Comparison of Happy Moments for Different Kinds of People and Under Different Time Horizons

Yinpei Wang

```r
packages.used=c("tidyverse", "tidytext",
                "DT", "scales", "wordcloud2",
                "gridExtra", "textdata")

# check packages that need to be installed.
packages.needed=setdiff(packages.used,
                        intersect(installed.packages()[,1],
                                  packages.used))
# install additional packages
if(length(packages.needed)>0){
  install.packages(packages.needed, dependencies = TRUE)
}

# load packages
library(tidyverse)
library(tidytext)
library(DT)
library(scales)
library(wordcloud2)
library(gridExtra)
library(textdata)
```

```r
# read data files
demo_data <- read_csv("../data/demographic.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for deta
ils,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```r
hm_data <- read_csv("../output/processed_data.csv")
hm_data_24h <- read_csv("../output/processed_24h.csv")
hm_data_3m <- read_csv("../output/processed_3m.csv")
hm_data_female <- read_csv("../output/processed_female.csv")
hm_data_male <- read_csv("../output/processed_male.csv")
hm_data_married <- read_csv("../output/processed_married.csv")
hm_data_unmarried <- read_csv("../output/processed_unmarried.csv")
hm_data_parent <- read_csv("../output/processed_parent.csv")
hm_data_notparent <- read_csv("../output/processed_notparent.csv")
```

Introduction

What makes people happy? This age-old question has perplexed philosophers, psychologists, and average Joes alike. While the answer is no doubt complex, mining a dataset of 100,000 crowdsourced happy moments provides tantalizing clues into the roots of human joy.

In this analysis, I explore patterns and trends in HappyDB, searching for insights among the stars of positivity. Using tools like word frequencies and sentiment analysis, I uncover variations in happy moments between different groups of people and across timeframes. Specifically, I compare the joyful moments of females versus males, married versus unmarried individuals, parents versus the childless, and experiences over the past 24 hours versus the past 3 months.

The findings reveal fascinating contrasts in the sources of happiness across demographics and periods of time. Married folk derive bliss from family time, while the unmarried are more likely to experience childlike glee. Women tend to find joy in relationships and pleasures, while male happiness skews towards self-satisfaction and leisure activities. Parents bask in their children's glow, whereas the childless are more self-focused. And recent joy derives from sharable moments, while satisfaction from the past quarter year comes from achievements.

What makes the human heart swell up with joy? The answer is complicated, variable, and deeply personal. Join me as we delve into the HappyDB data for telling insights into this universally sought-after state of being: happiness.

Word Frequency Analysis

Word clouds provide an intuitive yet informative way to visualize the results of word frequency analysis. By displaying the most common words in a textual dataset in various font sizes, word clouds give at-a-glance insights into the prevalent themes and topics. Thus, I use word clouds to make comparison for word frequencies.

Comparison between different time horizons

```
word_count_24h <- hm_data_24h %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE)
word_count_3m <- hm_data_3m %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE)
wc_24h <- wordcloud2(word_count_24h, size = 0.5)
wc_3m <- wordcloud2(word_count_24h, size = 0.5)
wc_24h
```

wc_3m

The word clouds of different time horizon show little difference.

Comparison between females and males

```
word_count_female <- hm_data_female %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE)
word_count_male <- hm_data_male %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE)
wc_female <- wordcloud2(word_count_female, size = 0.5)
wc_male <- wordcloud2(word_count_male, size = 0.5)
wc_female
```
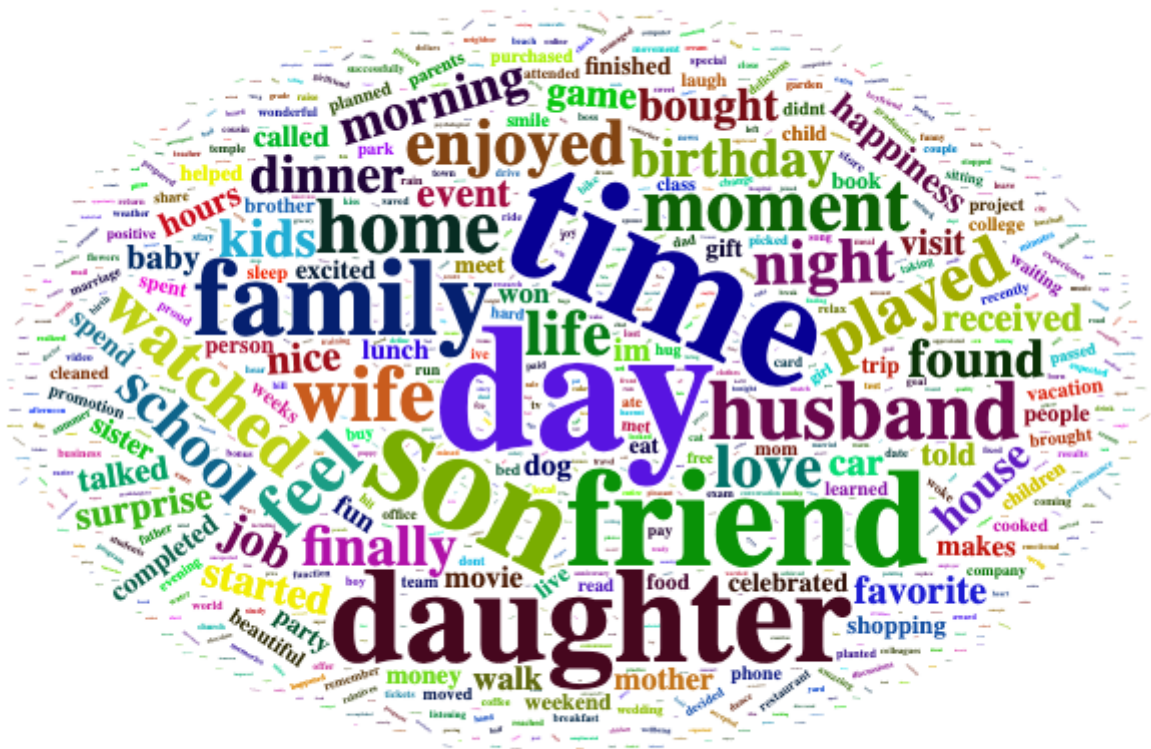


```
wc_male
```

