Amazon Review Analysis - Airpods and Alternative

Chae Yeon Ryu

1 Summary

The results of the analysis provide some insights as to how producers can improve their products and target specific customers who will purchase the authentic Airpods over alternative earbuds and vice versa. First of all, a lot of customers were found to have compared their products with some other well-known earphones in their reviews, indicating that they are knowledgeable about cutting-edge technologies in earphone industry. In Sentiment Analysis, alternative earbuds customers displayed more fear, surprise, and stronger hatred than Airpods customers. These feelings may be associated, and these may have been caused by prevalent issues with battery and charging unlike Airpods. However, customers appreciate the fact that alternative earbuds have similar multi-functionalities as Airpods with much cheaper price. On the other hand, customers seem to dislike customer service and policies in both products. The ML models built on these reviews may be helpful in predicting the review type (positive or critical) of other earphones to some extent, but in a restricted context.

2 Motivation

In this project, I decided to analyze Amazon reviews on two products: Apple Airpods and alternative earbuds made after Airpods. The motivational question is whether a consumer behavior can be traced by text analysis. Specifically, this question is subdivided into: How do customers comment on products when they score higher/lower? Do customers compare alternative earbuds with authentic Airpods? If so, how do they compare them? Do customers buy/review on alternative earbuds knowing that it's an Airpods imitation? What features of earphones do customers value? Are those features same across two products? What makes certain customers buy an Airpods/alternative products over another? Lastly, can we predict whether a random earphones review is positive or critical based on the model built from these Amazon reviews?

3 Data

The initial data were obtained from the Amazon reviews on 2 products: "Apple Airpods (2nd Generation)" and "Wireless Earbuds, Bluetooth Earbuds Deep Bass Hi-Fi Stereo 30H Cycle Playtime in-Ear Wireless Headphones Sweatproof Earphones Sport Headsets Built-in Mic for Workout/Driving/Home Office", which is considered to be a cheaper alternative to Airpods with a very similar appearance and options. In the analysis, Apple Airpods's acronym "App" and the alternative product's acronym "Alt" will be used in figures and tables.

For each product, 50 reviews from Top Positive Reviews and 50 reviews from Top Critical Reviews were obtained by extracting HTML scripts. There are 200 reviews total, 100 for Apple Airpods (App) and 100 for Wireless Earbuds (Alt). Typically, top Positive Reviews include reviews with rating 4-5 and Critical Reviews with rating 1-3. From now on, positive reviews will be represented as "good", and critical reviews will be represented as "bad" for simplicity. The distribution of length of reviews by rating and product type

(App vs. Alt) is shown in Figure 1. Reviews were tokenized by splitting on the space character, removing unnecessary punctuation, lowering capital letters, and excluding stopwords, presented in Table 1.

3.1 Exploratory Data Analysis

Figure 1 shows that the reviews for Apple Airpods tend to be longer than those for alternative earbuds throughout all ratings from 1 to 5. It might be due to the fact that there are many Apple's fans wanting to share/know its products in detail through more descriptive reviews. In addition, at rating 3 and 4, there are two abnormal outliers for Airpods. Overall, positive reviews or very critical reviews appear to be longer than reviews of average rating.

