



Yale-New Haven
Teachers Institute®

Curriculum Units by Fellows of the Yale-New Haven Teachers Institute
2008 Volume VI: Depicting and Analyzing Data: Enriching Science and Math Curricula through Graphical Displays and Mapping

Graphics Creation and Statistical Interpretation: Relating Local Economics and the Global Environment

Curriculum Unit 08.06.07
by Jonathan Knickerbocker

Introduction

Statistics is a powerful and relatively new branch of mathematics that enables analysis and interpretation of data and summarization of data as graphical displays. Graphical displays are relatively new, invented in late 18th century, and can instantly convey a large amount of information (Tufté, 2001). In an era of increasing amounts of information, and increasing rate of collection of information, the ability to represent large data sets with a graphic and to analyze large data sets is important. Statistics is now said to “encompass ... the entire problem of making decisions in the face of uncertainty (Miller, 1999)”. In an era of ever-increasing energy and resource demands, the ability to live efficiently with minimal environmental impact is important for students in both economical and environmental senses.

Purpose

The purpose of this unit is to enable students to use statistics and data representation to make observations in trends of data, and to use these observations to make predictions, comparisons, and inferences about the world around them, both on a large scale and a small scale. The ultimate goal is for students to be aware of and have an appreciation for local and global aspects of environment and economy. This appreciation is particularly useful for students to be more economical consumers, and thus maintain a comfortable standard of living, while being as efficient as possible with their consumption.

With global warming and severe weather disasters regularly making the news, students may wonder how their lives affect the larger picture. They should be made aware of what they can do to curb any ill effects humans have witnessed as a result of the exploitation of natural resources. “The global population is precariously large ... Homo sapiens is reaching the limit of its food and water supply. The emulation cannot be sustained, not with the same levels of consumption and waste” (Wilson, 1998). The comparative exploration of data from

other geographic regions may help students empathize with the general human condition, particularly environmental and economic crisis throughout the world.

Goals and Objectives

In this unit, students will apply statistics, graphical interpretation and representation to the areas of economics, environment, and population. The topics of economics, environment, and population will each be researched and analyzed on a macroscopic scale (global, continental, or national) and a more local scale (back yard, neighborhood, city, county, state, New England). Students will achieve understanding of the statistical skills and concepts through the exploration of these topics. By the end of this unit, students should be able to demonstrate familiarity with the following concepts, and mastery of the following skills:

Statistical and Graphical Topics

Descriptive Statistics

Students will classify data as nominal, ordinal, interval, or ratio (including units of measurement), classify data sets (from tables or graphics) as univariate or bivariate. Students will summarize univariate data sets with measures of center, variation, and position, including median, mean, mode, range, interquartile range, standard deviation, variance, and quartiles. Also, students will describe the effect of an outlier on the measures of center, and interpret graphical displays of data using center, spread, clusters, gaps, outliers, and other unusual features and shapes.

Graphics

Students will construct graphical displays of univariate and bivariate data, including dotplots, histograms, cumulative frequency plots, stem and leaf plots, box plots, scatter plots, bar charts, pie charts, frequency tables and distributions, and ogive (continuous cumulative frequency graph or curve). Students will use scatter plots to estimate a least-squares regression line. Also, students will choose an appropriate form of data display given a data set. Students will compare clusters, gaps, outliers, unusual features, shapes of distributions, and compare distributions using various data representations. Finally, students will describe, analyze, and interpret frequency distributions, and describe properties of a normal distribution.

Analysis of Bivariate Data

Students will choose and acquire linear or nonlinear bivariate data to represent graphically in both time series and scatter plots. They will explore correlation and regression graphically and algebraically. Finally, students will interpolate from and extrapolate on (linearized) bivariate data sets using regression equations.

Key Concepts

This section describes in detail the main statistical, graphical, economical and environmental concepts that will be discussed in this unit. Terms that may be included in a vocabulary list are italicized. These concepts are referred to in the individual lesson plans. The economical and environmental concepts are related to the statistical and graphical applications that will be used to study the behavior of and relationships between the natural resources and human populations. References to resources for data sets involving the economical and environmental concepts are provided in the individual lessons.

Statistical Concepts

Data is measured or observed information. The two main *data types* are *quantitative* and *qualitative*. Quantitative data is numerical, while qualitative data is non-numerical. The four common *data types* include nominal, ordinal, interval and ratio measurements. *Nominal data* is purely qualitative, and is essentially categorical, i.e. does not have order or magnitude. *Ordinal data* can be qualitative or quantitative, has implicit order, but has no common difference. *Interval data* are purely quantitative, but the difference between two data has meaning. *Ratio data* is interval data in which a zero implies a quantity of zero for some measurement or observation (Larson & Farber, 2000). Comparison of the characteristics of data types can be eased by use of a matrix ¹.

Data summary measures of central tendency include *mean*, *median*, *mode*. The mean is the sum of all data divided by the size of the data set. The median is the middle entry in an ordered data set of odd size n . If n is even, the median is the mean of the two middle data points of an ordered data set. The mode is the datum with the highest frequency.

Data summary measures of dispersion, spread, or variability include *range*, *variance*, *standard deviation*, *interquartile range*. In an ordered data set, the lowest value is the *minimum* and the largest value is the *maximum*. The range is the difference of the maximum and minimum. The variance (of a sample of size n) is the sum of the squared difference of each data point from the mean, divided by one less than the size of the data set. The standard deviation is the square root of the variance ². The interquartile range is found by first separating the ordered data set into four *quartiles*, each with the same size. The divisions of the quartiles are known as Q_1 , the median (Q_2), and Q_3 . The interquartile range (IQR) is $Q_3 - Q_1$. Taken with the maximum and minimum, the median and quartiles are collectively known as the *five number summary*, which is useful for creating box plots.

