Assignment 8

Sonali Singh

# Lab-Logistic Regression

Run Lab\_LDA01.R in R

# Knowledge Mining: Linear Discriminant Analysis  
# File: Lab\_LDA01.R  
# Theme: Linear Discriminant Analysis  
# Adapted from ISLR Chapter 4 Lab  
  
# Setup  
require(ISLR)

## Loading required package: ISLR

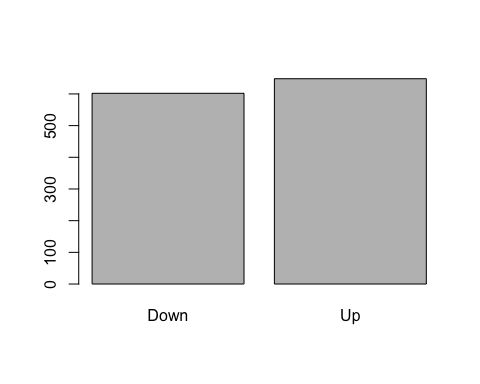
require(MASS)

## Loading required package: MASS

require(descr)

## Loading required package: descr

attach(Smarket)  
  
library(haven)  
TEDS\_2016 <- read\_stata("https://github.com/datageneration/home/blob/master/DataProgramming/data/TEDS\_2016.dta?raw=true")  
  
  
## Linear Discriminant Analysis  
freq(Direction)

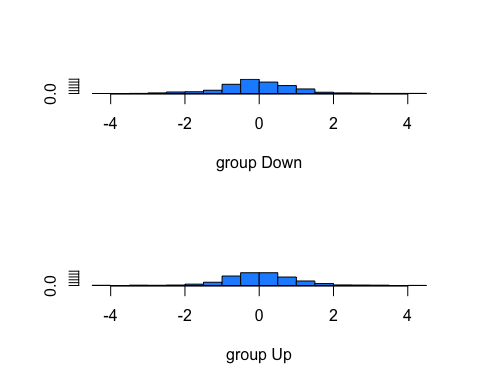


## Direction   
## Frequency Percent  
## Down 602 48.16  
## Up 648 51.84  
## Total 1250 100.00

train = Year<2005  
lda.fit=lda(Direction~Lag1+Lag2,data=Smarket, subset=Year<2005)  
lda.fit

## 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

### **LDA Results Interpretation**

1. **Data Summary**: The **Smarket** data contains daily returns of the S&P 500 index with predictors such as **Lag1** and **Lag2**. The data is split based on the **Direction** (Up or Down) of the market.
2. **Model Setup and Fitting**: You used Linear Discriminant Analysis (LDA) with predictors **Lag1** and **Lag2**, based on data from years before 2005. LDA tries to model the difference in means between two groups, weighted by the inverse of their covariance, to find a linear combination that best separates the groups.
3. **Model Outputs**:
   * **Prior Probabilities**: Indicates the prior beliefs about the probabilities of directions being ‘Up’ or ‘Down’ before seeing the data. For your model, these were approximately balanced (49.2% for Down and 50.8% for Up).
   * **Group Means**: Shows the average values of **Lag1** and **Lag2** for each market direction. The means suggest slight differences in lags between days the market went up versus days it went down.
   * **Coefficients of Linear Discriminants**: These coefficients (-0.642 for **Lag1** and -0.513 for **Lag2**) represent the linear combination of **Lag1** and **Lag2** that best separates the Up and Down classes.
4. **Model Evaluation**:
   * The confusion matrix from your LDA model on the 2005 data (**Direction.2005**) indicates the counts of true and predicted classifications:
     + True Down predicted as Down: 35
     + True Down predicted as Up: 76
     + True Up predicted as Down: 35
     + True Up predicted as Up: 106
   * **Accuracy Calculation**: The calculation **0.5595238** indicates that the model correctly predicted the market direction about 56% of the time for the year 2005.
5. **Interpretation**:
   * The LDA model performs slightly better than chance, but not by a wide margin. The relatively low predictive power may suggest that the variables **Lag1** and **Lag2** provide limited information for predicting market direction.
   * The symmetry in the confusion matrix (both wrong and right predictions are balanced across the actual classes) suggests the model does not favor one class over another, which is good for unbiased predictions.

### **Perform Model Selection Methods**

**2a. What is/are the requirement(s) of LDA?** Linear Discriminant Analysis (LDA) requires that the predictors are normally distributed within each class, and that the classes have similar covariance matrices.

# Load necessary libraries if not already installed  
packages <- c("leaps", "haven")  
install.packages(packages[!sapply(packages, requireNamespace)])

## Loading required namespace: leaps

library(leaps)  
library(haven)  
  
# Load the dataset  
TEDS\_2016 <- read\_stata("https://github.com/datageneration/home/blob/master/DataProgramming/data/TEDS\_2016.dta?raw=true")  
  
# Convert labeled variables to appropriate types  
TEDS\_2016$Age <- as.numeric(as.character(TEDS\_2016$Age))  
TEDS\_2016$Edu <- as.factor(as.character(TEDS\_2016$Edu))  
TEDS\_2016$income <- as.numeric(as.character(TEDS\_2016$income))  
  
# Remove missing values  
TEDS\_2016 <- na.omit(TEDS\_2016)  
  
# Run the regsubsets command  
subset\_fit <- regsubsets(voteblue ~ Age + Edu + income, data = TEDS\_2016, nbest = 1, method = "exhaustive")

## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =  
## force.in, : 1 linear dependencies found

## Reordering variables and trying again:

# Summarize the results  
cat("\nSummary of Best Subset Selection:\n")

##   
## Summary of Best Subset Selection:

print(summary(subset\_fit))

## Subset selection object  
## Call: regsubsets.formula(voteblue ~ Age + Edu + income, data = TEDS\_2016,   
## nbest = 1, method = "exhaustive")  
## 7 Variables (and intercept)  
## Forced in Forced out  
## Age FALSE FALSE  
## Edu2 FALSE FALSE  
## Edu3 FALSE FALSE  
## Edu4 FALSE FALSE  
## Edu5 FALSE FALSE  
## income FALSE FALSE  
## Edu9 FALSE FALSE  
## 1 subsets of each size up to 6  
## Selection Algorithm: exhaustive  
## Age Edu2 Edu3 Edu4 Edu5 Edu9 income  
## 1 ( 1 ) " " " " " " "\*" " " " " " "   
## 2 ( 1 ) " " " " "\*" "\*" " " " " " "   
## 3 ( 1 ) "\*" " " "\*" "\*" " " " " " "   
## 4 ( 1 ) "\*" " " "\*" "\*" "\*" " " " "   
## 5 ( 1 ) "\*" "\*" "\*" "\*" "\*" " " " "   
## 6 ( 1 ) "\*" "\*" "\*" "\*" "\*" " " "\*"

# Perform stepwise forward selection  
stepwise\_forward <- stepAIC(lm(voteblue ~ 1, data = TEDS\_2016), direction = "forward")

## Start: AIC=-1540.96  
## voteblue ~ 1

# Perform stepwise backward selection  
stepwise\_backward <- stepAIC(lm(voteblue ~ ., data = TEDS\_2016), direction = "backward")

