Business Insight Report:

Text Analysis in the Fast Fashion Industry

Maria P. Lopez Moreno

Hult International Business School

Cohort: MBAN 1

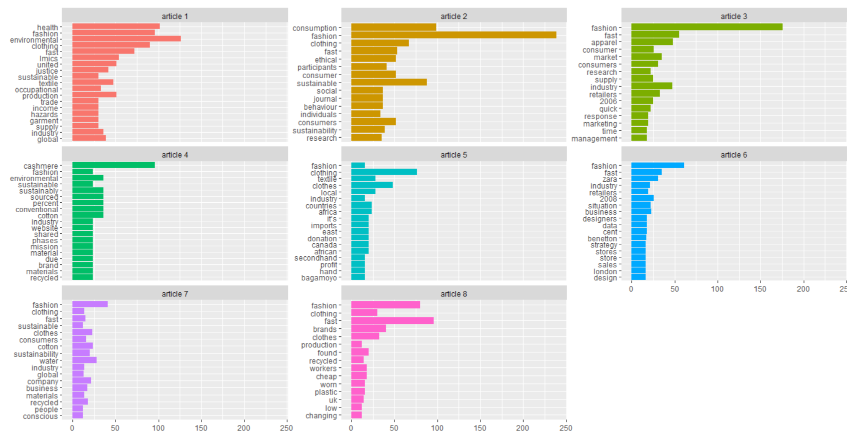
DAT-5317: Text Analytics and Natural Language Processing (NLP)

Professor Thomas Kurnicki

December 5, 2021

Business Insight Report: Text Analysis in the Fast Fashion Industry

According to McNeill & Moore (2015) over the last decade, the fast fashion phenomenon has been impacting the clothing industry (p. 213). Fashion used to be seen as a topic related only to vanity and that was focused on women. Nevertheless, the text analysis shows us that fashion, especially fast fashion, is having more attention and its scope is reaching everyone, even if they are interested or not in the industry. This study was developed over eight different articles related to the fashion industry and the impact, challenges, and some proposals that fast fashion consumption has created.

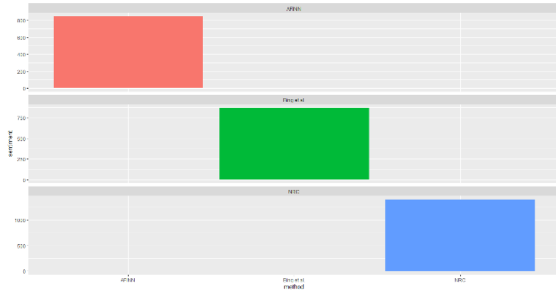


Graph A 1: Frequencies of the words per article

When structuring the data, the outcomes shown words as environmental, justice, sustainable, and ethical, were shown as frequent[1]. This glimpse the importance of being aware of what is going on in the fashion industry and why this fast-paced consumption is gaining force day by day. "The global fashion industry generates a huge amount of waste - one full garbage truck of clothes is burned or sent to a landfill every second, according to a report by the Ellen MacArthur Foundation." (Gerretsen & Kottasová, 2020). This gives us a glimpse of the importance of being aware of what is going on in the fashion industry and why this fast-paced consumption is gaining force day by day. With that in mind, the sentiment analysis of the articles was run[2].

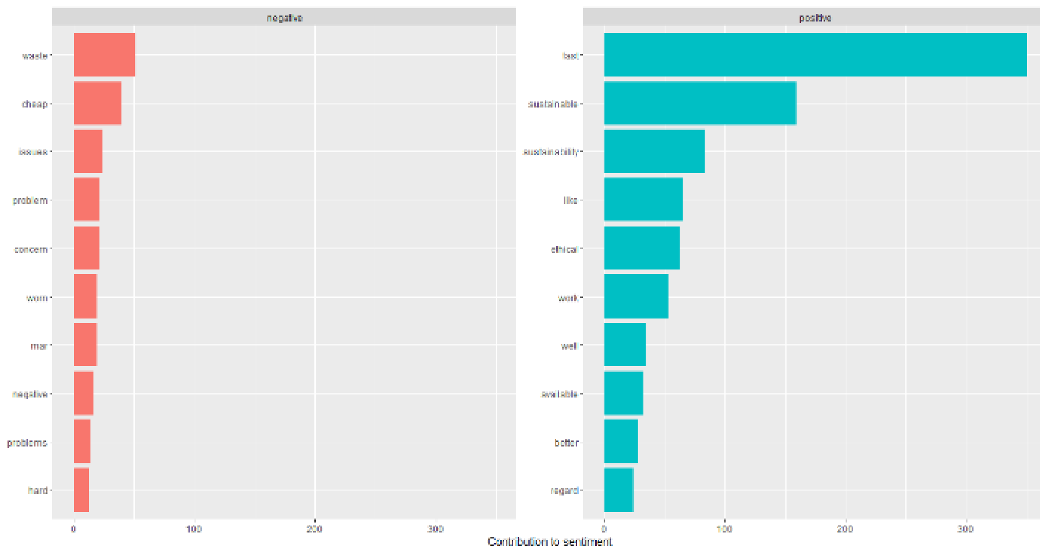
---

[1] See Graph A2 in the appendix.
[2] See Graphs A3 and A4 in the appendix.



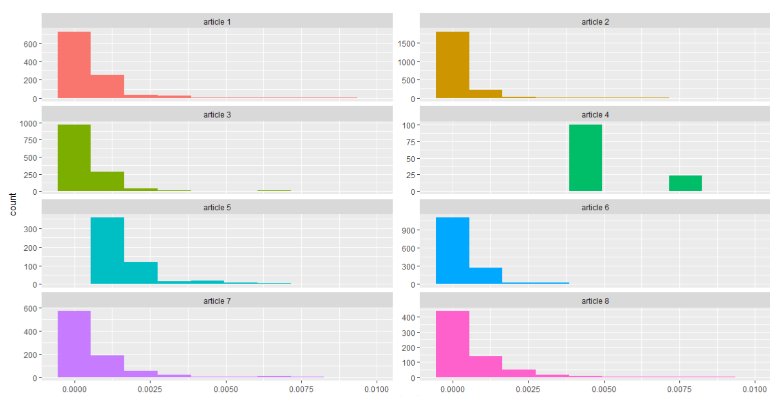
Graph A 3: Sentiment Count per sentiment library in the overall text of the articles

The results showed a positive overall of the data. Nevertheless, this result is affected by the "fast" word, which is one of the most frequent ones and has a positive classification in the data used, but in the context of the research, it is related to the fast consumption in the fashion industry and has a negative connotation.



Graph A 2: Most frequent positive and negative sentiments in the Fast Fashion articles

Within the analysis carried out, words such as fast fashion and environment generated a lot of noise in the possible results that can be seen when running this type of analysis. That can be seen in the histogram graph of the tokens per article[3]. For this reason, a study by TF-IDF that allowed to have more information about what was happening in-depth with this



Graph A 4: Token Histogram per Article Adjusted

[3] See Graph A6 in the appendix.

fast consumption of fashion, explained in the general articles, was carried out.

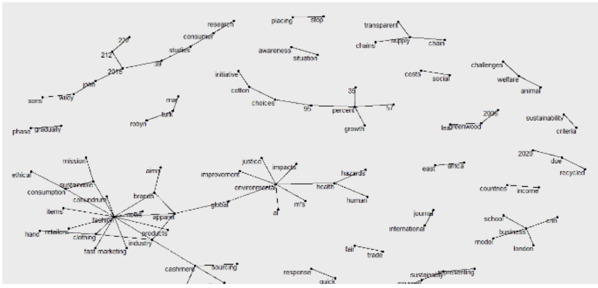
The results presented important words such as health, behavior, workers, and hazards; showed names of some brands that are related to the fast fashion industry such as Zara and Benetton; and mentioned different textiles and materials such as cashmere, piñatex, and plastic[4].

Graph A 5: Most Unique and Most Frequent Tokens per Article

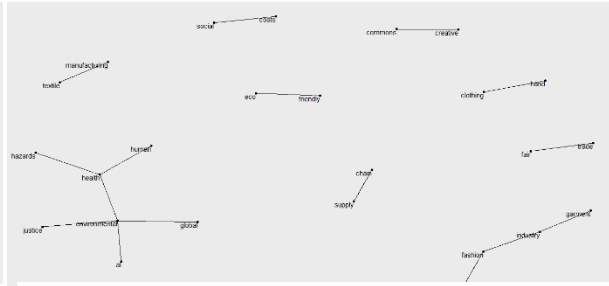
Although, with this analysis, you can have some general ideas about fast fashion consumption and its impact. Limiting the study to words was insufficient. Due to the length of the articles and the number of tokens that were had in total after removing stop words, a bigram analysis was carried out. This allowed having indications about the type of document and the in-depth content of each analyzed text.

The bi-gram analysis was applied to the entire dataset and article by article too[5]. As it can be seen in all the different bigram networks the importance of a transparent supply chain, social cost and fair trade are some of the topics that surround the fast fashion industry and represent a challenge to solve. Phrases as sustainable mission, recycler textile, plastic fibers, organic cotton, recycled polyester, give us the idea that the industry is aware of the problems that are creating, and it is looking for alternatives to mitigate the impact that they are having.

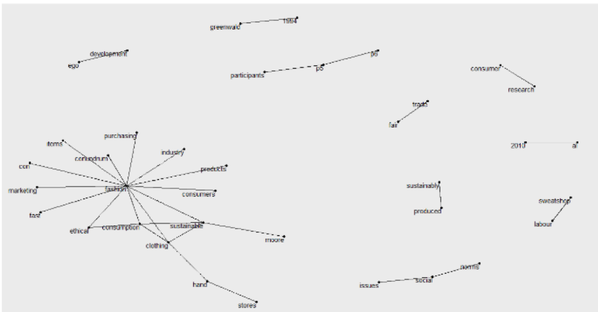
[4] See Graph A7 in the appendix.
[5] See Graphs A8 to A21 in the appendix.



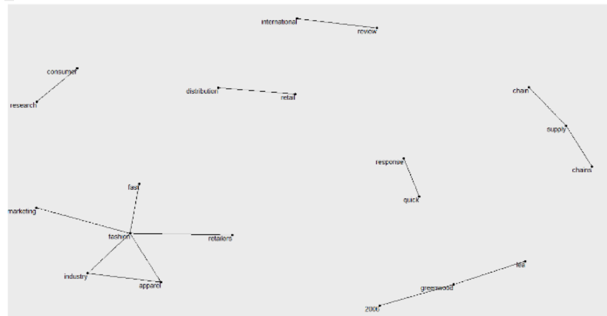
Graph A 8: Bigram Network



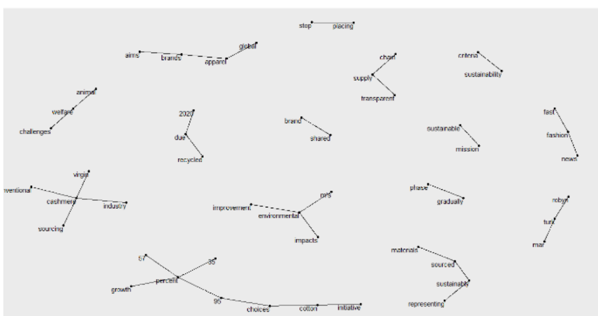
Graph A 7: Bigram Network Article 1



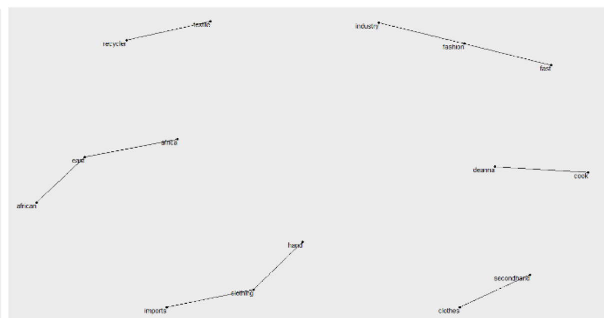
Graph A 15: Cleaned Bigram Network Article 2



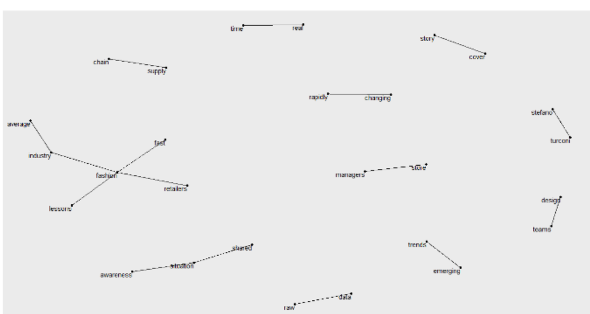
Graph A 13: Bigram Network Article 3



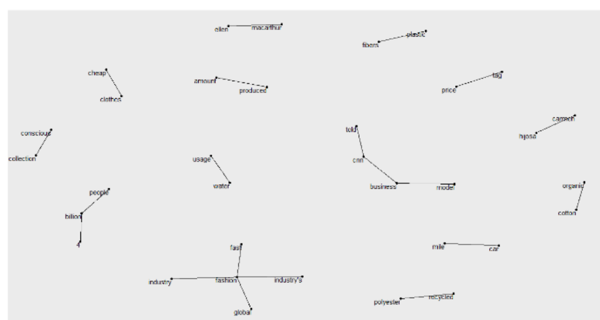
Graph A 11: Bigram Network Article 4



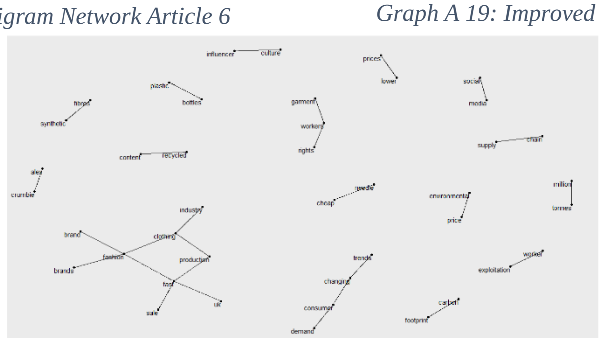
Graph A 9: Bigram Network Article 5



Graph A 17: Cleaned Bigram Network Article 6



Graph A 19: Improved Bigram Network Article 7

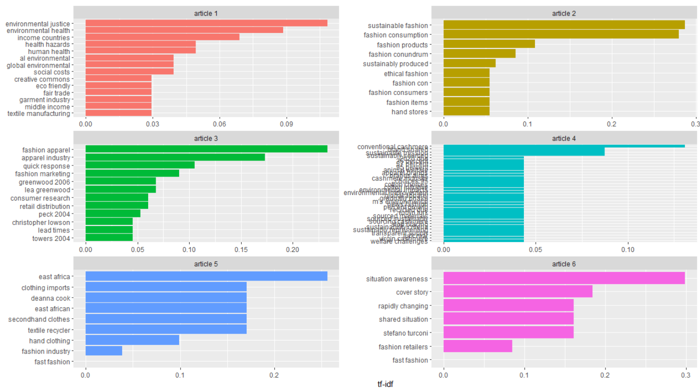


Graph A 6: Improved Bigram Network Article

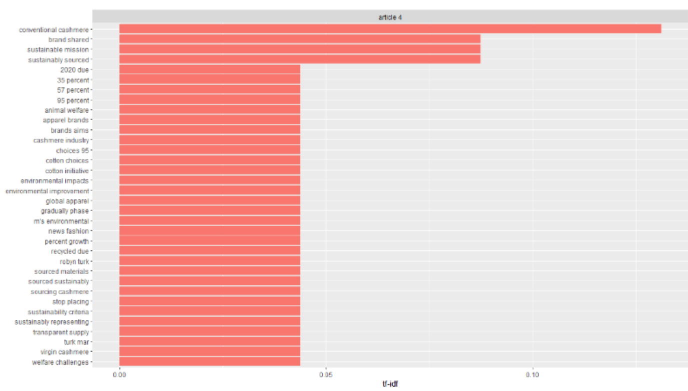
Based on the results and seeking to take full advantage of the tools already used, a TF-IDF model was generated on the bigrams of the different articles. From the analysis on bigrams carried out, it was discovered that there is a need to separate our articles between longer and shorter and filter according to their size to have enough business insights.[6]

Overall, the articles present topics as environmental justice, health hazards, social cost, animal welfare, conscious collection, carbon footprint, among others as frequent and unique phrases that describe fast fashion as a topic of general awareness.

