```r

# define root path
basis <- ''

library(gdata)
library(XLConnect)
library(utils)

# load socio-economic factors from NOMIS database
gender_density <- read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Resident Population')
age <- read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Age Structure')
ethnic <- read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Ethnic Group')
house <- read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Tenure')
socgrade <- read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Social Grade')
claimants <-read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='Claimant Count Mar 2015')
business<-read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='UK Business Counts 2014')
jobsdensity<-read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='Jobs Density 2013')
educ<-read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Qualifications')
APsheet <- read.xls(paste(basis,'/The Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.xlsx',sep=''),stringsAsFactors=FALSE,sheet='AP sheet')

# normalise data
business$Micro.Perc <- business$Micro..0.to.9./business$Total
business$Small.Perc <- business$Small..10.to.49./business$Total
business$Medium.Perc <- business$Medium.sized..50.to.249./business$Total
business$Large.Perc <- business$Large..250../business$Total
business<-business[,c('Code','Micro.Perc','Small.Perc','Medium.Perc','Large.Perc')]

socgrade$AB.Perc <- socgrade$AB.Higher.and.intermediate.managerial.administrative.professional.occupations/socgrade$All.categories..Approximated.social.grade
socgrade$C1.Perc <- socgrade$C1.Supervisory..clerical.and.junior.managerial.administrative.professional.occupations/socgrade$All.categories..Approximated.social.grade
```

```
31 socgrade$C2.Perc <-
   socgrade$C2.Skilled.manual.occupations/socgrade$All.categories..Approximated.social.gra
   de
32 socgrade$DE.Perc <-
   socgrade$DE.Semi.skilled.and.unskilled.manual.occupations..unemployed.and.lowest.grade.
   occupations/socgrade$All.categories..Approximated.social.grade
33 socgrade<-socgrade[,c('Code','AB.Perc','C1.Perc','C2.Perc','DE.Perc')]
34
35 names(gender_density)[5:7]<-c('Male','Female','Density')
36 gender_density<-gender_density[,c(1:3,5:7)]
37 age <- age[,c(2:4)]
38
39 data_all<-merge(gender_density,age)
40 ethnic<-ethnic[,c(2,6,8,10,12,14)]
41 names(ethnic)<-c('Code','White','Mixed','Asian.British','Black.British','Other')
42 data_all<-merge(data_all,ethnic)
43 names(house)[c(6,8,10,12,14)]<-
   c('Owner.Perc','Shared.Owner','Soc.Rent','Private.Renter','Rent.Free')
44 house <- house[,c(2,6,8,10,12,14)]
45 data_all<-merge(data_all,house)
46 data_all<-merge(data_all,socgrade)
47 names(claimants)[[4]]<-'Claimants'
48 claimants<-claimants[,c(2,4)]
49 data_all<-merge(data_all,claimants)
50 data_all<-merge(data_all,business)
51 jobsdensity<-jobsdensity[,c(2,3)]
52 data_all<-merge(data_all,jobsdensity)
53 educ <- educ[,c(2,4,8,14)]
54 names(educ)<-c('Code','QualNO','GCSE.Max','HE.Above')
55 data_all<-merge(data_all,educ,all.x=TRUE)
56 names(APsheet)[[5]]<-'Code'
57 APsheet <- APsheet[,c(2,3,5)]
58 data_all <- merge(data_all,APsheet)
59 data_all$AP.Name <- sapply(as.character(data_all$AP.Name),function(x)
   gsub('&amp;','&',x),USE.NAMES = FALSE)
60 #
61 # #### get elections data ####
62 #
63 # library(RJSONIO)
64 # results <- fromJSON(url,simplify = FALSE)
65 #
66 # results2015 <- data.frame()
67 # for (result in results) {
68 #
69 #   names <- result[[1]]$name
70 #   winners <- result[[1]]$winningParty
71 #   before <- result[[1]]$sittingParty
72 #
73 #   Votes.C <- as.numeric(result[[1]]$results$C$votes)
74 #   Votes.Lab <- as.numeric(result[[1]]$results$Lab$votes)
75 #   Votes.LD <- as.numeric(result[[1]]$results$LD$votes)
76 #   Votes.UKIP <- as.numeric(result[[1]]$results$UKIP$votes)
77 #   Votes.Green <- as.numeric(result[[1]]$results$Green$votes)
78 #   Votes.SNP <- as.numeric(result[[1]]$results$SNP$votes)
79 #   Votes.DUP <- as.numeric(result[[1]]$results$DUP$votes)
80 #   Votes.C.Share <- as.numeric(result[[1]]$results$C$percentageShare)
