BE REVOLUTIONARY™

HOW ETHICAL STANDARDS ARE DISCUSSED IN PEER-REVIEWED PUBLICATIONS: *QUALITATIVE VS QUANTITATIVE APPROACHES*

LISSIE BATES-HAUS, PH.D.

DACSS 697: TEXT AS DATA

12 MAY 2022

University of Massachusetts Amherst

PROJECT OVERVIEW

- The ethical considerations related to the use of human subjects in experimental research are of critical importance *because experimental research involves, at the heart of it, the manipulation of real-world circumstances in real people's lives*.
- Ethical standards exist to protect participants in experimental research, as well as to ensure that researchers maximize the benefits derived from their research and minimize the potential harm done.
- There are multiple ways that these standards can be articulated and communicated within a scholarly community. One important way is through the inclusion of these considerations in publications in peer-reviewed journals.
- Utilizing a text as data approach, this project will explore the different tools that may help in understanding the context of ethic*-related terms included in peer-reviewed publications.

RESEARCH FRAMEWORK

- Extension of a pilot study I performed in Spring 2021 for Causal Inference (final paper here)
- Descriptive Statistics
- What proportion of articles in corpus mention a term related to ethics (ethic*)?
- Break down by Journal Title
- Word Cloud to visualize terms near ethics
- Topic Modeling
- Look at topics related to ethic* discussions
- Look at topics modeled from abstracts
- Look at possible connections

DATA COLLECTION DECISION MAKING

- Focus only on peer-reviewed journals in the domain of Political Science
- Articles collected from the following journals via the U Mass Library and Jstor:
  - American Journal of Political Science
  - American Political Science Review
  - American Politics Research
  - Journal of Experimental Political Science
  - Political Science Research and Methods
  - Research and Politics
- Articles were pulled by the following criteria:
- The search term used was field experiment (not in quotes)
- Data range was limited to 2012-2022
- All articles returned were downloaded as pdfs, and all citations were downloaded to EndNote.

WORK FLOW

- Articles collected and all citations brought into EndNote
- Made sure I had a pdf of all articles
- Created code to bring into R and preprocess
- Ended up realizing this was just not going to work
- Final data analysis (more details published on Rpubs)
- Created excel csv that contained all author information, title, journal, abstract and all ethic* paragraphs for articles that mentioned
- Descriptives: proportion of yeses, and types of mentions
- Word cloud of ethics paragraphs
- Word cloud of abstracts?
- LDA topic modeling exploration

TECHNICAL CHALLENGES

- Many and varied
- This project was difficult for me to conceptualize (this will be discussed after the results)
- Figuring out how to pull pdfs into R and convert them to text, especially when I decided to focus on ethic* paragraphs
- Multiple steps to format data so that I could run the analytics I wanted
- Struggled with stemming. When I lemmatized, it identified each variation separately rather than seeing them all coming from ethic*
- Spent a lot of time pursuing avenues that did not work
- Code constantly failing – figuring out why? GOOGLE YOUR ERROR MESSAGE!!

EXPLORATORY MODELING DISCUSSION

- All code here on my Rpubs.
- This has been a very educational project – but also very frustrating, and it wasn't until May 10th that I really understood why.
- The biggest reason I think this project never really came together is because my overall research question is not one that is well-suited to be answered by Text As Data tools – or at least, not by the tools as I understand them.
- I did a lot of wrestling with the data, trying to get it into a place where I could work with it. When I ran my first LDA topic models on my entire corpus, I did not get anything that made any sense. I'd been asking all along "is topic modeling the right tool?" and I believe that I have answered that question, which is no. It isn't.
- In part, as I've discussed, I am not actually interested in the topics of these papers. I'm interested in whether the authors include a substantive discussion of the ethical considerations of human subjects research. No matter how many times I ran the models, the term ethic* never appeared when I modeled on the entire corpus.

DISCUSSION CONT.

- After several conversations with someone in my life who uses a lot of these tools professionally, I decided to change my approach. I decided that I would break my documents into paragraphs, select only the paragraphs that mentioned the term ethic*, and model on those paragraphs alone. I decided also instead of modeling on my corpus as a whole, I would model on only the abstracts.
- This did make a smaller dataset, but I was hopeful that I might get something more coherent from my models.
- I spent significant time (approximately 12 hours or so) working solely on the problem of how to bring in my corpus and break it into paragraphs. When I finally was able to do this and run some models, they were really not comprehensible, and I think that I understand why.

DISCUSSION CONT.

- When I went in and looked at all of the things I'd put together for paragraphs, it turns out, they did not pull in correctly. I tried multiple different tools to bring pdfs into R, but when I went back to the original documents, it appeared to me that R (at least all the tools I found) was unable to handle pfs in a two-column format, of which many of my documents were, meaning that I did not have the data I thought I did.
- I ended up having to do a huge amount of this work by hand. I ended up needing to manually put the abstracts into the excel spreadsheet. There may very well have been an easier way to do this, but because I'd lost significant time on the previous technical issues, I decided it would probably be faster simply to do it by hand.
- Because it had looked to me that the pdf were mis-rendered due to the two-column formatting (not to mention other encoding errors), I decided to look at the ethics texts in context as well.

