Python Tsunami

Part 1: Intro to Python by HeaDS

Who are we?

Who are we

UCPH HeaDS

Center for Health Data Science (HeaDS)

- The Data Lab
  - Provides data science support for all research groups at SUND
  - Organizes workshops/seminars

- Research Units
  - work on different areas and topics within the field of health data science

- The Sandbox
  - Develops HPC environments for training, teaching and testing

https://heads.ku.dk

Who are we

Center for Health Data Science (HeaDS)

- Upcoming events:
  - Introduction to Machine Learning: 24-28 April
  - GDPR for Biomedical Researchers: 24-25 May
  - Just Bash It: 5-6 June
  - Excel to R: 12 + 14 June
  - RNAseq analysis: 19-21 June

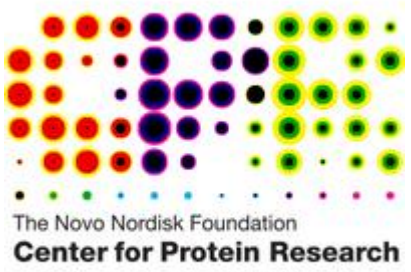
Check our website calendar! ->https://heads.ku.dk

About this course

About this course

Originally developed at the Center for Protein Research (CPR) by:

- Alberto Santos Delgardo (University of Oxford)
- Henry Webel (NNF CPR)
- Annelaura Bach Nielsen (NNF CPR)
- Rita Colaço (PRI)



We say thank you for the course material which we have adapted.

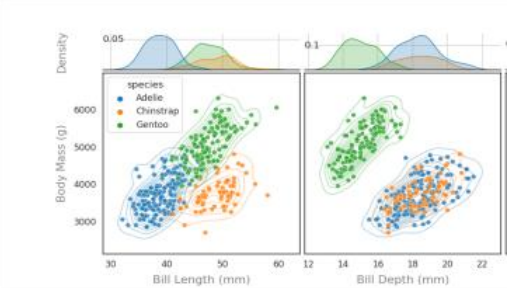
About this course

Starting Time	Day 1	Day 2	Day 3
8:30	Morning Coffee (optional)		
8:45	Intro	Pandas I	Quiz
9:15	Variables & Datatypes		
9:50	Break		Break
10:00	Iterables I		Visualization
10:15		Break	
10:45	Break	Pandas II	
11:00	Iterables II		
12:15	Lunch		
13:15	Booleans, Operators and Conditions	Pandas III	Dataset exercise
14:15	Break	Break	
14:30	Loops	Pandas IV	
15:30			
16:00	Finished		

What is programming?

What is programming?

Programming is a set of **machine-readable** instructions that transform your input into your desired output.



input → program → output

We will attempt to shed some light on this part

Why is programming nice?

UCPH HeaDS

- Learning by doing:

  Difficult to ‘break’ a computer with wrong programming
- Reproducibility:

  The same thing should happen every time you run (*though some tasks involve some randomness)
- Transferable:

  Easily share your work with colleagues
- Many useful online resources
- Automate complex analysis workflows
- Important tool for working since we live in a data driven world

Why Python?

