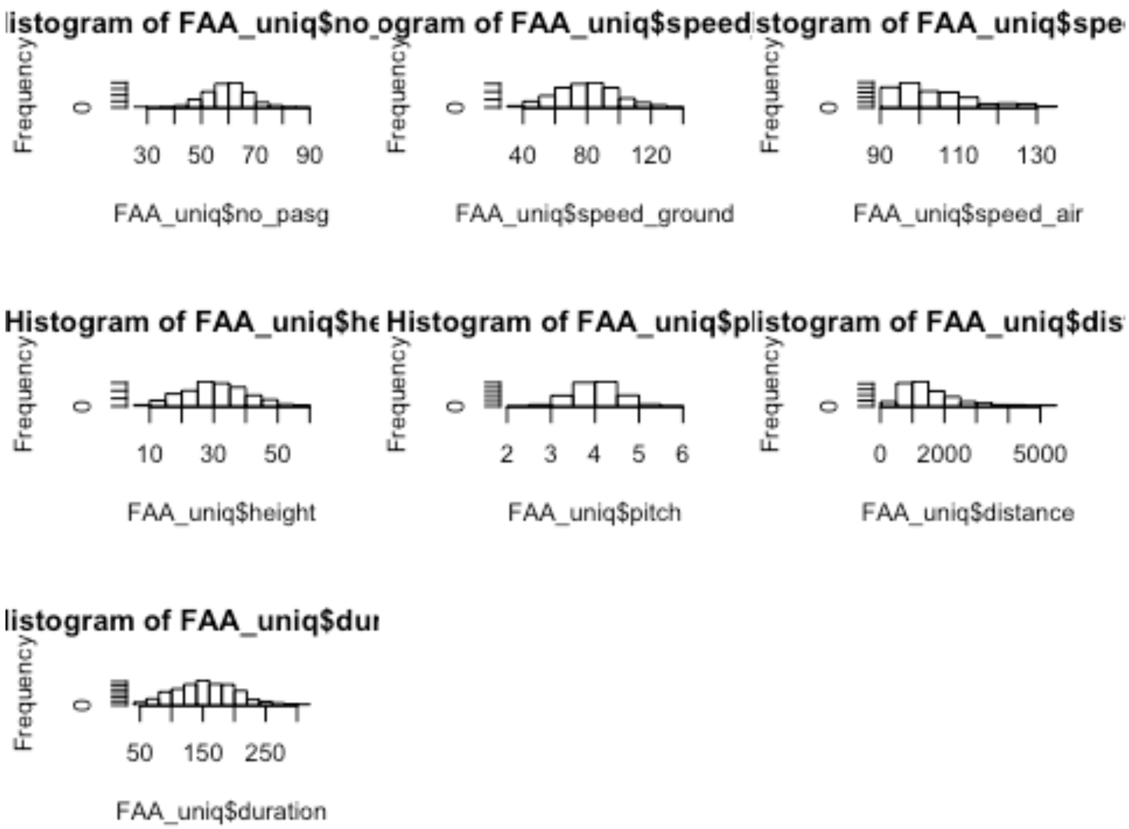
Redid the structure and calculated if the number of aircraft rows for each of the aircrafts is significant enough for analysis.

```r
str(FAA_uniq)

## 'data.frame':    781 obs. of  8 variables:
##  $ aircraft    : chr  "airbus" "airbus" "airbus" "airbus" ...
##  $ no_pasg     : num  36 38 40 41 43 44 45 45 45 45 ...
##  $ speed_ground: num  47.5 85.2 80.6 97.6 82.5 ...
##  $ speed_air   : num  NA NA NA 97 NA ...
##  $ height      : num  14 37 28.6 38.4 30.1 ...
##  $ pitch       : num  4.3 4.12 3.62 3.53 4.09 ...
##  $ distance    : num  251 1257 1021 2168 1321 ...
##  $ duration    : num  172 188 93.5 123.3 109.2 ...

summary(FAA_uniq)

##    aircraft            no_pasg       speed_ground      speed_air     
##  Length:781         Min.   :29.00   Min.   : 33.57   Min.   : 90.00  
##  Class :character   1st Qu.:55.00   1st Qu.: 66.19   1st Qu.: 96.15  
##  Mode  :character   Median :60.00   Median : 79.79   Median :100.89  
##                     Mean   :60.08   Mean   : 79.64   Mean   :103.50  
##                     3rd Qu.:65.00   3rd Qu.: 92.13   3rd Qu.:109.42  
##                     Max.   :87.00   Max.   :132.78   Max.   :132.91  
##                                                      NA's   :586     
##      height           pitch          distance          duration     
##  Min.   : 6.228   Min.   :2.284   Min.   :  41.72   Min.   : 41.95  
##  1st Qu.:23.594   1st Qu.:3.653   1st Qu.: 919.05   1st Qu.:119.63  
##  Median :30.217   Median :4.014   Median :1273.66   Median :154.28  
##  Mean   :30.455   Mean   :4.014   Mean   :1541.20   Mean   :154.78  
##  3rd Qu.:36.988   3rd Qu.:4.382   3rd Qu.:1960.43   3rd Qu.:189.66  
##  Max.   :59.946   Max.   :5.927   Max.   :5381.96   Max.   :305.62  
## 

sum(FAA_uniq$aircraft=="airbus")

## [1] 394

sum(FAA_uniq$aircraft=="boeing")

## [1] 387
```

