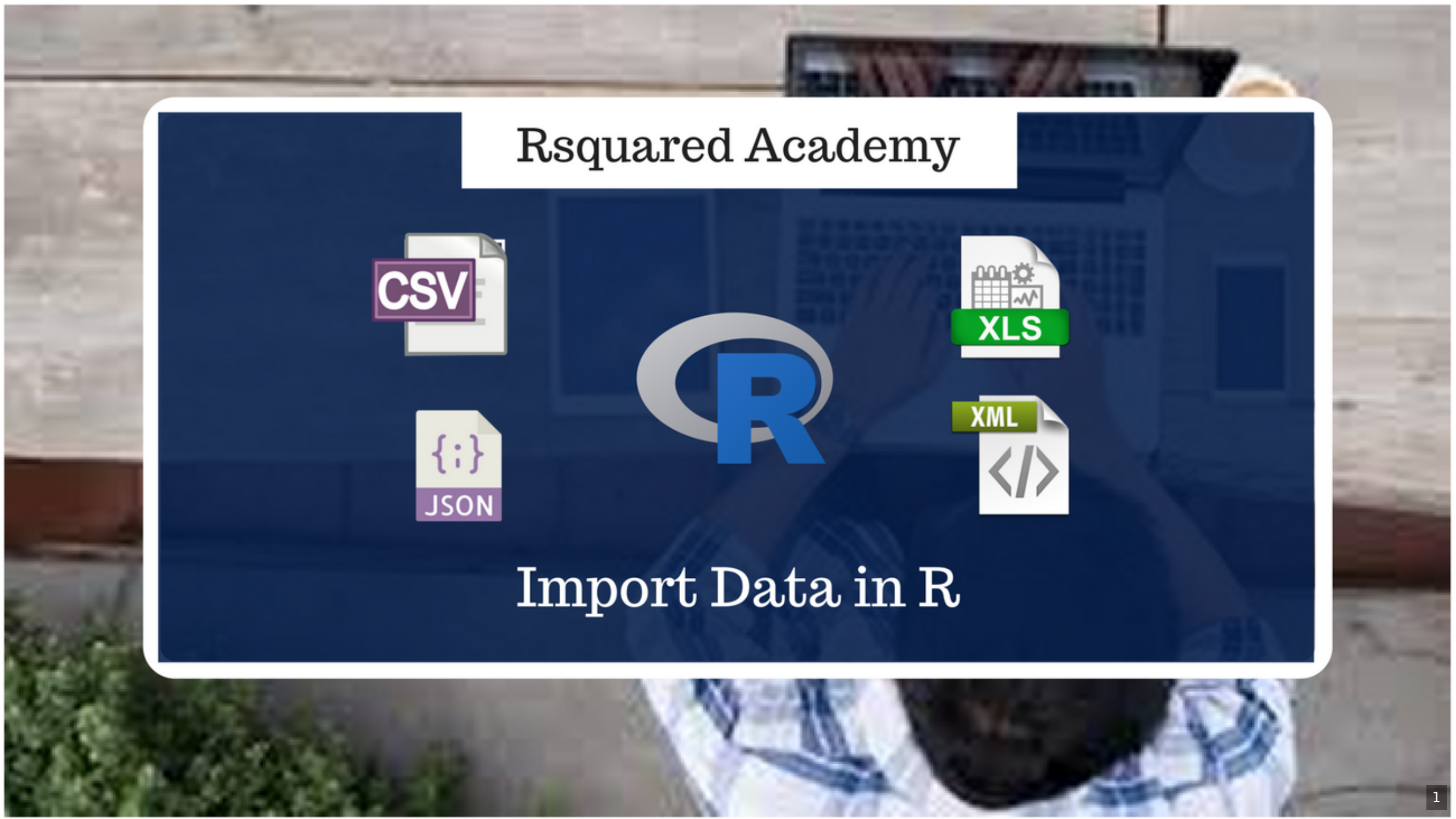

Rsquared Academy
CSV
XLS
JSON
XML
Import Data in R

Agenda

Read data from

- xls
- xlsx
- SAS
- SPSS
- STATA

Libraries

```
library(readxl)
library(haven)
```

Overview

- list sheets in an excel file
- read data from an excel sheet
- read specific cells
- read specific rows
- read specific columns

