



## Data Analysis

1 Find the mean of these numbers

**13, 2, 10, 3**

|         |        |         |
|---------|--------|---------|
| A<br>10 | B<br>7 | C<br>12 |
| D<br>11 | E<br>5 | F<br>3  |

2 What is the median of this data set?

36, 44, 44, 57, 90, 39, 37, 64, 30, 51

|         |         |
|---------|---------|
| A<br>57 | B<br>37 |
| C<br>47 | D<br>44 |

3 Find the median of these numbers

**8, 14, 18, 9, 8, 4, 15**

|         |         |        |
|---------|---------|--------|
| A<br>11 | B<br>15 | C<br>9 |
| D<br>14 | E<br>18 | F<br>4 |

4 Find the range of these numbers

**7, 7, 5**

|        |        |        |
|--------|--------|--------|
| A<br>2 | B<br>4 | C<br>1 |
| D<br>6 | E<br>7 | F<br>3 |

5 Would the mean or the median be a better measure of center for this data set?

**Data set: 26, 29, 23, 26, 3, 26, 28**

|             |           |
|-------------|-----------|
| A<br>Median | B<br>Mean |
|-------------|-----------|

6 What is the interquartile range (IQR) of this data set?

34, 54, 70, 45, 80, 70, 73, 76, 27, 14

|         |          |
|---------|----------|
| A<br>66 | B<br>107 |
| C<br>39 | D<br>73  |

7 What is the interquartile range (IQR) of this data set?

46, 46, 66, 24, 55, 57, 58, 59

|            |           |
|------------|-----------|
| A<br>12.5  | B<br>42   |
| C<br>104.5 | D<br>58.5 |

8 How is this variable best classified?

**Distance travelled during a road trip**

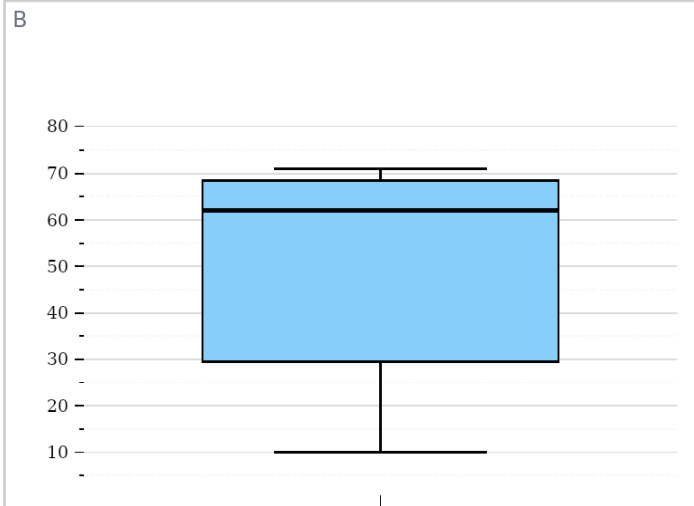
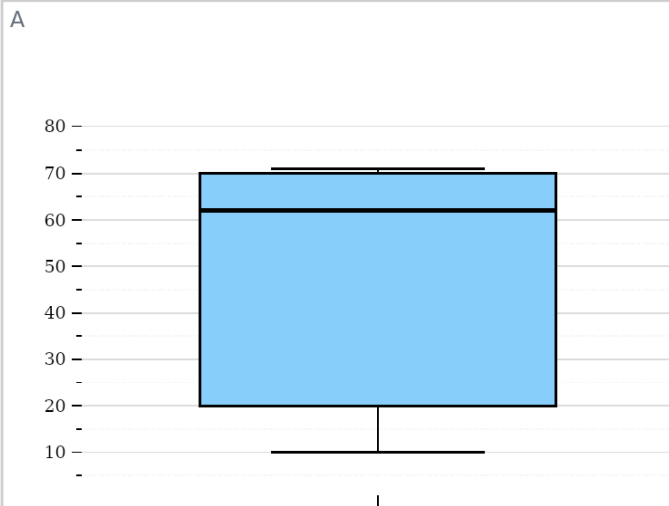
|                                |
|--------------------------------|
| A<br>Quantitative (continuous) |
| B<br>Quantitative (discrete)   |
| C<br>Qualitative               |