Compare the word clouds for females and males. We can see that both females and males get most happiness ses from friends. However, females get more happiness es from family as there are high frequency words like husband, family, home, daughter, son, etc. While the frequencies of these words for males are not that high. Males are more satisfying with leisure activities as the words like played, game, feel, etc have higher frequencies.

Comparison between married and unmarried

```
word_count_married <- hm_data_married %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE)
word_count_unmarried <- hm_data_unmarried %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE)
wc_married <- wordcloud2(word_count_married, size = 0.5)
wc_unmarried <- wordcloud2(word_count_unmarried, size = 0.5)
wc_married
```

wc_unmarried

For unmarried people, their main happiness are from friends. However, for married ones, their main focus moved from friends to families since the frequencies of family-related words exceeds that of friend.

Comparison between parents and non-parents

```
word_count_parent <- hm_data_parent %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE)
word_count_notparent <- hm_data_notparent %>%
  unnest_tokens(word, text) %>%
  count(word, sort = TRUE)
wc_parent <- wordcloud2(word_count_parent, size = 0.5)
wc_notparent <- wordcloud2(word_count_notparent, size = 0.5)
wc_parent
```



```
wc_notparent
```

Check the word clouds, we see the same pattern as the above section. That is, the childless focus mostly on friends and self-satisfaction while the parents focus mainly on families.

Sentiment Analysis

Sentiment analysis proves to be an invaluable tool in unraveling the intricacies of happy moments among different demographics and time horizons. By employing natural language processing techniques, sentiment analysis allows us to not only quantify the positive emotions encapsulated in the HappyDB corpus but also discern subtle variations in the emotional landscape. This analytical approach enables us to uncover underlying patterns in the expression of happiness. In this section, I apply sentiment analysis to compare most frequent positive/negative words as well as the sentiment scores among different demographic categories.

Most frequent positive/negative words for different kinds of people

Examining the most frequent positive and negative words associated with various demographics adds a granular layer to our analysis of happy moments. Identifying these words provides a nuanced understanding of the unique elements that contribute to happiness for different groups. For example, pinpointing the most common positive words for females versus males, or for married versus unmarried individuals, can unveil specific themes that resonate more strongly within each demographic.

```
# load sentiment words
bing <- get_sentiments('bing')
# source function to plot results
source("../lib/plot_sentiment.R")
```

