1. Which graph is most appropriate for displaying the distribution of a single quantitative variable?

A. Pie chart

B. Bar chart

C. Histogram

D. Scatterplot

---

2. Which statement is true about a boxplot?

A. It shows individual data points clearly

B. It displays categories and proportions

C. It shows median, quartiles, and possible outliers

D. It is best for showing relationships between two variables

1. Which graph is most appropriate for displaying the distribution of a single quantitative variable?

A. Pie chart

B. Bar chart

C. Histogram

D. Scatterplot

C

---

2. Which statement is true about a boxplot?

A. It shows individual data points clearly

B. It displays categories and proportions

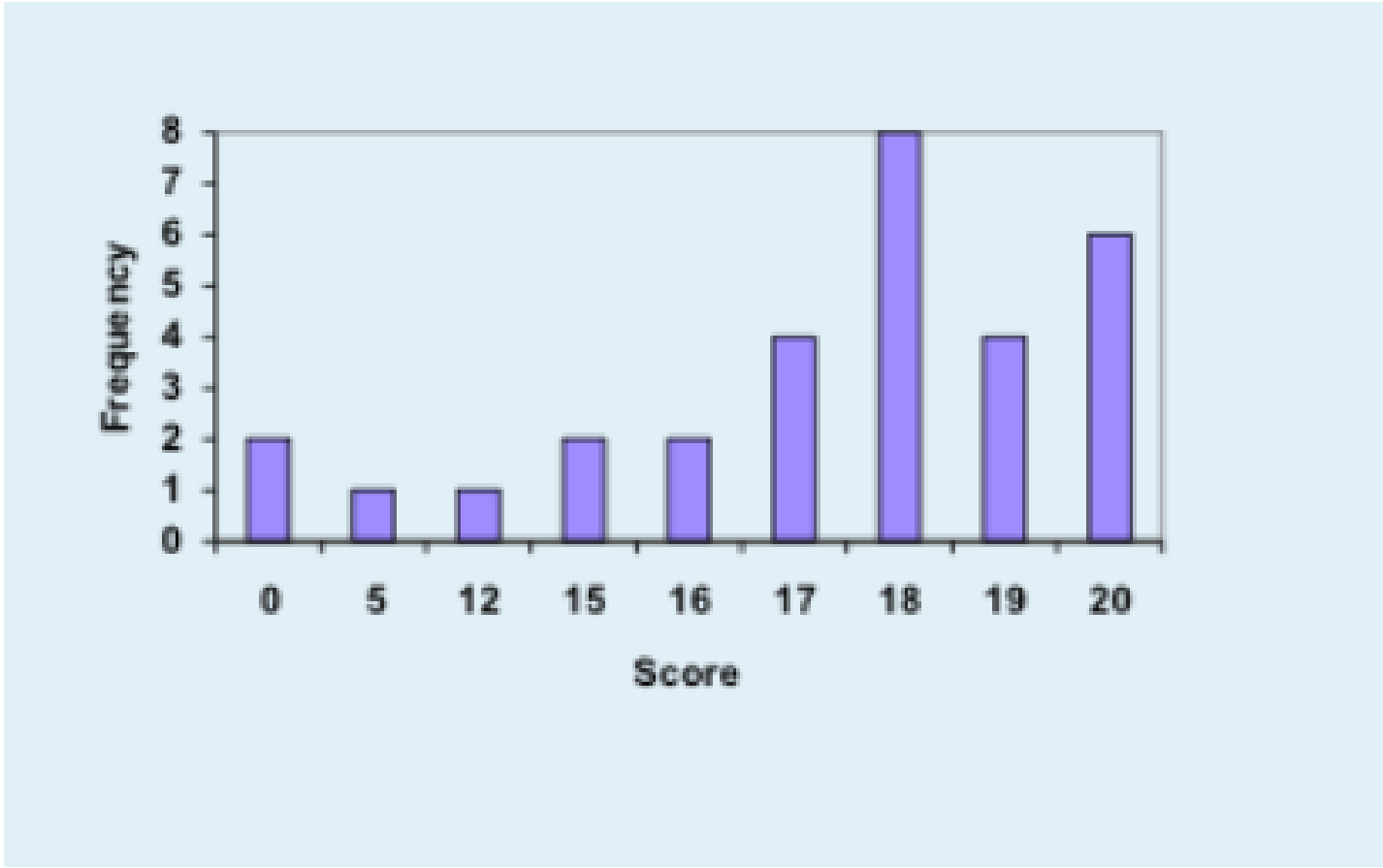
C. It shows median, quartiles, and possible outliers

