4-1-2004

Health Information and Numerical Data

Kevin Kawamoto

Follow this and additional works at: http://commons.pacificu.edu/inter04

Health Information and Numerical Data

Rights

Health Information and Numerical Data

Posted on **April 1, 2004** by **Editor**

By **Kevin Kawamoto** <kawamoto@u.washington.edu>

INDEX:

.01 **Numbers Tell (or Help Tell) Stories** (return to index)

Whenever I used to teach courses involving computer-assisted research, one thing I would point out to students is the wealth of numerical data on the Internet that is freely available to the public. "Numbers are not just numbers," I would say as I showed them mind-numbing rows and columns of numbers on the overhead projector just to make a point. "Numbers often have stories to tell us. The challenge is to find the stories in this mass of numbers." And then we would go about finding those stories.

A popular and potentially interesting electronic database that one can use to introduce students to numerical data is the U.S. Census Bureau's Web site at www.census.gov. This amazing resource is full of rich demographic data, much more than one could ever digest in a single quarter or semester. Using census data on the Web, teachers can design exercises in which students isolate certain columns of numbers (e.g., population change data from one year to another year), download this data onto a computer's hard drive, "clean up" the data so that they are in the proper format and free of "debris," and then upload that cleaned-up data into an Excel spreadsheet for computer analysis.

For example, students can look at population change in the U.S. from 1990 to 2000. They would have three columns in their Excel spreadsheet to start with: Column 1) State name; Column 2) 1990 population per state; and Column 3) 2000 population per state. There would be 51 rows of

data, each row representing a state, plus the District of Columbia. Then they would "tell" Excel (using a simple mathematical formula) to compute the population change from 1990 to 2000 (i.e., [1990 population subtracted from 2000 population] divided by [2000 population]), which would quickly give them an answer in decimal form for each state, plus the District of Columbia. (Their answers for each state would appear in the fourth column, if that is where they decide to put them.) The fifth column could be a conversion of the decimal (e.g., 0.167967311) to a simple percentage (e.g., 17%).

Then the human analysis comes into play. Students would study the fifth column (percentages) and see which states have had the biggest growth and which states have had the smallest growth (or even negative growth) in the ten years of under study. This is the beginning of their story. If the analysis showed, for example, that Nevada and Arizona had the biggest growth in population from 1990 to 2000, that's an interesting piece of a story that can be developed and expanded in many different directions. Usually, once they have the results of the computer analysis and are confident that the answers are accurate, and they isolate some specific results that have the potential for story development, the next question they need to ask is "WHY?"

WHY did the population in Nevada and Arizona grow so much, outpacing the other 48 states and the District of Columbia? What cultural, social, technological, economic, and political factors may have led to this? (Not all of these may be answerable, but they are worth asking.) Numerical data can probably help to answer some of these questions, but also more qualitative methods of inquiry begin to be useful as well, such as expert interviews. A journalist, for example, might take this finding to the people knowledgeable about development in that state and ask for informed opinions about what led to the growth.

Working with numerical data also teaches students the importance of looking at real numbers, as well as looking at numbers as proportions of a whole. Looking at real numbers exclusively, the population of California or New York grew by many more people in that ten-year period than either Nevada or Arizona. But as a percentage (of change) of the total population, Nevada and Arizona easily came out on top. Students learn about comparing groups – such as state populations – where the group sizes are quite different from each other.

Crime databases such as the Uniform Crime Report are another excellent teaching tool. Students compare crime rates in cities, learn about the importance of knowing "rates" of crime as opposed to simply looking at real numbers, and understand the methods used to collect this kind of information. In fact, an understanding of data collection methods is critical before using any numerical data found on the Internet or elsewhere.

There are all kinds of interesting and even fun numerical analysis tools on the Web. Salary calculators can help students compare the purchasing power of their dollars from one city to another. They allow you to enter the salary in one city (say $55,000 in Seattle, Washington) and tell you what the equivalent salary (based on a whole range of factors) in another specified city would be (say $63,671 in Evanston, Illinois, for example).

A lesson on numerical data would also likely include at least an introduction to statistics, especially inferential statistics, and under what conditions statistics can be generalizable to the larger population.

In addition to finding tremendous volumes of numerical data on the Internet, student should learn that many library reference sections or government documents divisions have extensive databases with numerical data on a wide range of subjects. Searching a library catalog under subject headings such as "statistics," "judicial statistics," "criminal statistics," "vital statistics," "educational statistics," "commercial statistics," "military statistics," and so forth can start them on a journey of discovery. Almanacs, books of facts, and statistical yearbooks are often good collections of numerical data. [1]

02. **Health Statistics** (return to index)

When it comes to health stories, numerical data can provide a level of detail and specificity that general words alone cannot convey. Saying that "diabetes is a growing problem in the United States" and "large numbers of people in the U.S. have diabetes" may be true, but these lack the specificity that make the information meaningful. More compelling would be something that quoted an official source, with specific numbers, such as:

"More than 13 million Americans have been diagnosed with diabetes, according to the National Center for Chronic Disease Prevention and Health Promotion. This is more than double the number of people who were diagnosed in 1980. The American Diabetes Association says that 5.2 million others are unaware they have the disease."

This is just a nutshell paragraph that needs much more elaboration, but it moves the story about diabetes from broad generalities to more specific information. This could then go in many other directions, perhaps focusing on a person or several people who have the disease, explaining the difference between Type I and Type II diabetes, relating the increase to the growing problem of obesity (especially in children), or talking about screening or diagnostic methods, or symptoms.

