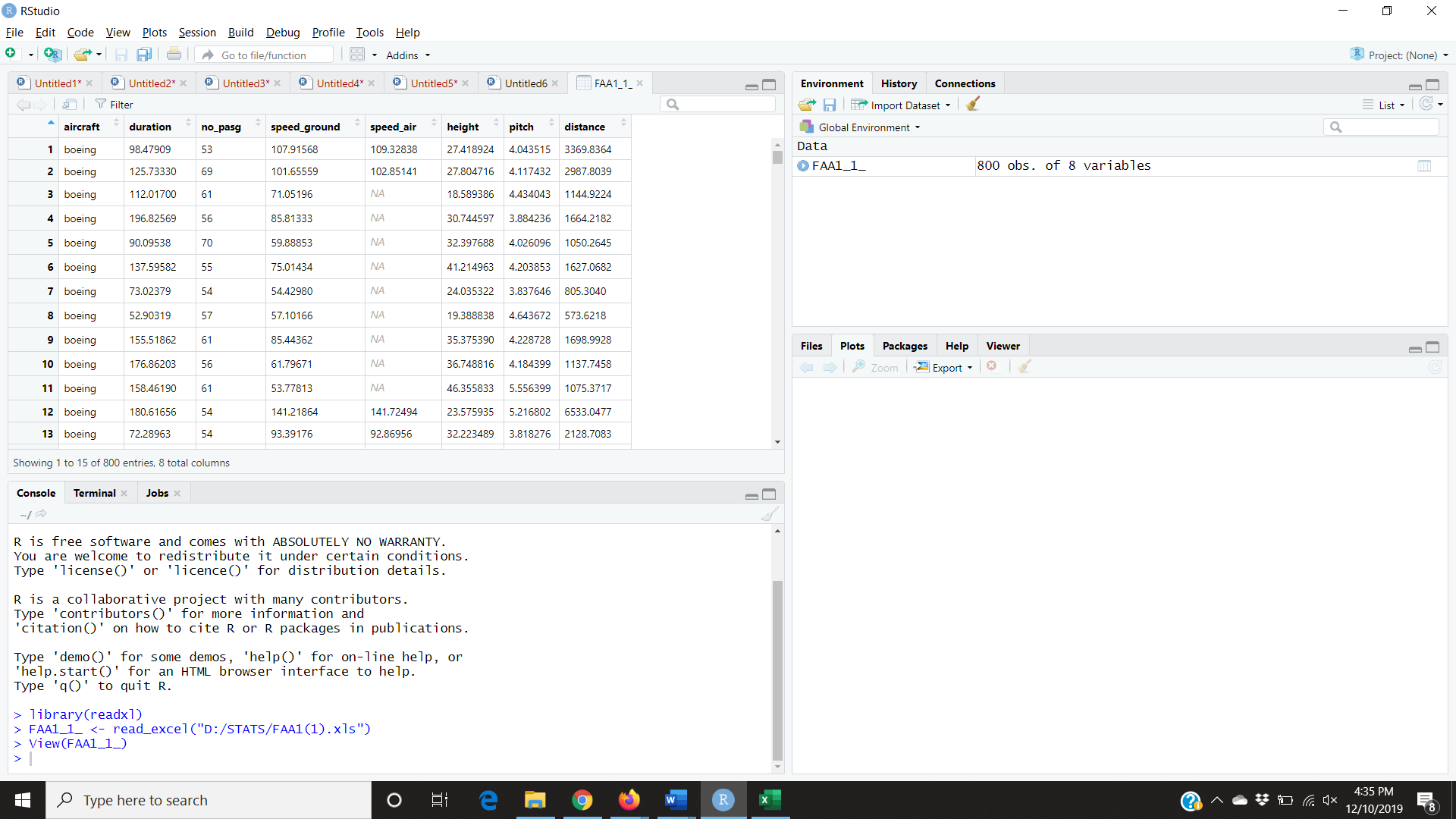
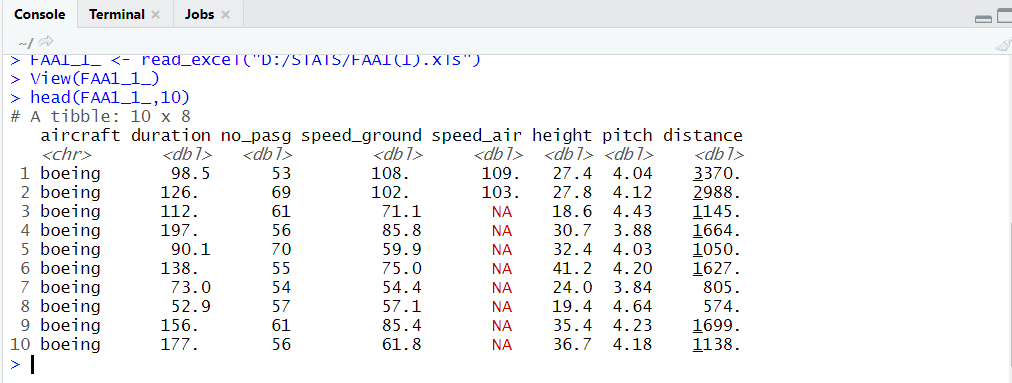
**CHAPTER -1**

