```
setwd("C:/R")
load('fdata.RData')
attach(final)
head(final)
final$gp <- runif(dim(final)[1])
testSet <- subset(final, final$gp <= 0.1)
trainSet <- subset(final, final$gp > 0.1)
trainSet["gp"] <- NULL
testSet["gp"] <- NULL
rm(final)
library(MASS)
attach(trainSet)
fit <- lm(ssc ~ age + gender + location + ethnicity + coder
          + som1 + som2 + som3 + som4 + som5 + som6 + som7
          + som8 + som9 + som10 + som11 + som12 + som13 + som14)
summary(fit)
step <- stepAIC(fit, direction = "both")
step
step$anova
rm(step)
rm(fit)
fit1 <- lm( ssc ~ age + location + ethnicity + coder + som1 + som2 + som3 +
            som4 + som5 + som10 + som11 + som12 + som13 + som14)
summary(fit1)
detach(trainSet)
rm(trainSet)
testSet$ssc_pred <- predict(fit1, newdata = testSet)
rm(fit1)
library(ggplot2)
```

```
ggplot(data = testSet, aes(x = ssc_pred, y = ssc)) +
  geom_point(color = "red") +
  geom_line(aes(x = ssc, y = ssc), color = "blue")
```