Lab 4: Dictionary-Based Analysis

Srishti Saha (ss1078)

27 February, 2020

Exercise 1: Using the link here, load the review dataset into your workspace. We will need the columns containing which product the review is for as well as the review text, indicate which columns they are and create a new cleaned dataframe with only these two columns. Print the first few rows of data.

The column 'name' mentions the product the review is for along with other descriptions/specifications. However, there are a few missing values in this column which we might have to treat for. (Normally, ASINs are unique product IDs assigned by Amazon to a product- There are 41 unique ASIN combinations (along with 1 missing value.))
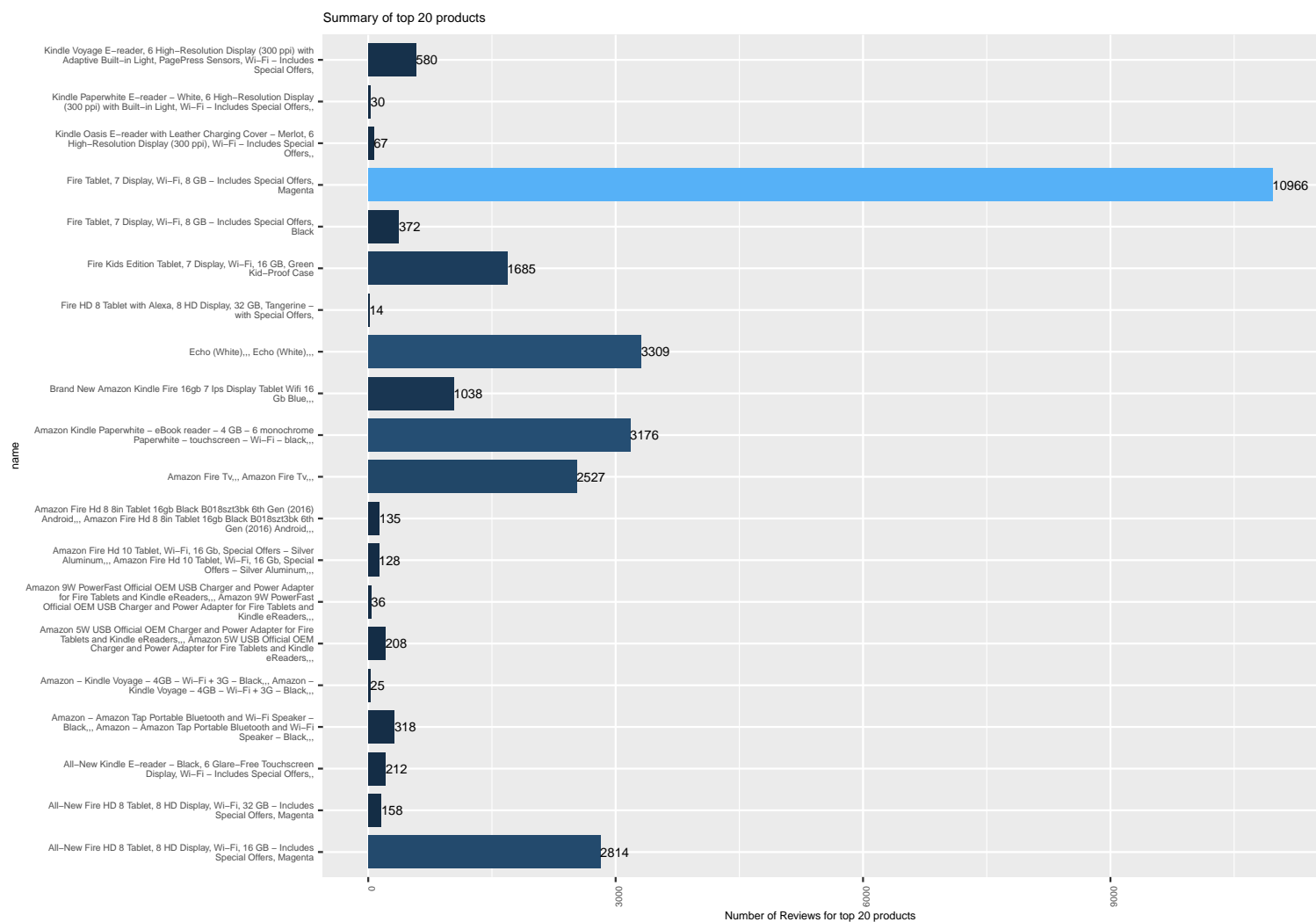
Table 1: Top 10 rows from Amazon product review data

