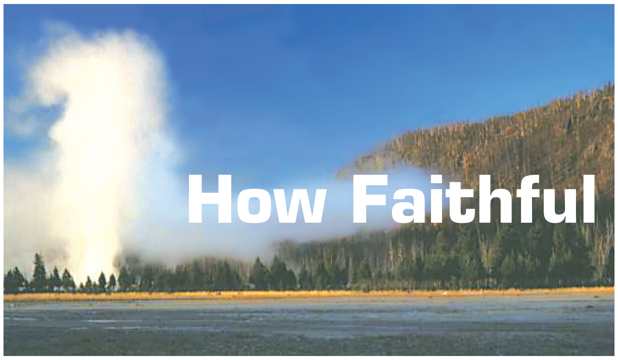
J. Michael Shaughnessy and Maxine Pfannkuch



How Faithful Is Old Faithful?

Statistical Thinking: A Story of Variation and Prediction

Statistics is a relatively new discipline. Only in the last one hundred years have common methods and common reasoning evolved that can be applied to data from many fields. In the early years, the field of statistics was influenced by the work of Ronald A. Fisher, Karl Pearson, and Jerz Neyman. They focused on developing tools and methods that primarily focused on randomization More recently, exploratory data analysis has been emphasized (Tukey 1977). As statistics continues to mature as a discipline, statistics educators are paying more attention to developing overall models of statistical thinking (Wild and Pfannkuch 1999). This shift in statistics means refocusing the emphasis in teaching from how to do statistics to how to think about statistics. In this next step in the evolution of statistics and statistics teaching, two questions arise: What is statistical thinking? and How can we develop students' statistical thinking?

The authors of this article have found that data sets from the Old Faithful geyser in Yellowstone Park furnish a rich context for introducing such important aspects of statistical thinking as the central role of variation and the importance of asking our students what they would predict. In this article, we first discuss the context of the data, next present a classroom exploration of the data, and then discuss the nature of statistical thinking as it pertains to this Old Faithful data set.

This shift in statistics means refocusing the emphasis from how to do statistics to how to think about statistics

THE CONTEXT

We imagine that we have just arrived at Yellowstone National Park, the home of geyser basins, thermal mud pots, hot springs, acid lakes, and a multitude of fascinating animals and plants. Furthermore, as has actually happened to one of the authors, we have just missed the most recent eruption of Old Faithful Geyser, which has periodically been spewing streams of hot water high into the sky for centuries. How long would we have to wait until the next eruption of Old Faithful? Before reading any further, write down your best estimate.

Readers who have visited Yellowstone Park might have some basis for making an informed estimate of the wait time until the next eruption. However, someone who is unfamiliar with geysers or who has not been to Yellowstone might not have any basis for estimating the wait time. Some geysers are dormant between eruptions for many hours or days. Other geysers erupt almost continuously. How could we obtain better information?

One strategy might be to appeal to a higher authority, by asking a park ranger or by reading the sign that indicates the approximate time of the next eruption. However, a strategy that can help those of us who are not actually at the park is to make a prediction on the basis of past data on Old Faithful. This latter strategy opens the door for an adventure in statistical thinking. We encourage readers to first work through our investigation of the Old Faithful data so that they can experience it in the manner in which we have used it with our own students. We then return to a deeper discussion of the aspects of statistical thinking that can arise while exploring this data set.

THE INVESTIGATION—EXPLORING DATA ON OLD FAITHFUL

Data on wait times between eruptions of Old Faithful are available through such sources as Hand et al. (1994). The following is approximately a day's worth of data on wait times for Old Faithful. Old Faithful erupts approximately twenty times each day. The data show the numbers of minutes between the time when Old Faithful stopped erupting to when it first began to erupt again.

