
Assignment 3

Yuqi Liao

802642182

April 13, 2015

#Introduction:

This project tries to explore the relationship between ideology and words. I build a validated model of words and ideology of the congressperson. I've gone through the following process.

#Process:

1- Use the sunlight foundation API to download the congressional record for the latest (113th) congress (Senate and House). Using the script by Gaurav Sood, the data is well structured to reflect information including the speaker, his or her party and the speech.

2- Clean up data, using code from Gaurav Sood. In particular, remove stop words, stem the words, remove special characters, numbers, capitalization etc.

3- Download ideology scores from voterview.com.

4- Merge the cleaned speech data with ideology scores from voterview.com (DW-Nominate, 1st dimension). Regarding the fact that the names in the two datasets are not the same, I use the a csv file named Charles Stewart's data found online as a bridge that matches up ICPSR IDs to Bioguide IDs.

5- Turn the merged data into bigram and trigram using similar codes. Keep only top 500 bigrams, top 500 trigrams sorted by the chi-square statistics. (The chi-square test function is based on the paper suggested by Gaurav Sood)

6- Come up with the top 100 republican and democrat bigrams and trigrams in a text file following on the chi-square test results. They indicate that there are statistically significantly different usage of bigrams and trigrams by the two party members. For example, when "climate change" is mentioned, it is more likely to be said from a democratic congressperson other than a republican one. The following modeling part have more details to support this argument.

7- Modeling. I've used a simplest OLS model as well as a more nuanced Ridge and Lasso model to approach this. The three models unsurprisingly yields similar results, which again supports some preliminary findings by the chi-square tests.

To try to use these three models, the first thing to do is to come up with a matrix with 504 columns. The first four columns include the name of the congressperson, their party, their speeches and their ideology score (These four columns all come from the merged data set that has been generated, which has more than one row for each congressperson). The rest 500 columns consist the top 500 bigram by chi-square statistics (there will be similar logic for the top 500 trigram analysis, due to the time constraint, this part is skipped.) Then I try to construct a function in r to read through the speech cells of each row and assign value "1" to a bigram or trigram cell for that row if the bigram or trigram phrase in mentioned in the speech. For example, the congressperson in the first row has a cell about speech. If the term "climate change" is mentioned in the speech, the value in that row under the dummy variable "climate change" is defined as "1", otherwise it is "0". After this process, I try to run a OLS regression, in which I regress the ideology score on the 500 dummy variables. I can then try to interpret the output as to which bigrams/trigrams have the most significant impact on the ideology score. Following the guidance of the Introduction to Statistical Learning textbook, I build the Ridged and Lasso model with cross-validation and foudn similar results.

