



Lecture "Data Science Infrastructures"

Exercise 08 (DSI E08 ST24) – Related Lecture DSI L09

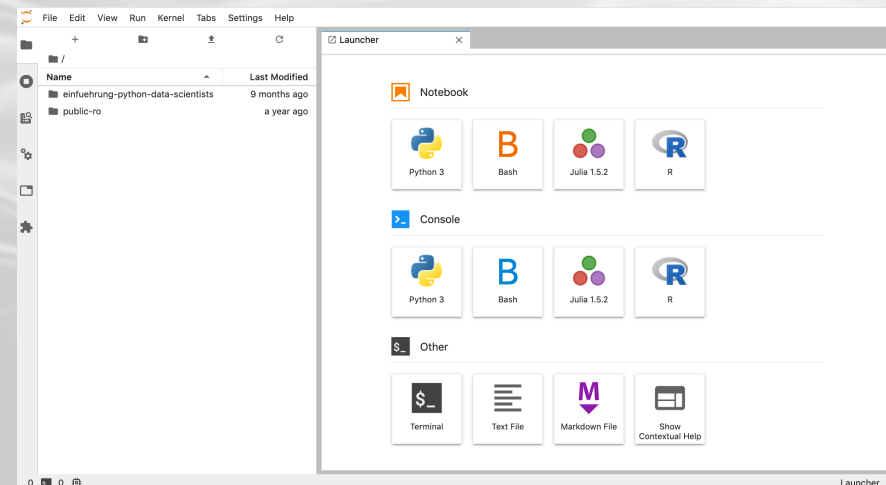
13.06.2024, Philipp Wieder, GWDG

- I. Data Visualization – Find the right representation for selected data sources
  - Solutions can be both submitted in German and English
  - Please submit the solutions to the “Exercise Solutions” folder in Stud.IP

# General Information

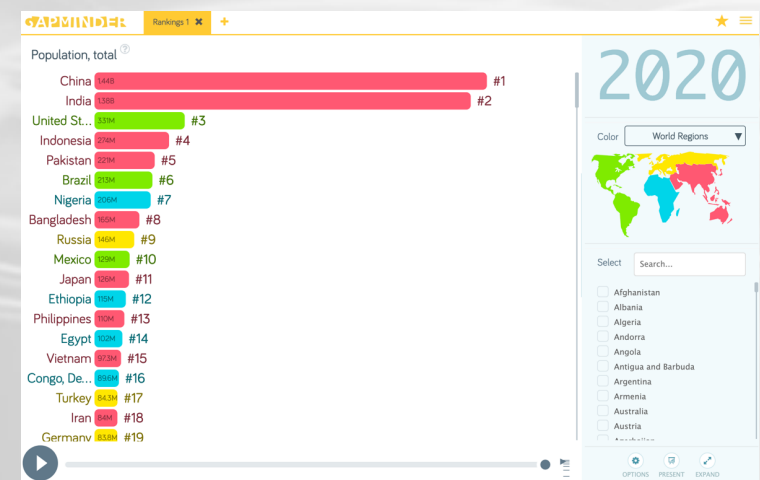


- You can use Jupyter Notebooks for this exercise. You can install it local or log into the GWDG Jupyter Cloud
  - Offers Jupyter Lab as cloud service
  - <https://jupyter-cloud.gwdg.de>
  - Works with your student account (NO further steps needed)
- Step-by-step assignment will be made available via Stud.IP.



# Data Sets

- We use data sets from <https://www.gapminder.org/>
  - Properly checked data
  - Freely available as csv
  - You can compare your results with the online version
- We work with data sets with different characteristics





# Assignment I

## Compare Different Charts



- Task: Compare different representations of one particular data set
  - Pandas & matplotlib libraries (in case of Jupyter/Python; other languages/tools are possible)
  - Create three different charts
  - Which one do you prefer?
- Tool: e.g. Jupyter Lab / Notebook
- Programming Language: e.g. Python

# Assignment II

## Choose “Best” Representation



- Task: Program most suitable representation for a given data set
  - Visualization library of your choice
  - Analyse data set and choose representation
  - Include relevant data (e.g. source) to the notebook
- Tool: e.g. Jupyter Lab / Notebook
- Programming Language: e.g. Python