Mike Shaughnessy, mike@mth.pdx.edu, teaches mathematics to preservice and in-service teachers at Portland State University, Portland, OR 97207. His primary teaching and research interests are in probability and statistics, geometry, and problem solving. Maxine Pfannkuch, pfannkuc@pop.math.auckland.ac.nz, is a lecturer in the mathematics department at the University of Auckland, Auckland 1, New Zealand. She works with undergraduate and graduate mathematics and statistics students. One of her primary teaching and research interests is the nature of statistical thinking.

Copyright © 2002 The National Council of Teachers of Mathematics, Inc. www.nctm.org. All rights reserved.
This material may not be copied or distributed electronically or in any other format without written permission from NCTM.

First day (minutes between eruptions):

51	82	58	81	49	92	50	88	62
93	56	89	51	79	58	82	52	88

If we had some friends who were planning to visit Yellowstone National Park, how long should we tell them to expect to wait between eruptions of Old Faithful? Before reading further, readers should make an estimate and give some justification for the prediction. They can also construct a graph of the first day's data on Old Faithful's eruptions.

Of course, one day's worth of data does not give much basis for a prediction. Two more days of Old Faithful eruption data, picked at random from a larger data set, follow:

Day 2:

86	78	71	77	76	94	75	50	83
82	72	77	75	65	79	72	78	77

Day 3:

65	89	49	88	51	78	85	65	75
77	69	92	68	87	61	81	55	93

Readers should construct graphical representations of the second and third days' data, similar to the representation of the first day's data. Compare the data for the three days. At this point, what could we predict for our friends? How long should they expect to wait for the next eruption of Old Faithful?

The three days for which data are given are only a small part of a data set, given in the **appendix,** for the wait times for 300 consecutive Old Faithful eruptions (Hand et al. 1994). One strategy that we have used is to give each student a strip showing a day's worth of data; have them analyze, graph, and make predictions from it; then have students trade several times with other students; and repeat this process with data for several other days.

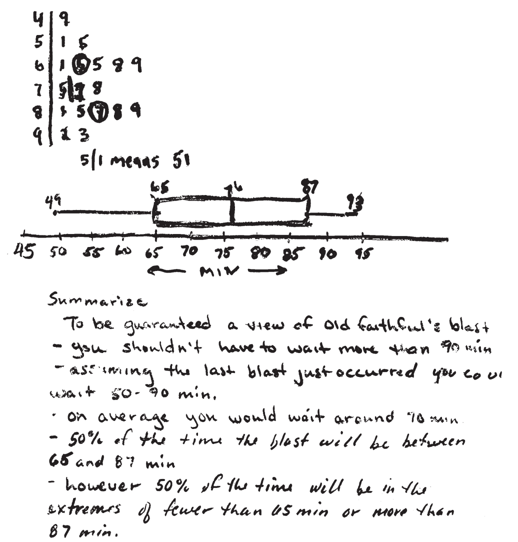


Fig. 1
Student's stem-and-leaf plot and box plot of one day of Old Faithful data

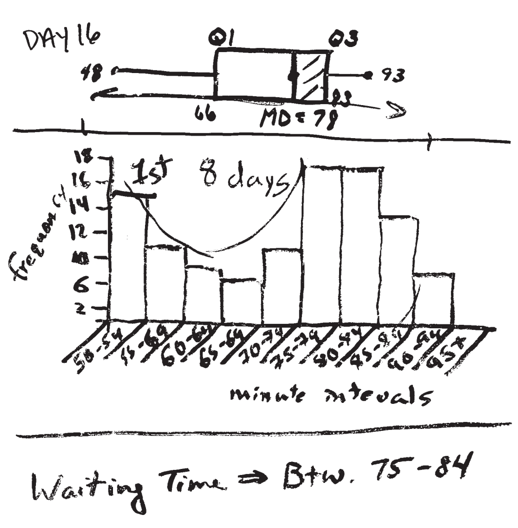


Fig. 2
Student's box plot of one day and histogram of eight days of Old Faithful data

One day's worth of data does not give much basis for a prediction

TYPICAL STUDENT RESPONSES

Many students first just calculate a mean or determine a median for a day's worth of Old Faithful data and base their initial prediction on a measure of central tendency. The mean of the first day's data is 70.1, and the median is 70.5; the mean of the second day's data is 79.9, and the median is 77; and so forth. Although the mean does furnish a one-number summary of a data set, it can also mask important features in the distribution of the data.

