Python for data science

Dani Brake

October 12, 2017

UWEC Data Analytics
This presentation is available at my homepage, danibrake.org/papers

The notebooks are available at github.com/ofloveandhate/data_science_python
Outline

1. Environments & tutorial
2. Python packages for data science
3. Do something
An environment is a way of interacting with software

In Python, we have these critical elements:
- Text editor
- Terminal – IO interface
- Working install of Python

Two major alternatives
1. Terminal-texteditor pair
2. IDE – Integrated Development Environment
Terminal & text editor

Your choice of text editor – I really like Sublime and Atom

Your choice of terminal program – your favorite shell, too

Empowering for those environments without a GUI connection

A steeper learning curve

My preferred method for library authoring
Jupyter Notebook

- All-in-one REPL
- Read-eval-print loop
- Markdown cells and code cells
- Full power of Python.
- Github can view them online. Easy to share.

My preferred method for interactive presentations

Welcome to Python!
I am super glad you are here with me. This is a Jupyter notebook, a way of interacting with the Python interpreter, and writing a presentation at the same time.

Traits of Python
Python is...

Interactive
Let’s make a cell, and run it right in the worksheet.

```python
In [1]:
1 #this is a code cell.
2 #You can write arbitrary Python code.
3 #This is a great place for demo code, and a poor place for library code.
4 #the first function we should write
5 def hello():
6 print("hello" + "world") #colons is required
7 #python accepts both double and single quotes to define string literals
```

```python
In [2]:
1 #now we call the function we just wrote
2 hello()
```

```
helloworld
```

^^^ the output from the function we just ran.

We just defined a function, and ran it, right in our browser, without being connected to the internet (necessarily).

eFl’s explore the constraints of the system. You should know some things.

Whitespace matters
If you don’t indent correctly, your code won’t run. It’ll be all like ‘unexpected indent’ and stuff.

```python
In [3]:
1 a = 1
2 b = 2
3 c = a+b+1.1
```
There are tons of libraries to help you solve your problems.

Core computation
- NumPy
- SciPy
- Matplotlib
- Sympy

Plotting & vis
- Pandas
- Seaborn
- Bokeh
- Plotly

Machine learning & AI
- Tensorflow
- NLTK
- Scrapy

this is not at all exhaustive!
Resources are there for you

Cheatsheets
- python
- another python
- pandas
- matplotlib
- markdown

Websites and tutorials
- https://learnpython.org/
- https://docs.python.org/3/tutorial/
- http://sthurlow.com/python/
- https://pythonprogramming.net/math-basics-python-3-beginner-tutorial/
Live demos

because i heart you
Thank you for your kind attention!