SOSC 3700A **Quantitative Social Analysis Practicum**

Fall 2020

Wednesday, 1:30-2:50pm

Zoom

Instructor: Dr. WANG, Hongbo (hbwang@ust.hk)

Office: Academic Building, Room 2372 (Ext. 7804)

Office Hours: Thursday, 2-3:00pm

TA: Ren, Yunzhu Bamboo (yrenae@connect.ust.hk)

Office: Academic Building, Room 3001

Office Hours: By appointment

Description and Objectives

This course is intended to help students gain practical, real-world experience with the skills that they have learned in their QSA methods courses. The goal is to give students practice independently choosing a topic, identifying a dataset, selecting a method, carrying out analysis, and reporting results outside of a methods class where the final paper or project is highly structured. At the same time, it is intended to help students to develop ideas for their capstone project.

Students will conduct a small project that involves quantitative analysis of social data. Under the supervision of the instructor, students will choose a topic, locate suitable data, and design and conduct an analysis.

The work done for the project should be above and beyond any work done for the QSA topical elective. Students may not submit for 3700 credit any work that is being submitted or has been submitted for credit in another class.

They should apply basic techniques for managing and analyzing datasets that they have learned in their methodological courses, acquiring the experience working independently that they will need for their capstone project. Students will probably find it easiest to work with a well-documented publicly available dataset downloaded from a site like ICPSR or IPUMS but are certainly welcome to construct their own dataset and analyze it. Students are encouraged to use the opportunity afforded by 3700 to explore different sources of data that they might use for their Capstone Project and conduct analysis with it.

Students may work in groups. For a group project to be approved by the instructor, it must be more ambitious than an individual project, with expectations set according to the size of the

group. Roles of group member will need to be specified in the proposal and contributions to the final project will need to be documented in the final report.

The class will meet on an occasional basis during the term. A detailed schedule can be found later in this syllabus. At these meetings, students will be required to present progress on their projects and participate in discussions of other students' projects.

Projects

The following types of projects are consistent with the goals of the course:

1) Quantitative analysis relevant to a question covered in a SOSC topical class that the student has taken or is currently enrolled in. For example, a student who is taking or has taken Gender and Society might follow up on that class by locating survey or census datasets and examining gender differences in some outcome of interest. The final report should present the analysis as tables or figures and interpret them in light of what they learned in the class.
2) Partial replication of an analysis from a published paper, possibly one that you have read in a SOSC topical class. This should be possible if they paper relied on publicly available data, or the authors made their data available. How much of the analysis you can replicate would depend on the methods used in the published paper? For a paper that used very advanced methods that you have not yet learned, it may be sufficient to do a similar analysis with the simpler methods you have used,
3) Exploration of a variety of datasets to identify one that could be used in a capstone project on a topic of interest. This can be preparatory to SOSC 4100, providing an opportunity to narrow down to one or two datasets that could be used for the capstone project, allowing for SOSC 4100 to focus on literature review or preliminary analysis in preparation for SOSC 4110.
4) Quantitative analysis of existing social data in connection with a topic of personal interest.
5) Creation and analysis of a small dataset relevant to a topic in society, politics or economics, for example by coding of an existing set of materials. This could be useful experience in the most hands-on and sometimes time-consuming aspect of analysis, which is the preparation of a dataset. Please consult with me before choosing a topic involving data collection in case there are human subject issues to be considered.

Your project should make use of what you have learned in your methods classes, especially 1100 and 2400, but hopefully other methods classes you have taken. At the very least, it should demonstrate that you can load a dataset, transform variables via categorization and create new ones, filter the data to work on subsets according to the needs of the analysis, carry out different kinds of analysis including univariate tabulations, cross-tabulations, histograms, and scatterplots. It should not repeat anything you have done in a previous 3700 offering or in one of your methods classes.

Expectations for group projects will be scaled up according to the number of students in the group.

Prerequisites

SOSC 2400

This is intended primarily for QSA students in their third year.

Written Submissions

In addition to attending all five online meetings (See "Scheduled Meetings" below), you will need to complete the following over the course of the term.

