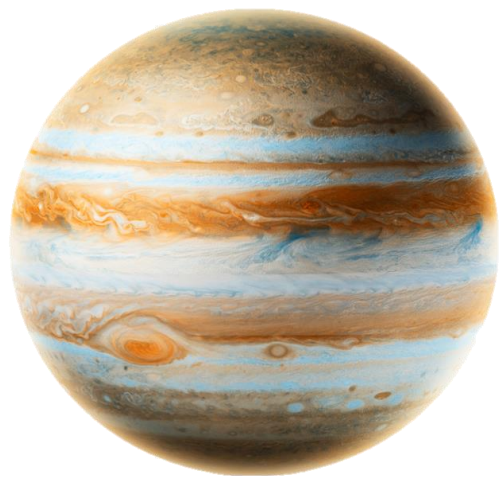


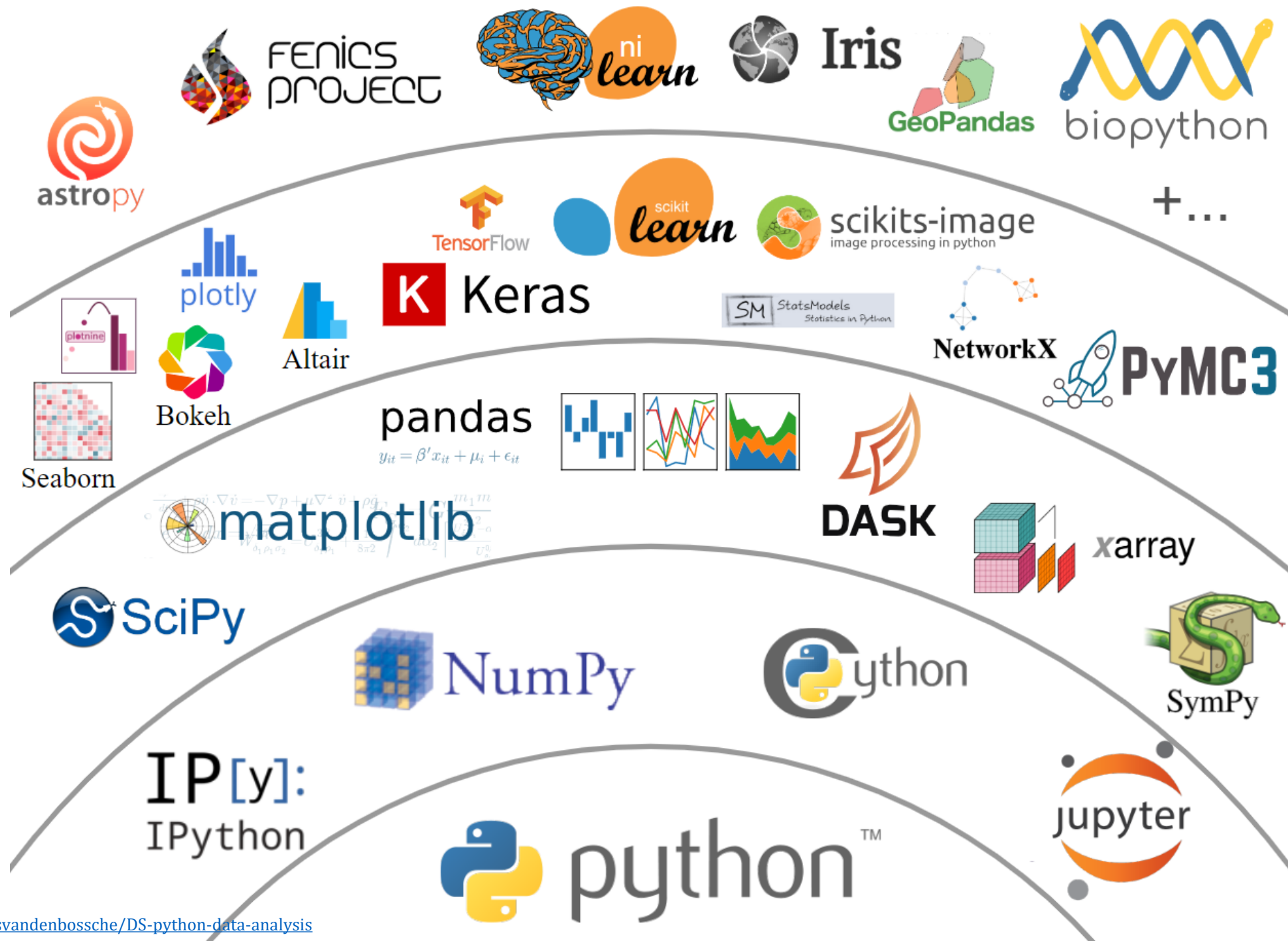
Data analysis and programming with Python

