Statistics and Probability Standards in the CCSS-M Grades 6-HS

Grade 6

Develop understanding of statistical variability.

-6.SP.A.1 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. *For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages*.

-6.SP.A.2 Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.

-6.SP.A.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

Summarize and describe distributions.

-6.SP.B.4 Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

-6.SP.B.5 Summarize numerical data sets in relation to their context, such as by:

-6.SP.B.5a Reporting the number of observations.

-6.SP.B.5b Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

-6.SP.B.5c Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

-6.SP.B.5d Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

Grade 7

Use random sampling to draw inferences about a population.

-7.SP.A.1 Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.

-7.SP.A.2 Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. *For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or* *prediction might be.*

Draw informal comparative inferences about two populations.

-7.SP.B.3 Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *For example, the mean height of players on the basketball team is 10 cm greater than the mean height of players on the soccer team, about twice the variability (mean absolute deviation) on either team; on a dot plot, the separation between the two distributions of heights is noticeable.*

-7.SP.B.4 Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *For example, decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book.*

Investigate chance processes and develop, use, and evaluate probability models.

-7.SP.C.5 Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event.

-7.SP.C.6 Approximate the probability of a chance event by collecting data on the chance process that produces it and observing its long-run relative frequency, and predict the approximate relative frequency given the probability. *For example, when rolling a number cube 600 times, predict that a 3 or 6 would be rolled roughly 200 times, but probably not exactly 200 times.*

-7.SP.C.7 Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy.

-7.SP.C.7a Develop a uniform probability model by assigning equal probability to all outcomes, and use the model to determine probabilities of events. *For example, if a student is selected at random from a class, find the probability that Jane will be selected and the probability that a girl will be selected.*

-7.SP.C.7b Develop a probability model (which may not be uniform) by observing frequencies in data generated from a chance process. *For example, find the approximate probability that a spinning penny will land heads up or that a tossed paper cup will land open-end down. Do the outcomes for the spinning penny appear to be equally likely based on the observed frequencies?*

-7.SP.C.8 Find probabilities of compound events using organized lists, tables, tree diagrams, and simulation.

-7.SP.C.8a Understand that, just as with simple events, the probability of a compound event is the fraction of outcomes in the sample space for which the compound event occurs.

-7.SP.C.8b Represent sample spaces for compound events using methods such as organized lists, tables and tree diagrams. For an event described in everyday language (e.g., “rolling double sixes”), identify the outcomes in the sample space which compose the event.

-7.SP.C.8c Design and use a simulation to generate frequencies for compound events. *For example, use random digits as a simulation tool to approximate the answer to the question: If 40% of donors have type A blood, what is the probability that it will take at least 4 donors to find one with type A blood?*

Grade 8

Investigate patterns of association in bivariate data.

-8.SP.A.1 Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

-8.SP.A.2 Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.

-8.SP.A.3 Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height*.

-8.SP.A.4 Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables. *For example, collect data from students in your class on whether or not they have a curfew on school nights and whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores?*

High School

Summarize, represent, and interpret data on a single count or measurement variable

-HSS-ID.A.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).(Alg 1/Int 1)

-HSS-ID.A.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets. (Alg 1/Int 1)

-HSS-ID.A.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers). (Alg 1/Int 1)

-HSS-ID.A.4 Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve. (Alg 2/Int 3)

Summarize, represent, and interpret data on two categorical and quantitative variables

-HSS-ID.B.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data. (Alg 1/Int 1)

-HSS-ID.B.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related. (Alg 1/Int 1)

-HSS-ID.B.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models. (Alg 1/Int 1)

-HSS-ID.B.6b Informally assess the fit of a function by plotting and analyzing residuals. (Alg 1/Int 1)

-HSS-ID.B.6c Fit a linear function for a scatter plot that suggests a linear association. (Alg 1/Int 1)

Interpret linear models

-HSS-ID.C.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data. (Alg 1/Int 1)

-HSS-ID.C.8 Compute (using technology) and interpret the correlation coefficient of a linear fit. (Alg 1/Int 1)

-HSS-ID.C.9 Distinguish between correlation and causation. (Alg 1/Int 1)

Understand and evaluate random processes underlying statistical experiments

-HSS-IC.A.1 Understand statistics as a process for making inferences about population parameters based on a random sample from that population.(Alg 2/Int 3)

-HSS-IC.A.2 Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation. *For example, a model says a spinning coin falls heads up with probability 0.5. Would a result of 5 tails in a row cause you to question the model*? (Alg 2/Int 3)

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

-HSS-IC.B.3 Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each. (Alg 2/Int 3)

-HSS-IC.B.4 Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling. (Alg 2/Int 3)

-HSS-IC.B.5 Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant. (Alg 2/Int 3)

-HSS-IC.B.6 Evaluate reports based on data. (Alg 2/Int 3)

Understand independence and conditional probability and use them to interpret data

-HSS-CP.A.1 Describe events as subsets of a sample space (the set of outcomes) using characteristics (or categories) of the outcomes, or as unions, intersections, or complements of other events ("or," "and," "not").(Geom/Int 2)

-HSS-CP.A.2 Understand that two events *A* and *B* are independent if the probability of *A* and *B* occurring together is the product of their probabilities, and use this characterization to determine if they are independent. (Geom/Int 2)

-HSS-CP.A.3 Understand the conditional probability of *A* given *B* as *P*(*A* and *B*)/*P*(*B*), and interpret independence of *A* and *B* as saying that the conditional probability of *A* given *B* is the same as the probability of *A*, and the conditional probability of *B* given *A* is the same as the probability of *B*. (Geom/Int 2)

-HSS-CP.A.4 Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities. *For example, collect data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

-HSS-CP.A.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.* (Geom/Int 2)

Use the rules of probability to compute probabilities of compound events.

-HSS-CP.B.6 Find the conditional probability of *A* given *B* as the fraction of *B*'s outcomes that also belong to *A*, and interpret the answer in terms of the model. (Geom/Int 2)

-HSS-CP.B.7 Apply the Addition Rule, P(A or B) = P(A) + P(B) – P(A and B), and interpret the answer in terms of the model. (Geom/Int 2)

-HSS-CP.B.8 (+) Apply the general Multiplication Rule in a uniform probability model, P(A and B) = P(A)P(B|A) = P(B)P(A|B), and interpret the answer in terms of the model. (Geom/Int 2)

-HSS-CP.B.9 (+) Use permutations and combinations to compute probabilities of compound events and solve problems. (Geom/Int 2)

Calculate expected values and use them to solve problems

-HSS-MD.A.1 (+) Define a random variable for a quantity of interest by assigning a numerical value to each event in a sample space; graph the corresponding probability distribution using the same graphical displays as for data distributions.(4^{th} year)

-HSS-MD.A.2 (+) Calculate the expected value of a random variable; interpret it as the mean of the probability distribution. (4^{th} year)

-HSS-MD.A.3 (+) Develop a probability distribution for a random variable defined for a sample space in which theoretical probabilities can be calculated; find the expected value. *For example, find the theoretical probability distribution for the number of correct answers obtained by guessing on all five questions of a multiple-choice test where each question has four choices, and find the expected grade under various grading schemes.* (4^{th} year)

-HSS-MD.A.4 (+) Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value. *For example, find a current data distribution on the number of TV sets per household in the United States, and calculate the expected number of sets per household. How many TV sets would you expect to find in 100 randomly selected households?* (4^{th} year)

Use probability to evaluate outcomes of decisions

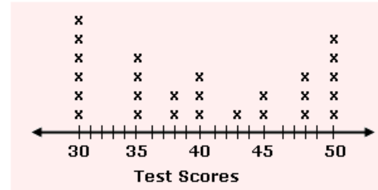
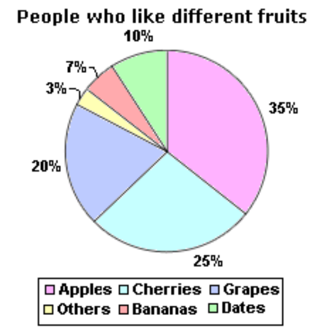
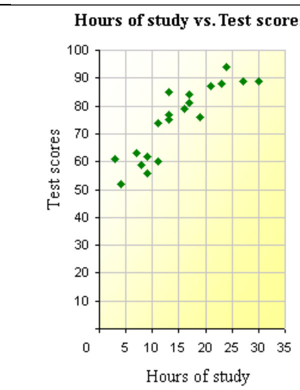
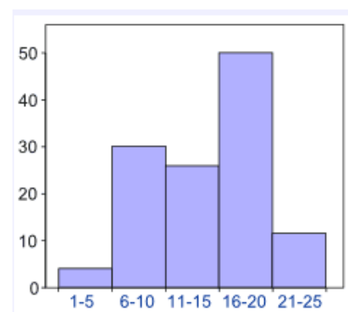
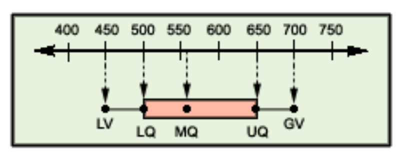
-HSS-MD.B.5 (+) Weigh the possible outcomes of a decision by assigning probabilities to payoff values and finding expected values. (4^{th} year)

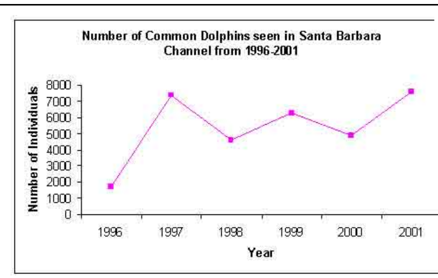
-HSS-MD.B.5a Find the expected payoff for a game of chance. *For example, find the expected winnings from a state lottery ticket or a game at a fast-food restaurant.* (4^{th} year)

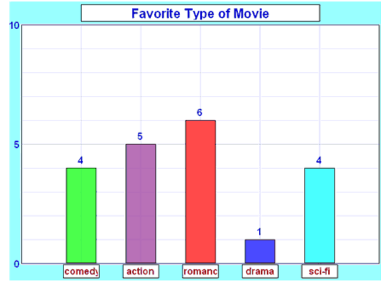
-HSS-MD.B.5b Evaluate and compare strategies on the basis of expected values. *For example, compare a high-deductible versus a low-deductible automobile insurance policy using various, but reasonable, chances of having a minor or a major accident.* (4^{th} year)

-HSS-MD.B.6 (+) Use probabilities to make fair decisions (e.g., drawing by lots, using a random number generator).(Alg 1, Geom/Int 1, Int 2)

-HSS-MD.B.7 (+) Analyze decisions and strategies using probability concepts (e.g., product testing, medical testing, pulling a hockey goalie at the end of a game). (Alg 1, Geom/Int 1, Int 2)

<table>
<tr>
<td>Dot Plot
or Line Plot</td>
<td>Shows data on a number line with dot or other marks to show frequency. Generally used when the data is discrete and numerous pieces on data fall on most values.</td>
<td>
</td>
</tr>
<tr>
<td>Circle Graph</td>
<td>Graph in the form of a circle that is divided into sectors, with each sector representing a part of a set of data. Best used to compare parts of a whole.</td>
<td>
</td>
</tr>
<tr>
<td>Scatter Plot</td>
<td>Graph made by plotting ordered pairs in a coordinate plane to show the correlation between two sets of data</td>
<td>
</td>
</tr>
<tr>
<td>Histogram</td>
<td>Bar graph that shows how frequently data occur within certain ranges or intervals. The height of each bar gives the frequency in the respective interval.</td>
<td>Ages of people purchasing snow cones at the school carnival.