#Model Outputs

1- OLS

Call:

```
lm(formula = dwnom1 ~ . - dwnom1 - Name.in.Bridge.dataset - speaker_party -
speaking, data = mergeddataset_new_clean_sample)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.9544	-0.2477	0.0000	0.1547	1.1214

Coefficients: (251 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.027615	0.066187	0.417	0.6769
climat.chang	-1.370795	5.145171	-0.266	0.7901
mr.presid	0.050797	0.131567	0.386	0.6998
unanim.consent	-0.403299	2.446616	-0.165	0.8692
presid.ask	0.324445	0.215355	1.507	0.1332
minimum.wage	0.115009	6.166129	0.019	0.9851
care.law	-0.871423	0.932433	-0.935	0.3509
ask.unanim	0.026667	2.461447	0.011	0.9914
senat.proceed	0.507493	0.819217	0.619	0.5362
republican.colleagu	1.296391	2.119514	0.612	0.5413
session.senat	-0.036351	0.391489	-0.093	0.9261

meet.session	NA	NA	NA	NA
rhode.island	-1.712523	0.735501	-2.328	0.0207 *
author.meet	NA	NA	NA	NA
obama.administr	-0.328097	1.223194	-0.268	0.7887
eagl.scout	-0.369068	2.084750	-0.177	0.8596
new.york	-0.290983	0.243179	-1.197	0.2326
gun.violenc	-1.691932	1.643846	-1.029	0.3044
immigr.reform	-1.985701	1.404354	-1.414	0.1586
consent.member	2.810853	1.251565	2.246	0.0256 *
senat.offic	-0.063920	0.705343	-0.091	0.9279
student.loan	0.819199	1.421075	0.576	0.5648
action.debat	0.544108	1.249736	0.435	0.6637
interven.action	NA	NA	NA	NA
consent.senat	-0.222609	0.688075	-0.324	0.7466
consent.committe	-0.319954	0.555298	-0.576	0.5650
civil.right	NA	NA	NA	NA
white.hous	0.386208	0.446109	0.866	0.3875
afford.care	-1.506976	0.973176	-1.549	0.1228
read.titl	0.664040	0.973170	0.682	0.4957
american.peopl	0.120415	0.277012	0.435	0.6642
offic.build	NA	NA	NA	NA
carbon.pollut	NA	NA	NA	NA
day.revis	-0.273000	0.775478	-0.352	0.7251
african.american	0.230241	0.570654	0.403	0.6869
dirksen.senat	NA	NA	NA	NA
clerk.read	-0.606611	0.560245	-1.083	0.2800
major.leader	1.982005	1.057609	1.874	0.0621 .
motion.reconsid	-0.506412	0.560529	-0.903	0.3672
care.act	2.122055	1.127435	1.882	0.0610 .
presid.health	1.482691	1.436602	1.032	0.3030
tabl.interven	NA	NA	NA	NA
legisl.day	NA	NA	NA	NA
unemploy.insur	1.681885	1.219431	1.379	0.1691
illeg.immigr	0.819930	0.627645	1.306	0.1926
comprehens.immigr	4.402690	4.112948	1.070	0.2855
may.legisl	-0.839874	1.023182	-0.821	0.4125
speaker.move	0.667079	0.395633	1.686	0.0930 .
announc.senat	-0.523615	0.393348	-1.331	0.1843
execut.session	0.449978	1.089908	0.413	0.6801
rollcal.vote	-0.481014	0.393648	-1.222	0.2229
member.may	NA	NA	NA	NA
employ.mandat	-1.481969	2.582903	-0.574	0.5666

youth.award	-15.066110	6.952844	-2.167	0.0312 *
rais.minimum	NA	NA	NA	NA
hear.entitl	NA	NA	NA	NA
upon.tabl	NA	NA	NA	NA
madam.presid	-0.101179	0.196649	-0.515	0.6073
mr.speaker	0.042306	0.096862	0.437	0.6627
muslim.brotherhood	NA	NA	NA	NA
laid.upon	NA	NA	NA	NA
san.francisco	-0.820329	0.554718	-1.479	0.1404
remark.includ	-0.247000	0.775478	-0.319	0.7504
includ.extran	NA	NA	NA	NA
interest.rate	-0.989211	2.884925	-0.343	0.7320
presid.offic	0.052377	0.321153	0.163	0.8706
absent.senat	0.617385	0.393348	1.570	0.1178
boy.scout	0.044931	1.253740	0.036	0.9714
morn.busi	0.220311	0.316862	0.695	0.4875
tea.parti	-0.882315	0.861421	-1.024	0.3067
speaker.boehner	-2.505620	1.185773	-2.113	0.0356 *
trillion.debt	-1.429935	4.972001	-0.288	0.7739
period.morn	-0.835802	1.721296	-0.486	0.6277
tax.break	-7.042344	7.266482	-0.969	0.3334
republican.budget	4.703074	2.796763	1.682	0.0939 .
proceed.execut	-1.899363	1.378498	-1.378	0.1695
public.health	-0.419803	0.875328	-0.480	0.6319
shut.govern	1.165442	0.785958	1.483	0.1394
republican.leadership	5.316522	4.422592	1.202	0.2305
web.site	2.160258	1.294029	1.669	0.0963 .
titl.bill	NA	NA	NA	NA
