Guidebook for Student Projects in Data Analysis

produced by

Southern Nevada Regional Professional Development Program

and

Nevada Chapter of the American Statistical Association

Principal Authors:
Maria Dufek, Jeffers ES, Las Vegas
David Thiel, SNRPDP

September 2006
(rev. 09/07)

Guidebook for Student Projects in Data Analysis

Produced by: Southern Nevada Regional Professional Development Program
and Nevada Chapter of the American Statistical Association

Principal authors: Maria Dufek, Jeffers ES, Las Vegas and David Thiel, SNRPDP

Table of Contents

Chapter I: Asking the Question

What is *Asking the Question*?

Probably the most difficult aspect of doing a project in data analysis is determining its subject. Who is to be studied? What about them interests us? Why? All of these are important in doing a study. One statistics textbook calls these the "W's." It is what the authors of this document refer to as *Asking the Question*.

Asking the Question means determining ***what*** *we are interested in studying about a population and* ***why***. It may be that we want to know the preferences of music by students in our school because we are going to hold a dance. We might like to garden and are curious about how plant growth responds to different amounts of sunlight. *Asking the Question* does ***not*** refer to the specific question that will be asked of respondents in a survey. That is, "What is your favorite type of music?" is ***not*** what is meant by *Asking the Question*.

Note: When we write "who is to be studied," the *who* are not necessarily people. The *who* could be bean sprouts or vehicles. *What* refers to the information of interest. In the case of bean sprouts, *what* could be how tall they are after several weeks. If we're studying vehicles, *what* could be their type, color, gas mileage, or all three.

Deterministic Answers vs. Data that Vary

Students must recognize that a data analysis question is based on data that vary from individual to individual. "How tall is Maria?" and "What is Dave's favorite style of music?" are not data analysis questions. The answers are a single height and a single style. However, "How tall are students in my class?" and "What are my classmates' favorite styles of music?" are data analysis questions. Height varies from student to student, as does music preference. The anticipation of *variation* is the key to asking a question in data analysis. After all, statistics *is* the study of variation. Heights vary from student to student, and their tastes in music are not all identical. This is *variation*.

So *what* do we want to know? About *whom*? *Why*? These are often difficult for students to generate. The process can be made easier by teaching students to "notice" and "wonder."

Noticing and Wondering

Students need to be given opportunities to observe their world using all of their senses (or as many as possible!) Use notebooks to have students record what they notice about their classroom. Have them work independently writing, illustrating and describing what they notice. This may take the form of an open-ended noticing or may be focused on specific criteria. These can be charted and shared.

From these noticings the wonderings will come. It may need to begin with *Teacher Talk*—"Dave, you noticed that the desks are arranged in 6 rows of 4 desks in each. You also noticed that we are really crowded in here. I'm wondering if there is another way to arrange the desks to give us more room?"

"Sara, you noticed that we have 3 boys and 19 girls in our class. I'm wondering if the other 3rd grade classes have more girls than boys."

"Manuel noticed that the teacher has pictures of dogs and horses on her desk. I'm wondering if those are her pets and what pets the others in the class might have."

Encourage the students to record their wonderings. Some may not be data analysis questions. You will need to guide the work from here. A focused noticing and wondering—"What do you notice that can be measured or counted?" can bring up the awareness level that some things have variability and some do not. If a student wonders how many chairs and desks are in each classroom, that is a curiosity question that may lead to some counting experiences. If the student wonders if all classrooms have the same area of floor space that may lead to some measuring and comparing. The question is, "Why do you want to know about chairs?" Is there a reason to know that? Does the number of chairs tell you anything else about the class? What is it about floor space that interests you?

Having students do many forms of noticing and wondering generates interest in, and curiosity about, a variety of topics.

Formulating Grade-level Appropriate Questions

When students take on a project, it's important that they consider a question that is appropriate for their grade level. For instance, let's take the case of favorite music. Simply surveying students in a class and asking each what is his or her favorite style of music is a fairly simple, low-level study. This would be appropriate for students in the primary grades.

Intermediate students may wish to not only know what students' favorites are, but also if students like more than one style of music. They may also compare the music tastes of students of different ages, or classes, or even schools. This turns up the complexity a bit by making comparisons and looking for relationships.

Older students—middle and high school—should be exploring several (or many) aspects of music preference. Do people who prefer rock music also like rap? Do they tend to dislike country? Do they like jazz, but not classical? Is there a connection between music preference and the rhythm of the beat? The possibilities are nearly endless, and are limited only by the students' imaginations, noticings, and wonderings.

Let's look at another example where one might do a study that does not involve surveys or people—plant growth. Primary students might be interested in how light affects plant growth. They could plant two flowers, beans, or some other flora in separate pots, place one near a window, and place the other on the side of the room opposite the window. After some period of time, they could compare growth.

Students in intermediate grades might do the same, but with multiple plants. They would average the growth, consider its variability, and even compare it at multiple points in time.

As students get into middle and high school, they could consider multiple settings with various degrees of "lightness," using multiple plants in each case. They might also consider other factors that could come into play, like soil moisture, fertilization, etc.

Questions Based on the Type of Data

Data comes in two flavors: ***categorical*** (qualitative) and ***numerical*** (quantitative). ***Categorical*** data describes a quality about the individuals or things being studied. Favorite style of music is categorical data: rock, jazz, country, etc. The type or color of car a teacher drives is categorical information.