name	reviews.text
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	This product so far has not disappointed. My children love to use it and I like the ability to monitor control what content they see with ease.
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	great for beginner or experienced person. Bought as a gift and she loves it
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	Inexpensive tablet for him to use and learn on, step up from the NABI. He was thrilled with it, learn how to Skype on it already...
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	I've had my Fire HD 8 two weeks now and I love it. This tablet is a great value.We are Prime Members and that is where this tablet SHINES. I love being able to easily access all of the Prime content as well as movies you can download and watch laterThis has a 1280/800 screen which has some really nice look to it its nice and crisp and very bright infact it is brighter then the ipad pro costing $900 base model. The build on this fire is INSANELY AWESOME running at only 7.7mm thick and the smooth glossy feel on the back it is really amazing to hold its like the futuristic tab in ur hands.
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	I bought this for my grand daughter when she comes over to visit. I set it up with her as the user, entered her age and name and now Amazon makes sure that she only accesses sites and content that are appropriate to her age. Simple to do and she loves the capabilities. I also bought and installed a 64gig SD card which gives this little tablet plenty of storage. For the price I think this tablet is best one out there. You can spend hundreds of dollars more for additional speed and capacity but when it comes to the basics this tablets does everything that most people will ever need at a fraction of the cost.
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	This amazon fire 8 inch tablet is the perfect size. I purchased it for my husband so that he has a bigger screen than just his phone. He had gotten me one a few years ago so I knew it would be a good purchase.
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	Great for e-reading on the go, nice and light weight, and for the price point given, definitely worth the purchase.
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	I gave this as a Christmas gift to my inlaws, husband and uncle. They loved it and how easy they are to use with fantastic features!
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	Great as a device to read books. I like that it links with my borrowed library e-books. Switched from another popular tablet brand and I am happy with the choice I made. It took some time to get books from my previous non-Kindle reader, but finally figured out a way!
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	I love ordering books and reading them with the reader.

Exercise 2: Summarize the data by outputting the products reviewed in this dataset as well as the number of reviews per product.

