30 Worked Answers and Explanations

Notes for SHS • Data Science • Topic 82 • All research examples are fictional

Question 1 — Define visualization

Question: What is data visualization?

Hint: It represents data visually.

Worked answer: Graphical representation of data to explore, explain or communicate findings.

Common mistake: A graph cannot create evidence missing from its underlying data.

Question 2 — Categorical type

Question: Classify DeviceType in the profiles.

Hint: Labels without intrinsic number.

Worked answer: Nominal categorical: Smartphone, Laptop or Tablet.

Common mistake: Do not connect its categories with a time-series line.

Question 3 — Numerical type

Question: Classify WeeklyStudyHours.

Hint: It measures a quantity.

Worked answer: Numerical measurement, suitably treated as continuous.

Common mistake: Rounded recording does not inherently make the underlying quantity categorical.

Question 4 — Compare categories

Question: Which graph compares 120, 72 and 48 students?

Hint: Use separate bars.

Worked answer: A zero-baseline bar or column chart.

Common mistake: A line graph suggests unjustified continuity between levels.

Question 5 — Trend over time

Question: Which graph best communicates sessions over 12 weeks?

Hint: Time ordered points.

Worked answer: A dated line chart, or period columns if comparing totals.

Common mistake: Preserve chronological order and any missing weeks.

Question 6 — Identify histogram

Question: Which chart groups score observations into numeric bins?

Hint: It shows a distribution.

Worked answer: A histogram.

Common mistake: Unlike a bar chart, bins represent numeric intervals.

Question 7 — Scatter plot

Question: What does a scatter plot use for its two axes?

Hint: One dot per paired record.

Worked answer: Two quantitative variables, such as hours X and score Y.

Common mistake: Association does not imply causation.

Question 8 — Pie vs bar

Question: When might a pie show Level A/B/C shares?

Hint: A legitimate whole.

Worked answer: When three mutually exclusive categories sum to 100%; bars are often better for exact comparison.

Common mistake: Multiple-response categories cannot be a valid exclusive pie.

Question 9 — Chart axes

Question: What do the x and y axes represent in the score scatter?

Hint: Hours horizontal, score vertical.

Worked answer: X: weekly study hours; Y: assessment score (0-100).

Common mistake: State measurement units.

Question 10 — Chart title

Question: Improve “Student Data” for the three-level count chart.

Hint: Say what is shown.

Worked answer: “240 Fictional Students by Study Level (A 120, B 72, C 48).”

Common mistake: Avoid vague or causal headlines.

Question 11 — Bar-chart comparison

Question: Which level contains the most student profiles and by how many more than C?

Hint: Subtract count values.

Worked answer: Level A; 120 - 48 = 72 more students than Level C.

Common mistake: Do not confuse a count difference with percentage points.

Question 12 — Percentages

Question: Compute the shares for levels A, B and C.

Hint: Divide by 240.

Worked answer: A 50%, B 30%, C 20%; total 100%.

Common mistake: The denominator must be all 240 unique profiles.

Question 13 — Weekly line

Question: Name the first and last weekly event totals.

Hint: Read verified summary.

Worked answer: Week 1: 45; Week 12: 70 learning activity events.

Common mistake: These are events, not unique students.

Question 14 — Nonmonotonic four months

Question: For visits 120,150,135,195, is the trend increasing every month?

Hint: Look Feb to Mar.

Worked answer: No: it falls 150 to 135 before rising to 195; Jan-Apr net change +75 (+62.5%).

Common mistake: A positive overall change does not mean monotonic increase.

Question 15 — Histogram bins

Question: Place eight scores 60,65,70,70,75,80,85,95 into bins [60,70), [70,80), [80,90), [90,100).

Hint: The 70 is included in its starting bin.

Worked answer: Frequencies 2,3,2,1, summing to eight.

Common mistake: Never double count observations on boundaries.

Question 16 — Box-plot interpretation

Question: What information is contained in a standard box plot?

Hint: Median and quartiles.

Worked answer: Median, Q1, Q3, IQR, whiskers and separately identified potential outliers.

