Exercises for the course
"An introduction to R"
Sheet 08

Exercise 40: The file "olympics.meter.txt" contains the jump height and the leap distance of the olympia winners between 1896 and 1992 in the disciplines 'high jump' and 'long jump'. All measurements are given in the unit 'meter'. Plot the data points for the discipline 'high jump' against the years (true years, not years since 1900) and add the regression line. Your result should resemble the following figure.
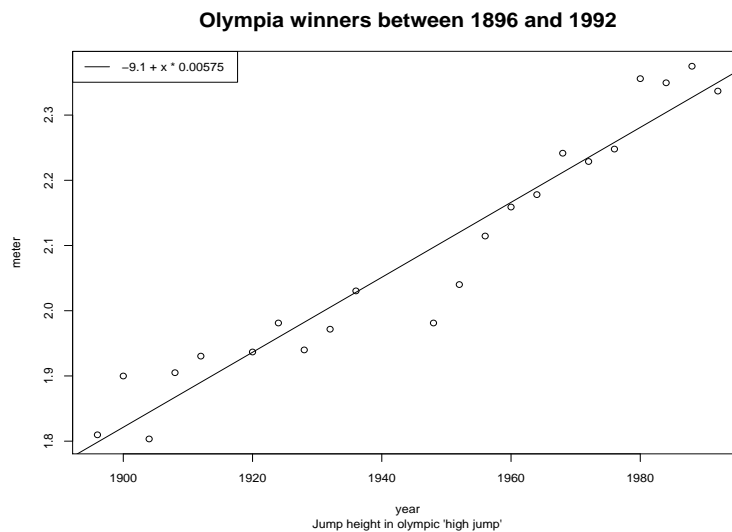
Hints: The main title is magnified with factor 1.8. The values in the legend are rounded to 3 significant digits. It is useful to first define a vector of the true years.

Exercise 41: *Regular expressions*
Define `cols <- colors()` to be the vector of color names. Using regular expressions find the following elements of `cols`:

- All color names which contain "yellow".

- All colors whose name starts with "yellow".

- All colors whose name starts with "yellow" or with "orange".

- All color names whose third character is 'c' (This is not interesting for color names. However when looking at DNA data one might want to find sequences with a certain polymorphism at a given locus).

Exercise 42: The sample mean is sensitive to outliers, e.g. `mean(1:4)` is 2.5 whereas `mean(c(1:4,1000))` is 202. To reduce the impact of outliers write a function `trim.mean()` which removes the smallest and the largest element of its argument and returns the mean of the remaining vector. Assume that the argument is numeric and contains no `NA`s. Here is how it should work:

```
> trim.mean(c(-100,1:4,1000))
[1] 2.5
> mean(1:4)
[1] 2.5
```

Exercise 43: Execute `data()` and observe the output. Now assign the object returned by `data()` as `dat`. What is the class of `dat`? Use `unclass()` to have a look at the contents of `dat`. What are the names of the elements of `dat`? One element is a matrix with name `results`. Denote the column 'Item' of `results` as `s`. Find all elements of `s` which

- contain "men"

- contain "air"

- start with "euro"

- have an 'a', 'c', 'g' or 't' at the fourth position

- contain both "sect" and "pray".

Exercise 44: Define a class 'species'. Each object of that class shall have the following three elements: a species *name*, a sample *dna* sequence and a variable which indicates whether the species is *endangered* or not. For simplicity let's use the S3 approach. Objects are created as lists and by setting the `class` attribute. Create a first object and let for simplicity "atcg" be the value of *dna*. Then define the print method for the class. Here is an example how the output of the print command could look like:

```
Species:     Elephant
DNA:         atcg
Endangered:  No
```

Hint: You need tabulators (\t) and newlines (\n) to produce such an output.