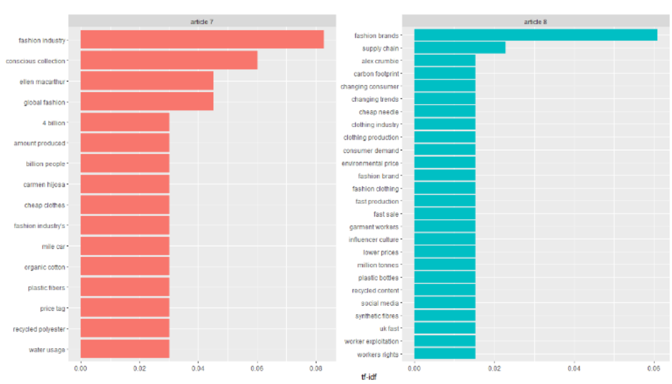
Finally, after applying the different frameworks to the analyzed texts and seeing that in effect the massive consumption of clothing and its accelerated change in trends and consumption represent a great challenge for today's society. Two important points need to be made. The first is that as it could be seen in the final analysis, we cannot forget the context of the situations when



Graph A 21: Most unique and frequent bigrams per article – long articles



Graph A 23: Most Frequent and unique bigrams article 4



Graph A 22: Most unique and frequent bigrams for the shorter articles

[6] See Graphs A22 to A24 in the appendix.

analyzing text, since the semantics of our discourse is what will give us a complete picture of what we have. in front. Second, directly related to the textile industry, is that I consider it important to make a call not only to brands that promote accelerated and massive consumption of clothing but also to consumers who support it and the way they are consuming it. After analyzing the text about fast fashion, it is important to consider that if actions are not taken in time to mitigate the impact that the high consumption generated by the fashion industry, the environmental and social consequences will be greater than what we can imagine and perhaps manage to handle.

REFERENCE LIST & BIBLIOGRAPHY

Bick, R., Halsey, E. & Ekenga, C. (2018). The global environmental injustice of fast Fashion. *Environmental Health. 17*(92). https://doi.org/10.1186/s12940-018-0433-7

Bhardwaj, V. & Ann Fairhurst, A. (2010). Fast fashion: response to changes in the fashion industry. *The International Review of Retail, Distribution and Consumer Research*. *20*(1), 165-173, https://doi.org/10.1080/09593960903498300

Cook, D. (September 2, 2019). Second-Hand Clothing in East Africa: What the Fashion Industry Doesn't Want You to Know. *Forage Sustain*. https://forageandsustain.com/second-hand-clothing-in-east-africa-what-the-fast-fashion-industry-doesnt-want-you-to-know/

Crumbie, A. (October 5, 2021). What is fast fashion and why is it a problem?. Ethical Consumer. https://www.ethicalconsumer.org/fashion-clothing/what-fast-fashion-why-it-problem#:~:text=Fast%20fashion%20is%20

Gerretsen, I. & Kottasová, I. (May 6, 2020). The world is paying a high price for cheap clothes. *CNN Business*. https://edition.cnn.com/2020/05/03/business/cheap-clothing-fast-fashion-climate-change-intl/index.html

McNeill, L. & Moore, R. (2015). Sustainable fashion consumption and the fast fashion conundrum: fashionable consumers and attitudes to sustainability in clothing choice. *International Journal of Consumer Studies, 39*. 212–222. https://doi.org/10.1111/ijcs.12169

Sull, D. and Turconi, S. (2008). Fast fashion lessons. *Business Strategy Review, 19*(2). 4-11. https://doi.org/10.1111/j.1467-8616.2008.00527.x

Turk, R. (Mar 20, 2019). H&M phases out conventional cashmere for sustainable mission. *FASHIONUNITED.* https://fashionunited.com/news/fashion/h-m-phases-out-conventional-cashmere-for-sustainable-mission/2019032026843#.XJSutUm_kyo.twitter

APPENDIX

R Code and R Outputs

```
#################################
### MSBA 1 HULT 2021-2022
### Business Insight Assignment
### "The Fast Fashion Revolution"
### Created by: Maria Paula Lopez Moreno
### Date: 12.02.2021
### Version 1.0
################################
# This project is going to be focused in analyze different text related with fast fashion and the
# impact that is having around the world.
### Calling Libraries ###
library(pdftools)
library(tm)
library(dplyr)
library(tidytext)
library(tidyverse)
library(ggplot2)
library(tidyr)
library(igraph)
library(ggraph)
library(textdata)
```

library(stringr)

#######################

Importing my data

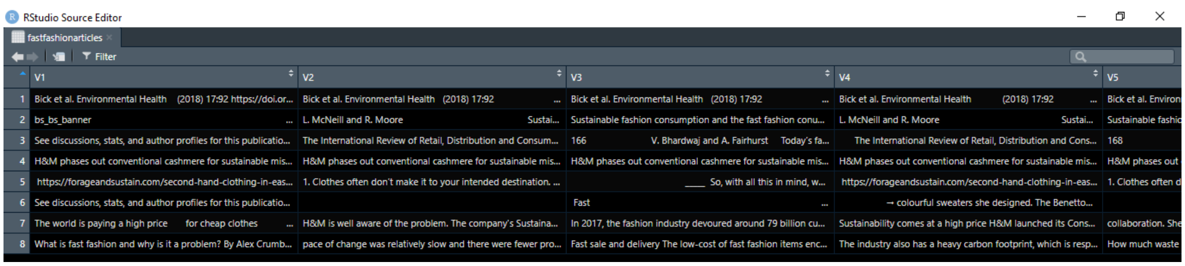
#######################

setwd("D:/Documentos/HULT/UNIVERSITY/TEXT ANALYTICS & NPL/Business Case/PDF files")

nm <- list.files(path="D:/Documentos/HULT/UNIVERSITY/TEXT ANALYTICS & NPL/Business Case/PDF files")

fastfashionarticles <- do.call(rbind, lapply(nm, function(x) pdf_text(x)))

R Code Output:


RStudio Source Editor

fastfashionarticles

Filter

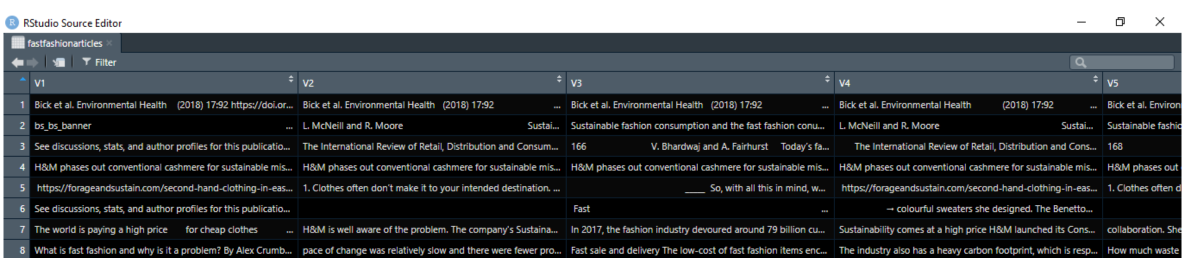
	V1	V2	V3	V4	V5
1	Bick et al. Environmental Health (2018) 17:92 https://doi.or...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environ
2	bs_bs_banner ...	L. McNeill and R. Moore Sustai...	Sustainable fashion consumption and the fast fashion conu...	L. McNeill and R. Moore Sustai...	Sustainable fashio
3	See discussions, stats, and author profiles for this publicatio...	The International Review of Retail, Distribution and Consum...	166 V. Bhardwaj and A. Fairhurst Today's fa...	The International Review of Retail, Distribution and Cons...	168
4	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out
5	https://forageandsustain.com/second-hand-clothing-in-eas...	1. Clothes often don't make it to your intended destination. ...	____ So, with all this in mind, w...	https://forageandsustain.com/second-hand-clothing-in-eas...	1. Clothes often d
6	See discussions, stats, and author profiles for this publicatio...		Fast ...	→ colourful sweaters she designed. The Benetto...	
7	The world is paying a high price for cheap clothes ...	H&M is well aware of the problem. The company's Sustaina...	In 2017, the fashion industry devoured around 79 billion cu...	Sustainability comes at a high price H&M launched its Cons...	collaboration. She
8	What is fast fashion and why is it a problem? By Alex Crumb...	pace of change was relatively slow and there were fewer pro...	Fast sale and delivery The low-cost of fast fashion items enc...	The industry also has a heavy carbon footprint, which is resp...	How much waste

Figure A 1: Fast Fashion Articles Matrix

- The code creates a matrix in the environment. The text divided into multiple columns due to the appearance of images within the different articles (See Figure A1).

fastfashiondf <- data.frame(fastfashionarticles)

R Code Output:


RStudio Source Editor

fastfashionarticles

Filter

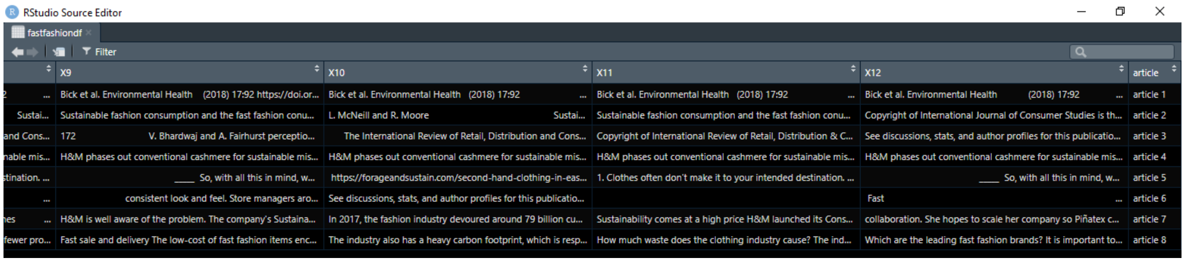
	V1	V2	V3	V4	V5
1	Bick et al. Environmental Health (2018) 17:92 https://doi.or...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environ
2	bs_bs_banner ...	L. McNeill and R. Moore Sustai...	Sustainable fashion consumption and the fast fashion conu...	L. McNeill and R. Moore Sustai...	Sustainable fashio
3	See discussions, stats, and author profiles for this publicatio...	The International Review of Retail, Distribution and Consum...	166 V. Bhardwaj and A. Fairhurst Today's fa...	The International Review of Retail, Distribution and Cons...	168
4	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out
5	https://forageandsustain.com/second-hand-clothing-in-eas...	1. Clothes often don't make it to your intended destination. ...	____ So, with all this in mind, w...	https://forageandsustain.com/second-hand-clothing-in-eas...	1. Clothes often d
6	See discussions, stats, and author profiles for this publicatio...		Fast ...	→ colourful sweaters she designed. The Benetto...	
7	The world is paying a high price for cheap clothes ...	H&M is well aware of the problem. The company's Sustaina...	In 2017, the fashion industry devoured around 79 billion cu...	Sustainability comes at a high price H&M launched its Cons...	collaboration. She
8	What is fast fashion and why is it a problem? By Alex Crumb...	pace of change was relatively slow and there were fewer pro...	Fast sale and delivery The low-cost of fast fashion items enc...	The industry also has a heavy carbon footprint, which is resp...	How much waste

Figure A 2: Fast Fashion Articles Data Frame

- The line of code converts the matrix into a data frame (See Figure A2).

```
fastfashiondf$article <- c("article 1", "article 2", "article 3", "article 4", "article 5",
                           "article 6", "article 7", "article 8")
```

R Code Output:



RStudio Source Editor

fastfashiondf

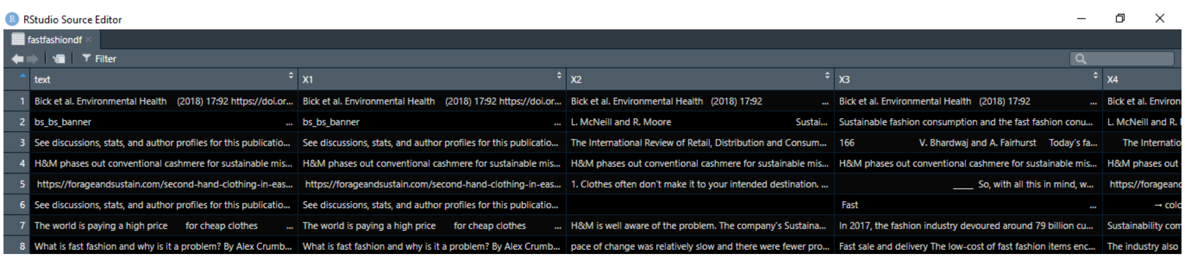
	X9	X10	X11	X12	article
2 ...	Bick et al. Environmental Health (2018) 17:92 https://doi.or...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environmental Health (2018) 17:92 ...	article 1
Sustai...	Sustainable fashion consumption and the fast fashion conu...	L. McNeill and R. Moore Sustai...	Sustainable fashion consumption and the fast fashion conu...	Copyright of International Journal of Consumer Studies is th...	article 2
and Cons...	172 V. Bhardwaj and A. Fairhurst perceptio...	The International Review of Retail, Distribution and Cons...	Copyright of International Review of Retail, Distribution & C...	See discussions, stats, and author profiles for this publicatio...	article 3
nable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	article 4
stination. ...	____ So, with all this in mind, w...	https://forageandsustain.com/second-hand-clothing-in-eas...	1. Clothes often don't make it to your intended destination. ...	____ So, with all this in mind, w...	article 5
...	consistent look and feel. Store managers aro...	See discussions, stats, and author profiles for this publicatio...		Fast ...	article 6
nes ...	H&M is well aware of the problem. The company's Sustaina...	In 2017, the fashion industry devoured around 79 billion cu...	Sustainability comes at a high price H&M launched its Cons...	collaboration. She hopes to scale her company so Piñatex c...	article 7
fewer pro...	Fast sale and delivery The low-cost of fast fashion items enc...	The industry also has a heavy carbon footprint, which is resp...	How much waste does the clothing industry cause? The ind...	Which are the leading fast fashion brands? It is important to...	article 8

Figure A 3: Fast Fashion Articles Data Frame, Articles Column

- The line of code adds a new column to my data frame called articles, this column is going to be used for the analysis to be able to manipulate our data depending on the article (See Figure A3)

```
fastfashiondf <- fastfashiondf %>%
  unite ("text", 1:12 ,sep = " ", remove = FALSE, na.rm = FALSE)
```

R Code Output:



RStudio Source Editor

fastfashiondf

	text	X1	X2	X3	X4
1	Bick et al. Environmental Health (2018) 17:92 https://doi.or...	Bick et al. Environmental Health (2018) 17:92 https://doi.or...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environmental Health (2018) 17:92 ...	Bick et al. Environ
2	bs_bs_banner ...	bs_bs_banner ...	L. McNeill and R. Moore Sustai...	Sustainable fashion consumption and the fast fashion conu...	L. McNeill and R.
3	See discussions, stats, and author profiles for this publicatio...	See discussions, stats, and author profiles for this publicatio...	The International Review of Retail, Distribution and Consum...	166 V. Bhardwaj and A. Fairhurst Today's fa...	The Internatio
4	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out conventional cashmere for sustainable mis...	H&M phases out
5	https://forageandsustain.com/second-hand-clothing-in-eas...	https://forageandsustain.com/second-hand-clothing-in-eas...	1. Clothes often don't make it to your intended destination. ...	____ So, with all this in mind, w...	https://forageand
6	See discussions, stats, and author profiles for this publicatio...	See discussions, stats, and author profiles for this publicatio...		Fast ...	→ colo
7	The world is paying a high price for cheap clothes ...	The world is paying a high price for cheap clothes ...	H&M is well aware of the problem. The company's Sustaina...	In 2017, the fashion industry devoured around 79 billion cu...	Sustainability com
8	What is fast fashion and why is it a problem? By Alex Crumb...	What is fast fashion and why is it a problem? By Alex Crumb...	pace of change was relatively slow and there were fewer pro...	Fast sale and delivery The low-cost of fast fashion items enc...	The industry also

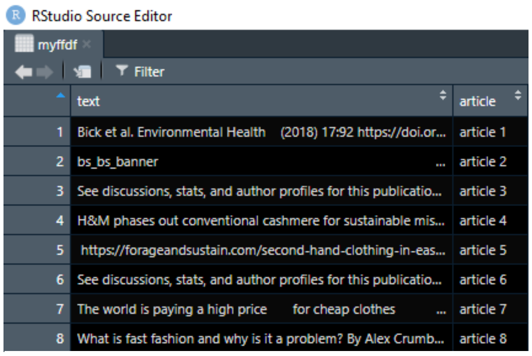
Figure A 4: Fast Fashion Article Data Frame, Text Variable

- The line of code creates a new column to my data frame called text, this column is the union of all the parts of my article that are represented in the different columns in just one variable per article (See Figure A4).

