6.10 PROJECTS

Project 6.1 Polarized politics

The legislative branch of the United States government, with its two-party system, goes through phases in which the two parties work together productively to pass laws, and other phases in which partisan lawmaking is closer to the norm. In this project, we will use data available on the website of the Clerk of the U.S. House of Representatives (`https://clerk.house.gov/Votes`) to analyze the voting behavior and polarization of the U.S. House over time.

Part 1: Party line votes

We will call a vote *party line* when at least half of the voting members of one party vote one way and at least half of the voting members of the other party vote the other way. In the House, when the vote of every member is recorded, it is called a *roll call* vote. Individual roll call votes are available online at URLs that look like the following:

`https://clerk.house.gov/cgi-bin/vote.asp?year=<year>&rollnumber=<number>`

The placeholder `<year>` is replaced with the year of the vote and the placeholder `<number>` is replaced with the roll call number. For example, the results of roll call vote 194 from 2010 (the final vote on the Affordable Care Act) are available at

`https://clerk.house.gov/cgi-bin/vote.asp?year=2010&rollnumber=194`

Run the following program to view one of these files:

```python
import urllib.request as web

def main():
    url = 'https://clerk.house.gov/cgi-bin/vote.asp?year=2010&rollnumber=194'
    webpage = web.urlopen(url)
    for line in webpage:
        line = line.decode('utf-8')
        print(line.rstrip())
    webpage.close()

main()
```

(All of the roll call votes from 1997 to 2019 are also archived on the book website, if this is more convenient.)

This file is in a format called XML (short for extensible markup language). In XML, data elements are enclosed in matching pairs of *tags*. In this file, all of the data for this particular vote is enclosed in a pair of `<vote-data>` ⋯ `</vote-data>` tags, beginning on line 56. The lines at the beginning of the file are metadata that store information like the name and date of the vote. Between the `<vote-data>` ⋯ `</vote-data>` tags, each vote is enclosed in a pair of `<recorded-vote>` ⋯ `</recorded-vote>` tags, like these (shortened considerably):

```
<recorded-vote><legislator name-id="A000022" ... </recorded-vote>
<recorded-vote><legislator name-id="A000055" ... </recorded-vote>
<recorded-vote><legislator name-id="A000364" ... </recorded-vote>
```

Look carefully at the file to find where the actual votes are cast within these lines. Then write a function

```
partyLine(year, number)
```

that determines whether House roll call vote `number` from the given `year` was a party line vote. Use only the recorded vote elements, and no other parts of the XML files. Do not use any built-in string methods. Be aware that some votes are recorded as Yea/Nay and some are recorded as Aye/No. Do not count votes that are recorded in any other way (e.g., "Present" or "Not Voting"), even toward the total numbers of votes. Test your function thoroughly before moving to the next step.

Question 6.1.1 *What type of data should this function return? (Think about making your life easier for the next function.)*

Question 6.1.2 *Was the vote on the Affordable Care Act a party line vote?*

Question 6.1.3 *Choose another vote that you care about. Was it a party line vote?*

Second, write another function

```
countPartyLine(year, maxNumber)
```

that uses the `partyLine` function to return the fraction of votes that were party line votes during the given `year`. The parameter `maxNumber` is the number of the last roll call vote of the year.

Finally, write a function

```
plotPartyLine()
```

that plots the fractions of party line votes for the last 20 years. To keep things simple, you may assume that there were 450 roll call votes each year. If you would prefer to count all of the roll call votes, here are the numbers for the last several years:

Year	Number	Year	Number
2019	701	2006	543
2018	500	2005	671
2017	710	2004	544
2016	622	2003	677
2015	705	2002	484
2014	564	2001	512
2013	641	2000	603
2012	659	1999	611
2011	949	1998	547
2010	664	1997	640
2009	991	1996	455
2008	690	1995	885
2007	1186	1994	507

Question 6.1.4 *What fraction of votes from each of the last twenty years went along party lines?*

Note that collecting this data may take a long time, so make sure that your functions are correct for a single year first.

Question 6.1.5 *Describe your plot. Has American politics become more polarized over the last 20 years?*

Question 6.1.6 *Many news outlets report on the issue of polarization. Find a news story about this topic online, and compare your results to the story. It might be helpful to think about the motivations of news outlets.*

Part 2: Blue state, red state

We can use the roll call vote files to infer other statistics as well. For this part, write a function

```
stateDivide(state)
```

that plots the number of Democratic and Republican representatives for the given `state` every year for the last 20 years. Your plot should have two different curves (one for Democrat and one for Republican). You may also include curves for other political parties if you would like. Test your function on several different states.

Call your `stateDivide` function for at least five states, including your home state.

Question 6.1.7 *Describe the curves you get for each state. Interpret your results in light of the results from Part 1. Has the House of Representatives become more polarized?*