Organizing & Presenting Data: A How To Primer (updated Summer 2011)

CAROLE COCHRAN, MSW & DEBRA S. NORRIS, EDD, MSW

We would like to thank for reviewing the manual:
Danielle Syndey Dearborn,
SD KIDS COUNT staff assistant, summer 2011

Contents

Introduction

Why a manual about organizing & presenting data?

Have you ever:

- Calculated how long it would take to drive from point A to point B?
- Figured your tax return?
- Calculated your GPA?
- Prepared your household budget for the month?
- Balanced your checkbook?

If so, you are well prepared to use data effectively.

Information is the key to effective change. Changing policies and conditions requires facts. Successful change must have at its foundation reliable data that make a case or prove a point. Unfortunately, the very term "data" intimidates many because they assume that using data require knowledge of complicated mathematical and statistical procedures. To use data effectively, you only need to know how to select the right facts and numbers and perform the mathematical basics learned in elementary school— counting, adding, subtracting, multiplying, and dividing. The Primer includes finding credible facts (data) from Journals and Web resources to help find the facts needed to prove a point.

The Primer presents six essential steps for working with data. They are:

1. Know how to use Journal articles and Web resources.
2. Know the purpose for data: Why build a data story?
3. Know what data you need.
4. Know where to get the data & how to organize the data.
5. Know how to calculate the numbers.
6. Know how to present the data.

The Primer is designed to help students, advocates, social/human service workers, and others learn how to select, analyze, and present data. The essential steps presented throughout the manual can apply to any field.

Objectives of the manual

1. Identify credible web-based literature and journal articles
    a. Search for best practices
    b. Understand how to read journal articles
2. Identify the purpose of a data project
    a. Know what you want the data to illustrate
    b. Know what your end product will be
3. Match data needed to the purpose
    a. Identify what data are needed
    b. Map your data search
4. Master simple organizational tasks
    a. Know how to set up a spreadsheet for a primary database
    b. Know how to set up a spreadsheet for importing secondary data
5. Identify data sources
    a. Choose the appropriate source
    b. Choose the appropriate data within the data source
6. Know how to perform appropriate calculations
    a. Match calculations to the purpose
    b. Match calculations to the data
7. Know how to present the data
    a. Match data display to the purpose
    b. Maximize data presentation options

Chapter One - Get Curious

Summary of Chapter

The chapter will introduce you to ways to get curious about issues. The chapter will introduce you to professional journals and credible Website tools available for issue searches.

Objectives of Chapter

I. To provide Websites to explore "evidence-based" programs.

II. To provide information on professional organizations and networks.

III. To provide information on how journal articles are organized.

IV. To provide information about statistics often reported in journal articles.

The Search

The Internet is a collection of interconnected information resources. There is no central indexing. However, there are ways of getting information. The internet is like a library in that way.

In a library you can find information by looking for a book on the catalog system or browsing the shelves. Once you have a book, you need to scan it to see if it contains helpful information. You can also look for information in magazines by flipping through magazines on the shelf or using a periodical guide. Again, you must read the article to see if it contains the information you need. The same is true with the Internet. There are several places and ways to look for information. However, only by actually looking at the information can you know if it is helpful.

The Internet contains a tremendous amount of information. It is easy to get sidetracked during a search. That is why it is best to focus and limit your search before you begin. So, before you begin: 1) determine the time you will spend online; 2) formulate a question(s) to answer; and 3) develop a search plan; think of related sites you have visited, and think of key words.

The quickest way to perform an Internet search is by using a search engine. Search engines allow you to perform key word searches. Search engines will work best with specific and narrow topics. With a search engine, you can perform a simple search by typing in the word or words that describe your topic. Be as specific as possible. The search engine will return sites that contain the words you have submitted. Most engines will "weigh" returns by putting the sites they feel will be most relevant to the searcher first. Different search engines have different qualifications for weighing items.

Be Curious

The best way to become curious is to read what other people have written. The internet is full of easy access information but, be careful, it may not be credible. A credible source of information is usually information that has been looked at or reviewed by other people who have studied the same or related issue area.

Information on research linked to helping practices can be found in the evidence-based practices or best practices tool kits and manuals produced by government agencies. Evidence-based practices reports provide the most recent research findings and are sources of credible information. These resources have become a critical component to any program planning and grant effort. The websites offer search windows so a topic area can be accessed easily.

To insure credibility in you search start with government agencies such as :

SAMHSA's National Mental Health Information Center (Adult MH Toolkits)

http://www.mentalhealth.org/cmhs

National Institute of Mental Health **http://www.nimh.gov**

The National Institute for Health **http://www.nih.gov**

Another helpful way to search the internet is through Professional Organizations:

The National Extension Association of Family & Consumer Sciences

http://www.neafcs.org

American Association of Family & Consumer Sciences **http://www.aafcs.org**

Many Networks and specific research and practice groups provide a clearinghouse of information

The National Child Welfare Bureau **http:www.ncwb.org**

Promising Practices Network on Children, Families and Communities

http://www.promisingpractices.org

Rural Assistance Center **www.raconline.org**

Through the professional organization and network sites many credible journals can also be searched.

To help insure the research articles provide reliable information look for the article to follow a standard template. A reviewed "refereed" research journal article will use the following general format.

Review of Literature

•Show support for research question

Methodology or Method

•How question was investigated

•Sometimes an article may investigate an impact a particular practice or intervention has had on a group, in that case the article may have an Intervention section

Analysis & Findings

•How data or question information was "looked at" or specific data analysis performed. Don't let that scare you. We will look at some general terms that will help us to understand some of these procedures

Results or Data Findings

Discussion or Conclusion

•Often some recommendations may be found in this section.

•All the parts needed to build the case for the question investigated and explain what the investigators found.

•Make sure to understand the language and complexity of what is being studied.

•Find articles that are written clearly and use headers for the different sections.

Some Simple Statistics

Some simple statistical information will help in understanding the purpose of an article's research methods section. Most research attempts to explain an interaction or relationship between two variables. A variable is something (concept) that can be measured. A discrete variable is an either/or concept, e.g. adult or child. Other factors are considered continuous because there are many variations of responses within the range of factors being explored.

Independent variables are the factors the researcher suspects may influence or cause an effect on the situation being studied. The dependent variable is the factor being studied. For example, if a researcher is studying Body Mass Index (BMI) and the effect of a nutrition curriculum on BMI, the BMI would be the dependent variable and the nutrition curriculum would be the independent variable. What type of data is collected determines what a researcher can do to investigate the relationship or effect of study variables.

Nominal Data: often called demographics: sex, racial or ethnic group, Age goes in here as well even though it can also serve as interval data. It is often reported within a range (20-25) to protect the identity of a study participant.

- Name sort of data: people, place, things data, and either/or responses to a question such as yes or no
- Mode: the most common response
- Statistical Analysis such as Chi Square and Cross Tabulation
- Correlation: does one factor or variable increase when another variable increases? Does liking potato chips go up as age increases?
- Goodness of Fit

Ordinal Data explores relationships:

- Rank: 1st, 2nd, 3rd
- Likert scale common tools: Strongly agree, Agree, Neutral, Disagree & Strongly disagree

Spearman r test (Correlation)

- Relationship

Mann-Whitney U-test

- 2 samples independent

Interval Data Looks for differences

- Numeric data all the time where data streams (variables) have equal intervals between them such as age or a standard test score
- Mean, the average (X)
- ANOVA's look for effects

t-tests: (Differences)

- One sample comparing data to a population

ANOVA (Effect)

- Comparing more than 2 samples or groups of data

Probability level

- A p of .05 means that in the universe randomly everything has a 5 in 100 chance of occurring
- For a researcher to have any confidence in their data really showing something the p value needs to be less than .01 $P < .01$

Symbols used

- n= number of responses (participants in the study)
- M= mean the average of responses
- SD = standard deviation, The square root of the differences of how responses or are distributed from the mean
- p= probability, how likely is this effect to happen randomly

Chapter Two - Know the purpose for data: Why build a data story?