</td>
</tr>
<tr>
<td>Box Plot</td>
<td>A plot that shows the distribution of a set of data along a number line, dividing the data into four parts using the median and quartiles.</td>
<td>The amount of calories present in some bakery items


LV - LEAST VALUE
LQ - LOWER QUARTILE
MQ - MIDDLE QUARTILE
GV - GREATEST VALUE
UQ - UPPER QUARTILE</td>
</tr>
<tr>
<td>Line Graph</td>
<td>A graph that uses line segments to connect data points and generally shows changes in data over time.</td>
<td>
</td>
</tr>
</table>

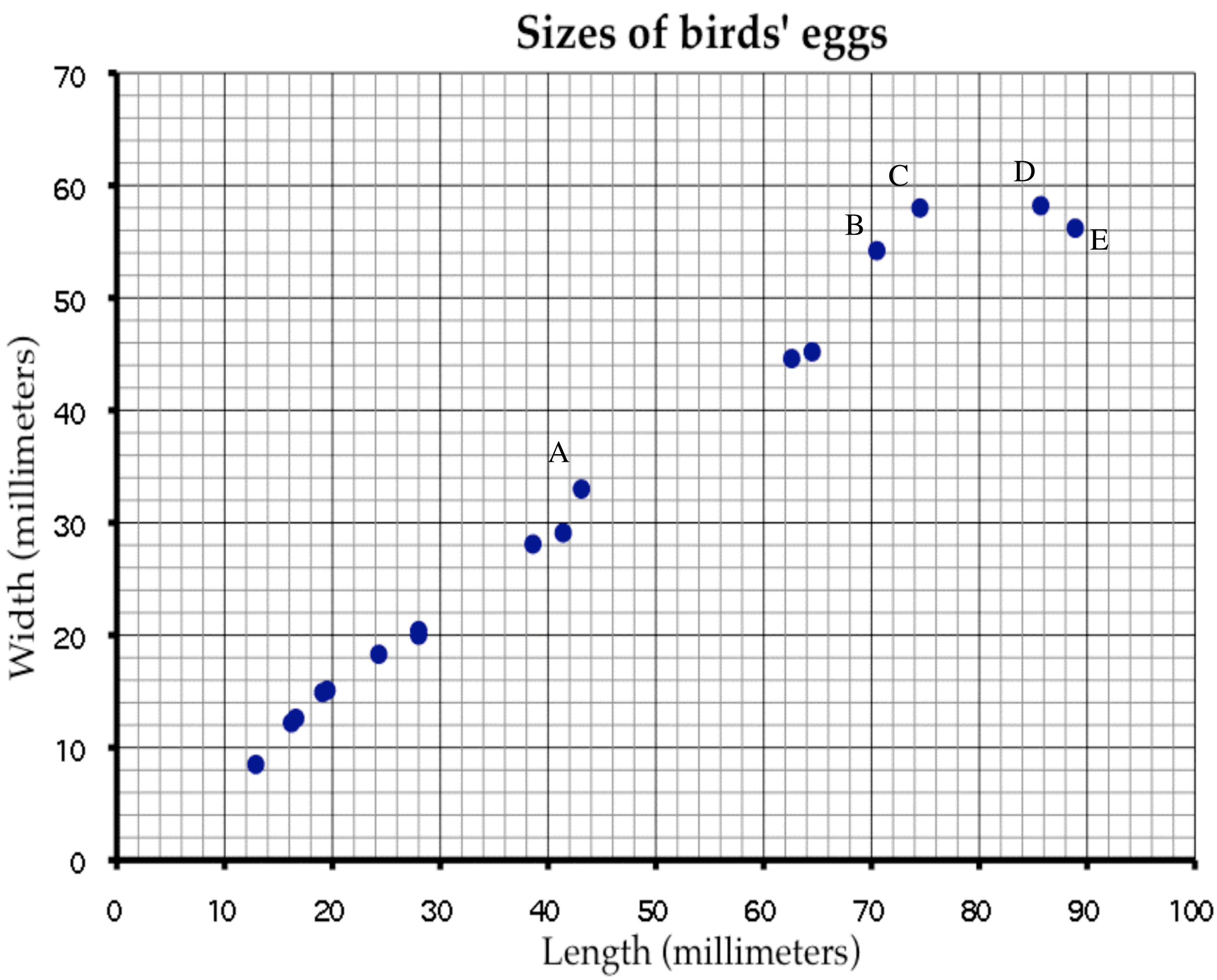
Bar Graph	A graph used when data falls in categories that typically have no numeric order. The height of each bar gives the frequency in the respective category.	

Directions: Print 1 set for each group cutting them apart. As an ice breaker, give each participant a different graph piece (of the first 4) and instructions for them to find three other different graphs to form a group. When they sit down, there is an envelope with the other graph pieces and all of the descriptions and titles. The group has to match each title with its description and graph.

When we did this activity it was a great review/refresher for the teachers….middle school teachers did better than the High School teachers.

Birds' Eggs

This scatter diagram shows the lengths and the widths of the eggs of some American birds.



1. A biologist measured a sample of one hundred Mallard duck eggs and found they had an average length of 57.8 millimeters and average width 41.6 millimeters.

Use a **X** to mark a point that represents this on the scatter diagram.

2. What does the graph show about the connection between the lengths of birds' eggs and their widths?

3. Another sample of similar birds has eggs with a length of 35 millimeters on average.
If these birds follow the trend in the scatterplot, about what width would you expect these eggs to be, on average?

4. Describe the differences in shape of the two eggs C and D.

5. Which of the eggs A, B, C, D, and E has the greatest ratio of length to width?
Explain how you decided.

MAT.08.ER.3.000SP.B.134 Claim 3

Sample Item ID:	MAT.08.ER.3.000SP.B.134
Grade:	08
Primary Claim:	**Claim 3: Communicating Reasoning** Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.
Secondary Claim(s):	Claim 1: Concepts and Procedures Students can explain and apply mathematical concepts and carry out mathematical procedures with precision and fluency. Claim 2: Problem Solving Students can solve a range of complex well-posed problems in pure and applied mathematics, making productive use of knowledge and problem solving strategies.
Primary Content Domain:	Statistics and Probability
Secondary Content Domain(s):	
Assessment Target(s):	3 B: Construct, autonomously, chains of reasoning that will justify or refute propositions or conjectures. 1 J: Investigate patterns of association in bivariate data. 2 A: Apply mathematics to solve well-posed problems arising in everyday life, society, and the workplace. 2 D: Identify important quantities in a practical situation and map their relationships (e.g., using diagrams, two-way tables, graphs, flowcharts, or formulas).
Standard(s):	8.SP.4
Mathematical Practice(s):	1, 2, 3, 5, 6, 7
DOK:	3
Item Type:	ER
Score Points:	4
Difficulty:	M
Key:	See Sample Top-Score Response.
Stimulus/Source:	
Target-Specific Attributes (e.g., accessibility issues):	Calculators may be used for this item.
Notes:	The table consists of a response space in each blank cell. Each response space allows for a maximum of 2 digits.

Jacob surveyed 25 adults to ask whether they had at least one child under the age of 18 and whether they had at least one pet. This table shows the results of the survey.

Adult	**At Least One Child Under the Age of 18**		**At Least One Pet**	
	Yes	No	Yes	No
A	✓		✓	
B		✓		✓
C	✓			✓
D		✓		✓
E		✓	✓	
F		✓		✓
G		✓	✓	
H	✓		✓	
I	✓		✓	
J	✓			✓
K		✓		✓
L	✓		✓	
M		✓	✓	
N	✓		✓	
O		✓		✓
P	✓		✓	
Q		✓		✓
R	✓		✓	
S		✓		✓
T	✓		✓	
U		✓	✓	
V	✓		✓	
W		✓		✓
X		✓	✓	
Y		✓	✓	

Part A

Use the results of the survey to complete this table.

At Least One Pet	At Least One Child Under the Age of 18	
	Yes	**No**
Yes		
No		

Part B

Jacob made the conjecture that there is a possible association between whether an adult has at least one child under the age of 18 and whether the adult has at least one pet.

State whether the results of the survey provide evidence that adults who have at least one child under the age of 18 also tend to have at least one pet. Explain your answer.

Sample Top-Score Response:

Part A:

At Least One Pet	At Least One Child Under the Age of 18	
	Yes	No
Yes	9	6
No	2	8

Part B:
Yes, there is evidence that the adults who have at least one child under the age of 18 also tend to have at least one pet. I found the relative frequencies for whether the adult had at least one pet or not given that the adult had at least one child and then given that the adult did not have any children, and 82% of the adults who had at least one child also had at least one pet. My work is shown below.

At Least One Pet	At Least One Child Under the Age of 18		Totals
	Yes	No	
Yes	$\frac{9}{11} \approx 82\%$	$\frac{6}{14} \approx 43\%$	$\frac{15}{25} = 60\%$
No	$\frac{2}{11} \approx 18\%$	$\frac{8}{14} \approx 57\%$	$\frac{10}{25} = 40\%$

Scoring Rubric:

Responses to this item will receive 0-4 points, based on the following:

Part A

2 points: The student thoroughly understands how to construct two-way tables. The student's response in *Part A* matches the sample top-score response.

1 point: The student shows a partial understanding of how to construct two-way tables. The student makes one or two minor mathematical errors when constructing the table.

0 points: The student shows inconsistent or no understanding of how to construct two-way tables. The student makes several mathematical errors when constructing the table or one or more conceptual errors.

Part B

2 points: The student thoroughly understands how to interpret two-way tables. The student uses relative frequencies or a similar method in an explanation that corresponds with the table the student provided in *Part A*.

1 point: The student shows a partial understanding of how to interpret two-way tables. The student uses the numbers in the table without using relative frequencies or a similar method in an explanation that corresponds with the table the student provided in *Part A*.

0 points: The student shows inconsistent or no understanding of how to interpret two-way tables. The student does not use the table to help construct the explanation OR the student writes an explanation that does not correspond with the table the student provided in *Part A*.

PROBLEM SOLVING

Mathematics Assessment Project

CLASSROOM CHALLENGES

A Formative Assessment Lesson

Interpreting Statistics: A Case of Muddying the Waters

Mathematics Assessment Resource Service
University of Nottingham & UC Berkeley
Beta Version

Interpreting Statistics:
A Case of Muddying the Waters

MATHEMATICAL GOALS

This lesson unit is intended to help you assess how well students are able to:

- Interpret data and evaluate statistical summaries.
- Critique someone else's interpretations of data and evaluations of statistical summaries.

The lesson also introduces students to the dangers of misapplying simple statistics in real-world contexts, and illustrates some of the common abuses of statistics and charts found in the media.

COMMON CORE STATE STANDARDS

This lesson relates to the following *Mathematical Practices* in the *Common Core State Standards for Mathematics*:

3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

This lesson gives students the opportunity to apply their knowledge of the following *Standards for Mathematical Content* in the *Common Core State Standards for Mathematics:*

S-ID: Summarize, represent, and interpret data on two categorical and quantitative variables. Interpret linear models.

S-IC: Make inferences and justify conclusions from sample surveys, experiments, and observational studies.

INTRODUCTION

This lesson unit is structured in the following way:

- Before the lesson, students work individually on a task that is designed to reveal their current understandings and difficulties. You review their work and create some questions that will help them to improve their solutions.
- At the start of the lesson, students role-play a scene that provides the background to the lesson task. Next, students work in small groups on a collaborative discussion task. They extract information from evidence and critically assess how it was gathered and presented.
- After a whole-class discussion students again work individually, on a new task. This has similar content to the original task, allowing them to demonstrate the progress they have made during the lesson.

MATERIALS REQUIRED

- Each student will need a copy of the initial assessment task *Muddying the Waters*, and a copy of the assessment task *Unhappy Campers*.
- Each small group of students will need a copy of the role-play script *A Case of Muddying the Waters,* and a copy of the worksheet *Case Notes*.
- There are some projector resources to support the role-play and whole-class discussion. Use of a computer and projector is recommended. You could print the images onto overhead projector slides instead. If you do not have access to a projector, print a copy of these slides for each small group of students: *Exhibits 1, 2, 3, 4, Scene 1: The Assistant DA's office,* and *Scene 2: At the Factory*.
- A teaspoon may be useful to help illustrate the concept of 'mg per cubic meter'.

