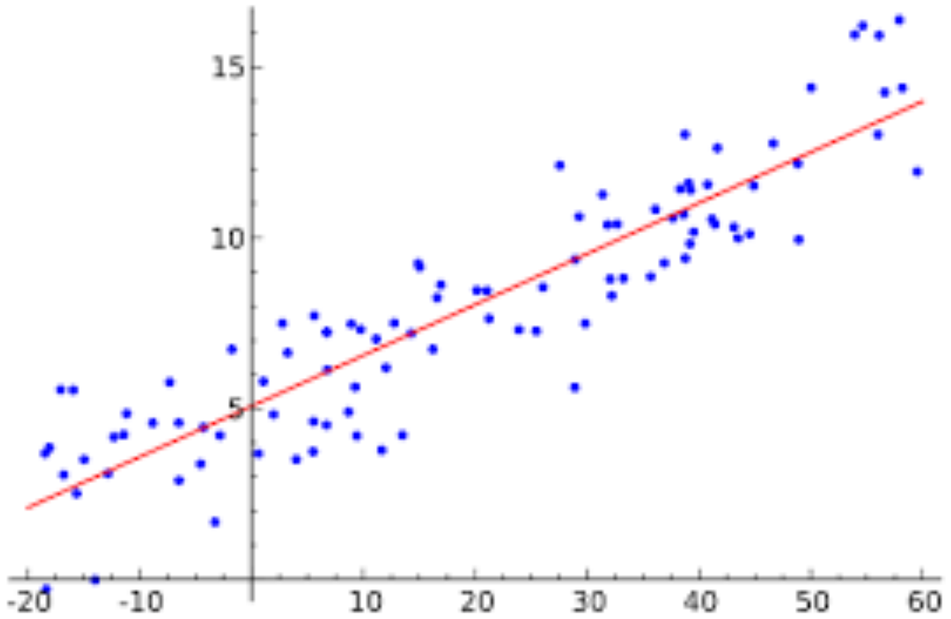
Data visualisation

TU Delft

Why create visuals?

- Visuals provide a better way for observation
- Explore data
- Discover patterns
- Discover trends
- Story-telling



Value of visualisation

- **Exploratory visualization**
  - Develop and assess hypotheses
  - Find patterns / Discover errors in data
  - Identify trends and clusters, spotting local patterns, evaluating modeling output

- **Explanatory Visualization**
  - Share and Present Results
  - Stimulate Research, Collaborations and Ideas

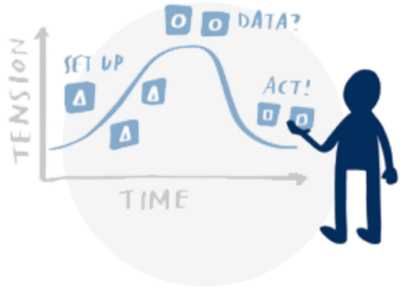
Ideas to story

understand the
context

choose an
effective visual

eliminate
clutter

focus
attention



tell a
story

TU Delft

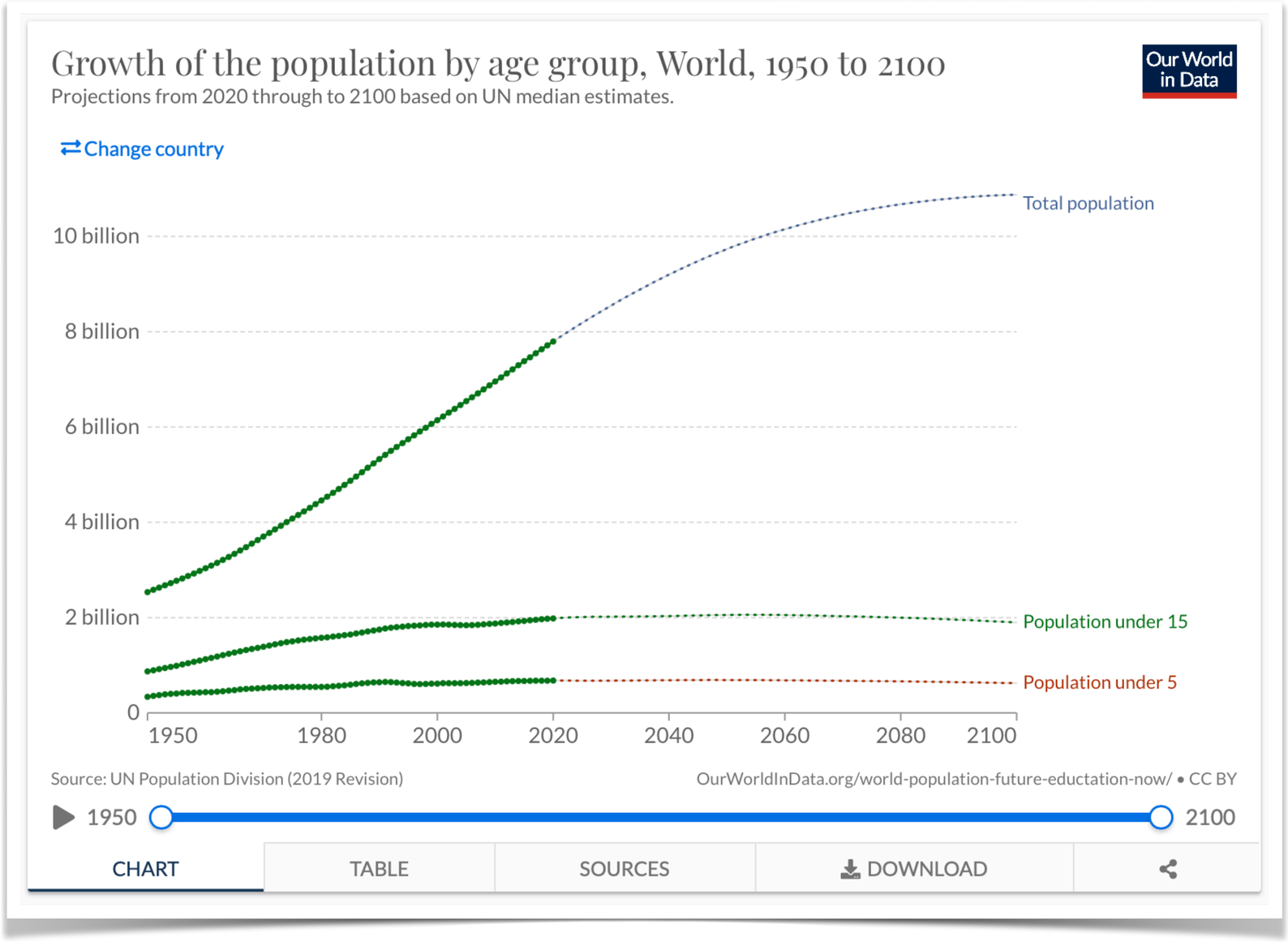
Lab
Artificial Intelligence
for Mobility

Data stories

- New York Times
- Nature/Lancet articles
- Wealth to scale
- Medium articles

Visualisation

- Static
- Dynamic
- Interactive



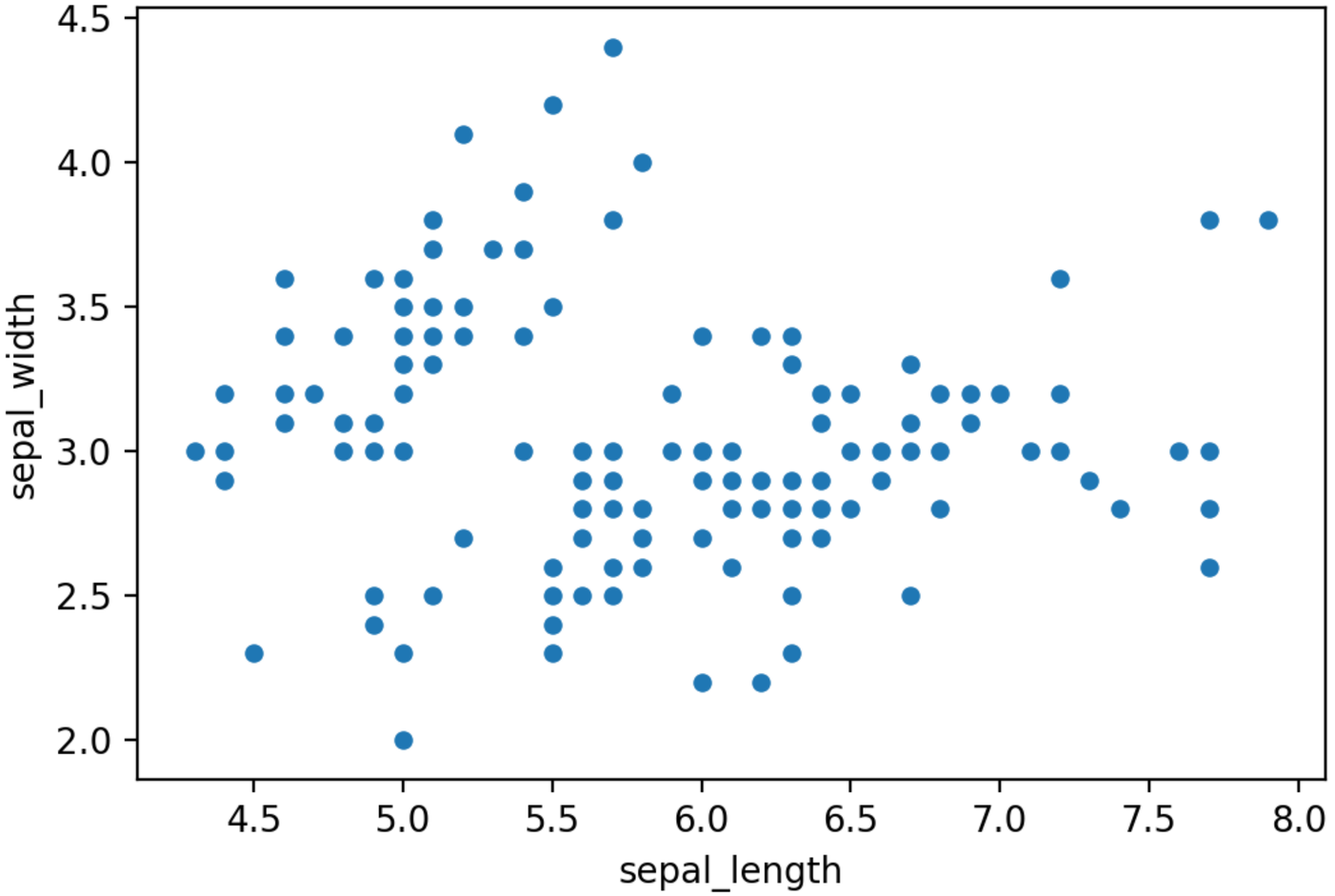
Which visualisation to use?

