

IMD0105 - Special Issues in Information Technology VI

Introduction to Python for Data Science

Natal-RN
March 2017



Previously on last class (...)



<https://goo.gl/2dC811>

BIG DATA



Agenda

- Data Science War
- Anaconda
- My First Notebook
- Intro to Python for Data Science



DataCamp
Learn data analysis for free,
interactively

DATA SCIENCE WARS

























VS.



python

R and Python are waging war:
while both programming languages are gaining prominence
in the data analytics community, they are fighting
to become data scientists' language of choice.

Which side are you taking?

Language Rank	Types	Spectrum Ranking
1. C	  	100.0
2. Java	  	98.1
3. Python	 	98.0
4. C++	  	95.9
5. R		87.9
6. C#	  	86.7
7. PHP		82.8
8. JavaScript	 	82.2
9. Ruby	 	74.5
10. Go	 	71.9



Guido van Rossum wrote in 1996.

"Over six years ago, in December 1989, I was looking for a "hobby" programming project that would keep me occupied during the week around Christmas. My office ... would be closed, but I had a home computer, and not much else on my hands. I decided to write an interpreter for the new scripting language..."



Version 3.x (<https://www.python.org/downloads/>)



ANACONDA®

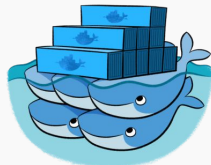
Modern open source analytics platform
powered by Python



<https://www.continuum.io/downloads>

Why Anaconda?

- [Anaconda](#) is a distribution of packages built for data science.
- It comes with **conda**, a package and environment manager.
- You'll be using conda to **create environments for isolating** your projects that use different versions of Python and/or different packages.





Home

Environments

Learning

Community

Documentation

Developer Blog

Feedback



Applications on

py3



Channels

Refresh

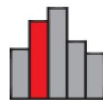


notebook

4.3.1

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

Launch



glueviz

0.9.1

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

Install



qtconsole

4.2.1

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Install



spyder

3.1.3

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

Install

Hands on

Install packages:

```
$conda install pandas matplotlib scikit-learn numpy
```

<https://conda.io/docs/>

