GR5243-Project1

Chenghao Yu cy2475

September 18, 2018

HappyDB is a corpus of 100,000 crowd-sourced happy moments via Amazon's Mechanical Turk. You can read more about it on https://arxiv.org/abs/1801.07746.

Here, we do some analysis on the cleaned data and try to dig some deep relation between the happy moments and users' demographic information, especially the relation between happy moments and marital status and parenthood.

Step 0 - Load all the required libraries

```
##                _
## platform       x86_64-w64-mingw32
## arch           x86_64
## os             mingw32
## system         x86_64, mingw32
## status
## major          3
## minor          4.4
## year           2018
## month          03
## day            15
## svn rev        74408
## language       R
## version.string R version 3.4.4 (2018-03-15)
## nickname       Someone to Lean On
```

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

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

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

Create a bag of words using the text data for different marital status group

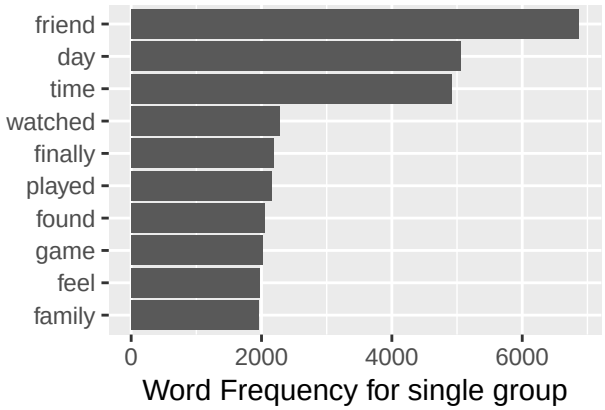
Create bigrams using the text data

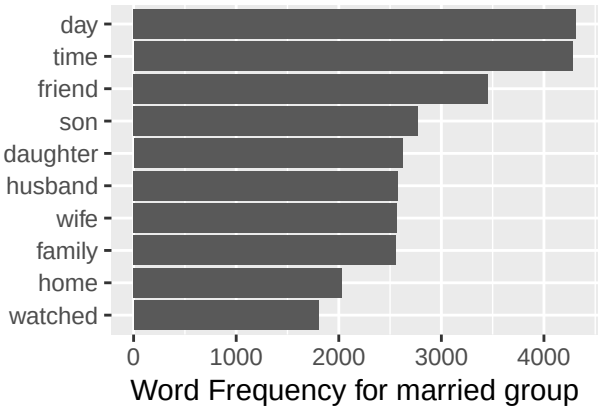
Data Bolg

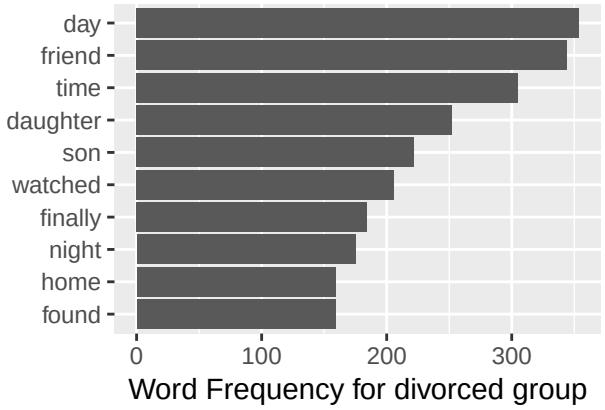
How Being A Parent Make Your Life Happier.