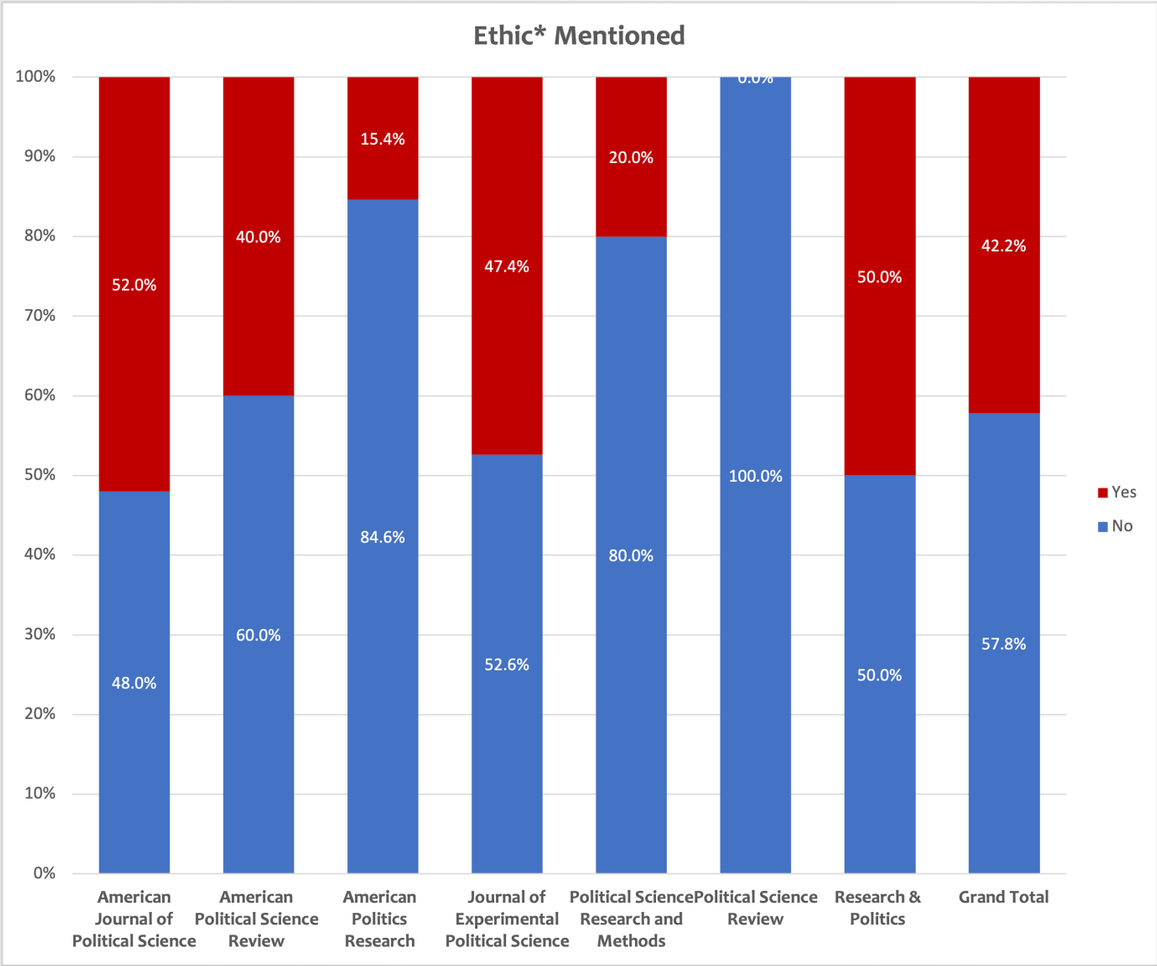
DISCUSSION CONT.