PhD course, 20/04/2020

Giacomo Bellandi

giacomo.bellandi@am-team.com





Sources for this course

Installation instructions:

- <https://jorisvandenbossche.github.io/DS-python-data-analysis/setup.html>

Many options to get started:

- <https://docs.scipy.org/doc/numpy/user/quickstart.html>
- https://www.w3schools.com/python/python_getstarted.asp
- <https://www.python.org/about/gettingstarted/>

Documentation and exercises:

- <http://scipy-lectures.org/>
- <https://github.com/jorisvandenbossche/DS-python-data-analysis>
- https://github.com/stijnvanhoey/course_python_introduction

Software

- <https://anaconda.org/>

Why Python

Large community

Fast development

Standard for GIS

Multi-platform (Win, Linux, OS, ..)

Multi-functional (web development, data analysis, automate recurring tasks, ..)

Open source

On the other hand, commercial tools are:

Well supported with direct help

Client friendly

Conda:

Robust and updated library of supported packages

Less installation/compatibility issues

Conda envs:

Separate dependencies for specific packages of a specific project

Different versions of python possible

Easily shareable and reproducible with others

Environments

ANACONDA NAVIGATOR

Sign in to Anaconda Cloud

Home

Environments

Learning

Community

Documentation

Developer Blog

Search Environments

base (root)

COVID19

pysumoGSA

data-analysis-course

Installed Channels Update index... Search Packages

Name	T	Description	Version
backports			1.0
backports.funtool...		Backport of funtools.lru_cache from python 3.3 as published at activestate.	1.6.1
backports.shutil_g...		A backport of the get_terminal_size function from python 3.3's shutil.	1.0.0
backports.shutil_w...		Backport of shutil.which from python 3.3	3.5.2
backports_abc		A backport of recent additions to the 'collections.abc' module.	0.5
blas			1.0
bleach		Easy, whitelist-based html-sanitizing tool	3.1.1
ca-certificates		Certificates for use with other packages.	2018.4.16
certifi		Python package for providing mozilla's ca bundle.	2018.4.16
colorama		Cross-platform colored terminal text.	0.4.3
configparser		This library brings the updated configparser from python 3.5 to python 2.6-3.5	3.5.0
cycier		Composable style cycles.	0.10.0
decorator		Better living through python with decorators.	4.4.1
defusedxml		Xml bomb protection for python stdlib modules	0.6.0
entrypoints		Discover and load entry points from installed packages.	0.2.3
enum34		Python 3.4 enum backported to 3.3, 3.2, 3.1, 2.7, 2.6, 2.5, and 2.4	1.1.6
freetype		A free, high-quality, and portable font engine	2.9.1
funtools32		Backport of the funtools module from python 3.2.3 for use on 2.7 and pypy.	3.2.3.2
futures		Backport of the concurrent.futures package from python 3.2	3.3.0
icc_rt		Intel runtime libraries for c, c++ and fortran compilers	2019.0.0
icu		International components for unicode.	58.2

