FREE EDITION · NOTES + 3 SAMPLE MCQS

CUET · COMPUTER SCIENCE · CLASS XII · CODE 308

Plotting Data using Matplotlib

CUET unit: Plotting Data using Matplotlib

By **UniDrill** · NCERT-grounded study material

WWW.UNIDRILL.IN

📌 Snapshot

- Data visualisation uses Python's Matplotlib library — the key tool for converting tabular/numerical data into charts that reveal trends, comparisons, and distributions.
- The full workflow: importing pyplot, creating line charts, bar charts, histograms, scatter plots, box plots, and pie charts — all the standard chart types.
- Customisation parameters (markers, colours, linewidth, linestyle, grid, title, labels, ticks) are tested directly by NTA through code-reading MCQs.
- Pandas' built-in `.plot()` method is a wrapper around pyplot, including the `kind=` keyword that selects chart type — a frequently tested concept.
- Open data concepts and quartile/box-plot theory (outliers, whiskers, IQR) appear in data-interpretation questions.

📖 Detailed Notes

2.1 Core concepts

- **Data visualisation** means graphical or pictorial representation of data using graphs, charts, etc. Its purpose is to visualise variation and show relationships between variables. Visualisation helps communicate results effectively to intended users. (NCERT §4.1, p. 105–106)
- **Matplotlib library** is used for creating static, animated, and interactive 2D plots in Python. It is installed with `pip install matplotlib` and its pyplot module is imported with `import matplotlib.pyplot as plt`, where `plt` is a commonly used alias. (NCERT §4.2, p. 106)
- **Figure and pyplot module:** The pyplot module contains a collection of functions that work on a figure. A figure is the overall window where outputs of pyplot functions are plotted. A figure contains a plotting area, legend, axis labels, ticks, and title (Figure 4.1). Each pyplot function makes some change to the figure. (NCERT §4.2, p. 106–107)
- **plot() function:** `plt.plot(x, y)` plots x versus y as a line chart by default. `plt.show()` displays the figure. A figure can be saved with `plt.savefig('filename.png')`. (NCERT §4.2, p. 107–108)

- **Table 4.1 — Pyplot functions for different chart types:** `plot()` for line/markers, `bar()` for bar plot, `boxplot()` for box-and-whisker plot, `hist()` for histogram, `pie()` for pie chart, `scatter()` for scatter plot. (NCERT §4.2, p. 108)
- **Table 4.2 — Customisation functions:** `grid()` configures grid lines; `legend()` places a legend; `savefig()` saves the figure; `show()` displays all figures; `title()` sets chart title; `xlabel()` / `ylabel()` set axis labels; `xticks()` / `yticks()` get or set tick locations and labels. (NCERT §4.3, p. 108)
- **Marker (§4.3.1):** A marker is any symbol that represents a data value in a line chart or scatter plot. Marker codes include `"."` (point), `","` (pixel), `"o"` (circle), `"v"` (triangle down), `"^"` (triangle up), `"*"` (star), `"D"` (diamond), `"s"` (square), `"p"` (pentagon), among others listed in Table 4.3. (NCERT §4.3.1, p. 109–110)
- **Colour (§4.3.2):** Colour can be specified in the `color` parameter using character codes: `'b'` =blue, `'g'` =green, `'r'` =red, `'c'` =cyan, `'m'` =magenta, `'y'` =yellow, `'k'` =black, `'w'` =white (Table 4.4). Full colour names may also be used. (NCERT §4.3.2, p. 110)
- **Linewidth and Line Style (§4.3.3):** `linewidth` is specified in pixels; default is 1 pixel. Values greater than 1 produce thicker lines. The `linestyle` parameter accepts strings such as `"solid"`, `"dashed"`, `"dashdot"`. (NCERT §4.3.3, p. 111)
- **Pandas .plot() method (§4.4):** From Pandas version 0.17.0, Series and DataFrame objects have a `.plot()` method. Syntax: `df.plot(kind='...')` where `kind` accepts values listed in Table 4.5: `line` (default), `bar`, `barh`, `hist`, `box`, `area`, `pie`, `scatter`. This is a wrapper around matplotlib.pyplot. (NCERT §4.4, p. 112–113)
- **Plotting a Line Chart (§4.4.1):** A line plot shows frequency of data along a number line; used for continuous datasets to visualise growth or decline over time. When a DataFrame is plotted with `kind='line'`, the x-axis defaults to the numeric index (row labels). Custom x-tick labels can be set using `plt.xticks(ticks, labels)`. (NCERT §4.4.1, p. 113–116)
- **Plotting Bar Chart (§4.4.2):** Bar charts are preferred to show comparisons. Unlike line plots, bar charts can plot string values on the x-axis. `kind='bar'` plots a vertical bar chart; `kind='barh'` plots a horizontal bar chart. The `x=` parameter specifies the DataFrame column for x-axis; `edgecolor`, `linewidth`, `linestyle` can further customise bars. (NCERT §4.4.2, p. 116–118)
- **Plotting Histogram (§4.4.3):** A histogram is a column-chart where each column represents a range of values (bins), and height corresponds to the count of values in that bin. `df.plot(kind='hist')` auto-selects bin size. The `bins` parameter can be an integer, list, or range. The `fill` parameter (boolean) controls whether bars are filled; `hatch` fills bars with a pattern (`'-'`, `'+'`, `'x'`, `'\\'`, `'*'`, `'o'`, `'O'`, `'.'`). (NCERT §4.4.3, p. 119–120)
- **Open Data (§4.4.3 subsection):** Websites that provide data freely for anyone to download are called Open Data sources. "Open Government Data (OGD) Platform

