

SOC 4930/5050: Final Project Instructions

Christopher Prener, Ph.D.

December 7th, 2017

Overview

All students will select a set of variables from the 2012 General Social Survey and perform an original data analysis culminating in a series of linear regression models. Each student will be responsible for selecting a research question, establishing a hypothesis, conducting an analysis to test that hypothesis, and presenting the results. This process and project mirrors the steps taken to author a quantitative conference presentation (for all students) and a quantitative journal article (for students enrolled in SOC 5050).

Data

Data for the final project is available in the [Final-Project-Data](#) repository that is part of our [SOC 4930 & SOC 5050 organization](#). The data have been obtained from the [Inter-university Consortium for Political and Social Research \(ICPSR\)](#). Their license precludes making the data publicly available, and so these data are only available to current students.

Memo



Due by September 25th



Submit as a GitHub issue

All students should submit a memo as a GitHub issue that identifies a topic and key variables you are interested in exploring. Make sure these variables are present in the 2012 edition of the GSS - not all variables are available for every year. Ideally, you should also identify a continuous dependent variable. All memos should be accompanied

by two or three key citations that describe the social science research on your topic.

Annotated Bibliography



SOC 5050 students only



Due by October 16th



Submit as a Markdown (.md) file

An annotated bibliography with twenty to twenty-five scholarly, peer-reviewed sources should be submitted as part of the preparation of the final project. Peer-reviewed sources include articles from reputable academic journals and books from University presses and other scholarly publishing houses.

Before you begin, you should select either the [American Psychological Association \(APA\)](#) or the [Chicago Manual of Style \(CMS\)](#) style guidelines.¹ If you use Chicago, please use parenthetical citations in the Author Date in-text citation format. **Do not mix citation styles - you must select one and apply it consistently.**

Your annotated bibliography should include a properly formatted bibliography entry for each source as well as a paragraph length description of the source and its relevance to your project.

¹ Public and Social Policy doctoral students are strongly encouraged to use CMS.

Conducting a Literature Search

I recommend following the steps below as part of a comprehensive strategy for conducting a literature search.

1. **Define your research topic and question:** What is it that you are trying to identify sources in the literature for?
2. **Identify appropriate search terms:** What are the keywords that might help you find relevant articles?

3. **Identify the “state of the art”:** Start by searching Annual Review journals (or similar, such as Sociology Compass) for recent articles that summarize the state of the literature in your area. You may not find an exact match, but you may find reviews of related topics. Use these articles to get a sense of what the “state of the art” is for your topic area(s).
4. **Citations from review articles:** Use the review articles you find to identify both seminal publications that define a field of study (may not be recent) as well as more recent empirical studies. Find and download these articles.
5. **Search other social science publications:** Whether or not you find helpful review articles, you will want to also search other key journals. Start with generalist journals and then move to more specific journals focused on particular types of research. Use the same search terms described above. In sociology, for example, the major generalist journals are *American Sociological Review*, *American Journal of Sociology*, *Social Forces*, and *Social Problems*. In one of my sub-fields, medical sociology, *Social Science and Medicine*, *Journal of Health and Social Behavior*, and *Health and Place* are all key journals.
6. **Search related journals:** Are there journals from other disciplines that are related? If so, search those as well using your search terms. In my research, the *American Journal of Public Health* and the *Annals of the American Association of Geographers* are both important journals from related disciplines.
7. **Search Sociological Abstracts:** If you are not finding many citations, you may want to search the Sociological Abstracts database, which catalogs many more sociological journals.
8. **Iterative Processing of Articles:** As you read through the articles you find and download in steps 3 through 7, seek out the studies that they cite that seem related. Download them, read them, and identify more related studies that they cite, then start the process over. Also use this process to identify books.

