Abstract:
I used the sqlite database of noun phrases I built in the previous assignment and R to create a back to back histogram. This graphic represents the top 10 phrases ranked by the number of bills they appear in where the right hand side represents the Senate and the left the House. The phrase is rendered in the bar itself.

Data Set:
The data set represented here is the same as the one computed from last week's assignment.

Method:
I used the 'RSQLite' package to build an SQL connection to my database of noun phrases. Using a query that returned the most popular phrase and it's associated count for the House and Senate, I wrote these value into data frames. I then created a vector that represented the rank of the phrase the number of times the phrase was repeated. I did this because R was unwilling to compute a histogram of string values. So instead I aliased each string value to an integer rank and then plotted the histogram. I utilized the histbackback() function from the 'Hmisc' library for basic rendering of the graph and it's labels and axes. I then programmatically added text labels which represented the actual noun phrase being portrayed.

Code:

```r
require('DBI')
require('RSQLite')
require('Hmisc')
m <- dbDriver('SQLite')
con <- dbConnect(m, '/home/foo/618hw3.db')
rss <- dbSendQuery(con,
  'select foo.billnp as np, sum(foo.bar) as count
  from (select billnp, count(distinct(billno)) as bar
       from billnp
       where billno like "s%"
       group by billnp, billyear) foo
  group by foo.billnp
  order by sum(foo.bar) desc
  limit 10')
senate_phrasing <- fetch(rss)
dbClearResult(rss)

rsh <- dbSendQuery(con,
  'select foo.billnp as np, sum(foo.bar) as count
  from (select billnp, count(distinct(billno)) as bar
       from billnp
       where billno like "h%"
       group by billnp, billyear) foo
  group by foo.billnp
  order by sum(foo.bar) desc
  limit 10')
house_phrasing <- fetch(rsh)
dbClearResult(rsh)
```
House <- c()
Senate <- c()
for (i in 1:10) {
    House <- append(House, rep(i, house_phrasing$count[i]))
    Senate <- append(Senate, rep(i, senate_phrasing$count[i]))
}

hist_plot <- histbackback(
    House, Senate,
    brks=0:10,
    sub='Number of Bills Phrase Appears In',
    ylab='Rank of Phrase by Frequency',
    main='Number of Bills Top 10 Phrases Appear in by Branch')
barplot(-hist_plot$left, col='blue', horiz=T, add=T, space=0, axes=F)
barplot(hist_plot$right, col='red', horiz=T, add=T, space=0, axes=F)

par(cex=.7)
for (i in 1:10){
    text(-400, i-.5, house_phrasing$np[i])
    text(200, i-.5, senate_phrasing$np[i])
}

Resulting Plot:
Closing Thoughts:

Using unweighted word frequency results in capturing the form of the congressional documents, but not the content.