Numerical data describes a quantity about the individuals or things being studied. How many hours each day students listen to their favorite music is numerical. The height of a bean sprout plant after three weeks is numerical, as is the number of cars a teacher owns. Generally, if one had to ***count*** or ***measure*** to gather the data, it is **numerical**.

The level of sophistication of a project is often a function of whether the data collected is categorical or numerical. It can be difficult, but not impossible, to formulate erudite questions if the data to be collected is only categorical. By extending a categorical-inspired project to include numerical data adds a level of complexity to the project and can make it more interesting. Consider the following situations that begin with a question that generates categorical data and are broadened to include numerical data.

- What activities do students like to do in their free time? How much time do they spend each week on them? How much time is spent doing the activity in a single session?
- What are students' favorite restaurants? How much do they spend each time they eat at those restaurants?
- What kinds of pets do students own? How many pets do students own? How large are the pets?
- What varieties of fruit do students like to eat? Which do they dislike? What are the levels of sugar, acid, etc. in various fruits?
- Do students wash their hands before/after ________? How often do students wash their hands each day? How long do students spend washing their hands?

Even categorical questions can be extended by using other categorical questions, to get more detailed information, or look for relationships among the different categories.

- What styles of potatoes do people like? If they like French fries, which fast food chain is the best? If they like baked, what toppings do they like?
- What varieties of fruit do students like to eat? Which do they dislike? Does it depend on their texture? On their peel/rind?

- What kinds of pets do students own? Do the students live in apartments or houses?

Helping students learn to formulate a “question” takes some time, but having them notice and wonder makes things go easier. As students try to determine a subject for their project, have them consider the following question stems. These are just a start. They generally become more sophisticated as one moves down the list.

- What are…favorite …?
- How often/how many/how much…?
- Does…differ between…?
- What are the similarities and differences between…?
- What is the relationship between…?
- Is there a trend…?
- How would one predict…?

Chapter II: Collecting the Data

What is *Collecting the Data*?

Once students have decided upon a question, the next step is *Collecting the Data* to answer the question. What method will be used to collect the data? How much data will be collected? Are there any concerns that will require certain techniques? Each of these must be considered in the data collection process. Usually, the question will dictate what method of data collection is required.

Data Collection Methods

Data collection methods can be divided into three basic categories: *surveys*, *observational studies*, and *experiments*.

Surveys. A *survey* is a method where information is gathered from individuals about habits, preferences, opinions, attitudes, etc. The survey may be a brief series of orally-posed questions, like a *poll*, or may last longer and take the form of an *interview*. Paper or electronic *questionnaires* may be used to allow more time for respondents to answer, access more potential respondents, or simplify the data collection process. (To be precise, a *survey* is actually a specific form of the next category, the *observational study*.)

Observational Studies. An *observational study* is the method by which a researcher tries to establish relationships between variables by making observations of the target population. Data may come from direct observations of the individuals in the population, or may come from examinations of records in books, almanacs, databases, or the Internet. In an observational study, the researcher does nothing to influence the individuals being studied. When doing an observational study, the data already exists; it simply needs to be gathered.

Experiments. An *experiment* is a study where the researcher deliberately influences certain events and observes the reaction to them. Its goal is to establish a cause-effect relationship between two or more variables. When doing an experiment, the data does not exist and needs to be created.

Experiments also include *simulations*, a study where actual conditions are imitated using chance outcomes based on predetermined models. An example of a simulation would be to explore variation in the percentage of male newborns in a hospital by using a coin toss to determine gender.

To understand the difference between *observational studies* and *experiments*, consider the following scenario. Students may notice that the bean sprouts in each of several classrooms are of different heights. They also notice that the amount of sunlight reaching the plants is different in each room. The students could gather data about plant height and sunlight, then attempt to find a relationship between the two. This is an *observational study*. In an *experiment*, the students would deliberately expose plants to various degrees of sunlight, and then record how plant growth was influence by it. In the first case, the data was there, the students just had to record it. In the second case, they actually created the data by influencing certain conditions.

The Quantity of Data to be Collected

When collecting data for a survey or observational study, there are two terms that apply to how much data is collected: *census* and *sample*. A *census* is a study where data is collected from all individuals in a population. A *sample* studies only part of the population.

If a sample is done, because it is impossible or impractical to do a census, the issue becomes how large of a sample is needed. Larger samples are preferable to small ones, as the researcher is more likely to get a clear picture of the population's characteristics. However, there is no magic number as to how many individuals should be in a sample. It depends on the complexity of the study and how accessible the individuals in the population are.

Considerations in Sampling

If a census cannot be done, then a sample must be taken. A sample must be as representative of the population as possible. The group of individuals that has the potential to be in a sample must reflect the population. If we were interested in the music preferences of students in our school, it would not make sense to survey only those in our classroom, only our grade, or only the boys.

Those chosen to be in the sample must be selected as impartially as possible. People have natural biases that may guide them to include or exclude certain individuals from a sample, even unconsciously. To overcome this, and to increase the chance of a representative sample, individuals should be selected randomly. The use of chance devices—slips of paper in a hat, dice, spinners, and random number tables—allows for selection of individuals in an unbiased manner.

Considerations in Surveys

Often when surveys are done, unexpected things occur that hamper exploring the data or drawing conclusions from it. A datum may be missing, unreadable, or so unusual that it makes one wonder if it were recorded correctly. Results may be much different from what was anticipated. An individual chosen to be in the sample or census may have been missed or surveyed multiple times. Variables that were not considered before a study was done may now be of interest to us when it is over. These issues can usually be addressed by maintaining detailed records about who provided what information, in the event a follow-up survey is needed.