Step 8: Plotting of the variables

```r
#Step 8
par(mfrow = c(3, 3))
```

```r
hist(FAA_uniq$no_pasg)
hist(FAA_uniq$speed_ground)
hist(FAA_uniq$speed_air)
hist(FAA_uniq$height)
hist(FAA_uniq$pitch)
hist(FAA_uniq$distance)
hist(FAA_uniq$duration)
```



Step 9

- Data reduced to 781 observations after removing abnormal values
- Speed_air still has 586 rows having no value
- Airbus and Boeing records are 394 and 387 respectively after cleaning

Initial analysis for identifying important factors that impact the response variable “landing distance”

Step 10: Pairwise correlation Table:Table 1

```r
#Step 10
FAA_uniq$aircraft <- ifelse(FAA_uniq$aircraft == "airbus", 0, 1)
```

```
cor_faa <- cor(FAA_uniq)
round(cor_faa, 2)
```

```
##              aircraft no_pasg speed_ground speed_air height pitch distance
## aircraft         1.00   -0.03        -0.05        NA  -0.02  0.36     0.23
## no_pasg         -0.03    1.00         0.00        NA   0.04 -0.01    -0.02
## speed_ground    -0.05    0.00         1.00        NA  -0.05 -0.05     0.87
## speed_air          NA      NA           NA         1     NA    NA       NA
## height          -0.02    0.04        -0.05        NA   1.00  0.03     0.10
## pitch            0.36   -0.01        -0.05        NA   0.03  1.00     0.07
## distance         0.23   -0.02         0.87        NA   0.10  0.07     1.00
## duration        -0.04   -0.04        -0.05        NA   0.01 -0.05    -0.05
##              duration
## aircraft        -0.04
## no_pasg         -0.04
## speed_ground    -0.05
## speed_air          NA
## height           0.01
## pitch           -0.05
## distance        -0.05
## duration         1.00
```

```
tab <- cor_faa[,7,drop=FALSE]
Table1 <- data.frame("Name of the Variable" =
c("aircraft","no_pasg","speed_ground","speed_air","height","pitch","Distance"
,"duration"),
                     "Corr_Coefficient"=c(tab[1:8]),
                     "Direction of corr_coefficient" =
c(ifelse(is.positive(tab[1])=="TRUE","Positive","Negative"),

ifelse(is.positive(tab[2])=="TRUE","Positive","Negative"),

ifelse(is.positive(tab[3])=="TRUE","Positive","Negative"),

ifelse(tab[4]=="","NA","NA"),

ifelse(is.positive(tab[5])=="TRUE","Positive","Negative"),

ifelse(is.positive(tab[6])=="TRUE","Positive","Negative"),

ifelse(is.positive(tab[7])=="TRUE","Positive","Negative"),

ifelse(is.positive(tab[8])=="TRUE","Positive","Negative")))
Table1 <- Table1[order(-abs(Table1$Corr_Coefficient)),]
Table1
```

