TEXT

MPA 635: Data Visualization

November 13, 2018

PLAN FOR TODAY

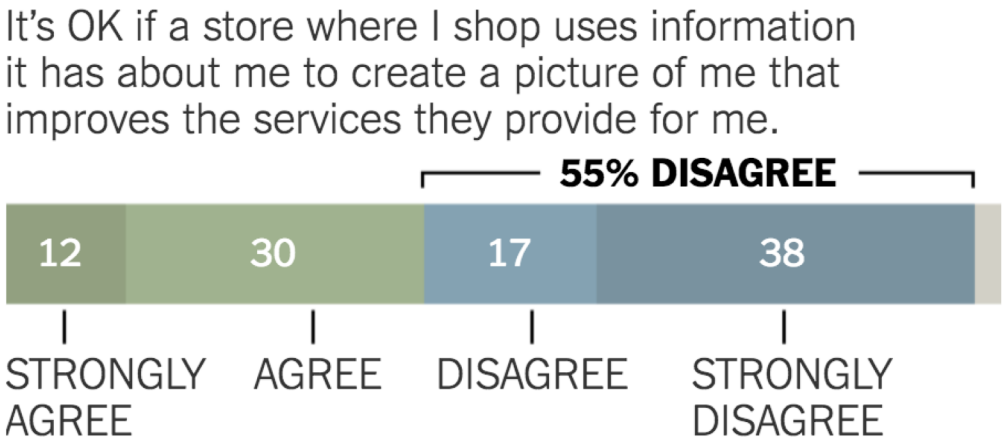
Surveys and qualitative data

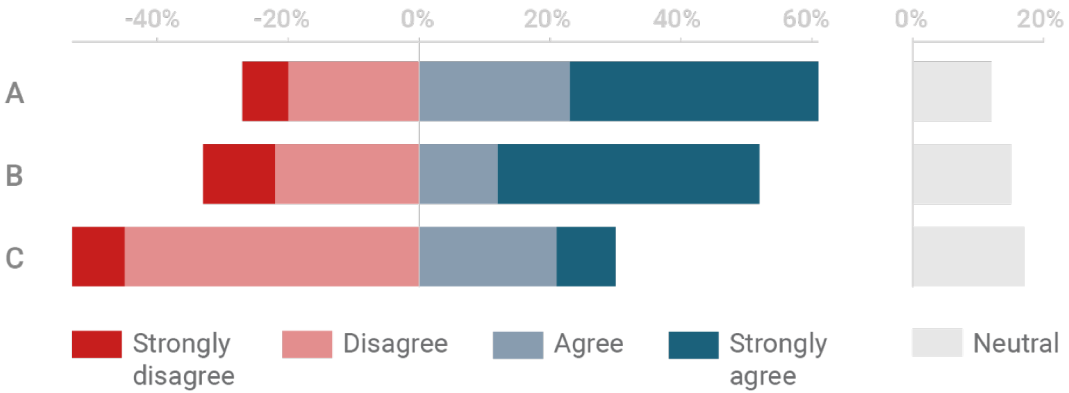
Digital humanities

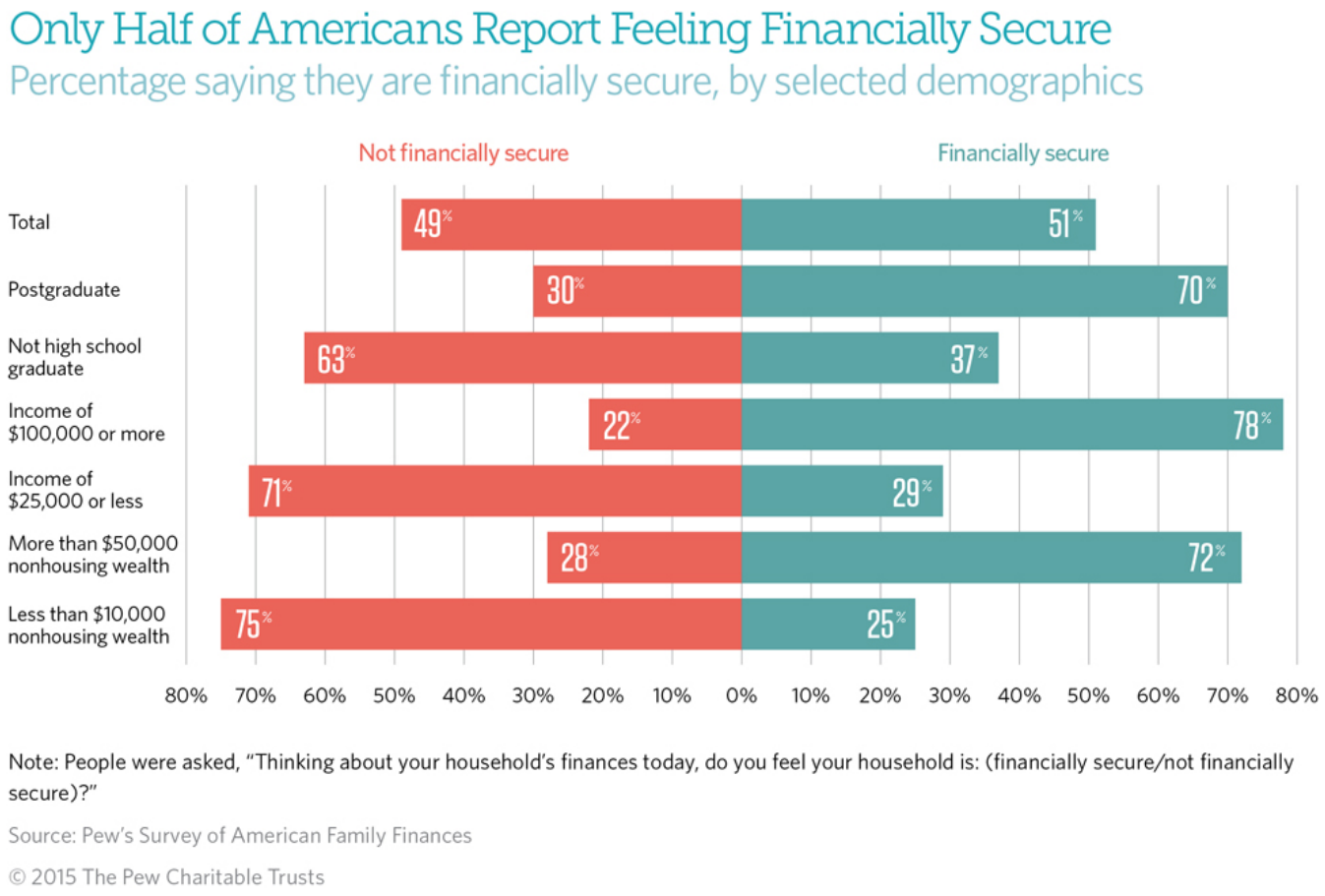
Visualizing text with R

SURVEYS AND QUALITATIVE DATA

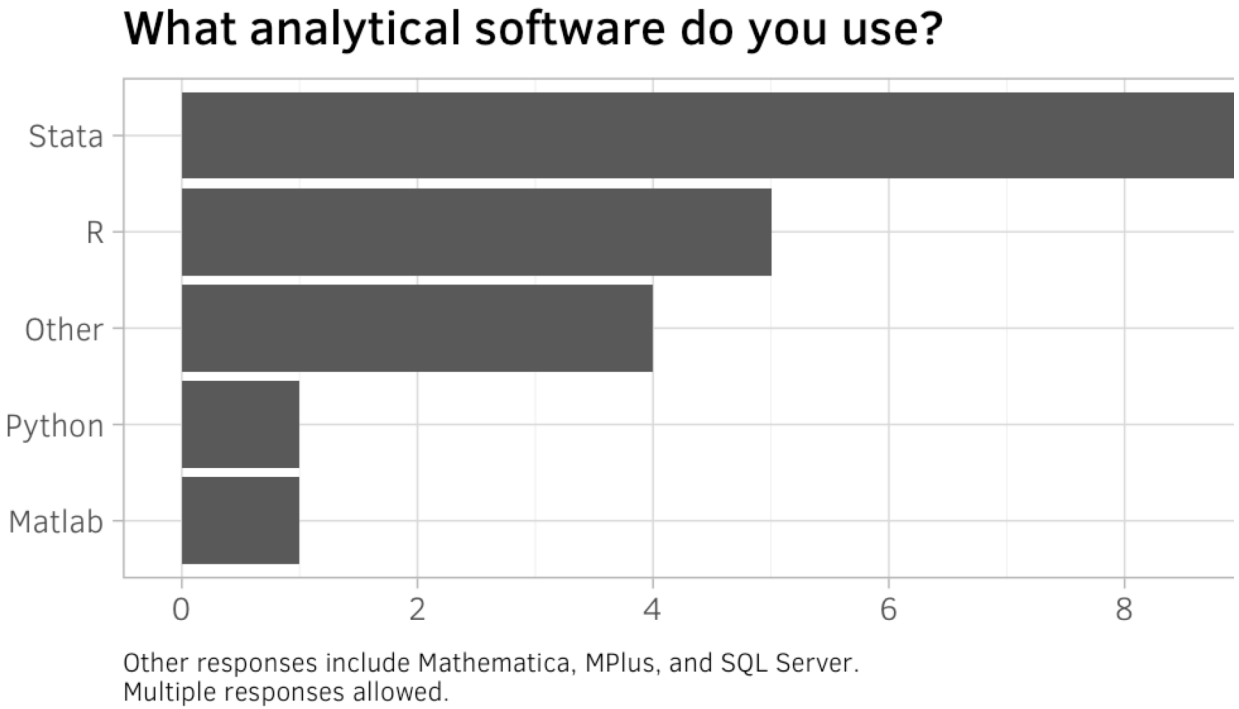
SINGLE RESPONSE QUESTIONS

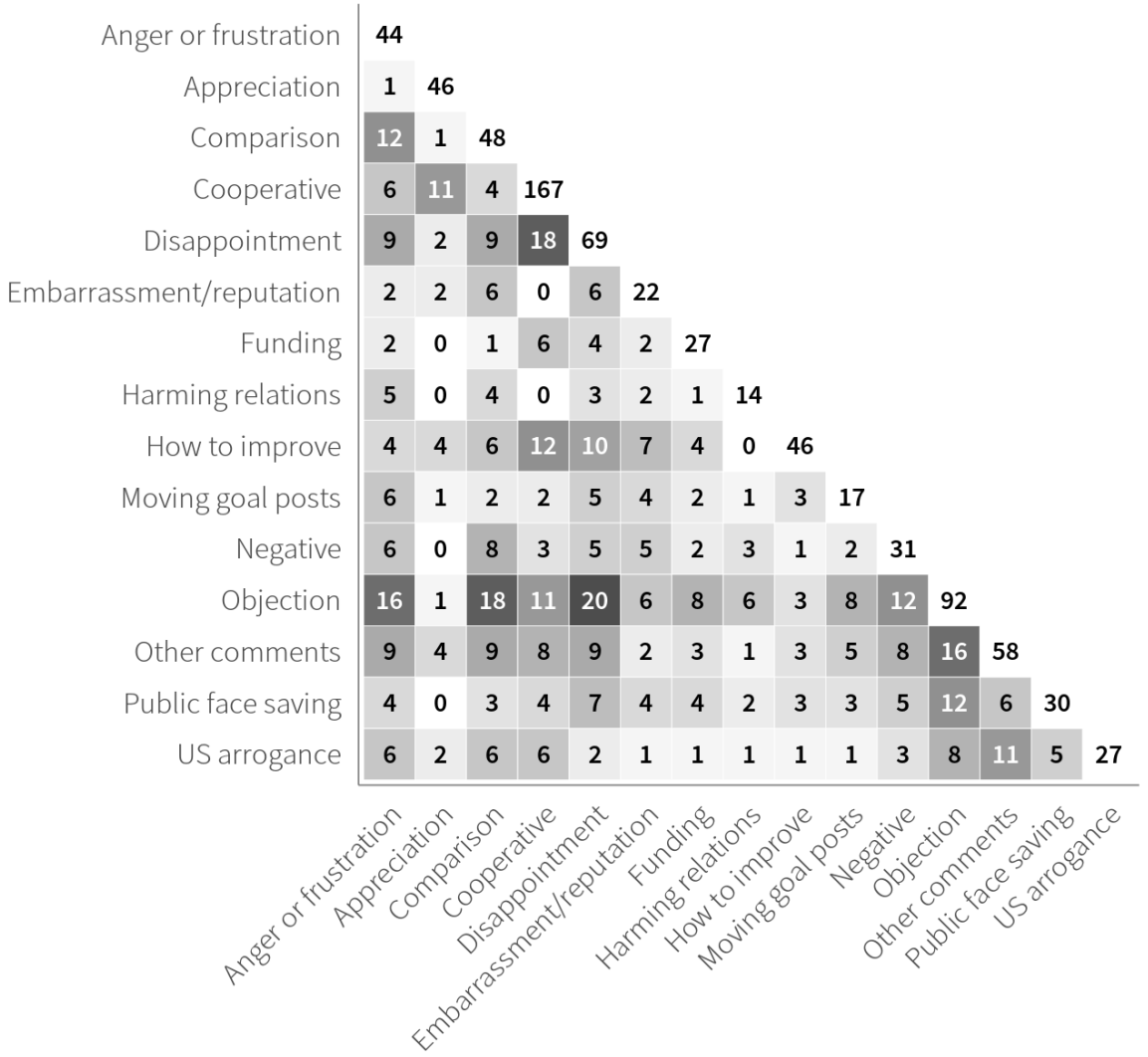




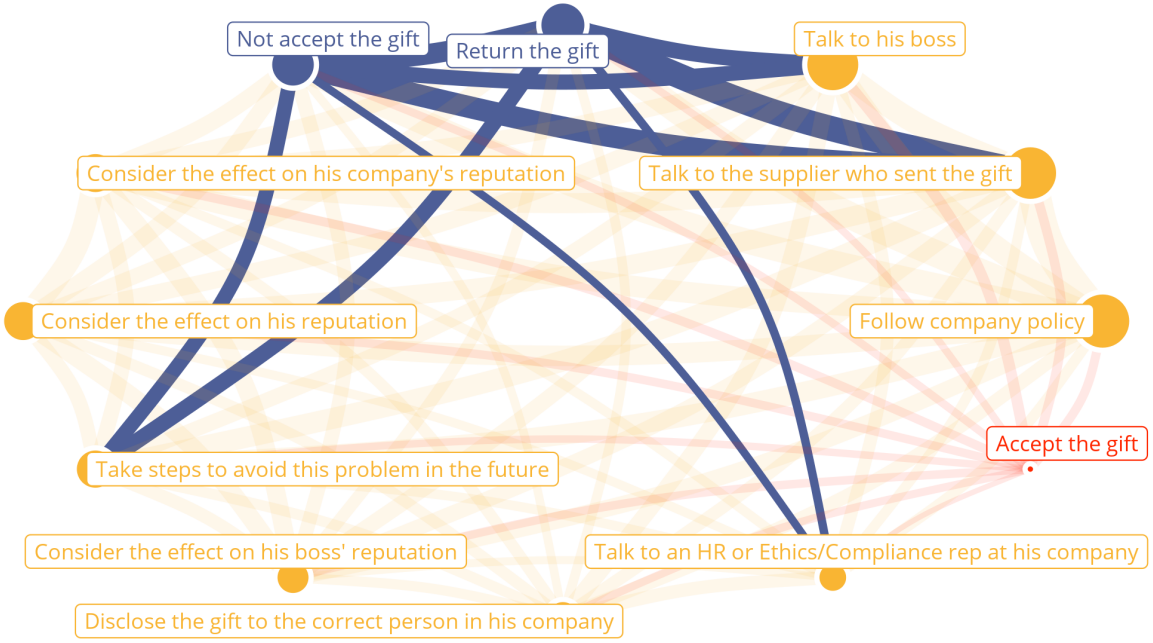


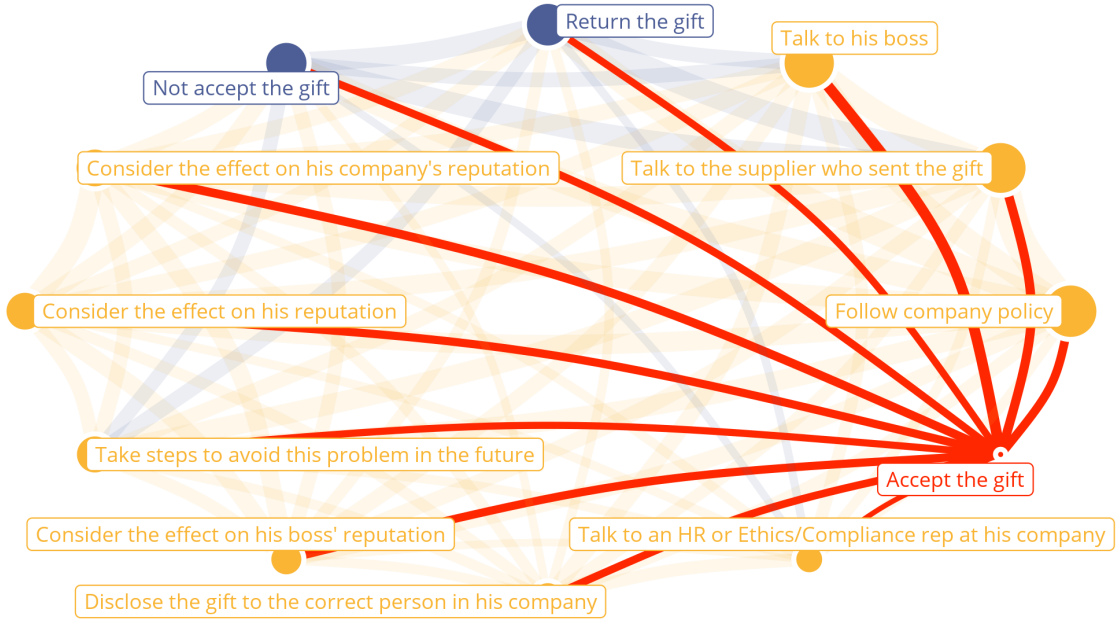
MULTIPLE RESPONSE QUESTIONS





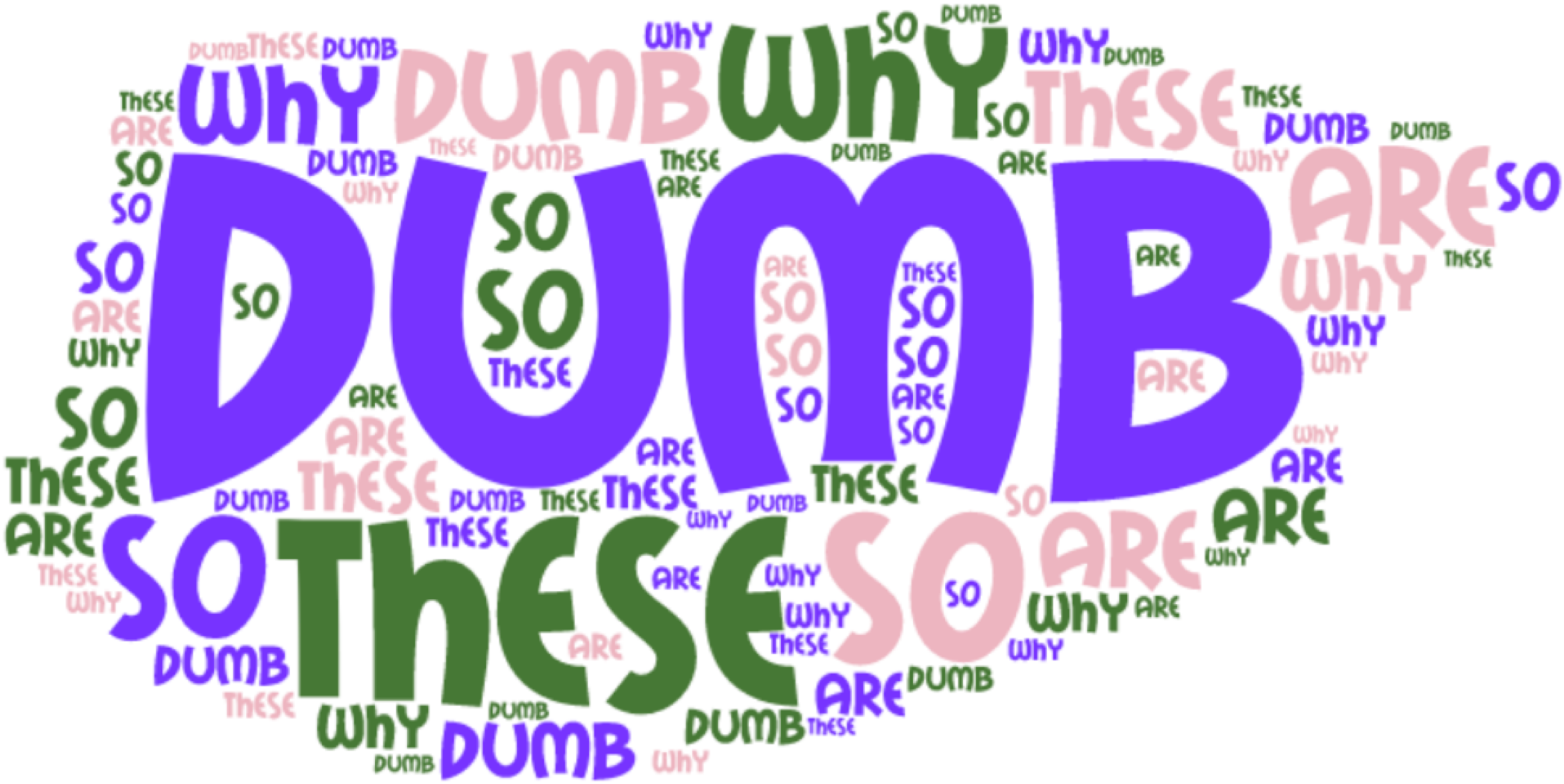
COOCCURRENCE ANALYSIS



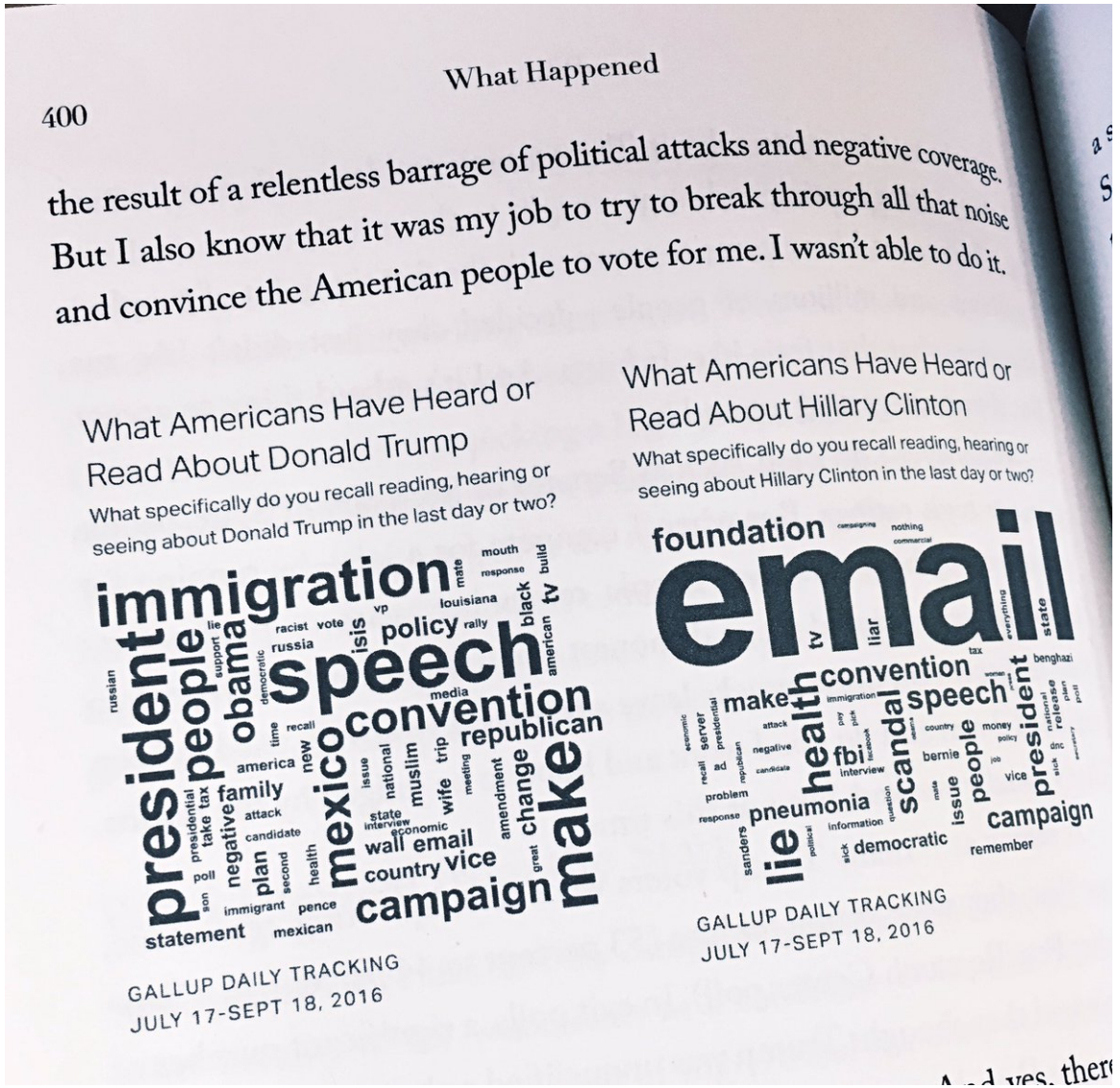


FREE RESPONSES

N	O	P	
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 wo
Somewhat likely	10	90	I donate the amount that I usually would
Somewhat unlikely	0	100	i'm poor
Neither likely nor unlikely	10	90	It is not a cause that is very important to me. i have other things t
Extremely likely	29	71	I want to contribute to the cause, but also keep some of the mone
Somewhat likely	20	80	It's a reasonable amount of money for an individual to donate to a
Extremely unlikely	0	100	I don't fully agree with their mission
Somewhat likely	10	90	I am pretty poor so I need to keep some for myself, but I also war
Extremely likely	5	95	I think it would be a good amount to give from the money I have a
Neither likely nor unlikely	69	31	to help with their cause
Somewhat unlikely	0	100	My dad always told me to give until it hurts, and right now I am hu
