Topic 3

Loading data into R

Sergey Mastitsky ©
Klaipeda, 28-30 September 2011

3.1. Using Excel for data preparation

Preparing data for R

- *Microsoft Excel* is one of the most common software used to prepare datasets for R
- An *Excel* file is usually saved as a `.txt` or `.csv` file before loading it into R
- Other formats (e.g., SPSS, SAS, Access, SQL, etc.) are possible – see package `foreign`

Some rules to remember when preparing data for R

- No empty cells; missing values should be replaced with as `NA`
- Name the columns, e.g.:

	A	B	C	D
1	Treatment	Barrel	Length	Weight
2	Control	Control3	29.28	0.992
3	Control	Control3	29.83	0.772
4	Control	Control3	31.93	0.894
5	Control	Control3	26.63	0.822

- No spaces in the variable names
- No names starting with numbers or dots
- Put the file into the current R's working directory

Author: Sergey Mastitsky

Example dataset we will begin with

- Experiment on the influence of pH on the shell length / tissue dry weight in *Dreissena polymorpha*
- Four pH treatments: `Control` (Lake Ontario water), `A` (7.3), `B` (7.1), and `C` (6.9)
- Each treatment has three replicates: `Control1-Control3`, `A1-A3`, `B1-B3`, and `C1-C3`

3.2. Reading data in

Set up the working directory

- **Use the command**

```
> setwd("~/Introductory R
 Course/R_Course_Datasets")
```

- **Or in RStudio do**
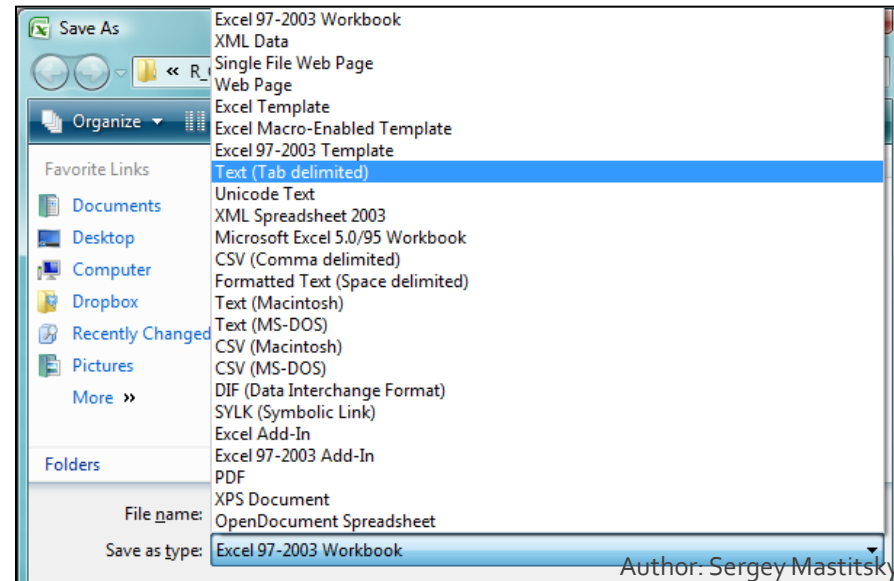
Tools -> Set Working Directory -> Choose Directory -> …your Desktop -> folder "Introductory R Course" -> folder "R_Course_Datasets"

Preparing the file in *Excel*

- Open the file `pH_down_experiment.xls`
- Quickly examine the data
- Save the file as a tab-delimited text file with the same name:
  Save as -> Save as type -> Text (Tab delimited)
- Open the file in a text editor to see how it looks like

The `read.table()` function

- **The main "working horse" to read data in R**
- **Its main arguments:**
  - `file = "some.name.txt"`: name of the file to read in (can be a URL link as well)
  - `header = TRUE`: indicates if the columns have names
  - `sep = "\t"` or `sep = ","`: separator of values in the file

Reading our example dataset in:

```
> LWdata <- read.table(
 file = "pH_down_experiment.txt",
 header = TRUE,
 sep = "\t")
> names(LWdata)
> head(LWdata)
> LWdata
```

Author: Sergey Mastitsky

Saving the workspace

```
> save.image(file =
  "pH_experiment.rda")
```

- Or in **RStudio** do
Workspace -> Save Workspace As…
```
pH_experiment.rda
```