

What are the most popular themes in Science Fiction?

Philosophy of Science Fiction - March 2020 Thomas de Mareuil - DSB

As part of the Philosophy of Science Fiction course (HFC569) taught by J. Degenève at Ecole Polytechnique, I tried to conduct a short statistical analysis in order to identify the most common themes in science-fiction books.

The *Sci-Fi Books data set* that I used was collected by Dr. Kathleen M. Carley in 2017, and is available for download from the [CASOS website \(http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php\)](http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php). The dataset contains information about a sample of 157 science-fiction books written over the past 3 centuries and listed among the most popular of the genre. Information includes date, author, popularity index (number of lists of top N science fiction books that these appeared in circa 2016) and content of the story. For content, each book was read by 2 people, which then graded 11 possible themes (robots, time travel, psychic powers, post-apocalyptic, etc.) on a scale from 0 (theme not present at all in the story) to 3 (theme present, strong and integral to the story). I built this analysis on top of an online article by Zarrina Niyazova, available on [towardsdatascience.com \(https://towardsdatascience.com/which-content-of-the-story-is-the-most-popular-in-top-rated-sci-fi-books-6f3644246ac3\)](https://towardsdatascience.com/which-content-of-the-story-is-the-most-popular-in-top-rated-sci-fi-books-6f3644246ac3).

Let's now dive into the data to see what are the most common themes in popular SF books, and how these have evolved over time!

Most popular themes

There are 157 books and 18 features for each in the dataset.

	Book	Frequency	date	author	century	quarter century	author gender	robots, androids, AI	battles	
0	Frankenstein	6	1818	Mary Shelley	1	1	2	1	1	
1	Twenty Thousand Leagues Under the Sea	7	1870	Jules Verne	1	3	1	0	1	
2	Journey to the Center of the Earth	2	1874	Jules Verne	1	3	1	0	2	
3	The Time Machine	6	1895	H. G. Wells	1	4	1	0	2	
4	The War of the Worlds	3	1898	H. G. Wells	1	4	1	0	3	
5	Princess of Mars	2	1917	Edgar Rice Burroughs	2	5	1	0	3	

The most famous authors in the dataset (frequency in top sci-fi books list superior to 5) are:

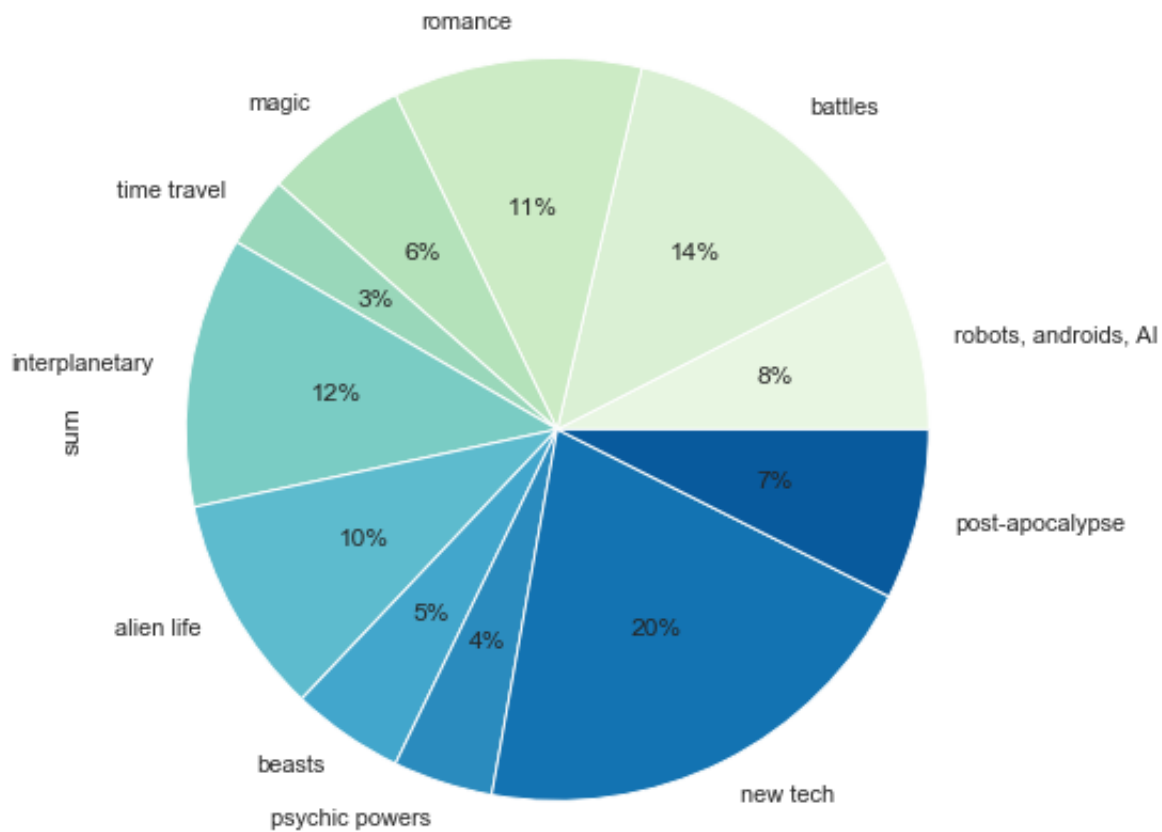
```
array(['Mary Shelley', 'Jules Verne', 'H. G. Wells', 'Aldous Huxley',
      'Isaac Asimov', 'George Orwell', 'Ray Bradbury', 'Alfred Bester',
      'Walter M. Miller Jr', 'Frank Herbert', 'Robert A. Heinlein',
      'Arthur C. Clarke', 'Kurt Vonnegut', 'Ursula K. Le Guin',
      'Larry Niven', 'Joe Haldeman', 'Samuel R. Delany', 'Douglas Adams',
      'William Gibson', 'Margaret Atwood', 'Orson Scott Card',
      'Dan Simmons', 'Connie Willis', 'Kim Stanley Robinson',
      'Neal Stephenson', 'Vernor Vinge', 'Robert Zemeckis',
      'Paolo Bacigalupi'], dtype=object)
```

Most popular overall

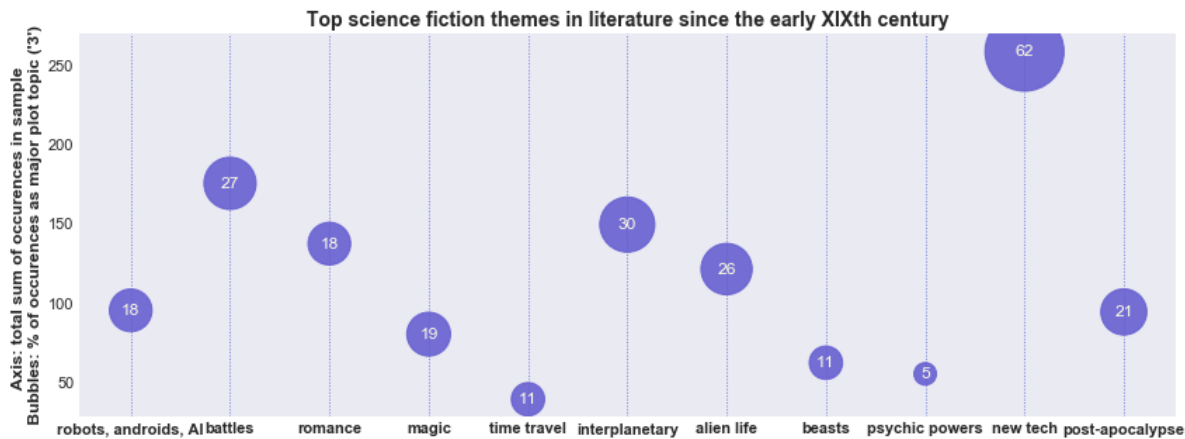
As evidenced in the table and graph below, the most frequent themes in sci-fi books (overall) are new technologies (not AI), battles and interplanetary settings.

	content	total sum	sum top theme	% top theme
0	robots, androids, AI	95	18	11%
1	battles	175	27	17%
2	romance	137	18	11%
3	magic	80	19	12%
4	time travel	39	11	7%
5	interplanetary	149	30	19%
6	alien life	121	26	16%
7	beasts	62	11	7%
8	psychic powers	55	5	3%
9	new tech	258	62	39%
10	post-apocalypse	94	21	13%