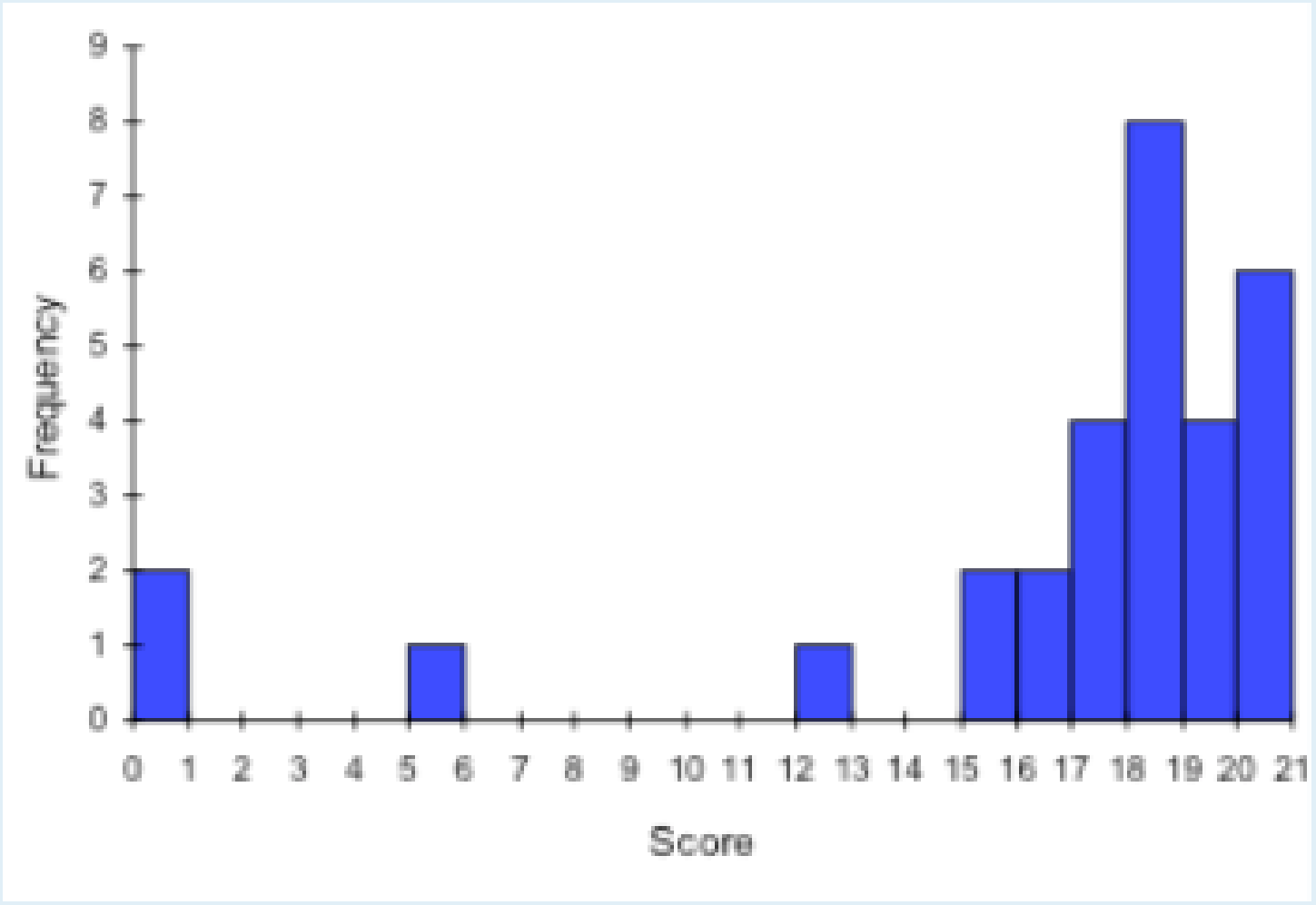
D. It is best for showing relationships between two variables

