Universität Zürich[UZH]

Soziologisches Institut

Data Analysis – Advanced Statistics with Python

Dr. Julia Jerke

jerke@soziologie.uzh.ch

Thursday, 12.15pm – 13.45pm, AND 2.46

Session 1 – Introduction to Python

Agenda

1. Covid certificate details
2. Introduction round
3. Course overview
4. Why Python?
5. Python installation
6. Getting to know Python and Spyder

1. Covid certificate Details

UZH regulations

- Students are required to have **Covid certificates** for on-site courses at all levels (Bachelor's, Master's, PhD) and further education programs
- Vaccination **certificates from abroad** will also be recognized if the vaccine used is approved in the EU or listed in the WHO's Emergency Use Listing
- UZH will offer **free tests** (until the end of October) to students and staff who do not have a Covid certificate. Testing is currently available at UZH in the test center on **Irchel Campus** and in the **Corona Center on Hirschengraben**
- Teaching staff are not responsible for carrying out spot checks, but **may do so** if they believe it is necessary

Specific guidelines for that course

- Respect the UZH regulations
- **We probably all know vulnerable people that we want to protect!**
- Specific rules for the course:
  - Always bring your Covid certificate to the session
  - Participants that do not have a certificate, are required to leave the course
  - If you have symptoms, stay at home (fever, coughing, shortness of breath, etc.)
  - If possible, keep distance
  - Please wear a mask for additional protection
  - While speaking, you may take the mask off (this particularly implies that the lecturer is usually not wearing a mask)

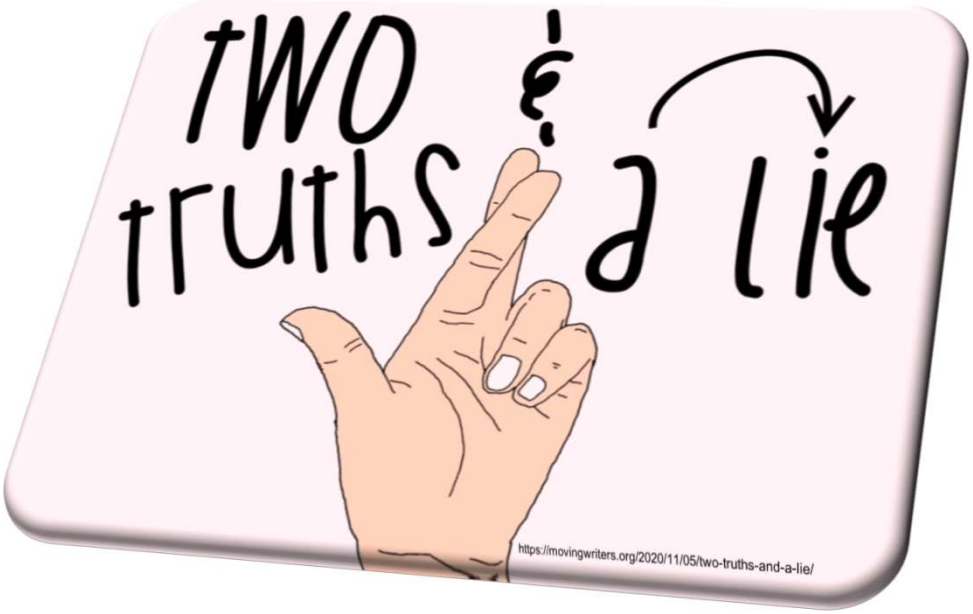
2. Introduction round

Two Truths and a Lie

1. For each one of you: think of two statements that are true about you and come up with one lie

   Choose statements that you are comfortable sharing, because we may need to talk about some of them ;-)

2. Share your two truths and the lie in the plenum without telling which is which
3. Let's find the lie and get to know each other!



