Business Insight Report:

Shoe Brands and Running

Noor Hejeeali

Hult International Business School

Text Analytics and Natural Language Processing (NLP) – DAT -5317

Professor Thomas Kurnicki

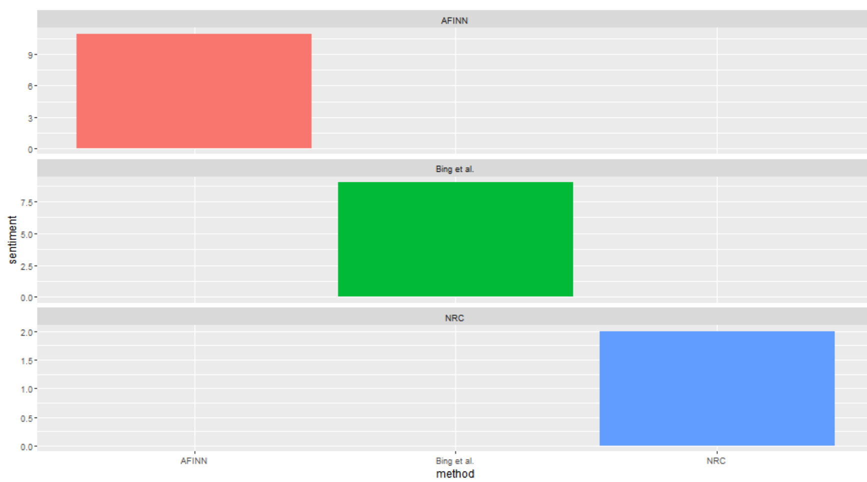
December 5th, 2021

Abstract

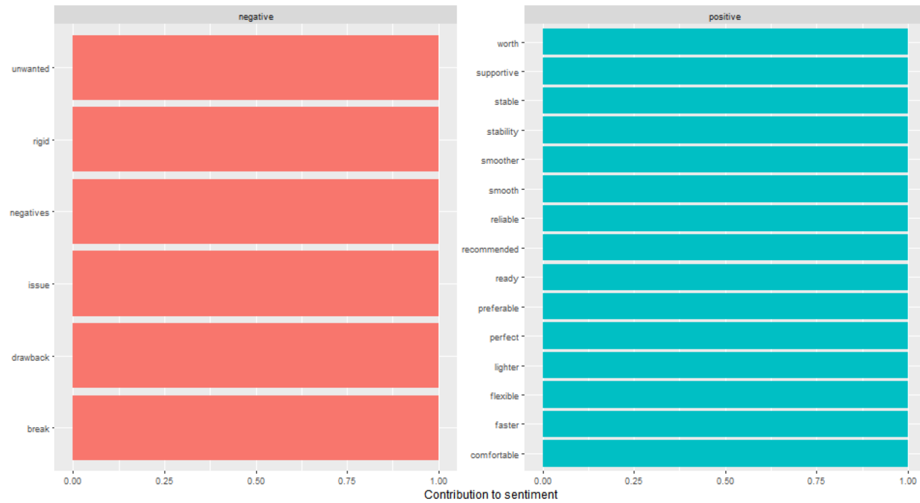
This business insight report is about four different running shoes from four different brands. Each shoe is analyzed through various frameworks to give recommendations on which shoe to purchase. The R code and R output is included in the appendix. The abstract, references page, and appendix are not included in the word count.

Finding the right running shoes is difficult but very important. Every person has a different preference on what they are looking for. Some look for comfort and stability while others look for a lightweight minimal support shoe. Some of the popular running brands are Brooks, Mizuno, Asics, and Nike. I picked 4 shoes which were Mizuno Wave Rider, Nike Pegasus, Asics Gel Kayano, and Brooks Adrenaline to analyze. The sentiment analysis, term frequency, tf-idf, n-grams analysis was used to get a better understanding of each brand. This analysis helps to understand each shoe and recommend the best one to purchase.

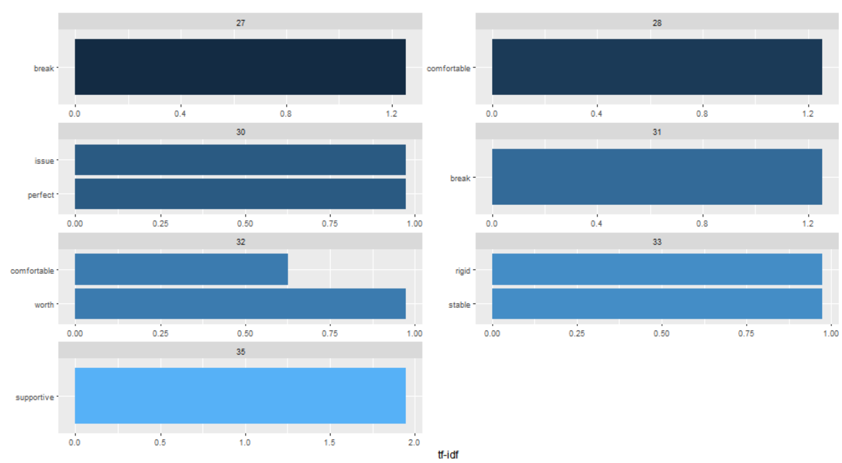
The first brand is Brooks. From the sentiment analysis, a variety of positive and negative words were shown. Even with the negative words, the AFINN and Bing charts showed no negative numbers. This is shown in the chart below.



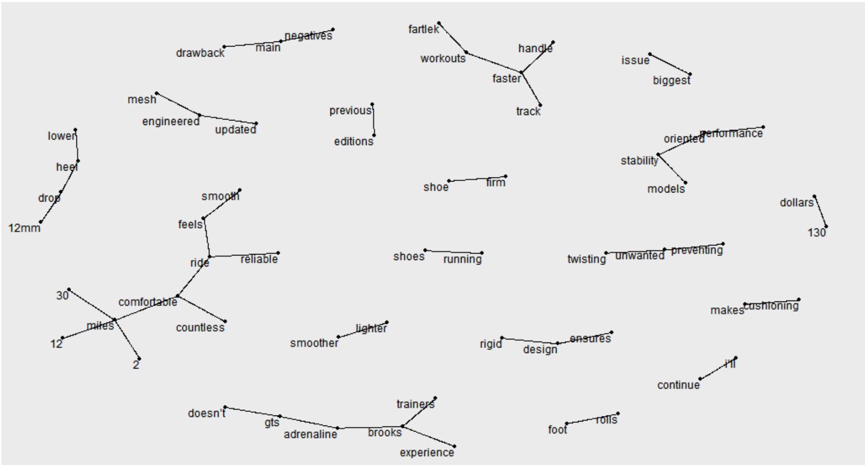
The NRC gives us insight into what types of words were used. It helps us understand the emotions but also whether a word is a positive or negative word. A variety of positive and negative words were used. The chart below shows the different negative and positive words. The negative words which were surprising were break, rigid, and unwanted. Rigid is an interesting negative word because comfortable was a positive word.



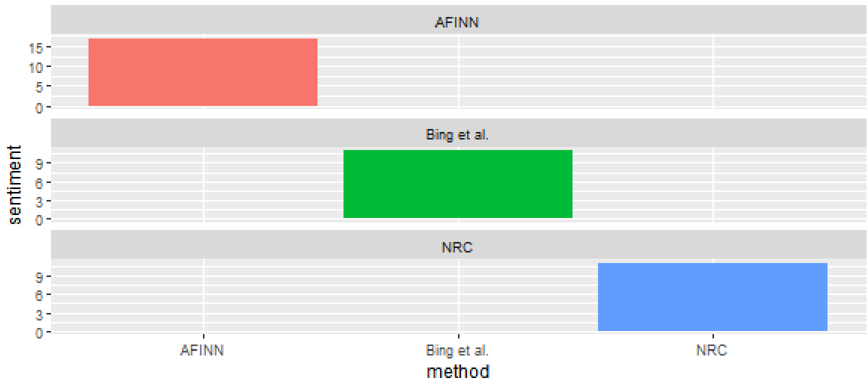
To dig deeper into the positive and negative words from the NRC, term frequency and tf-idf was done. Comfortable was used 51 times. Rigid was used 35 times. Issue and break were used frequently too. Rigid had a term frequency of 0.028 while comfortable had a term frequency of 0.038. This means that these terms were unique words and not common words used.The tf-idf help to define unique words which were used. The chart below shows the different unique terms used along with the calculated tf-idf.



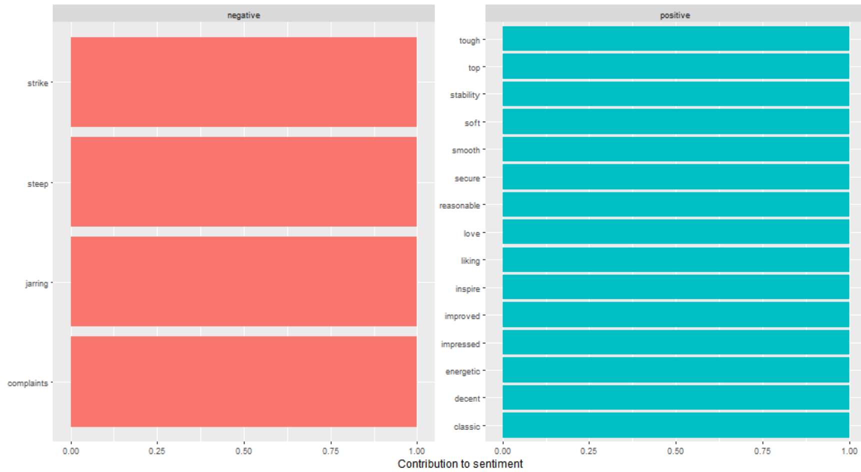
Besides this, the n-gram analysis was done to understand different words which were used together. The bigram was used because I wanted to see which 2 tokens are used together. The chart below shows the different words used together. The words that stood out were foot rolls, rigid design, and shoe firm. Foot rolls stood out because it seems this word was related to an injury. If a shoe is known for injuries, then a person may not want to purchase. It could be that the rigid design and shoe firmness caused the injury.



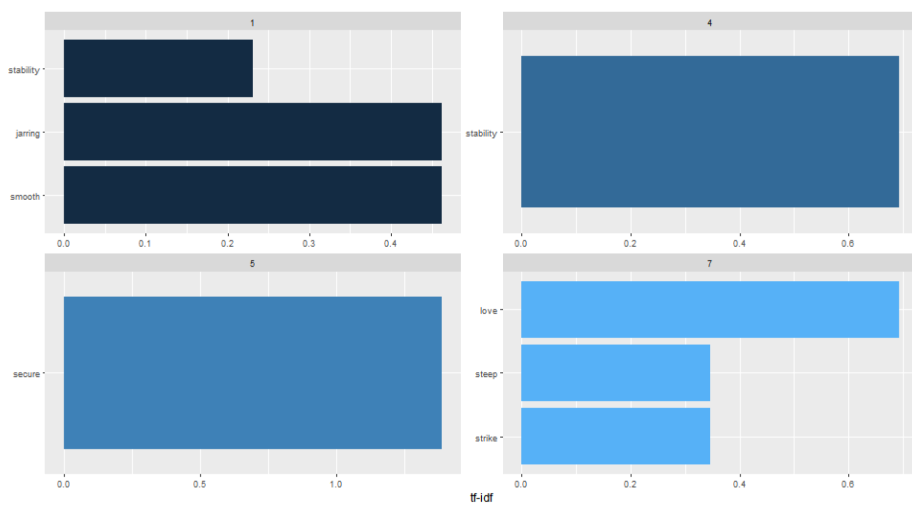
The second brand is Mizuno. The chart below shows that AFINN had a higher score for Mizuno compared to Brooks. There were no words that were associated with -1 for it. Similar to Brooks, Mizuno had no negative amounts associated with it.



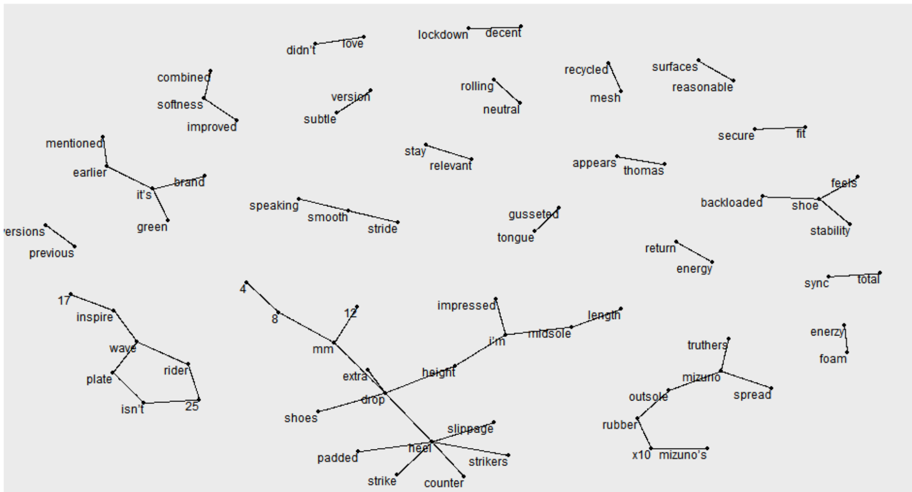
The chart below shows the different positive and negative words associated with Mizuno Wave Rider shoe. Strike and complaints were negative words that struck out. This means that there were complaints about this specific shoe. Some of the positive words were secure, reasonable, and classic. Even though the negative word complaints was used, people believed the shoe was a classic too.



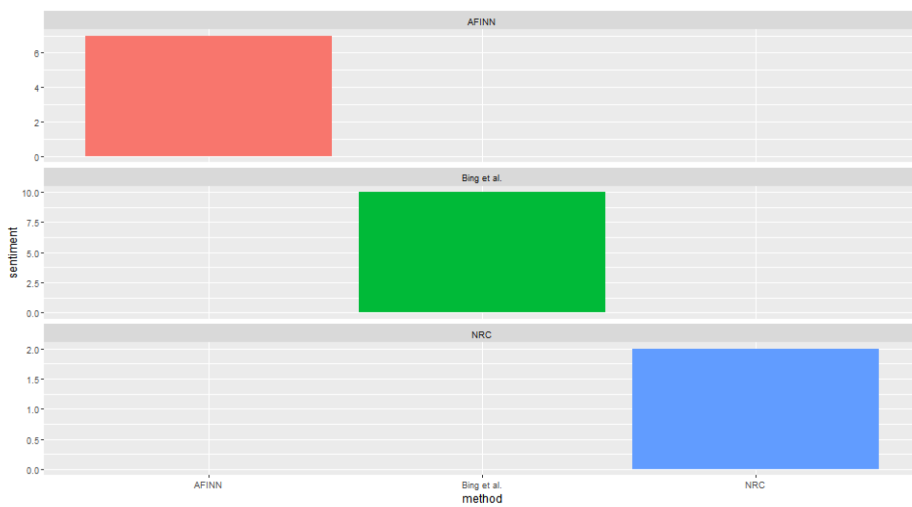
Stability and jarring were two of the most used terms. Jarring was used 72 times while stability was used 79 times. Strike was used 60 times. It is bad because two of the negative words are used frequently. Jarring and stability were in rank 1 but stability had a high tf-idf. Stability was not a unique word, but a frequent word used. Strike had a low tf-idf which was 0.346 meaning it was a unique word.



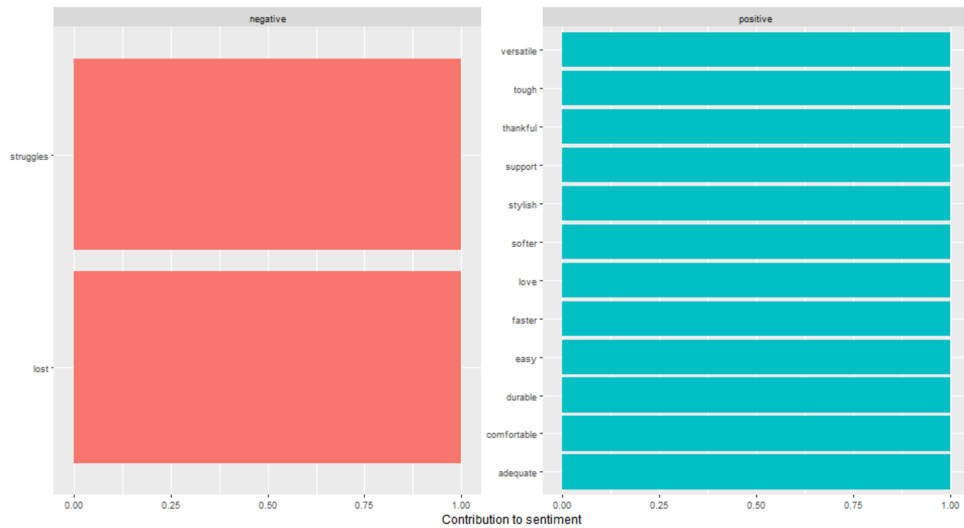
The words that struck out in the bigram chart shown below are heel strikers, heel slippage, and extra drop. From the sentiment analysis and tf-idf, strike was unique negative word used. It seems that many people had issues with the heel design of this shoe. The extra drop may have caused it.



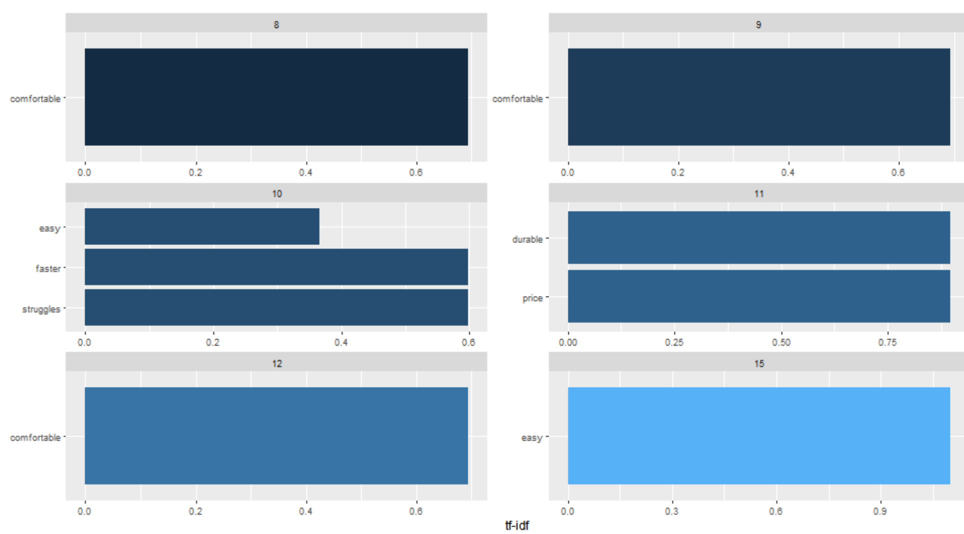
The third brand is Nike. The chart below shows that the AFINN is like Brooks and Mizuno. Also, Nike had no negative amounts too in Bing and AFINN. The NRC range was lower compared to Brooks and Mizuno.



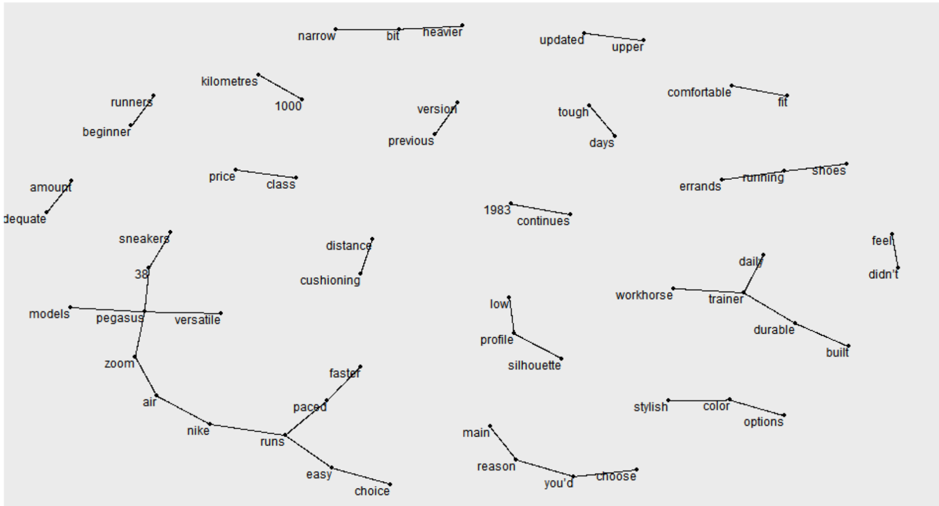
The chart below shows the various positive and negative words associated with the Nike Pegasus shoe. The negative word that struck out was struggles because it could mean something is wrong with design or fit of the shoe. Some of the positive words were versatile, comfortable, and support. Versatile was interesting because people may use a shoe for other activities besides running.



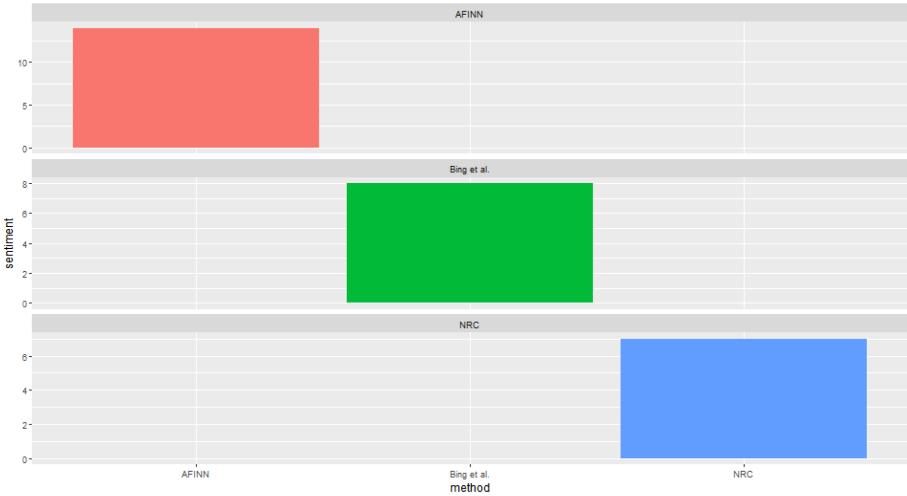
Easy was one of the most used terms with an amount of 51. Durable and price were used often too with an amount of 38. Struggles was used 28 times. The negative words were used less frequently here compared to the other 2 brands. Even though easy was used many times, it had a high tf-idf of 1.09. This means it is not a unique word. Struggle is a unique word used because the tf-idf was 0.59 for it. This is shown in the chart below.



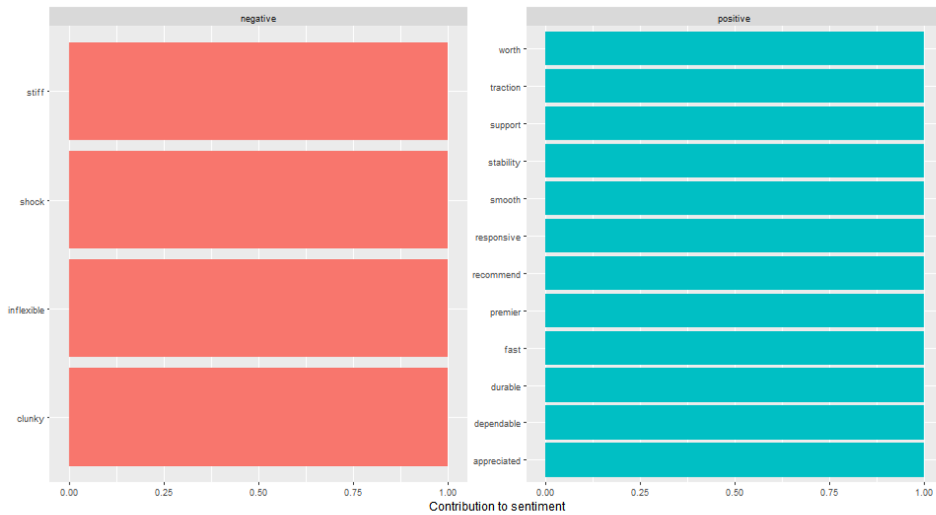
From the bigram chart below, the words that struck out were beginner runners, bit narrow, and daily trainer. If a person is has just started running or is looking for a multipurpose shoe, then the Nike Pegasus shoe may be a good fit. The downside is the narrow fit which may not work for everyone.



