# **Online Appendix**

### Government websites as data:

# A methodological pipeline with application to the websites of municipalities in the United States

Markus Neumann\*

Fridolin Linder<sup>†</sup>

Bruce Desmarais<sup>‡</sup>

August 1, 2019

#### **Abstract**

A local government's website is an important source of information about policies and procedures for residents, community stakeholders and scholars. Existing research in public administration, public policy, and political science has relied on manual methods of website content collection and processing, limiting the scale and scope of website content analysis. We develop a methodological pipeline that researchers can follow in order to gather, process, and analyze website content. Our approach, which represents a considerable improvement in scalability, involves downloading the entire contents of a website, extracting the text and discarding redundant information. We provide an R package that can be used to apply our proposed pipeline. We illustrate our methodological pipeline through the collection and analysis of a new and innovative dataset—the websites of over two hundred municipal governments in the United States. We build upon recent research that analyzes how variation in the partisan control of government relates to content made available on the government's website. Using a structural topic model, we find that cities with Democratic mayors provide more information on policy deliberation and crime control, whereas Republicans prioritize basic utilities and services such as water, electricity and fire safety.

#### 1 Overview

In this online appendix we include supporting information about our data, the data collection process, and some additional analyses. In the first section, we present additional details on data

<sup>\*</sup>Department of Political Science, The Pennsylvania State University, University Park, PA 16802, USA. Email: mvn5218@psu.edu. Corresponding author.

<sup>&</sup>lt;sup>†</sup>Department of Political Science, Social Media and Political Participation Lab, New York University, New York, NY 10012, USA. Email: fridolin.linder@nyu.edu

<sup>&</sup>lt;sup>‡</sup>Department of Political Science, The Pennsylvania State University, University Park, PA 16802, USA. Email: bdes-marais@psu.edu

collection, along with some additional descriptive statistics. In the second section, we provide additional details on, and results from, our topic modeling analysis.

#### 2 Data collection methods and sources

We acquired the municipal website URLs from two sources: One, we scraped the URLs of city websites from their respective Wikipedia pages, which we found from lists of cities contained within each state. Two, the General Services Administration (GSA) maintains all '.gov' addresses, and provides a complete list of all such domains to the public. The data from the GSA contains the following variables: (1) domain name, specifically, the all-uppercase version of domain and top-level domain (for example, 'ABERDEENMD.GOV'); (2) the type of government entity to which the domain is registered, such as city, county, federal agency, etc; (3) for federal agencies, the name is specified; (4) the city in which the domain is registered. Naturally, the GSA's list does not contain cities which do not use a '.gov' website (or, in many cases, a city owns a registered '.gov' address, but uses a different one). Furthermore, some of the links are non-functional, and some of the county websites on the list are incorrectly marked as city websites (and vice versa). Since the GSA data is less complete and less reliable than the URLs found on Wikipedia, we mainly rely on the latter and only supplement them with the GSA data if a specific city doesn't have a URL recorded on Wikipedia, or our tests (see below) find it to be non-functional.

Not all of the URLs contained in these archives are functional. To test the URLs' functionality, we use a web driver-controlled browser - a browser that is automatically controlled by a program rather than a human user. We use the Python bindings for the program Selenium, which we use to control Firefox through the web driver Geckodriver. This is advantageous compared to conventional scraping tools such as Beautiful Soup or Rvest because most websites are

<sup>&</sup>lt;sup>1</sup>The dataset is made available at https://github.com/GSA/data/tree/gh-pages/dotgov-domains. This list is updated once per month—we rely on the version released on January 16, 2017.

designed to be explored by browsers. Modern browsers perform a lot of actions behind the scenes, such as URL resolution and redirection. The use of a web driver-controlled browser is necessary in our case because a) some city websites simply don't work, but they don't always output an error code correctly (this can fail, for example, if a webmaster simply stops maintaining a site without removing it entirely) which would throw off an automatic scraper, and more often, b) cities sometimes change their websites' URLs, in which case they redirect from the old to the new URL. A web driver-controlled browser, unlike the more rigid conventional scraping tools, will simply follow this redirection. This allows us to subsequently record and use the new URL for the actual website scraping. Consequently, an automated browser allows us to robustly answer the following questions: Is the website actually there? Does it work? If not, is it somewhere else or is it broken? We record this information and construct a list of verified URLs.

To download the websites, we rely on the Unix command line tool wget.<sup>2</sup> This program is used to download files from the Internet, and with the use of a recursive option, acts like a web crawler and scraper. This means that wget downloads HTML files, parses them and then follows the links contained therein. Then it follows those links and repeats the process until it has constructed a complete tree of the website (note that the program is instructed to stay on the same domain, i.e.

<sup>&</sup>lt;sup>2</sup>An alternative source of web data which has attained some popularity within the digital humanities is the Internet Archive's Wayback machine (https://archive.org/), which preserves snapshots of websites over time. In theory, this would be very useful to projects such as ours, since it would allow us to measure the evolution of websites over time, in response to changes in city executives. However, the Wayback Machine suffers from some limitations that, in our eyes, inhibit its usefulness for scientific research in general, and our project in particular. First of all, the Internet Archive does not conduct all of its web crawling itself. Rather, third parties donate their crawling data to the Internet Archive. This means that the source of the data varies and the crawling process often remains opaque, which is a problem with respect to scientific transparency. Second, this also means that the crawling isn't done in regular intervals unless the website in question is extremely prominent. Our research is focused on municipal websites, many of which are fairly obscure, preventing the Wayback Machine from being useful to us. For example, as of the time of writing, the website of Attica, IN, has only been crawled 25 times since the Wayback Machine's inception in 1996 (and in this case, all of these crawls are from 2014 on). By contrast, the website of New York City has been crawled over 7,000 times. Third, webcrawls are not instant and are not always done on an entire website at the same time. This means that for example, a website's frontpage might have been scraped one day, and its page on the city's mayor only a month later – within the same crawl. Once again, this problem is exacerbated for small, less prominent websites. While we appreciate the Internet Archive's efforts to preserve snapshots of the web and recommend its API to practitioners who can work around the limitations outlined here, these problems preclude its usefulness for our purposes.

it does not follow external links). This way, all the files that make up a website are downloaded. For some cities, whose websites make heavy use of JavaScript to serve content dynamically, such content is not reachable with our methodology and would require additional steps to obtain. For this paper, we ignore such sites and restricted our corpus to cities with at least three successfully downloaded pages.<sup>3</sup>

The partisanship of the mayor of each city is coded in different ways, depending on the state. For Indiana, where elections are nominally partisan, this information is accessible through the state government's website<sup>4</sup>. For Louisiana, we received data on the outcomes of mayoral elections from the Local Elections in America Project (LEAP) (Marschall and Shah 2013). For the other states, where mayoral elections are not nominally partisan (but the partisanship of the mayor is still well-known), we employed different means: For New York and Washington, we searched the state campaign finance websites, and coded the parties of the candidates based on the party committees from which they received donations. For California and Texas, where our data consists of highly populated cities, partisanship information was acquired from Ballotpedia<sup>5</sup>. Finally, we also scraped mayoral partisanship from the cities' Wikipedia pages. When compared to the other data sources above, (and manual searches in case of conflicts) Wikipedia proved to be very reliable and added additional cases to our dataset even for Indiana and Louisiana. Generally speaking, we found data scraped from Wikipedia, aided by manual corrections in case of missing or conflicting data, to be more reliable than data from governmental sources.<sup>6</sup>

Information on other covariates (population and median household income - from the American

<sup>&</sup>lt;sup>3</sup>There is a possibility that this leads to a small bias in selecting against cities with the resources to build more elaborate websites. However, given that our sample is generally more on the wealthy side, this, if anything, should lead to a more balanced sample.

<sup>&</sup>lt;sup>4</sup>http://www.in.gov/apps/sos/election/general/general2015?page=office&countyID=1&officeID=32&districtID=-1&candidate=

<sup>&</sup>lt;sup>5</sup>https://ballotpedia.org/List\_of\_current\_mayors\_of\_the\_top\_100\_cities\_in\_the\_United\_States

<sup>&</sup>lt;sup>6</sup>In Indiana, the data includes only cities - incorporated municipalities with at least 2,000 inhabitants - as opposed to towns.

