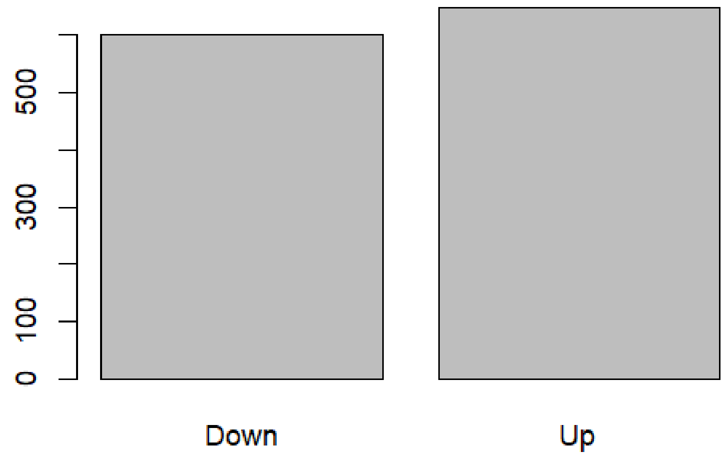
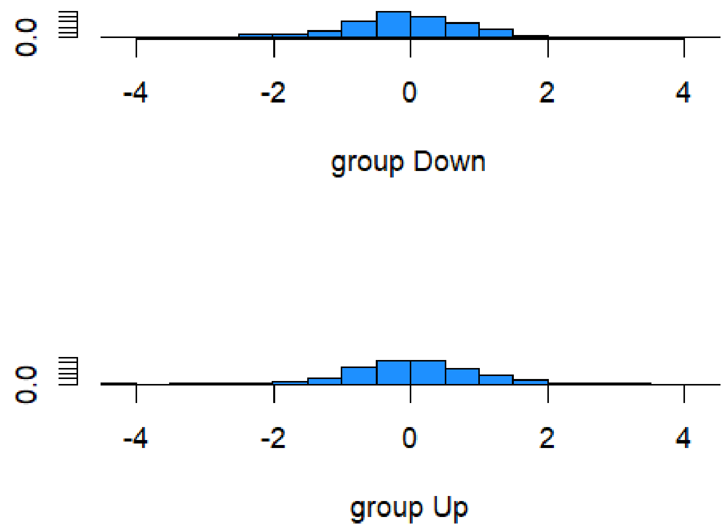
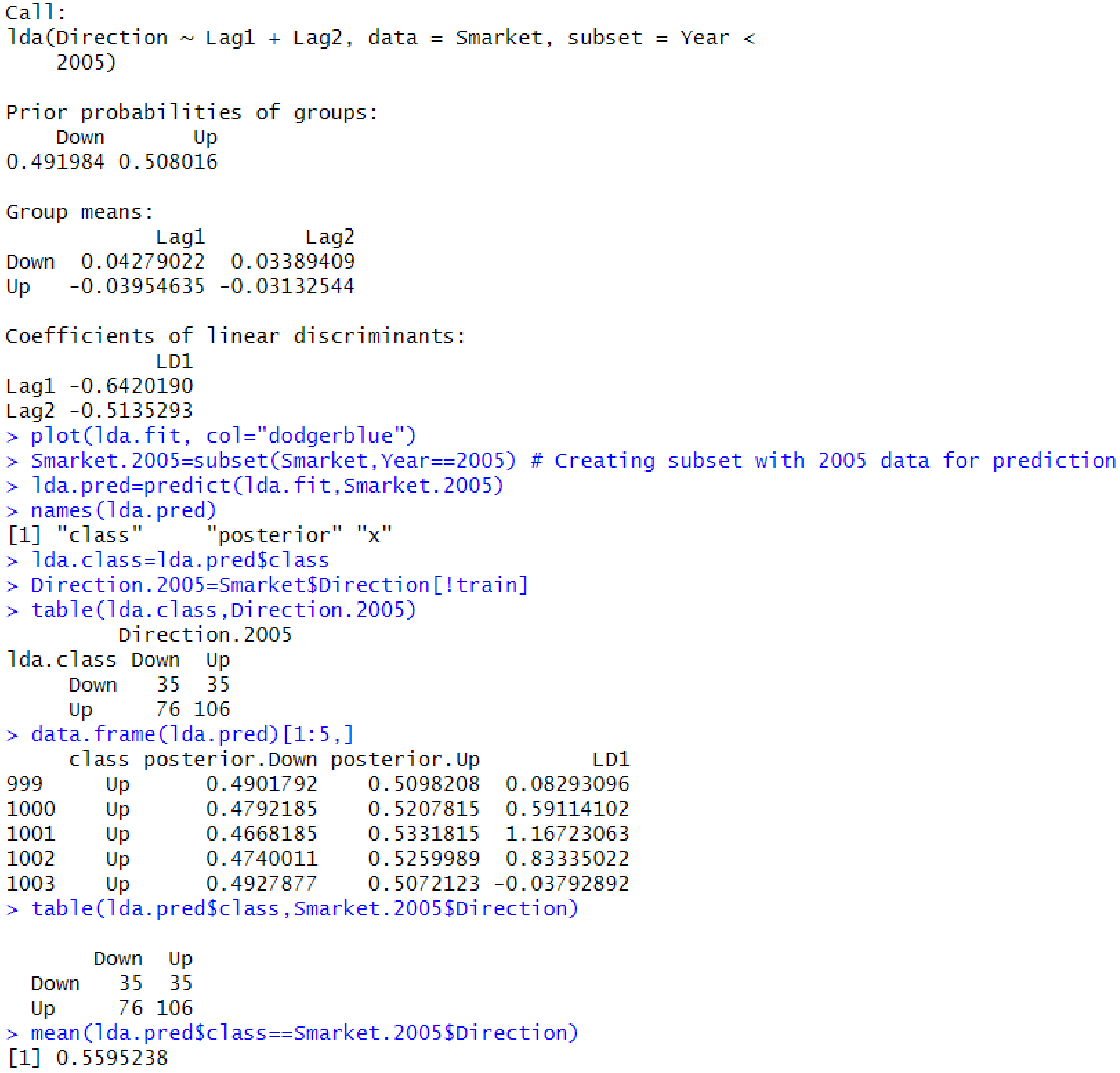
Assignment 8

Christina Lam

Lab_LDA01







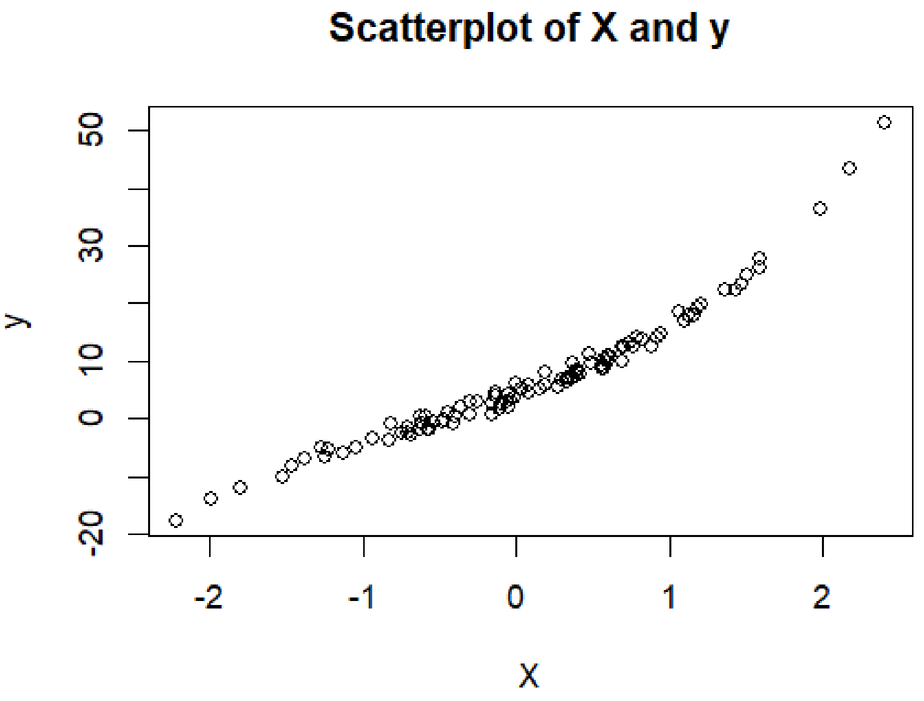
```
Call:
lda(Direction ~ Lag1 + Lag2, data = Smarket, subset = Year <
    2005)

Prior probabilities of groups:
    Down       Up
0.491984 0.508016

Group means:
            Lag1        Lag2
Down  0.04279022  0.03389409
Up   -0.03954635 -0.03132544

Coefficients of linear discriminants:
            LD1
Lag1 -0.6420190
Lag2 -0.5135293
> plot(lda.fit, col="dodgerblue")
> Smarket.2005=subset(Smarket,Year==2005) # Creating subset with 2005 data for prediction
> lda.pred=predict(lda.fit,Smarket.2005)
> names(lda.pred)
[1] "class"     "posterior" "x"
> lda.class=lda.pred$class
> Direction.2005=Smarket$Direction[!train]
> table(lda.class,Direction.2005)
         Direction.2005
lda.class Down  Up
     Down   35  35
     Up     76 106
> data.frame(lda.pred)[1:5,]
     class posterior.Down posterior.Up         LD1
999     Up      0.4901792    0.5098208  0.08293096
1000    Up      0.4792185    0.5207815  0.59114102
1001    Up      0.4668185    0.5331815  1.16723063
1002    Up      0.4740011    0.5259989  0.83335022
1003    Up      0.4927877    0.5072123 -0.03792892
> table(lda.pred$class,Smarket.2005$Direction)

       Down  Up
  Down   35  35
  Up     76 106
> mean(lda.pred$class==Smarket.2005$Direction)
[1] 0.5595238
```

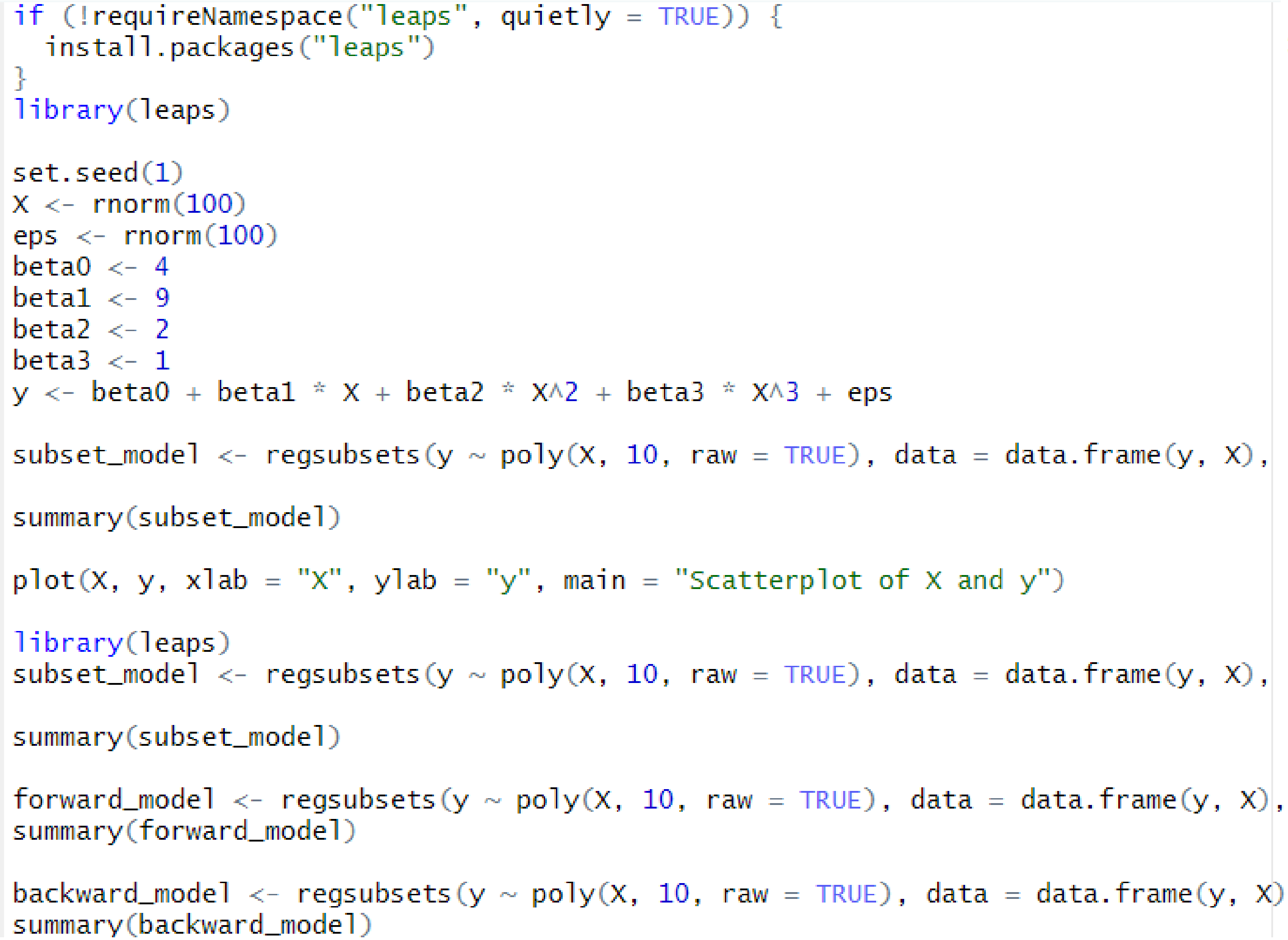
