Project 1

Noreen Mayat

Question: How do romantic relationships, experiences and interactions impact happiness across gender groups?

In this project, I explore how romantic relationships, experiences and interactions impact happiness across gender groups of male and female, through computing the average word frequency for a bag of “romance”-related words for male and female, LDA topic modeling, and word2vec.

Make sure to install all packages below before running any code; I installed them by running:

```
# install.packages('package') for each package listed.
```

Set ur directory. I set mine running this in the console:

```
# setwd('Desktop/GitHub/ads-fall2023-project1-nm3224/doc/')
```

You may check your directory by running getwd() in the console.

Let’s take a look at some happy moments, and the accompanying data:

```
## [1] "I went on a successful date with someone I felt sympathy and connection with.
"
## [2] "I was happy when my son got 90% marks in his examination"
## [3] "I went to the gym this morning and did yoga."
## [4] "We had a serious talk with some friends of ours who have been flaky lately. T
hey understood and we had a good evening hanging out."
## [5] "I went with grandchildren to butterfly display at Crohn Conservatory\n"
## [6] "I meditated last night."
```

```
##   wid  age country gender marital parenthood
## 1   1 37.0     USA      m married          y
## 2   2 29.0     IND      m married          y
## 3   3   25     IND      m  single          n
## 4   4   32     USA      m married          y
## 5   5   29     USA      m married          y
## 6   6   35     IND      m married          y
```

Here is the frequency of all words in our documents: these are the top 10 with the highest frequencies.

```
##            word  freq
## friend   friend 10892
## day         day  9930
## time       time  9692
## family   family  4692
## watched watched  4385
## home       home  4211
## played   played  4058
## feel       feel  3946
## finally finally  3922
## found     found  3720
```

Now, let's compare the frequency of words, stratified by each gender group. Note that we have roughly the same number of observations for each group, with slightly more testimonies from males. There are some 42,019 observations labeled female, and 57,597 observations labeled female.

Here are the most frequent terms for females discussing happy moments:

```
##             word freq
## day           day 4460
## time         time 4337
## friend     friend 4269
## husband   husband 2572
## son           son 2188
## family     family 2109
## daughter daughter 2054
## home         home 1893
## watched   watched 1731
## feel         feel 1723
```

Here are the most frequent terms for males discussing happy moments:

