```
#Austin Dickerson

#Inclass 9-2


getwd()

setwd("C:\\Users\\Austin\\Desktop\\Virginia Tech\\CS 3654\\CMDA")

load('fdata.RData')

attach(final)


final$gp <- runif(dim(final)[1])

testSet <- subset(final, final$gp <= 0.1)

trainSet <- subset(final, final$gp > 0.1)

rm(final)

install.packages("MASS")

library(MASS)

attach(trainSet)

fit <- lm(disorder ~ som1 + som2 + som3 + som4 + som5 + som6 + som7 + som8 + som9 + som10 +
som11 + som12 + som13 + som14 + age + gender + location + ethnicity + coder)

step <- stepAIC(fit, direction = "both")

step

step$ssc

rm(step)

rm(fit)

fit1 <- lm(disorder ~ som1 + som2 + som3 + som4 + som5 + som6 + som7 + som8 + som9 + som10 +
som11 + som12 + som13 + som14 + age + gender + location + ethnicity + coder)

testSet$casualpred <- predict(fit1, newdata = testSet)

head(testSet)

library(ggplot2)

ggplot(data = testSet, aes(x = casualpred, y = casual)) +
```

```
  geom_point(color = "red")+
  geom_line(aes(x = casual, y = casual), color = "blue")
#This model is not very accurate, the “line of best fit” doesn’t really match up with the scatterplot dots
```