

A Retrospective Exploration of the US Senate through Data

Jordan Poles, Tony Chen and Aparna Narendrula

December 5, 2015

1 An Introduction to the Data

Our analysis covers a dataset which has been extracted from several different sources covering the timespan from the 101st to 113th US Congress. This corresponds to years of 1989 through 2014. The goal of our analysis was to assess trends in the members and bills passed by the US Senate. Our group found it most sensible to focus our efforts on 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.
 - (a) Each state is assigned 2 Senators, while the House of Representatives determines the number of representatives for a state based upon population.
 - (b) The Senate is smaller in size, 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 congress, using the [Congress API](#) provided by the NY Times. This dataset was made retrieved as JSON from a URI based API. These JSON files - as well as all of the other datafiles we fetched - were cleaned by an R script, 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 like percentage of votes with party and vote miss rate.

The member data we retrieved got us interested in examining the voting behaviour of the senators. We proceed with our analysis by examining the bills which were voted on in each session of the senate. We focus our analysis on roll-call bills - those bills where 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.

While the roll-call vote data from the senate tells us about passage rates for individual bills, it does not reveal individual voting behaviour. Indeed, this data was initially retrieved in the interest of compiling a complete list of roll-call votes for lookup on our third and final data-source, [govtrack.org](#). 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 use this information to examine each senators voting behavior individually, and in comparison to the corresponding parties or congresses.

2 The Importance of Assessing our Government

The US 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. It was for this reason that our group was 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 very straightforward. We hope that the publication of workflows such as this will make US Government 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.

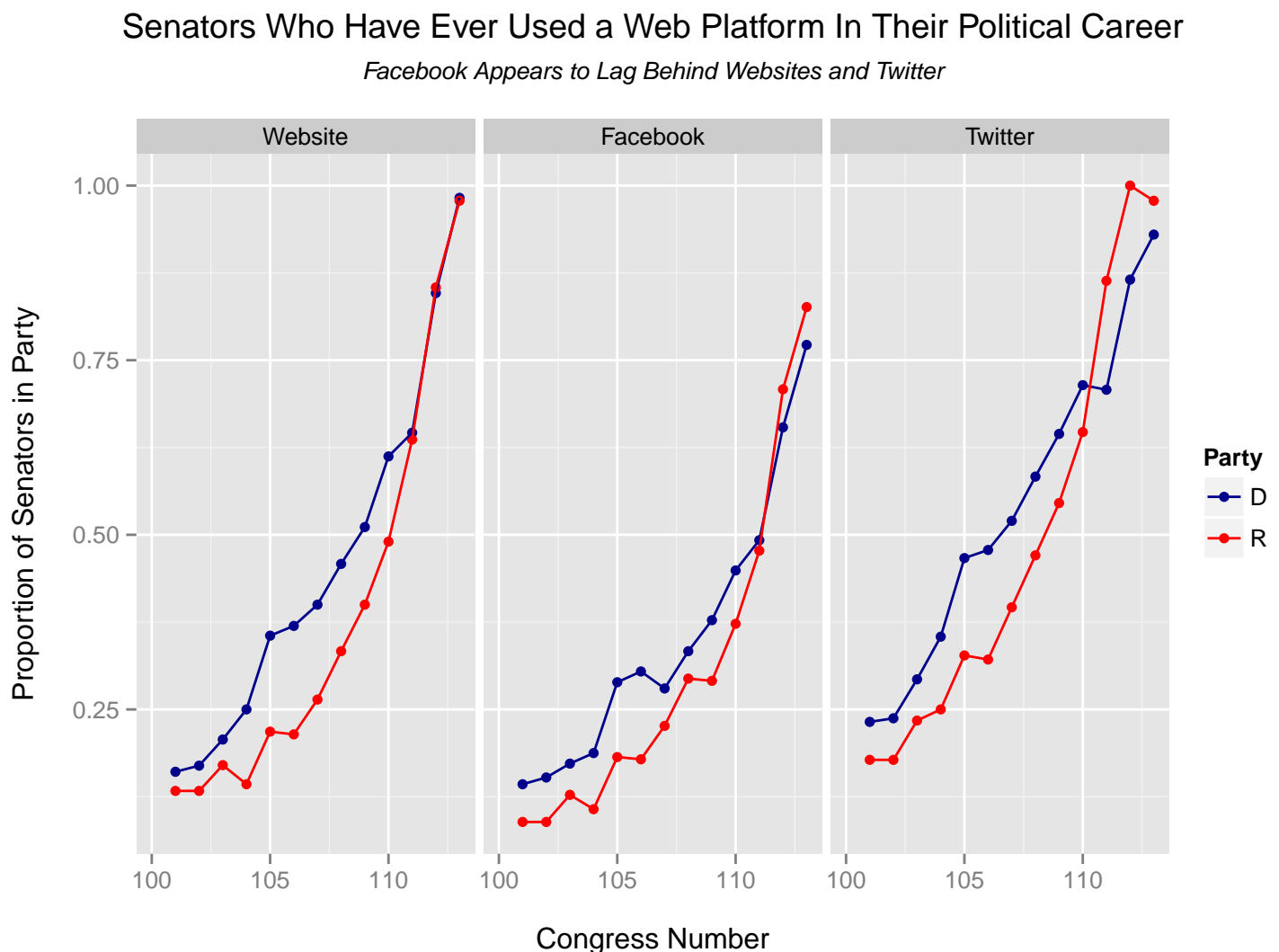
3 Examining The Members of Senate

The NY Times Congress API provides an interface for accessing a number of characteristics on US Senators including their web presence, party and certain voting patterns. This provides us some introductory insights into the 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. Of course, it was technology (the electronic cataloging and publication of Senate roll-call votes) which has made all of this analysis possible in the first place; allowing for greater government transparency.

