


# ASMAS SS2024 - Exercise 2

AUTHOR  
Peter von Rohr

PUBLISHED  
February 26, 2024

WEBR STATUS  
 Ready!

## Problem 1: Reading Data

The first step of a data analysis in R is to read the data. This can be done in different ways which are described below.

### Direct Assignment

As done in Exercise 1, we have assigned the data directly to different R-objects. To recap, this was done with

```
vec_width <- c(82,65,76,80,78,70,72,70,65,73)
vec_height <- c(185,168,168,193,180,181,182,169,165,170)
```

### Reading Files

- From local storage: read data from local file
- From website: specify link directly

### Different Formats

- Excel: An excel file has to be downloaded first and can then be imported.

```
# download first
s_wh_data <- "https://charlotte-ngs.github.io/asmasss2024/data/asm_width_height.xlsx"
s_down_dir <- tempdir()
s_dest_file <- file.path(s_down_dir, basename(s_wh_data))
download.file(url = s_wh_data, destfile = s_dest_file)
# read from local file
tbl_wh <- readxl::read_excel(s_dest_file)
# delete downloaded file
unlink(s_dest_file)
# show table read from xls
tbl_wh
```

```
# A tibble: 10 × 2
```

	Width	Height
	<dbl>	<dbl>
1	82	185
2	65	168
3	76	168
4	80	193
5	78	180
6	70	181
7	72	182
8	70	169
9	65	165
10	73	170

- CSV

```
s_wh_data <- "https://charlotte-ngs.github.io/asmasss2024/data/asm_width_height.csv"
tbl_wh <- readr::read_delim(s_wh_data, delim = ",")
```

Rows: 10 Columns: 2

— Column specification —————

Delimiter: ","

dbl (2): Width, Height

i Use `spec()` to retrieve the full column specification for this data.  
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

```
tbl_wh
```

# A tibble: 10 × 2

```
  Width Height
  <dbl> <dbl>
1     82    185
2     65    168
3     76    168
4     80    193
5     78    180
6     70    181
7     72    182
8     70    169
9     65    165
10    73    170
```

The downloaded data can be summarized using the function `summary()`.

```
summary(tbl_wh)
```

	Width	Height
Min.	:65.0	Min. :165.0
1st Qu.:	70.0	1st Qu.:168.2
Median :	72.5	Median :175.0
Mean :	73.1	Mean :176.1
3rd Qu.:	77.5	3rd Qu.:181.8
Max. :	82.0	Max. :193.0

## Problem 2: Download Beef-Cattle Data

There is a dataset on `Breast Circumference` and `Body Weight` for beef cattle animals available in two different formats.

1. Excel: [https://charlotte-ngs.github.io/asmasss2024/data/asm\\_bw\\_bc\\_reg.xlsx](https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_bc_reg.xlsx)
2. CSV: [https://charlotte-ngs.github.io/asmasss2024/data/asm\\_bw\\_bc\\_reg.csv](https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_bc_reg.csv)

### Tasks

- Read the data from both formats
- Provide summary statistics of the variables `Breast Circumference` and `Body Weight`
- Plot `Breast Circumference` on the x-axis and `Body Weight` on the y-axis

### Solutions

- Read the data Start by reading from Excel workbook

▶ Run Code



```

1 # read data from xlsx workbook
2 s_bw_bc_url_xlsx <- "https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_bc_reg.xlsx"
3 s_down_dir <- tempdir()
4 s_bw_bc_path <- file.path(s_down_dir, basename(s_bw_bc_url_xlsx))
5 download.file(s_bw_bc_url_xlsx, s_bw_bc_path)
6 tbl_bw_bc_xlsx <- readxl::read_excel(s_bw_bc_path)
7 unlink(s_bw_bc_path)
8 tbl_bw_bc_xlsx

```

trying URL 'https://charlotte-ngs.github.io/asmasss2024/data/asm\_bw\_bc\_reg.xlsx'

# A tibble: 10 × 3

	Animal	`Breast Circumference` <dbl>	`Body Weight` <dbl>
1	1	176	471
2	2	177	463
3	3	178	481
4	4	179	470
5	5	179	496
6	6	180	491
7	7	181	518
8	8	182	511
9	9	183	510
10	10	184	541

Read data from CSV-file

▶ Run Code



```

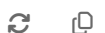
1 # read data from CSV-file
2 s_bw_bc_url_csv <- "https://charlotte-ngs.github.io/asmasss2024/data/asm_bw_bc_reg.csv"
3 df_bw_bc_csv <- read.table(s_bw_bc_url_csv, header = TRUE, sep = ",", as.is = TRUE)
4 df_bw_bc_csv

```

	Animal	Breast.Circumference	Body.Weight
1	1	176	471
2	2	177	463
3	3	178	481
4	4	179	470
5	5	179	496
6	6	180	491
7	7	181	518
8	8	182	511
9	9	183	510
10	10	184	541

- Summary statistics for Breast Circumference and Body Weight

▶ Run Code



```

1 # summary statistics
2 summary(tbl_bw_bc_xlsx)

```

	Animal	Breast Circumference	Body Weight
Min.	: 1.00	Min. :176.0	Min. :463.0
1st Qu.:	3.25	1st Qu.:178.2	1st Qu.:473.5
Median :	5.50	Median :179.5	Median :493.5
Mean :	5.50	Mean :179.9	Mean :495.2
3rd Qu.:	7.75	3rd Qu.:181.8	3rd Qu.:510.8
Max. :	10.00	Max. :184.0	Max. :541.0

- Plot

[▶ Run Code](#)

```
1 # plot
2 plot(tbl_bw_bc_xlsx$`Breast Circumference`, tbl_bw_bc_xlsx$`Body Weight`)
```

