HANDS-ON DATA WORKSHOPS

PYTHON PLAYGROUND

While you are waiting to start please download the files for this afternoon

https://github.com/epfl-exts/PythonPlaygroundAfternoon

Sue Cheatham

HANDS-ON DATA WORKSHOPS

PYTHON PLAYGROUND

Beginner skill level workshop 26 + 27 May 2018

Sue Cheatham

Timetable

Welcome back!

Afternoon session

13:30 Data Analysis Introduction
Walk through data analysis
Data sets to investigate with code to run or templates to code yourself

15:00 Coffee Break

15:30. Maps
Data sets and code to run to create some choropleth maps

16: 30 Review
17: 00 End of day

Data Analysis

Data analysis often involves the creation and study of graphical representations of the data
Eg bar graphs, scatter plots

The goal of data analysis is to

- discover useful information,
- suggest conclusions,
- support decision-making.

Data

Open data is freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control

We will use open data from a number of sources

Data comes in various shapes, sizes and formats

Today will just concentrate on extracting some insights from comma separated files

The aim of this afternoon is to plot some basic graphs and data visualisations

Data in tables

Data is often organised in tables, which form **rows** and **columns**
Columns are sometimes called fields or attributes
An **index** uniquely identifies a row of data
A missing value can result in an empty **cell**, NULL or NaN

Index

	Variety	Color	Synonym	Percentage Area	Area hectares
0	Pinot noir	red	Blauburgunder	29.7	4402
1	Chasselas	white	Gutedel	27.1	4013
2	Gamay	red	NaN	10.2	1514
3	Merlot	red	NaN	6.9	1028
4	Müller-Thurgau	white	NaN	3.3	493
5	Gamaret	red	NaN	2.6	380
6	Chardonnay	white	NaN	2.2	321
7	Sylvaner	white	Rhin	1.6	241

Column names

Row

Column

Percentage of seats held by women in national parliament

Data taken from The Humanitarian Data Exchange https://data.humdata.org/

Looking at evolution of numbers for Switzerland, Germany, France and Italy from 1997 - 2015