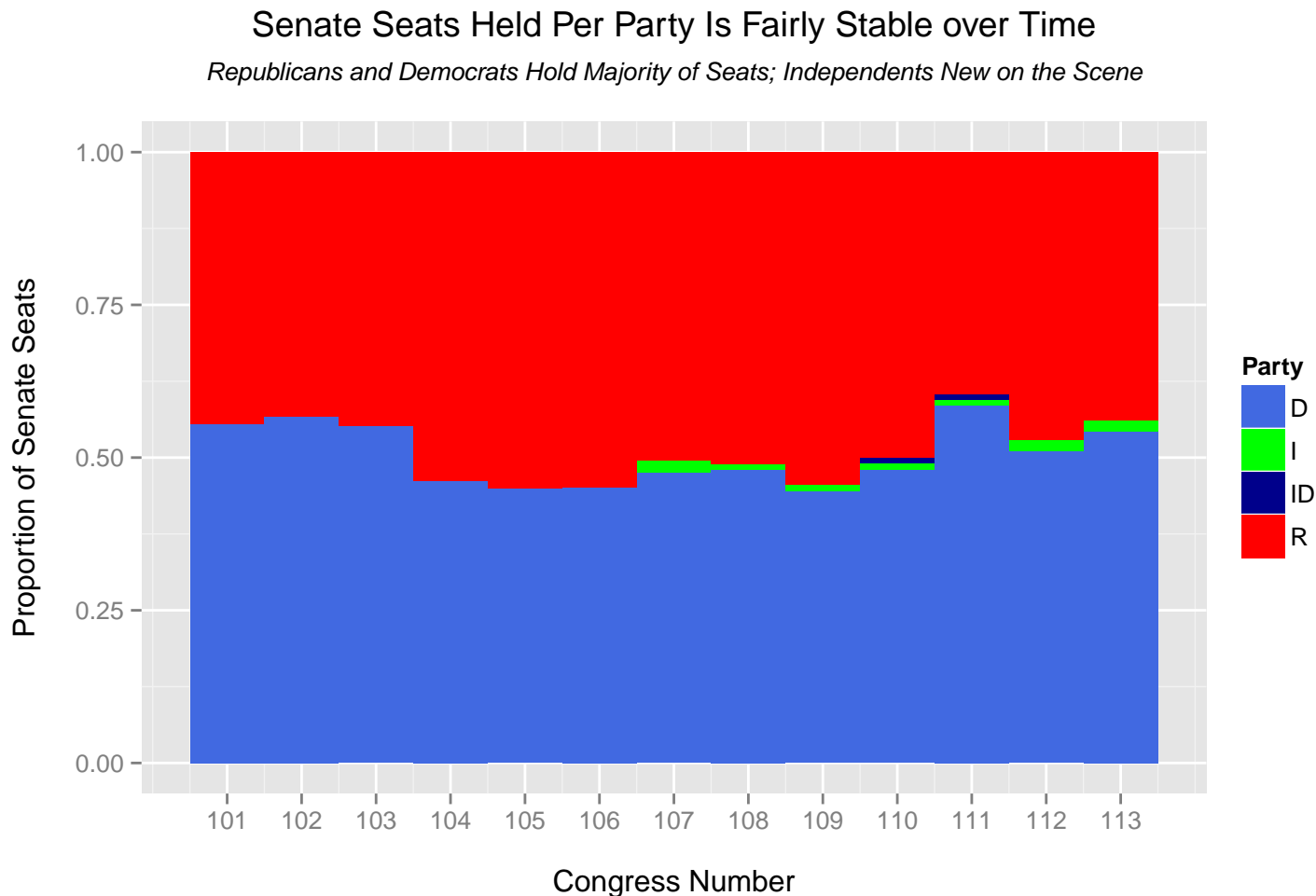
To begin our analysis, we turn to the use of web platforms by Senators. These platforms (including websites, Facebook, and Twitter) have become more common over time. Since the inception of FDR's fireside chats, electronic media has permitted mass communication of political thought; this capacity has been greatly expanded by the internet. In our plot we display lifetime usage of each platform. We show that usage has grown rapidly for all platforms, though Facebook has lagged behind. High web-platform usage is not surprising, given the number of senators that host their websites under senate.gov sub-domains (name.senate.gov). Twitter also seems to be popular, and is often utilized to disseminate political ideology and news.¹



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

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

The subsequent 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; they have never held more than 2 seats. There is one minor party featured in this plot, the Iowa Democrats, who are an offshoot of the Democratic party. They have only ever held two seats in the US Senate, but do hold a good deal of power in the state; the party also has a noteworthy involvement in the Iowa Caucuses, which are important for presidential elections².