Most frequent content overall (%)



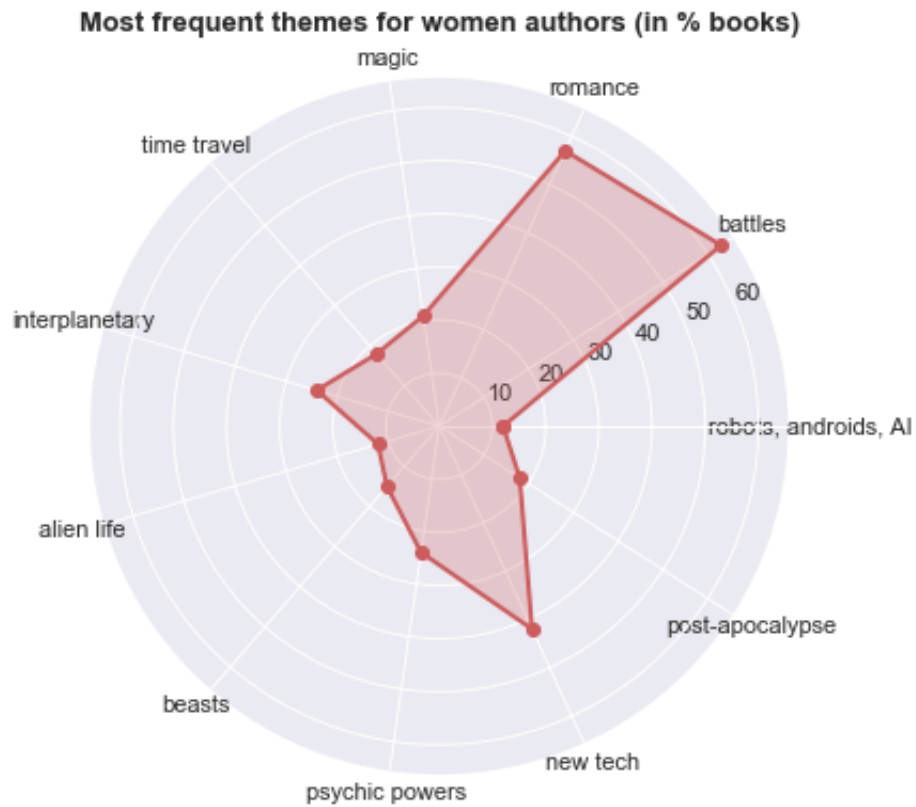
In this more visual representation, we can see both the overall occurrence (y axis) and popularity (size of bubble, with % of grades '3') of science fiction themes in literature. We see again that new techs, battles and interplanetary settings are the most frequent themes, both overall and in terms of top themes (grade = 3).



