Reading Human Fertility Database and Human Mortality Database data into `R`

Tim Riffe

Department of Demography, University of California, Berkeley

May 22, 2015

Abstract

This is about

1 Motivation

The Human Fertility Database (2015) and Human Mortality Database (2015) are two widely used data sources for the comparative and historical 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. The `HMDHFDplus` package provides easy direct access to the databases using a simple standard set of arguments. Issues such as authentication and fixing column classes are handled automatically. This report outlines the basic features and provides usage examples.

2 Installation

The `HMDHFDplus` package is hosted by and can be installed directly from `github`.[1] Two external dependencies help `R` handle `html` parsing and database authentication, and these must be installed first in order for `HMDHFDplus` to properly install. These two dependencies are `cURL` and `XML`, and their installation unfortunately depends on one's operating system at this time.

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

---

[1] The package is under the following url: `http://github.com/timriffe/TR1/tree/master/TR1/HMDHFDplus`. This report, as well as a useful `README` document, can be found under `http://github.com/timriffe/TR1`.

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

Then install `HMDHFDplus` using:

```
library(devtools)
install_github("timriffe/TR1/TR1/HMDHFDplus")
```

This is the simplest case. For Windows, Mac, or other special cases, some further steps are required. Please consult the `README`[2] file on the `github` repository for more details, and contact the maintainer if necessary.

3   Usage and examples

Load the package using:

```
library(HMDHFDplus)
```

The two main functions of interest are `readHMDweb()` and `readHFDweb()`, and both functions have the same essential arguments. These two functions only require the user to supply country codes, data item names, and database authentication parameters. The HFD also allows users to extract data from former updates, which may be useful for strict replication purposes. In order to make use of this feature, users must note the 8-digit date code associated with the specific country series update. By default, `readHFDweb()` extracts the most recent update. It helps to be familiar with HMD and HFD file naming conventions. Analagous functions are also made available for databases using similar formatting standards. At this time, these databases include the Japanese Mortality Database (2015), the Canadian Human Mortality Database (2015), and the Human Fertility Collection (2015).[3]   To get the population codes used in either database, run

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

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

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

4   Acknowledgements

---

[2]The `README` is at the foot of the main repository page: `https://github.com/timriffe/TR1`

[3]The Human Life-Table Database (2015) may also be incorporated in the future.

References

Canadian Human Mortality Database. Department of Demography, Université de Montréal, 2015. Available at http://www.bdlc.umontreal.ca/chmd/.

Human Fertility Collection. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria), 2015. Available at http://www.fertilitydata.org/cgi-bin/index.php.

Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria), 2015. Available at www.humanfertility.org.

Human Life-Table Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany) and Institut national d'études démographiques (France), 2015. Available at http://www.ipss.go.jp/p-toukei/JMD/index-en.html.

Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany), 2015. Available at www.mortality.org or www.humanmortality.de (data downloaded on July 10, 2014).

Japanese Mortality Database. National Institute of Population and Social Security Research (Japan), 2015. Available at http://www.ipss.go.jp/p-toukei/JMD/index-en.html.

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.