

Are All These Trends from 'Booktok'?

Exploring Applications for Metadata Analysis in Public Library Collection Development

By Maddy Casey, MsLIS



Project GitHub page

<https://github.com/mc2398/metadata-analysis-vis/>

contact me:

mcase393@pratt.edu



Project write up (in my portfolio)

https://mc2398.github.io/graduate_portfolio/goodreads_article.html

Project Design

PROJECT INSPIRATION

- “The Goodreads Classics”
 - Digital Humanities project by Melanie Walsh and Maria Antoniak
- Interest In applying DH methods to public library work
- *The Library Book* by Susan Orlean
 - Can I use data to the “gold prospecting” books of today?

DATA SOURCE

- Explores metadata from books nominated for the Goodreads’ Readers Favorite Awards
- Awards ran from 2011-2024
- Nominated books are selected by Goodreads based on data analysis and published in the year of the awards
 - Must have a minimum average rating of 3.5 at the time of the awards
- Primarily analysed “genre” tags (subject terms) from the nominated books

RESEARCH DESIGN

Goal:

Ease of Replication in Public Library Setting

Research Questions:

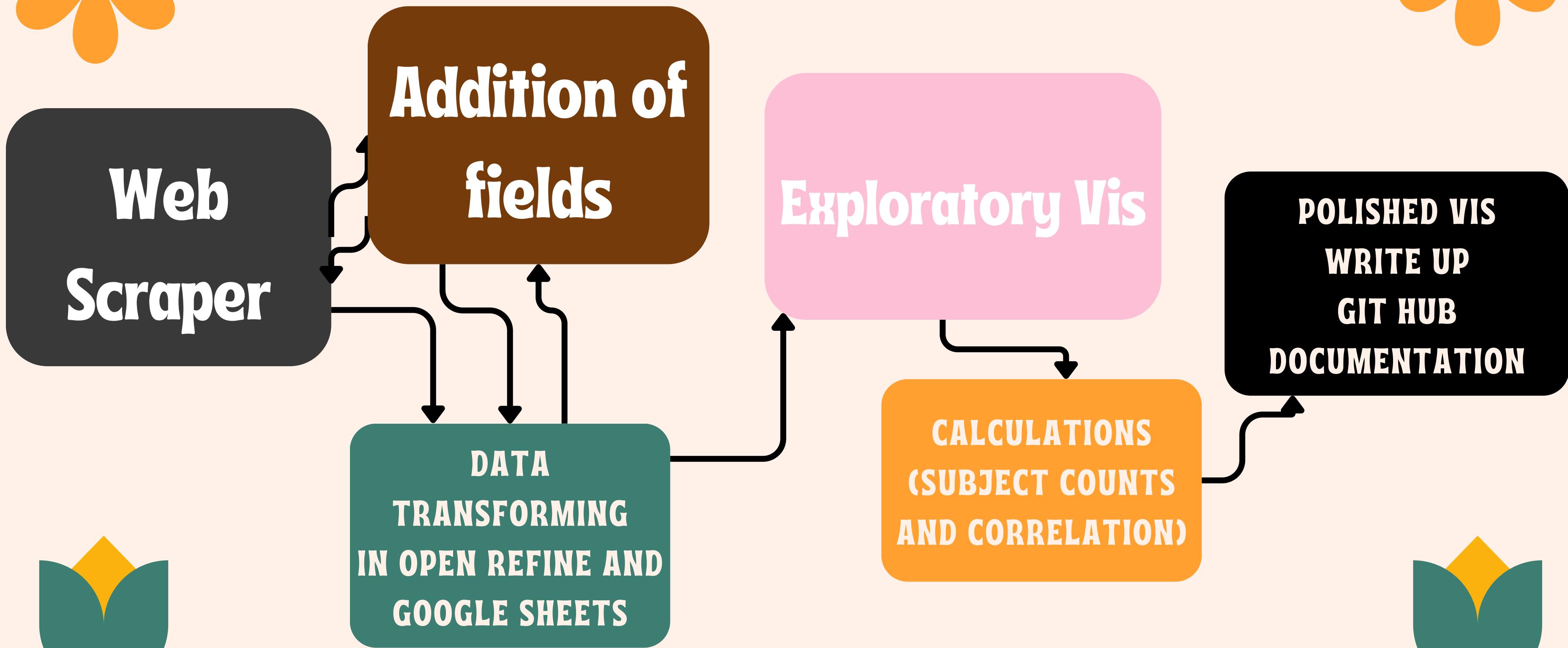
What subject-matter trends existed in the books that won the readers' favorite awards across the studied years?

To what degree are the trends in popular topics persistent across the studied time period?

Pipeline



Pipeline



Goodreads Metadata	Library Metadata
Inclusion in Goodreads Favorite Awards	Annual Circulation Count
Title	Title
Goodreads id	unique id, determined by the institution
Author	Creator
Genre	Subject Terms
Description	Abstract, as applicable
Category of Awards	Call Number

Findings

Reader's Favorite Contextual Data

year	number of categories	original number of books	deduplicated number of books
2024	15	300	272
2023	15	299	284
2022	17	340	320
2021	17	340	323
2020	20	399	381
2019	20	400	384
2018	21	420	385
2017	20	400	390
2016	20	400	393
2015	20	400	390
2014	20	399	392
2013	20	400	391
2012	20	399	379
2011	22	440	401

Repeat Contributors

28% of the authors/contributors whose books were nominated received multiple nominations

Stephen King- 19 unique books

Rick Riordan- 17 unique books

Sarah J. Maas- 15 unique books

Colleen Hoover- 13 unique books

Alice Oseman- 6 unique books

Elin Hilderbrand- 6 unique books

Kristin Hannah- 6 unique books

Margaret Atwood- 6 unique books

Liane Moriarty- 5 unique books

Emily Henry- 5 unique books

Gwyneth Paltrow- 5 books

Genre Frequency Calculations

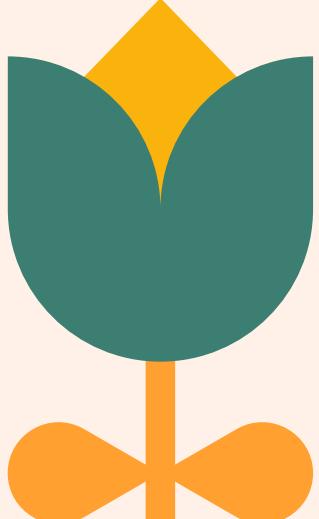
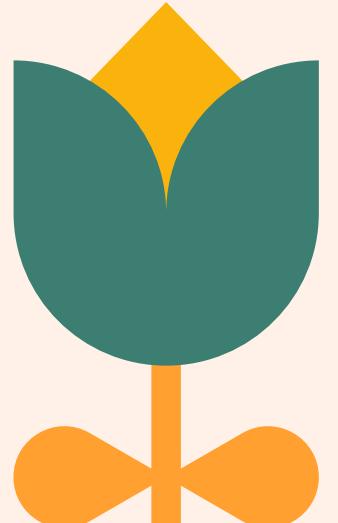
=REGEXMATCH (C2, "^Fiction,l, Fiction,l, Fiction\$")

