A Retrospective Exploration of the US Senate through Data

Jordan Poles, Tony Chen, and Aparna Narendrula

December 7, 2015

1 An Introduction to the Data

This report details our analysis of a dataset on politics that covers the 101st to the 113th US Congresses (1989 - 2014). The dataset was extracted from several sources, and allowed us 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 two Senators, while the number of representatives for a state in the House of Representatives is based on the state's 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 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 as well as the bills they passed. 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 behavior. 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.

1.1 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 behaviors. 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.

2 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.

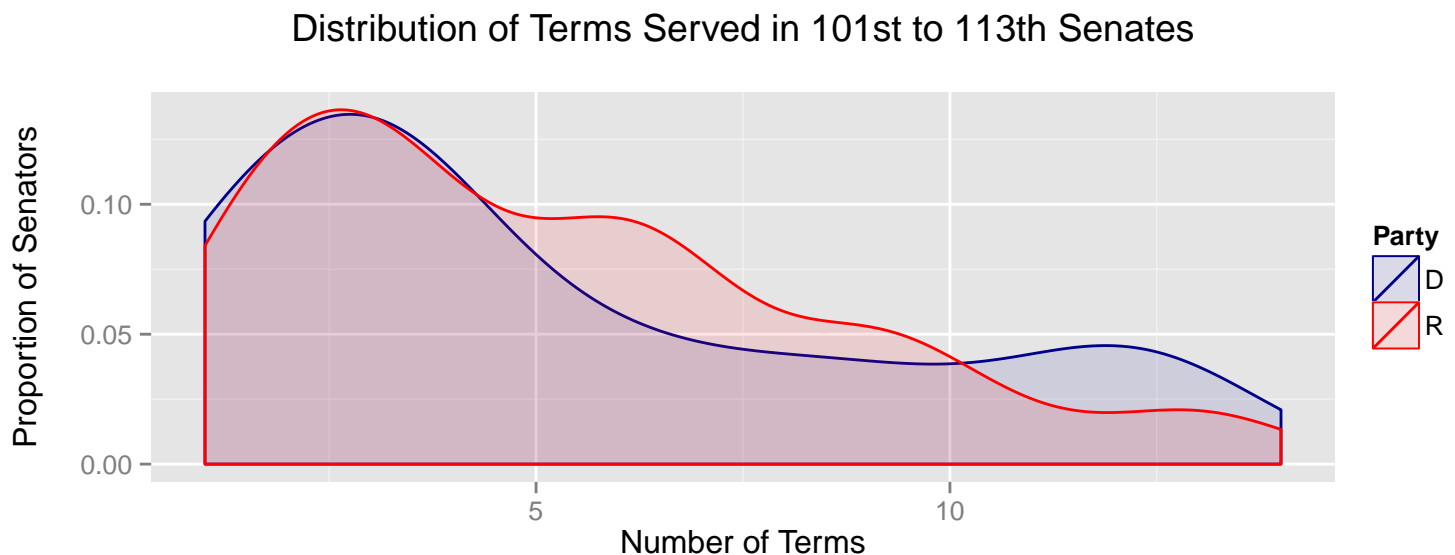## 2.1 How Have Senators Adopted Web Platforms Over Time?

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 FDRs 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
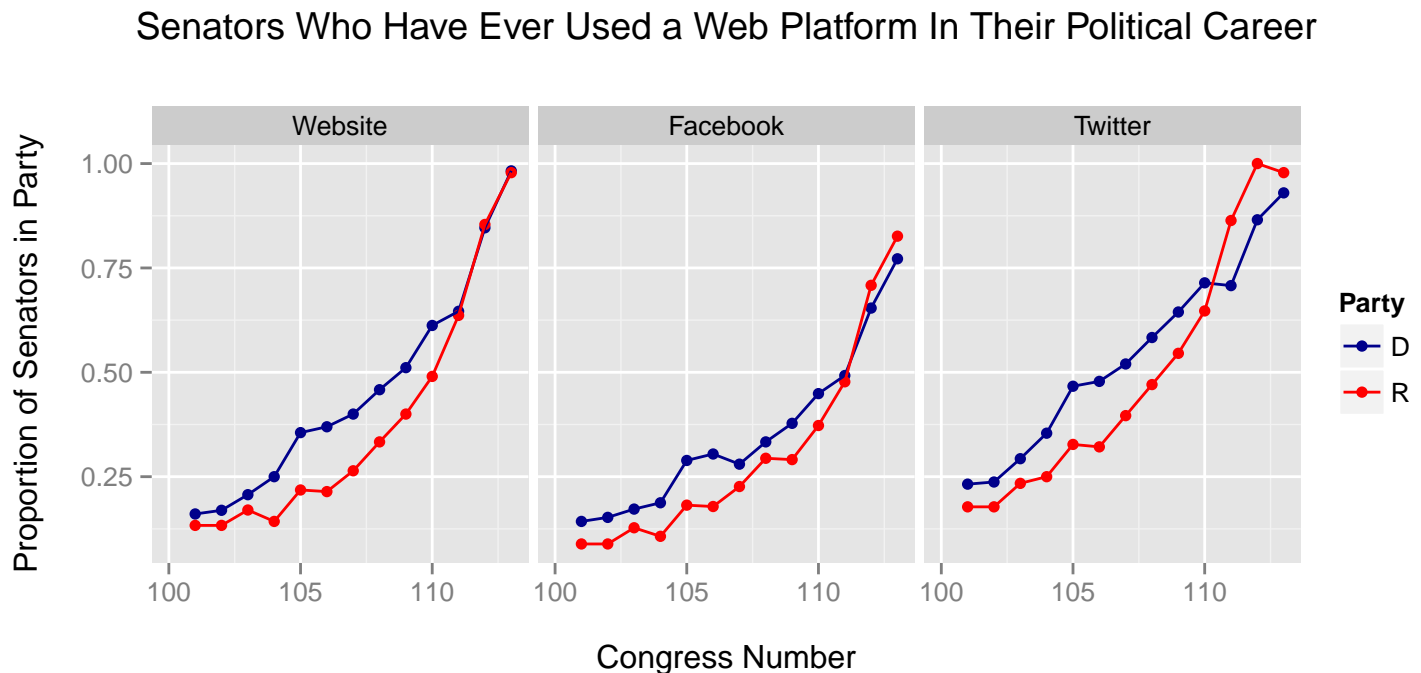
2.2 How many Terms do Senators Typically Serve?

