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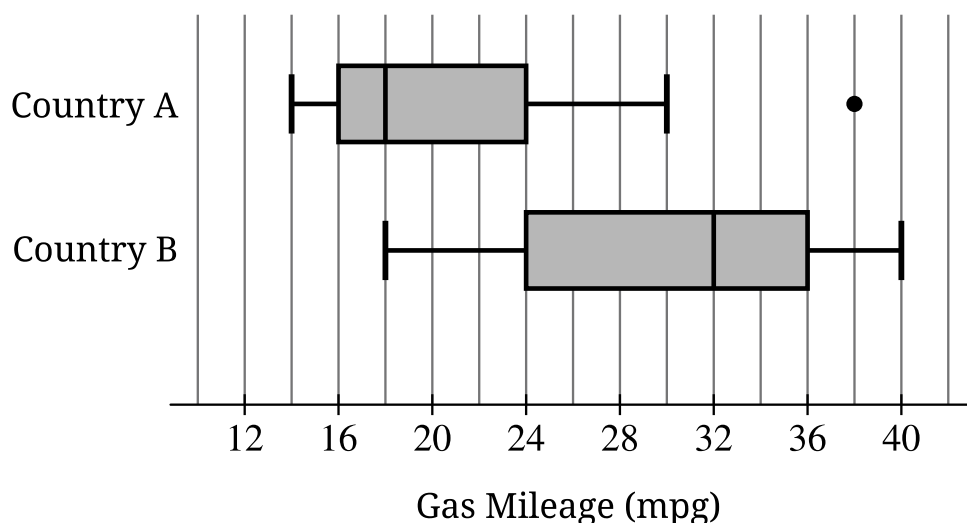* The manager of an automotive company is interested in comparing the gas mileages for cars manufactured in Country A and cars manufactured in Country B. The manager selected a random sample of 100 cars manufactured in Country A and a random sample of 100 cars manufactured in Country B. The gas mileages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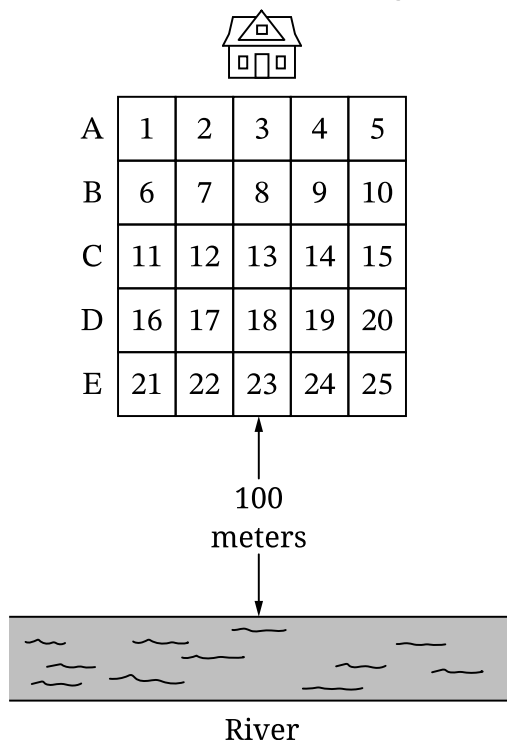
Boxplots of Gas Mileage for Each Country

Gas Mileage (mpg)

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.

2. Aphids are tiny insects that feed on plants such as cabbage plants. A farmer wants to reduce the number of aphids in a cabbage field. A river is located 100 meters south of the cabbage field. The farmer divides the field into 25 regions of equal size, as shown in the diagram. Each region has approximately the same number of cabbage plants.

Farmer's House and Cabbage Field

A	1	2	3	4	5
B	6	7	8	9	10
C	11	12	13	14	15
D	16	17	18	19	20
E	21	22	23	24	25

100 meters

River

The farmer would like to estimate the proportion of cabbage plants in the field that are affected by aphids and believes that the extent of aphid damage is greater for the regions in the cabbage field closer to the river. To obtain the estimate, the farmer is considering three sampling methods.

- Sampling method I: Select region 3, which is closest to the farmer's house and farthest from the river. Examine every cabbage plant in the region for aphid damage.

- Sampling method II: Randomly select one row (A, B, C, D, or E). For every region in the selected row, examine every cabbage plant for aphid damage.

- Sampling method III: Randomly select one region from each of rows A, B, C, D, and E. For each selected region, examine every cabbage plant for aphid damage.

