UNIVERSITY OF CALGARY

R Wizardry, Week 2: Importing data Into R

January 18th, 2017

Importing Data into R

Types of files:

- .txt: text files
- .tsv: tab-separated (or delimited) files
- .csv : comma separated value
- .csv2: same as .csv but for countries who use comma as decimal point and semicolon as field separator
- Excel files

Importing data into R

- *read.table()*: basic, enough for *.txt* and tab-delimited files
- *read.delim()*: for tab-delimited files
- *read.csv()*: comma-separated and semicolon-separated files
- *read.csv* is based on *read.table* (the latter will import a lighter but can import data sometimes give different results due to the default parameters

To import Excel files (.xls, .xlsx, .xlsm, etc.):

- variable_name <- loadWorkbook("<name and extension of your file>")
- variable_name <- readWorksheet(wb, sheet=1)

To import from a web page:

- data <- readHTMLTable(urldata, stringsAsFactors = FALSE)

Checklist to make it easier to import data correctly into R:

If you work with spreadsheets, the first row is usually reserved for the header, while the first column is used to identify the sampling unit;

Avoid names, values or fields with blank spaces, otherwise each word will be interpreted as a separate variable, resulting in errors that are related to the number of elements per line in your data set;

If you want to concatenate words, inserting a . in between to words instead of a space;

https://www.datacamp.com/community/tutorials/r-data-import-tutorial

Short names are prefered over longer names;

Try to avoid using names that contain symbols such as ?, $,%, ^, &, *, (,),-,#, ?,,,<,>, /, |, \, [,] ,{, and };

Delete any comments that you have made in your Excel file to avoid extra columns or NA's to be added to your file; and

Make sure that any missing values in your data set are indicated with NA.

https://www.datacamp.com/community/tutorials/r-data-import-tutorial

Preparing Your R Workspace

```
rm(list=ls(all=TRUE))
```

```
getwd()
```

```
setwd("<location of your dataset>")
```

Watch out for the orientation of the slashes (/ or \) when setting up the directories!

Introduction to Report Generation with RStudio

Reproducibility is a fundamental concept in science:
“Talk is cheap. Show me the code.” Linus Torvalds.

Rmarkdown!