- Story
- Audience
- Size of the data
- Data types
- Functionality and relationship between data elements

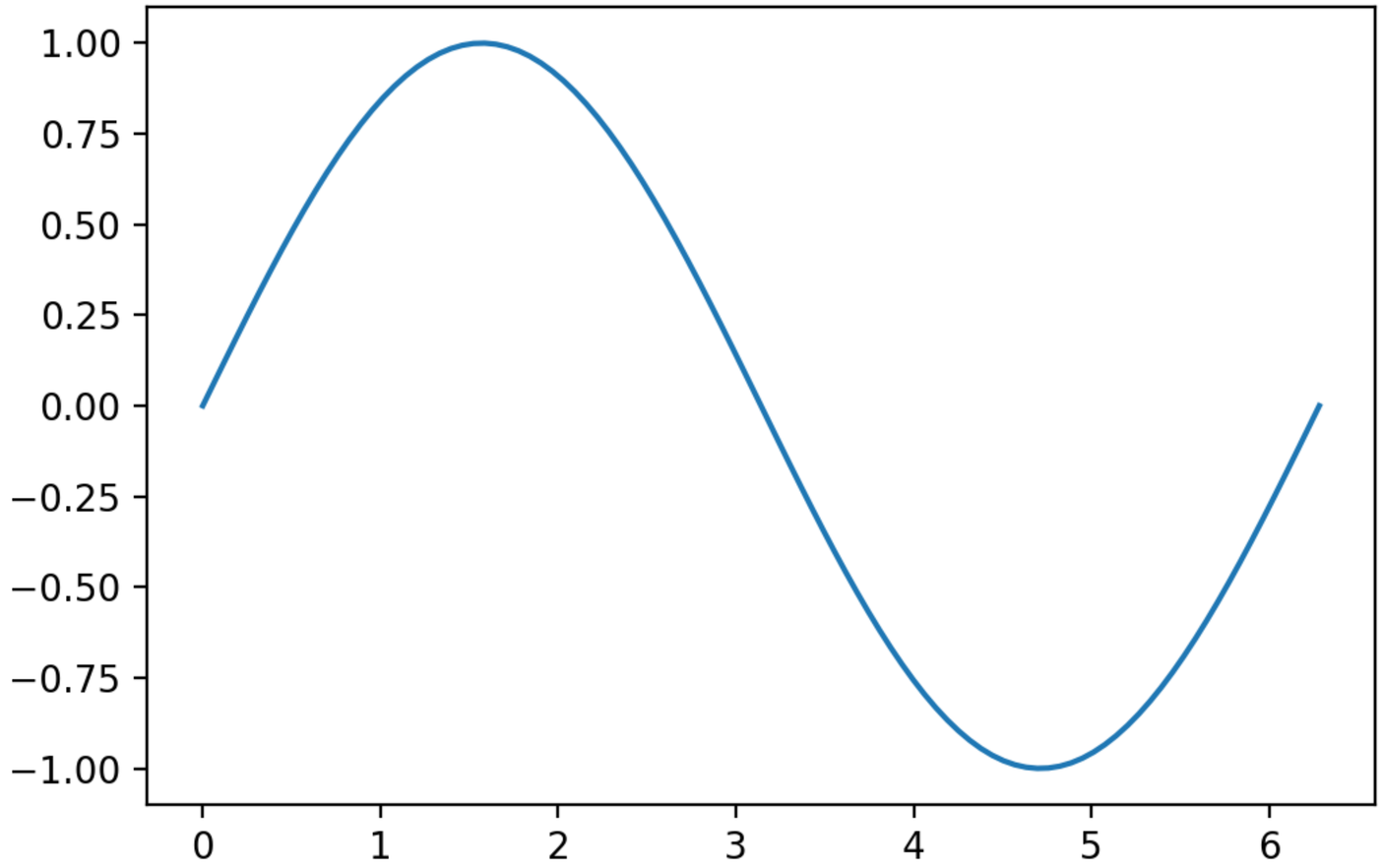
Data visualisation catalog

Basic plots

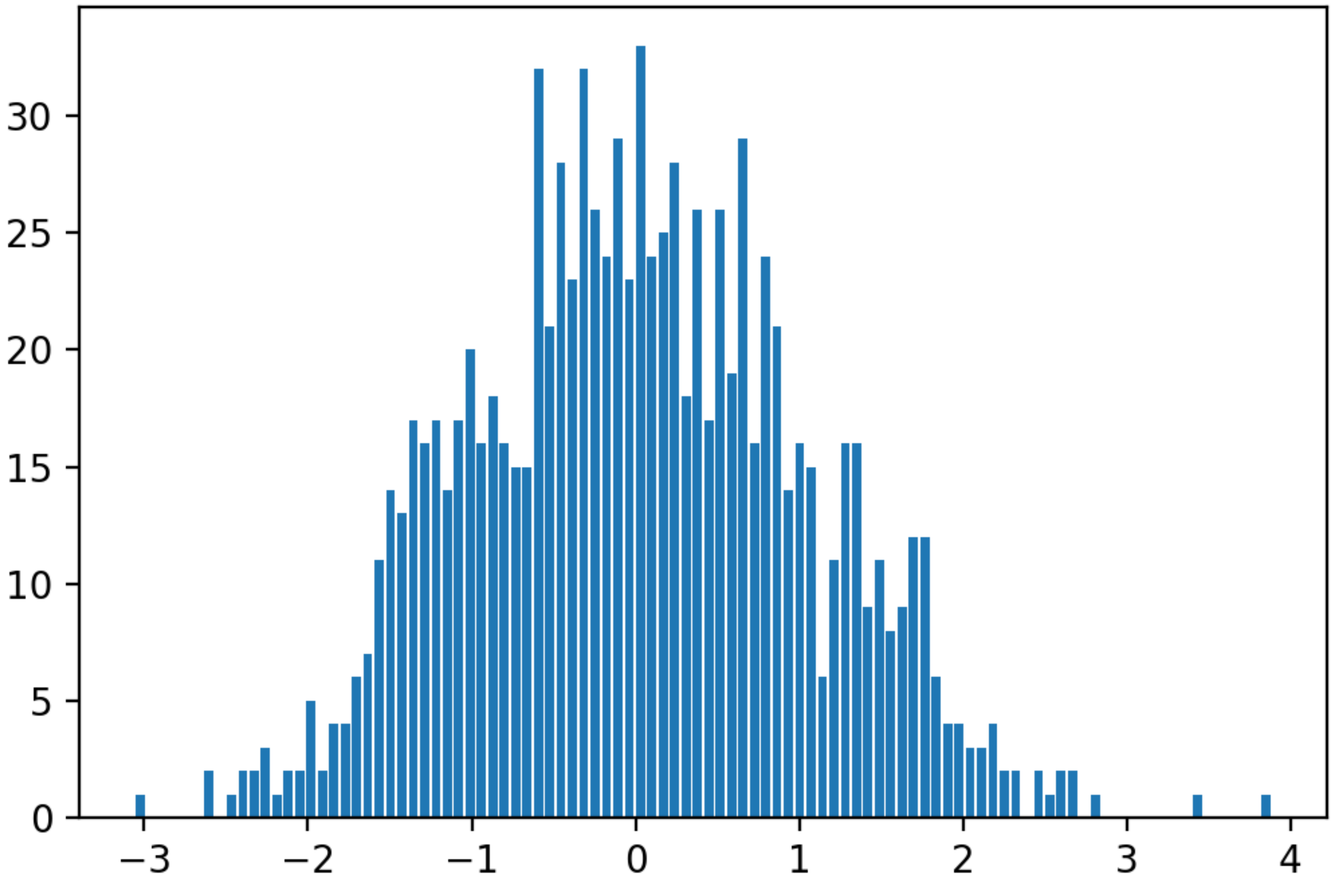
Scatter plot



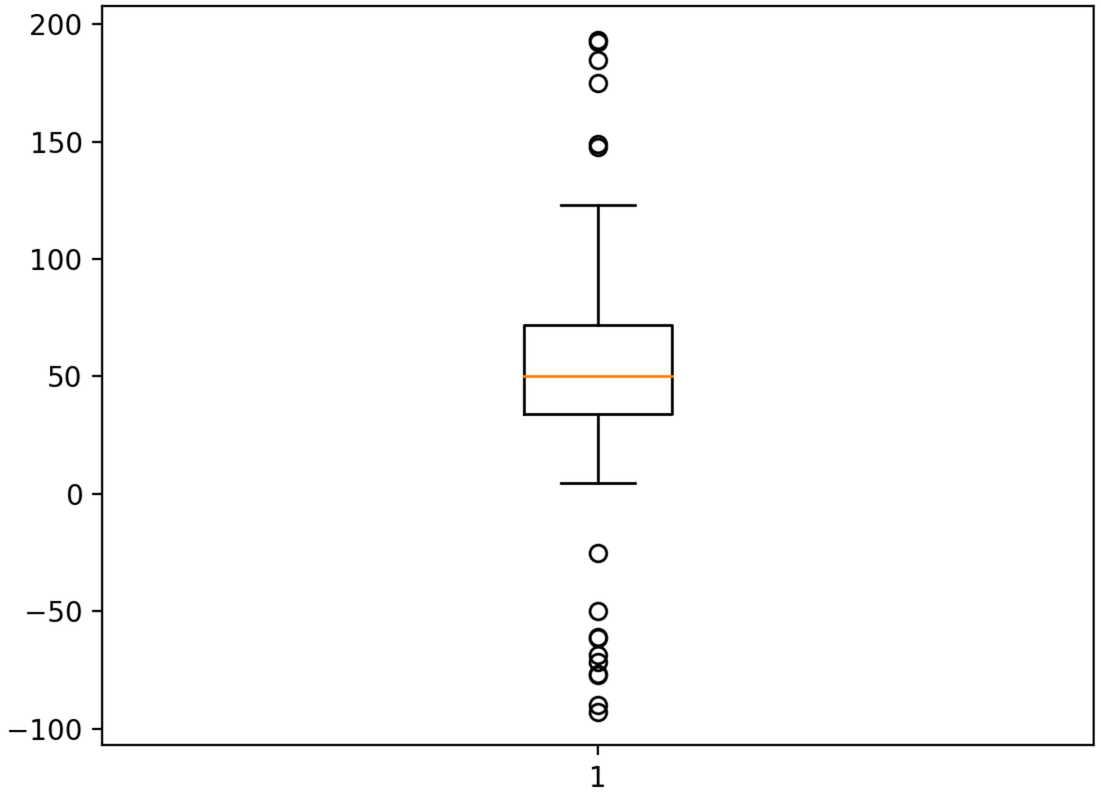
Line plot