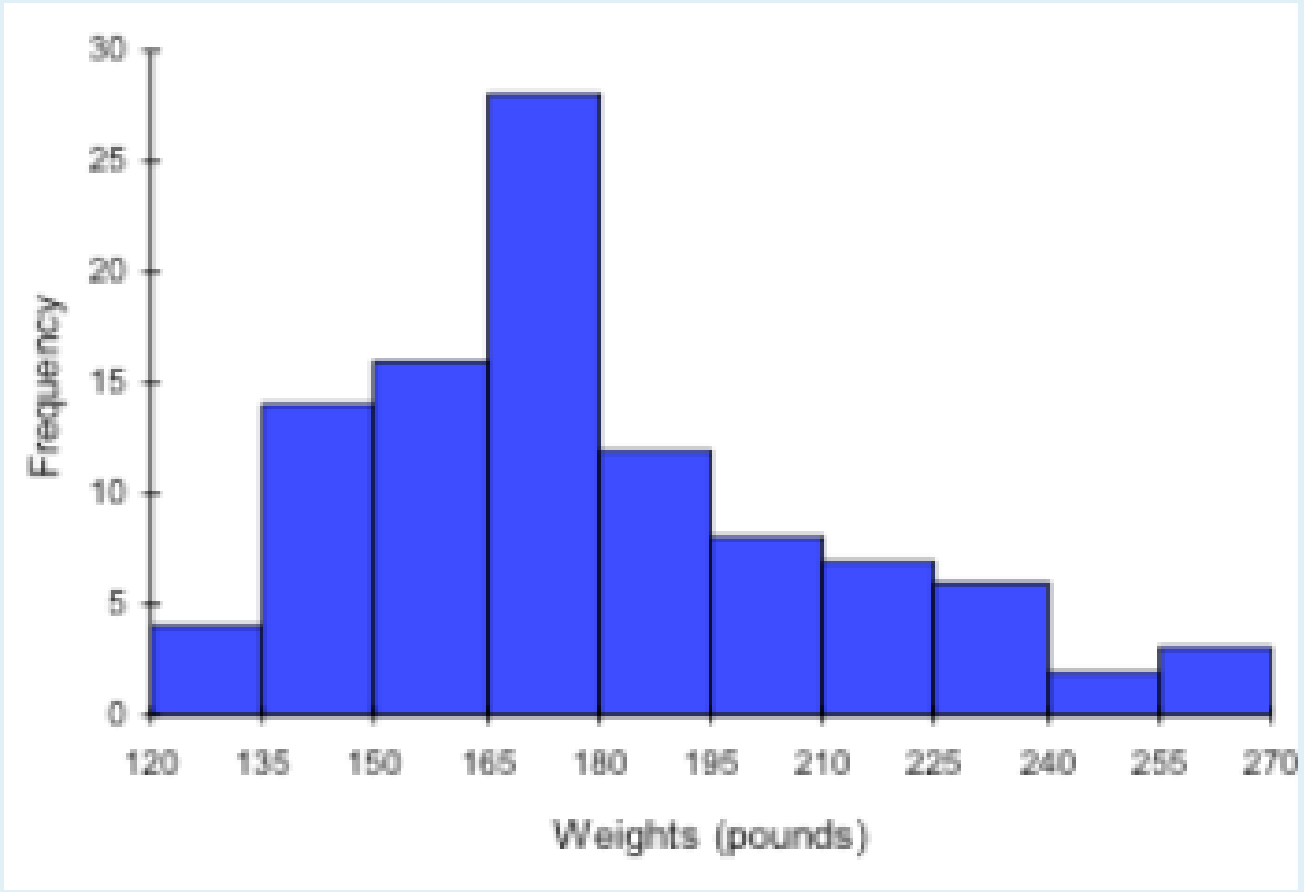
C

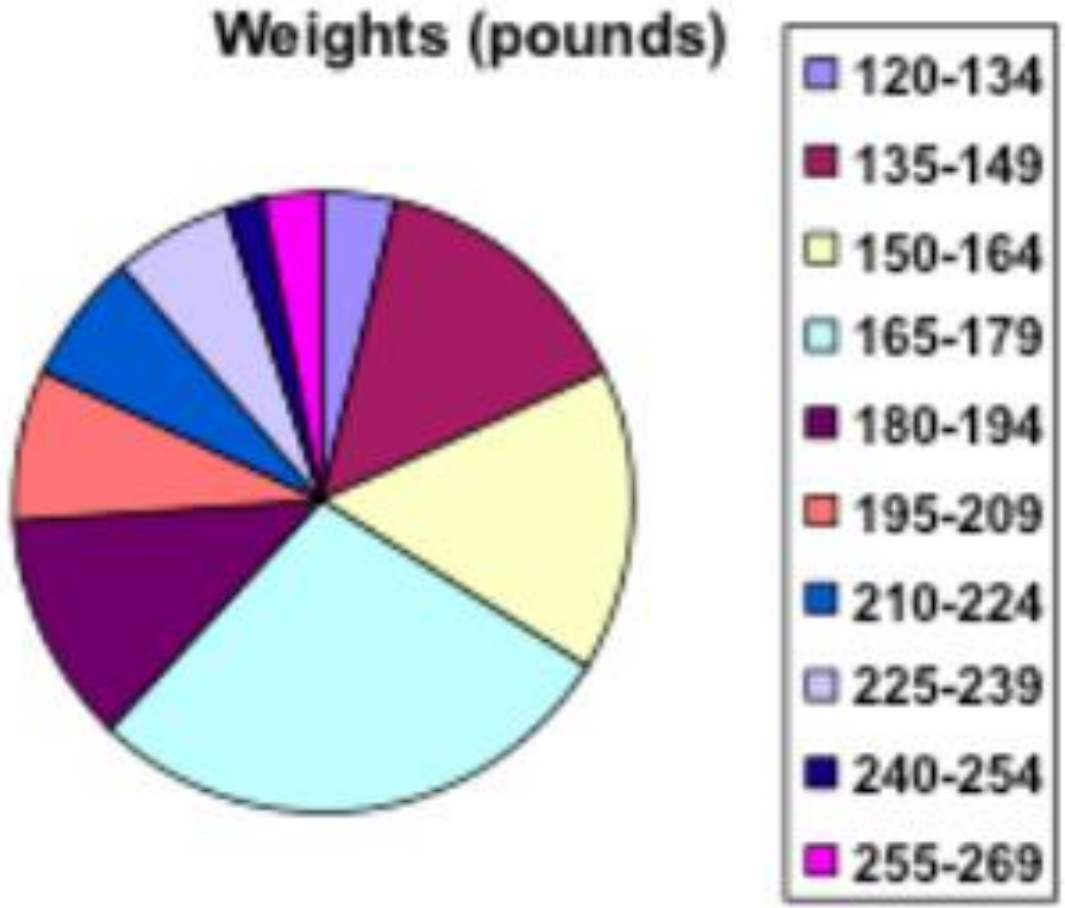
- Quantitative data are displayed using **dotplots, stemplots, histograms, or boxplots**.
    - The choice of graph depends on the **data size and purpose**.
- Graphs help reveal the **shape, center, spread, and outliers** of a distribution.
    - Histograms show the overall pattern using intervals (bins).
- All graphs should be clearly labeled and interpreted **in context**.

