

References for some of the Software Carpentry lessons

Summary: A list of Software Carpentry related free online resources that I have compiled along time.

Contents

bash	1
git	1
Python	2
general references / guides	2
Scientific Python	2
pandas	2
Data Science / Analysis	2
IPython - Jupyter	3
Integrated Development Enviroments - IDEs	3
HPC	3

bash

For a tutorial like guide to bash <http://www.bash.academy/>

Once you get more advanced, or just for reference or looking for specific things, the "Advanced Bash-Scripting Guide" is a very useful resource (and it comes first in google many times when searching for bash related subjects)

<http://tldp.org/LDP/abs/html/>

For an interactive explanation of bash commands see <https://explainshell.com/>. Try any of the classroom examples we did, or just the examples you can find there.

git

A good Mercurial introduction <http://hginit.com/index.html> . Mercurial is "the other" Distributed Version Control System, and its syntax is very similar to git. All the concepts apply to both, and it has good basic insights on how Mercurial (and git) works.

Later on, the "Pro Git book" has anything you need as a reference for git: from basic commands, up to how setting up your own central git repository server with ssh access (or locally) in a few lines <https://git-scm.com/book>

CodeAcademy free TryGit online tutorial, created by GitHub and useful for reviewing the course sillabus, and a bit more <https://try.github.io/levels/1/challenges/1>

Branching explained in an Interactive tutorial <https://learngitbranching.js.org/>

For a list of Git graphic user interface (GUI) clients see <https://git-scm.com/downloads/guis>

Python

general references / guides

If you want to jump straight into Python from zero (even if you know other programming languages), I strongly recommend "Python for Everybody". Really clear approach and very easy to follow, it also gives the most affordable approach to Object Oriented Programming I've ever seen. It's free, but you can buy a hard copy really cheap. Also the author has a very good Coursera track based on the book.

<https://www.py4e.com/book.php>

Also, "The Hitchhiker's Guide to Python" <http://docs.python-guide.org/en/latest/> is a very good all-around Python reference and study guide.

"A Whirlwind Tour of Python" is a fast-paced introduction to essential components of the Python language for researchers and developers who are already familiar with programming in another language. Book and Jupyter notebooks.

<https://github.com/jakevdp/WhirlwindTourOfPython>

A quite extensive community list of resources for the Python avid ones. From the very basic for people new to programming, to specific occupations as "Python for the Humanities" references:

<https://www.fullstackpython.com/best-python-resources.html>

A curated collection of Jupyter/IPython notebooks that are notable, from the Jupyter project GitHub page itself <https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks>

Scientific Python

For a more scientific oriented goal, perhaps for people with some at least basic python knowledge, the "Scipy Lecture Notes" are a great resource. With lessons on how to interface Python with C and Fortran, and other more advanced topics.

<http://www.scipy-lectures.org/>

Notebooks from J.R. Johansson, "Scientific Computing with Python", cover a wide variety of subjects, even some advanced ones as HPC or interfacing Python with C and Fortran

<https://github.com/jrjohansson/scientific-python-lectures>

For those of you using Matlab, see an online seminar entitled "Python for MATLAB users" and a related whitepaper, "MATLAB to Python Migration Guide", from Enthought.

<https://www.enthought.com/webinar/python-for-matlab-users>

pandas

"Pandas in a nutshell" notebook <http://kanoki.org/2017/07/16/pandas-in-a-nutshell/>

A Beginner's Guide to Optimizing Pandas Code for Speed

<https://engineering.upside.com/a-beginners-guide-to-optimizing-pandas-code-for-speed-c09ef2c6a4d6?gi=789797286ed>

Data Science / Analysis

Python Data Science Book, from Jake VanderPlas

<https://jakevdp.github.io/PythonDataScienceHandbook/>

Data science Python notebooks: Deep learning (TensorFlow, Theano, Caffe, Keras), scikit-learn, Kaggle, big data (Spark, Hadoop MapReduce, HDFS), matplotlib, pandas, NumPy, SciPy, Python essentials, AWS, and various command lines.

<https://github.com/donnemartin/data-science-ipython-notebooks>

IPython - Jupyter

Slideshow about IPython and Jupyter, 34 slides, very good

<http://eueung.github.io/python/ipython-intro>

28 Jupyter Notebook tips, tricks and shortcuts

<http://www.pybloggers.com/2016/10/28-jupyter-notebook-tips-tricks-and-shortcuts>

Building Interactive Dashboards with Jupyter

<https://blog.dominodatalab.com/interactive-dashboards-in-jupyter/>

Integrated Development Enviroments - IDEs

The most popular IDEs for Python are

- Pycharm: proprietary software with a good-enough community version, specifically designed for Python development.
- Microsoft's Visual Studio Code: open source general IDE, with goog Python support through its plugins system.
- Atom and SublimeText: text editors which can become kind of IDEs by means of pluggins and addons.
- VIM (I had to do it :-)

By the moment there is no R-Studio equivalent.

HPC

SGE - SLURM migration guide / table

<https://confluence.csiro.au/display/SC/Reference+Guide%3A+Migrating+from+SGE+to+SLURM>