Creating Effective Graphs, Tables, and Charts

Meg Gorzycki, Ed.D.

The purpose of these exercises is to improve students' understanding of graphic representations of data and information, and to assist them in making decisions about how to best represent data and information in their own research.

Standards for Graphic Illustration of Data

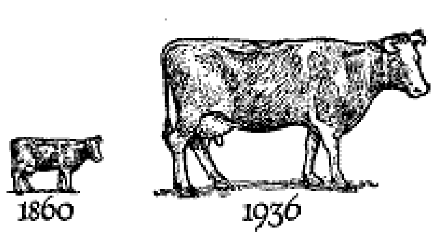
The advantage of using charts, graphs and tables when presenting information is that graphic illustrations offer a visual summary of data that is often dense and complex. Graphic illustrations also may suggest cause-effect relationships, patterns over time, and changes in natural conditions or in the behavior of organisms and humans.

The effective display of data regards three standards (Klass, 2002). Frist the author must have data that is meaningful. A researcher, for example, might find that the average amount of sodium daily consumed by men is different than the amount daily consumed by women, but unless the information is offered in the context of a meaningful discussion about the significance of the findings or the implications of the finds, the data loses its relevance to our lives. In addition, data should be displayed efficiently, which means the graphics should be easy to read. Finally, the data should be unambiguous. Ambiguity in graphic illustrations occur when authors use elements of design that distort lines, bars, and sections that represent parts of a whole. Sometimes data is made ambiguous because the labels in a graph or chart are not accurate.

In his classic text, *How to Lie with Statistics* (1954), Darrell Huff illustrated how it is possible for researchers to examine the same data and arrive upon radically different conclusions and how it is possible to "spin" the data in order for researchers to say what they want it to say. It is wise, therefore, for readers to study the text that accompanies graphic illustrations; and, if text does not accompany the graphics, readers may do well to suspend their belief and judgment until additional information is collected to verify what the graphic is "saying."

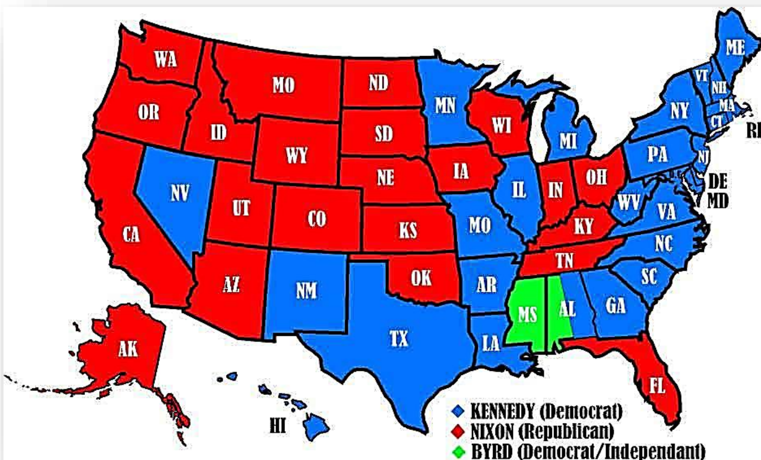
Huff also illuminates how pictures can lie. Consider the graphic illustration of the Crescive Cow, (Huff, p. 72) which was intended to show readers that the number of milk cows in the U.S. increased from about 8 million in 1860 to about 25 million in 1936. The 1936 cow should be three times the size of the 1860 cow, but has been distorted to represent a cow that is four to five times the size of the 1860 cow.

Figure 1: The Crescive Cow



Graphic illustrations of election results are sometimes misleading. Consider the illustration here of the 1960 presidential election.

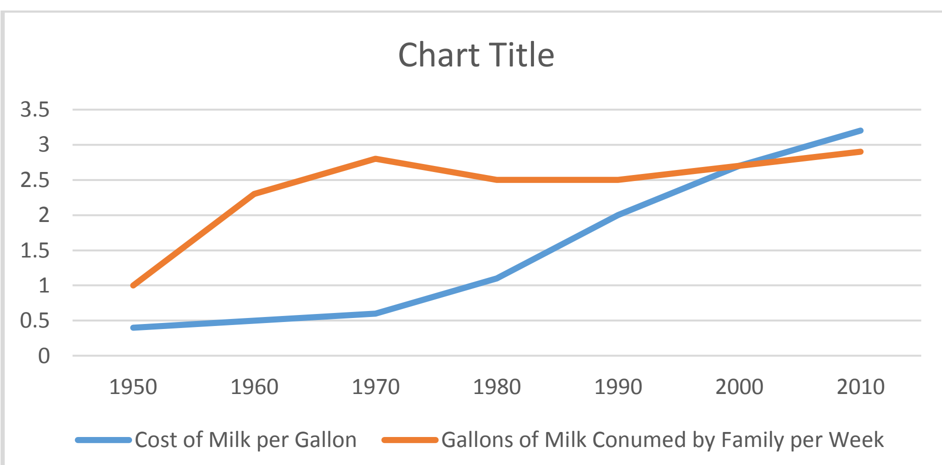
Figure 2: Presidential Election, 1960 States Carried
(http://www.historycentral.com/elections/1960state.html).



The red states appear to take up more space than do the blue states, suggesting that Nixon won the race. The illustration does not, however tell the reader what percentage of voters in each state voted for which candidate, and so what appears to be a victory for Nixon—or at least a close race—cannot be verified until researcher document whether a candidate captured states' electoral votes by a margin of 51% or by a margin of 93%.

Distortions can also occur when graphs use more than one standard of measurement in a single illustration. Carefully examine the following graph (Figure 3)

Figure 3: Time Series of the Cost of Milk and Number of Gallons Consumed per Family



Can you detect the problems with the chart in figure 3?

The problem is the result of a scaling effect, or using two different scales of measurement in a single chart or graph.

The most problematic aspect of the chart is that the Y axis, (the vertical line containing whole numbers and fractions) is not labeled and so readers cannot tell whether they represent the price of milk in dollars and cents or how many gallons of milk families drink on average per week. It is nearly impossible to draw meaning from this data. In addition, readers have no idea whether the prices reflect a national average or represent prices in a given region of the country, which can vary greatly. Readers do not know what the researcher meant by "family," which could be defined a group of four people or six people, or something else; readers also do not know if all families were included in the study or whether the data was taken from working class and poor families or more affluent and middle class families.

Chart Types and Components

There are several types of charts used to illustrate data, and the type of data often determines which chart will be best suited for illustrating data clearly and accurately. The types of charts include: pie charts, bar charts, line graphs, scatterplots, and boxplots. All charts have the same components:

- Labels for the Y and X axis
- Legends or keys for identifying various grouping of data
- The source of the data
- The graphical elements of the particular form, such as lines, bars, pie wedges, etc.
- The scale of the ranges of data used in the Y axis (sometimes also needed for the X axis)

The organization of these components in the actual illustration should not interfere with the reader's ability to read the chart. Sometimes gridlines, legends and labels are essential to for the reader to interpret the graph and so crowding the components visually is problematic. Figure 4 presents readers with problems.

Figure 4: Percentage of Walleye with Dermal Sarcoma (Caution: Based on Fictitious Data)