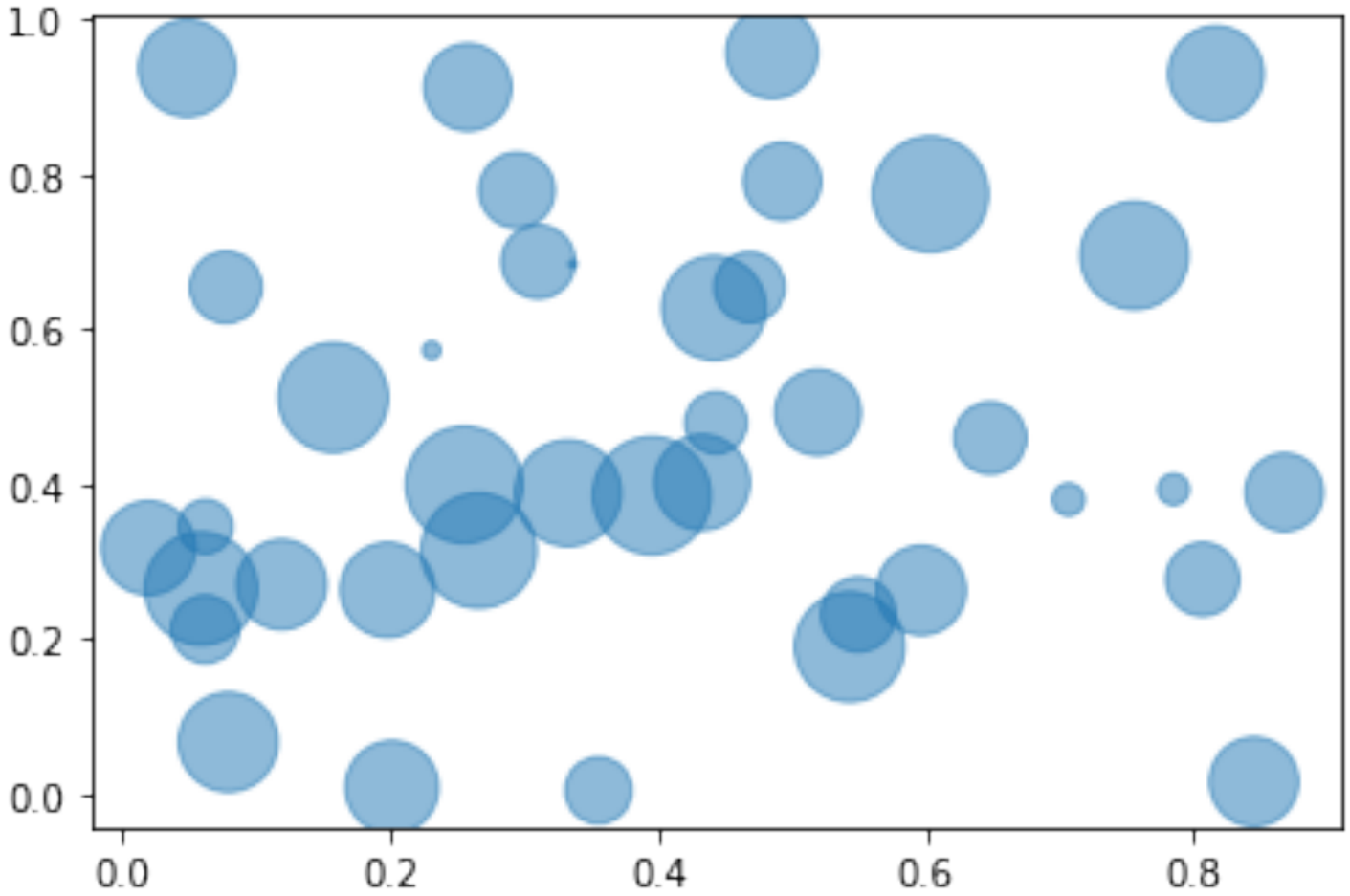
Histogram



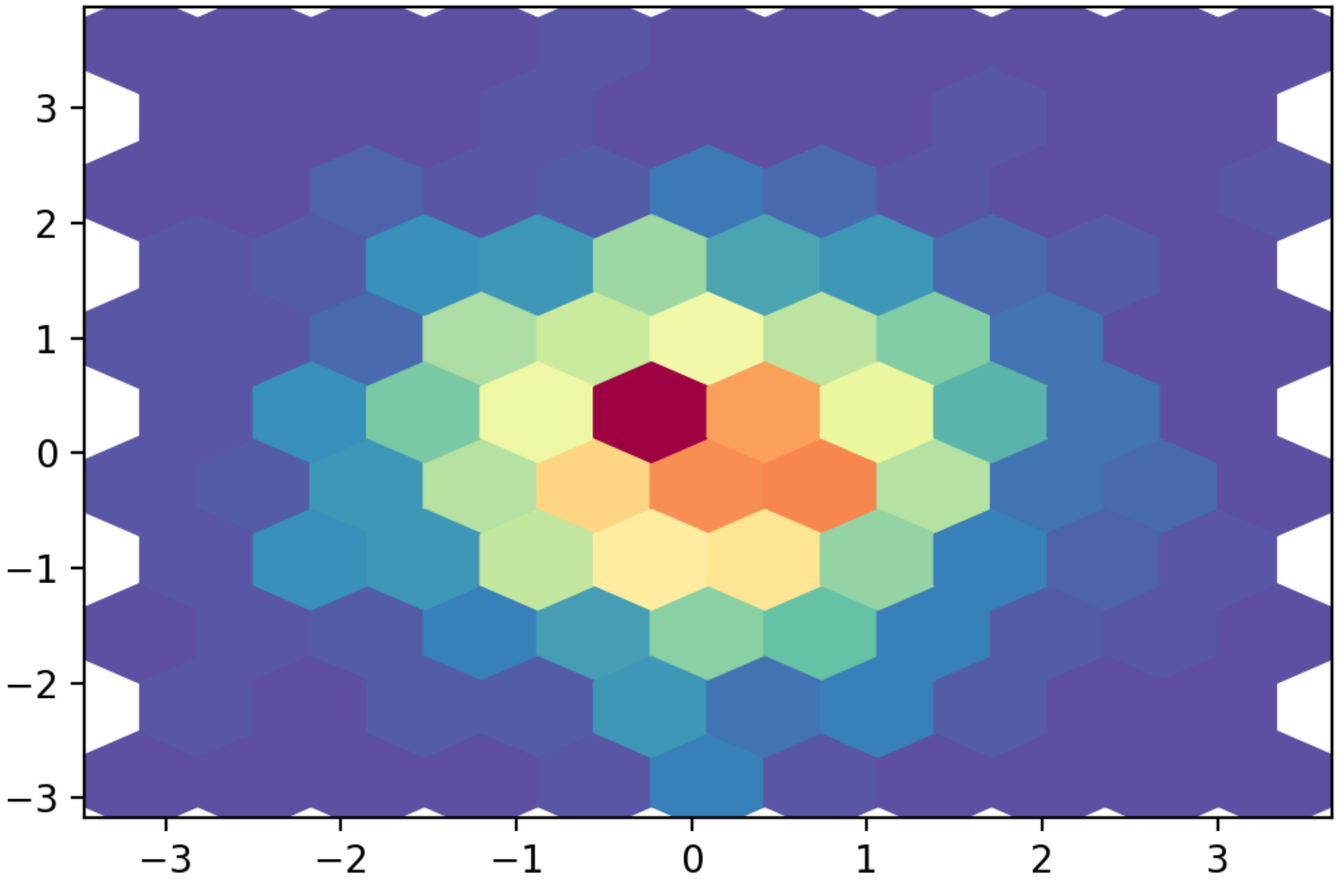
Boxplot



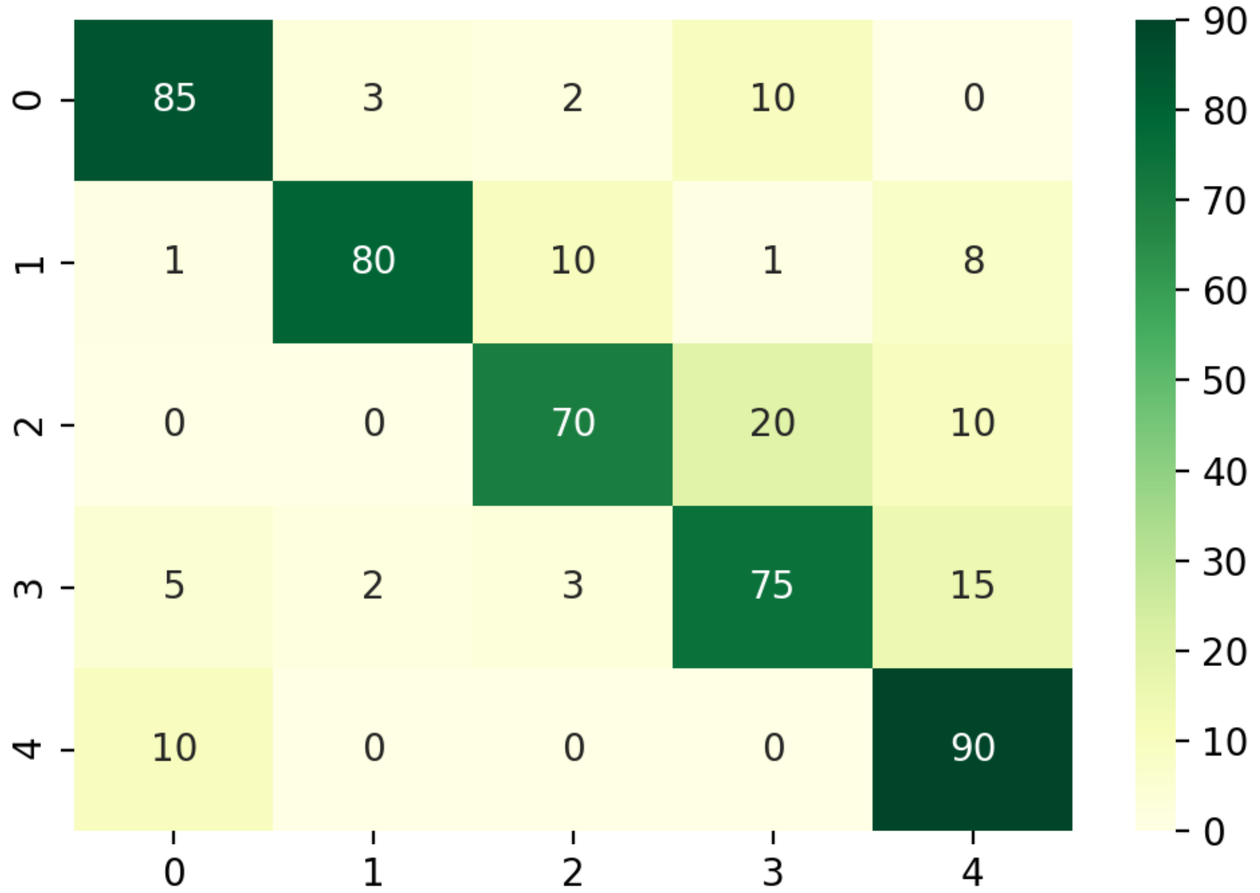
Bubble chart



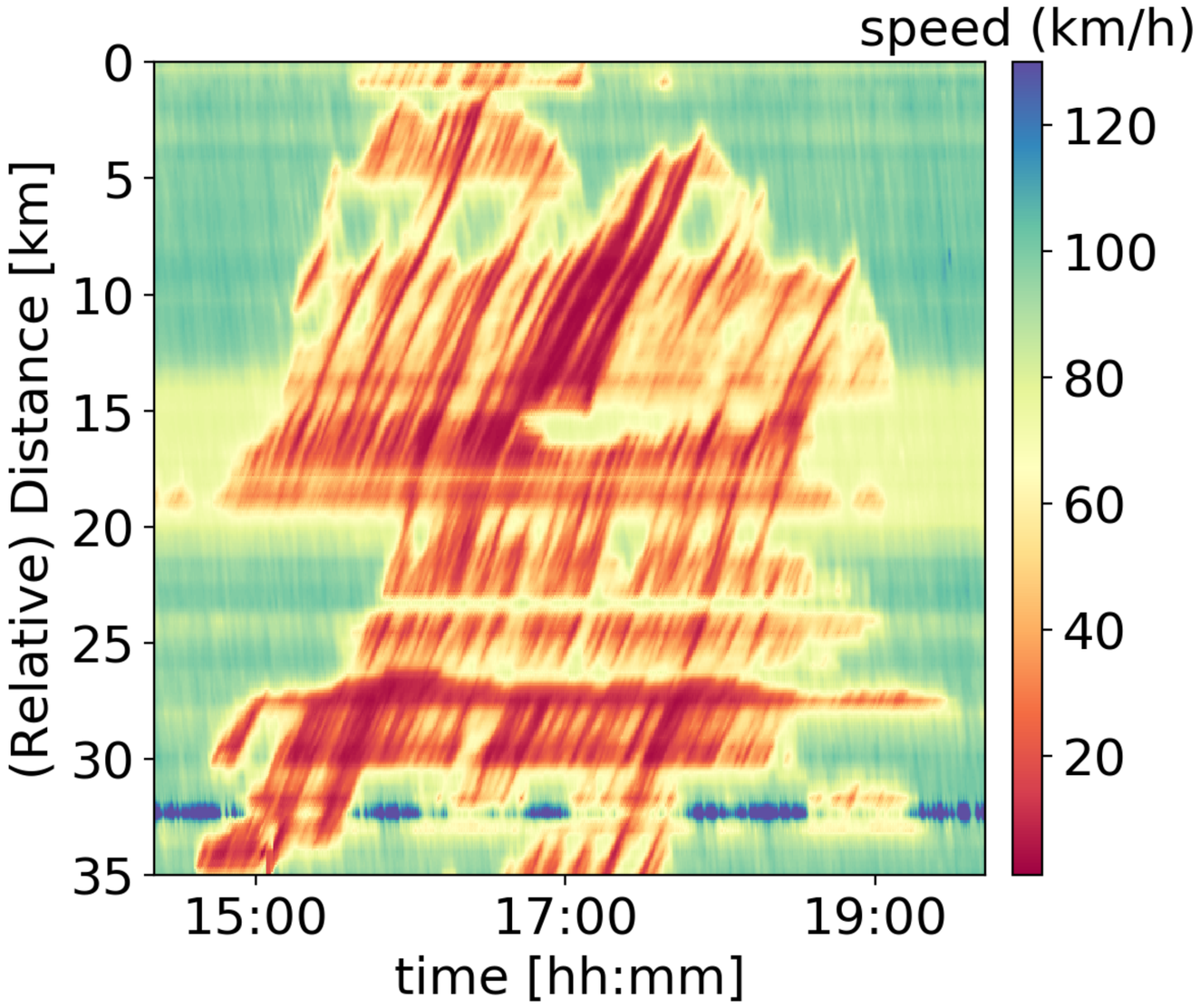
Hexbin



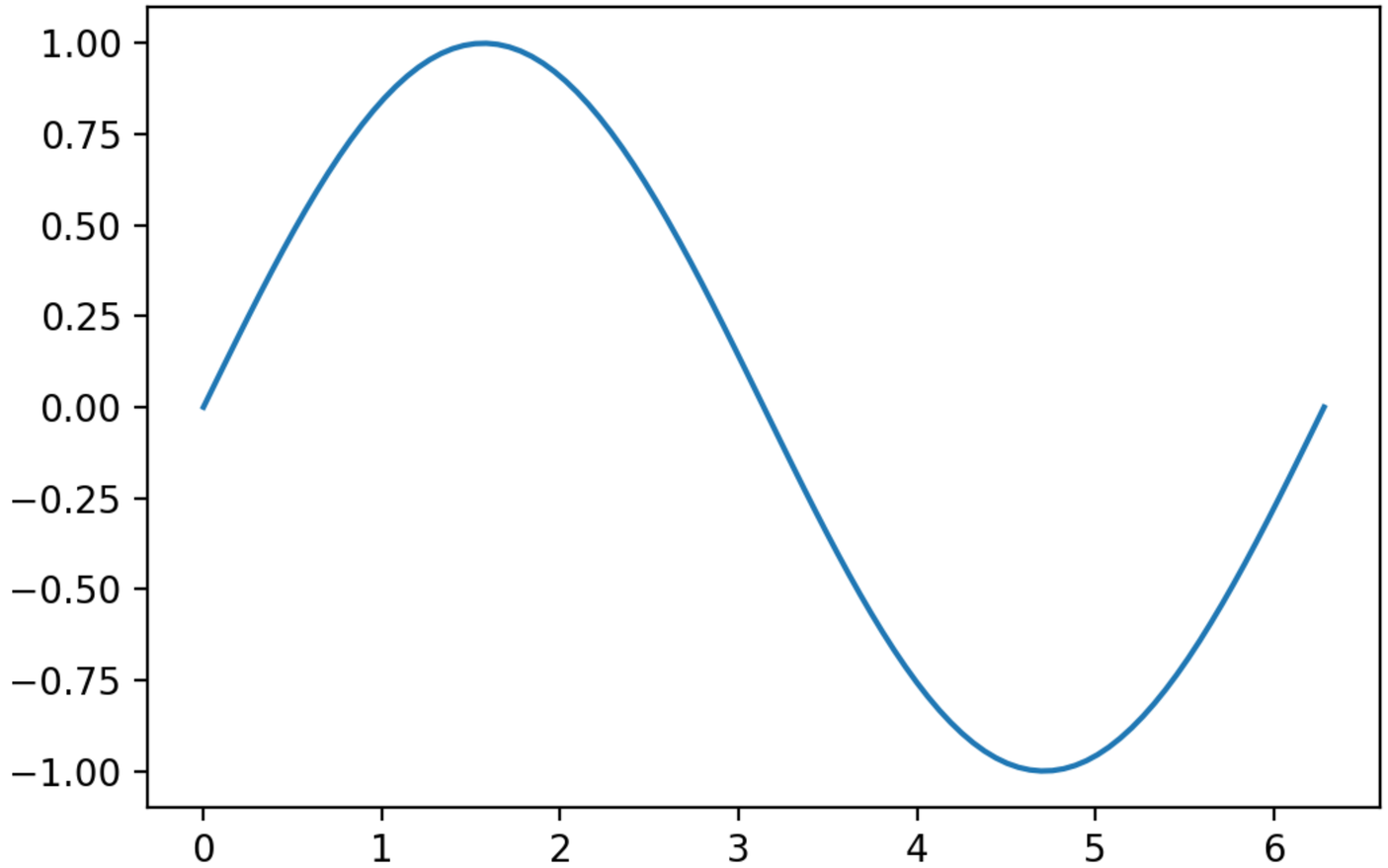
Heatmap


0
1
2
3
4
85
3
2
10
0
1
80
10
1
8
0
