SOSC 4100
Research Pro-seminar

Fall 2020

Tuesdays
T1: 10:30AM-11:20AM
T2: 1:30PM-2:20PM

Check Canvas for Zoom meeting ID

15 September 2020 version

Instructor
Cameron Campbell
Appointments: https://www.doodle.com/mm/cameroncampbell/studentappointments

This class prepares students for their QSA Capstone Project. By the end of the term, students will have a detailed plan for their Capstone Project that they will execute in SOSC 4110. We will meet weekly, during which time students will discuss their ideas. We may also have sessions devoted to topics relevant for fourth year students, including PG studies, guest presentations from faculty or RPG students, and possibly other invited guests.

Required Elements

The key elements of the class include 1) attendance and participation, 2) assignments, 3) submission of a capstone project plan at the end of the term, and 4) a presentation of the plan to the class. Students must complete all four elements to earn a passing grade.

Attendance and Participation

Students must attend all class sessions and participate in discussion. For Zoom sessions, cameras must be turned on. Students may miss up to two sessions without an explanation or documentation. Documented absences use up the two allowed absences first, so that for example, if a student has two documented absences and one undocumented absence, the documented absence is beyond the allowance and may result in attendance and participation being considered insufficient, leading to an F.

Assignments

Over the course of the term, I will ask students to post responses to prompt in the Discussion at the Canvas site in advance of a class meeting, as a basis for discussion. There will not be too many of these, but they will need to be completed to earn a passing grade.

Capstone Project Plan

At the end of the term, each student will submit a Capstone Project Plan. This should be realistic and attainable. This will include all the key elements of a proposal including a specification of the aims and significance of the research, a brief discussion of relevant literature and a description of the data and methods.

In most cases the data will be something available from a public repository like the Inter-University Consortium for Political and Social Research (https://www.icpsr.umich.edu/web/pages/), but it could be data collected by the student, or provided by a faculty member. For public data from ICPSR or another repository, or data provided by a faculty member, the plan will need to confirm that the contents of the data are appropriate to the task at hand.

The plan should be around 1750 words of text. References should be included but are not included in the word count requirement. Pages should be numbered. All work must be your own. If when describing a method or a dataset you believe it is necessary to use text from elsewhere, clearly identify it as a quotation and provide proper attribution.

You are welcome to exceed the recommended word length if you want to get a head start on your Capstone Project by writing more literature review, presenting descriptive results from your dataset, or otherwise writing material to be re-used in your Capstone Project.

The general organization may resemble a proposal you have written for SOSC 2140 or another class. The difference is that it must be realistic, and something that you are confident you can complete in the available time in SOSC 4110. It must also involve quantitative analysis using methods you have learned in SOSC 1100, SOSC 2400, and other methods classes. Ideally the dataset will be large and complex enough that your analysis will successfully showcase your skills for potential employers or postgraduate programs.

The plan must include the following elements:

Aims – 250 words
A very brief statement of the goal of the Capstone Project, roughly 250 words. For someone who doesn't read the rest of your proposal, this should provide a basic idea of your plan.

Significance – 250 words
Explain why the analysis is important. Convince the reader that the work is worth doing.

Background – 500 words
Anything about the topic or the setting you think it is important for the reader to know. May include discussion of previous relevant studies.

Relationships of interest - 250 words
What specific relationships are you interested in? If you already have hypotheses, you may state them here.

Data – 250 words
Describe the data you will use, including basic information about the target population, the sample size, when it was created and so forth. List the names of the variables that you will use in your study.

Methods -250 words
Describe the methods you will use. In addition to basic bivariate analysis of relationships between pairs of variables via tabulations or scatterplots, you should plan on using some more advanced methods that will help showcase the skills you have learned for employers or postgraduate programs. These may include multiple regression,

Schedule (may be adjusted)

Week	Topic	Assignment
1 - 8/9/2020	Introduction	
2 - 15/9/2020	Discussion of ideas	Respond to prompt
3 - 22/9/2020	Discussion of ideas	
4 - 29/9/2020	Sources of data	
5 - 6/10/2020	Faculty research opportunity Franziska Barbara-Keller	Everyone please attend T2 at 1:30pm
6 - 13/10/2020	Faculty research opportunity Cameron Campbell	Paper on CGED-Q https://doi.org/10.1017/jch.2020.15
7 - 20/10/2020	Research ethics	
8 - 27/10/2020	Progress reports	
9 - 3/11/2020	Progress reports	
10 - 10/11/2020		
11- 17/11/2020	Presentations of Plans	PPT and draft text for plan
12 - 24/11/2020	Presentations of Plans	
13- 1/12/2020	Presentations of Plans	
11/12/2020		*Final version of plan due*

Publicly accessible data

The following are resources that you should explore this semester to help identify the dataset(s) that you may use for your Capstone Project in 4110.

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

This is a good starting point if you are looking for datasets that include documentation that might be the basis of a Capstone Project. ICPSR has a vast keyword 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.

Harvard Dataverse
https://dataverse.harvard.edu/

Vast collection of datasets deposited by researchers in a variety of areas, and keyword searchable.

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

This is a vast online repository of Census and related data from various countries and time periods. Census data from North America is especially complete. Custom extracts may be downloaded according to your needs.

Gateway to Global Aging Data
https://g2aging.org/

A platform for population-representative longitudinal survey data relevant to aging from countries and regions around the world.

UK Data Archive
https://www.data-archive.ac.uk/

Large collection of social science and population datasets for the UK.

Consortium of European Social Science Data Archives (CESSDA)
https://www.cessda.eu/

Another repository. Searchable with keywords.

Surveys

China

- China Family Panel Survey (CFPS)
- China Health and Retirement Longitudinal Survey (CHARLS)
- China Health and Nutrition Survey (CHNS)
- China Health and Family Life Survey (CHFLS)
- Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan
- Taiwan Educational Panel Survey (TEPS)
- Taiwan Longitudinal Study of Aging (TLSA)
- Taiwan Panel Study of Family Dynamics (PSFD)
- Taiwan Youth Project (TYP)

United States

- Add Health The National Longitudinal Study of Adolescent to Adult Health
- General Social Survey
- Health and Retirement Survey
- National Longitudinal Survey of Youth 1979
- National Longitudinal Survey of Youth 1997

Datasets

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.