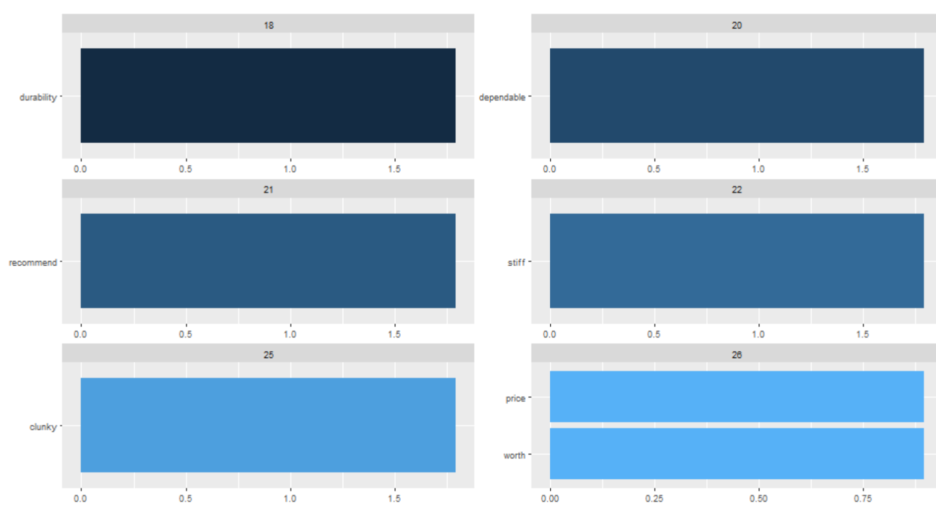
The final brand is Asics. Similar to the other brands, the AFINN and Bing had no negative amounts. The NRC range low compared to Brooks and Mizuno. This is shown in the chart below.



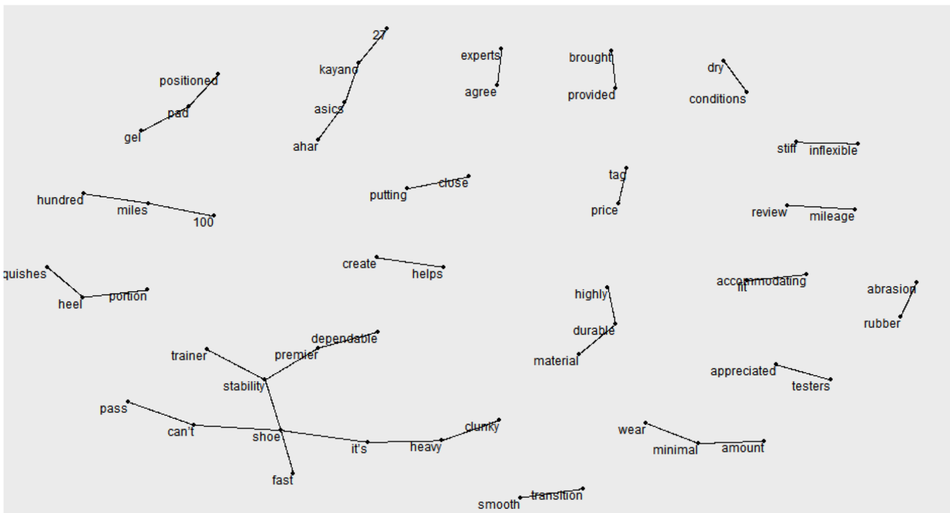
In the chart below it shows the different positive and negative words associated with the Asics Gel Kayano shoe. The negative words stiff and clunky were interesting. Stiff is like the negative word rigid used for Brooks. Some of the positive words used were support, dependable, and durable.



The terms durability and clunky were the most used. Durability was used 33 times while clunky was used 28 times. Even though these two terms were used the most, the tf-idf was very high. The tf-idf was 1.791 which means these words are frequent and not unique. The unique words were price and worth which had a tf-idf of 0.895. The chart below shows the tf-idf for the different words.



The bigram chart below shows the different 2 token terms used together. The words that caught my eye were appreciated testers, hundred miles, experts agree, and stability trainer. If a runner who is looking for a long-distance shoe, then the Asics Gel Kayano may be a good fit. Also, experts agree and appreciated testers stood out because it shows credibility for the brand and shoe.



From the different analysis done, I would recommend not purchasing the shoes Brooks Adrenaline and Mizuno Wave Rider. The reasons why I would recommend against purchasing is because of the impact of the negative words such as strike and rigid. The negative words were used a high amount of time. The negative words were used more frequently for these 2 shoes compared to others. Also, the bigrams chart showed words that stuck out. These words include heel slippage, foot rolls, rigid design, and extra drop. Rigid was a negative word used frequently for Brooks Adrenaline. Since strike was a negative word used for Mizuno then this may be why the heel slippage may be occurring.

Depending on what a person is looking for I would recommend purchasing the Nike Pegasus or Asics Gel Kayano shoe. I would recommend a person who is new to running and looking for a versatile shoe to purchase the Nike Pegasus. Through the sentiment and tf-idf analysis, price and durable were used. A new runner may be on a budget and wants a cheaper durable shoe. Also, the bigram chart showed that beginner runner and daily trainer were used. The Nike Pegasus would be a good fit for them. I would recommend a person who runs often or is a marathon runner to purchase the Asics Gel Kayano shoe. Through the sentiment and tf-idf analysis, durability was a term used often. Worth was not a frequent term used. Also, the bigrams chart shows words used such as hundred miles and experts agree.

In the end I would recommend a beginner runner to purchase the Nike Pegasus Shoe and a regular runner to purchase the Asics Gel Kayano shoe. I would not recommend purchasing the Brooks Adrenaline and Mizuno Wave Rider shoes. This recommendation was given after completing the different analysis and visualizations. These analyses include sentiment, term frequency, tf-idf, and n-grams.

References

Eisinger, A. (2021). *The Nike Air Zoom Pegasus 38 is the perfect shoe for all types of runners*. SELF. https://www.self.com/review/nike-pegasus-38-review.

Ellis, C. (2021). *Brooks Adrenaline GTS 21 review*. TechRadar. https://www.techradar.com/in/reviews/brooks-adrenaline-gts-21.

Jones, T. J. (2020). *Asics Gel Kayano 27 Review*. Running Shoes Guru. https://www.runningshoesguru.com/2020/06/asics-gel-kayano-27-review/.

Langelier, A., Haines, R. (2021). *Mizuno Wave Rider 25 performance review " believe in the run*. Believe in the Run. https://www.believeintherun.com/mizuno-wave-rider-25-performance-review/.

Law, B. (2021). *Nike Zoom Pegasus 38 review*. Running Shoes Guru. https://www.runningshoesguru.com/2021/05/nike-zoom-pegasus-38-review/.

Law, B. (2021). *Mizuno Wave Rider 25 review*. Running Shoes Guru. https://www.runningshoesguru.com/2021/08/mizuno-wave-rider-25-review/.

Matsumoto, E. (2020). *Shoe review: Asics Gel-Kayano 27*. Fleet Feet. https://www.fleetfeet.com/blog/shoe-review-asics-gel-kayano-27.

Metzler, B. (2020). *BROOKS ADRENALINE GTS 21 REVIEW*. JackRabbit.https://www.jackrabbit.com/info/blog/brooks-adrenaline-gts-21-review.html.

Petruny, M. (2021). *Nike gives the 38th Air Zoom Pegasus just the quick tune-up it needed*. Runner's World. https://www.runnersworld.com/gear/a37282601/nike-air-zoom-pegasus-38-review/.

Ronto, P. (2020). *Asics Gel Kayano 27 - review 2021 - facts, deals ($89)*. Athletic shoe reviews. https://runrepeat.com/asics-gel-kayano-27.

Unknown. (2021). *Gel-Kayano*. ASICS. https://www.asics.com/us/en-us/gel-kayano/c/aa50101000/.

Unknown. (2021). *How to choose the right shoes for you*. Runners Need. https://www.runnersneed.com/expert-advice/gear-guides/choosing-the-right-running-shoes.html.

Unknown. (2021). *Running Shoes Reviews*. Running Shoes Guru. https://www.runningshoesguru.com/reviews/nike/pegasus/.

Unknown. (2021). *The new adrenaline GTS 22*. Brooks Running. https://www.brooksrunning.com/en_us/brooks-adrenaline-gts-running-shoes/.

Unknown. (2021). *Wave rider 24*. Mizuno USA Official Website. https://www.mizunousa.com/wave-rider-24.

Unknown. (2021). *Brooks Adrenaline GTS 21 review*. Running Shoes Guru. https://www.runningshoesguru.com/2021/01/brooks-adrenaline-gts-21-review/.

Unknown. (2021). *Brooks Adrenaline GTS 21: Reviews and full analysis*. Runner's Lab. https://runnerslab.com/review/brooks-adrenaline-gts-21/.

Appendix

R Code:

```
.libPaths("C:/Rlibs")

#Mizuno Reviews Text
library(dplyr)
library(tidytext)

#Mizuno Reviews
text1 <- c("Enerzy is what Mizuno needed to stay relevant in the game. The Wave Rider 25 is energetic with the energy return and bounce that hits just right without being too soft or too jarring. The midsole and Wave Plate are in total sync with each other to make every stride smooth. Speaking of that plate, it provides just a little bit of stability for those who don't quite pronate to the extreme.")
text2 <- c("I'm impressed that Mizuno kept the slipper-like step-in feel that has decent lockdown. When I first tried these on, I was reminded of what made the previous versions good for their era. However, these aren't the Riders of the early 2000s, as Mizuno does a good job of blending the new with the classic. If you like Mizuno, you will like this shoe, and it'll probably last forever.")
text3 <- c("Mizuno's X10 rubber outsole is tough as a bowl of nails, without any milk. I put it through plenty of reasonable surfaces around town, yet it keeps coming back like it's brand new.")
text4 <- c("On top of the outsole, Mizuno spread its Enerzy foam from the heel to the rest of the shoe for a full-length midsole. I'm a big fan of the improved softness combined with this more subtle version of the Wave Plate. The Wave Rider 25 isn't as much of a stability shoe, so the Wave Plate isn't nearly as extreme as the one on the Wave Inspire 17. This shoe feels much more neutral rolling through your stride.")
text5 <- c("Moving to the upper, I have no complaints about the recycled mesh. As mentioned earlier, it's green in more ways than one which is a huge plus for me. A well-padded heel counter keeps my foot in place and eliminates any heel slippage. The gusseted tongue also contributes to the secure fit, but let's not mention it too much (say "gusseted tongue" three times and Thomas appears, just like Beetlejuice).")
text6 <- c("This is a backloaded shoe, with much more stack in the heel (hello, 12 mm drop). Heel strikers will like this shoe, but the midfooters of the world, like myself, may feel a little left out. That 12 mm drop is noticeable, especially if you're used to 4-8 mm drop shoes.")
text7 <- c("The Mizuno truthers out there won't like this take: I don't like the 12 mm drop. I didn't love it in the Wave Inspire 17, and it doesn't work here either. It pushes me to heel strike and is a bit too steep for my liking. If you can teach me to love the extra drop height, I'm all ears.")

mizuno_rating <- rbind(text1, text2, text3, text4, text5, text6, text7)
colnames(mizuno_rating)[1] <- "text"
mizuno_rating <- as.data.frame(mizuno_rating)

id <- c(1,2,3,4,5,6,7)
```

```
id <- matrix(id, ncol=1)

mizuno_rating <- cbind(id,mizuno_rating)

tidy_mizuno <- mizuno_rating %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word) %>%
  arrange(desc(n))

#Nike Pegasus Review
text8 <- c("When I tried it on for the first time, the updated upper felt way softer and more comfortable than the previous version. It fit like a glove.")
text9 <- c("On that first run, the ride felt really comfortable with plenty of long-distance cushioning but it didn’t feel like the versatile Pegasus models of days gone by.")
text10 <- c("The Pegasus 38 is more suited to easy runs and struggles with faster-paced runs because of how much softer it is. This workhorse has lost its wings.")
text11 <- c("The main reason you’d choose the Pegasus 38 over any other daily trainer in its price class is durability. It feels like a really well-built, durable trainer with an outsole that could easily do over 1000 kilometres.")
text12 <- c("It’s also a great option for beginner runners who only want to buy one pair of shoes that has a comfortable fit and will last a very, very long time.")
text13 <- c("It’s on those tough days that I’m especially thankful for shoes like the Nike Air Zoom Pegasus 38. The workhorse trainer—which was first introduced in 1983—continues to be a shoe that any runner, at any level, can use and love.")
text14 <- c("I always find that Nike runs a bit narrow and a bit small. ")
text15 <- c("In many ways the Nike Air Zoom Pegasus 38 sneakers really are the Goldilocks of running shoes: Just the right amount of cushion and responsiveness, plus a low-profile silhouette and several stylish color options that make these an easy choice to wear while running errands or with any athleisure look.")
text16 <- c("Though they’re a bit heavier than one might expect, I didn’t really notice this until I was on slightly longer runs. Perhaps best of all, they provided an adequate amount of support so that my knees and hips felt good from start to finish.")

nike_rating <- rbind(text8, text9, text10, text11, text12,text13,text14,text15,text16)
colnames(nike_rating)[1] <- "text"
nike_rating <- as.data.frame(nike_rating)

id <- c(8,9,10,11,12,13,14,15,16)
id <- matrix(id, ncol=1)

nike_rating <- cbind(id,nike_rating)

tidy_nike <- nike_rating %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
```

```
  count(word) %>%
  arrange(desc(n))

#Asics Gel Kayano 27 Review
text17 <- c(" The heel portion is firm and responsive while the forefoot is plushy and springy.")
text18 <- c("Durability is a strength of the Kayano 27. They show minimal wear even after putting close to 100 miles on them. These are built to last through many miles at many different paces.")
text19 <- c("The material used for the outsole is AHAR (ASICS High Abrasion Rubber), a highly durable material that provides traction in both wet and dry conditions. After a hundred miles, a minimal amount of wear showed on the outsole.")
text20 <- c("Overall, the ASICS Kayano 27 is a dependable premier stability trainer. The ride they provided brought me back to them again and again after my review mileage.")
text21 <- c("If you are a runner looking for a premier stability shoe that delivers every time you lace them up I would recommend the ASICS Kayano 27.")
text22 <- c("These aspects can cause the back portion of the trainers to be inflexible, stiff and heavy.")
text23 <- c("Two of the most noticeable things about the Kayano, though, are its step-in feel and accommodating fit.")
text24 <- c("The GEL pad, positioned in the heel, squishes when you land to soak up shock and disperses it evenly throughout the shoe. It also helps create a smooth transition from heel to toe, which testers appreciated.")
text25 <- c("All experts agree this shoe can’t pass as a fast shoe. It’s heavy, clunky, has a lot of support and, therefore, can’t help you pick up the pace. ")
text26 <- c("While the price tag is high, those who commented on it said Kayano 27 is worth the money. ")

asics_rating <- rbind(text17, text18, text19,text20,text21,text22,text23,text24,text25,text26)
colnames(asics_rating)[1] <- "text"
asics_rating <- as.data.frame(asics_rating)

id <- c(17,18,19,20,21,22,23,24,25,26)
id <- matrix(id, ncol=1)

asics_rating <- cbind(id,asics_rating)

tidy_asics <- asics_rating %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word) %>%
  arrange(desc(n))

#Brooks Adrenaline GTS 21 Review
text27 <- c("I knew from experience Brooks trainers take a little longer to break-in. ")
text28 <- c("They provided a natural and comfortable ride from the first stride. By the end of the run, I was looking forward to more miles in them.")
```

```
text29 <- c("It was a reliable ride whether it was 2 miles or 12 miles. The midsole was built to last for hundreds of miles.")
text30 <- c("Although a great shoe, and one I’ll continue to use, the shoe is not perfect. There are two main negatives that linger in my mind. The biggest issue for me was the 12mm drop.")
text31 <- c("The other main drawback for me was the break-in time. Most running shoes are ready to go straight out of the box, yet these took more than 30 miles to get a good feel for them.")
text32 <- c("Priced at 130 dollars, the Brooks Adrenaline is worth every penny. They provided me with countless comfortable miles because of the execution of all aspects.")
text33 <- c("As expected, the ride feels smooth and very stable. This is definitely a firm shoe, but not unpleasantly so. The rigid design ensures good transfer of energy as your foot rolls, while preventing unwanted twisting.")
text34 <- c("While this shoe is lighter, smoother and more flexible than previous editions, the Adrenaline GTS doesn’t have the liveliness to handle faster track and fartlek workouts as well as some performance-oriented stability models.")
text35 <- c("The cushioning makes it a great shoe for longer rides and daily use. The upper is breathable and supportive with an updated engineered mesh. ")
text36 <- c(" It's not recommended for sprints or faster workouts. Some users also suggested that a lower heel drop-off would be preferable.")

brooks_rating <- rbind(text27,text28,text29,text30,text31,text32,text33,text34,text35,text36)
colnames(brooks_rating)[1] <- "text"
brooks_rating <- as.data.frame(brooks_rating)

id <- c(27,28,29,30,31,32,33,34,35,36)
id <- matrix(id, ncol=1)

brooks_rating <- cbind(id,brooks_rating)

tidy_brooks <- brooks_rating %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word) %>%
  arrange(desc(n))

###SENTIMENT ANALYSIS

###BROOKS####
#Getting sentiment analysis for Brooks
brooks_afinn <- tidy_brooks %>%
  inner_join(get_sentiments("afinn"))%>%
  summarise(sentiment=sum(value)) %>%
  mutate(method="AFINN")

brooks_bing_and_nrc <- bind_rows(
  tidy_brooks%>%
```

```
    inner_join(get_sentiments("bing"))%>%
    mutate(method = "Bing et al."),
  tidy_brooks %>%
    inner_join(get_sentiments("nrc") %>%
                filter(sentiment %in% c("positive", "negative"))) %>%
    mutate(method = "NRC")) %>%
  count(method,  sentiment) %>%
  spread(sentiment, n, fill=0) %>%
  mutate(sentiment = positive-negative)

library(ggplot2)
bind_rows(brooks_afinn, brooks_bing_and_nrc) %>%
  ggplot(aes(method, sentiment, fill=method))+
  geom_col(show.legend=FALSE)+
  facet_wrap(~method, ncol =1, scales= "free_y")

#Most common positive and negative words for brooks
brooks_bing_counts <- tidy_brooks %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=T) %>%
  ungroup()

brooks_bing_counts

brooks_bing_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="Contribution to sentiment", x=NULL)+
  coord_flip()

###ASICS######
#Getting sentiment analysis for Asics
asics_afinn <- tidy_asics %>%
  inner_join(get_sentiments("afinn"))%>%
  summarise(sentiment=sum(value)) %>%
  mutate(method="AFINN")

asics_bing_and_nrc <- bind_rows(
  tidy_asics%>%
    inner_join(get_sentiments("bing"))%>%
    mutate(method = "Bing et al."),
```

```
  tidy_asics %>%
    inner_join(get_sentiments("nrc") %>%
                 filter(sentiment %in% c("positive", "negative"))) %>%
    mutate(method = "NRC")) %>%
  count(method,  sentiment) %>%
  spread(sentiment, n, fill=0) %>%
  mutate(sentiment = positive-negative)

library(ggplot2)
bind_rows(asics_afinn, asics_bing_and_nrc) %>%
  ggplot(aes(method, sentiment, fill=method))+
  geom_col(show.legend=FALSE)+
  facet_wrap(~method, ncol =1, scales= "free_y")

#Most common positive and negative words for asics
asics_bing_counts <- tidy_asics %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=T) %>%
  ungroup()

asics_bing_counts

asics_bing_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="Contribution to sentiment", x=NULL)+
  coord_flip()

###NIKE####
#Getting sentiment analysis for Nike
nike_afinn <- tidy_nike %>%
  inner_join(get_sentiments("afinn"))%>%
  summarise(sentiment=sum(value)) %>%
  mutate(method="AFINN")

nike_bing_and_nrc <- bind_rows(
  tidy_nike%>%
    inner_join(get_sentiments("bing"))%>%
    mutate(method = "Bing et al."),
  tidy_brooks %>%
    inner_join(get_sentiments("nrc") %>%
```

```
              filter(sentiment %in% c("positive", "negative"))) %>%
    mutate(method = "NRC")) %>%
  count(method,  sentiment) %>%
  spread(sentiment, n, fill=0) %>%
  mutate(sentiment = positive-negative)

library(ggplot2)
bind_rows(nike_afinn, nike_bing_and_nrc) %>%
  ggplot(aes(method, sentiment, fill=method))+
  geom_col(show.legend=FALSE)+
  facet_wrap(~method, ncol =1, scales= "free_y")

#Most common positive and negative words for nike
nike_bing_counts <- tidy_nike %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=T) %>%
  ungroup()

nike_bing_counts

nike_bing_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="Contribution to sentiment", x=NULL)+
  coord_flip()

###MIZUNO####
#Getting sentiment analysis for mizuno
mizuno_afinn <- tidy_mizuno %>%
  inner_join(get_sentiments("afinn"))%>%
  summarise(sentiment=sum(value)) %>%
  mutate(method="AFINN")

mizuno_bing_and_nrc <- bind_rows(
  tidy_mizuno%>%
    inner_join(get_sentiments("bing"))%>%
    mutate(method = "Bing et al."),
  tidy_mizuno %>%
    inner_join(get_sentiments("nrc") %>%
              filter(sentiment %in% c("positive", "negative"))) %>%
    mutate(method = "NRC")) %>%
```

```
  count(method,  sentiment) %>%
  spread(sentiment, n, fill=0) %>%
  mutate(sentiment = positive-negative)

library(ggplot2)
bind_rows(mizuno_afinn, mizuno_bing_and_nrc) %>%
  ggplot(aes(method, sentiment, fill=method))+
  geom_col(show.legend=FALSE)+
  facet_wrap(~method, ncol =1, scales= "free_y")

#Most common positive and negative words for mizuno
mizuno_bing_counts <- tidy_mizuno %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=T) %>%
  ungroup()

mizuno_bing_counts

mizuno_bing_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n, fill=sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y")+
  labs(y="Contribution to sentiment", x=NULL)+
  coord_flip()

##ZIPFs Law and TF-IDF###
###MIZUNO####
mizuno <- mizuno_rating %>%
  unnest_tokens(word, text) %>%
  count(id, word, sort=TRUE) %>%
  ungroup()

mizuno_total_words <- mizuno %>%
  group_by(id) %>%
  summarize(total=sum(n))

mizuno_words <- left_join(mizuno, mizuno_total_words)%>%
  filter(word %in% c("smooth", "secure", "strike", "jarring", "steep", "love", "stability"))

print(mizuno_words)

mizuno_freq_by_rank <- mizuno_words %>%
```

```
  group_by(id) %>%
  mutate(rank = row_number(),
         `term frequency` = n/total)
mizuno_freq_by_rank

mizuno_freq_by_rank %>%
  ggplot(aes(rank, `term frequency`, color=word))+
  geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
  geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
  scale_x_log10()+
  scale_y_log10()

#TF-IDF
mizuno_brand_words <- mizuno_words %>%
  bind_tf_idf(word, id,n)

mizuno_brand_words

mizuno_brand_words %>%
  arrange(desc(tf_idf))

mizuno_uniqueness <- mizuno_brand_words %>%
  arrange(desc(tf_idf))

#Graph for Term Frequency
mizuno_brand_words %>%
  arrange(desc(tf_idf)) %>%
  mutate(word=factor(word, levels=rev(unique(word)))) %>%
  group_by(id) %>%
  top_n(15) %>%
  ungroup %>%
  ggplot(aes(word, tf_idf, fill=id))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="tf-idf")+
  facet_wrap(~id, ncol=2, scales="free")+
  coord_flip()

##NIKE###
nike <- nike_rating %>%
  unnest_tokens(word, text) %>%
  count(id, word, sort=TRUE) %>%
  ungroup()

nike_total_words <- nike %>%
  group_by(id) %>%
  summarize(total=sum(n))
```

```
nike_words <- left_join(nike, nike_total_words)%>%
  filter(word %in% c("easy", "faster", "struggles", "soft", "price", "comfortable", "durable"))

print(nike_words)

nike_freq_by_rank <- nike_words %>%
  group_by(id) %>%
  mutate(rank = row_number(),
         `term frequency` = n/total)
nike_freq_by_rank

nike_freq_by_rank %>%
  ggplot(aes(rank, `term frequency`, color=word))+
  geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
  geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
  scale_x_log10()+
  scale_y_log10()

#TF-IDF
nike_brand_words <- nike_words %>%
  bind_tf_idf(word, id,n)

nike_brand_words

nike_brand_words %>%
  arrange(desc(tf_idf))

nike_uniqueness <- nike_brand_words %>%
  arrange(desc(tf_idf))

#Graph for Term Frequency
nike_brand_words %>%
  arrange(desc(tf_idf)) %>%
  mutate(word=factor(word, levels=rev(unique(word)))) %>%
  group_by(id) %>%
  top_n(15) %>%
  ungroup %>%
  ggplot(aes(word, tf_idf, fill=id))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="tf-idf")+
  facet_wrap(~id, ncol=2, scales="free")+
  coord_flip()

##Asics###
asics <- asics_rating %>%
```

```
  unnest_tokens(word, text) %>%
  count(id, word, sort=TRUE) %>%
  ungroup()

asics_total_words <- asics %>%
  group_by(id) %>%
  summarize(total=sum(n))

asics_words <- left_join(asics, asics_total_words)%>%
  filter(word %in% c("durability", "dependable", "worth","recommend", "stiff", "price",
"clunky"))

print(asics_words)

asics_freq_by_rank <- asics_words %>%
  group_by(id) %>%
  mutate(rank = row_number(),
        `term frequency` = n/total)
asics_freq_by_rank

asics_freq_by_rank %>%
  ggplot(aes(rank, `term frequency`, color=word))+
  geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
  geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
  scale_x_log10()+
  scale_y_log10()

#TF-IDF
asics_brand_words <- asics_words %>%
  bind_tf_idf(word, id,n)

asics_brand_words

asics_brand_words %>%
  arrange(desc(tf_idf))

asics_uniqueness <- asics_brand_words %>%
  arrange(desc(tf_idf))

#Graph for Term Frequency
asics_brand_words %>%
  arrange(desc(tf_idf)) %>%
  mutate(word=factor(word, levels=rev(unique(word)))) %>%
  group_by(id) %>%
  top_n(15) %>%
  ungroup %>%
```

```
  ggplot(aes(word, tf_idf, fill=id))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="tf-idf")+
  facet_wrap(~id, ncol=2, scales="free")+
  coord_flip()

##Brooks###
brooks <- brooks_rating %>%
  unnest_tokens(word, text) %>%
  count(id, word, sort=TRUE) %>%
  ungroup()

brooks_total_words <- brooks %>%
  group_by(id) %>%
  summarize(total=sum(n))

brooks_words <- left_join(brooks, brooks_total_words)%>%
  filter(word %in% c("comfortable", "stable","worth", "perfect", "supportive", "rigid", "issue",
"break"))

print(brooks_words)

brooks_freq_by_rank <- brooks_words %>%
  group_by(id) %>%
  mutate(rank = row_number(),
         `term frequency` = n/total)
brooks_freq_by_rank

brooks_freq_by_rank %>%
  ggplot(aes(rank, `term frequency`, color=word))+
  geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
  geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
  scale_x_log10()+
  scale_y_log10()

#TF-IDF
brooks_brand_words <- brooks_words %>%
  bind_tf_idf(word, id,n)

brooks_brand_words

brooks_brand_words %>%
  arrange(desc(tf_idf))

brooks_uniqueness <- brooks_brand_words %>%
  arrange(desc(tf_idf))
```