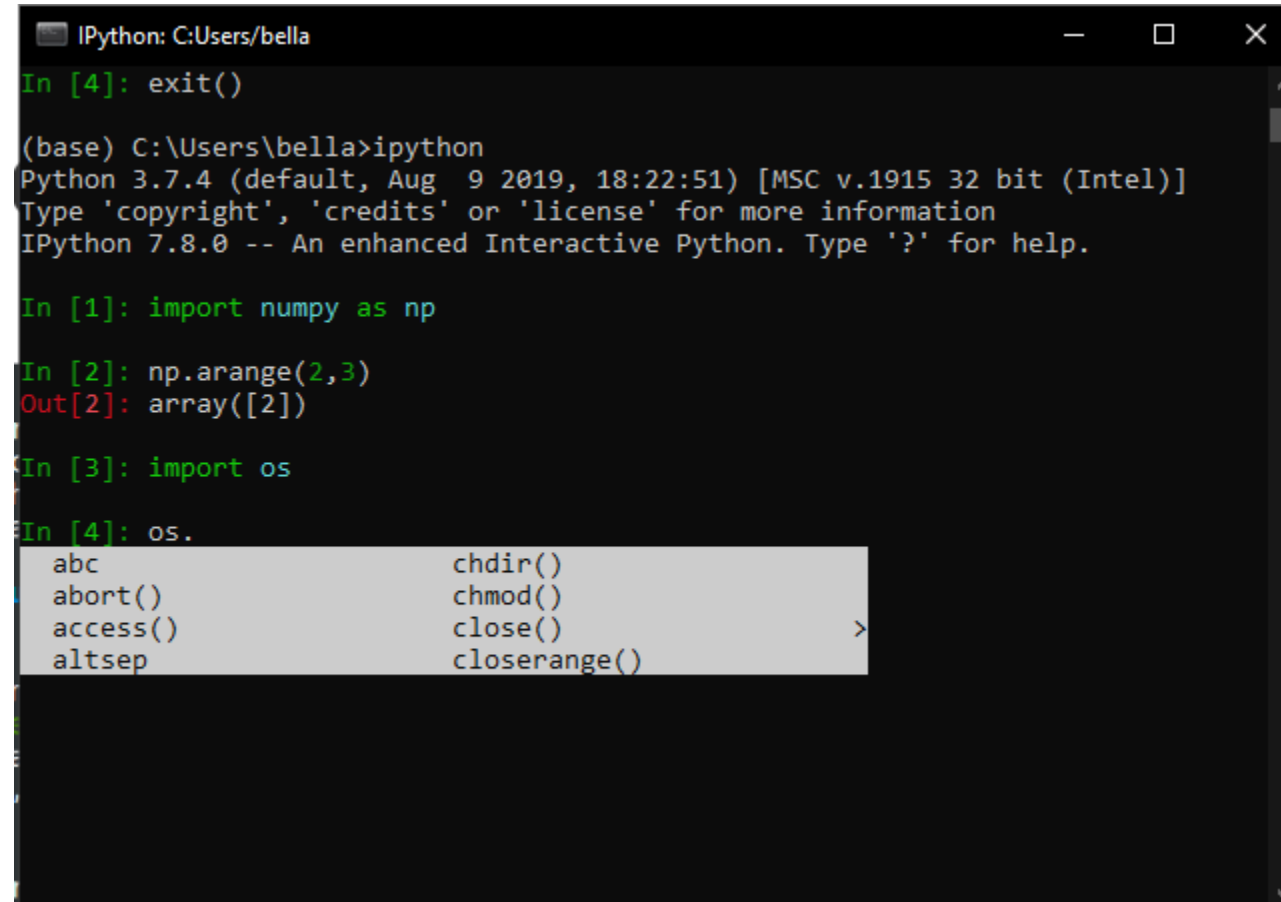
Create Clone Import Remove

Importing environment C:\Users\bella\conda\envs\data-... Fetching pycparser-2... Cancel

Popular commands from cmd/Terminal

<https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html>

iPython



```
IPython: C:\Users\bella
In [4]: exit()

(base) C:\Users\bella>ipython
Python 3.7.4 (default, Aug  9 2019, 18:22:51) [MSC v.1915 32 bit (Intel)]
Type 'copyright', 'credits' or 'license' for more information
IPython 7.8.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: import numpy as np

In [2]: np.arange(2,3)
Out[2]: array([2])

In [3]: import os

In [4]: os.
abc      chdir()
abort()  chmod()
access() close()
altsep   closerange()
```

GUI for python?