Table 2: Summary of Amazon product reviews data

name	nbr__of__reviews
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Magenta	10966
	6760
Echo (White),,, Echo (White),,,	3309
Amazon Kindle Paperwhite - eBook reader - 4 GB - 6 monochrome Paperwhite - touchscreen - Wi-Fi - black,,,	3176
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	2814
Amazon Fire Tv,,, Amazon Fire Tv,,,	2527
Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Green Kid-Proof Case	1685
Brand New Amazon Kindle Fire 16gb 7 Ips Display Tablet Wifi 16 Gb Blue,,,	1038
Kindle Voyage E-reader, 6 High-Resolution Display (300 ppi) with Adaptive Built-in Light, PagePress Sensors, Wi-Fi - Includes Special Offers,	580
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Black	372
Amazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker - Black,,, Amazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker - Black,,,	318
All-New Kindle E-reader - Black, 6 Glare-Free Touchscreen Display, Wi-Fi - Includes Special Offers,,	212
Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,, Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,	208
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta	158
Amazon Fire Hd 8 8in Tablet 16gb Black B018szt3bk 6th Gen (2016) Android,,, Amazon Fire Hd 8 8in Tablet 16gb Black B018szt3bk 6th Gen (2016) Android,,,	135
Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Aluminum,,, Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Aluminum,,,	128
Kindle Oasis E-reader with Leather Charging Cover - Merlot, 6 High-Resolution Display (300 ppi), Wi-Fi - Includes Special Offers,,	67
Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,, Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,	36
Kindle Paperwhite E-reader - White, 6 High-Resolution Display (300 ppi) with Built-in Light, Wi-Fi - Includes Special Offers,,	30
Amazon - Kindle Voyage - 4GB - Wi-Fi + 3G - Black,,, Amazon - Kindle Voyage - 4GB - Wi-Fi + 3G - Black,,,	25
Fire HD 8 Tablet with Alexa, 8 HD Display, 32 GB, Tangerine - with Special Offers,	14
Kindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs),,	9
Kindle Keyboard,,, Kindle Keyboard,,,	9
Kindle Paperwhite,,, Kindle Paperwhite,,,	9
Amazon Echo and Fire TV Power Adapter,,, Amazon Echo and Fire TV Power Adapter,,,	8
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red,,, Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red,,,	6
Amazon Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Blue Kid-Proof Case - Blue	6
Fire HD 8 Tablet with Alexa, 8 HD Display, 16 GB, Tangerine - with Special Offers,	6
Kindle Oasis E-reader with Leather Charging Cover - Black, 6 High-Resolution Display (300 ppi), Wi-Fi - Includes Special Offers,,	6
New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable,,, New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable,,,	6
Amazon - Fire 16GB (5th Gen, 2015 Release) - Black,,, Amazon - Fire 16GB (5th Gen, 2015 Release) - Black,,,	5
Amazon Kindle Fire 5ft USB to Micro-USB Cable (works with most Micro-USB Tablets),,, Amazon Kindle Fire 5ft USB to Micro-USB Cable (works with most Micro-USB Tablets),,,	4
Certified Refurbished Amazon Fire TV with Alexa Voice Remote,,, Certified Refurbished Amazon Fire TV with Alexa Voice Remote,,,	4
Amazon Kindle Touch Leather Case (4th Generation - 2011 Release), Olive Green,,, Amazon Kindle Touch Leather Case (4th Generation - 2011 Release), Olive Green,,,	3
Certified Refurbished Amazon Fire TV (Previous Generation - 1st),,, Certified Refurbished Amazon Fire TV (Previous Generation - 1st),,,	3
Echo (Black),,, Echo (Black),,,	3
Amazon Kindle Fire Hd (3rd Generation) 8gb,,, Amazon Kindle Fire Hd (3rd Generation) 8gb,,,	2
Amazon Kindle Lighted Leather Cover,,, Amazon Kindle Lighted Leather Cover,,,	2
Amazon - Kindle Voyage - 4GB - Wi-Fi + 3G - Black,,, Fire HD 8 Tablet with Alexa, 8 HD Display, 16 GB, Tangerine - with Special Offers",	1
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red,,, Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,	1
Amazon Fire Tv,,, Kindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs)",,	1
Amazon Kindle Lighted Leather Cover,,, Kindle Keyboard,,,	1
Amazon Standing Protective Case for Fire HD 6 (4th Generation) - Black,,, Amazon Standing Protective Case for Fire HD 6 (4th Generation) - Black,,,	1
Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st),,, Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st),,,	1
Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st),,, Kindle Paperwhite,,,	1
Coconut Water Red Tea 16.5 Oz (pack of 12),,, Amazon Fire Tv,,,	1
Echo (Black),,, Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,	1
Echo (White),,, Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Tangerine"	1
New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable,,,	1

There are 49 distinct items (however, one of them is a blank value with 6760 reviews) with 34660 reviews. The item with the highest number of reviews is 'Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Magenta' with 10966 reviews. Following is a plot of the top 20 items with the most number of reviews (excluding the item with blank name)

Summary of top 20 products

Exercise 3: You will see there are 49 products with various review counts. Imagine the case that you are a consumer deciding between two similar products to purchase and are using reviews to help inform this purchase. From our data, select two products that are reasonable to compare and create new dataframes for each product containing only the reviews for that product. Identify which product you chose and print the first few rows of data for each product's dataframe. Make sure each product has at least 50 reviews. Hint: sort products alphabetically then use the split function.

The two similar products that I am considering for further investigation are:

- **Product 1**: All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta (2814 reviews)
- **Product 2**: All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta (158 reviews)

The difference between the two products here is the storage capacity of these devices. Sample of the reviews are given below

Table 3: Sample reviews for Product 1 (All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta)

SNo	name	reviews.text
961	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	You cannot top the value at this price point with any other tablet. The only reason I give it 4 stars instead of 5 is the limited app selection through the amazon app store. If this came with the google play store, I would give it a 5 in a heartbeat. Overall it is a great tablet at this price point. I would highly recommend it.
1008	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	YOU PAY FOR THE BEST AND YOU GET THE BEST. THANK YOU BEST BUY FOR PROVIDING THIS PRODUCT
1220	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	You will not find another tablet like this with the affordable price.
1486	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	You cannot get a better tablet for the price and Best Buy had it on sale.
2667	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	You get what you pay for... This is not great, the functionality and limitations leave much to be desired. Should have spent more and got another Samsung Galaxy Tab.