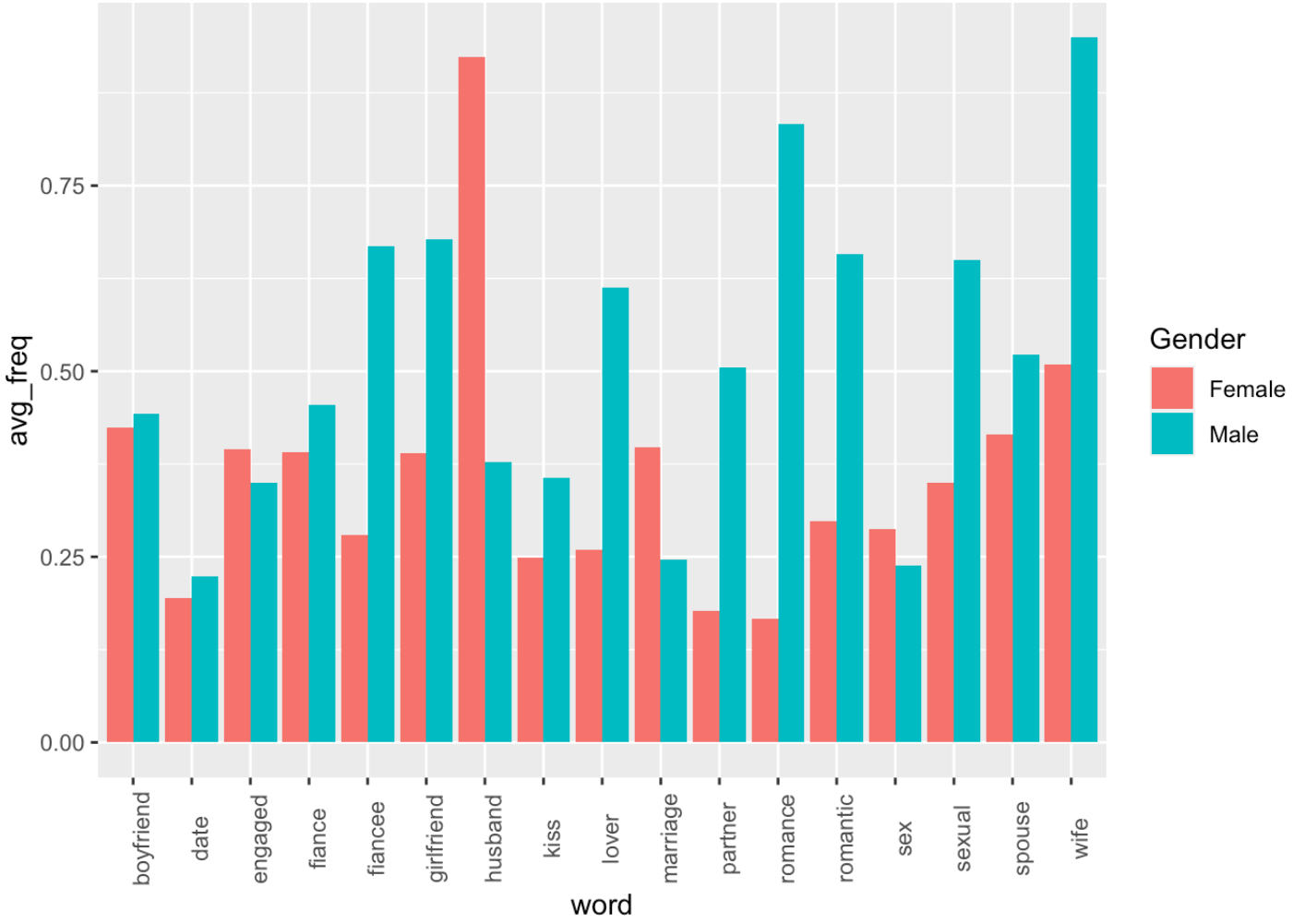
```
##            word freq
## friend   friend 6528
## day         day 5418
## time       time 5295
## played   played 2645
## wife       wife 2644
## watched watched 2611
## family   family 2556
## game       game 2395
## home       home 2291
## finally finally 2267
```

I don't see any major discrepancies off the bat; it may be more useful to now zone in on a specific bag of words I came up with related to romance, relationships, intimacy, and partnership.

```
love <- c('wife', 'husband', 'kiss', 'date', 'boyfriend', 'girlfriend', 'fiance', 'fi
ancee', 'engaged', 'sex', 'sexual', 'dating', 'romance', 'romantic', 'spouse', 'partn
er', 'lover', 'marriage')
```

Word Frequency

Let's take a look at the average word frequencies (frequency for gender m vs. f / overall frequency) for romance related words for both genders:




```
## Saving 7 x 5 in image
```

It seems men that the average frequency for romance-related words is higher in the male category than the female category. This means men discuss their romantic lives and relationships when talking about their happy moments more than women. This could imply that romantic relationships, interactions, and experiences are more impactful and related to male happiness than female happiness.

