

R Notebook

Code

Part1 Data Clean

Code

Code

```
      USA      IND      CAN      VEN      GBR      PHL      MEX
AUS      BRA
0.861015690 0.089499582 0.006127565 0.005013462 0.004456411 0.002970940 0.002042522 0.0012
99786 0.001299786
      NGA
0.001114103
```

Code

```
USA  IND  CAN  VEN  GBR  PHL  MEX  AUS  BRA  NGA
9274 964   66   54   48   32   22   14   14   12
```

American are the majority of this data and compared to American, the samples from other country are too small to represent their population, so here we only analyse people who come from America.

Code

```
NAs introduced by coercion
```

Code

```
[1] 233 227  98  95  88  84
```

Code

age0_20	age20_30	age30_40	age40_50	age50_100
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
0.04285053	0.4776492	0.2933827	0.1022675	0.08385007
1 row				

We can know from the result that the range of normal age is 2-98, and we should also eliminate the age which seems imposible. we seprate the age range into 5 groups, they are 0-20,20-30,30-40,40-50,50-100.

Code

```
  f    m    o
5445 5311   56
```

Code

```

divorced    married    separated    single    widowed
0.052377862 0.405209975 0.009455826 0.526374339 0.006581997

```

Code

```

      n      y
0.5963192 0.4036808

```

Code

We choose the gender: male and female, which are majority of the gender kind. For marital, we use single and married data. For parenthood, there are just yes and no.

Code

```

24h    3m
49746 50646

```

Code

```

achievement    affection    bonding    enjoy_the_moment    exercise
leisure
33897          34164          10726          11109          1196
7457
nature
1843

```

Code

```

achievement    affection    bonding    enjoy_the_moment    exercise
leisure
4268          4810          1750          1506          217
1304
nature
252

```

Code

We can find that the reflection period are almostly equal to each other. It is better for us to use predicted category data because there are no NA in this data set.

Code

Code

```
[1] 73475
```

Code

```
[1] 10844
```

Code

[1] 100392

Code

[1] 1 13839

Code

[1] 132

Code

2	3	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
33	34	35	36	37													
9	3	3	346	779	1092	1912	2419	3147	3414	3849	4419	3977	3964	4630	4004	3252	3455
2554	3012	2483	2224	1966													
38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
56	57	58	59	60													
1708	1283	1180	1289	944	986	918	582	477	456	542	571	419	404	528	384	545	309
343	350	215	213	222													
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	78	83	84
88	95	227	233														
354	192	110	90	108	135	51	81	87	84	27	63	15	81	6	3	63	3
6	3	9	51														

Code

[1] 72

Code

[1] 0

Code

[1] 73343

There are 100392 data in the original data set and there we just use 73343 of them which contain the total information that we are interested in. There are 10844 people are involved in this survey, but the range of their wid is from 1 to 13839.

Code

Code

Code

[1] 416712

Code

friend	time	day	watched	played	finally	found	dinner	home	night
game	family	daughter							
6914	6162	5805	3383	3102	3079	3074	2872	2790	2747
2620	2610	2596							
son	job	nice	favorite	husband	wife	dog			
2516	2496	2369	2349	2298	2215	2140			

Code

```
[1] 398058
```

Code

friend	time	day	watched	found	finally	played	dinner	home	night
daughter	family	son							
6616	5684	5239	3236	3026	2959	2875	2818	2610	2604
2537	2428	2419							
game	nice	favorite	job	husband	wife	received			
2402	2288	2287	2263	2247	2169	2060			

Code

Here we need to clean the text because there are duplicated words in one moment, and the duplicated words will make us count one moment more than once. So, here we need to remove this kind of error. From the result we can find some differences which means this work is necessary.

Code

```
Parsed with column specification:
cols(
  aunt = [31mcol_character()][39m
)
```

Code

```
[1] " auntie " "^auntie " " auntie$"
```

Code

```
[1] 28127
```

Code

```
[1] 21240
```

Code

In this section, family_dict is cleaned to make it more exact to find the moments which contain family members. For example, “mom” can be a beginning of some other words which are not related to family member. However, for word “friend”, even though there are other versions but all of them must be related to friend. In addition, we also get 2 very important vectors here, familyrows and friendrow. We use these two vectors do lots of job later.