Neither likely nor unlikely	0	100	I would rather keep the money for myself and find a charity that I
Extremely unlikely	0	100	I want the most for myself.
Neither likely nor unlikely	5	95	Can afford to give a little
Extremely unlikely	0	100	Because I would then have 100$ more dollars.
Extremely unlikely	0	100	I'm a broke boi. If anyone need humanitarian aid, it's me.
Somewhat likely	10	90	I'm in a position where I would need the extra money, but I also w
Somewhat unlikely	90	10	I think it is a worthy cause and I think donating 90% of the amoun
Extremely likely	50	50	I feel splitting it 50/50 would be a fair deal. I get to help make a di
Extremely likely	20	80	I feel that my contribution is enough. I would gladly donate again
Somewhat likely	9	91	give a little
Somewhat likely	1	99	I like money
Somewhat unlikely	0	100	I do not really know what they will do with the money.


DUMB
WHY
THESE
SO
ARE

IS THIS OKAY?



the result of a relentless barrage of political attacks and negative coverage. But I also know that it was my job to try to break through all that noise and convince the American people to vote for me. I wasn't able to do it.

What Americans Have Heard or Read About Donald Trump

What specifically do you recall reading, hearing or seeing about Donald Trump in the last day or two?

GALLUP DAILY TRACKING
JULY 17-SEPT 18, 2016

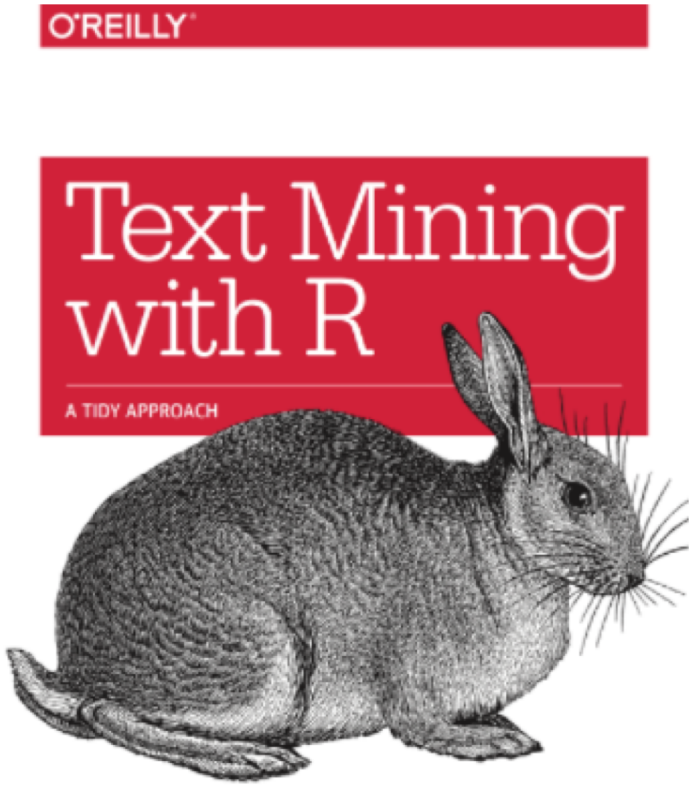
What Americans Have Heard or Read About Hillary Clinton

What specifically do you recall reading, hearing or seeing about Hillary Clinton in the last day or two?