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

---

[1]For instance: https://twitter.com/McConnellPress

2.3 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]

Senate Seats Held Per Party Is Fairly Stable over Time
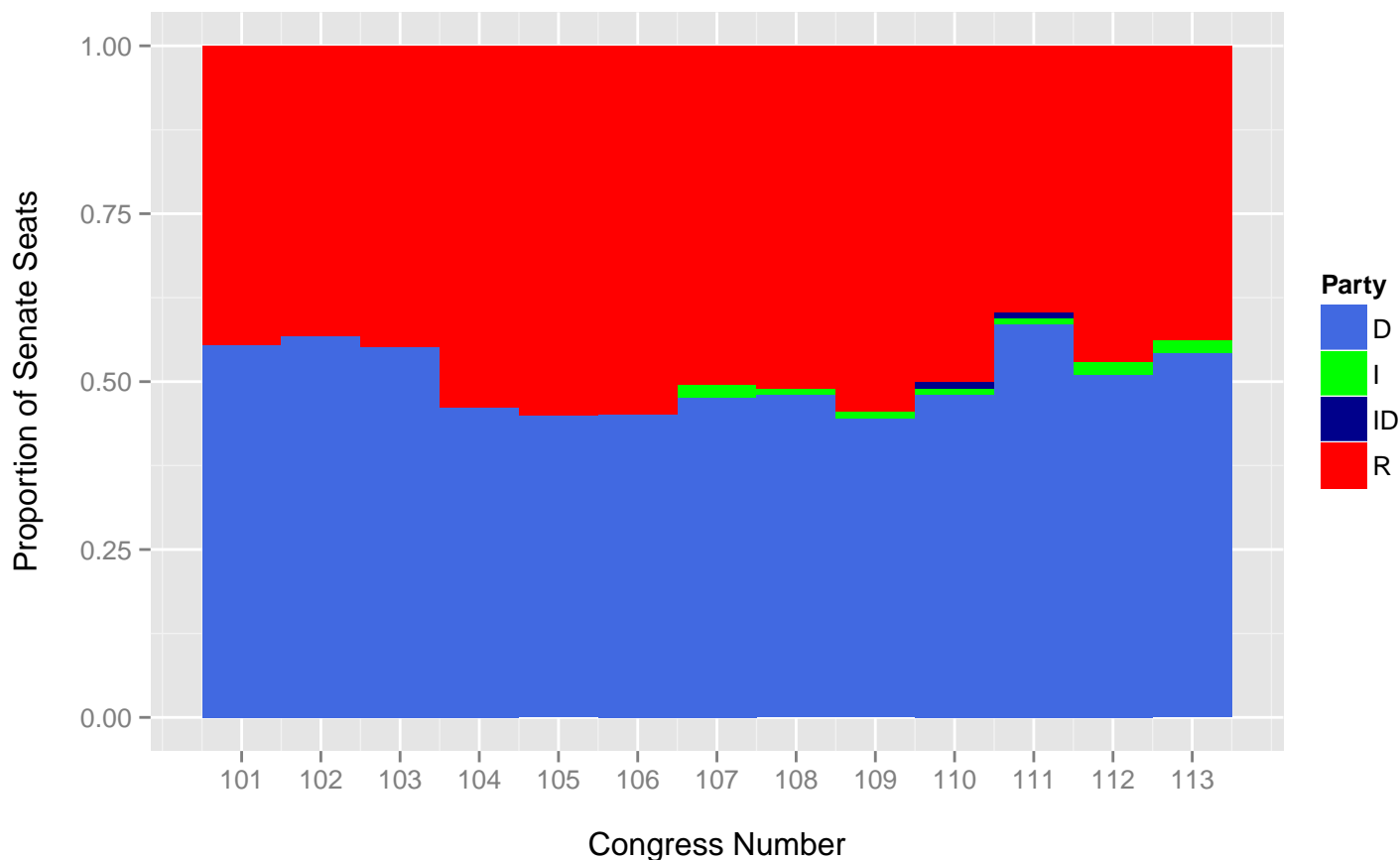
2.4 How Often do Multiple Senators Share a Surname?

There are 12 surnames held by more than one Senator over course of 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).

Number of Senators who Share the Same Last Name

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

2.5   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 (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.

Party Preferences By State as Determined by Senate Seats

Republicans (Red) vs. Democrats (Blue)
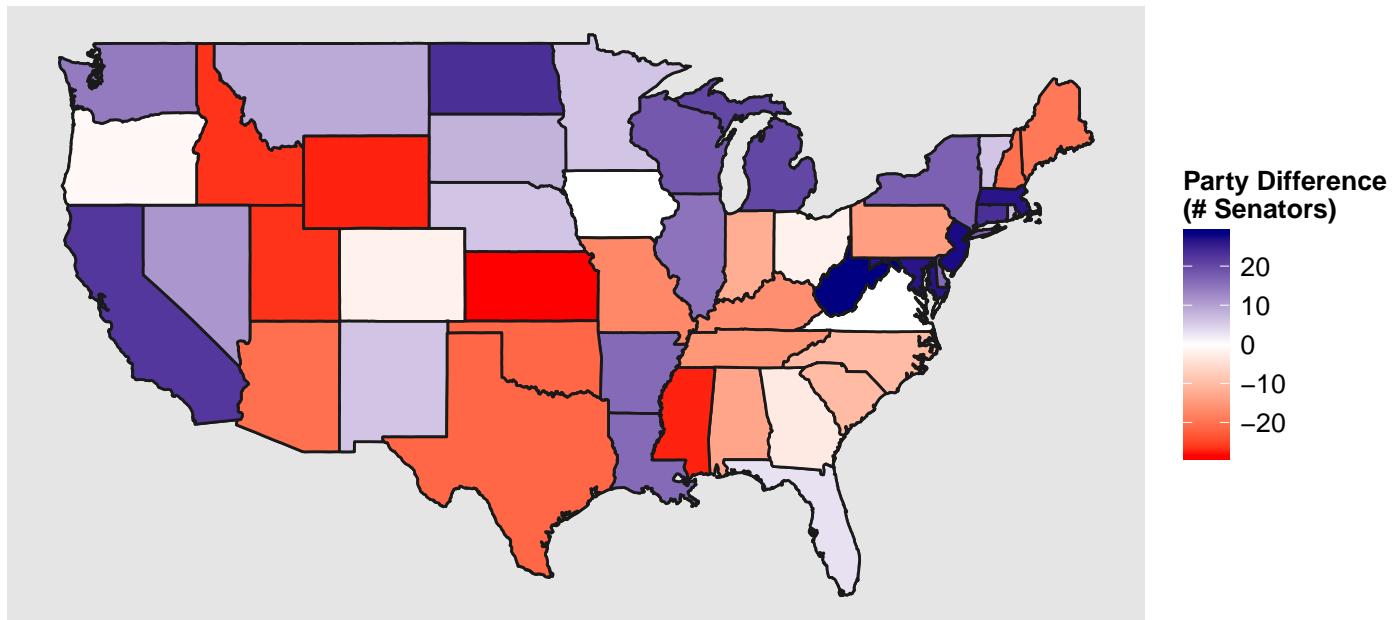
2.6   Who are the Independent Senators? Where are They Elected?