A *frequency distribution* is a tabular representation of the *frequency* of a finite number of categories, or *classes*, where the classes are interval data. *Frequency* is the number of data entries in a class. The types of distributions include *normal or symmetric*, *skewed right*, *skewed left*, and *uniform or flat*. In a normal distribution, all measures of center coincide. In a skewed-left distribution, mean median mode, and the “peak” of the data occurs on the right side of the distribution. In a skewed-right distribution, mode median mean, and the “peak” of the data occurs on the left side of the distribution ³. When a frequency distribution is normal, probabilities or percentiles can be known using the standard deviation: 68% of the data lies within 1 standard deviation of the mean, 95% of the data is within two standard deviations from the mean, and virtually all (99.7%) of the data lies within three standard deviations from the mean ⁴.

A *variable*, for the purposes of this unit, is a phenomenon measured or recorded as quantitative data. In this way, a data set consisting of discrete nominal categories with frequencies is considered *univariate*. A data set with two corresponding sets of data (ordered pairs) is considered a *bivariate* data set.

A *population* is the set of all outcomes or measurements. A *sample* is a part of the population. A *statistic* is a measurement of some facet of a sample, while a *parameter* measures a characteristic of the population (Larson & Farber, 2000). One of the main goals of *inferential statistics* is to use statistics to estimate parameters.

Missing data points within the range of a bivariate data set can be *interpolated* or *extrapolated* using a linear regression formula ⁵.

Graphical Concepts

Graphical displays used to show frequency of univariate data that is classified nominally include dot plots, bar graphs, pie charts, and frequency tables. A stem and leaf plot and a histogram display frequency and distribution shape of data that is classified by interval data. Cumulative frequency plots or curves show the rate of increase in total frequency for all previous classes. Graphics lending themselves to display of distribution shape and variability include box plots and histograms. Graphical displays used to highlight the relationship between two interval or ratio variables include time series graphs and scatter plots. In both of these types of displays, data are considered as ordered pairs.

Economic and Environmental Concepts

Some of the economic and environmental concepts that students may explore include the basics for human existence on a large scale: climate, weather, freshwater, food, energy, health care, child care, housing, transportation, and taxes. Other data that may be explored includes human population of different areas, land area, production of waste and consumption of energy, carbon footprint, and water footprint. Preliminary data or introduction to the idea of natural resources may be acquired by viewing data maps and cartograms, which will reinforce geographical knowledge and practice spatial reasoning.

Economical concepts consist primarily of costs, which students will investigate and to which students will apply the statistics and graphics. Some basic cost categories that may be explored include household costs, such as the costs of water, food, electric and fuel (energy) costs, health care costs and childcare costs. The results of these costs or rates of consumption may be compared to the known quantities of resources that are available, and conclusions or predictions about the future availability of the resources will be explored ⁶.

Students

This unit is written for students at James Hillhouse High School, an urban comprehensive high school in the City of New Haven, CT. The audience includes students in grades 9-12. Students are placed in 9th grade courses based on middle school performance and Connecticut Mastery Test (CMT) scores. The student population of Hillhouse is about 1300. The unit could be incorporated into a more advanced course, such as Advanced Placement Statistics. When applicable, relevant extension opportunities for advanced students are

mentioned.

Rationale

This unit places an emphasis on interpreting and applying statistical methods to real data sets acquired by the students. Within certain limitations, the data explored should reflect student's interest in phenomena that may be viewed at both a local level and a national or global level "scale". Doing statistics with real data sets acquired by the students helps by making the data relevant to the students, especially when they are the ones who acquired the data. By analyzing household economic factors, students can become aware of the different facets of the cost of living, as well as the rates of increase of prices. In the context of studying statistics and data representation, household factors can be related to larger scale environmental factors, like the production of raw materials, rate of consumption of raw and manufactured materials and goods, and the end result of consumerism, waste management. Classroom discussion of both economics and environmentalism should help students to become more economical as consumers, and more aware as citizens of the 21st century, where environmental and economic policies and behaviors may effect generations to come.

Teaching Strategies

This section discusses some of the overall strategies that will be used in instruction of this unit. If strategies for a particular lesson vary from those given here, it will be noted in the particular lesson in which the deviation occurs.

Prior to this series of lessons, students will be separated into groups by interest. The basic interest groups could include food, water, housing, health care, and transportation. These choices may provide an opportunity for students to explore potential career paths. Students in a particular group will follow a theme involving their selected variables, and will frequently report back to the class with their findings. This reporting may be done via email, as in the case of sharing graphics, data sets, or data set resources. This may also be accomplished through classroom discussion, as in the case of observations made or questions raised while interpreting an environmental, economic, statistical, or graphical concept or data set.

Convenient data sources will be used to acquire relatively small data sets about local phenomena. Then, students will search for large scale or global trends in the same data, usually using the Internet, as there is a vast amount of information available in graphical or tabular form. These data may include the costs of housing, food, transportation, health care and education.

Students will be encouraged to identify similarities and differences in measures of center and variability in different data sets that measure the same variables, particularly in comparison of large populations and small populations. Students will also be encouraged to take notes to summarize the main ideas of statistical and graphical concepts, through used of the primary text, as well as online resources about these concepts.

Local and Global Comparisons

While most of the example data sets and graphical representations deal with national or world data, the teacher needs to ensure that students acquire data on a smaller, more local scale, perhaps the city of New

Haven, or the state of Connecticut. For the purposes of this unit, it is not necessary that the local population be a subset of the larger population. This could be made so in the case of a more advanced class, so that inferential statistical methods could be used to compare sample and population parameters.