(1) Work plan (Due **9/30**)

This should be 3-4 paragraphs. You will submit a description of your topic that also specifies the dataset(s) and methods (s) that you will use. You should also explain how you chose your topic and what it is interesting to you. If you are doing a group project, the names of all group members and division of work should be included.

(2) Bi-weekly Reports on Progress: Weeks 4, 6, 8, 10, 12

Please keep a diary of your work on the project which describes all related activities, including exploration of literature and data, downloading and processing of data including data that you decided not to use, any calculations you are doing, including ones that didn't work out. This should document that you have spent the expected amount of time on the project, including on data and calculations that are not reflected in the final report.

Each individual or group should submit a brief report on progress via Canvas *every two weeks starting in Week 4*. We will create assignments on Canvas accordingly.

(3) Midterm Report (Presentation): Week 7, **10/21**

This should summarize your progress midway through the term. You should describe any calculations you have carried out. If you have run into any obstacles, describe them, and explain how you overcame them. You may list any questions you have that you would like to raise in discussion with the instructor.

(4) Final Report (Presentation): Week 13, **12/2**

Your final report should summarize the topic and your findings, but it is not expected to be a fully-developed research paper. More important is that it demonstrates that you have mastered the preparation and analysis of data and the interpretation and presentation of results.

Your final report should include a series of visually appealing and appropriately titled and labeled tables and figures, with appropriately chosen axis ranges, symbols, and color schemes. They should be suitable for use in a business or academic presentation.

Your final report should include all the detail on how you prepare your dataset and produce your tables and figures, among other things.

Grading

P/F 1 credit.

To earn a P, students must attend class meetings, submit all materials required by the instructor, make a project presentation, and submit a final written report. For group projects, the report must describe the contributions of each member. The presentation and report must include a description of the steps taken as part of the analysis, including the statistical package used and include the code or scripts used to transform the data.

The project will be evaluated regarding merits in design, analysis and interpretation. Feedback will be given regarding prospects for development into a Capstone Project.

Scheduled Meetings (Subject to adjustment)

Calendar Week	Topic	Presentation	Discussion
Week 1 (**9/9**)	Introduction		
Week 2 (**9/16**)	Choice of Topic and Study Design		√
Week 3 (**9/30**)	Locate Data		√
Week 7 (**10/21**)	Midterm Report	√	√
Week 13 (**12/2**)	Project Presentation	√	√

Sources of Data

Many datasets are readily available for you to download. Major examples include:

Interuniversity Consortium for Political and Social Research
https://www.icpsr.umich.edu/web/pages/

ICPSR has a vast searchable collection of surveys and other datasets on a wide range of topics including political, social, and economics. If you would like to try analyzing survey data, you may visit ICPSR to do a keyword search for surveys that cover topics you are interested in.

IPUMS (Integrated Public Use Microsamples)
https://ipums.org/

This is a vast online repository of Census data from various countries and time periods. Census data from North America is especially complete, with United States Census materials available back to 1850.

China Multigenerational Panel Databases (CMGPD)
https://www.icpsr.umich.edu/web/ICPSR/series/265

These are datasets that my collaborators created and publicly released. They cover rural populations in northeast China from the mid-18th century to the beginning of the 20th century. They include life histories and information about household structure. The data are extensively documented in English and in Chinese and have already been subject to extensive processing to make them straightforward to use. Doing a case study of a community would be good practice working with a large, rich dataset.

China Government Employee Dataset-Qing (CGED-Q)
https://dataspace.ust.hk/dataset.xhtml?persistentId=doi:10.14711/dataset/E9GKRS

This is a subset of a dataset that my collaborators and I are building that records information about civil servants in the Qing dynasty. They are recorded every three months. They can be used for comparisons of groups of officials, for example according to their province of origin or their type of qualification. Records of an official can be linked across time straightforwardly by sorting on surname, given name, province of origin and county. This data is more raw than the CMGPD data and would provide useful experience in working with the 'messy' data that you are likely to encounter in the private or public sector.