Descriptive Statistics:
Concepts, Purpose, and Examples

Roadmap: what descriptive statistics are, why they matter, and how they are used

Objectives: summarize a dataset, pick valid measures by data type, interpret and report results

For students, early-career analysts, and professionals building a reliable foundation

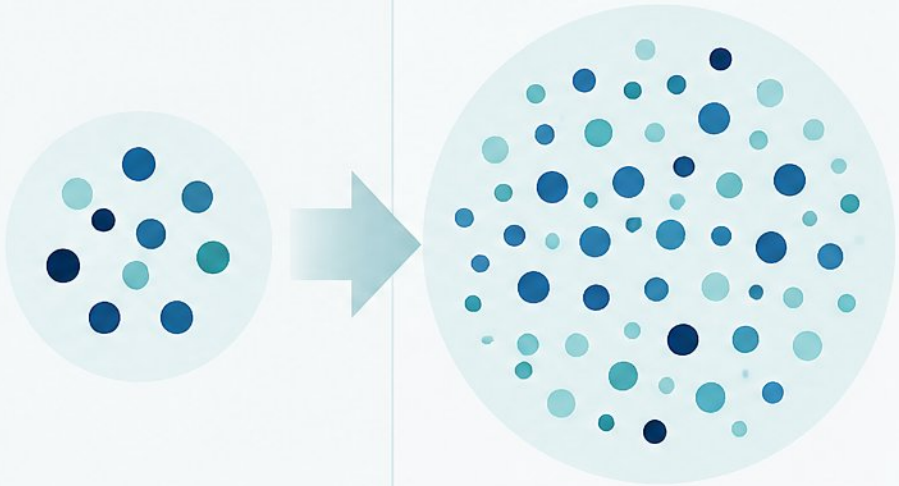
Motivating example: 2,000 raw customer transactions versus a three-number summary

Core question: what is typical, and how much does it vary?



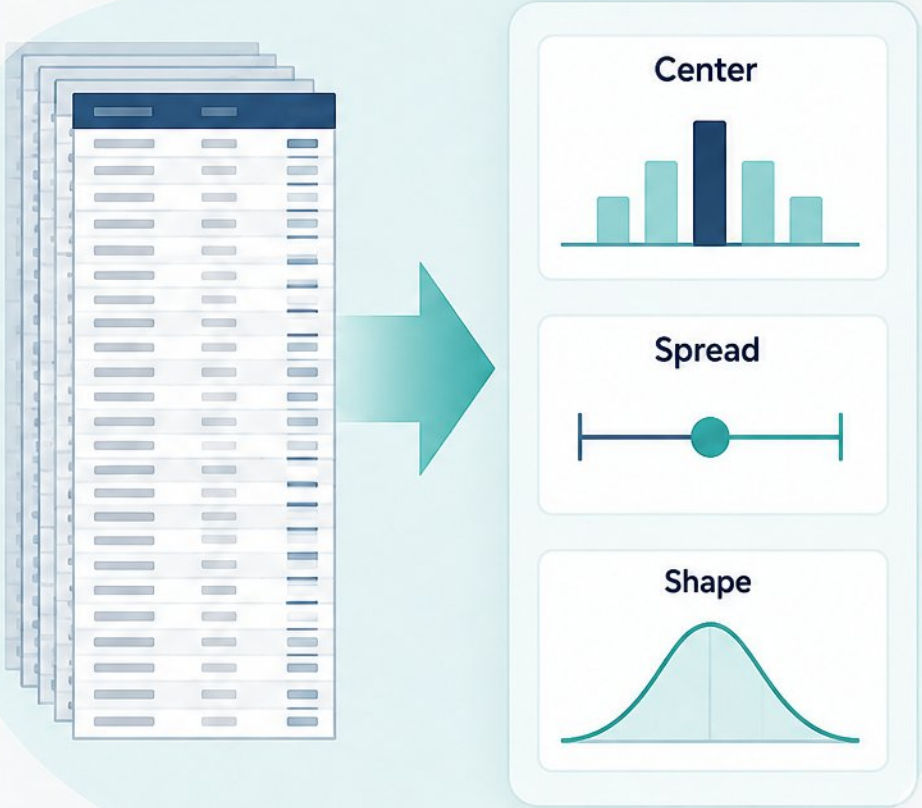
What Descriptive Statistics Are and Are Not