0
70
20
10
5
2
3
75
15
10
0
0
0
90
0
1
2
3
4
90
80
70
60
50
40
30
20
10
0

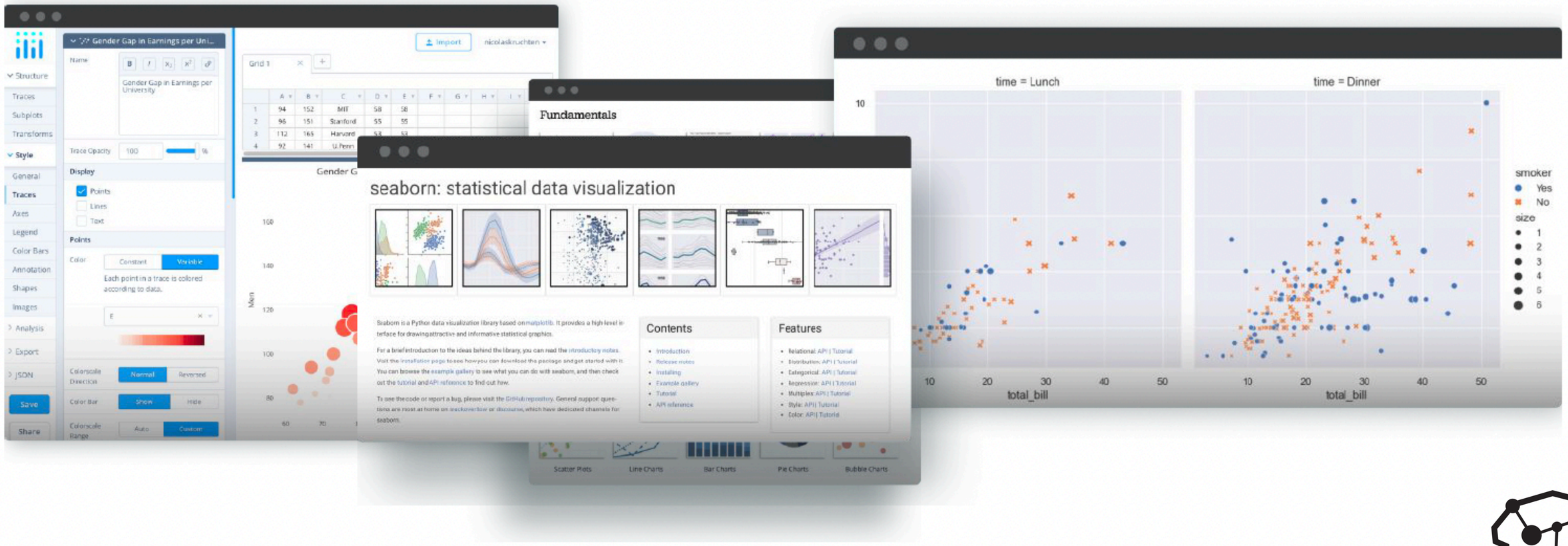
Image


speed (km/h)
120
100
80
60
40
20
0
5
10
15
20
25
30
35
(Relative) Distance [km]
15:00
17:00
19:00
time [hh:mm]

What is wrong with this plot?



Which Python library to use?