Technology

As much as possible, technology will be incorporated into the processes of acquiring, organizing, displaying, and interpreting data. Technologies that may be used include graphing calculators, spreadsheet programs, graphing software, and slideshow presentation software. Data representations (graphics) will be created using both manual methods and technological methods. When possible, the introduction to an environmental or economic concept will be aided by teacher presentation of data maps or cartograms, reinforcing student geographic knowledge, and practicing map reading skills. Practicing map reading skills will foster spatial reasoning. Looking at maps will help students grasp the idea of being part of a much larger picture, and provoke discussion about the reasons for the differences in environmental and economic data for different geographic areas, and populations.

Since students will be acquiring much of their data through the use of the Internet, electronic communication should be the norm for individual discourse between the teacher and the student. Students will need email accounts, and parents will be notified of the requirement. Accommodations will have to be made for students without computer or Internet access outside of school, perhaps notification of public or school library hours. School computer labs can be made available during class time, as well as outside of school hours. Students will be required to keep an electronic portfolio of all sources used, data acquired, graphics created, and interpretations made. They will use materials in this portfolio to create the final cumulative assessment, which consists of a slideshow presentation.

In the creation of graphical displays of data sets, particularly with technology, students should be prompted to adhere to the following recommendations by Tufte: “Principles of Graphical Excellence”, including giving the viewer the greatest number of insights in the shortest time with the “least ink in the smallest space”, “Graphical Integrity”, particularly “the representation of numbers should be directly proportional to the numerical quantities represented”, and “clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity”. The “Principles in the theory of data graphics” can be used as a process model for creating efficient graphical displays 7.

Materials

The following is a list of materials that are needed throughout this unit: graph paper, rulers, graphing calculators, computers, spreadsheet software, Internet access, email accounts, slideshow presentation software, floppy disks or flashdrives. An almanac or encyclopedia may also be useful for finding data sets. Other materials needed for a specific lesson are mentioned in that lesson.

Whenever possible, electronic data sets should be downloaded in a spreadsheet format, so the data can be manipulated easily without having to input the data into a spreadsheet.

Classroom Activities and Lesson Plans

Lesson 1 -- Univariate Data Classification and Descriptive Statistics

Introduction

Prior to this introductory lesson, students should know the definition of data, and be familiar with some environmental and economical aspects that affect living, for example income, income source, geography, weather. This lesson poses the main question: How will we survive? This question, and the many answers that it may have, is relevant to the students as they become independent from their parents and have to fend for themselves in a world offering little compromise when it comes to household economics. This lesson provides the basis for future learning about data display interpretation, particularly central tendency and variability or dispersion (spread).

Objectives - students will be able to

1. Classify data as nominal, ordinal, interval, ratio (without graphics)
2. Give examples of each of these data types.
3. Calculate descriptive statistics for univariate data sets
4. Interpret frequency tables and distributions

Strategy

The teacher should provide some examples of data sets that relate to the general categories chosen by the students, and provide definitions of the main data types, and give examples of different units used to measure the same variables. Looking at populations of very different sizes may be useful to exemplify the use of different units. Prompt students to recall any formulas and terminology about descriptive statistics.

Student task

1. Use the internet or an almanac to find univariate data sets (without graphics) for weather data, water data, and food data, energy data, housing ⁸.
2. Identify the variables and classify them ⁹.
3. Calculate one-variable statistics by hand, with calculator, and with spreadsheet software: mean, median, mode, range, quartiles, IQR.
4. Discuss possible reasons and implications for the calculated descriptive statistics, e.g. which countries have a gasoline price above the mean, etc.

Figure 1 contains an example of a simple data table and some descriptive statistics. ¹⁰

Figure 1

June 2008 Gasoline Prices by Country					
Country	Price per liter (\$)	Central Tendency		Variability & Position	
FRANCE	2.30	mean	1.91	max	2.42
GERMANY	2.34	median	2.10	min	1.07
ITALY	2.42			range	1.36
SPAIN	1.91			standard deviation	0.51709
UK	2.30			variance	0.267382
JAPAN	1.63			Q1	1.48
CANADA	1.33			Q3	2.32
USA	1.07			IQR	0.84

Discussion and Assessment Questions

Discussion and assessment prompts and sample response are included in the next few paragraphs. Students may be expected to discuss or answer some of the following questions. The teacher may decide which questions are used for discussion and which are for assessment.

For nominal data, the teacher could ask “What type of data are the food groups?”, “What are the basic household cost categories?” Possible responses include the following: rent, mortgage, food, water, electricity, oil, gas, other.

For ordinal data, the teacher could ask “In what way can households, communities and nations be ranked (ordered)?” Some possible responses include family (population) size, area (land size)

For interval data, the teacher could ask “How do we measure the date”, “How is temperature measured?” The first question could stimulate a discussion about an interval or unit of a “one day”. Interval units of temperature measurement include degrees F, C.

For ratio data, the teacher could ask “How are elevation and depth measured?”, “How are savings and debt measured? “How is temperature measured?”, “What does $t = 0$ mean if t describes time?”, “What would a negative value for t mean?”

Many sets of ratio data include both positive and negative numbers. A ratio measure of temperature is the Kelvin scale; this could stimulate a conversation about absolute zero.

Lesson 2 -- Univariate Graphical Analysis and Data Classification from Graphics

Introduction

In this lesson, students search for graphical displays of their chosen environmental and economic factors. Then, students describe the types of data displayed in the graphics. Finally, students describe and interpret distribution shapes of frequency distributions. These skills will aid students in the creation of graphical displays in the next lesson.

