DOCUMENT RESUME

ED 328 451 SE 051 931

AUTHOR	Russell, Susan Jo; Corwin, Rebecca B.
TITLE	Statistics: The Shape of the Data. Used Numbers: Real Data in the Classroom. Grades 4-6.
INSTITUTION	Lesley Coll., Cambridge, Mass.
SPONS AGENCY	National Science Foundation, Washington, D.C.
REPORT NO	ISBN-0-86651-514-3
PUB DATE	89
CONTRACT	MDR-8651649
NOTE	152p.; For related documents, see SE 051 929-930. In cooperation with the Technical Education Research Centers (TERC) and the Consortium for Mathematics and Its Applications (COMAP).
AVAILABLE FROM	Addison-Wesley Publishing Co., The Alternative Publishing Group, P.O. Box 10888, Palo Alto, CA 94303-0879 ($12.95).
PUB TYPE	Guides - Classroom Use - Guides (For Teachers) (052)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Calculators; *Classification; Computer Assisted Instruction; Cooperative Learning; *Data Analysis; Elementary School Mathematics; *Graphs; Interdisciplinary Approach; *Intermediate Grades; *Learning Activities; Mathematics Education; Problem Solving; *Statistics; Student Projects; Teaching Guides; Teaching Methods

ABSTRACT

A unit of study that introduces collecting, representing, describing, and interpreting data is presented. Suitable for students in grades 4 through 6, it provides a foundation for further work in statistics and data analysis. The investigations may extend from one to four class sessions and are grouped into three parts: "Introduction to Data Analysis"; "Learning About Landmarks in the Data"; and "A Project in Data Analysis." An overview of the investigation, session activities, dialogue boxes, and teacher notes are included in each investigation. The major goals developed in each part of this guide are: (1) describing the shape of the data; (2) defining the way data will be collected; (3) summarizing what is typical of the data; (4) making quick sketches of the data; (5) inventing ways to compare two sets of data; (6) representing data first through sketches, then through a presentation graph or chart; (7) using the median as a landmark in the data; (8) understanding that the median is only one landmark in the data; and (9) experiencing all the stages of a data analysis investigation. Attached are 10 student sheets. (KR)


U N
DATA
DALE SEYMOUR PUBLICATIONS
2
3

Used Numbers: Real Data in the Classroom is a project of Technical Education Research Centers (TERC), Lesley College, and the Consortium for Mathematics and Its Applications (COMAP).

The *Used Numbers* materials focus on the processes of data analysis—collecting, organizing, representing, and interpreting data. Through collecting and analyzing real data, students develop and use ideas and tools from key areas of mathematics—counting, measuring, sorting and classification, estimation, graphing, computation in context, and statistics—and they are introduced to appropriate uses of computers and calculators for data analysis.

The primary goal of the *Used Numbers* project is to help teachers and students become intelligent and critical users of data: to understand the limitations and the power of using numbers to reveal patterns and trends, to compare and predict, and to make informed decisions about complex issues.

Statistics: The Shape of the Data

A unit of study for grades 4–6
from USED NUMBERS: REAL DATA IN THE CLASSROOM

Developed at Technical Education Research Centers and Lesley College

Susan Jo Russell and Rebecca B. Corwin

DALE SEYMOUR PUBLICATIONS

The *Used Numbers* materials were prepared with the support of National Science Foundation Grant No. MDR-8651649. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily represent the views of the National Science Foundation. These materials shall be subject to a royalty-free, irrevocable, worldwide, nonexclusive license in the United States Government to reproduce, perform, translate, and otherwise use and to authorize others to use such materials for Government purposes.

Cover design and illustrations: Rachel Gage

Order number DS01026
ISBN 0-86651-514-3

DALE
SEYMOUR
PUBLICATIONS
P.O. BOX 10888
PALO ALTO, CA 94303

abcdefghi-MA-9543210

USED NUMBERS STAFF

Co-principal investigators

Susan Jo Russell
Technical Education Research Centers (TERC)

Susan N. Friel
Lesley College

Curriculum development

Rebecca B. Corwin (TERC and Lesley College)
Tim Barclay (TERC)
Antonia Stone (Playing to Win)

Research and evaluation

Janice R. Mokros (TERC)
Alana Parkes (TERC)
Debra Gustafson (TERC)
John Olive (University of Georgia)
Deborah Ruff (University of Georgia)
Heide Wiegel (University of Georgia)
Bonnie Brownstein (Institute for Schools of the Future)
Ellen Bialo (Institute for Schools of the Future)
Michele Arsenault (Institute for Schools of the Future)
Mary Fullmer (University of Chicago)

Design and production

Elisabeth Roberts (TERC)
Jessica Goldberg (TERC)
LeAnn Davis (TERC)
John Abbe (TERC)
Laurie Aragon (COMAP)

Cooperating classrooms for this unit

Fern Fisher and Francine Hiller
Cambridge Public Schools, Massachusetts

Dorothy Spahr
Wellesley Public Schools, Massachusetts

Reina Huerta
New York City Public Schools

Christine Fuentes
Clarke County Public Schools, Georgia

Advisory board

Joan Akers, California State Department of Education
Bonnie Brownstein, Institute for Schools of the Future
James Landwehr, AT&T Bell Laboratories
Steven Leinwand, Connecticut State Department of Education
John Olive, University of Georgia
David Pillemer, Wellesley College
Andee Rubin, Bolt Beranek and Newman Laboratories
Cindy Stephens, D. C. Heath
Marion Walter, University of Oregon
Virginia Wooley, Boston Museum of Science

Thanks also to advice and comment from Marilyn Burns, Solomon A. Garfunkel (COMAP), and Bob Willcutt.

CONTENTS

PREFACE

In an information-rich society such as ours, statistics are an increasingly important aspect of daily life. We are constantly bombarded with information about everything around us. This wealth of data can become confusing, or it can help us make choices about our actions.

Educators and mathematicians now stress the importance of incorporating data analysis and statistics into the elementary mathematics curriculum to prepare students for living and working in a world filled with information based on data. The *Curriculum and Evaluation Standards for School Mathematics*, published by the National Council of Teachers of Mathematics in 1989, highlights statistics as one of the key content strands for all grade levels.

Many teachers see the need to support students in becoming better problem solvers in mathematics. However, it is difficult to find problems that give students the kind of experiences they need, are manageable in the classroom, and lead to the learning of essential mathematics. The area of data analysis—collecting, organizing, graphing, and interpreting data—provides a feasible, engaging context in which elementary grade students can do real mathematics. Students of all ages are interested in real data about themselves and the world around them.

Teaching statistics: Pedagogical issues

We introduce students to good literature in their early years. We do not reserve great literature until they are older—on the contrary, we encourage them to read it or we read it to them. Similarly, we can give young students experience with real mathematical processes rather than save the good mathematics for later.

Through collecting and analyzing real data, students encounter the uncertainty and intrigue of real mathematics. Mathematicians do not sit at desks doing isolated problems. Instead, they discuss, debate, and argue—building theories and collecting data to support them, working cooperatively (and sometimes competitively) to refine and develop such theories further. Mathematicians and scientists use information or data like snapshots to look at, describe, and better understand the world. They cope with the real-world "messiness" of the data they encounter, which often do not lead to a single, clear answer.

Because statistics is an application of real mathematics skills, it provides the opportunity to model real mathematical behaviors. As students engage in the study of statistics, they, like scientists and statisticians, participate in:

- cooperative learning
- theory building
- discussing and defining terms and procedures
- working with messy data
- dealing with uncertainty

We want elementary school students to have the opportunity to engage in such real mathematical behavior, discussing, describing, challenging each other, and building theories about real-world phenomena based on their work.

Data analysis in the mathematics curriculum

Exploring data involves students directly in many aspects of mathematics. Data are collected through counting and measuring; they are sorted and classified; they are represented through graphs, pictures, tables, and charts. In summarizing and comparing data, students calculate, estimate, and choose appropriate units. In the primary grades, work with data is closely tied to the number relationships and measuring processes that students are learning. In the upper elementary grades, students encounter some of the approaches used in statistics for describing data and making inferences. Throughout the data analyis process, students make decisions about how to count and measure, what degree of accuracy is appropriate, and how much information is enough; they continually make connections between the numbers and what those numbers represent.

Instead of doing mathematics as an isolated set of skills unrelated to the world of reality, students can understand statistics as the vibrant study of the world in which they live, where numbers can tell them many different stories about aspects of their own lives. The computation they do is for a purpose, and the analysis they do helps them to understand how mathematics can function as a significant tool in describing, comparing, predicting, and making decisions. ■

TEACHING DATA ANALYSIS

The nature of data analysis

In data analysis, students use numbers to describe, compare, predict, and make decisions. When they analyze data, they search for patterns and attempt to understand what those patterns tell them about the phenomena the data represent.

A data analysis investigation generally includes recognizable phases:

- considering the problem
- collecting and recording data
- representing the data
- describing and interpreting the data
- developing hypotheses and theories based on the data

These phases often occur in a cycle: the development of a theory based on the data often leads to a new question, which may begin the data analysis cycle all over again.

Elementary students can collect, represent, and interpret real data. Although their work differs in many ways from that of adult statisticians, their processes are very similar. Elementary school students can both analyze data and use those data to describe and make decisions about real situations.

Because real data are the basis for investigations in data analysis, there are no predetermined "answers." For example, if your class collects data on the ages of the students' siblings, the students understand that their job is more than simply coming up with an answer that you knew all along. Not only do you *not* know the answer in advance, but, without seeing the data, you may not even know what the most interesting questions are going to be!

While this situation encourages students to do their own mathematical thinking, it can also feel risky for you. Many teachers welcome a little uncertainty in their mathematics classes, when it prods their students to be more independent thinkers. To support you, the authors provide sample experiences from teachers who have used the activities described here so that you can be prepared for the kinds of issues that are likely to arise. You will soon build your own repertoire of experiences with data analysis activities and will be able to anticipate likely questions, confusions, and opportunities.

The importance of discussion in mathematics

A central activity in data analysis is dialogue and discussion. While it is easy for you and your students to become engaged and enthusiastic in collecting data and making graphs, a significant amount of time should also be devoted to reflection about the meaning of the data.

Since students are not used to talking much during their mathematics work, it is important to support active decisionmaking by the students from the very beginning of the investigation. Students' participation in framing the initial question, choosing the methods of investigation, and deciding on ways to organize their data is essential. Once the data are collected and organized, the students must grapple with interpreting the results. If you have the outcome of a discussion or the "teaching points" you want to make too clearly in mind, you may guide students' observations too quickly into predetermined channels. When student ideas are ignored, misinterpreted, or rejected, they soon understand that their job is to second-guess the "answer" you had in mind.

On the other hand, if students find that *anything* they say is accepted in the same way, if every contribution is "a good idea" and no idea is ever challenged, they can lose motivation to participate. Ask students to reflect on, clarify, and extend their ideas and to listen to and ask questions of each other. Discussions in mathematics should encourage students to interpret numbers, make conjectures, develop theories, consider opposing views, and support their ideas with reasons.

Sensitive issues in data analysis

Students of all ages are interested in data about themselves and the issues they care about. Topics that matter enough to students to make them compelling topics for study often have very personal aspects. Investigations about families, heights, or students' chores, for example, can all bring up sensitive issues. After trying many topics in many classrooms, we have concluded that the potential sensitivity of a topic is not a reason to avoid it; on the contrary, these are the very topics that most engage student interest. All teachers deal with difficult or sensitive issues in their classroom, and the skills demanded of a teacher in handling issues that arise during data analysis activities are no different. Keep in mind that students may sometimes want their data to be anonymous. Focusing on the patterns and shape of the class data, rather than on individual pieces of data, is particularly helpful, especially for upper elementary students.

Small-group work

Many of the investigations involve students working in teams. At first, keep small-group sessions short and focused. For students not used to working in small groups, assign specific tasks that encourage the participation of all the group members. For example, instead of, "Have a discussion in your group to decide what you want to ask the second graders about their bedtimes," you might say, "Come up with three possible questions you could ask the second graders."

Materials

Students need materials to represent their data during their investigations. These range from Unifix cubes to pencil and paper to computer software. What is most important is that students are able to construct multiple views of the data quickly and easily and that they do not become bogged down in drawing and coloring elaborate graphs (which are appropriate only at the very end of an investigation when students are ready to "publish" their findings).

Any material that can be moved easily and rearranged quickly offers possibilities for looking at data. For example, students might write or draw their data on *index cards* (or any paper rectangles); then these can be arranged and rearranged. *Unifix cubes* (or other interconnecting cubes) are another good material for making representations throughout the grades. We have found that *stick-on notes* (such as Post-it notes), with each note representing one piece of data, are an excellent material for making rough drafts of graphs. They can be moved around easily and adhere to tables, desks, paper, or the chalkboard. *Pencil and unlined paper* should always be available for tallies, line plots, and other quick sketch graphs.

Calculators

Calculators should be available, if possible, throughout the activities. Their use is specifically suggested in some of the investigations. It is no secret to students that calculators are readily available in the world and that adults use them often. But many students do not know how to use a calculator accurately, do not check their results for reasonableness, and do not make sensible choices about when to use a calculator. Only through using calculators with appropriate guidance in the context of real problems can they gain these skills.

Computers

Computers are a key tool in data analysis in the world outside of school. Graphing software, for example, enables scientists and statisticians to display large sets of data quickly and to construct multiple views of the data easily. Some software for the elementary grades allows this flexibility as well. A finished graph made by the computer may, for some students, be an appropriate illustration for a final report of their findings. But keep in mind that students also make interesting and creative graphs by hand that would not be possible with the software available to them. Other computer software, including software for sorting and classifying and data base software, is particularly useful for some data analysis investigations. Where the use of a software tool would particularly enhance a data analysis investigation, recommendations for incorporating its use are made in the text.

Home-school connections

Many opportunities arise in data analysis investigations for communicating with parents about the work going on in the classroom and for including them as participants in your data investigations. When you begin this unit, you may want to send a note home to parents explaining that students will be studying data analyis in their mathematics class and that, from time to time, parents can be of assistance in helping students collect data from home. Parents or other family members often provide an available comparison group. Studies of age, family size, height, and so forth can be extended to include parents. If students are studying their own families, they may be interested in collecting comparison data about their parents' families. Including parents and other significant family members as participants in your data analysis investigations can stimulate their interest and enthusiasm for the work students are doing in school and, at the same time, help students see that the mathematics they do in school is connected to their life outside of school.

Interdisciplinary connections

Many teachers find ways to connect the data analysis experiences students have in mathematics to other areas of the curriculum. Data analysis is, after all, a tool for investigating phenomena of all kinds. The same approaches that students use in this unit can be called on for an investigation in science or social studies. Making these connections explicit and helping students transfer what they have learned here to new areas will give them an appreciation of the usefulness of mathematics throughout the curriculum. ■

STATISTICS: THE SHAPE OF THE DATA
UNIT OVERVIEW

Statistics: The Shape of the Data is a unit of study that introduces collecting, representing, describing, and interpreting data. Suitable for students in grades 4 through 6, it provides a foundation for further work in statistics and data analysis, including the other two upper-grade units in the *Used Numbers* series. *Statistics: The Shape of the Data* involves students in

- collecting real data through experimentation, observation, and surveys
- representing data, using a variety of materials to construct models, graphs, tables, and diagrams
- describing landmarks and features of the data
- formulating hypotheses and building theories about the reality represented by the data

How to use this unit

Like all the *Used Numbers* units, *Statistics: The Shape of the Data* is organized into investigations that may extend from one to four class sessions. To cover the entire unit requires approximately 17 class sessions of about 45 minutes each. Teachers who have used this unit have found that a schedule of 2–3 sessions per week works best to maintain continuity while allowing enough time for reflection and consolidation between sessions. The activities are sequenced so that students move gradually from more straightforward to more complex investigations. The investigations are grouped into three parts:

- **Part 1: Introduction to data analysis**
  How many raisins are in a box?
  How many people are in a family?
  How long can we hold our breath?
- **Part 2: Learning about landmarks in the data**
  How much taller is a fourth (fifth, sixth) grader than a first grader?
  Looking at Mystery Data
  Finding landmarks in the data
- **Part 3: A project in data analysis**
  Investigating sleep

The three parts work well as a single five-to-six-week unit. Some teachers have substituted this unit for their textbook chapters on statistics or graphing. Others have used it late in the year as a way to consolidate students' mathematical learning, knowing that it brings together work in graphing, measurement, estimation, computation, and statistics in a problem-solving context. The parts can also be spaced over the entire school year. For example, some teachers use Part 1 in September to start off their work in mathematics. They return to Part 2 in January and use Part 3 in May when students have been together for most of a school year and are more able to work independently. Within each part, it is important that 2–3 sessions take place each week so that the experiences build on each other, allowing students gradually to acquire skills and understanding in data analysis.

Planning the investigations

This guide gives you all the information you'll need to guide students through the investigations in *Statistics: The Shape of the Data*. The information is organized as follows:

Investigation overview. This section includes (1) a summary of the student activity, (2) materials you will need for the investigation and any special arrangements you may need to make, and (3) a list of the important mathematical ideas you will be emphasizing. Plan to look carefully at this overview a day or two before launching the investigation.

Session activities. For each session, you will find step-by-step suggestions that outline the students' explorations and the teacher's role. Although suggestions for questions and instructions are given, you will of course modify what you say to reflect your own style and the needs of your students. In all cases, the teacher's words are intended to be guidelines, *not* word-for-word scripts. Plan to read through this section before each session to establish in your mind the general flow of the activities.

Dialogue Boxes. The Dialogue Boxes illustrate the special role of discussion in these investigations and convey the nature of typical student-teacher interactions. Examples are drawn from the actual experiences of classes that have used these investigations. They call attention to issues that are likely to arise, typical student confusions and difficulties, and ways in which you can guide and support students in their mathematical thinking. Plan to read the relevant Dialogue Boxes before each session to help prepare for interactions with your students.

Teacher Notes. These sections provide important information you will need about various mathematical concepts and data sets. Here you will find explanations of key aspects of the process of collecting and analyzing data, including ways to graph data and how and when to introduce basic mathematical terms. The Teacher Notes are listed in the contents because many are useful as references throughout the unit, not just where they first appear. You might plan to read them all for background information before starting the unit, then review them as needed when they come up in particular investigations.

Goals for students

The "Important mathematical ideas" listed in the overviews highlight the particular student goals for each investigation. The major goals for *Statistics: The Shape of the Data* are as follows:

Part 1: Introduction to data analysis

Describing the shape of the data. Students gradually move from noticing individual features of the data ("Two boxes had 33 raisins, three boxes had 34 raisins") to describing the overall shape of the data distribution ("Over half of the boxes had between 34 and 37 raisins").

Defining the way data will be collected. Students make decisions about how to count or measure—a key aspect of data analysis. They begin to see that such decisions can have a profound effect on the eventual outcome of their statistical work, even before they collect any data.

Summarizing what is typical of the data. In these first investigations, students do not use formal measures such as the *median* or *mean*, but they do develop their own informal ways of summarizing the key features of a data set.

Making quick sketches of the data. Students use a variety of graphs for displaying data during the process of analysis. They build up a repertoire of useful "first draft" graphs that include line plots, stem-and-leaf plots, tables, and tallies.

Part 2: Learning about landmarks in the data

Inventing ways to compare two sets of data. Students use the ideas they have developed about describing the shape of the data and summarizing what is typical of a particular set of data as they find ways to compare two data sets.

Representing data first through sketches, then through a presentation graph or chart. Students experience the progression from a "rough draft" stage, through revision and refinement, to presentation of their findings in a more finished form.

Using the median as a landmark in the data. Students discover the usefulness of the median in describing a set of data and in comparing one data set with another. They develop an understanding of what the median describes; that is, the median is the value that divides the data set in half.

Understanding that the median is only one landmark in the data. Students explore the limitations of the median and find that while it gives us some information about the data set, it does not reveal all the important features of the data.

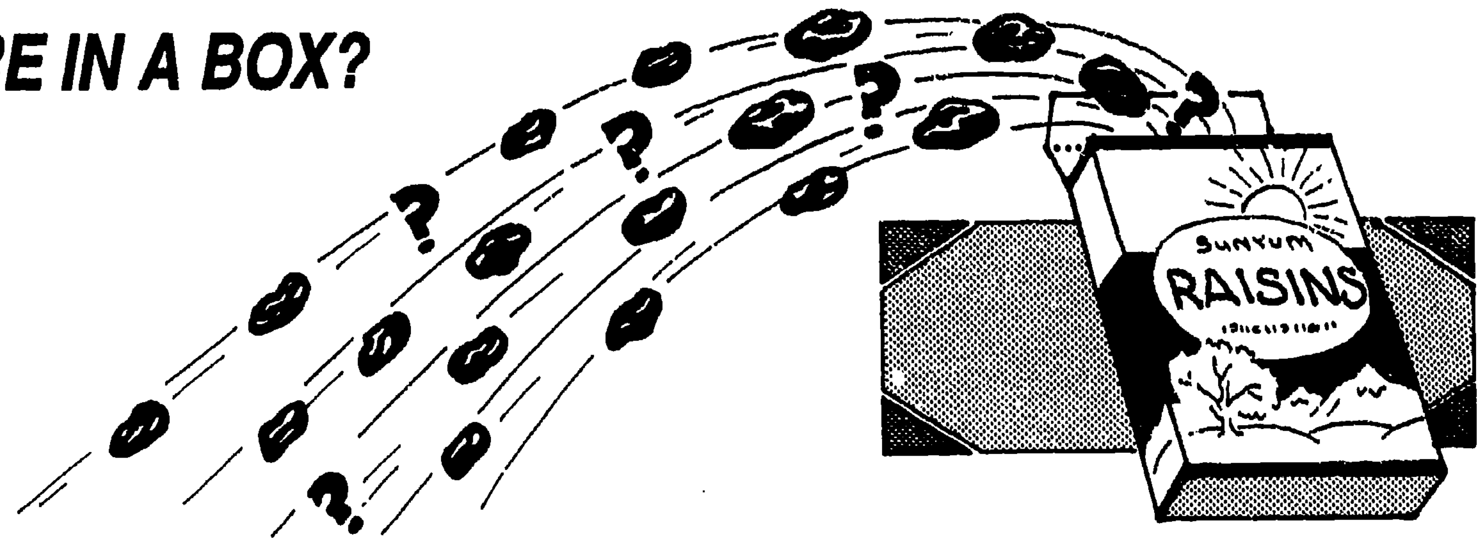
Part 3: A project in data analysis

Experiencing all the stages of a data analysis investigation. Like the writing process, data analysis includes the stages of brainstorming, first draft, revision, subsequent drafts, and publication. Students experience all these in their concluding project, which includes choosing and refining a research question. ■

Statistics: The Shape of the Data

PART 1
Introduction to data analysis

HOW MANY RAISINS ARE IN A BOX?



INVESTIGATION OVERVIEW

What happens

Students count the number of raisins in a sample of small boxes of raisins (one box for each student), record and organize the results, and describe the shape of the data distribution. The line plot is introduced as a useful type of "sketch graph"—a rough, first draft visual representation of the data.

The activities take one class session of about 45 minutes.

What to plan ahead of time

- Provide small boxes of raisins (the half-ounce size), at least one for each student. (Alternatively, use other small packages of easily countable things that are packed by weight, such as peanuts.)
- Provide unlined paper for making sketch graphs.
- Become familiar with making a line plot. See the Teacher Note, *Line plot: A quick way to show the shape of the data* (page 17).

☛ Save the empty raisin boxes to use as a data set for a later investigation, *Finding landmarks in the data.*

Important mathematical ideas

Making quick sketches of the data. Graphs are more than a clear and tidy presentation of final results; they are working tools, used to represent data during the process of analysis. Support your students' inventions of ways to display data quickly and clearly, as discussed in the Teacher Note, *Sketch graphs: Quick to make, easy to read.* Encourage them to build up a repertoire of useful "first draft" graphs such as line plots, tables, and tallies. In this session you will demonstrate one useful sketch graph, a line plot, as described in the Teacher Note, *Line plot: A quick way to show the shape of the data.*

Describing the shape of the data. Describing data distributions is the focus of this entire unit. Help students gradually move from noticing individual features of the data ("Two boxes had 33 raisins, three boxes had 34 raisins") to describing the overall shape of the distribution ("Over half of the boxes had between 34 and 37 raisins"). This process is explained further in the Teacher Note, *The shape of the data: Clumps, bumps, and holes.* ■

SESSION ACTIVITIES

Introducing the unit

This year, as part of our mathematics work, we will be studying statistics. Have you ever heard the word *statistics*? Can you give me any examples of statistics?

Students may know about the use of statistics in sports or in opinion polls.

Statistics is the study of data. Data are numbers that give us information about something in the real world. We can collect some data right now. How many people in this room have a pet (have brown eyes, speak Spanish, once lived in a different country, take the bus to school)?

Ask several of these questions, count the student responses, and point out that these are data (e.g., "Our data show that 12 students in this class take the bus to school").

People collect data by counting, like we just did, or by measuring or by doing experiments. Who can think of some data we can collect by measuring?

Students may think of examples such as the size of the classroom, their own heights, or the distance from home to school. They may also think of measures involving weight, volume, time, or temperature.

After mathematicians or scientists have collected their data, they study the data carefully and look for patterns that could tell them something important. For example, studying data about traffic accidents might provide information about which kinds of cars are the safest or whether seat belts make a difference. Data about the size of fish in certain lakes and rivers could give clues about the effects of water pollution.

If possible, provide an example of the use of statistics in your school or community. For example, in one school, a particular piece of playground equipment had been forbidden to students younger than fourth grade. The principal used data about injuries on this piece of equipment to make his decision about who could use it. Many of the younger children, it turned out, did not have hands big enough to grasp the bars securely.

Just like mathematicians and scientists who use statistics, we can collect data to find out new things about ourselves or other things around us. Today, we are going to start by collecting data about something familiar—a box of raisins.

Considering the problem: Estimating the number of raisins

Give a box of raisins to each student. Ask them to keep the boxes closed.

Does anybody have an idea about how many raisins there are in a box this size?

Let students offer their ideas. Then have them open their boxes so they can see the top layer of raisins.

What do you think now? Do you want to revise your estimate?

Students may have a variety of ideas. Allow enough time for them to talk about these. Possible follow-up questions:

Why do you think there will be about [50]? Your idea is very different from Jan's; how are you thinking about your estimate? Will the number of raisins in each box be the same or different? Why do you think so?

Collecting data: Counting raisins

Students open their boxes and count the raisins.

Recording and organizing the data: Sketch graphs

As students finish their counts, they report their data. Record the numbers in a list on the chalkboard, in whatever order they are reported.

If we wanted to organize these data better, what could we do?

Take a few suggestions from the students. Then have them work in pairs or groups of three, each group choosing one way to organize the data quickly. Emphasize that this is a rough draft sketch; it need not be done meticulously. Once they have organized the data, each group should write down three important things they can say about their data.