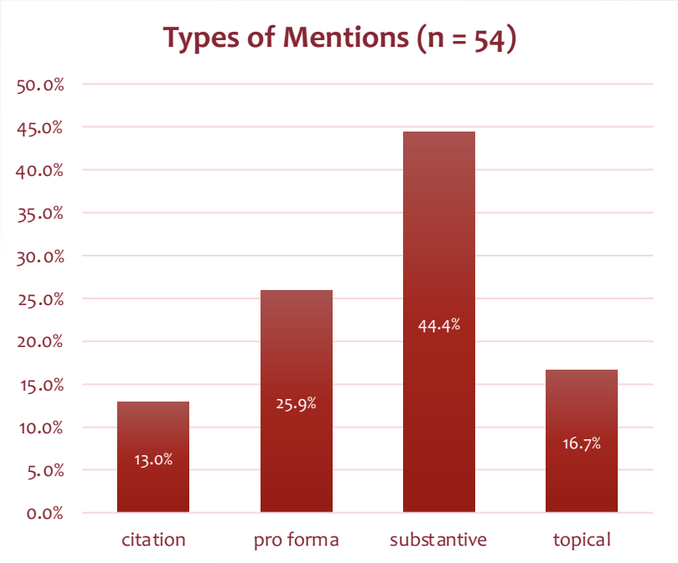
- Because I had run KWIC searches, I was easily able to identify the specific articles that had my term of interest. I then reviewed each article to identify the context of the mention, and this was where it became clear to me that automated tools were not able to give me the information that I wanted, because I observed that even though I was able to identify which articles had the mentions, there were different types of mentions that the automated tools were not able to distinguish – at least, not at my level of technical skill. Especially because I could not trust that the "paragraphs" I identified were accurately represented in R.
- What I found was that there were four kinds of mentions:
    - Substantive – there was a real discussion of the ethical considerations made by the authors in their research design process
    - Topical – the term related to the subject under investigation
    - Pro Forma – the term was mentioned as part of a brief acknowledgement that the design had been approved by an ethics committee or IRB
    - Citation – the term appeared in the bibliography

DISCUSSION CONT.

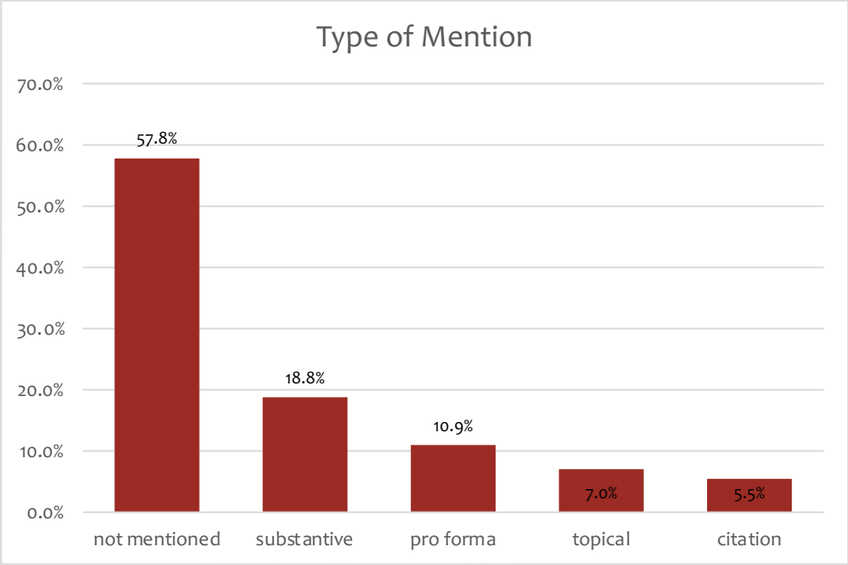
- So where does that leave the project?
- Using the data I created by hand, I was able to make some summary descriptions of ethics, in particular, looking at the breakdown by type of mention
- I was able to create cleaner/more cohesive topic models for both the Abstracts and the Ethics Paragraphs, though I'm not sure they hang together as well as I would like, and I haven't found much, if any relationship between the models and the issue of ethics.
- Ultimately, I think a different project would have been better served by this approach.
- What I'll be presenting:
  - Descriptives
  - Word Clouds (final process documented here)
  - Drill Down to Substantive Ethics – Word Cloud (I think it's not enough data to model effectively)
  - LDA Models (code and processes documented here)

DESCRIPTIVE RESULTS





These will be discussed in later slides.



ABSTRACTS WORDCLOUD

Comments:

- Words drawn from all **abstracts**
- N = top 40 words mentioned
- Stop words, numbers, punctuation removed
- Overall seems to be representing general political-science adjacent words
- Importance of "field experiment" understandable as that was the search term
- Stemming might be useful – I did not know how to combine terms based on stem

ETHIC* WORDCLOUD (ALL MENTIONS)

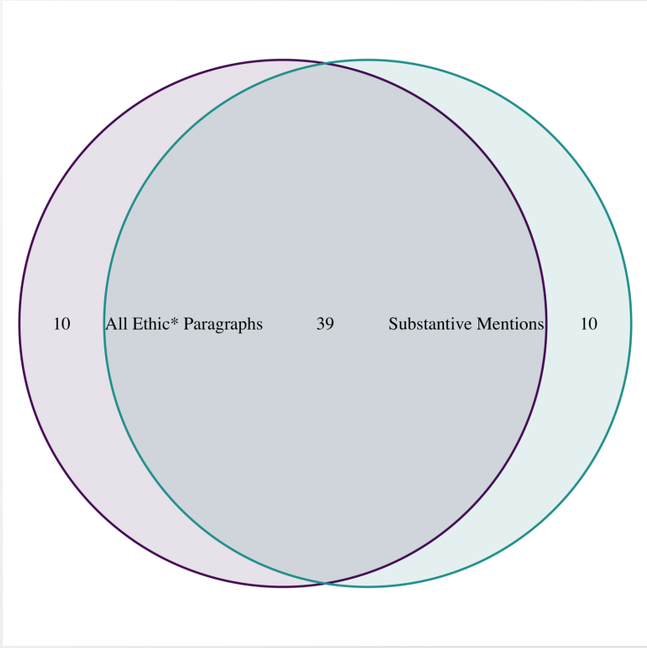
Comments:

- These words are drawn only from the paragraphs that mention the term ethic*
- N = top 40 words mentioned
- Stop words, numbers, punctuation removed
- There doesn't seem to be a big difference between the two sets of words, leading me to think that it is the substantive paragraphs that are driving the discussion of ethics overall

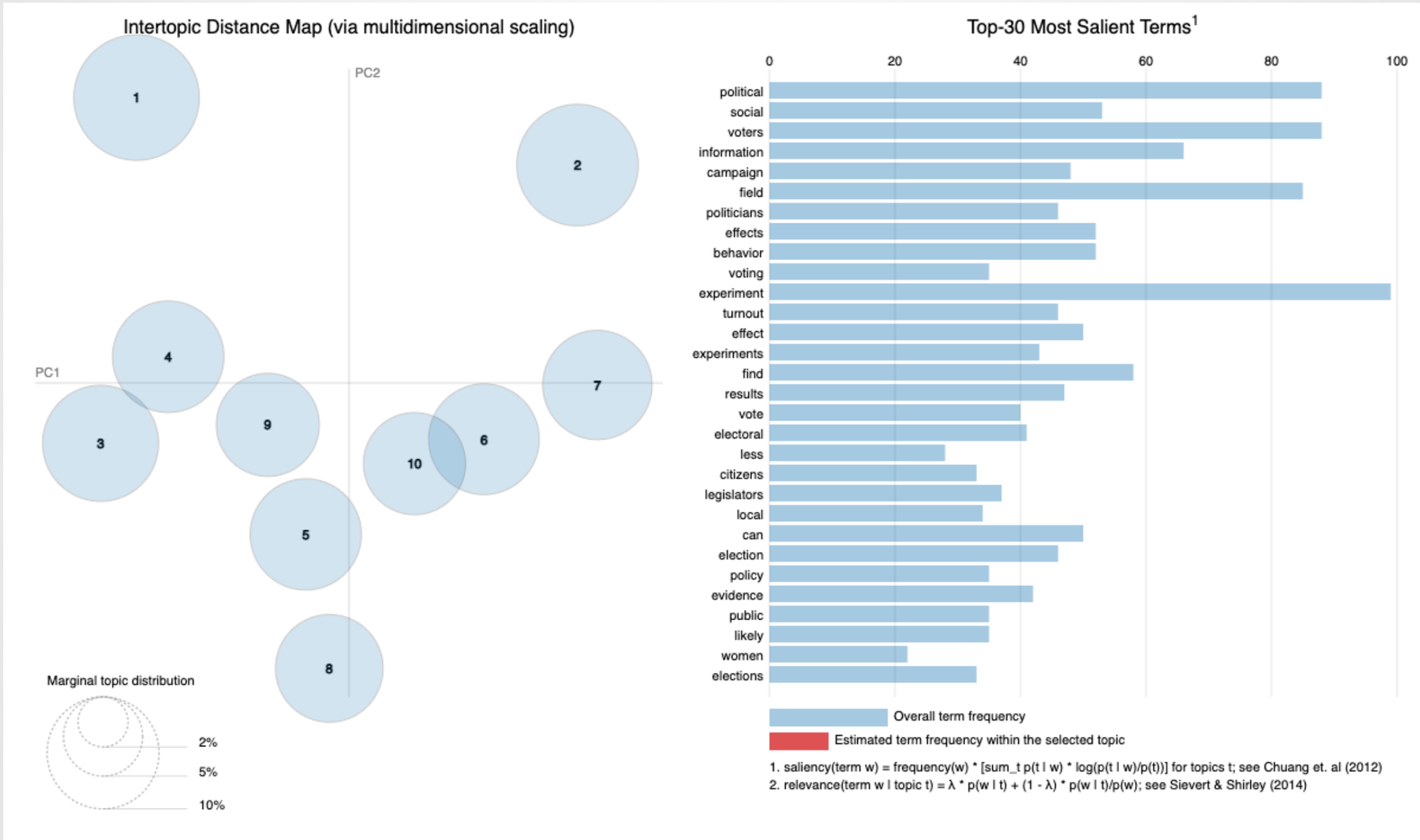
ETHIC* WORDCLOUD (SUBSTANTIVE MENTIONS ONLY)