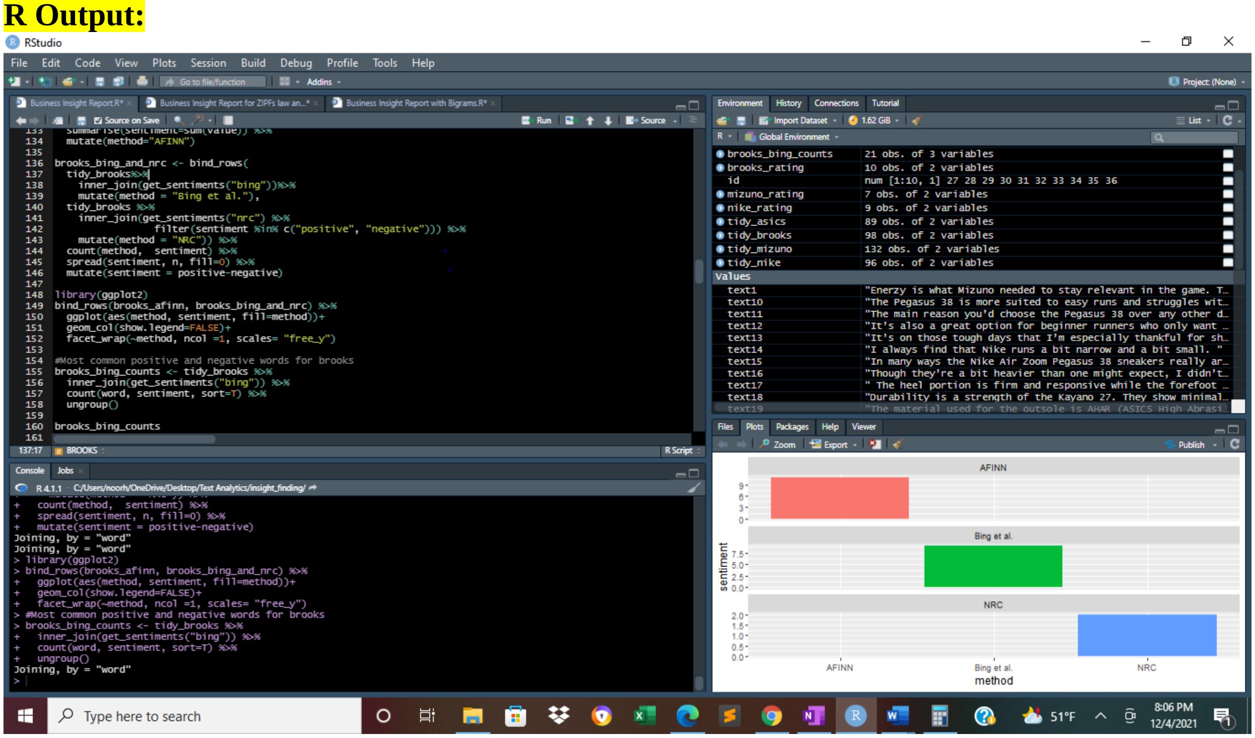
```
#Graph for Term Frequency
brooks_brand_words %>%
  arrange(desc(tf_idf)) %>%
  mutate(word=factor(word, levels=rev(unique(word)))) %>%
  group_by(id) %>%
  top_n(15) %>%
  ungroup %>%
  ggplot(aes(word, tf_idf, fill=id))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="tf-idf")+
  facet_wrap(~id, ncol=2, scales="free")+
  coord_flip()
##Bigrams##
###MIZUNO Bigrams
mizuno_bigram <- mizuno_rating %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2)

mizuno_bigram

mizuno_bigram%>%
  count(bigram, sort = TRUE)

library(tidyr)
mizuno_bigrams_separated <- mizuno_bigram%>%
  separate(bigram, c("word1", "word2"), sep = " ")

mizuno_bigrams_filtered <- mizuno_bigrams_separated %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)

mizuno_bigram_counts <- mizuno_bigrams_filtered %>%
  count(word1, word2, sort = TRUE)

mizuno_bigram_counts

###BROOKS

brooks_bigram <- brooks_rating %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2)

brooks_bigram

brooks_bigram%>%
  count(bigram, sort = TRUE)
```

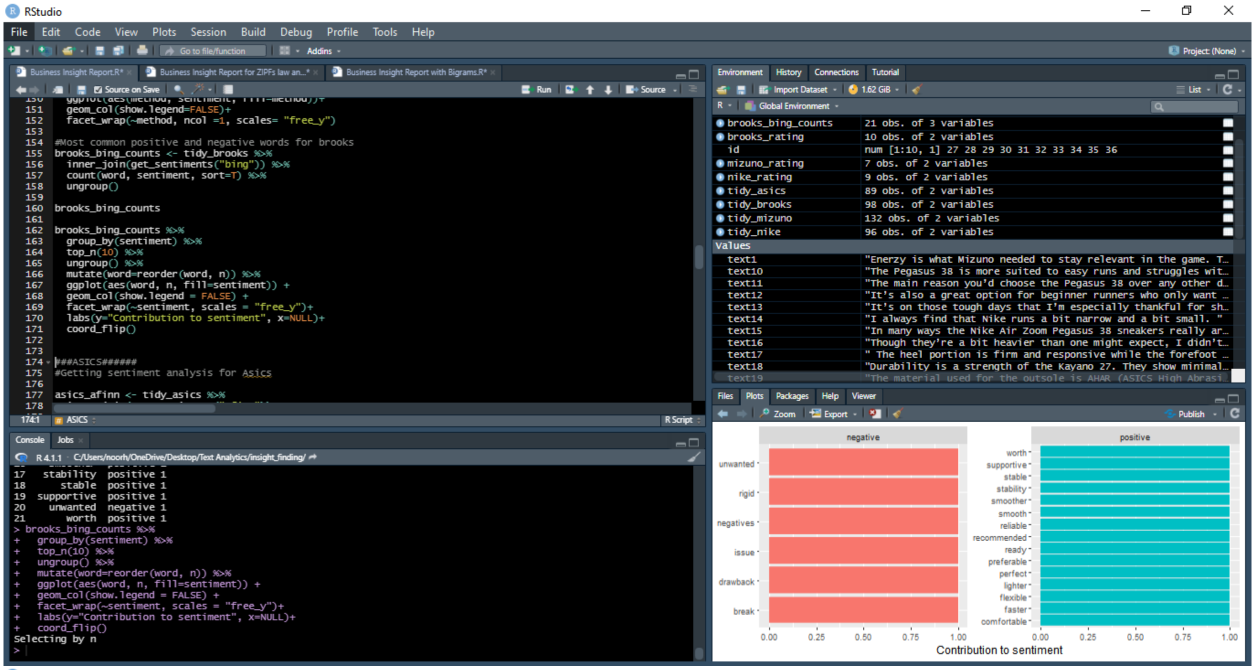
```
library(tidyr)
brooks_bigrams_separated <- brooks_bigram%>%
  separate(bigram, c("word1", "word2"), sep = " ")

brooks_bigrams_filtered <- brooks_bigrams_separated %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)

#creating the new bigram, "no-stop-words":
brooks_bigram_counts <- brooks_bigrams_filtered %>%
  count(word1, word2, sort = TRUE)

brooks_bigram_counts

###ASICS
asics_bigram <- asics_rating %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2)

asics_bigram

asics_bigram%>%
  count(bigram, sort = TRUE)

library(tidyr)

asics_bigrams_separated <- asics_bigram%>%
  separate(bigram, c("word1", "word2"), sep = " ")

asics_bigrams_filtered <- asics_bigrams_separated %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)

asics_bigram_counts <- asics_bigrams_filtered %>%
  count(word1, word2, sort = TRUE)

asics_bigram_counts

###NIKE

nike_bigram <- nike_rating %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2)

nike_bigram

nike_bigram%>%
  count(bigram, sort = TRUE)
```

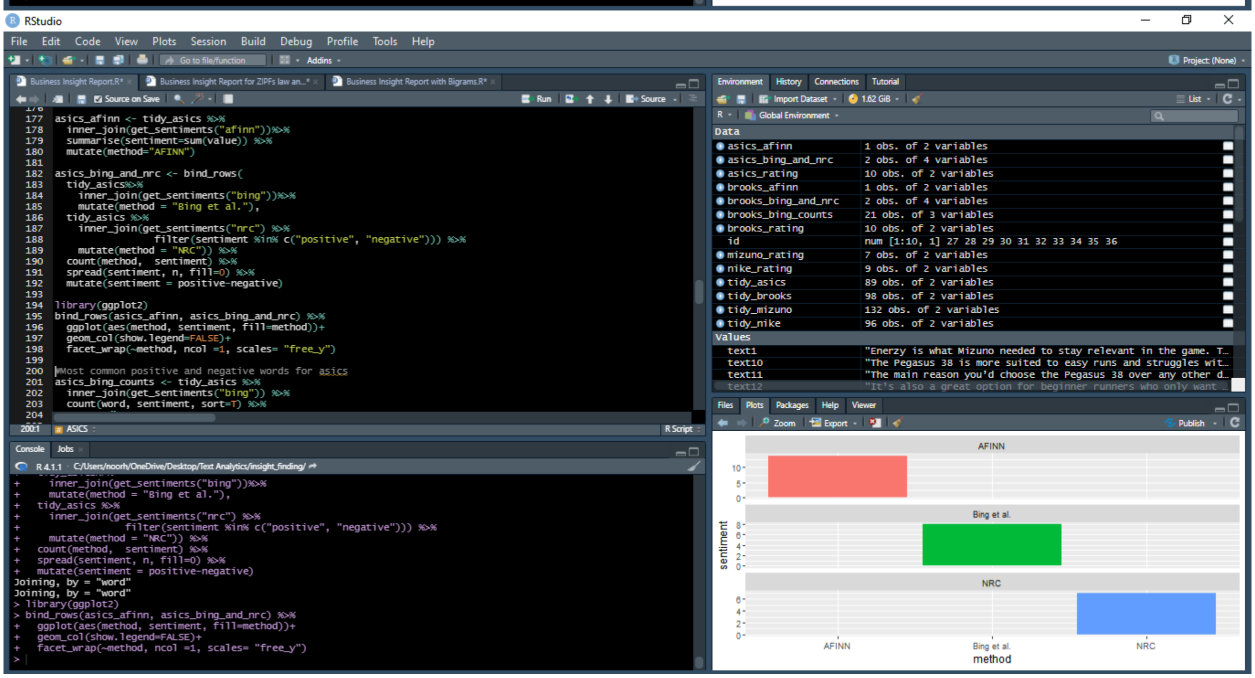
```
library(tidyr)
nike_bigrams_separated <- nike_bigram%>%
  separate(bigram, c("word1", "word2"), sep = " ")

nike_bigrams_filtered <- nike_bigrams_separated %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)

nike_bigram_counts <- nike_bigrams_filtered %>%
  count(word1, word2, sort = TRUE)

nike_bigram_counts

###CHARTS FOR BIGRAMS
##MIZUNO
library(igraph)
mizuno_bigram_graph <- mizuno_bigram_counts %>%
  filter(n>0) %>%
  graph_from_data_frame()

mizuno_bigram_graph

#install.packages("ggraph")
library(ggraph)
ggraph(mizuno_bigram_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)

###Brooks
library(igraph)
brooks_bigram_graph <- brooks_bigram_counts %>%
  filter(n>0) %>%
  graph_from_data_frame()

brooks_bigram_graph

library(ggraph)
ggraph(brooks_bigram_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)

##Asics
library(igraph)
```