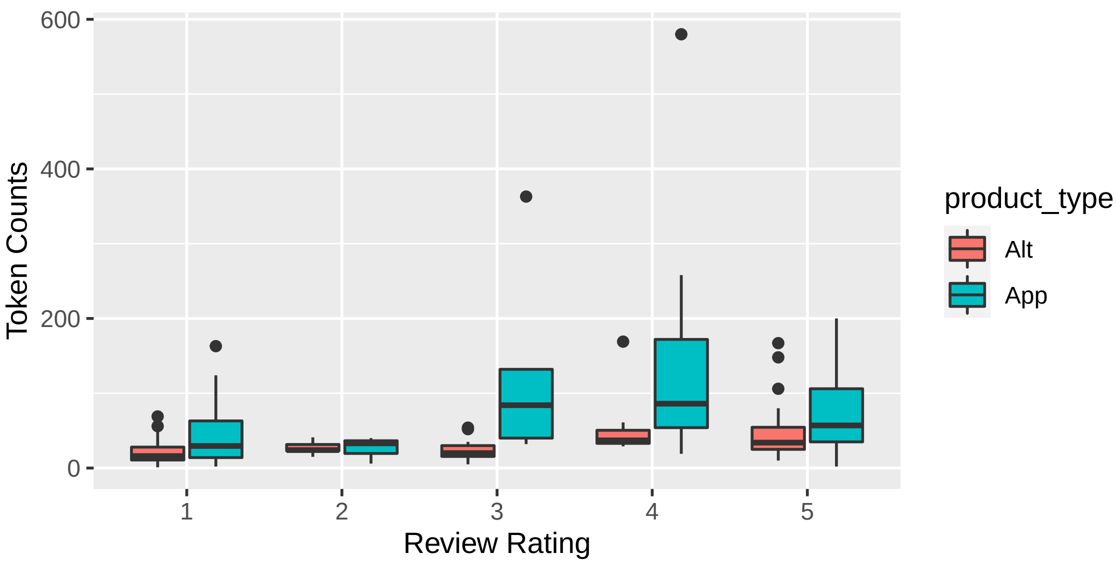

Figure 1: Review Length by Rating and Product Type

Table 1 further corroborates that the Airpods reviews are generally longer than those for alternative earbuds. The total counts of tokens for Airpods in positive (good) reviews and critical (bad) reviews exceed those for alternative earbuds.

Table 1: Review Subcorpora on AirPods and Alternative Earbuds

	App_good	App_bad	Alt_good	Alt_bad
Words Counts (tokens)	4370	2494	2354	1138
Number of Reviews	50	50	50	50

In addition, Chi-square test was conducted to confirm if Apple Airpods users tend to write longer reviews and whether it's related to rating. Then, frequency tables were obtained to see keywords specific to a certain category of reviews, guiding further analysis.

Based on Chi-square test, there is insufficient statistical evidence that the relationships between length of reviews, review rating and product type are significant: p-value of product type vs. length of reviews was 0.2733561 and that of review rating vs. length of reviews was 0.4024940.

Tokens from four subcorpora are represented in a simplified frequency table 2. What is first noticeable is that from the alternative earbuds reviews, "apple" or "airpods" were frequently mentioned, implying that some customers are aware of the similarities between two products.

Table 2: Most Frequent Tokens (descending order)

App_good	App_bad	Alt_good	Alt_bad
airpods	apple	sound	case
just	get	one	charging
use	like	apple	airpods
one	just	case	amazon
sound	new	pair	get
earbuds	sound	quality	great
good	one	apple	ear
price	use	sound	battery
get	quality	earbuds	pair
just	bought	ear	charge

4 Method

The following methods will be used for analysis. First, sentiment analysis with syuzhet and nrc will be conducted to see how customers respond to products and how their reactions are different between two products and between positive/critical. Common sentiment words will be obtained to measure the customer dissatisfaction.
Furthermore, a bigram network will be visualized using a Markov chain to see the co-occurrence of certain tokens in pairs. This will help explain the topics discussed in each type of reviews.
Along with bigram network analysis, collocational networks will be plotted for some meaningful tokens in order to see their implication in the review. As suggested from the frequency table, network analysis will be conducted on alternative earbuds reviews to see what customers like/dislike about this product compared to Apple Airpods.
Lastly, machine learning models with 6 different algorithms (Naive Bayes, Bagging, Boosting, RF, SVM, Tree) will be built based on 200 reviews with the classifier "Review Rating" - positive or critical. They will be used to predict the rating of other earphones' reviews and the accuracy will be measured to see if these models are effective and generalizable for other earphone reviews.

5 Results

5.1 Sentiment Analysis

Figure 2 shows sentiment counts in reviews categorized in 4 sections - Airpods in positive and critical reviews and alternative earbuds in positive and critical reviews. All graphs show a similar trend across sentiments. Quantitatively, almost an equal proportion of negative sentiments is found in both Airpods and alternative earbuds reviews, showing that the degree of dissatisfaction may be similar across all customers regardless of product type.