Objectives - students will be able to

1. Classify data sets as univariate or bivariate from a graphic, identify data types
2. Estimate and discuss measures of center and variability from graphical displays
3. Interpret various graphical displays of data using center, spread, clusters, gaps, outliers, other unusual features, shape.
4. Describe, analyze and interpret frequency distributions and histograms

Strategy

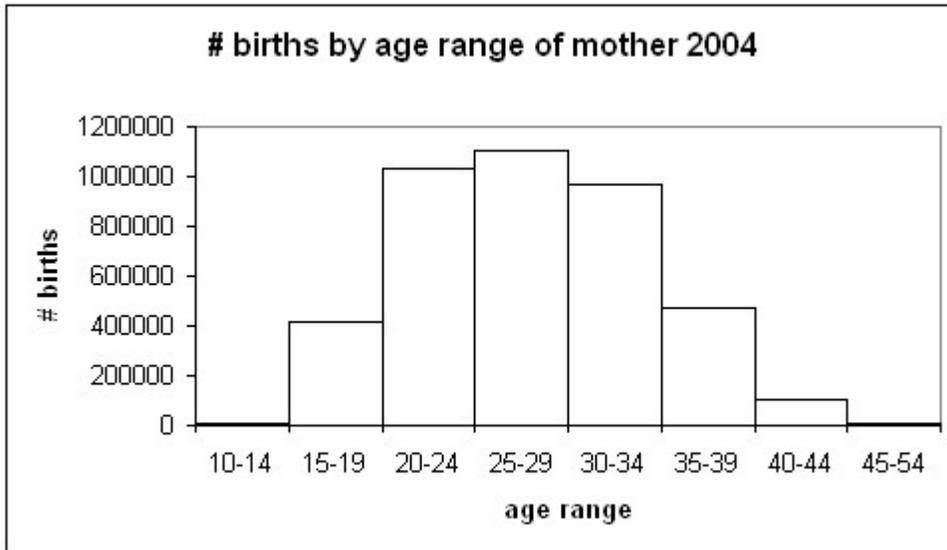
Teacher should promote students to identify examples of different shapes of distributions, including normal, skewed, and “flat” normal (high and low variability) skewed left and right. The teacher should also provide examples of different types of frequency distributions: Interpreting a stem-and-leaf plot is a good starting point for this lesson, as many students are familiar with this type of display.

Student task

1. Use the internet or an almanac to find univariate data sets represented in a graphical format.
2. Interpret frequency tables, frequency histograms (using data sets that represent each of the data types).
3. Describe the central tendency and spread of a data set for one particular graphic.
4. Compare two frequency distributions describing similar parameters for two different populations ¹¹

Figure 2 is a sample histogram, showing the numbers of births to selected age ranges of women in the US in 2004. ¹²

Figure 2



Discussion and Assessment Questions

Using this histogram, or a similar graphic, students will make observations such as the following:

It should be clear by looking at the histogram that most children are born to mothers of ages 25-29. Slightly more children are born to mothers ages 20-24 than of mothers ages 30-34. But since about 100,000 children are born to mothers ages 40-44, the histogram is slightly skewed right. The spread of the data has obvious biological limitations; note the very small number of children born to mothers below 15 or above 44.

Lesson 3 -- Univariate Graphical Analysis and Construction -- Frequency Distribution Types and the Normal Distribution

Introduction

In this lesson, students will continue analyzing graphics describing univariate data sets, and use their analysis skills to interpret and critique peers constructed data displays. The students experience a variety of data representations in this lesson. A more detailed discussion and presentation of distribution types occurs in this lesson. This lesson culminates with a discussion of frequency distribution types, in particular the normal distribution.

Objectives - students will be able to

1. Construct graphical displays of univariate categorical data, including dotplots, histograms, cumulative frequency plots, stem and leaf plots, box plots, bar charts, pie charts, frequency tables and distributions
2. Describe, analyze, and interpret frequency distributions
3. Compare clusters, gaps, outliers, unusual features, shapes of distributions
4. Compare distributions using various data representations (table, histogram, stem-and-leaf, box plots)

Strategy

The teacher should present examples of one-variable data that is represented in tables, pie charts, bar graphs, emphasizing the type of data represented in these graphical displays (nominal and quantitative).

Student task

1. Graph nominal and quantitative data with pie charts, bar graphs, ...
2. Graph frequency distributions (by hand, with graphing calculator, graphing software, graphing applets) and relate their shapes to measures of center and spread. Using technology, students should experiment with different class widths.
3. Possible extension: Assume normality of population and find standard scores using formula.

A good exercise for students is to calculate their household or family budget, and then graph the results in a bar graph ¹³. Note that the budget is for one month. Figure 3 contains sample output of the Family Budget Calculator while Figure 4 is a possible bar graph representation of the data. Figure 5 is a pie chart of the family budget data. This data might be compared to the U.S. averages household expenses, or another location in the U.S. Students may decide on the other location, perhaps using a location of their favorite sports team or the location of their favorite television show ¹⁴. Obviously, for comparison, the monthly expenses need to be converted into annual expenses or vice-versa.