tar.sand	NA	NA	NA	NA
two.leader	-0.917251	1.125652	-0.815	0.4159
colleagu.join	0.293676	0.326193	0.900	0.3688
conduct.hear	NA	NA	NA	NA
presid.obama	0.040407	0.418098	0.097	0.9231
equal.pay	2.512750	3.783421	0.664	0.5072
middl.class	-1.379461	2.050407	-0.673	0.5017
vote.right	NA	NA	NA	NA
sea.level	NA	NA	NA	NA
extran.materi	NA	NA	NA	NA
senat.resum	NA	NA	NA	NA
open.govern	1.025633	2.280962	0.450	0.6534
move.proceed	0.563385	0.552326	1.020	0.3087
republican.leader	-0.936835	1.063025	-0.881	0.3790

work.hard	0.711391	0.669120	1.063	0.2887
head.start	NA	NA	NA	NA
farm.bill	-0.568476	0.417949	-1.360	0.1750
fair.shot	NA	NA	NA	NA
radic.islam	0.528593	1.074243	0.492	0.6231
deepest.congratul	NA	NA	NA	NA
previou.order	0.872017	0.578017	1.509	0.1327
scout.america	NA	NA	NA	NA
ask.colleagu	-0.422312	0.372446	-1.134	0.2579
session.consider	NA	NA	NA	NA
legisl.session	-0.442615	0.552326	-0.801	0.4237
equal.divid	0.988126	0.592256	1.668	0.0965 .
join.recogn	-0.379652	0.364191	-1.042	0.2982
nation.debt	-0.906129	1.927384	-0.470	0.6387
consid.made	-0.034000	0.775478	-0.044	0.9651
govern.shutdown	4.148302	6.472443	0.641	0.5222
illeg.alien	NA	NA	NA	NA
made.laid	NA	NA	NA	NA
reconsid.consider	NA	NA	NA	NA
democrat.republican	0.030928	0.383464	0.081	0.9358
al.qaeda	-0.434563	0.485617	-0.895	0.3717
senat.republican	-0.951461	1.956219	-0.486	0.6271
peopl.disabl	-0.831087	2.270720	-0.366	0.7147
loan.debt	-2.564107	3.456221	-0.742	0.4589
energi.effici	-0.667539	0.624715	-1.069	0.2863
republican.friend	-1.608006	2.183533	-0.736	0.4622
balanc.time	0.310416	0.174599	1.778	0.0766 .
repeal.afford	1.929877	2.419491	0.798	0.4258
fairfax.counti	-1.773257	1.512168	-1.173	0.2420
work.togeth	-0.210940	0.341643	-0.617	0.5375
work.famili	0.200701	0.384637	0.522	0.6023
let.us	0.213827	0.469828	0.455	0.6494
go.confer	-4.630040	4.561507	-1.015	0.3111
enforc.law	-1.562333	1.358011	-1.150	0.2511
pay.bill	0.112970	0.456272	0.248	0.8047
rule.law	-0.092545	0.837068	-0.111	0.9121
follow.leader	1.046929	1.654664	0.633	0.5275
women.health	0.876283	0.744640	1.177	0.2404
bill.hr	-0.495788	0.523759	-0.947	0.3448
took.offic	-1.783603	3.233274	-0.552	0.5817
care.plan	-0.305381	0.797772	-0.383	0.7022
futur.accomplish	NA	NA	NA	NA

leader.remark	NA	NA	NA	NA
health.center	-3.268217	2.757033	-1.185	0.2370
permit.speak	NA	NA	NA	NA
presid.promis	0.413385	0.552326	0.748	0.4549
senat.democrat	-0.086992	0.737965	-0.118	0.9063
hous.republican	-3.407865	2.682317	-1.270	0.2051
ad.debt	NA	NA	NA	NA
exhibit.dedic	NA	NA	NA	NA
senat.permit	NA	NA	NA	NA
tax.increas	0.887737	3.232041	0.275	0.7838
unemploy.benefit	NA	NA	NA	NA
proceed.vote	NA	NA	NA	NA
speaker.rise	0.406607	0.221872	1.833	0.0680
background.check	1.196771	0.886408	1.350	0.1782
reserv.balanc	-0.111641	0.209866	-0.532	0.5952
end.hunger	-3.846369	4.284987	-0.898	0.3702
student.debt	-0.120687	1.570931	-0.077	0.9388
busi.senat	-0.341436	0.579832	-0.589	0.5565
million.american	-0.368236	0.452431	-0.814	0.4165
agre.motion	0.094976	0.797168	0.119	0.9053
rais.tax	-0.160248	2.316262	-0.069	0.9449
busi.record	-0.798547	1.030833	-0.775	0.4393
paus.recogn	NA	NA	NA	NA
preexist.condit	-1.156907	2.149540	-0.538	0.5909
scout.earn	NA	NA	NA	NA
educ.develop	NA	NA	NA	NA
consent.request	NA	NA	NA	NA
train.wreck	3.180416	3.298476	0.964	0.3359
republican.senat	-2.282729	2.081235	-1.097	0.2738
pass.clean	NA	NA	NA	NA
radic.islamist	NA	NA	NA	NA
make.sure	0.294924	0.290491	1.015	0.3110
dedic.demonstr	NA	NA	NA	NA
south.carolina	-0.610009	0.599534	-1.017	0.3099
presid.unit	0.481593	0.691625	0.696	0.4869
follow.senat	NA	NA	NA	NA
attain.hard	NA	NA	NA	NA
kill.american	0.795558	0.726667	1.095	0.2747
tax.loophol	-2.231351	1.679362	-1.329	0.1852
