

## Modeling assignment

Mengfan Cheng  
04/13/2015

The model is about the relationship between congressmen's speeches and their ideologies. We are trying to parse and analyze the ngrams of congressmen's speeches to see the top 100 bigrams and trigrams for both Republicans and Democrats. Also, a model is to be built approaching the relationship between words and ideology. Here are the exact steps:

Steps:

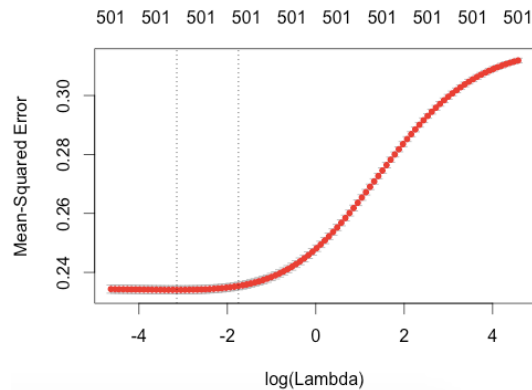
- 1) Download the data of the most recent Congress via Sunlight Foundation API (credits to Dr. Gaurav Sood <https://gist.github.com/soodoku/85d79275c5880f67b4cf>)
- 2) Clean the speeches in the downloaded dataset: remove stop words, stem the words, remove special characters, numbers, capitalization etc (credits to Dr. Gaurav Sood <https://github.com/soodoku/Text-as-Data>)
- 3) Download ideology data from voteview.com (DW-Nominate, 1st dimension)
- 4) Merge the two datasets by ID#
- 5) Produce a list of all possible 'bigrams' and 'trigrams' in the data.
- 6) Produce a list of the top 100 bigrams and trigrams for both Republicans and Democrats
- 7) Build a ridge model about the relationship between people's words and their ideology.

Results:

- 1) Top 500 bigrams and trigrams as listed in the Appendix Top Bigrams and Top Trigrams
- 2) Top 100 bigrams and trigrams for both Republicans and Democrats as listed in Appendix Top Bigrams of Republicans, Top Trigrams of Republicans, Top Bigrams of Democrats and Top Trigrams of Democrats.

- 3) Ridge Model:

As is shown in the last part of R script, top bigrams are checked in each person's speech to see whether they have involved certain topic in their speeches. Based on this, a ridge model is created where the dependent variable is the ideology score and the independent variables are whether people's speeches include the top 500 bigrams. And cross-validation is adopted to tune the model. In cross validation, it is found that the best  $\lambda = 0.04307892$ . To further expand the model, top 500 trigrams and other n-grams may also be adopted.



```
> predict(out,type="coefficients",s=bestlam)[1:20,]  
(Intercept)      Name_Party      climat.chang      mr.presid  
-0.0006955137  0.0002955027 -0.2168090075 -0.0831234149  
unanim.consent  presid.ask      minimum.wage      care.law  
-0.0611493362  0.0009078932 -0.0991155242  0.0944811657  
ask.unanim      senat.proceed  republican.colleagu session.senat  
-0.0255829898 -0.0717401725 -0.1419230142 -0.0371710179  
meet.session    rhode.island    author.meet      obama.administr  
-0.0531362061 -0.1447501142 -0.0803064003  0.0984196321  
eagl.scout      new.york        gun.violenc      immigr.reform  
0.1352411791 -0.1114423172 -0.2352216012 -0.1046244924
```

### Appendix Top Bigrams

	X	bigram	bigram_d_fre q	bigram_r_fre q	chisq()
	84688	84688 climat chang	2095	256	7.87E-05
	287967	287967 mr presid	9043	4758	7.28E-05
	372713	372713 unanim consent	6633	3186	6.62E-05
	313121	313121 presid ask	4198	1560	6.61E-05
	285218	285218 minimum wage	1822	264	6.37E-05
	69090	69090 care law	157	1530	6.11E-05
	36441	36441 ask unanim	6011	2903	5.93E-05
	341982	341982 senat proceed	1263	162	4.65E-05
	329692	329692 republican colleagu	1018	78	4.41E-05
	344487	344487 session senat	1144	157	4.10E-05
	282638	282638 meet session	1075	137	3.97E-05
	333274	333274 rhode island	998	116	3.82E-05
	40413	40413 author meet	1035	134	3.80E-05
	295470	295470 obama administr	285	1350	3.80E-05
	146896	146896 eagl scout	15	710	3.64E-05
	292709	292709 new york	3155	1464	3.39E-05
	215612	215612 gun violenc	624	15	3.18E-05
	234998	234998 immigr reform	1457	417	3.16E-05
	101901	101901 consent member	9	573	2.99E-05
	341908	341908 senat offic	755	81	2.97E-05
	356406	356406 student loan	1476	465	2.88E-05
	6177	6177 action debat	652	55	2.76E-05
	249237	249237 interven action	647	54	2.74E-05
	101972	101972 consent senat	955	196	2.74E-05
	101813	101813 consent committe	833	143	2.67E-05
	82628	82628 civil right	1111	291	2.62E-05
	383893	383893 white hous	609	1633	2.56E-05
	12660	12660 afford care	2646	1292	2.55E-05
	323232	323232 read titl	3	469	2.52E-05
	26272	26272 american peopl	3357	5351	2.50E-05
	296724	296724 offic build	764	131	2.45E-05
	68462	68462 carbon pollut	484	15	2.41E-05
	123218	123218 day revis	8	463	2.40E-05
	13355	13355 african american	797	154	2.38E-05
	138852	138852 dirksen senat	547	43	2.36E-05
	84578	84578 clerk read	24	498	2.35E-05
	275961	275961 major leader	995	2140	2.29E-05
	287240	287240 motion reconsid	680	108	2.27E-05
	68618	68618 care act	2689	1395	2.24E-05
	313394	313394 presid health	11	439	2.23E-05