Creating a new dataframe with Articles and Descriptions

myffdf <- cbind.data.frame(text = fastfashiondf$text, article = fastfashiondf$article)

R Code Output:


RStudio Source Editor

myffdf

	text	article
1	Bick et al. Environmental Health (2018) 17:92 https://doi.or...	article 1
2	bs_bs_banner ...	article 2
3	See discussions, stats, and author profiles for this publicatio...	article 3
4	H&M phases out conventional cashmere for sustainable mis...	article 4
5	https://forageandsustain.com/second-hand-clothing-in-eas...	article 5
6	See discussions, stats, and author profiles for this publicatio...	article 6
7	The world is paying a high price for cheap clothes ...	article 7
8	What is fast fashion and why is it a problem? By Alex Crumb...	article 8

Figure A 5: Fast Fashion Data Frame

- The line of code creates a new data frame with just two variables, one with the text variable an the other with the article variable (See Figure A5). This one is going to be the data frame that is going to be used for the text analysis.

Understanding my data

########################

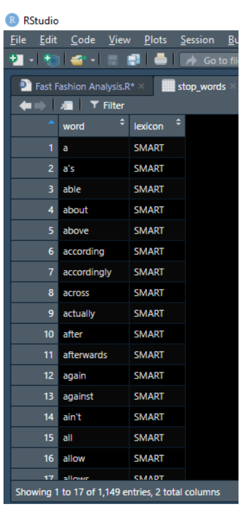
Tokenizing my data

########################

Calling data

data("stop_words")

R Code Output


RStudio

stop_words

	word	lexicon
1	a	SMART
2	a's	SMART
3	able	SMART
4	about	SMART
5	above	SMART
6	according	SMART
7	accordingly	SMART
8	across	SMART
9	actually	SMART
10	after	SMART
11	afterwards	SMART
12	again	SMART
13	against	SMART
14	ain't	SMART
15	all	SMART
16	allow	SMART

Showing 1 to 17 of 1,149 entries, 2 total columns

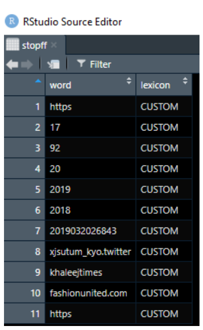
Figure A 6: Stop Words

- This line of code creates a table data frame in the environment with the most common stop_words in the English language and it's lexicon (See Figure A6).

Adding some words that we found that are frequent but has no business insights for the analysis

```
stopff <- tribble(~word,~lexicon,
                  "https", "CUSTOM",
                  "17", "CUSTOM",
                  "92", "CUSTOM",
                  "20", "CUSTOM",
                  "2019", "CUSTOM",
                  "2018", "CUSTOM",
                  "2019032026843", "CUSTOM",
                  "xjsutum_kyo.twitter", "CUSTOM",
                  "khaleejtimes", "CUSTOM",
                  "fashionunited.com", "CUSTOM",
                  "https", "CUSTOM",
)
```

R Code Output:



RStudio Source Editor

stopff

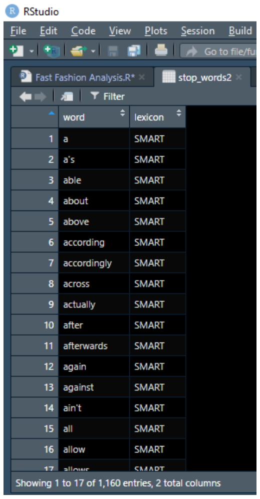
	word	lexicon
1	https	CUSTOM
2	17	CUSTOM
3	92	CUSTOM
4	20	CUSTOM
5	2019	CUSTOM
6	2018	CUSTOM
7	2019032026843	CUSTOM
8	xjsutum_kyo.twitter	CUSTOM
9	khaleejtimes	CUSTOM
10	fashionunited.com	CUSTOM
11	https	CUSTOM

Figure A 7: Fast Fashion Stop Words

- This line of code creates a table data frame in the environment with the newest words that I want to include to the preloaded data of stop_words in the English language (See Figure A7).

```
stop_words2 <- stop_words %>%
  bind_rows(stopff)
```

R Code Output:



	word	lexicon
1	a	SMART
2	a's	SMART
3	able	SMART
4	about	SMART
5	above	SMART
6	according	SMART
7	accordingly	SMART
8	across	SMART
9	actually	SMART
10	after	SMART
11	afterwards	SMART
12	again	SMART
13	against	SMART
14	ain't	SMART
15	all	SMART
16	allow	SMART

Showing 1 to 17 of 1,160 entries, 2 total columns

Figure A 8: Combined Stop Words

- This line of code creates a table data frame in the environment with the combination of the stop_words data frame and the stopff data frame (See Figure A8).
- This is going to be the stop_words data that is going to be used for the analysis of the articles.
- The new words were selected during the first steps of the understanding of the data, and now it is presented at the beginning of the code to be able to analyze the most important business insights and not the findings of unuseful words.

```
token_ff <- myffdf %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words2) %>%
  count(word, sort = TRUE)
```

R Code Output:

	word	n
1	fashion	733
2	fast	350
3	clothing	297
4	environmental	194
5	industry	194
6	sustainable	159
7	consumers	140
8	consumption	135
9	clothes	122
10	consumer	116
11	production	111
12	health	102
13	cashmere	96
14	cotton	88
15	textile	88
16	supply	84

Showing 1 to 17 of 4,274 entries, 2 total columns

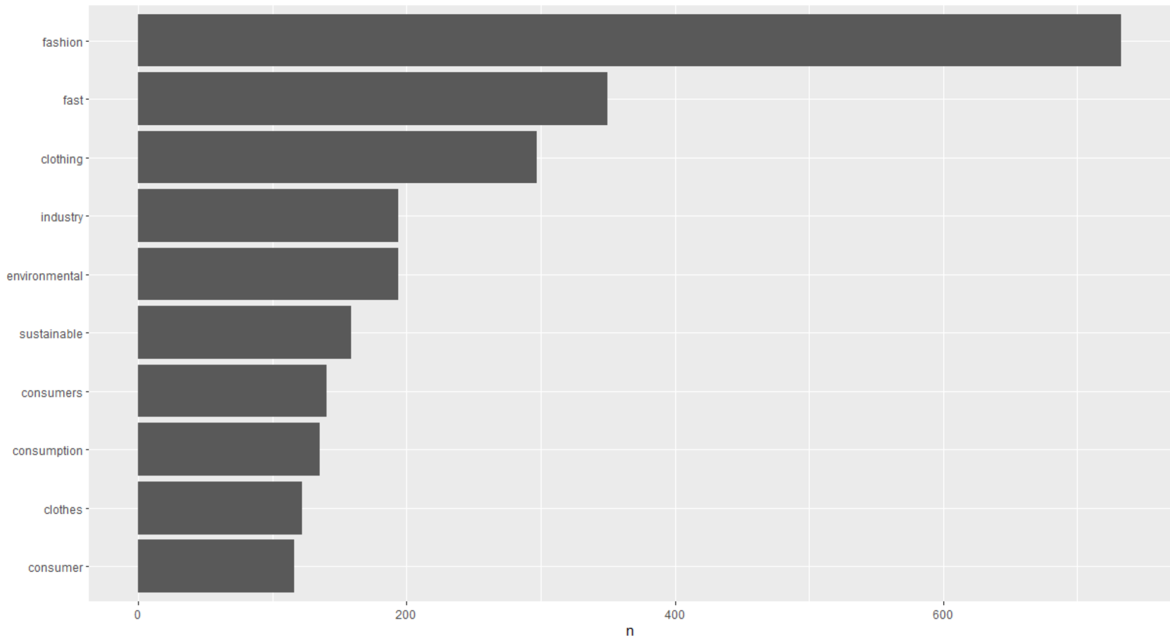
Figure A 9: Total Token Counts without Stop Words

- This line of code creates a data frame in the environment with the count of the words in all the articles without the stop_words previously selected (See Figure A9).

```
### Seeing and analyzing the Total Frequency in all the documents ###
tfhist_ff <- token_ff %>%
  mutate(word=reorder(word, n)) %>%
  top_n(10,n) %>%
  arrange(desc(n)) %>%
  ggplot(aes(word, n))+
  geom_col()+
  xlab(NULL)+
  coord_flip()
print(tfhist_ff)
```

R Code Output:

- This line of code creates the next bar chart with the most frequent words in the articles (See Graph A1).



Graph A 24: Token Frequencies Overall

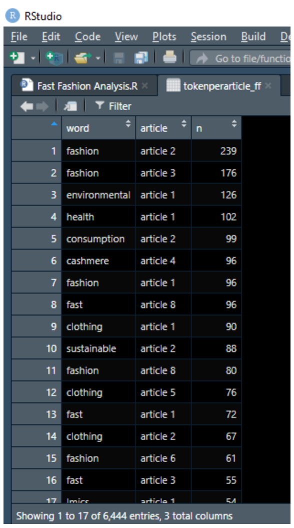
```
# As expected, the most common words in the article are fashion, fast, clothing and industry
```

It is important to see that environmental, sustainable, consumers and consumption are also common words due that problem that the fast fashion industry represents

Seeing and Analyzing Frequency per Article

```
tokenperarticle_ff <- myffdf %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words2) %>%
  count(word, article, sort = T)
```

R Code Output:



	word	article	n
1	fashion	article 2	239
2	fashion	article 3	176
3	environmental	article 1	126
4	health	article 1	102
5	consumption	article 2	99
6	cashmere	article 4	96
7	fashion	article 1	96
8	fast	article 8	96
9	clothing	article 1	90
10	sustainable	article 2	88
11	fashion	article 8	80
12	clothing	article 5	76
13	fast	article 1	72
14	clothing	article 2	67
15	fashion	article 6	61
16	fast	article 3	55

Showing 1 to 17 of 6,444 entries, 3 total columns

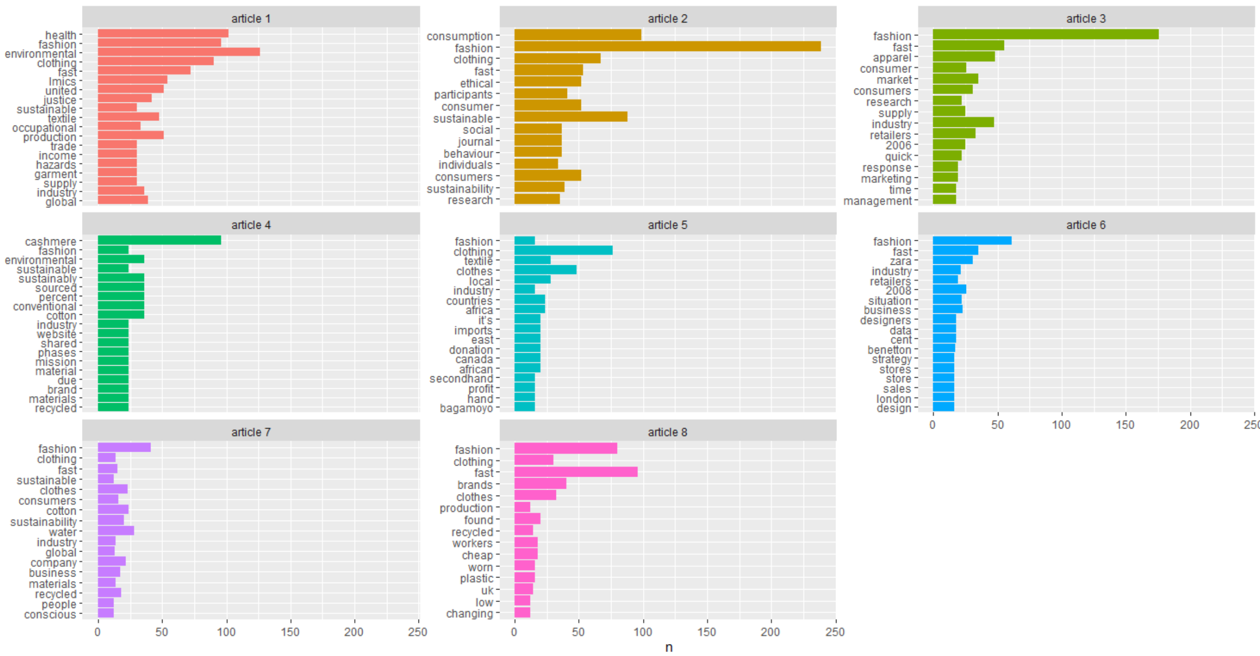
Figure A 10: Sample of the Tokenized data frame including article variable

- This line of code creates a data frame with the tokenization of the words including the variable of the article for each word (See Figure A10).

```
freq_hist_ff <- tokenperarticle_ff %>%
  group_by(article) %>%
  top_n(15) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = article))+
  geom_col(show.legend = FALSE)+
```

```
  facet_wrap(~article, scales = "free_y")+
  xlab(NULL)+
  coord_flip()
print(freq_hist_ff)
```

R Code Output:



Graph A 25: Frequencies of the words per article

- These lines of code create a graph showing the most frequent words per article (See Graph A2).

Based on the frequency per article I assume:

1st Article is related with health and environmental problems related to fashion industry

2nd Article is related with the consumption of clothing and the ethics behind the fashion industry

3rd Article is can be an overall research of the market and the changes with the consumer behavior and the fashion industry

4th Article is related with the cashmere (material used in many clothes) and probably with it's impact in the environment and the use of maybe other type of materials

5th Article is related with the textile industry in Africa

6th Article is related with the company Zara and it's role in the fast fashion industry

7th Article is probably related with an overall document about fast fashion and it's impact in the world

8th Article is probably related with the principal brand that are doing something or are someway related with the fast fashion industry

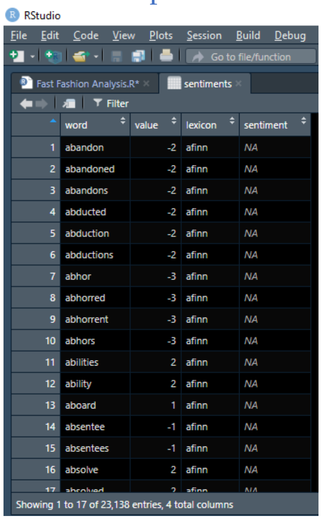
```
##########################
### Sentiment Analysis ###
##########################
# Calling data
afinn <- get_sentiments("afinn")
nrc <- get_sentiments("nrc")
bing <- get_sentiments("bing")
```

R Code Output:

- These lines of code create 3 table data frames in the environment with the sentiment data that are storage in the afinn, nrc and bing data frames.

```
sentiments <- bind_rows(mutate(afinn, lexicon="afinn"),
            mutate(nrc, lexicon= "nrc"),
            mutate(bing, lexicon="bing")
)
```

R Code Output:



	word	value	lexicon	sentiment
1	abandon	-2	afinn	NA
2	abandoned	-2	afinn	NA
3	abandons	-2	afinn	NA
4	abducted	-2	afinn	NA
5	abduction	-2	afinn	NA
6	abductions	-2	afinn	NA
7	abhor	-3	afinn	NA
8	abhorred	-3	afinn	NA
9	abhorrent	-3	afinn	NA
10	abhors	-3	afinn	NA
11	abilities	2	afinn	NA
12	ability	2	afinn	NA
13	aboard	1	afinn	NA
14	absentee	-1	afinn	NA
15	absentees	-1	afinn	NA
16	absolve	2	afinn	NA

Showing 1 to 17 of 23,138 entries, 4 total columns

Figure A 11: Sentiments Data Frame

- These lines of code combine the 3 table data frames previously created into a single data frame called sentiments (See Figure A11).

token_ffs <- myffdf %>%

unnest_tokens(word, text)

R Code Output:

- These lines of code create a data frame in the environment with the tokenized words per article.

afinn_ff <- token_ffs %>%

inner_join(get_sentiments("afinn"))%>%

summarise(sentiment=sum(value)) %>%

mutate(method="AFINN")

R Code Output:

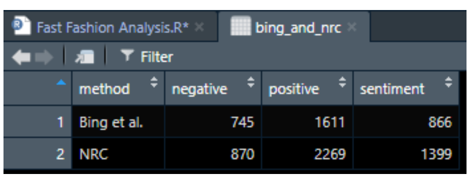
	sentiment	method
1	842	AFINN

Figure A 12: Summarize inner join between AFFIN data and tokenized words of all the articles.