When students begin to create their own graphs, a variety of features of the distribution appear in their graphical representations. **Figures 1** through **4** are examples of students' work depicting data for one or more days of Old Faithful's eruptions. As shown in **figure 1,** some students make stem-and-leaf plots or box plots. Others make histograms, as in **figure 2,** perhaps accompanied by a box plot. Still others create dot plots or bar plots, in which the bar's height represents the length of the wait time between eruptions, as shown in **figure 3,** or plots of the length of the wait versus the number of the wait, as shown in **figure 4.**

Each of these types of representations can highlight or mask particular patterns in the data. Box plots furnish a good visual representation of the range and of the middle 50 percent of the data, as well as allow comparisons of the box size and position for several days. However, stem-and-leaf plots and histograms yield a clearer picture of the data's distribution. Box plots involve data reduction to

We have concentrated heavily on measures of central tendency, and we have neglected variation

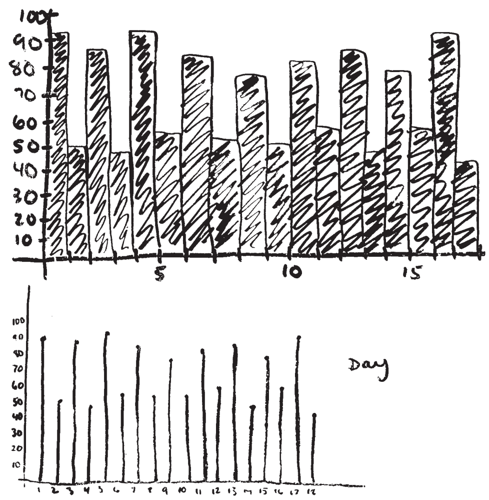


Fig. 3
Student's bar graph and dot plot of wait times for eighteen consecutive eruptions of Old Faithful

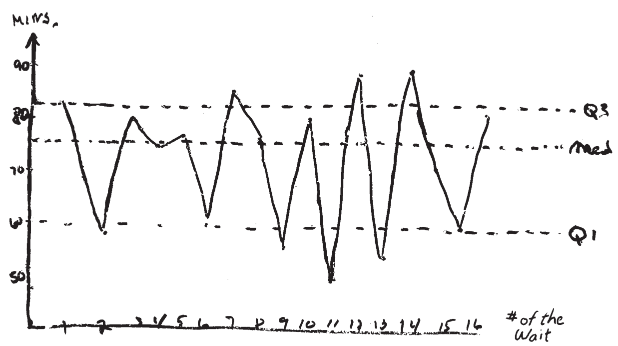


Fig. 4
Student's connected plot of sixteen consecutive wait times for Old Faithful eruptions