Table 4: Sample reviews for Product 2 (All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta)

SNo	name	reviews.text
30	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta	Wife really likes it. Works great& likes the front & back camera.
96	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta	We replaced and aging fire HDX. She noticed the WiFi connection is improved and the screen is larger. She prefers this to my iPad Air
98	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta	Wanted something with a kids mode on it so we could limit time and content. This works. Saw it for a really low price so couldn't resist otherwise we would have gotten the padded kids version.
118	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta	Wanted a tablet but did not want to pay for an iPad. The fire is a great tablet with a crystal clear hd videos. The perfect inexpensive alternative to the iPad
154	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta	Works very well. Video quality is very good. It is easy to navigate.

Exercise 4: Create a tidy text object for each product's review data.

```
tidy_reviews1<-selected_product1 %>%
    unnest_tokens("word", reviews.text)
tidy_reviews2<-selected_product2 %>%
    unnest_tokens("word", reviews.text)
```

Exercise 5: Clean the text of each dataset's review column as you see fit and find the top 20 frequent words for each product's reviews. Which words are unique to each product's top 20 list? Which words appear in both?

We will conduct the following steps to preprocess this text data:

- Removal of Stop words
- Removal of white spaces
- Removing numbers
- Removing intra-word punctiation
- Lemmatization
- Converting to lower case words (done automatically for tidytext objects)

Stop words

```r
data("stop_words")

# remove the stop words in the list above
tidy_reviews1<-tidy_reviews1 %>%
      anti_join(stop_words)

tidy_reviews2<-tidy_reviews2 %>%
      anti_join(stop_words)
```

White Spaces

```r
# removing white spaces
tidy_reviews1$word <- gsub("\\s+","",tidy_reviews1$word)
tidy_reviews2$word <- gsub("\\s+","",tidy_reviews2$word)
```

Removing Numbers

```r
# removing numbers from the words list
tidy_reviews1<-tidy_reviews1[-grep("\\b\\d+\\b", tidy_reviews1$word),]
tidy_reviews2<-tidy_reviews2[-grep("\\b\\d+\\b", tidy_reviews2$word),]
```

Punctuation

Although interword punctuation is removed in tidytext automatically, we will remove intraword punctuations separately.

```r
# removing punctuation (intraword)
tidy_reviews1$word<-removePunctuation(tidy_reviews1$word,preserve_intra_word_contractions = FALSE,
                preserve_intra_word_dashes = FALSE)

tidy_reviews2$word<-removePunctuation(tidy_reviews2$word,preserve_intra_word_contractions = FALSE,
                preserve_intra_word_dashes = FALSE)
```

Lemmatization

```r
#Lemmatization
tidy_reviews1<-tidy_reviews1  %>%
  mutate(word = textstem::lemmatize_words(word))

tidy_reviews2<-tidy_reviews2  %>%
  mutate(word = textstem::lemmatize_words(word))
```

Converting to lower case

This step happens automatically in the tidytext format

Top 20 words per product: Table 5 and Table 6

Common words from top 20 list of both products: Table 7 (17 words)

Unique words in top 20 word list for both products: Table 8 and Table 9 (3 words each)

Table 5: Top 20 words of product 1: (16 GB version of Fire tablet)

word	n
tablet	1476
love	747
buy	703
price	573
fire	548
kindle	541
easy	522
amazon	496
read	449
screen	372
app	360
purchase	333
book	310
game	276
play	257
product	237
gift	218
size	205
kid	203
ipad	195

Table 6: Top 20 words of product 2: (32 GB version of Fire tablet)

word	n
tablet	76
kindle	75
love	59
read	52
app	50
amazon	48
easy	43
buy	39
fire	38
device	34
screen	32
book	31
game	23
size	22
purchase	21
store	21
time	21
ipad	18
product	18
play	17

Table 7: Common words from both products' top 20 list

word	n1	n2
tablet	1476	76
love	747	59
buy	703	39
fire	548	38
kindle	541	75
easy	522	43
amazon	496	48
read	449	52
screen	372	32
app	360	50
purchase	333	21
book	310	31
game	276	23
play	257	17
product	237	18
size	205	22
ipad	195	18

