Cleaning - Agregador - V1

Natalia Tosi

10/16/2021

1. Cleaning Brazil Data (saved as Raw)

```
#Importing raw data
d <- read_csv("DATA/Popularity-LatinAmerica-BR.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character(),
##   day = col_double(),
##   year = col_double(),
##   Positive = col_double(),
##   Negative = col_double(),
##   `DK-Neutral` = col_double(),
##   Neutral = col_double(),
##   DK = col_double(),
##   SampleSize = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
#Or read from saved as windows csv, so read with latin1. If saving as (mac) csv,
# read with macroman
#d <-read.csv(paste("./DATA/_popularity_BR_25mar21.csv",sep=""),
#            fileEncoding="latin1",na.strings = ".")

dd <- subset(d, President != "Figueiredo")

#Get rid of spaces in pollster names
d$Institute <- gsub("\\s","\\.",d$Institute,perl=T)

#Use short presidential names, and rder factors cronologically
d$PresidentS <- factor(toupper(d$President),levels=c("FIGUEIREDO","SARNEY","COLLOR",
                                                     "FRANCO","CARDOSO","LULA","DILMA"
                                                     ,"TEMER","BOLSONARO"))

#Adjust IBOPE (IPEC is former IBOPE)
d$Institute <- gsub("IPEC-ExIBOPE","IBOPE",
                    gsub("Ibope","IBOPE",d$Institute))

#Check for missing data in relevant vars
```

```
tmp <- apply(is.na(subset(d,select=c(raw.date,President,PresidentS,Positive,Institute))),
             2,sum)
if(sum(tmp)>0){cat("Attention! Data missing in source\n");print(tmp)}

#Fix dates #
#Always a problem in Xls/csv
d$day[which(is.na(d$day))]<-15
d$month <- gsub("ago","aug",d$month)
d$month <- gsub("mai","may",d$month)
d$date <- as.Date(paste(d$year,d$month,d$day,sep="-"),format="%Y-%b-%d")
d$month <- d$day <-  NULL

d<-d[sort(as.character(d$date),index.return=TRUE)$ix,] #sort by date
d$Q <- paste(substr(d$date,1,5),quarters(d$date),sep="") #quarter indicator
d$Q <- gsub("Q","",d$Q)
d$M <- substr(d$date,1,7)#month indicator
d$raw.date <- NULL


save(d, file="R/popularity_raw_BR.RData")
write.csv(d,'DATA/popularity_raw_BR.csv')

#fork: save Bolsonaro's popularity separately
dd <- subset(d,President=="Bolsonaro"&year>=2019)
save(dd,file="R/popularity_raw_bolsonaro_BR.RData")
write.csv(dd,'DATA/popularity_raw_bolsonaro_BR.csv')
```