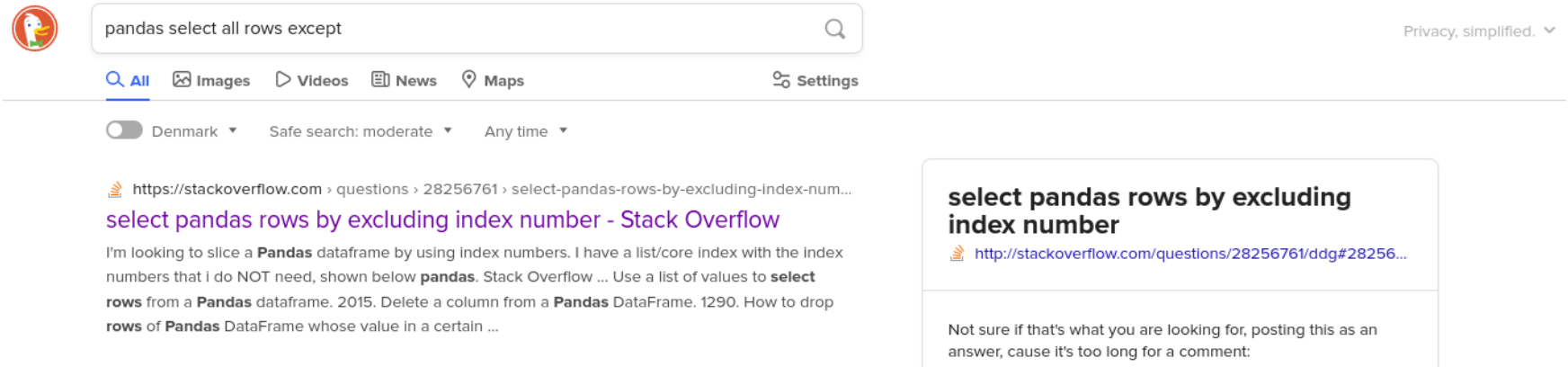
Python is a great programming language for both beginners and advanced programmers:

- Easy to grasp, close to natural language
- Many learning resources available
- Large community (i.e. stackoverflow for questions)
- Libraries
- Can do very advanced things like neural networks

Online communities - Where to get help

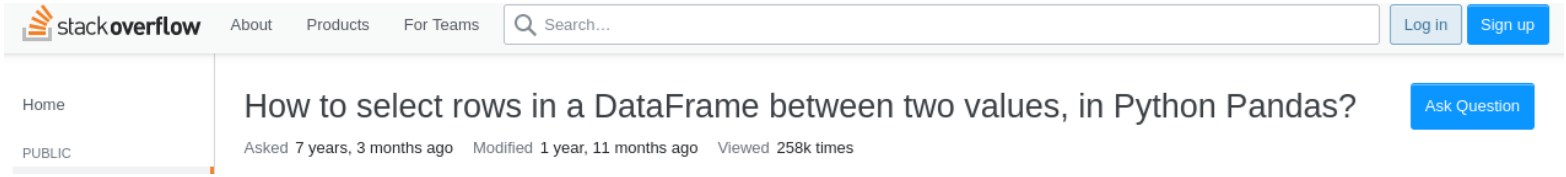
Online communities such as stackoverflow are an important tool in programming.

Nobody knows everything, but together we know more than ever before!

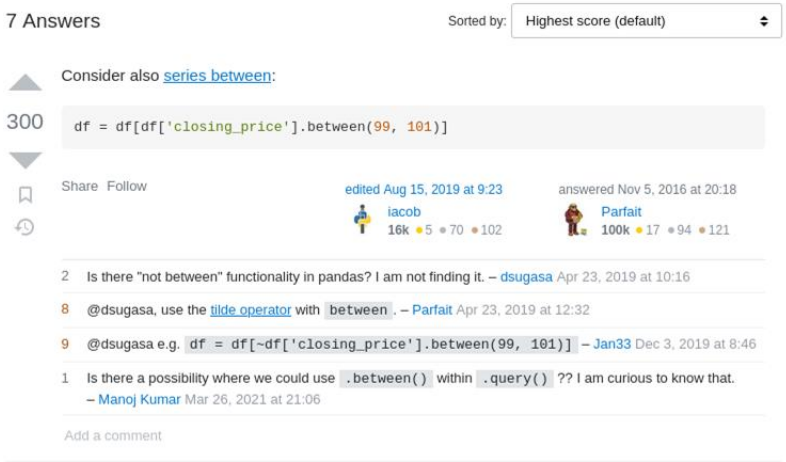


Online communities - Where to get help

You have a question ....



Someone out there has an answer!



Using libraries/packages

Libraries

Python has many libraries, also called packages, that other programmers have developed. Find and **use** them!