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]

Independents of the 101st to 113th Senates

Name	State	Congress #	Total Votes
Joseph Lieberman	CT	112	487
Angus King	ME	113	657
Dean Barkley	MN	107	14
James Jeffords	VT	107	463
James Jeffords	VT	108	675
James Jeffords	VT	109	645
Bernard Sanders	VT	110	657
Bernard Sanders	VT	111	677
Bernard Sanders	VT	112	487
Bernard Sanders	VT	113	657

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

3   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. This archive is accessible via a series of folders on their servers containing XML files. We parsed these XML files using the code found in Appendix A in order to gain insights into the types of bills reviewed in the US Senate, and the patterns of bill passage.

3.1   What Types of Roll-Call Votes are Handled by the 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: motions, where a vote is called in order to direct the proceedings of the Senate (end discussion on a topic or consider a bill), and ammendements in which the senators vote on additions/removals of text from bills. 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.
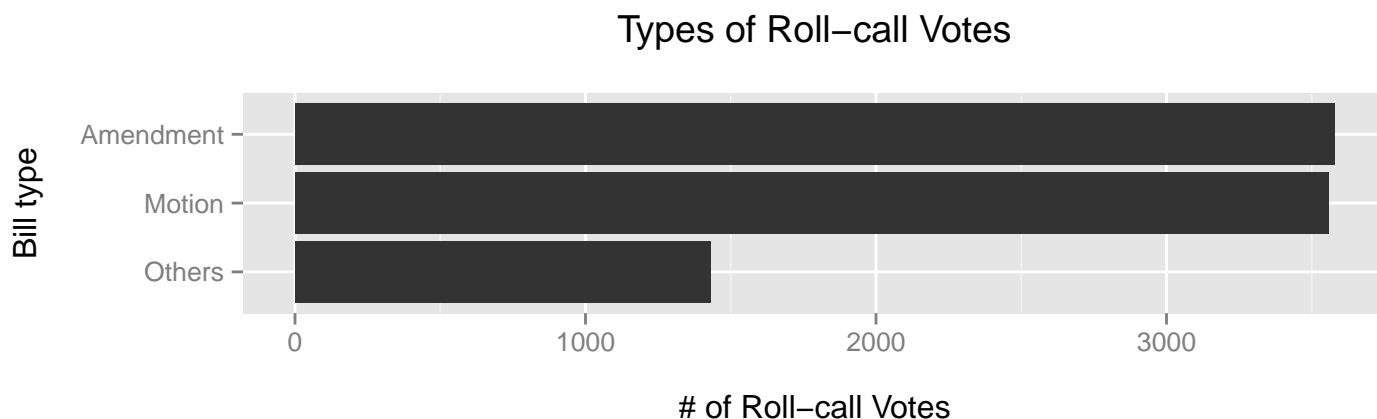

3.2   How Have Trends in Bill Type Changed over Time?

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 this 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 our previous plot. In addition, there are peaks for both Amendment and Motion in 1995, indicating that the Senate was particularly busy that year, perhaps as a consequence of the 1995 Republican Revolution (see 5.2). Also, in 2014, the count of bills regarding amendments dropped significantly, while the count of other bills increased significantly, perhaps due to Obamacare debates during that year.
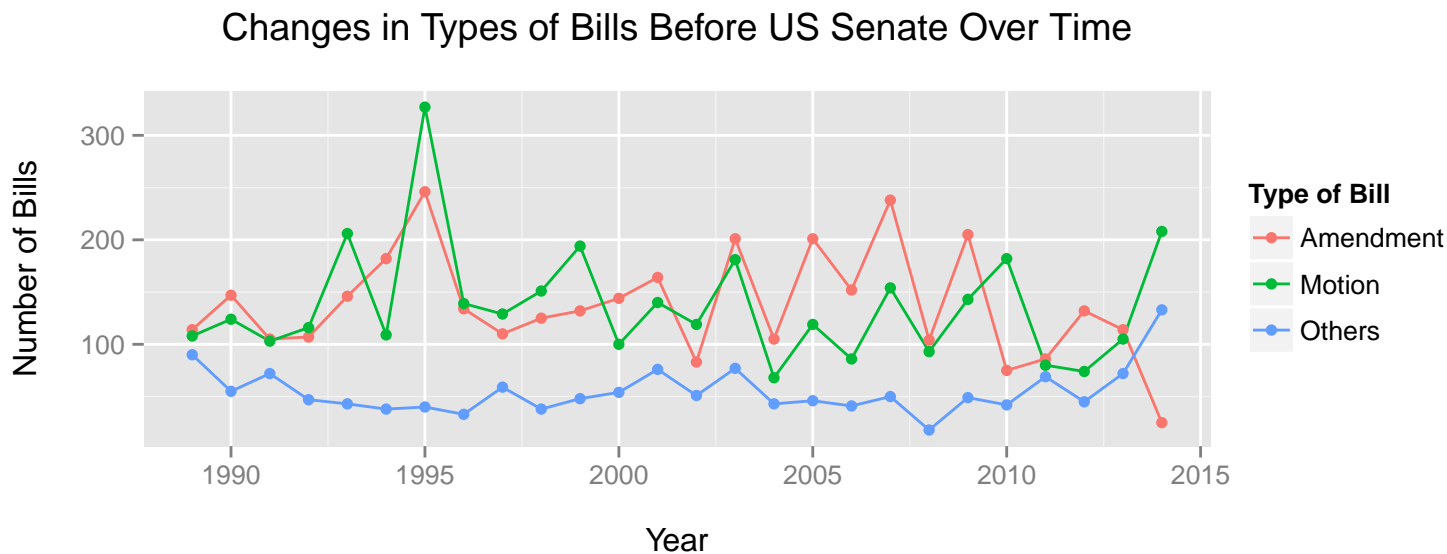
3.3 How Does the Percentage of Passed Bills Change Over Time?

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 bills consistently, but not significantly, decreased between Congress 107 and Congress 113. However, the percentage of passed bills soared in Congress 113. This may be explained as the effect of the debate on Obamacare or other important policies. 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).