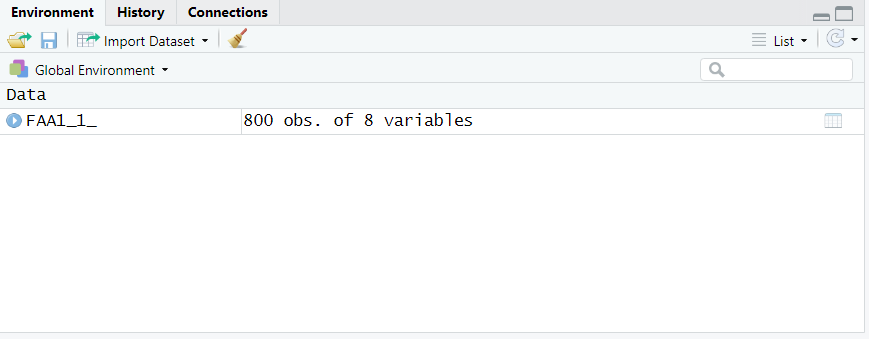
IMPORTING THE DATASET

1. Import “FAA-1.xls” into R.



head(FAA1\_1\_,10) -> to view the dataset





OBSERVATION:

There are a total of 800 observations from 8 variables.

**CHAPTER-2**

**DATA CLEANING**

install.packages("dplyr")

library(dplyr)

#to check missing values

colSums(is.na(FAA1\_1\_))

#to remove missing values

FAA1 = na.omit(FAA1\_1\_)

#to remove duplicates

distinct(FAA1)

#to remove abnormal values

nrow(FAA1)

faa1 = filter(FAA1,duration > 40)

nrow(faa1)

faa2 = filter(faa1,speed\_ground > 30 & speed\_ground <140)

nrow(faa2)

faa3 = filter(faa2,speed\_air>30 & speed\_air <140)

nrow(faa3)

faa4 = filter(faa3,height >= 6)

nrow(faa4)

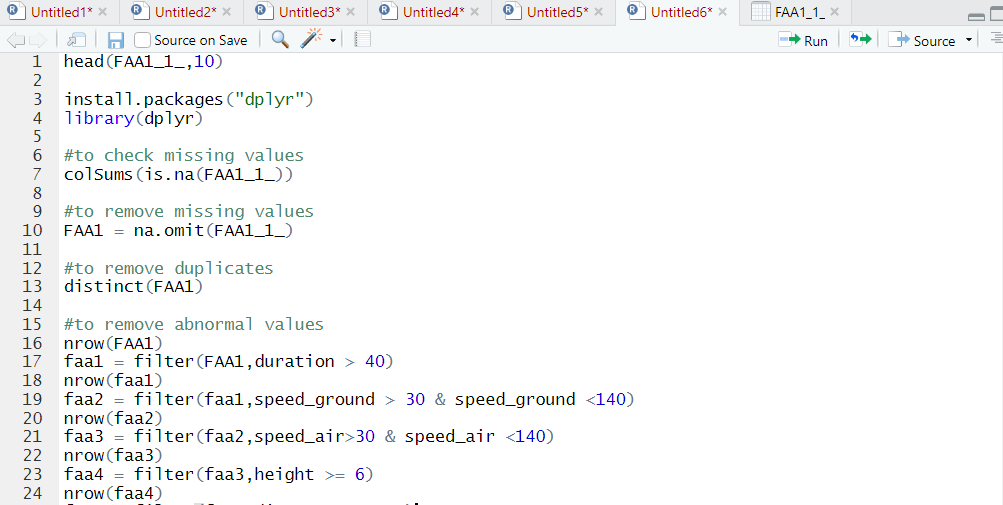
faa5 = filter(faa4,distance < 6000)

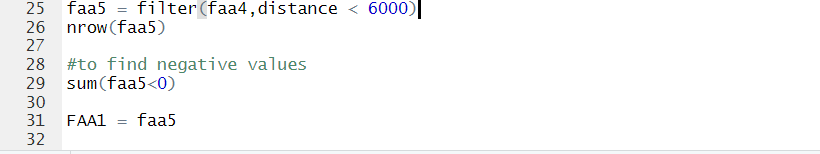
nrow(faa5)