Summary of Chapter

The chapter will provide a brief overview of what data are and how data enrich our abilities to understand and describe our world. The difference between primary and secondary data are discussed. Through an exercise and example, learners enhance their appreciation and daily use of data.

Objectives of Chapter

I. To understand what data are

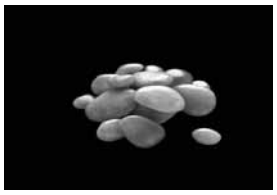
II. To understand the difference between primary and secondary data

III. To appreciate what data do for us

What are data?

Webster's Dictionary (Merriam-Webster, 2005) defines data as *"factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation."* Data must be measurable and lay the foundation for discussion, calculation, and/or reasoning. You determine the purpose for your data to guide what you hope to discover or support. What you are investigating will define your data sources. The purpose for your inquiry will also help determine how you will assemble, not just gather, data. The assembly of the correct data leads to sound decision-making. Consider the difference between gathering data and assembling data.

Gathering or assembling data



Gather

The definition (Merriam-Webster, 2005) of gather is *"to cause to come together; to accumulate gradually, amass, harvest, pick."* What you end up with is a pile of data much like the pile of stones in the picture. It is gathered but what it is intended to show us is unclear. A well defined purpose for data is critical.

Assemble

The definition (Merriam-Webster, 2005) of assemble is "*to bring together into a group or a whole; to fit together or join together the parts*." By assembling data you are working towards a purpose. You have an end product in mind and you are building toward that goal. You are telling the story, illustrating the purpose of your data or research with your data.

To put it another way, when you order something, such as a bicycle, and it comes in a box, what does it say? Some **assembly** required. They have already gathered the necessary parts. You have to assemble it to look like the picture on the front. Similarly, you don't gather a puzzle together, you assemble it. Without a clear purpose (order and form) for data you are just gathering it. You bring order and form to the chaos by assembling data in a meaningful way.

Data as evidence

Data as evidence are the building blocks for "proof". Funding authorities want proof, via evidence, to substantiate claims of need and effectiveness of services or programs. This in turn leads to a more integrated and coherent policy response to issues. Evidence is a statement of fact substantiated by some sort of data.

The data most common to service provision are:

1) Data to substantiate a claim of need

2) Data to describe an agency's or a community's provision of services or lack of services

3) Data to describe a region, people, or events

Most professionals and students are familiar with primary data gathering procedures. Fewer professionals and students are comfortable with searching for secondary data sources. The manual has been prepared to raise the comfort level and use of secondary data sources.

Primary versus Secondary data

Primary data are gathered when an agency or individual designs the questions used to solicit specific information. Often in the professional service arena an individual or agency will design research questions to create a survey or questionnaire. Surveys and questionnaires are often used to help solicit information (data) that is assembled into a report, review, or grant proposal. These data are specific and limited to the practice and programs of an individual and/or agency. Professional practitioners and agencies use primary data to plan, evaluate and expand their efforts.

Primary data gathering tools are:

- direct observation - lets you focus on details of importance to you.
- surveys - written surveys let you collect considerable quantities of detailed data. They can be telephone surveys, record reviews, computer generated surveys or a questionnaire.

- interviews - allow in-depth questioning and follow-up questions.
- focus groups - allow in-depth questioning and discussion on a topic.

What you learn in setting up a spreadsheet in Chapter Three can apply to managing primary data, however this manual will focus on secondary data use. A secondary data source means that the information is simply second-hand, existing databases are secondary data sources.

For example, the information from the U.S. Census Bureau is secondary unless you are looking at copies of the original returns. The list of purely secondary source material could include:

- databases
- TV, radio, internet
- magazines, newspapers
- reviews
- research articles

There are many secondary data sources. Secondary data are less expensive and (sometimes) easier to acquire than primary data. However, problems may arise around questions of the reliability, accuracy and integrity of the data. Who collected it? Can they be trusted? How old is it? Where was it collected? Can the data be verified? A general rule of thumb is that databases maintained by government agencies and other agencies of reputable standing provide the researcher with an accepted level of credibility.

Often secondary data has been pre-processed to give totals or aggregates and the original details are lost so you can't verify it by replicating the methods used by the original data collectors. Attention to totals, aggregates, and percentages is key to your use of the data.

Exercise

The first task is to get comfortable with data. Work through the exercises below to focus on data and its utility in professional communication.

Try to accurately describe a person or object using general observations. Really try to accurately describe the person or object with just vague general descriptors. How successful or unsuccessful were your attempts? Now explain the same person or object with specific data/details. Notice the difference in having data versus not having data; including data add focus and increases credibility.

Consider the two paragraphs below and see how adding data helps make your case more persuasive.

Statement without data: *In recent years, South Dakota has seen both a need and a demand for low-cost, high quality child care. South Dakota has a high number of females over age 16, with children, in the workforce. The need for child care touches the lives of every South Dakotan. The child care industry is an essential part of the state's economic development strategy.*

Statement with data: *Without child care, most South Dakota businesses would be hard-pressed to find enough employees. That's because in South Dakota 73 percent of children under age 6 have one or both parents who are in the workforce. South Dakota leads the nation in the percentage of women in the workforce who have children younger than 6. In South Dakota, the percentage is 77.5%, compared to 63.5% for the United States as a whole. At 47 percent, South Dakota also leads the nation with the highest percentage of children under age 6 in paid child care. That's almost twice the national average of 26 percent.*

Licensed or registered child care itself is a significant industry in South Dakota that:

- *Generates $100.6 million in gross annual receipts.*
- *Creates 4,410 jobs in South Dakota.*
- *Yields $124.5 million in direct economic activity.*

(Cochran, C. and Stuefen, R., 2004)

Chapter Three - Know what data you need

Summary of Chapter

The chapter will discuss how the purpose for data guides the data assembly process. The chapter also leads the learner through a step-by-step process for importing data from secondary sources into their own working databases. Methods for primary data sources are also discussed.

Objective of Chapter

I. To identify specifics in a data search that informs the assembly of the data.

The purpose for your data

The purpose for your data: what you want to show through data is critical to setting parameters around what data you want. What data you want sets limits on where you go for the data. Everyone can successfully assemble data that accurately answers questions, adds credibility to a claim or powerful descriptors to a situation. The key to successful data assembly begins with clarity of the data search. Data gathering requires identifying and defining the facts needed to satisfy the purpose for your data search.

How do you begin to assemble data?

What do you hope to prove or discover, discuss, or make decisions about? Look at the big picture first. The "big" picture reveals the specifics and will guide your search. Here is where you begin to use the description of a situation to identify where numbers would add credibility to your words.

Decide what data you need to assemble

Remember there are three distinct types of data that overlap from time to time when we are gathering data.

1) Data to substantiate a claim or need

2) Data to describe an agency's or a community's provision or lack of services

3) Data to describe a region, people, or events

Data about people, or "demographic" data, are such pieces of information as age, race, gender, income, employment status, and grade in school. Some of these characteristics do not change—such as gender or race. Others, such as age, education, and income do change. Data about events are such occurrences as births, deaths, graduations, traffic crashes, and immunizations. Data about things

can be places, organizations, families, programs, and objects such as houses. All three types of data quantify something about a time and a place. They refer to a specific time, perhaps a day, a calendar year, or a state or federal fiscal year.

"Cross-sectional" data refer to one point in time; "longitudinal" data (sometimes called time-series data) cover at least two points in time. Data also refer to a specific geographic area.

We will now look at the mechanics necessary to assemble data to investigate a simple question.

Chapter Four - Know where to get data & how to organize the data

Summary of Chapter

The chapter will instruct learners on understanding databases and spreadsheets.

The chapter will assist learners in setting up a database and a spreadsheet to be able to manipulate the data.

The chapter will instruct learners on obtaining secondary data from databases found on the Internet.

Objectives of Chapter

I. To understand the basics of a database and a spreadsheet.

II. To begin to build a database with information from various sources .

III. To know how to set up a spreadsheet of data.

IV. To understand the very basics of how the Internet works.

V. To begin to build a data source sheet with information on where to get the specific data you need.

VI. To know how to obtain secondary data

What is a database?