overcom.advers	NA	NA	NA	NA
make.educ	NA	NA	NA	NA
leader.designe	NA	NA	NA	NA

proudli.ask	NA	NA	NA	NA
ve.ad	NA	NA	NA	NA
higher.educ	0.445309	0.425046	1.048	0.2958
clean.energi	0.222593	0.995852	0.224	0.8233
live.extend	NA	NA	NA	NA
level.rise	NA	NA	NA	NA
build.conduct	NA	NA	NA	NA
defeat.previou	-2.118204	0.952194	-2.225	0.0270 *
health.care	-0.430197	0.271002	-1.587	0.1137
determin.hard	NA	NA	NA	NA
paycheck.fair	NA	NA	NA	NA
extend.deepest	0.013127	0.829680	0.016	0.9874
finest.qualiti	NA	NA	NA	NA
across.countri	0.085758	0.242183	0.354	0.7236
right.act	NA	NA	NA	NA
earn.prestigi	NA	NA	NA	NA
economi.children	NA	NA	NA	NA
medic.devic	NA	NA	NA	NA
budget.amend	-0.095724	0.615024	-0.156	0.8764
young.peopl	1.287475	0.579904	2.220	0.0273 *
famili.peer	NA	NA	NA	NA
rest.live	12.850789	6.106433	2.104	0.0363 *
princ.william	1.449218	2.144052	0.676	0.4997
commun.notabl	NA	NA	NA	NA
commun.eagl	NA	NA	NA	NA
highest.distinct	NA	NA	NA	NA
mr.chairman	0.177996	0.127025	1.401	0.1624
commun.across	1.084673	0.612780	1.770	0.0779 .
proceed.consider	-0.739109	0.688075	-1.074	0.2838
work.allow	1.579146	2.383577	0.663	0.5083
motion.order	NA	NA	NA	NA
revis.extend	-0.559268	0.391508	-1.428	0.1544
special.young	NA	NA	NA	NA
troop.earn	NA	NA	NA	NA
januari.day	NA	NA	NA	NA
extend.remark	NA	NA	NA	NA
histori.month	-0.301713	0.962410	-0.313	0.7542
domest.violenc	0.042484	1.148154	0.037	0.9705
activ.troop	NA	NA	NA	NA
reduc.deficit	0.384518	3.439656	0.112	0.9111
develop.work	NA	NA	NA	NA
strive.make	NA	NA	NA	NA

debt.nation	NA	NA	NA	NA
religi.freedom	NA	NA	NA	NA
lo.angel	0.004933	0.501076	0.010	0.9922
public.servic	-0.110813	0.357931	-0.310	0.7571
extrem.weather	NA	NA	NA	NA
particip.mani	NA	NA	NA	NA
keep.doctor	NA	NA	NA	NA
food.tabl	NA	NA	NA	NA
suspend.rule	NA	NA	NA	NA
today.ve	NA	NA	NA	NA
snap.benefit	5.440010	2.990994	1.819	0.0701 .
america.troop	NA	NA	NA	NA
ask.immedi	1.218641	0.729859	1.670	0.0962 .
ryan.budget	NA	NA	NA	NA
speak.minut	0.418580	0.359635	1.164	0.2456
prescript.drug	NA	NA	NA	NA
year.trillion	1.194149	4.526420	0.264	0.7921
balanc.budget	NA	NA	NA	NA
extend.unemploy	NA	NA	NA	NA
america.effort	NA	NA	NA	NA
debt.today	NA	NA	NA	NA
activ.part	1.067816	0.555318	1.923	0.0556 .
leadership.take	NA	NA	NA	NA
previou.question	1.272867	0.767515	1.658	0.0985 .
speaker.januari	-1.479789	1.264659	-1.170	0.2431
black.caucu	NA	NA	NA	NA
new.mexico	NA	NA	NA	NA
cut.snap	NA	NA	NA	NA
take.activ	NA	NA	NA	NA
energi.natur	NA	NA	NA	NA
broken.immigr	-0.107309	2.036358	-0.053	0.9580
proceed.period	NA	NA	NA	NA
consent.adjourn	-1.100797	0.786560	-1.400	0.1629
hous.resolut	-0.316562	0.487403	-0.649	0.5166
reopen.govern	-1.447248	2.680611	-0.540	0.5898
institut.health	2.478563	2.192348	1.131	0.2593
drone.strike	NA	NA	NA	NA
american.citizen	-0.293958	0.725587	-0.405	0.6857
school.receiv	0.062104	0.521896	0.119	0.9054
pm.senat	NA	NA	NA	NA
koch.brother	NA	NA	NA	NA
respect.famili	0.321812	1.089836	0.295	0.7680

resum.legisl	NA	NA	NA	NA
rule.xxii	-0.454226	0.584325	-0.777	0.4377
purpos.debat	NA	NA	NA	NA
work.persever	NA	NA	NA	NA
proceed.calendar	-1.019000	0.775478	-1.314	0.1900
achiev.highest	NA	NA	NA	NA
heritag.month	NA	NA	NA	NA
tax.credit	NA	NA	NA	NA
activ.mani	NA	NA	NA	NA
program.help	-1.093933	1.650468	-0.663	0.5081
homeland.secur	0.185336	0.490575	0.378	0.7059
pleas.contact	NA	NA	NA	NA
remark.senat	-0.785874	0.667483	-1.177	0.2402
tax.cut	-1.170277	6.361649	-0.184	0.8542
notifi.senat	NA	NA	NA	NA
reconsid.laid	NA	NA	NA	NA
prior.vote	NA	NA	NA	NA
congression.black	NA	NA	NA	NA
earn.numer	NA	NA	NA	NA
debt.year	NA	NA	NA	NA
prestigi.award	-0.325412	0.560529	-0.581	0.5621
come.togeth	-0.101863	0.447227	-0.228	0.8200
obama.took	NA	NA	NA	NA