360066	360066	tabl interven	510	38	2.22E-05
265810	265810	legisl day	85	614	2.19E-05
373284	373284	unemploy insur	801	176	2.19E-05
233932	233932	illeg immigr comprehens	90	624	2.18E-05
96778	96778	immigr	639	98	2.17E-05
280455	280455	may legisl	6	413	2.16E-05
350361	350361	speaker move	50	521	2.13E-05
28195	28195	announc senat	395	3	2.11E-05
168835	168835	execut session	417	11	2.11E-05
334873	334873	rollcal vote	1045	324	2.08E-05
283097	283097	member may	20	433	2.06E-05
154683	154683	employ mandat	22	438	2.06E-05
391862	391862	youth award	385	3	2.06E-05
322084	322084	rais minimum	493	46	2.03E-05
220060	220060	hear entitl	569	79	2.03E-05
374485	374485	upon tabl	593	94	1.98E-05
274814	274814	madam presid	2928	1644	1.97E-05
288033	288033	mr speaker muslim	15319	18811	1.96E-05
288768	288768	brotherhood	3	361	1.93E-05
260574	260574	laid upon	586	97	1.92E-05
336987	336987	san francisco	511	65	1.89E-05
328081	328081	remark includ	4	355	1.88E-05
238751	238751	includ extran	6	360	1.87E-05
248396	248396	interest rate	945	296	1.86E-05
313569	313569	presid offic	930	289	1.84E-05
1216	1216	absent senat	6	354	1.84E-05
58917	58917	boy scout	77	528	1.84E-05
287018	287018	morn busi	808	221	1.83E-05
361939	361939	tea parti	612	119	1.82E-05
350112	350112	speaker boehner	515	73	1.82E-05
371079	371079	trillion debt	38	433	1.81E-05
307484	307484	period morn	444	43	1.81E-05
361483	361483	tax break	394	23	1.81E-05
329669	329669	republican budget	549	90	1.80E-05
320305	320305	public health	757	197	1.80E-05
315651	315651	proceed execut	352	8	1.80E-05
345841	345841	shut govern republican	766	203	1.79E-05
329789	329789	leadership	457	50	1.79E-05
382072	382072	web site	255	857	1.78E-05
367280	367280	titl bill	40	432	1.78E-05

361341	361341	tar sand	336	4	1.77E-05
372145	372145	two leader	475	60	1.76E-05
87172	87172	colleagu join	1677	792	1.74E-05
98519	98519	conduct hear	486	72	1.68E-05
313564	313564	presid obama	1721	2905	1.66E-05
161482	161482	equal pay	359	21	1.64E-05
284058	284058	middl class	1403	625	1.63E-05
379089	379089	vote right	470	71	1.61E-05
339081	339081	sea level	308	6	1.59E-05
172501	172501	extran materi	65	452	1.58E-05
342061	342061	senat resum	377	34	1.57E-05
299177	299177	open govern	400	44	1.56E-05
287486	287486	move proceed	338	19	1.56E-05
329788	329788	republican leader	629	156	1.56E-05
386500	386500	work hard	1221	518	1.55E-05
219200	219200	head start	518	102	1.53E-05
178005	178005	farm bill	1632	807	1.53E-05
175679	175679	fair shot	342	23	1.53E-05
321947	321947	radic islam	4	290	1.52E-05
128234	128234	deepest congratul	281	1	1.52E-05
314309	314309	previou order	382	41	1.50E-05
35952	35952	ask colleagu	1445	683	1.49E-05
339026	339026	scout america	48	396	1.49E-05
344462	344462	session consid	291	7	1.48E-05
266254	266254	legisl session	318	17	1.48E-05
161417	161417	equal divid	522	110	1.47E-05
256252	256252	join recogn	511	105	1.46E-05
289568	289568	nation debt	146	587	1.45E-05
209937	209937	govern shutdown	1054	431	1.43E-05
102703	102703	consid made	393	52	1.43E-05
233867	233867	illeg alien	3	269	1.42E-05
275083	275083	made laid	422	66	1.42E-05
325452	325452	reconsid consid	390	53	1.40E-05
		democrat			
131476	131476	republican	869	318	1.40E-05
17974	17974	al qaeda	124	529	1.37E-05
342051	342051	senat republican	600	163	1.37E-05
305824	305824	peopl disabl	307	21	1.36E-05
271547	271547	loan debt	341	37	1.34E-05
157146	157146	energi effici	635	189	1.32E-05
329746	329746	republican friend	291	18	1.32E-05
44641	44641	balanc time	2197	3353	1.32E-05
328557	328557	repeal afford	286	17	1.31E-05