Figure 3

Categories of monthly household expenses for New Haven Area	
Category	Annual Cost
housing	1142
food	288
child care	734
transportation	339
health care	211
other	344
total	3058

Figure 4

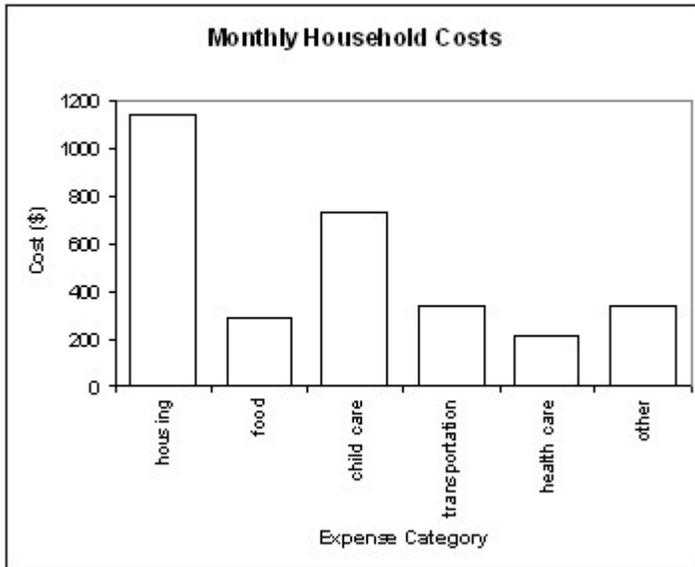
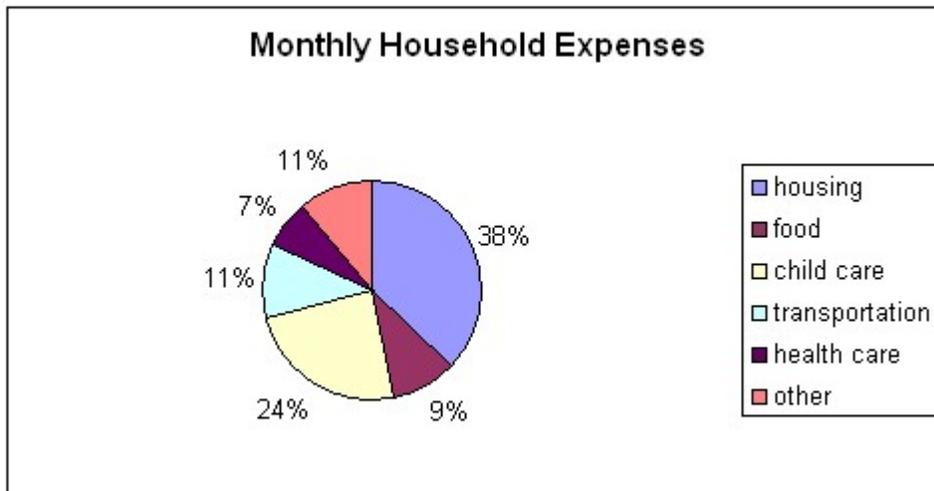


Figure 5



Discussion and Assessment Questions

- Which distribution is the most "balanced" or "symmetric"?
- Which distribution has more/less of the data above/below the mean?
- Which is the largest/smallest household expense category?

Lesson 4 -- Bivariate Graphical Construction and Analysis

Introduction

In this lesson, it is assumed students are familiar with the coordinate plane and know how to graph ordered pairs. The focus will be on creation of bivariate data displays, and the interpretation of these displays. It is also

assumed students are familiar with the concepts of slope and linear function. The main focus of this lesson is scatter plots, while lesson 5 focuses on time series plots. Informal (nonalgebraic) interpretation and analysis of a scatterplot, and its corresponding trendline, forms the bulk of this lesson.

Objectives - students will be able to

1. Represent bivariate quantitative data in a graphical format
2. Represent linear trends in bivariate data graphically and algebraically

Strategy

The teacher should provide examples of both time series plots and scatter plots, and have students compare differences between the two. Particular emphasis should be placed on the interpretation of slope, as the change in the dependent variable over the change in the independent variable. Close attention should be paid to the “scaling” used on the axis.

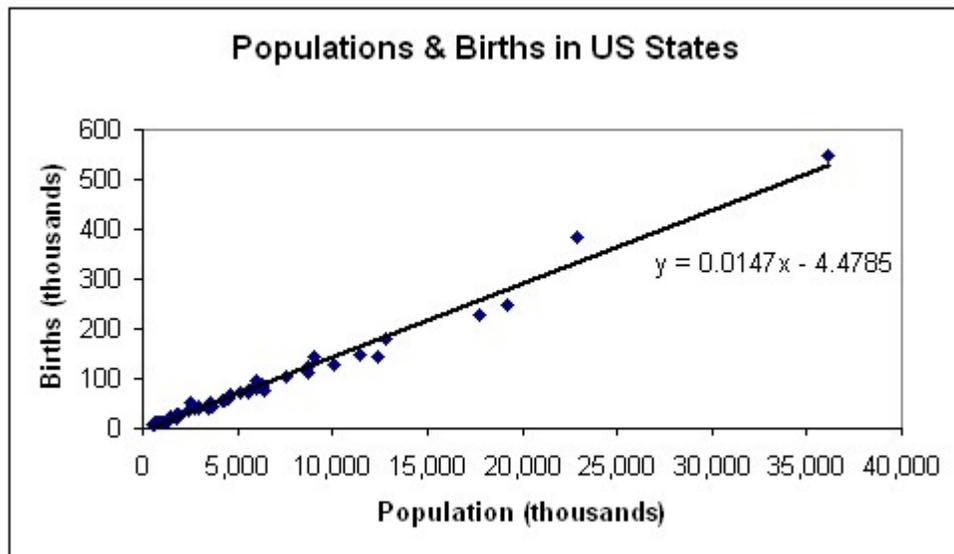
The teacher may begin with time series data, as units whose denominators are time are more easily understood by students, e.g. dollars per hour and miles per hour. Continue with histograms of varying bin sizes or class widths and end with scatter plot of two variables, the units of which students will not be familiar with (e.g. miles per gallon per dollar in the topic of fuel economy). Within linear regression data, start with direct variation, and work toward positive, then negative y-intercepts.