GALLUP DAILY TRACKING
JULY 17-SEPT 18, 2016

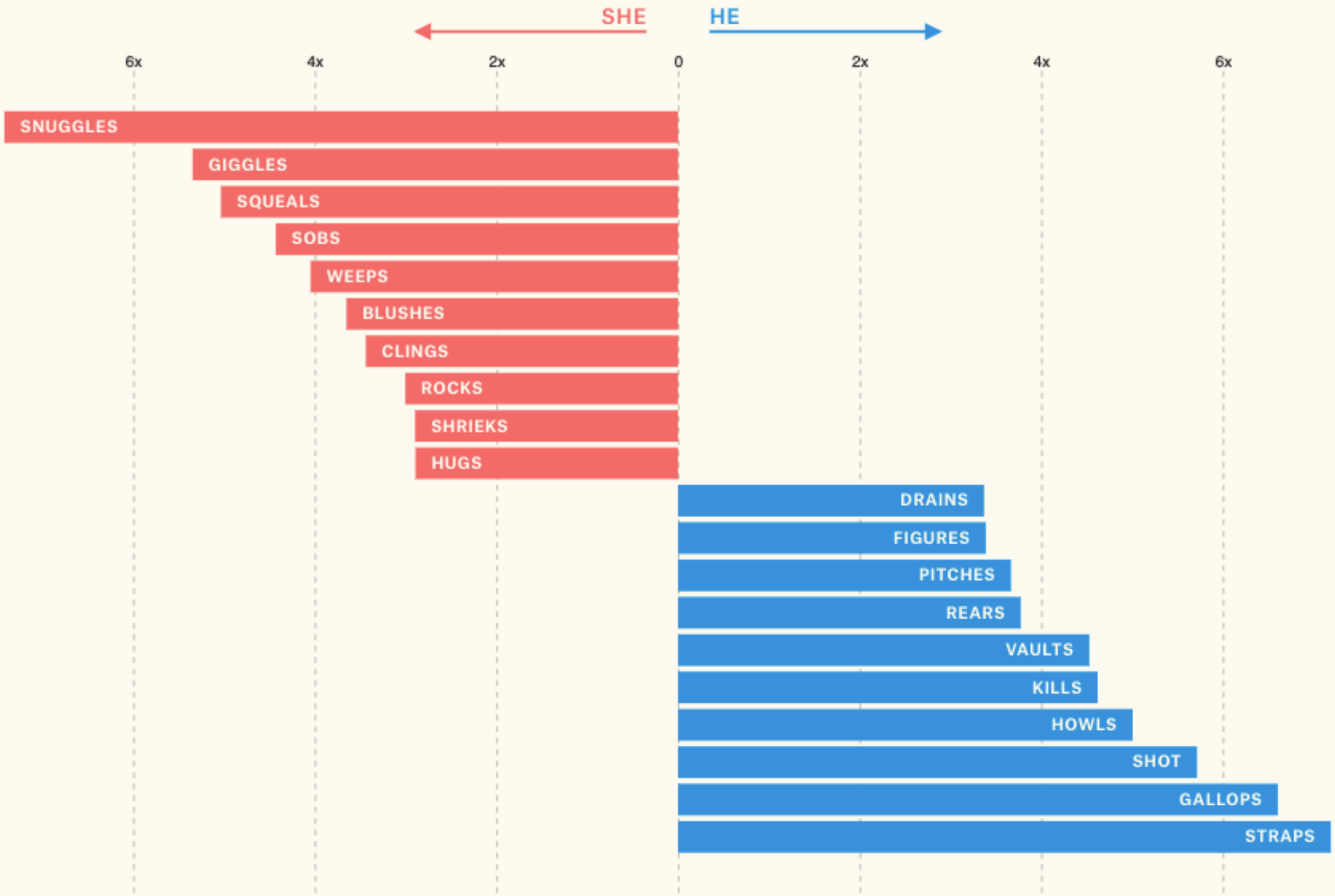
WORD CLOUDS FOR GROWNUPS

Counting words, but in fancier ways



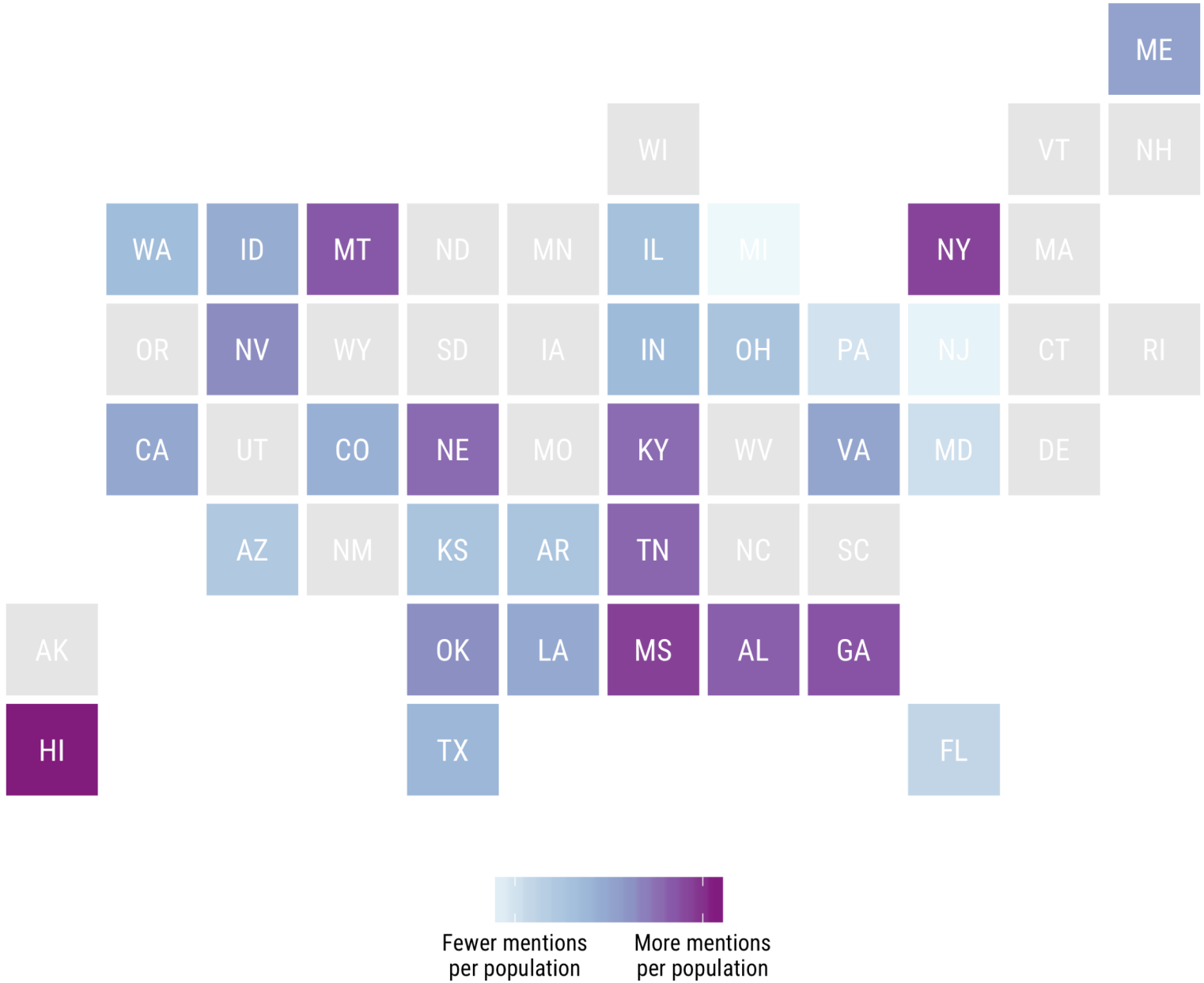
The most used words for women vs. men

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



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 any neck, although he did have a very large mustache. Mrs. Dursley was thin and blonde and had nearly twice the
    usual amount of neck, which came in very useful as she spent so much of her time craning over garden fences, spying on the
    neighbors. The Dursleys had a small son called Dudley and in their opinion there was no finer boy anywhere.    The Dursleys had
    everything they 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
    several years; 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 hummed as he picked out his most boring
    tie for work, and Mrs. Dursley gossiped away happily as she wrestled a screaming Dudley into his high chair.    None of them
    noticed a large, tawny owl flutter past the window.    At half past eight, Mr. Dursley picked up 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 out of number
```

	text	chapter
1	THE BOY WHO LIVED Mr. and Mrs. Dursley, of nu...	1
2	THE VANISHING GLASS Nearly ten years had pas...	2
3	THE LETTERS FROM NO ONE The escape of the B...	3
4	THE KEEPER OF THE KEYS BOOM. They knocked ...	4
5	DIAGON ALLEY Harry woke early the next mornin...	5
