

TEXT

1 , 201

PLAN FOR TODAY

Surveys and qualitative data

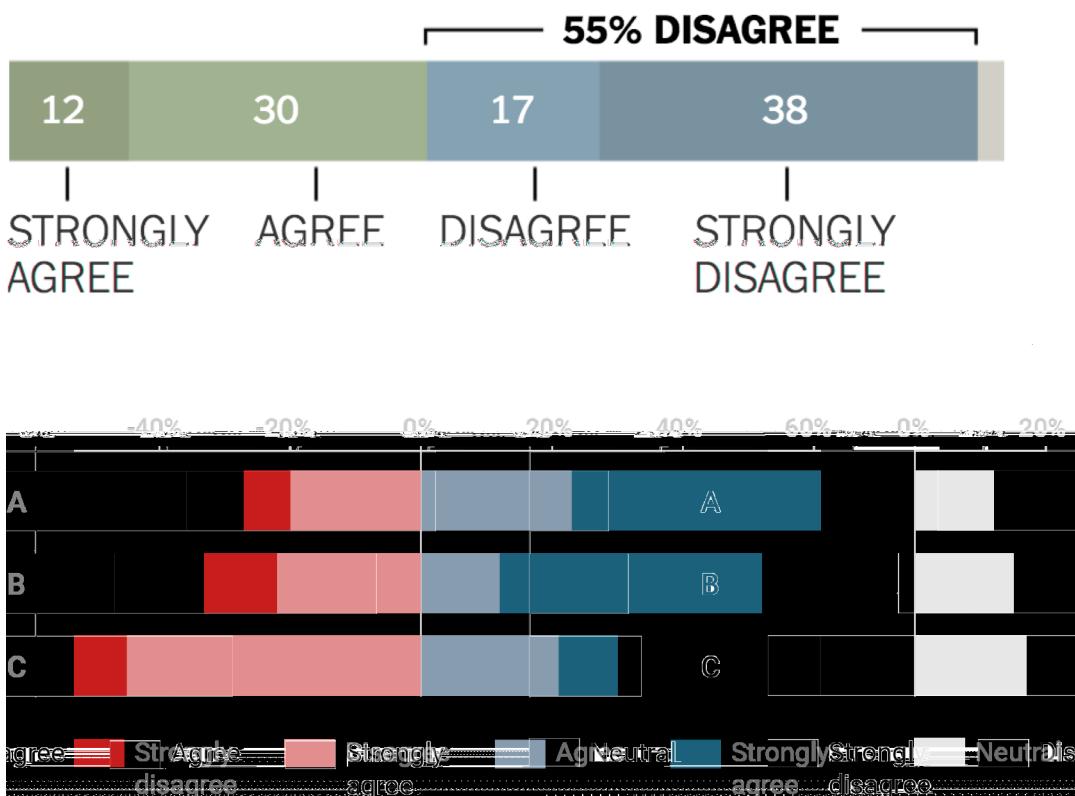
Digital humanities

Visualizing text with R

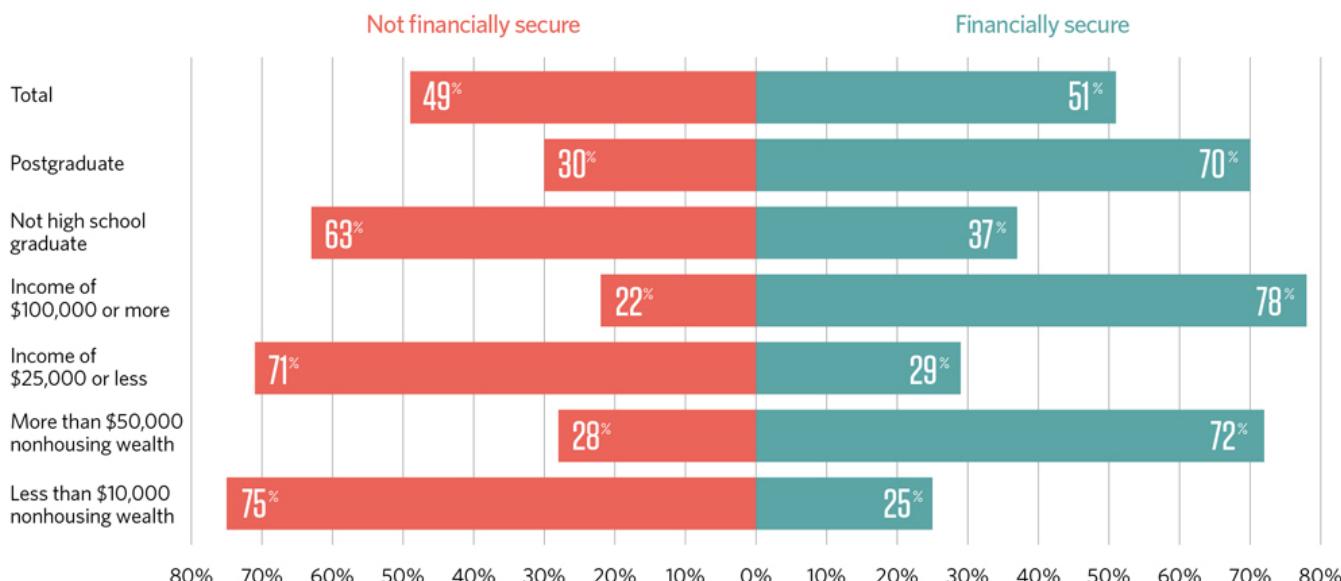
E AND
Q ALI A I E DA A

SINGLE RESPONSE QUESTIONS

It's OK if a store where I shop uses information it has about me to create a picture of me that improves the services they provide for me.



Only Half of Americans Report Feeling Financially Secure
Percentage saying they are financially secure, by selected demographics

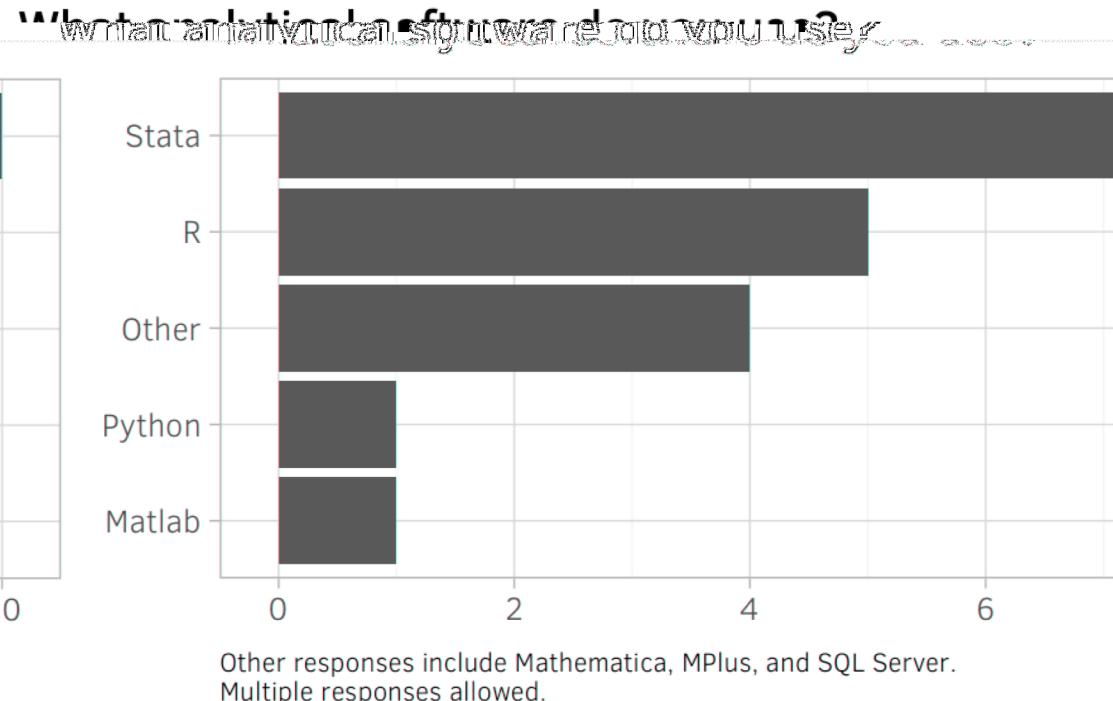


Note: People were asked, "Thinking about your household's finances today, do you feel your household is: (financially secure/not financially secure)?"

Source: Pew's Survey of American Family Finances

© 2015 The Pew Charitable Trusts

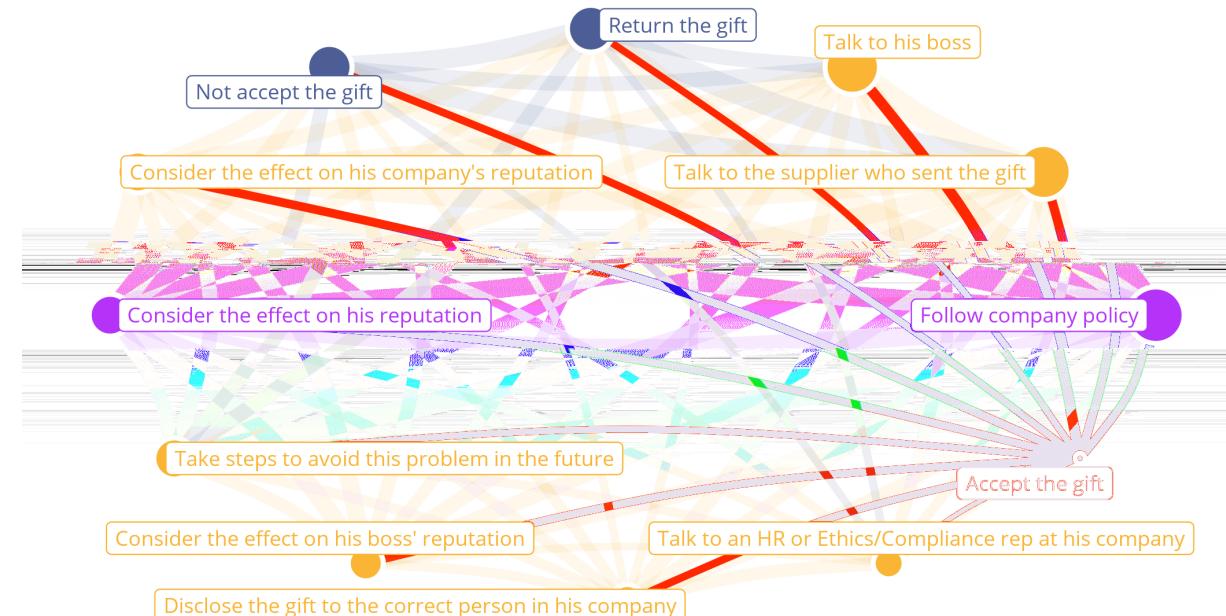
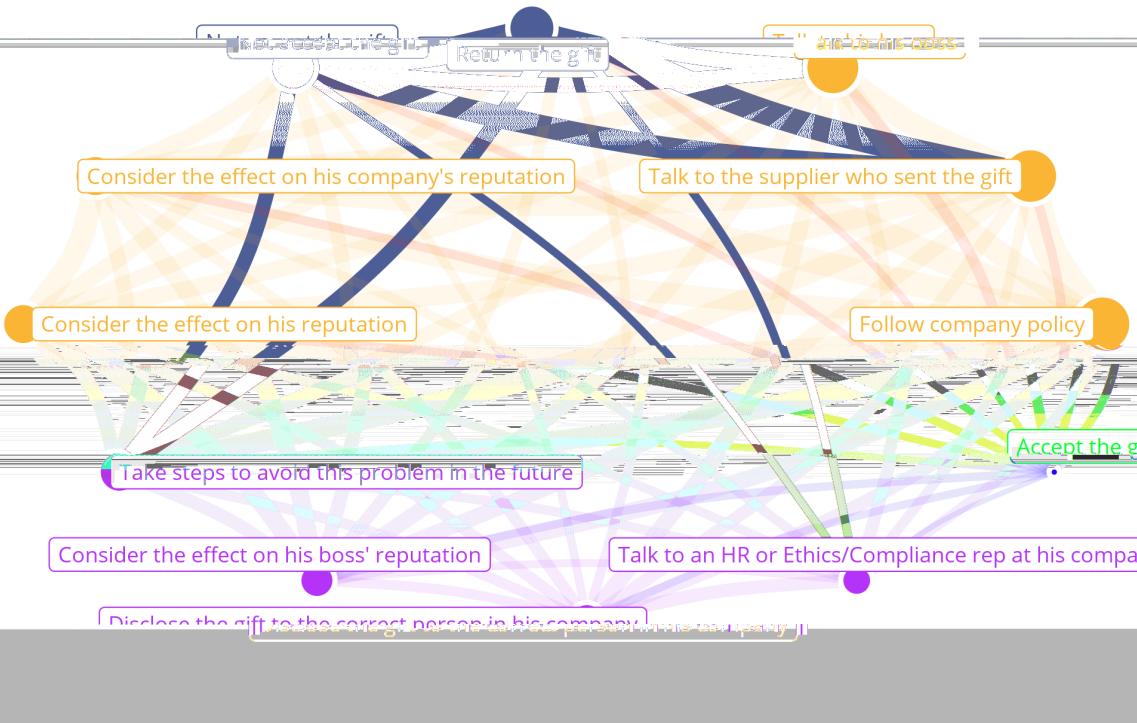
MULTIPLE RESPONSE QUESTIONS