My two truths and one lie

1. I already did Zorbing.

2. I have already broken my hand more than two times.

3. I used to be able to do a flip.

Go to **www.menti.com** and use the code **6767 8466**

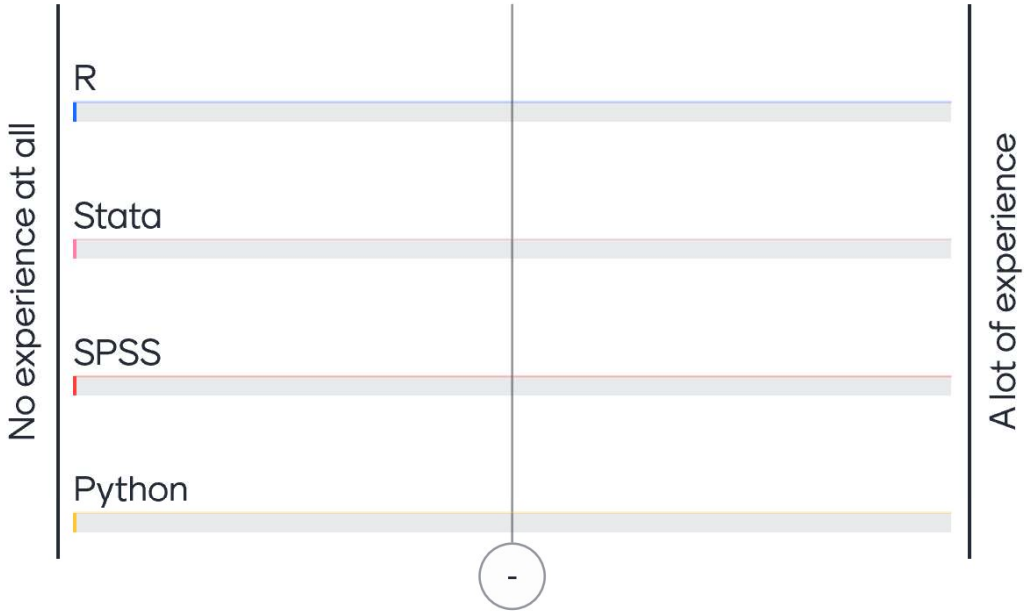
Let me know what you are studying!

Press S to show image

Go to **www.menti.com** and use the code **6767 8466**

How much experience do you have with the following statistical programs/languages?

Mentimeter



Press S to show image

3. Course overview

Course objectives

- Refreshment of basic statistical knowledge
- Get an overview (!) of advanced data analytical methods
  - We will introduce various specific methods and their practical applications
  - Some of them we will practice in more depth
  - However, we can not always go into full detail
  - *But*: In the end, you should be able to identify the optimal method given a specific data setting
- Introduction to statistics with Python
- Getting to know the Python workflow (e.g., working with *Spyder* and *Jupyter Notebook*)
- Learn how to properly backup your code (e.g., Git and Github)

Program

Date	Session
23.09.2021	Introduction to Python
30.09.2021	Getting and exploring data
07.10.2021	Linear regression
14.10.2021	Logistic regression
21.10.2021	Advanced regression methods
28.10.2021	Data visualization I
04.11.2021	*Special topic:* Git and Github
11.11.2021	Cluster analysis
18.11.2021	Principal component analysis (PCA)
25.11.2021	Data visualization II
02.12.2021	*Special topic*: Introduction to machine learning
09.12.2021	*Special topic:* Natural language processing
16.12.2021	*Special topic:* Databases and SQL

OLAT

- You will find all material of the course online at OLAT
- This includes:
    - Syllabus
    - Slides
    - Exercise sheets
    - Further material such as helpful texts, cheat sheets, example scripts