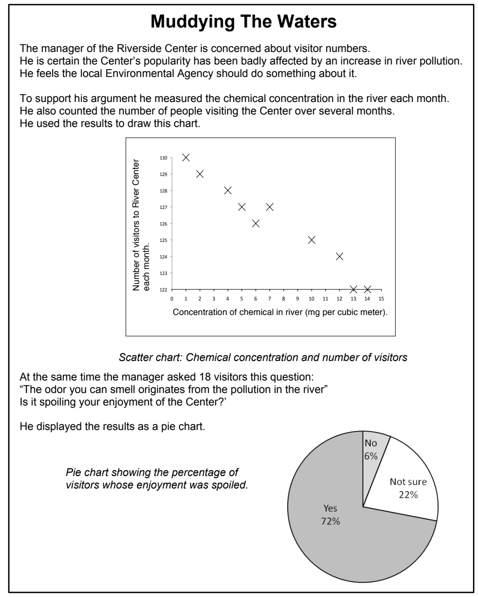
TIME NEEDED

30 minutes of one lesson, a second 1-hour lesson, and 15 minutes in a third lesson (or for homework). All timings are approximate. Exact timings will depend on the needs of the students.

SUGGESTED LESSON OUTLINE: LESSON 1

Interactive whole-class introduction (10 minutes)

Have the students do this task in class a day or more before the formative assessment lesson. This will give you an opportunity to assess the work, and to find out the kinds of difficulties students have with it. You will then be able to target your help more effectively in the follow-up lesson.


Muddying The Waters

The manager of the Riverside Center is concerned about visitor numbers.
He is certain the Center's popularity has been badly affected by an increase in river pollution.
He feels the local Environmental Agency should do something about it.

To support his argument he measured the chemical concentration in the river each month.
He also counted the number of people visiting the Center over several months.
He used the results to draw this chart.

Scatter chart: Chemical concentration and number of visitors

At the same time the manager asked 18 visitors this question:
"The odor you can smell originates from the pollution in the river"
Is it spoiling your enjoyment of the Center?'

He displayed the results as a pie chart.

Pie chart showing the percentage of visitors whose enjoyment was spoiled.

Give each student a copy of the task *Muddying the Waters*. Ask students to read through the task carefully. Some teachers ask students to take turns to read parts of the task aloud.

Introduce the task, asking questions to help students to understand the problem and its context.

This task is concerned with the river pollution and its effect on the environment. What do I mean by river pollution?

What does it mean when someone says that the level of pollution in a river is illegal?

Does anybody know of a river that is polluted? What was the source of the pollution?

How can you tell that a river is polluted?

In particular, explain how chemical pollution in a river is measured. You could use a teaspoon to help illustrate this.

Chemical pollution is measured in milligrams per cubic meter of river water.

Does anyone know how much a milligram is? A cubic meter is?

A teaspoon of sugar is about 4,000 mg. This classroom is about 300 cubic meters. [Use figures for your own room.]

Assessment task: *Muddying the Waters* (15 minutes)

Ask students to work through the task.

Spend fifteen minutes working individually on this task.

Don't worry if you can't complete everything.

There will be a lesson [tomorrow] that should help you understand the math better.

Your goal is to be able to answer questions like these confidently by the end of the next lesson.

At this stage, do not help students with the task or comment on their work. Stop them after 15 minutes, whether or not they have finished. Collect in students' papers for analysis.

Preparation for the next lesson (5 minutes)

Explain the theme for the next lesson. You need three copies of the role-play script *A Case of Muddying the Waters*.

Next lesson we are going to continue the theme of river pollution with some role-play.

The role-play sets the scene for the task. An Assistant DA talks to a local Environmental officer and a Factory owner about the pollution of a river. Who wants to play these three roles?

Give each volunteer a copy of the role-play script *A Case of Muddying the Waters*.

Ask the volunteers to read over their parts before the next lesson. If you wish, enrich the role-play by asking the volunteers to bring in a few theatrical props.

Assessing students' responses

Collect students' responses to the task. Make some notes on what their work reveals about their current levels of understanding. The purpose of this is to forewarn you of issues that will arise during the lesson itself, so that you may prepare carefully.

We suggest that you do not score students' work. The research shows that this will be counterproductive as it will encourage students to compare their scores, and will distract their attention from what they can do to improve their mathematics.

Instead, help students to make further progress by summarizing their difficulties as a series of questions. Some suggestions for these are given in the *Common issues* table (below). These have been drawn from common difficulties observed in trials of this lesson unit.

We suggest that you write a list of your own questions, based on your students' work, using the ideas below. You may choose to write questions on each student's work. If you do not have time to do this, select a few questions that will be of help to the majority of students. These can be written on the board at the beginning of the lesson.

Common issues:	Suggested questions and prompts:
Student does not understand that there are alternative interpretations of data and statistics, some of which may be biased. For example: The student copies or paraphrases the Riverside Manager's interpretations.	• What does the chart show? What does the Riverside Manager say the chart shows? Is there a difference? • In what ways is what the Riverside Manager said misleading?
Student does not recognize that things can happen together without one causing the other For example: The student does not contradict the causal claim made about the scatter plot.	• If two things happen at the same time, does that mean one made the other happen? • Write down another reason that there might be a correlation. • What other interpretations of the correlation can you find?
Student does not understand that survey questions may push respondents towards a particular response For example: The student does not recognize that the phrasing of the statement biases respondents towards thinking of the river as polluted and smelly.	• Does the way this question is asked make a 'yes' response more likely than 'no' or 'maybe'? Why do you think that? Does it matter? • Try writing this question in a way that doesn't push the respondent towards a particular answer.
Student does not recognize that statistics may be compiled in ways that push readers towards a biased interpretation For example: The student does not recognize that on the scatter plot, starting the 'number of visitors' scale at 122 (rather than 0) distorts perceptions of the proportional change in the number of visitors. Or: The student does not recognize that it is inappropriate to draw conclusions about the whole population from such a small sample size.	• Notice that the scale on this graph starts at 122. How different would the graph look if you drew the axis showing the whole range? How might that affect your interpretation? • How many visitors were there overall? When was the survey conducted? Can you think of a more convincing way to set the survey? • How do the ways data was collected affect your interpretation of the results?

SUGGESTED LESSON OUTLINE: LESSON 2

Review individual solutions to *Muddying the Waters* (10 minutes)

Remind students of their work on the assessment task.

Recall the work you did in the last lesson on river pollution.

In this lesson you will build on that work.

Return the papers to the students. If you chose to write questions on the board rather than on individual papers, display them now.

I read your papers and I've some questions about them.

I'd like you to work on improving your answers for a few minutes, using my questions.

Ask students to work on their own for a few minutes, answering your questions.

Interactive role-play introduction (10 minutes)

To introduce the lesson task, use a projector and slides P-3 to P-9 from the projector resource: *Exhibits 1, 2, 3, 4; Scene 1: The assistant DA's office; Scene 2: At the Factory*. If you do not have a projector, hand out the printed copies of these slides. You also need four copies of the role-play script, *A Case of Muddying The Waters* (one copy for each actor and one copy for yourself.)

In the next section of this lesson you will be working on river pollution again.

The role-play sets the scene.

Ask the actors to read out the script. Advise them to talk slowly, and to pause at the end of each sentence, as the script contains a lot of information. Encourage the students to listen carefully to the facts being presented about the river pollution.

Collaborative small-group work (25 minutes)

Once the students have acted out the scenes (and the applause has died down) turn to the class and say:

The case goes to court. The Assistant DA prosecutes the Factory Owner for polluting the river.

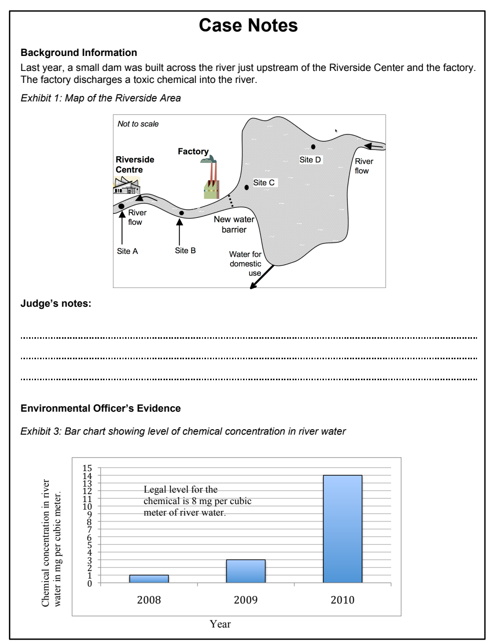
What does 'prosecution' mean?

Your task now is to be the judges. You have to reach a fair judgment about who wins the case.

Organize students into groups of two or three. Provide each small group with a copy of the worksheet *Case Notes*. *Case Notes* contains the map, information from the script, and arguments made in court by the Environmental Officer and the Factory Owner.

Help students to understand the task and its context.

I'm giving you a copy of the arguments presented in court.



Case Notes

Background Information

Last year, a small dam was built across the river just upstream of the Riverside Center and the factory. The factory discharges a toxic chemical into the river.

Exhibit 1: Map of the Riverside Area

Judge's notes:

Environmental Officer's Evidence

Exhibit 3: Bar chart showing level of chemical concentration in river water

Read through the information carefully. Write notes on what you think the data and statistics show.

In particular, ask them to focus on critical analysis of the information presented.

Explain why you agree or disagree with the arguments people made, using math.

The important thing is to look critically at all of the information. Do not just accept what people say as fact.

At the end of the lesson, you will use your work to decide together whether the factory owner is guilty, or not guilty, of polluting the river.

These instructions are reproduced on slide P-8, *Judge's Instructions*.

During small-group work you have two tasks, to notice strengths and weaknesses you see in students' work, and to support their thinking.

Notice strengths and weaknesses in students' work

Find out about students' current levels of understanding and the difficulties they encounter in the task. Students may be used to interpreting statistical diagrams, but may find it more difficult to critique someone else's biased reading of information. Students may fail to notice a bias in a question, or may struggle to understand the issue of small sample size. You can use the information about common difficulties to focus the whole-class discussion towards the end of the lesson.

Support student thinking

Try not to solve students' difficulties for them. Instead, ask them questions to help them move their thinking on.

You could strengthen your argument if you did some math on the data you've been given.

Is there another way to present this data?

Could you redraw that chart so it displays the important features of the data better?

Questions similar to those in the *Common issues* table on p. 3 were found to be useful in lesson trials.

For students who are struggling, it may help to ask some specific questions about aspects of the mathematics:

Describe this chart.

Is there another way to present this data?

The Environmental Officer/Factory Owner drew this conclusion. Can you draw any different conclusions from this evidence?

Encourage students to explain their reasoning to others in the group before writing it down. Other group members may question and refine the explanations.

Whole-class discussion: reaching a judgment (10 minutes)

Organize a whole-class discussion, focusing on the mathematical practice of critiquing the reasoning of others.

Choose a group to present their argument about one piece of evidence. Instruct the other groups to listen and write down questions about the group's argument.

Hani, does this evidence help the factory owner show he is not guilty? Tell us why you think that.

If you disagree with the group's interpretation of the evidence, write why, and challenge them at the end of their presentation.

Once the group has presented their case, other groups get a chance to challenge the details of their argument. If the challenge is not based on mathematics, you can rule it out of court.

That is not a mathematical argument. As there is no good evidence, it can't be accepted in court.

Once you have modeled this process for students, give them responsibility for deciding whether there is an evidential base for each claim and challenge.

Is that a mathematical argument? Is there good evidence for what [Shelley] has just said?

Does the evidence support her conclusion?

Once students have worked through the four pieces of evidence, ask them to come to a collective decision about the verdict.

Do you find the factory owner guilty or not guilty?

If there is disagreement, take arguments from both sides. You may find you cannot reach a collective decision. In that case, suggest students send the defendant for a retrial.

Whole-class discussion: Summing up (5 minutes)

Point out that an important message of this lesson is that it is easy to 'get it wrong' when interpreting statistics, especially in complicated real-world situations. In reality, most of the 'evidence' in this lesson is too vague to draw any firm scientific conclusions. A lot of questions are left unanswered.