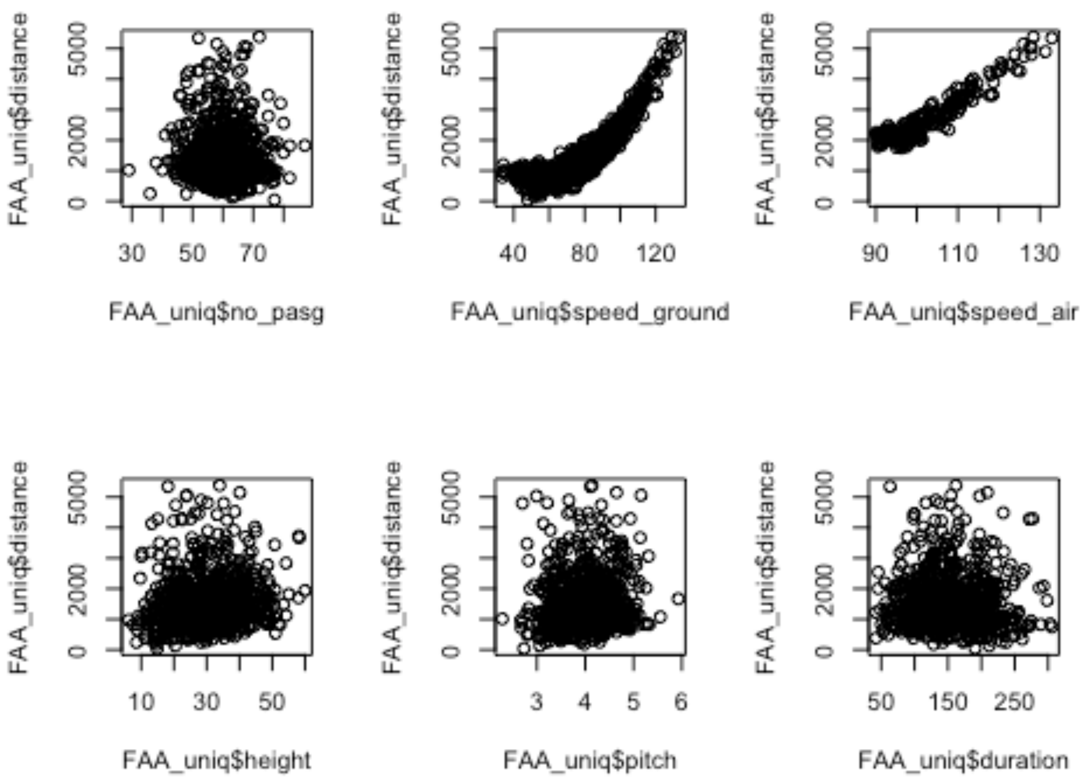
```
##   Name.of.the.Variable Corr_Coefficient Direction.of.corr_coefficient
## 7             Distance       1.00000000                      Positive
## 3         speed_ground       0.86771145                      Positive
## 1             aircraft       0.22998335                      Positive
```

```
## 5          height   0.10372080        Positive
## 6           pitch   0.06868102        Positive
## 8        duration  -0.05138252        Negative
## 2         no_pasg  -0.01685312        Negative
## 4       speed_air          NA            <NA>
```

Step 11

As per the plots the correlation strength observed is consistent with the respective values.

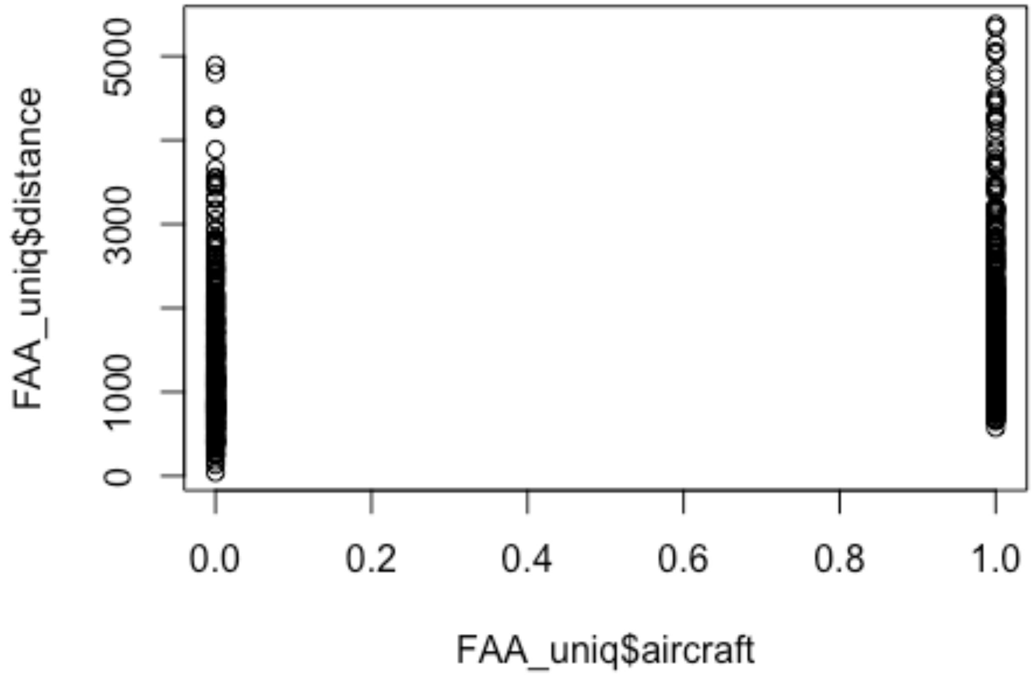
```r
#Step 11
par(mfrow = c(2, 3))
plot(FAA_uniq$distance ~ FAA_uniq$no_pasg)
plot(FAA_uniq$distance ~ FAA_uniq$speed_ground)
plot(FAA_uniq$distance ~ FAA_uniq$speed_air)
plot(FAA_uniq$distance ~ FAA_uniq$height)
plot(FAA_uniq$distance ~ FAA_uniq$pitch)
plot(FAA_uniq$distance ~ FAA_uniq$duration)
```