Well-maintained libraries generally are:

- Tested
- Optimized
- Documented

There is no need to reinvent the wheel. During this course we will use:

- Pandas (all the data analysis!)
- Math (basic math)
- Plotly express (visualization)

Libraries

If you are running Python from a local installation, you need to have libraries **installed** before you can use them.

On Google Colab you can generally just import, they are already installed.

- Import the math library:

```
import math
```

- Now I can use functions from that library, i.e. calculating the logarithm or square root:

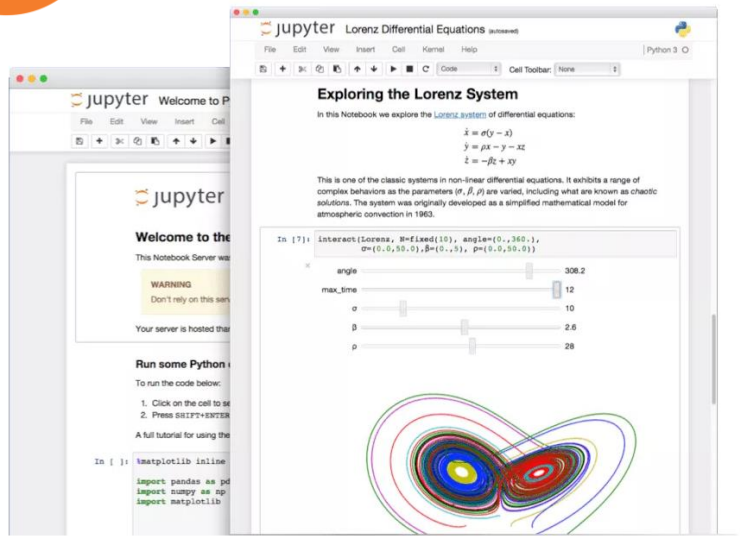
```
math.log(3)
```

```
math.sqrt(4)
```

Python environments

Jupyter notebook

The Jupyter Notebook is an **open-source application** to create and share documents that contain code, equations, visualizations and text (markdown).



- Browser-based development environment for creating, running and sharing Python code
- Combine code with text and output
- Runs from your **local installation**. I.e., you need Python and the libraries you want to use installed on your computer

Google Colab

Google Colab is a Jupyter Notebook hosted on Google's servers, not your own machine. It still runs in your browser.