175745	175745	fairfax counti	256	6	1.30E-05
387061	387061	work togeth	1825	1004	1.30E-05
386405	386405	work famili	609	180	1.28E-05
267152	267152	let us	835	317	1.27E-05
205133	205133	go confer	629	192	1.27E-05
157673	157673	enforc law	129	512	1.25E-05
304798	304798	pay bill	594	176	1.24E-05
335240	335240	rule law	222	669	1.23E-05
53612	53612	bill hr	148	540	1.22E-05
188970	188970	follow leader	283	22	1.22E-05
385484	385484	women health	283	22	1.22E-05
368807	368807	took offic	59	365	1.21E-05
69252	69252	care plan	94	436	1.21E-05
198235	198235	futur accomplish	223	1	1.20E-05
263699	263699	leader remark	286	26	1.19E-05
219379	219379	health center	263	17	1.18E-05
307639	307639	permit speak	339	50	1.17E-05
313631	313631	presid promis	5	228	1.17E-05
230152	230152	hous republican	992	440	1.16E-05
341511	341511	senat democrat	190	600	1.16E-05
7914	7914	ad debt	6	230	1.16E-05
169189	169189	exhibit dedic	221	3	1.16E-05
341951	341951	senat permit	326	46	1.15E-05
361594	361594	tax increas	77	390	1.15E-05
373259	373259	unemploy benefit	538	158	1.13E-05
315697	315697	proceed vote	329	49	1.13E-05
350484	350484	speaker rise	4017	2828	1.13E-05
330773	330773	reserv balanc	963	1701	1.12E-05
44016	44016	background check	589	190	1.12E-05
156321	156321	end hunger	218	5	1.11E-05
356307	356307	student debt	237	12	1.11E-05
64317	64317	busi senat	346	60	1.10E-05
284730	284730	million american	1701	970	1.09E-05
15822	15822	agre motion	305	42	1.09E-05
322132	322132	rais tax	84	392	1.09E-05
64253	64253	busi record	7	219	1.09E-05
304759	304759	paus recogn	3	207	1.08E-05
312658	312658	preexist condit	498	142	1.08E-05
339032	339032	scout earn	2	203	1.08E-05
149620	149620	educ develop	217	8	1.06E-05
101958	101958	consent request	450	118	1.06E-05
369957	369957	train wreck	10	222	1.06E-05
329895	329895	republican senat	508	151	1.06E-05

276867	276867	make sure	3272	2242	1.05E-05
303806	303806	pass clean	218	9	1.05E-05
321948	321948	radic islamist	4	204	1.05E-05
127718	127718	dedic demonstr	195	1	1.05E-05
349663	349663	south carolina	265	694	1.05E-05
313839	313839	presid unit	337	805	1.05E-05
189132	189132	follow senat	80	374	1.04E-05
38983	38983	attain hard	192	1	1.03E-05
258308	258308	kill american	24	252	1.03E-05
361623	361623	tax loophol	227	14	1.03E-05
301802	301802	overcom advers	197	3	1.03E-05
276397	276397	make educ	208	7	1.03E-05
263412	263412	leader designe	251	24	1.03E-05
319278	319278	proudli ask	3	196	1.02E-05
376954	376954	ve ad	3	196	1.02E-05
223438	223438	higher educ	694	270	1.02E-05
83722	83722	clean energi	280	37	1.02E-05
270908	270908	live extend	192	2	1.02E-05
267615	267615	level rise	189	1	1.02E-05
62259	62259	build conduct	299	46	1.01E-05
219376	219376	health care	4640	6047	1.01E-05
128454	128454	defeat previou	191	2	1.01E-05
135090	135090	determin hard	196	4	1.01E-05
305147	305147	paycheck fair	233	18	1.01E-05
172013	172013	extend deepest	335	64	1.01E-05
183907	183907	finest qualiti	1	187	1.01E-05
4546	4546	across countri	2001	1230	1.01E-05
333379	333379	right act	410	103	1.01E-05
147444	147444	earn prestigi	3	192	1.00E-05
148626	148626	economi children	3	192	1.00E-05
282029	282029	medic devic	67	338	9.92E-06
61403	61403	budget amend	25	247	9.91E-06
391808	391808	young peopl	960	454	9.91E-06
177110	177110	famili peer	1	183	9.85E-06
332192	332192	rest live	221	15	9.84E-06
314560	314560	princ william	199	7	9.79E-06
287669	287669	mr chairman	2045	2992	9.74E-06
94001	94001	commun notabl	2	184	9.74E-06
93568	93568	commun eagl	1	181	9.74E-06
223596	223596	highest distinct	1	181	9.74E-06
93248	93248	commun across	328	64	9.73E-06
315648	315648	proceed consider	297	49	9.72E-06
386079	386079	work allow	216	14	9.70E-06

