Algebra 1

ELG HS.S.1: Summarize, represent, and interpret data on a single count or measurement variable

Vertical Progression:

6th Grade	**6.SP.B Summarize and describe distributions.** ○ **6.SP.B.4** Display numerical data in plots on a number line, including dot plots, histograms, and box plots. ○ **6.SP.B.5** Summarize numerical data sets in relation to their context, such as by: ○ **6.SP.B.5.b** Describing the nature of the attribute under investigation, including how it was measured and its units of measurement. ○ **6.SP.B.5.c** Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered. ○ **6.SP.B.5.d** Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.
7th Grade	**7.SP.B Draw informal comparative inferences about two populations** ○ **7.SP.B.3** Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. ○ **7.SP.B.4** Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.
Algebra 1	**ELG.MA.HS.S.1: Summarize, represent, and interpret data on a single count or measurement variable.** ○ **S-ID.1** Represent data with plots on the real number line (dot plots, histograms, and box plots). ○ **S-ID.2** Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets. ○ S-ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).
Algebra 2	**ELG.MA.HS.S.1: Summarize, represent, and interpret data on a single count or measurement variable.** ○ **S-ID.4** Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.

Students will demonstrate command of the ELG by:

- Accurately representing data in a number of different ways (dot plots, histograms, and box plots)
- Comparing the center and spread of two or more sets of data.
- Interpreting differences in the center and spread of data sets in terms of their context.

Vocabulary:

- Interquartile range
- Mean
- Median
- Mode
- Quartile
- Range
- Standard deviation
- Variance

Sample Instructional/Assessment Tasks:

1) **Standard(s): S-ID.1 -3**

Source: Illustrated Mathematics
https://www.illustrativemathematics.org/content-standards/HSS/ID/A/3/tasks/1027)

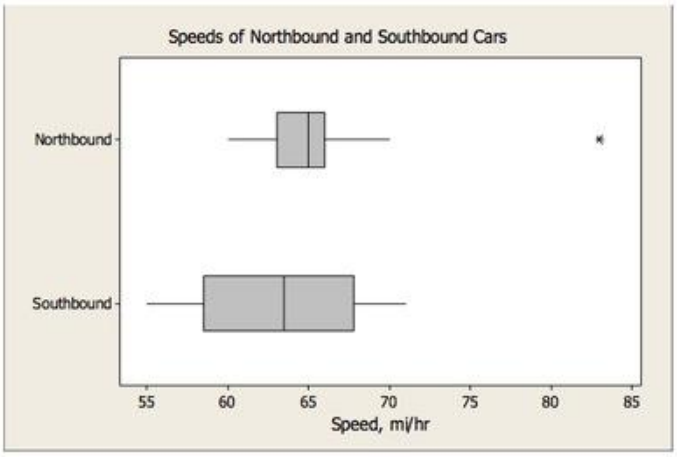
Item Prompt: Speed Trap

A statistically-minded state trooper wondered if the speed distributions are similar for cars traveling northbound and for cars traveling southbound on an isolated stretch of interstate highway. He uses a radar gun to measure the speed of all northbound cars and all southbound cars passing a particular location during a fifteen-minute period. Here are his results:

Northbound Cars				
60	62	62	63	63
63	64	64	64	65
65	65	65	66	66
67	68	70	83	
Southbound Cars				
55	56	57	57	58
60	61	61	62	63
64	65	65	67	67
68	68	68	68	71

Draw box plots of these two data sets, and then use the plots and appropriate numerical summaries of the data to write a few sentences comparing the speeds of northbound cars and southbound cars at this location during the fifteen minute time period.

Correct Answer:



Using the median to describe typical speed, we would say that typical speed is about the same (median of 65 mph for northbound and 63.5 mph for southbound) for northbound cars and southbound cars. One noticeable difference between the two speed distributions is that the southbound speeds are more variable than the northbound speeds. This means that the northbound speeds tended to be more consistent than the southbound speeds, which tended to differ more from one car to another. Other than the outlier in the northbound speeds, both speed distributions appear to be approximately symmetric.

2) **Standard(s): S-ID.1 -3**

Source: Illustrative Mathematics
https://www.illustrativemathematics.org/content-standards/HSS/ID/A/2/tasks/942

Item Prompt: Haircut Costs

Seventy-five female college students and 24 male college students reported the cost (in dollars) of his or her most recent haircut. The resulting data are summarized in the following table.

	Females	Males
No. of observations	75	24
Minimum	0	0
Maximum	150	35
1st Quartile	20	9.25
Median	31	17
3rd Quartile	75	20
Mean	52.53	20.13

a. Using the minimum, maximum, quartiles and median, sketch two side-by-side box plots to compare the hair cut costs between males and females in this student's school.

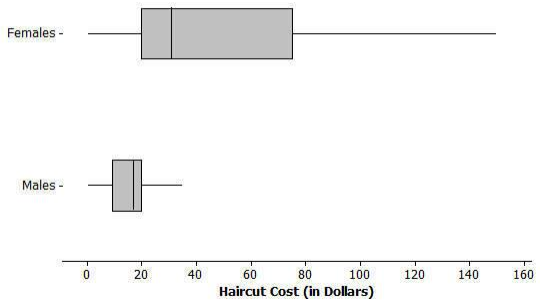
b. How would you describe the difference in haircut costs between males and females? Be sure you discuss differences/similarities in shape, center and spread.

c. Why is the mean greater than the median both for males and for females? Explain your reasoning.

d. Is the median or mean a more appropriate choice for describing the "centers" of these two distributions?

Correct Answer(s):

a. Students can sketch out a basic box plot with whiskers extending to the min and max, a box extending from the first quartile to the third quartile, and a line at the median, as shown below. In order to compare haircut costs of males and females, the two boxplots should be plotted side by side on the same scale.



b. Both boxplots show distributions that are skewed to the right. It makes sense that most haircuts will not cost too much, but a few students will spend a large amount. Since the cost will always be a positive number, the minimum cannot be less than 0 and there is a long right tail. The centers and spreads are quite different. The median cost for females is about twice that of males, and there is much more variability in the haircut costs for women. The interquartile range (IQR) for women is $55, while for men it is $10.75.

c. We should not be surprised that the mean is larger than the median because the distribution appears to be skewed to the right. The mean averages all the values in the data, so is "pulled" toward the high ones. The median is the 50th percentile and is resistant to the extreme values.

d. Since the median gives a better description of the center, or a "typical" haircut cost, it is more appropriate. Note that the mean for males is about equal to the 3rd quartile, indicating that 75% of the males paid less than the mean haircut cost for males. For women, the median is $31, indicating that half of women spent $31 or less, but the mean haircut cost for women is $52.53. The mean doesn't give us a good idea of what we could expect for a typical student's haircut cost. It is best to only use the mean when the data distribution is reasonably symmetric.