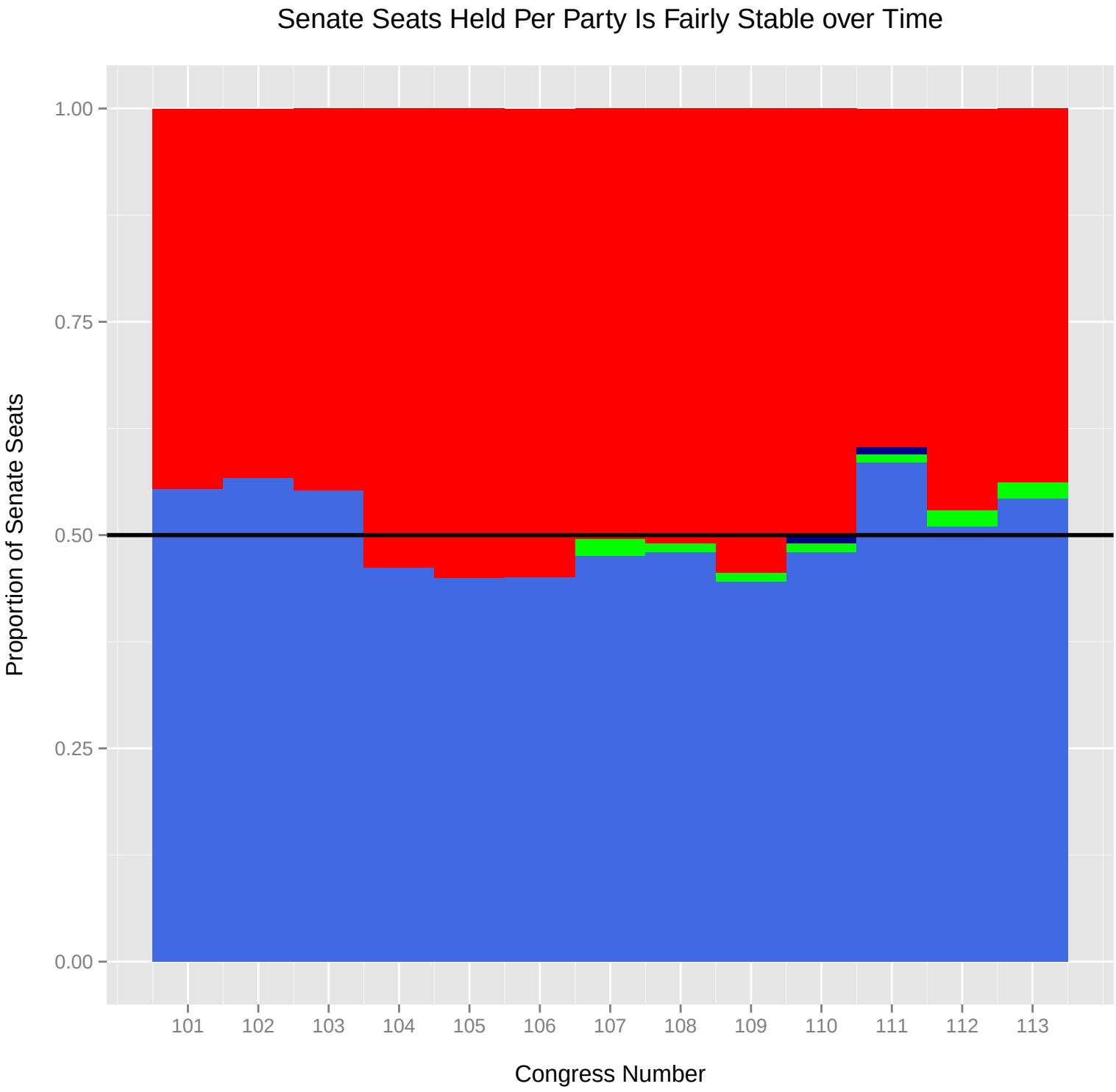
Analyzing Roll Call Votes from the US Senate

Jordan Poles

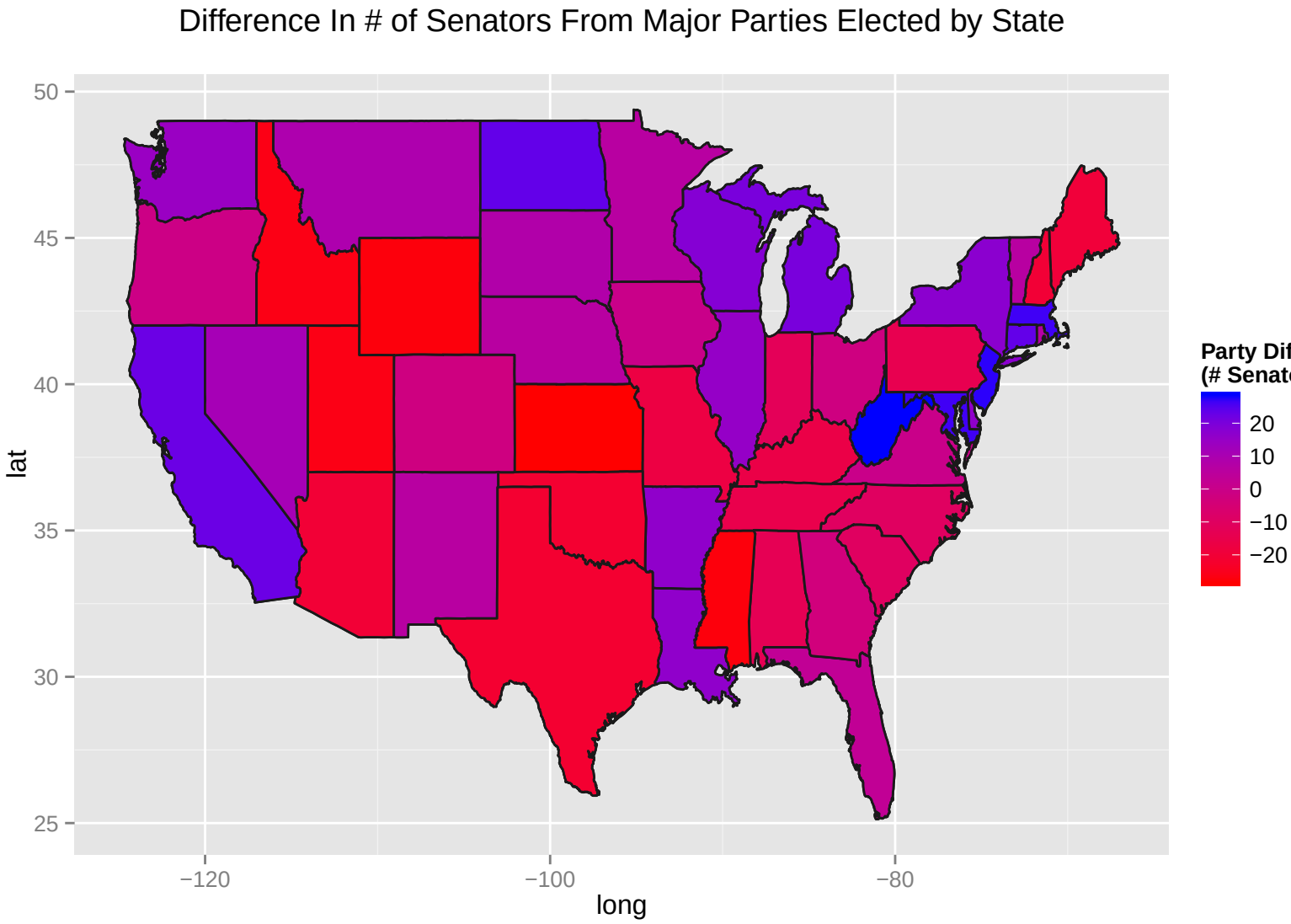
November 23, 2015

```
#Setting up libraries and imports
  require(RSQLite)

## Loading required package:  RSQLite
## Loading required package:  DBI

  require(ggplot2)

## Loading required package:  ggplot2

  require(grid)

## Loading required package:  grid

  require(reshape2)

## Loading required package:  reshape2

  source("../config.R")
  source("../voteAnalysis.R")
```

