supreme_court_report

Connor Putnam

2/14/2021

Obtain

The first step in this project was to import the data, with the focus being on the questions regarding the United States Supreme Court. The data used can be found at the following Git repository: supreme court transcripts. The repo contains records for every supreme court case see by the justices of the United States Supreme Court. It is spread throughout 15,375 `JSON` files and takes up 3.17 GB. In order to make this usable for data analysis it must be converted into a data frame with the relevant information withdrawn.

Scrub

In order to clean everything up the first step was to upload the data into python. This was challenging given the amount of **JSON** files present. The next task was to create a folder with all of the **JSON** files in it. Then python was used to iterated through all the files, this process can be seen in the code snippet below.

```
def load_json_files(list_of_files):
    vals = []

    list_of_files = [x for x in list_of_files if ".json" in x]
    for i, file in enumerate(list_of_files):
        print(i, len(list_of_files), file)
        with open(
            f"/Users/connorputnam/Documents/CS512/cases/{file}", errors="ignore"
        ) as f:
            d = json.load(f)
            vals.append(d)
    return vals
```

After that I was interested in the files that deal with the **JSON** key `decisions`. Approximately 16000 of the **JSON** files were dealing with the individual case decisions. The following code indexed all the **JSON** not containing the key `decisions` into a list and then subtracted those files from the list containing all the files and stored them under the variable `val_updated`.

```
    df2 = df[df["decisions"].notna()]  # drop them
    list(df2.index)  # list of number to feed to vals
    val_updated = [vals[i] for i in list(df2.index)]  # feed them
```

After filtering out those files the next was nest was to flatten the **JSON** data by using the `json_normalize` function in the `pandas` library. Now the data was in a useable `pandas` dataframe. The code below shows this, additionally the `record_path` argument allows you to *unnest* the **JSON** data to obtain what you are looking for.

```
    df_decisions2 = pd.json_normalize(
        val_updated, record_path=["decisions", "votes"], errors="ignore"
```

```
)
```

In addition to the data obtained from the code above I was also interested in obtaining the information contained in the `member.roles` column of the `pandas` data frame `df_decisions2`. This column was expanded by using the `concat` function which is also found in the `pandas` package.

```
df_appoint = (
    pd.concat(
        {i: pd.DataFrame(x) for i, x in df_decisions2.pop("member.roles").items()}
    )
    .reset_index(level=1, drop=True)
    .join(df_decisions2, lsuffix="_left")
    .reset_index(drop=True)
)
```

The last step in getting the **JSON** file in a useable format was to convert the newly created `pandas` data frame into a `.csv` file, suitable for analysis, by using the `to.csv` function.

Once the files were in a usable format, **csv** file, it was uploaded into `R` and the `tidyverse` package was used to perform data manipulation and get the desired information for visualization.

