

## Transforming and Visualizing Data with Python

Learn how to accelerate your data analyses using Pandas and Plotnine, two Python libraries specifically designed for transforming and visualizing small to medium-sized data sets. Together with JupyterLab it enables a convenient environment for interactive data analysis.

We're going to have a close look at Plotnine, a Python package based on the widely used R package ggplot2, that implements the so-called grammar of graphics. Its concise and consistent syntax allows you to create high-quality data visualisations in a quick and iterative manner that are suitable for both exploration and communication.

Pandas offers many features, and in one day, through a good balance of presentation and interactive exercises, we're going to cover the most important ones, including: importing, filtering, grouping, joining, exploring, and visualising data. By the end of this workshop, you'll understand the fundamentals of Pandas and be ready to perform your own analyses, You'll also have a solid understanding of the grammar of graphics and how to create data visualisations in Python for your daily work.

### What you'll learn

- Load data from text files, spreadsheets, databases, and APIs
- Performing advanced joins and merges

- Generating insightful pivot tables
- Transforming data between wide and long formats
- Master the fundamentals of data visualization and their use in the grammar of graphics
- Learn the kinds of visualisations that are applicable for each data type
- Explore bar charts, line plots, scatter plots, histograms, and boxplots

### This workshop is for you because

You have experience in Excel or R and want to learn about Pandas and the PyData ecosystem

You have programming experience in Python and want to start analysing data using pandas

You want to develop attractive and useful graphs and charts

You want to correctly use data visualization techniques to better explore your datasets

## Prerequisites

You're expected to have some experience with programming in Python. Roughly speaking, if you're familiar with the following Python syntax and concepts, then you'll be fine:

- assignment, arithmetic, boolean expression, tuple unpacking
- bool, int, float, list, tuple, dict, str, type casting
- in operator, indexing, slicing
- if, elif, else, for, while
- range(), len(), zip()
- def, (keyword) arguments, default values
- import, import as, from import ...
- lambda functions, list comprehension
- JupyterLab or Jupyter Notebook

## Schedule

### Day 1:

Essential data structures

- Numpy data types
- Numpy arrays
- Pandas Series
- Pandas DataFrame

- Pandas Index

Importing data

Manipulating data

- Selecting rows and columns
- Filtering rows
- Joining and concatenating
- Missing values, duplicates
- Converting data types
- Working with categorical data
- String manipulation

Exploring data

- Computing aggregate statistics
- Pivot tables
- Correlations

### Day 2:

Examples of data visualizations

- Histogram
- Densityplot
- Boxplot
- Bar chart

### A layered grammar of graphics

- Aesthetics
- Geometries
- Under the hood of layers
- Statistical transformations
- Scales
- Coordinate systems
- Facets
- Labels
- Themes

**Trainer:** Jeroen Janssen, PhD

**This course will be held in English.**

**Please note:**

This workshop cannot be credited as qualification measures at the Department of Mechanical Engineering. However, you are welcome to participate.