

5243_proj1_sj2854

Shichao Jia UNL:sj2854

2018年9月19日

what's make your girlfriend/wife happy

do you feel some time you just don't get it why your wife/ girlfriend want? or why suddenly they tell you you don't love them enough? lets use data science to solve this mystery!

Step 0 - Load all the required libraries

```
library(tidyverse)
library(tidytext)
library(DT)
library(scales)
library(wordcloud2)
library(gridExtra)
library(ngram)
library(shiny)
```

Step 1 - Load the processed text data along with demographic information on contributors

We use the processed data for our analysis and combine it with the demographic information available.

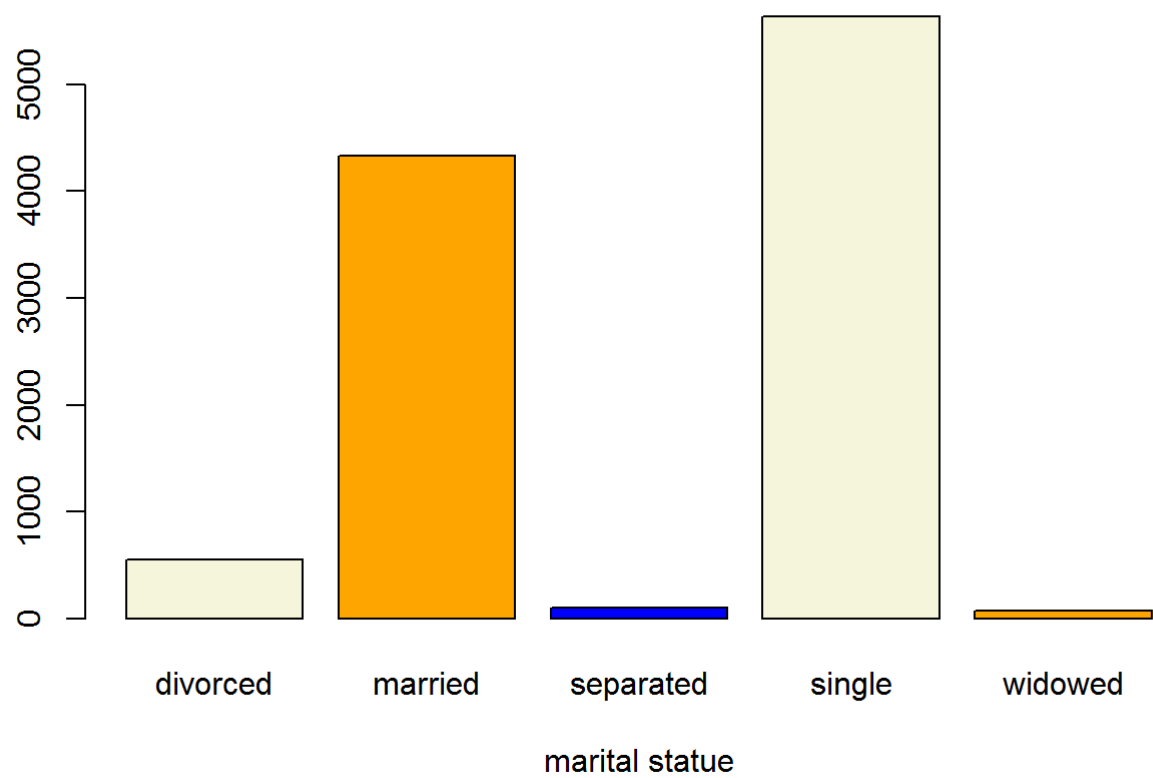
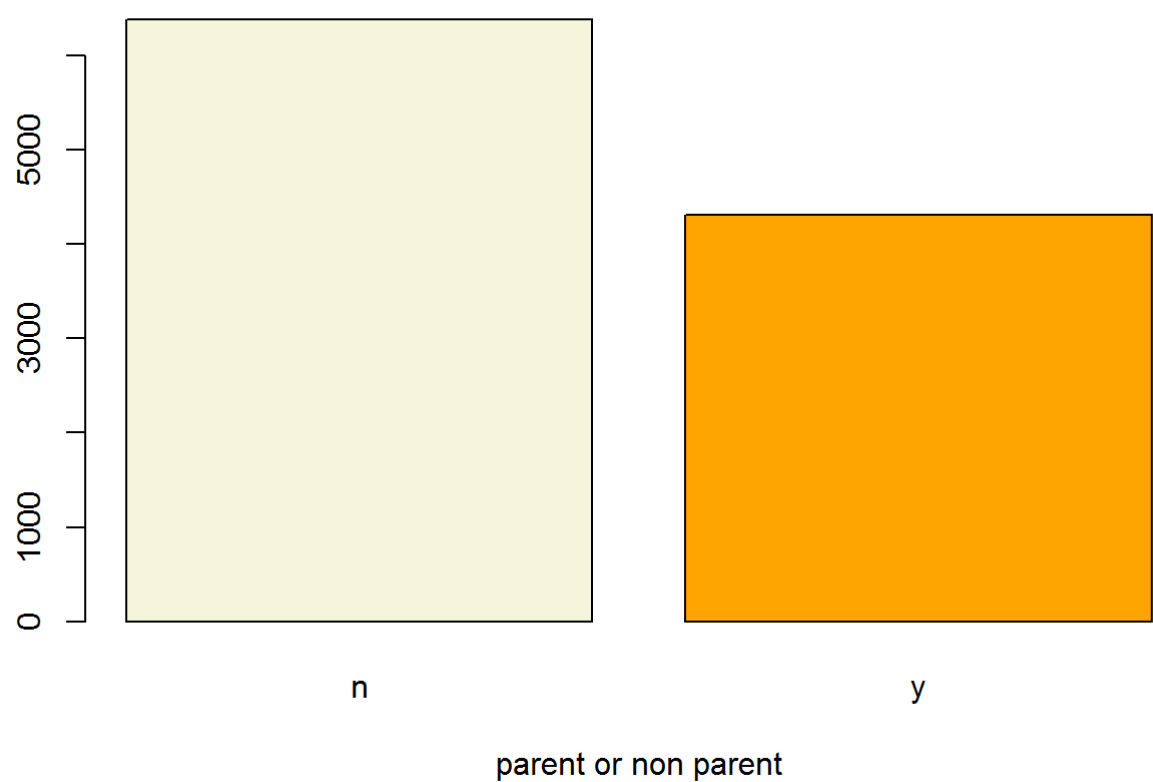
```
## # A tibble: 6 x 6
##   wid age  country gender marital parenthood
##   <int> <chr> <chr>   <chr>  <chr>   <chr>
## 1  2964 57.0  USA     f      <NA>    n
## 2  3457 52.0  USA     <NA>   married y
## 3  3655 <NA>   USA     m      single n
## 4  3715 23    <NA>    f      single n
## 5  3776 37.0  USA     f      <NA>    n
## 6  3897 <NA>   USA     f      <NA>    <NA>
```

```
## # A tibble: 0 x 6
## # ... with 6 variables: wid <int>, age <chr>, country <chr>, gender <chr>,
## #   marital <chr>, parenthood <chr>
```

