Reading Human Fertility Database and Human Mortality Database data into R

Tim Riffe

Department of Demography, University of California, Berkeley

May 14, 2015

Abstract

This is about

1 Motivation

The Human Fertility Database and Human Mortality Database are likely the two most widely used data sources for the comparative study of fertiltiy and mortality. Both databases offer the option to download data in bulk in a few different formats. After a bulk download, users can then set up local databases in a variety of convenient ways, such as that described by Minton (2015) for R (R Development Core Team 2012) users. Another option for R users is to read data directly into an interactive session from the respective database websites. This option is a bit tricky for the typical R user to program from scratch, but it is a convenient way to quickly start working with data and generating results on the fly. Some other common pitfalls for working with these data sources in R stem from age-class and year coding, which often require coersion to use as an integer in analyses. We present a package, `DemogBerkeley`, which handles these issues automatically for R users. The goal is to reduce the coding time and effort between acquiring HMD or HFD source data and conducting analysis.

2 Installation

The main package is hosted on `github`. Two external dependencies help R handle web authentification to the databases, and these must be installed first in order for `DemogBerkeley` to properly install. These two dependencies are `cURL` and `XML`, and their installation unfortunately depends on the operation system.

In `Linux` and similar systems, open the Terminal and run:

```
sudo apt-get install libcurl
sudo apt-get install libxml2-dev
```

Then install `DemogBerkeley` using:

```
library(devtools)
install_github("UCBdemography/DemogBerkeley", subdir = "DemogBerkeley")
```

This is the simplest case. For Windows, Mac, or other special cases, some further steps are required. Please consult the `README`[1] file on the `github` repository for more details, and contact the maintainer if necessary.

3   Usage and examples

Load the package using:

```
library(DemogBerkeley)
```

The two main functions of interest are `readHMDweb()` and `readHFDweb()`, and both functions have identical arguments. It helps to be a bit familiar with HMD and HFD file naming conventions. To get the population codes used in either database, run

```
getHFDcountries()
getHMDcountries()
```

To see the data products available for a given country, run

```
getHFDitemavail("USA")
getHMDitemavail("USA")
```

The HMD and HFD both require user registration, and users will have usernames and passwords for these databases.

4   Acknowledgements

Thanks to Joshua Goldstein and Carl Boe for supporting development of this R functionality, and to Vladimir Shkolnikov, Dmitri Jdanov, and Tomáŝ Sobotka for the invitation to present this material at the HFD side meeting to the 2015 PAA Annual Meeting. This work was supported by the National Institute On Aging of the U.S. National Institutes of Health (NIH) under Award Numbers R01-AG011552 and R01-AG040245. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH.

---

[1]The `README` is at the foot of the main repository page: `https://github.com/UCBdemography/DemogBerkeley`

References

Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available at www.humanfertility.org.

Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on July 10, 2014).

Jon Minton. Merging, exploring, and batch processing data from the human fertility database and human mortality database. Technical Report TR-2015-001, Max Planck Institute for Demographic Research (MPIDR), April 2015. URL `http://www.demogr.mpg.de/papers/technicalreports/tr-2015-001.pdf`.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.