Anger or frustration	44
Appreciation	1 46
Comparison	12 1 48
Cooperative	6 11 4 167
Disappointment	9 2 9 18 69
Embarrassment/reputation	2 2 6 0 6 22
Funding	2 0 1 6 4 2 27
Harming relations	5 0 4 0 3 2 1 14
How to improve	4 4 6 12 10 7 4 0 46
Moving goal posts	6 1 2 2 5 4 2 1 3 17
Negative	6 0 8 3 5 5 2 3 1 2 31
Objection	16 1 18 11 20 6 8 6 3 8 12 92
Other comments	9 4 9 8 9 2 3 1 3 5 8 16 58
Public face saving	4 0 3 4 7 4 4 2 3 3 5 12 6 30
US arrogance	6 2 6 6 2 1 1 1 1 3 8 11 5 27

Anger or frustration
Appreciation
Comparison
Cooperative
Disappointment
Embarrassment/reputation
Funding
Harming relations
How to improve
Moving goal posts
Negative
Objection
Other comments
Public face saving
US arrogance

COOCCURRENCE ANALYSIS

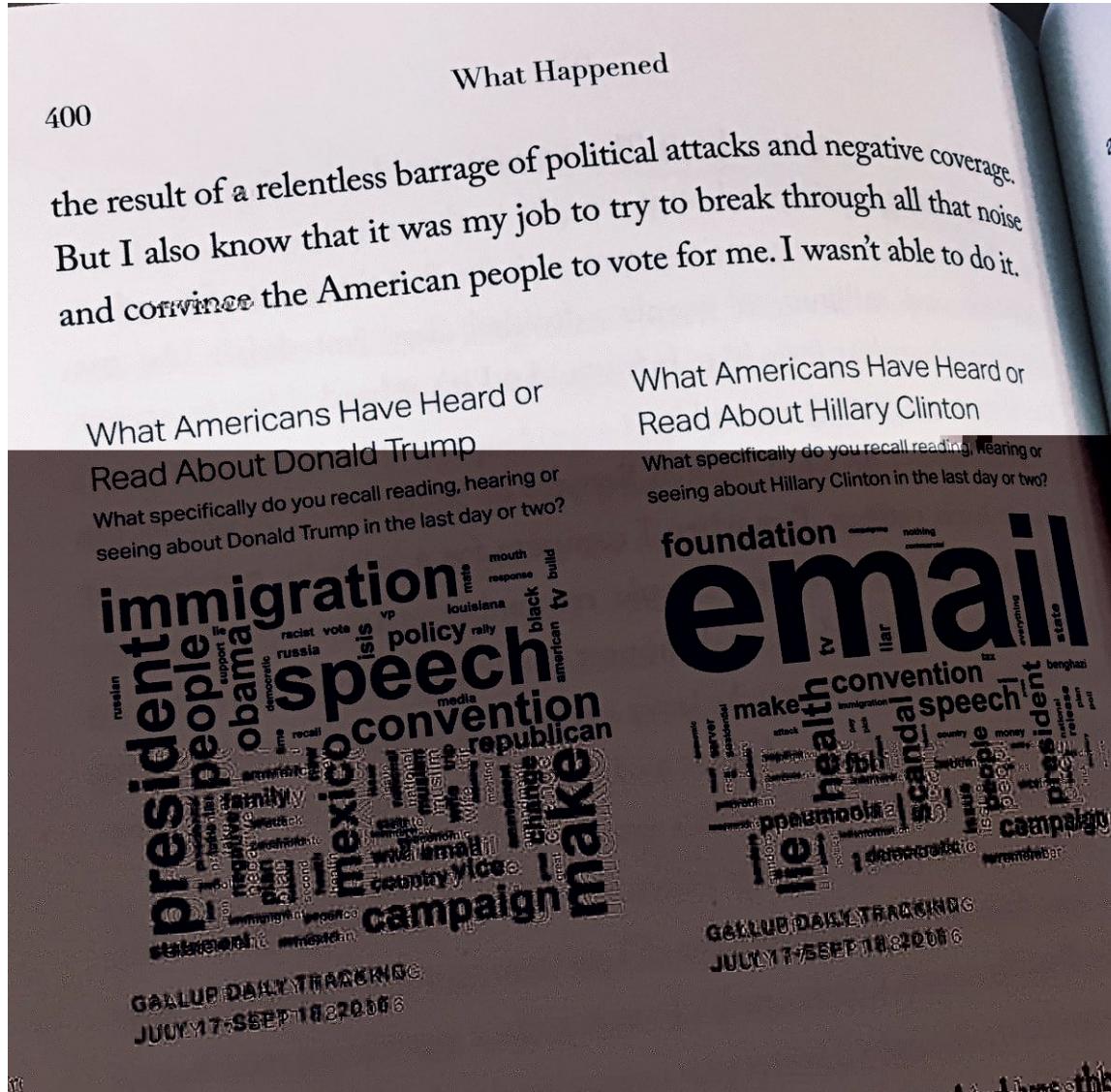


FREE RESPONSES

N	O	P	amount	donate	amount	keep	amount	why
donate_likely	amount	donate	amount	keep	amount	why		
Somewhat unlikely			0		100	I am poor		
Somewhat unlikely			0		100	I really feel like I deserve to treat myself recently. I have been working hard.		
Somewhat likely			10		90	I donate the amount that I usually would.		
Extremely likely			0		100	I'm poor	100	poor
Extremely unlikely			10		90	It is not a cause that is very important to me. I have other things to do.	Neither likely	
Extremely unlikely			29		71	I want to contribute to the cause, but also keep some of the money.	Extremely likely	
Extremely likely			20		80	It's a reasonable amount of money for an individual to donate to a charity.	Somewhat likely	
Extremely unlikely			0		100	I don't fully agree with their mission.	Extremely unlikely	
Extremely likely			10		90	I am pretty poor and need to keep some for myself, but I also want to help others.	Somewhat likely	
Somewhat likely			5		95	I think it would be a good amount to give from the money I have.	Extremely likely	
Extremely likely			69		31	to help with their cause		Neither likely
Extremely unlikely			0		100	My dad always told me to give until it hurts, and right now I am here.	Somewhat likely	
Extremely unlikely			0		100	I would rather keep the money for myself and find a charity that I like.	Neither likely	
Extremely unlikely			0		100	I want the most for myself.	Extremely unlikely	
Extremely unlikely			5		95	Can afford to give a little		Neither likely
Extremely unlikely			0		100	Because I would then have 100\$ more dollars.		Extremely unlikely
Extremely unlikely			0		100	I'm a broke boi. If anyone need humanitarian aid, it's me.		Extremely unlikely
Somewhat likely			10		90	I'm in a position where I would need the extra money, but I also want to help others.	Somewhat likely	
Somewhat unlikely			90		10	I think it is a worthy cause and I think donating 90% of the amount is fair.	Somewhat likely	
Somewhat likely			50		50	I feel splitting it 50/50 would be a fair deal. I get to help make a difference.	Extremely unlikely	
Somewhat likely			20		80	I feel that my contribution is enough. I would gladly donate again.	Extremely unlikely	
Somewhat likely			9		91	give a little		Somewhat likely
Somewhat likely			1		99	I like money		Somewhat likely
Extremely unlikely			0		100	I do not really know what they will do with the money.		Extremely unlikely

This image is a dense, abstract collage composed of numerous words and shapes, primarily in shades of green, orange, and purple, set against a background of white lined paper. The words are repeated in various sizes and orientations, creating a layered and chaotic effect. Key words visible include 'ARE', 'SO', 'DUMB', 'WHY', 'THESE', 'SOMETIMES', 'DUMBBEDUMB', and 'DUIMB'. The overall composition is organic and lacks a clear, organized structure.

IS THIS OKAY?



