

A Retrospective Exploration of the US Senate through Data

Jordan Poles, Tony Chen and Aparna Narendrula

December 6, 2015

1 An Introduction to the Data

This report details our analysis of a dataset covering the 101st to the 113th US Congresses (1989 - 2014). The dataset was extracted from several sources, and our goal was to use this data to assess trends in the members of the US Congress, their voting patterns and bills passed by this body. We focused on understanding the US Senate, which is one of the two chambers of Congress (the other being the House of Representatives). We selected the Senate for two reasons:

1. The Senate provides a more balanced overview of the political sentiments present on the state scale than does the House of Representatives.
 - (a) Each state is assigned 2 Senators, while the number of representatives for a state in the House of Representatives is based on the states population.
 - (b) The Senate is smaller in size and less susceptible to partisan issues and bureaucratic gridlock.
2. The Senate provided the lists of roll call votes used in our analysis in a much more accessible format (XML) than the House of Representatives (HTML).

To begin our analysis, we examined the characteristics of the senators who participated in each of Congresses 101 - 113 using the NY Times [Congress API](#). We refer to this data as member data for the remainder of this report. The member dataset was made retrieved as JSON via several calls to a URI based API. These JSON files - as well as all of the other datafiles we fetched - were cleaned by an R script (Appendix A) prior to storage in a SQLite database for analysis. This data was stored in the members table, and contains a variety of variables including identifiers such as name, ID, party, state, and seniority in addition to metrics such as percentage of votes with party and vote miss rate.

The member data motivated us to examine the voting behavior of senators. We proceeded to examine the bills which were voted on in each session of the Senate. Our analysis focuses on roll-call bills - those bills for which the vote of each senator is individually recorded. In order to examine these bills we first fetched a series of files from [senate.gov](#) containing lists of roll-call votes (and metadata) for each session (or year) of congress in an XML format. These data were stored in the senateRollCalls table, which contains the date of each vote, the issue at hand, and the outcome (overall number of yeas and nays). We refer to this as roll-call vote data.

The roll-call vote data from the Senate offers information about passage rates for individual bills but it does not reveal individual voting behaviour. This data was retrieved using our third and final data-source, [govtrack.org](#), in the interest of compiling a complete list of roll-call votes for lookup. The govtrack group provides a bulk data API which can be used to lookup the vote of each individual senator on a given bill. This data was retrieved as JSON before being stored in our third table votes. We used this information to examine each Senators voting behavior individually in relation to corresponding parties or congresses.

2 The Importance of Assessing our Government

The United States government is said to be by the people, of the people, and for the people. The importance of transparency cannot be overstated in order for such a government to remain accountable to the people. Therefore we were surprised that there was no readily published workflow for citizen analysis of data on important governmental bodies like the Senate. In this report we detail a protocol which can be used to fetch and store data regarding Senate members and their voting behaviours. The three-step process we detail for acquiring this data, is not straightforward, so we hope that the publication of workflows such as this will make US governmental data more accessible to the public. This workflow may also serve as an entry point into more advanced analysis, such as corpus analysis and machine learning, of Senate proceedings and bills.

3 Examining The Members of Senate

The NY Times Congress API provides an interface for accessing a number of characteristics on US Senators including their web presence, party and voting patterns. This provides us some introductory insights into the behavior of politicians elected by the US citizenry to represent them in the federal government.

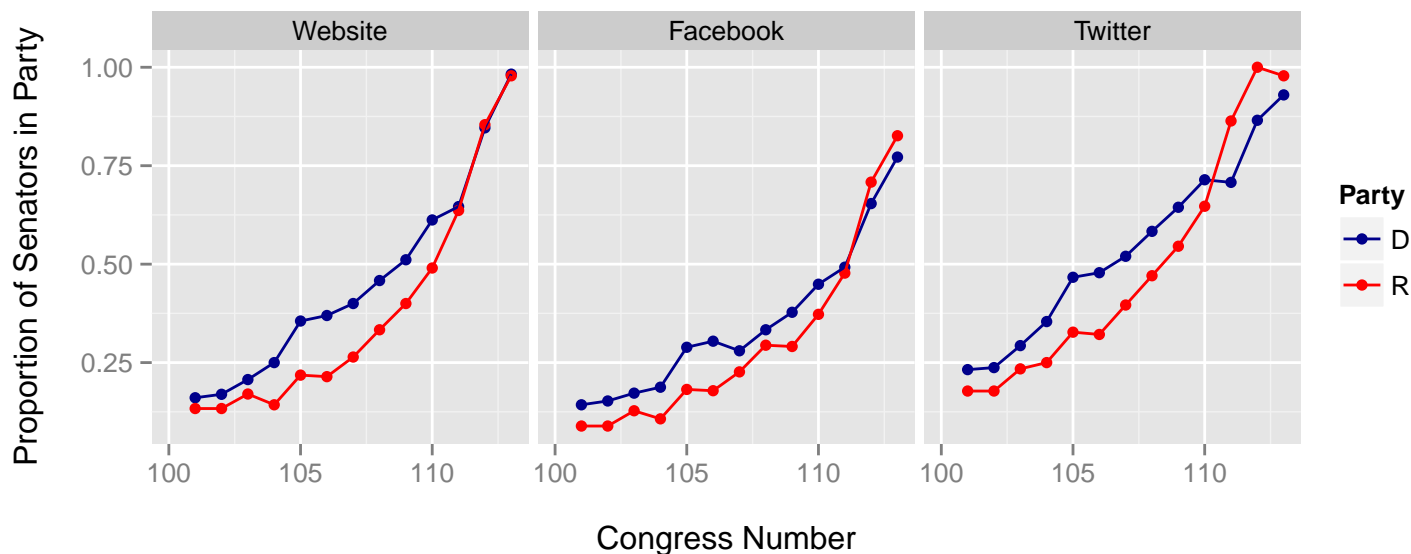
3.1 How have senators adopted web platforms over time?

