Installation Guide for R and Python

Manuel Lleonart & Luis Ignacio Menéndez García

June 28, 2023

1 Anaconda:

Anaconda is an specific software designed to use python applied for data science. Within Anaconda you have Conda, a way to manage your python packages and organize projects. Anaconda will handle python in your computer and already gives you the default packages needed to do the basics. One of its main advantages is its ability to store results in memory holding your python session opened. This allows you, as we will see in a second, to explore and modigy your code interactively. Last but not least, environments are a fundamental tool for python programming. We will cover them in detail during the brush up course.

1.1 Installation

You can follow the steps in their official guide.

We've seen that Anaconda operates via browser. But where can you actually type your code and see results? You use what is called an IDE. An Integrated development environment is nothing but a software that helps you writing your code. IDEs are typically customizable and allow the programmer to code in different languages, organise projects etc. There are many IDEs available: Pycharm, Spyder, Jupyter Lab... Some programs, such as R or MATLAB, come with their predefined one. We will now see how to install the one that has become most popular recently: VSCode.

2 VSCode installtion guide:

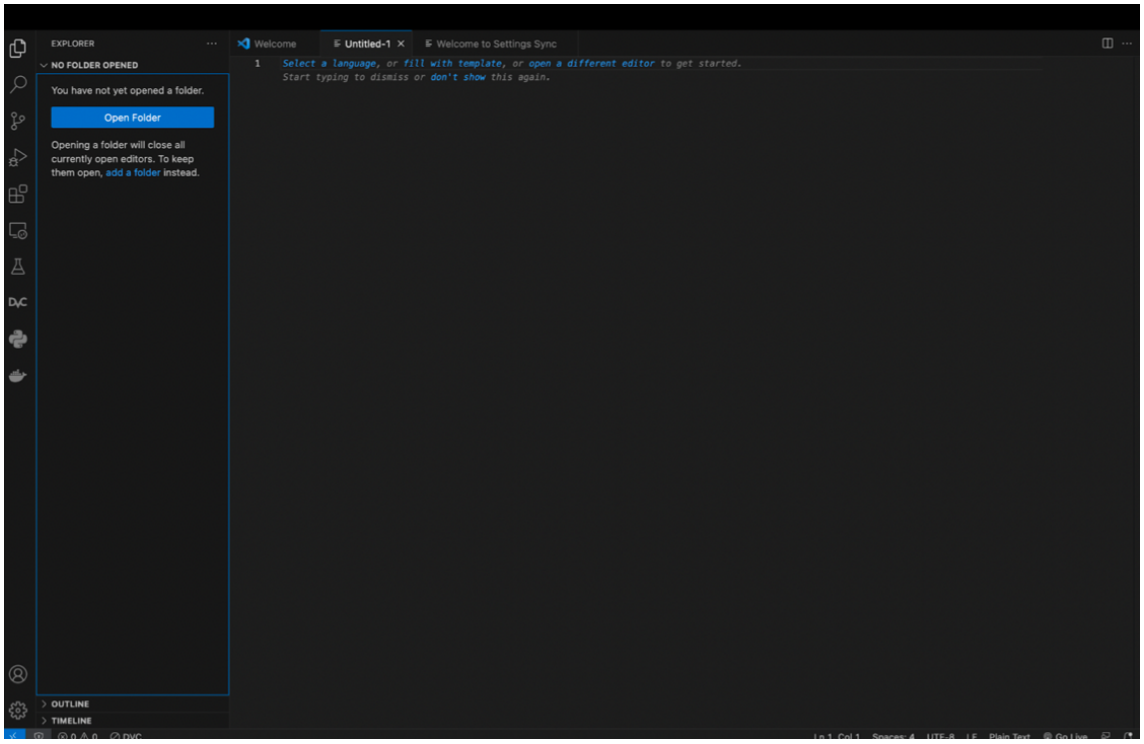
Vscode has recently integrated Atom, from the github team. It combines features from Microsoft and Github that will make your life much, much easier when coding. Similarly, it can be used to deploy things in GitHub, render html web pages, operate with cloud services such as Google Cloud or AWS and much more!

