Compile data for Rio de Janeiro Municipality between 2012-2016

Marcelo Ferreira da Costa Gomes

3 de fevereiro de 2017

Loading data

We'll use dengue data from Rio de Janeiro Municipality from 2012 to 2016, taken from the DENGEON database.

```
library(foreign)
target.cols <- c('NU_NOTIFIC', 'ID_MUNICIP', 'ID_UNIDADE', 'DT_NOTIFIC', 'DT_DIGITA')
flist <- list.files('./data/', pattern='*.dbf$', full.names = T)
df.dbf <- read.dbf(flist[1], as.is=T)[, target.cols]
for (fname in flist[2:length(flist)]){
  df.dbf <- rbind(df.dbf, read.dbf(fname, as.is=T)[, target.cols])
}
geocod <- 330455
df.dbf <- df.dbf[df.dbf$ID_MUNICIP == geocod, ]
```

Filter by list of columns of interest and drop possible duplicates:

```
filter.cols <- c('NU_NOTIFIC', 'ID_MUNICIP', 'ID_UNIDADE', 'DT_NOTIFIC')
df.dbf.clean <- df.dbf[(!duplicated(df.dbf[, filter.cols]) &
                                !is.na(df.dbf$DT_DIGITA)), ]

nrow(df.dbf.clean)
```

```
## [1] 310341
```

Next we are going to generate epidemiological week information to both notification and digitization dates

```
require(leos.opportunity.estimator)
# Create columns with epiweek, epiyear, and epiyearweek for notification and digitalization ones:
target.cols <- c('DT_NOTIFIC_epiyearweek', 'DT_NOTIFIC_epiweek', 'DT_NOTIFIC_epiyear')
if (!all(target.cols %in% names(df.dbf.clean))){
  df.dbf.clean <- generate.columns.from.date(df.dbf.clean, 'DT_NOTIFIC')
  names(df.dbf.clean) <- sub("^epi", "DT_NOTIFIC_epi", names(df.dbf.clean))
}
target.cols <- c('DT_DIGITA_epiyearweek', 'DT_DIGITA_epiweek', 'DT_DIGITA_epiyear')
if (!all(target.cols %in% names(df.dbf.clean))){
  df.dbf.clean <- generate.columns.from.date(df.dbf.clean, 'DT_DIGITA')
  names(df.dbf.clean) <- sub("^epi", "DT_DIGITA_epi", names(df.dbf.clean))
}
names(df.dbf.clean)
```

```
##  [1] "NU_NOTIFIC"             "ID_MUNICIP"
##  [3] "ID_UNIDADE"             "DT_NOTIFIC"
##  [5] "DT_DIGITA"              "DT_NOTIFIC_epiyearweek"
##  [7] "DT_NOTIFIC_epiweek"     "DT_NOTIFIC_epiyear"
##  [9] "DT_DIGITA_epiyearweek"  "DT_DIGITA_epiweek"
## [11] "DT_DIGITA_epiyear"
```

Save data on disk:

```
saveRDS(df.dbf.clean, file='data/dengue.munRJ.dbf.2012.2016.epiyearweek.rds')
```