A database is a collection of related data stored together in an electronic file that can be easily retrieved. The collected data could be in any number of formats (electronic, printed, graphic, audio, statistical, combinations). There are physical databases (paper/print) and electronic databases.

Examples of databases:

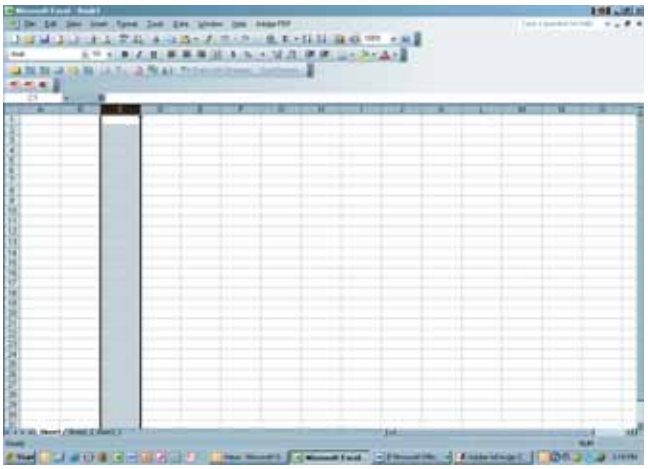
- phone book
- voter registration rolls
- immunization records of children in a family
- Census Bureau data
- library catalogue system

Understanding how databases are organized may help retrieve information more efficiently. The easiest way to explain how a database is organized is by explaining a spreadsheet.

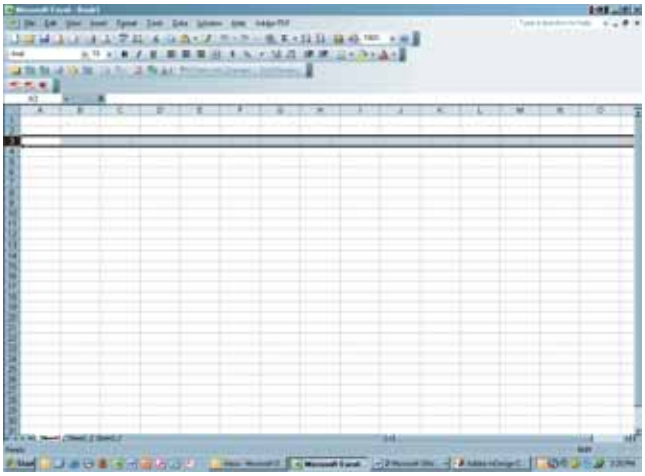
What is a spreadsheet?[1]

A spreadsheet (James, 2006) consists of a grid made from columns and rows. It is an environment that can make number manipulation easy and somewhat painless.

In a spreadsheet the column is defined as the vertical space that is going up and down the window. *Letters* are used to designate each column's location. In the spreadsheet below, the COLUMN labeled C is highlighted.

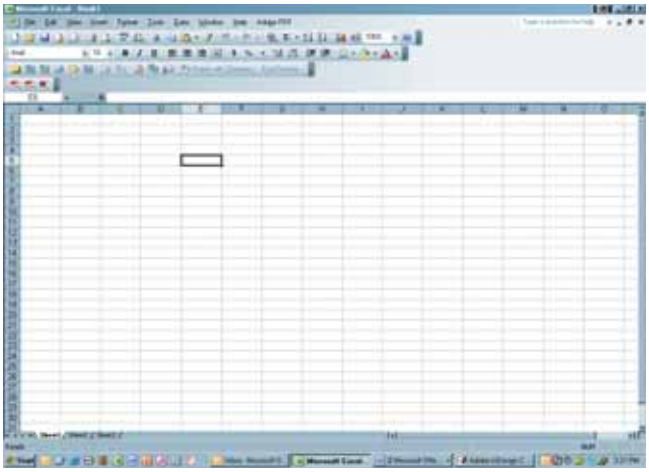


The row is defined as the horizontal space that is going across the window. *Numbers* are used to designate each row's location. In the spreadsheet below, the ROW labeled 3 is highlighted.



[1] *Note* Information in this section is from "A basic tutorial of Excel" by Brad James, (2006). Adapted with permission.

The cell is defined as the space where a specified row and column intersect. Each cell is assigned a name according to its column letter and row number. In the spreadsheet below the CELL labeled E5 is highlighted. When referencing a cell, you should put the column first and the row second.



There are three basic types of data that can be entered.

- Labels - (text with no numerical value). Labels are text entries. They do not have a value associated with them. We typically use labels to identify what we are talking about.

- Constants - (just a number -- constant value). Constants are entries that have a specific fixed value. If someone asks you how old you are, you would answer with a specific answer. Other people will have different answers, but it is a fixed value for each person. Sometimes constants are referring to dollars, sometimes referring to percentages, and other times referring to a number of events. These are typed into the computer with just the numbers and are changed to display their type of number by formatting. You will need to know what the numbers are representing, e.g. a fraction, a data, a percentage.

- Formulas - (a mathematical equation used to calculate). Formulas are entries that have an equation that calculate the value to display. The equation will be updated upon the change or entry of any data that is referenced in the equation. There is a list of the functions available for formulas within Microsoft Excel, under the menu insert down to function. Formulas or functions must begin with an equal sign (=). When entering formulas into a spreadsheet, make as many references as possible to existing data. If you can reference that information you do not have to type

it in again. And more importantly, if that other information changes, you do not have to change the equations.

Spreadsheets have math functions built into them. The most basic operations are the standard multiply, divide, add and subtract. These operations follow the order of operations (just like algebra). Let's look at some examples.

Selecting cells in an equation is a very important concept of a spreadsheet. You need to know how to reference the data in other parts of the spreadsheet. You can select several cells together if we can specify a starting cell and a stopping cell. This will select ALL the cells within this specified BLOCK of cells. If the cells that we want to work with are not together (non-contiguous cells) you can use a comma to separate the cells or by holding down the control-key (command key on a MAC) and selecting cells or blocks of cells. A comma will be inserted automatically to separate these chunks of data. Consider the data below:

	A	B
1	3,005	3,135
2	16,637	18,253
3	3,311	3,206
4	7,185	7,089
5	25,931	25,207
6	35,231	35,580
7	5,503	5,485
8	1,773	1,759
9	8,763	7,914

The most common function in any spreadsheet is the sum function. The sum function takes all of the values in each of the specified cells and totals their values. To obtain the sum of the numbers in columns A1 -A9 highlight the cell below A9, click on the Σ on the toolbar and hit enter on the keyboard. This is also a drop-down box that you can select other functions such as average, maximum, minimum, and standard deviation.

You often need to format the numbers to display the appropriate number of decimals, dollar signs, percentage, red (for negative dollars), etc. It is best to keep numbers describing similar items as uniform as possible. If we have the number 3.53262624672423, we would probably have to make the column wider. We need to set the number of decimal places to what is important. If this was a dollar figure it should be $3.53.

A question that everyone (who has ever worked on a spreadsheet) has asked at one time or another is, “Where did all my numbers go?” or "Where did all of those ####### come from and why are they in my spreadsheet?” The problem is the number trying to be displayed in a particular cell does not have enough width to display properly. To clear up the problem you just need to make the column wider.

Here are two ways to change the column width:

1. Select the column (or columns) with the problem by clicking on their labels (letters). Then you choose the MENU FORMAT. Go down to COLUMN and over to WIDTH and type in a new number for the column width.

2. Move the arrow to the right side of the column label and click and drag the mouse to the right (to make wider) or left (to make smaller). Let up on the mouse button when the column is wide enough.

If you have a spreadsheet designed and you forgot to include some important information, you can insert a column into an existing spreadsheet. What you must do is click on the column label (letter) and choose Columns from the Insert menu. This will insert a column immediately to the left of the selected column.

Numbers can usually be represented quicker and to a larger audience in a picture format. Excel has a chart program built into its main program. The Chart Wizard will step you through questions that will (basically) draw the chart from the data that you have selected. There are many types of charts. The two most widely used are the bar chart and the pie chart. The bar chart is usually used to display a change (growth or decline) over a time period. You can quickly compare the numbers of two different bar charts to each other. The pie chart is usually used to look at what makes up a whole. If you had a pie chart of where you spent your money you could look at the percentages of dollars spent on food, transportation, housing, and other categories. You can add legends, titles, and change many of the display variables.