The official Vscode webpage already gives some quick installation guidelines for your OS. Installation follows as any other software and shouldn't give you any trouble. Therefore, I will focus on how to set up python and conda once vscode is ready and give you some tricks to personalise it (I take for granted you have python and conda already installed at this point).

VScode can be synchronized in different machines in case you operate with multiple laptops or ever change to a new computer. For this, I recommend you to create a GitHub account. Microsoft users can also sign up with their own accounts. Once you've done this, VScode will ask you to sign in the initial installation steps.

Now let's run some code!

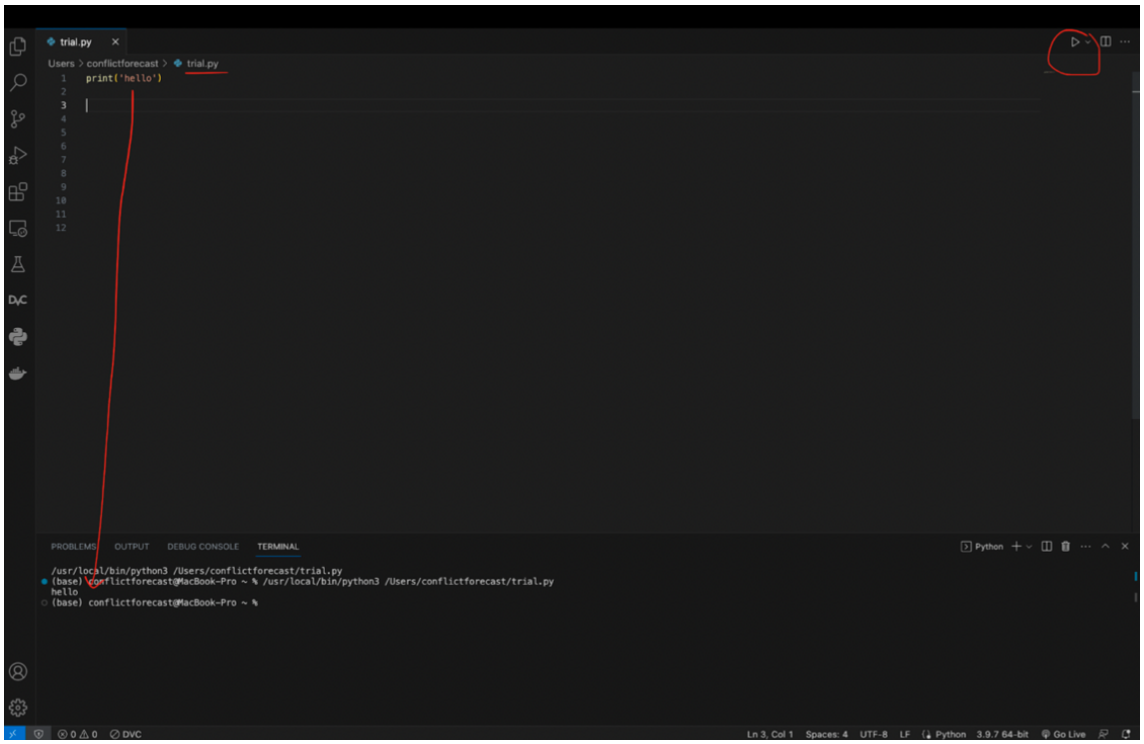
Once you open a new text file, you will see something like this:



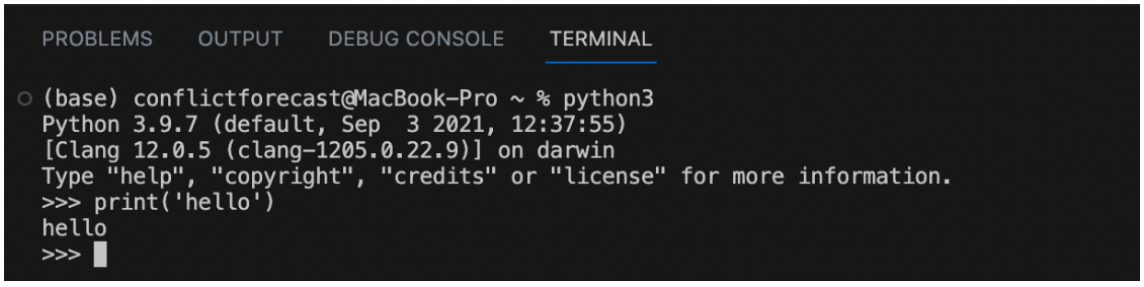
On the left panel, you have the explorer. Once we cover how to operate with environments, VSCode will remember which folder and environment you are working in and it will make it easier to initialise projects. Notice that the new text file, Untitled-1, has no default language. As we saw before, an IDE is not only for python but you can support multiple languages (even latex!). Selecting the python language will make VScode aware and able to highlight your syntax mistakes (you will see howcool this can get once you integrate copilot in). So, let's select Python.

.py vs .ipynb

I've saved the file as .py but often you will see python files with the extension .ipynb. What is the difference? The .py file is just a regular python text file. You can run it in **any** setup that has python installed and that's all it takes. When you click on the play button on the right top panel, the terminal pops up and your python code runs.



Another way is to run the code via CLI. A Command Line Interface basically refers to instructions that you can tell your computer to do through the terminal. Although we will see some handy commands to operate with the terminal during the brush up sessions(bash language) , this will not be the focus now. Simply, in order to run python via CLI, just write *python3* or *python* in your terminal.



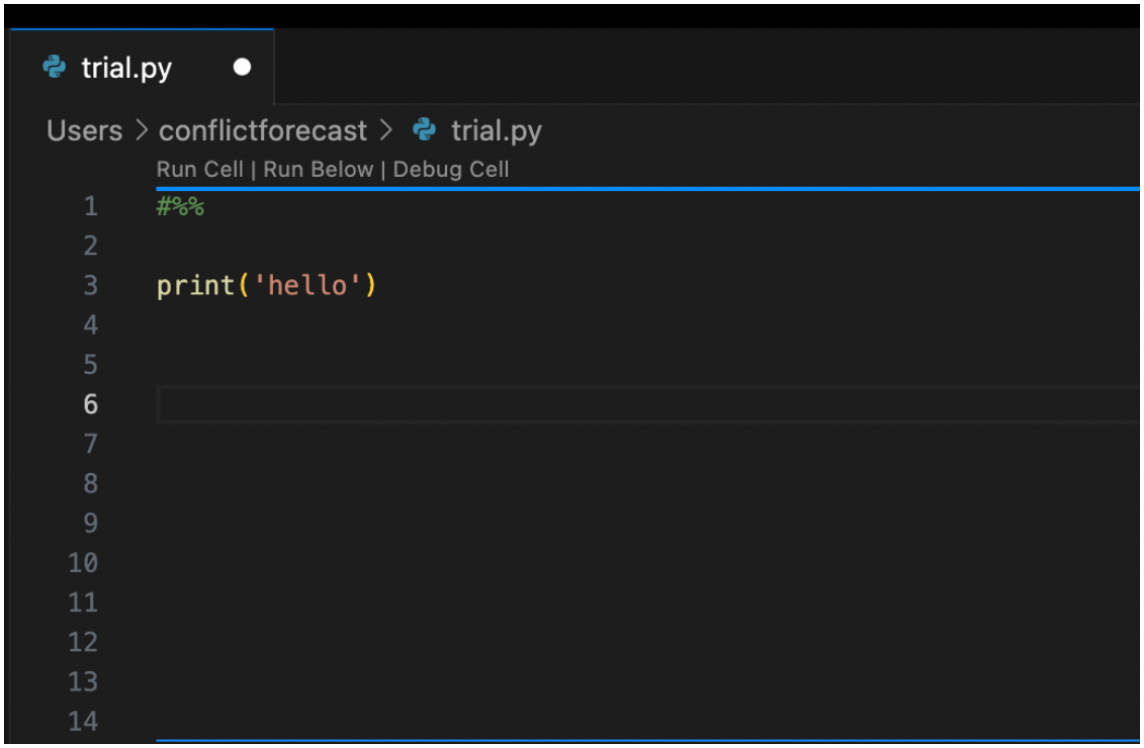
Now the >>> indicates you're inside python (you don't even need an IDE for this!). Another way to run codes is by calling it directly with the python3 command:

```
(base) conflictforecast@MacBook-Pro ~ % python3
Python 3.9.7 (default, Sep  3 2021, 12:37:55)
[Clang 12.0.5 (clang-1205.0.22.9)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> print('hello')
hello
>>> exit()
(base) conflictforecast@MacBook-Pro ~ % python3 trial.py
hello
(base) conflictforecast@MacBook-Pro ~ %
```