3.3 How often do multiple senators share a surname?

We find that there are 12 surnames held by more than one senator; these are shown below. Several of these pairs of senators are relatives of one another, 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. Both Republicans, John Chafee held the seat from 1976-1999,³ and his son Lincoln Chafee (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 that comes to mind is the Clinton Family).

	1	2	3	4	5	6	7	8	9	10	11	12
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>

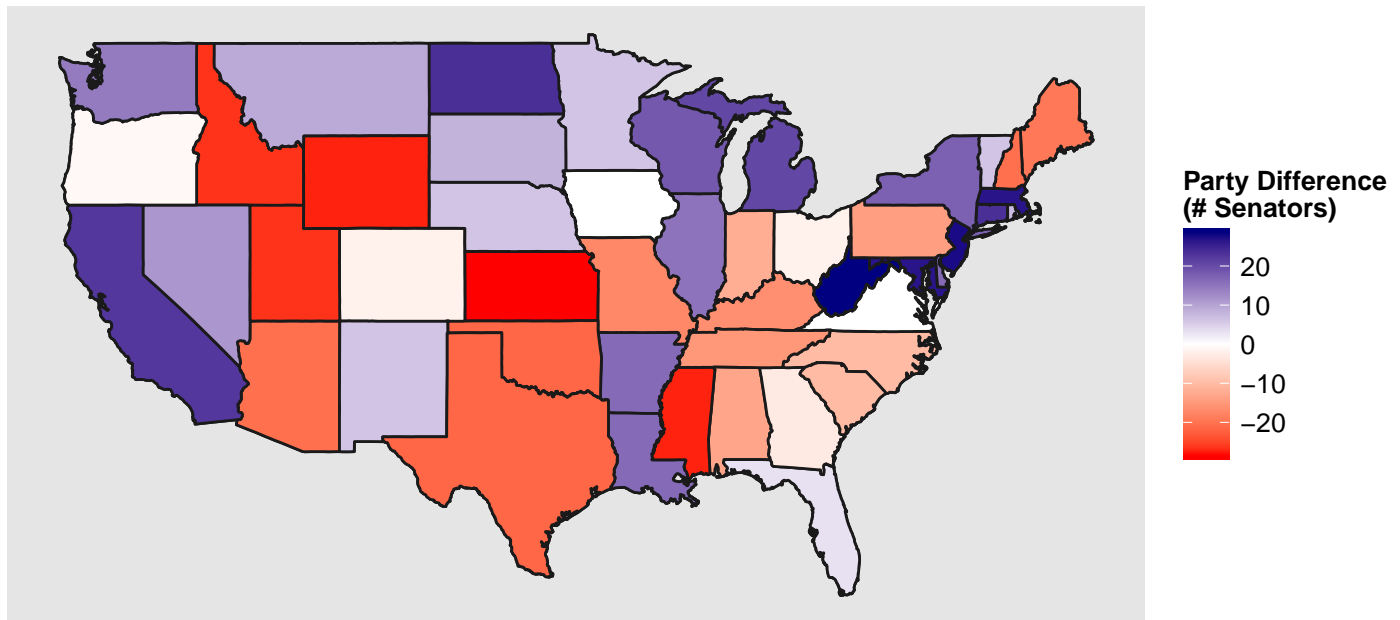
⁴<http://bioguide.congress.gov/scripts/biodisplay.pl?index=C001040>

3.4 Do certain states prefer certain political parties?

Over the course of the period examined, many states elected similar numbers of senators from both major parties (white); these are states in which senatorial elections may be more contentious. Those states shown in red represent a Republican majority; blue a Democratic majority. The steadfast democratic states are primarily found in the Northeast, while several states in the South and Middle of America are dominated by the Republican party.

Party Preferences By State as Determined by Senate Seats

Republicans (Red) vs. Democrats (Blue)



3.5 Who are the independent senators? Where are they elected?

Given that they are a small cohort, we thought 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 had a small number of votes; further research revealed that he temporarily held the chair of another senator who had died in a plane crash (Paul Wellstone)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