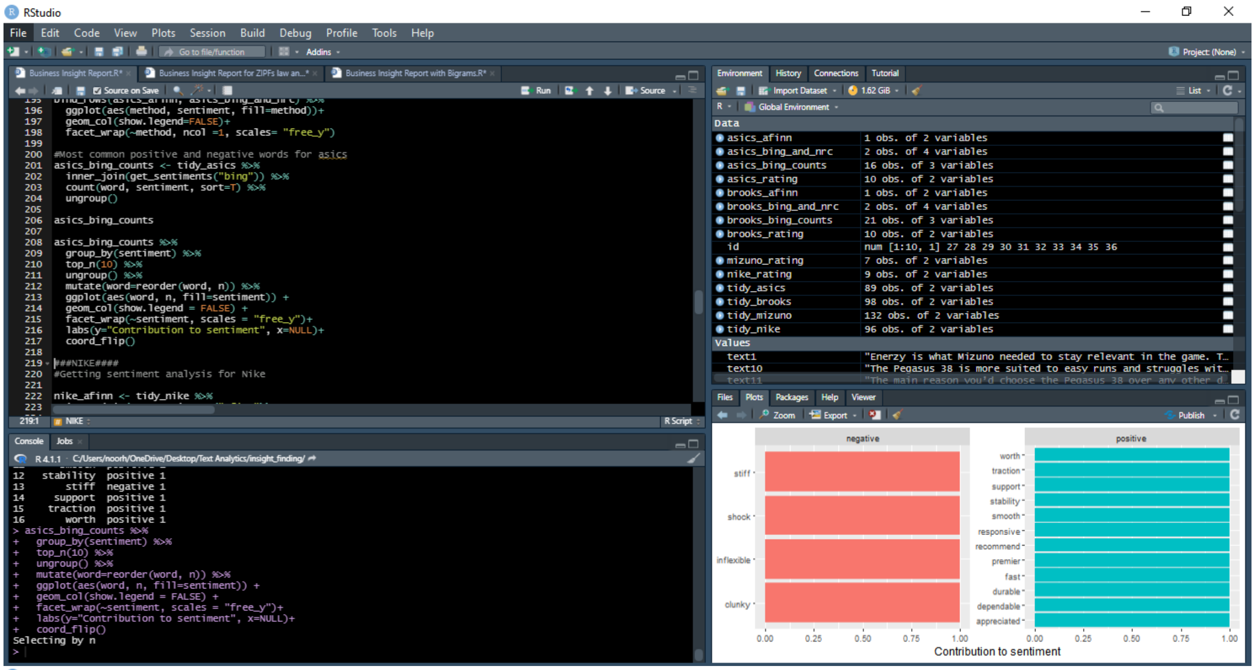
```
asics_bigram_graph <- asics_bigram_counts %>%
  filter(n>0) %>%
  graph_from_data_frame()

asics_bigram_graph

library(ggraph)
ggraph(asics_bigram_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)

##Nike
library(igraph)
nike_bigram_graph <- nike_bigram_counts %>%
  filter(n>0) %>%
  graph_from_data_frame()

nike_bigram_graph

library(ggraph)
ggraph(nike_bigram_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

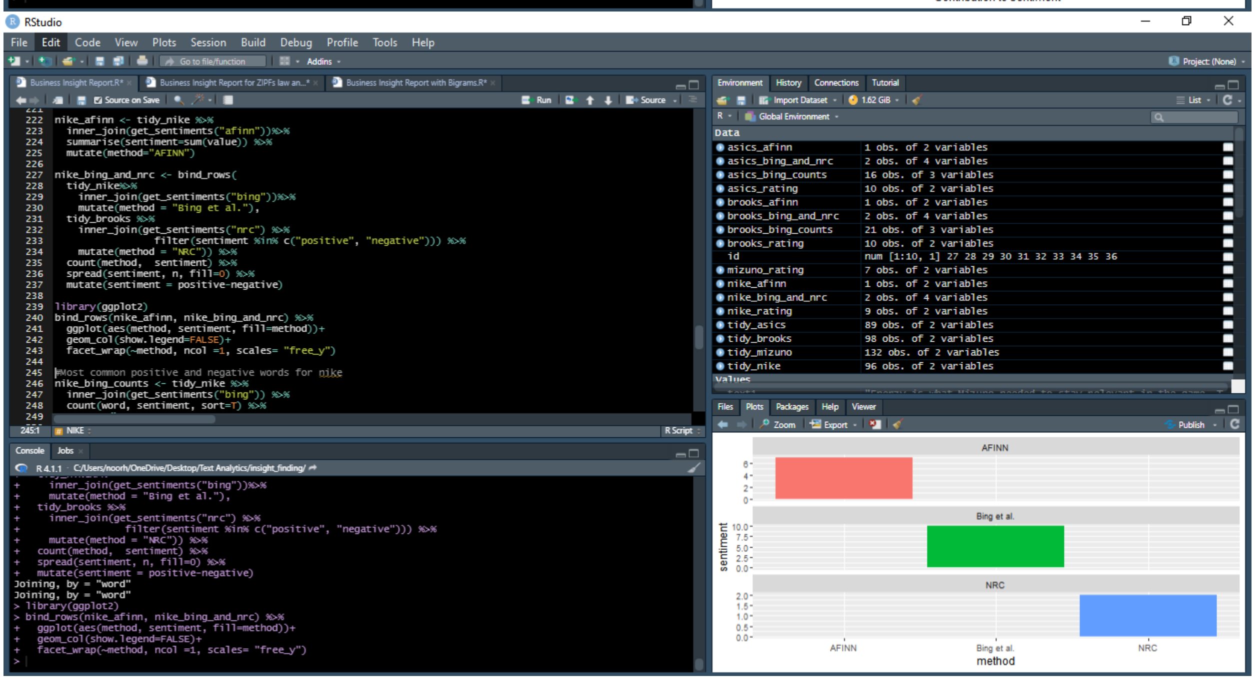
R Output:

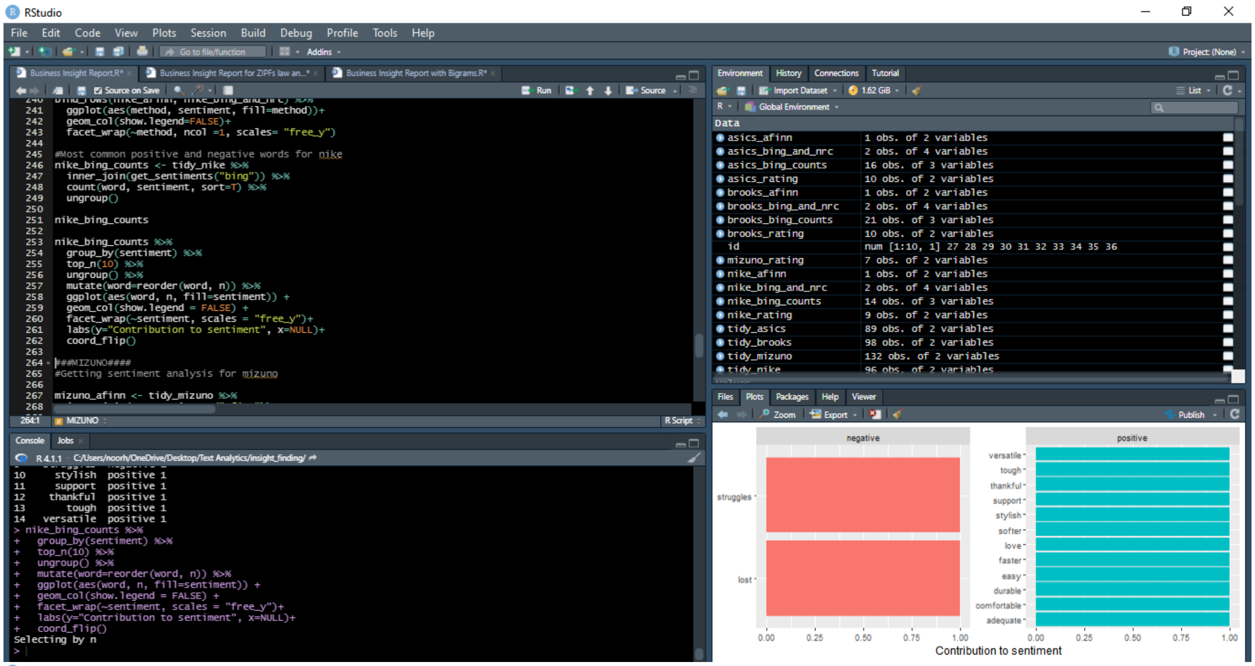


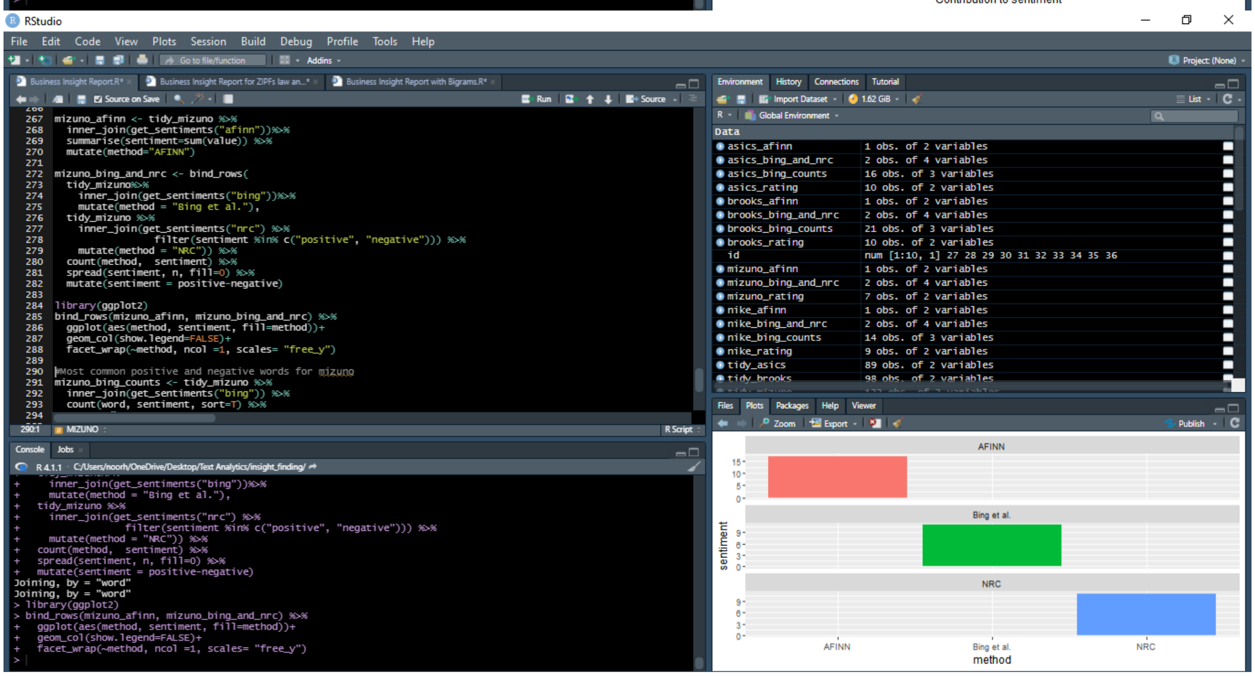

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
geom_col(show.legend=FALSE)+
facet_wrap(~method, ncol =1, scales= "free_y")
#Most common positive and negative words for brooks
brooks_bing_counts <- tidy_brooks %>%
inner_join(get_sentiments("bing")) %>%
count(word, sentiment, sort=T) %>%
ungroup()
brooks_bing_counts
brooks_bing_counts %>%
group_by(sentiment) %>%
top_n(10) %>%
ungroup() %>%
mutate(word=reorder(word, n)) %>%
ggplot(aes(word, n, fill=sentiment)) +
geom_col(show.legend = FALSE) +
facet_wrap(~sentiment, scales = "free_y")+
labs(y="Contribution to sentiment", x=NULL)+
coord_flip()
###ASICS######
#Getting sentiment analysis for Asics
asics_afinn <- tidy_asics %>%
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
17 stability positive 1
18 stable positive 1
19 supportive positive 1
20 unwanted negative 1
21 worth positive 1
> brooks_bing_counts %>%
+ group_by(sentiment) %>%
+ top_n(10) %>%
+ ungroup() %>%
+ mutate(word=reorder(word, n)) %>%
+ ggplot(aes(word, n, fill=sentiment)) +
+ geom_col(show.legend = FALSE) +
+ facet_wrap(~sentiment, scales = "free_y")+
+ labs(y="Contribution to sentiment", x=NULL)+
+ coord_flip()
Selecting by n
Environment History Connections Tutorial
Global Environment
brooks_bing_counts 21 obs. of 3 variables
brooks_rating 10 obs. of 2 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno_rating 7 obs. of 2 variables
nike_rating 9 obs. of 2 variables
tidy_asics 89 obs. of 2 variables
tidy_brooks 98 obs. of 2 variables
tidy_mizuno 132 obs. of 2 variables
tidy_nike 96 obs. of 2 variables
Values
Files Plots Packages Help Viewer
negative
positive
unwanted
rigid
negatives
issue
drawback
break
worth
supportive
stable
stability
smoother
smooth
reliable
recommended
ready
preferable
perfect
lighter
flexible
faster
comfortable
0.00 0.25 0.50 0.75 1.00
Contribution to sentiment

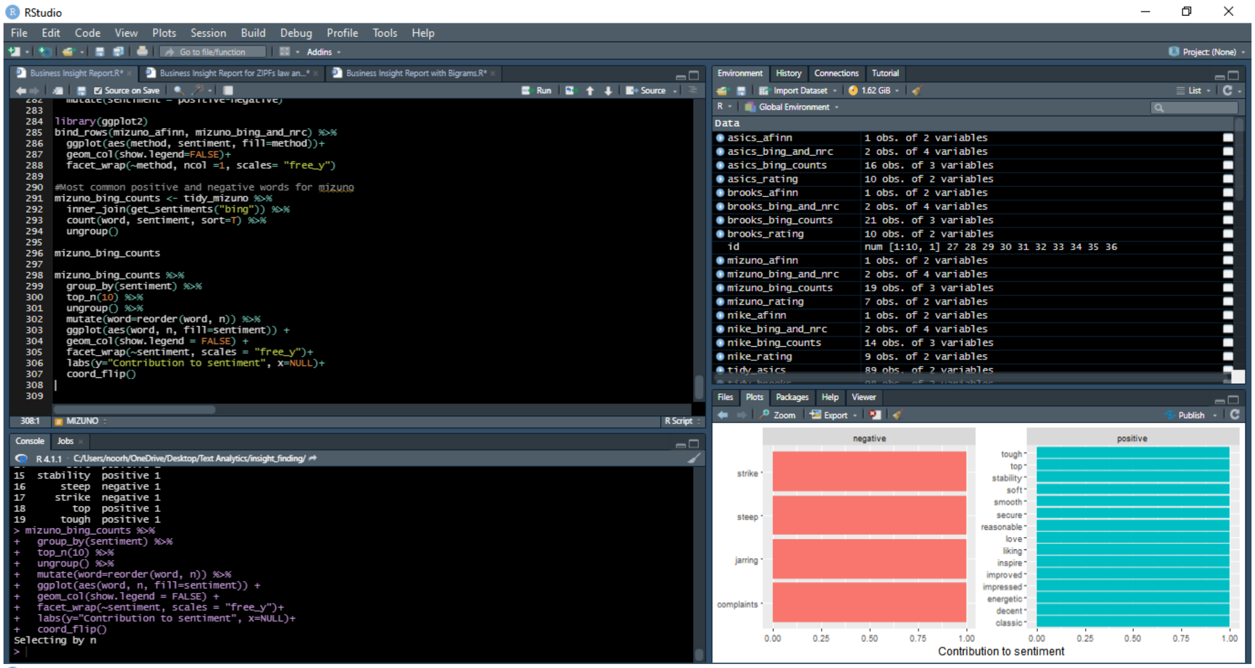

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
asics_afinn <- tidy_asics %>%
inner_join(get_sentiments("afinn"))%>%
summarise(sentiment=sum(value)) %>%
mutate(method="AFINN")
asics_bing_and_nrc <- bind_rows(
tidy_asics%>%
inner_join(get_sentiments("bing"))%>%
mutate(method = "Bing et al."),
tidy_asics %>%
inner_join(get_sentiments("nrc") %>%
filter(sentiment %in% c("positive", "negative"))) %>%
mutate(method = "NRC")) %>%
count(method, sentiment) %>%
spread(sentiment, n, fill=0) %>%
mutate(sentiment = positive-negative)
library(ggplot2)
bind_rows(asics_afinn, asics_bing_and_nrc) %>%
ggplot(aes(method, sentiment, fill=method))+
geom_col(show.legend=FALSE)+
facet_wrap(~method, ncol =1, scales= "free_y")
#Most common positive and negative words for asics
asics_bing_counts <- tidy_asics %>%
inner_join(get_sentiments("bing")) %>%
count(word, sentiment, sort=T) %>%
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
+ inner_join(get_sentiments("bing"))%>%
+ mutate(method = "Bing et al."),
+ tidy_asics %>%
+ inner_join(get_sentiments("nrc") %>%
+ filter(sentiment %in% c("positive", "negative"))) %>%
+ mutate(method = "NRC")) %>%
+ count(method, sentiment) %>%
+ spread(sentiment, n, fill=0) %>%
+ mutate(sentiment = positive-negative)
Joining, by = "word"
Joining, by = "word"
> library(ggplot2)
> bind_rows(asics_afinn, asics_bing_and_nrc) %>%
+ ggplot(aes(method, sentiment, fill=method))+
+ geom_col(show.legend=FALSE)+
+ facet_wrap(~method, ncol =1, scales= "free_y")
Environment History Connections Tutorial
Global Environment
Data
asics_afinn 1 obs. of 2 variables
asics_bing_and_nrc 2 obs. of 4 variables
asics_rating 10 obs. of 2 variables
brooks_afinn 1 obs. of 2 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_rating 10 obs. of 2 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno_rating 7 obs. of 2 variables
nike_rating 9 obs. of 2 variables
tidy_asics 89 obs. of 2 variables
tidy_brooks 98 obs. of 2 variables
tidy_mizuno 132 obs. of 2 variables
tidy_nike 96 obs. of 2 variables
Values
Files Plots Packages Help Viewer
AFINN
Bing et al.
NRC
sentiment
AFINN Bing et al. NRC
method

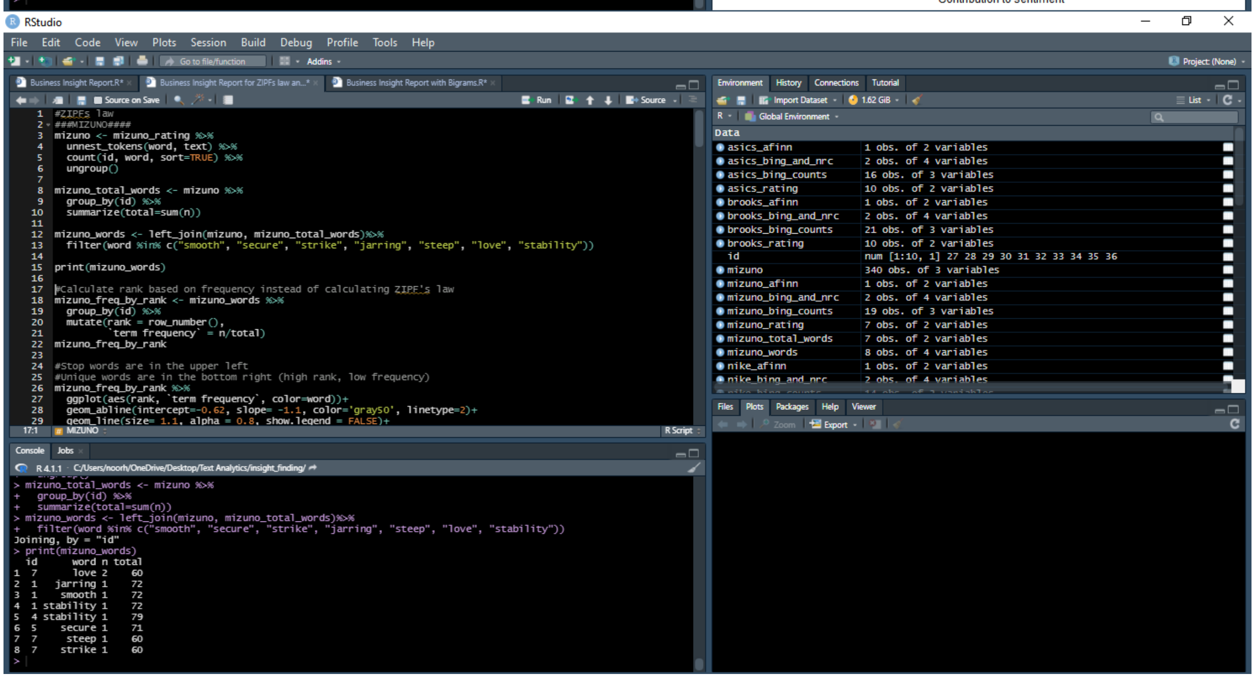

RStudio
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
ggplot(aes(method, sentiment, fill=method))+
geom_col(show.legend=FALSE)+
facet_wrap(~method, ncol =1, scales= "free_y")
#Most common positive and negative words for asics
asics_bing_counts <- tidy_asics %>%
inner_join(get_sentiments("bing")) %>%
count(word, sentiment, sort=T) %>%
ungroup()
asics_bing_counts
asics_bing_counts %>%
group_by(sentiment) %>%
top_n(10) %>%
ungroup() %>%
mutate(word=reorder(word, n)) %>%
ggplot(aes(word, n, fill=sentiment)) +
geom_col(show.legend = FALSE) +
facet_wrap(~sentiment, scales = "free_y")+
labs(y="Contribution to sentiment", x=NULL)+
coord_flip()
###NIKE####
#Getting sentiment analysis for Nike
nike_afinn <- tidy_nike %>%
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
negative
positive
stiff
shock
inflexible
clunky
worth
traction
support
stability
smooth
responsive
recommend
premier
fast
durable
dependable
appreciated
Contribution to sentiment

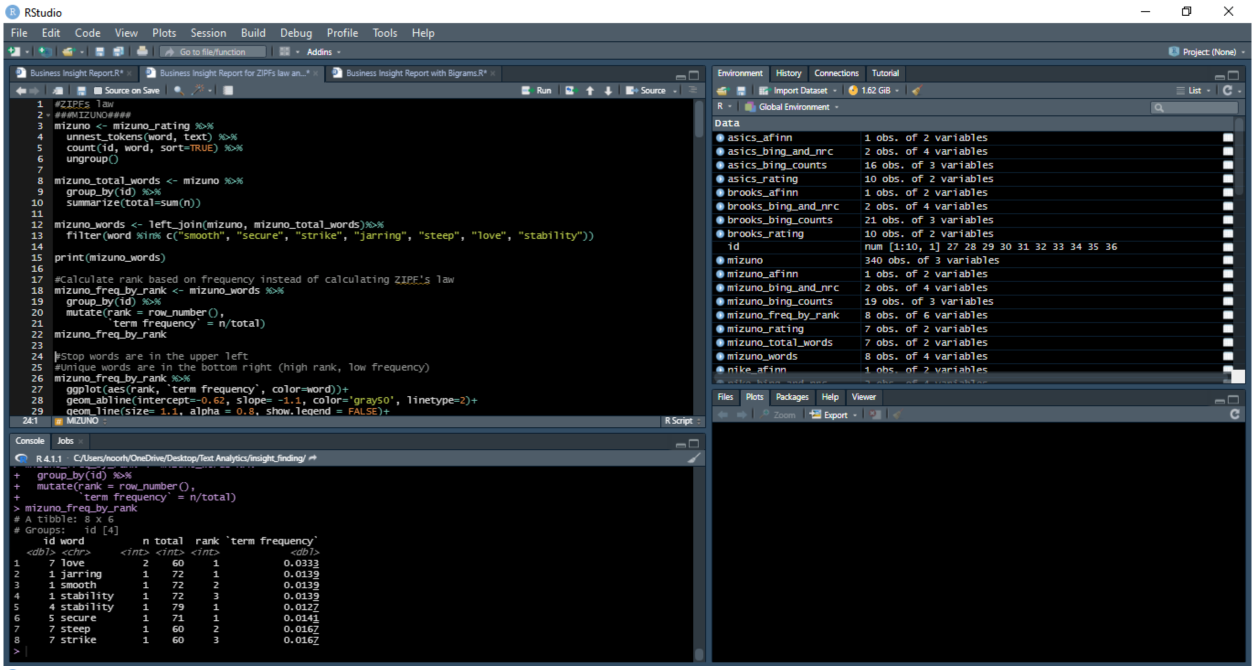

RStudio
nike_afinn <- tidy_nike %>%
inner_join(get_sentiments("afinn"))%>%
summarise(sentiment=sum(value)) %>%
mutate(method="AFINN")
nike_bing_and_nrc <- bind_rows(
tidy_nike%>%
inner_join(get_sentiments("bing"))%>%
mutate(method = "Bing et al."),
tidy_brooks %>%
inner_join(get_sentiments("nrc") %>%
filter(sentiment %in% c("positive", "negative"))) %>%
mutate(method = "NRC")) %>%
count(method, sentiment) %>%
spread(sentiment, n, fill=0) %>%
mutate(sentiment = positive-negative)
library(ggplot2)
bind_rows(nike_afinn, nike_bing_and_nrc) %>%
ggplot(aes(method, sentiment, fill=method))+
geom_col(show.legend=FALSE)+
facet_wrap(~method, ncol =1, scales= "free_y")
#Most common positive and negative words for nike
nike_bing_counts <- tidy_nike %>%
inner_join(get_sentiments("bing")) %>%
count(word, sentiment, sort=T) %>%
AFINN
Bing et al.
NRC
sentiment
method

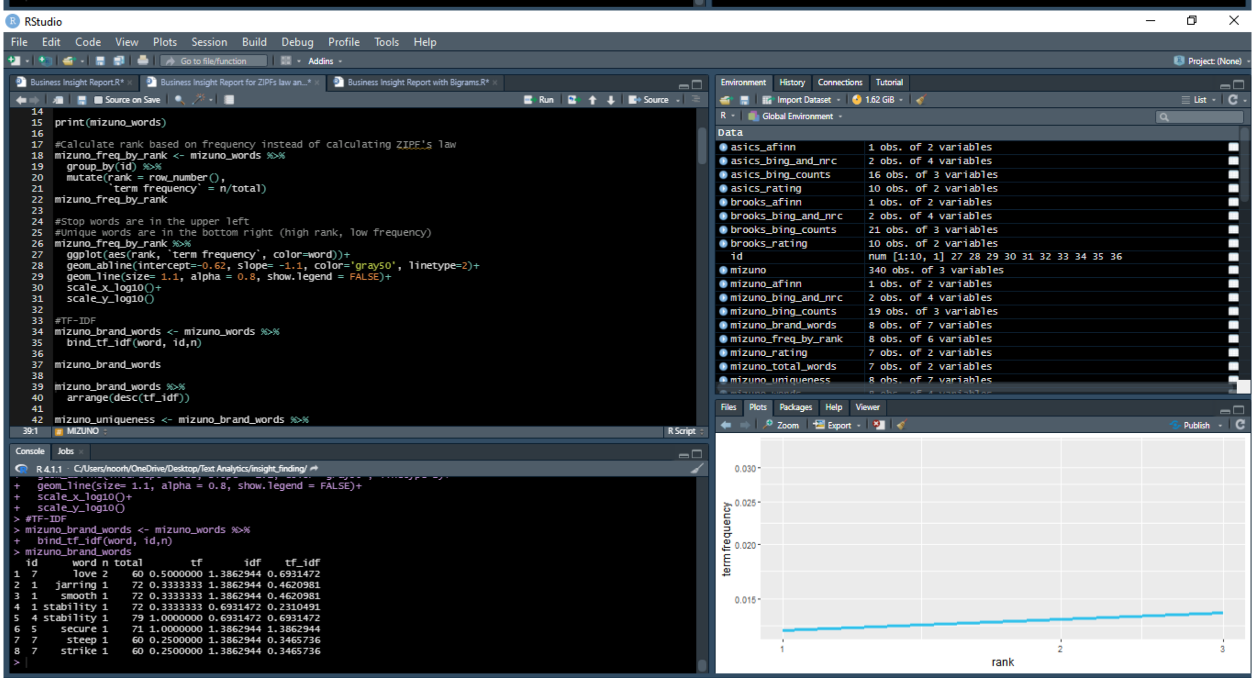

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
ggplot(aes(method, sentiment, fill=method))+
geom_col(show.legend=FALSE)+
facet_wrap(~method, ncol =1, scales= "free_y")
#Most common positive and negative words for nike
nike_bing_counts <- tidy_nike %>%
inner_join(get_sentiments("bing")) %>%
count(word, sentiment, sort=T) %>%
ungroup()
nike_bing_counts
nike_bing_counts %>%
group_by(sentiment) %>%
top_n(10) %>%
ungroup() %>%
mutate(word=reorder(word, n)) %>%
ggplot(aes(word, n, fill=sentiment)) +
geom_col(show.legend = FALSE) +
facet_wrap(~sentiment, scales = "free_y")+
labs(y="Contribution to sentiment", x=NULL)+
coord_flip()
###MIZUNO####
#Getting sentiment analysis for mizuno
mizuno_afinn <- tidy_mizuno %>%
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
10 stylish positive 1
11 support positive 1
12 thankful positive 1
13 tough positive 1
14 versatile positive 1
Selecting by n
Environment History Connections Tutorial
Import Dataset 1.62 GiB
Global Environment
Data
asics_afinn 1 obs. of 2 variables
asics_bing_and_nrc 2 obs. of 4 variables
asics_bing_counts 16 obs. of 3 variables
asics_rating 10 obs. of 2 variables
brooks_afinn 1 obs. of 2 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_rating 10 obs. of 2 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno_rating 7 obs. of 2 variables
nike_afinn 1 obs. of 2 variables
nike_bing_and_nrc 2 obs. of 4 variables
nike_bing_counts 14 obs. of 3 variables
nike_rating 9 obs. of 2 variables
tidy_asics 89 obs. of 2 variables
tidy_brooks 98 obs. of 2 variables
tidy_mizuno 132 obs. of 2 variables
Files Plots Packages Help Viewer
Zoom Export Publish
negative
positive
struggles
lost
versatile
tough
thankful
support
stylish
softer
love
faster
easy
durable
comfortable
adequate
0.00 0.25 0.50 0.75 1.00
Contribution to sentiment

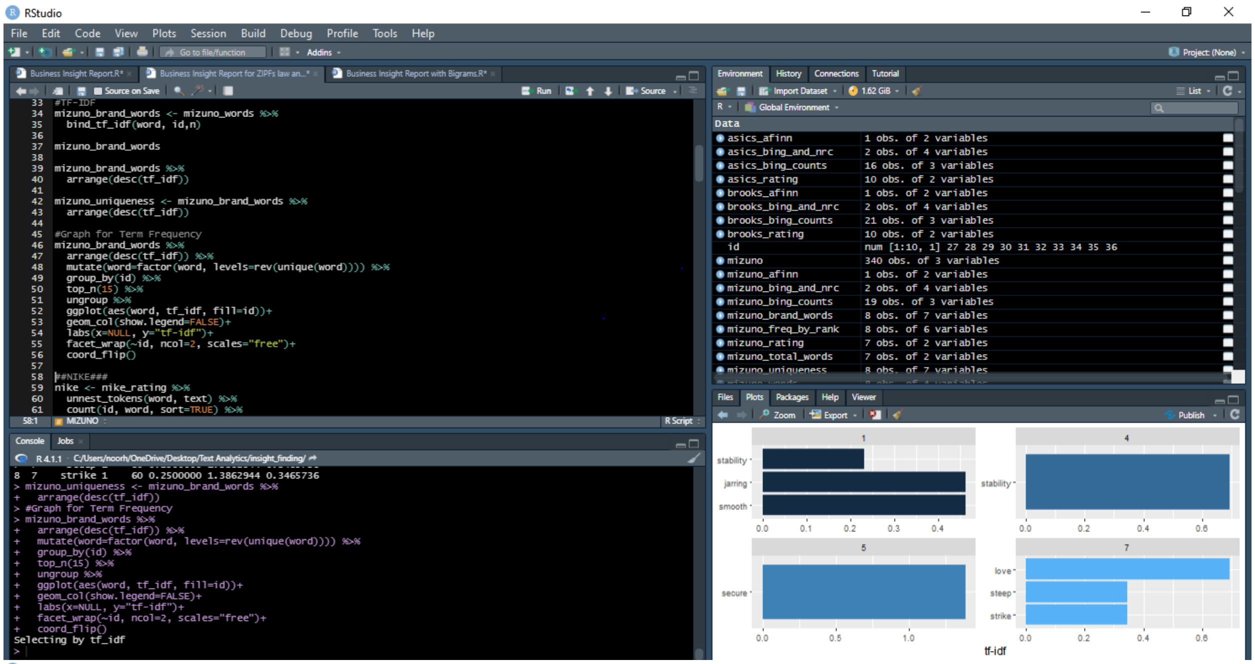

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
mizuno_afinn <- tidy_mizuno %>%
inner_join(get_sentiments("afinn"))%>%
summarise(sentiment=sum(value)) %>%