TU Delft

Lab
Artificial Intelligence for Mobility

Data visualisation with python

- Matplotlib
- Seaborn
- Plotly
- etc

Matplotlib

Python Libraries

- **Matplotlib**
  - It is the first library for Python data visualization. Most of the other libraries took Matplotlib as their base. Some libraries exist only to extend the functionality of Matplotlib and work together with less code.
  - Versatile
  - Easy to see the property of the information
  - Aesthetics need to be improved
  - Low-level and less user-friendly

TU Delft

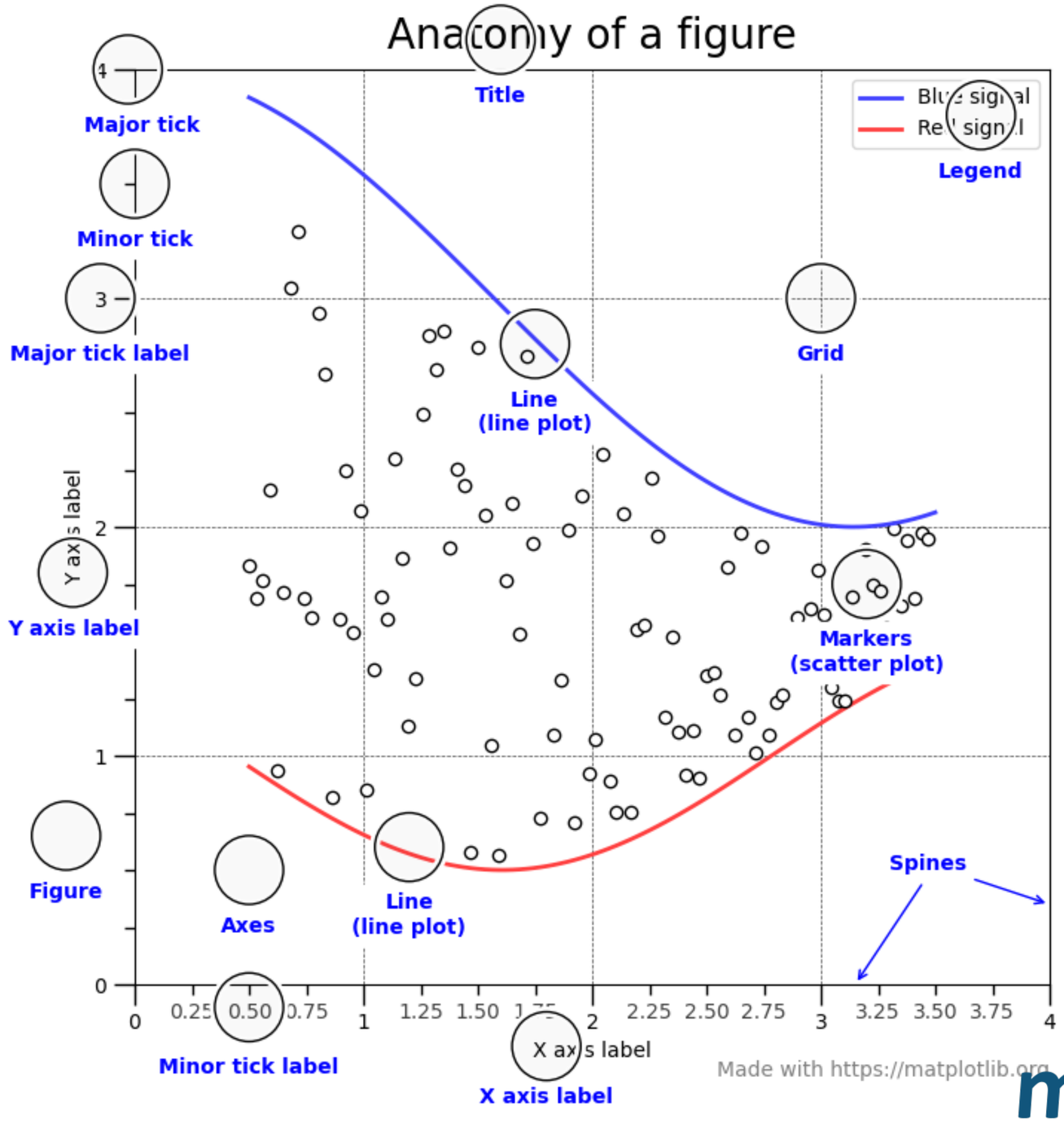
Lab
Artificial Intelligence for Mobility

Matplotlib

- The most fundamental visualisation library in Python
- The base for other ‘advanced’, ‘more abstract’ visualisation libraries
- Make plots quickly
- Pros:
  - Full controls to low-level configurations of plots
- Cons:
  - Take time and efforts to generate ‘good’ plots

Matplotlib

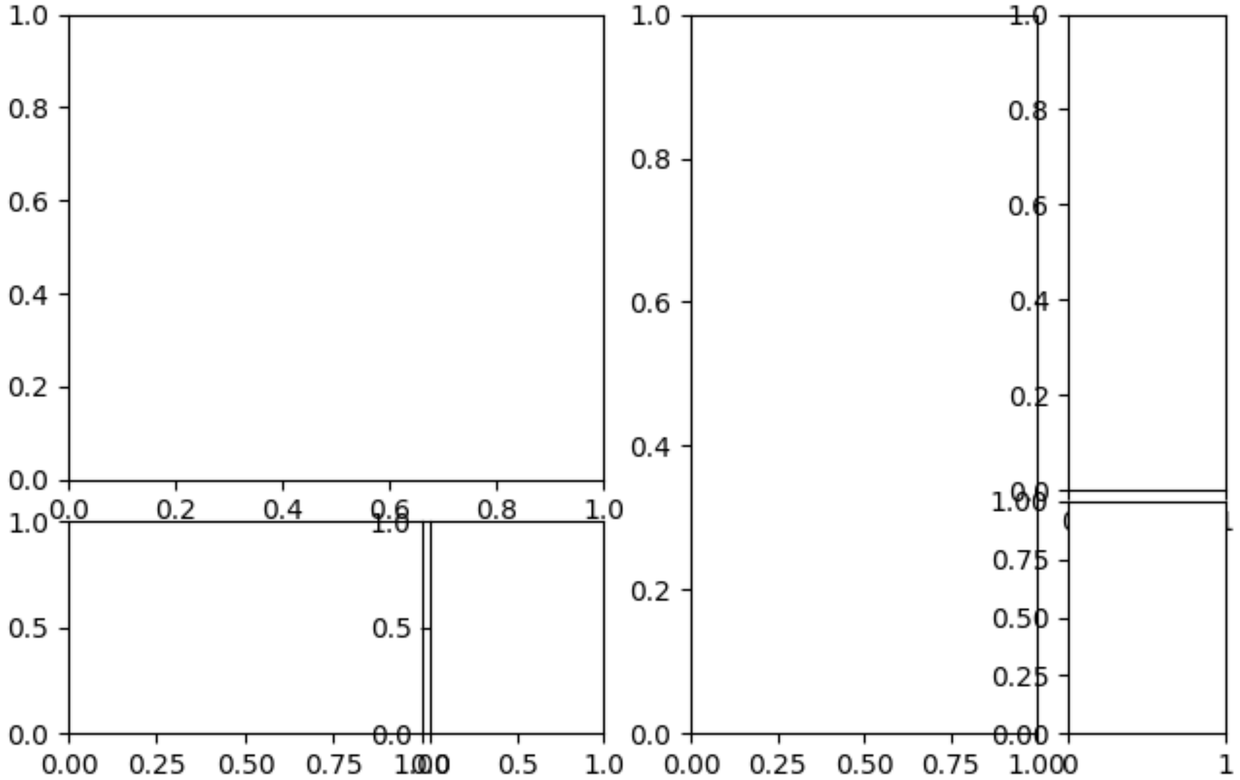
Figure anatomy



Matplotlib

Subplots

- Multiple *sub*-plots in a single figure
- Customisable layouts



matplotlib

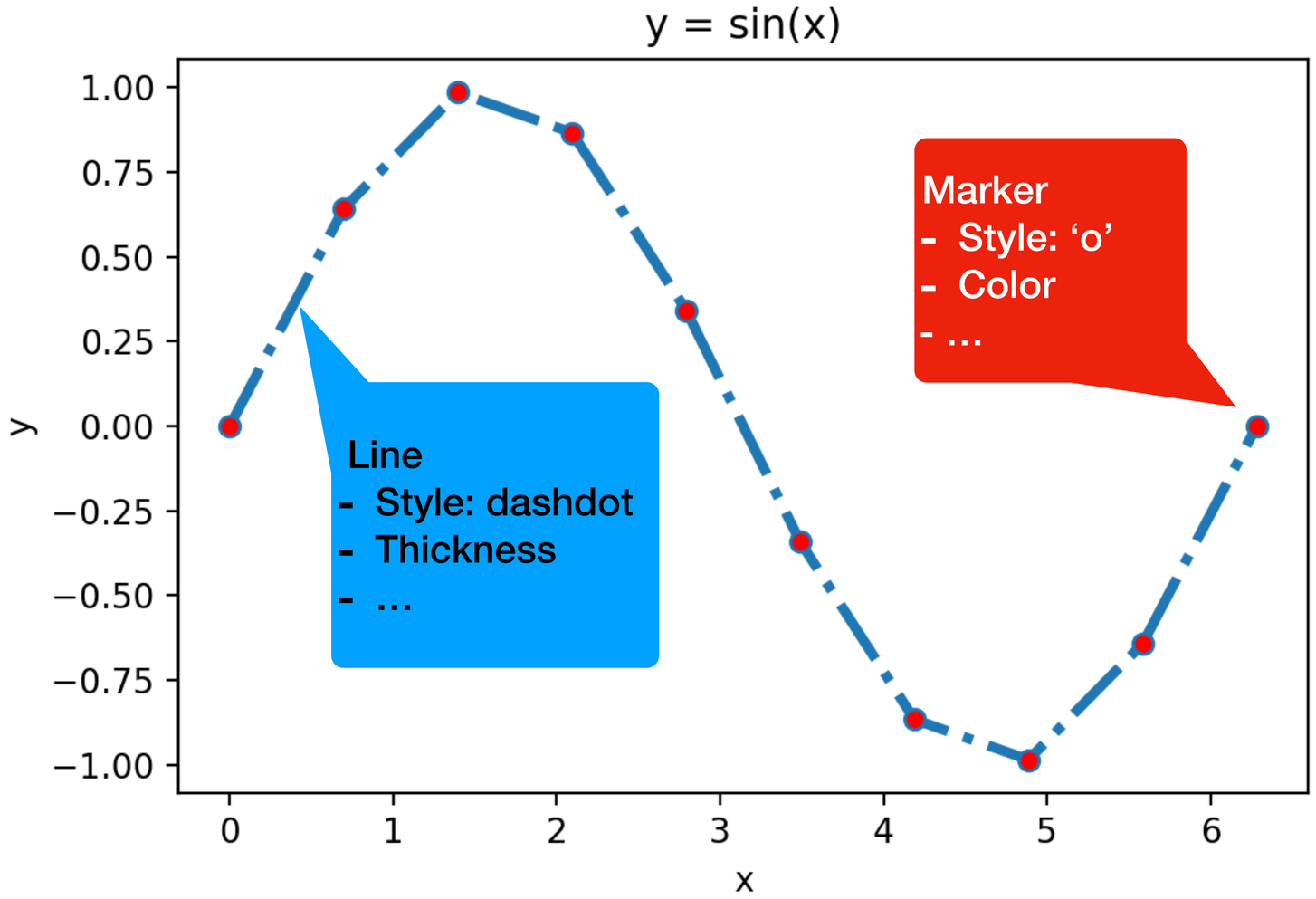
Line plot

Properties

figure	Figure
fillstyle	{'full', 'left', 'right', 'bottom', 'top', 'none'}
gid	str
in_layout	bool
label	object
linestyle or ls	{'-', '--', '-.', ':', '', (offset, on-off-seq), ...}
linewidth or lw	float
marker	marker style string, Path or MarkerStyle
markeredgecolor or mec	color
markeredgewidth or mew	float
markerfacecolor or mfc	color
markerfacecoloralt or mfcalt	color
markersize or ms	float
markevery	None or int or (int, int) or slice or list[int] or float or (float, float) or list[bool]
path_effects	AbstractPathEffect
picker	float or callable[[Artist, Event], tuple[bool, dict]]
pickradius	float
rasterized	bool
sketch_params	(scale: float, length: float, randomness: float)
snap	bool or None
solid_capstyle	CapStyle or {'butt', 'projecting', 'round'}
solid_joinstyle	JoinStyle or {'miter', 'round', 'bevel'}