List Sheets

```
excel_sheets('sample.xls')
```

```
## [1] "ecom"
```

Read Sheet

```
read_excel('sample.xls', sheet = 1)
```

```
## # A tibble: 7 x 5
##   channel        users new_users sessions bounce_rate
##   <chr>          <dbl>     <dbl>    <dbl> <chr>
## 1 Organic Search 43296     40238    50810 48.72%
## 2 Direct         12916     12311    16419 49.27%
## 3 Referral       10983      7636    18105 22.26%
## 4 Social         10346     10029    11101 61.92%
## 5 Display         5564      4790     7220 83.30%
## 6 Paid Search     2687      2205     3438 38.02%
## 7 Affiliates      1773      1585     2167 55.75%
```

Read Specific Cells

	A	B	C	D	E
1	channel	users	new_users	sessions	bounce_rate
2	Organic Search	43296	40238	50810	48.72%
3	Direct	12916	12311	16419	49.27%
4	Referral	10983	7636	18105	22.26%
5	Social	10346	10029	11101	61.92%
6	Display	5564	4790	7220	83.30%
7	Paid Search	2687	2205	3438	38.02%
8	Affiliates	1773	1585	2167	55.75%
9					

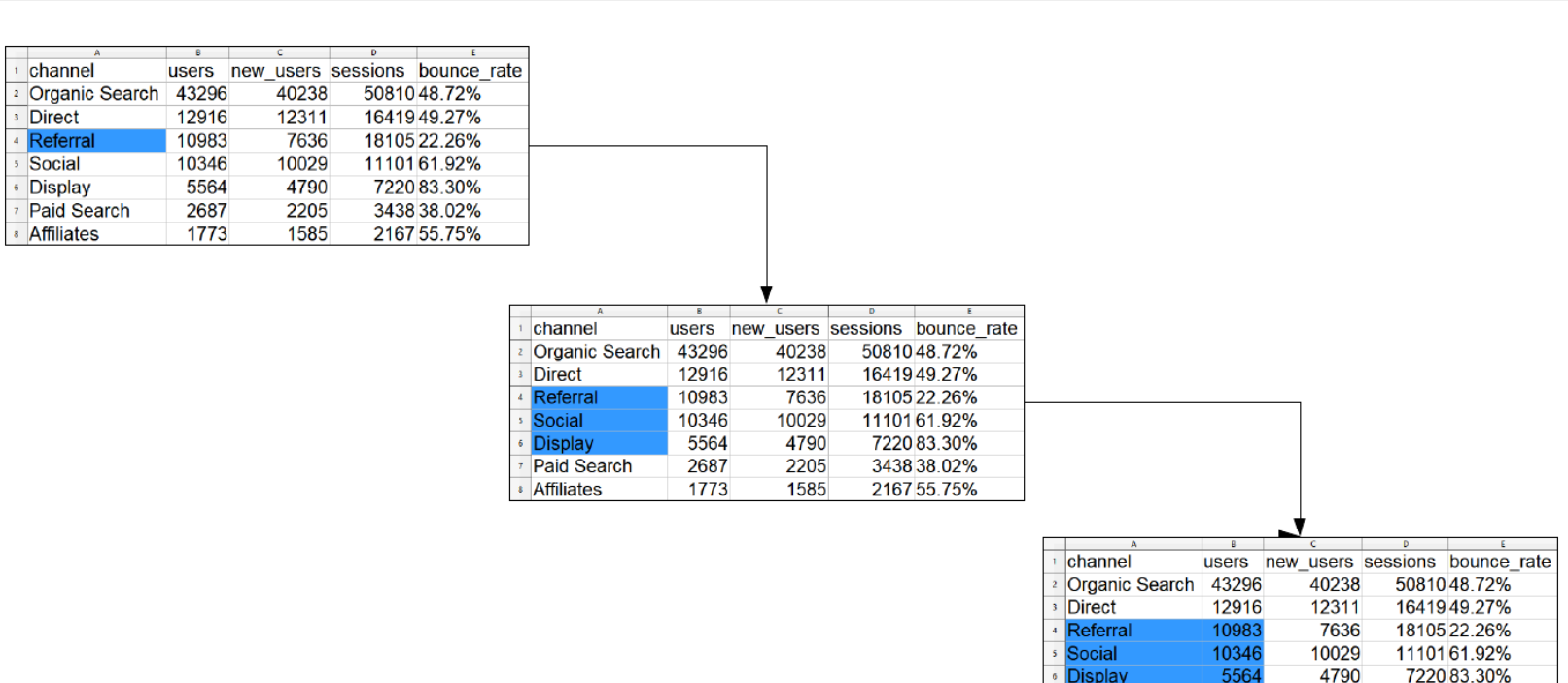
B	C
users	new_users
43296	40238
12916	12311
10983	7636