Men vs. women authors

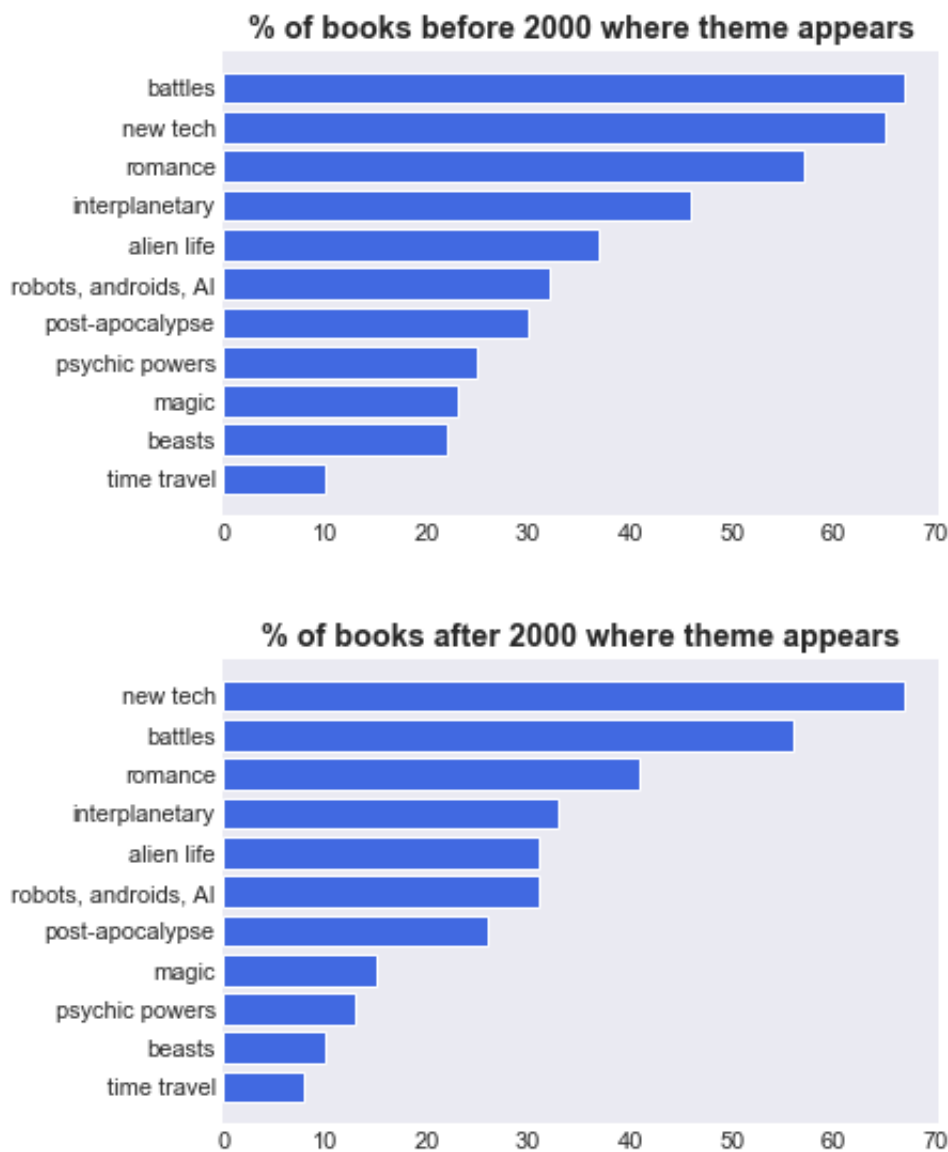
Below, we can see in "radar plots" the themes preferred by women vs. men authors.

For female authors, the preferred contents are romance, battles and new technologies. New technologies and battles are also top themes for men authors, but romance not so much! Men talk more about interplanetary adventures and aliens than women, while women tend to write more about psychic powers and time travel than men.



