Appendix for Level A

What Are Common Name Lengths?

Formulate Questions

During the first week of school, a third-grade teacher is trying to help her students learn one another's names by playing various games. During one of the games, a student named MacKenzie noticed she and her classmate Zacharius each have nine letters in their names. MacKenzie conjectured that their names were longer than everyone else's names. The teacher decided that this observation by the student provided an excellent opening for a statistics lesson.

The next school day, the teacher reminds students of MacKenzie's comment from the day before and asks the class what they would like to know about their classmates' names. The class generates a list of questions, which the teacher records on the board as follows:

→ Who has the longest name? The shortest?

→ Are there more nine-letter names or six-letter names? How many more?

→ What's the most common name length?

→ How many letters are in all of our names?

→ If you put all of the eight- and nine-letter names together, will there be as many as the five-letter names?

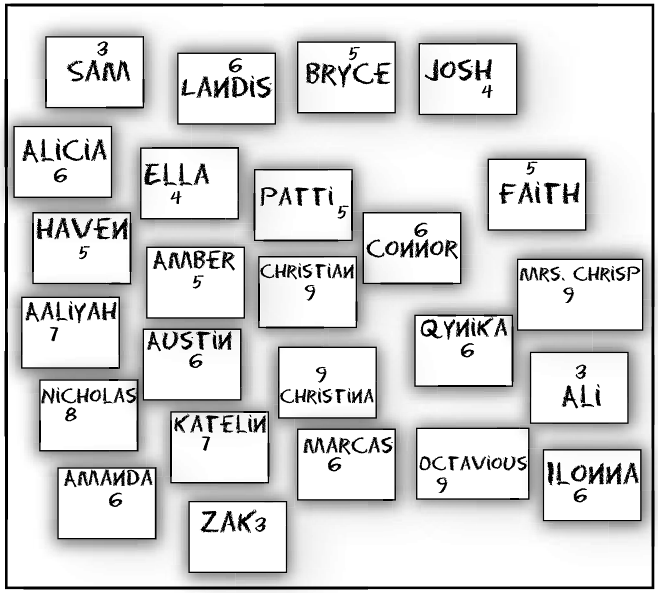


Figure 36: Random placement of names

Collect Data

The statistics lesson begins with students writing their names on sticky notes and posting them on the white board at the front of the room. This is a census of the classroom because they are gathering data from all students in the class.

Given no direction about how to organize the notes, the students arbitrarily place them on the board.

In order to help students think about how to use graphical tools to analyze data, the teacher asks the students if they are easily able to answer any of the

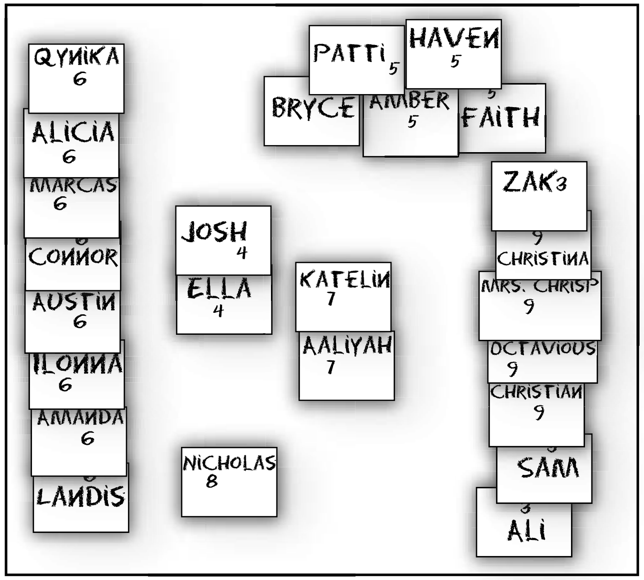


Figure 37: Names clustered by length

posed questions now by looking at the sticky notes, and the students say they cannot. The teacher then suggests that they think of ways to better organize the notes. A student suggests grouping the names according to how many letters are in each name.

The teacher again asks if they can easily answer the questions that are posed. The students say they can answer *some* of the questions, but not easily. The teacher asks what they can do to make it easier to answer the questions. Because the students have been constructing graphs since kindergarten, they readily an-

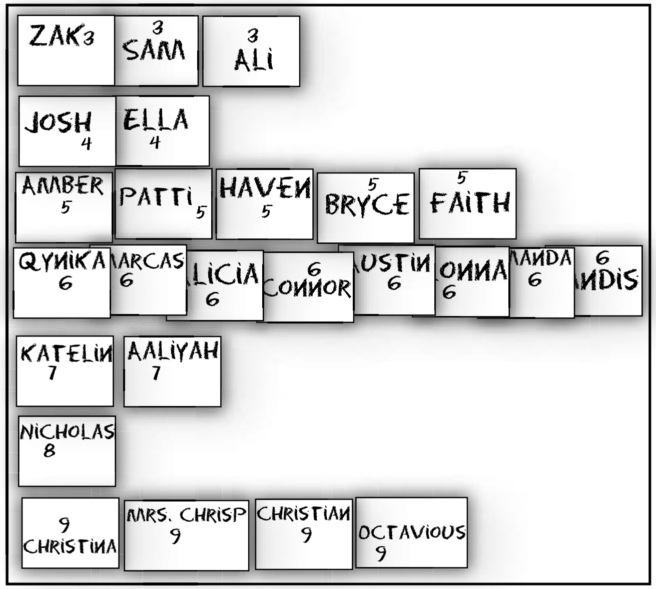


Figure 38: Preliminary dotplot

swer, "Make a graph!" The teacher then facilitates a discussion of what kind of graph they will make, and the class decides on a dotplot, given the fact that their names are already on sticky notes and given the available space on the board. Note that this display is *not* a bar graph because bar graphs are made when the data represent a categorical variable (such as favorite color). A dotplot is appropriate for a numerical variable, such as the number of letters in a name.