mutate(method="AFINN")
mizuno_bing_and_nrc <- bind_rows(
tidy_mizuno%>%
inner_join(get_sentiments("bing"))%>%
mutate(method = "Bing et al."),
tidy_mizuno %>%
inner_join(get_sentiments("nrc") %>%
filter(sentiment %in% c("positive", "negative"))) %>%
mutate(method = "NRC")) %>%
count(method, sentiment) %>%
spread(sentiment, n, fill=0) %>%
mutate(sentiment = positive-negative)
library(ggplot2)
bind_rows(mizuno_afinn, mizuno_bing_and_nrc) %>%
ggplot(aes(method, sentiment, fill=method))+
geom_col(show.legend=FALSE)+
facet_wrap(~method, ncol =1, scales= "free_y")
#Most common positive and negative words for mizuno
mizuno_bing_counts <- tidy_mizuno %>%
inner_join(get_sentiments("bing")) %>%
count(word, sentiment, sort=T) %>%
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
Joining, by = "word"
Joining, by = "word"
Environment History Connections Tutorial
Global Environment
AFINN
Bing et al.
NRC
sentiment
method

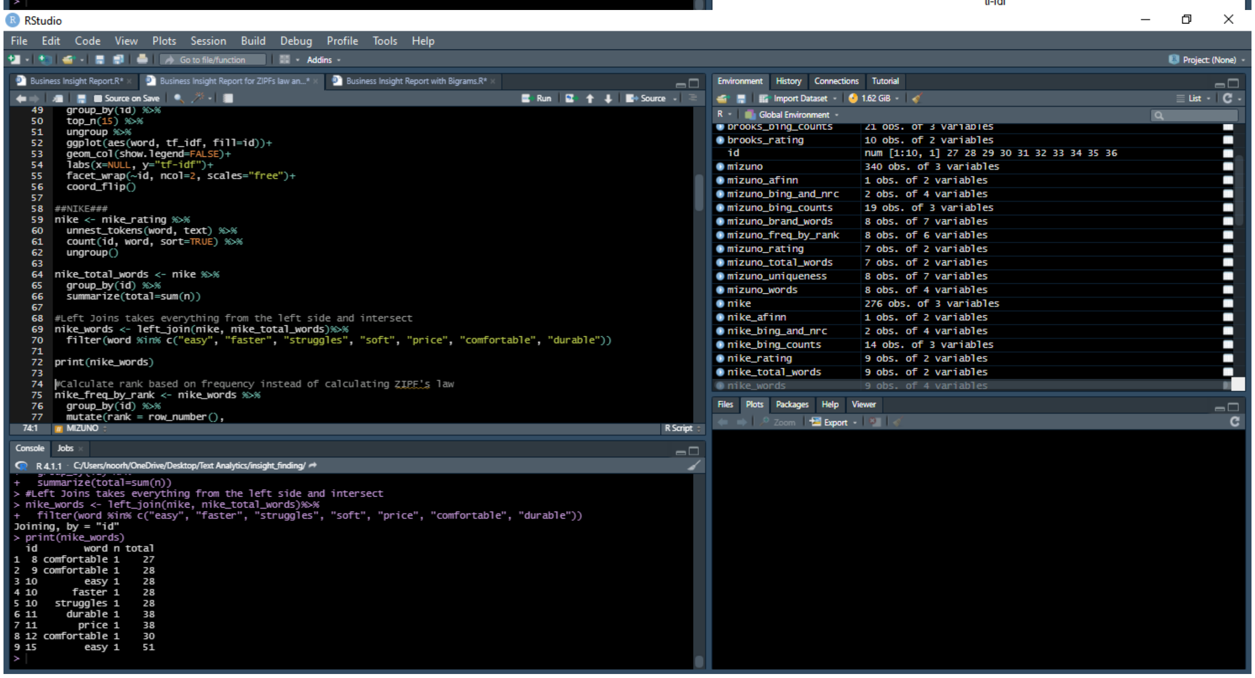

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
library(ggplot2)
bind_rows(mizuno_afinn, mizuno_bing_and_nrc) %>%
ggplot(aes(method, sentiment, fill=method))+
geom_col(show.legend=FALSE)+
facet_wrap(~method, ncol =1, scales= "free_y")
#Most common positive and negative words for mizuno
mizuno_bing_counts <- tidy_mizuno %>%
inner_join(get_sentiments("bing")) %>%
count(word, sentiment, sort=T) %>%
ungroup()
mizuno_bing_counts
mizuno_bing_counts %>%
group_by(sentiment) %>%
top_n(10) %>%
ungroup() %>%
mutate(word=reorder(word, n)) %>%
ggplot(aes(word, n, fill=sentiment)) +
geom_col(show.legend = FALSE) +
facet_wrap(~sentiment, scales = "free_y")+
labs(y="Contribution to sentiment", x=NULL)+
coord_flip()
Environment History Connections Tutorial
Global Environment
Data
asics_afinn 1 obs. of 2 variables
asics_bing_and_nrc 2 obs. of 4 variables
asics_bing_counts 16 obs. of 3 variables
asics_rating 10 obs. of 2 variables
brooks_afinn 1 obs. of 2 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_rating 10 obs. of 2 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno_afinn 1 obs. of 2 variables
mizuno_bing_and_nrc 2 obs. of 4 variables
mizuno_bing_counts 19 obs. of 3 variables
mizuno_rating 7 obs. of 2 variables
nike_afinn 1 obs. of 2 variables
nike_bing_and_nrc 2 obs. of 4 variables
nike_bing_counts 14 obs. of 3 variables
nike_rating 9 obs. of 2 variables
tidy_asics 89 obs. of 2 variables
Console Jobs
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
15 stability positive 1
16 steep negative 1
17 strike negative 1
18 top positive 1
19 tough positive 1
> mizuno_bing_counts %>%
+ group_by(sentiment) %>%
+ top_n(10) %>%
+ ungroup() %>%
+ mutate(word=reorder(word, n)) %>%
+ ggplot(aes(word, n, fill=sentiment)) +
+ geom_col(show.legend = FALSE) +
+ facet_wrap(~sentiment, scales = "free_y")+
+ labs(y="Contribution to sentiment", x=NULL)+
+ coord_flip()
Selecting by n
Files Plots Packages Help Viewer
negative
positive
strike
steep
jarring
complaints
tough
top
stability
soft
smooth
secure
reasonable
love
liking
inspire
improved
impressed
energetic
decent
classic
0.00 0.25 0.50 0.75 1.00
Contribution to sentiment

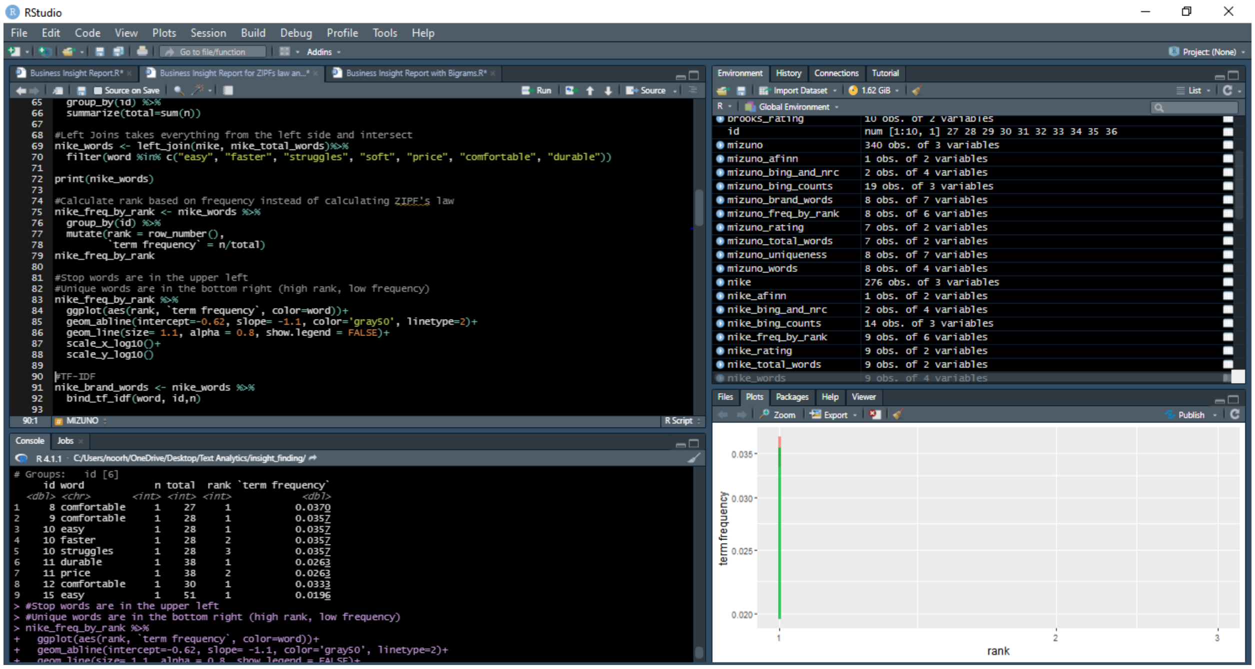

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
#ZIPFs law
###MIZUNO####
mizuno <- mizuno_rating %>%
unnest_tokens(word, text) %>%
count(id, word, sort=TRUE) %>%
ungroup()
mizuno_total_words <- mizuno %>%
group_by(id) %>%
summarize(total=sum(n))
mizuno_words <- left_join(mizuno, mizuno_total_words)%>%
filter(word %in% c("smooth", "secure", "strike", "jarring", "steep", "love", "stability"))
print(mizuno_words)
#Calculate rank based on frequency instead of calculating ZIPF's law
mizuno_freq_by_rank <- mizuno_words %>%
group_by(id) %>%
mutate(rank = row_number(),
`term frequency` = n/total)
mizuno_freq_by_rank
#Stop words are in the upper left
#Unique words are in the bottom right (high rank, low frequency)
mizuno_freq_by_rank %>%
ggplot(aes(rank, `term frequency`, color=word))+
geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
Global Environment
Data
asics_afinn 1 obs. of 2 variables
asics_bing_and_nrc 2 obs. of 4 variables
asics_bing_counts 16 obs. of 3 variables
asics_rating 10 obs. of 2 variables
brooks_afinn 1 obs. of 2 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_rating 10 obs. of 2 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno 340 obs. of 3 variables
mizuno_afinn 1 obs. of 2 variables
mizuno_bing_and_nrc 2 obs. of 4 variables
mizuno_bing_counts 19 obs. of 3 variables
mizuno_rating 7 obs. of 2 variables
mizuno_total_words 7 obs. of 2 variables
mizuno_words 8 obs. of 4 variables
nike_afinn 1 obs. of 2 variables
nike_bing_and_nrc 2 obs. of 4 variables
Console Jobs
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
> mizuno_total_words <- mizuno %>%
+ group_by(id) %>%
+ summarize(total=sum(n))
> mizuno_words <- left_join(mizuno, mizuno_total_words)%>%
+ filter(word %in% c("smooth", "secure", "strike", "jarring", "steep", "love", "stability"))
Joining, by = "id"
> print(mizuno_words)
id word n total
1 7 love 2 60
2 1 jarring 1 72
3 1 smooth 1 72
4 1 stability 1 72
5 4 stability 1 79
6 5 secure 1 71
7 7 steep 1 60
8 7 strike 1 60

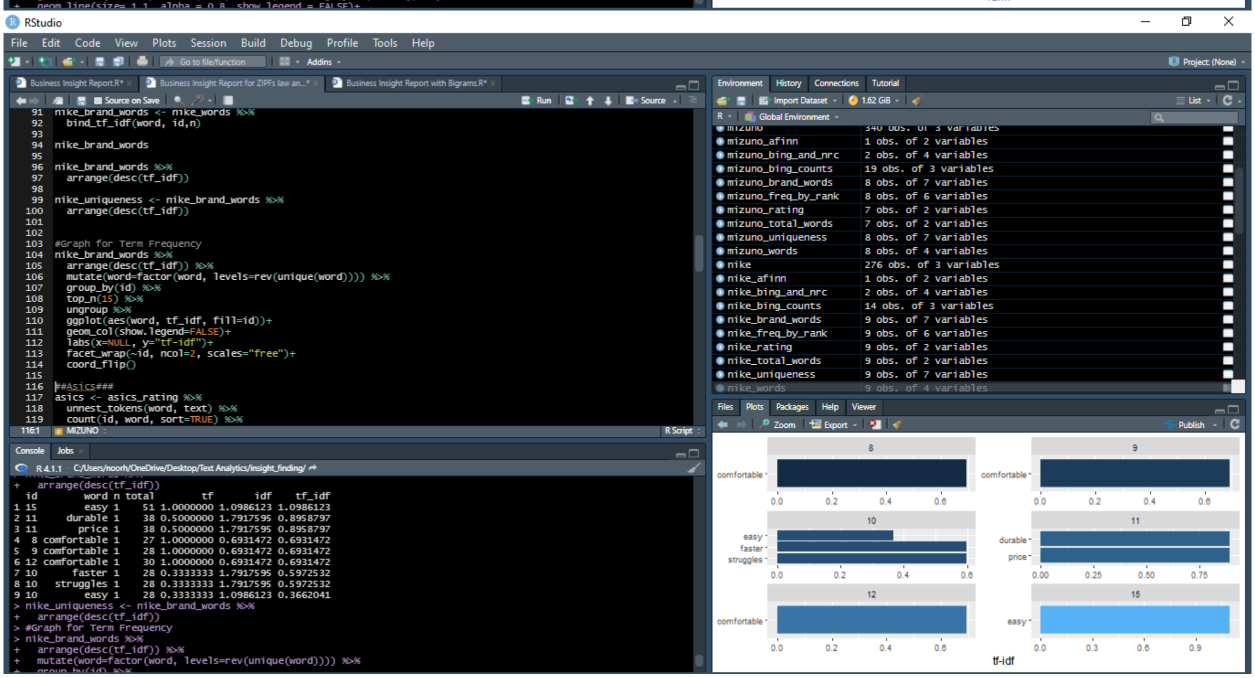

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
#ZIPFs law
###MIZUNO####
mizuno <- mizuno_rating %>%
unnest_tokens(word, text) %>%
count(id, word, sort=TRUE) %>%
ungroup()
mizuno_total_words <- mizuno %>%
group_by(id) %>%
summarize(total=sum(n))
mizuno_words <- left_join(mizuno, mizuno_total_words)%>%
filter(word %in% c("smooth", "secure", "strike", "jarring", "steep", "love", "stability"))
print(mizuno_words)
#Calculate rank based on frequency instead of calculating ZIPF's law
mizuno_freq_by_rank <- mizuno_words %>%
group_by(id) %>%
mutate(rank = row_number(),
`term frequency` = n/total)
mizuno_freq_by_rank
#Stop words are in the upper left
#Unique words are in the bottom right (high rank, low frequency)
mizuno_freq_by_rank %>%
ggplot(aes(rank, `term frequency`, color=word))+
geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
A tibble: 8 x 6
Groups: id [4]
id word n total rank `term frequency`
1 7 love 2 60 1 0.0333
2 1 jarring 1 72 1 0.0139
3 1 smooth 1 72 2 0.0139
4 1 stability 1 72 3 0.0139
5 4 stability 1 79 1 0.0127
6 5 secure 1 71 1 0.0141
7 7 steep 1 60 2 0.0167
8 7 strike 1 60 3 0.0167
Environment History Connections Tutorial
Global Environment
Data
asics_afinn 1 obs. of 2 variables
asics_bing_and_nrc 2 obs. of 4 variables
asics_bing_counts 16 obs. of 3 variables
asics_rating 10 obs. of 2 variables
brooks_afinn 1 obs. of 2 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_rating 10 obs. of 2 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno 340 obs. of 3 variables
mizuno_afinn 1 obs. of 2 variables
mizuno_bing_and_nrc 2 obs. of 4 variables
mizuno_bing_counts 19 obs. of 3 variables
mizuno_freq_by_rank 8 obs. of 6 variables
mizuno_rating 7 obs. of 2 variables
mizuno_total_words 7 obs. of 2 variables
mizuno_words 8 obs. of 4 variables

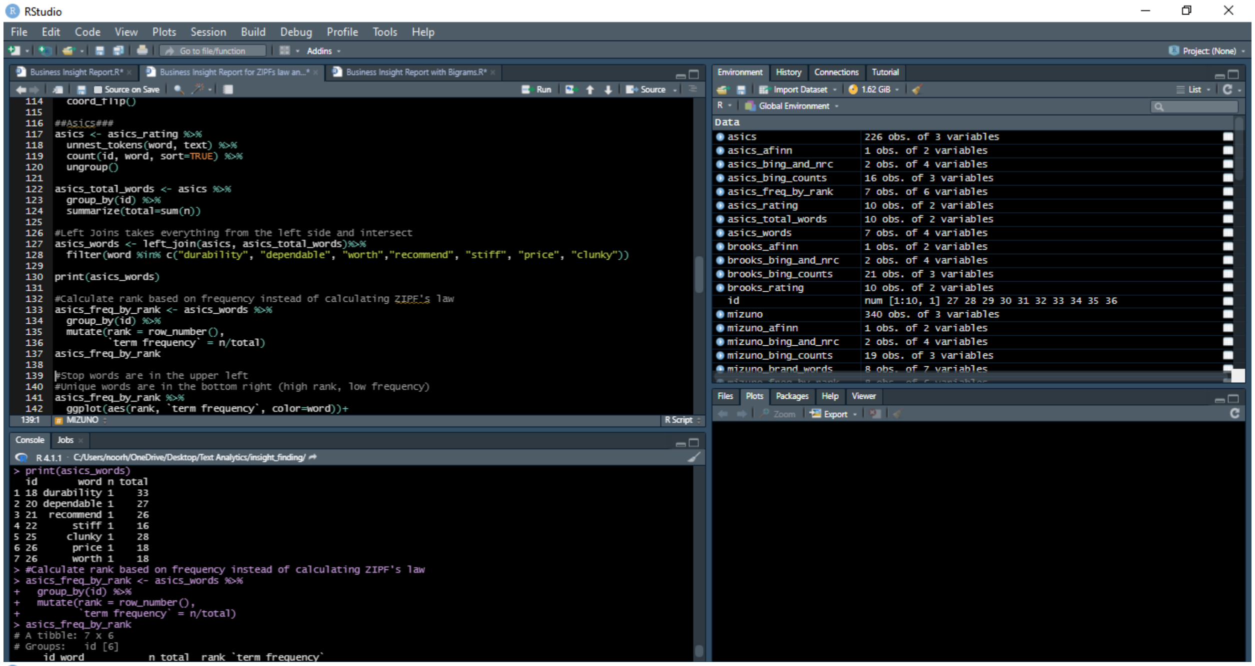

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
print(mizuno_words)
#Calculate rank based on frequency instead of calculating ZIPF's law
mizuno_freq_by_rank <- mizuno_words %>%
group_by(id) %>%
mutate(rank = row_number(),
`term frequency` = n/total)
mizuno_freq_by_rank
#Stop words are in the upper left
#Unique words are in the bottom right (high rank, low frequency)
mizuno_freq_by_rank %>%
ggplot(aes(rank, `term frequency`, color=word))+
geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
scale_x_log10()+
scale_y_log10()
#TF-IDF
mizuno_brand_words <- mizuno_words %>%
bind_tf_idf(word, id,n)
mizuno_brand_words
mizuno_brand_words %>%
arrange(desc(tf_idf))
mizuno_uniqueness <- mizuno_brand_words %>%
id word n total tf idf tf_idf
1 7 love 2 60 0.5000000 1.3862944 0.6931472
2 1 jarring 1 72 0.3333333 1.3862944 0.4620981
3 1 smooth 1 72 0.3333333 1.3862944 0.4620981
4 1 stability 1 72 0.3333333 0.6931472 0.2310491
5 4 stability 1 79 1.0000000 0.6931472 0.6931472
6 5 secure 1 71 1.0000000 1.3862944 1.3862944
7 7 steep 1 60 0.2500000 1.3862944 0.3465736
8 7 strike 1 60 0.2500000 1.3862944 0.3465736
mizuno_brand_words 8 obs. of 7 variables
mizuno_uniqueness 8 obs. of 7 variables
term frequency
rank

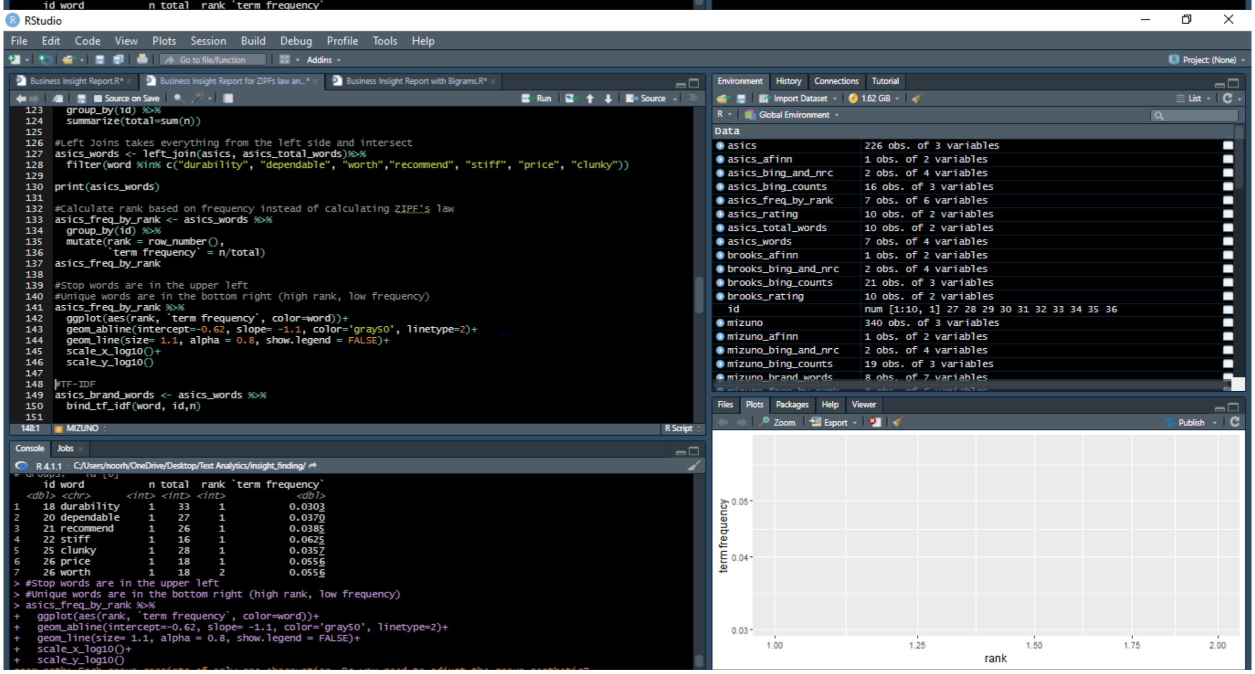

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
#TF-IDF
mizuno_brand_words <- mizuno_words %>%
bind_tf_idf(word, id,n)
mizuno_brand_words
mizuno_brand_words %>%
arrange(desc(tf_idf))
mizuno_uniqueness <- mizuno_brand_words %>%
arrange(desc(tf_idf))
#Graph for Term Frequency
mizuno_brand_words %>%
