Stigler Center: 2022-23 Research Professional Task

George J. Stigler Center for Study of the Economy and the State[*]

Fall, 2022

Instructions

Please complete the following tasks and submit the required materials by the designated deadline. Ensure that you follow the directions carefully and submit all requested materials.

General

Both the code and the results that you submit should be legible to your peers and those who would be your supervisors. That is, ensuring that your code is well-documented and well-commented; and that your written reports and data visualizations clearly convey your analysis is of the utmost importance.

Additionally, writing easily manageable and reproducible code is critical for research. For each task, please write a README that describes each of the files that you have created/used and what each does. This should be written so that a user with a technical background but with no knowledge of the task itself can understand your code and reproduce all of the output products with minimal (if any!) changes to the code itself.

When submitting your materials please submit all your materials in a single compressed `.zip` file named as follows:

- `[firstname]_[lastname].zip`

That `.zip` file should contain three folders (`task1a`, `task1b`, and `task2`). See below for additional submission details.
If examples are given, follow them closely. Good luck and have fun!

Disclaimers

The contents of this coding activity are **confidential**. Do NOT share this document or any of the attached material with anyone. Sharing of these instructions or any associated materials will result in your immediate disqualification and may preclude you from future opportunities with the University of Chicago.

---

[*]https://www.chicagobooth.edu/research/stigler

1   Python and Web Scraping

1A   *American Economic Review*:

Using Python, collect publication data from the *American Economic Review*. Using Python and your choice of packages (for example BeautifulSoup, Selenium, or requests) and Pandas, collect the following information for all available issues:

- Volume and issue number;

- a hyperlink; and

- issue-date information.

Store the collected issue-level data in a `.csv` file. Your resultant spreadsheet should be structured follows:

Table 1: Example Task 1A Output Structure

volume	link	issue_date
Volume 112	https://www.aeaweb.org/issues/692	September 2022 (Vol. 112, No.9)
Volume 112	https://www.aeaweb.org/issues/689	August 2022 (Vol. 112, No.8)
Volume 112	https://www.aeaweb.org/issues/685	July 2022 (Vol. 112, No.7)
...	...	...

Inside a folder named `task1a` save your README (use any widely used text format you desire, such as markdown: `.md`), code, and spreadsheet as follows:

- `task1a_[firstname]_[lastname]_README.md`

- `task1a_[firstname]_[lastname].csv`

- `task1a_[firstname]_[lastname].[py/ipynb]`

1B   *American Economic Review*: Redux

Note that each article published by the *American Economic Review* is labeled with JEL codes which describe to which field(s) and subfield(s) the article contributes. Each alphanumeric JEL code is associated with a human-readable description. Using the same packages as above, construct a *second* spreadsheet that contains the following bibliographic information:

- volume and issue number;

- issue-date information;

- the article's title;

- the article's author(s);

- the article's print page-numbers;

- a *permanent link* to the article;

- the article's JEL code(s); and

- the article's JEL code description(s).

Store the collected issue-JEL-code-level data in a `.csv` file. Your resultant spreadsheet should be structured as follows:

Inside a folder named `task1b` save your README (use any widely used text format you desire, such as plaintext: `.txt`), code, and spreadsheet as follows:

- `task1b_[firstname]_[lastname]_README.txt`

- `task1b_[firstname]_[lastname].csv`

- `task1b_[firstname]_[lastname].[py/ipynb]`

Table 2: Example Task 1B Output Structure

volume	issue_date	article_title	authors	page_numbers	article_link	jel_code	jel_description
Volume 112	September 2022 (Vol. 112, No.9)	Belief Elicitation and Behavioral Incentive Compatibility	David Danz; Lise Vesterlund; Alistair J. Wilson	pp. 2851-83	https://doi.org/10.1257/aer.20201248	D83	Search; Learning; Information and Knowledge; Communication; Belief; Unawareness
Volume 112	September 2022 (Vol. 112, No.9)	Belief Elicitation and Behavioral Incentive Compatibility	David Danz; Lise Vesterlund; Alistair J. Wilson	pp. 2851-83	https://doi.org/10.1257/aer.20201248	D91	Micro-Based Behavioral Economics: Role and Effects of Psychological, Emotional, Social, and Cognitive Factors on Decision Making
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	D22	Firm Behavior: Empirical Analysis
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	G31	Capital Budgeting; Fixed Investment and Inventory Studies; Capacity
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	G35	Payout Policy
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	H25	Business Taxes and Subsidies including sales and value-added (VAT)
Volume 112	September 2022 (Vol. 112, No.9)	Dividend Taxes and the Allocation of Capital	Charles Boissel; Adrien Matray	pp. 2884-2920	https://doi.org/10.1257/aer.20210369	H32	Fiscal Policies and Behavior of Economic Agents: Firm
...	...	...	...	...	...	...	...

2 Data Analysis

For this task feel free to use the software program/suite of your choice. However, for purposes of reproducibility, Stata, R, or Python are preferred.