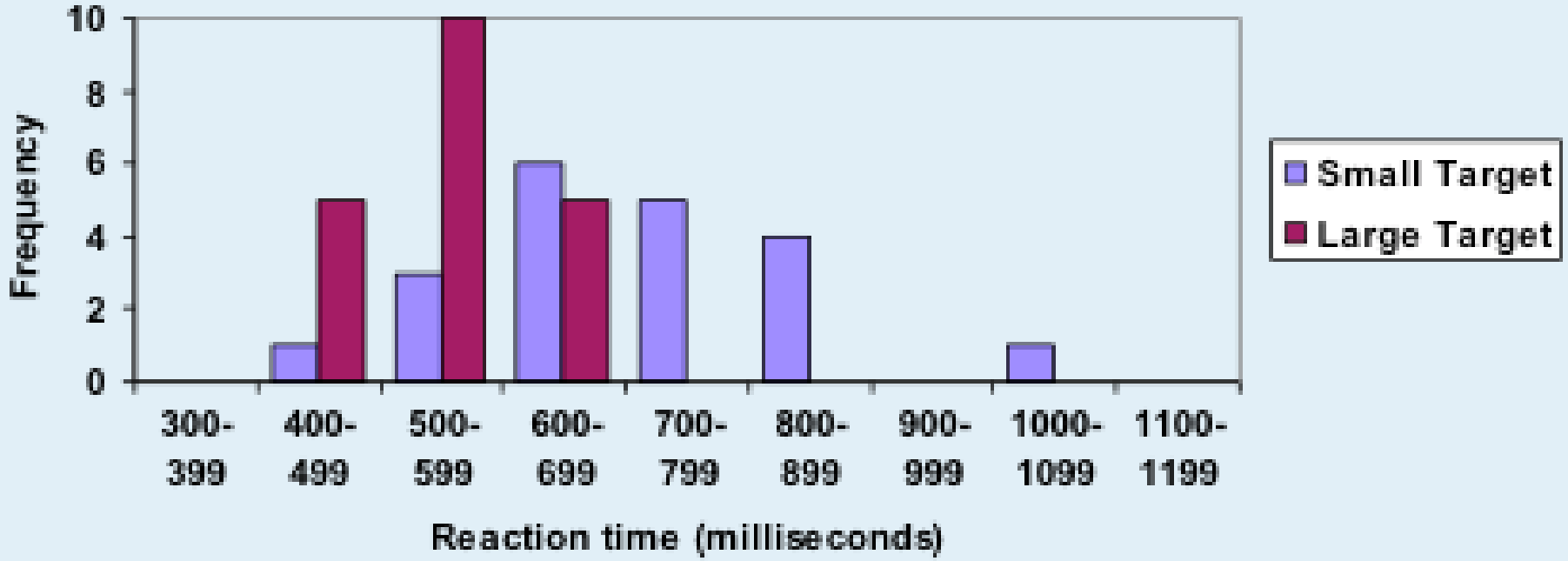
Quantitative variable	定量变量
Histogram	直方图
Distribution	分布
Center	中心
Spread	离散程度


