Chapter 11
Learning to Reason About Comparing Groups

As statistics moves to the forefront in education, much interest is developing around the process of comparing two groups ... (which) previews an important concept later developed in introductory college statistics courses: statistical inference.
(Makar & Confrey, 2002, p. 2)

Snapshot of a Research-Based Activity on Comparing Groups

Students are shown a bag of gummy bears (a rubbery-textured confectionery, roughly two cm long, shaped in the form of little bears) and two stacks of books: one is short (one book) and one is high (four stacked books). They are shown a launcher made with tongue depressors and rubber bands (see Fig. 11.1), and are asked to make a conjecture about how the height of a launching pad will result in different distances when gummy bears are launched. The students discuss different rationales for launches traveling farther from either of the height conditions. They are then randomly assigned to small groups to set up and gather data in one of the two conditions, each small group launching gummy bears 10 times to collect data for their assigned height (short or high stack of books).

Once the data are recorded, they are analyzed using boxplots to compare the results for the two conditions. The boxplots are used to determine that the higher launch resulted in further distances.

Students had previously completed an activity that showed them how dot plots can be transformed into boxplots, and are reminded again of the dots (individual data values) hidden within or represented by the boxplot. Their attention is drawn to two types of variability, the variability between the two sets of data (resulting from the two conditions) and the variability in the data: within each group (in each boxplot). Students recall earlier discussions in the variability unit on error variability (noise) and signals in comparing these groups, and they realize the need for an experimental protocol that will help to keep the noise small and reveal clearer signals, so that true difference can be revealed. This experiment is revisited in a later activity when they are able to use a protocol to gather data with less variability and analyze the difference using a t-test (in the Inference unit, see Chapter 13).

J.B. Garfield, D. Ben-Zvi, *Developing Students' Statistical Reasoning*,

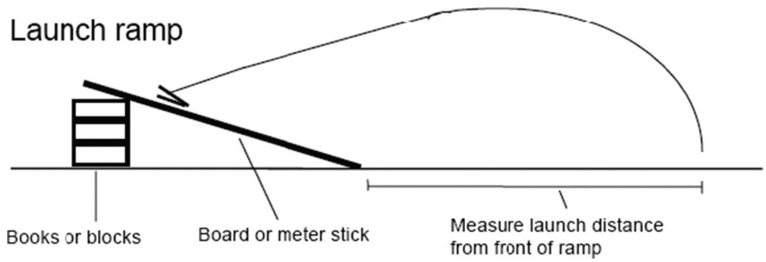

Fig. 11.1 Example of the launching set-up for gummy bears (from Cobb & Miao, 1998)

Rationale for This Activity

While this activity is often used to illustrate the principles of experimental design, we use it in this unit on comparing groups with boxplots for many reasons. First, it provides an interesting and motivating context for comparing two groups of data: to answer a question using an experiment. Second, this activity helps students deepen their understanding of variability, what causes it, how it affects an experiment, how it is revealed in graphs, and its role in comparing two groups of data. Finally, we believe that it is important to revisit principles of experimental design and methods of collecting data so that students can deepen their understanding of these concepts in different contexts and connect these principles to the new topics being studied.

The Importance of Reasoning About Comparing Groups

Comparing two groups of data is an intuitive and interesting task frequently used to engage students in reasoning about data. Many research studies compare two or more groups, either on an experimental variable (e.g., use of a new drug) or on an observational variable (e.g., gender, or age group). While many statistics courses first teach students to graph, summarize, and interpret data for a single group, often activities involving comparisons of more than one group are more interesting and provide the context for meaningful learning. While dotplots can be useful for comparing small data sets, we believe that boxplots are a very useful graphical representation for comparing larger data sets. Although boxplots are often very difficult for students to understand, we think that this graph is extremely useful because it facilitates the comparison of two or more groups, allowing for easy comparisons of center (median), variability, (range and interquartile range) and other measures of location (upper and lower quartile) as well as identifying outliers that may not be revealed in a histogram. While our lessons are designed to help students construct an understanding of boxplots as they may be used to compare data sets, we believe that it is always helpful for students to use different graphical representations in exploring and analyzing data.

The reasons for including "Comparing Groups" as a separate topic of instruction include:

1. Comparing two or more groups can be structured as an informal and early version of statistical inference, and can help prepare students for formal methods of statistical inference.
2. Problems that involve group comparisons are often more interesting than ones that involve a single group.
3. Research shows that students at all ages do not have good intuitive strategies for comparing groups and may have some common misconceptions regarding group comparisons, that need to be explicitly addressed in instruction (e.g., Konold et al., 1997).
4. Comparing groups motivates the need for and use of advanced data representations such as *boxplots*, a graphical display that is best employed in group comparison situations, but which is not easily understood or interpreted by students.

Comparing Groups with Boxplots

Boxplots are part of various graphical tools developed by Tukey (1977) for the purpose of analyzing data. In their review of the literature, Bakker, Biehler, and Konold (2004) suggested why educators began to introduce even young students to boxplots.

> First, the boxplot incorporates the median as the measure of center, and some early research had suggested that the median is easier for students to understand as a measure of center than is the mean (Mokros & Russell, 1995). Boxplots also provide, in the Interquartile Range (IQR), a measure of the degree of spread and an alternative to the computationally more challenging standard deviation (SD). (Besides, a clear geometrical interpretation of the SD can only be developed in the context of normal distributions.) Furthermore, boxplots depict both the measure of spread and center pictorially, which is largely why boxplots are such a powerful way to quickly compare several groups at once. Therefore the boxplot and the interquartile range promised to provide better tools for developing an initial feeling for spread than other graphs and measures of spread. (p. 164–5)

Bakker et al. (2004) describe boxplots as "conceptually rich" tools. To understand them, interpreters need at least to know what minimum, first quartile, median, third quartile, and maximum are. In many situations, they need to understand that the median is used as a measure of the center of a distribution; that the length of the box (not its width) is a measure of the spread of the data; and that the range is another measure of spread" (p. 166).