Jupyter notebook: `DataAnalysis/ParliamentSeatsWomen.ipynb`

Data: `DataAnalysis/Data/percentageWomenNationalParliament.csv`

NB file endings

Population density

Data taken from World Bank https://data.worldbank.org/

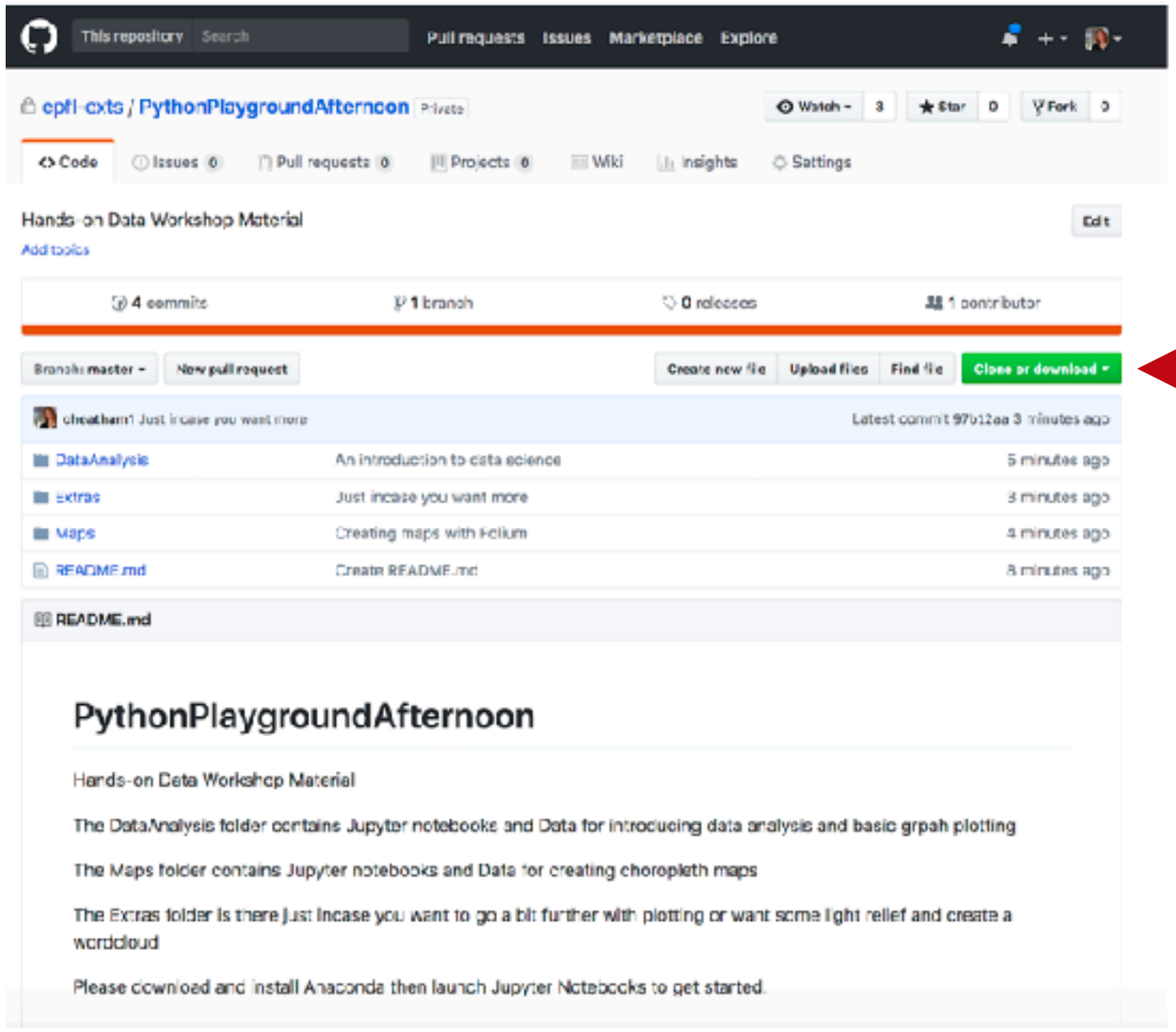
Looking at evolution of population density for 54 countries from 1961 - 2013