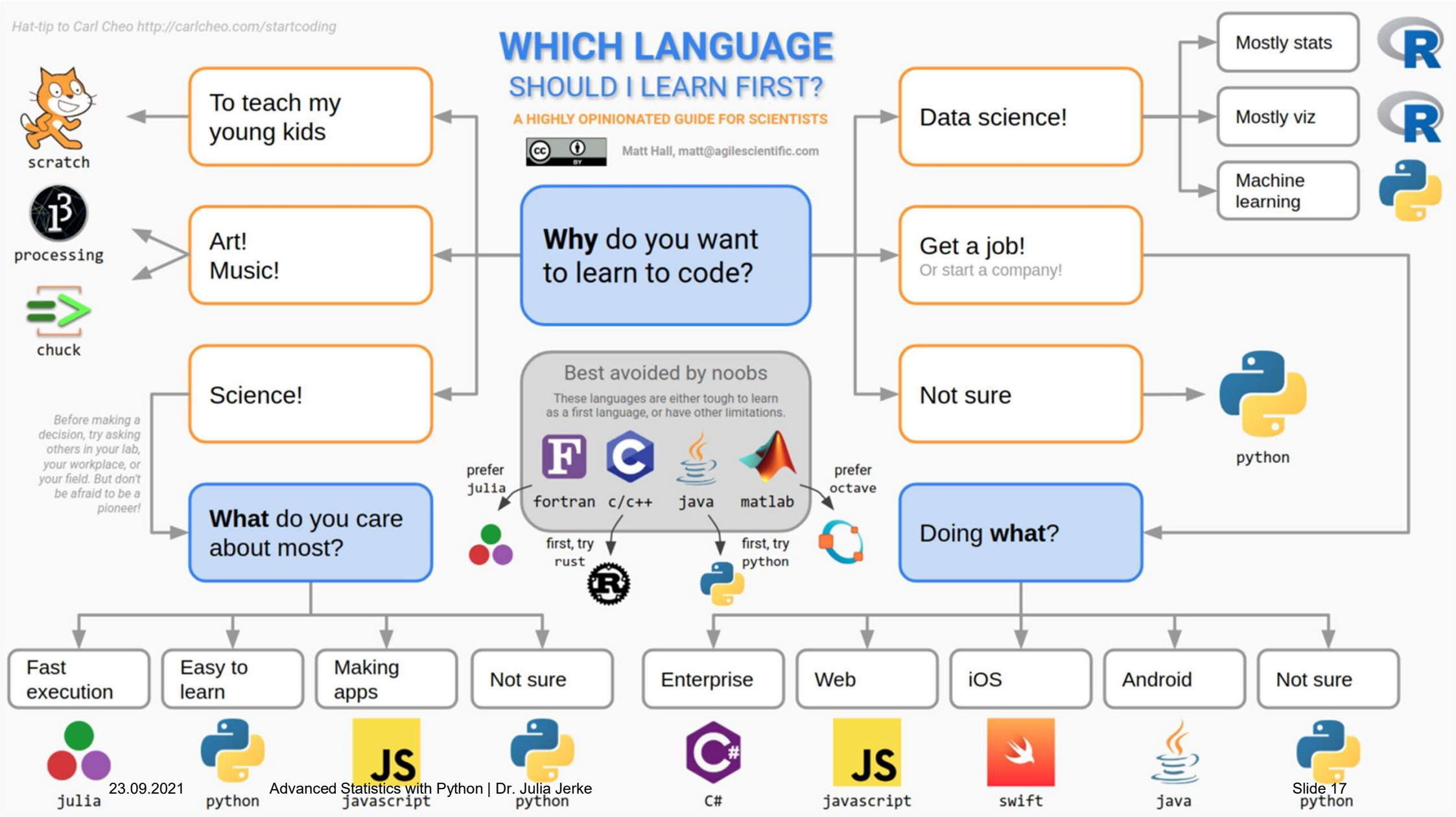
Examination

- Upon successful completion of this course, **6 ECTS** are credited
- Students are required to **actively contribute** to the sessions
- Throughout the course, **8 exercise sheets** will be distributed that cover the topics of the sessions, to pass the course you need to meet the following two requirements:

  (1) Successfully **complete 6 out of the 8** exercise sheets (> 60% of the points)

  (2) Obtain at least **60% of the total score** across all exercise sheets
- The exercise sheets must be submitted electronically 7 days after publication