Community Survey 5-Year Data (2015)) was acquired through the API of the U.S. Census Bureau<sup>7</sup>.

Tables 1 and 2 provide additional information about the data collected for this project. In Table 1, we present the state-by-state breakdown of the mayoral partisanship of the cities collected in the respective state. In Table 2, we present the distribution of file extensions before and after processing.

State	Democratic	Republican
California	9	6
Indiana	46	54
Louisiana	28	17
New York	36	16
Texas	2	7
Washington	11	2

Table 1: Descriptive statistics on the partisanship of the cities in the corpus.

Filetype	Occurances Before	Occurances After		
html	211682	887362		
pdf	464842	638802		
jpg	0	36958		
xml	0	29638		
Other	162681	9475		
ics	435	8950		
png	0	8863		
doc	6972	8430		
txt	317	6025		
	793990	5234		
docx	3137	4319		
TOTAL	1644056	1644056		

Table 2: Number of files per type, before and after detecing them via their magic number. The table shows that a lot of files originally have the wrong type, and that converting them correctly has a large impact on how many of them end up being usable.

<sup>&</sup>lt;sup>7</sup>https://www.census.gov/data/developers/data-sets.html

# 3 Supplemental Informational on Topic Modeling Application

The structural topic model is implemented in the R package STM (Roberts et al. 2018). We use 60 topics—the number recommended by the authors<sup>8</sup> for medium- to large-sized corpora.<sup>9</sup> We use four covariates: First, *party*, to estimate the difference in topic prevalence based on whether mayors are Republican or Democratic. Second, *city population*, which the literature frequently emphasizes as a determinant of the issues a city faces (see, for example, Guillamón et al. (2013)). Third, we control for wealth by relying on *median income* as a covariate, which we use as a proxy for the tax base in a city. Fourth and finally, we include state dummy variables, which should account for language that is associated with state-specific issues, and general background variables that vary across states.<sup>10</sup>

#### 3.1 Supplemental results

In Table 3, we present topic results from an STM, and mayoral party effect, estimated on data processed using a minimal HTML parser, implemented as the 'KeepEvertything' algorithm in the boilerpipeR package. An illustration of the content removed by the full boilerpipe procedure is given by Topic 56. This is the fourth-most Republican topic, and is a website boilerplate topic, with top words 'click', 'reserved', 'trademark', 'password', 'search', and 'connected'. As further illustration, the second-most Democratic topic is another website boilerplate topic, with top words 'sitemap', 'clerk', 'calendar', 'online', 'bureau', and 'alert'. We do not see any clear boilerplate topics when using data processed using boilerpipe.

In Tables 4 and 5 we present results of the STM with 120 topics organized according to the ef-

 $<sup>^8</sup>$ For this recommendation, see the documentation for the function stm() in version 1.3.0 of the R package stm (Roberts et al. 2018).

<sup>&</sup>lt;sup>9</sup>Since our corpus is at the larger end of that spectrum, we also estimated a model with 120 topics, but found no notable differences.

<sup>&</sup>lt;sup>10</sup>The "Fightin' Words" methodology developed by Monroe et al. (2008) could also be used to analyze word-frequency differences between cities based on mayors' partisanship, but we elected to use the structural topic model since, unlike "Fightin' Words", the structural topic model enables us to adjust for several other features through multiple regression.

fect of mayoral partisanship on topic prevalence. The partisan themes in the 120-topic STM mirror those in the 60-topic model, with cities led by Democratic mayors focused disproportionately on finances (e.g., Topics 75, 71, 61) and social problems (e.g., Topics 93, 101, 39, 52), and cities led by Republican mayors focused disproportionately on basic services and utilities (e.g., Topics 114, 11, 86, 116).

In Tables 6 and 7 we present the topics ordered according to the effects of median income and population, respectively. Considering income-based variation in topics, there are several topics prevalent in more wealthy cities that focus on initiatives that go well-beyond standard city services—downtown and building revitalization (Topics 39 and 12), renewable energy (Topic 20), bike/pedestrian-oriented development (Topic 60), wildlife conservation (Topic 3). Such topics cannot be found among those that are more prevalent in less wealthy cities. When it comes to population, more populous cities deal disproportionately with issues related to public health (Topics 2 and 11), crime (Topic 59), homelessness (Topic 52), and diversity (Topic 6).

## **Funding**

This work was supported by the National Science Foundation [1320219, 1637089, 1641047].

#### References

Guillamón, M. D., F. Bastida, and B. Benito (2013). The electoral budget cycle on municipal police expenditure. *European Journal of Law and Economics* 36(3), 447–469.

Marschall, M. and P. Shah (2013). Local elections in america project. *Center for Local Elections in American Politics. Kinder Institute for Urban Research, Rice University.* (Database).

Monroe, B. L., M. P. Colaresi, and K. M. Quinn (2008). Fightin' words: Lexical feature selection

and evaluation for identifying the content of political conflict. *Political Analysis 16*(4 SPEC. ISS.), 372–403.

Roberts, M. E., B. M. Stewart, and D. Tingley (2018). *stm: R Package for Structural Topic Models*. R package version 1.3.3.

#	Top Word 1	Top Word 2	Top Word 3	Top Word 4	Top Word 5	Top Word 6	Tokens assigned
30	riverfront	department	authority	transit	police	enjoy	212
55	posted	dream	celebration	broadcast	ballpark	football	544
5	trust	revocable	leisure	learn	mfr	living	327
56	click	reserved	trademark	password	search	connected	387
49	chair	agenda	subcommittee	briefing	presentation	committee	416
47	motion	second	adjourn	carry	whiting	unanimous	451
51	storm	drain	sanitary	water	sewer	infiltration	391
43	article	subsection	shall	provision	chapter	unlawful	467
2	virus	tuberculosis	infection	influenza	hepatitis	cannabis	2555
1	councilman	whereas	alderman	resolved	yea	resolution	576
59	subcontractor	proposer	bidder	bid	consultant	subcontract	508
33	eff	inf	effluent	ether	batch	isomer	1240
54	dwelling	alteration	plumbing	canceled	plumb	mechanical	311
12	think	something	somebody	appreciate	seem	everything	2911
21	artist	ceremony	jazz	celebrate	prize	yoga	3326
11	disaster	evacuation	marshal	apparatus	tornado	aircraft	1174
4	craftsman	architecture	facade	distinctive	architectural	historic	1634
34	contributor	filer	officeholder	political	payee	candidate	272
17	assessor	informal	taxpayer	doc	viewing	determination	442
18	setback	variance	plat	height	thence	frontage	472
19	findings	tank	carcinogen	string		yon	247
3	vend	meat	utensil	towel	qty fat	cheese	2791
38	application	applicant				mail	390
50 60		1.1	must	copy	tenant habitat		
	wetland	shoreline	vernal	riparian		marsh	1986
26	student	teacher	classroom	beech	academic	doe	832
7	obesity	sugary	epidemic	drink	sensible	ounce	98
20	garland	invoice	assoc	rouge	baton	vendor	524
31	credit	docket	bbl	agent	month	app	61
40	slideshow	arrow	printer	stumble	blogger	google	189
8	deductible	outpatient	prescription	coinsurance	copay	inpatient	809
41	playground	park	tennis	picnic	viewpoint	ravine	405
45	taxable	res	deed	value	homestead	star	106
36	thickness	fitting	conduit	conductor	ductile	trench	1756
22	noise	mitigation	impact	significant	adverse	sensitive	346
35	householder	universe	margin	poverty	race	census	251
15	imp	amt	micron	rend	land	sustain	117
16	dist	applied	col	occupancy	monoxide	valuation	123
58	pickup	bag	recyclable	bin	landfill	curbside	651
46	contracted	medicare	allocation	subtotal	unencumbered	payroll	236
14	perm	queue	delay	peak	adj	volume	232
39	successor	franchisee	covenant	redemption	bankruptcy	obligation	678
44	prune	circumference	tree	planting	shrub	root	1798
23	tax	increment	deduction	abatement	assessed	levy	397
6	fugitive	exhaust	renewable	bio	emission	coal	859
57	savings	costs	capital	ltd	improvement	excise	172
27	density	mixed	planned	infill	orient	retail	365
52	bicycle	pedestrian	bike	route	bus	curb	572
48	actuarial	asset	governmental	assets	investment	debt	342
37	supervisor	technician	aide	incumbent	employee	trainee	758
32	introduced	absent	councilor	preside	digest	legislator	615
28	persona	para	sin	ante	junta	combo	2376
24	audit	procedure	effectiveness	ensure	software	timely	497
42	budget	endorse	revenue	endorsed	balance	expenditure	232
13	homeless	affordable	affordability	housing	supportive	homelessness	380
10	impound	taxicab	license	cat	dog	neuter	700
9	strategy	stakeholder	focus	goal	engagement	outreach	732
25	profile	executive	bend	sustainability	cleanup	rates	110
53	theft	burglary	fwy	aggravated	aggravate	robbery	253
29	sitemap	clerk	calendar	online	bureau	alert	124
50	arrest	complainant	allegation	shooting	homicide	victim	1665
				9			

Table 3: Top words from a structural topic model with 60 topics and FREX scoring, with data processed using a minimal HTML parser. Colors depict partisanship based on coefficient size. White cells are non-significant topics.

