

Exercise 7

This is voluntary exercise, but it's obligatory for those who have not reported any of exercises 1-6. If you are missing more than one exercise, it's also obligatory to attend the lecture by visiting professor Jiří Horák on Wednesday 4.5. and write a short report about what you learned there. Attending the lecture by prof. Horák is strongly recommended also for other students.

Submit your solutions to Moodle by Friday the 8th of May. Solve the both problems, but you can ignore the last part of Problem 2 if you find it too challenging.

1. Read again the presentation *Data Analysis and Tools*. Explore internet and possibly literature to find more information about the topics of the presentation. Write 1-2 pages about applications of data analysis and tools for it. Try to choose applications of which you are interested, and avoid repeating the contents of file *Data Analysis and Tools*. Consider also possibilities to use the methods and tools in your own work. **Use your own words, don't copy and paste from internet or other sources.**
2. This exercise provides you an introduction to very basics of Python programming language and Anaconda environment for Python development. Python is a modern very powerful high-level language, and among other applications it's widely used in data analysis. There are also several additional Python modules for data analysis and scientific computing, like SciPy (<https://www.scipy.org/>) and Pandas (<http://pandas.pydata.org/>). Anaconda (<https://www.continuum.io/>) is a versatile development environment including many Python modules and essential modules for R, a very popular open-source statistical software and programming language.
 - Download and install Anaconda on your own computer from <https://www.continuum.io/>. Alternatively, if you are using Metropolia's computers (when you don't have privileges to install it) you can use Anaconda which is already installed on Media virtual machine (see below).
 - Now, if you installed Anaconda on your own computer, start Spyder development environment (in Windows: go to Start Menu and write *spyder*).
 - If you are using classroom desktop, go to start menu and write *vmware*, and you'll find VMware Horizon Client. Open it and log in to desktop.metropolia.fi. Choose virtual machine *Media*. Start Spyder development environment (go to Start Menu and write *spyder*).
 - Save Python file *data_example.py* and data file *Qdata.txt* in your own directory (on your own computer, or on Z-disk, which is also visible from Media virtual machine). Open file *data_example.py* in Spyder.
 - Run the file (green button which looks like 'play').

- Try to understand how the programming code works. Write something about that in your report. If you don't understand something, report also that.
- (This can be skipped if you find it too challenging. If you have some earlier experience of programming with any language, it should be quite easy.) Modify the code such as the average is printed for each year in form:

 "The average discharge in year 2000 was..."

 "The average discharge in year 2001 was..."

 "The average discharge in year 2002 was..."

 etc.

You should use the date in the first column somehow to check out it's the end of year, when it's time to print out the average.

- If you want, you can learn more about capabilities of Python and Anaconda and make some modifications, additions and improvements to the program. It's up to you how far do you want to go.