American's Happy Moments with Family v.s. Friends

[illegible]

Definitely No!

friend	watched	found	finally	played	dinner	home	night	daughter	family
son	game	nice							
6616	3236	3026	2959	2875	2818	2610	2604	2537	2428
2419	2402	2288							
favorite	job	husband	wife	received	bought	morning			
2287	2263	2247	2169	2060	1986	1960			

[1] 398058

From the word cloud and table we can find that the number of “friend” appeared in the happy moments is greatly larger than the “family”, but is that the truth? People will be happier with friends rather than family? Let’s see what is the truth under the data.

1 “Family” vs “Friend” !

Actually, when we use family_dict to find other words like “mother”, “father”, “kid”, we can get far more “family” which are expressed in other ways. From the data, there are 17941 happy moments contain the words related to “family”! Compared to 9160 happy moments contain “friend”, “family” wins! We love family! Of course, friends are also very important to everyone.

Code

In addition, we need to figure out that there are some happy moments contain both “family” and “friend”. In this section we found there are 943 happy moments with both “family” and “friend”, so the number of happy moments only contain “family” is 16998 and the the number of happy moments only contain “friend” is 8217.

2 Who will be happy with family? vs Who will be happy with friend?

Here, we analyse that if there exists some difference between people who are happy with family and those who are happy with friend. ### (a) How old are you?

Code

```
Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing scale.
```

Code

```
Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing scale.
```

Code

```
Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing scale.
```

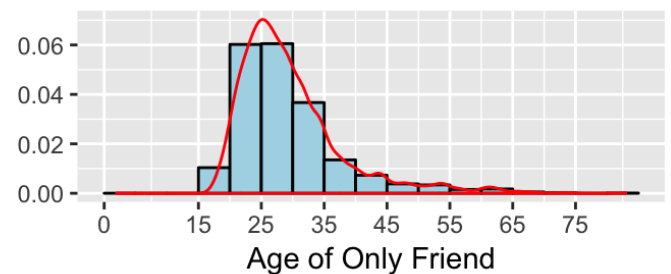
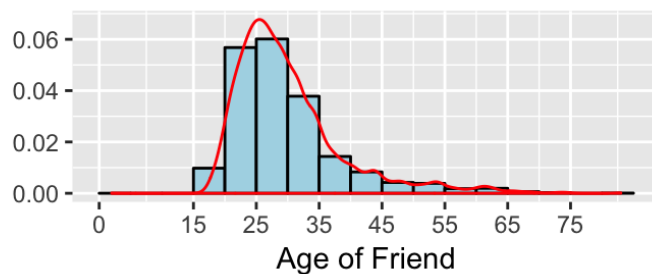
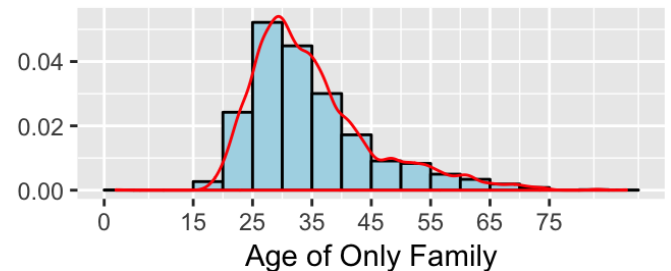
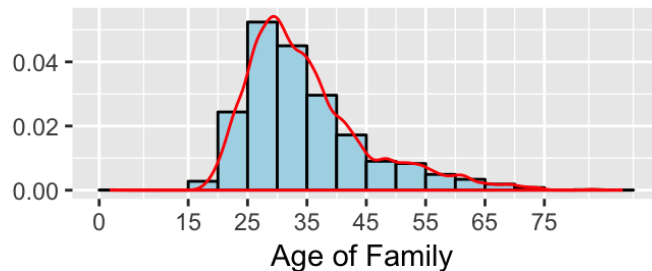
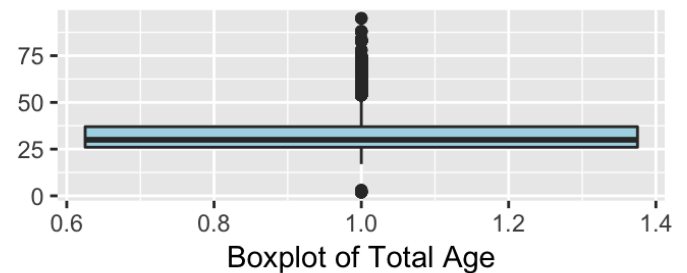
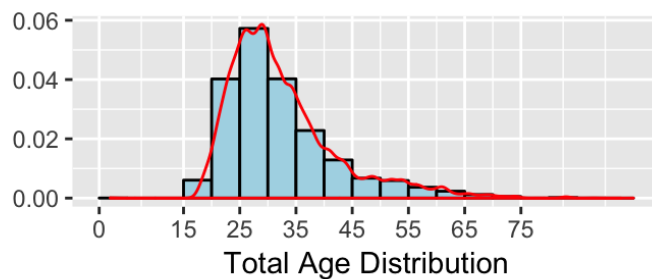
Code

```
Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing scale.
```