also.respect	NA	NA	NA	NA
nation.institut	NA	NA	NA	NA
agre.preambl	NA	NA	NA	NA
republican.shutdown	NA	NA	NA	NA
presid.move	0.379588	0.337249	1.126	0.2614
date.time	NA	NA	NA	NA
invest.infrastructur	0.353220	0.909897	0.388	0.6982
recogn.applaud	NA	NA	NA	NA
resum.consider	NA	NA	NA	NA
counti.receiv	NA	NA	NA	NA
senat.juli	NA	NA	NA	NA
rule.pass	NA	NA	NA	NA
york.citi	-0.001588	1.378269	-0.001	0.9991
law.school	-1.006073	0.551456	-1.824	0.0693 .
pm.room	NA	NA	NA	NA
today.honor	-0.392796	0.237279	-1.655	0.0991 .
resolut.agre	-0.457921	0.553956	-0.827	0.4092
everi.day	-0.184594	0.261636	-0.706	0.4811
drink.water	0.750007	0.677355	1.107	0.2692

hous.adjourn	NA	NA	NA	NA
consent.resolut	NA	NA	NA	NA
american.energi	2.373312	4.500092	0.527	0.5984
minut.gentlewoman	0.065234	0.394749	0.165	0.8689
preambl.agre	NA	NA	NA	NA
immedi.notifi	NA	NA	NA	NA
proceed.approv	NA	NA	NA	NA
reserv.use	NA	NA	NA	NA
sexual.orient	NA	NA	NA	NA
journal.proceed	NA	NA	NA	NA
prayer.pledg	NA	NA	NA	NA
xxii.waiv	0.151976	0.797168	0.191	0.8490
american.disabl	NA	NA	NA	NA
disabl.act	NA	NA	NA	NA
member.legisl	NA	NA	NA	NA
ask.join	NA	NA	NA	NA
make.america	1.819997	1.563526	1.164	0.2455
leader.reserv	NA	NA	NA	NA
control.two	NA	NA	NA	NA
northern.virginia	-0.193040	1.473614	-0.131	0.8959
war.terror	1.769199	1.350022	1.310	0.1912
increas.minimum	NA	NA	NA	NA
middl.east	-0.248337	0.309866	-0.801	0.4236
foreign.relat	-0.266840	0.431201	-0.619	0.5366
presid.said	-0.941145	1.335645	-0.705	0.4817
baptist.church	-0.086896	0.497479	-0.175	0.8615
child.care	NA	NA	NA	NA
senat.complet	NA	NA	NA	NA
debat.motion	NA	NA	NA	NA
approv.date	NA	NA	NA	NA
goe.say	0.617385	0.552326	1.118	0.2647
could.avoid	NA	NA	NA	NA
republican.refus	NA	NA	NA	NA
immigr.law	1.659948	2.018016	0.823	0.4115
pay.equal	NA	NA	NA	NA
affair.author	-0.351256	0.861181	-0.408	0.6837
emerg.unemploy	NA	NA	NA	NA
obamacar.exchang	0.235897	1.603098	0.147	0.8831
depart.homeland	0.271783	0.775722	0.350	0.7264
time.equal	NA	NA	NA	NA
month.recogn	-1.096700	1.114822	-0.984	0.3262
follow.prayer	NA	NA	NA	NA

great.lake	0.405579	0.916865	0.442	0.6586
equal.work	NA	NA	NA	NA
use.later	NA	NA	NA	NA
jefferson.counti	NA	NA	NA	NA
presid.say	-0.739765	1.175644	-0.629	0.5298
million.children	NA	NA	NA	NA
day.follow	NA	NA	NA	NA
oil.compani	0.002268	1.076955	0.002	0.9983
complet.busi	NA	NA	NA	NA
state.congress	0.652310	0.476928	1.368	0.1726
senat.april	NA	NA	NA	NA
senat.action	NA	NA	NA	NA
year.old	0.077940	0.367973	0.212	0.8324
govern.open	-0.230741	1.467766	-0.157	0.8752
sandi.hook	NA	NA	NA	NA
commun.health	NA	NA	NA	NA
assault.weapon	NA	NA	NA	NA
put.food	NA	NA	NA	NA
right.movement	NA	NA	NA	NA
individu.mandat	-3.180875	1.660885	-1.915	0.0566 .
execut.branch	0.792307	0.872418	0.908	0.3647
day.presid	0.586455	1.001479	0.586	0.5587
california.ms	-0.019989	0.820508	-0.024	0.9806
keyston.xl	-0.662387	2.708454	-0.245	0.8070
democrat.friend	NA	NA	NA	NA
american.women	-0.045091	0.725896	-0.062	0.9505
legal.statu	NA	NA	NA	NA
deem.expir	NA	NA	NA	NA
strong.opposit	NA	NA	NA	NA
senat.mikulski	-0.364615	0.393348	-0.927	0.3548
children.could	NA	NA	NA	NA
hunger.america	NA	NA	NA	NA
busi.come	NA	NA	NA	NA
adjourn.previou	NA	NA	NA	NA
expir.journal	NA	NA	NA	NA
hour.deem	NA	NA	NA	NA
xl.pipelin	NA	NA	NA	NA
rise.today	-0.175883	0.208505	-0.844	0.3997
move.forward	0.031277	0.251270	0.124	0.9010
effort.put	NA	NA	NA	NA
pledg.morn	NA	NA	NA	NA
big.govern	-0.234675	0.651611	-0.360	0.7190

keep.health	1.177221	1.218375	0.966	0.3349
senat.period	NA	NA	NA	NA
broken.promis	-1.157600	2.691656	-0.430	0.6675
gender.ident	NA	NA	NA	NA
account.work	NA	NA	NA	NA
clean.continu	NA	NA	NA	NA
budget.confer	NA	NA	NA	NA
continu.resolut	-2.603858	4.012891	-0.649	0.5170