Common mistake: Do not claim whiskers always mark min/max.

Question 17 — Choosing scatter

Question: Which chart reveals a possible hours-score relationship?

Hint: Paired numeric measurements.

Worked answer: Scatter plot, with optional justified fitted trend line.

Common mistake: Not a pair of unrelated bar charts or proof of causation.

Question 18 — 100% stacked bar

Question: What does it show for different levels?

Hint: Within-group composition.

Worked answer: Percentages of resource preferences by study level, noting denominators A120, B72, C48.

Common mistake: Equal-length 100% bars conceal unequal raw group sizes.

Question 19 — Heatmap values

Question: What does a resource-type-by-subject heatmap color represent?

Hint: Read legend.

Worked answer: Counts of the synthetic activity events for each cell.

Common mistake: A heatmap is not automatically correlations or rates.

Question 20 — Counts and rates

Question: Is 720 learning events equal to 720 active students?

Hint: Multiple events per person.

Worked answer: No: 720 events belong to 225 distinct students in the generated activity table.

Common mistake: Use unique counts for people.

Question 21 — Truncated bar

Question: Why should student-count bars normally start at zero?

Hint: Length encodes magnitude.

Worked answer: A nonzero baseline visually exaggerates relative length differences; zero gives proportional bars.

Common mistake: This is not a universal zero rule for every line graph.

Question 22 — Inappropriate pie

Question: A survey permits selecting multiple preferred resources. Can the selections be exclusive pie slices?

Hint: Percentages can exceed 100%.

Worked answer: Not without a valid mutually exclusive whole; show separate category bars or another suitable chart.

Common mistake: Do not force overlapping answers to sum to 100%.

Question 23 — Dual-axis risk

Question: How can dual y-axes mislead?

Hint: Separate scales can be manipulated.

Worked answer: Different ranges may produce spurious visual alignment; use labels and aligned small multiples where possible.

Common mistake: Visual correlation can be engineered with scales.

Question 24 — Accessible colors

Question: What prevents reliance solely on red and green?

Hint: Use more than hue.

Worked answer: Distinguishable contrast with direct labels and symbols/patterns where appropriate.

Common mistake: Color alone must not communicate essential categories.

Question 25 — Alt text

Question: Write chart alt text for 120/72/48 students.

Hint: Name quantities and category order.

Worked answer: “Bar chart: Level A 120 students, Level B 72, Level C 48.”

Common mistake: “Graph of data” is not meaningful.

Question 26 — Aggregation error

Question: Why is counting all joined profile-event rows a bad way to find students?

Hint: One-to-many joins.

Worked answer: Each student can appear across several events; count distinct StudentCode in profile frame or grouped events.

Common mistake: Do not confuse sessions with people.

Question 27 — Causality warning

Question: If scatter points slope upward, are extra study hours proven to raise scores?

Hint: Think confounding.

Worked answer: No. The graph shows association in synthetic observations, not randomized intervention effects.

Common mistake: Trend lines cannot establish cause.

Question 28 — Uncertainty display

Question: Which chart feature can show justified mean-estimate uncertainty?

Hint: Point plus interval.

Worked answer: A dot with a correctly computed and labelled confidence interval, or a suitable error-bar plot.

Common mistake: Never label SD or SE bars as confidence intervals without method.

Question 29 — Debug a chart

Question: A graph of weekly activity sorts dates as text and omits the zero-activity weeks. What should be fixed?

Hint: Date parsing and full calendar.

Worked answer: Parse real dates, sort chronologically, reindex all twelve expected weeks, and distinguish absent observations from actual zeros.

Common mistake: Do not connect missing values as recorded.

Question 30 — Four-chart story

Question: Choose a four-chart story for the generated 240 profiles and 720 activities.

Hint: Ask four different questions.

Worked answer: Bar: A/B/C 120/72/48; line: 12 weekly event totals; histogram: profile scores; scatter: paired hours and scores; report sources and limitations.

Common mistake: A chart quartet is not causal proof.