

Self–Study Resources For Students Joining DS 210 (Spring 2022)

DS 210 (Programming for Data Science) is a part of the data science major curriculum. The main prerequisite for students taking the course is DS 110 (Intro to Data Science with Python) or an equivalent course. To help students prepare for DS 210, especially those who did not take DS 110, we list here various self–study resources.

Note that these resources cover much more than you would learn from DS 110. So don't be overwhelmed by the amount of information in them, but also make sure you understand the basics of all tools listed below and know where to look for additional information if you need it.

Python

Knowledge of basic features of Python is required. Online courses around the web that cover this and way beyond:

- <https://www.learnpython.org/>
- <https://www.codecademy.com/catalog/language/python>

Jupyter Notebook

You should be able to use Jupyter Notebooks for interactive data processing. There are many online tutorials that introduce Jupyter Notebooks, including:

- <https://realpython.com/jupyter-notebook-introduction/>
- <https://www.dataquest.io/blog/jupyter-notebook-tutorial/>

NumPy & Pandas

Python libraries NumPy and Pandas are very useful for efficiently processing real–world data. This includes reading and outputting data in popular data formats.

Resources:

- The first few sections of the NumPy user guide: <https://numpy.org/doc/stable/user/index.html>
- The first few sections of the Pandas user guide: https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html

Matplotlib

You should be able to use Matplotlib to visualize your data and things you discover about it. A good starting point is the “Basic Usage” section of the Matplotlib tutorial: <https://matplotlib.org/stable/tutorials/introductory/usage.html>