- These lines of code create a data frame in the environment with the total afinn sentiments that are in the tokenized words of all the articles (See Figure A12).

```
bing_and_nrc <- bind_rows(
  token_ffs%>%
    inner_join(get_sentiments("bing"))%>%
    mutate(method = "Bing et al."),
  token_ffs %>%
    inner_join(get_sentiments("nrc") %>%
              filter(sentiment %in% c("positive", "negative"))) %>% #I'm extracting the sentiments as a binary variable
    mutate(method = "NRC")) %>%
  count(method,  sentiment) %>%
  spread(sentiment, n, fill=0) %>%
  mutate(sentiment = positive-negative)
```

R Code Output:



	method	negative	positive	sentiment
1	Bing et al.	745	1611	866
2	NRC	870	2269	1399

Figure A 13: Summarize inner join between Bing and NRC data and tokenized words of all the articles

- These lines of code create a data frame in the environment with the total combination of bing and nrc sentiments that are in the tokenized words of all the articles (See Figure A13).

```
bind_rows(afinn_ff, bing_and_nrc) %>%
  ggplot(aes(method, sentiment, fill=method))+
  geom_col(show.legend=FALSE)+
  facet_wrap(~method, ncol =1, scales= "free_y")
```

R Code Output:

Graph A 26: Sentiment Count per sentiment library in the overall text of the articles

- These lines of code create a plot with the scores of the number of sentiments depending of its source include in all the documents (See Graph A3).

```
# In general, the articles have positive sentiments.

# Analyzing the sentiments
bing_counts <- token_ffs %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=T) %>%
  ungroup()
```

R Code Output:

- These lines of code create a data frame with the count of the tokenized words per type of sentiment. The sentiments were classified as positive or negative for this classification using the bing sentiments library.

```
bing_counts %>%
  group_by(sentiment) %>%
```

top_n(10) %>%

ungroup() %>%

mutate(word=reorder(word, n)) %>%

ggplot(aes(word, n, fill=sentiment)) +

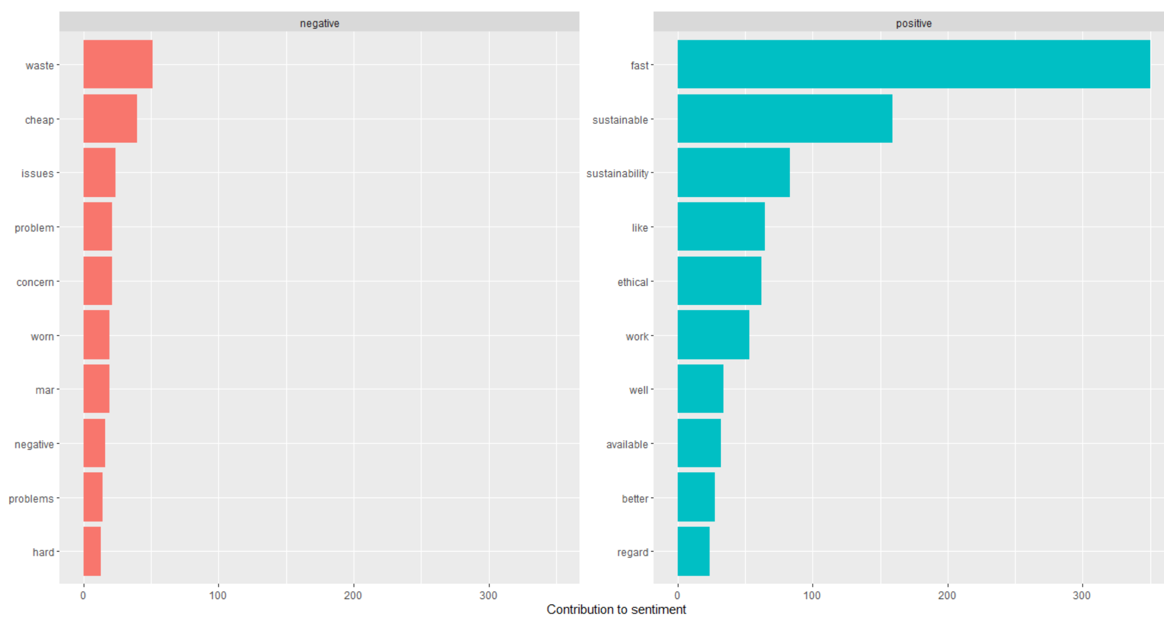
geom_col(show.legend = FALSE) +

facet_wrap(~sentiment, scales = "free_y")+ #this line of code puts our sentiment in columns (we are going to have as many columns as sentiments)

labs(y="Contribution to sentiment", x=NULL)+

coord_flip()

R Code Output:



Graph A 27: Most frequent positive and negative sentiments in the Fast Fashion articles.

- These lines of code create a plot of the classified sentiments and shows the most frequent ones in the overall documents (See Graph A4).

In general, we can see that there are a lot of different sentiments in the articles used for this

analysis. Nevertheless, analyzing the results there are some sentiments that are classified in the

wrong group, for example, fast. This one of the ones with higher frequency and it is show as a

positive one. Nevertheless, in our context it represents a negative sentiment.

################

TF - IDF

################

```
ff_token <- myffdf %>%
  unnest_tokens(word, text) %>%
  count(article, word, sort=TRUE) %>%
  ungroup()
```

R Code Output:

	article	word	n
1	article 1	the	477
2	article 2	the	394
3	article 2	of	370
4	article 1	and	348
5	article 2	and	326
6	article 1	of	291
7	article 2	to	291
8	article 3	the	277
9	article 5	the	256
10	article 6	the	245
11	article 2	fashion	239

Figure A 14: Token Frequency with the article variable

- These lines of code create a data frame with the frequency of the words of all the documents and includes the article as a variable (See Figure A14).

```
total_words <- ff_token %>%
  group_by(article) %>%
  summarize(total=sum(n))
```

R Code Output:

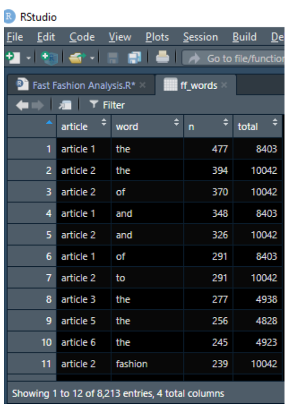
	article	total
1	article 1	8403
2	article 2	10042
3	article 3	4938
4	article 4	3048
5	article 5	4828
6	article 6	4923
7	article 7	4022
8	article 8	3728

Figure A 15: Total words per article data frame

- These lines of code create a table data frame with the total words per article (See Figure A15).

```
ff_words <- left_join(ff_token, total_words)
```

R Code Output:



	article	word	n	total
1	article 1	the	477	8403
2	article 2	the	394	10042
3	article 2	of	370	10042
4	article 1	and	348	8403
5	article 2	and	326	10042
6	article 1	of	291	8403
7	article 2	to	291	10042
8	article 3	the	277	4938
9	article 5	the	256	4828
10	article 6	the	245	4923
11	article 2	fashion	239	10042

Showing 1 to 12 of 8,213 entries, 4 total columns

Figure A 16: Tokenized Data Frame with article, frequencies, and total amount of words per article.

- This line of code joins the total ff_token with the inner join with total words. Giving us a data_frame with the frequencies of each word with the total words per article next to the other (See Figure A16).

```
ggplot(ff_words, aes(n/total, fill = article))+
  geom_histogram(show.legend=FALSE)+
  xlim(NA, 0.001) +
  facet_wrap(~article, ncol=2, scales="free_y")
```

R Code Output:

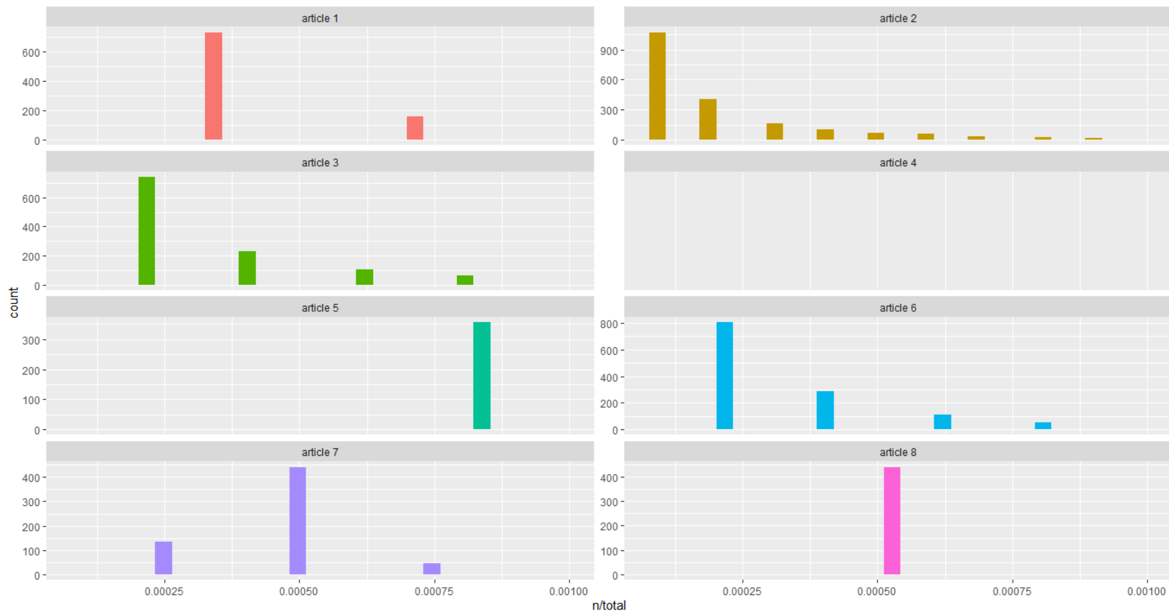
Console Message:

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning messages:

1: Removed 1453 rows containing non-finite values (stat_bin).

2: Removed 7 rows containing missing values (geom_bar).



Graph A 28: Token Histogram per Article

- These lines of code give us a histogram per article, the bigger lines represent the most frequent words (See Graph A5).

The result is not the expected. Adjust the formula and re-run the code

Adjusting bins to a lower level and xlim y to a higher one. This must be done because the frequencies of the words are small due of the total amount of words.

```
ggplot(ff_words, aes(n/total, fill = article))+
  geom_histogram(bins = 10,show.legend=FALSE)+
  xlim(NA, 0.01) +
  facet_wrap(~article, ncol=2, scales="free_y")
```

R Code Output:

Warning messages:

1: Removed 86 rows containing non-finite values (stat_bin).

2: Removed 8 rows containing missing values (geom_bar).

Graph A 29: Token Histogram per Article Adjusted

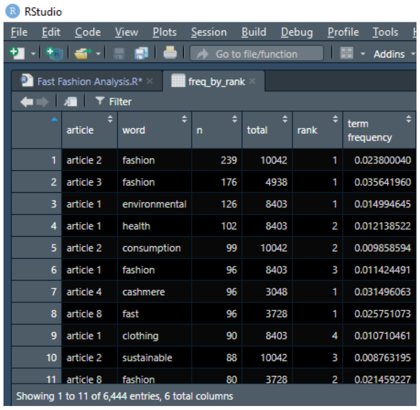
- These lines of code give us a histogram per article, the bigger lines represent the most frequent words.

The result is better. The articles, in general, are normally distributed and right skewed. There are other words that can give us business insights about the fast fashion industry.

```
################
### ZIPF's law ###
################
freq_by_rank <- ff_words %>%
  anti_join(stop_words2) %>%
  group_by(article) %>%
  mutate(rank = row_number(),
```

`term frequency` = n/total)

R Code Output:



RStudio

File Edit Code View Plots Session Build Debug Profile Tools

Fast Fashion Analysis.R* | freq_by_rank

	article	word	n	total	rank	term frequency
1	article 2	fashion	239	10042	1	0.023800040
2	article 3	fashion	176	4938	1	0.035641960
3	article 1	environmental	126	8403	1	0.014994645
4	article 1	health	102	8403	2	0.012138522
5	article 2	consumption	99	10042	2	0.009858594
6	article 1	fashion	96	8403	3	0.011424491
7	article 4	cashmere	96	3048	1	0.031496063
8	article 8	fast	96	3728	1	0.025751073
9	article 1	clothing	90	8403	4	0.010710461
10	article 2	sustainable	88	10042	3	0.008763195
11	article 8	fashion	80	3728	2	0.021459227

Showing 1 to 11 of 6,444 entries, 6 total columns

Figure A 17: Frequency by Rank Data Frame

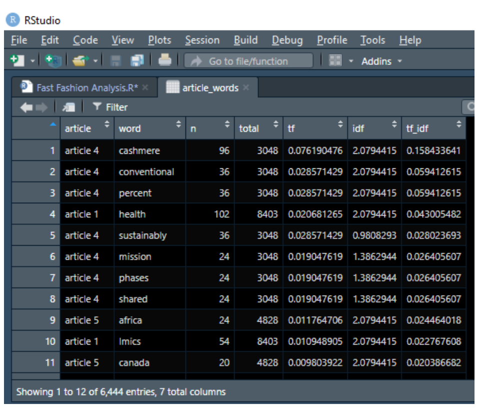
- These lines of code give us a grouped data frame with the rank and the term frequency on it (See Figure A17).

As we saw in the first part fashion is the most common word. There are other words as fashion, environmental, consumption, clothing that we know that are frequent, but are not giving us different business insights.

Seeing the most unique but most frequent words in the articles

```
article_words <- ff_words %>%
  anti_join(stop_words2) %>%
  bind_tf_idf(word, article, n) %>%
  arrange(desc(tf_idf))
```

R Code Output:



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Fast Fashion Analysis.R* | article_words

	article	word	n	total	tf	idf	tf_idf
1	article 4	cashmere	96	3048	0.076190476	2.0794415	0.158433641
2	article 4	conventional	36	3048	0.028571429	2.0794415	0.059412615
3	article 4	percent	36	3048	0.028571429	2.0794415	0.059412615
4	article 1	health	102	8403	0.020681265	2.0794415	0.043005482
5	article 4	sustainably	36	3048	0.028571429	0.9808293	0.028023693
6	article 4	mission	24	3048	0.019047619	1.3862944	0.026405607
7	article 4	phases	24	3048	0.019047619	1.3862944	0.026405607
8	article 4	shared	24	3048	0.019047619	1.3862944	0.026405607
9	article 5	africa	24	4828	0.011764706	2.0794415	0.024464018
10	article 1	lmics	54	8403	0.010948905	2.0794415	0.022767608
11	article 5	canada	20	4828	0.009803922	2.0794415	0.020386682

Showing 1 to 12 of 6,444 entries, 7 total columns

Figure A 18: tf_idf data frame

- These lines of code clean the stop words in the ranked frequency data frame and calculate the tf_idf and the idf for each word per article. The data frame is organized based on the tf_idf score (See Figure A18).

```
article_words %>%
  arrange(desc(tf_idf)) %>%
  mutate(word=factor(word, levels=rev(unique(word)))) %>%
  group_by(article) %>%
  top_n(5) %>%
  ungroup %>%
  ggplot(aes(word, tf_idf, fill=article))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="tf-idf")+
  facet_wrap(~article, ncol=2, scales="free")+
  coord_flip()
```

R Code Output:

Graph A 30: Most Unique and Most Frequent Tokens per Article

- These lines of code create a graph with the most frequent and unique words per article (See Graph A7).

there are some numbers as 2006 that can mean something important about data.

names of brands as Benetton or Zara can give us new insights about the article 6

we have new sustainable materials as piñatex, recycle materials as plastic, that tell us about more important innovations.

about our article from Africa, we have an specific location Bagamoyo in Tanzania.

###############

Bigrams

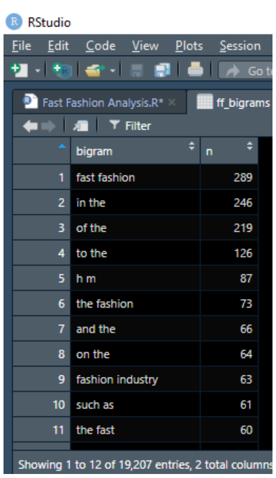
###############

The semantic structure is something important. Thus, the analysis of the documents is going to be run with some parts of the speech that can give more insights about the fast fashion industry.

