Project 1: Data Analysis Project

Christian Carstensen

Vision: Programming is more than writing code. The ultimate goal of the projects in this course is that you learn to formulate a programming problem of your own choice, and find your own way to solve it, and present the results. The bullets below are minimum requirements, but otherwise it is very much up to you, what you will like to do with your project. Remember to write your code in a manner that is apt for others to run and read when providing feedback. I hope to see some creative ideas!

- **Objectives:** In your data analysis project, you should show that you can:
  1. Apply data cleaning and data structuring methods (L07 and L08).
  2. Apply data analysis methods (L07 and L08).
  3. Structure a code project (L05).
  4. Document code (L05).
  5. Present results in text form and in figures (L03, L06 and L07).
- **Content:** Find a subject you are interested in or perhaps have knowledge about already. For example the subject of your BA thesis (if you are that far), something related to your student job, or perhaps what you would like write about in your next thesis or seminar paper. The important thing here is the quality of the code and the presentation of results, not whether you acquire new economic knowledge. In your data analysis project, you should at a minimum:
  1. Import data from an online source of your own choosing (through download or an API).
  2. Present the data visually (and perhaps interactively).
  3. Apply some method(s) from descriptive economics (»samfundsbeskrivelse«). That is, make a report that tells a story in numbers and graphs about an economic phenomenon or trend.

  Example of structure: See this repository.
- **Structure:** Your data analysis project should consist of:
  1. A README.md with a short introduction to your project. It is written in Markdown like a Jupyter Notebook.
  2. A single self-contained notebook (.ipynb) presenting the analysis.
  3. Fully documented Python module files (.py).

4. Data. If you do not access data through an API call in your Python scripts, then you must store it in a .csv file that is uploaded to your repository along code files. Note that Github won't allow you to upload very large data sets (ie. multiple GBs).

- **Size:** *Quality before quantity.* Cleaning and joining a couple of data sets and then making some interactive figures might be just fine. Else you will be asked to extend it for the exam.
- **Hand-in:** On GitHub by uploading it to your repository in the dataproject folder:

  github.com/projects-2022-YOURGROUPNAME/dataproject/

- **Deadline:** 17th of April 23.59
- **Peer feedback:** After handing in, you will be asked to give peer feedback on the projects of two other groups. We will go over this in the lectures.
- **Exam:** Your data analysis project will be a part of your exam portfolio.
  You can incorporate feedback before handing in the final version.
- **Note** : You can find a few suggestions for APIs to use in the notebook for L08. You are very welcome to find data elsewhere!