Step 12

Already performed before so as to have a consistency in the table 1 and the others.

```r
plot(FAA_uniq$distance ~ FAA_uniq$aircraft)
```



Regression using a single factor each time

Step 13: Regression using each variable independently and creating Table2 with the observations

```r
model1 <- lm(FAA_uniq$distance ~ FAA_uniq$aircraft)
model2 <- lm(FAA_uniq$distance ~ FAA_uniq$no_pasg)
model3 <- lm(FAA_uniq$distance ~ FAA_uniq$speed_ground)
model4 <- lm(FAA_uniq$distance ~ FAA_uniq$speed_air)
model5 <- lm(FAA_uniq$distance ~ FAA_uniq$height)
model6 <- lm(FAA_uniq$distance ~ FAA_uniq$pitch)
model7 <- lm(FAA_uniq$distance ~ FAA_uniq$duration)

Table2 <- data.frame("Name of the Variable" = 
c("aircraft","no_pasg","speed_ground","speed_air","height","pitch","duration"
), "p-value" = 

c(summary(model1)$coefficients[2,4],summary(model2)$coefficients[2,4],summary
(model3)$coefficients[2,4],
```

```
summary(model4)$coefficients[2,4],summary(model5)$coefficients[2,4],summary(m
odel6)$coefficients[2,4],summary(model7)$coefficients[2,4]),
                          "Direction of coefficient" =
c(ifelse(is.positive(summary(model1)$coefficients[2,1])=="TRUE","Positive","N
egative"),

ifelse(is.positive(summary(model2)$coefficients[2,1])=="TRUE","Positive","Neg
ative"),

ifelse(is.positive(summary(model3)$coefficients[2,1])=="TRUE","Positive","Neg
ative"),

ifelse(is.positive(summary(model4)$coefficients[2,1])=="TRUE","Positive","Neg
ative"),

ifelse(is.positive(summary(model5)$coefficients[2,1])=="TRUE","Positive","Neg
ative"),

ifelse(is.positive(summary(model6)$coefficients[2,1])=="TRUE","Positive","Neg
ative"),

ifelse(is.positive(summary(model7)$coefficients[2,1])=="TRUE","Positive","Neg
ative")))
Table2

##   Name.of.the.Variable       p.value Direction.of.coefficient
## 1             aircraft  7.806173e-11                 Positive
## 2              no_pasg  6.381673e-01                 Negative
## 3         speed_ground 1.060591e-238                 Positive
## 4            speed_air  2.550801e-94                 Positive
## 5               height  3.710162e-03                 Positive
## 6                pitch  5.503786e-02                 Positive
## 7             duration  1.514002e-01                 Negative
```