#	Top Word 1	Top Word 2	Top Word 3	Top Word 4	Top Word 5	Top Word 6	Tokens assigned
96	subcommittee	agenda	forum	speaker	item	adjournment	217
49	prize	celebration	ceremony	parade	follower	favorite	2043
102	motion	second	adjourn	unanimous	carry	whiting	207
73	legislator	player	football	leg	town	stadium	695
95	online	email	website	browser	contact	server	351
70	election	ballot	lobbyist	voter	candidate	campaign	407
74	tentative	conditional	approval	grading	attachment	deviation	177
79	snow	remember	plow	lock	scam	sure	888
28	craftsman	revival	historic	gabled	bungalow	historical	882
114	park	playground	recreation	picnic	mesa	trail	235
11	tuberculosis	infection	hepatitis	overdose	influenza	vaccine	1515
21	think	something	want	thing	talk	everybody	1155
86	sewer	sanitary	water	pipeline	drinking	wastewater	176
59	fort	worth	plot	tad	falls	demo	192
20	subsection	licensee	article	chapter	sec	shall	214
47	inf	micron	effluent	eff	sludge	isomer	591
62	bid	buyer	seller	bidder	price	quote	357
48	contributor	instruction	filer	political	officeholder	payee	79
04	provisions	subcontractor	surety	rev	bidder	supplementary	232
27	breach	franchisee	hereunder	agreement	remedy	agree	213
12	youth	camp	teach	teen	lesson	yoga	722
23	dog	rabies	euthanasia	euthanized	pet		1710
25 35	trust	revocable	mfr	apportionment	living	spay assn	285
				1.1	_		
16	emergency	preparedness	null	dispatch	rescue	fire	
80	energy	efficiency	customer	saving	rebate	renewable	382
13	proud	leadership	honor	pleased	grateful	passion	1168
18	garland	invoice	assoc	check	firefighter	association	152
81	page	last	sub	update	prime	award	17
2	mosquito	insecticide	spray	bait	repellent	pesticide	997
20	project	improvement	funding	justification	completion	acquisition	47 I
05	thence	plat	easement	annexation	pud	westerly	255
18	comment	concern	suggest	clarify	suggestion	dear	307
34	library	campus	doe	branch	center	arena	208
40	portfolio	treasury	investment	maturity	yield	liquidity	250
15	masonry	plaster	joist	stud	sheathing	ceiling	875
53	department	authority	dpt	correction	citywide	transit	109
3	vend	utensil	meat	fat	cheese	salad	1325
8	assessor	taxpayer	determination	informal	petition	notification	39
58	recycling	bag	garbage	recycle	recyclable	recyclables	318
87	sign	billboard	pole	speeding	illuminate	banner	472
31	student	elementary	school	college	graduate	academic	233
32	dwelling	alteration	plumbing	plumb	canceled	mechanical	143
51	combustible	vent	piping	conductor	duct	flammable	517
91	app	credit	download	post	issued	agent	57
66	wetland	vernal	riparian	habitat	specie	species	1040
44	findings	string	tank	carcinogen	qty	lust	128
42	contamination	spill	remediation	groundwater	asbestos	hazardous	343
99	prep	batch	qualifier	analytical	surrogate	sample	313
84	airport	facility	aviation	maintenance	operation	aircraft	150
19	accessory	height	dwell	frontage	setback	subsection	218
6	householder	poverty	disability	married	husband	universe	93
98			epidemic	soda	sensible	drink	
	obesity	sugary					
33	avenue	street	west	east	boulevard	south	98
10	deductible	copay	prescription	coinsurance	outpatient	inpatient	488
50	ductile	trench	pipe	manhole	coupling	compaction	705
17	margin	error	occupied	race	occupy	islander	79
5	earthquake	flood	floodplain	flooding	landslide	fault	723
76	variance	setback	yard	exception	fence	front	94
16	business	marijuana	cannabis	manufacturing	industry	collective	319
.08	fugitive	bio	exhaust	unmitigated	noise	receptor	262

Table 4: Top words from a structural topic model with 120 topics (first 60 topics displayed here) and FREX scoring. Colors depict partisanship based on coefficient size. White cells are non-significant topics.

	TD 337 1.1	T W 10	T W 12	TD 337 1.4	TD XX 15	T W 16	T. 1	. 1
#	Top Word 1	Top Word 2	Top Word 3	Top Word 4	Top Word 5	Top Word 6		assigned
29	labor	worker	force	unemployed	earnings	civilian	80	I
111	discharge	pollutant	inspection	inspect	pollution	inspector	109	•
68	contractual	parts	duke	outside	postage	receipts	274	
77	curb	pavement	sidewalk	ramp	gutter	asphalt	390	
65	draft	update	process	review	staff	progress	67	1
24	landlord	tenant	renewal	rent	lease	expired	255	
106	consultant	proposer	procurement	contract	firm	subcontractor	179	
43	blanket	medicare	payroll	premium	undistributed	refund	107	ī
103	urban	mixed	density	redevelopment	development	industrial	115	1
89	taxable	res	deed	value	homestead	star	41	•
83		demolition		demolish	floor		82	
	building		story			build		
119	cost	estimate	estimated	initial	costs	change	52	T.
109	respondent	satisfied	dissatisfied	survey	satisfaction	disagree	403	
64	must	signature	copy	application	applicant	submission	139	
26	tax	deduction	amt	assessed	bill	abatement	171	
78	yes	worksheet	text	pic	font	button	476	
7	greenhouse	emission	coal	climate	ozone	dioxide	334	
54	parking	tow	taxi	vehicle	shuttle	passenger	236	
41	assistant	analyst	technician	aide	specialist	asst	119	ī
22	allocation	val	cove	acct	glen	subtotal	79	i i
63	fee	charge	license	reservation	surcharge	refundable	143	
117		•			flt	detector	113	
	delay	perm	queue	peak				1
4	datum	database	copyright	accuracy	data	compile	193	
45	audit	auditor	auditing	internal	implemented	procedure	222	
100	mitigation	impact	significant	adverse	significance	unavoidable	136	
88	gender	discrimination	transgender	immigrant	immigration	religion	859	
9	district	zoning	maker	vacancy	speaker	planner	45	1
12	artist	artwork	art	arts	mural	sculpture	1055	
94	contracted	encumbrance	unencumbered	exp	expend	bud	71	1
110	rouge	parish	baton	thereto	sewerage	adjudicate	464	
46	commissioner	chair	commission	committee	briefing	advisory	187	_
85	sch	min	tin	hump	carpool	qua	390	
15	complainant	allegation	allege	complaint	doc	misconduct	963	
30	incumbent	examination	supervision	knowledge	exam		410	
				Č		ability		_
107	savings	ltd	village	neighborhood	excise	costs	81	1
72	imp	burglary	theft	testify	petitioner	mischief	116	
60	bike	bicycle	bicyclist	pedestrian	route	mobility	336	
82	accomplishment	narrative	grantee	outcome	objective	mod	101	1
36	decline	trend	recession	average	rate	percentage	265	
52	homeless	homelessness	supportive	consolidated	transitional	counseling	193	
1	alderman	resolved	whereas	resolution	authorizing	authorize	245	
92	concept	design	realm	visual	character	conceptual	433	
71	bond	obligation	proceeds	redemption	debt	series	174	_
67	dist	applied	col	occupancy	valuation	monoxide	62	Ī
25	scenario	figure	appendix	assume	assumption	model	162	
38	horas	persona	* *	yon	sou		1350	
		1	para	•		ante	90	-
14	federal	agency	entity	recipient	grant	eligible		
56	waterfront	shoreline	marina	beach	port	boat	844	
61	revenue	balance	expenditure	reserve	forecast	budget	101	•
75	governmental	asset	liability	assets	statement	pension	142	•
37	endorse	endorsed	budget	proposed	adopted	adopt	111	
69	tree	planned	circumference	gross	density	infill	211	
90	councilman	introduced	ordain	ordinance	digest	yea	244	
97	actuarial	grievance	employee	retirement	bargaining	actuary	250	
39	affordable	housing	affordability	homeowner	income	bedroom	150	ī .
55	ave	combo	blossom	pearl	cir	olive	1091	_
13	strategy	goal	strategic	stakeholder	focus	initiative	162	
57	absent	int	preside	ordained	tag	numbers	194	
101								
101 93	violent shooting	gang suspect	firearm	offender gunshot	crime flee	patrol shoot	511 730	

