

Personality-analysis

2025-06-25

Khai báo Thư viện

```
library(reshape2)
library(ggplot2)
```

Đọc dữ liệu

```
data <- read.csv("E:/Học Tập/Đại Học/Năm 4 HK 2 2024-2025/thống kê nhiều chiều/thong_ke_chieu_chieu/PRO
```

Thông tin tổng quan của bộ dữ liệu

```
head(data, 10)
```

```
##      personality_type social_energy alone_time_preference talkativeness
## 1      Extrovert      6.794295          3.854670          8.725446
## 2      Ambivert      6.378988          5.731157          7.029529
## 3      Ambivert      7.459421          6.322263          3.922269
## 4      Extrovert      6.159626          3.097837          6.019093
## 5      Introvert      5.568462          6.986722          3.913240
## 6      Introvert      2.807173          6.326097          2.770085
## 7      Introvert      1.537468          7.605496          4.294666
## 8      Ambivert      6.636632          4.350586          7.131511
## 9      Extrovert      7.330318          2.820421          5.224400
## 10     Ambivert      5.377395          4.762184          4.244447
##      deep_reflection group_comfort party_liking listening_skill empathy
## 1      2.515151      7.097368      8.5887616      6.774799 6.430132
## 2      7.274493      4.111199      3.2582475      5.550909 3.958179
## 3      4.622261      5.343276      7.4521522      9.483990 6.127654
## 4      1.965440      7.837140     10.0000000      9.436733 8.949684
## 5      9.926161      1.650483      0.3622982      7.470387 6.756837
## 6      5.617670      4.752116      0.3026297      7.137657 5.399156
## 7      9.166947      3.085335      1.0428744     10.000000 5.899131
## 8      5.455061      2.530413      3.0816555      5.177344 5.692996
## 9      5.053459      7.969032      9.1069343      6.717555 7.061661
## 10     6.399885      4.781680      6.0745921     10.000000 5.601565
##      creativity organization leadership risk_taking public_speaking_comfort
## 1      6.142968      0.000000      8.486191      6.888588          8.618033
## 2      6.149457      4.530322      6.214742      4.877224          5.018526
## 3      7.032017      7.092078      3.034192      4.312376          6.474164
## 4      8.923875      5.232694     10.000000      8.203100          7.269755
## 5      9.507803      6.378051      4.452973      2.456879          2.779744
## 6      4.564805      3.360418      2.977163      4.983358          2.789303
## 7      5.423302      8.319798      2.759467      2.067699          2.285145
## 8      3.951603      9.236595      5.938200      7.731858          4.759397
## 9      7.591541      7.961516      3.016811      6.048440          7.021811
## 10     4.927993      5.285973      2.003761      4.414151          4.616793
##      curiosity routine_preference excitement_seeking friendliness
```

```
## 1 5.514367 4.247436 8.916668 8.370776
## 2 6.170892 6.304219 6.494122 9.127521
## 3 6.179085 4.930449 4.054730 7.602897
## 4 5.477594 2.284054 9.442942 10.000000
## 5 5.138220 7.043162 4.149461 4.628954
## 6 6.383135 9.444365 4.377930 5.771033
## 7 6.160209 5.133382 2.708449 6.213928
## 8 6.275260 4.118357 7.183344 8.332522
## 9 8.816983 3.661762 6.790167 6.105775
## 10 7.698937 5.413746 5.351088 10.000000
## emotional_stability planning spontaneity adventurousness reading_habit
## 1 4.919931 4.426422 4.853313 8.257134 5.270555
## 2 4.585075 4.604019 6.067201 6.289347 5.753165
## 3 5.651975 6.269467 5.524244 9.238784 5.250405
## 4 6.622884 5.566069 4.327018 8.489791 5.312617
## 5 5.283958 4.711944 5.187689 3.167217 7.060235
## 6 5.781525 5.464827 3.888344 1.868406 9.188567
## 7 6.950514 6.822373 7.906636 3.173683 9.458013
## 8 5.006115 6.244406 3.643867 7.601721 5.803109
## 9 4.966374 5.620067 4.459566 6.996722 5.980063
## 10 5.107616 4.530871 7.586472 5.513111 5.950501
## sports_interest online_social_usage travel_desire gadget_usage
## 1 10.000000 9.154296 4.816422 9.191711
## 2 5.334303 4.683781 4.725666 5.956141
## 3 3.153540 5.000338 6.139166 6.033048
## 4 8.379936 7.601946 6.370056 5.410145
## 5 2.333388 7.771569 5.534336 5.704598
## 6 4.817574 3.427446 4.674632 4.027299
## 7 2.272047 5.218917 2.745851 4.056359
## 8 5.024010 7.029383 6.990230 7.147859
## 9 9.638288 5.509409 7.513645 8.788241
## 10 2.328848 5.644438 6.251067 3.539273
## work_style_collaborative decision_speed stress_handling
## 1 8.313590 8.032376 7.176905
## 2 5.890619 3.158988 3.423577
## 3 5.807500 4.571003 5.647480
## 4 6.671781 6.600233 5.870088
## 5 5.832968 5.813099 3.758084
## 6 3.048724 2.774809 4.560962
## 7 5.695608 3.230287 6.761266
## 8 4.680890 5.107637 7.246737
## 9 9.300327 8.086807 5.784908
## 10 5.698559 4.079389 6.609341
```

```
# Kiểm tra số chiều của dữ liệu
dim(data)
```

```
## [1] 20000 30
```

Bộ dữ liệu tổng hợp này được thiết kế để mô phỏng các loại tính cách của con người — Hướng nội, Hướng ngoại và cả người có cả hướng ngoại lẫn hướng nội — dựa trên nhiều đặc điểm hành vi và tâm lý khác nhau.

Bộ dữ liệu này chứa 20.000 mục nhập và 30 cột, bao gồm 29 đặc điểm số biểu thị các chỉ số tính cách và 1 cột nhãn (personality_type).

Trong đó

- `personality_type`: biến mục tiêu gồm 3 kiểu tính cách: Introvert(hướng nội), Extrovert(hướng ngoại), or Ambivert (hỗn hợp).
- `social_energy`: Xu hướng đón nhận năng lượng từ tương tác xã hội (0–10).
- `alone_time_preference`: thể hiện sự thoải mái với sự cô đơn.
- `talkativeness`: Xu hướng tham gia và các cuộc trò chuyện.
- `deep_reflection`: Tần suất suy nghĩ sâu hoặc suy nghĩ nội tâm.
- `group_comfort`: Dễ hoà nhập với các nhóm/ hội.
- `party_liking`: Sự thích thú đối với các bữa tiệc và các sự kiện.
- `listening_skill`: Khả năng lắng nghe
- `empathy`: khả năng đồng cảm hay thấu hiểu cảm xúc của người khác.
- `creativity`: sự tư duy sáng tạo.
- `organization`: Mức độ ưu tiên cho các trật tự, kế hoạch.
- `leadership`: Sự thoải mái khi làm lãnh đạo.
- `risk_taking`: Sự chấp nhận rủi ro.
- `public_speaking_comfort`: Mức độ thoải mái khi nói chuyện trước đám đông hoặc công chúng.
- `curiosity`: Mức độ thích học tập và khám phá những cái mới.
- `routine_preference`: Thể hiện thói quen và sự tự phát.
- `excitement_seeking`: Mức độ mong muốn những trải nghiệm mới mẻ.
- `friendliness`: Mức độ thân thiện.
- `emotional_stability`: Khả năng kiểm soát và cân bằng khi căng thẳng.
- `planning`: Xu hướng lập kế hoạch từ trước.
- `spontaneity`: Hành động tự phát không có kế hoạch từ trước.
- `adventurousness`: Sẵn sàng thử những hoạt động mới và mạo hiểm.
- `reading_habit`: Tần suất đọc sách báo.
- `sports_interest`: Mức độ quan tâm tới thể thao hoặc các hoạt động thể chất.
- `online_social_usage`: Thời gian dành cho mạng xã hội.
- `travel_desire`: Sở thích du lịch và khám phá những địa điểm mới.
- `gadget_usage`: Tần suất sử dụng các thiết bị công nghệ và tiện ích.
- `work_style_collaborative`: mức độ làm việc nhóm so với việc làm cá nhân.
- `decision_speed`: Thời gian đưa ra các quyết định.
- `stress_handling`: Khả năng quản lý căng thẳng cách hiệu quả.

Khảo sát dữ liệu

```
str(data)
```

```
## 'data.frame':   20000 obs. of  30 variables:
## $ personality_type      : Factor w/ 3 levels "Ambivert","Extrovert",...: 2 1 1 2 3 3 3 1 2 1 ...
## $ social_energy         : num  6.79 6.38 7.46 6.16 5.57 ...
## $ alone_time_preference : num  3.85 5.73 6.32 3.1 6.99 ...
## $ talkativeness         : num  8.73 7.03 3.92 6.02 3.91 ...
```

```
## $ deep_reflection      : num  2.52 7.27 4.62 1.97 9.93 ...
## $ group_comfort       : num  7.1 4.11 5.34 7.84 1.65 ...
## $ party_liking        : num  8.589 3.258 7.452 10 0.362 ...
## $ listening_skill     : num  6.77 5.55 9.48 9.44 7.47 ...
## $ empathy             : num  6.43 3.96 6.13 8.95 6.76 ...
## $ creativity          : num  6.14 6.15 7.03 8.92 9.51 ...
## $ organization        : num  0 4.53 7.09 5.23 6.38 ...
## $ leadership          : num  8.49 6.21 3.03 10 4.45 ...
## $ risk_taking         : num  6.89 4.88 4.31 8.2 2.46 ...
## $ public_speaking_comfort : num  8.62 5.02 6.47 7.27 2.78 ...
## $ curiosity           : num  5.51 6.17 6.18 5.48 5.14 ...
## $ routine_preference   : num  4.25 6.3 4.93 2.28 7.04 ...
## $ excitement_seeking  : num  8.92 6.49 4.05 9.44 4.15 ...
## $ friendliness        : num  8.37 9.13 7.6 10 4.63 ...
## $ emotional_stability  : num  4.92 4.59 5.65 6.62 5.28 ...
## $ planning            : num  4.43 4.6 6.27 5.57 4.71 ...
## $ spontaneity         : num  4.85 6.07 5.52 4.33 5.19 ...
## $ adventurousness     : num  8.26 6.29 9.24 8.49 3.17 ...
## $ reading_habit       : num  5.27 5.75 5.25 5.31 7.06 ...
## $ sports_interest     : num  10 5.33 3.15 8.38 2.33 ...
## $ online_social_usage  : num  9.15 4.68 5 7.6 7.77 ...
## $ travel_desire       : num  4.82 4.73 6.14 6.37 5.53 ...
## $ gadget_usage        : num  9.19 5.96 6.03 5.41 5.7 ...
## $ work_style_collaborative: num  8.31 5.89 5.81 6.67 5.83 ...
## $ decision_speed      : num  8.03 3.16 4.57 6.6 5.81 ...
## $ stress_handling     : num  7.18 3.42 5.65 5.87 3.76 ...
```