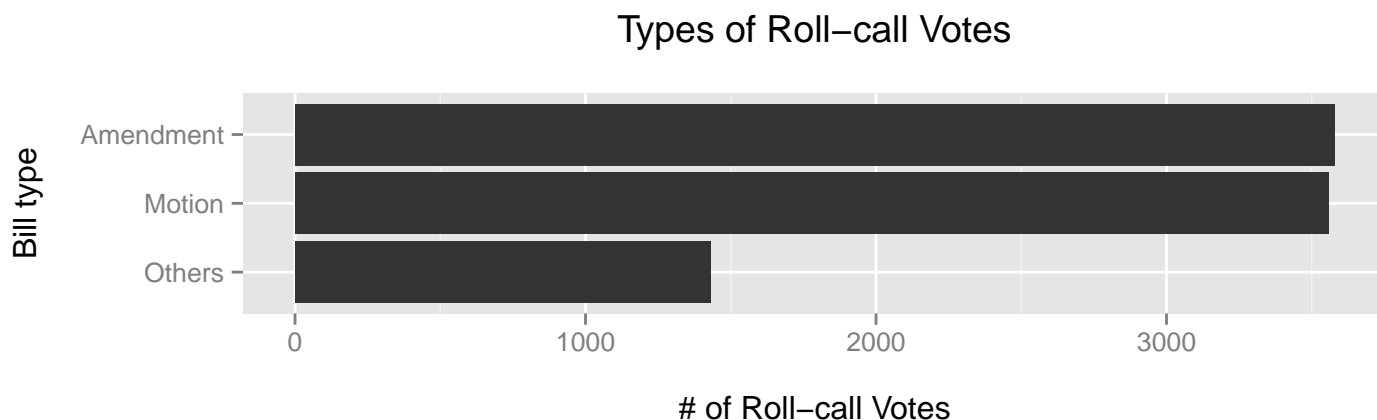
⁵<http://www.nytimes.com/2002/10/25/politics/25CND-CRASH.html>

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

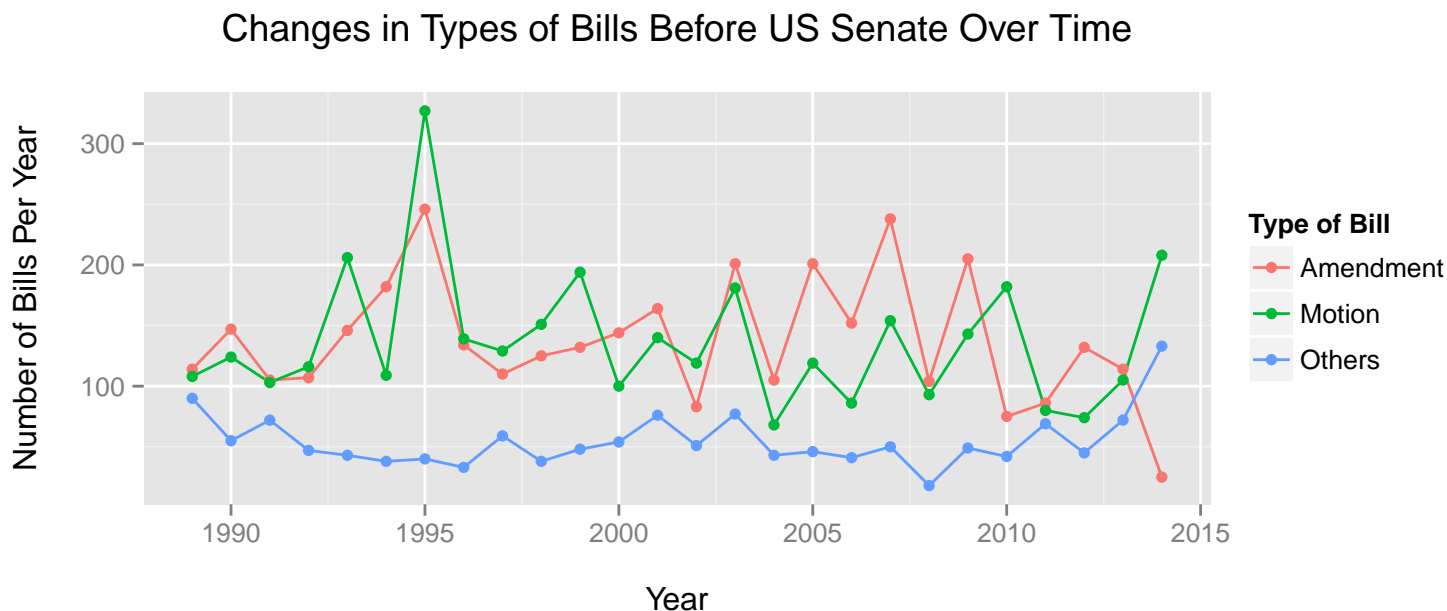
4.1 What types of roll-call votes are handled By US Senate?

There are all kinds of bills introduced to the Senate, but most of them fall into a couple of broad categories. After plotting counts of bills out, we notice that the two majority categories are the one that initiate a motion to the table and the one that is related to certain amendments, each of which has a count of more than 3500; while all other kinds of bills only count to about 1400 in total, so we consider it is appropriate to synthesize them into one distinct category: "Others".



