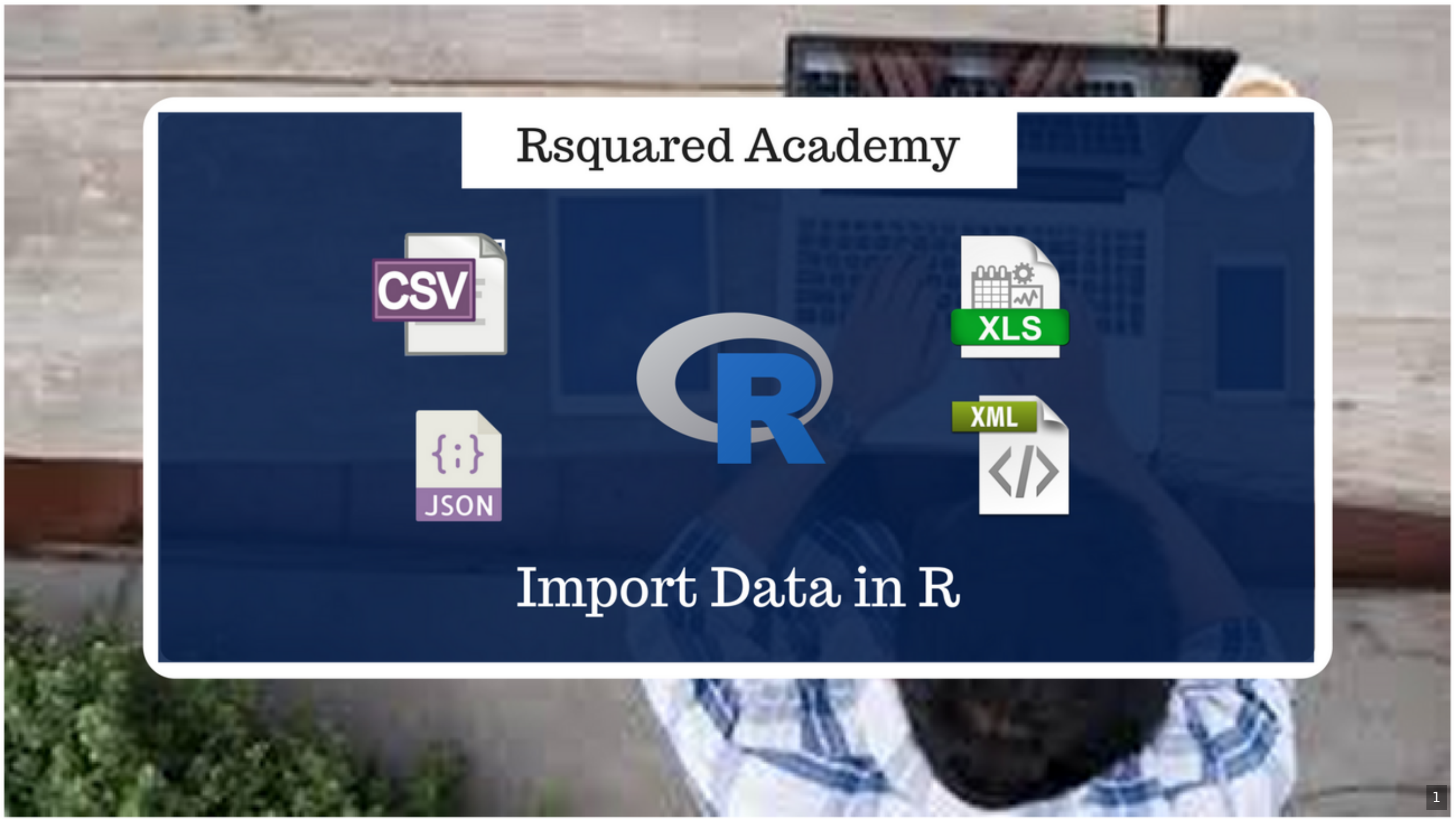

Rsquared Academy
CSV
XLS
JSON
XML
Import Data in R

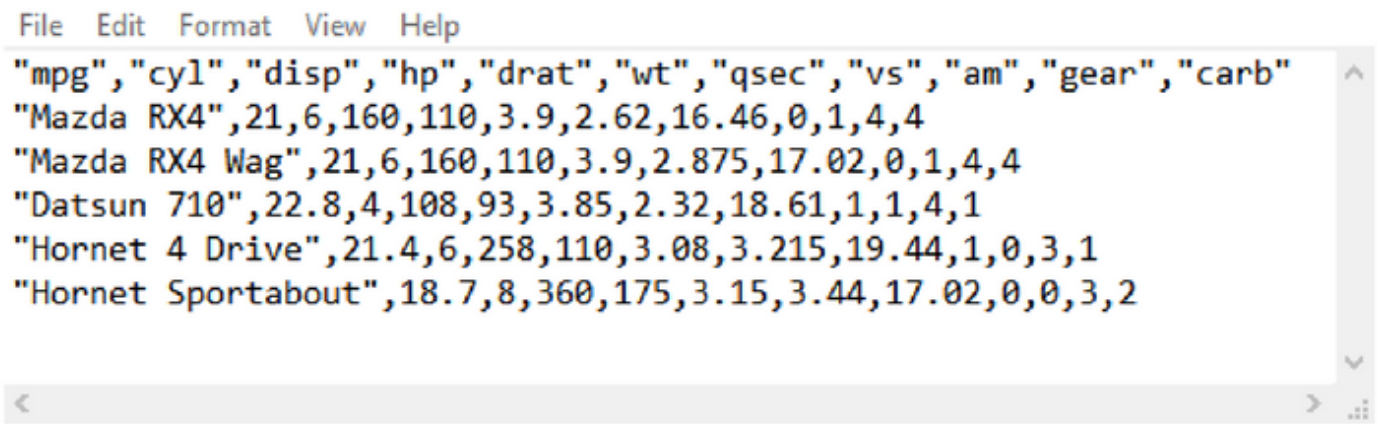
Agenda

- read data from flat or delimited files
- handle column names/header
- skip text/info
- specify column/variable types
- read specific columns/variables