Ask a few students to demonstrate their methods for organizing the data or quickly demonstrate them yourself on the board. Make sure that all the different types of representation they have invented are demonstrated. See the Teacher Note, *Sketch graphs: Quick to make, easy to read* (page 17).

Mathematicians have invented ways of displaying data, too. Here's one way that's easy to use. It's called a *line plot*.

Organize the raisin data on a line plot large enough for everyone to see. See the Teacher Note, *Line plot: A quick way to show the shape of the data* (page 17).

Describing the data: What's the shape of these data?

What are some of the things you decided you could say about these data?

Help students express their initial ideas. See the Teacher Note, *The shape of the data: Clumps, bumps, and holes* (page 18). Follow up with questions such as:

What else can you say about these data? Does anyone have another way to describe this representation? Suppose someone asked you, "About how many raisins are in a box?" What could you say?

See the Dialogue Box, *Describing the shape of the data* (page 16), for a sample discussion.

Developing theories: Making predictions

If we opened five more boxes of raisins, what is your best guess about how many raisins would be in them, based on the data we already have?

Students work on this question for a few minutes in small groups, then report their theories back to the whole class. Encourage students to ask each other questions. Expect them to give reasons for their ideas.

Mary's group's theory seems very different from Joe's group's theory. Do some of you have thoughts about that?

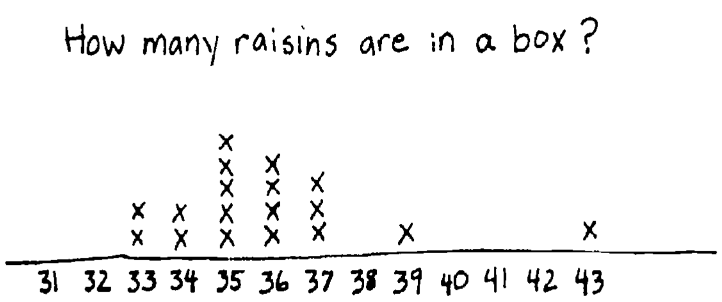
☛ At the end of the session, allow students to eat the raisins. Then have each student write *inside the top flap* the number of raisins found in that box. Collect and save the boxes for use as a data set in the investigation *Finding landmarks in the data*.

Extensions

Students are very interested in what happens if they add more boxes of raisins to their data. If you have extra boxes of raisins, students might count them and add their findings to the class data. Or, keep the data posted and add to them periodically. You could use the raisin boxes themselves to make a more permanent display. Do these additional data change the shape of the data distribution in any way?

Try a similar activity with some other prepackaged material—peanuts, sunflower seeds, fresh peapods. What is the shape of the data this time? Is it harder or easier to predict how much will be in a new package? ■

DIALOGUE BOX
Describing the shape of the data



So what can you say about the raisin data? Let's hear a few of your ideas.

MARIA: Well, there are a lot at 35.

ALAN: There was only one at 39 and one at 43.

JANE: There are two at 33 and 34.

What else did you notice?

KAREN: Thirty-three is the lowest.

So no boxes had fewer than 33 raisins?

DAVID: Yeah. And 43 was the highest.

So the range was from 33 to 43. What else?

ANNIE: There's nothing at 38, 40, 41, or 42.

Annie's noticing that there are a lot of holes in this part of the data. Can anyone say any more about that?

SUE: Well, there's nothing at 31 or 32 either.

Yes, 33 is the lowest count and there's nothing below it. But this situation, that Annie noticed up here, is a little different. What can you say about that?

JESSIE: Mostly, the raisins go from 33 to 37, but sometimes you get something higher.

Can anyone add to that?

BEN: You'd be really lucky if you were the one who got 43!

In fact, mathematicians have a name for a piece of data that is far away from all the rest. They call it an *outlier*. An outlier is an unusual piece of data—sometimes it might actually be an error, but sometimes it's just an unusual piece of data. It's usually interesting to try to find out more about an outlier. Who had the outlier in this case?

CHRIS: I did. And I counted twice, and Kyle checked it, too, so I know it was 43.

JANE: Maybe he's got smaller raisins.

Any other theories about Chris's box?

MARY: Maybe it doesn't really weigh the same as the other boxes. Maybe too many raisins got dropped in when it was going through the factory.

[Later] . . .

So if someone asked you, "What's the typical number of raisins in a box?"—what would you say?

KIM: Well, I'd say 35.

[Addressing the class as a whole, not just the student who answered]: **Why would 35 be a reasonable description of how many raisins are in a box?**

ALICE: Because the most boxes had 35.

Any other ways to say this? Or any different ideas?

ARRIE: Well, I wouldn't say just 35.

Why not?

ARRIE: Well, there's really not that much difference between 33, 34, 35, 36. They're all really close together. I'd say 33 to 37, 'cause the 39 and 43 aren't what you'd usually get.

So Arrie is saying he'd use an *interval* to describe the raisins, from 33 to 37, and Kim said she'd say 35 was typical. What do other people think about that?

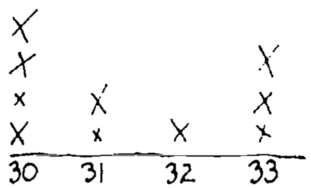
In this discussion, the class has moved gradually from describing individual features of the data to looking at the shape of the data as a whole. The teacher introduced the ideas of *interval*, *range*, and *outlier* because they came up in the discussion and were appropriate in describing these data. [For explanations of *range* and *outlier*, see the Teacher Note, *Statistical words to introduce as appropriate*, page 32.] Throughout the conversation, the teacher tries to have students give reasons for their ideas and pushes them to think further by asking for additions or alternatives to ideas students have raised. ■

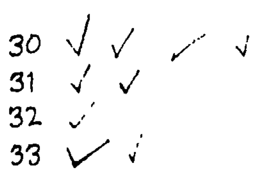
TEACHER NOTE

Sketch graphs: Quick to make, easy to read

Graphing is often taught as an art of presentation, as the endpoint of the data analysis process, as the means for communicating what has been found. Certainly, a pictorial representation is an effective way to present data to an audience at the end of an investigation. But graphs, tables, diagrams, and charts are also data analysis tools. A user of statistics employs pictures and graphs frequently during the process of analysis as a means of better understanding the data.

Many working graphs need never be shown to anyone else or posted on the wall. Students can make and use them just to help uncover the story of the data. We call such representations used during the process of data analysis "sketch graphs" or "rough draft graphs."





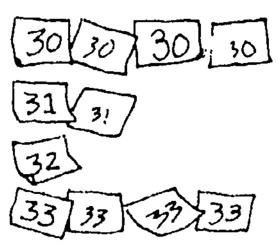
We want students to become comfortable with a variety of such working graphs. Sketch graphs should be easy to make and easy to read; they should not challenge students' patience or fine motor skills. Unlike graphs for presentation, sketch graphs do not require neatness, careful measurement or scaling, use of clear titles or labels, or decorative work.

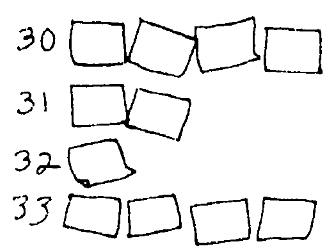
Sketch graphs:

- can be made rapidly
- reveal aspects of the shape of the data
- are clear, but not necessarily neat
- don't require labels or titles (as long as students are clear about what they are looking at)
- don't require time-consuming attention to color or design

Encourage students to invent different forms until they discover some that work well in organizing their data. Sketch graphs might be made with pencil and paper, with Unifix or other connecting cubes, or with stick-on notes. Cubes and stick-on notes in particular offer flexibility because they can easily be rearranged.

Two standard forms of representation are particularly useful for a first look at the data—line plots and stem-and-leaf plots. Line plots are introduced during the investigation *How many raisins are in a box?*; stem-and-leaf plots during the investigation *How long can we hold our breath?* ■



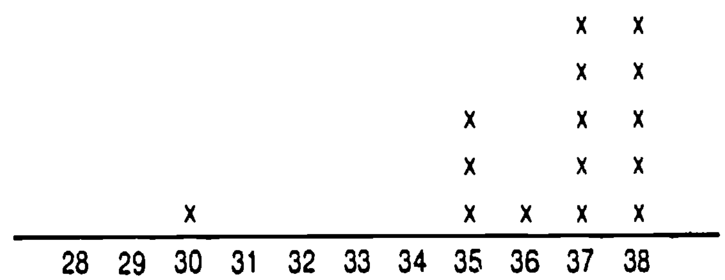


TEACHER NOTE

Line plot: A quick way to show the shape of the data

A line plot is a quick way to organize numerical data. It clearly shows the range of the data and how the data are distributed over that range. Line plots work especially well for numerical data with a small range, such as the number of raisins in a box.

This representation is often used as a working graph during data analysis. It is an initial organizing tool for beginning work with a data set, not a careful, formal picture used to present the data to someone else. Therefore, it need not include a title, labels, or a vertical axis. A line plot is simply a sketch showing the values of the data along a horizontal axis and X's to mark the frequency of those values in the data set. For example, if students have just collected data on the number of raisins in 15 boxes, a line plot showing these data might look like this:



From this display, we can quickly see that two-thirds of the boxes have either 37 or 38 raisins. Although the range is from 30 to 38, the interval in which most data falls is from about 35 to 38. The outlier, at 30, appears to be an unusual value, separated by a considerable gap from the rest of the data.

One advantage of a line plot is that each piece of data can be recorded directly on the graph as it is collected. To set up a line plot, students start with an initial guess about what the range of the data is likely to be: What should we put as the lowest number? How high should we go? Leave some room on each end of the line plot so that you can lengthen the line later if the range includes lower or higher values than you expected.

By quickly sketching data in line plots on the chalkboard you provide a model of using such graphs to get a quick, clear picture of the shape of the data. ■

✎ TEACHER NOTE

The shape of the data: Clumps, bumps, and holes

Describing and interpreting data is a skill that must be acquired. Too often, students simply read numbers or other information from a graph or table without any interpretation or understanding. It is easy for students to notice only isolated bits of information (e.g., "Vanilla got the most votes," "Five people were 50 inches tall") without developing any overall sense of what the graph shows. Looking at individual numbers in a data set without looking for patterns and trends is something like decoding the individual words in a sentence without comprehending the meaning of the sentence.

To help students pay attention to the shape of the data—the patterns and special features of the data—we have found it useful to use such words as *clumps, clusters, bumps, gaps, holes, spread out, bunched together*, and so forth. Encourage students to use this casual language about shape to describe where most of the data are, where there are no data, and where there are isolated pieces of data.

A discussion of the shape of the data often breaks down into two stages. First, we decide what are the special features of the shape: Where are the clumps or clusters, the gaps, the outliers? Are the data spread out, or are lots of the data clustered around a few values? Second, we decide how we can interpret the shape of these data: Do we have theories or experience that might account for how the data are distributed?

As an example, consider the graph below which shows the weights (in pounds) of 23 lions in U.S. zoos.

25–49	✓ ✓
50–74	
75–99	
100–124	✓ ✓
125–149	
150–174	
175–199	
200–224	✓
225–249	✓
250–274	
275–299	✓ ✓ ✓
300–324	✓ ✓ ✓
325–349	✓ ✓
350–374	✓
375–399	✓
400–424	✓ ✓ ✓
425–449	✓
450–474	✓ ✓ ✓

(Source: Zoos in Atlanta, Cleveland, Little Rock, Memphis, Miami, the Bronx, Philadelphia, Rochester, San Antonio, and Washington, DC. Data collected in 1987.)

☛ Note that this example is included here for teacher use only; the same data will be presented as a Mystery Set in a later investigation, *Finding landmarks in the data*.

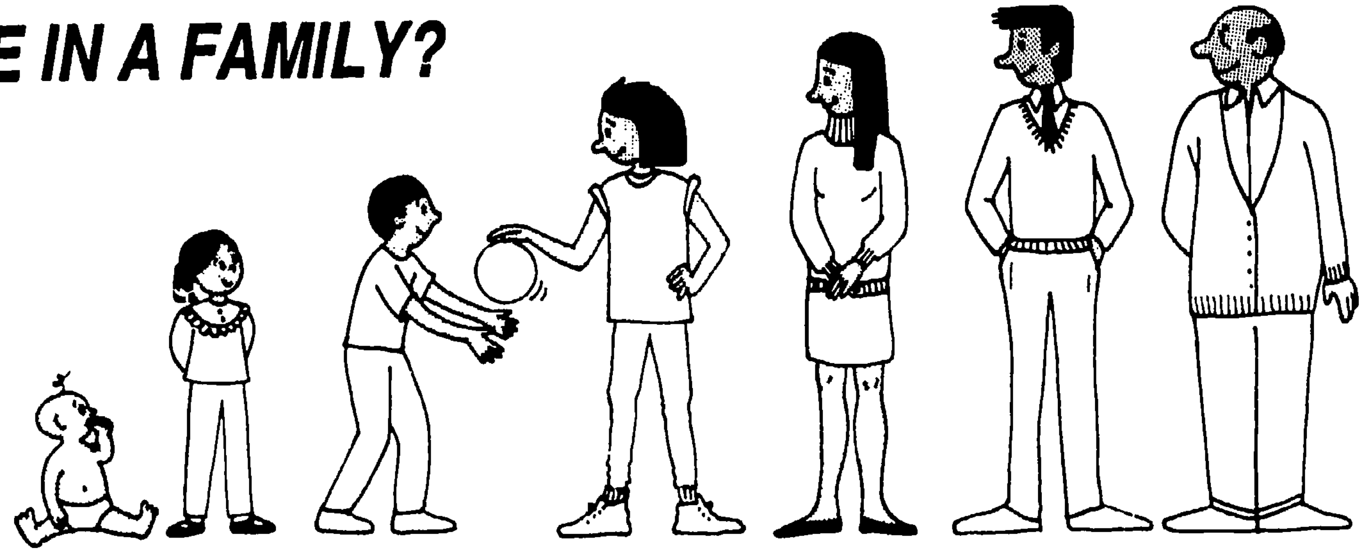
In the first stage of discussion, students observed the following special features:

- There is a clump of lions between 400 and 475 pounds (about a third of the lions).
- There is another cluster centering around 300 pounds (another third).
- There are two pairs of much lighter lions, separated by a gap from the rest of the data.

In the second stage of discussion, students considered what might account for the shape of these data. They immediately theorized that the four lightest lions must be cubs. They were, in fact, one litter of 4-month-old cubs in the Miami Zoo. The other two clusters turned out to reflect the difference between the weights of adult male and female lions.

Throughout this unit, we strive to steer students away from merely reading or calculating numbers drawn from their data (e.g., the range was 23 to 48, the median was 90, the biggest height was 52 inches). These numbers are useful only when they are seen in the context of the overall shape and patterns of the data set and when they lead to questioning and theory-building. By focusing instead on the broader picture—the shape of the data—we discover what those data have to tell us about the world. ■

HOW MANY PEOPLE ARE IN A FAMILY?



INVESTIGATION OVERVIEW

What happens

Students discuss information about the typical size of a family in their community. They decide how to count family size for their class; then they collect the data, describe the shape of the distribution, and determine typical family size for the class.

The activities take two class sessions of about 35 minutes each.

What to plan ahead of time

- Find out the typical family size for your community. This information can be obtained from your city's or town's government offices. They have census data on record and are usually glad to provide this information over the phone. They may also send you further demographic information about the community. The information available is usually the mean family size, and it is typically in decimal form (e.g., 3.24). The use of decimals is not recommended for this investigation; "about 3" or "between 3 and 4" is accurate enough.
- Provide Unifix cubes or a similar concrete material for representing the data (optional, Sessions 1 and 2).
- Provide unlined paper for sketch graphs and data descriptions (Session 2).
- Provide copies of the class list with student data added, or display the data prominently where everyone can see them (Sessions 2 and 3).

☛ Save the data for use as another data set in the investigation, *Finding landmarks in the data.*

Important mathematical ideas

Defining the way data will be collected. Making decisions about how to count or measure is a key aspect of data analysis. Such decisions can have a profound effect on the eventual outcome of statistical work, even before any data is collected. For the consumer of statistics, knowing how such decisions have been made is critical for interpreting the findings of data analysis.

Making quick sketches of the data and describing the shape of the data. Continue to reinforce the processes introduced in the preceding investigation, *How many raisins are in a box?*

Summarizing what is typical of the data. Encourage students to develop their informal ideas about typicalness. Students often invent important concepts about summarizing data, expressed informally as ideas about the middle of the data and where data appear to be "clumped." Ideas to encourage and those to avoid are discussed in the Teacher Note, *Summarizing data: What's typical?* ■

SESSION 1 ACTIVITIES

Considering the problem: Defining family size

Introduce the data you obtained from your local community.

I called up City Hall in [name of your town] and I asked them what the typical family size is in [your town]. They told me the typical family size is about 3. What do you think they mean by that? How do you think they found out?

After a brief discussion, ask:

Do you think the typical family size for our class might be the same as it is for our town?

Allow students to express their opinions about this, encouraging them to give reasons for their opinions.

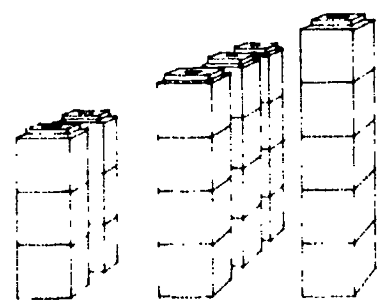
Let's take a look at the data for our class. If we're going to study family size for this class, how would we count who is in a family?

How to count the people in a family has provoked a lively and extended discussion in every group that has investigated this problem, from third graders to adults. Defining how to count or measure is a critical part of data analysis. Initial decisions about definition profoundly affect the outcomes of many statistical studies. Allow 15–20 minutes for initial discussion and 5–10 minutes for coming to a decision about how to count.

Help the students arrive at a consensus about a definition for family size. Try to avoid voting as a means of settling disagreements; try to help them to determine instead what aspects of family size they are most interested in. Note that the object is not to arrive at the same definition the U.S. census uses, but to arrive at a definition that the class as a whole decides is reasonable for their own data collection. See the Dialogue Box, *Who is in your family?* (page 23).

Collecting and recording data: Counting who is in your family

Using the consensus definition, students record their names and family size either in a display where everyone can see it or on a class list that will be duplicated for the small-group work in Session 2. If you have Unifix or other interlocking cubes, each student might build a tower to represent his or her family; that is, a tower of 4 cubes represents a family of 4 people. Then the towers can be arranged in a sequence or in groups to show how many families of each size there are. ■



“”DIALOGUE BOX
Who is in your family?

Students are usually eager to discuss their families. As this is a topic of intense personal interest and concern, allow adequate time for everyone to participate in the discussion. Diversity in family structure is to be expected and respected. Students may bring up many kinds of family situations. What about grandparents, aunts, or uncles who live with them? If we're going to count Eddie's grandmother who lives with him, what about Irena's grandfather who lives upstairs from her, or Rafael's aunt with whom he spends the summers? Students often talk about parts of their families living in different locations ("When I'm at my dad's house, there are three of us there, but when I'm at my mom's, there are four of us there"). Older siblings who no longer live at home, foster children, people temporarily living with the family, permanent members of the household who are not relatives, and even pets have been brought up in these discussions ("My sister only comes home to borrow the car, so can I count her as a quarter of a person?").

Teachers can handle these discussions sensitively so that all family styles and arrangements are acknowledged and accepted. But eventually students must come to some consensus about a definition of family. There is, of course, no single right way to construct this definition; in fact, different groups have settled on different definitions, depending on how they wanted to pursue their investigation. These have included: "You and your parents and your sisters and brothers, regardless of where they live"; "Everyone who lives in your house right now"; or, as one group of adults finally decided, "Everyone who uses the same bathroom as you!" The definition can be constructed to reflect what the students are most interested in. For example, if they want to know how many people actually share their living space, they would want to count everyone in the household; but if they are more interested in how many brothers and sisters people have, they might want to count all siblings, no matter where they live. Students may want to vote on a definition, but discourage this approach. As amateur statisticians, students shouldn't select a definition because it is the most popular, but because it will lead them to collect the data that will best give them the information they want. ■

SESSION 2 ACTIVITIES

Organizing and describing the data: What's the shape of these data?

Students work in small groups to represent and describe the data. Each group makes a sketch or picture of the data in at least two different ways. Encourage students to use the line plot, demonstrated in the raisins investigation. A concrete material, such as Unifix cubes, provides another excellent way of displaying family sizes. This is a good opportunity for students to invent new ways to display their data.

Each group writes down a description of the data, based on their representations, and makes a joint decision about how best to describe the typical family size for the class. See the Dialogue Box, *Some student descriptions of typical family size* (page 24).

Interpreting the data: Family size in our class, family size in the community

Ask for reports from the small groups:

What did you notice about our data? What did you choose for a typical size? Why?

Expect students to give reasons for their decisions, and encourage them to ask questions of each other about their choices. In this discussion, students often express informally some important ideas about how to summarize data, such as looking at the middle of the data, or at where the data are "clumped." See the Teacher Note, *Summarizing data: What's typical?* (page 25).

How does the typical family size in this class compare to the statistic about typical family size in our community? Are they different or similar? Why might this be true?

Guide the discussion with additional questions as needed:

What kinds of people live in the community who aren't represented in our class or even in our school? [people without children, people whose children have grown up and left home, single people] Do you think the family size in our class is like the family size of other classes in the school?

Extensions

Many classes have become interested in surveying other classes in the school, comparing other classes to their own, and compiling the data from many classes to see what the typical family size is for a larger number of students.

Students might also enjoy a study of the census. At this writing, the national 1990 census is just around the corner, and there will be a lot of public information available. Students may be aware of census materials coming to their homes. Educational materials about the census are available from: U.S. Census Bureau, 2601 North Howard Street, Baltimore, MD 21218, Attention: Lorraine Taylor. Invite a working statistician who deals with demographics in your community to come talk with the class about what he or she does and what the population trends in your area seem to be. ■

"" DIALOGUE BOX
Some student descriptions of typical family size

CATHY: In ou· class the smallest family is two and the largest is ten, and I think you should pick a number right in the middle, so I picked six because it's right in the middle between two and ten.

SARAH: The most families in our class have four people, so I think four is the most usual family size.

RICARDO: I think around three or four, because there's a big clump of families around there.

KYLE: More than half of the kids have either three or four people in their families, so I think three and four.

IRENA: You could say "three and a half" because there are a lot of kids at three and a lot at four, so it's sort of between three and four. ■

✎TEACHER NOTE
Summarizing data: What's typical?

Summarizing data is one of the main tasks of data analysis. A data set starts out as an unordered set of many values. In data analysis, we need to capture the essence of the shape of the data through a few key numbers so that we can describe and compare data sets without referring to all the values.

As consumers of statistics, our encounters with data sets are often through these key numbers: the average (mean) number of people in a U.S. household in 1985 was 2.69 (a record low); in the same year, the median age of the population was 31.5; also in 1985, more women in the U.S. work force were employed in "administrative support" than in any other category of occupations, while the mode for men was the category of "precision production, craft, and repair."*

Statisticians try to capture the essence of the data by identifying its center, or average, and then describing how the data are spread around that center. We are used to thinking of "average" as the arithmetic mean, the number obtained by adding all the values and dividing by the number of values. Actually, there are many possible measures of average. The *mean* (used above as a measure of people per household) is one; *median* (used above for the age of the population) and *mode* (used above in the occupational statistics) are other measures of average.

In order for students to understand these summary methods, they need experience with situations in which they feel a need for summarizing the data. Deciding what is usual, typical, or central for a group is one reason to summarize data. Students will encounter the question "What's typical?" in many of their data analysis investigations. For comparing two sets of data, they will also find it necessary to summarize each data set. For example, to determine whether the number of people per household in the U.S. is increasing or decreasing, we must capture the shape of large data sets in a few numbers to help us compare data from different years.

Students begin to understand how to summarize data by developing their own approaches during the investigations. For example, when comparing the heights of first and fourth graders, one fourth grader hit upon an important notion about how to summarize a data set: "We should find the number that's maybe in the middle or that all the other numbers are crowded around."

In a later investigation, *Finding landmarks in the data*, you will develop the idea of a *median*, a powerful and developmentally appropriate average for use at the upper elementary grades. Understanding of the median builds on the informal ideas about the center of the data that students develop in the first few investigations. *Mode* will also be touched on in *Finding landmarks in the data*.

Chances are your students already know how to compute the arithmetic mean or, as they may call it, the "average." Although this type of average is often taught in elementary school, research has shown that the nature and significance of the mean is often not understood, even by older students and adults.

We recommend that teachers discourage any use of the mean until students have had a lot of exposure to data analysis. Students need experience with a great variety of data sets before they are ready to understand how the arithmetic mean relates to the data it represents. A later unit in the *Used Numbers* series, *Statistics: Middles, Means, and In-Betweens*, provides appropriate experiences for learning about the mean.

In the meantime, discourage your students from applying the add-'em-all-up-and-divide-by-the-number approach they may have learned with little understanding. If a student says, "We could find the average," you can respond something like this: "Yes, we could. Actually, there are many kinds of averages that you'll learn about as you go on in mathematics. Averages are often ways of saying what's typical about a set of data. Right now we're going to be inventing our own ways of deciding what's typical, and later in this unit you're going to learn about one kind of average." ■

*U.S. Bureau of the Census, Current Population Reports, Special Studies Series P-23, No. 150, *Population Profile of the United States: 1984/85*, U.S. Government Printing Office, Washington, DC, 1987.

HOW LONG CAN WE HOLD OUR BREATH?

INVESTIGATION OVERVIEW

What happens

Students plan and carry out an investigation into how long the members of their class can hold their breath. In the first session, they experiment with breath-holding, come up with a plan for collecting data, and learn how to use a stem-and-leaf plot for representing their data. In the second session, they collect data and work in small groups to decide on a typical breath-holding time for the class.

The activities take two class sessions of about 40 minutes each.

What to plan ahead of time

- If the classroom clock does not have a clearly visible second hand, provide a timer, stopwatch, or regular watch with a second hand for each group of students (Sessions 1 and 2).
- Have calculators available (Sessions 1 and 2).
- Provide unlined paper for sketch graphs and data descriptions (Session 2).
- Become familiar with making a stem-and-leaf plot. See the Teacher Note, *Stem-and-leaf plot: Another quick way to organize data* (page 29).

☛ Save the data for use again in a later investigation, *Finding landmarks in the data.*

Important mathematical ideas

Defining the way data will be collected. In this investigation, students design an experiment to collect data. They will have to come to a consensus on one method, but should be aware that different methods of collecting data can lead to different findings. For practical reasons, the many variations students think of cannot all be tried in a classroom. It is important that they keep in mind, though, the possible effect of their chosen data collection method on the outcome of their statistical investigation.

Using more than one representation to view data. Throughout the unit, students should be developing a repertoire of quick and clear ways to represent their data for analysis. In this investigation, students will learn about a new representation, the stem-and-leaf plot (described in the Teacher Note, *Stem-and-leaf plot: Another quick way to organize data*). Encourage them to use this new representation as well as to continue developing their own ways of displaying data. Different representations work best for different data, and individuals may develop preferences for certain representations that work well for them.

