

Preferencje kulturalne - PCA

Szymon Talaga

13.12.2014

Raport przedstawia procedurę redukcji danych dotyczących preferencji kulturalnych. Dokładniej rzecz biorąc zredukowane zostaną wymiary wszystkich baterii pytań z 7-stopniową skalą lickertowską dotyczących preferencji względem książek, muzyki, prasy itp. Posłuży do tego Analiza Głównych Składowych. Podyktowane jest to tym, że celem jest przede wszystkim redukcja danych przy zachowaniu jak największej ilości wariacji oraz struktur korelacji (dlatego wybrano PCA a nie analizę czynnikową). Również z tego powodu zastosowana zostanie rotacja ukośna quartimin, która daje zwykle najklarowniejsze rozwiązania (simple structure) zachowując przy tym najwierniej korelacje międzyczynnikowe (Browne, 2001).

Redukcja danych jest konieczna z tego względu, że informacje dotyczące preferencji kulturalnych będą później użyte do klasyfikacji respondentów na podstawie analizy skupień. W związku z tym ważne jest, aby zmniejszyć liczbę wymiarów (przy minimalnej utracie informacji), żeby uchronić się przed kłutwą wymiarowości, która polega na tym, że wraz ze zwiększaniem się wymiarowości danej przestrzeni zaczynają się zacierać różnice między dużymi i małymi dystansami między obiektami, a właśnie na tych dystansach bazują wszystkie procedury analizy skupień. Jednocześnie redukcja danych przy zastosowaniu rotacji ortogonalnych może doprowadzić do przekłamania rzeczywistych relacji między obiektami (Walesiak, 2009). Stąd wynika, że optymalnym wyborem jest rotacja quartimin.

We wszystkich przypadkach optymalna ilość wymiarów rozwiązania PCA będzie ustalana na podstawie porównywania wartości własnych wymiarów z wartościami własnymi rozwiązania dla tzw. “bardzo prostej struktury”, czyli dla danych o tych samych średnich i wariancjach, ale nieskorelowanych ze sobą. Rozwiązanie to jest zaimplementowane w postaci symulacji Monte Carlo w pakiecie *psych* dostępnym w środowisku R (Revelle, 2014).

Finalnym produktem tej analizy będą wyniki czynnikowe respondentów, które potem posłużą za podstawę analizy skupień.

Jak zwykle opisy zmiennych można znaleźć w plikach “Kwestionariusz.pdf” i “Codebook.pdf”.

Wczytanie głównego zbioru danych oraz niezbędnych pakietów i funkcji:

```
load("../MainData/MainData4.RData")
D.back = D
library(psych)
library(doBy)

## Loading required package: survival
## Loading required package: splines

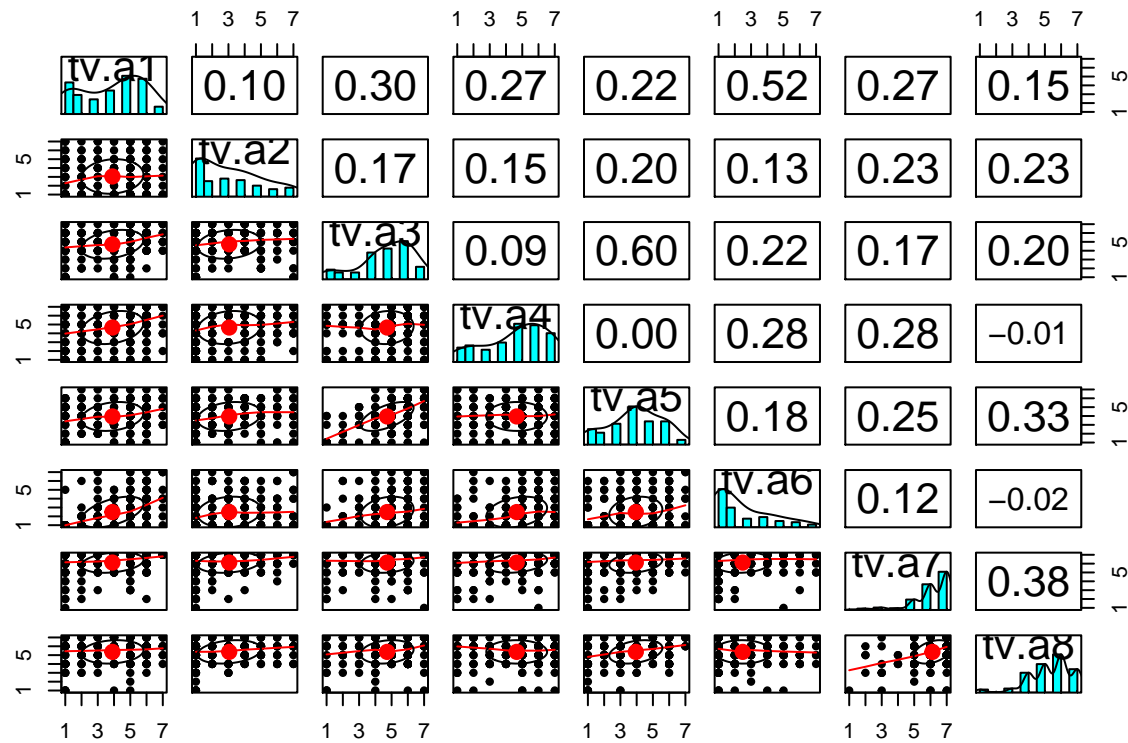
library(lattice)
library(GPArotation)
source("../HelperFunctionsMisc/ComputingMisc.R")
```

Preferencje względem telewizji

Bateria: tv.a (1-8) : preferencje względem różnych rodzajów programów telewizyjnych

Przygotowanie danych i korelogramy dla pozycji:

```
D = D[,which(names(D)=="tv.a1"):which(names(D)=="tv.b9")]
D.a = D[,which(names(D)=="tv.a1"):which(names(D)=="tv.a8")]
pairs.panels(D.a)
```

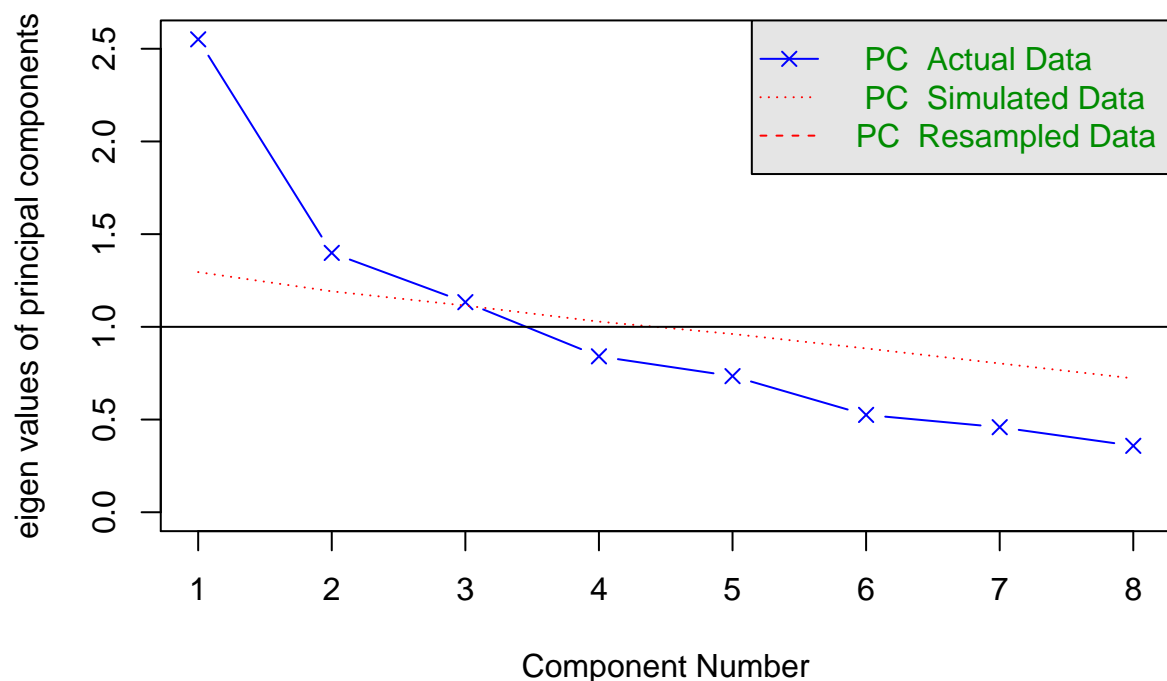


Optymalna ilość wymiarów:

```
fa.parallel(D.a, fa="pc")
```

```
## Loading required package: parallel
## Loading required package: MASS
```

Parallel Analysis Scree Plots



Parallel analysis suggests that the number of factors = NA and the number of components = 3

Jak widać rozwiązanie 3-czynnikowe jest optymalne.

