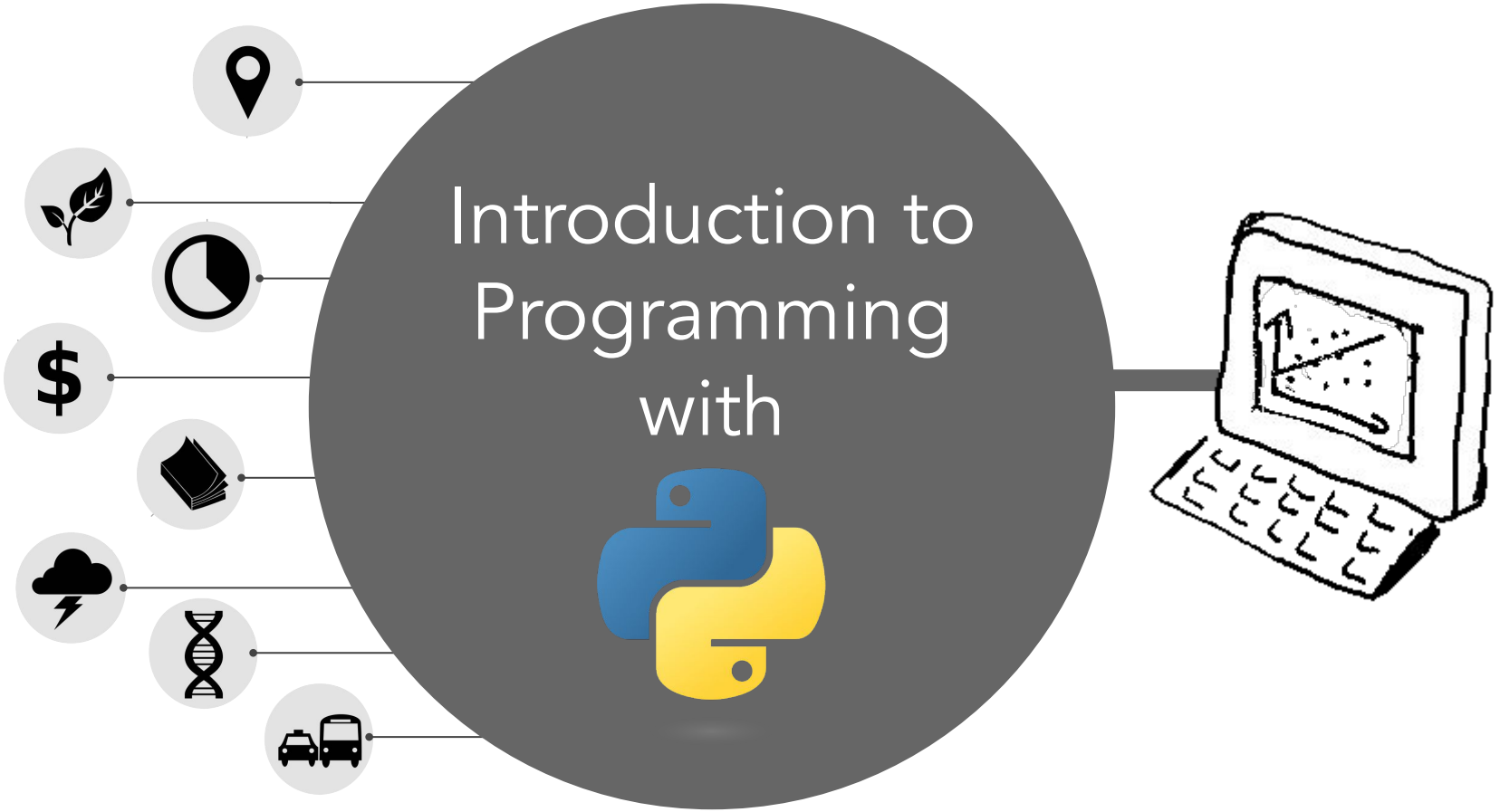


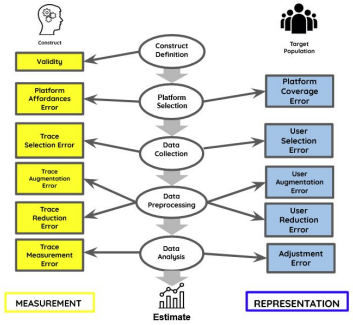
Indira Sen, Ana Sanchez Acosta
- eKOMEX -
Konstanz, 22.02.2024

Indira Sen: In a nutshell

Junior Faculty@Uni Mannheim [CSS + NLP]

Before:

- Postdoc@Uni-Konstanz
- Ph.D Student in Computational Social Science@GESIS, RWTH-Aachen



TED-On: A Total Error Framework for Digital Traces of Human

Something I worked on: A framework for diagnosing errors in studies using digital traces based on measurement theory and surveys

Areas of Interest

- Measuring social phenomena or 'constructs' (attitudes and behaviors) from digital traces (tweets, reddit posts, Wikipedia) with computational methods
- Measuring and Mitigating Harmful Communication Online (hate speech, harassment)
- Designing computational methods that are:
  - Grounded in theory
  - Robust and generalizable

If you're curious about my work, here are the slides from a recent talk

And you?

tell us about yourself in a few sentences

what makes you interested in Python / programming?