Spyder (similar to Matlab, Rstudio,..) is installed with Anaconda but normally used as testing environment

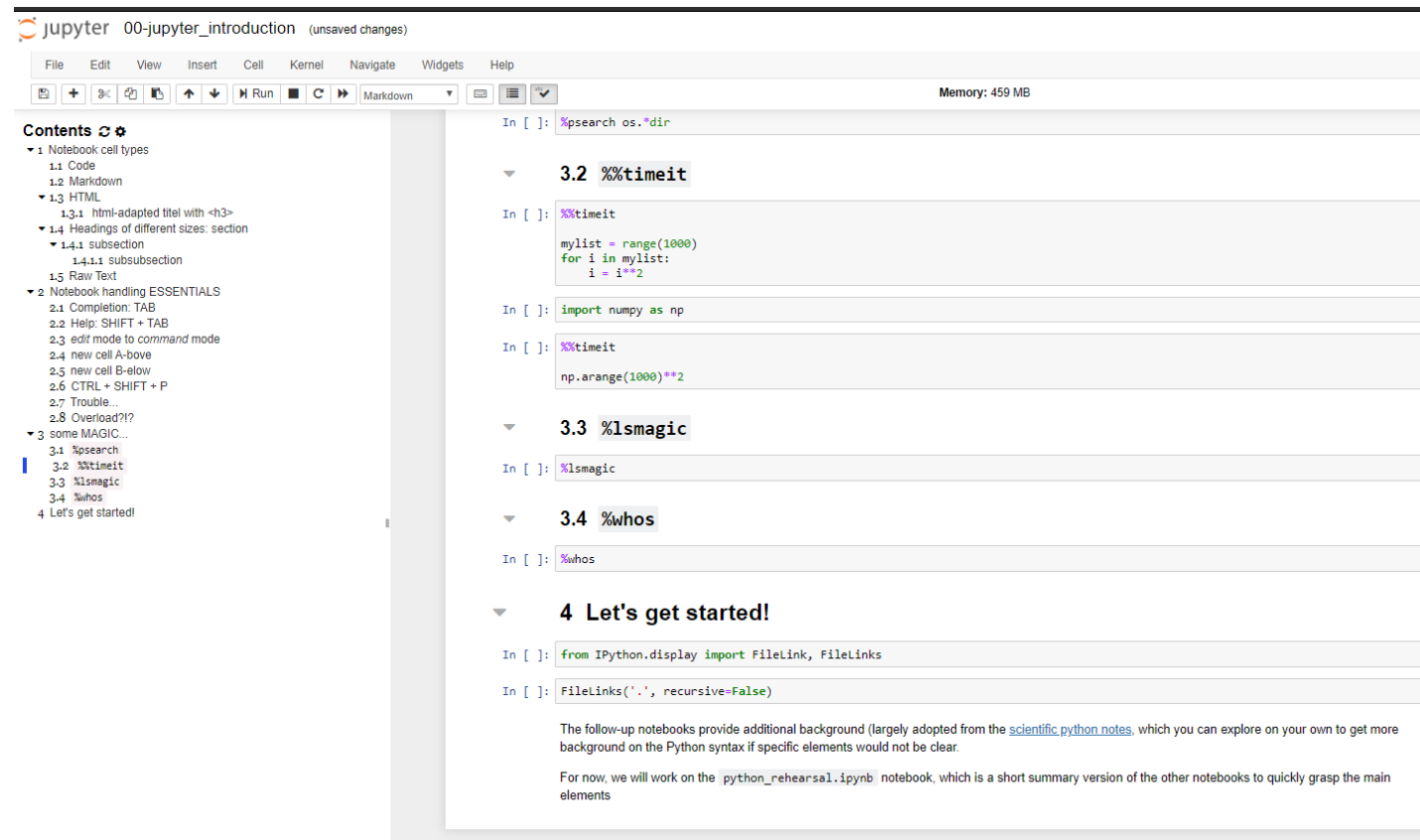
PyCharm, Eclipse, VS Code, are alternatives for web development etc.

