Project1–What made you happy today?

Xinyi Chen xc2464

There are countless moments in our daily life that can bring different emotions: happy, sad, excited, angry,etc. In this project, we are interested in one question: What makes people happy? In order to figure out the categories of main reasons that make people happy and the characteristics of happy moments for different groups of people, natrual language processing and text mining methods were applied.

Overview: causes of happiness

Statistic	Data
Total Happy Moments	100,535
Happy Moments in past 24 hours	49,831
Distinct Respondents	10844
Average Happy Moments/Person	9.27

The dataset used for this analysis is HappyDB. The HappyDB is a corpus of 100,000 crowd-sourced happy moments. It contains happy moments (in past 24 hours or 3 months ago) and some additional information in addition to original happy moments, like work id, age, gender, marital status for respondants. The table above is an overview of the happydb dataset. Table view is not enough to show key words of the happy moments, so it is time to show a wordcloud plot.

```
wordcloud(tdm.overall$term, tdm.overall$`sum(count)`,
          scale=c(5,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Blues"))
```


From the wordcloud plot, it looks that "work" made a majority of people happy, at least in the past 24 hours, which is unexpected. We can also tell that people like "friend", "daughter", food, family are also strong factors that can bring people happiness.

After the overview, it is obvious that things which made people happy can be approximately classified into several classes: people, family, food, pets, amusements, which almost perfectly matches the exsiting topics in topic_dict file from the original happydb dataset. Thus, I decided to use those as the topics for terms of happy moments.