81 #   Votes.Lab.Share <- as.numeric(result[[1]]$results$Lab$percentageShare)
```

```
#   Votes.LD.Share <- as.numeric(result[[1]]$results$LD$percentageShare)
#   Votes.UKIP.Share <- as.numeric(result[[1]]$results$UKIP$percentageShare)
#   Votes.Green.Share <- as.numeric(result[[1]]$results$Green$percentageShare)
#   Votes.SNP.Share <- as.numeric(result[[1]]$results$SNP$percentageShare)
#   Votes.DUP.Share <- as.numeric(result[[1]]$results$DUP$percentageShare)
#
#   if (length(Votes.UKIP)==0) {Votes.UKIP=0}
#   if (length(Votes.UKIP.Share)==0) {Votes.UKIP.Share=0}
#   if (length(Votes.Lab)==0) {Votes.Lab=0}
#   if (length(Votes.Lab.Share)==0) {Votes.Lab.Share=0}
#   if (length(Votes.C)==0) {Votes.C=0}
#   if (length(Votes.C.Share)==0) {Votes.C.Share=0}
#   if (length(Votes.LD)==0) {Votes.LD=0}
#   if (length(Votes.LD.Share)==0) {Votes.LD.Share=0}
#   if (length(Votes.Green)==0) {Votes.Green=0}
#   if (length(Votes.Green.Share)==0) {Votes.Green.Share=0}
#   if (length(Votes.SNP)==0) {Votes.SNP=0}
#   if (length(Votes.SNP.Share)==0) {Votes.SNP.Share=0}
#   if (length(Votes.DUP)==0) {Votes.DUP=0}
#   if (length(Votes.DUP.Share)==0) {Votes.DUP.Share=0}
#
#   tots_temp <-
Votes.SNP.Share+Votes.Green.Share+Votes.LD.Share+Votes.C.Share+Votes.Lab.Share+Votes.UK
IP.Share+Votes.DUP.Share
#   tots_temp_2 <-
Votes.SNP+Votes.Green+Votes.LD+Votes.C+Votes.Lab+Votes.UKIP+Votes.DUP
#   tots_others <- (100*tots_temp_2/tots_temp)-tots_temp_2
#   tots_others_perc <- 100*tots_others/(100*tots_temp_2/tots_temp)
#
#   temp <-
data.frame(Names=names,X2015.result=winners,X2010.result=before,Votes.C=Votes.C,
#
Votes.Lab=Votes.Lab,Votes.LD=Votes.LD,Votes.UKIP=Votes.UKIP,Votes.Green=Votes.Green,Vot
es.SNP=Votes.SNP,
#
Votes.C.Share=Votes.C.Share,Votes.Lab.Share=Votes.Lab.Share,Votes.LD.Share=Votes.LD.Sha
re,
#
Votes.UKIP.Share=Votes.UKIP.Share,Votes.Green.Share=Votes.Green.Share,Votes.SNP.Share=V
otes.SNP.Share,Votes.Others=tots_others,Votes.Others.Share=tots_others_perc)
#   if (nrow(results2015)>0) {
#     results2015<-rbind(results2015,temp)
#   } else {
#     results2015<-temp
#   }
#
# }
#
#
# results2015$X2015.result<-as.character(results2015$X2015.result)
# results2015$Names<-as.character(results2015$Names)
# results2015$X2010.result<-as.character(results2015$X2010.result)
# nrow(results2015)
# # fix the parties names
# parties <- c('Conservative','Labour','Liberal Democrats','Ukip','Scottish National
Party')
# results2015$X2015.result[results2015$X2015.result=='C']<-'Conservative'
```

```
# results2015$X2015.result[results2015$X2015.result=='Lab']<-'Labour'
# results2015$X2015.result[results2015$X2015.result=='LD']<-'Liberal Democrats'
# results2015$X2015.result[results2015$X2015.result=='UKIP']<-'Ukip'
# results2015$X2015.result[results2015$X2015.result=='SNP']<-'Scottish National Party'
# results2015$X2015.result[!(results2015$X2015.result %in% parties)] <- 'Other'
#
# results2015$X2010.result[results2015$X2010.result=='C']<-'Conservative'
# results2015$X2010.result[results2015$X2010.result=='Lab']<-'Labour'
# results2015$X2010.result[results2015$X2010.result=='LD']<-'Liberal Democrats'
# results2015$X2010.result[results2015$X2010.result=='UKIP']<-'Ukip'
# results2015$X2010.result[results2015$X2010.result=='SNP']<-'Scottish National Party'
# results2015$X2010.result[!(results2015$X2010.result %in% parties)] <- 'Other'
#
# names(results2015)[[1]]<-'AP.Name'
# data_all<-merge(data_all,results2015,all.x=TRUE)
# data_all[is.na(data_all$X2015.result),]
#
# # match new results to codes data
# names(results2010)
# table(results2015$X2015.result)
# table(results2015[,c('X2015.result','X2010.result')])
# sum(table(results2015$X2015.result))
#
# table(data_all[,c('X2015.result','Region')])


#### get turnout by const #####

load(paste(basis,'/The Times/Elections/PostElection/turnout.RData',sep=''))
turnout2010<-read.csv(paste(basis,'/The
Times/Elections/PostElection/turnout2010.csv',sep=''))

names(turnout_data)[[2]]<-'Turnout.2015'
data_all<-merge(data_all,turnout_data)

data_all$Electorate <- 100*