Nhận xét: Các biến đều có giá trị numeric. Chỉ có biến “personality_type” là biến mục tiêu.

```
summary(data)
```

```
## personality_type social_energy alone_time_preference talkativeness
## Ambivert :6573 Min. : 0.000 Min. : 0.000 Min. : 0.000
## Extrovert:6857 1st Qu.: 3.591 1st Qu.: 3.510 1st Qu.: 3.596
## Introvert:6570 Median : 5.541 Median : 5.464 Median : 5.575
## Mean : 5.509 Mean : 5.455 Mean : 5.526
## 3rd Qu.: 7.424 3rd Qu.: 7.416 3rd Qu.: 7.472
## Max. :10.000 Max. :10.000 Max. :10.000
## deep_reflection group_comfort party_liking listening_skill
## Min. : 0.000 Min. : 0.000 Min. : 0.000 Min. : 1.308
## 1st Qu.: 4.383 1st Qu.: 4.416 1st Qu.: 2.728 1st Qu.: 5.813
## Median : 5.972 Median : 6.051 Median : 5.053 Median : 6.990
## Mean : 5.959 Mean : 6.014 Mean : 5.040 Mean : 6.964
## 3rd Qu.: 7.598 3rd Qu.: 7.669 3rd Qu.: 7.364 3rd Qu.: 8.156
## Max. :10.000 Max. :10.000 Max. :10.000 Max. :10.000
## empathy creativity organization leadership
## Min. : 0.8049 Min. : 0.000 Min. : 0.000 Min. : 0.000
## 1st Qu.: 5.4485 1st Qu.: 4.993 1st Qu.: 4.806 1st Qu.: 4.441
## Median : 6.4814 Median : 6.003 Median : 5.987 Median : 6.043
## Mean : 6.4832 Mean : 5.996 Mean : 5.973 Mean : 6.020
## 3rd Qu.: 7.5215 3rd Qu.: 7.013 3rd Qu.: 7.142 3rd Qu.: 7.649
## Max. :10.0000 Max. :10.000 Max. :10.000 Max. :10.000
## risk_taking public_speaking_comfort curiosity routine_preference
## Min. : 0.000 Min. : 0.000 Min. : 0.6422 Min. : 0.000
## 1st Qu.: 3.428 1st Qu.: 3.614 1st Qu.: 5.4595 1st Qu.: 4.094
```

```
## Median : 5.012    Median : 5.535            Median : 6.4984    Median : 5.471
## Mean   : 5.012    Mean   : 5.512            Mean   : 6.5038    Mean   : 5.453
## 3rd Qu.: 6.602    3rd Qu.: 7.442            3rd Qu.: 7.5652    3rd Qu.: 6.817
## Max.   :10.000    Max.   :10.000            Max.   :10.0000    Max.   :10.000
## excitement_seeking  friendliness    emotional_stability    planning
## Min.   : 0.000    Min.   : 0.9932    Min.   : 0.4373    Min.   : 0.000
## 1st Qu.: 3.604    1st Qu.: 5.8432    1st Qu.: 5.0101    1st Qu.: 4.832
## Median : 5.545    Median : 6.9980    Median : 6.0265    Median : 5.981
## Mean   : 5.529    Mean   : 6.9757    Mean   : 6.0158    Mean   : 5.993
## 3rd Qu.: 7.502    3rd Qu.: 8.1747    3rd Qu.: 7.0241    3rd Qu.: 7.167
## Max.   :10.000    Max.   :10.0000    Max.   :10.0000    Max.   :10.000
## spontaneity    adventurousness    reading_habit    sports_interest
## Min.   : 0.000    Min.   : 0.000    Min.   : 0.000    Min.   : 0.000
## 1st Qu.: 4.156    1st Qu.: 3.429    1st Qu.: 4.325    1st Qu.: 4.127
## Median : 5.500    Median : 5.045    Median : 5.963    Median : 5.508
## Mean   : 5.495    Mean   : 5.034    Mean   : 5.941    Mean   : 5.492
## 3rd Qu.: 6.853    3rd Qu.: 6.656    3rd Qu.: 7.565    3rd Qu.: 6.861
## Max.   :10.000    Max.   :10.000    Max.   :10.000    Max.   :10.000
## online_social_usage    travel_desire    gadget_usage    work_style_collaborative
## Min.   : 0.000    Min.   : 0.000    Min.   : 0.000    Min.   : 0.000
## 1st Qu.: 5.158    1st Qu.: 4.155    1st Qu.: 4.868    1st Qu.: 5.161
## Median : 6.529    Median : 5.499    Median : 6.007    Median : 6.526
## Mean   : 6.499    Mean   : 5.509    Mean   : 6.007    Mean   : 6.502
## 3rd Qu.: 7.885    3rd Qu.: 6.877    3rd Qu.: 7.161    3rd Qu.: 7.879
## Max.   :10.000    Max.   :10.000    Max.   :10.000    Max.   :10.000
## decision_speed    stress_handling
## Min.   : 0.000    Min.   : 0.000
## 1st Qu.: 4.150    1st Qu.: 4.973
## Median : 5.528    Median : 5.979
## Mean   : 5.516    Mean   : 5.982
## 3rd Qu.: 6.906    3rd Qu.: 6.991
## Max.   :10.000    Max.   :10.000
```

Nhận xét: Ta thấy bộ dữ liệu này khảo sát với số lượng người thuộc từng nhóm ~ nhau:

- Ambivert :6573
- Extrovert:6857
- Introvert:657

Các biến còn lại đều là thang đo từ 1-10.

Kiểm tra giá trị bị thiếu

```
missing_values <- colSums(is.na(data))
print(missing_values)
```

```
##      personality_type    social_energy    alone_time_preference
##                0                0                0
##      talkativeness    deep_reflection    group_comfort
##                0                0                0
##      party_liking    listening_skill    empathy
##                0                0                0
##      creativity    organization    leadership
##                0                0                0
##      risk_taking    public_speaking_comfort    curiosity
##                0                0                0
```

```
##      routine_preference      excitement_seeking      friendliness
##              0              0              0
##      emotional_stability      planning      spontaneity
##              0              0              0
##      adventurousness      reading_habit      sports_interest
##              0              0              0
##      online_social_usage      travel_desire      gadget_usage
##              0              0              0
##      work_style_collaborative      decision_speed      stress_handling
##              0              0              0
```

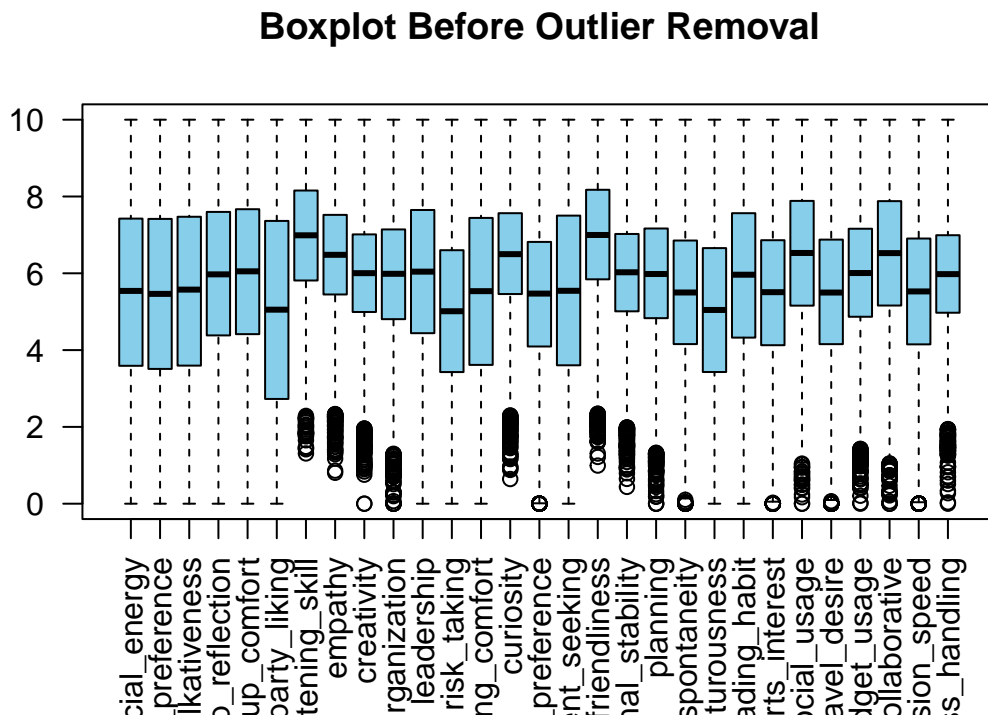
Nhận xét: Bộ dữ liệu không có giá trị nào bị thiếu.