Table 8: Words unique to Product 1

word	n
price	573
gift	218
kid	203

Table 9: Words unique to Product 2

word	n
device	34
store	21
time	21

Exercise 6: Create a dictionary of relevant terms that might aid in our comparison of these two products. Subset the tidytext dataframe that contain these words. What percentage of reviews did you capture with your dictionary for each product?
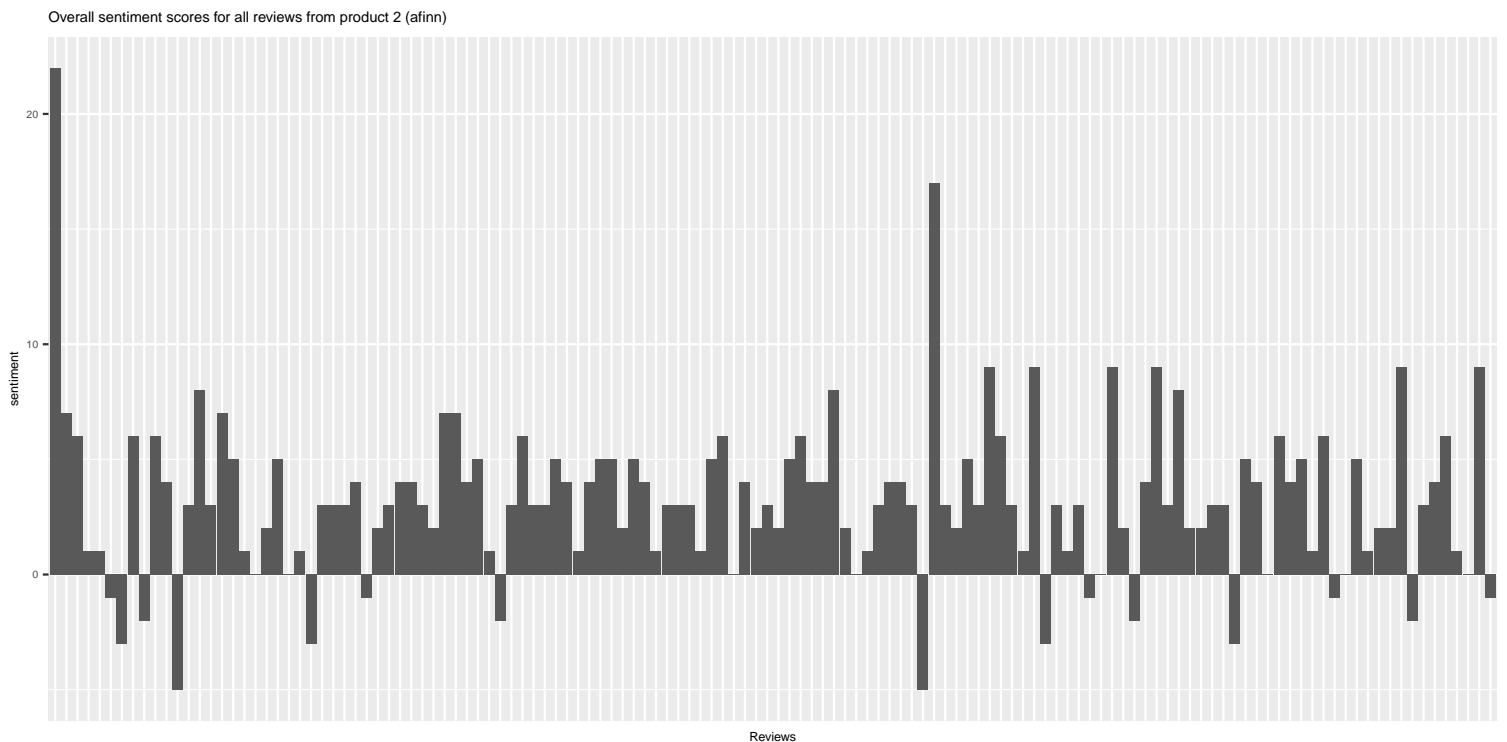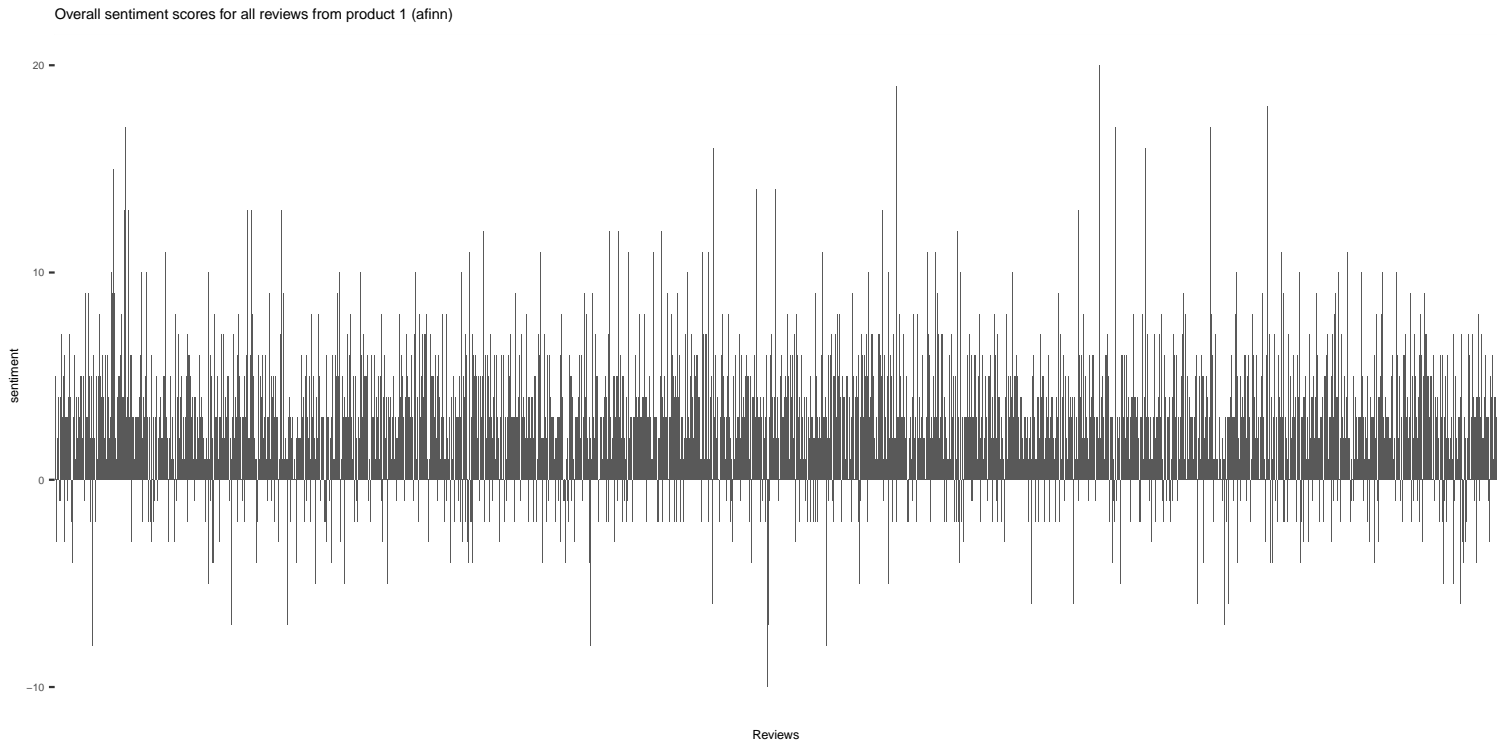
```
## The dictionary consists of the words:
##  tablet love store gift time easy memory download trade price app fast storage card upgrade


## Number of reviews selected from Product 1 (16 GB) is:  232
##   and number of reviews from Product 2 (32 GB) is:  18
```