## Data Analysis

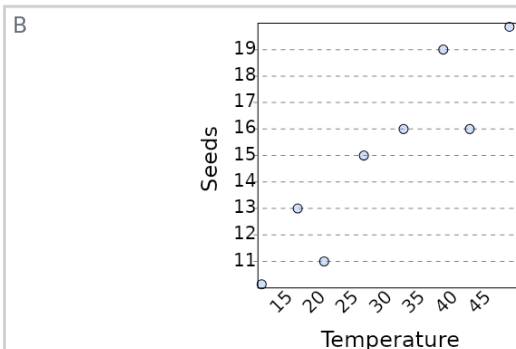
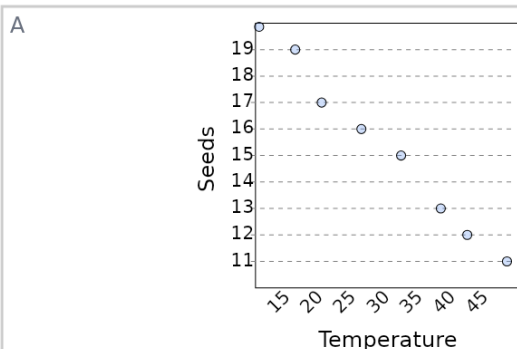
9 Which box plot represents this data set?

10, 16, 58, 43, 62, 66, 70, 71, 67



10 Which scatter plot matches this data set?

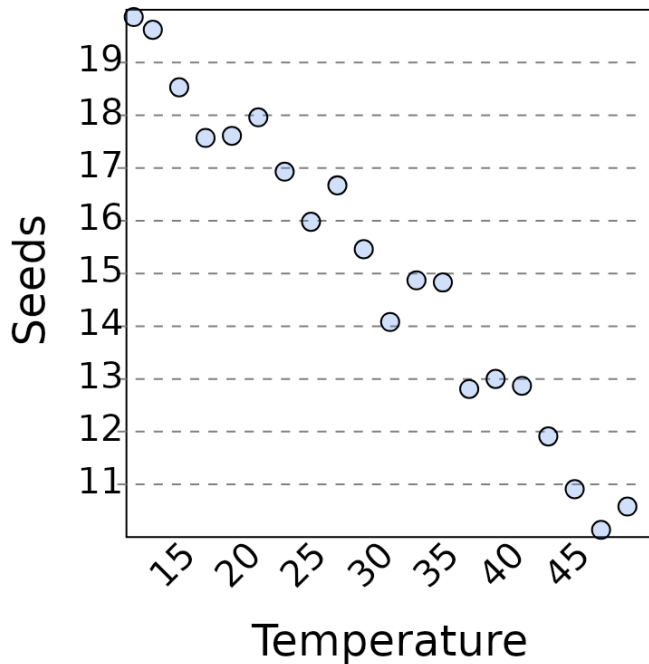
| Temperature | Seeds |
|-------------|-------|
| 10          | 10    |
| 16          | 13    |
| 20          | 11    |
| 26          | 15    |
| 32          | 16    |
| 38          | 19    |
| 42          | 16    |
| 48          | 20    |





## Data Analysis

**11** What is the correlation direction and strength on this scatter plot?



A

strong positive correlation

B

strong negative correlation

C

no correlation

D

weak negative correlation

**12** Which of these variables is best classified as quantitative (discrete)?

**Quantitative (discrete)**

A Mode of transportation used to travel to work

B Distance travelled during a road trip

C Number of correct answers on a quiz

**13** Is this random variable discrete or continuous?

**Air pressure inside a bicycle tire**

A

Discrete

B

Continuous

**14** Which random variable has this type of probability distribution?

**Discrete**

A Blood pressure of a patient

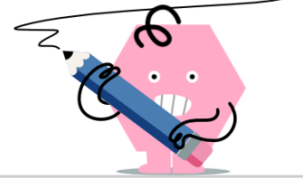
B Total goals scored during a hockey game

**15** Why does this sampling approach introduce bias?

**Goal: Average homework time for students at a high school**  
**Sampling approach: Survey students in the library after school**

A Students in the library after school may study more than the typical student.