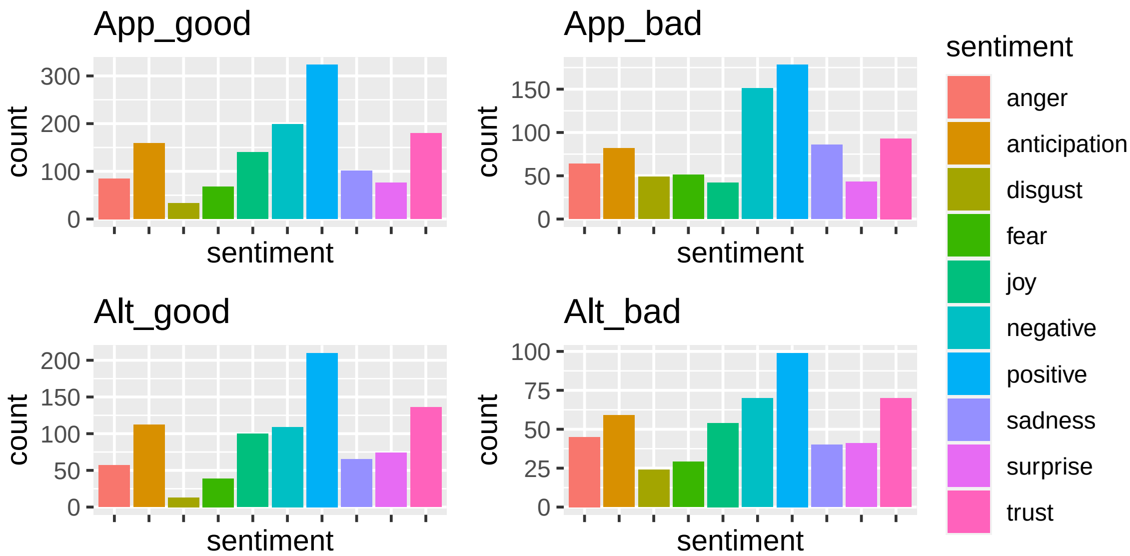

Figure 2: Review Sentiment

However, words in alternative earbuds reviews imply stronger negative connotations, as seen from Figure 3 displaying top words contributing to negative sentiment. Hence, from the qualitative perspective, alternative earbuds customers may be more upset than Airpods customers. In addition, comparing reviews on both products, alternative earbuds display more fear and surprise proportionally. This seems to imply that customers found more unexpected features from alternative earbuds as opposed to well-known features of Airpods.

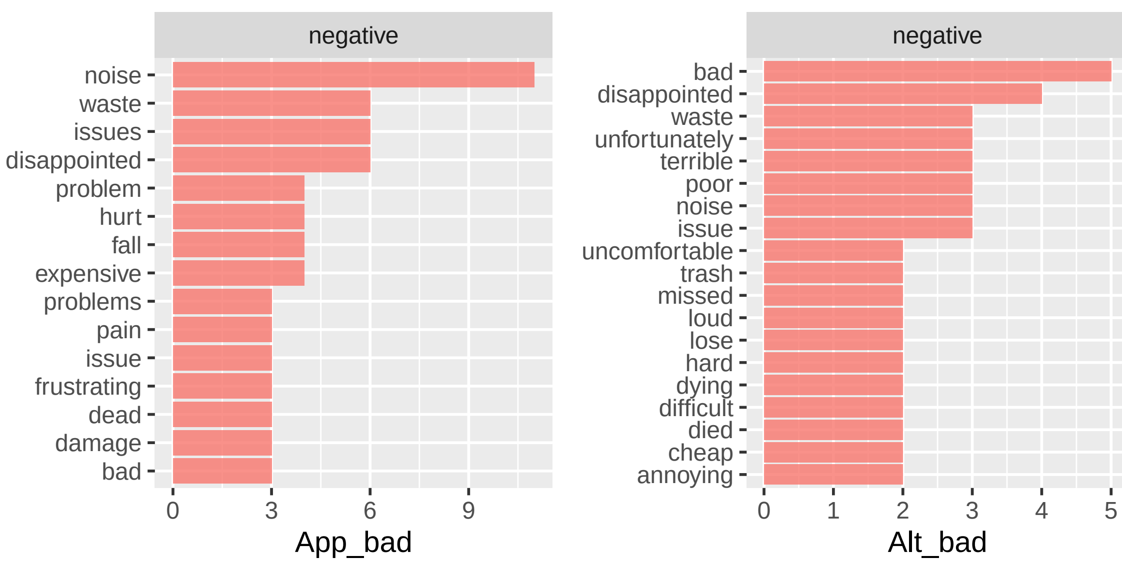

Figure 3: Contribution to Sentiment

5.2 Bigram Analysis

Figure 4 displays a network of bigrams of 4 review types: positive Airpods reviews (App_good), critical Airpods reviews (App_bad), positive alternative reviews (Alt_good), critical alternative reviews (Alt_bad).