We'll write down our code in **Notebooks**



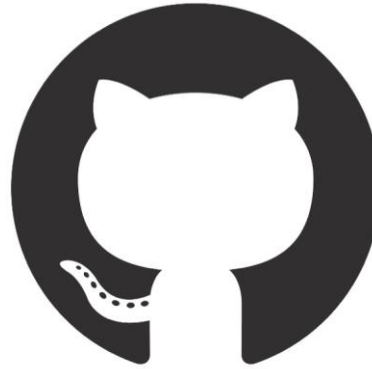
Jupyter Notebook

This is an interactive scripting environment, perfect for prototyping your scripts, but also for keeping track of workflow in specific projects, writing reports minimizing recursive data handling, ...



<https://jupyter-contrib-nbextensions.readthedocs.io/en/latest/install.html>

Managing scripts



Current repository
Pyhon_basicsCourse

Current branch
master

Fetch origin
Last fetched 22 minutes ago

Changes 1

History

notebooks\4_Time_operations.ipynb

@@ -0,0 +1,539 @@

1 changed file

notebo...4_Time_operations.ipynb

Create 4_Time_operations.ipynb

Description

Commit to master

IPYNB File

IPYNB File

Search or jump to...

Pull requests

Issues

Marketplace

Explore

gbellandi / Pyhon_basicsCourse

Unwatch 1

Star 0

Fork 0

Code

Issues 0

Pull requests 0

Actions

Projects 0

Wiki

Security

Insights

Settings

First tools to start using python for data analysis and modelling

Manage topics

3 commits

1 branch

0 packages

0 releases

1 contributor

MIT

Branch: master

New pull request

Create new file

Upload files

Find file

Clone or download

gbellandi a condensation of exercises for starting with python

Latest commit a37f8e7 27 minutes ago

data

img

notebooks

.gitattributes