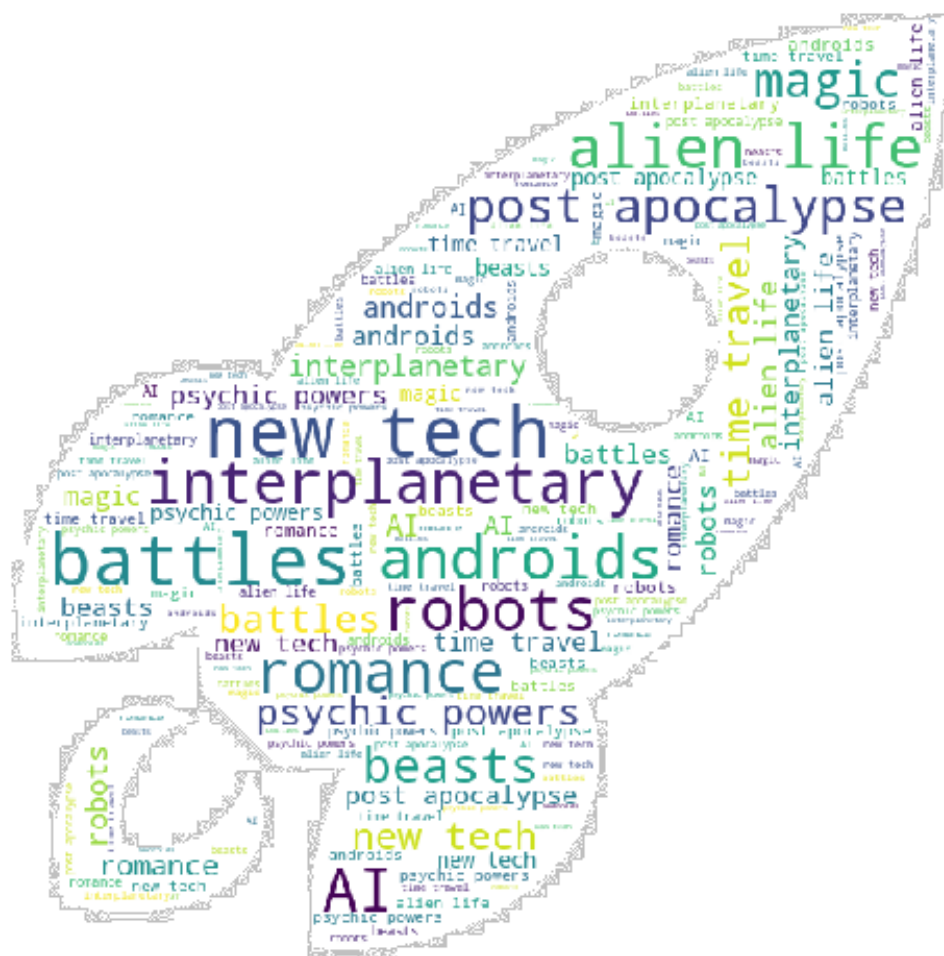
Before vs. after 2000

Let's take a look at differences in content since 2000 compared to before 2000. We can spot interesting evolutions: older books are more about romance, interplanetary trips, aliens, and battles, while more recent books focus more on new tech and androids/robots/AI.



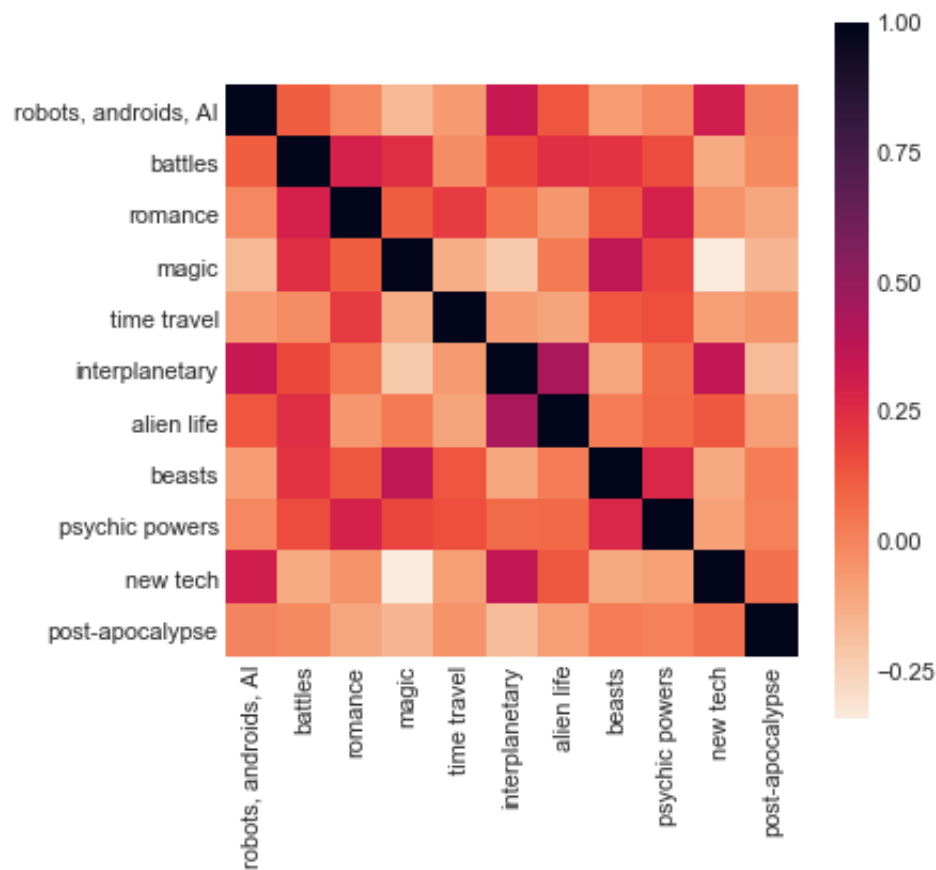
WordCloud

In this WordCloud representation, the most frequent themes are repeated more often... in the shape of a spaceship!

