Data Analysis - 2024

Exercise sheet no 1: 17. September 2024
Data visualization

Useful commands:

- `np.loadtxt('<filename>')` — **load data from the file `<filename>` into an array**
- `plt.plot(...)`
- `plt.hist(...)`
- `plt.xlabel(...)`, `plt.ylabel(...)`
- `plt.savefig(...)`

Exercise 1: Ironman Zürich (5 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` script, 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.

Label the axes. Save the plots. (3 Points)

(b) 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 (min,max) and the number of bins. Label the axes. Save the plots. (2 Points)

Deadline for submission: Friday, 20 September 2024 14:00
Form: Submission of solutions as a single python script to OLAT. Make sure to adhere to the "exercise rules".