

DATA SCIENCE

TECHNICAL INSTALLATION GUIDE

HARDWARE SPECIFICATIONS

Follow the guidelines below to ensure your machine is fully prepared for our upcoming Data Science course.

System Requirements

Make sure your machine is running with administrator permissions so that you can install the relevant programs and files. We also recommend that you use a laptop with at least 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.

Next, check your machine against the following hardware requirements:

- At least 8GB RAM
- At least 10GB free Hard Drive space (after installing Anaconda). More space is better!
- Your machine should be running on updated, patched operating systems no more than two years old.

Mac Users

General Assembly **strongly recommends Mac computers**. We recommend students use a Mac with an updated version of OS X in order to run all of the programs necessary for the course. *Macs with versions of OS X over two years old will have difficulty in-class.*

PC Users

While you can be a data scientist with *any* machine, unfortunately, some **Python libraries have compatibility issues with older versions of Windows**. For example, Anaconda users report trouble with *Windows 7 x64* machines.

Therefore, 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](#) to run Anaconda in a Linux environment via [Ubuntu Desktop](#). [See more information here](#).

IT Support

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

REQUIRED TOOLS

Before the course starts, familiarize yourself with the following technologies:

- [Anaconda](#) - We will be using Anaconda as our primary development environment.
- [Python 3.6](#) - This is installed by default on most Macs - and also comes with Anaconda!
- [Github](#) - We'll be using GA's Github servers on a daily basis to store and share our code.
- [Git](#) (Mac) / [Git Bash](#) (PC) - Students should install command line tools for Git.

HOW WE USE ANACONDA

- [Anaconda](#) includes many of the Python packages we'll be using in-class:
 - *Python 3.6*: The current stable version of the Python language.
 - *iPython / Jupyter*: Required tools for creating notebooks.
 - *Pandas*: Your go-to library for organizing and managing data.
 - *Matplotlib*: The king of all python plotting packages.
 - *Gensim*: Framework for vector modeling.
 - *NLTK*: 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).
 - *PyMC*: Common stats tool for simulation and optimization.
 - *Sqlite*: Standalone, lightweight SQL database engine.
 - *Statsmodels*: Simple statistical computation (used with SciPy).

ADDITIONAL TOOLS

We also ask that you install the following programs:

- [Atom](#), [Sublime](#), or [Visual Studio Code](#): Popular text editors for writing code.
 - [Slack](#): Messaging platform that we will use to chat in-class.
 - [Chrome](#): We'll use Google's web browser for its built-in developer tools.
-

GETTING STARTED

Here is a checklist you can use to ensure you are ready for class:

1. Look for an email invite to join your cohort's Slack group!
2. Create a Github Enterprise account on GA's Github server.
 - a. Go to git.generalassemb.ly
 - b. Create a new account
 - c. Enter a username that includes your first name and last name, so that instructors can easily identify you!
 - d. Share your username with your instructional team.
3. Install [Git Bash](#) on your PC (or [Git](#) on your Mac).
4. [Download and install Anaconda](#) for Python 3.6.
 - a. Test that your installation was successful by opening a Jupyter Notebook.
 - i. You can do this by opening Anaconda Navigator and clicking on Jupyter Notebook, which will open in your default browser.
 - ii. You can also do this from your computer's command line by typing "jupyter notebook"
 1. PC users should open the **Git Bash** command line terminal to enter their commands.
 2. Mac users should open the **Terminal** application to enter their commands.

STRETCH GOALS

The following sections are **optional**. Feel free to try these on your own; however, we will also review them together during class:

1. Confirm your version of python by typing `python -V` in your command line interface.
 - a. It should say: `Python 3.6`
2. Jupyter notebooks also let us open older versions of Python if desired. For example, you could type the following to create a Python 2.7 kernel:
 - a. `conda create -n py27 python=2.7 ipykernel`
`conda create -n py27 python=2.7`
 - b. Next, type the following (depending on your operating system):
 - i. Windows users: `activate py27`

- ii. Mac users: `source activate py27`
- c. `conda install notebook ipykernel`
`ipython kernel install --user.`
 - i. ^ Enter the username that you use to sign into your operating system here.
 - ii. Voila! Now you can run older versions of Python! This may be helpful for users who have to interact with legacy codebases.
- 3. If you like, you can also go ahead and install some additional libraries that we will use later on in the course.
 - a. **TextBlob:** This is a library we'll use when discussing natural language processing:
 - i. In your command line application, type in: `pip install -U textblob`
 - ii. Then type in: `python -m textblob.download_corpora lite`