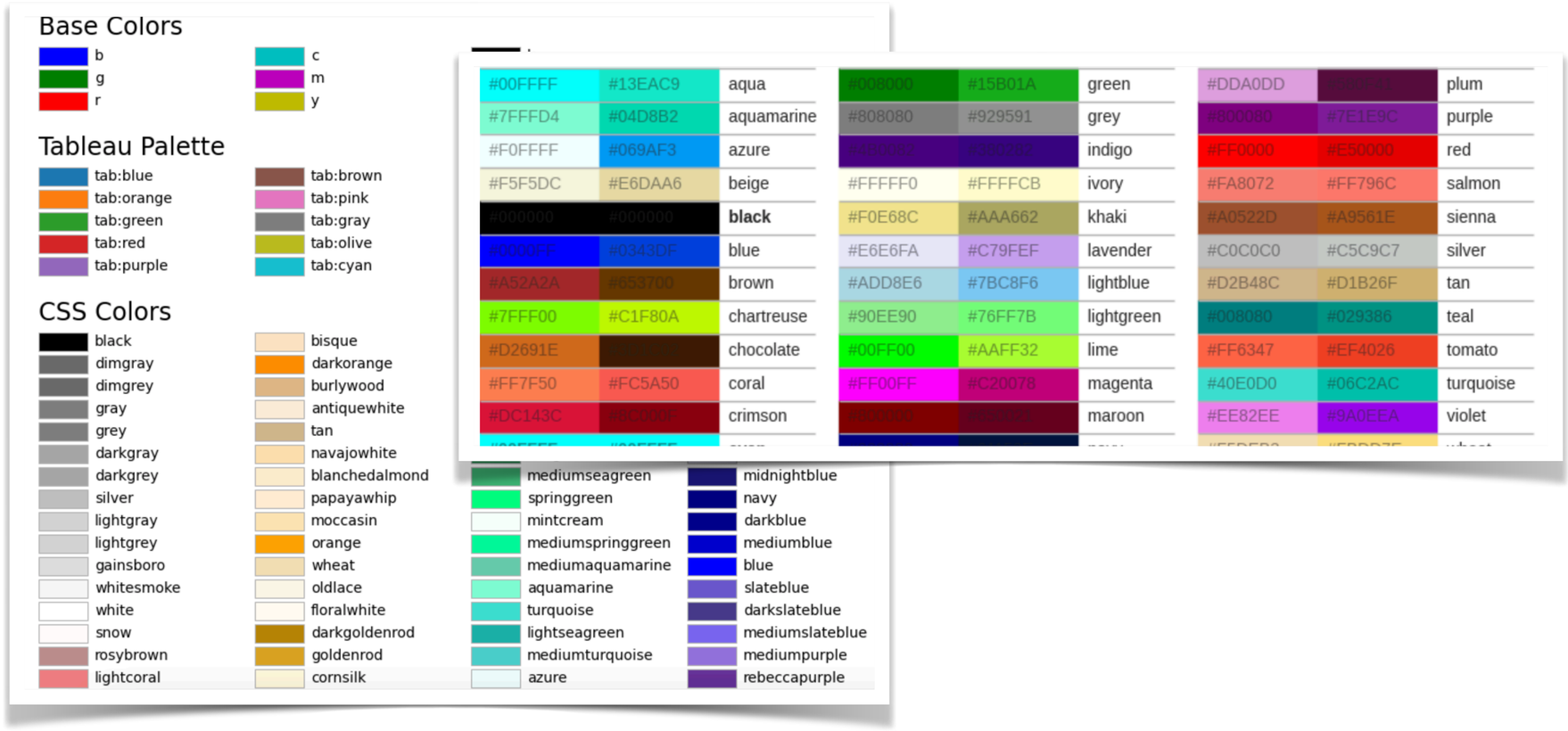
Line plot

Properties



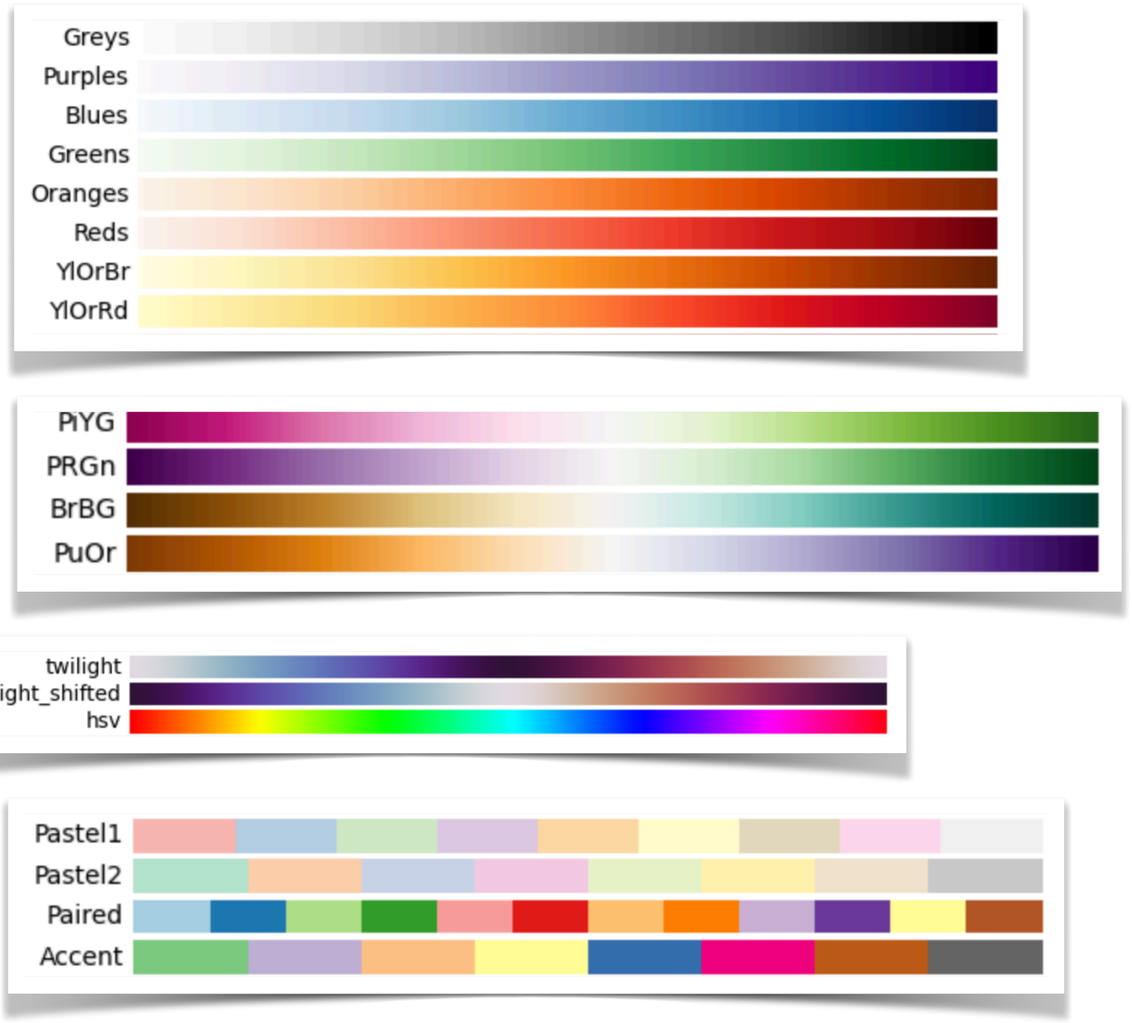
Colors

- Named colors
- Hex color codes
- RGB tuples
  - (0.1, 0.2, 0.5)



Colormap

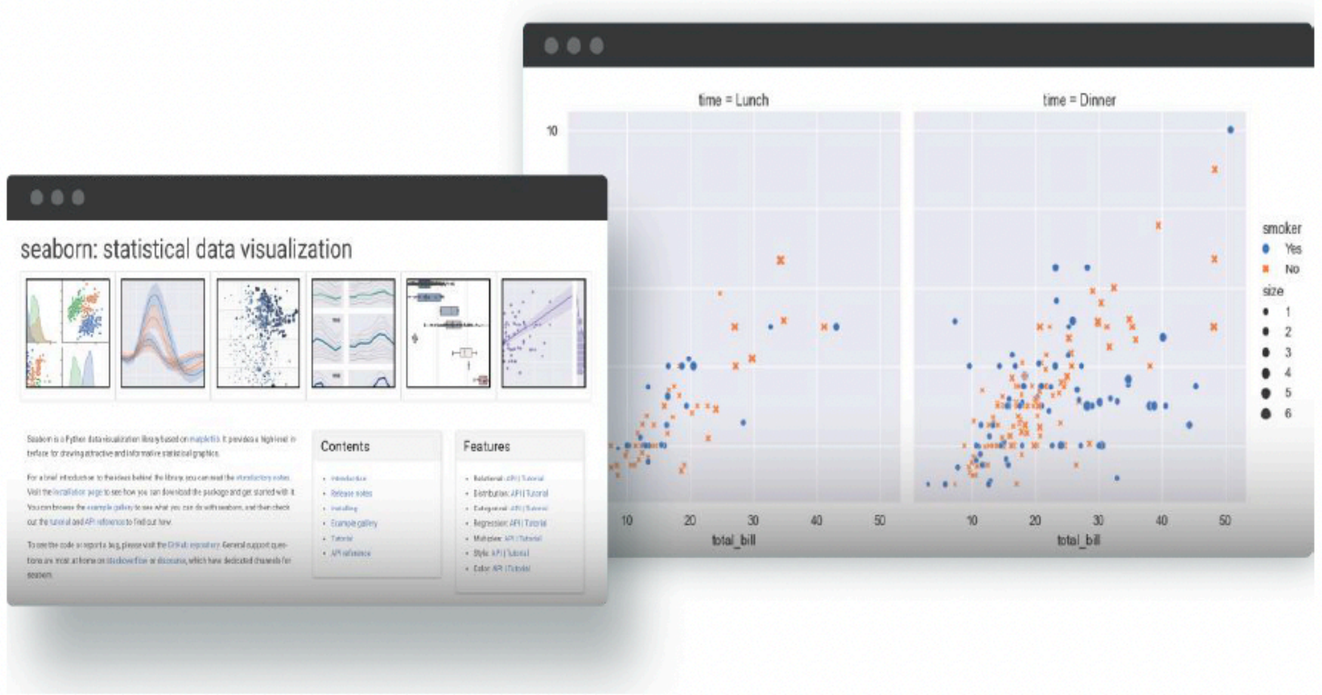
- Sequential
- Diverging
- Cyclic
- Qualitative



Seaborn

Python Libraries

- **Seaborn**
  - Seaborn is a library for visualizing data arrays based on a Matplotlib python plot package.
  - Very easy to create individual graphs and heat maps.
  - Beautifully presents processed data.
  - Less Code
  - Less Extensive Collection
  - Better aesthetics