Frequency
8
7
6
5
4
3
2
1
0
0
5
12
15
16
17
18
19
20
Score

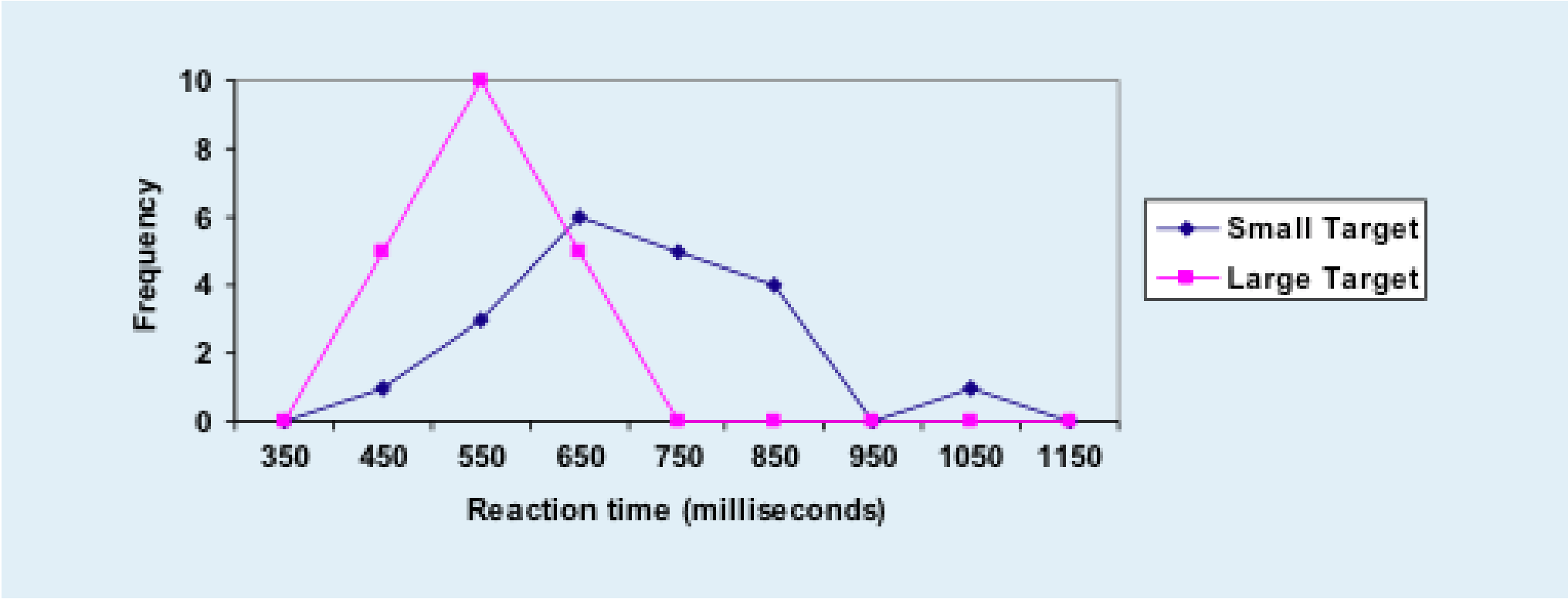

Frequency
0
1
2
3
4
5
6
7
8
9
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Score


30
25
20
15
10
5
0
Frequency
120
135
150
165
180
195
210
225
240
255
270
Weights (pounds)


Weights (pounds)
120-134
135-149
150-164
165-179
180-194
195-209
210-224
225-239
240-254
255-269


10
8
6
4
2
0
Frequency
300-399
400-499
500-599
600-699
700-799
800-899
900-999
1000-1099
1100-1199
Reaction time (milliseconds)
Small Target
Large Target

Frequency Polygon



T/F. Frequency polygons are commonly used with grouped data.

A teacher records the scores of 25 students on a math test. The scores range from 55 to 100.

(a) What type of variable is the test score?

(b) Suggest an appropriate graph to display the distribution of scores and explain why.

(c) Name two features of the distribution that can be described from the graph.

(a) The test score is a **quantitative variable** because it represents numerical measurements.

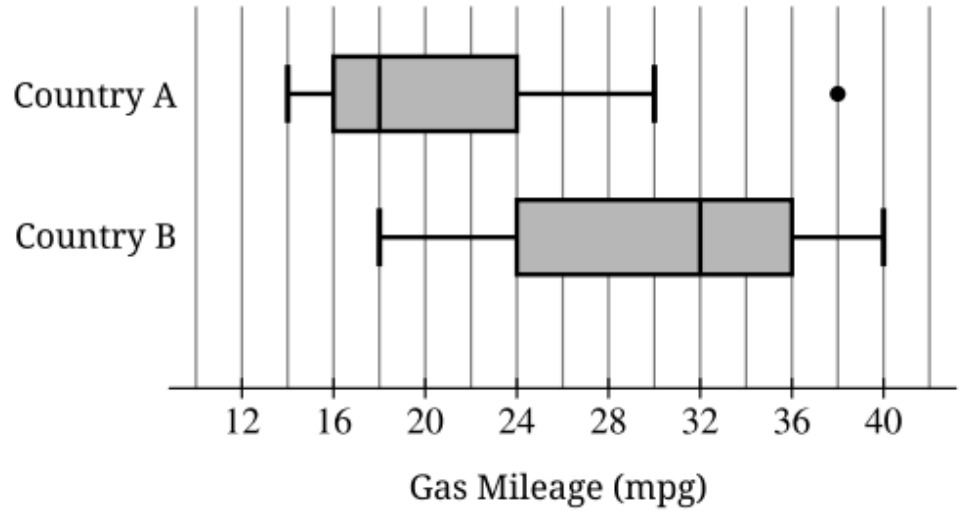
(b) A **histogram** is appropriate because it shows the **distribution of numerical data** and how frequently values occur in different intervals.