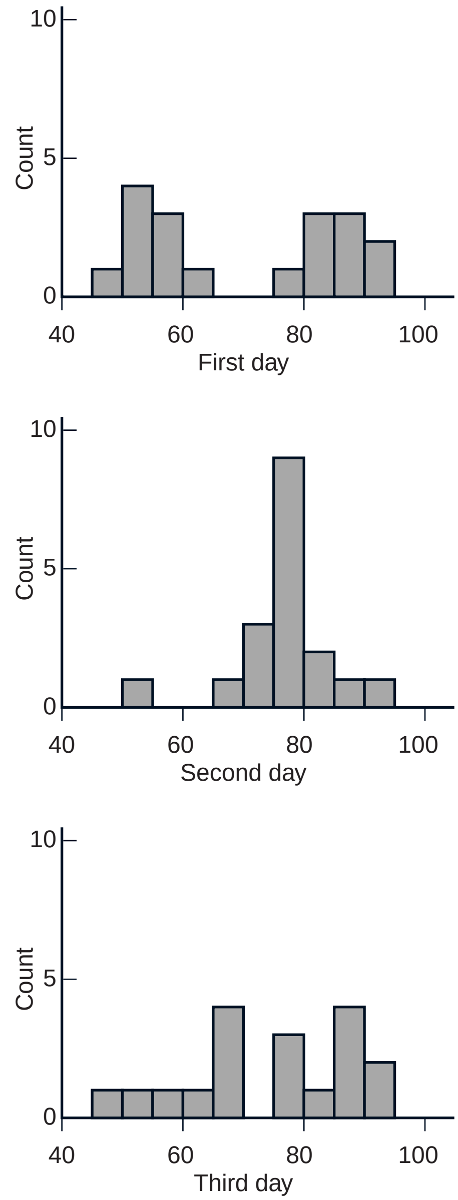


Fig. 5
Minutes between eruptions of Old Faithful; histograms for three different days

summarize the data, whereas stem-and-leaf plots and histograms display the actual data. Stem-and-leaf plots or histograms can reveal gaps that are masked in a box plot. Data for the first and third days for Old Faithful appear somewhat bimodal in a histogram, whereas the second day's data are more moundlike, as indicated in **figure 5.**

Even more telling, an alternating short-long pattern in Old Faithful's eruptions is visually highlighted by students who create plots of consecutive wait times, dot plots, or bar graphs (**figs. 4** and **5**). This oscillating pattern can be completely missed by students who just calculate a mean or draw a box plot for the data.

Our past teaching, our textbooks, and many state and national assessments have concentrated heavily on measures of central tendency (mean, median, and mode), and we have neglected variation (Shaughnessy et al. 1999). However, variation is the essential signature in the Old Faithful data. Variation exists among days and within each day, and patterns in the variation can go completely unnoticed if we, or our students, concentrate only on centers and neglect variation. Students who pay attention to the variability in the data are much more likely to predict a range of outcomes or an interval for the wait time for Old Faithful. Such students make predictions similar to "Most of the time you'll wait between fifty and ninety minutes" rather than a single value of seventy minutes, as shown in the student work in **figure 1.**

The real power in exploring the Old Faithful data set arises when we ask students to share their graphical representations and predictions with one another. Students share a wide variety of graphical representations, and some students express surprise at some of their fellow students' graphs. We have even heard applause for some students who put up a plot of consecutive wait times after students have shared many box plots, stem-and-leaf plots, and histograms. The plot of consecutive wait times is a powerful visual characterization of the alternating pattern in the Old Faithful data. Students who create plots of consecutive wait times often ask for more information. They want to know the length of the previous wait so that they can more accurately predict the next wait interval. They begin to make conjectures about the reasons that the data for Old Faithful alternate. They begin to ask what aspect of the geyser system causes this pattern of variation. They are beginning to show statistical thinking.

WHAT IS STATISTICAL THINKING?

The question "What is statistical thinking?" has provoked considerable debate. However, the central element of any definition of statistical thinking is an understanding of variation. According to Moore (1990, p. 135) the core element of statistical thinking is variation: "the omnipresence of variation in processes . . . the design of data production with variation in mind . . . the explanation of variation." Moore believes that students in the future will have a structure of thought that whispers "variation matters." The quality-management field believes that statistical thinking has three key principles: all work occurs in a system of interconnected processes, variation exists in all processes, and understanding and reducing variation are keys to success. Mallows (1998) believes that any definition of statistical thinking that does not include the relevance of the data to the problem is inadequate. Wild and Pfannkuch (1999) believe that statistical thinking is a complex activity, and they have identified five elements that are fundamental to statistical thinking in empirical inquiry in any field:

- Recognition of the need for data
- Transnumeration
- Consideration of variation
- Reasoning with statistical models
- Integrating the statistical and contextual

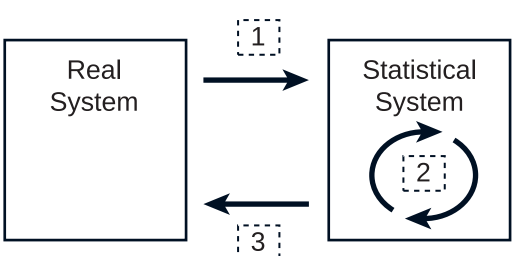


Fig. 6
The process of transnumeration

We are back at the second question posed at the beginning of this article. If these five elements are at the core of statistical thinking, then how can we develop statistical thinking in our students?