## Start: AIC=-69109.69  
## voteblue ~ District + Sex + Age + Edu + Arear + Career + Career8 +   
## Ethnic + Party + PartyID + Tondu + Tondu3 + nI2 + votetsai +   
## green + votetsai\_nm + votetsai\_all + Independence + Unification +   
## sq + Taiwanese + edu + female + whitecollar + lowincome +   
## income + income\_nm + age + KMT + DPP + npp + noparty + pfp +   
## South + north + Minnan\_father + Mainland\_father + Econ\_worse +   
## Inequality + inequality5 + econworse5 + Govt\_for\_public +   
## pubwelf5 + Govt\_dont\_care + highincome + votekmt + votekmt\_nm +   
## Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense

## Warning in stepAIC(lm(voteblue ~ ., data = TEDS\_2016), direction = "backward"):  
## 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Edu 3 -3.1850e-28 1.1195e-25 -69119  
## - Career8 1 -6.8704e-28 1.1158e-25 -69118  
## - Career 1 -6.3084e-28 1.1164e-25 -69118  
## - Party 1 -5.3330e-28 1.1174e-25 -69117  
## - nI2 1 -4.5169e-28 1.1182e-25 -69116  
## - Arear 1 -1.7754e-28 1.1209e-25 -69113  
## - Age 1 -8.3210e-29 1.1219e-25 -69112  
## - Govt\_dont\_care 1 -5.2000e-30 1.1227e-25 -69112  
## - Mainland\_father 1 -4.8600e-30 1.1227e-25 -69112  
## - north 1 -4.3800e-30 1.1227e-25 -69112  
## - lowincome 1 -6.2000e-31 1.1227e-25 -69112  
## - South 1 2.0000e-31 1.1227e-25 -69112  
## - Inequality 1 2.4000e-31 1.1227e-25 -69112  
## - Ethnic 1 9.8000e-30 1.1228e-25 -69112  
## - Minnan\_father 1 2.7300e-29 1.1230e-25 -69111  
## - econworse5 1 3.5510e-29 1.1231e-25 -69111  
## - Econ\_worse 1 5.0930e-29 1.1232e-25 -69111  
## - highincome 1 6.6500e-29 1.1234e-25 -69111  
## - inequality5 1 8.7030e-29 1.1236e-25 -69111  
## - whitecollar 1 1.0917e-28 1.1238e-25 -69111  
## - age 1 1.7541e-28 1.1245e-25 -69110  
## - Govt\_for\_public 1 2.0557e-28 1.1248e-25 -69110  
## <none> 1.1227e-25 -69110  
## - Tondu 1 2.9935e-28 1.1257e-25 -69109  
## - pubwelf5 1 4.6463e-28 1.1274e-25 -69107  
## - Taiwanese 1 5.3068e-28 1.1280e-25 -69107  
## - District 1 2.6759e-27 1.1495e-25 -69086  
##   
## Step: AIC=-69119.03  
## voteblue ~ District + Sex + Age + Arear + Career + Career8 +   
## Ethnic + Party + PartyID + Tondu + Tondu3 + nI2 + votetsai +   
## green + votetsai\_nm + votetsai\_all + Independence + Unification +   
## sq + Taiwanese + edu + female + whitecollar + lowincome +   
## income + income\_nm + age + KMT + DPP + npp + noparty + pfp +   
## South + north + Minnan\_father + Mainland\_father + Econ\_worse +   
## Inequality + inequality5 + econworse5 + Govt\_for\_public +   
## pubwelf5 + Govt\_dont\_care + highincome + votekmt + votekmt\_nm +   
## Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Career 1 -4.1351e-28 1.1151e-25 -69125  
## - Career8 1 -2.4522e-28 1.1168e-25 -69123  
## - Party 1 -2.0588e-28 1.1172e-25 -69123  
## - nI2 1 -1.9613e-28 1.1173e-25 -69123  
## - Age 1 -1.8974e-28 1.1173e-25 -69123  
## - Govt\_dont\_care 1 2.2440e-29 1.1195e-25 -69121  
## - Mainland\_father 1 2.3270e-29 1.1195e-25 -69121  
## - north 1 2.5000e-29 1.1195e-25 -69121  
## - Inequality 1 2.7300e-29 1.1195e-25 -69121  
## - South 1 2.8190e-29 1.1195e-25 -69121  
## - lowincome 1 2.9090e-29 1.1195e-25 -69121  
## - econworse5 1 6.5830e-29 1.1199e-25 -69120  
## - Minnan\_father 1 6.7600e-29 1.1199e-25 -69120  
## - highincome 1 7.5360e-29 1.1200e-25 -69120  
## - Econ\_worse 1 8.8860e-29 1.1201e-25 -69120  
## - edu 1 9.7780e-29 1.1202e-25 -69120  
## - inequality5 1 9.9540e-29 1.1202e-25 -69120  
## - whitecollar 1 1.2866e-28 1.1205e-25 -69120  
## - Ethnic 1 1.5609e-28 1.1208e-25 -69120  
## - age 1 1.7628e-28 1.1210e-25 -69119  
## - Govt\_for\_public 1 2.0169e-28 1.1213e-25 -69119  
## <none> 1.1192e-25 -69119  
## - pubwelf5 1 4.3387e-28 1.1236e-25 -69117  
## - Taiwanese 1 5.2647e-28 1.1245e-25 -69116  
## - Arear 1 7.9051e-28 1.1271e-25 -69113  
## - Tondu 1 7.9430e-28 1.1272e-25 -69113  
## - District 1 1.8212e-27 1.1374e-25 -69104  
##   
## Step: AIC=-69125.43  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## Party + PartyID + Tondu + Tondu3 + nI2 + votetsai + green +   
## votetsai\_nm + votetsai\_all + Independence + Unification +   
## sq + Taiwanese + edu + female + whitecollar + lowincome +   
## income + income\_nm + age + KMT + DPP + npp + noparty + pfp +   
## South + north + Minnan\_father + Mainland\_father + Econ\_worse +   
## Inequality + inequality5 + econworse5 + Govt\_for\_public +   
## pubwelf5 + Govt\_dont\_care + highincome + votekmt + votekmt\_nm +   
## Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - whitecollar 1 -5.4590e-29 1.1141e-25 -69128  
## - Govt\_dont\_care 1 -5.4560e-29 1.1141e-25 -69128  
## - Mainland\_father 1 -5.4110e-29 1.1141e-25 -69128  
## - lowincome 1 -4.7620e-29 1.1142e-25 -69128  
## - north 1 -4.6740e-29 1.1142e-25 -69128  
## - Inequality 1 -4.0510e-29 1.1143e-25 -69128  
## - South 1 -3.9600e-29 1.1143e-25 -69128  
## - Party 1 -3.3620e-29 1.1143e-25 -69128  
## - econworse5 1 -1.8420e-29 1.1145e-25 -69128  
## - Minnan\_father 1 -1.3600e-29 1.1145e-25 -69128  
## - Econ\_worse 1 3.2500e-30 1.1147e-25 -69127  
## - highincome 1 5.8800e-30 1.1147e-25 -69127  
## - nI2 1 1.0260e-29 1.1148e-25 -69127  
## - edu 1 1.8740e-29 1.1148e-25 -69127  
## - Tondu 1 2.9070e-29 1.1149e-25 -69127  
## - inequality5 1 3.2440e-29 1.1150e-25 -69127  
## - age 1 9.3090e-29 1.1156e-25 -69127  
## - Govt\_for\_public 1 1.3274e-28 1.1160e-25 -69126  
## <none> 1.1147e-25 -69125  
## - Age 1 2.8796e-28 1.1175e-25 -69125  
## - pubwelf5 1 3.5127e-28 1.1182e-25 -69124  
## - Career8 1 3.7160e-28 1.1184e-25 -69124  
## - Taiwanese 1 4.7562e-28 1.1194e-25 -69123  
## - Ethnic 1 1.0281e-27 1.1249e-25 -69118  
## - Arear 1 2.1545e-27 1.1362e-25 -69107  
## - District 1 2.7875e-27 1.1425e-25 -69101  
##   
## Step: AIC=-69127.43  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## Party + PartyID + Tondu + Tondu3 + nI2 + votetsai + green +   
## votetsai\_nm + votetsai\_all + Independence + Unification +   
## sq + Taiwanese + edu + female + lowincome + income + income\_nm +   
## age + KMT + DPP + npp + noparty + pfp + South + north + Minnan\_father +   
## Mainland\_father + Econ\_worse + Inequality + inequality5 +   
## econworse5 + Govt\_for\_public + pubwelf5 + Govt\_dont\_care +   
## highincome + votekmt + votekmt\_nm + Blue + Green + No\_Party +   
## voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Govt\_dont\_care 1 -4.7210e-29 1.1142e-25 -69130  
## - Mainland\_father 1 -4.6780e-29 1.1142e-25 -69130  
## - lowincome 1 -4.0260e-29 1.1143e-25 -69130  
## - north 1 -3.9400e-29 1.1143e-25 -69130  
## - Inequality 1 -3.3130e-29 1.1143e-25 -69130  
## - South 1 -3.2250e-29 1.1143e-25 -69130  
## - Party 1 -2.6470e-29 1.1144e-25 -69130  
## - econworse5 1 -1.1020e-29 1.1145e-25 -69130  
## - Minnan\_father 1 -6.3500e-30 1.1146e-25 -69129  
## - Econ\_worse 1 1.0560e-29 1.1148e-25 -69129  
## - highincome 1 1.3250e-29 1.1148e-25 -69129  
## - nI2 1 1.7670e-29 1.1148e-25 -69129  
## - edu 1 3.4080e-29 1.1150e-25 -69129  
## - Tondu 1 3.8070e-29 1.1150e-25 -69129  
## - inequality5 1 3.9730e-29 1.1151e-25 -69129  
## - age 1 1.0070e-28 1.1157e-25 -69128  
## - Govt\_for\_public 1 1.4040e-28 1.1161e-25 -69128  
## <none> 1.1147e-25 -69127  
## - Career8 1 2.4694e-28 1.1171e-25 -69127  
## - Age 1 2.9732e-28 1.1176e-25 -69127  
## - pubwelf5 1 3.5973e-28 1.1183e-25 -69126  
## - Taiwanese 1 4.8380e-28 1.1195e-25 -69125  
## - Ethnic 1 1.0381e-27 1.1250e-25 -69119  
## - Arear 1 2.1641e-27 1.1363e-25 -69109  
## - District 1 2.8046e-27 1.1427e-25 -69103  
##   
## Step: AIC=-69129.43  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## Party + PartyID + Tondu + Tondu3 + nI2 + votetsai + green +   
## votetsai\_nm + votetsai\_all + Independence + Unification +   
## sq + Taiwanese + edu + female + lowincome + income + income\_nm +   
## age + KMT + DPP + npp + noparty + pfp + South + north + Minnan\_father +   
## Mainland\_father + Econ\_worse + Inequality + inequality5 +   
## econworse5 + Govt\_for\_public + pubwelf5 + highincome + votekmt +   
## votekmt\_nm + Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 +   
## votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Mainland\_father 1 -4.5150e-29 1.1142e-25 -69132  
## - lowincome 1 -3.8580e-29 1.1143e-25 -69132  
## - north 1 -3.7740e-29 1.1143e-25 -69132  
## - Inequality 1 -3.1480e-29 1.1143e-25 -69132  
## - South 1 -3.0600e-29 1.1144e-25 -69132  
## - Party 1 -2.4900e-29 1.1144e-25 -69132  
## - econworse5 1 -9.2100e-30 1.1146e-25 -69132  
## - Minnan\_father 1 -4.5500e-30 1.1146e-25 -69131  
## - Econ\_worse 1 1.2590e-29 1.1148e-25 -69131  
## - highincome 1 1.5030e-29 1.1148e-25 -69131  
## - nI2 1 1.9840e-29 1.1149e-25 -69131  
## - edu 1 3.6550e-29 1.1150e-25 -69131  
## - Tondu 1 4.0700e-29 1.1151e-25 -69131  
## - inequality5 1 4.1340e-29 1.1151e-25 -69131  
## - age 1 1.0250e-28 1.1157e-25 -69130  
## - Govt\_for\_public 1 1.4361e-28 1.1161e-25 -69130  
## <none> 1.1147e-25 -69129  
## - Career8 1 2.4967e-28 1.1172e-25 -69129  
## - Age 1 2.9969e-28 1.1177e-25 -69129  
## - pubwelf5 1 3.7804e-28 1.1184e-25 -69128  
## - Taiwanese 1 4.8557e-28 1.1195e-25 -69127  
## - Ethnic 1 1.0437e-27 1.1251e-25 -69121  
## - Arear 1 2.1657e-27 1.1363e-25 -69111  
## - District 1 2.8067e-27 1.1427e-25 -69105  
##   
## Step: AIC=-69131.42  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## Party + PartyID + Tondu + Tondu3 + nI2 + votetsai + green +   
## votetsai\_nm + votetsai\_all + Independence + Unification +   
## sq + Taiwanese + edu + female + lowincome + income + income\_nm +   
## age + KMT + DPP + npp + noparty + pfp + South + north + Minnan\_father +   
## Econ\_worse + Inequality + inequality5 + econworse5 + Govt\_for\_public +   
## pubwelf5 + highincome + votekmt + votekmt\_nm + Blue + Green +   
## No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - lowincome 1 -5.2330e-29 1.1141e-25 -69134  
## - north 1 -5.0180e-29 1.1142e-25 -69134  
## - Inequality 1 -4.4190e-29 1.1142e-25 -69134  
## - South 1 -4.2870e-29 1.1142e-25 -69134  
## - Party 1 -2.4840e-29 1.1144e-25 -69134  
## - econworse5 1 -2.1360e-29 1.1144e-25 -69134  
## - Minnan\_father 1 -5.6200e-30 1.1146e-25 -69133  
## - Econ\_worse 1 3.4000e-31 1.1147e-25 -69133  
## - highincome 1 3.6800e-30 1.1147e-25 -69133  
## - inequality5 1 2.8970e-29 1.1150e-25 -69133  
## - edu 1 2.9710e-29 1.1150e-25 -69133  
## - age 1 8.8220e-29 1.1155e-25 -69133  
## - nI2 1 9.0870e-29 1.1156e-25 -69133  
## - Tondu 1 1.0154e-28 1.1157e-25 -69132  
## - Govt\_for\_public 1 1.3031e-28 1.1160e-25 -69132  
## <none> 1.1147e-25 -69131  
## - Age 1 2.7183e-28 1.1174e-25 -69131  
## - Career8 1 3.6286e-28 1.1183e-25 -69130  
## - pubwelf5 1 3.6495e-28 1.1183e-25 -69130  
## - Taiwanese 1 4.7536e-28 1.1194e-25 -69129  
## - Ethnic 1 1.0277e-27 1.1249e-25 -69124  
## - Arear 1 2.1570e-27 1.1362e-25 -69113  
## - District 1 2.8034e-27 1.1427e-25 -69107  
##   
## Step: AIC=-69133.36  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## Party + PartyID + Tondu + Tondu3 + nI2 + votetsai + green +   
## votetsai\_nm + votetsai\_all + Independence + Unification +   
## sq + Taiwanese + edu + female + income + income\_nm + age +   
## KMT + DPP + npp + noparty + pfp + South + north + Minnan\_father +   
## Econ\_worse + Inequality + inequality5 + econworse5 + Govt\_for\_public +   
## pubwelf5 + highincome + votekmt + votekmt\_nm + Blue + Green +   
## No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Party 1 -7.0600e-29 1.1140e-25 -69136  
## - north 1 -6.3600e-30 1.1147e-25 -69135  
## - South 1 1.0900e-30 1.1147e-25 -69135  
## - Inequality 1 2.3900e-30 1.1148e-25 -69135  