Percentage of Bills Passed Per Year
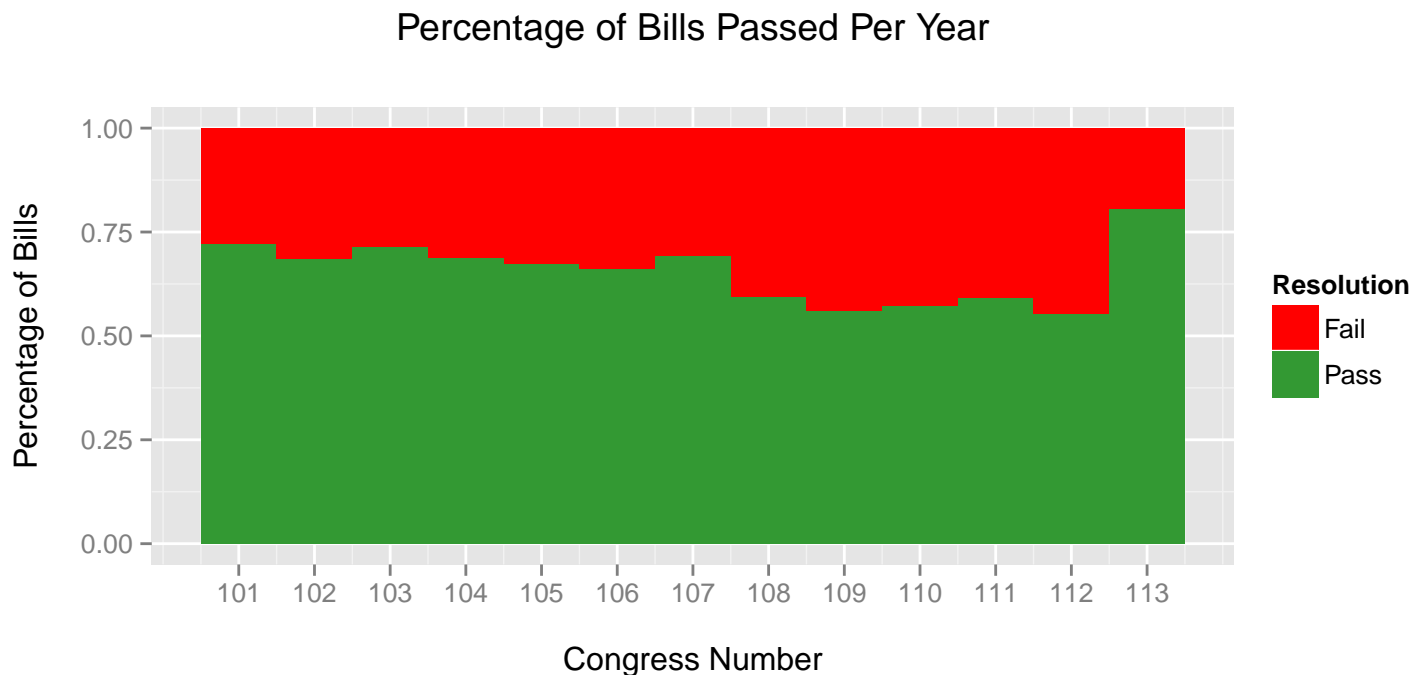
3.4 Are Bills that Received Unanimous Votes Typically Passed or Failed?

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 non-partisan nature of many bills.

Number of Bills Resolved Unanimously Per Congress

*Bills with 5 or Fewer Dissenting Votes*
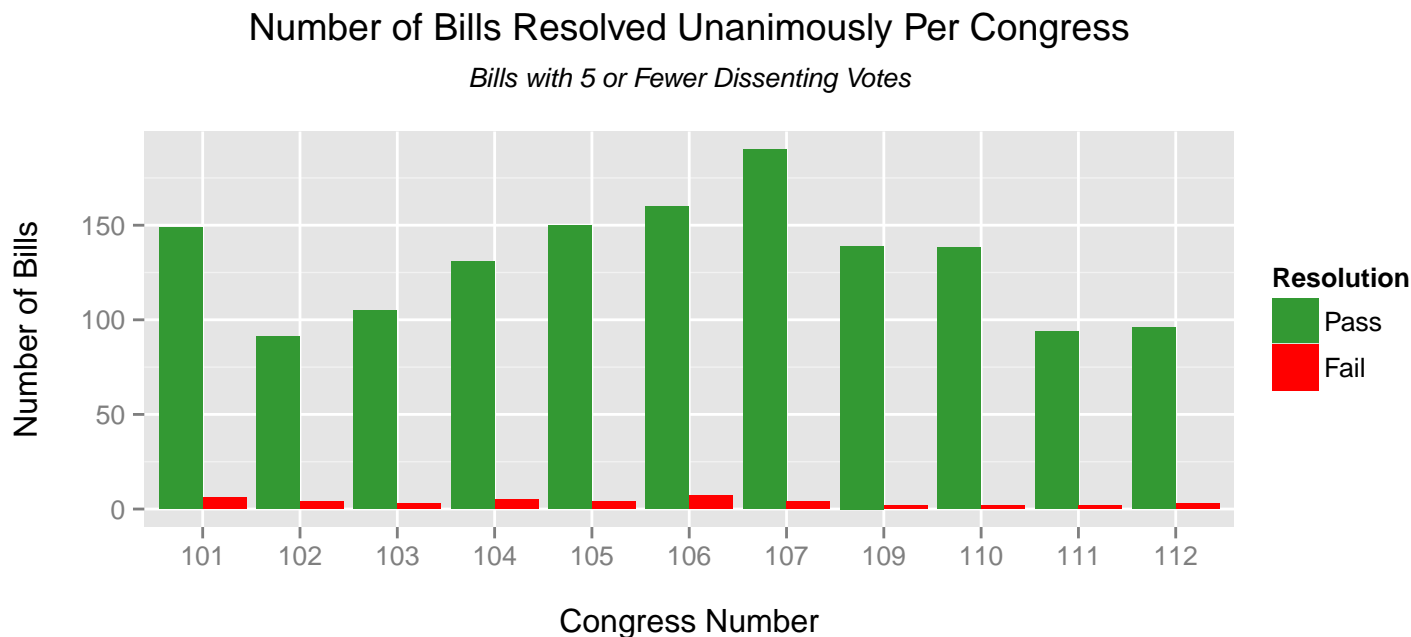
4 An Analysis of Voting Patterns Using Roll-Call Votes with Individual Senator Resolutions

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. We retrieved this data from govtrack.org. Examining individual votes allows us to examine trends in variance of votes within a party and between parties, and this variance may be due to factors such as domestic affairs, politics, and foreign relations.

## 4.1 Do Senators Typically Vote Along Their Party Lines?

For all of the Congresses examined, the majority of both Democratic and Republican Senators voted over 80 percent of the time with their party. There is a small and steady increase between Congresses 101 and 109 in Senators voting with their party. In Congress 110, a discrepancy emerged: most Democrats began to vote 90 percent of the time or more with their party, while most Republicans fell to voting under 90 percent of the time with their party. This trend may correlate with the Obama presidency, which began in 2008 (Congress 110 began in 2007).

