###Code####

#setting directory

setwd("C:/Users/jmkel/OneDrive - University of Cambridge/Courses/Seminar/coding")

library("ggvis")

library("plyr")

library("dplyr")

library("stringr")

library("tidyr")

library("foreign")

#####reformulating data######

UNCTADdataset<-read.csv("state-investor-dispute.csv",sep=",",header=TRUE,na.string="")

#removing the two signifiers

UNCTADdataset[,1:2]<-list(NULL)

#################Coding variables####################

#####Cording whether a ruling been reached#####

UNCTADdataset<-mutate(UNCTADdataset,ruling=ifelse(IIA\_breaches\_found1 %in% c("Not applicable - settled or discontinued before decision on liability",

"None - jurisdiction declined"), 0,

ifelse(IIA\_breaches\_found1 %in% c("Data not available", "Pending"), NA, 1)))

#Removing NAs

UNCTADdataset<- UNCTADdataset %>% drop\_na(ruling)

UNCTADdataset$ruling <- as.factor(UNCTADdataset$ruling)

#######Amount claimed private#######

UNCTADdataset <- mutate(UNCTADdataset, noamountclaimed = ifelse(claimed\_by\_investor %in% "", "1", "0"))

####tribunal's average experience#####

##total per arbitrator##

#President#

x<-UNCTADdataset %>%

drop\_na(outcome1) %>%

group\_by(president, outcome1) %>%

tally()%>%

group\_by(president) %>%

mutate(pct = n / sum(n))

#Arbitrator claimant#

y <-UNCTADdataset%>%

drop\_na(outcome1) %>%

group\_by(arbitrator\_claimant, outcome1) %>%

tally() %>%

group\_by(arbitrator\_claimant) %>%

mutate(pct = n / sum(n))

#Arbitrator respondent#

z <- UNCTADdataset %>%

drop\_na(outcome1) %>%

group\_by(arbitrator\_respondent, outcome1) %>%

tally() %>%

group\_by(arbitrator\_respondent)

x1 <- merge(x, y, by.x= c("president", "outcome1"), by.y = c("arbitrator\_claimant", "outcome1"), all = TRUE)

x1 <- merge(x1, z, by.x= c("president", "outcome1"), by.y = c("arbitrator\_respondent", "outcome1"), all= TRUE)

x1<- dplyr::filter(x1, !grepl("(replaced)", president))

##total for panel##

x1$total <-rowSums(x1[,c("n.x", "n.y", "n")], na.rm=TRUE)

proportion <- x1 %>%

group\_by(president) %>%

mutate(pct = total / sum(total)) %>%

filter()

##proportion for panel##

proportion <- proportion %>%

group\_by(president) %>%

mutate(label\_y = cumsum(pct)) %>%

filter(outcome1=="1")

A1<- proportion %>%

filter(pct==1) %>%

mutate(outcome1=0, pct=0)

A1$outcome1 <-as.factor(A1$outcome1)

proportion<- rbind(proportion, A1)

##This matches the arbitrator’s experience to each case###

A1<- proportion[, c("president", "pct")]

#If they're the president of the tribunal...

colnames(A1)[colnames(A1)=="pct"] <- "presidentpct"

UNCTADdataset<- merge(UNCTADdataset, A1, by.x = "president", by.y = "president", all.x=TRUE)

#...appointed by the respondent

colnames(x)[colnames(x)=="presidentpct"] <- "arbrespct"

UNCTADdataset<- merge(UNCTADdataset, x, by.x = "arbitrator\_respondent", by.y = "president", all.x=TRUE)

#...or by the claimant

colnames(x)[colnames(x)=="arbrespct"] <- "arbclaim"

UNCTADdataset<- merge(UNCTADdataset, x, by.x = "arbitrator\_claimant", by.y = "president", all.x=TRUE)

#then creates an average score for the tribunal

UNCTADdataset$averageexp <- rowMeans(UNCTADdataset[, c("presidentpct", "arbrespct", "arbclaim")], na.rm = TRUE)

######Sentiment analysis#####

### BRYSBAERT ANALYSIS ###

#https://www.ncbi.nlm.nih.gov/pubmed/23404613#

## AROUSAL ##

ratings <- brys[,c(2,6)]

y<-ratings

arousal <- sapply(UNCTADdataset$summary,simplify=FALSE,FUN=function(x,ratings=y){

#y = ratings

word.list <- strsplit(x,'\\s+')

words <- unlist(word.list)

pos.matches <- match(words,y$Word)

indices <- subset(pos.matches,!is.na(pos.matches))

sentiment.score <- sum(y[indices,2])/length(indices)

return(sentiment.score)

})

df <- data.frame(matrix(unlist(arousal), nrow=nrow(UNCTADdataset), byrow=T))

UNCTADdataset$pos.arousal <- df[,1]