B Students in the library after school may complete homework more accurately than other students.

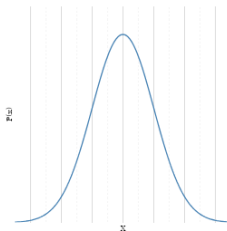


## Data Analysis

**16** Which model could represent this random variable's probability distribution?

### Distance travelled during a bicycle ride

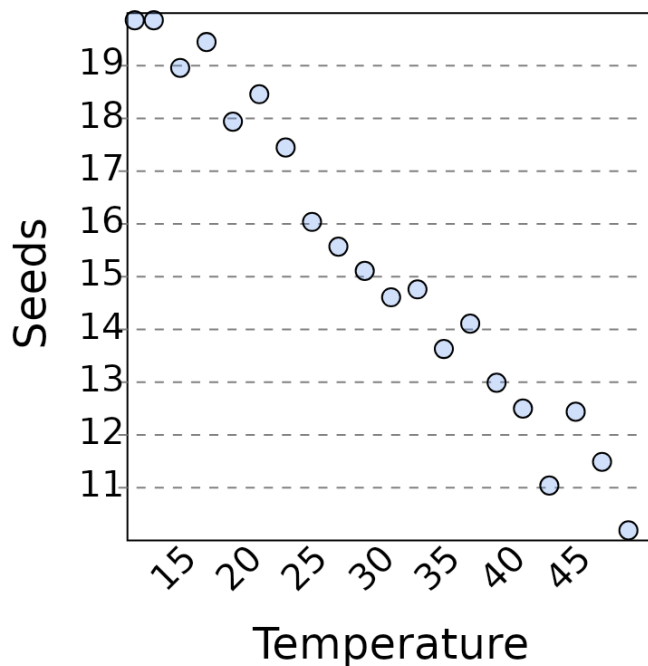
A



B

| $X$   | $P(X)$   |
|-------|----------|
| $X_1$ | $P(X_1)$ |
| $X_2$ | $P(X_2)$ |
| $X_3$ | $P(X_3)$ |

**17** Using this data, would predicting seeds when temperature is 0 be interpolation or extrapolation?

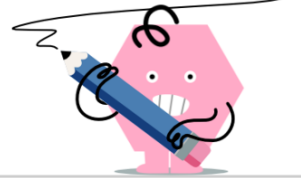


A

Interpolation

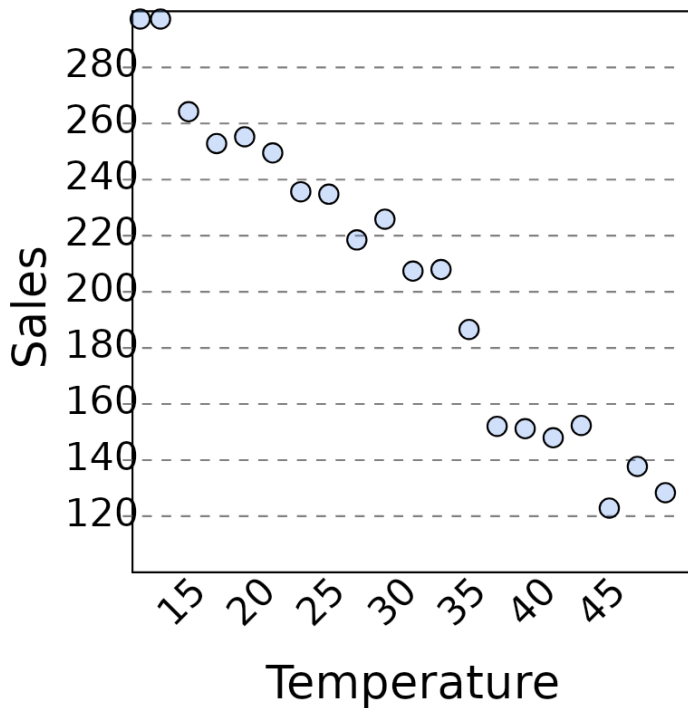
B

Extrapolation



## Data Analysis

**18** Using this data, would predicting sales when temperature is 57 be interpolation?



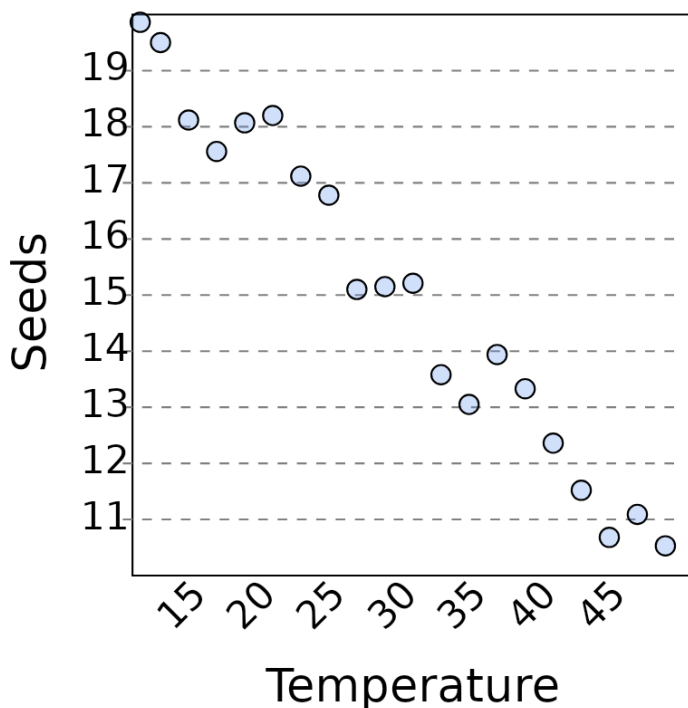
A

No

B

Yes

**19** If temperature is 24, would 13 be a reasonable value of seeds?

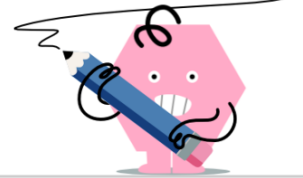


A

No

B

Yes



## Data Analysis

**20** What type of bias does this sampling approach introduce?

**Goal: Percentage of people who exercise weekly**  
**Sampling approach: Survey people leaving a gym**

- |   |                         |
|---|-------------------------|
| A | Selection bias          |
| B | Measurement bias        |
| C | Voluntary response bias |
| D | Response bias           |