OLD FAITHFUL AND DEVELOPING STUDENTS' STATISTICAL THINKING—A DEEPER LOOK

Understanding variation is central to statistical thinking. This Old Faithful activity seeks to promote variation as the "big idea" to which students' attention should be drawn. An aspect that should be considered in developing statistical thinking is that the reasoning processes are fundamentally different from those of mathematics, since statistical thinking deals with uncertain, empirical data. The student is placed in the role of a data detective. Students must look for patterns, deal with the variation, and make judgments and predictions on the basis of the data. We next discuss the five elements of statistical thinking and further illustrate the use of the Old Faithful data in promoting statistical thinking.

Recognition of the need for data

The foundations of statistical inquiry rest on the assumption that many real situations cannot be judged without gathering and analyzing properly collected data. Anecdotal evidence or one's own experience may be unreliable and misleading for judgments and for decision making. Therefore, data are considered a prime requirement for judgments about real situations. Our initial situation in the Old Faithful activity emphasizes the need for data, since students' predictions of the wait time for the next eruption are probably not within their own experience. A teaching focus should be on the need for data, because much research suggests that students think that their own judgments and beliefs are more reliable than data. In addition, students do not see any purpose in analyzing data, since they already know the "answer." This activity is one in which they are unlikely to know the answer and must therefore look at data to make a judgment.

In statistical thinking, the reasoning processes are fundamentally different from those of mathematics

Transnumeration

Transnumeration is a coined word, meaning "numeracy transformation for facilitating understanding." Transnumeration occurs in three specific phases in a statistical problem and can be viewed from a modeling perspective. A diagram is given in **figure 6.** Transnumeration is a dynamic process of changing representations to engender understanding. If we consider the real system and the statistical system, then transnumeration-type thinking occurs through—

- capturing measures of the real system that are relevant,
- constructing multiple statistical representations of the real system, and
- communicating to others what the statistical system suggests about the real system.

In our Old Faithful activity, some relevant measures have already been "captured," namely, the students receive wait-time data on Old Faithful. The second phase of transnumerative thinking starts with a strip of one day's data, when students need to consider ways to change the data representation to facilitate a prediction. It also occurs when students are asked to draw a graph but are not told what type of graph to draw. Then, when they are asked to share their graphs with the whole class, they soon recognize that different representations convey different types of information about the geyser. The students' dot plots, stem-and-leaf plots, box plots, histograms with varying class intervals, and plots of consecutive wait times all reveal different messages about Old Faithful. Histograms with large class intervals can actually obscure information, and the students must therefore recognize this problem and try several different class-interval widths. To expedite matters, students may want to use Sturges's guideline, which suggests that the ideal number of class intervals is about $1 + \log_2 n$, where n is the number of data values. This sharing of graphs promotes transnumerative thinking and the need to look at multiple representations to detect messages in the data. The third phase of transnumeration begins to occur when the students are asked to communicate their predictions of wait time.

Sharing graphs promotes trans-numerative thinking and the need to look at multiple representations to detect messages in the data

Consideration of variation

Making a judgment from data requires an understanding of variation during the process of statistical inquiry. To make an informed prediction, we first must notice that variation exists, either directly from the data or from the graphs of the data. When students first look at one day's data and then see that a classmate's data for a day look quite different, they begin to see variability from day to day, as well as within a single day. The variation that occurs in the data encourages students to request more data to improve their prediction.