How, exactly, were the wildlife surveys conducted? Why did the second survey look at so many more fish than the first? How do you 'count the number of invertebrates' at a site?

The dam has reduced the flow of the river by 80%. Even without the pollution, is it possible that this could affect the wildlife or the popularity of the Riverside Center?

Can you see any other problems with the data collection? With the statistics that have been calculated?

If you think it is appropriate to the class, you could mention that if they study statistics further they will learn how to calculate **significance**: the likelihood that a difference in two results is not just 'the luck of the draw'. Most serious scientific studies will do this, but you do not often find it in news reports!

Next lesson: assessment task *Unhappy Campers* (15 minutes)

Ask students to do this task in the next lesson, or for homework.

Give each student a copy of the assessment task, *Unhappy Campers*. Explain that this task uses very similar math to the lessons on river pollution, but in a new context.

Help students to read through the task sheet, and use questions to help them understand the context.

What is a wind turbine?

What are they used for?

What do decibels measure?

How loud is 50 decibels? 30 decibels? [0 decibels is the threshold for human hearing. A whisper in a quiet library is about 25 decibels. Normal conversation at about five feet is around 60 decibels.]

Ask students to work on their own on the assessment task, bearing in mind what they have learned during the previous lesson.

I want you to work on this task, using those same ideas about a fair, mathematical critique.

Remember not to believe all the arguments someone gives you using statistics.

After the assessment, you may find it useful to ask students to compare their responses to the first and second assessment tasks, so they can see the progress they have made.

SOLUTIONS

Assessment task: Muddying the Waters

Question 1.

Interpreting the scatter chart

- The water was tested on a monthly basis for 10 months and at the same time the number of visitors to the Center was recorded.
- There is a negative correlation.
- The number of visitors per month varies from 122 to 130. The range is 8 and the mean is 126.
- Chemical concentration varies from 1 to 14 mg per m^3. The range is 13 mg per m^3 and the mean is 7.4 mg per m^3.

Interpreting the pie chart

- Eighteen people are involved in the survey. 13 replied yes, 1 replied no and 4 were unsure.

Question 2.

The Riverside Manager's argument is misleading in various ways.

- The scatter plot has a misleading scale. It gives the impression that correlated with a rise in pollution there has been a massive drop in visitor numbers. In fact, there is a fall of only 8. Overall the decrease is 6%.
- There is a negative correlation on the scatter chart. This may not be causal as there are many other reasons why the visitor numbers fell, such as change in season. If the dam was reducing the amount of water in the river, this might have made it less attractive to visitors. The survey was over 10 months, not a year.
- The pie chart is based on a survey that uses a biased question: people may not have noticed a smell until they were asked about it.
- The sample size for the pie chart is small. The results of the survey are unlikely to be a true representation of all the visitors to the Center. Providing the number of people as well as the percentages in each response category would be helpful.

Lesson task: *Case Notes*

The concentration of the chemical in the river has risen above the legal limit.

The bar chart is appropriate, and is clearly shows that three distinct tests have been carried out. The concentration of the chemical in the river **is** now above the legal limit.

The levels were within the limit in the previous two years. The factory is discharging the same amount of chemical, but the flow rate of the water has reduced, meaning that the **concentration** is now above the legal limit.

Students may have calculated the concentration of the chemical in the water:

Last year:

$$\frac{60}{20} = 3mg/m^3.$$

This year:

$$\frac{60}{4} = 15mg/m^3.$$

Arguing that the mean concentration is within the legal limit is a misuse of statistics, because the low measurements in the first two years disguise the much higher figure in the last year:

$$\frac{1+3+14}{4} = 6mg/m^3.$$

A more sympathetic judge might argue that there are not enough test sites to prove that the pollution was not caused by another source: it might have been useful to test the water upstream of the factory, to find out whether that water was polluted to start with.

There has been an increase in the number of diseased fish due to the rise in chemical pollution.

The survey is misleading because the sample sizes are different. Arguing that there are now ten times more diseased fish is incorrect because it ignores the sample sizes. If students have calculated proportions or percentages, they will get a better sense of the data than if they rely on numbers:

Two years ago:

$$\frac{6}{300} \times 100 = 2\%$$

This year:

$$\frac{64}{1600} \times 100 = 4\%$$

Using this to argue that the number of diseased fish has doubled is still a misleading use of statistics: finding slightly more or fewer diseased fish in either survey (due to weather, the way the fish were caught, or just 'the luck of the draw') would make a big difference to the percentages.

The judge could argue that the survey is poor evidence because the sample sizes are too small to detect a difference in such a small percentage of diseased fish. Or the judge might argue that there are not enough survey sites to show whether being downstream of the factory makes a difference. The reduced flow rate of the river might have affected the health of the fish regardless of the pollution. Why did the second survey look at so many more fish than the first: were the fish harder to find the first time?

The number of invertebrates has not changed.

There has been hardly any change in the mean number of invertebrates. Two years ago the mean across four sites was 21, and now it is 19.

However, two years ago the range was 4. Now the range is 20. This is a big increase.

The sites most likely to have been affected by the pollution are A and B, downstream from the factory. Two years ago the mean number of invertebrates at these sites was 21, now it is 12. This is quite a large decrease. In contrast, the mean at sites C and D has increased.

Arguing that the mean number of invertebrates has hardly changed is a misuse of statistics: taking the mean of all four sites (including two which would not have been affected by pollution from the factory) hides the possibly significant reduction at the polluted sites.

The number of birds has increased.

Using a line graph to represent this data is inappropriate because it gives the impression that the birds were continually monitored. A bar chart with two bars would be more appropriate.

The scale on the line graph is misleading because it gives the impression that there has been a dramatic increase in the number of birds. The data show that there has only been an increase of 6 birds (about 7%). This is insignificant, especially without more details of how the birds were counted or at what time of year.

Arguing that the chart shows that the number of birds has increased dramatically in the last two years is a misleading use of statistics.

Assessment task: *Unhappy Campers*

Question 1.

Interpreting the scatter chart:

- There are fourteen data points on the scatter chart; the survey took place over a two-week period. There is a negative correlation.
- The number of visitors ranges from 70 to 78, a range of 8 with a mean average of 75. The noise level ranges from 10 to 60 decibels, with a range of 50 decibels and a mean average of 35 decibels[1].

Interpreting the pie chart:

- The number of campers surveyed was 50.
- 80% of the sample responded yes, 16% unsure, and 4% no.
- The numbers of respondents are 40, 8, and 2 respectively.

Question 2.

The camp manager's argument is biased in several ways.

Her choice of math introduces bias:

- The scatter plot has a misleading scale. The scale on the 'number of campers' axis starts at 40 rather than 0. It gives the impression that correlated with the rise in noise level there has been a large drop in visitor numbers. The number of campers only varies by 8 across the fourteen-day period, decreasing by about 10% between the quietest and noisiest day.
- The survey statement and question is biased. Stating that the noise is 'loud' and assuming that the respondent can hear the noise pushes the respondent towards a positive response; the use of 'spoiling enjoyment' in the question also introduces potential bias.
- The pie chart is based on a relatively small sample (50 campers). The survey took place on one day. On only 2 days on the scatter chart were there 50 campers. Both days were particularly noisy. Surveying only on a noisy day produces potential bias in the survey responses. It would have been helpful to show the number of respondents, not just percentages, on the pie chart, to aid interpretation of the results.

Her interpretations of her data and statistics are incorrect.

Peggy claims that the noise from the turbines has **caused** a drop in camper numbers. The correlation between the noise level in decibels and the number of campers does not show there is a causal relationship between the two variables. There may be other explanations of why the number of campers and the noise level correlate. For example, the turbine noise increases with the wind level, so you would expect fewer people to want to camp at noisy times, because it is windier then.

Although most of the people surveyed did state that the wind turbines spoilt their enjoyment, the questionnaire was biased, the sample was small, and the survey took place on a noisy day. As the results of the survey are dubious, there is no evidence to support Peggy's interpretation that most people coming to the camp would agree with the results of the survey. She cannot generalize from a small, biased sample, and she cannot rely on responses to a biased question.

[1]From the US Environmental Protection Agency website:

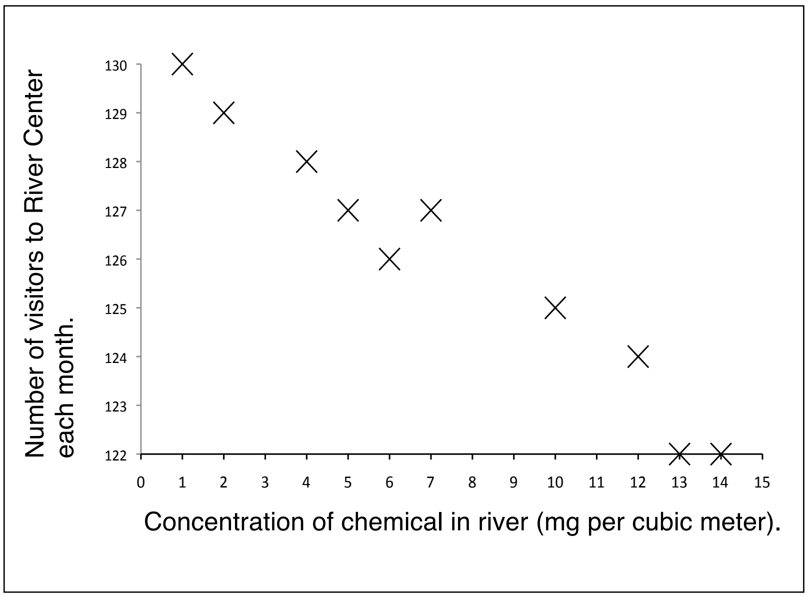
The document identifies a 24-hour exposure level of 70 decibels as the level of environmental noise which will prevent any measurable hearing loss over a lifetime. Likewise, levels of 55 decibels outdoors and 45 decibels indoors are identified as preventing activity interference and annoyance. These levels of noise are considered those which will permit spoken conversation and other activities such as sleeping, working and recreation, which are part of the daily human condition.

http://www.epa.gov/history/topics/noise/01.htm

Muddying The Waters

The manager of the Riverside Center is concerned about visitor numbers.
He is certain the Center's popularity has been badly affected by an increase in river pollution.
He feels the local Environmental Agency should do something about it.

To support his argument he measured the chemical concentration in the river each month.
He also counted the number of people visiting the Center over several months.
He used the results to draw this chart.



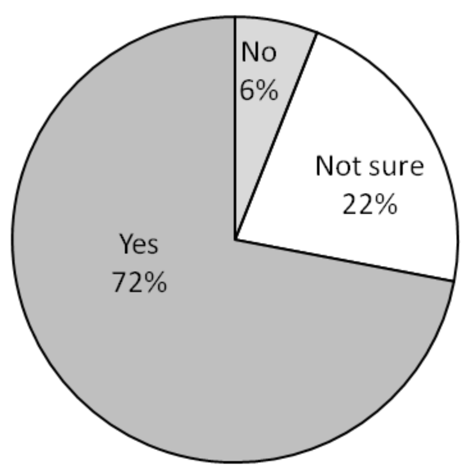
Scatter chart: Chemical concentration and number of visitors

At the same time the manager asked 18 visitors this question:

'The odor you can smell originates from the pollution in the river.
Is it spoiling your enjoyment of the Center?'

He displayed the results as a pie chart.

Pie chart showing the percentage of visitors whose enjoyment was spoiled.



The centre manager writes to the Environmental Officer to try to get something done about the river pollution.

