American, Mexican-DATA

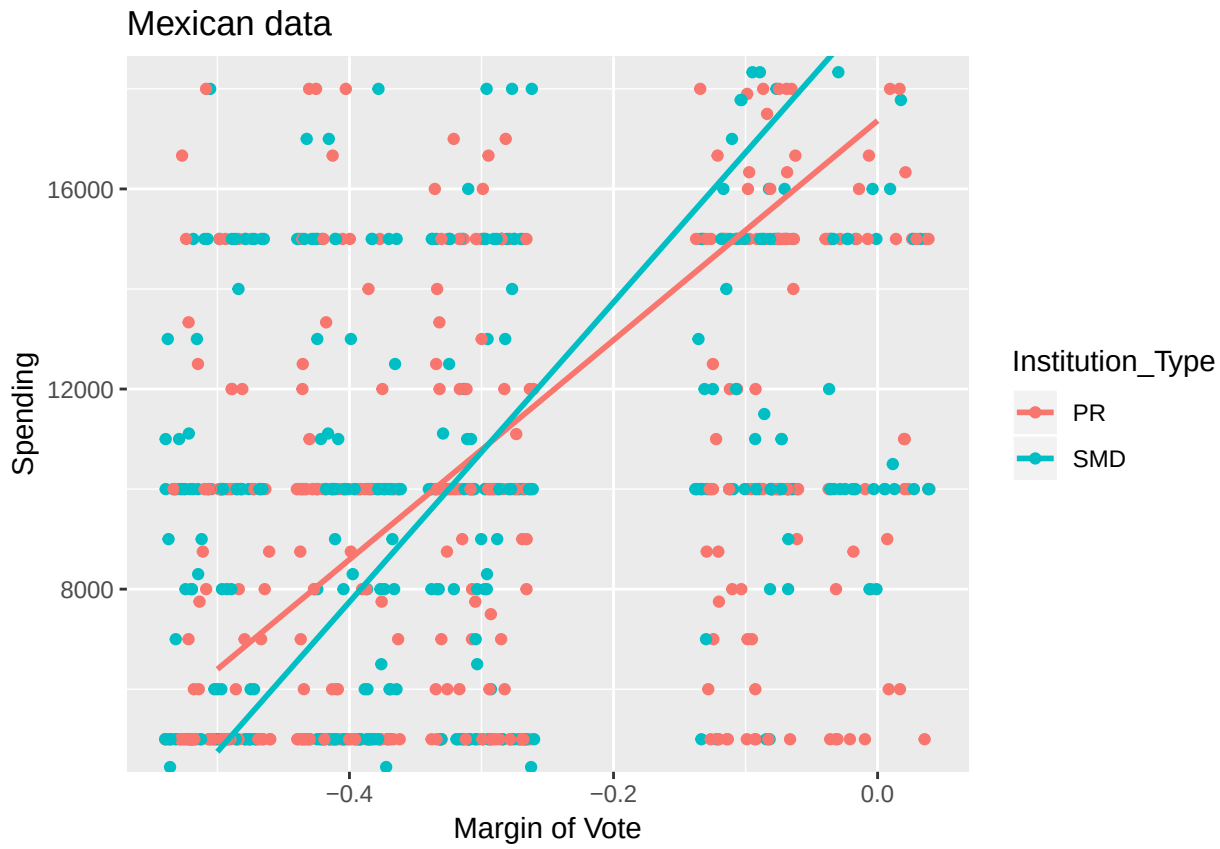
Tolgahan Dilgin, Alan Yu

February 22, 2019

```
#load the relevant libraries

library(ggplot2)
library(reshape2)
library(dplyr)
library(data.table)

data = read.csv("Mexican_Data_Combined.csv")
#this mxdata.csv is the subset from my Alan_Data_Entry_1.xlsx containing P12s, Respondant_ID and Instit

subdata = as.data.frame(data)

subdata = subdata[subdata$Actual.Total>=90000 & subdata$Actual.Total<=100000,]
#0 and -1 are competitive and interaction term between competitive and type



#subset the columns that we want to use to analyze Question 12
subdata_12 = select(subdata, Respondent_ID, Institution_Type, P12_Distrito_1, P12_Distrito_2, P12_Distri

#then to make the graph we need to change Districts 1-9 to there corresponding competitiveness level

#rearrange data, # I use the melt function because we want the column names to be on the x-axis. We wan
subdata_12.melt = melt(subdata_12, id=c("Respondent_ID", "Institution_Type"))

#Rename the districts to level of competitiveness and Institution Type to English abbreviation.
