Introduction to Data Science using Python

Installing Python

In this workshop we will be using the Anaconda Python Distribution, available from Anaconda at this link. We will be using Python version 3.7, so make sure you install that version. The Anaconda Python Distribution is available for Windows, Mac OSX and Linux.

See this screencast for a walkthrough on how to install Anaconda Python on your computer.

Setting up the Python environment

Anaconda allows you to set up `conda` environments, which are similar to traditional `virtualenv` virtual environments in Python. We create environments so that we can have a setup of Python and Python packages that is reproducible for a project, that we can port to other computers to create identical computing environments, and that provides a sandbox that is separate from other environments we may create for other projects.

We will create an environment called `python_ds` that will include the following packages:

- anaconda
- altair
- plotly
- plotnine
- graphviz
- pygraphviz

This can be done at the command line (the Anaconda Prompt that you can find on the Windows Start menu, or a Terminal on Mac OSX or Linux) using the command

```
conda create -n python_ds anaconda altair plotly plotnine graphviz pygraphviz
```

However, it may be easier using the Anaconda Navigator GUI that comes with Anaconda Python Distribution. I provide a walkthrough of this process at this link.

Programming environments

We will be using two programs, Spyder and JupyterLab, to interact with Python within the environments we have created. Both of these should be opened from within the `python_ds` environment so that we can access the packages we have installed there. At the command line, you can activate the `python_ds` package by typing

```
conda activate python_ds
```

You will see a `(python_ds)` before your cursor to let you know this environment is activated.

This can also be done using Anaconda Navigator, as seen in this screencast.

Notebooks

The Jupyter Notebooks that will accompany this workshop will be available at https://www.github.com/DistrictDataLabs/Brookings_Python_DS. These notebooks can be downloaded on your own computer, or run online using the Binder link available on that page.

These notebooks will be available the morning of the first day of the workshop.