## - econworse5 1 2.0820e-29 1.1149e-25 -69135  
## - nI2 1 2.1850e-29 1.1149e-25 -69135  
## - Tondu 1 2.8950e-29 1.1150e-25 -69135  
## - Minnan\_father 1 3.9560e-29 1.1151e-25 -69135  
## - Econ\_worse 1 4.3580e-29 1.1152e-25 -69135  
## - highincome 1 4.5730e-29 1.1152e-25 -69135  
## - inequality5 1 6.7110e-29 1.1154e-25 -69135  
## - edu 1 7.3730e-29 1.1155e-25 -69135  
## - age 1 1.3818e-28 1.1161e-25 -69134  
## - Govt\_for\_public 1 1.7989e-28 1.1165e-25 -69134  
## <none> 1.1147e-25 -69133  
## - Career8 1 2.3893e-28 1.1171e-25 -69133  
## - Age 1 3.0624e-28 1.1178e-25 -69132  
## - pubwelf5 1 4.2137e-28 1.1189e-25 -69131  
## - Taiwanese 1 5.3027e-28 1.1200e-25 -69130  
## - Ethnic 1 9.2312e-28 1.1240e-25 -69127  
## - Arear 1 2.1804e-27 1.1365e-25 -69115  
## - District 1 2.8555e-27 1.1433e-25 -69108  
##   
## Step: AIC=-69134.77  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## PartyID + Tondu + Tondu3 + nI2 + votetsai + green + votetsai\_nm +   
## votetsai\_all + Independence + Unification + sq + Taiwanese +   
## edu + female + income + income\_nm + age + KMT + DPP + npp +   
## noparty + pfp + South + north + Minnan\_father + Econ\_worse +   
## Inequality + inequality5 + econworse5 + Govt\_for\_public +   
## pubwelf5 + highincome + votekmt + votekmt\_nm + Blue + Green +   
## No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Inequality 1 -5.5370e-29 1.1148e-25 -69137  
## - north 1 -5.1800e-29 1.1148e-25 -69137  
## - South 1 -4.5480e-29 1.1149e-25 -69137  
## - Tondu 1 -3.9080e-29 1.1150e-25 -69137  
## - econworse5 1 -2.5640e-29 1.1151e-25 -69137  
## - Minnan\_father 1 -9.5900e-30 1.1153e-25 -69137  
## - Econ\_worse 1 -4.2000e-31 1.1153e-25 -69137  
## - highincome 1 2.6000e-30 1.1154e-25 -69137  
## - edu 1 1.3560e-29 1.1155e-25 -69137  
## - inequality5 1 1.7570e-29 1.1155e-25 -69137  
## - Arear 1 8.0510e-29 1.1162e-25 -69136  
## - age 1 9.4680e-29 1.1163e-25 -69136  
## - Govt\_for\_public 1 1.3638e-28 1.1167e-25 -69135  
## - Ethnic 1 1.5579e-28 1.1169e-25 -69135  
## - nI2 1 1.9928e-28 1.1173e-25 -69135  
## <none> 1.1153e-25 -69135  
## - pubwelf5 1 3.8744e-28 1.1192e-25 -69133  
## - Career8 1 4.6111e-28 1.1200e-25 -69132  
## - Taiwanese 1 4.7258e-28 1.1201e-25 -69132  
## - Age 1 8.0175e-28 1.1234e-25 -69129  
## - District 1 2.9751e-27 1.1451e-25 -69108  
##   
## Step: AIC=-69136.73  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## PartyID + Tondu + Tondu3 + nI2 + votetsai + green + votetsai\_nm +   
## votetsai\_all + Independence + Unification + sq + Taiwanese +   
## edu + female + income + income\_nm + age + KMT + DPP + npp +   
## noparty + pfp + South + north + Minnan\_father + Econ\_worse +   
## inequality5 + econworse5 + Govt\_for\_public + pubwelf5 + highincome +   
## votekmt + votekmt\_nm + Blue + Green + No\_Party + voteblue\_nm +   
## votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - north 1 -5.3190e-29 1.1149e-25 -69139  
## - South 1 -4.7020e-29 1.1149e-25 -69139  
## - Tondu 1 -3.2010e-29 1.1151e-25 -69139  
## - econworse5 1 -2.6520e-29 1.1151e-25 -69139  
## - Minnan\_father 1 -1.2390e-29 1.1153e-25 -69139  
## - Econ\_worse 1 -1.6500e-30 1.1154e-25 -69139  
## - highincome 1 1.0000e-30 1.1154e-25 -69139  
## - edu 1 8.3200e-30 1.1155e-25 -69139  
## - inequality5 1 5.9910e-29 1.1160e-25 -69138  
## - age 1 9.1380e-29 1.1163e-25 -69138  
## - Govt\_for\_public 1 1.4049e-28 1.1168e-25 -69137  
## - Arear 1 1.6336e-28 1.1170e-25 -69137  
## - nI2 1 1.9000e-28 1.1173e-25 -69137  
## <none> 1.1154e-25 -69137  
## - Ethnic 1 2.0942e-28 1.1175e-25 -69137  
## - pubwelf5 1 3.9114e-28 1.1193e-25 -69135  
## - Taiwanese 1 4.7337e-28 1.1201e-25 -69134  
## - Career8 1 4.8953e-28 1.1203e-25 -69134  
## - Age 1 9.1496e-28 1.1245e-25 -69130  
## - District 1 2.9819e-27 1.1452e-25 -69110  
##   
## Step: AIC=-69138.66  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## PartyID + Tondu + Tondu3 + nI2 + votetsai + green + votetsai\_nm +   
## votetsai\_all + Independence + Unification + sq + Taiwanese +   
## edu + female + income + income\_nm + age + KMT + DPP + npp +   
## noparty + pfp + South + Minnan\_father + Econ\_worse + inequality5 +   
## econworse5 + Govt\_for\_public + pubwelf5 + highincome + votekmt +   
## votekmt\_nm + Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 +   
## votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Tondu 1 1.5290e-29 1.1156e-25 -69141  
## - South 1 5.0440e-29 1.1160e-25 -69140  
## - econworse5 1 7.4860e-29 1.1162e-25 -69140  
## - Minnan\_father 1 8.6360e-29 1.1163e-25 -69140  
## - Econ\_worse 1 1.0087e-28 1.1165e-25 -69140  
## - highincome 1 1.0295e-28 1.1165e-25 -69140  
## - edu 1 1.1345e-28 1.1166e-25 -69140  
## - inequality5 1 1.6460e-28 1.1171e-25 -69139  
## - age 1 1.9397e-28 1.1174e-25 -69139  
## <none> 1.1155e-25 -69139  
## - Govt\_for\_public 1 2.4413e-28 1.1179e-25 -69138  
## - Ethnic 1 2.5598e-28 1.1180e-25 -69138  
## - Arear 1 2.6317e-28 1.1181e-25 -69138  
## - nI2 1 2.8374e-28 1.1183e-25 -69138  
## - pubwelf5 1 4.9369e-28 1.1204e-25 -69136  
## - Taiwanese 1 5.6547e-28 1.1211e-25 -69135  
## - Career8 1 5.9124e-28 1.1214e-25 -69135  
## - Age 1 9.5672e-28 1.1250e-25 -69131  
## - District 1 2.8967e-27 1.1444e-25 -69113  
##   
## Step: AIC=-69140.69  
## voteblue ~ District + Sex + Age + Arear + Career8 + Ethnic +   
## PartyID + Tondu3 + nI2 + votetsai + green + votetsai\_nm +   
## votetsai\_all + Independence + Unification + sq + Taiwanese +   
## edu + female + income + income\_nm + age + KMT + DPP + npp +   
## noparty + pfp + South + Minnan\_father + Econ\_worse + inequality5 +   
## econworse5 + Govt\_for\_public + pubwelf5 + highincome + votekmt +   
## votekmt\_nm + Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 +   
## votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Career8 