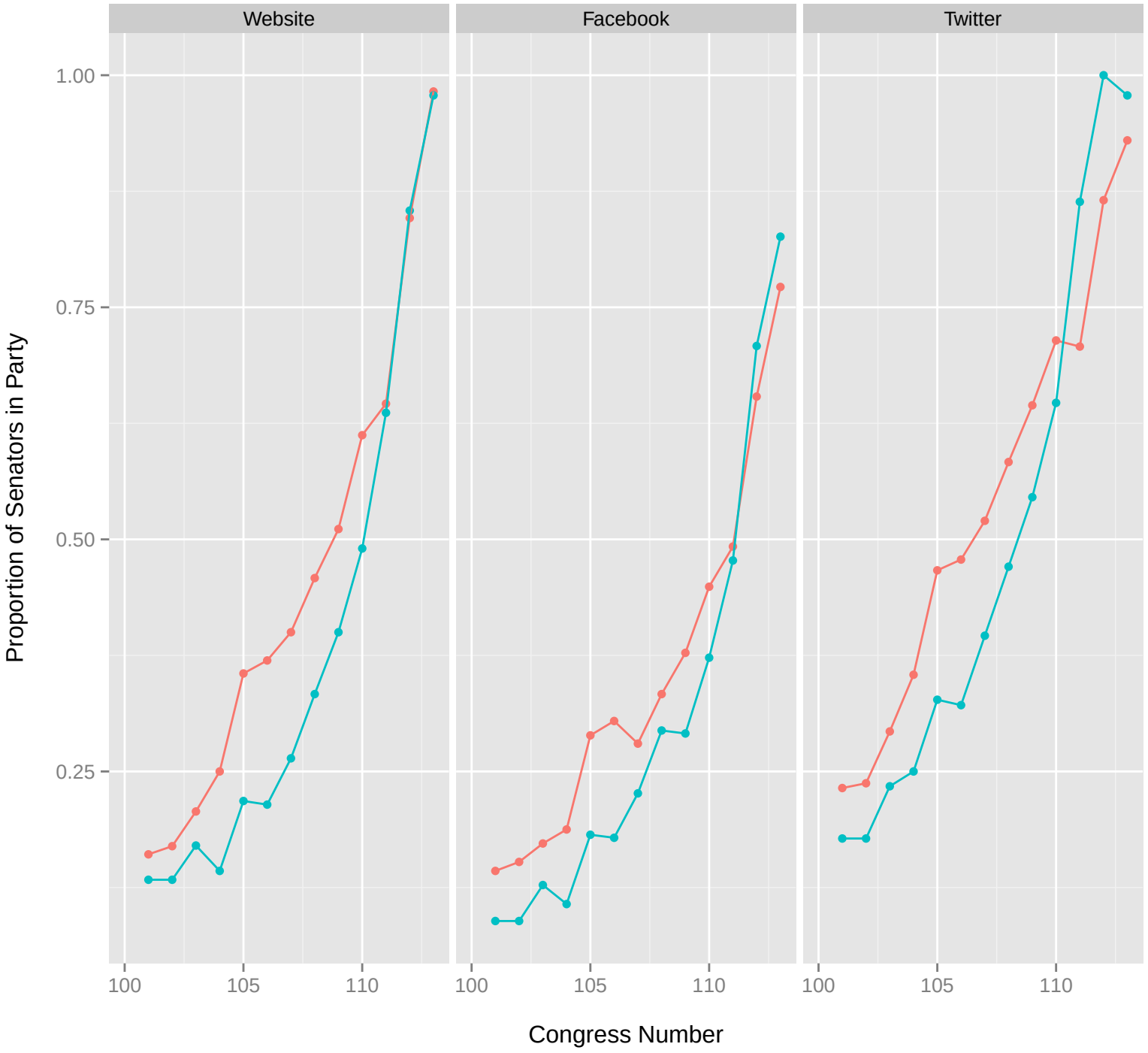
```
partyDataByYear = queryDB(paste("SELECT party as Party, congressNumber, count(*) as ct",
    "FROM members", "GROUP BY party, congressNumber"))
ggplot(partyDataByYear, aes(x = congressNumber, y = ct, fill = Party)) + scale_fill_manual(name = "Party",
    values = c("#4169E1", "green", "dark blue", "#FF0000")) + geom_histogram(position = "fill",
    stat = "identity", width = 1) + xlab("\nCongress Number") + scale_x_continuous(breaks = 101:113) +
    ylab("Proportion of Senate Seats\n") + ggtitle("Senate Seats Held Per Party Is Fairly Stable over Time\n") +
    geom_hline(yintercept = 0.5, size = 1)
```