.gitignore

LICENSE

README.md

environment.yml

README.md

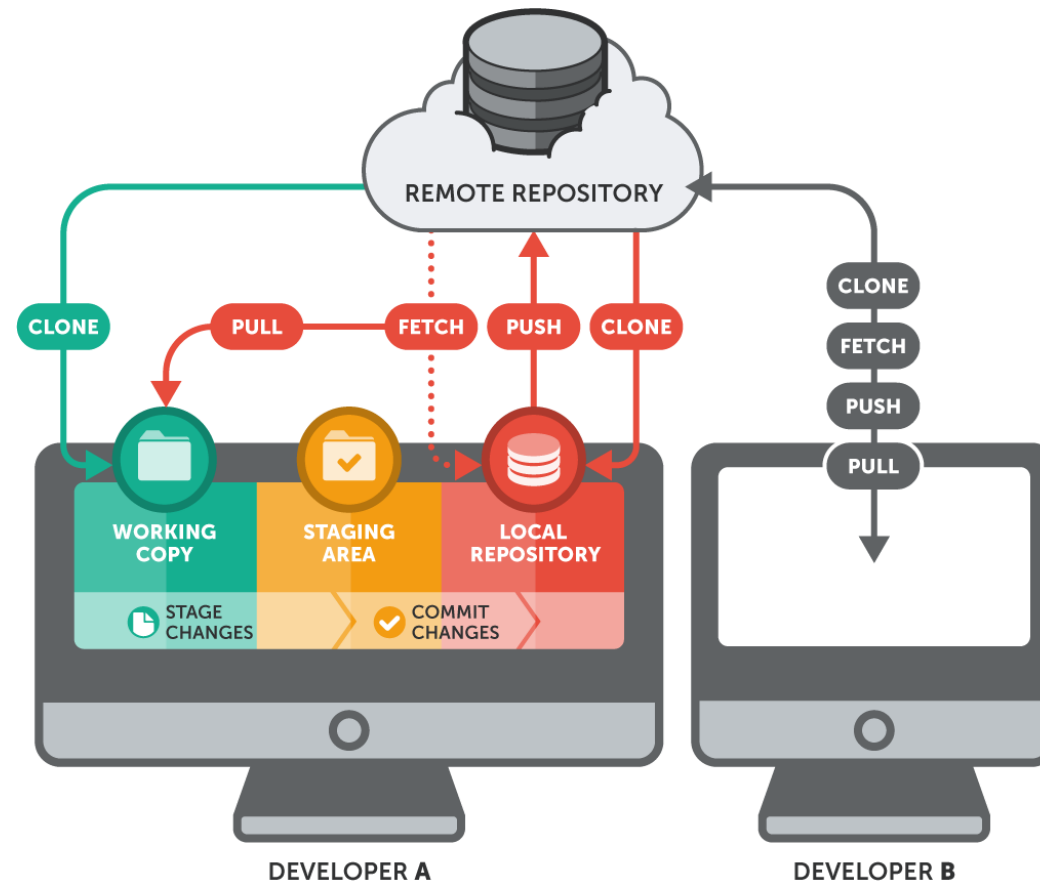
Pyhon_basicsCourse

First tools to start using python for data analysis and modelling

The material shown in these notebooks is a condensed version of the material gathered in the years from different sources that I consider very useful for a beginner. Among the most inspiring sources:

MC_Crema_Raceway2016.ipynb	21-Apr-17 11:43 AM
MC_report_notebook_EHV16.ipynb	28-Nov-19 3:04 PM
MC_RomaEst_Sept2017.ipynb	19-Sep-18 10:49 AM
MC_RomaEst2017.ipynb	14-Sep-18 10:49 AM
MC_RomaEst2018.ipynb	19-Sep-18 10:48 AM
MC_SColombano2015.ipynb	19-Dec-19 4:52 PM
MC_SColombano2017.ipynb	19-Dec-19 4:51 PM
Metric_linearity_check.ipynb	20-Nov-17 6:08 PM
MIA composition overview.ipynb	12-Nov-18 5:42 PM
N2O_PCA.ipynb	11-Sep-18 5:55 PM
N2O_PCA_chap2.ipynb	09-Dec-19 12:58 PM
N2Orisk_Aer.ipynb	23-Apr-19 4:01 PM
N2Orisk_paper_figures.ipynb	23-Apr-19 4:03 PM
OldDICEA_offgas_provaLunga.ipynb	16-May-19 7:20 AM
OldDICEA_offgas_provaPunti.ipynb	12-Oct-17 4:59 PM
pca.pdf	12-Feb-16 3:42 PM
plotLindaData.ipynb	18-Apr-16 3:55 PM
raw4KMO.xlsx	19-Mar-18 4:10 PM
readLiqN2O.py	03-Apr-17 10:47 AM
README.md	02-Feb-17 8:28 AM
RMSE_ofQairData.ipynb	11-Aug-17 4:28 PM
SA_trial.ipynb	03-May-17 6:30 PM
sample size on aSOTE.ipynb	07-Mar-16 5:48 PM