subdata_12.melt$variable = as.character(subdata_12.melt$variable)
subdata_12.melt$Institution_Type = as.character(subdata_12.melt$Institution_Type)
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_1"] = "-.3"
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_2"] = "-.4"
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_3"] = "-.5"
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_4"] = "-.1"
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_5"] = "0"
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_6"] = "-.1"
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_7"] = "-.5"
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_8"] = "-.4"
subdata_12.melt$variable[subdata_12.melt$variable == "P12_Distrito_9"] = "-.3"
subdata_12.melt$Institution_Type[subdata_12.melt$Institution_Type == "RP"] = "PR"
subdata_12.melt$Institution_Type[subdata_12.melt$Institution_Type == "MR"] = "SMD"

subdata_12.melt = subdata_12.melt[complete.cases(subdata_12.melt), ] #to remove NAs
subdata_12.melt = transform(subdata_12.melt, variable = as.numeric(variable)) #change variable from cha

Competitive = c()
for (i in subdata_12.melt$variable) {
  if (i == "0" | i == "-0.1") {
    Competitive = c(Competitive, 1)
```

```
  } else {
    Competitive = c(Competitive, 0)
  }
}
Binary_IT = c()
for (i in subdata_12.melt$Institution_Type) {
  if (i == "SMD") {
    Binary_IT = c(Binary_IT, 1)
  } else {
    Binary_IT = c(Binary_IT, 0)
  }
}

subdata_12.melt = cbind(subdata_12.melt, Binary_IT, Competitive)
```

```
ggplot(subdata_12.melt, aes(x = variable, y = value, color = Institution_Type)) + geom_jitter() + labs(
```




```
InstitutionTypeSubSet = split(subdata_12.melt, subdata_12.melt$Institution_Type)

#regression coefficients
lm(value ~ variable, data = InstitutionTypeSubSet$SMD)
```

```
##
## Call:
## lm(formula = value ~ variable, data = InstitutionTypeSubSet$SMD)
##
```

```
## Coefficients:
## (Intercept)     variable
##       19721        29941
```

```
lm(value ~ variable, data = InstitutionTypeSubSet$PR)
```

```
##
## Call:
## lm(formula = value ~ variable, data = InstitutionTypeSubSet$PR)
##
## Coefficients:
## (Intercept)     variable
##       17365        21932
```