As you find sources, I recommend keeping track of them in a spreadsheet. This makes it easier to figure out if you have already found and downloaded an article that you may come across multiple times. Once you identify an article you might think is relevant, read the abstract then read the introduction and the conclusion. Does the article still seem related? If not, put the article to the side and mark it as not likely relevant in your spreadsheet. Do not spend a lot of time reading or taking notes on articles that are not likely to be useful to you.

For articles that are relevant, focus on trying to connect articles to each other - who is citing who? Who is working on similar topics or making similar arguments? Who is contradicting or refuting other scholars? These connections will be useful to develop as you think about *synthesizing* the literature for your paper.

Reviews that contain the requested number of citations (20 to 25) at a minimum will be eligible for a check-plus. To earn the check-plus, each of these citations should be from peer reviewed sources. These sources should largely be drawn from sociology journals or journals that have a clear relationship to your topic. The reviews you write should be rich but succinct summaries of each source and note connections. fiNot having enough sources, having sources that are not peer reviewed or not drawn from sociology sources, or having reviews that are not clear, concise, and informative are all reasons for having a check or check-minus grade.

Analysis



Complete using an RNotebook (.Rmd) file

Your analysis should be completed in a well-organized RNotebook (.Rmd) file that uses literate programming techniques, RMarkdown text syntax, and tidyverse packages (including piped code where possible). It should include the following elements:

1. A well-constructed hypothesis: what are you trying to test with your final regression models?
2. Descriptive statistics for all variables, as appropriate for their level of measurement.
3. A histogram for your dependent variable.
4. A full normality assessment of your dependent variable, including q-q plots.
5. Independent t-tests of your main independent variable of interest and your dependent variable.
6. Correlation analysis of all of your variables.
7. A scatter plot showing your main independent variable of interest and your dependent variable.

8. An initial set of set of three OLS regression models, each specified both as a \LaTeX equation and the appropriate R functions.
9. A full assessment of the OLS regression assumptions on your third and final model.
10. If adjustments are warranted, re-fit your initial three models (make sure to update your equations as necessary) on the adjusted set of variables and/or observations.

Draft Presentation



Due by November 13th



Submit your slides as a .pdf file

Use presentation software (either Microsoft PowerPoint or Apple Keynote) to draft a short, six to eight minute presentation of your findings. Please refer to my super-meta [presentation on presentations](#) for guidance on constructing a good presentation. Your draft will be graded using the “check” system outlined in the syllabus. At the draft stage, presentations should include:

1. A title slide (presentation title, name, date, etc.)
2. An agenda slide with the main topics you want to cover
3. A literature slide (SOC 4930 students should include a few key citations [two to three]); SOC 5050 students should avoid the temptation to summarize each of their sources)
4. Plots, including your histogram, q-q plot, and scatter plot described in the *Analysis* section above.

Draft Handout**Due by November 13th****Submit your handout as a .pdf file**

Use \LaTeX to produce a handout that includes your title slide information (presentation title, name, date, etc.), a abstract that summarizes your research question and hypothesis, a descriptive statistics table, and a few key citations. These should all fit on one page. Your draft will be graded using the “check” system outlined in the syllabus.

Draft Paper**SOC 5050 students only****Due by November 13th****Submit your paper as a .pdf file**

Use \LaTeX to produce a paper that, at a minimum, includes a draft “Introduction”, “Background”, and “Data and Methods” sections. Keep in mind that your final paper should be approximately 5000 words long, and will also contain “Results”, “Discussion”, and “Conclusion” sections. Your initial sections should be written with this final length and content mind. Your draft paper should include a well-formatted bibliography. Your draft will be graded using the “check” system outlined in the syllabus.

Final Presentation



Due on December 18th



Present orally; submit slides as a .pdf file

During our finals week session, you will present their research findings. Each student will have eight minutes to present their work. Use the presentation software of your choice (e.g. PowerPoint or Keynote).² Finally presentations should include:

² Please do not use Prezi.

