matplotlib

Pictures of your data: look before you model, check after you fit

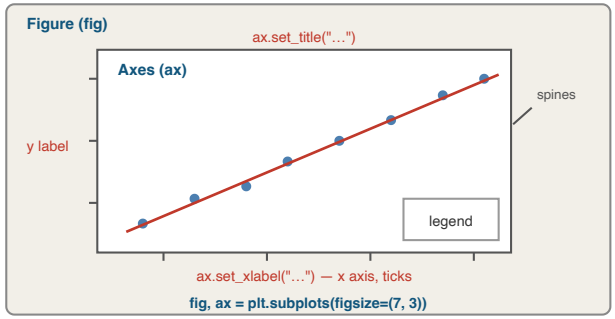
Python cheatsheet · MATH60033A · Quantitative Methods in IB · used in sessions 3 to 13

```
import matplotlib.pyplot as plt
```

one figure, one or more axes; draw on the axes

1 · The anatomy of a figure

*A **figure** is the page; an **axes** is one plot on it. You draw on the axes.*



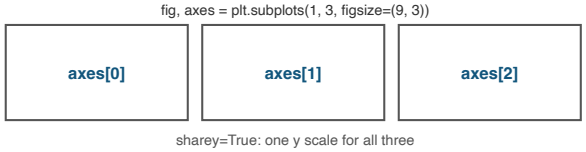
`fig, ax = plt.subplots(figsize=(7, 3))`
one figure with one axes, 7 × 3 inches — wide and short fits a slide.

`ax.scatter(...)` · `ax.plot(...)` · `ax.hist(...)`
every drawing call is a method of `ax`.

`plt.show()`
display the figure (in a script, it opens a window).

> `plt.scatter(...)` also works: it draws on the *current* axes. With several panels, use `ax.` — it says which one.

2 · Several panels



`fig, axes = plt.subplots(1, 3, figsize=(9, 3))`
a row of three axes; `axes[0]`, `axes[1]`, `axes[2]`.

`plt.subplots(2, 2)` → `axes[0, 1]`
a grid: row, then column.

`plt.tight_layout()`
stop labels from overlapping between panels.

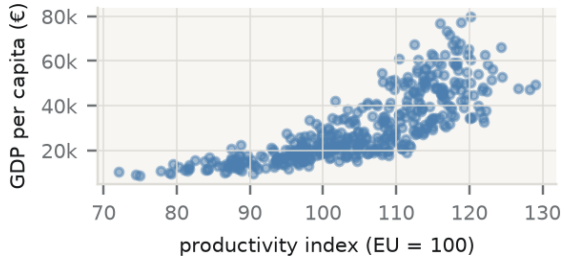
3 · Points and lines

`ax.scatter(x, y, s=12, alpha=0.6)`
one dot per row. `s` is the marker size, `alpha` the transparency: with hundreds of points, see-through dots show density.

```
ax.scatter(d["productivity_idx"], d["gdp_pc_eur"],
s=12, alpha=0.6)
```

`ax.plot(x, y, color="…", lw=2, ls="--")`
a line through the points, in order of `x`. `lw` width, `ls` style: `"-"`, `"--"`, `":"`.

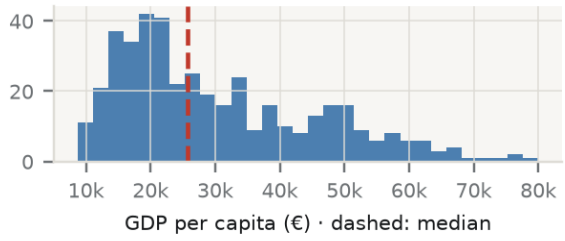
`ax.plot(x, y, label="ridge"); ax.legend()`
name each line, then draw the legend.



4 · Distributions

`ax.hist(x, bins=30)`
how a variable is spread: count of rows per bin. Try several `bins`.

`d.boxplot(column="alcohol", by="good", ax=ax)`
pandas draws a box per group on a matplotlib axes: median, quartiles, outliers.



Before any model: look

The habits of session 3, on every plot you make.

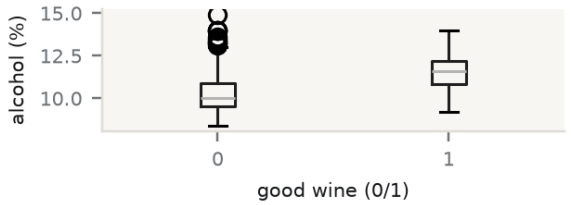
`1 · draw the data first`
a straight line through a curved cloud is still a straight line — the software will not warn you.

`2 · say the units`
on both axes, every time: € per head, % of GDP, index points.

`3 · one question per plot`
a residual plot asks *what did the line miss?*; a histogram asks *how is it spread?*

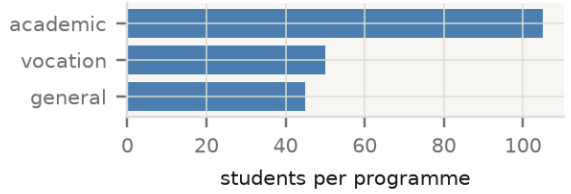
`4 · a line is a prompt`
a threshold or a zero line says where to look, not what to delete.

