

# Representing Categorical Data with Graphs

1. Which type of graph is most appropriate for displaying categorical data?

- A. Histogram
  - B. Scatter plot
  - C. Bar chart
  - D. Line graph
- 

2. A pie chart is mainly used to show:

- A. Relationships between two variables
- B. Changes over time
- C. Parts of a whole
- D. Continuous intervals of data

# Representing Categorical Data with Graphs

1. Which type of graph is most appropriate for displaying categorical data?

- A. Histogram
- B. Scatter plot
- C. Bar chart
- D. Line graph

C

---

2. A pie chart is mainly used to show:

- A. Relationships between two variables
- B. Changes over time
- C. Parts of a whole
- D. Continuous intervals of data

C

- Categorical data are best displayed using **bar graphs** or **pie charts**.
  - Bar graphs show counts or percentages for each category.
  - Pie charts show how each category contributes to the whole.
- Bars should have **equal width and separated gaps** (unlike histograms).
- Graphs should be clearly labeled with an appropriate title and scale.

Categorical variable

---

分类变量

---

Bar graph

---

条形图

---

Pie chart

---

饼图

---

Frequency

---

频数

---

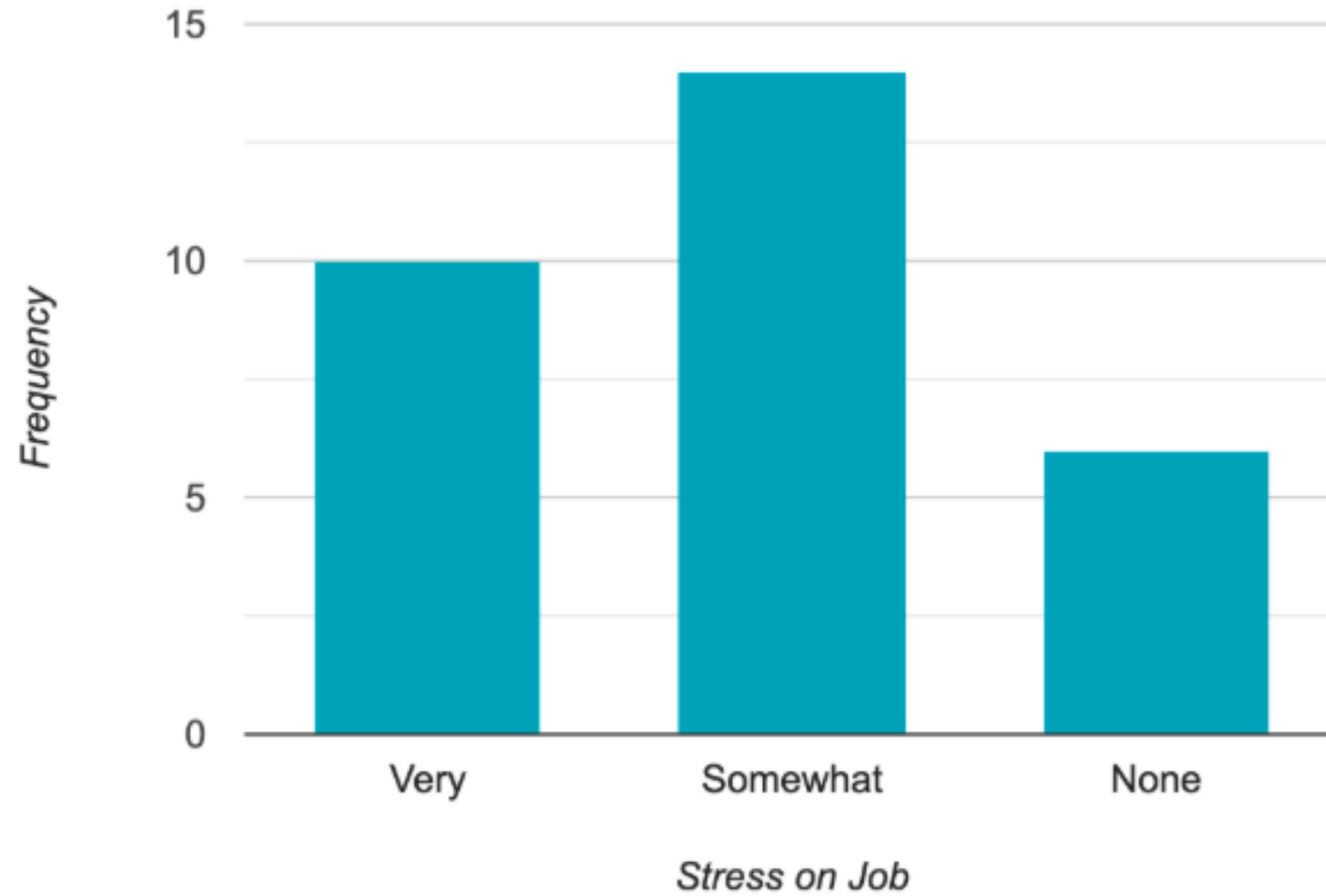
Relative frequency

相对频率

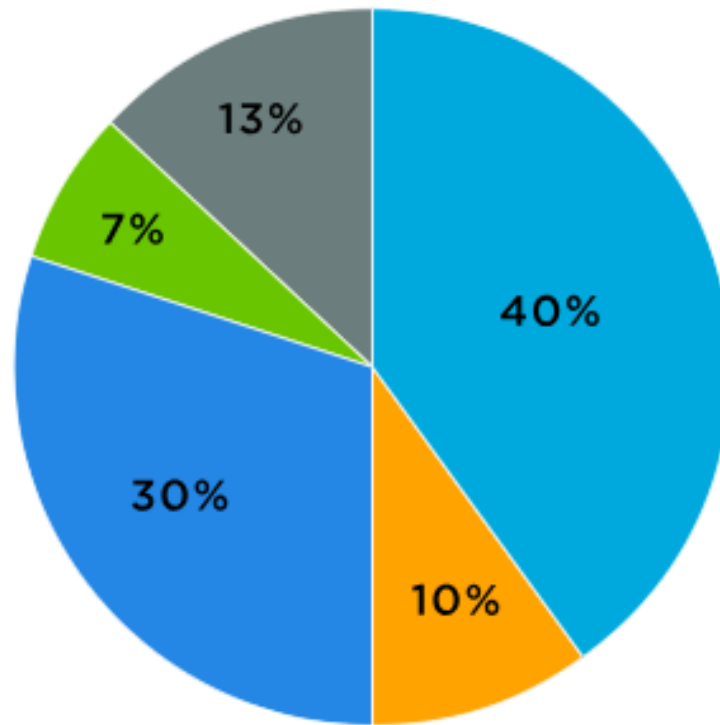
## Mnemonic: **BLCS** — **B**ars, **L**abels, **C**ompare, **S**pecific

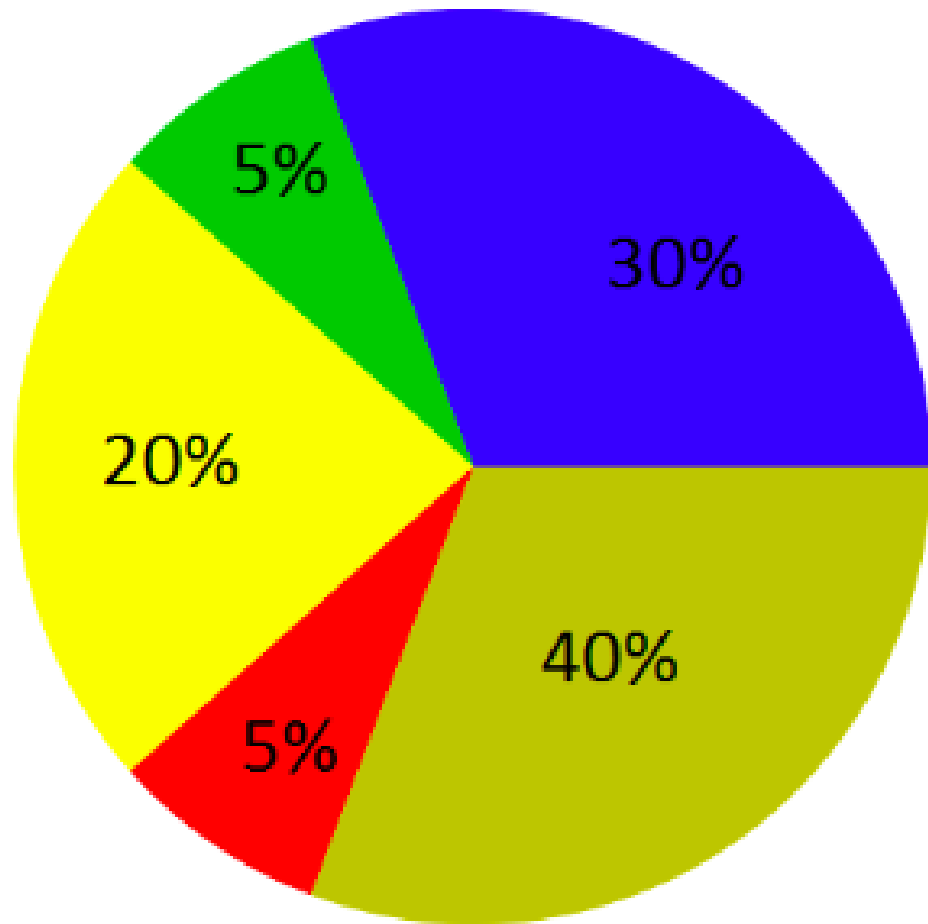
- **B:** Choose the right type of bar graph (regular for single group, side-by-side or segmented for multiple groups)
- **L:** Label all axes and include a title — missing labels cost points
- **C:** Compare using relative frequencies when group sizes differ
- **S:** Be specific in every interpretation — cite actual numbers