arrange(desc(tf_idf)) %>%
mutate(word=factor(word, levels=rev(unique(word)))) %>%
group_by(id) %>%
top_n(15) %>%
ungroup %>%
ggplot(aes(word, tf_idf, fill=id))+
geom_col(show.legend=FALSE)+
labs(x=NULL, y="tf-idf")+
facet_wrap(~id, ncol=2, scales="free")+
coord_flip()
##NIKE###
nike <- nike_rating %>%
unnest_tokens(word, text) %>%
count(id, word, sort=TRUE) %>%
Environment History Connections Tutorial
Global Environment
Data
asics_afinn 1 obs. of 2 variables
asics_bing_and_nrc 2 obs. of 4 variables
asics_bing_counts 16 obs. of 3 variables
asics_rating 10 obs. of 2 variables
brooks_afinn 1 obs. of 2 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_rating 10 obs. of 2 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno 340 obs. of 3 variables
mizuno_afinn 1 obs. of 2 variables
mizuno_bing_and_nrc 2 obs. of 4 variables
mizuno_bing_counts 19 obs. of 3 variables
mizuno_brand_words 8 obs. of 7 variables
mizuno_freq_by_rank 8 obs. of 6 variables
mizuno_rating 7 obs. of 2 variables
mizuno_total_words 7 obs. of 2 variables
mizuno_uniqueness 8 obs. of 7 variables
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
8 7 strike 1 60 0.2500000 1.3862944 0.3465736
> mizuno_uniqueness <- mizuno_brand_words %>%
+ arrange(desc(tf_idf))
> #Graph for Term Frequency
> mizuno_brand_words %>%
+ arrange(desc(tf_idf)) %>%
+ mutate(word=factor(word, levels=rev(unique(word)))) %>%
+ group_by(id) %>%
+ top_n(15) %>%
+ ungroup %>%
+ ggplot(aes(word, tf_idf, fill=id))+
+ geom_col(show.legend=FALSE)+
+ labs(x=NULL, y="tf-idf")+
+ facet_wrap(~id, ncol=2, scales="free")+
+ coord_flip()
Selecting by tf_idf
1
stability
jarring
smooth
4
stability
5
secure
7
love
steep
strike
tf-idf

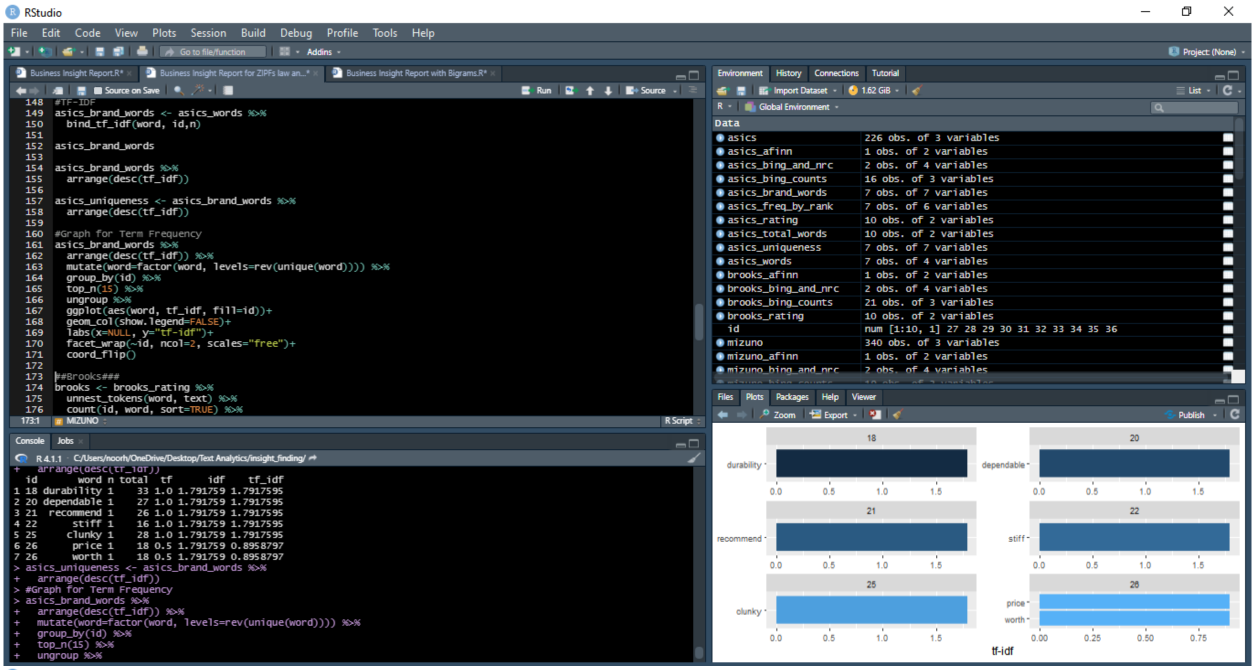

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
group_by(id) %>%
top_n(15) %>%
ungroup %>%
ggplot(aes(word, tf_idf, fill=id))+
geom_col(show.legend=FALSE)+
labs(x=NULL, y="tf-idf")+
facet_wrap(~id, ncol=2, scales="free")+
coord_flip()
##NIKE###
nike <- nike_rating %>%
unnest_tokens(word, text) %>%
count(id, word, sort=TRUE) %>%
ungroup()
nike_total_words <- nike %>%
group_by(id) %>%
summarize(total=sum(n))
#Left Joins takes everything from the left side and intersect
nike_words <- left_join(nike, nike_total_words)%>%
filter(word %in% c("easy", "faster", "struggles", "soft", "price", "comfortable", "durable"))
print(nike_words)
#Calculate rank based on frequency instead of calculating ZIPF's law
nike_freq_by_rank <- nike_words %>%
group_by(id) %>%
mutate(rank = row_number(),
brooks_rating 10 obs. of 2 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno 340 obs. of 3 variables
mizuno_afinn 1 obs. of 2 variables
mizuno_bing_and_nrc 2 obs. of 4 variables
mizuno_bing_counts 19 obs. of 3 variables
mizuno_brand_words 8 obs. of 7 variables
mizuno_freq_by_rank 8 obs. of 6 variables
mizuno_rating 7 obs. of 2 variables
mizuno_total_words 7 obs. of 2 variables
mizuno_uniqueness 8 obs. of 7 variables
mizuno_words 8 obs. of 4 variables
nike 276 obs. of 3 variables
nike_afinn 1 obs. of 2 variables
nike_bing_and_nrc 2 obs. of 4 variables
nike_bing_counts 14 obs. of 3 variables
nike_rating 9 obs. of 2 variables
nike_total_words 9 obs. of 2 variables
nike_words 9 obs. of 4 variables
Console Jobs
+ summarize(total=sum(n))
> #Left Joins takes everything from the left side and intersect
> nike_words <- left_join(nike, nike_total_words)%>%
+ filter(word %in% c("easy", "faster", "struggles", "soft", "price", "comfortable", "durable"))
Joining, by = "id"
> print(nike_words)
id word n total
1 8 comfortable 1 27
2 9 comfortable 1 28
3 10 easy 1 28
4 10 faster 1 28
5 10 struggles 1 28
6 11 durable 1 38
7 11 price 1 38
8 12 comfortable 1 30
9 15 easy 1 51

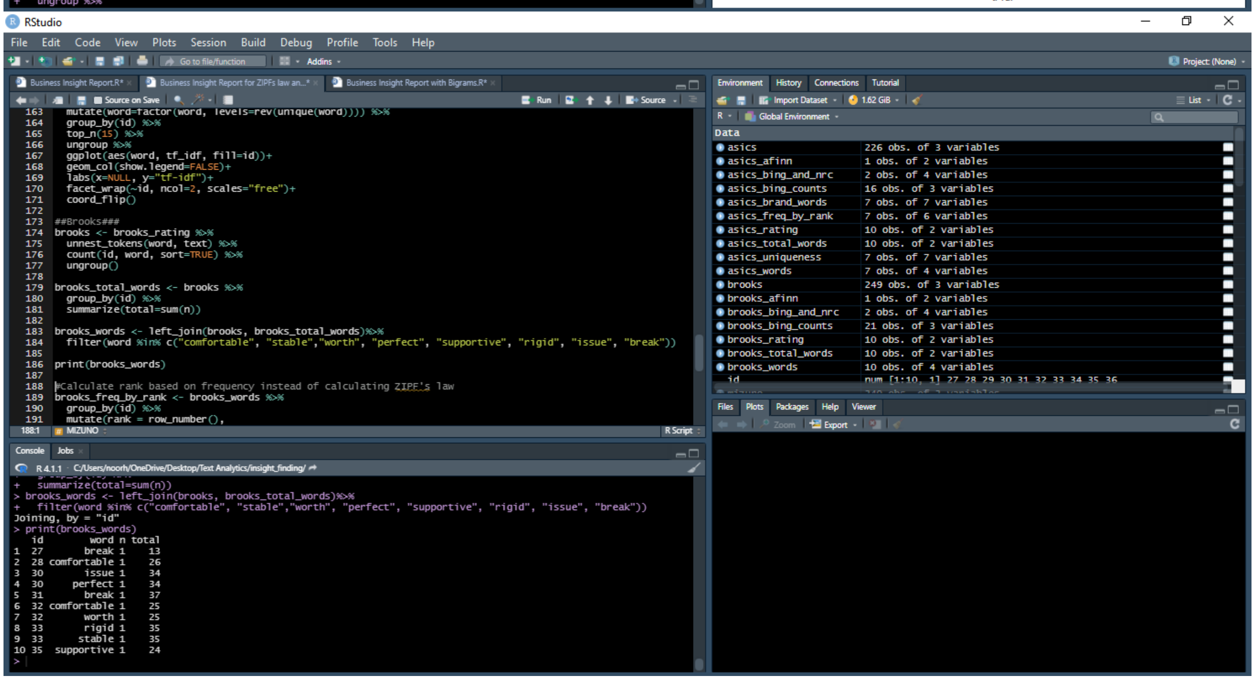

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
group_by(id) %>%
summarize(total=sum(n))
#Left Joins takes everything from the left side and intersect
nike_words <- left_join(nike, nike_total_words)%>%
filter(word %in% c("easy", "faster", "struggles", "soft", "price", "comfortable", "durable"))
print(nike_words)
#Calculate rank based on frequency instead of calculating ZIPF's law
nike_freq_by_rank <- nike_words %>%
group_by(id) %>%
mutate(rank = row_number(),
`term frequency` = n/total)
nike_freq_by_rank
#Stop words are in the upper left
#Unique words are in the bottom right (high rank, low frequency)
nike_freq_by_rank %>%
ggplot(aes(rank, `term frequency`, color=word))+
geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
scale_x_log10()+
scale_y_log10()
#TF-IDF
nike_brand_words <- nike_words %>%
bind_tf_idf(word, id,n)
mizuno
mizuno_afinn 1 obs. of 2 variables
mizuno_bing_and_nrc 2 obs. of 4 variables
mizuno_bing_counts 19 obs. of 3 variables
mizuno_brand_words 8 obs. of 7 variables
mizuno_freq_by_rank 8 obs. of 6 variables
mizuno_rating 7 obs. of 2 variables
mizuno_total_words 7 obs. of 2 variables
mizuno_uniqueness 8 obs. of 7 variables
mizuno_words 8 obs. of 4 variables
nike 276 obs. of 3 variables
nike_afinn 1 obs. of 2 variables
nike_bing_and_nrc 2 obs. of 4 variables
nike_bing_counts 14 obs. of 3 variables
nike_freq_by_rank 9 obs. of 6 variables
nike_rating 9 obs. of 2 variables
nike_total_words 9 obs. of 2 variables
Groups: id [6]
id word n total rank `term frequency`
1 8 comfortable 1 27 1 0.0370
2 9 comfortable 1 28 1 0.0357
3 10 easy 1 28 1 0.0357
4 10 faster 1 28 2 0.0357
5 10 struggles 1 28 3 0.0357
6 11 durable 1 38 1 0.0263
7 11 price 1 38 2 0.0263
8 12 comfortable 1 30 1 0.0333
9 15 easy 1 51 1 0.0196
term frequency
rank

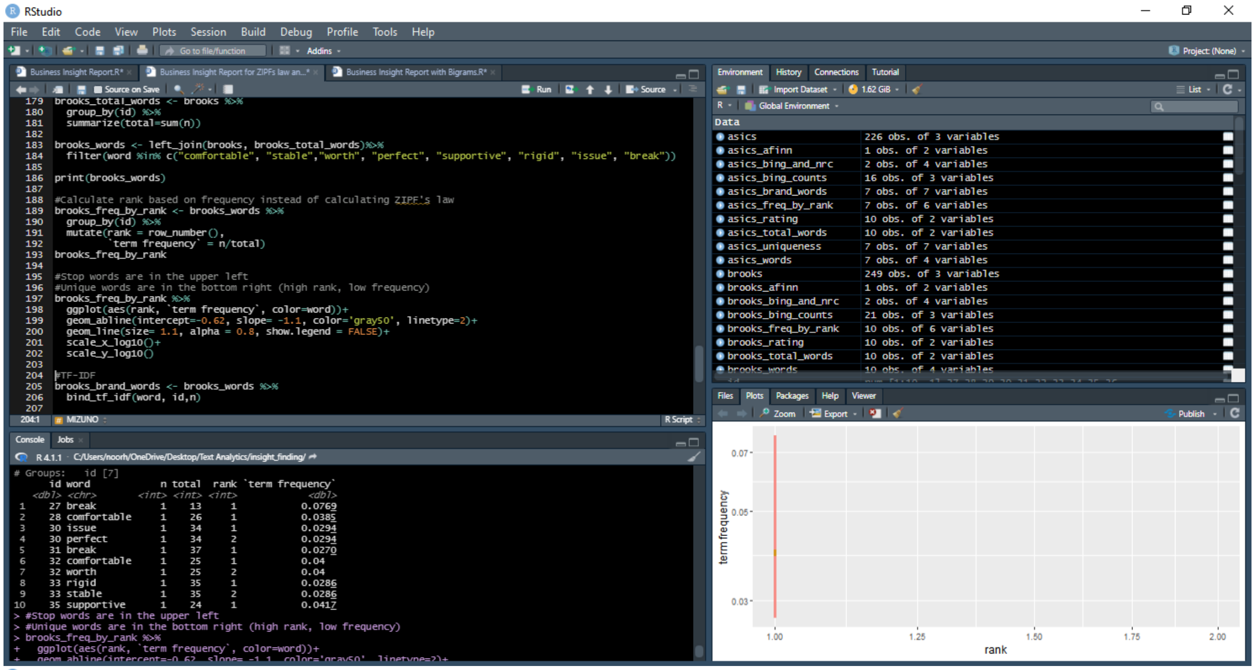

RStudio
nike_brand_words <- nike_words %>%
bind_tf_idf(word, id,n)
nike_brand_words
nike_brand_words %>%
arrange(desc(tf_idf))
nike_uniqueness <- nike_brand_words %>%
arrange(desc(tf_idf))
#Graph for Term Frequency
nike_brand_words %>%
arrange(desc(tf_idf)) %>%
mutate(word=factor(word, levels=rev(unique(word)))) %>%
group_by(id) %>%
top_n(15) %>%
ungroup %>%
ggplot(aes(word, tf_idf, fill=id))+
geom_col(show.legend=FALSE)+
labs(x=NULL, y="tf-idf")+
facet_wrap(~id, ncol=2, scales="free")+
coord_flip()
##asics###
asics <- asics_rating %>%
unnest_tokens(word, text) %>%
count(id, word, sort=TRUE) %>%
id word n total tf idf tf_idf
1 15 easy 1 51 1.0000000 1.0986123 1.0986123
2 11 durable 1 38 0.5000000 1.7917595 0.8958797
3 11 price 1 38 0.5000000 1.7917595 0.8958797
4 8 comfortable 1 27 1.0000000 0.6931472 0.6931472
5 9 comfortable 1 28 1.0000000 0.6931472 0.6931472
6 12 comfortable 1 30 1.0000000 0.6931472 0.6931472
7 10 faster 1 28 0.3333333 1.7917595 0.5972532
8 10 struggles 1 28 0.3333333 1.7917595 0.5972532
9 10 easy 1 28 0.3333333 1.0986123 0.3662041
comfortable
easy
faster
struggles
durable
price
tf-idf

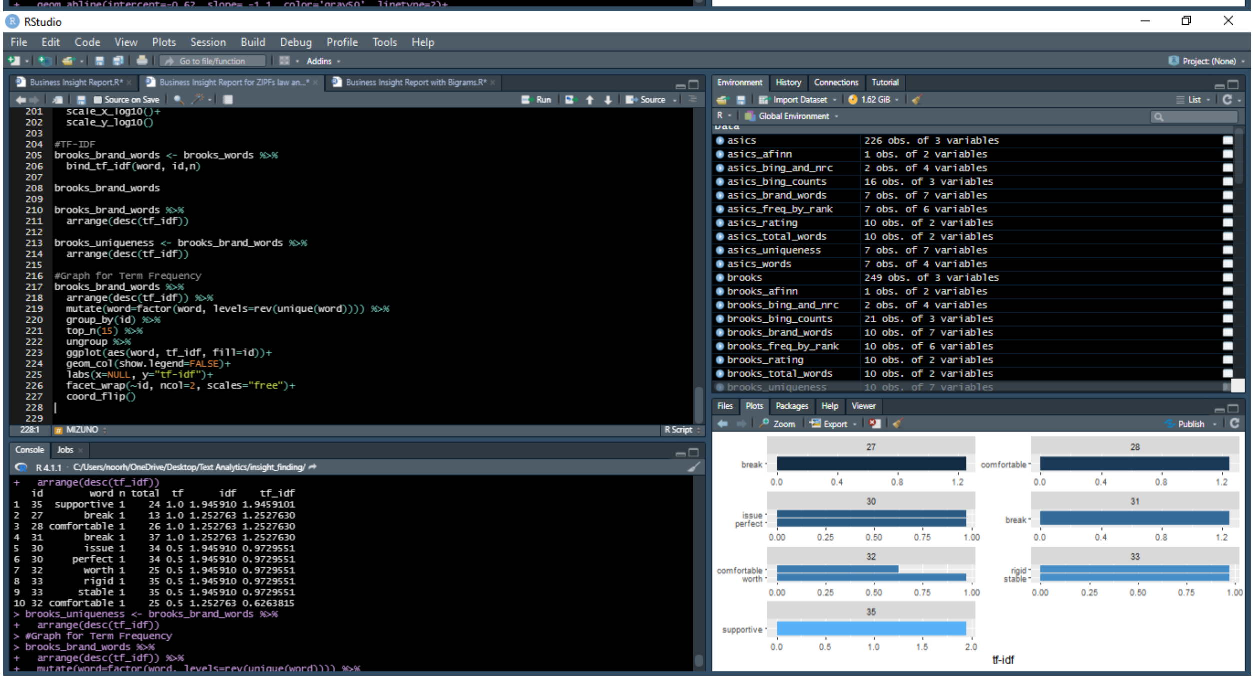


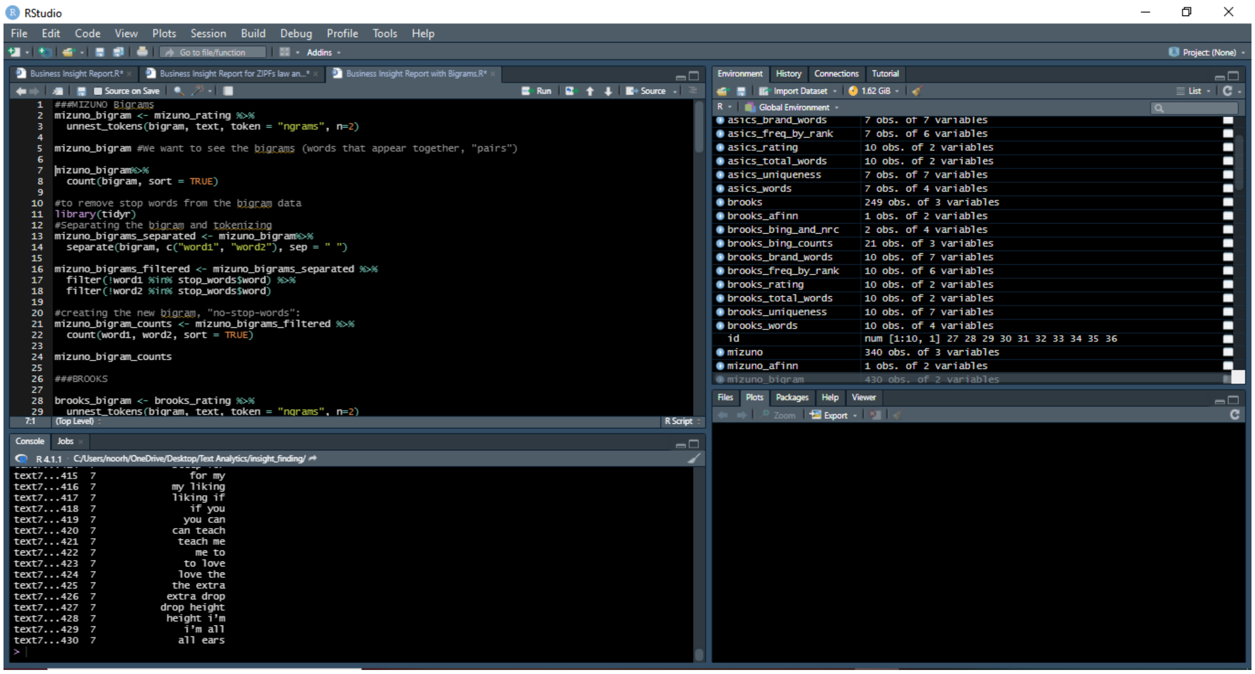


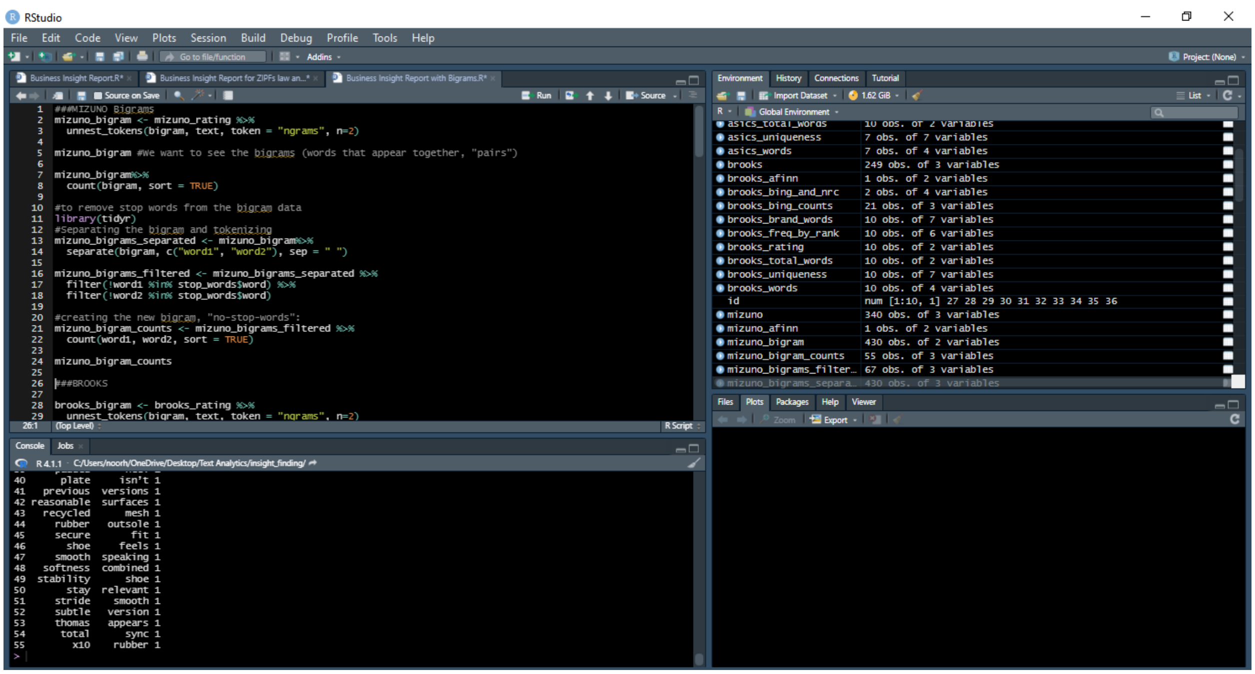

RStudio
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
#TF-IDF
asics_brand_words <- asics_words %>%
bind_tf_idf(word, id,n)
asics_brand_words
asics_brand_words %>%
arrange(desc(tf_idf))
asics_uniqueness <- asics_brand_words %>%
arrange(desc(tf_idf))
#Graph for Term Frequency
asics_brand_words %>%
arrange(desc(tf_idf)) %>%
mutate(word=factor(word, levels=rev(unique(word)))) %>%
group_by(id) %>%
top_n(15) %>%
ungroup %>%
ggplot(aes(word, tf_idf, fill=id))+
geom_col(show.legend=FALSE)+
labs(x=NULL, y="tf-idf")+
facet_wrap(~id, ncol=2, scales="free")+
coord_flip()
##Brooks###
brooks <- brooks_rating %>%
unnest_tokens(word, text) %>%
count(id, word, sort=TRUE) %>%
Console
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
id word n total tf idf tf_idf
1 18 durability 1 33 1.0 1.791759 1.7917595
2 20 dependable 1 27 1.0 1.791759 1.7917595
3 21 recommend 1 26 1.0 1.791759 1.7917595
4 22 stiff 1 16 1.0 1.791759 1.7917595
5 25 clunky 1 28 1.0 1.791759 1.7917595
6 26 price 1 18 0.5 1.791759 0.8958797
7 26 worth 1 18 0.5 1.791759 0.8958797
> asics_uniqueness <- asics_brand_words %>%
+ arrange(desc(tf_idf))
> #Graph for Term Frequency
> asics_brand_words %>%
+ arrange(desc(tf_idf)) %>%
+ mutate(word=factor(word, levels=rev(unique(word)))) %>%
+ group_by(id) %>%
+ top_n(15) %>%
+ ungroup %>%
18
20
21
22
25
26
durability
dependable
recommend
stiff
clunky
price
worth
tf-idf

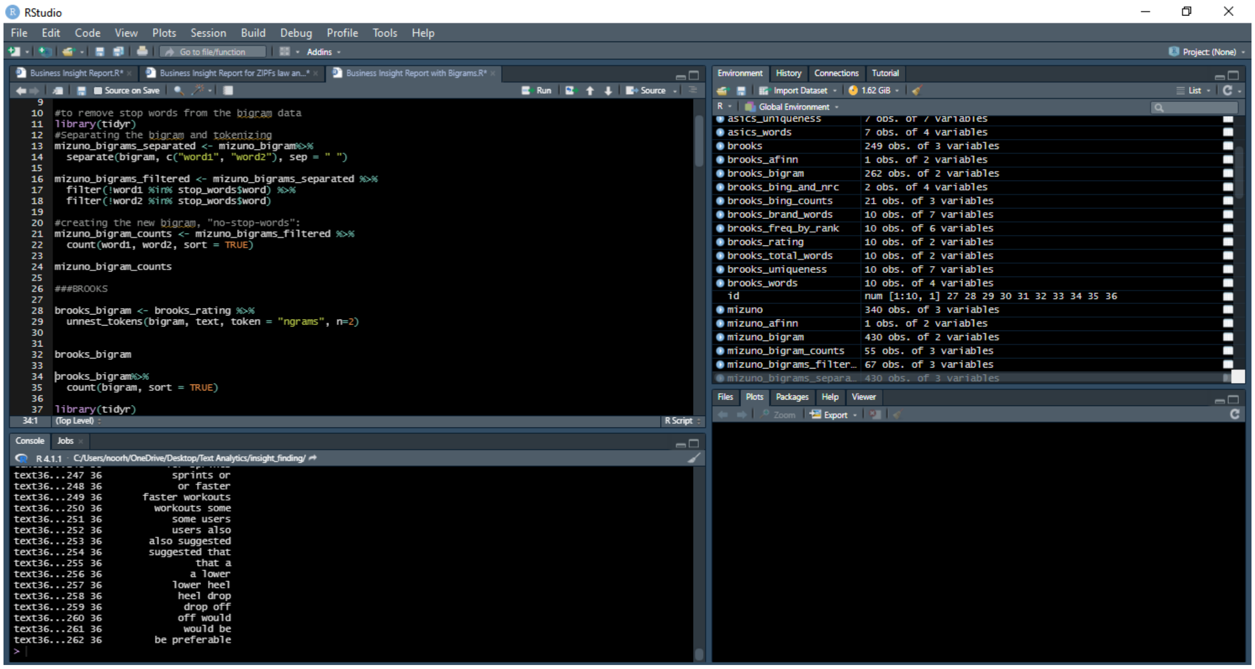

RStudio
##Brooks###
brooks <- brooks_rating %>%
unnest_tokens(word, text) %>%
count(id, word, sort=TRUE) %>%
ungroup()
brooks_total_words <- brooks %>%
group_by(id) %>%
summarize(total=sum(n))
brooks_words <- left_join(brooks, brooks_total_words)%>%
filter(word %in% c("comfortable", "stable","worth", "perfect", "supportive", "rigid", "issue", "break"))