The teacher then uses computer software to translate this information into a more abstract dotplot, as shown in Figure 39. This helps the students focus on the general shape of the data, rather than on the particular names of the students.

Interpret Results

The teacher then facilitates a discussion of each question posed by the students, using the data displayed in the graph to answer the questions. Students also add appropriate labels and titles to the graph. The teacher helps students use the word "mode" to answer the question about the most common name length. She introduces the term "range" to help students answer the questions about shortest and longest names. Students visualize from the dotplot that there is variability in name length from individual to individual. The range gives a sense of the amount of variability in name length within the class. Using the range, we know that if the name for any two students are compared, the name lengths cannot differ by more than the value for the range.

The teacher then tells the students that there is another useful question they can answer from this data. Sometimes it is helpful to know "about how long most names are." For instance, if you were making place cards for a class lunch party, you might want to know how long the typical name is in order to decide which size of place cards to buy. The typical or average name length is called the mean. Another way to think of this is, "If all of our names were the same length, how long would they be?" To illustrate this new idea, the teach-

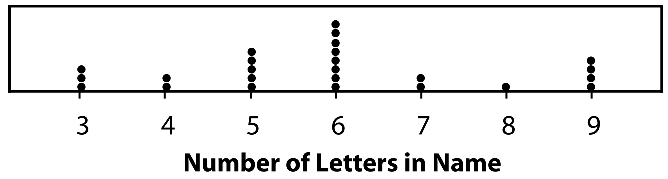


Figure 39: Computer-generated dotplot

er has students work in groups of four, and each child takes a number of snap cubes equal to the number of letters in his/her name. Then all four children at one table put all of their snap cubes in a pile in the middle of the table. They count how many cubes they have in total. Then they share the cubes fairly, with each child taking one at a time until they are all gone or there are not enough left to share. They record how many cubes each child received. (Students at some tables are able to use fractions to show that, for example, when there are two cubes left, each person could get half a cube. At other tables, the students simply leave the remaining two cubes undistributed.) The teacher then helps the students symbolize what they have done by using addition to reflect putting all the cubes in the middle of the table and using division to reflect sharing the cubes fairly among everyone at the table. They attach the words "mean" and "average" to this idea.

Finally, the students are asked to transfer the data from the sticky notes on the board to their own graphs. The class helps the teacher generate additional questions about the data that can be answered for homework. Because the students' graphs look different, the next

Figure 40: Student-drawn graphs

day the teacher will lead a discussion about the features of the various graphs the students have constructed and the pros and cons of each.

Valentine's Day and Candy Hearts

Formulate Questions

As Valentine's Day approaches, a teacher decides to plan a lesson in which children will analyze the characteristics of a bag of candy hearts. To begin the lesson, the teacher holds up a large bag of candy hearts and asks the children what they know about them from prior experience. The children know that the hearts are different colors and that they have words on them. The teacher asks the children what they wonder about the bag of hearts she is holding. The children want to know how many hearts are in the bag, what they say, and whether there are a lot of pink hearts, because most people like pink ones the best. The teacher tells the children that they will be able to answer some of those questions about their own bags of candy.

Collect Data

Each child receives a small packet of candy hearts. Students are asked how they can sort their hearts, and the students suggest sorting them by color—a categorical variable. The teacher asks students what question this will help them answer, and the students readily recognize that this will tell them which color candy appears most often in the bag.

Analyze Data

After sorting the candies into piles and counting and recording the number of candies in each pile, the teacher guides the students to make a bar graph with their candies on a blank sheet of paper. The children construct individual bar graphs by lining up all of their pink candies, all of their white candies, etc. The