287234	287234	motion order	189	4	9.70E-06
333216	333216	revis extend	219	600	9.70E-06
350773	350773	special young	3	186	9.69E-06
371194	371194	troop earn	2	183	9.69E-06
254206	254206	januari day	6	194	9.67E-06
172148	172148	extend remark	219	599	9.66E-06
224892	224892	histori month	242	25	9.65E-06
143348	143348	domest violenc	563	197	9.64E-06
7135	7135	activ troop	3	185	9.64E-06
325887	325887	reduc deficit	446	128	9.64E-06
136132	136132	develop work	211	13	9.57E-06
355887	355887	strive make	208	12	9.55E-06
125723	125723	debt nation	14	213	9.54E-06
327810	327810	religi freedom	102	396	9.49E-06
271490	271490	lo angel	448	131	9.49E-06
320414	320414	public servic	951	459	9.39E-06
172756	172756	extrem weather	231	23	9.32E-06
303326	303326	particip mani	10	198	9.30E-06
257526	257526	keep doctor	6	187	9.29E-06
189485	189485	food tabl	294	52	9.26E-06
359267	359267	suspend rule	106	398	9.25E-06
368023	368023	today ve	11	199	9.21E-06
348184	348184	snap benefit	219	19	9.19E-06
25289	25289	america troop	5	182	9.16E-06
36125	36125	ask immedi	30	244	9.14E-06
335794	335794	ryan budget	253	34	9.14E-06
349947	349947	speak minut	360	87	9.12E-06
312884	312884	prescript drug	348	81	9.09E-06
391097	391097	year trillion	27	236	9.09E-06
44463	44463	balanc budget	281	680	9.06E-06
172185	172185	extend unemploy	323	69	9.00E-06
24560	24560	america effort	5	179	9.00E-06
125866	125866	debt today	17	210	8.98E-06
6985	6985	activ part	12	197	8.96E-06
264225	264225	leadership take	10	191	8.92E-06
314311	314311	previou question	455	143	8.91E-06
350296	350296	speaker januari	14	201	8.90E-06
55880	55880	black caucu	193	11	8.88E-06
292506	292506	new mexico	402	113	8.87E-06
120416	120416	cut snap	168	2	8.87E-06
360163	360163	take activ	12	195	8.85E-06
157279	157279	energi natur	302	60	8.85E-06
60690	60690	broken immigr	336	78	8.80E-06

315682	315682	proceed period	199	14	8.79E-06
101782	101782	consent adjourn	172	4	8.77E-06
230156	230156	hous resolut	51	280	8.67E-06
328519	328519	reopen govern	323	73	8.63E-06
246190	246190	institut health	366	96	8.63E-06
145714	145714	drone strike	11	188	8.61E-06
25586	25586	american citizen	221	573	8.54E-06
338705	338705	school receiv	223	26	8.53E-06
309923	309923	pm senat	204	18	8.52E-06
260193	260193	koch brother	181	9	8.52E-06
331487	331487	respect famili	11	186	8.50E-06
332607	332607	resum legisl	167	4	8.50E-06
335436	335436	rule xxii	201	17	8.50E-06
320636	320636	purpos debat	3	164	8.49E-06
386783	386783	work persever	198	16	8.47E-06
315643	315643	proceed calendar	160	2	8.43E-06
3969	3969	achiev highest	12	187	8.42E-06
222776	222776	heritag month	168	5	8.40E-06
361521	361521	tax credit	834	399	8.40E-06
6942	6942	activ mani	14	191	8.36E-06
317042	317042	program help	413	126	8.36E-06
226683	226683	homeland secur	1952	1253	8.34E-06
309758	309758	pleas contact	172	7	8.32E-06
328123	328123	remark senat	329	80	8.29E-06
361523	361523	tax cut	274	52	8.27E-06
294016	294016	notifi senat	168	6	8.25E-06
325455	325455	reconsid laid	231	32	8.24E-06
314710	314710	prior vote	196	17	8.23E-06
101093	101093	congression black	178	10	8.21E-06
147421	147421	earn numer	10	178	8.21E-06
125905	125905	debt year	12	183	8.20E-06
89967	89967	come togeth	995	519	8.19E-06
313989	313989	prestigi award	46	260	8.19E-06
295643	295643	obama took	36	238	8.15E-06
21986	21986	also respect	17	194	8.12E-06
289779	289779	nation institut	419	133	8.11E-06
15881	15881	agre preamb	228	32	8.08E-06
		republican			
329904	329904	shutdown	162	5	8.07E-06
313539	313539	presid move	282	58	8.07E-06
122010	122010	date time	221	30	7.95E-06
249896	249896	invest infrastructur	188	16	7.93E-06
324987	324987	recogn applaud	285	61	7.93E-06