Another option is to do pilot studies. Pilot studies are usually small surveys, observational studies, or experiments done to establish what types of results would be obtained in the main study, to become aware of other variables of interest, and to work out problems in the study's design.

When conducting surveys, how questions are asked is a very important consideration. Should the question be open-ended or closed-ended? Is the question understandable? Is the question being posed in a manner that does not tend to elicit a particular response? If a question is not comprehensible, if it tends to sway respondents to a certain answer or answers, or if it is so open-ended as to result in a different response from every respondent, then drawing conclusions from the data may be difficult or impossible.

Considerations in Experiments

Just as random selection is important in samples, randomization is important in experiments. In a study of how sunlight affects plant growth, the plants available for the experiment must be randomly assigned to the various levels of sunlight. The chance assignment filters out problems that may be hidden from us, the plants' health for example.

The number of subjects in the experiment is a key consideration. If one has a single plant in each of the different levels of sunlight, or even two or three, it becomes more difficult to determine if differences in growth are actually due to the differences in light or some other variable. The more subjects involved in the experiment, the more we can attribute differences between groups to the variables under study.

The final consideration is the control of other variables. Even though our plants may be exposed to different amounts of sunlight, other variables should be controlled as much as possible. That is, experimental conditions should be as identical as possible for all plants, except for the sunlight. All plants should receive the same amount of water and food at the same time, artificial lighting should be identical, etc. Anything that could potentially affect growth should be controlled.

Designing for Differences

Asking the Question is about *anticipating* variation. *Collecting the Data* is about *acknowledging* that variation. As students progress in their understanding of data analysis, their studies should be designed to expose those differences. The principles described above assist in doing that.

An Example of a Survey

Question: Do students at our middle school who prefer one style of music tend to like/dislike certain other styles of music?

The population of interest is all 1800± students at the school. It would be nearly impossible to do a census, so a survey using random sampling will be done. A sample size of 300 is desired, with 100 from each of the 6th, 7th, and 8th grades. The teachers of the Math Department have agreed to help in accessing the respondents.

There are 20 math classes in each grade, so five students will be sampled in each class to obtain the desired 100 students from each. Each teacher has provided a count of the number of students in each class. Five students are chosen at random using the following as an example:

> If the class count is 30, the numbers 1 to 30 are written on slips of paper and placed in a hat. Five of the slips are drawn, at random, and the numbers on those slips recorded. Those numbers identify the potential respondents by the order they appear in the teacher's roll book. If the number drawn is 5, then one of the potential respondents will be the fifth student listed on the teacher's class list.

The teacher will administer the survey questionnaire to the five students during the last minute of class on November 1. If the potential respondent is absent, the teacher will attempt it only once again on November 4. The questionnaires will be collected at the end of the day, November 4.

The questionnaire appears as follows:

ID: 823605

(1) What is your *favorite* style of music? (Check only ONE)

___ country ___ classical ___ hip-hop ___ rap ___ rock ___ metal
___ jazz ___ Gregorian chants ___ easy listening ___ blues
___________________ other (write-in)

(2) For each of the following styles of music, write a "D" in the space if you *dislike* that style of music or write an "L" in the space if you *like* that style of music. Please complete all spaces with "D" or "L."

___ country ___ classical ___ hip-hop ___ rap ___ rock ___ metal
___ jazz ___ Gregorian chants ___ easy listening ___ blues

The first three digits of the ID is the teacher's room number, the fourth is the class period, and the last two identify the randomly selected student by position on the class list. If a follow-up is needed, the student can be identified.

An example of the raw data from the survey, before analysis, is given below:

ID	favorite	country	classical	hip-hop	rap	rock	metal	jazz	G. chant	easy list.	blues
823605	rap	D	D	L	L	L	D	L	D	D	L
823612	jazz	D	L	D	D	L	D	L	D	L	L
823613	rock	L	L	L	D	L	L	L	D	D	D
823620	rap	D	L	L	L	L	D	D	L	L	L
823647	reggae	D	D	D	D	D	D	D	D	D	L

An Example of an Experiment

Question: Does the amount of sunlight a bean plant receives affect its growth rate?

The process begins by planting several dozen bean seeds at equal depth in identical pots and placing them under a grow light in the multi-purpose room. During this time, all containers will be covered, and receive an equal amount of water each morning. The pots will be placed to receive equal amounts of light. After several days, we will “harvest” the 16 plants that have heights closest to 5 cm in height.

The experiment will take place in the school’s quad area. It is well-shielded from wind, but is exposed to open air and sunlight. Plants will be placed next to the walls of the quad, halfway down each wall. The area next to the south wall receives no direct sunlight during the day. The north wall receives several hours of direct sunlight. The east and west walls both receive some direct sunlight, but not as much as the north wall.

The 16 chosen plants will be numbered from one to 16 and their heights to tenths of centimeters recorded. Slips of paper numbered from 1–16 will be placed in a hat and mixed. Four slips will be drawn and the plants with the same numbers as those slips will be placed on the north wall. Four more will be drawn for the west wall and the east wall. The remaining four plants will be placed on the south wall.

The plants will receive equal amounts of water (about 1 oz.) each morning before school. The teacher will water them on weekends. Every Monday, Wednesday, and Friday, the plants’ heights will be measured to the nearest tenth of a centimeter. The experiment will run for three weeks, at which time the plants’ heights will be compared.