Explore

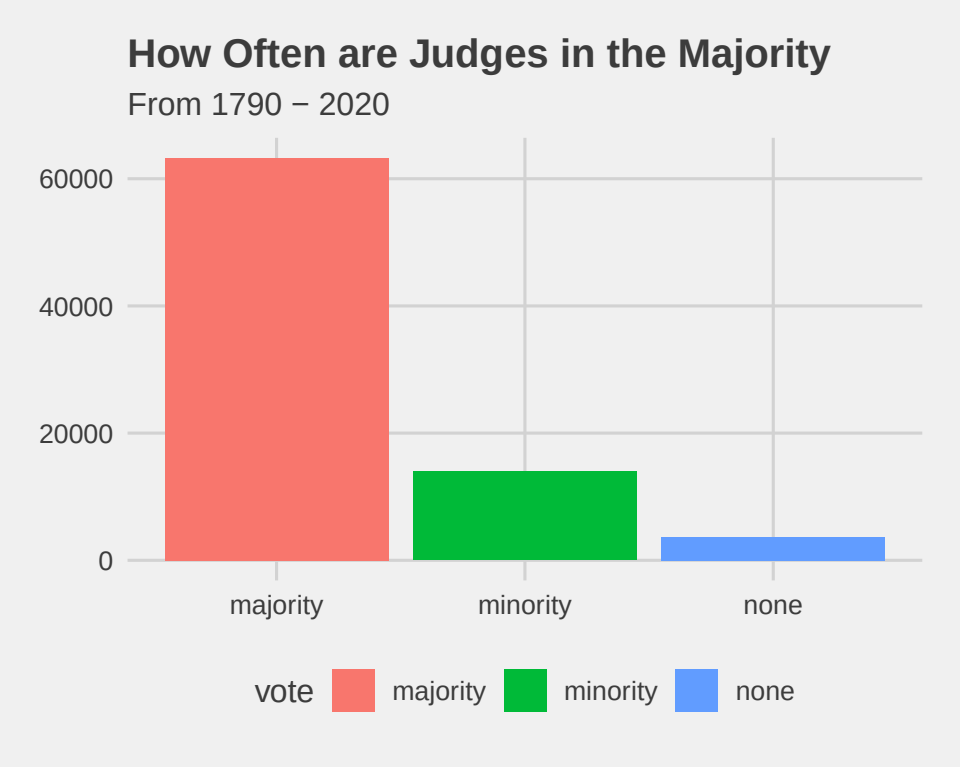
There was not much exploration to do in this project besides a few summary statistics. It can be argued that visualizations might be under the umbrella of exploring but I choose to put those in the modeling section do to the complexity of them and because there was no model building in this assignment. So, for this section I just included a couple summary statistics as well a graph that helped influence my questions.

Table 1: Summary Statistics for Ideology

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-6.599	-1.198	0.48	0.1551	1.528	4.414

Table 2: Summary Statistics for Member Length of Service(Days)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
49	6839	8894	9302	12293	13358



Model

Three Data Analysis Questions

1. Which president has appointed the most judges? And have Republican or Democratic Presidents picked the most Judges? (Here just considering the post World War 2 era given that before then there were more parties like the Wigs and the parties do not fit into the modern "liberal' or"conservative" classifications)
2. In modern history how do the ideologies of judges compare to one another? Are Republican appointed judges more ideological or are Democratic appointed judges more ideological? Are any of the judge's ideologies the opposite of what one would expect?
3. Has the level of ideology changed throughout different times in American history?

