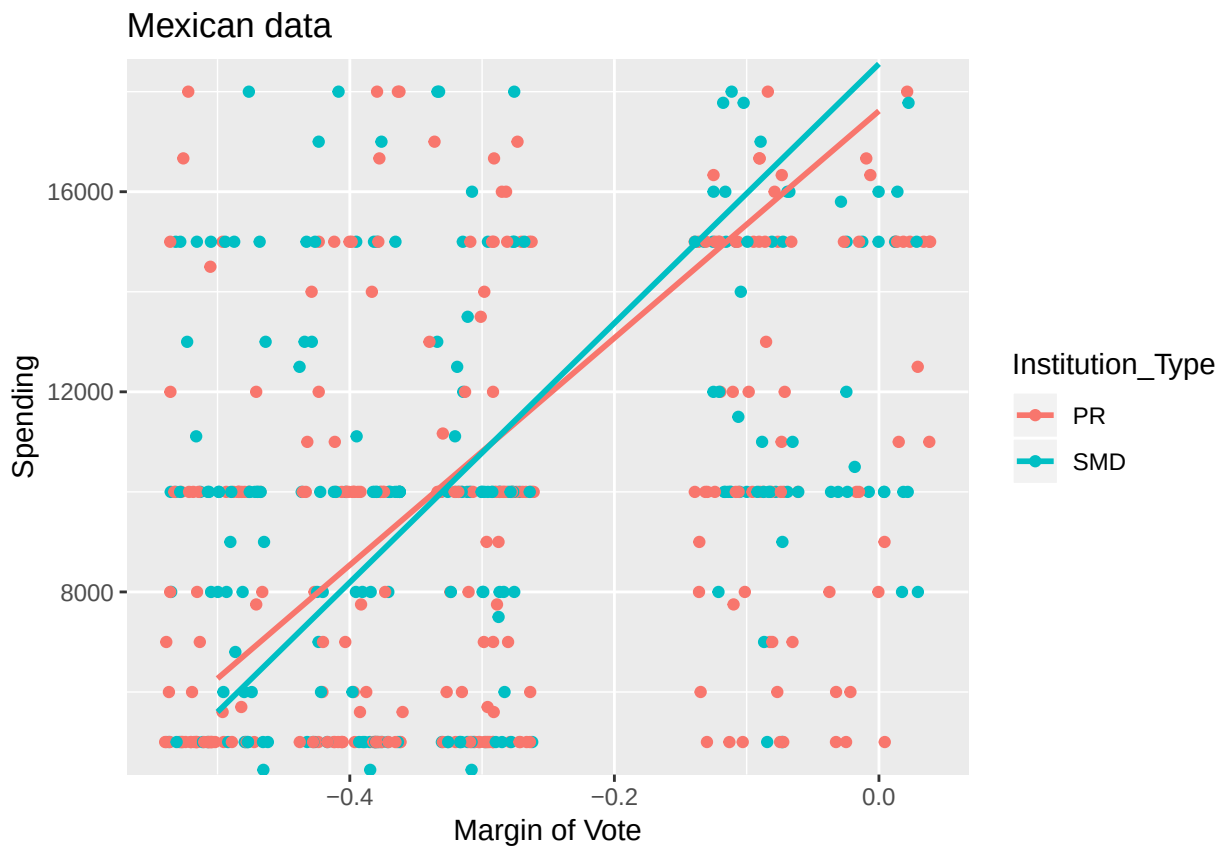
MX-DATA

Tolgahan Dilgin, Alan Yu

February 22, 2019

```
#load the relevant libraries

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

data = read.csv("Alan_Data_Entry_1.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)

#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
```

```
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
##       18555        25909
```

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

```
##
## Call:
## lm(formula = value ~ variable, data = InstitutionTypeSubSet$PR)
##
## Coefficients:
## (Intercept)     variable
##       17614        22690
```

```
amdata = read.csv("Statalik_Multiple.csv")

amdata = as.data.frame(amdata)
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

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
##           15549            22497
```