WORD CLOUDS FOR GROWNUPS

Counting words, but in fancier ways

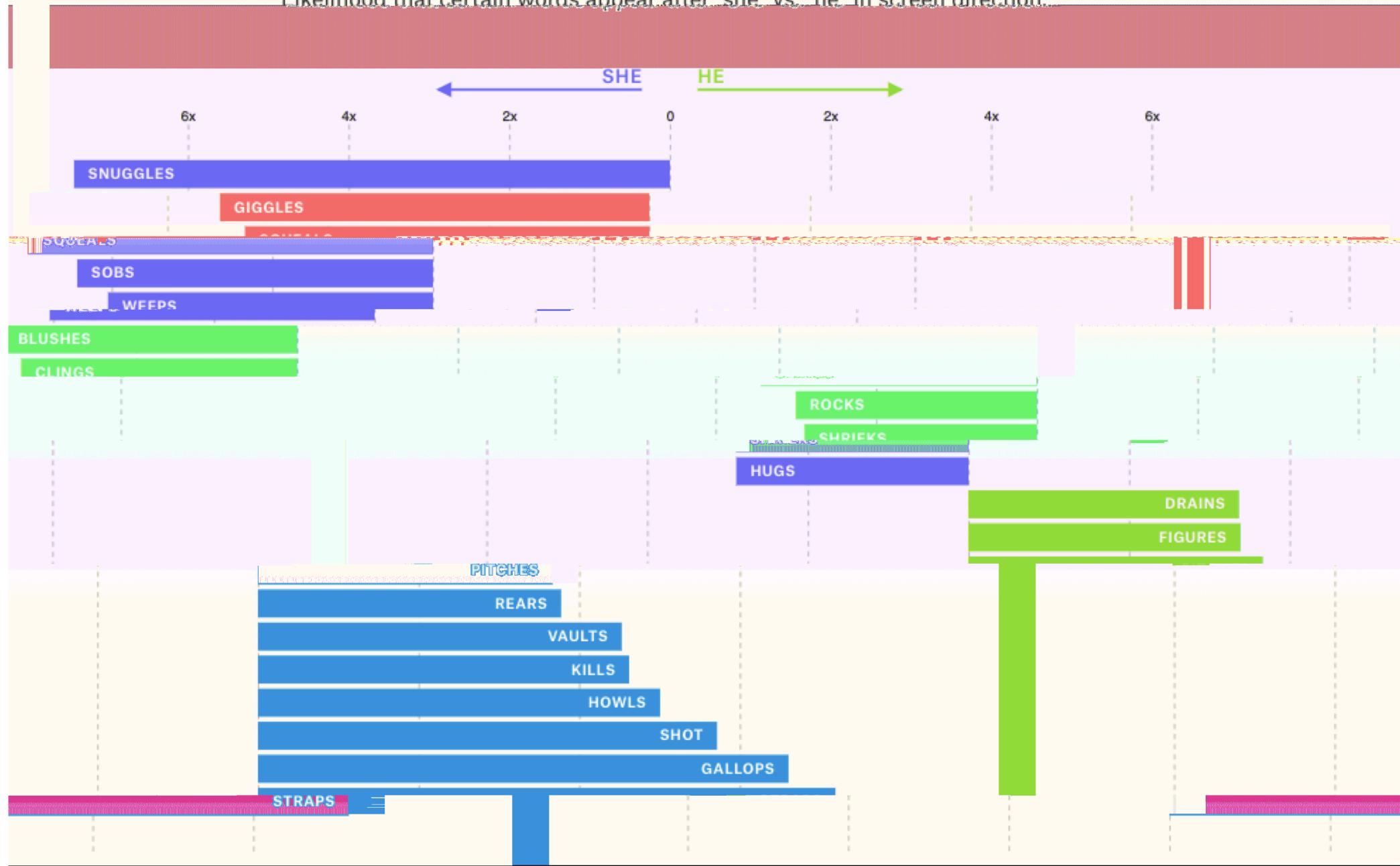
O'REILLY®



Julia Silge & David Robinson

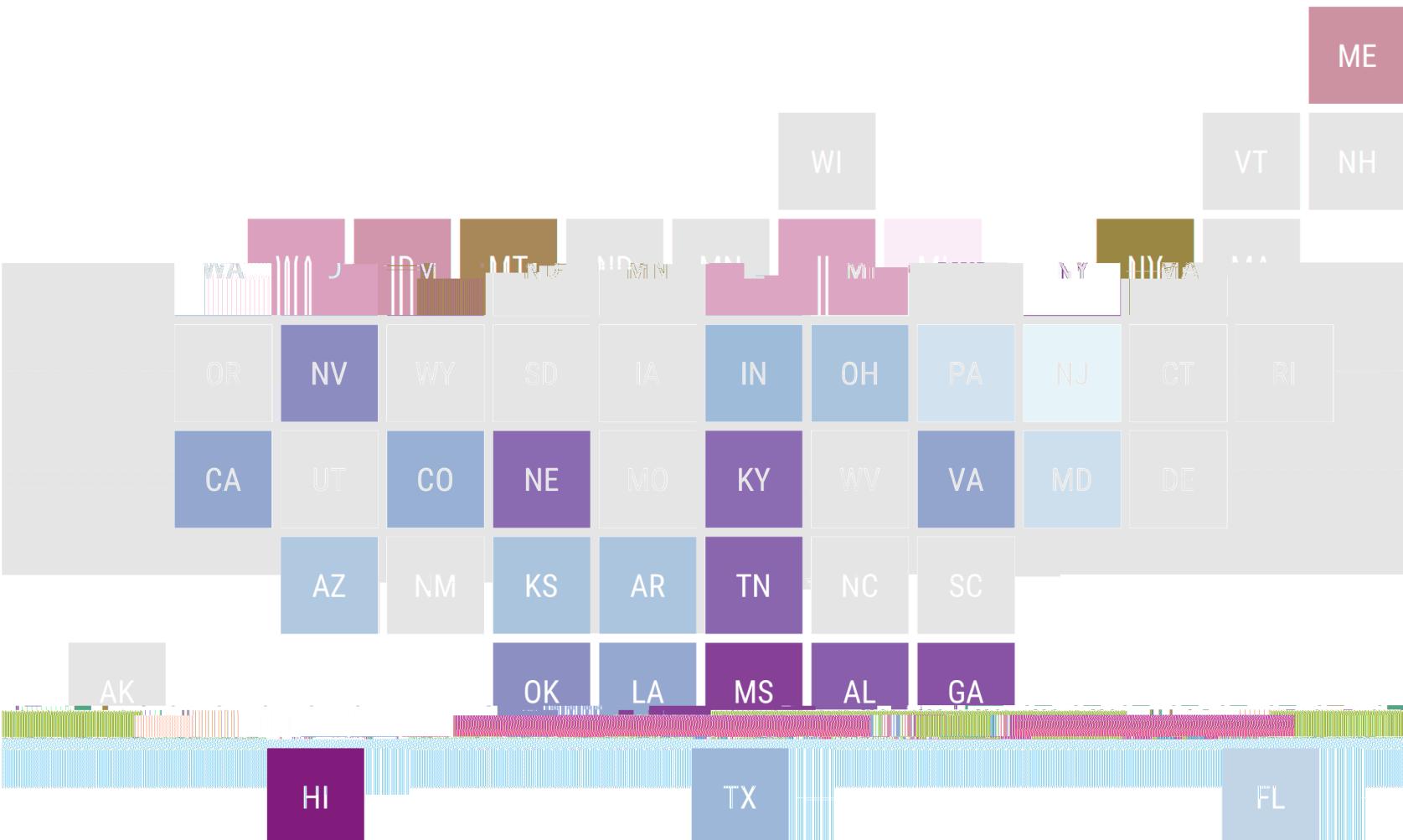
The most used words for women vs. men

Likelihood that certain words appear after "she" vs. "he" in screen direction.



Which States Are Mentioned Most in Song Lyrics?

States like Hawaii and Montana are mentioned more often relative to their population



Fewer mentions
per population

More mentions
per population

DIGITAL HUMANITIES

CRASH COURSE IN COMPUTATIONAL LINGUISTICS

Tokens, lemmas, and
parts of speech

Sentiment analysis

tf-idf

Topics and LDA

Fingerprinting

TIDY TEXT

THE BOY WHO LIVED Mr. and Mrs. Dursley of number four Privet Drive were proud to say that they were perfectly normal, thank you very much. They were the last people you'd expect to be involved in anything strange or mysterious, because they just didn't hold with such nonsense. Mr. Dursley was the director of a firm called Grunnings, which made drills. He was a big, beefy man with hardly an unusual amount of neck, although he did have a very large mustache. Mrs. Dursley was thin and blonde and had nearly twice the neck, which came in very useful as she spent so much of her time craning over garden fences, spying on the Dursleys had a small son called Dudley and in their opinion there was no finer boy anywhere. The Dursleys had wanted, but they also had a secret, and their greatest fear was that somebody would discover it. They didn't think they could bear it if anyone found out about the Potters. Mrs. Potter was Mrs. Dursley's sister, but they hadn't met for in fact, Mrs. Dursley pretended she didn't have a sister, because her sister and her good-for-nothing husband were as unDursleyish as it was possible to be. The Dursleys shuddered to think what the neighbors would say if the Potters arrived in the street. The Dursleys knew that the Potters had a small son, too, but they had never even seen him. This boy was another good reason for keeping the Potters away; they didn't want Dudley mixing with a child like that. When Mr. and Mrs. Dursley woke up on the dull, gray Tuesday our story starts, there was nothing about the cloudy sky outside to suggest that strange and mysterious things would soon be happening all over the country. Mr. Dursley, hunched over his briefcase, pecked Mrs. Dursley on the cheek, and tried to kiss Dudley good-bye but missed, because Dudley was now having a tantrum and throwing his cereal at the walls. "Little tyke," chortled Mr. Dursley as he left the house. He got into his car and backed

text	chapter
1 THE BOY WHO LIVED Mr. and Mrs. Dursley, of number four Privet Drive were proud to say that they were perfectly normal, thank you very much. They were the last people you'd expect to be involved in anything strange or mysterious, because they just didn't hold with such nonsense. Mr. Dursley was the director of a firm called Grunnings, which made drills. He was a big, beefy man with hardly an unusual amount of neck, although he did have a very large mustache. Mrs. Dursley was thin and blonde and had nearly twice the neck, which came in very useful as she spent so much of her time craning over garden fences, spying on the Dursleys had a small son called Dudley and in their opinion there was no finer boy anywhere. The Dursleys had wanted, but they also had a secret, and their greatest fear was that somebody would discover it. They didn't think they could bear it if anyone found out about the Potters. Mrs. Potter was Mrs. Dursley's sister, but they hadn't met for in fact, Mrs. Dursley pretended she didn't have a sister, because her sister and her good-for-nothing husband were as unDursleyish as it was possible to be. The Dursleys shuddered to think what the neighbors would say if the Potters arrived in the street. The Dursleys knew that the Potters had a small son, too, but they had never even seen him. This boy was another good reason for keeping the Potters away; they didn't want Dudley mixing with a child like that. When Mr. and Mrs. Dursley woke up on the dull, gray Tuesday our story starts, there was nothing about the cloudy sky outside to suggest that strange and mysterious things would soon be happening all over the country. Mr. Dursley, hunched over his briefcase, pecked Mrs. Dursley on the cheek, and tried to kiss Dudley good-bye but missed, because Dudley was now having a tantrum and throwing his cereal at the walls. "Little tyke," chortled Mr. Dursley as he left the house. He got into his car and backed	1
2 THE VANISHING GLASS Nearly ten years had passed since the day Harry first stepped through the archway into Diagon Alley.	2
3 THE LETTERS FROM NO ONE The escape of the Boggart in the cupboard.	3
4 THE KEEPER OF THE KEYS BOOM. They knocked on the door of number twelve, Privet Drive.	4
5 DIAGON ALLEY Harry woke early the next morning, and the first thing he saw was a small owl perched on his bed.	5
6 THE JOURNEY FROM PLATFORM NINE AND THREE-QUAARTERS Harry and Ron had to go through the Marauder's Map to get to Platform Nine and Three-Quarters.	6
7 THE SORTING HAT The door swung open at once.	7
8 THE POTIONS MASTER There, look." "Where?" "It's here."	8
9 THE MIDNIGHT DUEL Harry had never believed he had been chosen for the task.	9
10 HALLOWEEN Malfoy couldn't believe his eyes when he saw the three boys.	10
11 QUIDDITCH As they entered November, the weather turned cold.	11
12 THE MIRROR OF ERISSED Christmas was coming.	12
13 NICOLAS FLAMEL Dumebol had convinced Harry that he was the last person to tell him the truth.	13
14 NORBERT THE NORWEGIAN RIDGEBACK Quirrell, who had been sent to collect the dragon.	14
15 THE FORBIDDEN FOREST Things couldn't have been worse for Harry.	15
16 THROUGH THE TRAPDOOR In years to come, Harry would always remember the trapdoor.	16
17 THE MAN WITH TWO FACES It was Quirrell. "You're not going to tell anyone, are you?"	17