# Bar Chart

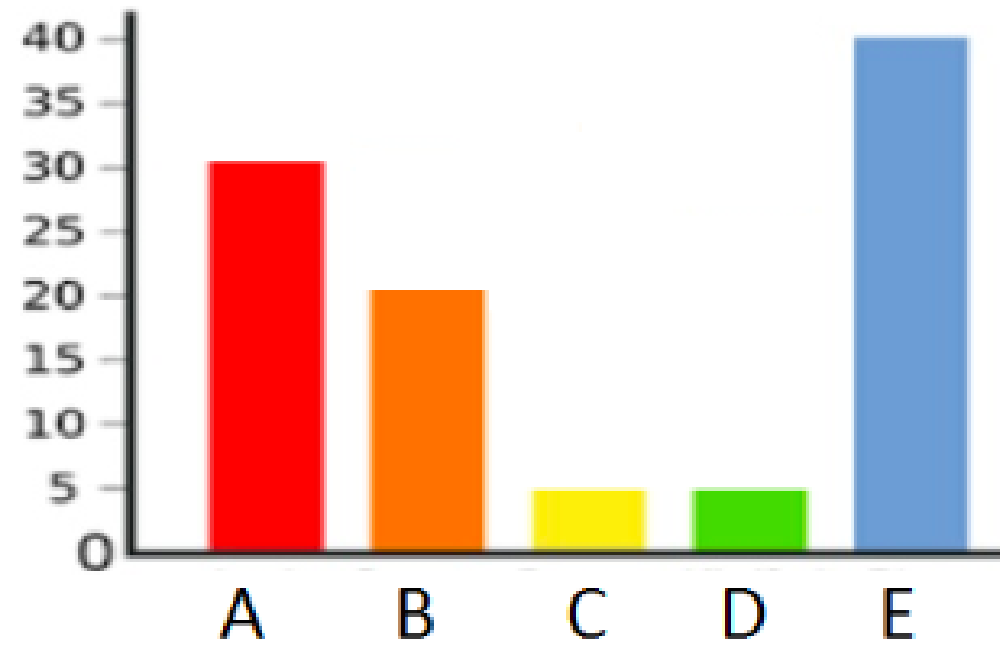


A **pie chart** is a circular graph that is divided into slices, with each slice representing a different category. The size of each slice is proportional to the fraction of the whole that is represented by that category. Pie charts are often used to show the relative proportions of different categories within a dataset.