Comments:

- These words are drawn only from the paragraphs that were identified as mentioning the term ethic* in a substantive way
- N = top 40 words mentioned
- Stop words, numbers, punctuation removed
- As the Venn diagram demonstrates, there is significant overlap between the two word sets



ABSTRACTS LDA MODELS K=10

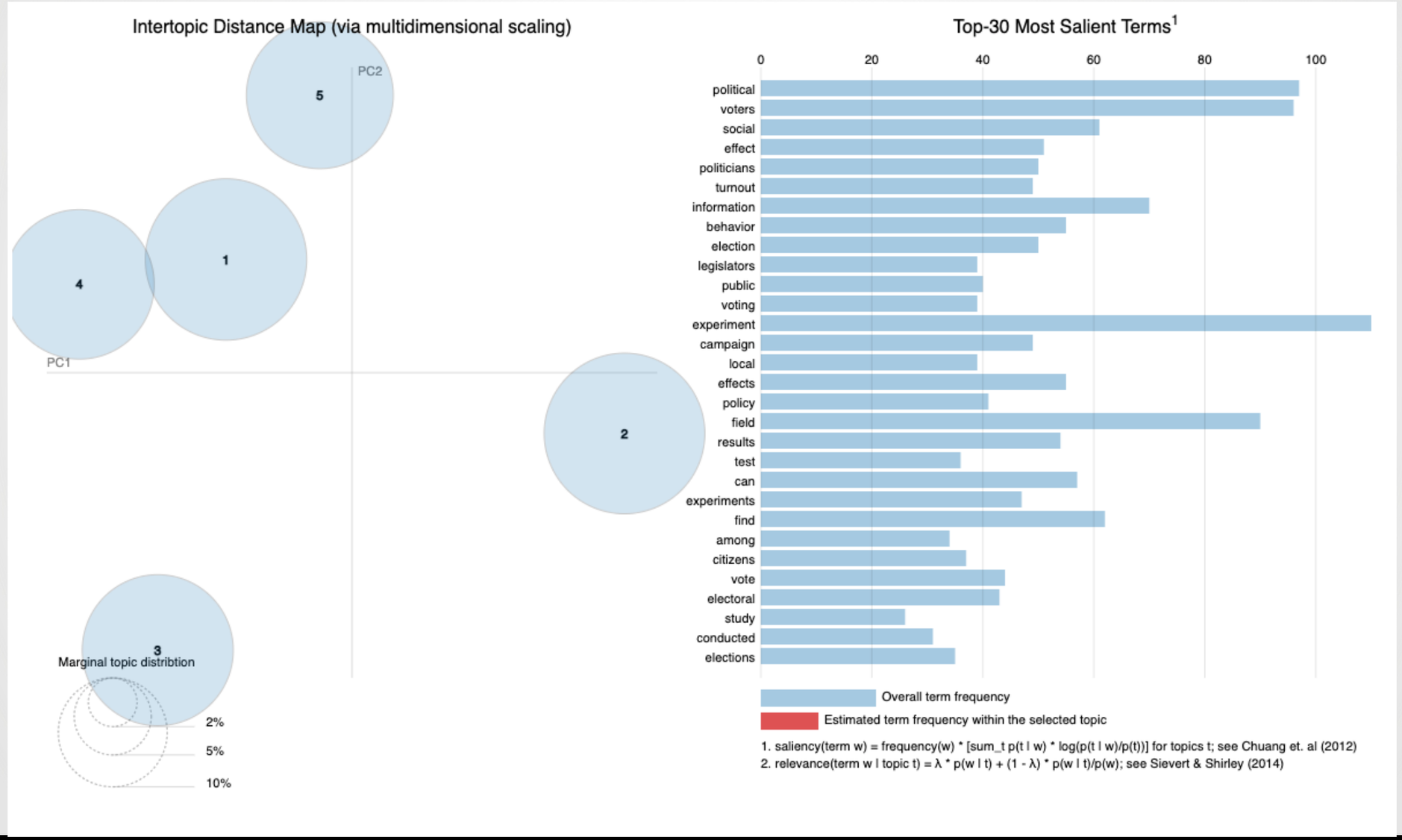


ABSTRACTS LDA MODELS K=10

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
experiments	cooperation	one	women	party	politicians	election	campaign	attitudes	data
mobilization	many	economic	context	states	messages	contact	campaigns	immigrants	survey
experimental	randomly	examine	services	direct	citizens	partisan	finance	exclusionary	communities
participation	network	outcomes	respond	increased	candidate	members	us	mechanism	help
new	leaders	traditional	incentives	receive	parties	affects	negative	natives	media
prior	countries	representatives	citizen	united	responsiveness	authoritarian	transparency	diffusion	support
democracy	particular	black	causal	ballot	preferences	candidates	audits	muslim	exposure
contribute	theoretical	via	supporters	education	constituents	elections	toward	conversations	work
effective	performance	organization	gender	second	pressure	group	federal	effectiveness	control
participants	however	contexts	conflict	similar	governments	norms	received	elite	email
interventions	ability	development	corruption	important	interest	suggesting	conducted	native	requests
particularly	access	explore	finding	different	contrast	subjects	issue	online	role
secrecy	community	whereas	firms	appear	responsive	existing	revealed	shares	areas
evidence	self-predicted	white	form	debates	voting	observational	sent	strong	average
calls	treated	connections	individual	measure	policy	stated	audited	forms	strategies
general	expectations	opinion	intergroup	play	efforts	positive	provides	problems	need
future	active	u.s	others	civic	elected	groups	s	revealing	representation
previous	brokers	bribe	ways	beliefs	less	accountability	incumbents	understanding	communication
reminder	developing	cycle	individuals	without	voters	audit	random	amplified	might
settings	empirical	matching	behaviors	capital	although	burdens	administrative	central	misinformation

- Top words definitely make sense to me
- While the topics are fairly distinct on the map, I was not able to really distinguish them from one another
- This may be because the actual topics of the research projects varied widely, but all were within the discipline of political science
- Would it make sense to prune more tokens?