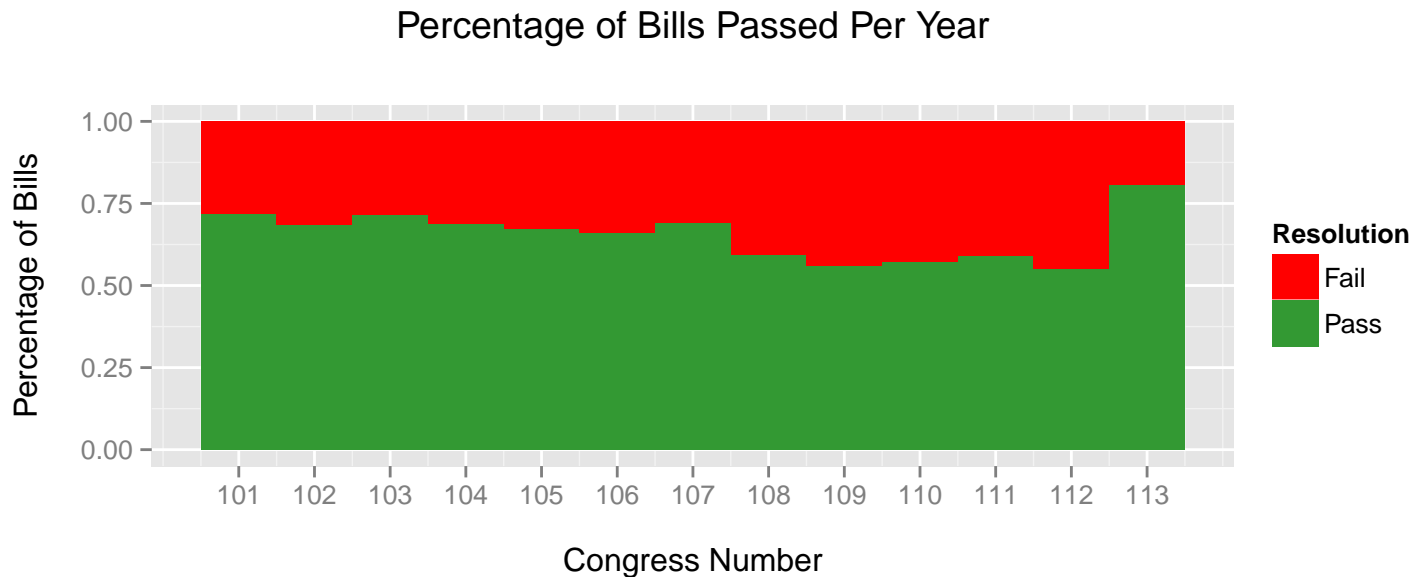
4.2 Changes of bill types over year

After examining the all-time distribution of bill types, we then tried to investigate if there is any trend in the changing of bill types. We expect that bill types distribution should be fairly stable over time, but it turns out that the counts of the two majority categories ("Amendment" and "Motion") are randomly oscillating, indicating the lack of a clear pattern. However, there are still some noteworthy points. First, There are peaks for both "Amendment" and "Motion" in 1995, so the senate was really busy in that year. This phenomenon might be related to the 1995 Republican Revolution. Also, in the most recent year, 2014, the count of bills regarding amendments has dropped significantly, while the count of other bills increased significantly. This could be due to Obamacare taking effect in that year.



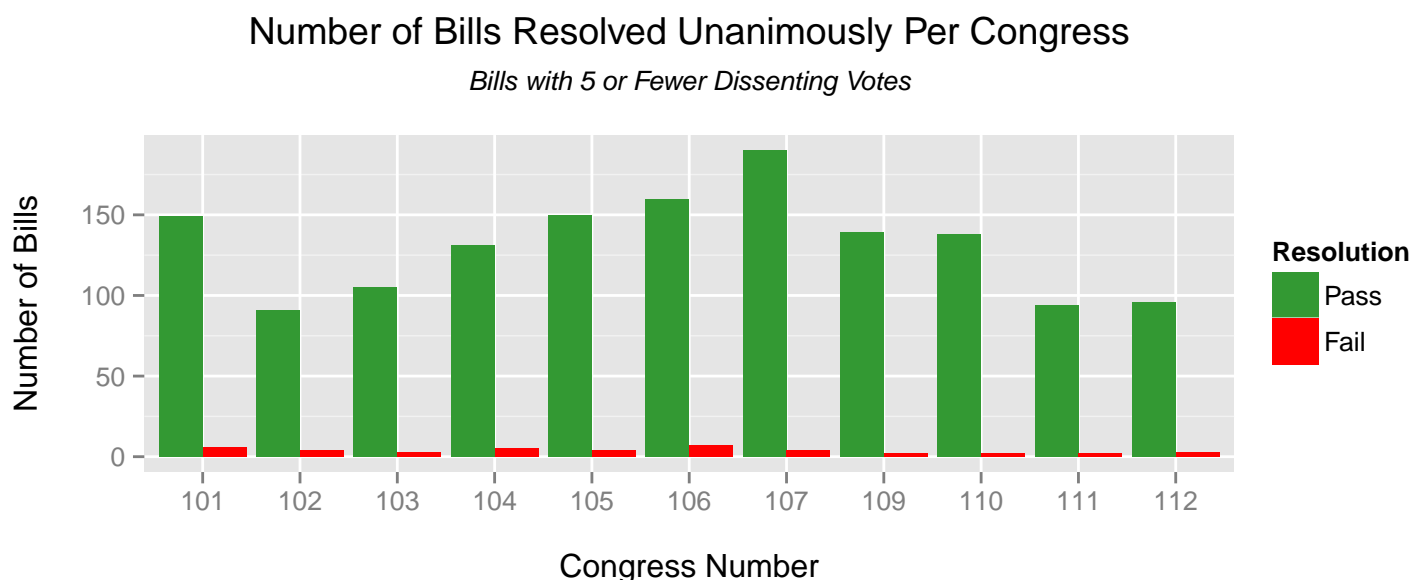
4.3 How does the percentage of passed bills change over time?