6	THE JOURNEY FROM PLATFORM NINE AND THREE-Q...	6
7	THE SORTING HAT The door swung open at once....	7
8	THE POTIONS MASTER There, look." "Where?"...	8
9	THE MIDNIGHT DUEL Harry had never believed h...	9
10	HALLOWEEN Malfoy couldn't believe his eyes wh...	10
11	QUIDDITCH As they entered November, the weat...	11
12	THE MIRROR OF ERISED Christmas was coming. ...	12
13	NICOLAS FLAMEL Dumbledore had convinced Ha...	13
14	NORBERT THE NORWEGIAN RIDGEBACK Quirrell, ...	14
15	THE FORIBIDDEN FOREST Things couldn't have b...	15
16	THROUGH THE TRAPDOOR In years to come, Har...	16
17	THE MAN WITH TWO FACES It was Quirrell. "Y...	17

TOKENS

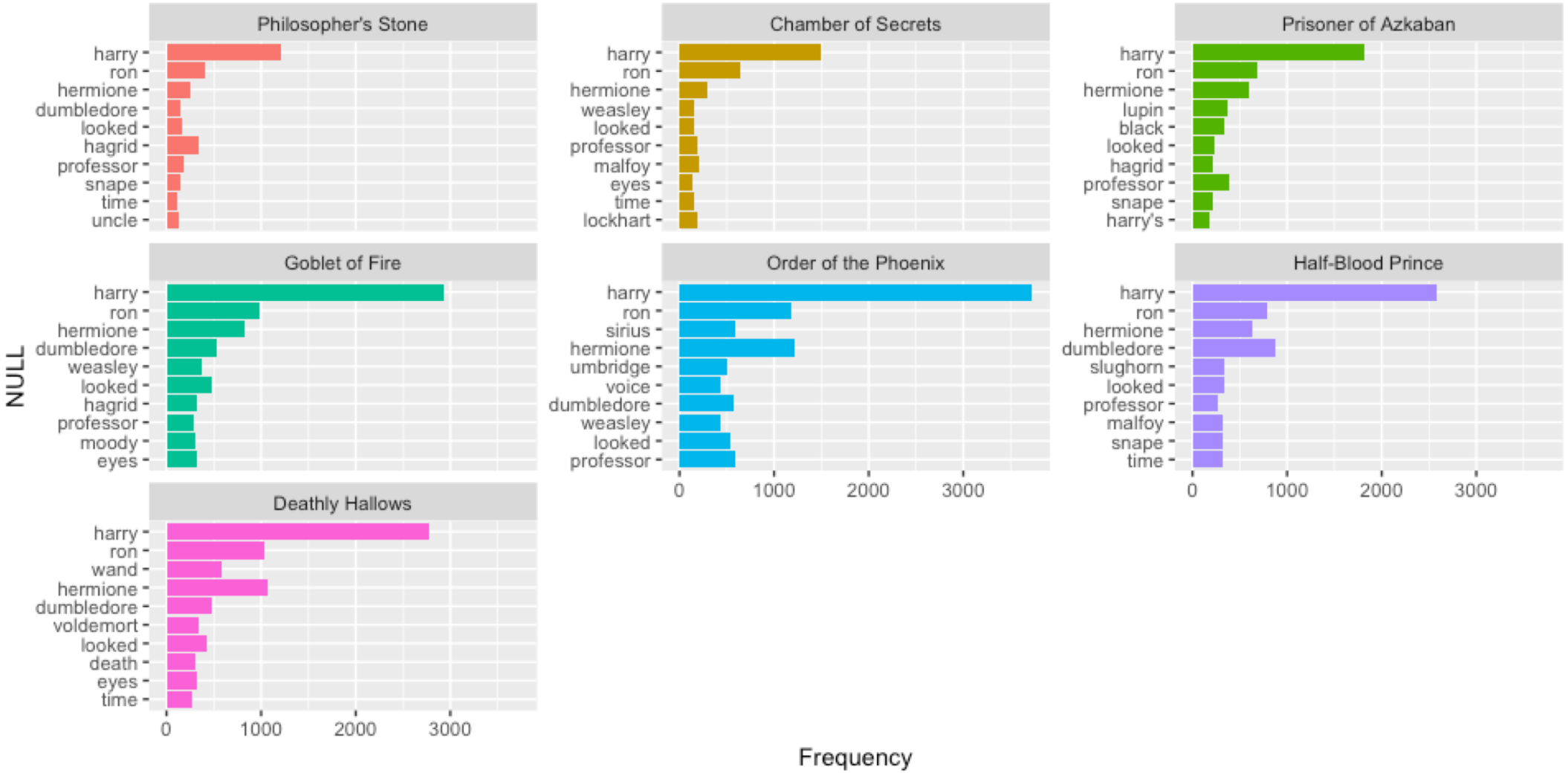
Element of the text

Word | n-gram | Sentence | Verse | Line | Paragraph

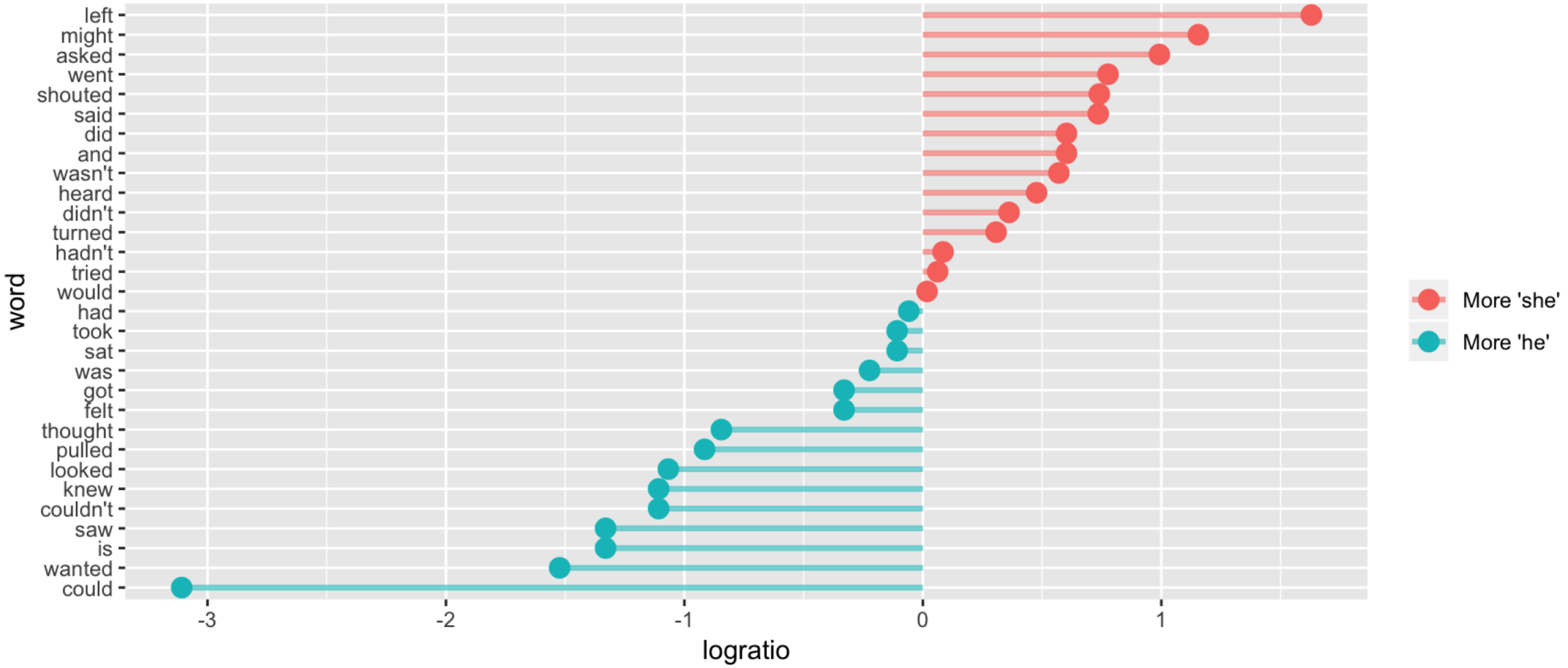
	chapter	word	book
1	1	the	Harry Potter and the Philosopher's Stone
2	1	boy	Harry Potter and the Philosopher's Stone
3	1	who	Harry Potter and the Philosopher's Stone
4	1	lived	Harry Potter and the Philosopher's Stone
5	1	mr	Harry Potter and the Philosopher's Stone
6	1	and	Harry Potter and the Philosopher's 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

	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
23	and the	82
24	from the	80

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
2	1	1	2	BOY	boy	PROPN	NNP	4
3	1	1	3	WHO	who	NOUN	WP	8
4	1	1	4	LIVED	live	VERB	VBD	12
5	1	1	5			SPACE	_SP	17
6	1	1	6	Mr.	mr.	PROPN	NNP	19
