



DATA VISUALISATION PROJECT ON BOOK REVIEW DATASET

BY: ASIF IKBAL & MAHENDRA NANDI

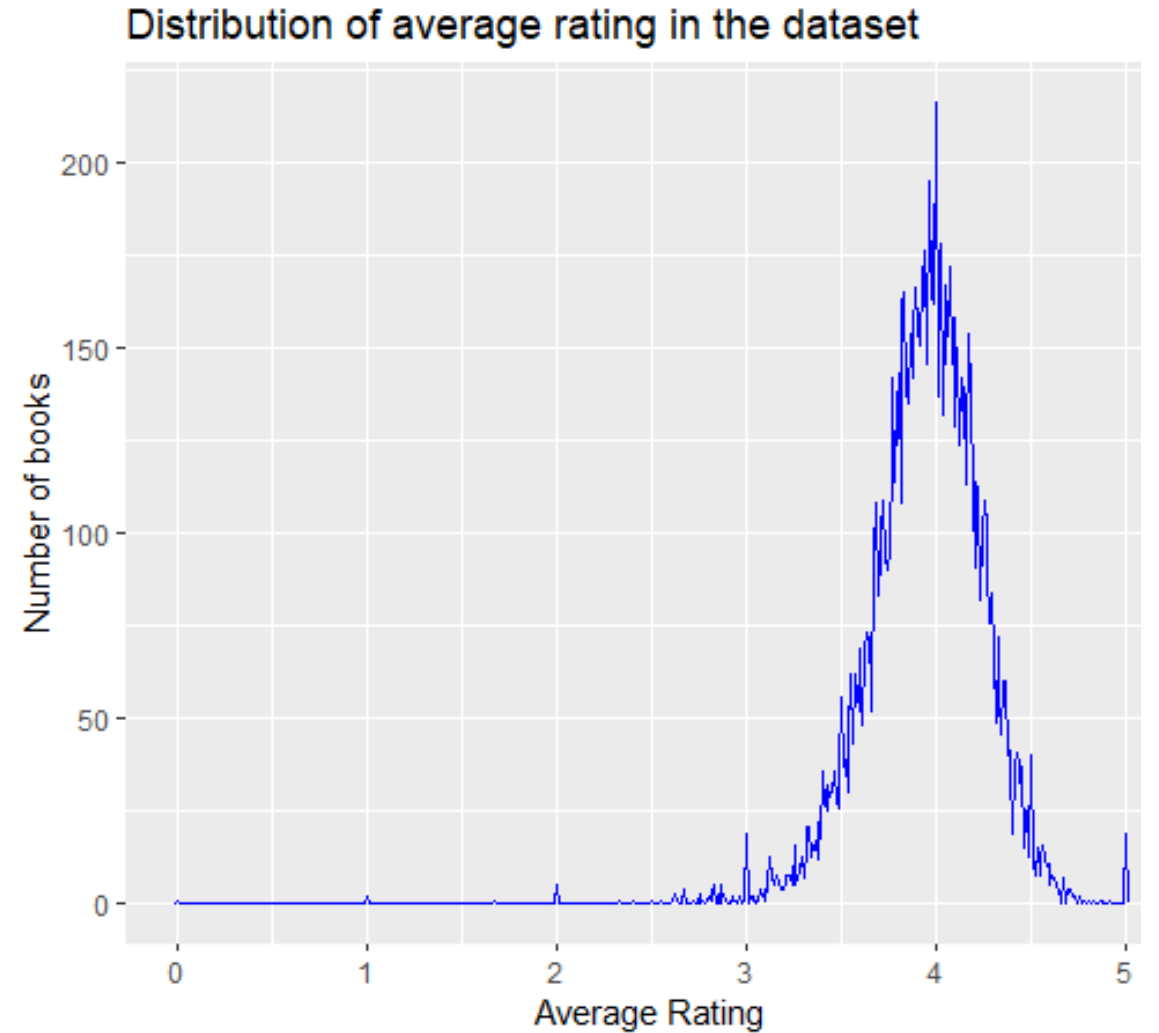


Visualizing Book Review Dataset

The dataset consists 13 columns. Our primary goal is to visualize all the dependencies among them and finding out the key factors for a good review of the books. A good review in the sense that it may have a higher average rating, higher number of reviews.

Dependencies:

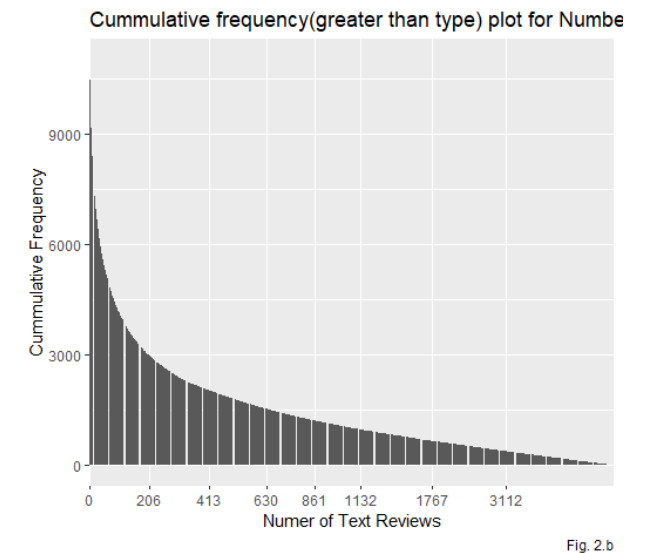
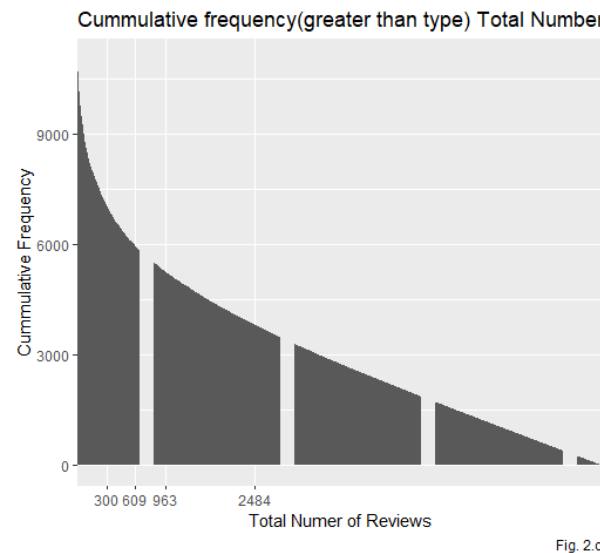
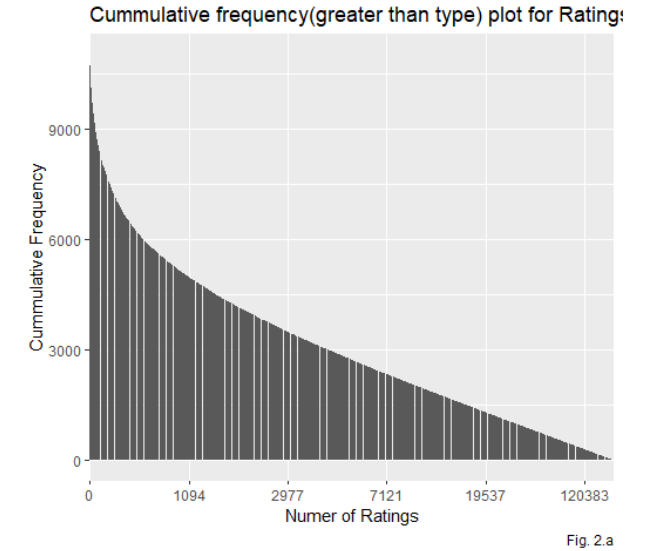
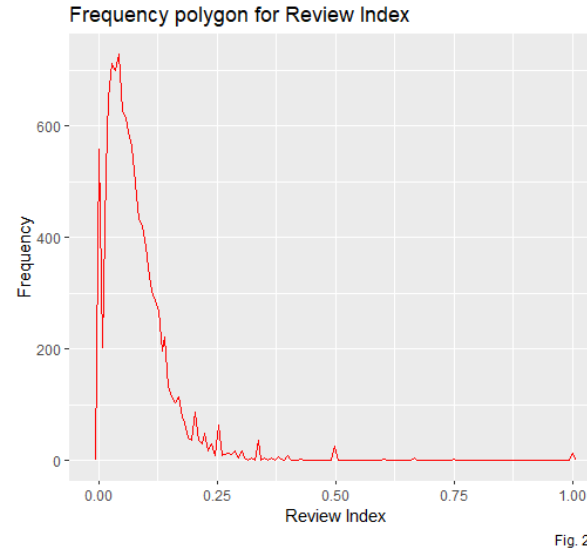
- (1) Distribution of average rating
- From the distribution of the average rating in Fig. 1 it is clear that almost all the books have been rated in the range 3.5-4.5. Out of almost 11000 books 22 books is rated as "5".