Out of a total list of 2972 reviews, the dictionary filtered 250 reviews for both products combined. The original dataset had 2814 reviews for 'All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta' and 158 reviews for 'All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta'. After filtering as per the dictionary, we had 232 and 18 products for them respectively. This is a 91.75% reduction in the number of reviews for the first product (16 GB version) and a 88.60% reduction in the number of reviews for the second product (32 GB version).

Exercise 7: Perform a sentiment analysis on each individual review and then average your results for the entire set using the bing sentiment dictionary. Based on your results, which product would you choose to purchase?

Using 'afinn' dictionary

Overall sentiment scores for all reviews from product 1 (afinn)

Overall sentiment scores for all reviews from product 2 (afinn)

From the above, just by eye-balling we see that both products seems to have a similar proportion of negative reviews. The negative reviews are lower than the positive reviews. 396 out of 2250 reviews (17.6%) have a negative sentiment score assigned to them. For product 2, 23 out of 130 (17.7%) reviews are negative. However, Product 2 (32 GB version) has a higher maximum positive sentiment score (22) than Product 1 (20). It also has a higher average sentiment score (3.22) than Product 1 (2.76). Thus, just from the above analysis, **I would lean towards Product 2 (32 GB version) than Product 1.**

Using 'bing' dictionary
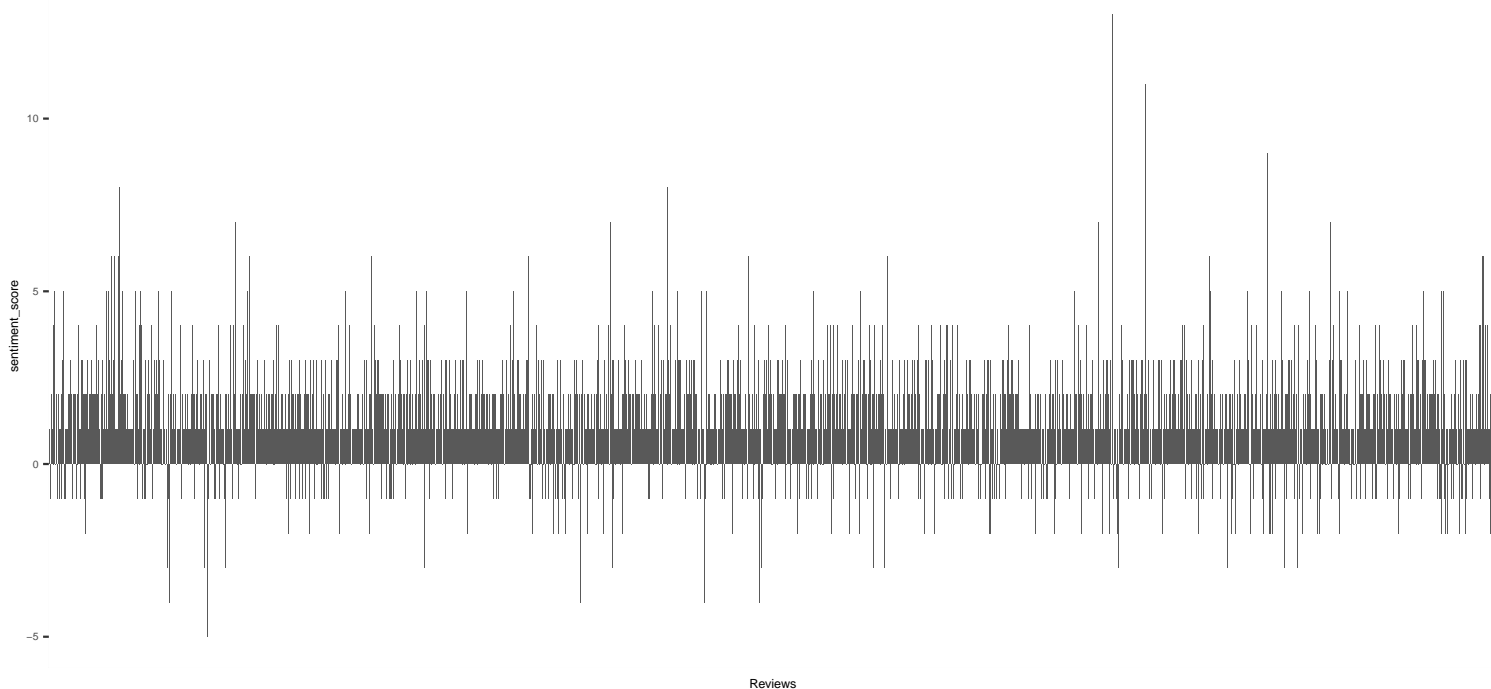