Question 1: Focus on Exploring Data   4 points

General Scoring Notes

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each part of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

Model Solution	Scoring
A 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.	**Essentially correct (E)** if the response satisfies at least three of the following four components: 1. Directly compares the center for the two distributions 2. Directly compares the spread (either IQR or range) for the two distributions 3. Indicates that the gas mileage of one of the cars manufactured in Country A is an outlier 4. Provides sufficient context, which includes the manufacturing countries ("Country A" and "Country B") AND the dependent variable ("gas mileage" or "mpg") **Partially correct (P)** if the response satisfies only two of the four components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Scoring Notes:

- The response need not include specific numerical values to satisfy any given component or to score E on part A.
- A response that only uses the "means" in the comparison of center would not satisfy component 1.
- Any acceptable mention of shape in the response should be ignored because complete shape information cannot be determined from a boxplot. Acceptable mentions of shape include:
  - The shape of the distribution of the gas mileage for cars manufactured in Country A can be described as skewed, positively skewed, or right skewed.
  - The shape of the distribution of the gas mileage for cars manufactured in Country B can be described as skewed, negatively skewed, left skewed, or **approximately** symmetric.

- If the response describes the shape of either distribution as just "symmetric," "normal," "unimodal," or an incorrect shape (e.g., "the distribution of gas mileages for Country A is left skewed" or "the distribution of gas mileages for Country B is right skewed"), then part A cannot be scored E.
- A response that only lists values for center and/or spread and does not directly compare them does not satisfy components 1 and/or 2.
- A response that just refers to "A" or "B" AND the dependent variable ("gas mileage" or "mpg") may satisfy component 4.

Model Solution	Scoring
B 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.	**Essentially correct (E)** if the response satisfies the following three components: 1. Indicates that 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 2. Indicates 18 as the value of the median 3. Provides a reasonable justification for the relationship between the mean and the median based on a right-skewed distribution or the expectation that an outlier in the right tail will pull the mean above the median **Partially correct (P)** if the response satisfies component 1 AND either component 2 or component 3. **Incorrect (I)** if the response does not meet the criteria for E or P.

Scoring Notes:

- A response that states the median is 18 mpg in either part A or part B satisfies component 2.
- A response that provides a weak justification (e.g., "because the distribution is skewed") does not satisfy component 3. However, a response that provides a justification based on the distribution being right skewed may satisfy component 3.
- Component 3 may be satisfied with an appropriate numerical argument that shows that the mean must be greater than 18 mpg.
    - An example of an inappropriate numerical argument would be averaging portions of the five-number summary.

Model Solution	Scoring
C 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.	**Essentially correct (E)** if the response satisfies the following four components: 1. Correctly calculates 26 mpg as the range of the combined data set 2. Provides a justification for the range with an argument based on identifying 40 as the maximum of the combined data set and identifying 14 as the minimum of the combined data set 3. Calculates 24 mpg as a possible value of the median of the combined data set 4. Provides a justification for the median by indicating that at least half of the combined data values are at most 24 and at least half of the combined data values are at least 24 **Partially correct (P)** if the response satisfies only two or three of the four components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Scoring Notes:

- A response that solely indicates a range of values (e.g., "the gas mileages range from 14 mpg to 40 mpg") does not satisfy component 1 but may satisfy component 2.
- A response that states the range as a value of 25 or 27 from misinterpreting the minimum value in the distribution of gas mileages for Country A as 13 or 15 may be considered a minor error and satisfies components 1 and 2.

- A response that gives the correct value for the median (24) and justifies the value by looking at the quartiles on one side of 24 (e.g., identifying that one-fourth of Country B cars and three-fourths of Country A cars have mpg less than or equal to 24) satisfies component 4.
- A response that provides justification of the median based on a value near the 100th data value satisfies component 4.
- A response that provides justification of the median based on counting the segments in the boxplots may satisfy component 4. The response is not required to specify equal sample sizes.
  - For example, one-half of the boxplot segments are no greater than 24 mpg (four of eight segments— three segments from Country A and one segment from Country B) and one-half of the boxplot segments are at least 24 mpg (four of eight segments—one segment from Country A and three segments from Country B), which indicates that 24 mpg is the median of the combined data set.

Scoring for Question 1	Score
Complete Response Three parts essentially correct	4
Substantial Response Two parts essentially correct and one part partially correct	3

Developing Response
Two parts essentially correct and no part partially correct
  OR
One part essentially correct and one or two parts partially correct
  OR
Three parts partially correct
	2

Minimal Response One part essentially correct and no parts partially correct *OR* No part essentially correct and two parts partially correct	1