An example of the raw data from the experiment, before analysis, is given below:

Plant	Wall	Apr1	Apr3	Apr5	Apr8	Apr10	Apr12	Apr15	Apr17	Apr19	Apr22
1	E	4.7	6.5	8.3	11.0	12.8	14.4	17.4	20.0	22.8	26.1
2	S	5.2	6.2	7.6	9.1	9.5	10.5	11.7	13.3	14.3	15.8
3	S	5.0	5.8	6.8	8.0	9.2	10.2	11.7	12.7	14.1	16.5
4	W	5.6	7.4	8.4	10.5	12.3	14.1	16.2	18.2	19.6	22.0
5	N	4.9	7.1	9.7	12.7	15.5	18.5	21.8	24.4	26.8	30.7
6	W	4.9	6.1	7.9	10.9	12.3	14.1	16.5	18.1	18.9	21.3
7	N	5.1	7.7	9.7	13.6	16.0	18.4	22.0	24.6	26.8	30.1
8	W	4.8	6.4	8.0	9.8	11.8	14.2	16.9	18.3	19.7	22.1
9	E	5.1	7.3	9.7	13.0	15.4	17.4	20.1	22.3	24.3	27.6
10	N	5.0	7.4	9.6	13.2	15.4	17.6	20.6	22.8	25.4	29.0
11	S	5.1	6.3	7.5	9.9	11.5	12.7	14.8	15.6	17.0	18.2
12	S	4.7	6.1	7.1	8.3	9.1	10.3	11.8	13.2	14.6	16.7
13	W	5.1	6.1	7.9	10.3	12.1	13.7	16.4	17.8	19.6	22.9
14	N	5.1	6.9	9.3	12.9	15.1	17.7	21.0	23.6	26.0	29.6
15	E	5.1	6.9	8.5	11.8	13.4	15.0	17.4	19.4	21.0	24.3
16	E	4.7	6.9	8.9	12.5	14.7	16.9	19.9	21.5	23.5	26.5

Chapter III: Organizing and Analyzing the Data

What is *Organizing the Data*?

Once data has been collected, it must be organized into a form from which patterns can be recognized, conclusions made, and inferences drawn. This is the heart of the poster competition—creating a graphical display that leads the viewer to these goals. The structure of such displays depends in large part on the question asked and the type of data collected, and in some measure the creativity of the author. Just as the study should be designed to expose differences (variation) within and between groups, graphical displays should reveal those differences.

Forms of Data Display—Categorical Data

The simplest form of categorical data display is a *frequency table*. A frequency table simply counts how many times each category has occurred and displays it in tabular form. For example, a student may ask respondents to identify their favorite type of music. The responses range from classical to rap. A frequency table may look like the one shown.

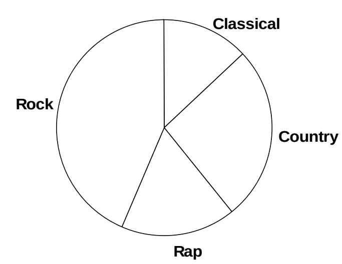
Type of Music	Frequency
Classical	3
Country	6
Rap	4
Rock	9

If an association between two categorical variables is being investigated, then a *two-way frequency table* can be made. Each cell in the table shows the frequencies of combinations of categories. For instance, perhaps students are wondering if people who like plain vanilla ice cream prefer other plain foods, like cheese pizza. The table shows how preferred pizza and ice cream styles are related.

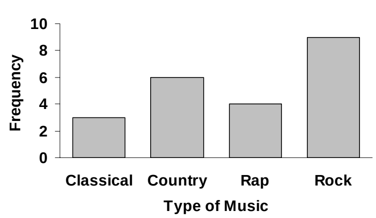
		Ice Cream		
		Vanilla	Chocolate	Total
Pizza	Cheese	8	2	10
	Pepperoni	3	6	9
	Total	11	8	19

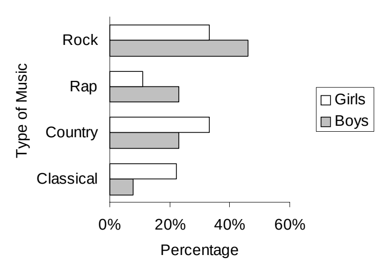
Often, a table is insufficient to get a clear picture of the distribution of data, or see patterns or associations. What is called for is a "picture" of the data. The most common pictures of categorical data are *circle graphs* (or *pie graphs*) and *bar graphs*.

Circle graphs can show how a whole group is broken into smaller parts. The circle graph shown represents the music data from above.



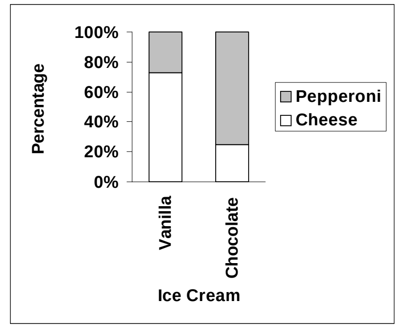
Bar graphs show similar information to circle graphs. While comparing part to whole is harder, sometimes seeing differences between parts is easier. Multiple-bar graphs can be used to compare groups by breaking data into sub-groups. The music data is shown as a whole group and split by gender.