4. Why Python?

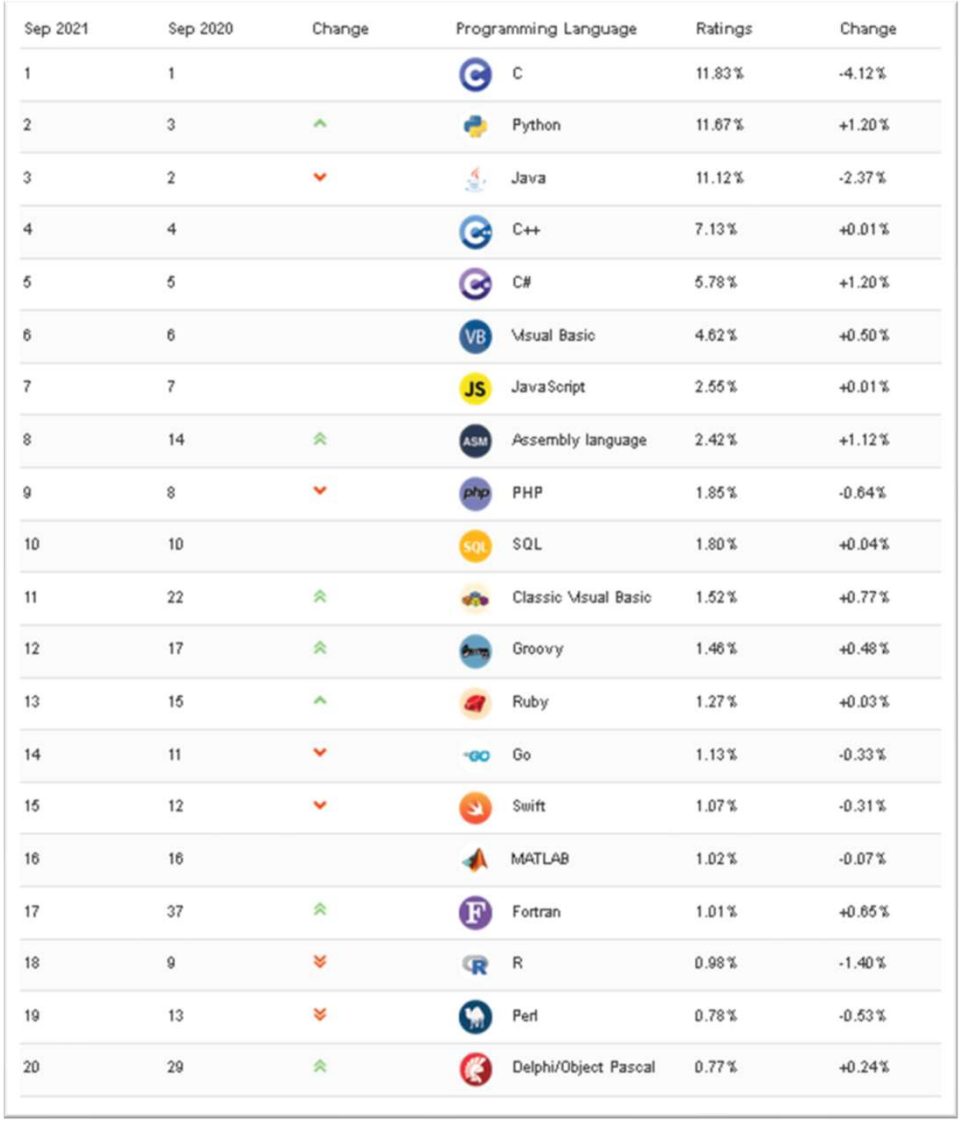

Hat-tip to Carl Cheo http://carlcheo.com/startcoding
WHICH LANGUAGE
SHOULD I LEARN FIRST?
A HIGHLY OPINIONATED GUIDE FOR SCIENTISTS
Matt Hall, matt@agilescientific.com
Why do you want to learn to code?
To teach my young kids
scratch
Art! Music!
processing
chuck
Science!
What do you care about most?
Before making a decision, try asking others in your lab, your workplace, or your field. But don't be afraid to be a pioneer!
Data science!
Mostly stats
Mostly viz
Machine learning
Get a job!
Or start a company!
Not sure
python
Doing what?
Best avoided by noobs
These languages are either tough to learn as a first language, or have other limitations.
fortran
c/c++
java
matlab
prefer julia
prefer octave
first, try rust
first, try python
Fast execution
Easy to learn
Making apps
Not sure
julia
python
javascript
python
Enterprise
Web
iOS
Android
Not sure
C#
javascript
swift
java
python

The strengths of Python

- Easy to learn, it has a very steep learning curve
- Easy to read: very simple and not complicated syntax
- Multi-purpose: you can use Python for mostly anything
  - Statistics (basic statistics, but also advanced stuff such as machine learning)
  - Scientific computing
  - Web development
  - Software development
  - Robotics
  - …
- Particularly powerful for big data and machine learning
- Huge community, plenty of guides, documentations and video tutorials (high chance that every problem you may encounter, will already have been solved and documented by someone)
- Very large archive of packages
- Free and open-source