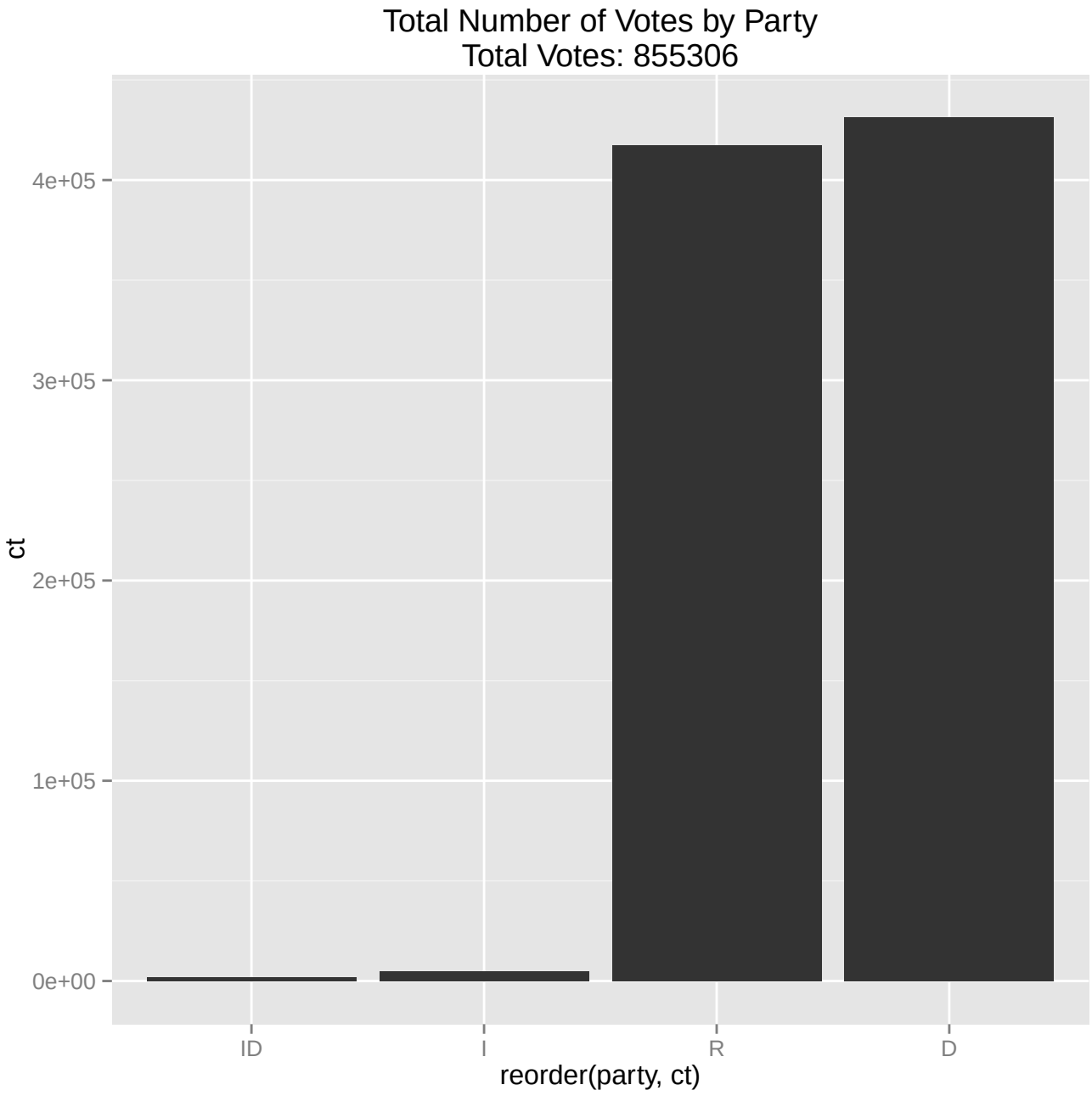
```
stateParty = queryDB(oaste("SELECT state as stateAbrev, party, count(*) as ct",
    "FROM members", "WHERE party IN ('D', 'R') GROUP BY party, state"))

## Error in dbSendQuery(dbcon, query):  error in evaluating the argument 'statement' in selecting a method for fun
'dbSendQuery':  Error:  could not find function "oaste"

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_gradient(name = "Party Differ
    low = "red", high = "blue") + ggtitle(paste("Difference In # of Senators",
    "From Major Parties Elected by State\n"))
```



