Exercises for the course
# "An introduction to R"
Sheet 09

Exercise 45: The file "olympics.inch.txt" contains the jump height and the throwing range of the olympia winners between 1896 and 1992 in the disciplines 'high jump', 'long jump' and 'discus'. All measurements are given in the unit 'inch'. Read the data, convert it into the unit 'meter' (1 inch = 0.0254 meter) and store the converted data set into the file "olympics.meter.txt".

Exercise 46: The data set `ChickWeight` contains the weights of chicks in the first 21 days after birth. The chicks are on four different diets. The file `ChickWeightIncr.txt` contains the weight increase of the data set `ChickWeight` between birth and day 21. Load the data from `ChickWeightIncr.txt` into the variable `chickw`. Produce a picture which resembles the following multi-figure. Use seed 1111 for that.
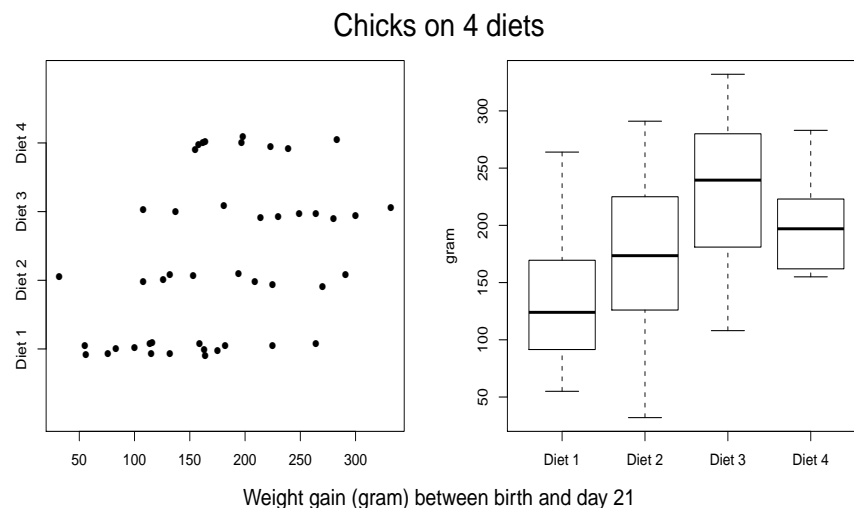
Hints: The figure on the left-hand side is produced with `method="jitter"` and `ylim=c(0,5)`. As method="jitter" adds a bit of noise you should set the seed to 1111 in order to produce the same figure. The point character is 16. The main title is magnified with factor 2, the subtitle is magnified with 1.4. The main title might be well-placed at line -3 and the subtitle might be well-placed at line -2. After producing this picture, store it into the pdf-file 'chicksOn4diets.pdf'.

Exercise 47: Download the file `riscfactorsTable.txt` and load the data. Produce a picture which resembles the following figure.
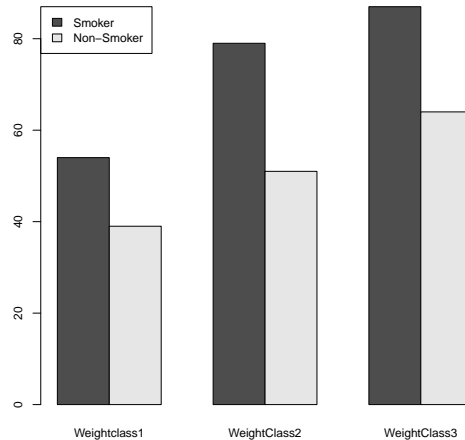
Hints: Use `args.legend=list()` (requires `legend.text=`) to place the legend at position `x="topleft"`.

Exercise 48: *Search for genes*
Download the file `sequences.txt` from the web page. Read the lines of the file into a vector `sequences` of strings. We want to find specific genes and store them into a file.

- In each line of `sequences.txt`, find the first occurrence of a substring which starts with (the start codon) AUG and end with (a stop codon) UAG.

- Repeat the last step, but the match should contain UAG at most once (minimal match). Store the positions into the variable `pos`. Hint: *?.

- From `pos`, obtain a vector `firstpos` of the first positions of the matches and a vector `lastpos` of the last positions of the matches.

- Extract the matches from `sequences`, e.g., using `substr`.

- Store the matches as lines into the file `genes.txt`.

Exercise 49: Open a file with name "confinterval.R". Store the following function into this file. Write a function `conf.interval` which takes a numeric vector as argument and which returns the mean and the confidence interval

$$\Big[\texttt{mean(x)} - \frac{\texttt{sd(x)}}{\texttt{sqrt(length(x))}} * \texttt{qt}(0.975, \texttt{length(x)} - 1),$$
$$\texttt{mean(x)} + \frac{\texttt{sd(x)}}{\texttt{sqrt(length(x))}} * \texttt{qt}(0.975, \texttt{length(x)} - 1)\Big]$$

for the mean. The confidence interval is to be represented through its lower and upper interval boundaries. The function should check whether its argument is an numeric vector. If the argument is not numeric, then the function should print the warning "Argument is not numeric: returning NA" and return 'NA'. Moreover the additional argument 'na.rm' shall indicate whether or not missing values are to be removed from the argument. By default missing arguments shall be removed. Having completed the definition of the function, save the file and source it with

```
> source("confinterval.R")
```

Then execute the function `conv.interval` on the command line to test it. Here are examples for how the function should work.

```
> conf.interval(c(1,3))
$mean
[1] 2
$lower.conf
[1] 1
$upper.conf
[1] 3

> conf.interval(c(1,NA,3))
$mean
[1] 2
$lower.conf
[1] 1
$upper.conf
[1] 3

> conf.interval(c(1,"a",3))
[1] NA
Warning message:
In conf.interval(c(1, "a", 3)) : Argument is not numeric: returning NA
```