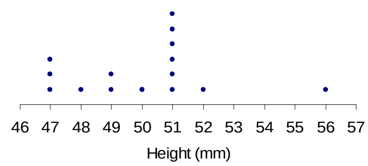
Note: when sub-groups are different sizes—there were 14 boys and 9 girls in the favorite music survey—it is often better to graph the data as percentages than frequencies.

One can compare how categories are distributed between two groups using circle graphs. As an alternative, a *stacked bar graph* is equivalent to comparing multiple circle graphs. Such a graph is shown using the ice cream/pizza data from the previous page.



Forms of Data Display—Numerical Data

Displaying numerical data usually takes the form of a picture based on a number line. The simplest form of numerical display is the *line plot* (or *dotplot*). Each observation is represented by a dot (or some other symbol) and like observations are stacked on a number line. For example, a dotplot of the heights of plants measured to the nearest millimeter is shown.

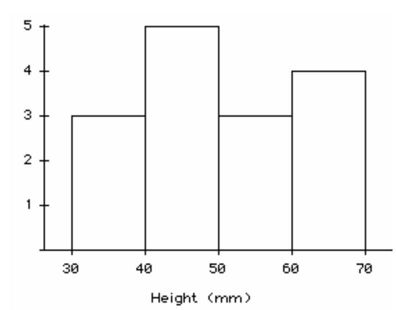


When data is much more spread out, or takes on many different values, it may be done as a *stem-and-leaf plot* or as a *histogram*. A stem-and-leaf plot shows the actual values of the data using consecutive place values. In the stem-and-leaf plot shown, the heights of fifteen plants range from 33 mm to 69 mm. It is important that the leaves on each stem (of the plot, not the actual plant) have equal spacing.

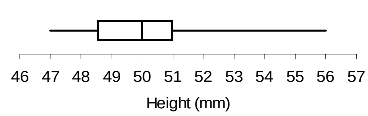
```
3|349
4|01137
5|244
6|0139

Key: 3|4 represents
34 mm
```

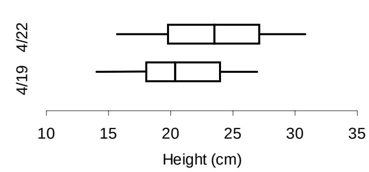
A histogram shows the distribution of data in intervals, but the individual values cannot be determined from the display, like in a stem-and-leaf plot. It is preferable to a stem-and-leaf plot for very large sets of data. A histogram is shown that displays the same data as the stem-and-leaf plot.



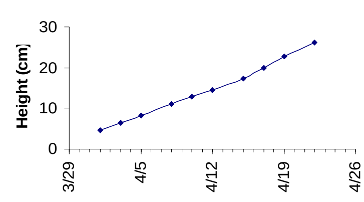
A *box-and-whisker plo*t (or *boxplot*) is a form of graph that divides a set of ordered data into four equally-sized groups, then visually shows the ranges of those groups. The spread of the lowest and highest fourths are shown with whiskers; the 2nd and 3rd fourths are shown with boxes. It provides a quick visual representation of the center, spread, and to some extent, the shape of the data. The boxplot shown is of the same data as the dotplot above.



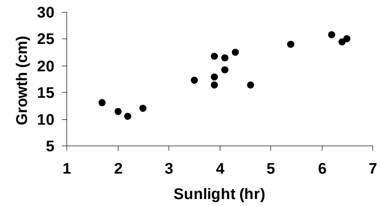
The power of the boxplot is in comparing the center and spread of multiple groups. The *parallel boxplot* shown displays a comparison between the heights of plants on two consecutive measuring days.



When numerical data is gathered over a period of time, it can best be displayed on a *time plot*. In the time plot shown, the height of a particular plant has been recorded regularly over several weeks. The graph shows the trend of its growth.



When an association between two numerical variables is being explored, the type of graph most commonly used is a *scatterplot*. Each observation has two values, which are plotted as an ordered pair on a coordinate plane. For example, one may be wondering if there is a relationship between the amount of sunlight a plant gets each day and its growth over a three-week period.



Summary Measures

Summary measures are single numerical values that are used to describe the center or spread of numerical data. The most commonly used measures of center are *mean* and *median*. Frequently used measures of spread are the *range*, *interquartile range*, and *standard deviation*. When examining bivariate numerical data, *correlation* and the *slope* and *intercept* of a fitted line may be used. Which measures are used is dependent on the nature of the data, its shape, and whether or not extreme values (*outliers*) are present.

Type of Music	Frequency
Classical	3
Country	6
Rap	4
Rock	9

Summary measures like the mean and median are NOT used with categorical data. It is common to see students take data as shown in the table, add the frequencies, and conclude that the mean is 5.5. The raw data is categorical: classical, country, rap, and rock. One cannot do arithmetic with categories and the value 5.5 makes no sense in this context. The same is true for the median, range, and all other measures of numerical data.

Statistical Inference

High school students in honors or Advanced Placement statistics courses may formulate a question that can be answered using inferential procedures (*t*-tests, chi-square tests, analysis of variance, etc.). This should be encouraged and the results of inferential procedures should be included in the presentation of the project. The timing of project deadlines may require some out-of-class instruction on inferential procedures, as these tend to be taught in the spring.

Final Notes about Graphical Displays and Summary Measures

Any graphs and measures students choose to provide in their analysis should be germane to the question asked. Graphs should not be made just for the sake of making a graph. Additionally, multiple representations of the same data need not be provided if the additional graphics do not give any more information. That is, a stem-and-leaf plot and a histogram of the same data are redundant, as potentially are a circle graph and a bar graph.