1. A title slide (presentation title, name, date, etc.)
2. An agenda slide with the main topics you want to cover
3. A literature slide (SOC 4930 students should include a few key citations [two to three]); SOC 5050 students should avoid the temptation to summarize each of their sources)
4. A data and methods slide that describes the source of your data (the 2012 General Social Survey) and the general techniques you used to produce your analyses
5. A descriptive statistics slide with mean and, as appropriate, the standard deviation and range for continuous variables *or* the meaning of values for binary variables.
6. Plots, including your histogram, q-q plot, and scatter plot described in the *Analysis* section above.
7. A summary of your regression findings in a well-formatted table that is appropriate for your medium (do not copy and paste an image of a \LaTeX table).
8. A limitations slide - what cautionary notes do you have about your analyses or your findings?
9. A discussion and conclusion slide that summarizes your findings, places them in the context of the literature you have identified, and offers some concluding remarks.

Final Handout**Due on December 18th****Submit your handout as a .pdf file; bring 14 copies**

Your presentation should be accompanied by a well-formatted \LaTeX handout that includes, on the first page:

1. A title block with relevant information (presentation title, name, date, affiliation).
2. An abstract that summarizes your research question and hypothesis.
3. A descriptive statistics table created using the `stargazer` package for R.
4. A few key citations (2-3) formatted using either APA or Chicago.

Your second page should contain a regression table showing multiple models that build upon each other. It should be created using the `stargazer` package for R. Make sure you add AIC and BIC to the table and remove un-adjusted R^2 .

Final Paper**SOC 5050 students only****Due by November 18th****Submit your paper as a .pdf file**

Use L^AT_EX to produce an empirical paper that is formatted like an article manuscript and contains the following sections:

1. Cover page with a title, word count, your name, your affiliation, and your contact information (professional address, email).³
2. Abstract of no more than 300 words on its own page
3. The main body of the paper, no more than 5,000 words in length, containing:
 - (a) Introduction
 - (b) Background
 - (c) Data and Methods
 - (d) Results
 - (e) Discussion
 - (f) Conclusion
4. Any expository endnotes
5. Tables, one per page
6. Figures, one per page
7. Works cited

³ Use the Department of Sociology and Anthropology, Saint Louis University, 3700 Lindell Blvd., St. Louis MO 63108 if you need a professional address.

See the [L^AT_EX Paper Guide](#) for details on what should be included in each of these sections and technical details for formatting your paper using L^AT_EX.

HANDOUT RUBRIC	Poor (F)	Fair (D)	Satisfactory (C)	Good (B / A-)	Excellent (A)
Handout	<ul style="list-style-type: none"> ▶ No handout submitted 	<ul style="list-style-type: none"> ▶ LaTeX handout style or LaTeX not used ▶ Handout contains less than 4 of the required elements ▶ Handout has major detractors in its design and organization 	<ul style="list-style-type: none"> ▶ LaTeX handout style not used ▶ Handout contains only 4 required elements ▶ Handout has moderate organizational flaws 	<ul style="list-style-type: none"> ▶ LaTeX handout style used ▶ Handout contains all 5 required elements ▶ Handout has only minor flaws but is generally well organized 	<ul style="list-style-type: none"> ▶ LaTeX handout style used ▶ Handout contains all 5 required elements ▶ Handout is exceptionally well designed and organized