Comparison between different time horizons

```
# 24h
sent_sum_24h <- word_count_24h %>% inner_join(bing, by = 'word')
plot_sent_24h <- plot_sentiment(sent_sum_24h,
                                "Most Frequent Words of Happy Moments within 24h")
# 3m
sent_sum_3m <- word_count_3m %>% inner_join(bing, by = 'word')
plot_sent_3m <- plot_sentiment(sent_sum_3m,
                               "Most Frequent Words of Happy Moments within 3 month
s")
# show plot
grid.arrange(plot_sent_24h, plot_sent_3m)
```
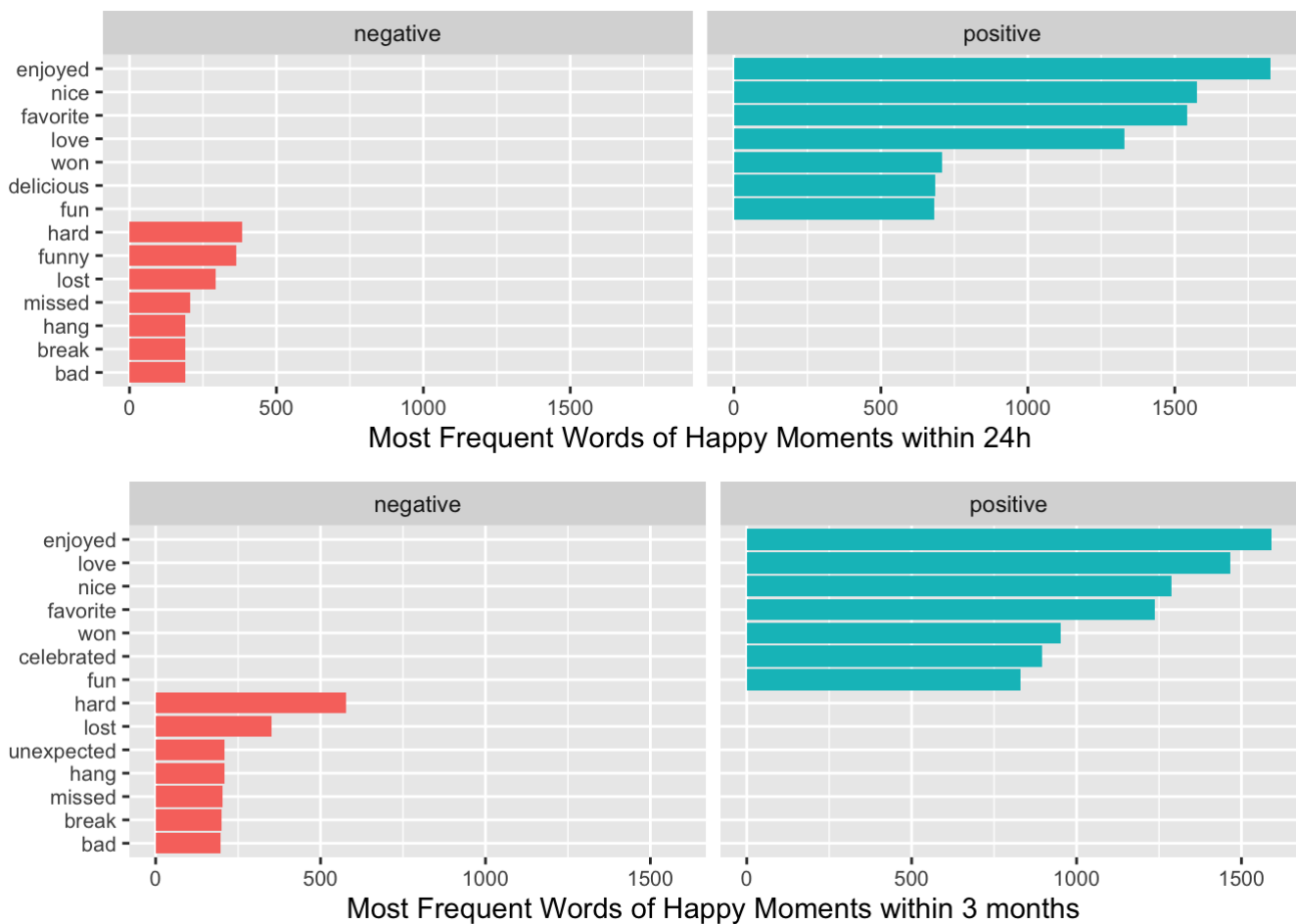
Compare the most frequent positive/negative words among different time horizons, we can see that love becomes more important as a source of happiness when the time horizon is longer. Also, 'hard' becomes more frequent during a longer time horizon. This may imply that during longer time horizon, people are more satisfied through overcoming hardness.

Comparison between females and males

```
# female
sent_sum_female <- word_count_female %>% inner_join(bing, by = 'word')
plot_sent_female <- plot_sentiment(sent_sum_female,
                                   "Most Frequent Words of Happy Moments for female
s")
# male
sent_sum_male <- word_count_male %>% inner_join(bing, by = 'word')
plot_sent_male <- plot_sentiment(sent_sum_male,
                                 "Most Frequent Words of Happy Moments for males")
# show plot
grid.arrange(plot_sent_female, plot_sent_male)
```
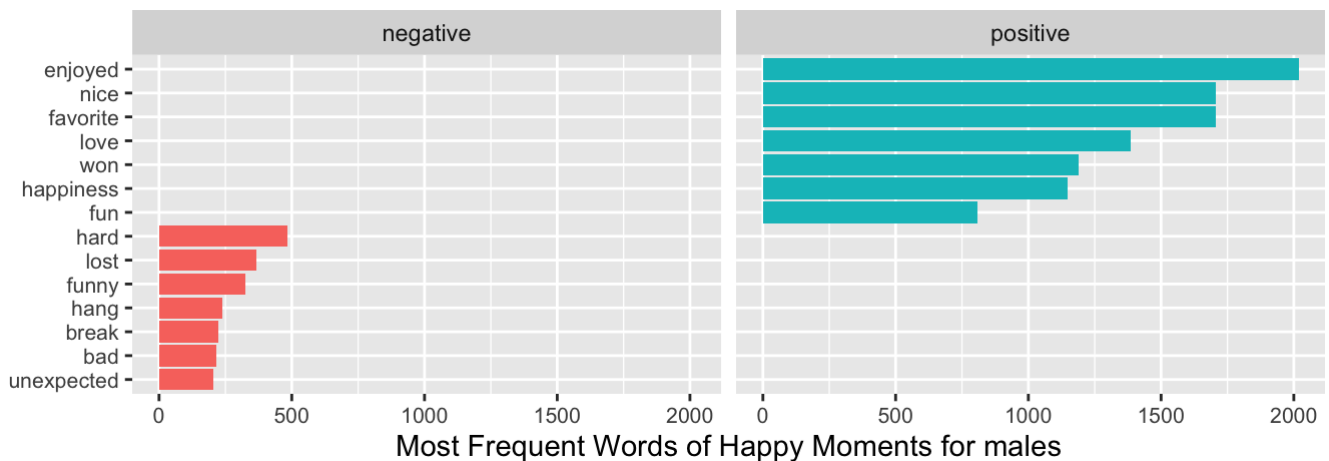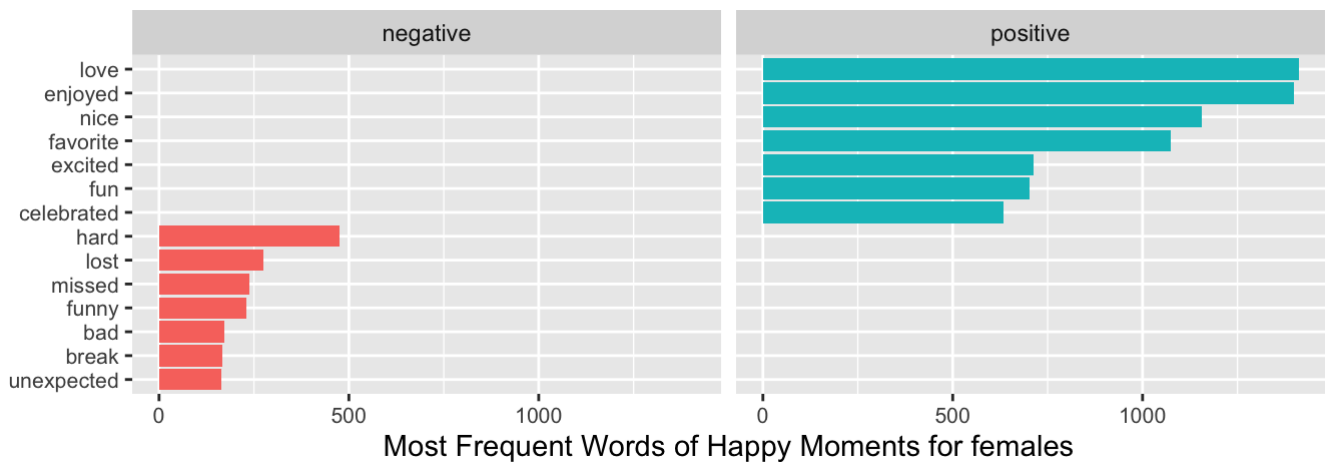