The first part showed that the documents are not too extensive (based on the frequency of the words). Analyzing bigrams is a good method to find out more business insights and can clear if the sentiment analysis is accurate or not.

```
ff_bigrams <- myffdf %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2) %>%
  count(bigram, sort = TRUE)
```

R Code Output:



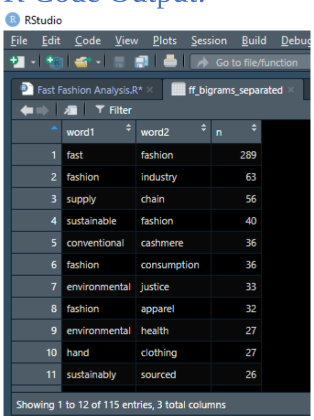
	bigram	n
1	fast fashion	289
2	in the	246
3	of the	219
4	to the	126
5	h m	87
6	the fashion	73
7	and the	66
8	on the	64
9	fashion industry	63
10	such as	61
11	the fast	60

Figure A 19: Bigrams Frequency

- These lines of code create a data frame with the frequencies of each bigram in all the articles (See Figure A19).

```
ff_bigrams_separated <- ff_bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ") %>%
  filter(!word1 %in% stop_words2$word) %>%
  filter(!word2 %in% stop_words2$word) %>%
  top_n(100,n)
```

R Code Output:



	word1	word2	n
1	fast	fashion	289
2	fashion	industry	63
3	supply	chain	56
4	sustainable	fashion	40
5	conventional	cashmere	36
6	fashion	consumption	36
7	environmental	justice	33
8	fashion	apparel	32
9	environmental	health	27
10	hand	clothing	27
11	sustainably	sourced	26

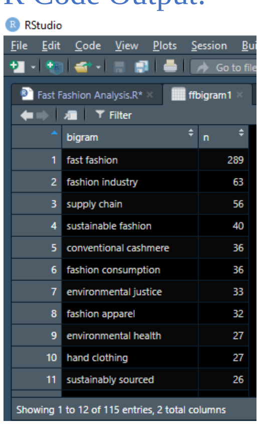
Showing 1 to 12 of 115 entries, 3 total columns

Figure A 20: Cleaned and Separated Bigram Dataframe

- These lines of code separate the bigrams into 2 columns and clean all the observations that have a stop word (See Figure A20).

```
ffbigram <- ff_bigrams_separated %>%
  unite ("bigram", 1:2 ,sep = " ", remove = FALSE, na.rm = FALSE)
ffbigram1 <- cbind.data.frame(bigram = ffbigram$bigram, n = ffbigram$n)
```

R Code Output:



	bigram	n
1	fast fashion	289
2	fashion industry	63
3	supply chain	56
4	sustainable fashion	40
5	conventional cashmere	36
6	fashion consumption	36
7	environmental justice	33
8	fashion apparel	32
9	environmental health	27
10	hand clothing	27
11	sustainably sourced	26

Showing 1 to 12 of 115 entries, 2 total columns

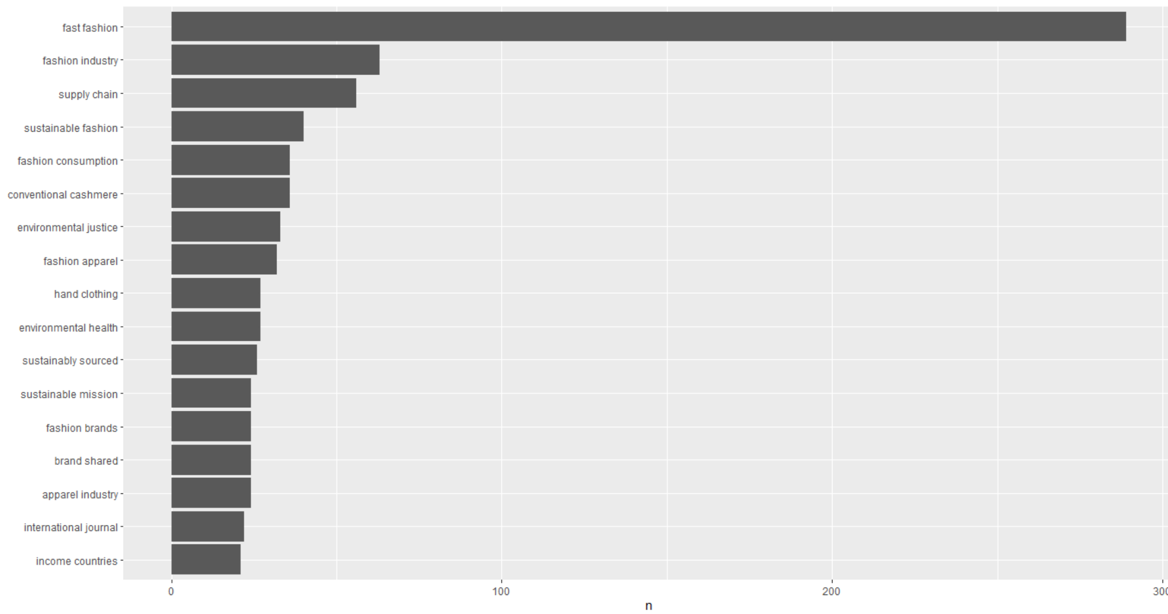
Figure A 21: Cleaned Bigrams

-These lines of code combine the cleaned columns into one and then extract the columns of interest bigram and frequency of each one into a new data frame (See Figure A21).

```
freq_ffb <- ffbigram1 %>%
```

```
  mutate(bigram=reorder(bigram, n)) %>%
  filter(n > 20) %>%
  ggplot(aes(bigram, n))+
  geom_col()+
  xlab(NULL)+
  coord_flip()
print(freq_ffb)
```

R Code Output:



Graph A 31: Bigrams Frequency

- These lines of code create a graph that shows the most frequent bigrams in all the articles combined (See Graph A8).

Different results as the supply chain, the environmental justice, the hand clothing, international journal and income countries.

Can give us more insights about what fast fashion is and why it is important to know more about it as consumers.

```
# Bigram Networks

bigram_graph <- ff_bigrams_separated %>%
  filter(n>10) %>%
  graph_from_data_frame()

bigram_graph
```

R Code Output:

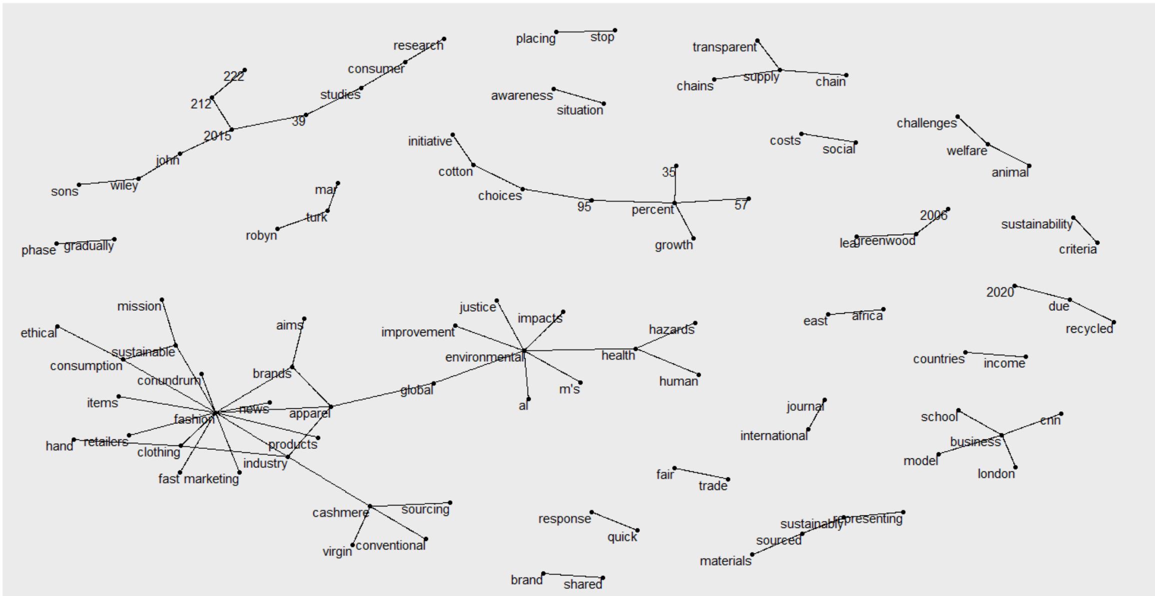
Console Output:

```
IGRAPH f5ccb79 DN-- 97 81 --
+ attr: name (v/c), n (e/n)
+ edges from f5ccb79 (vertex names):
 [1] fast          ->fashion      fashion       ->industry
 [3] supply        ->chain        sustainable   ->fashion
 [5] conventional  ->cashmere     fashion       ->consumption
 [7] environmental->justice       fashion       ->apparel
 [9] environmental->health        hand          ->clothing
[11] sustainably   ->sourced      apparel       ->industry
[13] brand         ->shared       fashion       ->brands
[15] sustainable   ->mission      international->journal
+ ... omitted several edges
>
```

```
ggraph(bigram_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
```

geom_node_text(aes(label=name), vjust =1, hjust=1)

R Code Output:



Graph A 32: Bigram Network

- These lines of code create a bigram network that shows more semantic structures that can help to understand more about fast fashion industry (See Graph A9).

With this analysis we can see more things about the documents and the frequent phrases.

An example is that John Wiley & Sons are the editors of one of the documents that is an international journal of consumer studies.

Nevertheless, we cannot clean all this words or the numbers because can be some way related also with the business insights that can be taken from other documents.

Let's analyze the same framework per article

```
ff_articlebigrams <- myffdf %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2) %>%
  count(bigram, article, sort = T)
ff_articlebigrams_separated <- ff_articlebigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ") %>%
```

```
  filter(!word1 %in% stop_words2$word) %>%
  filter(!word2 %in% stop_words2$word) %>%
  top_n(100,n)
```

R Code Output:

- These lines of code create a data frame without stop words with the article variable as part of the data frame.

```
## article 1 ##
bigrama1_graph <- ff_articlebigrams_separated %>%
  filter(article == "article 1") %>%
  graph_from_data_frame()
bigrama1_graph
```

R Code Output:

Console Output:

```
IGRAPH b6d442c DN-- 28 18 --
+ attr: name (v/c), article (e/c), n (e/n)
+ edges from b6d442c (vertex names):
 [1] fast         ->fashion       environmental->justice
 [3] environmental->health        income       ->countries
 [5] supply       ->chain         health       ->hazards
 [7] human        ->health        al           ->environmental
 [9] fashion      ->industry      global       ->environmental
[11] social       ->costs         creative     ->commons
[13] eco          ->friendly      fair         ->trade
```

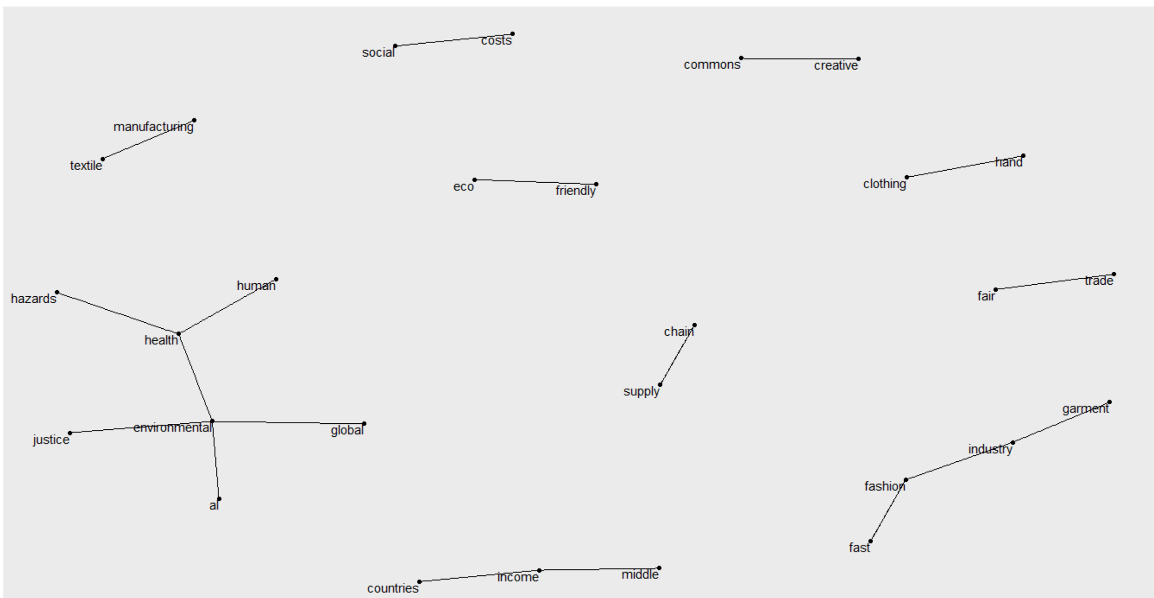
[15] garment ->industry hand ->clothing

+ ... omitted several edges

>

```
ggraph(bigrama1_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 33: Bigram Network Article 1

- These lines of code create a bigram network for the first article (See Graph A10)

```
## article 2 ##
bigrama2_graph <- ff_articlebigrams_separated %>%
  filter(article == "article 2") %>%
  graph_from_data_frame()
bigrama2_graph
```

R Code Output:

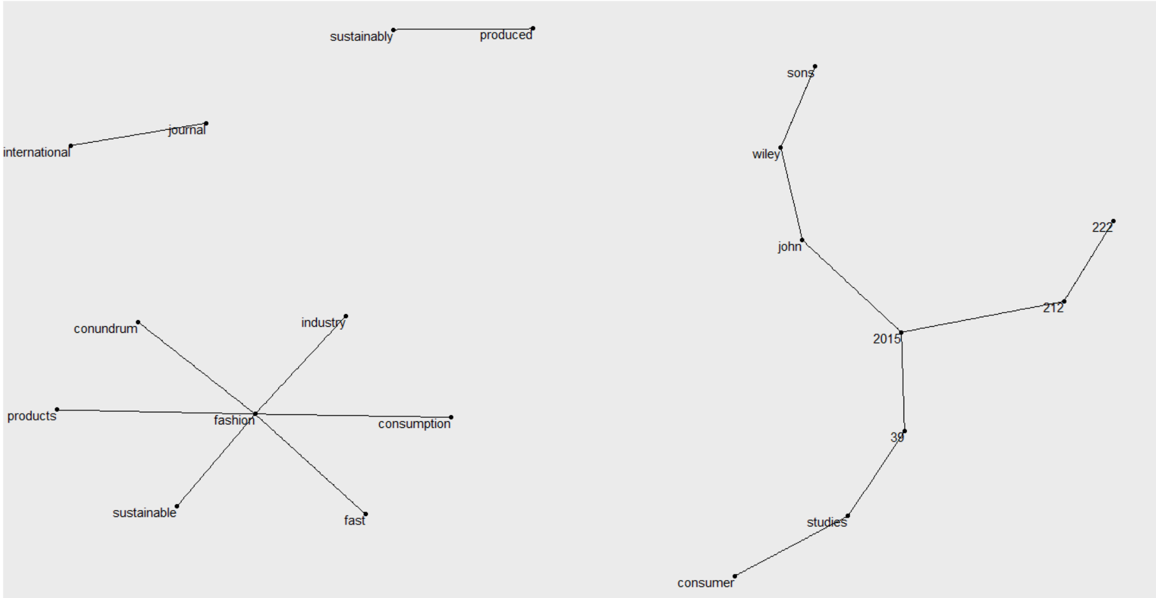
Console Output:

```
IGRAPH cf22582 DN-- 20 16 --
+ attr: name (v/c), article (e/c), n (e/n)
+ edges from cf22582 (vertex names):
 [1] fast         ->fashion     sustainable  ->fashion
 [3] fashion      ->consumption international->journal
 [5] consumer     ->studies     fashion      ->products
 [7] 2015         ->212         2015         ->john
 [9] 212          ->222         39           ->2015
[11] fashion      ->conundrum   john         ->wiley
[13] studies      ->39          wiley        ->sons
[15] fashion      ->industry    sustainably  ->produced
>
```

```
ggraph(bigrama2_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:

- These lines of code create a bigram network for the first article (See Graph A11)