Gallery · a box per group (session 4)



`w.boxplot(column="alcohol", by="good", ax=ax)`
if the two boxes overlap completely, no coefficient will rescue the predictor.

Gallery · a bar chart



`ax.barh(s.index, s.values); ax.invert_yaxis()`
counts or scores per category; horizontal bars keep long names readable.

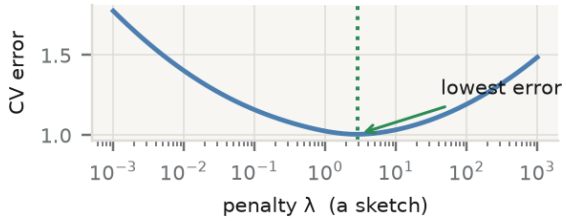
5 · Reference lines and words

`ax.axhline(0, color="black", lw=1)`
a horizontal line across the plot — zero, a threshold, a mean.

`ax.axvline(x, ls=":")`
a vertical line — the chosen λ, the median.

`ax.text(x, y, "sweet spot", fontsize=9)`
words placed at data coordinates.

`ax.annotate("lowest", xy=(x, y), xytext=(x2, y2), arrowprops=dict(arrowstyle="->"))`
words with an arrow pointing at a point.



> A threshold line is a prompt to **look** at the rows above it — not a rule to delete them.

6 · Labels, scales, limits

`ax.set_xlabel("GDP per capita (€)") · ax.set_ylabel(…)`
say the units, every time.

`ax.set_title("…") · fig.suptitle("…")`
a title per panel · one for the figure.

`ax.set_xscale("log")`
a log axis: λ grids, incomes, FDI — anything spanning orders of magnitude.

`ax.set_xlim(0, 10) · ax.set_ylim(…)`
fix the range shown.

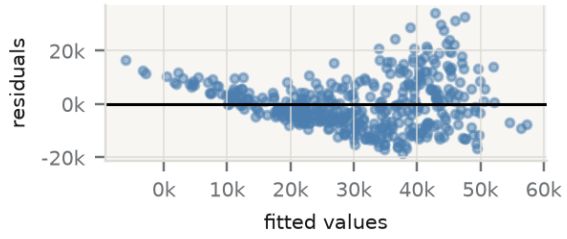
`ax.set_xticks(pos); ax.set_xticklabels(names, rotation=60)`
name the ticks, tilted when the names are long.

`ax.invert_yaxis()`
largest bar at the top in a horizontal bar chart.

7 · Save

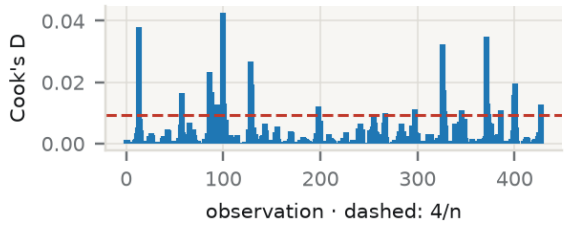
`plt.savefig("out.png", dpi=110, bbox_inches="tight")`
a file for a paper or a slide. Call it **before** `plt.show()`, which empties the figure.

Gallery · residuals vs fitted (session 3)



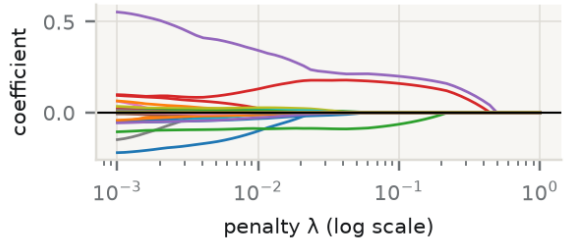
`ax.scatter(fit.fittedvalues, fit.resid, s=12, alpha=0.6); ax.axhline(0)`
random scatter around zero: fine. A curve or a funnel: the model missed something.

Gallery · Cook's distance (session 3)



`ax.stem(cooks_d); ax.axhline(4 / len(d), ls="--")`
one spike per row: a tower above the rest is a row carrying the result.

Gallery · a coefficient path (session 5)



`for k in range(P): ax.plot(alphas, coefs[k]); ax.set_xscale("log")`
one line per predictor as λ grows: the lasso switches them off one by one.

8 · Colours, markers, styles

`color="tab:blue" · color="#11557c"`
a named colour or a hex code; keep one colour per meaning across panels.

`marker="o" · ls="--" · lw=2`
dots, dashed lines, line width — for lines that must print in black and white.

`alpha=0.3`
transparency: overlapping points show where the data are dense.

`ax.grid(axis="y", visible=False)`
remove grid lines that compete with the data.

`cmap="Blues" in ax.imshow(corr)`
a colour scale, e.g. a correlation matrix as a heat map.

9 · When a plot looks wrong

`nothing appears`
in a script, call `plt.show()` at the end.

`the saved file is blank`
`plt.savefig` must come **before** `plt.show()`.

`labels overlap`
`plt.tight_layout()`, a wider `figsize`, or `rotation=60` on the ticks.

`a solid blob of points`
smaller `s=`, lower `alpha=`, or a histogram instead.

`a line zig-zags`
`ax.plot` joins points in row order: sort by x first.

`one panel's scale hides another's`
`plt.subplots(..., sharey=False)`, or a log axis.

> A plot without units on its axes is not finished.