> Dear Environmental Officer,
>
> Please find enclosed two charts.
>
> The scatter plot clearly shows that the increase in the concentration of the chemical in the river has caused a real drop-off in visitor numbers to the Center over the last year.
>
> The pie chart proves that people (not surprisingly) don't like the acrid smell of pollution wafting up from the river.
>
> The river needs to be cleaned up; it's not good for the environment and it's certainly not good for my business. Please let me know what action you intend to take.
>
> Yours faithfully,
>
> Manager, Riverside Center

Tasks

1. Describe in detail what you think the two charts show.

----

2. Do you think the Riverside Center Manager's argument is fair?
Explain your reasoning.

A Case Of Muddying The Waters: Role-Play Scripts

Scene 1: Environmental Officer talking to the Assistant DA in the DA's office

Teacher	A year ago a small dam was built across the river to provide extra water for local domestic use. It was positioned just upstream of the Riverside Center and a factory [*show class **Exhibit 1** and point out the landmarks*]. This factory continually discharges a toxic chemical into the river. Prompted by the letter [*show class **Exhibit 2***] from the manager of the Riverside Center, the local Environmental Officer [*indicate the student playing this role*] checks the concentration of the chemical in the river. Unhappy with the result he arranges to meet the Assistant DA at his office [*indicate the student playing this role*]. Here is how the meeting went *[show class the photograph of the DAs office*]:
Environmental Officer	Each year I test the river water for this toxic chemical, and up until this year the concentration levels have been fine.
Assistant DA	But there's a problem now?
Environmental Officer	That's right. As you can see from this chart, [*show class **Exhibit 3***] the concentration of the chemical in the river is now disturbingly high.
Assistant DA	What can account for this change?
Environmental Officer	Well, downstream of the barrier the flow of the river has decreased from 20 to 4 cubic meters per second.
Assistant DA	Go on.
Environmental Officer	This decrease does affect the concentration of the chemical in the river. To calculate this concentration you divide the chemical discharge from the factory by the rate of flow of the river.
Assistant DA	Oh yeah, right [*not understanding a word*]. Have you noticed any other changes?
Environmental Officer	Yes, the number of diseased fish in the river. Two years ago we only found 6 diseased fish in the waters near the factory, but just last week we found 64. I'm sure you'll agree that's a massive increase.
Assistant DA	Mmm…. I think I better take a ride out to the factory and see what the owner has to say about all this.

Scene 2: Assistant DA interviewing the Factory Owner in the Factory Owner's office

Show the photograph of the factory.

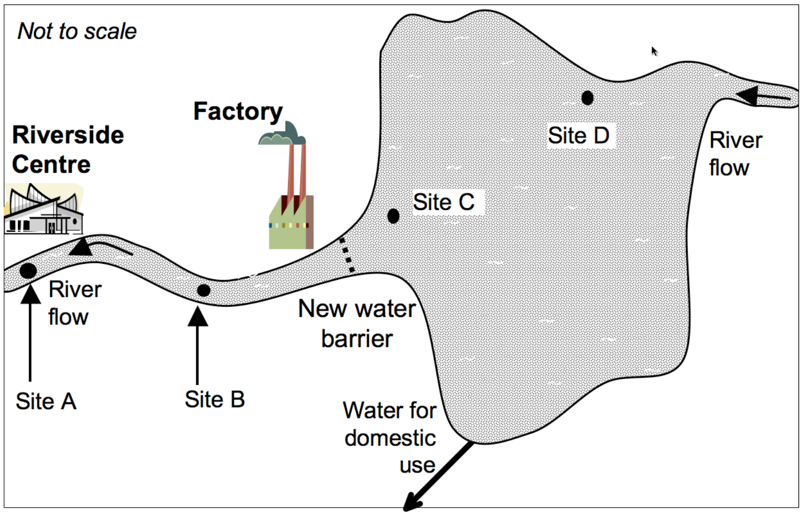
Teacher	So the assistant DA then drove out to the factory to interview its owner [*indicate the student playing this role to the class*].
Assistant DA	Do you know you are discharging a toxic chemical into the river?
Factory owner	Sure we know that, but we're only discharging it at the rate of 60 milligrams per second. When I first opened the factory some years ago, my manager checked that this was within the legal limits, and this rate has not changed since then.
Assistant DA	I see. But the fact is, the concentration of the chemical in the river is now above the legal limit, and you are the only factory in the area producing the stuff.
Factory owner	Well, it's a mystery to me. Something else must have caused the increase.
Assistant DA	I hear what you're saying, but I still think you're responsible.
Factory owner	Do you reckon – because I don't. What is important, though, is the health of the river. As you probably know, a good sign of a healthy river is the variety of the invertebrates in it. I've been keeping an eye them in our river.
Assistant DA	What, you mean you've actually been counting the number of snails and water insects in the river?
Factory owner	Well yes, my people have counted all animals without a backbone. They've been counted at four different sites and the average number has barely changed in the last two years [*show class Exhibit 1 and point to the 4 sites*]. That's a good sign, don't you think?
Assistant DA	That is interesting. Have you checked anything else?
Factory owner	I've also kept a close watch on the number of birds around the factory, and, as you can see from the chart, there has been a dramatic increase [*show class Exhibit 4*]. Another sure sign of a healthy river.
Assistant DA	Well, I will get my experts to check all this. But from what I've heard, there is a case to answer. So I'll see you in court!

Case Notes

Background Information

Last year, a small dam was built across the river just upstream of the Riverside Center and the factory. The factory discharges a toxic chemical into the river.

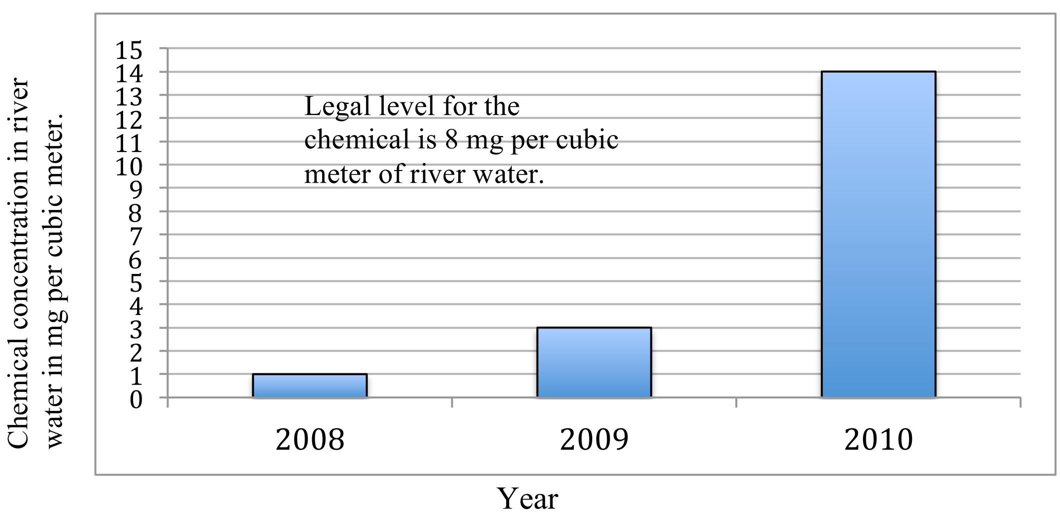
Exhibit 1: Map of the Riverside Area



Judge's notes:

Environmental Officer's Evidence

Exhibit 3: Bar chart showing level of chemical concentration in river water



"Each year, I test the river water for this toxic chemical, and up until this year the concentration levels have been fine.

The chart shows that the concentration of the chemical in the river is now disturbingly high.

Downstream of the barrier, the flow of the river has decreased from 20 to 4 cubic meters per second. This affects the concentration of the chemical in the river.

To calculate this concentration you divide the chemical discharge from the factory by the rate of flow of the river."

Judge's notes:

Factory Owner's Evidence

"We are discharging a toxic chemical, but only at the rate of 60 milligrams per second.

When I first opened my factory the manager checked that this is within the legal limits.

The rate has not changed since then.

If the increase is now outside the legal limits, someone else must have caused the change."

Judge's notes:

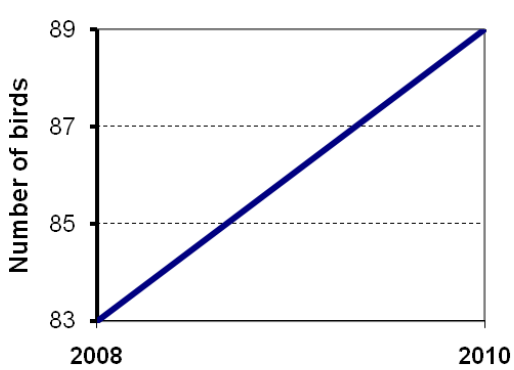
Factory Owner's Evidence

"I've kept a close watch on the number of birds around the factory.

You can see from the chart that there has been a dramatic increase in the number of birds around here.

This is another sure sign of a healthy river."

Exhibit 4: Survey of the number of birds seen next to the factory in one afternoon



Judge's notes:

Environmental Officer's Evidence

"There has been an increase in the number of diseased fish in the river."

Exhibit 5: Survey of diseased fish near site A

Both surveys covered a period of five days.

- Two years ago 6 fish out of 300 were diseased.
- Last week 64 fish out of 1,600 were diseased.

Judge's notes:

The Factory Owner's Evidence

"It's important to think about the health of the river.
A sign of a good healthy river is the variety of invertebrates in it.
I've been counting the number of animals without a backbone – things like snails and water insects."

Exhibit 6: Survey of numbers of invertebrates

	Two years ago	Now
Site A	20	15
Site B	22	9
Site C	19	23
Site D	23	29

"Invertebrates were counted at four different sites.
On the map (Exhibit 1), you can see where the sites are.
The average number of invertebrates has barely changed in the last two years - a good sign."

Judge's notes:

Judge's summing up:

Unhappy Campers

Peggy is the manager of a campground.

Some new wind turbines are built next to the camp.

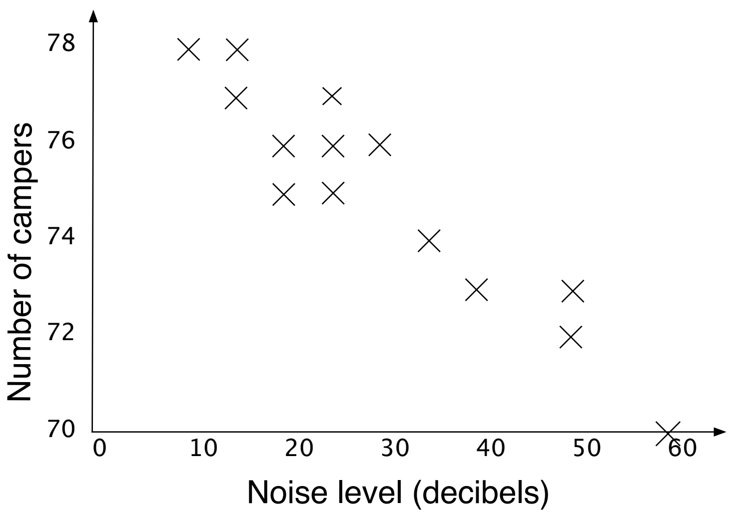
Peggy feels sure fewer people are visiting her site because of the turbines.

Peggy wants the Environmental Agency to get rid of the turbines.

To make her case she does some math.

For two weeks, Peggy records the number of visitors to the campsite each day. She also records the highest level of noise the turbines make each day.

She uses the information to draw a scatter chart.



Peggy also conducts a survey of the 50 campers who visit one day.

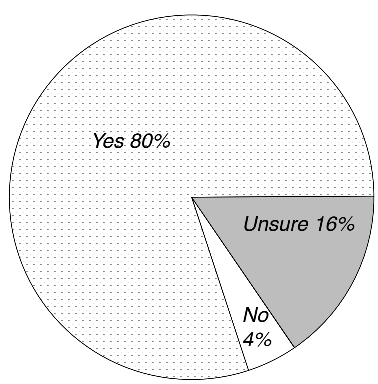
She asks the campers this question:

The loud noise you can hear comes from turbines on the wind farm.

Is the wind farm spoiling your enjoyment of the campsite? *Yes* *No* *Unsure*

Peggy uses the survey results to draw a pie chart.