7	1	1	7	and	and	CCONJ	CC	23
8	1	1	8	Mrs.	mrs.	PROPN	NNP	27
9	1	1	9	Dursley	dursley	PROPN	NNP	32
10	1	1	10	,	,	PUNCT	,	39
11	1	1	11	of	of	ADP	IN	41
12	1	1	12	number	number	NOUN	NN	44
13	1	1	13	four	four	NUM	CD	51
14	1	1	14	,	,	PUNCT	,	55
15	1	1	15	Privet	privet	PROPN	NNP	57
16	1	1	16	Drive	drive	PROPN	NNP	64
17	1	1	17	,	,	PUNCT	,	69
18	1	1	18	were	be	VERB	VBD	71

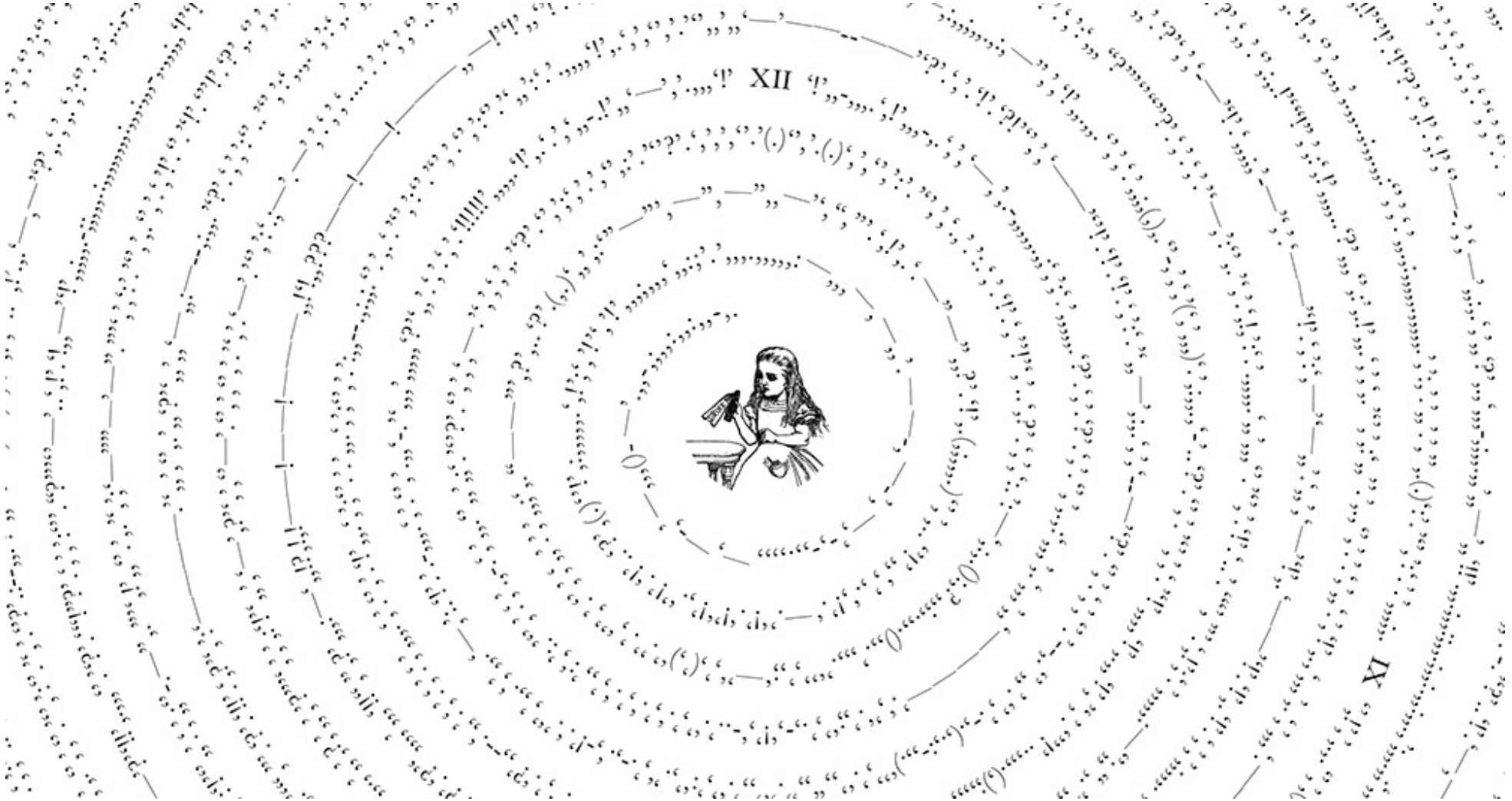
PART OF SPEECH FREQUENCY

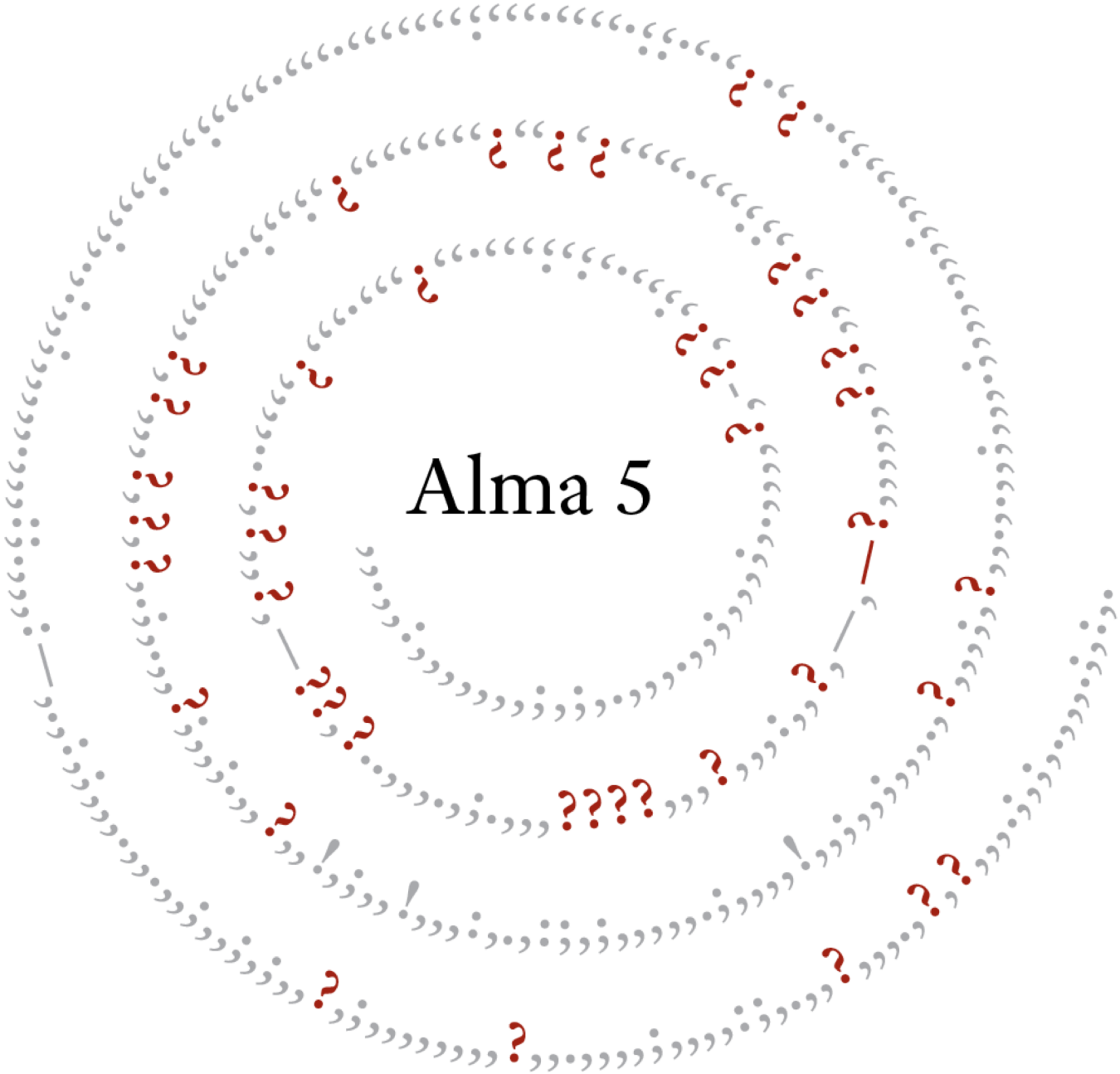
	lemma	n
1	be	3300
2	have	1298
3	say	925
4	do	721
5	get	460
6	would	409
7	go	406
8	look	381
9	could	303
10	see	300
11	know	293
12	will	260
13	think	232
14	come	204
15	tell	181
16	take	173

	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	hagrid	364
4	hermione	269
5	professor	181
6	snape	172
7	dumbledore	153
8	dudley	138
9	malfoy	122
10	neville	117
11	vernon	116
12	quirrell	111
13	uncle	111
14	mcgonagall	101
15	potter	96
16	gryffindor	85

ARTSY STUFF




Alma 5

SENTIMENT ANALYSIS

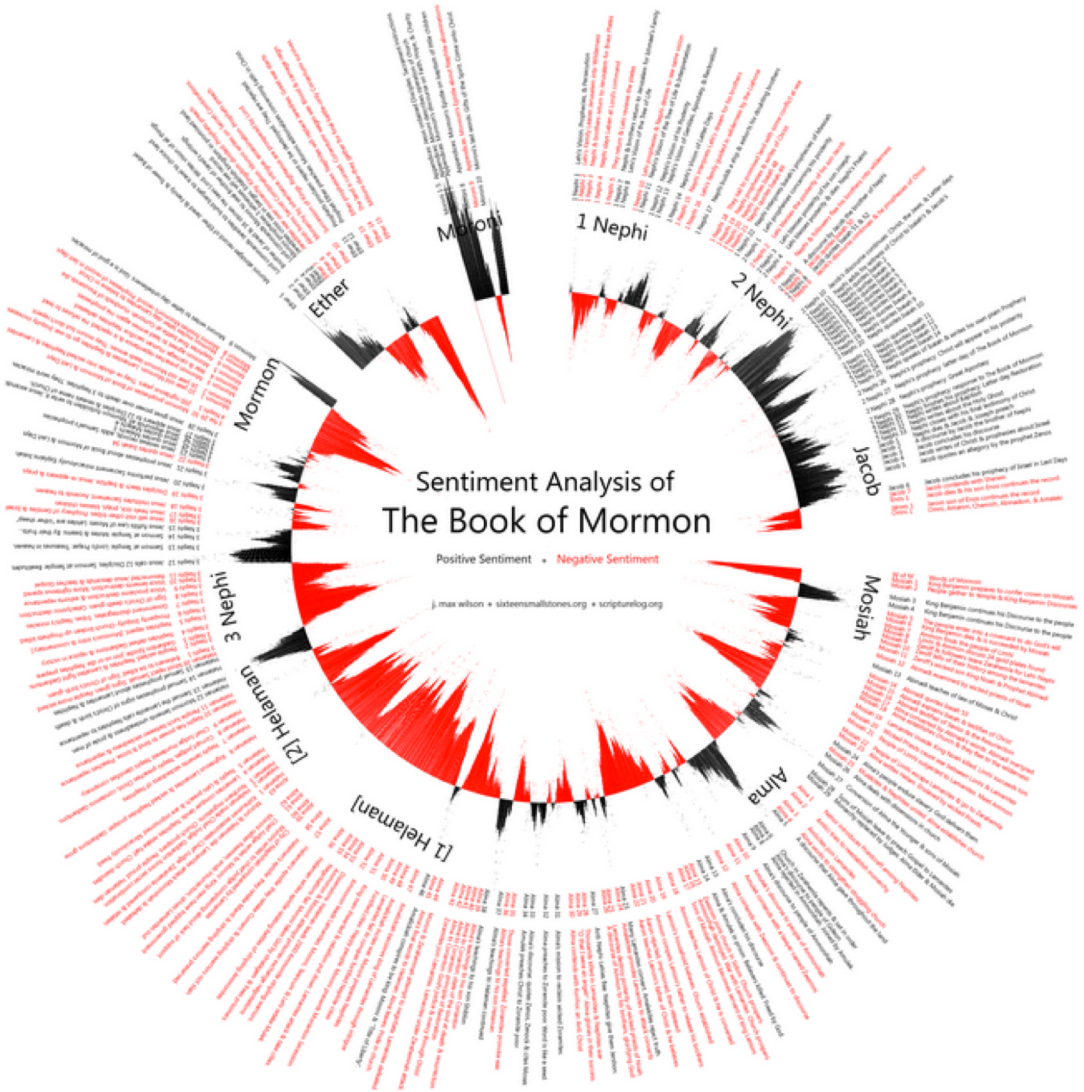
How positive or negative a text is

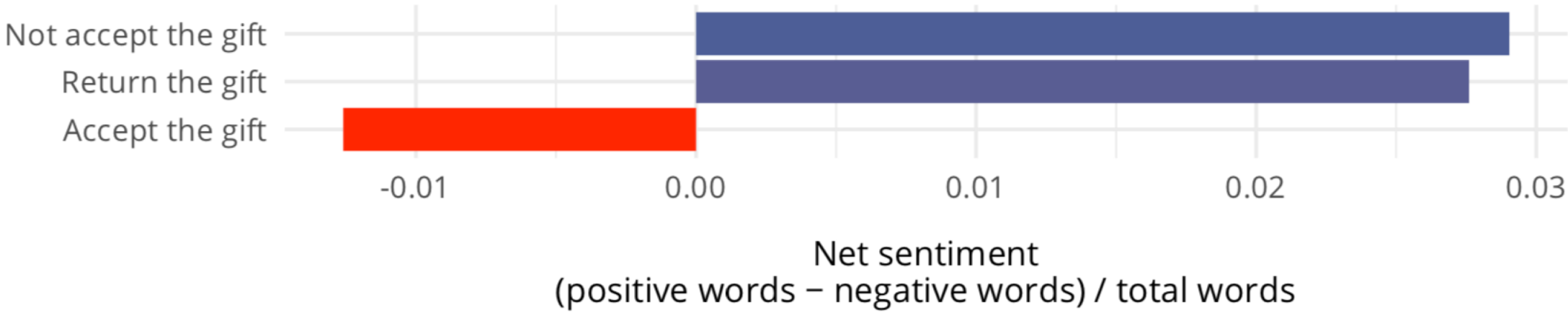
```
> get_sentiments("bing")
# A tibble: 6,788 x 2
   word        sentiment
   <chr>       <chr>
 1 2-faced     negative
 2 2-faces     negative
 3 a+          positive
 4 abnormal    negative
 5 abolish     negative
 6 abominable  negative
 7 abominably  negative
 8 abominate   negative
 9 abomination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