TOKENS

Element of the text

Word

n-gram

Sentence

Verse

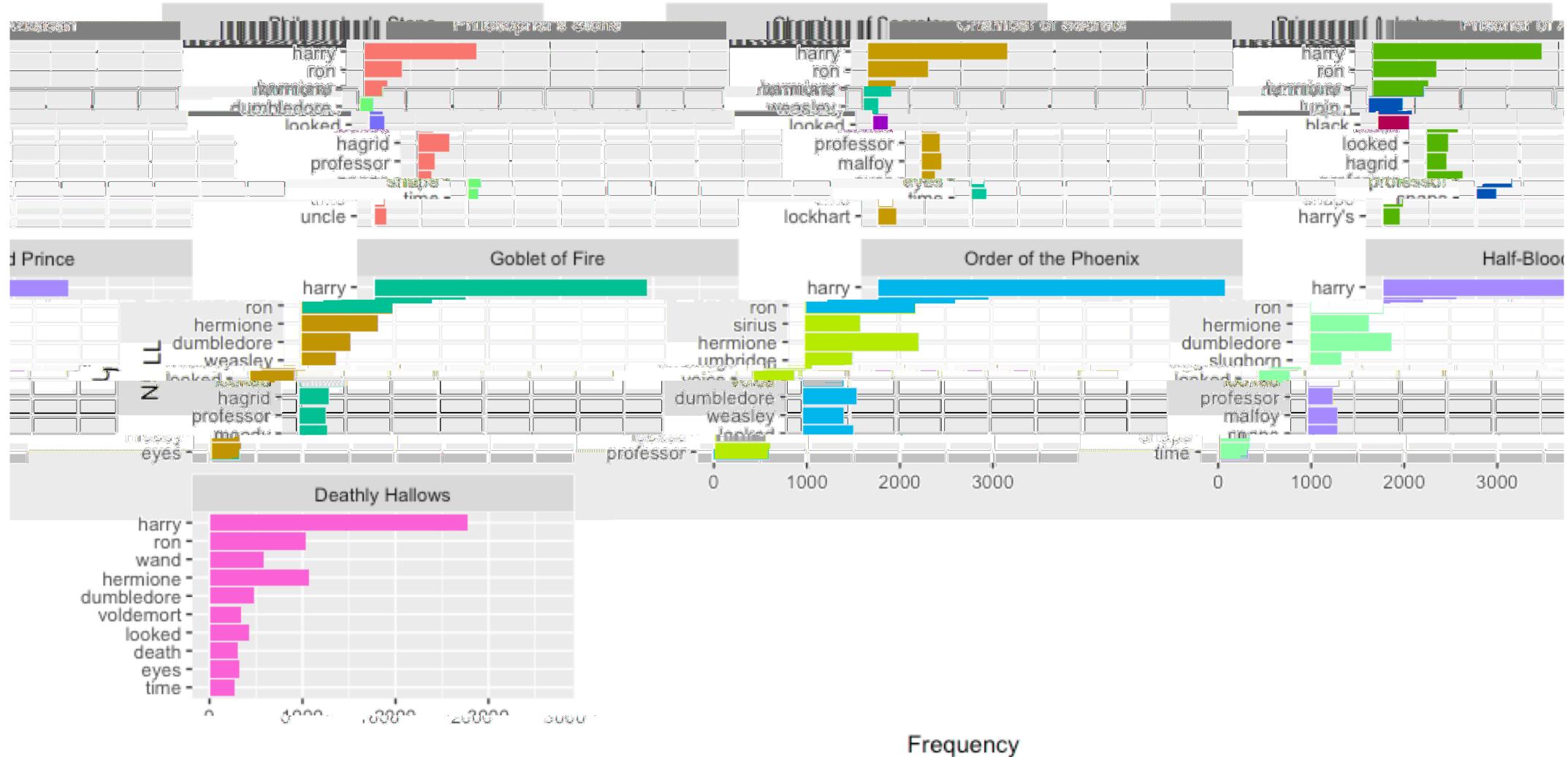
Line

Paragraph

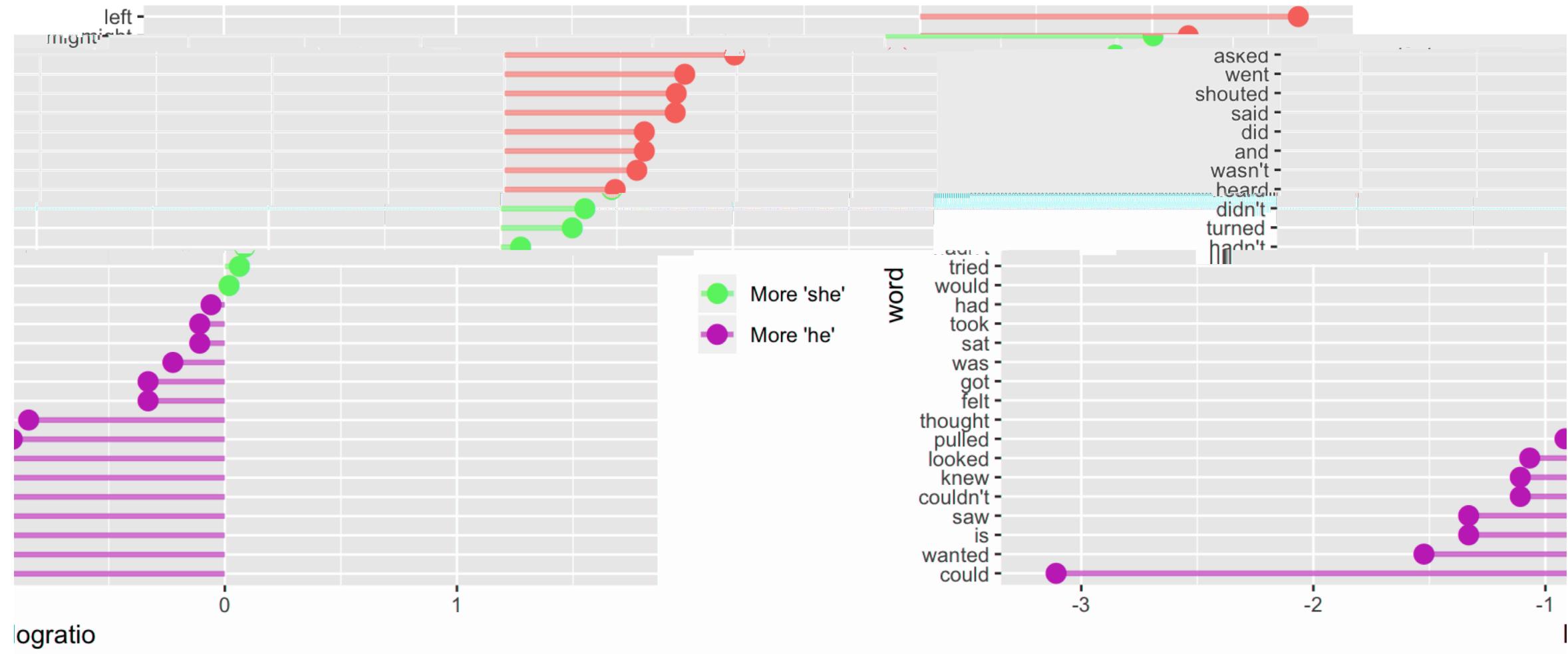
rank	chapter	word	book
1	1	the	Harry Potter and the Philosopher's Stone
2	1	boy	Harry Potter and the Philosopher's Stone
3	1	who	Harry Potter and the Philosopher's Stone
4	1	lived	Harry Potter and the Philosopher's Stone
5	1	mr	Harry Potter and the Philosopher's Stone
6	1	and	Harry Potter and the Philosopher's Stone
7	1	mrs	Harry Potter and the Philosopher's Stone
8	1	dursley	Harry Potter and the Philosopher's Stone
9	1	of	Harry Potter and the Philosopher's Stone
10	1	number	Harry Potter and the Philosopher's Stone
11	1	four	Harry Potter and the Philosopher's Stone
12	1	privet	Harry Potter and the Philosopher's Stone
13	1	drive	Harry Potter and the Philosopher's Stone
14	1	were	Harry Potter and the Philosopher's Stone
15	1	proud	Harry Potter and the Philosopher's Stone

rank	bigram	n
6	to the	173
7	out of	148
8	at the	141
9	said harry	136
10	he had	115
11	said ron	109
12	to be	109
13	he said	106
14	in a	105
15	was a	99
16	uncle vernon	97
17	they were	92
18	professor mcgonagall	90
19	said hagrid	89
20	in his	86
21	going to	85
22	into the	85
	the	82
	in the	80
	and	23
	from	24

TOKEN FREQUENCY



N-GRAM FREQUENCY



PARTS OF SPEECH

	id	sid	tid	word	lemma	upos	pos	cid
1	1	1	1	THE	the	DET	DT	0
				POY	buy	PPRON	NNP	
	1	1	3	WHO	who	NOUN	WP	8 3
			4	LIVED	live	VERB	VBD	12 4
	5	5	5	SPACE	SPACE			
6	Mr.	mr.		PROPN	NNP	19	6 1	1
7	and	and		CCONJ	CC	23	7 1	1
8	Mrs.	mrs.		PROPN	NNP	27	8 1	1
9	Dursley	dursley		PROPN	NNP	32	9 1	1
	,	,		PUNCT	,	39	10 1	
	f	o	r	ADP	in	41	11 1	
	NOUN	NN		44	12 1	12	number	number
	NUM	CD		51	13 1	1	four	four
	PUNCT	,		55	14 1	1	,	,
	PROPN	NNP		57	15 1	1	Privet	privet
	PROPN	NNP		64	16 1	1	Drive	drive
	PUNCT	,		69	17 1	1	,	,
	VERB	VBD		71	18 1	1	were	be

