PLSC 476: Empirical Legal Studies

Research Module Assignment #1

Spring 2021

Overview

This is the first of four "research module" assignments for this class. *This module is different from the other three*, in two important respects. First, unlike the other three research modules, this assignment is a "common" assignment – all members of the class will complete the same assignment. For the other three research modules, students will select a topic and complete the research module on their own. Second, this assignment has a specific due date, after which submissions will not be accepted. Students not submitting their research module #1 by the due date specified below will receive zero points on the assignment. In contrast, the other three research modules are all due on or before the final day of the course.

Goals

There are a few general goals for this module, including:

- to (re-)familiarize you with using R for data analysis,
- to begin to get you accustomed to working with court and legal data, and
- to establish expectations regarding the content and format of the class research modules.

More specifically, this assignment will require that you acquire, merge, clean / recode, and visualize data in a basic / descriptive way, and that you conduct a series of simple graphical and statistical analyses.

Data

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

1. The 2020 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 1946-2019, inclusive;

   · the justice-centered data (which contains one line of data for each justice voting in each case), with

   · cases organized by docket, yielding $N = 93967$ as of February 1, 2021.

   The link to a zipped `.csv` of the specific data you'll be using is here. 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. Hint: This requires the use of the `merge` command, and should yield a data frame with 93967 rows and 269 columns.

2. Plot a figure showing the number of male and female justices that served on the U.S. Supreme Court during this period (OT 1946-2019). Note that I am *not* asking for the number of votes cast, just the number of justices. Hint: You can use the `aggregate` command – or `dplyr` – to aggregate the SCDB to the individual (`justiceName`) level.

3. Plot a histogram of the *birth years* of the justices that served during the OT 1946-2019 era. Hint: This is similar to what you're asked to do in item 2.

4. Generate a frequency table for the distribution of `issueAreas` in the cases during the 1946-2019 period. 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 10493 "cases" identified by the `docketId` variable.

5. Generate the same frequency table, this time expressing each cell as a proportion of the total number of cases. Hint: `prop.table` is your friend here; also consider using `round()` to make things less messy.

6. Plot a time-series of the number of civil rights and liberties cases (that is, cases for which `issueArea` < 6) heard each term. 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. Plot a similar time-series of the proportion of the Court's decisions that were decided in a liberal direction, by term. 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. 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. 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.

Format, Due Date, Etc.

For this research module, the format is straightforward: Demonstrate completion of and/or provide answers to each of the items in the *Assignment* section, above. In this case, these need be no more than a sentence or two, or a single figure or small table.[1]

You should submit your finished module electronically, via email, both to Professor Zorn (at `zorn@psu.edu`) and to Mr. Suman (at `wts5131@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 6:00 p.m. ET on **Tuesday, February 16, 2021**.

[1] Your future research modules will have a somewhat different format; as noted in the description of the research module assignments, "research modules should generally be 3-4 pages long, and will include 2-3 graphs and between 500 and 1000 words of text. They will be written in sections, with the first setting out the question to be investigated, the second describing the data and the empirical approach, and the third setting forth the findings."