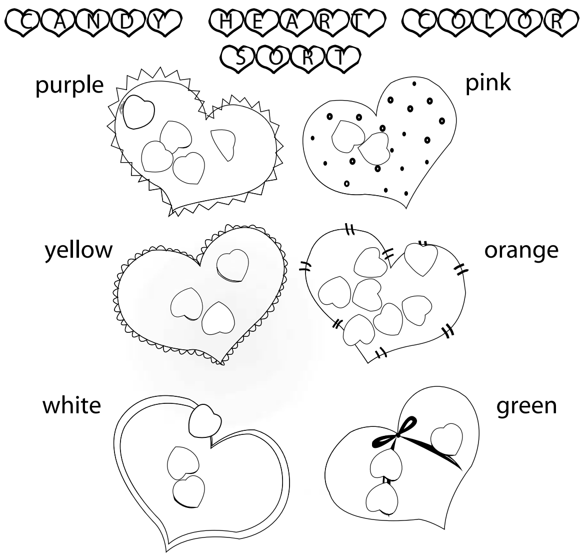


Figure 41: Initial sorting of candies

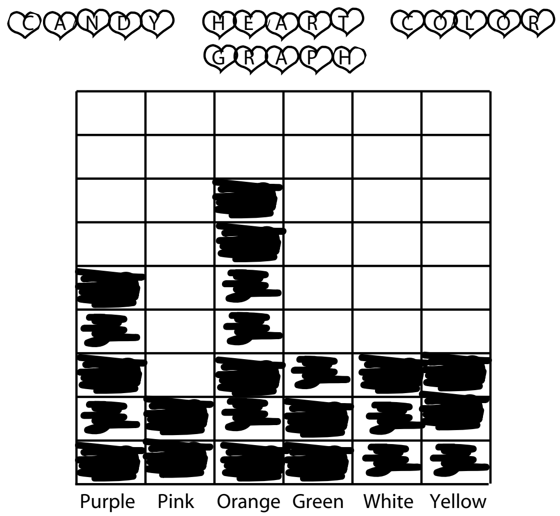


Figure 42: Bar graph of candy color

teacher then provides a grid with color labels on the *x*-axis and numerical labels on the *y*-axis so the students can transfer their data from the actual candies to a more permanent bar graph.

Interpret Results

After students construct their individual graphs, the teacher distributes a recording sheet on which each student records what color occurred the most frequently (the modal category) and how many of each color they had. This is followed by a class discussion in which the teacher highlights issues of variability. First, the students recognize that the number of each color varies within a package. Students also recognize that their packets of candy are not identical, noting that some students had no green hearts while others had no purple hearts. Some students had more pink hearts than any other color, while other students had more white hearts. At Level A, students are acknowledging variability between packages—the concept of between group variability that will be explored in more detail at Level B. The students hypothesize that these variations in packages were due to how the candies were packed by machines. The students also noted differences in the total number of candies per packet, but found this difference to be small. The student with the fewest candies had 12, while the student with the greatest number of candies had 15. The teacher asked students if they had ever read the phrase “packed by weight, not by volume” on the side of a package. The class then discussed what this meant and how it might relate to the number of candies in a bag.

(Note: Images in this example were adapted from *www.littlegiraffes.com/valentines.html*.)

Appendix for Level B

Many questionnaires ask for a "Yes" or "No" response. For example, in the Level B document, we explored connections between whether students like rap music and whether they like rock music. To investigate possible connections between these two categorical variables, the data were summarized in the following *two-way frequency table,* or *contingency table.*

Table 4: Two-Way Frequency Table

		Like Rap Music?		
		Yes	No	**Row Totals**
Like Rock Music?	Yes	27	6	33
	No	4	17	21
Column Totals		31	23	54

Since 82% (27/33) of the students who like rock music also like rap music, students who like rock music tend to like rap music as well. Because students who like rock music tend to like rap music, there is an *association* between liking rock music and liking rap music.

At Level B, we explored the association between height and arm span by examining the data in a scatterplot, and we measured the strength of the association with the Quadrant Count Ratio, or QCR. For the height/arm span problem, both variables are numerical. It also is possible to measure the strength and direction of association between certain types of categorical variables. Recall that two numerical variables are positively associated when above-average values of one variable tend to occur with above-average values of the other and when below-average values of one variable tend to occur with below-average values of the other. Two numerical variables are negatively associated when below-average values of one variable tend to occur with above-average values of the other and when above-average values of one variable tend to occur with below-average values of the other.

The scatterplot below for the height/arm span data includes a vertical line ($x = 172.8$) drawn through the mean height and a horizontal line ($y = 169.3$) drawn through the mean arm span.

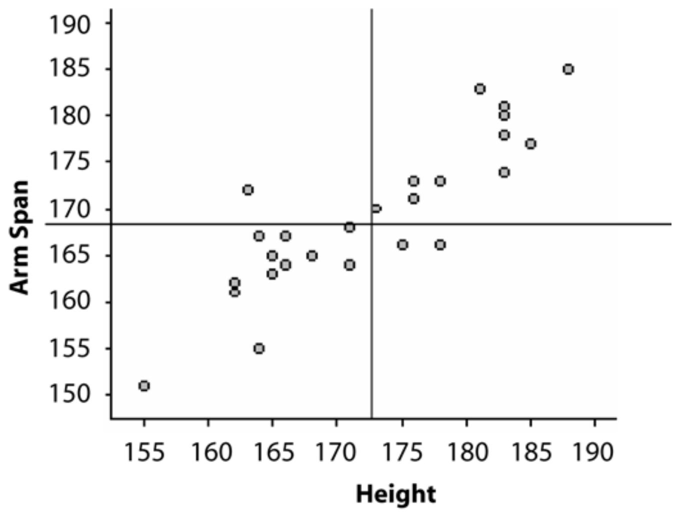


Figure 43: Scatterplot of arm span/height data

An alternative way to summarize the data would have been to ask each student the following two questions:

Is your height above average?

Is your arm span above average?

Note that for these data, the response to each question is either "Yes" or "No."