Leveraging word2vec for Partner-related words and happiness

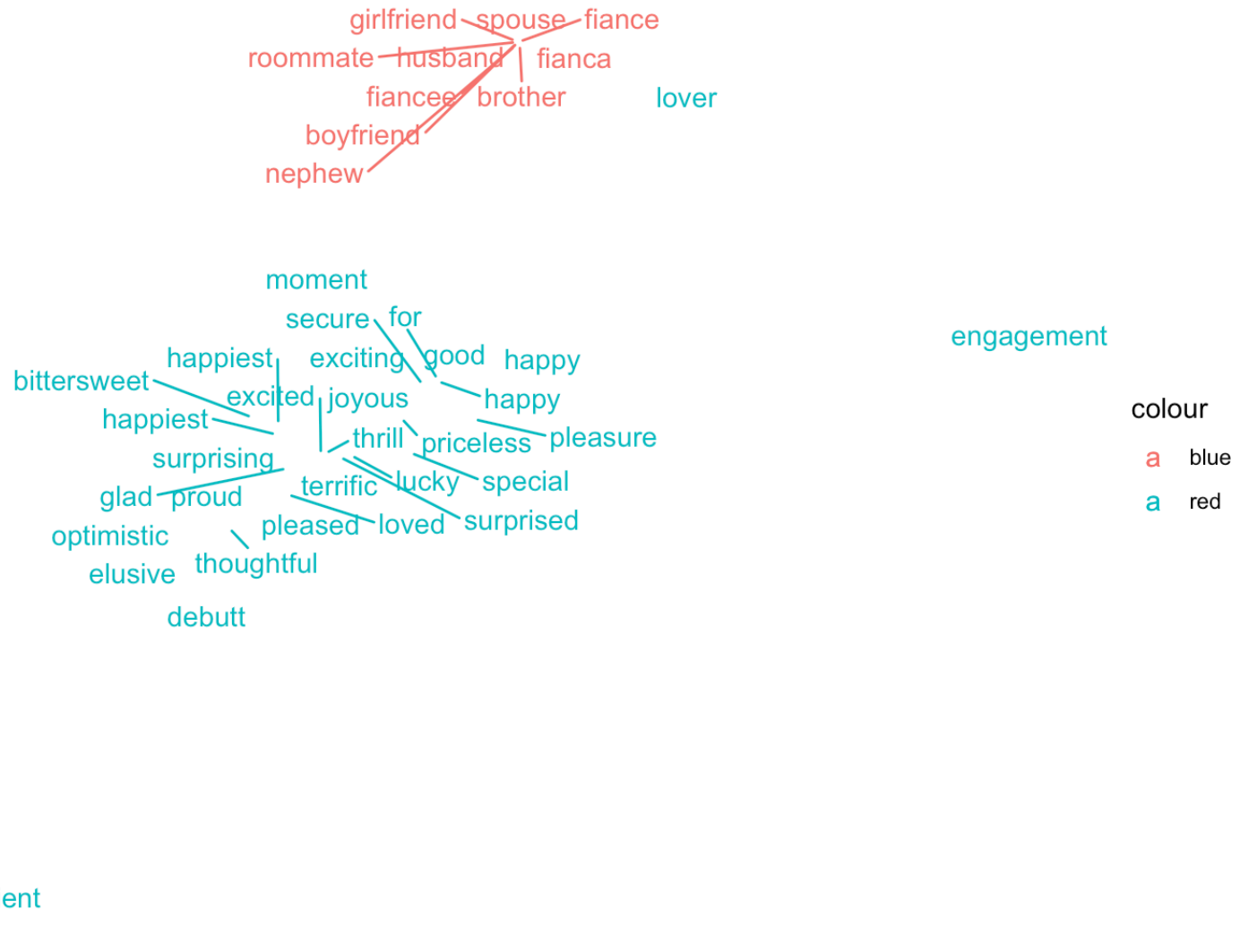
Now, using word2vec: first, we convert the data to a list of characters to input into our model.

```
## i went on a successful date with someone i felt sympathy and connection with.
```

Lemmatizing our text and using speech tag (verb, adverb, noun, adjective) will make representation easier (let's say we want to see all adjectives and nouns relative to the topic of animals). Note this cell takes a while to run; followed the same lemmatization process as the TA.

We now want to get the words most similar to partner in the embedding and we compare them to the words most similar to happy to observe a relationship.