Step 14: Standardizing the independent variables and creating Table3

```
FAA_stan <- FAA_uniq
FAA_stan$aircraftsd <- (FAA_uniq$aircraft-
mean(FAA_uniq$aircraft))/sd(FAA_uniq$aircraft)
FAA_stan$no_pasgsd <- (FAA_uniq$no_pasg-
mean(FAA_uniq$no_pasg))/sd(FAA_uniq$no_pasg)
FAA_stan$speed_groundsd <- (FAA_uniq$speed_ground-
mean(FAA_uniq$speed_ground))/sd(FAA_uniq$speed_ground)
FAA_stan$heightsd <- (FAA_uniq$height-
mean(FAA_uniq$height))/sd(FAA_uniq$height)
FAA_stan$pitchsd <- (FAA_uniq$pitch-mean(FAA_uniq$pitch))/sd(FAA_uniq$pitch)
FAA_stan$durationsd <- (FAA_uniq$duration-
mean(FAA_uniq$duration))/sd(FAA_uniq$duration)

model11 <- lm(FAA_stan$distance ~ FAA_stan$aircraftsd)
model21 <- lm(FAA_stan$distance ~ FAA_stan$no_pasgsd)
```

```
model31 <- lm(FAA_stan$distance ~ FAA_stan$speed_groundsd)
model41 <- lm(FAA_stan$distance ~ FAA_stan$speed_air)
model51 <- lm(FAA_stan$distance ~ FAA_stan$heightsd)
model61 <- lm(FAA_stan$distance ~ FAA_stan$pitchsd)
model71 <- lm(FAA_stan$distance ~ FAA_stan$durationsd)

Table3 <- data.frame("Name of the Variable" =
c("aircraft","no_pasg","speed_ground","speed_air","height","pitch","duration"
), "Standardized_coeff_value" =

c(abs(summary(model11)$coefficients[2,1]),abs(summary(model21)$coefficients[2
,1]),abs(summary(model31)$coefficients[2,1]),

abs(summary(model41)$coefficients[2,1]),abs(summary(model51)$coefficients[2,1
]),abs(summary(model61)$coefficients[2,1]),abs(summary(model71)$coefficients[
2,1])),
                     "Standardized_coeff_direction" =
c(ifelse(is.positive(summary(model11)$coefficients[2,1])=="TRUE","Positive","
Negative"),

ifelse(is.positive(summary(model21)$coefficients[2,1])=="TRUE","Positive","Ne
gative"),

ifelse(is.positive(summary(model31)$coefficients[2,1])=="TRUE","Positive","Ne
gative"),

ifelse(is.positive(summary(model41)$coefficients[2,1])=="TRUE","Positive","Ne
gative"),

ifelse(is.positive(summary(model51)$coefficients[2,1])=="TRUE","Positive","Ne
gative"),

ifelse(is.positive(summary(model61)$coefficients[2,1])=="TRUE","Positive","Ne
gative"),

ifelse(is.positive(summary(model71)$coefficients[2,1])=="TRUE","Positive","Ne
gative")))


Table3 <- Table3[order(-abs(Table3$Standardized_coeff_value)),]
Table3

##   Name.of.the.Variable Standardized_coeff_value
## 3         speed_ground                784.92339
## 1             aircraft                208.04071
## 5               height                 93.82483
## 4            speed_air                 79.24368
## 6                pitch                 62.12818
## 7             duration                 46.48013
## 2              no_pasg                 15.24517
```

```
##   Standardized_coeff_direction
## 3                     Positive
## 1                     Positive
## 5                     Positive
## 4                     Positive
## 6                     Positive
## 7                     Negative
## 2                     Negative
```

Step 15:Comparing the three tables and creating a common Table0

```
Table0 <- Reduce(function(x,y) merge(x = x, y = y),
       list(Table1, Table2, Table3))

Table0 <- Table0[order(-abs(Table0$Corr_Coefficient),Table0$p.value,-
Table0$Standardized_coeff_value),]
Table0

##   Name.of.the.Variable Corr_Coefficient Direction.of.corr_coefficient
## 7         speed_ground       0.86771145                      Positive
## 1             aircraft       0.22998335                      Positive
## 3               height       0.10372080                      Positive
## 5                pitch       0.06868102                      Positive
## 2             duration      -0.05138252                      Negative
## 4              no_pasg      -0.01685312                      Negative
## 6            speed_air               NA                          <NA>
##          p.value Direction.of.coefficient Standardized_coeff_value
## 7 1.060591e-238                 Positive                784.92339
## 1  7.806173e-11                 Positive                208.04071
## 3  3.710162e-03                 Positive                 93.82483
## 5  5.503786e-02                 Positive                 62.12818
## 2  1.514002e-01                 Negative                 46.48013
## 4  6.381673e-01                 Negative                 15.24517
## 6  2.550801e-94                 Positive                 79.24368
##   Standardized_coeff_direction
## 7                     Positive
## 1                     Positive
## 3                     Positive
## 5                     Positive
## 2                     Negative
## 4                     Negative
## 6                     Positive
```

Step 16: Compare the regression coefficients of the three models below:

- Model.sg: LD ~ Speed_ground
- Model.sa: LD ~ Speed_air
- Model.comb: LD ~ Speed_ground + Speed_air

