Cool Statistics! Econ B2000 Kevin R Foster, CCNY Fall 2012

Learning Objective: Students will understand that statistics is important, exciting, and worth hard work.

Stats pervade our everyday life, particularly online. Google is dominant in search because they've figured out how to give people what they want, using sophisticated statistics-based algorithms to do so. After you type even a single letter the site is making suggestions, based on your past browsing.

Smart uses of stats drive many top companies, from Google's advertisements to Facebook's friend suggestions and advertisements to Netflix with its movie suggestions to Amazon's "people who bought ... also bought ..." Even game designers use stats on attention spans – Angry Birds didn't just happen! Every company is always looking to improve, to get a competitive edge.

Hal Varian, chief economist at Google and previous Dean of the business school at Berkeley, notes "I keep saying that the sexy job in the next 10 years will be statisticians. And I'm not kidding." (*New York Times*, August 9, 2009, "For Today's Graduate, Just One Word: Statistics.")

Of course economics is built on statistics, from GDP to unemployment to interest rates or consumer confidence.

Companies battle to best use available data to forecast trends. Walmart is dominant because they built an inventory management system that used real-time data. In finance, quants rule: every trader is looking for new algorithms to figure out market moves, even (with high-frequency trading) to calculate the answer just a split second faster than their competitors. Risk managers then construct models to protect against those traders. In sports, data-driven managers are dominating the old guard who made decisions based on their gut instinct. Politicians depend on polls for nearly every utterance. Market research informs smart business decisions. Manufacturer's quality control is based on stats. Of course statistics is a bedrock of science and was developed in parallel to the scientific revolution.

We can use statistics to learn about the world and try to overcome our misconceptions, as in this presentation from the TED talks by Hans Rosling, 2006,

http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html



Where he explains why his students are dumber than chimps.

One of the earliest time-series charts, showing the effect of war on Great Britain's national debt, is due to William Playfair (credited as being the inventor of the pie chart, bar chart, and other displays still in use today):



The Divisions at the Bottom are Years, & those on the Right hand Money.

This chart shows the British debt from "the Revolution" (i.e. the Glorious Revolution when William and Mary ascended) to the end of the American Revolution.

Some charts from the NYTimes looked at the costs of American wars (Week In Review July 25, 2010):

The Current Conflicts Are the Second-Most Expensive...

The combined costs of Iraq, Afghanistan and other post-9/11 operations are exceeded only by those of World War II.



But They Have Involved Far Fewer People...

Each figure represents 500,000 people who served during the course of each conflict (includes those fighting in theaters of war and those serving in support areas).

World War II	TOTAL SERVING: 16.1 million	***********	***************	
Vietnam War	8.7	************	!!!	
Korean War	5.7	***********		
World War I	4.7		Spanish-American War	0.3
Gulf War	2.2	****	War of 1812	0.3
Civil War (Union)	2.2	****	Revolutionary War up to	0.25
Current Wars	2.1	****	Mexican War	0.08

At a Smaller Cost to the Economy.

War costs as a percentage of United States gross domestic product in the peak year of each conflict. Estimate not available for Revolutionary War-era G.D.P.



The modern guru of graphical display is Edward Tufte. This illustration, which he commonly uses and indentifies as "probably the best statistical graphic ever drawn", is by Charles Joseph Minard:



This shows Napoleon's invasion of Russia. The width of the line indicates the number of soldiers as well as how they moved to Moscow. The bottom graph shows the cold on the retreat (in degrees Reaumur; multiply by 4/5 to get Celsius so it bottomed at -24°C).

Wainer (among others) also pick out another important innovation, the early graphical display which had the most immediate improving effect on human welfare, by John Snow.

Snow's map shows deaths from cholera (the dots) as well as the location of well pumps (the x's) in London.



Figure 21.1. A map drawn by John Snow in 1854. The dots represent where a cholera death occurred. The crosses show the location of water pumps. Snow noticed that most of the deaths were centered on the Broad Street pump. Princeton University Library, Department of Rare Books and Special Collections.

The story tells that Snow went to the Broad Street pump, removed the handle, and shortly cholera deaths began to diminish. Prevailing public health theory had told that cholera was transmitted through the air; this chart and the associated experiment provide strong evidence that the problem was water quality.

Sometimes statistics reveal surprising things about the world; my colleague Prof. Bengoa noted a recent economic study showing that "size matters": there is a surprising correlation between economic growth and the average size of what the paper coyly refers to as "the male organ."

Figure 1: GDP in 1985 and the size of male organ in 76 countries, ORGAN in linear and quadratic form, $\bar{R}^2=0.15$



Male organ (cm)



There are many modern uses and surveys. The NY Times had this graphic showing the different uses of time during the day, gathered from the American Time Use Survey (which we'll use in class). Here http://www.nytimes.com/interactive/2009/07/31/business/20080801-metrics-graphic.html is the full interactive chart where you can compare the time use patterns of men and women, employed and unemployed, and other groups . (The article is http://www.nytimes.com/2009/07/31/business/20080801-metrics-graphic.html is the full interactive chart where you can compare the time use patterns of men and women, employed and unemployed, and other groups . (The article is http://www.nytimes.com/2009/08/02/business/20080801-metrics-graphic.html is the full interactive chart where you can compare the time use patterns of men and women, employed and unemployed, and other groups . (The article is http://www.nytimes.com/2009/08/02/business/20080801-metrics-graphic.html is the full interactive chart where you can



You can just browse online to find many beautiful examples of data presentation. Some are more effective as art; some are better at presenting the data.