Table 5: Top words from a structural topic model with 120 topics (second 60 topics displayed here) and FREX scoring. Colors depict partisanship based on coefficient size. White cells are non-significant topics.

february subcommitte speaker agenda committee commission 446 and a committee commission 446 and and a committee commission 447 and	#	Top Word 1	Top Word 2	Top Word 3	Top Word 4	Top Word 5	Top Word 6		Tokens assigned
Second		chair	subcommittee	speaker	agenda	committee	commission		_
budget evenue adopted balance transfer expenditure 176   trench manhole ductile excavation pipe grout   trench manhole ductile excavation pipe grout   trench manhole ductile excavation   trench manhole ductile excavation   trench manhole   trench manhole   trench manhole   trench manhole   trench manhole   trench   t	19	setback	variance	zoning	height	yard	accessory	453	
downtown mixed retail waterfront orient density 419 rtrust revocable planned mfr revocable planned mfr audit auditor procedure timely implemented oversight 422 mitigation impact significant aleverse governmental asset actuarial liability implemented oversight 422 mitigation impact significant aleverse governmental asset actuarial liability implemented oversight 422 mitigation impact significant aleverse experimental asset actuarial liability implemented oversight 422 mitigation impact significant aleverse experimental asset actuarial liability implemented oversight 422 mitigation impact significant aleverse experimental asset actuarial liability implemented oversight 422 mitigation impact significant aleverse experimental inferitation infiltration infiltration distinctive revival storefront 1731 mitigation impact significant aleverse experimental infiltration infi	5	draft	comment	review	revision	clarify	process	356	
trench manhole ductile excavation pipe grout 1436	58	budget	revenue	adopted	balance	transfer	expenditure	176	•
y trust revocable planned mfr audit to gay 377 and and the presence and th	39	downtown	mixed	retail	waterfront	orient	density	419	_
absent addit anditor procedure timely implemented oversight 472 andit andit anditor procedure timely implemented oversight 472 andit anditor procedure timely implemented oversight 472 andit anditor procedure timely implemented oversight 472 and the significant adverse environmental measure 217 and the significant adverse environmental measure 218 and the scharge sladege 751 and the significant adverse environmental measure 218 and the scharge sladege 751 and the scharge	50	trench	manhole	ductile	excavation	pipe	grout	1436	
4 audit auditor procedure timely implemented oversight 472 5 mitigation impact significant adverse environmental asset actuarial liability financial statement 235   4 cell'uent inf eff infiltration discharge sludge 751   7 craftsman architecture brick distinctive revival storefront 1731   8 grievance deductible coinsurance dependent employee copay 583   9 crespondent compare figure trend appendix satisfied 696   10 customer renewable efficiency energy saving conservation 652   10 customer renewable efficiency energy saving conservation 652   11 customer renewable efficiency energy saving conservation 652   12 customer renewable efficiency energy saving conservation 652   13 wetland specie species vernal sidewalk bicyclist 561   14 medicare payroll blanket combustible flammable egress   15 vent combustible flammable egress   15 vent combustible flammable egress   15 vent combustible flammable egress   15 student teacher preschool academic   16 amagin error disability speak employed   17 fugitive bio emission coal unmitigated   18 margin error disability speak employed   19 margin error disability speak employed   10 playground recreation picnic   10 playground recreation picnic   11 pobesity sugary epidemic dirink calorire exhaust 773   11 postition miteron maturity treasury yield   12 taxable deed   res   18 perm   queue delay   peak   adj   flt   187   19 portfolio micron maturity treasury yield   19 portfolio micron maturity treasury yield   11 postition   11 postition   12 portfolio micron maturity treasury yield   13 payra persona horas bad contracted unine   14 accomplishment   fax   application   tow   fermale   floo   15 ercedit docket   app   post   during dermination   post   flindings   fant   flindings   flam   flindings   flam   flindings	9	trust	revocable	planned	mfr	apportionment	exhibit	361	•
25         mitigation         impact         significant         adverse         environmental         measure         217           47         effluent         inf         eff         infiltration         discharge         sludge         751         —           47         effluent         inf         eff         infiltration         discharge         sludge         751         —           48         corafisman         archiccture         coinsurance         dependent         erepondent         copay         583         —           48         contributor         filer         officiencle         energy         saving         conservation         602         —           56         savings         neighborhood         bike         pedestrian         route         sidewalk         bicyclist         561         —           28         garland         assoc         association         firefighter         duke         xerox         480         —           21         medicare         payroll         blanket         contractual         undistributed         dept         116         —           22         allocation         subtotal         suce         sacademic         subtogerses <td>1</td> <td>absent</td> <td>preside</td> <td>authorize</td> <td>ordained</td> <td>int</td> <td>tag</td> <td>377</td> <td>-</td>	1	absent	preside	authorize	ordained	int	tag	377	-
governmental asset actuarial liability financial statement 235   12 craftsman architecture brick distinctive revival storefront 1731   13 craftsman architecture brick distinctive revival storefront 1731   14 craftsman architecture brick distinctive revival storefront 1731   15 craftsman architecture brick distinctive revival storefront 1731   16 craftsman architecture brick distinctive revival storefront 1731   17 craftsman architecture brick distinctive revival storefront 1731   18 craftsman architecture brick distinctive revival storefront 1731   19 craftsman architecture brick distinctive revival storefront 1731   10 grievance dependent employee copay 583   10 customer renewable efficiency energy saving conservation 652   10 customer renewable officeholder oblitical or rouge payce 293   13 swelland specie species vernal cological matrix 131   10 customer or oute sidewalk bicyclist 561   13 wetland specie species vernal cological riparian 2293   14 went combustible flammable egress ceiling extinguisher 1160   15 vent embessness student teacher preschool academic kindergarten youth 855   15 taudent teacher preschool academic kindergarten youth 855   15 taudent teacher preschool academic kindergarten youth 855   16 taudent dwelling suite ave tad alteration 491   17 fuglitive bio emission coal unmitigated exhaust 773   18 playground recreation picnic park restroom 200   18 perm queue delay peak adj flit 187   18 perm queue delay peak adj flit 187   19 delay ground recreation picnic park restroom 200   18 perm queue delay peak adj flit 187   19 delay ground precreation picnic park restroom 200   10 delay peak adj flit 187   11 delay ground precreation picnic park restroom 200   12 credit docket app post download month 61   13 anatashle deed res householder islander census occupied female 160   14 accomplishment grantee narrative outcome grant recipient 255   15 accedit docket app post download month 61   16 credit docket app post download month 61   17 credit docket app post download month 61   18 perm post of th	4	audit	auditor	procedure	timely	implemented	oversight	472	_
governmental asset actuarial liability financial statement 235   12 craftsman architecture brick distinctive revival storefront 1731   13 craftsman architecture brick distinctive revival storefront 1731   14 craftsman architecture brick distinctive revival storefront 1731   15 craftsman architecture brick distinctive revival storefront 1731   16 craftsman architecture brick distinctive revival storefront 1731   17 craftsman architecture brick distinctive revival storefront 1731   18 craftsman architecture brick distinctive revival storefront 1731   19 craftsman architecture brick distinctive revival storefront 1731   10 grievance dependent employee copay 583   10 customer renewable efficiency energy saving conservation 652   10 customer renewable officeholder oblitical or rouge payce 293   13 swelland specie species vernal cological matrix 131   10 customer or oute sidewalk bicyclist 561   13 wetland specie species vernal cological riparian 2293   14 went combustible flammable egress ceiling extinguisher 1160   15 vent embessness student teacher preschool academic kindergarten youth 855   15 taudent teacher preschool academic kindergarten youth 855   15 taudent teacher preschool academic kindergarten youth 855   16 taudent dwelling suite ave tad alteration 491   17 fuglitive bio emission coal unmitigated exhaust 773   18 playground recreation picnic park restroom 200   18 perm queue delay peak adj flit 187   18 perm queue delay peak adj flit 187   19 delay ground recreation picnic park restroom 200   18 perm queue delay peak adj flit 187   19 delay ground precreation picnic park restroom 200   10 delay peak adj flit 187   11 delay ground precreation picnic park restroom 200   12 credit docket app post download month 61   13 anatashle deed res householder islander census occupied female 160   14 accomplishment grantee narrative outcome grant recipient 255   15 accedit docket app post download month 61   16 credit docket app post download month 61   17 credit docket app post download month 61   18 perm post of th	25	mitigation	impact	significant	adverse	environmental	measure	217	