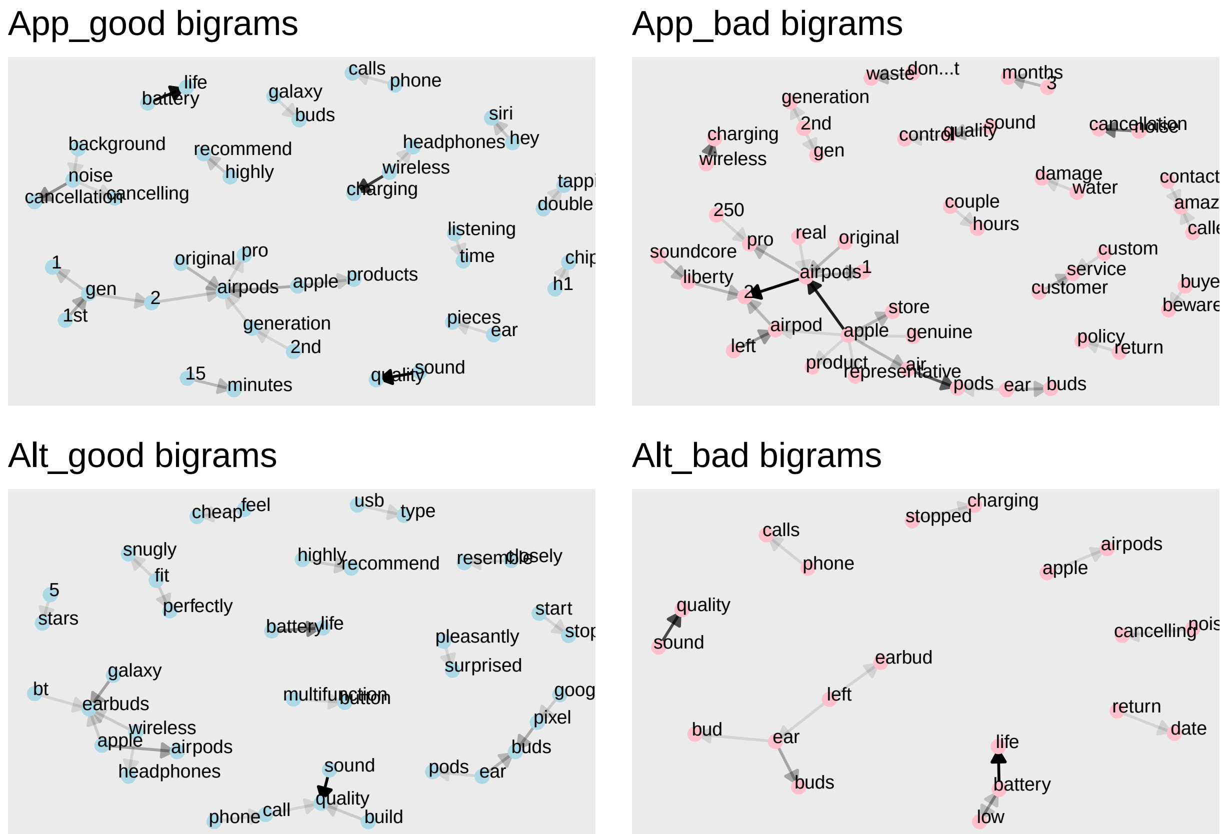

Figure 4: Bigram Network

From App_good bigrams, we can see that many positive Airpods reviews include comparisons between other Airpods generations (i.e. 1st, 2nd, pro, air) as well as Galaxy buds from Samsung, but no other cheaper alternatives. In addition, they discuss Airpods' functionality such as noise cancellation, double tapping, wireless charging, and siri - these seem to be the reasons that customers were satisfied with Airpods.
These options appear again in App_bad bigrams, but in addition to those, “customer service”, “damage water”, “return policy”, and “250” appear in App_bad bigram. Hence, it seems as customers were dissatisfied by customer policies (return policies, request service, etc.), no water resistance, and the high price - $250 used to be an approximate price of Airpods when first released.