Combine both the data sets and keep the required columns for analysis

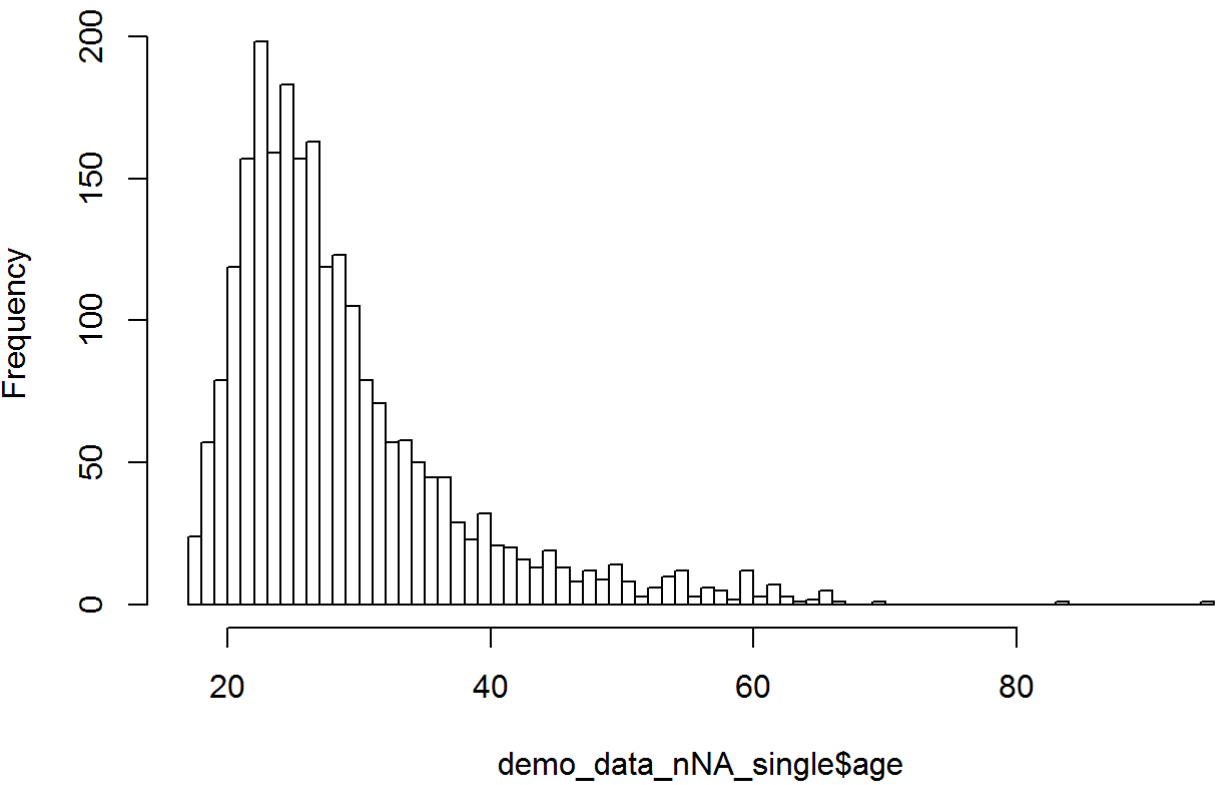
We select a subset of the data that satisfies specific row conditions.

we analyst the data that what is the certain age we want to persue



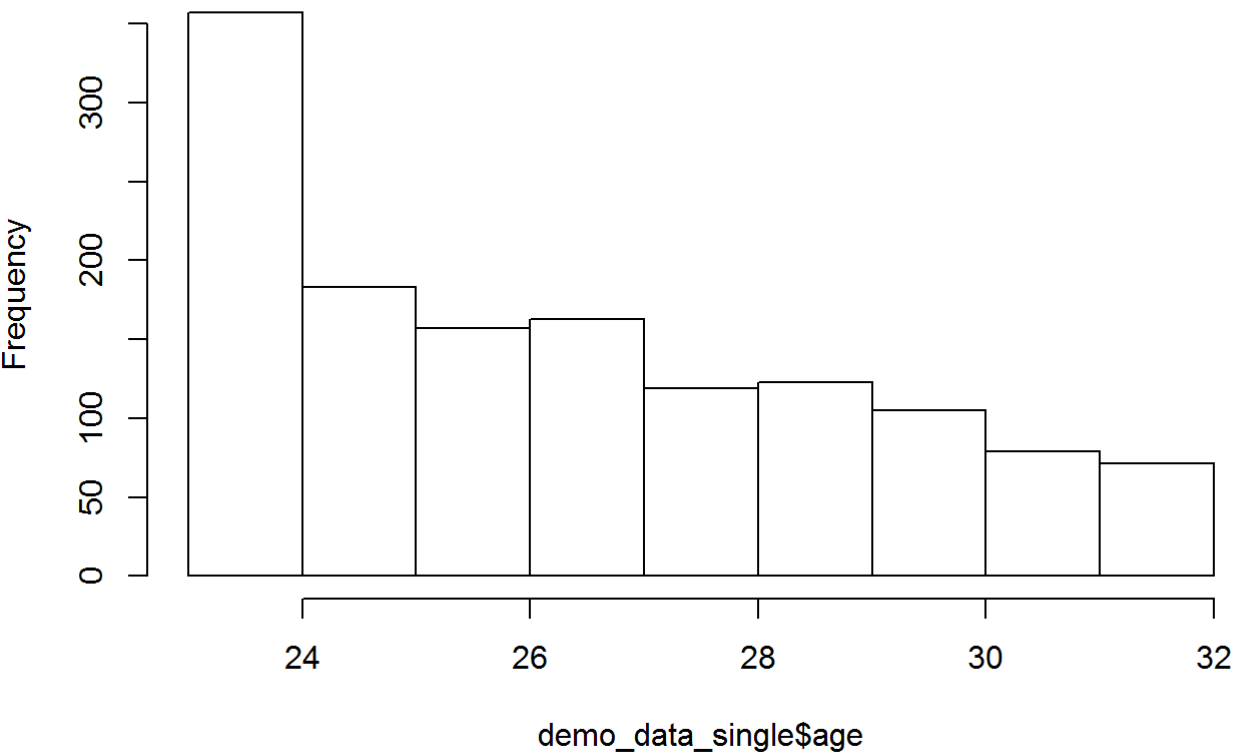
marital

Histogram of demo_data_nNA_single\$age

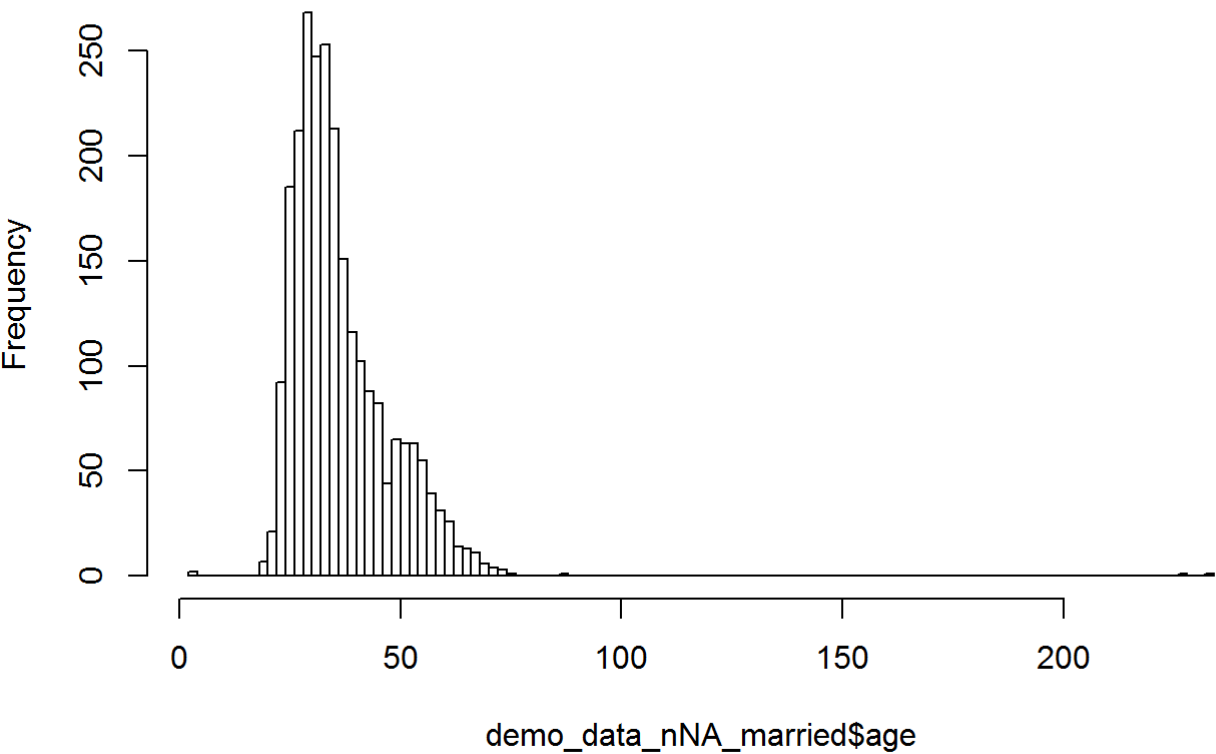


##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	17.00	23.00	27.00	29.25	32.00	95.00	1