Jupyter notebook: `DataAnalysis/PopulationDensity.ipynb`

Data: `DataAnalysis/Data/populationDensity.csv`

Data and Files

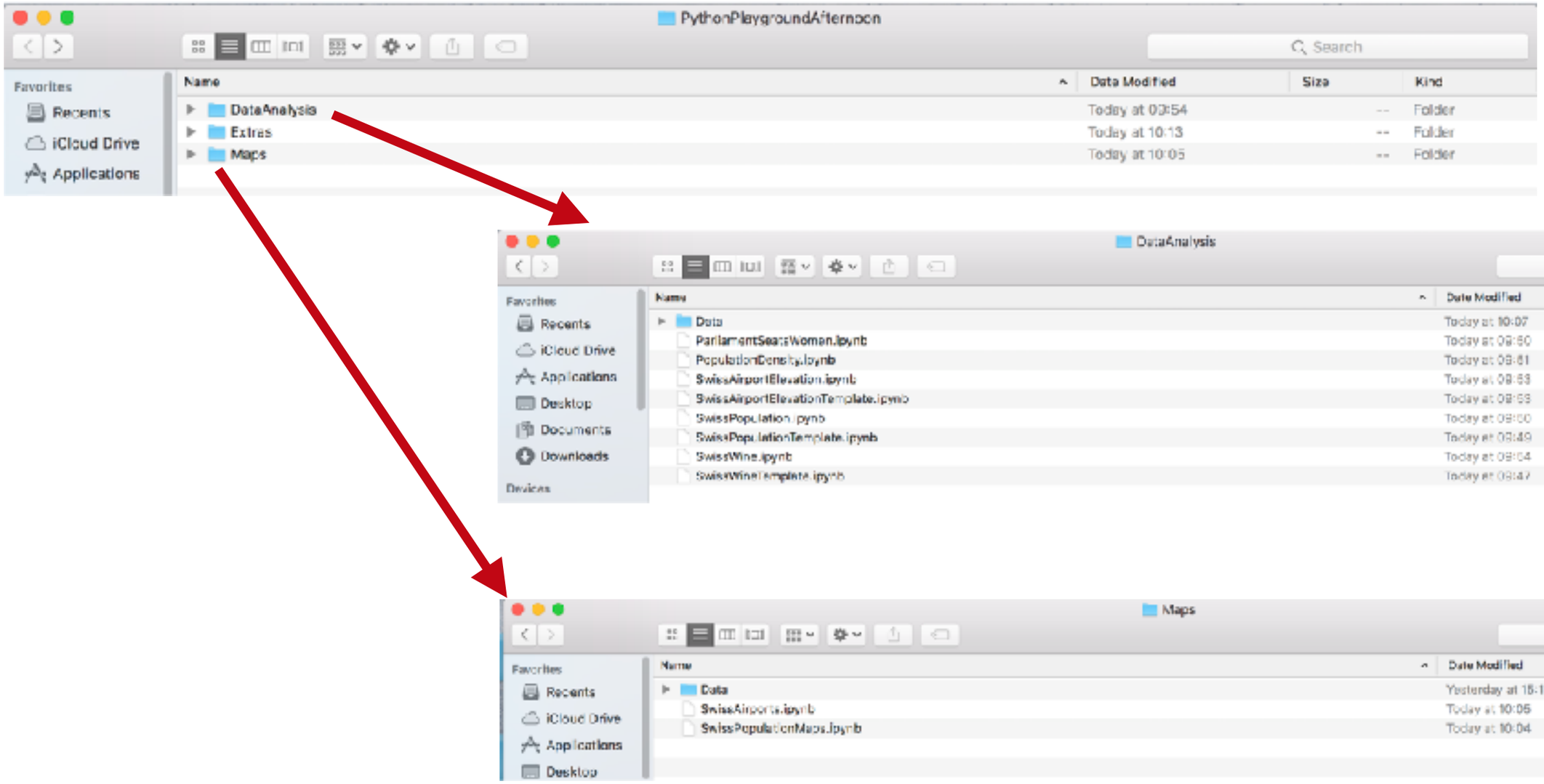
https://github.com/epfl-exts/PythonPlaygroundAfternoon