Graph A 34: Bigram Network Article 2

As this one is the article related with jhon wiley we are going to remove the words that we found that are not giving us business insights just for this one.

```
stopffa2 <- tribble(~word,~lexicon,
        "https", "CUSTOM",
        "17", "CUSTOM",
        "92", "CUSTOM",
        "20", "CUSTOM",
        "2019", "CUSTOM",
        "2018", "CUSTOM",
        "2019032026843", "CUSTOM",
        "xjsutum_kyo.twitter", "CUSTOM",
        "khaleejtimes", "CUSTOM",
        "fashionunited.com", "CUSTOM",
        "https", "CUSTOM",
```

```
          "john", "CUSTOM",
          "wiley", "CUSTOM",
          "sons", "CUSTOM",
          "212", "CUSTOM",
          "222", "CUSTOM",
          "39", "CUSTOM",
          "2015", "CUSTOM",
          "studies", "CUSTOM",
          "international", "CUSTOM",
          "journal", "CUSTOM"
)
stop_words2a2 <- stop_words %>%
  bind_rows(stopffa2)
ff_article2bigrams <- myffdf %>%
  filter(article == "article 2") %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2) %>%
  count(bigram, sort = T)
ff_article2bigrams_separated <- ff_article2bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ") %>%
  filter(!word1 %in% stop_words2a2$word) %>%
  filter(!word2 %in% stop_words2a2$word) %>%
  top_n(100,n)
```

R Code Output:

- These lines of code are the same process that were applied before, with a change in the stop words to exclude just the ones that are affecting the outcomes for this article.

```
cleanbigrama2_graph <- ff_article2bigrams_separated %>%
  filter(n > 3) %>%
  graph_from_data_frame()
cleanbigrama2_graph
```

R Code Output:

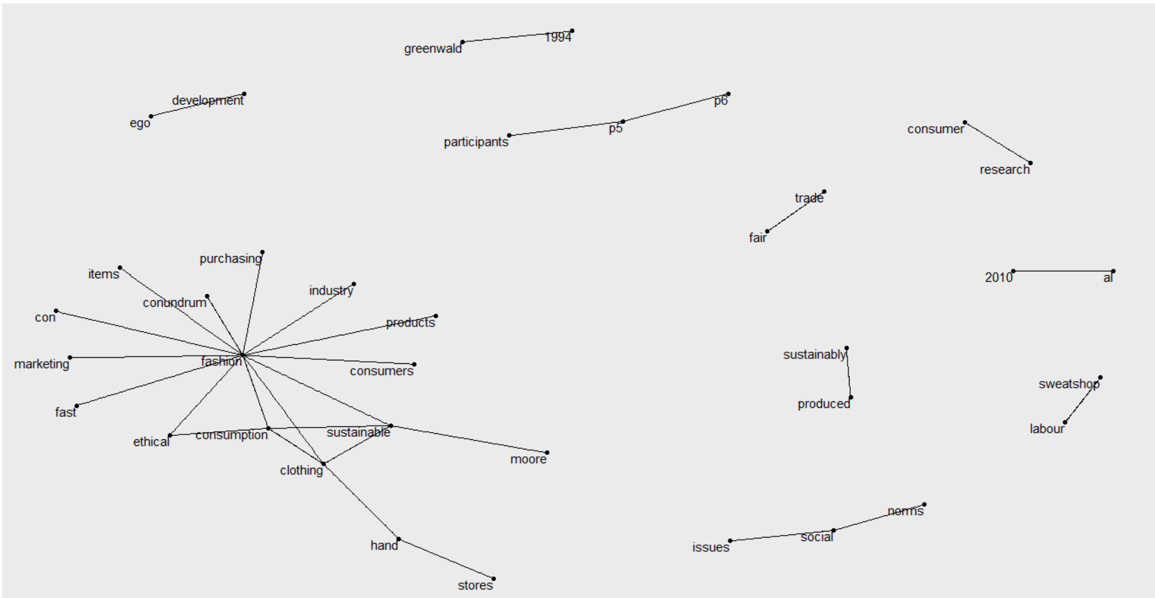
Console Output:

```
IGRAPH 5a7ad17 DN-- 37 31 --
+ attr: name (v/c), n (e/n)
+ edges from 5a7ad17 (vertex names):
 [1] fast       ->fashion     sustainable->fashion
 [3] fashion    ->consumption fashion    ->products
 [5] fashion    ->conundrum   fashion    ->industry
 [7] sustainably->produced    ethical    ->fashion
 [9] fashion    ->con         fashion    ->consumers
[11] fashion    ->items       hand       ->stores
[13] ethical    ->consumption fashion    ->clothing
[15] fashion    ->purchasing  hand       ->clothing
+ ... omitted several edges
>
```

```
ggraph(cleanbigrama2_graph, layout = "fr") +
  geom_edge_link()+
```

```
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 35: Cleaned Bigram Network Article 2

- These lines of code create the clean bigram network for the second article (See Graph A12)

```
# Now we are getting more insights about this particular article.
## article 3 ##
bigrama3_graph <- ff_articlebigrams_separated %>%
  filter(article == "article 3") %>%
  graph_from_data_frame()
bigrama3_graph
```

R Code Output:

Console Output:

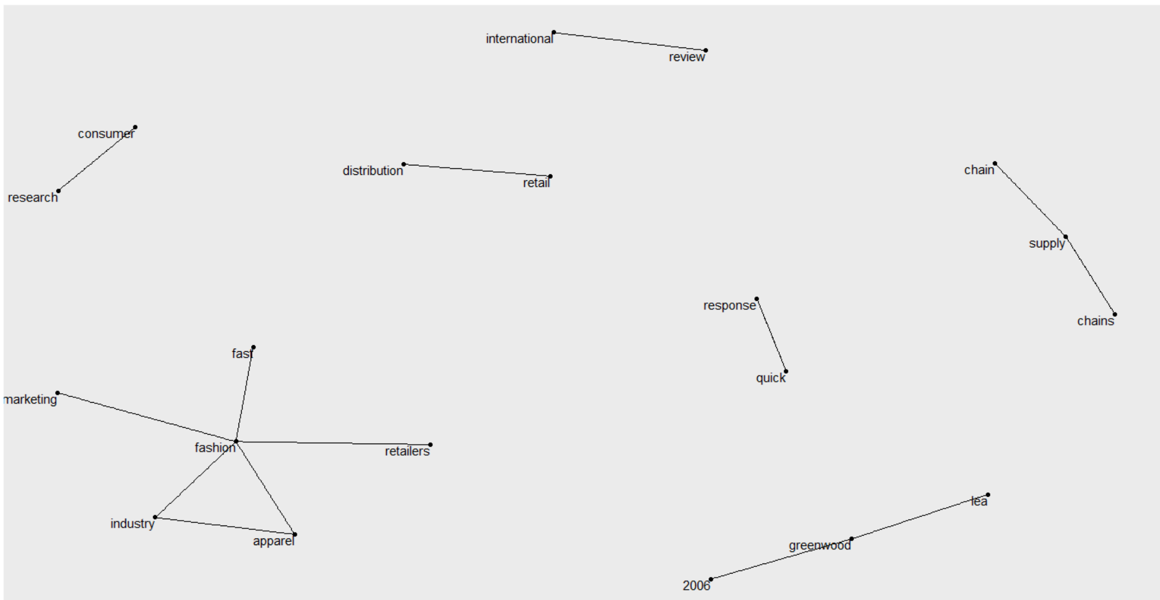
IGRAPH da5e77a DN-- 20 14 --

+ attr: name (v/c), article (e/c), n (e/n)

```
+ edges from da5e77a (vertex names):
 [1] fast         ->fashion      fashion      ->apparel
 [3] apparel      ->industry     fashion      ->industry
 [5] quick        ->response     supply       ->chain
 [7] fashion      ->marketing    fashion      ->retailers
 [9] greenwood    ->2006         lea          ->greenwood
[11] supply       ->chains       consumer     ->research
[13] international->review       retail       ->distribution
>
```

```
ggraph(bigrama3_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Result:



Graph A 36: Bigram Network Article 3

- These lines of code create a bigram network for the third article (See Graph A13)

This article gives us one interesting insight. They cite a lot of information from other research from Barnes, L., and G. Lea-Greenwood that was published in 2006 related with the fast fashion supply chain.

```
## article 4 ##
bigrama4_graph <- ff_articlebigrams_separated %>%
  filter(article == "article 4") %>%
  graph_from_data_frame()
bigrama4_graph
```

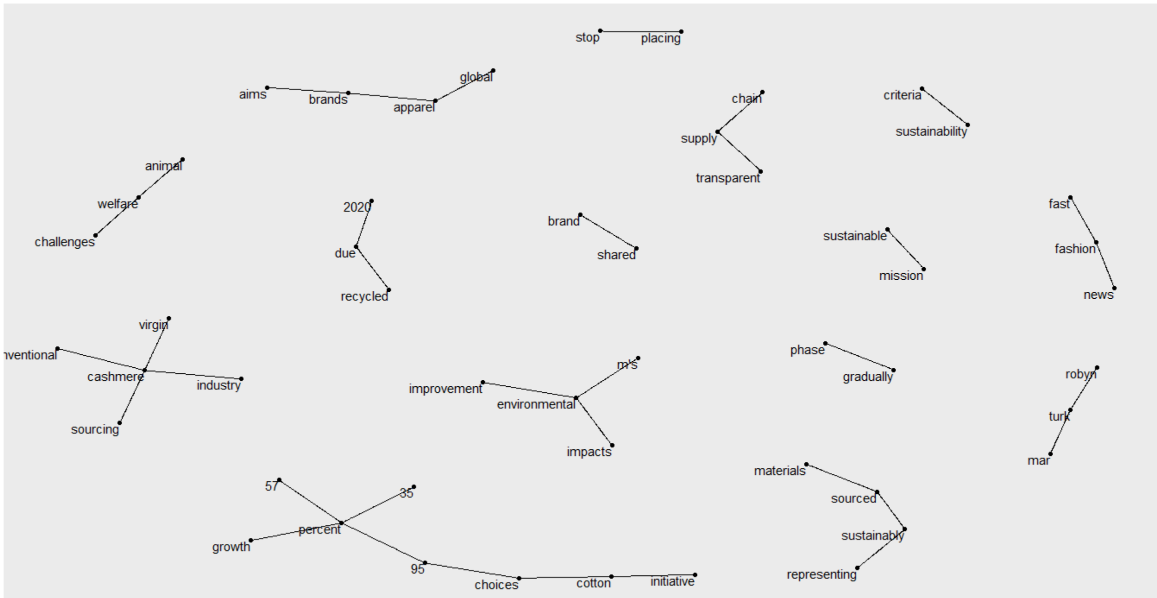
R Code Output:

Console Output:

```
IGRAPH 0fb2fe1 DN-- 50 36 --
+ attr: name (v/c), article (e/c), n (e/n)
+ edges from 0fb2fe1 (vertex names):
 [1] conventional ->cashmere    brand        ->shared
 [3] sustainable  ->mission     sustainably  ->sourced
 [5] 2020         ->due         35           ->percent
 [7] 57           ->percent     95           ->percent
 [9] animal       ->welfare     apparel      ->brands
[11] brands       ->aims        cashmere     ->industry
[13] choices      ->95          cotton       ->choices
[15] cotton       ->initiative  environmental->impacts
+ ... omitted several edges
>
```

```
ggraph(bigrama4_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 37: Bigram Network Article 4

- These lines of code create a bigram network for the article number 4 (See Graph A14).

Here we can see that the article is related with a gradually phase of the cashmere industry to cotton or more sustainable sourced materials.

```
## article 5 ##
bigrama5_graph <- ff_articlebigrams_separated %>%
  filter(article == "article 5") %>%
  graph_from_data_frame()
bigrama5_graph
```

R Code Output:

Console Output:

IGRAPH c4867a4 DN-- 15 9 --

+ attr: name (v/c), article (e/c), n (e/n)

+ edges from c4867a4 (vertex names):

[1] east ->africa fast ->fashion hand ->clothing

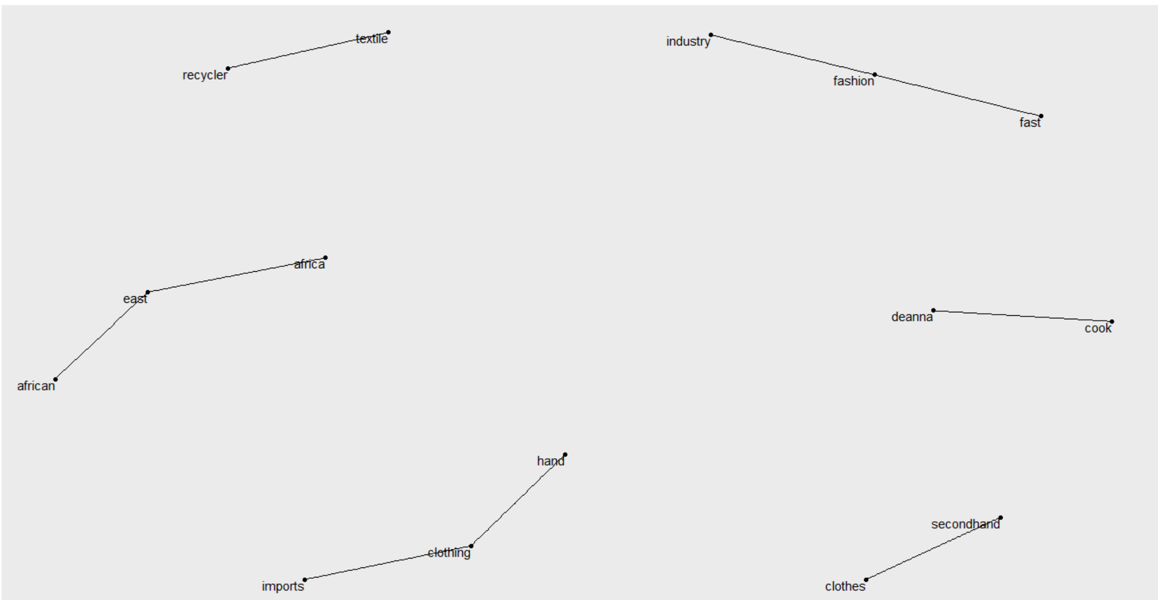
[4] clothing ->imports deanna ->cook east ->african

[7] fashion ->industry secondhand->clothes textile ->recycler

>

```
ggraph(bigrama5_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 38: Bigram Network Article 5

- These lines of code create a bigram network for the article number 5 (See Graph A15).

article 6

```
bigrama6graph <- ff_articlebigrams_separated %>%
  filter(article == "article 6") %>%
  graph_from_data_frame()
bigrama6graph
```

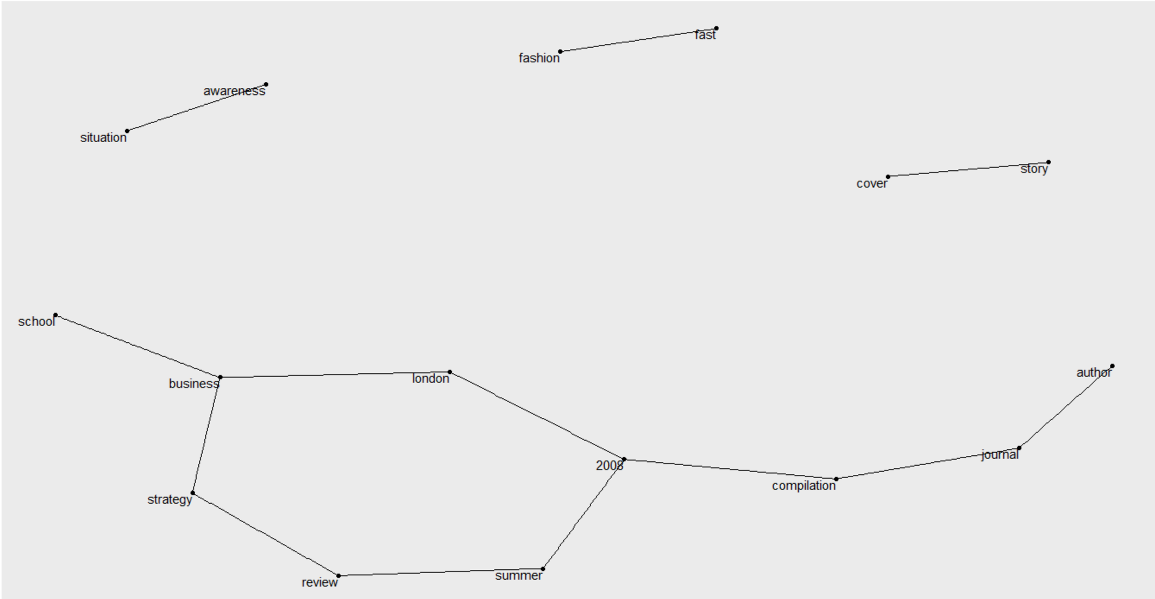
R Code Output:

Console Output:

```
IGRAPH 011d4bd DN-- 16 13 --
+ attr: name (v/c), article (e/c), n (e/n)
+ edges from 011d4bd (vertex names):
 [1] fast       ->fashion     business   ->school
 [3] london     ->business    situation  ->awareness
 [5] business   ->strategy    strategy   ->review
 [7] 2008       ->london      author     ->journal
 [9] compilation->2008        cover      ->story
[11] journal    ->compilation review     ->summer
[13] summer     ->2008
>
```

```
ggraph(bigrama6graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 39: Bigram Network Article 6

- These lines of code create a bigram network for the article number 6 (See Graph A16).

In this article we have a similar situation that in the article 2. It is a publication of a journal of the London Business School.

