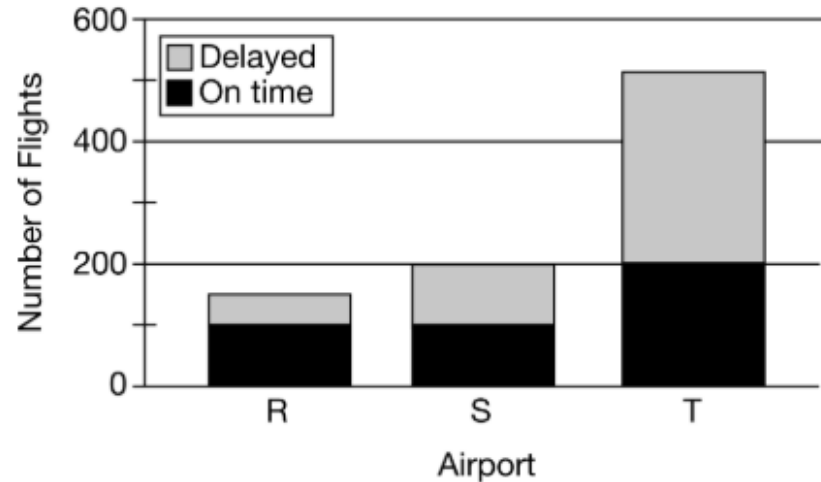


Representing Two Categorical Variables

The following segmented bar chart shows the number of flights that were either on time or delayed at three different airports on one day.



Which of the following statements is supported by the bar chart?

- (A)** Airport T has the greatest percentage of on-time flights compared to the other two airports.
- (B)** Airport R has the least percentage of on-time flights compared to the other two airports.
- (C)** The number of on-time flights at Airport S is half the number of on-time flights at Airport T.
- (D)** The number of on-time flights at Airport R is less than the number of on-time flights at Airport S.
- (E)** The number of flights at Airport T is equal to the total number of flights at Airports R and S combined.

A survey of 100 students records their **favorite type of music** (Pop, Rock, Classical) and whether they **own a musical instrument** (Yes, No). The results are summarized in a two-way table.

- (a) Identify the **row variable** and **column variable**.
 - (b) Calculate the **marginal distribution** of students who own a musical instrument.
 - (c) Describe the **conditional distribution** of music preference for students who own a musical instrument.
-

Example

A survey asked 100 students about *gender* (male/female) and whether they *prefer sports* (yes/no). The results are shown:

	Sports: Yes	Sports: No	Total
Male	30	10	40
Female	20	40	60
Total	50	50	100

Is there an association between gender and sports preference?

Researchers study exercise frequency (low/medium/high) among smokers vs. non-smokers. Each group is shown as a segmented bar that totals 100%.

How does the segmented bar graph show association?

Stretch Exercise (AP Statistics level+)

A school surveys 300 students about **whether they play sports** and **whether they prefer online or in-person classes**.

	Online	In-person	Total
Play sports	30	120	150
Don't play sports	90	60	150
Total	120	180	300

(a) Calculate the **conditional distribution** of class preference for:

- students who play sports
- students who don't play sports

(b) Based on your results, describe the **association** between playing sports and class preference.

(c) A student claims:

"Playing sports causes students to prefer in-person classes."

Is this justified? Explain.

(d) Suppose the data were collected separately for **Grade 11 and Grade 12**, and in both grades there is little to no association, but in the combined table there is a strong association.

- What phenomenon could explain this?
- Briefly explain how this is possible.

(e) Without doing calculations, explain how the conclusion might change if the table showed **relative frequencies instead of counts**.

(a) By the CLT, the sampling distribution of the sample mean is approximately normal because $n = 36 \geq 30$, even though the population is skewed.

(b) Standard error: $SE = \sigma / \sqrt{n} = 12 / 6 = 2$

(c) $z = (47 - 50) / 2 = -1.5 \rightarrow P(X < 47) \approx 0.067$ (6.7%)

(d) For $P < 0.01$: $z = -2.33 \rightarrow n \approx 87$