Python vs. R

- R and Python share many advantages (open source, easy to learn and read, large community, hundreds of helpful packages)
- According to the TIOBE index, Python is the second most popular programming language, R is on position 18
- However, R is specifically developed for statistical modelling and computation and, therefore, still is the primary tool!
- Python, as a versatile and easy to learn language, is a perfect complement to R
- In particular, Python should be preferred over R in the context of big data and machine learning



Sep 2021	Sep 2020	Change	Programming Language	Ratings	Change
1	1		C	11.83%	-4.12%
2	3	↑	Python	11.67%	+1.20%
3	2	↓	Java	11.12%	-2.37%
4	4		C++	7.13%	+0.01%
5	5		C#	5.78%	+1.20%
6	6		Visual Basic	4.62%	+0.50%
7	7		JavaScript	2.55%	+0.01%
8	14	⇈	Assembly language	2.42%	+1.12%
9	8	↓	PHP	1.85%	-0.64%
10	10		SQL	1.80%	+0.04%
11	22	⇈	Classic Visual Basic	1.52%	+0.77%
12	17	⇈	Groovy	1.46%	+0.48%
13	15	↑	Ruby	1.27%	+0.03%
14	11	↓	Go	1.13%	-0.33%
15	12	↓	Swift	1.07%	-0.31%
16	16		MATLAB	1.02%	-0.07%
17	37	⇈	Fortran	1.01%	+0.65%
18	9	⇊	R	0.98%	-1.40%
19	13	⇊	Perl	0.78%	-0.53%
20	29	⇈	Delphi/Object Pascal	0.77%	+0.24%

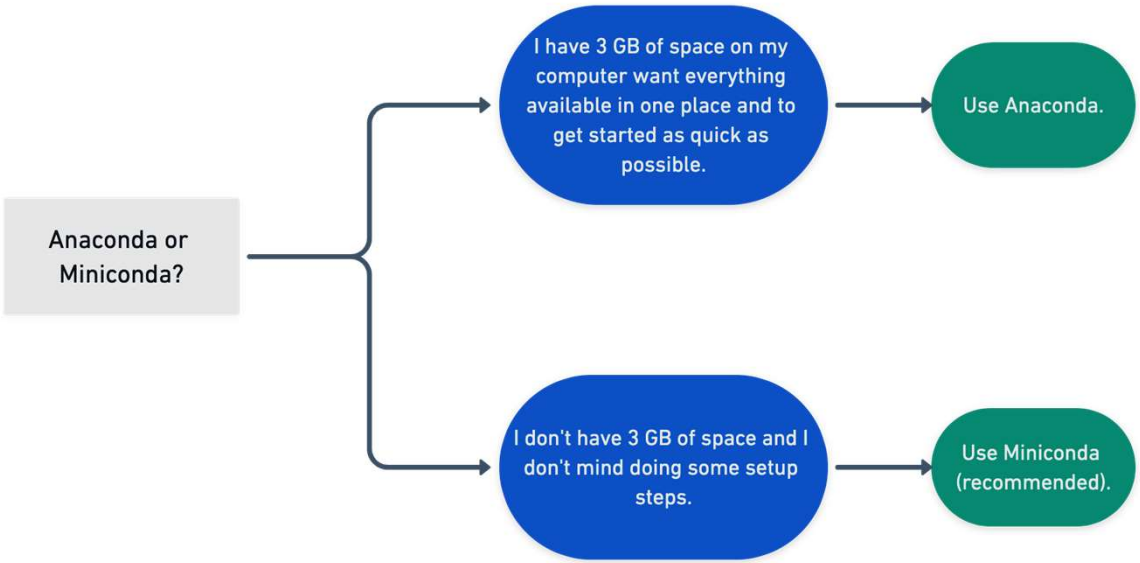
5. Python installation

Various ways to install Python

Solo Python installation	Installation of a distribution
• Separate installation of the Python interpreter • Optional: installation of an editor to write comprehensive scripts or an IDE (integrated development environment)	• Installation of a collection of software components to work with Python (interpreter, package manager, IDE, ect.) • The most common distributions are Anaconda or Miniconda

Anaconda or Miniconda

- *Anaconda* is particularly beginner friendly and convenient
- The major difference between the two distributions is that with *Anaconda* a multitude of packages will already be installed
- With *Miniconda* these packages have to be installed later when needed with the package manager *Conda*
- However, *Anaconda* takes longer to install and requires more space