India" (data.gov.in) is the Government of India's open data platform. (NCERT §4.4.3, p. 121)

- **Plotting Scatter Chart (§4.4.4):** A scatter chart is a two-dimensional visualisation that uses dots for two different variables — one on each axis. Scatter plots are also called correlation plots. The size `s` of the bubble can encode a third variable. Syntax: `plt.scatter(x=..., y=..., s=size, color=..., marker=..., edgecolor=...)`. (NCERT §4.4.4, p. 124–125)
- **Quartiles and Box Plot (§4.4.5):** Quartiles divide data into four equal parts. A Box Plot is the visual representation of a statistical summary: Minimum, Q1, Q2 (Median), Q3, Q4, Maximum, and Outliers. Whiskers extend to highest and lowest non-outlier values. Shorter whisker distance = small variation; longer = large variation. `kind='box'`; `vert=False` makes it horizontal. (NCERT §4.4.5, p. 126–129)
- **Plotting Pie Chart (§4.4.6):** A pie chart divides a circle into sectors, each representing a part of the whole. Use `df.plot(kind='pie', y='column_name')` or set `subplots=True`. `explode` specifies fraction to offset each slice; `autopct` displays percentage labels (e.g., `"%.2f"`). Default labels are index values of the DataFrame. (NCERT §4.4.6, p. 130–133)
- **When NOT to use each chart (NCERT §4.4).** Line charts assume continuity along the x-axis — they should not be used for categorical x-data (use bar instead). Pie charts become unreadable with more than 5-6 slices — use a bar chart for many categories. Histograms need numeric continuous data — they should not be used for categorical counts (use bar). NTA scenario-based questions test this judgment.
- **Scatter for correlation (NCERT §4.4.4, p. 124).** A positive slope in the scatter cloud suggests positive correlation; a negative slope suggests negative correlation; a shapeless cloud suggests no correlation. NCERT introduces the concept; CUET sometimes tests interpretation through diagram-reading MCQs.
- **Box plot strengths (NCERT §4.4.5, p. 126).** A single chart compresses min, Q1, median, Q3, max, and outliers into one view — making side-by-side comparison of multiple subjects (or groups) very efficient. The "Performance Analysis" example shows this.
- `subplots=True` **for pie charts (NCERT §4.4.6, p. 130).** Without `subplots=True`, Pandas may refuse to plot a multi-column pie chart. Specifying `y='col'` and `subplots=True` together is the safest pattern.

2.2 Definitions to memorise

Term	Definition	Page
Data visualisation	Graphical or pictorial representation of data using graphs, charts, etc., to visualise variation or show relationships between variables	105
Matplotlib	Python library for creating static, animated, and interactive 2D plots	106

Term	Definition	Page
pyplot module	Sub-module of Matplotlib containing functions to create and customise figures and plots	106
Figure	The overall window where outputs of pyplot functions are plotted; contains plotting area, legend, axis labels, ticks, title	106
Marker	Any symbol that represents a data value in a line chart or scatter plot	110
Linewidth	Width of the line in a chart, specified in pixels; default is 1	111
Bin	An interval range into which data is sorted in a histogram; height of each bin is proportional to count of data points in it	119
Open Data	Data freely available for anyone to download and use, primarily for educational purposes; data.gov.in is an example	121
Scatter chart	Two-dimensional visualisation using dots to show the relationship (correlation) between two variables	124
Quartile	A measure that divides data into four equal parts, each containing an equal number of observations	126
Box Plot	Visual representation of the statistical summary of a dataset: Min, Q1, Q2 (Median), Q3, Q4, Max, and Outliers	126
Outlier	An observation that is numerically distant from the rest of the data; shown as individual points beyond the whiskers in a box plot	126
Whisker	The two lines outside the box in a box plot that extend to the highest and lowest non-outlier values	126
Explode (pie chart)	Parameter that specifies the fraction of radius by which each pie slice is offset/expanded	132
Autopct	Parameter in pie chart that displays each slice's percentage as a label	132
`plt.show()`	Pyplot function that renders/displays figures	107
`plt.savefig()`	Pyplot function that saves a figure to disk	108
`plt.title()`	Adds title to the current plot	108
`plt.xlabel()` / `plt.ylabel()`	Set axis titles	108
`plt.xticks()` / `plt.yticks()`	Set tick positions and labels	108
`plt.legend()`	Add a legend to the plot	108
`bar()`	Pyplot function for vertical bar chart	108
`hist()`	Pyplot function for histogram	108