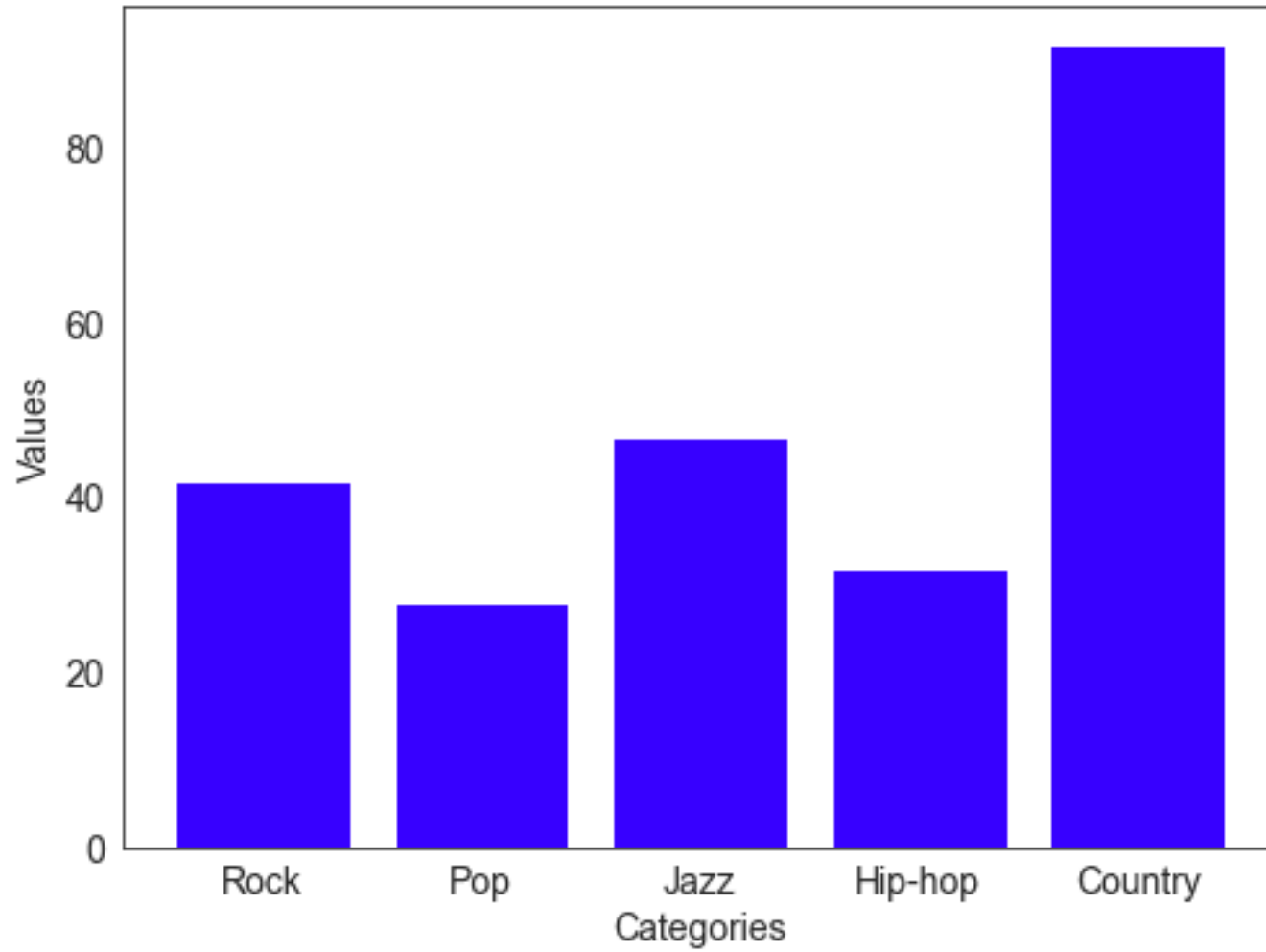
## Categorical Data



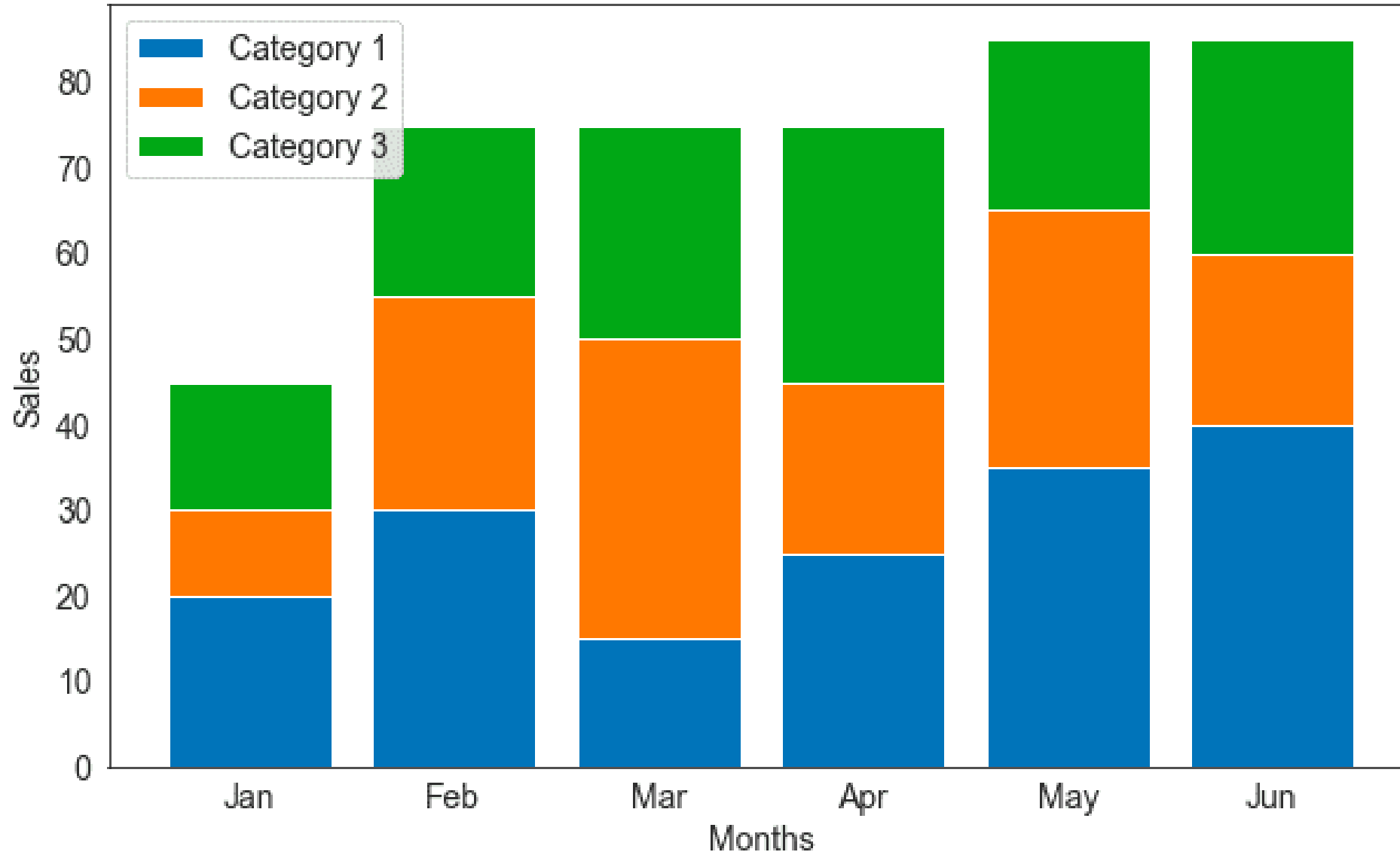
# 4. Stacked Bar Chart



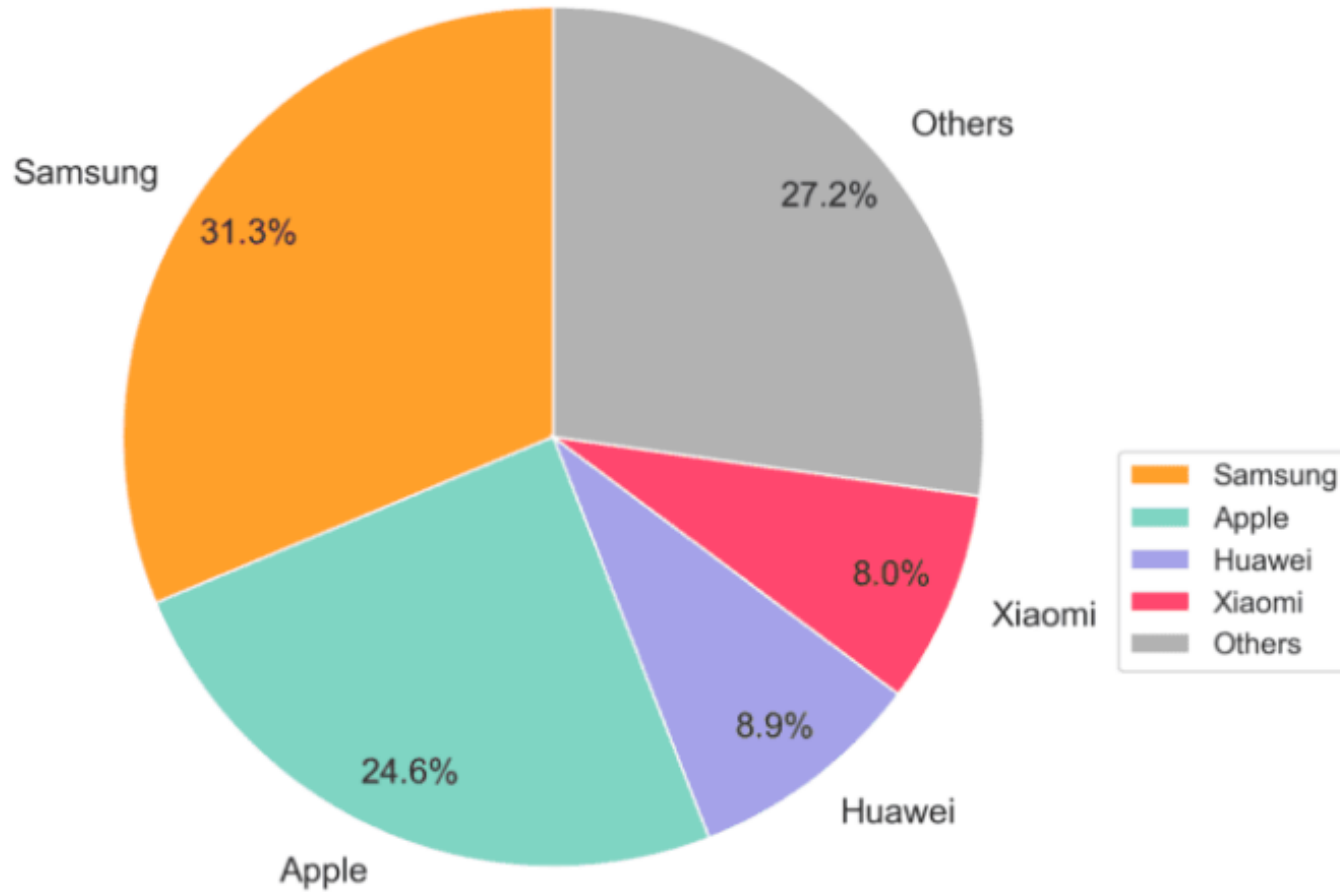
Frequency of Music Preferences



### Monthly Sales by Product Category

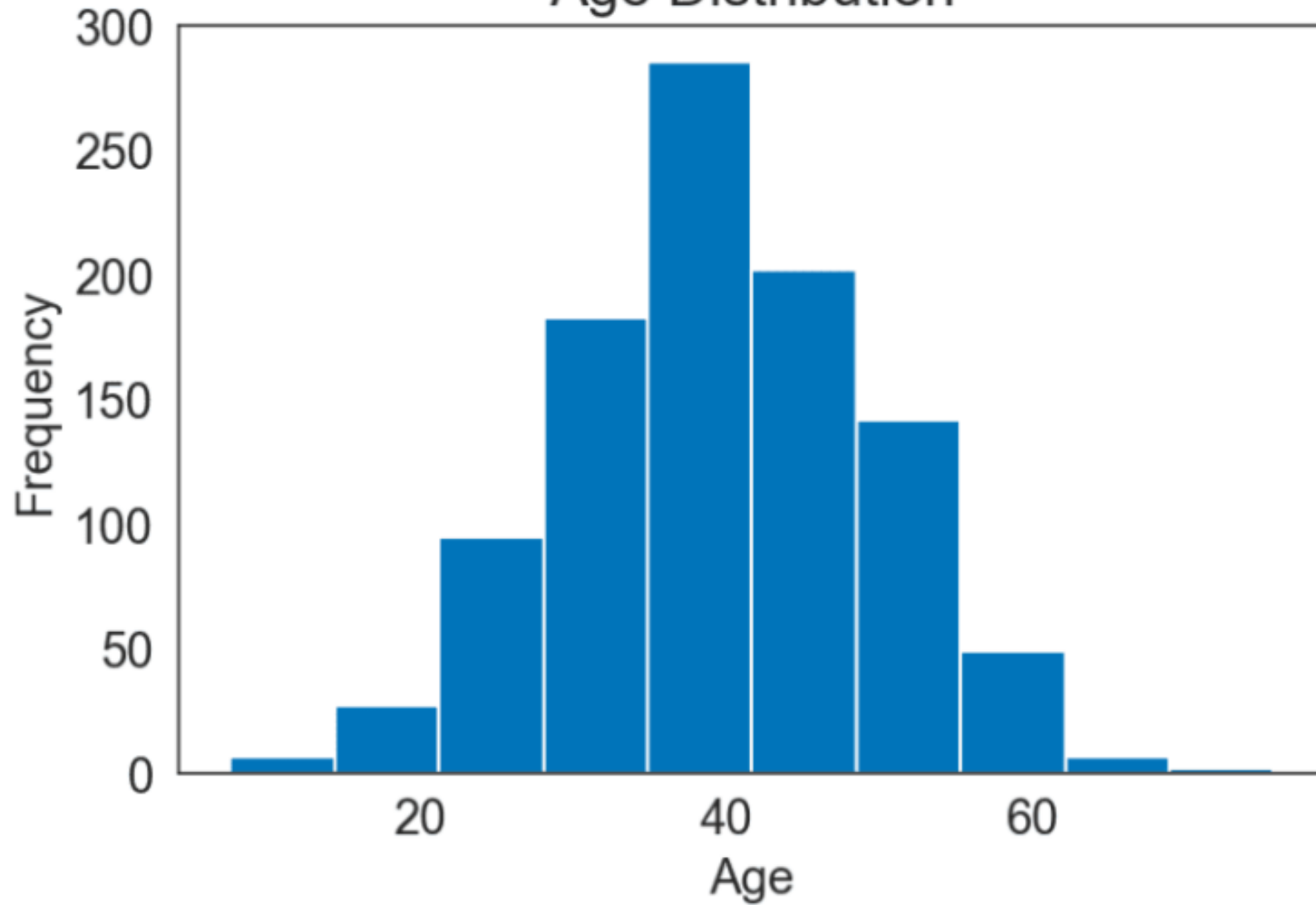


Market Share of Smartphone Brands



T/F Pie charts are usually better than bar charts for comparing very small differences.

### Age Distribution



T/F A frequency graph can help identify patterns such as skewness or clusters.

A school surveys 200 students and records their preferred mode of transportation to school: **bus, car, bike, or walk.**

- (a)** Name the type of variable being measured.
- (b)** Identify the most appropriate graph to display this data and explain why.
- (c)** Explain why a histogram would *not* be appropriate for this data.

(a) The variable is a **categorical variable** because it classifies students into distinct groups.