#to find negative values

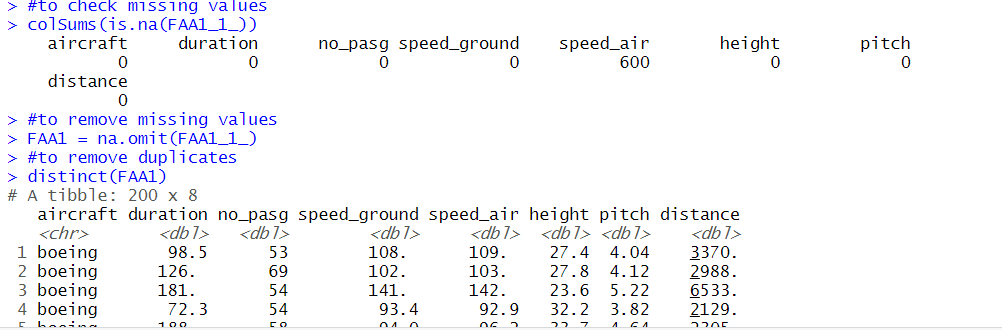
sum(faa5<0)

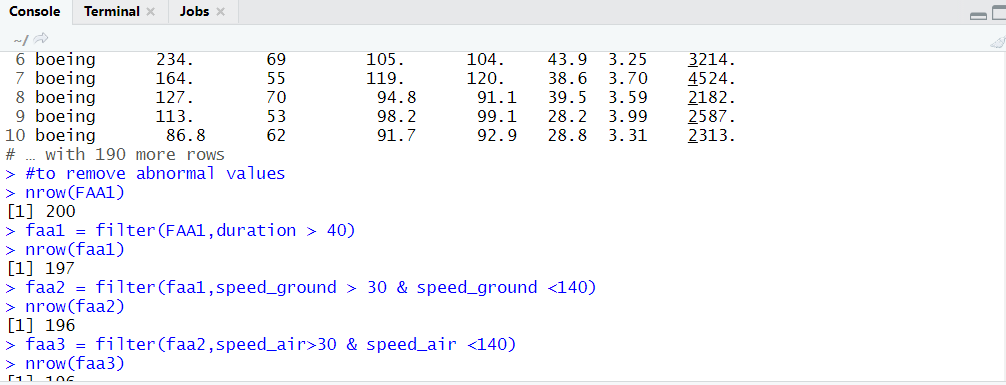
FAA1 = faa5

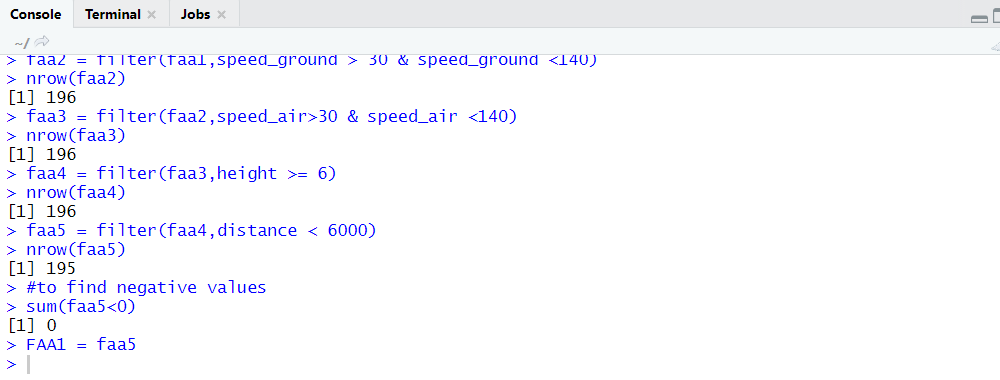




OUTPUT :







OBSERVATIONS:

1. There were 3 abnormal values in duration
2. There was 1 abnormal value in speed\_ground.
3. There was 1 abnormal value in distance.
4. There were 600 missing values in speed\_air and removing them gave a total of 200 observations.