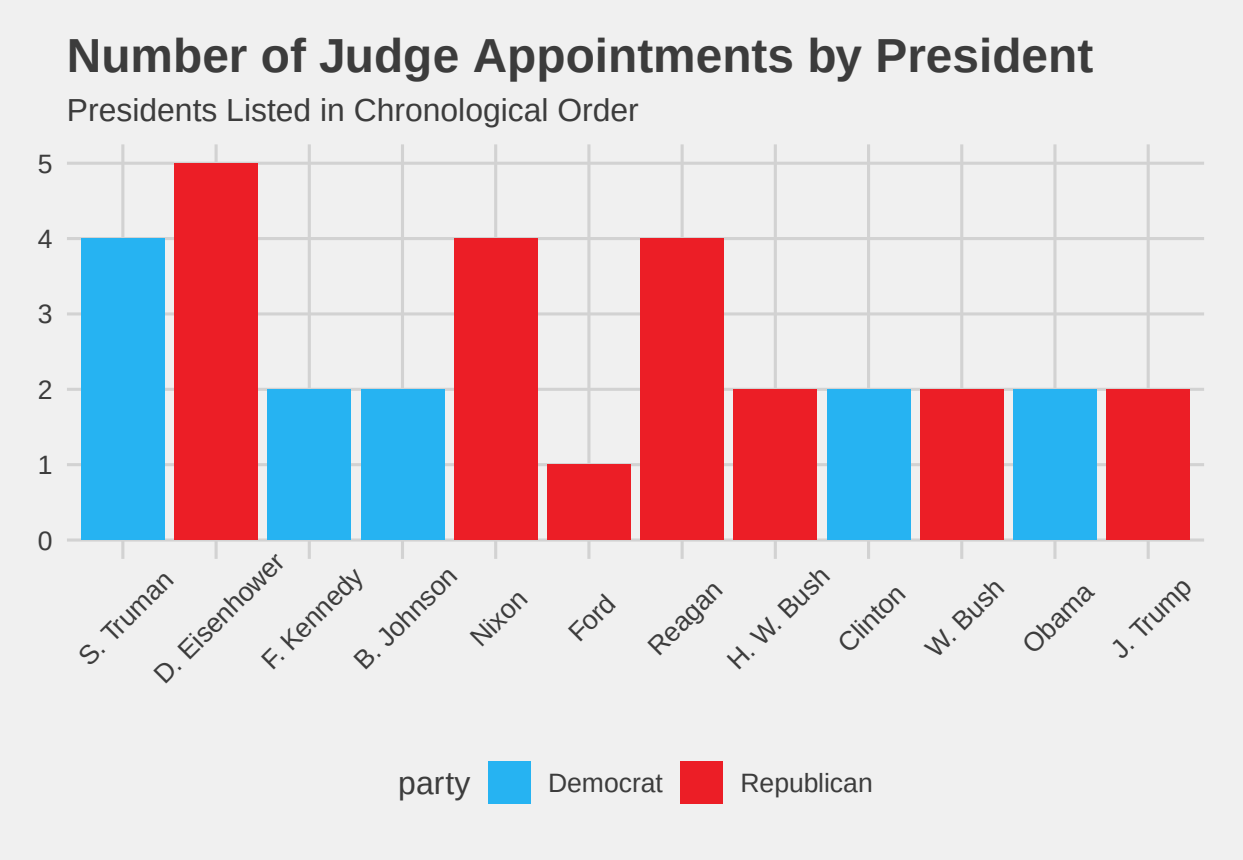
Question 1

For the first question I performed variable selection as well as some data manipulation using the `dpylr` functions such `filter`, `mutate`, `case_when` and `group_by` in order to make classifications. A sample of the created dataset is shown below:

presidental_year	FirstName	LastName	party	JudgeFullName
2017	Donald	J. Trump	Republican	Neil Gorsuch
2017	Donald	J. Trump	Republican	Brett M. Kavanaugh
2009	Barack	Obama	Democrat	Sonia Sotomayor
2009	Barack	Obama	Democrat	Elena Kagan
2001	George	W. Bush	Republican	Samuel A. Alito, Jr.
2001	George	W. Bush	Republican	John G. Roberts, Jr.