Not accept the gift
Return the gift
Accept the gift
-0.01
0.00
0.01
0.02
0.03
Net sentiment
(positive words − negative words) / total words

TF-IDF

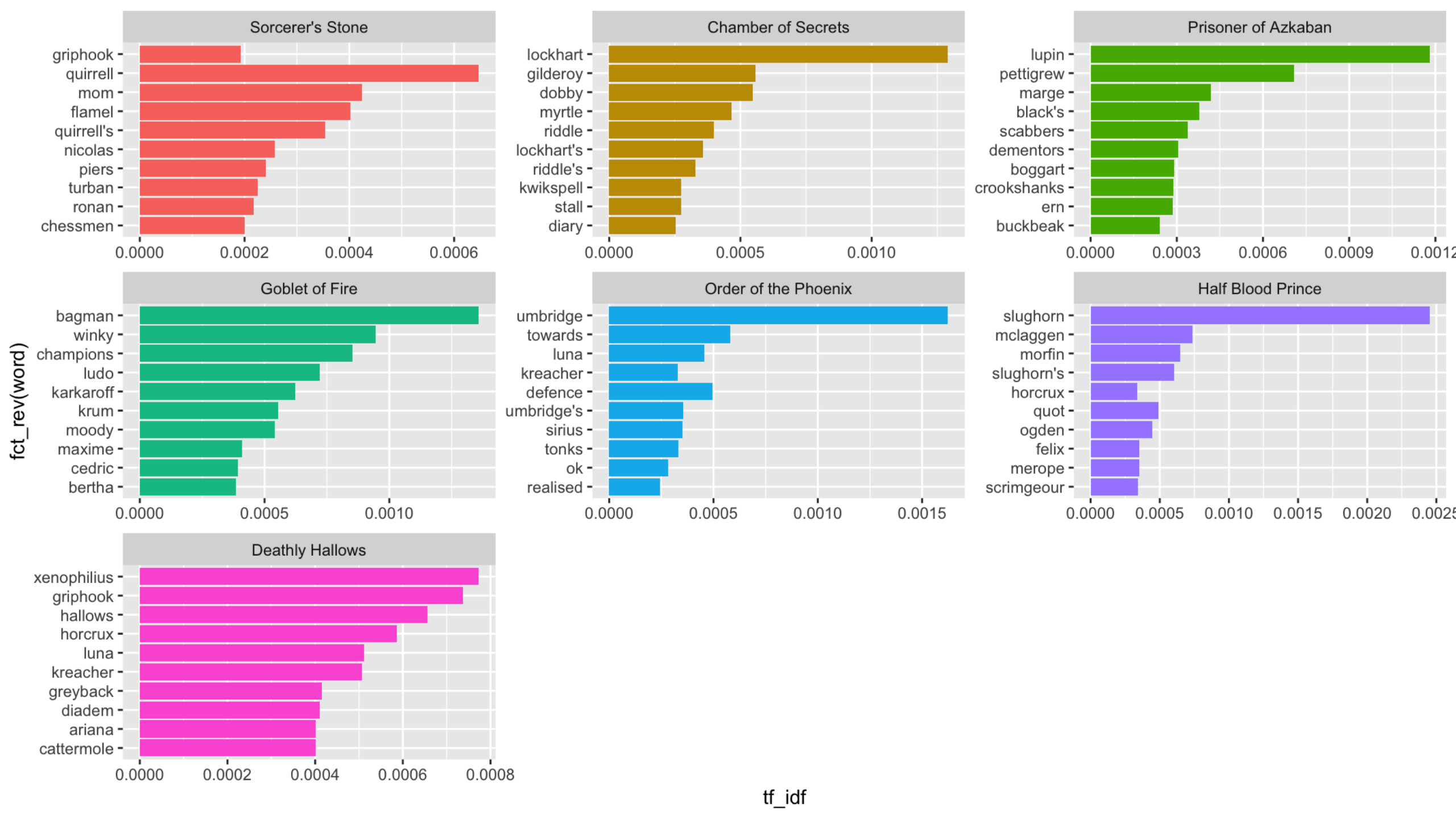
Term frequency-inverse document frequency

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

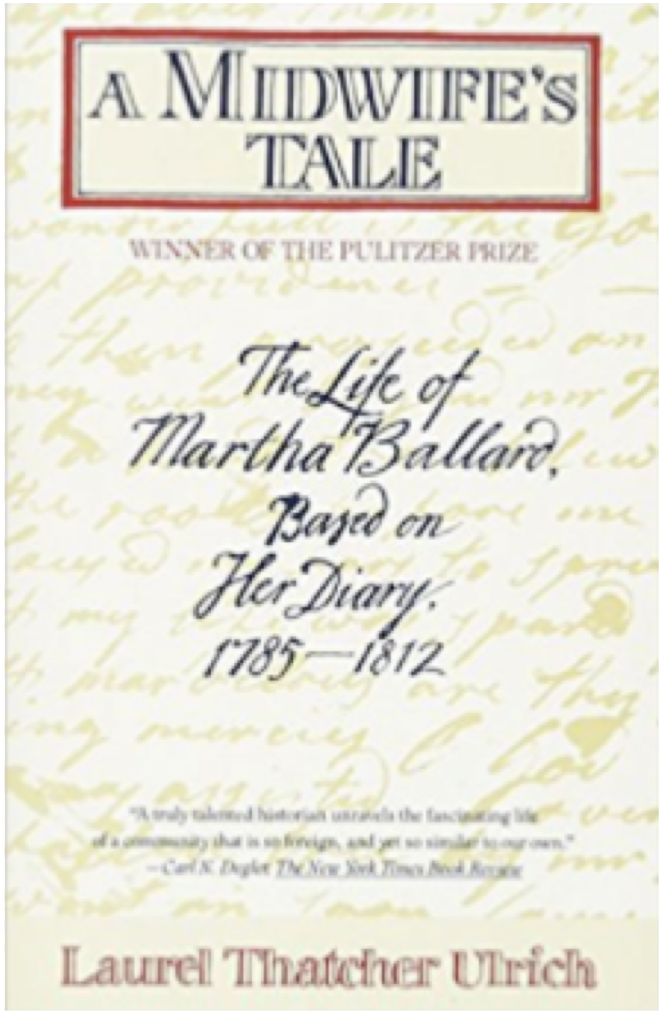
$$tf(\text{term}) = \frac{n_{\text{term}}}{n_{\text{terms in document}}}$$

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

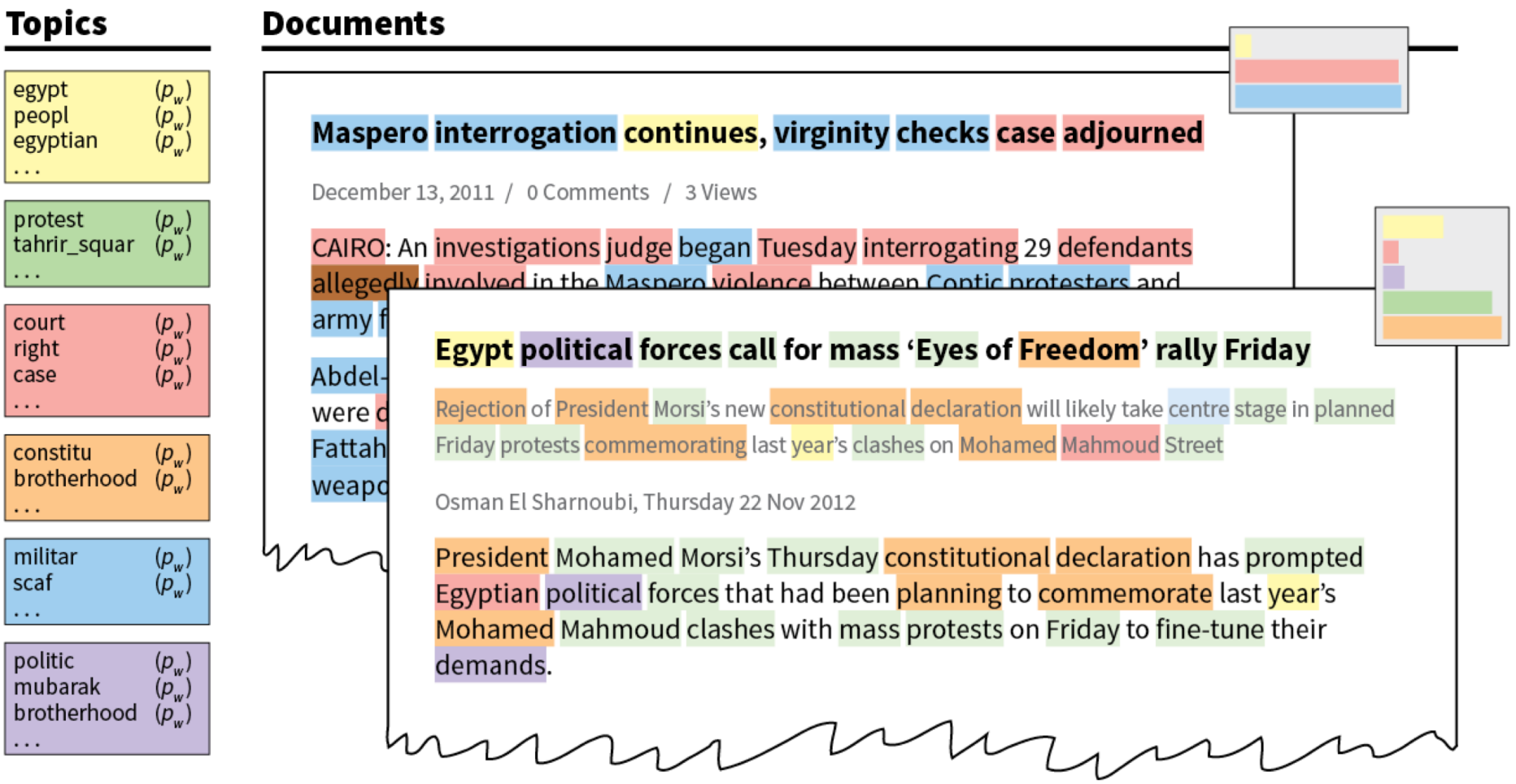
$$tf\text{-}idf(\text{term}) = tf(\text{term}) \times idf(\text{term})$$