- Summarize observed data with numbers and graphs; no generalization
- Three families: central tendency, spread, and frequency distribution
- Description describes the sample; inference generalizes to a population
- Statistics describe samples; parameters describe populations — wording matters
- Older than inference; central to exploratory data analysis as a first pass



Why Bother Summarizing Data

- Raw rows overwhelm; humans miss patterns beyond a few dozen values
- Every summary answers three questions: center, spread, and shape
- Few numbers make comparisons easy: classes, regions, or time periods
- Used daily: data quality checks, baseline tables, dashboards, decision briefs
- Caution: summaries compress and can hide detail — pair numbers with visuals



Data Types and Measurement Scales

- Categorical vs. numerical: this distinction drives every later choice
- Nominal, ordinal, interval, ratio — department names, ratings, temperature, revenue
- Scale limits valid statistics: no means on nominal labels; medians suit ordinal
- Default for categorical data: frequency tables and crosstabulations
- Check early: missing values, implausible values, miscoded categories, duplicates

Nominal	Ordinal	Interval	Ratio
			
Department names	Ratings	Temperature	Revenue

Measures of Central Tendency

Mean: arithmetic average; efficient for symmetric data but pulled by extreme values

Median: middle value with high breakdown point; preferred for income, latency, skewed data

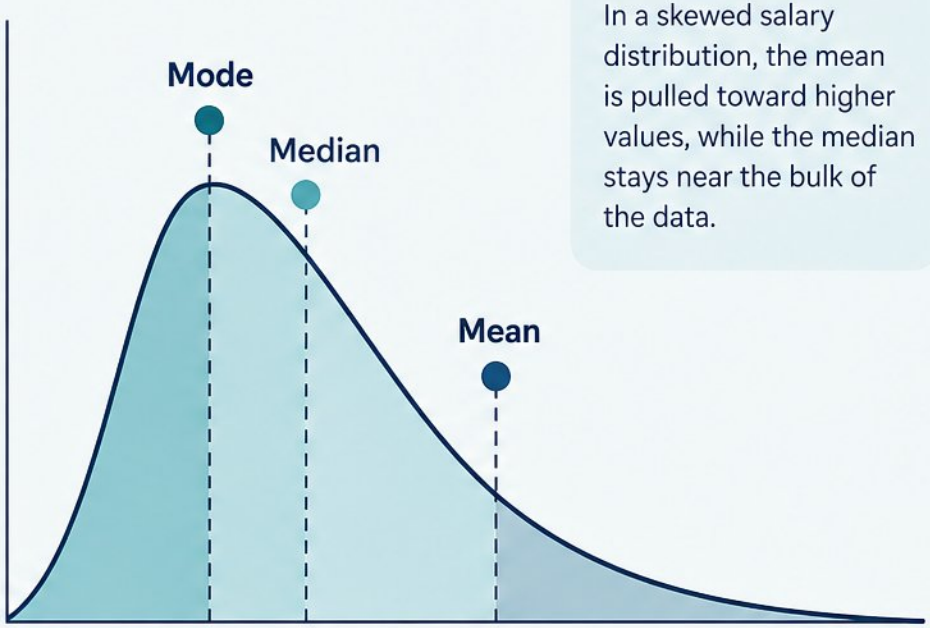
Mode: most frequent value; useful for categorical and multimodal data