Most bills initiated at the Senate are either passed or rejected. Since its reasonable to think that the percentage of bills got passed should be fairly stable over time and that roughly half of all bills should be passed while the other half negated, its really surprising to see that statistics suggest otherwise: the percentage of passed "Amendment" bills was constantly, but not significantly, dropping from the congress 107 through 113, and the percentage of passed "Motion" experienced a soar during congress 113. The soar could possibly be explained as the effect of the launch of Obamacare. Moreover, there are generally more bills got passed than bills got 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?

Our definition of an unanimously resolved bill is a bill that is either passed with less than 5 vetoes or rejected with less than 5 approvals. The graph is surprising in two ways. First, there are significantly more unanimously passed bills than failed bills; second, there are a lot of unanimously resolved bills each year, while we previously assumed that amount should be pretty low. One reason that may explain both strangenesses is that there are many bills whose passage will hurt no one, and they are likely the ones that are neither related to money nor certain political parties.



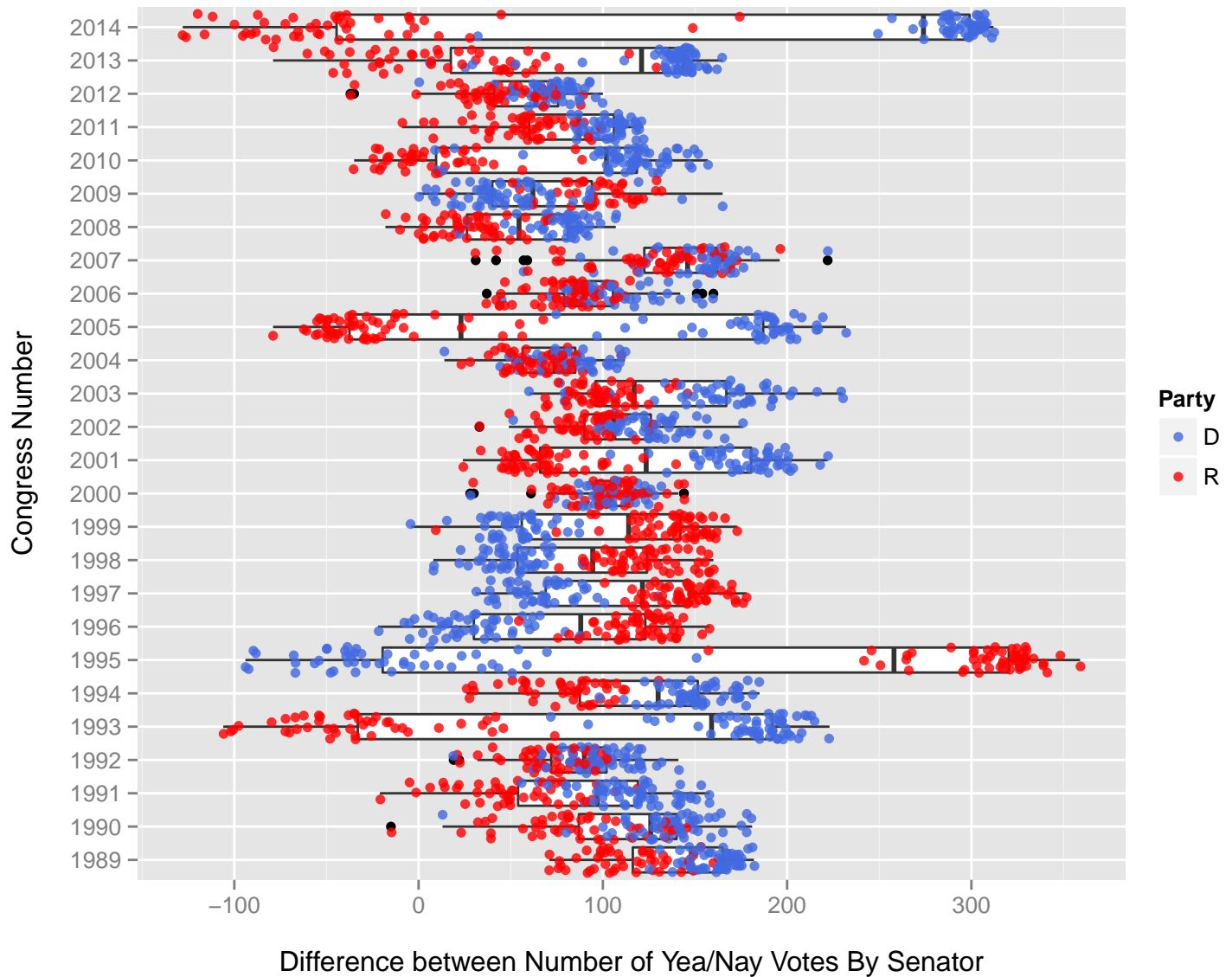
5 An Analysis of Voting Patterns Using Roll-Call Votes with Individual Senator Resolutions