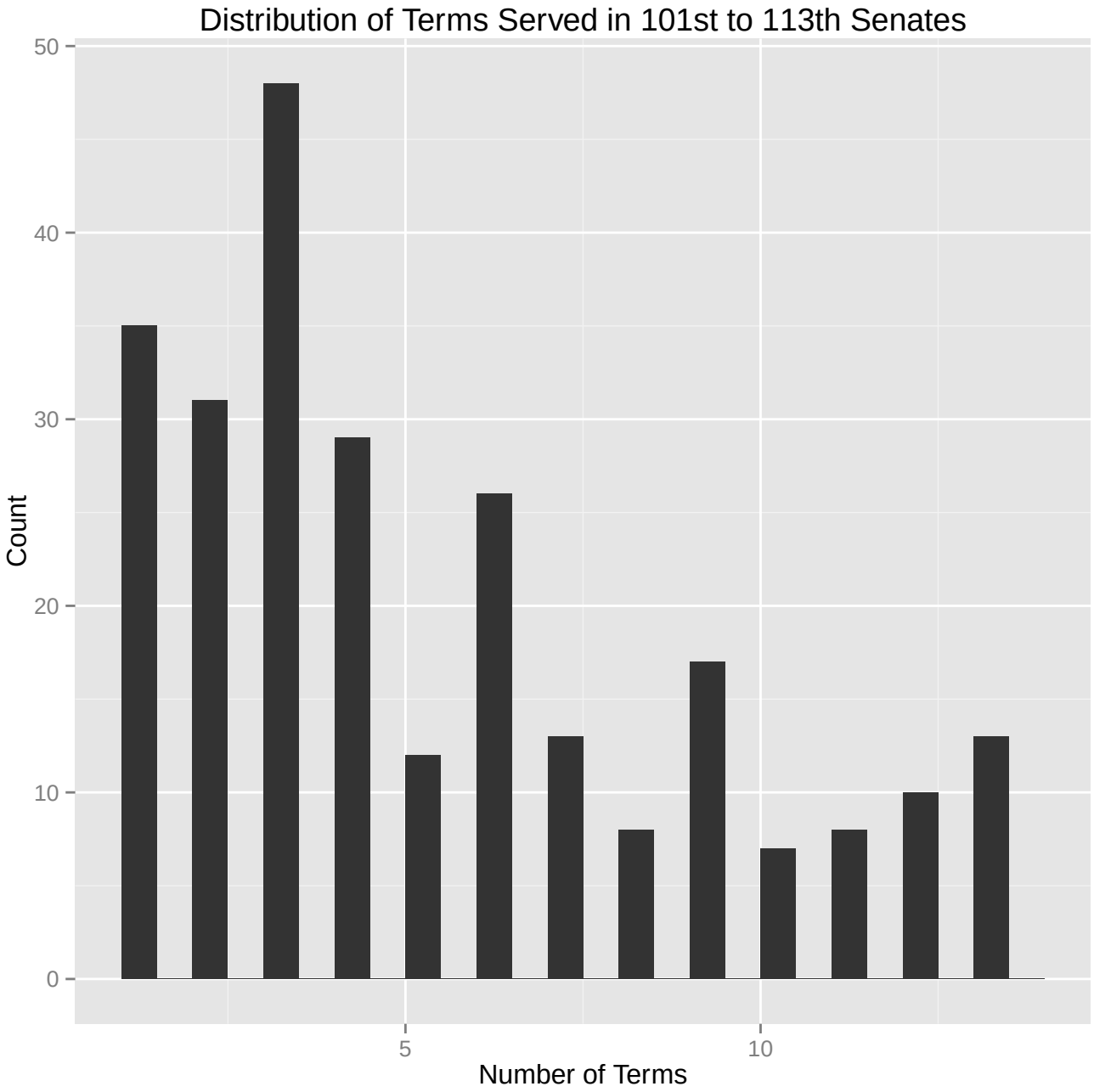
```
websiteCt = queryDB("SELECT party as Party, congressNumber, count(*) as webct FROM members WHERE URL!='' AND Party
twitterCt = queryDB("SELECT party as Party, congressNumber, count(*) as twitterct FROM members WHERE twitter_accou
fbCt = queryDB("SELECT party as Party, congressNumber, count(*) as fbct FROM members WHERE facebook_account!='' AN
totalCt = queryDB("SELECT party as Party, congressNumber, count(*) as totalct FROM members WHERE Party!='I' GROUP
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") +
    ylab("Proportion of Senators in Party\n") + ggtitle("Proportion of Senators Who Have Ever Used a Web Platform
```


Proportion of Senators Who Have Ever Used a Web Platform In Their Political Career
Website
Facebook
Twitter
1.00
0.75
0.50
0.25
Proportion of Senators in Party
100
105
110
Congress Number

```
query = "SELECT party, count(*) as ct FROM votes GROUP BY party"
partyVoteTotals = queryDB(query, "data.sqlite")
ggplot(partyVoteTotals, aes(x = reorder(party, ct), y = ct)) + geom_bar(stat = "identity") +
    ggtitle(sprintf("Total Number of Votes by Party\nTotal Votes: %s", sum(partyVoteTotals$ct)))
```

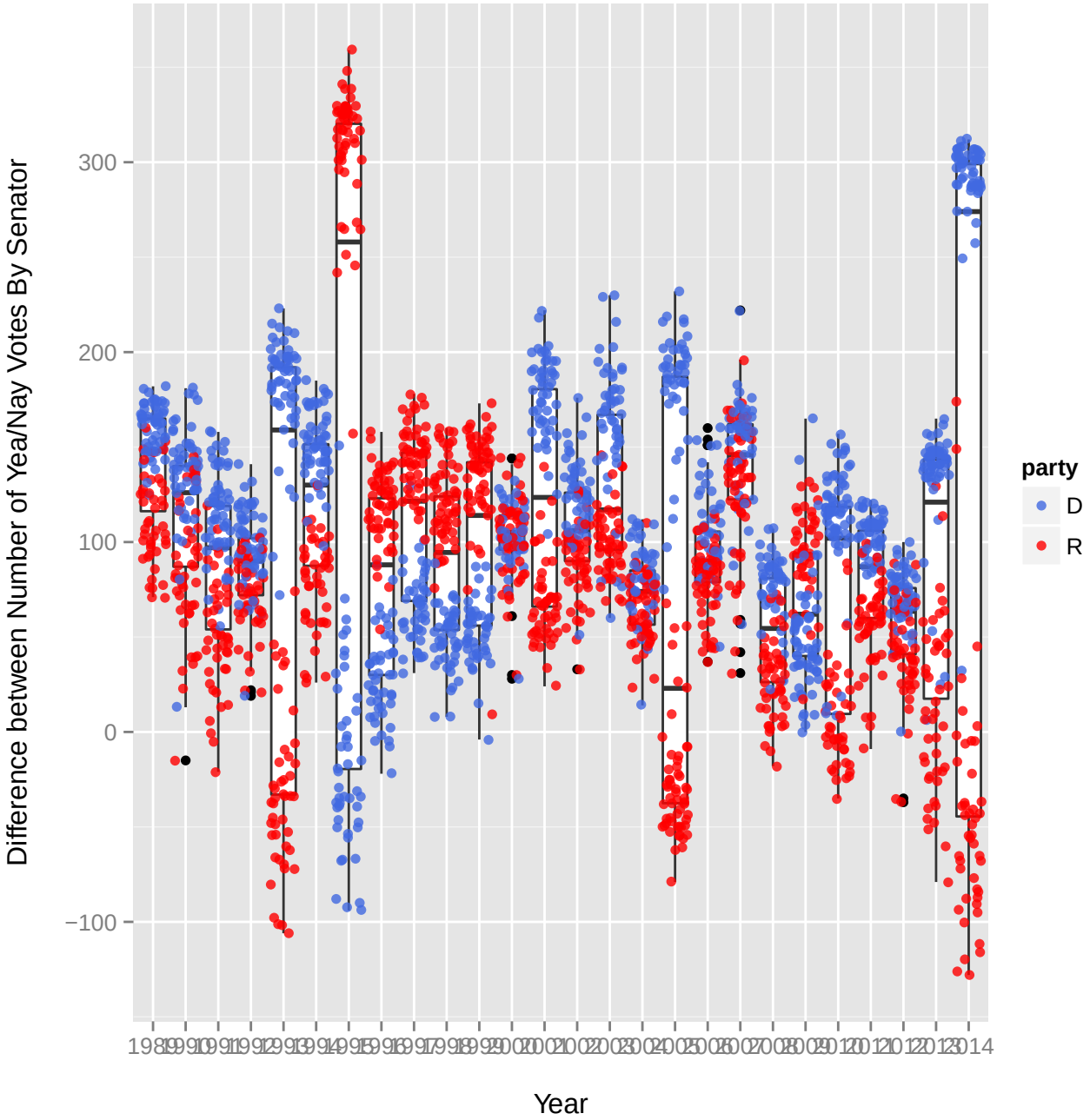


```
query = "SELECT id, first_name, last_name, party, seniority, count(*) AS ct FROM members GROUP BY id ORDER BY ct DE
senatorTotals = queryDB(query, "data.sqlite")
ggplot(senatorTotals, aes(x = ct)) + xlim(1, 14) + geom_bar(binwidth = 0.5) +
    xlab("Number of Terms") + ylab("Count") + ggtitle("Distribution of Terms Served in 101st to 113th Senates")
```