The integration of technology into the modern American political process presents a number of interesting paradigms for exploration. For example, the electronic cataloguing and publication of Senate roll-call votes made all of this analysis possible in the first place, allowing for greater government transparency.

To begin our analysis, we turned to the use of web platforms by Senators. These platforms (including websites, Facebook, and Twitter) have become more common over time. Since the inception of FDR's fireside chats, electronic media has permitted mass communication of political thought. This capacity has been greatly expanded by the internet. This plot shows that usage of the three platforms has grown rapidly, although the use of Facebook has lagged behind the use of websites and Twitter. High web-platform usage is not surprising given the number of Senators who host their websites under senate.gov sub-domains (name.senate.gov). Twitter also appears to be popular and is often utilized to disseminate political ideology and news.¹

Senators Who Have Ever Used a Web Platform In Their Political Career

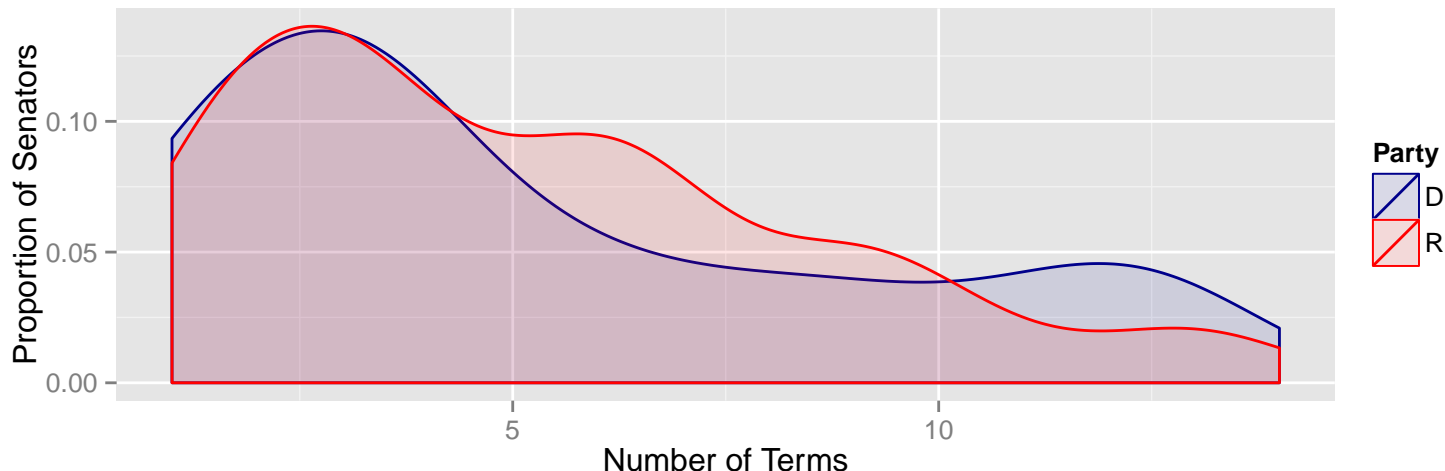
Facebook Appears to Lag Behind Websites and Twitter



3.2 For how many terms do Senators typically serve?

Many Senators serve no more than 5 terms. Democrats and Republicans have similar distributions, although more Republicans serve 5-10 terms, while more Democrats serve greater than 10 terms.

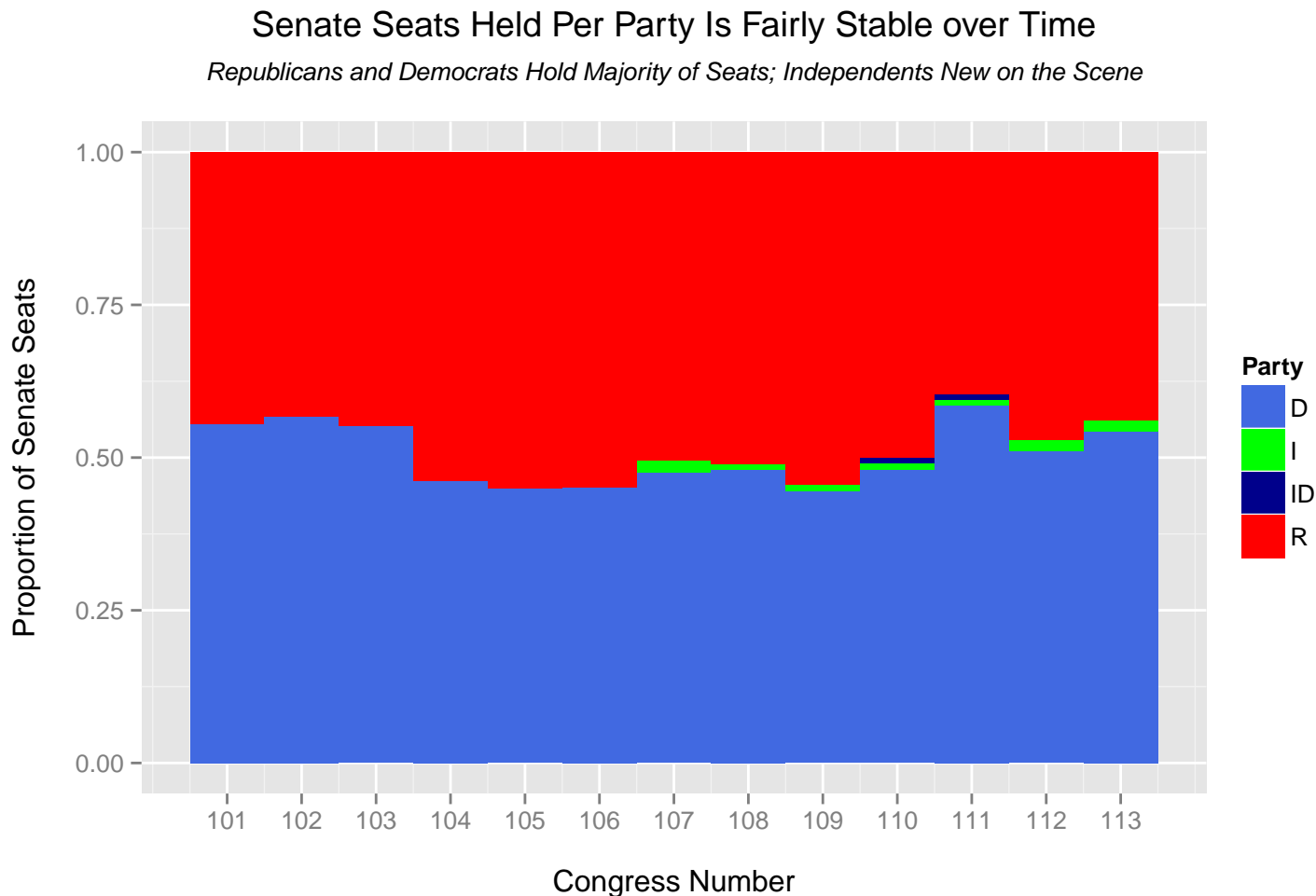
Distribution of Terms Served in 101st to 113th Senates