- Correlation between Speed_Air and Speed_Ground is 98.83 and hence keeping both the variables in our model would have a negative impact on our model as there would be interaction.
- I would keep **Speed_ground** as there are more data points present for this variable.

```
#Step 16
model.sg <- lm(FAA_uniq$distance ~ FAA_uniq$speed_ground)
model.sa <- lm(FAA_uniq$distance ~ FAA_uniq$speed_air)
model.comb <- lm(FAA_uniq$distance ~
FAA_uniq$speed_ground+FAA_uniq$speed_air)

FAA_ns <- na.omit(FAA_uniq)
cor(FAA_ns)

##                  aircraft       no_pasg speed_ground     speed_air
## aircraft       1.00000000 -5.594062e-02 -0.079619768 -7.710399e-02
## no_pasg       -0.05594062  1.000000e+00  0.003570599  2.242971e-05
## speed_ground  -0.07961977  3.570599e-03  1.000000000  9.883475e-01
## speed_air     -0.07710399  2.242971e-05  0.988347471  1.000000e+00
## height        -0.05299014 -6.625455e-03 -0.095483596 -8.672929e-02
## pitch          0.36550660 -3.766471e-02 -0.063161271 -4.826810e-02
## distance       0.17251928 -3.258255e-02  0.928771947  9.432190e-01
## duration      -0.04798222 -6.917843e-02  0.023885892  4.454351e-02
##                    height       pitch    distance    duration
## aircraft      -0.052990141  0.36550660  0.17251928 -0.04798222
## no_pasg       -0.006625455 -0.03766471 -0.03258255 -0.06917843
## speed_ground  -0.095483596 -0.06316127  0.92877195  0.02388589
## speed_air     -0.086729286 -0.04826810  0.94321897  0.04454351
## height         1.000000000 -0.03321763  0.05775639  0.07377549
## pitch         -0.033217630  1.00000000  0.03402263 -0.05627519
## distance       0.057756386  0.03402263  1.00000000  0.05241698
## duration       0.073775491 -0.05627519  0.05241698  1.00000000
```

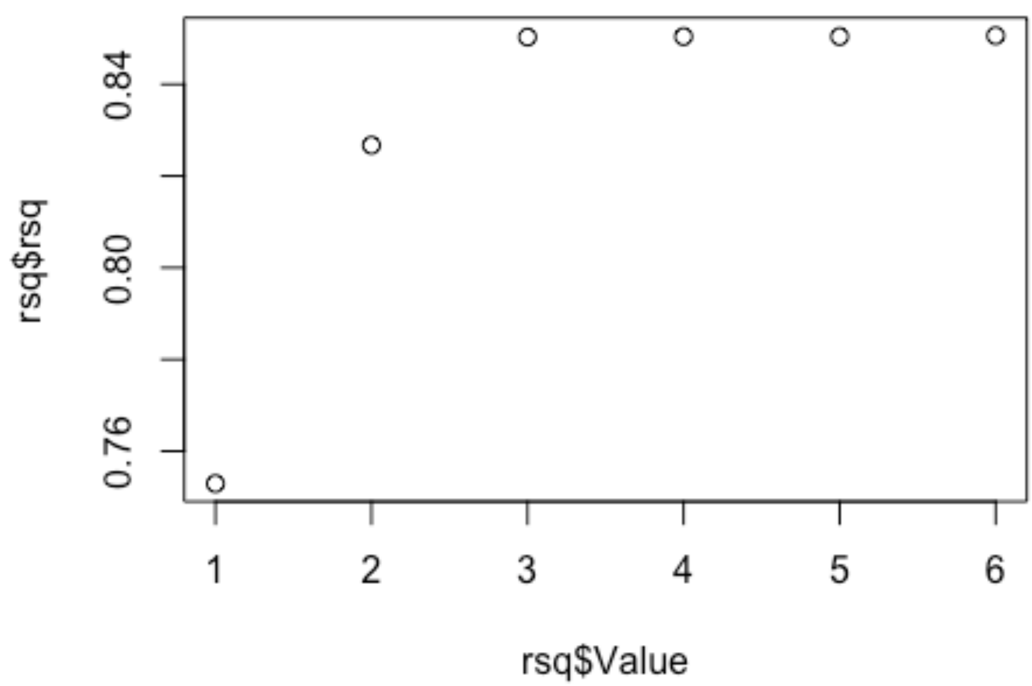
Variable selection based on our ranking in Table 0

Step 17

*We observe that r-squared values are increasing with increase in the number of variables, however the increase is insignificant after number of variables increase to 4 and above.

