

Foundations of R Software

Lecture 46

Data Handling

⋮⋮⋮

**Importing and Reading EXCEL and other Data
Files**

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Importing data files from other software:

```
> setwd("C:/RCourse/")
```

Spreadsheet (Excel) file data

The `readxl` package has the function `read_excel()` for reading Excel files.

This will read the first sheet of an Excel spreadsheet.

To read Excel files, we first need to install the package

```
install.packages("readxl")
```

```
library(readxl)
```

Importing data files from other software:

Spreadsheet (Excel) file data

```
read_excel("datafile.xlsx")
```

```
read_excel("datafile.xls")
```

Specify sheet either by position or by name

```
read_excel(datasets, sheet_number)
```

```
read_excel(datasets, "sheet_name")
```

Importing data files:

spexcel - Excel

File Home Insert Page Layout Formulas Data Review View

Clipboard Font Alignment Number

A1 Variable 1

	A	B	C	D	E
1	Variable 1	Variable 2	Variable 3		
2	1	10	100		
3	2	20	200		
4	3	30	300		
5	4	40	400		
6	5	50	500		
7					

Importing data files:

To extract variable, write

`object_name$Variable_name`

Example:

```
> dataspexcel = read_excel("spexcel.xlsx",  
sheet=1)
```

```
> dataspexcel
```

```
# A tibble: 5 x 3
```

	<code>`Variable 1`</code>	<code>`Variable 2`</code>	<code>`Variable 3`</code>
	<code><dbl></code>	<code><dbl></code>	<code><dbl></code>
1	1	10	100
2	2	20	200
3	3	30	300
4	4	40	400
5	5	50	500

Importing data files:

Excel data

Example:

```
> dataspexcel$`Variable 1`  
[1] 1 2 3 4 5
```

```
> dataspexcel$`Variable 2`  
[1] 10 20 30 40 50
```

```
> mean(dataspexcel$`Variable 1`)  
[1] 3
```

	A	B	C	D
1	Variable 1	Variable 2	Variable 3	
2	1	10	100	
3	2	20	200	
4	3	30	300	
5	4	40	400	
6	5	50	500	
7				

Sheet1 | Sheet2 | Sheet3

Importing data files:

Excel data Example:

```
R Console
> dataspexcel = read_excel("spexcel.xlsx", sheet=1)
> dataspexcel
# A tibble: 5 x 3
  `Variable 1` `Variable 2` `Variable 3`
      <dbl>      <dbl>      <dbl>
1           1         10         100
2           2         20         200
3           3         30         300
4           4         40         400
5           5         50         500
>
> dataspexcel$`Variable 1`
[1] 1 2 3 4 5
> dataspexcel$`Variable 2`
[1] 10 20 30 40 50
>
> mean(dataspexcel$`Variable 1`)
[1] 3
>
```

Importing data files:

Excel data

Example:

```
> dataspexcel2 = read_excel("spexcel.xlsx",  
sheet=2)
```

```
> dataspexcel2
```

```
# A tibble: 5 x 3
```

	<code>`Variable 4`</code>	<code>`Variable 5`</code>	<code>`Variable 6`</code>
	<code><dbl></code>	<code><dbl></code>	<code><dbl></code>
1	6	110	110
2	7	120	210
3	8	130	310
4	9	140	410
5	10	150	510

Importing data files:

Excel data

Example:

```
> dataspexcel2$`Variable 4`  
[1] 6 7 8 9 10
```

```
> dataspexcel2$`Variable 5`  
[1] 110 120 130 140 150
```

```
> dataspexcel2$`Variable 6`  
[1] 110 210 310 410 510
```

```
> mean(dataspexcel2$`Variable 6`)  
[1] 310
```

	A	B	C	D
1	Variable 4	Variable 5	Variable 6	
2	6	110	110	
3	7	120	210	
4	8	130	310	
5	9	140	410	
6	10	150	510	
7				
		Sheet1	Sheet2	Sheet3

Importing data files:

Excel data Example:

```
R Console
> dataspexcel2 = read_excel("spexcel.xlsx", sheet=2)
> dataspexcel2
# A tibble: 5 x 3
  `Variable 4` `Variable 5` `Variable 6`
    <dbl>         <dbl>         <dbl>
1           6         110         110
2           7         120         210
3           8         130         310
4           9         140         410
5          10         150         510
>
> dataspexcel2$`Variable 4`
[1]  6  7  8  9 10
>
> dataspexcel2$`Variable 5`
[1] 110 120 130 140 150
>
> dataspexcel2$`Variable 6`
[1] 110 210 310 410 510
>
> mean(dataspexcel2$`Variable 6`)
[1] 310
```

Importing data files from other software:

Spreadsheet (Excel) file data

Limit the number of data rows read

```
read_excel(datasets, n_max = 3)
```

Read from an Excel range using A1 or R1C1 notation

```
read_excel(datasets, range = "C1:E7")
```

```
read_excel(datasets, range = "R1C2:R2C5")
```

R1C1 notation : Row-Column notation

R2C3 refers to the cell at the second row and third column

Importing data files from other software:

Spreadsheet (Excel) file data

Limit the number of data rows read

```
read_excel(datasets, n_max = 3)
```

```
dataspexcel4 = read_excel("spexcel.xlsx",  
n_max=3)
```

```
dataspexcel4
```

```
# A tibble: 3 x 3
```

	`Variable 1` <dbl>	`Variable 2` <dbl>	`Variable 3` <dbl>
1	1	10	100
2	2	20	200
3	3	30	300

Importing data files from other software:

```
dataspexcel4 = read_excel("spexcel.xlsx",  
n_max=3)
```

```
R Console  
> dataspexcel4 = read_excel("spexcel.xlsx", n_max=3)  
> dataspexcel4  
# A tibble: 3 x 3  
  `Variable 1` `Variable 2` `Variable 3`  
    <dbl>      <dbl>      <dbl>  
1         1         10         100  
2         2         20         200  
3         3         30         300
```

Importing data files from other software:

Spreadsheet (Excel) file data

Read from an Excel range using A1 or R1C1 notation

```
R Console
> dataspexcel5 = read_excel("spexcel.xlsx", range="A2:B3", sheet=1)
> dataspexcel5
# A tibble: 1 x 2
  `1`    `10`
  <dbl> <dbl>
1     2    20
>
> dataspexcel6 = read_excel("spexcel.xlsx", range="A2:B3", sheet=2)
> dataspexcel6
# A tibble: 1 x 2
  `6`    `110`
  <dbl> <dbl>
1     7   120
>
```

Importing data files from other software:

SPSS data file

For reading SPSS data files, use **foreign** package and function **read.spss()**

To read SPSS files, we first need to install the package

```
install.packages(" foreign ")  
library(foreign)  
data = read.spss("datafile.sav")
```

Importing data files:

HTML data file

For reading HTML data files, use **XML** package and function

readHTMLTable

To read HTML data files, we first need to install the package

```
install.packages("XML")
```

```
library(XML)
```

```
data = readHTMLTable("filename")
```

Importing data files:

Other data files

The **foreign** package also includes functions to load from other formats, including:

- read.octave("<Path to file>"): Octave and MATLAB
- read.systat("<Path to file>"): SYSTAT
- read.xport("<Path to file>"): SAS XPORT
- read.dta("<Path to file>"): Stata