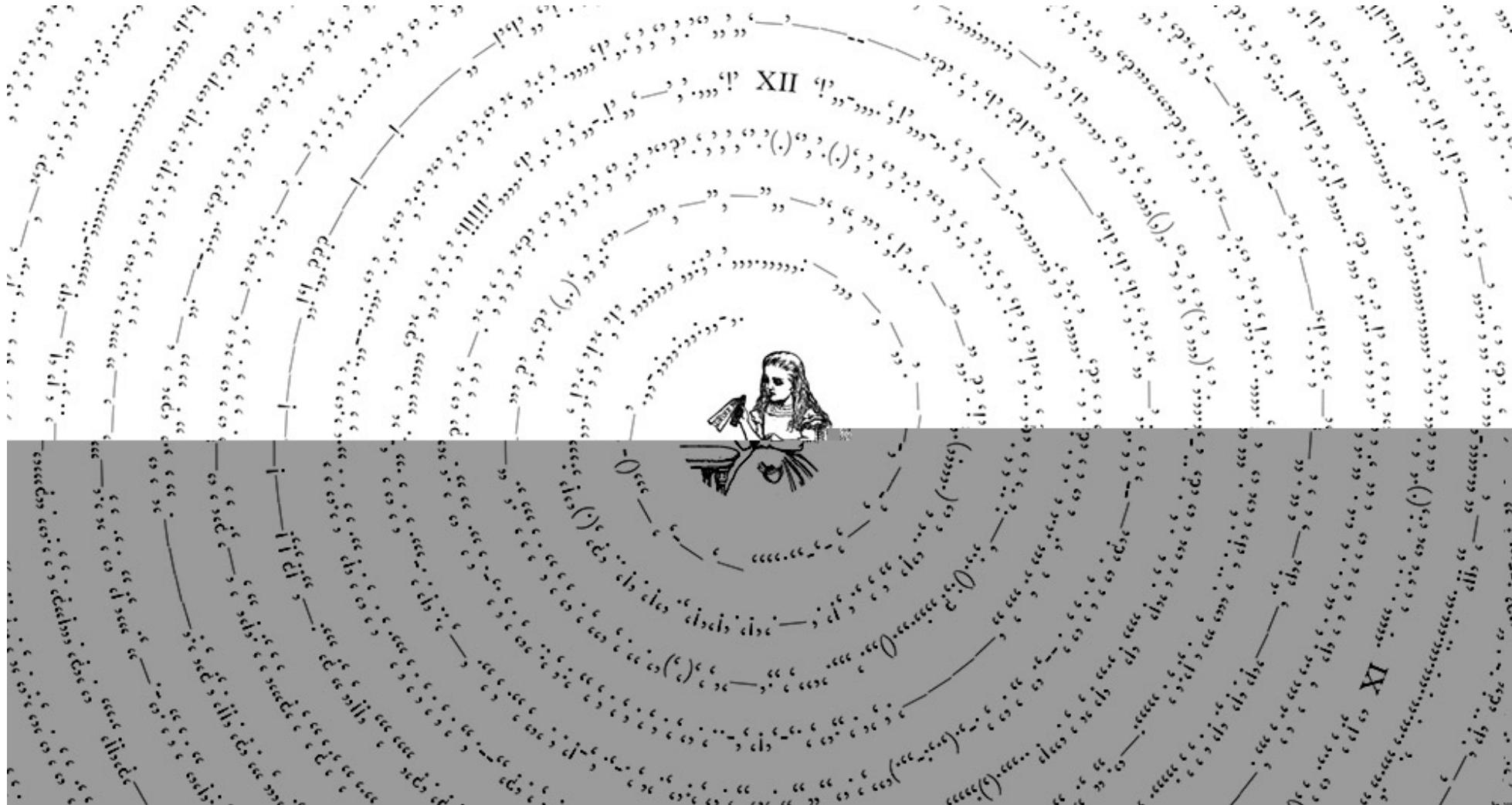
PART OF SPEECH FREQUENCY

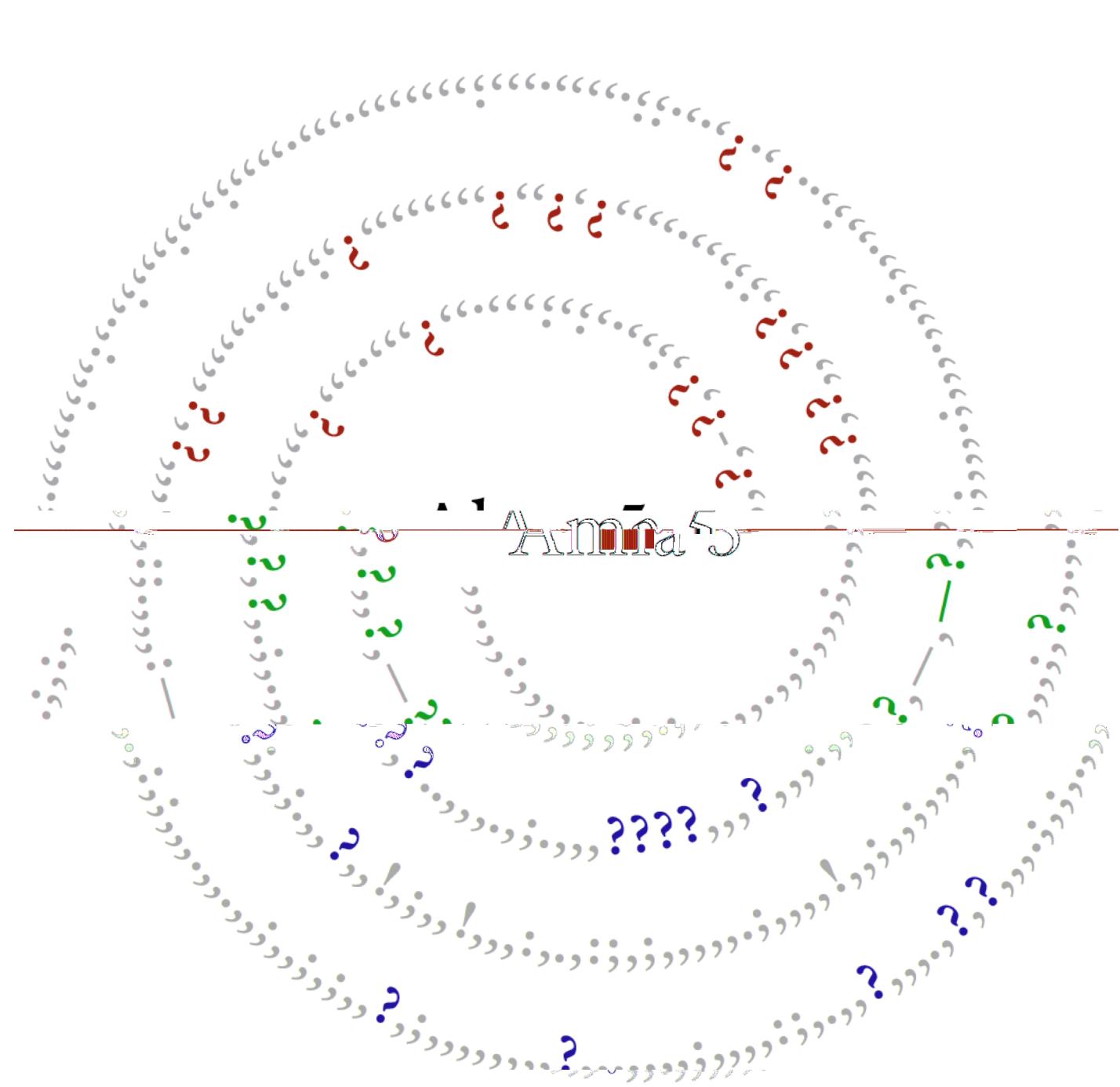
	lemma	n
300	1 be	33
298	2 have	12
925	3 say	9
	4 do	721
5	get	460
	6	406
	7 go	381
	8 look	303
	9 could	300
	10 see	293
	11 know	260
	12 will	232
	13 think	204
	14 come	181
tell		15
	16 take	110
	17 take	10

	lemma	n
1	go	142
2	look	109
3	try	64
4	say	44
5	be	38
6	do	35
7	come	34
8	get	33
9	stand	33
10	talk	28
11	hold	26
12	sit	26
13	star	23
14	smile	19
15	take	19
16	watch	19

	lemma	n
1	harry	1325
2	ron	429
3	hermione	261
4	hermione	260
181	5 professor	
172	6 snape	
153	7 dumbledore	
138	8 dudley	
122	9 malfoy	
117	10 potter	
11	vernon	116
12	quirrell	111
13	uncle	111
	mcgonagall	101
	otter	96
	ryffindor	85

ARTSY STUFF





SENTIMENT ANALYSIS

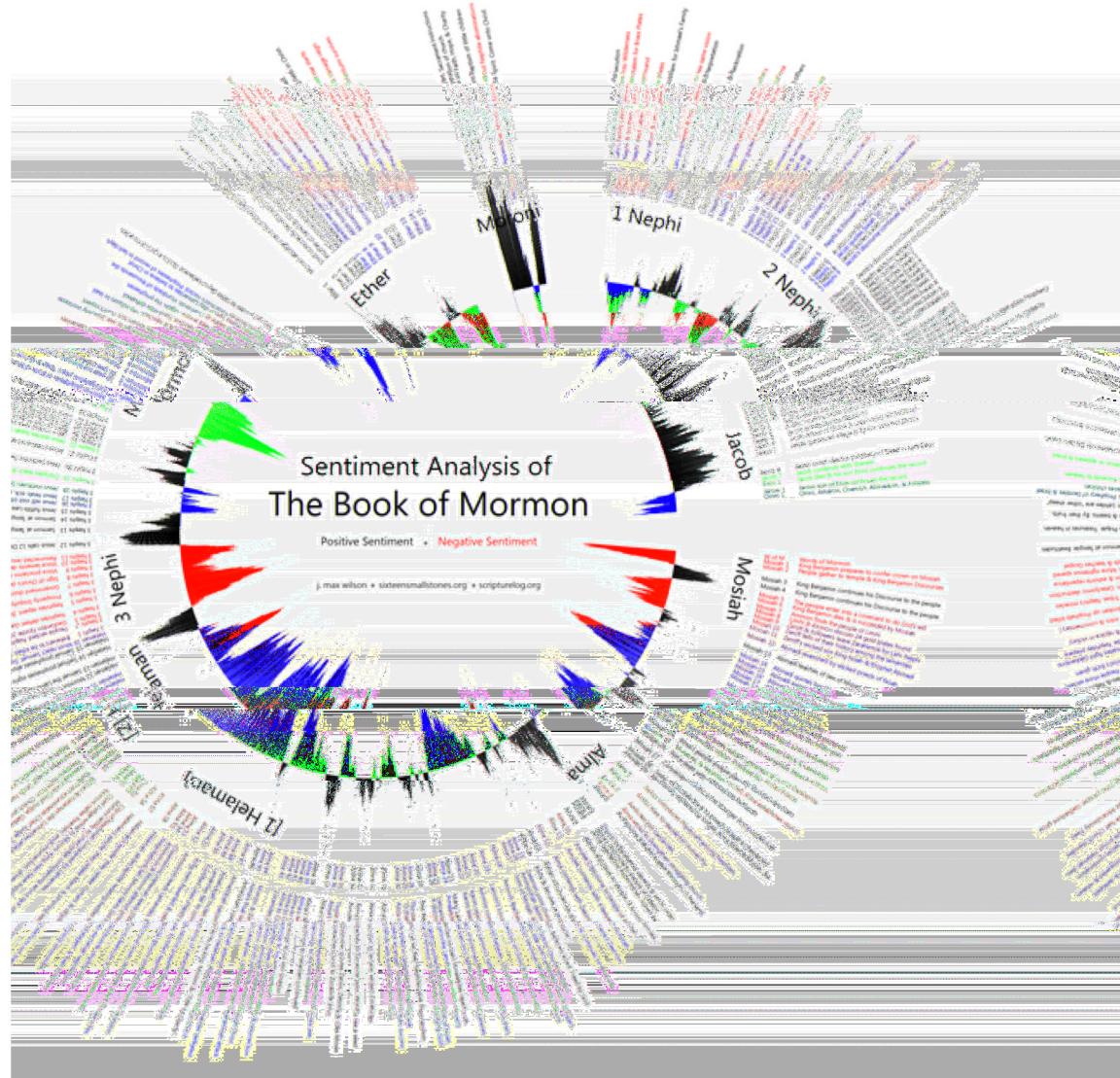
How positive or negative a text is

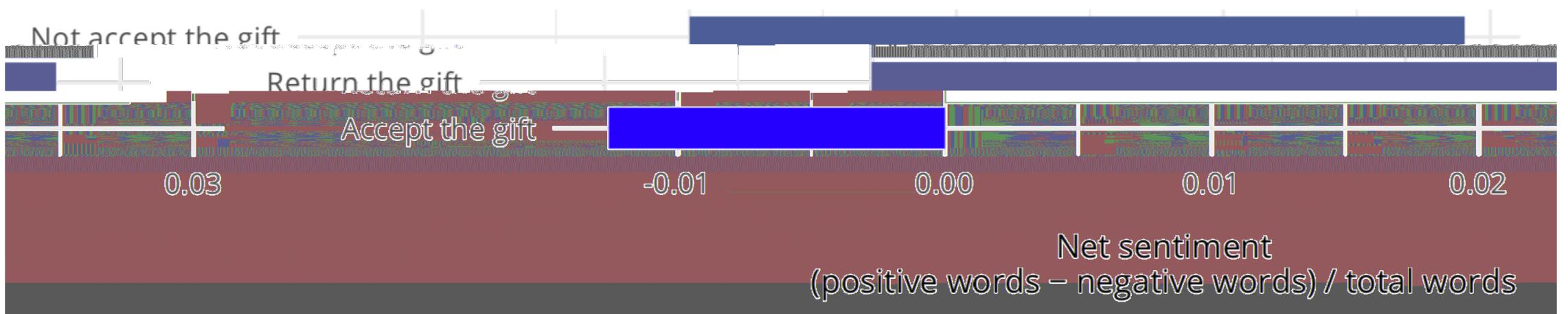
```
> get_sentiments("bing")
# A tibble: 6,788 x 2
  word      sentiment
  <chr>    <chr>
1 2-faced   negative
2 2-faces   negative
3 a+         positive
4 abnormal   negative
5 abolish   negative
6 abominable negative
7 abominably negative
8 abominate  negative
9 abomination negative
10 abort     negative
# ... with 6,778 more rows
```

```
> get_sentiments("afinn")
# A tibble: 2,476 x 2
  word      score
  <chr>     <int>
1 abandon    -2
2 abandoned  -2
3 abandons   -2
4 abducted   -2
5 abduction  -2
6 abductions -2
7 abhor      -3
8 abhorred   -3
9 abhorrent  -3
10 abhors    -3
# ... with 2,466 more rows
```

```
> get_sentiments("loughran")
# A tibble: 4,149 x 2
  word      sentiment
  <chr>    <chr>
1 abandon   negative
2 abandoned negative
3 abandoning negative
4 abandonment negative
5 abandonments negative
6 abandons   negative
7 abdicated  negative
8 abdicates  negative
9 abdicating negative
10 abdication negative
# ... with 4,139 more rows
```