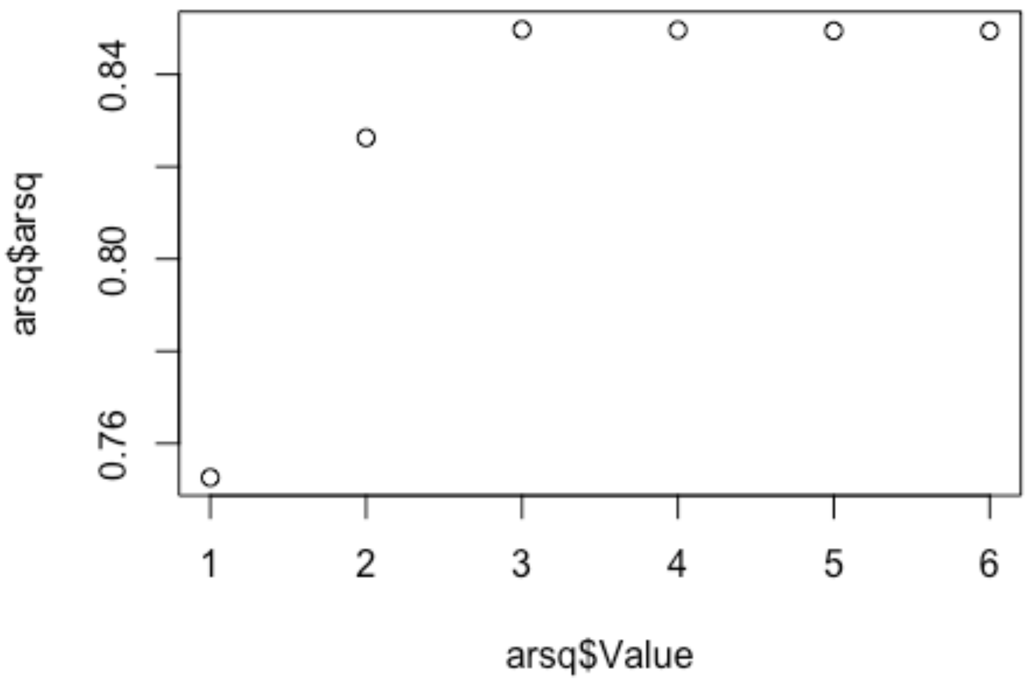
```
model.x1 <- lm(FAA_uniq$distance ~ FAA_uniq$speed_ground)
model.x2 <- lm(FAA_uniq$distance ~ FAA_uniq$speed_ground+FAA_uniq$aircraft)
model.x3 <- lm(FAA_uniq$distance ~
FAA_uniq$speed_ground+FAA_uniq$aircraft+FAA_uniq$height)
model.x4 <- lm(FAA_uniq$distance ~
FAA_uniq$speed_ground+FAA_uniq$aircraft+FAA_uniq$height+FAA_uniq$pitch)
model.x5 <- lm(FAA_uniq$distance ~
FAA_uniq$speed_ground+FAA_uniq$aircraft+FAA_uniq$height+FAA_uniq$pitch+FAA_un
```

```
iq$duration)
model.x6 <- lm(FAA_uniq$distance ~
FAA_uniq$speed_ground+FAA_uniq$aircraft+FAA_uniq$height+FAA_uniq$pitch+FAA_un
iq$duration+FAA_uniq$no_pasg)
rsq <-
data.frame("Value"=1:6,"rsq"=c(summary(model.x1)$r.squared,summary(model.x2)$
r.squared,summary(model.x3)$r.squared,

summary(model.x4)$r.squared,summary(model.x5)$r.squared,summary(model.x6)$r.s
quared))
plot(rsq$Value,rsq$rsq)
```



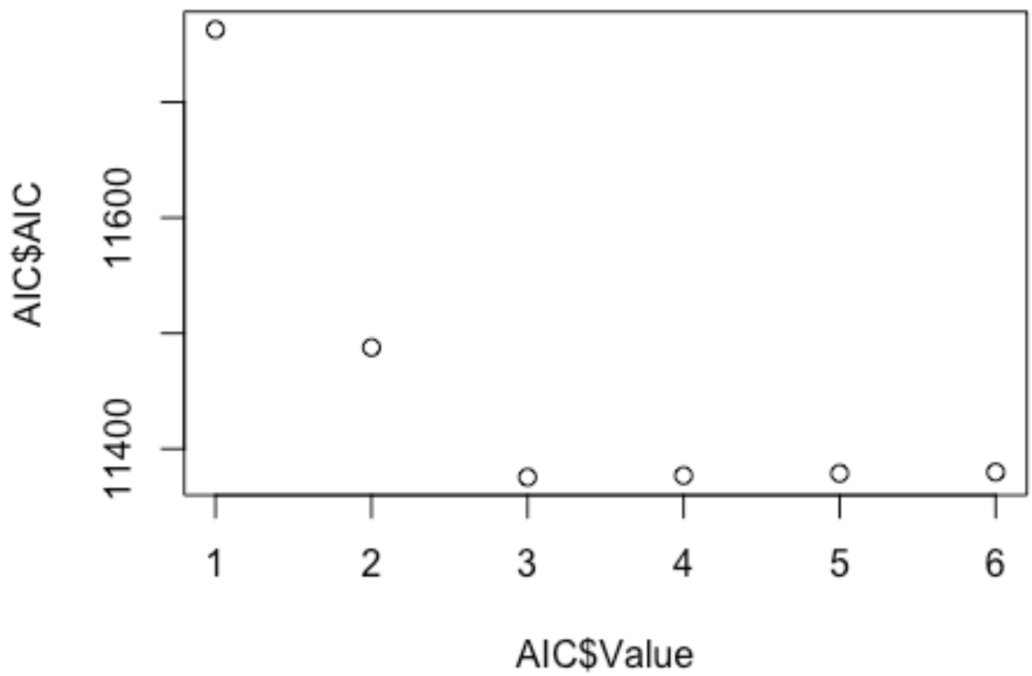
Step 18

```
arsq <-
data.frame("Value"=1:6,"arsq"=c(summary(model.x1)$adj.r.squared,summary(model
.x2)$adj.r.squared,summary(model.x3)$adj.r.squared,

summary(model.x4)$adj.r.squared,summary(model.x5)$adj.r.squared,summary(model
.x6)$adj.r.squared))
plot(arsq$Value,arsq$arsq)
```