From Alt_good bigrams, we can see many comparisons with well-known earphones such as Airpods or Galaxy buds. In addition, customers seem to like the functionality, fit, and the price as “multifunctional button”, “fit perfectly”, “wireless”, and “cheap” appear. On the other hand, in Alt_bad bigrams, “return date”, “low battery”, “battery life”, and “stop charging” are found, implying that customers were dissatisfied by customer policies (return policies) and battery/charging issues.

In both bigrams for alternative earbuds, either Apple or Airpods appears, and it shows that customers buying cheaper, alternative earbuds frequently compare this product to Airpods. Furthermore, unlike Airpods which doesn't seem to have many issues with battery or charging, such issues are frequently brought up in alternative earbuds reviews.

5.3 Collocational Analysis

In order to see how cheap alternative earbuds are compared to Airpods and how many customers appreciate it, collocational netowrks are plotted.

Figure 5 displays collocational networks for “airpods” and “price” on alternative earbuds positive reviews. In the network, “1/4”, “1/3”, “cheap”, “money”, and “low” appear as collocating with “airpods” on the alternative earbuds reviews. “1/4”, “1/3” would mean 1/3 or 1/4 of the Airpods price. Then, it seems as many customers are satisfied with the lower price even after they tried for music quality, etc. Furthermore, “alternative” , “similar” and “comparable” associated with Airpods suggest that this alternative earbuds may be as good as authentic Airpods with much cheaper price for many users. This result is corroborated by previous bigram analysis where we found that part of customer satisfaction is driven by price level.

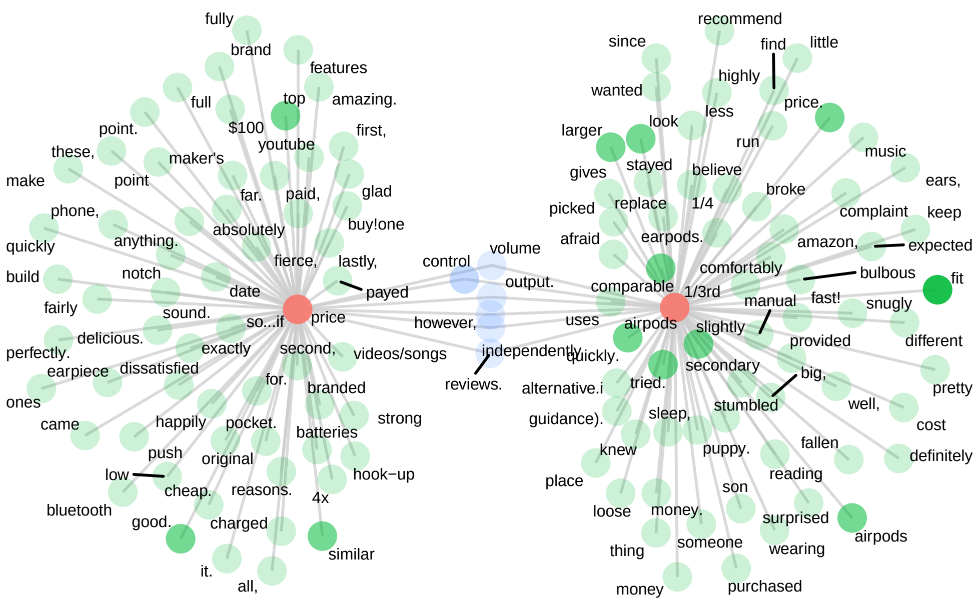

Figure 5: Collocational Networks with ‘airpods’ and ‘price’ from Alternative Earbuds Reviews

5.4 ML Predictive Modeling

Given 200 reviews from Airpods and alternative earbuds, Six ML models are built using algorithms for ensemble classification: Naive Bayes, Bagging, Boosting, RF, SVM, and Tree. They were tested against 20 test reviews from a different product: Beats Studio Buds. 10 were positive reviews and 10 were critical reviews, and the accuracy and predictions are shown below.