Best Subset, Forward Stepwise, & Backward Stepwise

- **Which of the 3 models with k predictors has the smallest training RSS?**
  - Best Subset
    - Because forward and backward stepwise determine models that depend on which predictors they pick first as they iterate toward the kth model (a poor choice early on cannot be undone)

- **Which of the 3 models with k predictors has the smallest test RSS?**
  - Best Subset
  - Because it considers more models than the other methods, however the other models might have better luck picking a model that fits the test data better since it would be less subject to overfitting
  - Outcome depends more heavily on the choice of test set/validation method than on the selection method

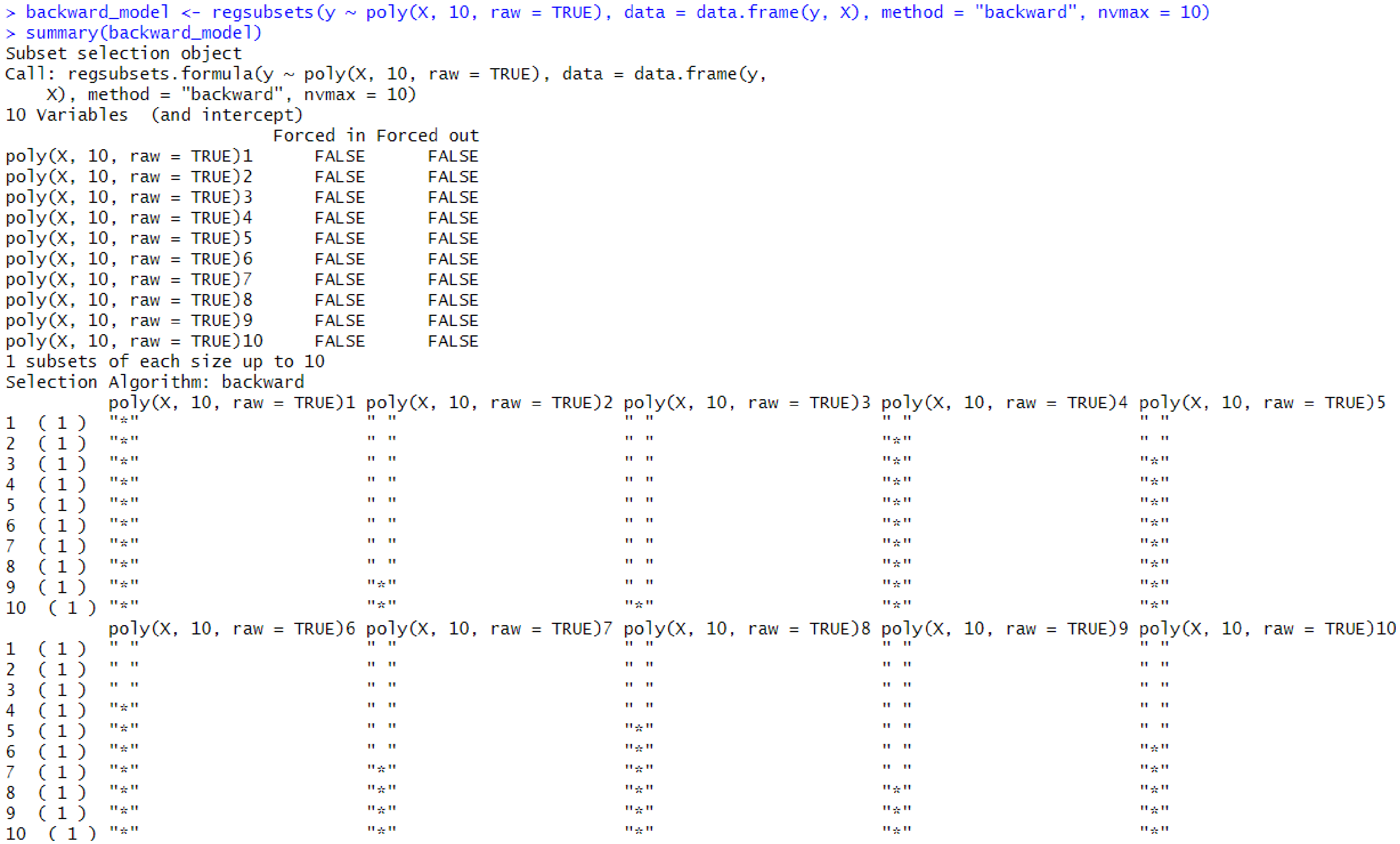
Application Exercise





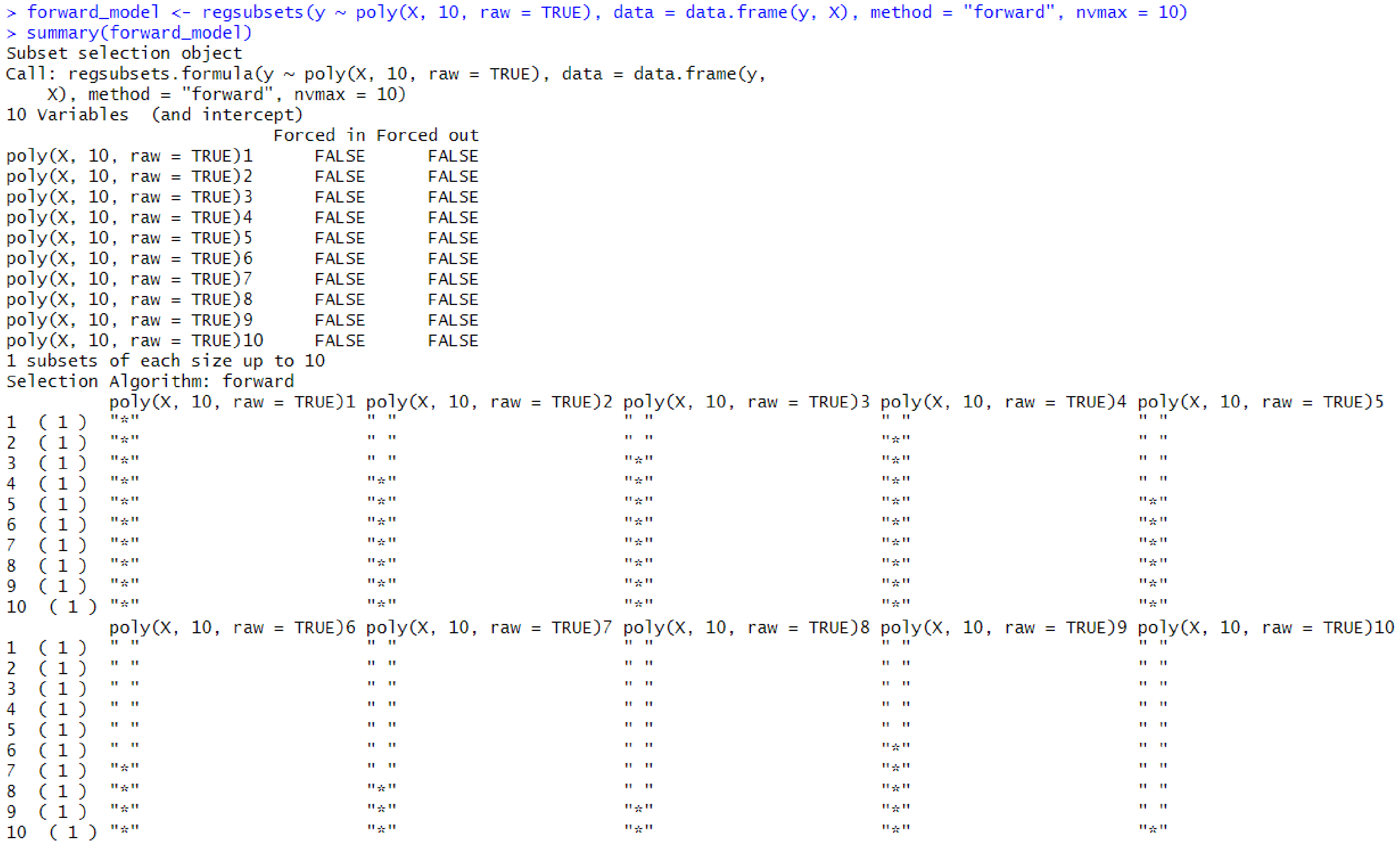
```
if (!requireNamespace("leaps", quietly = TRUE)) {
  install.packages("leaps")
}
library(leaps)

set.seed(1)
X <- rnorm(100)
eps <- rnorm(100)
beta0 <- 4
beta1 <- 9
beta2 <- 2
beta3 <- 1
y <- beta0 + beta1 * X + beta2 * X^2 + beta3 * X^3 + eps

subset_model <- regsubsets(y ~ poly(X, 10, raw = TRUE), data = data.frame(y, X), nvmax = 10)

summary(subset_model)

plot(X, y, xlab = "X", ylab = "y", main = "Scatterplot of X and y")

library(leaps)
subset_model <- regsubsets(y ~ poly(X, 10, raw = TRUE), data = data.frame(y, X), nvmax = 10)

summary(subset_model)

forward_model <- regsubsets(y ~ poly(X, 10, raw = TRUE), data = data.frame(y, X), method = "forward", nvmax = 10)