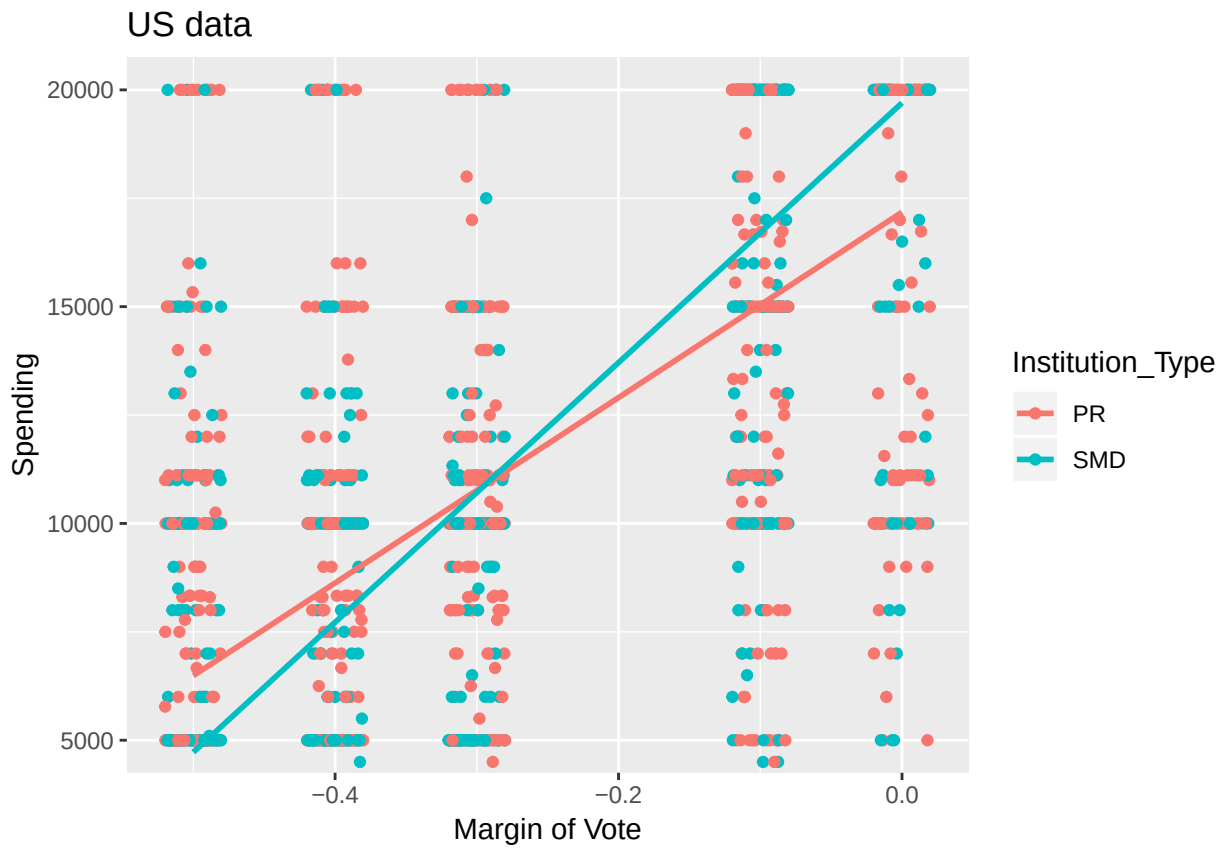
```
summary(lm(value ~ Competitive + Binary_IT + Competitive*Binary_IT, data = subdata_12.melt))
```

```
##
## Call:
## lm(formula = value ~ Competitive + Binary_IT + Competitive *
##     Binary_IT, data = subdata_12.melt)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -18617  -5299  -1458   3353  41780
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8219.6      366.3  22.437  < 2e-16 ***
## Competitive             8427.3      634.5  13.281  < 2e-16 ***
## Binary_IT               -920.9      510.9  -1.802  0.07173 .
## Competitive:Binary_IT   2890.8      885.0   3.267  0.00112 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7454 on 1274 degrees of freedom
## Multiple R-squared:  0.2871, Adjusted R-squared:  0.2854
## F-statistic:   171 on 3 and 1274 DF,  p-value: < 2.2e-16
```

```
####
```

```
amdata = read.csv("Statalik_Multiple.csv")
amdata = as.data.frame(amdata)
amdata = amdata[amdata$P12_Total>=90000 & amdata$P12_Total<=100000,]
keep = c("District_Spending", "Competitiveness", "Institution_Type")
amdata = (amdata[keep])