PRESENTATION RUBRIC (1 of 2)	Poor (F)	Fair (D)	Satisfactory (C)	Good (B / A-)	Excellent (A)
Slide Design (30%)	<ul style="list-style-type: none"> ▶ Does not meet assignment criteria ▶ There are a significant number of syntax, grammar, word choice and/or spelling errors ▶ No tables or figures included on slides 	<ul style="list-style-type: none"> ▶ No visible consideration given to design ▶ Major detractors, such as small font size, misaligned text, or inappropriate graphics ▶ There are numerous syntax, grammar, word choice and/or spelling errors ▶ Tables and figures are difficult to read 	<ul style="list-style-type: none"> ▶ Moderate consideration given to design, including some use of colors and shapes to accent information ▶ Only minor detractors in design ▶ There are a moderate number of syntax, grammar, word choice and/or spelling errors ▶ Tables and figures are legible but have significant detracting elements 	<ul style="list-style-type: none"> ▶ Clear consideration seen design, including use of color and shape elements to accent information ▶ No significant detracting elements ▶ There are a small number of syntax, grammar, word choice and/or spelling errors ▶ Tables and figures show clear consideration for legibility and design with some minor flaws 	<ul style="list-style-type: none"> ▶ Exceptional design that shows deep thought and attention to detail ▶ No significant detracting elements ▶ Little to no syntax, grammar, word choice and/or spelling errors ▶ Tables and figures are exceptionally well designed

PRESENTATION RUBRIC (2 of 2)	Poor (F)	Fair (D)	Satisfactory (C)	Good (B / A-)	Excellent (A)
Research Details (40%)	<ul style="list-style-type: none"> ▶ Missing 4 or more of the elements listed on the previous page ▶ Research process is poorly described 	<ul style="list-style-type: none"> ▶ Missing 3 of the elements listed on the previous page ▶ Research process is missing major descriptive elements 	<ul style="list-style-type: none"> ▶ Missing 1 or 2 of the elements listed on the previous page ▶ Research process missing some key descriptive elements 	<ul style="list-style-type: none"> ▶ Contains all 9 required elements ▶ Research process is well defined and communicated 	<ul style="list-style-type: none"> ▶ Contains all 9 required elements ▶ Research process and conclusions are exceptionally well communicated
Format (10%)	▶ Presentation does not adhere to the written guidelines			<ul style="list-style-type: none"> ▶ Presentation deviates from the written guidelines in only a minor way 	<ul style="list-style-type: none"> ▶ Presentation follows written guidelines for formatting with no deviations
Presentation (10%)	<ul style="list-style-type: none"> ▶ Not possible to hear or follow presenter ▶ Presentation is too short (≤ 2 minutes) or too long (≥ 12 minutes) 	<ul style="list-style-type: none"> ▶ Presentation is generally hard to follow and understand ▶ Presentation is within 4 minutes of time-limits 	<ul style="list-style-type: none"> ▶ Presenter occasionally hard to understand or follow ▶ Presentation is within 2 minutes of time-limits 	<ul style="list-style-type: none"> ▶ Presenter speaks clearly and generally concisely ▶ Presentation is within time-limits (6 to 7 minutes) 	<ul style="list-style-type: none"> ▶ Presenter speaks exceptionally clearly and concisely ▶ Presentation is within time-limits (6 to 7 minutes)
Citations (10%)	▶ Presentation and handout do not adhere to the written guidelines			<ul style="list-style-type: none"> ▶ Citations are formatted correctly with only a small number of errors 	<ul style="list-style-type: none"> ▶ Citations are formatted correctly with little to no error

SOC 4930 POINT TOTALS	Poor (F / 0%)	Fair (D / 65%)	Satisfactory (C / 75%)	Good (B / 85% or A- / 90%)	Excellent (A / 100%)
Handout	0 points	32.5 points	37.5 points	42.5 or 45 points	50 points
Slide Design	0 points	35.1 points	40.5 points	45.9 or 48.6 points	54 points
Research Details	0 points	46.8 points	54 points	61.2 or 64.8 points	72 points
Format	0 on the assignment			15.3 or 16.2 points	18 points
Presentation	0 points	11.7 points	13.5 points	15.3 or 16.2 points	18 points
Citations	0 on the assignment			15.3 or 16.2 points	18 points