Step 19

```
AIC <-
data.frame("Value"=1:6,"AIC"=c(AIC(model.x1),AIC(model.x2),AIC(model.x3),

AIC(model.x4),AIC(model.x5),AIC(model.x6)))
plot(AIC$Value,AIC$AIC)
```



Step20

- I will use Speed_ground and aircraft as adding more variables will be over fitting the model and the adjusted rsquare and AIC score does not increase significantly after the addition of the second variable.

Step 21

```
Model_LM <- lm(FAA_uniq$distance
~FAA_uniq$speed_ground+FAA_uniq$aircraft+FAA_uniq$height+FAA_uniq$pitch+FAA_u
niq$duration+FAA_uniq$no_pasg ,data = FAA_uniq)
fit1_LM <- stepAIC(Model_LM, direction = 'backward',trace=TRUE)

## Start:  AIC=9161.49
## FAA_uniq$distance ~ FAA_uniq$speed_ground + FAA_uniq$aircraft +
##     FAA_uniq$height + FAA_uniq$pitch + FAA_uniq$duration +
FAA_uniq$no_pasg
##
##                          Df Sum of Sq       RSS    AIC
## - FAA_uniq$duration       1      3958  95358717 9159.5
## - FAA_uniq$pitch          1     71084  95425843 9160.1
## - FAA_uniq$no_pasg        1    117353  95472112 9160.5
## <none>                               95354759 9161.5
## - FAA_uniq$height         1  15009018 110363777 9273.7
```

```
## - FAA_uniq$aircraft      1  40386495 135741254  9435.3
## - FAA_uniq$speed_ground  1 500199716 595554475 10590.2
##
## Step:  AIC=9159.53
## FAA_uniq$distance ~ FAA_uniq$speed_ground + FAA_uniq$aircraft +
##     FAA_uniq$height + FAA_uniq$pitch + FAA_uniq$no_pasg
##
##                         Df Sum of Sq       RSS     AIC
## - FAA_uniq$pitch         1     69985  95428702  9158.1
## - FAA_uniq$no_pasg       1    119190  95477907  9158.5
## <none>                               95358717  9159.5
## - FAA_uniq$height        1  15015934 110374650  9271.7
## - FAA_uniq$aircraft      1  40401899 135760616  9433.4
## - FAA_uniq$speed_ground  1 501410372 596769088 10589.8
##
## Step:  AIC=9158.1
## FAA_uniq$distance ~ FAA_uniq$speed_ground + FAA_uniq$aircraft +
##     FAA_uniq$height + FAA_uniq$no_pasg
##
##                         Df Sum of Sq       RSS     AIC
## - FAA_uniq$no_pasg       1    120379  95549081  9157.1
## <none>                               95428702  9158.1
## - FAA_uniq$height        1  15127102 110555804  9271.0
## - FAA_uniq$aircraft      1  47862507 143291209  9473.6
## - FAA_uniq$speed_ground  1 501605243 597033945 10588.1
##
## Step:  AIC=9157.08
## FAA_uniq$distance ~ FAA_uniq$speed_ground + FAA_uniq$aircraft +
##     FAA_uniq$height
##
##                         Df Sum of Sq       RSS     AIC
## <none>                               95549081  9157.1
## - FAA_uniq$height        1  15048298 110597379  9269.3
## - FAA_uniq$aircraft      1  48026872 143575953  9473.1
## - FAA_uniq$speed_ground  1 501619125 597168206 10586.3

Model_lm1 <- lm(FAA_uniq$distance ~ FAA_uniq$speed_ground)
fit2_LM <- stepAIC(Model_LM, direction = 'forward',trace=TRUE)

## Start:  AIC=9161.49
## FAA_uniq$distance ~ FAA_uniq$speed_ground + FAA_uniq$aircraft +
##     FAA_uniq$height + FAA_uniq$pitch + FAA_uniq$duration +
FAA_uniq$no_pasg
```