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
id	title	genres	fiction_subject	audiobook_subject	nonfiction_subject	fantasy_subject	romance_subject	contemporary_subject	young_adult_subject	adult_subject	mystery_subject	biography_subject	memoir_subject	historical_subject
198902277	The Wedding Pe	Fiction, Romance	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
208931300	Intermezzo	Fiction, Contemp	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
133938826	Welcome to the I	Contemporary, B	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
195430687	Blue Sisters	Fiction, Contemp	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
208516656	Here One Mome	Fiction, Audiobo	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
200484931	Swan Song	Fiction, Romanc	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
198688460	Good Material	Fiction, Romanc	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
200987323	I Hope This Find	Fiction, Romanc	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
199534613	Margo's Got Mor	Fiction, Audiobo	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
139400713	Marty!	Fiction, Literary F	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
174147294	Wandering Stars	Fiction, Historica	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
62929342	Real Americans	Fiction, Historica	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
200028726	Sandwich	Fiction, Audiobo	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
204811915	Tell Me Everythir	Fiction, Literary F	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
197798168	All Fours	Fiction, Audiobo	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
205478762	Playground	Fiction, Science	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
199344873	Same As It Ever	Fiction, Audiobo	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
199793431	The Mighty Red	Fiction, Historica	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
176443453	Greta & Valdin	Fiction, Queer, L	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
220572240	Colored Televisic	Fiction, Literary F	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
127305853	The Women	Historical Fiction	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE
173754979	James	Fiction, Historica	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
112975658	The Frozen Rive	Historical Fiction	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE
207909446	The Booklover's	Historical Fiction	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
199798217	The Lion Womer	Historical Fiction	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
203164368	By Any Other Na	Historical Fiction	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
181110045	The Briar Club	Historical Fiction	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE
199223201	Husbands & Lov	Historical Fiction	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
217362723	The Seventh Vei	Historical Fiction	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE
127306192	Anita de Monte	L Fiction, Historica	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE
202468422	There Are Rivers	Historical Fiction	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE

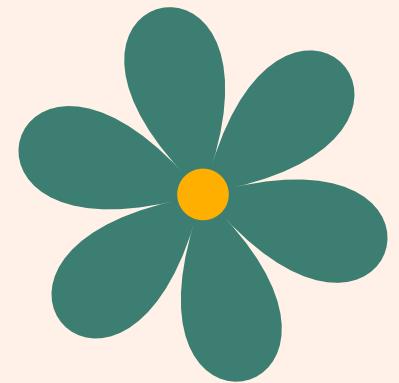
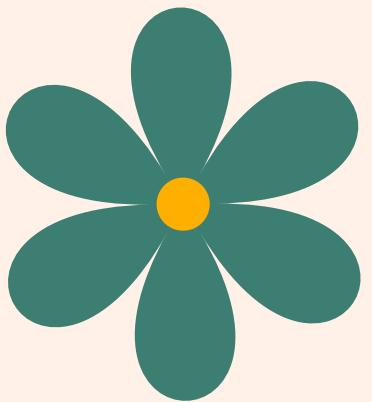
Genre Frequency Calculations

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	title	year	paranormal_subj	biography_memoir	childrens_subject	literary_fiction_story	poetry_subject	crime_subject	queer_subject	graphic_novels	contemporary_romance	magic_subject
246	181346634	There's Always Us	2024	0	0	0	0	1	0	0	0	0	0
247	203708995	The Friday Afternoon	2024	0	1	0	0	0	0	0	0	0	0
248	196585876	An Unfinished Love	2024	0	1	0	0	0	0	0	0	0	0
249	176443920	The Wives: A Memoir	2024	0	1	0	0	0	0	0	0	0	0
250	182095845	Here After	2024	0	1	0	0	0	0	0	0	0	0
251	126918693	My Side of the River	2024	0	1	0	0	0	0	0	0	0	0
252	127305940	The Manicurist's Secret	2024	0	1	0	0	0	0	0	0	0	0
253	200100950	Wanted: Toddler	2024	0	1	0	0	0	0	0	0	0	0
254	201751300	The Bookshop: A Novel	2024	0	0	0	0	0	0	0	0	0	0
255	203956674	The Small and the Mighty	2024	0	0	0	0	0	0	0	0	0	0
256	195608683	The Demon of Us	2024	0	0	0	0	0	0	0	0	0	0
257	208580608	The Elements of Style	2024	0	0	0	0	0	0	0	0	0	0
258	207567691	American Scary: A Novel	2024	0	0	0	0	0	0	0	0	0	0
259	195791072	Women in the Woods	2024	0	0	0	0	0	0	0	0	0	0
260	204927599	Nexus: A Brief History	2024	0	0	0	0	0	0	0	0	0	0
261	201626940	Ask Not: The Kennedy Years	2024	0	0	0	0	0	0	0	0	0	0
262	199798785	Challenger: A True Story	2024	0	0	0	0	0	0	0	0	0	0
263	62315849	The Lost Tomb: A Novel	2024	0	0	0	0	0	1	0	0	0	0
264	198563695	When Women Rule: A Novel	2024	0	0	0	0	0	0	0	0	0	0
265	177188695	Not Your China Girl	2024	0	0	0	0	0	0	0	0	0	0
266	145624993	Madness: Race and Survival in the Age of AIDS	2024	0	0	0	0	0	0	0	0	0	0
267	199393033	By the Fire We Cried: A Novel	2024	0	0	0	0	0	0	0	0	0	0
268	198493808	The Situation Room: A Novel	2024	0	0	0	0	0	0	0	0	0	0
269	191746386	The Wide Wide World: A Novel	2024	0	0	0	0	0	0	0	0	0	0
270	199797834	When the Sea Came In: A Novel	2024	0	0	0	0	0	0	0	0	0	0
271	203767605	The Siege: A Six-Year Story	2024	0	0	0	0	0	0	0	0	0	0
272	145624514	Everyone Who Is: A Novel	2024	0	0	0	0	0	0	0	0	0	0
273	210246743	Eden Undone: A Novel	2024	0	0	0	0	0	1	0	0	0	0
274				17	15	0	35	2	17	23	1	21	
275													
276													

