

# readxl (1)

INTRODUCTION TO IMPORTING DATA IN R



**Filip Schouwenaars**  
Instructor, DataCamp

# Microsoft Excel

- Common data analysis tool
- Many R packages to interact with Excel
- readxl - Hadley Wickham

# Typical Structure Excel Data

- Different sheets with tabular data

The diagram illustrates a data transformation process. On the left, a wide table with 2 columns (Capital, Population) and 5 rows is shown. An Excel icon is positioned to its right. An arrow labeled "Pivot" points to a long table on the right, which has 2 columns (Capital, Population) and 10 rows. The long table is divided into two sections by a horizontal line, representing data for two different years (1990 and 2000). The top section (rows 1-5) shows data for the year 1990, and the bottom section (rows 6-10) shows data for the year 2000. The long table has a header row for each section, with the year indicated in the first column.

Capital	Population
New York	16044000
Berlin	3433695
Madrid	3010492
Stockholm	1683713

year\_1990

Capital	Population
New York	17800000
Berlin	3382169
Madrid	2938723
Stockholm	1942362

year\_2000



# readxl

- `excel_sheets()`
  - list different sheets
- `read_excel()`
  - actually import data into R

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

# excel\_sheets()

```
dir()
```

```
"cities.xlsx" "the_rest_is_secret.txt"
```

```
excel_sheets("cities.xlsx")
```

```
"year_1990" "year_2000"
```

# read\_excel()

```
read_excel("cities.xlsx")
```

```
# A tibble: 4 × 2
  Capital Population
  <chr>      <dbl>
1 New York  16044000
2 Berlin    3433695
3 Madrid    3010492
4 Stockholm 1683713
```

```
read_excel("cities.xlsx", sheet = 2)
```

```
read_excel("cities.xlsx", sheet = "year_2000")
```

```
# A tibble: 4 × 2
  Capital Population
  <chr>      <dbl>
1 New York  17800000
2 Berlin    3382169
3 Madrid    2938723
4 Stockholm 1942362
```

# Let's practice!

INTRODUCTION TO IMPORTING DATA IN R

# readxl (2)

INTRODUCTION TO IMPORTING DATA IN R



**Filip Schouwenaars**  
Instructor, DataCamp



# read\_excel()

```
read_excel(path, sheet = 1,  
           col_names = TRUE,  
           col_types = NULL,  
           skip = 0)
```

Capital	Population		
New York	16044000		
Berlin	3433695		
Madrid	3010492		
Stockholm	1683713		
year_1990			
year_2000			



# read\_excel() - col\_names

```
read_excel(path, sheet = 1,  
           col_names = TRUE,  
           col_types = NULL,  
           skip = 0)
```

- `col_names = FALSE`: R assigns names itself
- `col_names = character vector`: manually specify

# read\_excel() - col\_types

```
read_excel(path, sheet = 1,  
           col_names = TRUE,  
           col_types = NULL,  
           skip = 0)
```

```
read_excel("cities.xlsx", col_types = c("text", "text"))
```

```
# A tibble: 4 × 2  
  Capital Population  
  <chr>      <chr>  
1 New York  16044000  
2 Berlin   3433695  
3 Madrid   3010492  
4 Stockholm 1683713
```

# read\_excel() - col\_types

```
read_excel(path, sheet = 1,  
           col_names = TRUE,  
           col_types = NULL,  
           skip = 0)`
```

```
read_excel("cities.xlsx",  
           col_types = c("text", "blank"))
```

```
# A tibble: 4 × 1  
  Capital  
  <chr>  
1 New York  
2 Berlin  
3 Madrid  
4 Stockholm
```

# read\_excel() - skip

```
read_excel(path, sheet = 1,  
           col_names = TRUE,  
           col_types = NULL,  
           skip = 0)
```

```
read_excel("cities.xlsx",  
           col_names = c("Capital", "Population"),  
           skip = 2)
```

```
# A tibble: 3 × 2  
  Capital Population  
  <chr>      <dbl>  
1 Berlin    3433695  
2 Madrid    3010492  
3 Stockholm 1683713
```

- n\_max not (yet) available

# Wrap-up

- `excel_sheets()`
- `read_excel()`
- Everything you need!
- Fast
- Same arguments as in `readr` package
- Consistency

# Let's practice!

INTRODUCTION TO IMPORTING DATA IN R

# gdata

INTRODUCTION TO IMPORTING DATA IN R



**Filip Schouwenaars**  
Instructor, DataCamp



# gdata

- Gregory Warnes
- Entire suite of tools for data manipulation
- Supercharges basic R
- `read.xls()`
- Support for XLS
- Support for XLSX with additional driver
- No `readxl::excel_sheets()` equivalent

# gdata

XLS  $\xrightarrow{\text{Perl}}$  CSV

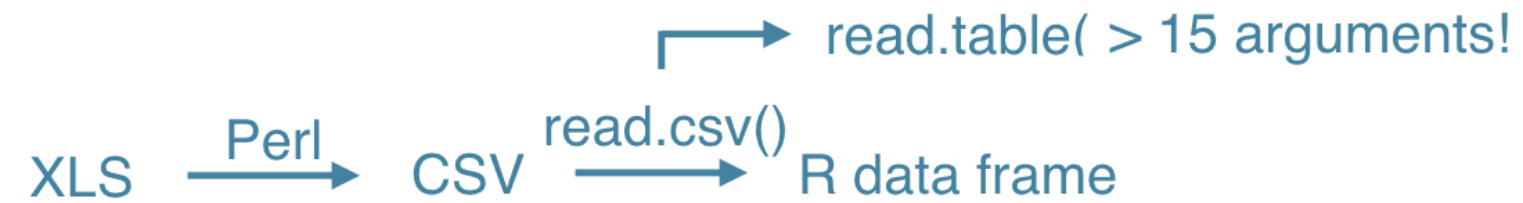
# gdata

XLS  $\xrightarrow{\text{Perl}}$  CSV  $\xrightarrow{\text{read.csv()}}$  R data frame

# gdata



# gdata



- Elegant extension of utils package
- Easy if familiar with utils
- Extremely inefficient
- readxl < v1.x


# cities.xls

Capital	Population
New York	16044000
Berlin	3433695
Madrid	3010492
Stockholm	1683713

year\_1990

	Population
	17800000
	3382169
Madrid	2938723
Stockholm	1942362

year\_2000



# read.xls()

```
install.packages("gdata")  
library(gdata)
```

```
read.xls("cities.xls")
```

```
      Capital Population  
1 New York    16044000  
2   Berlin     3433695  
3   Madrid     3010492  
4 Stockholm     1683713
```

```
read.xls("cities.xls", sheet = "year_2000")
```

```
      Capital Population  
1 New York    17800000  
2   Berlin     3382169  
3   Madrid     2938723  
4 Stockholm     1942362
```

# Let's practice!

INTRODUCTION TO IMPORTING DATA IN R