Data on the Internet

The Primer will provide information on accessing the U.S. Census Bureau and the KIDS COUNT Datacenter. These will give the reader the tools needed to access and download other data from other sites. It is important to remember that each database has a particular set of attributes that need to be kept in mind when selecting one to use:

- What subjects are covered by the database?
- What does the database include?

- What year was the data collected?
- Is the data from a sample or 100% data?

The type of data you need will determine the data source you use and each data source has its own unique strengths and weaknesses.

U. S. Census Bureau

There are various data surveys and censuses available from the U. S. Census Bureau. The Primer will review the data that a general user would access most often. These are: the Decennial Census, the American Community Survey, and the Population Estimates Program

Decennial Census

"The U.S. Census counts every resident in the United States. It is mandated by Article I, Section 2 of the Constitution and takes place every 10 years. The data collected by the decennial census determine the number of seats each state has in the U.S. House of Representatives and is also used to distribute billions in federal funds to local communities".

(Source: U. S. Census Bureau website: http://2010.census.gov/2010census/about/ Accessed 6/23/2011).

American Community Survey

The American Community Survey (ACS) is an ongoing survey that provides data every year -- giving communities the current information they need to plan investments and services. Information from the survey generates data that help determine how more than $400 billion in federal and state funds are distributed each year.

To help communities, state governments, and federal programs, the ACS asks about:

Age	Family & Relationships	Education
Sex	Income & Benefits	Veteran Status
Race	Health Insurance	Disabilities
Where you work & how you get there	Where you live and how much you pay for some essentials	

All this detail is combined into statistics,which are used to help decide everything from school lunch programs to new hospitals.

Tips for using American Community Survey data

1. ***All American Community Survey (ACS) data are estimates.***
   - The Census Bureau collects American Community Survey data from a sample of the population in the United States and Puerto Rico--rather than from the whole population All ACS data are survey estimates. To help you interpret the reliability of the estimate, the Census Bureau publishes a margin of error (MOE) for every ACS estimate.
2. ***American Community Survey collects and releases data by the calendar year for geographic areas that meet specific population thresholds.***

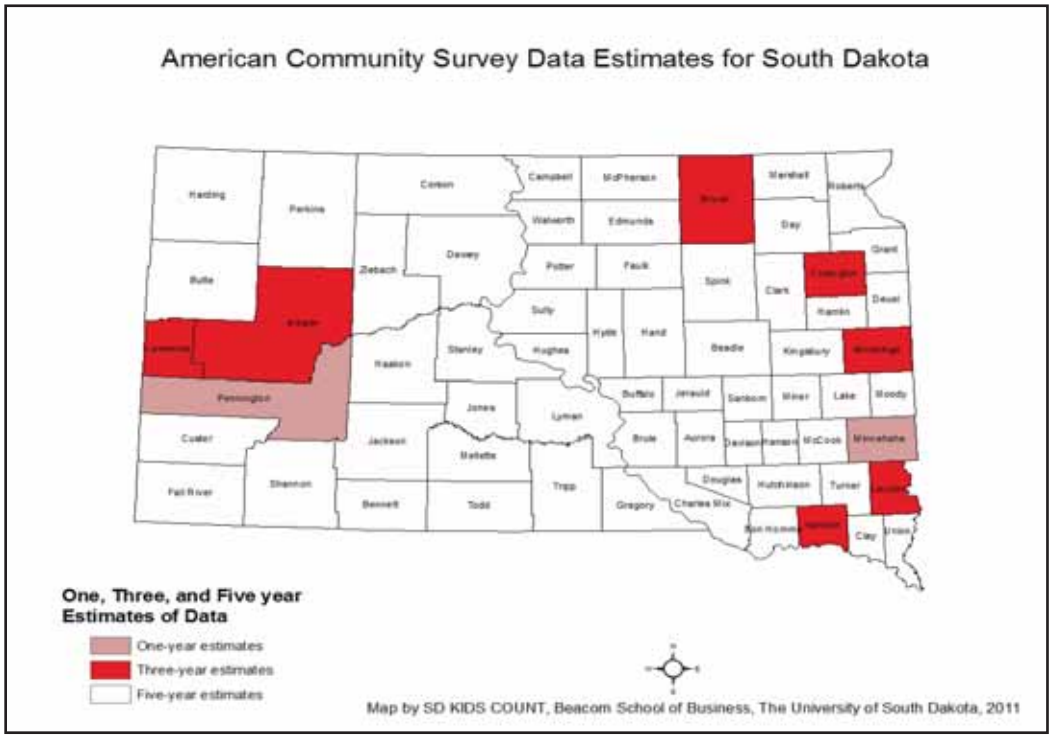
(Source: U. S. Census Bureau website: http://www.census.gov/acs/www/about_the_survey/american_community_survey/ Accessed 6/23/2011).

Distinguishing features of ACS 1-year, 3-year, and 5-year estimates

(Source U.S.Census Bureau website: http://www.census.gov/acs/www/guidance_for_data_users/estimates/ Accessed 6/23/2011)

1-year estimates	3-year estimates	5-year estimates
12 months of collected data	36 months of collected data	60 months of collected data
Data for areas with populations of 65,000+	Data for areas with populations of 20,000+	Data for all areas
Smallest sample size	Larger sample size than 1-year	Largest sample size
Less reliable than 3-year or 5-year	More reliable than 1-year; less reliable than 5-year	Most reliable
Most current data	Less current than 1-year estimates; more current than 5-year	Least current
Best used when	**Best used when**	**Best used when**
Currency is more important than precision	More precise than 1-year, more current than 5-year	Precision is more important than currency
Analyzing large populations	Analyzing smaller populations	Analyzing very small populations
	Examining smaller geographies because 1-year estimates are not available	Examining tracts and other smaller geographies because 1-year estimates are not available

Based on the geographies in the table above, American Community Survey data for South Dakota counties would be as follows. Minnehaha and Penninton Counties will have one-year estimates; and Brown, Bookings, Codington,Lawrence, Lincoln, Meade, and Yanton will have three-year estimates. The remaining counties will have five-year estimates. (See map on the next page). This means there is timely data for places with greater population but for places with less population (rural areas) data are averaged over time, either three- or five-year averages to increase the sample size in hopes of reducing the margin of error.



Population Estimates

The Census Bureau's Population Estimates Program (PEP) produces July 1 estimates for years after the last published decennial census (2000), as well as for past decades. Existing data series such as births, deaths, Federal tax returns, medicare enrollment, and immigration, are used to update the decennial census base counts. PEP estimates are used in federal funding allocations, in setting the levels of national surveys, and in monitoring recent demographic changes.

A methodology reference accompanies most of our population estimates offerings.With each new issue of July 1 estimates, the Population Estimates Program revises estimates for years back to the last census. Previously released estimates are superseded. Revisions to estimates are usually due to revised or updated administrative input data, changes in methodology, or legal boundary changes, especially for subcounty units of government, such as incorporated places. The frequency of estimates and availability of demographic detail vary by geographic level. (Source U.S.Census Bureau website: http://www.census.gov/popest/topics/terms/Accessed 6/23/2011)

The margin of error in the data for rural areas provides a range by which the actual number could vary. For example , the estimated total households in Hughes County, South Dakota in the 2005-2009 American Community Survey (ACS) was an estimated 6,907 with a margin of error of +/-207. This means there

could be as many as 7,114 households or as few as 6,700 households.

Which data source should you use?
The 2010 Census shows the number of people who live in the U.S. and the American Community Survey shows how people live.

- Use numbers from the 2010 Census to obtain counts of the population and their basic characteristics (sex, age, race, Hispanic origin, and homeowner status).
- Use data from the American Community Survey to obtain demographic, social, economic, and housing characteristics.
- Use data from the Census Bureau's Population Estimates Program in the years between censuses. The Census Bureau's Population Estimates Program produces official population estimates for the nation, states, counties, cities and towns, plus housing unit estimates for states and counties.