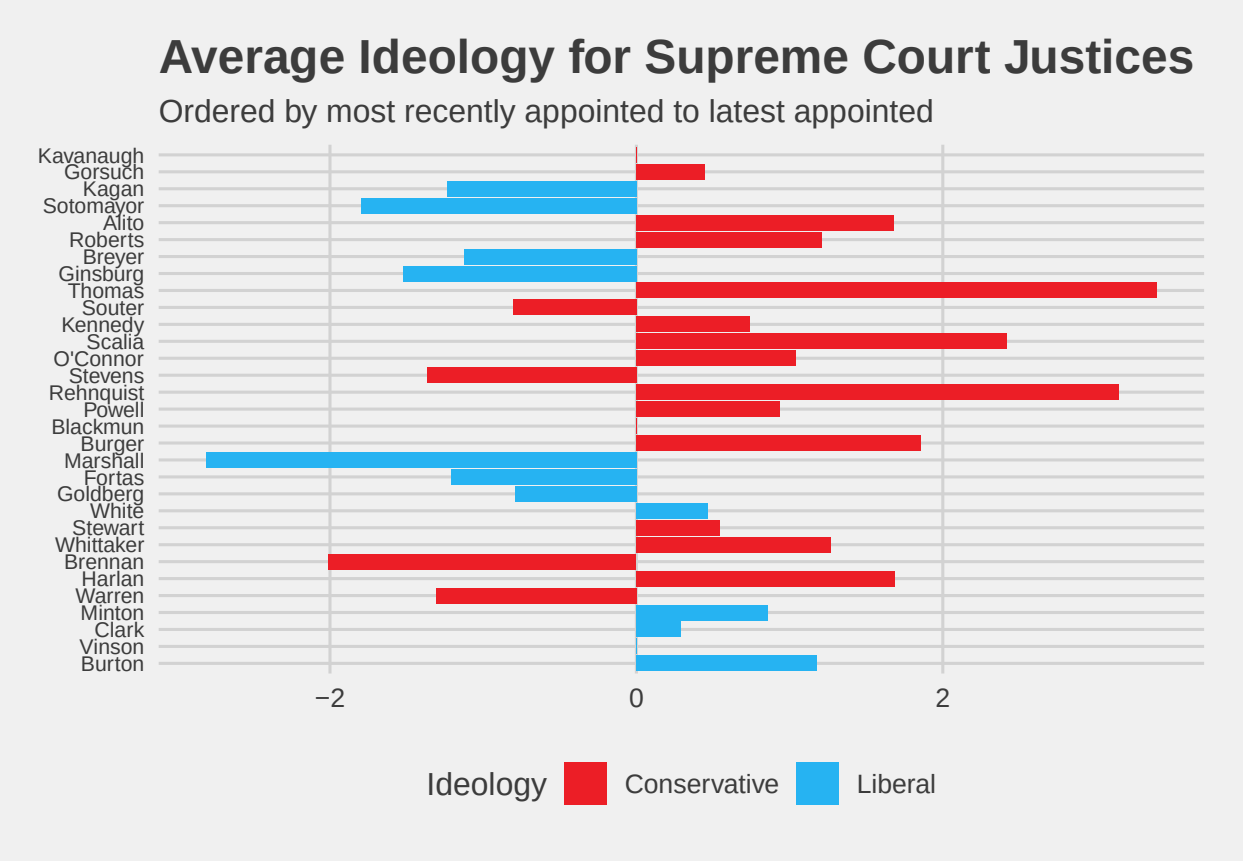
Based on the visualization below it looks like some presidents get to appoint more judges than others. In particular it appears as though Republican judges have been able to appoint more judges than Democrats. President Eisenhower(R) appointed the most judges at five whereas all of the democratic presidents besides Truman only appointed 2 judges. Note that Ford although a Republican did not serve a full term in office.

*At the time of this project the data did not contain any decisions made by the newly Trump appointed judge Amy Coney Barrett.



Question 2

In order to address the next question, I used similar data manipulation techniques as in the first question. Though this time I was interested in the ideology score of every judge on every one of the recorded decisions. To do this I took an average of every judge's ideology and then plotted them below. Where negative values indicate more liberal leaning and positive values indicate more conservative leanings.