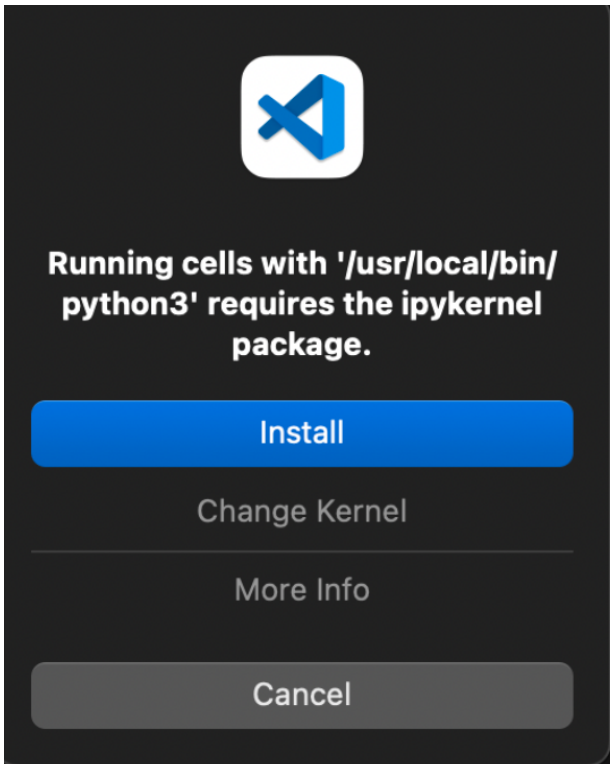
Now let's focus on .ipynb formats. A python notebook requires Jupyter to be installed. Typically, you will see them together with Jupyter lab but Vscode can handle them without any problem. A file with a .ipynb format cannot be run with ordinary python and vice versa. This typically creates problems if you're collaborating in a project or working with multiple extensions. Is there a workaround?

Yes! Vscode integrated the best of Atom in order to operate with Conda in a .py file. How?