(Source: U. S. Census Bureau website: http://www.census.gov/acs/www/about_the_survey/american_community_survey/ Accessed 6/23/2011).

You can access data for the Decennial Census, the American Community Survey, and population estimates from American FactFinder (AFF). American FactFinder allows users to search and analyze demographic and economic information about the nation and communities using a variety of methods. They include:

- Keyword Search
- Predefined Topics
- Geographic Search
- Population & Ancestry Group Search
- Industry Search

How to download the data you need:
Information about downloading data from American FactFinder can be found on the SD KIDS COUNT website: **http://www.usd.edu/sdkidscount/Toolbox.cfm**

KIDS COUNT Data Center
http://datacenter.kidscount.org/Explore hundreds of measures of well-being for kids across the nation, or in your state, city, or community. Create customized maps, graphs, and charts.

How to download the data you need: The information can be found on the SD KIDS COUNT website: **http://www.usd.edu/business/south-dakota-kids-count/upload/KIDS-COUNT-Data-Center.pdf**

Examples of Other Data Sites that have downloadable data

http://www.childtrendsdatabank.org/ *The Child Trends DataBank* - National trends and research on over 100 key indicators of child and youth well-being

http://webappa.cdc.gov/sasweb/ncipc/mortrate10_sy.html *WISQARS - (Web-based Injury Statistics Query and Reporting System)* - An interactive database system that provides customized reports of injury-related data.

http://www.ers.usda.gov/Data/SNAP/index.htm *Foodstamp Machine Local Data* - State and county-level estimates of Supplemental Nutrition Assistance Program (SNAP) participation and benefit levels, combined with area estimates of total population and the number of persons in poverty.

Examples of Other Data Sites that have data already compiled by state and/or county

http://nces.ed.gov/ *The National Center for Education Statistics (NCES)* - the primary federal entity for collecting and analyzing data related to education.

http://www.statehealthfacts.org/ *Statehealthfacts.org* - A project of the Henry J. Kaiser Family Foundation that provides free, up-to-date, and easy-to-use health data for all 50 states.

http://www.marchofdimes.com/Peristats/ *PeriStats* - Developed by the March of Dimes, the site provides free access to US, state, county, and city maternal & infant health data.

http://www.fedstats.gov/ *Fedstats* - provides access to the full range of official statistical information produced by the Federal Government without having to know in advance which Federal agency produces which particular statistic.

http://feedingamerica.org/hunger-in-america/hunger-studies/map-the-meal-gap.aspx *Feeding America* - Map the Meal Gap project provides data about hunger at the local community level.

Organizing your Data Sources

Once you have data sources you may want to keep a listing of the organization, website address, type of data obtained, and the date data were obtained. This is useful if different people are responsible for collecting or finding data. This type of data can also be used to keep track of your data sources.

Organization	Website	Type of Data obtained	Date obtained
SD KIDS COUNT	http://www.sdkidscount.org	County data on children and families in South Dakota	June 9, 2011
KIDS COUNT Data Center	http://datacenter.kidscount.org/	Children in extreme poverty in South Dakota	June 12, 2011
2010 Census	http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml	2010 Demographic Profile Data	June 17, 2011

Chapter Five - Know how to calculate the numbers

Summary of Chapter

The chapter will instruct learners on manipulating the data using simple math and statistical procedures.

Objectives of Chapter

I. To understand and perform basic mathematical calculations

II. To understand and perform basic statistical calculations.

Math Tools

Many people dislike math. They dismiss math with "I'm a word person, not a numbers person". Anyone who works with data has to be able to do math, but the math you need to know is very basic.

Convert a fraction into a decimal:
Divide the top number of the fraction by the bottom number
General example: 5/8 = "5 divided by 8" = 0.625
Example: The state data center newsletter noted five-eighths of the 1,750 families in County X are living below the poverty level. You calculate the decimal by selecting the "5", then the divide button, then the "8", then the "=" button, and discover that 5/8= 0.625. You then multiply 0.625 by 1,750 to learn that there are about 1,094 families living in poverty in County X.

Convert a decimal into a percentage:
Multiply by 100, or simply move the decimal point two places to the RIGHT.
0.858 = 85.8% or 1.255 = 125.5%
Example: If five-eighths of the families are living in poverty that is 0.625, which is 62.5% of them.

Convert a percentage into a decimal:
Divide by 100, or simply move the decimal point two places to the LEFT.
43.7% = 0.437 or 148% = 1.48
Example: If 62.5% of families are living in poverty that is 0.625

Convert a percentage into a fraction:
Often, the best way to express a percentage in a story is by turning it into a fraction, such as "About two-thirds of the adults in County Y have college degrees". The simplest way is to remember that "percent" really means "per 100". Therefore, for example, 43% is the fraction 43/100.

For converting percentages into common fractions or useful phrases refer to the information below. You can be approximate; for example, if the percentage in question is 35%, it's perfectly accurate to say "More than a third…"

5% "one out of every 20"
10% "a tenth" or "one in ten"
20% "a fifth" or "one in five"
25% "a fourth" or "one in four"
30% "three out of ten"
33% "a third" or "one in three"
40% "two out of every five"
50% "half"
60% "three-fifths" or "three out of five"
66% "two-thirds" or "two out of three"
70% "seven out of 10"
75% "three-fourths" or "three out of four"
80% "four out of five"
90% "nine out of ten"

Calculate X% of Y:
Convert X% into a decimal, then multiply that decimal by Y
20% of 90 = 0.20 x 90 = 18 or 130.5% of 45 = 1.305 x 45 = 58.7
Example: As above, if 62.5% of 1,750 families are living in poverty, then 0.625 x 1,750 = 1,094 families living in poverty.

Compare two numbers using percentages (A is what percent of B?):
A divided by B, multiplied by 100 turns the decimal into a percentage
5 is what percent of 8?: 5/8 = .625 = 62.5%, so 5 is 62.5% of 8
8 is what percent of 5?: 8/5 = 1.6 = 160%, so 8 is 160% of 5
Example: The median household income in one neighborhood is $20,000, compared to the county's overall median household income of $24,000. Therefore, the neighborhood's median income is (20000/24000) x 100 = 0.833 x 100=83.3% of the county median household income.

Compare numbers using percentage difference (A is what percentage MORE or LESS than B?):
(A divided by B) - 1, then multiply by 100 to turn the decimal into a percentage. Use MORE THAN if the answer is positive, and LESS THAN if it's negative.
5 and 8: (5/8) - 1 = 0.625 - 1 = -0.375 = -37.5%, so 5 is 37.5% less than 8
8 and 5: (8/5) - 1 = 1.6 - 1 = .6 = 60%, so 8 is 60% more than 5

Example: The median household income in one neighborhood is $20,000, compared to the county's overall median household income of $24,000. Therefore, the neighborhood's median household income is (20000/24000)-1 = 0.833-1=-0.167=-16.7%. So you can say the neighborhood's median household income is about 17% less than, or below, the county's median household income.

How to Perform Data Calculations

There are numerous computations that KIDS COUNT uses that involve rates, percents, and percent change over time. Many of these calculations are similar. The computation used depends on what we want to say about the data. Basically, each of these calculations involves dividing one piece of data by another.

Percent

A 'percent' means one part in one hundred. Ten percent means 10 out of 100. To calculate a percent, divide the number in a sub-group (or smaller number) by the number in the total group and multiply by 100.

Example: Percent: (number in subgroup / number in whole group)
(number of low birth weight births in South Dakota, 2008 [783]) / (number of all live births in South Dakota [12,066]) x100.

$$783 / 12{,}066 = 0.0648 \times 100 = 6.5\%$$

Almost 7% of babies born in South Dakota were low birth weight in 2008.