332606	332606	resum consider	211	26	7.90E-06
111892	111892	counti receiv	152	3	7.84E-06
341752	341752	senat juli	149	2	7.83E-06
335294	335294	rule pass	109	370	7.78E-06
391652	391652	york citi	394	123	7.77E-06
262701	262701	law school	436	148	7.77E-06
367637	367637	today honor	1202	685	7.75E-06
309922	309922	pm room	172	11	7.75E-06
330961	330961	resolut agre	271	56	7.73E-06
165640	165640	everi day	1439	868	7.73E-06
145462	145462	drink water	288	65	7.71E-06
229410	229410	hous adjourn	44	246	7.70E-06
101961	101961	consent resolut	216	30	7.69E-06
25776	25776	american energi minut	52	261	7.63E-06
285427	285427	gentlewoman	340	94	7.63E-06
312580	312580	preambl agre	219	32	7.62E-06
234580	234580	immedi notifi	164	9	7.60E-06
315638	315638	proceed approv	207	27	7.57E-06
330840	330840	reserv use	209	28	7.56E-06
344944	344944	sexual orient	197	23	7.53E-06
256465	256465	journal proceed	206	27	7.52E-06
312568	312568	prayer pledg	208	28	7.51E-06
389676	389676	xxii waiv	143	2	7.50E-06
25725	25725	american disabl	172	13	7.48E-06
138867	138867	disabl act	156	7	7.45E-06
283071	283071	member legisl	14	174	7.45E-06
36148	36148	ask join	51	255	7.44E-06
276206	276206	make america	369	113	7.44E-06
263707	263707	leader reserv	204	27	7.42E-06
107343	107343	control two	147	4	7.41E-06
293704	293704	northern virginia	173	14	7.40E-06
380513	380513	war terror	46	244	7.40E-06
240373	240373	increas minimum	225	37	7.38E-06
284061	284061	middl east	347	727	7.36E-06
190638	190638	foreign relat	384	123	7.35E-06
313706	313706	presid said	185	485	7.35E-06
45132	45132	baptist church	448	162	7.34E-06
78611	78611	child care	306	79	7.32E-06
341441	341441	senat complet	213	32	7.31E-06
125171	125171	debat motion	156	8	7.31E-06
32388	32388	approv date	204	28	7.30E-06
207201	207201	goe say	45	240	7.30E-06

110132	110132	could avoid	29	206	7.29E-06
329877	329877	republican refus	142	3	7.29E-06
234899	234899	immigr law	136	404	7.28E-06
304859	304859	pay equal	150	6	7.27E-06
12114	12114	affair author	201	27	7.26E-06
154228	154228	emerg unemploy	218	35	7.24E-06
132275	132275	depart homeland	1003	550	7.23E-06
295702	295702	obamacar exchang	2	138	7.23E-06
366419	366419	time equal	172	15	7.21E-06
286850	286850	month recogn	143	4	7.19E-06
189066	189066	follow prayer	208	31	7.17E-06
211991	211991	great lake	324	91	7.16E-06
161547	161547	equal work	153	8	7.14E-06
376055	376055	use later	205	30	7.13E-06
254508	254508	jefferson counti	163	12	7.13E-06
313711	313711	presid say	58	262	7.11E-06
284766	284766	million children	246	50	7.10E-06
122762	122762	day follow	213	34	7.10E-06
297637	297637	oil compani	193	25	7.08E-06
95970	95970	complet busi	204	30	7.08E-06
353008	353008	state congress	147	417	7.07E-06
341321	341321	senat april	135	2	7.06E-06
341276	341276	senat action	181	20	7.05E-06
390683	390683	year old	703	337	7.05E-06
209672	209672	govern open	393	133	7.03E-06
337094	337094	sandi hook	194	26	7.02E-06
93750	93750	commun health	187	23	7.01E-06
320962	320962	put food	242	49	7.00E-06
36832	36832	assault weapon	153	9	7.00E-06
333809	333809	right movement	211	34	7.00E-06
242100	242100	individu mandat	88	315	6.99E-06
168720	168720	execut branch	211	515	6.96E-06
123132	123132	day presid	45	233	6.95E-06
65242	65242	california ms	172	17	6.95E-06
258107	258107	keyston xl	197	493	6.95E-06
131278	131278	democrat friend	4	138	6.92E-06
26732	26732	american women	196	28	6.89E-06
265601	265601	legal statu	74	287	6.88E-06
128115	128115	deem expir	193	27	6.85E-06
355994	355994	strong opposit	210	35	6.84E-06
341854	341854	senat mikulski	273	67	6.83E-06
79040	79040	children could	31	201	6.81E-06
231960	231960	hunger america	133	3	6.80E-06