Describing the data. In addition to using such informal descriptive language as *clusters, clumps, gaps, spread out,* and *bunched together*, use the statistical terms *range* and *outlier* whenever they come up in the data, as discussed in the Teacher Note, *Statistical words to introduce as appropriate.*

Summarizing what is typical of the data. Continue to reinforce the data summary methods that students developed in the preceding investigation, *How many people are in a family?* ■

SESSION 1 ACTIVITIES

Considering the problem: Introducing the investigation

When we learn to swim, we practice holding our breath under water. Today we're going to do an experiment to find out something about the human population in this room: How long can fourth [fifth, sixth] graders hold their breath?

Different kinds of mammal populations can hold their breaths for different amounts of time. Beavers can hold their breath under water for about 15 minutes. Humpback whales can stay under water for about 25 minutes, while sperm whales, which dive very deep to find food, have developed the ability to hold their breath up to an hour and a half. But these times are at the high end of what a whale can do; whales typically come to the surface to breathe about every two minutes. An active harbor seal takes a breath about every 5–7 minutes. Sea otters typically stay underwater for 4–5 minutes.*

We are going to try to determine the typical time for the class—not the longest or the shortest, but what might generally be expected. So our question is: What's the typical amount of time people in this class can hold their breath?

Ask for some estimates from students about how long they can hold their breath and what they think might be typical for the class.

Considering the problem: Planning the investigation

Today we're going to plan the investigation and in the next session we'll carry it out. To get us into thinking about the best way to collect these data, have each person in your group try holding his or her breath once, time the person, and write down the time. Then, as a group, think about how we should do the experiment.

Are there different ways we could do this? Should each of you hold your breath once or perhaps a few times? Should we take the highest times? What other problems can you think of that we might run into in collecting these data? Come back with some suggestions for a plan for how we might do our experiment.

In small working groups, students time each other while everyone has a chance to hold his or her breath once. They record these data. Then they discuss and formulate a plan for collecting class data.

All groups report on their questions and ideas. Listen to students' ideas, then support the class in coming to a consensus about

* Statistics from B. N. Irby, M. K. McEwen, S. A. Brown, & E. M. Meek (eds.), *Diversity of marine animals*, published for the Mississippi-Alabama Sea Grant Consortium (Jackson, MI: University Press of Mississippi, 1984); *Seals*, grades 3–6 (Boston, MA: New England Aquarium, n.d.); Roy Nickerson, *Brother whale: A pacific whalewatcher's log* (San Francisco: Chronicle Books, 1977); Dorothy Patent, *All about whales* (New York: Holiday House, 1987).

how they will carry out the experiment.

Organizing data: Using a stem-and-leaf plot

Let's take a look at your preliminary data. First we'll use a line plot; then I'm going to show you a new way to sketch your data that you might want to use tomorrow.

Make a line plot of the breath-holding data students collected in their groups, asking students for ideas about where the line plot should begin and end and what intervals to use in marking it. If some data is in seconds and some is in minutes and seconds, have students change all their data to seconds. Calculators should be available for this conversion. You may want to have a couple of students share their methods for converting.

What can you say about these preliminary data?

Students describe the shape of these data. Usually, breath-holding data will be much more spread out than the raisin or family-size data.

The raisin data and the family-size data had a much smaller range than these data.

Ask students if they recall what the range (lowest to highest value) of the data was in the raisin and family-size investigations.

But here the range is from [22 to 91 seconds], and if we use a line plot, the data are really spread out and it's hard to see whether there are any real clumps. Another way mathematicians sketch their data is by using a stem-and-leaf plot.

Demonstrate how to make a stem-and-leaf plot with the students' data. To involve the class in making this new representation, have each student in quick succession tell you again his or her breath-holding time. Record the times on the stem-and-leaf plot in the order that the students call them out. Then demonstrate how to reorder the numbers in each row of the plot so that each row is in sequence. See the Teacher Note, *Stem-and-leaf plot: Another quick way to organize data* (following).

What can you see about the shape of the data now? Are there any differences between what you can see in the stem-and-leaf plot and the line plot?

Seek honest student opinions about whether the stem-and-leaf works better for these data than the line plot. How well each represents the data depends both on the particular data and personal preference. Regardless of which plot students prefer, emphasize that looking at data in more than one way can often give us further insight into the reality which the data represent. Mathematicians and statisticians frequently try many ways of viewing the data during analysis—for any investigation, we may not know in advance which representations will work best. ■

TEACHER NOTE
Stem-and-leaf plot: Another quick way to organize data

The stem-and-leaf plot (also called a stem-plot) is, like the line plot, an easy way to represent the shape of a data set. A stem-and-leaf plot works best for data with a range of several decades, since the plot is most frequently organized by tens.

For example, here is a collection of data showing how long 21 children could hold their breath (in seconds):

55 43 30 22 91 48 79 38 32 59 79 45
35 62 55 39 47 42 80 47 79

To make a stem-and-leaf plot of these data, we divide each value into tens and units. The tens become the "stem" of the plot, and the units are the "leaves":

2	2
3	0 8 2 5 9
4	3 8 5 7 2 7
5	5 9 5
6	2
7	9 9 9
8	0
9	1

In this plot, the first line shows one data point in the twenties: 22. The second line shows five data points in the thirties: 30, 38, 32, 35, 39.

Once we have quickly organized the data into a stem-and-leaf plot, we can rearrange the leaves in each decade to put them in order:

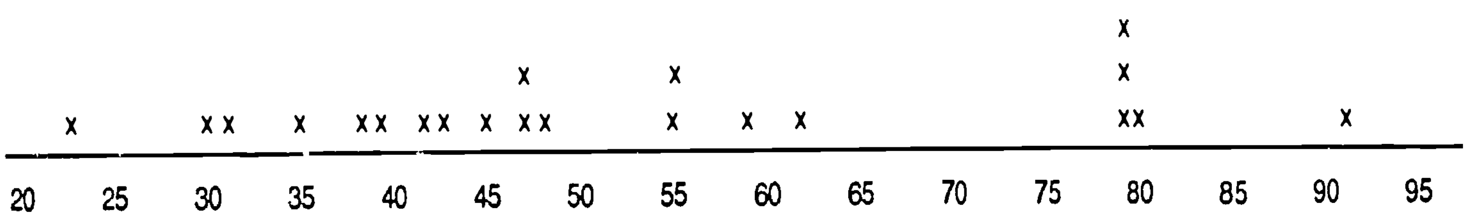
2	2
3	02589
4	235778
5	559
6	2
7	999
8	0
9	1

It is now easy to see many features of the data. The data range from 22 to 91 seconds; there is a large cluster of data—about half the data—in the thirties and forties. If we include the fifties in that cluster, we account for two-thirds of the data. There is an interesting small cluster at 79–80 seconds; data points above 60 seconds are very scattered.

For this particular data set, a stem-and-leaf plot helps to show how the data are clustered. Because of the large range involved, these data would appear very spread out on a line plot (see below). While it is harder to see clusters of data on the line plot, it is interesting to note that one of the characteristics of the shape of these data, the gap between 62 and 79, shows up more clearly on the line plot than on the stem-and-leaf plot.

Because different representations may highlight different aspects of the data, encourage students to use more than one sketch graph to represent any given data set. ■

A LINE PLOT SHOWING THE BREATH-HOLDING DATA FOR 21 STUDENTS



SESSION 2 ACTIVITIES

Collecting and recording data: Timing breath-holding

According to the plan formulated in the last session, students work in pairs or small groups to collect their breath-holding data.

Record their data, without ordering it, on the chalkboard or on a chart visible to everyone. One teacher gave a class list to each group. They recorded their own data next to their names, then passed their list to the next group, which recorded its data and passed on the list, until every group had everyone's data.

Organizing and describing the data: Deciding on a typical value

Working in small groups, students do three things. First, they make sketches or pictures of the data in at least two ways. Encourage them to try a stem-and-leaf plot as one of their sketches, but to also try other graphs, diagrams, or pictures. Second, they write down a description of the data: Where are the data clumped? Is there more than one clump? Are there holes in the data? Are there any outliers? Third, they decide what a typical time is for breath-holding in the class. Stress that they should have good reasons for their choices.

Students report their findings to the class. Expect them to give reasons supporting their choices for typical values. See the Dialogue Box, *Discussing invented methods for finding typical values.*

Extensions

Try a different design for the experiment. One class thought it might make a difference whether or not people watched the timer while holding their breath. In fact, some thought they would hold their breath longer if they could see the seconds going by, and some thought they would hold their breath for a shorter time under those circumstances. Does looking at the clock make a difference?

Some groups have come up with theories about other factors that might affect breath-holding time. For example, pearl divers and other trained divers can hold their breath for up to 3 minutes. Can people train themselves to hold their breath longer? One group wondered if adults who don't smoke would be able to hold their breath longer than adults who do smoke. Another group thought that people who exercise regularly might be able to hold their breath longer than others. Some students might pursue an investigation in one of these areas. ■

“”DIALOGUE BOX

Discussing invented methods for finding typical values

CHRIS: We think you should pick a number that comes up the most, so we picked 55, because there are more 55s than any other number.

What does everyone else think about that method?

CARMEN: We did the same thing. There are a lot of 55s, so that seemed like what was typical.

Did anybody make a different choice?

ALICE: We came out with 63.

So your choice is a little higher than what Chris's group picked. Why do you think that's reasonable?

ALICE: I don't know. It just seemed like that would be it.

But I'm interested in your reason for 63. *[Pause. Still no response from the student.]* **I see a big clump of data in the 60s.**

ALICE: Yeah, the clump seems like it's crowded around 63.

Yes, that's an interesting method. I can see your reasons for both of these methods. Does anyone have a good argument for choosing one over the other for the most typical value for our class?

JOEY: Well, even though 55 has the most, there are still only five kids with 55, but there are 12 kids bunched around the low 60s.

So you'd choose the biggest clump of data?

[Later in the same discussion] . . .

BEN: Our group chose 90.

What were your reasons?

BEN: A lot of kids got 90, and also it's the highest number.

You chose the highest number?

BEN: Yes, because a bunch of kids got it.

What do some of you who picked numbers in the 50s or 60s think about that?

RAFAEL: I don't think you should pick the highest, because that's not what's typical. Most kids couldn't hold their breath that long.

SUE: Yeah, like with the whales, the typical time wasn't the shortest or the longest, but somewhere in the middle.

So you think a middle value is more typical. What do you think about that argument, Ben?

BEN: Yeah, but maybe more people could really hold their breath that long, but they didn't try really hard.

That would be a good experiment. You're asking a new question: If you got people to try really hard, would more times come out in the 90s?

(Dialogue Box continued)

BEN: Yeah.

So you noticed that four kids got 90; that's an important part of the shape of these data. But Sue and Rafael are arguing that with these data, a *typical* value wouldn't be that high.

BEN: But can we try my experiment?

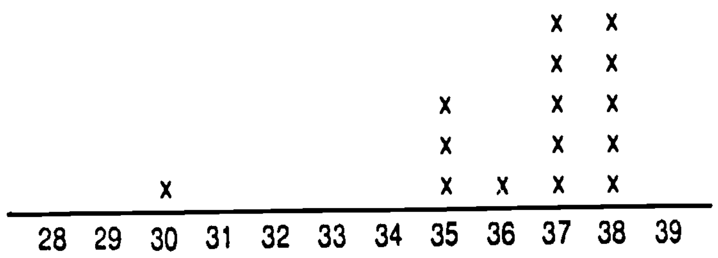
☛This discussion can be a difficult one for the teacher because, while there is certainly no single right answer, students may come up with unreasonable approaches. Encourage all invented methods; many of the students' ideas will help them understand standard measures of center such as the median and mode, which they will encounter in a later investigation. However, do not let students get the message that any method is as good as any other; expect them to reflect on whether or not their results are reasonable. Juxtaposing one student method with another, as the teacher does in this discussion, is often a good way to help students think about the reasonableness of their method. ■

TEACHER NOTE

Statistical words to introduce as appropriate

Range and *outlier* are two statistical ideas that come up naturally in discussing data with students.

The range of the data is simply the interval from the lowest value to the highest value in the data set. The range of the data in the line plot below, showing how many raisins were in each of 15 boxes of raisins, is from 30 to 38:



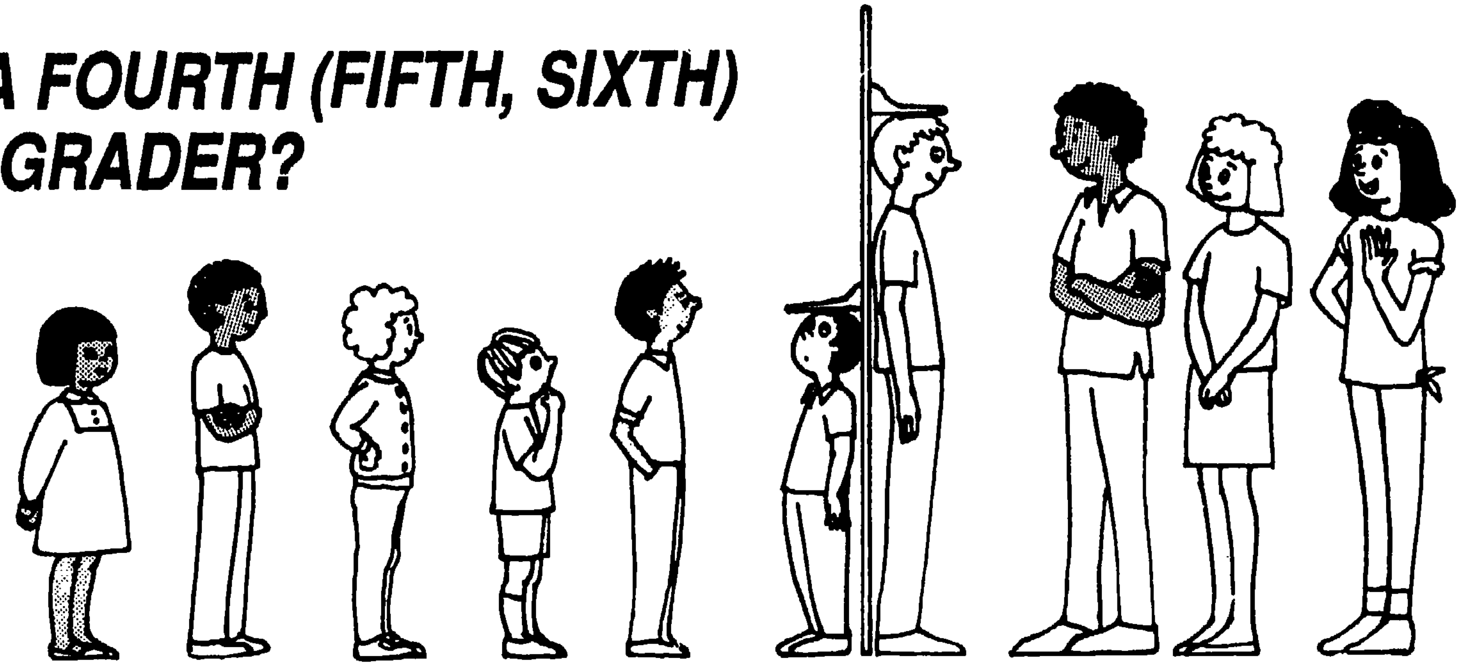
An outlier is an individual piece of data that has an unusual value, much lower or much higher than most of the data. That is, it "lies outside" the overall shape and pattern of the data. There is no one definition of how far away from the rest of the data a value must be to rate mention as an outlier. Although statisticians have rules of thumb for finding outliers, these are always subject to judgment about a particular data set. As you view the shape of the data, you and your students must judge whether there are values that really don't seem to fit with the rest of the data. For example, in the raisin data, the box containing 30 raisins seems to be an outlier. In the breath-holding data shown in a stem-and-leaf plot (page 30), it is not clear whether any values are outliers. The few pieces of data in the 60s, 70s, 80s, and 90s are so scattered that it is hard to say whether the value 91 really falls outside the pattern of the data or is just a part of the scatter at the upper end of the plot.

Both *range* and *outliers* are ideas that will come up naturally in the work with different data sets in this unit. They can be introduced as soon as they arise in the students' descriptions of their data. Students easily learn the correct terms for these ideas and are particularly interested in outliers. Outliers should be examined closely. Sometimes they turn out to be mistakes—someone counted, measured, or recorded incorrectly—but other times they are simply unusual values. Students are generally very interested in building theories about these odd values: What might account for them? ■

Statistics: The Shape of the Data

PART 2
Learning about landmarks in the data

HOW MUCH TALLER IS A FOURTH (FIFTH, SIXTH) GRADER THAN A FIRST GRADER?



INVESTIGATION OVERVIEW

What happens

Students measure their own height and the height of students in a first grade class, compare the heights of the two classes, and find ways to describe and represent the comparison. With this problem, you and your students move into a more complex investigation requiring several class sessions.

The activities take four class sessions of about 40 minutes each (minimum). One way of breaking the investigation into four sessions is suggested in the following outline; you can vary this as appropriate for your students. If students decide to remeasure the class or to take measurements of more classes, you may need an additional session. Also, some students may need more time to finish their presentation graphs.

What to plan ahead of time

- Provide measuring tools—yardsticks, metersticks, or tape measures—for each group of students (Sessions 1 and 3).
- Have calculators available.
- Provide unlined paper for sketching the data.
- Duplicate Student Sheet 1 (page 69) for each small group (Session 2).
- Arrange with the first grade teacher(s) a way for your students to collect the heights of first graders (Session 3).
- Provide materials for making presentation graphs: squared paper with inch squares or centimeter squares (a reproducible sheet of centimeter squares is provided on page 79), and colored markers or crayons (Session 4).
- Copies of the height data (for both their own class and the first graders) for each small group (Sessions 2, 3, and 4).

Important mathematical ideas

Inventing ways to quantify differences between two sets of data. In order to compare two data sets, students use the ideas they have developed about describing the shape of the data and summarizing what is typical of a particular set of data. Comparison motivates students to capture the data in a single number or a small interval ("Fourth graders are about 55 inches tall" or "Fourth graders are between 54 and 56 inches tall") without ignoring the context for that number

("Typically, fourth graders' height is from 50 to 58 inches tall, but a few fourth graders are shorter or taller").

Using linear measurement in the context of a mathematical investigation. The measurement process is discussed in the Teacher Note, *Measuring heights: Using tools.*

Inventing representations for comparing two sets of data. Some of the variety possible in this creative task is illustrated in the Teacher Note, *Presentation graphs: Inventiveness and clarity.*

Representing data first through sketches, then developing a presentation graph or chart. In this investigation, students move from a "rough draft" stage through revision and refinement to presentation of their findings in a more finished form. The steps of the process are described in the Teacher Note, *Stages in data analysis: Learning from the process approach to writing.* ■

SESSION 1 ACTIVITIES

Considering the problem: Deciding how to measure

Sometimes we collect data in order to compare one group to another. For example, we could use your raisin data and compare them with a different brand of raisins, or we could compare your breath-holding data with data for a younger or older group of students. Today we're going to start working on a problem that involves comparing two sets of data: How much taller is a fourth (fifth, sixth) grader than a first grader? First we'll collect some initial data—your heights; then we'll plan how to do the rest of the investigation.

Have a couple of pairs of students demonstrate measuring each other's height. Use this demonstration to raise questions about how they are going to measure each other. Will shoes be on or off? What units of measurement will they use? Will they measure to the closest inch, half-inch, quarter-inch? Resist the temptation to raise questions yourself; support students as they raise and resolve their own questions.

Collecting initial data: Measuring height for our class

Students work in groups of three to measure their own heights. One student is measured, one student does the measuring, and the third checks the accuracy of the measurement and records it. Large pieces of paper tacked on a wall or bulletin board on which students can mark their heights greatly facilitate the measuring process. See the Teacher Note, *Measuring heights: Using tools* (page 38).

Discuss briefly any problems that arise during the measurement. See the Dialogue Box, *I'm 48 inches and 1 centimeter tall!* (page 37).

Describing the data: How tall are we?

Ask students how you can record their data on the chalkboard so everyone can see it. Students might suggest a line plot, a table, or tallies.

Using their suggestions, record the class data. If they have more than one suggestion, make more than one representation so they can look at the data in different ways.

What can you say about how tall people in our class are?

Support students in describing both the overall shape of the data—its range, how it clumps or spreads out, whether there are any outliers—and how they might summarize what's typical of them as a group (e.g., "Most of us are between 54 and 56 inches tall, but there's also a little clump around 50 to 51 inches"). ■

DIALOGUE BOX

I'm 48 inches and 1 centimeter tall!

During this investigation, you will probably discover that many students have difficulty with linear measurement. Teachers commonly find that some of their upper elementary students:

- combine metric and English systems ("I'm 48 inches and 1 centimeter tall.")
- do not distinguish between metric and English units ("I'm 125 inches tall.")
- do not know how to combine two parts of the measurement ("I'm 6 inches taller than one yardstick—how do I do that?")
- measure from the wrong end of the ruler or yardstick

Don't be alarmed! Such difficulties are common among students who have had little experience with measurement in a real context. Use this investigation as an opportunity for students to gain experience and to help each other be accurate. Having students discuss and check accuracy as they work in small groups is often more effective than teaching a lesson on linear measurement. Have students share the difficulties they encounter; support students who use their own experience and knowledge to check the reasonableness of their measurements.

What were some of the problems you noticed in trying to get accurate measurements of your height?

SARAH: Well, in our group we got 50 inches for Mary, and then we got 125 inches for Karen, and we knew that didn't make any sense because Karen and Mary are about the same.

KAREN: And, anyway, no one could be 125 inches!

SUE: Yeah, that would be past the ceiling. *[Laughter.]*

I'm not sure it would be past the ceiling, but you're right—it doesn't sound like a fifth grader! So you paid attention to whether your measurements seemed reasonable. Before you tell us what was wrong, can anyone else guess how they got 125 inches for someone's height?

DAVE: I know, because we did the same thing. Because on one side is centimeters and on one side is inches, and we were using the centimeter side at first. Then we realized it couldn't be over 100 inches for somebody's height.

Is Dave right? Was that the problem?

ROBERTO: Yeah, we got confused about the two sides of the yardstick.

Anyone else come up with a problem they had to solve when they were measuring?

JANE: Yeah, we did. We measured me and we came out to 57 inches, and I *know* I'm not 57 inches.

How did you know?

JANE: Because my mom keeps a chart at home of all the kids and we mark it on our birthdays, and I know I'm shorter than that. On my birthday I was 51.

KIM: And also Jane's one of the shortest in our class, and 57 wouldn't be near to the shortest height in the class.

So you realized that 57 inches wasn't reasonable? Did you figure out what happened?

JANE: Yeah. When we measured the first part and then we moved the ruler up to measure the rest, we flipped it around and we had the ruler going in the wrong direction.

Oh, you mean you measured from the end of the yardstick that says 36?

KIM: Yeah, we got to 36 and then we swung it around like this, so we were measuring from 36 *up*, and we marked the second part at 21, so we had 36 plus 21 inches and we got 57. But we were going the wrong way. ■

✎TEACHER NOTE
Measuring heights: Using tools

Measuring heights seems simple enough, but for upper elementary students it can pose a real challenge. Even though students can do measurement worksheets and manipulate measurement data on paper, they may not have had much experience using rulers and other measurement tools. Students who have done woodworking, who have built things at home, who have played with and built models (including dollhouses) will be the most expert at this activity. They have some physical experience to draw from—they are familiar with the tools and know how to use them, and they have internalized the sizes of the measurement units.

Conducting measurement activities in small groups allows you to take the observer's role, to see what understandings and misconceptions your students have. Listening to their comments and questions as you circulate will give you some real insight into the mechanical and conceptual problems that measurement often presents to upper elementary students.

Some predictable problems arise when students use measurement tools. The need to line up the ruler at zero is not always obvious. Additionally, students may start from the wrong end when they pick up and move a ruler. They may combine units, using both metric and English systems. They may not notice that their "yardstick" is a meterstick. All these skills depend to some degree on prior measurement experience.

For many teachers, it helps to think of students' initial measurements as *first approximations* to their answers, rather than final results. A vital part of their learning is the opportunity to discuss the reasonableness of their measurements, to measure several times, and to correct their measuring mistakes. When students feel the results matter, they become much more precise.

A teacher's role in this process is delicate. Your students will need to discuss their methods (Should they use the yardstick? the foot ruler? Should they stand against the wall?) as well as their results. Asking questions that focus on whether their results make sense (Is Jeremy about the same height as Eliza? Is Imani that much taller than Sandra?) will probably be most helpful.

Measurement work provides a natural place to use the calculator. If some students measure in inches and others in feet and inches, you have a good opportunity to work with conversion methods. Although conversion seems straightforward to adults, it is hard work for upper elementary students. Students will have a variety of methods and will want to spend some time trying them out. Be sure that they estimate first, so that they begin to check the calculator results with common sense. This step will further reinforce the role of estimation in the measurement process. ■

SESSION 2 ACTIVITIES

Considering the problem: How would we figure out how much taller fourth (fifth, sixth) graders are than first graders?

Encourage students to suggest methods for solving this problem. Sudents will begin thinking about a method for comparing two data sets, but need not yet choose the method to use in this investigation. See the Dialogue Box, *How can we compare our class with a first grade class?* (page 39).

Give students a chance to reevaluate their data on heights in their class. They might suggest modifications that would require remeasuring everyone. Or, maybe they need more data from students in their own grade. If possible, support any decisions along these lines. See the Teacher Note, *Stages of data analysis: Learning from the process approach to writing* (page 39).

Considering the problem: An example of comparing two groups

Determine the small groups that will work together to complete the height investigation. Give out copies of Student Sheet 1, *Heights of two fourth grade classes.* In small groups, students work on comparing their own data with these fourth grade data. How would they describe the differences and similarities? Ask groups to write a brief description of their comparison. ■

DIALOGUE BOX
How can we compare our class with a first grade class?

So what kinds of plans can we make for figuring out how much taller our class is than Ms. Rivera's first graders?

MARCIA: We should measure all the first graders and all of us.

JESSIE: We could pair everyone up.

Pair everyone? How would that work?

JESSIE: Like the shortest with the shortest and the tallest with the tallest.

Oh, so you'd put everyone in order by height and pair up the shortest first grader with the shortest fifth grader? Then what?

JESSIE: You could subtract.

Jessie's suggesting we pair up first graders and fifth graders in order. *[Writes on the board: Pair up first graders and fifth graders.]* **Does anyone have some other ideas about that?**

EDDIE: Well, it's not exactly like that. But my idea is that you'd find the medium ones, like there's the shortest and there's the tallest, but that's not what's really typical, so you look at the mediums.

You'd compare the medium first graders and the medium fifth graders? *[Writes on board: Compare medium 1st and 5th.]* **Other ideas?**

ANNIE: We could find the one number that's sort of in the middle or that the other numbers are crowded around, and then do the same thing for the first grade and then subtract.