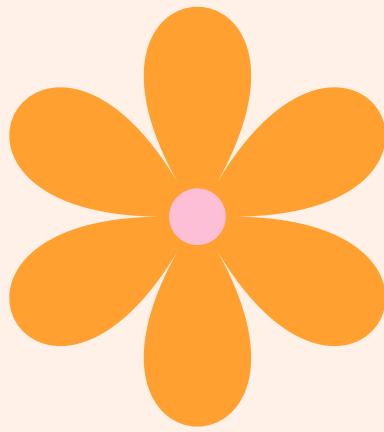
Convert to table



Genre Frequency Calculations



	A	B	C	D	E	F	G	H
1	Year	Subject	Book Count	Year Book Total	Percent_True_F	Percent_True	Percent-Rounde	Percent of Books
2		2024 fiction	184	272	67.64705882	67.64705882	67.65	67.65
3		2024 audiobook	215	272	79.04411765	79.04411765	79.05	79.05
4		2024 nonfiction	61	272	22.42647059	22.42647059	22.43	22.43
5		2024 fantasy	89	272	32.72058824	32.72058824	32.73	32.73
6		2024 romance	93	272	34.19117647	34.19117647	34.2	34.2
7		2024 contemporary	67	272	24.63235294	24.63235294	24.64	24.64
8		2024 young adult	41	272	15.07352941	15.07352941	15.08	15.08
9		2024 adult	67	272	24.63235294	24.63235294	24.64	24.64
10		2024 mystery	70	272	25.73529412	25.73529412	25.74	25.74
11		2024 biography	30	272	11.02941176	11.02941176	11.03	11.03
12		2024 memoir	28	272	10.29411765	10.29411765	10.3	10.3
13		2024 historical fiction	49	272	18.01470588	18.01470588	18.02	18.02
14		2024 thriller	53	272	19.48529412	19.48529412	19.49	19.49
15		2024 historical	39	272	14.33823529	14.33823529	14.34	14.34
16		2024 science fiction	24	272	8.823529412	8.823529412	8.83	8.83
17		2024 history	31	272	11.39705882	11.39705882	11.4	11.4
18		2024 humor	15	272	5.514705882	5.514705882	5.52	5.52
19		2024 horror	37	272	13.60294118	13.60294118	13.61	13.61
20		2024 lgbt	27	272	9.926470588	9.926470588	9.93	9.93
21		2024 mystery thriller	39	272	14.33823529	14.33823529	14.34	14.34
22		2024 paranormal	17	272	6.25	6.25	6.25	6.25
23		2024 biography memo	15	272	5.514705882	5.514705882	5.52	5.52
24		2024 childrens	0	272	0	0	0	0
25		2024 literary fiction	35	272	12.86764706	12.86764706	12.87	12.87
26		2024 poetry	2	272	0.7352941176	0.7352941176	0.74	0.74
27		2024 crime	17	272	6.25	6.25	6.25	6.25
28		2024 queer	23	272	8.455882353	8.455882353	8.46	8.46
29		2024 graphic novels	1	272	0.3676470588	0.3676470588	0.37	0.37
30		2024 contemporary ro	21	272	7.720588235	7.720588235	7.73	7.73
31		2024 magic	9	272	3.308823529	3.308823529	3.31	3.31
32		2024 middle grade	0	272	0	0	0	0



Calculating Correlations

A	B
lgbt_subject	romance_subject
0	1
0	1
0	0
1	0
0	0
0	1
0	1
0	1
0	0
1	0
0	0
0	0
0	0
1	0
0	0
0	0
0	0
1	1
0	0
0	0
0	0
0	0
0	0
0	0
0	1
0	0
0	1
0	1
0	0
0	0

```
▶ lgbt_romance_csv_file_path = '/content/correlation_comparisons_lgbt_x_romance.csv'

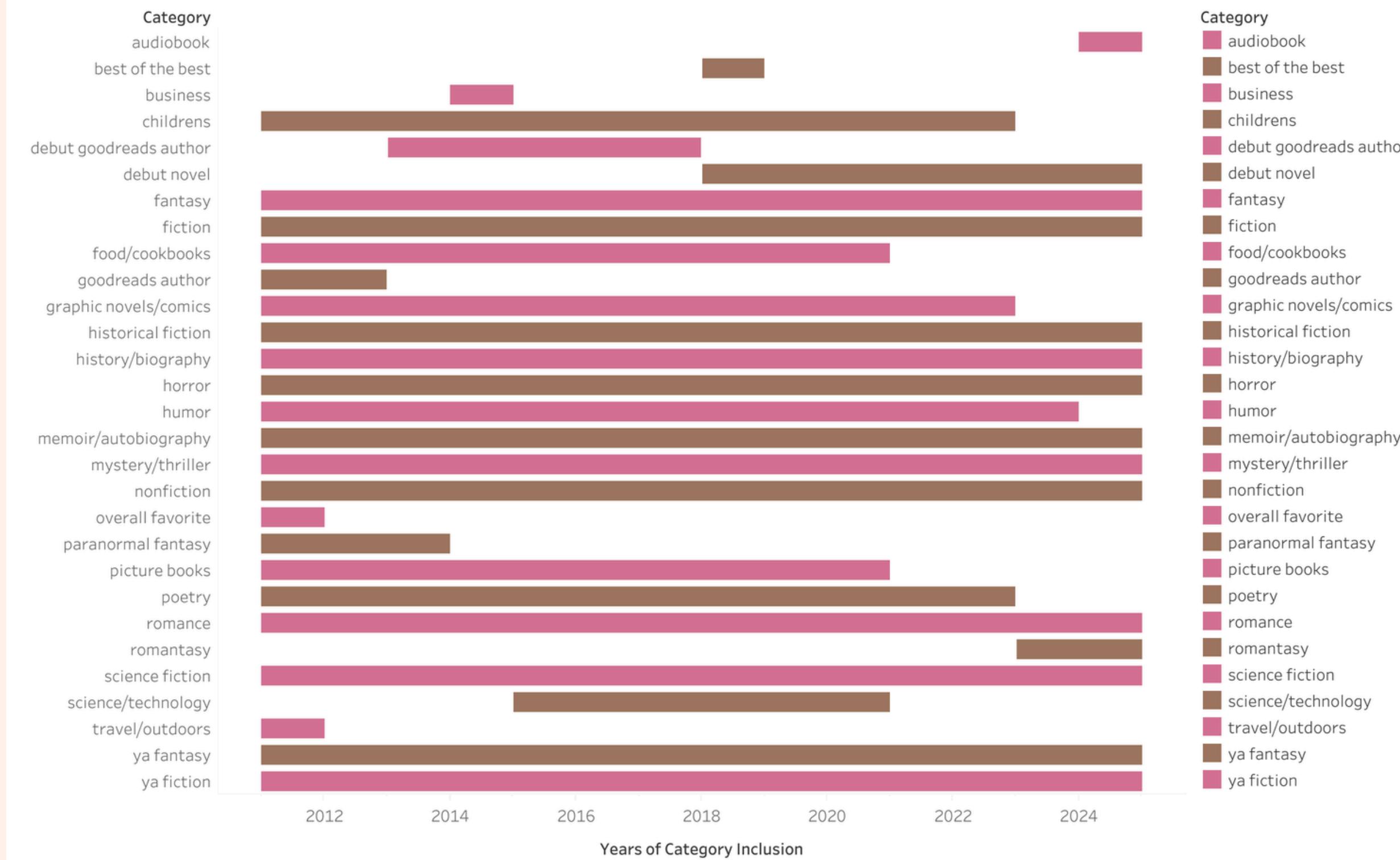
[ ] lgbt_romance_df = pd.read_csv(lgbt_romance_csv_file_path)

[ ] correlation, p_val = p_value_calc(lgbt_romance_df, 'lgbt_subject', 'romance_subject')
print(f'{correlation}')
print(f'{p_val}')

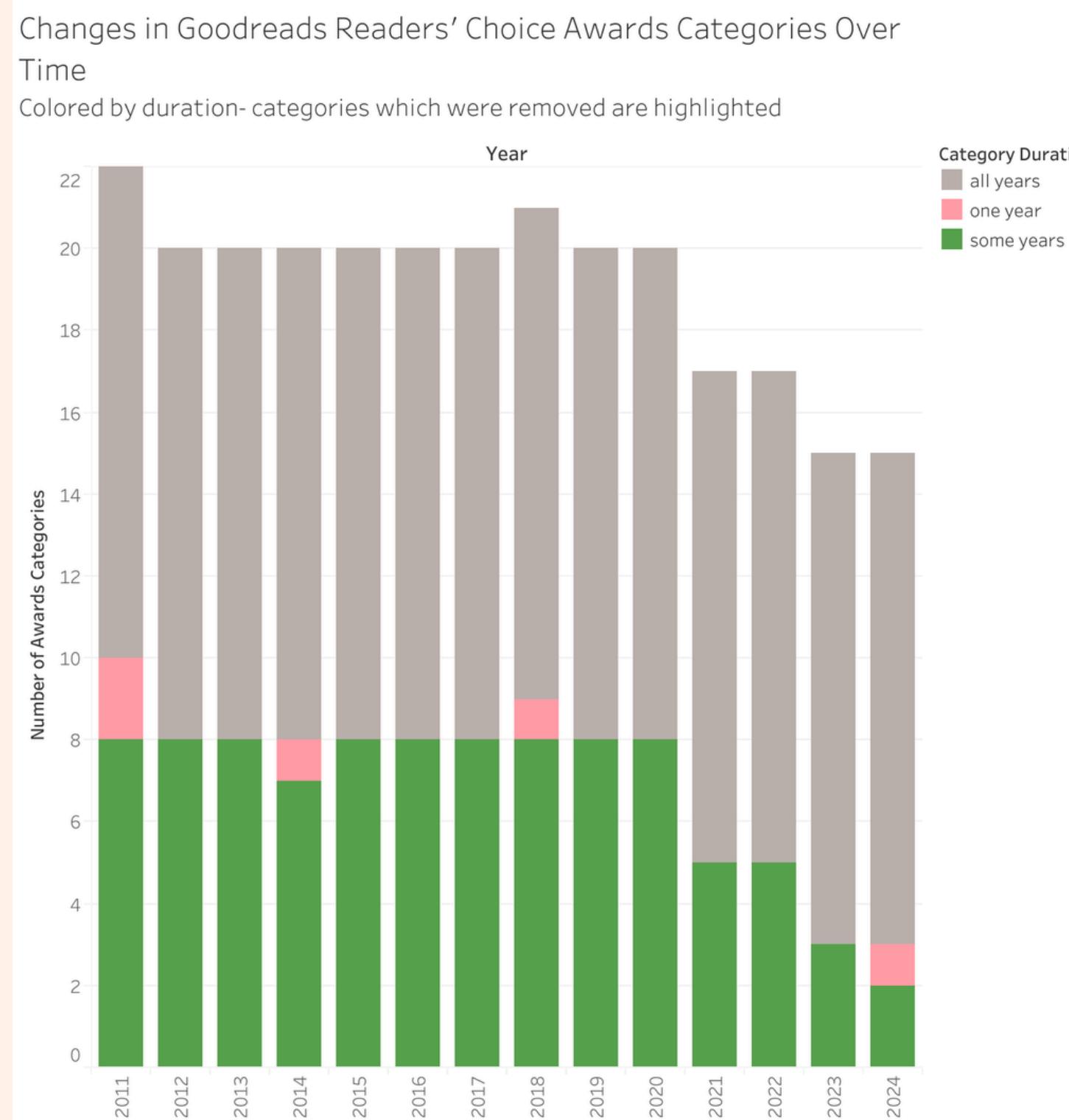
→ 0.0615544229845321
    1.12063878632218e-05
```