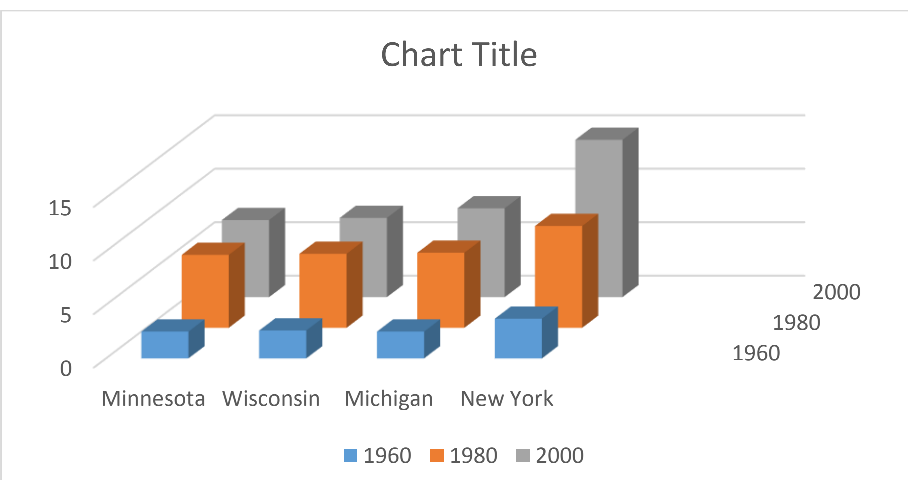
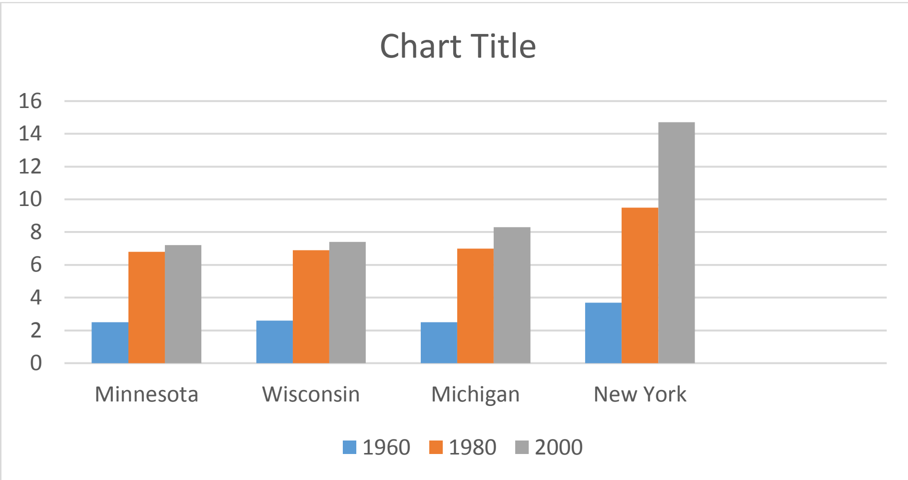


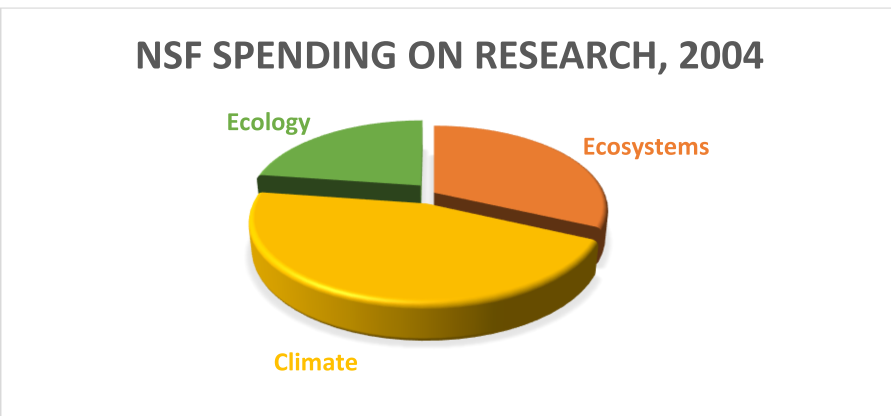
Figure 3 attempts to give readers a look at how the viral disease, Dermal Sarcoma, has infected walleye (a fish native to freshwater lakes in the Great Lakes region), but may be confusing readers because the scale on the Y axis is not well labeled, nor does its scale allow for readers to see subtle changes in data. The incremental jump from 5% to 10% does not clearly illustrate points between those percentages. In addition, the bars create the illusion that the percentages for Minnesota and Wisconsin during 1980 and 2000 were identical and they are not. Note what happens in Figure 5 when the chart is revised.