Folders and Files

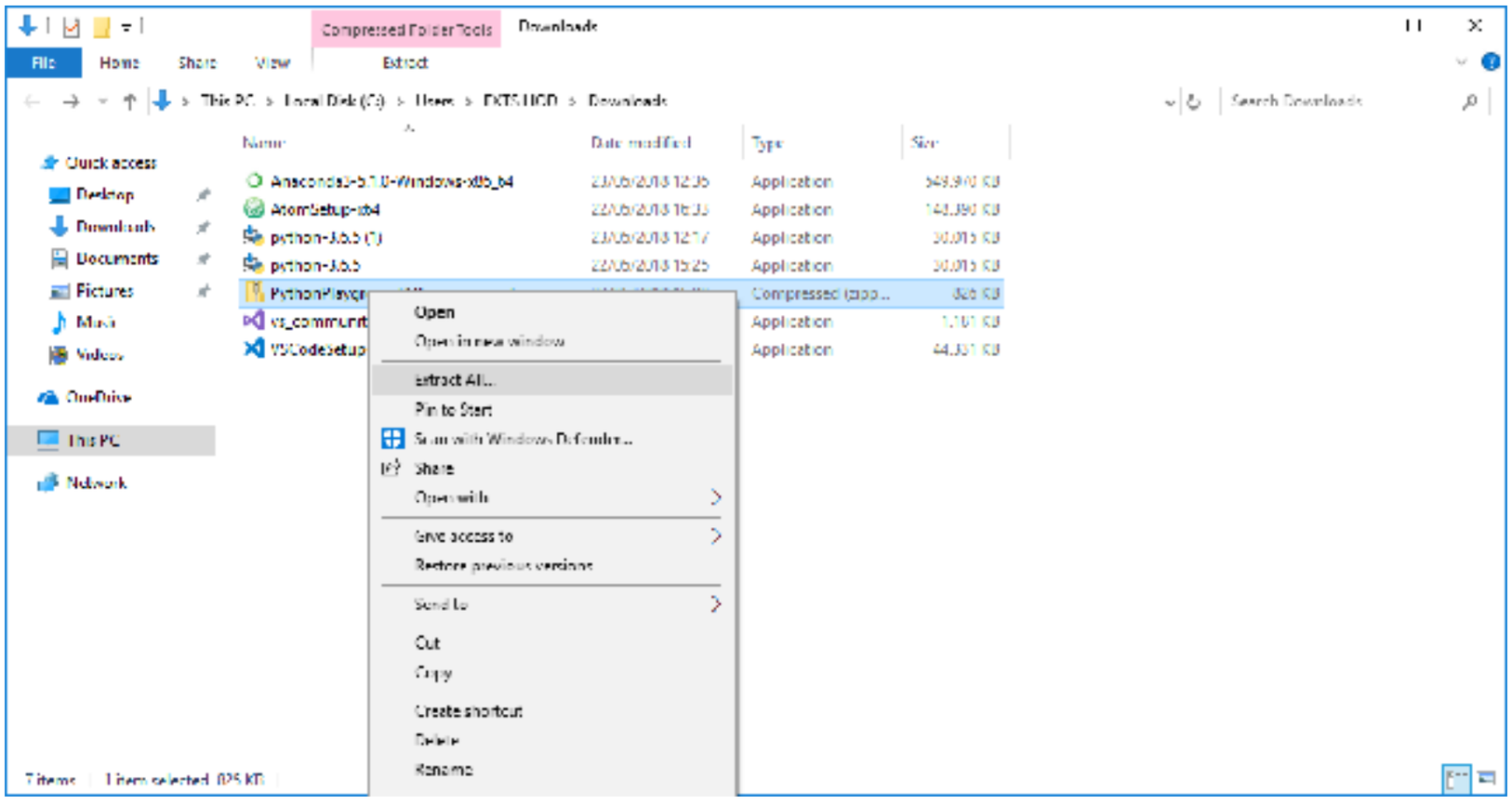
When you have downloaded your files onto your laptop you can see the same folders and files

Please move PythonPlaygroundAfternoon folder to your Desktop



Folders and Files

Download ZIP file, save and open
The downloaded folder will be called PythonPlaygroundAfternoon-master
This is compressed(zipped)
Right click on file and ‘Extract All’