print(brooks_words)
#Calculate rank based on frequency instead of calculating ZIPF's law
brooks_freq_by_rank <- brooks_words %>%
group_by(id) %>%
mutate(rank = row_number(),
Console
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
+ summarize(total=sum(n))
> brooks_words <- left_join(brooks, brooks_total_words)%>%
+ filter(word %in% c("comfortable", "stable","worth", "perfect", "supportive", "rigid", "issue", "break"))
Joining, by = "id"
> print(brooks_words)
id word n total
1 27 break 1 13
2 28 comfortable 1 26
3 30 issue 1 34
4 30 perfect 1 34
5 31 break 1 37
6 32 comfortable 1 25
7 32 worth 1 25
8 33 rigid 1 35
9 33 stable 1 35
10 35 supportive 1 24

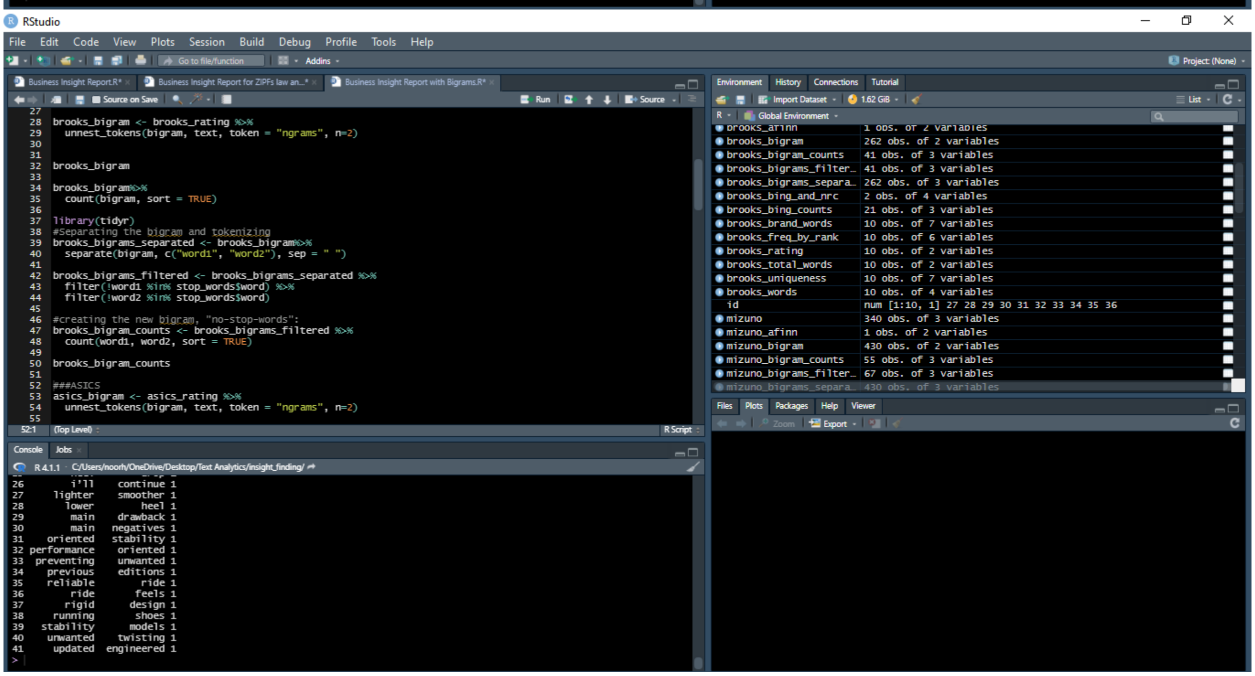


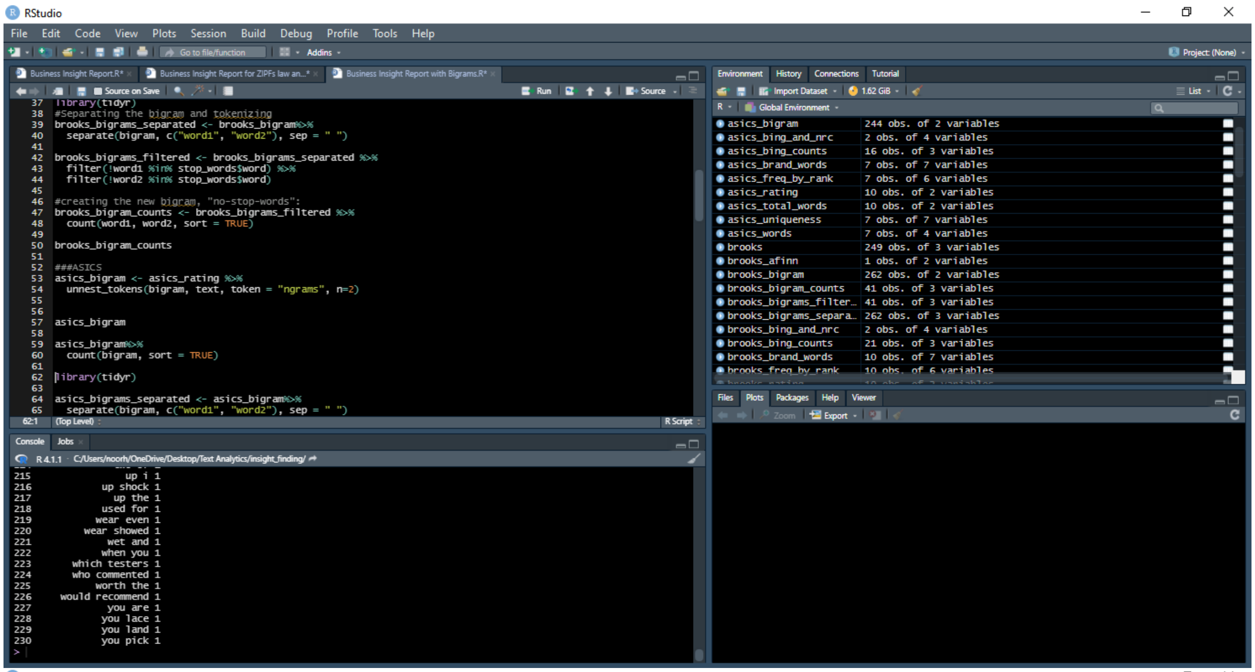


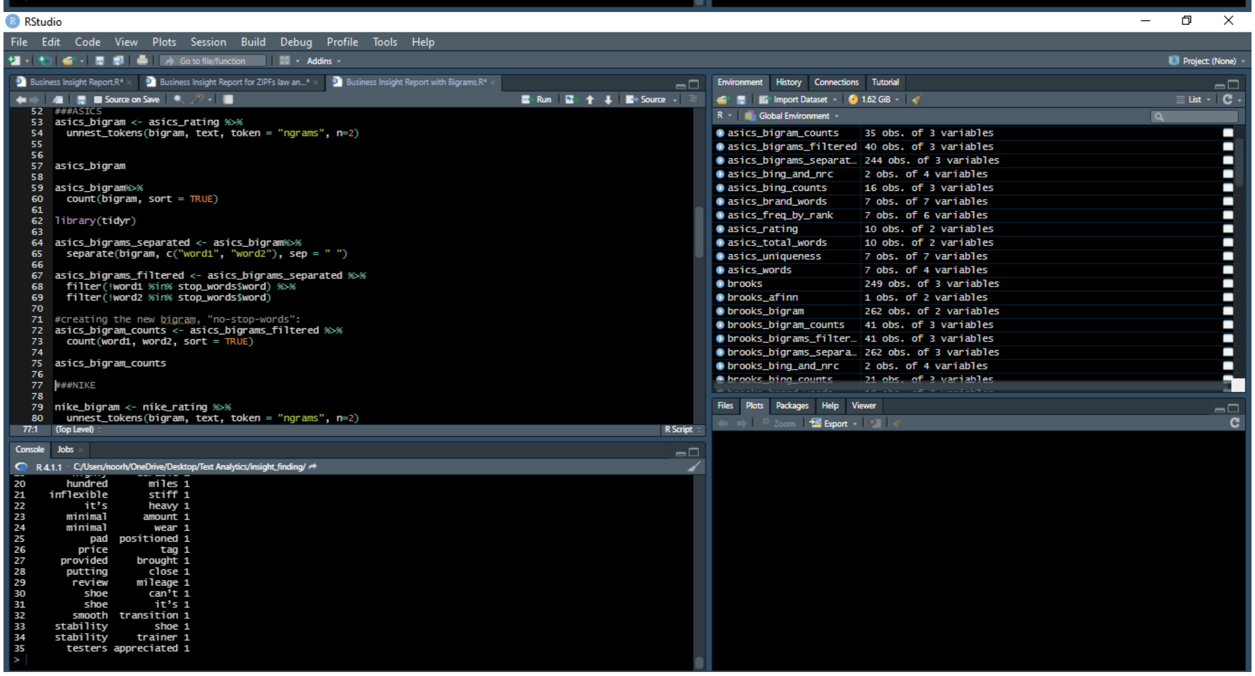

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
###MIZUNO Bigrams
mizuno_bigram <- mizuno_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
mizuno_bigram #We want to see the bigrams (words that appear together, "pairs")
mizuno_bigram%>%
count(bigram, sort = TRUE)
#to remove stop words from the bigram data
library(tidyr)
#Separating the bigram and tokenizing
mizuno_bigrams_separated <- mizuno_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
mizuno_bigrams_filtered <- mizuno_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
mizuno_bigram_counts <- mizuno_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
mizuno_bigram_counts
###BROOKS
brooks_bigram <- brooks_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
text7...415 7 for my
text7...416 7 my liking
text7...417 7 liking if
text7...418 7 if you
text7...419 7 you can
text7...420 7 can teach
text7...421 7 teach me
text7...422 7 me to
text7...423 7 to love
text7...424 7 love the
text7...425 7 the extra
text7...426 7 extra drop
text7...427 7 drop height
text7...428 7 height i'm
text7...429 7 i'm all
text7...430 7 all ears
Environment History Connections Tutorial
asics_brand_words 7 obs. of 7 variables
asics_freq_by_rank 7 obs. of 6 variables
asics_rating 10 obs. of 2 variables
asics_total_words 10 obs. of 2 variables
asics_uniqueness 7 obs. of 7 variables
asics_words 7 obs. of 4 variables
brooks 249 obs. of 3 variables
brooks_afinn 1 obs. of 2 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_brand_words 10 obs. of 7 variables
brooks_freq_by_rank 10 obs. of 6 variables
brooks_rating 10 obs. of 2 variables
brooks_total_words 10 obs. of 2 variables
brooks_uniqueness 10 obs. of 7 variables
brooks_words 10 obs. of 4 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno 340 obs. of 3 variables
mizuno_afinn 1 obs. of 2 variables
mizuno_bigram 430 obs. of 2 variables

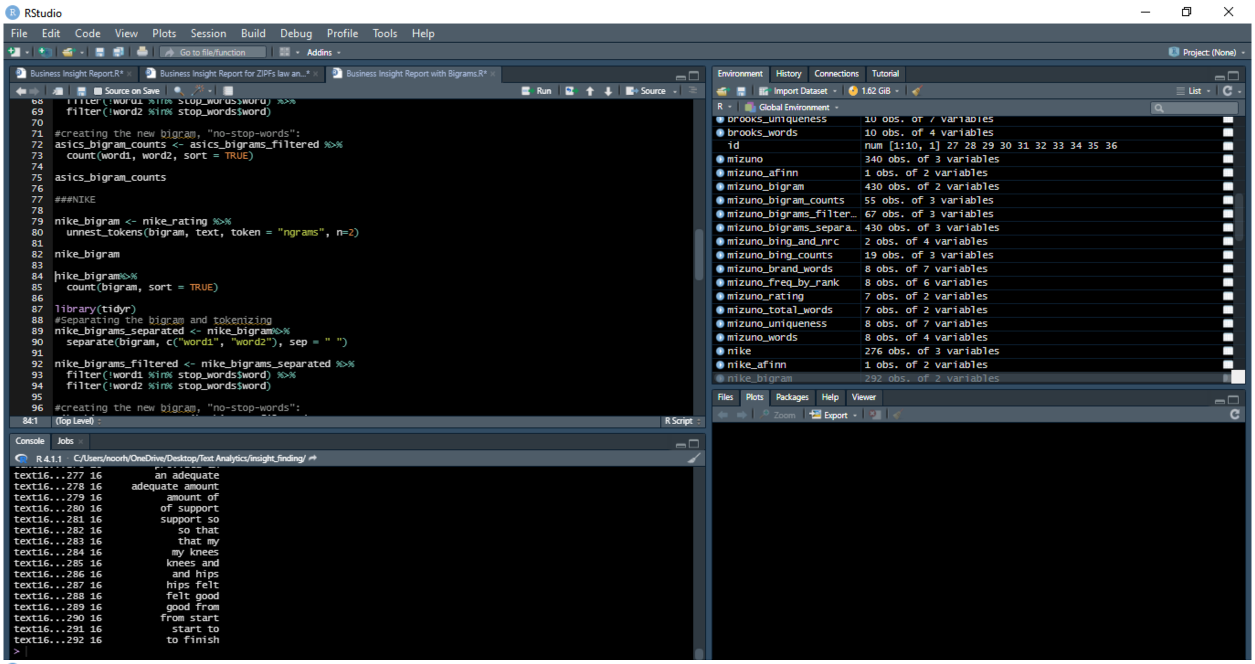

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
###MIZUNO Bigrams
mizuno_bigram <- mizuno_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
mizuno_bigram #We want to see the bigrams (words that appear together, "pairs")
mizuno_bigram%>%
count(bigram, sort = TRUE)
#to remove stop words from the bigram data
library(tidyr)
#Separating the bigram and tokenizing
mizuno_bigrams_separated <- mizuno_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
mizuno_bigrams_filtered <- mizuno_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
mizuno_bigram_counts <- mizuno_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
mizuno_bigram_counts
###BROOKS
brooks_bigram <- brooks_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
40 plate isn't 1
41 previous versions 1
42 reasonable surfaces 1
43 recycled mesh 1
44 rubber outsole 1
45 secure fit 1
46 shoe feels 1
47 smooth speaking 1
48 softness combined 1
49 stability shoe 1
50 stay relevant 1
51 stride smooth 1
52 subtle version 1
53 thomas appears 1
54 total sync 1
55 x10 rubber 1
Environment History Connections Tutorial
asics_uniqueness 7 obs. of 7 variables
asics_words 7 obs. of 4 variables
brooks 249 obs. of 3 variables
brooks_afinn 1 obs. of 2 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_brand_words 10 obs. of 7 variables
brooks_freq_by_rank 10 obs. of 6 variables
brooks_rating 10 obs. of 2 variables
brooks_total_words 10 obs. of 2 variables
brooks_uniqueness 10 obs. of 7 variables
brooks_words 10 obs. of 4 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno 340 obs. of 3 variables
mizuno_afinn 1 obs. of 2 variables
mizuno_bigram 430 obs. of 2 variables
mizuno_bigram_counts 55 obs. of 3 variables
mizuno_bigrams_filter… 67 obs. of 3 variables

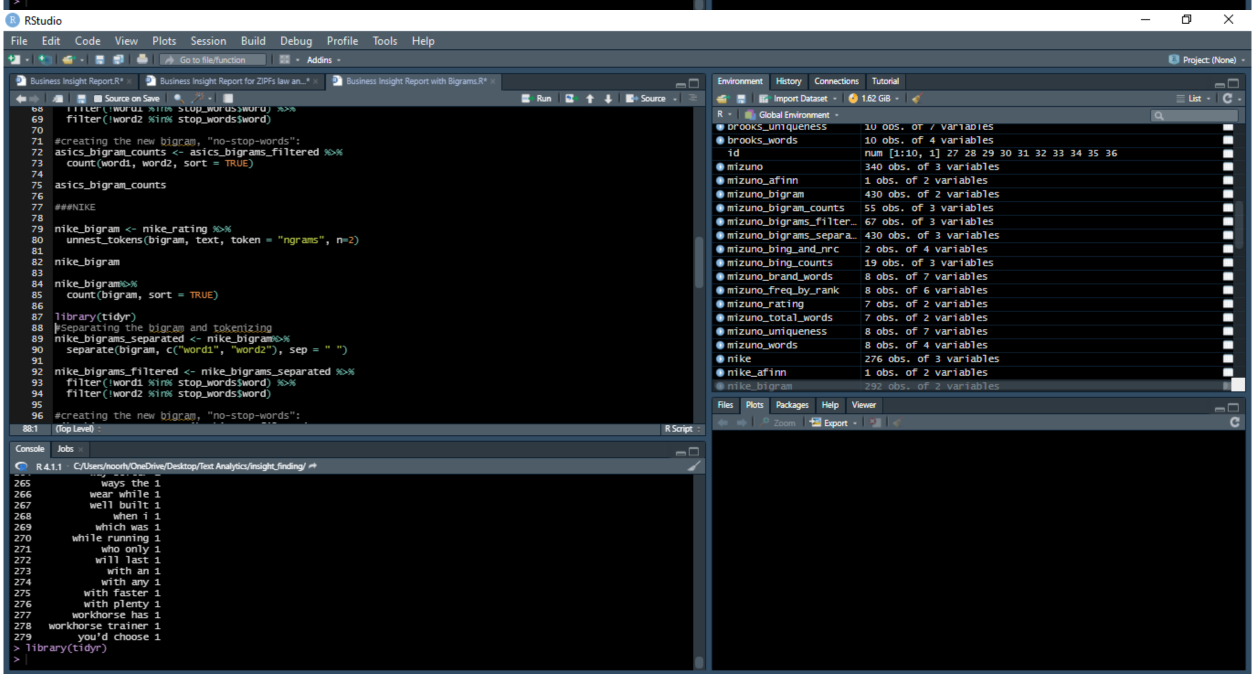

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
#to remove stop words from the bigram data
library(tidyr)
#Separating the bigram and tokenizing
mizuno_bigrams_separated <- mizuno_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
mizuno_bigrams_filtered <- mizuno_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
mizuno_bigram_counts <- mizuno_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
mizuno_bigram_counts
###BROOKS
brooks_bigram <- brooks_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
brooks_bigram
brooks_bigram%>%
count(bigram, sort = TRUE)
library(tidyr)
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
text36...247 36 sprints or
text36...248 36 or faster
text36...249 36 faster workouts
text36...250 36 workouts some
text36...251 36 some users
text36...252 36 users also
text36...253 36 also suggested
text36...254 36 suggested that
text36...255 36 that a
text36...256 36 a lower
text36...257 36 lower heel
text36...258 36 heel drop
text36...259 36 drop off
text36...260 36 off would
text36...261 36 would be
text36...262 36 be preferable

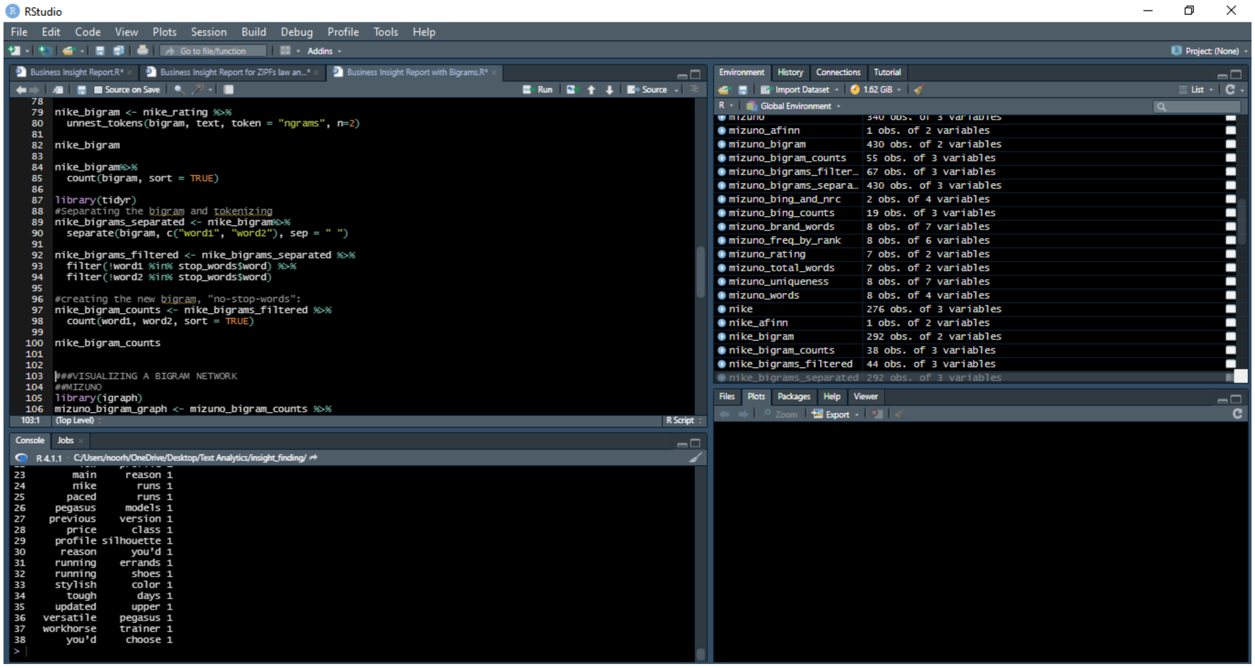

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
brooks_bigram <- brooks_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
brooks_bigram
brooks_bigram%>%
count(bigram, sort = TRUE)
library(tidyr)
#Separating the bigram and tokenizing
brooks_bigrams_separated <- brooks_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
brooks_bigrams_filtered <- brooks_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
brooks_bigram_counts <- brooks_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
brooks_bigram_counts
###ASICS
asics_bigram <- asics_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
26 i'll continue 1
27 lighter smoother 1
28 lower heel 1
29 main drawback 1
30 main negatives 1
31 oriented stability 1
32 performance oriented 1
33 preventing unwanted 1
34 previous editions 1
35 reliable ride 1
36 ride feels 1
37 rigid design 1
38 running shoes 1
39 stability models 1
40 unwanted twisting 1
41 updated engineered 1

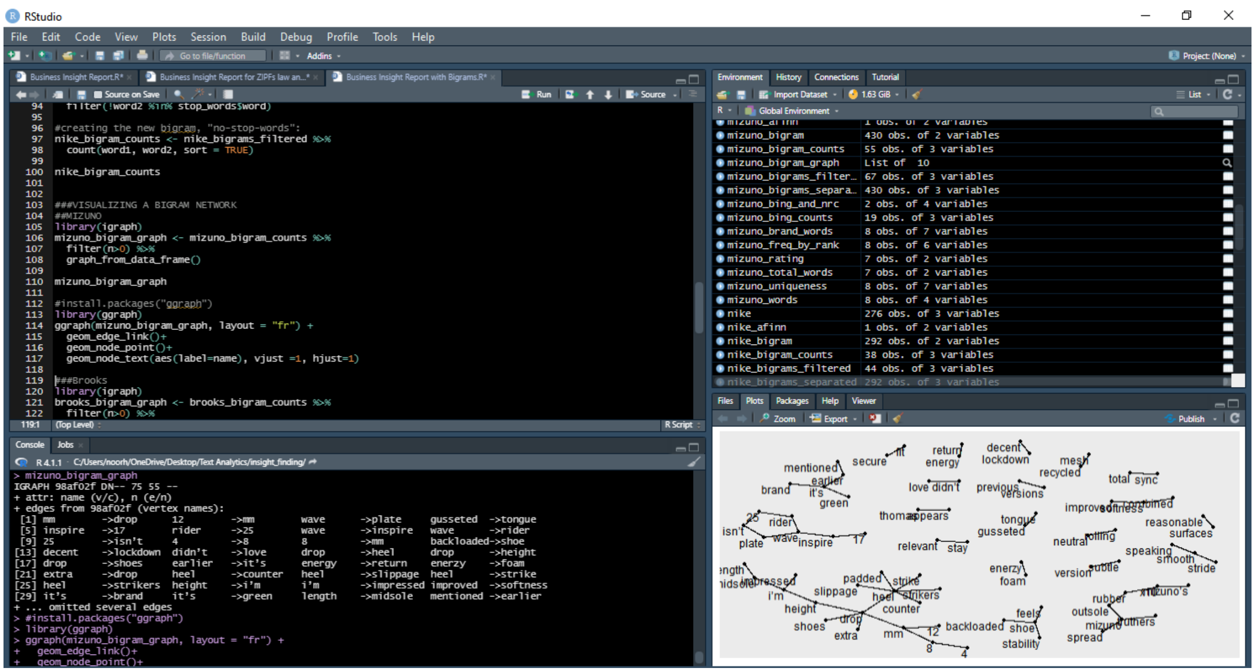

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
library(tidyr)
#Separating the bigram and tokenizing
brooks_bigrams_separated <- brooks_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
brooks_bigrams_filtered <- brooks_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
brooks_bigram_counts <- brooks_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
brooks_bigram_counts
###ASICS
asics_bigram <- asics_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
asics_bigram
asics_bigram%>%
count(bigram, sort = TRUE)
library(tidyr)
asics_bigrams_separated <- asics_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
215 up i 1
216 up shock 1
217 up the 1
218 used for 1
219 wear even 1
220 wear showed 1
221 wet and 1
222 when you 1
223 which testers 1
224 who commented 1
225 worth the 1
226 would recommend 1
227 you are 1
228 you lace 1
229 you land 1
230 you pick 1
Environment History Connections Tutorial
Global Environment
asics_bigram 244 obs. of 2 variables
asics_bing_and_nrc 2 obs. of 4 variables
asics_bing_counts 16 obs. of 3 variables
asics_brand_words 7 obs. of 7 variables
asics_freq_by_rank 7 obs. of 6 variables
asics_rating 10 obs. of 2 variables
asics_total_words 10 obs. of 2 variables
asics_uniqueness 7 obs. of 7 variables
asics_words 7 obs. of 4 variables
brooks 249 obs. of 3 variables
brooks_afinn 1 obs. of 2 variables
brooks_bigram 262 obs. of 2 variables
brooks_bigram_counts 41 obs. of 3 variables
brooks_bigrams_filter… 41 obs. of 3 variables
brooks_bigrams_separa… 262 obs. of 3 variables
brooks_bing_and_nrc 2 obs. of 4 variables
brooks_bing_counts 21 obs. of 3 variables
brooks_brand_words 10 obs. of 7 variables
Files Plots Packages Help Viewer

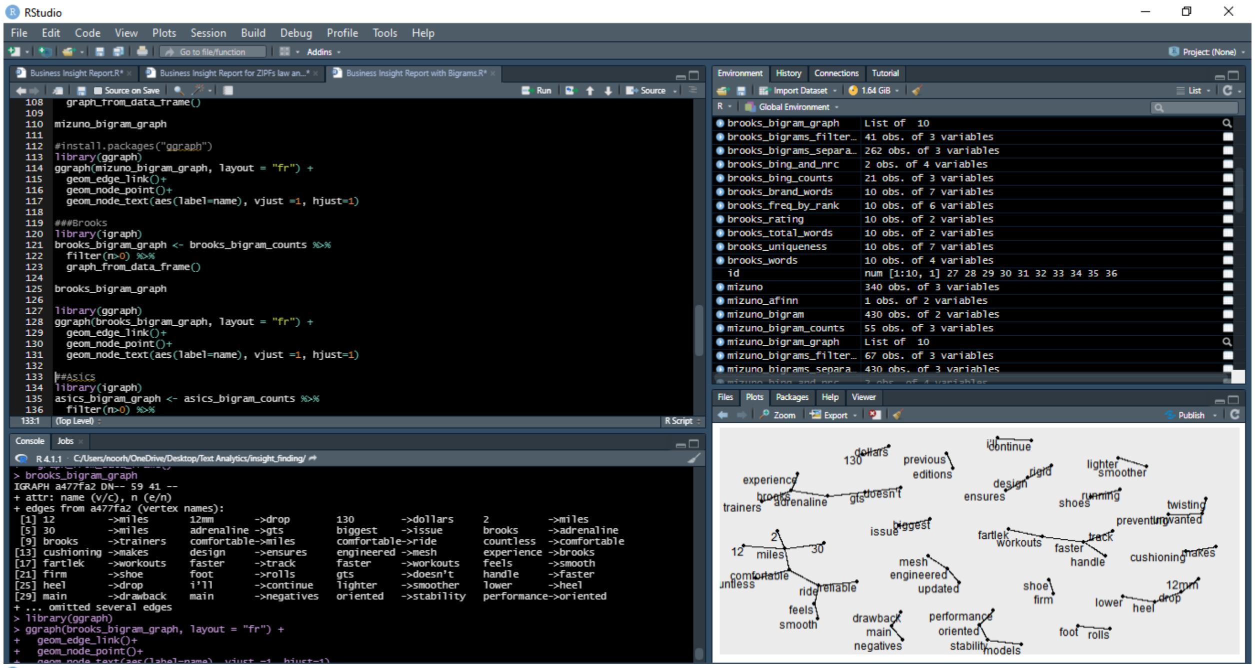

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
###ASICS
asics_bigram <- asics_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
asics_bigram
asics_bigram%>%
count(bigram, sort = TRUE)
library(tidyr)