(data_all$Votes.C+data_all$Votes.Lab+data_all$Votes.LD+data_all$Votes.UKIP+data_all$Votes.Green+

 data_all$Votes.SNP+data_all$Votes.Others)/data_all$Turnout.2015

data_all<-merge(data_all,turnout2010)

######## TUNROUT ANALYSIS #######

data_all$ElectorateHasIncreased <- data_all$Electorate>data_all$Votes2010
paste(sum(data_all$Electorate),sum(data_all$Votes2010),sep=',')
data_all$TurnoutHasIncreased <- data_all$Turnout.2015>data_all$Turnout.2010
data_all$AllVoted.2015 <- data_all$Electorate*data_all$Turnout.2015/100
data_all$AllVoted.2010 <- data_all$Votes2010*data_all$Turnout.2010/100
paste(sum(data_all$AllVoted.2015),sum(data_all$AllVoted.2010),sep=',')
summary(data_all$Turnout.2015)

data_all$TurnoutDifference <- data_all$Turnout.2015-data_all$Turnout.2010
table(data_all[,c('ElectorateHasIncreased','X2015.result')])
table(data_all[,c('TurnoutHasIncreased','X2015.result')])
aggregate(TurnoutDifference~X2015.result,median,data=data_all)
```

```r
aggregate(TurnoutDifference~X2015.result,mean,data=data_all)
aggregate(TurnoutDifference~X2015.result,sd,data=data_all)
data_all$SwingSeats <- data_all$X2015.result!=data_all$X2010.result
quantile(data_all$TurnoutDifference,seq(0,1,0.1))
data_all$Engaged <- 'Normal'
data_all$Engaged[data_all$TurnoutDifference>3] <- 'Engaged'
data_all$Engaged[data_all$TurnoutDifference<=-1.1] <- 'Disengaged'
table(data_all[,c('X2015.result','Engaged')])
table(data_all[,c('SwingSeats','Engaged')])
temp<-data_all[data_all$Engaged=='Disengaged' & data_all$SwingSeats==TRUE,]

data_structured <- data_all[,c(34,42:47,49,50)]
library(reshape)
data_structured<-melt(data_structured,id=c(1,9))

data_structured <- data_all[,c(34,42:47,49,50,30)]
data_structured<-melt(data_structured,id=c(1,9,10))

data_structured <- data_all[,c(34,42:47,49,50,53)]
data_structured<-melt(data_structured,id=c(1,9,10))
data_structured$TurnoutDifference<-data_structured$Turnout.2015-
data_structured$Turnout.2010

#### prepare data for tree ####

node.fun1 <- function(x, labs, digits, varlen)
{
  paste(labs,paste(',',paste((x$frame$n-x$frame$dev),paste("over", x$frame$n))))

}

data_all$Mean.Age<-NULL
names(data_all)[[8]] <- 'Age'
names(data_all)[24:25] <- c('Micro.Bsns','Small.Bsns')
data_all$Big.Bsns <- data_all$Medium.Perc+data_all$Large.Perc
data_all$OneMinusSmallBsns <- data_all$Medium.Perc+data_all$Large.Perc
names(data_all)[19:22]<-c('AB','C1','C2','DE')
names(data_all)[[13]]<-'Other.Eth'
data_all$Owner <- data_all$Owner.Perc+data_all$Shared.Owner
data_all$Owner.Perc<-NULL
data_all$Shared.Owner<-NULL
data_all$Region[data_all$Region=='North East' | data_all$Region=='North West' |
data_all$Region=='Yorkshire and The Humber'] <- 'North'
data_all$Region[data_all$Region=='West Midlands' | data_all$Region=='East Midlands'] <-
'Midlands'
data_all$Region[data_all$Region=='South West' | data_all$Region=='South East'] <-
'South'
data_all$Scotland<-'RestUK'
data_all$Scotland[data_all$Region=='Scotland']<-'Scotland'
data_all$SmMcr.Bsns <- data_all$Micro.Bsns+data_all$Small.Bsns
data_all$Prv.Rent <- data_all$Private.Renter+data_all$Rent.Free
data_all$NS.Renter <- data_all$Private.Renter+data_all$Rent.Free


scotland_data <- data_all[data_all$Region=='Scotland',]
summary(scotland_data)

```

```r
library(rpart)
library(rpart.plot)

######## Explore trees as results come in ###########

# without scotland
tree <- rpart(X2015.result~Soc.Rent+GCSE.Max+Male+Female+Density+Age+
                Claimants,data=data_all[data_all$Region!='Scotland',],
              control=rpart.control(minsplit=2,
cp=0.0001,minbucket=3,maxdepth=3),method='class')

prp(tree, extra=100, under=T, yesno=T, node.fun=node.fun1,main ="All Constituencies (No Scotland)",branch.type=5)
print(tree)

#### Swing ####

tree <-
rpart(X2015.result~Male+Female+Density+Age+White+Mixed+Asian.British+Black.British+Other.Eth+

 Scotland+Big.Bsns+Claimants,data=data_all[data_all$X2010.result=='Liberal Democrats',],
              control=rpart.control(minsplit=3, cp=0.01,minbucket=3))

prp(tree, extra=100, under=T, yesno=F, node.fun=node.fun1,main ="Lib Dem in 2010",branch.type=5)
```