Assignment 2

Sonali Singh

2024-05-07

# Knowledge Mining: Linear Discriminant Analysis

File: Lab\_LDA01.R Theme: Linear Discriminant Analysis Adapted from ISLR Chapter 4 Lab

### **Overview**

The TEDS\_2016 dataset contains a rich set of variables aimed at understanding political opinions in Taiwan. Key variables include demographic factors, political preferences, and views on unification with China.

### **Summary Statistics**

The dataset encapsulates responses from a diverse demographic, spanning various districts, age groups, and education levels. Notably, the data shows:

* A nearly balanced sex distribution.
* A wide age range, allowing for generational analysis in political beliefs.
* A focus on voters’ education levels that vary from below elementary to college and above.

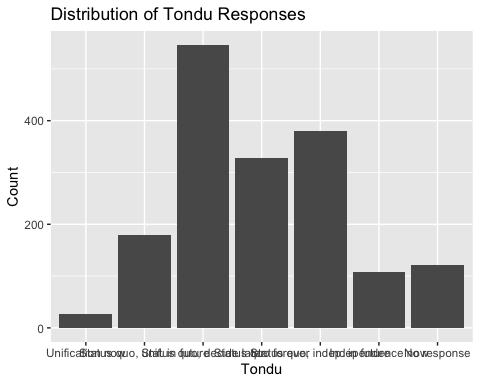
library(haven)  
TEDS\_2016 <- read\_stata("https://github.com/datageneration/home/blob/master/DataProgramming/data/TEDS\_2016.dta?raw=true")  
  
summary(TEDS\_2016)