Figure 5: Revision of Figure 4



Figures 6 and 7 illustrate that the choice of graphic design may distort the reader's understanding of the data.

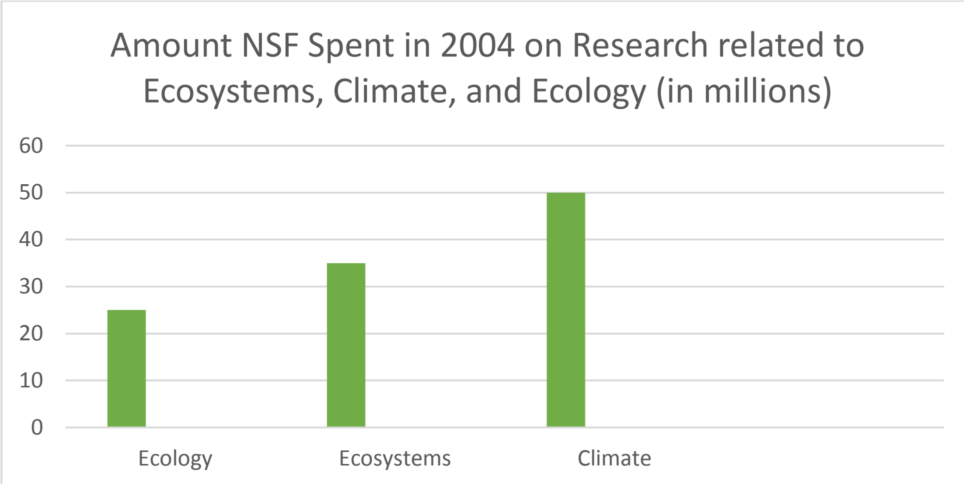
Figure 6: National Science Foundation Spending in 2004 on Research by Category (Adapted from Neff, M. (2005). How climate changed ecology: Recent trends in NSF funded ecology research. Arizona State University. Retrieved from http://archive.cspo.org/_old_ourlibrary/presentations/neff_final.pdf.)



Apart from the fact that the pie chart in figure 6 is not properly labeled so the actual percentages are displayed, it is probably the wrong chart to use. Readers also do not know how much money is represented in the chart. Did the NSF spend a total of a billion dollars on research in 2004, or did they spend $607.00? A pie chart tells readers that the researcher is representing pieces of 100% of something. In this case readers are asked to assume that whatever the NSF spent on research in 2004 is represented by spending in three categories, and this may not be accurate because the NSF sponsors research in many other areas.

Using the same data, researchers can create a more accurate chart to show readers what the NSF spent in 2004. Note the changes in Figure 7.

Figure 7: Revised figure 6.



Pie charts should be used when the researcher wants to illustrate percentages of a given phenomenon within a whole population. This holds true even if the whole population is represented by a sample population, such as a sample of 800 undergraduates on a single campus to represent the 24,000 undergraduates on that campus. Pie charts are best used when they are two-dimensional, as three-dimensional (such as Figure 6) pies sometimes distort the weight or value of each "piece."

Klass (2002) asserts that scatterplots are the best type of graphs to use when illustrating the relationship between two variables. In Figure 8, readers see that there exists a strong relationship, or correlation, between two variables. Note that each dot on the chart represents one county.

Figure 8: Obesity Rates of Children Under Age 18 and Distribution of Fresh Produce Markets by County (Based on fictitious data)