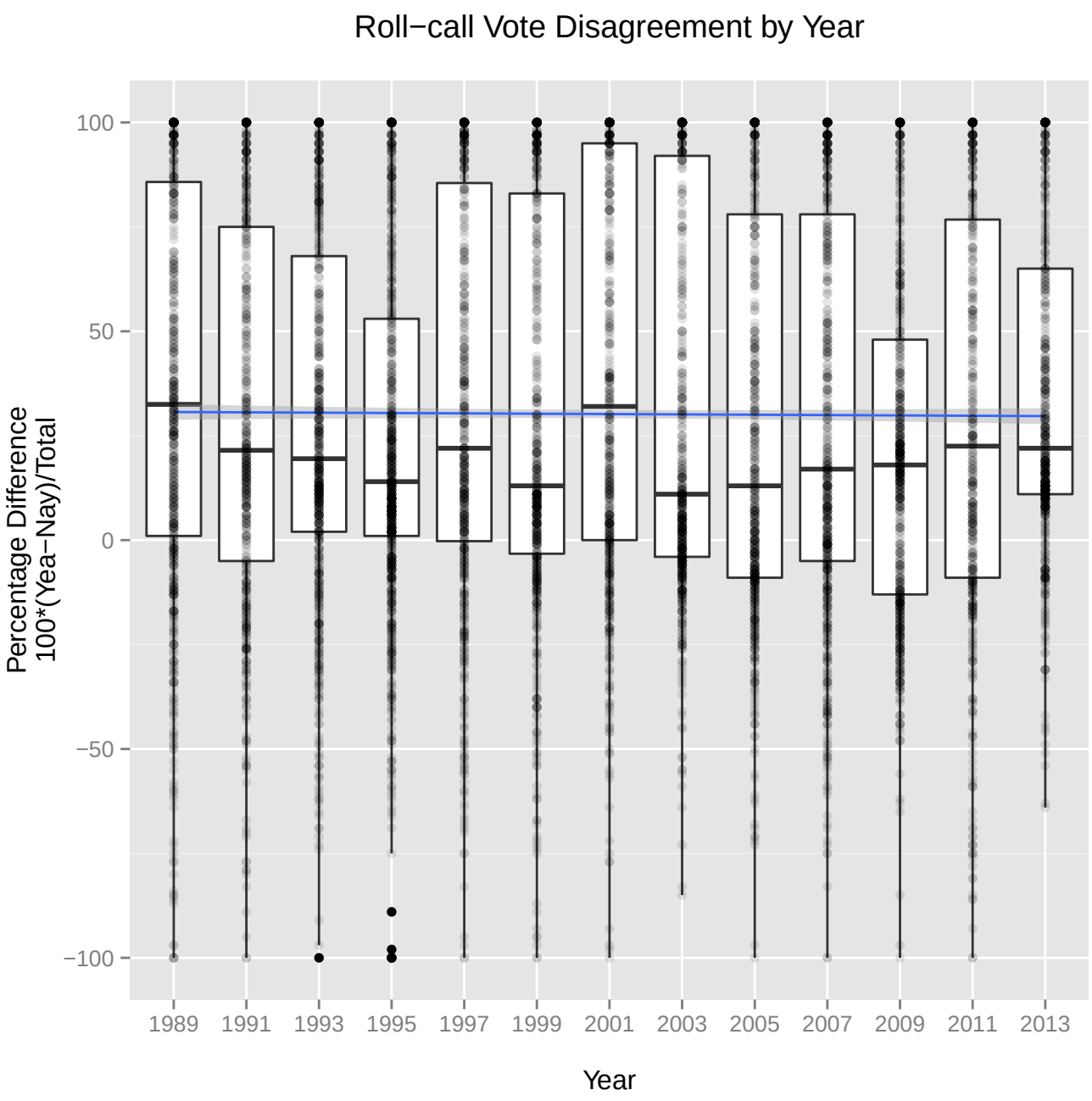


```
  votedata = queryDB("SELECT id, party, vote, count(*) as ct, year FROM votes WHERE party in ('D', 'R') GROUP BY id
  votedatawide = dcast(votedata, id+year+party~vote, value.var="ct")
  votedatawide$diff = (votedatawide$Yea-votedatawide$Nay)
  ggplot(votedatawide, aes(x=year, y=diff))+geom_boxplot()+geom_jitter(aes(color=party), alpha=.8)+scale_color_manu
```