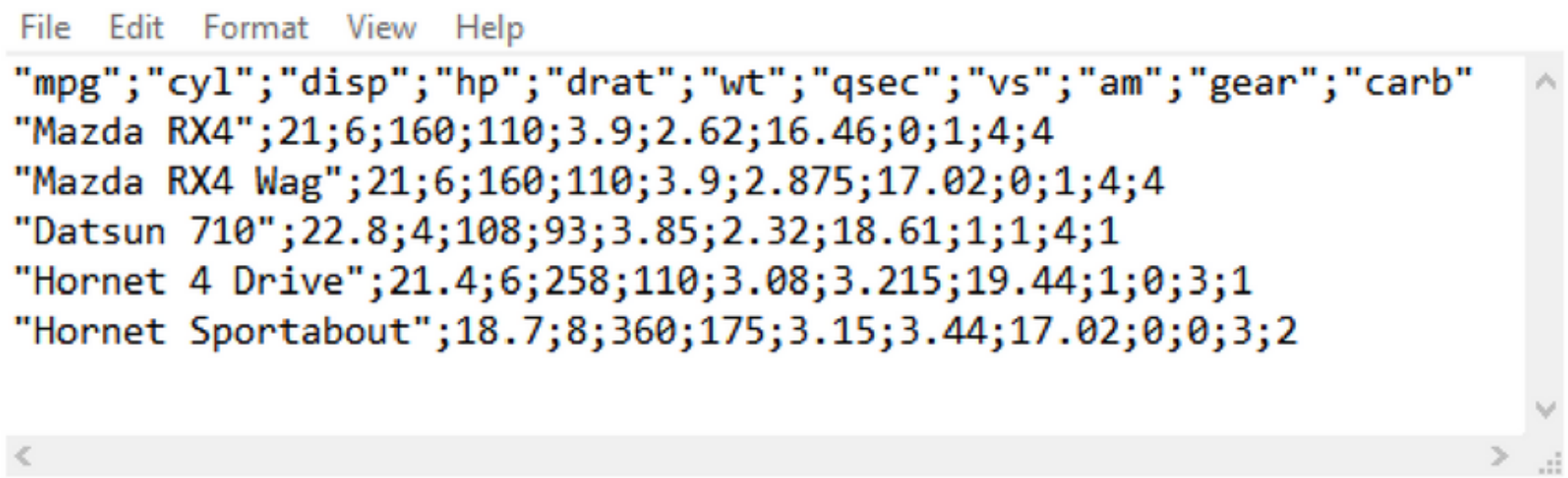
Libraries

```
library(readr)
```

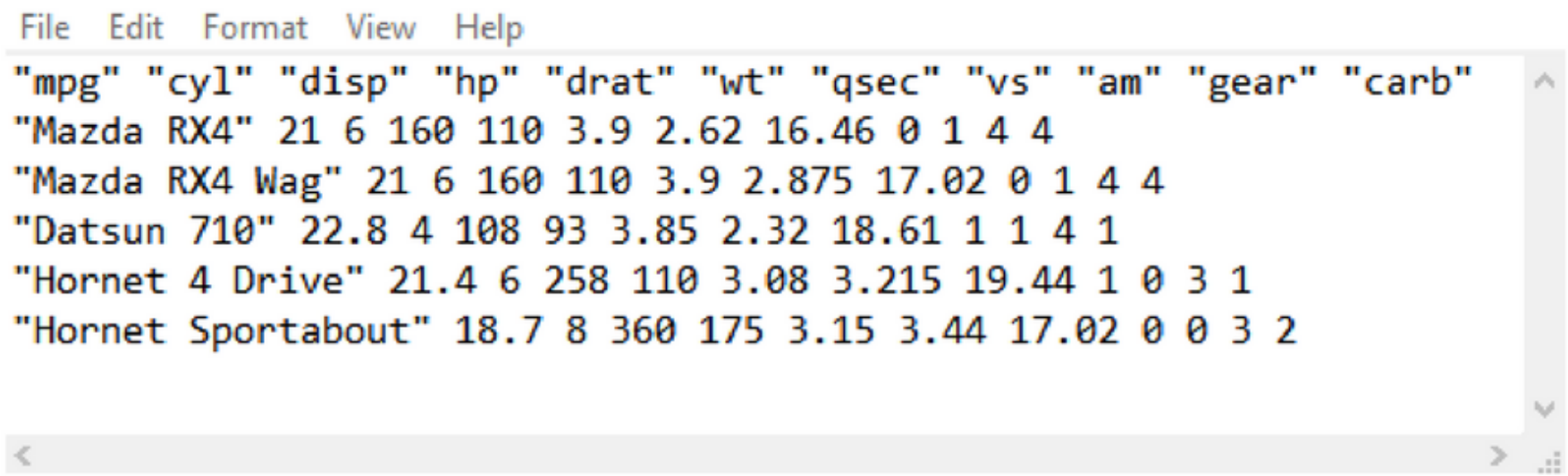
Comma Separated Values



Semi Colon Separated Values



Space Separated Values



Tab Separated Values

File Edit Format View Help

"mpg"	"cyl"	"disp"	"hp"	"drat"	"wt"	"qsec"	"vs"	"am"	"gear"	"carb"
"Mazda RX4"	21	6	160	110	3.9	2.62	16.46	0	1	4	4
"Mazda RX4 Wag"	21	6	160	110	3.9	2.875	17.02	0	1	4	4
"Datsun 710"	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
"Hornet 4 Drive"	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
"Hornet Sportabout"	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2

Read CSV File

```
read_csv('mtcars.csv')
```

```
## # A tibble: 32 x 11
##      mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  21       6  160    110  3.9   2.62  16.5     0     1     4     4
##  2  21       6  160    110  3.9   2.88  17.0     0     1     4     4
##  3  22.8     4  108     93  3.85  2.32  18.6     1     1     4     1
##  4  21.4     6  258    110  3.08  3.22  19.4     1     0     3     1
##  5  18.7     8  360    175  3.15  3.44  17.0     0     0     3     2
##  6  18.1     6  225    105  2.76  3.46  20.2     1     0     3     1
##  7  14.3     8  360    245  3.21  3.57  15.8     0     0     3     4
##  8  24.4     4  147.    62  3.69  3.19  20       1     0     4     2
##  9  22.8     4  141.    95  3.92  3.15  22.9     1     0     4     2
## 10  19.2     6  168.   123  3.92  3.44  18.3     1     0     4     4
## # ... with 22 more rows
```

Read CSV File

```
read_delim('mtcars.csv', delim = ",")
```

```
## # A tibble: 32 x 11
##      mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  21       6  160    110  3.9   2.62  16.5     0     1     4     4
##  2  21       6  160    110  3.9   2.88  17.0     0     1     4     4
##  3  22.8     4  108     93  3.85  2.32  18.6     1     1     4     1
##  4  21.4     6  258    110  3.08  3.22  19.4     1     0     3     1
##  5  18.7     8  360    175  3.15  3.44  17.0     0     0     3     2
##  6  18.1     6  225    105  2.76  3.46  20.2     1     0     3     1
##  7  14.3     8  360    245  3.21  3.57  15.8     0     0     3     4
##  8  24.4     4  147.    62  3.69  3.19  20       1     0     4     2
##  9  22.8     4  141.    95  3.92  3.15  22.9     1     0     4     2
## 10  19.2     6  168.   123  3.92  3.44  18.3     1     0     4     4
## # ... with 22 more rows
```

Column Names

File Edit Format View Help

```
"mpg","cyl","disp","hp","drat","wt","qsec","vs","am","gear","carb"
"Mazda RX4",21,6,160,110,3.9,2.62,16.46,0,1,4,4
"Mazda RX4 Wag",21,6,160,110,3.9,2.875,17.02,0,1,4,4
"Datsun 710",22.8,4,108,93,3.85,2.32,18.61,1,1,4,1
"Hornet 4 Drive",21.4,6,258,110,3.08,3.215,19.44,1,0,3,1
"Hornet Sportabout",18.7,8,360,175,3.15,3.44,17.02,0,0,3,2
```

File Edit Format View Help

```
"Mazda RX4",21,6,160,110,3.9,2.62,16.46,0,1,4,4
"Mazda RX4 Wag",21,6,160,110,3.9,2.875,17.02,0,1,4,4
"Datsun 710",22.8,4,108,93,3.85,2.32,18.61,1,1,4,1
"Hornet 4 Drive",21.4,6,258,110,3.08,3.215,19.44,1,0,3,1
"Hornet Sportabout",18.7,8,360,175,3.15,3.44,17.02,0,0,3,2
```

Column Names

```
read_csv('mtcars1.csv')
```

```
## Warning: Duplicated column names deduplicated: '4' => '4_1' [11]
```

```
## # A tibble: 31 x 11
##     `21`   `6` `160` `110` `3.9` `2.62` `16.46`   `0`   `1`   `4` `4_
##    <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>   <dbl> <dbl> <dbl> <dbl> <db
##  1  21       6  160    110  3.9    2.88    17.0     0     1     4
##  2  22.8     4  108     93  3.85   2.32    18.6     1     1     4
##  3  21.4     6  258    110  3.08   3.22    19.4     1     0     3
##  4  18.7     8  360    175  3.15   3.44    17.0     0     0     3
##  5  18.1     6  225    105  2.76   3.46    20.2     1     0     3
##  6  14.3     8  360    245  3.21   3.57    15.8     0     0     3
##  7  24.4     4  147.    62  3.69   3.19    20       1     0     4
##  8  22.8     4  141.    95  3.92   3.15    22.9     1     0     4
##  9  19.2     6  168.   123  3.92   3.44    18.3     1     0     4
## 10  17.8     6  168.   123  3.92   3.44    18.9     1     0     4
## # ... with 21 more rows
```

Column Names

```
read_csv('mtcars1.csv', col_names = FALSE)
```

```
## # A tibble: 32 x 11
##       X1    X2    X3    X4    X5    X6    X7    X8    X9   X10   X11
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  21       6  160    110  3.9   2.62  16.5     0     1     4     4
##  2  21       6  160    110  3.9   2.88  17.0     0     1     4     4
##  3  22.8     4  108     93  3.85  2.32  18.6     1     1     4     1
##  4  21.4     6  258    110  3.08  3.22  19.4     1     0     3     1
##  5  18.7     8  360    175  3.15  3.44  17.0     0     0     3     2
##  6  18.1     6  225    105  2.76  3.46  20.2     1     0     3     1
##  7  14.3     8  360    245  3.21  3.57  15.8     0     0     3     4
##  8  24.4     4  147.    62  3.69  3.19  20       1     0     4     2
##  9  22.8     4  141.    95  3.92  3.15  22.9     1     0     4     2
## 10  19.2     6  168.   123  3.92  3.44  18.3     1     0     4     4
## # ... with 22 more rows
```

