

# readr: read\_csv & read\_tsv

INTRODUCTION TO IMPORTING DATA IN R



**Filip Schouwenaars**  
Instructor, DataCamp

# Overview

- Before: utils package
- Specific R packages
  - readr
  - data.table

# readr

- Hadley Wickham
- Fast, easy to use, consistent
- utils: verbose, slower

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

# CSV files

```
read.csv("states.csv", stringsAsFactors = FALSE)
```

```
      state    capital pop_mill area_sqm
1 South Dakota    Pierre    0.853   77116
2   New York    Albany   19.746   54555
3    Oregon     Salem    3.970   98381
4   Vermont Montpelier    0.627    9616
5    Hawaii  Honolulu    1.420   10931
```

```
read_csv("states.csv")
```

```
# A tibble: 5 × 4
      state    capital pop_mill area_sqm
  <chr>      <chr>    <dbl>   <int>
1 South Dakota    Pierre    0.853   77116
2   New York    Albany   19.746   54555
3    Oregon     Salem    3.970   98381
4   Vermont Montpelier    0.627    9616
5    Hawaii  Honolulu    1.420   10931
```

# TSV files

```
read.delim("states.txt", stringsAsFactors = FALSE)
```

```
      state    capital pop_mill area_sqm
1 South Dakota    Pierre    0.853   77116
2   New York    Albany   19.746   54555
3    Oregon     Salem    3.970   98381
4   Vermont Montpelier    0.627    9616
5    Hawaii  Honolulu    1.420   10931
```

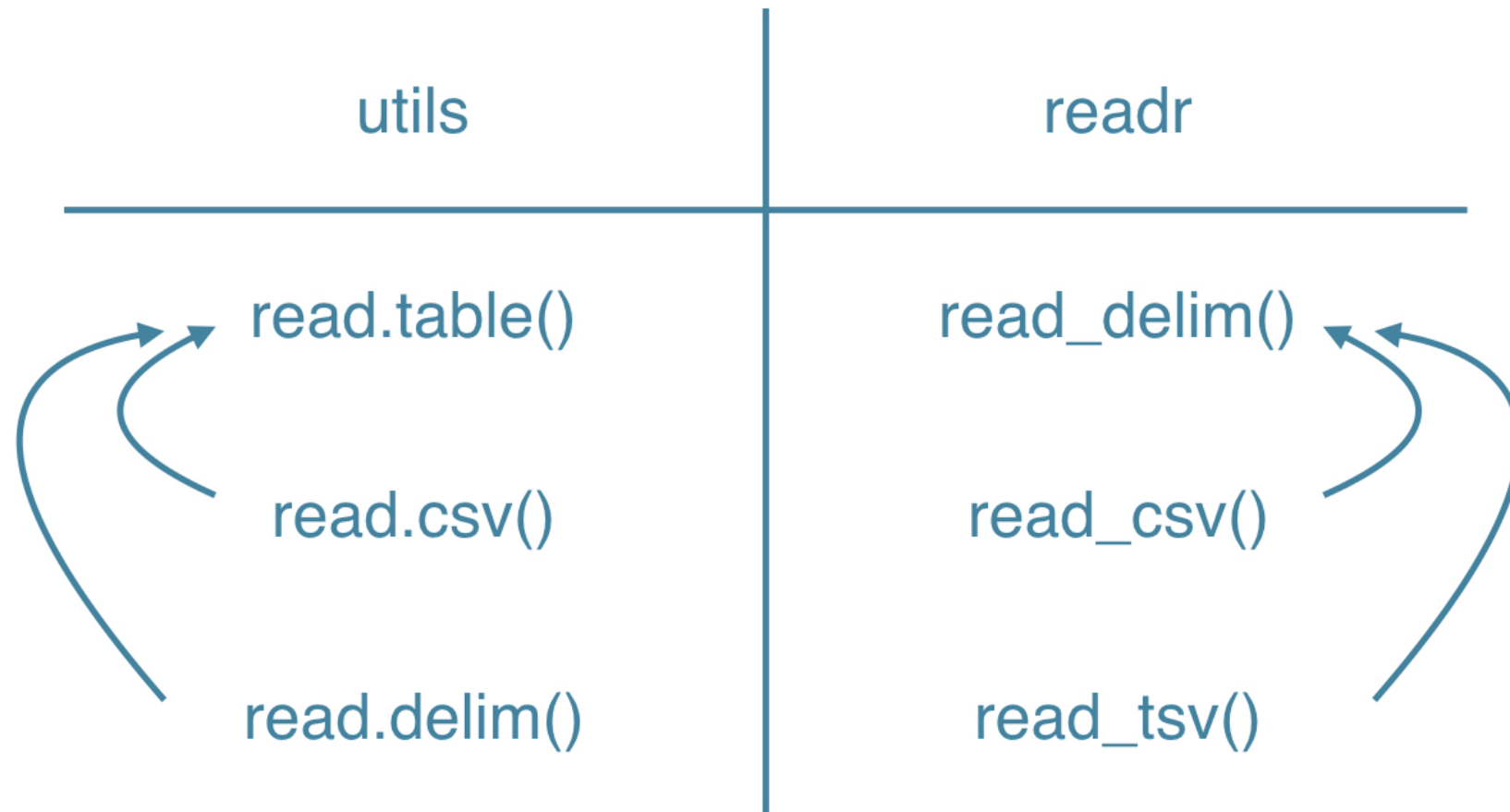
```
read_tsv("states.txt")
```

```
# A tibble: 5 × 4
      state    capital pop_mill area_sqm
  <chr>      <chr>    <dbl>   <int>
1 South Dakota    Pierre    0.853   77116
2   New York    Albany   19.746   54555
3    Oregon     Salem    3.970   98381
4   Vermont Montpelier    0.627    9616
5    Hawaii  Honolulu    1.420   10931
```

