

# Reading Data

Direct entry, Data frames

# Inputting Data

- **R** allows users to input data using a wide range methods.
  - ✓ Directly by typing the data into **R** (using `scan()`)
  - ✓ Reading external files: txt, csv, SAS, SPSS, Excel.

I encourage you to learn different methods, but we will cover a common and robust use case: handling `csv` files<sup>1</sup>.

<sup>1</sup> Find out more: <https://www.datacamp.com/community/tutorials/r-data-import-tutorial>

# Inputting Data

## Direct Method

You can directly input data points one by one using `scan()`

### Do it yourself

```
> x <- scan() ↵  
13 2 1.2 3 18 6 ↵  
↵  
> x ↵
```

This is called a *base function*.

# Inputting Data

## External Files

External files come in various formats and a number of *convenience functions* are available:

```
read.table()
```

```
read.csv()
```

```
read.delim()
```

Before we need to find out our working directory:

### **Do it yourself**

```
> getwd() ↵
```

# Inputting Data

## Setting Paths

You can use `dir()` to find what is in each directory and `setwd()` to change to a new working directory.

**Do it yourself:** We are going to change working directory to the revariant-master folder we put in the OneDrive folder earlier:

```
> dir() ↵
```

```
> setwd("C://PATH/TO/revariant-  
master") ↵
```

```
> dir() ↵
```

**Do it yourself:** Read the `simple.txt` data set and store it in a *data frame* called `easy`.

```
> easy <- read.table('simple.txt',  
header = TRUE, sep = "\t")
```

Let us look at the first 6 lines of the data:

```
> head(easy)
```

There's also an RStudio command (note uppercase letter) for looking at a data frame:

```
> View(easy)
```

Now plot the data!

```
> plot(easy)
```

# Inputting Data

## Comma Separated

**Do it yourself:** Read the `smoking.csv` data set and store it in a *data frame* called `smoking`.

```
> smoking <- read.csv('smoking.csv',  
header = TRUE) ←
```

Let's look at the data too:

```
> View(smoking) ←
```



# Data Frames

- A data frame is a list of variables, each of the same length but **not** necessarily of the **same type**.
- The top line of the table, called the *header*, contains the column names.
- Each horizontal line afterward denotes a *data row*, which begins with the name of the row, and then followed by the actual data.



# Built-in Data Frames

- We can also call built-in data frames in **R** for our tutorials.
- This can be done by using the `data ( )` command.
- For example, here is a built-in data frame in **R**, called `mtcars`.

**Do it yourself:** Call the R built-in data set `mtcars` as follows:

```
> data(mtcars) ↵
```

Let us look at the first 6 lines of the data

```
> head(mtcars) ↵
```

Find out more about it:

```
> ?mtcars ↵
```

# The Environment

**Do it yourself:** See all the objects and data in your environment:

```
> ls () ↵
```

Or you can see it in the top right corner of RStudio (Environment tab).