Donor Insights : using factor analysis

Author: Simon Liu

Background: an organization wants to know if there are some relationships in characteristics of potential donors. They could explore these variables and see is there any pattern.

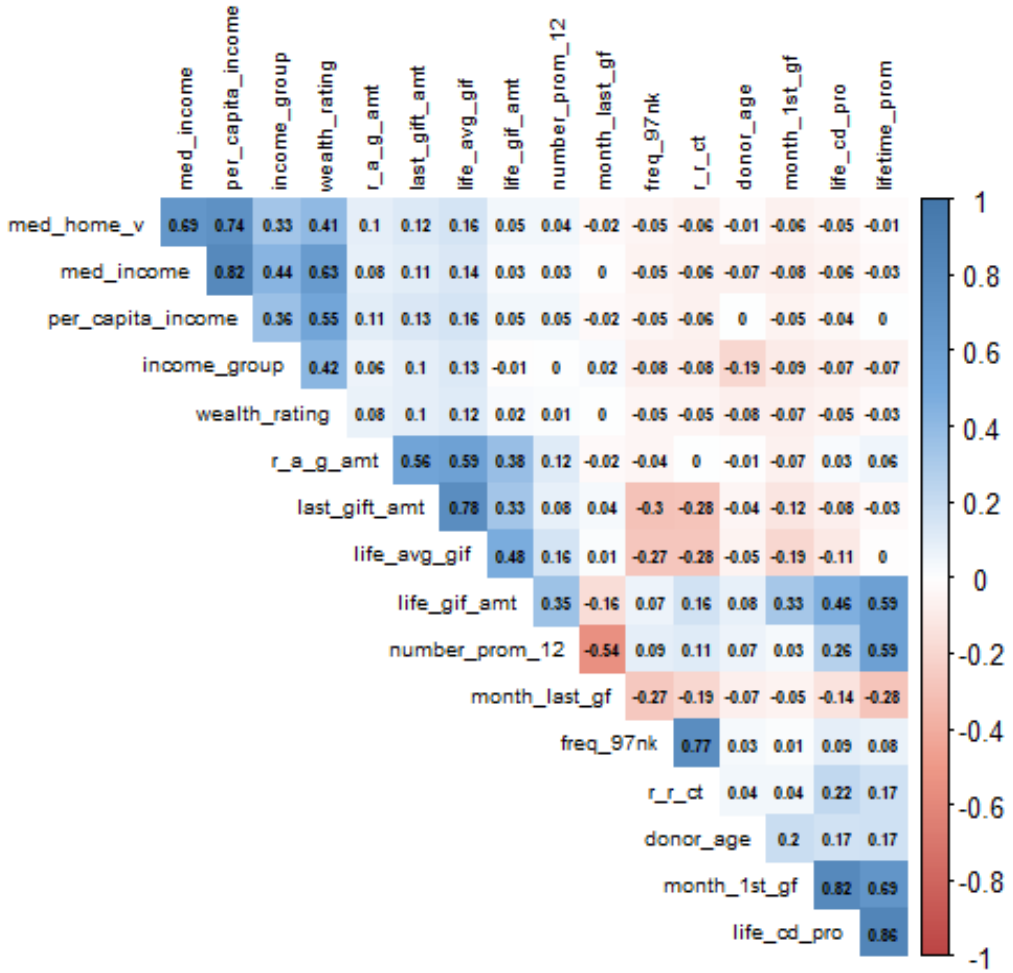
- Variable description: all patrons' data is from their database. Just to list a few:
  - donor_age
  - income_group: from 1 to 7.
  - last_gift_amt: the amount donor gave last time.
  - lifetime_gift_amount: the total amount a donor gave in life time.
  - lifetime_prom: the number of solicitations ever sent.
  - median_home_value
  - median_household_income
  - per_capita_income: census data.
  - months_since_first_gift
  - number_prom_12: the number of promotions in the last 12 months.
  - wealth_rating: from 1 to 9.

Methodology: consider doing a factor analysis to all variables and see if there is any latent factors among donors, using R package stats, psych().

- Before doing factor analysis, needs to see if these variables have some linear relationships. (in other word, if the correlation matrix is different from an identity matrix)
  - if variables have no linear relationships then factor analysis is not appropriate.

