Fredrik deBoer
Office TK **Email TK**
Hours TK **Phone TK**

Senior Capstone Seminar
Data Journalism: Reading and Reporting Responsibly

Course Description

In the last several years, the term "data journalism" has made a splash in the media. With the increasing availability of both large data sets and freeware tools to analyze them, as well as the proliferation of firms like Google and Amazon that perform so-called "Big Data" analytics, journalists and writers have begun to apply statistical analysis in ways previously undertaken only in universities and research labs. Websites such as Ezra Klein's Vox and Nate Silver's 538 brought stats to a broader audience than previously possible. Legacy media such as *The New York Times* launched their own data-based offshoots to compete in this growing space. Advanced statistical analysis has become commonplace in journalism concerning education, sports, economics, and more. Readers have found themselves, suddenly, buffeted with a new kind of news and commentary.

But with new kinds of information comes new risks and new challenges. Statistics are both famously powerful and, potentially, notoriously deceptive. The old saying "lies, damn lies, and statistics," frequently misattributed to Mark Twain, demonstrates the pervasive sense that statistics are used as often to misinform as to inform. If a new generation of reporters and writers is eager to share statistical arguments with us, a new generation of readers must develop an appropriate skepticism, without descending into nihilism. How can we as consumers of news and commentary wade through this thicket of statistical data that's presented to us, separating informative data from misleading data? Responsible claims from misleading claims? In this course, we will confront exactly these questions. We'll read about research methods, journalistic ethics, and philosophy of knowledge, then apply these concepts to real-world publications, including newspaper articles, blog posts and articles from data journalism publications, and peer-reviewed academic journals. We'll develop and practice techniques for critically reading and summarizing empirical data and claims. And we'll strive to strike a balance between being overly credulous and overly skeptical towards empirical data and arguments.

This writing-intensive course is not a statistics course, nor a research methods course. Rather, it is a course in being both an effective consumer of empirical data of various forms, and an effective communicator of empirical data. From education to cancer, from baseball to obesity, we'll explore the nexus of data, argument, and communication, as both readers and writers. And we'll place those issues in the socio-cultural context of news media, examining how data journalism has changed the economics of digital media and whether these websites have a gender and racial diversity problem.

Required Texts

To rent or purchase:

Leavitt, F. (2001). *Evaluating Scientific Research: Separating Fact from Fiction*. Upper Saddle River, NJ: Pearson Education. ISBN: 0-13-012845-7

To be distributed (non-exhaustive):

Anderson, C. (2008). The end of theory: the data deluge makes the scientific method obsolete. *Wired*.

Charney, D. (1996). Empiricism is not a four-letter word. *College Composition and Communication*, 567-593.

deBoer, F. (2011). Notes on a revolution in stats. *Wunderkammer Magazine*.

Harford, T. (2014). Big Data: are we making a big mistake? *Financial Times Magazine*.

Ioffe, J. (2013). Ezra Klein: The wise boy. *The New Republic*.

Katz, Y. (2012). Noam Chomsky on where artificial intelligence went wrong. *The Atlantic*.

Marcus, G. (2013). Steamrolled by Big Data. *The New Yorker*.

Mukherjee, S. (2010). A thief in the night. *The Emperor of All Maladies: A Biography of Cancer* (250 – 257). New York: Scribner.

Norvig, P. On Chomsky and the two cultures of statistical learning. *Norvig.Com*.

Standen, A. (2015). Big Data not a cure-all in medicine. *National Public Radio*.

Sternbergh, A. (2008). The spreadsheet psychic. *New York Magazine*.

We will also read contemporary pieces of data journalism, to be selected by myself and class members over the course of the semester.

Assignments

Synopses (8 x 5% = 40%): Students will regularly develop synopses of newspaper, magazine, website, and academic journal articles that involve claims based on data. These synopses will include defining the kind of data being used, its source, how it connects with the author's claim(s), and critical evaluations of the credibility of the data and how responsibly the author is using it. These written responses will typically be between one and a half to two pages.

Issue Papers (2 x 15% = 30%): Twice in the semester, students will write a four or five page paper in which they examine a particular methodological technique, controversy, or topic. These papers will define the area of interest, explain why it is important, and summarize various attitudes about that area of interest. Examples might include competing statistics in sports like baseball and basketball, such as Wins Above Replacement or Value Over Replacement Player; the relationship between socioeconomic class and educational outcomes; whether there is a shortage of STEM graduates in American colleges and universities; the accuracy of Google Translate; and many more. These issue papers cannot possibly be exhaustive, given their length. Instead, they are intended to provide an introduction to the given issue and enough information for a reader to explore it further.

Final Paper & Presentation (20%): At the end of the semester, students will expand one of their two issue papers into a 8-10 page paper and a 8-10 minute presentation. In contrast with the Issue Papers, these Final Papers are expected to include the writer's personal evaluation of the issue at hand. Final Papers should weigh in on controversies, noting if one side is more convincing or the other, and evaluate particular metrics, noting whether competing metrics are more effective. Presentations should be designed for a general audience and include an audio-visual component.

Class Participation (10%)