“Big Data” ?

= Big in chances and challenges

= “Found/Organic data”: observational, non-reactive, non-probabilistic samples, data-generating process unclear (not “designed” by a researcher)

= Not in one single place; but openly accessible

= A lot of unstructured content (often text) + metadata

= Big enough to train Machine Learning models on → Data Science

Scientific Programming

- Strong emphasis on analysis of data (vs. development of systems)
  - Not about building software
  - Result-oriented, not performance-driven
  - Using different libraries suited for the Data Science purpose
  - we will only lightly touch on object oriented programming

Why  python™ ?

- Well-established
- Easy-to-read, easy-to-learn
- Cross-platform
- Broad, multi-purpose language (e.g. Web apps)
- Interactive programming

Course goals

- Get an idea of the “Data Science Workflow”
- Enable you to apply Python to solve your problem
- Give an overview on some important libraries for academia
- Gives you basic background on:
  - Foundational programming concepts
  - Data Visualization
- Not: standard QuantSoc statistical procedures in Python

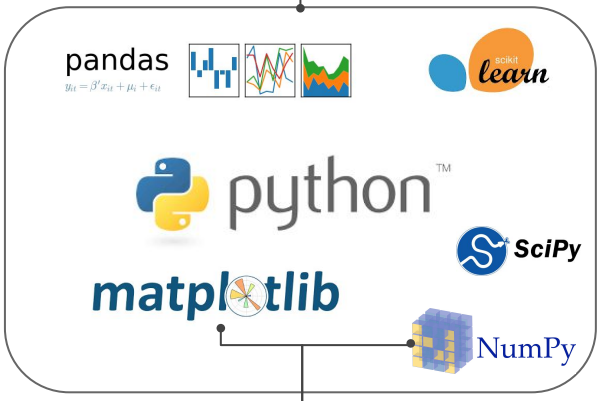
So after this course, you should be able …

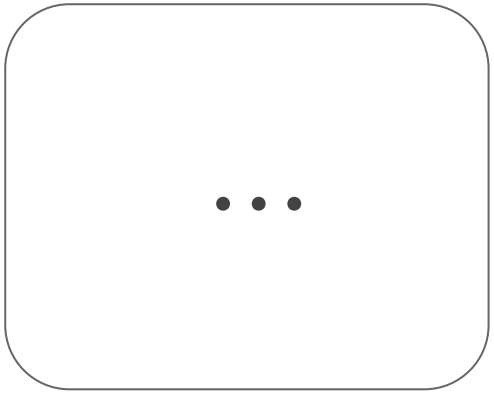
- … learn the basic building blocks of analysis with Python
- … to produce some visualizations for your data
- … use Jupyter Notebooks proficiently