Categories over Time

Goodreads Favorite Awards Categories, 2011-2024



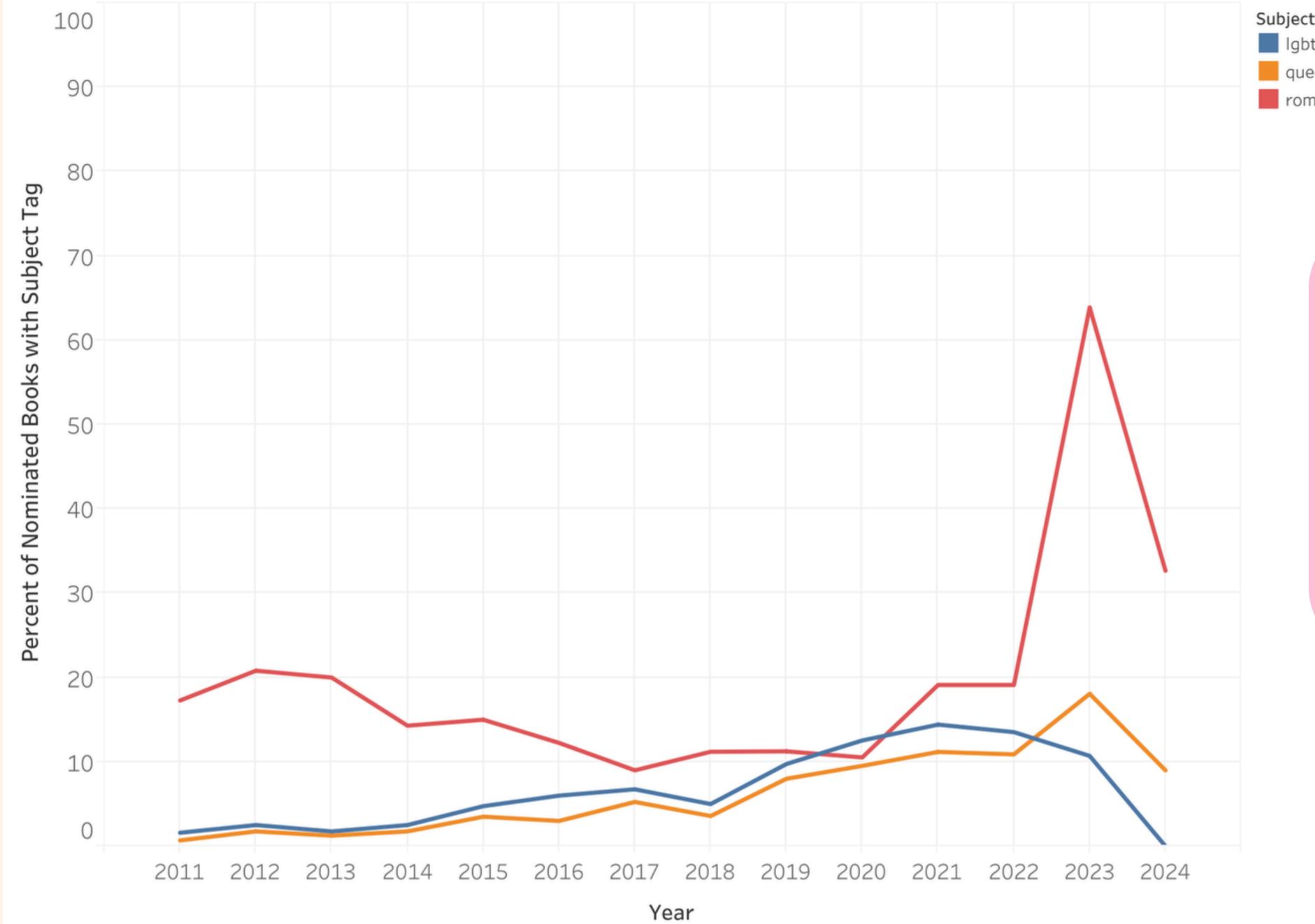
Categories over time, presented another way



**Link to “Booktok”
(user testing feedback)**

Draft Visualization

Topics of Popular Goodreads Books- 2011-2024

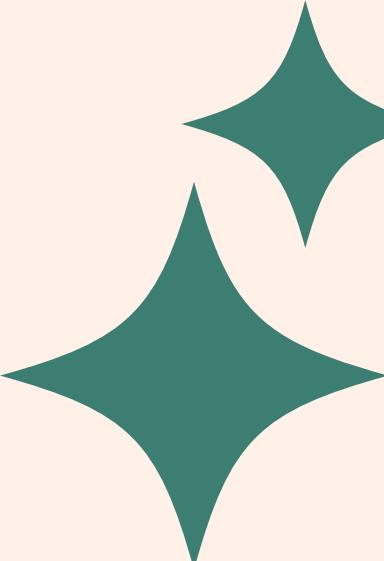


In early-stage user testing, when asked to interpret the real-world circumstances that may be informing the visualized trend, one user immediately replied “Booktok”