TinkerPlots, software for precollege-level students (Konold & Miller, 2005; http://www.keypress.com/tinkerplots), includes a simple graphic display called the "hatplot" that can be used to guide students to the more sophisticated idea of a boxplot. Each hat is composed of two parts: a central "crown" and two "brims" on each side of the crown. The "crown" is a rectangle that, in the case of *percentile hatplot*, shows the location of the middle 50% of the data – the Interquartile Range (IQR). The brims are lines that extend out to the minimum and maximum values of the data set. There are four different options for how the crown of a hatplot is formed: based on percentiles (the default, see example in Fig. 11.2), the range,

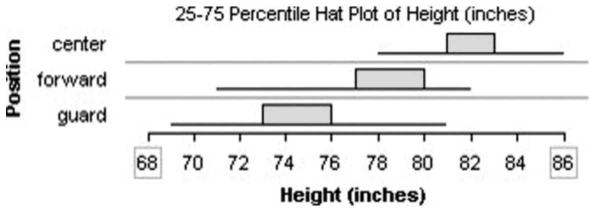

Fig. 11.2 Three parallel percentile hatplots of basketball player's height by their position in *TinkerPlots*

the average deviation, and the standard deviation. Thus, hatplots can be seen as a generalization of a boxplot, and may provide means for allowing students to build on intuitive ideas they have about distributions (Konold, 2002a).

The Place of Comparing Groups in the Curriculum

While this topic can be introduced early in a class, the formal study of comparing groups and boxplots usually takes place after students have studied measures of center and spread, as well as the topic of graphical representations of distributions. However, the ideas of quartile and interquartile range may be introduced at the same time students are learning about boxplots.

After students study this topic, it is helpful to combine all the topics of data analysis and examine them together before moving on to topics leading to statistical inference. We, therefore, offer suggestions in this chapter for activities that integrate ideas of distribution shape, center, and spread, along with comparison of different methods of graphically representing data. Informal inferences are made when comparing groups in this part of a course, laying the foundations for more formal study of statistical inference (see Chapter 13). When groups are compared later on, such as in two sample tests of significance, boxplots are used again to help examine variability between and within groups. Boxplots are revisited again in the unit on covariation (Chapter 14) when multiple boxplots are graphed over time, and the medians help students spot a linear trend.

Review of the Literature Related to Reasoning About Comparing Groups

Studies on comparing groups have focused on how learners approach this topic, what their typical strategies and difficulties are, and how to help them develop their reasoning about comparing groups. Early works indicated and demonstrated that the group comparison problem is one that students do not initially know how to approach and encounter many difficulties with negotiating comparison strategies. Various strategies to improve students reasoning about comparing groups were studied, as well as the role of graphical representations (emphasis on boxplots) in supporting making sense of comparing data sets situations.

Difficulties in Reasoning About Comparing Groups

Primary school students' use of intuitions and statistical strategies to compare simple data sets using line plots was explored by Gal et al. (1989, 1990). Although some students in their study used statistical strategies for comparison, many others focused only on some features of the data, but did not offer a complete synthesis. Others have used incorrect strategies, such as finding totals when they were inappropriate due to different sample sizes, or inventing qualitative explanations such as being better because data are more spread out.

In a follow up study, Watson and Moritz (1999) further identified and detailed categories of school students' reasoning in comparing two data sets. In their study, 88 students in grades three to nine initially compared data sets of equal sizes, but were not able to attend to the issue of unequal sample size. Only in higher reasoning levels, the issue of unequal sample size was resolved with some proportional strategy employed for handling different sizes. The researchers recommend the use of a combination of visual and numerical strategies in comparisons of data sets, "hopefully avoiding the tendency to "apply a formula" without first obtaining an intuitive feeling for the data sets involved" (p. 166).

This recommendation is supported by additional studies showing that students who appear to use averages to describe a single group or know how to compute means did not use them to compare two groups. Gal et al. (1990) found that sixth and ninth grade students did not resort to proportional reasoning or visual comparison of graphs to reach appropriate comparing groups conclusions. Difficulties were found in a case study of two pairs of high school students who were interviewed after a year-long course in which they had used a number of statistics including means, medians, and percents to make group comparisons (Konold et al., 1997). In this study, students did not use any of these comparison techniques during the interview. The researchers claimed that the students' failure to use averages when comparing two groups "was due in part to their having not made the transition from thinking about and comparing properties of individual cases, or properties of collections of homogeneous cases, to thinking about and comparing group propensities" (p. 165). It seems, therefore, that one challenge in instruction of this topic is to make students comfortable summarizing a difference by comparing two groups using some representative measure of center (see Chapter 9 on reasoning about center), a prerequisite to understand the rationale of statistical inference in their advanced studies.

Konold and Higgins (2003) suggested that students' difficulties in comparing groups stemmed from their initial inability to apply "aggregate-based reasoning" – understanding a distribution as a whole, an entity that has many features such as center, spread, and shape (See Chapter 8 on the concept of distribution). Bright and Friel (1998), for example, found that eighth grade students using a stem-and-leaf plot to compare groups, could identify a "middle clump" (where the majority of values are) in a single distribution, but could not use this information to make comparisons. Several students compared just selected individuals from each group. Ben-Zvi (2004b) similarly describes how seventh grade students attend to local details of the comparisons, such as comparing the difference between two cells in frequency table, the difference in heights of two adjacent bars in a double bar chart, or comparing disjoint edge values in the distributions, but find it hard to spot and describe the difference between the two distributions as a whole.

