

COM6018 Data Science with Python

Week 3 - Introducing NumPy

Varvara Papazoglou

Overview

- Review of Lab Class 2
- Introducing NumPy
- Preview of Lab Class 3

Review of Lab Class 2

- Atmospheric Carbon dioxide (CO₂) and Methane (CH₄) concentrations.
- We want to plot the global warming effect of these gases combined.
- Methane is a more potent greenhouse gas than CO₂ so we can't just add the concentrations together.

Global Warming Potential

Global Warming Potential (GWP) - the relative ability of one molecule of a greenhouse gas to contribute to warming.

- GWP for CO₂ is defined to be 1.
- GWP for CH₄ is 25.

Can convert all greenhouse gases to CO₂ equivalent (CO₂e) by multiplying by their GWP.

- $\text{Total CO}_2\text{e} = \text{CO}_2 + 25 \times \text{CH}_4$

The Problem

Our CO₂ and CH₄ concentrations are stored in two different datasets from different sources.

Problems:

- The datasets are in different formats.
- The datasets are not aligned in time.
- The datasets do not use consistent naming conventions.
- The datasets may contain missing values.

Tasks

- Reading datasets from their files
- Making sure the data is stored with the correct types
- Selecting and renaming the relevant fields
- Merging the two datasets into one dataset
- Dealing with missing values
- Making new CO₂e field by combining existing CO₂ and CH₄ fields
- Plotting the result and interpreting the plot

Hidden aim of the lab class

Purpose of the lab class was to show

- these tasks are not as straightforward as they sound! 🌀
- implementations written in pure Python are slow. 🐌
- there are better ways to do this. 👍

Motivation for introducing NumPy and Pandas 🐼

The Solution Notebook

[Open the Solution Notebook](#)

Introducing NumPy

NumPy is a core Python package for scientific computing that

- provides a powerful N-dimensional array object,
- provides highly optimised linear algebra tools,
- has tight integration with C/C++ and Fortran code,
- is licensed under a BSD license, i.e., it is freely reusable.
- it is very fast. 

(Pronounced "numb pie" not "numpee" ! Short for 'Numerical Python')

NumPy Tutorial

[Link to the NumPy tutorial](#)

Lab Class 3 Preview

- We will use NumPy to do some analysis of the Lab 2 greenhouse gas data.

Lab Class 3

[Link to lab class](#)

Next Steps

- Review the Week 2 materials:
 - Review the Lab 2 solution notebook if you have not already done so.
 - Read the Week 2 tutorial on reading data
- Prepare for the Week 3 lab class:
 - Read through the NumPy tutorial.
- Checkout the reading list for Week 3 on Blackboard.