¹For instance: <https://twitter.com/McConnellPress>

3.3 What are the dynamics of political parties in the US Senate over time?

This figure is the culmination of an analysis of the power held by each of the political parties in the US Senate. We confirm the prevailing understanding of American politics as a system dominated by two major political parties (a bipartisan system). In recent years, an independent party has emerged. However, since then the party has not grown significantly, as they have never held more than 2 seats at a time. The one minor party featured in this plot is the Iowa Democrats (an offshoot of the Democratic party). They have only ever held two seats in the US Senate but hold a good deal of power in the state. The party also has noteworthy involvement in the Iowa Caucuses, which are important for presidential elections.²



3.4 How often do multiple senators share a surname?

There are 12 surnames held by more than one Senator over time in the Congresses examined. Several Senators who share a surname are related, indicating the possibility of some degree of nepotism in the Senate. For instance, in Rhode Island, the Chafee family has held Senate seats for over three decades. John Chafee (R) held the seat from 1976-1999,³ and his son Lincoln Chafee (R) (who made a run for President in 2016 but has already dropped out) held the seat directly after him from 1999-2007.⁴ This suggests that some families have a high level of power in US governance (another instance is the Clinton Family).

Last Name	Brown	Burdick	Chafee	Dole	Graham	Johnson	Murkowski	Nelson	Pryor	Smith	Udall	Warner
# of Senators	3	2	2	2	2	2	2	2	2	2	2	2

²<http://iowademocrats.org/>

³<http://bioguide.congress.gov/scripts/biodisplay.pl?index=c000269>

⁴<http://bioguide.congress.gov/scripts/biodisplay.pl?index=C001040>

Over the course of the period examined, many states elected similar numbers of Senators from both major parties (shown in white). These are states in which senatorial elections may be more contentious. The states shown in red represent a Republican majority, and the states shown in blue represent a Democratic majority. Steadfast democratic states are primarily found in the Northeast, while states dominated by the Republican party are typically found in the South and Middle of America. The political, cultural, and economic histories of different regions of the United States account for the trends shown.

Republicans (Red) vs. Democrats (Blue)



Given that independent Senators are a small cohort, we decided that the independent senators deserved some analytical consideration. We note that the majority of elections won by independent Senators are in the state of Vermont (VT) and that there are a small overall number of candidates per number of elections won. Senator Barkley (I) had a small number of votes. Further research revealed that he temporarily held the chair of Senator Paul Wellstone, who died in a plane crash.⁵

	Name	State	Congress #	Total Votes
1	Joseph Lieberman	CT	112	487
2	Angus King	ME	113	657
3	Dean Barkley	MN	107	14
4	James Jeffords	VT	107	463
5	James Jeffords	VT	108	675
6	James Jeffords	VT	109	645
7	Bernard Sanders	VT	110	657
8	Bernard Sanders	VT	111	677
9	Bernard Sanders	VT	112	487
10	Bernard Sanders	VT	113	657

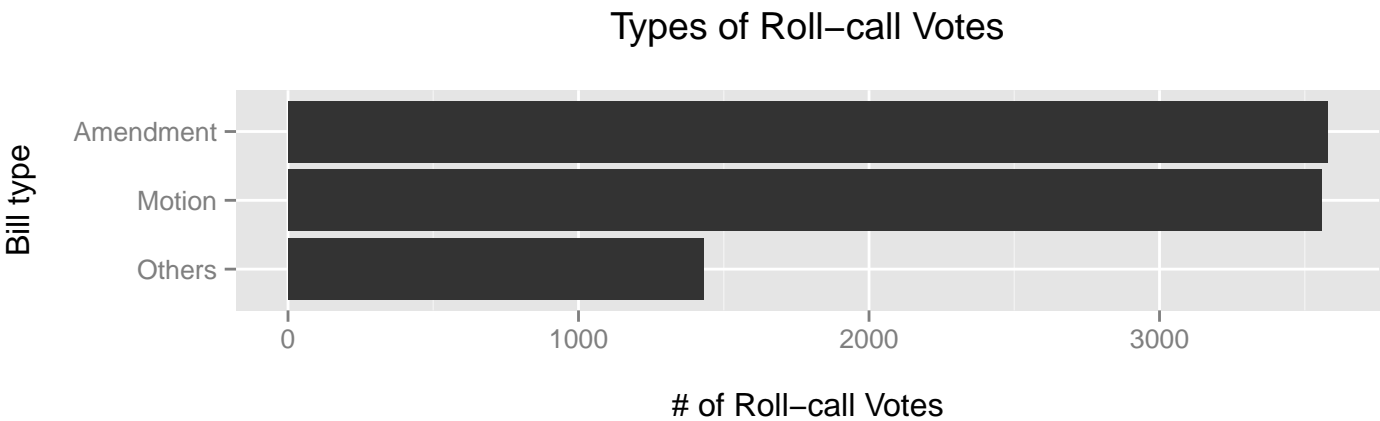
⁵<http://www.nytimes.com/2002/10/25/politics/25CND-CRASH.html>