Ta lấy các cột biến để thực hiện PCA (trừ biến personality_type)

```
numeric_data <- data[, -1]
```

Vẽ boxplot

```
boxplot(numeric_data, main = "Boxplot Before Outlier Removal", las = 2, col = "skyblue")
```



Nhận

xét: Ta nhận thấy có các giá trị ngoại lai ở một số biến.

Kiểm tra giá trị ngoại lai và tiến hành xử lý

Vì đây là bộ dữ liệu nhiều chiều nên không thể sử dụng IQR để xác định các giá trị ngoại lai mà cần sử dụng khoảng cách Mahalanobis để xác định các giá trị ngoại lai.

```
mahalanobis_distance <- mahalanobis(numeric_data, colMeans(numeric_data), cov(numeric_data))
cutoff_value <- qchisq(0.95, df = ncol(numeric_data)) # 95% quantile for chi-square distribution
outliers <- which(mahalanobis_distance > cutoff_value)
```

In số lượng ngoại lai và tính tỉ lệ ngoại lai trong bộ dữ liệu

```
cat("số lượng ngoại lai:", length(outliers), "\n")
```

```
## số lượng ngoại lai: 821
```

```
cat("Tỷ lệ ngoại lai:", length(outliers)/nrow(data)*100, "%", "\n")
```

```
## Tỷ lệ ngoại lai: 4.105 %
```

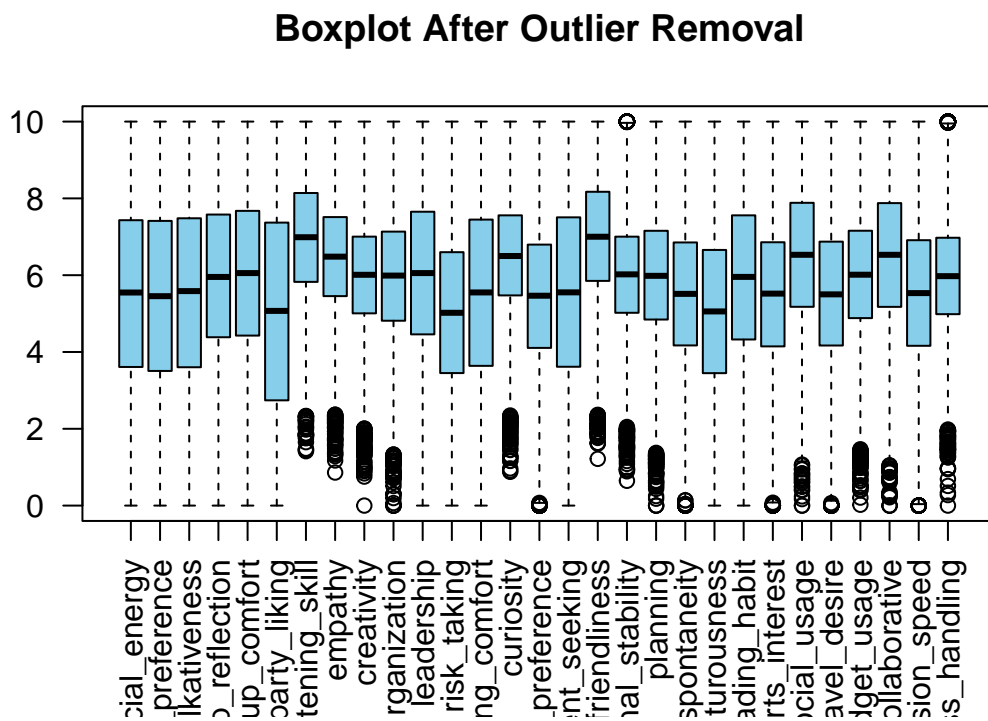
Nhận xét: có 821 giá trị ngoại lai chiếm 4,105% giá trị quan trắc bộ dữ liệu. Vì số lượng ngoại lai không lớn nên ta tiến hành loại bỏ các giá trị này.

Loại bỏ các giá trị ngoại lai.

```
data_no_outliers <- numeric_data[-outliers, ]
```

Vẽ lại boxplot sau khi loại bỏ các giá trị ngoại lai

```
boxplot(data_no_outliers, main = "Boxplot After Outlier Removal", las = 2, col = "skyblue")
```



Nhận xét: mặc dù có thể thấy vẫn còn nhiều giá trị ngoại lai, nhưng đã có thể nhìn thấy các đặc trưng của bộ dữ liệu

Kiểm tra lại bộ dữ liệu sau khi loại bỏ giá trị ngoại lai

```
nrow(data_no_outliers)
```

```
## [1] 19179
```

Nhận xét: còn 19,179 quan trắc

Thực hiện PCA trên bộ dữ liệu

```
pca_result <- prcomp(data_no_outliers, scale. = TRUE)
```

Hiển thị kết quả PCA.

```
summary(pca_result)
```

```
## Importance of components:
##          PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  3.4743 1.00879 1.00226 0.99460 0.96404 0.96298 0.88075
## Proportion of Variance 0.4162 0.03509 0.03464 0.03411 0.03205 0.03198 0.02675
## Cumulative Proportion 0.4162 0.45132 0.48596 0.52007 0.55212 0.58410 0.61085
##          PC8      PC9     PC10     PC11     PC12     PC13     PC14
## Standard deviation  0.87784 0.87198 0.8650 0.85581 0.77528 0.77209 0.77019
## Proportion of Variance 0.02657 0.02622 0.0258 0.02526 0.02073 0.02056 0.02046
## Cumulative Proportion 0.63742 0.66364 0.6894 0.71470 0.73542 0.75598 0.77643
##          PC15     PC16     PC17     PC18     PC19     PC20     PC21
## Standard deviation  0.76691 0.7635 0.75748 0.73347 0.67616 0.67019 0.66767
## Proportion of Variance 0.02028 0.0201 0.01979 0.01855 0.01577 0.01549 0.01537
## Cumulative Proportion 0.79671 0.8168 0.83660 0.85515 0.87092 0.88641 0.90178
##          PC22     PC23     PC24     PC25     PC26     PC27     PC28
## Standard deviation  0.66345 0.6596 0.61951 0.58181 0.57866 0.57463 0.57346
## Proportion of Variance 0.01518 0.0150 0.01323 0.01167 0.01155 0.01139 0.01134
## Cumulative Proportion 0.91696 0.9320 0.94519 0.95687 0.96841 0.97980 0.99114
##          PC29
## Standard deviation  0.50697
## Proportion of Variance 0.00886
## Cumulative Proportion 1.00000
```

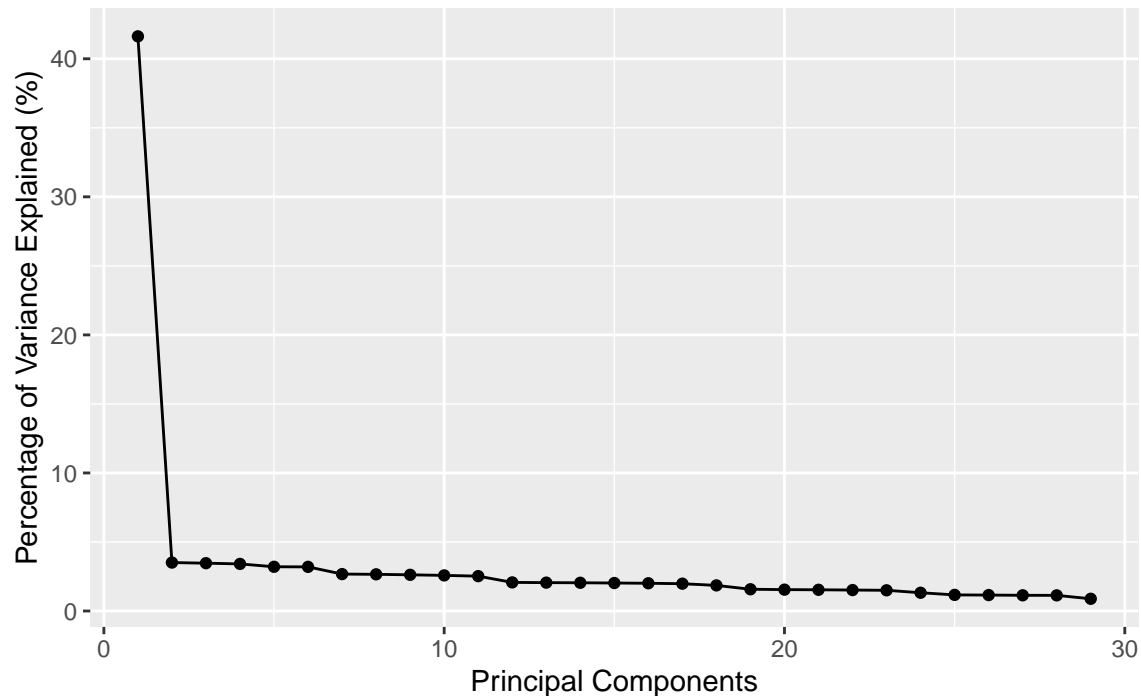
Nhận xét:

- Theo nguyên tắc Kaiser, nên giữ lại 3 thành phần chính (PC1, PC2, PC3), vì chỉ các thành phần này có trị riêng lớn hơn 1.
- Ba thành phần chính đầu tiên thoả nguyên tắc Kaiser tỉ lệ phương sai đóng góp là 48,596%
- Ta thấy từ thành phần chính thứ nhất (PC1) chiếm 41,62% tỉ lệ phương sai, chiếm phần lớn biến động
- Nhưng từ thành phần chính thứ 2 trở đi thì tỉ lệ đóng góp phương sai giảm mạnh chỉ còn dưới 4%. Ta có thể suy đoán từ thành phần chính thứ 2 sẽ có các đóng góp vào phương sai của bộ dữ liệu là rất nhỏ. Dựa vào tỉ lệ đóng góp phương sai có thể nhận 2 thành phần chính là PC1 và PC2 bỏ qua nguyên tắc Kaiser

```
# Scree plot
# Custom scree plot using ggplot2
pca_variance <- data.frame(
  PC = 1:length(pca_result$sdev),
  Variance = pca_result$sdev^2 / sum(pca_result$sdev^2) * 100
)
p3 <- ggplot(pca_variance, aes(x = PC, y = Variance)) +
  geom_line() +
  geom_point() +
  labs(title = "Scree Plot", x = "Principal Components", y = "Percentage of Variance Explained (%)") +
  theme(plot.title = element_text(hjust = 0.5, size = 14))

# Display scree plot
print(p3)
```


Scree Plot



Nhận xét:

- PC1: Giải thích khoảng 40% phương sai.
- PC2 từ trở đi: Giảm mạnh xuống dưới mức 4% phương sai, đồ thị bắt đầu thoải ra. Phù hợp với nhận định ở trên là chỉ nên giữ lại 2 thành phần chính là PC1 và PC2.
- Điểm elbow có thể xác định là điểm PC2.