In schools, the emphasis has historically been on descriptive statistics, especially on measures of center, and variation has been neglected. Most students initially report that the mean time of about seventy minutes is the length of time that they would expect to wait before the next eruption. Although it could be argued that this response is appropriate, focusing on one number as the solution addresses neither the variability in the system nor the pattern in the variability. As soon as we ask the question *why,* for example, "Why do patterns appear in the Old Faithful data? we enter the realm of what we call *analytical statistics*. Analytical statistics attempts to find explanations, seeks causes, makes predictions, and looks behind the data. Information about the *why* questions can be sought in patterns in variation in the data. Since the variation in the Old Faithful data is not random, underlying geological causes or relationships are likely; reasons exist for the variation. Thus, when the students share their graphs, we must encourage them to look at the graphs through a "variation lens" and so must encourage them to search for patterns in variation.

Reasoning with statistical models

According to the research of Konold and others (1997), when dealing with data, students have difficulty making the transition from thinking about and comparing individual cases to thinking about and comparing group propensities. Konold and his colleagues point out that reasoning with statistical models requires the ability to carry out both aggregate-based and individual-based reasoning and to recognize the power and limitations of such reasoning across a variety of situations. This aggregate-based reasoning, coupled with recognizing the patterns in the data set, is fundamental to statistical thinking.

When students are asked to draw a graph, they need to see that different days can produce different or similar patterns, but overall they need to see that the pattern or group propensity in Old Faithful is bimodal rather than unimodal. Teaching should focus on the patterns in distributions and patterns in the centers and spreads, not on the individual pieces of data for Old Faithful. Asking students whether knowing one wait time between eruptions would be sufficient for making a prediction may make them aware of the inadequacies of looking at individual examples.

After students have shared their graphs, we have found that they often decide that plots of consecutive wait times are the most appropriate models for depicting the distribution of the Old Faithful data, since such plots emphasize the oscillating character of the data. Thus, part of reasoning with statistical models involves selecting or creating a model that optimally represents and communicates the nature of the real problem and that focuses our reasoning about the data.

Integrating the statistical and contextual

The integration of statistical knowledge and contextual knowledge is a fundamental element of statistical thinking. The statistical model must capture elements of the real situation, and the resultant data carry their own literature base (Cobb and Moore 1997), that is, the data tell a story. Information about the real situation is contained in the statistical summaries, and a synthesis of statistical and contextual knowledge must therefore occur to draw out what can be learned in the context sphere.

At the beginning of the Old Faithful classroom activity, we discuss the context and general behavior of geysers to enable students to understand the meaning of the data. Such contextual knowledge is essential for seeing and interpreting any messages contained in the data. Students play the role of detectives who are looking for patterns to form their predictions; a continuous dialogue should exist between the data and context, as indicated in **figure 7.** For example, the pattern in this particular data set is bimodal and oscillating. Why do the wait times between eruptions oscillate about two mean times? What is the source of this variation? How does this geyser system work?

Another example occurs in a scatterplot of eruption wait times plotted against previous eruption wait time. This plot appears nonlinear, as shown in **figure 8.** This graph was generated using data from the first eight days. The sample correlation coefficient is –0.727. With the high influence point (108, 50) removed, the sample correlation coefficient is –0.722. The evidence of nonlinearity is weak. When a quadratic is fitted, with the high influence point removed, the *P*-value is 0.08. The nature of this relationship furnishes an opportunity for further investigation in an advanced class. Why do we find more variability in the wait time until the next eruption when the wait time since the previous eruption was a long one? Should other factors about Old Faithful be considered? What is the story contained in the data? Are other geyser systems the same?

What is the story contained in the data?