Fig

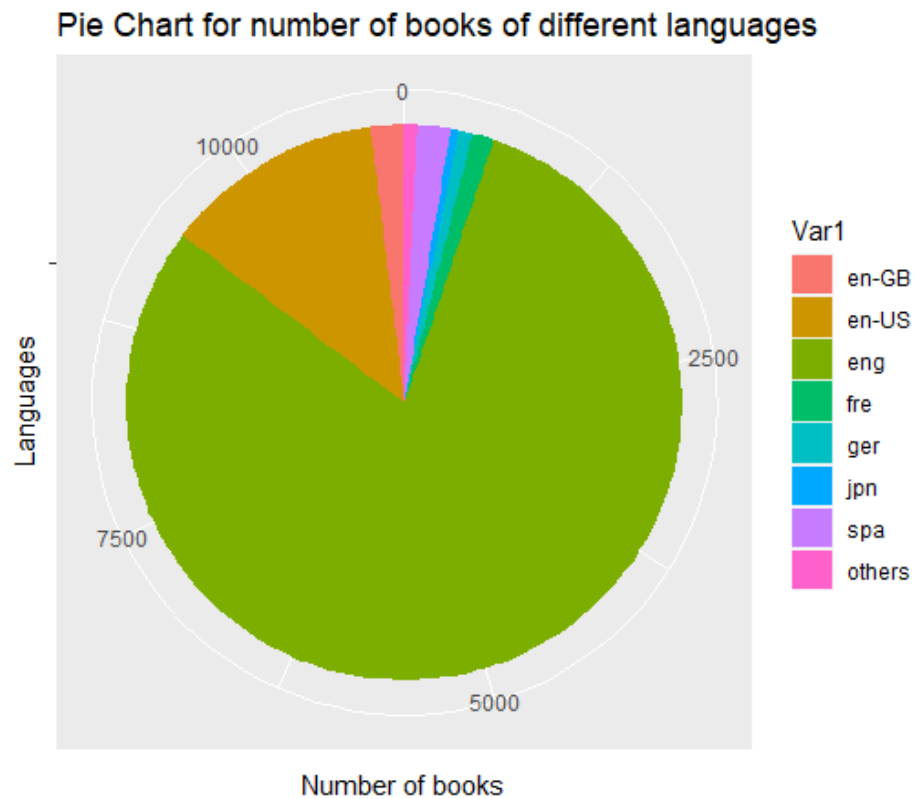
(2) Dependencies between different review parameters and average ratings

We can see that maximum books got reviews less than 20000 . Fig. 2.a,2.b,2.c and 2.d shows the distribution.



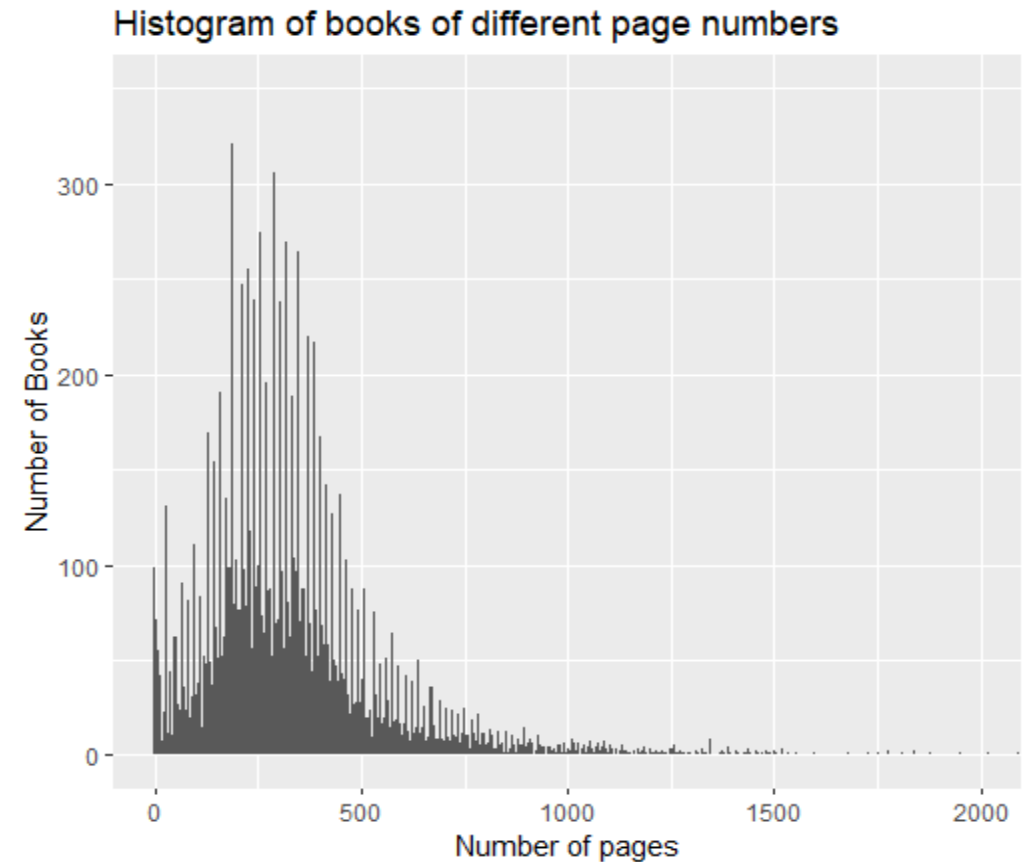
- (3) Number of books in different languages

- Here maximum books are written in “eng” language . 80% book is written in “eng” 13% in “en-US” and the remaining 7% is in other language.



- (4) Number of books in different page ranges

- Pages of maximum books(around 60%) are in the range of 200-400



- (5) Number of authors having exact number of books
- There are more than 6000 authors who wrote only one book. And 1000 authors wrote two books. The number of authors who wrote more than 5 books is very low.