Rate

A rate is simply the number of things per some other number, usually 100, 1,000 or other multiples of 10. A percentage is a rate-per 100. Depending on the size of the subgroup, a rate greater than 100 is often used. This is the case in health statistics such as infant mortality, child death, etc. To calculate a rate you need three pieces of information-(1) the total group number, (2) the number in the sub group and (3) the 'per' number—per 1,000, 10,000, or 100,000. The per number is your multiplier.
Example: Rate: (number in sub-group / number in whole group) x multiplier
(number of child deaths in South Dakota, 2008 [26] / number of children ages 1-14 in South Dakota [154,779]) x multiplier.

$$26 / 154{,}779 = 0.0001679 \times 100{,}000 = 16.8$$

The child death rate for South Dakota in 2007 was 16.8 per 100,000 children ages 1-14.

Rates can also be obtained per month or per day. For 2008 the figures would show:

12,074 births /12 months = 1,006 infant births per month in South Dakota

12,074 births / 365 days = 33 infant births per day in South Dakota

Percent Change

Change from one time period to the next can be calculated. This is called the 'percent change'. It is calculated in the following manner: (newer year number - older year number) / older year number x 100

Example:

2008 est population under age 20 in South Dakota = 221,461 (newer number)
2000 population under age 20 in South Dakota = 234,385 (older number)

(221,461 - 234,385) = - 12,924
-12,924 / 234,385 = -.0055 x 100 = - 5.5%

This figure indicates that between 2000 and the 2008 estimates for the number of children under age 20 *decreased* by 5.5%.

Statistical Tools

There are statistical terms that you need to understand and be able to calculate when working with data. Such statistics are particularly useful in helping summarize and put into context the numbers you'll be examining.

For instance, it would be cumbersome to describe the economic status of a city by listing the incomes of every single resident. Instead, we take all those incomes and collapse them into a more meaningful number.

Aggregates

Aggregate is just the fancy way of saying total. Certain data are supplied as aggregates within a given piece of geography, such as the total value of housing or the total number of apartments with 1, 2, or 3 bedrooms.

Mean

The mean, also called the average, is the sum of a group of values, divided by the number of values in the group. For instance, assume you have the total populations of each of 210 census tracts, total the column containing the population count for each tract and then divide by 210 to find that the average size of the census tracts.

Median

The median simply is the middle value in a list of values that have been sorted in numerical order. Say there are N values in the list: If N is an odd number, the median is the value located at number [(N-1)/2)+1. If N is even, the median is the average of item N/2 and (N/2)+1.] Excel and other spreadsheet software packages will figure the median of a group of numbers without you having to sort and count halfway down the list.

Don't confuse the median with the mean, although it's quite possible for the mean and the median of a group of numbers to be the same. The median is often used in data tabulations instead of the average, particularly with variables that may not be distributed evenly, such as age, income, or home value. The reason is that the median isn't as likely to be affected by extreme values.

Mode

The "mode" is the value that occurs most often. If no number is repeated, then there is no mode for the list.

Rate

This is a way of standardizing values so that different areas can be compared fairly. You can recognize that a rate is being talked about if the word "per" is in the description, such as "deaths per 1,000 live births."

To calculate a rate you need three pieces of information: (1) the total group number, (2) the number in the sub group and (3) the 'per' number--per 1,000, 10,000, or 100,000. The "per" number is your multiplier.

Example: Rate: (number in sub-group / number in whole group) x multiplier (number of infant deaths in the state [70] / number of live births in the state [10,698]) x multiplier (70 / 10,698) = 0.0065 x 1,000 = 6.5
Thus the infant mortality rate for South Dakota, in 2002, was 6.5. This means that for every 1,000 live births that occurred in the state in 2002, there were 6.5 infant deaths.

Weighted Average

This is for figuring out an average for a larger area when all you have is information from the smaller areas that comprise it. For example, imagine that three tracts have these median home values: $60,000, $80,000, and $120,000. The simple average of those values is $86,667. But let's say that the three tracts have this many houses respectively: 3,000, 2,000, and 500. The weighted average is calculated by multiplying the individual averages by the number of values that

created each of them, getting the total, and then dividing by the total number of values. So it looks like this:
(($60000*3000) + ($80000*2000) + ($120000*500)) / (3000 + 2000 + 500) = $72,727. This is only an approximation of what you could get if you had the individual values for each of the 5,500 homes. But $72,727 is likely to be a lot closer to the real value than $86,667.

Percentiles

Like the median, percentiles are just values at specified intervals in an ordered list of value; a value in the 90th percentile, for instance, is greater than 90% of the rest of the values. The median is the 50th percentile. Commonly used "n-tiles" are quartiles (four segments divided at 75%, 50% and 25% of the list) and quintiles (five segments divided at 80%, 60%, 40% and 20% of the list.)

- *Series breaks*- changes in the way data are defined, classified, or collected from one time period to another. How we define data variables can change. The concept of "person" seems straightforward yet the 1790 census counted slaves as three-fifths of a person and American Indians not at all.
- *Confidentiality*- small numbers breach confidentiality, hence data are aggregated
- *Time lags*- collecting, analyzing and disseminating is time consuming Census 2000 data is not ready to be released until 2001!
- *Differing definitions and methods*- similar concepts are defined differently, applied differently and thus there are different results. Over 70 different federal agencies collect, analyze, and disseminate data. "Employment" and "Income" mean something different to the Bureau of Economic Analysis, Bureau of Census, and the Bureau of Labor Statistics.
- *Population figures*-more than one statistical agency produces own series on this variable. Results are different and thus confusing because different methodologies are used.
- *Revisions*-estimates are revised and keeping up with revisions is an ongoing battle

Chapter Six - Know how to present your assembled data

Summary of Chapter

The chapter will assist learners in pulling together research information into concise presentation form. Learners will have the opportunity to develop and present their data reports.

Objective of Chapter

I. To present data in a manner that is understandable to your audience.

Data in a spreadsheet

Instructions for creating a chart using Microsoft Excel 2003, 2007 or 2010:

(*Note: By using the search engine of your choice, Google, Yahoo, Ask.com, GoodSearch.com, to name a few you can find information on how to make a chart using Microsoft Excel in whatever version you have*.)

An audio tutorial of how to make a chart using Microsoft Office Excel 2003, 2007, or 2010 can be found at:

http://office.microsoft.com/en-us/support/training-FX101782702.aspx

The training has the following goals:

* Create a chart using Excel.

* Make changes to a chart after you create it.

* Understand basic chart terminology.

The course includes:

• Self-paced lessons and practice sessions for hands-on experience. Practices require the Excel version you have on your computer.

• A short test at the end of each lesson; tests are not scored.

• A Quick Reference Card you can take away from the course.

Creating a chart is not difficult, however you need to understand these aspects when making a chart:

1. What type of chart to use?

2. How should the chart look, i.e., positioning of rows and columns of information?

Chart Type

What type of chart do you want?

An explanation of the different chart types available in Microsoft Excel can be found here:

http://office.microsoft.com/en-us/excel/HA012337371033.aspx

A brief explanation of the more common charts used follows.

Bar Graphs

The Bar graph displays the real-time value of specific variables and is mainly used for comparisons. Bar graphs consist of an axis and a series of labeled horizontal or vertical bars that show different values for each bar. The numbers along a side of the bar graph are called the scale. A double bar graph gives two pieces of information for each item on the vertical axis, rather than just one.

Line Graph

A line graph is a way to summarize how two pieces of information are related and how they vary depending on one another. The numbers along a side of the line graph are called the scale. The line graph follows a set of data over time and is used for analyzing trends in a specific variable.

Pie Charts

A pie chart is a circle graph divided into pieces, each displaying the size of some related piece of information. Pie charts are used to display the sizes of parts that make up some whole.

Use numbers & words to tell your story

"Words and numbers are of equal value, for, in the cloak of knowledge, one is warp and the other woof. It is no more important to count the sands than it is to name the stars. Therefore, let both kingdoms live in peace." Juster, 1961, p. 77.

There are many guidelines available to follow for presenting data. A resource from the Population Reference Bureau is below.

Guidelines for Effective Data Presentations - www.prb.org
then select PRB Library. The guide gives practical advice and examples in the art of presenting data to nonspecialist audiences.

Also, keep the following in mind:
• Check your numbers, then check them again. Have someone else proof your work and your arithmetic. Go back and check your numbers against the original source (be sure to keep a copy of all your sources). Remember, one error in one table can kill the overall credibility of your material—and of your organizational efforts.