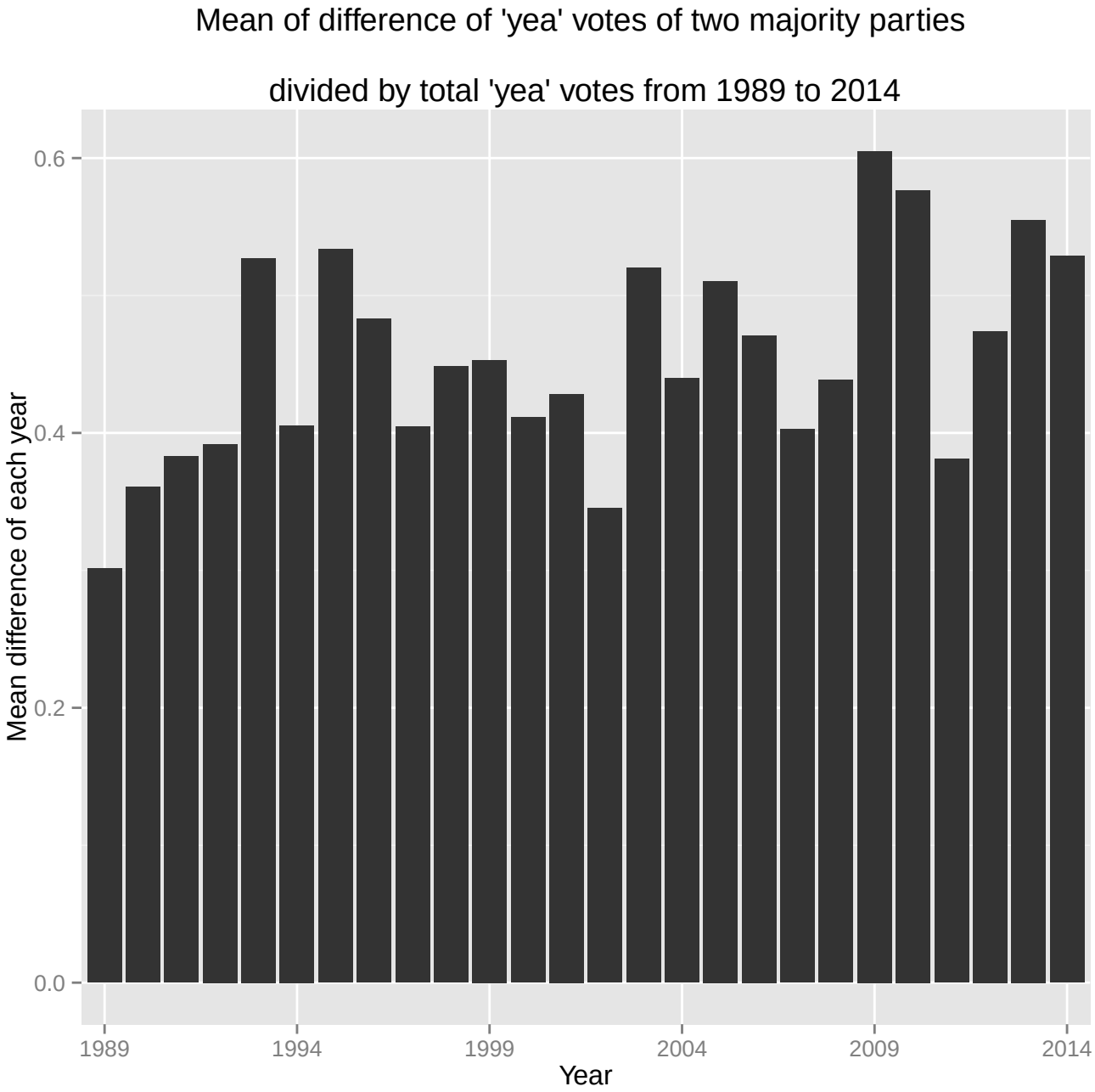
```
## Warning:  Removed 3 rows containing non-finite values (stat_boxplot).
## Warning:  Removed 3 rows containing missing values (geom_point).
```



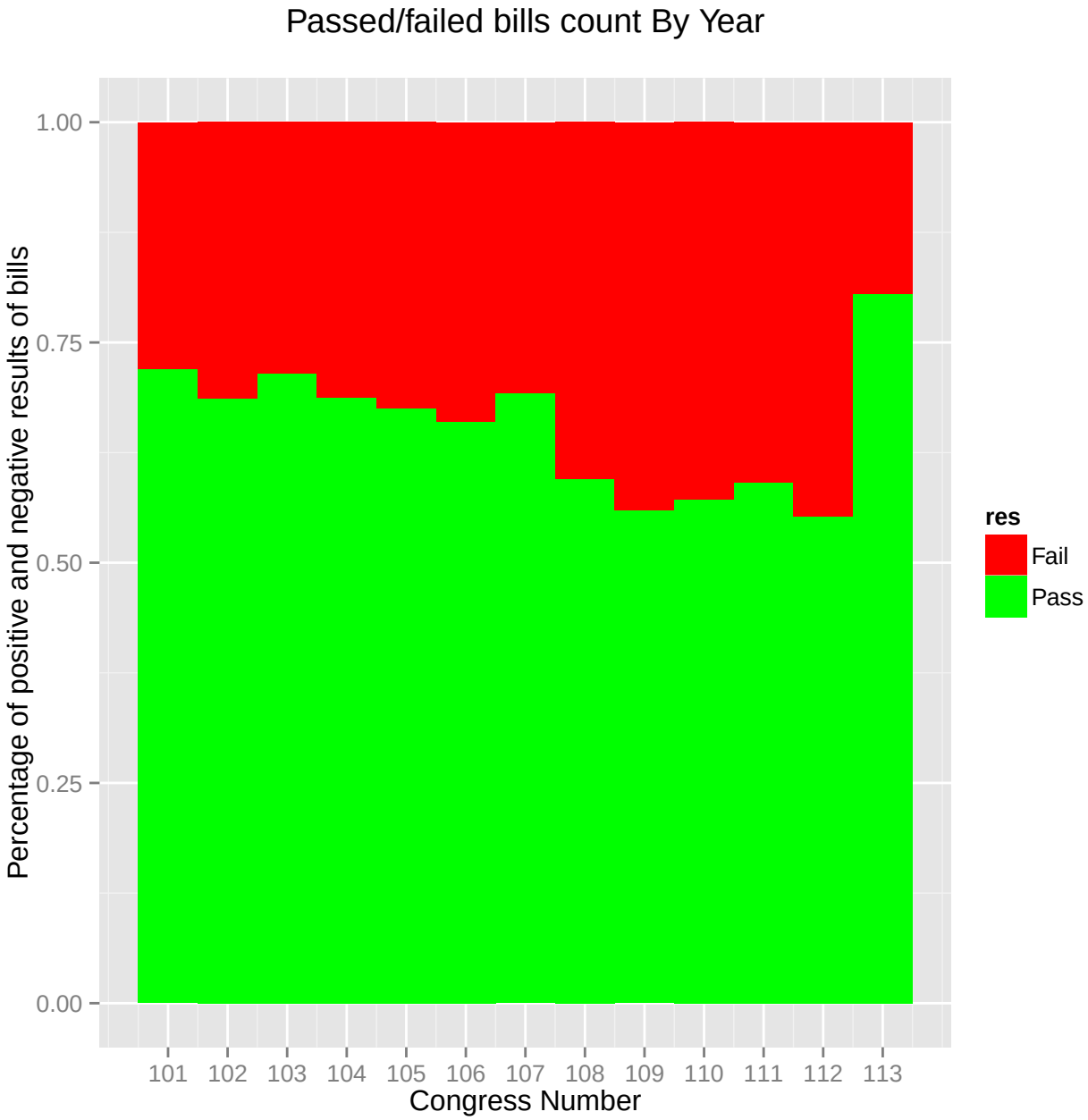
```
query = "SELECT yeas, nays, (yeas+nays) as total, (100*(yeas-nays)/(yeas+nays)) as voteDiff, congressNumber, sessi
rollCallStats = queryDB(query, "data.sqlite")
rollCallStats$year = apply(rollCallStats, 1, function(x) {
    congressToYear(x["congressNumber"], x["sessionNumber"])
})
ggplot(rollCallStats, aes(x = as.factor(year), y = voteDiff)) + geom_boxplot() +
    geom_smooth(aes(group = 1), method = "lm") + geom_point(alpha = 0.1) + ggtitle("Roll-call Vote Disagreement by
    xlab("\nYear") + ylab("Percentage Difference\n100*(Yea-Nay)/Total")
```