amdata$Competitiveness = abs(amdata$Competitiveness) #normalize the competitiveness
amdata$Competitiveness = -1 * (amdata$Competitiveness) #make all values negative

Competitive = c()
for (i in amdata$Competitiveness) {
  if (i == "0" | i == "-0.1") {
    Competitive = c(Competitive, 1)
  } else {
    Competitive = c(Competitive, 0)
  }
}
```

```
Binary_IT = c()
for (i in amdata$Institution_Type) {
  if (i == "SMD") {
    Binary_IT = c(Binary_IT, 1)
  } else {
    Binary_IT = c(Binary_IT, 0)
  }
}

#cbind(amdata, Binary_IT, Competitive)

ggplot(amdata, aes(x = Competitiveness, y = District_Spending, color = Institution_Type)) + geom_jitter
```



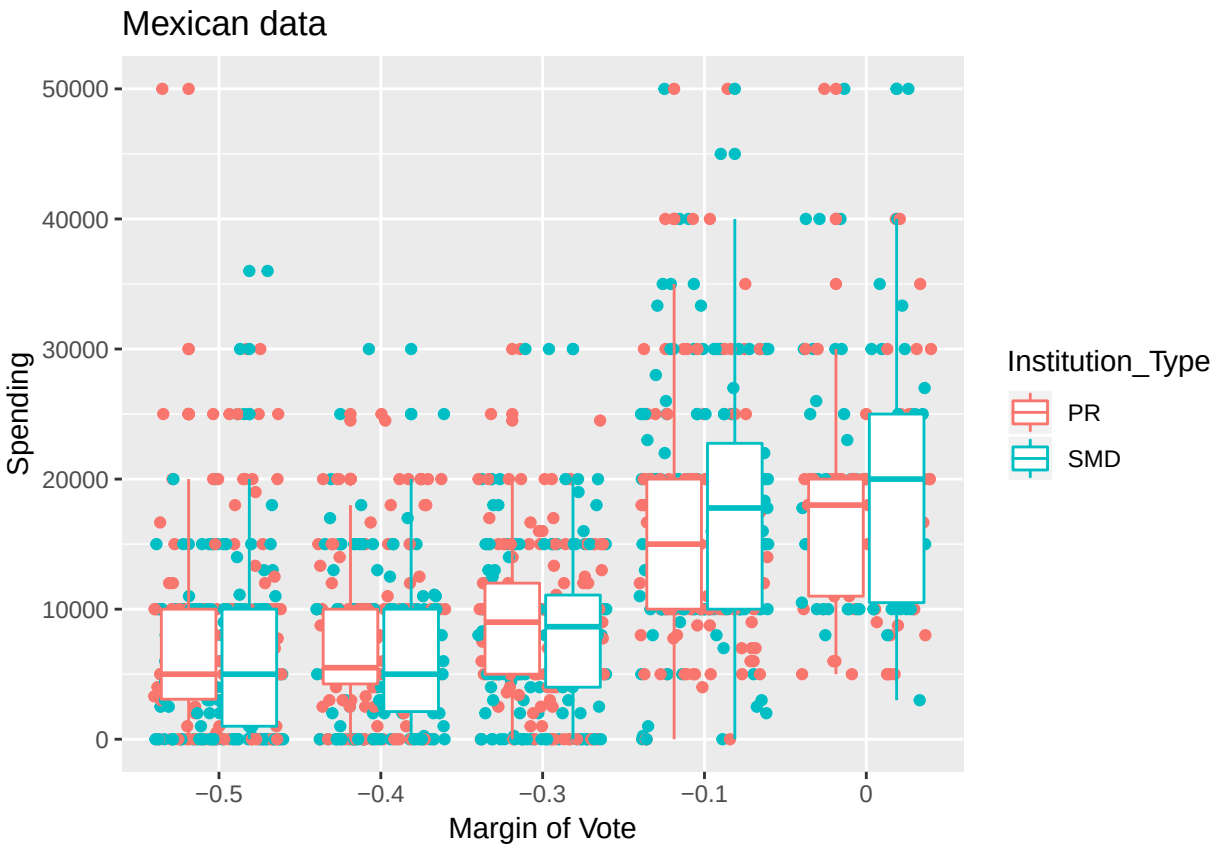

```
AMInstitutionTypeSubSet = split(amdata, amdata$Institution_Type)

#regression coefficients without adjustment
lm(District_Spending ~ Competitiveness, data = AMInstitutionTypeSubSet$SMD)
```

```
##
## Call:
## lm(formula = District_Spending ~ Competitiveness, data = AMInstitutionTypeSubSet$SMD)
##
## Coefficients:
##     (Intercept)  Competitiveness
##           19700            29950
```