Topic Analysis

The below graph shows the frequency of each topic. Unlike that showed in wordcloud plot, this bar plot tells that "family" is the most important influencing factor of making people happy, rather than "work". Thus, the reason that "work" is so conspicuous in wordcloud plot is that, although mentioning "work" in the moment of happiness, the true meaning that the sentence would like to convey may be quite different than "working makes people happy". For example, "I went to shopping with my friends after work." The main point of this sentence is "after-work activity make me happy" instead of "work" itself. Thus, topic selection and modeling is indispensable in natural language processing and data analysis.

```
my_df$topic <- as.character(topic$x)

my_df <- my_df[my_df$topic != "none",]
term_count <- table(my_df$topic)
barplot(term_count[order(term_count)], las=2, horiz = T, main = "Topic Frequency", co
l="lightpink",border = "lightseagreen")
```
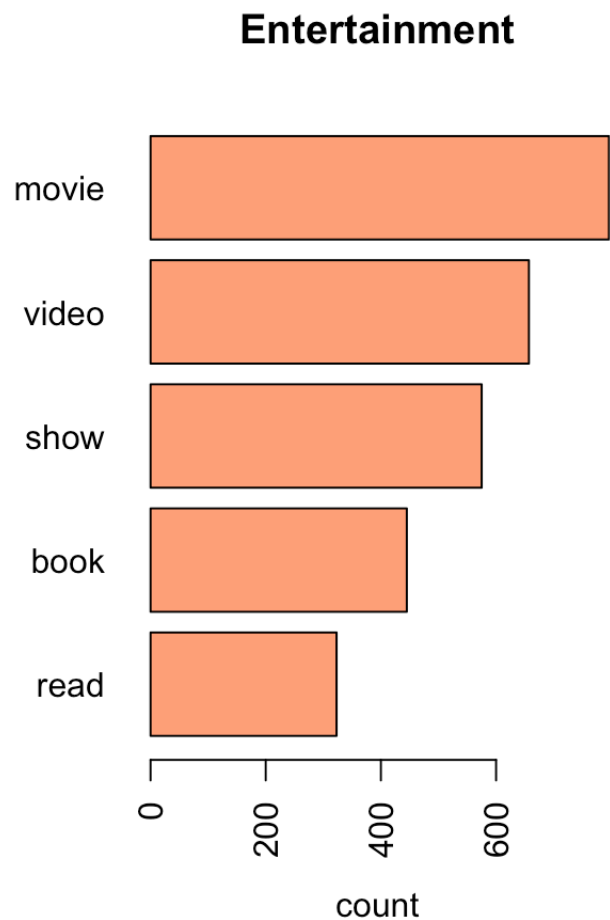
Topic Frequency
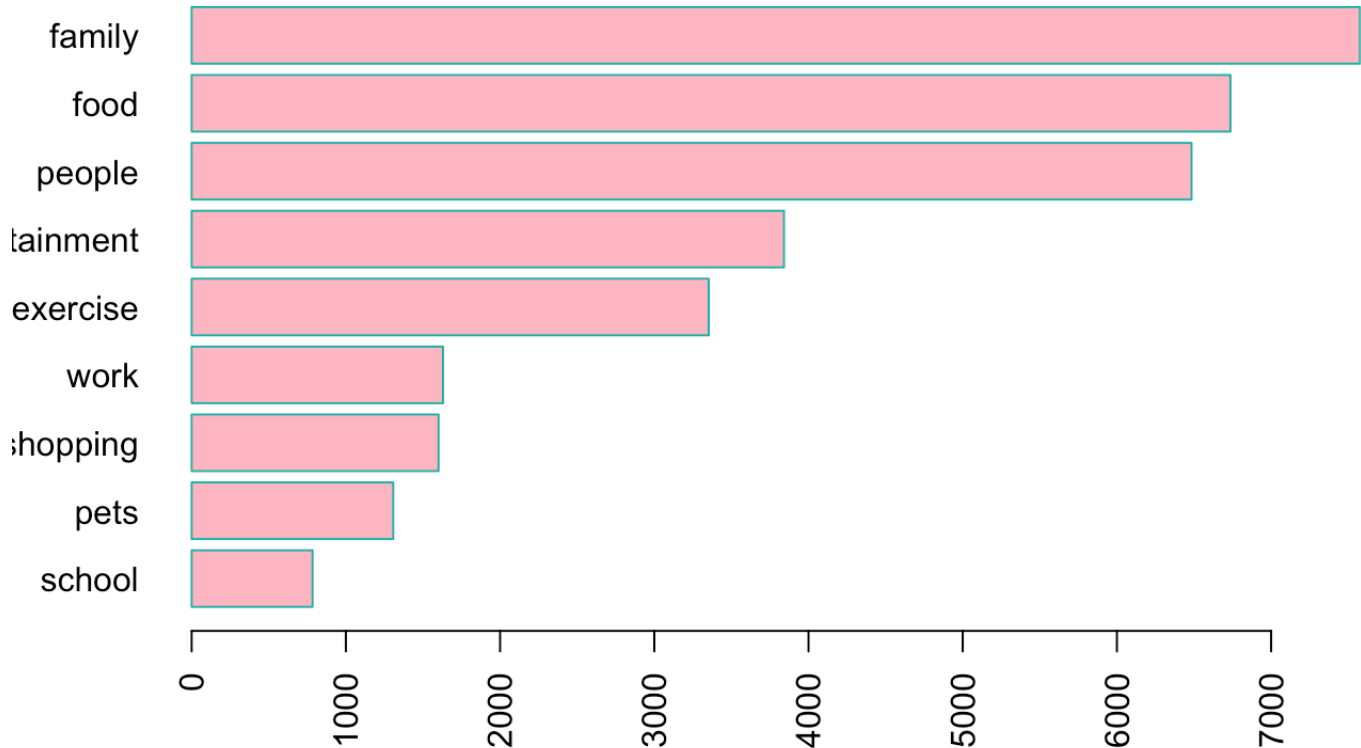
What is the most popular entertainment ways/sports/people that made you happy?

As topics were selected, we are now interested in the most popular activities/people under some representative topics. ####Top prevalent entertainment ways
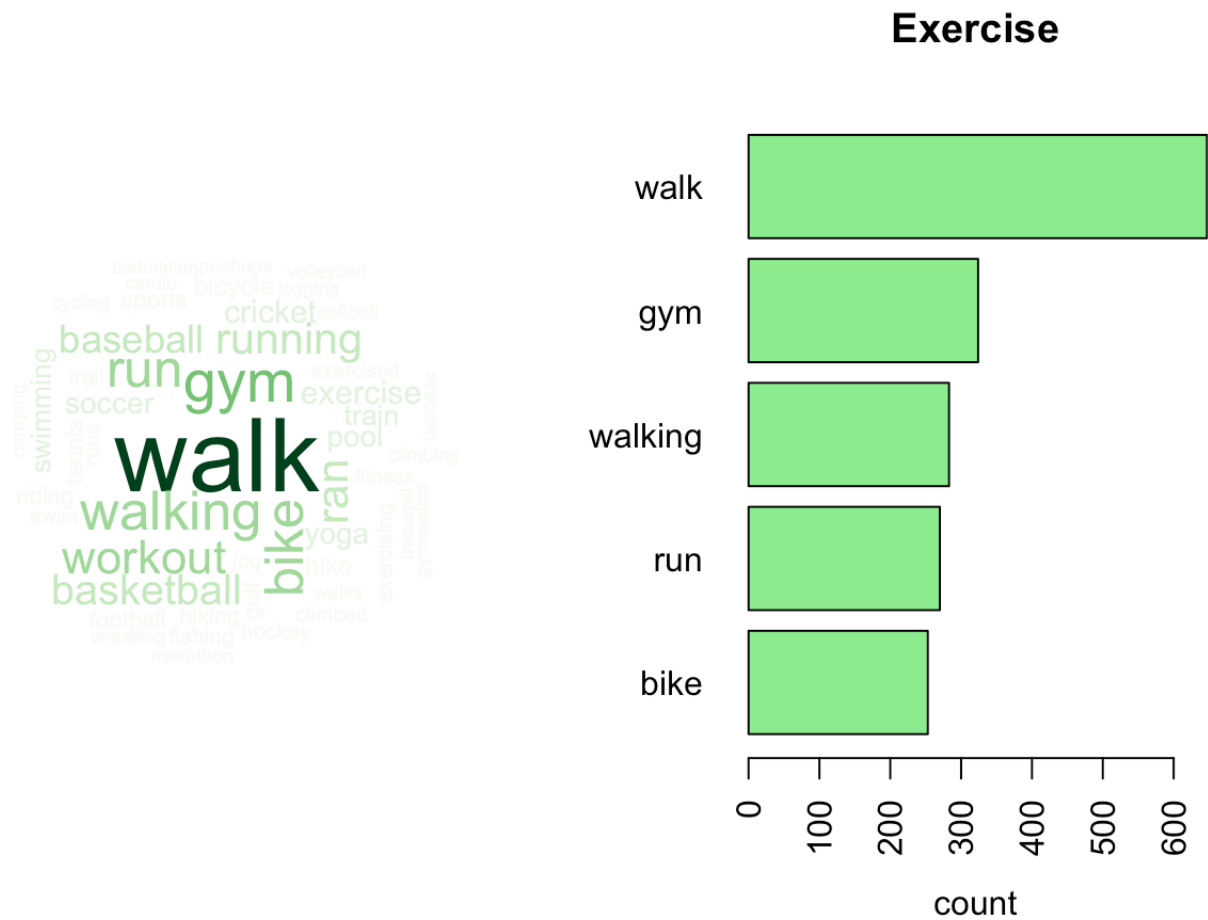
```
cloud_graph <- function(x,color1, color2,t){
  x.tidy <- tdm.tidy[tdm.tidy$term %in% unlist(x),]
  x.overall <- x.tidy %>% group_by(term) %>% summarise(sum(count))
  freq <- table(x.tidy$term)
  freq <- freq[order(freq,decreasing = T)][1:5]
  par(mfrow = c(1,2))
  wordcloud(x.overall$term, x.overall$`sum(count)`,
          scale=c(3,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,color1))
  barplot(rev(freq), horiz = T, xlab = "count", col = color2,las=2, main = t)
}

cloud_graph(entertainment,"Oranges","lightsalmon","Entertainment")
```

From the wordcloud plot and barplot above, most people prefer visual enjoyment like watching movies, videos. Next is reading books. Films ans shows are usually easy with few complicated thoughts needed. Moreover, they have attractive visual appeal, which can won the hearts of audiences. In addition, the respondents of the survey are workers, generally, watching something is always the first choice of relaxing after work.

popular sports

```
cloud_graph(exercise, "Greens", "lightgreen", "Exercise")
```


For exercises, walking won the first prize. In this busy society, walking is one of the most costless and efficient way of keeping healthy. People can walk anytime and anywhere. Replacing the 10-minute driving to work with 30-minute walk would be a good start of a nice day. Besides, going to gym and workout to get a better-look body shape is also the favorites these days. There is nothing better than have muscular body shape, not only healthy, but also beautiful.

family member

```
cloud_graph(familyd,"Reds","lightcoral", "Family")
```

Family

people that make us happy

```
cloud_graph(people,"Accent", "honeydew", "People")
```
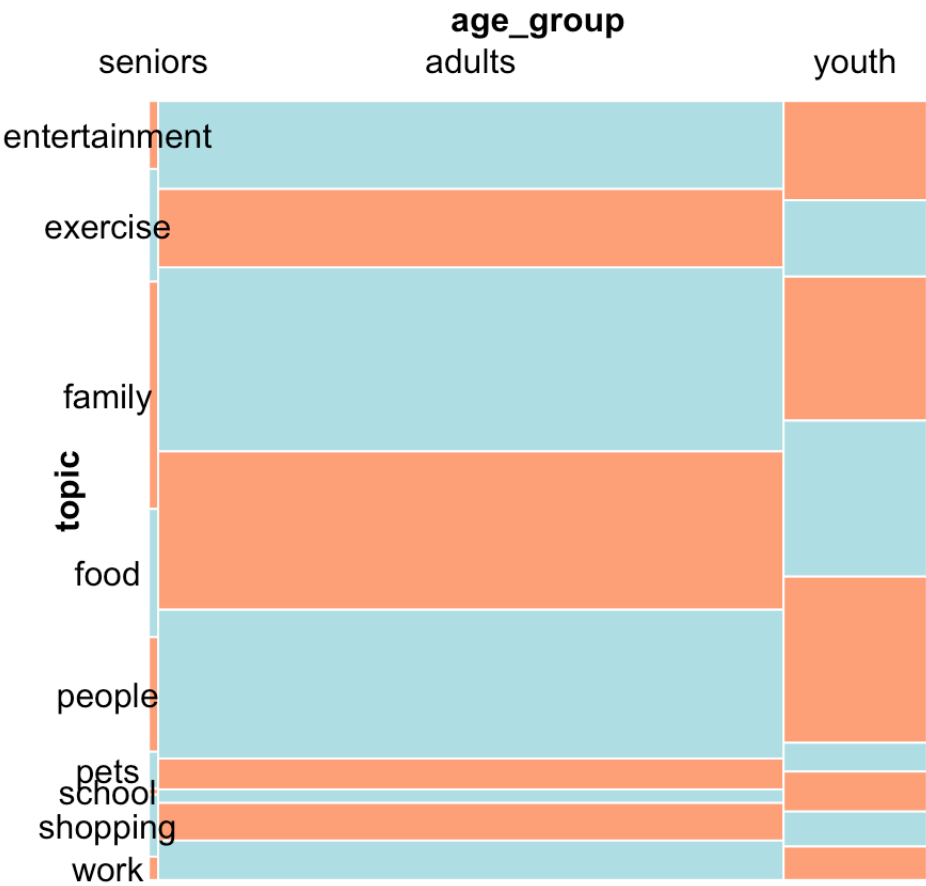
People

The terms of topic "People" and "Family" are quite similar. Most of them are marital partners, kids and friends. Spending time with family and friends can always bring happiness.
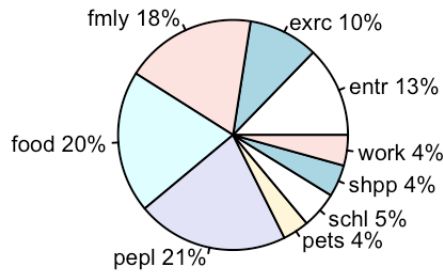
What make different groups of people happy?

Although taking a look into each topics is needed, analyzing between groups is also inevitable.

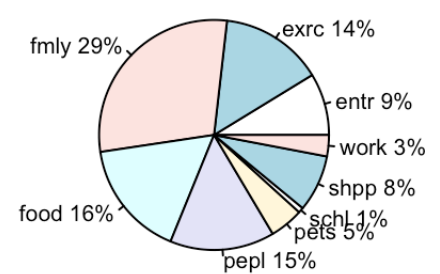## Reasons for different Age Groups

age_group
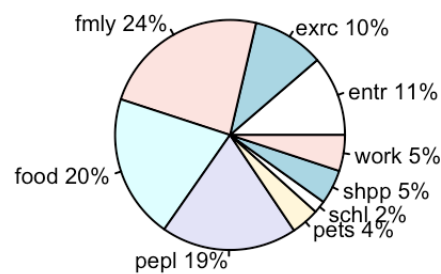
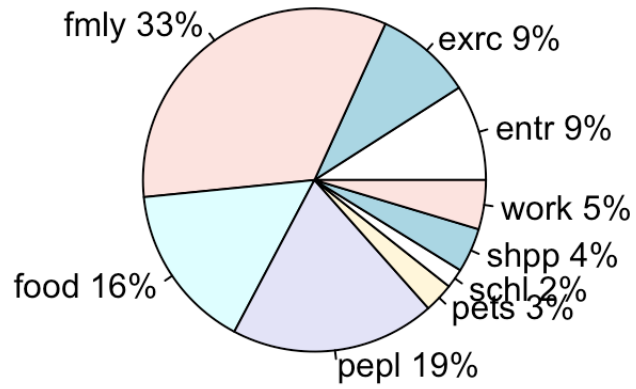**youth**   **adults**   **seniors**
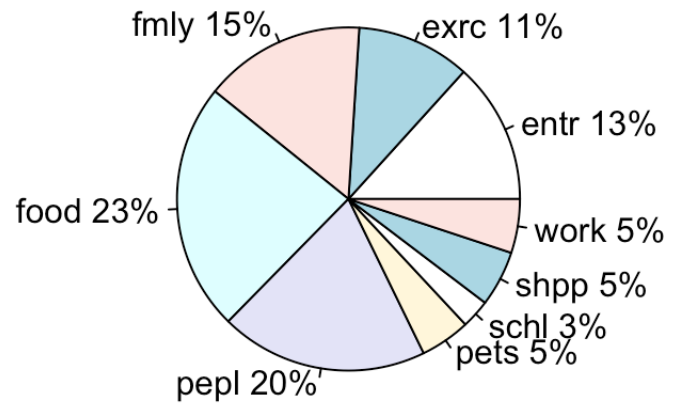

The mosaic plot provided a rough distribution of topics in each age group. For adults (25~65) and seniors(>65), happiness mostly comes from family; nevertheless, young people are more likely to to get gratification through food and friends. In other words, as getting older, the proportion of "family" accounts more of the reasons for feeling happy.

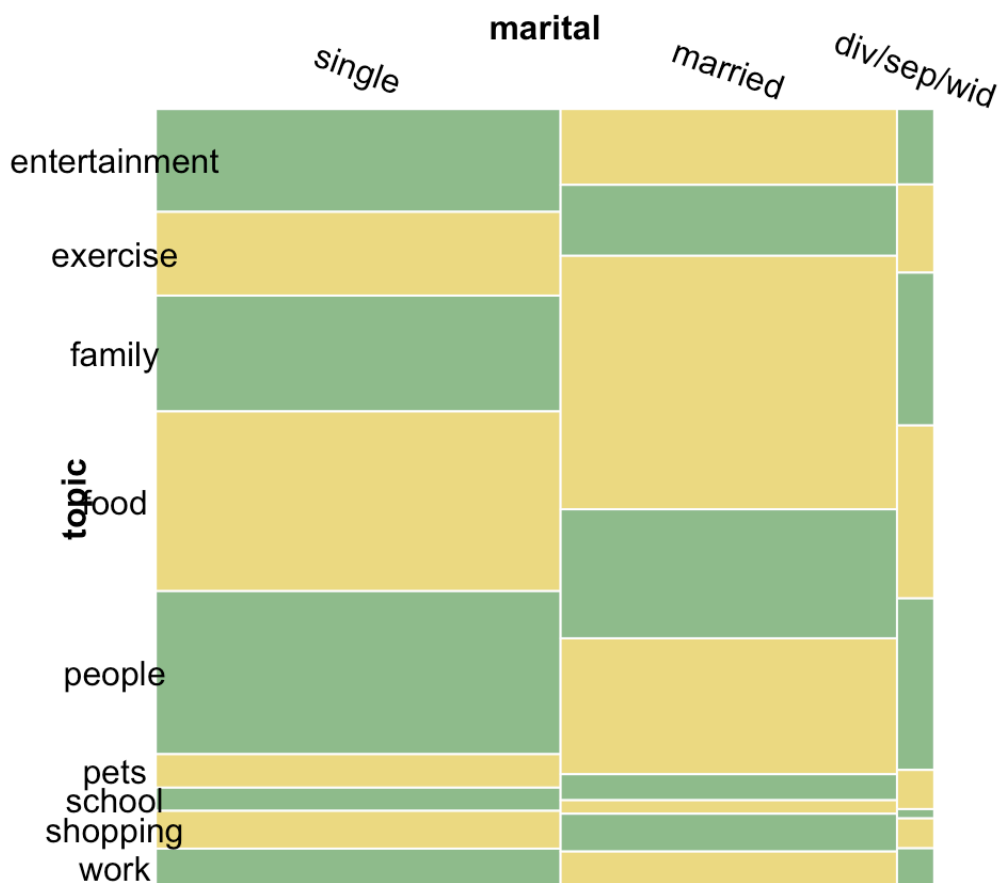## Topic vs Parenthood

```
mosaic(topic~parenthood,direction = c("v", "h"),
       labeling = labeling_border(rot_labels = c(0, 90, 0, 0)),
       gp = gpar(fill = c("chocolate", "thistle3"), col = "black"),
       spacing = spacing_equal(sp = unit(0.2, "lines")),my_df[my_df$parenthood!="",])
```
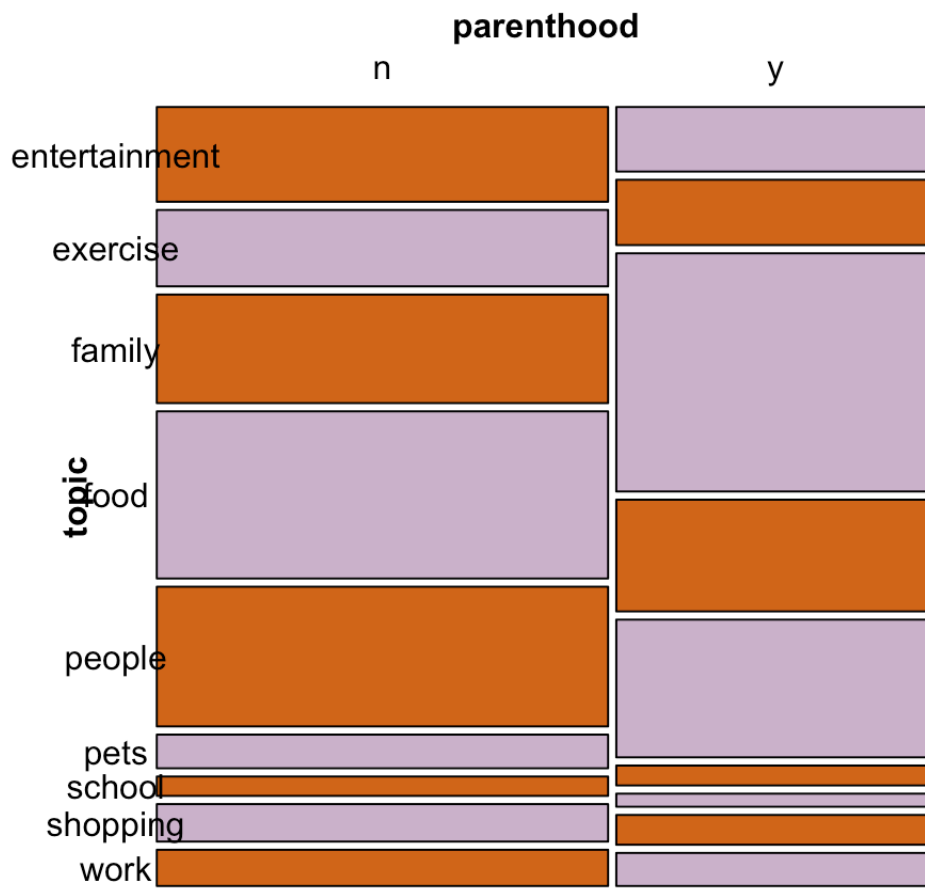
parenthood


```
par(mfrow = c(1,2))
pie_graph(table(my_df[my_df$parenthood=="y",]$topic), "yes")
pie_graph(table(my_df[my_df$parenthood=="n",]$topic), "no")
```
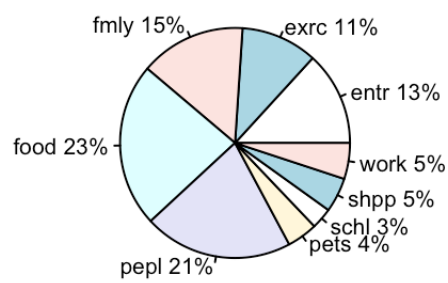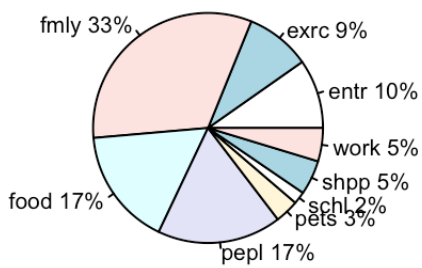
yes   **no**