**21** Is this sample representative of the population?

**Population: All restaurants in a city**  
**Sample: 250 restaurants that offer food delivery**

- |   |                                                |
|---|------------------------------------------------|
| A | No, it is not representative of the population |
| B | Yes, it is representative of the population    |

**22** Which sample is most representative of this population?

**Population: All trees in a provincial park**

- |   |                                                     |
|---|-----------------------------------------------------|
| A | 300 trees located within 50 metres of hiking trails |
| B | 300 trees randomly selected throughout the park     |

**23** Would a value of 8 be an outlier for this data set?

Median = 40,  $Q_1 = 27$ ,  $Q_3 = 41$

- |   |     |
|---|-----|
| A | Yes |
| B | No  |

**24** Is the value 97 an outlier in this data set?  
54, 29, 32, 30, 58, 45, 39, 97

- |   |     |
|---|-----|
| A | No  |
| B | Yes |

**25** Which population does this sample best represent?

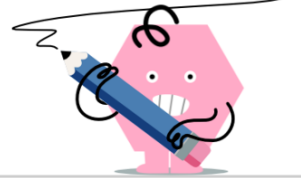
**Sample: 2,500 voters randomly selected from the provincial voter registry**

- |   |                                                           |
|---|-----------------------------------------------------------|
| A | People motivated to participate in online political polls |
| B | All eligible voters in a province                         |