Many Eyes, from IBM, http://manyeyes.alphaworks.ibm.com/manyeyes/page/Visualization_Options.html, gives an overview of both the classics and more recent innovations such as word clouds.

Here is a word-cloud image of Shakespeare's most-used words,

age art back bad bear beauteous beauty beauty's behold black blessed blood born breast breath bright bring brow call canst care change child cold cruel day days dead dear death decay deeds deep delight desire disgrace dost doth duil dumb earth end ere excuse eyeeyees face fair false faults fear find fire flowers form foul found fresh friend full gentle give glass glory good grace great grow hand happy hate hath head hear heart's heaven heaven's heavy hell hold honour hours ill joy kind leave leaves lie lies life light lines lips live lives living long lose loss lov'st love love's lovely loves loving made make makes making mayst memory men mind mine mistress muse nature night o'er outward part parts past pen pity place pleasure poor power praise precious pride prove putrich rose roses sad sake save set shadow shalt shame shouldst Show sight sing skill slave soul speak spirit stand stay store straight strong summer's sun swear

Sweets taught ten tender thing things thought thoughts thyself till time time's times tis tongue treasure true true truth verse view virtue waste whilst white wilt wit woe

words WOrld worst Worth write wrong youth

Visual Complexity has some beautiful images, <u>http://www.visualcomplexity.com/vc/</u>.

Such as this one on NYC Subway Ridership 1905-2006, http://diametunim.com/shashi/nyc_subways/



Note the slider on the right to change the time back to 1905.

This is <u>an interesting site</u> that uses Amazon's Mechanical Turk (which blasts simple calculations to many human workers) to do things like rate pictures, to determine how attractiveness changes with age for men and women:

Attractive



Where, except for babies, women are consistently rated as more attractive although the variation across age is interesting. The <u>whole paper</u> about perceptions of people based on thumbnail pictures is here.

Then there are the awful statistics. From *Damned Lies and Statistics*, Joel Best nominates the single worst, http://books.google.com/books?id=685UteNN_4AC&dg=best-statistics&printsec=frontcover&source=in&hl=en&ei=ospNTMK1MYH5&AaGooko&sa=X&oi=book result&ct=result&resnum=11&ved=oCEsOGAEwCg#v=onepage&g&f=false

When I read the quotation, I assumed the student had made an error in copying it. I went to the library and looked up the article the student had cited. There, in the journal's 1995 volume, was exactly the same sentence: "Every year since 1950, the number of American children gunned down has doubled."

(Note that 2⁴⁵ is more than 35 trillion.)

Of course there are many more traditional uses of statistics in business, economics, and finance. Financial analysis is built around the idea that a careful look at the data can distinguish assets which might outperform the market, from those that might underperform.

How?

Hopefully you're convinced that you want to learn stats. How can you best do it? Previous students have provided advice on a survey: "Don't take the class lightly. Make sure you take your time and make sure that you understand everything," "Study every day," "do every homework," "form a study group."

Bottom line: it takes work. If you work hard, you'll succeed. If you don't, you'll fail. It's your choice.

The class syllabus explains how we'll get there. Below I provide some specifics on the skills that you will be expected to have by the end of the term.

Want to learn more, about how to do good and avoid bad?

If you begin a love affair with Statistics and want to read more, here are some suggestions of books:

- Leonard Mlodinow, Drunkard's Walk
- Edward R. Tufte The Visual Display of Quantitative Information, Visual Explanations: Images and Quantities, Evidence and Narrative (in library)
- Howard Wainer, Graphic Discovery: A Trout in the Milk and Other Visual Adventures
- David Salsburg, Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century
- James Stock & Mark Watson, Introduction to Econometrics and Peter Kennedy, A Guide to Econometrics
- Jane E. Miller, *The Chicago Guide to Writing about Numbers* (in library)
- John W. Tukey, *Exploratory Data Analysis* (in library)
- Stephen Stigler, Statistics on the Table (in library) and The History of Statistics: The Measurement of Uncertainty before 1900 (in library)
- Dierdre McCloskey, *Economical Writing* and *The Rhetoric of Economics* (in library)

Visualization/Stats/Econometrics blogs

Visualization

- <u>http://www.visualisingdata.com/</u>
- <u>http://infosthetics.com/</u>
- <u>http://www.informationisbeautiful.net/</u>
- <u>http://www.analyticsbridge.com</u>

Stats

- <u>http://junkcharts.typepad.com/numbersruleyourworld/</u>
- <u>http://statisticsforum.wordpress.com/</u>
- <u>http://andrewgelman.com/</u>

Econometrics

• <u>http://davegiles.blogspot.com/</u>

Economics

Of course just about every economics blog is intensely focused on data, trying to make sense of how the economy works, such as

- <u>http://www.calculatedriskblog.com</u>
- <u>http://economistsview.typepad.com/</u>
- <u>http://www.econbrowser.com</u>.