**CHAPTER-3**

**DATA VISUALISATION**

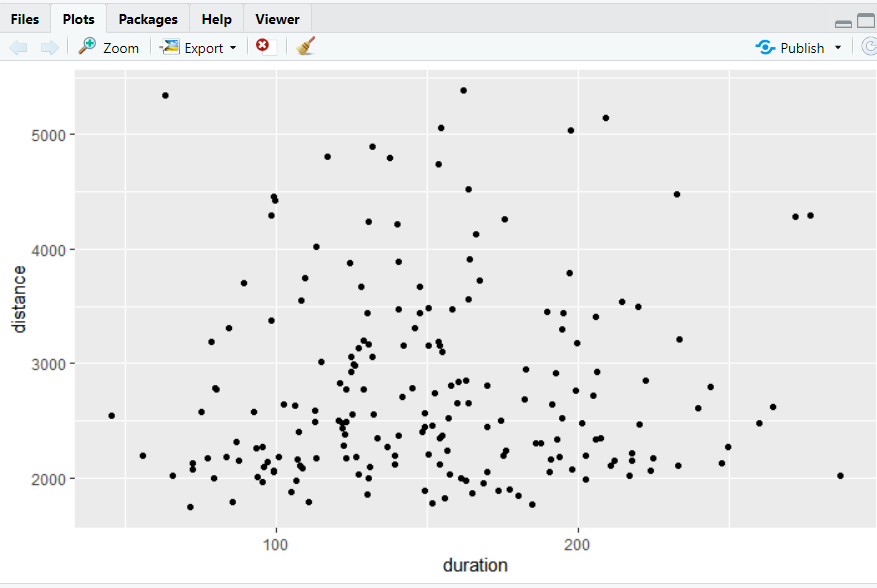
Plotting distance against various predictor variables

1. **Distance Vs. Duration**

install.packages("ggplot2")

library(ggplot2)

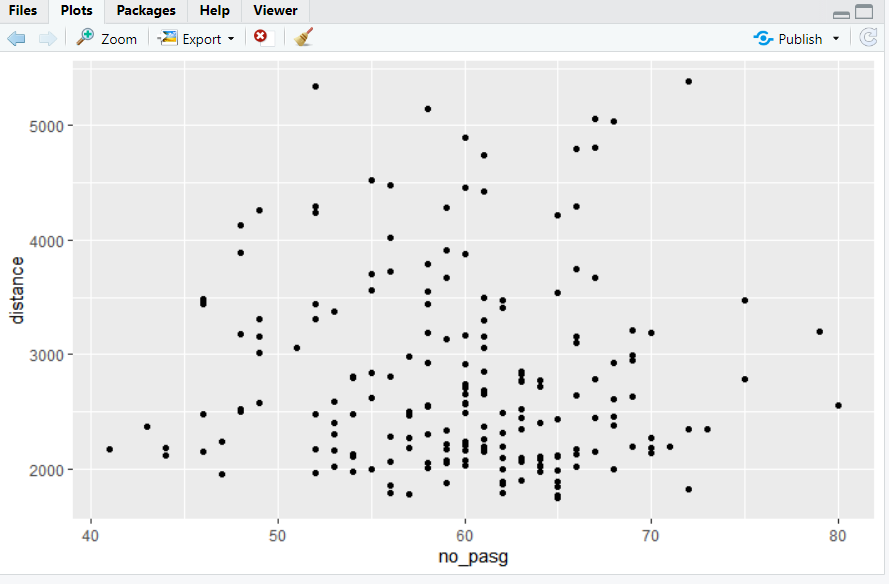
ggplot(data = FAA1, mapping = aes(x = duration, y = distance)) + geom\_point()



1. **Distance Vs. no\_pasg**

#plot distance against no\_pasg

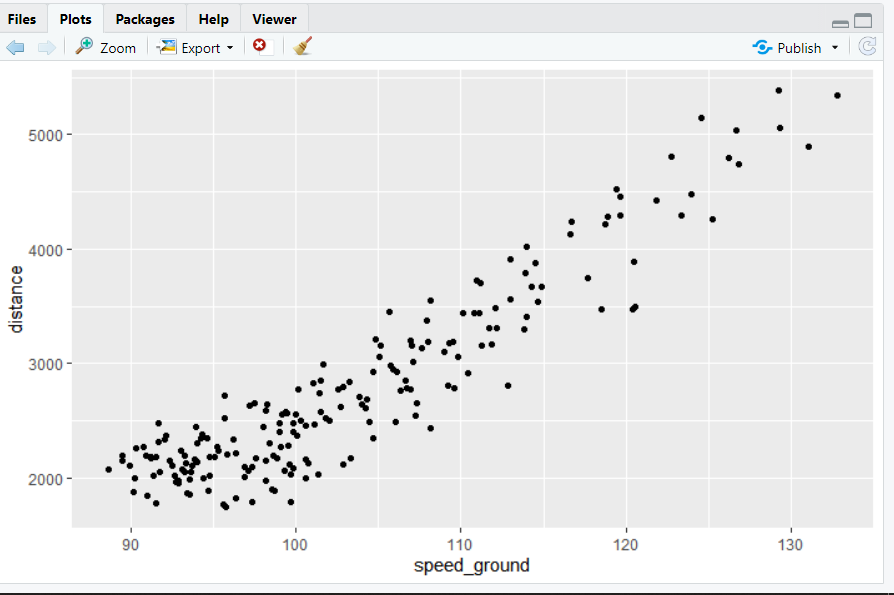
ggplot(data = FAA1, mapping = aes(x = no\_pasg, y = distance)) + geom\_point()