Senators Typically Vote Along their Party Lines
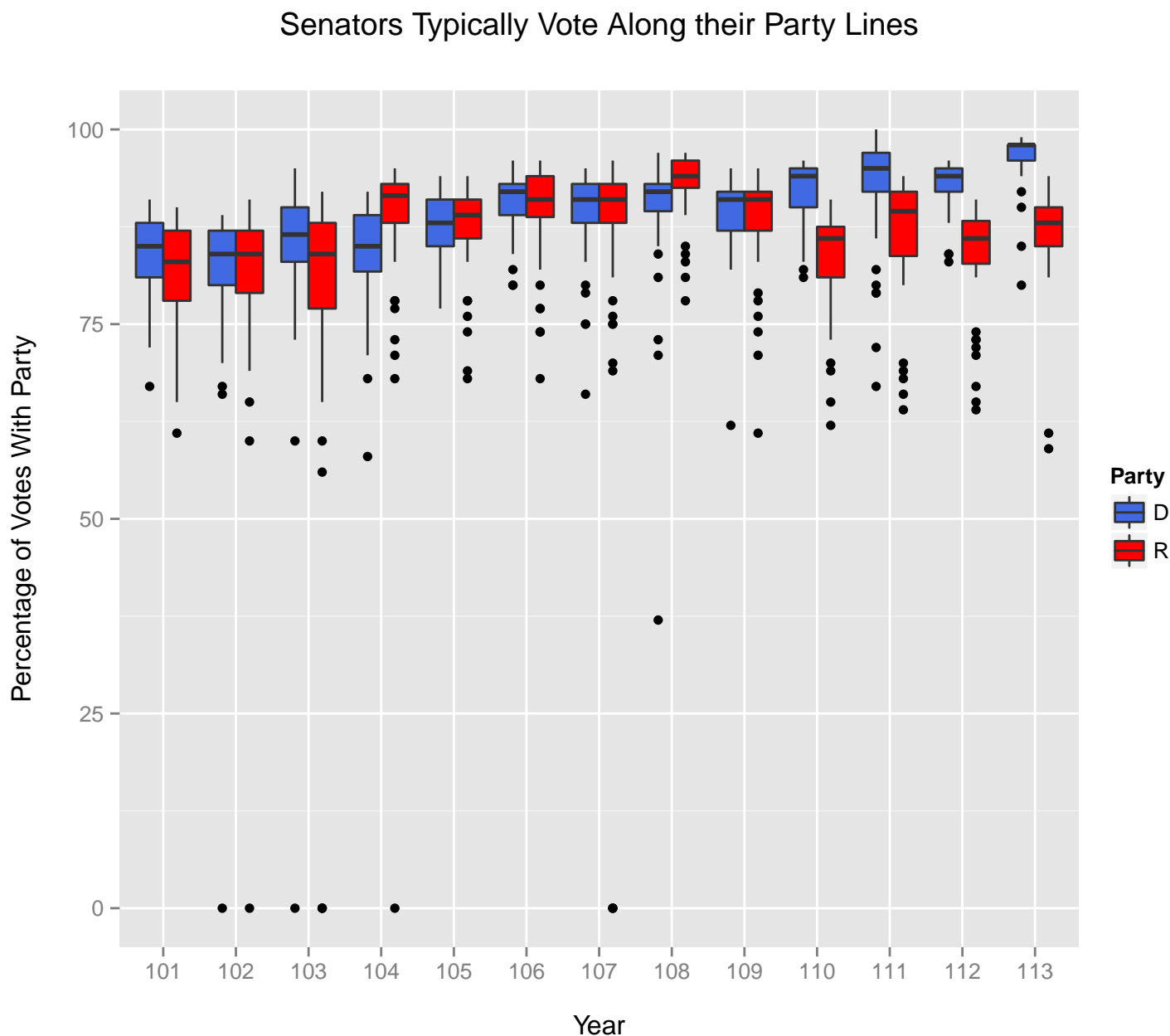
4.2 Do Senators Vote Yea/Nay more in a Given Year?

Data on individual votes is consistent with roll-call vote data in that the majority of Senators in nearly if not all years from 1989-2014 vote yea more often than nay (4.3). However, investigation of the number of yae/nay votes of individual Senators is correlated with political party, as shown through the clustering of red dots in the figure below, each dot symbolizing one Republican Senator and the clustering of blue dots, each dot symbolizing one Democratic Senator, for each year. Certain years have more variance, which follows from a greater divide between party voting for a particular year. Years with a particularly great divide between the difference in yae and nae votes per Senator are 1993, 1995, 2005, and 2014. These difference may be due to political events, such as the Republican Revolution in 1995, during which the GOP experienced great success in the US Midterm Elections (the plot shows a significantly greater number of yaes-nays for Republicans than for Democrats. For the majority of years from 1989-2015, the difference yae - nay for Democrats is greater than for Republicans, except between 1995 and 1999.