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

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

```
##           15873             18373
```

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


```
AM_SMD = AMInstitutionTypeSubSet$SMD
AM_PR = AMInstitutionTypeSubSet$PR

#regression coefficients with adjustment
lm(District_Spending ~ Competitiveness, data = AM_SMD[AM_SMD$District_Spending>=5000 & AM_SMD$District_S
```

```
##
## Call:
## lm(formula = District_Spending ~ Competitiveness, data = AM_SMD[AM_SMD$District_Spending >=
##     5000 & AM_SMD$District_Spending <= 50000, ])
##
## Coefficients:
##     (Intercept)  Competitiveness
##           19996            23055
```

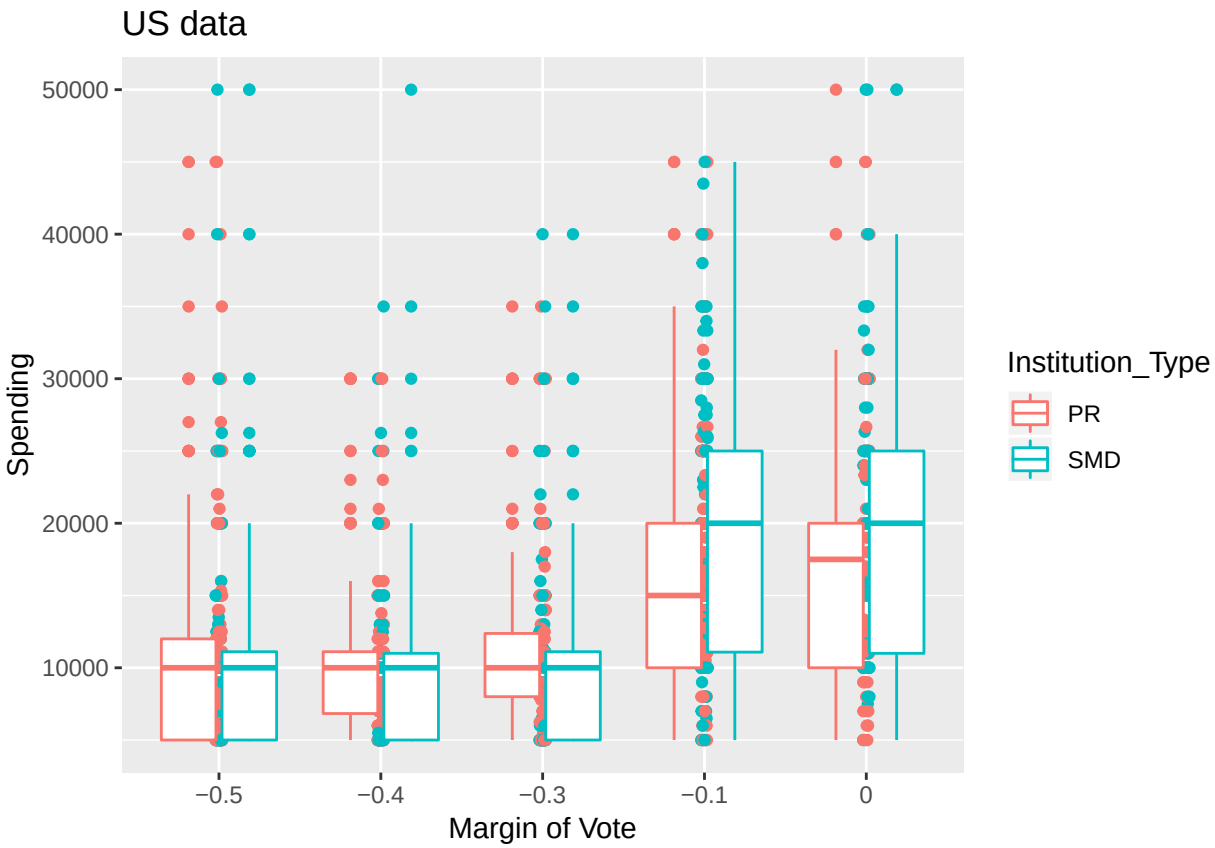
```
lm(District_Spending ~ Competitiveness, data = AM_PR[AM_PR$District_Spending>=5000 & AM_PR$District_Spe
```

```
##
## Call:
## lm(formula = District_Spending ~ Competitiveness, data = AM_PR[AM_PR$District_Spending >=
##     5000 & AM_PR$District_Spending <= 50000, ])
##
## Coefficients:
##     (Intercept)  Competitiveness
```