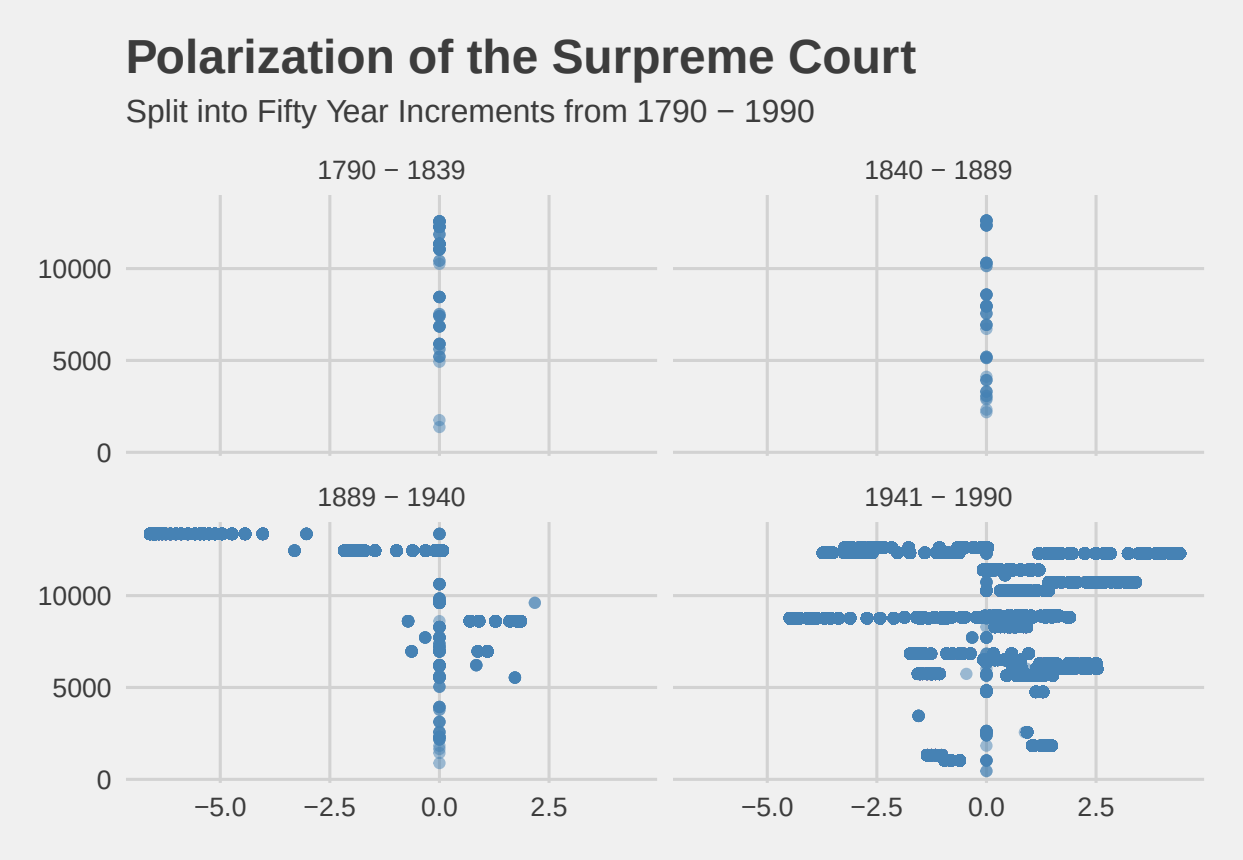


On the conversative side it looks like both `Thomas` and `Rehnquist` were both very much to the right and on the liberal side it looks like `Marshall` was very much to the left. I am particularly interested in the judges that have a `0` ranking, such as `Vinson` or `Blackmun`. To me a judge that is impartial to one side of the spectrum is a good thing, but it seems as though that is a pretty rare occurrence.

Question 3

The last question I tasked myself with was trying to gain some insight into if the ideology rates have changed throughout American history. Here is an I used all the ideology scoring available in the data. In addition to this there was data consisting of system time data representing the data in which a judge was appointed and another representing the time they retired I then took that data and aggregate into fifty-year time frame. I choose to stop at the year 1990, because the last judge to leave the bench was appointed in 1990.

From the plot below it can be seen that indeed the modern era of American history has seen much more ideologically driven judges than in previous eras.



Interpreting

For this last section I will review the process I went through in the previous steps, I discussed answers to my questions and provided context throughout the report. After going through everything the data wrangling portion of the project was the most challenging and required a good amount of time. I had to further unnest the **JSON** files so that way I got the right data for the questions I wanted to ask. The main challenge with the visualizations was adding in supplementary data, (i.e., president's party affiliation), and to wrangle the data into a columns that could be visualized. Grouping variables were key to this segment of the report.