COM6018 Data Science with Python

Week 4: Introducing Pandas

Jon Barker

Copyright © 2023–2025 Jon Barker, University of Sheffield. All rights reserved.

In this lab

- Reading datasets from CSV files
- Dealing with missing data
- Merging datasets
- Plotting data
- Comparing Pandas with previous Python-only approaches

The Task

- You will analyse atmospheric gas concentration data.
- Extending our previous analysis to include more greenhouse gases.
- We will be using Pandas to make the data processing easier.

The Data

We will be reusing the `co2.csv` and `ch4.csv` files from the previous lab.

Introducing two new datasets:

- `data/sf6.csv` (Sulphur Hexafluoride) and
- `data/n2o.csv` (Nitrous Oxide).

The Aim

We aim to estimate the combined global warming potential of all four gases.

We will produce a figure like the one below.

Obtaining the Jupyter Notebook

If you have cloned and pulled the module's GitHub repository, you should see,

```
materials/labs/
├── 040_working_with_pandas.ipynb
├-- ... etc
├── data
│   ├── ch4.csv
│   ├── co2.csv
│   ├── n2o.csv
│   ├── sf6.csv
│   ├── ... etc
```

The lab is `040_working_with_pandas.ipynb` and it will need the files `data/ch4.csv`, `data/co2.csv`, `data/n2o.csv`, and `data/sf6.csv`.

Alternatively, you can download the notebook and data via links on Blackboard.

Getting Help

- If you are stuck, just raise a hand to ask for help.
- Feel free to discuss the lab with your neighbours.
- Re-read the Pandas tutorial notes
  - In the Git repo at `tutorials/040_Introducing_Pandas.ipynb`
  - Or online at https://uos-com-6018.github.io/COM6018
- Use the Pandas documentation for reference, https://pandas.pydata.org/