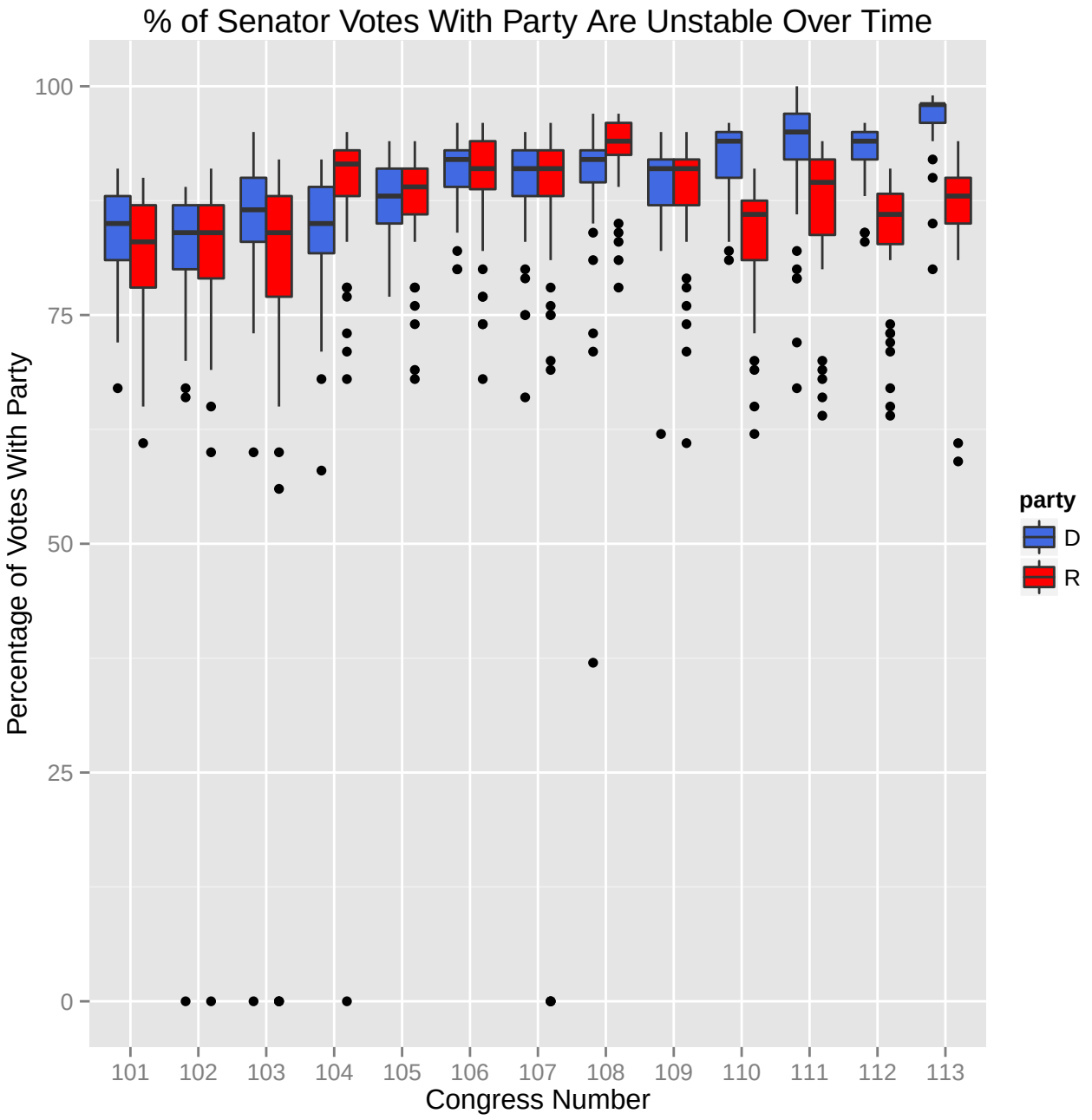
```
query = "select r.year as year, r.voteNumber as voteNumber,
            abs(r.c - d.c) * 1.0 / (r.c + d.c) as diff
         from (select voteNumber, year, count(*) as c
               from votes
               where vote == 'Yea' and party == 'R' group by year, voteNumber)
               as r
              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
Mean difference of each year
0.6
0.4
0.2
0.0
1989
1994
1999
2004
2009
2014
Year

