Data Analysis - 2020

Exercise sheet no 1: 15. September 2020
Data visualization

Useful commands:

- `help(<command>)` or `<command>?` — Help for command `<command>`
- `np.loadtxt('<filename>')` — load data from the file `<filename>` into an array
- `plt.hist(...)`, `plt.plot(...)`
- `plt.xlabel(...)`, `plt.ylabel(...)`

Exercise 1: Ironman Zürich (10 Points)

The file

`ironman.txt`

contains the results of the male finishers of the "Ironman Zürich" 2010 as a table. The table contains the following columns for each participant:

- column 1: total rank
- column 2: year of birth
- column 3: total time
- column 4: swimming time
- column 5: swimming rank
- column 6: cycling time
- column 7: cycling rank
- column 8: running time
- column 9: running rank

All times are given in minutes.
(Source: `http://services.datasport.com/2010/tri/ironman/RANG091.PDF`).

Write a `python` program, which reads the data from the file and creates

(a) a "scatter-plot" for

1) the total rank versus the total time,
2) the age of the participant versus the total time,
3) the running time versus the swimming time,
4) the swimming time versus the total time,
5) the cycling time versus the total time,
6) the running time versus the total time.

(2 Points)

(b) Do you see correlations between the different variables you looked at in a)? Did you expect any? (3 Points)

(c) a histogram for

1) the total time
2) the age of the participants (at the time of the race in 2010).

Define for both histograms the range of values and bin widths. Label the axes. (2 Points)

(d) Add error bars to your histograms according to the statistical uncertainty in each bin (3 points).

Deadline for submission: Friday, 25 September 2020 14:00
Form: Submission of solutions as python script to da@physik.uzh.ch. Answers for exercise 1b) as comment in the script.