The 12 individuals in the scatterplot with below-average height and below-average arm span (Quadrant 3) responded "No" to both questions. Because their responses to both questions are the same, these 12 responses are in *agreement*. The 11 individuals in the scatterplot with above-average height and above-average arm span (Quadrant 1) responded "Yes" to both questions. Since their responses to both questions are the same, these 11 responses are in agreement. When the responses to two "Yes/No" questions are the same (No/No) or (Yes/Yes), the responses are in agreement.

The one individual with below-average height and above-average arm span (Quadrant 2) responded "No" to the first question and "Yes" to the second question, (No/Yes). Since her/his responses to the two questions are different, these two responses are in *disagreement*. The two individuals with above-average height and below-average arm span (Quadrant 4) responded "Yes" to the first question and "No" to the second question (Yes/No). Since their responses to the two questions are different, their responses are in disagreement. When the responses to two "Yes/No" questions are different (No/Yes) or (Yes/No), the responses are in disagreement.

For the data in the scatterplot in Figure 43, the results to the above two questions can be summarized in the following 2x2 two-way frequency table:

Table 19: 2x2 Two-Way Frequency Table

		Height above Average?		
		No	Yes	**Row Totals**
Arm Span above Average?	No	12	2	14
	Yes	1	11	12
Column Totals		13	13	26

Notice that there are a total of 23 responses in agreement (12 No/No and 11 Yes/Yes to the height/arm span questions), and that these correspond to the points in Quadrants 3 and 1, respectively, in the scatterplot. Also, there are a total of three responses in disagreement (two Yes/No and one No/Yes), and these correspond to the points in Quadrants 4 and 2, respectively. Recall that the QCR is determined as follows:

$$\frac{(\text{Number of Points in Quadrants 1 and 3}) - (\text{Number of Points in Quadrants 2 and 4})}{\text{Number of Points in all Four Quadrants}}$$

Restated in terms of Table 19:

$$QCR = \frac{\text{(Number of Points in Agreement)} - \text{(Number of Points in Disagreement)}}{\text{Number of Points in all Four Quadrants}}$$

Based on this, we can say that two "Yes/No" categorical variables are positively associated when the responses tend to be in agreement—the more observations in agreement, the stronger the positive association. Negative association between two "Yes/No" categorical variables occurs when the responses tend to be in disagreement—the more observations in disagreement, the stronger the negative association.

The responses to two "Yes/No" questions can be summarized as follows in a two-way frequency table:

Table 20: Two-Way Frequency Table

		Question 1		**Row Totals**
		No	Yes	
Question 2	No	a	b	r_1=a+b
	Yes	c	d	r_2=c+d
Column Totals		c_1=a+c	c_2=b+d	T= a+b+c+d

Note: a = the number who respond No/No; b = the number who respond Yes/No; c = the number who respond No/Yes; d = the number who respond Yes/Yes.

Conover (1999) suggests the following measure of association based on a 2x2 table summarized as above.

$$\frac{(a+d) - (b+c)}{T}$$

Let's call this measure the *Agreement-Disagreement Ratio* (ADR). Note that this measure of association is analogous to the QCR correlation coefficient for two numerical variables.

The ADR for the height/arm span data is:

$$ADR = \frac{(12+11) - (2+1)}{26} = .77$$

An ADR of .77 indicates a strong positive association between height and arm span measurements.

Recall the music example data, which were summarized as follows:

Table 21: Two-Way Frequency Table

		Like Rap Music?		
		No	Yes	**Row Totals**
Like Rock Music?	No	17	4	21
	Yes	6	27	33
Column Totals		23	31	54

The ADR for the rap/rock data is:

$$ADR = \frac{(17 + 27) - (4+6)}{54} = .63$$

An ADR of .63 indicates a fairly strong association between liking rock and liking rap music.

Another question presented in Level B was:

Do students who like country music tend to like or dislike rap music?

Data collected on 54 students are summarized in the following two-way frequency table:

Table 22: Two-Way Frequency Table

		Like Rap Music?		
		No	Yes	**Row Totals**
Like Country Music?	No	10	22	32
	Yes	13	9	22
Column Totals		23	31	54

For these data,

$$\text{ADR} = \frac{(10+9) - (22+13)}{54} = -.30$$

An ADR of –.30 indicates a negative association between liking country music and liking rap music.

The QCR and the ADR are additive in nature, in that they are based on "how many" data values are in each quadrant or cell. Conover (1999) suggests the *phi coefficient* as another possible measure of association for data summarized in a 2x2 table.

$$\text{Phi} = \frac{ad - bc}{\sqrt{r_1 r_2 c_1 c_2}}$$

Conover points out that Phi is analogous to Pearson's correlation coefficient for numerical data. Both Phi and Pearson's correlation coefficient are multiplicative, and Pearson's correlation coefficient is based on "how far" the points in each quadrant are from the center point.

Recall that in Example 6 of Level C, students investigated the relationship between height and forearm length. The observed data are shown again here as Table 14, and the resulting plots and regression analysis are given in Figure 35.