## District Sex Age Edu Arear   
## Min. : 201 Min. :1.000 Min. :1.0 Min. :1.000 Min. :1.000   
## 1st Qu.:1401 1st Qu.:1.000 1st Qu.:2.0 1st Qu.:2.000 1st Qu.:1.000   
## Median :6406 Median :1.000 Median :3.0 Median :3.000 Median :3.000   
## Mean :4661 Mean :1.486 Mean :3.3 Mean :3.334 Mean :2.744   
## 3rd Qu.:6604 3rd Qu.:2.000 3rd Qu.:5.0 3rd Qu.:5.000 3rd Qu.:4.000   
## Max. :6806 Max. :2.000 Max. :5.0 Max. :9.000 Max. :6.000   
##   
## Career Career8 Ethnic Party   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. : 1.00   
## 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000 1st Qu.: 5.00   
## Median :2.000 Median :4.000 Median :1.000 Median : 7.00   
## Mean :2.683 Mean :3.811 Mean :1.658 Mean :13.02   
## 3rd Qu.:4.000 3rd Qu.:5.000 3rd Qu.:2.000 3rd Qu.:25.00   
## Max. :5.000 Max. :8.000 Max. :9.000 Max. :26.00   
##   
## PartyID Tondu Tondu3 nI2   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. : 1.00   
## 1st Qu.:2.000 1st Qu.:3.000 1st Qu.:2.000 1st Qu.: 1.00   
## Median :2.000 Median :4.000 Median :2.000 Median : 3.00   
## Mean :4.522 Mean :4.127 Mean :2.667 Mean :35.13   
## 3rd Qu.:9.000 3rd Qu.:5.000 3rd Qu.:3.000 3rd Qu.:98.00   
## Max. :9.000 Max. :9.000 Max. :9.000 Max. :98.00   
##   
## votetsai green votetsai\_nm votetsai\_all   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :1.0000 Median :0.0000 Median :1.0000 Median :1.0000   
## Mean :0.6265 Mean :0.3781 Mean :0.6265 Mean :0.5478   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## NA's :429 NA's :429 NA's :248   
## Independence Unification sq Taiwanese   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :1.0000 Median :1.0000   
## Mean :0.2888 Mean :0.1225 Mean :0.5172 Mean :0.6272   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
##   
## edu female whitecollar lowincome   
## Min. :1.000 Min. :0.0000 Min. :0.0000 Min. :1.000   
## 1st Qu.:2.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:4.000   
## Median :3.000 Median :0.0000 Median :1.0000 Median :5.000   
## Mean :3.301 Mean :0.4864 Mean :0.5373 Mean :4.343   
## 3rd Qu.:5.000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:5.000   
## Max. :5.000 Max. :1.0000 Max. :1.0000 Max. :5.000   
## NA's :10   
## income income\_nm age KMT   
## Min. : 1.000 Min. : 1.000 Min. : 20.00 Min. :0.0000   
## 1st Qu.: 3.000 1st Qu.: 2.000 1st Qu.: 35.00 1st Qu.:0.0000   
## Median : 5.500 Median : 5.000 Median : 49.00 Median :0.0000   
## Mean : 5.324 Mean : 5.281 Mean : 49.11 Mean :0.2296   
## 3rd Qu.: 7.000 3rd Qu.: 8.000 3rd Qu.: 61.00 3rd Qu.:0.0000   
## Max. :10.000 Max. :10.000 Max. :100.00 Max. :1.0000   
## NA's :330   
## DPP npp noparty pfp   
## Min. :0.0000 Min. :0.00000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :0.0000 Median :0.00000 Median :0.0000 Median :0.00000   
## Mean :0.3497 Mean :0.02544 Mean :0.3716 Mean :0.01893   
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.00000 Max. :1.0000 Max. :1.00000   
##   
## South north Minnan\_father Mainland\_father   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :1.0000 Median :0.0000   
## Mean :0.4947 Mean :0.4799 Mean :0.7225 Mean :0.1024   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
##   
## Econ\_worse Inequality inequality5 econworse5   
## Min. :0.0000 Min. :0.0000 Min. :1.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:4.000 1st Qu.:3.000   
## Median :1.0000 Median :1.0000 Median :5.000 Median :4.000   
## Mean :0.5544 Mean :0.9355 Mean :4.495 Mean :3.644   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:5.000 3rd Qu.:4.000   
## Max. :1.0000 Max. :1.0000 Max. :5.000 Max. :5.000   
##   
## Govt\_for\_public pubwelf5 Govt\_dont\_care highincome   
## Min. :0.0000 Min. :1.000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :3.000 Median :0.0000 Median :1.0000   
## Mean :0.4249 Mean :2.877 Mean :0.4988 Mean :0.5765   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :5.000 Max. :1.0000 Max. :1.0000   
## NA's :330   
## votekmt votekmt\_nm Blue Green No\_Party  
## Min. :0.0000 Min. :0.0000 Min. :0 Min. :0 Min. :0   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0 1st Qu.:0 1st Qu.:0   
## Median :0.0000 Median :0.0000 Median :0 Median :0 Median :0   
## Mean :0.2053 Mean :0.2752 Mean :0 Mean :0 Mean :0   
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0   
## Max. :1.0000 Max. :1.0000 Max. :0 Max. :0 Max. :0   
## NA's :429   
## voteblue voteblue\_nm votedpp\_1 votekmt\_1   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :1.0000 Median :0.0000   
## Mean :0.2787 Mean :0.3735 Mean :0.5256 Mean :0.2309   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## NA's :429 NA's :187 NA's :187

str(TEDS\_2016)

