Assignment 3.1

1.

|  |  |
| --- | --- |
| Read.xport {foreign} |  |

**Read a SAS XPORT Format Library**

**Description**

Reads a file as a SAS XPORT format library and returns a list of data.frames.

**Usage**

read.xport(file)

**Arguments**

|  |  |
| --- | --- |
| file | character variable with the name of the file to read. The file must be in SAS XPORT format. |

**Value**

If there is a more than one dataset in the XPORT format library, a named list of data frames, otherwise a data frame. The columns of the data frames will be either numeric (corresponding to numeric in SAS) or factor (corresponding to character in SAS). All SAS numeric missing values (including special missing values represented by .\_, .A to .Z by SAS) are mapped to **R** .

Trailing blanks are removed from character columns before conversion to a factor. Some sources claim that character missing values in SAS are represented by ' ' or '': these are not treated as R missing files.

The read.xport function reads SAS xport formatted files, augmenting the functionality of the read.xport function provided in the 'foreign' package with additional features, borrowed from sasxport.get in Frank Harrell's 'Hmisc' package. Namely, variables are properly coerced into the types specified by the format field.

2. Haven is designed to faciliate the transfer of data between R and SAS, SPSS, and Stata. It makes it easy to read SAS, SPSS, and Stata file formats in to R.

install.packages("haven")

The haven package provides functions for importing from SAS, SPSS and Stata file formats, read\_sas(), read\_sav() and read\_dta(). This functionality is similar to that available in the base **R** foreign package but is often faster, can read SAS7BDAT files and formats, works with Stata 14 and 14 files. First we have to install haven package and select the path.

install.packages("haven")

library(haven)

dat = read\_sas("path to file", "path to formats catalog")

The returned object will be a data frame where SAS variable labels are attached as an attribute to each variable. When a variable is attached to a format in SAS and the formats are stored in a library, its path also needs to be supplied. Missing values in numeric variables should be seamlessly converted. Missing values in character variables are converted to the empty string.

SAS, Stata and SPSS all have the notion of a “labelled”" variable. These are similar to categorical [factor variables in R](http://www.stat.berkeley.edu/classes/s133/factors.html), but integer, numeric and character vectors can be labelled and not every value must be associated with a label. To turn a labelled variable into a standard factor **R** variable use the as\_factor() function,

3. **Description read.arff(foreign)**

Reads data from Weka Attribute-Relation File Format (ARFF) files.

**Usage**

read.arff(file)

**Arguments**

**File**

A character string with the name of the ARFF file to read from, or a [connection](http://stat.ethz.ch/R-manual/R-patched/library/base/html/connections.html) which will be opened if necessary, and if so closed at the end of the function call.

**Value**

A data frame containing the data from the ARFF file.

4. For reading large csv files, we can use either use readr::read\_csv() or data.table::fread(), as both are much faster than base::read.table().

readr::read\_csv\_chunked supports reading csv files in chunks.We can us this if we do not want to load whole data at once