defluent inf eff inititration discharge studge 751 craftsman architecture brick distinctive revival storefront 1731 craftsman architecture brick distinctive revival storefront 1731 craftsman architecture brick distinctive revival storefront 1731 dependix satisfied 696 created appendix satisfied 696 contributor filer officeholder political rouge payce 293 colitical rouge payce 293	45	governmental	•	-	liability	financial	statement	235	•
cardisman   architecture   brick   distinctive   revival   storefront   1731	47	effluent	inf	eff	infiltration	discharge	sludge	751	
grievance deductible coinsurance dependnt employee copay 583 ferspondent compare figure trend appendix satisfied 696 customer renewable efficiency energy saving conservation 652 customer renewable efficiency energy saving conservation 652 deficiency excise tild matrix 131 deficiency excise tild matrix 131 deficiency excise tild matrix 131 duck excise tild matrix 1	12	craftsman	architecture	brick			0		
respondent compare figure trend appendix satisfied 696 customer renewable efficiency energy saving conservation 652 construction filer officeholder political rouge payee 293 conservation 652 co									
20 customer renewable efficiency energy saving conservation 652 = 48 contributor filer officeholder political rouge payee 293 = 56 savings neighborhood village excise ltd matrix 131   1   1   1   1   1   1   1   1   1		C				1 *			_
48 contributor filer officeholder political rouge payee 293   56 savings neighborhood village excise ltd matrix 131   57 savings neighborhood village excise ltd matrix 131   58 savings neighborhood village excise ltd matrix 131   59 savings neighborhood village excise ltd matrix 131   50 bicycle bike pedestrian route sidewalk bicyclist 561   50 welland specie species vernal ecological riparian 2293   58 garland assoc association firefighter duke xerox 480   59 vent combustible flammable egress ceiling extinguisher 11160   51 vent combustible flammable egress ceiling extinguisher 11160   52 homeless homelessness affordable supportive housing affordability 394   53 student teacher preschool academic kindergarten youth \$55   53 canceled dwelling suite ave tad alteration 491   54 fugitive bio emission coal unmitigated exhaust 773   55 fugitive bio emission coal unmitigated exhaust 773   56 savings sugary epidemic drink calorie sensible 96   57 fugitive bio emission coal unmitigated exhaust 773   58 papplied col dist occupancy monoxide valuation 128   58 perm queue delay peak adj fit 187   59 taxable deed res homestead value bowed valuation 128   51 taxable deed res homestead value when the proposer subcontract bidding 152   58 imp assessor taxpayer peitition preliminary determination 91   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res homestead value sond the proposer subcontract bidding 152   51 taxable deed res			•	-					
So savings   neighborhood   village   excise   ltd   matrix   131				•					
bicycle bicycle bicycles pedestrian route sidewalk bicyclists 561   3 wetland specie species vermal ecological riparian 2293   3 wetland specie species vermal ecological riparian 2293   3   3 wetland sassoc association firefighter duke xerox 480   3   4   3 medicare payroll blanket contractual undistributed dept 322   4   4   4   4   4   4   4   4   4					•	-			
28 garland assoc association firefighter duke xerox 480   1293   1294   1406		_	0	_					
garland assoc association firefighter duke xerox 480  1 vent combustible flammable egress ceiling extinguisher 1160  1 went combustible flammable egress ceiling extinguisher 1160  2 banker payroll blanket contractual undistributed dept 322 bomeless homelessness affordable supportive housing affordability 394  3 student teacher preschool academic kindergarten youth 855  2 allocation subtotal admin cost yon allocate 190  2 canceled dwelling suite ave tad alteration 491  3 canceled dwelling suite ave tad alteration 491  4 playground recreation picnic park restroom coal unmitigated exhaust 773  11 obesity sugary epidemic drink calorie sensible 96  1 park restroom zoo 546  40 amt invoice acct exp unencumbered encumbrance 116  3 applied col dist occupancy monoxide valuation 128  18 perm queue delay peak adj flt 187  55 taxable deed res homestead value book 87  1 imp assessor taxpayer petition preliminary determination 128  1 imp assessor taxpayer petition preliminary determination 91  1 portfolio micron maturity treasury yield investment 538  3 papra persona horas bud contracted ante 1334  44 findings tank string carcinogen lust sic 255  3 osubcontractor bid bidder proposer subcontract biddings tank string carcinogen lust sic 255  4 accomplishment grantee narrative outcome grant recipient 255  4 accomplishment grantee narrative outcome grant recipient 255  4 incumbent prep batch qualifier analystic examination 1091  3 thence east south corner west avenue 340  4 influenza infection vaccine patient tuberculosis hepatitis 2254  4 influenza infection vaccine patient tuberculosis hepatitis 2254  4 influenza infection vaccine patient tuberculosis hepatitis paradise laske analyst complaint allegation defendant offender commander complaint 149  4 complainant allegation defendant offender commander complaint 149  4 complainant allegation defendant offender commander complaint 149  4 complainant allegation defendant offender commander complaint 149  4 parts fun music beginner grante commander complaint 149  4 parts fun music b		•					•		<u> </u>
set vent combustible flammable egress ceiling extinguisher and medicare payroll blanket contractual undistributed dept 322 moderates affordable supportive housing affordablity 324 allocation subtotal admin cost yon allocate 190 margin error disability speak employed language 180 margin error disability speak employed exhaust 773 margin error erro			•	•			•		
medicare payroll blanket contractual undistributed dept 322 homeless homelessness affordable supportive louising affordability 394 allocation subtotal admin cost yon allocate 190 academic kindergarten youth 855 wouth 855 allocation subtotal admin cost yon allocate 190 allocate 190 margin error disability speak employed language 180 pmargin error disability speak employed language 180 pmargin error disability speak employed language 180 playground recreation picnic park restroom zoo 546 playground recreation picnic expark restroom zoo 546 playground recreation picnic park restroom zoo 546 playground playground playground recreation picnic park restroom zoo 546 playground playgroun		•							
Source   S					•		•		_
31 student teacher preschool academic kindergarten youth 855 allocation subtotal admin cost yon allocate 190 allocate 190 acanceled dwelling suite ave tad alteration 491 pages margin error disability speak employed language 180 pages fugitive bio emission coal unmitigated exhaust 773 pages playground recreation picnic park restroom zoo 546 playground playground recreation picnic park restroom zoo 546 playground playground recreation picnic park restroom zoo 546			1 7						
22 allocation subtotal admin cost yon allocate 190 are canceled dwelling suite ave tad alteration 491 are remore disability speak employed language 180 alteration being the proposer subcontractor bid bidder proposer subcontractor bid bidder proposer activation to maturity teasury yield investment 538 para persona horas bud contracted ante 1334 findings tank string carcinogen lust sic 255 alteration being the proposer subcontractor bid bidder proposer subcontractor bid bidder proposer activation to month of 1 census activation to meaning 658 means accomplishment grantee marrative outcome grant recipient 255 alteration allegation defendant offender commander complaint 1695 means 149 prize activation seems 100 prize activation and odd unidentified paradise 1886 means 149 prize activation seems 149 prize activation seems 149 prize activation and odd unidentified paradise 1886 means 149 prize activation activation seems 149 prize activation and odd unidentified paradise 1886 means 149 prize activation activation activation activation activation activation activation and adjourn carry unanimous chairman 419 accomplaint allegation adjourn carry unanimous chairman 419 accomplaint allegation adjourn carry unanimous chairman 419 prize accomplaint 1695 budglary robbery theft bomicide murder complaint 1695 purples accomplaint 1695 purples accomplishment 1600 purples accomplaint 1695 pu					* *	0	•		-
aze canceled dwelling suite ave tad alteration 491 margin error disability speak employed language 180 margin error disability speak employed language 180 margin error disability speak employed language 180 membrane 1773 membrane 180 memb							•		
margin error disability speak employed language 180 fugitive bio emission coal unmitigated exhaust 773 playeround recreation picnic park restroom zoo 546 numitigated exhaust 773 playeround recreation picnic park restroom zoo 546 numitigated exhaust 773 playeround recreation picnic park restroom zoo 546 numitigated exhaust 773 numitigated ex						•			<del>-</del>
7 fugitive bio emission coal unmitigated exhaust 773   11 obesity sugary epidemic drink calorie sensible 96   1   14   15   15   16   16   16   16   16   16			_						
11 obesity sugary epidemic drink calorie sensible 96 1		_		•	•				•
