PLSC 541: American Political Institutions

Exercise #1

Spring 2022

Overview

This is the first of three research exercises for this class. There are a few general goals for this exercise, including:

- to begin to get you accustomed to working with court and legal data,
- to develop skills in data management and visualization, and
- to establish expectations regarding the content and format of the class exercises.

More specifically, this assignment will require that you acquire, merge, clean / recode, and visualize data in a basic / descriptive way, that you conduct a series of straightforward graphical and statistical analyses, and that you present them in a manner and format that are appropriate for a professional audience in political science.

Data

You'll be working with data on the U.S. Supreme Court. More specifically, you'll use two data sources for this exercise:

1. The 2021 release of the Supreme Court Database (SCDB). This is a database containing data on the decisions of the Court. The database has several components and formats; for this module, you'll be working with:

   · the modern database, which contains data on decisions handed down during October Terms from 1946-2020, inclusive;
   · the justice-centered data (which contains one line of data for each justice voting in each case), with
   · cases organized by docket, yielding $N = 94687$ as of February 1, 2021.

   The codebook for the SCDB is available here and here; you should familiarize yourself with its contents.

2. The Biographical Directory of Article III Federal Judges, compiled by the Federal Judicial Center (FJC). These data contain biographical information on every individual who has ever served on an Article III court, including all federal district court judges, court of appeals judges, and Supreme Court justices. You can and should familiarize yourself with the contents of those data (e.g., by looking around a bit here).

   For purposes of this assignment, I've created a smaller version of the FJC biographical data (called "`FJC-SCOTUS.csv`") that is available on the course github repo, in the "Data" folder. Those data also contain identifiers (`justice` and `justiceName` that are consistent with the SCDB, allowing the two data files to be match-merged.

Assignment

Your assignment for this module is as follows:

1. Acquire the two datasets, read them into R , and match-merge the biographical information in the FJC data with the case- and vote-level information in the SCDB.[1]

2. Plot a figure showing the number of male and female justices that served on the U.S. Supreme Court during this period (OT 1946-2020). Note that I am *not* asking for the number of votes cast, just the number of justices.[2]

3. Plot a histogram of the *birth years* of the justices that served during the OT 1946-2020 era.[3]

4. Generate a frequency table for the distribution of `issueArea`s in the cases during the 1946-2020 period.[4]

5. Generate the same frequency table, this time expressing each cell as a proportion of the total number of cases.[5]

6. Plot a time-series of the number of civil rights and liberties cases (that is, cases for which `issueArea` < 6) heard each term.[6]

7. Plot a similar time-series of the proportion of the Court's decisions that were decided in a liberal direction, by term.[7]

8. Conduct and report a t-test for the hypothesis that the mean `direction` of justices' votes is different for white justices (`Race.or.Ethnicity = "White"`) than for nonwhite justices.[8]

9. Repeat the immediately previous t-test, but conduct the tests separately for each of the 14 `issueArea` codes (to the extent that is possible).

10. Finally, fit, present, and discuss a regression model[9] of the `direction` of justices' votes as a function of their gender, the party of the president who appointed them, and the `issueArea` of the case.

---

[1] Hint: This requires the use of the `merge` command, and should yield a data frame with 94687 rows and 269 columns.

[2] Hint: You can use the `aggregate` command – or `dplyr` – to aggregate the SCDB to the individual (`justiceName`) level, or you can just word directly with the FJC data.

[3] Hint: This is similar to what you're asked to do in item 2.

[4] Hint: The justice-centered SCDB has up to nine lines of data for each case – one for each justice voting in that case. To create this frequency table, you will need to extract one line of data for each of the 10573 "cases" identified by the `docketId` variable.

[5] Hint: `prop.table` is your friend here; also consider using `round()` to make things less messy.

[6] Hint: First, reduce the data to the case level (as in item 4); then extract only the civil rights and liberties cases; finally, aggregate the data by `term` (e.g., using `table`).

[7] Hint: The variable indicating the direction of the Court's decision is called `decisionDirection`, and is coded `1` for a conservative decision and `2` for a liberal one. Creating a variable (for example, called `liberalDecision`) by subtracting 1 from `decisionDirection` will turn that into a more conventional 0/1 "dummy" variable.

[8] Hint: Like `decisionDirection` above, the `direction` variable – which indicates the ideological direction of each justices' vote in each case – is also coded `1` for a conservative vote and `2` for a liberal vote.

[9] Here, use whatever regression tool(s) you are familiar with. There's a diverse range of levels of methods training in the class; I'll consider that fact when evaluating your assignment.

Format, Due Date, Etc.

For this exercise, formatting is important. In this case, please ensure that your finished exercise follows conventions for a political science conference paper: Adobe PDF format; a separate title page with title, date, and author + affiliation; and clearly labeled responses to each of the items above. In your exercise, please demonstrate completion of and/or provide answers to each of the items in the *Assignment* section, above. These need be no more than a sentence or two (or maybe three), or a single figure or small table.

You should submit your finished module electronically, via email, to Professor Zorn (at `zorn@psu.edu`). Please be sure to submit all the code and data necessary to replicate the analyses you conducted, either as (a) a separate `.R` file containing computer code, (b) an in-line appendix within your PDF, or (c) a component of the R Markdown file you used to create the module itself. Modules must be submitted *in either Adobe* `.pdf` *format or as an R Markdown* `.Rmd` *file* no later than 11:59 p.m. ET on **Tuesday, February 22, 2022**.