```r
1
2
3  # define root path
4  basis <- ''
5
6  library(gdata)
7  library(XLConnect)
8  library(utils)
9
10 # load socio-economic factors from NOMIS database
11 gender_density <- read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Resident Population')
12 age <- read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Age Structure')
13 ethnic <- read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Ethnic Group')
14 house <- read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Tenure')
15 socgrade <- read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Social Grade')
16 claimants <-read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='Claimant Count Mar 2015')
17 business<-read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='UK Business Counts 2014')
18 jobsdensity<-read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='Jobs Density 2013')
19 educ<-read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='Census 2011 Qualifications')
20 APsheet <- read.xls(paste(basis,'/The
   Times/Elections/PostElection/20150429_PostElectionsAnalysis_NOMIS_DDJ_DecisionTreeData.
   xlsx',sep=''),stringsAsFactors=FALSE,sheet='AP sheet')
21
22 # normalise data
23 business$Micro.Perc <- business$Micro..0.to.9./business$Total
24 business$Small.Perc <- business$Small..10.to.49./business$Total
25 business$Medium.Perc <- business$Medium.sized..50.to.249./business$Total
26 business$Large.Perc <- business$Large..250../business$Total
27 business<-business[,c('Code','Micro.Perc','Small.Perc','Medium.Perc','Large.Perc')]
28
29 socgrade$AB.Perc <-
   socgrade$AB.Higher.and.intermediate.managerial.administrative.professional.occupations/
   socgrade$All.categories..Approximated.social.grade
30 socgrade$C1.Perc <-
   socgrade$C1.Supervisory..clerical.and.junior.managerial.administrative.professional.occ
   upations/socgrade$All.categories..Approximated.social.grade
```

```r
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)
```

```
 82  #     Votes.LD.Share <- as.numeric(result[[1]]$results$LD$percentageShare)
 83  #     Votes.UKIP.Share <- as.numeric(result[[1]]$results$UKIP$percentageShare)
 84  #     Votes.Green.Share <- as.numeric(result[[1]]$results$Green$percentageShare)
 85  #     Votes.SNP.Share <- as.numeric(result[[1]]$results$SNP$percentageShare)
 86  #     Votes.DUP.Share <- as.numeric(result[[1]]$results$DUP$percentageShare)
 87  #
 88  #     if (length(Votes.UKIP)==0) {Votes.UKIP=0}
 89  #     if (length(Votes.UKIP.Share)==0) {Votes.UKIP.Share=0}
 90  #     if (length(Votes.Lab)==0) {Votes.Lab=0}
 91  #     if (length(Votes.Lab.Share)==0) {Votes.Lab.Share=0}
 92  #     if (length(Votes.C)==0) {Votes.C=0}
 93  #     if (length(Votes.C.Share)==0) {Votes.C.Share=0}
 94  #     if (length(Votes.LD)==0) {Votes.LD=0}
 95  #     if (length(Votes.LD.Share)==0) {Votes.LD.Share=0}
 96  #     if (length(Votes.Green)==0) {Votes.Green=0}
 97  #     if (length(Votes.Green.Share)==0) {Votes.Green.Share=0}
 98  #     if (length(Votes.SNP)==0) {Votes.SNP=0}
 99  #     if (length(Votes.SNP.Share)==0) {Votes.SNP.Share=0}
100  #     if (length(Votes.DUP)==0) {Votes.DUP=0}
101  #     if (length(Votes.DUP.Share)==0) {Votes.DUP.Share=0}
102  #
103  #     tots_temp <-
     Votes.SNP.Share+Votes.Green.Share+Votes.LD.Share+Votes.C.Share+Votes.Lab.Share+Votes.UK
     IP.Share+Votes.DUP.Share
104  #     tots_temp_2 <-
     Votes.SNP+Votes.Green+Votes.LD+Votes.C+Votes.Lab+Votes.UKIP+Votes.DUP
105  #     tots_others <- (100*tots_temp_2/tots_temp)-tots_temp_2
106  #     tots_others_perc <- 100*tots_others/(100*tots_temp_2/tots_temp)
107  #
108  #     temp <-
     data.frame(Names=names,X2015.result=winners,X2010.result=before,Votes.C=Votes.C,
