GET1030

Computers and the humanities

Lecture 4

Gentle introduction to python

Dr Miguel Escobar Varela

Objectives

To learn how to produce the most common types of scientific data visualizations (as seen in lecture #3) in Python

To turn you into a Python fan (if you aren't one already)

To pique your interest and encourage you to seek Python mastery beyond what this course offers

Python

Created by Guido van Rossum and first released in 1991

The language's philosophy emphasizes code readability

Increasingly used for scientific and academic programming

It is named after the British comedy group Monty Python.

A task-oriented approach

We will not cover all the basics in sequence (as it is often done in intro programming courses) but jump right into a task and just learn the things we need for that task.

If you want to know more about the bits we skipped, I recommend this free resource:

https://allendowney.github.io/ElementsOfDataScience/

‘Survival’ python

Let’s say I’m teaching you how to say a few words in a new language before a trip. I won’t teach grammar or basic vocabulary. Just some survival phrases. That’s what I’m doing here with Python.

A task-oriented approach

The task: to load an Excel file and visualize it in different ways, and export the results as stand-alone images.

Minimum concepts you need to know

- Variables
- Functions
- Libraries
- Data frames

Setting up

Requirements Python, Pandas, Matplotlib, Seaborn and Jupyter.

If you have never installed Python, it is recommended that you install all packages at once by installing Anaconda:
https://docs.continuum.io/anaconda/install/

If you are an experienced programmer just make sure that you have Python and the packages above installed.

Jupyter

- Notebook combine executable code and comments (written in markdown).

Visualizing data

For visualizing data, we are going to use two other libraries, called **Seaborn** and **Matplotlib**. The steps should now be familiar.

In [12]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

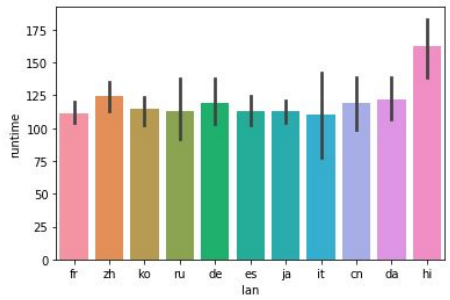
*For more information, please refer to the official Seaborn tutorials*Seaborn is built on top of Matplotlib. To understand how the visualizations work at a deeper level, please see the official Matplotlib tutorial

Barplots

We are now going to see a few types of data visualization. Each one will take a few parameters (similar to arguments in the functions we saw earlier). To create a bar plot, we specify three parameters:

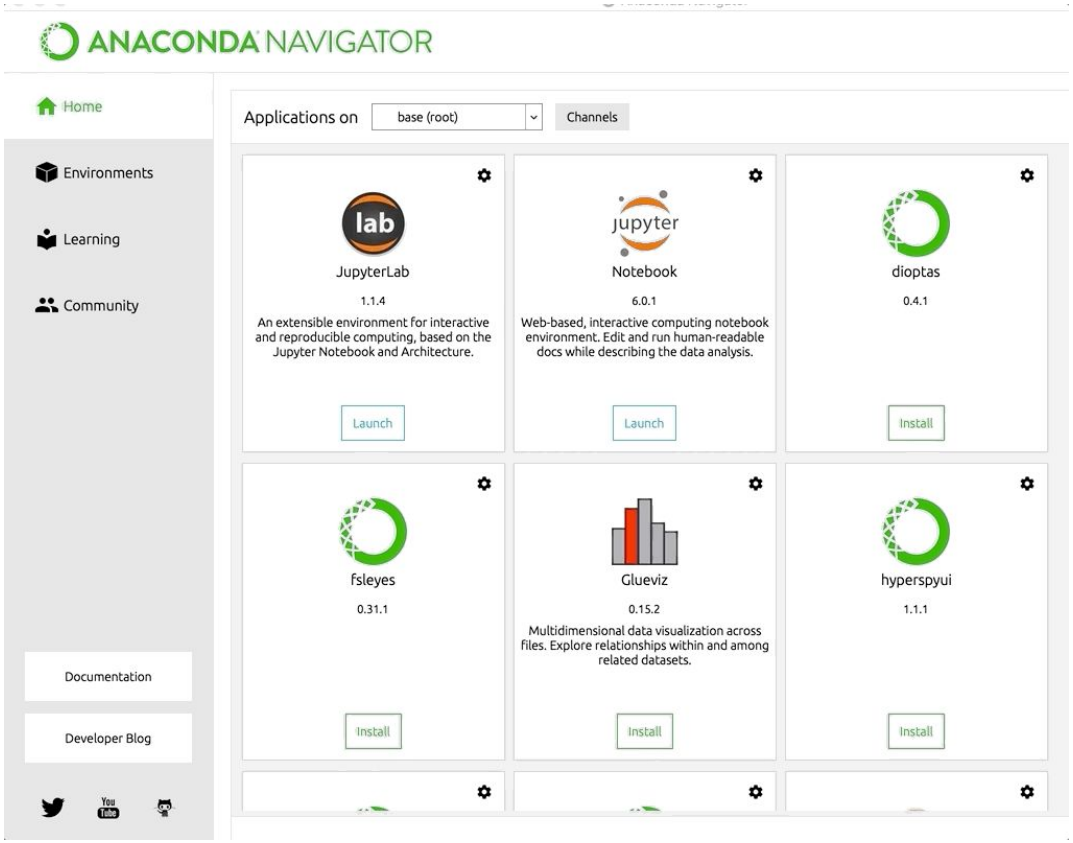
- x = the categories to be displayed in the x axis
- y = the values to be graphed
- data = the source of the data, in this case the data frame we created earlier, by exporting data from Excel.

In [13]:
```
ax = sns.barplot(x="lan",y="runtime", data=df)
```



Jupyter App

- If you are on windows, the Jupyter Notebook App can be launched by clicking on the Jupyter Notebook icon in the start menu or by the Anaconda Navigator.



Jupyter from Command Line

You can also launch Jupyter from the command line in Windows, Mac and Linux with the following command:

```
jupyter notebook
```

It is recommended to first navigate to the folder where you have your current project stored.

Windows
https://devblogs.microsoft.com/commandline/windows-terminal-preview-1-1-release/

Mac

https://lifehacker.com/launch-an-os-x-terminal-window-from-a-specific-folder-1466745514

For troubleshooting, see here: https://jupyter.readthedocs.io/en/latest/running.html

Things to learn in Jupyter

- Open and run an existing file
- Create a new file
- Add comments and execute code
- Understand errors (read the traceback)
- Load data from csv, excel

Jupyter shortcuts

- Ctrl + Enter to execute a cell
- B to add a cell below the current one
- A to add a cell above the current one
- Pres d + d to delete a cell