Table 3: Predictions Accuracy

Bayes	Bagging	Boosting	RF	SVM	Tree
0.5	0.65	0.55	0.6	0.65	0.55

Table 4: Predictions Comparison

Actual	Bayes	Bagging	Boosting	RF	SVM	Tree
bad	bad	bad	bad	bad	bad	bad
bad	bad	good	good	good	good	bad
good	bad	good	good	good	good	good
good	bad	good	good	good	good	good
good	bad	good	good	good	good	good
good	bad	good	good	good	good	bad
bad	bad	good	good	good	good	good
good	good	good	good	good	good	good
bad	bad	good	good	good	good	good
bad	bad	bad	bad	bad	bad	good
bad	good	good	bad	good	bad	good
bad	bad	good	good	good	good	bad
good	bad	good	good	bad	good	bad
bad	bad	good	good	good	good	good
good	bad	good	bad	good	bad	good
bad	bad	good	good	good	good	good
good	bad	good	good	good	good	good
good	bad	good	bad	good	good	bad
good	bad	good	bad	good	good	good
bad	bad	bad	bad	bad	bad	bad

Model predictions accuracy ranges from 50% to 65%, indicating that they will get about half the review types right. Their prediction accuracy is similar across all algorithms, indicating that no particular algorithm is especially effective in predicting the type of reviews for earphones. From Table 4, it is observed that Bayes model tends to be strict on good reviews: a lot of positive (good) reviews were predicted as critical (bad). Bagging model is the opposite: a lot of critical (bad) reviews were predicted as positive (good). These models may be useful if we want to know which reviews are distinctively positive or critical.

6 Discussion

In conclusion, there were some similarities and differences between Airpods and alternative earbuds reviews. Both customers appreciated the sound quality, functionality and wireless earbuds while complaining about customer service and policies. As for differences, Airpods customers appreciated noise-cancelling, wireless charging and battery life, which were not/rarely mentioned in alternative earbuds reviews. On the other hand, more emotions like fear and surprise were detected along with hatred (which seems to be associated with non-tolerable defective features, possibly battery and charging) in alternative earbuds reviews. Alternative earbuds customers were mainly concerned about battery/charging issues while appreciating the cheaper price and almost equivalent functionality. Those customers were also knowledgeable about Airpods and included comparisons in their reviews. Hence, they knew how much cheaper alternative earbuds were compared to Airpods, as discovered in collocational analysis. This implies that price is one of the deciding factors in purchase. The ML models built on these reviews are moderately predictive and may be helpful in predicting the review type (positive or critical) of other earphones in a restricted context where we want to find particularly positive or negative reviews.

This project leaves room for further experimentation. First, based upon what customers value, I wish to explore more regrets through more complex ngrams (i.e."I wish I bought"). It will help us narrow down to more specific features making/changing purchase decision. In addition, I hope to build a more comprehensive training/testing model with higher accuracy. This will require experimentation with more data and various algorithms.

7 References

7.1 Data

- Apple Airpods reviews

https://www.amazon.com/Apple-AirPods-Charging-Latest-Model/product-reviews/B07PXGQC1Q/ref=cm_cr_dp_d_show_all_btm?ie=UTF8&reviewerType=all_reviews

- Wireless Earbuds reviews (alternative earbuds)

https://www.amazon.com/Bluetooth-Wireless-Headphones-Sweatproof-Earphones/product-reviews/B088ZTB1Z7/ref=cm_cr_dp_d_show_all_btm?ie=UTF8&reviewerType=all_reviews

- Beats Studio Budsreviews (test)

https://www.amazon.com/Beats-Studio-Cancelling-Earbuds-Built-Bluetooth-Headphones/product-reviews/B096SV8SJG/ref=cm_cr_dp_d_show_all_btm?ie=UTF8&reviewerType=all_reviews

7.2 Method

Introduction to Tidy Text Mining, http://rstudio-pubs-static.s3.amazonaws.com/355276_5845790bca154871b7b9507d34ca212f.html.

Jockers, Matthew. "Introduction to the Syuzhet Package." Introduction to the Syuzhet Package, 24 Nov. 2020, https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html.

"Text Mining: Sentiment Analysis." Text Mining: Sentiment Analysis · UC Business Analytics R Programming Guide, 2 Oct. 2018, https://uc-r.github.io/sentiment_analysis.

Wang, Cheng-Jun. "Sentiment Analysis with Machine Learning in R." DataScience+, 16 Jan. 2016, https://datascienceplus.com/sentiment-analysis-with-machine-learning-in-r/.

Yan, Qiushi. "Notes for 'Text Mining with R: A Tidy Approach.' " Bookdown, 11 May 2020, https://bookdown.org/Maxine/tidy-text-mining/.