Histogram of demo_data_single\$age

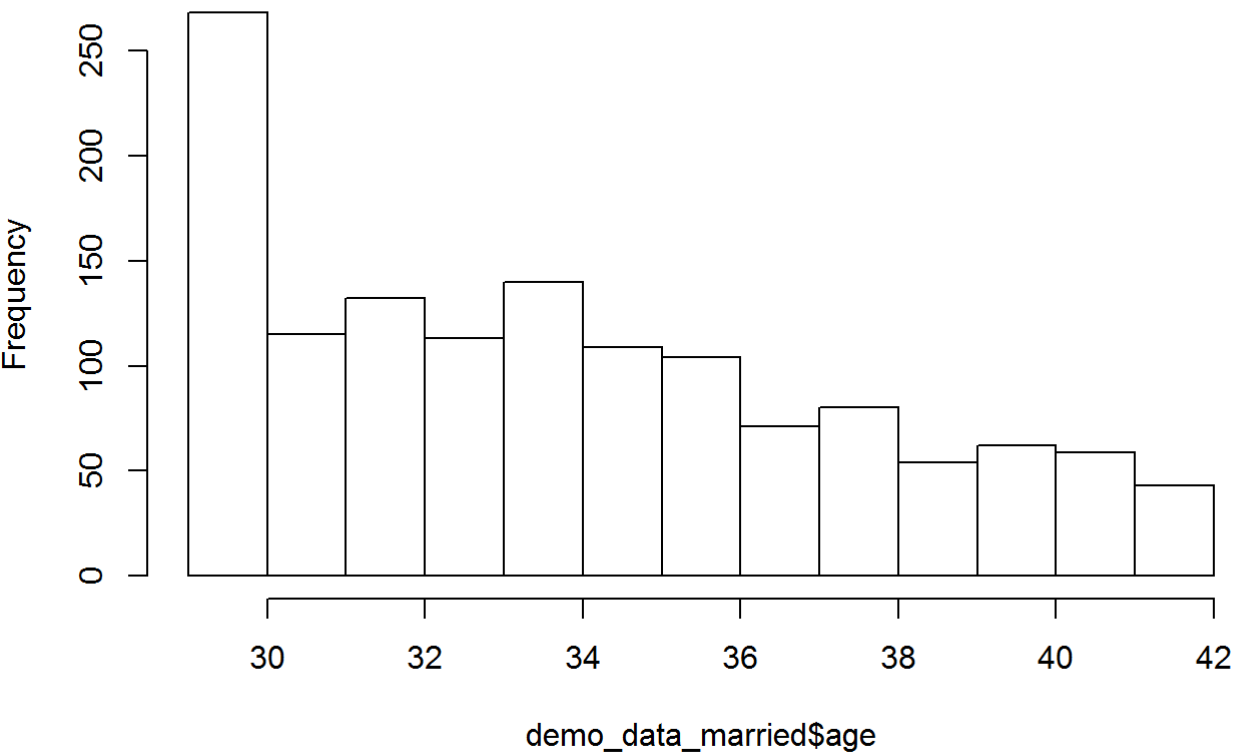


Histogram of demo_data_nNA_married\$age



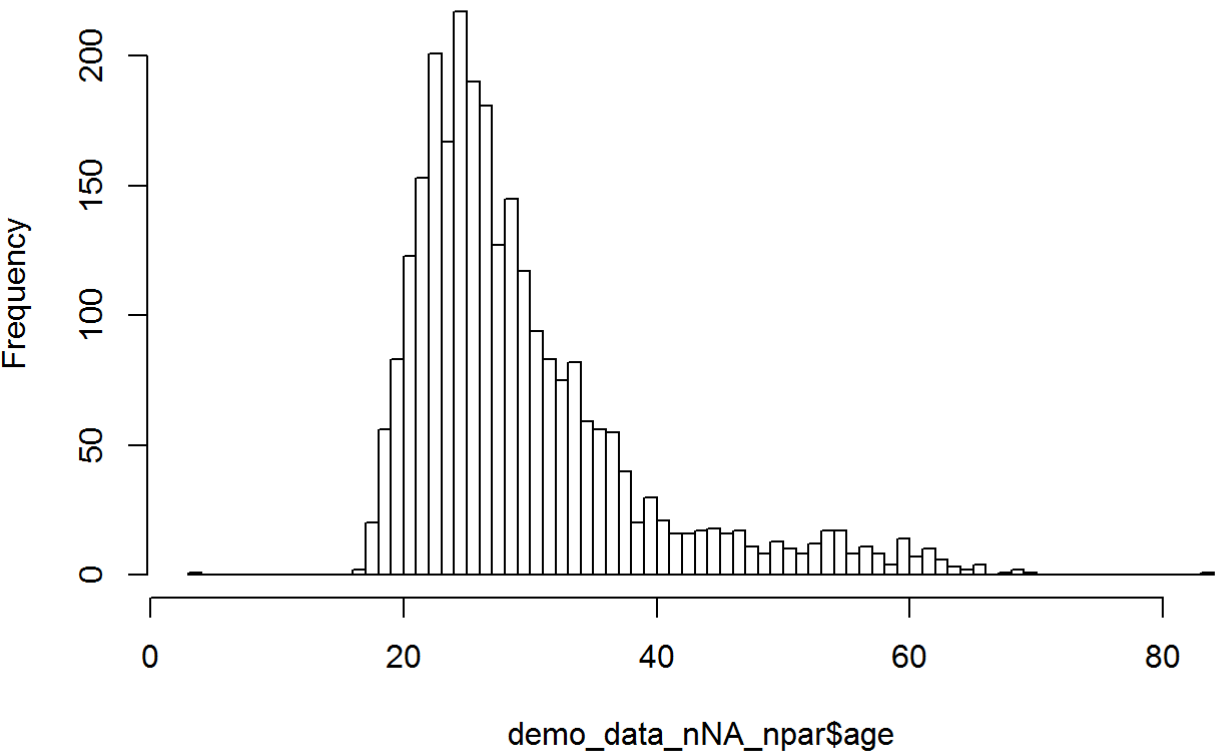
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	3.00	29.00	34.00	37.01	42.00	233.00