Instructional Approaches to Develop Reasoning About Comparing Groups

Researchers have offered different methods, instructional materials and sequences, and technological tools to overcome these difficulties. For example, Cobb (1999) suggests that the idea of "middle clumps" ("hills") helps students gradually develop their reasoning about comparing groups. Students in seven and eight grades, who used the *Minitools* software (Cobb et al., 1997), began to make decisions about group difference by comparing the numbers of cases in each group within narrow intervals of the range, and gradually moved to referring to global features of the distributions such as shape, center, and spread.

The introduction of new technological tools to support students' reasoning about comparing distributions has created new opportunities in the pedagogy and research of this topic. Hammerman and Rubin (2004) describe how teachers used a new dynamic data visualization tool (*TinkerPlots*) to divide distributions into slices and consequently compare frequencies and percentages within these slices to make inferences (see grey-shaded "slices" in Fig. 11.3). The type of thinking observed was "slice-wise comparison across groups," which tended to ignore the distribution as a whole. The researchers suggest that using this new tool engendered and made visible thinking that had previously lain dormant or invisible. These teachers' slice-wise comparison reasoning seemed to be an extension of the "pair-wise comparison" type of reasoning, which involves comparisons of two individual cases or data values, that other researchers have documented (e.g., Ben-Zvi, 2004b; Moritz, 2004).

In a follow up study, Rubin et al. (2005) found that teachers characterized data using both traditional aggregate measures such as the mean and median as well as novel methods for looking at data such as numbers or percentages around cut points, modal clumps, and overall shape. Teachers using *TinkerPlots* and *Fathom* (Key Curriculum Press, 2006; http://www.keypress.com/fathom) increased their confidence in what these measures were telling them when the stories each measure or characterization told pointed in the same direction. Similarly, when multiple samples from the same population gave some consistency in measures, their confidence in the measure was increased (such as in Fig. 11.3). By contrast, when measures pointed in different directions, teachers were found to spend time further exploring the data so that they can better understand what story was really being told, and

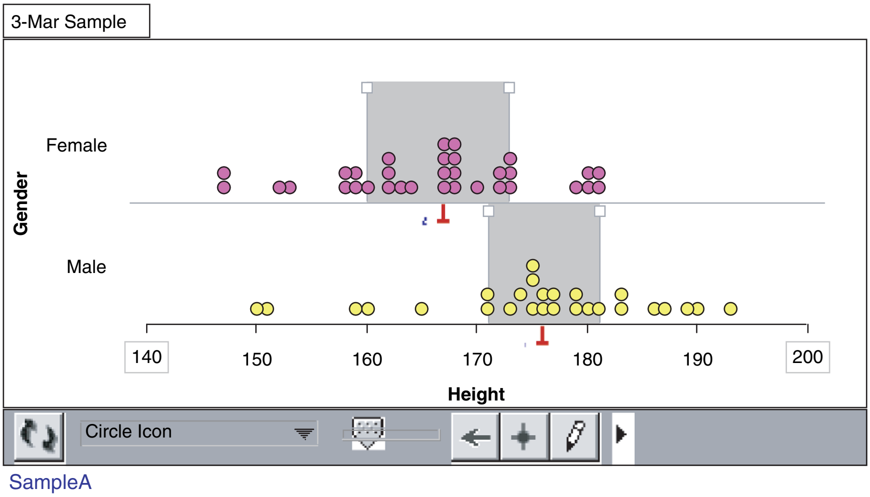

Fig. 11.3 Height data by gender, with mean, median, and IQR marked (From Rubin, Hammerman, Campbell, and Puttick, 2005)

exploring the meaning of the measures to understand what each was telling them. Such explorations become necessary in part when the shape of data is not bell-shaped and symmetric (e.g., skewed distributions), suggesting the importance of learners working with data sets of various shapes in order to more robustly understand the meaning behind various data analytic tools.

Two different kinds of measures of data distributions – *rule-driven* measures and *value-driven* measures were identified by Rubin et al. (2005) as they studied the development of teachers' reasoning about comparing groups. While both of these can describe data in an aggregate way, the researchers believe that value-driven measures are easier to use at first, perhaps until the meaning and implications of the rule that produce a rule-driven measure are clear. They also described how some people use rule-driven measures to create a value around which to make a value-driven comparison, and speculated about the relative power of using such a value rather than one chosen at random, although context driven values might be more powerful still.

Research on Learning to Understand Boxplots in Comparing Groups Situations

As described earlier, the boxplot is a valuable tool for data analysis. The use of boxplots allows students to compare groups of data by examining both center and spread, and to contrast from within group variability to between groups variability. However, several research studies have identified problems students having understanding and reasoning about boxplots. For example, Bakker et al. (2004) claim that several features of boxplots make them particularly difficult for young students to use in authentic contexts. For example, boxplots obscure information on individual cases, the median (shown by the line in the box) does not appear to be as intuitive to students as a measure of center, and the use of quartiles (to construct the box and show the upper and lower boundaries of the box) are difficult for students to fully understand. Bakker et al. (2004) suggest an explanation for these difficulties.