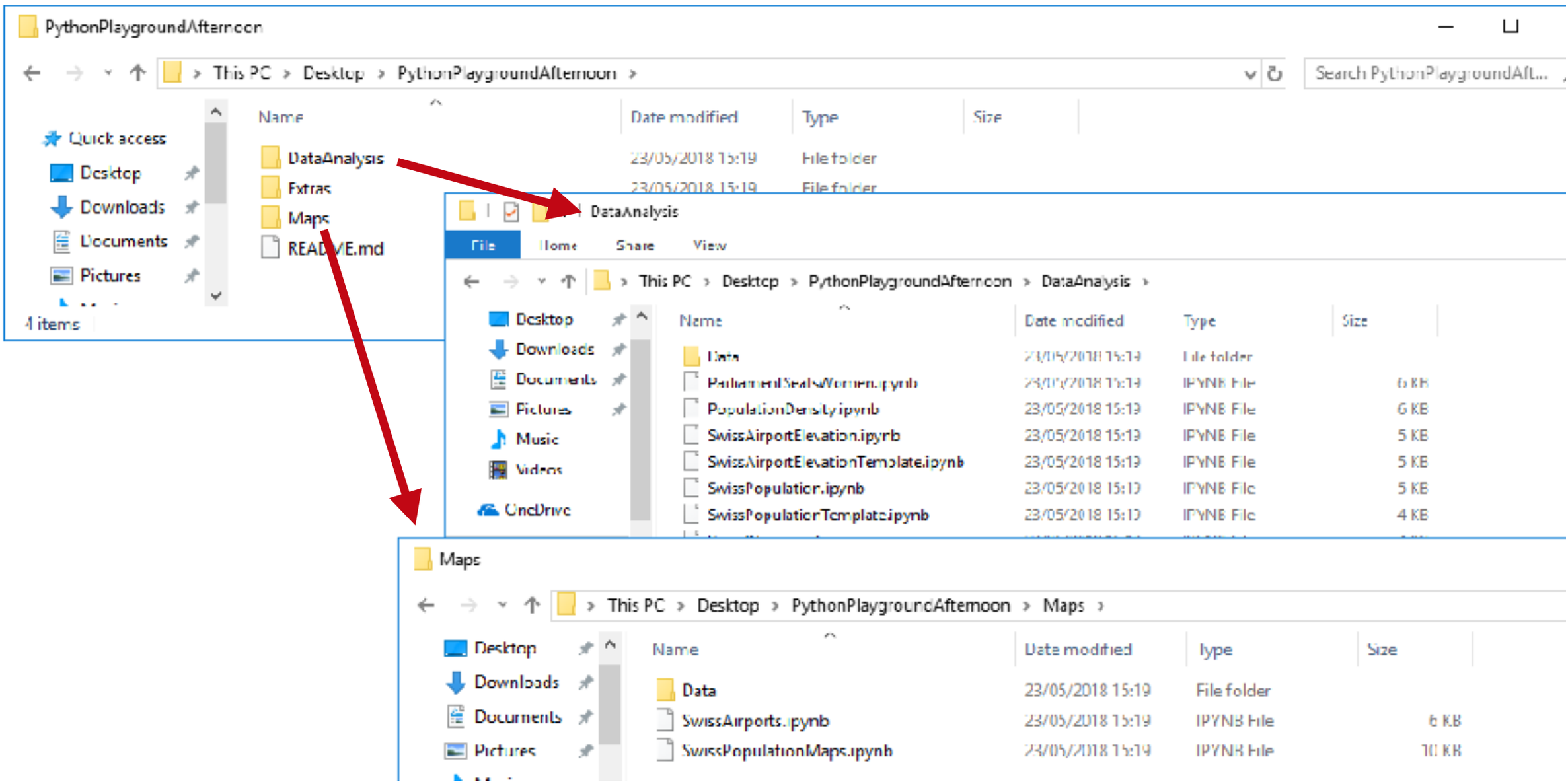
Drag the PythonPlaygroundAfternoon-master folder and drop onto the Desktop.
Rename the folder PythonPlaygroundAfternoon

EPFL EXTENSION SCHOOL

Folders and Files

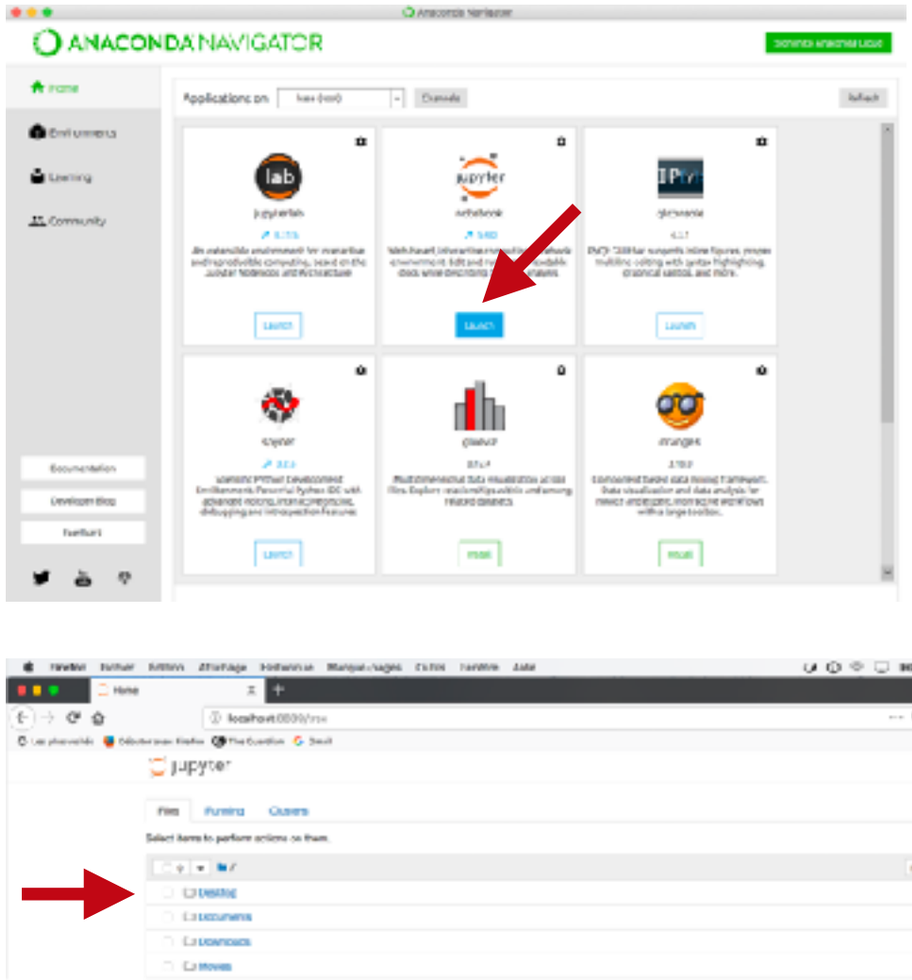
When you have downloaded your files onto your laptop you can see the same folders and files