playground recreation picnic park restroom zoo 546 ant invoice acct exp unencumbered encumbrance 116 applied col dist occupancy monoxide valuation 128 applied col dist occupancy monoxide valuation 128 applied col delay peak adj fit 187 applicant 187 amile for race householder islander census occupied female 160 amile fax application click applicant copy 367 amile fax application click applicant copy 367 amile fax application preliminary determination predemination proposed lease 339 amile findings tank string carcinogen lust sic 255 amile findings tank string carcinogen lust sic 255 amile findings tank string carcinogen lust sic 255 amile findings article subsection shall franchisee paragraph meaning 658 amile endorse rescue assistant analyst technician 355 amile endorsed endorse rescue assistant analyst technician 355 amile endorsed endorse rescue assistant analyst technician 355 amile increment prep batch qualifier analytical examination 1091 amile applicant tuberculosis hepatitis 2980 amile findings tank strategy leadership engagement focus 502 amile focus 502 amile findings initiative outreach strategy leadership engagement focus 502 amile focus 502 amile focus focus focus for analytical examination 1091 amile focus for a function added and amile analytical examination 1091 amile focus article analytical examination 1091 amile focus article paradise 1886 amile focus 609 amile finding try amile focus 609 amile finding try amile finding try amile finding try amile finding try amile finding finding and defendant offender commander complaint 1695 amile finding finding finding finding and defendant offender commander complaint 1695 amile finding findi		-				-			
amt invoice acct exp unencumbered encumbrance 116 applied col dist occupancy monoxide valuation 128 applied color res homestead value book 87 l staxable deed res homestead value book 87 l occupied female 160 amplication click applicant copy 367 applicant copy 36		obesity	sugary	epidemic	drink	calorie	sensible		1
applied col dist occupancy monoxide valuation 128   18   perm queue delay peak adj flt 187   187		playground	recreation	picnic	park	restroom		546	
perm queue delay peak adj fit 187   55 taxable deed res homestead value book 87   6 race householder islander census occupied female 160   24 mail fax application click applicant copy 367   8 imp assessor taxpayer petition preliminary determination 91   17 portfolio micron maturity treasury yield investment 538   35 redemption bond increment obligation proceeds lease 339   38 para persona horas bud contracted ante 1334   44 findings tank string carcinogen lust sic 255   30 subcontractor bid bidder proposer subcontract bidding 512   47 article subsection shall franchisee paragraph meaning 658   15 credit docket app post download month 61   137 endorsed endorse rescue assistant analyst technician 355   14 accomplishment grantee narrative outcome grant recipient 255   15 license licensee citation tow fee taxicab 710   13 initiative outreach strategy leadership engagement focus 502   33 thence east south corner west avenue 340   42 incumbent prep batch qualifier analytical examination 1091   57 councilman introduced alderman whereas resolved councilwoman 615   23 bag recyclable recyclables reusable vegetable bait 2254   24 influenza infection vaccine patient tuberculosis hepatitis 2980   25 mesa canyon via odd unidentified paradise 1886   26 mesa canyon via odd unidentified paradise 1886   27 mile 1695   28 para persona horas bud contracted ante 1334   29 mile 13 para persona horas bud contracted ante 1334   31 para persona horas bud contracted ante 1334   32 para persona horas bud contracted ante 1334   33 para persona horas bud contracted ante 1334   34 findings tank string carcinogen lust sic case 339   35 para persona horas bud contracted ante 1334   36 para persona horas bud contracted ante 1334   37 paradise female 1360   38 para persona horas bud contracted ante 1334   39 para persona horas bud contracted ante 1334   44 findings tank string carcinogen lust sic carcinogen					exp				1
taxable deed res homestead value book 87   1   6   7   7   7   7   7   7   7   7   7		applied	col		occupancy	monoxide			•
race householder islander census occupied female 160    A mail fax application click applicant copy 367		perm	queue	delay	peak	adj	flt		•
mail fax application click applicant copy 367  mip assessor taxpayer petition preliminary determination 91  redemption bond increment obligation proceeds lease 339  para persona horas bud contracted ante 1334  findings tank string carcinogen lust sic 255  subcontractor bid bidder proposer subcontract bidding 512  redeit docket app post download month 61  redeit docket app post download month 61  redeit docket app post download month 61  redicese citation tow fee taxicab 710  initiative outreach strategy leadership engagement focus 502  thence east south corner west avenue 340  thence east south corner west avenue 340  thence east south corner resolved councilwoman 615  councilman introduced alderman whereas resolved councilwoman 615  influenza infection vaccine patient tuberculosis hepatitis 2980  mesa canyon via odd unidentified paradise 1886  mesa canyon via odd unidentified paradise 1886  motion second adjourn carry unanimous chairman 419  burglary robbery theft homicide murder gunshot 945	55	taxable	deed	res	homestead	value	book		1
imp assessor taxpayer petition preliminary determination 91 portfolio micron maturity treasury yield investment 538 predemption bond increment obligation proceeds lease 339 para persona horas bud contracted ante 1334 findings tank string carcinogen lust sic 255 subcontractor bid bidder proposer subcontract bidding 512 redit docket app post download month 61 redit docket app post download month 61 redit accomplishment grantee narrative outcome grant recipient 255 license license citation tow fee taxicab 710 linitiative outreach strategy leadership engagement focus 502 linimbent prep batch qualifier analytical examination 1091 councilman introduced alderman whereas resolved councilwoman 615 limitiative infection vaccine patient tuberculosis hepatitis 2980 linfluenza infection vaccine patient tuberculosis hepatitis 2980 lingluary robbery theft homicide murder gunshot 945 license unspective field in muster player prize 4565 log motion second adjourn carry unanimous chairman 419 log termination 91 linvestment 538 linvestment 538 lease 339 lease	6	race	householder	islander	census	occupied	female	160	•
portfolio micron maturity treasury yield investment 538    35 redemption bond increment obligation proceeds lease 339    38 para persona horas bud contracted ante 1334    44 findings tank string carcinogen lust sic 255    30 subcontractor bid bidder proposer subcontract bidding 512    27 article subsection shall franchisee paragraph meaning 658    15 credit docket app post download month 61    37 endorsed endorse rescue assistant analyst technician 355    14 accomplishment grantee narrative outcome grant recipient 255    54 license licensee citation tow fee taxicab 710    13 initiative outreach strategy leadership engagement focus 502    33 thence east south corner west avenue 340    42 incumbent prep batch qualifier analytical examination 1091    57 councilman introduced alderman whereas resolved councilwoman 615    23 bag recyclable recyclables reusable vegetable bait 2254    24 influenza infection vaccine patient tuberculosis hepatitis 2980    25 mesa canyon via odd unidentified paradise 1886    49 artist fun music beginner player prize 4565    59 burglary robbery theft homicide murder gunshot 945	24	mail	fax	application	click	applicant	сору	367	-
portfolio micron maturity treasury yield investment 538    35 redemption bond increment obligation proceeds lease 339    38 para persona horas bud contracted ante 1334    44 findings tank string carcinogen lust sic 255    30 subcontractor bid bidder proposer subcontract bidding 512    27 article subsection shall franchisee paragraph meaning 658    15 credit docket app post download month 61    37 endorsed endorse rescue assistant analyst technician 355    14 accomplishment grantee narrative outcome grant recipient 255    54 license licensee citation tow fee taxicab 710    13 initiative outreach strategy leadership engagement focus 502    33 thence east south corner west avenue 340    42 incumbent prep batch qualifier analytical examination 1091    57 councilman introduced alderman whereas resolved councilwoman 615    23 bag recyclable recyclables reusable vegetable bait 2254    24 influenza infection vaccine patient tuberculosis hepatitis 2980    25 mesa canyon via odd unidentified paradise 1886    49 artist fun music beginner player prize 4565    59 burglary robbery theft homicide murder gunshot 945	8	imp	assessor	taxpayer	petition	preliminary	determination	91	1
35 redemption bond increment obligation proceeds lease 339  38 para persona horas bud contracted ante 1334  44 findings tank string carcinogen lust sic 255  30 subcontractor bid bidder proposer subcontract bidding 512  27 article subsection shall franchisee paragraph meaning 658  15 credit docket app post download month 61  37 endorsed endorse rescue assistant analyst technician 355  14 accomplishment grantee narrative outcome grant recipient 255  4 license licensee citation tow fee taxicab 710  13 initiative outreach strategy leadership engagement focus 502  33 thence east south corner west avenue 340  42 incumbent prep batch qualifier analytical examination 1091  57 councilman introduced alderman whereas resolved councilwoman 615  23 bag recyclable recyclables reusable vegetable bait 2254  2 influenza infection vaccine patient tuberculosis hepatitis 2980  21 everybody think something thing try want 2609  26 mesa canyon via odd unidentified paradise 1886  49 artist fun music beginner player prize 4565  50 motion second adjourn carry unanimous chairman 419  50 burglary robbery theft homicide murder gunshot	17	portfolio	micron	maturity	treasury		investment	538	
para persona horas bud contracted ante 1334  44 findings tank string carcinogen lust sic 255  30 subcontractor bid bidder proposer subcontract bidding 512  27 article subsection shall franchisee paragraph meaning 658  15 credit docket app post download month 61  37 endorsed endorse rescue assistant analyst technician 355  14 accomplishment grantee narrative outcome grant recipient 255  4 license licensee citation tow fee taxicab 710  13 initiative outreach strategy leadership engagement focus 502  33 thence east south corner west avenue 340  42 incumbent prep batch qualifier analytical examination 1091  57 councilman introduced alderman whereas resolved councilwoman 615  23 bag recyclable recyclables reusable vegetable bait 2254  2 influenza infection vaccine patient tuberculosis hepatitis 2980  21 everybody think something thing try want 2609  26 mesa canyon via odd unidentified paradise 1886  49 artist fun music beginner player prize 4565  6 motion second adjourn carry unanimous chairman 419  41 complainant allegation defendant offender commander complaint 1695  59 burglary robbery theft homicide murder gunshot	35	redemption	bond	•		proceeds	lease	339	-