Histogram of demo_data_married\$age



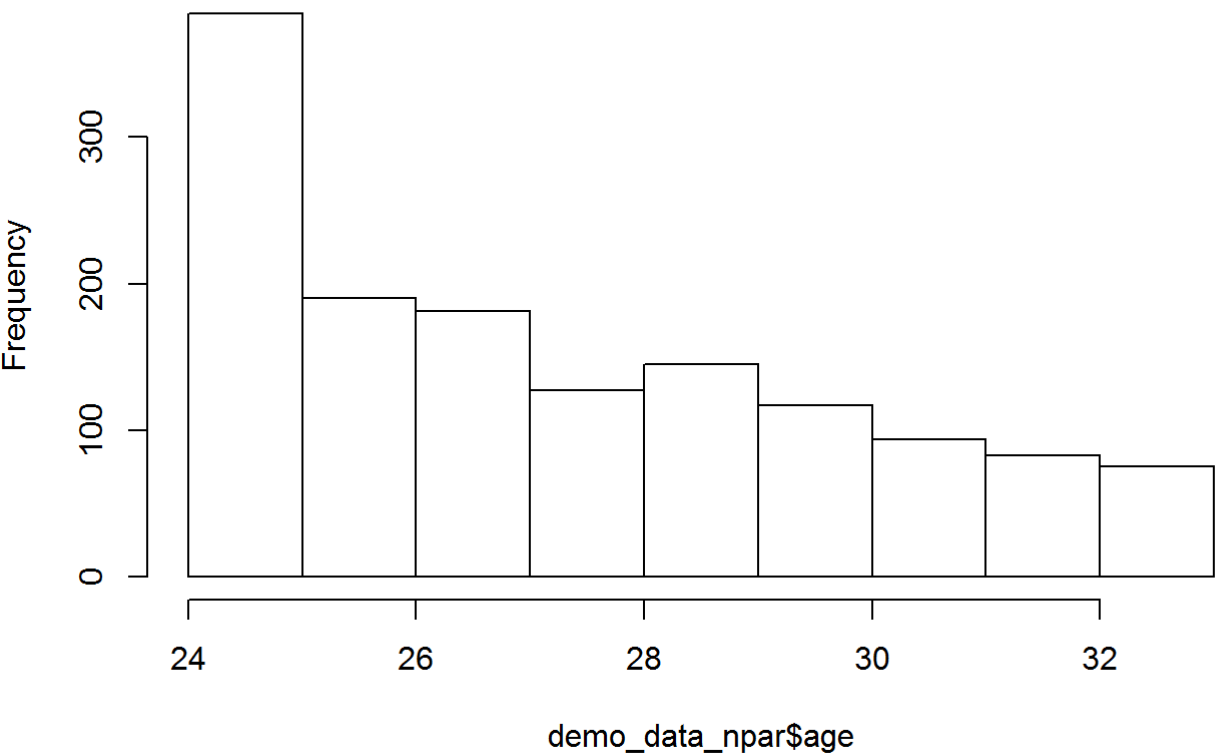
parenthood

Histogram of demo_data_nNA_npar\$age

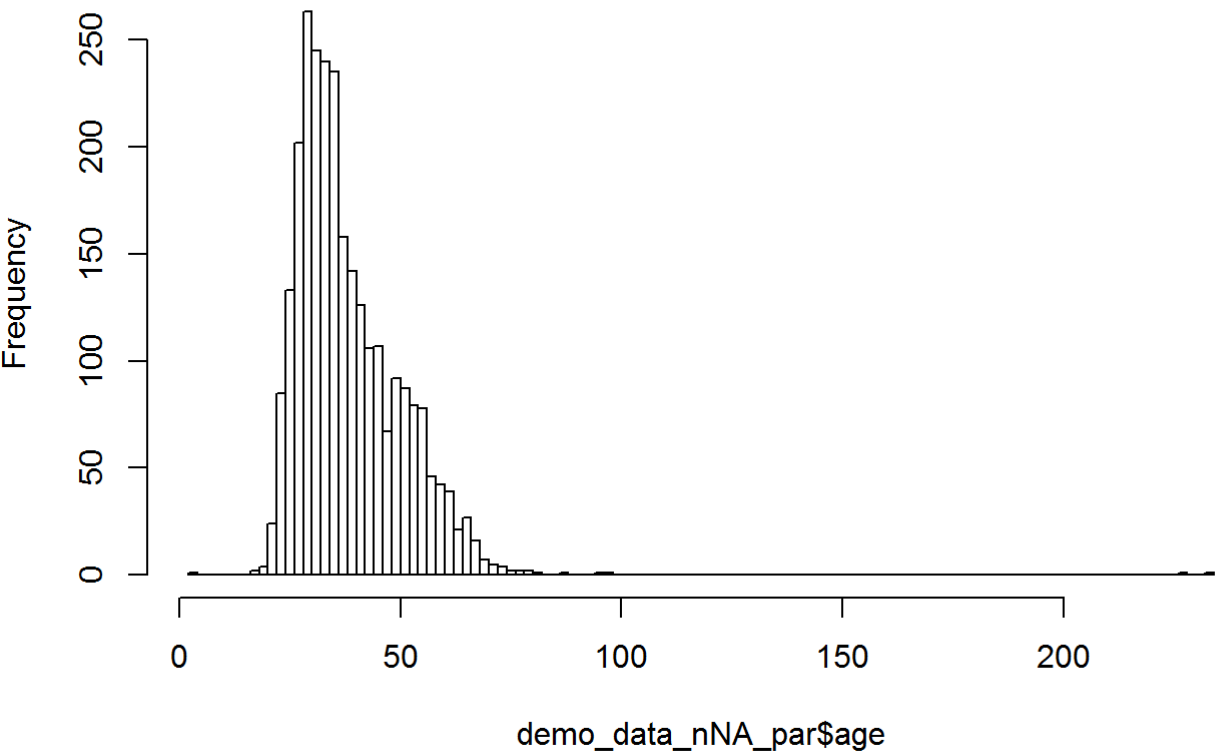


##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	3.00	24.00	27.00	29.88	33.00	84.00

Histogram of demo_data_npar\$age

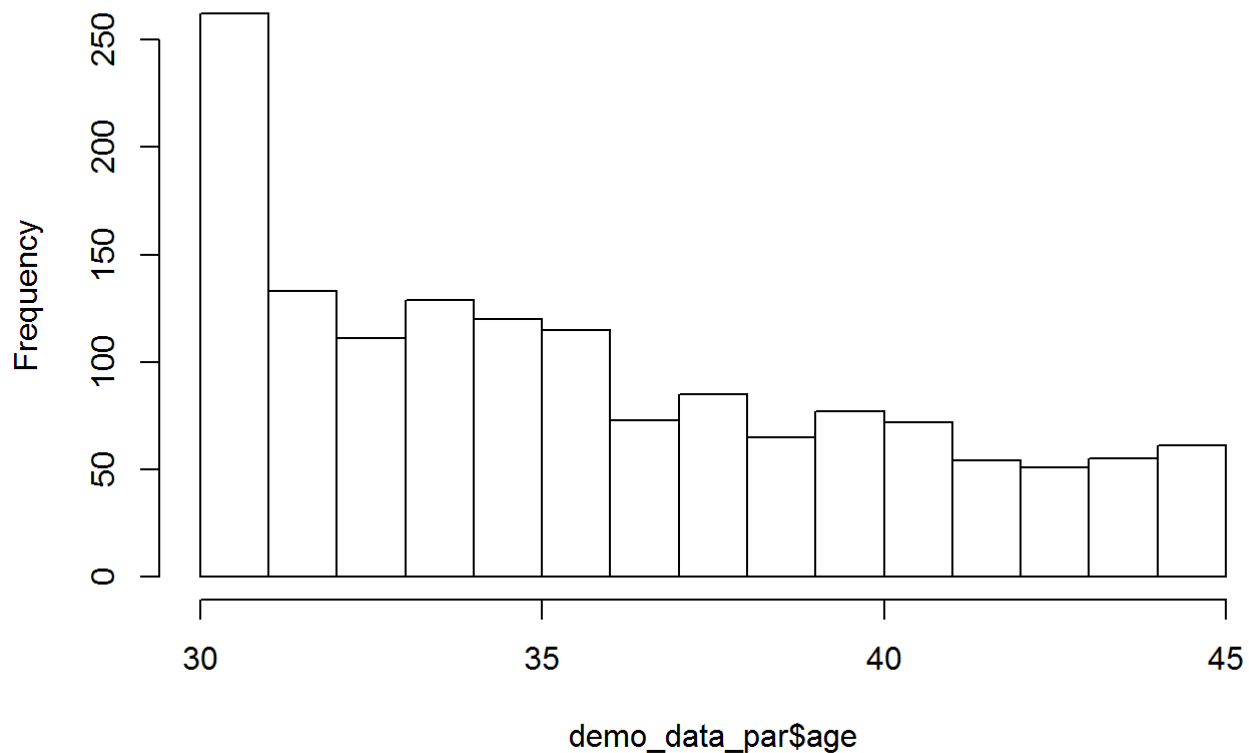


Histogram of demo_data_nNA_par\$age



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	3.00	30.00	36.00	38.83	45.00	233.00	1

