SCI 199Y: "Lies, damned lies, and statistics"

from the First Year Seminar Handbook

The title of the course is a quotation attributed to Disraeli, a British statesman of the nineteenth century. Does it still apply today? This course will consider how statistics and statistical thinking get used (and abused) in a variety of activities, including polling, public health, marketing, advertising, lotteries. Some questions that will be addressed are: Why do newspapers report a "margin of error" for poll results, and what does it mean? How can graphs and charts provide information (or misinformation)? What makes a good graph? How do new cancer drugs get tested, and why doesn't the same protocol work for AIDS? How do studies on mice get extrapolated to humans, and do the results make any sense? What is quality control, and why is it currently so fashionable in North American industry?

Examples of 'current' interest in which statistical thinking is a relevant component

- O.J. Simpson trial
- closing the fisheries in Newfoundland
- polling results for Quebec's vote on sovereignty
- Harris' policy on employment equity
- discovery of a new Alzheimer's gene
- Statistics Canada will include a question on race in the next census (Globe and Mail, Monday Sept. 10, 1995)

What's required in this course – what you'll do

- regular attendance, participation and discussion (15%)
- lots of reading
- copy of Tufte's book *The Visual Display of Quantitative Information* (*VDQI*)
- class presentations (2) of approximately 15 minutes (15%)
- short projects regularly (approximately 6 in total) (30%)
- final essay (40%)

What's offered in this course – what I'll do

- organize discussion around selected themes
- provide background information
- provide directions to further material for reading
- try to keep discussion lively and topical
- try not to get too technical
- provide a sample solution for the first short project, before it's due
- provide guidelines for the final essay
- hold office hours Tuesday 2-3 and Thursday 3-4 (SS6006)

How will the course be organized

We will consider between five and ten themes for discussion, background, and further investigation. These will be selected from current topics according to interests of the class participants, and may consider any of the topics already mentioned in this outline, and any of the following: statistics in sports (is there a hot streak in basketball, how are tennis players seeded, what would have been the outcome of the 94 World Series, is figure skating judging fair, ...); how to assess risks (should bicycle helmets be required, do power lines cause cancer, ...); probability in everyday life (the Monty Hall problem, picking winning lottery numbers,..); environment vs. heredity (*The Bell Curve*, twin studies, genetics and crime); ... Please feel free to suggest other topics for consideration, and be warned that you will be asked about your preferences for choice of topics.

The first theme which will be treated in depth is the use of pictures (usually graphs) to display numerical information. Because numbers are often difficult to assimilate, especially if there are a lot of them, and because statistics is often considered to be very mysterious, many popular accounts in newspapers, magazines, and so on summarize information with some type of graphic. This can be done well or badly, as we will see, and there is a scholarly field of inquiry into how to do it well.

Sources of information

- daily newspapers: pick your favourite; fairly good discussions of science issues appear in the *New York Times*, *The Times*, and the *Globe and Mail*.
- magazines: *Chance*, *New Yorker*, *Saturday Night*, *Scientific American*, *Science*, *Nature*, *J. Amer. Medical Assoc.*, *New England J. Medicine*, *Lancet*,...
- books:

  Visual Display of Quantitative Information, E. Tufte [required]

  The Bell Curve, R. Herrndorff and C. Murray

  Statistics: Concepts and Controversies, D. Moore

  Statistics: A Guide to the Unknown, J. Tanur et al.

More specific reading lists will be provided for each theme.

Required for next week

- Find a graph in a newspaper article that interests you. Bring a copy of the graph and accompanying article to class.
- Find out where you can read/borrow/buy current issues of the *New York Times*, *The Times*, the *Washington Post* and the *Globe and Mail*
- Find out where you can read/borrow/buy current issues of *Chance*, *New Yorker*, *Saturday Night*, *Scientific American* and *Lancet*
- Read the chapter on "Graphical Excellence" in *VDQI*. (Feel free to start on p.16: pp.13–15 have a few tedious technical bits.) Come to class prepared to note the parts you didn't follow, and the parts you found particularly interesting.