



Tutorial Information and Participant Instructions

Tutorial set up instructions will be added in June. There will be a slack channel set for each tutorial if you have trouble with set up or have additional questions.

Introductory Tutorials

Intro to Python Programming

Matt Davis

This tutorial is a gentle introduction to Python for folks who are completely new to it and may not have much experience programming. We'll work in a Jupyter Notebook, one of the most popular tools in scientific Python. You'll learn how to write beautiful Python while practicing loops, if's, functions, and usage of Python's built-in features in a series of fun, interactive exercises. By the end of the tutorial we think you'll be ready to write your own basic Python -- but most importantly, we want you to learn the form and vocabulary of Python so that you can understand Python documentation, interpret code written by others, and get the most out of other SciPy tutorials.

Tutorial Prerequisites: None

Set up instructions:

Introduction to Numerical Computing With NumPy

Eric Olsen

NumPy provides Python with a powerful array processing library and an elegant syntax that is well suited to expressing computational algorithms clearly and efficiently. We'll introduce basic array syntax and array indexing, review some of the available mathematical functions in NumPy, and discuss how to write your own routines. Along the way, we'll learn just enough about matplotlib to display results from our examples. Tutorial Prerequisites: The tutorial is intended for people new to the scientific Python ecosystem.

Tutorial Prerequisites: The tutorial is intended for people new to the scientific Python ecosystem. Previous experience in Python or another programming language is useful but not required.

Set up instructions:

Learn Python through Data Processing in Pandas

Daniel Chen

The goal of this tutorial is to guide new learners into Python, especially those who are first-time attendees of SciPy. We'll introduce how to program in Python using data cleaning in pandas as the teaching example. This will help transition new learners who work with data in spreadsheets but want to utilize the power of the scientific python stack.

Tutorial Prerequisites: The tutorial is intended for people new to the scientific Python ecosystem. Previous experience in Python or another programming language is useful but not required.

Set up instructions:

The Jupyter Interactive Widget Ecosystem



Set up instructions:

Introduction to Conda for (Data) Scientists

David Pugh

This tutorial is a Software Carpentry-style introduction to Conda for (data) scientists. This tutorial motivates the use of Conda as a development tool for building and sharing project specific software environments that facilitate reproducible (data) science workflows. Particular attention is given to using Conda to create reproducible environments with NVIDIA GPU dependencies (including environments for Horovod, TensorFlow, PyTorch, and NVIDIA RAPIDS) as well as a discussion of best practices for using Conda in HPC environments.

Tutorial Prerequisites: Basic familiarity with Python programming and Bash shell concepts (i.e., basic commands, environment variables, etc). Familiarity installing NVIDIA CUDA Toolkit would be beneficial for NVIDIA GPU focused episodes.

Set up instructions:

Intermediate/Advanced Tutorials

Xarray for Scalable Scientific Data Analysis

Joseph Hamman, Ryan Abernathy, Deepak Cherian, Stephan Hoyer

Xarray provides data structures for multi-dimensional labeled arrays and a toolkit for scalable data analysis on large, complex datasets with many related variables. Xarray combines the convenience of labeled data structures inspired by Pandas with the multi-dimensional arrays of NumPy and parallel out-of-core computation from Dask to provide an intuitive, powerful and scalable platform for scientific analysis. This tutorial will introduce data scientists already familiar with Numpy and Pandas to the Xarray package and will guide participants through the process of using Xarray from small to big data applications. The tutorial also highlights how Xarray interacts with the greater scientific Python ecosystem and a wide range of common array storage formats.

Tutorial Prerequisites: Students are expected to have some familiarity with Jupyter, Numpy, and Pandas. No specific domain knowledge is required to effectively participate in this tutorial.

Set up instructions:

Parallel and Distributed Computing in Python with Dask

James Bourbeau, Mike McCarty, Dharhas Pothina

Dask is a library for scaling and parallelizing Python code on a single machine or across a cluster. Dask provides familiar, high-level interfaces to extend the SciPy ecosystem (e.g. NumPy, Pandas, Scikit-Learn) to larger-than-memory or distributed environments, as well as lower-level interfaces for parallelizing custom algorithms and workflows. This tutorial will cover the ins and outs of Dask for new users, including the Dask Array and Dask DataFrame collections, low-level Dask Delayed and Futures interfaces, pros and cons of Dask's task schedulers, and interactive diagnostic tools to help users better understand their computational performance.

Tutorial Prerequisites: No prior experience with Dask is required. Familiarity with Python, NumPy, and Pandas is preferred.

Set up instructions:

Deep Learning from Scratch with PyTorch

Hugo Bowne-Anderson, Dhavide Aruliah

This tutorial introduces deep learning (also called neural networks) to intermediate-level Pythonistas. The goal is for participants to develop a sound conceptual foundation for deep learning and to obtain some hands-on experience using an industry-ready toolkit. They will do this in two parts: (1) implementing a neural network classifier from scratch (following a quick review of NumPy array-based computing & supervised learning with Scikit-Learn); and (2) a tour of the PyTorch library building more sophisticated, industry-grade neural networks of varying depth & complexity.

Tutorial Prerequisites: Attendees should be comfortable with basic Python programming (e.g., data structures, functions, etc.) Some prior exposure to Python data science libraries (e.g., NumPy, Pandas, Scikit-Learn) is helpful.

Set up instructions:

Bayesian Data Science by Simulation

Eric Ma, Hugo Bowne-Anderson



Tutorial Prerequisites: Knowledge of `numpy`, `matplotlib`, and `Python` are prerequisites for this tutorial; in addition to curiosity and an excitement to learn new things!

Set up instructions:

Spatial Data Analysis with PySAL

Serge Rey, Elijah Knaap

This tutorial is an introduction to geospatial data analysis in Python, with a focus on the Python Spatial Analysis Library (PySAL). It introduces participants to the different libraries to work with vector geospatial data, and will cover munging geo-data and exploring relations over space. This includes importing data in different formats (e.g. shapefile, GeoJSON), visualizing, combining, and tidying them up for analysis. The second part of the workshop focuses on applications of spatial analytical methods to geodemographics and segregation.

Tutorial Prerequisites: No previous experience with those geospatial python libraries is needed, but basic familiarity with geospatial data and concepts (shapefiles, vector vs raster data) and pandas will be helpful.

Set up instructions:



Institutional Sponsor