Let's analyze partner and happy associated words for our corpus. ## Overall

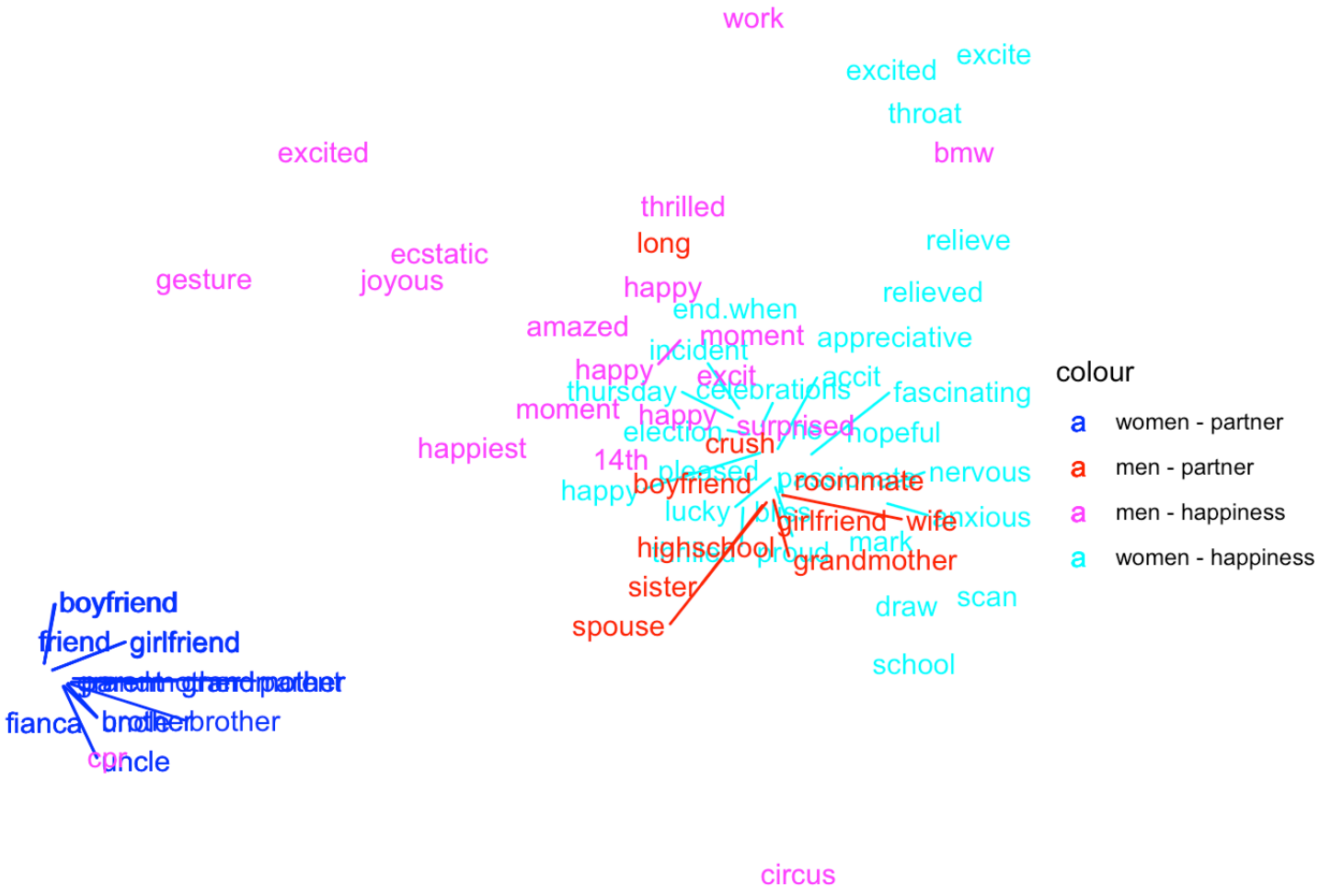


```
## $title
## [1] "100 most similar words to partner with word2vec - umap"
##
## attr(,"class")
## [1] "labels"
```

```
## Saving 7 x 5 in image
```

Per gender

Now, let's look at each per gender. Note this cell also takes a while to run; followed the same process as the TA. Apologies!



```
## $title
## [1] "Most similar words to partner with word2vec - umap"
##
## attr(,"class")
## [1] "labels"
```

```
## Saving 7 x 5 in image
```