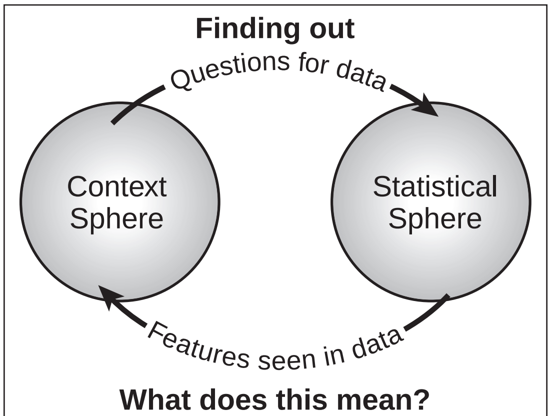


Fig. 7
The integration of statistical and contextual knowledge

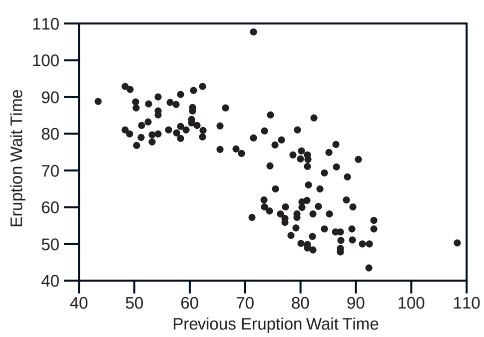


Fig. 8
Comparing adjacent eruption wait times

ADDITIONAL QUESTIONS TO PROMOTE STUDENTS' STATISTICAL THINKING

Many opportunities occur throughout the Old Faithful investigation to promote further statistical thinking with students. Many of these opportunities can be tapped by asking well-placed questions that prompt students to think about and discuss ideas. A few of those opportunities follow:

Before students are given the data

How much data would you need for a prediction? One wait time? Two wait times? One day's worth of data? Two days' worth of data? A year's worth of data?

When they are given the data

How does your prediction using data compare with your first prediction without the data?

After they draw a graph of the first day's worth of data

What patterns, if any, do you notice?

After they draw graphs of two more days' data

How do your predictions compare with your previous ones?

When they share graphs

Compare and contrast the information revealed by each graph. What is gained or lost with the various graphical representations? How are the graphs related to one another?

When they are ready to interpret the information

What other data would be helpful, for example, duration time of eruptions, to enable you to further understand the wait time between blasts? What does the oscillating pattern tell you about how this geyser works? What other information on this geyser system would help you interpret this pattern?

When they are ready to make a final prediction

Will this prediction be true for the whole year? Will seasonal variation occur? Will this prediction be true over several years? Will yearly variation occur? What, if any, limitations should you put on your prediction? Is your prediction of oscillation valid? In one day, for example, the alternating pattern may be long, long, short, long, long, short, short, long, long, short, and so on. What graphical representation could you use to verify whether the alternating pattern of short and long wait times is generally true?

When they are ready to communicate the information to others

What graphical representation would best communicate your prediction? What other information, besides your prediction, should be communicated?

At the end of the inquiry

What have you learned about making predictions? About variation? How well do your predicted times compare with the actual wait times? If you continued with this problem, what would you investigate next? Can you find an explanation for the pattern? Does a relationship exist between duration times of eruptions and wait times between eruptions? Eruption-duration data for Old Faithful can be found in Foreman and Bennett (1999).

Additional project

Data are available on eruptions of other geysers and volcanoes, such as Kilauea on Hawaii from 1750 to the present (see www.jason.org and hvo.wr.usgs.gov/kilauea/history/historytable). Gather information on wait times between eruptions and duration of eruptions for Kilauea or some other volcanoes or geysers. Compare the information with that for Old Faithful, and determine similarities and differences in the patterns of variability.

SO HOW FAITHFUL IS OLD FAITHFUL?

