

Week 4 — Visual Types II: Bars, Distribution & Maps

STA191 Fundamentals of Data Visualization · Topic 2 · Suriyati Ujang

The rest of the toolkit: bars for comparison, histograms and boxplots for distribution, geospatial for place — and the traps that make good data lie.

By the end of this week you can: build bar charts correctly (zero baseline); choose among bar types; show a distribution with a histogram and a boxplot; map data; and avoid common visual traps.

Session 1 — Bars & the zero-baseline rule

4.1 Bars — comparing categories

Bars are for categorical comparison: the eye compares endpoints, so it's easy to see which is largest. Because we judge by length, bar charts **must start at zero** — otherwise the lengths lie.

The zero-baseline trap: quarterly sales RM102k→105k→108k→110k drawn with the axis starting at 100k look like a fivefold jump; drawn from zero, they show the honest 8% rise. The truncated axis was the lie. Keep bars wider than the gaps between them.

4.2 The bar family

- **Vertical / clustered** — compare categories, or several series side by side.
- **Stacked** — show a total and its parts; but only the bottom segment shares a baseline, so upper parts are hard to compare.
- **Horizontal** — best when category names are long.
- **Stacked horizontal** — ideal for Likert survey data (strongly disagree → strongly agree).

Example: course-feedback Likert responses sit naturally on a stacked horizontal bar.

Session 2 — Distribution, maps & cautions

4.3 Histogram — the shape of one variable

Groups numbers into bins and shows how many values fall in each; bars touch because the scale is continuous. Use it to see the distribution — normal, skewed, or with outliers.

Example: bin final marks 0–20...80–100 to see whether the class clusters in the middle or leans low.

4.4 Boxplot — a five-number summary you can see

Draws minimum, Q1, median, Q3, maximum, with outliers as dots. Fastest way to compare spread across groups. You already compute this by hand — a boxplot is that summary drawn. Side-by-side boxplots compare, e.g., CGPA spread across programmes. Judge shape from the box, not the whiskers.

4.5 Geospatial — data on a map

When location is the point, map it. A choropleth shades each region by value (darker = higher); bubbles size each place by value. *Example:* shading states by dengue cases lights up the hotspots instantly.

4.6 Cautious elements — how good data lies

- **Pie traps** — too many slices, or 3D/shadow effects. Prefer bars for comparison.
- **Never use 3D** for a single dimension — it warps values.
- **Avoid a secondary y-axis** — split into two charts instead.
- **Cut clutter** — heavy gridlines, shading, rainbow palettes, bold everywhere.

The two biggest everyday lies: the truncated bar axis and the 3D pie. Both make small differences look dramatic.

Key terms: zero baseline · clustered/stacked bar · Likert · histogram & bins · boxplot/five-number summary · choropleth · chartjunk

Review: (1) A bar y-axis starts at 100 not 0 — effect? (2) Which chart shows the distribution of one numeric variable? (3) What five numbers does a boxplot display?

Glossary

- **Zero baseline** — a bar chart's y-axis must start at zero so lengths compare honestly.
- **Clustered / stacked bar** — clustered compares series; stacked shows a total and its parts.
- **Likert (stacked horizontal)** — a survey scale shown as one stacked horizontal bar per question.
- **Histogram & bins** — groups numeric values into touching intervals showing how many fall in each.
- **Boxplot / five-number summary** — min, Q1, median, Q3, max (+ outliers).
- **Choropleth** — a map shading regions by value (darker = higher).
- **Chartjunk** — decorative clutter that lowers clarity and trust.