*Note- this visualization contains a data error which has since been rectified

Findings **“Vignettes”**

My Portfolio



Explaining Correlation

correlation (pearson's R): strength + direction of a relationship between values. Range of -1 to 1

p-value: Less than 0.05 typically significant

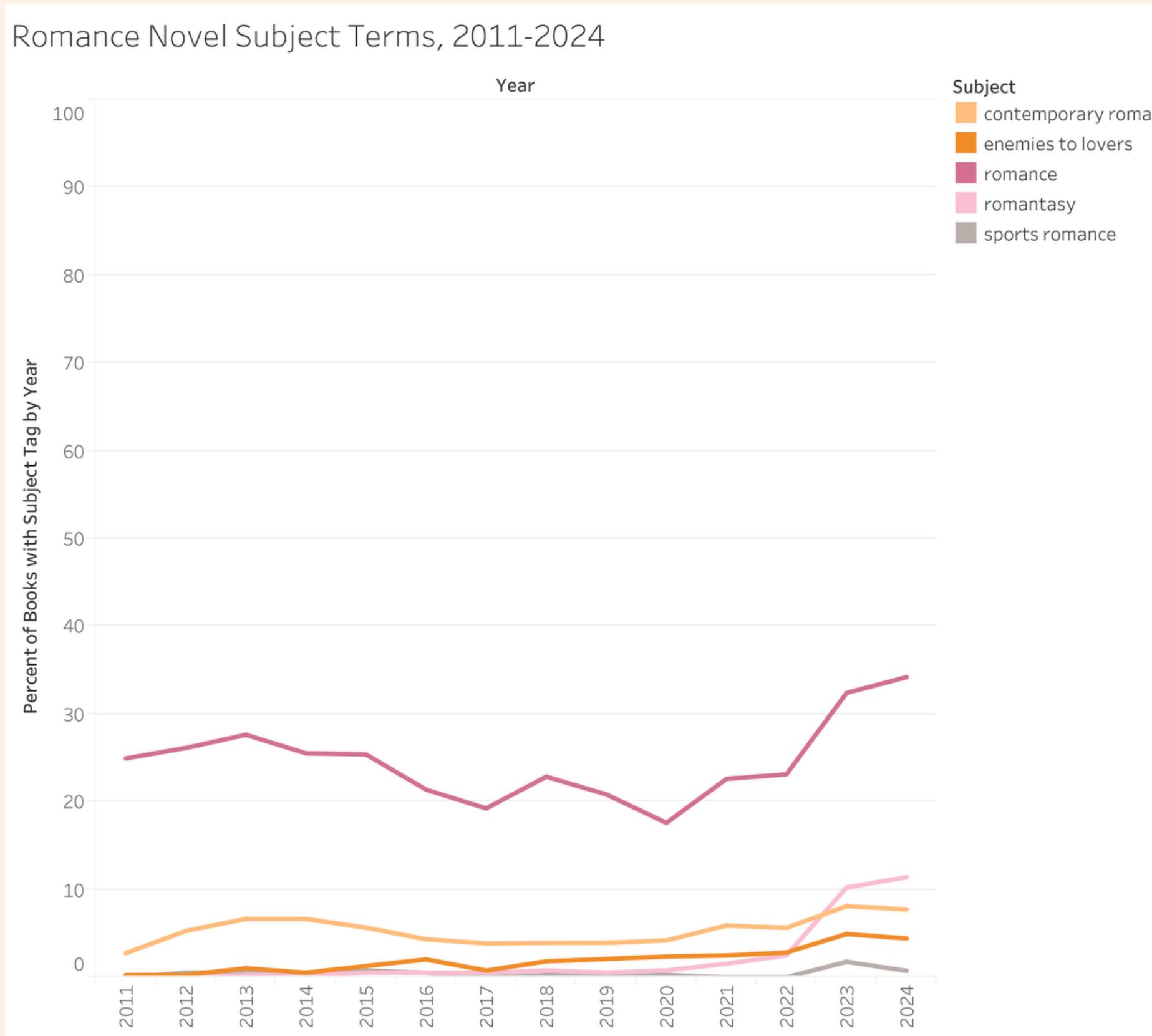
Fields	Correlation	P-value
lgbt, queer	0.8166515870289031	0.0
war, cookbooks	-0.038032679821485824	0.006679914902734242
cooking category, cooking	0.8941807186296845	0.0