econom.develop	-0.217262	0.504741	-0.430	0.6672
food.stamp	NA	NA	NA	NA
war.coal	NA	NA	NA	NA
medic.research	-0.682351	1.239055	-0.551	0.5823
buy.gun	NA	NA	NA	NA
district.texa	0.141362	0.478684	0.295	0.7680
mental.health	-0.104897	0.497944	-0.211	0.8333
sexual.assault	-0.794146	0.458860	-1.731	0.0847 .
divid.control	-1.013525	0.898667	-1.128	0.2605
gentleman.pennsylvania	0.341507	0.232972	1.466	0.1439
therein.minut	NA	NA	NA	NA
women.famili	NA	NA	NA	NA
littl.guy	NA	NA	NA	NA
senat.pass	-0.426097	0.457812	-0.931	0.3529
energi.product	NA	NA	NA	NA
prevent.servic	NA	NA	NA	NA
consent.mandatori	NA	NA	NA	NA
speaker.presid	NA	NA	NA	NA
bill.consider	NA	NA	NA	NA
second.read	-0.455615	0.552326	-0.825	0.4102
speak.therein	NA	NA	NA	NA
announc.committe	NA	NA	NA	NA
bay.area	1.811436	0.805082	2.250	0.0253 *
presid.follow	NA	NA	NA	NA
mandatori.quorum	NA	NA	NA	NA
go.colleg	NA	NA	NA	NA
offic.nation	-0.588805	1.020628	-0.577	0.5645
vote.motion	NA	NA	NA	NA
time.two	NA	NA	NA	NA
western.new	0.650284	0.956617	0.680	0.4973
pass.comprehens	NA	NA	NA	NA
new.tax	0.859019	0.651672	1.318	0.1887
like.health	NA	NA	NA	NA
vote.clean	NA	NA	NA	NA

vladimir.putin	0.598970	0.617525	0.970	0.3330
colleg.afford	NA	NA	NA	NA
intern.revenu	NA	NA	NA	NA
hispan.heritag	NA	NA	NA	NA
permiss.revis	NA	NA	NA	NA
puerto.rico	NA	NA	NA	NA
us.attorney	0.703167	0.483688	1.454	0.1473
tax.hike	NA	NA	NA	NA
hurrican.sandi	-0.476058	0.712110	-0.669	0.5044
new.regul	NA	NA	NA	NA
faith.credit	NA	NA	NA	NA
implement.obamacar	1.941598	1.160872	1.673	0.0957 .
clean.bill	NA	NA	NA	NA
energi.secur	1.364637	0.968482	1.409	0.1601
faith.execut	NA	NA	NA	NA
food.assist	NA	NA	NA	NA
gentlewoman.california	-0.456166	0.466313	-0.978	0.3289
like.announc	NA	NA	NA	NA
busi.today	NA	NA	NA	NA
devic.tax	NA	NA	NA	NA
renew.energi	NA	NA	NA	NA
global.war	NA	NA	NA	NA
increas.spend	NA	NA	NA	NA
feder.govern	-0.266851	0.304276	-0.877	0.3813
new.haven	NA	NA	NA	NA
promis.american	NA	NA	NA	NA
caucu.meet	NA	NA	NA	NA
water.infrastructur	NA	NA	NA	NA
proceed.legisl	NA	NA	NA	NA
border.secur	2.205341	1.532027	1.439	0.1513
legal.immigr	NA	NA	NA	NA
harri.reid	NA	NA	NA	NA
given.permiss	NA	NA	NA	NA
oklahoma.mr	NA	NA	NA	NA
fifth.amend	NA	NA	NA	NA
democrat.colleagu	NA	NA	NA	NA
abl.keep	NA	NA	NA	NA
motion.proceed	NA	NA	NA	NA
mateo.counti	NA	NA	NA	NA
god.bless	-0.200683	0.874910	-0.229	0.8188
keep.govern	NA	NA	NA	NA
diseas.control	NA	NA	NA	NA

social.justic	-0.547185	0.693314	-0.789	0.4307
ask.given	NA	NA	NA	NA
speaker.behalf	0.816079	0.553956	1.473	0.1420
record.presid	NA	NA	NA	NA
vote.confirm	-0.165007	0.561002	-0.294	0.7689
day.within	NA	NA	NA	NA
straw.purchas	NA	NA	NA	NA
madam.chairman	0.662712	0.271192	2.444	0.0152 *
full.particip	NA	NA	NA	NA
quorum.rule	NA	NA	NA	NA
protect.public	NA	NA	NA	NA
print.record	0.085080	0.333086	0.255	0.7986
b.outlay	NA	NA	NA	NA
guest.worker	NA	NA	NA	NA
senat.collin	0.375491	0.791515	0.474	0.6356
new.england	NA	NA	NA	NA
kansa.mr	0.610009	0.900259	0.678	0.4987
feder.land	0.186887	0.393820	0.475	0.6355
district.georgia	-0.491921	0.553956	-0.888	0.3754
de.moin	0.336591	0.741191	0.454	0.6501
ye.rollcal	NA	NA	NA	NA
infrastructur.invest	NA	NA	NA	NA
across.america	0.055805	0.361478	0.154	0.8774
superstorm.sandi	NA	NA	NA	NA
local.economi	NA	NA	NA	NA
author.b	0.078834	0.840787	0.094	0.9254
home.state	0.229558	0.322153	0.713	0.4768
quorum.call	-0.120110	0.242731	-0.495	0.6212

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5483 on 250 degrees of freedom

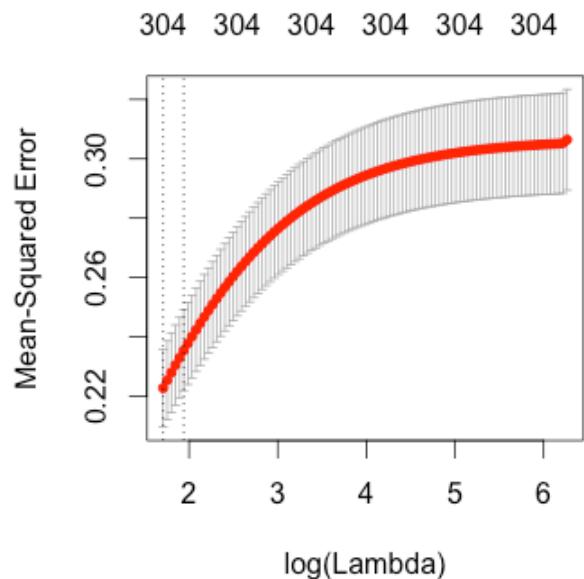
Multiple R-squared: 0.5236, Adjusted R-squared: 0.0491

F-statistic: 1.103 on 249 and 250 DF, p-value: 0.2186

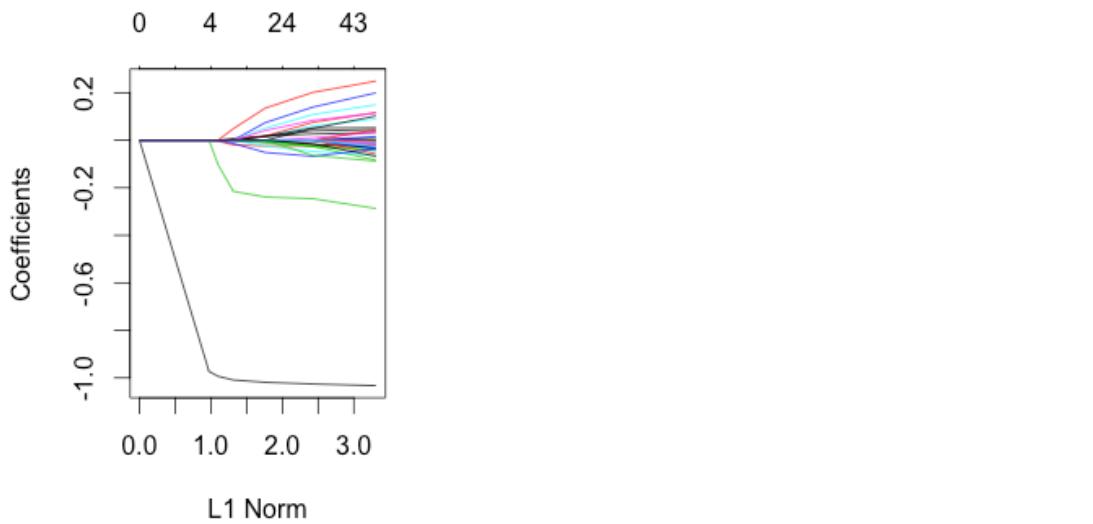
2 - Ridged