1 -1.0172e-28 1.1144e-25 -69144  
## - South 1 -5.1960e-29 1.1149e-25 -69143  
## - econworse5 1 -3.0680e-29 1.1151e-25 -69143  
## - Minnan\_father 1 -1.3540e-29 1.1153e-25 -69143  
## - Econ\_worse 1 -8.8200e-30 1.1153e-25 -69143  
## - Arear 1 -7.8300e-30 1.1153e-25 -69143  
## - highincome 1 -4.6000e-31 1.1154e-25 -69143  
## - Ethnic 1 5.2600e-30 1.1155e-25 -69143  
## - edu 1 2.6280e-29 1.1157e-25 -69142  
## - inequality5 1 4.3610e-29 1.1159e-25 -69142  
## - age 1 9.5990e-29 1.1164e-25 -69142  
## - Govt\_for\_public 1 1.4188e-28 1.1168e-25 -69141  
## <none> 1.1154e-25 -69141  
## - nI2 1 3.0498e-28 1.1185e-25 -69140  
## - pubwelf5 1 3.8756e-28 1.1193e-25 -69139  
## - Taiwanese 1 4.7652e-28 1.1202e-25 -69138  
## - Age 1 6.1009e-28 1.1215e-25 -69137  
## - District 1 2.2967e-27 1.1384e-25 -69121  
##   
## Step: AIC=-69142.56  
## voteblue ~ District + Sex + Age + Arear + Ethnic + PartyID +   
## Tondu3 + nI2 + votetsai + green + votetsai\_nm + votetsai\_all +   
## Independence + Unification + sq + Taiwanese + edu + female +   
## income + income\_nm + age + KMT + DPP + npp + noparty + pfp +   
## South + Minnan\_father + Econ\_worse + inequality5 + econworse5 +   
## Govt\_for\_public + pubwelf5 + highincome + votekmt + votekmt\_nm +   
## Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - South 1 -1.0120e-28 1.1145e-25 -69146  
## - econworse5 1 -7.2860e-29 1.1148e-25 -69145  
## - Minnan\_father 1 -6.4740e-29 1.1149e-25 -69145  
## - Econ\_worse 1 -4.8790e-29 1.1151e-25 -69145  
## - highincome 1 -4.1570e-29 1.1152e-25 -69145  
## - edu 1 -1.5010e-29 1.1154e-25 -69145  
## - Age 1 -8.2000e-31 1.1156e-25 -69145  
## - inequality5 1 2.0000e-30 1.1156e-25 -69145  
## - nI2 1 1.1870e-29 1.1157e-25 -69144  
## - age 1 4.5210e-29 1.1160e-25 -69144  
## - Arear 1 8.4110e-29 1.1164e-25 -69144  
## - Govt\_for\_public 1 8.9670e-29 1.1165e-25 -69144  
## <none> 1.1156e-25 -69143  
## - pubwelf5 1 3.3636e-28 1.1189e-25 -69141  
## - Taiwanese 1 4.1312e-28 1.1197e-25 -69141  
## - Ethnic 1 1.8654e-27 1.1342e-25 -69127  
## - District 1 2.0777e-27 1.1363e-25 -69125  
##   
## Step: AIC=-69144.47  
## voteblue ~ District + Sex + Age + Arear + Ethnic + PartyID +   
## Tondu3 + nI2 + votetsai + green + votetsai\_nm + votetsai\_all +   
## Independence + Unification + sq + Taiwanese + edu + female +   
## income + income\_nm + age + KMT + DPP + npp + noparty + pfp +   
## Minnan\_father + Econ\_worse + inequality5 + econworse5 + Govt\_for\_public +   
## pubwelf5 + highincome + votekmt + votekmt\_nm + Blue + Green +   
## No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - econworse5 1 -8.5260e-29 1.1148e-25 -69147  
## - Minnan\_father 1 -6.2150e-29 1.1150e-25 -69147  
## - Econ\_worse 1 -6.1520e-29 1.1150e-25 -69147  
## - highincome 1 -5.3820e-29 1.1151e-25 -69147  
## - Age 1 -3.9160e-29 1.1153e-25 -69147  
## - edu 1 -2.4190e-29 1.1154e-25 -69147  
## - inequality5 1 -7.4900e-30 1.1156e-25 -69147  
## - nI2 1 1.0400e-30 1.1157e-25 -69146  
## - age 1 3.6810e-29 1.1160e-25 -69146  
## - Govt\_for\_public 1 7.3010e-29 1.1164e-25 -69146  
## - Arear 1 1.3973e-28 1.1170e-25 -69145  
## <none> 1.1157e-25 -69144  
## - pubwelf5 1 3.1398e-28 1.1188e-25 -69143  
## - Taiwanese 1 4.0107e-28 1.1197e-25 -69143  
## - Ethnic 1 1.9633e-27 1.1353e-25 -69128  
## - District 1 2.1153e-27 1.1368e-25 -69126  
##   
## Step: AIC=-69146.13  
## voteblue ~ District + Sex + Age + Arear + Ethnic + PartyID +   
## Tondu3 + nI2 + votetsai + green + votetsai\_nm + votetsai\_all +   
## Independence + Unification + sq + Taiwanese + edu + female +   
## income + income\_nm + age + KMT + DPP + npp + noparty + pfp +   
## Minnan\_father + Econ\_worse + inequality5 + Govt\_for\_public +   
## pubwelf5 + highincome + votekmt + votekmt\_nm + Blue + Green +   
## No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Econ\_worse 1 -6.7910e-29 1.1153e-25 -69149  
## - highincome 1 -2.6750e-29 1.1158e-25 -69148  
## - Minnan\_father 1 -2.5290e-29 1.1158e-25 -69148  
## - edu 1 4.0900e-30 1.1161e-25 -69148  
## - Age 1 1.7610e-29 1.1162e-25 -69148  
## - inequality5 1 3.2250e-29 1.1163e-25 -69148  
## - nI2 1 4.6250e-29 1.1165e-25 -69148  
## - age 1 6.0860e-29 1.1166e-25 -69148  
## - Govt\_for\_public 1 9.5510e-29 1.1170e-25 -69147  
## - Arear 1 1.3720e-28 1.1174e-25 -69147  
## <none> 1.1160e-25 -69146  
## - pubwelf5 1 3.2706e-28 1.1193e-25 -69145  
## - Taiwanese 1 4.3388e-28 1.1204e-25 -69144  
## - Ethnic 1 2.0161e-27 1.1362e-25 -69129  
## - District 1 2.0763e-27 1.1368e-25 -69128  
##   
## Step: AIC=-69147.89  
## voteblue ~ District + Sex + Age + Arear + Ethnic + PartyID +   
## Tondu3 + nI2 + votetsai + green + votetsai\_nm + votetsai\_all +   
## Independence + Unification + sq + Taiwanese + edu + female +   
## income + income\_nm + age + KMT + DPP + npp + noparty + pfp +   
## Minnan\_father + inequality5 + Govt\_for\_public + pubwelf5 +   
## highincome + votekmt + votekmt\_nm + Blue + Green + No\_Party +   
## voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Age 1 -5.1520e-29 1.1158e-25 -69150  
## - highincome 1 -5.0140e-29 1.1158e-25 -69150  
## - Minnan\_father 1 -4.9620e-29 1.1158e-25 -69150  
## - edu 1 -2.4410e-29 1.1160e-25 -69150  
## - inequality5 1 -2.3900e-30 1.1162e-25 -69150  
## - nI2 1 -6.5000e-31 1.1163e-25 -69150  
## - age 1 4.2170e-29 1.1167e-25 -69149  
## - Govt\_for\_public 1 7.4550e-29 1.1170e-25 -69149  
## - Arear 1 9.0530e-29 1.1172e-25 -69149  
## <none> 1.1163e-25 -69148  
## - pubwelf5 1 3.2388e-28 1.1195e-25 -69147  
## - Taiwanese 1 4.2204e-28 1.1205e-25 -69146  
## - Ethnic 1 1.8508e-27 1.1348e-25 -69132  
## - District 1 2.1063e-27 1.1373e-25 -69130  
##   
## Step: AIC=-69149.59  
## voteblue ~ District + Sex + Arear + Ethnic + PartyID + Tondu3 +   
## nI2 + votetsai + green + votetsai\_nm + votetsai\_all + Independence +   
## Unification + sq + Taiwanese + edu + female + income + income\_nm +   
## age + KMT + DPP + npp + noparty + pfp + Minnan\_father + inequality5 +   
## Govt\_for\_public + pubwelf5 + highincome + votekmt + votekmt\_nm +   
## Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - highincome 1 -1.3800e-29 1.1164e-25 -69152  
