111# read in the data file

df <- jsonlite::fromJSON("fall2019-survey-M04.json")

New categorical column for Likelihood.to.recommend

# create a new column called score that records whether this customer is detractor, passive or detractor based on the likelihood to recommend

df$score <- "Passive"

df$score[df$Likelihood.to.recommend>8] <- "Promoter"

df$score[df$Likelihood.to.recommend < 7] <- "Detractor"

**Code to make individual datasets for each partner airline:**

library(dplyr)

library(stringr)

# Subset of df for each airline

SouthEastAirlines <- df%>%

filter(Partner.Name=="Southeast Airlines Co.")

OursinAirlines <- df%>%

filter(Partner.Name=="Oursin Airlines Inc.")

NorthwestBusinessAirlines <- df%>%

filter(Partner.Name=="Northwest Business Airlines Inc.")

CheapseatsAirlines <- df%>%

filter(Partner.Name=="Cheapseats Airlines Inc.")

FlyToSunAirlines <- df%>%

filter(Partner.Name=="FlyToSun Airlines Inc.")

SigmaAirlines <- df%>%

filter(Partner.Name=="Sigma Airlines Inc.")

EnjoyFlyingAirlines <- df%>%

filter(Partner.Name=="EnjoyFlying Air Services")

OnlyJetsAirlines <- df%>%

filter(Partner.Name=="OnlyJets Airlines Inc.")

------------------------------------------------------------------

FlyFastAirlines <- df %>%

filter(Partner.Name == "FlyFast Airways Inc." )

PaulSmithAirlines <- df%>%

filter(Partner.Name=="Paul Smith Airlines Inc.")

FlyHereAirways <- df%>%

filter(Partner.Name=="FlyHere Airways")

CoolYoungAirlines <- df%>%

filter(Partner.Name=="Cool&Young Airlines Inc.")

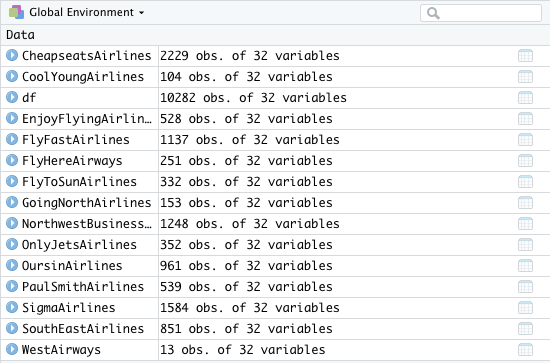
GoingNorthAirlines <- df%>%

filter(Partner.Name=="GoingNorth Airlines Inc.")

WestAirways <- df%>%

filter(Partner.Name=="West Airways Inc.")

We should have 14 subsets and the main dataset like this:



Things/actions to perform on each individual dataset:

* histogram summary
* Number of observations
* Descriptive stats (mean, median, mode..)

# a function that takes in a subset of data of certain airlines and output its **NPS score**

NPS.calculator <- function(air\_sub)

{

detractors.num <- air\_sub %>%

filter(air\_sub$Likelihood.to.recommend < 7) %>%

count()

promoters.num <- air\_sub %>%

filter(air\_sub$Likelihood.to.recommend> 8) %>%

count()

total.num <- count(air\_sub)

return((promoters.num-detractors.num)/total.num\*100)

}

1. Find similar coefficient
2. Compare histogram/ NPS score

Age in these groups.