Code

```
Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing scale.
```

Code



This set of plots contain both histograms and boxplots. We analyse the difference of people's age who spend their happy moments with their family and who spend thier happy moments with their firends. The first line gives us the total distribution of age of the data. We can find that most people who do this survey are between 20-45. From the boxplot we can find that there are many outliers and the median of the data is around 30. In the second line of this graph. First one plots the distribution of age from those whose happy moments are with their family and the second one only plots the ages of one whose happy moments only contain family. In the third line of this graph. First one plots the distribution of the age from those whose happy moments are with their friends and the other one plots the ages of those whose happy moments only metion friends.

To compare with these set of histograms and density curves, we can find that people who feel happier with their family are older than those are happy with their friends. In addition, we can also find that there are more younger people who only mention friend in their happy moments(compare the 2 graphs in the third line).

Code

```
[1] 32.58691 35.00925 30.10786
```

Code

```
Pairwise comparisons using t tests with pooled SD

data:  X and A

 1      2
2 <2e-16 -
3 <2e-16 <2e-16

P value adjustment method: none
```

[Code](#)

	p.value <dbl>	conf.int <dbl>	estimate <dbl>	p.value.1 <dbl>	conf.int.1 <dbl>	estimate.1 <dbl>
mean of x	0.6256793	-0.1629385	35.06324	0.0005600155	-0.7068639	29.65705
mean of y	0.6256793	0.2709189	35.00925	0.0005600155	-0.1947517	30.10786
2 rows						

To give more specific values, we use t.test to display the data. Firstly, by using multiple-t-test, we find that the means of age from total data, family data, friend data are all difference with each other. Using t-test to analyse the mean of age from family data and only family data, we find that the mean of them are equal can be accept under 95% confidence interval. However, the mean of age from friend data is larger than only friend data.

Conclusion: Younger people are easier to be happy with thier friends!

[Code](#)

gen.family <fctr>	Freq <dbl>	gen.friend <fctr>	Freq.1 <dbl>
f	0.5241625	f	0.4126638
m	0.4758375	m	0.5873362
2 rows			

[Code](#)

mar.family <fctr>	Freq <dbl>	mar.friend <fctr>	Freq.1 <dbl>
married	0.6584917	married	0.2267467
single	0.3415083	single	0.7732533
2 rows			

[Code](#)

par.family <fctr>	Freq <dbl>	par.friend <fctr>	Freq.1 <dbl>
n	0.4001449	n	0.8004367
y	0.5998551	y	0.1995633
2 rows			

[Code](#)

peri.family <fctr>	Freq <dbl>	peri.friend <fctr>	Freq.1 <dbl>
hours_24	0.4507553	hours_24	0.4509825