(c) Two features that can be described are:

- **Center** (e.g., mean or median score)
- **Spread** (e.g., range, variability, or standard deviation)
- Optionally, **shape** (symmetric, skewed) and **outliers** can also be identified.

1. The manager of an automotive company is interested in comparing the gas mileages for cars manufactured in Country A and cars manufactured in Country B. The manager selected a random sample of 100 cars manufactured in Country A and a random sample of 100 cars manufactured in Country B. The gas mileages for each sample, in miles per gallon (mpg), are summarized in the boxplots.

Boxplots of Gas Mileage for Each Country



A. Compare the distributions of gas mileage for the sample of cars manufactured in Country A and the sample of cars manufactured in Country B.

B. For the distribution of gas mileage for the sample of cars manufactured in Country A, would you expect the mean to be greater than 18 mpg, less than 18 mpg, or equal to 18 mpg? Justify your answer.

C. The manager will create a new boxplot with the combined data from the sample of cars manufactured in Country A and the sample of cars manufactured in Country B.

i. What is the range of the combined data set? Justify your answer.

ii. What is a possible value of the median of the combined data set? Justify your answer by referencing the boxplots shown.

The distribution of gas mileage for the sample of cars manufactured in Country A has a lower center than the distribution of gas mileage for the sample of cars manufactured in Country B. The median gas mileage for the sample of cars manufactured in Country A (18 mpg) is less than the median gas mileage for the sample of cars manufactured in Country B (32 mpg).

The range of the gas mileages for the sample of cars manufactured in Country A (24 mpg) is slightly greater than the range of the gas mileages for the sample of cars manufactured in Country B (22 mpg). However, the IQR of the gas mileages for the sample of cars manufactured in Country A (8 mpg) is less than the IQR of the gas mileages for the sample of the cars manufactured in Country B (12 mpg).

The car manufactured in Country A with 38 mpg (the maximum of the sample of cars manufactured in Country A) is an outlier, while the distribution of gas mileage for the sample of cars manufactured in Country B has no outliers.

The mean of the distribution of gas mileage for the sample of cars manufactured in Country A is expected to be greater than 18 mpg, the median of the distribution. Because the distribution of gas mileage for the sample of cars manufactured in Country A has an outlier to the right (or is skewed to the right), the mean of the distribution (which is not resistant) is expected to be pulled above the median (which is resistant) toward the higher values of gas mileage.

i. The maximum value in the combined data is 40 mpg because 40 mpg is the maximum gas mileage for the sample of cars manufactured in Country B, and as shown in the boxplot, all the gas mileages for the sample of cars manufactured in Country A are less than 40 mpg. The minimum value in the combined data is 14 mpg, because 14 mpg is the minimum mpg for the sample of cars manufactured in Country A, and as shown in the boxplot, all the gas mileages for the sample of cars manufactured in Country B are greater than 14. Thus, the range of the combined data set is $40 - 14 = 26$ mpg.

ii. In the combined data, there are 200 gas mileages. The median is a value where at least half, or 100, of the gas mileages in the combined data are less than or equal to the median value and at least half, or 100, of the gas mileages in the combined data are greater than or equal to the median value. From the boxplot for the sample of cars manufactured in Country A, the third quartile, Q3, is 24 mpg indicating there are at least 75 gas mileages less than or equal to 24 mpg and at least 25 gas mileages greater than or equal to 24 mpg. From the boxplot for the sample of cars manufactured in Country B, the first quartile, Q1, is 24 mpg indicating there are at least 25 gas mileages less than or equal to 24 mpg and at least 75 gas mileages greater than or equal to 24 mpg. Thus, in the combined data set, there are at least 100 gas mileages less than or equal to 24 mpg and at least 100 gas mileages greater than or equal to 24 mpg, which implies 24 is the value of the median of the combined data set.