```
> out=glmnet(x,y,alpha=0)
> predict(out,type="coefficients",s=bestlam)[1:20,]
  (Intercept) speaker_partyD speaker_partyI
  0.0868079655 -0.0857478339  0.0000000000
  speaker_partyR   climat.chang      mr.presid
```

0.0857477294	-0.0040521634	-0.0095258438
unanim.consent	presid.ask	minimum.wage
-0.0111628599	-0.0115607724	-0.0360394301
care.law	ask.unanim	senat.proceed
0.0109106062	-0.0120100353	-0.0215779642
republican.colleagu	session.senat	meet.session
-0.0185957708	-0.0156655247	-0.0156654596
rhode.island	author.meet	obama.administr
-0.0007068405	-0.0156654351	0.0025567446
eagl.scout	new.york	
0.0180866225	-0.0258908013	



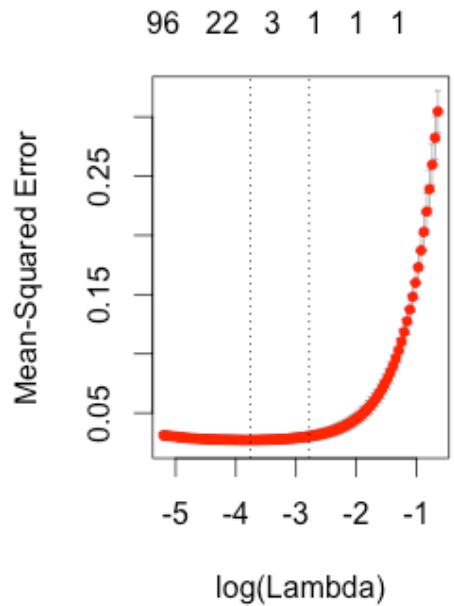
3 – Lasso