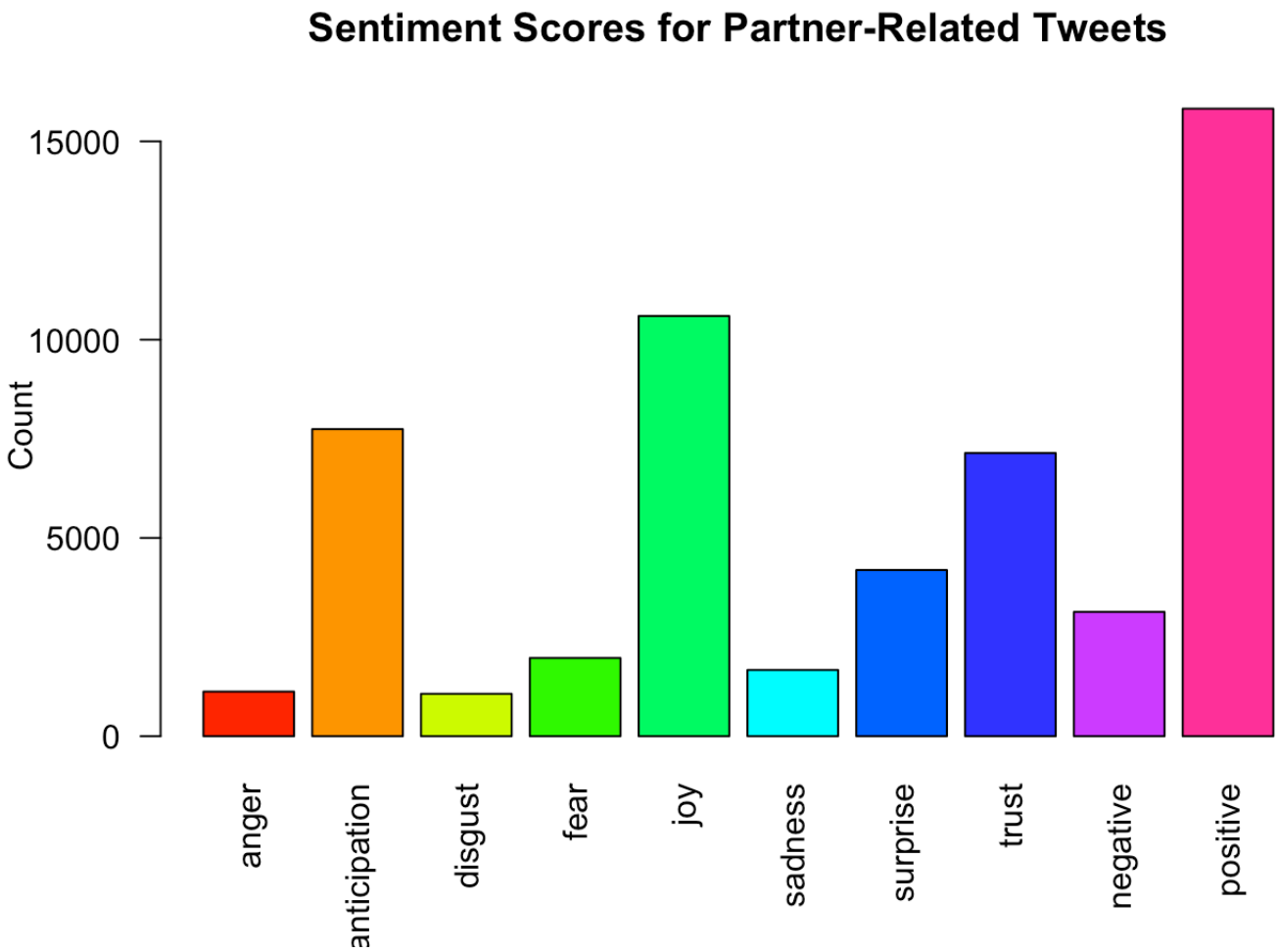
From this analysis we can see that male happiness is very closely associated with partnership/partner words, while female happiness is not.

Sentiment Analysis

Now, let's conduct a sentiment analysis on text documents referring to "love" and see how the sentiments vary between men and women.

```
##   anger anticipation disgust fear joy sadness surprise trust negative positive
## 1     0            0       0    0   1       0        0     0        0        2
## 2     0            1       0    0   1       0        0     0        0        1
## 3     0            0       0    0   1       0        0     0        0        1
## 4     0            0       0    0   0       0        0     0        0        1
## 5     0            3       0    0   2       0        1     1        0        2
## 6     1            0       0    1   0       0        1     0        1        0
```

