

Chapter 1: Exploring Data — Distributions

AP Statistics • Chapter 1 • Updated March 2026 • ~45 min read

Learning Objectives

- Distinguish between categorical and quantitative variables
- Display distributions using dotplots, stemplots, and histograms
- Describe distribution shape: symmetric, skewed left, skewed right
- Identify center and spread from a graphical display
- Construct and interpret boxplots; identify outliers using the IQR rule
- Compare distributions using appropriate graphical and numerical summaries

1.1 Types of Variables

Statistics begins with **data** — information collected about individuals. Before analyzing data, we must identify the type of variable being measured, since different variable types require different methods.

Definition: Types of Variables

A **categorical variable** (also called qualitative) places each individual into one of several groups or categories. Examples: eye color, gender, country of birth, AP exam score (A/B/C/D/F).

A **quantitative variable** takes numerical values for which arithmetic makes sense. Examples: height in cm, SAT score, temperature, number of siblings.

Quantitative variables can be further divided:

- **Discrete**: takes countable values (e.g., number of pets: 0, 1, 2, 3, ...)
- **Continuous**: can take any value in an interval (e.g., time, weight, length)

Example 1.1 — Identifying Variable Types

A survey of 30 AP Statistics students records the following information. Classify each variable.

Variable	Type	Reason
Favorite subject	Categorical	Places student in a category (Math, English, ...)
Hours studied per week	Quantitative (continuous)	Numerical, arithmetic makes sense
Number of AP exams taken	Quantitative (discrete)	Countable whole numbers
Grade in AP Stats (A/B/C)	Categorical	Letter grades are categories, not numbers

TRY IT

A researcher records: (a) blood type of each patient, (b) systolic blood pressure, (c) number of hospitalizations. Classify each variable.

1.2 Displaying Distributions with Graphs

To understand a dataset, we start by making a graph. The graph reveals the **distribution** of a variable — what values occur and how often.

Dotplots

A **dotplot** places each data value as a dot above a number line. Dotplots work well for small datasets and show individual values clearly.

Example 1.2 — Reading a Dotplot

The number of text messages sent by 12 students in one hour: **3, 5, 5, 7, 8, 8, 8, 10, 12, 12, 15, 20**

Each value gets one dot. Stacked dots indicate repeated values. We can see immediately that most students sent 5–12 messages, with one outlier at 20.

Histograms

A **histogram** divides the range of data into equal-width intervals (called *bins*) and displays the count or percent of observations in each bin. Histograms work well for large datasets.

How to Construct a Histogram

1. Choose a convenient number of bins (typically 5–10)
2. Make the bins equal in width, covering the full range
3. Count the observations in each bin
4. Draw bars of height = frequency (or relative frequency); bars touch each other

Describing Shape

When you look at a histogram (or any distribution graph), describe its **shape**:

Distribution Shapes

- **Symmetric**: left and right sides are roughly mirror images; the mean \approx median
- **Skewed right** (positively skewed): long tail extends to the right; mean $>$ median
- **Skewed left** (negatively skewed): long tail extends to the left; mean $<$ median
- **Unimodal**: one peak; **Bimodal**: two peaks; **Uniform**: roughly flat

- ★ **AP Exam Tip:** When describing a distribution, always address **S**hape, **C**enter, **S**pread, and any **O**utliers (SCSO or "SOCS"). Free-response graders look for all four components.

1.3 Boxplots and the Five-Number Summary

A **boxplot** (box-and-whisker plot) summarizes a distribution using five key values called the **five-number summary**: Minimum, Q1, Median (Q2), Q3, Maximum.

Five-Number Summary

Given a dataset sorted in order:

- **Minimum:** smallest value
- **Q1** (first quartile): median of the lower half of the data
- **Median (Q2):** middle value (or average of two middle values)
- **Q3** (third quartile): median of the upper half of the data
- **Maximum:** largest value

The **Interquartile Range (IQR)** = $Q_3 - Q_1$ measures the spread of the middle 50% of data.

Example 1.3 — Computing the Five-Number Summary

AP exam scores for 15 students (sorted):

1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5

Step 1 — Median: The 8th value = **4**

Step 2 — Q1: Lower half = {1, 2, 2, 3, 3, 3, 4}; median = **3**

Step 3 — Q3: Upper half = {4, 4, 4, 5, 5, 5, 5}; median = **5**

Five-number summary: Min = 1, Q1 = 3, Median = 4, Q3 = 5, Max = 5

IQR = Q3 - Q1 = 5 - 3 = 2

Identifying Outliers

The $1.5 \times \text{IQR}$ Rule for Outliers

An observation is a **suspected outlier** if it falls:

- Below $Q_1 - 1.5 \times \text{IQR}$, or
- Above $Q_3 + 1.5 \times \text{IQR}$

On a modified boxplot, outliers are plotted as individual points; whiskers extend only to the last non-outlier value.

TRY IT

A dataset has $Q_1 = 12$, $Q_3 = 20$. Calculate the IQR and the outlier fences.