ABSTRACTS LDA MODELS K=5

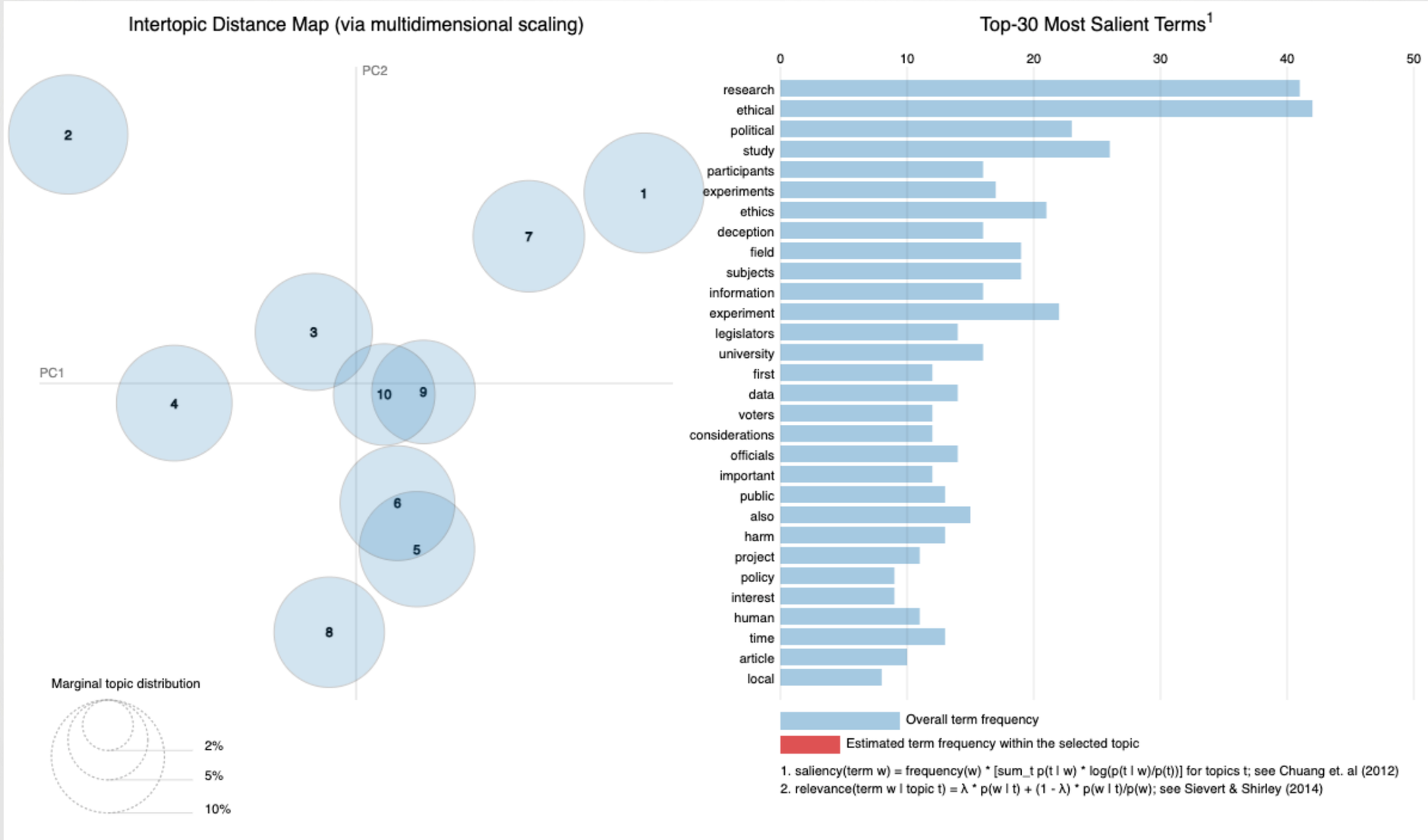


ABSTRACTS LDA MODELS K=5

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
politicians	political	legislators	effect	legislators
voters	turnout	discrimination	study	discrimination
officials	election	women	party	women
use	voting	behavior	campaign	behavior
candidate	increase	us	new	us
first	mobilization	conducted	direct	conducted
responsiveness	partisan	among	increases	among
constituents	cooperation	negative	citizens	negative
help	bias	incumbent	finance	incumbent
points	experimental	group	citizen	group
provide	effects	large	elections	large
sent	contact	transparency	messages	transparency
government	parties	audits	assigned	audits
lower	experiments	media	different	media
rates	knowledge	muslim	exposure	muslim
received	vote	action	federal	action
service	economic	institutions	randomized	institutions
services	expectations	high	designed	high
work	little	immigrants	female	immigrants
make	providing	members	report	members

- Similar conclusions as to the K=10 – that the individual topics are not that easily distinguished from one another
- Ran these models multiple times, didn't find much – it was all pretty general