63699	63699	busi come	217	39	6.77E-06
9657	9657	adjourn previou	186	25	6.72E-06
171250	171250	expir journal	190	27	6.70E-06
229073	229073	hour deem	190	27	6.70E-06
389669	389669	xl pipelin	158	425	6.69E-06
334240	334240	rise today	3829	2921	6.68E-06
287392	287392	move forward	1511	962	6.67E-06
151613	151613	effort put	32	200	6.66E-06
309876	309876	pledg morn	189	27	6.65E-06
52629	52629	big govern	11	151	6.62E-06
257581	257581	keep health	27	188	6.60E-06
341949	341949	senat period	188	27	6.60E-06
60720	60720	broken promis	12	153	6.59E-06
200235	200235	gender ident	126	2	6.57E-06
3676	3676	account work	145	9	6.57E-06
83712	83712	clean continu	142	8	6.55E-06
61491	61491	budget confer	187	27	6.54E-06
106009	106009	continu resolut	944	525	6.54E-06
148265	148265	econom develop	524	225	6.53E-06
189477	189477	food stamp	716	358	6.53E-06
380336	380336	war coal	9	144	6.52E-06
282091	282091	medic research	279	74	6.51E-06
64623	64623	buy gun	133	5	6.49E-06
141814	141814	district texa	14	156	6.49E-06
283553	283553	mental health	595	274	6.49E-06
344934	344934	sexual assault	954	534	6.49E-06
142037	142037	divid control	211	39	6.47E-06
		gentleman			
201361	201361	pennsylvania	49	231	6.47E-06
363839	363839	therein minut	192	30	6.47E-06
385455	385455	women famili	183	26	6.45E-06
341937	341937	senat pass	514	220	6.44E-06
270588	270588	littl guy	19	167	6.44E-06
157333	157333	energi product	68	266	6.42E-06
314244	314244	prevent servic	157	15	6.41E-06
101899	101899	consent mandatori	134	6	6.40E-06
350424	350424	speaker presid	19	166	6.39E-06
53202	53202	bill consider	28	186	6.38E-06
339287	339287	second read	156	15	6.36E-06
350012	350012	speak therein	192	31	6.36E-06
28095	28095	announc committe	119	1	6.35E-06
46737	46737	bay area	172	22	6.34E-06
313347	313347	presid follow	176	24	6.32E-06

277633	277633	mandatori quorum	140	9	6.30E-06
205107	205107	go colleg	276	75	6.30E-06
296998	296998	offic nation	41	211	6.27E-06
378948	378948	vote motion	286	81	6.26E-06
367117	367117	time two	242	57	6.26E-06
383534	383534	western new	144	11	6.24E-06
303812	303812	pass comprehens	163	19	6.23E-06
292660	292660	new tax	47	222	6.23E-06
269065	269065	like health	15	153	6.20E-06
378697	378697	vote clean	130	6	6.18E-06
378441	378441	vladimir putin	37	201	6.18E-06
87823	87823	colleg afford	145	12	6.16E-06
248940	248940	intern revenu	77	276	6.14E-06
224338	224338	hispan heritag	161	19	6.13E-06
307585	307585	permiss revis	172	24	6.11E-06
320505	320505	puerto rico	233	54	6.11E-06
374700	374700	us attorney	355	124	6.09E-06
361587	361587	tax hike	13	146	6.09E-06
232123	232123	hurrican sandi	205	40	6.08E-06
292592	292592	new regul	20	162	6.06E-06
175845	175845	faith credit	166	22	6.03E-06
		implement			
235954	235954	obamacar	9	135	6.03E-06
157382	157382	energi secur	87	291	6.02E-06
83702	83702	clean bill	152	16	6.02E-06
175867	175867	faith execut	10	137	6.00E-06
189271	189271	food assist	172	25	6.00E-06
		gentlewoman			
201469	201469	california	213	45	5.98E-06
268758	268758	like announc	118	3	5.98E-06
64426	64426	busi today	232	55	5.97E-06
136201	136201	devic tax	31	184	5.96E-06
328482	328482	renew energi	361	130	5.95E-06
204781	204781	global war	26	173	5.94E-06
240580	240580	increas spend	26	173	5.94E-06
179788	179788	feder govern	2065	2791	5.94E-06
292422	292422	new haven	133	9	5.92E-06
317978	317978	promis american	11	138	5.92E-06
71672	71672	caucu meet	114	2	5.92E-06
381050	381050	water infrastructur	159	20	5.90E-06
315666	315666	proceed legisl	130	8	5.90E-06
58192	58192	border secur	389	737	5.88E-06
265529	265529	legal immigr	67	252	5.87E-06