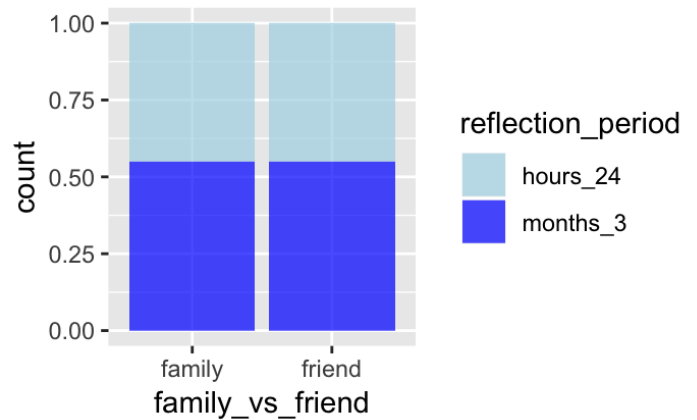
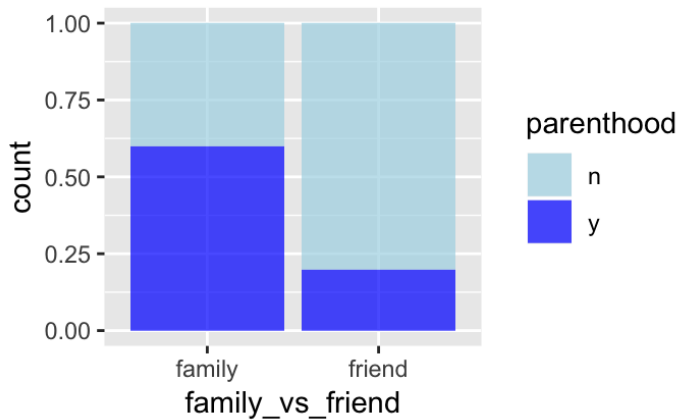
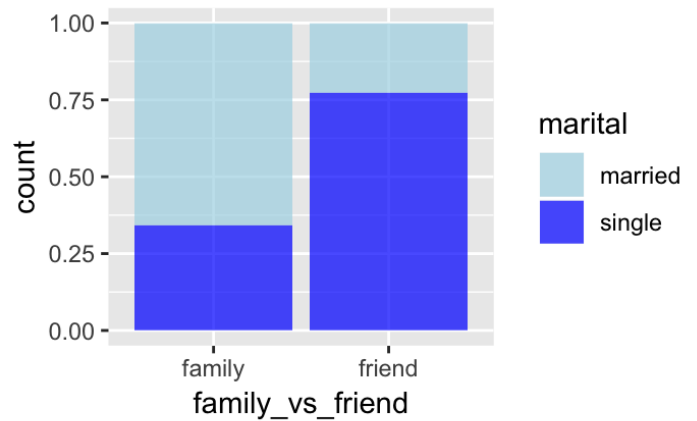
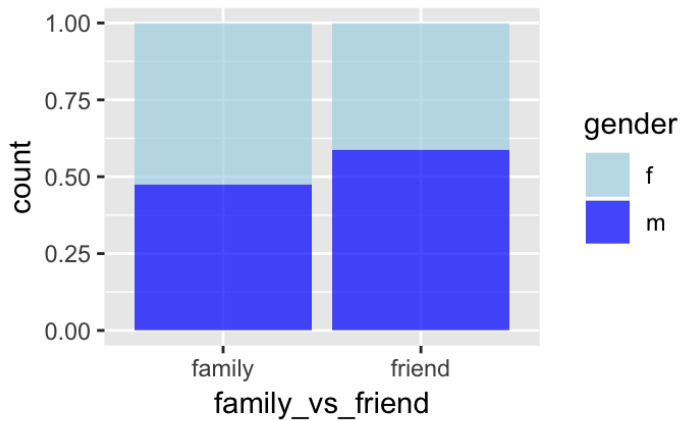
peri.family <fctr>	Freq <dbl>	peri.friend <fctr>	Freq.1 <dbl>
months_3	0.5492447	months_3	0.5490175
2 rows			

Code

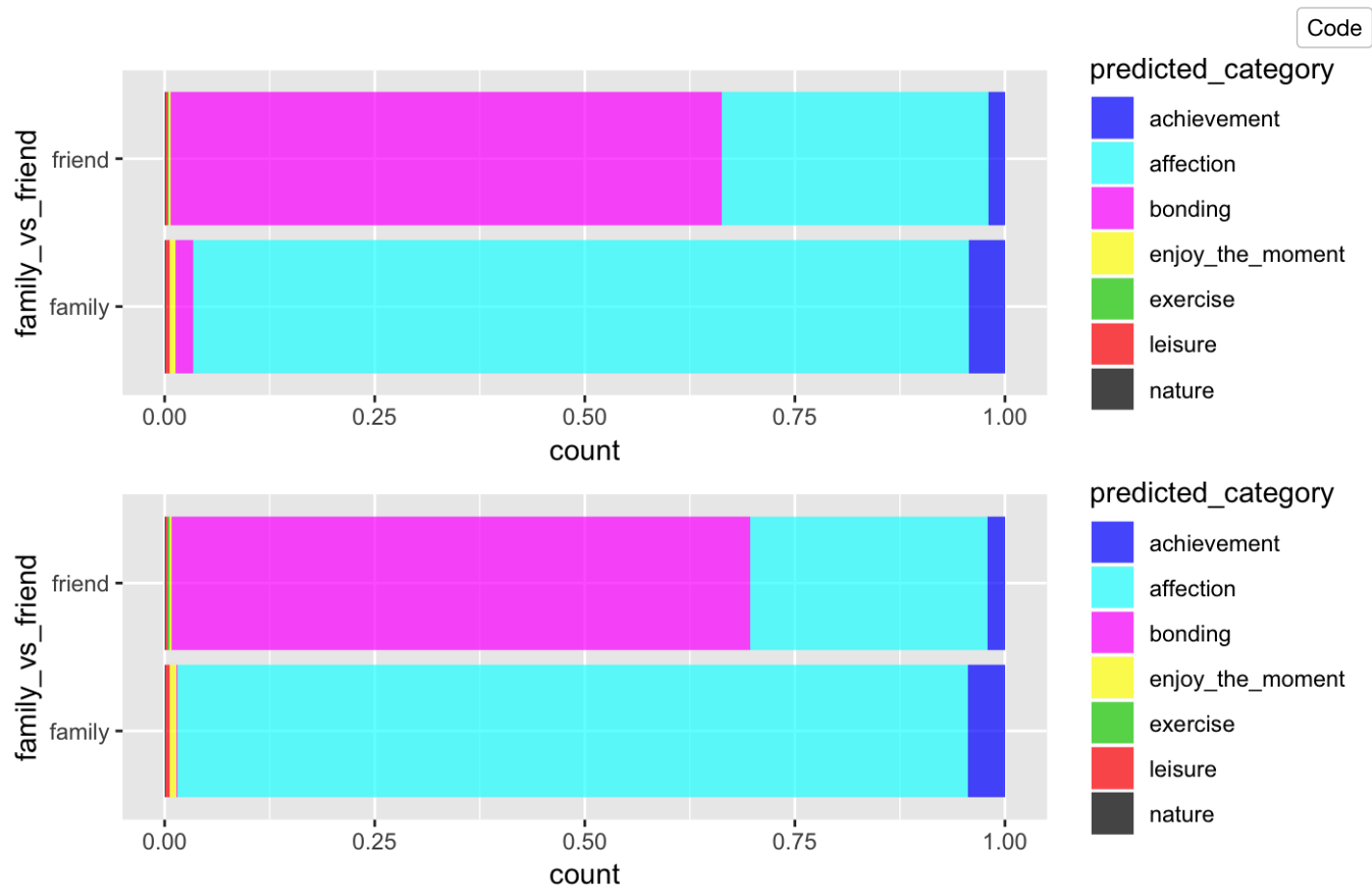
precate.family <fctr>	Freq <dbl>	precate.friend <fctr>	Freq.1 <dbl>
achievement	0.0425282872	achievement	0.019323144
affection	0.9240287609	affection	0.318122271
bonding	0.0200657711	bonding	0.655131004
enjoy_the_moment	0.0073017112	enjoy_the_moment	0.002292576
exercise	0.0002786913	exercise	0.001637555
leisure	0.0041246307	leisure	0.002292576
nature	0.0016721476	nature	0.001200873
7 rows			

Code

Code



From this set of chart we can tell that famale spend more happy moments with their family than the male. Married people's happy moments are more related to family than those who are single. When people have children their happy moments contain more matters about their family. As for the reflection_period, that do not influence the people's happy moments.



In this chart we analyse the different category when they are happy with different people. Obviously, when people spend thier happy moments with friends, they feel more about bonding. When people spend thier happy moments with family, they feel more about affection. In addition, feeling achievement will accurs more when they are happy with their family than with friends. Enjoy the moment and leisure will also more common when they are happy with their family.

The upper graph use total data from family and friend and the lower graph use data from only family and only friend. We can find little difference between them. Feeling bonding is seldom occurs in the happy moments only with family.

3 what activities people do with family? vs What activities people do with friend?

Code

Code

Code

daughter	son	husband	wife	dinner	home	mother	sister	mom	watched
visit	birthday	brother							
2527	2416	2236	2163	1260	1124	1065	978	971	923
908	860	835							
played	love	baby	night	friend	nice	spend			
813	809	807	788	724	647	627			

Code

Code

[Code](#)

girlfriend	boyfriend	dinner	talked	played	night	birthday	visit
game	watched						
1555	1011	567	559	526	468	461	443
422	362						
nice	family	havent	met	fun	event	weekend	home
spend	love						
357	297	295	293	281	275	272	264
257	253						

We can find some common point from these graphs that is people is the most important points and then have dinner together! That sounds good! It is happy moment!