109  #
     Votes.Lab=Votes.Lab,Votes.LD=Votes.LD,Votes.UKIP=Votes.UKIP,Votes.Green=Votes.Green,Vot
     es.SNP=Votes.SNP,
110  #
     Votes.C.Share=Votes.C.Share,Votes.Lab.Share=Votes.Lab.Share,Votes.LD.Share=Votes.LD.Sha
     re,
111  #
     Votes.UKIP.Share=Votes.UKIP.Share,Votes.Green.Share=Votes.Green.Share,Votes.SNP.Share=V
     otes.SNP.Share,Votes.Others=tots_others,Votes.Others.Share=tots_others_perc)
112  #     if (nrow(results2015)>0) {
113  #        results2015<-rbind(results2015,temp)
114  #     } else {
115  #        results2015<-temp
116  #     }
117  #
118  # }
119  #
120  #
121  # results2015$X2015.result<-as.character(results2015$X2015.result)
122  # results2015$Names<-as.character(results2015$Names)
123  # results2015$X2010.result<-as.character(results2015$X2010.result)
124  # nrow(results2015)
125  # # fix the parties names
126  # parties <- c('Conservative','Labour','Liberal Democrats','Ukip','Scottish National
     Party')
127  # results2015$X2015.result[results2015$X2015.result=='C']<-'Conservative'
```

```
128  # results2015$X2015.result[results2015$X2015.result=='Lab']<-'Labour'
129  # results2015$X2015.result[results2015$X2015.result=='LD']<-'Liberal Democrats'
130  # results2015$X2015.result[results2015$X2015.result=='UKIP']<-'Ukip'
131  # results2015$X2015.result[results2015$X2015.result=='SNP']<-'Scottish National Party'
132  # results2015$X2015.result[!(results2015$X2015.result %in% parties)] <- 'Other'
133  #
134  # results2015$X2010.result[results2015$X2010.result=='C']<-'Conservative'
135  # results2015$X2010.result[results2015$X2010.result=='Lab']<-'Labour'
136  # results2015$X2010.result[results2015$X2010.result=='LD']<-'Liberal Democrats'
137  # results2015$X2010.result[results2015$X2010.result=='UKIP']<-'Ukip'
138  # results2015$X2010.result[results2015$X2010.result=='SNP']<-'Scottish National Party'
139  # results2015$X2010.result[!(results2015$X2010.result %in% parties)] <- 'Other'
140  #
141  # names(results2015)[[1]]<-'AP.Name'
142  # data_all<-merge(data_all,results2015,all.x=TRUE)
143  # data_all[is.na(data_all$X2015.result),]
144  #
145  # # match new results to codes data
146  # names(results2010)
147  # table(results2015$X2015.result)
148  # table(results2015[,c('X2015.result','X2010.result')])
149  # sum(table(results2015$X2015.result))
150  #
151  # table(data_all[,c('X2015.result','Region')])
152
153
154  #### get turnout by const #####
155
156  load(paste(basis,'/The Times/Elections/PostElection/turnout.RData',sep=''))
157  turnout2010<-read.csv(paste(basis,'/The
     Times/Elections/PostElection/turnout2010.csv',sep=''))
158
159  names(turnout_data)[[2]]<-'Turnout.2015'
160  data_all<-merge(data_all,turnout_data)
161
162  data_all$Electorate <- 100*
     (data_all$Votes.C+data_all$Votes.Lab+data_all$Votes.LD+data_all$Votes.UKIP+data_all$Vot
     es.Green+
163
      data_all$Votes.SNP+data_all$Votes.Others)/data_all$Turnout.2015
164
165  data_all<-merge(data_all,turnout2010)
166
167  ######## TUNROUT ANALYSIS #######
168
169  data_all$ElectorateHasIncreased <- data_all$Electorate>data_all$Votes2010
170  paste(sum(data_all$Electorate),sum(data_all$Votes2010),sep=',')
171  data_all$TurnoutHasIncreased <- data_all$Turnout.2015>data_all$Turnout.2010
172  data_all$AllVoted.2015 <- data_all$Electorate*data_all$Turnout.2015/100
173  data_all$AllVoted.2010 <- data_all$Votes2010*data_all$Turnout.2010/100
174  paste(sum(data_all$AllVoted.2015),sum(data_all$AllVoted.2010),sep=',')
175  summary(data_all$Turnout.2015)
176
177  data_all$TurnoutDifference <- data_all$Turnout.2015-data_all$Turnout.2010
178  table(data_all[,c('ElectorateHasIncreased','X2015.result')])