Histogram of demo_data_par\$age



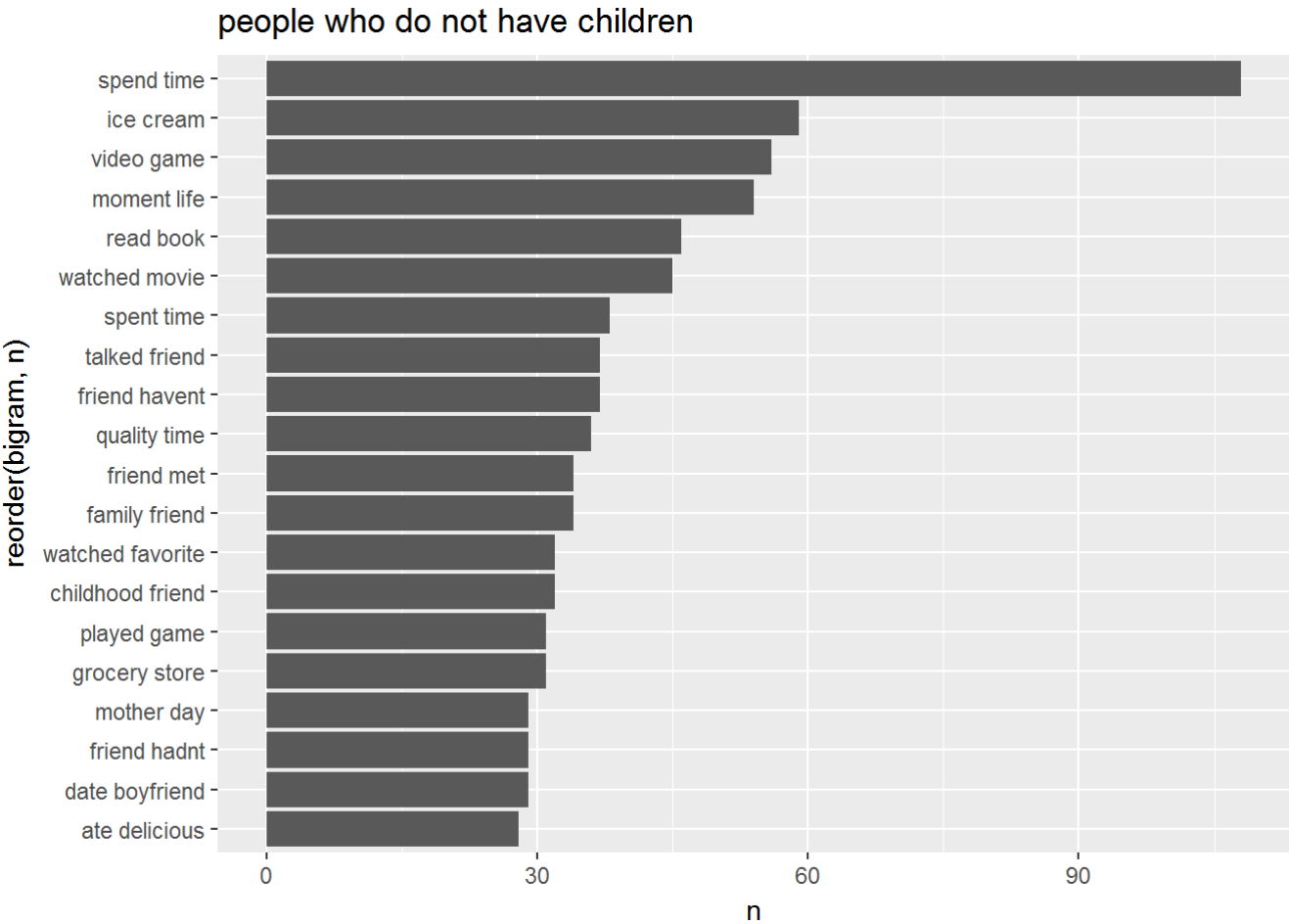
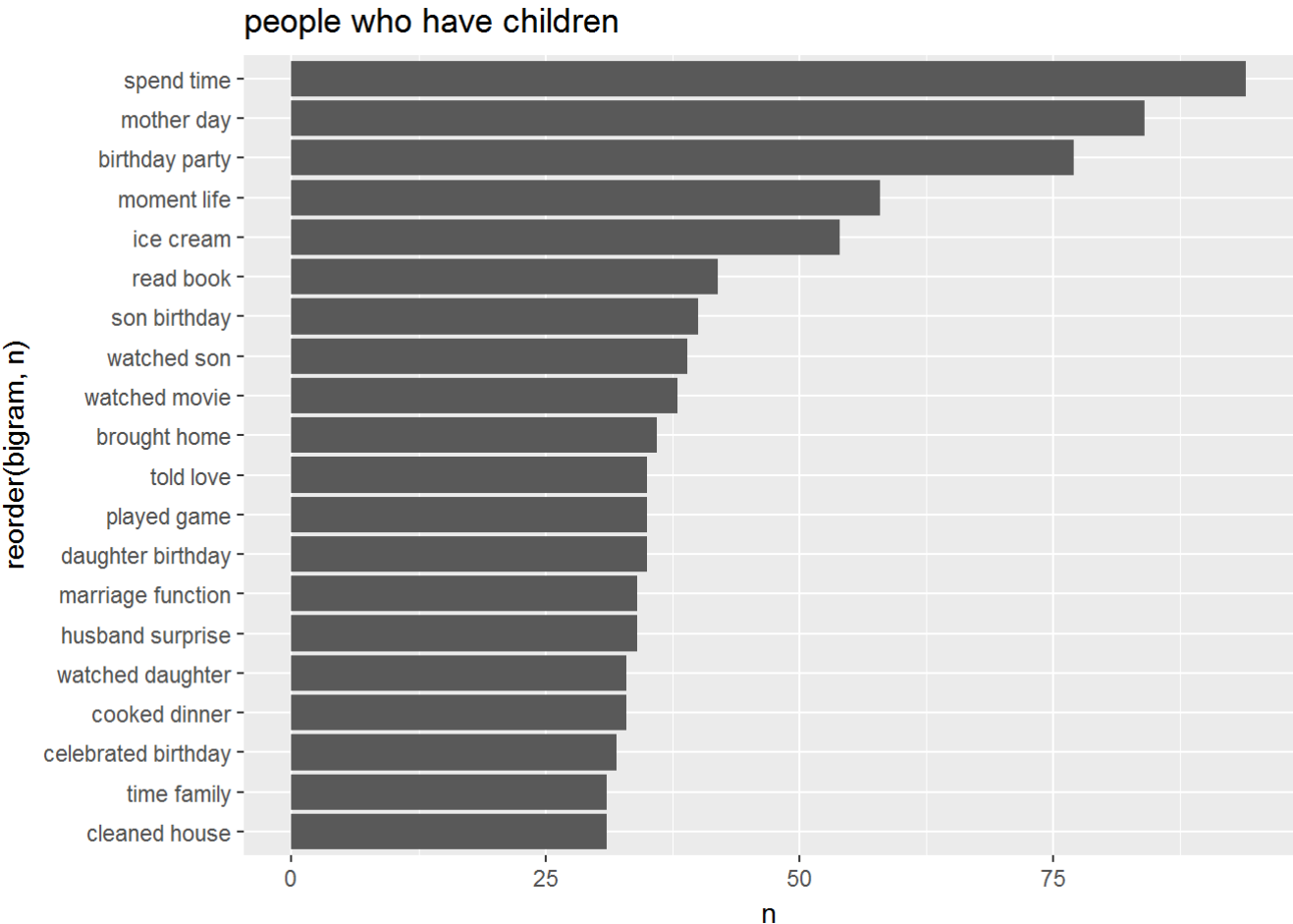
subset the data

```
#subset of singles in certain age range
demo_data_single_wid <- demo_data_single$wid
hm_data_single <- hm_data[hm_data$wid %in% demo_data_single_wid, ]
#subset of married in certain age range
demo_data_married_wid <- demo_data_married$wid
hm_data_married <- hm_data[hm_data$wid %in% demo_data_married_wid, ]
#subset of people who do not have children in certain age range
demo_data_npar_wid <- demo_data_npar$wid
hm_data_npar <- hm_data[hm_data$wid %in% demo_data_npar_wid, ]
#subset of people who have children in certain age range
demo_data_par_wid <- demo_data_par$wid
hm_data_par <- hm_data[hm_data$wid %in% demo_data_par_wid, ]
```