4 Examining Patterns in Bill Passage

The United States Senate maintains an archive of roll-call votes - those votes in which the vote of each individual senator is recorded, rather than just having an overall tally - accessible via a series of folders on their servers containing XML files. We parsed these XML files using the code found in Appendix A

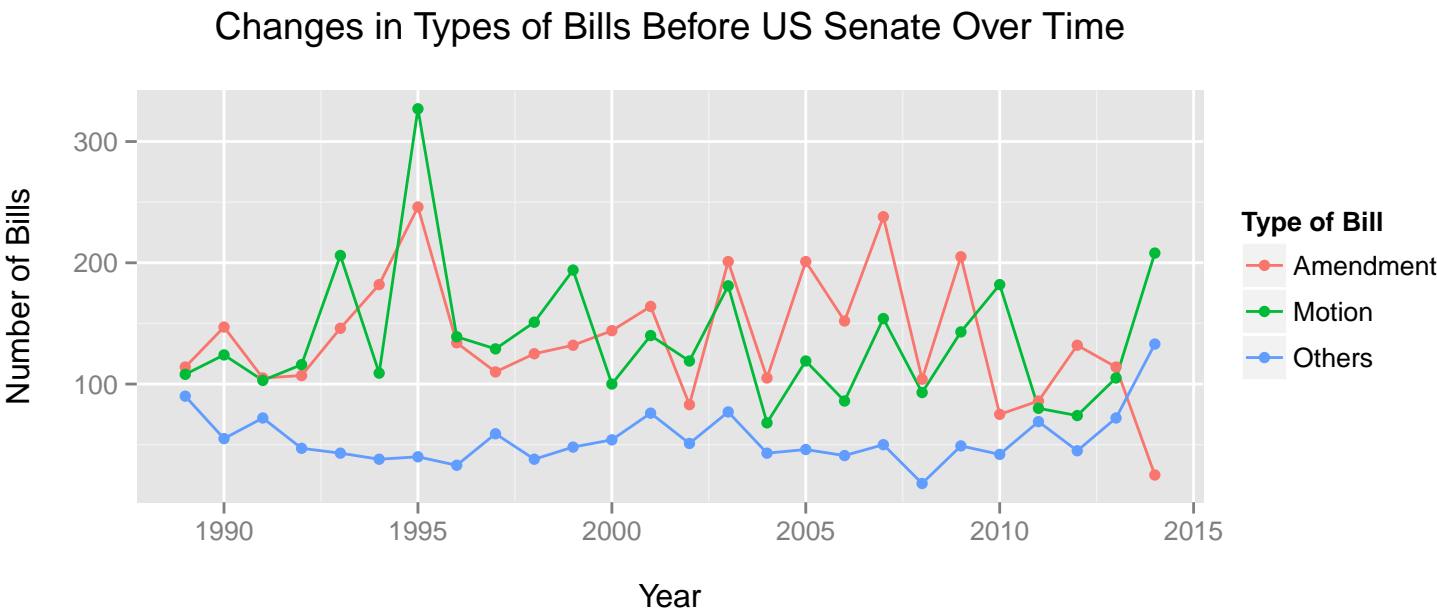
4.1 What types of roll-call votes are handled By US Senate?

Various bills are introduced to the Senate, but most fall into broad categories. After plotting counts of bills, we found that two categories of bills account for most bills in the Senate: one that initiates a motion to the table and one that is related to certain amendments. Both of these categories has a count of more than 3500. All other kinds of bills only count to about 1400 in total, so we synthesized them into one distinct category: Others.