The displays and summaries outlined here are only some of the possible ways to organize and analyze data. Additionally, only brief summaries of the different methods have been provided. See the appendix for resources that provide more detail in graphical display and data analysis.

Chapter IV: Completing the Project

How Does One *Complete the Project*?

The purpose of a project in data analysis is to answer a question by the collection and analysis of data. It is in a sense, the telling of a visual, statistical story with a beginning (asking the question), a middle (collecting the data), and a climax at the end (displaying and analyzing the data). The story must be told in such a way that the reader or viewer can follow all parts of the story and be able to see or draw the correct conclusion at the end.

The mechanism for telling the story may take on various forms. Depending on the complexity of the project and any external factors (e.g. competition requirements, teacher directions), the project may be a paper, a poster, or an oral presentation.

Whatever the means of delivery, the question must be understandable, text must be readable, and graphics must be clear, concise, and link back to the question. Any conclusions presented must be sound and based on the data, or if not directly communicated, at least easily drawn from the information presented. The use of color and other accouterments should enhance the presentation without distracting from the overall message.

Is There Advice Specifically for Statistical Posters?

John W. Tukey (Statistical Science, 5(3): 327-339) said,

> "The purpose of a statistical poster, then, is to visually tell a story from the data about some phenomena revealing the conclusions that can be drawn. Because there is no narrator to tell the story, nor an accompanying report to discuss the data, the poster must be able to stand alone; it should not have to be explained.
>
> For this reason, special care must be taken to present ideas clearly. Not only must the viewers understand the individual graphics, but they must also understand the relationships among the graphics and how the graphics address the question(s) being studied."

Here are some guidelines from the Poster and Project Competition webpage at www.amstat.org:

> BASIC GUIDELINES
>
> While constructing a poster, it is important to keep in mind that the central idea of the study should be the prominent feature of the poster. To bring the main idea into focus, questions such as the following should be asked.
>
> ◊ What is the purpose for displaying this information?
>
> ◊ What comparisons should be made?
>
> ◊ Which trends should be shown?
>
> Questions should be asked until the central idea of the study becomes clear. This becomes the focal point of the poster. The poster must reveal what the data have to say. It must allow the viewer to see the data; that is, to see the variation in the data, the structure of the data, the important patterns in the data (or lack of a pattern), the data points that do not fit the pattern, and the conclusions to be drawn from the data. Further, each graphic on the poster should convey new information about the data-a pattern or structure, for example, that cannot be seen in the other graphics.
>
> The poster title should be informative in order to reduce the need for additional explanatory text. For example, the title may indicate the questions addressed by the graphics or even convey the major conclusion to be drawn from the data.

Each graphic's legend should be positioned so there is no question which graphic and legend go together. Further, each graphic and its legend should stand alone. If the graphics need to be viewed in a certain sequence, however, then the viewer's eyes must be guided in that sequence.

Try to eliminate trivial and extraneous information, gridlines, or lettering. In particular, redundancy in titles and legends should be omitted. Only explanations needed to make the conclusions clear should be included. Data tables should not be shown on the poster, as reading off numbers is not the point of the display.

Choose a few harmonious colors that are easily visible. The key to using color effectively is restraint. Colors should not distract the viewer, but should enhance recognition of the structure of the data and the conclusions.

Edward Tufte (*The Visual Display of Quantitative Information*) said graphics may "…reveal the data at several levels of detail, from a broad overview to the fine structure." In a similar sense, a poster may do the same. At a distance, only a broad overview of the poster and the data is possible. Therefore, main titles should be visible, and overall outlines of the data-as revealed by the graphics-should be seen. On closer inspection, however, aspects such as individual data labels and legends should be more apparent.

Chapter V: Assessing the Project

How does one *Assess the Project*?

The assessment of a project will depend on its guidelines and requirements set forth by the teacher or sponsoring organization. For example, the following is a project assessment outline by teachers Amy Strube (Neal ES, Las Vegas) and Chris O'Neill (Lied MS, Las Vegas):

GRAPH PROJECT

Name: ____________________

Project Due Date: ____________________

- Each student will create and present an original graph project to the class. The poster and presentation must depict an understanding of the data and the graphs used.
- The graph project will be assembled and displayed on a poster board measuring between 18 and 24 inches high and 24 and 30 inches wide.
- Computer graphics may be used to illustrate, decorate, or label the poster. The student must make all graphs and tables. Neatness and organization of the poster is important and will factor into the final grade.
- **Tape** this form to the back of the poster.
- Be prepared to present your poster, discuss why you chose this topic and explain how you collected your data.

__________ 1. **Formulate a survey question** (10 pts.)
Think of something you can measure, or get an opinion about. Phrase it as a question. You might have to offer a limited choice of responses. Form a secondary question that further explores or examines the topic. This is also displayed on the poster.
Display a title for your investigation on the poster.

__________ 2. **Gather data for your survey questions** (20 pts)
You will collect data about your questions ether by observing and recording it over a period of time, or by interviewing or observing at least 50 people. **Display the results of the survey question in appropriate tables on the poster.**
Be sure to properly title and label the tables.

__________ 3. **Present a statistical analysis of the data.** (10 pts.)
Calculate the mean, median, mode, and range of the data. Remember, this cannot be done with all types of data. **Display a summary of the statistical analysis on the poster.**
If your data does not have this, then write a statement explaining why.