- We could use correlation coefficient matrix to get a general idea and Bartlett's test to verify.
  - use psych() for Bartlett's test.
  - use corrplot() for correlation matrix plot.

Correlation Coefficient Matrix



- Correlation coefficients show that there are some positive correlation between:
  - home value and household income.
  - recent avg card gift amount and life gift amount.
  - last gift amount and life avg gift amount.
  - count of donations and count of response to promotions.
- Some negative correlation between:
  - count of promotion in last 12 months and months since last gift donation.
  - last gift amount and count of response to promotion.
- Infers the more frequent you contribute, the less your contribution amount is!

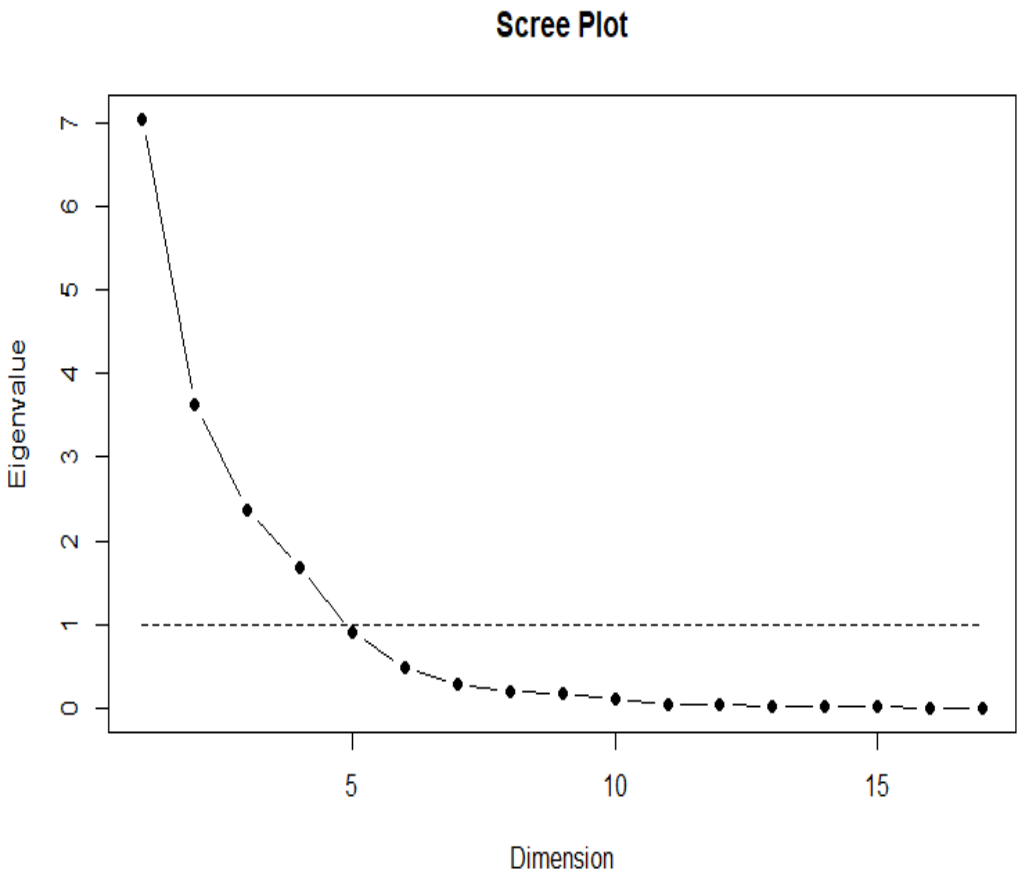
Bartlett's Test

- Test shows that the null hypothesis is rejected at 0.05 level. Therefore, the correlation coefficients matrix is significantly differ from an identify matrix. We could go to next step.

```
> cortest.bartlett(cor(dona03),n=1000)
$chisq
[1] 10553.27

$p.value
[1] 0
```

Determine the number of factors using scree plot.



- Scree plot shows need to keep 4 factors (eigenvalue >1)

Extracting factors: factanal()

- Set the number of factors to be 4 and using factor rotation to make explanation easier.

```
factor_num <-4
fit <- factanal(dona03,factor_num,scores=c("regression"),
                rotation="varimax",lower=0.1)
```

- Use varimax rotation method to make the explanation simpler.

- Factor result is on the next page.