summary(forward_model)

backward_model <- regsubsets(y ~ poly(X, 10, raw = TRUE), data = data.frame(y, X), method = "backward", nvmax = 10)
summary(backward_model)
```

Application Exercise



```
> backward_model <- regsubsets(y ~ poly(X, 10, raw = TRUE), data = data.frame(y, X), method = "backward", nvmax = 10)
> summary(backward_model)
Subset selection object
Call: regsubsets.formula(y ~ poly(X, 10, raw = TRUE), data = data.frame(y, 
    X), method = "backward", nvmax = 10)
10 Variables  (and intercept)
                          Forced in Forced out
poly(X, 10, raw = TRUE)1      FALSE      FALSE
poly(X, 10, raw = TRUE)2      FALSE      FALSE
poly(X, 10, raw = TRUE)3      FALSE      FALSE
poly(X, 10, raw = TRUE)4      FALSE      FALSE
poly(X, 10, raw = TRUE)5      FALSE      FALSE
poly(X, 10, raw = TRUE)6      FALSE      FALSE
poly(X, 10, raw = TRUE)7      FALSE      FALSE
poly(X, 10, raw = TRUE)8      FALSE      FALSE
poly(X, 10, raw = TRUE)9      FALSE      FALSE
poly(X, 10, raw = TRUE)10     FALSE      FALSE
1 subsets of each size up to 10
Selection Algorithm: backward
          poly(X, 10, raw = TRUE)1 poly(X, 10, raw = TRUE)2 poly(X, 10, raw = TRUE)3 poly(X, 10, raw = TRUE)4 poly(X, 10, raw = TRUE)5
1  ( 1 )  "*"                      " "                      " "                      " "                      " "
2  ( 1 )  "*"                      " "                      " "                      "*"                      " "
3  ( 1 )  "*"                      " "                      " "                      "*"                      "*"