Pie chart showing percentage of campers whose enjoyment was spoiled by turbine noise.



Peggy writes to the Environmental Officer.

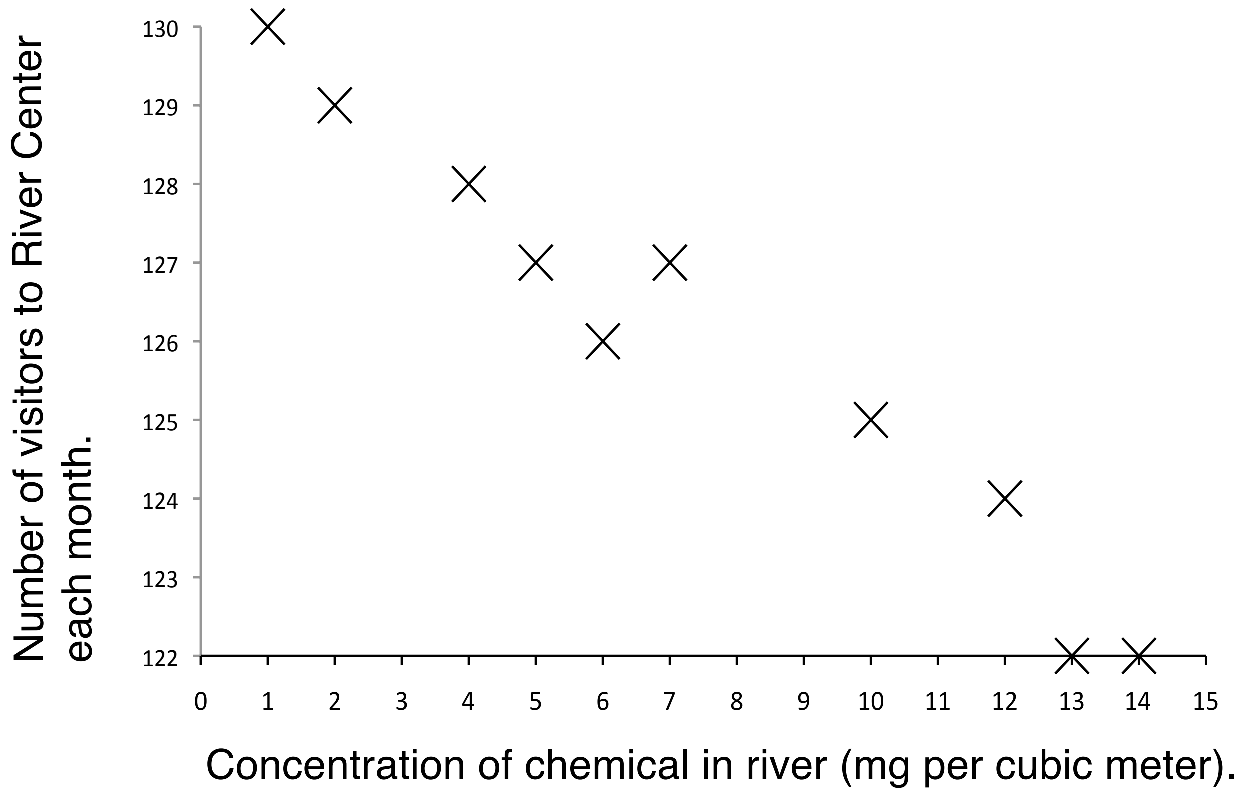
> Dear Environmental Officer,
>
> Please find enclosed two charts.
>
> The new wind farm is noisy and ugly, and I have evidence to prove it.
>
> The scatter chart shows that the wind farm drives campers away. The noise causes fewer people to come to camp.
>
> The pie chart shows that most people don't like the wind farm. It's affecting their enjoyment of the camp.
>
> You should close the wind farm because it is spoiling the local environment.
>
> Yours faithfully,
>
> Camp Manager

1. Describe what you think is shown by the scatter graph and pie chart.

----

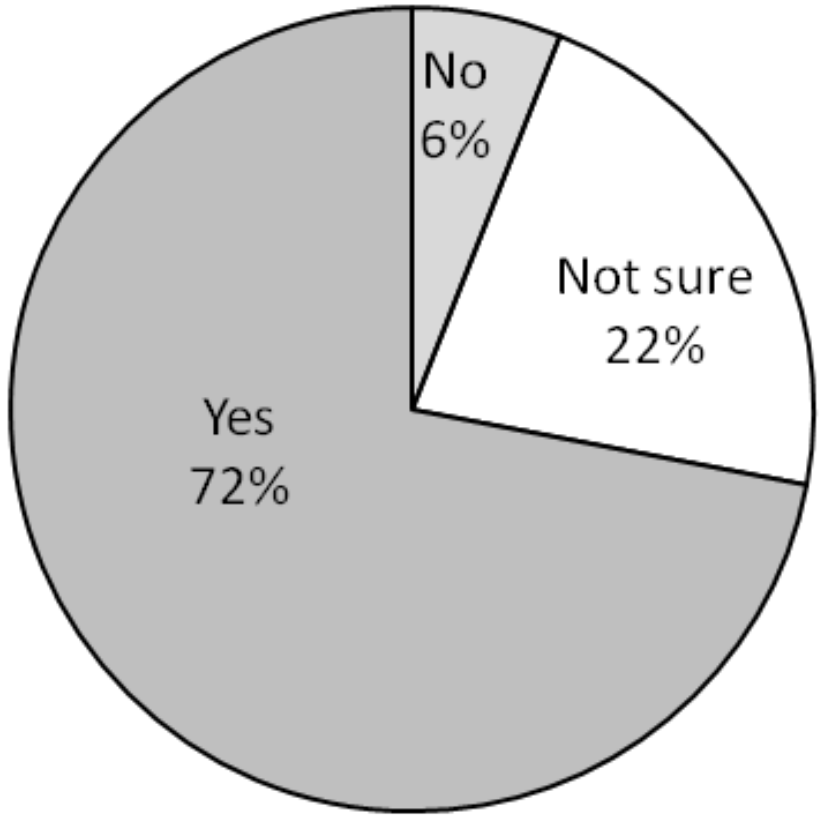
2. Do you think the camp manager makes a fair argument? Explain your answer.

Muddying the Waters: Scatter Chart



Scatter chart: Chemical concentration and number of visitors.

Muddying the Waters: Pie Chart



Pie chart showing the percentage of visitors whose enjoyment was spoiled.