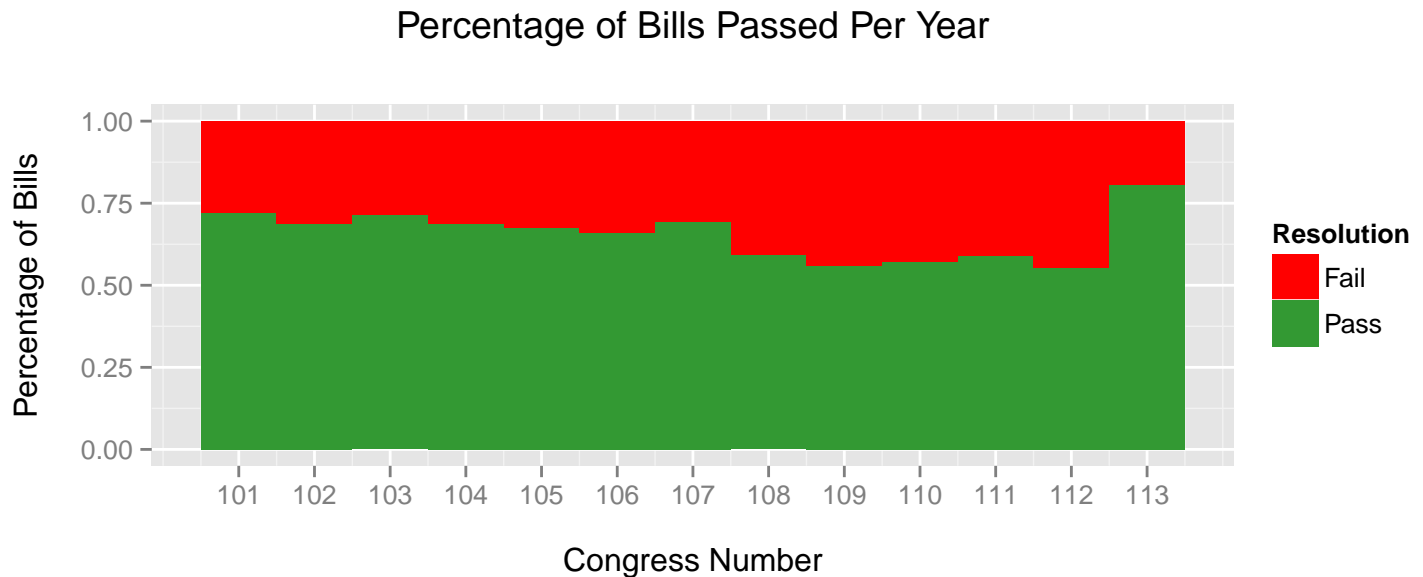
4.2 How have Trends in Bill Type Changed over Time?

After examining the overall distribution of bill types in Congresses 101-113, we investigated trends in the changing number of bills of the three bill types (see 4.1) over time. We expected that bill types distribution would be fairly stable over time, but Plot 7 shows the counts of the two majority categories (Amendment and Motion) are randomly oscillating, indicating the lack of a clear pattern. However, in the majority of years examined, between 100 and 200 bills were Amendment and Motion, while typically under 100 bills were in the category Other, which is consistent with bill type data over time. In addition, there are peaks for both Amendment and Motion in 1995, indicated that the Senate was particularly busy that year, perhaps as a consequence of the 1995 Republican Revolution (see Plot 10). Also, in 2014, the count of bills regarding amendments dropped significantly, while the count of other bills increased significantly, perhaps due to Obamacare taking effect in that year.



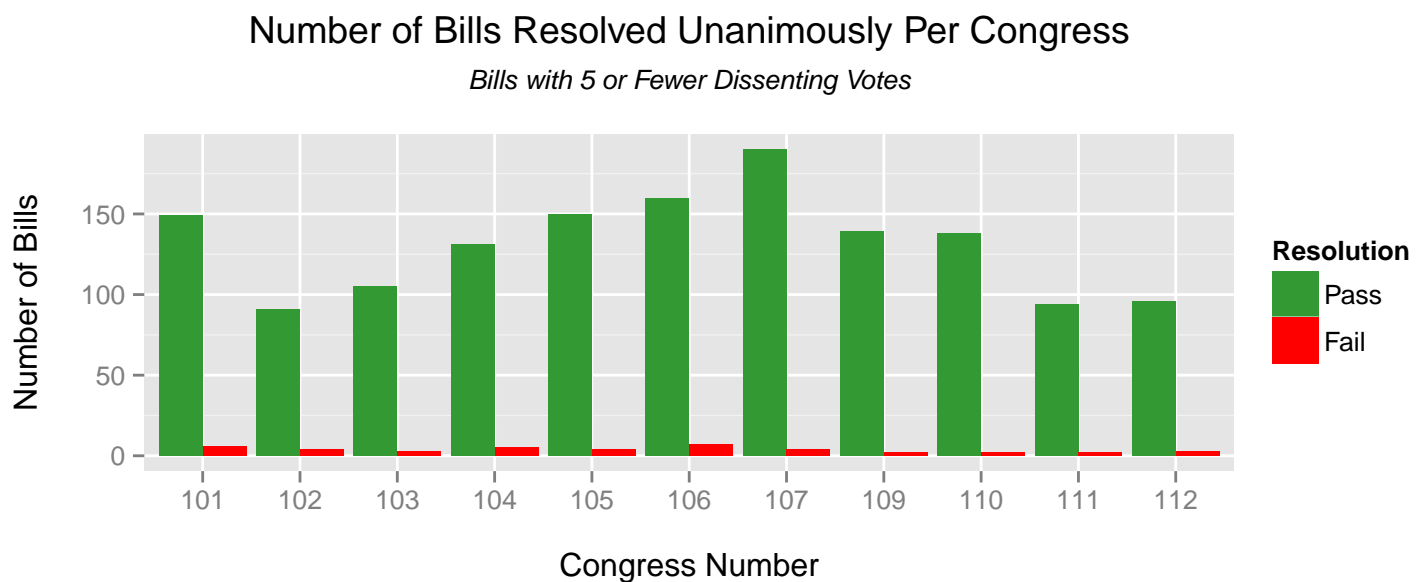
4.3 How does the percentage of passed bills change over time?

Most bills initiated in the Senate are either passed or rejected. One would think that the percentage of bills passed would be fairly stable over time but the statistics suggest otherwise: the percentage of passed Amendment bills consistently, but not significantly, decreased between Congress 107 and Congress 113; the percentage of passed Motion experienced a soar in Congress 113. The soar could possibly be explained as the effect of the launch of Obamacare. Moreover, in general more bills were passed than were rejected, especially for the Others category whose passing rate are constantly near 100 percent. The majority of passed bills may be explained by the fact that most bills are routine or do not have a serious impact (ex. changing the name of a national park).