Selection rule: symmetric data → mean; skewed data → median; categories → mode

Worked salary example: mean and median tell different stories

Reporting habit: name the measure used; pair it with a spread statistic

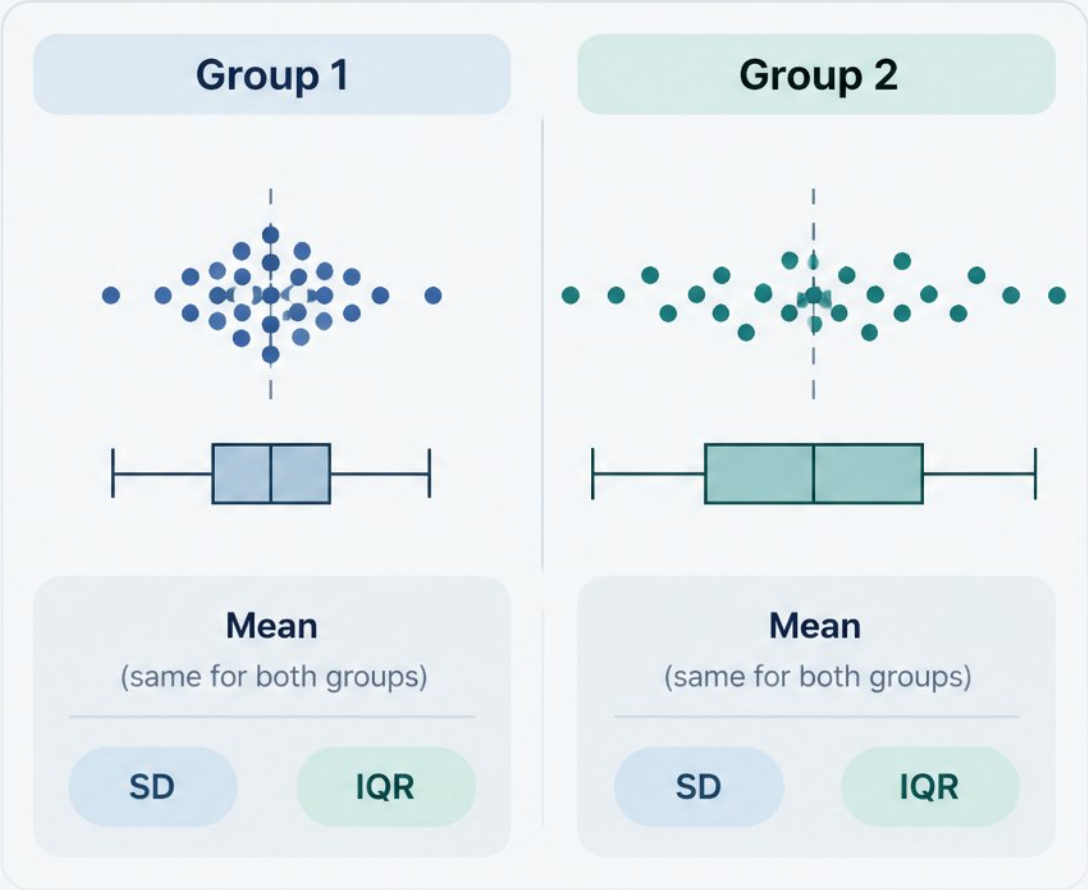
Salary Distribution



In a skewed salary distribution, the mean is pulled toward higher values, while the median stays near the bulk of the data.

Measures of Spread and Variability

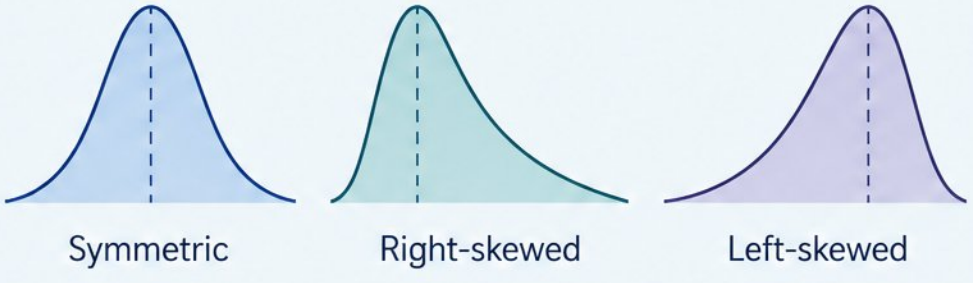
- Range, IQR, variance, SD: what each measures and its units
- How wide is the data, not just where the middle sits
- Same mean, different spread: two groups can behave differently
- Standard deviation: the average distance from the mean
- Coefficient of variation compares spread across different units
- Report mean ± SD, or median with IQR—never a lone center value

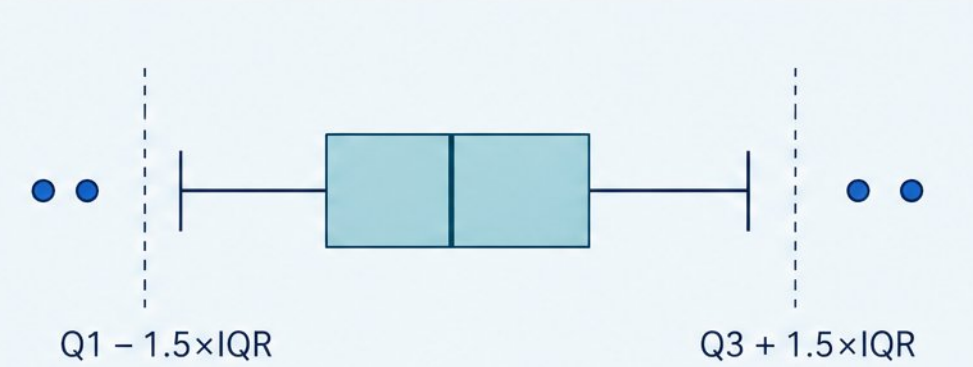


Same mean, different spread:
two groups can behave differently.