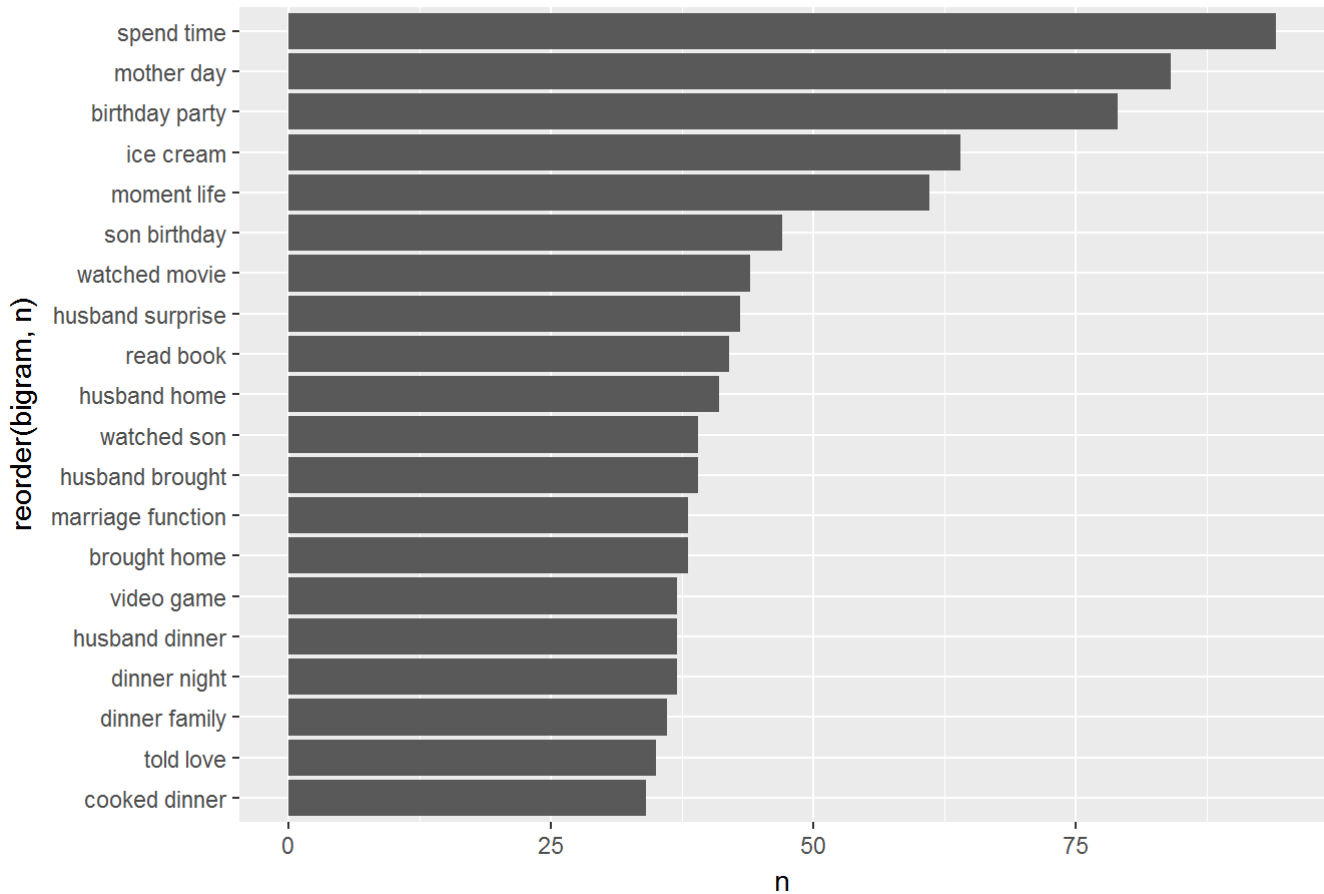
```
### subset of singles
#### Create a bag of words using the text data
bag_of_words_single <- hm_data_single %>%
  unnest_tokens(word, text)

word_count_single <- bag_of_words_single %>%
  count(word, sort = TRUE)
```