4.4 Are bills that received unanimous votes typically passed or failed?

A unanimously resolved bill is a bill that is either passed with fewer than 5 vetoes or rejected with fewer than 5 approvals. This plot shows that there are significantly more unanimously passed bills than failed bills, which is consistent with the fact that the majority of bills proposed in each of Congresses 101 through 113 were passed (4.2). The number of unanimously resolved bills per year is surprisingly high, perhaps because of the bipartisan nature of many bills.

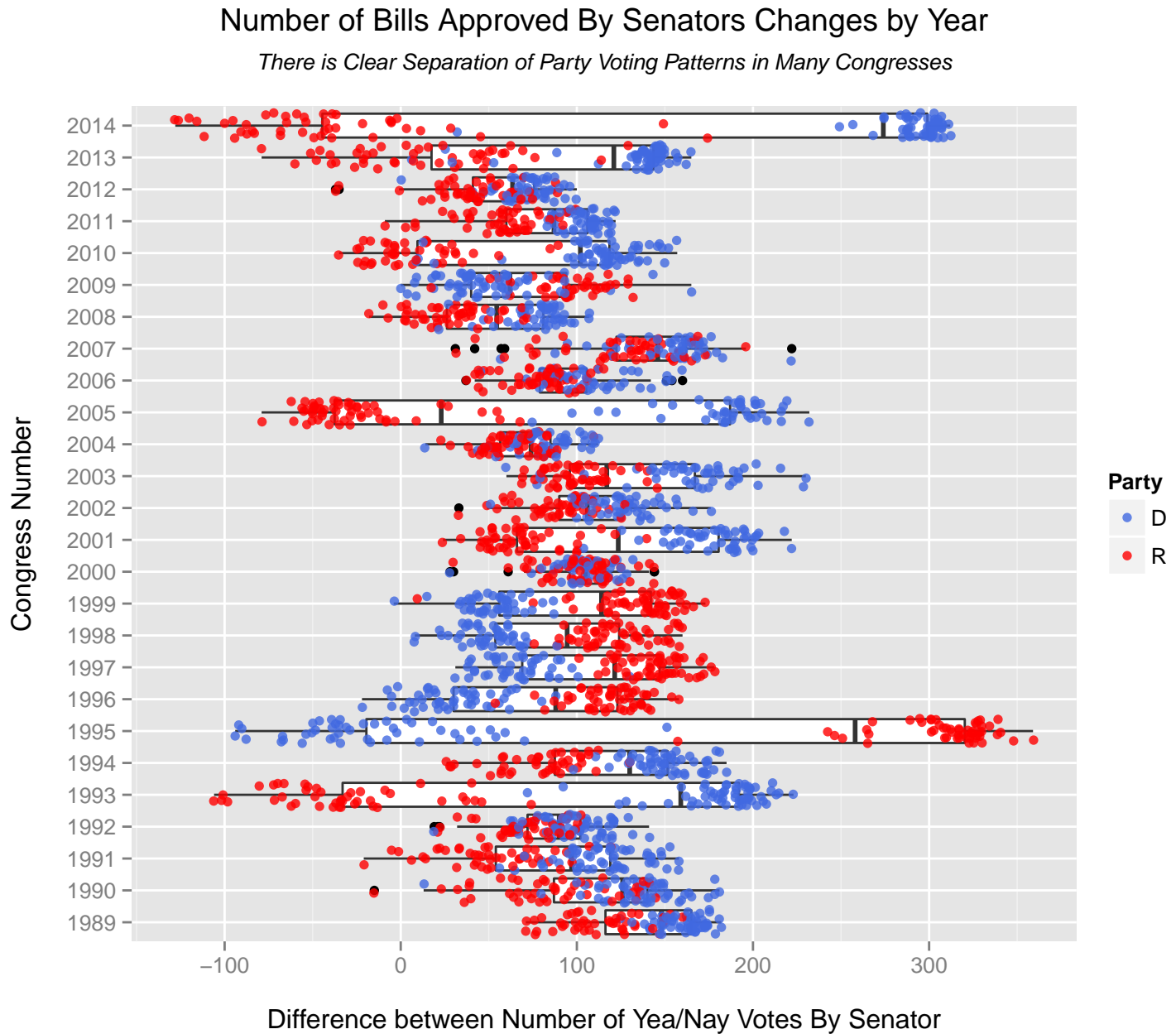


5 An Analysis of Voting Patterns Using Roll-Call Votes with Individual Senator Resolutions

In the previous section, we examined the overall trends in bill passage in the US Senate. For each of the roll-call votes we examined in our analysis, there is a corresponding list of each Senators vote: Yea or Nay. More content needed here