SOC 5050 POINT TOTALS	Poor (F / 0%)	Fair (D / 65%)	Satisfactory (C / 75%)	Good (B / 85% or A- / 90%)	Excellent (A / 100%)
Handout	0 points	13 points	15 points	17 or 18 points	20 points
Slide Design	0 points	17.5 points	20.25 points	23 or 24.3 points	27 points
Research Details	0 points	23.4 points	27 points	30.6 or 32.4 points	36 points
Format	0 on the assignment			7.7 or 8.1 points	9 points
Presentation	0 points	5.9 points	6.75 points	7.7 or 8.1 points	9 points
Citations	0 on the assignment			7.7 or 8.1 points	9 points

RUBRIC (1 of 2)	Poor (F)	Fair (D)	Satisfactory (C)	Good (B or A-)	Excellent (A)
Code (10 points)	<ul style="list-style-type: none"> ▶ RNotebook not submitted, or: <ul style="list-style-type: none"> • RNotebook contains little to no analysis • RNotebook contains no narrative text • Widespread use of non-tidyverse code 	<ul style="list-style-type: none"> ▶ RNotebook does not clearly document process ▶ Limited use of Markdown ▶ Limited narrative text included in notebook ▶ Widespread use of non-tidyverse code 	<ul style="list-style-type: none"> ▶ RNotebook has some degree of organization but is hard to follow ▶ Markdown has some syntax errors ▶ Some narrative text but incomplete or vague ▶ Moderate use of non-tidyverse code 	<ul style="list-style-type: none"> ▶ RNotebook is well organized using headings and Markdown syntax ▶ Generally good implementation of the literate programming paradigm ▶ Minor divergence from tidyverse approach 	<ul style="list-style-type: none"> ▶ RNotebook is exceptionally well organized using headings and Markdown syntax ▶ Exceptional implementation of the literate programming paradigm ▶ Tidyverse principles used exclusively
Content and Development (30 points)	<ul style="list-style-type: none"> ▶ No thesis statement ▶ No cogent or supported analysis ▶ 5 or more relevant aspects of an article are missing 	<ul style="list-style-type: none"> ▶ No thesis statement ▶ Analysis and argument offer limited support, insight and/or logical progression ▶ 3 or 4 aspects of an article as missing 	<ul style="list-style-type: none"> ▶ Unclear thesis statement ▶ Background provides an incomplete assessment of the relevant literatures ▶ 1 or 2 aspects of an article are missing 	<ul style="list-style-type: none"> ▶ Clear thesis statement ▶ Background provides a generally thorough assessment of the relevant literatures ▶ All relevant aspects of an article are present and generally well-executed 	<ul style="list-style-type: none"> ▶ Clear thesis statement ▶ Background provides an exceptionally well-rounded assessment of the relevant literatures ▶ All relevant aspects of an article are present and exceptionally well-executed
Analysis (30 points)	<ul style="list-style-type: none"> ▶ No hypotheses included ▶ Analysis is not related to the selected topic in a meaningful way ▶ No statistical tests included or are not at all relevant 	<ul style="list-style-type: none"> ▶ No hypotheses included ▶ Analysis is only tangentially related to the selected topic ▶ Significant concerns with the relevance of the included statistical tests, their execution, and how their findings are communicated 	<ul style="list-style-type: none"> ▶ Hypotheses are not explicitly rooted in the literature review and/or not clearly stated ▶ Analysis and argument are moderately supported, insightful and cogent ▶ Moderate concerns with statistical tests' relevance, execution, or communication 	<ul style="list-style-type: none"> ▶ Hypotheses are rooted in the literature review and generally well stated ▶ Analysis and argument are generally supported, insightful and cogent ▶ Minor concerns with statistical tests' relevance, execution, or communication 	<ul style="list-style-type: none"> ▶ Hypotheses are rooted in the literature review and clearly stated ▶ Analysis and argument are supported, exceptionally original, insightful and cogent ▶ Chosen statistical tests are appropriate, correctly executed, and exceptionally well-communicated