# Wrapping in utils and readr

utils	readr
read.table()	read_delim()
read.csv()	read_csv()
read.delim()	read_tsv()

# Wrapping in utils and readr



# Let's practice!

INTRODUCTION TO IMPORTING DATA IN R



# readr: read\_delim

INTRODUCTION TO IMPORTING DATA IN R



**Filip Schouwenaars**  
Instructor, DataCamp

# states2.txt

```
states2.txt
```

```
state/capital/pop_mill/area_sqm
```

```
South Dakota/Pierre/0.853/77116
```

```
New York/Albany/19.746/54555
```

```
Oregon/Salem/3.970/98381
```

```
Vermont/Montpelier/0.627/9616
```

```
Hawaii/Honolulu/1.420/10931
```

# states2.txt

```
read.table("states2.txt", header = TRUE, sep = "/",  
           stringsAsFactors = FALSE)
```

```
      state capital pop_mill area_sqm  
1 South Dakota   Pierre    0.853   77116  
2   New York    Albany   19.746   54555  
3    Oregon     Salem    3.970   98381  
4   Vermont Montpelier    0.627    9616  
5    Hawaii    Honolulu    1.420   10931
```

```
read_delim("states2.txt", delim = "/")
```

```
# A tibble: 5 x 4  
      state capital pop_mill area_sqm  
  <chr>    <chr>    <dbl>   <int>  
1 South Dakota   Pierre    0.853   77116  
2   New York    Albany   19.746   54555  
3    Oregon     Salem    3.970   98381  
4   Vermont Montpelier    0.627    9616  
5    Hawaii    Honolulu    1.420   10931
```

# col\_names

```
states3.txt
```

```
South Dakota/Pierre/0.853/77116
```

```
New York/Albany/19.746/54555
```

```
Oregon/Salem/3.970/98381
```

```
Vermont/Montpelier/0.627/9616
```

```
Hawaii/Honolulu/1.420/10931
```