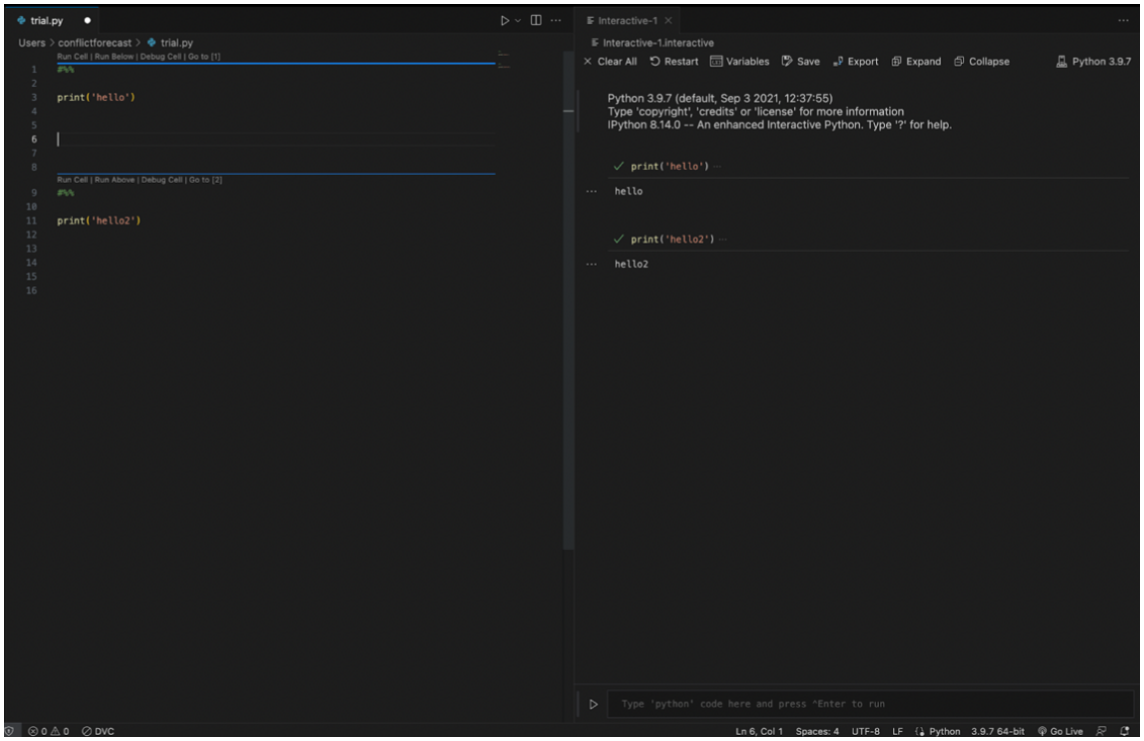
Just type #%% at the beginning of the textfile:



You immediately see how Vscode detected this as a cell. If you click on run cell (or do command + enter), you will be guided to the necessary installation to run this code via Conda. You need to install a kernel to interact with Jupyter:



Now look at the right panel. An interactive shell appears where you will see the output of your code. Basically, you will write your code on the left and see results of the cells on the right (of course, this layout is totally customizable).



But then, what should I use .py or .ipynb?

There is no close answer to this. Typically, whenever you're exploring data or creating new codes you will need to change your code often, correct mistakes etc. This is where .ipynb comes handy. You won't need to run your code from the beginning for every change you do. However, from personal experience I can tell you notebooks operate through Jupyter and Conda and are much more prone to break due to external errors (kernel errors). On the other hand, .py files do not require Conda, can be easily integrated with other .py files and are extremely useful if you want to operate in Cloud services.

But, notice the file I was running with #%% is still .py! This is the power of vscode. It allows you to explore via Conda in the same format as a .py file. Once your code is ready to deploy or run, you can just use python CLI and leave it running safely.

Handy extensions for Vscode:

Extensions are typically open code contributions to customise IDES. This is one of the main reasons why people choose Vscode vs others. Here, I will just tell you my favorite ones:

- **Atom One Dark Theme**
- **DVC** : Very handy to run machine learning experiments, store results etc.
- **Python Indent:** It will highlight to you if there are indentation problems in your files.
- **Ident-rainbow**: Colors indentation levels.
- **Github Copilot**: We will talk more about this in the sessions. You can apply for a free copilot license with your BSE account here.
- **autoDocstring**: To document codes and keep track of your functions.
- **Peacock**: One we operate with environments, this extension allows you to customise different colours for each one. It is very handy once you're coding on multiple projects.
- **Github Pull requests**: Mandatory to deploy projects in github.

3 Installation for R

Programmers and data scientists typically have low attachments to specific programming languages. Languages change over time, they specialize for different tasks and its often good

to be versatile [1] This is partly why you will be using different languages in this program. Another popular one for data science and statistics is R. Downloading it is very easy, just follow their installation guide. (warning, choose the windows one if you're not a Mac user). As we saw before, you need an IDE. Vscode can handle R too but their have their own specific IDE so let's use that! Donwload it here.

If you have any questions don't hesitate to send me an email to **luisigmenendez@gmail.com**.

[1] Just see the rankings for old low level programs such as C++ or Fortran. These languages have switched their function to specialize in system development over the years but they are still used and combined.