# **Data and Statistics Discussion Guide (for use during or after reading)**

1. What is the study of statistics? Provide an example and a non-example of a statistical question. (Data and Statistics, p. 4-5)
   1. Statistics is the study of data, which are a set of measurements or other facts. Statistical questions are answered by data that vary or differ. For example, “how fast did the sprinters run?” is a statistical question because there will be multiple answers to the question, and those answers will likely be different from one another. Non-statistical questions have just one answer. For example, “how fast did Molly run?” would be a non-statistical question because even though it can be answered with a number, there is only one correct answer for the question.
2. What are bar graphs? What are histograms? (Bar Graphs, Double Bar Graphs, and Histograms, p. 10-15)
   1. A bar graph uses lengths or heights of bars to show numerical data. Bar graphs help us visualize and compare data points easily. Double bar graphs are an extension of bar graphs and show two sets of data using lengths or heights of bars. Double bar graphs are great tools for visualizing and comparing data.
   2. Histograms are special types of bar graphs that represent data with lengths and heights of bars but show the number of data values within various intervals, or ranges. Histograms help us visualize data and describe the distribution of a data set.
3. What are line graphs and double line graphs? When are they useful and how do they represent data? (Line Graphs and Double Line Graphs, p. 16-19)
   1. Line graphs show the relationship between two variables and are plotted on coordinate grids. Line graphs are useful for tracking changes over short or long periods of time. Line graphs can help us visualize and see trends in the data that might be more difficult to see on a bar graph.
   2. A double line graph shows two sets of data, often using different colors to represent each set. Double line graphs can be useful when comparing two sets of data to one another.
4. What are circle graphs? When are they useful and how do they represent data? (Circle Graphs, p. 20-21)
   1. A circle graph, sometimes called a pie chart, uses a circle to show a whole and sections of the circle to show parts. We can think about a circle graph showing 100% of the data, and each piece showing how much of that 100% is made up of certain data points. Circle graphs can help us visualize and compare data within a data set.
5. Why might stem-and-leaf plots be useful tools for organizing data? (Line Plots and Stem-and-Leaf Plots, p. 22-23)
   1. A stem-and-leaf plot is a shorthand way of presenting data. Stem-and-leaf plots can be used to show a lot of data in a compressed and organized way. The stem represents the first digit (or set of digits) of each data value and the leaves represent the last digit of each data value. Sometimes, stem-and-leaf plots can be more useful than frequency tables or other representations because it is easier to see the distribution of data.
6. What does the word “distribution” mean in terms of statistics? What do mathematicians look for when describing distributions of data? (Distribution of Data, p. 24-25)
   1. In statistics, the term “distribution” refers to the shape of the data and the information we can infer from it. When mathematicians describe distributions, they consider how spread out the data points are from one another. In addition, they look for trends in the data, such as many points bunched together or whether the data seems evenly distributed. Understanding the distribution of data can help us better interpret data sets.
7. What are measures of central tendency? Describe the three main measures of central tendency and how they help us understand data sets. (Measures of Central Tendency, p. 26-27)
   1. Measures of central tendency are used to determine the central, or middle value, of a data set. There are three main measures of center: mean, median, and mode.
   2. The mean is the sum of the data divided by the number of data values. The mean is often called the average. When people talk about the average, they are often referring to a single number that can be used to generally represent the entire set of data.
   3. The median is another measure of central tendency. To find the median, line up all the data values and identify the one in the middle. If the median is close to the mean, the data set is likely somewhat evenly distributed.
   4. The mode describes the most common value in a data set. The mode is the only main measure of central tendency that can be used to describe categorical data.
8. What are measures of variability? Describe each and explain how they help us better understand data sets. (Measures of Variability: Range and Shape, p. 28-31)
   1. Measures of variability refer to how alike or different the values in a data set are. To describe the shape of a data set, think about how spread out the values are. If the data are all bunched together, they have low variability and if they are all spread out, they have high variability. A data value that is unusually large or small compared to the others is considered an outlier.
   2. The range describes the difference between the least and greatest values of a data set. When a data set does not have any outliers, the range can be used as a good indicator of the set’s variability.
9. How are box plots used to represent data? Why or when might this representation be useful? (Box Plots, p. 32-33)
   1. Box plots are used to summarize the statistics of a given data set. Box plots include the minimum, maximum, and median. In addition, they show the range of the entire data set as well as the range of the middle half of the data values.
   2. Box plots are useful because they not only provide a visual summary of the data, but also provide a quick look into a measure of central tendency, the distribution and shape of the data, and the range of the entire set.
10. What is statistical reasoning? Why is it important? (Statistical Reasoning, p. 34-35)
    1. Statistical reasoning describes the thought processes mathematicians use to analyze data and make inferences about it. Statistical reasoning includes considering different measures of central tendency and variability in order to make claims about the data set. Often, mathematicians will combine their knowledge of statistics with the context of the data set itself to make inferences and claims.