1. **Distance Vs. speed\_ground**

#plot distance against speed\_ground

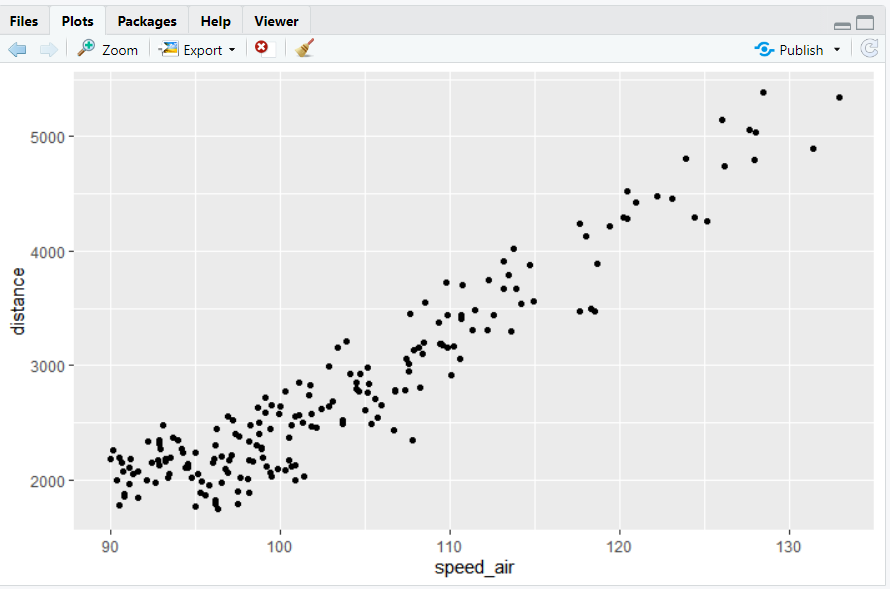
ggplot(data = FAA1, mapping = aes(x = speed\_ground, y = distance)) + geom\_point()



1. **Distance Vs. speed\_air**

#plot distance against speed\_air

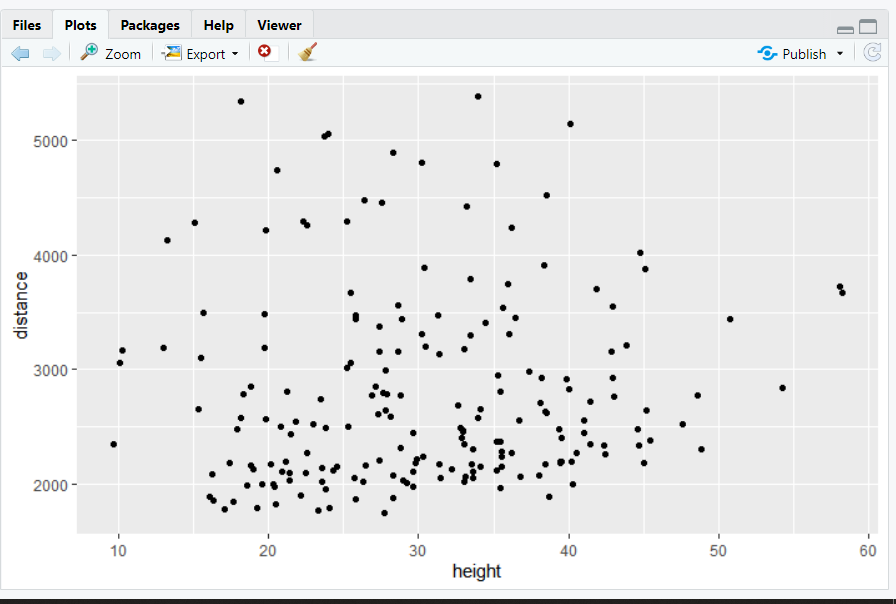
ggplot(data = FAA1, mapping = aes(x = speed\_air, y = distance)) + geom\_point()