__________ 4. **Display the data using 2 different graphs.** (50 pts.)
Create 2 different graphs that appropriately display the results of the survey questions. Be sure to title and label the graphs. Also be sure to show the scale and intervals. **Display the 2 graphs on the poster.** Neatness and accuracy are very important parts of these graphs. Also, this is to be done in color.

__________ 5. Write a summary, conclusion and prediction based on the data (10 pts.)
Summarize the results of the survey data, draw a conclusion about the data, and make a prediction using the data. **Include the summary, conclusion, and prediction on the poster.**

__________ Total: 100 pts.

The Nevada K-12 Statistics Poster Competition uses the following guidelines in assessing posters submitted to it:

- ♦ Overall Impact of the Display
    - ▪ Poster beckons the viewer.
        - It is eye-catching and has visual attractiveness.
        - Spelling and grammar are correct.
        - Construction of the poster is neat and balanced.
        - Poster is free of clutter.
        - Color is pleasing and well-used (not over used or under used).
        - Type is readable.
    - ▪ Points that will reduce ranking on a poster include:
        - Errors in spelling and grammar.
        - Errors in construction of the poster.
        - Clutter.
        - Inappropriate or overuse of color.
        - Lack of readability of type.

- ♠ Thoughtfulness of the Purpose and Method of Collection
    - ▪ Poster conveys a purposeful question properly investigated.
        - The question shows depth of thought for the grade level.
        - The methods and volume of data collection are appropriate and sufficient to answer the question.
        - The raw data is included.
    - ▪ Points that will reduce ranking on a poster include:
        - Questions that lack depth of thought or are not grade level appropriate.
        - Errors in the methods or volume of data collection.

- ♥ Clarity of the Message
    - ▪ Poster tells a story.
        - The message demonstrates important relationships and patterns, evident conclusions and the ability to stand alone, even without the explanatory paragraph on the back.
        - The question is clear.
        - The answer to the question is obvious from the components.
        - All components of the poster lead to answering the question.
        - The display may lead to further questions.
    - ▪ Points that will reduce ranking on a poster include:
        - The question is not clear and requires examination of the back of the poster, or is not clear at all.
        - Determining the answer to the question requires some effort such as: examination of the back of the poster or it is indeterminable.
        - One or more components are irrelevant to answering the question.

- ♣ Appropriateness of Graphs
    - ▪ Graphs (and measures) are correctly created.
        - All graphs are correctly constructed.
        - Any measures are correctly computed.
        - All graphs and measures are appropriate for the data.
    - ▪ Points that will reduce ranking on a poster include:
        - One or more graphs have errors in construction.
        - Measures are incorrectly computed.
        - Graphs and/or measures are inappropriate for the data.

- • Creativity

The following are some examples of well-designed and poorly-designed posters and graphs from the Nevada K–12 Statistics Poster Competition sponsored by the Nevada Chapter of the American Statistical Association.

Well-designed

Figure 1

Figure 2

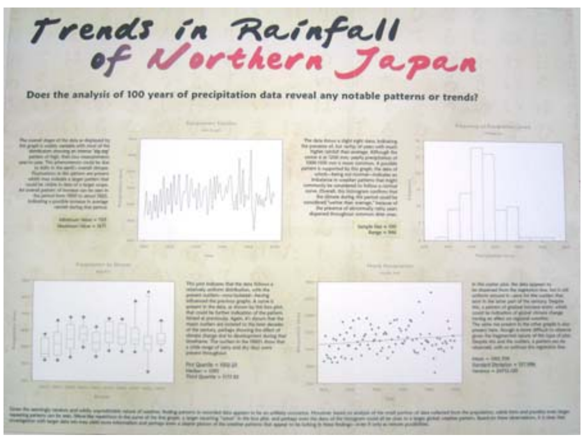


Figure 3

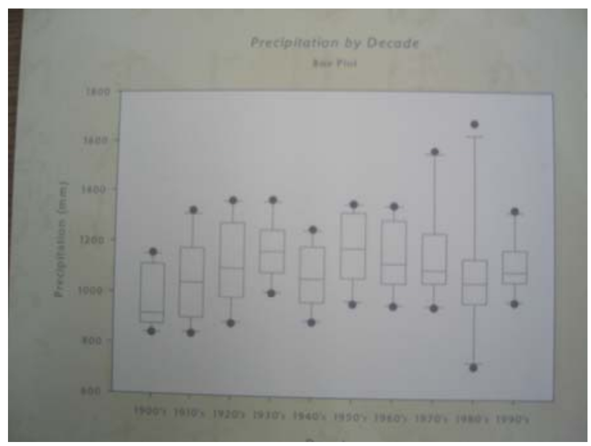


Figure 4

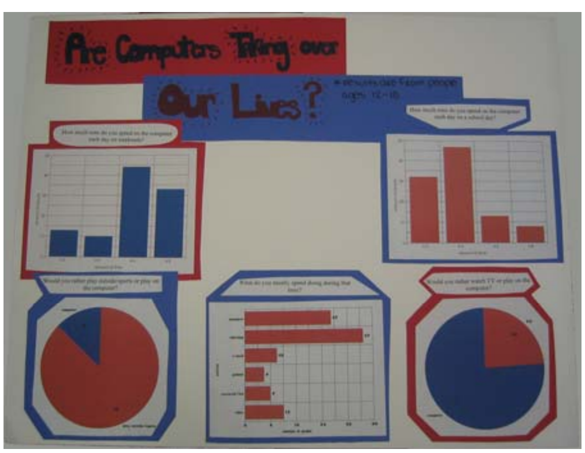


Figure 5

Figure 6

Figure 1: A first-grade poster that is well-organized and the display clearly communicates how animals move.