As the plot shows, love is a more important source of happinesses for females than for males. Women tend to find joy in relationships and pleasures, while male happiness skews towards self-satisfy.

Comparison between married and unmarried

```
# married
sent_sum_married <- word_count_married %>% inner_join(bing, by = 'word')
plot_sent_married <- plot_sentiment(sent_sum_married,
                                    "Most Frequent Words of Happy Moments for married
people")
# unmarried
sent_sum_unmarried <- word_count_unmarried %>% inner_join(bing, by = 'word')
plot_sent_unmarried <- plot_sentiment(sent_sum_unmarried,
                                      "Most Frequent Words of Happy Moments for unmar
ried people")
# show plot
grid.arrange(plot_sent_married, plot_sent_unmarried)
```
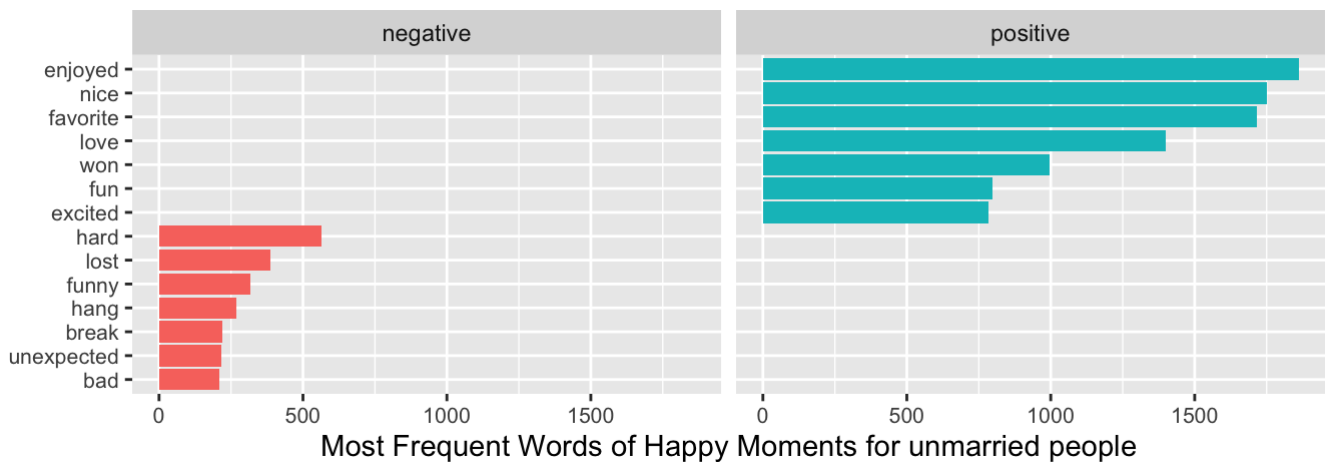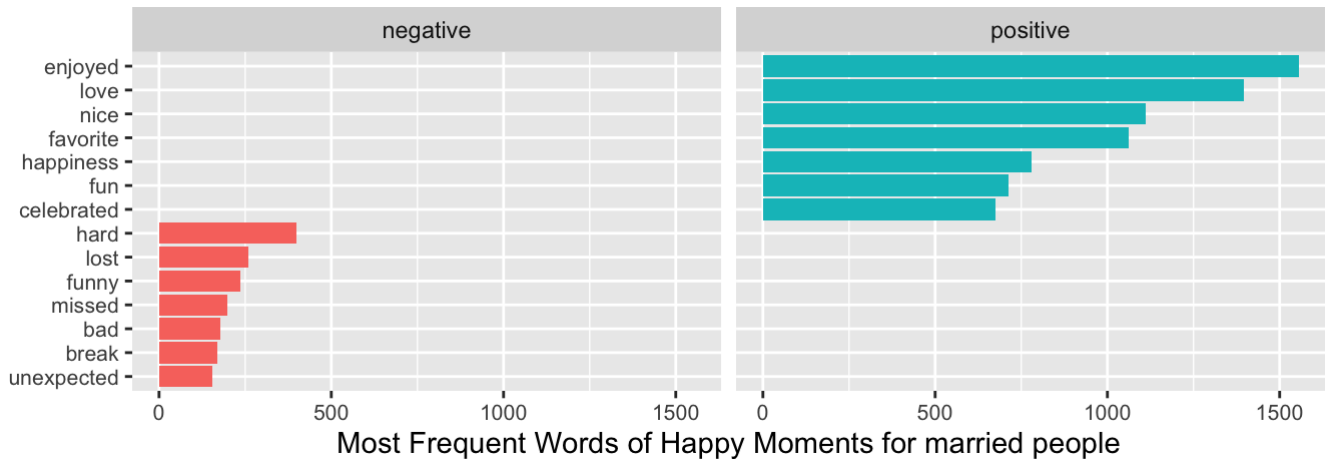
'love' is more important for married people than for unmarried people. This confirmed my result of word frequency analysis again: married ones get more happiness from families while unmarried ones get more happiness hanging out with friends.