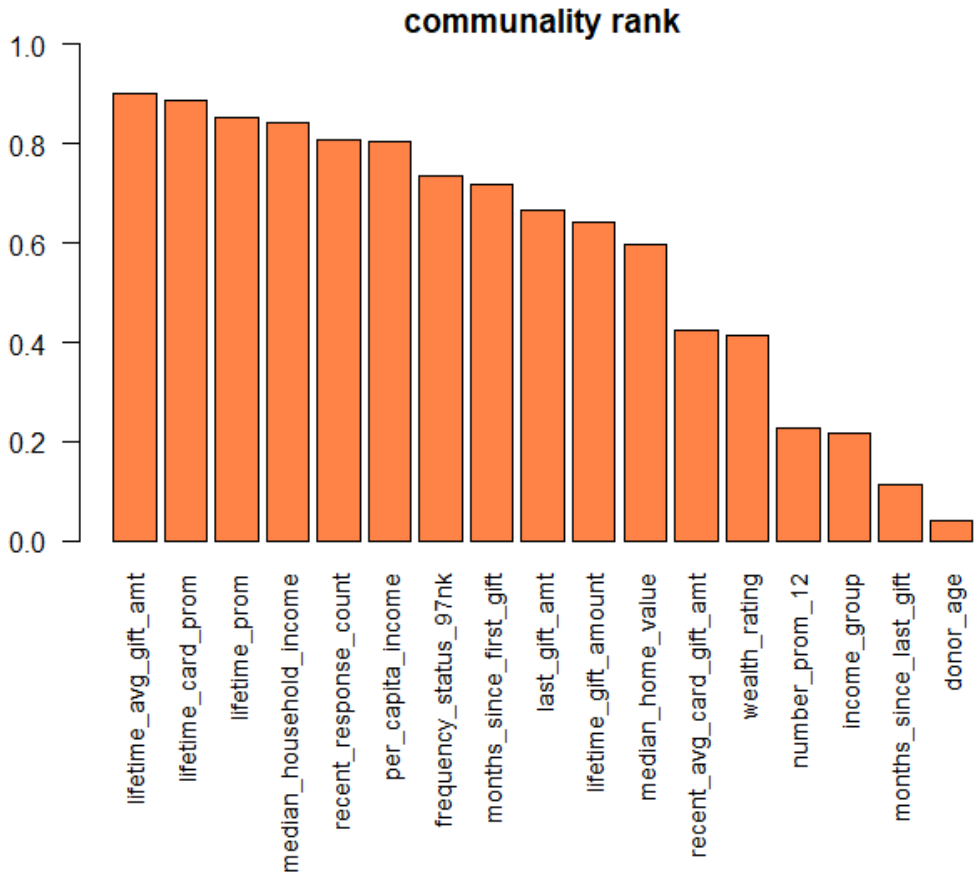
Result discussion: 4 factors explained about 60% of variance!

```
Loadings:
                        Factor1 Factor2 Factor3 Factor4
median_home_value        0.77
median_household_income  0.92
per_capita_income        0.89
wealth_rating            0.64
lifetime_card_prom               0.94
lifetime_prom                    0.91
months_since_first_gift          0.83
last_gift_amt                            0.78
lifetime_avg_gift_amt                    0.92
lifetime_gift_amount             0.51    0.59
recent_avg_card_gift_amt                 0.64
frequency_status_97nk                                    0.84
recent_response_count                                    0.88
donor_age
income_group             0.45
months_since_last_gift
number_prom_12                   0.36

                Factor1 Factor2 Factor3 Factor4
SS loadings        2.89    2.87    2.37    1.76
Proportion Var     0.17    0.17    0.14    0.10
Cumulative Var     0.17    0.34    0.48    0.58
```