Bảng giá trị Loading của bộ dữ liệu.

```
pca_loadings <- pca_result$rotation
print("PCA Loadings:")
```

```
## [1] "PCA Loadings:"
```

```
print(pca_loadings)
```

	PC1	PC2	PC3	PC4
## social_energy	-0.2354584797	-0.003230447	0.0089488402	-0.0042293311
## alone_time_preference	0.2377749517	0.007432720	0.0001628922	0.0147584702
## talkativeness	-0.2372408347	0.006574858	0.0037555831	-0.0048204969
## deep_reflection	0.2166162385	-0.001820952	0.0095025283	0.0206507891
## group_comfort	-0.2195160341	-0.013787039	0.0077433819	-0.0068214203
## party_liking	-0.2493799053	-0.005017505	-0.0075050892	-0.0023049277
## listening_skill	0.1468725243	0.014067368	-0.0564493072	0.0202680841
## empathy	0.0801071331	0.059369963	-0.1686416276	-0.1037094559
## creativity	-0.0004821603	-0.595868071	0.5609243127	0.5309138296
## organization	0.1463626540	-0.006812303	0.0479412644	-0.0218985580
## leadership	-0.2188241182	0.003498648	0.0056999341	0.0085358383
## risk_taking	-0.2162371489	-0.005453788	0.0043354777	-0.0071092794
## public_speaking_comfort	-0.2369338572	0.008759231	-0.0018120175	0.0008787959
## curiosity	-0.0789811459	0.182968366	-0.0078833501	0.0501370454

## routine_preference	0.1896374154	-0.003830423	0.0214836475	0.0155830884
## excitement_seeking	-0.2361408472	-0.006016336	0.0064426953	0.0077796029
## friendliness	-0.1478245989	-0.013370500	0.0130007514	0.0163475041
## emotional_stability	0.0017478281	-0.696548820	-0.0999051163	-0.6861695288
## planning	0.1453969364	0.028014436	0.0172714632	0.0334453077
## spontaneity	-0.1895430937	0.012875210	0.0050047836	0.0068856811
## adventurousness	-0.2180031878	-0.005368661	-0.0116719936	0.0002835990
## reading_habit	0.2184463722	-0.009105386	0.0066669262	-0.0066313027
## sports_interest	-0.1919115299	0.005687154	-0.0156160504	0.0099142558
## online_social_usage	-0.1888350443	0.003783110	-0.0240690467	-0.0203554800
## travel_desire	-0.1891205760	-0.010082285	-0.0045409474	-0.0010329808
## gadget_usage	-0.1476171313	-0.011465669	-0.0383898185	0.0045425790
## work_style_collaborative	-0.1886869605	-0.016427512	0.0034867044	-0.0071158787
## decision_speed	-0.1909480023	0.015889926	0.0255022826	-0.0008063213
## stress_handling	0.0029519266	0.346372929	0.7980502933	-0.4794939811
##	PC5	PC6	PC7	PC8
## social_energy	0.017550985	-0.010384951	-0.0032912990	-0.017138566
## alone_time_preference	-0.017201923	0.016590026	0.0177071272	0.021212970
## talkativeness	0.003785184	-0.016930986	-0.0206211102	-0.017498024
## deep_reflection	-0.017153930	0.011455465	-0.0028409109	0.002180901
## group_comfort	-0.008479773	-0.022998927	-0.0045131721	-0.011942829
## party_liking	0.024780835	-0.011077792	-0.0119434377	-0.009636668
## listening_skill	-0.019588464	0.043563634	-0.1888332710	0.609126865
## empathy	-0.073082742	-0.968366293	-0.0334008679	0.018114692
## creativity	-0.103524714	-0.182771005	-0.0338574773	0.033276127
## organization	0.003753556	-0.037884494	0.5357371137	-0.507187283
## leadership	0.013933779	0.002320159	-0.0057858023	-0.011331054
## risk_taking	0.016504523	-0.025124084	-0.0508480733	-0.007947540
## public_speaking_comfort	0.011052474	-0.010749883	-0.0178312970	-0.019340906
## curiosity	-0.972313400	0.070846574	0.0567434821	0.012474166
## routine_preference	-0.003218821	0.038661321	0.0601747113	0.042580963
## excitement_seeking	0.016054459	-0.022736217	-0.0183165635	-0.019650699
## friendliness	0.044894582	-0.023269025	0.6243600675	0.182937590
## emotional_stability	-0.162305280	0.063135316	-0.0143387878	-0.002676133
## planning	-0.049184231	0.043719125	-0.4507092789	-0.562199620
## spontaneity	0.004163952	-0.011090507	0.0118044918	-0.011813344
## adventurousness	-0.011908523	-0.027676706	-0.0150489402	-0.010053056
## reading_habit	-0.009462591	0.009149553	-0.0004598184	-0.004038720
## sports_interest	0.049789891	-0.018720524	0.0049431630	-0.037760926
## online_social_usage	0.017703035	-0.033040411	-0.0344270517	-0.070744550
## travel_desire	0.025440926	-0.020392523	-0.0252225932	-0.035778726
## gadget_usage	0.015226833	-0.045629816	-0.2455904008	-0.041020712
## work_style_collaborative	0.006033024	0.012150964	-0.0633580936	-0.042478650
## decision_speed	0.013221208	-0.017701253	-0.0498308270	-0.040432701
## stress_handling	0.028036232	-0.066933996	-0.0428720671	0.058719719
##	PC9	PC10	PC11	PC12
## social_energy	0.029983897	0.010621363	-0.049640665	-0.0118495766
## alone_time_preference	-0.027932150	-0.013687982	0.023403016	-0.0185831902
## talkativeness	0.030008256	-0.002436067	-0.029808368	-0.0228083149
## deep_reflection	-0.023462144	0.003480261	0.028531241	0.0156123986
## group_comfort	0.015849151	-0.006605581	-0.029177197	-0.0098647034
## party_liking	0.033057209	0.009954208	-0.032114859	0.0051430966
## listening_skill	0.284064187	0.643532407	-0.236908475	-0.0379662778
## empathy	0.051751055	-0.014781740	0.007163343	0.0192922788

## creativity	-0.000099662	0.006559388	-0.015429710	-0.0208925758
## organization	-0.099200526	0.487046096	-0.412738523	-0.0069128398
## leadership	0.024032792	0.022507453	-0.035980955	-0.0192081673
## risk_taking	0.042229869	0.025359388	-0.050981909	-0.0265675237
## public_speaking_comfort	0.033055105	0.016211473	-0.036852907	-0.0088573792
## curiosity	-0.026663564	-0.001190888	0.006985740	-0.0054544279
## routine_preference	-0.112357148	-0.025381551	0.096713290	0.0304805238
## excitement_seeking	0.035226340	0.015470346	-0.034593273	0.0025972371
## friendliness	0.291443482	0.171972232	0.653996940	-0.0309455530
## emotional_stability	0.028303601	0.012817745	0.019573242	0.0190755624
## planning	0.505316669	0.250500763	0.349277353	0.0139456044
## spontaneity	0.055135494	0.039219368	-0.054697237	0.4887717385
## adventurousness	0.035934211	0.028165015	-0.052506320	-0.0098761962
## reading_habit	-0.038257773	-0.006209130	0.031639627	0.0264676740
## sports_interest	0.067253322	0.055301251	-0.071198292	0.0594519918
## online_social_usage	0.043142738	0.029410504	-0.073507427	-0.7978420866
## travel_desire	0.056078115	0.033246287	-0.064160152	0.2404966217
## gadget_usage	-0.721452039	0.470813313	0.404447735	0.0007340383
## work_style_collaborative	0.070617114	0.090435080	-0.083854689	-0.0414667344
## decision_speed	0.048565944	0.124006144	-0.091076888	0.2305893785
## stress_handling	-0.006096783	0.022871924	0.021229524	-0.0134641275
##	PC13	PC14	PC15	PC16
## social_energy	-0.0048592754	-0.0317060651	0.007975429	0.0421556712
## alone_time_preference	0.0083608901	-0.0187721477	-0.017652605	-0.0410708963
## talkativeness	-0.0002670947	0.0036306313	0.020267674	0.0197876537
## deep_reflection	0.0131861132	-0.0038832661	-0.063596615	-0.0265709954
## group_comfort	-0.0056628480	-0.0163209352	0.055953510	-0.0423096711
## party_liking	-0.0093408471	-0.0070409281	0.012238332	0.0145981262
## listening_skill	-0.0270187301	-0.0134674803	-0.029903473	0.0332929723
## empathy	0.0365837441	0.0101714737	0.010428645	-0.0054678914
## creativity	-0.0142283275	-0.0113961392	-0.003724472	0.0027834390
## organization	-0.0168619029	-0.0197152008	-0.019517959	-0.0017384635
## leadership	-0.0106314047	0.0148726413	0.050451059	-0.0167484325
## risk_taking	0.0388402419	-0.0385405535	0.016291353	0.0620697563
## public_speaking_comfort	0.0260141270	-0.0012120466	-0.011717712	0.0378701513
## curiosity	-0.0134749359	0.0198339987	-0.001591621	-0.0062061881
## routine_preference	0.5387125089	0.3781150233	0.097042042	0.4454950430
## excitement_seeking	-0.0079278666	0.0069083662	0.001110816	0.0005183836
## friendliness	0.0082423087	0.0129938398	-0.004211540	-0.0284432329
## emotional_stability	-0.0093050927	-0.0006452529	0.015061907	0.0186390864
## planning	-0.0114823020	-0.0108646300	0.001349840	0.0421296555
## spontaneity	0.2312712970	-0.6038990422	-0.232489051	0.3814818861
## adventurousness	-0.0346163071	0.0135474349	0.017182849	-0.0003730590
## reading_habit	0.0350293983	-0.0095495876	-0.066359777	-0.0619102827
## sports_interest	-0.3249396957	0.0755787835	0.283339642	-0.2086385805
## online_social_usage	0.0799257895	-0.0691049710	-0.167552666	0.3591255967
## travel_desire	-0.1387659726	0.5888597916	-0.700511810	0.0371656117
## gadget_usage	-0.0408351332	-0.0351341611	-0.032371997	-0.0307215337
## work_style_collaborative	0.7148781591	0.0349188449	-0.033845264	-0.6128731710
## decision_speed	0.0631378336	0.3569178609	0.563509599	0.2920510412
## stress_handling	-0.0130925978	-0.0055616557	-0.017808513	-0.0137041258
##	PC17	PC18	PC19	PC20
## social_energy	0.011677806	-0.1234072712	-1.184470e-02	-0.062282704
## alone_time_preference	0.001989953	0.0821437255	5.955568e-03	-0.032741872