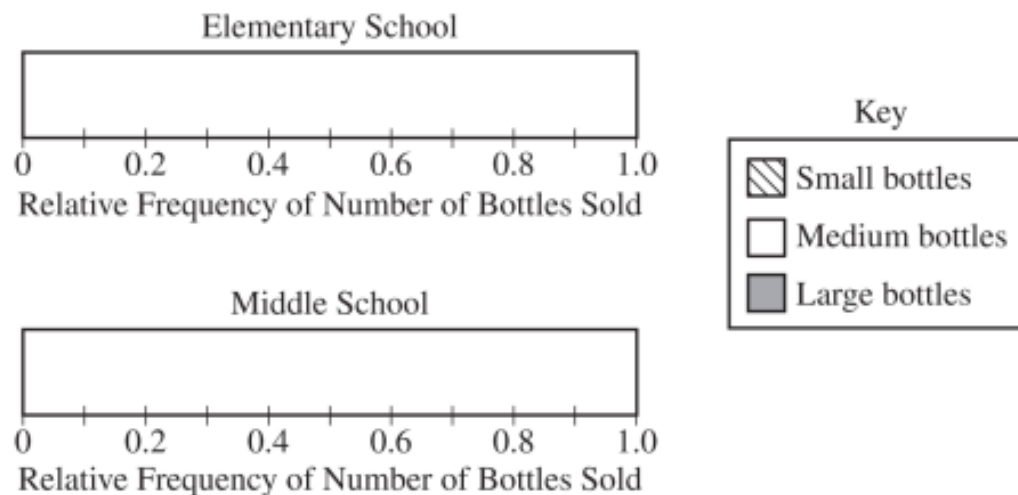
(b) A **bar graph** is the most appropriate because it displays and compares frequencies or relative frequencies for each category.

(c) A histogram is not appropriate because histograms are used for **quantitative variables**, not categorical variables.

1. A local elementary school decided to sell bottles printed with the school district's logo as a fund-raiser. The students in the elementary school were asked to sell bottles in three different sizes (small, medium, and large). The relative frequencies of the number of bottles sold for each size by the elementary school were 0.5 for small bottles, 0.3 for medium bottles, and 0.2 for large bottles.

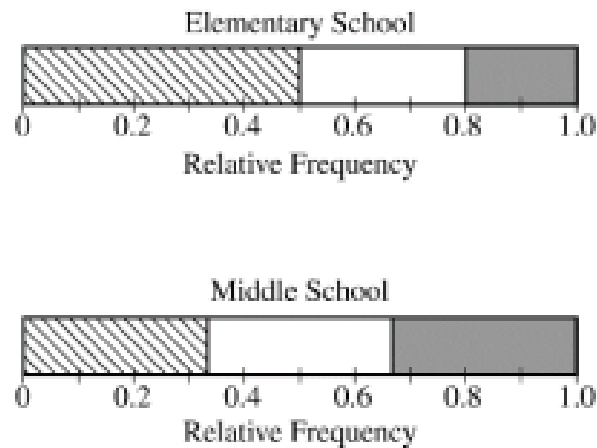
A local middle school also decided to sell bottles as a fund-raiser, using the same three sizes (small, medium, and large). The middle school students sold three times the number of bottles that the elementary school students sold. For the middle school students, the proportion of bottles sold was equal for all three sizes.

- (a) Complete the segmented bar graphs representing the relative frequencies of the number of bottles sold for each size by students at each school.



- (b) An administrator at the elementary school concluded that the elementary school students sold more small bottles than the middle school students did. Is the elementary school administrator's conclusion correct? Explain your response.

(a)



(b)

No, the segment for small bottles for the elementary school is wider than the segment for small bottles for the middle school; however, the middle school students sold three times as many bottles as the elementary school students. So, if the elementary school sold  $x$  number of bottles, the middle school sold  $3x$  number of bottles.

For example, if the elementary students sold 100 bottles total, then they sold  $0.5(100)$  or 50 small bottles. However, because the middle school students sold three times the total number of bottles as the elementary students, they would have sold 300 bottles total and

$(0.\bar{3})(300) = 100$  small bottles. Because

$100 > 50$ , the middle school sold more small bottles, and the elementary school's administrator is not correct.

## Stretch & Challenge Questions: Representing Categorical Data with Graphs (AP Statistics)

### 1. Misleading Graph Analysis

A company displays a bar chart showing sales of three products (A, B, C). The y-axis starts at 80 instead of 0, and Product B appears "twice as large" as Product A.

- Explain why this graph may be misleading.
  - Redraw (describe) how a corrected version should look.
- 

### 2. Choosing the Best Graph

A survey records students' favourite transport methods to school (walk, bus, car, bike).