```

> out=glmnet(x,y,alpha=1,lambda=grid)
> lasso.coef=predict(out,type="coefficients",s=bestlam)[1:20,]
> lasso.coef
            (Intercept) speaker_partyD speaker_partyI
6.440303e-01      -1.039351e+00 0.000000e+00
speaker_partyR      climat.chang     mr.presid
2.798155e-18      0.000000e+00 -1.168799e-02
unanim.consent     presid.ask     minimum.wage
0.000000e+00      0.000000e+00 0.000000e+00
care.law           ask.unanim    senat.proceed
0.000000e+00      0.000000e+00 0.000000e+00
republican.colleagu session.senat   meet.session
0.000000e+00      0.000000e+00 0.000000e+00
rhode.island        author.meet   obama.administr
0.000000e+00      0.000000e+00 0.000000e+00
eagl.scout          new.york
0.000000e+00      0.000000e+00

```



#Top 100 republican and democrat bigrams and trigrams

Top 100 republican and democrat bigrams and trigrams							
bigram_democrat	bigram_r_freq	bigram_republican	bigram_d_freq	trigram_democrat	trigram_d_freq	trigram_republican	trigram_r_freq
mr speaker	18811	mr speaker	15319	ask unanim consent	6009	ask unanim consent	2900
health care	6047	mr presid	9043	presid ask unanim	3968	mr speaker rise	2691
american peopl	5351	unanim consent	6633	mr speaker rise	3867	speaker rise today	2286
mr presid	4758	ask unanim	6011	mr presid ask	3305	reserv balanc time	1688
balanc time	3353	health care	4640	speaker rise today	3103	yield back balanc	1467
unanim consent	3186	presid ask	4198	afford care act	2586	back balanc time	1466
mr chairman	2992	speaker rise	4017	rise today honor	1092	health care law	1451
rise today	2921	rise today	3829	yield back balanc	1086	presid ask unanim	1342
presid obama	2905	american peopl	3357	back balanc time	1086	afford care act	1254
ask unanim	2903	make sure	3272	meet session senat	1075	mr presid ask	1114
speaker rise	2828	new york	3155	ask colleague join	1061	yield minut gentleman	1079
feder govern	2791	madam presid	2928	author meet session	1030	presid unit state	800
make sure	2242	care act	2689	quorum call rescind	1017	speaker ask unanim	764
major leader	2140	afford care	2646	unanim consent order	1014	unanim consent order	652
reserv balanc	1701	balanc time	2197	order quorum call	1003	quorum call rescind	645
madam presid	1644	climat chang	2095	depart homeland secur	1002	rise today recogn	637
white hous	1633	feder govern	2065	consent order quorum	1001	order quorum call	633
presid ask	1560	mr chairman	2045	rise today recogn	965	consent order quorum	632
care law	1530	across countri	2001	reserv balanc time	961	suggest absenc quorum	599
new york	1464	homeland secur	1952	unanim consent senat	942	revis extend remark	597
care act	1395	work togeth	1825	suggest absenc quorum	881	unanim consent member	573
obama administr	1350	minimum wage	1822	madam presid ask	860	rise today honor	566
afford care	1292	presid obame	1721	unanim consent committe	832	depart homeland secur	549
homeland secur	1253	million american	1701	yield minut gentleman	753	mr speaker move	469
across countri	1230	colleagu join	1677	senat offic build	708	clerk read titl	467
work togeth	1004	farm bill	1632	interven action debat	646	legisl day revis	463
million american	970	move forward	1511	speaker ask colleagu	623	day revis extend	461
move forward	962	student loan	1476	laid upon tabl	584	health care plan	425
everi day	868	immigr reform	1457	consent senat proceed	575	presid health care	424
web site	857	ask colleagu	1445	comprehens immigr reform	571	keyston xl pipelin	418
farm bill	807	everi day	1439	dirksen senat offic	542	may legisl day	411
presid unit	805	middl class	1403	tabl interven action	510	member may legisl	411
colleagu join	792	senat proceed	1263	rais minimum wage	487	consent member may	410
border secur	737	work hard	1221	conduct hear entitl	462	unit state congress	398
middl east	727	today honor	1202	colleagu join recogn	461	boy scout america	394
eagl scout	710	session senat	1144	upon tabl interven	456	speaker reserv balanc	373
south carolina	694	civil right	1111	period morn busi	444	ask colleagu join	372
today honor	685	quorum call	1101	unanim consent request	429	includ extran materi	360
ask colleagu	683	meet session	1075	new york citi	394	suspend rule pass	356
balanc budget	680	govern shutdown	1054	consid made laid	391	remark includ extran	355
quorum call	675	rollical vote	1045	reconsid consid made	390	extend remark includ	353
rule law	669	author meet	1035	motion reconsid consid	390	madam presid ask	353
middl class	625	republican colleagu	1018	made laid upon	381	health human servic	350
illeg immigr	624	home state	1018	nation institut health	359	necessarli absent senat	349
legisl day	614	depart homeland	1003	proceed execut session	352	mr speaker reserv	337
home state	610	rhode island	998	yield minut gentlewoman	331	rule pass bill	318
senat democrat	600	major leader	995	presid unit state	325	wall street journal	296
revis extend	600	come togeth	995	broken immigr system	325	rise today support	294
extend remark	599	hous republican	992	senat permit speak	322	pass bill hr	290
nation debt	587	reserv balanc	963	student loan debt	314	presid obama took	237
consent member	573	young peopl	960	morn busi senat	304	balance budget amend	233
american citizen	573	consent senat	955	new budget author	302	ask immedi consider	232
depart homeland	550	sexual assault	954	agr motion reconsid	301	obama took offic	231
bill hr	540	public servic	951	offic build conduct	299	continu reserv balanc	216
sexual assault	534	interest rate	945	energi natur resourc	299	intern revenu servic	212
al qaeda	529	contine resolut	944	senat proceed vote	298	mr speaker januari	201
boy scout	528	presid offic	930	year new budget	291	proudli ask join	196
continu resolut	525	democrat republican	869	fiscal year new	291	day presid obama	186
speaker move	521	let us	835	execut session consid	289	take activ part	185
come togeth	519	tax credit	834	author b outlay	288	special young man	184
work hard	518	consent committe	833	budget author b	288	scout america troop	182
execut branch	515	morn busi	808	repeat afford care	282	unit state code	182
enfor law	512	unemploy insur	801	senat proceed consider	282	ask join command	181
clerk read	498	african american	797	follow leader remark	281	mr speaker behalf	175
keyston xl	493	shut govern	766	demand record vote	274	global war terror	171
presid said	485	offic build	764	mr presid would	273	friend across aisn	167
read titl	469	public health	757	year fiscal year	269	unanim consent senat	164
student loan	465	senat offic	755	fiscal year fiscal	266	member legisl day	161
day revis	463	food stamp	716	outlay fiscal year	264	consent member legisl	159
public servic	459	year old	703	b outlay fiscal	261	medic devic tax	159
young peopl	454	higher educ	694	build conduct hear	260	senat south carolina	158
extran materi	452	motion reconsid	680	vote right act	256	senat major leader	155
hous republican	440	action debat	652	would vote ye	243	mr speaker presid	149
presid health	439	interven action	647	senat proceed execut	241	move hous adjourn	149
employ mandat	438	comprehens immigr	639	speaker ask unanim	240	open amend process	145
care plan	436	energi effici	635	new york mr	240	secretari homeland secur	145
member may	433	republican leader	629	leader remark senat	239	unanim consent committe	143
trillion debt	433	go confer	629	keep govern open	237	speaker move hous	141
titl bill	432	print record	628	commite energi natur	232	major leader senat	139
govern shutdown	431	gun violenc	624	motion reconsid laid	231	meet session senat	137
xl pipelin	425	tea parti	612	martin luther king	230	colleagu hous join	135
immigr reform	417	white hous	609	two leader designe	229	author meet session	133
state congress	417	work famili	609	put food tabl	229	gentleman pennsylvania mr	132
feder land	416	senat republican	600	foreign relat committie	225	speaker ask colleagu	131
may legisl	413	mental health	595	increas minimum wage	223	date enact act	129
immigr law	404	pay bill	594	revis extend remark	217	bill hr amend	127
tax credit	399	upon tabl	593	consent resolut agre	214	mr presid would	126
suspend rule	398	background check	589	unanim consent resolut	214	new york citi	123
scout america	396	laid upon	586	mr presid move	214	hous join congratul	121
religi freedom	396	hear entitl	569	yield back time	214	speaker continu reserv	120
rais tax	392	domest violenc	563	immigr reform bill	212	resolut ask immedi	116
tax increas	390	republican budget	549	follow prayer pledg	208	like health care	114
follow senat	374	dirksen senat	547	equal divid control	208	keep health care	112
rule pass	370	unemploy benefit	538	colleagu join honor	208	mr speaker privileg	112
took offic	365	econom develop	524	senat ask unanim	206	promis american peopl	110
muslim brotherhood	361	equal divid	522	journal proceed approv	205	south carolina mr	110
includ extran	360	head start	518	time two leader	205	demand record vote	109
food stamp	358	speaker boehner	515	resolut agre preamb	205	would vote ye	108
remark includ	355	senat pass	514	proceed approv date	204	unanim consent request	107
absent senat	354	san francisco	511	reserv use later	204	delay employ mandat	107