findings tank string carcinogen lust sic 255  30 subcontractor bid bidder proposer subcontract bidding 512  27 article subsection shall franchisee paragraph meaning 658  15 credit docket app post download month 61  37 endorsed endorse rescue assistant analyst technician 355  14 accomplishment grantee narrative outcome grant recipient 255  54 license licensee citation tow fee taxicab 710  13 initiative outreach strategy leadership engagement focus 502  33 thence east south corner west avenue 340  42 incumbent prep batch qualifier analytical examination 1091  57 councilman introduced alderman whereas resolved councilwoman 615  23 bag recyclable recyclables reusable vegetable bait 2254  2 influenza infection vaccine patient tuberculosis hepatitis 2980  21 everybody think something thing try want 2609  26 mesa canyon via odd unidentified paradise 1886  49 artist fun music beginner player prize 4565  6 motion second adjourn carry unanimous chairman 419  41 complainant allegation defendant offender commander complaint 1695  59 burglary robbery theft homicide murder gunshot	38	•	persona	horas	-	contracted	ante	1334	
subcontractor bid bidder proposer subcontract bidding 512 article subsection shall franchisee paragraph meaning 658 credit docket app post download month 61 leadership endorsed endorse rescue assistant analyst technician 355 leadership engagement focus 502 leadership engagement focus 602 leadership engagement focus 6			•	string	carcinogen			255	•
article subsection shall franchisee paragraph meaning 658  credit docket app post download month 61  redit docket app post download month 61  accomplishment grantee narrative outcome grant recipient 255  license licensee citation tow fee taxicab 710  initiative outreach strategy leadership engagement focus 502  at thence east south corner west avenue 340  incumbent prep batch qualifier analytical examination 1091  councilman introduced alderman whereas resolved councilwoman 615  abag recyclable recyclables reusable vegetable bait 2254  influenza infection vaccine patient tuberculosis hepatitis 2980  influenza infection vaccine patient tuberculosis hepatitis 2980  everybody think something thing try want 2609  mesa canyon via odd unidentified paradise 1886  motion second adjourn carry unanimous chairman 419  complainant allegation defendant offender commander complaint 1695  burglary robbery theft homicide murder gunshot	30	•	bid		_	subcontract	bidding	512	
15 credit docket app post download month 61  37 endorsed endorse rescue assistant analyst technician 355  14 accomplishment grantee narrative outcome grant recipient 255  54 license licensee citation tow fee taxicab 710  13 initiative outreach strategy leadership engagement focus 502  33 thence east south corner west avenue 340  42 incumbent prep batch qualifier analytical examination 1091  57 councilman introduced alderman whereas resolved councilwoman 615  23 bag recyclable recyclables reusable vegetable bait 2254  2 influenza infection vaccine patient tuberculosis hepatitis 2980  21 everybody think something thing try want 2609  26 mesa canyon via odd unidentified paradise 1886  49 artist fun music beginner player prize 4565  16 motion second adjourn carry unanimous chairman 419  41 complainant allegation defendant offender commander complaint 1695  59 burglary robbery theft homicide murder gunshot 945	27		subsection				•		
artist fun music beginner player rescue assistant analyst technician 355 accomplishment grantee narrative outcome grant recipient 255 accomplish accomplishment grantee narrative outcome grant recipient 255 accomplishment grantee narrative outcome grant recipient 255 accomplish grantee narrative outcome grant recipient recipient 255 accomplish grantee narrative outcome grant recipient recipient page and secure accomplish grantee narrative outcome grant recipient recipient page accomplish grantee narrative outcome grant recipient recipient page accomplish grantee narrative outcome grant recipient page accomplished accomplished page accomplished page accompl							_		1
14 accomplishment grantee narrative outcome grant recipient 255 54 license licensee citation tow fee taxicab 710 13 initiative outreach strategy leadership engagement focus 502 33 thence east south corner west avenue 340 42 incumbent prep batch qualifier analytical examination 1091 57 councilman introduced alderman whereas resolved councilwoman 615 23 bag recyclable recyclables reusable vegetable bait 2254 2 influenza infection vaccine patient tuberculosis hepatitis 2980 21 everybody think something thing try want 2609 26 mesa canyon via odd unidentified paradise 1886 49 artist fun music beginner player prize 4565 16 motion second adjourn carry unanimous chairman 419 41 complainant allegation defendant offender commander complaint 1695 59 burglary robbery theft homicide murder gunshot 945					1				<u> </u>
license license citation tow fee taxicab 710 initiative outreach strategy leadership engagement focus 502 incumbent prep batch qualifier analytical examination 1091 councilman introduced alderman whereas resolved councilwoman 615 influenza infection vaccine patient tuberculosis hepatitis 2980 influenza canyon via odd unidentified paradise 1886 mesa canyon via odd unidentified paradise 1886 motion second adjourn carry unanimous chairman 419 incomplainant allegation defendant offender commander complaint 1695 incomplainant allegation defendant offender commander gunshot 945						•			•
13 initiative outreach strategy leadership engagement focus 502  33 thence east south corner west avenue 340  42 incumbent prep batch qualifier analytical examination 1091  57 councilman introduced alderman whereas resolved councilwoman 615  23 bag recyclable recyclables reusable vegetable bait 2254  2 influenza infection vaccine patient tuberculosis hepatitis 2980  21 everybody think something thing try want 2609  26 mesa canyon via odd unidentified paradise 1886  49 artist fun music beginner player prize 4565  6 motion second adjourn carry unanimous chairman 419  41 complainant allegation defendant offender commander complaint 1695  59 burglary robbery theft homicide murder gunshot 945			•			•			_
thence east south corner west avenue 340 corner west incumbent prep batch qualifier analytical examination 1091 strength particular analytical examination 1091 strength prep batch qualifier analytical examination 1091 strength prep batch qualifier analytical examination 1091 strength prep batch qualifier analytical examination 1091 strength prep stre									_
incumbent prep batch qualifier analytical examination 1091  councilman introduced alderman whereas resolved councilwoman 615  bag recyclable recyclables reusable vegetable bait 2254  influenza infection vaccine patient tuberculosis hepatitis 2980  everybody think something thing try want 2609  mesa canyon via odd unidentified paradise 1886  artist fun music beginner player prize 4565  motion second adjourn carry unanimous chairman 419  complainant allegation defendant offender commander complaint 1695  burglary robbery theft homicide murder gunshot 945									
57 councilman introduced alderman whereas resolved councilwoman 615 23 bag recyclable recyclables reusable vegetable bait 2254 2 influenza infection vaccine patient tuberculosis hepatitis 2980 21 everybody think something thing try want 2609 26 mesa canyon via odd unidentified paradise 1886 49 artist fun music beginner player prize 4565 16 motion second adjourn carry unanimous chairman 419 41 complainant allegation defendant offender commander complaint 1695 59 burglary robbery theft homicide murder gunshot 945									_
bag recyclable recyclables reusable vegetable bait 2254  influenza infection vaccine patient tuberculosis hepatitis 2980  everybody think something thing try want 2609  mesa canyon via odd unidentified paradise 1886  artist fun music beginner player prize 4565  motion second adjourn carry unanimous chairman 419  complainant allegation defendant offender commander complaint 1695  burglary robbery theft homicide murder gunshot 945						•			
2 influenza infection vaccine patient tuberculosis hepatitis 2980 21 everybody think something thing try want 2609 26 mesa canyon via odd unidentified paradise 1886 49 artist fun music beginner player prize 4565 16 motion second adjourn carry unanimous chairman 419 41 complainant allegation defendant offender commander complaint 1695 59 burglary robbery theft homicide murder gunshot 945									
21 everybody think something thing try want 2609 26 mesa canyon via odd unidentified paradise 1886 49 artist fun music beginner player prize 4565 16 motion second adjourn carry unanimous chairman 419 41 complainant allegation defendant offender commander complaint 1695 59 burglary robbery theft homicide murder gunshot 945		_	•	•		_			
mesa canyon via odd unidentified paradise 1886 49 artist fun music beginner player prize 4565 16 motion second adjourn carry unanimous chairman 419 41 complainant allegation defendant offender commander complaint 1695 59 burglary robbery theft homicide murder gunshot 945									
16 motion second adjourn carry unanimous chairman 419 41 complainant allegation defendant offender commander complaint 1695 59 burglary robbery theft homicide murder gunshot 945		• •		_		•			
motion second adjourn carry unanimous chairman 419 complainant allegation defendant offender commander complaint 1695 burglary robbery theft homicide murder gunshot 945			•				1		
41 complainant allegation defendant offender commander complaint 1695 59 burglary robbery theft homicide murder gunshot 945						1 *			
59 burglary robbery theft homicide murder gunshot 945				•					<u> </u>
[7]	39	odigialy	robbery	uleit	12	muruer	gunsnot	743	