[Code](#)

```
Parsed with column specification:
cols(
  movie = [31mcol_character()][39m
)
```

[Code](#)

```
Parsed with column specification:
cols(
  `10k` = [31mcol_character()][39m
)
```

[Code](#)

```
Parsed with column specification:
cols(
  `Applebee's` = [31mcol_character()][39m
)
```

[Code](#)

```
Parsed with column specification:
cols(
  aunt = [31mcol_character()][39m
)
```

[Code](#)

```
Parsed with column specification:
cols(
  `26mon` = [31mcol_character()][39m
)
```

[Code](#)

```
Parsed with column specification:
cols(
  school = [31mcol_character()][39m
)
```

[Code](#)

```
Parsed with column specification:
cols(
  acquired = [31mcol_character()][39m
)
```

[Code](#)

```
Parsed with column specification:
cols(
  CEO = [31mcol_character()][39m
)
```

[Code](#)[Code](#)[Code](#)

	entertainment <dbl>	exercise <dbl>	food <dbl>	pets <dbl>	school <dbl>	shopping <dbl>	work <dbl>
family	909	1227	3543	482	709	977	730
friend	577	700	1952	238	283	357	320

2 rows

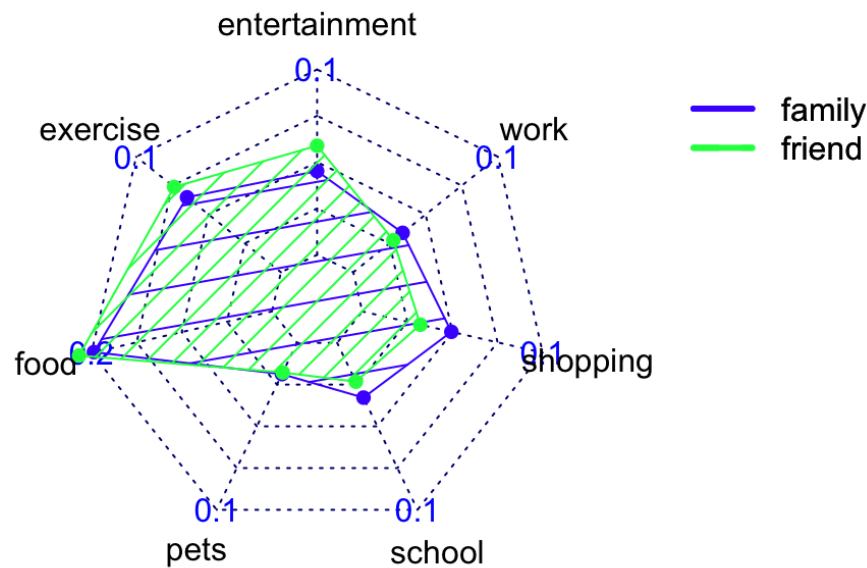
[Code](#)

	entertainment <dbl>	exercise <dbl>	food <dbl>	pets <dbl>	school <dbl>	shopping <dbl>	work <dbl>
family	0.05066607	0.06839084	0.1974806	0.02686584	0.03951842	0.05445627	0.04068892
friend	0.06299127	0.07641921	0.2131004	0.02598253	0.03089520	0.03897380	0.03493450

2 rows

[Code](#)

activities with family vs friend



In this part, we are going to find out what people prefer do with different people(“family” vs “friend”). Using the percentage number and the radar plot, we can find that with family and friends people like do something about food, maybe having food or making food. Then people prefer to do exercise with their family and friend. The least things people do with family and friend are pets and work and that make scense, because these things usually can be done by self.

To compare the different activities with family and firend, the radar plot also shows us that when people have fun with their friend the term of “food”,“exercise”,“enterainment” take more percentage than when they with their family. On the other hand, when people spending their happy time with their family, they are prefer to go shopping or talk about school or work, they also so spend more happy moments with family on “pets” than with friend.

4 Who are Family?!

Code

```
Parsed with column specification:
cols(
  fiancée = [31mcol_character()][39m,
  children = [31mcol_character()][39m,
  brother = [31mcol_character()][39m,
  aunt = [31mcol_character()][39m
)
```

Code

total.length	total.length.1	total.length.2	total.length.3
<int>	<int>	<int>	<int>

total.length	total.length.1	total.length.2	total.length.3
<int>	<int>	<int>	<int>
4676	9568	4128	1464

1 row

Code

summember.row	summember.row.1	summember.row.2	summember.row.3
<int>	<int>	<int>	<int>
4665	8611	3984	1378