Skip Lines

File Edit Format View Help

```
"The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design
,,,,,,,,,,
A data frame with 32 observations on 11 variables.,,,,,,,,,,
,,,,,,,,,,
"[, 1]", mpg, Miles/(US) gallon,,,,,,,,
"[, 2]", cyl, Number of cylinders,,,,,,,,
"[, 3]", disp, Displacement (cu.in.),,,,,,,,
"[, 4]", hp, Gross horsepower,,,,,,,,
"[, 5]", drat, Rear axle ratio,,,,,,,,
"[, 6]", wt, Weight (1000 lbs),,,,,,,,
"[, 7]", qsec, 1/4 mile time,,,,,,,,
"[, 8]", vs, V/S,,,,,,,,
"[, 9]", am," Transmission (0 = automatic, 1 = manual)",,,,,,,,
"[,10]", gear, Number of forward gears,,,,,,,,
"[,11]", carb, Number of carburetors,,,,,,,,
,,,,,,,,,,
,,,,,,,,,,
"Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.",,,,,,,,,,
,,,,,,,,,,
mpg,cyl,disp,hp,drat,wt,qsec,vs,am,gear,carb
21,6,160,110,3.9,2.62,16.46,0,1,4,4
21,6,160,110,3.9,2.875,17.02,0,1,4,4
22.8,4,108,93,3.85,2.32,18.61,1,1,4,1
21.4,6,258,110,3.08,3.215,19.44,1,0,3,1
18.7,8,360,175,3.15,3.44,17.02,0,0,3,2
18.1,6,225,105,2.76,3.46,20.22,1,0,3,1
14.3,8,360,245,3.21,3.57,15.84,0,0,3,4
```

Skip Lines

```
read_csv('mtcars2.csv')
```

```
## Warning: Missing column names filled in: 'X2' [2], 'X3' [3], 'X4' [4]
## 'X5' [5], 'X6' [6], 'X7' [7], 'X8' [8], 'X9' [9], 'X10' [10], 'X11' [
```

```
## # A tibble: 51 x 11
##    `The data was e~ X2    X3    X4    X5    X6    X7    X8    X9    X
##    <chr>            <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <
##  1 <NA>             <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
##  2 A data frame wi~ <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
##  3 <NA>             <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
##  4 [, 1]            mpg   Mile~ <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
##  5 [, 2]            cyl   Numb~ <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
##  6 [, 3]            disp  Disp~ <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
##  7 [, 4]            hp    Gros~ <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
##  8 [, 5]            drat  Rear~ <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
##  9 [, 6]            wt    Weig~ <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
## 10 [, 7]            qsec  1/4 ~ <NA>  <NA>  <NA>  <NA>  <NA>  <NA>  <
## # ... with 41 more rows, and 1 more variable: X11 <chr>
```

Skip Lines

```
read_csv('mtcars2.csv', skip = 19)
```

```
## # A tibble: 32 x 11
##      mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  21       6  160    110  3.9   2.62  16.5     0     1     4     4
##  2  21       6  160    110  3.9   2.88  17.0     0     1     4     4
##  3  22.8     4  108     93  3.85  2.32  18.6     1     1     4     1
##  4  21.4     6  258    110  3.08  3.22  19.4     1     0     3     1
##  5  18.7     8  360    175  3.15  3.44  17.0     0     0     3     2
##  6  18.1     6  225    105  2.76  3.46  20.2     1     0     3     1
##  7  14.3     8  360    245  3.21  3.57  15.8     0     0     3     4
##  8  24.4     4  147.    62  3.69  3.19  20       1     0     4     2
##  9  22.8     4  141.    95  3.92  3.15  22.9     1     0     4     2
## 10  19.2     6  168.   123  3.92  3.44  18.3     1     0     4     4
## # ... with 22 more rows
```