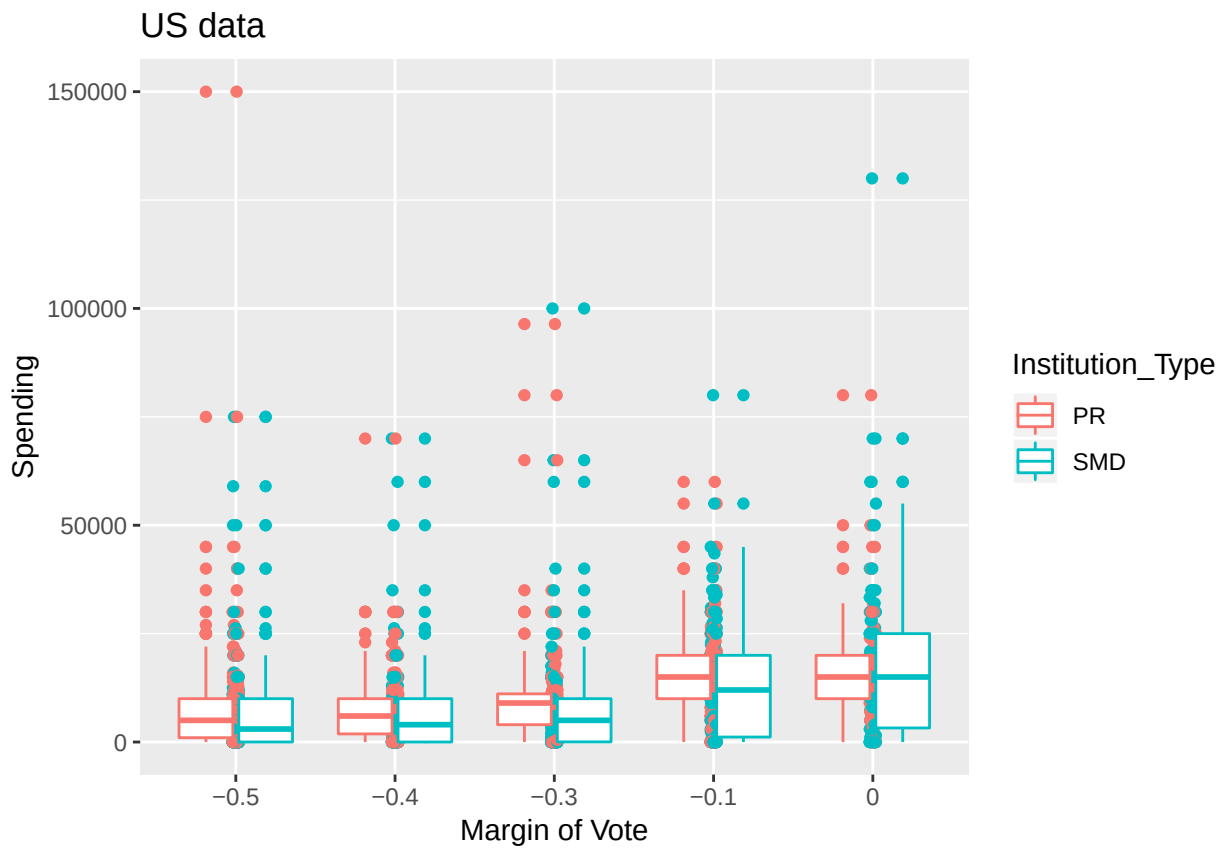
```
##           17473             15383
```