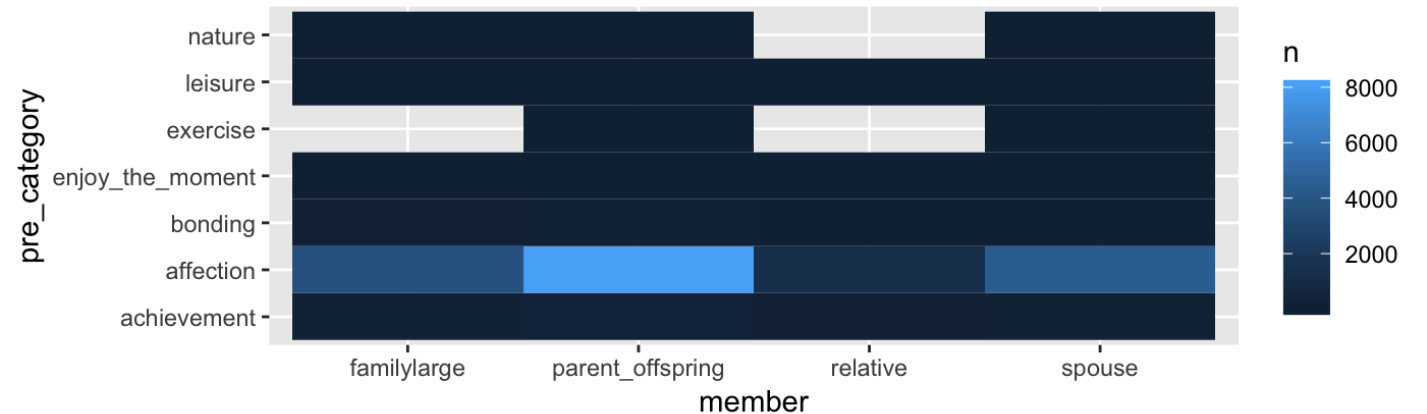
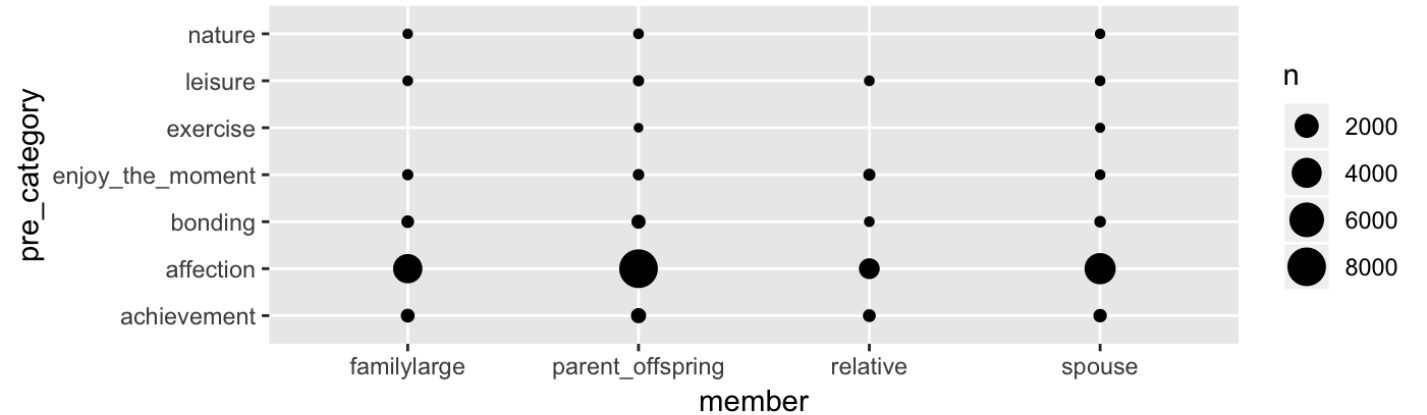
1 row

In this section, we go futher studies on “family”. We categorize family into four groups by using family_dict. The first group of family is spouse; the second one is parent_offspring; the third one is familylarge which contain people like “brother”, “sister”, “grandmother”, “grandfather”, those peoply may live together; the fourth one is relatives, like “cousin”, “aunt”.

Here our job is to find difference of category of happy moments that people will have on different family member they are with. From earlier work, we know that people will more likely feel affection and achievement with their family compared with friend. Now we are going to figer out what knids of group of family will give us what kind of feelings.