1.4 Comparing Distributions

A common AP Statistics task is to **compare two or more distributions**. Use side-by-side boxplots or back-to-back stemplots. Always compare shape, center, spread, and outliers in context.

Example 1.4 — Comparing Two Distributions

Two classes take the same quiz. Class A: min=52, Q1=68, median=74, Q3=82, max=96. Class B: min=60, Q1=72, median=80, Q3=85, max=92.

Center: Class B has a higher median (80 vs 74), suggesting Class B performed better on average.

Spread: Class A has a larger IQR ($82 - 68 = 14$) vs Class B ($85 - 72 = 13$), so Class A is slightly more variable.

Shape: Both distributions appear roughly symmetric based on the summary values.

Outliers: No outliers visible from the five-number summary.

★ **AP Exam Tip:** When comparing distributions, always write comparisons in **context** and use **comparative language** ("Class B's median is higher than Class A's median"). Simply listing each distribution's statistics without comparing earns partial credit only.

Practice Problems

Problem 1

A sample of 10 students recorded how many hours they sleep per night: **6, 7, 7, 8, 8, 8, 9, 9, 10, 12**

- Find the five-number summary.
- Calculate the IQR.
- Identify any outliers using the $1.5 \times \text{IQR}$ rule.

► [Show Solution](#)

Problem 2

A histogram shows that the distribution of household incomes in a city is **strongly skewed right**.

- (a) What does the skewed-right shape tell us about most households vs. a few households?
- (b) Would you expect the mean income to be greater than or less than the median income? Explain.

► [Show Solution](#)

Problem 3

Classify each variable as categorical or quantitative:

- (a) ZIP code
- (b) Annual rainfall in mm
- (c) Shirt size (S/M/L/XL)
- (d) Number of siblings

► [Show Solution](#)

Problem 4 — AP Free Response Style

Two competing tutoring programs (Program A and Program B) report the following SAT Math score gains for a sample of students:

Program A: Min=20, Q1=40, Median=60, Q3=80, Max=150

Program B: Min=30, Q1=50, Median=65, Q3=75, Max=100

Compare the distributions of score gains for the two programs. Write a complete response using the SOCS framework.

► [Show Solution](#)

Problem 5 — Multiple Choice Style

A distribution has $Q1 = 45$ and $Q3 = 65$. Which of the following values would be classified as an outlier?

(A) 20 (B) 35 (C) 70 (D) 80

► [Show Solution](#)

Problem 6

A dataset of 20 values is described: the mean is 55 and the median is 42. What does this tell you about the shape of the distribution? Explain your reasoning.

► [Show Solution](#)

Problem 7

Identify an appropriate graph for each situation:

- (a) Display the distribution of birth months (Jan–Dec) for 50 students
- (b) Compare the heights of male and female students in a class of 60
- (c) Show the distribution of 200 SAT scores

► [Show Solution](#)

Problem 8 — Challenge

A dataset has the property that the mean, median, and mode are all equal.

- (a) What shape does the distribution likely have?
- (b) Give a specific example of such a dataset with 5 values.

► [Show Solution](#)

Chapter Summary

TYPES OF DATA

Categorical Variable

Records which group or category an individual belongs to. Examples: gender, color, region. Summarized with frequency tables and bar charts.

Quantitative Variable

Records numerical values where arithmetic makes sense. Examples: height, temperature, income. Summarized with histograms, dotplots, boxplots.

Distribution

Describes the pattern of values: shape (symmetric, skewed, bimodal), center (mean/median), spread (range/IQR/SD), and outliers.

Comparing Distributions

Use parallel boxplots or back-to-back stemplots. Compare shape, center, spread, and outliers in context. Always use comparative language.

GRAPH TYPES

Histogram

Groups quantitative data into intervals (bins). Shows shape clearly. Use for large datasets.

Boxplot

Shows the five-number summary: min, Q1, median, Q3, max. Outliers plotted individually beyond $1.5 \times \text{IQR}$ from Q1/Q3.

Dotplot / Stemplot

Shows every data value. Useful for small datasets to see exact values and identify gaps or clusters.

Bar Chart

For categorical data. Bars represent frequency or relative frequency for each category. Bars should NOT touch.

SHAPE DESCRIPTIONS

1. **Symmetric** — roughly mirror-image on both sides of center
2. **Skewed right** — tail extends to the right; mean > median
3. **Skewed left** — tail extends to the left; mean < median

4. Unimodal / Bimodal — one or two distinct peaks

Key Terms

Individual

A person or object described by data.

Variable

A characteristic that takes different values for different individuals.

Categorical Variable

Places individuals into groups or categories; values are labels, not numbers.

Quantitative Variable

Takes numerical values with a meaningful scale; arithmetic makes sense.

Distribution

Pattern of values in data — shape, center, spread, and any outliers.

Skewness

Asymmetry in a distribution. Right-skewed: long tail right; left-skewed: long tail left.