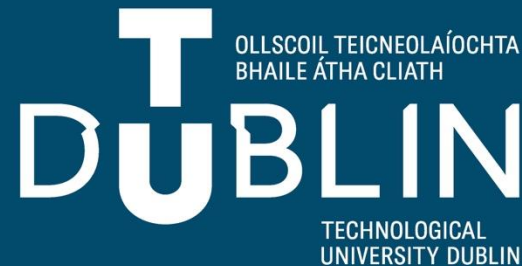# Programming for Analytics

Lecture 7: Data Visualisation

Bojan Božić
School of Computer Science
TU Dublin, Grangegorman

bojan.bozic@tudublin.ie

Overview

- Why Visualise Data?

- Charts

- Matplotlib

- Line Plot, Bar Chart, Histogram, Scatter Plot, Box Plot

- Choosing the Right Chart

Why Visualise Data?

- Communicate insights clearly

- Detect patterns, outliers, trends

- Help decision-making through clarity

- Essential part of exploratory data analysis (EDA(

Types of Charts

- Line chart – trends over time

- Bar chart – comparisons

- Histogram – distribution

- Scatter plot – relationships

- Box plot – summary statistics

Activity: Chart Matching

- Which charts would be suitable for the following data:
  - Heart rate during a run   ➜ Line Chart
  - All star players per NBA team (historical)

    ➜ Bar chart
  - Age of survey participants   ➜ Histogram
  - Study hours vs grades   ➜ Scatter plot
  - House prices per region   ➜ Box plot

Introduction to matplotlib

- `matplotlib.pyplot` is the main plotting module

- Import using

`import matplotlib.pyplot as plt`

- Use `%matplotlib inline` in notebooks

Basic Line Plot

- `plt.plot(x, y)` for a line chart
- Use `plt.title()`,`plt.xlabel()`, `plt.ylabel()`
- Show with `plt.show()`

Example

```
import matplotlib.pyplot as plt

plt.plot([1,2,3], [1,2,3])
```

Activity: Line Plot

- Plot temperature for Dublin (average per month for one year) with two simple lists.

Markers and Line Styles

Colours

black	k	dimgrey	dimgray
grey	gray	darkgray	darkgrey
silver	lightgrey	lightgray	gainsboro
whitesmoke	white	w	snow
rosybrown	lightcoral	indianred	brown
firebrick	maroon	darkred	red
r	mistyrose	salmon	tomato
darksalmon	coral	orangered	lightsalmon
sienna	seashell	chocolate	saddlebrown
sandybrown	peachpuff	peru	linen
bisque	darkorange	burlywood	antiquewhite
tan	navajowhite	blanchedalmond	papayawhip
moccasin	orange	wheat	oldlace
floralwhite	darkgoldenrod	goldenrod	cornsilk
gold	lemonchiffon	khaki	palegoldenrod
darkkhaki	ivory	beige	lightyellow
lightgoldenrodyellow	olive	y	yellow
olivedrab	yellowgreen	darkolivegreen	greenyellow
chartreuse	lawngreen	sage	lightsage
darksage	honeydew	darkseagreen	palegreen
lightgreen	forestgreen	limegreen	darkgreen
green	g	lime	seagreen
mediumseagreen	springgreen	mintcream	mediumspringgreer
mediumaquamarine	aquamarine	turquoise	lightseagreen
mediumturquoise	azure	lightcyan	paleturquoise
darkslategray	darkslategrey	teal	darkcyan
c	cyan	aqua	darkturquoise
cadetblue	powderblue	lightblue	deepskyblue
skyblue	lightskyblue	steelblue	aliceblue
dodgerblue	lightslategray	lightslategray	slategray
slategrey	lightsteelblue	cornflowerblue	royalblue
ghostwhite	lavender	midnightblue	navy
darkblue	mediumblue	blue	b
slateblue	darkslateblue	mediumslateblue	mediumpurple
blueviolet	indigo	darkorchid	darkviolet
mediumorchid	thistle	plum	violet
purple	darkmagenta	m	fuchsia
magenta	orchid	mediumvioletred	deeppink
hotpink	lavenderblush	palevioletred	crimson
pink	lightpink		

Activity: Add Another City

- Add the temperature for another city
- Add different line styles, markers, colours for each
- Add a legend
- Add a grid

Bar Charts

- `plt.bar(categories, values)` for vertical bars

- `plt.barh()` for horizontal bars

- Good for comparing categories

Activity: Bar Chart

- Plot a list of 5 students with their respective average grades in a bar chart (horizontal and vertical).

Histograms

- `plt.hist(data, bins=n)` shows distribution

- Useful for frequency counts of numeric data

Activity: Histogram

- Create a list of values for the age of students and plot a histogram to show the distribution.

Scatter Plots

- `plt.scatter(x, y)` for relationships
- Good for correlation or clustering

Activity: Scatter Plot

- Compare height and weight of 5,000 individuals (normal distribution).

- Height: mean = 170cm, std = 7cm

- Weight: mean = 80.7kg, std = 5kg

- Correlation = 0.6

Box Plots

- `plt.boxplot(data)` shows median, quartiles, outliers

- Useful for understanding spread and outliers

Activity: Box Plot

- Create a normal distribution of yearly income for Malazan cadre mages and visualise it with a box plot.

- Hint: There's around 80 mages in the whole army.

1 gold falcon $\simeq$ 20 silver falcons

Make reasonable assumptions for everything else you need.

💰 **Malazan Cadre Mage Income Estimate**

Category	Value
Base pay (Imperial mage rank)	4–6 silver falcons per week
Hazard pay (campaigns, demons, undead desert snakes)	2–10 falcons per week
Survival bonus (if you live)	1–2 gold falcons per campaign
Loot/ "unofficial acquisitions"	wildly variable
Pension	*lol*

Using pandas Plotting

- `df.plot()` uses matplotlib under the hood
- Quick way to create line, bar, histogram etc.
- Pass `kind=bar`, `kind=hist`, etc.

Activity: pandas Plotting

- Load the `malazan_armies.json` file into a DataFrame.

- Experiment with different types of charts using `df.plot(kind=…)`

Plot Styling and Formatting

- Change colours, figure size, labels, font size
- `plt.figure(figsize=(10,6))` for bigger plots
- Rotate labels: `plt.xticks(rotation=45)`
- Set colour through color property in the plot method (e.g. `plt.bar(color=…)`)

Activity: Styled Chart

- Use one of the bar charts you've created earlier.
- Add rotated labels and colours.

Saving Figures

- Use `plt.savefig('filename.png')` to export

- Useful for reports and presentations

Activity: Export Plot

- Save one of your plots as PNG or PDF

Combining Multiple Plots

- `plt.subplot()` allows multiple plots on one figure

- Useful for comparisons or dashboards

- `plt.subplot(rows, columns, which plot)`

Activity: Two Plots Side-by-Side

- Use `subplot(1,2,i)` to show some of your line and bar charts together

Choosing the Right Chart

- Line -> trend, time series

- Bar -> categories

- Histogram -> distribution

- Box plot -> spread, outliers

- Scatter -> relationships

Best Practices

- Always label axes and titles

- Use readable fonts and colours

- Avoid clutter and distortion

- Match chart type to data and purpose

Recap and Next Week

- You learned to build charts with matplotlib and pandas

- Next: Data Cleaning and Preprocessing

Questions?