1. **Distance Vs. height**

#plot distance against height

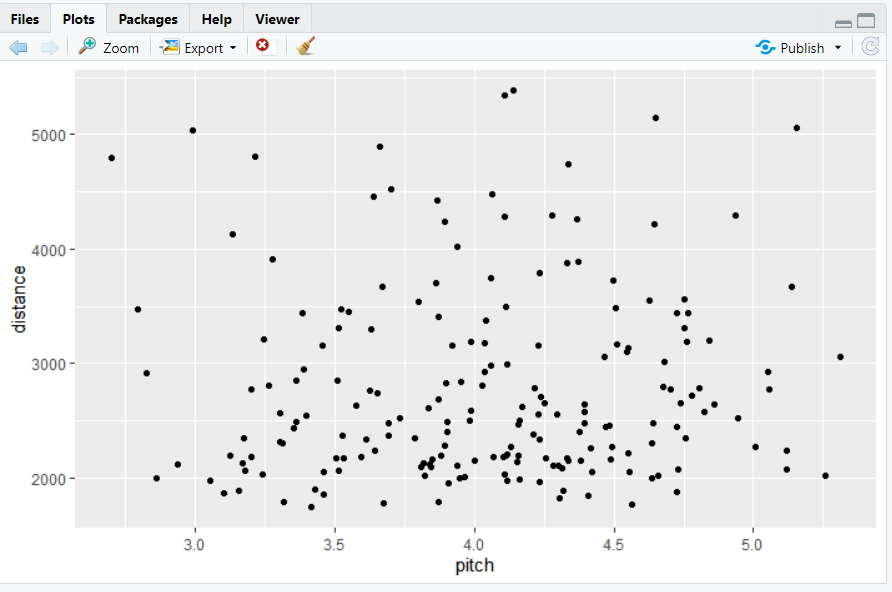
ggplot(data = FAA1, mapping = aes(x = height, y = distance)) + geom\_point()



1. **Distance Vs. pitch**

#plot distance against pitch

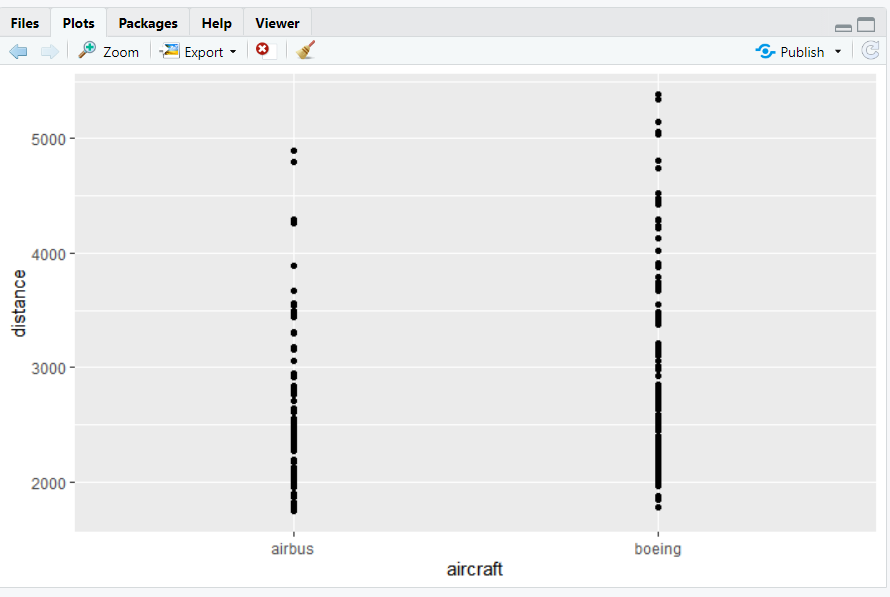
ggplot(data = FAA1, mapping = aes(x = pitch, y = distance)) + geom\_point()



1. **Distance Vs. Aircraft**

#plot distance against aircraft make

ggplot(data = FAA1, mapping = aes(x = aircraft, y = distance)) + geom\_point()



CORRELATION MATRICES

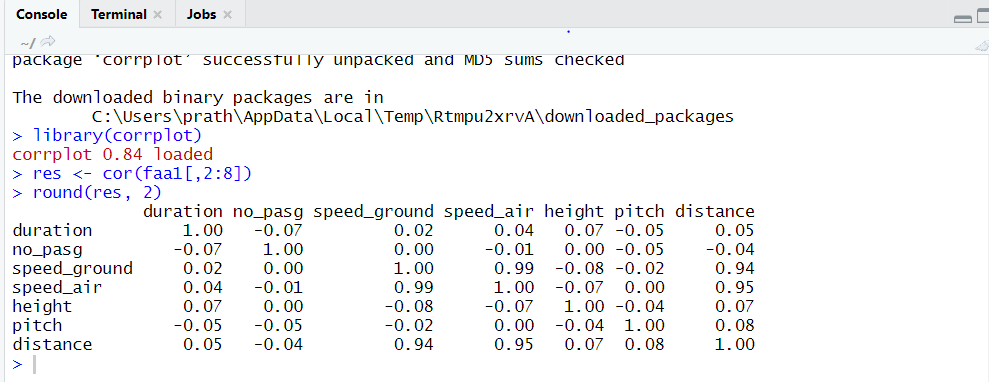
# correlation matrix

install.packages("corrplot")

library(corrplot)

res <- cor(faa1[,2:8])

round(res, 2)



OBSERVATIONS:

There is a strong correlation between distance-speed\_ground, distance-speed\_air and speed\_air-speed\_ground.