## - age 1 1.3300e-29 1.1167e-25 -69151  
## - inequality5 1 1.4200e-29 1.1167e-25 -69151  
## - Minnan\_father 1 2.0200e-29 1.1168e-25 -69151  
## - edu 1 6.3300e-29 1.1172e-25 -69151  
## - Govt\_for\_public 1 1.5460e-28 1.1181e-25 -69150  
## - nI2 1 1.9490e-28 1.1185e-25 -69150  
## <none> 1.1166e-25 -69150  
## - Arear 1 3.4730e-28 1.1200e-25 -69148  
## - pubwelf5 1 4.0800e-28 1.1206e-25 -69148  
## - Taiwanese 1 4.7950e-28 1.1214e-25 -69147  
## - Ethnic 1 7.6490e-28 1.1242e-25 -69144  
## - District 1 4.5682e-27 1.1623e-25 -69109  
##   
## Step: AIC=-69151.08  
## voteblue ~ District + Sex + Arear + Ethnic + PartyID + Tondu3 +   
## nI2 + votetsai + green + votetsai\_nm + votetsai\_all + Independence +   
## Unification + sq + Taiwanese + edu + female + income + income\_nm +   
## age + KMT + DPP + npp + noparty + pfp + Minnan\_father + inequality5 +   
## Govt\_for\_public + pubwelf5 + votekmt + votekmt\_nm + Blue +   
## Green + No\_Party + voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - inequality5 1 8.5900e-29 1.1180e-25 -69152  
## - age 1 8.7200e-29 1.1180e-25 -69152  
## - Minnan\_father 1 9.9400e-29 1.1181e-25 -69152  
## - edu 1 1.4250e-28 1.1185e-25 -69152  
## <none> 1.1171e-25 -69151  
## - nI2 1 2.1850e-28 1.1193e-25 -69151  
## - Govt\_for\_public 1 2.3520e-28 1.1194e-25 -69151  
## - Arear 1 3.2490e-28 1.1204e-25 -69150  
## - pubwelf5 1 4.8910e-28 1.1220e-25 -69148  
## - Taiwanese 1 5.5680e-28 1.1227e-25 -69148  
## - Ethnic 1 6.4510e-28 1.1235e-25 -69147  
## - District 1 4.6721e-27 1.1638e-25 -69109  
##   
## Step: AIC=-69152.32  
## voteblue ~ District + Sex + Arear + Ethnic + PartyID + Tondu3 +   
## nI2 + votetsai + green + votetsai\_nm + votetsai\_all + Independence +   
## Unification + sq + Taiwanese + edu + female + income + income\_nm +   
## age + KMT + DPP + npp + noparty + pfp + Minnan\_father + Govt\_for\_public +   
## pubwelf5 + votekmt + votekmt\_nm + Blue + Green + No\_Party +   
## voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - age 1 -3.4000e-30 1.1178e-25 -69154  
## - Minnan\_father 1 2.2200e-29 1.1181e-25 -69154  
## - edu 1 3.7200e-29 1.1183e-25 -69154  
## - Govt\_for\_public 1 1.2840e-28 1.1192e-25 -69153  
## - nI2 1 1.7360e-28 1.1196e-25 -69153  
## <none> 1.1179e-25 -69152  
## - pubwelf5 1 3.6370e-28 1.1215e-25 -69151  
## - Arear 1 3.7610e-28 1.1216e-25 -69151  
## - Taiwanese 1 4.4890e-28 1.1224e-25 -69150  
## - Ethnic 1 6.5060e-28 1.1244e-25 -69148  
## - District 1 4.6606e-27 1.1645e-25 -69110  
##   
## Step: AIC=-69153.56  
## voteblue ~ District + Sex + Arear + Ethnic + PartyID + Tondu3 +   
## nI2 + votetsai + green + votetsai\_nm + votetsai\_all + Independence +   
## Unification + sq + Taiwanese + edu + female + income + income\_nm +   
## KMT + DPP + npp + noparty + pfp + Minnan\_father + Govt\_for\_public +   
## pubwelf5 + votekmt + votekmt\_nm + Blue + Green + No\_Party +   
## voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - edu 1 3.4500e-29 1.1190e-25 -69155  
## - Minnan\_father 1 8.0700e-29 1.1195e-25 -69155  
## - nI2 1 1.4610e-28 1.1201e-25 -69154  
## <none> 1.1187e-25 -69154  
## - Govt\_for\_public 1 2.1070e-28 1.1208e-25 -69154  
## - Arear 1 3.9260e-28 1.1226e-25 -69152  
## - pubwelf5 1 4.4390e-28 1.1231e-25 -69151  
## - Taiwanese 1 5.6260e-28 1.1243e-25 -69150  
## - Ethnic 1 7.9430e-28 1.1266e-25 -69148  
## - District 1 4.6788e-27 1.1655e-25 -69112  
##   
## Step: AIC=-69155.06  
## voteblue ~ District + Sex + Arear + Ethnic + PartyID + Tondu3 +   
## nI2 + votetsai + green + votetsai\_nm + votetsai\_all + Independence +   
## Unification + sq + Taiwanese + female + income + income\_nm +   
## KMT + DPP + npp + noparty + pfp + Minnan\_father + Govt\_for\_public +   
## pubwelf5 + votekmt + votekmt\_nm + Blue + Green + No\_Party +   
## voteblue\_nm + votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Minnan\_father 1 8.350e-29 1.1200e-25 -69156  
## - nI2 1 1.236e-28 1.1204e-25 -69156  
## - Govt\_for\_public 1 2.003e-28 1.1212e-25 -69155  
## <none> 1.1192e-25 -69155  
## - Arear 1 3.762e-28 1.1230e-25 -69153  
## - pubwelf5 1 4.477e-28 1.1237e-25 -69153  
## - Taiwanese 1 5.629e-28 1.1248e-25 -69152  
## - Ethnic 1 7.635e-28 1.1268e-25 -69150  
## - District 1 4.668e-27 1.1659e-25 -69113  
##   
## Step: AIC=-69156.06  
## voteblue ~ District + Sex + Arear + Ethnic + PartyID + Tondu3 +   
## nI2 + votetsai + green + votetsai\_nm + votetsai\_all + Independence +   
## Unification + sq + Taiwanese + female + income + income\_nm +   
## KMT + DPP + npp + noparty + pfp + Govt\_for\_public + pubwelf5 +   
## votekmt + votekmt\_nm + Blue + Green + No\_Party + voteblue\_nm +   
## votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - nI2 1 1.3060e-28 1.1216e-25 -69157  
## <none> 1.1203e-25 -69156  
## - Govt\_for\_public 1 2.6710e-28 1.1229e-25 -69155  
## - Arear 1 3.1930e-28 1.1234e-25 -69155  
## - pubwelf5 1 5.1430e-28 1.1254e-25 -69153  
## - Taiwanese 1 6.1600e-28 1.1264e-25 -69152  
## - Ethnic 1 6.6710e-28 1.1269e-25 -69152  
## - District 1 4.4705e-27 1.1650e-25 -69116  
##   
## Step: AIC=-69156.87  
## voteblue ~ District + Sex + Arear + Ethnic + PartyID + Tondu3 +   
## votetsai + green + votetsai\_nm + votetsai\_all + Independence +   
## Unification + sq + Taiwanese + female + income + income\_nm +   
## KMT + DPP + npp + noparty + pfp + Govt\_for\_public + pubwelf5 +   
## votekmt + votekmt\_nm + Blue + Green + No\_Party + voteblue\_nm +   
## votedpp\_1 + votekmt\_1