Sorcerer's Stone
griphook
quirrell
mom
flamel
quirrell's
nicolas
piers
turban
ronan
chessmen
0.0000
0.0002
0.0004
0.0006
Chamber of Secrets
lockhart
gilderoy
dobby
myrtle
riddle
lockhart's
riddle's
kwikspell
stall
diary
0.0000
0.0005
0.0010
Prisoner of Azkaban
lupin
pettigrew
marge
black's
scabbers
dementors
boggart
crookshanks
ern
buckbeak
0.0000
0.0003
0.0006
0.0009
0.0012
Goblet of Fire
bagman
winky
champions
ludo
karkaroff
krum
moody
maxime
cedric
bertha
0.0000
0.0005
0.0010
Order of the Phoenix
umbridge
towards
luna
kreacher
defence
umbridge's
sirius
tonks
ok
realised
0.0000
0.0005
0.0010
0.0015
Half Blood Prince
slughorn
mclaggen
morfin
slughorn's
horcrux
quot
ogden
felix
merope
scrimgeour
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
Deathly Hallows
xenophilius
griphook
hallows
horcrux
luna
kreacher
greyback
diadem
ariana
cattermole
0.0000
0.0002
0.0004
0.0006
0.0008
fct_rev(word)
tf_idf

TOPIC MODELING



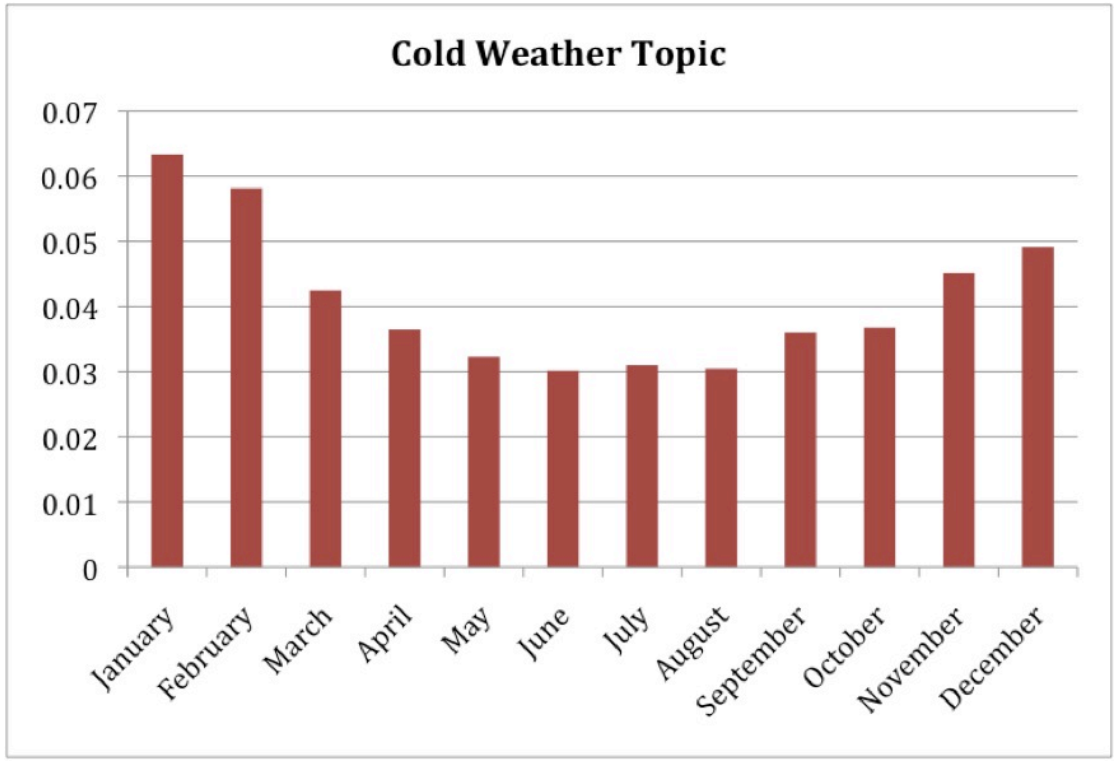
LATENT DIRICHLET ALLOCATION

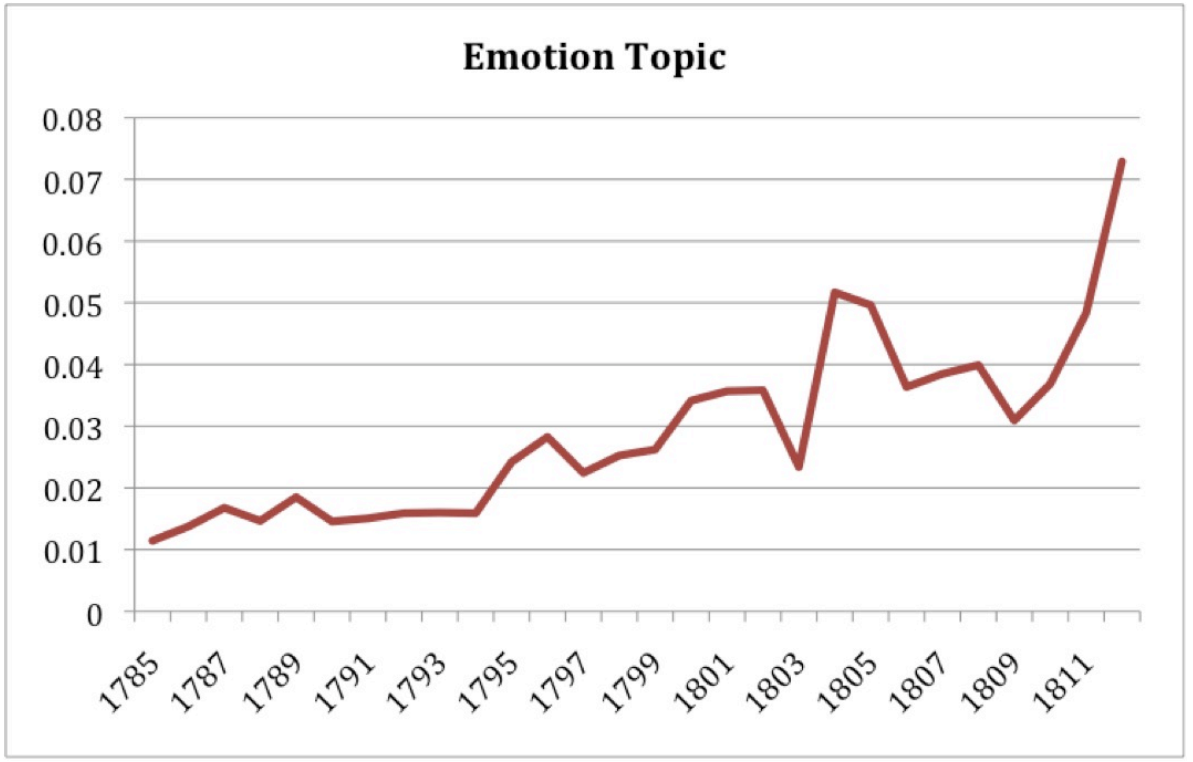


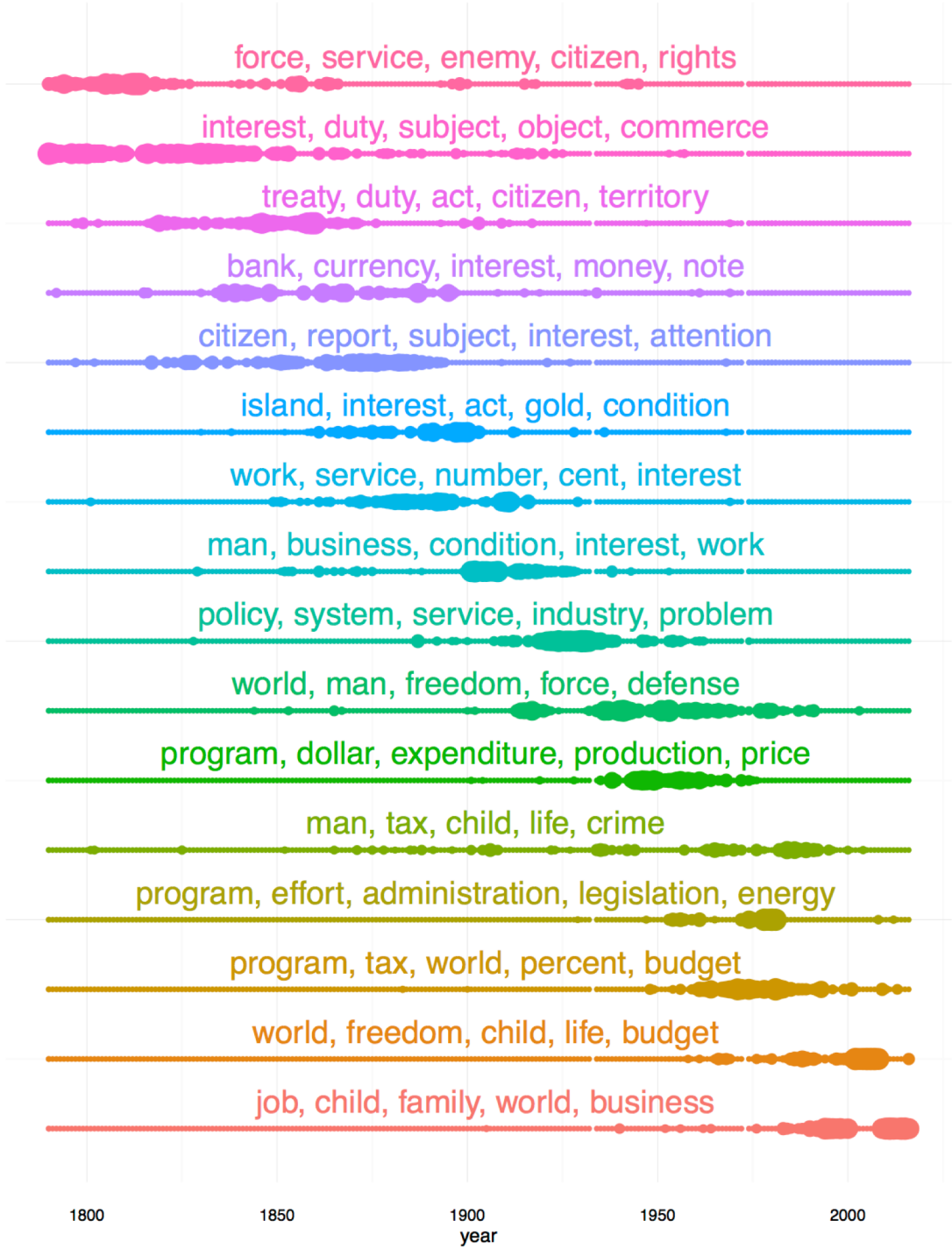
CLUSTERS OF RELATED WORDS

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 expired expird weak dead las past heard days drowned departed evinn
GARDENING	gardin sett worked clear beens corn warm planted matters cucumbers gatherd potatoes plants ou sowd door squash wed seeds
SHOPPING	lb made brot bot tea butter sugar carried oz chees pork candles wheat store pr beef spirit churnd flower
ILLNESS	unwell mr sick gave dr rainy easier care head neighbor feet relief made throat poorly takeing medisin ts stomach

TRACK TOPICS OVER TIME






force, service, enemy, citizen, rights
interest, duty, subject, object, commerce
treaty, duty, act, citizen, territory
bank, currency, interest, money, note
citizen, report, subject, interest, attention
island, interest, act, gold, condition
work, service, number, cent, interest
man, business, condition, interest, work
policy, system, service, industry, problem
world, man, freedom, force, defense
program, dollar, expenditure, production, price
man, tax, child, life, crime
program, effort, administration, legislation, energy
program, tax, world, percent, budget
world, freedom, child, life, budget
job, child, family, world, business
1800
1850
1900
1950
2000
year