Regression Analysis: Height versus Forearm

The regression equation is:

Predicted Height = 45.8 + 2.76 (Forearm)

Table 14: Heights vs. Forearm Lengths

Forearm (cm)	Height (cm)	Forearm (cm)	Height (cm)
45.0	180.0	41.0	163.0
44.5	173.2	39.5	155.0
39.5	155.0	43.5	166.0
43.9	168.0	41.0	158.0
47.0	170.0	42.0	165.0
49.1	185.2	45.5	167.0
48.0	181.1	46.0	162.0
47.9	181.9	42.0	161.0
40.6	156.8	46.0	181.0
45.5	171.0	45.6	156.0
46.5	175.5	43.9	172.0
43.0	158.5	44.1	167.0

Is the slope of 2.8 "real," or simply a result of the chance variation from the random selection process? This question can be investigated using simulation.

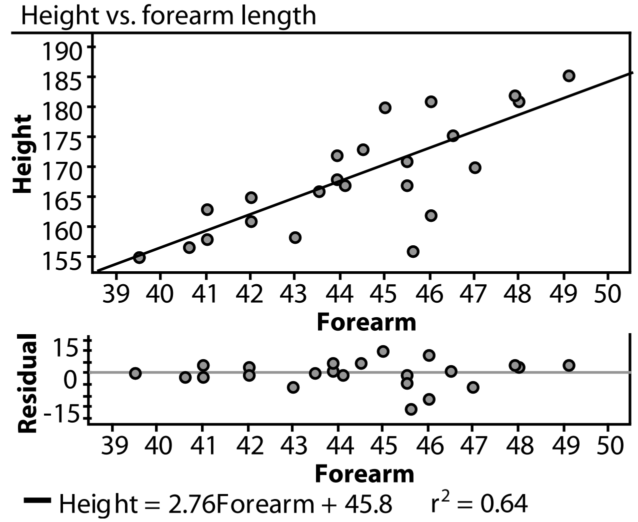


Figure 35: Scatterplot and residual plot

If there were no real relationship between height and forearm length, then any of the height values could be paired with any of the forearm values with no loss of information. In the spirit of the comparison of means in the radish experiment, you could then randomly mix up the heights (while leaving the forearm lengths as-is), calculate a new slope, and repeat this process many times to see if the observed slope could be generated simply by randomization. The results of 200 such randomizations are shown in Figure 44. A slope as large as 2.8 is never reached by randomization, which provides strong evidence that the observed slope is not due simply to chance variation. An appropriate conclusion is that there is significant evidence of a linear relationship between forearm length and height.

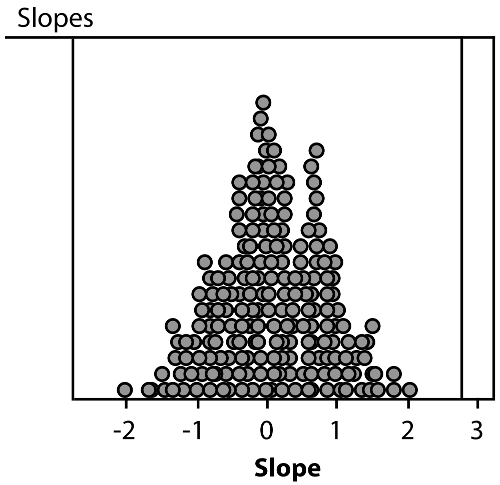


Figure 44: Dotplot showing association

Example 1: A Survey of Healthy Lifestyles

A high-school class interested in healthy lifestyles carried out a survey to investigate various questions they thought were related to that issue. A random sample of 50 students selected from those attending a high school on a particular day were asked a variety of health-related questions, including these two:

Do you think you have a healthy lifestyle?

Do you eat breakfast at least three times a week?

The data are given in Table 23.

Table 23: Result of Lifestyle Question

	Eat Breakfast		
Healthy Lifestyle	Yes	No	Total
Yes	19	15	34
No	5	11	16
Total	24	26	50

From these data, collected in a well-designed sample survey, it is possible to estimate the proportion of students in the school who think they have a healthy lifestyle and the proportion who eat breakfast at least three times a week. It also is possible to assess the degree of association between these two categorical variables.

For example, in the lifestyle survey previously described, 24 students in a random sample of 50 students attending a particular high school reported they eat breakfast at least three times per week. Based on this sample survey, it is estimated that the proportion of students at this school who eat breakfast at least three times per week is 24/50 = .48 with a margin of error of:

$$2\sqrt{\frac{(.48)(.52)}{50}} = .14$$

Using the margin of error result from above (.14), the interval of plausible values for the population proportion of students who eat breakfast at least three times a week is (0.34, 0.62). Any population proportion in this interval is consistent with the sample data in the sense that the sample result could reasonably have come from a population having this proportion of students eating breakfast.

To see if the answers to the breakfast and lifestyle questions are associated with each other, you can compare the proportions of *yes* answers to the healthy lifestyle question for those who regularly eat breakfast with those who do not, much like the comparison of means for a randomized experiment. In fact, if a 1 is recorded for each *yes* answer and a 0 for each *no* answer, the sample proportion of *yes* answers is precisely the sample mean. For the observed data, there is a total of 34 1s and 16 0s. Re-randomizing these 50 observations to the groups of size 24 and 26 (corresponding to the yes and no groups on the breakfast question) and calculating the difference in the resulting proportions gave the results in Figure 45. The observed difference in sample proportions (19/24) – (15/26) = 0.21 was matched or exceeded 13 times out of 200 times, for an estimated p-value of 0.065. This is moderately small, so there is some evidence that the difference between the two proprtions might not be a result of chance variation. In other words, the responses to the health lifestyle question and the eating breakfast question appear to be related in the sense that those who think they have a healthy lifestyle also have a tendency to eat breakfast regularly.

