

COM6018 Data Science with Python

Week 5 - Introducing Matplotlib and Seaborn

Varvara Papazoglou

Overview

- Principles of Data Visualisation
- Introducing Matplotlib and Seaborn
- Introducing Assignment 1
- Review of Lab Class 4
- Preview of Lab Class 5

Principles of Data Visualisation

- Data Visualisation is a key part of Data Science.
- We need to understand how to present data in a way that is meaningful and informative.

<https://www.biostat.wisc.edu/~kbroman/presentations/graphs2017.pdf>

Introducing Matplotlib and Seaborn

What is Matplotlib?

- Matplotlib is a low-level Python library for creating plots.
- It is the most widely used plotting library for Python.
- It is very powerful and flexible.

What is Seaborn?

- Seaborn is a high-level Python data visualisation library built on Matplotlib.
- It is particularly good for statistical data visualisation (i.e., examining the distribution of data).

Matplotlib and Seaborn Tutorial

[Link to the Matplotlib tutorial](#)

Assignment 1

- This has now been released.
- Worth 40% of the module mark.
- Due on Friday 6th November 2026, 3pm (Week 6).
- **No lectures/labs in week 6!**

Review of Lab Class 4

- We continued our analysis of the atmospheric gas concentration data.
- We used Pandas to repeat the work we had done in Lab Class 2.
- We introduced more datasets, i.e., SF₆ and N₂O.
- We solved together steps 1-7.
- We will have a look at the solutions for steps 8 and 9 today.

The Problem

We want to combine the gas concentration datasets to produce a single plot like this:

Tasks

- Reading the data from the CSV files.
- Renaming fields to be consistent across the datasets.
- Dealing with missing data.
- Converting all gas concentrations to the same units.
- Merging the datasets
- Computing the carbon dioxide equivalent (CO₂e) concentration.
- Creating the plot.

We did all of this in NumPy in Lab Class 2 (lots of code!). We now did it in Pandas (much less code!).

The Solution Notebook

[Open the Solution Notebook \(Week 4\)](#)

Lab Class 5 Preview

- You will be provided with several complex datasets.
- You will use Matplotlib to reconstruct a series of plots.

Lab Class 5

[Link to Lab Class 5](#)

Next Steps

- **Review** the Week 4 materials:
 - Review the Lab Class 4 solution notebook if you have not already done so.
 - Make sure you have read the Week 4 tutorial on Pandas.
- **Prepare** for the Week 5 lab class:
 - Read through the Matplotlib tutorial.
- **Read** through the assignment instructions. Bring any questions to the lab class on Tuesday.
- **No lectures/labs next week!**