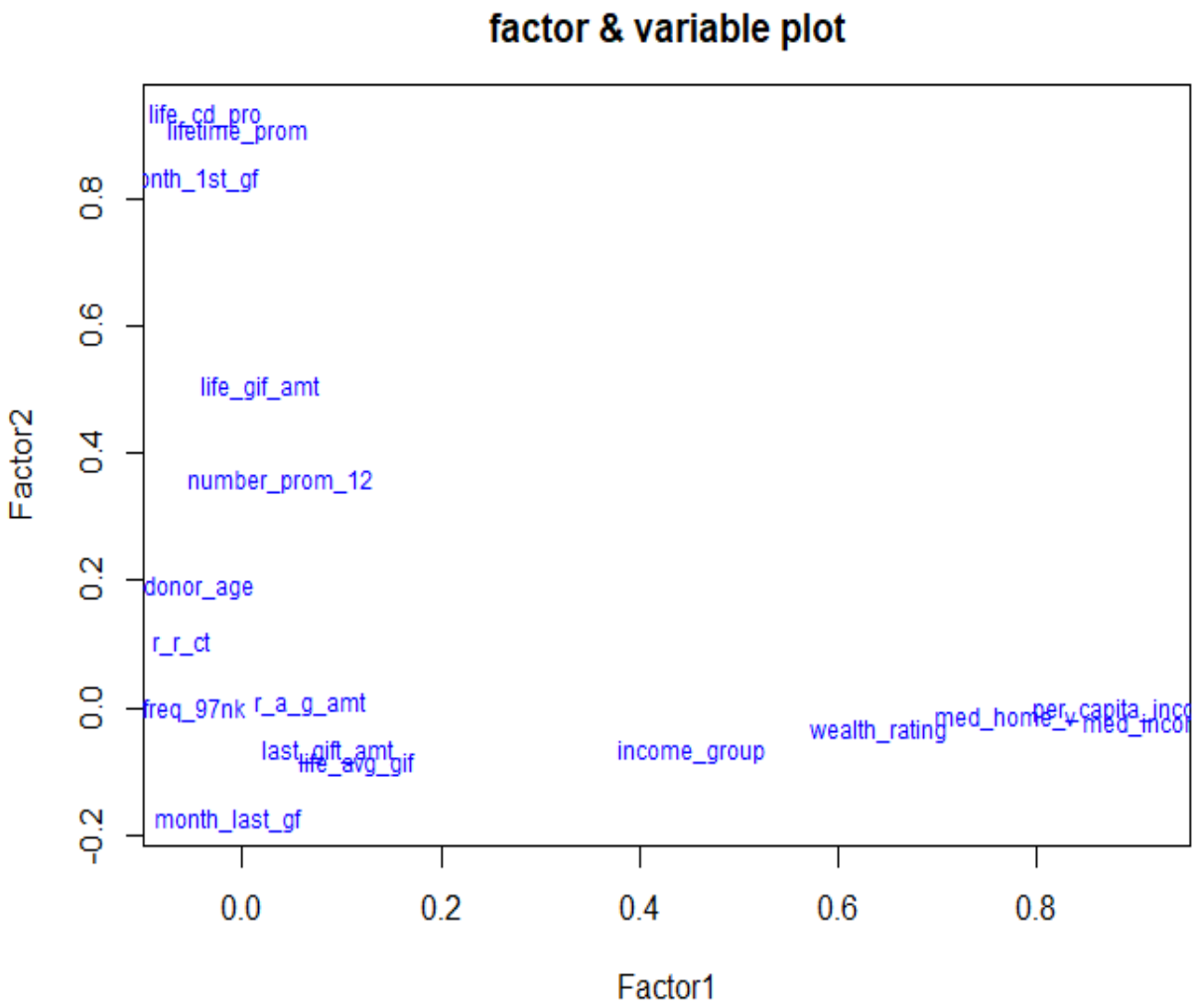
- Factor explanation:
  - 1st factor is related to donor's wealth: income, home value etc.
  - 2nd factor relates to historic participation of promotion events.
  - 3rd factor measures historic gift amounts contributed by the donor.
  - 4th factor measures response to historical promotion and donations.
- Factor cumulative :
  - The total variance explained by 4 factor model is 58%.

Result discussion: ranking of communalities



- Rank of communalities of variables.
  - The higher the bar, the better the variable is explained by factors. The lower the bar, the harder variable can be explained by factors.
  - Age and month since last gift are not easily explained by our 4 factors.
  - Life time gift amount, life time card promotion, life time promotion, and median household income are better represented by factors.

Factor1(wealth of donor) vs Factor2 (past involvement in promotion) plot



- Plot F1(wealth) vs F2(past promotion amt)
  - Some variables reflect wealth, others measure past involve in promotions.
  - Note recent gift amount does not have higher loading on either F1 or F2, meaning recent gift amt does not indicate wealth or past participation.
  - Also, life gift amt does not load high on F1(wealth), indicating there is no strong relationship between wealth and total donation.