Our stack


jupyter

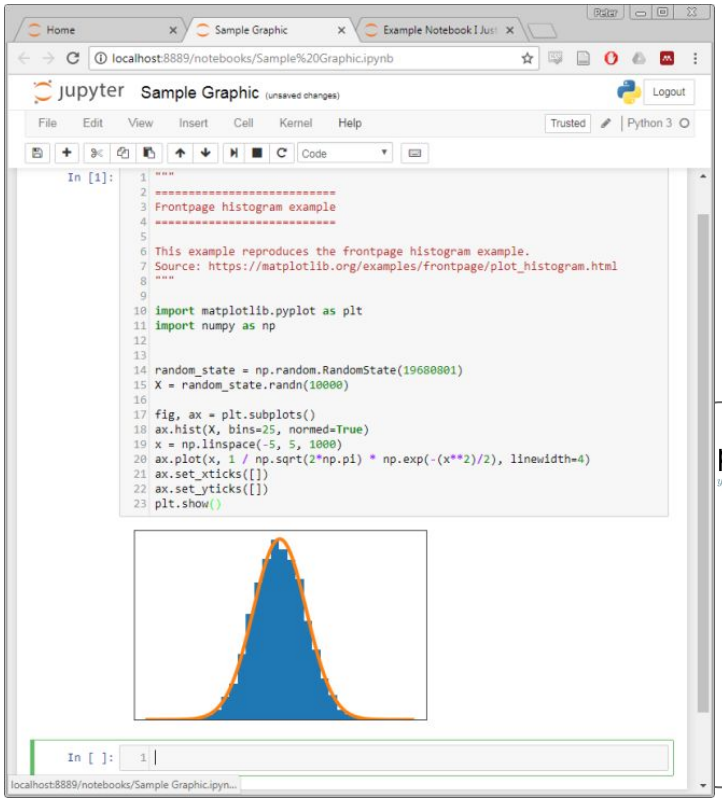

spyder

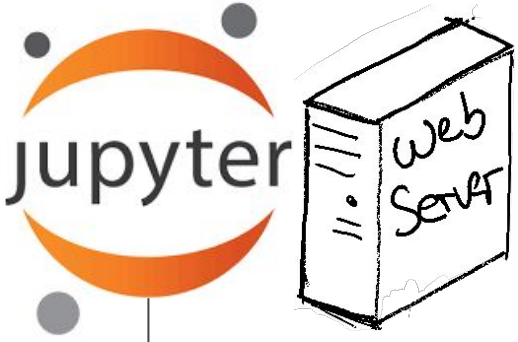

pandas
scikit learn
python
SciPy
matplotlib
NumPy


• • •

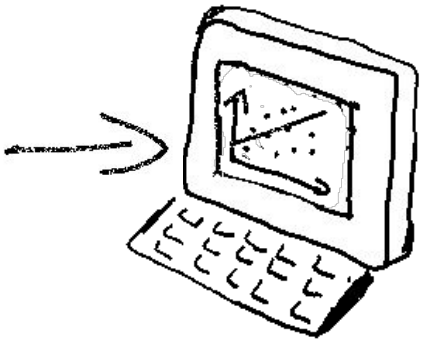

ANACONDA

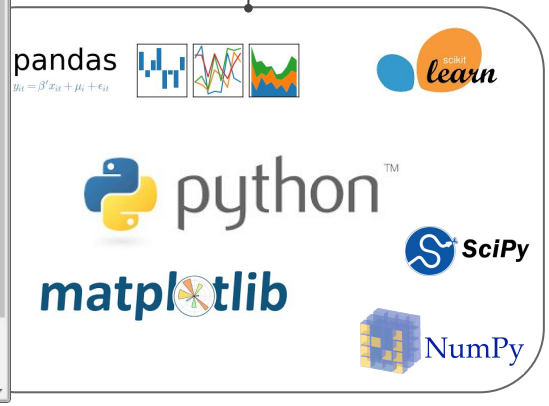
Our stack


Home
Sample Graphic
Example Notebook I Just
localhost:8889/notebooks/Sample%20Graphic.ipynb
jupyter Sample Graphic (unsaved changes)
Logout
File Edit View Insert Cell Kernel Help
Trusted
Python 3
Code
In [1]:
"""
Frontpage histogram example
This example reproduces the frontpage histogram example.
Source: https://matplotlib.org/examples/frontpage/plot_histogram.html
"""
import matplotlib.pyplot as plt
import numpy as np
random_state = np.random.RandomState(19680801)
X = random_state.randn(10000)
fig, ax = plt.subplots()
ax.hist(X, bins=25, normed=True)
x = np.linspace(-5, 5, 1000)
ax.plot(x, 1 / np.sqrt(2*np.pi) * np.exp(-(x**2)/2), linewidth=4)
ax.set_xticks([])
ax.set_yticks([])
plt.show()
In [ ]:
localhost:8889/notebooks/Sample Graphic.ipyn...


jupyter

web
Server




pandas
scikit learn
python
SciPy
matplotlib
NumPy

- Current Python version: Python 3.12.0, Python 2.7.14 (considered legacy now)
- Python 3
  - Released 2008
  - backward-incompatible (at least not 100%)
  - For several years some libraries did not transition
- Nowadays, one should use Python 3 (unless…)

Logistics

- lectures and in-class exercises 9.00-12:00, 13:30-15:30
- “assignments” + office hours: 15:40-16:45
  - feel free to troubleshoot installation issues
  - other questions you might have
- solutions of the assignment will be uploaded after 16:45