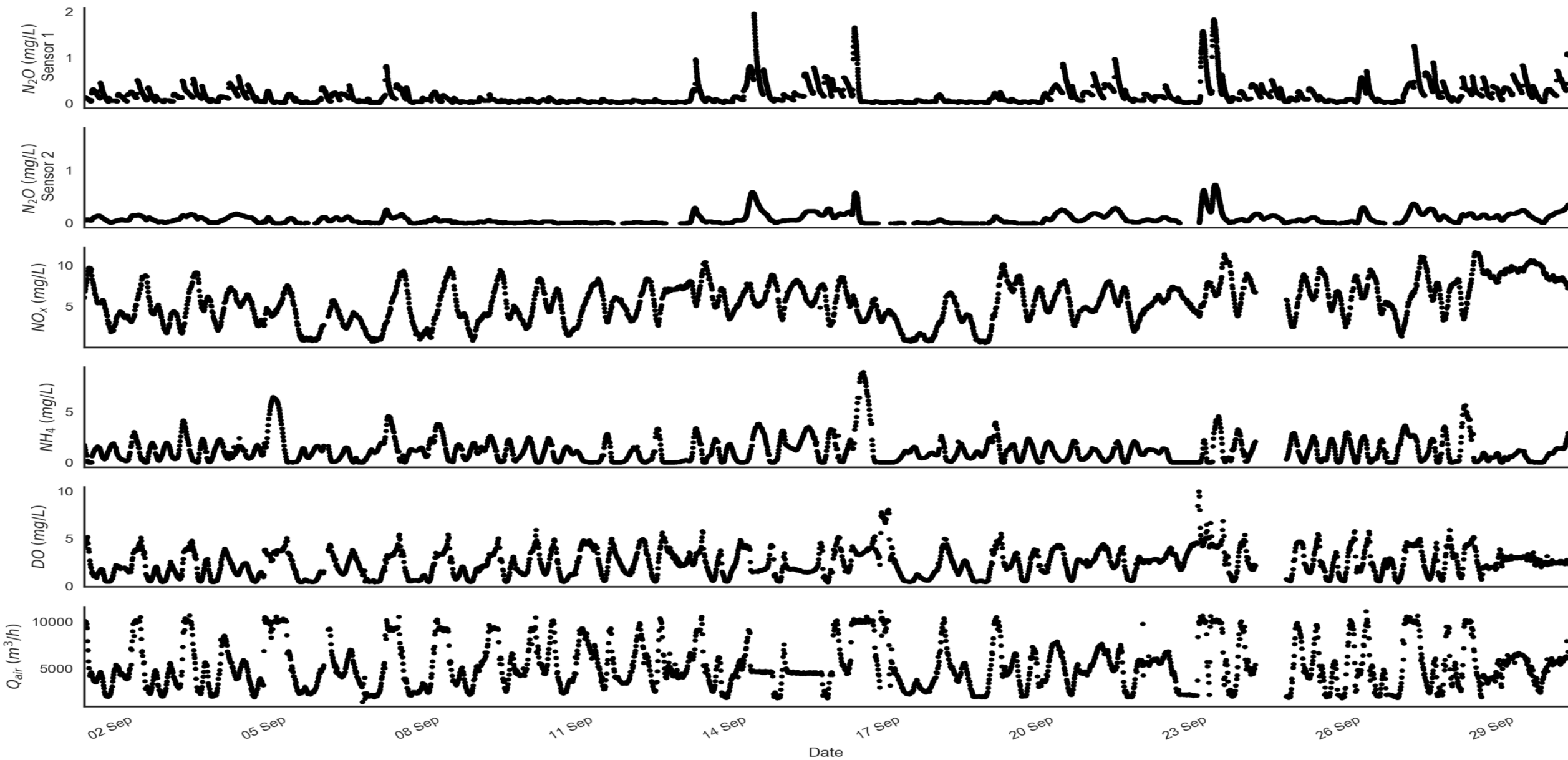
Managing scripts

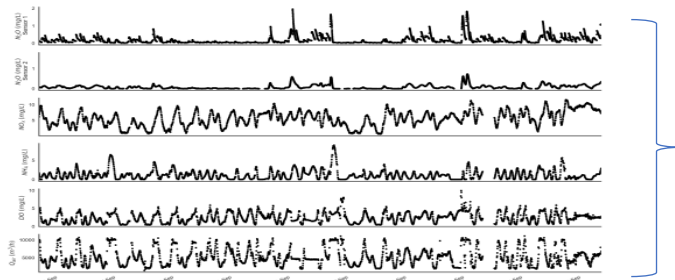


<https://guides.github.com/>

<https://medium.com/mindorks/what-is-git-commit-push-pull-log-aliases-fetch-config-clone-56bc52a3601c>

Some theory on PCA for better
understanding the dedicated notebook





X
 $n \times p$

Centering

$$\bar{X} = X - \frac{1}{n} \sum_{i=1}^n X[i, j]$$



$p \times p$
covariance matrix

$$S = \text{cov}(\bar{X}) = \frac{\bar{X}^T \cdot \bar{X}}{n - 1}$$



Eigen decomposition

$$V^{-1} \cdot S \cdot V = D$$



*Rearrange
eigenvectors and
eigenvalues*

