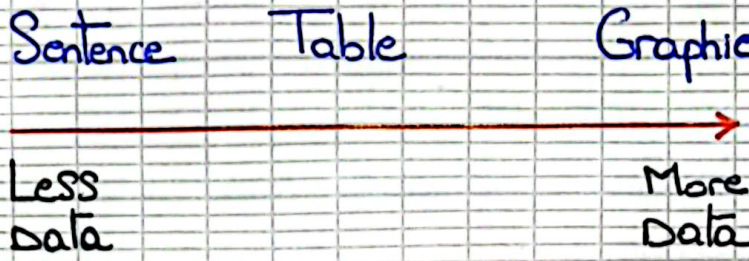
I_ Plots: Find the correct plot for your data

→ Do you really need a plot?



- Sentence → use when to display 1-2 numerical values
- Table → use when exact value matter; arrange for readability
- Graph → Use when comparing trends, patterns, or proportions, not specific numbers.

→ Types of Plots

1) Displaying Distributions
   - ↳ many observations for one variable
   - ↳ Summary statistics are not informative
   - ↳ Histograms / Density plots / Box plots / Violin plot Ridgeline plot

2) Relationship
   - ↳ Compare two (max three) variables
   - ↳ Line plots / Scatter plots / Heat Maps

3) Ranking
   - ↳ Comparing several variables
   - ↳ interested which one is bigger, but not how much bigger
   - ↳ Bar plots / Paralled plots / Word Cloud

4) Part to Whole
- Knowing the predominant category
- Pie charts / Dendograms / Treemaps

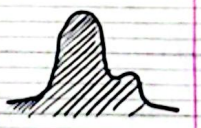
5) Map
- Data is geographical
- Background Map / Choropleth Maps

→ Plotting Distributions

1) Histogram

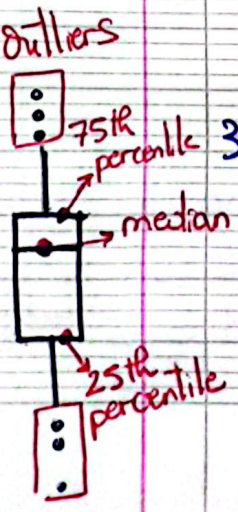
- Represents the distribution of a numerical variable
- Divide data into bins and count the nb. of observ. in each bin
- Different bin sizes → different conclusions
- only for numerical data, not categorical
  ↳ use bar plot

2) Density Plot



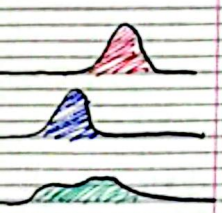
- Similar to histogram, provides a smooth estimate of the distribution
- More suitable for continuous variables
- Good for comparing multiple distributions (use transparency to overlay multiple densities)

3) Box Plot (Box-and-Whiskers Plot)



- Summarize a distribution using percentiles (25th, 50th/median, 75th) and outliers
- Helps in comparing multiple distributions
- Downside: Doesn't show exact sample size
- Can overlay individual data points for better clarity

4) Violin Plot
- Combination of box plot and density plot
- Displays both summary statistics and the full distribution (in a symmetric way)
- Useful when dealing with large datasets
- More informative than box plots in scientific papers



5) Ridgeline Plot
- Shows distributions of multiple variable in a stacked manner
- Uses either density plots or histograms
- Best for few variables with clear patterns (too many categories lead to clutter)

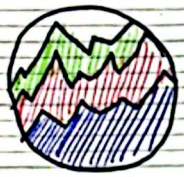
→ Plotting Relationships

1) Line Plot

- best for showing trends overtime (Time series)
- X_axis should be continuous
- Can be used for discrete data with high resolution
- Avoid for categorical data (misleading)

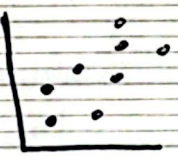
2) The area Plot
- Shaded line plot
- Not recommended unless representing distributions
- Violates the principal of minimum ink (unnecessary shading)
- Overlapping area can reduce readability

3) Stacked area plot



- Displays distributions of variables so as to avoid overlap between them
- Not recommended for relationships between variables
- Difficult to interpret

4) Scatter Plot

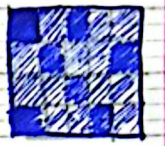


- Displays relationships between two numeric variables
- For each datapoint, value of its 1st variable is represented on X, the 2nd on the Y

5) Bubble Plot

- A scatter plot with a third variable mapped to the size of the dot
- Prioritizes X and Y axes over the size variable
- Using color adds a fourth variable but may reduce readability.

6) Heatmap



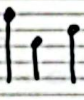
- A matrix where values are represented using colors
- Useful for spotting patterns in large datasets
- A better alternative to 3D plots

→ Plotting Ranking

1) BarPlot

- Most efficient way to compare rankings.
- Rlt. between numeric and categoric variable
- Each categorical variable = bar
- Numeric value = length of the bar
- Order categories to improve readability
- Grouped bar charts work for multiple observations
- Avoid stacked bar charts (hard to compare values)

2) Lollipop Plot



- A lighter alternative to bar plots
- Good for dense barplots with similar value
- Uses thin lines with dots, making it cleaner
- Reduces ink usage

3) Cleveland Dot (Dumbbell) Plot

- Useful for comparing two values per category

4) The circular barplots

- Bar plot where the x-axis is wrapped around a circle
- Avoid using them

5) Parallel Plot

- Displays multivariate data by connecting values across variables

6) Avoid Radar (Spider) plots
- Converts linear data into a difficult-to-read circular format
- Area-based interpretation is misleading

7) WordCloud
- Font size proportional to the time that they appearing in a text
- Area is hard to decode
- Longer words appear bigger

→ Comparing Part to whole

1) Pie Charts
- Circle divided into sectors, each representing a proportion of the whole.
- Often used to show proportion, where the sum of sectors equal 100%
- Poor at showing differences, angles and areas are hard to compare
- Only acceptable:
  - ↳ We're comparing one part to the whole = only 2 slices
  - ↳ Exact values are irrelevant
  - ↳ Slices are easily compared at a glance

2) Doughnut Chart
- Same as pie chart with a whole in the middle.

3) Treemap
- Nested rectangles, good for hierarchical data
- Each group represented by a rectangle whose area is proportial to its value
- Not for precise comparisons

4) Dendrogram
- Hierarchical rlt. between categories

→ Plotting Spatial Data Maps
- Starting point of any geospatial visualization
  (1) Find spatial data
  (2) Plot it

1) Choropleth map
- Displays divided geographical areas or regions that coloured in relation to a numeric variable
- Problem: Bigger regions attract more attention
  [variable should be normalized

2) others
- Connection map
- Hexbin map
- Cartogram