218695	218695	harri reid	96	302	5.83E-06
204368	204368	given permiss	173	27	5.83E-06
297751	297751	oklahoma mr	18	153	5.83E-06
181306	181306	fifth amend	7	126	5.82E-06
131185	131185	democrat colleagu	67	251	5.82E-06
732	732	abl keep	50	219	5.81E-06
287238	287238	motion proceed	351	126	5.81E-06
279816	279816	mateo counti	109	1	5.80E-06
206866	206866	god bless	106	318	5.80E-06
257572	257572	keep govern	449	189	5.80E-06
140394	140394	diseas control	189	35	5.79E-06
348282	348282	social justic	138	12	5.79E-06
36088	36088	ask given	176	29	5.77E-06
350102	350102	speaker behalf	29	176	5.77E-06
325606	325606	record presid	130	9	5.76E-06
378716	378716	vote confirm	244	64	5.75E-06
123485	123485	day within	2	111	5.75E-06
355600	355600	straw purchas	149	17	5.74E-06
274812	274812	madam chairman	98	303	5.73E-06
196476	196476	full particip	116	4	5.72E-06
321843	321843	quorum rule	110	2	5.70E-06
318999	318999	protect public	189	36	5.69E-06
314665	314665	print record	628	315	5.68E-06
42876	42876	b outlay	288	90	5.67E-06
215278	215278	guest worker	15	143	5.67E-06
341432	341432	senat collin	218	51	5.67E-06
292370	292370	new england	216	50	5.67E-06
257405	257405	kansa mr	3	112	5.65E-06
179834	179834	feder land	170	416	5.65E-06
141513	141513	district georgia	186	35	5.64E-06
123741	123741	de moin	42	200	5.64E-06
389773	389773	ye rolcal	167	26	5.64E-06
243925	243925	infrastructur invest	117	5	5.62E-06
4512	4512	across america	468	205	5.62E-06
357811	357811	superstorm sandi	171	28	5.62E-06
271805	271805	local economi	250	69	5.62E-06
40090	40090	author b	288	91	5.60E-06
226537	226537	home state	1018	610	5.59E-06
321840	321840	quorum call	1101	675	5.59E-06

## Appendix Top 100 Bigrams of Republicans

mr speaker	18811
unit state	9670
health care	6047
american peopl	5351
mr presid	4758
balanc time	3353
unanim consent	3186
ca nt	3111
mr chairman	2992
rise today	2921
presid obama	2905
ask unanim	2903
would like	2831
speaker rise	2828
feder govern	2791
last year	2330
make sure	2242
major leader	2140
madam speaker	1941
speaker yield	1842
year ago	1818
small busi	1765
high school	1749
reserv balanc	1701
yield back	1673
side aisl	1671
madam presid	1644
white hous	1633
hous repres	1621
yield floor	1599
men women	1579
urg colleagu	1571
presid ask	1560
yield minut	1548
care law	1530
back balanc	1481
new york	1464
health insur	1426
care act	1395
obama administr	1350
nation secur	1320
afford care	1292

nt want	1277
homeland secur	1253
law enforc	1246
nt know	1245
last week	1234
yield time	1231
across countri	1230
time may	1186
air forc	1165
wo nt	1163
may consum	1160
act chair	1150
minut gentleman	1149
fiscal year	1123
suprem court	1091
next year	1064
world war	1059
rank member	1056
pass bill	1042
thank gentleman	1039
hard work	1032
bill would	1021
want thank	1015
member	
congress	1008
work togeth	1004
speaker ask	988
natur ga	982
social secur	977
million american	970
move forward	962
creat job	958
million peopl	942
piec legisl	939
look forward	922
colleagu support	910
amend would	904
even though	885
one thing	870
everi day	868
war ii	866
web site	857
would say	848

would vote	846
nt get	823
nt think	818
would nt	810
around world	809
farm bill	807
presid unit	805
today recogn	798
colleagu join	792
mani year	774
mani peopl	773
washington dc	771
offer amend	770
congression	
district	761
come floor	760
re go	755



## Appendix Top 100 Trigrams of Republicans

ask unanim consent	2900
mr speaker rise	2691
speaker rise today	2286
reserv balanc time	1688
mr speaker yield	1551
yield back balanc	1467
back balanc time	1466
health care law	1451
presid ask unanim	1342
afford care act	1254
time may consum	1157
mr presid ask	1114
yield minut gentleman	1079
yield time may	1032
mr speaker ask	882
world war ii	849
presid unit state	800
speaker ask unanim	764
unit state america	725
urg colleagu support	673
speaker yield time	662
unanim consent order	652
quorum call rescind	645
rise today recogn	637
order quorum call	633
consent order quorum	632
suggest absenc	
quorum	599
revis extend remark	597
speaker yield minut	592
unanim consent	
member	573
rise today honor	566
congression budget	
offic	552
depart homeland	
secur	549
mr chairman yield	495
mr speaker move	469
present would vote	468
clerk read titl	467

legisl day revis	463
day revis extend	461
health care plan	425
presid health care	424
keyston xl pipelin	418
may legisl day	411
member may legisl	411
consent member may	410
mr speaker proudli	409
colleagu side aisl	402
unit state congress	398
boy scout america	394
read titl bill	393
speaker reserv balanc	373
ask colleagu join	372
mr speaker would	369
includ extran materi	360
suspend rule pass	356
remark includ extran	355
extend remark includ	353
madam presid ask	353
move suspend rule	353
speaker move suspend	351
health human servic	350
follow senat	
necessarili	349
necessarili absent	
senat	349
senat necessarili	
absent	349
mr speaker today	346
floor suggest absenc	344
yield floor suggest	343
mr speaker reserv	337
health care system	331
arm servic committe	327
act chair gentleman	318
rule pass bill	318
new york time	315
speaker yield back	315
mr speaker thank	314
would like thank	312
amend offer	310

gentleman	
mr presid rise	305
may consum mr	298
wall street journal	296
rise today support	294
pass bill hr	290
presid yield floor	287
thank gentleman yield	286
mr speaker want	283
rise today congratul	281
mr chairman rise	280
madam speaker yield	279
unit state armi	276
chair question amend	275
question amend offer	271
time act chair	268
men women uniform	267
health care cost	259
balanc time act	256
consum mr speaker	253
budget control act	249
mr speaker pleas	248
look forward work	243
mr chairman amend	241