```
stopffa6<- tribble(~word,~lexicon,
                "https", "CUSTOM",
                "17", "CUSTOM",
                "92", "CUSTOM",
                "20", "CUSTOM",
                "2019", "CUSTOM",
                "2018", "CUSTOM",
                "2019032026843", "CUSTOM",
                "xjsutum_kyo.twitter", "CUSTOM",
                "khaleejtimes", "CUSTOM",
                "fashionunited.com", "CUSTOM",
```

```
                    "https", "CUSTOM",
                    "journal", "CUSTOM",
                    "compilation", "CUSTOM",
                    "london", "CUSTOM",
                    "business", "CUSTOM",
                    "school", "CUSTOM",
                    "strategy", "CUSTOM",
                    "review", "CUSTOM",
                    "summer", "CUSTOM",
)
stop_words2a6 <- stop_words %>%
  bind_rows(stopffa6)
ff_article6bigrams <- myffdf %>%
  filter(article == "article 6") %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2) %>%
  count(bigram, sort = T)
ff_article6bigrams_separated <- ff_article6bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ") %>%
  filter(!word1 %in% stop_words2a6$word) %>%
  filter(!word2 %in% stop_words2a6$word) %>%
  top_n(100,n)
```

R Code Outcome:

- These lines of code are the same process that were applied before, with a change in the stop words to exclude just the ones that are affecting the outcomes for this article.

```
cleanbigrama6_graph <- ff_article6bigrams_separated %>%
  filter(n > 3) %>%
  graph_from_data_frame()
cleanbigrama6_graph
```

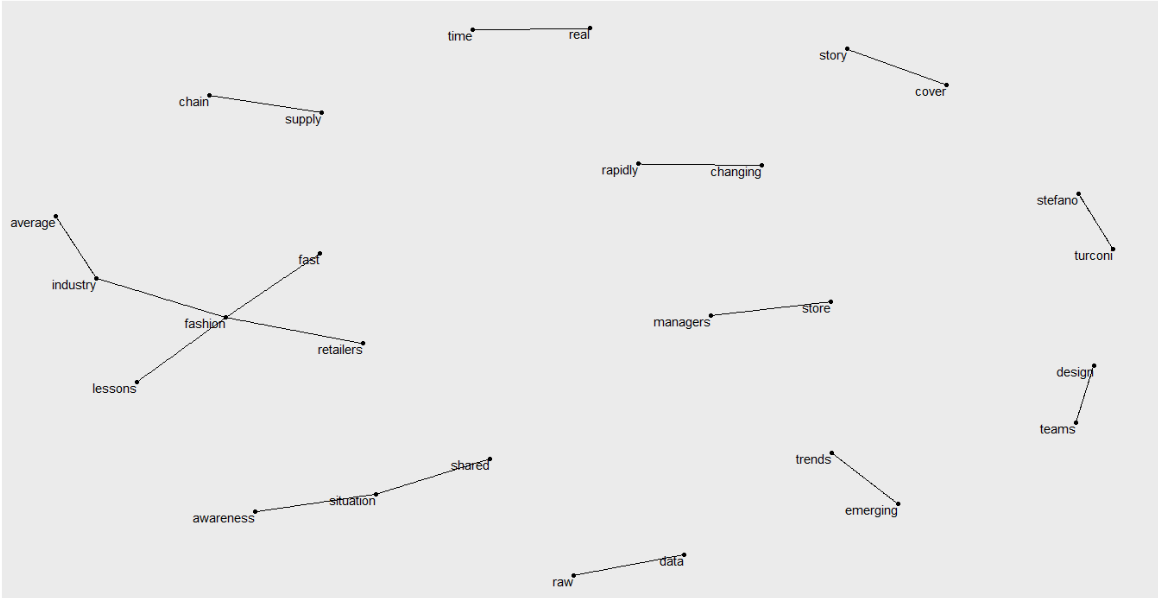
R Code Output:

Console Output:

```
IGRAPH 475632f DN-- 27 16 --
+ attr: name (v/c), n (e/n)
+ edges from 475632f (vertex names):
 [1] fast     ->fashion   situation->awareness cover    ->story
 [4] rapidly  ->changing  shared   ->situation stefano  ->turconi
 [7] fashion  ->retailers fashion  ->industry  industry ->average
[10] raw      ->data      store    ->managers  design   ->teams
[13] emerging ->trends    fashion  ->lessons   real     ->time
[16] supply   ->chain
>
```

```
ggraph(cleanbigrama6_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 40: Cleaned Bigram Network Article 6

- These lines of code create the clean bigram network for the article number 6 (See Graph A17).

The first thing that we can see cleaning this article, is that the bigrams have low frequencies so it is probably a short document related with the lessons about fast fashion retailers.

article 7

```
bigrama7_graph <- ff_articlebigrams_separated %>%
  filter(article == "article 7") %>%
  graph_from_data_frame()
bigrama7_graph
```

R Code Output:

Console Output:

IGRAPH 9d7f374 DN-- 8 5 --

+ attr: name (v/c), article (e/c), n (e/n)

+ edges from 9d7f374 (vertex names):

[1] fast ->fashion cnn ->business fashion ->industry

[4] told ->cnn conscious->collection

>

ggraph(bigrama7_graph, layout = "fr") +

geom_edge_link()+

geom_node_point()+

geom_node_text(aes(label=name), vjust =1, hjust=1)

R Code Output:

Graph A 41: Bigram Network Article 7

- These lines of code create a bigram network for the article number 7 (See Graph A18).

This one is a short note about fast fashion and the conscious collection

we are going to change our source so we can have more words for our business insights

ff_articlebigrams_separated_short <- ff_articlebigrams %>%

separate(bigram, c("word1", "word2"), sep = " ") %>%

filter(!word1 %in% stop_words2$word) %>%

filter(!word2 %in% stop_words2$word) %>%

```
  filter(n>3)
```

R Code Outcome:

- These lines of code are the same process that were applied before, with a change in frequency in the filter command. Smaller frequency.

```
bigrama7_graph_short <- ff_articlebigrams_separated_short %>%
  filter(article == "article 7") %>%
  graph_from_data_frame()
bigrama7_graph_short
```

R Code Output:

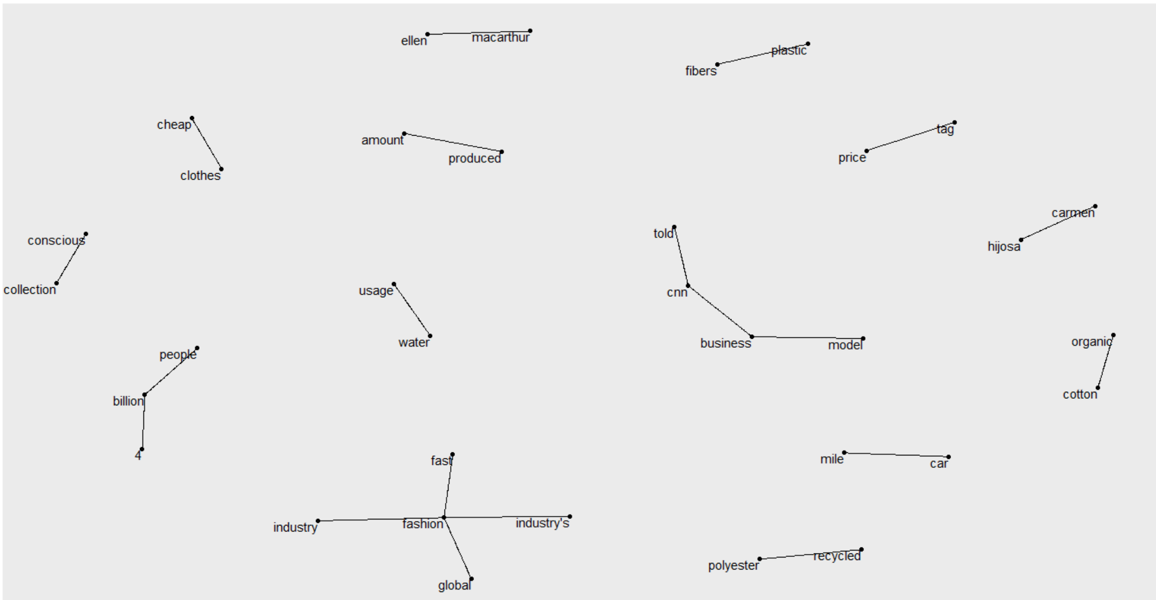
Console Output:

```
IGRAPH cf6bc46 DN-- 34 20 --
+ attr: name (v/c), article (e/c), n (e/n)
+ edges from cf6bc46 (vertex names):
 [1] fast     ->fashion    cnn      ->business   fashion  ->industry
 [4] told     ->cnn        conscious->collection ellen    ->macarthur
 [7] global   ->fashion    4        ->billion    amount   ->produced
[10] billion  ->people     business ->model      carmen   ->hijosa
[13] cheap    ->clothes    fashion  ->industry's mile     ->car
[16] organic  ->cotton     plastic  ->fibers     price    ->tag
[19] recycled ->polyester  water    ->usage
>
```

```
ggraph(bigrama7_graph_short, layout = "fr") +
  geom_edge_link()+
```

```
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 42: Improved Bigram Network Article 7

- These lines of code create the improved bigram network for the article number 7 (See Graph A19).

```
# This article analyzes the amount of people around the globe and the usage of different materials an resources and its relation with the price
## article 8 ##
bigrama8_graph <- ff_articlebigrams_separated %>%
  filter(article == "article 8") %>%
  graph_from_data_frame()
bigrama8_graph
```

R Code Output:

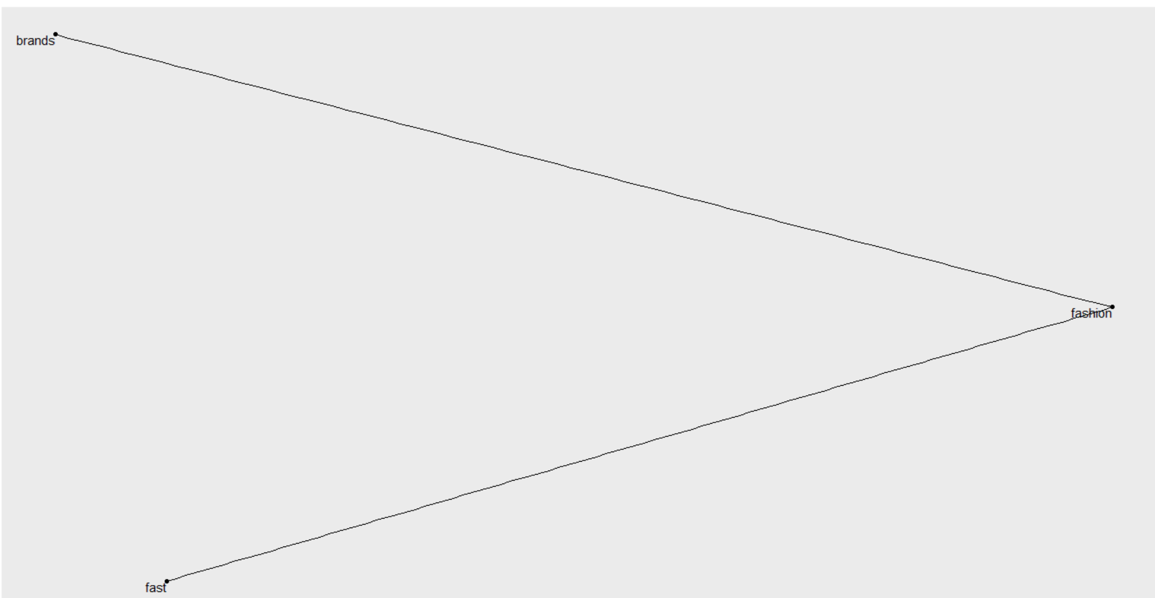
Console Output:

IGRAPH 046a784 DN-- 3 2 --

```
+ attr: name (v/c), article (e/c), n (e/n)
+ edges from 046a784 (vertex names):
[1] fast   ->fashion fashion->brands
>
```

```
ggraph(bigrama8_graph, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 43: Bigram Network Article 8

- These lines of code create a bigram network for the article number 8 (See Graph A20).

```
# Same problem as last article
bigrama8_graph_short <- ff_articlebigrams_separated_short %>%
  filter(article == "article 8") %>%
  graph_from_data_frame()
```

bigrama8_graph_short

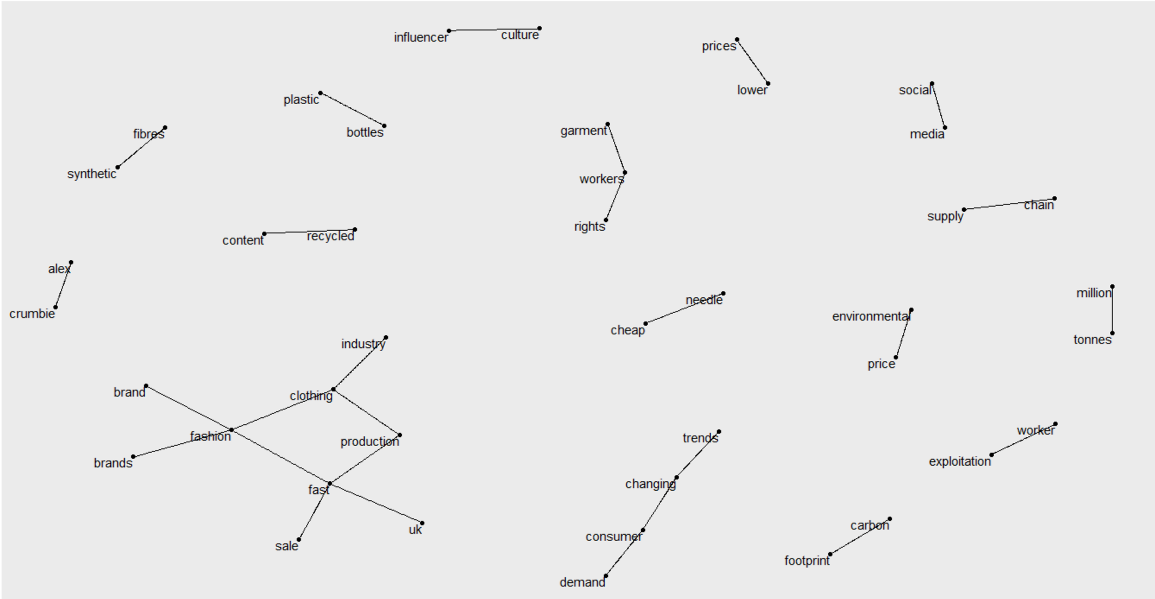
R Code Output:

Console Output:

```
IGRAPH 26b1743 DN-- 42 27 --
+ attr: name (v/c), article (e/c), n (e/n)
+ edges from 26b1743 (vertex names):
 [1] fast         ->fashion    fashion      ->brands
 [3] supply       ->chain      alex         ->crumbie
 [5] carbon       ->footprint  changing     ->consumer
 [7] changing     ->trends     cheap        ->needle
 [9] clothing     ->industry   clothing     ->production
[11] consumer     ->demand     environmental->price
[13] fashion      ->brand      fashion      ->clothing
[15] fast         ->production fast         ->sale
+ ... omitted several edges
>
```

```
ggraph(bigrama8_graph_short, layout = "fr") +
  geom_edge_link()+
  geom_node_point()+
  geom_node_text(aes(label=name), vjust =1, hjust=1)
```

R Code Output:



Graph A 44: Improved Bigram Network Article 8

- These lines of code create the improved bigram network for the article number 8 (See Graph A21).