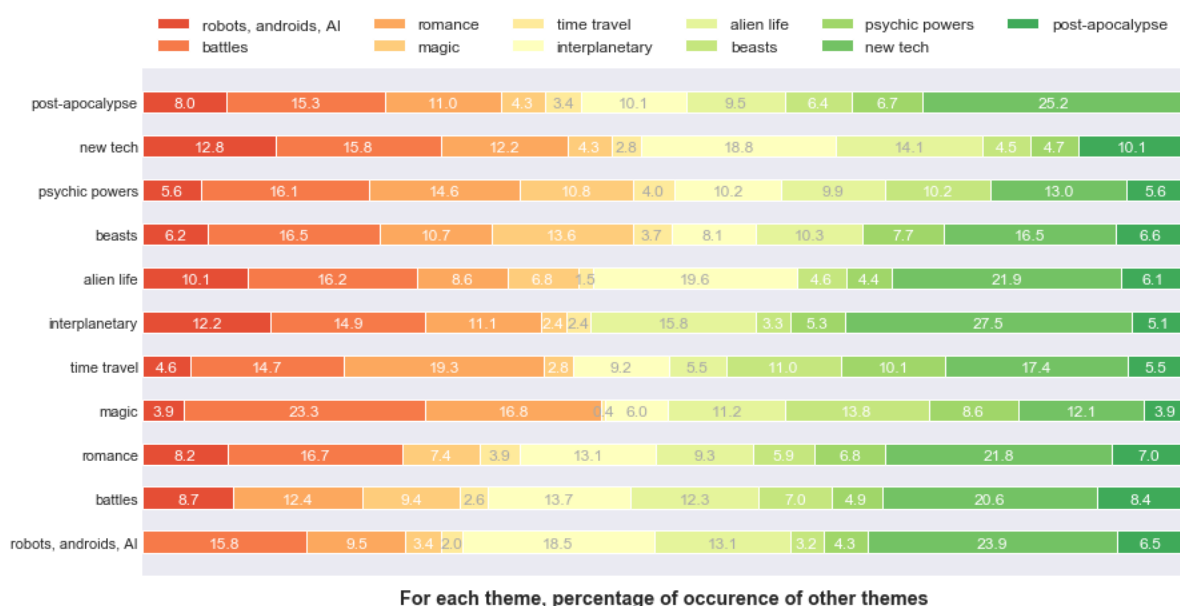


What themes are most commonly associated in science fiction books?

On the correlation plot below, the plotting color corresponds to correlation; the darker the colour, the more likely the association of 2 themes.



The correlation plot isn't easy to read: for a more visual graph, the horizontal bar chart below shows for each theme the % of occurrence of other themes. For example, the most common second content for time travel is romance. Also, robots and magic or psychic powers are rarely found together in the same book, while alien life is commonly associated with battles. Unsurprisingly, most of the times the second content for post-apocalypse and interplanetary settings are new technologies.



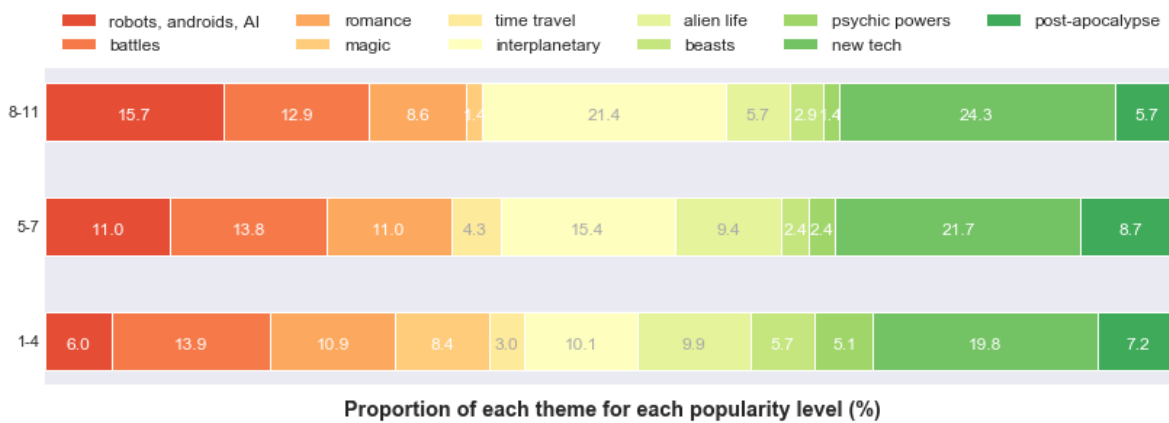
For each theme, percentage of occurrence of other themes