Student task

1. Acquire bivariate tabular data about two environmental or economic phenomena.
2. Establish hypothesis about the relationship between the two variables.
3. Create a scatter plot showing the relationship between the two variables, draw a trend line, interpret the meaning of slope (including units) and y-intercept.
4. Verbally describe relationship of variables (a related to b AND b related to a -- concept of inverse function for advanced students).
5. Use technology to create scatter plots, and calculate equations of trend lines (include med-med regression), interpolate, extrapolate using equation.
6. Verbally describe relationship of two variables, based on data collection, statistics calculation, and graphical construction.
7. Peer critique students graphical displays and interpretations.

Two variables that can be related using a scatter plot are number of births by state and population by state. In this way, the birth rate may be found for each state, or an average birth rate for the U.S. may be obtained. Figure 8 shows a scatter plot with the least-squares regression line ¹⁵. The method of drawing the regression line varies, depending on the software available, and to some extent, the computer literacy of the students. When comparing data for a specific year or month, especially when the two data sets are found from different sources, it is important for students to make sure that the data reflects the same time period.

Figure 6



Discussion and Assessment Questions

- What type of data is each graphic displaying?
- What are the measurement units?
- How many variables are depicted?
- What is the relationship between the variables?
- Is amount increasing or decreasing?
- What type of correlation exists between the two variables? Positive, negative, linear, strong, weak

Lesson 5 -- Graphical Construction and Interpretation -- Univariate and Bivariate

Introduction

In this lesson, the relationships between economic and environmental variables from one or more areas will be explored and related in depth. The students are allowed to choose which variables they are trying to relate,

and these choices should follow their initial topic of interest. This is a good opportunity to differentiate the data sets by student's interest. The decisions will have little guidance from instructor (formative assessment). The graphics created in this lesson serve as a formative assessment on graphics creation.

For the univariate constructions and analyses, students will focus on frequency distributions and comparisons thereof. The bivariate constructions and analyses will focus on time series. In lesson 6 the students will be calculating equations of regression lines. The interpretations expected of students are general, i.e. "is the graph increasing or decreasing?"

A time series graph may stimulate discussion about (petroleum) energy consumption ¹⁶. Other possible graphics to interpret address some of the following questions: "Where does electricity come from?" ¹⁷, "What are petroleum products used for?" ¹⁸, "Where does energy come from?" ¹⁹

Objectives - students will be able to

1. Identify normally distributed phenomena
2. Identify strong and weak relationships between environmental and economic variables
3. Describe relationships between variables

Strategy

The teacher should use technology almost exclusively and incorporate more than one graphic into the graphical depiction of a data set, i.e. look at the variables separately and together.

Student task

1. If students' hypotheses about the relationship between two variables were not evident in the scatter plot that they constructed, they should explore the variables independently, perhaps by means of histograms.
2. Students should then try to find two variables which are strongly correlated, and write observations about the relationships between the variables.
3. If the students' hypotheses about the relationship between two variables are affirmed in creating the scatter plot, they may proceed to writing notes about the regression equation formulas and the calculation and interpretation of the correlation coefficient.

Discussion and Assessment Questions

Which variable did you choose to graphically represent as a frequency distribution? Why?

Is this variable normally distributed?

Which two variables did you graph together in a scatter plot? Why?

Which variable is the independent variable? The dependent variable?

Describe the relationship between the two variables in the scatter plot.

Lesson 6 -- Regression: Interpolation and Extrapolation

Introduction

In this lesson, algebraic models will be derived from scatter plots of data that students have previously acquired and represented graphically, correlation will be discussed and student informal hypotheses about relationships between variables made. There is a focus on concept of slope as ratio of two variables, emphasizing units of measurement. This lesson continues discussion of statistical and graphical topics in lesson 4.

Objectives - students will be able to

1. Explore correlation and regression graphically and algebraically
2. Interpolate from and extrapolate on relationships found in bivariate data sets, using regression equation

Strategy

The teacher should encourage students to explore the relationship between energy (petroleum) costs and costs of everyday items, e.g. milk, bread, sneakers, bicycles, and provide students with procedure for calculating regression line by hand, using graphing calculator, and using spreadsheet software (right click & draw trend line, or use functions and additional table entries).

Student task

1. Students will use their bivariate data sets to find linear regression models for their data.
2. Then students will graph these models with the graph of the data set.
3. Students will use their regression models to find missing values and values outside of the data set.

Tabular data about college graduation rates and per capita income can be depicted in a scatter plot. ²⁰

Discussion and Assessment Questions

Describe the relationship between the two variables.

How strong is the relationship? How do you know?

Which variable is the independent variable? The dependent variable?

Cumulative Assessment

Students will prepare visual, written, and oral presentations of their research question, hypotheses, data gathering, and conclusions about a pair of variables related to the theme of economy and/or the environment. They should include discussion of predictions using a linear model and extrapolation or interpolation.

The slideshow presentation will contain the following items: title, question(s) about the relationship of at least two different variables for at least two different populations, sample data sets about the relevant variables, sources of data, univariate graphical representation of data (some type of display of a frequency distribution), bivariate graphical display of the relationship between two variables. Each graphical display requires a verbal interpretation of the data. Finally, students have to answer their question about the relationship between two variables, or differences between two populations.

Figure 7 contains rubrics for both graphics interpretation and graphics creation that could be made more specific for each type of data display.