> Quartiles are particularly tricky. Not all integers can be divided by 4, and there is the additional complexity of how to deal with cases that have the same value. There are different ways of doing this, and thus different definitions of quartiles. Computer programs use different definitions, and these definitions are not always well-documented (Freund & Perles, 1987) … Quartiles do not match well the way students tend to conceive of distributions. In several recent studies, researchers noted that students tend to think of a distribution as comprising three parts, rather than four. They think about (a) the majority in the middle (which usually includes more than 50% of the cases); (b) low values; and (c) high values (Bakker & Gravemeijer, 2004; Konold, Robinson, Khalil, Pollatsek, Well, Wing, & Mayr, 2002). Students also referred to the center majorities as "clumps," which was why Konold and colleagues (2002) propose calling them "modal clumps." (p. 167–8)

In light of these major hurdles, Bakker et al. (2004) recommend that educators consider the various features of boxplots and carefully determine whether, how, and when to introduce boxplots to students at a particular grade level.

Boxplots are difficult even for teachers to fully understand. In a study of secondary teachers at the end of a professional development sequence, Makar and Confrey (2004) used interviews to study how teachers reasoned with boxplots with *Fathom* to address the research question, "How do you decide whether two groups are different?" The researchers found that the teachers were generally comfortable working with and examining traditional descriptive statistical measures as a means of informal comparison. However, they had major difficulties in regard to variability, in particular how to (1) interpret variability within a group; (2) interpret variability between groups; and (3) distinguish between these two types of variability.

Implications of the Research: Teaching Students to Reason About Comparing Groups

The research studies highlight that students have many difficulties understanding comparing groups, boxplots and the related ideas of quartiles, median, and interquartile range. It is not intuitive for students to look at data as an aggregate when comparing groups, so they need to be guided in this process. There are many times in an introductory statistics course when it is appropriate to compare two or more sets of data, and over time the guidance can be decreased. It may help to begin with more informal intuitive comparing methods first and then eventually move to more formal methods.

Group comparisons require students to revisit and integrate previously learned ideas about distribution: shape, center, and spread. Difficulties students have in roughness of quartiles and various methods for finding them, can be helped if students find quartiles by dividing data "roughly" into four groups, and not worrying about more precise computational details.

In order to understand the ideas of center and variability represented in the boxplots, students should have many opportunities to look at multiple graphs and multiple statistics for the same variables, so that they may see how these ideas are reflected using these different types of summaries. This also helps them to see that we do not just compare means or medians when comparing groups, but we also need to examine variability. Statistical thinking should be modeled for students as comparison of groups involves discussions of variability between and variability within groups, and how that affects inferences, even if they are informal. For example, even though one group has a high mean than the other, there is so much scatter and spread in the groups that it is hard to tell what the "trend" or "signal in the noise" is.

Since students often do not see the data values hidden in a boxplot, they tend to equate length of whiskers or width of the box with amount of data. Therefore, students need opportunities to see the data behind the box, using physical and computer examples. Students may confuse the height of a horizontal boxplot with frequency of data, so it is important to have students notice and play with this dimension so they realize that it does not indicate anything about the variable or its frequency. Finally, counterintuitive examples may help students improve their understanding and reasoning, such as presenting students with two groups of data where one has a higher interquartile range, but the other has a higher standard deviation, and why these are different.

The Role of Technology in Helping Students to Reason About Comparing Groups

Research suggests that students should be scaffolded to reason with boxplots through keeping the data in dotplot form, under the boxplots (Bakker et al., 2004). We find *TinkerPlots* useful for helping students learn to understand and reason about boxplots. This tool allows students to see how a dotplot can be transformed to a boxplot, first showing where the dots are in boxplot before they are hidden.

While different Web applets exist for boxplots, they typically show the five number summary for a boxplot of data, allow one group to be expanded into several boxplots based on a categorical variable (e.g., http://www.shodor.org/interactivate/activities/boxplot/?version=1.4.2 & browser=MSIE & vendor=Sun _ Microsystems _ Inc.0), or rotate back and forth between a boxplot and histogram of data (e.g., http://nlvm.usu.edu/en/nav/frames_asid_200_g_4_t_5.html?open=instructions). These applets can be useful to help students interpret boxplots and learn how different features of data sets are presented differently in a histogram or a boxplot.

Progression of Ideas: Connecting Research to Teaching

Introduction to the Sequence of Activities to Develop Reasoning About Comparing Groups

Students begin comparing groups informally when they examine graphs for different variables and data sets in the earlier unit on distribution, and as they examine graphs of data in units on center and variability. Boxplots are introduced as a more formal method of comparing two groups of data, once students have already studied basic ideas of center and spread. Discussions can be focused on what we compare, when we compare two dotplots, and which is center and spread. Students discuss where the middle half of the data is in the groups being compared, which can be graphically illustrated using a hatplot (from *TinkerPlots*). This allows students to see the data set as an entity rather than as points or slices of data (see Chapter 8), and to compare the middles of the two data sets. Medians can then be added to the hatplots, which transforms them into boxplots. Once boxplots have been introduced, students should be encouraged to make the connections back to dotplots, seeing how the dots in a dotplot map to a boxplot, an idea that is often hidden and confusing to students. Advantages of using boxplots to compare groups can be examined, as students see that it is easy to compare both center and spread simultaneously when comparing boxplots for data sets.

In order to further develop students' reasoning about boxplots, students can be given sets of boxplots and histograms and match the two graphs that are for the same set of data, allowing then to think about how features of a histogram would show up in a boxplot (e.g., symmetry, skewness, outliers) and vice versa. Students can then be given different sets of boxplots to compare as they answer research questions about how these plots reveal group differences. This can lead to informal inferences with boxplots as students consider differences in means relative to variability. Table 11.1 shows a series of steps that can be used to help students first build informal ideas and then formal ideas of comparing groups with boxplots.