Please move PythonPlaygroundAfternoon folder to your Desktop



Running a Jupyter Notebook

Step1
Launch Anaconda-Navigator



Step 2
Then launch a Jupyter notebook

This opens a window in your browser

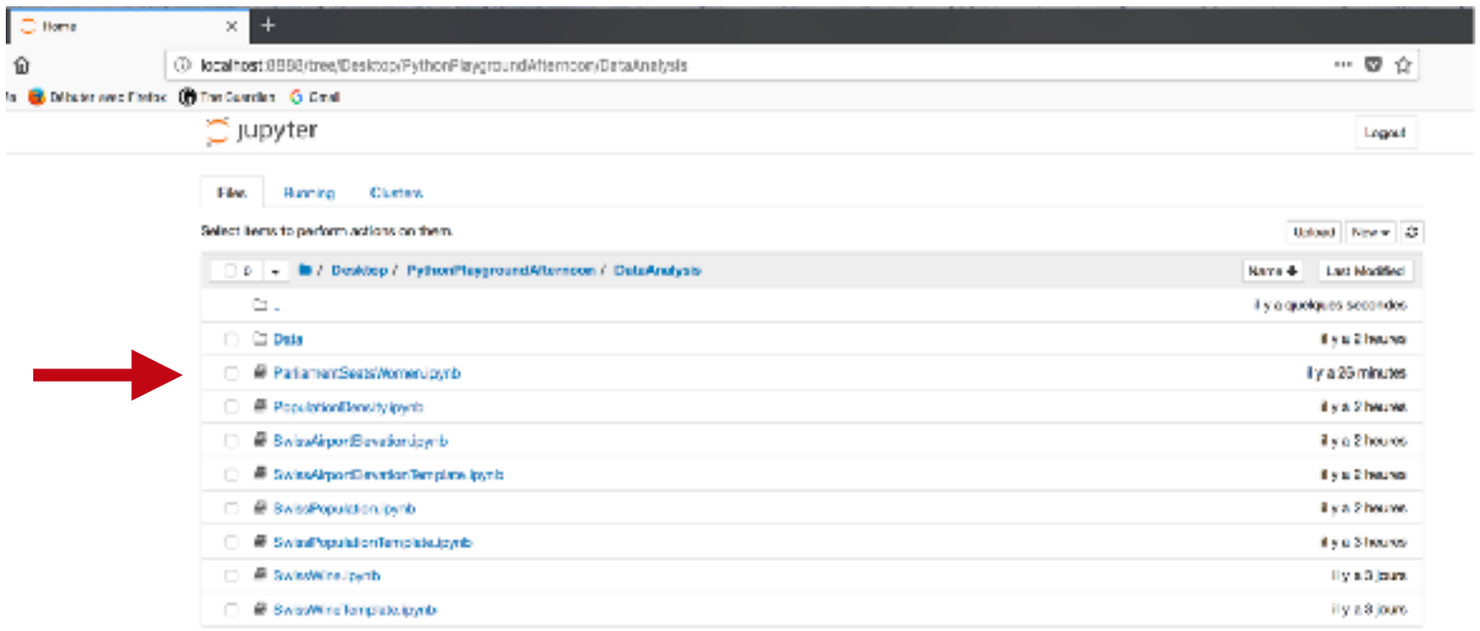
Step 3
Select Desktop
Then PythonPlaygroundAfteroon