Figure 7

Rubric for Graphics Interpretation
The number of variables is correctly identified
The type of each variable is correctly identified (including quantitative / qualitative, and nominal, ordinal, interval, ratio)
The units of each variable are correctly identified
Trends in data are identified, including discussion of center, variability, and distribution shape
For bivariate displays: relationship between variables is verbalized as positive or negative correlation (including relatively strong or weak correlation)
For bivariate display, the units of the slope of a regression line are identified and interpreted
Rubric for Graphics Construction
The type of graphical display is appropriate for the data set
The graphical display has a title that refers to all variables
The axes are labeled with the correct variables, units are included (Graphical Integrity)
Scaling on each axis is appropriate for the data set (Graphical Integrity)
The data is accurately represented graphically (Graphical Excellence)
The data is also represented in a table, including units
Where applicable, a regression line and equation is included
The units of the slope of the regression line are identified and interpreted
Interpolations and extrapolations are made using the regression equation and trendline
Note: The parenthetical phrases refer to (Tufte, 2001). See footnote 7.

Resources

Most of the resources below are appropriate for both teachers and students. The online resources are separated into the following categories: graphics, background text, Internet applications, and data tables.

Printed Sources

Larson, R. & Betsy Farber (2000). *Elementary Statistics - Picturing the World* . Upper Saddle River, NJ: Prentice-Hall.

Used for definitions of many statistical key concepts and formulas

Miller, Irwin & M. Miller (1999). *John E. Freund's Mathematical Statistics* . Upper Saddle River, NJ: Prentice Hall.

Page 1 "encompass decision making" used in introduction

Tufte, Edward R. (2001). *The Visual Display of Quantitative Information* . Cheshire, CT: Graphics Press, LLC.

Used in creating rubrics

Wilson, E.O. (1998). *Consilience - The Unity of Knowledge*. New York, NY: Vintage Books - Random House.

Quote in introduction from page 307

Electronic Resources

Graphics

Archer, C, M Jacobson, (2005 feb 3). Evaluation of global wind power. Retrieved July 20, 2008, from Stanford University Web site: http://www.stanford.edu/group/efmh/winds/global_winds.html

Map of wind speeds on earth (for wind generated power)

Bloch, M, S Carter, A Cox / New York Times, (2008 May 3). All of inflation's little parts. *The New York Times*, Retrieved July 20, 2008, from http://www.nytimes.com/interactive/2008/05/03/business/20080403_SPENDING_GRAPHIC.html

Graphic that shows detail of inflation rate (2007-2008). Well related to family budget.

Friendly, M (2001). Gallery of Data Visualization: The Best and Worst of Statistical Graphics. Retrieved June 27, 2008, from Department of Mathematics and Statistics | Faculty of Science and Engineering | York University Web site: <http://www.math.yorku.ca/SCS/Gallery/>

Some graphics are not resolved well, could refer to them for general trends (time series). See laurels and darts for good & bad graphics.

Gapminder, (2008). Gapminder. Retrieved June 27, 2008, from Gapminder Web site: <http://www.gapminder.org/>

Excellent site -- use for geography, animating population growth, animate elec consumption, can use to calculate growth rate (linear) of a variable in a particular country (by having it traced over time).

Global Virtual University, (2008). Globalis - an interactive world map. Retrieved July 1, 2008, from Global Virtual University Web site: <http://globalis.gvu.unu.edu/>

Map applet showing, among other things, human impact, population density, landcover, nightlights, climate forecasts.

Loster, M (2006-07-02). Total primary energy supply:. Retrieved June 27, 2008, from ez2c Web site: http://www.ez2c.de/ml/solar_land_area/

World map of solar power per square meter, has table for main deserts.

UNEP/GRID-Arendal, (2008). Maps and Graphics at UNEP/GRID-Arendal. Retrieved June 27, 2008, from Maps and Graphics at UNEP/GRID-Arendal Web site: <http://maps.grida.no/>

Good for data map interpretation, environment.

UNEP Grid - Arendal, (2008). UNEP Shelf Programme. Retrieved July 21, 2008, from UNEP Shelf Programme: Data Inventory Map Web site: <http://maps.continentalshelf.org/viewer.htm>

45 N, -75 W, -70 E, 40 S

Neat for viewing coastlines & getting familiar with Cartesian coordinates.

United Nations Environment , (2002). Vital Water Graphics, United Nations Environment Programme. Retrieved June 27, 2008, from Vital Water Graphics, United Nations Environment Programme Web site:
<http://www.unep.org/dewa/assessments/ecosystems/water/vitalwater/19.htm>

World's freshwater supplies per river basin.

woodheat.org, (30/05/2008). The Carbon Cycle. Retrieved June 27, 2008, from Woodheat Home Web site:
<http://www.woodheat.org/environment/carbon.htm>

Nice carbon cycle graphic.

Background Text

Easton, V.J. & John H. McColl (1997). Statistics Glossary - presenting data. Retrieved June 26, 2008, from University of Glasgow - Department of Statistics Web site: http://www.stats.gla.ac.uk/steps/glossary/presenting_data.html

Good source for some basic statistics definitions.

Northern Territory Government, Australia, (2005). Make The Switch. Retrieved June 26, 2008, from Northern Territory Government, Australia Web site: http://www.nt.gov.au/dpifm/Minerals_Energy/MakeTheSwitch/index.cfm?header=Glossary

Glossary of units, environmental terminology.

Internet Applications

The Economic Policy Institute, (2008). Economic Policy Institute: Datazone online Calculators. Retrieved July 20, 2008, from Economic Policy Institute Web site: http://www.epi.org/content.cfm/datazone_calculators

See Family budget calculator, output could be related to NYT inflation graphic.

FedStats, (2007, Mar 12). FedStats. Retrieved July 20, 2008, from FedStats Web site: <http://www.fedstats.gov/>

See left side: stats by state or topic or by topic: energy. Mapstats is nice to use online, has clickable us map by state with state stats, including area in square miles.