Term	Definition	Page
`scatter()`	Pyplot function for scatter plot	108
`boxplot()`	Pyplot function for box-and-whisker plot	108
`pie()`	Pyplot function for pie chart	108
Bins	Intervals used to group data in a histogram	119
Hatch	Pattern fill on histogram or bar	120
Correlation Plot	Alternative name for scatter plot	124
OGD Platform India	data.gov.in — India's open data portal	121

2.3 Diagrams / processes to remember

- **Figure 4.1 — Components of a plot (p. 106):** Shows a complete labelled diagram of a chart with Chart Title, y-axis, x-axis, y ticks, x ticks, axis titles, Plotted Data, and Legend. Students must know all component names.
- **Figure 4.2 — Line chart output of Program 4-1 (p. 107):** Basic line chart of date vs. temperature with no labels or title — illustrates the minimal `plt.plot(x, y); plt.show()` usage.
- **Figure 4.3 — Line chart with labels and grid (p. 109):** Same data with `xlabel`, `ylabel`, `title`, `grid(True)`, `yticks` added — shows how customisation functions change appearance.
- **Figure 4.16 — A Box Plot structure (p. 126):** Labelled diagram showing Minimum, Q1 (lower quartile), Q2 (middle/median), Q3 (upper quartile), Q4, Maximum, and Outliers. Critical for understanding box plot interpretation questions.
- **Table 4.5 — kind= values for Pandas .plot() (p. 113):** line, bar, barh, hist, box, area, pie, scatter. Must be memorised as NTA frequently tests which `kind=` value produces which chart type.

2.4 Common confusions / NTA trap points

- `plt.plot()` **vs.** `df.plot(kind='line')`**:** Both produce line charts. `plt.plot()` is a direct pyplot call; `df.plot()` is Pandas' wrapper. The `kind='line'` is the default for `df.plot()`. Confusing the two syntaxes is a common error.
- `bar` **vs.** `barh`**:** `kind='bar'` produces a vertical bar chart; `kind='barh'` produces a horizontal bar chart. NTA distractors often swap these.
- **Default x-axis in Pandas bar/line plots:** If the `x=` parameter is not specified in `df.plot()`, the bar plot uses the DataFrame index (numeric, starting from 0) as x-axis — not column names. Students miss this and select wrong answers in output-prediction questions.

- `fill=True` **vs.** `fill=False` **in histograms:** `fill=True` (default) means bars are colour-filled; `fill=False` means bars are empty (outline only). The `hatch` parameter adds a pattern regardless of `fill`.
- **Box plot components (NCERT §4.4.5, p. 126).** Q1 = 25th, Q2 = 50th (Median), Q3 = 75th percentile. Whiskers reach min/max of non-outlier data; outliers appear as individual dots.
- **Default x-axis is the index (NCERT §4.4.2, p. 116-117).** Use `x=` parameter to select another column.
- `plt.show()` **must be called (NCERT §4.2, p. 107).** Without it the figure may not display in script mode.
- `savefig()` **before** `show()` — once a figure is shown and closed it cannot be saved. NCERT note implicitly.
- **Pie chart needs a single column (NCERT §4.4.6, p. 130-132).** Use `y='col'` to specify which column to plot.
- `legend()` **placement** — accepts a `loc` argument; default is 'best' (NCERT Table 4.2, p. 108).
- **Bar vs Histogram (NCERT §4.4.2 vs §4.4.3).** Bar compares categories; histogram shows frequency over intervals — they are NOT interchangeable.

🎯 Practice MCQs

Q1. Which of the following commands is used to import the pyplot module of Matplotlib with the alias `plt`?

A. `import matplotlib as plt`

B. `import pyplot.matplotlib as plt`

C. `import matplotlib.pyplot as plt`

D. `from matplotlib import plot as plt`

Q2. Consider the following code: ```python import matplotlib.pyplot as plt date = ["25/12", "26/12", "27/12"] temp = [8.5, 10.5, 6.8] plt.plot(date, temp) plt.show() ``` What type of chart does this code produce by default?

A. Bar chart

B. Scatter plot

C. Histogram

D. Line chart

Q3. Match the following Pandas `df.plot(kind=...)` values with their corresponding chart types: | List I (kind=) | List II (Chart type) | |---|---| | P. `'barh'` | 1. Pie chart | | Q. `'hist'` | 2. Horizontal bar plot | | R. `'pie'` | 3. Scatter plot | | S. `'scatter'` | 4. Histogram |

A. P-2, Q-4, R-1, S-3

B. P-1, Q-4, R-2, S-3

C. P-2, Q-3, R-1, S-4

D. P-4, Q-2, R-3, S-1

📊 PYQ Alignment

Data visualisation has appeared consistently in CUET (UG) Computer Science/Informatics Practices papers because Matplotlib is a core practical skill in the Class XII syllabus; questions most frequently target correct function names (`plt.plot()` vs. `df.plot()`), the `kind=` parameter values (especially bar vs. barh, hist, scatter, box), customisation parameters (marker, color, linewidth, linestyle), and box-plot interpretation (identifying median, outliers, and variation from whisker length). See PYQ archive for Computer Science.