PCA:

```
pca.tv.a = principal(D.a, nfactors=3, rotate="quartimin")
pca.tv.a = fa.sort(pca.tv.a)
pca.tv.a
```

```
## Principal Components Analysis
## Call: principal(r = D.a, nfactors = 3, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          TC2  TC1  TC3  h2  u2
## tv.a6  0.82  0.16 -0.14 0.70 0.30
## tv.a1  0.73  0.23  0.05 0.64 0.36
## tv.a4  0.64 -0.35  0.33 0.61 0.39
## tv.a5  0.00  0.83  0.16 0.77 0.23
## tv.a3  0.18  0.82 -0.02 0.73 0.27
## tv.a7  0.16 -0.03  0.77 0.64 0.36
## tv.a8 -0.25  0.28  0.70 0.63 0.37
## tv.a2  0.05  0.04  0.58 0.36 0.64
##
##
##          TC2  TC1  TC3
## SS loadings      1.76 1.70 1.62
## Proportion Var    0.22 0.21 0.20
## Cumulative Var    0.22 0.43 0.64
## Proportion Explained 0.35 0.33 0.32
```

```
## Cumulative Proportion 0.35 0.68 1.00
##
## With component correlations of
##      TC2  TC1  TC3
## TC2 1.00 0.12 0.15
## TC1 0.12 1.00 0.20
## TC3 0.15 0.20 1.00
##
## Test of the hypothesis that 3 components are sufficient.
##
## The degrees of freedom for the null model are 28 and the objective function was 1.53
## The degrees of freedom for the model are 7 and the objective function was 0.79
## The total number of observations was 231 with MLE Chi Square = 177.15 with prob < 7.8e-35
##
## Fit based upon off diagonal values = 0.81
```

Pytanie dotyczące oglądania kanałów sportowych (tv.a2) ma niski zasób zmienności wspólnej (h^2), więc najlepiej je odrzucić i powtórzyć analizę:

```
D.a = D.a[, -which(names(D.a)=="tv.a2")]
pca.tv.a = principal(D.a, nfactors=3, rotate="quartimin")
pca.tv.a = fa.sort(pca.tv.a)
pca.tv.a
```

```
## Principal Components Analysis
## Call: principal(r = D.a, nfactors = 3, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      TC2  TC1  TC3  h2  u2
## tv.a6 0.81 0.18 -0.18 0.71 0.29
## tv.a1 0.72 0.23 0.09 0.64 0.36
## tv.a4 0.66 -0.33 0.32 0.61 0.39
## tv.a5 0.00 0.83 0.16 0.77 0.23
## tv.a3 0.17 0.83 -0.03 0.74 0.26
## tv.a7 0.18 -0.04 0.83 0.74 0.26
## tv.a8 -0.24 0.27 0.74 0.69 0.31
##
##      TC2  TC1  TC3
## SS loadings      1.76 1.69 1.45
## Proportion Var    0.25 0.24 0.21
## Cumulative Var    0.25 0.49 0.70
## Proportion Explained 0.36 0.35 0.30
## Cumulative Proportion 0.36 0.70 1.00
##
## With component correlations of
##      TC2  TC1  TC3
## TC2 1.00 0.12 0.13
## TC1 0.12 1.00 0.18
## TC3 0.13 0.18 1.00
##
## Test of the hypothesis that 3 components are sufficient.
##
## The degrees of freedom for the null model are 21 and the objective function was 1.42
## The degrees of freedom for the model are 3 and the objective function was 0.87
```

```
## The total number of observations was 231 with MLE Chi Square = 195.94 with prob < 3.2e-42
##
## Fit based upon off diagonal values = 0.83
```

Teraz rozwiązanie jest dobrej jakości (zachowuje 70% wariancji) i można przystąpić do jego interpretacji.

Wymiar I (24% wariancji) : tv.a5, tv.a3 (wiadomości lokalne i ogólne) : jednoznacznie widać, że odnosi się on preferowania telewizji jako źródła informacji i tak właśnie będzie rozumiany. Nazwany zostaje tvinfo.

Wymiar II (25% wariancji) : tv.a6, tv.a1, tv.a4 (programy reality show, rozrywkowe i seriale) : ten wymiar zdaje się odpowiadać gustowi zorientowanemu na proste i lekkie programy rozrywkowe i seriale. Nazwany zostanie tvlowbrow.

Wymiar III (21% wariancji) : tv.a7, tv.a8 (filmy i programy kulturalne/edukacyjne) : ten czynnik najbliższy jest preferowaniu bardziej wymagających produktów telewizyjnych, które wymagają bardziej wyętej uwagi. Zostanie nazwany tvhighbrow.

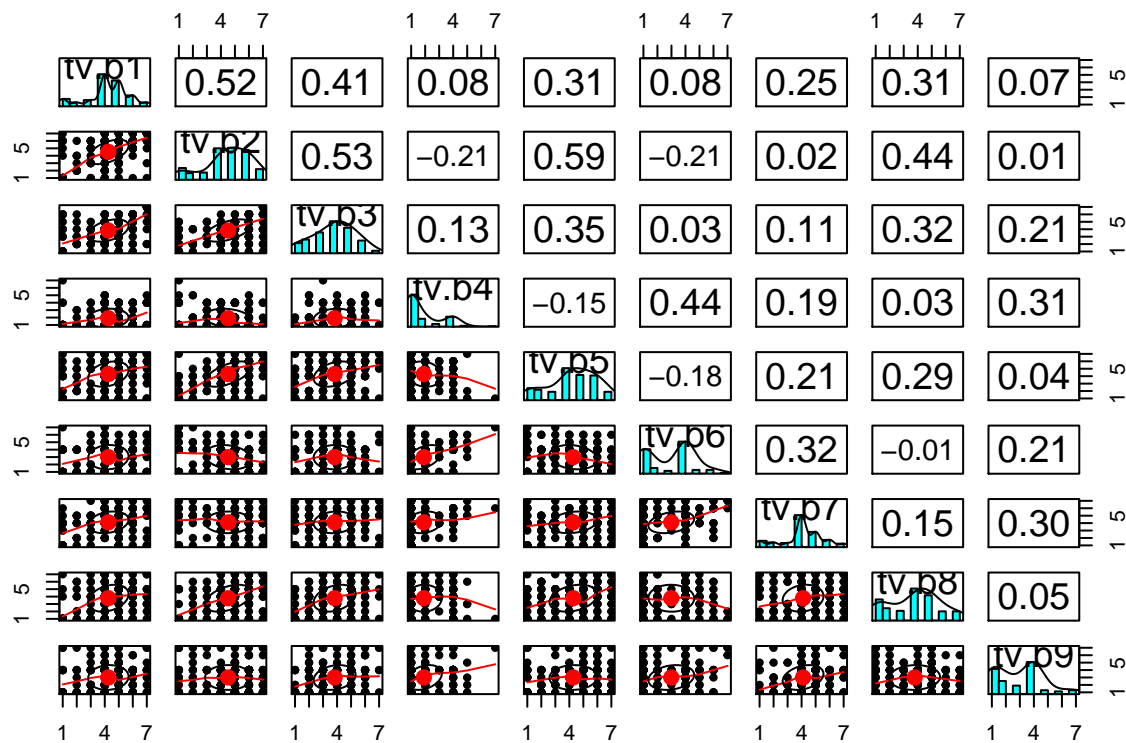
Zapisanie wyników czynnikowych respondentów:

```
D.back$tvinfo = pca.tv.a$scores[,1]
D.back$tvlowbrow = pca.tv.a$scores[,2]
D.back$tvhighbrow = pca.tv.a$scores[,3]
```

Bateria: tv.b1 do tv.b9 - preferencje względem polskich stacji telewizyjnych

Przygotowanie danych i korelogramy dla pozycji:

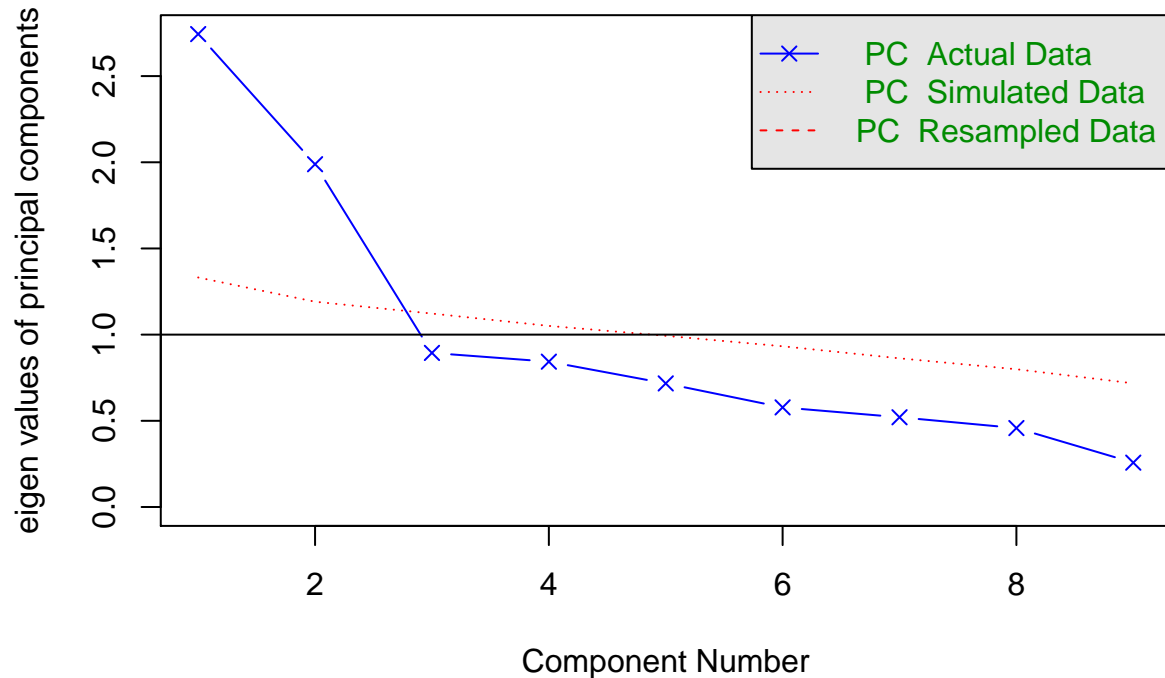
```
D.b = D[, which(names(D)=="tv.b1"):which(names(D)=="tv.b9")]
pairs.panels(D.b)
```



Optymalna ilość wymiarów:

```
fa.parallel(D.b, fa="pc")
```

Parallel Analysis Scree Plots



Parallel analysis suggests that the number of factors = NA and the number of components = 2

Jak widać optymalną liczbą wymiarów rozwiązania jest 2.

PCA:

```
pca.tv.b = principal(D.b, nfactors=2, rotate="quartimin")
pca.tv.b = fa.sort(pca.tv.b)
```

Rozwiązanie jest dobrej jakości (zachowuje 53% wariancji) i można przystąpić do jego interpretacji:

Wymiar I (30% wariancji) : tv.b2, tv.b5, tv.b3, tv.b1, tv.b8 (TVN, TVN24, Polsat, TVP, TVN Style) : ten wymiar raczej dosyć jednoznacznie odpowiada preferowaniu standardowych ogólnodostępnych polskich stacji telewizyjnych. Zostanie nazwany tvpop.

Wymiar II (23% wariancji) : tv.b4, tv.b6, tv.b9, tv.b7 (TV Trwam, TV Republika, Polsat Sport, TVP Historia) : ten czynnik ma odrobinę mniej oczywistą interpretację. Przede wszystkim skupia dwie najważniejsze stacje telewizyjne o bardziej konserwatywnym. Jednocześnie odnosi się do preferowania Polsat Sport i TVP Historia, co w pewnym stopniu odpowiada wybieraniu bardziej specyficznych programów zamiast typowej rozrywki telewizyjnej. Zostanie nazwany tvconst.

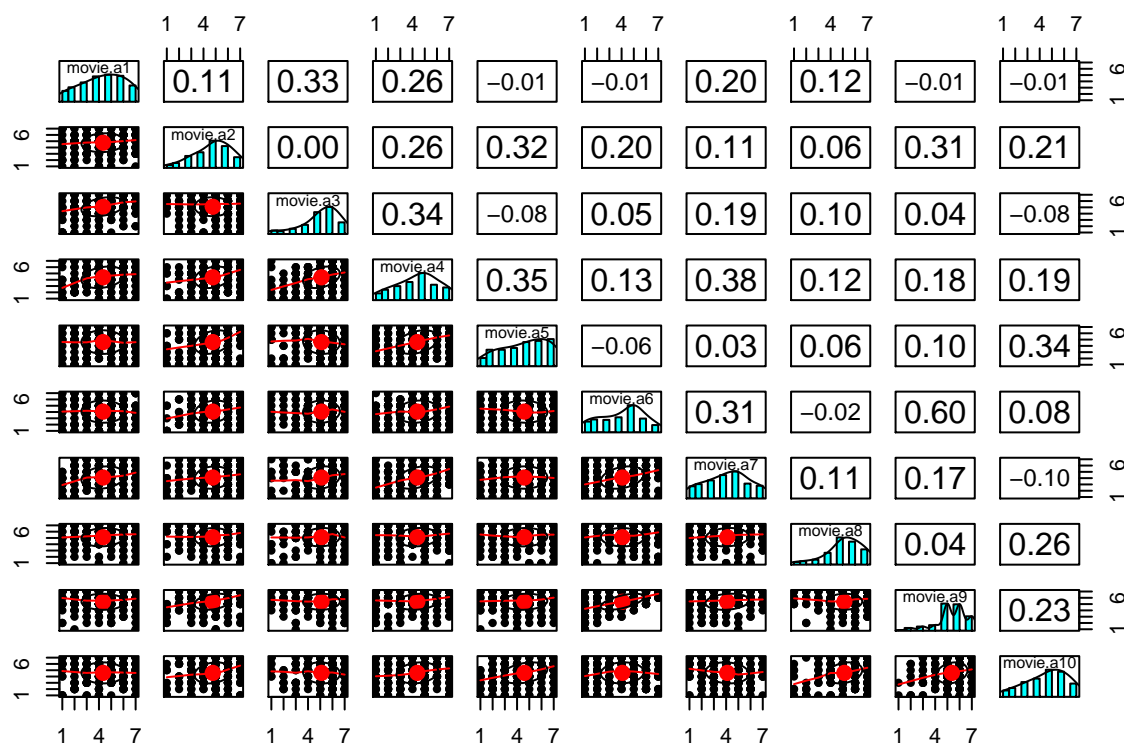
Zapisanie wyników czynnikowych respondentów:

```
D.back$tvpop = pca.tv.b$scores[,1]
D.back$tvconst = pca.tv.b$scores[,2]
```

Preferencje względem filmów i kina

Przygotowanie danych i korelogramy dla pozycji:

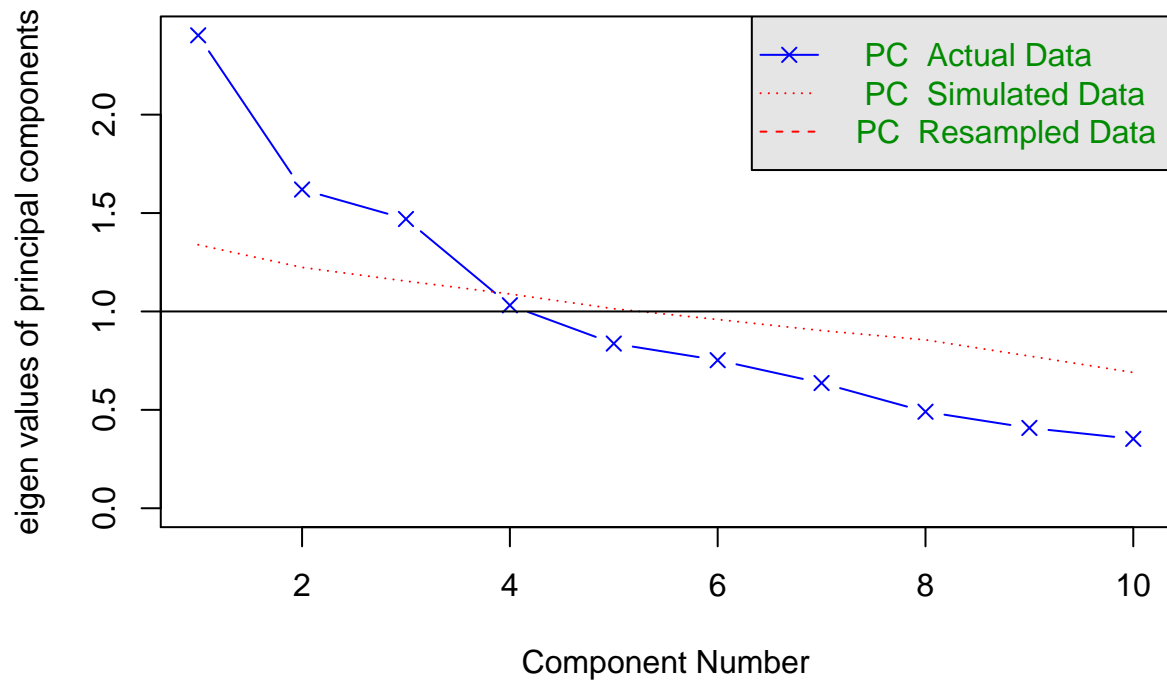
```
D = D.back[, which(names(D.back)== "movie.a1"):which(names(D.back)== "movie.a10")]
pairs.panels(D)
```



Optymalna ilość wymiarów:

```
fa.parallel(D, fa="pc")
```

Parallel Analysis Scree Plots



Parallel analysis suggests that the number of factors = NA and the number of components = 3

PCA:

```
pca.movies = principal(D, nfactors=3, rotate="quartimin")
pca.movies = fa.sort(pca.movies)
pca.movies
```

```
## Principal Components Analysis
## Call: principal(r = D, nfactors = 3, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          TC2  TC3  TC1  h2  u2
## movie.a3  0.74 -0.06 -0.16 0.55 0.45
## movie.a1  0.70 -0.15 -0.01 0.49 0.51
## movie.a4  0.64  0.10  0.39 0.63 0.37
## movie.a7  0.56  0.37 -0.13 0.50 0.50
## movie.a6  0.01  0.90 -0.12 0.79 0.21
## movie.a9 -0.05  0.82  0.16 0.73 0.27
## movie.a5  0.02 -0.09  0.78 0.60 0.40
## movie.a10 -0.12  0.04  0.75 0.57 0.43
## movie.a2  0.08  0.32  0.51 0.42 0.58
## movie.a8  0.27 -0.16  0.35 0.21 0.79
##
##          TC2  TC3  TC1
## SS loadings      1.88 1.81 1.80
## Proportion Var    0.19 0.18 0.18
## Cumulative Var    0.19 0.37 0.55
## Proportion Explained 0.34 0.33 0.33
```



```
## Cumulative Proportion 0.34 0.67 1.00
##
## With component correlations of
##      TC2  TC3  TC1
## TC2 1.00 0.12 0.06
## TC3 0.12 1.00 0.13
## TC1 0.06 0.13 1.00
##
## Test of the hypothesis that 3 components are sufficient.
##
## The degrees of freedom for the null model are 45 and the objective function was 1.79
## The degrees of freedom for the model are 18 and the objective function was 0.76
## The total number of observations was 231 with MLE Chi Square = 170.69 with prob < 6.6e-27
##
## Fit based upon off diagonal values = 0.74
```

Kino noir i kryminały mają bardzo niski zasób zmienności wspólnej, więc zostają usunięta a analiza powtórzona:

PCA:

```
D = D[,-which(names(D)=="movie.a8")]
pca.movies = principal(D, nfactors=3, rotate="quartimin")
pca.movies = fa.sort(pca.movies)
pca.movies
```

```
## Principal Components Analysis
## Call: principal(r = D, nfactors = 3, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      TC2  TC3  TC1  h2  u2
## movie.a3  0.74 -0.06 -0.16 0.55 0.45
## movie.a1  0.70 -0.15 -0.01 0.49 0.51
## movie.a4  0.65  0.08  0.42 0.67 0.33
## movie.a7  0.56  0.37 -0.13 0.50 0.50
## movie.a6  0.01  0.90 -0.12 0.81 0.19
## movie.a9 -0.05  0.83  0.16 0.74 0.26
## movie.a5  0.04 -0.12  0.83 0.68 0.32
## movie.a10 -0.16  0.08  0.70 0.52 0.48
## movie.a2  0.09  0.30  0.55 0.45 0.55
##
##      TC2  TC3  TC1
## SS loadings      1.84 1.81 1.75
## Proportion Var    0.20 0.20 0.19
## Cumulative Var    0.20 0.41 0.60
## Proportion Explained 0.34 0.34 0.32
## Cumulative Proportion 0.34 0.68 1.00
##
## With component correlations of
##      TC2  TC3  TC1
## TC2 1.00 0.12 0.08
## TC3 0.12 1.00 0.11
## TC1 0.08 0.11 1.00
##
## Test of the hypothesis that 3 components are sufficient.
```

```
##
## The degrees of freedom for the null model are 36 and the objective function was 1.68
## The degrees of freedom for the model are 12 and the objective function was 0.7
## The total number of observations was 231 with MLE Chi Square = 156.68 with prob < 2.5e-27
##
## Fit based upon off diagonal values = 0.78
```

Teraz rozwiązanie jest dobrej jakości (zachowuje 59% wariancji), zatem można przystąpić do jego interpretacji:

Wymiar I (19% wariancji) : movie.a5, movie.a10, movie.a2, movie.a4 (fantasy/sci-fi, kino wojenne, animacje etc., filmy historyczne/kostiumowe) : ten wymiar wydaje się odpowiadać dosyć jednoznacznie preferowaniu fantastyki i science fiction oraz generalnie wszelkiego rodzaju epickich narracji i fabuł. Zostanie nazwany moviesepic.

Wymiar II (20% wariancji) : movie.a3, movie.a1, movie.a4, movie.a7 (kino artystyczne, dramaty, filmy historyczne/kostiumowe, musicale) : ten wymiar jest odrobinę trudniejszy do interpretacji. Przed wszystkim odnosi się do preferowania bardziej wymagających filmów. Jednocześnie musicale niekoniecznie można zaklasyfikować do grupy kina wymagającego. Z drugiej strony jest to gatunek dosyć różnorodny, więc niektóre z nich na pewno można uznać za pasujące do tego wymiaru. Zostanie on zatem nazwany movieshigh.

Wymiar III (20% wariancji) : tv.a6, tv.a9 (komedie romantyczne i komedie) : ten wymiar z pewnością związany jest z preferowaniem filmów lżejszych i łatwiejszych w odbiorze. Nazwany zostanie w takim razie movieslight.

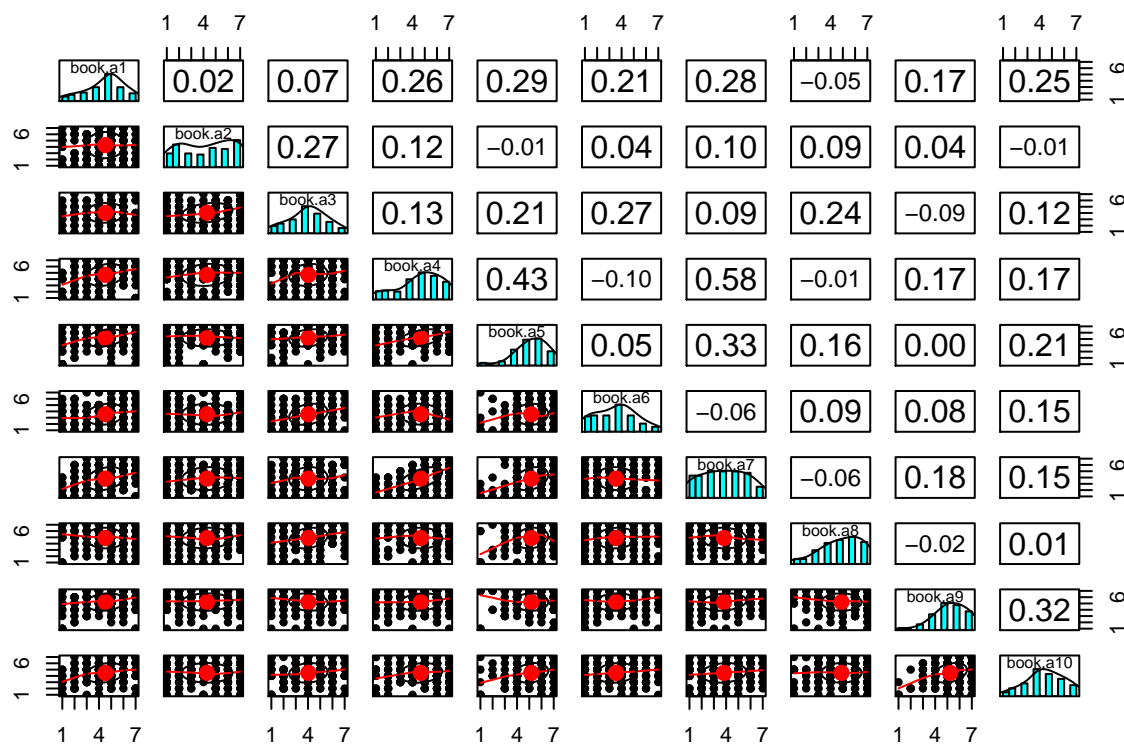
Zapisanie wyników czynnikowych respondentów:

```
D.back$moviesepic = pca.movies$scores[,1]
D.back$movieshigh = pca.movies$scores[,2]
D.back$movieslight = pca.movies$scores[,3]
```

Preferencje wobec książek i literatury

Przygotowanie danych i korelogramy dla pozycji:

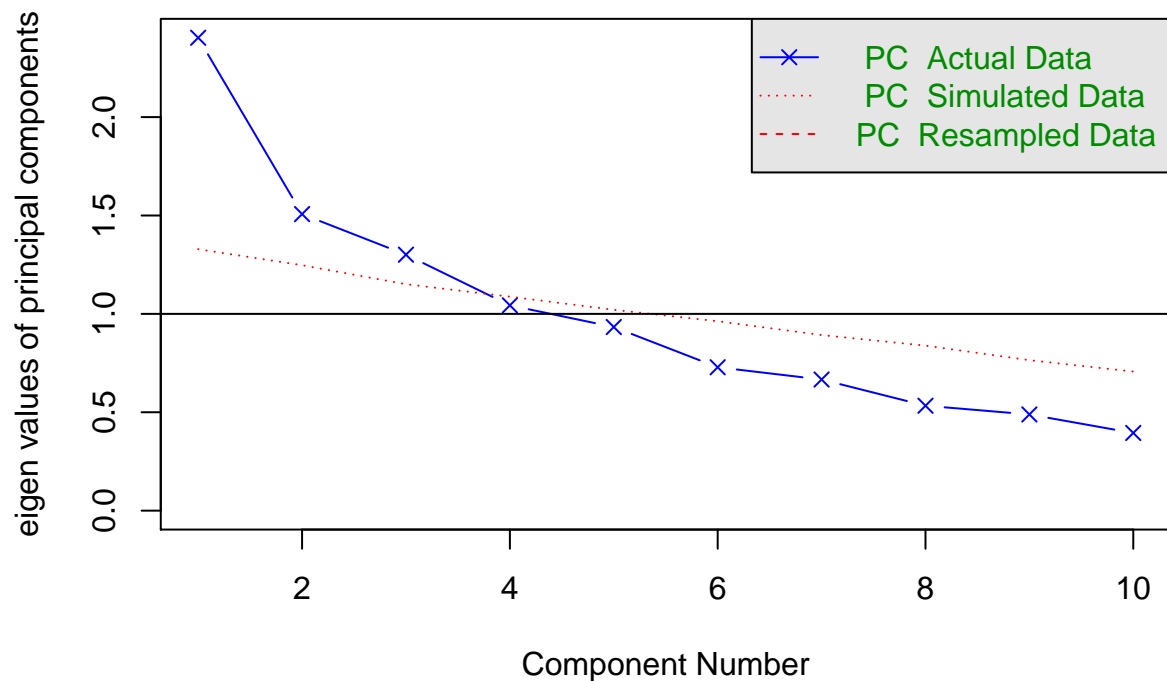
```
D = D.back[,which(names(D.back)=="book.a1"):which(names(D.back)=="book.a10")]
pairs.panels(D)
```



Optymalna liczba wymiarów:

```
fa.parallel(D, fa="pc")
```

Parallel Analysis Scree Plots



Parallel analysis suggests that the number of factors = NA and the number of components = 3

Symulacja sugeruje rozwiązanie 3-wymiarowe, ale jak widać na wykresie 4-ty wymiar jest minimalnie poniżej wartości dla zmiennych losowych.