RUBRIC (2 of 2)	Poor (F)	Fair (D)	Satisfactory (C)	Good (B or A-)	Excellent (A)
Tables and Figures (10 points)	<ul style="list-style-type: none"> ▶ No tables or figures, or exceptionally limited use of them 	<ul style="list-style-type: none"> ▶ ggplot2 and stargazer are not used for most outputs ▶ Tables and figures are have significant issues with labeling and organization ▶ No use of color on plots 	<ul style="list-style-type: none"> ▶ ggplot2 and stargazer are used for most outputs ▶ Tables and figures are have some moderate issues with labeling and organization ▶ Limited use of color on plots 	<ul style="list-style-type: none"> ▶ ggplot2 and stargazer are used exclusively ▶ Tables and figures are generally well organized and clearly labeled ▶ Figures use color to emphasize relevant points 	<ul style="list-style-type: none"> ▶ ggplot2 and stargazer are used exclusively ▶ Tables and figures are exceptionally well organized and clearly labeled ▶ Figures use color to emphasize relevant points
Organization and Cohesion (5 points)	<ul style="list-style-type: none"> ▶ Paper has no obvious organizational strategy 	<ul style="list-style-type: none"> ▶ Paper has significant organizational or flow discrepancies 	<ul style="list-style-type: none"> ▶ Paper has moderate organizational or flow discrepancies 	<ul style="list-style-type: none"> ▶ Paper is generally well organized and easy to follow 	<ul style="list-style-type: none"> ▶ Paper is clearly organized and easy to follow
Format (5 points)	<ul style="list-style-type: none"> ▶ Paper does not adhere to the written guidelines 	<ul style="list-style-type: none"> ▶ Paper does not adhere to the written guidelines 			<ul style="list-style-type: none"> ▶ Paper follows written guidelines for formatting with no deviations
Quality of Writing, Grammar, and Spelling (5 points)	<ul style="list-style-type: none"> ▶ Writing quality is poor and seriously detracts from the message of the paper ▶ There are a significant number of syntax, grammar, word choice and/or spelling errors 	<ul style="list-style-type: none"> ▶ Writing quality is below average and detracts from the message of the paper ▶ There are numerous syntax, grammar, word choice and/or spelling errors 	<ul style="list-style-type: none"> ▶ Writing quality is average but does not detract from the message of the paper ▶ There are a moderate number of syntax, grammar, word choice and/or spelling errors 	<ul style="list-style-type: none"> ▶ Writing quality is generally high ▶ There are a small number of syntax, grammar, word choice and/or spelling errors 	<ul style="list-style-type: none"> ▶ High quality writing ▶ Little to no syntax, grammar, word choice and/or spelling errors
Citations (5 points)	<ul style="list-style-type: none"> ▶ Paper does not adhere to the written guidelines 	<ul style="list-style-type: none"> ▶ Paper does not adhere to the written guidelines 			<ul style="list-style-type: none"> ▶ Citations and bibliography are formatted correctly with little to no error

POINT TOTALS	Poor (F / 0%)	Fair (D / 65%)	Satisfactory (C / 75%)	Good (B / 85% or A- / 90%)	Excellent (A / 100%)
Code	0 points	6.5 points	7.5 points	8.5 or 9 points	10 points
Content and Development	0 points	19.5 points	22.5 points	25.5 or 27 points	30 points
Analysis	0 points	19.5 points	22.5 points	25.5 or 27 points	30 points
Tables and Figures	0 points	6.5 points	7.5 points	8.5 or 9 points	10 points
Organization and Cohesion	0 points	3.25 points	3.75 points	4.25 or 4.5 points	5 points
Format	0 on the assignment			4.25 or 4.5 points	5 points
Quality of Writing, Grammar, and Spelling	0 points	3.25 points	3.75 points	4.25 or 4.5 points	5 points
Citations	0 on the assignment			4.25 or 4.5 points	5 points