A short document about the rights of the workers, analyze the social media and the influencer culture related with the fast fashion production and diferent brands, principally located in uk.

#####################

IDF + Bigrams

#####################

Based on the analysis, the data is going to be filtered again with the stop words that are not relevant in overall.

This last analysis is a combination of the frameworks TF_IFD and Bigrams and it is going to be applied to the overall information.

```
stopfinal <- tribble(~word,~lexicon,
        "https", "CUSTOM",
        "17", "CUSTOM",
```

"92", "CUSTOM",

"20", "CUSTOM",

"2019", "CUSTOM",

"2018", "CUSTOM",

"2019032026843", "CUSTOM",

"xjsutum_kyo.twitter", "CUSTOM",

"khaleejtimes", "CUSTOM",

"fashionunited.com", "CUSTOM",

"https", "CUSTOM",

"john", "CUSTOM",

"wiley", "CUSTOM",

"sons", "CUSTOM",

"212", "CUSTOM",

"222", "CUSTOM",

"39", "CUSTOM",

"2015", "CUSTOM",

"studies", "CUSTOM",

"international", "CUSTOM",

"journal", "CUSTOM",

"compilation", "CUSTOM",

"london", "CUSTOM",

"business", "CUSTOM",

"school", "CUSTOM",

```
                "strategy", "CUSTOM",
                "review", "CUSTOM",
                "summer", "CUSTOM",
                "cnn", "CUSTOM"
)
stopffwords <- stop_words %>%
  bind_rows(stopfinal)
finalbigram <- myffdf %>%
  unnest_tokens(bigram, text, token = "ngrams", n=2) %>%
  count(bigram, article, sort = TRUE)
```

R Code Outcome:

- These lines of code are the same process that were applied before when creating bigrams. The only difference is that the data for stop_words have change.

```
# Separating the data frame based on the analysis made in the bigrams per article (long and short documents)
# Filtering the longer articles
long_articles <- finalbigram %>%
  filter(!article == "article 7" & !article == "article 8")
```

R Code Outcome:

- These lines of code filter the finalbigram data frame for all the long articles (from the bigrams conclusions, 7 and 8 are the short ones).

```
short_articles <- finalbigram %>%
  filter(article == "article 7" | article == "article 8")
```

R Code Outcome:

- These lines of code filter the finalbigram data frame for the 2 short articles (from the bigrams conclusions, 7 and 8 are the short ones).

Important to notice that here we must use | instead of & or we are going to have an empty data frame.

Analyzing long articles

```
final_bigrams_separated_l <- long_articles %>%
  separate(bigram, c("word1", "word2"), sep = " ") %>%
  filter(!word1 %in% stopffwords$word) %>%
  filter(!word2 %in% stopffwords$word) %>%
  top_n(100,n)
fffinalbigram_l <- final_bigrams_separated_l %>%
  unite ("bigram", 1:2 ,sep = " ", remove = FALSE, na.rm = FALSE)
fffinalbigram1 <- cbind.data.frame(bigram = fffinalbigram_l$bigram, article =
fffinalbigram_l$article, n = fffinalbigram_l$n)
```

R Code Outcome:

- These lines of code create the final data frame with the frequencies of the bigrams for only the long articles.

```
total_bigrams <- fffinalbigram1 %>%
  group_by(article) %>%
  summarize(total=sum(n))
```

R Code Outcome:

- These lines of code create a table data frame with the total bigrams per article.

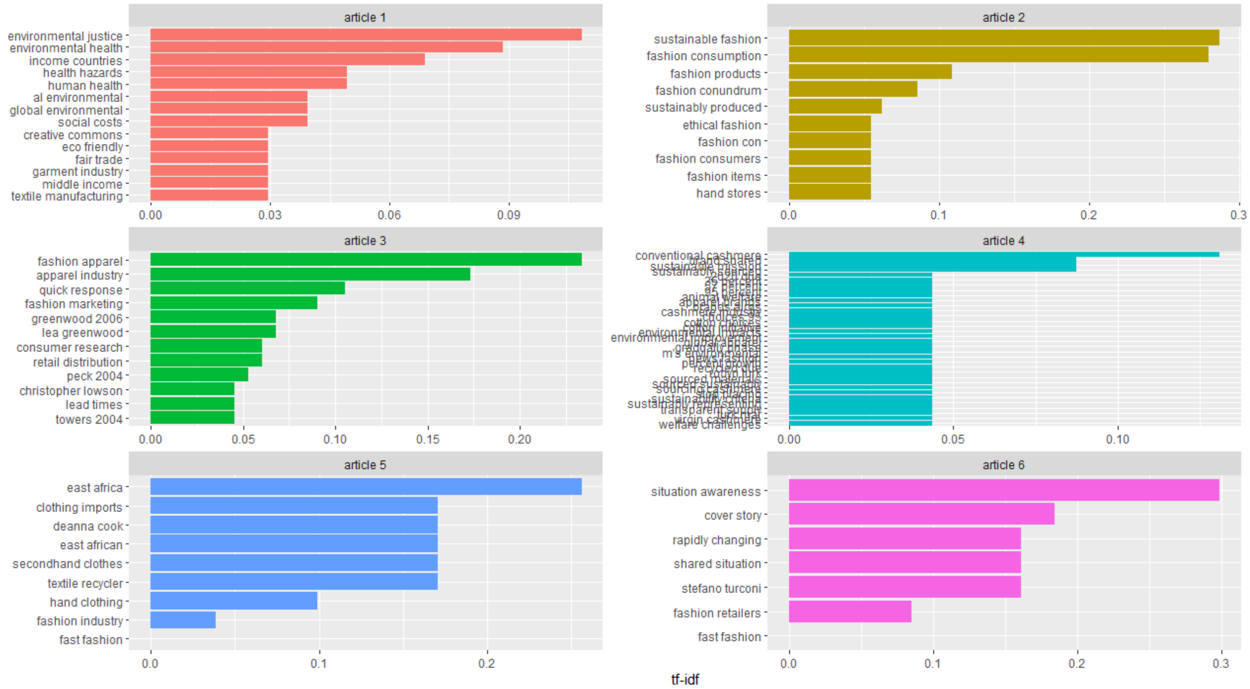
```
fffinal_bigrams <- left_join(fffinalbigram1, total_bigrams)
```

R Code Outcome:

- This line of code joins the total fffinalbigram1 with the inner join with total bigrams per article. Giving us a data_frame with the frequencies of each bigram with the total bigrams per article next to the other.

```
article_bigrams <- fffinal_bigrams %>%
  bind_tf_idf(bigram, article, n) %>%
  arrange(desc(tf_idf))
article_bigrams %>%
  arrange(desc(tf_idf)) %>%
  mutate(bigram=factor(bigram, levels=rev(unique(bigram)))) %>%
  group_by(article) %>%
  top_n(10) %>%
  ungroup %>%
  ggplot(aes(bigram, tf_idf, fill=article))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="tf-idf")+
  facet_wrap(~article, ncol=2, scales="free")+
  coord_flip()
```

R Code Output:



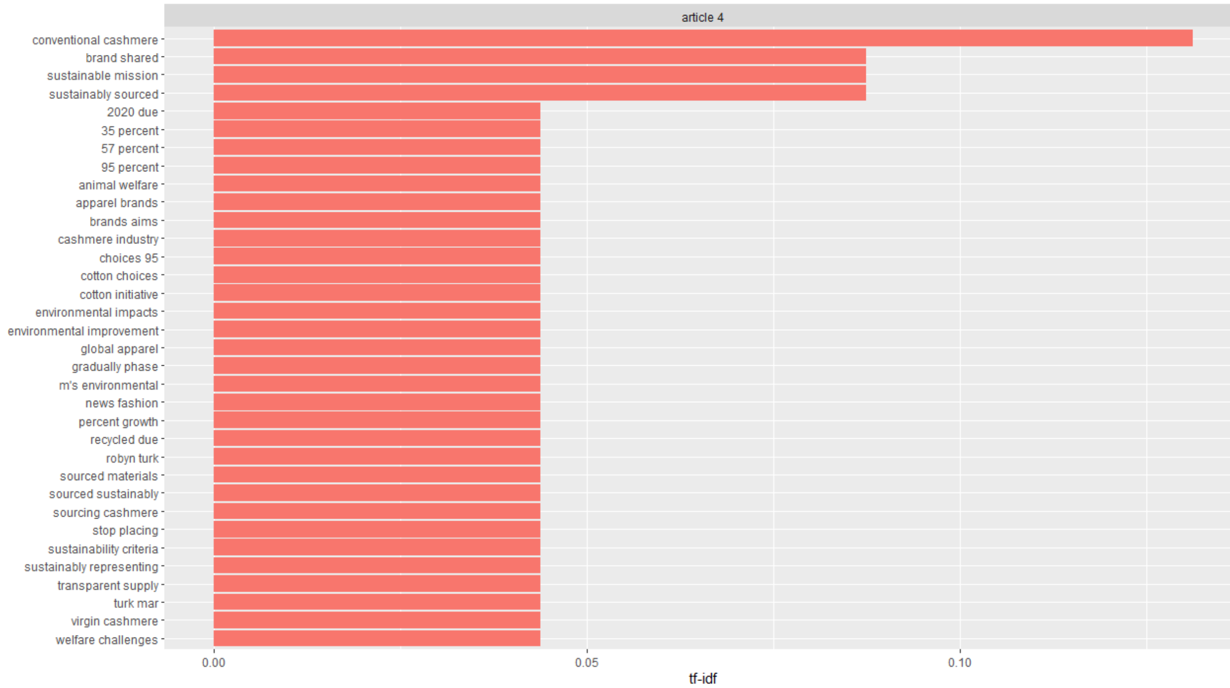
Graph A 45: Most unique and frequent bigrams per article – long articles

- These lines of code create a graph with the most frequent and unique words per article for the long articles (See Graph A22).

```
# The graph for the article 4th is not easy to understand. Filtering the data to have just this article

article_bigrams %>%

  arrange(desc(tf_idf)) %>%

  mutate(bigram=factor(bigram, levels=rev(unique(bigram)))) %>%

  group_by(article) %>%

  filter(article == "article 4") %>%

  top_n(10) %>%

  ungroup %>%

  ggplot(aes(bigram, tf_idf, fill=article))+

  geom_col(show.legend=FALSE)+

  labs(x=NULL, y="tf-idf")+

  facet_wrap(~article, ncol=2, scales="free")+
```

coord_flip()

R Code Output:



Graph A 46: Most Frequent and unique bigrams article 4

- These lines of code create a graph with the most frequent and unique words for the 4th article (See Graph A23).

Analyzing short articles

```
final_bigrams_separated_s <- short_articles %>%
  separate(bigram, c("word1", "word2"), sep = " ") %>%
  filter(!word1 %in% stopffwords$word) %>%
  filter(!word2 %in% stopffwords$word) %>%
  filter(n>3)
fffinalbigram_s <- final_bigrams_separated_s %>%
  unite ("bigram", 1:2 ,sep = " ", remove = FALSE, na.rm = FALSE)
```

R Code Outcome:

- These lines of code create the final data frame with the frequencies of the bigrams for only the short articles.

```
fffinalbigram2 <- cbind.data.frame(bigram = fffinalbigram_s$bigram, article =
fffinalbigram_s$article, n = fffinalbigram_s$n)
```

R Code Outcome:

- These lines of code create the final data frame with the frequencies of the bigrams for only the shorter articles.

```
total_bigrams_S <- fffinalbigram2 %>%
  group_by(article) %>%
  summarize(total=sum(n))
```

R Code Outcome:

- These lines of code create a table data frame with the total bigrams per short article.

```
fffinal_bigrams_s <- left_join(fffinalbigram2, total_bigrams_S)
```

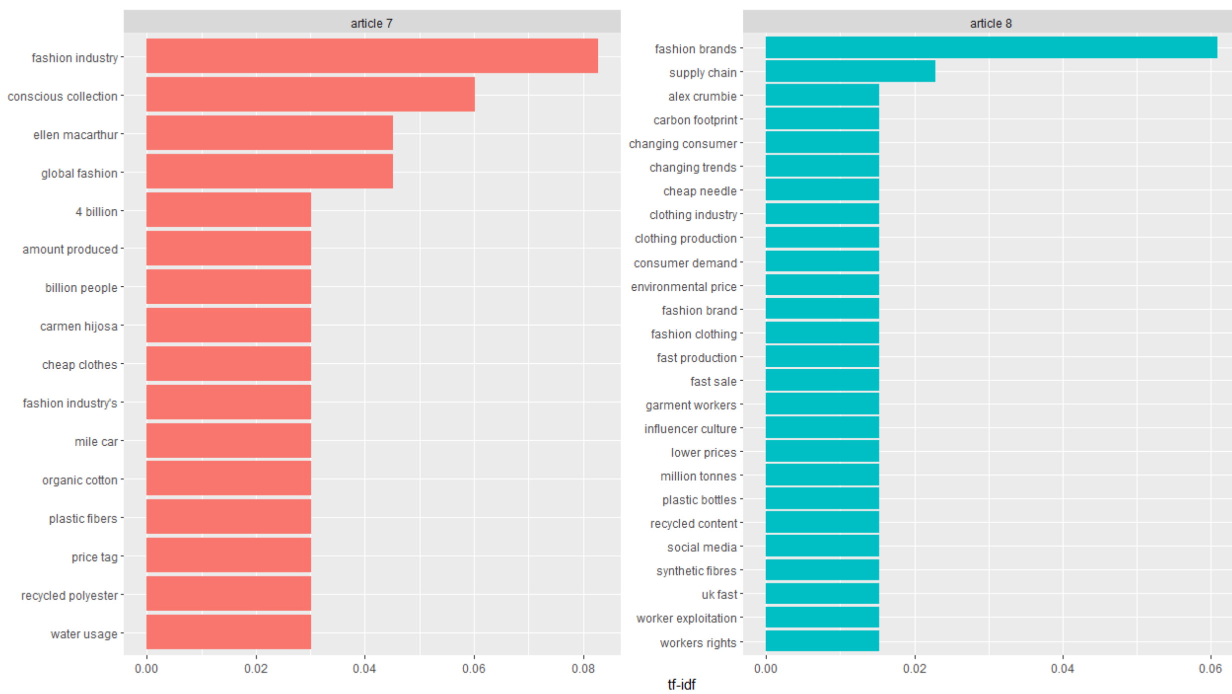
R Code Outcome:

- This line of code joins the total fffinalbigram2 with the inner join with total bigrams per short article. Giving us a data_frame with the frequencies of each bigram with the total bigrams per article next to the other.

```
article_bigrams_s <- fffinal_bigrams_s %>%
  bind_tf_idf(bigram, article, n) %>%
  arrange(desc(tf_idf))
article_bigrams_s %>%
  arrange(desc(tf_idf)) %>%
  mutate(bigram=factor(bigram, levels=rev(unique(bigram)))) %>%
```

```
group_by(article) %>%
top_n(10) %>%
ungroup %>%
ggplot(aes(bigram, tf_idf, fill=article))+
geom_col(show.legend=FALSE)+
labs(x=NULL, y="tf-idf")+
facet_wrap(~article, ncol=2, scales="free")+
coord_flip()
```

R Code Outcome:



Graph A 47: Most unique and frequent bigrams for the shorter articles

- These lines of code create a graph with the most frequent and unique words for the shorter articles (See Graph A24).

Conclusions from the text analysis:

The fast fashion industry represents a challenge specially for the environmental impact that it has.

Brands as Zara, HM and Benetton are part of this industry, and now they are focusing their efforts in trying to replace their materials to create more sustainable fashion clothing, even when the tendencies are rapidly changing.

The impact of the fashion industry is not only over the environment, but also over political matters as workers rights.

The consumers are responsible of the way there are consuming and have to be aware of the ethical fashion behind their clothe.

The general analysis of the sentiment of the articles related to fast fashion shows a positive sentiment. Nevertheless, after complementing the analysis with the semantic structure of the words, the “fast” labeled as a positive sentiment, actually represents a negative one. That changes the general sentiment, reinforcing the importance of look over the challenges that the fast fashion industry represents.