Course Title: APSTA-GE 2331: Data Science for Social Impact

Number of Credits: 3

Meeting Pattern: twice a week, 1.25 hours for each class, Fall offering

Course Instructor: Jennifer Hill

Course Description:
Course focuses on the competencies required and the issues that arise Course focuses on how analysts use data and quantitative evidence to impact policy and practice. Students will learn how to gather and analyze data to address questions about program efficacy and efficient targeting of resources. Topics will include how to choose organizational partners, implement change, build trust with organizations and civic agencies, satisfy the needs of stakeholders and manage legal, ethical, and logistical constraints. Students will discuss real case studies and appropriate ways to address them.

Course Prerequisites:
- REQUIRED: Two semesters of Statistics or Data Science (for instance APSTA-GE 2003/2004 or DS-GA 1001/1002 or equivalent as approved by the instructor).

Learning Objectives:
Upon completion of the course, students will be able to
1) integrate organizational goals and available data into an actionable research plan
2) identify and articulate basic legal, ethical, logistical and financial constraints in research involving human subjects and human-generated data
3) critically assess and develop applicable frameworks for an ethically sound balance between privacy protection and transparency/reproducibility
4) identify the basics of best practice in reproducible research
5) articulate how organizations may have to serve the needs of various stakeholders and outline potential pathways to negotiate potentially competing desiderata
6) identify and demonstrate how to address complications associated with combining datasets and satisfying human subjects requirements

Course Format:
2.5 hours of class time per week divided between two class sessions. The course will primarily follow a lecture format with time allowed for small group discussion. Students will be exposed to these issues and potential solutions through case studies presented by guest lecturers of the following types: researchers who have directly encountered such challenges in the field, practitioners working to promote social good at organizations that hire or partner with data scientists, experts on particular topics (e.g. IRB compliance or reproducibility in research).
Students will be expected to be active, respectful, and intellectually additive classroom participants.

**Course Requirements**
The grade for this course will be determined as follows: class participation (15% of grade), written/computational assignments (60%), and one group final project (25%).

Students will be required to participate in class activities (small or large group) by asking questions or making comments a minimum of 20 times during the course of the semester. Participation will be documented by teaching assistants. This is a requirement in part because communicating the ideas learned in the course is an important learning goal.

Assignments (each spaced a few weeks apart) will be the following: 1) practice dissemination findings in the form of both an Op-Ed and a report to program administrators for the same program, 2) creation of replicable computer code for a program evaluation with messy data and multiple datasets that need to be combined, 3) interviewing someone who works at the intersection of Data Science and Social Impact and writing a blog post about research-practice partnerships, 4) completing CITI human subjects certification. In the final group project the group will be presented with a hypothetical situation defined by the available data, resources, organizational goals and constraints and the group will be required to devise a research and dissemination strategy (these case studies will be created in consultation with the speakers recruited each term and thus will vary each year).

**Required Readings and/or Text**
- O'Neil, Cathy and Schutt, Rachel (2003) *Doing Data Science: Straight Talk from the Frontline*, O'Reilly Media. (Chapters 1, 4, 9, 11, 14, 16) [DDS]
- Foster, I. et al. (2016) *Big Data and Social Science*, CRC Press [BDSS]

The following readings may be supplemented by materials suggested by guest speakers.

**Course Outline** (list of topics by week)
1. **Defining Data Science and Social Impact.**
   - DSS, Chapter 1

3. **Opportunities and challenges in research-practice partnerships**
   [plus associated case studies from WT Grant site]

4. **(More) Opportunities and challenges in research-practice partnerships**

5. **Incorporating qualitative and quantitative data**

6. **Mechanics and politics of heterogeneous data integration**

7. **Best practice in transparent and reproducible programming**

8. **Understanding and balancing the needs of different stakeholders and target audiences**

9. **Disseminating results to multiple stakeholders; handling null or negative results**

10. **Translating results into practice**

11. Careful use of data and data-driven decision making: Messy data, significance, and confidentiality
BDSS, Chapter 11

12. Legal framework for data collection, analysis, and sharing (IRB, FERPA, HIPAA)
NYU human subjects tutorial: [http://www.nyu.edu/ucaihs/tutorial/](http://www.nyu.edu/ucaihs/tutorial/)

13. Scaling social impact

14. Wrap-up and presentations

Academic Integrity:
All students are responsible for understanding and complying with the New York University Steinhardt School Statement on Academic Integrity. A copy of this statement is available at: http://steinhardt.nyu.edu/policies/academic_integrity.

Students with Disabilities:
Students with physical or learning disabilities are required to register with the Moses Center for Students with Disabilities, 726 Broadway, 2nd Floor, (212-998-4980 and online at http://www.nyu.edu/csd) and are required to present a letter from the Center to the instructor at the start of the semester in order to be considered for appropriate accommodation.