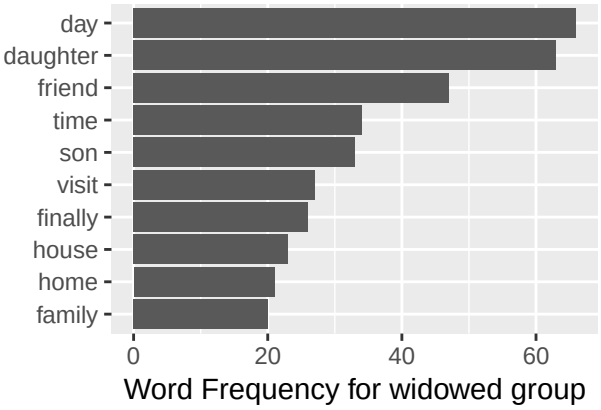
Nowadays, the increasing number of well educated people tend to be a DINK(Double Income No Kids). Their reasons are variance but the common sense in this group is that the parenthood cannot bring them more happiness than other status in their lives.

However, the analysis on the HappyDB data shows that the subject's status do not make much difference, except the parenthood.



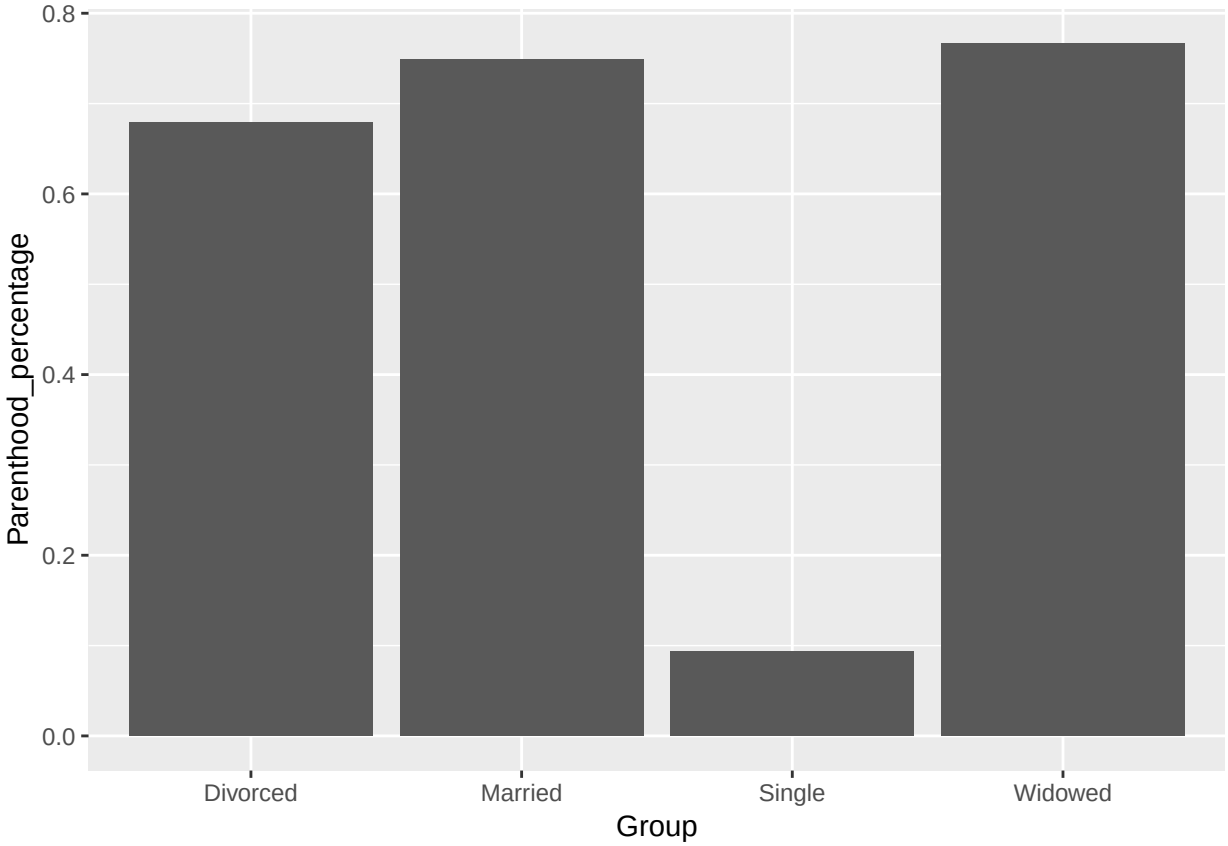




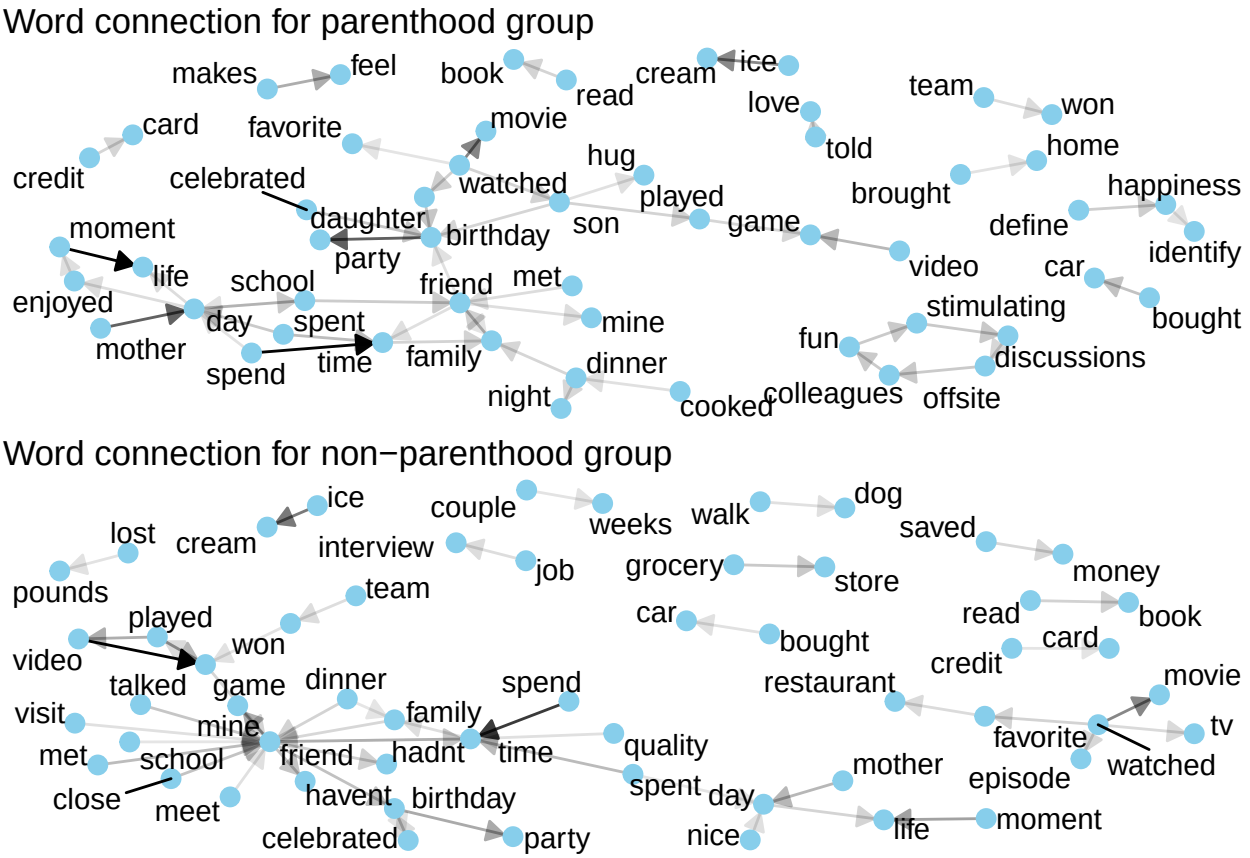


By compaing the bar chart for four groups with different martial status, there are no remarkable differences about the top rank of happiness topics between four groups. The single group seems do not have the topic related to parenthood in top ranking, which because the percentage of parenthood in single group in much lower than it in the other three group.

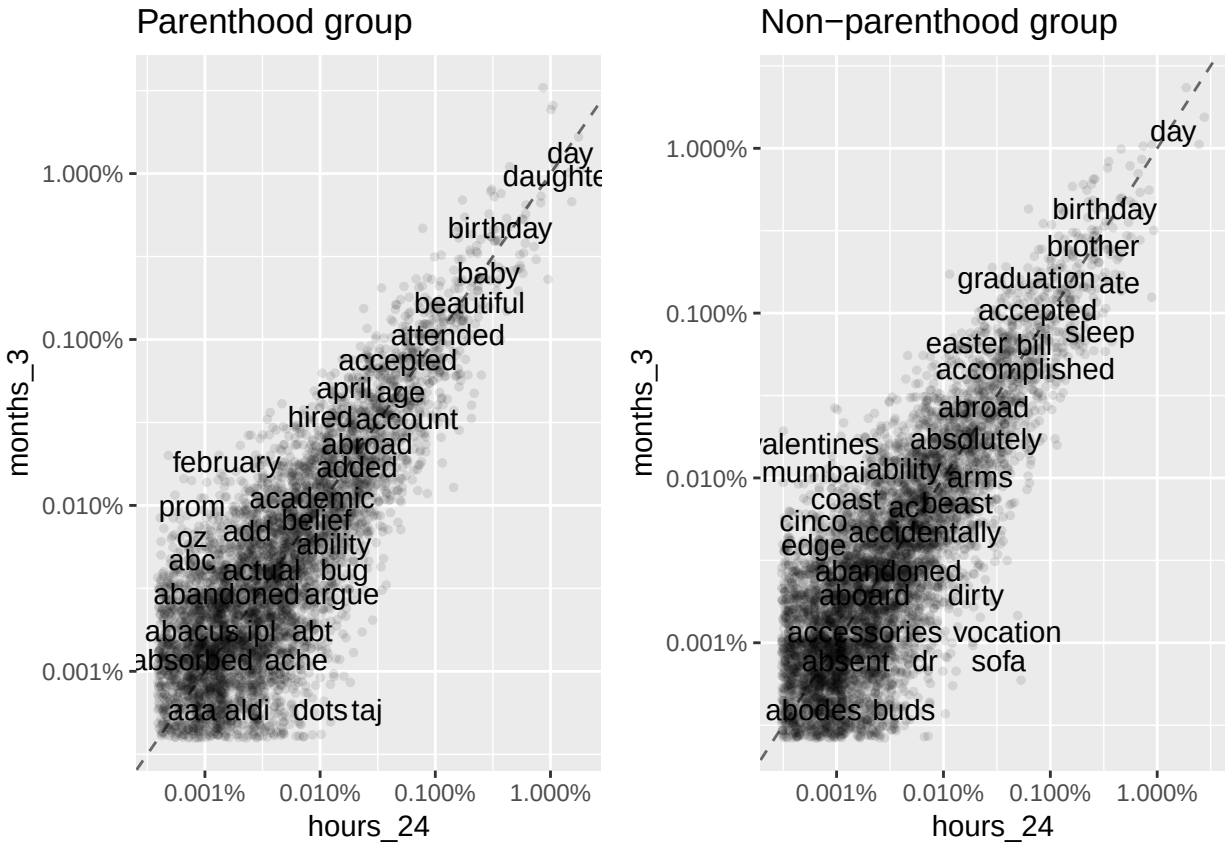

0
50
100
150
200
Single
Married
Divorced
Widowed



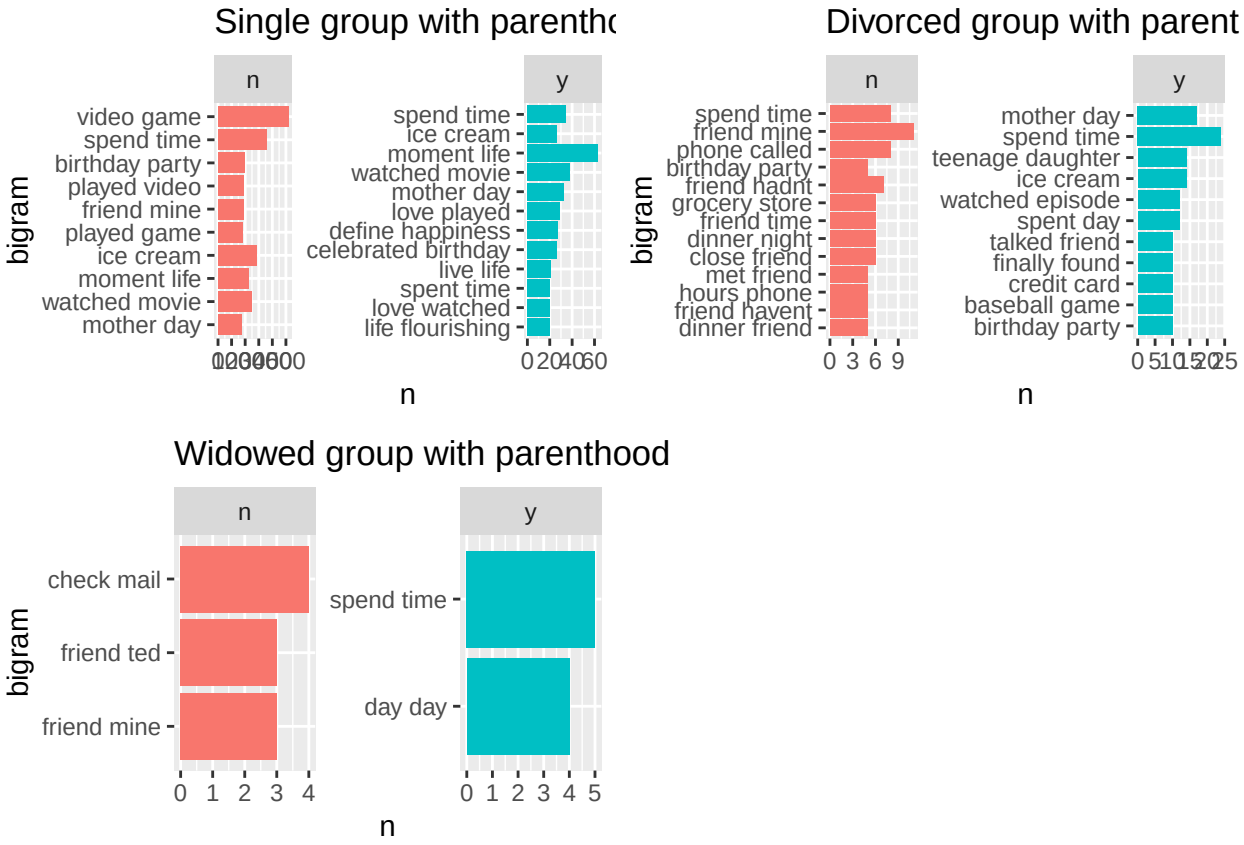
Also, the range of age is highly related to the martial status, but according to the bar chart above, the only thing influenced by age is that the people in single group are not old enought to be a parent.



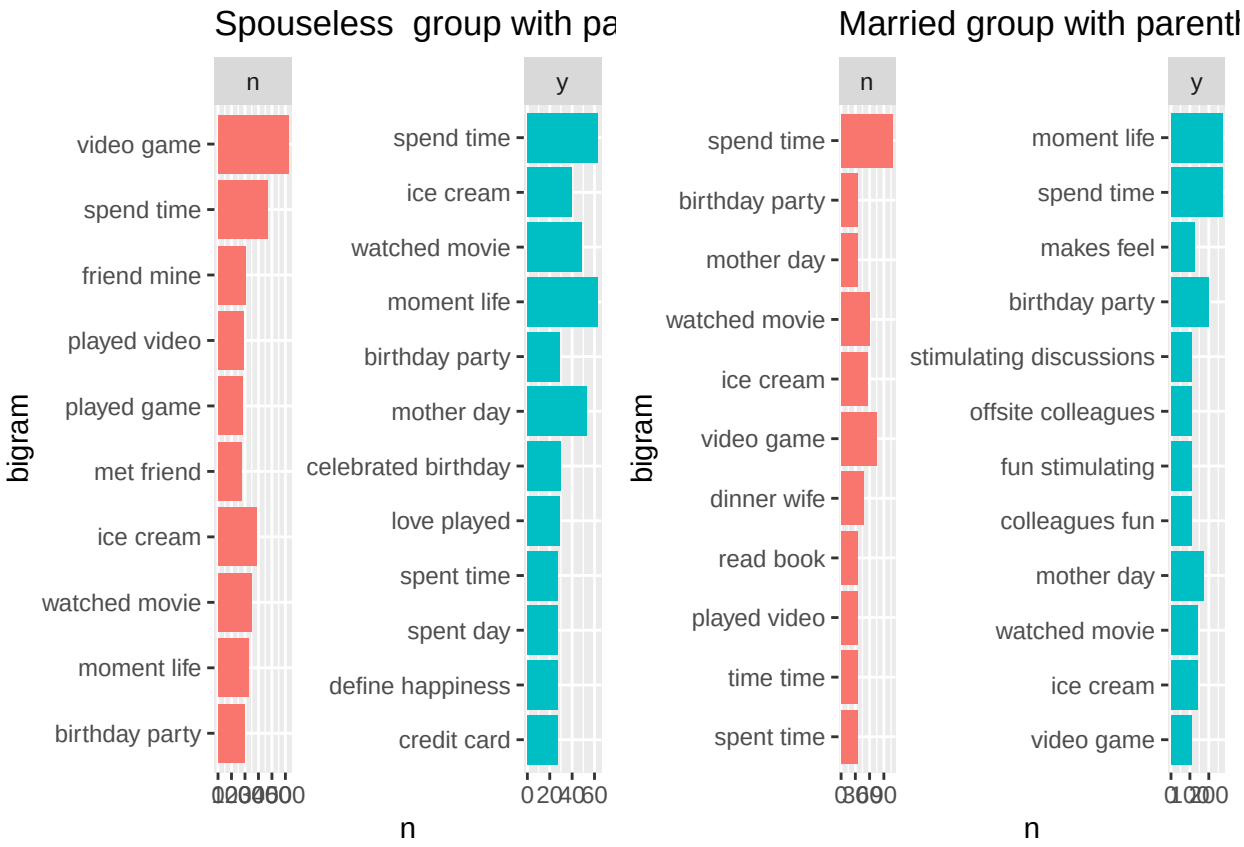
As the words connection chart shows, after being a parenthood, the number of topices which have path to the topic “family” are increased. The connections between “family” and its left side topics such as “mine”, “game”, “team”, etc is one-way path. These topics cannot be considered as a part of family topics. Meanwhile, in the chart of parenthood group, the topic “birthday” is more closer to the topic “family” and it connected to the “daughter” and “son”, and these are all point to the “family”, they are parts of family topics.



The family group topics like “Birthday”, are both long period topics as well as short period topics, so any topics join the family group topics will be the resources of happiness for long period and short period. Thus, after “Son” and “daughter” linked to the “fmaily” because a person become a parent, he or she is likely to get more happiness resource and oppotunities.



If a person do not have children and spouse at the same time, there are two topics may be used as a offset. The "friend" and "game" are new topics appears only in the spouseless group without parenthood.



For more detailed, people may get happiness from "game" when they do not have children. It may infer that the happiness to be a parent is similar as raise a child. The sense of achievement may be the root of these happiness. The "friend" is likle to be the offset of spouseless. It is reasonable because both the friendship and marriage relationship are the emotion relationship.