Number of Bills Approved By Senators Changes by Year
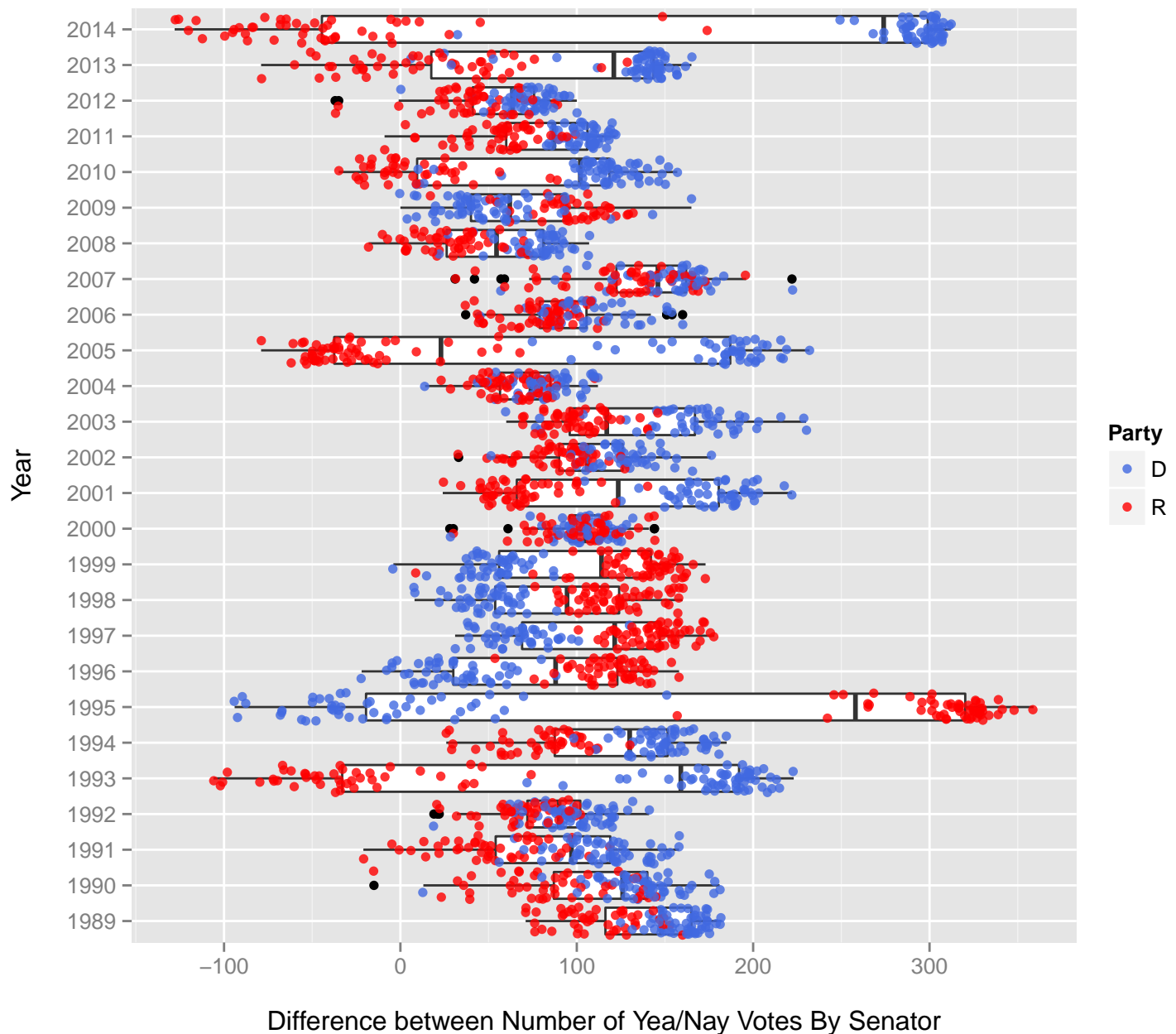
5 Concluding Thoughts

In accord with popular belief, we show that the United States Senate is a bipartisan political body. The Republican and Democratic parties constantly grapple for majority seats in the Senate, while smaller parties like the Independents try to hold on to the few seats they can win.

With our plot on web platforms (3.1), we showed that the parties both adopt the technologies and in turn, perhaps social norms of their time. This provides insight into the possibility of the parties evolution over time. Republican and Democratic political platforms have transformed ever since the parties became the authorities of American politics in the 1800s. The parties have become more similar in certain ways, such as by adopting a position of anti-slavery, but they have switched general positions on certain major issues such as on economic ideology. Our plot is a preliminary analysis of the fluid modernization of behavior and adoption of social norms of the US government. Techniques similar to the ones we used in our analysis of web platforms can be used on data about Republican and Democratic ideology and other political tools over time to increase transparency of the political parties that dominate US politics and their interactions with one another.

There are several advantages to having a two-party system. Bipartisan governments generally tend to be more efficient and more easily managed than those with multiple parties, and two-party system may be the reason for the majority of bills being passed in the US Senate (4.3). In addition, a two-party system intuitively makes the voting system less complicated for voters. However, with only two options, voters may not entirely agree with either party. Gridlock is a major concern, as is corruption.

With Senate seats held by Republicans and Democrats almost evenly with power shifting between parties over terms (3.3), gridlock has the potential to occur often and lead to the discontent of the American public with their government. Further analysis on the frequency of gridlock within the Senate and between the Senate, House of Republicans, and Executive Branch will reveal truths about the efficiency of the US bipartisan system.

The table in 3.4 suggests the possibility of corruption in the US Senate through nepotism. In a country of 318 million individuals, the objective likelihood of relatives serving in the Senate is incredibly low, but the data shows that the phenomenon does in fact occur. Further analysis into individuals voting patterns on specific votes, the effects lobbying, and other political persuasion can gain more insight into the true extent of corruption in the US government.

Politics between Republicans and Democrats bleed into controversial domestic issues and topics such as gun control, abortion, and gay marriage as well as into foreign policy issues such as isolationism versus interventionism, wars, and international trade and economics. In order to fully achieve governmental transparency, a lengthy and deep analysis should be done on the hundreds or thousands of variables that may affect the legislative outcomes of Congress, as well as the outcomes of the Executive and Judicial branches of the US government. Our analysis of the US Senate is a preliminary investigation of the possibilities for achieving governmental transparency. The thought-process and techniques created and utilized in our report can sufficiently be used to model countless datasets in order to entirely lift the barrier between the US government and the people of the United States of America.

6 Appendix

6.1 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 role-call vote APIs:

1. fetchJSON - takes a URL, processes JSON from URL, and returns a raw membership dataframe

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 (generated by fetchXML.R) and uses the above function to retreive all votes. Saves to SQLite.

fetchXML.R - This script contains the functions required to download data from the list of XML files containing roll-call votes hosted on senate.gov:

1. fetchXML - takes a URL, processes XML from URL, and returns a raw senate roll-call dataframe

2. senateRollCall - takes a congress number and chamber; returns a cleaned Senate membership dataframe

3. fetchSenateRollCalls - Takes a range and list of sessions, and retreives the data from senate.gov using the above functions. ORganizes data into a neat dataframe.

fetchData.R - This Script fetches all of the datasets and stores them in a SQLite database file (data.sqlite) using functions from the previous two files.

