

Stat 139: Project Guidelines, Spring 2015

Overview:

The course project is your opportunity to explore a topic of your academic or personal interest. You will work in groups of 2-4 students (graduate students should try to form groups with 2-3 undergraduates). If you are looking for group mates and have a project in mind, we recommend that you post your ideas on course website under “Discussions” (make sure to post under the “Project” discussion board).

The project can take two forms: data analysis or theoretical extension. The project is supposed to be interesting for you: choose a topic you like! A final work product of the project will be a paper (4-6 pages of text), which you will hand in during reading period on **Wednesday, May 6, 1:39pm**.

Requirements and Deadlines:

1. Deadline #1: Email the names of group members and a short one-paragraph description of the project proposal to stat139harvard@gmail.com (cc: all group members) by **Friday, April 17, 1:39pm**. Use subject line “Stat 139: Project Proposal”.
2. Deadline #2: Submit a 1-page (or slightly longer) project update, including outline of the paper, explanation of what you have done so far, and what is left to do, and description of any challenges you have faced (for example, difficulties obtaining data). Send one email per group to stat139harvard@gmail.com by (cc: all group members) **Monday, April 27th**. Use subject line “Stat 139: Project Update”.
3. Deadline #3 (final project documentation): Submit an electronic version of your paper and any R code (with comments) to stat139harvard@gmail.com (one per group and cc: all group members) by **Wednesday, May 6, 1:39pm**. Use subject line “Stat 139: Final Project Submission”. Also, hand one paper copy (to the usual 3rd floor HW submission box) by 1:39pm on that day as well.

General Advice:

- The text of your paper (not including tables, graphics, appendices (R code, output, or mathematical derivations), etc...) should be about 4-6 pages. The length may vary widely from project to project.
- Thoughtful graphics are often more illuminating than tables or pages of R output.
- Your grade does not depend on whether your original hypotheses are correct - you will probably learn more if your hypotheses are incorrect!
- Your research question, motivation, hypotheses, methods, assumptions, results, limitations, and conclusions & further discussions should be presented clearly in your paper, if applicable (you may not need a separate section for each of these topics, and not all of these topics are relevant for all projects). Your poster should also include a short discussion of any challenges you faced. Be sure to cite sources, if applicable, and include a reference list.

Details on Theoretical Extension Project:

State a question that extends the theoretical material covered in this course, and explore it. Exploring the question might involve performing simulations, reading published statistics papers, writing out some calculations of your own, and/or demonstrating the principles on a small data set (such as the data sets you have used on homework problems).

When selecting a topic, ask yourself whether the question is within the scope of this course and project. We encourage you to choose a topic that you might continue to explore after this course; however, be sure that you will be able to make appropriate progress within the project timeframe.

Possible theoretical projects include extended simulations to answer a question that you propose; explanation and application of Bayesian versions of the models discussed in this class (note that you cannot use the same project for this course and another course); a critical review of recent statistics papers on a topic related to this class; or a careful reading of an older, foundational paper that first proposed a method discussed in this course.

Details on Data Analysis Project:

State an applied research question and answer it by analyzing a data set, using the tools emphasized in this course. Focus on assessing assumptions, choosing appropriate tools, and evaluating the validity of your results.

If you are in a field other than statistics, already working on applied research, or considering future applied work, we encourage you to choose a topic that relates to your interests. However, if you use a data set that has been analyzed before by you or others (that you know of), the research question that you address for this project must be new. If the data set is associated with a published paper, for example, you might explore a different model or ask a different question.

When choosing a data set, ask yourselves: Can you address your research question using this data set? Are the statistical tools you'll need within the scope of this course?

This is not a project about conducting experiments or surveys. We strongly encourage you to use data that already exists. If you do generate data yourself, keep in mind the general principles we've discussed.

You may find the following data resources helpful:

Feel free to share other useful sources that you come across by posting them on the course website under "Discussions → Projects".