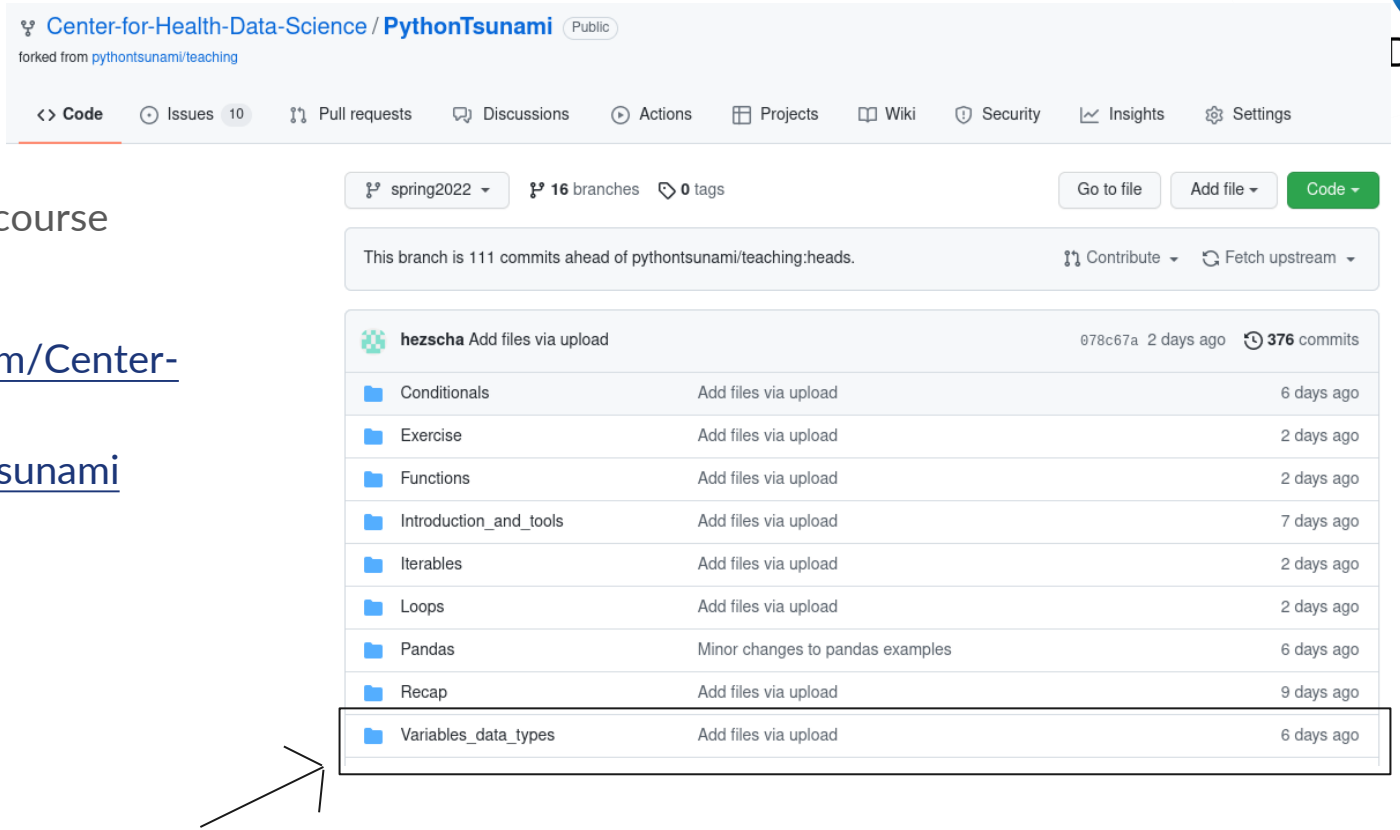
- tool to write, execute and share Python code through the browser
- requires no setup to use and provides free access to computing resources on Google's servers including GPUs
- is connected to a Google account and data and notebooks can be accessed through Google Drive.

We'll use Colab during the course.