In the previous section, we examined the overall changes in bill passage in the US Senate. For each of the roll-call votes we examine in our analysis, there is a corresponding list of each Senator's vote: Yea or Nay. We

```
## Warning: Removed 3 rows containing non-finite values (stat.boxplot).  
## Warning: Removed 3 rows containing missing values (geom_point).
```

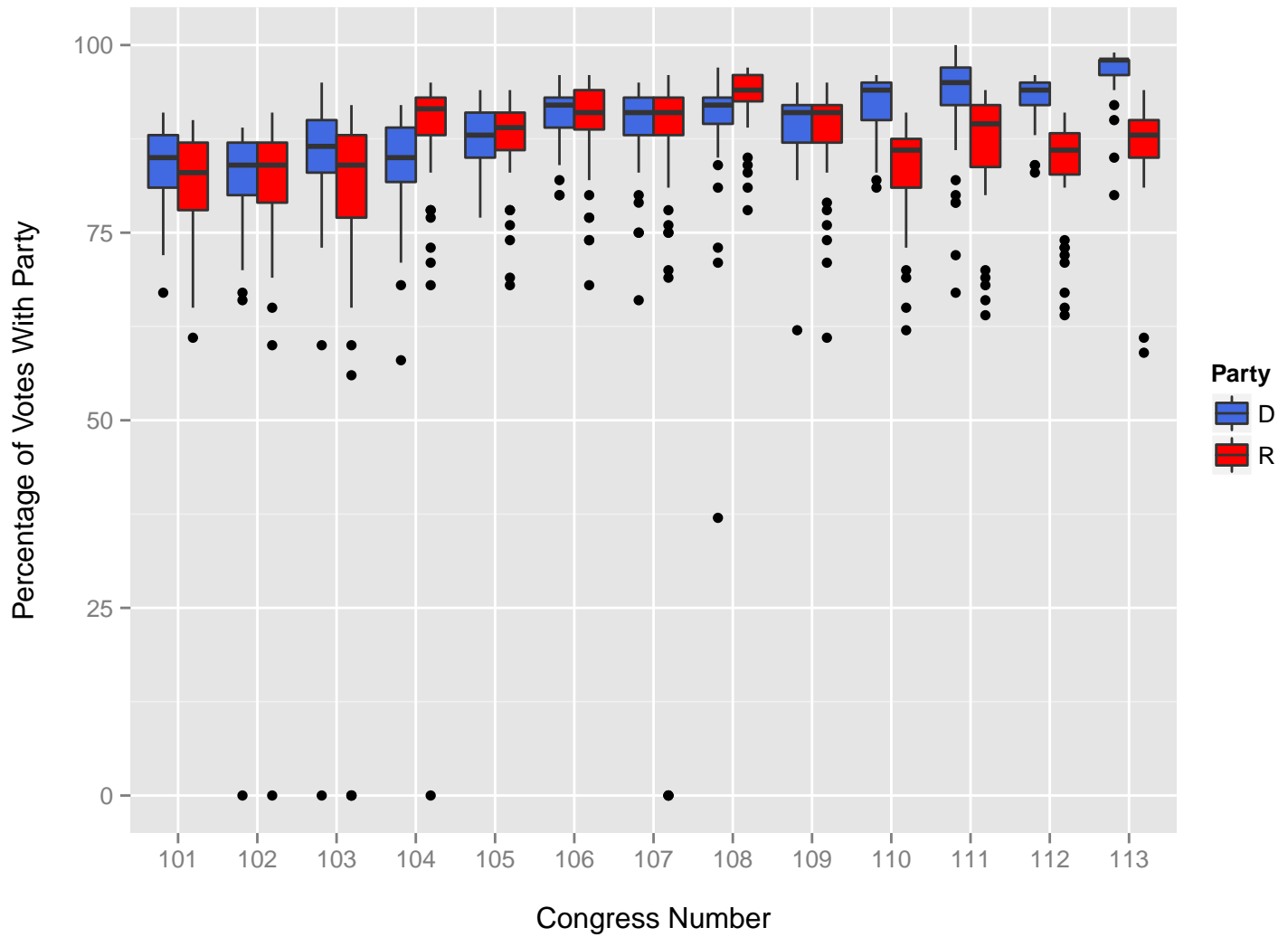
Number of Bills Approved By Senators Changes by Year