Can you say a little more about that?

ANNIE: You'd get all the heights and you'd put them on a graph, all in order, and then you'd look for maybe the middle number or a big clump where most of the heights are.

OK. I'm not sure if that's one idea or two different ideas. *[Writes: Use middle height; Look for a clump where most heights are.]* **Who else has a plan for comparing this class and the first graders?**

At this point, the teacher is not pushing for complete, detailed plans, but is trying to get out a variety of basic approaches. Students will be working in small groups to develop a more detailed plan later. The list from this brainstorming session will provide starting points for the small-group work. ■

TEACHER NOTE
Phases of data analysis: Learning from the process approach to writing

The process of data analysis is similar to many other creative processes. Students doing data analysis follow the same processes that adults do; the analyses may be less complex, but the procedures are the same. In data analysis, as in writing or art, teachers help children do real work rather than stilted school assignments requiring fill-in-the-blank responses. The teacher's role is relatively subtle—shaping the process, asking questions that guide the students' progress toward their goals, hearing and responding to their ideas and theories. Students are expected to have something original and interesting to say, and the teacher provides an environment that enriches and supports students' self-expression.

Data analysis has many similarities to the process approach to writing, which typically includes four phases. The process starts with a *planning phase* (often called pre-writing or brainstorming). This is followed by the *writing phase*, when a very rough draft of ideas is first put down on paper. The third phase is the *revision* or *rewriting phase* when the writer elaborates, clarifies, restructures, and edits the piece. The final phase is the *publication phase*, when the writer's completed piece is shared with others. These processes may be reiterated until the piece of writing is finished.

Data analysis has four phases parallel to those in the writing process:

Phase One: Brainstorming and planning. During this time, students discuss, debate, and think about their research question. In some cases, defining and agreeing upon the question may take a considerable amount of time. Having defined the question and agreed upon terms, students consider possible sources of data, ways of recording them, and how they might organize themselves to collect needed information.

Phase Two: Putting it on paper. For the collection and representation of data, students develop their discovery drafts—what we call "sketch graphs"—the first draft of the information on which they base their developing theories. Students represent the data in a variety of ways to help them describe the important features. They use their first drafts as tools as they look for relationships and patterns in the data.

Phase Three: Revision. Writers are encouraged to share their drafts with their peers in order to determine how an audience perceives their work. Similarly, in the data analysis process, the students often present their sketch graphs, preliminary findings, and beginning theories to their working group in order to see whether their interpretations seem supported by the data, and whether others see things they haven't noticed. Revision in data analysis may include finding new ways to organize and represent the data, developing better descriptions of the data, collecting additional data, or refining the research questions and collecting a different kind of data.

Phase Four: Publication or display. The nature of "publishing" the results of data analysis varies, just as it does for a story or essay. Sometimes students develop a theory that is the basis for a report on a particular topic; at other times they may develop a theory that inspires further investigation. A completed report of a data analysis investigation may involve a written description of the study with conclusions and recommendations, final presentation graphs of information previously displayed in working graphs, and a verbal or written presentation of the report to an interested audience.

When teachers think about the writing process, their role as facilitator and helper seems familiar and obvious. Of course students need time to think and revise their work! Of course they need to be challenged and led, sensitively, to the next level of awareness. The writing process seems more familiar to most of us than the mathematics process because we, too, have done writing.

The process of data analysis needs the same kind of teacher support. Students need to try their ideas, to rough them out, to be challenged and encouraged to go further in their thinking. It is important that they have time to think and to consider options—and vitally important that they see their work as part of a process. Data analysis, like writing, is not cut and dried. There are many ways to approach a question and many conclusions to be drawn. Like writing, mathematical investigation is a creative blend of precision and imagination. ■

SESSION 3 ACTIVITIES

Collecting more data: Measuring first graders' heights

You will have to decide with the students and the first grade teacher(s) how to collect the first grade height data. The first grade class(es) could come for a visit to your classroom, or teams of students could go to the first grade classroom. Because of the number of students involved, some teachers may feel tempted to select a measuring team to collect the data. However, it is very important that all students participate in some way in measuring first graders so that they have a feel for the data.

Have the data recorded in a central location or on a class list that can be duplicated for each small group.

Organizing, describing, and interpreting the data: Comparing two sets of data

In small groups, students do the following:

- Make quick sketches of the data, one for first grade heights, and one for their own heights. Each group might sketch the data in several ways.
- Analyze the shape of each data set. Where are the clumps, bumps, holes? What is the range? Are there any outliers?
- Decide on a method to compare the two data sets.
- Try out the chosen method. See if the result is reasonable. Try a different method, if needed.

Circulate while students work in their small groups. Emphasize that they are to come up with a good method for solving the problem, not just an answer. Remind them of methods they used to identify a typical value in the breath-holding problem that might work here as well. ■

SESSION 4 ACTIVITIES

Publishing findings: How much taller is a fourth (fifth, sixth) grader than a first grader?

Once each group of students has worked with the data long enough to settle on a way of comparing the two data sets, they decide how to represent and report their findings.

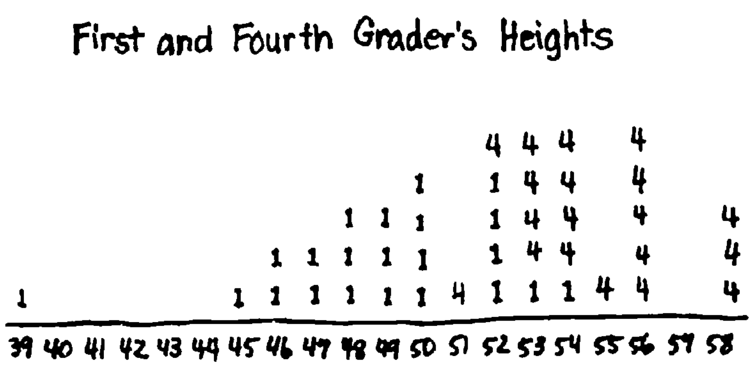
Each small group makes a presentation graph or chart to show the data and to illustrate their method for comparing the two sets. At the rough draft stage, encourage students to experiment with different ways to present their data before starting a final presentation graph. A group might find more than one good way of showing the data. Encourage inventiveness and clarity. Each group also writes a description of their method to go with their representation of the data. See the Teacher Note, *Presentation graphs: Inventiveness and clarity* (page 42).

You may or may not have each group present their findings to the rest of the class. Often, such presentations are much less interesting than the discussions and interaction that went on within the groups. Posting the representations and explanations on the bulletin board and giving students time to examine them may be sufficient. As an alternative, hold a brief evaluation conference with each group, just as you might do to evaluate a piece of writing. ■

✎TEACHER NOTE
Presentation graphs: Inventiveness and clarity

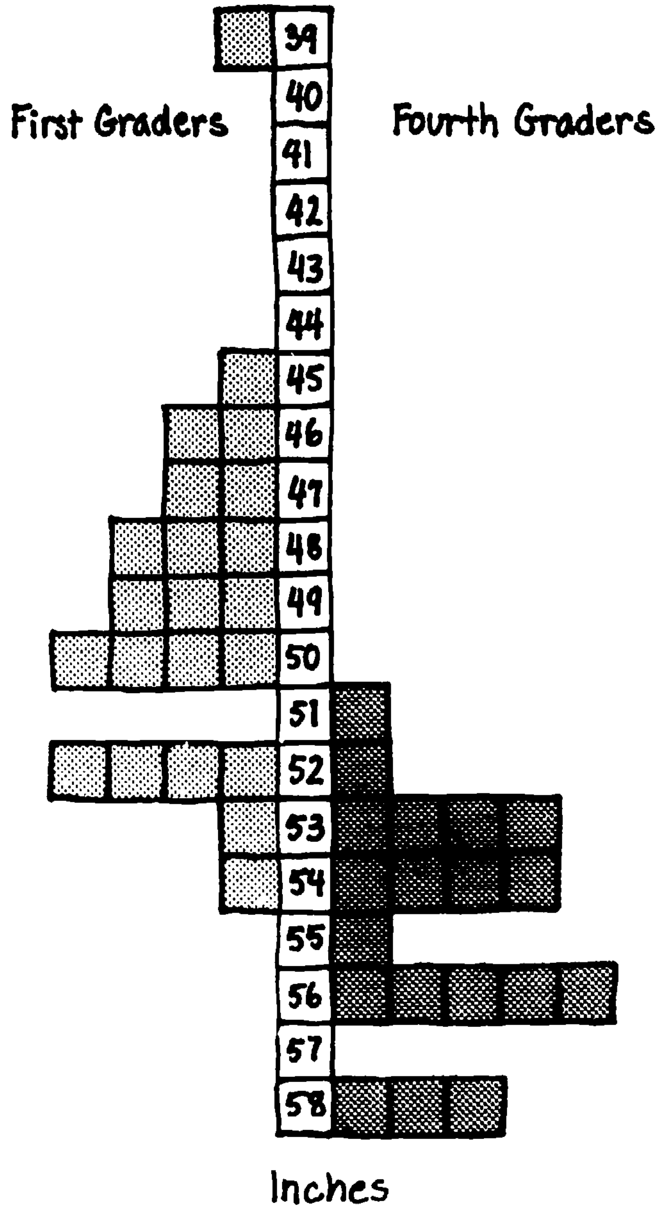
Like the final draft of an essay, a presentation graph is directed at a particular audience. Its purpose is to display the data so that they are organized, clear, and accessible. It supports what students have found in their analysis of the data by directing attention to important features of the data.

Constructing presentation graphs is an opportunity for students to be inventive. Comparing two sets of data is a situation with lots of room for creativity and seems to inspire students and teachers alike. For example, below is a graph that fourth grade students made to show the comparison of first graders' and fourth graders' heights. Their variation on a standard bar graph format provides a clear picture of the distribution of fourth and first graders' heights, showing where the two data sets do and do not overlap.



75

Another group invented the following way, also quite effective, to show the comparison between data sets.

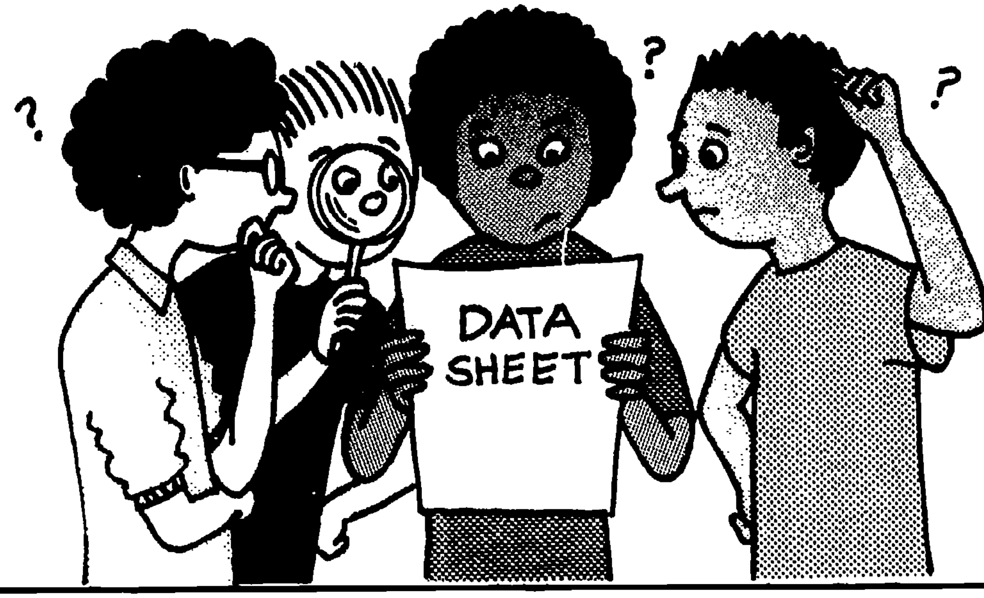


What is powerful about these graphs is that they clearly present what is important about the data. Some students become bogged down in coloring or decorating graphs in ways that obscure rather than illuminate the data. Color is a wonderful tool for revealing features of the data; in the first height graph shown here, students could have used two different colors rather than writing 1s and 4s to differentiate the two data sets. However, color and design for their own sake may detract from the clarity of the data themselves. Encourage students to design their graphs carefully to highlight important features of the data and to make the points they want to make.

If a computer is available, graphing software provides one option for students as they create their presentation graphs. Some students may enjoy and benefit from the clarity, neatness, and precision that graphing software provides. However, the software appropriate for the elementary grades limits students to straightforward representations of standard tables and graphs, such as bar graphs. The invented graphs shown here could not have been made using graphing software. If you use a computer, encourage a mixture of student-generated and computer-generated graphs in your classroom, and help students see the limitations as well as the advantages of graphs made on the computer. ■

76

LOOKING AT MYSTERY DATA



INVESTIGATION OVERVIEW

What happens

Students examine three sets of Mystery Data, each of which gives the measurements (length or height) of individuals in some group of living things. Students describe the data and construct a theory about what the data might represent. A fourth set of Mystery Data can be used for further group work or individual homework.

Keep this session playful. The emphasis is on fitting the theory to the data, not on finding the right answer. You can even encourage fantasy, as long as the fantasies fit the data!

The activities take one class session of about 45 minutes, with an optional second session for library research.

What to plan ahead of time

- Become familiar with the Mystery Data sets yourself. In order to understand the task facing your students, you might try solving the mysteries of Student Sheets 2–5 before you read about them in the Teacher Note, *About the mystery data.*
- Duplicate Student Sheets 2, 3, and 4 (pages 70–72) for the class investigation. Duplicate Student Sheet 5 (page 73) for homework or an extension of this investigation (optional).
- Provide measurement tools—rulers, yardsticks, metersticks, tape measures—for each group of students.

☛ Tell the students to save Student Sheet 2 for use in the next investigation, *Finding landmarks in the data.*

Important mathematical ideas

Describing the shape of the data. Continue to use the approach to describing data that was introduced in the first investigation, *How many raisins are in a box?*

Summarizing what is typical of the data. In order to develop a theory about the data, students summarize the characteristics of the group represented by the data (e.g., "Most of the things are between 80 and 84 inches long").

Examining the relationship between grouped and ungrouped data. Each set of Mystery Data is presented in two ways: a bar graph of all the individual values (ungrouped data), and a line plot showing a count of how often each value occurs (grouped data). Both representations are commonly used to convey information in newspapers and elsewhere; students should be aware of the differences between them.

Visualizing and estimating lengths and heights. By using their own height and other familiar measurements as benchmarks, students can estimate what other lengths would look like. For example, if we know Sandy is 4 feet tall, then we can use her as a benchmark for visualizing 8 feet. ■

SESSION ACTIVITIES

Describing data: Mystery Data A

Give out copies of Student Sheet 2, *Mystery Data A.*

This page shows the measurements of 24 individual members of some group of living things. Your task is to develop a theory about what this group might be by looking carefully at the data. In the graphs on the left, the bars show the measurement of each thing that was measured. On the right is a line plot that shows the same data in a different way. Don't tell me any of your theories yet about what this group of things might be. First, what can you say about these data? How would you describe them?

Allow time for students to describe the data as completely as possible. In the discussion, encourage them to use what they know about their own heights and the heights of the first graders they measured to estimate and visualize how tall or long the mystery beings might be. See the Dialogue Box, *Visualizing measurement data* (page 45).

Now that we have a good sense of how tall or long these mystery beings are, let's brainstorm about what they could be. I'm interested in your theories, but I'm also interested in your reasons for your theories and whether you agree or disagree with theories other people come up with.

Record students' theories in a place where they can remain posted.

Keep the answer to this first set a mystery for now. Give out Student Sheets 3 and 4, *Mystery Data B* and *C.* Be sure to have measurement tools available for all the groups.

These Mystery Data sets represent two more groups of living things. Your job is to describe the data carefully and then figure out what these things could possibly be. It may be that no one will discover the right answer—and that's OK. Your theory can be strange and unusual or ordinary and down-to-earth, as long as you can make a good argument that your theory really fits the data.

Developing theories: Mystery Data B and C

Working in small groups, students describe each data set and develop theories about what living things the data could represent.

Developing theories: Researching data collected by others (optional)

If you have access to a school or public library, students can do some research to confirm or disprove their theories. Such research could also be a homework assignment.

Publishing findings

For each set of Mystery Data, students write "the story of the data," including a description of the data, how the measurements compare to others they know about (e.g., their own heights), their theories about what the data might represent, and reasons for their theories. You might make a display showing the data and student stories.

Final discussion: The mysteries revealed

Emphasize that there was no way to know for sure what these Mystery Data represent. The task was to develop theories that really matched the data.

Now we're going to try to get closer to what these things really are. To narrow down the possibilities, I'm going to give you some clues.

Discuss one data set at a time. Using the clues, students can eliminate some of the theories they have developed and feel more sure about others. You may want to devise your own clues, or you might use these:

- Mystery Data A: They were all in Seattle at the same time on February 12, 1989.
- Mystery Data B: They are not human. They live in zoos and museums in the United States.
- Mystery Data C: These data were collected in 1989. If we measured the same things now, we would get different results.

See the Teacher Note, *About the Mystery Data* (page 46), to get further ideas for clues. You could play a game of "Twenty Questions" to reveal the mysteries for each data set, or you might give more clues until the students uncover the mystery. You may want to reveal the solution dramatically, having the answer in a sealed envelope that a student can open and read to the class.

Extensions

Student Sheet 5, *Mystery Data D*, can be used for homework or small-group work. ■

DIALOGUE BOX
Visualizing measurement data

So you've told me a lot about the shape of these data. But what about these measurements? About how long or tall are these things?

JORGE: They're taller than we are.

How do you know that?

BEN: Well, I'm 52 inches tall and the shortest one is 72, so that's a lot taller than me.

The shortest is 72 inches? About how tall is that? Who has an estimate?

MARY: It's as tall as you.

ARPIE: It's as tall as my Dad.

KIM: It's about up to the doorway.

Who has a way we can get a good estimate of how big 72 inches is?

JESSIE: We could measure it.

Yes, we could, and we can do that in a minute. But is there any way we can get a good estimate just by imagining about how big 72 inches is?

CARMEN: I have a way. If Ben is 52 inches, then 72 inches is 20 inches taller than him. He can stand up, and then we can just show about 20 inches taller.

How would you show about 20 inches taller? Who has an idea?

RAFAEL: That's close to two feet, so we could put two rulers in a line over his head and that would be close.

JANE: Also, 72 inches is the same as 6 feet.

The same as 6 feet. How would that help?

JANE: Well, if we knew something that was 6 feet, then we could tell 72 inches.

Anyone know what might be 6 feet?

CHRIS: If we knew how tall the doorway is, that might be close.

EDDIE: Six feet is two yardsticks.

KAREN: My dad's exactly 6 feet, and there's still about that much space *[showing with hands]* between him and the doorway.

So the smallest thing is about as tall as your dad, about up to here? What about the biggest—how big is 88 inches? ■

✎ TEACHER NOTE
About the mystery data

To the teacher: STOP! Don't look!

☛ Before you read the following descriptions of the Mystery Data sets (Student Sheets 2–5) and ruin the surprise, don't you want to try solving the mysteries yourself?

Mystery Data sets A, B, C, and D each contain measurements of individuals in some real group of living things. The living things in each set are closely related in some way. Each group is something about which data might be logically collected—not some random assortment of disparate things, such as a bear, a camel, a moose, and a zebra. The students' job is to develop a theory about what each set of Mystery Data might represent.

The first three data sets are displayed in two types of graphs: (1) a bar graph of the ungrouped data, in which each individual value is shown as a single bar; and (2) a line plot of the data, with the values grouped to show the frequency with which each value occurs in the data set. Mystery Data sets A, B, and C are to be used for small-group work in the investigation, *Looking at mystery data.* Mystery Data D is an optional set that you might want to use for homework or further group work at the end of the *Looking at mystery data* investigation.

Mystery Data A

These values are the heights of the 24 basketball players who were selected to play in the 1989 NBA All-Star Game, played on February 12, 1989 in Seattle, Washington. The West team won 143 to 134.

The 1989 All-Star players are listed below with their heights, positions, and teams. The first five players in each list were the starters.

West Team	Height	Position	Team
Alex English	80" (6' 8")	forward	Denver Nuggets
Karl Malone	81" (6' 9")	forward	Utah Jazz
Akeem Olajuwon	84" (7')	center	Houston Rockets
Dale Ellis	77" (6' 5")	guard	Seattle SuperSonics
John Stockton	74" (6' 2")	guard	Utah Jazz
Kareem Abdul-Jabbar	86" (7' 2")	center	Los Angeles Lakers
Clyde Drexler	79" (6' 7")	guard	Portland Trail Blazers
Tom Chambers	82" (6' 10")	forward	Phoenix Suns
Chris Mullin	79" (6' 7")	forward	Golden State Warriors
James Worthy	81" (6' 9")	forward	Los Angeles Lakers
Mark Eaton	88" (7' 4")	center	Utah Jazz
Kevin Duckworth	82" (6' 10")	center	Portland Trail Blazers
East Team			
Charles Barkley	78" (6' 6")	forward	Philadelphia 76ers
Dominique Wilkins	79" (6' 7")	forward	Atlanta Hawks
Moses Malone	83" (6' 11")	center	Atlanta Hawks
Michael Jordan	78" (6' 6")	guard	Chicago Bulls
Isiah Thomas	73" (6' 1")	guard	Detroit Pistons
Patrick Ewing	84" (7')	center	New York Knicks
Pat Cummings	80" (6' 8")	forward	New York Knicks
Larry Nance	82" (6' 10")	forward	Cleveland Cavaliers
Mark Price	72" (6')	guard	Cleveland Cavaliers
Mark Jackson	76" (6' 4")	guard	New York Knicks
Pat Daugherty	84" (7')	center	Cleveland Cavaliers
Kevin McHale	82" (6' 10")	forward	Boston Celtics

(Source: *Boston Globe* sports department.)

Mystery Data B

These values are the lengths of 18 boa constrictors living in various museums or zoos in the United States. Their names, lengths, and locations are as follows:

Name	Length		Location
Shannon	116"	(9'8")	Boston Museum of Science
Tony	86"	(7'2")	Boston Museum of Science
Bambi	94"	(7'10")	Boston Museum of Science
Bob	54"	(4'6")	Boston Museum of Science
Tiger	79"	(6'7")	Boston Museum of Science
Saulette	108"	(9')	Franklin Park Children's Zoo
Jake	114"	(9'6")	Franklin Park Children's Zoo
Bella	84"	(7')	Worcester Science Center
Floyd	72"	(6')	Worcester Science Center
Boa	54"	(4'6")	Worcester Science Center
Lady	96"	(8')	Earlham College
Sleeper	72"	(6')	Boston University

Malcolm	80"	(6'8")	Science Museum of Connecticut
Godzilla	93"	(7'9")	Science Museum of Connecticut
Alexis	72"	(6')	Science Museum of Connecticut
Julius	64"	(5'4")	Science Museum of Connecticut
(unnamed)	63"	(5'3")	Busch Gardens
(unnamed)	78"	(6'6")	Busch Gardens

(Source: Boston Museum of Science; Franklin Park Children's Zoo, Boston; Worcester Science Center, Worcester, MA; Earlham College, Richmond, IN; Boston University; Science Museum of Connecticut, Hartford, CT; and Busch Gardens, Tampa, FL. Lengths given by the Worcester Science Center and by Boston University are estimates.)

Mystery Data C

These values are the lengths at birth of a group of 14 babies born at Mount Auburn Hospital in Cambridge, Massachusetts in late August, 1989.

(Source: Community Relations Department, Mount Auburn Hospital, Cambridge, MA.)

Mystery Data D

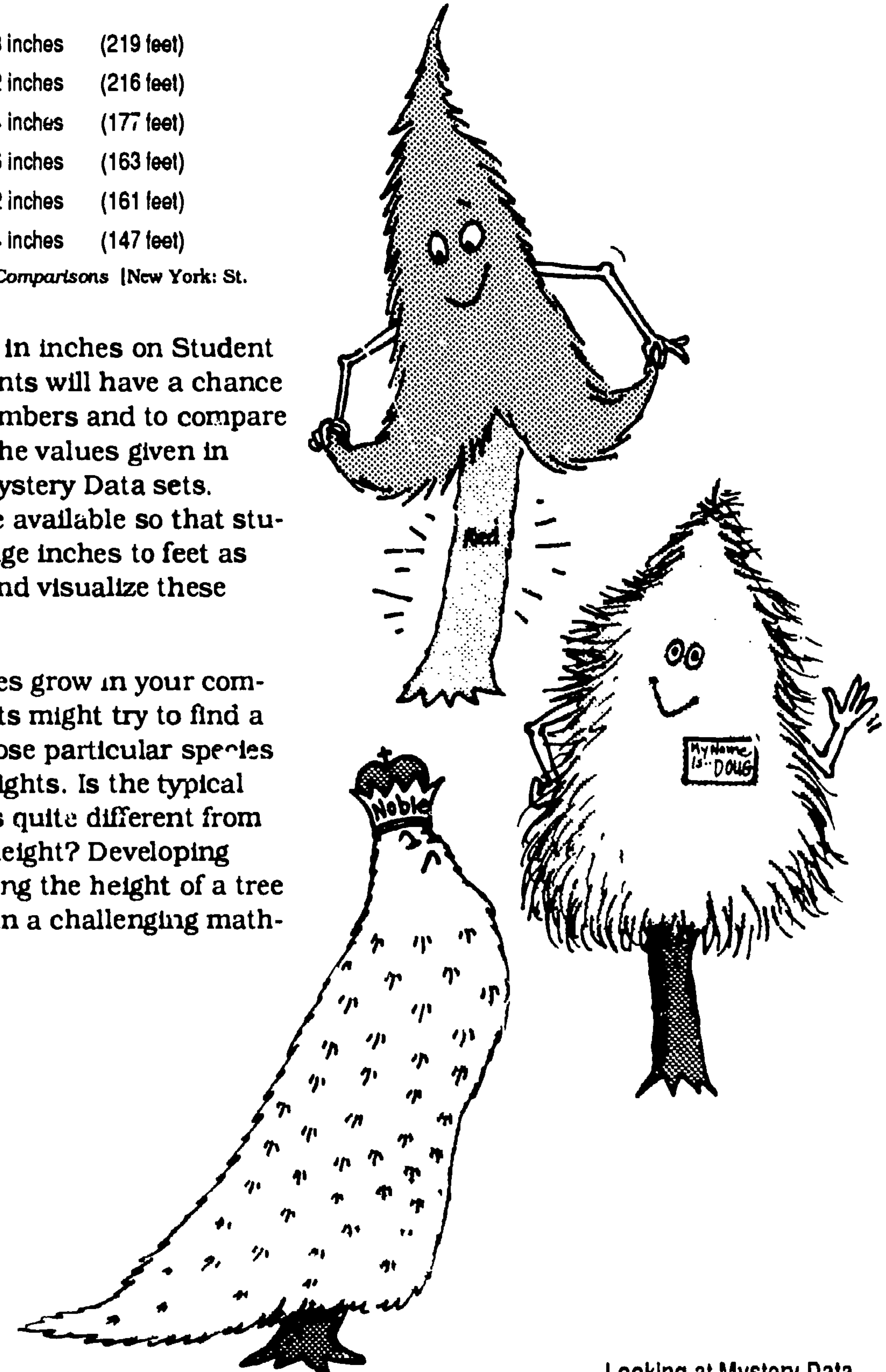
These values are the heights of 11 trees, each of which reached a record-breaking height for its species. (We might consider these the "Tree All-Stars.") All 11 trees were found growing in the United States. Their species and heights are as follows:

Name	Height	
California redwood	4392 inches	(366 feet)
Douglas fir	3624 inches	(302 feet)
Noble fir	3336 inches	(278 feet)
Giant sequoia	3264 inches	(272 feet)
Ponderosa pine	2676 inches	(223 feet)
Cedar	2628 inches	(219 feet)
Sitka spruce	2592 inches	(216 feet)
Western larch	2124 inches	(177 feet)
Hemlock	1956 inches	(163 feet)
Beech	1932 inches	(161 feet)
Black cottonwood	1764 inches	(147 feet)

(Source: The Diagram Group, *Comparisons* [New York: St. Martin's Press, 1980], p. 59.)

