Data Scientifique

Introduction to Python

Instructor:

Yara Abu Awad

What is Data Scientifique?

https://www.datascientifique.ca/

Data Scientifique

Plan for today

- First half:
  - We review some basic concepts and code
- Second half:
  - Hands on practice

Workshop Materials

https://github.com/YaraRAA/python

What is python?

- A programming language
- A programming "language" is really a translation

Human intent --> machine code

Some coding terms

- Program: sequence of instructions to your computer (also referred to as a script)
- Debugging: tracking down and correcting program errors
- Syntax: structure of the program and rules about that structure e.g. grammar

this sentence contains a syntax error. So does this one

- A note about word editors

Scripts

- Remember: a script = a program = a collection of code = instructions to the computer to do x,y & z
- Can be saved in a file with the extension:

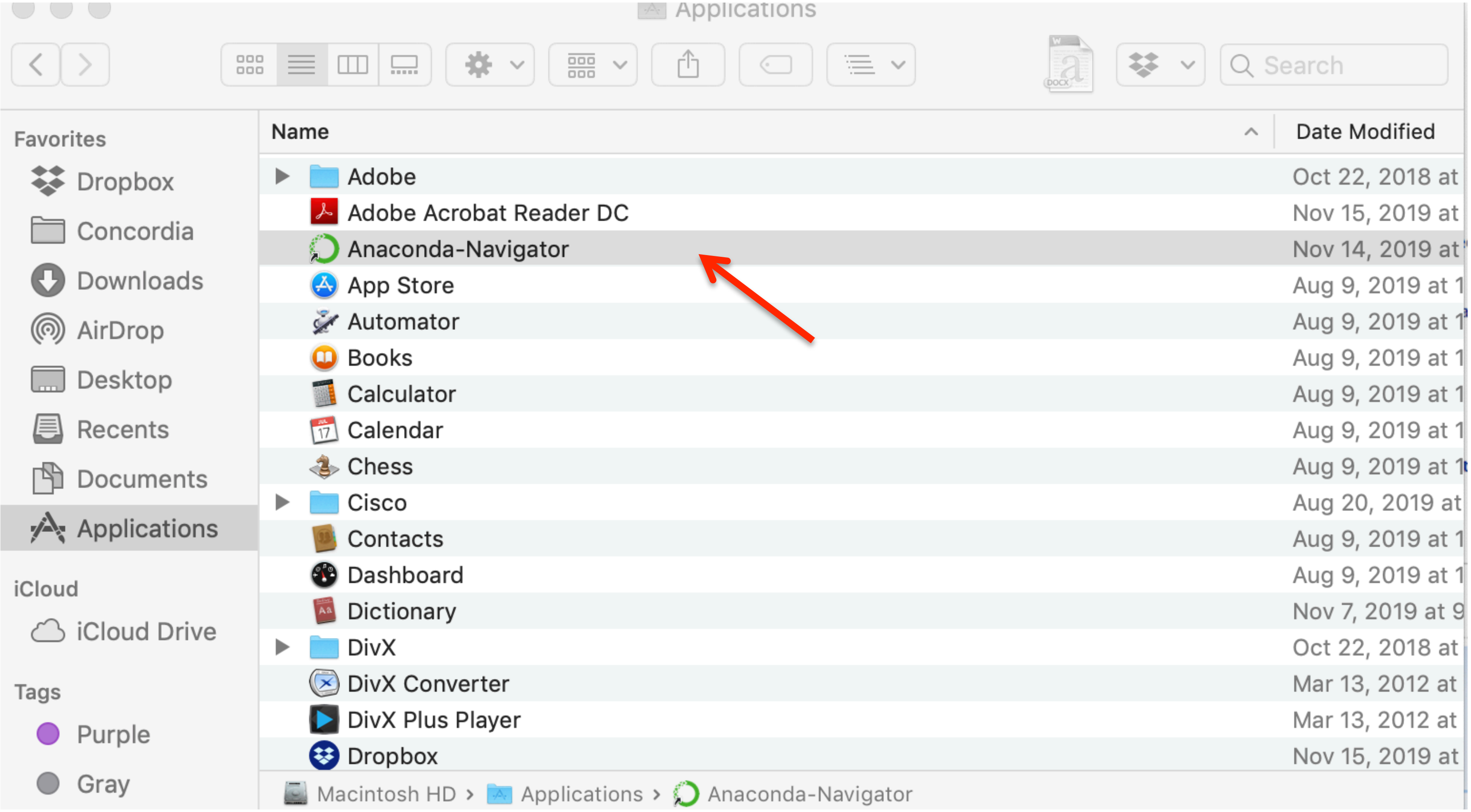
  .py if created in Spyder or Idle or

  .ipynb if created in jupyter notebook

- Let’s create a .py file

Data Scientifique

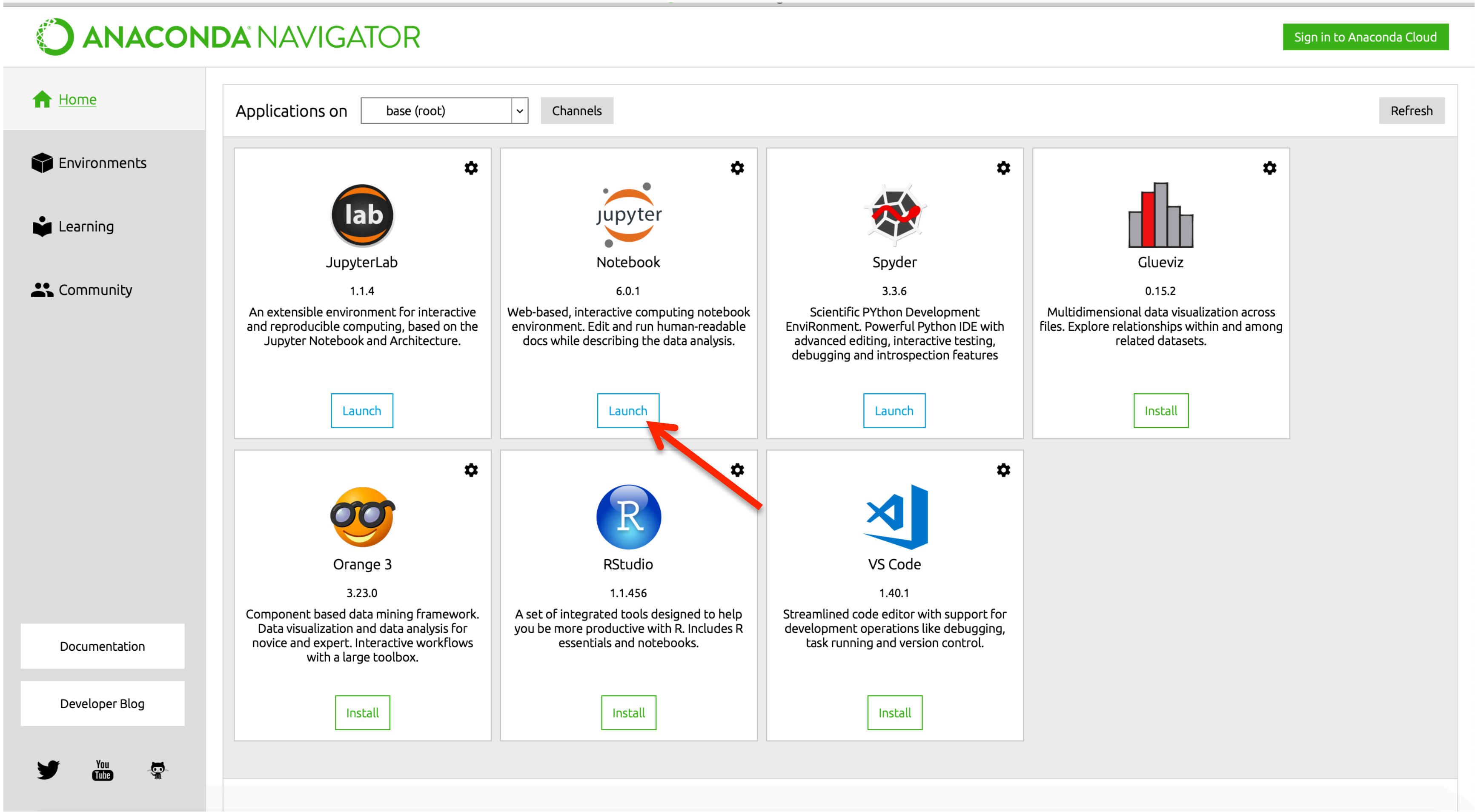
Open Anaconda Navigator



Name	Date Modified
Adobe	Oct 22, 2018 at
Adobe Acrobat Reader DC	Nov 15, 2019 at