Read Specific Cells

```
read_excel('sample.xls', sheet = 1, range = "B1:C4")
```

```
## # A tibble: 3 x 2
##   users new_users
##   <dbl>     <dbl>
## 1 43296     40238
## 2 12916     12311
## 3 10983      7636
```

Read Specific Cells

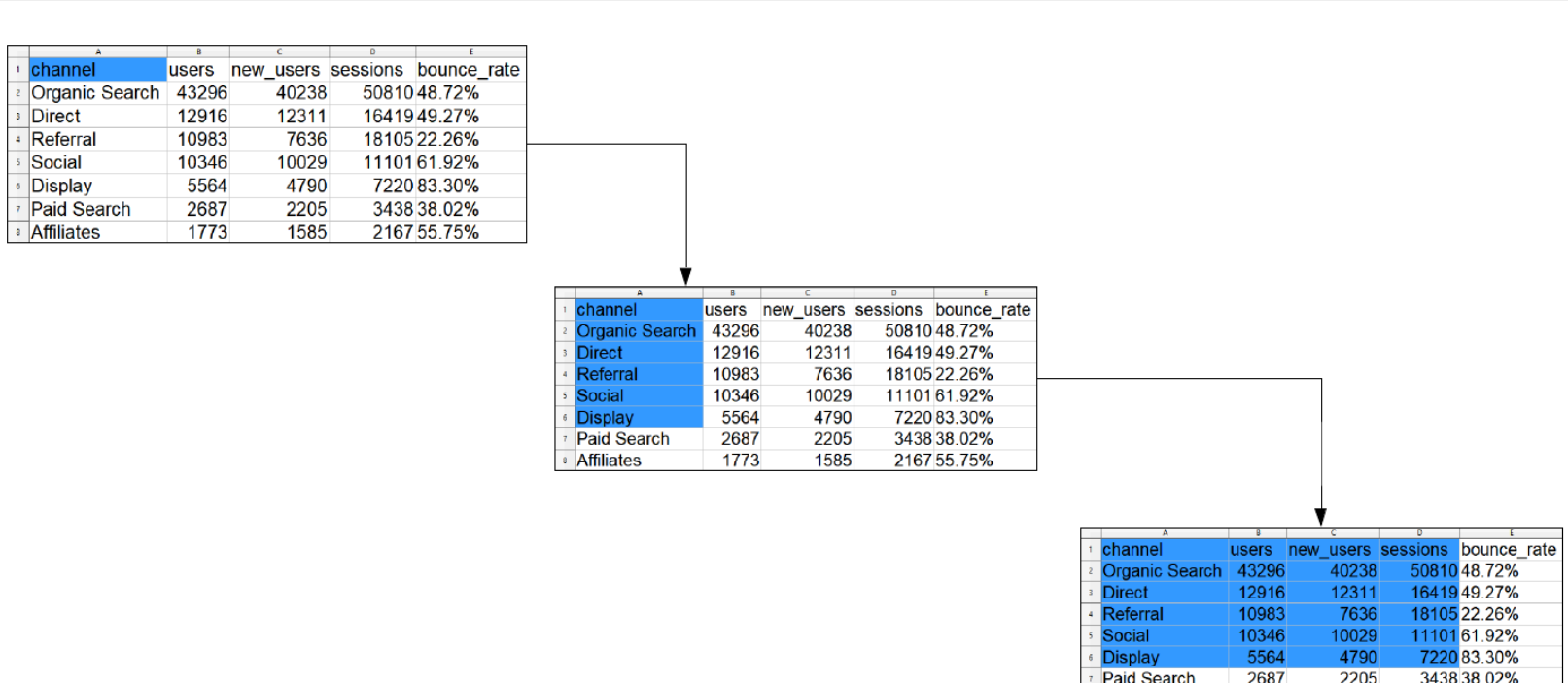


Read Specific Cells

```
read_excel('sample.xls', sheet = 1, col_names = FALSE,
  range = anchored("A4", dim = c(3, 2)))
```

```
## # A tibble: 3 x 2
##   X__1      X__2
##   <chr>    <dbl>
## 1 Referral 10983
## 2 Social   10346
## 3 Display   5564
```

Read Specific Cells



	A	B	C	D	E
1	channel	users	new_users	sessions	bounce_rate
2	Organic Search	43296	40238	50810	48.72%
3	Direct	12916	12311	16419	49.27%
4	Referral	10983	7636	18105	22.26%
5	Social	10346	10029	11101	61.92%
6	Display	5564	4790	7220	83.30%
7	Paid Search	2687	2205	3438	38.02%
8	Affiliates	1773	1585	2167	55.75%

Read Specific Cells

```
read_excel('sample.xls', sheet = 1,
  range = cell_limits(c(1, 1), c(6, 4)))
```

```
## # A tibble: 5 x 4
##   channel        users new_users sessions
##   <chr>          <dbl>     <dbl>    <dbl>
## 1 Organic Search 43296     40238    50810
## 2 Direct         12916     12311    16419
## 3 Referral       10983      7636    18105

## 4 Social         10346     10029    11101
## 5 Display         5564      4790     7220
```

