DATA SCIENCE @ GA

TECHNICAL GUIDE

REQUIRED TOOLS

Before the course starts, you should have the following technologies installed:

- **Anaconda** - We will be using Anaconda as our primary development environment.
- **Python 2.7** - This is installed by default on most Macs and also comes with Anaconda.
- **Github** - We'll be using Github on a daily basis to store and share our code.
- **Git** (mac) / **Git Bash** (pc) - Students should install command line tools for Git.

COMMON TOOLS

- **Anaconda** bundles many of the Python packages we'll be using, including:
    - *Python 2.7*: The most widely used, stable, enterprise version of Python.
    - *Ipython* / *Jupyter* / *Pandas*: Required tools for creating notebooks.
    - *Matplotlib*: The king of all python plotting packages.
    - *Gensim*: A framework for vector modeling.
    - *NLTK* & *Spacy*: Used for natural language processing.
    - *NumPy*: Fundamental array processing tool.
    - *Scikit-learn*: Modules for machine learning & data modeling.
    - *SciPy*: Scientific library for python.
    - *Seaborn*: Statistical data visualizer.
    - *Pip* & *Setuptools*: package installer & version manager (Mac only).
    - *Sqlite*: Standalone, lightweight SQL database engine.
    - *Statsmodels*: Simple statistical computation (used with SciPy).

OPTIONAL TOOLS

These tools aren't required, but many students have found them helpful:

- Atom or Sublime - Popular text editors for writing code.
- Import.io - Simple web scraping tool with a graphic interface.
- Plot.ly - User-friendly tool for plotting graphs.

HARDWARE SPECIFICATIONS

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 organization. Our instructors will be teaching the course using Macs, and we strongly recommend students use a Mac with OS X 10.11 ("El Capitan") in order to run all of the programs necessary for the course. *This rules out some older MacBooks*.

If in doubt, compare your machine against the following hardware requirements:

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

PC Users

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.

We **strongly recommend** that PC 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.

IT Support

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***.