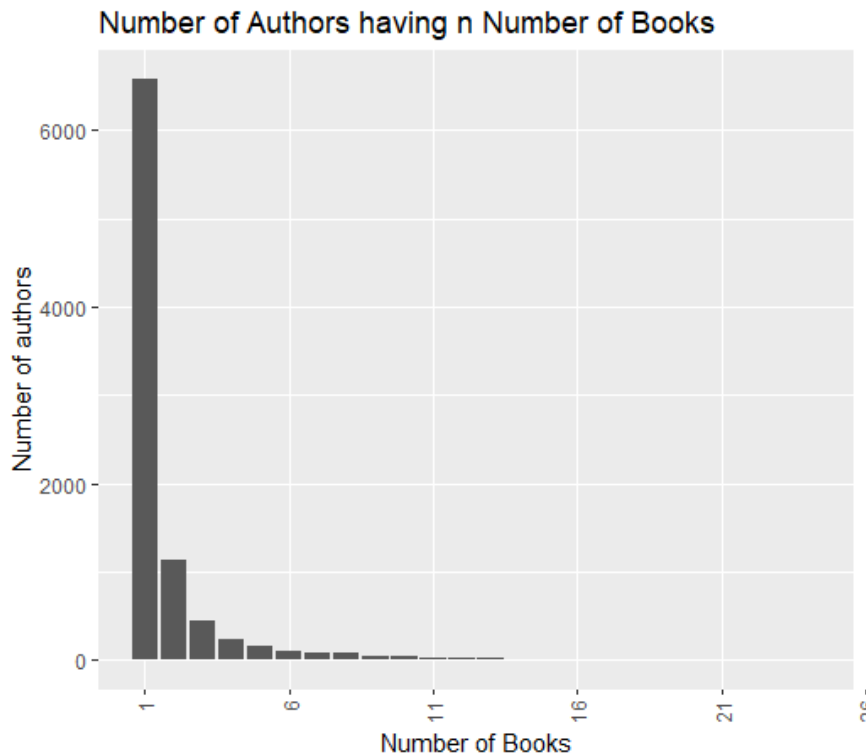


Fig. 5

- (6) Books published in a different years
- It is seen that in every year no of published books is increased largely. But after 2016 it suddenly goes down.

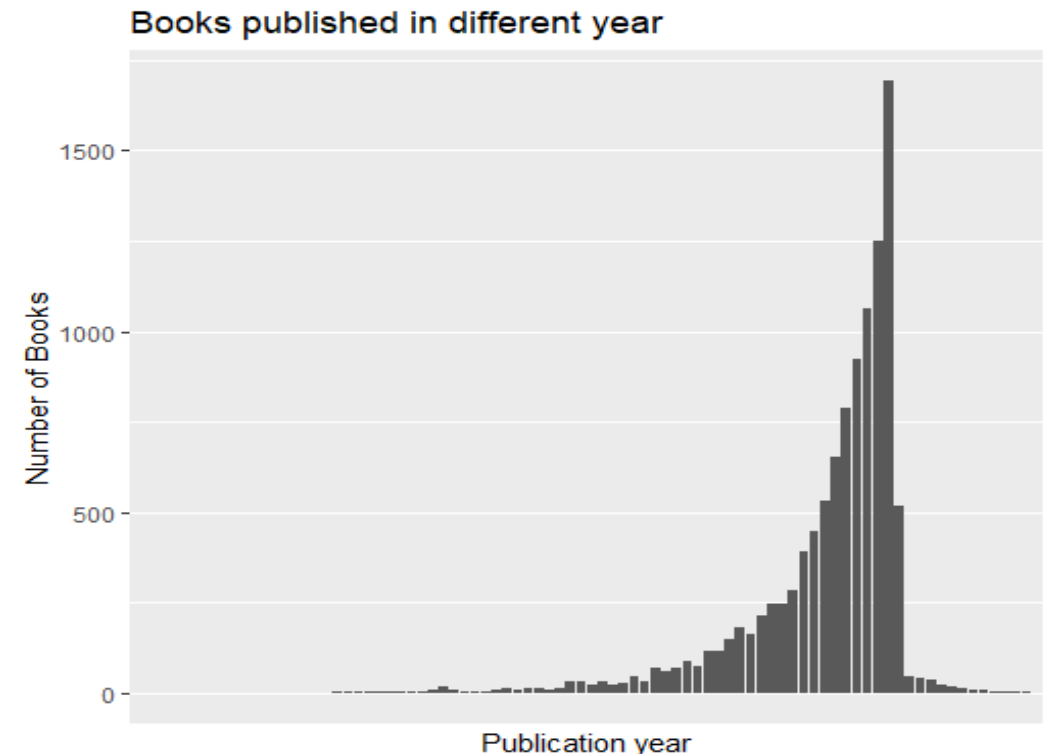


Fig.6

(7) Dependency between Average rating and different languages

- Though the number of books written in “eng” is much greater than other languages , average rating of books written in 'eng' is ~3.9 . Whereas very few books are written in “wel” but the average rating is 5 .

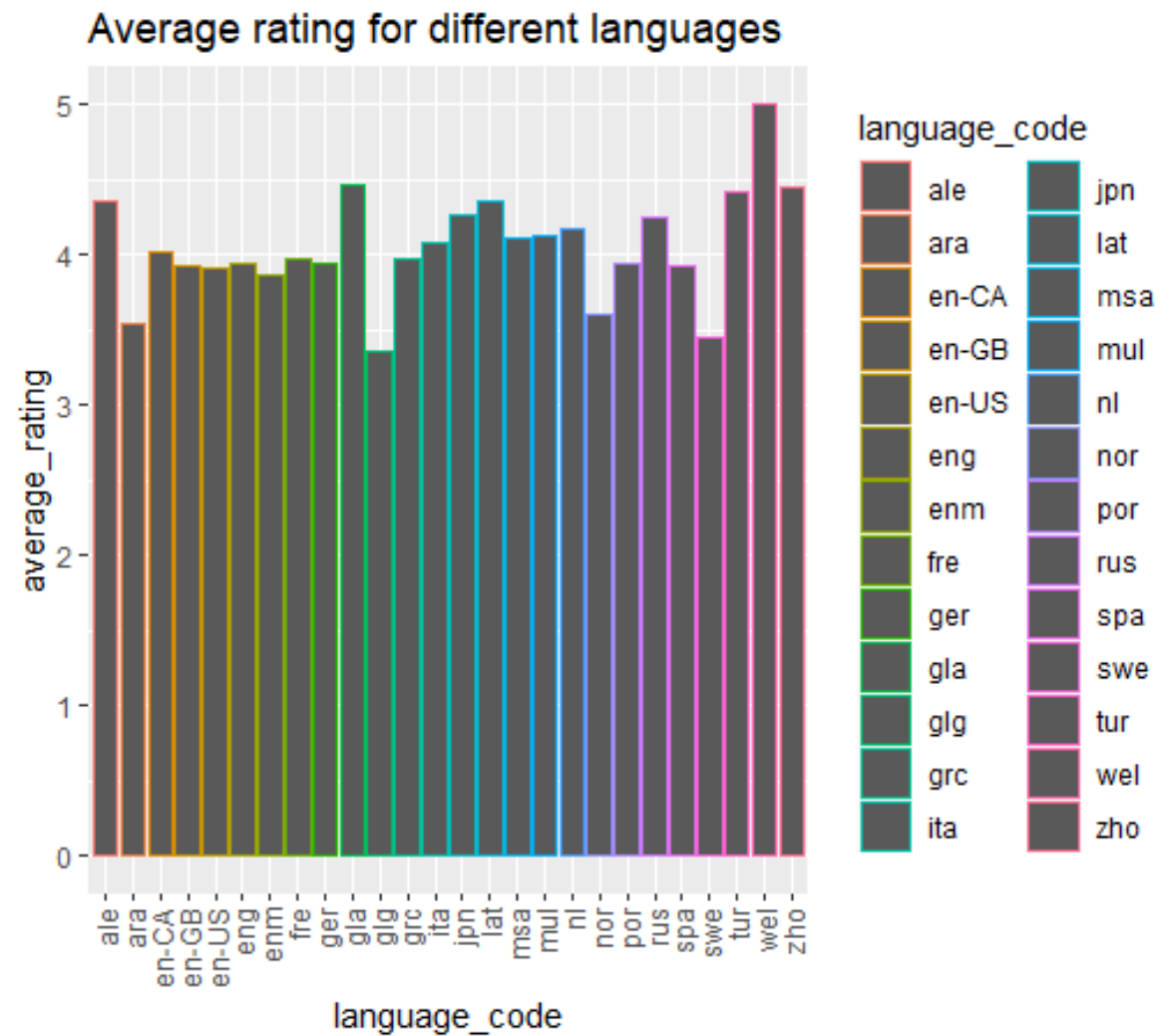


Fig. 7

- (8) Relation between Average Rating and Number of pages
- We can find here that rating is almost varies as the num of pages of the book . Book with no of pages 2600-2700 is of high rated . Though there are few books in this range , It actually ensures that if a book is interesting to the readers, total number pages of book does not matter.

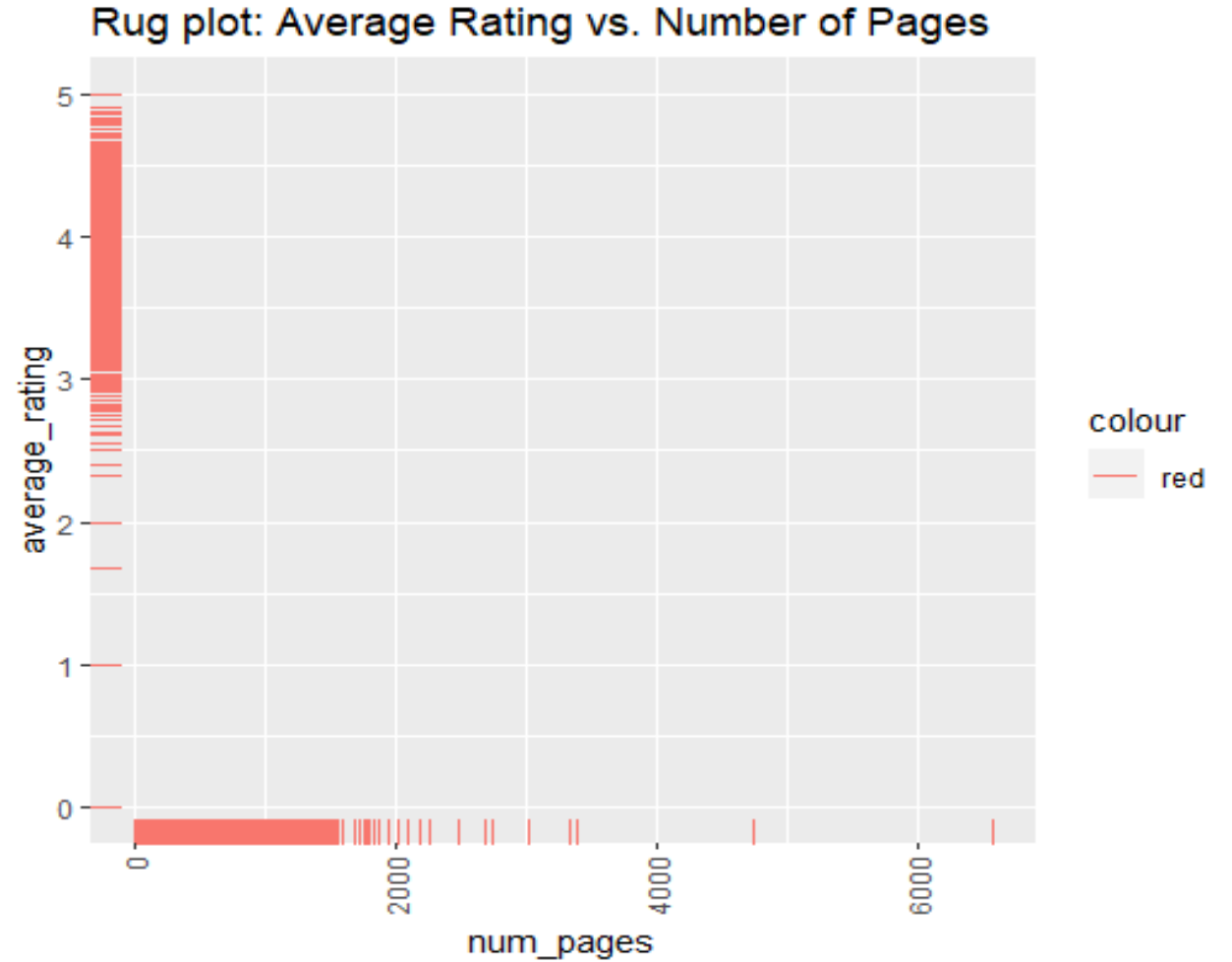
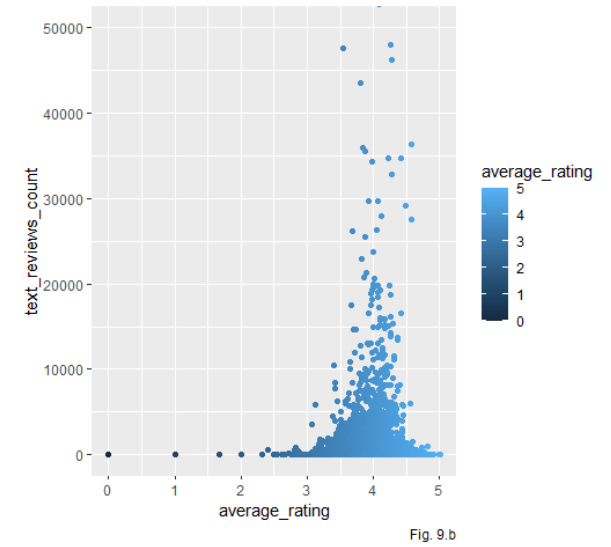
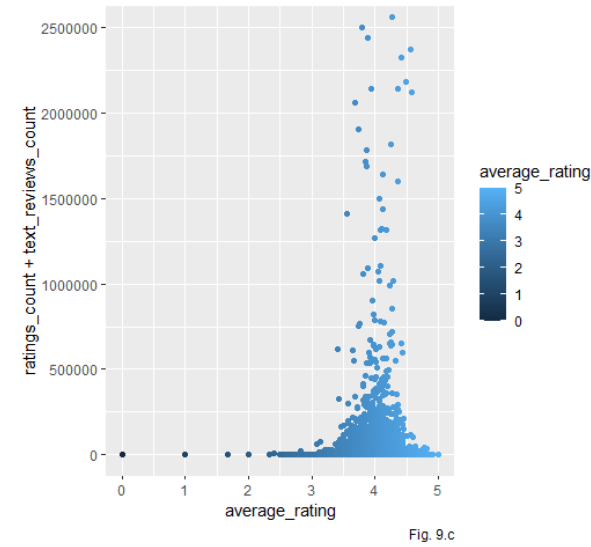
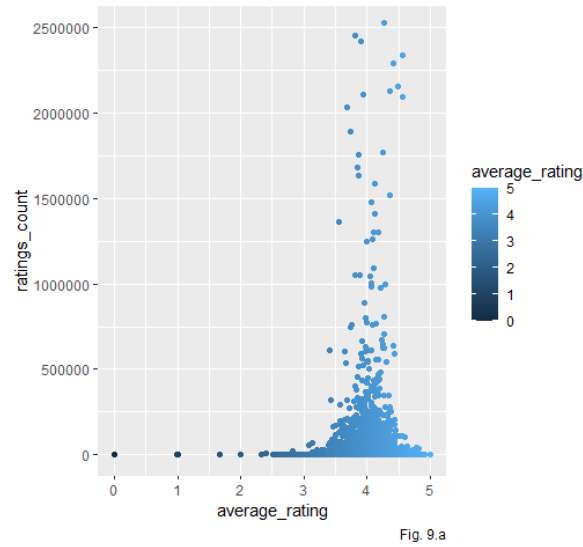
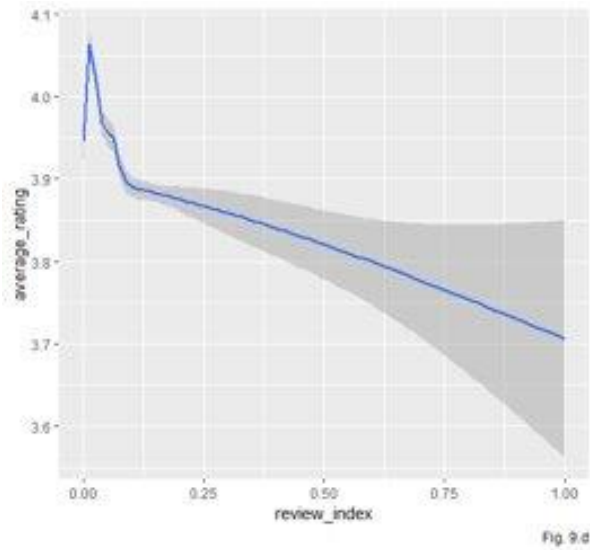


Fig. 8

- (9) Dependencies of the various review parameters with average ratings

- As we earlier saw that there are very few books with very high number of reviews . The books with very high reviews also fall under the rating range 3.5-4.5 . So, we can say maximum people, who gives a review , responds to a book with an average rating of 3.5-4.5 ,unless they get a good satisfaction with the book or become totally hopeless .Which is intuitively acceptable.



(10) How average rating is differed for different publishers

There are 2290 publishers published the books . Out of 2290 publishers 47 publishers got average rating greater than 4.5 , 2190 got in the range [3.5,4.5] , and the remaining 134 publishers got rating less than 3.5.

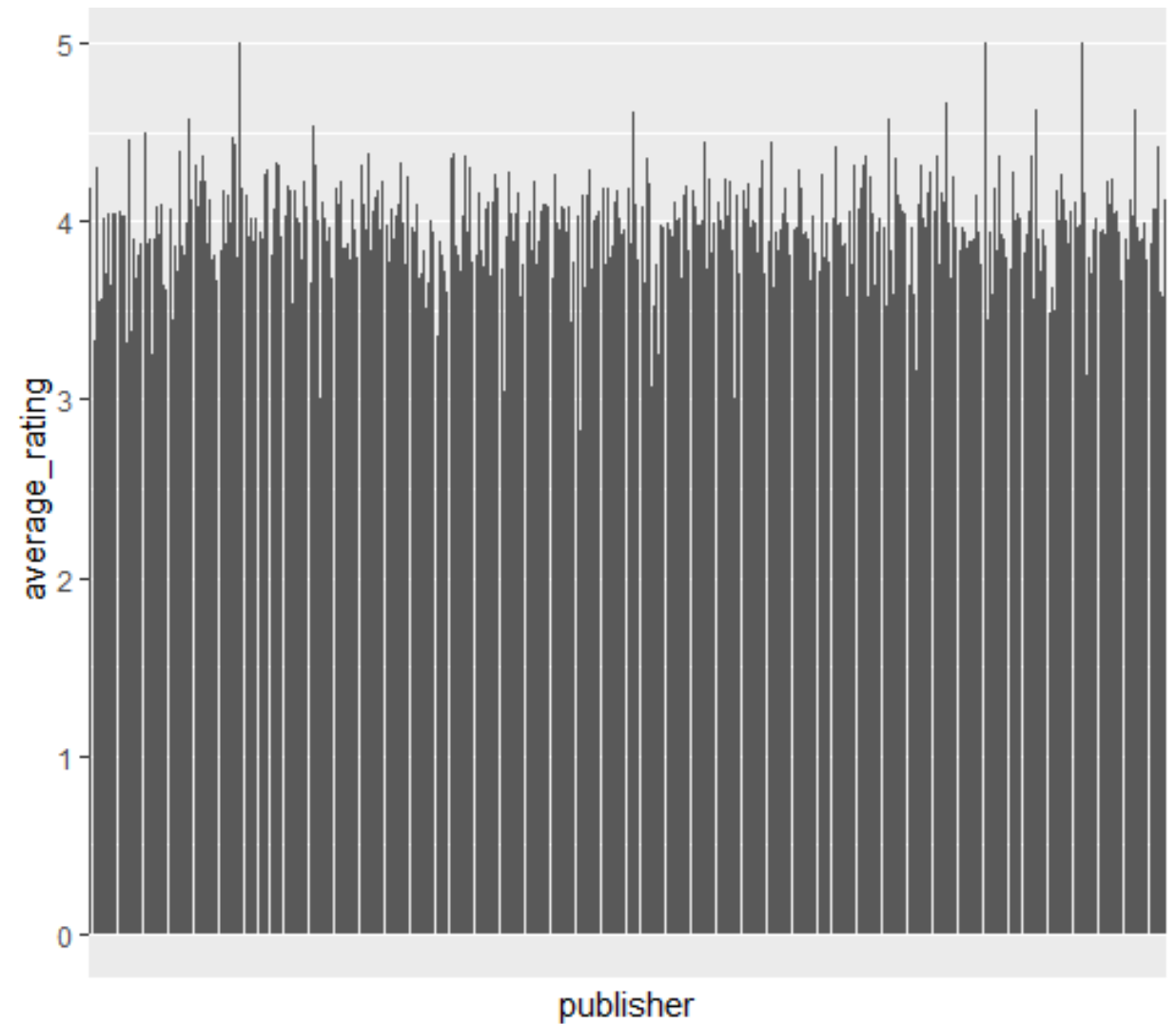


Fig 10

(11) Average number of pages for different languages

- Among the languages, the most used language is 'eng'. And we can see that books with "eng" language have 300 pages in average. Whereas books written in "enm" has maximum no of average pages (around 1100). But it is also a point to note that books of most average-rated language "wel" are of total average pages 150 (which is minimum among the used languages)

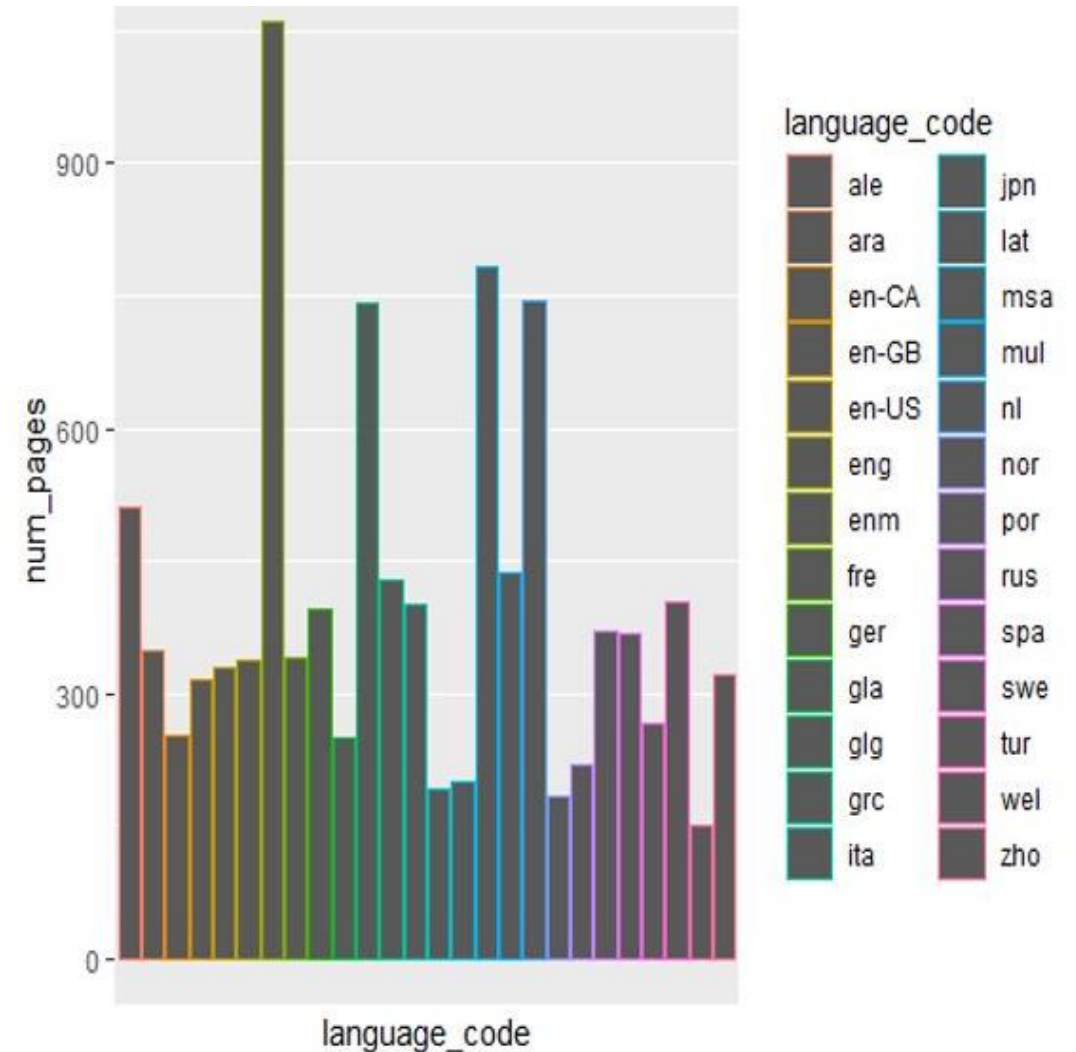


Fig. 11

(12) Number of reviews for different languages

- As the most used language to write a book is “eng”, it has highest total reviews . But also it has a highest number of reviews which is far more from others . So we can conclude that most of the people who read books of ‘eng’ language used to give reviews after reading it .

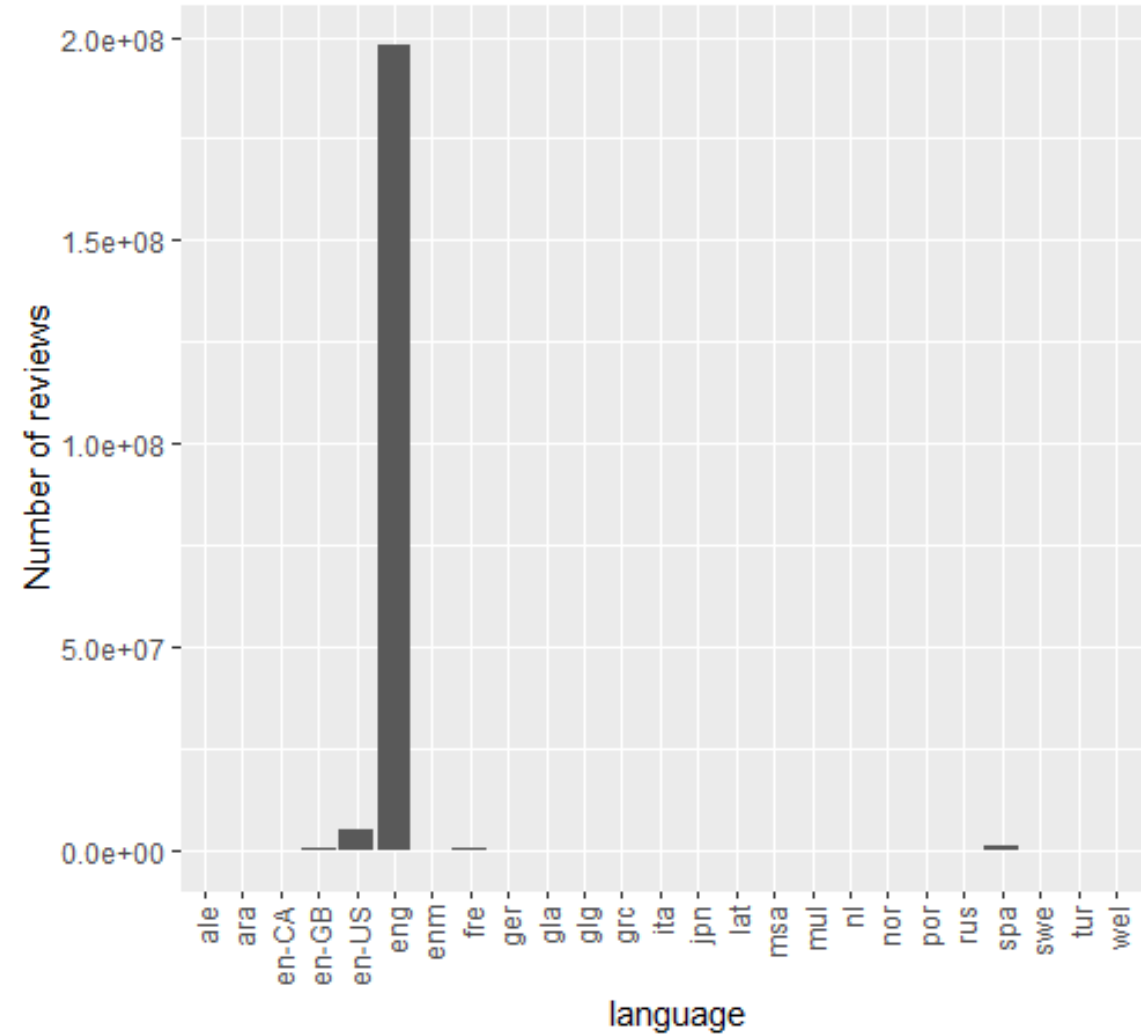


Fig.

(13) Dependency between Number of pages and Review parameters

- Books are reviewed most for the books having pages less than 1000. Number of reviews for books having more number of pages are very low.

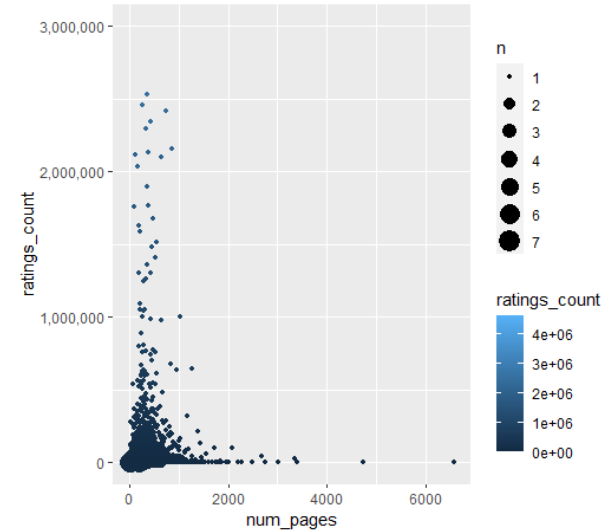


Fig. 13.a

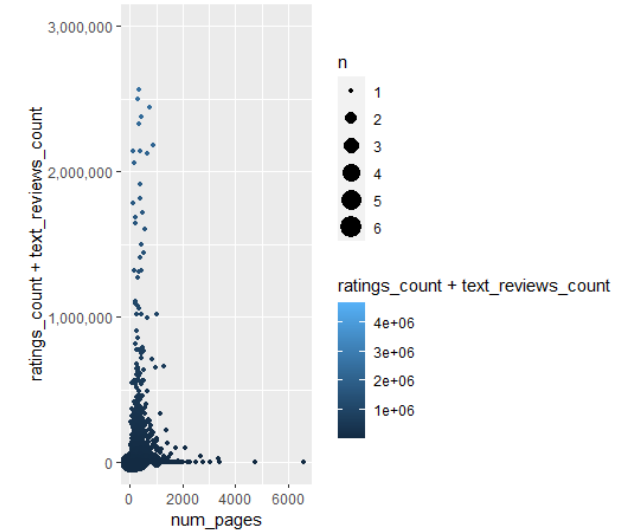


Fig. 13.c

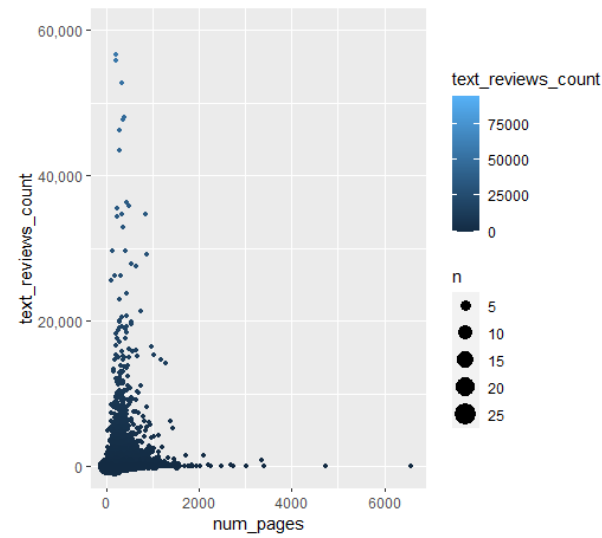


Fig. 13.b

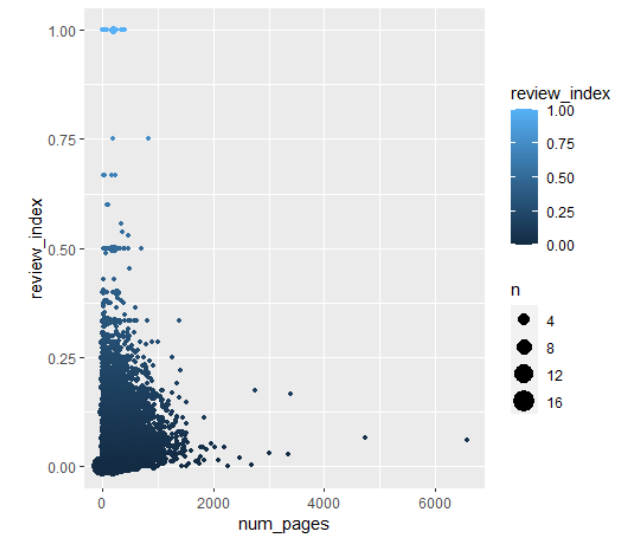
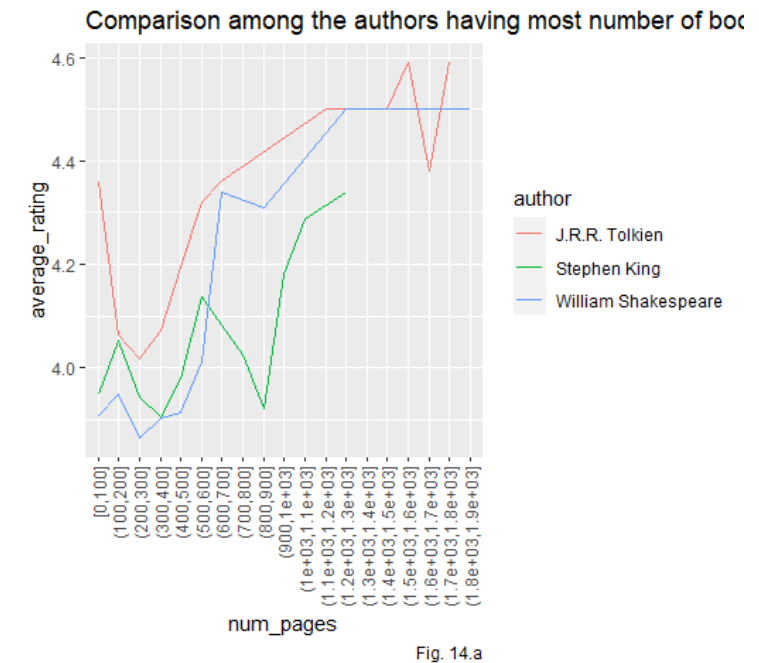
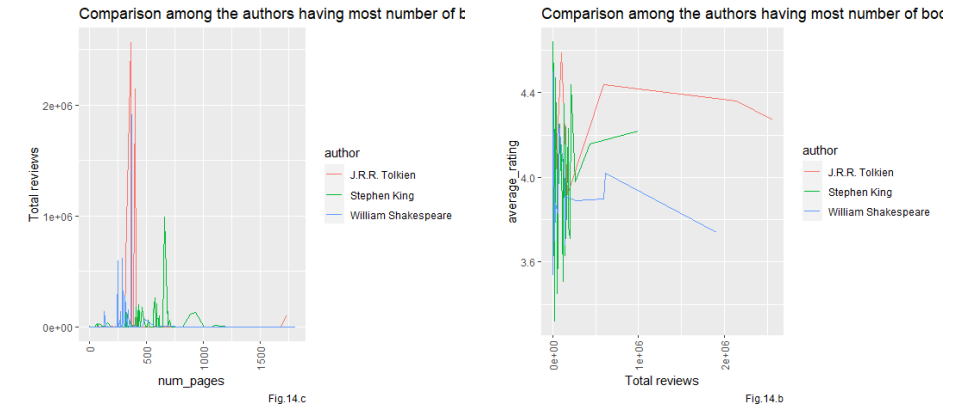
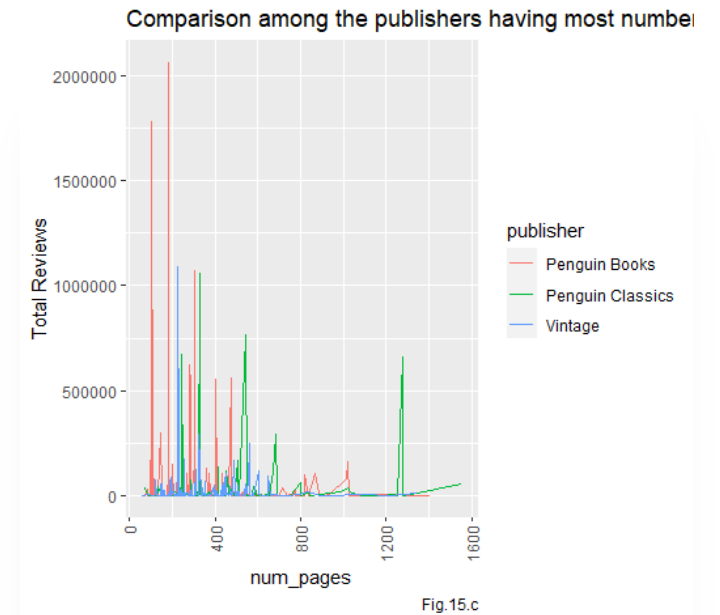
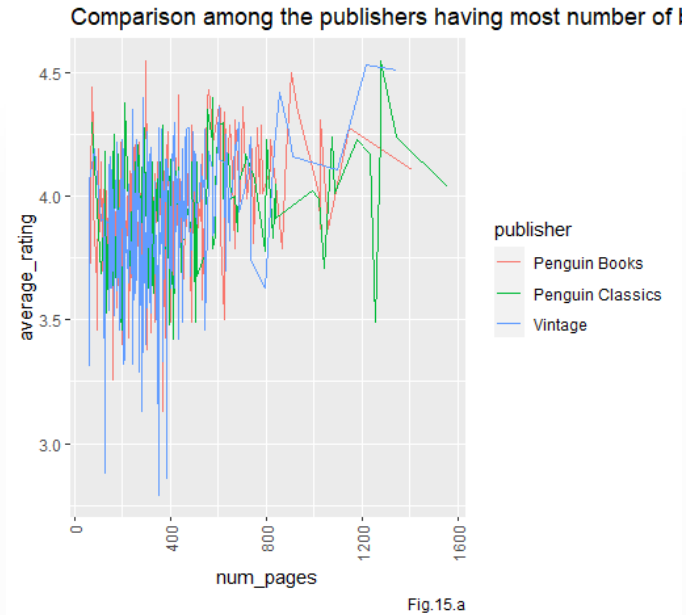
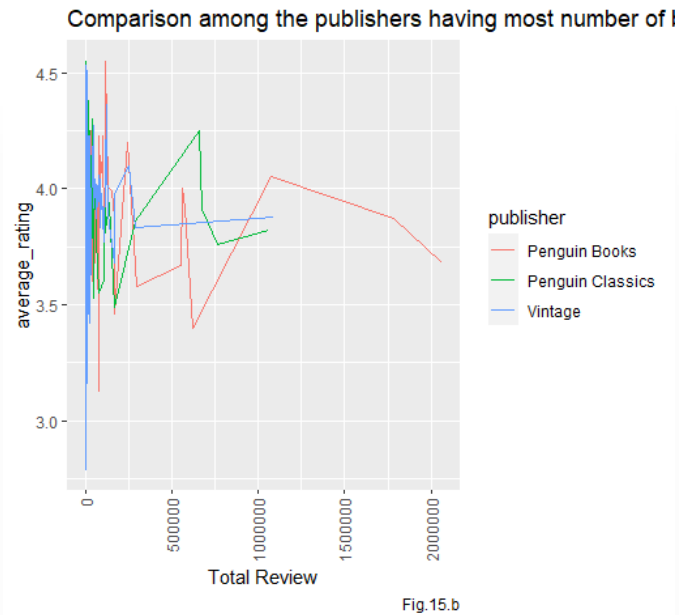


Fig. 13.d

(14) Comparison between top three authors having most number of books

- Initially average rating decreases for more number of pages, but later on it increases for all 3 authors.
- For low number of reviews all the authors have a mixed average rating. But when total number of reviews increases, average rating decreased for J.R.R. Tolkien and William Shakespeare but increases for Stephen King.
- All three authors have more reviews in pages between 250 to 750.

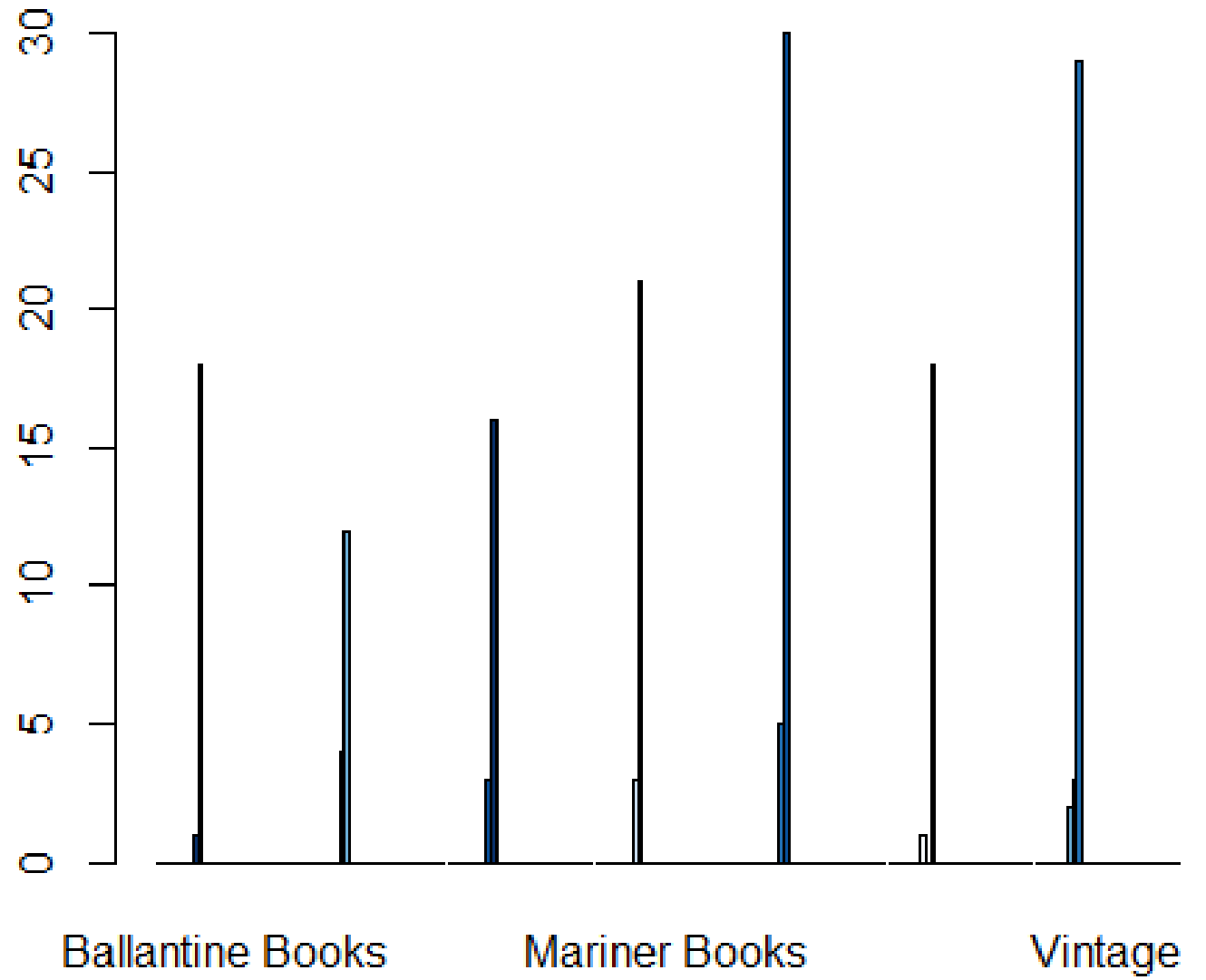




(15) Comparison between top three publishers

- For higher number of reviews, average ratings are below 4 for these publishers
- Average rating has not such pattern for different number of pages
- For higher number of pages, total reviews decreases for all publishers

(16) Books
published by
top publishers
in different
languages



Conclusion



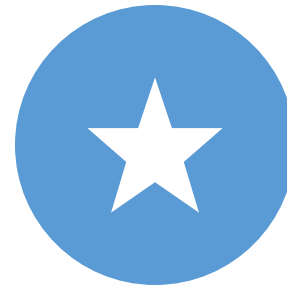
Good average ratings are for the books having page ranges 250-750



Number of reviews are mostly for the books written in english. Probably because of the reader of the language are more than others. Also most books in the dataset are written in this language.



Review index (ratio between text reviews and ratings count) are mostly below 0.25. So only a quarter of the reviewer gives a text review.



For higher review index, average rating decreases.