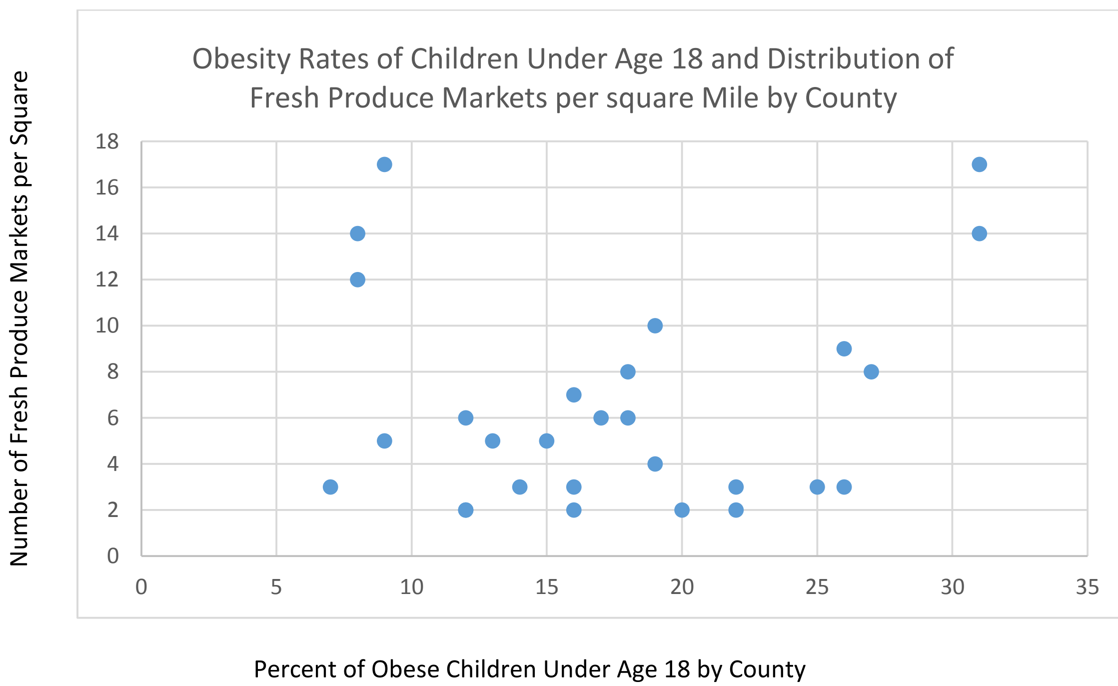


Figure 8 reveals ***no correlation*** bewteen the distribution of fresh produce markets and obesity in children under age 18. Readers see that there is no pattern linking obesity and the distribution of fresh produce markets. Figure 9, however, shows a ***positive correlation*** as the number of obese children increase as does the number of fresh produce markets. Figure 10 shows a negative correlation as the percentage of children with obesity increases as the number of fresh porduce markets decreases.

Figure 9: Obesity Rates of Children Under Age 18 and Distribution of Fresh Produce Markets by County (Based on fictitious data)