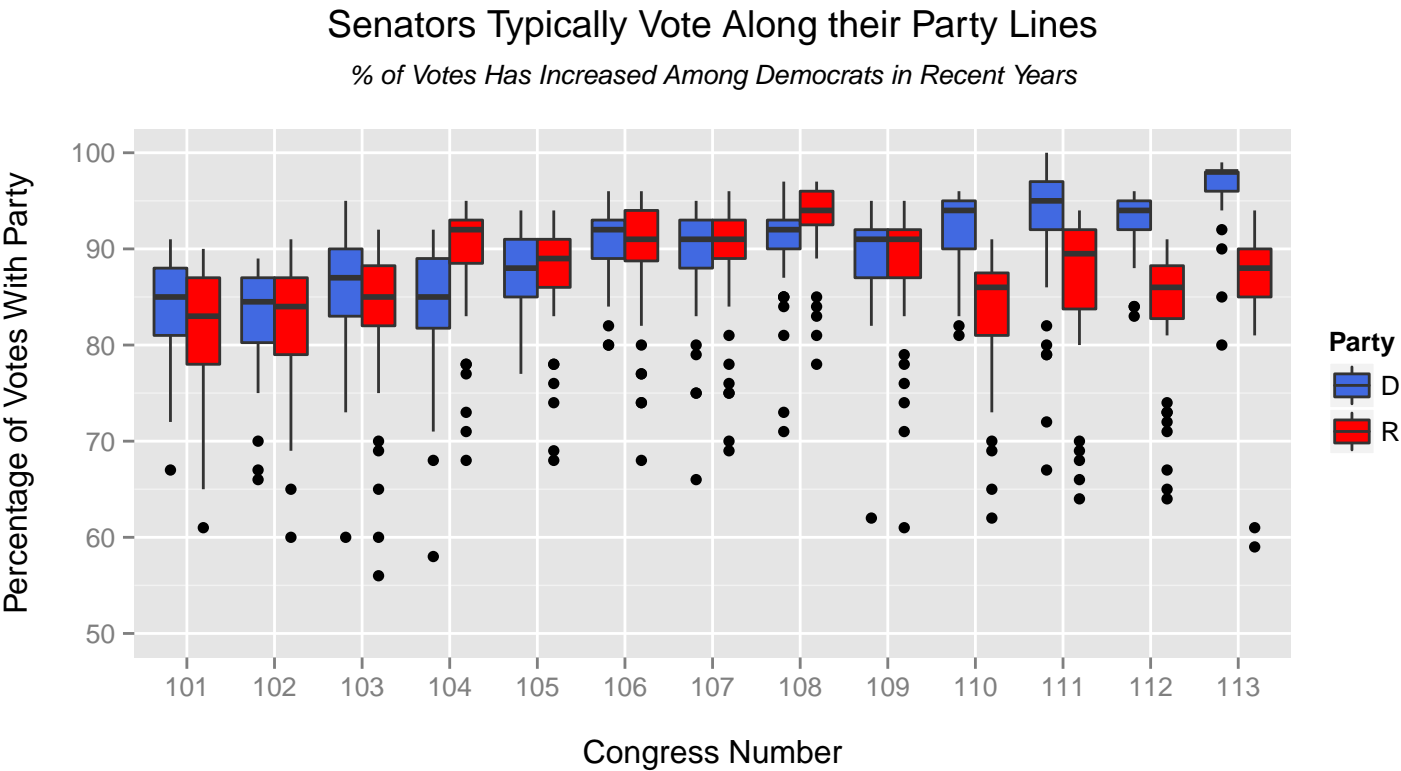
5.1 Do Senators Vote Yea/Nay more in a given year?

Content needed here.