Read Specific Cells

```
read_excel('sample.xls', sheet = 1,
  range = cell_limits(c(1, 2), c(NA, NA)))
```

```
## # A tibble: 7 x 4
##   users new_users sessions bounce_rate
##   <dbl>     <dbl>    <dbl> <chr>
## 1 43296     40238    50810 48.72%
## 2 12916     12311    16419 49.27%
## 3 10983      7636    18105 22.26%
## 4 10346     10029    11101 61.92%
## 5  5564      4790     7220 83.30%
## 6  2687      2205     3438 38.02%
## 7  1773      1585     2167 55.75%
```

Read Specific Cells

```
read_excel('sample.xls', sheet = 1,
  range = cell_limits(c(1, NA), c(NA, 2)))
```

```
## # A tibble: 7 x 2
##   channel        users
##   <chr>          <dbl>
## 1 Organic Search 43296
## 2 Direct         12916
## 3 Referral       10983
## 4 Social         10346
## 5 Display         5564
## 6 Paid Search     2687
## 7 Affiliates      1773
```

Read Single Column

```
read_excel('sample.xls', sheet = 1, range = cell_cols(2))
```

```
## # A tibble: 7 x 1
##   users
##   <dbl>
## 1 43296
## 2 12916
## 3 10983
## 4 10346
## 5  5564
## 6  2687
## 7  1773
```