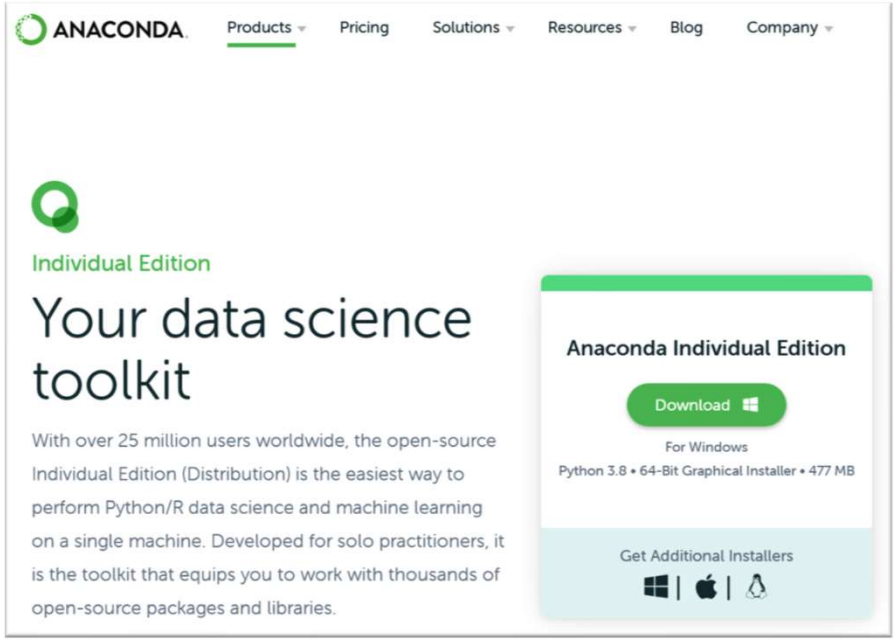


Installation steps with Anaconda

1. **Go to the *Anaconda* website and scroll to the bottom:**

   https://www.anaconda.com/products/individual

2. Choose the appropriate installer and download it
3. Open the installer to install *Anaconda* and *Python*
4. While installing, do not change the default settings
5. Open the *Anaconda Navigator* (either with the Finder for Mac or the Start Menu for Windows)
6. Click to launch *Spyder*



If you prefer to install *Miniconda*, the steps above remain the same.
Download the installer here:

https://docs.conda.io/en/latest/miniconda.html

Installation steps with Anaconda

1. Go to the *Anaconda* website and scroll to the bottom:

   https://www.anaconda.com/products/individual

2. **Choose the appropriate installer and download it**
3. Open the installer to install *Anaconda* and *Python*
4. While installing, do not change the default settings
5. Open the *Anaconda Navigator* (either with the Finder for Mac or the Start Menu for Windows)
6. Click to launch *Spyder*

If you prefer to install *Miniconda*, the steps above remain the same.
Download the installer here:

https://docs.conda.io/en/latest/miniconda.html

Installation steps with Anaconda

1. Go to the *Anaconda* website and scroll to the bottom:

   https://www.anaconda.com/products/individual

2. Choose the appropriate installer and download it
3. **Open the installer to install *Anaconda* and *Python***
4. While installing, do not change the default settings
5. Open the *Anaconda Navigator* (either with the Finder for Mac or the Start Menu for Windows)
6. Click to launch *Spyder*

If you prefer to install *Miniconda*, the steps above remain the same.
Download the installer here:

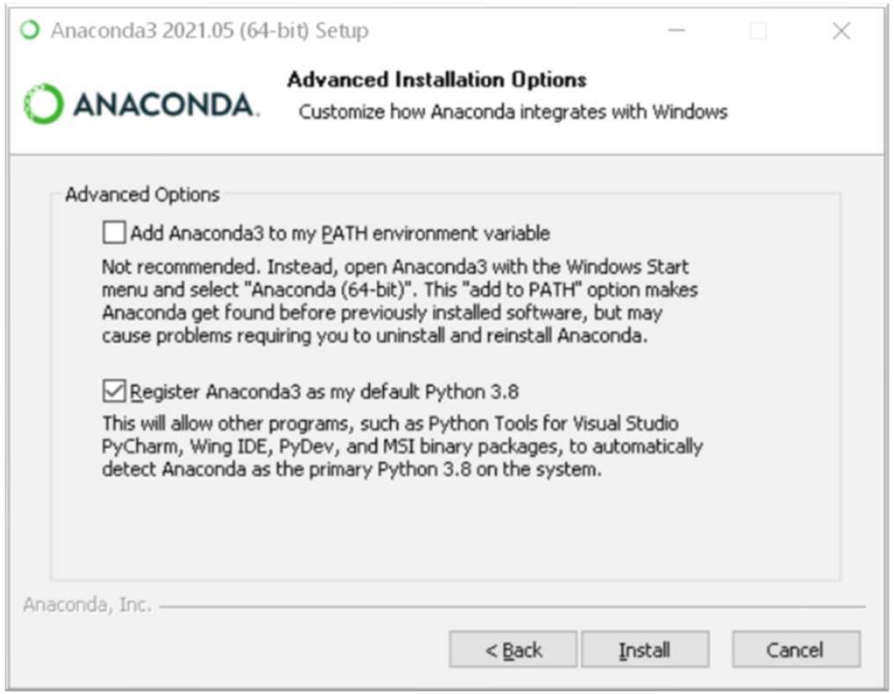
https://docs.conda.io/en/latest/miniconda.html

Installation steps with Anaconda

1. Go to the *Anaconda* website and scroll to the bottom:

   https://www.anaconda.com/products/individual

2. Choose the appropriate installer and download it
3. Open the installer to install *Anaconda* and *Python*
4. **While installing, do not change the default settings**
5. Open the *Anaconda Navigator* (either with the Finder for Mac or the Start Menu for Windows)
6. Click to launch *Spyder*

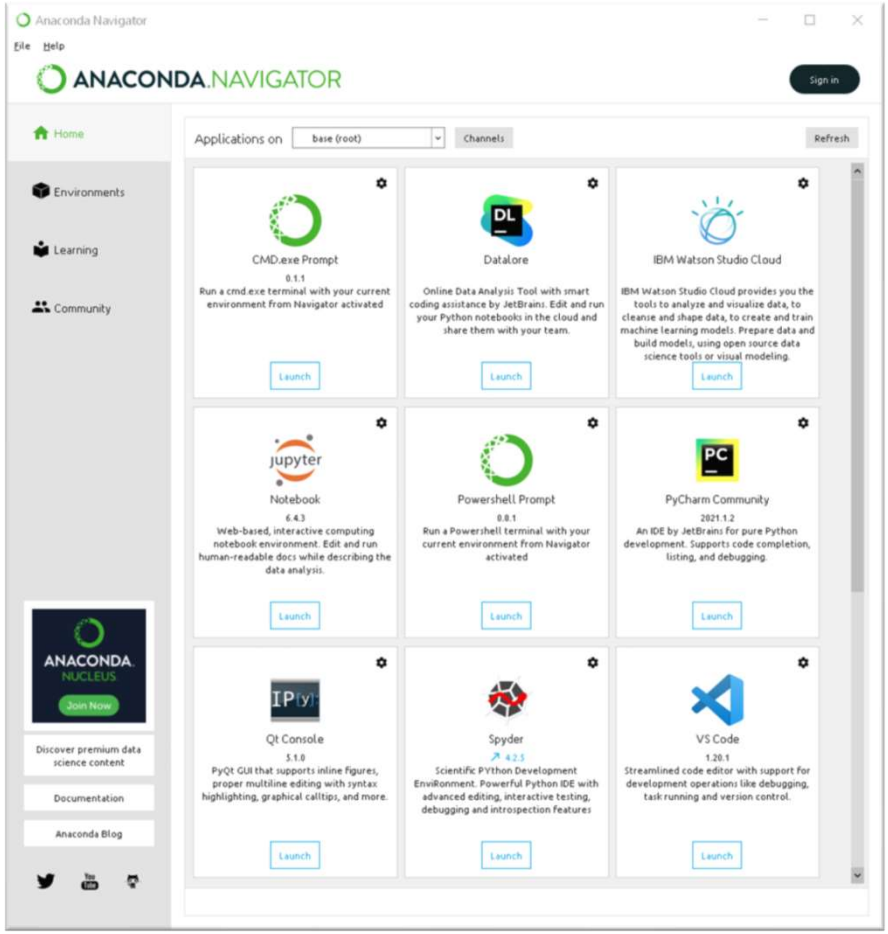


If you prefer to install *Miniconda*, the steps above remain the same.
Download the installer here:

https://docs.conda.io/en/latest/miniconda.html

Installation steps with Anaconda

1. Go to the *Anaconda* website and scroll to the bottom:
   https://www.anaconda.com/products/individual
2. Choose the appropriate installer and download it
3. Open the installer to install *Anaconda* and *Python*
4. While installing, do not change the default settings
5. **Open the *Anaconda Navigator* (either with the Finder for Mac or the Start Menu for Windows)**
6. Click to launch *Spyder*



If you prefer to install *Miniconda*, the steps above remain the same.
Download the installer here:

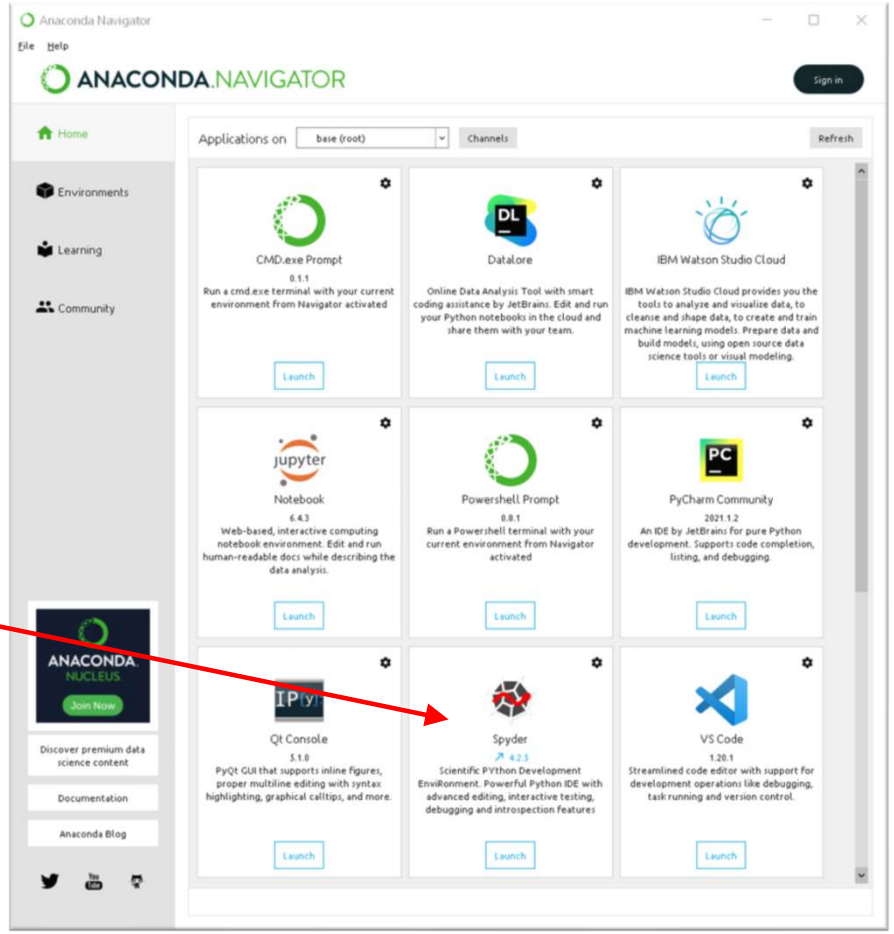
https://docs.conda.io/en/latest/miniconda.html

Installation steps with *Anaconda*

1. Go to the *Anaconda* website and scroll to the bottom:

   https://www.anaconda.com/products/individual

2. Choose the appropriate installer and download it
3. Open the installer to install *Anaconda* and *Python*
4. While installing, do not change the default settings
5. Open the *Anaconda Navigator* (either with the Finder for Mac or the Start Menu for Windows)
6. **Click to launch *Spyder***



If you prefer to install *Miniconda*, the steps above remain the same.
Download the installer here:

https://docs.conda.io/en/latest/miniconda.html

6. Getting to know Python and Spyder