ETHIC* LDA MODELS K=10

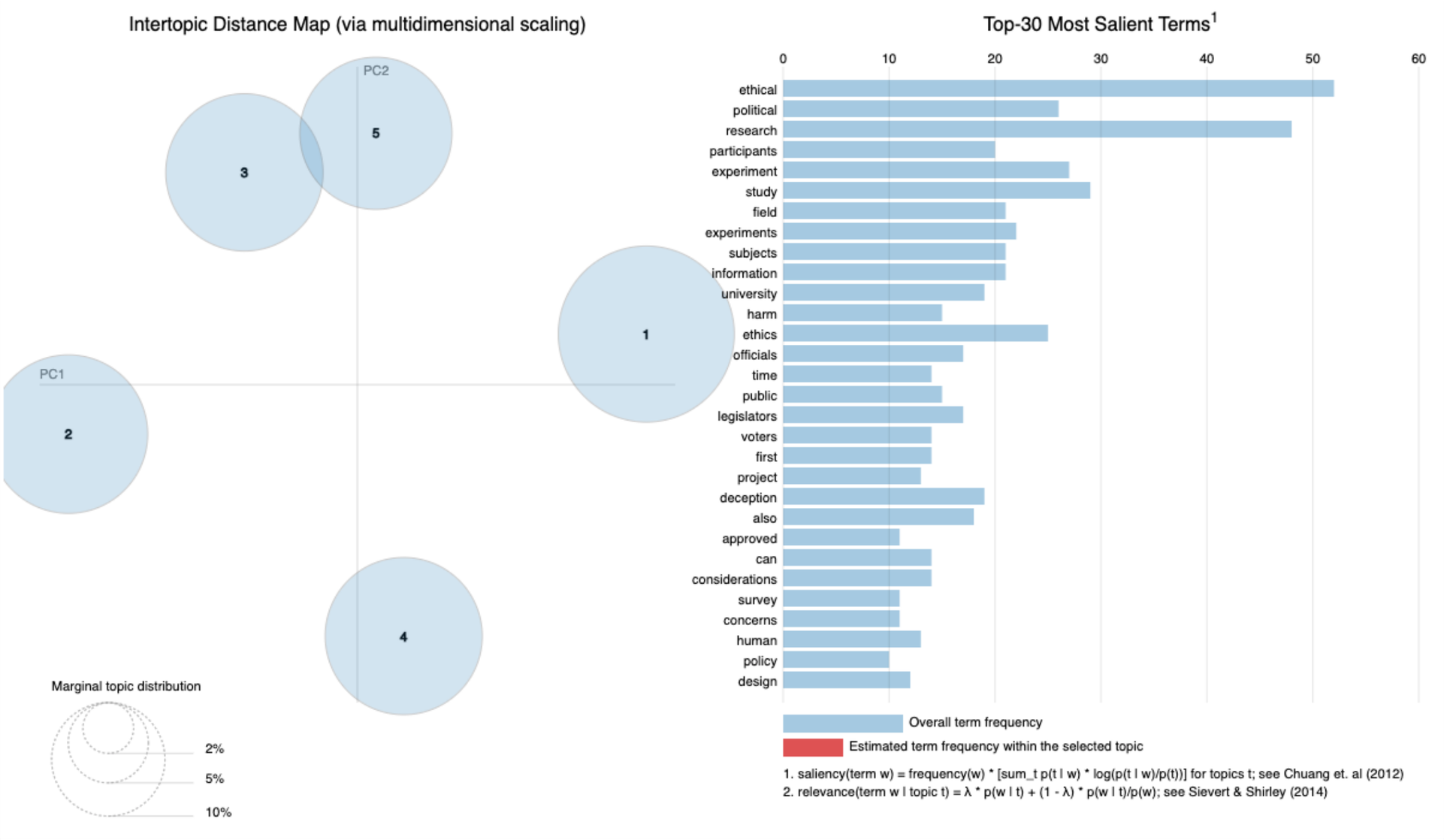


ETHIC* LDA MODELS K=10

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
data	officials	committee	also	voters	participants	university	design	subjects	deception
project	responses	electoral	harm	considerations	interest	approved	information	conducted	legislators
obtained	requests	politics	time	local	potential	behavior	experimental	et	public
consent	third	sample	concerns	within	received	reviewed	may	use	can
finance	context	uganda	minimize	risk	authors	board	issues	answer	policy
approval	demands	violence	whether	second	al	institutional	outcomes	might	benefits
assistance	grose	study	researchers	social	broockman	include	call	especially	finally
candidate	new	additional	politicians	school	bystanders	thus	necessary	including	respond
scandals	regime	brief	request	team	provided	ethical	asked	guidance	short
campaign	desired	implementation	work	evidence	subject	see	make	several	help
consider	discrimination	thank	audit	three	declare	research	principles	text	words
economic	given	two	possible	norms	participation	business	transparency	apsa's	countries
high	housing	concern	important	number	ensure	conflicts	article	dataverse	findings
selecting	love	carefully	individual	respondents	intervention	review	fowler	e	way
services	mps	democracy	costs	analyses	adheres	firms	candidates	easy	science
analysis	address	identify	end	interview	assigned	irb	author	effect	butler
approach	attendees	national	future	low	cases	minimized	believe	ideas	among
appropriate	even	paper	last	provide	left	service	character	plan	anonymity
become	kept	pennsylvania	minimizing	us	reduced	similar	designs	prevent	made
council	one	safety	present	d	used	conducting	experimentation	question	outcome

- As with the word cloud, even though there aren't that many substantive paragraphs, those do seem in general to be driving the discussion.
- With this K, there's some overlap as well as some distinct models, but again, from examination, there didn't appear to be any clear topics that seemed to gel together.
- In addition – this data is still pretty messy. Might need more qualitative assessment for pruning.

ETHIC* LDA MODELS K=5



ETHIC* LDA MODELS K=5

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
political	public	field	experiments	field
participants	first	approved	information	approved
project	policy	local	experiment	local
survey	legislators	whether	university	whether
social	behavior	et	officials	et
outcomes	received	experimental	can	experimental
team	requests	obtained	considerations	obtained
election	risk	second	human	second
finance	email	consent	design	consent
might	may	board	deception	board
possible	one	institutional	article	institutional