## tibble [1,690 × 54] (S3: tbl\_df/tbl/data.frame)  
## $ District : dbl+lbl [1:1690] 201, 201, 201, 201, 201, 201, 201, 201, 201, 201, 201...  
## ..@ label : chr "District"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:73] 201 401 501 502 701 702 703 704 801 802 ...  
## .. ..- attr(\*, "names")= chr [1:73] "Yi Lan County Single District" "Hsinchu County Single District" "Miaoli County 1st District" "Miaoli County 2nd District" ...  
## $ Sex : dbl+lbl [1:1690] 2, 2, 1, 1, 2, 2, 1, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 1,...  
## ..@ label : chr "Sex"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Male" "Female"  
## $ Age : dbl+lbl [1:1690] 4, 2, 5, 4, 5, 5, 5, 4, 5, 4, 5, 1, 5, 3, 4, 5, 4, 5,...  
## ..@ label : chr "Age"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:5] 1 2 3 4 5  
## .. ..- attr(\*, "names")= chr [1:5] "20-29" "30-39" "40-49" "50-59" ...  
## $ Edu : dbl+lbl [1:1690] 4, 5, 5, 2, 1, 2, 1, 5, 1, 1, 1, 2, 1, 5, 5, 1, 3, 4,...  
## ..@ label : chr "Education"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:6] 1 2 3 4 5 9  
## .. ..- attr(\*, "names")= chr [1:6] "Below elementary school" "Junior high school" "Senior high school" "College" ...  
## $ Arear : dbl+lbl [1:1690] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Area"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:6] 1 2 3 4 5 6  
## .. ..- attr(\*, "names")= chr [1:6] "Taipei, New Taipei, Keelung and Yi Lan" "Taoyuan, Hsinchu and Miaoli" "Taichung, Changhua and Nantou" "Yunlin, Chiayi and Tainan" ...  
## $ Career : dbl+lbl [1:1690] 1, 2, 1, 4, 3, 2, 4, 1, 4, 3, 3, 5, 5, 4, 1, 5, 2, 2,...  
## ..@ label : chr "Occupations5"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:5] 1 2 3 4 5  
## .. ..- attr(\*, "names")= chr [1:5] "Hight-class WHITE COLLAR" "Low-class WHITE COLLAR" "FARMER" "WORKER" ...  
## $ Career8 : dbl+lbl [1:1690] 1, 3, 1, 4, 5, 7, 4, 2, 4, 5, 5, 7, 7, 7, 2, 7, 3, 1,...  
## ..@ label : chr "Occupation8"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:8] 1 2 3 4 5 6 7 8  
## .. ..- attr(\*, "names")= chr [1:8] "Civil servants" "Managers and Professionals (priv.)" "CLERKS (priv.)" "Labor (priv.)" ...  
## $ Ethnic : dbl+lbl [1:1690] 1, 2, 2, 1, 9, 1, 2, 1, 1, 2, 1, 1, 2, 1, 2, 9, 2, 2,...  
## ..@ label : chr "Ethnic"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:4] 1 2 3 9  
## .. ..- attr(\*, "names")= chr [1:4] "Taiwanese" "Both" "Chinese" "Noresponse"  
## $ Party : dbl+lbl [1:1690] 25, 25, 3, 25, 25, 6, 25, 24, 25, 25, 6, 5, 25, ...  
## ..@ label : chr "Party Preference"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:26] 1 2 3 4 5 6 7 8 9 10 ...  
## .. ..- attr(\*, "names")= chr [1:26] "Strongly support KMT" "Somewhat support KMT" "Lean to KMT" "Somewhat lean to KMT" ...  
## $ PartyID : dbl+lbl [1:1690] 9, 9, 1, 9, 9, 2, 9, 6, 9, 9, 2, 2, 9, 1, 1, 9, 9, 9,...  
## ..@ label : chr "Party Identification"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:7] 1 2 3 4 5 6 9  
## .. ..- attr(\*, "names")= chr [1:7] "KMT" "DPP" "NP" "PFP" ...  
## $ Tondu : dbl+lbl [1:1690] 3, 5, 3, 5, 9, 4, 9, 6, 9, 9, 5, 5, 9, 5, 4, 9, 9, 4,...  
## ..@ label : chr "Position on unification and independence"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:7] 1 2 3 4 5 6 9  
## .. ..- attr(\*, "names")= chr [1:7] "Immediate unification" "Maintain the status quo,move toward unification" "Maintain the status quo, decide either unification or independence" "Maintain the status quo forever" ...  
## $ Tondu3 : dbl+lbl [1:1690] 2, 3, 2, 3, 9, 2, 9, 3, 9, 9, 3, 3, 9, 3, 2, 9, 9, 2,...  
## ..@ label : chr "3 categories of TONDU"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:4] 1 2 3 9  
## .. ..- attr(\*, "names")= chr [1:4] "Unification" "Maintain the status quo" "Independence" "Nonresponse"  
## $ nI2 : dbl+lbl [1:1690] 3, 98, 98, 3, 98, 98, 98, 3, 98, 1, 2, 98, 98, ...  
## ..@ label : chr "Who is the current the premier of our country?"