Outline IHME Lifetable Data and Materials

Lucas Helal

2023-05-11

This is the raw documentation to guide the steps which are done by me (Lucas Helal), as a part of an training-exploratory analysis of data wrangling and GBD data. This is important to highlight because that, given the nature of training for big data sets manipulation, no *a priori* hypothesis were specified neither a statistical analysis plan.

Please do not hesitate to contact me further for any reason.

1.**All that you need to begin!**

1.1. Raw data sets:

The raw data sets were downloaded one by one in the IHME/GHDx website, through the following URL:

`https://ghdx.healthdata.org/record/ihme-data/gbd-2019-life-tables-1950-2019>`

All the 24 raw data sets, in `".csv`" file, can be found at my Zenodo repository with a `"README.txt`" file explaining what they are, under the DOI 10.5072/zenodo.1200943. It is a nearly 1GB panel data data set so I recommend to run them directly in your statistical software if you want.

1.2. Processed data set:

You can find my manipulated data set in my GitHub repository, also written as `".csv`“, where you will find other important documents.

`https://github.com/lhelal1/git-remote-2/life-tables-project>`

1.3. Software and packages:

All procedures were done in R for MacOS (Monterey 12.5, Intel Core i7, 16GB RAM, 1TB SSD). R version was R version 4.2.3 (2023-03-15) -- “Shortstop Beagle Platform: x86_64-apple-darwin17.0 (64-bit); and R Studio version was Version 2023.03.0+386 (2023.03.0+386). All details of installed packages, dependencies, versions were extracted through – `"renv"` package and were locked before `"commit`”.

The link to install the R software is displayed below, via R webpage and the Fiocruz (BR) CRAN:

`https://cran.fiocruz.br/bin/macosx/big-sur-x86_64/base/R-4.3.0-x86_64.pkg`

You may also want to download the very same version of RStudio I'm using:

`https://download1.rstudio.org/electron/macos/RStudio-2023.03.0-386.dmg`

For packages, you will find all of what is on-board shared in GitHub as well and, for each script, I will signalize what are the necessary packages to reproduce what was done.

1.4. Scripts:

All scripts are shared in a time-line manner, accompanied by this raw "`markdown`" as guidance for analysis step-by-step.

2.**Analysis**

2.1. For analysis, you should first run the "`renv`" package to load all libraries and depencies that I'm using. If it doesn't work, it will pointed out in each script/markdown I'm sharing with you. The command to be used should be "`renv::init()`", just in case (or "`renv::restore()`" after you have uploaded all the necessary stuff - including the package).

2.2. You will find several scripts and datasets. As they were a product of an exploratory analysis, please do not hesitate in case of a data set not match with your directory herein.

2.3. All of the necessary stuff (additional datasets, manipulated datasets, plots previously printed are in an specific folder in my GitHub).

Please have fun and I hope you have at least a minimal reproducible experience. I promise to be better the next (and suppose from the very beginning it will be shared with collaborators/third parties.)

PPG Epidemiologia (UFRGS, Brazil) | PPG Cardiologia (UFRGS, Brazil)

lucas.helal@ufrgs.br | lhelal@hcpa.edu.br | lucas@lucashelal.me

+ 55 51 99991 5551