

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 10.12 “Sierra” or higher in order to run all of the programs necessary for the course. *Macs with older operating systems or machines over two years old will have difficulty in-class.*

PC USERS

If you are a PC user, we **require** that you use the most recent version of windows (“Windows 10”) in order to run the programs for our course. Students have reported a number of compatibility issues running our python libraries with older versions of Windows.*

*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](#).

IT SUPPORT

Please note that most of our instructors will be conducting the course using Macs, and may not be able to help Windows or Linux users troubleshoot any issues you might encounter. While

we're happy to try and lend a hand, you will ultimately be responsible for your own tech support!

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).
-

CONTEXT ON TOOLS

These are the technologies we'll use in class:

- [Anaconda](#) - We will be using Anaconda as our primary development environment. When you get to class, you'll install Python on your machine using this tool. It's important that you, your instructor, and your classmates all use the same versions of Anaconda and Python, so install Anaconda and Python 3+.
- [Python](#) - This is included with your Anaconda installation.
- [Github](#) - We'll be using GA's Github servers on a daily basis to store and share our code. Set yourself up with [an account](#) on GA's internal GitHub server so that you're ready to use GitHub in your pre-work and can access course content in class. **Note: Set up an account even if you already have a GitHub! We'll be using a version called "Github Enterprise" that's separate from any GitHub account you've previously created.**
- **Git (Mac/PC):** Git is a version control system that tracks changes in computer files and makes it easy to coordinate work across multiple machines and people. PC users should install [GitBash](#); Mac users should install [Homebrew](#).
- [Atom](#), [Sublime](#), or [Visual Studio Code](#): Popular text editors for writing code.
- [Chrome](#): We'll use Google's web browser for its built-in developer tools.

HOW WE USE ANACONDA

- [Anaconda](#) includes many of the Python packages we'll be using in-class:
 - *Python*: 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).
-