surveyors	politicians	results	interest	results
assistance	request	third	appendix	third
bias	science	approval	burden	approval
three	butler	conducting	responses	conducting
used	evidence	participation	authors	participation
politics	electoral	thus	researchers	thus
sample	asked	desposato	reviewed	desposato
transparency	especially	knowledge	see	knowledge
violence	grose	party	broockman	party

- As with the word cloud, even though there aren't that many substantive paragraphs, those do seem in general to be driving the discussion.
- Overall, the top 30 words do seem to make a lot of sense.
- With this K, the models seem more distinct models, but again, from examination, there didn't appear to be any clear topics that seemed to gel together.
- Would this also benefit from more pruning of words?

CONCLUSIONS

- Overall, I felt that I learned an enormous amount from this project.
- First and foremost, this experience reiterated the absolute cruciality of research design from start to finish - meaning, making sure that there is a very clear, concrete and manageable research question, and then ensuring that the tools you choose are able to answer that question.
- I think my challenges arose from not having a clear enough question and choosing the wrong tools.
- I do think, in fact, I got some really good data – particularly around the analysis of how many of the ethic* mentions were *actually substantive* as opposed to pro forma or citations, but I'm not sure how I might have gotten that information from the automated tools, given that my word clouds weren't really that different.
- I think in the future, when I keep hitting dead ends on a project, I will know to re-evaluate much earlier in the process as to whether I've picked the appropriate tools for the analysis.
- In addition, I learned a tremendous amount technically, in terms of working with my data. The challenges I faced every step of the way pushed me to better skills and a deeper understanding of the challenges of this kind of data work.
- I've put many of the files of things I tried that did not work on my github here

NEXT STEPS

- Continue to work on technical skills – especially around working with pdfs
- Continue to practice the TaD tools and visualizations
- Focus on a making a stronger connection between research question and analytic tools
- Continue to explore topic modeling (with a more appropriate dataset)
- Maybe run a Twitter scrape on a topic of interest and perform sentiment analysis and topic modeling on that data to get a better sense of how the topic modeling can work
- Looking ahead to the Capstone Project – working on how to incorporate these tools because I think they're really fascinating