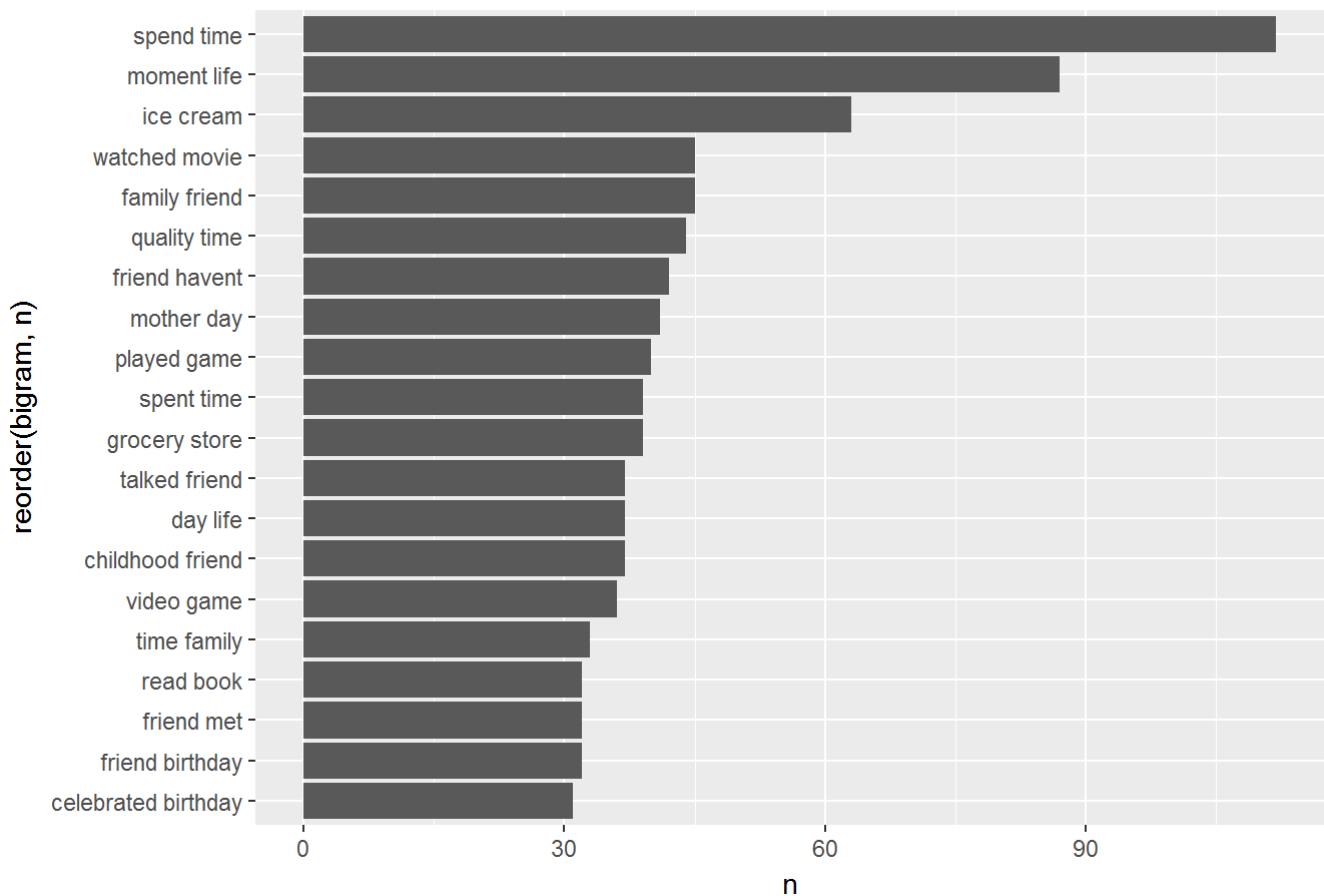
have vs not have children



people who married



people who are still single

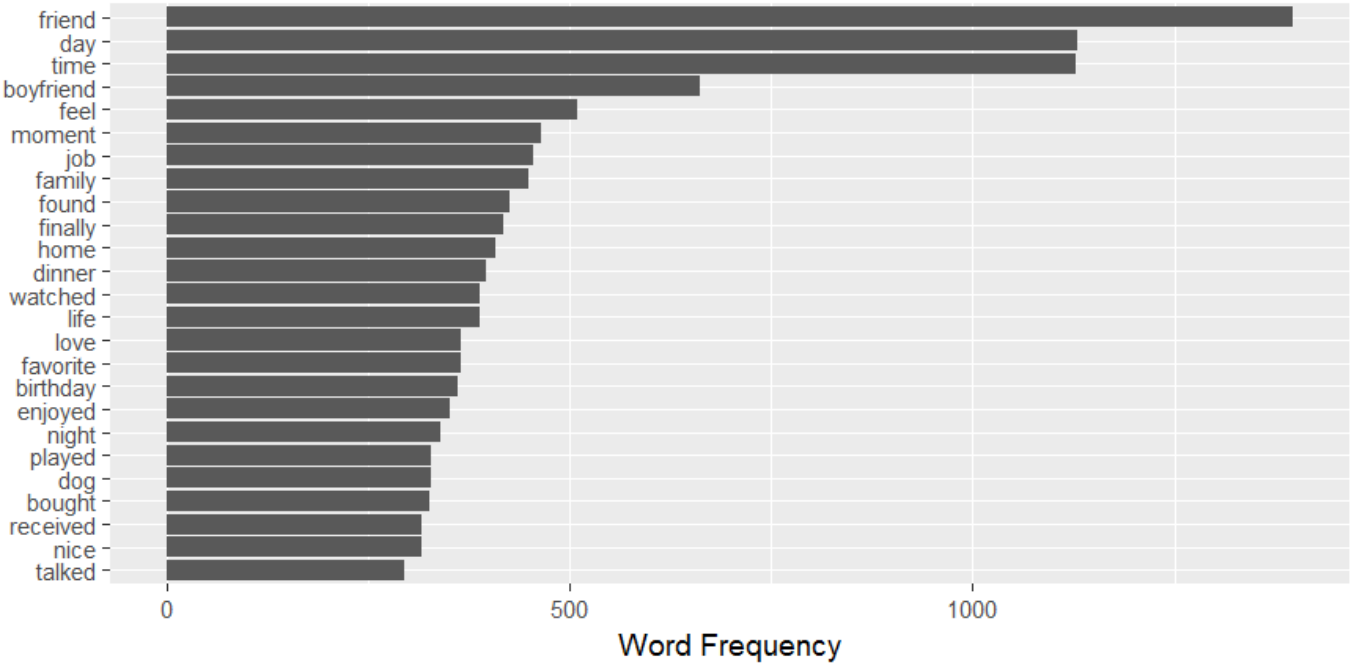
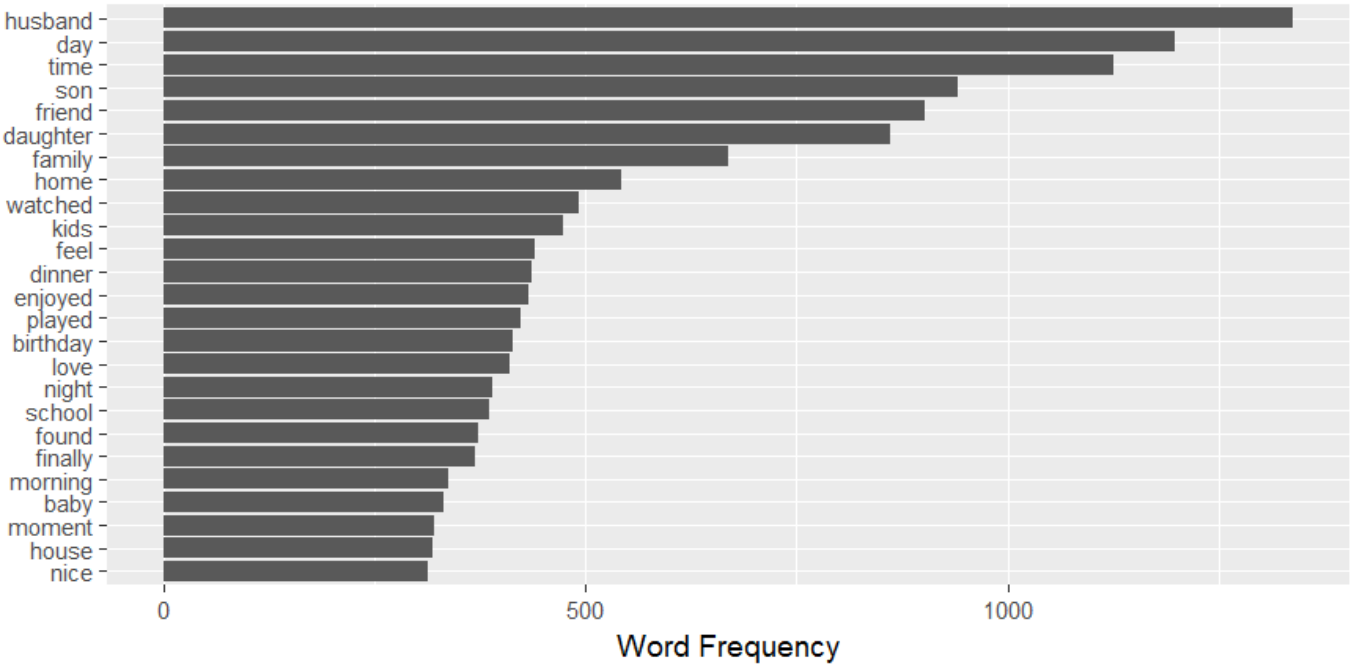


we find out that for the single women in major age range are more focusing on their personal life like their own family and friends, their friends birthday. But when women getting married, we find out that more and more people gain their happiness from their husband and their childrens, also marriage function. and compare to the children chart