Introduction to the Lessons

There are four lessons that lead students to compare groups and develop the idea of boxplot as a graphical representation of data that reveals both center and spread and facilitates comparisons of two or more samples of data. The lessons begin with a comparison of two brands of raisins to show that boxplots help in making comparisons and informal inferences. Then students are guided to examine more carefully the characteristics of a boxplot, moving from a dotplot to a hatplot to boxplot, to show how the dots are hidden by the plot, and what the parts of the box represent. The second lesson has students make informal inferences using boxplots to compare distances for Gummy Bears launched using two different heights for launching pads and focuses on comparing groups of data using boxplots. The third lesson develops students' understanding and use of boxplots by having them interpret boxplots in

Table 11.1 Sequence of activities to develop reasoning about comparing groups with boxplots[1]

Milestones: ideas and concepts	Suggested activities
Informal ideas of comparing groups	
• Informal comparisons of dot plots and histograms	• Activities in Lessons 1 and 2 of the Distribution Unit (Chapter 8)
• Comparison of graphs to determine which has a higher and lower standard deviation	• What Makes the Standard Deviation Larger or Smaller Activity? (Lesson 2, Variability Unit, Chapter 10)
Formal ideas of comparing groups with boxplots	
• Data as an aggregate rather than points and slices when comparing groups	• How Many Raisins in a Box Activity (Lesson 1: "Understanding Boxplots")
• How a boxplot represents a data set, how points are "hidden" in a boxplot	• How Many Raisins in a Box Activity (Lesson 1)
• Coordination of comparisons of center and spread in comparing groups	• Gummy Bears Activity (Lesson 2: "Comparing Groups with Boxplots")
• How variability between groups and variability within groups are used in comparing groups	• Gummy Bears Activity (Lesson 2)
• Advantages of using boxplots to compare groups	• Comparing Boxplots Activity (Lesson 2)
• How to make informal inferences from comparisons of samples of data using boxplots	• Interpreting Boxplots Activity (Lesson 3: "Reasoning about Boxplots")
• Understanding how features of data are revealed in different graphs of the same data	• Matching Histograms to Boxplots Activity (Lesson 3)
• Integrating reasoning about shape, center, and spread in different graphical representations	• How do Students Spend Their Time Activity (Lesson 4: "Comparing Groups with Histograms, Boxplots, and Statistics")
Revisiting the idea of comparing groups in subsequent units	
• Variability between groups and variability within groups when making formal inferences involving two samples of data	• Gummy Bears Revisited Activity (Lesson 4, Statistical Inference Unit, Chapter 13)

answering different research questions, and then match boxplots to histograms. The final lesson, integrates all the main ideas in data analysis as students use boxplots (and other graphs and statistics) to analyze a multivariate data set, exploring which variables have larger and smaller amounts of variability.

[1] See page 391 for credit and reference to authors of activities on which these activities are based.

Lesson 1: Understanding Boxplots

This lesson introduces the boxplot as a way to graphically compare two or more groups of data. It has students progress from comparing groups with dotplots, to using hat plots (a feature of *TinkerPlots*) and finally moving to boxplots. By using *TinkerPlots*, students are able to see the data values "hidden" in a boxplot. Students then examine and compare two groups of data in a series of questions using boxplots. Student learning goals for this lesson include:

1. Understand that a boxplot shows where certain percentages of data lie.
2. Understand that a boxplot offers a good way to compare groups of data.
3. Begin to reason about comparing groups using boxplots.
4. Learn how to read and interpret boxplots.
5. Become more fluent in comparing groups of data by comparing shapes, centers, and spreads of two data sets given in boxplots.

Description of the Lesson

The lesson begins with a question about how different brands of the same food product vary, and whether all similar products (of the same size) give the same amount (e.g., number of M&M candies in a small bag, or "does the same size bag of potato chips from two competing companies, give the same amount of chips in each box?"). Students are asked how they can make an informed decision about which product to purchase, and this leads to the need to collect and examine some data.

In the *How Many Raisins in a Box* activity, students are given small boxes of raisins and data are collected on the number of raisins in each box for two competing brands. The data are first collected as two dotplots, but then the class discusses a better way to graphically compare the two data sets. *TinkerPlots* is used to help develop an understanding of a boxplot. First students talk about ways to compare the two data sets; one option is to compare where most of the data are, and then where the middle halves of the data are. The hatplots graphs in *TinkerPlots* are used, where the "hat" is the middle fifty percent of the data, and the outer brims are the remaining quarters of the data set, as shown in Fig. 11.4.

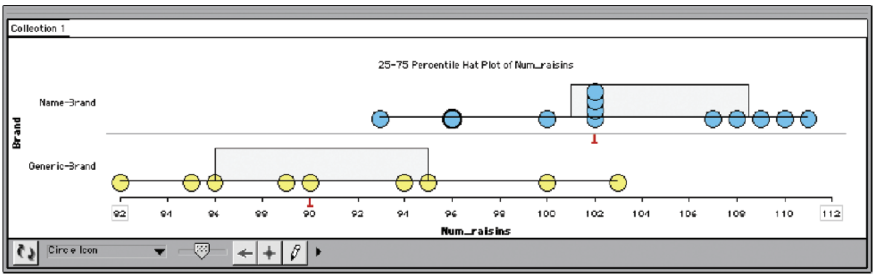

Fig. 11.4 Comparing two data sets (brand of raisins) using the hatplot graph in *TinkerPlots*

Next, the medians are added to the plots by clicking on the *Median* icon (shown in Fig. 11.4 as red "⊥"). Students can count the data values above and below the median and the values in each part of the hatplot.