- Compare when a bar chart, pie chart, and segmented bar chart would each be most appropriate.
  - Justify which graph best shows differences between two schools.
- 

### 3. Two-Way Table Interpretation

Two schools recorded students' preferred study method:

	Group Study	Solo Study
School A	40	60
School B	30	70

- ↳ Construct a relative frequency table.
  - ↳ Determine whether there is an association between school and study preference.
  - ↳ Explain your reasoning using percentages.
- 

### 4. Real-World Design Challenge

A café wants to display customer drink preferences (coffee, tea, juice, water) to investors.

- ↳ Design the most effective graph(s) to present this data clearly.
- ↳ Explain why your chosen graph improves communication compared to at least one alternative.

## Stretch & Challenge Questions: Representing Categorical Data with Graphs (AP Statistics)

### 1. Misleading Graph Analysis

A company displays a bar chart showing sales of three products (A, B, C). The y-axis starts at 80 instead of 0, and Product B appears “twice as large” as Product A.

- Explain why this graph may be misleading.
  - Redraw (describe) how a corrected version should look.
- 

### 2. Choosing the Best Graph

A survey records students’ favourite transport methods to school (walk, bus, car, bike).

- Compare when a bar chart, pie chart, and segmented bar chart would each be most appropriate.
  - Justify which graph best shows differences between two schools.
- 

### 3. Two-Way Table Interpretation

Two schools recorded students’ preferred study method:

	Group Study	Solo Study
School A	40	60
School B	30	70

- ↳ Construct a relative frequency table.
  - ↳ Determine whether there is an association between school and study preference.
  - ↳ Explain your reasoning using percentages.
- 

### 4. Real-World Design Challenge

A café wants to display customer drink preferences (coffee, tea, juice, water) to investors.

- ↳ Design the most effective graph(s) to present this data clearly.
- ↳ Explain why your chosen graph improves communication compared to at least one alternative.

## 1. Misleading Graph Analysis

- Y-axis not starting at 0 exaggerates differences.
  - Makes comparisons visually distorted.
  - Correct version: bar chart with true scale starting at 0 and equal intervals.
- 

## 2. Choosing the Best Graph

- Bar chart: best for simple comparison of categories.
  - Pie chart: shows proportion of whole.
  - Segmented bar chart: best for comparing two schools.
  - Best choice: segmented bar chart (direct comparison of distributions).
- 

## 3. Two-Way Table Interpretation

Relative frequencies:

- School A: Group 40%, Solo 60%
  - School B: Group 30%, Solo 70%
- Conclusion: slight association (School A more group study, School B more solo study).
- 

## 4. Real-World Design Challenge

- Best graph: bar chart or segmented bar chart.
- Bar chart shows clear comparison of drink counts.
- Segmented bar chart best if comparing groups (e.g., age or time).
- Better than pie chart because comparisons across categories are clearer.

# One variable data- descriptive statistics

1. Distribution (分布)
2. Mean (平均数)
3. Median (中位数)
4. Mode (众数)
5. Range (极差)
6. Standard Deviation (标准差)
7. Variance (方差)
8. Outlier (离群值)
9. Quartiles (四分位数)
10. Interquartile Range (IQR) (四分位距)

1. Data distributions describe how values are spread and shaped.
2. Center is measured using mean or median depending on symmetry.
3. Spread describes variability in a dataset.
4. Outliers can strongly affect summary statistics.
5. Skewness describes asymmetry in a distribution.
6. Symmetric distributions have similar left and right sides.
7. The IQR measures the middle 50% of data.
8. Standard deviation measures average distance from the mean.
9. Graphs such as histograms and boxplots display distributions.
10. Choosing appropriate measures depends on data shape and context.

A teacher collected the test scores (out of 100) of 12 students:

**52, 65, 70, 70, 74, 78, 80, 83, 85, 88, 92, 96**

- (a)** Construct a suitable graphical display for the data.
- (b)** Describe the shape, center, and spread of the distribution.
- (c)** Identify any outliers using the IQR method.
- (d)** Calculate and interpret the median score.
- (e)** Would the mean or median be a better measure of center? Justify your answer.

**(a)**

A histogram or boxplot is appropriate. (Data are quantitative and univariate.)

**(b)**

Shape: slightly left-skewed (tail toward lower scores).

Center: around 78–80.

Spread: from 52 to 96 (range = 44).

**(c)**

$Q1 = 70$ ,  $Q3 = 85$

$IQR = 15$

Lower fence =  $70 - 1.5(15) = 47.5$

Upper fence =  $85 + 1.5(15) = 107.5$

→ No outliers.

**(d)**

Median =  $(78 + 80) / 2 = 79$

Interpretation: 50% of students scored below 79 and 50% above 79.

**(e)**

Median is better because the data is slightly skewed and more robust to extreme values.

# TRUE or FALSE

- A bar chart is commonly used to represent categorical data.

# TRUE or FALSE

- In a pie chart, each category is shown as a proportion of the whole.

# TRUE or FALSE

- Histograms are the best graph for representing categorical data.