Importing Tables in R (2)

Stat 133 with Gaston Sanchez

R package "`readr`"

Data from google sheets

The package “readr” (by Wickham et al) is a recent package that makes it easy to read many types of tabular data

http://blog.rstudio.org/2015/04/09/readr-0-1-0

http://cran.r-project.org/web/packages/readr/vignettes/readr.html

Package “readr”

```
# remember to install it
install.packages("readr")

# load it
library(readr)
```

“readr” functions

Are around 10x faster than base functions

Are more consistent (better desgined)

Produce data frames that are easier to use

They have more flexible column specification

“readr” functions

Read delimited files with:

- `read_csv()`
- `read_csv2()`
- `read_delim()`
- `read_tsv()`

Read fixed width files with

- `read_table()`
- `read_fwf()`

Input arguments

`file_name`: (path) name of file

`col_names`: column names

`col_types`: data types of columns

`progress`: progress bar

Input arguments

`file` gives the file to read; a url or local path.

A local path can point to a a zipped, bzipped, xzipped, or gzipped file it'll be automatically uncompressed in memory before reading.

Input arguments

`col_names`: describes the column names (equivalent to header in base R). It has three possible values:

- `TRUE` will use the the first row of data as column names.
- `FALSE` will number the columns sequentially.
- A character vector to use as column names.

Input arguments

`col_types` (equivalent to `colClasses` automatically)

- `col_logical()`: contains only logical values
- `col_integer()`: integers
- `col_double()`: doubles (reals)
- `col_euro_double()` “Euro” doubles that use commas "," as decimal separator
- `col_date()`: Y-m-d dates
- `col_datetime()`: ISO8601 date times
- `col_character()`: everything else

Column types correspondence

Type	Abbreviation
`col_logical()`	`l`
`col_integer()`	`i`
`col_numeric()`	`n`
`col_double()`	`d`
`col_euro_double()`	`e`
`col_date()`	`D`
`col_datetime()`	`T`
`col_character()`	`c`
`col_skip()`	`_`

Column Types

Using a compact string: `"dc__d"`

Each letter corresponds to a column so this specification means: read first column as *double*, second as *character*, *skip* the next two, and read the last column as *double*.

Column Types

Another way to override the default choices of column types is by passing a list of col_… types

```
read_csv("iris.csv", col_types = list(
  Sepal.Length = col_double(),
  Sepal.Width = col_double(),
  Petal.Length = col_double(),
  Petal.Width = col_double(),
  Species = col_factor(c("setosa", "versicolor",
"virginica"))
))
```

Output

- Characters are never automatically converted to factors
- Column names are left as is (i.e. there is no `check.names = TRUE`)
- Use backticks to refer to variables with unusual names: `df$`Income ($000)``
- Row names are never set
- The output has class `c("tbl_df", "tbl", "data.frame")`

File: starwarstoy.csv

```
Name,Gender,Homeworld,Born,Jedi
Anakin,male,Tatooine,41.9BBY,yes
Amidala,female,Naboo,46BBY,no
Luke,male,Tatooine,19BBY,yes
Leia,female,Alderaan,19BBY,no
Obi-Wan,male,Stewjon,57BBY,yes
Han,male,Corellia,29BBY,no
Palpatine,male,Naboo,82BBY,no
R2-D2,unknown,Naboo,33BBY,no
```

String Columns as Factors

By default, functions in "readr" do not convert character strings into factors. But you can specify what columns to be imported as factors (you must specify the factor levels)

```
sw1 <- read_csv(
  file = "starwarstoy.csv",
  col_types = list(
    gender = col_factor(c("male", "female"))
))
```

String Columns as Factors

"readr" allows you to import specific columns of a data set in an easier way than R base functions:

```
# import the first four columns
sw2 <- read_csv(
  file = "starwarstoy.csv",
  col_types = "ccnn___"
)
```

Other packages

Files from other programs

Type	Package	Function
Excel	gdata	`read.xls()`
Excel	xlsx	`read.xlsx()`
Excel	readxl	`read_excel()`
Excel	XLConnect	`readWorksheet()`
SPSS	foreign	`read.spss()`
SAS	foreign	`read.ssd()`
SAS	foreign	`read.xport()`
Matlab	R.matlab	`readMat()`
Stata	foreign	`read.dta()`
Octave	foreign	`read.octave()`
Minitab	foreign	`read.mtp()`
Systat	foreign	`read.systat()`

Data from google sheets?

Data from google sheets

R package “googlesheets” by Jennifer Bryan and Joanna Zhao

https://github.com/jennybc/googlesheets