SENTIMENT ANALYSIS





TF - IDF

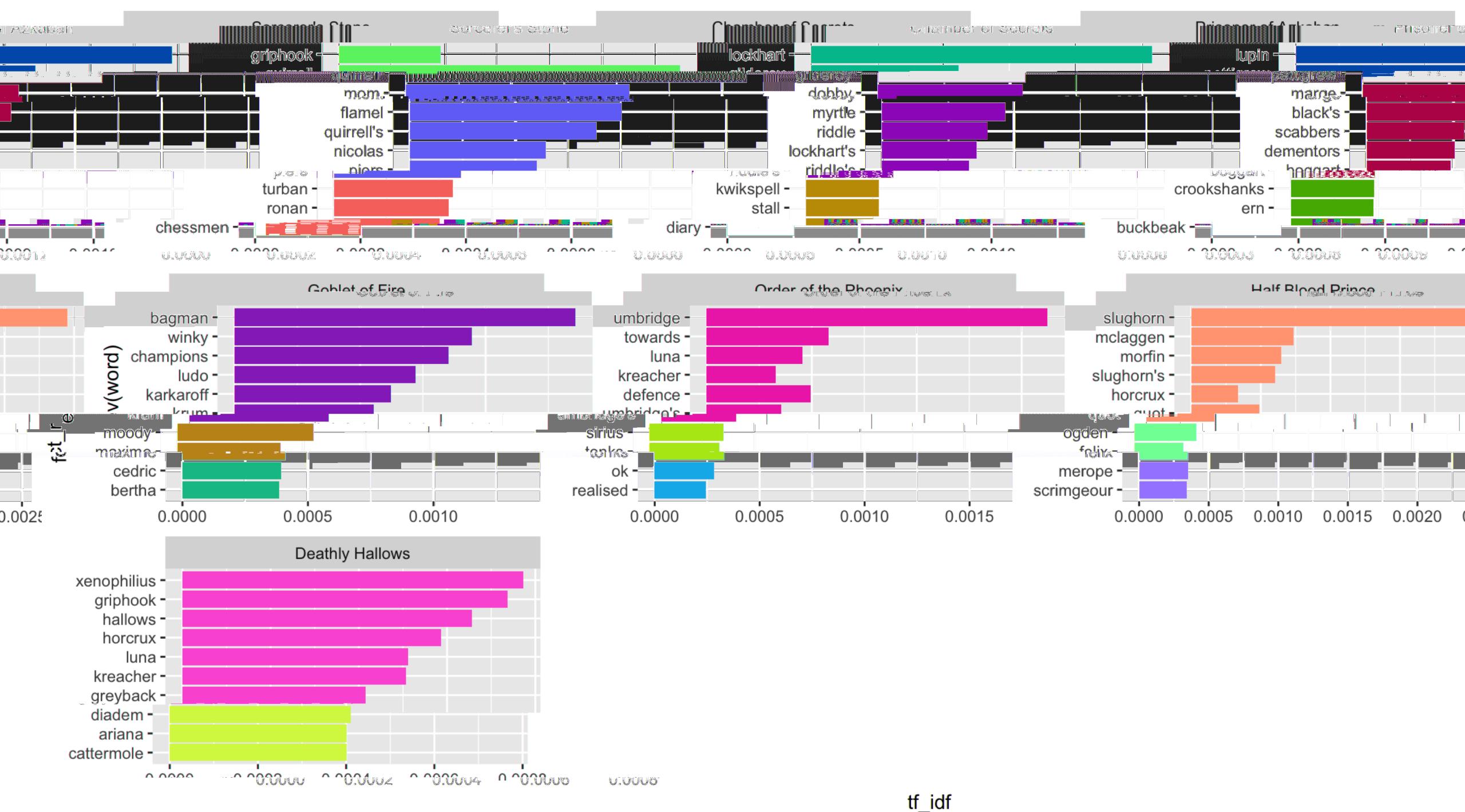
Term frequency-inverse document frequency

How important a term is compared to the rest of the documents

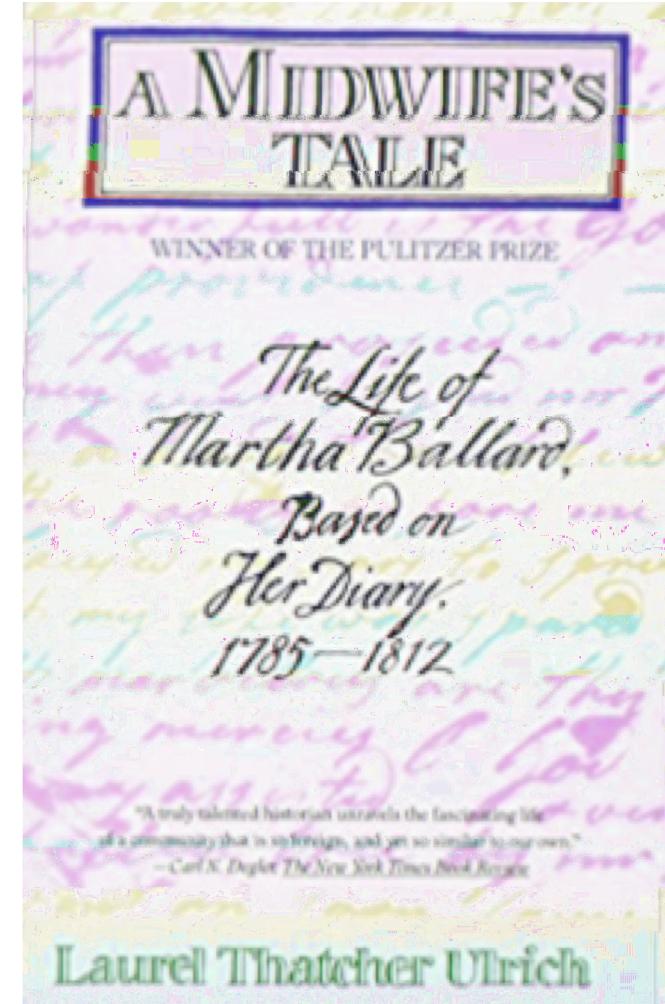
$$tf(\text{term}) = \frac{n_{\text{term}}}{n_{\text{terms in document}}}$$

$$idf(\text{term}) = \ln \left(\frac{n_{\text{documents}}}{n_{\text{documents containing term}}} \right)$$

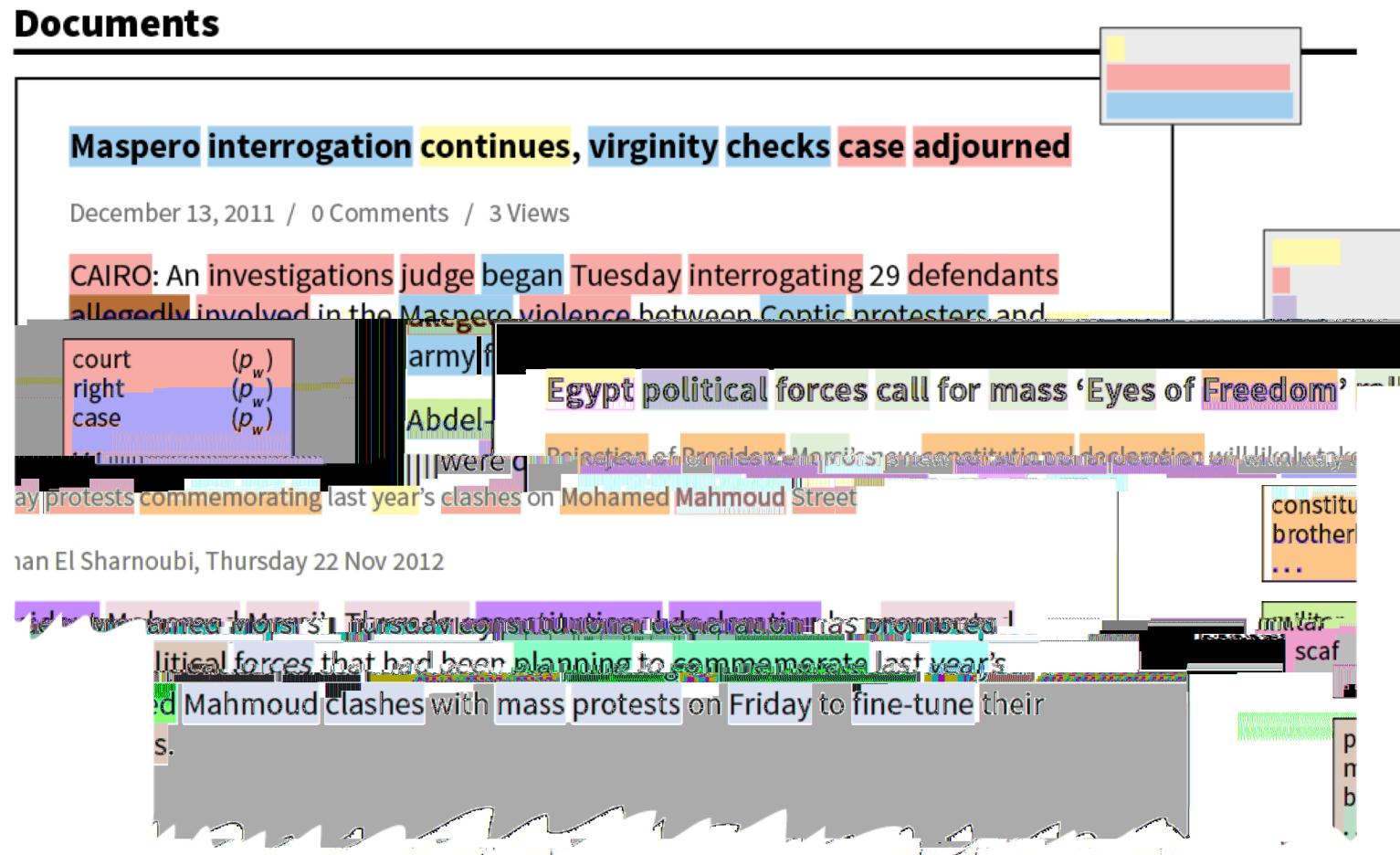
$$tf-idf(\text{term}) = tf(\text{term}) \times idf(\text{term})$$



TOPIC MODELING



LATENT DIRICHLET ALLOCATION



CLUSTERS OF RELATED WORDS

Topic Label Topic Words

MIDWIFERY birth deld safe morn receivd calld left cleverly pm labour fine reward arivd infant expected recd shee born patient

