Process for downloading variables

Pedro Girardi

December 11, 2018

PME

Use the package `lodown` to download the variables. This is the process recommended by IBGE on its website. This process saves several .rds files.

```r
library(lodown)
# examine all available PME microdata files
pme_cat <-
  get_catalog( "pme" ,
               output_dir = file.path( path.expand( "Data/" ) , "PME" ) )

# 2016 only
pme_cat <- subset( pme_cat , year < 2010 )
# download the microdata to your local computer
pme_cat <- lodown( "pme" , pme_cat,
                   output_dir = file.path( path.expand( "Data/" ) , "Data/PME" ))

# or for complete data:
lodown( "pme" ,output_dir = file.path( path.expand( "Data/" ) , "PME" ) )
```

Once you have the .rds files for the years required, I convert them to .csv files. This saved me some space so I could upload to mirage.

```r
reduced_list = list.files(pattern = ".rds")
reduced_df = map_dfr(reduced_list, readRDS)
fwrite(reduced_df, "~/Comps/reduced_df1.csv")
```

The next part can be computationally intensive. The goal is to read the .csvs into R, remove all the variables that are not of interest, and then finally merge all the datasets into one. The goal here is to have a dataset that is ready to run a simple regression. One potential issue can be to trace the same people over time. PUCRIO, a Brazilian university, has a code in STATA that takes care of that.

```r
# set directory to where all the current .csvs are located
# create a new directory where the new .csvs will be located
# read every old .csv, eliminate all variables that are not of interest
# write new csv (hopefully much smaller) in new directory
# merge all the csv files, reading them in R and binding the rows
```

Inflation

Manually download .csv from IBGE website. There will be multiple files for different years. Check why IBGE has different files. The code below should suffice to write a narrow csv for every inflation file. The narrow .csv can then be matched to specific individual observations in the PME file.

```r
library(tidyverse)
inflationData = read.csv("inflacaoMetro.csv", skip = 3)[-1,]
inflationData = inflationData[1:8,]
inflationData = inflationData %>% gather(2:ncol(inflationData), key = "Month", value = "Inflation")
colnames(inflationData)[1] = "Region"
inflationData = arrange(inflationData,Region)
```

```r
monthsPortuguese = c("jan", "fev", "mar", "abr", "mai",
                     "jun", "jul", "ago", "set", "out",
                     "nov", "dez")
inflationData = inflationData %>% mutate(year=as.integer(str_sub(Month, -4,-1)),
                                         Month=str_sub(Month, 1,3))
inflationData = inflationData %>% mutate(Month = match(inflationData$Month, monthsPortuguese))

write.csv(inflationData, "inflationCleaned.csv")
```

GDP Growth

Likely a similar process to inflation