## talkativeness	0.008560289	-0.0997404536	1.253367e-02	-0.022881365
## deep_reflection	0.062841141	0.2646925962	6.459360e-01	0.129092352
## group_comfort	0.005037200	-0.1372446067	-5.556975e-02	0.417670246
## party_liking	0.004973546	-0.0876370252	-2.234698e-02	0.014590472
## listening_skill	0.022105402	-0.0809934173	-2.722382e-02	0.018998336
## empathy	0.011646917	-0.0190399305	-9.261333e-03	0.009758191
## creativity	0.003728631	0.0092131029	-1.681298e-02	-0.001709650
## organization	-0.013957843	-0.0989037487	6.640051e-03	0.016039773
## leadership	-0.015035372	-0.2423181843	-1.983068e-03	-0.033918721
## risk_taking	0.050474693	-0.2111009093	7.507488e-01	-0.083050400
## public_speaking_comfort	-0.022139021	-0.1234923366	-2.225509e-03	0.015436605
## curiosity	0.028698957	0.0235600745	3.366388e-03	0.001837239
## routine_preference	0.445560098	-0.2903737800	-4.040742e-02	-0.021030518
## excitement_seeking	0.003547263	-0.1032859551	2.940786e-02	-0.007210046
## friendliness	-0.038185300	0.0659354919	2.972983e-02	-0.015352141
## emotional_stability	0.014197597	0.0049589393	2.996805e-03	0.004841133
## planning	0.026235973	-0.0935915876	-2.728169e-02	0.010572534
## spontaneity	0.141120126	0.2486017709	-6.399905e-02	-0.001833252
## adventurousness	-0.059622741	-0.2044639058	2.074260e-06	-0.683375106
## reading_habit	0.022181118	0.2451430675	-2.776018e-03	-0.569155466
## sports_interest	0.785117944	0.2803192592	-4.319920e-02	-0.035117500
## online_social_usage	0.054693731	0.3600494998	-4.856153e-02	-0.004664252
## travel_desire	0.003338620	0.1595610471	-8.770983e-03	0.031296315
## gadget_usage	0.002517255	0.0310792981	-3.744538e-03	0.003537011
## work_style_collaborative	-0.001273216	0.1939948780	-2.229578e-02	0.008076704
## decision_speed	-0.382430013	0.4376361762	3.111560e-02	-0.007260702
## stress_handling	0.018632079	-0.0007243687	-5.470727e-03	-0.004508885
##	PC21	PC22	PC23	PC24
## social_energy	-0.0545746795	0.0252388257	0.045020917	-0.380681932
## alone_time_preference	0.0227992806	0.0042020700	-0.017051049	0.264830824
## talkativeness	-0.0393016008	-0.0219733045	-0.013912789	-0.273166153
## deep_reflection	-0.5235430459	-0.3197960924	0.042724222	-0.234241736
## group_comfort	-0.3952401241	0.3292121251	0.646144739	0.239857392
## party_liking	-0.0264678630	0.0206521122	0.012397077	-0.158786727
## listening_skill	-0.0106794408	0.0082240839	0.025325093	-0.033024923
## empathy	0.0071051715	-0.0222371058	0.001368069	-0.009117392
## creativity	0.0131562613	0.0003487206	-0.013715114	0.004820667
## organization	0.0029145807	0.0152832964	0.003366820	-0.019570193
## leadership	0.1804231677	-0.8256694600	0.326597903	0.241826462
## risk_taking	0.4172739530	0.2927386575	0.013591361	0.240609692
## public_speaking_comfort	0.0397202003	-0.0345658100	-0.017483423	-0.292911013
## curiosity	0.0271539341	-0.0006519352	-0.005661573	-0.007476519
## routine_preference	-0.0622722864	0.0026398411	0.029736427	-0.036935341
## excitement_seeking	-0.0059703168	0.0252936841	-0.018130589	-0.431068623
## friendliness	0.0099313357	0.0081596380	0.002036195	0.013453744
## emotional_stability	0.0005419307	-0.0203317922	-0.011450884	-0.003506523
## planning	-0.0128868499	0.0035528117	0.010462907	-0.006058952
## spontaneity	-0.0111830071	-0.0412450546	-0.008360232	0.099421040
## adventurousness	-0.5562834270	0.0531893182	-0.164558427	0.293566961
## reading_habit	0.1959066254	0.1097576469	0.658420612	-0.249314799
## sports_interest	0.0260169861	-0.0049395883	-0.045350933	0.063461832
## online_social_usage	-0.0013891273	-0.0224833193	0.008716583	0.078918796
## travel_desire	0.0294520151	0.0069741791	0.004347305	0.087102470
## gadget_usage	0.0027013023	0.0036070726	-0.015218432	0.013089987

## work_style_collaborative	0.0132004084	-0.0055931997	-0.072536005	0.048344767
## decision_speed	0.0258307958	-0.0039471607	0.010634442	0.053586538
## stress_handling	-0.0091723532	-0.0037529615	-0.009638336	0.006744085
##	PC25	PC26	PC27	PC28
## social_energy	0.7701920653	0.127046097	0.1916517872	-0.3345650201
## alone_time_preference	0.1888762422	0.232313752	-0.7333837164	-0.4540675584
## talkativeness	-0.0264275620	0.649478919	-0.3504079304	0.5458581180
## deep_reflection	-0.0259508249	0.018408436	0.0264949861	-0.0252344252
## group_comfort	-0.0209544643	-0.006228447	-0.0524412317	-0.0236541047
## party_liking	-0.0750780863	-0.007159057	0.0386671922	-0.0183100924
## listening_skill	-0.0046306659	0.006917727	0.0021208909	0.0142281723
## empathy	0.0053228103	0.003645854	0.0040743082	-0.0061817239
## creativity	-0.0059978333	0.009829858	0.0086679366	0.0001646245
## organization	-0.0122044636	0.007363606	-0.0010451089	0.0172944880
## leadership	0.0447997858	-0.053183336	-0.0315366769	0.0046137227
## risk_taking	0.0290002382	-0.003436893	0.0089686989	0.0052287136
## public_speaking_comfort	-0.5885245152	0.270008467	0.0718659865	-0.6119626613
## curiosity	0.0053387586	-0.009914202	0.0073745191	-0.0023324201
## routine_preference	-0.0061846770	-0.010804650	-0.0092367978	0.0060225765
## excitement_seeking	-0.0544855459	-0.653968196	-0.5373509338	0.0637635135
## friendliness	0.0045758337	0.004006044	-0.0035191013	-0.0046247334
## emotional_stability	-0.0024908729	0.003402297	-0.0129564663	-0.0093817908
## planning	0.0088076108	0.003985089	0.0002171406	-0.0014353967
## spontaneity	-0.0008944853	-0.010197958	-0.0287786357	0.0121451609
## adventurousness	-0.0509261809	-0.045204024	-0.0006182725	-0.0182576144
## reading_habit	-0.0755601662	0.021023909	0.0142390412	0.0233809675
## sports_interest	-0.0130269850	0.006309652	0.0010547532	-0.0355074046
## online_social_usage	0.0014737218	-0.029846013	-0.0068954429	0.0013913029
## travel_desire	0.0344686053	0.012733119	-0.0131257621	-0.0081351371
## gadget_usage	0.0102749860	0.004874247	-0.0058241960	-0.0066707433
## work_style_collaborative	0.0266910939	-0.002598409	0.0161397436	0.0006425350
## decision_speed	0.0242727598	0.002499559	-0.0270933349	-0.0259817714
## stress_handling	-0.0052502771	-0.005398883	-0.0010765422	-0.0037711708
##	PC29			
## social_energy	0.096423776			
## alone_time_preference	-0.158361712			
## talkativeness	0.093070524			
## deep_reflection	-0.047301225			
## group_comfort	0.066889043			
## party_liking	-0.944230658			
## listening_skill	-0.003180696			
## empathy	-0.006168116			
## creativity	-0.003309320			
## organization	-0.010987590			
## leadership	0.018721921			
## risk_taking	0.016199300			
## public_speaking_comfort	0.184452360			
## curiosity	-0.010392319			
## routine_preference	-0.020922834			
## excitement_seeking	0.133905346			
## friendliness	0.009541914			
## emotional_stability	0.001483087			
## planning	-0.009206637			
## spontaneity	0.022584836			

```
## adventurousness      0.039799944
## reading_habit        -0.040326739
## sports_interest      0.029597851
## online_social_usage   0.009321625
## travel_desire         0.013333427
## gadget_usage          0.003438040
## work_style_collaborative 0.011578981
## decision_speed        0.011919526
## stress_handling       -0.007741294
```

- PC1 (Thành phần chính đầu tiên)

Biến đóng góp lớn (tuyệt đối > 0.3):

social_energy: -0.2354584797 (âm, liên quan ngược).

alone_time_preference: 0.237749517 (dương, ưa thích thời gian một mình).

talkativeness: -0.2372408347 (âm, ít nói chuyện).

deep_reflection: 0.2161623835 (dương, suy ngẫm sâu).

group_comfort: -0.2195160341 (âm, không thoải mái trong nhóm).

party_liking: -0.2493799053 (âm, không thích tiệc tùng).

listening_skill: 0.1468725243 (dương, nhẹ, kỹ năng lắng nghe).