There is Clear Separation of Party Voting Patterns in Many Congresses

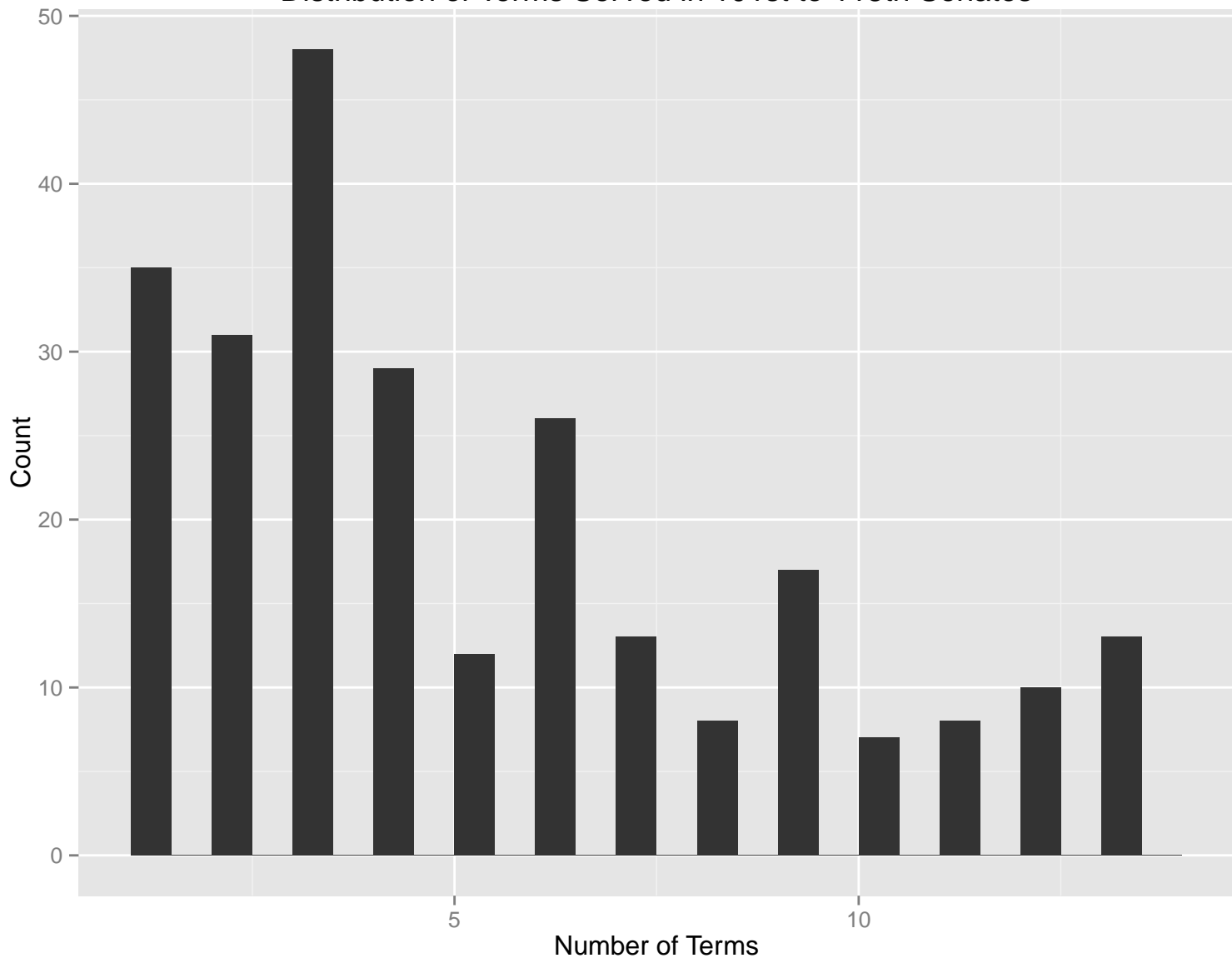


Senators Typically Vote Along their Party Lines

% of Votes Has Increased Among Democrats in Recent Years



Distribution of Terms Served in 101st to 113th Senates



```

query = "select r.year as year, r.voteNumber as voteNumber,
      abs(r.c - d.c) * 1.0 / (r.c + d.c) as diff
      from (select voteNumber, year, count(*) as c
            from votes
            where vote == 'Yea' and party == 'R' group by year, voteNumber)
      as r
      join
      (select voteNumber, year, count(*) as c
      from votes
      where vote == 'Yea' and party == 'D' group by year, voteNumber)
      as d
      on r.voteNumber == d.voteNumber and r.year == d.year"
yeaDiff = queryDB(query, "data.sqlite")
yeaDiffMean = setNames(aggregate(diff ~ year, yeaDiff, mean), c("year", "mean"))
yeaDiffSd = setNames(aggregate(diff ~ year, yeaDiff, sd), c("year", "std"))
yeaDiffDistribution = merge(yeaDiffMean, yeaDiffSd, by="year")
ggplot(yeaDiffDistribution) +
  aes(x = year, y = mean) +
  scale_x_discrete(breaks=c(1989,1994,1999,2004,2009,2014)) +
  labs(title="Mean of difference of 'yea' votes of two majority parties
        \ndivided by total 'yea' votes from 1989 to 2014") +
  xlab("Year") +
  ylab("Mean difference of each year") +
  geom_bar(stat="identity")

```

Mean of difference of 'yea' votes of two majority parties

divided by total 'yea' votes from 1989 to 2014