6.2 Appendix B: Code for Generating Figures

Loading Requirements:

```r
#Setting up libraries and imports
require(RSQLite)
require(ggplot2)
require(grid)
require(reshape2)
require(xtable)
source("../config.R")
source("../voteAnalysis.R")
```

Section 3 Figures:

How Have Senators Adopted Web Platforms Over Time?

```r
websiteCt = queryDB("
  SELECT party as Party, congressNumber, count(*) as webct
  FROM members
  WHERE URL!='' AND Party!='I'
  GROUP BY party, congressNumber
")
twitterCt = queryDB("
  SELECT party as Party, congressNumber, count(*) as twitterct
  FROM members
  WHERE twitter_account!=''
  AND Party!='I'
  GROUP BY party, congressNumber")
fbCt = queryDB("
  SELECT party as Party, congressNumber, count(*) as fbct
  FROM members
  WHERE facebook_account!='' AND Party!='I'
  GROUP BY party, congressNumber")
totalCt = queryDB("
  SELECT party as Party, congressNumber, count(*) as totalct
  FROM members
  WHERE Party!='I'
  GROUP BY party, congressNumber")
mediaCt = merge(websiteCt, twitterCt, by=c("Party", "congressNumber"))
mediaCt = merge(mediaCt, fbCt, by=c("Party", "congressNumber"))
mediaCt = merge(mediaCt, totalCt, by=c("Party", "congressNumber"))
mediaCt$Website = mediaCt$webct/mediaCt$totalct
mediaCt$Twitter = mediaCt$twitterct/mediaCt$totalct
mediaCt$Facebook = mediaCt$fbct/mediaCt$totalct
mediaPlotData = melt(mediaCt, measure.vars = c("Website", "Facebook", "Twitter"),
  id.vars = c("Party", "congressNumber"))
ggplot(mediaPlotData, aes(x=congressNumber, y=value, color=Party))+
  geom_point()+
  geom_line()+
  facet_wrap(~variable)+
  xlim(100, 113)+
  xlab("\nCongress Number\n")+
  ylab("\nProportion of Senators in Party\n")+
  scale_color_manual(values = c("dark blue", "#FF0000"))+
  ggtitle(expression(
    atop("Senators Who Have Ever Used a Web Platform In Their Political Career")
  ))
#ggsave("fig1.png", width = 11, height = 6, dpi = 600)
```

How many Terms do Senators Typically Serve?

```
query = "
  SELECT (first_name || ' ' || last_name) as fullName, party, count(*) AS ct
  FROM members
  WHERE party IN ('D', 'R')
  GROUP BY fullName
  ORDER BY ct DESC"
senatorTotals = queryDB(query, "data.sqlite")
ggplot(senatorTotals)+
  geom_density(aes(x=ct, color=party, fill=party), alpha=.1)+
  xlab("Number of Terms")+
  ylab("Proportion of Senators")+
  scale_color_manual(name="Party", values = c("dark blue", "#FF0000"))+
  scale_fill_manual(name="Party", values = c("dark blue", "#FF0000"))+
  ggtitle("Distribution of Terms Served in 101st to 113th Senates")
```

What are the Dynamics of Political Parties in the US Senate Over Time?

```
partyDataByYear = queryDB("
  SELECT party as Party, congressNumber, count(*) as ct
  FROM members
  GROUP BY party, congressNumber
", "data.sqlite")
ggplot(partyDataByYear, aes(x=congressNumber, y=ct, fill=Party))+
  scale_fill_manual(name = "Party", values = c("#4169E1", "green", "dark blue", "#FF0000"))+
  geom_histogram(position="fill", stat="identity", width=1)+
  xlab("\nCongress Number\n")+
  scale_x_continuous(breaks=101:113)+
  ylab("\nProportion of Senate Seats\n")+
  ggtitle(expression(atop("Senate Seats Held Per Party Is Fairly Stable over Time")))
#ggsave("fig2.png", width = 11, height = 6, dpi = 600)
```

How Often do Multiple Senators Share a Surname?

```
lastCounts=queryDB("
  SELECT tab.last_name as last, count(*) as lastct
  FROM (SELECT id, (last_name || ', ' || first_name) AS full_name, last_name, count(last_name) as ct
  FROM members GROUP BY full_name ORDER BY ct) AS tab
  GROUP BY last ORDER BY lastct DESC", "data.sqlite")
lastCounts = lastCounts[lastCounts$lastct>1,]
colnames(lastCounts) = c("Last Name", "# of Senators")
print(xtable(t(lastCounts)), include.colnames=FALSE)
```

Do Certain States Prefer Certain Political Parties?

```
stateParty = queryDB("
  SELECT state as stateAbrev, party, count(*) as ct
  FROM members
  WHERE party IN ('D', 'R') GROUP BY party, state
", "data.sqlite")
statePartyWide = dcast(stateParty, stateAbrev~party, value.var="ct")
statePartyWide[is.na(statePartyWide)]=0
statePartyWide$diff = statePartyWide$D-statePartyWide$R
statePartyWide$state = apply(statePartyWide, 1, FUN=function(x){stateAbrevToFull(x["stateAbrev"])})
stateMap = map_data("state")
ggplot(statePartyWide)+
  geom_map(data=stateMap, map=stateMap, aes(x=long, y=lat, map_id=region), fill="#ffffff", color="grey10")+
  geom_map(data=statePartyWide, map=stateMap, aes(fill=diff, map_id=state), color="grey10")+
  scale_fill_gradient2(name="Party Difference\n(# Senators)", low="red", mid="white", high="navy blue")+
  ggtitle(expression(
```

```
    atop("Party Preferences By State as Determined by Senate Seats",
    atop(italic("Republicans (Red) vs. Democrats (Blue)"))))
  ))+
  xlab("")+ylab("")+
  scale_x_continuous(breaks=NULL)+
  scale_y_continuous(breaks=NULL)
#ggsave("fig3.png", width = 10, height = 6, dpi = 600)
```

Who are the Independent Senators? Where are They Elected?

```
indep = queryDB("
  SELECT (first_name || ' ' || last_name) as fullName, state, congressNumber, total_votes
  FROM members
  WHERE party='I'
  ORDER BY state", "data.sqlite")
colnames(indep) = c("Name", "State", "Congress #", "Total Votes")
print(xtable(indep), include.rownames = FALSE)
```

Section 4 Figures:

What Types of Roll-Call Votes are Handled by the US Senate?

```
query <- "select * from
          (select type, count(*) as ct
           from senateRollCalls
           group by type)
          where ct > 10"
qtype <- queryDB(query, "data.sqlite")
ggplot(qtype) +
aes(x = reorder(type, ct), y = ct) +
labs(title="Types of Roll-call Votes\n") +
xlab("Bill type\n") +
ylab("\n# of Roll-call Votes") +
geom_bar(stat="identity") +
coord_flip()
```

How Have Trends in Bill Type Changed over Time?

```
query <- "select type, year, count(*) as ct
          from senateRollCalls
          group by type, year"
typeOverYear <- queryDB(query, "data.sqlite")
ggplot(typeOverYear) +
  aes(x=year, y=ct, colour=type) +
  geom_line() +
  geom_point() +
  scale_color_discrete(name="Type of Bill")+
  ylab("Number of Bills\n")+
  xlab("\nYear")+
  ggtitle("Changes in Types of Bills Before US Senate Over Time\n")
```