Nhận xét: PC1 dường như đại diện cho xu hướng hướng nội (introversion) so với hướng ngoại (extraversion). Các biến như ưa thích thời gian một mình, suy ngẫm sâu, và không thoải mái trong nhóm có tải dương, trong khi năng lượng xã hội, nói chuyện, và thích tiệc tùng có tải âm.

- PC2 (Thành phần chính thứ hai)

Biến đóng góp lớn:

empathy: 0.59369963 (dương, rất mạnh, đồng cảm).

creativity: 0.5966071 (dương, sáng tạo).

organization: -0.006812303 (gần 0, không ảnh hưởng).

leadership: 0.00393648 (gần 0, không ảnh hưởng).

risk_taking: -0.0163788 (gần 0).

Nhận xét: PC2 có thể đại diện cho khả năng cảm xúc và sáng tạo. Biến empathy và creativity có tải rất cao, trong khi các biến liên quan đến tổ chức hoặc rủi ro gần như không ảnh hưởng.

Kết luận: Dựa trên ma trận tải, chỉ 2 PC đầu tiên có các giá trị tải đáng kể (> 0.3 hoặc < -0.3) trên một số biến. Điều này phù hợp với phân tích Scree Plot (giữ 2 thành phần đầu).

```
# View PCA scores (optional)
```

```
pca_scores <- pca_result$x
head(pca_scores)
```

##	PC1	PC2	PC3	PC4	PC5	PC6	PC7
## 1	-4.95088554	0.5999770	0.4891240	0.1444526	1.16149946	-0.4541800	-1.8357171
## 2	0.01836153	-0.1719671	-0.9336756	1.7688506	0.48429265	1.6325181	1.1703302
## 3	0.43470228	-0.3817053	0.1990563	0.6744847	0.08598068	0.1158613	0.2628600
## 4	-3.73767338	-1.5505420	0.5898213	0.5874661	0.58840015	-2.2308770	0.3072722
## 5	3.37129356	-1.7210118	0.1638305	2.2629694	0.43322822	-0.3670358	-0.3117337
## 6	4.06680688	0.3056074	-1.1750224	0.1984983	-0.01043970	1.3401845	-0.2985862
##	PC8	PC9	PC10	PC11	PC12	PC13	PC14

```
## 1 1.9358290 -0.5371730 -0.4299197 1.55695192 -1.1549798 -0.30526355 0.3409464
## 2 0.6454624 -0.4535676 -1.2101079 1.30177495 0.5101126 0.01908338 -0.5741130
## 3 0.7016451 0.4840763 1.2850327 -0.26389166 0.4532659 -0.29223068 -0.2042601
## 4 1.3664440 1.9984536 1.2224723 -0.08013082 -0.7255313 -1.48117896 0.2873330
## 5 0.4122888 -1.1201225 -0.4537878 -0.95157307 -0.5306182 0.79546227 0.2815894
## 6 1.3468900 -0.2970477 -2.1251530 0.63233064 0.7416351 -0.24276168 0.6202183
##      PC15      PC16      PC17      PC18      PC19      PC20
## 1 1.6705710 -0.54284521 0.90876745 0.5325195 -0.48873715 -1.12068842
## 2 -0.3027956 -0.24528527 0.44229842 -0.6670295 0.51353381 -0.76196392
## 3 -0.8076508 0.08180701 -1.09336765 -1.2123305 -0.62237805 -1.19149085
## 4 0.6586500 -0.74310701 -0.01557478 -0.9664117 -0.15521181 -0.95375371
## 5 -0.7297615 1.04729249 -0.85145472 1.3606122 0.40374646 -0.17720956
## 6 -0.0444474 0.66553920 1.11409372 -0.8059357 0.01769597 -0.09703069
##      PC21      PC22      PC23      PC24      PC25      PC26
## 1 0.3713941 -0.04842183 -0.03959805 0.07370323 -0.5358582 -0.01794191
## 2 -0.4035717 -0.49024913 -0.66060435 -0.54392243 0.3616577 0.14809047
## 3 -1.1115715 1.30613036 -1.05422404 0.07626442 0.2979918 0.26936084
## 4 0.7778894 -0.14753323 0.55278328 0.58983262 -0.5050176 -1.06196186
## 5 -0.2543071 -0.96785218 -0.85887564 -0.72071600 0.8757206 -0.06579171
## 6 1.0401348 1.17892918 0.56930506 -0.15526342 -0.1365709 -0.58982298
##      PC27      PC28      PC29
## 1 -0.5875581 -0.073829921 -0.1200550
## 2 -0.3868965 0.353153465 0.6439832
## 3 0.5561247 -1.006520291 -0.8428700
## 4 -0.1709581 -0.000686744 -0.7614001
## 5 0.1202767 0.141038618 0.7634102
## 6 0.1524958 0.403526749 0.7775831
```

- PC1 (Thành phần chính đầu tiên)

Quan sát nổi bật:

Dòng 1: -4.95088554 (rất âm, nổi bật ở hướng ngược).

Dòng 2: 0.01836153 (gần 0, trung bình).

Dòng 3: 0.43470228 (dương nhẹ).

Dòng 5: 3.37129356 (rất dương, nổi bật ở hướng thuận).

Nhận xét: PC1 có biên độ lớn (-4.95 đến 3.37), phản ánh sự khác biệt mạnh giữa các quan sát. Dựa trên ma trận tải trước, PC1 liên quan đến hướng nội/hướng ngoại, nên các quan sát như dòng 1 (âm) có thể nghiêng về hướng ngoại, trong khi dòng 5 (dương) nghiêng về hướng nội.

- PC2 (Thành phần chính thứ hai)

Quan sát nổi bật:

Dòng 1: 0.5999770 (dương nhẹ).

Dòng 3: -1.5505420 (âm, nổi bật).

Dòng 5: -1.7210118 (âm, nổi bật).

Nhận xét: PC2 có biên độ từ -1.72 đến 0.60, liên quan đến cảm xúc/sáng tạo (theo ma trận tải). Quan sát dòng 3 và 5 có xu hướng thấp về đồng cảm/sáng tạo, trong khi dòng 1 cao hơn.

```
# Loading plot for PC1 and PC2 with arrows
pca_loadings <- as.data.frame(pca_result$rotation[, 1:2]) # Extract loadings for PC1 and PC2
pca_loadings$variable <- rownames(pca_loadings)
```

```

# Scale loadings for better visualization (optional, adjust multiplier as needed)
pca_loadings$PC1 <- pca_loadings$PC1 * 5
pca_loadings$PC2 <- pca_loadings$PC2 * 5

p4 <- ggplot() +
  geom_segment(data = pca_loadings, aes(x = 0, y = 0, xend = PC1, yend = PC2),
    arrow = arrow(length = unit(0.2, "cm")), color = "red") +
  geom_text(data = pca_loadings, aes(x = PC1, y = PC2, label = variable),
    vjust = 1.5, hjust = 0.5, size = 3, color = "red") +
  geom_vline(xintercept = 0, linetype = "dashed", color = "gray") +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray") +
  labs(title = "Loading Plot with Arrows (PC1 vs PC2)", x = "PC1 Loadings", y = "PC2 Loadings") +
  theme(plot.title = element_text(hjust = 0.5, size = 14)) +
  xlim(-1, 1) + ylim(-1, 1) # Adjust limits based on scaled loadings

# Display loading plot
print(p4)

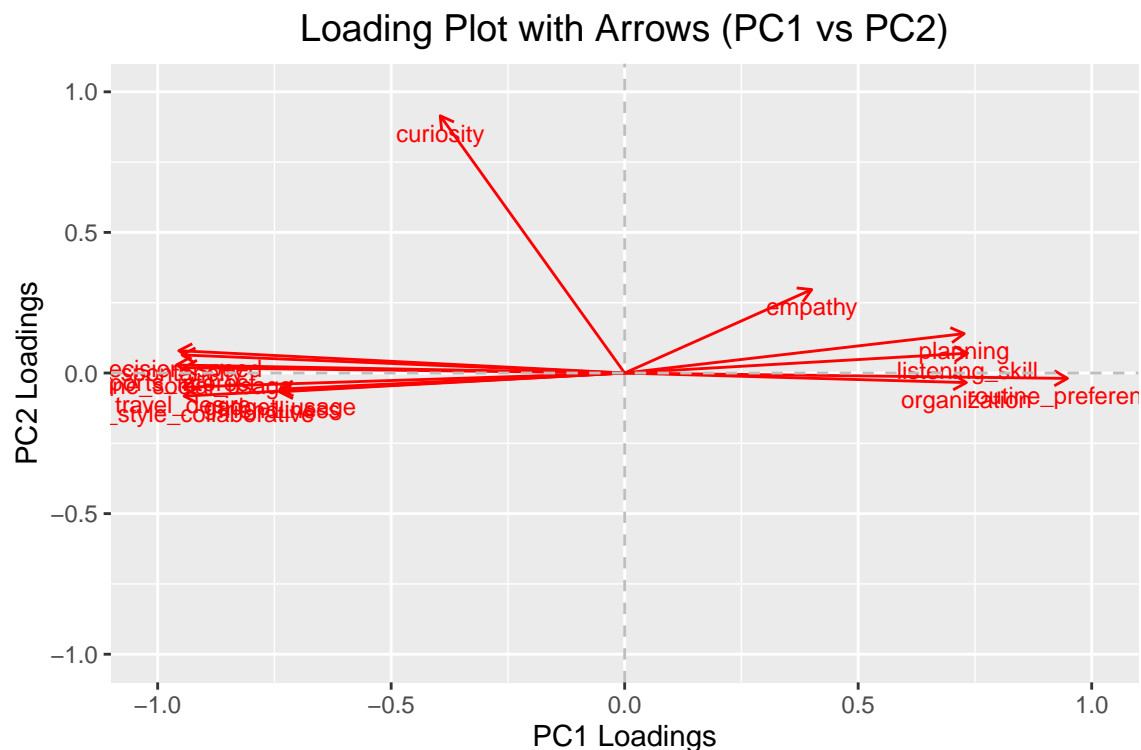
```

```

## Warning: Removed 15 rows containing missing values or values outside the scale range
## (`geom_segment()`).

## Warning: Removed 15 rows containing missing values or values outside the scale range
## (`geom_text()`).