asics_bigrams_separated <- asics_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
asics_bigrams_filtered <- asics_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
asics_bigram_counts <- asics_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
asics_bigram_counts
###NIKE
nike_bigram <- nike_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
R 4.1.1 · C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
20 hundred miles 1
21 inflexible stiff 1
22 it's heavy 1
23 minimal amount 1
24 minimal wear 1
25 pad positioned 1
26 price tag 1
27 provided brought 1
28 putting close 1
29 review mileage 1
30 shoe can't 1
31 shoe it's 1
32 smooth transition 1
33 stability shoe 1
34 stability trainer 1
35 testers appreciated 1
Environment History Connections Tutorial
Global Environment
asics_bigram_counts 35 obs. of 3 variables
asics_bigrams_filtered 40 obs. of 3 variables
asics_bigrams_separat… 244 obs. of 3 variables
asics_bing_and_nrc 2 obs. of 4 variables
asics_bing_counts 16 obs. of 3 variables
asics_brand_words 7 obs. of 7 variables
asics_freq_by_rank 7 obs. of 6 variables
asics_rating 10 obs. of 2 variables
asics_total_words 10 obs. of 2 variables
asics_uniqueness 7 obs. of 7 variables
asics_words 7 obs. of 4 variables
brooks 249 obs. of 3 variables
brooks_afinn 1 obs. of 2 variables
brooks_bigram 262 obs. of 2 variables
brooks_bigram_counts 41 obs. of 3 variables
brooks_bigrams_filter… 41 obs. of 3 variables
brooks_bigrams_separa… 262 obs. of 3 variables
brooks_bing_and_nrc 2 obs. of 4 variables
Files Plots Packages Help Viewer

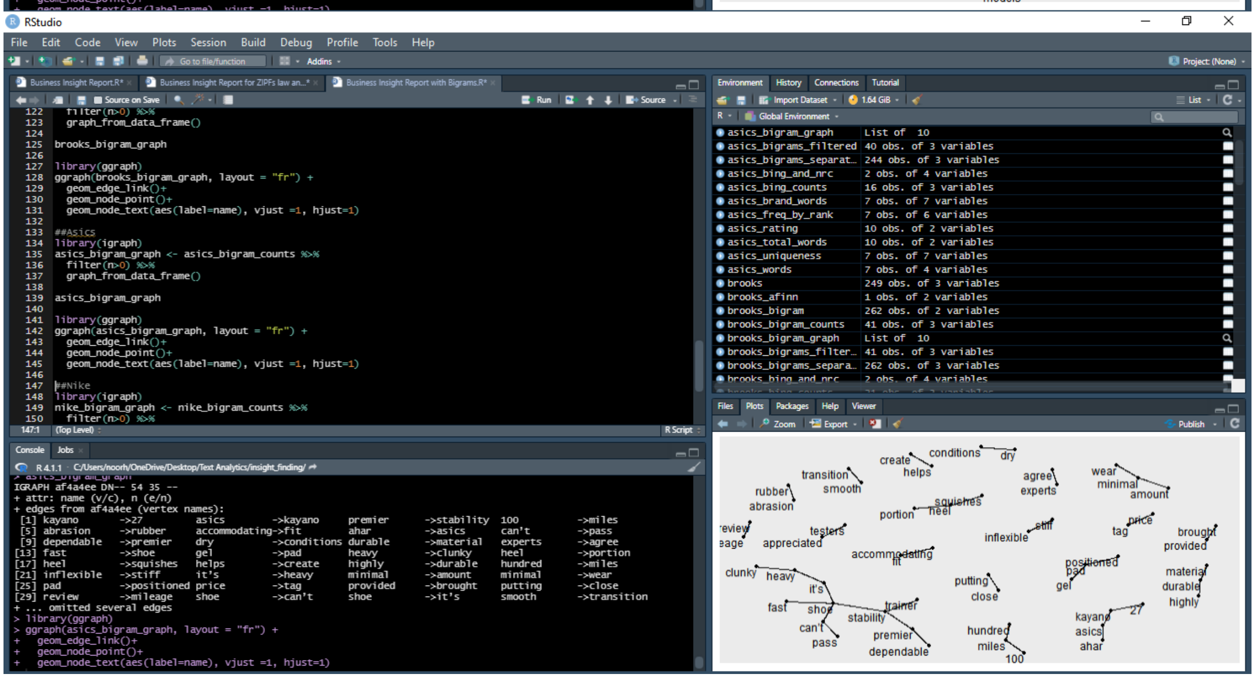

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
asics_bigram_counts <- asics_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
asics_bigram_counts
###NIKE
nike_bigram <- nike_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
nike_bigram
nike_bigram%>%
count(bigram, sort = TRUE)
library(tidyr)
#Separating the bigram and tokenizing
nike_bigrams_separated <- nike_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
nike_bigrams_filtered <- nike_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
text16...277 16 an adequate
text16...278 16 adequate amount
text16...279 16 amount of
text16...280 16 of support
text16...281 16 support so
text16...282 16 so that
text16...283 16 that my
text16...284 16 my knees
text16...285 16 knees and
text16...286 16 and hips
text16...287 16 hips felt
text16...288 16 felt good
text16...289 16 good from
text16...290 16 from start
text16...291 16 start to
text16...292 16 to finish
Environment History Connections Tutorial
Import Dataset 1.62 GiB List
Global Environment
brooks_uniqueness 10 obs. of 7 variables
brooks_words 10 obs. of 4 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno 340 obs. of 3 variables
mizuno_afinn 1 obs. of 2 variables
mizuno_bigram 430 obs. of 2 variables
mizuno_bigram_counts 55 obs. of 3 variables
mizuno_bigrams_filter... 67 obs. of 3 variables
mizuno_bigrams_separa... 430 obs. of 3 variables
mizuno_bing_and_nrc 2 obs. of 4 variables
mizuno_bing_counts 19 obs. of 3 variables
mizuno_brand_words 8 obs. of 7 variables
mizuno_freq_by_rank 8 obs. of 6 variables
mizuno_rating 7 obs. of 2 variables
mizuno_total_words 7 obs. of 2 variables
mizuno_uniqueness 8 obs. of 7 variables
mizuno_words 8 obs. of 4 variables
nike 276 obs. of 3 variables
nike_afinn 1 obs. of 2 variables
nike_bigram 292 obs. of 2 variables
Files Plots Packages Help Viewer


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
asics_bigram_counts <- asics_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
asics_bigram_counts
###NIKE
nike_bigram <- nike_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
nike_bigram
nike_bigram%>%
count(bigram, sort = TRUE)
library(tidyr)
#Separating the bigram and tokenizing
nike_bigrams_separated <- nike_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
nike_bigrams_filtered <- nike_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
265 ways the 1
266 wear while 1
267 well built 1
268 when i 1
269 which was 1
270 while running 1
271 who only 1
272 will last 1
273 with an 1
274 with any 1
275 with faster 1
276 with plenty 1
277 workhorse has 1
278 workhorse trainer 1
279 you'd choose 1
> library(tidyr)
Environment History Connections Tutorial
Global Environment
brooks_words 10 obs. of 4 variables
id num [1:10, 1] 27 28 29 30 31 32 33 34 35 36
mizuno 340 obs. of 3 variables
mizuno_afinn 1 obs. of 2 variables
mizuno_bigram 430 obs. of 2 variables
mizuno_bigram_counts 55 obs. of 3 variables
mizuno_bigrams_filter... 67 obs. of 3 variables
mizuno_bigrams_separa... 430 obs. of 3 variables
mizuno_bing_and_nrc 2 obs. of 4 variables
mizuno_bing_counts 19 obs. of 3 variables
mizuno_brand_words 8 obs. of 7 variables
mizuno_freq_by_rank 8 obs. of 6 variables
mizuno_rating 7 obs. of 2 variables
mizuno_total_words 7 obs. of 2 variables
mizuno_uniqueness 8 obs. of 7 variables
mizuno_words 8 obs. of 4 variables
nike 276 obs. of 3 variables
nike_afinn 1 obs. of 2 variables
nike_bigram 292 obs. of 2 variables
Files Plots Packages Help Viewer


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
nike_bigram <- nike_rating %>%
unnest_tokens(bigram, text, token = "ngrams", n=2)
nike_bigram
nike_bigram%>%
count(bigram, sort = TRUE)
library(tidyr)
#Separating the bigram and tokenizing
nike_bigrams_separated <- nike_bigram%>%
separate(bigram, c("word1", "word2"), sep = " ")
nike_bigrams_filtered <- nike_bigrams_separated %>%
filter(!word1 %in% stop_words$word) %>%
filter(!word2 %in% stop_words$word)
#creating the new bigram, "no-stop-words":
nike_bigram_counts <- nike_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
nike_bigram_counts
###VISUALIZING A BIGRAM NETWORK
##MIZUNO
library(igraph)
mizuno_bigram_graph <- mizuno_bigram_counts %>%
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Business Insight Report.R*
Business Insight Report for ZIPFs law an...*
Business Insight Report with Bigrams.R*
#creating the new bigram, "no-stop-words":
nike_bigram_counts <- nike_bigrams_filtered %>%
count(word1, word2, sort = TRUE)
nike_bigram_counts
###VISUALIZING A BIGRAM NETWORK
##MIZUNO
library(igraph)
mizuno_bigram_graph <- mizuno_bigram_counts %>%
filter(n>0) %>%
graph_from_data_frame()
mizuno_bigram_graph
#install.packages("ggraph")
library(ggraph)
ggraph(mizuno_bigram_graph, layout = "fr") +
geom_edge_link()+
geom_node_point()+
geom_node_text(aes(label=name), vjust =1, hjust=1)
###Brooks
library(igraph)
brooks_bigram_graph <- brooks_bigram_counts %>%
filter(n>0) %>%
Console Jobs
R 4.1.1 C:/Users/noorh/OneDrive/Desktop/Text Analytics/insight_finding/
> mizuno_bigram_graph
IGRAPH 98af02f DN-- 75 55 --
+ attr: name (v/c), n (e/n)
+ edges from 98af02f (vertex names):
+ ... omitted several edges
> #install.packages("ggraph")
> library(ggraph)
> ggraph(mizuno_bigram_graph, layout = "fr") +
+ geom_edge_link()+
+ geom_node_point()+


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
graph_from_data_frame()
mizuno_bigram_graph
#install.packages("ggraph")
library(ggraph)
ggraph(mizuno_bigram_graph, layout = "fr") +
geom_edge_link()+
geom_node_point()+
geom_node_text(aes(label=name), vjust =1, hjust=1)
###Brooks
library(igraph)
brooks_bigram_graph <- brooks_bigram_counts %>%
filter(n>0) %>%
graph_from_data_frame()
brooks_bigram_graph
library(ggraph)
ggraph(brooks_bigram_graph, layout = "fr") +
geom_edge_link()+
geom_node_point()+
geom_node_text(aes(label=name), vjust =1, hjust=1)
##Asics
library(igraph)
asics_bigram_graph <- asics_bigram_counts %>%
filter(n>0) %>%
> brooks_bigram_graph
IGRAPH a477fa2 DN-- 59 41 --
+ attr: name (v/c), n (e/n)
+ edges from a477fa2 (vertex names):
> library(ggraph)
> ggraph(brooks_bigram_graph, layout = "fr") +
+ geom_edge_link()+
+ geom_node_point()+


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
filter(n>0) %>%
graph_from_data_frame()
brooks_bigram_graph
library(ggraph)
ggraph(brooks_bigram_graph, layout = "fr") +
geom_edge_link()+
geom_node_point()+
geom_node_text(aes(label=name), vjust =1, hjust=1)
##Asics
library(igraph)
asics_bigram_graph <- asics_bigram_counts %>%
filter(n>0) %>%
graph_from_data_frame()
asics_bigram_graph
library(ggraph)
ggraph(asics_bigram_graph, layout = "fr") +
geom_edge_link()+
geom_node_point()+
geom_node_text(aes(label=name), vjust =1, hjust=1)
##Nike
library(igraph)
nike_bigram_graph <- nike_bigram_counts %>%
filter(n>0) %>%
IGRAPH af4a4ee DN-- 54 35 --
+ attr: name (v/c), n (e/n)
+ edges from af4a4ee (vertex names):
> library(ggraph)
> ggraph(asics_bigram_graph, layout = "fr") +
+ geom_edge_link()+
+ geom_node_point()+
+ geom_node_text(aes(label=name), vjust =1, hjust=1)