You will then see the two folders for this afternoon
DataAnalysis and Maps

Running a Jupyter Notebook

Step 4

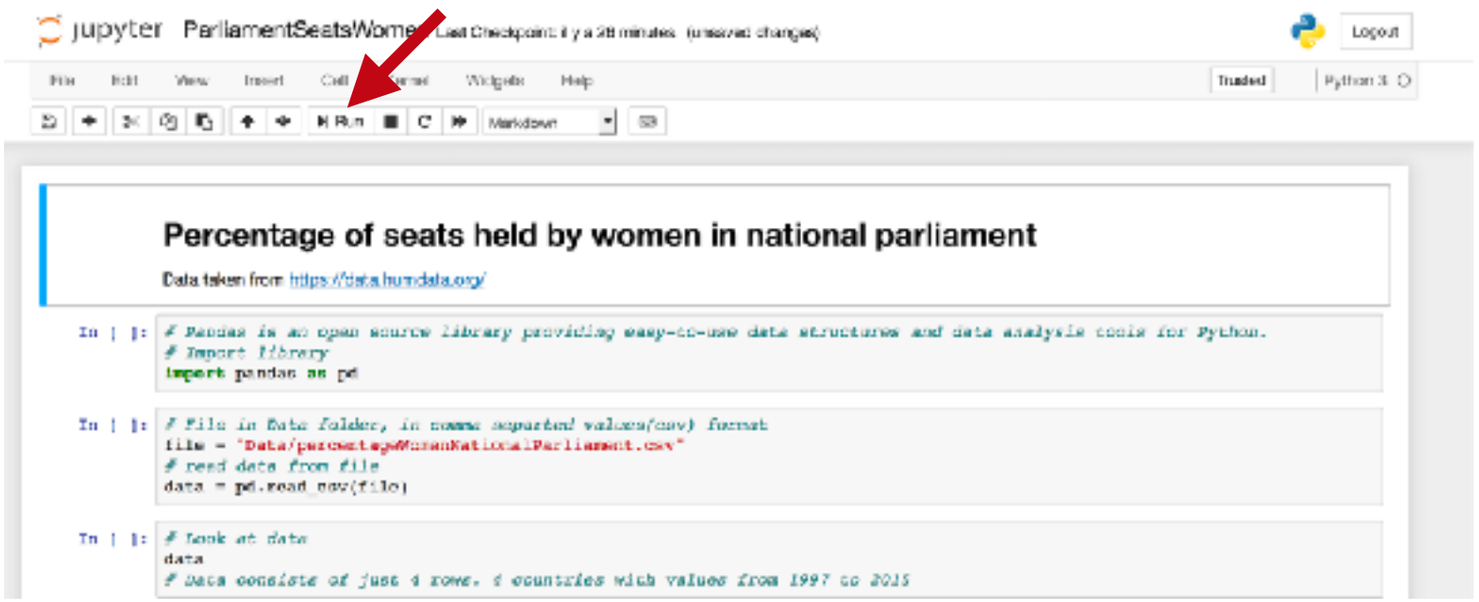
In the DataAnalysis folder
Open an interactive python notebook



Step 5

Step through the code by hitting the Run button

Before you hit run, think. What do you expect the code to do??



Run some code or write your own

We will step through some notebooks together
Then either run some of the prepared notebooks or
Follow the instructions in the templates and write your own code using the previous examples to help

1. Jupyter notebook template: `DataAnalysis/SwissPopulationTemplate.ipynb`

   Jupyter notebook worked example: `DataAnalysis/SwissPopulation.ipynb`

   Data: `DataAnalysis/Data/populationSwitzerland.csv`

2. Jupyter notebook template: `DataAnalysis/SwissAirportElevationTemplate.ipynb`

   Jupyter notebook worked example: `DataAnalysis/SwissAirportElevation.ipynb`

   Data: `DataAnalysis/Data/ch-airports.csv`

3. Jupyter notebook template: `DataAnalysis/SwissWineTemplate.ipynb`

   Jupyter notebook worked example: `DataAnalysis/SwissWine.ipynb`

   Data: `DataAnalysis/Data/swissWine.csv`

Map of Swiss airports

Produce map of Switzerland with Swiss airports marked on.