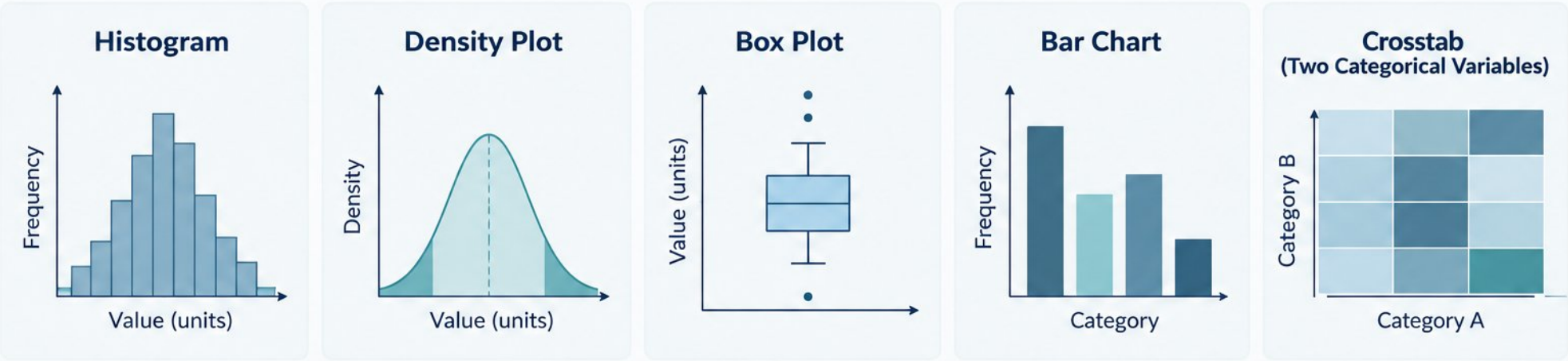
Distribution Shape, Skewness, and Outliers

- Symmetric, right-skewed, left-skewed shapes; kurtosis as a first look
- Skewness decides the summary: mean and SD for symmetric data
- For skewed data, report median with IQR instead
- IQR rule: flag values beyond Q1 − 1.5×IQR or Q3 + 1.5×IQR
- Z-score rule is weaker: mean and SD are themselves outlier-sensitive
- A flagged value is not automatically an error; pre-specify treatment





Visualizing Descriptive Statistics



- Histograms and density plots reveal modality and shape, not just center
- Box plots show median, quartiles, and outliers for group comparisons
- Bar charts for category frequencies; crosstabs for two categorical variables
- Design for honest comparison: consistent scales, no truncated axes, clear units
- Pair every visual with its numeric summary so the story stays unified

A Practical Workflow with a Worked Example





- Inspect, clean, summarize, visualize, report: five repeatable steps
- Customer-spend example: compute n, mean, median, SD, and IQR
- Business reading: typical customer plus how much spend varies
- Pitfalls: mean on skewed data, ignored missing values, excess decimals
- Same workflow in spreadsheets, Python or R, or a BI platform

Reporting Descriptive Results Correctly

Core principles:
clarity, consistency, transparency, accuracy in text, tables, figures

Standard formats:
mean with SD; median with IQR (25th–75th); *n* with percentage

Precision:
two decimals for statistics, one for percentages; always report sample size

Table 1 structure:
groups as columns, characteristics as rows, units stated

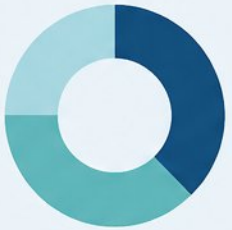
Avoid misleading presentation:
mismatched values, selective rounding, distorted charts

Table 1. Baseline Characteristics

Characteristic	Group 1	Group 2	Group 3
Age, years mean (SD)	XX.XX (XX.XX)	XX.XX (XX.XX)	XX.XX (XX.XX)
BMI, kg/m² median (IQR)	XX.XX (XX.XX–XX.XX)	XX.XX (XX.XX–XX.XX)	XX.XX (XX.XX–XX.XX)
Female, n (%)	XX (XX.X)	XX (XX.X)	XX (XX.X)
Total, n	XX	XX	XX