Finally, the hatplots are converted to boxplots and the individual data values can be hidden (Fig. 11.5). By going back and forth between the hatplot that shows the data values and the boxplot that hides the data values, students are led to see that the two plots include the same data points (number of raisins in a box), that the box of the boxplot is the hat of the hatplot, and that the "whiskers" of each boxplot also includes the same data points as the "brims" of the hat in the hatplot. They can see that the median is now included inside the box of the boxplot as well, and that this is the only important difference between the two plots other than hiding the data values. Students see the individual data points disappear as they go from hatplots to boxplots, illustrating how the boxplot represents the same number of data points (boxes of raisins), the median is still in the same place, and that there are equal numbers of boxes of raisins on either side of the median and in each whisker.

A discussion follows on how boxplots help compare the two brands of raisins showing differences in the center and spread of the numbers of raisins per box. They discuss why this difference exists as well as why there is variability from box to box, and come up with reasons for the two types of variability, within and between brands of raisins. They also make inferences about what they believe to be true for the larger population of boxes of raisins for each brand, based on these samples of data, making informal inferences.

Students then try to reason about and draw two boxplots, with 20 data values each, so that one has a long tail and one has a short tail, but both have five data values in the tail. Next, they reason about and draw a boxplot that would have the mean equal to a quartile, and then two different boxplots that both have ten data

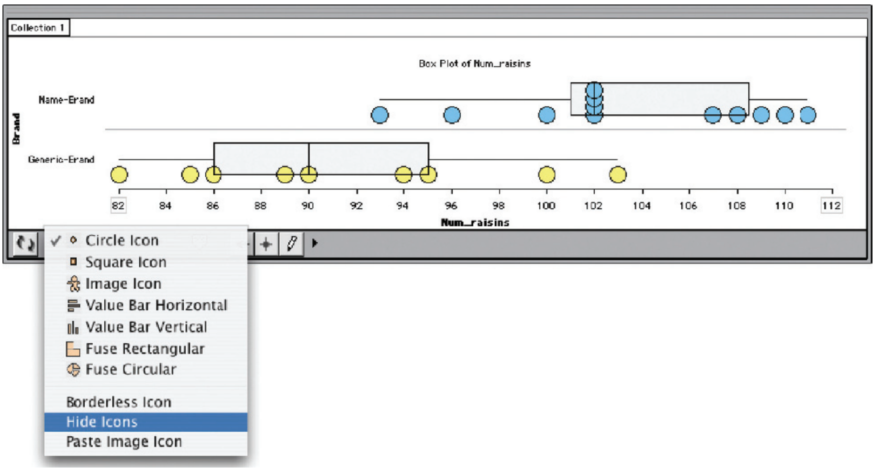

Fig. 11.5 Converting hatplots to boxplots and the option of hiding the individual data values in *TinkerPlots*

values lower than the median. How outliers are determined and represented may also be discussed along with how *Fathom* and other statistical software packages represent outliers on boxplots.

In a wrap-up discussion, students consider pairs of dotplots, histograms and boxplots, and discuss which type of graphical display makes it easier to identify shape, center, and spread, and which type makes it easier to compare groups of data.

Lesson 2: Comparing Groups with Boxplots

This lesson continues to use boxplots to compare groups, but this time the focus is on an experiment. Students make and test conjectures about how the height of a launching pad will result in distances when gummy bears are launched. Two types of variability are examined: the variability within each group and the variability between groups. This also help students distinguish between error variability (noise) and signals (trends), in comparing groups, and then realize the need for little noise and clearer signals, revisiting these ideas from the center and variability units (see Chapters 9 and 10). Student learning goals for this lesson include:

1. Use boxplots as a way to compare results of an experiment.
2. Deepen understanding of boxplots as a graphical representation of data.
3. Use boxplots to visually represent different types (sources) of variability (when it is desired and when it is noise).
4. Revisit the ideas of mean as signal and variability as noise, from repeated measurements in an experiment.
5. Recognize stability of measures of center as sample size increases. When sample grows, see how measure of center predict center of larger population, and how it stabilizes (varies less) as sample grows.
6. Distinguish between variability within treatments and variability between treatments.
7. Understand that it is desirable to reduce variability within treatments (by using experimental protocols).
8. Revisit idea that the only way to show cause and effect is with a randomized experiment.

Description of the Lesson

Students are shown a gummy bear and a launching system made from a tongue depressor and rubber bands. They make conjectures about the following question:

> *Will gummy bears travel a farther distance if they are launched from a steeper height or a lower height? (A stack of four books, or one book)*

They are then given supplies and told how to launch gummy bears and how to measure the distances that they travel. They are asked how students should be assigned to conditions, so that the results may be used to infer cause and effect relationships. Then, randomization is used to assign them to a group that will gather data for one of the two conditions. Students working in groups gather data for their condition: height of one book or height of four books. Data are gathered for 10 launches, and recorded in a table. Data are collected from each group and entered into *Fathom*. Students are asked how they think the data should be summarized and graphed so that they can compare the difference in distances for the two conditions. Various summaries can be generated and various graphs can be examined. Boxplots of a set of data are shown in Fig. 11.6.