4  ( 1 )  "*"                      " "                      " "                      "*"                      "*"
5  ( 1 )  "*"                      " "                      " "                      "*"                      "*"
6  ( 1 )  "*"                      " "                      " "                      "*"                      "*"
7  ( 1 )  "*"                      " "                      " "                      "*"                      "*"
8  ( 1 )  "*"                      " "                      " "                      "*"                      "*"
9  ( 1 )  "*"                      "*"                      " "                      "*"                      "*"
10  ( 1 ) "*"                      "*"                      "*"                      "*"                      "*"
          poly(X, 10, raw = TRUE)6 poly(X, 10, raw = TRUE)7 poly(X, 10, raw = TRUE)8 poly(X, 10, raw = TRUE)9 poly(X, 10, raw = TRUE)10
1  ( 1 )  " "                      " "                      " "                      " "                      " "
2  ( 1 )  " "                      " "                      " "                      " "                      " "
3  ( 1 )  " "                      " "                      " "                      " "                      " "
4  ( 1 )  "*"                      " "                      " "                      " "                      " "
5  ( 1 )  "*"                      " "                      "*"                      " "                      " "
6  ( 1 )  "*"                      " "                      "*"                      " "                      "*"
7  ( 1 )  "*"                      "*"                      "*"                      " "                      "*"
8  ( 1 )  "*"                      "*"                      "*"                      "*"                      "*"
9  ( 1 )  "*"                      "*"                      "*"                      "*"                      "*"
10  ( 1 ) "*"                      "*"                      "*"                      "*"                      "*"
```

Application Exercise



```
> forward_model <- regsubsets(y ~ poly(X, 10, raw = TRUE), data = data.frame(y, X), method = "forward", nvmax = 10)
> summary(forward_model)
Subset selection object
Call: regsubsets.formula(y ~ poly(X, 10, raw = TRUE), data = data.frame(y, 
    X), method = "forward", nvmax = 10)
10 Variables  (and intercept)
                          Forced in Forced out
poly(X, 10, raw = TRUE)1      FALSE      FALSE
poly(X, 10, raw = TRUE)2      FALSE      FALSE
poly(X, 10, raw = TRUE)3      FALSE      FALSE
poly(X, 10, raw = TRUE)4      FALSE      FALSE
poly(X, 10, raw = TRUE)5      FALSE      FALSE
poly(X, 10, raw = TRUE)6      FALSE      FALSE
poly(X, 10, raw = TRUE)7      FALSE      FALSE
poly(X, 10, raw = TRUE)8      FALSE      FALSE
poly(X, 10, raw = TRUE)9      FALSE      FALSE
poly(X, 10, raw = TRUE)10     FALSE      FALSE
1 subsets of each size up to 10