The answer to our original question depends on our statistical thinking. If we measure Old Faithful's wait times for "faithfulness" one eruption at a time, we might conclude that Old Faithful is not faithful at all. We might wait 49 minutes, or we might wait 102 minutes, 58 minutes, or 89 minutes. This pattern does not seem very faithful. However, perhaps wait time is not a good measure of "faithfulness." When we consider the overall pattern in the distribution of Old Faithful's wait times, we find that it is bimodal and oscillating. Using the pattern in the variation, we might be able to predict a time that is close to the wait time for the next eruption. Old Faithful is really very "faithful" to that overall pattern in the distribution of data. We hope that read-

APPENDIX

Old Faithful Daily Data Sets—Minutes between Eruptions

1)	86	71	57	80	75	77	60	86	77	56	81	50	89	54	90	73	60	83
2)	65	82	84	54	85	58	79	57	88	68	76	78	74	85	75	65	76	58
3)	91	50	87	48	93	54	86	53	78	52	83	60	87	49	80	60	92	43
4)	89	60	84	69	74	71	108	50	77	57	80	61	82	48	81	73	62	79
5)	54	80	73	81	62	81	71	79	81	74	59	81	66	87	53	80	50	87
6)	51	82	58	81	49	92	50	88	62	93	56	89	51	79	58	82	52	88
7)	52	78	69	75	77	53	80	55	87	53	85	61	93	54	76	80	81	59
8)	86	78	71	77	76	94	75	50	83	82	72	77	75	65	79	72	78	77
9)	79	75	78	64	80	49	88	54	85	51	96	50	80	78	81	72	75	78
10)	87	69	55	83	49	82	57	84	57	84	73	78	57	79	57	90	62	87
11)	78	52	98	48	78	79	65	84	50	83	60	80	50	88	50	84	74	76
12)	65	89	49	88	51	78	85	65	75	77	69	92	68	87	61	81	55	93
13)	53	84	70	73	93	50	87	77	74	72	82	74	80	49	91	53	86	49
14)	79	89	87	76	59	80	89	45	93	72	71	54	79	74	65	78	57	87
15)	72	84	47	84	57	87	68	86	75	73	53	82	93	77	54	96	48	89
16)	63	84	76	62	83	50	85	78	78	81	78	76	74	81	66	84	48	93

ers find this investigation as helpful for introducing and promoting statistical thinking with their students as we have.

REFERENCES

Cobb, George, and David Moore. "Mathematics, Statistics, and Teaching." *The American Mathematical Monthly* 104 (November 1997): 801–23.

Foreman, Linda, and Al Bennett. *Math Alive.* Course III. Salem, Ore.: The Math Learning Center, 1999.

Hand D. J., F. Daly, A. D. Lunn, Kevin J. McConway, and E. Ostrowski, eds. *Handbook of Small Data Sets.* London: Chapman & Hall, 1994.

"Jason Project." www.jason.org. World Wide Web.

Konold, Cliff, Alexander Pollatsek, Andrew Well, and Allen Gagnon. "Students Analyzing Data: Research of Critical Barriers." In *Research on the Role of Technology in Teaching and Learning Statistics,* Proceedings of the 1996 International Association of Statistics Education Round Table Conference, edited by Joan Garfield and Gail Burrill, pp. 151–67. Voorburg, The Netherlands: International Statistics Institute, 1997.

Mallows, Colin. "The Zeroth Problem." *The American Statistician* 52 (February 1998): 1–9.

Moore, David. "Uncertainty." In *On the Shoulders of Giants,* edited by Lynn Arthur Steen, pp. 95–137. Washington, D.C.: National Academy Press, 1990.

Shaughnessy, J. Michael, Jane Watson, Jonathon Moritz, and Christine Reading. "School Mathematics Students' Acknowledgment of Statistical Variation." Paper presented at the research presessions of the 77th annual meeting of the National Council of Teachers of Mathematics, San Francisco, April 1999.

Tukey, John. *Exploratory Data Analysis.* Reading, Mass.: Addison Wesley Longman, 1977.

U.S. Geological Survey, Hawaiian Volcano Observatory. "Summary of Historical Eruptions, 1750–Present." hvo.wr.usgs.gov/kilauea/history/historytable.html. World Wide Web.

Wild, Chris J., and Maxine Pfannkuch. "Statistical Thinking in Empirical Enquiry. *International Statistical Review* 67 (1999): 223–65.

Yellowstone National Park: National Park Service. www.nps.gov/yell/. World Wide Web.