IN-STK 5000

Today’s session

- General remarks
- Practicalities
- Some Python basics
- Exercises

About Dirk

- Associate Professor @ UiO
- Education in theoretical physics
  - MCMC simluations
  - Data analysis
- Day Job: VP Tech. Mgmt. in Equinor
- Worked many gigs in Data Science consulting
- Experience as senior software engineer

General Remarks

- Will try to be as hands-on as possible
  - Warning: Live coding
- Will draw on industry experience where possible
  - Aim: Make you hirable
- There will be practical exercises, voluntary, but recommended
  - If they are too boring for you, make your own!
  - Lots of hands-on experience with different data sets is important!
- Materials: https://github.com/dhesse/IN-STK5000-Autumn21

- Exercises: https://github.com/dhesse/IN-STK-5000-Autumn-21—Exercises

Python

- More specifically: Python 3.9
  - You have several options: **Pyenv**, Hombrew/other PMs, Anaconda, …
- You should be somewhat familiar with programming
  - Ideally in Python
- Work in groups, learn from each other
- If you don’t know python: Dive Into Python 3
- Why?
  - Real programming language
  - Widely used
    - You can get help
    - Makes you hirable

    - Plenty fast for us
- Get a good text editor (VSCode, Emacs, Vim …)

Jupyter

- `jupyter lab` will be used for teaching
- Graphical
- Interactive
- Great tool for exploring and teaching
- For any serious work, use a script instead!
  - C.f. the exercises

`git`

- Version control system
- Exercises, notes, etc. will be distributed via GitHub
- Use of git for project work strongly encouraged
  - Look at ‘pull requests’ to collaborate
- https://git-scm.com/docs/gittutorial

The plan

- **Session 1**
  - Python basics
  - Jupyter
- **Session 2**
  - Basic data loading and manipulation
- **Session 3**
  - More advanced data manipulation
  - Time series
- **Session 4**
  - Training machine learning models
  - Evaluating model performance
- **Session 5**
  - Automating ML workflows
- **Session 6**
  - Tensorflow/Keras

Exercises

- There will be exercises, voluntary but recommended
  - Get you some extra hands-on experience
- Solution skeleton on GitHub, together with solution templates and actual solutions
- Feel encouraged to experiment beyond those!