Hoare, R (2005 Jan 5). World Climate. Retrieved July 20, 2008, from World Climate Web site: <http://www.worldclimate.com/>

Compare local rainfall & temperature to "normal" for New Haven.

The Nature Conservancy, (Wed, 25 Jun 2008). Carbon Footprint Calculator. Retrieved June 27, 2008, from The Nature Conservancy Web site: <http://www.nature.org/initiatives/climatechange/calculator/>

Qudata.com, (2003). Qudata.com - Abstracts for Data mining resources. Retrieved June 27, 2008, from Qudata.com Web site:
http://qudata.com/lib/data_mining/

Has some good graphics, simulations: Stat calc inputs carriage return separated data, outputs distribution & all descriptive stats, Y(x) finder is curve fitter, inputting bivariate data, outputting the equation and error, 2d plotter simultaneously graphs 3 different

functions in different colors. Site also has a scientific calculator.

Statsoft, (2008). The Statistics Homepage. Retrieved June 27, 2008, from Data Mining, Statistical Analysis, Quality Control - Statistica Software Web site: <http://www.statsoft.com/textbook/stathome.html>

This site could be used as a primary source for statistical definitions and formulas.

Data Tables

Jensen, E. (2003, Feb 4). Anscombe's Quartet and Robust Fitting. Retrieved July 1, 2008, from Astronomy at Swarthmore College Web site: <http://astro.swarthmore.edu/astro121/anscombe.html>

National Center for Health Statistics, (June 26, 2008). NCHS - FASTATS. Retrieved June 27, 2008, from Center for Disease Control and Prevention Web site: <http://www.cdc.gov/nchs/fastats/default.htm>

Good health care resources, has indexed links by states and topics alphabetically.

OECD/IEA, (2008). International Energy Agency. Retrieved June 27, 2008, from International Energy Agency Web site: <http://www.iea.org/>

Many nice tables and graphics about energy.

US Census Bureau, (April 18, 2008). US Census Bureau: Teaching Materials. Retrieved July 20, 2008, from US Census Bureau Web site: <http://www.census.gov/dmd/www/schmat1.html>

See lessons using census data.

US Census Bureau, (2008 June 3). The US Census Bureau: The 2008 Statistical Abstract. Retrieved July 30, 2008, from The US Census Bureau: The US Statistical Abstract Web site: <http://www.census.gov/prod/2007pubs/08abstract/income.pdf>

US Dept. of Energy, (2003). A Consumer's Guide - Get Your Power From the Sun. Retrieved June 26, 2008, from National Renewable Energy Laboratory Homepage Web site: <http://www.nrel.gov/docs/fy04osti/35297.pdf>

See p11 for Calculating Electricity Bill Savings for Net Metered PV System.

Appendix: Implementing Standards

NCTM principles -- focus on equity, teaching, learning, technology

NCTM content standards -- focus on measurement, data analysis & probability

NCTM process standards -- focus on Communication, Connection, and Representation.

CT state dept of education curriculum framework - This unit focuses primarily on content area of Working with Data -- Probability and Statistics. Both core and extended items are included.

Footnotes

1. A matrix of characteristics of the data types, as well as examples of each, can be found in Larson & Farber, 2000, p11.
2. Larson & Farber, 2000, p71.
3. Larson & Farber, 2000, p61.
4. Larson & Farber, 2000, p197.
5. Larson & Farber, 2000, p433.
6. Source for some environmental terms and definitions: Northern Territory Government, Australia, 2005.
7. (Tuftte, 2001). Principles of Graphical Excellence are on p51, Graphical Integrity is found on p77, and Principles in the theory of data graphics is found on p105.
8. The following filenames (in quotes) and sources (in parentheses) contain the described energy data: (OECD/IEA, 2008), "Gas Web T1" - indigenous production or gross consumption by country, (OECD/IEA, 2008), "Oil Web T5" - Total OECD Imports of Gasoline by country, (OECD/IEA, 2008) "mps" page 3 and table 1 - June 2008 Average End User Gasoline Prices by country, (US Census Bureau, April 18, 2008), Teaching materials "stutable.pdf" - 2000 Resident Population of US by state, NCHS (CDC, 2007).
9. (Easton & McColl, 1997) has good examples of different variable types.
10. The data in this table is from (OECD/IEA, 2008).
11. A few Internet applications are a good place to have students begin searching for univariate data represented graphically. (Friendly, 2001) is a good site to find graphics for interpretation, as well as misleading graphics.
12. The data depicted in the histogram is from (National Center for Health Statistics, June 26, 2008 "CT birth data" p8).
13. The calculation of family budget can be made using (The Economic Policy Institute, 2008).
14. US Census Bureau, (2008 June 3), section 13 table 662 has expenditures somewhat itemized, so they may be related to the output from the family budget calculator. Table 663 has expense categories by state, which could be used to compare the costs of living in two different locations.
15. The data depicted is from both US Census Bureau, (2008 June 3) statpop table and National Center for Health Statistics, June 26, 2008 CT birth data on p13 table 4 columns 1 and 2.
16. (OECD/IEA, 2008) IEA energy tables, page 2, details prices in North America , Europe and Japan from 2005-2008.
17. (OECD/IEA, 2008) filename 19elec
18. (OECD/IEA, 2008) filename 19oil
19. (OECD/IEA, 2008) filenames 19prod, 19tpes, all from 1971 to 2005 for North America.
20. (US Census Bureau, April 18, 2008) teaching materials, filename 912ch2.pdf.

<https://teachersinstitute.yale.edu>

©2019 by the Yale-New Haven Teachers Institute, Yale University

For terms of use visit <https://teachersinstitute.yale.edu/terms>