Comparison between parents and others

```
# parent
sent_sum_parent <- word_count_parent %>% inner_join(bing, by = 'word')
plot_sent_parent <- plot_sentiment(sent_sum_parent,
                                    "Most Frequent Words of Happy Moments for parent
s")
# notparent
sent_sum_notparent <- word_count_notparent %>% inner_join(bing, by = 'word')
plot_sent_notparent <- plot_sentiment(sent_sum_notparent,
                                      "Most Frequent Words of Happy Moments for non-p
arents")
# show plot
grid.arrange(plot_sent_parent, plot_sent_notparent)
```
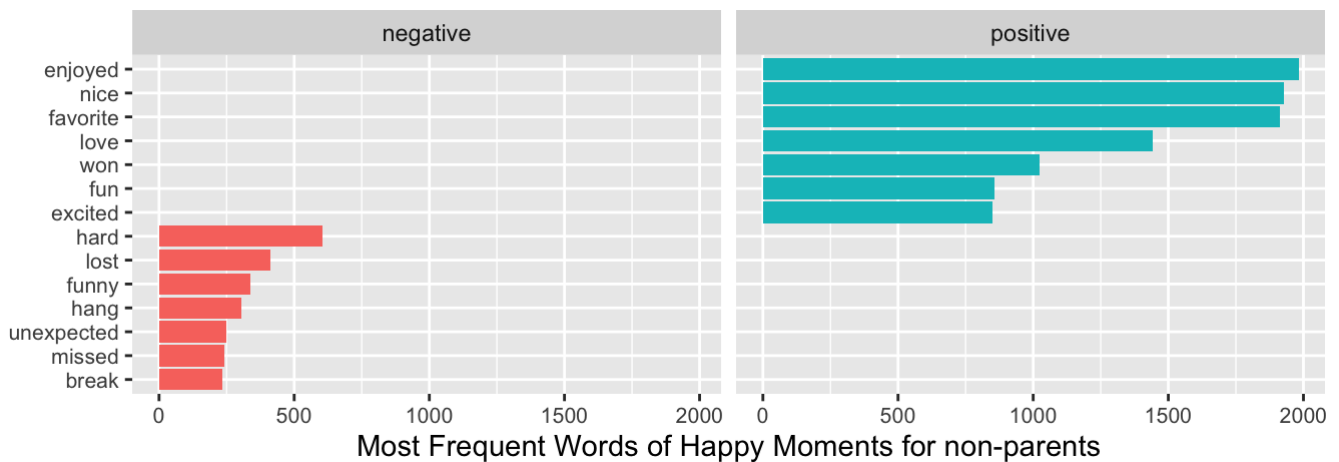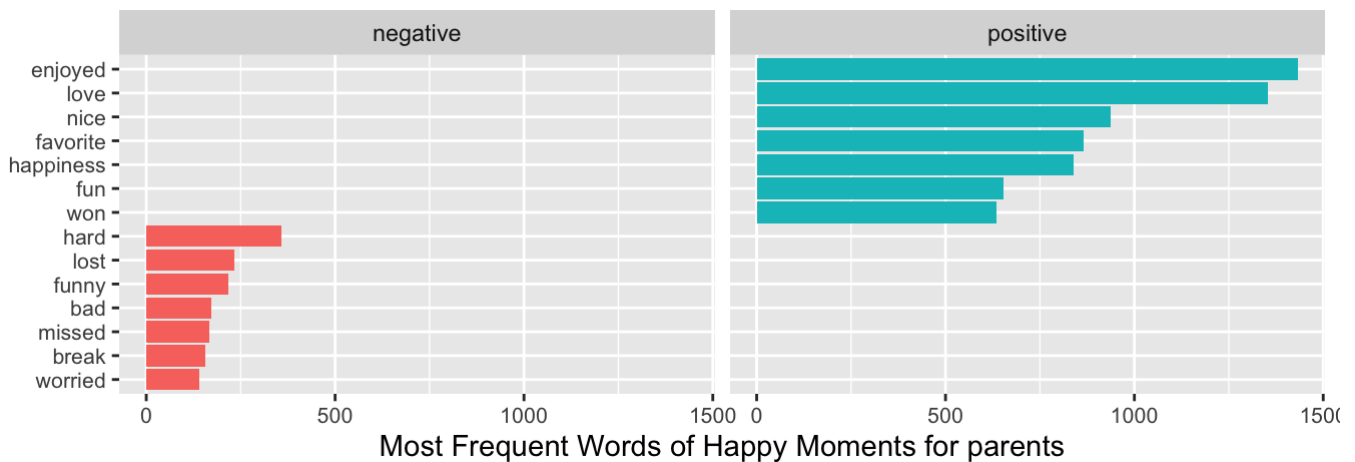