## Appendix Top 100 Bigrams of Democrats

mr speaker	15319
mr presid	9043
unit state	8402
unanim consent	6633
ask unanim	6011
health care	4640
presid ask	4198
speaker rise	4017
rise today	3829
american peopl	3357
make sure	3272
new york	3155
ca nt	2955
madam presid	2928
last year	2754
care act	2689
afford care	2646
would like	2451
high school	2336
year ago	2265
balanc time	2197
climat chang	2095
urg colleagu	2085
feder govern	2065
mr chairman	2045
across countri	2001
homeland secur	1952
hous repres	1920
yield floor	1911
side aisl	1875
work togeth	1825
minimum wage	1822
men women	1777
small busi	1759
presid obama	1721
health insur	1714
million american	1701
colleagu join	1677
speaker yield	1662
farm bill	1632
law enforc	1559

rank member	1557
move forward	1511
yield back	1507
yield minut	1479
student loan	1476
immigr reform	1457
ask colleagu	1445
bill would	1440
everi day	1439
fiscal year	1410
middl class	1403
hard work	1373
nation secur	1373
madam speaker	1363
social secur	1336
suprem court	1332
senat proceed	1263
million peopl	1247
work hard	1221
last week	1204
today honor	1202
session senat	1144
look forward	1142
civil right	1111
creat job	1104
quorum call	1101
back balanc	1090
nt know	1087
meet session	1075
govern shutdown	1054
rollcal vote	1045
new jersey	1036
author meet	1035
today recogn	1023
call rescind	1018
home state	1018
republican	
colleagu	1018
consent order	1016
absenc quorum	1015
speaker ask	1011
order quorum	1008
yield time	1005

want thank	1004
depart homeland	1003
rhode island	998
would vote	997
come floor	996
come togeth	995
major leader	995
act chair	992
hous republican	992
judiciari	
committe	977
nt want	975
wo nt	975
reserv balanc	963
young peopl	960
consent senat	955
sexual assault	954
around world	951

## Appendix Top 100 Trigrams of Democrats

ask unanim consent	6009
presid ask unanim	3968
mr speaker rise	3867
mr presid ask	3305
speaker rise today	3103
afford care act	2586
mr speaker yield	1410
rise today honor	1092
back balanc time	1086
yield back balanc	1086
meet session senat	1075
ask colleagu join	1061
author meet session	1030
quorum call rescind	1017
unanim consent order	1014
order quorum call	1003
depart homeland secur	1002
consent order quorum	1001
rise today recogn	965
reserv balanc time	961
mr speaker ask	954
unanim consent senat	942
time may consum	894
suggest absenc quorum	881
madam presid ask	860
unanim consent	
committe	832
yield time may	804
world war ii	776
yield minut gentleman	753
unit state america	717
senat offic build	708
interven action debat	646
speaker ask colleagu	623
urg colleagu support	610
colleagu side aisl	585
laid upon tabl	584
consent senat proceed	575
comprehens immigr	
reform	571
dirksen senat offic	542

speaker yield minut	524
tabl interven action	510
speaker yield time	499
rais minimum wage	487
conduct hear entitl	462
colleagu join recogn	461
upon tabl interven	456
mr presid rise	454
present would vote	445
period morn busi	444
floor suggest absenc	441
yield floor suggest	439
unanim consent request	429
new york citi	394
consid made laid	391
motion reconsid consid	390
reconsid consid made	390
ambassador youth award	382
arvada wheat ridg	382
ridg servic ambassador	382
servic ambassador youth	382
wheat ridg servic	382
made laid upon	381
mr chairman yield	369
mr speaker today	360
nation institut health	359
congression budget offic	358
proceed execut session	352
urg colleagu vote	351
look forward work	335
presid rise today	335
mr speaker pleas	332
yield minut gentlewoman	331
presid yield floor	329
broken immigr system	325
presid unit state	325
mr speaker thank	323
senat permit speak	322
want make sure	320
mr presid today	319
student loan debt	314
health care system	309
morn busi senat	304



new budget author	302
agre motion reconsid	301
energi natur resourc	299
offic build conduct	299
senat proceed vote	298
amend offer gentleman	293
mr presid yield	293
fiscal year new	291
year new budget	291
execut session consid	289
author b outlay	288
budget author b	288
today recogn applaud	283
repeal afford care	282
senat proceed consider	282
follow leader remark	281
extend deepest congratul	277
demand record vote	274