For this task, you should have received the following files[1]:

- `state_chars_text.csv`: some information about state-level institutional characteristics

- `law_firms_donations.csv`: donations data – donations made by *lawfirms*

- `lawyers_donations.csv`: donations data – donations made by *lawyers*

- `nonlaw_firms_donations.csv` donations data – donations made by non-lawfirm *firms*

- `nonlaw_indivs_donations.csv` donations data – donations made by non-lawyer *individuals*

Background

In this short task you will conduct a preliminary analysis regarding the role of state-level institutional characteristics and donations to the electoral campaigns of state supreme court judges in the United States. In more than 20 states, state supreme court judges (the term used to describe judges who sit on that state's court of final appeal) are elected by state residents, rather than being appointed by a state executive. The merits of having elected, rather than appointed judges, are unclear. At once, elected judges both embody the virtue of being responsive to a democratic electorate and are subject to those same democratic pressures that may undermine their impartiality.

Donations

As with other elections in the United States, individuals and corporations can make contributions to a candidate's campaign; those contributions can be spent on making advertisements, facilitating a candidate's travel for campaigning purposes, hiring support staff, etc.[2] Each state has slightly different rules about limits on donations and what must be disclosed when individuals or organization make those donations but in all states, those donations are a matter of public record.

Institutional Characteristics

In the United States, the structure of each state's government is determined by that state's own constitution. Consequently, there is considerable variation in the characteristics of each state's judiciary.[3] Examples of such characteristics include but are not limited to:

- Initial selection method. Some states' supreme court judges are only ever appointed by the executive. Others must win election.

---

[1]These files might require some cleaning before use. Make sure to keep copies of the originals so that your work can by replicated!

[2]Note that direct donations to campaigns are *different* from contributions to "independent expenditure" groups, so-called "super PACs." Donation data is for individuals and firms has been separated because different contribution limits may apply.

[3]For a bit more background or detail, follow this link.

- Subsequent selection method. In some states, judges can serve until retirement age or for life. In others, judges must consistently run for re-election. The types of those re-election competitions also vary from state to state.

- The length of the terms (and if there is a limit to the number of terms) that supreme court judges can serve.

- The number of judges who serve on the state's supreme court bench at any time.

Task

Your task is to reorganize and clean the provided data, perform a preliminary data analysis, and write a short report describing the results of your work. If a task is specified in the below instructions, it must be completed. However, a fully discursive set of instructions are not provided because we are interested in seeing how you independently work to ask and answer research questions.

1. For this task, rename variables as you wish but if you do rename variables make sure the new names are descriptive and that they are readily interpretable.

2. Add three new dichotomous variables to the state-characteristics dataset. If you think the coding could be ambiguous, describe your methodology and provide a justification for your decisions.

   a. `binding_commission`: coded as 1 if mid-term vacancies are filled by candidates suggested by a binding commission.
   b. `partisan_election`: coded as 1 if candidates face a partisan-election in their first election.
   c. `retention`: coded as 1 if incumbent candidates face a retention election for subsequent terms.

3. Compute the minimum, maximum, median, interquartile range, and mean donation size by individuals *for each state* over all races in states that have term-lengths of 8 years. Compute the same statistics only for lawyer-individuals. Present the results in a table.

4. Create a new dataframe/spreadsheet that aggregates all donations at the candidate-race level. Produce histograms that show the frequency of total-contributions by the candidates' political party affiliation.

5. Identify the top 10 largest law firm donors in Illinois. Compute a law firm's donations as the sum of all donations to all candidates in all races in the sample. Present the firm names and the sum of their donations in a table.

6. Identify the top 10 largest non-lawyer-individual donors who donated at least once to any race in a state that has a statutory/constitutional provision to run retention elections for a judge's subsequent terms. Compute an individual's donations as the sum of all donations to all candidates in all races in the sample. Present the individual names, the states in which they donated to races, and the sum of their donations in a table.[4]

---

[4]Hint: your table should only contain 10 rows.

7. Conduct some analysis to try to determine:

    a. if any of a state's institutional characteristics affect donation behavior; and

    b. whether that donation behavior affects election outcomes.

Use summary statistics, regression analysis, and data visualization as you see fit.

Please perform the above tasks and write a report discussing the key steps and decisions you made when exploring the data (and how those may have affected your results). Your objective here is to use the data *provided* to establish a compelling argument for the validity of your methodology and, therefore, whatever results you find. Your report can make use of tables with descriptive/summary statistics, data visualizations, regression analysis, or other qualitative information you see fit to include. If you do include regression analysis in your report, please ensure that each specification is at some point accompanied with an interpretation of those results.

Your report should also include a conclusion section that: 1) summarizes your results; 2) identifies the limitations of the dataset provided (and potentially your methodology); and 3) additional analysis you would perform if you had access to any dataset of your choice.

Inside a folder named `task2` save all of the data files you use, your README (use any widely used text format you desire, such as PDF: `.pdf`), any code you may have written to clean/analyze the data and/or compile your report, and any miscellaneous files that might be required to reproduce your report from scratch. Please ensure that your report is submitted as a widely used text format. Acceptable file types include `.docx`, `.pdf`, `.md` etc. Ensure that your *report* is named as follows.

- `task2_report_[firstname]_[lastname].[pdf/docx/md]`

Other auxillary files should follow a similar convention but can be named as you see fit.