**Table of Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | # observation | histogram | Descriptive stats (mean, median, mode..)  NPS |
| SouthEastAirlines | 851 |  | Mean = 7.4336  Median = 8  Mode = 9  NPS (%) = 11.9859 |
| OursinAirlines | 961 |  | Mean = 7.0343  Median = 7  Mode = 9  NPS (%) = 0.312 |
| NorthwestBusinessAirlines | 1248 |  | Mean = 5.4679  Median = 6  Mode = 7  NPS (%) = -57.9 |
| CheapseatsAirlines | 2229 |  | Mean = 7.113  Median = 8  Mode = 9  NPS (%) = 2.0188 |
| SigmaAirlines | 1584 |  | Mean = 7.477  Median = 8  Mode = 9  NPS (%) = 16.09848 |
| OnlyJetAirlines | 352 |  | Mean = 7.33  Median = 8  Mode = 9  NPS (%) = 12.5 |
| EnjoyFlyingAirlines | 528 |  | Mean = 7.309  Median = 8  Mode = 9  NPS (%) = 10.79545 |
| FlyToSunAirlines | 332 |  | Mean = 7.599  Median = 8  Mode = 9  NPS (%) = 19.27711 |
| FlyFast Airways Inc. | 1137 |  | Mean: 6.544  Median: 7  Mode: 8  NPS(%) = -19.26121 |
| Paul Smith Airlines Inc. | 539 |  | Mean: 7.553  Median: 8  Mode: 9  NPS(%) = 17.06865 |
| FlyHere Airways | 251 |  | Mean: 7.442  Median: 8  Mode: 9  NPS (%) = 14.34263 |
| Cool & Young Airlines Inc. | 104 |  | Mean: 7.78  Median: 9  Mode: 9  NPS (%) = 28.84615 |
| Going North Airlines Inc. | 153 |  | Mean: 6.732  Median: 7  Mode: 9  NPS (%) = -11.11111 |
| West Airways Inc. | 13 |  | Mean: 8.231  Median: 9  Mode: 9  NPS: 46.15385 |

## simple multi - linear regression

df <- subset(df, select = -freeText)

df <- na.omit(df)

df\_simple <- df[,c("Airline.Status”,"Age","Gender","Price.Sensitivity","Year.of.First.Flight","Flights.Per.Year","Loyalty","Type.of.Travel","Total.Freq.Flyer.Accts","Shopping.Amount.at.Airport","Eating.and.Drinking.at.Airport","Class","Day.of.Month","Partner.Name","Scheduled.Departure.Hour","Departure.Delay.in.Minutes","Arrival.Delay.in.Minutes","Flight.time.in.minutes","Flight.Distance","Likelihood.to.recommend")]

df\_simple$Gender.f <- factor(df\_simple$Gender)

df\_simple$Airline.Status.f <- factor(df\_simple$Airline.Status)

df\_simple$Type.of.Travel.f <- factor(df\_simple$Type.of.Travel)

df\_simple$Class.f <- factor(df\_simple$Class)

df\_simple$Partner.Name.f <- factor(df\_simple$Partner.Name)