# col\_names

```
read_delim("states3.txt", delim = "/", col_names = FALSE)
```

```
      X1      X2      X3      X4
      <chr>    <chr>  <dbl> <int>
1 South Dakota Pierre  0.853 77116
2   New York   Albany 19.746 54555
3    Oregon    Salem  3.970 98381
4   Vermont Montpelier 0.627  9616
5    Hawaii   Honolulu 1.420 10931
```

```
read_delim("states3.txt", delim = "/",
           col_names = c("state", "city", "pop", "area"))
```

```
      state      city      pop      area
      <chr>    <chr>  <dbl>  <int>
1 South Dakota Pierre  0.853 77116
2   New York   Albany 19.746 54555
3    Oregon    Salem  3.970 98381
4   Vermont Montpelier 0.627  9616
5    Hawaii   Honolulu 1.420 10931
```

# col\_types

```
read_delim("states2.txt", delim = "/")
```

	state	capital	pop_mill	area_sqm
	<chr>	<chr>	<dbl>	<int>
1	South Dakota	Pierre	0.853	77116
2	New York	Albany	19.746	54555
3	Oregon	Salem	3.970	98381
4	Vermont	Montpelier	0.627	9616
5	Hawaii	Honolulu	1.420	10931

```
read_delim("states2.txt", delim = "/", col_types = "ccdd")
```

	state	capital	pop_mill	area_sqm
	<chr>	<chr>	<dbl>	<dbl>
1	South Dakota	Pierre	0.853	77116
2	New York	Albany	19.746	54555
3	Oregon	Salem	3.970	98381
4	Vermont	Montpelier	0.627	9616
5	Hawaii	Honolulu	1.420	10931

# skip and n\_max

```
read_delim("states2.txt", delim = "/",  
           skip = 2, n_max = 3)
```

```
# A tibble: 3 x 4  
  New York    Albany 19.746 54555  
    <chr>      <chr> <dbl> <int>  
1   Oregon    Salem  3.970 98381  
2  Vermont Montpelier 0.627  9616  
3   Hawaii   Honolulu 1.420 10931
```

```
read_delim("states2.txt", delim = "/",  
           col_names = c("state", "city", "pop", "area"),  
           skip = 2, n_max = 3)
```

```
# A tibble: 3 x 4  
  state      city    pop  area  
  <chr>    <chr> <dbl> <int>  
1 New York    Albany 19.746 54555  
2   Oregon    Salem  3.970 98381  
3  Vermont Montpelier 0.627  9616
```

# Let's practice!

INTRODUCTION TO IMPORTING DATA IN R



# data.table: fread

INTRODUCTION TO IMPORTING DATA IN R



**Filip Schouwenaars**  
Instructor, DataCamp

# data.table

- Matt Dowle & Arun Srinivasan
- Key metric: speed
- Data manipulation in R
- Function to import data: fread()

```
install.packages("data.table")  
library(data.table)
```

- Similar to read.table()

# fread()

states.csv

```
state, capital, pop_mill, area_sqm  
South Dakota, Pierre, 0.853, 77116  
New York, Albany, 19.746, 54555  
Oregon, Salem, 3.970, 98381  
Vermont, Montpelier, 0.627, 9616  
Hawaii, Honolulu, 1.420, 10931
```

states2.csv

```
South Dakota, Pierre, 0.853, 77116  
New York, Albany, 19.746, 54555  
Oregon, Salem, 3.970, 98381  
Vermont, Montpelier, 0.627, 9616  
Hawaii, Honolulu, 1.420, 10931
```

# fread()

```
fread("states.csv")
```

```
      state capital pop_mill area_sqm
1: South Dakota   Pierre    0.853   77116
2:   New York    Albany   19.746   54555
3:   Oregon     Salem    3.970   98381
4:   Vermont Montpelier    0.627    9616
5:   Hawaii    Honolulu    1.420   10931
```

```
fread("states2.csv")
```

```
      V1      V2    V3    V4
1: South Dakota   Pierre 0.853 77116
2:   New York    Albany 19.746 54555
3:   Oregon     Salem 3.970 98381
4:   Vermont Montpelier 0.627 9616
5:   Hawaii    Honolulu 1.420 10931
```

# fread()

- Infer column types and separators
- It simply works
- Extremely fast
- Possible to specify numerous parameters
- Improved read.table()
- Fast, convenient, customizable

# Let's practice!

INTRODUCTION TO IMPORTING DATA IN R