FINGERPRINTING

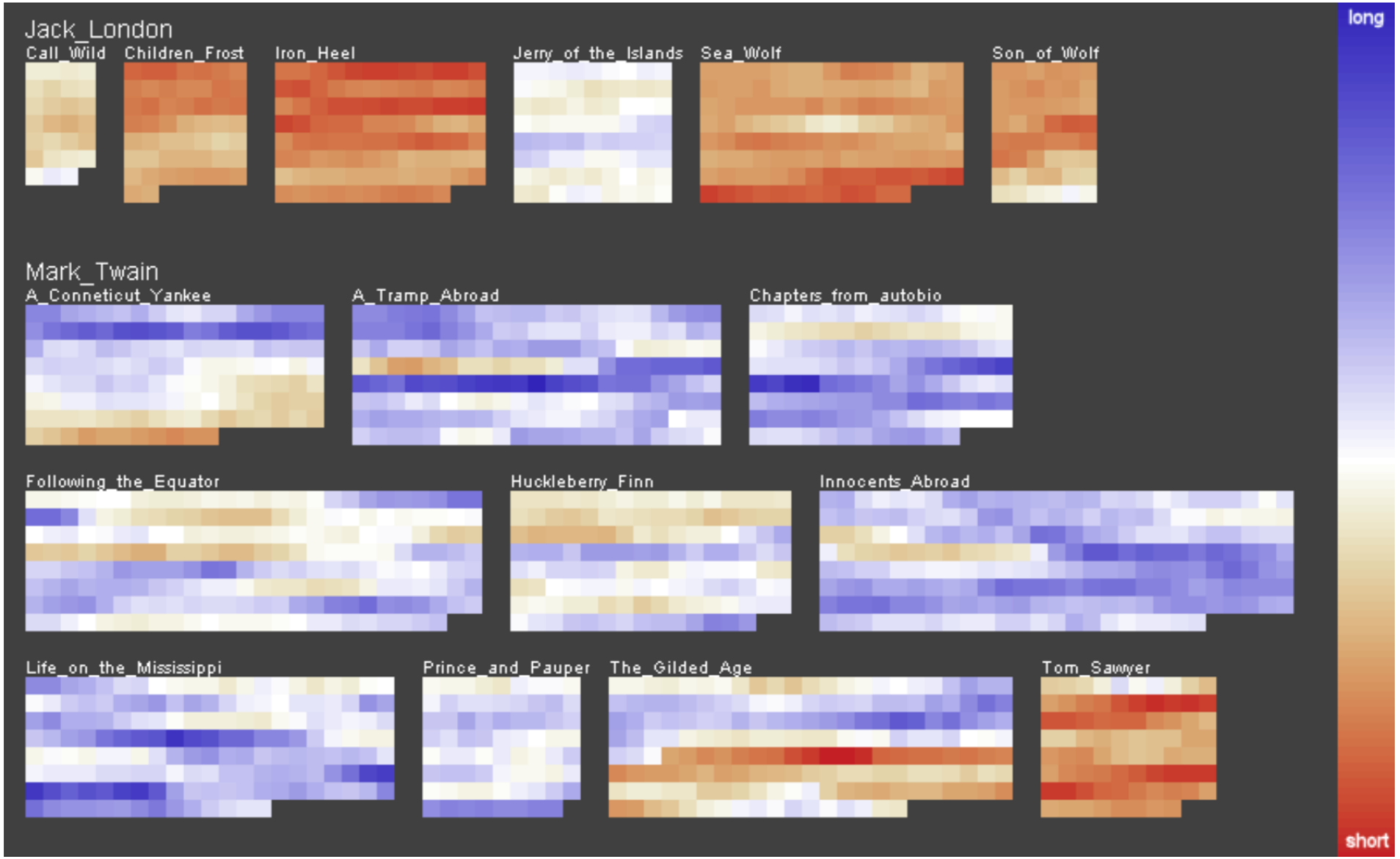
Analyze richness or uniqueness of document

Punctuation patterns, vocabulary choices, sentence length

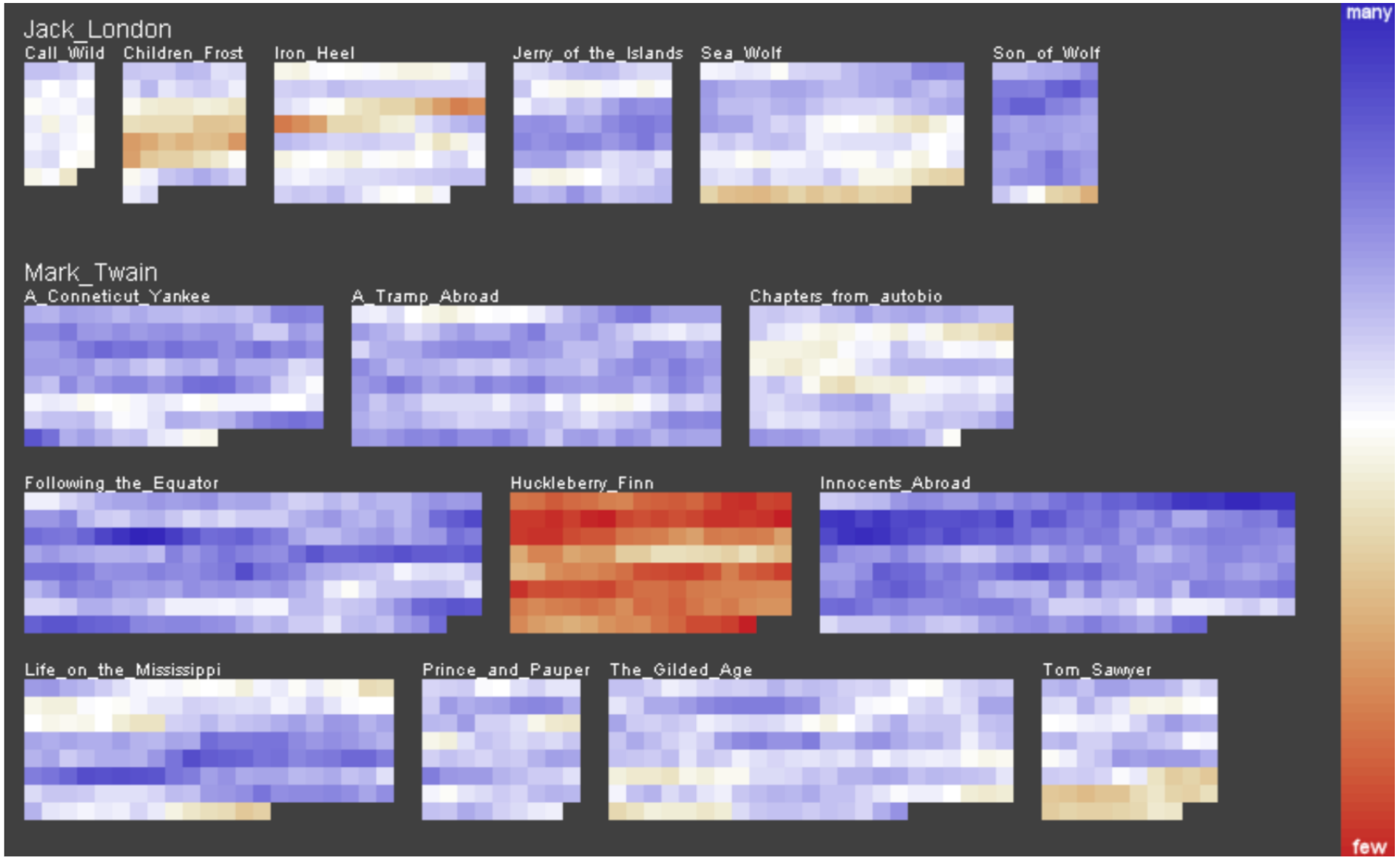
Hapax legomenon

	lemma	n
13	alicia	1
14	anne	1
15	anticheating	1
16	anythin	1
17	argus	1
18	arsenius	1

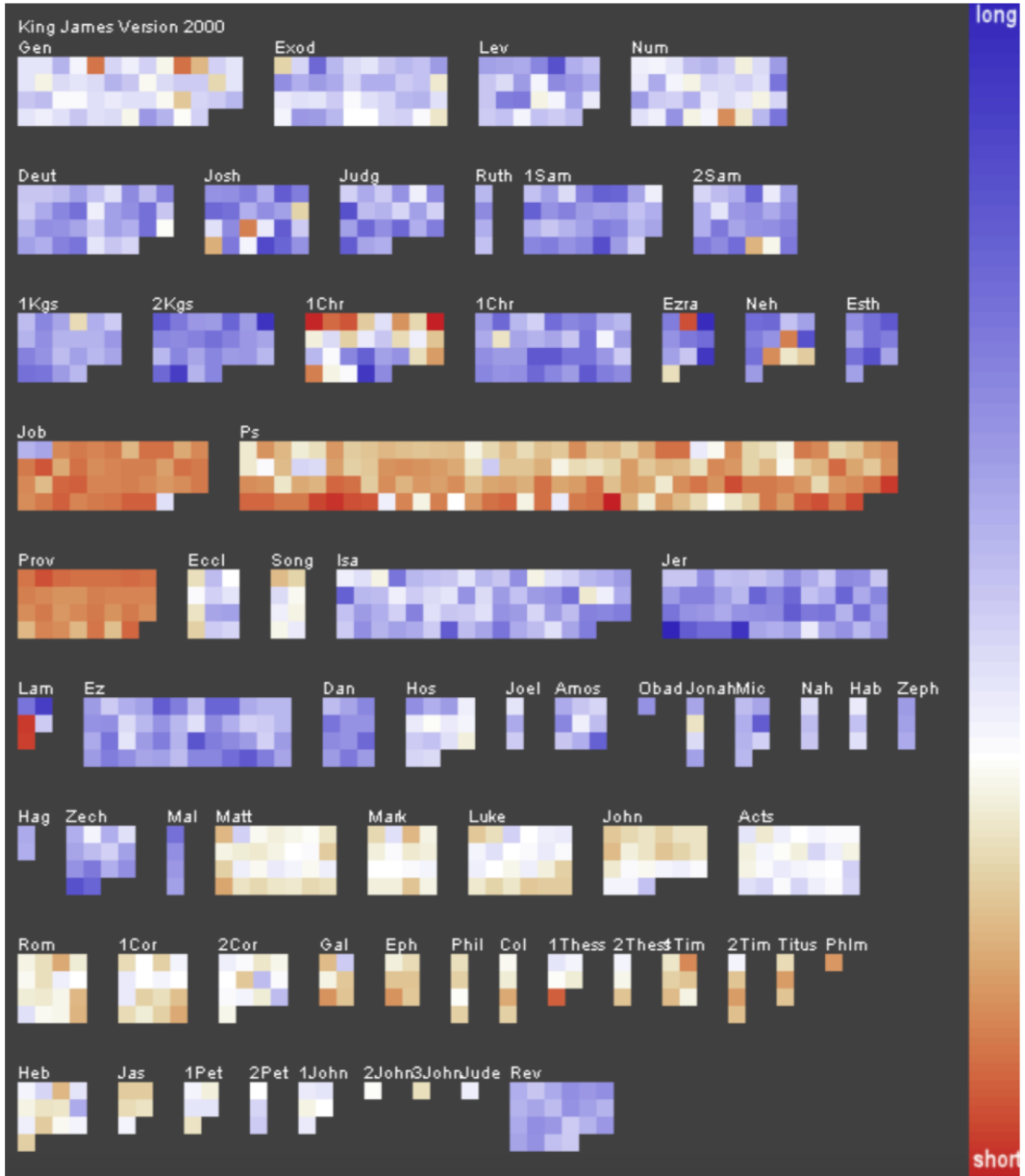
Sentence length



Hapax legomena



Verse length



VISUALIZING TEXT WITH R