**26** Which of these variables is qualitative?

**Qualitative**

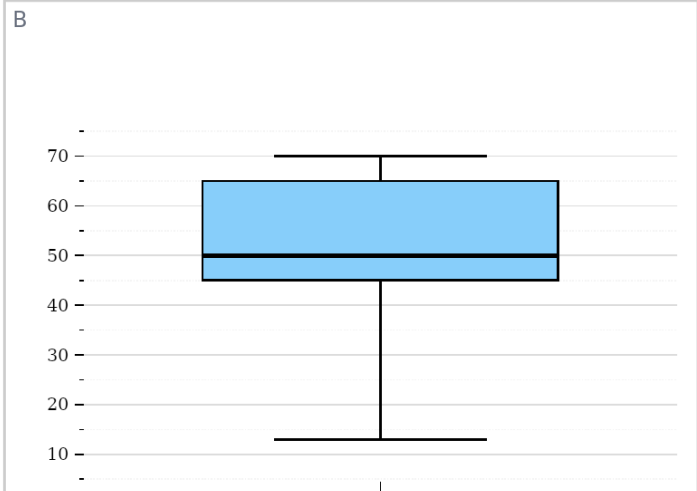
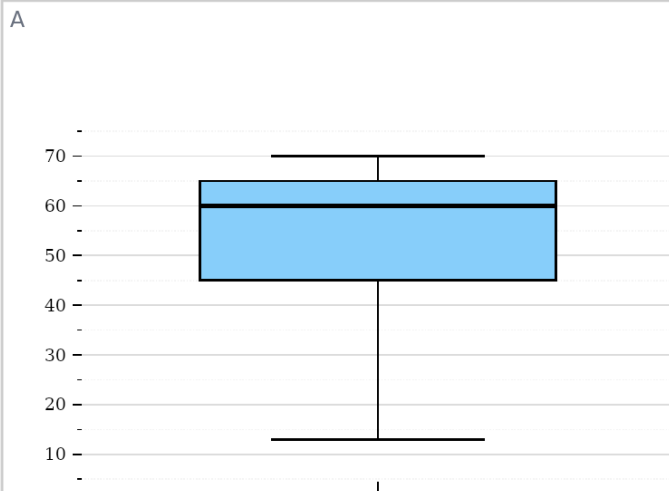
- |   |                                          |
|---|------------------------------------------|
| A | Number of pets owned by a household      |
| B | Number of goals scored in a hockey game  |
| C | Eye colour of a randomly selected person |