Chech the plot above, we can get the same conclusion as before: parents get more happiness from families while the childless get more happiness hanging out with friends.

# Are certain kind of people more happier than the counterparts?

Another advantage of sentiment analysis is that we can quantify the level of happiness. In this section, I use the lexicon 'afinn' to assign sentiment score for each word and the level of happiness for each response are quantified through average sentiment score to account for different number of word per response. Then, the level of happiness for people of different demographic status are compared using average sentiment scores.

```
# load lexicon "afinn"
afinn <- get_sentiments("afinn")
# merge a new data set containing sentiment scores and demographic status
avg_sent_scores <- hm_data %>% unnest_tokens(word, text) %>%
    inner_join(afinn, by = 'word') %>%
    group_by(wid) %>%
    summarise(avg_sent_score = mean(value)) %>%
    inner_join(demo_data, by = 'wid')
```

```
# load the function which computes the statistics in boxplot
source("../lib/get_box_stats.R")
```

Comparison between females and males

```
# boxplot of average sentiment scores for females and males
avg_sent_scores %>%
    ggplot(aes(x = gender, y = avg_sent_score, color = gender)) +
    geom_boxplot() +
    stat_summary(fun.data = get_box_stats, geom = "text", hjust = 0.5, vjust = 0.9) +
    ggtitle("Boxplot of Average Sentiment Scores for Females and Males") +
    ylab("Average Sentiment Score") +
    theme_bw()
```
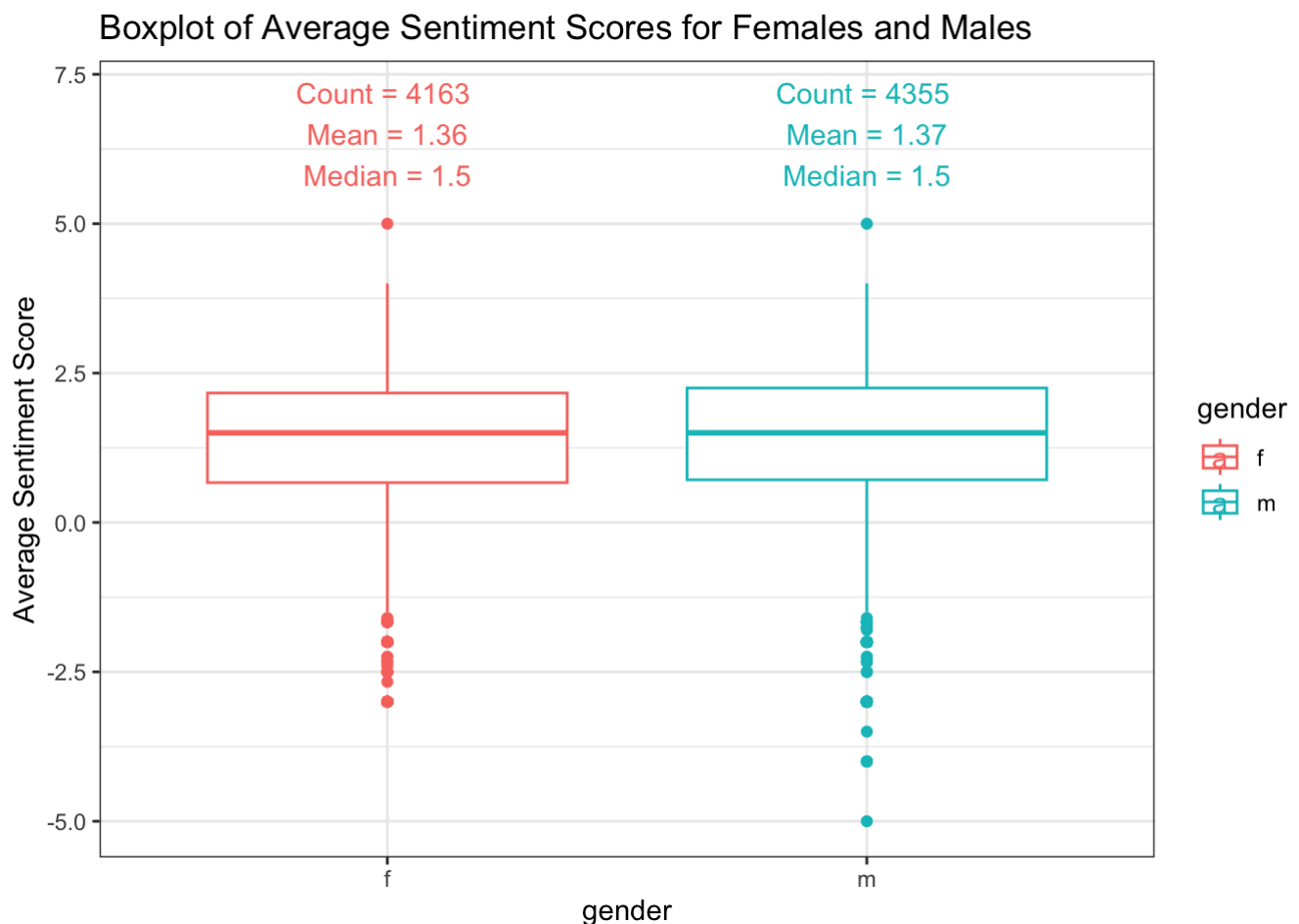