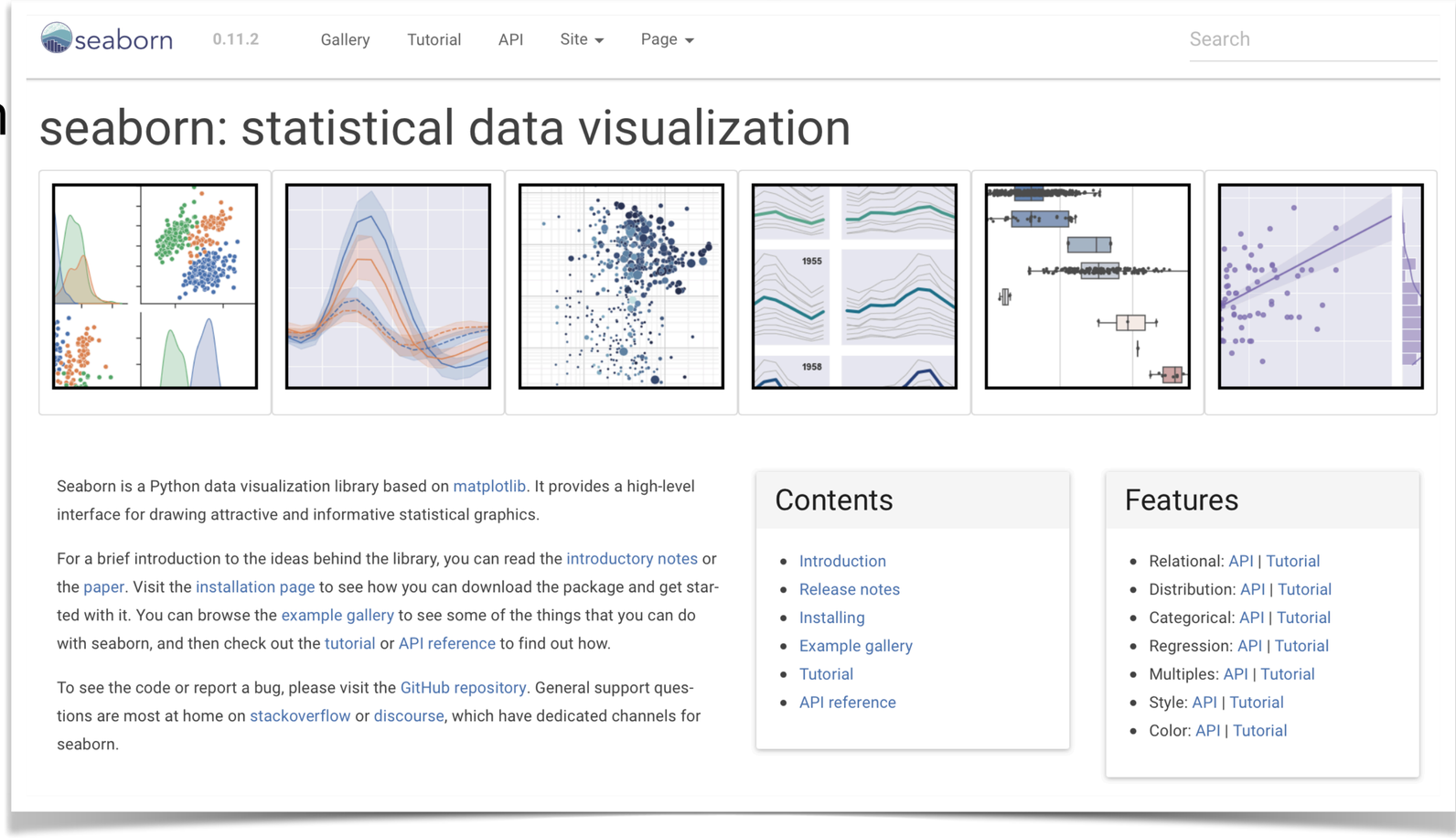


TU Delft

Lab
Artificial Intelligence for Mobility

Seaborn

- Statistical data visualisation
- Based on matplotlib
- High-level interface

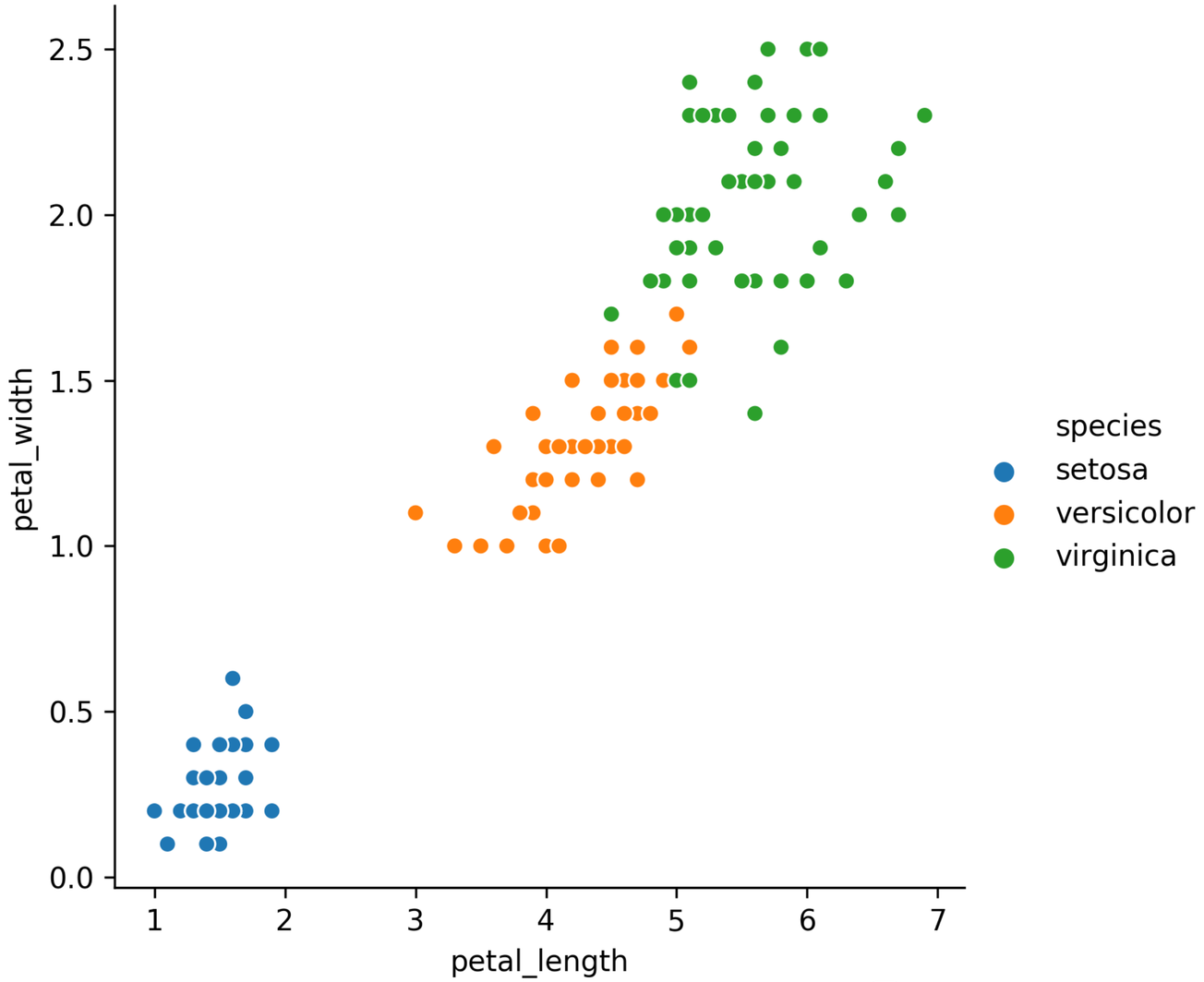


seaborn

Seaborn

Statistical relationships

- Scatter plot
- Line plot + confidence-interval
- Joint plot

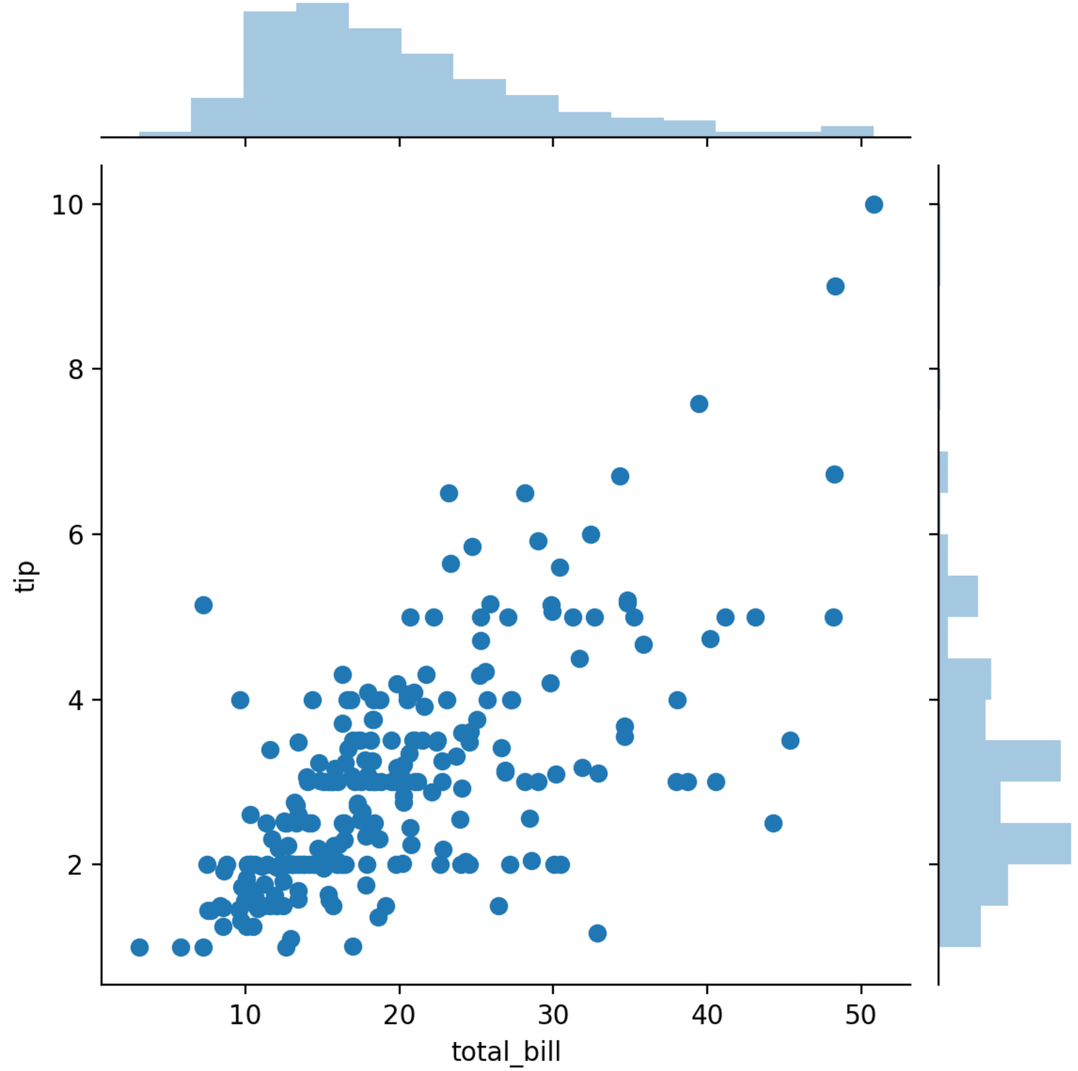


seaborn

Seaborn

Statistical relationships

- Scatter plot
- Line plot + confidence-interval
- Joint plot

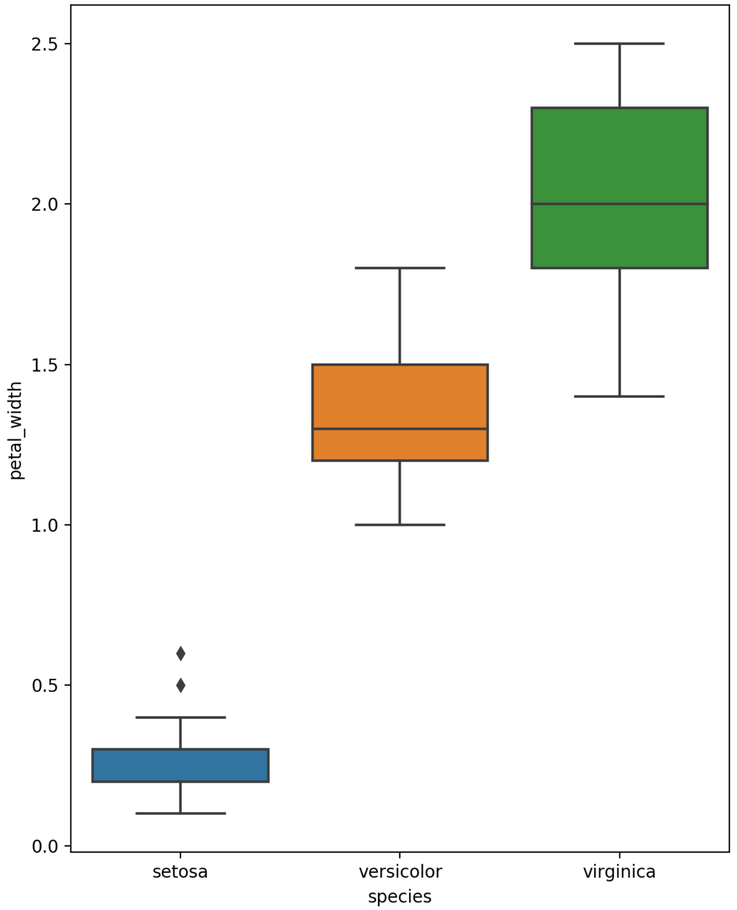


seaborn

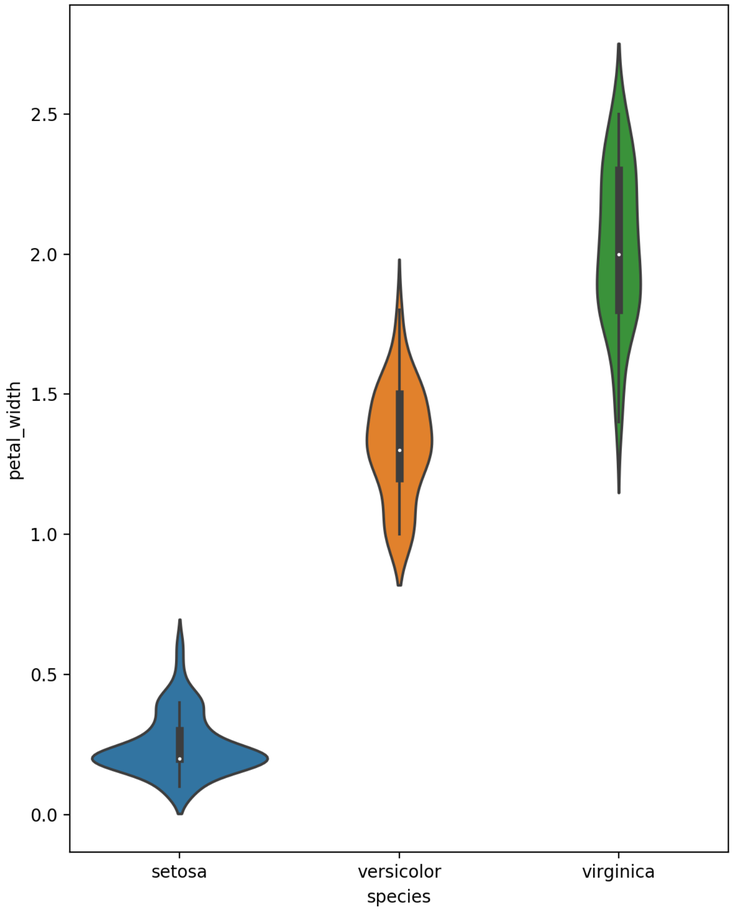
Seaborn

Categorial data

Boxplot