```
passedQuery <- "select 'Pass' as res, congressNumber, count(*) as cnt
          from senateRollCalls
          where result == 'Agreed to'
                or result == 'Confirmed'
                or result == 'Passed'
          group by congressNumber"
passedResults <- queryDB(passedQuery, "data.sqlite")
failedQuery <- "select 'Fail' as res, congressNumber, count(*) as cnt
          from senateRollCalls
          where result == 'Rejected'
          group by congressNumber"
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 = "res", values = c("red", "green")) +
  xlab("Congress Number\n") +
  scale_x_continuous(breaks=101:113) +
  ylab("\nPercentage of positive and negative results of bills") +
  ggtitle("Passed/failed bills count By Year\n")
```

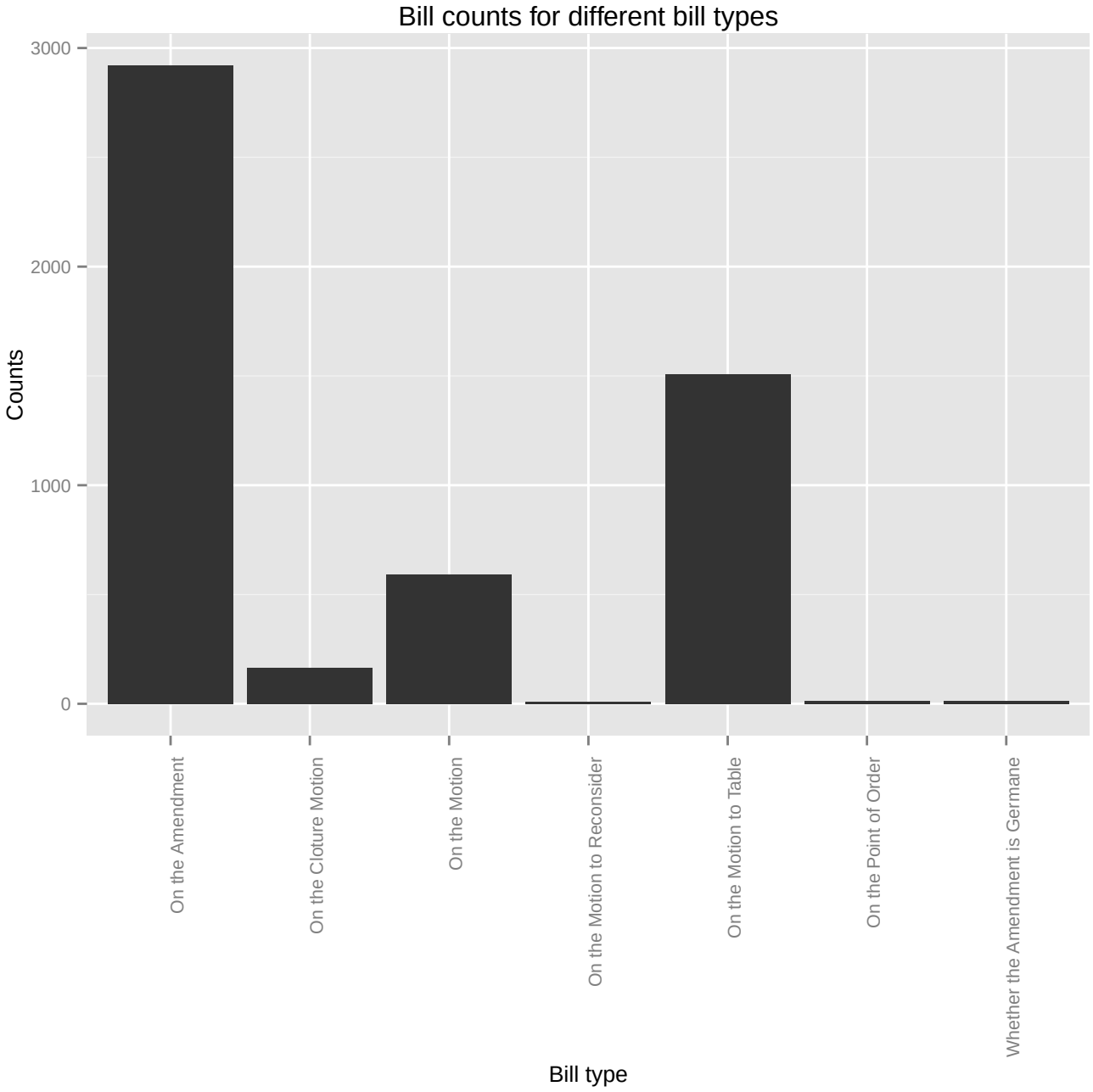

Passed/failed bills count By Year
Percentage of positive and negative results of bills
1.00
0.75
0.50
0.25
0.00
101
102
103
104
105
106
107
108
109
110
111
112
113
Congress Number
res
Fail
Pass

```
memberPct = queryDB("SELECT id, party, missed_votes_pct as missed, votes_with_party_pct as withParty, next_electi
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("Congress Number")+ylab("Pe
```




```
query <- "select * from
          (select questionType, count(*) as cnt
           from senateRollCalls
           where questionType not null
           group by questionType)
         where cnt > 10"
type <- queryDB(query, "data.sqlite")
ggplot(type) +
aes(x = questionType, y = cnt) +
labs(title="Bill counts for different bill types") +
xlab("Bill type") +
ylab("Counts") +
geom_bar(stat="identity") +
```

```
  theme(text = element_text(size=10),
      axis.text.x = element_text(angle=90, hjust = 1))
```




```
  stateMap = map_data("state")
  stateVote = queryDB(sprintf("SELECT state as stateAbrev, avg(votes_with_party_pct) as withParty FROM members GROU
  stateVote$state = apply(stateVote, 1, FUN=function(x){stateAbrevToFull(x["stateAbrev"])})
  ggplot()+geom_map(data=stateMap, map=stateMap, aes(x=long, y=lat, map_id=region), fill="#ffffff", color="grey10"
```

```
## Error in sprintf("Average % of Votes With Party By State"):  too few arguments
```