Installation Guide

REQUIRED TOOLS

Before the course starts, you may want to familiarise yourself with the following technologies:

- **Anaconda** - We will be using Anaconda as our primary development environment.
- **Python 2.7** - We will be using Python and its packages as our primary language.
- **Github** - We will be using Github on a daily basis to store and share our code.
- **Git**
  - Mac: **Git Bash**
  - PC: You will also need to install Git command line tools for your OS.
- **Postgres** - We will be using Postgres for local SQL-based data storage.
- **Tableau** is a popular dashboard creation system for visualising data.
- **Slack** - We will be using Slack on a daily basis to communicate with each other.

COMMON TOOLS

- Anaconda bundles many of the Python packages we will be using, including:
- **Python 2.7**: The most widely used, stable, enterprise version of Python.
- **Ipython / Jupyter** and **Pandas**: Core tools for notebooks and data analysis.
- **Matplotlib**: The king of all python plotting packages.
- **Gensim**: Framework for vector modelling.
- **NLTK** and **Spacy**: Used for natural language processing.
- **NumPy**: Array processing tool.
- **Scikit-learn**: Modules for machine learning and data modelling.
- **SciPy**: Scientific library for python.
- **Seaborn**: Statistical data visualiser.
- **Pip** and **Setuptools**: package installer and version manager (Mac only).
- **PyMC**: common stats tool for simulation and optimisation.
- **Sqlite**: Standalone, lightweight SQL database engine.
- **Statsmodels**: Simple statistical computation (used with SciPy).

ADDITIONAL TOOLS

These tools are not specifically required, but are highly recommended.

- **Atom** or **Sublime** are popular text editors for writing scripts to process data, perform analysis, and create visualisations.
- **Chrome** is Google's popular web browser, and comes with a complete set of developer tools built-in.
- **Import.io**: a useful web scraping tool with a graphic interface.
- **Plot.ly**: a user-friendly tool for plotting graphs.

A NOTE ABOUT TECHNOLOGY

Follow the guidelines below to ensure your machine is fully prepared for Data Science:

System Requirements

Make sure your machine is running with administrator permissions and has **at least 10+ GB of free disk space**. We also recommend that you use a laptop with a **13-inch screen or larger** in order to do your best work. In our experience, *students with an 11-inch screen have a harder time in class.*

Mac Users

General Assembly is a Mac-friendly organisation. Our instructors will be teaching the course using Macs, so we strongly recommend students use a Mac with OS X 10.12 (Sierra) in order to run all of the programs necessary for the course. **This rules out some older MacBooks.**

Check the following specs to make sure your machine can provide you with the performance you will need in this course:

- *1.6GHz dual-core Intel Core i5 processor*
- *Turbo Boost up to 2.7GHz*
- *Intel HD Graphics 6000*
- *At least 8GB RAM*
- *128GB flash storage*
- *10+ GB of free disk space*

PC Users

While you can be a data scientist with *almost any* machine, some students have found compatibility issues with older versions of Windows. While you

can be a data scientist with any machine, unfortunately, there are a number of compatibility issues with Python libraries and older versions of Windows. For example, Python and Anaconda users have identified multiple issues with Windows 7 x64 machines.

Therefore, we **strongly recommend** that PCs users adopt the latest version of Windows (Windows 10). PC users on older machines may consider installing a Virtual Machine like **Oracle's Virtualbox** and running Anaconda in a Linux environment via **Ubuntu Desktop**. **See more information here**.

Please note that our instructors will be conducting the course using Macs, and *may not be able to help PC or Linux users troubleshoot* any issues you might encounter. ***If you choose to use a PC or Linux machine, you will need to provide your own IT support.***

LINKS

- **Anaconda**, https://www.continuum.io/downloads
- **Python 2.7**, https://www.python.org/download/releases/2.7
- **Github**, http://github.com
- **Git**, https://git-scm.com/book/en/v2/Getting-Started-Installing-Git
- **Git Bash**, https://git-for-windows.github.io
- **Postgres**, http://www.postgresql.org/download
- **Tableau**, http://www.tableau.com
- **Slack**, http://slack.com
- **Atom**, https://atom.io
- **Sublime**, http://www.sublimetext.com
- **Chrome**, https://www.google.com/chrome
- **Import.io**, https://www.import.io
- **Plot.ly**, https://plot.ly
- **Oracle's Virtualbox**, https://www.virtualbox.org/wiki/Downloads
- **Ubuntu Desktop**, http://www.ubuntu.com/download
- **See more information here**, https://docs.continuum.io/anaconda/images