Violinplot

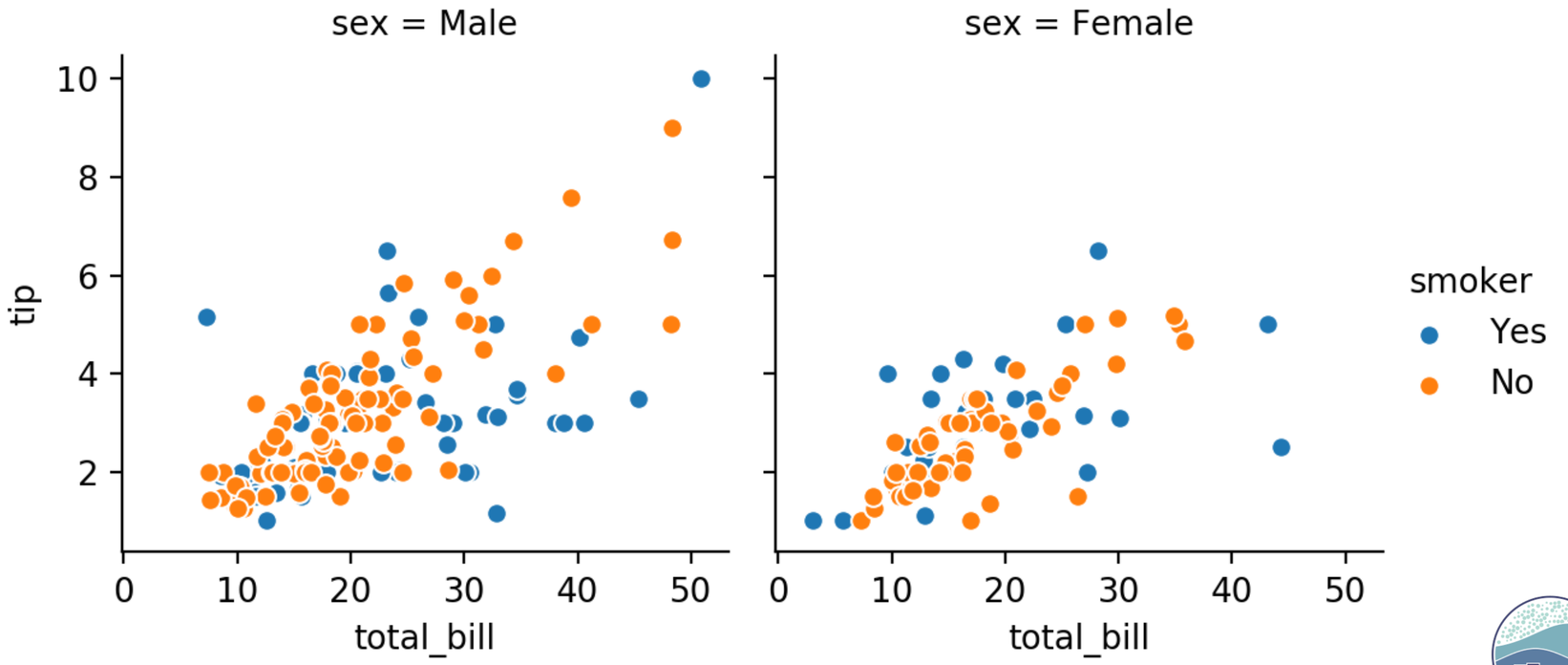


seaborn

Seaborn

Multi-plot grids

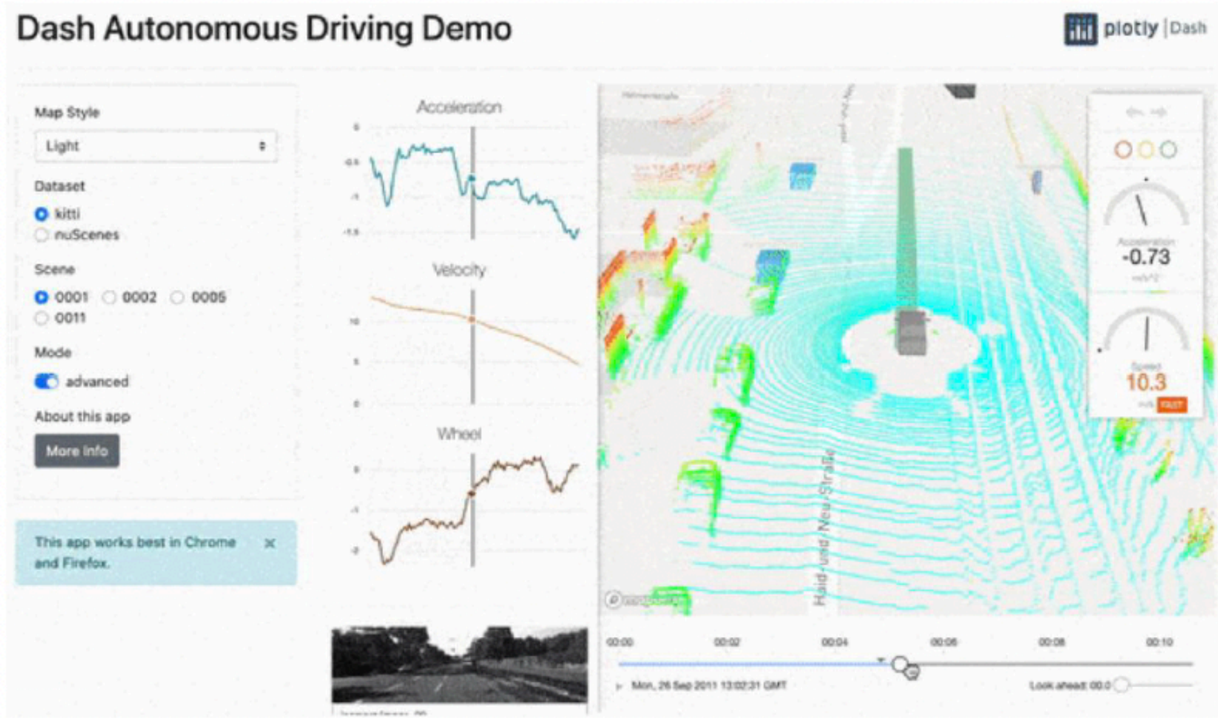
- Repeat the same plot on different subsets of dataset



seaborn

Python Libraries

- **Plotly**
  - Simple to create interactive plots
  - Easily create plots that are usually difficult to develop.
  - Plotly is the perfect tool for creating interactive plots with just a few lines of code.



Exam

- 4 types of questions - True/False, multiple choice, multi-select, coding
- 50 points in total
- Digital exam, nothing is allowed other than scrap paper for calculations (if needed)
- 3 hours