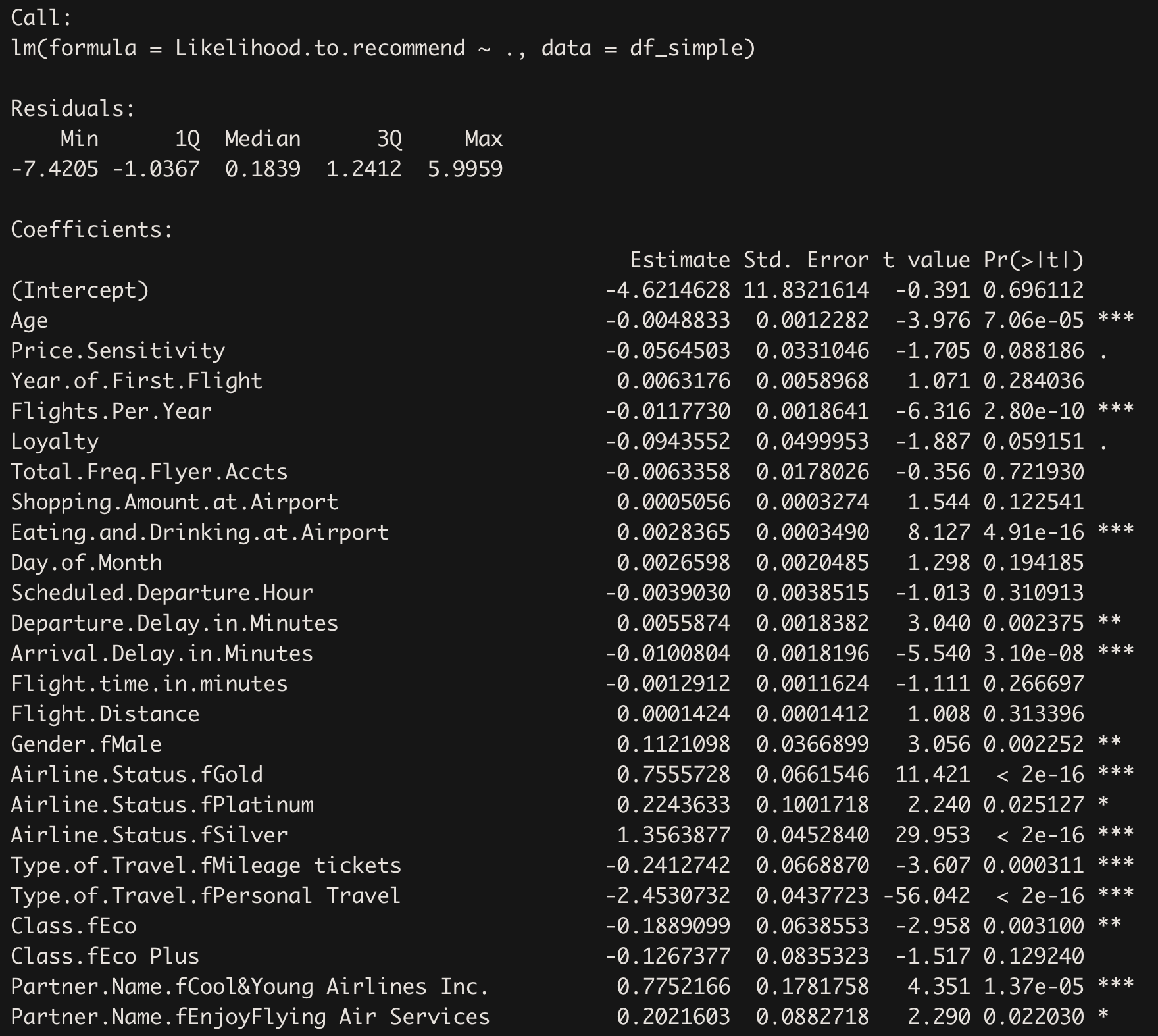
df\_simple <- subset(df\_simple, select = -c(Gender,Airline.Status))