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:5] 1 2 3 95 98  
## .. ..- attr(\*, "names")= chr [1:5] "Correct" "Incorrect" "I know but can't remember the name" "Refuse to answer" ...  
## $ votetsai : num [1:1690] NA 1 0 NA NA 1 1 1 1 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ green : num [1:1690] 0 0 0 0 0 1 0 1 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votetsai\_nm : num [1:1690] NA 1 0 NA NA 1 1 1 1 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votetsai\_all : num [1:1690] 0 1 0 0 0 1 1 1 1 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Independence : num [1:1690] 0 1 0 1 0 0 0 1 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Unification : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ sq : num [1:1690] 1 0 1 0 0 1 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Taiwanese : num [1:1690] 1 0 0 1 0 1 0 1 1 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ edu : num [1:1690] 4 5 5 2 1 2 1 5 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ female : num [1:1690] 1 1 0 0 1 1 0 1 1 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ whitecollar : num [1:1690] 1 1 1 0 0 1 0 1 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ lowincome : num [1:1690] 4 4 5 4 3 5 2 5 5 5 ...  
## ..- attr(\*, "label")= chr "How serious do you think low income of salaryman?"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ income : num [1:1690] 8 7 8 5 5.5 9 1 10 2 5.5 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ income\_nm : num [1:1690] 8 7 8 5 NA 9 1 10 2 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ age : num [1:1690] 59 39 63 55 76 64 75 54 64 59 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ KMT : num [1:1690] 0 0 1 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ DPP : num [1:1690] 0 0 0 0 0 1 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ npp : num [1:1690] 0 0 0 0 0 0 0 1 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ noparty : num [1:1690] 1 1 0 1 1 0 1 0 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ pfp : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ South : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ north : num [1:1690] 1 1 1 1 1 1 1 1 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Minnan\_father : num [1:1690] 1 1 1 1 1 1 1 1 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Mainland\_father: num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Econ\_worse : num [1:1690] 0 0 1 1 0 1 1 1 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Inequality : num [1:1690] 1 1 1 1 0 1 0 1 1 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ inequality5 : num [1:1690] 4 5 5 5 3 5 3 5 5 5 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ econworse5 : num [1:1690] 3 3 4 5 3 4 4 5 5 5 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Govt\_for\_public: num [1:1690] 1 1 1 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ pubwelf5 : num [1:1690] 5 5 4 1 3 2 2 1 3 2 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Govt\_dont\_care : num [1:1690] 0 0 1 1 0 1 1 1 0 1 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ highincome : num [1:1690] 1 1 1 1 NA 1 0 1 0 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votekmt : num [1:1690] 0 0 1 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votekmt\_nm : num [1:1690] NA 0 1 NA NA 0 0 0 0 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Blue : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ Green : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ No\_Party : num [1:1690] 0 0 0 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ voteblue : num [1:1690] 0 0 1 0 0 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ voteblue\_nm : num [1:1690] NA 0 1 NA NA 0 0 0 0 NA ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votedpp\_1 : num [1:1690] NA 1 0 NA NA 1 1 1 1 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ votekmt\_1 : num [1:1690] NA 0 1 NA NA 0 0 0 0 0 ...  
## ..- attr(\*, "format.stata")= chr "%9.0g"

