Visualizing data

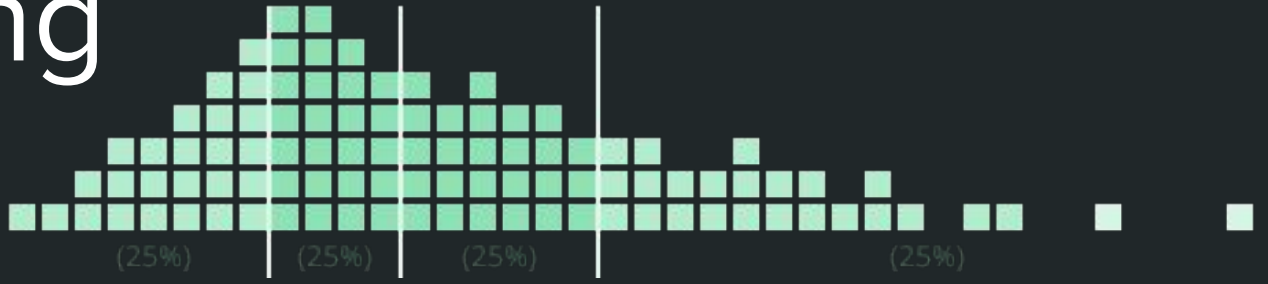


BILD 62

Objectives for today

- Load csv files into Python using NumPy
- Create and edit plots using `matplotlib`
- Dive into a notebook to plot our inflammation data

Reminder: we've started working with **modules** which we need to **`import`** into Python to use them

```
import numpy as np
```

full module name

nickname (and how we will refer to it in future code)

Module	Built-In	Description
csv	Yes	Aids in the reading, writing, and analysis of CSV files.
zipfile	Yes	Aids in the creation and extraction of compressed ZIP archive files.
matplotlib	No	Graphics library for plotting
plotly	No	A graphics library used for creating interactive plots for the web.
seaborn	No	A graphics library built on top of matplotlib with high-quality plots
pandas	No	A data processing library that specializes in data frames, which are analogous to spreadsheets.
scikit-learn	No	Contains basic tools for machine learning (i.e., helping to learn from data and make predictions).
numpy	No	Offers highly efficient data processing.
pygame	No	A game programming library that helps to build interactive, graphical games in Python.
django	No	Web development library that aids in designing websites and web applications.

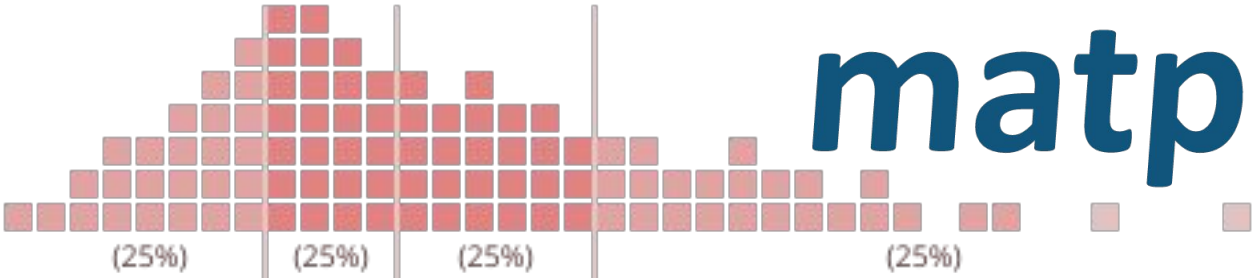
Common Python modules — ones we'll work with are highlighted

Working with data in Python

- In this class, we'll use NumPy to read **comma-separated values (csv)** and **tab-separated values (tsv)** files
  - You can tell by file extension!
  - Delimiter = separator
- Before plotting, we need to know the structure and type of our data
  - What is the data type?
  - Is it continuous or categorical?
- It can help to ***think through*** how you'd like to analyze and plot ***before*** **writing the code for it**