Maximum Lines

```
read_csv('mtcars.csv', n_max = 20)
```

```
## # A tibble: 20 x 11
##      mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  21       6 160     110  3.9   2.62  16.5     0     1     4     4
##  2  21       6 160     110  3.9   2.88  17.0     0     1     4     4
##  3  22.8     4 108      93  3.85  2.32  18.6     1     1     4     1
##  4  21.4     6 258     110  3.08  3.22  19.4     1     0     3     1
##  5  18.7     8 360     175  3.15  3.44  17.0     0     0     3     2
##  6  18.1     6 225     105  2.76  3.46  20.2     1     0     3     1
##  7  14.3     8 360     245  3.21  3.57  15.8     0     0     3     4
##  8  24.4     4 147.     62  3.69  3.19  20       1     0     4     2
##  9  22.8     4 141.     95  3.92  3.15  22.9     1     0     4     2
## 10  19.2     6 168.    123  3.92  3.44  18.3     1     0     4     4
## 11  17.8     6 168.    123  3.92  3.44  18.9     1     0     4     4
## 12  16.4     8 276.    180  3.07  4.07  17.4     0     0     3     3
## 13  17.3     8 276.    180  3.07  3.73  17.6     0     0     3     3
## 14  15.2     8 276.    180  3.07  3.78  18       0     0     3     3
## 15  10.4     8 472     205  2.93  5.25  18.0     0     0     3     4
```

Column Types

Data Type	Function
Integer	`col_integer()`
Double	col_double()
Logical	col_logical()
Categorical	col_factor()
Character	col_character()
Date/Time	col_datetime(), col_date(), col_time()
Skip	col_skip()

Column Types

```
spec_csv('mtcars5.csv')
```

```
## cols(
##   mpg = col_double(),
##   cyl = col_double(),
##   disp = col_double(),
##   hp = col_double()
## )
```

Column Types

Objective	Function
Specify column data types	`col_types()`
Skip column	col_skip()
Read spcecific columns	cols_only()

```
read_csv('mtcars5.csv',
         col_types = list(col_double(), col_factor(levels = c(4, 6, 8)),
                          col_double(), col_integer()))
```

```
## # A tibble: 32 x 4
##       mpg cyl    disp    hp
##     <dbl> <fct> <dbl> <int>
##  1  21   6       160    110

##  2  21   6       160    110
##  3  22.8 4       108     93
##  4  21.4 6       258    110
##  5  18.7 8       360    175
##  6  18.1 6       225    105
##  7  14.3 8       360    245
##  8  24.4 4       147.    62
##  9  22.8 4       141.    95
## 10  19.2 6       168.   123
## # ... with 22 more rows
```

Skip Columns

```
read_csv('mtcars5.csv',
         col_types = list(col_double(), col_factor(levels = c(4, 6, 8)),
                          col_skip(), col_integer()))
```

```
## # A tibble: 32 x 3
##      mpg cyl      hp
##    <dbl> <fct> <int>
##  1  21   6       110

##  2  21   6       110
##  3  22.8 4        93
##  4  21.4 6       110
##  5  18.7 8       175
##  6  18.1 6       105
##  7  14.3 8       245
##  8  24.4 4        62
##  9  22.8 4        95
## 10  19.2 6       123
## # ... with 22 more rows
```

Read Specific Columns

```
read_csv('mtcars5.csv',
         col_types = cols_only(mpg = col_double(),
                               cyl = col_factor(levels = c(4, 6, 8))))
```

```
## # A tibble: 32 x 2
##      mpg cyl
##    <dbl> <fct>
##  1  21   6

##  2  21   6
##  3  22.8 4
##  4  21.4 6
##  5  18.7 8
##  6  18.1 6
##  7  14.3 8
##  8  24.4 4
##  9  22.8 4
## 10  19.2 6
## # ... with 22 more rows
```

readr & Base R

Type	readr	Base R
comma	`read_csv()`	`read.csv()`
semicolon	read_csv2()	read.csv2()
tab	read_tsv()	read.delim() / read.table()
space	read_table()	read.table()
multiple spaces	read_table2()	read.table()
any delimiter	read_delim()	read.delim()

Thank You

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