The heights are given in inches on Student Sheet 5 so that students will have a chance to work with large numbers and to compare these numbers with the values given in inches in the other Mystery Data sets. Calculators should be available so that students can easily change inches to feet as they try to estimate and visualize these heights.

Do any of these species grow in your community? If so, students might try to find a number of trees of those particular species and estimate their heights. Is the typical height for this species quite different from the record-breaking height? Developing strategies for estimating the height of a tree will involve students in a challenging mathematical problem. ■



FINDING LANDMARKS IN THE DATA

INVESTIGATION OVERVIEW

What happens

Students are introduced to the *median* as a formal measure that statisticians use to summarize a set of data. They find the median in a number of concrete ways, such as lining up by height to find the median height of their class, and they find the median of data they collected in previous investigations. They use the median as one landmark to compare data sets: their heights with the heights of basketball players, the weights of a group of cats with those of a group of lions, and the number of cavities in their own class with those in a comparison group of elementary students. They learn that the median must be viewed in the context of other important features of the data, such as its range and any clusters. The *mode* is introduced and explored in a data set about pets.

The activities take three sessions of about 45 minutes each. The first session is likely to take a little longer than the others.

What to plan ahead of time

For Session 1:

- Duplicate and cut apart the '89 All-Star cards (pages 80–82), with the names and heights of the players in the 1989 All-Star basketball game (Only one set is needed.)
- Ask students to find their copies of Student Sheet 2 (the All-Star players' height graphs, saved from the preceding investigation, *Looking at mystery data.)*
- Bring out the class data saved from the previous investigations about raisins, family size, and breath-holding, and the empty boxes saved from the first investigation, *How many raisins are in a box?*
- Provide Unifix cubes, counting chips, base ten blocks, or other concrete materials for making models of the data.

For Session 2:

- Duplicate Student Sheets 6 and 7 (pages 74–75) for each group.

For Session 3:

- Duplicate Student Sheets 8 and 9 (pages 76–77) for each group.
- Before this session, make sure students know how many cavities they have had. They may need to ask parents or even call their dentists.
- Have calculators available.

Important mathematical ideas

Using the median as one landmark to describe a set of data and for comparing one data set to another. Using the median requires an understanding of what it describes; that is, the median is the value that divides the data set in half. Half of the data are less than or equal to the median value; half the data are equal to or greater than the median value. Background information is provided in the Teacher Note, *Finding and using the median.*

Understanding that the median is only one landmark in the data. While the median gives us some information about the data set, it does not necessarily tell us everything we should know about the data. Other features, including the range of the data and the way they are clustered or spread out, provide a context for understanding the median, as discussed in the Teacher Note, *Finding and using the median.*

Finding the median in a set of data stretched out in a line (e.g., when the students line up in order by height).

Finding the median in a set of data that is grouped by frequency (e.g., on a line plot or graph).

Describing categorical data, which has characteristics different from sequential numerical data. A graph of categorical data, such as a graph showing the kinds of pets people have, simply shows the number of items in each category (dogs, cats, fish, and so on). These data do not have a median, a range, or clusters. The *mode* is sometimes a useful landmark in this case—if there is a clear mode that seems to show something interesting about the data. This landmark is described in the Teacher Note, *Finding and using the mode.* ■

SESSION 1 ACTIVITIES

Considering the problem: Introducing a special landmark, the median

In the investigation about heights, when you decided how much taller a fourth (fifth, sixth) grader is than a first grader, you figured out ways to decide how tall a typical first grader or a typical fourth grader is. How did you pick one number or a few numbers to describe a group with so many different heights?

Allow students to briefly describe some of the methods they used.

I'm going to show you a method that statisticians use when they are describing data. They find a number called the median. Median means "middle." The median is the exact middle of the data when all the data are put in order.

If students have already come up with this method during previous investigations, point out that they invented this for themselves, just the way a mathematician once invented it. See the Teacher Note, *Finding and using the median* (page 54).

Describing data: Finding the median height for this class

What do you think the median height would be for our class?

Students offer suggestions about who the middle person might be and methods they could use to be sure of their answer. To check, they then line up in order of height. To find the middle, students might sit down in pairs—one from the short end, one from the tall end—until only one or two students are left.

If students suggested other strategies for finding the middle student, try one or two of these.

If you have an odd number of students, the median is the height of the middle student. If you have an even number of students, you may have two students left in the middle. Ask students how they might decide what the middle value is. Explain that when the median was invented, statisticians decided that the "official" method would be to find the value exactly between the two middle pieces of data, when there is an even number of values.

Is the median value, the middle height, about what we said was a typical height for our class?

Students compare the median with the ways they summarized the height data for the class. Usually the median will be a pretty good indicator of typical height for the class.

The median gives us some important information—that half the data is above and half the data is below that number. Our median height is [58] inches. But what doesn't the median tell you about the data?

The students brainstorm all the things the median does *not* show about their height. They might think about it this way: If another class only knew their median height, what *wouldn't* they know? For example, they wouldn't know the range of the data, and they wouldn't know how spread out or close together the data are. They wouldn't know where the clumps of data are, or if there are any really unusual values in the data. Could there be another class with the same median height as this class, which looked quite different when the students lined up in order? Let students try to describe such a situation. Guide them toward the conclusion that the median provides one important landmark, but doesn't give a complete picture.

Describing data: What's the median height of the '89 All-Star team?

Distribute the '89 All-Star cards randomly, one to a student, so the order of students' heights does *not* correspond to the order of the heights given on the cards.

Now you're going to line up in an order that will let us find the median height of the '89 All-Star basketball teams.

Students are to line up *according to the heights on their cards*, not according to their own heights. They may use the pairing-off method described above (or some other method suggested by students) to find the median height. Since there are 24 players on the All-Star teams, students will have to find the value exactly between the middle two players.

Interpreting the data: Using the median to compare our class and the All-Stars

So now you know about how to find a median. But what is a median good for? Why would anyone want to find a median?

Allow enough time for students to think hard about this. Referring to their copies of Student Sheet 2, the graphs of the All-Star players' heights, may help students see the median in the context of the whole data set during this discussion. See also the Dialogue Box, *What good is knowing the median?* (page 53).

One important use of the median is for comparing sets of data. For example, you know that the median height for our class is [58 inches] and the median height of the '89 All-Stars is 81-1/2 inches. How can we use this information to compare the heights of our class with the heights of the All-Stars?

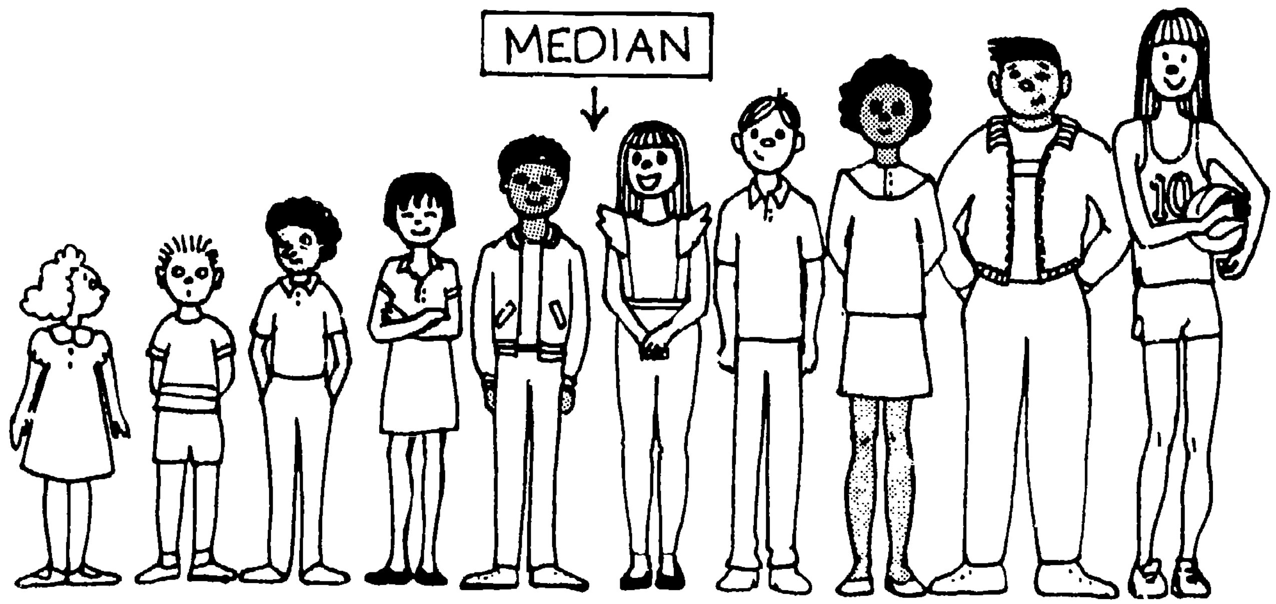
Take student suggestions. Remind students that this method is similar to some of the methods they devised when they compared first graders' heights with their own heights.

Describing data: Small group work on previously collected data sets

In small groups, students work on finding the median for the data they collected in previous investigations—the raisin data, the family-size data, and the breath-holding data. Each group starts with one data set. (More than one group can work on the same data.) Urge students to invent ways to find the middle for each data set in some concrete, physical way. Here are some possibilities:

- For the raisin data: Use the collection of raisin boxes you saved with the number of raisins written inside the lids. Line up the boxes themselves and find the middle box. Or, use concrete materials such as Unifix cubes, base ten blocks, or counting chips (as described below for the breath-holding data).
- For the family-size data: Use Unifix cubes to build a tower of cubes for each family in the class (e.g., a tower of 4 for Ellen, a tower of 6 for Ned, and so forth); line up the towers in order of height and find the middle tower.
- For the breath-holding data: It is harder to figure out a concrete way to show these data because the numbers are large. Students might write the numbers on scraps of paper or cards, arrange them in order, and find the middle one. Some groups have been able to build a concrete model by designating one color of cubes or chips to represent ten or twenty so the models aren't too unwieldy. Base ten blocks can be used, if students are familiar with them.

Once students have identified the median, they should make a graph or plot of the data (bar graph, stem-and-leaf plot, or line plot) and show where the median is on the graph. Although students have already plotted these data during previous investigations, they need to do it again now so that they begin to see how to find a median on a representation that does not show all the individual values lined up in a row. ■



93

94

DIALOGUE BOX
What good is knowing the median?

Now you've found the median height for our class and the median height of the All-Star team. So you know *how to find* a median. But what good is knowing the median? Why do you think statisticians or other scientists are interested in knowing the median of a set of data?

DAVE: It's the middle number.

Yes, it is the middle number; for example, the middle height in our class is 58 inches. But if you knew that the median height of some other class was 58 inches, too, what would you know about that other class?

SARAH: Well, if they lined up, the middle kid would be 58 inches tall.

Uh-huh, what else?

CATHY: I don't agree with Sarah, because it might not be the middle kid. It might be between the two kids in the middle.

OK, so you're both saying that 58 inches is the middle value either way. How does that help us know something about the height of the class?

IRENA: You know that half the kids are below 58 and half the kids are taller.

ALAN: They're like our class, about the same height.

They're like our class—what does that mean?

ALAN: The middle of their heights is 58 inches and so is ours.

What else could you say about whether they're like us or not?

BEN: They wouldn't have to be exactly alike. They could have some kids who were much taller than kids in our class.

How would that work?

BEN: The middle kid could still be 58 inches, but the kids taller than 58 inches could reach all the way up to 80 inches.

So the top half could be more spread out. What do other people think?

CARMEN: I don't think it would be that spread out. Fifth graders aren't 80 inches tall.

OTHERS: Yeah, that's silly. That's as tall as a basketball player.

So you're saying that you have some experience that tells you that another class with a median height of 58 inches wouldn't be as drastically different from ours as Ben was saying. What do you think, Ben?

BEN: They probably wouldn't be all the way up to 80 inches, but there still could be some kids who would be taller.

SUE: Yeah, or shorter.

Shorter? How would that work? . . .

[Later] . . .

It seems to me you've been saying that the median tells you some things about data but not everything. So, now, let's pretend again that there's another class with the same median height as we have. What would that tell us about how we're the same?

ANNIE: The middle is the same.

CHRIS: Well, it's kind of that we're clustered around the same height. It's, like, we're kind of the same, but not exactly.

Who can add something or say it differently?

CARMEN: Maybe the typical kid is around 58 inches in both classes, but there could still be things that are different. Like our tallest kid might be taller than their tallest kid.

KYLE: The middle is the same but the ends might be different.

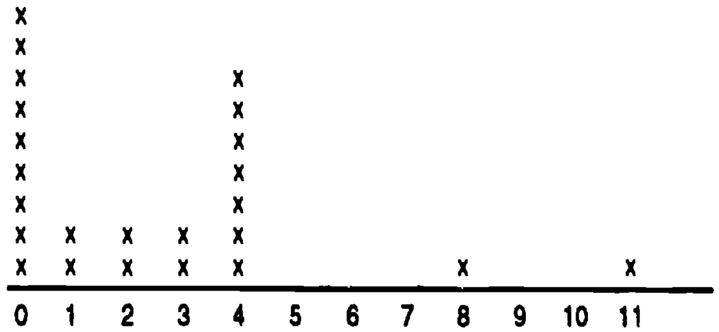
In this dialogue, the teacher attempts to move the students away from *how to find the median* to an explanation of what the median is good for—what it does and does not tell you about the data it represents. Students need to be able not only to *find* a median, but also to interpret one. Often a median will be stated with no surrounding context; for example, "The median age in the United States is 31.5 years." Understanding what the median tells you about the data as a whole depends in part on the prior knowledge of the context that you bring to interpreting this statistic. For example, students in this class have some knowledge about a reasonable range of fifth graders' heights, and this helps them interpret a given median height. ■

TEACHER NOTE
Finding and using the median

The median is an important landmark in a set of data. It is an *average* or a *measure of center* that helps summarize how the data are distributed. For example, the median age in the United States in 1985 was 31.5 years.* This statistic indicates that half the U.S. population was 31.5 years old or younger, while the remaining half of the population ranged in age from 31.5 years to the oldest living age. In other words, there are approximately as many people in the first three decades of life (0–30) as in the last, say, six decades of life (30–90). Notice that the median is *not* the middle of the range of the data; if the range of the data is from 0 to 90, the middle of this range would be 45 years. But the population is not spread symmetrically over this range. Just as many people are in the first third of the range as in the last two-thirds, so the median—the value that equally divides the data set—is at age 30 rather than age 45.

Another example, which the students explore in the investigation *Finding landmarks in the data*, is a data set showing the number of lifetime cavities of twenty-four 9-to-12-year-olds. These data are displayed in the following line plot.

* Data from U. S. Bureau of the Census, Current Population Reports, Special Studies Series P-23, No. 150, *Population Profile of the United States: 1984/85*, U. S. Government Printing Office, Washington, DC, 1987.

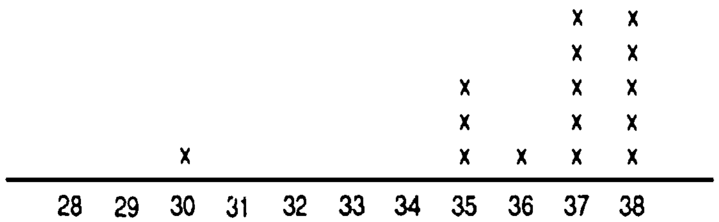


(Source: Dr. George W. McEachern III, DMD, Cambridge, MA.)

While the middle of the range is 5.5 (halfway between 0 and 11), the median is 2. Knowing that the range is 0 to 11 and that the median is 2 tells us that there are as many children in the group with 0 to 2 cavities as there are children with from 2 to 11 cavities.

The median is the midpoint of the data set. If all the pieces of data are lined up in order, and one person counts from one end while another person counts from the other end, the value where they meet is the median value. If there is an odd number of pieces of data, the median is the value of the middle piece. If there is an even number of pieces of data, the median is the value midway between the two middle pieces.

Here's a line plot showing the number of raisins in 15 boxes, similar to those made in the first investigation in this unit.



To find the median, imagine that we take all the values from the line plot and stretch them out in order:

30 35 35 35 36 37 37 37 37 37 38 38 38 38 38
↑
median = 37

The middle value is 37, so the median is 37 for this set of data.

If the data set had contained one more box of 35 raisins, it would have looked like this:

30 35 35 35 35 36 37 37 37 37 37 38 38 38 38 38
↑
median = 37

In this case, the median is between the two middle values; however, since the two middle values are the same, the median is still 37.

When the two middle values are *not* the same, as in the following data sets, the median is the value midway between the two middle numbers:

32 33 33 34 35 35 35 35 36 37 37 37 37 37 38 38
↑
median = 35.5

32 33 33 34 35 35 35 35 37 37 37 37 37 37 38 38
↑
median = 36

The following set of data (a survey of pets done in a classroom in California) does not have a median:

Dog	✔✔✔✔✔✔✔✔✔✔✔✔✔✔
Cat	✔✔✔✔✔✔✔✔✔✔✔
Fish	✔✔✔✔✔✔
Bird	✔✔✔✔
Hamster	✔✔✔
Mouse	✔✔✔
Horse	✔✔✔
Turtle	✔✔
Rabbit	✔
Guinea pig	✔
Snake	✔

While we can order these categories in different ways (in alphabetical order, from most to least in each category, or by size or weight), there is no intrinsic order to the items in the data set. When we line up the raisin data in order, we know that all the 35s always come after all the 34s and before the 36s. If we are finding the median height of students in a class, the students know how to line up according to height; 58 inches is always taller than 57 inches. But *fish* does not necessarily come before or after *dog*. If we had pictures of each animal in the pet data, lining up the pictures and finding the middle one would not tell us anything about the data set.

The median provides one landmark in the data, but it does not necessarily reveal all the important features of the data. For example, the fact that the median age in the United States is 31.5 years gives us some information about the shape of these data, especially because we bring to this piece of information some knowledge and experience about the context. That is, we know that ages range from birth to somewhere in the nineties or the low hundreds. We know that very few people reach ages beyond 90. So we can imagine to some extent what the age data look like. However, we still do not know, for example, what proportion of the 30-and-over group is over 65, or how many of the 0–30 group are in their teens. Experience with and knowledge of the context will help us interpret a statistic such as the median. However, for a data set to which we bring less of our own knowledge, the median or any other average, taken by itself, may not illuminate important aspects of the data. ■

SESSION 2 ACTIVITIES

Interpreting data: A tricky data set

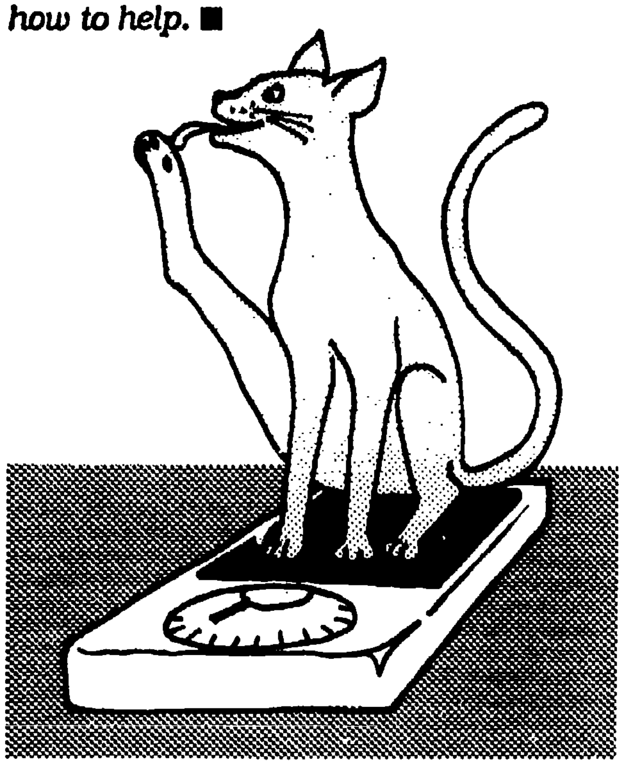
We've been talking about landmarks in the data and how to use them. You've used the median, the range, clumps, clusters, groups, intervals, holes, spaces, outliers. Now we're going to look at a tricky set of data that doesn't have the same kinds of landmarks as the data we've been investigating. These data were collected in a survey of student pets. Work in your groups and write down five important things you would say to describe these data. See if you can figure out why I say it doesn't have the same kinds of landmarks.

Give out copies of Student Sheet 6, *Pet data.* (For the source of this data set and the others in this investigation, see the Teacher Note, *About the data on Student Sheets 6–9,* page 60). Circulate as the groups work. If students think they can find landmarks such as clusters, outliers, or a median in these data, discuss with them what they mean by that. See the Dialogue Box, *Common misconceptions about the median . . . and how to help* (page 56).

Students will notice the large number of fish in this data set. As you circulate, introduce the word *mode* as another landmark that tells what occurs most frequently in the data. See the Teacher Note, *Finding and using the mode* (page 57).

Describing data: Finding the median value on a graph

As groups finish their work with the pet data, give them Student Sheet 7, *Cat weights*, with data for 23 domestic cats. Encourage them to organize the data in a graph or plot, write down as many important landmarks as they can see, then write a brief description (based on these landmarks) of the weights of the cats. Ask each group to find the median (if they don't do so on their own) and decide if the median is a good description of the typical cat in this group. Be on the alert for difficulties students have in finding the median from a graph, as described in the Dialogue Box, *Common misconceptions about the median . . . and how to help.* ■



DIALOGUE BOX

Common misconceptions about the median . . . and how to help

While the idea of "middle" seems straightforward and easily understood by students, the median can nevertheless be a difficult and complicated concept. As students encounter new examples, they may become confused about which "middle" they are trying to find. For example, when some students encounter the pet data, they think they can find the "median pet."

ARRIE: The median is guinea pig.

How does that work?

ARRIE: Well, there are 23 pets, so the middle bar on the graph would be the twelfth one. So we counted to the middle, like this, and the middle bar is guinea pigs.

I see how you found the middle of the graph, but I don't quite get it. Is it like when we lined up by height, and Cathy was our middle person, so the median height was 58 inches? Is guinea pig the middle pet?

KAREN: Yeah. It's in the middle here.

What does a median tell you about the data—for example, about the class heights when 58 inches was the median?

KAREN: That 58 is right in the middle. If we line up in order, then the middle person is 58 inches.

Uh-huh. Half of you were below the median and half were taller. So you had to put them in order? Could we line up in order according to our pets? If I have a cat and you have a dog and Ricardo has a fish, how would we line up?

KAREN: I don't know.

JOEY: We could do it in alphabetical order.

ARRIE: Or by size. The fish is the smallest and then the cat and then the dog.

Yes, there are several different orders. So if someone said "The median is fish," I wouldn't know for sure what came before fish and what came after. I don't even know what other animals there might be. It's different with the heights. When you say the median is 54 inches, I know what's below 54 and what's above 54 for sure. Try a quick sketch graph of how the pet graph would look in a different order; then see if you can come up with reasons why you think the pet data does or doesn't have a median. I'll come back around when I've checked with the other groups.

Work with the pet data will help surface this common misconception—that the median can be used for categorical data. In categorical data, there is no intrinsic order to the values of the data (i.e., *fish* does not come before or after *dogs*) and you can't tell for sure from one piece of data what other values are in the set (e.g., having dogs in the data does not tell us whether there are any

(Dialogue Box continued)

fish at all). One strategy some teachers have used to help students see that there is no unique order to these data, and therefore no "middle," is to have them show the data in different orders: alphabetical, in order by size of pet, and so forth.

Other misconceptions may surface when students develop a graph of cat weights (Student Sheet 7). For example, they may say that the median is the middle of the horizontal axis, rather than the middle piece of data. Sometimes students find the "middle" without putting the data in order first. Perhaps the most common student misconception is that the median is the middle of the *range* of the data.

ROBERTO: We said the typical cat is 11-1/2 pounds because 11-1/2 is exactly in the middle.

How do you know that's the middle of the cat weights?

ROBERTO: Because the smallest is 5 and the fattest is 18 pounds, so they're 13 apart.

ALAN: And 6-1/2 is half of 13, so if you add 6-1/2 to 5, that's 11-1/2. 11-1/2 is the exact middle between 5 and 18.

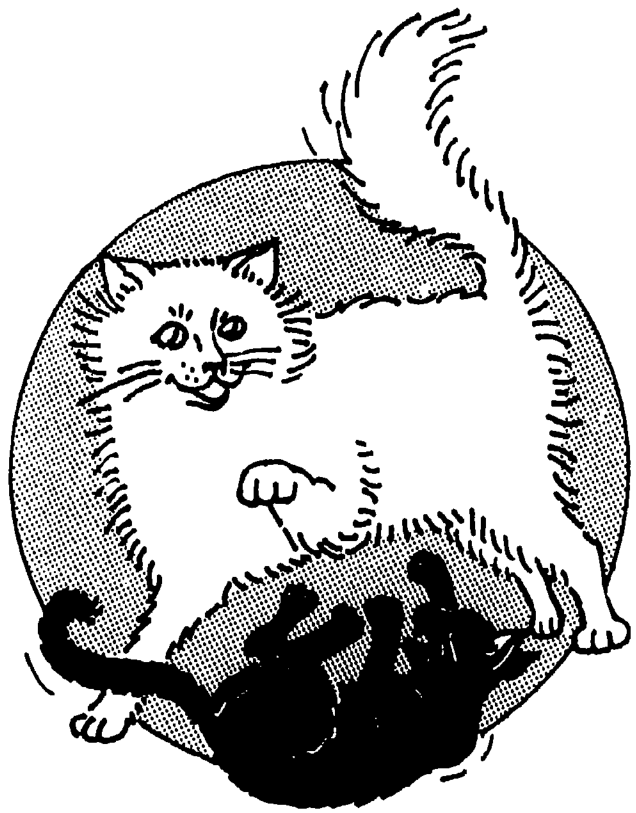
How would that work if we actually had all 23 cats here and we lined them up in order of weight?