```
## product 1
reviews_sentiment_bing1 <- tidy_reviews1 %>%
  inner_join(get_sentiments("bing")) %>%
    count(SNo, sentiment)

reviews_sentiment1_summary<-cast(reviews_sentiment_bing1, SNo ~ sentiment)%>%
  replace_na(list(positive = 0, negative = 0))%>%
  mutate(sentiment_score=positive-negative)


## product 2
reviews_sentiment_bing2 <- tidy_reviews2 %>%
  inner_join(get_sentiments("bing")) %>%
    count(SNo, sentiment)

reviews_sentiment2_summary<-cast(reviews_sentiment_bing2, SNo ~ sentiment)%>%
  replace_na(list(positive = 0, negative = 0))%>%
  mutate(sentiment_score=positive-negative)
```
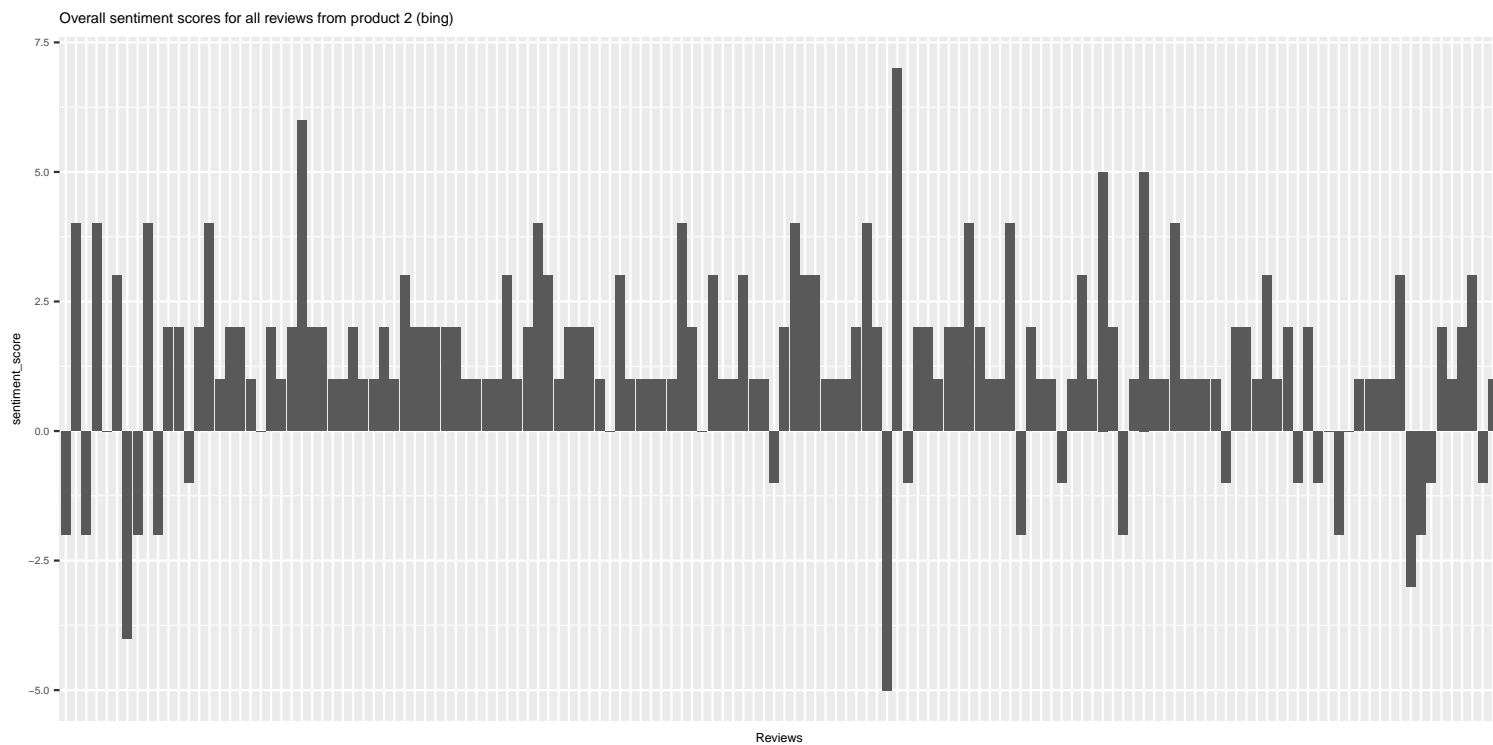
Overall sentiment scores for all reviews from product 1 (bing)

Overall sentiment scores for all reviews from product 2 (bing)

We see a similar trend using sentiment analysis plots from the bing disctionary. Let us look at the actual statistics of these scores to make a more informed comparison.

Table 10: Product Comparison

Product	Proportion.of.negative.reviews	Max.Positive.sentiment.score	Average.sentiment.score
Product 1 (16 GB)	0.184	13	1.311354
Product 2 (32 GB)	0.186	7	1.371429

From the above table and the plots, I am still favoring product 2 as it has a slightly higher average sentiment score and have almost similar proportion of negative reviews (as was seen in the case with 'afinn' dictionary). Moreover, by definition the second product has higher memory which is suitable for me and hence, as a customer, I would choose to purchase the second product **(All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta)**