Most highly rated content

First, let's sort out books by their frequency in top science-fiction books rankings. The most popular book in the dataset is *The Hitch-hiker's Guide to the Galaxy* (D. Adams, 1979).

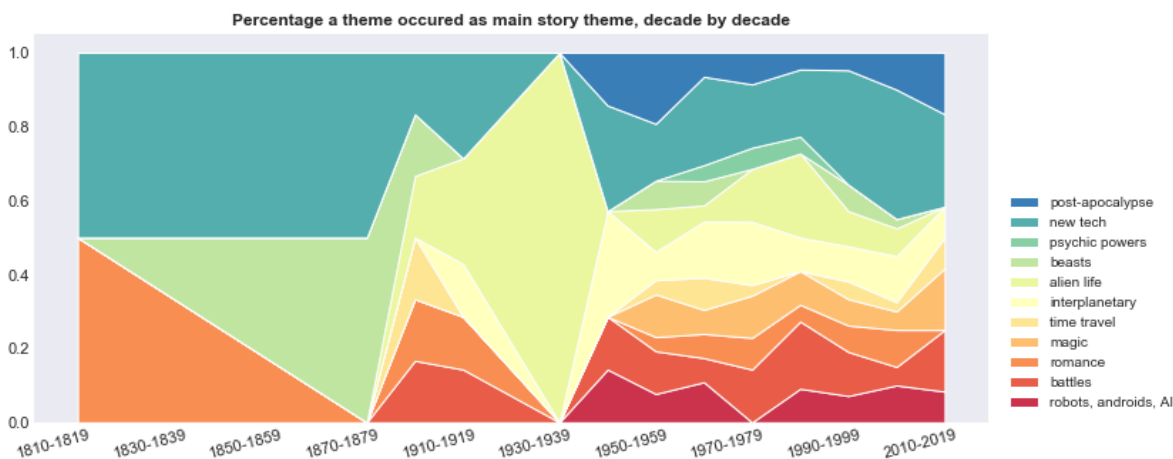
	Book	Frequency	date	author	century	quarter century	author gender	robots, androids, AI	battles
79	The Hitchhiker's Guide to the Galaxy	11	1979	Douglas Adams	2	8	1	1	0
85	Neuromancer	10	1984	William Gibson	2	8	1	3	1
46	The Dune Chronicles	9	1965	Frank Herbert	2	7	1	0	2
11	The Foundation Trilogy	8	1942	Isaac Asimov	2	6	1	3	2
70	The Forever War	8	1974	Joe Haldeman	2	7	1	2	3
62	Ringworld	8	1970	Larry Niven	2	7	1	2	1
52	2001: A Space Odyssey	7	1968	Arthur C. Clarke	2	7	1	3	0
89	Ender's Game	7	1985	Orson Scott Card	2	8	1	1	3

We can also plot the most frequent themes for each popularity level (1-4, 5-7 and 8-11). Interestingly, we can see that the presence of robots/androids/AI seems to be positively correlated with popularity! Same for interplanetary adventures. Conversely, magic, aliens, beasts and psychic powers are more present in less popular books.

