Course Description:
This course covers many of the common empirical tools used for research in Technology, Policy, and Innovation. Topics include: descriptive statistics, clustering, discrimination analysis, estimation, hypothesis testing, and regression analysis. To learn these topics, students will use modern statistical software programs to analyze data sets with socio-technological applications. After this course, students will have the tools to conduct robust data analyses and present the work in written and visually appealing formats. This course assumes that students have basic knowledge of statistics or data analysis.

Learning Objectives: Data is all around us. How do you sort through the vast amount of information to make valid conclusions and inferences? This is an introductory data analysis course that will teach you things like how to describe your data, types of data, linear regression, and principal component analysis. You will get practical data analysis experience using a variety of data sets ranging from stock prices to health statistics. This course uses R, an open-source statistical software that is increasingly becoming the most used data analysis software. At the end of the course, you will be equipped fully equipped to analyze a variety of data using the most modern software tools.

Class Readings:
- *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* by Cathy O’Neil

HELPFUL BOOKS:
- An Introduction to Applied Multivariate Analysis with R by Brian Everitt and Torsten Hothorn

ASSIGNMENTS:
**Class participation and attendance:** 10%
**Homework assignments:** 30% (two lowest homework assignments are dropped)
**Data Analysis paper:** 20% (PhD students)
**Exam 1:** 20% (Take home)
**Exam 2:** 20% (TBD)

**GRADES**
A: 94-100
A-: 90-93
B+: 87-89
B: 83-86
B-: 80-82
C+: 77-79
C: 73-76
C-: 70-72
D: 60-69
F: Less than 60

**Semester Schedule:**

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Preparation for Class</th>
<th>Homework</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29-Aug</td>
<td>Introduction to the Course</td>
<td>JW 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Installing R</td>
<td>Short Intro to R</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>How do we get a quick idea of what is in the data?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5-Sep</td>
<td>Review Linear Algebra</td>
<td>JW 2</td>
<td>Homework #1 due.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Read sections 2.5 and 2.6.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Read WMD: Ch. 1-3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12-Sep</td>
<td>How do you decrease the number of variables to analyze?</td>
<td>JW 8</td>
<td>Homework #2 due.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Read 8.1 to 8.5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>19-Sep</td>
<td>Distance measures and trees</td>
<td></td>
<td>Homework #3 due.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Read 12.1 to 12.4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>26-Sep</td>
<td>How do you test for normality and prepare the data for analysis?</td>
<td>JW 4</td>
<td>Homework #4 Due</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Read 4.1-4.8 (pay close attention to 4.4-4.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Don't worry about proofs.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Understand main concepts</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3-Oct</td>
<td>Review/Midterm</td>
<td></td>
<td>Homework #5 Due</td>
</tr>
<tr>
<td>7</td>
<td>10-Oct</td>
<td>What is the difference between simultaneous confidence interval and Bonferonni interval?</td>
<td>JW 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.1-5.2, 5.4-5.5</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>17-Oct</td>
<td>How do you do an Anova? What's the difference between big a small samples and how do you approach them differently</td>
<td>JW 6</td>
<td>Homework # 6 due.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Read 6.1 - 6.7, 6.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Read WMD Ch. 6-7</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Topic</td>
<td>Sections</td>
<td>Note</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>24-Oct</td>
<td>Ordinary least squares regression</td>
<td>JW 7 Project outline due Homework #7 due.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>31-Oct</td>
<td>Ordinary Least Squares Regression II</td>
<td>7.1-7.4, 7.5-7.6 Homework #8 due.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>7-Nov</td>
<td>How do you separate data into 2 groups?</td>
<td>JW 11 Memo outline due Homework #9 due. Read sections 12.1 to 11.8</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>14-Nov</td>
<td>Logit</td>
<td>Homework #10 due.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>21-Nov</td>
<td>TBD</td>
<td>Read WMD: Ch. 8-10 Homework #11 Due</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28-Nov</td>
<td>Thanksgiving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5-Dec</td>
<td>Review</td>
<td>Final Paper due December 12 Homework #12 Due</td>
<td></td>
</tr>
</tbody>
</table>

**Computers:** Please bring your laptop to class in order to do the class examples during class. If you do not have a laptop, you may share with another classmate.

**Software:** We will use the software program R for this course. R is an open source data analysis software that is growing in popularity. You can download it from [www.r-project.org](http://www.r-project.org).

**Late assignments:**
Unless otherwise noted, assignments are due BEFORE class on the day that they are due. If the assignment is turned in late, you automatically lose 3 points on the assignment and you continue to lose another 3 points each day. After 2 weeks, you will receive a 0 on that assignment.

**Attendance/Late Policy:**
Attendance to this class is mandatory. Having more than 1 unexcused absence will impact your grade. If you miss a class, you are still expected to do all the readings and assignments for that week. Be on time to class. If you are often late to class, you will lose class participation points.

**Electronics Policy:**
Silence/turn off your cell phones during the class. If you have an emergency where you need to keep you cell phone on, tell the professor before class. Please NO TEXTING during class. If you use a computer to take notes, please do not surf the web. It is distracting to the other students and the professor.
Student Accessibility Support Center Statement
If you have a physical, psychological, medical or learning disability that may impact your course work, please contact Student Accessibility Support Center, ECC (Educational Communications Center) Building, Room 128, (631)632-6748. They will determine with you what accommodations, if any, are necessary and appropriate. All information and documentation is confidential.

Students who require assistance during emergency evacuation are encouraged to discuss their needs with their professors and Student Accessibility Support Center. For procedures and information go to the following website: http://www.stonybrook.edu/ehs/fire/disabilities.

Academy Integrity Statement
Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty is required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at http://www.stonybrook.edu/commcms/academic_integrity/index.html

Critical Incident Management
Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of University Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures. Further information about most academic matters can be found in the Undergraduate Bulletin, the Undergraduate Class Schedule, and the Faculty-Employee Handbook.

Extra Statistics resources:

- Create data visualization: http://www.creativebloq.com/design-tools/data-visualization-712402
- DASL (The Data and Story Library): http://lib.stat.cmu.edu/DASL
- JASA (Journal of the American Statistical Association) Data Archive http://lib.stat.cmu.edu/jasadata/
- Statlib-Datasets Archive http://lib.stat.cmu.edu/datasets/
- University of California, Los Angeles Case Studies http://www.stat.ucla.edu/cases/
- U.S. Census Bureau
http://www.census.gov

Stats in the news, from George Mason University:
http://www.stats.org/

Online statistics textbooks and software:
- Computing for Data Analysis
  https://www.coursera.org/course/compdata
- Data Analysis
  https://www.coursera.org/course/dataanalysis
- Rice virtual lab in statistics
  http://onlinestatbook.com/rvls.html
- SISA simple interactive statistical analysis
  http://www.quantitativeskills.com/sisa/

Online resources for R:
- The main R project site:
  www.r-project.org
- An R online textbook
  Kickstarting R: http://cran.r-project.org/doc/contrib/Lemon-kickstart/
- Website for the Sarkar book:
  http://lmdvr.r-forge.r-project.org/figures/figures.html
- Quick R website (many helpful “how to” pages)
  http://www.statmethods.net/
- A U. of Wisconsin Stats professor’s site (B. Yandell)
  http://www.stat.wisc.edu/~yandell/software/
- Book on Regression and Anova on the R site:
  http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf