

Amazon Book Reviews

by Robert Huang

Advisor: Prof David Aldous

## Introduction

Amazon.com is the world's largest online retailer. Founded by Jeff Bezos in 1994 as an online bookstore, it has expanded its product line to include CDs, DVDs, electronics, apparel, and many other items.

Each item listed on Amazon has its own product page. One of the key features of the product page is the reader review. Anyone, whether he or she has purchased the item or not, is allowed to write a review for the product as long as he or she is a registered member of Amazon. The reviews include a rating from one star to five stars, with five being the highest, for the reviewed product. Along with this, visitors to the site have the ability to "review" the reviews themselves. Each user review comes with a "helpfulness" rating based on the number of people who answered yes or no to whether the particular review helpful was or not. So for example, one review might be such that "13 out of 15 people found this review helpful" or "0 out of 1 people found this review helpful." Bestselling books receive hundreds, some even thousands, of reader submitted reviews. Each review for a particular book might receive hundreds of "helpfulness" responses.

This is an exploratory data analysis project of books reviews, studying the relationships between the different variables associated with individual customer reviews. The original motivation for the project was whether early written reviews for a book get better feedback than later ones of the same quality. New questions that emerged include how do other factors affect the type of feedback a review receives? What effect do variables such as review rating and reviewer rank have on another?

## Data Collection

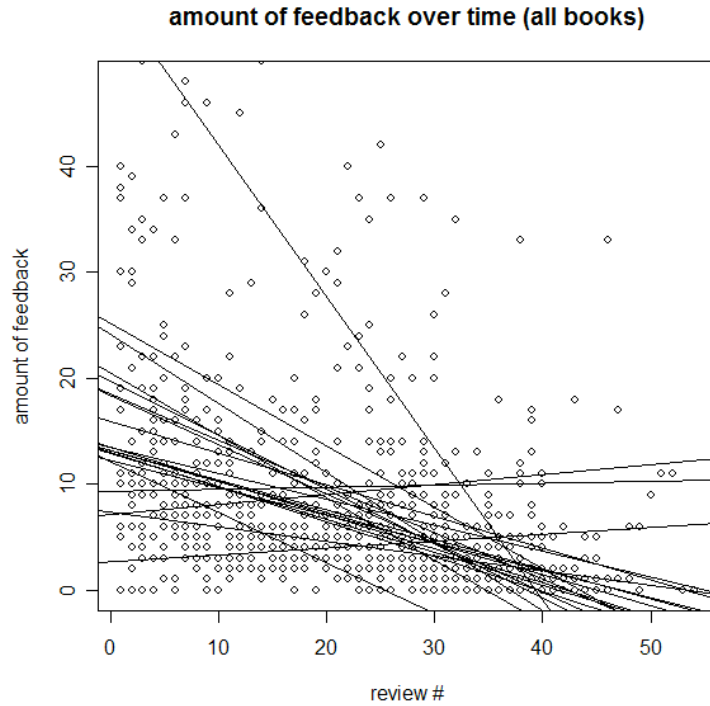
Data was collected for 20 books based on information on their product pages. Books with between 30 and 45 reviews were used. This provided enough data to study each book without being too much of a strain to collect information from the reviews. Also, books with these numbers of reviews had a good amount of feedback to look at. Most of the books chosen were released within the past two years. This was preferred as older book reviews sometimes did not contain all the available information that newer ones possessed. Many older reviews do not include any reviewer rank information and the amount of feedback for those reviews are inconsistent. The books chosen were a mixture of fiction and non-fiction on a variety of subjects. The following was collected for each book:

- the date the review was posted
- the star rating the review gave the book
- the amount of feedback the review received and how much of it was positive

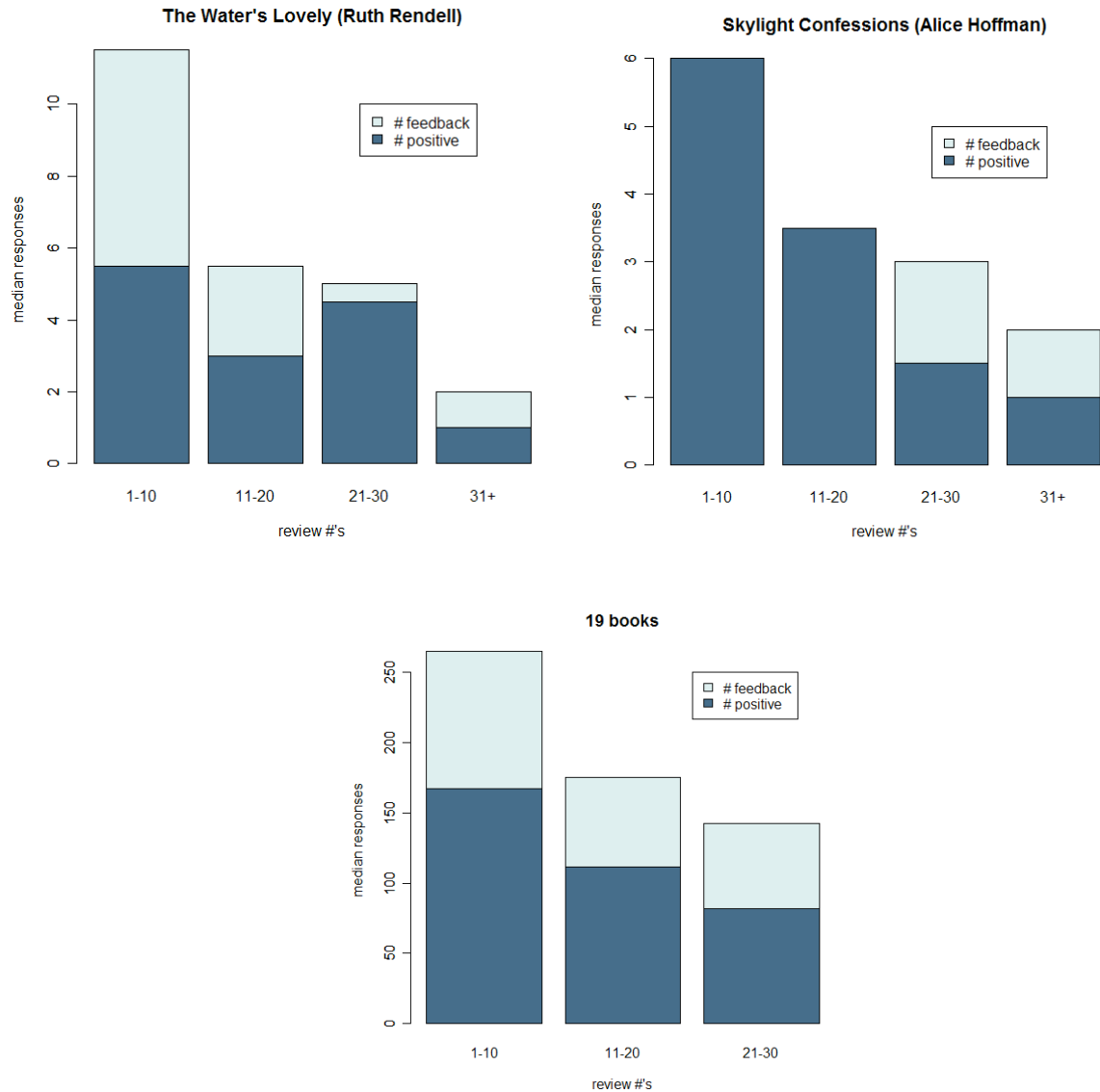
- the length of the review (number of words)
- the reviewer's rank
- two numbers attempting to quantify the quality of the review.

The reviewer rank is based on a rank Amazon gives to each registered member of the site. Amazon ranks every reviewer from one to about three million with one being the highest. This rank, according to the Amazon, is meant to signify who the “leading customer review writers” are and may be interpreted to mean who the best reviewers are. Thus there are labels such as “Top 10 Reviewer” or “Top 500 Reviewer” for reviewers with certain ranks. How this rank is determined is kept secret by Amazon but it is based on the “helpfulness” ratings of the reviewer's reviews.

The two numbers attempting to quantify the quality of each review range from one to three with three being the highest. Every review for every book was read and given two ratings. They attempt to gauge how informative the review is and how well it analyzes the book. For fiction, an informative review might describe the plot, characters, writing style, etc... of the book in good detail. A review getting high marks in analysis might analyze the themes of the book heavily or relate the book to something else that is relevant. So for example, a review that is very informative but provides little analysis will receive a three for “informative” and a one for “analysis” while another that is both uninformative and provides little analysis will receive ones in both “informative” and “analysis.” Receiving a one in one of the categories means little to no information/analysis. A two means some information/analysis and a three means anything more than that. This scale was chosen to keep things as simple as possible without going into too much analysis of the reviews themselves. These two measures are subjective, but something had to be done to somehow quantify the quality of the review in order to do an analysis of review quality.



In the above plot, for each book, the amount of feedback received for the  $n$ th review of each book is plotted along with least square lines fitting feedback and review number. There is a negative relation between the two variables in almost every case. It is an obvious result that earlier posted reviews receive more feedback, positive or not, than later posted reviews as older reviews have been on the page longer. Also, those reviews were likely on the product page close to the book's initial release date. One can assume that the page generates more visits at the product release time than anytime afterwards.



The above plots show the median amount of total (light blue) and positive (dark blue) feedback for each group of 10 successive reviews for two specific books and then for 19 of the 20 books (only the first 30 reviews were considered in this plot). While the declining feedback relationship is still apparent, there is also a large drop-off in the amount of feedback between the first ten reviews and the next ten before a steadier decline. Because the amount of feedback declines over time, so does the amount of positive feedback. Thus the amount of positive feedback for each successive group of 10 reviews cannot be compared absolutely as early reviews always receive more feedback.

One way to compare the amount of positive feedback over time is to use percentages. Using the bar plot for all 19 books from above, the amount positive divided by total feedback is 0.631,

0.637, and 0.575 for the first group of ten, second group of ten, and third group of ten reviews respectively. There is little difference between the percentage of positive feedback between the first two groups of reviews and a slight decline in the third. It appears that early reviews are better received with this analysis but there is not a great enough difference in the first two groups of successive reviews to say too much. Doing this same analysis for books individually leads to misleading results because early reviews receive so many more responses than later reviews. Most of the later reviews receive 0 or 1 responses making them hard to compare with the other reviews.

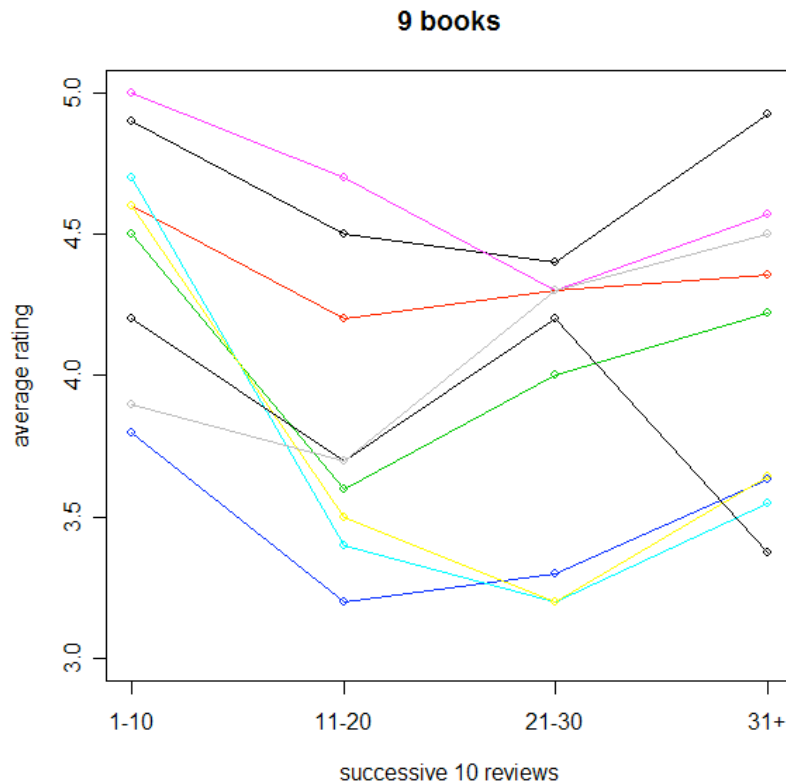
Another way to look at this is to plot the amount of positive feedback and negative feedback over time against review number. Then best fit lines are added to each of these two sets of data and the slopes compared. If the line fitting positive feedback has greater (absolute) slope than the line fitting total feedback, then earlier reviews are getting a better proportion of positive feedback as compared to later reviews. Before fitting the data for each book using this method, certain reviews were removed. Those reviews giving ratings of two stars or less were eliminated. Reviews giving low ratings are almost always negatively reacted to (as will be seen later) and distort the data. Also only reviews getting at least five feedback responses were kept. This removes the reviews receiving things such as 0 out of 0 or 0 out of 1 positive feedback responses which give little info and throw the fits of the lines off. The following is a table of the slopes of the best fit lines for each book:

<u>book</u>	<u>feedback</u> <u>slope</u>	<u>positive</u> <u>slope</u>
Liu	-0.087	-0.039
Yang	-2.734	-3.037
Harris	-0.1344	-0.0464
Stein	1.445	1.273
Dorsey	1.61	1.82
Chang	-1.27	-0.2488
Palahniuk	-0.0208	-0.3231
Gates	-0.9456	-0.5053
Anderesen	-1.423	-1.133
Barkley	-1.388	-0.7238
Feinstein	-3	-3.571
Lagnado	-0.711	-0.7988
Dillard	-1.362	-1.208
Galeano	-5.1	-4.7
Kriegel	-0.7668	-0.8978
Sofer	-0.7631	-0.6969
Hoffman	0.1	0.0818
Shriver	-1.088	-1.093
Smiley	-0.833	-0.3165
Rendell	-0.6605	-0.5697

mean	-0.9566	-0.8367
std dev	1.4353	1.4534

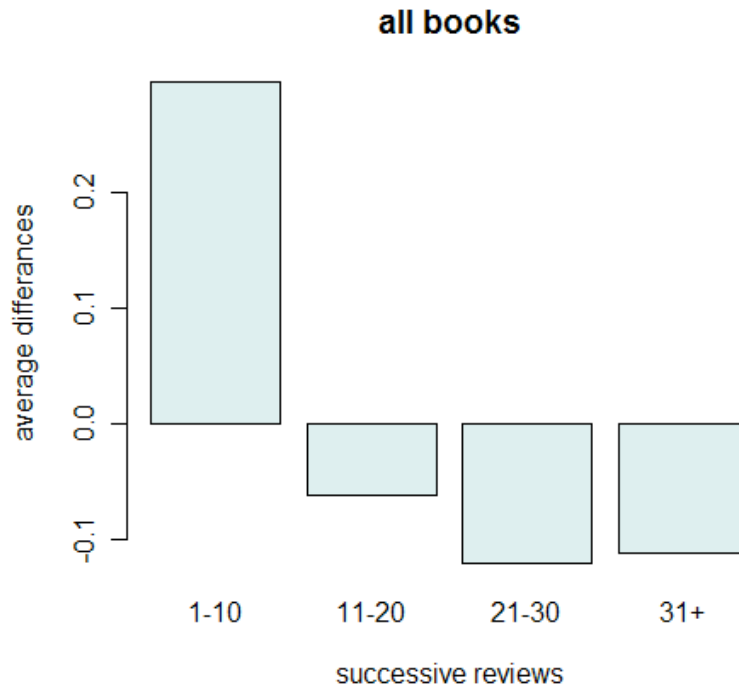
The mean (absolute) feedback slope is greater than the mean positive slope but in 7 of the 20 cases, the (absolute) positive slope was greater. Given this information, it is difficult to say that time has an effect on how well the reviews are received. As stated before, this analysis removed all reviews generating less than 5 feedback responses. Reviews receiving 0 responses do have meaning as a review getting 0 positive out of 0 responses can be interpreted to be worse than a review getting 50 positive out of 60 responses, but it is difficult to gauge what and how to effectively include them to do a proper analysis.

Looking at the average rating given by reviewers over time provides an interesting result. The following is a plot of the average ratings given by each group of 10 successive reviews for 8 of the books in the sample.



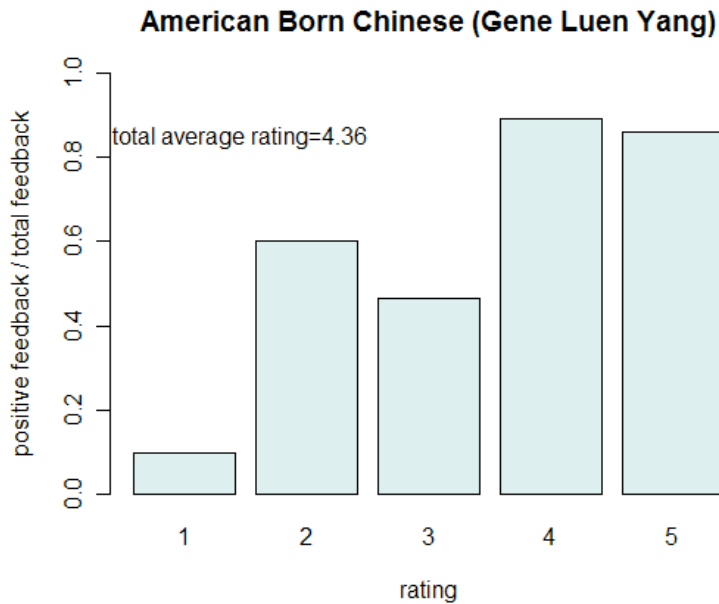
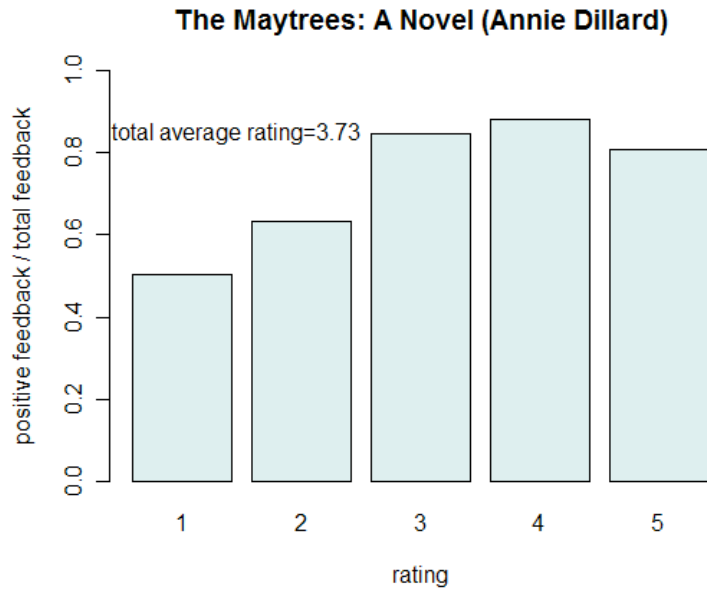
Most follow the same pattern with the first 10 reviews giving higher ratings than the second. The third group of 10 gives similar ratings to the previous ten and the last group give higher ratings than the previous group. The effect between the first 10 and second 10 groups of successive reviews is not unique for just this group of books. Taking the differences between the

average rating given by each group of 10 reviews with the total average rating overall given to the book for each book and then finding the average differences for all of the this leads to the following graph:



The first ten reviews on average give ratings that are about 0.3 points higher than the average rating received by the book. The next 3 groups of ten reviews give ratings lower than average, with the third and fourth groups giving a lot lower scores than the second ten. Why is it that the first ten reviewers give ratings so much higher than future overall rating of the book? One guess is that fans of the authors are more likely to rush and get in their review of the book first, giving the book a higher and more biased rating than it deserves. Maybe publishers even solicit people to give positive reviews to their book, hoping that the early positive responses leave the book with a better first impression for potential customers. Another guess is that reviewers are wary of giving the book a low rating until they see another bad review for fear of a backlash against them. All of this is just speculation though.

Breaking up the reviews by star rating provides another insight. The following two plots break up all reviews into 5 categories, those giving the book a rating of 5, 4, 3, 2, and 1 for specific books. They plot the proportion of positive responses over total responses for each of those 5 groups of reviews.



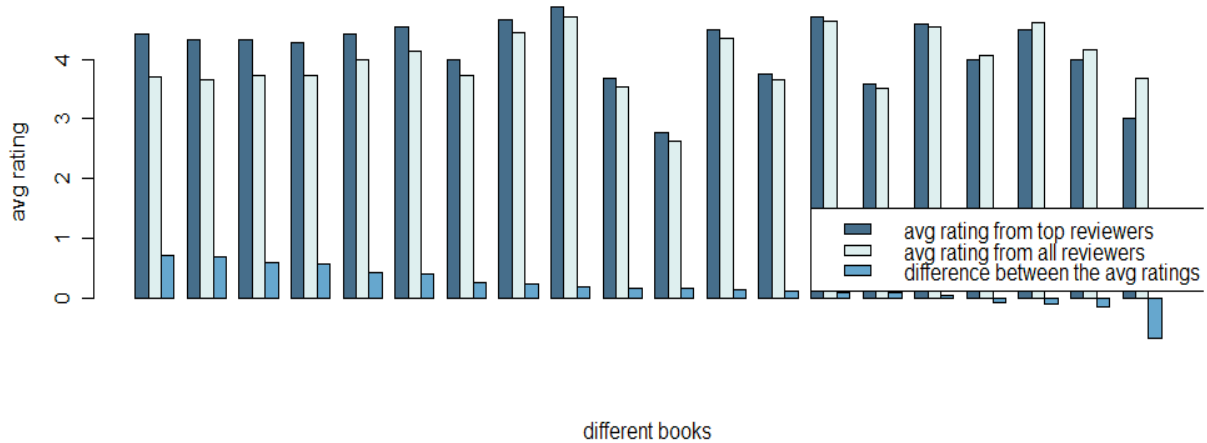
Reviews giving the books ratings closer to the books' actual ratings get better feedback responses from Amazon visitors. In the plot for *The Maytrees*, the book's total average star rating is a 3.73 and reviews giving the book 4 have received the best proportion of positive responses. In the plot for *American Born Chinese*, we see the same result. The book's average rating is a 4.36 and the reviews giving the book 4 stars received the best proportion of positive responses. This supports the argument that Amazon visitors give feedback on reviews based on how much they agree with them if the book's average star rating is taken to mean the public's

general opinion on the book. If the reviews for all 20 books are broken up the same way and the number of positive and total feedback responses summed, the following graph is produced:

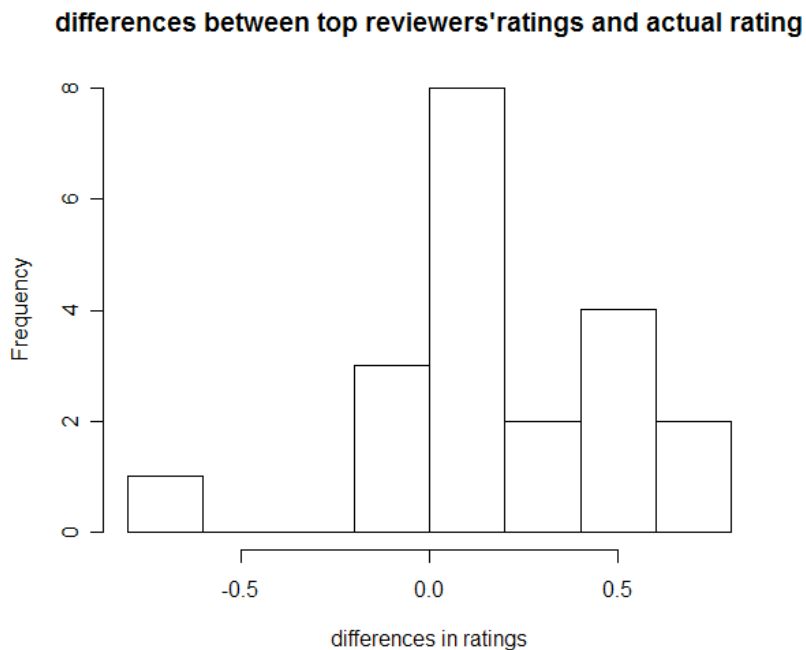


The effect is not as pronounced as when looking at the two previous graphs for individual books, but reviews giving 4 stars, that closest to the total average rating (found by taking the average rating of all 20 books), received a good number of positive responses. We see another major effect in this graph. Positive reviews generate the best feedback as is the case with 4's and 5's. Reviews giving ratings of 1 star receive high numbers of negative responses, especially in comparison with the reviews giving higher ratings. In fact, fewer than 40 percent of reviews giving 1 star get a positive response. For whatever reason, Amazon visitors are less likely to accept negative reviews. One possible explanation for this is that fans of the books visit the book's product page and express their dislike of low ratings on their favorite books. This might be a reason why reviewers do not give bad reviews early. Adding this effect with the fact that early reviews generate more responses, this no doubt hurts the rankings of reviewers giving negative reviews.

What effect does the previous result have on highly ranked reviewers? Top reviewers are defined to be those reviewers with ranks of 10,000 or higher. The following charts the average rating given by top reviewers for each of the 20 books along with the books' total average rating and the difference between the two ratings.

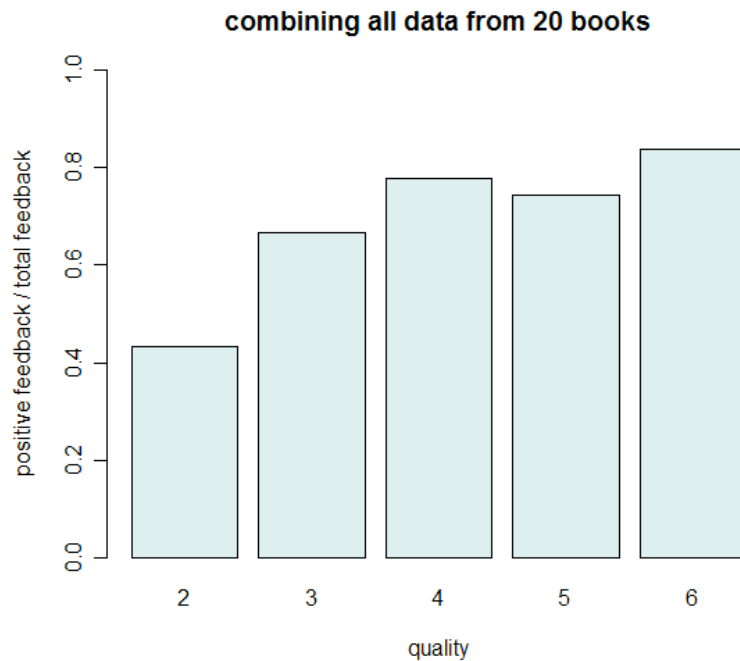


Top reviewers give higher ratings compared to the average rating in 16 of the 20 cases. In fact, top reviewers give .187 stars higher on average compared to the book's total rating. The following is a histogram of the differences reiterating this fact:



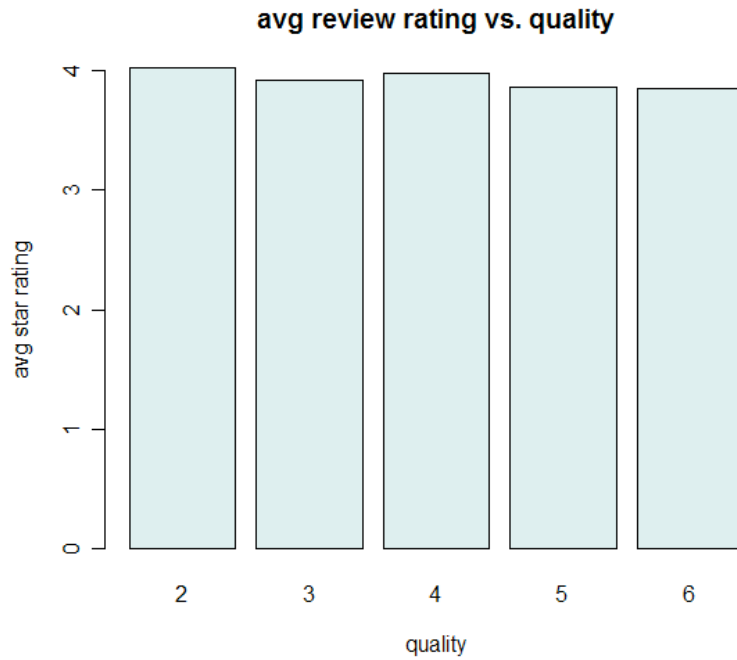
Is this a result of top reviewers knowing that giving high ratings result in better feedback? Do they use this fact to make sure they stay top reviewers? Or is it that those who review often and generally give high scores end up being top reviewers? How much weight can be placed on the reviews from top reviewers if they give certain ratings just to solicit positive responses? Are these reviews truly accurate?

The two quality ratings (described in data collection) are summed to get single number for quality (ranging from 2 to 6) to see if quality has any effect on the percentage of positive feedback received.



In general, better-written reviews get better feedback. Those getting 6's for quality (3 for informative and 3 for analytical) get the best responses while those getting 2's (1 for informative and 1 for analytical) get the worse responses. If this is indeed the case, then Amazon visitors recognize when reviews are well written in general and try not to let other factors affect their decisions on giving feedback.

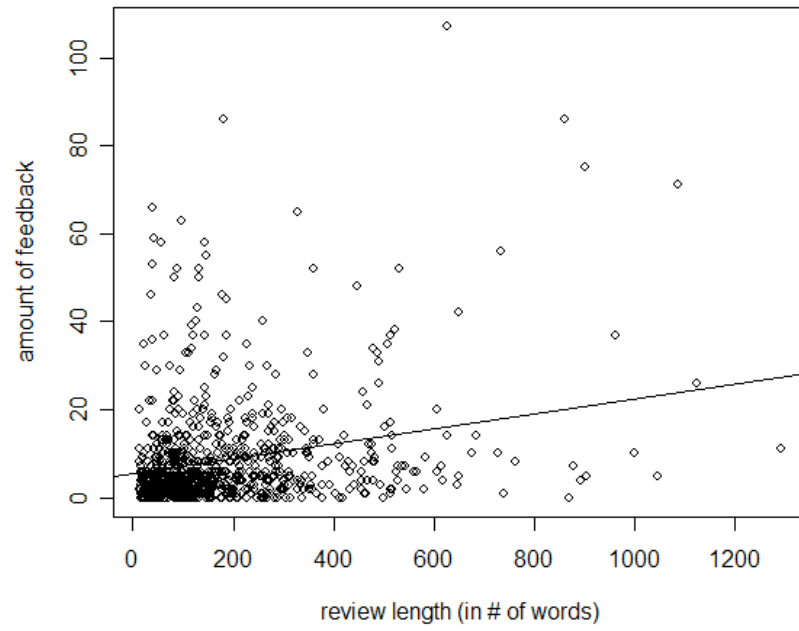
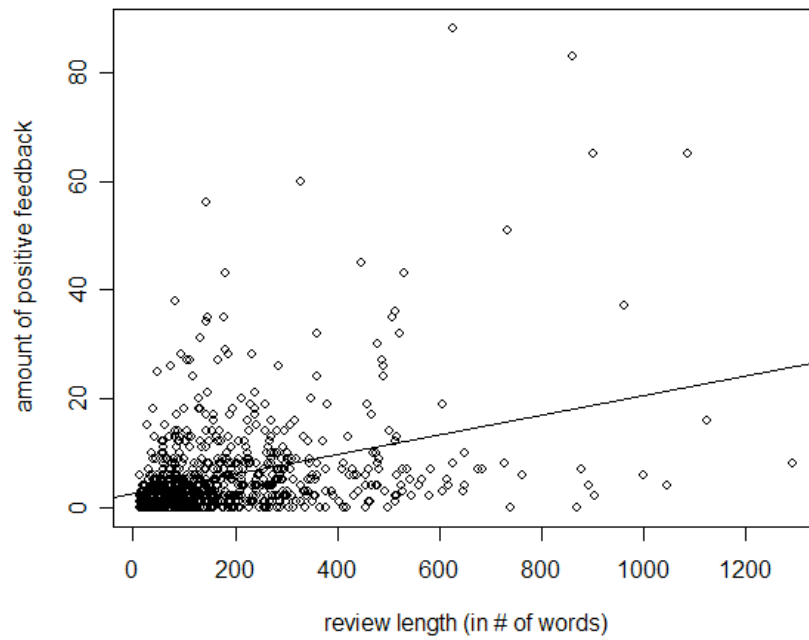
The next graph plots the average ratings found by separating the reviews of different quality as in the previous graph.



