Importing and exporting data

R

Application of R-Programming in Plant Pathology and Microbiology: Lecture 4. Dr. Aqleem Abbas

Background

- Data can be stored in a large variety of file formats.
- Each statistical package has its own format for data (xls for Microsoft Excel, dta for Stata, sas7bdat for SAS, TSV, CSV ...).
- R can read almost all file formats.

CSV file format

```
File Edit Format View Help
S.No,Spore.Length,Spore.Width,Hypha.Length,Hypha.Width,locations
1,5.1,3.5,1.4,0.2,Ichnee
2,4.9,3,1.4,0.2,Ichnee
```

Comma-separated values (CSV) file

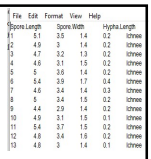
- It is a delimited text file that uses a **comma** to separate values.
- Several other CSVs such as **CSV2** use other field delimiters, for example, **semicolons**. These include tab-separated values and space-separated values.
- Import csv files: read.table(), read.csv(), read.delim()
- Import csv2 files: read.csv2() and read.delim2()
- Export csv files: write.table(), write.csv(),
- Export csv2 files: write.csv2(),

CSV2 file format

```
File Edit Format View Help
S.No;Spore.Length;Spore.Width;Hypha.Length;Hypha.Width;locations
1;5,1;3,5;1,4;0,2;Ichnee
2;4,9;3;1,4;0,2;Ichnee
```

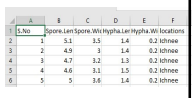

Tab-separated values (TSV)

- It is a simple text format for storing data in a **tabular structure**. e.g., database table or spreadsheet data,
- Each field value of a record is separated from the next by a tab character. txt



Excel spread sheets (xls,xlsx)

- Importing data from excel is **little bit complicated**. The easy and simple way to export Excel spreadsheets to CSV format and read the CSV in R.
- XLSX is the latest version of Microsoft Excel while XLS is the older version of Microsoft Excel.



Google Spread Sheets

- Make google spreadsheet public, publish it as a CSV file. Then you can read it in R. https://www.google.com/sheets/about/

"Statistical Analysis System" SAS (sas7bdat)

```
library("foreign")
mydata<-read.xport("SASData.xpt")
names(mydata)
```

"Statistical Package for the Social Sciences (SPSS) (sav)

read.spss() (foreign) and spss.get() (Hmisc)

```
library("foreign")
mydata<-read.spss("SPSSData.sav")
names(mydata)
```

JSON (JavaScript Object Notation)

It is a very common format on the internet.

Website import

- Machine learning repository https://archive.ics.uci.edu/ml/index.php

File Type	Delimiter/Origin of dataset	function(s) to import data (package name)	function(s) to export data (package name)
`.csv` , CSV	comma	`read.csv` (base), `read_csv` (readr)	`write.csv` (base), `write_csv` (readr)
depends (`.txt` , `.csv`)	tab	`read.table` , with sep = "\t"	`write.table` , with sep = "/t"
`.xlsx`	Excel sheet	`read.xlsx` (xlsx)	`write.xlsx` (xlsx)
`.sav`	SPSS	`spss.get` (Hmisc), `read_sav` (haven)	not advisable
`.sas7bdat`	SAS	`sasxport.get` (Hmisc), `read_sas` (haven)	not advisable
`.dta`	STATA	`read.dta` (foreign), `read_ dta` (haven)	not advisable

R-Object; Data frame

S.No	Spore.Len	Spore.Wi	Hypha.Le	Hypha.Wi	locations
1	5.1	3.5	1.4	0.2	Ichnee
2	4.9	3	1.4	0.2	Ichnee
3	4.7	3.2	1.3	0.2	Ichnee
4	4.6	3.1	1.5	0.2	Ichnee
5	5	3.6	1.4	0.2	Ichnee
6	5.4	3.9	1.7	0.4	Ichnee
7	4.6	3.4	1.4	0.3	Ichnee
8	5	3.4	1.5	0.2	Ichnee
9	4.4	2.9	1.4	0.2	Ichnee
10	4.9	3.1	1.5	0.1	Ichnee
11	5.4	3.7	1.5	0.2	Ichnee
12	4.8	3.4	1.6	0.2	Ichnee
13	4.8	3	1.4	0.1	Ichnee
14	4.3	3	1.1	0.1	Ichnee
15	5.8	4	1.2	0.2	Ichnee
16	5.7	4.4	1.5	0.4	Ichnee
17	5.4	3.9	1.3	0.4	Ichnee
18	5.1	3.5	1.4	0.3	Ichnee
19	5.7	3.8	1.7	0.3	Ichnee

Head

Maximum value

Mean

Minimum value

Range

Median

Tail

Structure (Str)

Row?
Column?
Variables names?
Datatype ?
Rows No
Column No

Descriptive statistics: Summarize, Describe and Present

Thank You