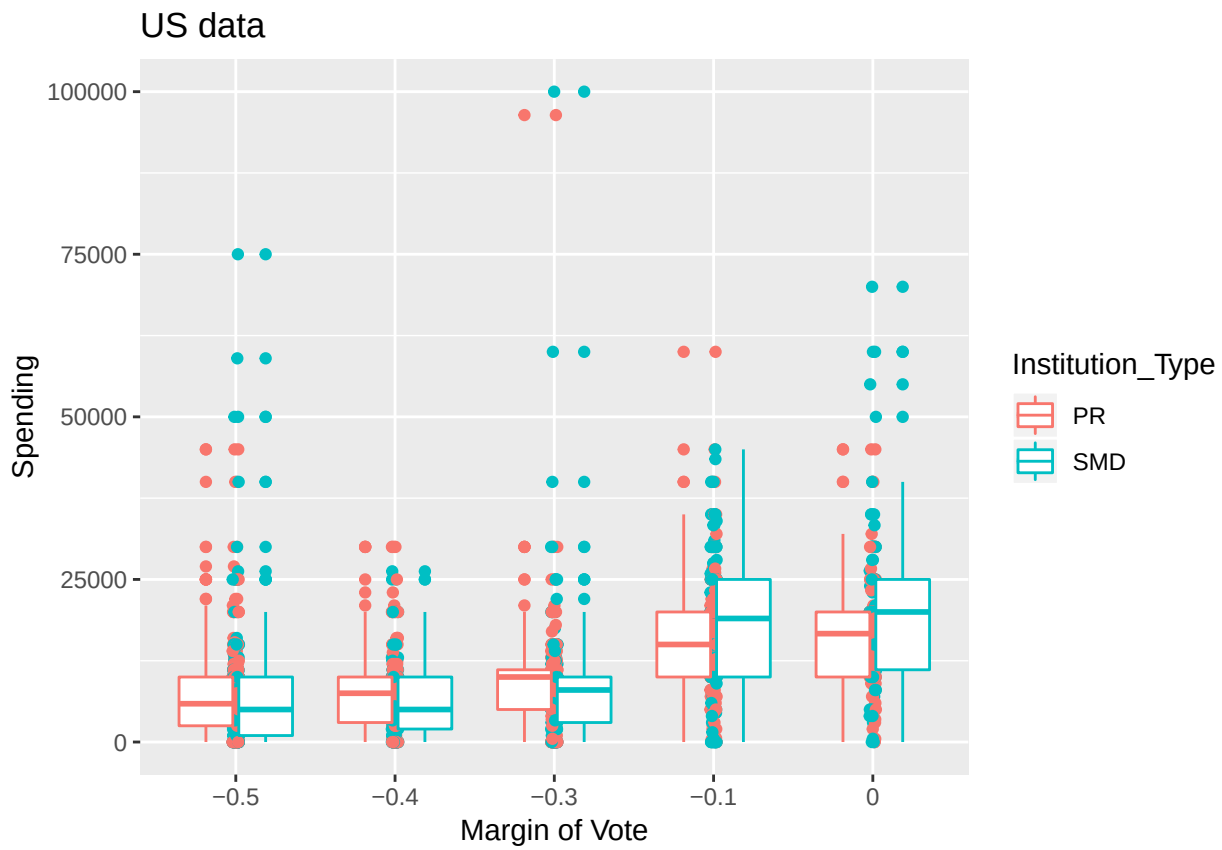
```
lm(District_Spending ~ Competitiveness, data = AMInstitutionTypeSubSet$PR)
```

```
##
## Call:
## lm(formula = District_Spending ~ Competitiveness, data = AMInstitutionTypeSubSet$PR)
##
## Coefficients:
##     (Intercept)  Competitiveness
##           17178            21379
```

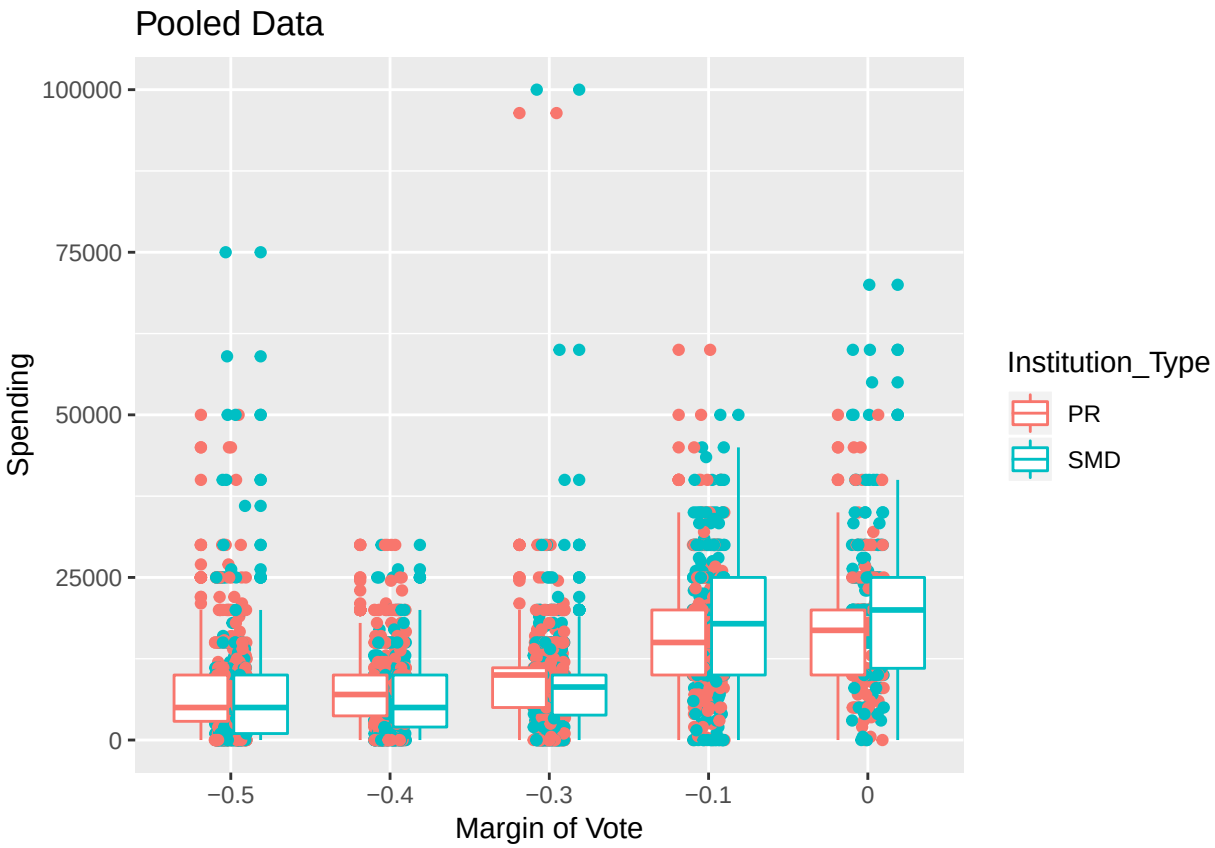
```
#mexican data boxplot
ggplot(subdata_12.melt, aes(x = factor(variable), y = value, color = Institution_Type)) + geom_jitter()
```




```
#american data boxplot without adjustment
ggplot(amdata, aes(x = factor(Competitiveness), y = District_Spending, color = Institution_Type)) + geo
```




```
names(subdata_12.melt)[3]<-"Competitiveness"
names(subdata_12.melt)[4]<-"District_Spending"
all_data = rbind(subdata_12.melt[c("District_Spending", "Competitiveness", "Institution_Type")], amdata

#combined data boxplot
ggplot(all_data, aes(x = factor(Competitiveness), y = District_Spending, color = Institution_Type)) + g
```


Pooled Data
100000
75000
50000
25000
0
Spending
−0.5
−0.4
−0.3
−0.1
0
Margin of Vote
Institution_Type
PR
SMD