Fig. 11.6 Boxplots comparing the difference in gummy bear distances for the two conditions in *Fathom*

The following questions are used to guide the discussion of results:

- Is there variability in the measurements for each condition? How do we show that variability?
- Is there variability between the two groups (conditions)? How do we look at and describe that variability?
- Why did we get different results for each group within a condition?
- What represents the signal and what represents the noise for each condition?
- How could we get the signal clearer? What would we have to do? (e.g., add more teams to each condition? Have each team launch more bears?)
- If we made a plot of the sample means from each group, how much variability would you expect to see in the distribution? Why?
- Based on our experiment, are we willing to say that a higher launch ramp caused the gummy bears to go farther? What are important parts of an experiment that are needed in order to show causation?

- What are some different sources of variability? There are two kinds of variability: “diversity” and “error, or noise.” Which do we like to have large? Which do we like to have small? Why?

The second activity, *Comparing Boxplots,* focuses student’s attention on the different kinds of information in a boxplot (e.g., quartiles) and how these can be used in comparing groups. In a wrap-up discussion, students summarize and explain how boxplots help make the comparison of results more visual and apparent, and how they help us examine signal and noise in this experiment.

Lesson 3: Reasoning About Boxplots

This lesson consists of activities that can be used to help students develop their reasoning about boxplots and to deepen their understanding of the concepts of distribution, center, and spread, and how they are interrelated. There are two activities. One has students practice comparing boxplots and second has students try to compare and match histograms to boxplots for the same variables. Student learning goals for this lesson include:

1. Gain experience in using boxplots to compare data sets and draw informal inferences about the populations represented.
2. Move from scaffolded questions to guide their interpretation and comparison of boxplots to situations where the scaffolding is removed and having to analyze the boxplot comparison without guidance.
3. Deepen their reasoning about different representations of data by having to match different graphs of the same data.

Description of the Lesson

In the first activity, *Interpreting Boxplots*, students compare and interpret boxplots. They are given different research questions along with two side by side boxplots. They are asked questions that guide them to make comparison based on the boxplots. The early questions direct their attention to percentages of data in different parts of the boxplot as shown in Fig. 11.7.

The following graph shows the distribution of ages for 72 recent Academy Award winners split up by gender (36 females and 36 males). Use the graph to help answer the following questions.

a) Estimate the percentage of female Oscar winners that were younger than 40.
b) The oldest 50% of male Oscar winners are between which two ages?
c) What would you expect the shape of the distribution to be for male Oscar winners? Explain.

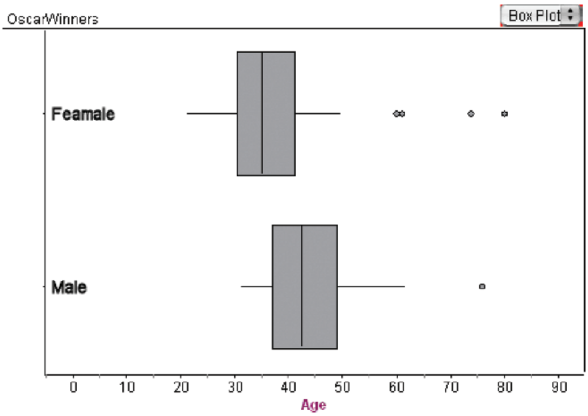

Fig. 11.7 The distribution of ages for 72 recent Academy Award winners split up by gender

d) Explain how to find the Interquartile Range (IQR) for the female Oscar winners.
e) Now, find the IQR for the female Oscar winners.
f) What information does the IQR of the female Oscar winners offer us? Why would a statistician be more interested in the IQR than in the range?
g) Compare the medians for male and female Oscar winners. What do you conclude about the ages of male and female Oscar winners? Explain.
h) Compare the IQR for the male and female Oscar winners. What do you conclude about the ages of male and female Oscar winners now? Explain.

Then other graphs are given with more open ended questions and students work in pairs to discuss and answer these questions. A class discussion allows comparison of answers and explanations of student reasoning.

In the second activity, *Matching Histograms to Boxplots*, students are given a set of five histograms and a set of five boxplots as shown in Fig. 11.8.

Students match each histogram to a boxplot of the same data. This activity requires them to think about how shape of a histogram might be represented in a boxplot, how the median shown in a boxplot might be located in a histogram, and how spread from the center is represented in both types of graphs (e.g., a histogram that is more bell shaped has more clustering to the center and therefore would show a smaller IQR as represented by the width of a boxplot).

A group discussion follows where students are asked which graphs were the easiest to match and why, and which were the most difficult to match and why. They identify how they made the matches, making their reasoning explicit. In a final wrap-up discussion, students are asked what different information is given by histograms and boxplots, and what similar information each provides. They comment on when it is better to use a histogram or a boxplot for a data set and they come to realize the importance of looking at more than one graph when analyzing data.