PCA:

```
pca.books = principal(D, nfactors=3, rotate="quartimin")
pca.books = fa.sort(pca.books)
pca.books

## Principal Components Analysis
## Call: principal(r = D, nfactors = 3, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          TC1  TC2  TC3  h2  u2
## book.a4   0.85  0.04  0.00  0.73  0.27
## book.a7   0.80 -0.05  0.06  0.66  0.34
## book.a5   0.61  0.30  0.07  0.49  0.51
## book.a3   0.12  0.80  0.01  0.65  0.35
## book.a8  -0.05  0.62 -0.09  0.38  0.62
## book.a2   0.17  0.47 -0.15  0.25  0.75
## book.a10  0.12 -0.01  0.70  0.52  0.48
## book.a9   0.09 -0.28  0.64  0.47  0.53
## book.a6  -0.33  0.41  0.57  0.61  0.39
## book.a1   0.33  0.00  0.54  0.44  0.56
##
##          TC1  TC2  TC3
## SS loadings      2.07 1.57 1.58
## Proportion Var    0.21 0.16 0.16
## Cumulative Var    0.21 0.36 0.52
## Proportion Explained 0.40 0.30 0.30
## Cumulative Proportion 0.40 0.70 1.00
##
## With component correlations of
##          TC1  TC2  TC3
## TC1 1.00 0.01 0.14
## TC2 0.01 1.00 0.13
## TC3 0.14 0.13 1.00
##
## Test of the hypothesis that 3 components are sufficient.
##
## The degrees of freedom for the null model are 45 and the objective function was 1.47
## The degrees of freedom for the model are 18 and the objective function was 0.74
## The total number of observations was 231 with MLE Chi Square = 166.19 with prob < 5.1e-26
##
## Fit based upon off diagonal values = 0.67
```

Widać, że aż dwie pozycje (book.a2 i book.a8) mają dosyć niskie zasoby zmienności wspólnej. Jednocześnie wiadomo, że rozwiązanie 4-czynnikowe również może być dobre, dlatego warto je sprawdzić.

PCA:

```
pca.books = principal(D, nfactors=4, rotate="quartimin")
pca.books = fa.sort(pca.books)
pca.books
```

```
## Principal Components Analysis
## Call: principal(r = D, nfactors = 4, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          TC1   TC2   TC3   TC4   h2   u2
## book.a4  0.84 -0.10  0.06  0.12  0.74  0.26
## book.a7  0.78 -0.16  0.13  0.11  0.67  0.33
## book.a5  0.71  0.34 -0.14 -0.25  0.69  0.31
## book.a3  0.14  0.69 -0.13  0.37  0.66  0.34
## book.a6 -0.27  0.67  0.36 -0.07  0.63  0.37
## book.a8  0.00  0.57 -0.26  0.12  0.40  0.60
## book.a9  0.03 -0.15  0.79  0.16  0.66  0.34
## book.a10 0.15  0.24  0.63 -0.11  0.52  0.48
## book.a1  0.39  0.22  0.41 -0.26  0.48  0.52
## book.a2  0.04  0.12  0.06  0.88  0.80  0.20
##
##          TC1   TC2   TC3   TC4
## SS loadings      2.10  1.55  1.47  1.13
## Proportion Var    0.21  0.16  0.15  0.11
## Cumulative Var    0.21  0.37  0.51  0.63
## Proportion Explained 0.34  0.25  0.23  0.18
## Cumulative Proportion 0.34  0.58  0.82  1.00
##
## With component correlations of
##          TC1   TC2   TC3   TC4
## TC1 1.00 0.09  0.11  0.06
## TC2 0.09 1.00  0.04  0.00
## TC3 0.11 0.04  1.00 -0.01
## TC4 0.06 0.00 -0.01  1.00
##
## Test of the hypothesis that 4 components are sufficient.
##
## The degrees of freedom for the null model are 45 and the objective function was 1.47
## The degrees of freedom for the model are 11 and the objective function was 1.08
## The total number of observations was 231 with MLE Chi Square = 240.38 with prob < 2.9e-45
##
## Fit based upon off diagonal values = 0.71
```

Teraz zasoby zmienności wspólnej są dosyć wysokie, ale book.a1 (biografie i książki historyczne) nie ma jednoznacznie przypisanego dominującego wymiaru, więc warto powtórzyć analizę bez tej pozycji.

PCA:

```
D = D[, -which(names(D)=="book.a1")]
pca.books = principal(D, nfactors=4, rotate="quartimin")
pca.books = fa.sort(pca.books)
```

Teraz rozwiązanie jest dobrej jakości (zachowuje 66% wariancji), zatem można przystąpić do jego interpretacji:

Wymiar I (22% wariancji) : book.a4, book.a7, book.a5 (klasyczny kanon literacki, literatura współczesna, poezja) : ten wymiar wydaje się jednoznacznie odpowiadać dosyć klasycznemu gustowi literackiemu. Zostanie zatem nazwany bookshigh.

Wymiar II (17% wariancji) : book.a3, book.a6, book.a8 (literatura młodzieżowa, poradniki, kryminały) : ten wymiar zdaje się odpowiadać preferencji względem literatury lekkiej i łatwiejszej w odbiorze. Zatem nazwano go bookslow.

Wymiar III (15% wariacji) : book.a6, book.a9, book.a10 (poradniki, książki popularno-naukowe, książki podróżnicze) : wymiar odpowiada gustowi skupionemu na informacji. Nazwany zostaje booksknow.

Wymiar IV (12% wariacji) : book.a2 (fantasy, science-fiction) : ten wymiar tworzony jest przez tylko jedną pozycję, jaką są książki fantasy i science-fiction. Odpowiada oczywiście silnym preferencjom w tym zakresie. Nazwany zostaje booksfant.

Zapisanie wyników czynnikowych respondentów:

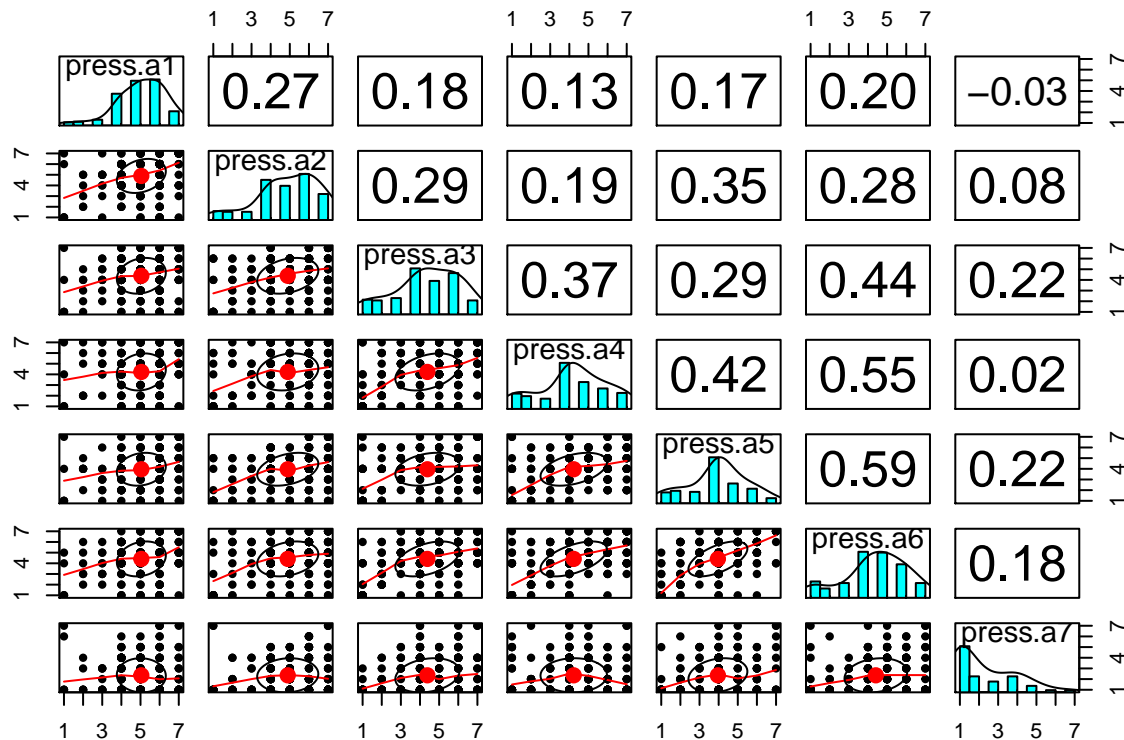
```
D.back$bookshigh = pca.books$scores[,1]
D.back$bookslow = pca.books$scores[,2]
D.back$booksknow = pca.books$scores[,3]
D.back$booksfant = pca.books$scores[,4]
```

Preferencje względem prasy

Bateria: press.a1 do press.a7 - preferencje względem rodzajów prasy

Przygotowanie danych i korelogramy dla pozycji:

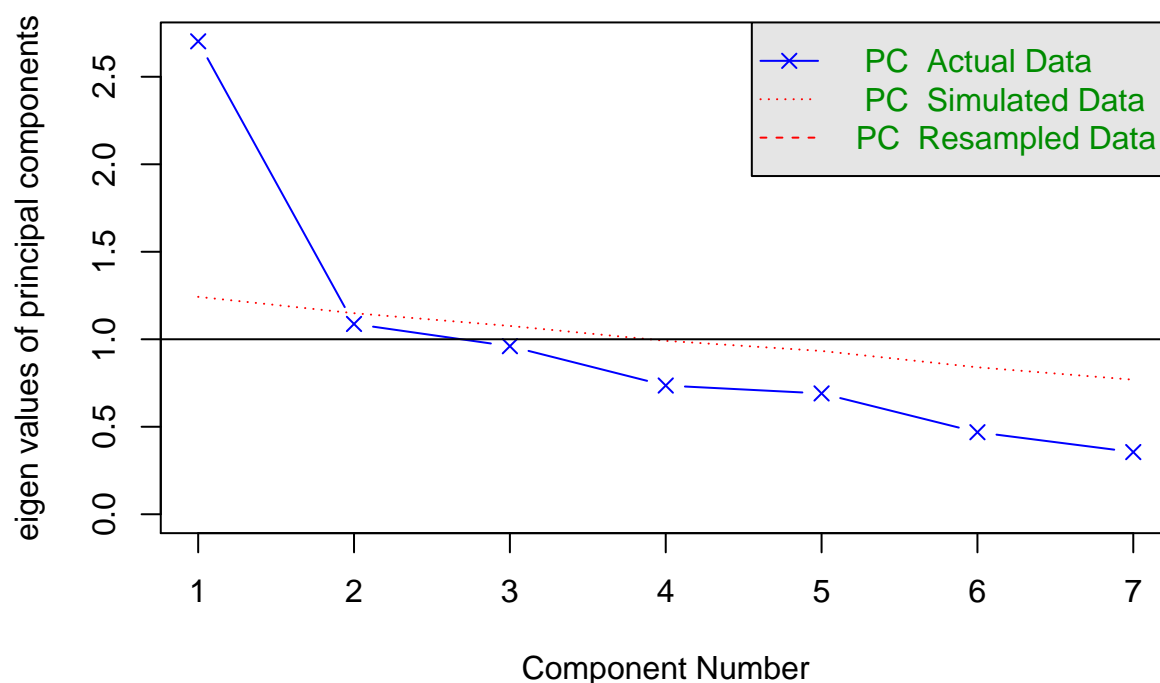
```
D = D.back[, which(names(D.back)=="press.a1"):which(names(D.back)=="press.b10")]
D.a = D[, 1:which(names(D)=="press.a7")]
pairs.panels(D.a)
```



Optymalna liczba czynników:

```
fa.parallel(D.a, fa="pc")
```

Parallel Analysis Scree Plots



Parallel analysis suggests that the number of factors = NA and the number of components = 1

Symulacja wskazuje na jeden wymiar, ale rozwiązanie dwuwymiarowe również wygląda na prawdopodobne.

PCA:

```
pca.press.a = principal(D.a, nfactors=1, rotate="quartimin")
pca.press.a
```

```
## Principal Components Analysis
## Call: principal(r = D.a, nfactors = 1, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          PC1    h2    u2
## press.a1 0.38 0.145 0.86
## press.a2 0.55 0.307 0.69
## press.a3 0.66 0.441 0.56
## press.a4 0.69 0.477 0.52
## press.a5 0.76 0.571 0.43
## press.a6 0.82 0.674 0.33
## press.a7 0.30 0.088 0.91
##
##          PC1
## SS loadings  2.70
## Proportion Var 0.39
##
## Test of the hypothesis that 1 component is sufficient.
##
## The degrees of freedom for the null model are  21  and the objective function was  1.43
```

```
## The degrees of freedom for the model are 14 and the objective function was 0.29
## The total number of observations was 231 with MLE Chi Square = 65.57 with prob < 1.2e-08
##
## Fit based upon off diagonal values = 0.85
```

Występują ekstremalnie niskie zasoby zmienności wspólnej. Należy powtórzyć analizę z rozwiązaniem dwuwymiarowym.

PCA:

```
pca.press.a = principal(D.a, nfactors=2, rotate="quartimin")
pca.press.a = fa.sort(pca.press.a)
pca.press.a
```

```
## Principal Components Analysis
## Call: principal(r = D.a, nfactors = 2, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          TC1   TC2   h2   u2
## press.a6 0.83 -0.01 0.68 0.32
## press.a5 0.77 -0.03 0.58 0.42
## press.a4 0.68  0.08 0.48 0.52
## press.a3 0.67 -0.03 0.45 0.55
## press.a2 0.49  0.40 0.43 0.57
## press.a1 0.26  0.71 0.59 0.41
## press.a7 0.41 -0.66 0.57 0.43
##
##          TC1   TC2
## SS loadings      2.67 1.12
## Proportion Var    0.38 0.16
## Cumulative Var    0.38 0.54
## Proportion Explained 0.71 0.29
## Cumulative Proportion 0.71 1.00
##
## With component correlations of
##          TC1   TC2
## TC1 1.00 0.06
## TC2 0.06 1.00
##
## Test of the hypothesis that 2 components are sufficient.
##
## The degrees of freedom for the null model are 21 and the objective function was 1.43
## The degrees of freedom for the model are 8 and the objective function was 0.62
## The total number of observations was 231 with MLE Chi Square = 138.81 with prob < 4.2e-26
##
## Fit based upon off diagonal values = 0.8
```

Teraz rozwiązanie jest dobre jakości (zachowuje 54% wariancji) i można zatem przystąpić do jego interpretacji:

Wymiar I (38% wariancji) : press.a6, press.a5, press.a4, press.a3, press.a2, press.a7 (prasa codzienna, prasa lokalna, opinie, kultura/styl życia, prasa hobbystyczna/branżowa, tabloidy) : ten czynnik odpowiada ogólnemu czytelnictwu prasy. Zostaje nazwany pressgen.

Wymiar II (16% wariancji) : press.a2, press.a1, -press.a7 (prasa hobbystyczna/branżowa, prasa popularno-naukowa, -tabloidy) (- oznacza silną ujemną korelację z czynnikiem) : ten wymiar zdaje się jednoznacznie odpowiadać preferowaniu prasy służącej zdobywaniu konkretnych informacji i wiedzy.

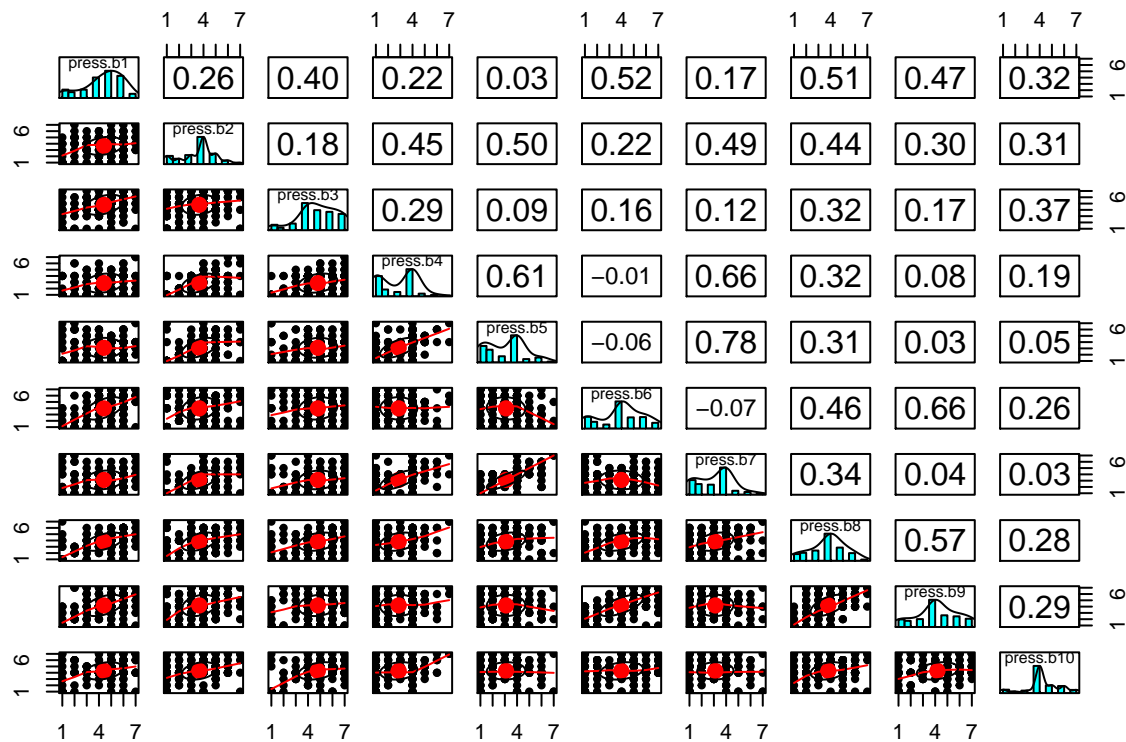
Zapisanie wyników czynnikowych respondentów:

```
D.back$pressgen = pca.press.a$scores[,1]
D.back$pressknow = pca.press.a$scores[,2]
```

Bateria: tv.b1 do tv.b10 - preferencje względem konkretnych polskich tytułów prasowych

Przygotowanie danych i korelogramy dla pozycji:

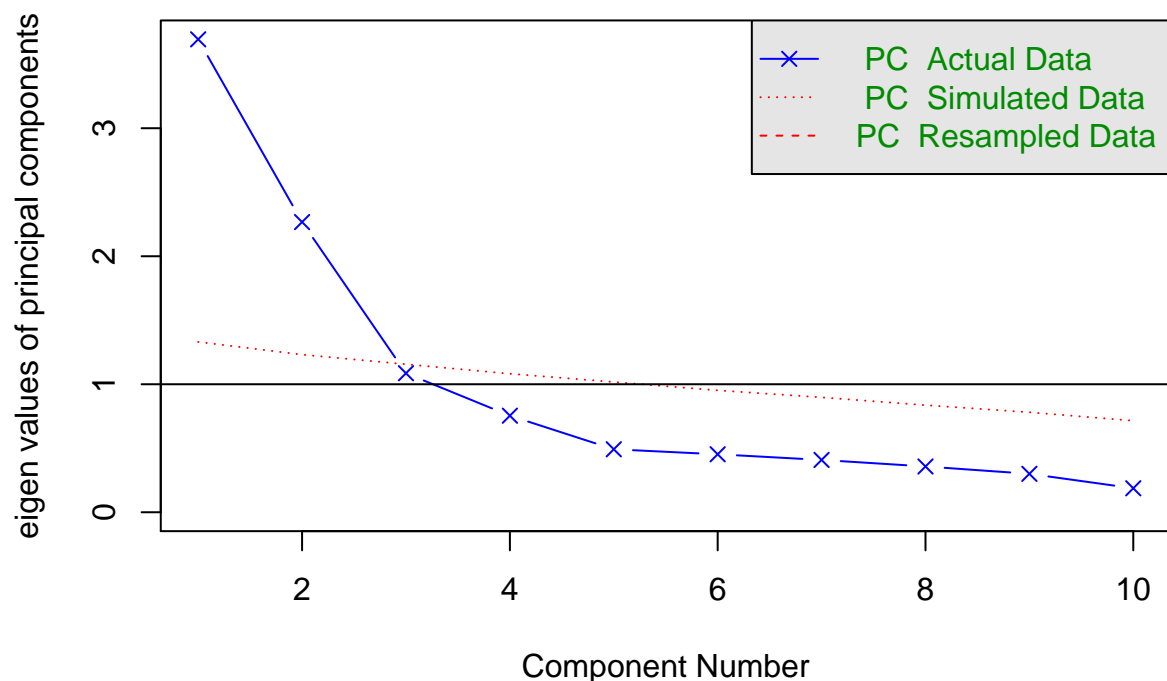
```
D.b = D[, which(names(D)=="press.b1"):which(names(D)=="press.b10")]
pairs.panels(D.b)
```



Optymalna liczba wymiarów:

```
fa.parallel(D.b, fa="pc")
```

Parallel Analysis Scree Plots



Parallel analysis suggests that the number of factors = NA and the number of components = 2

Jak widać symulacja sugeruje dwa rozwiązania, ale rozwiązanie 3-wymiarowe również może być dobre.

PCA:

```
pca.press.b = principal(D.b, nfactors=2, rotate="quartimin")
pca.press.b = fa.sort(pca.press.b)
pca.press.b
```

```
## Principal Components Analysis
## Call: principal(r = D.b, nfactors = 2, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          TC1  TC2  h2  u2
## press.b6  0.82 -0.21 0.67 0.33
## press.b9  0.82 -0.07 0.65 0.35
## press.b1  0.76  0.05 0.59 0.41
## press.b8  0.69  0.32 0.64 0.36
## press.b10 0.54  0.08 0.31 0.69
## press.b3  0.45  0.17 0.26 0.74
## press.b7 -0.05  0.91 0.81 0.19
## press.b5 -0.10  0.90 0.79 0.21
## press.b4  0.07  0.81 0.68 0.32
## press.b2  0.33  0.62 0.55 0.45
##
##          TC1  TC2
## SS loadings      3.06 2.90
## Proportion Var    0.31 0.29
```

```
## Cumulative Var      0.31 0.60
## Proportion Explained 0.51 0.49
## Cumulative Proportion 0.51 1.00
##
## With component correlations of
##      TC1  TC2
## TC1 1.00 0.15
## TC2 0.15 1.00
##
## Test of the hypothesis that 2 components are sufficient.
##
## The degrees of freedom for the null model are 45 and the objective function was 4.38
## The degrees of freedom for the model are 26 and the objective function was 0.68
## The total number of observations was 231 with MLE Chi Square = 153.09 with prob < 5.7e-20
##
## Fit based upon off diagonal values = 0.94
```

press.b3 i press.b10 (Forbes i Focus) mają bardzo niskie zasoby zmienności wspólnej. Sugeruje, to że rozwiązanie trójwymiarowe może być lepsze.

PCA:

```
pca.press.b = principal(D.b, nfactors=3, rotate="quartimin")
pca.press.b = fa.sort(pca.press.b)
pca.press.b
```

```
## Principal Components Analysis
## Call: principal(r = D.b, nfactors = 3, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           TC2  TC1  TC3  h2  u2
## press.b7  0.92 -0.02 -0.06 0.83 0.17
## press.b5  0.92 -0.04 -0.10 0.82 0.18
## press.b4  0.76 -0.08  0.25 0.71 0.29
## press.b2  0.63  0.31  0.06 0.57 0.43
## press.b9  0.00  0.89 -0.05 0.77 0.23
## press.b6 -0.14  0.89 -0.03 0.77 0.23
## press.b8  0.35  0.66  0.11 0.67 0.33
## press.b1  0.02  0.57  0.37 0.60 0.40
## press.b3  0.00 -0.07  0.89 0.75 0.25
## press.b10 -0.05  0.13  0.71 0.57 0.43
##
##           TC2  TC1  TC3
## SS loadings      2.86 2.56 1.63
## Proportion Var    0.29 0.26 0.16
## Cumulative Var    0.29 0.54 0.70
## Proportion Explained 0.41 0.36 0.23
## Cumulative Proportion 0.41 0.77 1.00
##
## With component correlations of
##      TC2 TC1 TC3
## TC2 1.0 0.1 0.2
## TC1 0.1 1.0 0.3
## TC3 0.2 0.3 1.0
##
```

```
## Test of the hypothesis that 3 components are sufficient.
##
## The degrees of freedom for the null model are 45 and the objective function was 4.38
## The degrees of freedom for the model are 18 and the objective function was 0.7
## The total number of observations was 231 with MLE Chi Square = 155.61 with prob < 6e-24
##
## Fit based upon off diagonal values = 0.96
```

Jak widać teraz rozwiązanie jest wysokiej jakości (zachowuje 70% wariancji), wobec czego można przystąpić do jego interpretacji:

Wymiar I (29% wariancji) : press.b7, press.b5, press.b4, press.b2 (Uważam Rze, Do Rzeczy, Gazeta Polska, Rzeczpospolita) : ten wymiar dosyć jednoznacznie odpowiada preferowaniu tytułów o bardziej konserwatywnym profilu. Zostaje nazwany presscons.

Wymiar II (26% wariancji) : press.b6, press.b9, press.b1, press.b8 (Gazeta Wyborcza, Polityka, Newsweek, Wprost) : ten wymiar odpowiada gustowaniu w najważniejszych, standardowych polskich tytułach prasowych.

Wymiar III (16% wariancji) : press.b3, press.b10 (Forbes, Focus) : ten wymiar wydaje się odpowiadać preferowaniu prasy specjalistycznej, która nie posiada politycznej linii redakcyjnej żadnego rodzaju.

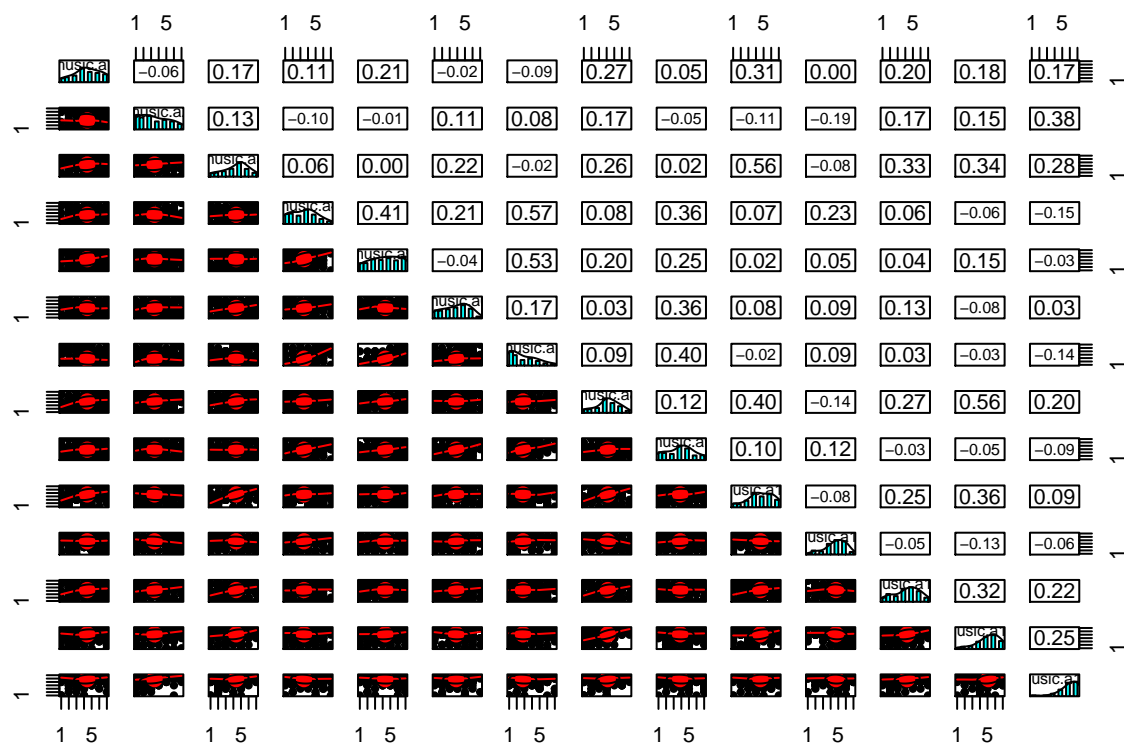
Zapisanie wyników czynnikowych dla respondentów:

```
D.back$presscons = pca.press.b$scores[,1]
D.back$presslib = pca.press.b$scores[,2]
D.back$pressspec = pca.press.b$scores[,3]
```