179  table(data_all[,c('TurnoutHasIncreased','X2015.result')])
180  aggregate(TurnoutDifference~X2015.result,median,data=data_all)
```

```r
181  aggregate(TurnoutDifference~X2015.result,mean,data=data_all)
182  aggregate(TurnoutDifference~X2015.result,sd,data=data_all)
183  data_all$SwingSeats <- data_all$X2015.result!=data_all$X2010.result
184  quantile(data_all$TurnoutDifference,seq(0,1,0.1))
185  data_all$Engaged <- 'Normal'
186  data_all$Engaged[data_all$TurnoutDifference>3] <- 'Engaged'
187  data_all$Engaged[data_all$TurnoutDifference<=-1.1] <- 'Disengaged'
188  table(data_all[,c('X2015.result','Engaged')])
189  table(data_all[,c('SwingSeats','Engaged')])
190  temp<-data_all[data_all$Engaged=='Disengaged' & data_all$SwingSeats==TRUE,]
191
192  data_structured <- data_all[,c(34,42:47,49,50)]
193  library(reshape)
194  data_structured<-melt(data_structured,id=c(1,9))
195
196  data_structured <- data_all[,c(34,42:47,49,50,30)]
197  data_structured<-melt(data_structured,id=c(1,9,10))
198
199  data_structured <- data_all[,c(34,42:47,49,50,53)]
200  data_structured<-melt(data_structured,id=c(1,9,10))
201  data_structured$TurnoutDifference<-data_structured$Turnout.2015-
     data_structured$Turnout.2010
202
203  #### prepare data for tree ####
204
205  node.fun1 <- function(x, labs, digits, varlen)
206  {
207    paste(labs,paste(',',paste((x$frame$n-x$frame$dev),paste("over", x$frame$n))))
208
209  }
210
211  data_all$Mean.Age<-NULL
212  names(data_all)[[8]] <- 'Age'
213  names(data_all)[24:25] <- c('Micro.Bsns','Small.Bsns')
214  data_all$Big.Bsns <- data_all$Medium.Perc+data_all$Large.Perc
215  data_all$OneMinusSmallBsns <- data_all$Medium.Perc+data_all$Large.Perc
216  names(data_all)[19:22]<-c('AB','C1','C2','DE')
217  names(data_all)[[13]]<-'Other.Eth'
218  data_all$Owner <- data_all$Owner.Perc+data_all$Shared.Owner
219  data_all$Owner.Perc<-NULL
220  data_all$Shared.Owner<-NULL
221  data_all$Region[data_all$Region=='North East' | data_all$Region=='North West' |
     data_all$Region=='Yorkshire and The Humber'] <- 'North'
222  data_all$Region[data_all$Region=='West Midlands' | data_all$Region=='East Midlands'] <-
     'Midlands'
223  data_all$Region[data_all$Region=='South West' | data_all$Region=='South East'] <-
     'South'
224  data_all$Scotland<-'RestUK'
225  data_all$Scotland[data_all$Region=='Scotland']<-'Scotland'
226  data_all$SmMcr.Bsns <- data_all$Micro.Bsns+data_all$Small.Bsns
227  data_all$Prv.Rent <- data_all$Private.Renter+data_all$Rent.Free
228  data_all$NS.Renter <- data_all$Private.Renter+data_all$Rent.Free
229
230
231  scotland_data <- data_all[data_all$Region=='Scotland',]
232  summary(scotland_data)
233
```

```
234  library(rpart)
235  library(rpart.plot)
236
237  ######## Explore trees as results come in ###########
238
239  # without scotland
240  tree <- rpart(X2015.result~Soc.Rent+GCSE.Max+Male+Female+Density+Age+
241                  Claimants,data=data_all[data_all$Region!='Scotland',],
242              control=rpart.control(minsplit=2,
     cp=0.0001,minbucket=3,maxdepth=3),method='class')
243
244  prp(tree, extra=100, under=T, yesno=T, node.fun=node.fun1,main ="All Constituencies (No
     Scotland)",branch.type=5)
245  print(tree)
246
247  #### Swing ####
248
249  tree <-
     rpart(X2015.result~Male+Female+Density+Age+White+Mixed+Asian.British+Black.British+Othe
     r.Eth+
250
      Scotland+Big.Bsns+Claimants,data=data_all[data_all$X2010.result=='Liberal
     Democrats',],
251              control=rpart.control(minsplit=3, cp=0.01,minbucket=3))
252
253  prp(tree, extra=100, under=T, yesno=F, node.fun=node.fun1,main ="Lib Dem in
     2010",branch.type=5)
254
255
256
257
```