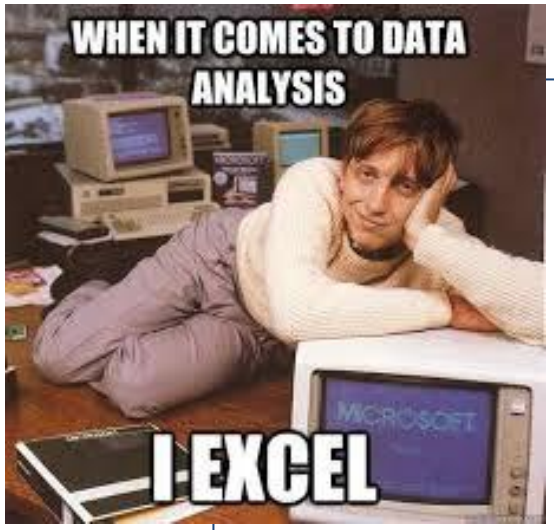
Material & Schedule

https://tinyurl.com/pythonrptu

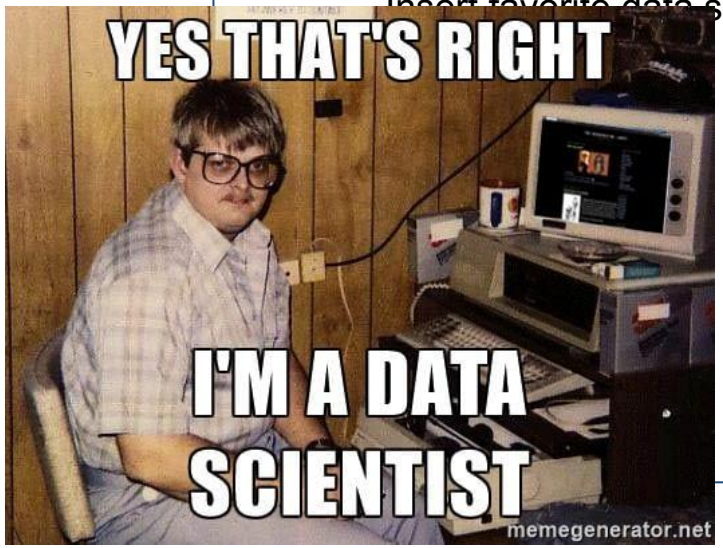
- Who had issues with Jupyter Notebooks?
  - if yes, use google colab for now
- Who hasn't been able to access Moodle?
- Who hasn't been able to access Github?

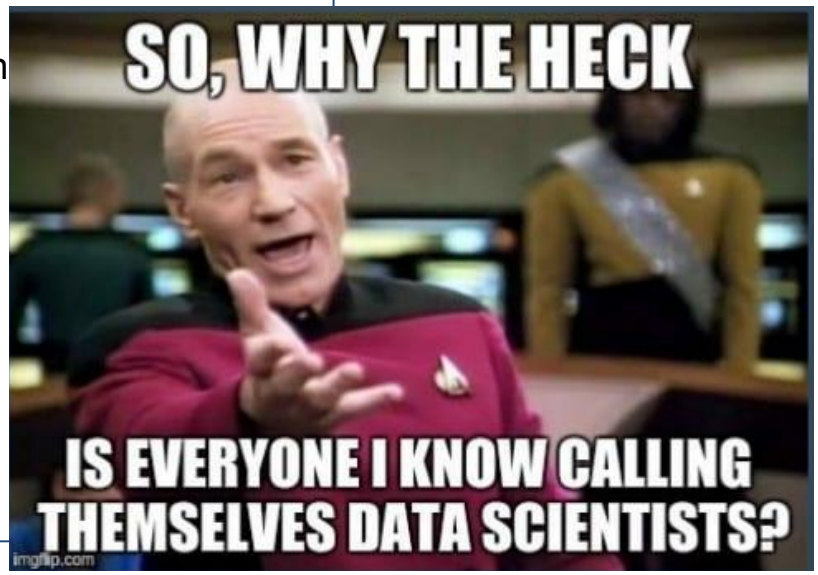
and now...

and now...