#####company experience#####

UNCTADdataset$company <- str\_extract(UNCTADdataset$SHORT.CASE.NAME, "(.\*)(?= v.)")

x <- UNCTADdataset %>%

group\_by(company) %>%

summarise(length(company))

UNCTADdataset<- merge(UNCTADdataset, x)

#rename

colnames(UNCTADdataset)[colnames(UNCTADdataset)=="length(company)"] <- "companyexp"

###########Coding Arbitrator’s final award########

UNCTADdataset <- mutate(UNCTADdataset, outcome1 = ifelse(outcome %in% "Decided in favour of State", "1",

ifelse(outcome %in% "Decided in favour of investor", "0",

"NA")))

#pro\_firm\_ruling

UNCTADdataset$pro\_firm\_ruling <- UNCTADdataset$outcome

UNCTADdataset$pro\_firm\_ruling <- revalue (UNCTADdataset$outcome, c("Decided in favour of State" = 0, "Decided in favour of investor" = 1,

"Decided in favour of neither party (liability found but no damages awarded)" = 0,

"Discontinued"= NA,

"Pending"= NA,

"Settled"= NA,

"Data not available" = NA

))

#pro\_state\_ruling

UNCTADdataset$pro\_state\_ruling <- UNCTADdataset$outcome

UNCTADdataset$pro\_state\_ruling <- revalue(UNCTADdataset$outcome, c("Decided in favour of State" = 1, "Decided in favour of investor" = 0,

"Decided in favour of neither party (liability found but no damages awarded)" = 1,

"Discontinued"= NA,

"Pending"= NA,

"Settled"= NA,

"Data not available" = NA))

#####Claimed\_by\_investor#####

##Standardising claim amounts (uSD Mln)

#captures all before USD

UNCTADdataset$claimed\_by\_investor <- str\_extract(UNCTADdataset$claimed\_by\_investor, "(.\*)(?= mln USD?)")

#removes other currencies

UNCTADdataset$claimed\_by\_investor <- str\_replace(UNCTADdataset$claimed\_by\_investor, "\\w(.\*)(?=[(])", "")

#gets rid of the bracket

UNCTADdataset$claimed\_by\_investor <- str\_replace(UNCTADdataset$claimed\_by\_investor, "[(]", "")

#####sectors######

#by primary secondary tertiary

UNCTADdataset$threesectormodel <- str\_extract(UNCTADdataset$sector, "(.\*)(?=:)")

UNCTADdataset <- mutate(UNCTADdataset, primary = ifelse(threesectormodel %in% "Primary", "1", "0"))

UNCTADdataset <- mutate(UNCTADdataset, othersector = ifelse(threesectormodel %in% c("Secondary", "Tertiary"), "1", "0"))

UNCTADdataset$primary <- as.factor(UNCTADdataset$primary)

UNCTADdataset$secondary <- as.factor(UNCTADdataset$secondary)

UNCTADdataset$Tertiary <- as.factor(UNCTADdataset$Tertiary)

UNCTADdataset$othersector <- as.factor(UNCTADdataset$othersector)

#####TIME#####

UNCTADdataset <- transform(UNCTADdataset, time = (YEAR.OF.INITIATION) - 1987)

#################Processing data####################

####the following converts the factors to the state needed to be analysed##

UNCTADdataset$ruling<-as.factor(UNCTADdataset$ruling)

UNCTADdataset$noamountclaimed<-as.factor(UNCTADdataset$noamountclaimed)

UNCTADdataset$averageexp<-as.numeric(UNCTADdataset$averageexp)

UNCTADdataset$pos.arousal<-as.numeric(UNCTADdataset$pos.arousal)

UNCTADdataset$companyexp<-as.factor(UNCTADdataset$companyexp)

UNCTADdataset$pro\_firm\_ruling<-as.factor(UNCTADdataset$pro\_firm\_ruling)

UNCTADdataset$pro\_state\_ruling<-as.factor(UNCTADdataset$pro\_state\_ruling)

UNCTADdataset$claimed\_by\_investor<-as.numeric(UNCTADdataset$claimed\_by\_investor)

UNCTADdataset$primary<-as.factor(UNCTADdataset$primary)

UNCTADdataset$othersector<-as.factor(UNCTADdataset$othersector)

UNCTADdataset<-transform(UNCTADdataset, time=(YEAR.OF.INITIATION)-1987)

#############heckman implementation##########

#pro-firm ruling

summary(selection(ruling ~ noamountclaimed+averageexp+pos.arousal+companyexp,

pro\_firm\_ruling ~ claimed\_by\_investor + averageexp + pos.arousal + primary + othersector + time +

noamountclaimed, UNCTADdataset))

#pro-state ruling

summary(selection(ruling ~ noamountclaimed + averageexp + pos.arousal + companyexp,

pro\_state\_ruling ~ claimed\_by\_investor + averageexp + pos.arousal + primary + othersector + time +

noamountclaimed, UNCTADdataset))