Anaconda-Navigator	Nov 14, 2019 at
App Store	Aug 9, 2019 at 1
Automator	Aug 9, 2019 at 1
Books	Aug 9, 2019 at 1
Calculator	Aug 9, 2019 at 1
Calendar	Aug 9, 2019 at 1
Chess	Aug 9, 2019 at 1
Cisco	Aug 20, 2019 at
Contacts	Aug 9, 2019 at 1
Dashboard	Aug 9, 2019 at 1
Dictionary	Nov 7, 2019 at 9
DivX	Oct 22, 2018 at
DivX Converter	Mar 13, 2012 at
DivX Plus Player	Mar 13, 2012 at
Dropbox	Nov 15, 2019 at

Macintosh HD › Applications › Anaconda-Navigator

Launch Spyder



ANACONDA NAVIGATOR

Sign in to Anaconda Cloud

Home

Environments

Learning

Community

Documentation

Developer Blog

Applications on base (root) Channels Refresh

JupyterLab
1.1.4
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.
Launch

Notebook
6.0.1
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
Launch

Spyder
3.3.6
Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features
Launch

Glueviz
0.15.2
Multidimensional data visualization across files. Explore relationships within and among related datasets.
Install

Orange 3
3.23.0
Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.
Install

RStudio
1.1.456
A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.
Install

VS Code
1.40.1
Streamlined code editor with support for development operations like debugging, task running and version control.
Install

The ‘Rules’ of Programming

First Rule of Programming:

First get it right – worry about speed later (or never)

Second Rule of Programming:

Be appropriately lazy & Don’t reinvent the wheel

What is a Module?

- A module (or package) is Python code designed to do something
- Sometimes referred to as “Libraries” in other languages
- Also helps with organizing code related to a particular domain
- Contributes to the “Second Rule of Programming”:
  - Be appropriately lazy & Don’t reinvent the wheel

Installing & Managing Modules

pip

Python Packaging Authority's recommended tool for installing packages from the Python Package Index, PyPI

ANACONDA®

Makes sure there is no conflict when installing modules

Comes preloaded with popular modules

Not just for python

Can create environments

Downloads packages from Anaconda repository and cloud

Using modules

- We use modules by importing them into our code
- Modules can contain both:
  - code (instructions to to something)
  - objects (information stored in some format that we can use)
- We then access code or objects within modules using a period .

(same for a column in a dataframe)