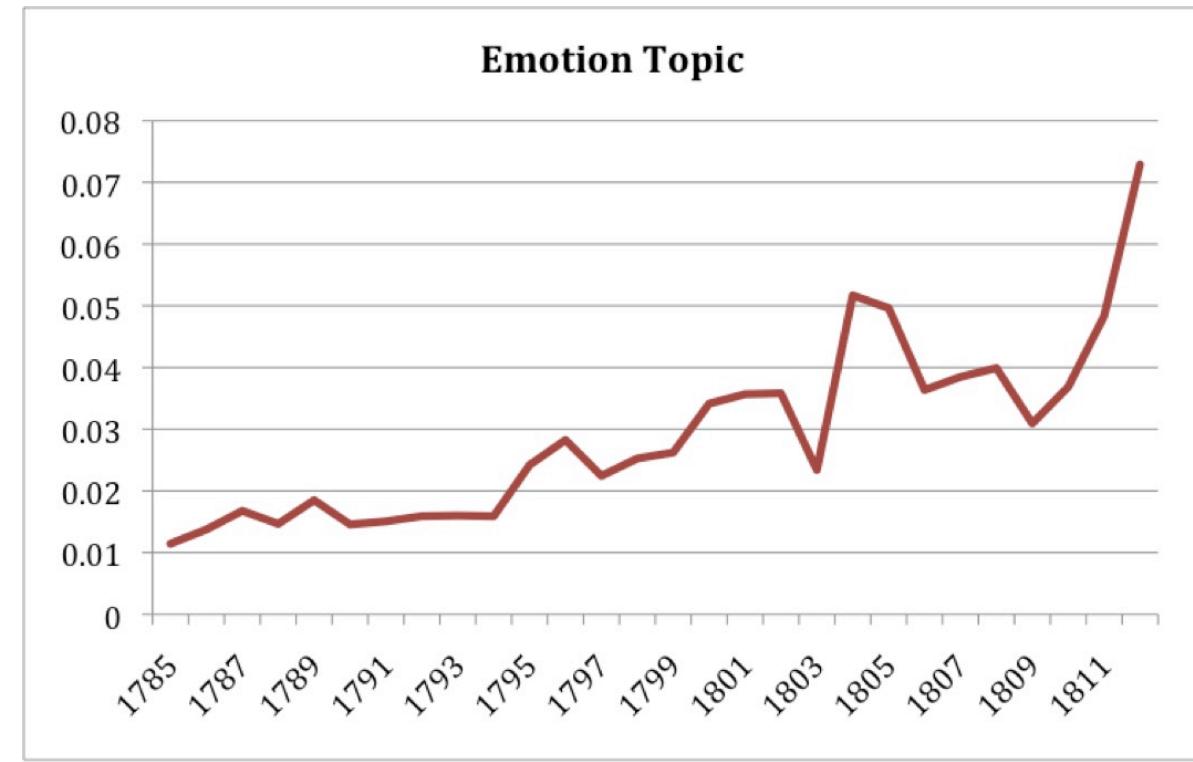
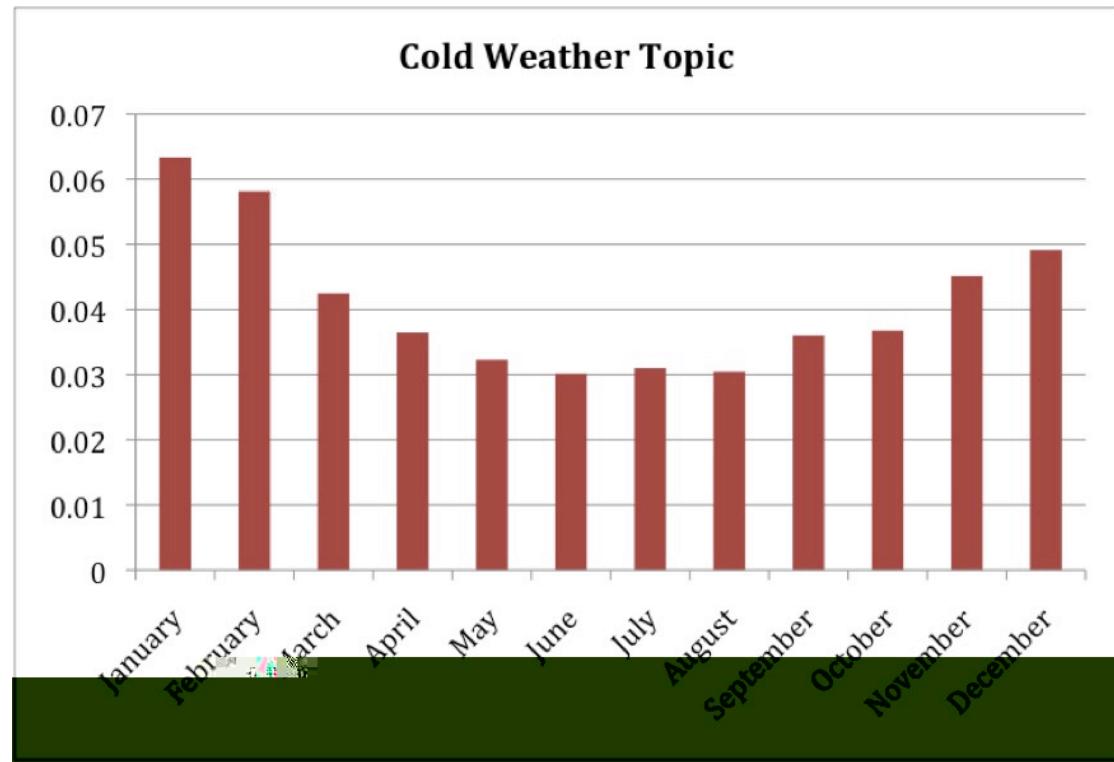
CHURCH meeting attended afternoon reverend worship foren mr famely performd vers attend public supper st service lecture discoarst a
yesterday informd morn years death ye hear expirid expird weak dead las past heard days drawnd departed evinn.

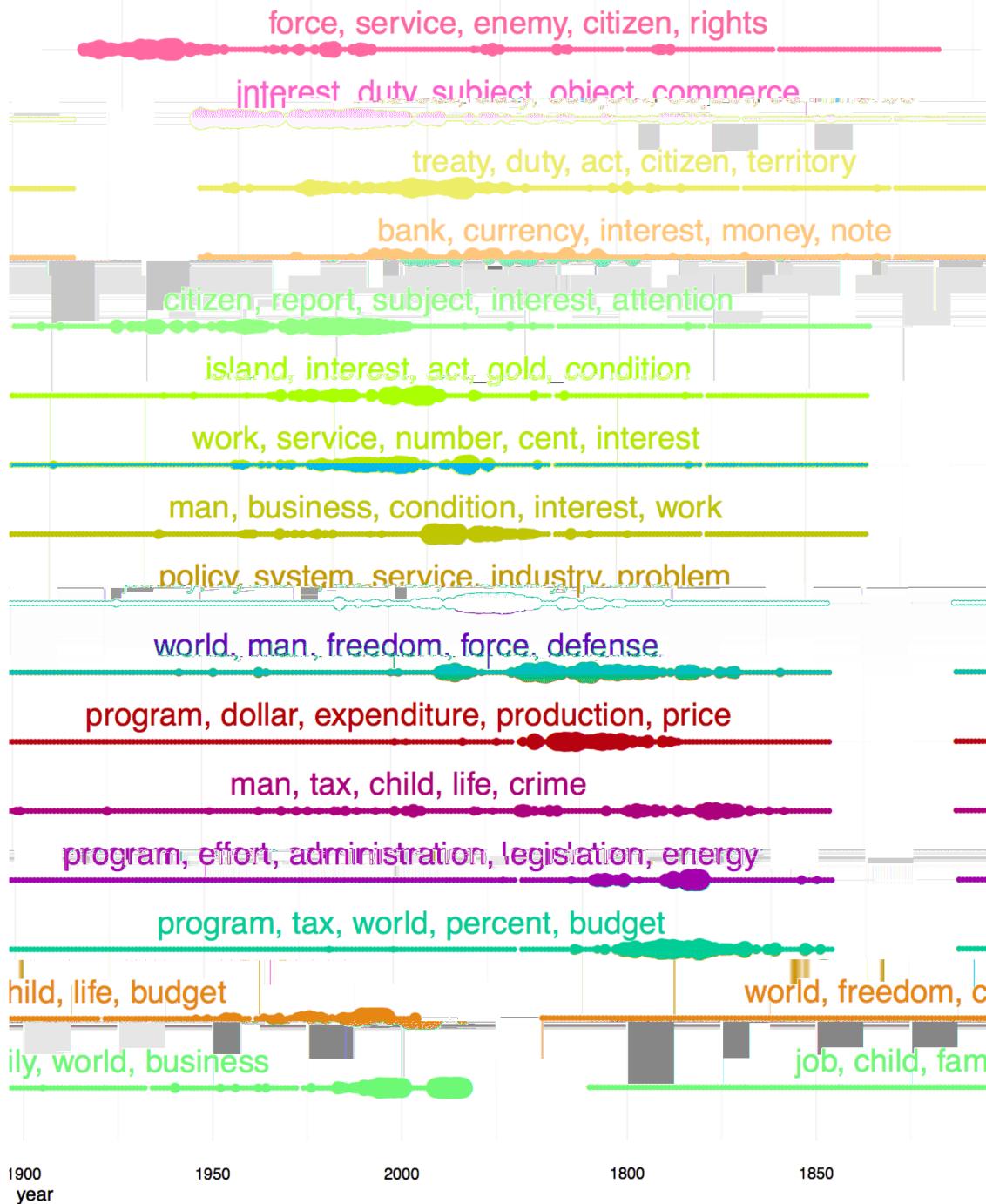
GARDENING gardin sett worked clear beans corn warr planted mattres cucumbers anthord potatoe y plants ou could does quach wud seeds

HOPPING lb made brot bot tea butter sugar carried oz chees pork candles wheat store pr beef spirit churnd flower S

ILLNESS unwell mr sick gave dr rainy easier care head neighbor feet relief made throat poorly takeing medisin rs stomach I