• If you can avoid it, do not hire an outside "expert" to do your work for you. If you do not do your own analysis, you can not explain it to your intended audience. If you feel the data are too complicated for you to work with, it may be too complicated for your audience to understand.

• Make your presentation simple. Percentages and rates, for example, are great statistics—just about everyone can grasp them well enough to get your point. If you believe there are two different audiences for your report that need different levels of information, consider preparing two separate reports.

• Try to show change over time. Remember, you want to use as many points in time as possible but at least two points in time separated by at least five years.

• Use the most recent data you can get. It is easier to convince your audience that a problem exists now if up-to-date numbers make up your case. When the most recent data are more than a year or two old, be sure you identify them as "the most recent data available".

• Always try to use data that show some intervention will make a difference.

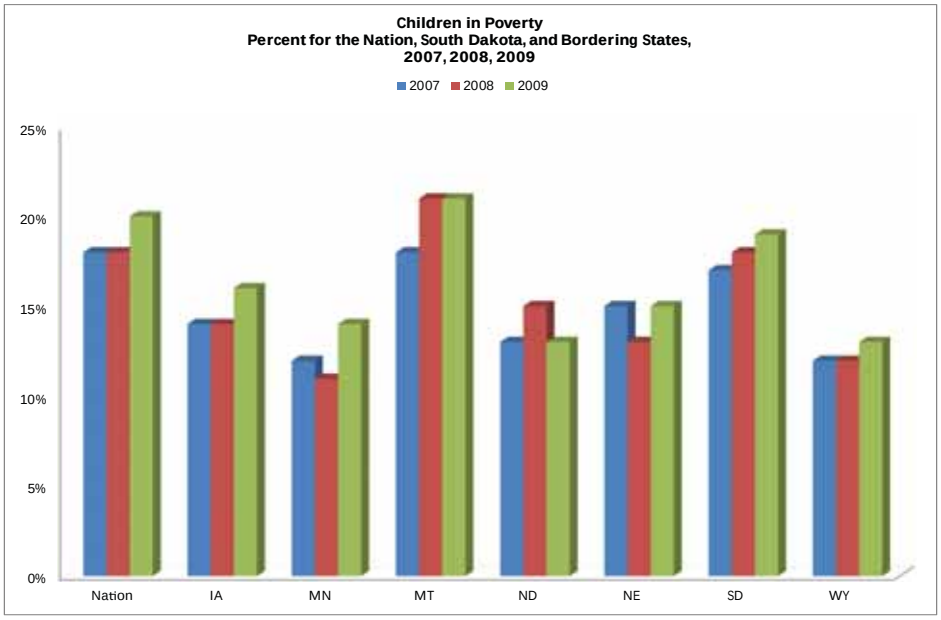
Using the principles outlined above, suppose that you work in a four county area. Your want to compile information on the percent of children in poverty for the past three years for the Nation, South Dakota and Bordering States. You could compile the information like this: Nation 2007-18%; 2008-18%; 2009-20%; IA 2007-14%; 2008-14%; 2009-16%; MN 2007-12%; 2008-11%; 2009-14%; MT 2007-18%; 2008-21%; 2009-21%; ND 2007-13%; 2008-15%; 2009-13%; NE 2007-15%; 2008-13%; 2009-15%; SD 2007-17%; 2008-18%; 2009-19%; WY 2007-12%; 2008-12%; 2009-13%.

A way that would make the data easier to understand is by placing the data in a table.

Percent of children in poverty (income below $21,756 for a family of two adults and two children in 2009)			
	2007	**2008**	2009
Nation	18%	18%	20%
IA	14%	14%	16%
MN	12%	11%	14%
MT	18%	21%	21%
ND	13%	15%	13%
NE	15%	13%	15%
SD	17%	18%	19%
WY	12%	12%	13%

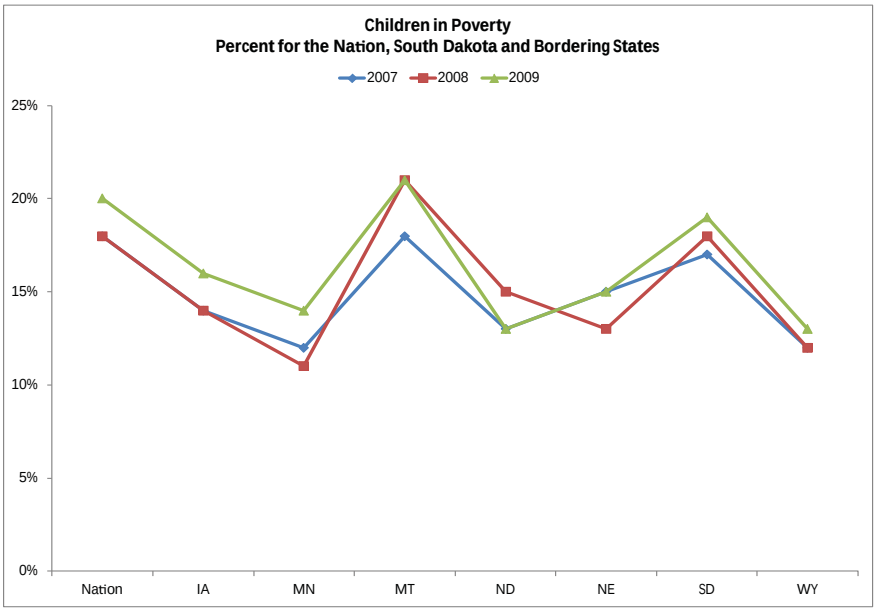
Source: The Annie E. Casey Foundation, KIDS COUNT Data Center, www.kidscount.org/datacenter.

Many things are easier to understand if portrayed visually and graphically, rather than in text or verbally. You need to determine the best way to display data so it is easily understandable to your audience. Would a bar chart or a line graph be more appropriate?



Source: The Annie E. Casey Foundation, KIDS COUNT Data Center, www.kidscount.org/datacenter.

A line graph would look like this:



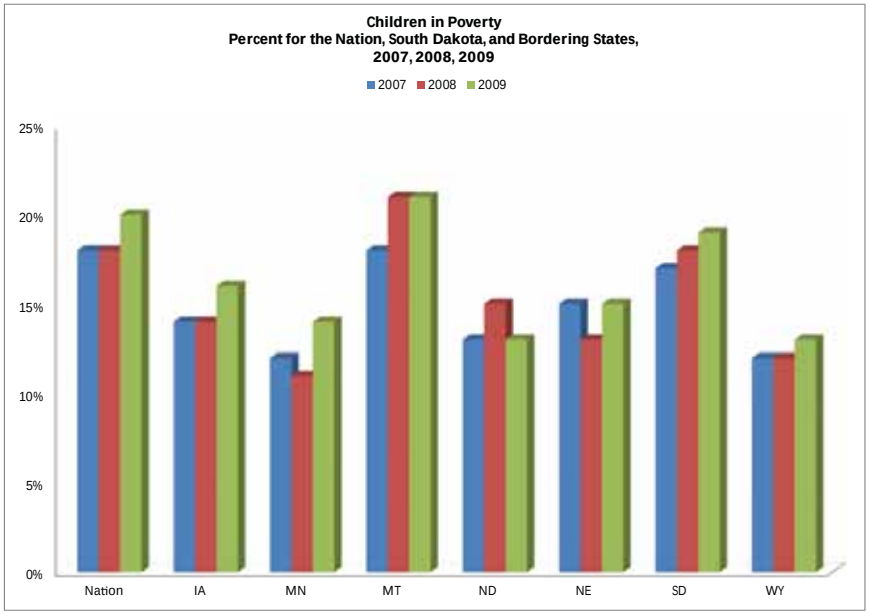
Source: The Annie E. Casey Foundation, KIDS COUNT Data Center, www.kidscount.org/datacenter.