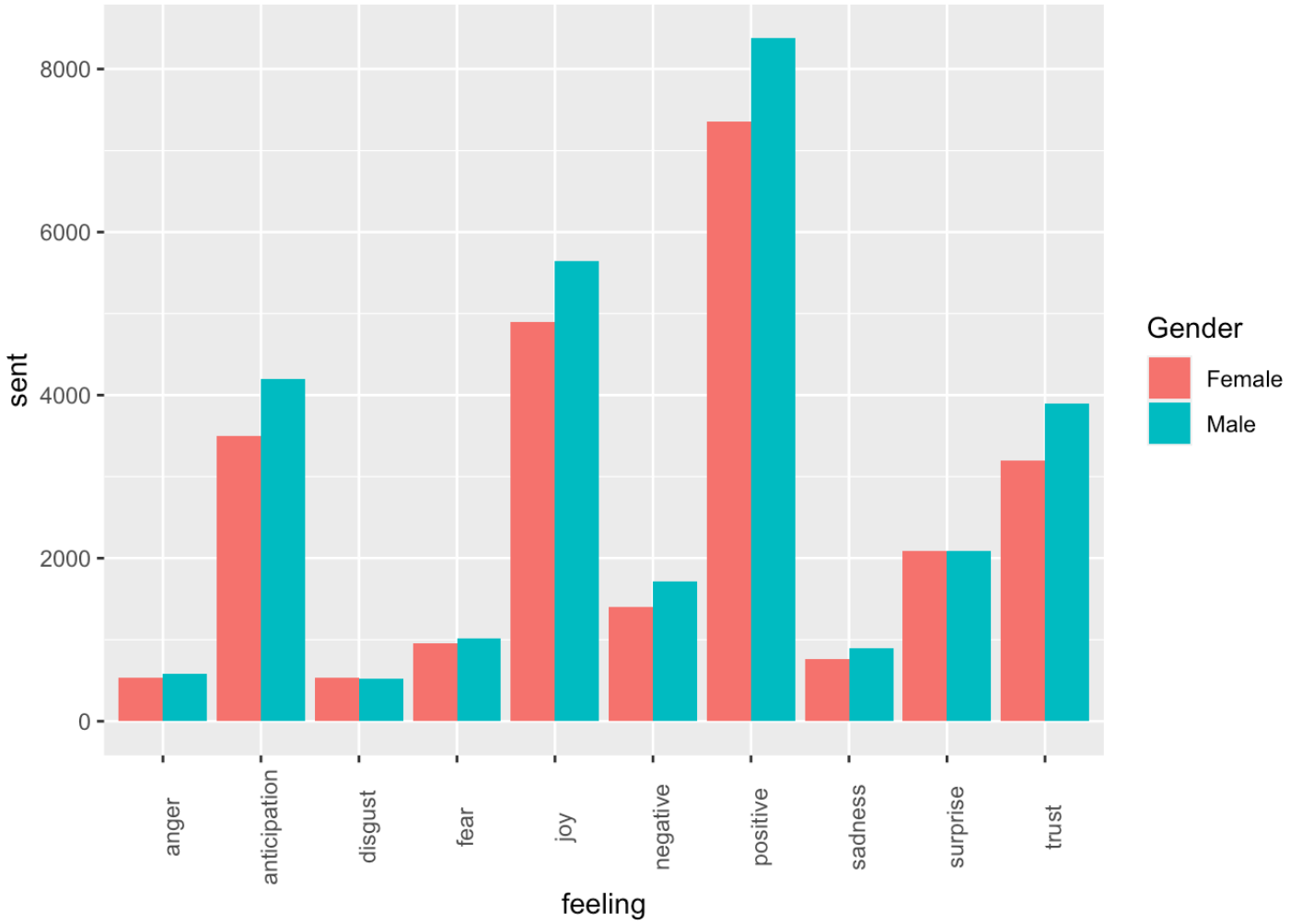
We have 10 sentiments to look at. Let's look at these sentiments for documents including the words listed in the "love" listed I provided. ### Overall




```
## quartz_off_screen
##                 2
```

Overall, sentiment related to partner text is positive and with joy. Let's test this analysis and compare for both genders. ## Per Gender

Let's plot both side by side:




```
## Saving 7 x 5 in image
```

Overall, documents referring to what I defined as "partner" or "relationship" words have higher sentiment values in categories positive, joy, trust, and more for males than for females.