## Warning: attempting model selection on an essentially perfect fit is nonsense  
  
## Warning: 0 df terms are changing AIC

## Df Sum of Sq RSS AIC  
## - Ethnic 1 1.5720e-28 1.1231e-25 -69157  
## <none> 1.1215e-25 -69157  
## - Arear 1 2.2960e-28 1.1238e-25 -69157  
## - Govt\_for\_public 1 2.2980e-28 1.1238e-25 -69157  
## - pubwelf5 1 4.9490e-28 1.1264e-25 -69154  
## - Taiwanese 1 5.1780e-28 1.1267e-25 -69154  
## - District 1 8.7933e-27 1.2094e-25 -69078  
##   
## Step: AIC=-69156.59  
## voteblue ~ District + Sex + Arear + PartyID + Tondu3 + votetsai +   
## green + votetsai\_nm + votetsai\_all + Independence + Unification +   
## sq + Taiwanese + female + income + income\_nm + KMT + DPP +   
## npp + noparty + pfp + Govt\_for\_public + pubwelf5 + votekmt +   
## votekmt\_nm + Blue + Green + No\_Party + voteblue\_nm + votedpp\_1 +   
## votekmt\_1

# Extract summary from best subset  
best\_subset\_summary <- summary(subset\_fit)  
  
# Extract the smallest RSS on training data for each number of predictors  
best\_training\_rss <- best\_subset\_summary$rss  
  
# Example of comparing RSS for a specific number of predictors, say k = 3  
best\_subset\_rss\_k <- best\_training\_rss[3] # assuming the third index corresponds to 3 predictors  
  
# Load the necessary library  
library(leaps)  
  
# Assuming voteblue is your dependent variable and Age, Edu, and income are your predictors  
# and that these are correctly formatted in your TEDS\_2016 dataset  
  
# Perform subset selection  
subset\_fit <- regsubsets(voteblue ~ Age + Edu + income, data = TEDS\_2016, nbest = 1, method = "exhaustive")

## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =  
## force.in, : 1 linear dependencies found