**CHAPTER-4**

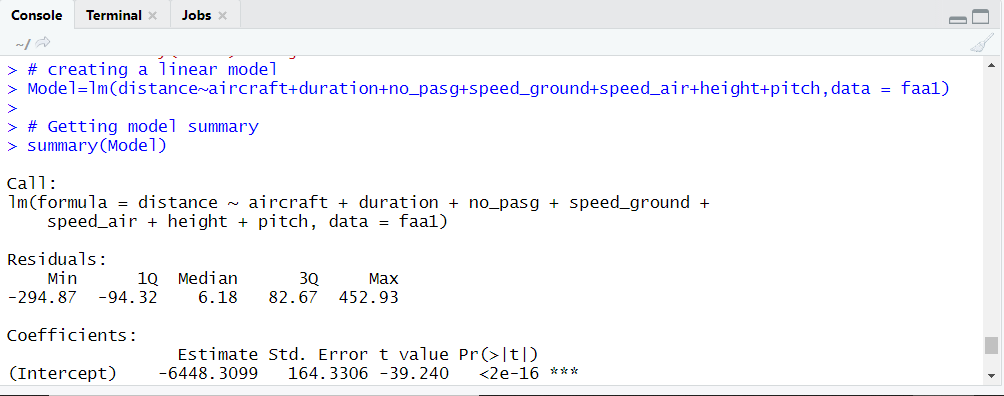
**MODEL FITTING**

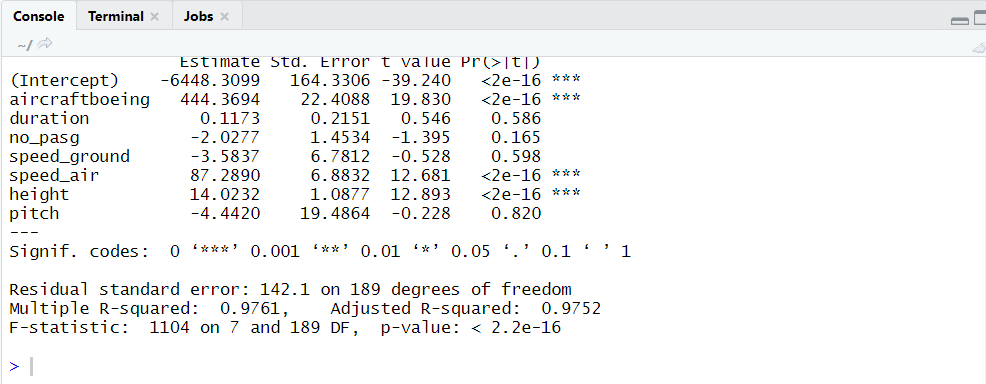
# to create a linear model

Model=lm(distance~aircraft+duration+no\_pasg+speed\_ground+speed\_air+height+pitch,data = faa1)

# to get model summary

summary(Model)





OBSERVATIONS:

From the above results, we can infer that aircraftboeing, speed\_air and height are the only variables that are showing significant values and hence we consider only those variables for modelling.

Additionally, we can note that out of the two aircraft makes, it takes only Boeing into consideration.

**CHAPTER-5**

**MODEL FITTING AND MODEL DIAGNOSTICS FOR AIRCRAFT MAKE**

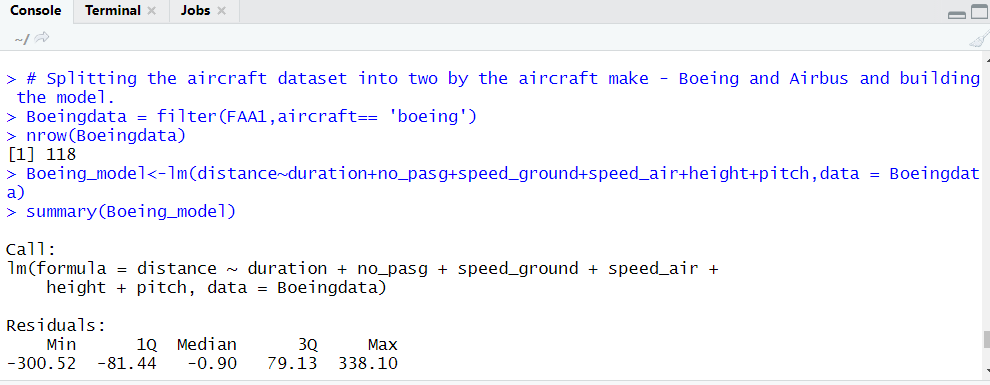
# Splitting the aircraft dataset into two by the aircraft make - Boeing and Airbus and building the model.