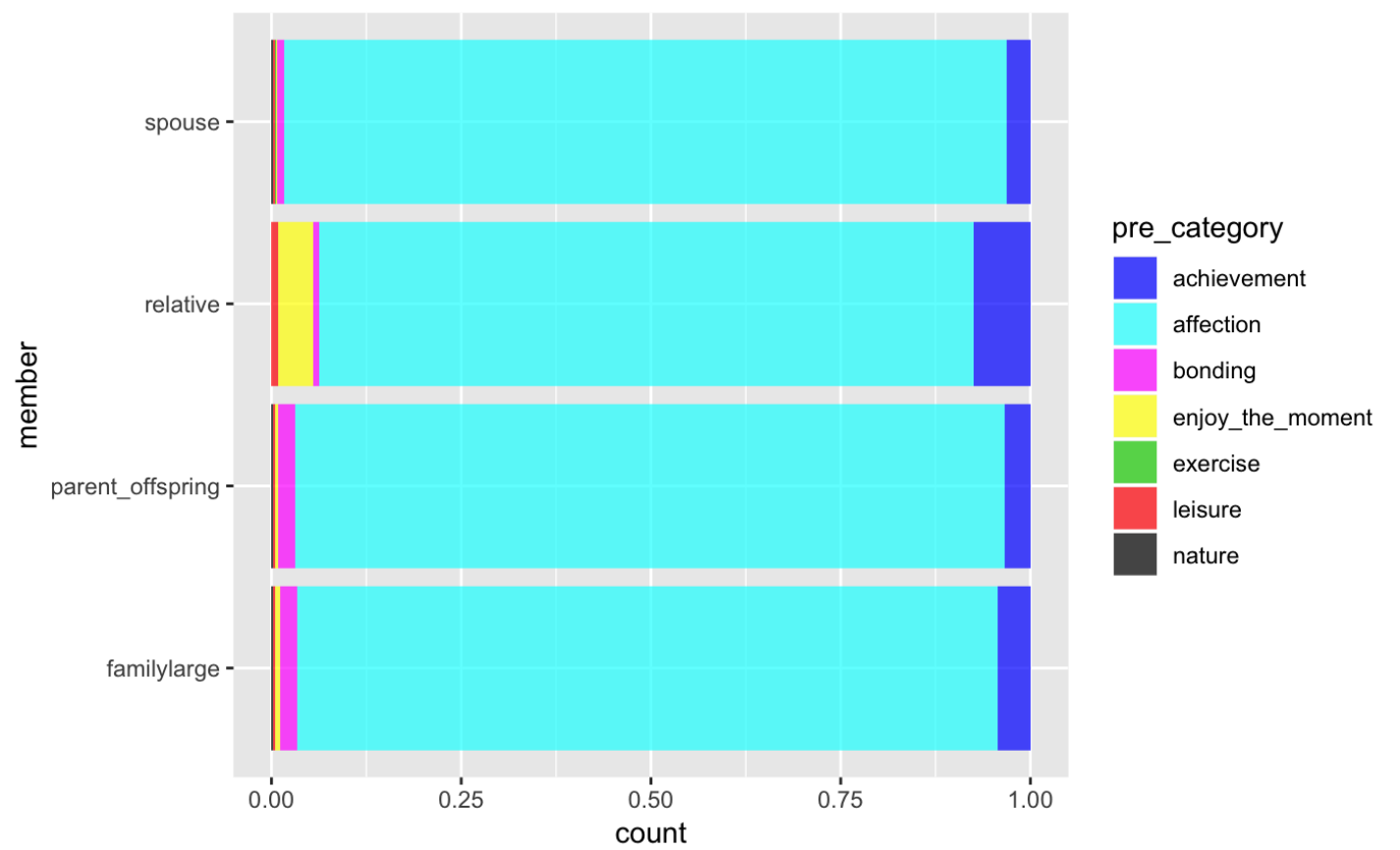
Code

Code



This pair of graph using absolute value shows us that people will strongly feel affection if they are parent_offspring relationship compared with other relationship. The secondary relationship give us most affectino feeling is spouse. For the achievement feeling, we can have the similar conclusion that parent_offspring is the first one. Compared with other relationship parent_offspring may contain more categories cause we can find in the lower graph the parent_offspring columnis brighter than others. There are also some blank in the relationship of familylarge and relatives. Not suprising, this kind of family are less contacted with when compared to parent_offspring and spouse.

Code



From this plot, even though we are analysing the same question but it shows us different information by using percentage value. With percentage value, we can excavate more other categories. We can find that when people spend happy moments with their relatives, they can be engaged in more categories like “affection”, “achievement”, “enjoy_the_moment”, “leisure”, “bonding”. And it is obviously that people like feeling achievement and enjoy_the_moment also leisure with their relatives than other group of family. Feeling achievement also occurs in “familylarge”. “Bonding” are more likely to appear when they spend time with their “parent_offspring” and “familylarge”. “Spouse” seldom have happy moments about “enjoy_the_moment”, and relatives seldom have happy moments about “nature”.

Above all we can conclude that different groups of family play different functions in our happy moments. However, the most of category in family’s happy moments is affection, which is totally different from the category when people have happy moments with their friends—“bonding”. Bonding and affection are the top 2 categories when people’s happy moments are related to family and friends.