fmly 33%   exrc 9%   fmly 15%   exrc 11%

entr 9%   entr 13%

work 5%   food 23%   work 5%

shpp 4%   shpp 5%

schl 2%   schl 3%

food 16%   pets 3%   pets 5%

pepl 19%   pepl 20%

Topics vs Marital Status

marital

single   married   div/sep/wid

topic:
entertainment
exercise
family
food
people
pets
school
shopping
work

div/sep/wid   **married**   **single**

fmly 20%   exrc 11%
entr 10%
food 22%   work 5%
            shpp 4%
            schl 1%
pepl 22%   pets 5%

fmly 33%   exrc 9%
            entr 10%
            work 5%
food 17%   shpp 5%
            schl 2%
pepl 17%   pets 3%

fmly 15%   exrc 11%
            entr 13%
food 23%   work 5%
   shpp 5%
   schl 3%
pepl 21%   pets 4%
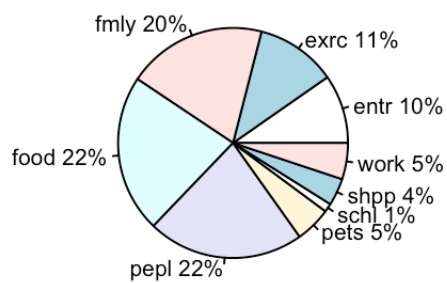
For people who have their own small family (married group and two-parent group), obviously family is the core of their life. Single people often live in a casual style, food, social network and leisure activities weights more in their happiest moments. For divorced/separated/widowed group, proportion of family, food and people are similar in their happy day.