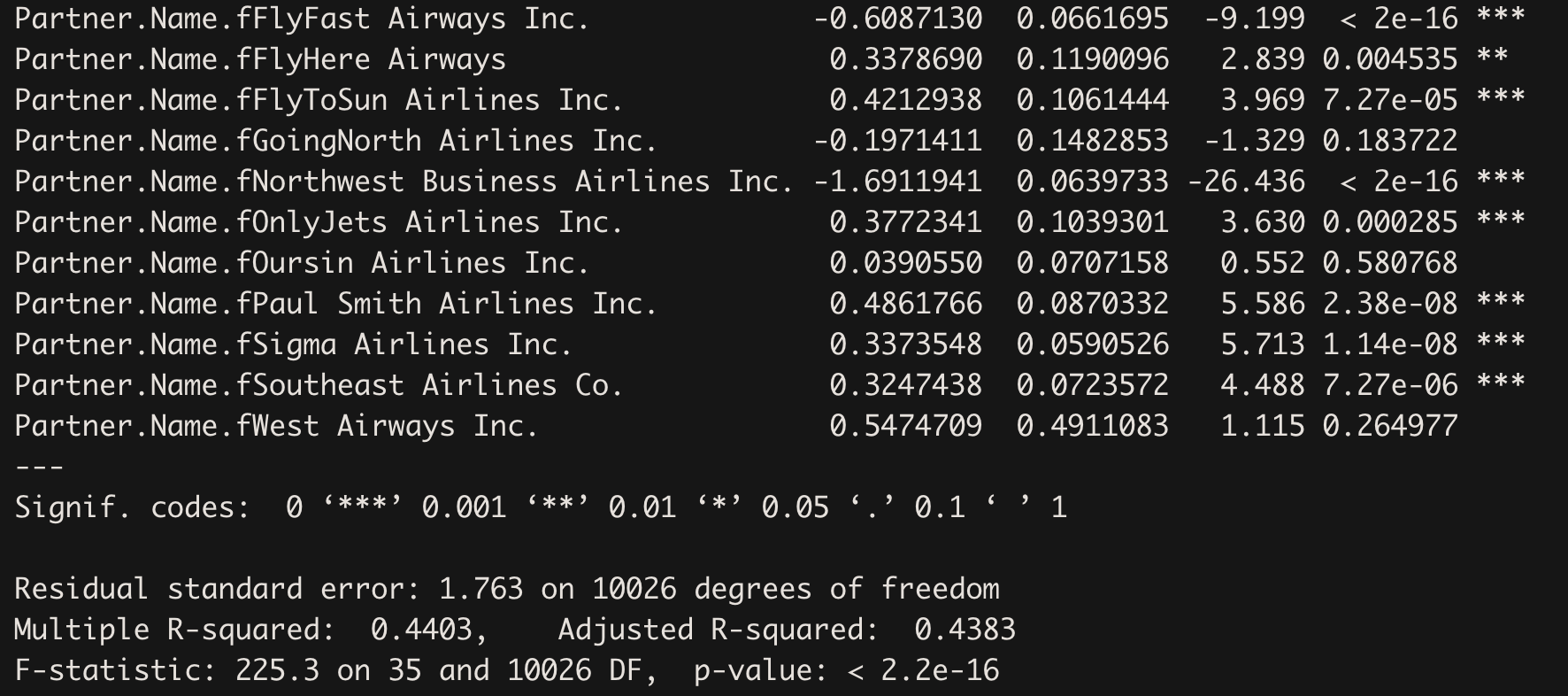
df\_simple <- subset(df\_simple, select = -c(Type.of.Travel,Class,Partner.Name))