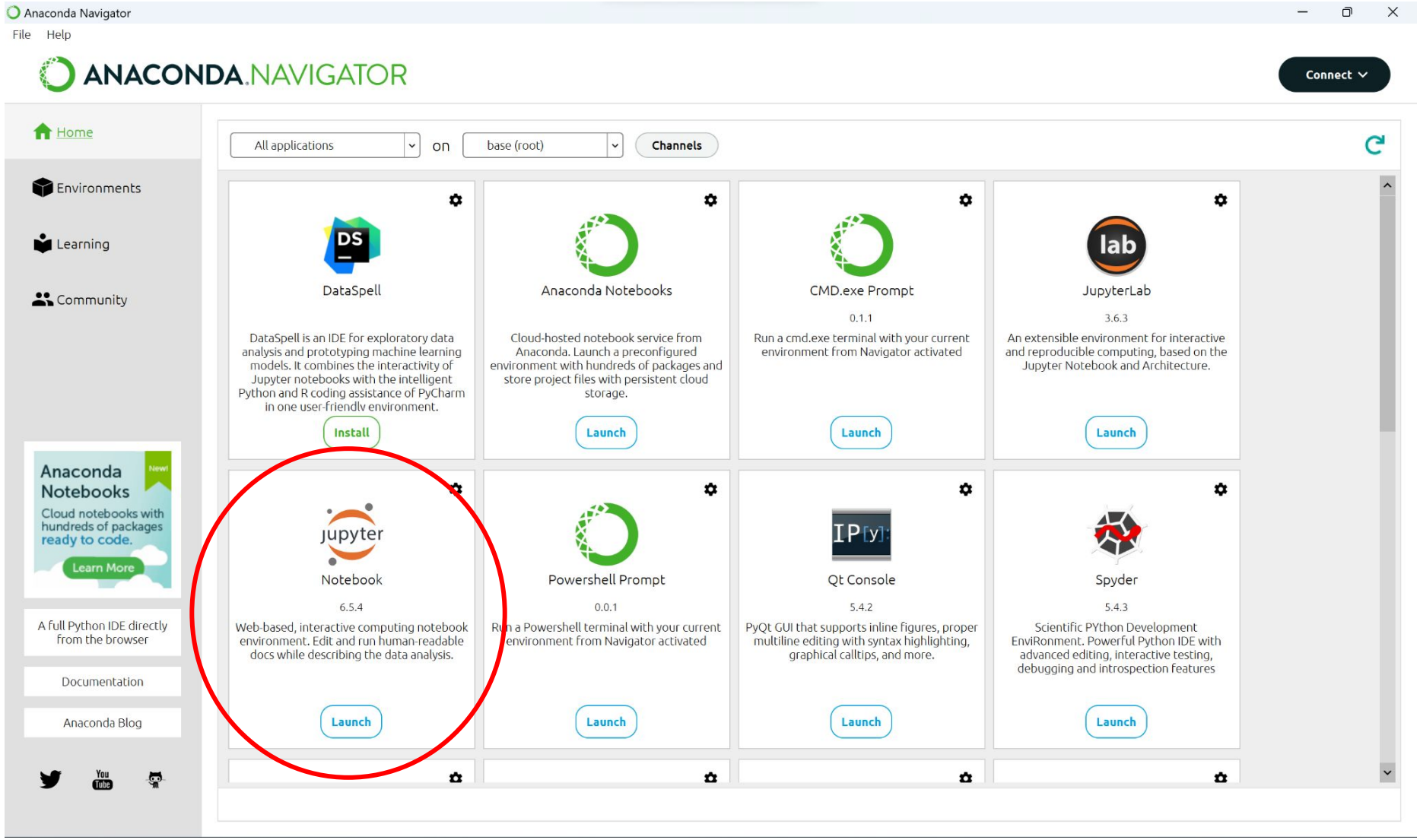


Insert favorite data scien





backup


Anaconda Navigator
File Help
ANACONDA.NAVIGATOR
Connect
Home
Environments
Learning
Community
All applications
on
base (root)
Channels
DataSpell
DataSpell is an IDE for exploratory data analysis and prototyping machine learning models. It combines the interactivity of Jupyter notebooks with the intelligent Python and R coding assistance of PyCharm in one user-friendly environment.
Install
Anaconda Notebooks
Cloud-hosted notebook service from Anaconda. Launch a preconfigured environment with hundreds of packages and store project files with persistent cloud storage.
Launch
CMD.exe Prompt
0.1.1
Run a cmd.exe terminal with your current environment from Navigator activated
Launch
JupyterLab
3.6.3
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.
Launch
Anaconda Notebooks
Cloud notebooks with hundreds of packages ready to code.
New!
Learn More
Notebook
6.5.4
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
Launch
Powershell Prompt
0.0.1
Run a Powershell terminal with your current environment from Navigator activated
Launch
Qt Console
5.4.2
PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.
Launch
Spyder
5.4.3
Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features
Launch
A full Python IDE directly from the browser
Documentation
Anaconda Blog


jupyter
Quit
Logout
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
0
/ OneDrive / Desktop / python_block_course
Name
Last Modified
File size
..
seconds ago
The notebook list is empty.
jupyter
Quit
Logout
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
0
/ OneDrive / Desktop / python_block_course
Name
Last Modified
File size
The notebook list is empty.
..
seconds ago
0_prep.ipynb
Upload
Cancel