Exhibit 1: Map of Riverside Area

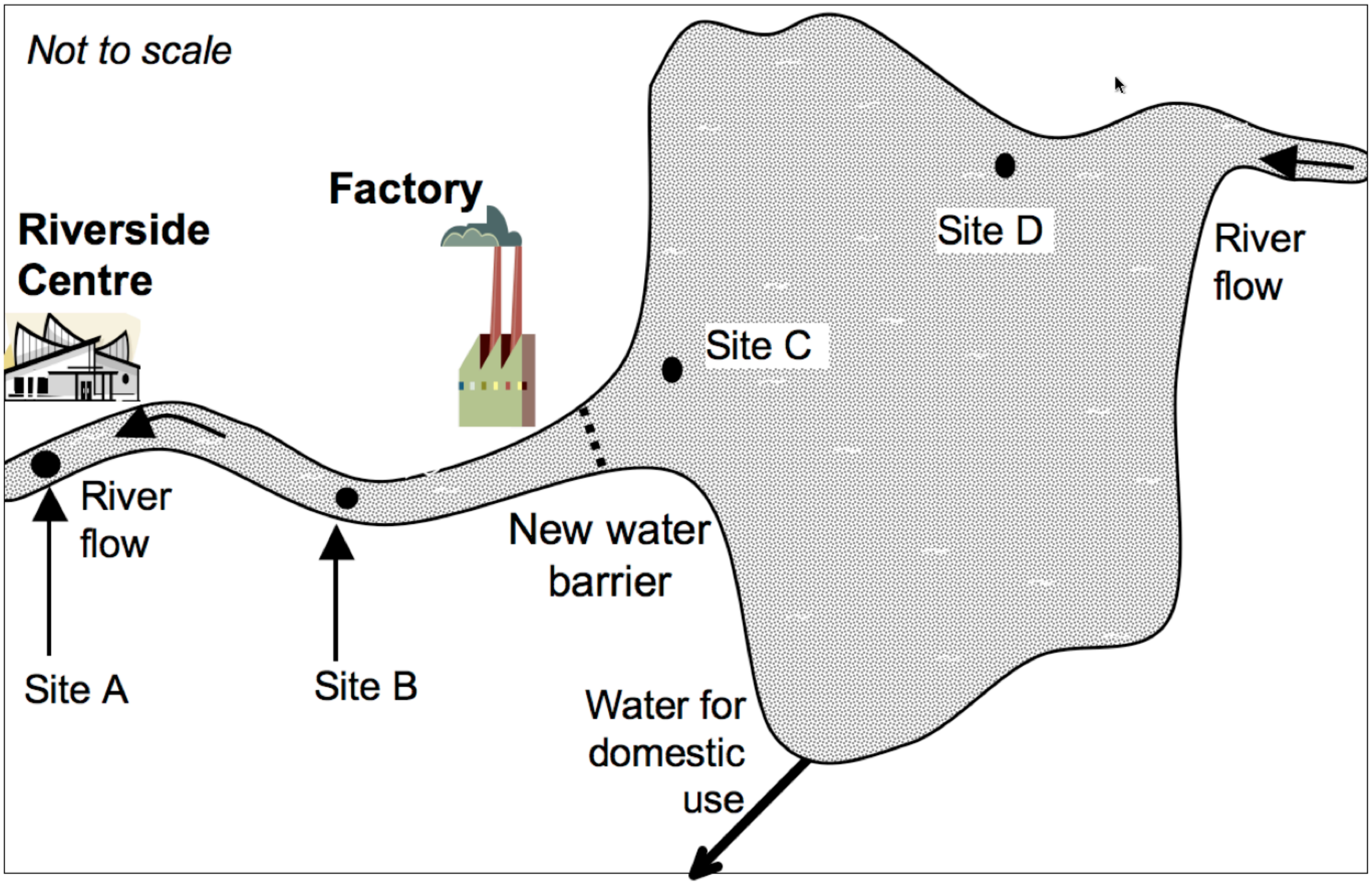


Exhibit 2: Riverside Manager's Letter

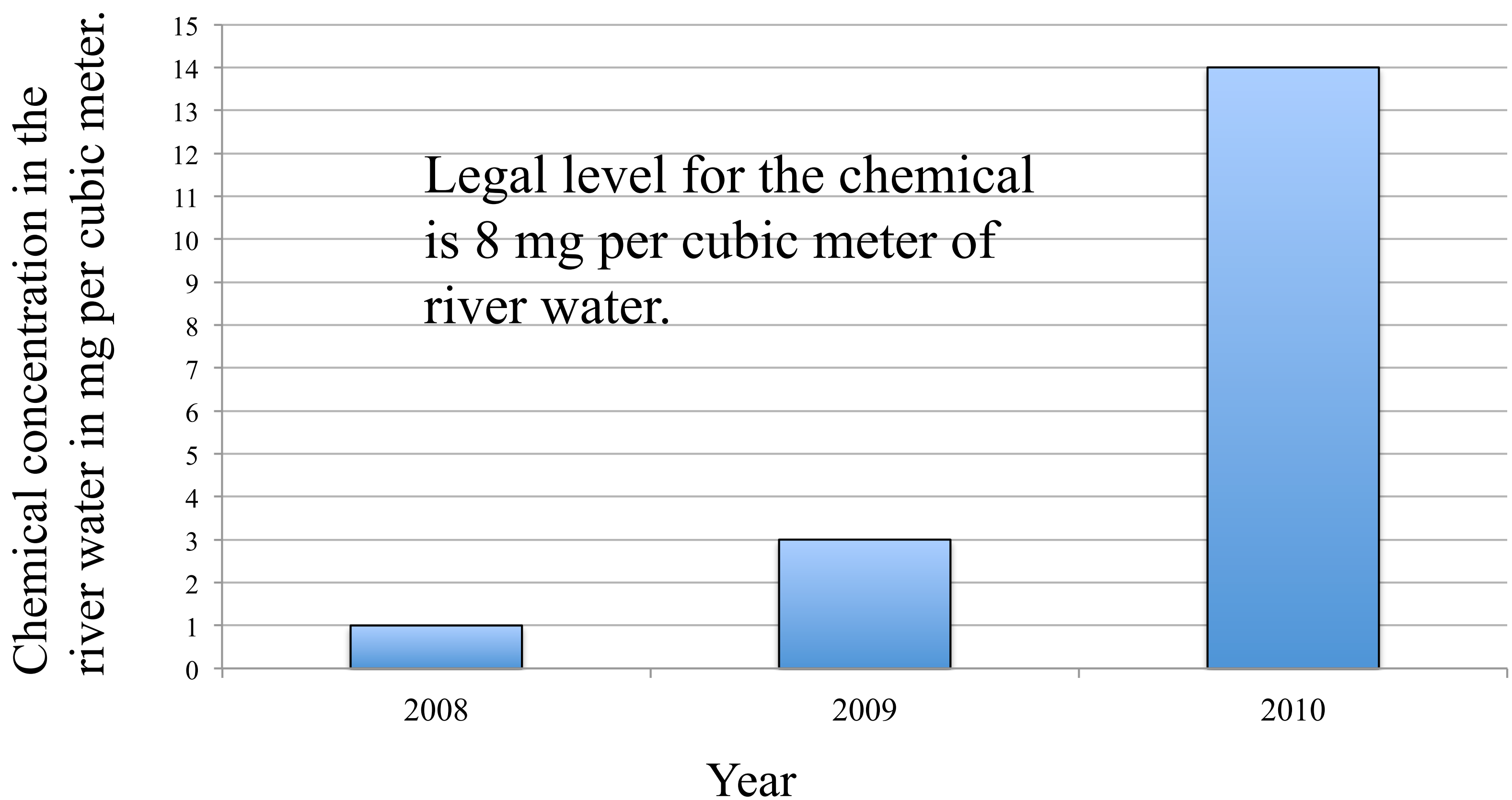
Dear Environmental Officer,
Please find enclosed two charts.
The scatter plot clearly shows that the increase in the concentration of the chemical in the river has caused a real drop-off in visitor numbers to the Center over the last year.
The pie chart proves that people (not surprisingly) don't like the acrid smell of pollution wafting up from the river.
The river needs to be cleaned up; it's not good for the environment and it's certainly not good for my business.
Please let me know what action you intend to take.
Yours faithfully,
Manager, Riverside Center

Scene 1: The Assistant DA's Office

Scene 2: At the factory

Exhibit 3

Bar chart showing level of chemical concentration in the river water.



Judge's Instructions

You have to reach a fair judgment about who wins the case.

- Read through the information carefully.
- Write notes on what you think the data and statistics show.
- Explain why you agree or disagree with the arguments, using math.

- Look critically at all of the information.
- Do not just accept what people say as *fact*.

Exhibit 4

Survey of the number of birds seen next to the factory in one afternoon.

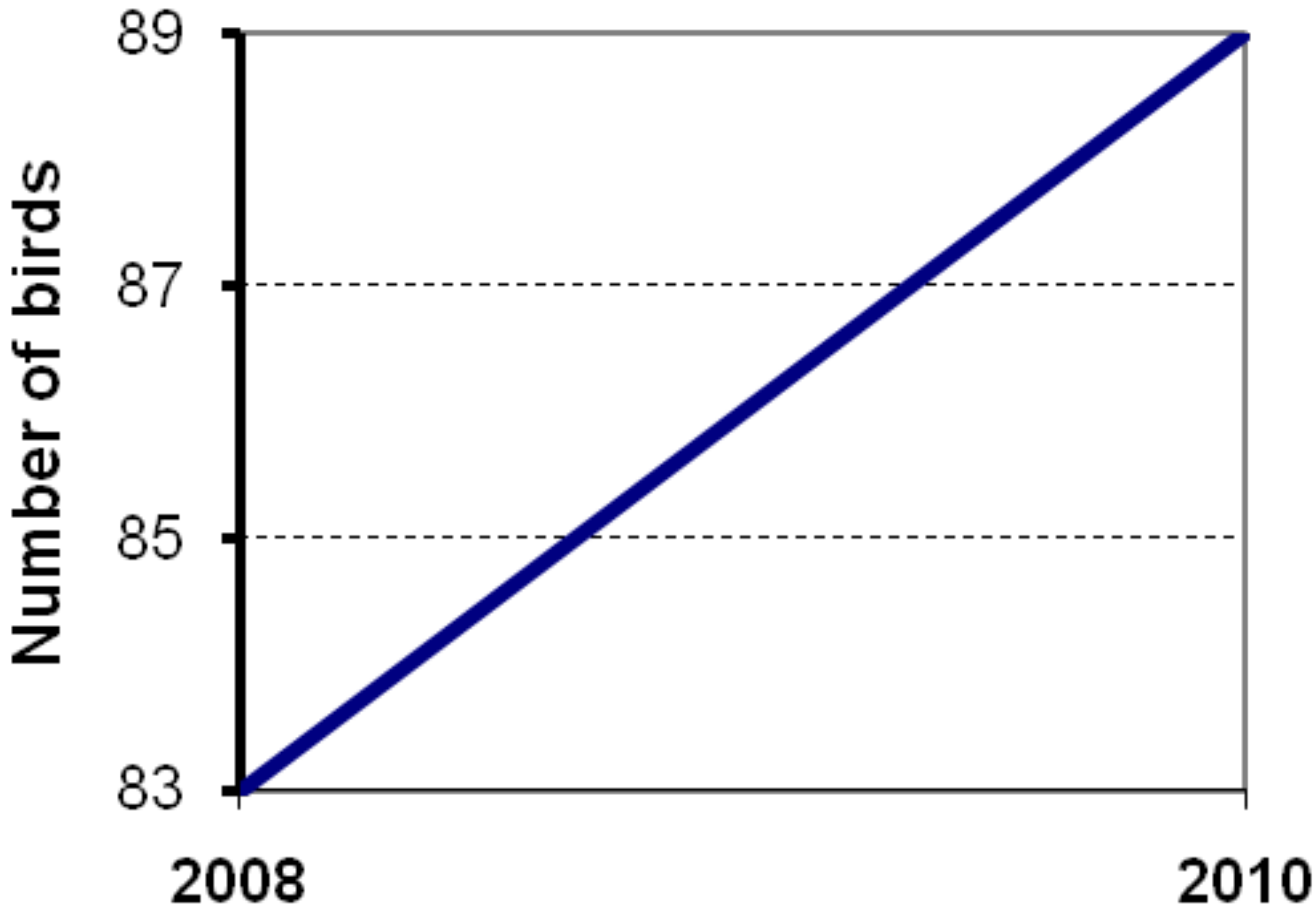


Exhibit 5

Survey of diseased fish near site A

Both surveys covered a period of five days.

- Two years ago 6 fish out of 300 were diseased.
- Last week 64 fish out of 1,600 were diseased.

Exhibit 6

Survey of numbers of invertebrates

	Two years ago	Now
Site A	20	15
Site B	22	9
Site C	19	23
Site D	23	29

Mathematics Assessment Project

CLASSROOM CHALLENGES

This lesson was designed and developed by the
Shell Center Team
at the
University of Nottingham
Malcolm Swan, Nichola Clarke, Clare Dawson, Sheila Evans
with
Hugh Burkhardt, Rita Crust, Andy Noyes, and Daniel Pead

It was refined on the basis of reports from teams of observers led by
David Foster, Mary Bouck, and Diane Schaefer
based on their observation of trials in US classrooms
along with comments from teachers and other users.

This project was conceived and directed for
MARS: Mathematics Assessment Resource Service
by
Alan Schoenfeld, Hugh Burkhardt, Daniel Pead, and Malcolm Swan
and based at the University of California, Berkeley

We are grateful to the many teachers, in the UK and the US, who trialed earlier versions
of these materials in their classrooms, to their students, and to
Judith Mills, Carol Hill, and Alvaro Villanueva who contributed to the design.

This development would not have been possible without the support of
Bill & Melinda Gates Foundation
We are particularly grateful to
Carina Wong, Melissa Chabran, and Jamie McKee

Investigation 4.4
Can You Roll Your Tongue?

Overview

This investigation focuses on students examining an **association** between two **categorical variables**. Specifically, they will investigate whether there is an association between gender and whether a person can roll their tongue. As part of this investigation, students will collect, organize, and analyze data in a **two-way table**; construct and analyze **segmented bar graphs**; and calculate the **percentages** of boys and girls who can roll their tongue. This investigation is based on an activity in *Probability Through Data*, a module in the Data-Driven Mathematics series (1999).

Association

Two categorical variables are associated if certain values of one variable are more likely to occur with certain values of the other variable.

GAISE Components

This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* The four components are formulate a statistical question that can be answered with data, design and implement a plan to collect appropriate data, analyze the collected data by graphical and numerical methods, and interpret the results of the analysis in the context of the original question. This is a GAISE Level B activity.

Learning Goals

Students will be able to do the following after completing this investigation:

- Organize data collected into a two-way table
- Analyze data in a two-way table

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

Common Core State Standards Grade Level Content

6RP3c Find a percent of a quantity as a rate per 100; solve problems involving finding the whole, given a part and the percent.

6SP3 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.

8.SP.4 Understand that patterns of association also can be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables.

Principles and Standards for School Mathematics

Data Analysis and Probability

Grades 6-8 Students should understand and use ratios and percentages to represent quantitative relationships and formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population.

Materials

- Data collection sheet (available on the CD)
- Data recording sheet (available on the CD)
- Grid paper
- Color markers

Estimated Time

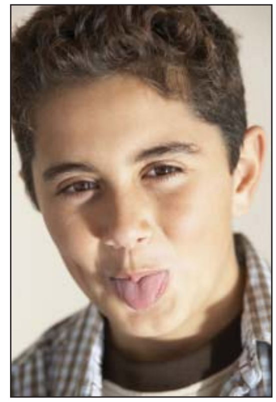
One day

Instructional Plan

✏ Formulate a Statistical Question

Ask your students to look around their classroom. Pose the question, "Is anyone in the room exactly like you?" "Are identical twins exactly the same?" Discuss that there are many traits or characteristics that make us different from each other. Have students list some of these traits. Examples are hair color, eye color, skin color, blood type, having double-jointed elbows, having "free" earlobes or "attached" earlobes, and whether they can roll their tongue. Discuss with your students that many of these traits are genetic (i.e., inherited or passed on from their parents). Ask which of the traits might have been inherited from their parents.

Tell students there are many traits they could investigate. Indicate that, for this activity, they will be investigating rolling one's tongue (even though it isn't genetic). The statistical question is, "Is gender associated with ability to roll one's tongue?"


Can't roll tongue


Can roll tongue

✏ Collect Appropriate Data

1. Have one student demonstrate how he/she is able to roll his/her tongue and another demonstrate that he/she is unable to roll his/her tongue.

2. Hand out a data collection sheet to each student. Your students should check whether they are a boy or girl and whether they can roll their tongue. Collect each of the data collection sheets. Figure 4.4.1 is an example of a data collection sheet.

Boy ________

Girl ________

Can roll tongue ________

Can't roll tongue ________

Figure 4.4.1 Data Collection Sheet

3. Hand out a recording sheet (available on the CD) to each student. Take each of the data collection sheets and read whether the sheet is checked boy or girl and whether the student can roll their tongue. As you read each data collection sheet, students should record the data on the recording sheet as shown in Table 4.4.1. Suggest that they write B for boy, G for girl, Y for yes they can roll their tongue, and N for no they cannot roll their tongue.