There appears to be little difference in the star ratings given by reviews of different quality. Thus well written reviews are not more critical than other reviews on average and badly written reviews are not always negative attacks on a book. There should not be too much weight placed on these results using the quality factor as the quality ratings were given arbitrarily and some bias might have slipped through when assigning ratings.

Other interesting graphs:

The length of the review (word count) correlates positively with the amount of total and positive feedback generated

**review length vs. amount of total feedback****review length vs. amount of positive feedback**

## Conclusion

The original question of whether early written reviews receive better feedback than later written reviews has not been answered clearly. There are conflicting results on which case it actually is with one result suggesting that early reviews (the first 20 as compared with reviews 21-30) get slightly better responses than later ones and another providing not much evidence for either case. But the evidence is not strong enough in either result to truly say anything. The main problem found in trying to answer that question was what to do with reviews generating no or just 1 response. Is getting 0 positive out of 1 response worse than getting 5 positive out of 20? If using percents, it is. Using correlations leads to the same result as reviews getting 1 positive out of 1 response fit better than reviews getting 40 out of 50 positive. In the analysis of positive feedback over time, the reviews generating few responses are removed. Most of these were the later reviews, so a lot of the data left over were just the early reviews. If these reviews were modified somehow and placed back for use in analysis, the results may certainly be different.

Despite this failure, other notable results were found. Reliability became a big issue. Early reviews may not always be reliable as they almost always overrate the book. How trustworthy the positive response data is came into question as reviews giving ratings close to the book's average star rating get better responses. Giving a negative review (unless the book has a low star rating) does the reviewer no favors. The dependency of the so-called top reviews was also questioned as top reviewers were always more generous than others in giving out high scores. Learning more about top reviewers specifically may make an interesting study. Maybe figuring out the algorithm that Amazon uses to determine top reviewers and seeing what type of writing or the frequency of writing reviews make a top reviewer. Perhaps more could be done with the quality rating as this project used that information sparingly because of the fact that they were obtained arbitrarily.

## Books used in project:

<b>Title</b>	<b>Author</b>	<b>Release Date</b>
The Accidental Asian: Notes of a Native Speaker	Eric Liu	5/12/1998
American Born Chinese	Gene Luen Yang	9/5/2006
The Becoming (The Anna Strong Chronicles, Book 1)	Jeanne C. Stein	11/28/2006
The Big Bamboo	Tim Dorsey	3/28/2006
The Chinese in America: A Narrative History	Iris Chang	3/31/2003
Fugitives and Refugees: A Walk in Portland, Oregon	Chuck Palahniuk	7/8/2003
The Gravedigger's Daughter	Joyce Carol Oates	6/1/2007
Heyday: A Novel	Kurt Andersen	3/6/2007
I May Be Wrong, But I Doubt It	Charles Barkley	10/1/2002
An Ice Cold Grave	Charlaine Harris	9/25/2007
Last Dance: Behind the Scenes at the Final Four	John Feinstein	2/7/2006
The Man in the White Sharkskin Suit	Lucette Lagnado	6/26/2007
The Maytrees	Annie Dillard	6/7/2007
Pistol: The Life of Pete Maravich	Mark Kriegel	2/5/2008
The Post-Birthday World	Lionel Shriver	3/13/2007
Septembers of Shiraz	Dalia Sofer	7/24/2007
Skylight Confessions	Alice Hoffman	1/11/2007
Soccer in Sun and Shadow	Eduardo Galeano	May-98
Ten Days in the Hills	Jane Smiley	2/13/2007
The Water's Lovely	Ruth Rendell	11/14/2006