ALAN: Um, well, the middle cat would be 11-1/2 pounds.

ROBERTO: Oh . . . no, it wouldn't . . . I don't think we found exactly the middle cat.

Find some way to show how you'd line up all the cats according to weight and which would be the middle cat exactly. I'll check back with you in a few minutes.

It may take repeated experiences with these ideas for students to develop a firm grasp of the concept of median. ■



TEACHER NOTE
Finding and using the mode

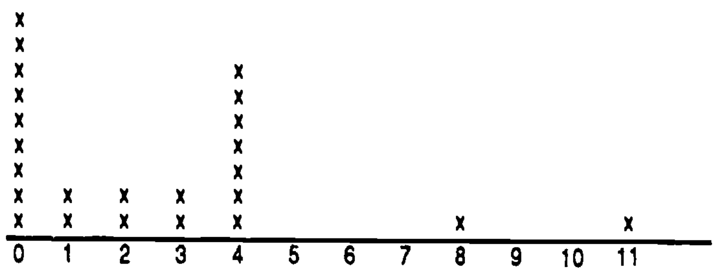
The mode is the value or values that appear most frequently in a data set. Students are already familiar with this idea, the "most." In the primary grades, they have probably had many experiences with simple graphs in which they compared favorite ice cream flavors, the kinds of transportation they used to get to school, and other data about themselves that were arranged in categories. We say that data sets of this type contain *categorical data*. Categorical data are classified (vanilla, chocolate; bicycle, bus, walk), and the number of items in each category can then be used as a basis for comparing categories (e.g., "More people walk than ride the bus to school"). Consider the categorical data shown in the table below, a portion of the results of a survey of class pets in 551 classrooms (which students explore in a graph on Student Sheet 6).

Fish	12,674
Cat	7,713
Dog	7,522
Bird	2,547
Rabbit	1,768
Hamster	1,347
Horse	1,166
Gerbil	696
Turtle	563

(Source: Technical Education Research Centers, Inc.; collected as part of the work of classrooms involved in the National Geographic Kids Network.)

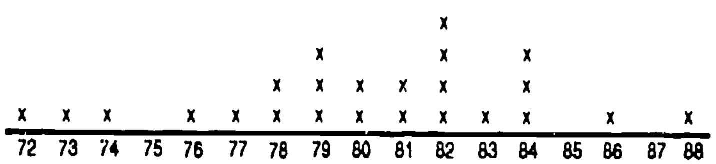
For categorical data, the mode is the only available summary statistic. The mode is the most frequent value. In these pet data, the mode is fish.

Numerical data may also have one or more modes. In this set of data from Student Sheet 9, showing the number of lifetime cavities among twenty-four 9- to 12-year-olds who visited a dentist in a two-week period, the mode is 0.



(Source: Dr. George W. McEachern III, DMD, Cambridge, MA.)

For the heights of the basketball players who played in the 1989 All-Star game, the mode is 82 inches.



(Source: *Boston Globe* sports department.)

The mode is not difficult to find. What is hard is deciding whether a mode is interesting. As a landmark, does it provide important information about the data? In the pet data, the mode is quite spectacular. There were over 12,000 fish in this student pet survey, about 5,000 more than the next closest category. However, simply to say, "The mode is fish" is only a first step. The real question is, "What is the significance of this mode?" Students who have studied this data set have raised interesting questions about the mode. Because people who have pet fish tend to have many in an aquarium, rather than a single fish, students have questioned whether this count is really comparable to the counts for dogs and cats. If the people with fish have even three fish each, the number of people with fish is quickly reduced to about 4,000. It may be that cats and dogs are actually more popular pets. If cat and dog are the modes (in these data, they are so close to each other and so much higher than the frequency of any other pet that we might consider them essentially equal), then what do these modes tell us about these data?

In the cavity data, the mode is also quite interesting. Of the 24 children, more than a third have no cavities. The mode of 0 is a striking landmark in the shape of these data. However, in the All-Star height data, the mode of 82 inches is not such an interesting landmark. The line plot of these data shows a central clump from about 78 to 84 inches, which includes about two-thirds of the players. The mode does not stand out as noteworthy; while four players are 82 inches tall, three are 79 inches and three others are 84 inches. In such a small data set, a value that occurs only slightly more frequently than others does not tell us much about the data. In this case, the range (from 72 to 88 inches) and the median (80.5 inches) provide a better summary of the data. ■

SESSION 3 ACTIVITIES

Considering the problem: Introducing another mystery data set

Give out copies of Student Sheet 8, *Another mystery*. Give the students time to say everything they can about what they notice about these data—clusters, clumps, gaps, the range, and so forth. Play "Twenty Questions" to establish what the data are (see page 60).

Interpreting data: Using the median to compare cats and lions

Working in pairs or small groups, students find the median of these data and then use the median lion and cat weights to answer the question, "About how many typical cats would it take to balance a typical lion?" Calculators should be available. If time permits, students can describe the strategies they used to figure out how many cats weigh as much as a lion.

Describing and interpreting data: Who has more cavities, the group on Student Sheet 9 or our class?

Give out copies of Student Sheet 9, *Cavity data*. Put the same data on the board in a line plot.

Here are some data that a dentist collected about the number of cavities some of his

9- to 12-year-old patients had had. What can you say about these data?

After students have described the data, continue:

We're going to find out how we compare to this group. Have we had more, fewer, or about the same number of cavities as a group?

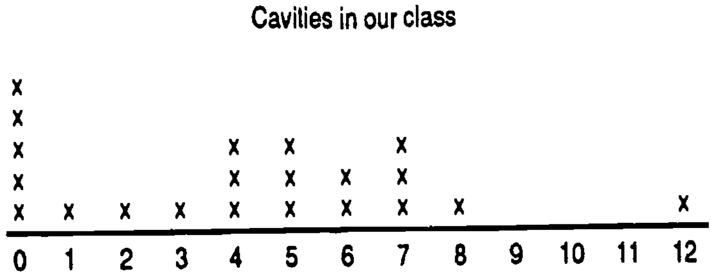
Collect data from each student, record it on the board, and have students add the class data to the line plot on the student sheet so that they can compare the shape of the two data sets. If some students do not know exactly how many cavities they have, they may have to estimate from memory.

How would you compare our class to this group? Does the median help? Does it tell you enough? What else would you want to say?

Encourage students to use the median for comparison, but also to look carefully at what the median shows and to add other information that will provide a context for the median. Now is the time to nip in the bud any tendency to give a rote response, simply because a formal measure (the median) has been introduced. You might hear, for example, "Our median is 4; their median is 2; the difference is 2." This kind of response can obscure interesting features of the data. A richer description might be as follows: "Our median is higher, so as a group we had more cavities; we had a lot more people with 5, 6, and 7 cavities, and we only had 5 people with no cavities." See the Dialogue Box, *How many cavities do we have?* ■

DIALOGUE BOX

How many cavities do we have?



JORGE: Our median is 4 and theirs is 2.

What do you think that tells us about comparing the number of cavities we have to the number they have?

JORGE: We had 2 more?

What do you mean, 2 more? Anyone have an idea?

SUE: Well, our middle person had 2 more cavities than their middle person.

ALICE: It's like on the average we had more cavities.

Can anyone add to that? What if you were reporting on this research about cavities?

KYLE: If we use Crest and they use Colgate, then we could say Colgate was better.

JOEY: Yeah, or maybe we don't go to the dentist as much as they do.

MARY: Or we eat more junk food.

So you have a few theories about why we have more cavities than they do. We don't know if those theories are correct, but you all seem to agree that typically we have more cavities as a group than the Massachusetts group. The median helps us see the difference. What else is the same or different about our data and their data that the median doesn't tell you?

KIM: Nobody had more than 12 cavities in both groups.

What do you think about that? Does that surprise you?

KIM: You might think that we'd have people who had lots more cavities than their top person, but we didn't.

DAVE: But even though we didn't have anyone higher than 12, they have more 1s, 2s, 3s and 4s than we do. We had more 5s, 6s, and 7s.

MARIA: And they had more people who didn't have any cavities.

JESSIE: Yeah, we only had 5 people with no cavities.

Do you think that's unusual for a group of kids your age? What do you think about that? ■

✎TEACHER NOTE
About the data on Student Sheets 6–9

The pet data on Student Sheet 6 were collected from a sample of 551 elementary school classrooms in May 1989. The classrooms were spread across 44 states, represented a variety of urban and rural locations, and also included classrooms in Puerto Rico, Japan, Canada, and one classroom in Moscow (USSR). (Technical Education Research Centers, Inc., collected the data as part of the work of classrooms involved in the National Geographic Kids Network.)

For the cat weight data on Student Sheet 7, the cats were unscientifically selected by the authors of this unit. They include Susan Jo's cat, Alexander, and Rebecca's cats, Strawberry and Lady Jane Grey. Other cats belong to colleagues and friends, and some data were collected at a local cat show. The cats were weighed and measured by the authors or by their owners.

The mystery data on Student Sheet 8 are the weights of lions in a number of U.S. zoos. Their names and weights are shown in the table that follows.

The four "nonames" were in a litter of 4-month-old cubs in the Miami Zoo. The two heavier cubs in the litter are males and the two lighter cubs are females.

Toza	460 pounds
K.C.	453 pounds
Firecracker	450 pounds
Charlie	432 pounds
Valentino	420 pounds
Linda	402 pounds
Mimi	400 pounds
Sunshine	378 pounds
Zike	350 pounds
Gilbert	325 pounds
Webster	325 pounds
Kimba	307 pounds
Gira	302 pounds
Alice	300 pounds
Rex	290 pounds
China	289 pounds
Hannah	280 pounds
Asha	240 pounds
Basra	212 pounds
Noname 1	105 pounds
Noname 2	100 pounds
Noname 3	45 pounds
Noname 4	41 pounds

(Source: Zoos in Atlanta, Cleveland, Little Rock, Memphis, Miami, The Bronx, Philadelphia, Rochester, San Antonio, and Washington, DC. Data collected in 1987.)

The cavity data on Student Sheet 9 come from the 9- to 12-year olds who visited a pediatric dentist in Cambridge, MA during about two weeks in August 1989. The dentist reports that urban children typically have more cavities than children from nonurban areas because they have easier access to junk food at the corner store.

0 cavities	1 9-year-old 4 10-year-olds 2 11-year-olds 2 12-year-olds
1 cavity	1 10-year-old 1 11-year-old
2 cavities	1 10-year-old 1 12-year-old
3 cavities	2 10-year-olds
4 cavities	4 11-year-olds 3 12-year-olds
8 cavities	1 11-year-old
11 cavities	1 12-year-old

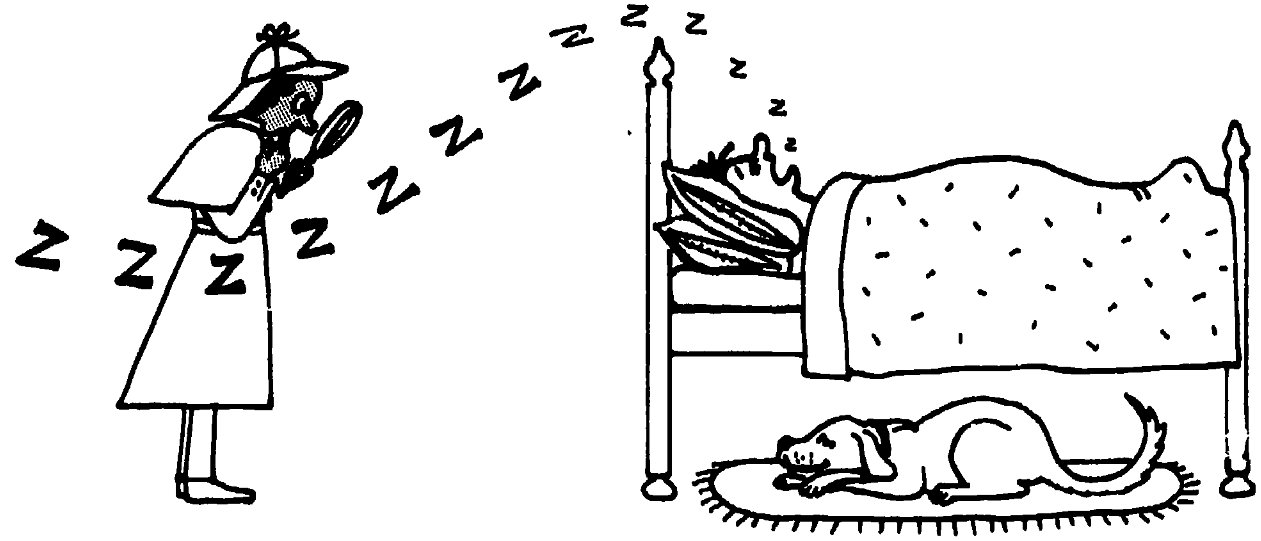
(Source: Dr. George W. McEachern III, DMD, Cambridge, MA)

The ages and number of cavities of the children in the data are shown in the table above. ■

Statistics: The Shape of the Data

PART 3
A project in data analysis

INVESTIGATING SLEEP



INVESTIGATION OVERVIEW

What happens

Students culminate their work in this unit by undertaking a more extended investigation in data analysis. After an initial whole-group activity and brainstorming session on the topic of sleep, the class breaks up into research teams of three or four students. Working together, the members of each team choose a question, collect and describe initial data, refine their question and data-collection procedures, collect additional data, organize and describe their data, and finally develop theories based on their data and publish their findings.

The activities take three or four sessions, not including time for data collection.

What to plan ahead of time

- Duplicate copies of Student Sheet 10 (page 78) for each student (Session 1).
- Provide materials for making sketch graphs and initial representations, including plain paper, squared paper, Unifix cubes, counting materials, and maybe stick-on notes for making and modifying rough draft graphs.
- Have calculators available.
- Provide materials for publishing findings, including a variety of paper, scissors, glue, and colored markers.

Important mathematical ideas

Undertaking a data analysis project from start to finish, from defining a question to publishing results.

Experiencing all the stages of a data analysis investigation. These stages, like the stages of writing, include brainstorming, first draft, revision, subsequent drafts, and publication. In particular, it is important for students to understand that the first data collected are usually not the final product in data analysis but, rather, indicate appropriate directions and refinements. The data analysis process is described in the Teacher Note, *Stages in data analysis: Learning from the process approach to writing.*

Choosing and refining a research question. In all data analysis, often the initial question must be changed or refined. This can occur for any number of reasons: the question was not clear; the initial data indicated that another question might be more interesting; the question did not get at the issue the researchers intended to explore; the researchers could not find a way of collecting good data to answer the question. How to help students struggle with these problems is discussed in the Teacher Note, *I wanna do it myself!*

Viewing the data in several different ways, using quick sketches and other representations to organize and display the data. It is easy for students to get bogged down in the mechanics of graph-making too early in the process. They must understand that representations used *during the process of analysis* should be simple and clear, but not necessarily neat and detailed. ■

SESSION ACTIVITIES

This project gives students the chance to carry out an extended investigation of data, from initial question posing through publication of results. Virtually every school child has already confronted issues about sleep: When do I have to go to bed? When do I have to get up? How much sleep do I need? Under what circumstances can I stay up late? These topics are often the basis for vigorous discussions between children and those responsible for their bedtimes. Because it is essential that students pursue questions and issues of particular interest to them within the general topic area, it is impossible to predict exactly what directions this investigation may take. The following outline suggests an appropriate process, but the particular ideas to be investigated will emerge from your students.

This extended investigation in data analysis is likely to take about four class sessions, not including the time needed for collection of data. Data collection can usually be done during non-class time, but some students may need your help arranging visits to other classrooms, if their investigation involves collecting data for a particular age group. Since the students will be working in research teams, much of the time will be spent in small-group activities. The following outline describes a whole-class activity and discussion to introduce the investigation, followed by four stages of the investigation that each research team must work through. The groups may proceed at different paces and will probably need periodic conferences with you about their progress.

Considering the problem: An introductory activity and discussion about sleep

Give out copies of Student Sheet 10, *Sleep data*, with information about the amount of sleep people need at different ages.

Here are some data collected by a group of scientists who do research about sleep. They show how much sleep people typically need at different ages. What do you notice about these data? . . . What else do you notice?

Encourage students to describe the data as completely as possible. Call attention to the data for their age group: Do they think it's accurate for their class?

Next, we're going to take a quick survey of how much sleep the people in this class typically get. How should we do that?

Students will need to consider what a typical night means. Should they choose last night? Are they considering a school night or a weekend night? Are they going to say how much sleep they are "supposed" to get, or will they figure out how much they actually do get? If someone usually turns over and goes back to sleep for half an hour after the alarm goes off, does that half hour count? Help students arrive at a consensus definition of a typical night's sleep.

Students then work together in pairs or groups of three to figure out the number of hours of sleep each person in the group gets. While students likely know their bedtimes and when they get up, they may never have thought about the actual number of hours they sleep.

Record all the students' data on a line plot displayed where everyone can see it.

What would you say about our data? How does our class data compare with research data about your age group?

After some discussion of the class data, continue:

We're going to be doing a study of sleep patterns in this school. In a few minutes, we're going to break into research groups so you can decide on particular questions you want to investigate. But before we do that, let's take a few minutes to talk about sleep.

Encourage students to discuss issues about sleep, giving personal examples that involve themselves, siblings, or friends. You might raise the following questions during this discussion: Do you think you get enough sleep? Do you think most people your age get enough sleep? Do you have arguments with your parents about sleep? What are the things that come up? How much sleep do you think you need? Do you think it's about the same for everyone your age? What about your younger or older brothers and sisters?

Allow this discussion to continue until you feel that students have raised most of the issues, as this will give them a good start on choosing a question in their research groups.

Let's brainstorm some questions that we might be able to investigate about sleep by collecting data ourselves. Questions might occur to you as you look at the research data, as you look at our own data on the line plot, or as you think about the discussion we just had. I'll list the questions as you come up with them.

At this point, list *all* suggested questions. Don't attempt to filter students' ideas because they are impractical. Part of their job as researchers will be to devise and refine a question for which they can easily collect data. See the Teacher Note, *I wanna do it myself!* (page 68).

Stage 1 of the investigation: Choosing a question

Divide the class into small research teams of three or four members who will work together for the duration of the investigation. Their first task is to meet as a team and choose a question they want to investigate. They might choose a question from the list already made by the class or, as they talk, they may think of something else that interests them. A question involving comparison (first vs. sixth graders, students in this school vs. another school, boys vs. girls) would provide an appropriate challenge for students who have done the other investigations in this unit.

As each team comes up with a question, they should have a brief conference with you about their idea. Your role is to help them refine their question, to encourage them to include some kind of comparison in their investigation, and to link them with resources for further refining their questions. Your role is *not* to simplify a question for the group prematurely, even if you see that they may have some trouble with their investigation. Revising their question further will be part of the task of the research team as they progress. See the Dialogue Box, *Helping students refine their questions* (page 67).

After they have checked in with you, each team should continue to work on refining their question and deciding what data they need to collect.

Here are some of the questions that student teams have selected:

- Do first graders get more sleep than students our age?
- Are there some students at each grade level who stay up much later than others?
- Do boys and girls have the same weekend bedtimes?
- Do the oldest students in the school get the least sleep?
- What's the typical weekend bedtime for each grade level?

Don't worry if some questions overlap.

Different perspectives on the same topic can add to the richness of what students learn.

Stage 2 of the investigation: Collecting, recording, organizing, describing, and interpreting preliminary data

Each team collects, records, organizes, and describes some initial data that they think w'll help answer their question. Emphasize to students that this data is preliminary, a "first try" to get some information to help them plan their study; they will collect more data later on. Suppose the question is, "Do boys and girls have the same weekend bedtimes?" That team could start by taking a survey in their own classroom to find each student's bedtime last Friday and Saturday. Once they collect their data, they organize it in some way, write a description of what they think the data show, and develop some theories about the patterns in their data. For example, one team investigated girls' and boys' Friday night bedtimes. They found out that the median bedtime for boys was an hour and a half later than the median bedtime for girls! They wanted to know why that was and they came up with several theories about their data: boys eat more sugar than girls so they are more active and stay up later; boys are more demanding and don't give in to their parents when they are told to go to bed; girls get up earlier, too, so they actually get the same amount of sleep.

Taking a survey of about 20 people (in their classroom, on the school bus, or among their friends) is a good way for students to collect their initial data. In Stage 3, they can collect data from a larger sample.

Making decisions about further investigation: A student-teacher conference

After each research team has completed a preliminary investigation, they again have a brief conference with you to help them decide what their final investigation will be. Some students may have found that there were problems with the way they formulated their question, so that people interpreted the question differently. Or, they may have found that younger students typically did not know the answer to the question they were asked. They may have discovered that the data they collected did not help them answer their question and they need a different kind of data. Or, they may have found that a more interesting question emerged from their data, one which they now want to pursue. They may have seen an interesting pattern in their data, as in the Friday-night bedtime example, but need to collect more data to see if the pattern really holds. The team that investigated boys' and girls' Friday bedtimes decided that they would also collect times students get up on Saturdays so they could compare the total amount of sleep for boys and girls. Encourage students to examine critically both the question they asked and the data they collected and to come up with new strategies for improving their research.

Stage 3 of the investigation: Collecting, organizing, and describing further data

Research teams now undertake their final investigation From this point, not all teams will go through exactly the same stages. Some teams may need to refine their questions and collect new data, while others are satisfied with their question but need additional data. As needed, the research teams will

- refine or change their question
- decide on their final data collection procedure
- decide how to record their data as they collect them
- collect and record their data
- organize and represent their data in several different ways
- write a description of their data

You may need to help coordinate some of the data collection, especially if it involves gathering information from other classes. Check in with each team periodically. After the data are collected, encourage students to represent their data in several different ways with sketch graphs or concrete materials (a line plot, bar graph, stem-and-leaf, Unifix cubes, and so forth) so they can see different aspects of the data.

Stage 4 of the investigation: Developing theories and publishing findings

The final task of the research team is to put their analysis into publishable form. This may be a written report, a class presentation, a display, or a letter to the principal or to parents about their findings. Publication of results, in whatever form, must include a statement of the research question, a presentation graph showing the data, a written description of the data, and the students' interpretation of their findings. Interpretation includes their hypotheses and theories about why the data look the way they do. Depending on their question, it could also include recommendations. For example, does it seem that about a third of the first graders are not getting enough sleep? Are too many students staying up late on Wednesday night because of a particular TV program? Do the researchers have any recommendations about these findings?

Your role is to ask questions, to encourage students to discuss their findings with each other, and to draw parallels between their work and the work of adult researchers who use statistics. All the difficulties students have encountered—defining or redefining their question, obtaining trustworthy data, and developing theories to explain their data—are the same problems encountered by statisticians and researchers in science, psychology, history, economics, and many other disciplines.

Conclude the project with a discussion in which students critique their investigation. You might ask: Were there any limitations to your study? Were there aspects of your research that did not satisfy you or that you might do differently next time? Are there now other questions you would ask? One team, for example, was unsure that the data from the youngest students in their study was reliable; they thought the students might have exaggerated their bedtimes. Since there seemed to be no other practical way of learning the bedtimes of these students, some team members thought a different question might have been better. One team member wondered if the young students' teacher might have cooperated in sending a survey home with each child, to get more reliable information from their parents. ■

“” DIALOGUE BOX
Helping students refine their questions

ARRIE: We're going to find out how much sleep a first grader gets.

How did you choose that idea? Why are you interested in that?

RAFAEL: Well, we were talking about the little kids who stay up really late.

EDDIE: Yeah, my friend's little sister stays up really late, even later than I go to bed.

JANE: And there's this kid Jonah from Ms. Weinberg's class who's always asleep on the bus going home.

EDDIE: I had to go to bed at 7 o'clock when I was in first grade.

RAFAEL: You never went to bed at 7 o'clock. You used to watch *Miami Vice*.

EDDIE: Yeah, but that was Friday. On the school nights I had to go to bed at 7.

Are you wondering whether first graders stay up later than you did at that age?

JANE: Yeah. We think that a lot of the little kids stay up too late.

That would make an interesting comparison—how late kids in fifth grade stayed up when they were in first grade compared to how late first graders stay up now. Have bedtimes changed since you were in first grade?

ARRIE: We could ask kids in our class and ask all the first graders.

JANE: Yeah, but the first graders can't all tell time. They probably don't know.

EDDIE: And some of them might just say they go to bed at 9 o'clock because they're bragging, but it might not really be true.

Maybe the first grade teachers would have some ideas about how you could find out the real times. I know that Ms. Cannelli has talked to me about how late some of her first graders go to bed on school nights, so I bet she'd be interested in talking to you about your idea. She might have information that would help you figure out exactly what to ask. ■

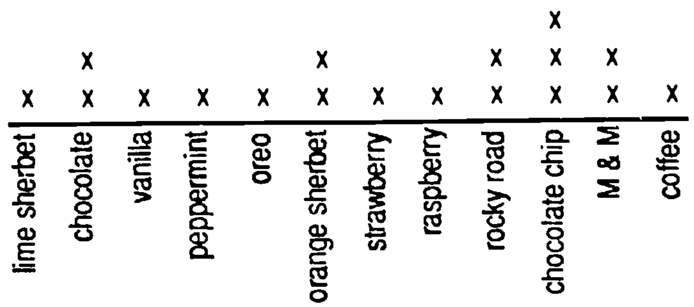
TEACHER NOTE
I wanna do it myself!

When upper elementary students start their own projects in data collection and analysis, they often select questions that you immediately recognize as problematic. It's very tempting to simplify their questions for them, to save them from wrestling with messy data or overwhelming amounts of information. Isn't it important to keep them from being overwhelmed? Shouldn't a teacher help simplify their questions in advance? The answer is no! Their final projects may solicit messy data, but wrestling with messy data is an important part of the analytic process.

Students must struggle with these issues themselves, but teachers have an important role in the process. It is vital that you encourage students to take plenty of time to think about their research questions. What data would help them answer their questions? How can they get those data? After students have collected their data, they must—with your support—carefully consider their results. Are they reasonable? Can they see any patterns in the data? Are there other avenues to take? Messy data lie at the heart of the process—in fact, most real data are messy. Tidying them up too early will not help students in the long run.

