

UNIVERSITY OF OREGON
Department of Economics

EC 320: Introduction to Econometrics
Instructor: P. Economides
Winter 2022

Data Project

As part of bolstering your programming skills and integrating them with your econometric knowledge, this project requires you to prepare data, visualize key stylized facts and evaluate an empirical question of interest. Alternatively you may send me a specific paper you wish to replicate the results of, using their provided data set. **Submit online via Canvas in PDF format by 11:59pm March 1st.** Upload your adjoining code separately in a .R or .RMD file. Failure to comply with the guidance above may result in reduced points, depending on the extent by which the student has deviated. In terms of structure, I expect a concise study which includes each of the following sections. Use double-spaced text. I would expect submissions between 10-15 pages in length for the final submission (excluding your citations and appendices).

- Introduction: Describe the data source type (panel, cross-sectional, etc.), its frequency (hourly, annual, etc.), and how representative the sample is. (0.5–1 page)
- Cleaning: Ultimately, you want to present this data in a clear and legible manner for its intended audience. This may require identifying outliers and treating for missing observations. List the challenges you encountered when preparing this data for its intended audience. Outline some of the limitations the data may face upon tackling these challenges. (0.5–1 page)
- Summary Statistics: This is the most creative portion of the assignment. While you are not addressing causal inference in this portion of the assignment, your presentation of the data may still hint at potential research questions underlying the material. Present a series of interesting facts about the data. Infer potential causal relationships and describe how you may go about testing those relationships. (4 – 8 pages)
- Regression: Provide inference based on a series of regressions carried out to explore your causal question of interest. Keep in mind that multiple stages of a regression are highly informative with respect to issues such as omitted variable bias. Discuss

the key assumptions necessary in order for your results to be considered unbiased and inference with respect to standard errors to be apt. (3 – 4 pages)

- Conclusion: For the more time-constraint reader, perhaps a future employer, you always want to be concise with your statements. Use this section to provide concluding remarks about the data, detailing how rich it appears to be and the potential uses it may provide in terms of empirical analysis. (1 page)

In addition to the steps above, bonus points will be provided to individuals who sufficiently motivate the importance of their study and provide adequate background information (this should be included in your introduction). Below I list a number of available data sources along with brief descriptions. Consider your own interests when choosing which to work with. Here I've included sources that would appeal to those interested in urban economics, international trade, health and crime.

- [NBER Archive](#) - An eclectic mix of public use economic, demographic, and enterprise data obtained over the years to satisfy the specific requests of NBER affiliated researchers for particular projects.
- [IPUMS USA](#) - Household Census data representative of the US population, detailing household demographics (sex, age, marriage status), economic characteristics (education, work status, income) by year and geographic regions (state, county, MSA)
- [Google Dataset Search](#) - A search engine for data sets. Mostly useful for finding previous studies that have made their data sets available publicly but are not large standard data sets of the kind you can get on IPUMS.
- [ITA](#) - US trade data at the national, state, metro and industry level. Third parties included, such as USA Trade Online, offer particularly granular data at a monthly frequency that allows you to separate by product code as well as distinguish by origin and destination partner countries.
- [BEA USDIA](#) - Provides data on US direct investment abroad, specifically looking at the activities of U.S. Multinational Enterprises. This annual data ranges across the last decade from 2009.
- [Bureau for Research and Economic Analysis of Development](#) - Duke University's Developing Country data, as maintained by Duncan Thomas. Given the spectacular increase in the availability and quality of data from developing countries in recent years, many of these datasets are in the public domain. This page is intended as a resource to help locate those data.

- [RAND FLS](#) - These family life surveys are provide detailed household and community surveys of developing countries. The currently available country surveys cover Malaysia (1976-77, 1988-89), Indonesia (1993, 1997, 2000, 2007, 2014), Guatemala (1995), and Bangladesh (1996).
- [Stanford Open Policing Project](#) - This project has gathered over 200 million records on traffic stops from dozens of state and local police departments across the country.
- [IHME](#) - The Institute for Health Metrics and Evaluation provides a wide range of health outcomes such as cholesterol levels, life expectancy, smoking prevalence and alcohol use. Measures can vary by race, state, county etc., depending on the specific data in mind.
- [CAIT Climate Data](#) - Highly recommend for those interested in environmental studies. These sets of data project information on historical emission levels, Paris agreement contributions, emission project performances, country pollution profiles and emission pathways going forward.
- [EPA AQS](#) - This air quality system uses local monitors across the country to track daily emission levels by CO₂, NO₂, Ozone, SO₂, PM_{2.5} and PM₁₀ emissions. The data is therefore extremely granular. For large downloads, I would recommend exploring the [R-integrated API tool](#).
- [Federal Reserve Economic Data](#) - Download, graph, and track 816,000 US and international time series from 107 sources. Frequencies in the data range from daily to annual and touch upon a wide variety of topics of interest.

Rachel Heath and Yu-chin Chin have also provided an [excellent list](#) of potential sources. If you would rather work with a different source, let me know by email so I can appraise the idea. I will be available at office hours, or by appointment, if you would like any feedback on your project idea or simply need to brainstorm. Your classmates and TA are also vital resources, so make sure to bounce your ideas around as much as you can.