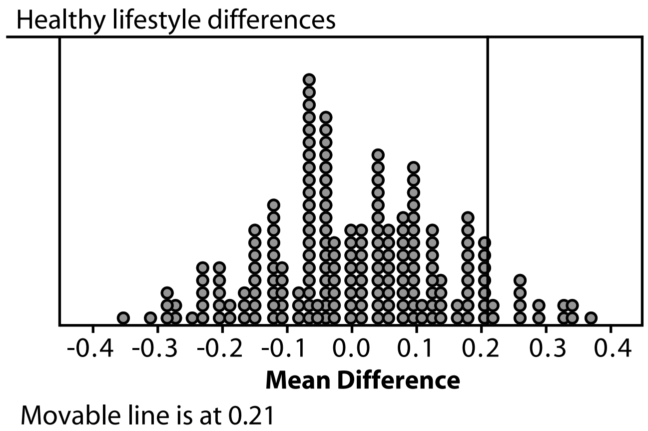


Figure 45: Dotplot showing differences in sample proportions

Example 2: An Experimental Investigation of Pulse Rates

On another health-related issue, a student decided to answer the question of whether simply standing for a few minutes increases people's pulses (heart rates) by an appreciable amount. Subjects available for the study were the 15 students in a particular class. The "sit" treatment was randomly assigned to eight of the students; the remaining seven were assigned the "stand" treatment. The measurement recorded was a pulse count for 30 seconds, which was then doubled to approximate a one-minute count. The data, arranged by treatment, are in Table 24. From these data, it is possible to either test the hypothesis that standing does not increase pulse rate, on the average, or to

Table 24: Pulse Data

	Pulse	Group	Category
1	62	1	sit
2	60	1	sit
3	72	1	sit
4	56	1	sit
5	80	1	sit
6	58	1	sit
7	60	1	sit
8	54	1	sit
9	58	2	stand
10	61	2	stand
11	60	2	stand
12	73	2	stand
13	62	2	stand
14	72	2	stand
15	82	2	stand

estimate the difference in mean pulse between those who stand and those who sit. The random assignment to treatments is intended to balance out the unmeasured and uncontrolled variables that could affect the results, such as gender and health conditions. This is called a *completely randomized design*.

However, randomly assigning 15 students to two groups may not be the best way to balance background

Table 25: Pulse Data in Matched Pairs

Pulse data: matched pairs

	MPSit	MPStand	Difference
=			
1	68	74	6
2	56	55	-1
3	60	72	12
4	62	64	2
5	56	64	8
6	60	59	-1
7	58	68	10

information that could affect results. It may be better to *block* on a variable related to pulse. Since people have different resting pulse rates, the students in the experiment were blocked by resting pulse rate by pairing the two students with the lowest resting pulse rates, then the two next lowest, and so on. One person in each pair was randomly assigned to sit and the other to stand. The matched pairs data are in Table 25. As in the completely randomized design, the mean difference between sitting and standing pulse rate can be estimated. The main advantage of the blocking is that the variation in the differences (which now form the basis of the analysis) is much less than the variation among the pulse measurements that form the basis of analysis for the completely randomized design.

In the first pulse rate experiment (Table 24), the treatments of "sit" or "stand" were randomly assigned to students. If there is no real difference in pulse rates for these two treatments, then the observed difference in means (4.1 beats per minute) is due to the randomization process itself. To check this out, the data resulting from the experiment can be re-randomized (reassigned to sit or stand after the fact) and a new difference in means recorded. Doing the re-randomization many times will generate a distribution of differences in sample means due to chance alone. Using this distribution, one can assess the likelihood of the original observed difference. Figure 46 shows the results of 200 such re-randomizations. The observed difference of 4.1 was matched or exceeded 48 times, which gives an estimated p-value of 0.24 of seeing a result of 4.1 or greater by chance alone. Because this is a fairly large p-value, it can be concluded that there is little evidence of any real difference in means pulse rates between the sitting and the standing positions based on the observed data.

In the matched pairs design, the randomization occurs within each pair—one person randomly assigned to sit while the other stands. To assess whether the observed difference could be due to chance alone and not due to treatment differences, the re-randomization must occur within the pairs. This implies that the re-randomization is merely a matter of randomly assigning a plus or minus sign to the numerical values of the observed differences. Figure 47 on the following page shows the distribution of the mean differences for 200 such re-randomizations; the observed mean difference of 5.14 was matched or exceeded eight times. Thus, the estimated probability of getting a mean difference of 5.1 or larger by chance alone is 0.04. This very small probability provides evidence that the mean difference can be attributed to something other than chance (induced by the initial randomization process) alone. A better explanation is that standing increases pulse rate, on average, over the sitting rate. The mean difference shows up as significant here, while it did not for the completely randomized design, because the matching reduced the variability. The differences in the matched pairs design have less variability than the individual measurements in the completely randomized design, making it easier to detect a difference in mean pulse for the two treatments.