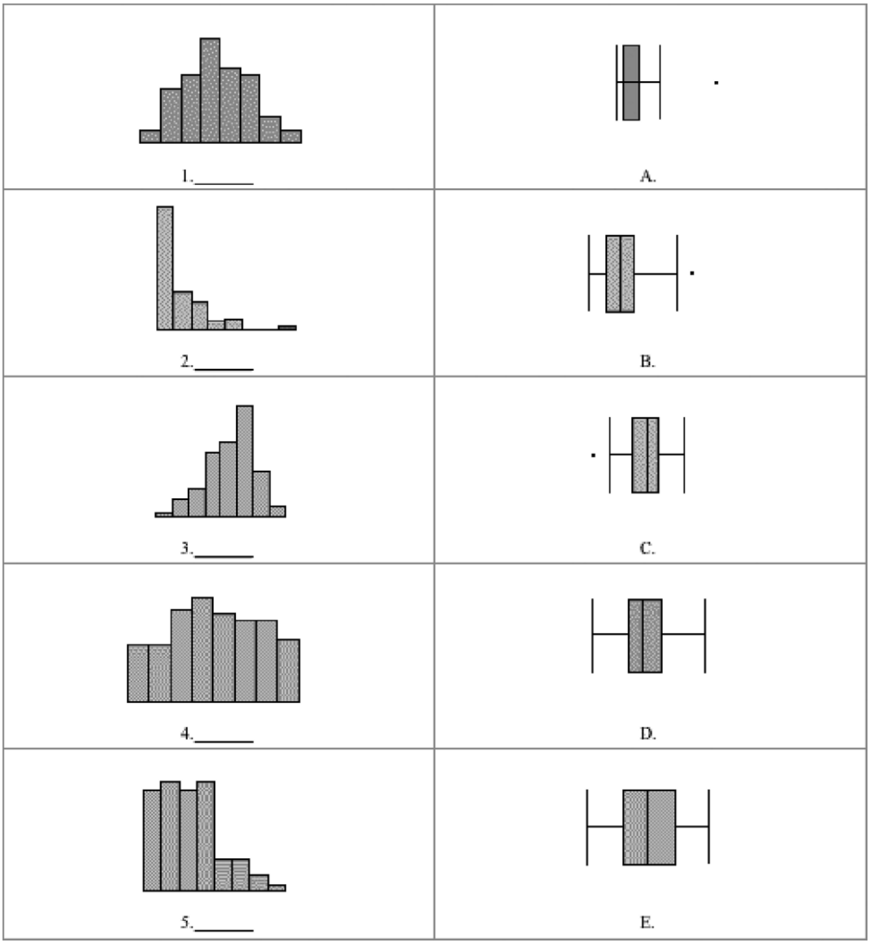

Fig. 11.8 Matching histograms to boxplots activity

Lesson 4: Comparing Groups with Histograms, Boxplots, and Statistics

This lesson builds upon and integrates the ideas of distribution: shape, center, and spread as they analyze a multivariate data set. Students make and test conjectures about variability expected for different variables, and then use graphs and statistics to test their conjectures. The lesson shows that in analyzing real data, we draw on a variety of methods and the answers we give depend on the methods we use. The analysis of multivariate data challenges students to see what they can learn from these data about how students spend their time. Student learning goals for this lesson include:

1. Review the concepts of distribution, center, and spread.
2. Understand how the concepts of distribution, center, and spread are related.
3. Know when to use each type of measure of center and variability.
4. Use boxplots to compare groups.
5. Realize that more than one graph is necessary to understand and analyze data, and that while boxplots are useful to compare groups, histograms (or dotplots) are also needed to better see the shape of the data.
6. Informally analyze a multivariate data set to find answers to open-ended questions that have different possible solutions.

Description of the Lesson

In the activity, *How do Students Spend their Time,* students consider and discuss how similar or different students are in their class in how they spend their time each day. Students are then divided into groups of three or four and predict the average number of minutes per day that students in this class spend on various activities. They record their predictions in a table (Table 11.2).

Students consider the variables and discuss with their group how much variability they would expect to see for each one as well as the shape of the distribution. Then they identify one of the variables that they think will have *little* variability and why they would expect this variable to have a *little* variability, discuss the shape of this distribution and then draw an outline of what they expect this graph to look like, labeling the horizontal axis with values and the variable name and where they expect the mean or median to be. After sharing their results in a whole class discussion, they repeat this activity for a variable that think would have *a lot* of variability.

Data gathered on the First Day of Class Survey (described in Chapter 6 on data, and converted from hours per week to minutes per day) are examined, using software, so that students can compare their predictions to the actual results, discussing any differences they found. In the last part of the activity, students compare side by side boxplots, histograms, and summary statistics for the entire data set of daily times. Students consider and discuss what information is shown in each graph, about

Table 11.2 Students' prediction table in the *How do students spend their time?* activity

Variable	Activity	Prediction of Average Time Spent (minutes per day)
Travel	Traveling to school	
Exercise	Exercise	
Parents	Communications with parents by email, phone, or in person	
Eating	Meals and snacks	
Internet	Time on the Internet	
Study	Study time	
Cell phone	Talk on cell phone	

the variability of the data, and what each summary statistics tells them as well as which graphs and statistics are most appropriate for summarizing each variable. Using all of these data, students then discuss and determine which variable has the smallest and which has the largest amount of variability and justify their answers.

In a wrap-up class discussion, results are compared and the issue emerges that you can answer this question in different ways, depending on the choice of graphs or statistics used. For example, interquartile range may be larger for one variable when you show boxplots, but the standard deviation may be larger for another variable because of outliers in the data set. Students then revisit what each measure of variability tells and how these relate to measures of center and shape of distribution. Students come to explain that there is no simple answer, and the shape, center, and spread are all interconnected. For example, for a skewed distribution with outliers, it is not helpful to use the standard deviation as a measure of variability. Also, it is not helpful to only consider variability; these measures need to be examined along with measures of center in order to meaningfully describe and analyze data. Students also may comment that side by side boxplots were much easier for comparing all the variables than individual dotplots or histograms.

Summary

The activities in this unit provide an important bridge from concepts of distribution, center, and variability (the elements of data analysis) to the ideas of statistical inference. At the same time, the topic of comparing groups helps students integrate and build on ideas of shape, center, and spread, learned in the previous units. Because research has suggested that students often fail to understand or correctly interpret boxplots, we have described a full sequence of activities that are designed to help students better understand and reason about boxplots as a method of graphically representing data as well as an efficient way to compare groups. Without such a careful progression of ideas along with software to help students see the points hidden by the graph, we do not believe most students will understand and correctly use and interpret these graphs.