- A range of data resources for academic community
<http://datalib.edina.ac.uk/catalogue/all>
- World Bank Open Data: free and open access to data about development in countries around the globe <http://data.worldbank.org/>.
- Sports Data: <http://it.stlawu.edu/~rlock/sports.html>

- Government data (from more than 70 agencies) <http://www.fedstats.gov/>
- 100+ Interesting Data Sets for Statistics
<http://rs.io/100-interesting-data-sets-for-statistics/>
- Real estate sales data in the US (it is free, although registration is required to get a full access)
<https://www.redfin.com/>
- Data available though Harvard Library (click “Data” on the left and explore!)
<http://library.harvard.edu/>
- Aid Data - Open Data for International Development
<http://www.aiddata.org/content/index>
- Center for Economic Policy Research - ceprDATA <http://ceprdata.org/>
- Correlates of War <http://www.correlatesofwar.org/>
- European Union - EUROSTAT <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>
- FBI - Crime Statistics <http://www.fbi.gov/stats-services/crimestats>
- Federal Reserve Economic Data (FRED) <http://research.stlouisfed.org/fred2/>
- Gapminder <http://www.gapminder.org/>
- Harvard University Library: Getting Started in Economic Research <http://guides.hcl.harvard.edu/content.php?pid=176637&sid=1487256> (Very comprehensive, many good links)
- IMF Data and Statistics <http://www.imf.org/external/data.htm>
- Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan <http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp>
- IQSS Dataverse Network <http://dvn.iq.harvard.edu/dvn/>
- Journal of Peace Research - Replication Data
<http://www.prio.no/Journals/Journal/?x=2&content=replicationData>
- NBER - Data <http://www.nber.org/data/>
- Paul Hensel’s International Relations Data Site <http://www.paulhensel.org/data.html>
- Peace Research Institute, Oslo, Norway <http://www.prio.no/Data/>
- Polity IV - Political Regime Characteristics and Transitions
<http://www.systemicpeace.org/polity/polity4.htm>
- Princeton University - Economics Data Links
<http://library.princeton.edu/catalogs/articles.php?subjectID=109>
- Resources for Economists (RFE) - Data http://rfe.org/showCat.php?cat_id=2
- Slovak Election Data Project (Marek Hlavac)
https://sites.google.com/site/marekhlavac/slovak_election_data_project

- UC-San Diego - Data and Statistics for Political Science
<http://ucsd.libguides.com/content.php?pid=62534&sid=567117>
- United Nations - UN Data <http://ucsd.libguides.com/content.php?pid=62534&sid=567117>
- University of Rochester - Guide to Political Science Data
http://www.lib.rochester.edu/index.cfm?page=3869&this_pageID=53
- World Bank - Barro-Lee Educational Attainment Data <http://go.worldbank.org/HKOH13Y5D0>
- World Bank - Database of Political Institutions <http://go.worldbank.org/2EAGGLRZ40>
- World Bank - World Development Indicators
<http://data.worldbank.org/data-catalog/world-development-indicators>
- World Health Organization (WHO) - Data
<http://data.worldbank.org/data-catalog/world-development-indicators>
- World Justice Project - Rule of Law Index
<http://worldjusticeproject.org/rule-of-law-index>
- Resources on Duke University web-site <https://stat.duke.edu/resources/datasets>
- Data surfing on the WWW, from Robin Lock (<http://it.stlawu.edu/~rlock/datasurf.html>)
- *Data Analysis Using Regression and Multilevel/Hierarchical Models*
<http://stat.columbia.edu/~gelman/arm/>
- StatLib at CMU <http://lib.stat.cmu.edu>, including the Data and Story Library
- *The Statistical Sleuth* (please do not use any data set that we have already explored on a homework problem or in class) <http://www.statisticalsleuth.com/> - under Data Sets.

Grading:

Each member of your group will receive the same grade on the following components:

- **40% - Statistical accuracy and appropriateness of statistical tools.** Does your project demonstrate a solid understanding of the course material? For a data project: Did you make sound statistical choices for your data analysis and justify these choices? For a theory project: Are your statistical claims accurate?
- **40% - Breadth, depth, and motivation.** Have you explained why your topic is important and/or interesting? For a data project: Did you fully explore the question(s) of interest and come to a satisfying conclusion? For a theory project: To what extent does your project extend the course material in a thought-provoking way? Did you cite any required references?
- **10% - Overall impression.** Does the project have a professional appearance and a polished look? Was it creative, unique, or insightful? Was it apparent that a lot of effort was put into the project? Was there an overall clear purpose and conclusions?
- **10% - Proposal and Project Update.** Did you submit the proposal and project update on time? Did you show you had thought about your project in these submissions?