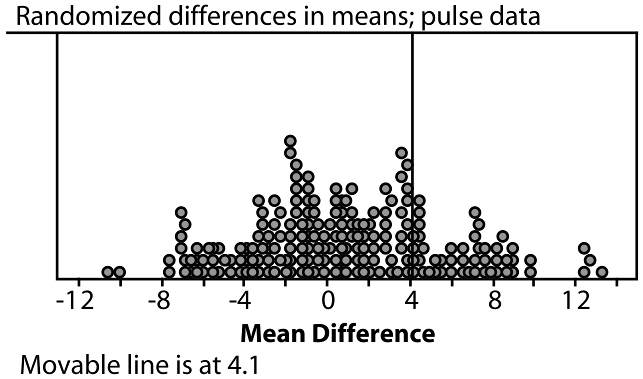
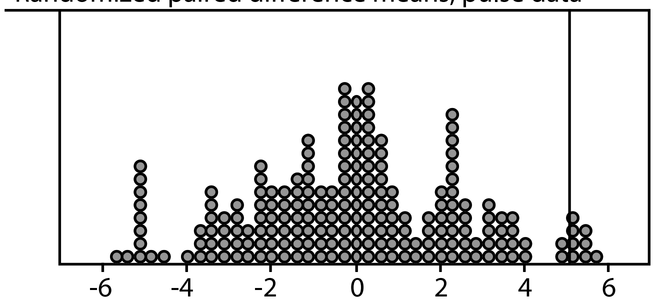


Figure 46: Dotplot of randomized differences in means

Randomized paired difference means; pulse data



Movable line is at 5.1

Figure 47: Dotplot of randomized pair difference means

Example 3: Observational Study—Rates over Time

Vital statistics are a good example of observational data that are used every day by people in various walks of life. Most of these statistics are reported as rates, so an understanding of rates is a critical skill for high-school graduates. Table 26 shows the U.S. population (in 1,000s) from 1990–2001. Table 27 shows the death rates for sections of the U.S. population over a period of 12 years. Such data recorded over time often are referred to as time series data.

Students' understanding of the rates in Table 27 can be established by posing problems such as:

→ Carefully explain the meaning of the number 1,029.1 in the lower left-hand data cell.

Table 26: U.S. Population (in 1,000s)

Year	Total Persons	Male	Female
1990	249,623	121,714	127,909
1991	252,981	123,416	129,565
1992	256,514	125,247	131,267
1993	259,919	126,971	132,948
1994	263,126	128,597	134,528
1995	266,278	130,215	136,063
1996	269,394	131,807	137,587
1997	272,647	133,474	139,173
1998	275,854	135,130	140,724
1999	279,040	136,803	142,237
2000	282,224	138,470	143,755
2001	285,318	140,076	145,242

→ Give at least two reasons why the White Male and Black Male entries do not add up to the All Races male entry.

→ Can you tell how many people died in 2001 based on Table 27 alone?

Hopefully, students will quickly realize that they cannot change from rates of death to frequencies of death without knowledge of the population sizes. Table 26 provides the population sizes overall, as well as for the male and female categories.

Noting that the population figures are in thousands but the rates are per 100,000, it takes a little thinking

Table 27: U.S. Death Rates (Deaths per 100,000 of Population)

Year	All Races		White		Black	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
1990	1202.8	750.9	1165.9	728.8	1644.5	975.1
1991	1180.5	738.2	1143.1	716.1	1626.1	963.3
1992	1158.3	725.5	1122.4	704.1	1587.8	942.5
1993	1177.3	745.9	1138.9	724.1	1632.2	969.5
1994	1155.5	738.6	1118.7	717.5	1592.8	954.6
1995	1143.9	739.4	1107.5	718.7	1585.7	955.9
1996	1115.7	733.0	1082.9	713.6	1524.2	940.3
1997	1088.1	725.6	1059.1	707.8	1458.8	922.1
1998	1069.4	724.7	1042.0	707.3	1430.5	921.6
1999	1067.0	734.0	1040.0	716.6	1432.6	933.6
2000	1053.8	731.4	1029.4	715.3	1403.5	927.6
2001	1029.1	721.8	1006.1	706.7	1375.0	912.5