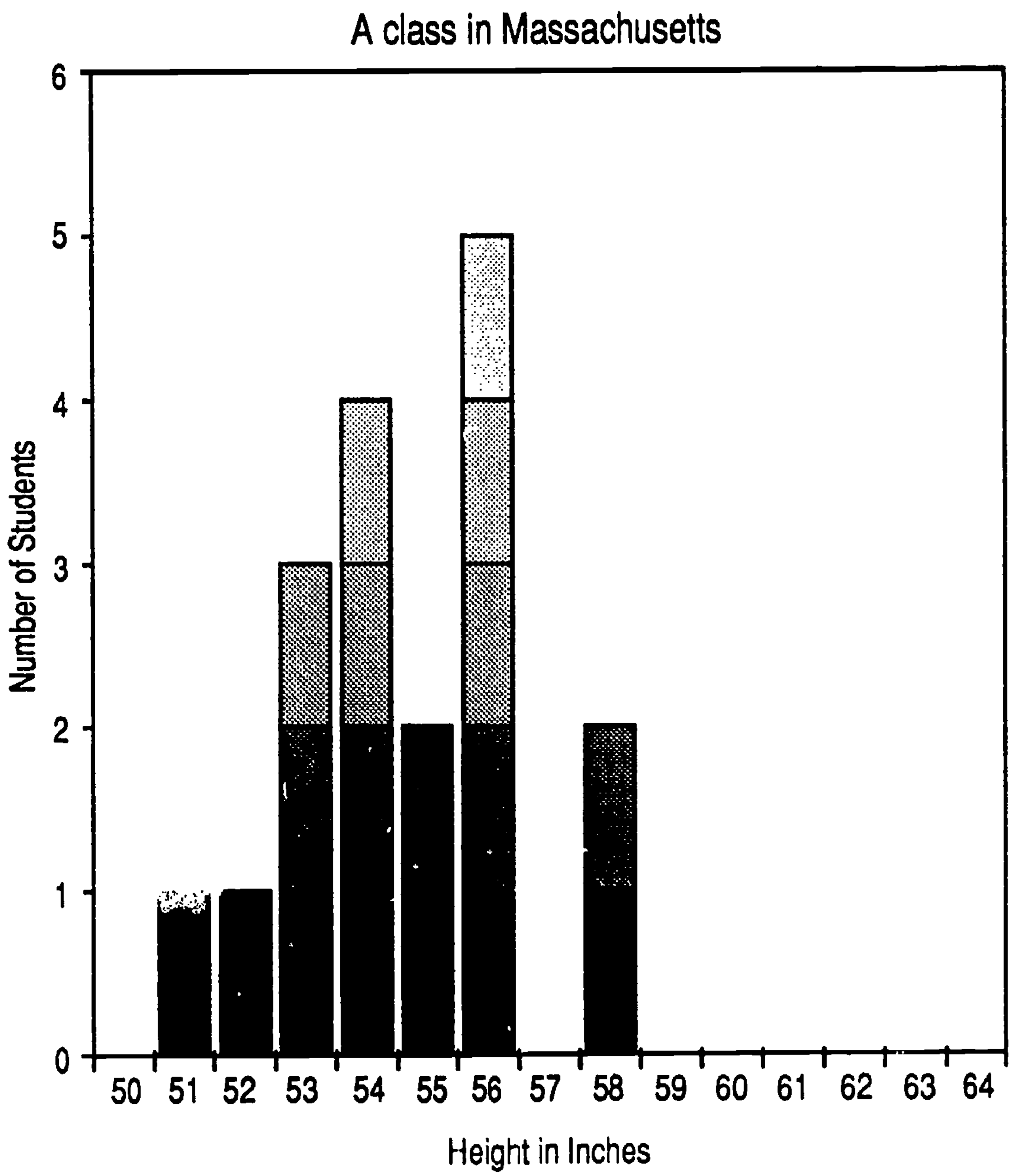
As an example, when students in one upper elementary class collected information about ways in which they had been injured while playing, they ended up with so much data that they could not easily organize and describe them. They concluded that they couldn't make any sense of their results unless they limited the sphere of their investigation. They decided to limit the study to play injuries incurred *at school*. This more focused study allowed them to find patterns in the data, and they were ultimately able to make important recommendations about safety on their own playground. The students were able to make good decisions on their own about how to focus their study *after* their first experiences with their data—and they learned something important about data analysis in the process.

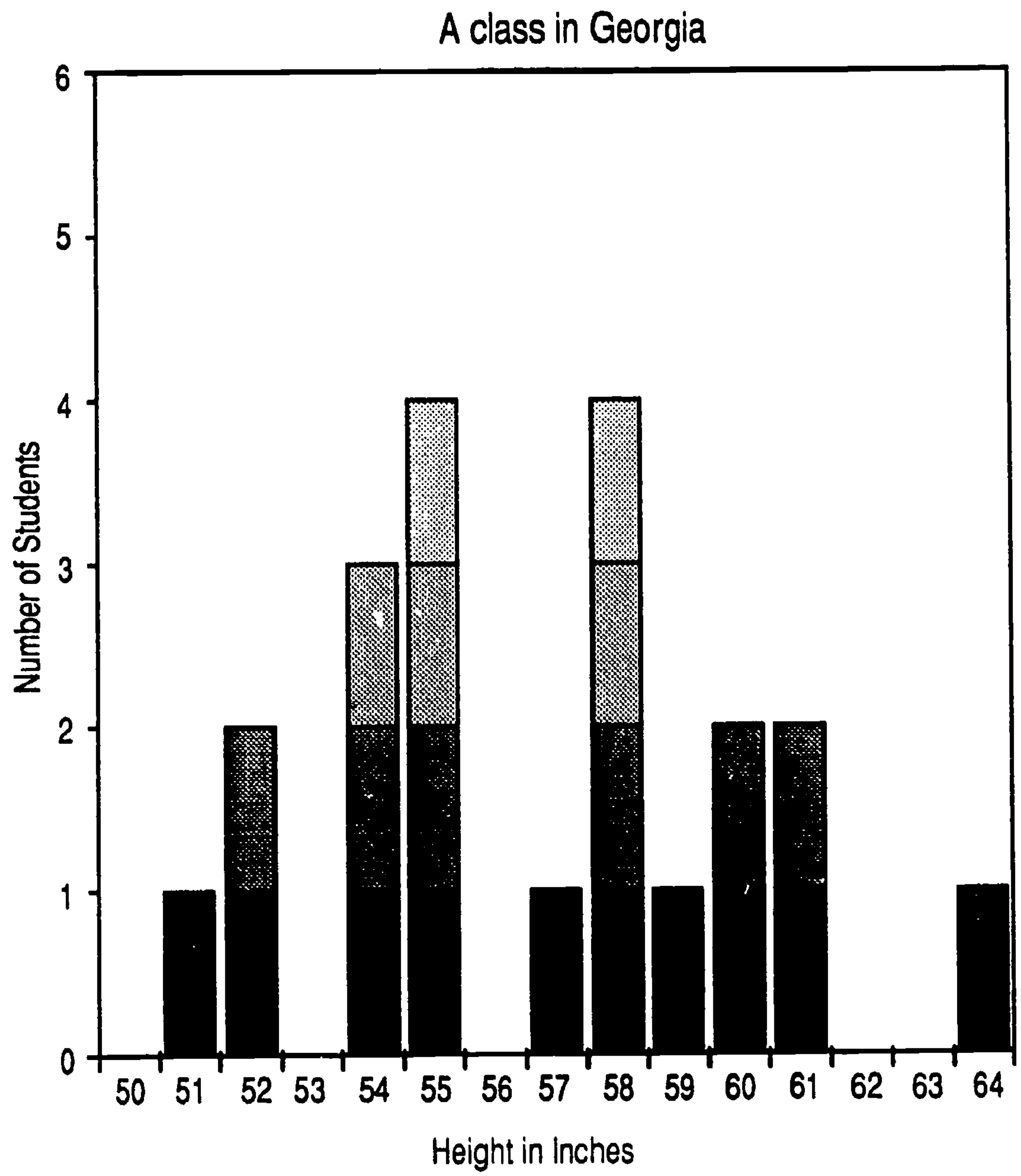
Another typical problem is a collection of data that does not reveal an obvious pattern. For example, students often collect data on favorite things and typically end up with "flat graphs." That is, a survey on favorite ice cream flavor may look like this:



Teachers often see in advance that students will need to categorize their data and may be tempted to suggest categories even before the data are collected: "You're going to get so many different answers. Maybe you just want to select five popular flavors and do your survey with those flavors." However, simplifying the question *ahead of time* may result in less interesting data, as well as lower student interest. After students have collected some data, you can encourage them to reclassify the information: "Can you find ways of reorganizing your results so that you can make some generalizations? If you grouped together all the flavors that include chocolate, are there some patterns?" Statisticians often need to think about reclassifying their collected information as they look for patterns in the data. This process is the very essence of data analysis. ■

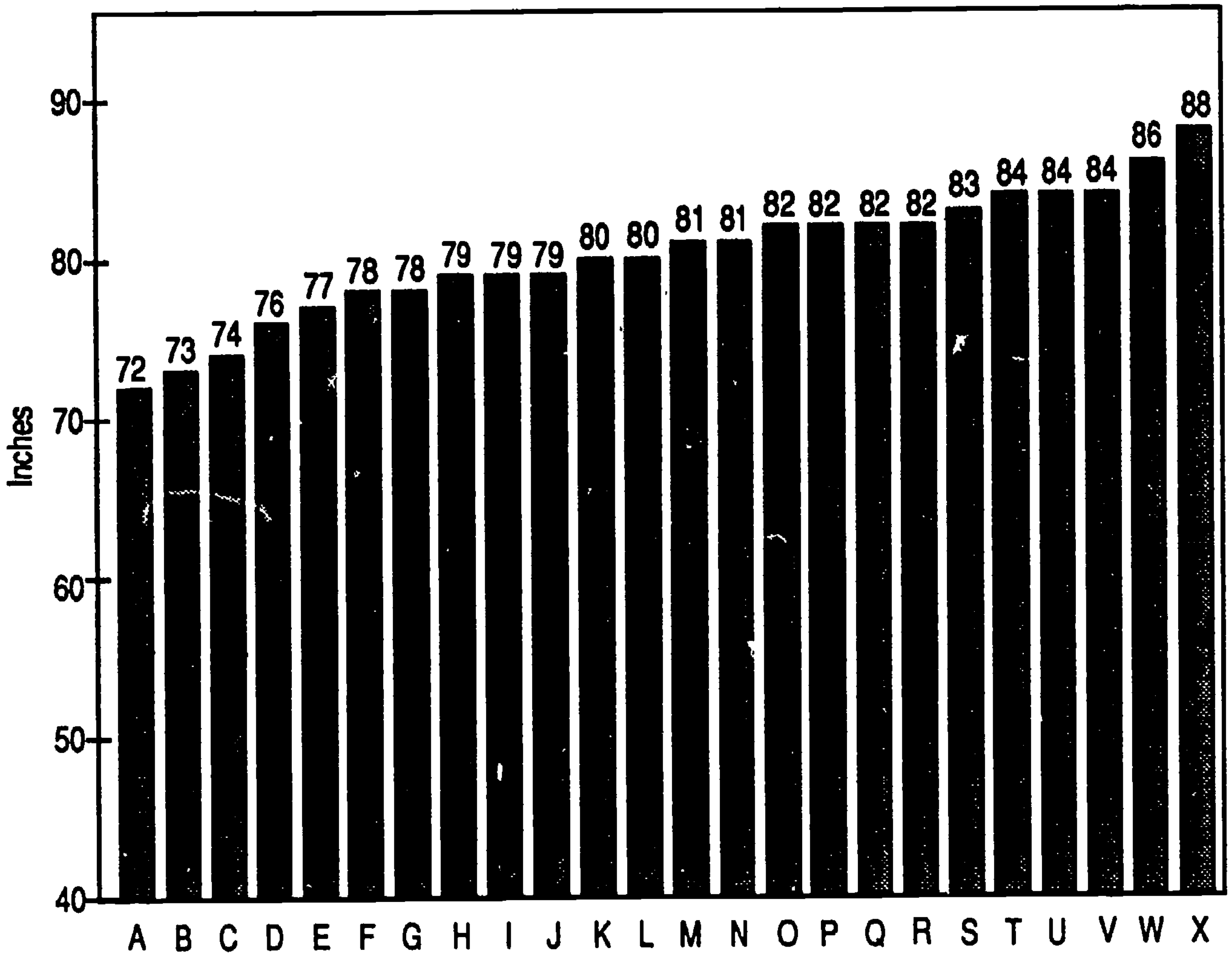
HEIGHTS OF TWO FOURTH GRADE CLASSES

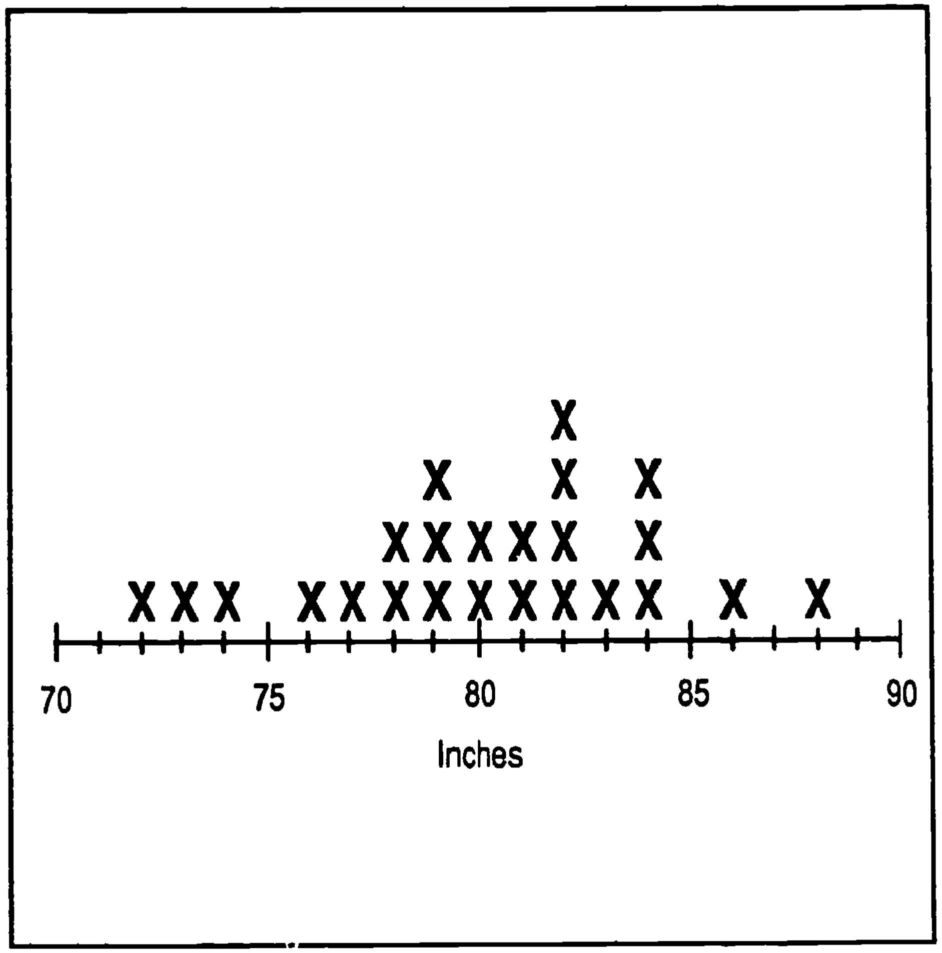




MYSTERY DATA A

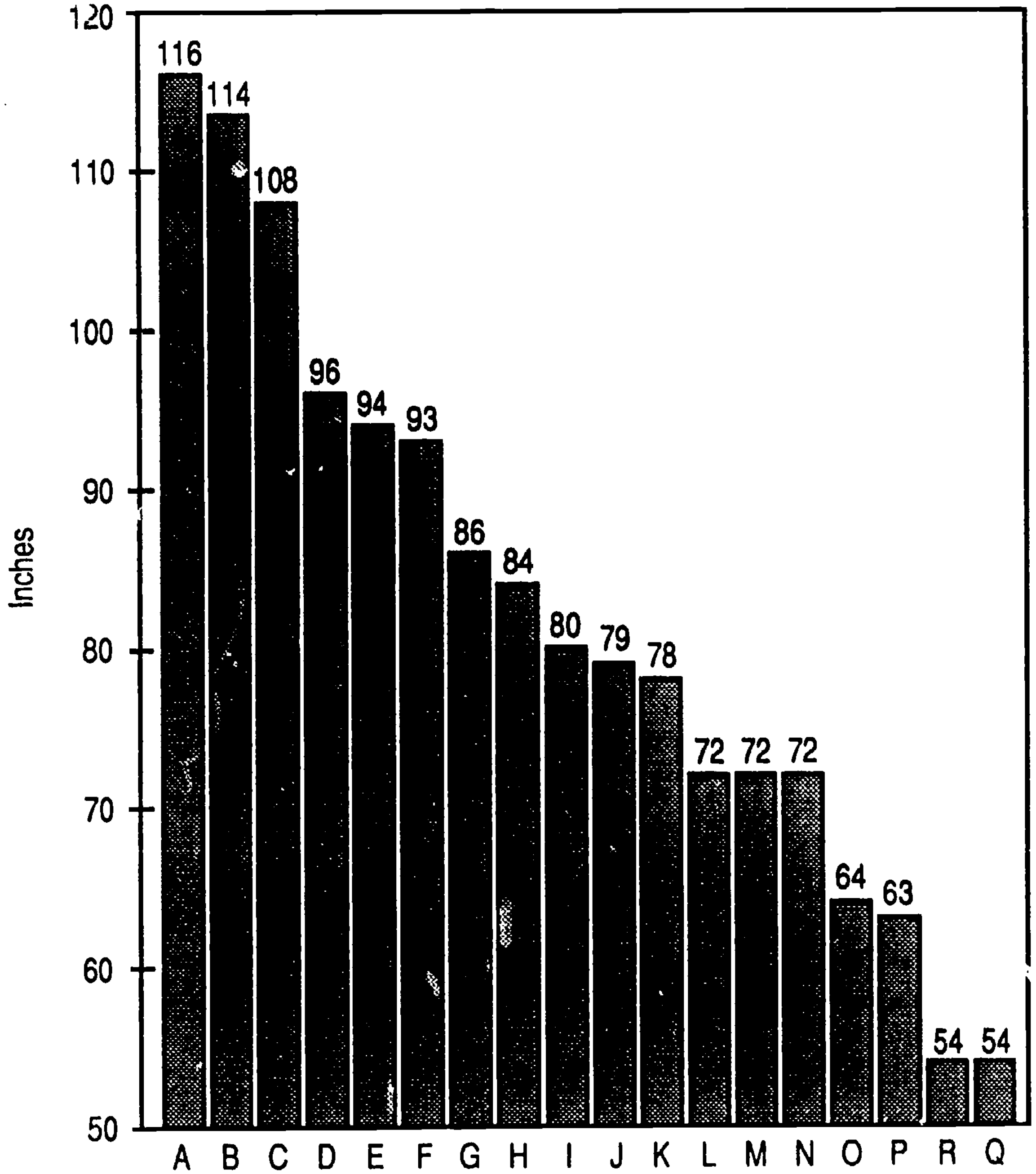
The two graphs below show the same data. These data represent some group of living things.

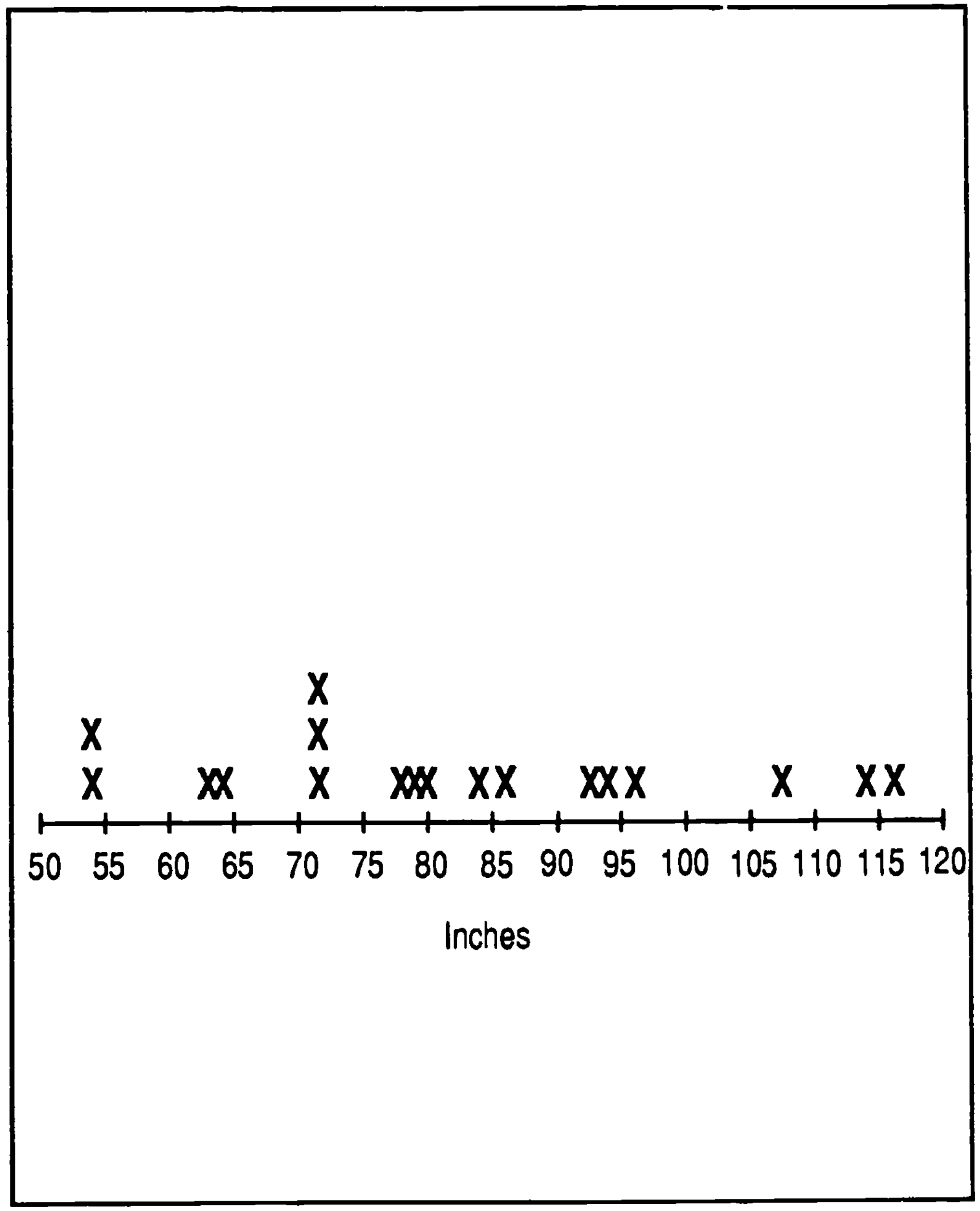




MYSTERY DATA B

The two graphs below show the same data. These data represent some group of living things.