To print the value of pi stored in the math module

Example 1:

```
import math
print(math.pi)
```

Example 2:

```
import math as m
print(m.pi)
```

To save the value of pi stored in the math module in a an object named x

Example 1:

```
import math
print(math.pi)


x = math.pi
x*3
```

Example 2:

```
import math as m
print(m.pi)


x = m.pi
x*3
```

To use the exp function from the math module & store the result

Example 1:

```
x= 3
import math
math.exp(x)


y = math.exp(x)
y * x
```

Example 2:

```
x= 3
import math as m
m.exp(x)


y = m.exp(x)
y * x
```

Data Scientifique

Module help

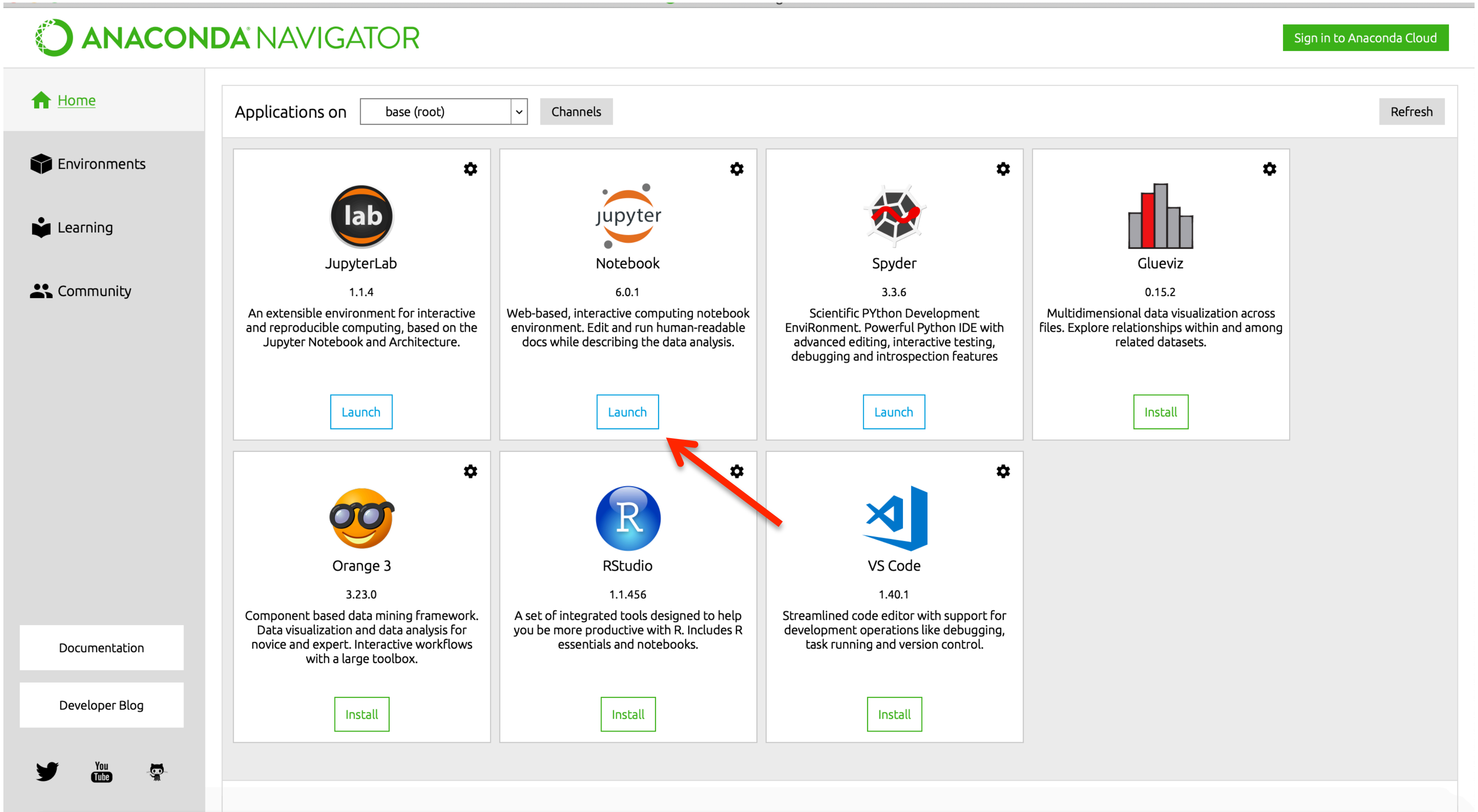
- Use code:

  ```
  help(module_name)
  ```

- OR Google:

  python math module

Now let’s look at Jupyter notebook



JupyterLab

1.1.4

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Notebook

6.0.1

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Spyder

3.3.6

Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Glueviz

0.15.2

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Orange 3

3.23.0

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.

RStudio

1.1.456

A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.

VS Code

1.40.1

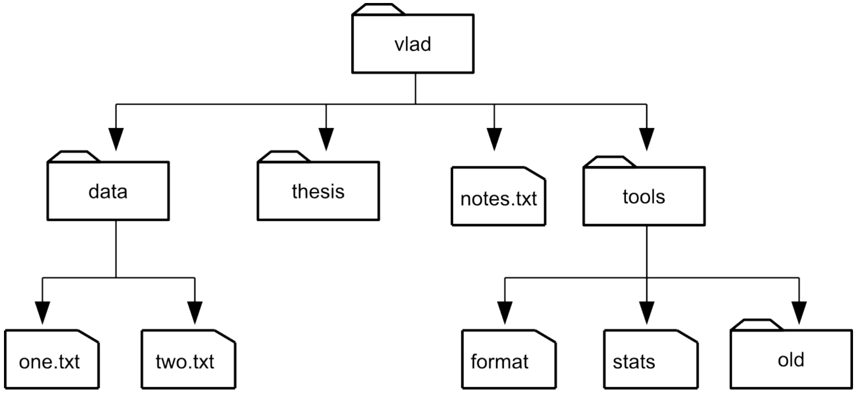
Streamlined code editor with support for development operations like debugging, task running and version control.

What is a directory?

- A directory is a folder
- A path is the location of your folder
- Windows:

  C:\Users\vlad\notes.txt
- Mac:

  /Users/vlad/notes.txt



In python a **working** directory is:
- the folder that all files are imported from or saved to, unless otherwise specified
- by default, this is the folder your script is saved in

Data Scientifique

Image source: http://swcarpentry.github.io