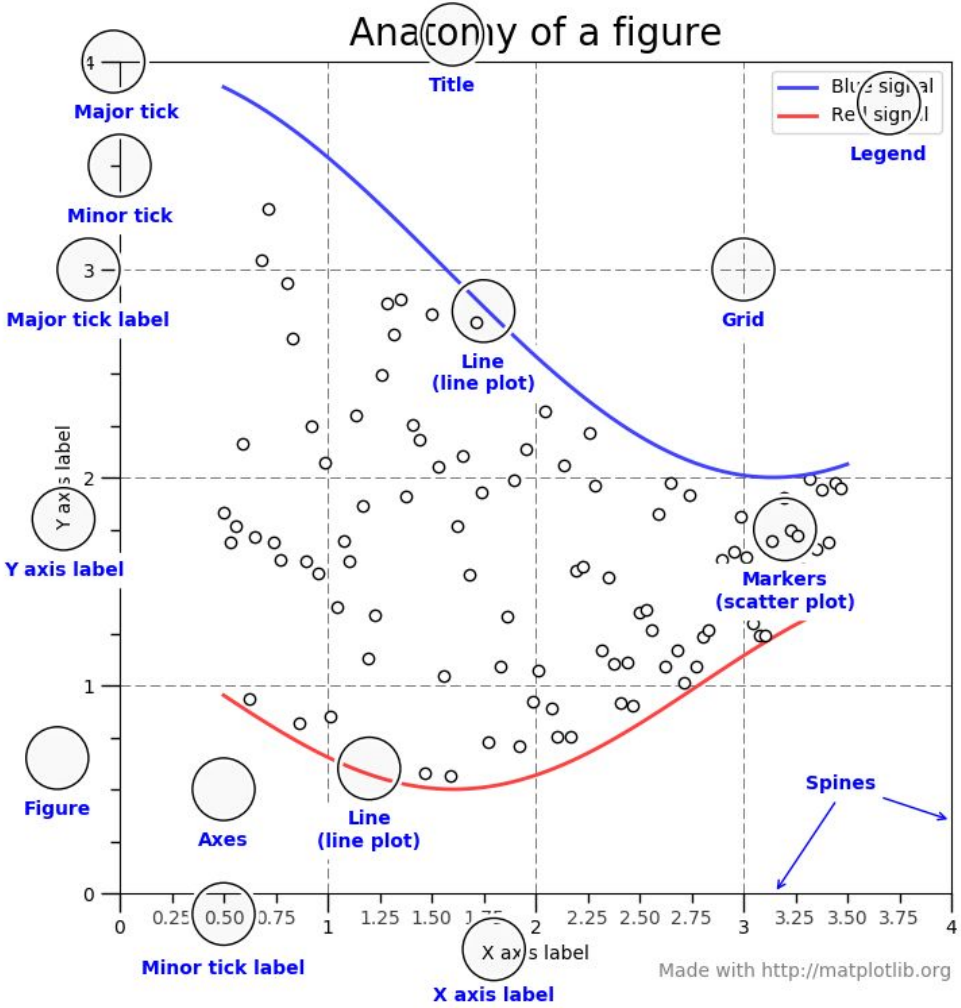
There are multiple ways to plot in Python

- Matplotlib (https://matplotlib.org/index.html)
  - Call to `pyplot` module
  - Through pandas (which uses pyplot)
- Seaborn (built on top of Matplotlib; https://seaborn.pydata.org/)
  - Loved by many #dataviz folks



There are (almost) endless things you can customize on your plot

... and once you write code to do so, you can reuse it!



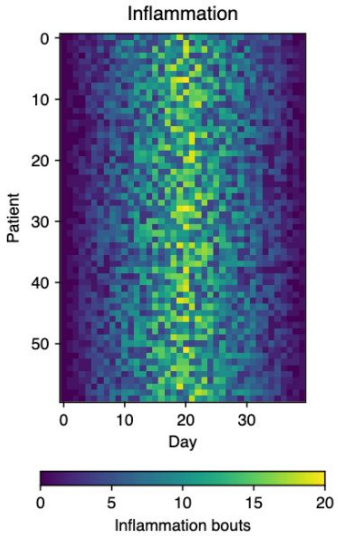
Objectives for today

- Load csv files into Python using NumPy
- Create and edit plots using `matplotlib`
- Dive into a notebook to plot our inflammation data

Inflammation data

Patients	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
	0	0	1	3	1	2	4
	0	1	2	1	2	1	3
	0	1	1	3	3	2	6
	0	0	2	0	4	2	2
	0	1	1	3	3	1	3

Analysis



Conclusion

How does the medication affect patients?

Today, visualizing our data as **line charts** and a **heatmap** will help us explore trends in the data

Resources

Barry Grant's visualizing data lecture:
https://www.youtube.com/watch?v=WfvBhFUzTQs

PyPlot tutorial https://matplotlib.org/stable/tutorials/pyplot.html

Matplotlib tutorials https://matplotlib.org/stable/tutorials/index.html

Top 50 Matplotlib Data Visualizations

Towards Data Science: Python Plotting Basics