lm\_simple <- lm(Likelihood.to.recommend~.,data = df\_simple)

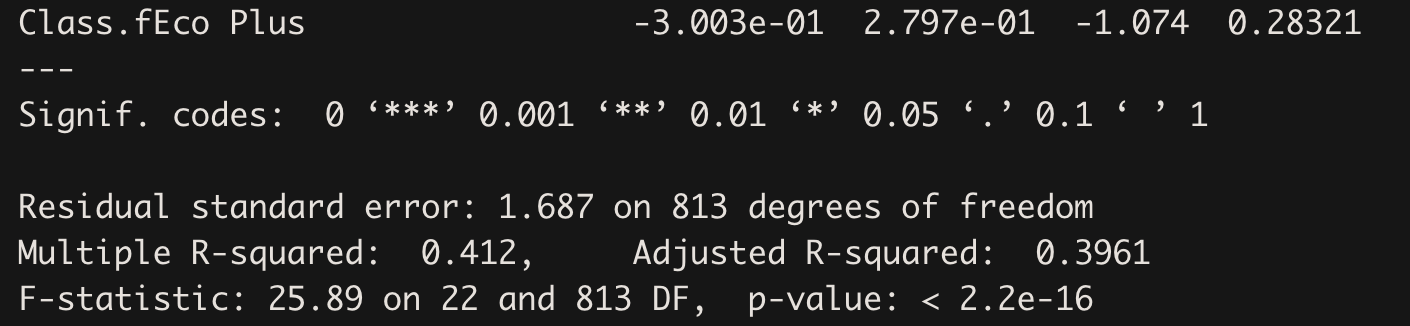
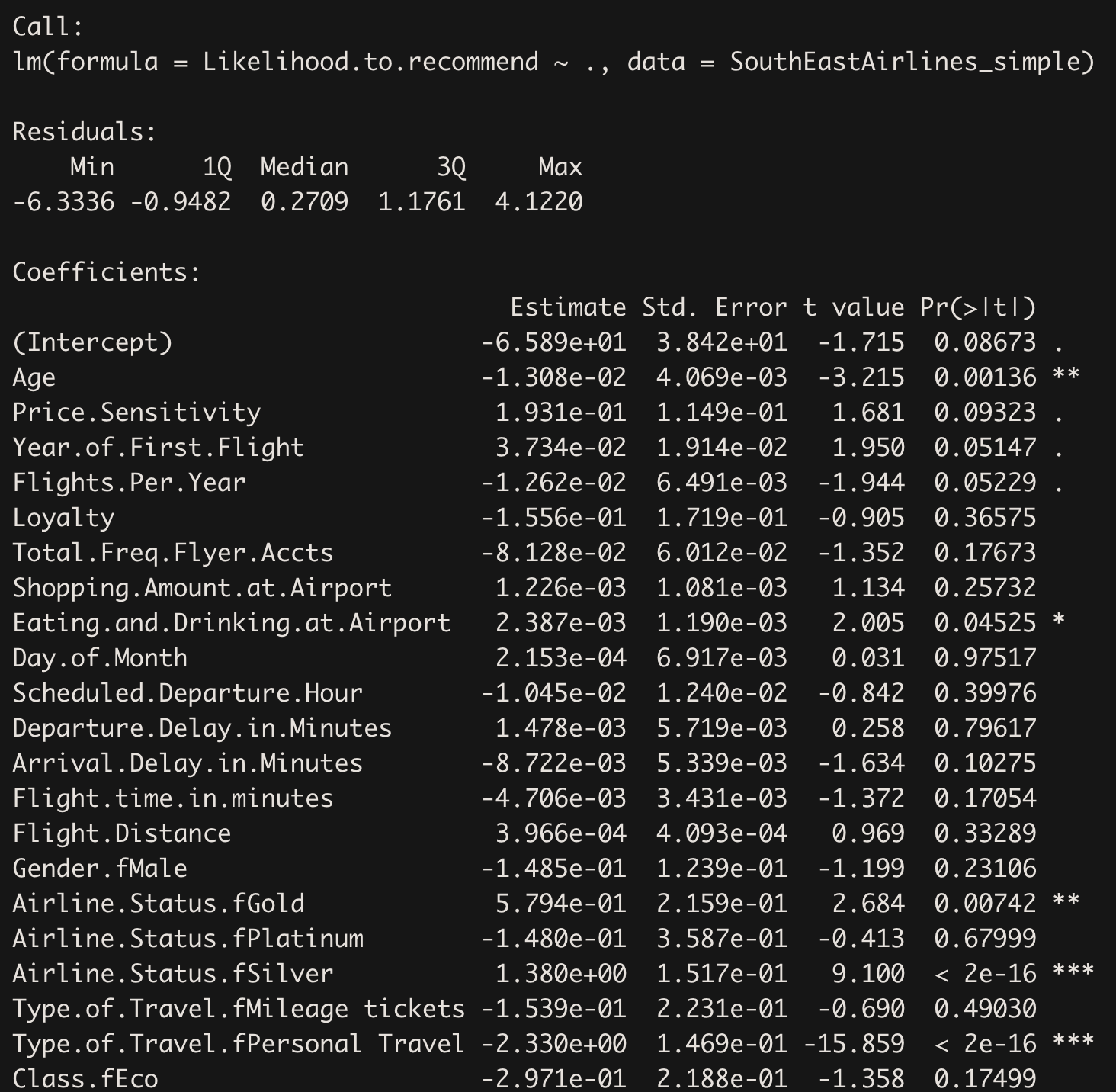
summary(lm\_simple)

### Total dataframe





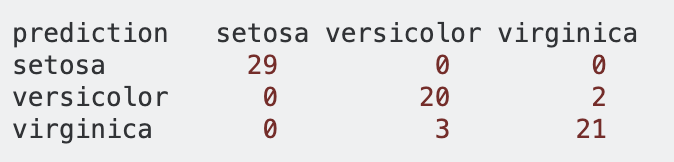
Southeast Airlines - Linear Regression of all fields



## confusion matrix for three classification

library(caret)

confusionMatrix(predict data,test data)



#Creating subset

Df <- Df%>%

filter(Partner.Name=="Oursin Airlines Inc."|Partner.Name=="Cheapseats Airlines Inc."|Partner.Name=="Northwest Business Airlines Inc."|Partner.Name=="FlyFast Airways Inc."|Partner.Name=="GoingNorth Airlines Inc.")