Data taken from The Humanitarian Data Exchange https://data.humdata.org/

Jupyter notebook: `Maps/swissAirports.ipynb`

Data: `Maps/Data/ch-airports.csv`

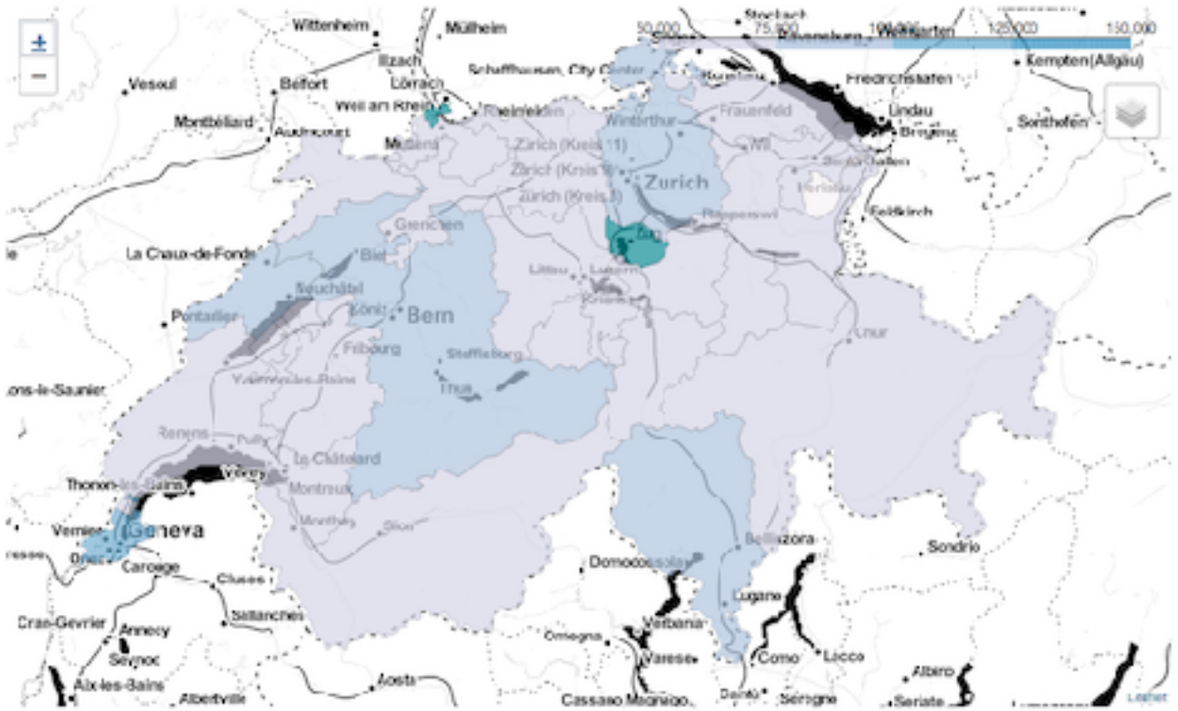
We will use Folium to create interactive maps
At the prompt type
>conda install folium

Choropleth maps of Swiss population

Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area or show the level of variability within a region.

Produce a map of Switzerland displaying information per canton: population, population density and GDP per capita.

Data taken from wikipedia



Jupyter notebook: `Maps/SwissPopulationMaps.ipynb`

Data: `Maps/Data/populationSwitzerland.csv`

Review

What activities did you like today? Why?
What activities did you not like today? Why?

Is there anything you would have liked to have been included that was not covered today?

Please share any good photos of the day

If you enjoyed the day don't forget to post on social media!

Please give me feedback and stay in touch: susan.cheatham@epfl.ch