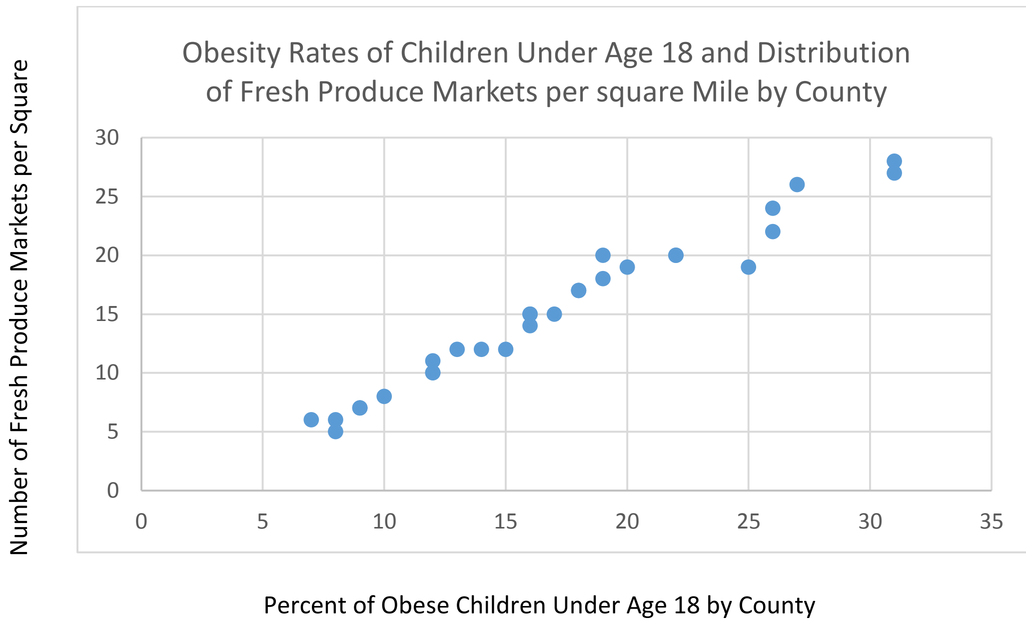
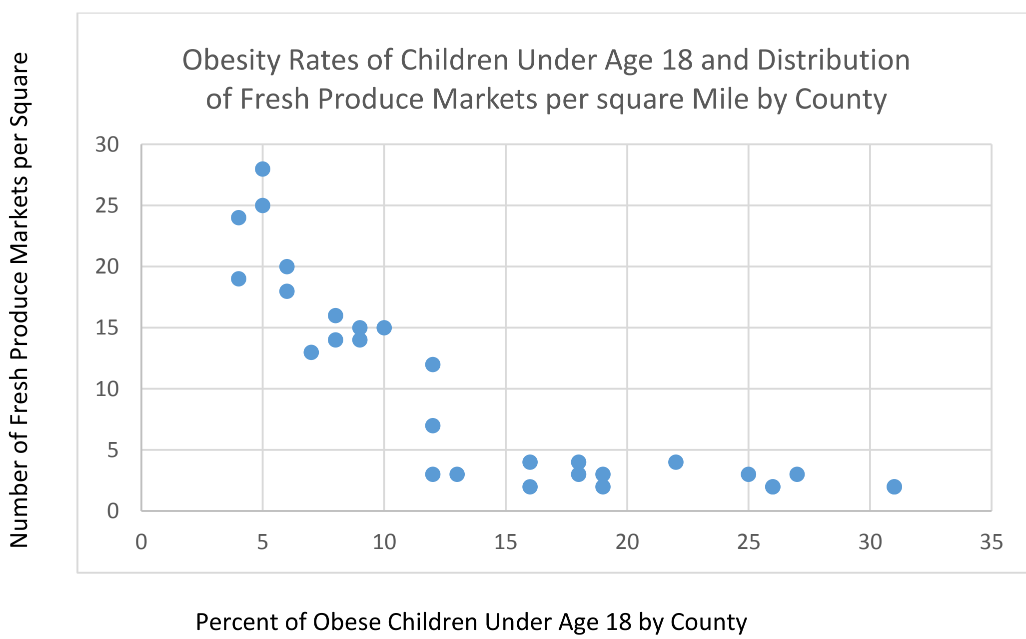


Figure 9: Obesity Rates of Children Under Age 18 and Distribution of Fresh Produce Markets by County (Based on fictitious data)



Tables

Tables provide a means of organizing data into categories of information that is easy to grasp. Like graphs and charts, tables can be very confusing as they may represent complex phenomenon and multiple variables.

Because tables can be complex, it is important to understand that they often describe parts of a whole. When a table reads, for example, that the recovery rate for sea birds contaminated with oil spilled from a damaged ship, the research is telling readers that out of all the contaminated birds in the study, a certain percentage of them survived and a certain number perished. The researcher may want to draw attention to what research has discovered about other aspects of survival after an oil spill at sea, and so may incorporate fields in the table that represent different variables.

Table 1 is based on fictional data. It illustrates how researchers might illustrate data regarding the survival rates of various birds impacted by an oil spill in the Gulf of Mexico. The purpose of the table is to illustrate the survival rates of different species of birds that were contaminated by the oil spill.