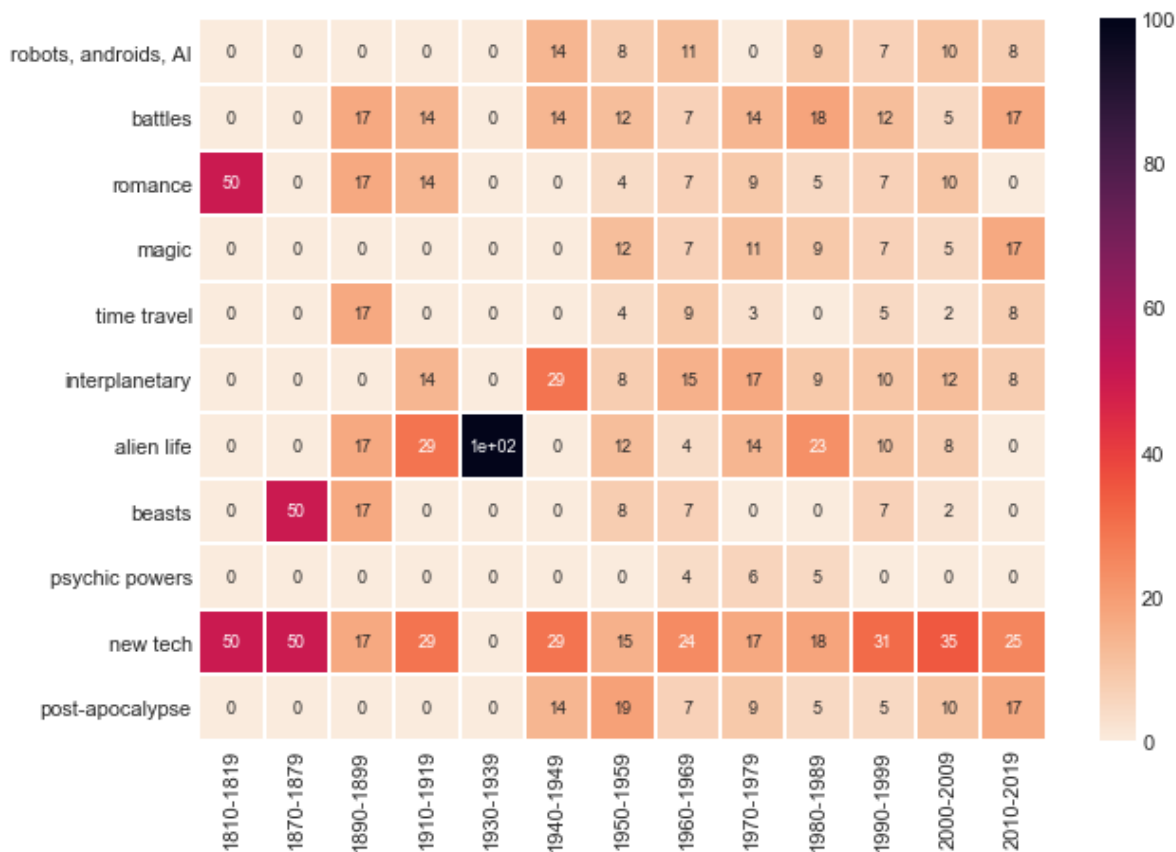


Evolution of themes over time

Last, let's take a look at content evolution over time (in terms of main book content, i.e. grade = 3). Thanks to the "stack plot" below, we can get a number of interesting insights. For example, we see that post-apocalyptic stories emerge only in the 30's. Same for stories about androids, robots and AI, which weren't present in science fiction literature before the 30's. There is less available data for the XIXth century, but we can still observe that stories about beasts or romance (as main plot theme) were more common in the past than in recent decades. In the XXIst century, stories are not about psychic powers or aliens any more, but more about apocalypse, new techs or robots/AI.



This last plot shows the same data as a heatmap (the higher the percentage of occurrence of a certain theme, the darker the colour - for each decade).



Conclusion

In conclusion, the results of this analysis are that the most frequent contents of sci-fi books overall are new technologies (not AI), other planets and battles. Male authors write more about new techs and space, while female authors write more about new techs and romance. Alien life is commonly associated with battles, while for post-apocalypse and battles the most frequent 2nd content in the same book is new technology. Beasts are no longer the main content of sci-fi stories, while apocalypse, robots or AI have emerged in the middle of the last century and remain steadily popular. The most popular content overall is new tech, but top rated books deal more frequently with AI, robots and other planets.