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

1. Graph Selection Justification

A researcher collects data on students' exam scores and study hours.

- Which graphs would you use to display each variable and why?
- Explain how your choice helps in analysing the data effectively.

---

2. Comparing Distributions

Two classes have the following exam score summaries:

- Class A: symmetric distribution, mean = 75, SD = 8
- Class B: right-skewed distribution, median = 78, IQR = 10
- Which class likely performed better overall?
- Which graph (histogram or boxplot) would better support your comparison? Explain.

---

3. Interpreting Shape from Graphs

A histogram shows most data clustered on the left with a long tail extending to the right.

- Describe the shape of the distribution.
- What would you expect about the relationship between mean and median?
- Which graph feature supports your conclusion?

---

4. Real-World Data Choice

A hospital tracks patient waiting times in minutes.

- Compare the usefulness of a histogram vs a boxplot for presenting this data.
- Explain which graph would be better for identifying outliers and why.

Condensed Answers

1. Graph Selection Justification

- Exam scores → histogram or boxplot (distribution of one quantitative variable).
- Study hours → histogram or dotplot (distribution of one variable).
- Helps show shape, center, spread, and variability clearly.

---

2. Comparing Distributions

- Likely better performance: Class B (higher median = 78 vs mean 75 in A).
- Better graph: boxplot (good for comparing median, IQR, and skewness).
- Histogram also useful but harder for direct comparison.

---

3. Interpreting Shape from Graphs

- Shape: right-skewed (positive skew).
- Mean > median due to long right tail.
- Long right tail in histogram shows extreme high values.

---

4. Real-World Data Choice

- Histogram: best for showing overall distribution shape.
- Boxplot: best for identifying outliers and comparing spread.
- Boxplot preferred for outliers due to IQR method and clear display of extreme values.

Exploring One-Variable Data (Descriptive Statistics)

1. Distribution（分布）
2. Variable（变量）
3. Quantitative Variable（定量变量）
4. Center（中心）
5. Spread（离散程度）
6. Mean（平均数）
7. Median（中位数）
8. Outlier（离群值）
9. Standard Deviation（标准差）
10. Interquartile Range, IQR（四分位距）

1. One-variable data describes a single quantitative or categorical variable.
2. Graphs such as histograms and boxplots display quantitative distributions.
3. Bar charts are used for categorical data, not quantitative data.
4. Center can be described using mean or median.
5. Spread describes how variable or dispersed the data are.
6. Standard deviation measures average distance from the mean.
7. IQR measures the spread of the middle 50% of data.
8. Outliers are extreme values that differ greatly from the rest of the data.
9. Shape describes whether a distribution is symmetric, skewed, or uniform.
10. Choosing summary statistics depends on the shape of the distribution.

AP Statistics FRQ: One-Variable Data (Descriptive Statistics)

A teacher records the number of hours students studied in a week:

2, 3, 3, 4, 5, 6, 6, 7, 8, 10, 12

(a) Construct a suitable graphical display for the data.

(b) Describe the shape, center, and spread of the distribution.

(c) Calculate the mean and median.

(d) Identify any outliers using the IQR method.

(e) Decide whether the mean or median is a better measure of center and justify your answer.

(a)

Histogram or boxplot (quantitative one-variable data).

(b)

Shape: right-skewed.

Center: around 6–7 hours.

Spread: range = 12 − 2 = 10.

(c)

Mean = 6.18 (approx.)

Median = 6

(d)

Q1 = 3.5, Q3 = 8

IQR = 4.5

Lower fence = −3.25, Upper fence = 14.75

→ No outliers.

(e)

Median is better because the distribution is slightly right-skewed and median is resistant to extreme values.

TRUE or FALSE

- A histogram is used to display categorical data.

TRUE or FALSE

- The median is always affected by extreme values (outliers).

TRUE or FALSE

- Standard deviation measures the spread of data around the mean.