Orientation of the data series

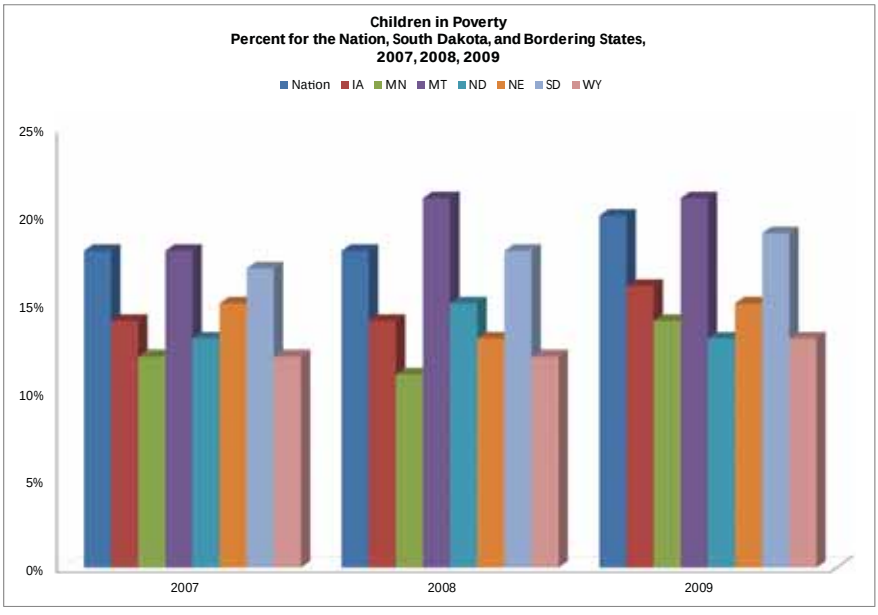
When you create a chart, Microsoft Office Excel determines the orientation of the data series. After you create a chart you can change the way that worksheet rows and columns are plotted in the chart by switching rows to columns or vice versa.

For example, when you create a chart for two rows and columns of worksheet data, Excel plots the data by rows, but you may want to plot the data by columns instead.

The chart on the next page shows Children in Poverty - Percent for the Nation, South Dakota, and Bordering States, 2007, 2008, 2009



Would switching the rows and columns in the bar chart above provide a "better" display of the information? If we switch our rows and columns, *[click the chart that contains the data that you want to plot differently. This displays the chart tools, adding the Design, Layout, and Format tabs. On the Design tab, in the Data group, click Switch Row/Column. When you click the Switch Row/Column button, Excel immediately changes the way the data in the chart is plotted by switching between the worksheet rows and the columns]*, the chart would look similar to this:

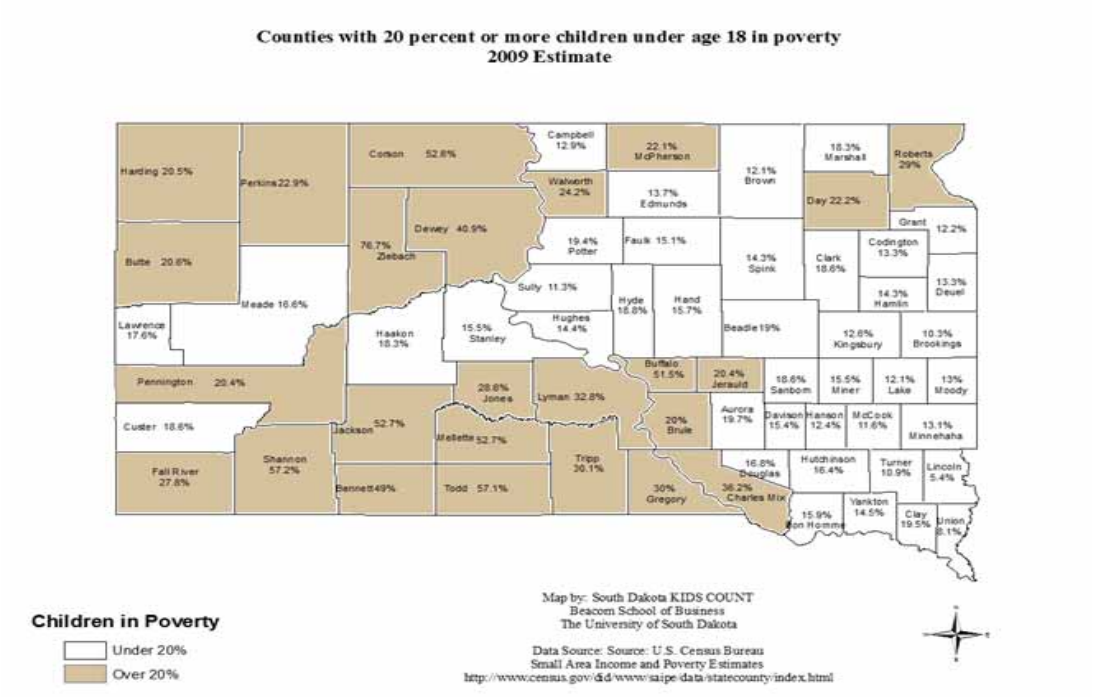


Using GIS to tell your story

Current technology the social sciences are using is Geographic Information Systems (GIS). GIS is powerful software technology, which allows unlimited amounts of information to be linked to a geographic location. With a digital map you can see a state, region, county, city, neighborhood, and block in terms of demographics, income levels, poverty rates, epidemics, high school dropout rates, and much more.

GIS can help people understand social problems, show discrepancies between needs and resources, or help legislators and other to see where the problem is occurring (as opposed to providing statistics, charts and graphs).

The map below shows how GIS can be used. The map shows counties with 20 percent or more of children under age 18 in poverty.



References

Merriam-Webster Online Dictionary (2005). Merriam-Webster, Incorporated. Retrieved January 13, 2006 from http://www.m-w.com/dictionary/data.

Merriam-Webster Online Dictionary (2005). Merriam-Webster, Incorporated. Retrieved January 13, 2006 from http://www.m-w.com/dictionary/gather

Merriam-Webster Online Dictionary (2005). Merriam-Webster, Incorporated. Retrieved January 13, 2006 from http://www.m-w.com/dictionary/assemble

Cochran, C. and Stuefen, R. (2004). The Economic Impact of the Child Care Industry in South Dakota. Vermillion, The University of South Dakota, Business Research Bureau, South Dakota KIDS COUNT.

James, B. W, (2006) A basic tutorial of Excel. Retrieved October 11, 2006 from The University of South Dakota, TRIO Web site: http://usd.edu/trio/tut.cfm.

Juster, N. (1961). The Phantom Toll Booth. New York: Random House.

Royse, D. (1991) Research methods in social work. Chicago: Nelson-Hall Publishers.

The University of South Dakota
www.usd.edu

Located in Vermillion, SD, USD was founded in 1862 by the Dakota Territorial Legislature, and is the state's oldest university. The University was accredited by the North Central Association of College and Schools in 1913. The University offers more than 100 academic programs in its eight schools and colleges. It is also home to South Dakota's only law and medical schools, and College of Fine Arts. The University also has a professionally accredited School of Business and the College of Arts and Sciences is a core for a liberal arts education.

USD Social Work Program
www.usd.edu/health-sciences/social-work/index.cfm

Social workers engage with children, families, schools, public and mental health systems and other agencies. Some social workers conduct research or are involved in planning or policy development. Social work is a profession for those with a passion for social justice and a desire to make a difference in improving people's lives. We offer the following academic programs: Bachelor of Science (BS) in Social Work and Master of Social Work (MSW.).

Department of Social Work
Julian Hall Room 253
The University of South Dakota
414 East Clark Street
Vermillion SD 57069
Phone: 605-677-5401
Fax: 605-677-7213
socialwork@usd.edu
www.usd.edu/socialwork

Beacom School of Business
www.usd.edu/business/index.cfm

The School of Business offers students a quality business education that goes beyond expectations. Our alumni have become global leaders in such fields as accounting, management and health services administration.

South Dakota KIDS COUNT Project
Beacom School of Business
The University of South Dakota
414 East Clark Street
Beacom Hall - Room 350
Vermillion, SD 57069-2390
Phone: 605.677.6432
Fax: 605.677.5427
E-mail: kidscount@usd.edu
Website: www.sdkidscount.org

August 2011