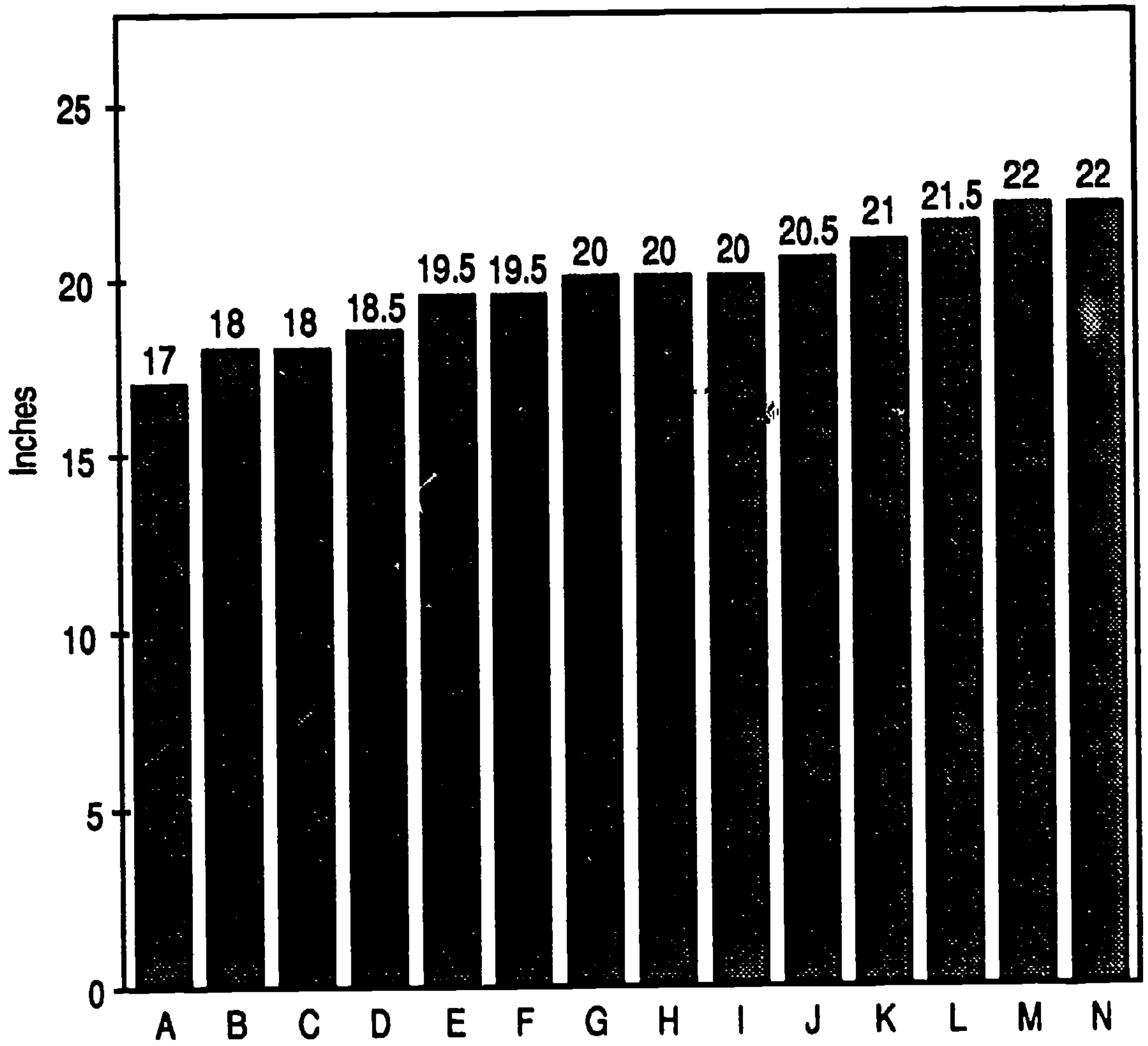


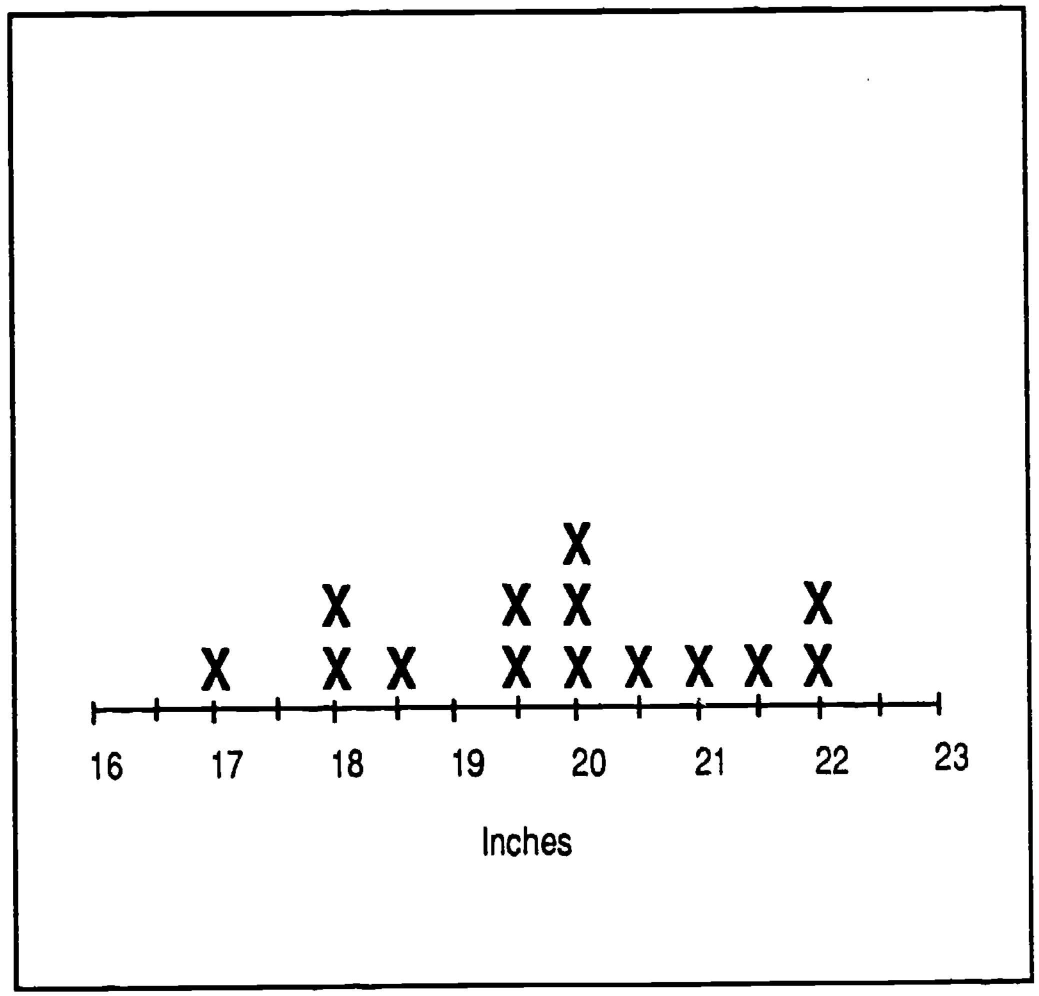


MYSTERY DATA C

STUDENT SHEET 4

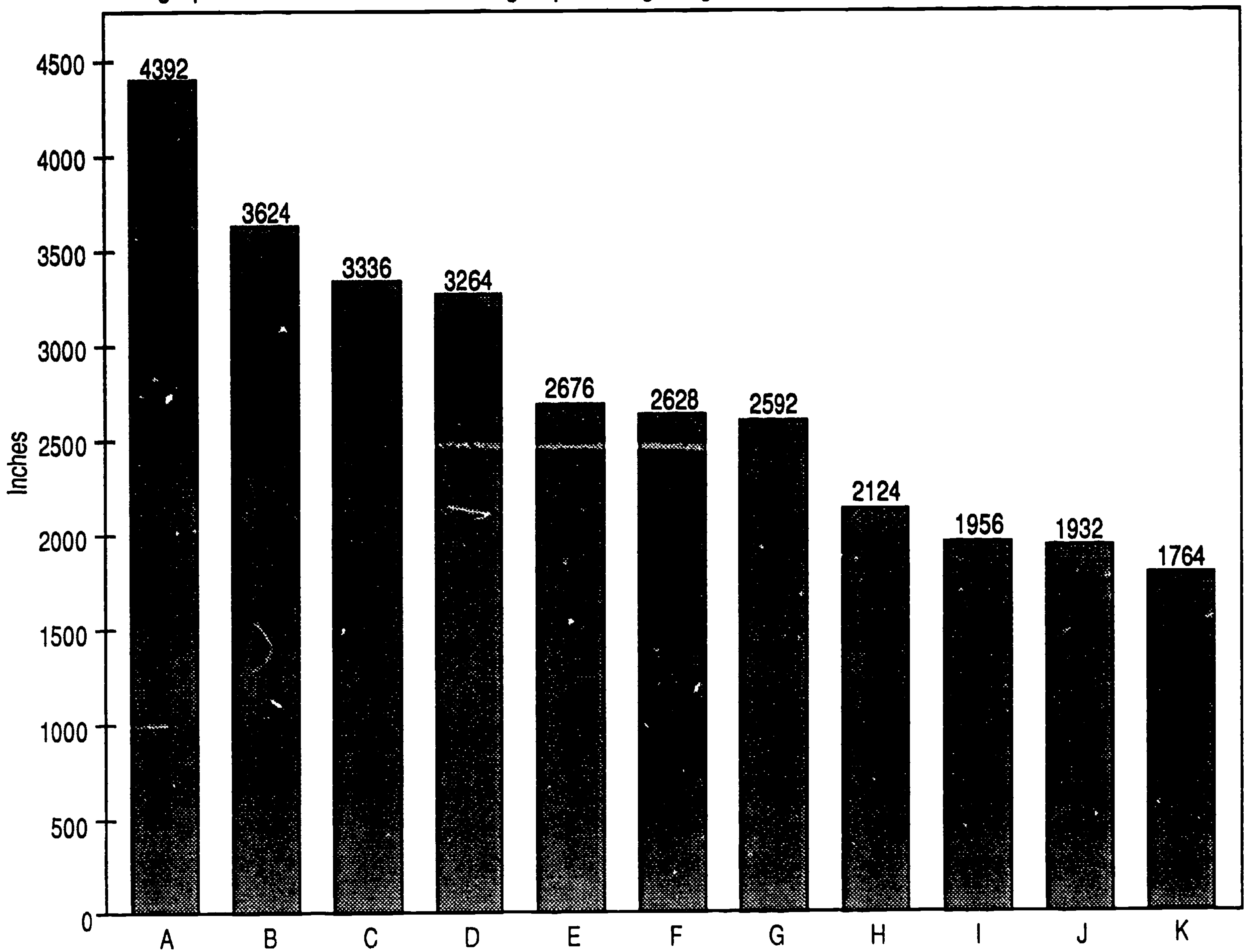
The two graphs below show the same data. These data represent some group of living things.



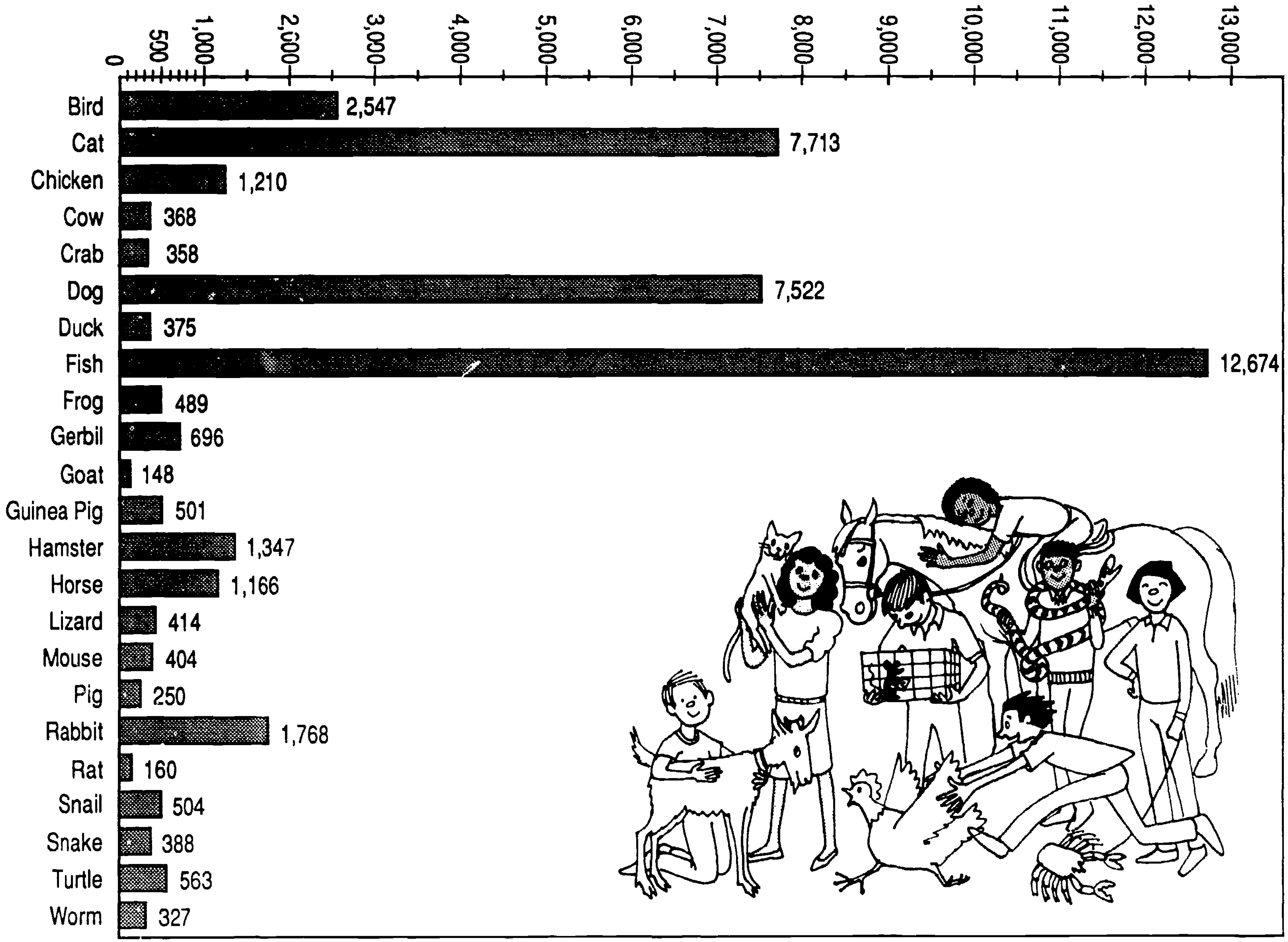


MYSTERY DATA D

This graph shows some data about a group of living things.



PET DATA

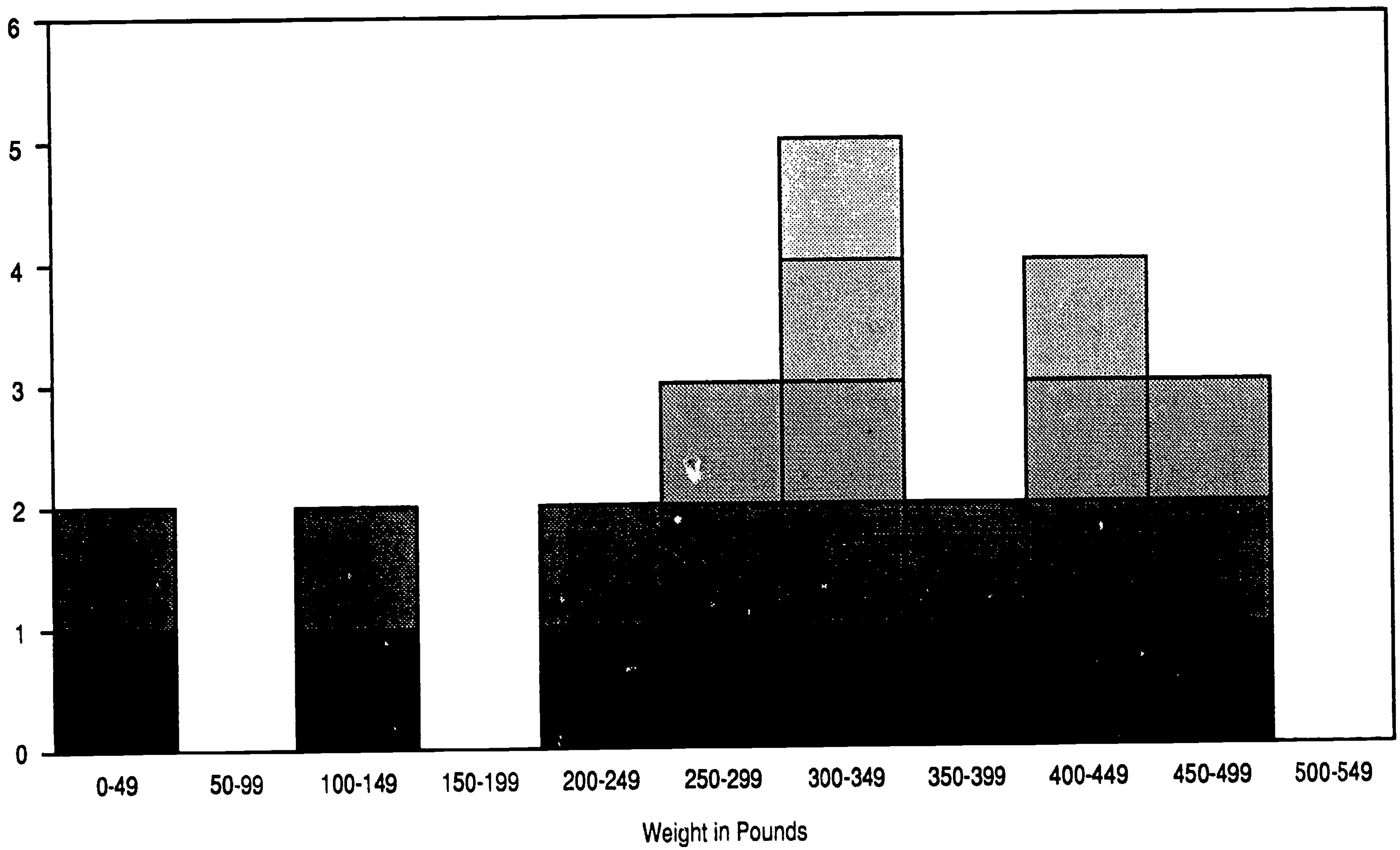

0
500
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
10,000
11,000
12,000
13,000
Bird 2,547
Cat 7,713
Chicken 1,210
Cow 368
Crab 358
Dog 7,522
Duck 375
Fish 12,674
Frog 489
Gerbil 696
Goat 148
Guinea Pig 501
Hamster 1,347
Horse 1,166
Lizard 414
Mouse 404
Pig 250
Rabbit 1,768
Rat 160
Snail 504
Snake 388
Turtle 563
Worm 327

CAT WEIGHTS

Ravena	14 pounds
Lady Jane Grey	8.5 pounds
Peebles	9 pounds
Grey Kitty	9 pounds
Weary	15 pounds
Misty	9 pounds
Wally	10 pounds
Amex	10 pounds
Melissa	11 pounds
Oddfuzz	18 pounds
Diva	11 pounds
Peau de Soie	7 pounds
Alexander	11 pounds
Lady	12 pounds
Mittens	10.5 pounds
George	14.5 pounds
Pepper	12 pounds
Strawberry	14.5 pounds
K.C.	16 pounds
Charcoal	12 pounds
Tigger	8 pounds
Tomonochi	8 pounds
Katenka	8.5 pounds

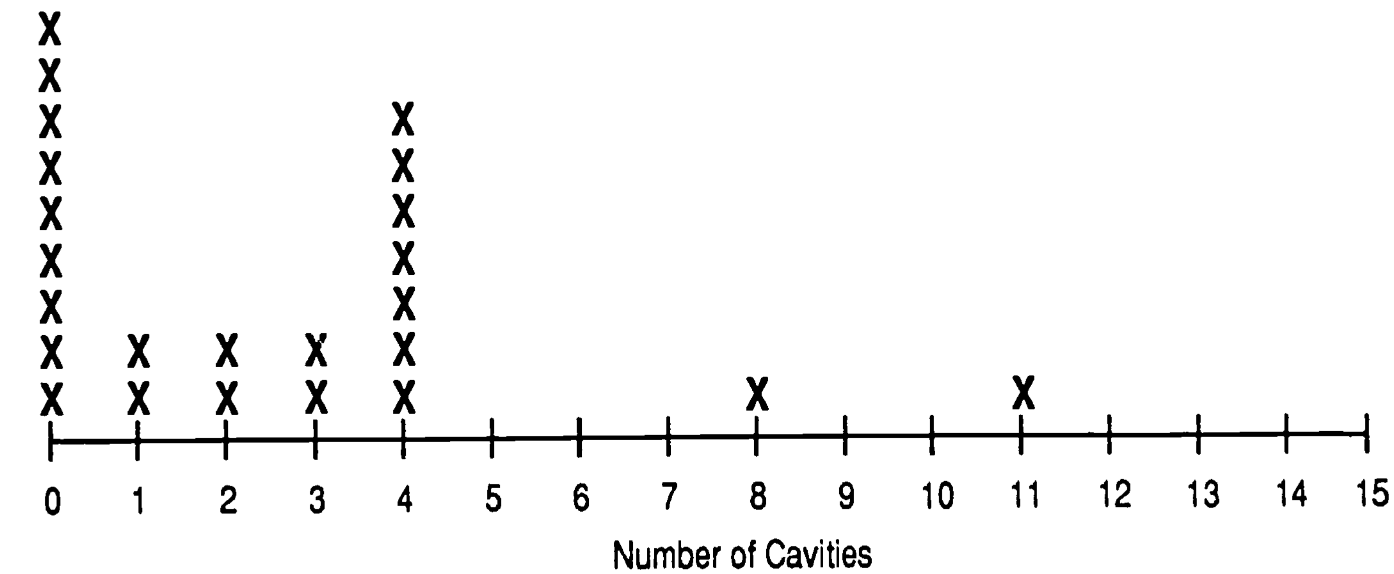
ANOTHER MYSTERY

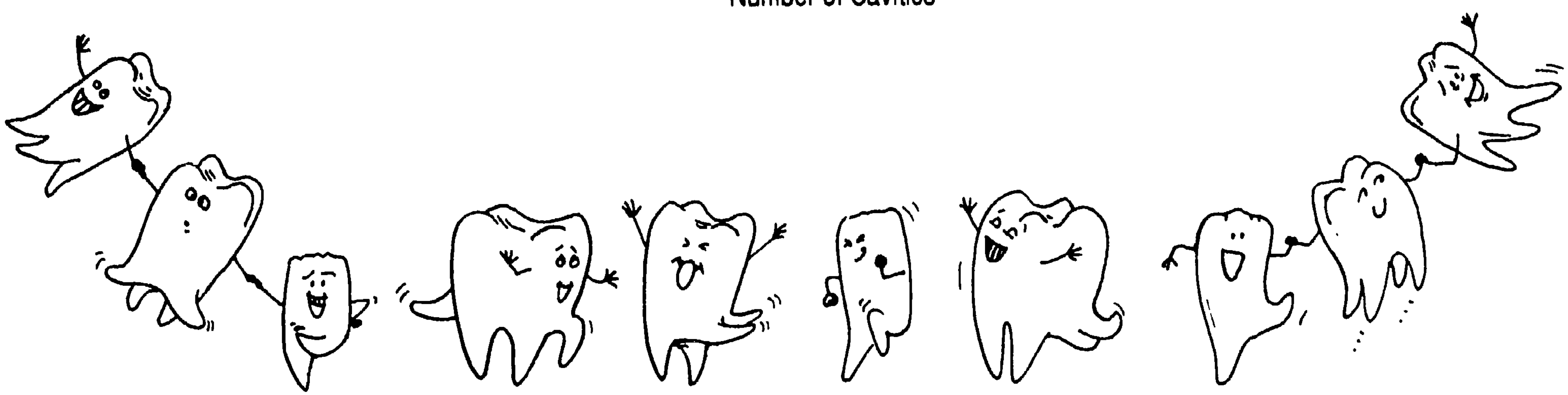
This graph shows some data about a group of living things. What do you think they are?



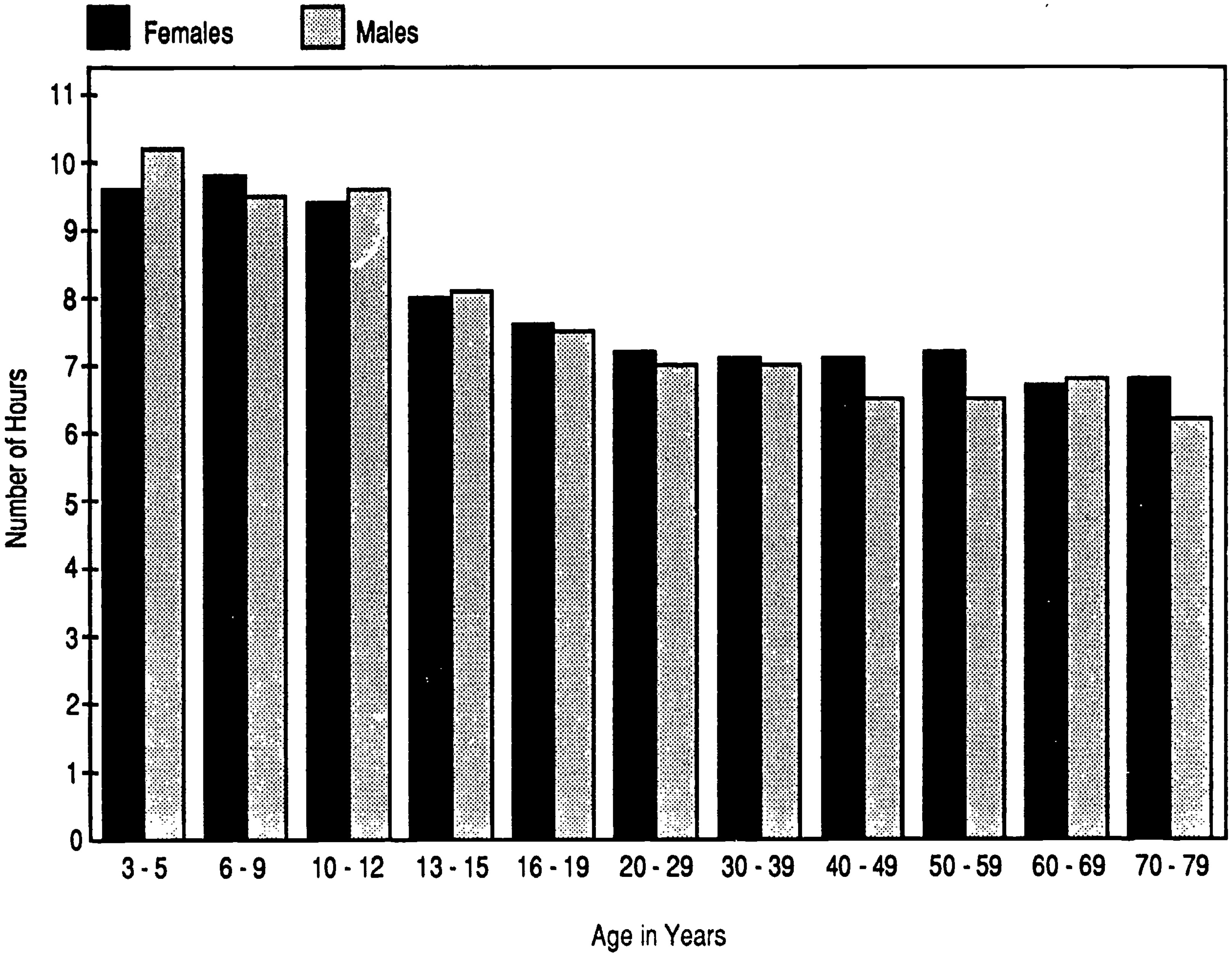
CAVITY DATA

These data represent cavities counted by a dentist in a group of 9- to 12-year-olds.

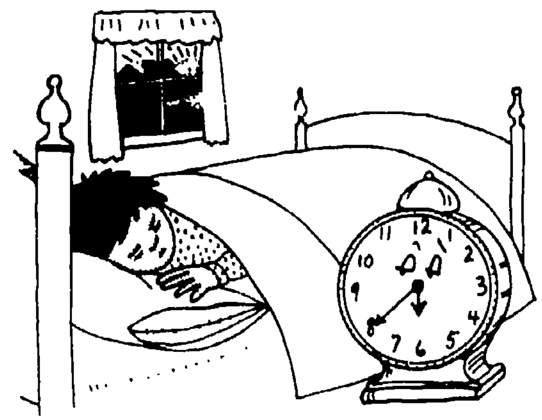




HOW LONG DO PEOPLE SLEEP?



Age	Females	Males
3-5	9.6	10.2
6-9	9.8	9.5
10-12	9.3	9.6
13-15	8.0	8.1
16-19	7.6	7.5
20-29	7.2	7.0
30-39	7.1	7.0
40-49	7.1	6.5
50-59	7.2	6.5
60-69	6.7	6.8
70-79	6.8	6.2



Source: Robert L. Williams, Ismet Karacan, Carolyn J. Hursch. *Electroencephalography (EEG) of human sleep: Clinical applications* (New York: Wiley, 1974).

'89 All-Star Alex English Denver Nuggets 80"	'89 All-Star Karl Malone Utah Jazz 81"	'89 All-Star Akeem Olajuwon Houston Rockets 84"	'89 All-Star Dale Ellis Seattle SuperSonics 77"
'89 All-Star John Stockton Utah Jazz 74"	'89 All-Star Clyde Drexler Portland Trail Blazers 79"	'89 All-Star Kareem Abdul-Jabbar Los Angeles Lakers 86"	'89 All-Star Tom Chambers Phoenix Suns 82"

'89 All-Star Chris Mullin Golden State Warriors 79"	'89 All-Star Mark Eaton Utah Jazz 88"	'89 All-Star Kevin Duckworth Portland Trail Blazers 82"	'89 All-Star James Worthy Los Angeles Lakers 81"
'89 All-Star Charles Barkley Philadelphia 76ers 78"	'89 All-Star Moses Malone Atlanta Hawks 83"	'89 All-Star Dominique Wilkins Atlanta Hawks 79"	'89 All-Star Michael Jordan Chicago Bulls 78"

'89 All-Star Isiah Thomas Detroit Pistons 73"	'89 All-Star Patrick Ewing New York Knicks 84"	'89 All-Star Pat Cummings New York Knicks 80"	'89 All-Star Larry Nance Cleveland Cavaliers 82"
'89 All-Star Mark Price Cleveland Cavaliers 72"	'89 All-Star Mark Jackson New York Knicks 76"	'89 All-Star Pat Dorhety Cleveland Cavaliers 84"	'89 All-Star Kevin McHale Boston Celtics 82"