How Does the Percentage of Passed Bills Change Over Time?

```
passedQuery <- "select 'Pass' as res, congressNumber, type, count(*) as cnt
          from senateRollCalls
          where result == 'Agreed to'
                or result == 'Confirmed'
                or result == 'Passed'
          group by congressNumber, type"
passedResults <- queryDB(passedQuery, "data.sqlite")
```

```
failedQuery <- "select 'Fail' as res, congressNumber, type, count(*) as cnt
          from senateRollCalls
          where result == 'Rejected'
          group by congressNumber, type"
failedResults <- queryDB(failedQuery, "data.sqlite")
results <- rbind(passedResults, failedResults)
ggplot(results) + aes(x=congressNumber, y=cnt, fill=res) +
  geom_histogram(position="fill", stat="identity", width=1) +
  scale_fill_manual(name = "Resolution", values = c("Pass"="#339933", "Fail"="red")) +
  #facet_grid(type ~ .) +
  xlab("\nCongress Number\n") +
  scale_x_continuous(breaks=101:113) +
  ylab("\nPercentage of Bills\n") +
  ggtitle(expression(atop("Percentage of Bills Passed Per Year")))
#ggsave("fig4.png", width = 11, height = 6, dpi = 600)
```

Are Bills that Received Unanimous Votes Typically Passed or Failed?

```
query <- "select pass, fail, p.congressNumber
          from
          (select congressNumber, count(*) as Pass
           from senateRollCalls
           where  CAST(nays as INTEGER) <= 5
           group by congressNumber) as p
          join
          (select congressNumber, count(*) as Fail
           from senateRollCalls
           where  CAST(yeas as INTEGER) <= 5
           group by congressNumber) as f
          on p.congressNumber == f.congressNumber"
unanimous <- queryDB(query, "data.sqlite")
unanimous$congressNumber <- factor(unanimous$congressNumber)
unanimous <- melt(unanimous[,c('congressNumber','Pass','Fail')],id.vars = 1)
ggplot(unanimous,aes(x = congressNumber,y = value)) +
  geom_bar(stat='identity',aes(fill = variable),position = "dodge")+
  scale_fill_manual(name = "Resolution", values = c("Pass"="#339933", "Fail"="red")) +
  xlab("\nCongress Number\n") +
  ylab("\nNumber of Bills\n") +
  ggtitle(expression(
    atop("Number of Bills Resolved Unanimously Per Congress",
    atop(italic("Bills with 5 or Fewer Dissenting Votes")))
  ))
#ggsave("fig5.png", width = 11, height = 6, dpi = 600)
```

Section 5 Figures:

Do Senators Typically Vote Along Their Party Lines?

```
memberPct = queryDB("
  SELECT id, party, missed_votes_pct as missed, votes_with_party_pct as withParty,
  next_election as nextElection, congressNumber
  FROM members
  WHERE party in ('D', 'R')")
memberPct$withParty = as.integer(as.character(memberPct$withParty))
memberPct$congressNumber = as.factor(memberPct$congressNumber)
ggplot(memberPct, aes(x=congressNumber, y=withParty, fill=party))+
  geom_boxplot()+
  xlab("\nYear\n")+
  ylab("\nPercentage of Votes With Party\n")+
  scale_fill_manual(name="Party", values = c("royal blue", "red"))+
```

```
  ggtitle(expression(
    atop("Senators Typically Vote Along their Party Lines")
  ))
#ggsave("fig7.png", width = 11, height = 6, dpi = 600)
```

Do Senators Vote Yea/Nay more in a Given Year?

```
votedata = queryDB("SELECT id, party, vote, count(*) as ct, year
                    FROM votes
                    WHERE party in ('D', 'R')
                    GROUP BY id, vote, year")
votedatawide = dcast(votedata, id+year+party~vote, value.var="ct")
votedatawide$diff = (votedatawide$Yea-votedatawide$Nay)
ggplot(votedatawide, aes(x=as.factor(year), y=diff))+
  geom_boxplot()+
  geom_jitter(aes(color=party), alpha=.8)+
  scale_color_manual(name="Party", values=c("royal blue", "red"))+
  xlab("\nYear\n")+
  ylab("\nDifference between Number of Yea/Nay Votes By Senator\n")+
  ggtitle(expression(
    atop("Number of Bills Approved By Senators Changes by Year")
  ))+
  coord_flip()

#ggsave("fig6.png", width = 11, height = 6, dpi = 600)
```

6.3 Appendix C: Code for Generating Animated .gifs for Time Series Vizualization

During the course of our analysis, we were struggling to find means of vizualizing time-scale state-level data like change in average missed votes by senators of each state over time; or state political preference over time. The typical chloropleth map (where each state is filled with a color determined by a variable) only allows for two dimensions to be vizualized: state, and variable level. Should we wish to use time we might instead have to use averages over time (which eliminates some resolution of the data during exploration).

In series.R we present a simple technique which generates (using the ggmap package) and saves sequential chloropleth maps. These maps are then stitched together by one of any number of applications in order to create an animated gif (we use the ImageMagick toolkit, which comes pre-installed on Xubuntu). For example, we will look at senator miss rate by state over time:

6.3.1 First we generate a series of plots and save them to disk:

```
stateMap = map_data("state")
congressRange = 101:113
alldata = queryDB("SELECT state as stateAbrev, avg(missed_votes_pct) as missed, congressNumber
                   FROM members
                   GROUP BY congressNumber, state
                   ORDER BY congressNumber DESC, missed DESC")
upperbound = ceiling(mean(alldata$missed)+(2*sd(alldata$missed)))
for(congressNumber in congressRange){
  stateVote = queryDB(sprintf("
    SELECT state as stateAbrev, avg(missed_votes_pct) as missed
    FROM members
    WHERE congressNumber=%s
    GROUP BY state",congressNumber))
  stateVote$state = apply(stateVote, 1, FUN=function(x){stateAbrevToFull(x["stateAbrev"])})
  uniScale = seq(60, 100, by=5);
  ggplot()+
    geom_map(data=stateMap, map=stateMap, aes(x=long, y=lat, map_id=region), fill="#ffffff", color="grey10")+
    geom_map(data=stateVote, map=stateMap, aes(fill=missed, map_id=state), color="grey10")+
    ggtitle(sprintf("Avg. Percentage of Votes Missed by State Senators\nCongress Number: %s\n", congressNumber))
    xlab("")+ylab("")+
    scale_x_continuous(breaks=NULL)+
    scale_y_continuous(breaks=NULL)+
    scale_fill_gradient(name="% Missed Votes", limits=c(0,upperbound), na.value="dark red")
  ggsave(file=sprintf("../figures/mapSeries/missed/%s.png", congressNumber), width= 10, height=6, dpi=200)
}
```

6.3.2 Then we generate the gif:

Execute animate.sh This file should be in the folder containing the generated images. The file contains the following command:

```
convert -delay 100 -loop 0 *.png map.gif
```

6.3.3 View the final product here!