Figure 2: Very clean and nicely laid out third-grade poster. Graphs are easy to read and interpret. Brief predictions and summary, while not necessary, communicate purpose and add more information.

Figures 3 & 4: This 12th-grade poster has a balanced layout and is easy to follow. Summary paragraphs were a bit verbose, but did not detract from or contradict the information in the graphs. Computer-produced graphs were easy to read and interpret, and clearly showed the trends in rainfall. Each graph showed a different aspect of the rainfall data.

Figures 5 & 6: Well-produced poster with graphs displaying multiple aspects of the information collected. Majority of the poster is dedicated to data display and graphs are easy to read. Notice that no summary paragraphs are present, or needed, as the graphs effectively stood alone in communicating how computers impact students' lives.

Poorly-designed

Figure 7

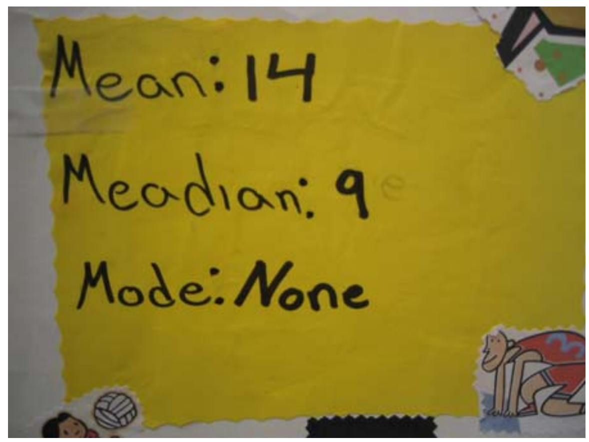


Figure 8

Figure 9

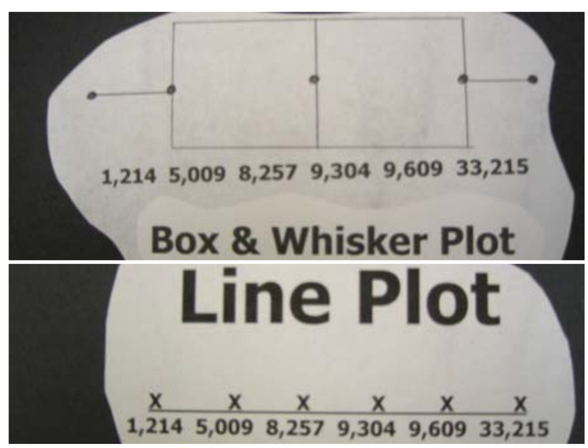


Figure 10

Figures 7 & 8: Not only does this sixth-grade poster fail to include two graphical displays relating different aspects of the data, but the included summary statistics are inappropriate. Favorite sport is a categorical variable, from which quantitative measures like mean and median cannot be computed. The mode here is actually the category "soccer." This poster has a lot of clutter and not much information.

Figure 9: More than three-fourths of this poster is dedicated to accouterments of neon paper, glitter, and glue. The graphs are practically unreadable. Like the poster in Figure 7, it is clear more time went into decorating this poster than went into collecting and analyzing the data.

Figure 10: The topic of this 10th-grade poster was a comparison of areas of six New England States. The line plot poorly displays the data as it is not drawn to scale. The box-and-whisker plot also suffers from scaling problems, is inappropriate for such a small data set, and generally conveys the same information as the line plot. For this reason, these were not considered to meet the competition's multiple graphs requirement. Finally, and most importantly, simply listing the areas of six states and making a graph of the areas is a feeble topic for a high school, or even middle school, poster.

Resources and References

- Nevada Chapter of ASA Poster Competition Web page, where one may view previous winning posters, download registration forms, and find this document. www.nevada.edu/~nvasa/
- American Statistical Association Poster Competition Web page, where one may view previous winning posters. www.amstat.org/education/index.cfm?fuseaction=poster1
- Bock, David E., Velleman, Paul F., and DeVeaux, Richard D. (2004). *Stats: Modeling the World.* Boston: Pearson/Addison-Wesley.
- Curcio. F.R. (2001), *Developing Data-Graph Comprehension in Grades K–8.* The National Council of Teachers of Mathematics. www.nctm.org
- Landwehr, J.M. and Watkins, A.E. (1987). *Exploring Data,* Palo Alto, CA: Dale Seymour Publications. www.pearsonlearning.com
- Tufte, E.R. (1983). *The Visual Display of Quantitative Information,* Cheshire, CT: Graphics Press. www.amazon.com
- *A Curriculum Framework for PreK-12 Statistics Education* (2005), Participants of the Guidelines for Assessment and Instruction in Statistics Education (GAISE) www.amstat.org/education/gaise
- *Math on Call: A Mathematics Handbook* (1998), Wilmington, MA. Great Source Education Group. www.greatsource.com
- *Navigating through Data Analysis* [*and Probability*] (2003), Reston, VA, National Council of Teachers of Mathematics.
- *Take It to the MAT* Newsletters. Click on the Math Newsletters link. www.rpdp.net
- Any good high school- or college-level statistics textbook.