Table 6: Top words from a structural topic model with 60 topics and FREX scoring. Colors depict city median income based on coefficient size (wealthier cities are orange, poorer cities are teal). White cells are non-significant topics.

#	Top Word 1	Top Word 2	Top Word 3	Top Word 4	Top Word 5	Top Word 6		Tokens assigned
2	influenza	infection	vaccine	patient	tuberculosis	hepatitis	2980	
38	para	persona	horas	bud	contracted	ante	1334	
59	burglary	robbery	theft	homicide	murder	gunshot	945	
52	homeless	homelessness	affordable	supportive	housing	affordability	394	-
24	mail	fax	application	click	applicant	copy	367	•
29	margin	error	disability	speak	employed	language	180	•
36	respondent	compare	figure	trend	appendix	satisfied	696	
41	complainant	allegation	defendant	offender	commander	complaint	1695	
13	initiative	outreach	strategy	leadership	engagement	focus	502	-
6	race	householder	islander	census	occupied	female	160	_
10	grievance	deductible	coinsurance	dependent	employee	copay	583	_
31 22	student allocation	teacher subtotal	preschool admin	academic	kindergarten	youth allocate	855 190	
11			epidemic	cost drink	yon calorie	sensible	96	
44	obesity findings	sugary tank	string		lust	sic	255	
23	bag	recyclable	recyclables	carcinogen reusable	vegetable	bait	2254	
17	portfolio	micron	maturity	treasury	yield	investment	538	_
4	audit	auditor	procedure	timely	implemented	oversight	472	_
42	incumbent	prep	batch	qualifier	analytical	examination	1091	_
27	article	subsection	shall	franchisee	paragraph	meaning	658	=
15	credit	docket	app	post	download	month	61	$\overline{}$
26	mesa	canyon	via	odd	unidentified	paradise	1886	
51	vent	combustible	flammable	egress	ceiling	extinguisher	1160	
7	fugitive	bio	emission	coal	unmitigated	exhaust	773	
18	perm	queue	delay	peak	adj	flt	187	•
54	license	licensee	citation	tow	fee	taxicab	710	
53	applied	col	dist	occupancy	monoxide	valuation	128	•
48	contributor	filer	officeholder	political	rouge	payee	293	•
25	mitigation	impact	significant	adverse	environmental	measure	217	•
9	trust	revocable	planned	mfr	apportionment	exhibit	361	-
8	imp	assessor	taxpayer	petition	preliminary	determination	91	•
20	customer	renewable	efficiency	energy	saving	conservation	652	
33	thence	east	south	corner	west	avenue	340	-
56	savings	neighborhood	village	excise	ltd	matrix	131	•
28	garland	assoc	association	firefighter	duke	xerox	480	_
12	craftsman	architecture	brick	distinctive	revival	storefront	1731	
21	everybody	think	something	thing	try	want	2609	
35	redemption	bond	increment	obligation	proceeds	lease	339	•
45	governmental	asset	actuarial	liability	financial	statement	235	•
30	subcontractor	bid 	bidder	proposer	subcontract	bidding	512	
40	amt	invoice	acct	exp	unencumbered	encumbrance	116	
55	taxable	deed	res	homestead	value	book	87	<u> </u>
3	wetland	specie	species	vernal	ecological	riparian	2293	_
37 32	endorsed	endorse dwelling	rescue	assistant	analyst	technician alteration	355 491	_
32 47	canceled	dwelling inf	suite eff	ave infiltration	tad discharge		751	_
5	effluent draft	inf comment	review	revision	discharge clarify	sludge process	356	
14	accomplishment	grantee	narrative	outcome	grant	recipient	255	•
39	downtown	mixed	retail	waterfront	orient	density	419	_
43	medicare	payroll	blanket	contractual	undistributed	dept	322	-
60	bicycle	bike	pedestrian	route	sidewalk	bicyclist	561	_
58	budget	revenue	adopted	balance	transfer	expenditure	176	7
50	trench	manhole	ductile	excavation	pipe	grout	1436	
19	setback	variance	zoning	height	yard	accessory	453	
34	playground	recreation	picnic	park	restroom	zoo	546	_
1	absent	preside	authorize	ordained	int	tag	377	<b>-</b>
46	chair	subcommittee	speaker	agenda	committee	commission	446	_
57	councilman	introduced	alderman	whereas	resolved	councilwoman	615	
16	motion	second	adjourn	carry	unanimous	chairman	419	
49	artist	fun	music	beginner	player	prize	4565	
				13				

Table 7: Top words from a structural topic model with 60 topics and FREX scoring. Colors depict city population based on coefficient size (larger cities are cyan, smaller cities are magenta). White cells are non-significant topics.