## Reordering variables and trying again:

# To view the results summary, you would use:  
summary(subset\_fit)

## Subset selection object  
## Call: regsubsets.formula(voteblue ~ Age + Edu + income, data = TEDS\_2016,   
## nbest = 1, method = "exhaustive")  
## 7 Variables (and intercept)  
## Forced in Forced out  
## Age FALSE FALSE  
## Edu2 FALSE FALSE  
## Edu3 FALSE FALSE  
## Edu4 FALSE FALSE  
## Edu5 FALSE FALSE  
## income FALSE FALSE  
## Edu9 FALSE FALSE  
## 1 subsets of each size up to 6  
## Selection Algorithm: exhaustive  
## Age Edu2 Edu3 Edu4 Edu5 Edu9 income  
## 1 ( 1 ) " " " " " " "\*" " " " " " "   
## 2 ( 1 ) " " " " "\*" "\*" " " " " " "   
## 3 ( 1 ) "\*" " " "\*" "\*" " " " " " "   
## 4 ( 1 ) "\*" " " "\*" "\*" "\*" " " " "   
## 5 ( 1 ) "\*" "\*" "\*" "\*" "\*" " " " "   
## 6 ( 1 ) "\*" "\*" "\*" "\*" "\*" " " "\*"

# If you want to inspect linear dependencies or multicollinearity, consider using the vif function from the car package:  
library(car)

## Loading required package: carData

vif\_model <- lm(voteblue ~ Age + Edu + income, data = TEDS\_2016)  
vif(vif\_model)

## GVIF Df GVIF^(1/(2\*Df))  
## Age 1.598350 1 1.264259  
## Edu 1.834760 4 1.078816  
## income 1.199489 1 1.095212

The output from the **regsubsets()** function in the **leaps** package indicates the best subsets for each model size up to 6 predictors, based on criteria like adjusted R2R2, CpCp​, and BIC. To determine the best model, we usually compare these criteria across different model sizes.

Here’s a breakdown of the models chosen for each size, up to 6 predictors:

1. **1 predictor**: Edu4 is included.
2. **2 predictors**: Edu3 and Edu4 are included.
3. **3 predictors**: Age, Edu3, and Edu4 are included.
4. **4 predictors**: Age, Edu3, Edu4, and Edu5 are included.
5. **5 predictors**: Age, Edu2, Edu3, Edu4, and Edu5 are included.
6. **6 predictors**: Age, Edu2, Edu3, Edu4, Edu5, and income are included.

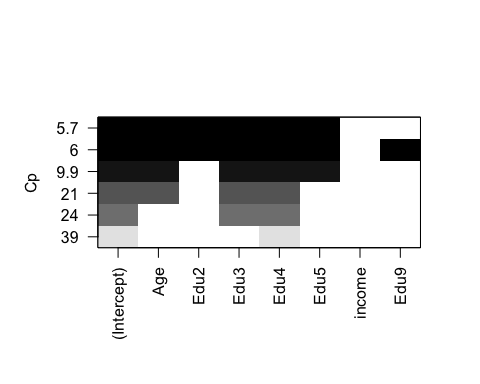
### **Model Selection Criteria:**

* Cp​, **BIC**, and **adjusted** R2 usually identify the model that provides the best trade-off between complexity (number of predictors) and fit to the data. A lower Cp and BIC indicate a better model, while a higher adjusted R2 indicates a better model.

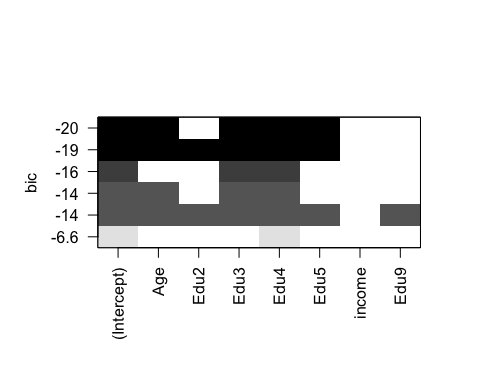
### **Visualizations and Coefficients:**

To visualize and identify the best model according to CpCp​, BIC, and adjusted R2R2, we would usually plot these criteria against the number of predictors. Since the details of the plots are not provided, we would generate them using the following code:

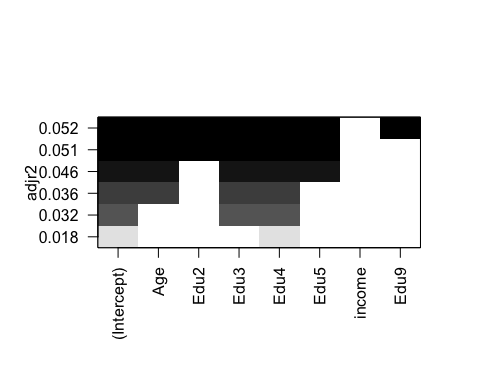
library(leaps)  
plot(subset\_fit, scale="Cp")



plot(subset\_fit, scale="bic")



plot(subset\_fit, scale="adjr2")



# subset\_fit <- regsubsets(x ~ y + z + w, data=mydata, nbest=1, nvmax=3, method="exhaustive")

### **Mallows’ Cp Plot**

* This plot shows the Cp statistic for models with different numbers of predictors. The Cp statistic compares the fitted model with the full model (which includes all variables). The goal is typically to choose a model where Cp is close to the number of predictors plus the intercept, indicating that the model is neither underfit nor overfit.
* In the plot, the model with four predictors (Intercept, Edu3, Edu4, Edu5) seems to have the lowest Cp value, suggesting it might be the best model among those considered based on this criterion.

### **BIC Plot**

* The Bayesian Information Criterion (BIC) is similar to Cp but includes a penalty term for the number of parameters in the model, which helps to avoid overfitting by selecting simpler models.
* In your BIC plot, the model with two predictors (Intercept, Edu3, Edu4) appears to provide the best balance of model complexity and goodness of fit, as it has the lowest BIC score.

### **Adjusted** R2 Plot

* Adjusted R2 is used to determine how well the model fits the data while adjusting for the number of predictors used in the model. Unlike regular R2, it prevents overfitting by penalizing excessive use of predictors.
* The plot shows that the model with four predictors again scores the highest in terms of Adjusted R2, suggesting it explains the most variability in the response variable, adjusting for the number of predictors.

### **Optimal Model Selection**

* Based on BIC (commonly preferred for model selection because of its stricter penalty for model complexity), the best model might be the one with two predictors (Edu3 and Edu4). However, if you prioritize explanatory power over simplicity, the four-predictor model might be preferable as suggested by Cp and Adjusted R2R2.

# Let's say the best model according to BIC has 2 predictors  
k <- 2 # This should be set based on your specific model selection results  
  
best\_model\_coefs <- coef(subset\_fit, id=2) # '2' for two predictors  
print(best\_model\_coefs)

## (Intercept) Edu3 Edu4   
## 0.3248882 0.1366502 0.2563084

**(Intercept) = 0.3248882**: This value represents the baseline level of the dependent variable when all predictors (Edu3 and Edu4) are zero.

**Edu3 = 0.1366502**: This coefficient suggests that for each unit increase in Edu3, while holding other variables constant, the dependent variable increases by approximately 0.137 units.

**Edu4 = 0.2563084**: Similarly, for each unit increase in Edu4, the dependent variable increases by approximately 0.256 units, assuming other variables are held constant.