Preferencje względem muzyki

Przygotowanie danych i korelogramy dla pozycji:

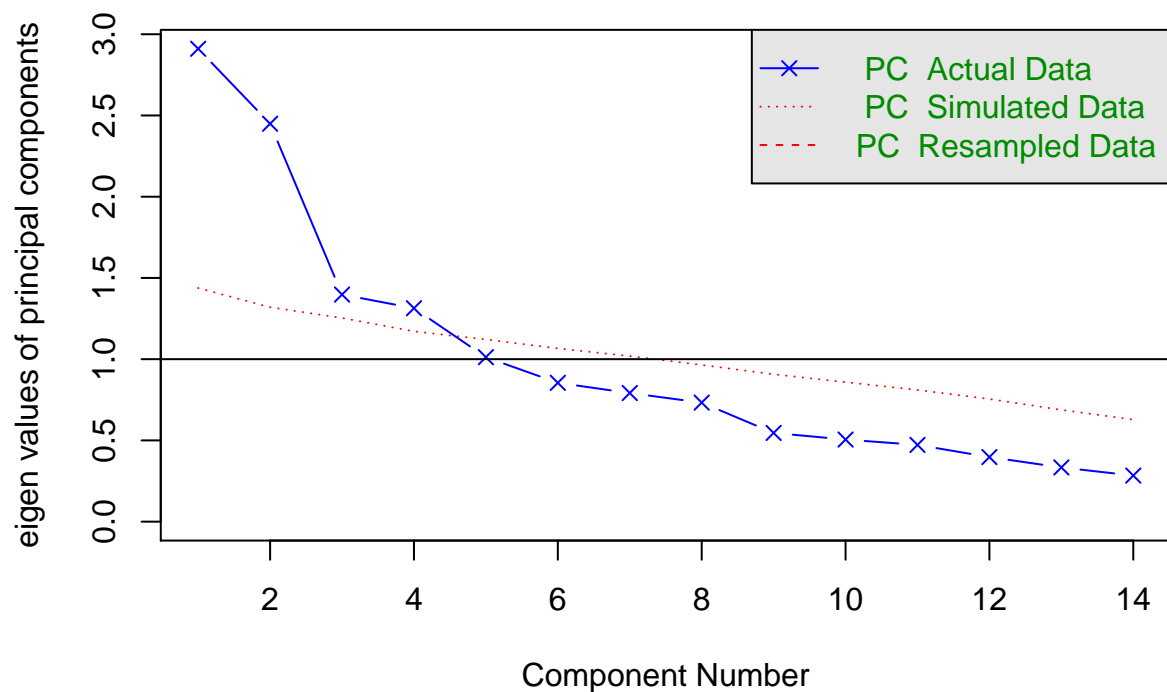
```
D = D.back[, which(names(D.back)=="music.a1"):which(names(D.back)=="music.a14")]
pairs.panels(D)
```



Optymalna liczba wymiarów:

```
fa.parallel(D, fa="pc")
```

Parallel Analysis Scree Plots



Parallel analysis suggests that the number of factors = NA and the number of components = 4

PCA:

```
pca.music = principal(D, nfactors=4, rotate="quartimin")
pca.music = fa.sort(pca.music)
pca.music
```

```
## Principal Components Analysis
## Call: principal(r = D, nfactors = 4, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          TC1  TC2  TC3  TC4  h2  u2
## music.a10  0.82 -0.07 -0.25  0.13 0.71 0.29
## music.a3   0.68 -0.14  0.10  0.36 0.62 0.38
## music.a13  0.65  0.07  0.24 -0.28 0.60 0.40
## music.a8   0.64  0.24  0.21 -0.19 0.58 0.42
## music.a1   0.57  0.10 -0.21 -0.18 0.38 0.62
## music.a12  0.49  0.00  0.26  0.15 0.36 0.64
## music.a7  -0.12  0.84  0.11  0.13 0.75 0.25
## music.a5   0.10  0.81  0.01 -0.28 0.72 0.28
## music.a4   0.05  0.71 -0.19  0.24 0.64 0.36
## music.a9   0.04  0.51 -0.10  0.45 0.51 0.49
## music.a2  -0.08  0.04  0.88  0.09 0.76 0.24
## music.a14  0.27 -0.16  0.58  0.05 0.47 0.53
## music.a11 -0.08  0.13 -0.37  0.27 0.25 0.75
## music.a6   0.05  0.07  0.12  0.83 0.72 0.28
##
##          TC1  TC2  TC3  TC4
## SS loadings      2.68 2.30 1.63 1.46
## Proportion Var    0.19 0.16 0.12 0.10
## Cumulative Var    0.19 0.36 0.47 0.58
## Proportion Explained 0.33 0.28 0.20 0.18
## Cumulative Proportion 0.33 0.62 0.82 1.00
##
## With component correlations of
##          TC1  TC2  TC3  TC4
## TC1 1.00  0.05  0.11  0.02
## TC2 0.05  1.00 -0.02  0.07
## TC3 0.11 -0.02  1.00 -0.02
## TC4 0.02  0.07 -0.02  1.00
##
## Test of the hypothesis that 4 components are sufficient.
##
## The degrees of freedom for the null model are 91 and the objective function was 3.44
## The degrees of freedom for the model are 41 and the objective function was 1.1
## The total number of observations was 231 with MLE Chi Square = 244.07 with prob < 1.1e-30
##
## Fit based upon off diagonal values = 0.85
```

Jak widać music.a11 (muzyka pop) ma bardzo niskie zasoby zmienności wspólnej, w związku z czym zostanie usunięta a analiza powtórzona:

PCA:

```
D = D[, -which(names(D)=="music.a11")]
pca.music = principal(D, nfactors=4, rotate="quartimin")
pca.music = fa.sort(pca.music)
pca.music
```

```
## Principal Components Analysis
## Call: principal(r = D, nfactors = 4, rotate = "quartimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          TC1   TC2   TC3   TC4   h2   u2
## music.a10  0.83 -0.07 -0.26  0.14 0.73 0.27
## music.a3   0.68 -0.13  0.11  0.35 0.63 0.37
## music.a13  0.64  0.07  0.26 -0.29 0.61 0.39
## music.a8   0.63  0.24  0.21 -0.18 0.57 0.43
## music.a1   0.56  0.11 -0.17 -0.21 0.38 0.62
## music.a12  0.47  0.01  0.30  0.12 0.37 0.63
## music.a7  -0.13  0.84  0.09  0.15 0.75 0.25
## music.a5   0.09  0.81  0.02 -0.29 0.73 0.27
## music.a4   0.06  0.72 -0.17  0.22 0.63 0.37
## music.a9   0.05  0.51 -0.12  0.47 0.54 0.46
## music.a2  -0.11  0.04  0.87  0.10 0.76 0.24
## music.a14  0.24 -0.15  0.65 -0.01 0.54 0.46
## music.a6   0.06  0.07  0.12  0.84 0.73 0.27
##
##          TC1   TC2   TC3   TC4
## SS loadings      2.64 2.28 1.60 1.43
## Proportion Var    0.20 0.18 0.12 0.11
## Cumulative Var    0.20 0.38 0.50 0.61
## Proportion Explained 0.33 0.29 0.20 0.18
## Cumulative Proportion 0.33 0.62 0.82 1.00
##
## With component correlations of
##          TC1   TC2   TC3   TC4
## TC1 1.00  0.05  0.12  0.02
## TC2 0.05  1.00 -0.02  0.07
## TC3 0.12 -0.02  1.00 -0.03
## TC4 0.02  0.07 -0.03  1.00
##
## Test of the hypothesis that 4 components are sufficient.
##
## The degrees of freedom for the null model are 78 and the objective function was 3.31
## The degrees of freedom for the model are 32 and the objective function was 1.11
## The total number of observations was 231 with MLE Chi Square = 247.02 with prob < 4.7e-35
##
## Fit based upon off diagonal values = 0.87
```

Teraz rozwiązanie jest dobrej jakości (zachowuje 61% wariacji), zatem można przystąpić do jego interpretacji:

Wymiar I (20% wariacji) : music.a10, music.a3, music.a13, music.a8, music.a1, music.a12 (jazz, blues, muzyka klasyczna, współczesna muzyka poważna, indie, muzyka etniczna/folkowa) : ten wymiar raczej jednoznacznie odpowiada preferencjom wobec bardziej wymagającej i skomplikowanej, tudzież nietypowej muzyki. Jako taki zostaje nazwany musichigh.

Wymiar II (18% wariacji) : music.a7, music.a5, music.a4, music.a9 (techno, muzyka elektroniczna,

house, hip-hop) : ten czynnik odpowiada preferencji względem współczesnych gatunków muzyki klubowej etc. Zostaje nazwany musicmodern.

Wymiar III (12% wariancji) : music.a2, music.a14 (heavy metal, rock) : ten wymiar jednoznacznie odpowiada preferencjom rockowo-metalowym. Zostaje nazwany musicrock.

Wymiar IV (11% wariancji) : music.a3, music.a9, music.a6 (blues, hip-hop, reggae) : ten wymiar dosyć wyraźnie odpowiada preferencji względem szeroko pojętej czarnej muzyce. Zostaje nazwany musicafro.

Zapisanie wyników czynnikowych dla respondentów:

```
D.back$musichigh = pca.music$scores[,1]
D.back$musicmodern = pca.music$scores[,2]
D.back$musicrock = pca.music$scores[,3]
D.back$musicafro = pca.music$scores[,4]
```

Koniec raportu