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

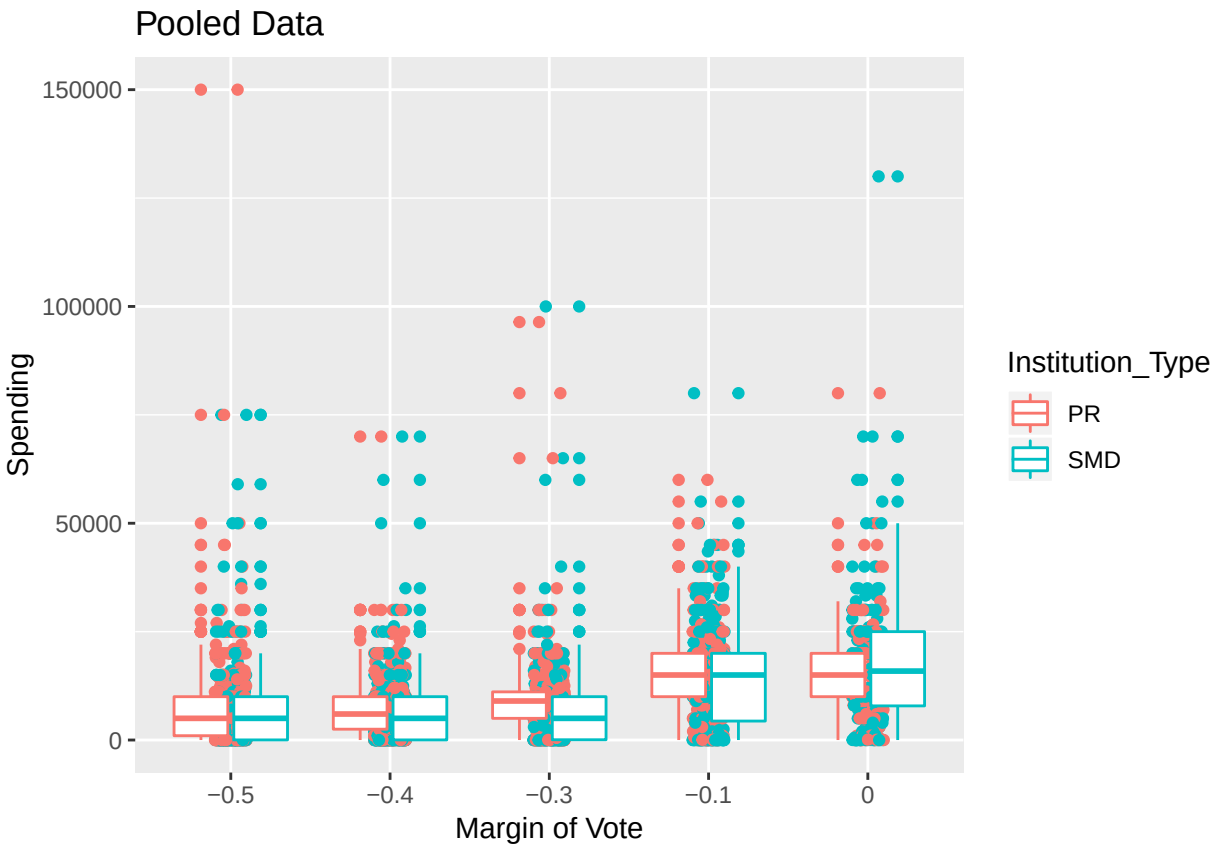

```
#american data boxplot with adjustment
ggplot(amdata, aes(x = factor(Competitiveness), y = District_Spending, color = Institution_Type)) + geo
```




```
#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
150000
100000
50000
0
Spending
−0.5
−0.4
−0.3
−0.1
0
Margin of Vote
Institution_Type
PR
SMD