Selection Algorithm: forward
          poly(X, 10, raw = TRUE)1 poly(X, 10, raw = TRUE)2 poly(X, 10, raw = TRUE)3 poly(X, 10, raw = TRUE)4 poly(X, 10, raw = TRUE)5
1  ( 1 )  "*"                      " "                      " "                      " "                      " "                     
2  ( 1 )  "*"                      " "                      " "                      "*"                      " "                     
3  ( 1 )  "*"                      " "                      "*"                      "*"                      " "                     
4  ( 1 )  "*"                      "*"                      "*"                      "*"                      " "                     
5  ( 1 )  "*"                      "*"                      "*"                      "*"                      "*"                     
6  ( 1 )  "*"                      "*"                      "*"                      "*"                      "*"                     
7  ( 1 )  "*"                      "*"                      "*"                      "*"                      "*"                     
8  ( 1 )  "*"                      "*"                      "*"                      "*"                      "*"                     
9  ( 1 )  "*"                      "*"                      "*"                      "*"                      "*"                     
10  ( 1 ) "*"                      "*"                      "*"                      "*"                      "*"                     
          poly(X, 10, raw = TRUE)6 poly(X, 10, raw = TRUE)7 poly(X, 10, raw = TRUE)8 poly(X, 10, raw = TRUE)9 poly(X, 10, raw = TRUE)10
1  ( 1 )  " "                      " "                      " "                      " "                      " "                      
2  ( 1 )  " "                      " "                      " "                      " "                      " "                      
3  ( 1 )  " "                      " "                      " "                      " "                      " "                      
4  ( 1 )  " "                      " "                      " "                      " "                      " "                      
5  ( 1 )  " "                      " "                      " "                      " "                      " "                      
6  ( 1 )  " "                      " "                      " "                      "*"                      " "                      
7  ( 1 )  "*"                      " "                      " "                      "*"                      " "                      
8  ( 1 )  "*"                      "*"                      " "                      "*"                      " "                      
9  ( 1 )  "*"                      "*"                      "*"                      "*"                      " "                      
10  ( 1 ) "*"                      "*"                      "*"                      "*"                      "*"                      
```