may want total correlation here

Slight negative relationship

Can explore relationships between subjects and categories

Sample Findings

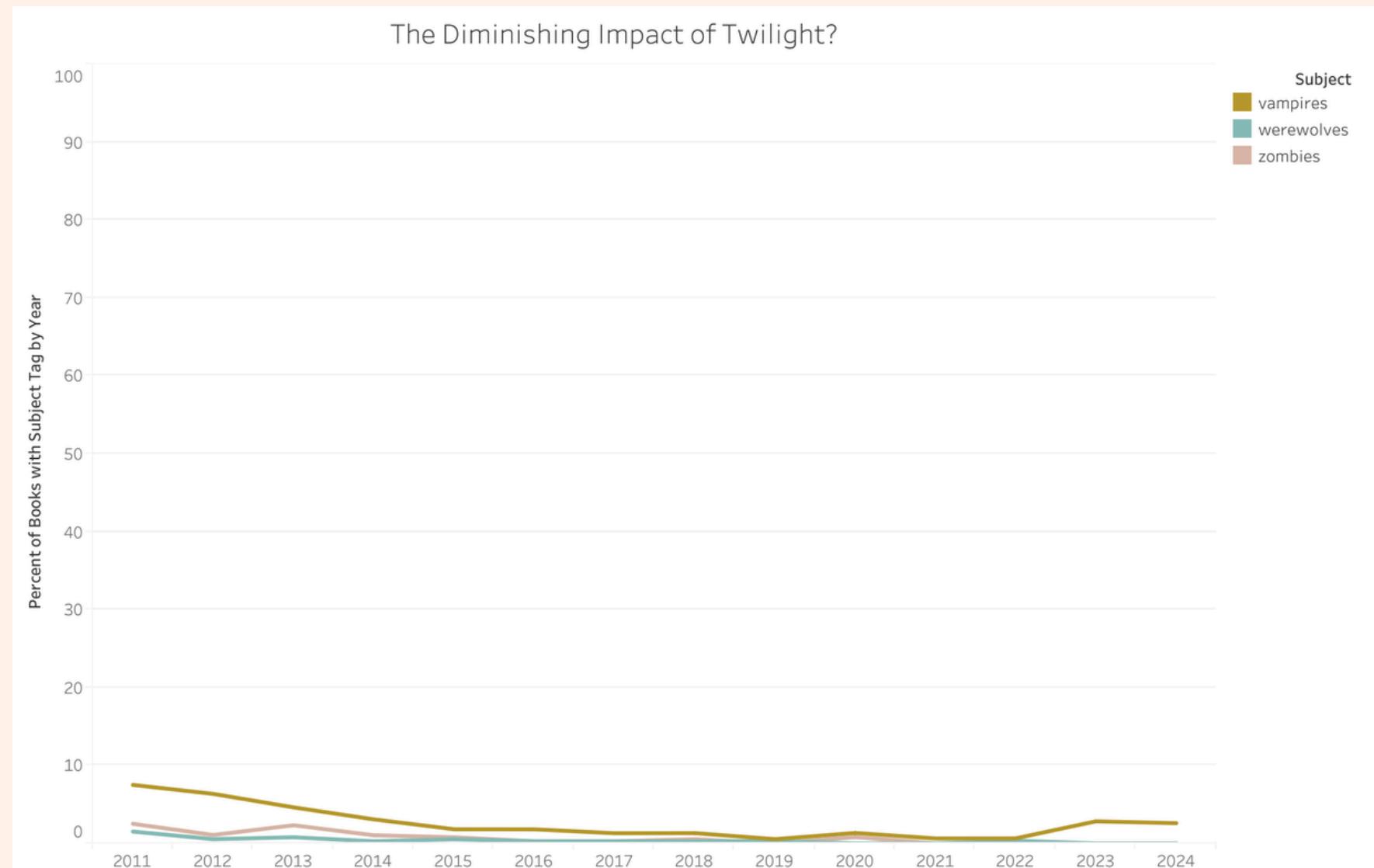


Potential Rise in
Romance novels
popularity *may be due
to category changes

Hypothesis: Taylor
Swift/Travis Kelce
sports romance spike?