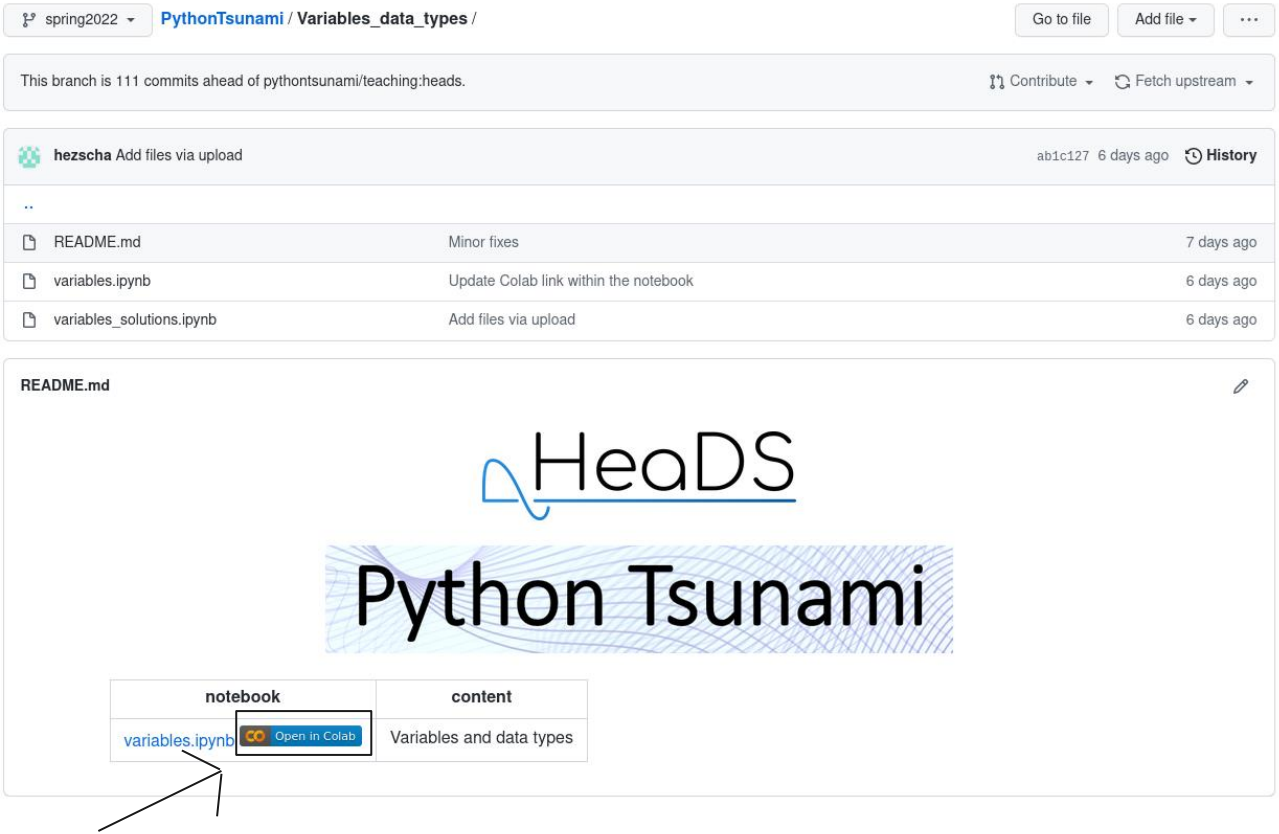
Course material

You can find the course material here:

https://github.com/Center-for-Health-Data-Science/PythonTsunami

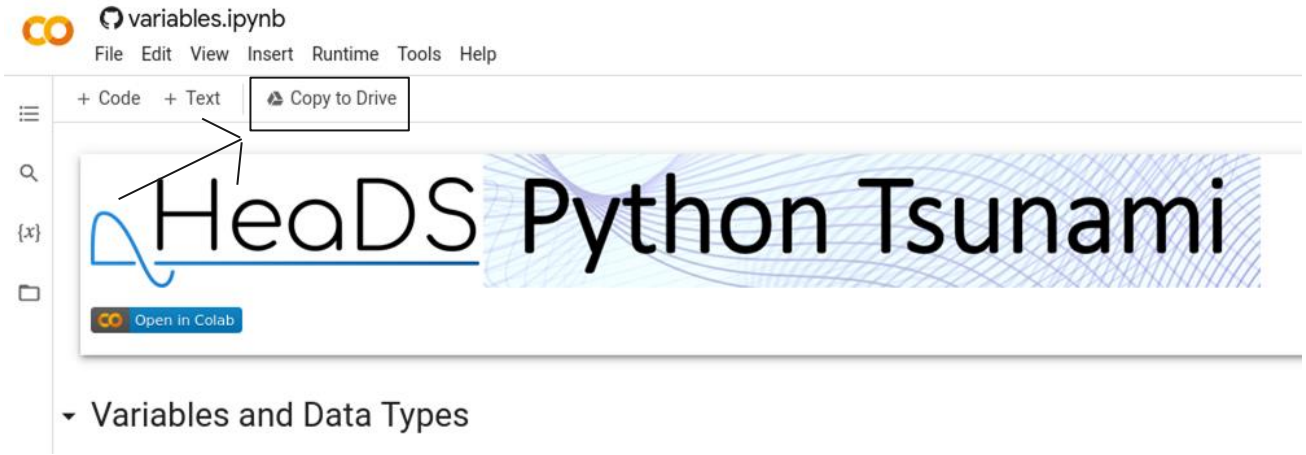


Course material



Course material

Remember to **save a copy** to your own google drive so you can save your notes and exercises!



Short Introduction

Take the next 5 mins to introduce yourself at your table:

- Name
- Position
- Unit
- What you do (very briefly!)