Read Specific Rows

```
read_excel('sample.xls', sheet = 1, range = cell_rows(1:4))
```

```
## # A tibble: 3 x 5
##   channel        users new_users sessions bounce_rate
##   <chr>          <dbl>     <dbl>    <dbl> <chr>
## 1 Organic Search 43296     40238    50810 48.72%
## 2 Direct         12916     12311    16419 49.27%
## 3 Referral       10983      7636    18105 22.26%
```

Read Specific Columns

```
read_excel('sample.xls', sheet = 1, range = cell_cols(2:3))
```

```
## # A tibble: 7 x 2
##   users new_users
##   <dbl>     <dbl>
## 1 43296     40238
## 2 12916     12311
## 3 10983      7636
## 4 10346     10029
## 5  5564      4790
## 6  2687      2205
## 7  1773      1585
```

Function	Description
`anchored()`	Range of cells
`cell_limits()`	Range of cells

Statistical Softwares

- SAS
- SPSS
- STATA

```
read_stata('airline.dta')
```

```
## # A tibble: 32 x 6
##     year     y     w     r     l     k
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  1948  1.21 0.243 0.145  1.41 0.612
##  2  1949  1.35 0.260 0.218  1.38 0.559
##  3  1950  1.57 0.278 0.316  1.39 0.573
##  4  1951  1.95 0.297 0.394  1.55 0.564
##  5  1952  2.27 0.310 0.356  1.80 0.574
##  6  1953  2.73 0.322 0.359  1.93 0.711
##  7  1954  3.03 0.335 0.403  1.96 0.776
##  8  1955  3.56 0.350 0.396  2.12 0.827
##  9  1956  3.98 0.361 0.382  2.43 0.800
## 10  1957  4.42 0.379 0.305  2.71 0.921
## # ... with 22 more rows
```

```
read_spss('employee.sav')
```

```
## # A tibble: 474 x 9
##       id gender    educ   jobcat  salary salbegin jobtime prevexp mir
##    <dbl> <chr+lbl> <dbl+> <dbl+l> <dbl+> <dbl+lb> <dbl+l> <dbl+l> <db
##  1     1 m         15     3       57000  27000    98       144     0
##  2     2 m         16     1       40200  18750    98       36      0
##  3     3 f         12     1       21450  12000    98       381     0
##  4     4 f         8      1       21900  13200    98       190     0
##  5     5 m         15     1       45000  21000    98       138     0
##  6     6 m         15     1       32100  13500    98       67      0
##  7     7 m         15     1       36000  18750    98       114     0
##  8     8 f         12     1       21900  9750     98       0       0
##  9     9 f         15     1       27900  12750    98       115     0
## 10    10 f         12     1       24000  13500    98       244     0
## # ... with 464 more rows
```

```
read_sas('airline.sas7bdat')
```

```
## # A tibble: 32 x 6
##     YEAR     Y     W     R     L     K
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  1948  1.21 0.243 0.145  1.41 0.612
##  2  1949  1.35 0.260 0.218  1.38 0.559
##  3  1950  1.57 0.278 0.316  1.39 0.573
##  4  1951  1.95 0.297 0.394  1.55 0.564
##  5  1952  2.27 0.310 0.356  1.80 0.574
##  6  1953  2.73 0.322 0.359  1.93 0.711
##  7  1954  3.03 0.335 0.403  1.96 0.776
##  8  1955  3.56 0.350 0.396  2.12 0.827
##  9  1956  3.98 0.361 0.382  2.43 0.800
## 10  1957  4.42 0.379 0.305  2.71 0.921
## # ... with 22 more rows
```

Summary

File Type	readr	foreign/sas7bdat
excel	`read_excel()`	
sas	read_sas()	read.sas7bdat()
spss	read_sav() / read_spss()	read.spss()
stata	read_dta() / read_stata()	read.dta()

Thank You

For more information please visit our website
www.rsquaredacademy.com