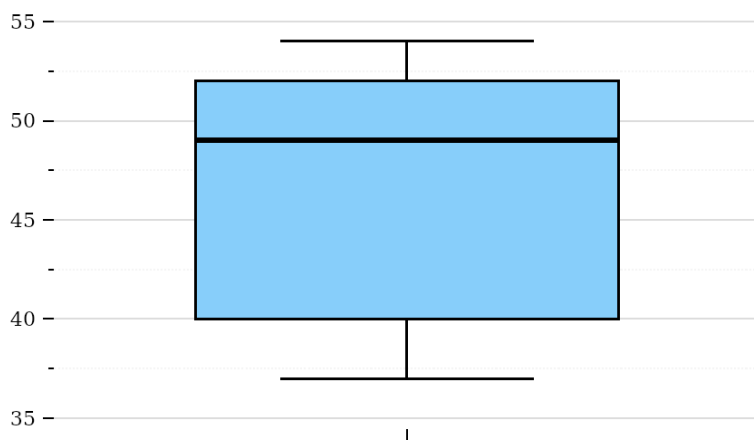
## Data Analysis

**27** Which box plot represents this data set?

50, 65, 13, 50, 44, 56, 50, 70, 68, 45



**28** Would a value of 72 be an outlier for this box plot?

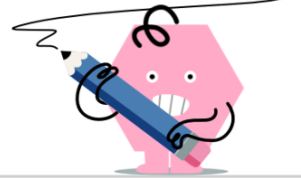


A

No

B

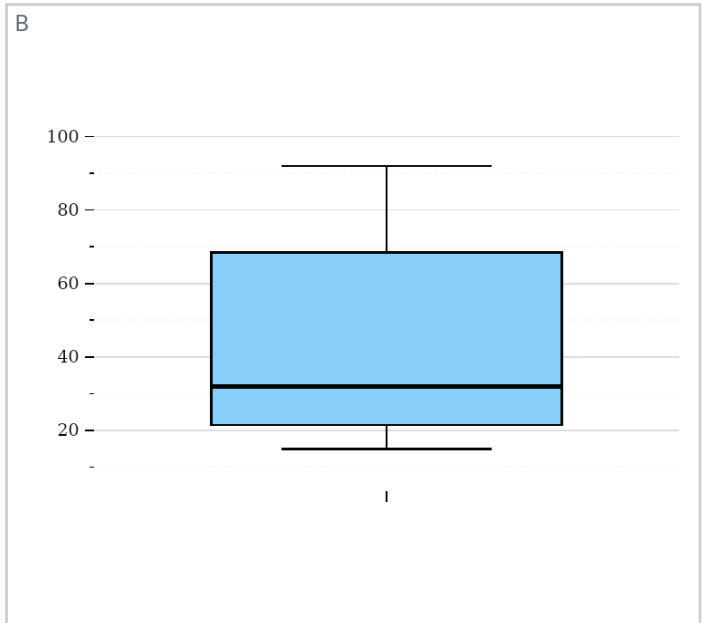
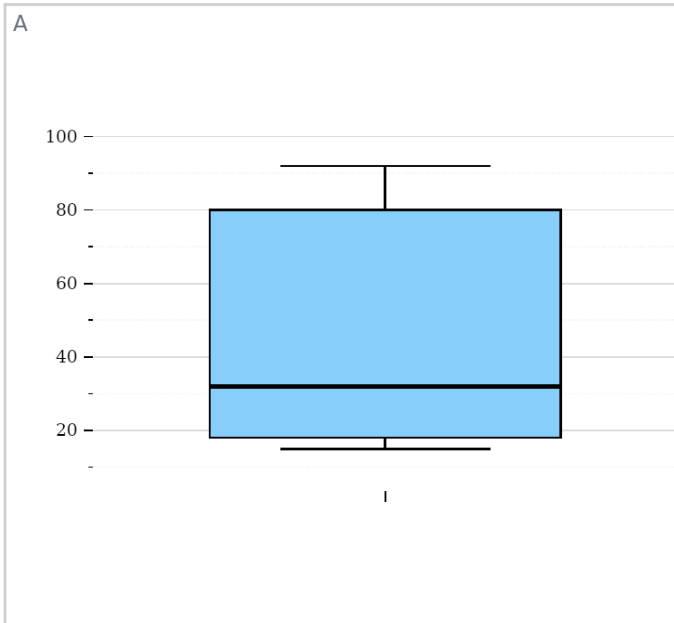
Yes



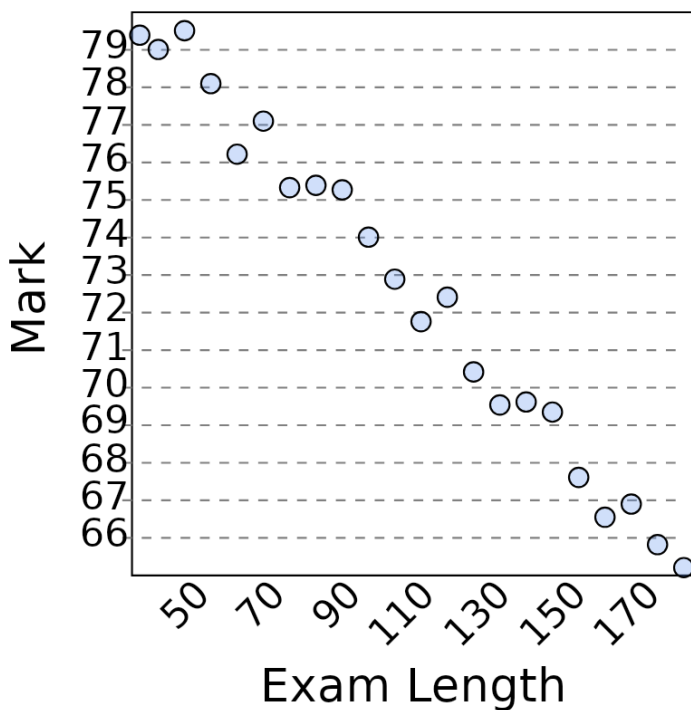
## Data Analysis

**29** Which box plot represents this data set?

15, 19, 24, 25, 32, 52, 63, 74, 92



**30** If exam length is 82.5, what would a reasonable value of mark be?



A

69

B

75