Matching the Summary to the Audience and Decision



1. Same data, different framing: executive brief gets headline plus uncertainty; appendix keeps full detail
2. Purpose first, then variables: describe for transparency, not routine significance testing
3. Describe results honestly; avoid phrases like "trending toward significance"
4. Add confidence intervals only when estimating a population quantity, not describing the sample
5. Reporting template: sample size, what was measured, measure with spread, key visual, plain takeaway

Guided Practice: Choosing Measures for a Dataset

Scenario A:
ordinal satisfaction survey —
median plus frequency counts



Center measure:

Spread measure:

Chart type:

Scenario B:
right-skewed income —
median and IQR, not the mean



Center measure:

Spread measure:

Chart type:

Scenario C:
bimodal process data —
one center value hides two regimes



Center measure:

Spread measure:

Chart type:

Task: pick center, spread, and chart for each; justify your choice

Decision rule: data type first, shape second, business question third

Guided Practice: Interpreting and Critiquing Descriptive Output

- Read a summary table: state typical case and variability in one sentence
- Spot errors: mean on skewed data, missing n, inconsistent decimals
- Judge outlier decisions: rule pre-specified? effect of exclusion disclosed?
- Critique charts: truncated axis, missing units, visuals contradicting numbers
- Apply a peer-review checklist drawn from reporting guidelines

Summary Table

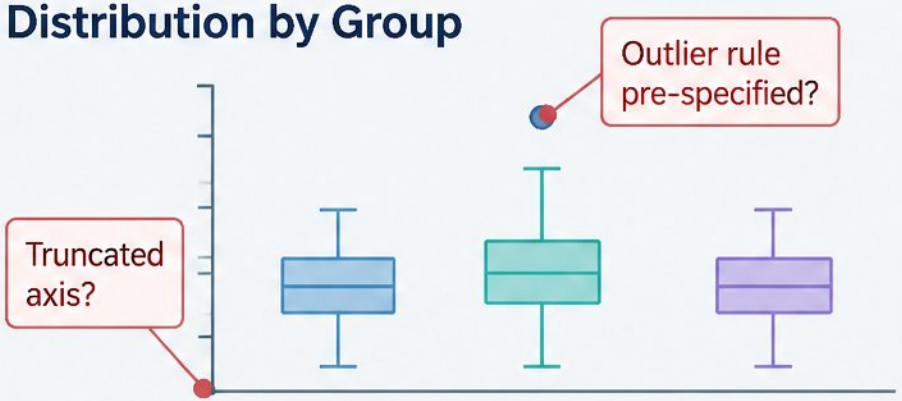
Missing n?

Mean on skewed data?

Inconsistent decimals?

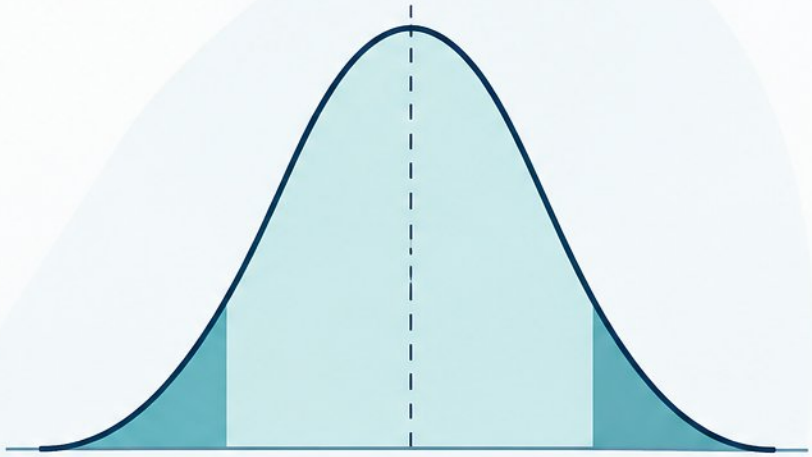
Group	n	Mean	SD	Median	IQR
A					
B					
C					

Distribution by Group



Key Takeaways and Next Steps

1. Descriptive statistics summarize center, spread, and shape of observed data
2. Data type and shape dictate the measure; report center, spread, and n together
3. Pair visuals with numbers; pre-specify outlier handling before analysis
4. Next: inferential statistics, where confidence intervals and tests build on these summaries
5. Try it: run the five-step workflow on a small dataset of your own





PersonWise.

PersonWise • AI digital-human course platform



Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/48d207f9/public