Collections takeaway-
consistent interest in
contemporary
romance

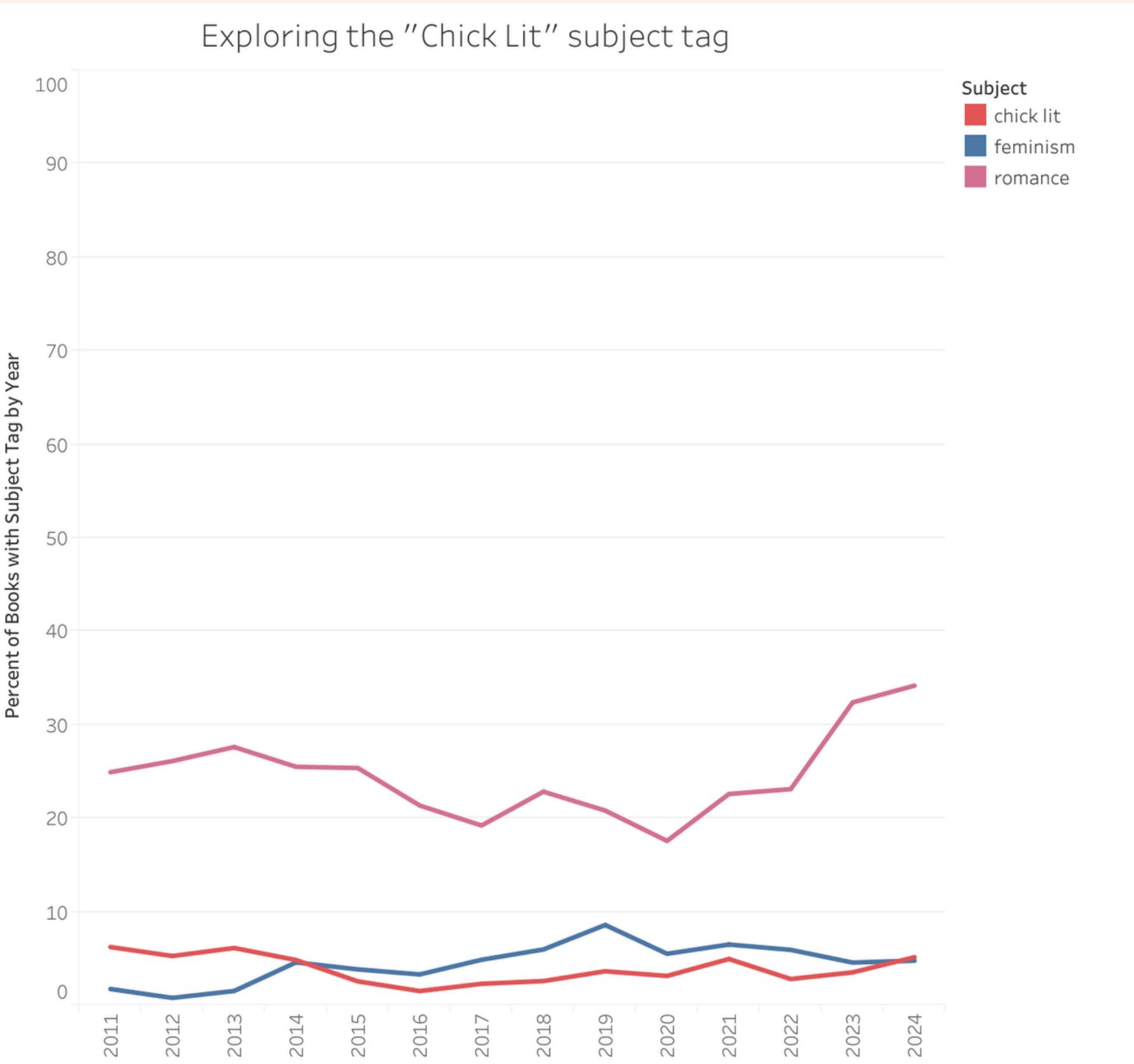
Sample Findings



Collections takeaway-
don't buy a bunch of
vampire books right
now

Sample Findings

Exploring the "Chick Lit" subject tag



chick lit, romance

Correlation: 0.24207674090157905

p-value: 1.0216723562656838e-68

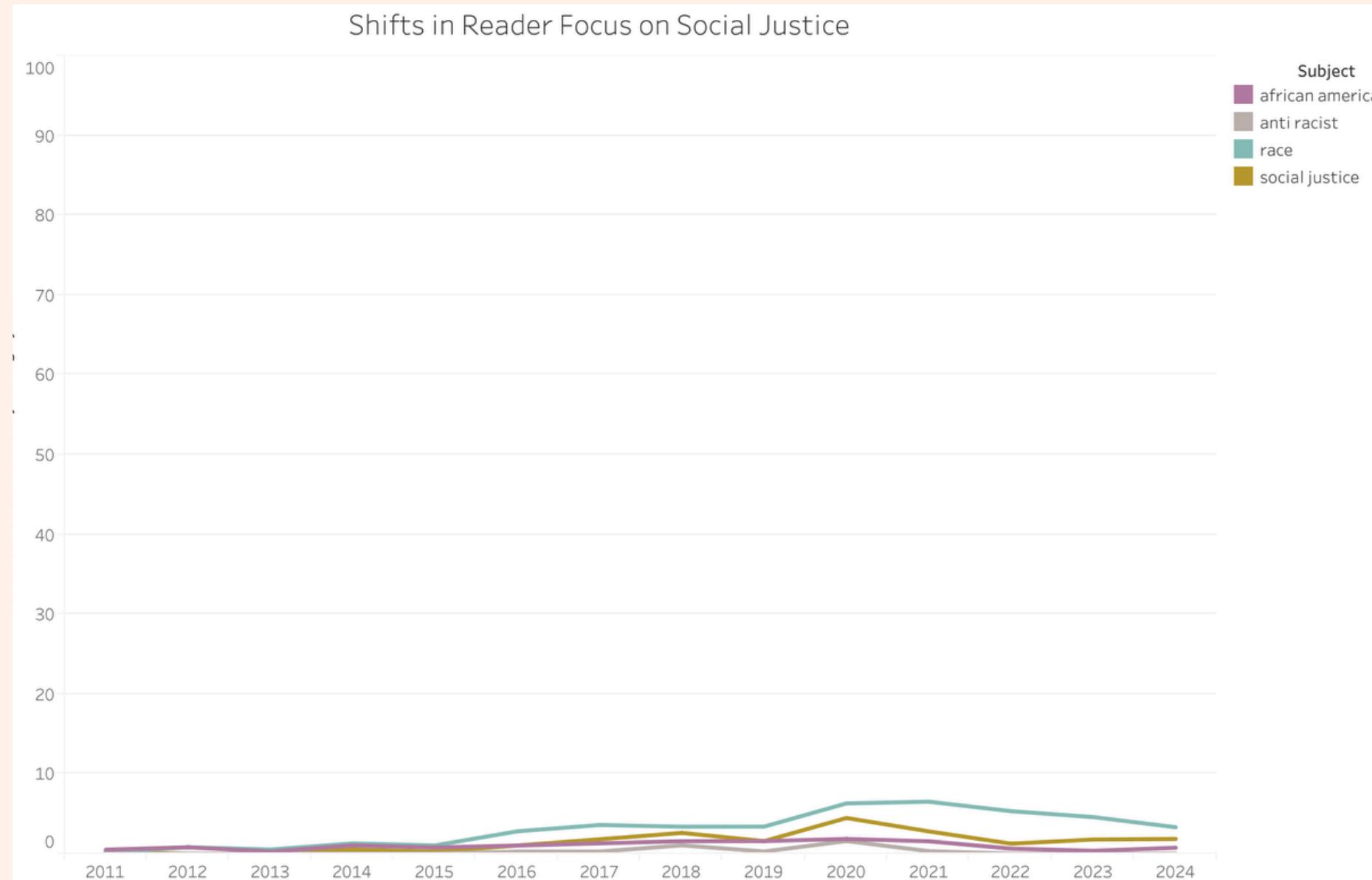
chick lit, feminism

Correlation: -0.03330097011877499

p-value: 0.017561322285696755

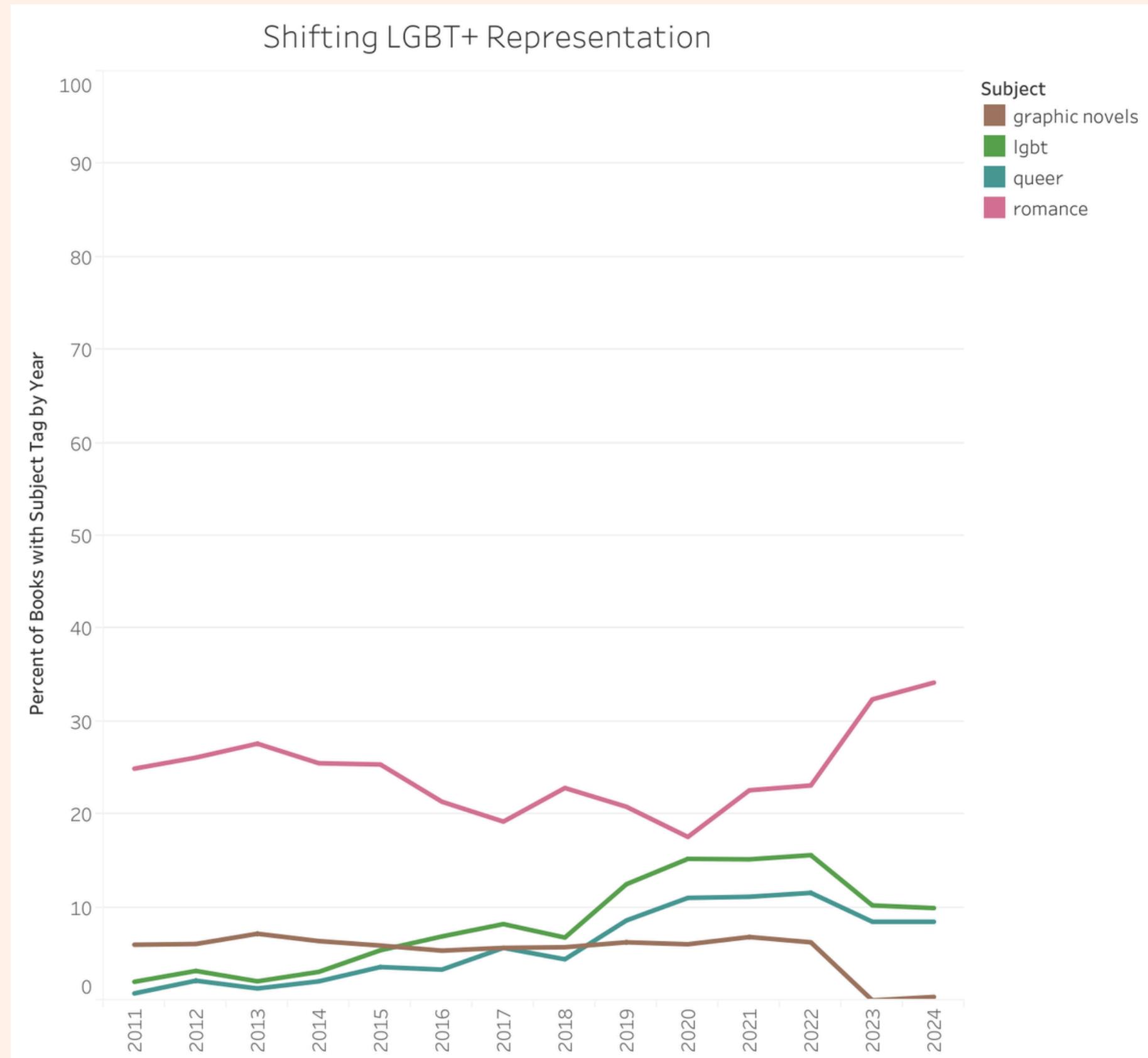
**Collections takeaway-
re-evaluate use of the
genre term chick lit**

Sample Findings



Collections takeaway-
look deeper at tagging
+ community interest in
social-justice related
books. Potentially
evaluate programming
+ book displays