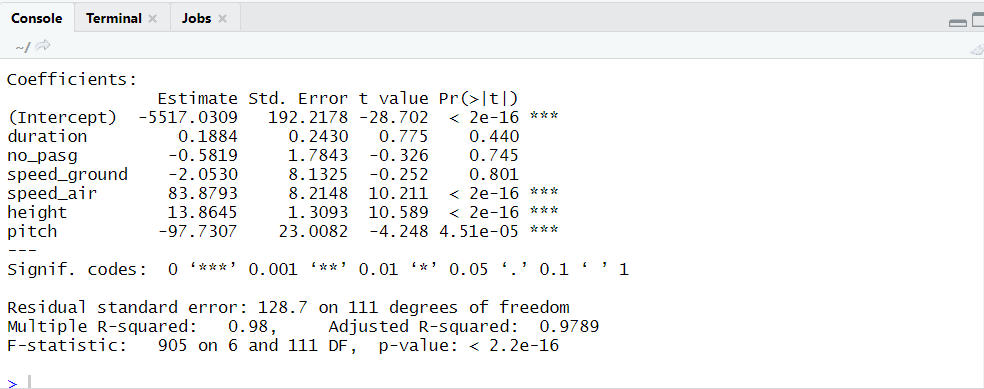
Boeingdata = filter(FAA1,aircraft== 'boeing')

nrow(Boeingdata)

Boeing\_model<-lm(distance~duration+no\_pasg+speed\_ground+speed\_air+height+pitch,data = Boeingdata)

summary(Boeing\_model)





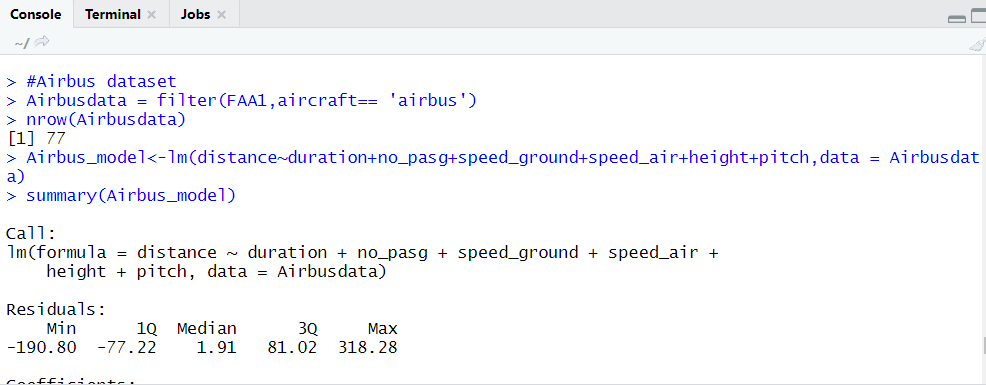
#Airbus dataset

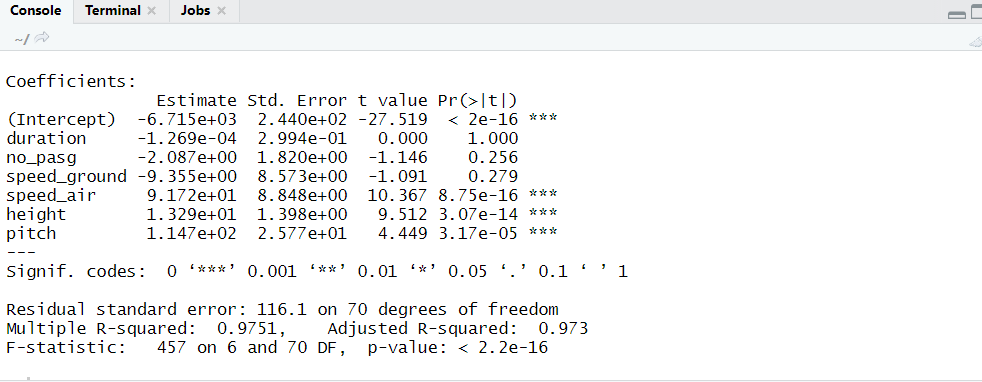
Airbusdata = filter(FAA1,aircraft== 'airbus')

nrow(Airbusdata)

Airbus\_model<-lm(distance~duration+no\_pasg+speed\_ground+speed\_air+height+pitch,data = Airbusdata)

summary(Airbus\_model)





OBSERVATIONS:

1. Boeing covers 97.89 % of dataset.
2. Airbus covers 97.3 % of dataset.
3. Both the aircraft makes have speed\_air, height and pitch as significant variables that can be seen from the coefficient values.
4. But there is a difference in the values and hence we can conclude that the make of an aircraft makes a difference or affects the landing distance of an aircraft.