Table 4.4.1 Example of Class Recording Sheet

Student	Boy or Girl	Roll Your Tongue Yes or No?
1	B	N
2	B	Y
3	G	Y
…	…	…

Analyze the Data

1. Discuss with your students that one way to help analyze the data is to organize the data into a table. Ask them what answers they could record when they were reading the data collection sheets. On the board, display Table 4.4.2. Ask your students to fill in the frequencies (counts) for the four possibilities based on their recording sheet.

Table 4.4.2 Frequency Table

Possibilities	Count/Frequency
Boy – Yes	
Boy – No	
Girl – Yes	
Girl – No	
Total	

2. Explain to your students that their frequency table can be displayed in a different way, called a two-way table. A two-way table organizes data about two categorical variables with rows labeled with the categories of one variable and the columns labeled with the categories of the other variable. In this investigation, the rows of the table are labeled with gender—boys and girls—and the columns are labeled with whether a person can roll their tongue. Demonstrate drawing and labeling the two-way table. The general form is shown in Table 4.4.3. Note that the two-way format is useful when investigating whether there is an association between two categorical variables.

Table 4.4.3 Two-Way Table

	Yes – Can Roll Tongue	No – Can't Roll Tongue	Total
Boy			
Girl			
Total			

3. Label each cell in Table 4.4.3 with letters representing frequencies, as shown in Table 4.4.4.

Table 4.4.4 Example of Completed Two-Way Table

	Yes – Can Roll Tongue	No – Can't Roll Tongue	Total
Boy	a	b	
Girl	c	d	
Total			

4. Explain to your students that the cell labeled "a" will contain the number of students who are both a boy and who said they could roll their tongue. Ask your students what the cell labeled "b" represents. Cell "c"? Cell "d"?

5. Ask your students how many boys are in the sample, using the letters in Table 4.4.4. **Note:** There are "a+b" boys. How many girls? There are "c+d." How many students can roll their tongues? "a+c" can roll their tongues. How many can't? "b+d" can't.

6. Have your students fill in the two-way table based on their class data as recorded in their frequency table, Table 4.4.2. An example of what their table may look like is given in Table 4.4.5.

Table 4.4.5 Row of the Boys' Data from the Two- Way Table

	Yes – Can Roll Tongue	No – Can't Roll Tongue	Total
Boy	8	7	**15**
Girl	6	4	**10**
Total	**14**	**11**	**25**

7. Ask your students to use Table 4.4.5 to answer the following questions. As students answer each question, have them point to the appropriate cell.

    a. How many students were in the class?

    b. How many students could roll their tongue?

    c. How many students were girls?

    d. How many students were boys?

    e. How many girls could roll their tongue?

    f. How many boys could roll their tongue?

    g. How many boys could not roll their tongue?

8. Remind your students of the question they are investigating: "Is gender associated with ability to roll one's tongue?" Ask them if they are ready to answer the question. Note that many of your students will say more boys can roll their tongues than girls. Keep asking until someone suggests they should be looking at percentages, not raw counts, as there are more boys than girls in the class.

9. Ask your students to find the percentage of boys who could roll their tongue. To help them answer this question, show them only the row with the boys' data. See Table 4.4.6. Have them find the fraction that answers the question, convert it to a decimal, and then convert it to a percentage. For example, for the boys who can roll their tongue, 8/15 = .53 = 53%.

Table 4.4.6 Row of the Boys' Data from the Two-Way Table

	Yes – Can Roll Tongue	No – Can't Roll Tongue	Total
Boy	8	7	15

10. Using Table 4.4.7, ask your students to find the percentage of girls who can and cannot roll their tongue.

Table 4.4.7 Row of the Girls' Data from the Two-Way Table

	Yes – Can Roll Tongue	No – Can't Roll Tongue	Total
Girl	6	4	10

11. Ask your students to put their percentages in a two-way table. See Table 4.4.8. Note that the Total row percentages are each 100%.

Table 4.4.8 Example of Row Percentages

	Yes – Can Roll Tongue	No – Can't Roll Tongue	Total
Boy	8/15 = .53 = 53%	7/15 = .47 = 47%	15/15 = 1.00 = 100%
Girl	6/10 = .60 = 60%	4/10 = .40 = 40%	10/10 = 1.00 = 100%
Total			

12. To help your students visualize the different percentages of boys and girls who can and cannot roll their tongue, demonstrate the construction of a segmented bar graph. Using Table 4.4.8, a segmented bar graph is shown in Figure 4.4.2. Note that the percentages could also be visualized in side-by-side bar graphs.

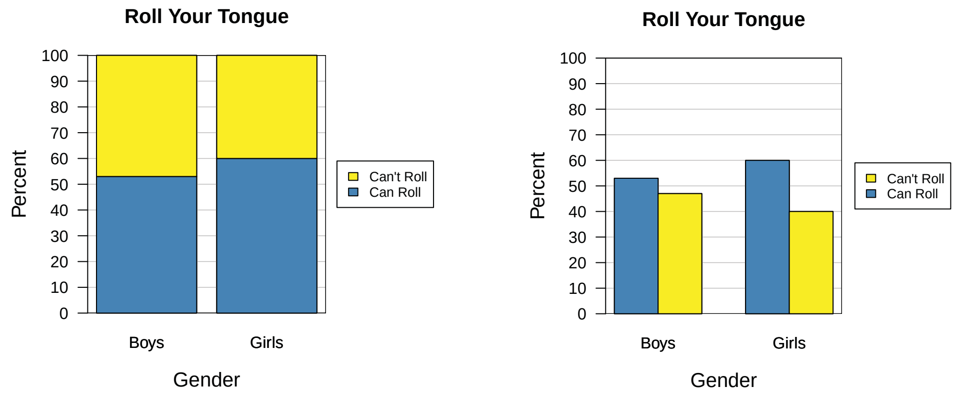


Figure 4.4.2 Segmented bar graph and side-by-side bar graph of example class data

Interpret the Results in the Context of the Original Question

1. Have your students recall the original statistical question, "Is gender associated with ability to roll one's tongue?" Have each group of students write an answer to the question and then justify it using the two-way table, appropriate calculations involving percentages, and the segmented bar graph. Suggest to your students that they should focus on the difference in the percentages and the heights of the bars in the segmented bar graph. Remind your students that an association exists between two categorical variables if knowing the response of one of the variables helps to know what the response might be of the other variable. Does knowing a girl was chosen from the group help know whether she can roll her tongue? Similarly, does knowing a boy was chosen help know whether he has the ability to roll his tongue? Have each group of students present their results to the class.

Example of 'Interpret the Results'

Note: The following is not an example of actual student work, but an example of all the parts that should be included in student work.

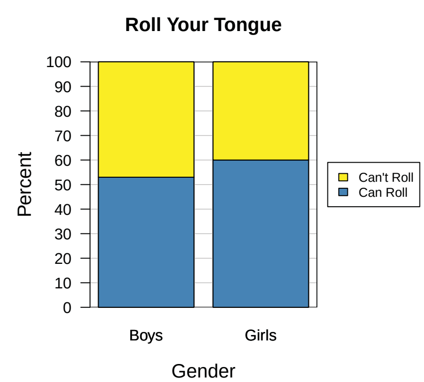
In our biology class, we often talk about genetics, so we thought a good statistics project in our mathematics class would be to take a genetic trait and see if it is associated with gender. We chose rolling our tongues. (After our study was complete, we found out that rolling one's tongue is not actually genetic. It is a learned trait. But it was fun doing the experiment anyway.) Our statistical question was "Is gender associated with ability to roll one's tongue?" We collected data by making a list of boys or girls and whether they could roll their tongue. We then counted how many there were in each of the four categories and organized the data in a two-way table like this one.

	Yes – Can Roll Tongue	**No – Can't Roll Tongue**	**Total**
Boy	8	7	15
Girl	6	4	10
Total	**14**	**11**	**25**

So, to answer the question, some of us say boys are more likely to roll their tongues than girls are. But, we messed up because there were more boys in class than girls. So, we should be looking at percentages, not counts. When we calculated the percentages, we almost based them on 25, but realized they had to be calculated within boys' and girls' totals. So, here is our table of row percentages.

	Yes – Can Roll Tongue	**No – Can't Roll Tongue**	**Total**
Boy	8/15 = .53 = 53%	7/15 = .47 = 47%	15/15 = 1.00 = 100%
Girl	6/10 = .60 = 60%	4/10 = .40 = 40%	10/10 = 1.00 = 100%
Total			

The actual answer to our question is that a higher percentage of girls can roll their tongues as compared to boys. Sixty percent of girls could roll their tongues compared to 53% of boys. Our teacher showed us how to visualize these results in what is called a segmented bar graph. It makes it clear that the percentage of girls is higher.



But we debated whether gender and ability to roll one's tongue are associated because some of us thought that 53% and 60% are kind of close and so the variables are not associated. Others thought the percentages were far enough apart to claim the variables are associated. Our teacher said we will learn more about association in high school.

Assessment with Answers

A survey asked a group of students if they participated in a sport and if they played a musical instrument. Table 4.4.7 shows the survey results.

Table 4.4.7 Survey Results

	Music Yes	Music No	Total
Sport Yes	18	2	20
Sport No	8	22	30
Total	**26**	**24**	**50**

Use the table to answer the following questions:

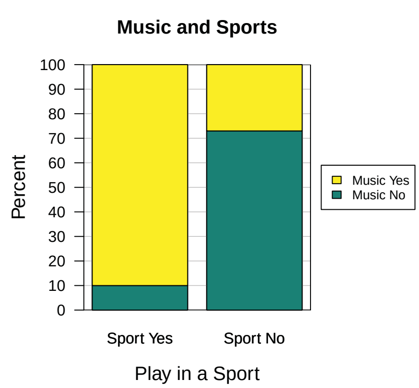
1. How many students said they participated in a sport? Twenty said they participated in a sport.

2. How many students said they did not play a musical instrument? Twenty-four said they did not play a musical instrument.

3. What does the number 8 represent in the table? The number 8 represents the number of students who said no to sports and yes to music.

4. What percentage of those who said they participated in a sport also played a musical instrument? 18/20 = .90.
5. What percentage of those who said they did not participate in a sport played a musical instrument? 8/30 = .27.

6. If a student participates in a sport, are they more likely to play a musical instrument than a student who does not participate in a sport? Use words, numbers, and graphs to explain your answer.



Students who do not participate in a sport are much less likely to play a musical instrument than those who do participate in a sport. Twenty-seven percent of students who do not participate in a sport also played an instrument while 90%

of those that did participate in a sport played an instrument. The segmented bar graph shows the big difference between the groups who do and don't play sports and whether they play an instrument. We can say that participation in sports and playing a musical instrument are associated.


Right-thumbed


Left-thumbed

Extensions

1. Ask students to collect data at home. Each student should ask one parent/guardian if he/she could roll his/her tongue. Collect data in a table during the next class period:

Possibilities	Number
Student yes – Parent/guardian yes	
Student yes – Parent/guardian no	
Student no – Parent/guardian yes	
Student no – Parent/guardian no	

   Your students should organize the data in a two-way table. Based on the table and calculated percents, students should determine if there appears to be an association between whether the parent/guardian can roll his/her tongue and whether the student can roll his/her tongue.

2. Your students could investigate if there appears to be an association between whether a person is left-handed or right-handed and whether they are left-thumbed or right-thumbed. **Note:** To determine whether one is left- or right-thumbed, have your students clasp their hands together immediately without thinking about it. Then look at the pictures to the left to determine the category. Students could collect class data and analyze the data to determine if there appears to be an association.

References

Franklin, C., G. Kader, D. Mewborn, J. Moreno, R. Peck, M. Perry, and R. Scheaffer. 2007. *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise*.

Hopfensperger, P., H. Kranendonk, and R. Scheaffer. 1999. *Probability through data.* New York, NY: Dale Seymour.

National Council of Teachers of Mathematics. 2000. *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Common Core State Standards for Mathematics, *www.corestandards.org*.