```



Nhận

xét:

Biến có tải lớn trên PC1 (hướng ngang)

Biến âm (bên trái, PC1 < 0):

social_energy, talkativeness, party_liking, group_comfort: Các biến này có tải âm trên PC1, gợi ý PC1 phân biệt hướng ngoại (âm) so với hướng nội (dương). Những người có giá trị âm trên PC1 có xu hướng năng động

xã hội.

Biến dương (bên phải, $PC1 > 0$):

alone_time_preference, deep_reflection: Các biến này có tải dương, liên quan đến xu hướng hướng nội, ưa thích thời gian một mình và suy ngẫm sâu.

Biến có tải lớn trên PC2 (hướng dọc)

Biến dương (trên, $PC2 > 0$):

curiosity, empathy, planning, listening_skill, organization, routine_preference: Các biến này có tải dương trên PC2, cho thấy PC2 có thể liên quan đến khả năng cảm xúc (empathy), tổ chức (planning, organization), và sự tò mò (curiosity).

Biến âm (dưới, $PC2 < 0$):

Ít biến nổi bật ở phía âm, nhưng một số như decision_speed có thể có tải nhẹ âm.

Biến có độ dài mũi tên lớn

curiosity, empathy, planning: Có mũi tên dài, cho thấy chúng đóng góp mạnh vào sự phân biệt giữa PC1 và PC2.

social_energy, alone_time_preference: Cũng có độ dài đáng kể, nhấn mạnh vai trò của chúng trong PC1.

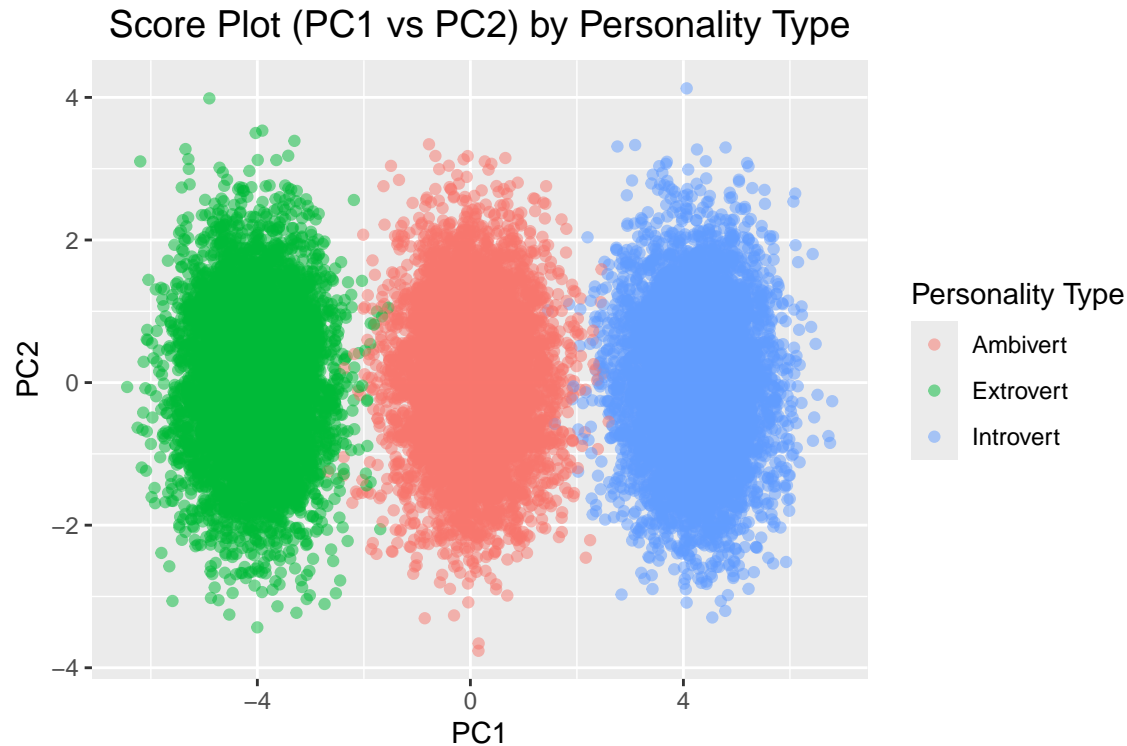
Ý nghĩa: PC1 chủ yếu phân biệt hướng nội (introversion) và hướng ngoại (extraversion), trong khi PC2 liên quan đến khả năng cảm xúc, tổ chức, và tò mò. Điều này phù hợp với ma trận tải trước.

```
# Score plot for PC1 and PC2 with personality_type
# Keep personality_type and sync with outliers
personality_types <- data$personality_type # Get personality_type from original data
personality_types_no_outliers <- personality_types[-outliers] # Remove outliers

pca_scores <- as.data.frame(pca_result$x)
pca_scores$personality_type <- personality_types_no_outliers # Assign synced personality types

p5 <- ggplot(pca_scores, aes(x = PC1, y = PC2, color = personality_type)) +
  geom_point(alpha = 0.5) +
  labs(title = "Score Plot (PC1 vs PC2) by Personality Type",
       x = "PC1", y = "PC2", color = "Personality Type") +
  theme(plot.title = element_text(hjust = 0.5, size = 14)) +
  scale_color_discrete(name = "Personality Type")

# Display score plot
print(p5)
```



Nhận xét:

Phân biệt rõ ràng: PC1 phân biệt tốt ba nhóm tính cách:

Extravert (âm) vs Introvert (dương), với Ambivert ở giữa, xác nhận PC1 là trục hướng nội/hướng ngoại.

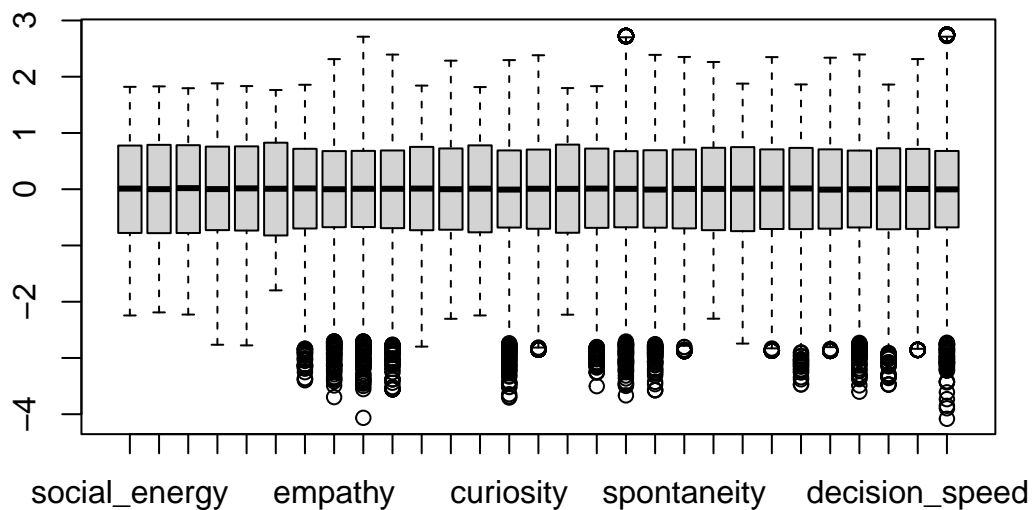
PC2 không phân biệt rõ ràng giữa các nhóm, nhưng cho thấy sự đa dạng nội bộ (ví dụ: Introvert có điểm cao trên PC2 có thể liên quan đến tò mò).

Độ chồng lấn: Có một số chồng lấn giữa Ambivert và hai nhóm còn lại, đặc biệt ở vùng PC1 gần 0, phản ánh tính chất trung gian của Ambivert.

Phân bố đều: Mỗi nhóm có số lượng quan sát đáng kể, với Extravert và Introvert có biên độ rộng hơn, trong khi Ambivert tập trung hơn.

