FE2015 - project6: explore dataset

Duc-Quang Nguyen

23 Aug 2015

Load libraries

```
library(dplyr)
library(ggplot2)
library(scales)
```

Settings

```
candidates.file <- "data/candidates_socio-demographics_2015-08-20.csv" #candidates data file

# which columns should be a factor
candidates.factorCols <- c('gender', 'city', 'country',
  'language', 'party_short', 'party_REC', 'district')

# residence related data
zipcode.file <- "data/indicators/PLZO_CSV_LV03.csv"
communePortrait.file <- "data/indicators/communesCH_2015_indicators_je-f-21.03.01.csv"
communeClass.file <- "data/indicators/be-b-00.04-rgs-01_urbanRural.csv"

# set ggplot theme
theme_set(theme_bw())
```

Load the data

```
## Warning: There are: 14 candidates with no age valid age! They will be
## replaced by NA
```

```
## Warning: 1 given zip does not have a BFS commune code!
## This represents: 0.082 %of the whole data
```

Subset the whole candidates data and combine with indicators

```
data <- data.read %>% select(one_of(c('ID_user', 'ID_Candidate', 'firstname', 'lastname', 'gender', 'yea
    'age', 'zip', 'city', 'country', 'party_short', 'party_REC', 'district', 'language')))

data$bfs <- qr[match(data$zip, qr$zip),'bfs']
# set to NA bfs code for candidates living not in Switzerland
data[which(data$country != "Schweiz"),'bfs'] <- NA

# add rural/countryside BFS classification
```

```
data$communeUrbanClass <- communeClass[match(data$bfs, communeClass[,1]), "Régions urbaines / rurales 2(

# add communal portrait data
data$communeBuildingSurface <- communePortrait[match(data$bfs, communePortrait[,1]), "Surfaces d'habita
data$communeForeigner <- communePortrait[match(data$bfs, communePortrait[,1]), "Etrangers en %"]
```