Topic vs Gender

```
mosaic(topic~gender,direction = c("v", "h"),
       labeling = labeling_border(rot_labels = c(0, 90, 0, 0)),
       gp = gpar(fill = c("lightcyan3", "indianred3"), col = "black"),
       spacing = spacing_equal(sp = unit(0.2, "lines")),my_df[my_df$gender!="o" & my_
df$gender!="" ,])
```
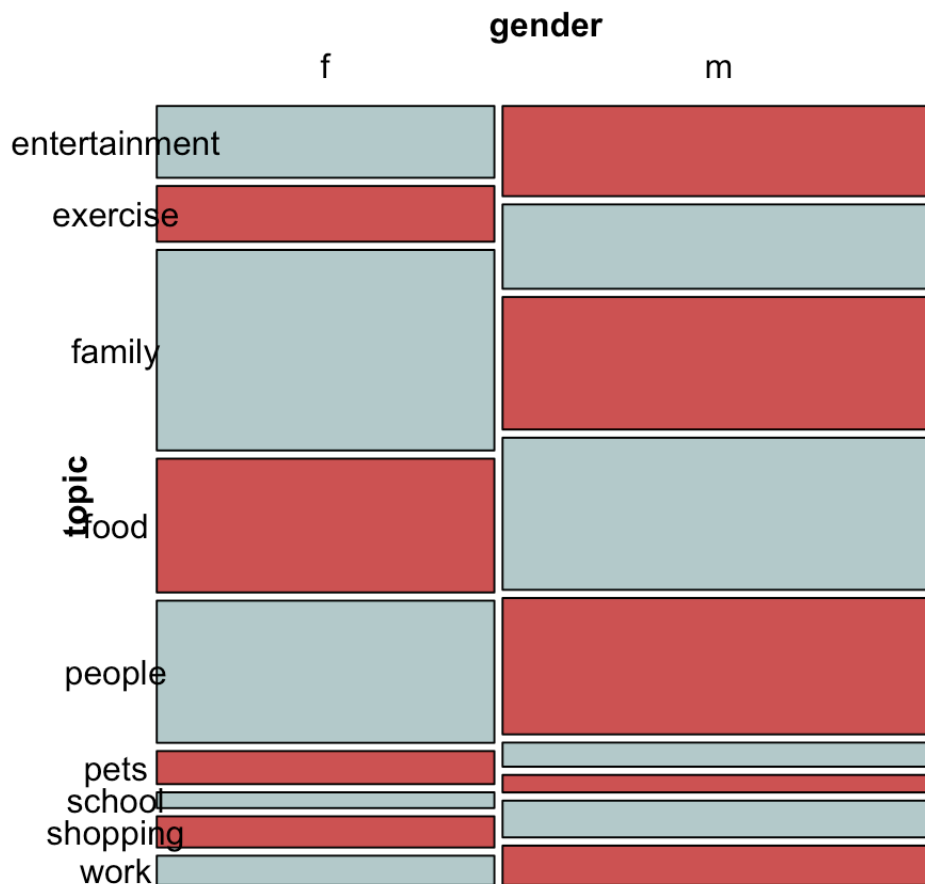
```
par(mfrow = c(1,2))
pie_graph(table(my_df[my_df$gender=="f",]$topic), "female")
pie_graph(table(my_df[my_df$gender=="m",]$topic), "male")
```
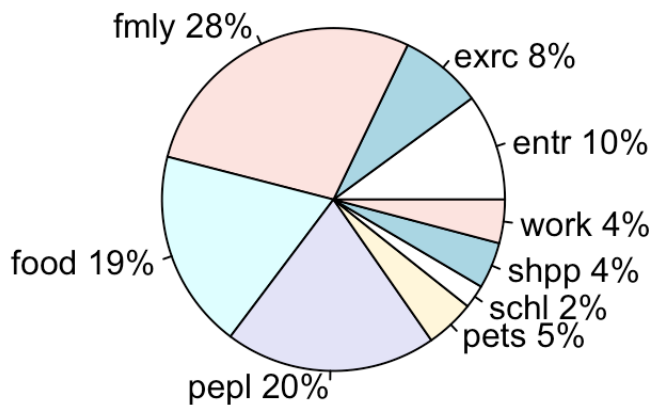
female   **male**
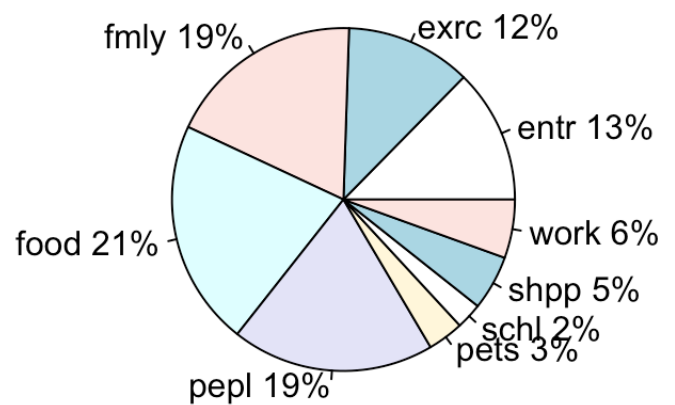
From plots above, evidently, females concentrate more in family life than males do. The percentage of family in females' happy moments is 9% more than that of males'. The 9% was separated to exercise, food, entartainment and work. That is to say, males focus more on themselves than females do, which might due to the different social roles they play.

Summary

By analyzing 100535 happy moments in past 24 hours, there are two insteresting findings: 1. Instead of entertainmens, family is the most consequential factor that make people. It becomes more and more influential as people getting older. 2. Different groups of people do feel happy because of different kind of things. We can make predictions on reason of feeling happy by the characters that a people have or inversely. For example, a married female adult is more likely to gain happiness from family related things; however, if the main topic of a happy moment is "food", the person will have high probability to be a single young male.

References:

DataSet: From HappyDB https://rit-public.github.io/HappyDB/ (https://rit-public.github.io/HappyDB/)

Happy Image: https://www.lifehack.org/570376/15-things-stop-doing-you-want-truly-happy (https://www.lifehack.org/570376/15-things-stop-doing-you-want-truly-happy) http://www.littlestarblog.com/music-monday-24-hours-of-happy/ (http://www.littlestarblog.com/music-monday-24-hours-of-happy/)