Table 1 Survival Rate of Birds Contaminated by an Oil Spill in the Gulf of Mexico, 2022:

Specie	Survived	Perished	Total
Reddish Egrets	40%	60%	100%
Brown Pelicans	35%	65%	100%
Least Terns	25%	75%	100%
Snowy Plover	20%	80%	100%
Total	30%	70%	100%

Note that there are two “Totals” in the Table 1. The total in the banner, last column, to the right of “Perished,” represents the total percent of percentages entered into each variable for each specie. Because the total percent of Reddish Egrets that perished and that survived add up to 100%, readers know that researchers have accounted for each subject in their sample population of Reddish Egrets in the study. The second “Total,” at the bottom of the first column, presents readers with the total percent of all subjects in the sample population that either survived or perished.

Thought tables that pack lots of data into one illustration may provide readers with lots of important information, they may also become harder to read. Table 2 represents data related to the data in Table 1. To increase the scope of information in the table, researchers wanted to tell readers about all the birds collected for the study—not just the ones that were contaminated by oil—but all that were living in a one-mile radius of where the oil spill occurred. Researchers also wanted to reveal the percentage of males and females in each specie that either survived or perished. Note what happens to the design of the table when these variables are added.

Table 2: Contamination and Survival Rate of Male and Female Birds Impacted by an Oil Spill in the Gulf of Mexico, 2022

Birds Impacted by Oil Spill in the Gulf of Mexico, 2022					
Specie	**Male Uncontaminated**	**Male Contaminated**	**Female Uncontaminated**	**Female Contaminated**	**Total**
Reddish Egret	2%	9%	2%	11%	24%
Brown Pelican	2%	8%	4%	10%	24%
Least Tern	1%	5%%	3%	8%	17%
Snowy Plover	3%	11%	5%	16%	35%
Total	8%	33%	14%	45%	100%

What does Table 2 tell readers?

- The banner indicates that the table will represent many birds in the table and the columns show readers how those birds have been organized into categories in the study.
- There are two sets of totals in the table; the last square of the last column (bottom far right) tells readers that the total percentage of each specie in the study tallies 100%, and that the number of uncontaminated and contaminated males and females in each species also tallies 100%. In other words, all the birds that were part of the study are accounted for.
- The table tells readers which birds were part of the study, but this does not mean that more types of birds or wildlife were not affected by the oil spill…

The table requires some explanations. For example: why is it that more Snowy Plovers were twice as likely as Least Terns to be impacted by the oil spill? Why was it that more females were contaminated?

The researchers are also obligated to tell readers how they gathered the data, as the data collections methods might have some influence in their capacity to have measured contamination and survival rates accurately. We do not know, for instance, whether the researchers included dead birds and counted them as “non-survivors,” or whether researchers collected only living birds, then classified them as “uncontaminated” or “contaminated,” and then tracked whether they lived or died after attempts to clean them and restore their health.

References

Huff, Darrell. (1954) *How to Lie With Statistics*. N.Y: W.W. Norton & Co. Inc.

Klass, G. (2002). *Constructing good Charts and Graphs*. Illinois State University. Retrieved from http://lilt.ilstu.edu/gmklass/pos138/datadisplay/sections/goodcharts.htm

Oldberg, Terry. An Ethical Problem in the Statistics of Defect Detection Test Reliability. Speech to the Golden Gate Chapter of the American Society for Non-destructive testing. March 10, 2005. Found at: http://www.ndt.net/article/v10n05/oldberg/oldberg.htm Accessed September 5, 2008.

Riegel, K. (1973) Dialectical Operations: The Final Period of cognitive Development. *Human Development*. (16) 346-370.

Schield, Milo. (1999) Statistical Literacy: Thinking Critically About Statistics as Evidence. Of Significance, Vol. I, Issue 1 Association of Public Data Users (APDU) Available at: www.StatLit.org/pdf/1999/SchieldAPDU.pdf.

Schield, Milo. (2004) Statistical Literacy curriculum Design. Curriculum Design Roundtable sponsored by the International Association of Statistical Educators (IASE) Available at: wwww.StatLit.org/pdf/2005SchieldASE.pdf.

Schield, Milo. (2005) Statistical Prevarication: Telling Half Truth Using Statistics. Communicating Statistics conference sponsored by the International Association of Statistical Educators (IASE) Available at: www.StatLit.org/pdf/2005ShieldIASE.pdf.