Sample Findings



LGBT, romance

Correlation: 0.0615544229845321
p-value: 1.12063878632218e-05

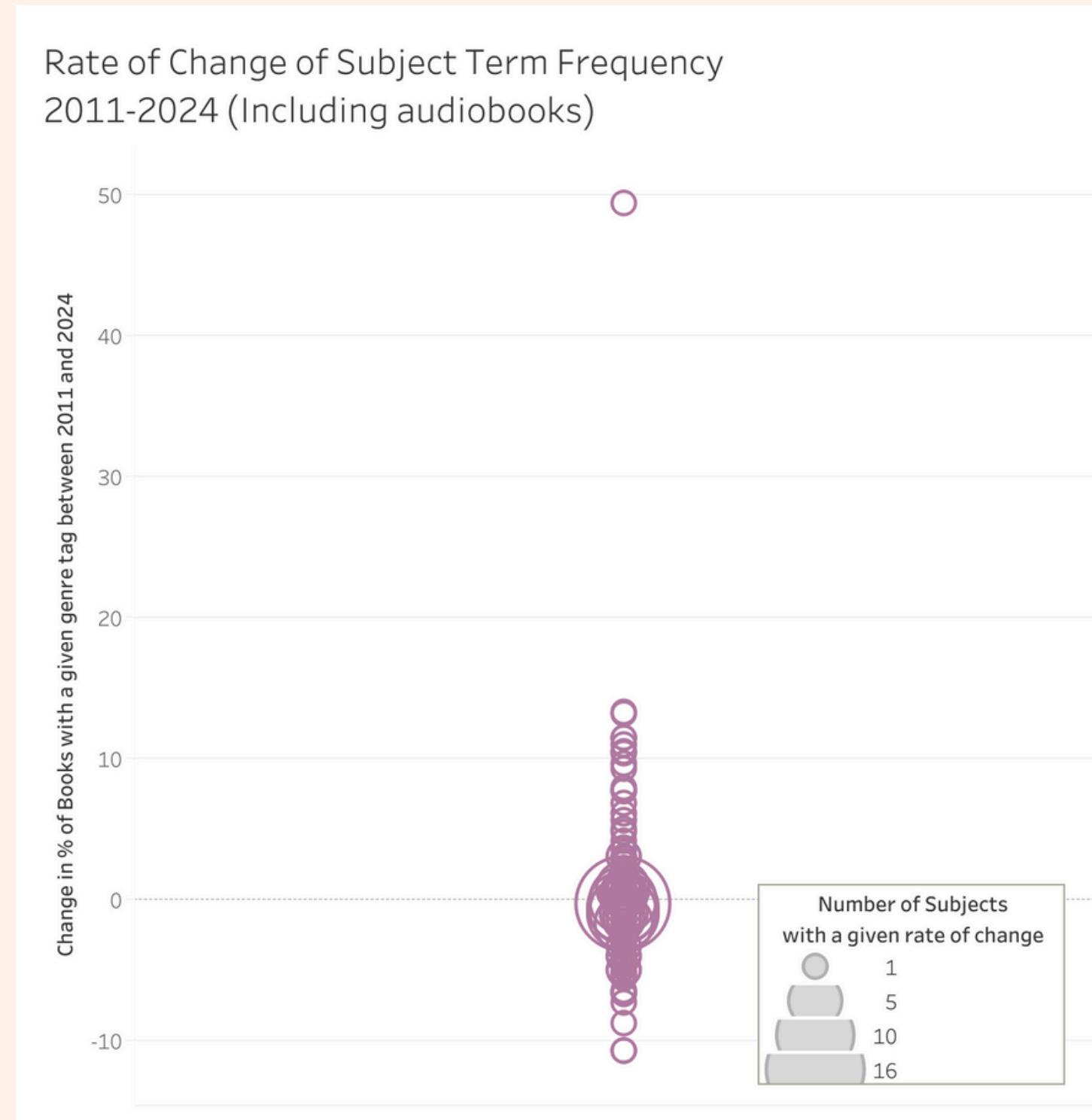
LGBT, graphic novels

Correlation: 0.07252166281769894
p-value: 2.2547387957731276e-07

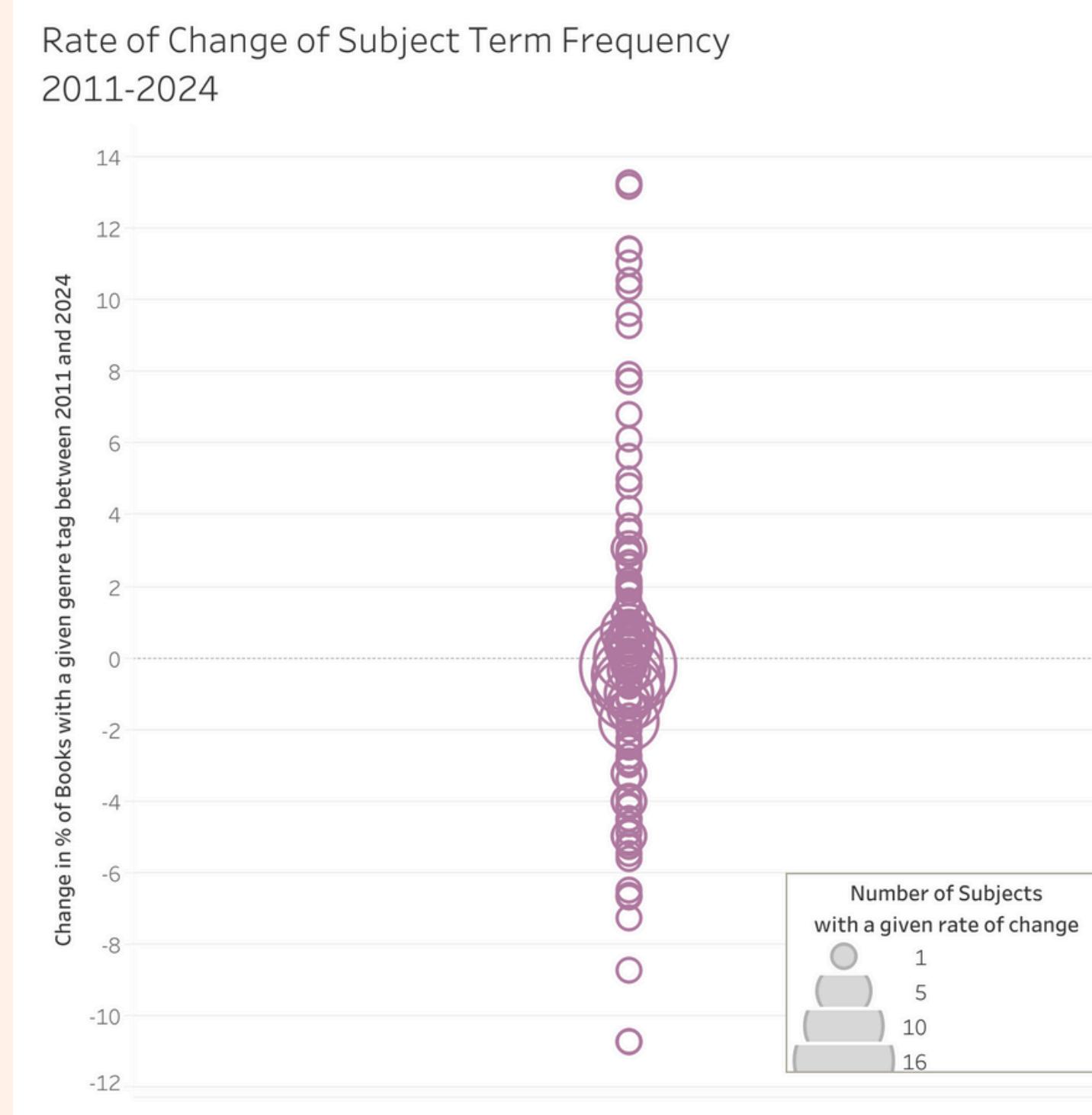
**Collections takeaway- explore
relationship btwn queer
representation and book
cateogories further**

Sample Findings

△ In subject
term frequency
(as % of total
books in each
year) between
2011 and 2024



Sample Findings



Collections takeaway-significant variation in persistence of trends.

Future Directions

- Write/develop further tutorials in GitHub on analysis processes
- Potentially develop workflows/tutorials for messier and larger datasets
- Topic modeling of description field

Thank you!



Project GitHub page

<https://github.com/mc2398/metadata-analysis-vis/>

contact me:

mcase393@pratt.edu



Project write up (in my portfolio)

https://mc2398.github.io/graduate_portfolio/goodreads_article.html