Stories about major health problems and phenomena can be enriched and enhanced with numbers. The level of detail one wants to go into in regard to numerical data (e.g., performing secondary data analysis on existing data) depends on the goals of the storyteller (which could be a student, teacher, journalist, crusader, advocate, politician, etc.) and the intended audience. Students in an advanced statistics class could take data collected from elsewhere and re-analyze them or conduct a related survey to see whether the new findings support the previous one(s). There are a number of studies that have been published, for example, looking at whether fiber in the diet may help prevent colon cancer. The findings have not been consistent. Other scientists may want to re-examine the data and analyze them in different ways, or find flaws in the method, or do new studies. While numbers may help illuminate our understanding of a problem, they do not represent "truth," as conflicting findings often remind us.

One good thing about numbers is that they can be summarized and presented in charts and graphs. One of the simplest graphs to start students off with is a population pie chart for the U.S. showing the breakdown between males and females. Gradually, more complex graphical depictions can be created, such as age groups (probably using a bar chart) and income levels. This may give students ideas about how to depict other information – e.g., the increase or decline of HIV infection from 1980 to the present or the increase or decrease of teen pregnancy over a period of years. Students can also speculate about what may have led to changes in the numbers over time. Was a particular intervention effective? Did people's eating habits change? Were detection methods more sensitive?

Where does one find health statistics on the Web? The following section provides some starting places to look up health-related numerical data. But this is just a sample of what's available. Some data in other locations may be proprietary because they have been collected and are being used for commercial research and development. Other data may be available to the public but not on the Web. Asking a librarian for help, especially one who works in a health sciences or medical library, would be a way to find more elusive data. But for starters, follow the links the below.

03. **The Numbers** (return to index)

Centers for Disease Control and Prevention, U.S. Dept. of Health and Human Services
National Center for Health Statistics
http://www.cdc.gov/nchs/

Centers for Disease Control and Prevention, U.S. Dept. of Health and Human Services
Data and Statistics Section
http://www.cdc.gov/node.do/id/0900f3ec8000ec28

U.S. Census Bureau
Statistical Abstracts of the United States
http://www.census.gov/statab/www/

National Institutes of Health, U.S. Dept. of Health and Human Services
Health Information
http://health.nih.gov/

American Cancer Society
Statistics
http://www.cancer.org/docroot/STT/stt_0.asp

The White House, U.S. Government
Social Statistic Briefing Room: Health
http://www.whitehouse.gov/fsbr/health.html

U.S. Food and Drug Administration, U.S. Dept. of Health and Human Services
http://www.fda.gov/

Substance Abuse and Mental Health Services Administration
U.S. Dept. of Health and Human Services
Office of Applied Statistics
http://www.oas.samhsa.gov/

National Library of Medicine
National Information Center on Health Services Research & Health Care Technology
http://www.nlm.nih.gov/nichsr/hsrsites.html

University of Michigan
Documents Center
http://www.lib.umich.edu/govdocs/sthealth.html

United Nations
UN Statistics Division
http://unstats.un.org/unsd/

04. **Conclusion** (return to index)

Numbers do not tell the whole story. In fact, as critics of statistics have said many times, numbers can be misleading – intentionally or unintentionally. The classic book, *How to Lie with Statistics* [2], is still a useful warning about how numbers may not be all they are cracked up to be, or to represent. Understanding how to work with numbers and statistics will help make students more critical readers of stories that use statistics.

Much of the numerical data that the government collects about any number of topics, including health, is made freely available to the public because the public has already paid for them through their taxes. Encourage students to make use of this store of valuable information.

Other health-related information on the Internet, not necessarily numerical, has been discussed extensively in past columns [3]. Using a wide range of health information sources is advisable.

REFERENCES (return to index)

[1] One organization that provides user-friendly advice to journalists and journalism students about computer-assisted research is the National Institute for Computer-Assisted Reporting, or NICAR, which has a useful Web site at www.nicar.org. While this site is geared to those in the news media industry, it is a good resource for the non-expert interested in data analysis. A quick search on the Web will yield many other sites that try to help non-experts understand basic data analysis.

[2] Darrell Huff, *How to Lie with Statistics* (New York: Norton, 1982).

[3] Past columns about health informatics include:

"Health Information Online Abundant and Varied"
http://bcis.pacificu.edu/journal/2002/11/kawamoto.php

"Teaching Students About Cyberhealth Information"
http://bcis.pacificu.edu/journal/2003/01/kawamoto.php

"Older Adults and the Internet"
http://bcis.pacificu.edu/journal/2003/02/kawamoto.php

"Computer Technology in Health Care Settings"
http://bcis.pacificu.edu/journal/2003/04/kawamoto.php

"Privacy and Personal Health Information"
http://bcis.pacificu.edu/journal/2003/06/kawamoto.php

"Healthy Learning Can Be Fun: Digital Media and Health Education"
http://bcis.pacificu.edu/journal/2003/07/kawamoto.php

"Compassion Knows No Border: The Research of Patricia Radin"
http://bcis.pacificu.edu/journal/2003/09/kawamoto.php

"Health Related Blogs"
http://bcis.pacificu.edu/journal/2004/01/kawamoto.php

This entry was posted in Uncategorized by **Editor**. Bookmark the **permalink [http://bcis.pacificu.edu/interface/?p=2950]** .

2 THOUGHTS ON "HEALTH INFORMATION AND NUMERICAL DATA"

Leo
on **October 2, 2013 at 9:12 AM** said:

You are so cool! I do not suppose I've read a single thing like this before.
So great to find another person with genuine thoughts on this topic.
Really.. many thanks for starting this up. This website is
one thing that is needed on the internet, someone with a bit of originality!

Ward Defee
on **February 6, 2014 at 12:07 AM** said:

Hallo, können Sie PM Sie mit ein paar Fragen über das, was Sie erwähnt haben auf dieser Webseite?