The plot above shows that on average, there is little difference of sentiment score between females and males. However, when we focus on extreme values, we can see that males are more likely to get extremely low sentiment scores.

Comparison between married and unmarried

```
# boxplot of average sentiment scores for married and unmarried
avg_sent_scores %>%
  ggplot(aes(x = marital, y = avg_sent_score, color = marital)) +
  geom_boxplot() +
  stat_summary(fun.data = get_box_stats, geom = "text", hjust = 0.5, vjust = 0.9) +
  ggtitle("Boxplot of Average Sentiment Scores for Married and Unmarried People") +
  ylab("Average Sentiment Score") +
  theme_bw()
```

## Boxplot of Average Sentiment Scores for Married and Unmarried People
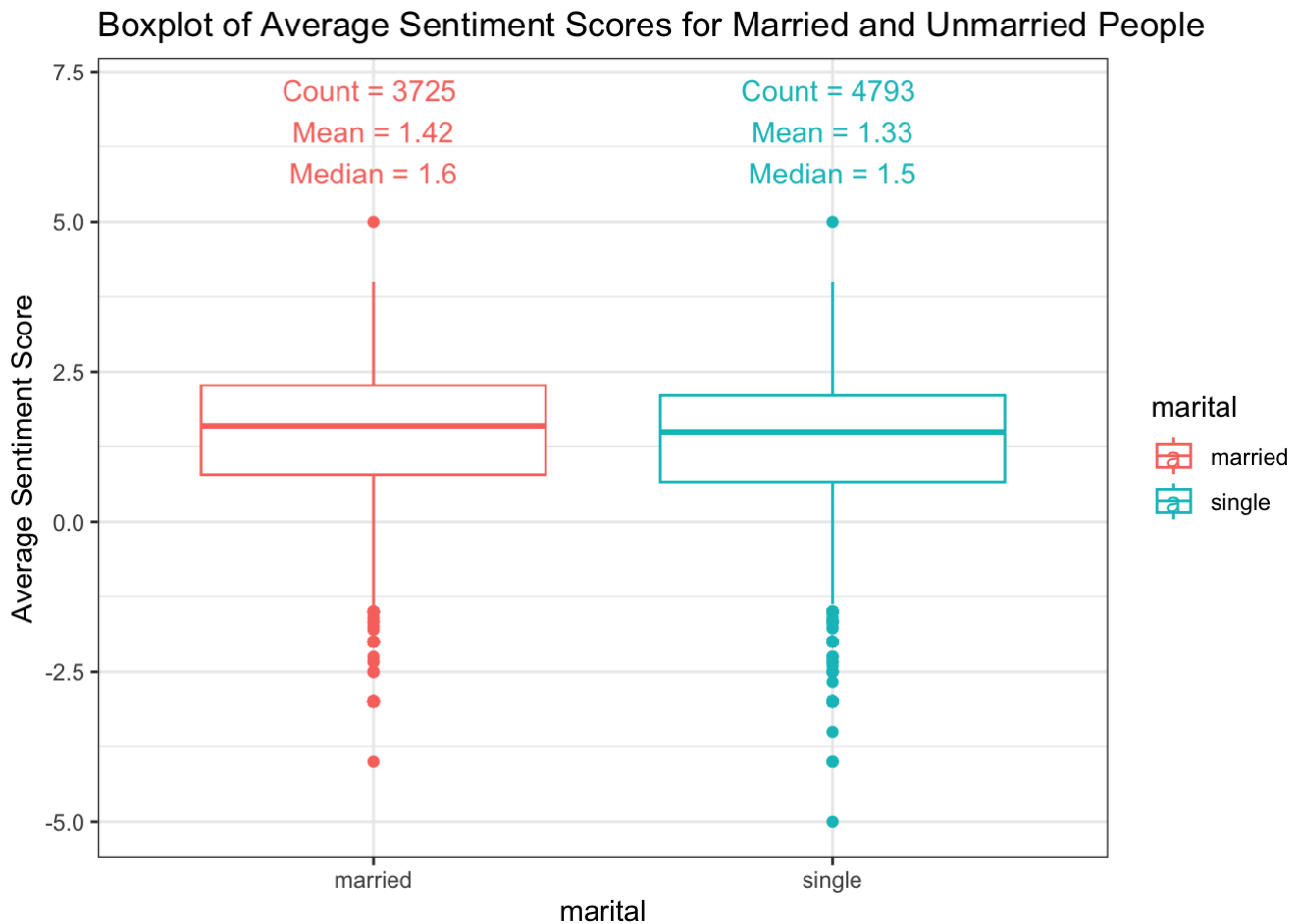
The plot above shows that the happiness level for married people are slightly higher than that of unmarried people. Also, unmarried people are more likely to have extremely low sentiment scores.