Trước khi phân tích nhân tố thì bộ dữ liệu xét phải được chuẩn hóa, nhưng hiện tại bộ dữ liệu chưa được scale nên ta thực hiện scale lại:

```
data_sc <- scale(data_no_outliers, scale = T)
boxplot(data_sc)
```



Nhận xét: Bộ dữ liệu đã được chuẩn hóa, ta có thể tiến hành phân tích nhân tố FA

```
fa.dat <- factanal(data_sc , factors = 2)
fa.dat
```

```
##
## Call:
## factanal(x = data_sc, factors = 2)
##
## Uniquenesses:
```

social_energy	alone_time_preference	talkativeness
0.342	0.328	0.331
deep_reflection	group_comfort	party_liking
0.451	0.436	0.249
listening_skill	empathy	creativity
0.760	0.931	1.000
organization	leadership	risk_taking
0.763	0.442	0.453
public_speaking_comfort	curiosity	routine_preference
0.331	0.933	0.591
excitement_seeking	friendliness	emotional_stability
0.336	0.750	0.998
planning	spontaneity	adventurousness
0.746	0.592	0.446
reading_habit	sports_interest	online_social_usage
0.442	0.581	0.593
travel_desire	gadget_usage	work_style_collaborative
0.593	0.760	0.593
decision_speed	stress_handling	
0.584	1.000	

```
##
## Loadings:
##
## Factor1 Factor2
## social_energy      0.808
## alone_time_preference -0.816
## talkativeness      0.816
## deep_reflection     -0.732 -0.115
## group_comfort      0.742  0.116
## party_liking       0.862
## listening_skill    -0.482
## empathy            -0.260
## creativity
## organization      -0.483
## leadership        0.743
## risk_taking       0.739
## public_speaking_comfort 0.817
## curiosity         0.259
## routine_preference -0.639
## excitement_seeking 0.813
## friendliness      0.480  0.138
## emotional_stability
## planning          -0.467 -0.189
## spontaneity       0.637
## adventurousness   0.742
## reading_habit     -0.740
## sports_interest   0.645
## online_social_usage 0.637
## travel_desire     0.637
## gadget_usage      0.486
## work_style_collaborative 0.638
## decision_speed    0.644
## stress_handling
##
## Factor1 Factor2
## SS loadings    11.489  0.155
## Proportion Var  0.396  0.005
## Cumulative Var  0.396  0.402
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 376.5 on 349 degrees of freedom.
## The p-value is 0.149
```

Nhận xét: với giá trị $p_value = 0.149 > 0.05$ chúng ta không đủ cơ sở bác bỏ việc phân tích dựa trên 2 nhân tố là không thể hiện đầy đủ thông tin của bộ dữ liệu. Vì vậy ta nhận việc phân tích bộ dữ liệu dựa trên 2 nhân tố.

KIỂM TRA ĐỘ CUNG CẤP THÔNG TIN CHO NHÂN TỐ

```
round(1 - fa.dat$uniquenesses,3)
```

##	social_energy	alone_time_preference	talkativeness
##	0.658	0.672	0.669
##	deep_reflection	group_comfort	party_liking
##	0.549	0.564	0.751
##	listening_skill	empathy	creativity
##	0.240	0.069	0.000

##	organization	leadership	risk_taking
##	0.237	0.558	0.547
##	public_speaking_comfort	curiosity	routine_preference
##	0.669	0.067	0.409
##	excitement_seeking	friendliness	emotional_stability
##	0.664	0.250	0.002
##	planning	spontaneity	adventurousness
##	0.254	0.408	0.554
##	reading_habit	sports_interest	online_social_usage
##	0.558	0.419	0.407
##	travel_desire	gadget_usage	work_style_collaborative
##	0.407	0.240	0.407
##	decision_speed	stress_handling	
##	0.416	0.000	

Nhận xét:

- đây là các biến có hệ số lớn, có nghĩa chúng chia sẻ rất nhiều phương sai của nó với các biến khác: social_energy, alone_time_preference, talkativeness, deep_reflection, group_comfort, party_liking, leadership, risk_taking, public_speaking_comfort, excitement_seeking, adventurousness, reading_habit.
- nhưng có một số biến gần như bằng 0 như là stress_handling, creativity, emotional_stability, empathy. gần như chúng không chia sẻ bất cứ một thông tin nào. vì có sự chênh lệch như vậy nên ta sẽ thử xoay nhân tố theo “varimax” xem thử có thay đổi chút gì được không.

```
fa.dat <- factanal(data_sc, factors = 2)
fa.dat
```

```
##
## Call:
## factanal(x = data_sc, factors = 2)
##
## Uniquenesses:
##      social_energy  alone_time_preference  talkativeness
##      0.342          0.328                 0.331
##      deep_reflection  group_comfort        party_liking
##      0.451           0.436                 0.249
##      listening_skill  empathy             creativity
##      0.760            0.931                 1.000
##      organization    leadership           risk_taking
##      0.763            0.442                 0.453
##      public_speaking_comfort  curiosity  routine_preference
##      0.331                  0.933          0.591
##      excitement_seeking      friendliness  emotional_stability
##      0.336                    0.750          0.998
##      planning                 spontaneity  adventurousness
##      0.746                    0.592          0.446
##      reading_habit            sports_interest  online_social_usage
##      0.442                    0.581          0.593
##      travel_desire            gadget_usage  work_style_collaborative
##      0.593                    0.760          0.593
##      decision_speed          stress_handling
##      0.584                    1.000
##
## Loadings:
##
```

	Factor1	Factor2
--	---------	---------

```

## social_energy          0.808
## alone_time_preference -0.816
## talkativeness         0.816
## deep_reflection        -0.732 -0.115
## group_comfort          0.742  0.116
## party_liking           0.862
## listening_skill        -0.482
## empathy                -0.260
## creativity
## organization           -0.483
## leadership             0.743
## risk_taking            0.739
## public_speaking_comfort 0.817
## curiosity              0.259
## routine_preference     -0.639
## excitement_seeking     0.813
## friendliness           0.480  0.138
## emotional_stability
## planning               -0.467 -0.189
## spontaneity            0.637
## adventurousness        0.742
## reading_habit          -0.740
## sports_interest        0.645
## online_social_usage     0.637
## travel_desire           0.637
## gadget_usage           0.486
## work_style_collaborative 0.638
## decision_speed         0.644
## stress_handling
##
##               Factor1 Factor2
## SS loadings    11.489  0.155
## Proportion Var  0.396  0.005
## Cumulative Var  0.396  0.402
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 376.5 on 349 degrees of freedom.
## The p-value is 0.149

```

```
round(1 - fa.dat$uniquenesses,3)
```

```

##          social_energy  alone_time_preference  talkativeness
##              0.658              0.672              0.669
##      deep_reflection      group_comfort      party_liking
##              0.549              0.564              0.751
##      listening_skill      empathy      creativity
##              0.240              0.069              0.000
##      organization      leadership      risk_taking
##              0.237              0.558              0.547
##      public_speaking_comfort      curiosity      routine_preference
##              0.669              0.067              0.409
##      excitement_seeking      friendliness      emotional_stability
##              0.664              0.250              0.002
##      planning      spontaneity      adventurousness
##              0.254              0.408              0.554

```

```
##          reading_habit          sports_interest          online_social_usage
##          0.558              0.419              0.407
##          travel_desire          gadget_usage work_style_collaborative
##          0.407              0.240              0.407
##          decision_speed          stress_handling
##          0.416              0.000
```

Nhận xét: Sau khi xoay ta cũng không nhận lại được gì nhiều thay đổi nên vì vậy ta sẽ phân tích hệ số tải của FA luôn.

```
fa.dat$loadings
```

```
##
## Loadings:
##          Factor1 Factor2
## social_energy      0.808
## alone_time_preference -0.816
## talkativeness      0.816
## deep_reflection    -0.732 -0.115
## group_comfort      0.742  0.116
## party_liking       0.862
## listening_skill    -0.482
## empathy            -0.260
## creativity
## organization      -0.483
## leadership         0.743
## risk_taking        0.739
## public_speaking_comfort 0.817
## curiosity          0.259
## routine_preference -0.639
## excitement_seeking 0.813
## friendliness       0.480  0.138
## emotional_stability
## planning           -0.467 -0.189
## spontaneity        0.637
## adventurousness    0.742
## reading_habit      -0.740
## sports_interest    0.645
## online_social_usage 0.637
## travel_desire      0.637
## gadget_usage       0.486
## work_style_collaborative 0.638
## decision_speed     0.644
## stress_handling
##
##          Factor1 Factor2
## SS loadings    11.489  0.155
## Proportion Var  0.396  0.005
## Cumulative Var  0.396  0.402
```

Nhận xét:

- Ở factor 1:
 - các biến có hệ số lớn hơn 0 như là: social_energy, talkativeness, group_comfort, party_liking, leadership, risk_taking, public_speaking_comfort, curiosity, excitement_seeking, friendliness, spontaneity, adventurousness, sports_interest, online_social_usage, travel_desire, gadget_usage,

- work_style_collaborative, decision_speed
- Các biến có hệ số < 0 : alone_time_preference, deep_reflection, listening_skill, empathy, organization, routine_preference, routine_preference, planning, reading_habit, stress_handling => đây là những tính cách thể hiện một người có tính cách hướng ngoại và đối lập với điều đó là một người hướng nội. Từ đây ta có thể thấy factor 1 chính là đại diện nhân tố của một người hướng ngoại và dễ dàng nhận biết và phân biệt với một người hướng nội.
 - Ở Factor 2:
 - Biến có hệ số > 0 : group_comfort, friendliness
 - Biến có hệ số < 0 : deep_reflection, planning => Đây là nhóm tính cách vừa có một chút của hướng ngoại là thoải mái trong nhóm, thân thiện nhưng cũng có chút của hướng nội là suy nghĩ nội tâm, lập kế hoạch. Từ đây ta có thể thấy factor 2 chính là nhân tố ẩn của một người khi họ vừa có chút tính cách hướng ngoại vừa có chút tính cách hướng nội hay còn gọi với một cái tên là Ambiver (người trung tính), là sự kết hợp giữa một chút hướng ngoại một chút hướng nội: đây là người có sự hòa đồng, thân thiện nhưng lại có cái nhìn rất sâu sắc đôi khi và bộ dữ liệu đã thể hiện đây là một người vừa có yếu tố xã hội vừa có yếu tố nội tâm. Mặc dù là hệ số của cả 4 biến không quá cao nhưng đầu đó ta có thể nhận dạng được sự pha trộn này.

Kết luận: Từ bộ dữ liệu ta đã nhận dạng được một cách rõ ràng của người hướng nội và hướng ngoại, nhưng phải qua phân tích nhân tố FA mà ta thấy rõ được một tính cách tiềm ẩn của một người trung tính là sự kết hợp một chút hướng ngoại và một chút hướng nội thông qua 2 yếu tố là xã hội và nội tâm.