TRACK TOPICS OVER TIME



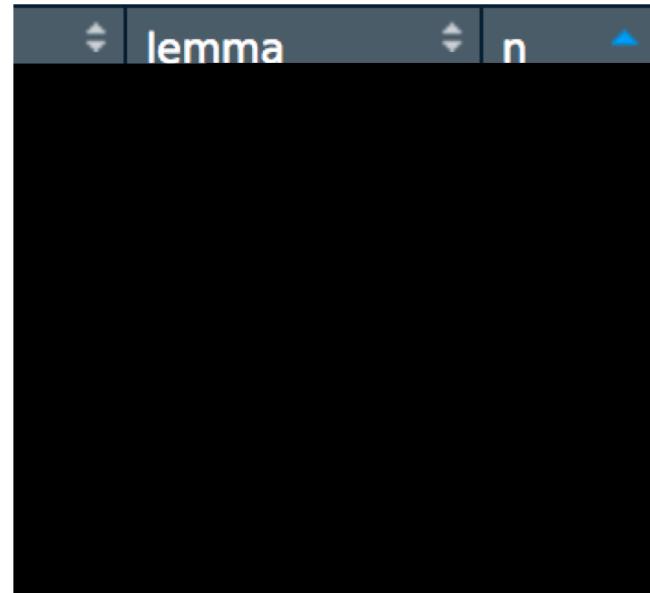


FINGERPRINTING

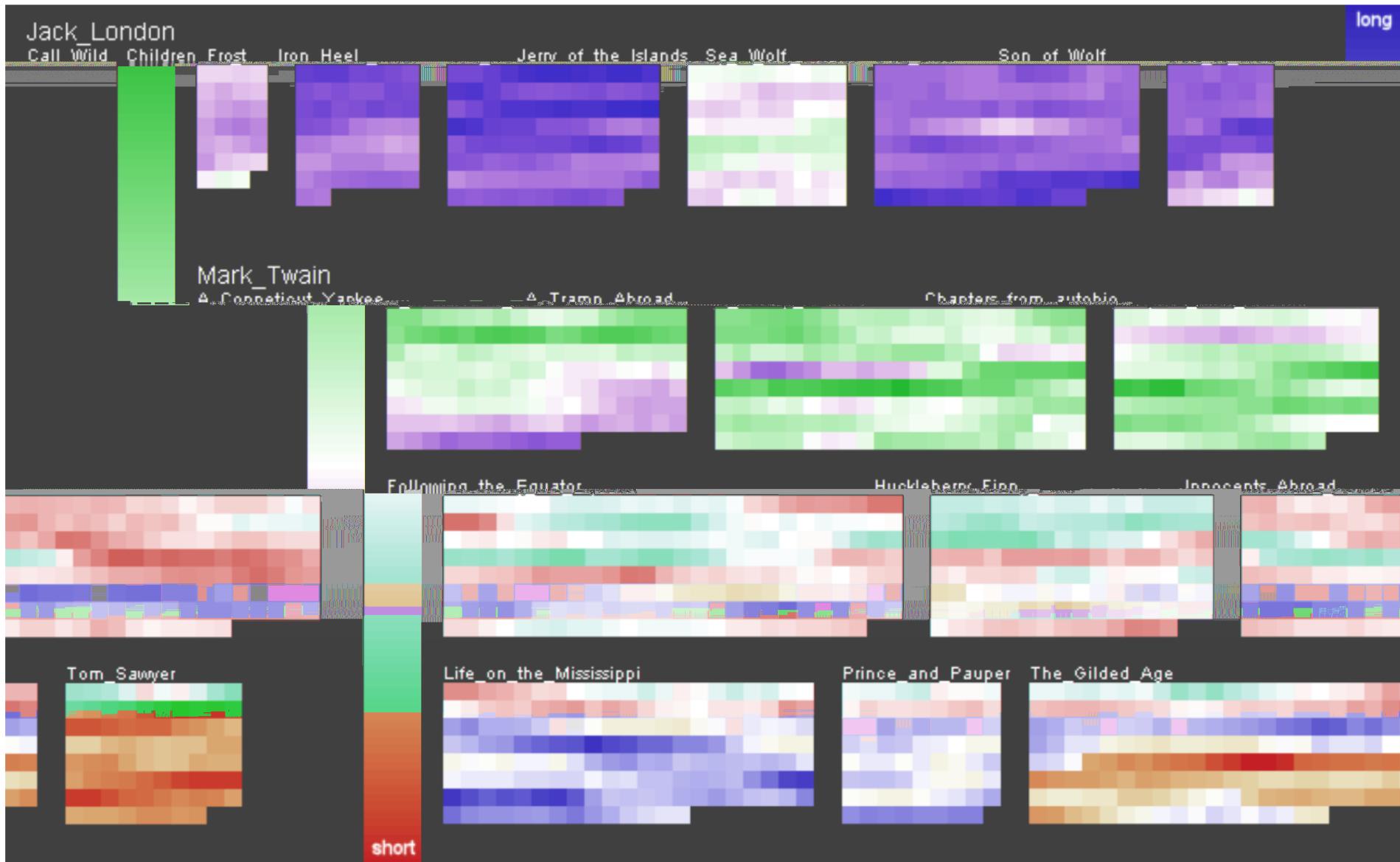
Analyze richness or uniqueness of document

Punctuation patterns, vocabulary choices, sentence length

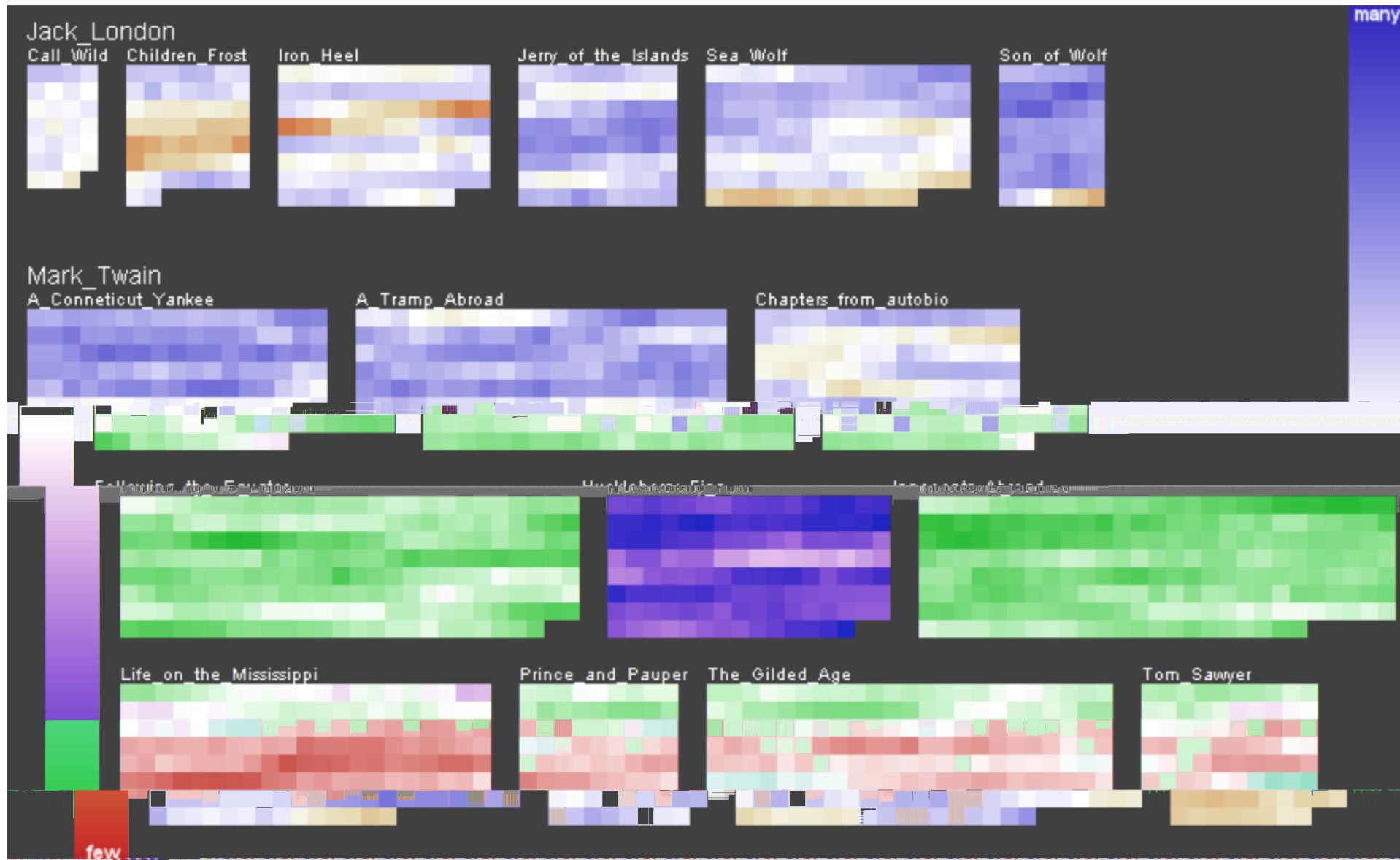
Hapax legomenon



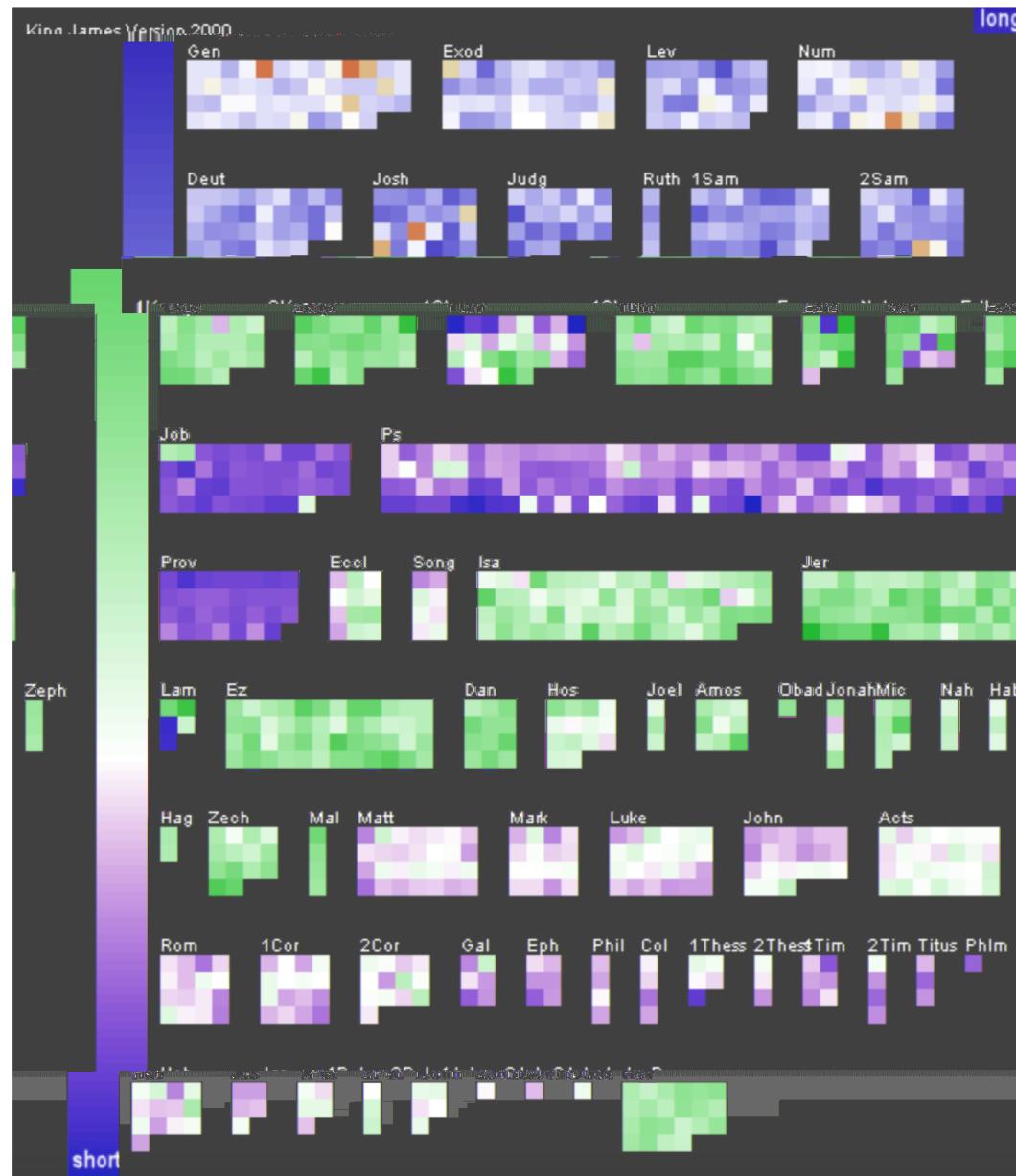
See See



Hapax legomena



Verse length



VISUALIZING TEXT WITH R