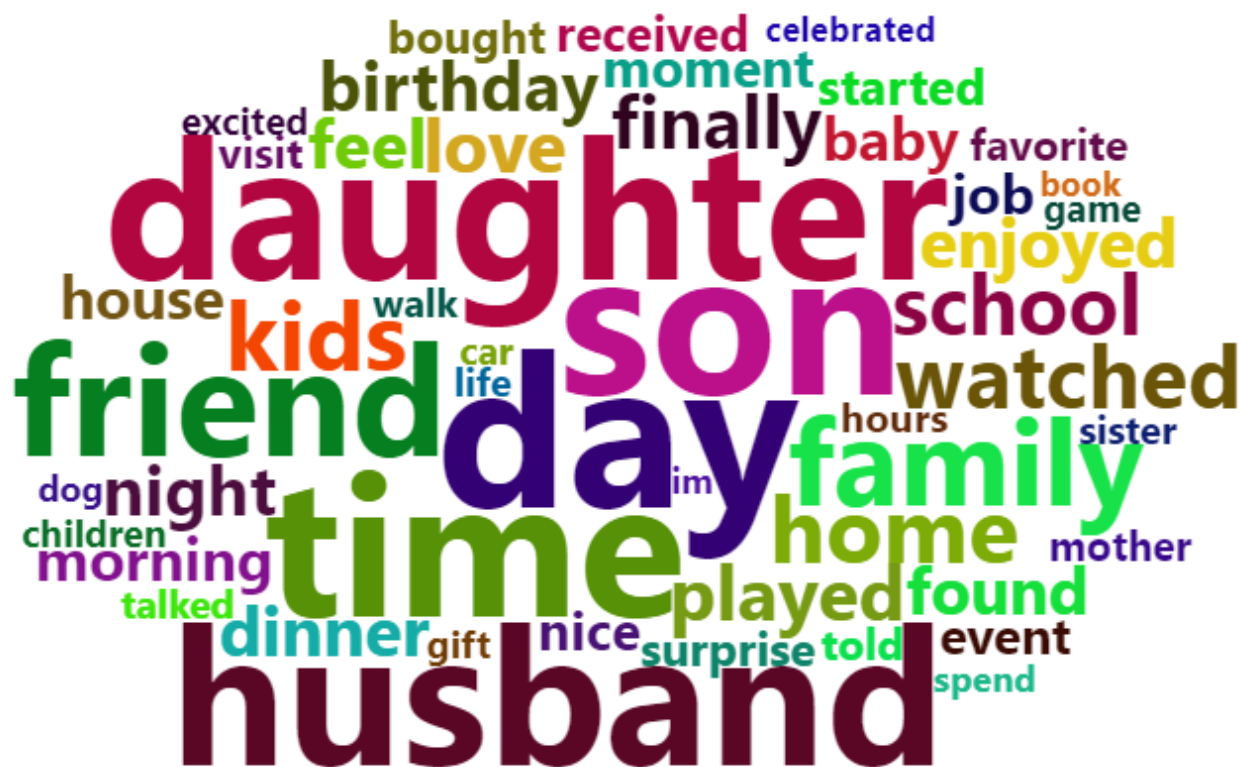
lastly using HappyDB_rShiny.rmd to produce the graph of word cloud

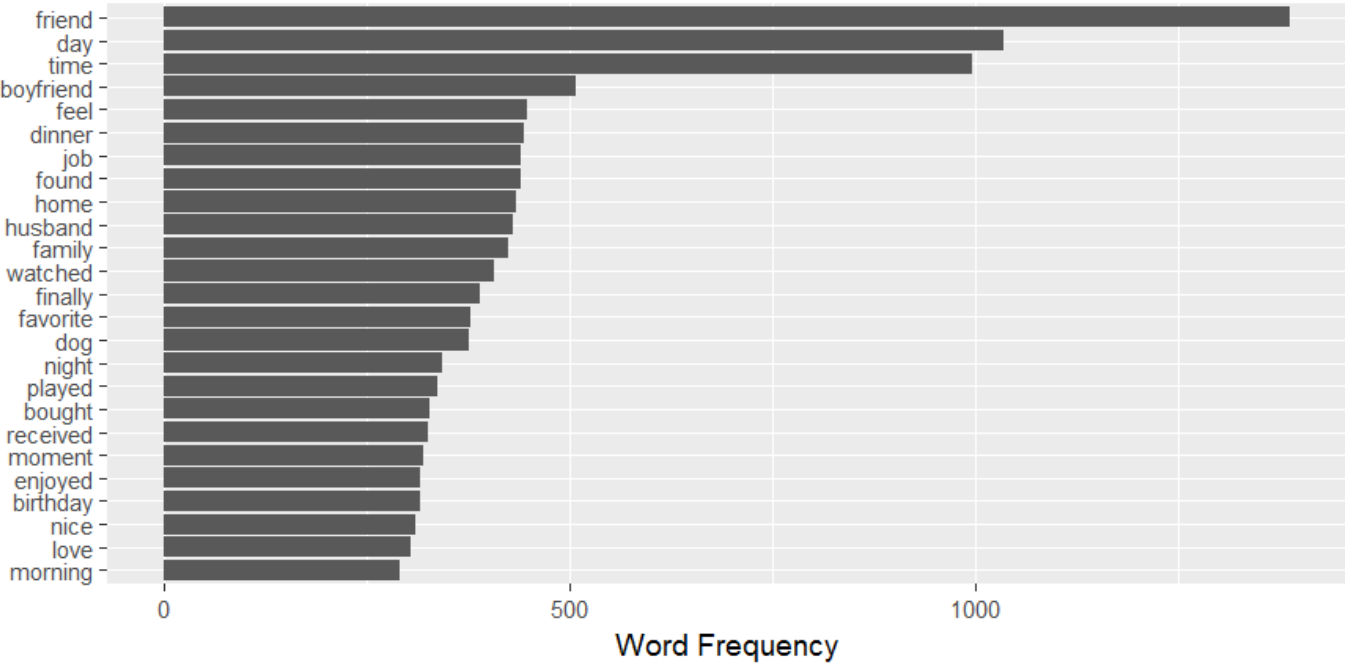
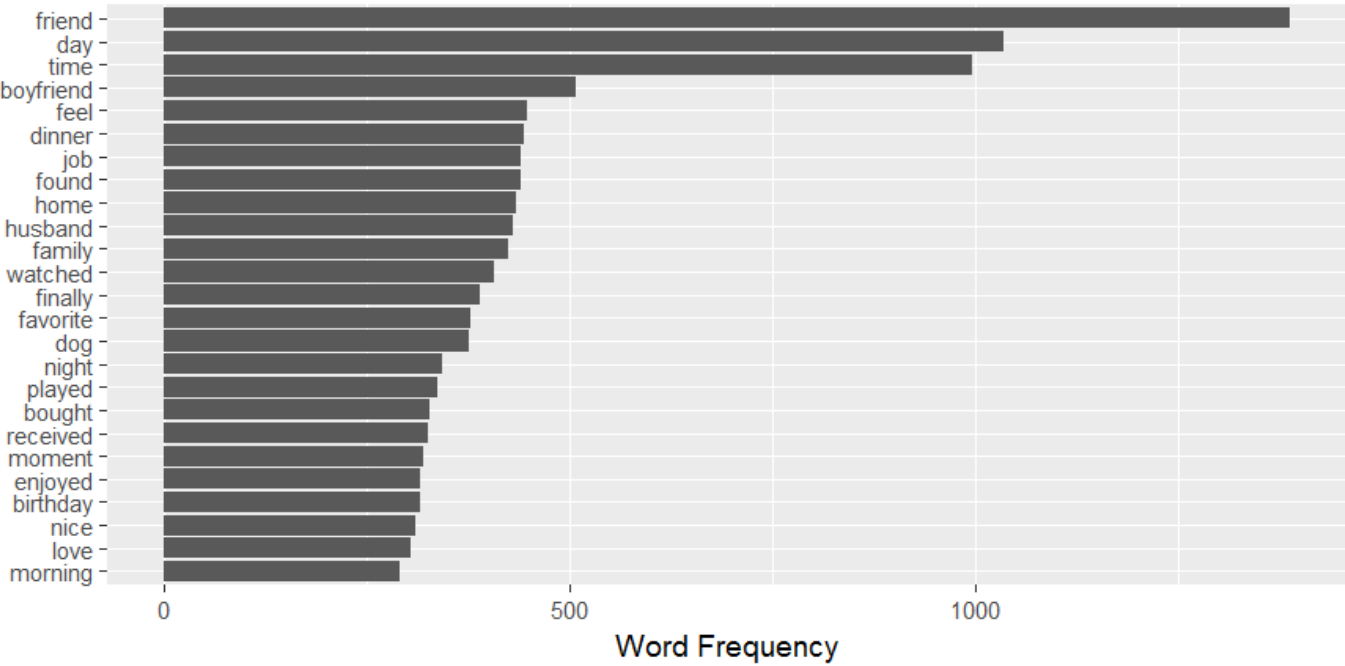
[illegible][illegible]

10/17



compare parent word cloud and nonparent
word cloud





in conclusion. for people don't have children or aren't married yet, we should focusing more on our girlfriends personal life like their feelings, their mental world. since in that case, more women are not generate happiness major from the matirial world yet.

For people who already have child or already married. we should focusing more on the real world problem like brought home and we should focusing on invest our assets more on our child and the time that shared by family since women at this segment tend to generate their happiness more from family and matirial world.

but the most important as both chart indicates the number one thing we should do is to spend more time with them.