library(ggplot2)  
TEDS\_2016$Tondu <- factor(TEDS\_2016$Tondu, labels=c("Unification now", "Status quo, unif. in future", "Status quo, decide later", "Status quo forever", "Status quo, indep. in future", "Independence now", "No response"))  
ggplot(TEDS\_2016, aes(x = Tondu)) + geom\_bar() + labs(title = "Distribution of Tondu Responses", x = "Tondu", y = "Count")



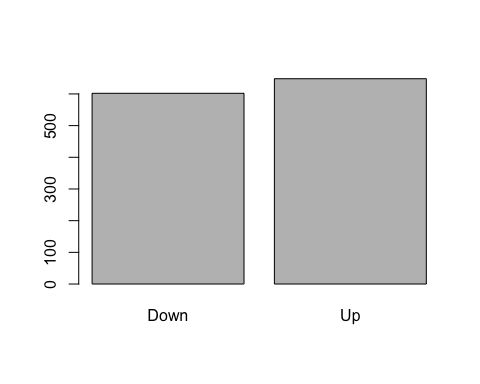
### **Encountered Issues**

* **Missing Values:** Some key political opinion variables like **votetsai** (vote for Tsai Ing-wen) have missing values, potentially skewing analysis related to election outcomes.
* **Outliers:** Extreme values in income and education levels might represent data entry errors or require transformation for meaningful analysis.

### **Handling Missing Values**

Given the nature of the study, imputation for missing values in **votetsai** could be considered using mode imputation, as it represents a categorical binary response. Alternatively, listwise deletion could be used if the missing data is not systematically biased.

## Linear Discriminant Analysis  
freq(Direction)

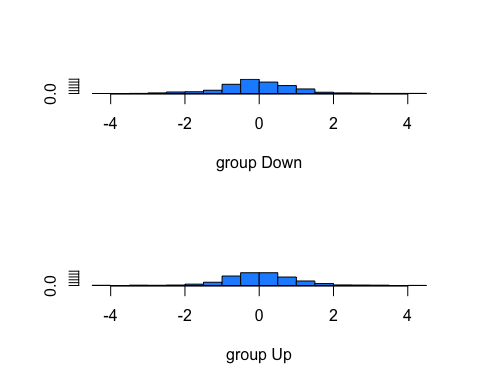


## Direction   
## Frequency Percent  
## Down 602 48.16  
## Up 648 51.84  
## Total 1250 100.00

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

### **Relationship Exploration**

#### **Tondu and Other Variables**

To explore the relationship between **Tondu** (stance on unification) and variables like **female**, **DPP**, **age**, **income**, **edu**, **Taiwanese**, and **Econ\_worse**, the following methods are proposed:

* **Correlation Analysis** for continuous variables to check for linear relationships.
* **Chi-Square Tests** for categorical variables to identify associations.
* **Logistic Regression** could be utilized to understand how these predictors influence the likelihood of favoring unification now versus later or never.

#### **Analysis of votetsai Variable**

Given the binary nature of **votetsai**, logistic regression is appropriate. This model will help understand the impact of demographics and political opinions on voting for Tsai Ing-wen.

### **Visual Analysis**

#### **Frequency Table and Bar Chart for Tondu Variable**

The **Tondu** variable was recoded with labels corresponding to different political stances regarding unification:

* Immediate unification
* Maintain the status quo, decide later
* Independence now
* No response

A bar chart was created to visualize the frequency of each category, revealing predominant preferences in the dataset. This visualization helps in quickly grasping the political leanings of the surveyed population.

### **Conclusions**

This exploratory analysis highlights the complex interplay of demographic and socio-political factors in shaping political opinions in Taiwan. The handling of missing data and the subsequent analyses provide foundational insights that could guide more detailed studies, such as predictive modeling of election outcomes or deeper analysis of policy support.

### **Recommendations for Further Analysis**

* **Predictive Modeling:** Utilize logistic regression to predict election outcomes based on socio-demographic and economic variables.
* **Deep Dive into Economic Factors:** Given the significant role of economic satisfaction (**Econ\_worse**), further analysis could quantify its impact on political stability and party support.

For detailed charts, tables, and further statistical tests, additional code execution and data wrangling in R are required as per the methodologies outlined.