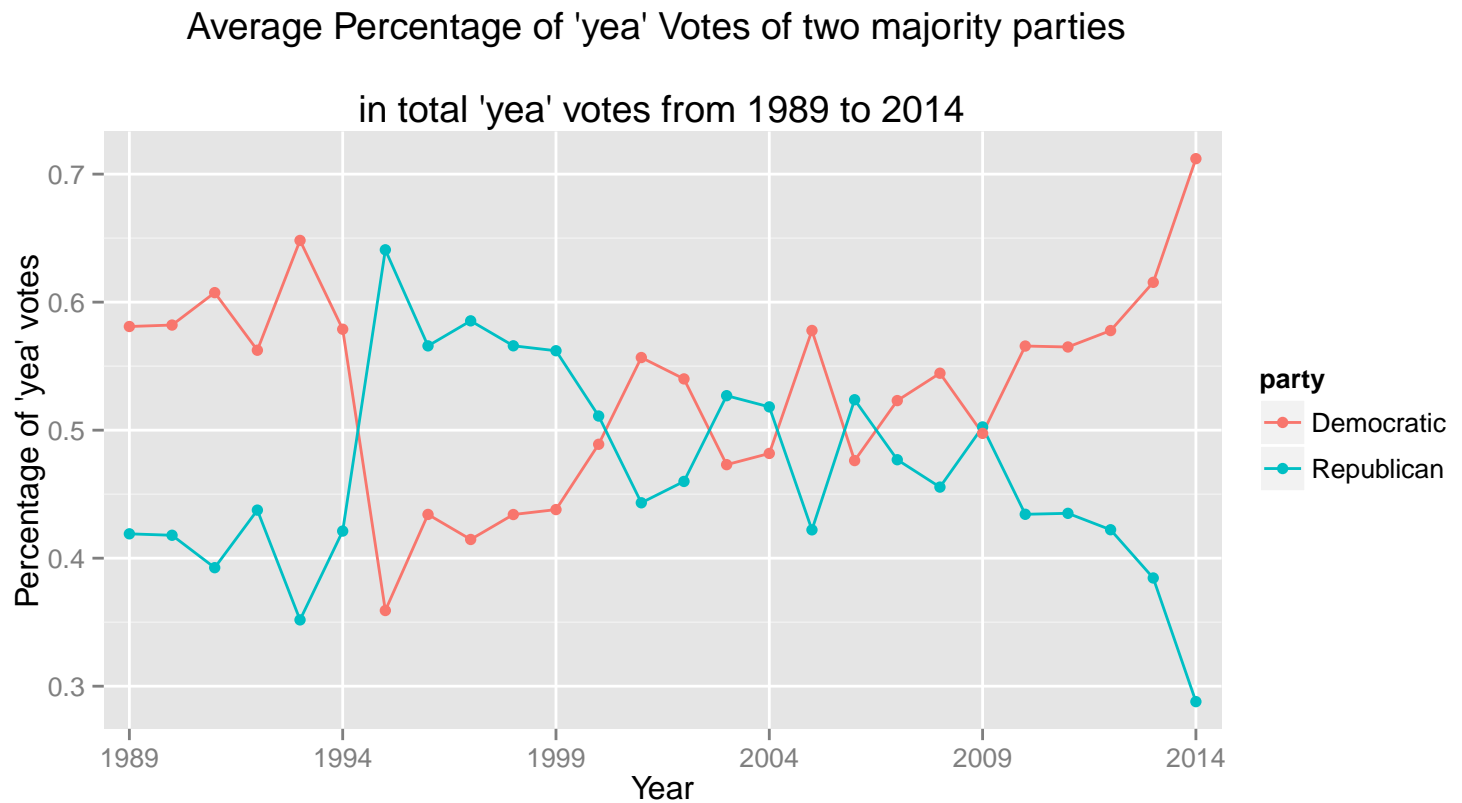
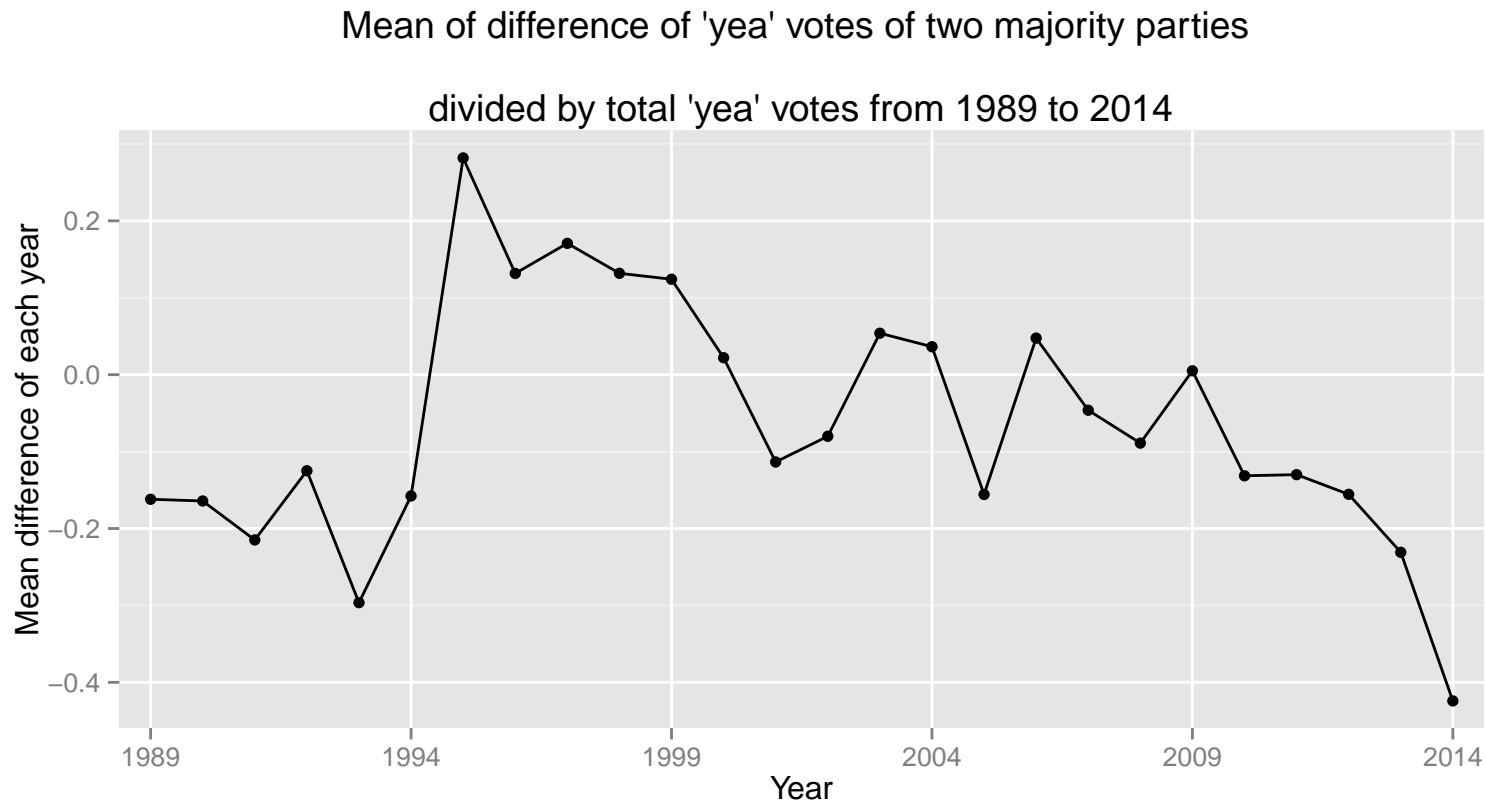
5.2 Do Senators Typically Vote Along Their Party Lines?

Content needed here.



5.3 How large is the disparity between yea votes by each major party?

Content needed here? Figure may need reworking.



6 Concluding Thoughts

Content needed here.

7 Appendix

7.1 Appendix A: Data fetching and Pre-processing

Three supplementary R files are included with this analysis. These contain the R code used to fetch and clean the data (divided by filetype) from our three sources:

[fetchJSON.R](#) - This script contains the functions required to download data from the New York Times Congress and gov-track.org roll-call vote APIs:

1. `fetchJSON` - takes a URL and returns a JSON formatted string.
2. `fetchCongressMembers` - takes a congress number and chamber; returns a cleaned Senate membership dataframe
3. `fetchAllMembers` - uses the above function to fetch senate members over a range of congresses; then assembles into one dataframe before exporting to SQLite.
4. `fetchVote` - Retrieves the roll-call record for a single vote from govtrack.org
5. `fetchVotes` - Takes a dataframe of roll-call votes and uses the above function to retrieve all votes. Saves to SQLite.

`fetchData.R` - This Script fetches all of the datasets and stores them in a SQLite database file (`data.sqlite`). The fetching and handling JSON files from