Comparison between parents and non-parents

```
# boxplot of average sentiment scores for parents and nonparents
avg_sent_scores %>%
  ggplot(aes(x = parenthood, y = avg_sent_score, color = parenthood)) +
  geom_boxplot() +
  stat_summary(fun.data = get_box_stats, geom = "text", hjust = 0.5, vjust = 0.9) +
  ggtitle("Boxplot of Average Sentiment Scores for Parents and Non-parents") +
  ylab("Average Sentiment Score") +
  theme_bw()
```

Boxplot of Average Sentiment Scores for Parents and Non-parents

Count = 5274
Mean = 1.33
Median = 1.5

Count = 3244
Mean = 1.43
Median = 1.6
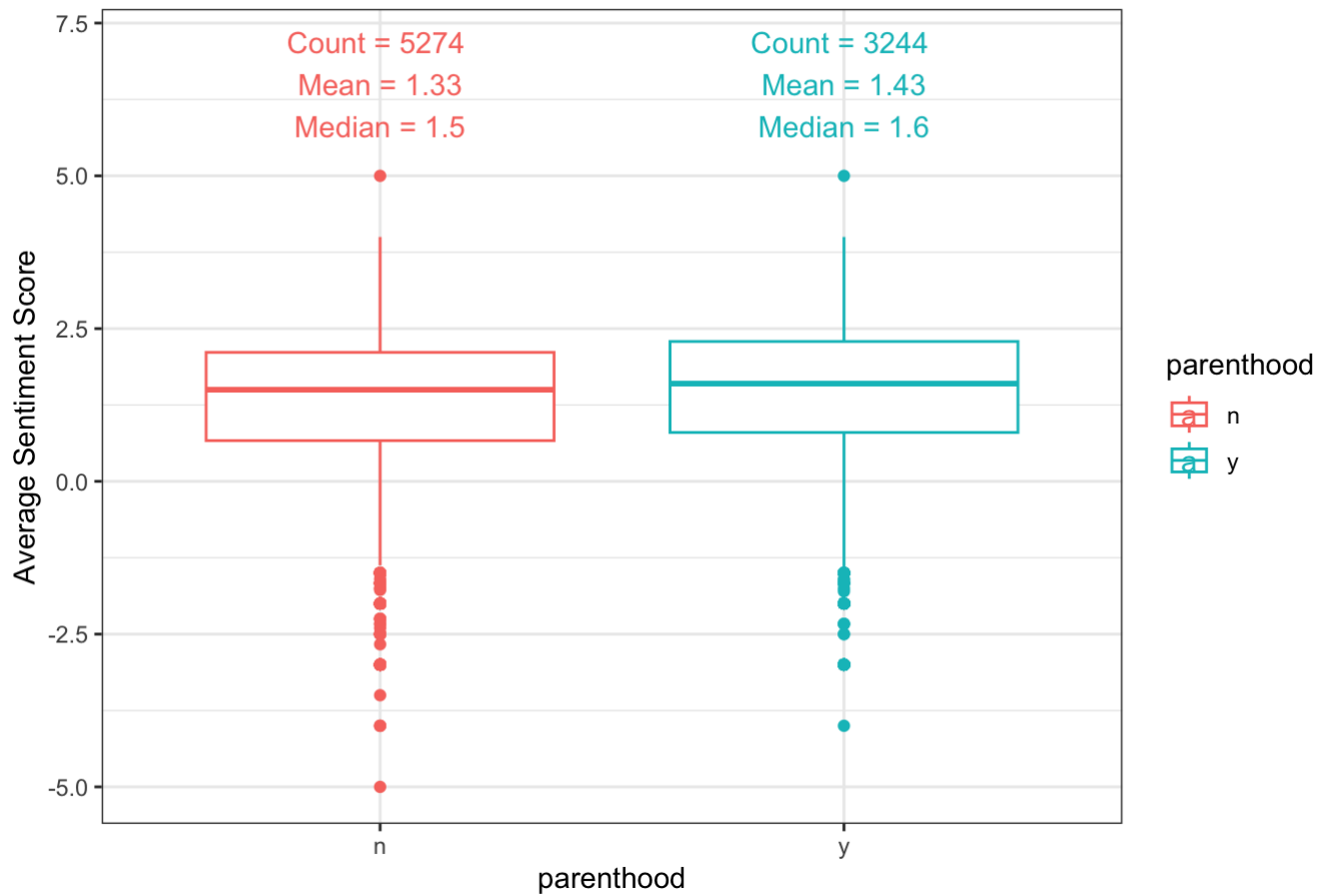
Compare the average sentiment scores between parents and non-parents, we can see that on average, parents have higher average sentiment scores than non-parents. This implies that children brings more happiness to parents and hence parents have higher level of happiness than the childless.