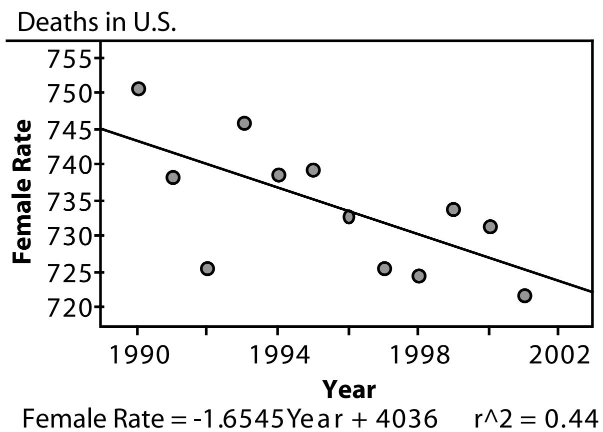


Figure 48: Scatterplot of death rates

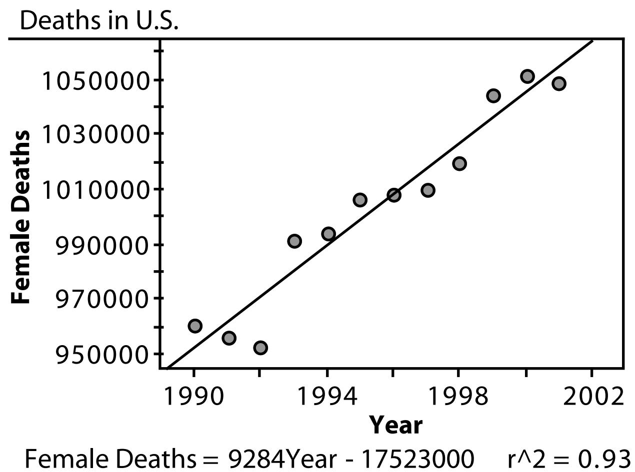


Figure 49: Scatterplot of actual deaths

on a student's part to go from rates to counts by making the computation shown in the formula:

$$\text{Female Deaths} = \text{Female Death Rate} \cdot \left(\frac{\text{Female Population}}{100} \right)$$

Some time series questions can now be explored. For example, how does the pattern of female death rates over time compare to the pattern of actual female deaths? The plots of Figures 48 and 49 provide a visual impression. The death rates are trending downward over time, with considerable variation, but the actual deaths are going up.

Students will discover that the picture for males is quite different, which can lead to interesting discussions.

Example 4: Graphs: Distortions of Reality?

Study the graph pictured in Figure 50. Do you see any weaknesses in this graphic presentation? If so, describe them and explain how they could be corrected.

Here are some plausible plots to correct errors of interpretation, and to raise other questions. Better presentations begin with a data table, such as Table 28, and then proceed to more standard graphical displays of such data.

Figure 50: Distorted graph [source: *Athens Banner-Herald*]

Table 28: Enrollment Data

Year	Total Students	African Americans
1996	29404	2003
1997	29693	1906
1998	30009	1871
1999	30912	1815
2000	31288	1856
2001	32317	1832
2002	32941	1825
2003	33878	1897
2004	33405	1845

The plot in Figure 51 shows total and African-American enrollments on the same scale. When viewed this way, one can see that the latter is a small part of the former, with little change, by comparison, over the years.

By viewing African-American enrollments by themselves, one can see that the marked decrease between 1996 and 2002 may be turning around—or leveling off.

However, the ratio of African American to total enrollment is still on the decrease!

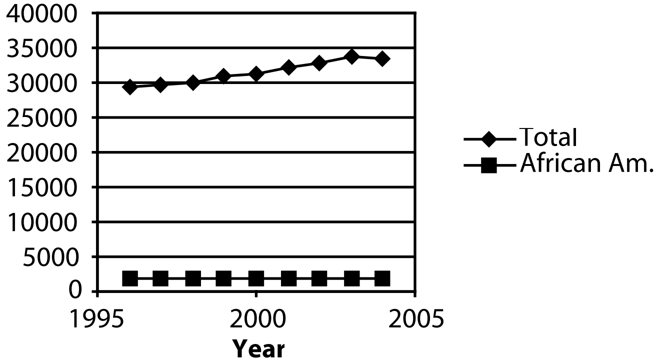


Figure 51: Plot of African-American vs. total enrollments

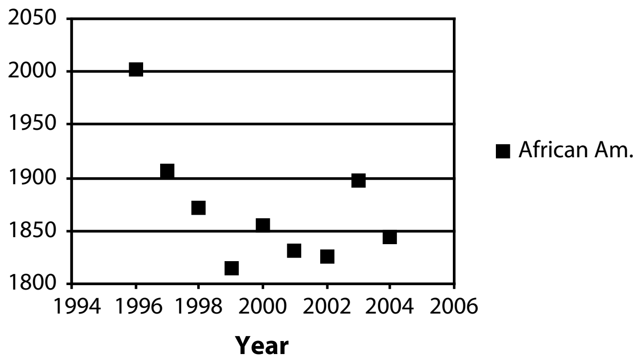


Figure 52: Plot of African-American enrollments only

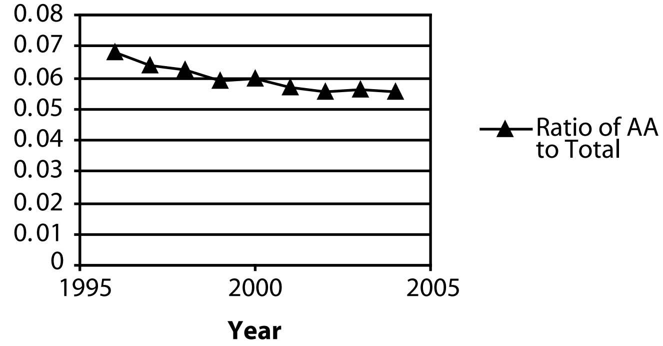


Figure 53: Ratio of African-American to total enrollments