R Notebook

#No1

library(readxl)  
library(psych)  
library(GPArotation)  
data10.1 <- read\_excel("C:/Users/lili/Downloads/Data10Penugasan.xlsx",sheet="Sheet1")  
data10.1\_new <- data10.1[,-1]  
data10.1\_new

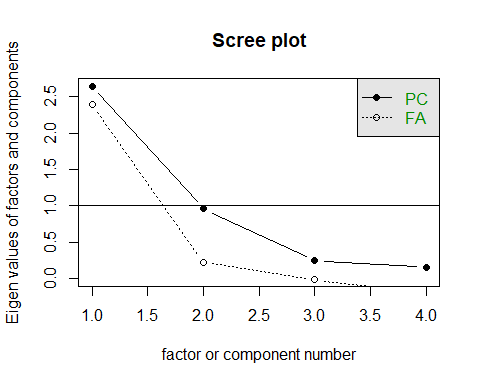
## # A tibble: 29 × 4  
## Miskin IPM TPT Gini  
## <dbl> <dbl> <dbl> <dbl>  
## 1 10.0 70.1 2.61 0.38   
## 2 37.2 58.0 2.39 0.342  
## 3 12.4 71.7 9.68 0.432  
## 4 24.2 68.8 6.31 0.349  
## 5 26.3 67.7 5.78 0.4   
## 6 24.6 72.2 10.4 10.4   
## 7 36.7 56.3 0.66 0.38   
## 8 34.7 48.4 1.78 0.376  
## 9 14.3 74.2 7.51 0.339  
## 10 19.4 61.5 3.08 0.448  
## # … with 19 more rows

eigen(cor(data10.1\_new))

## eigen() decomposition  
## $values  
## [1] 2.6448624 0.9562018 0.2476326 0.1513031  
##   
## $vectors  
## [,1] [,2] [,3] [,4]  
## [1,] 0.5210807 -0.38398472 0.7414383 0.17691771  
## [2,] -0.5745376 0.14487012 0.3004648 0.74742229  
## [3,] -0.5729825 -0.07458741 0.5151217 -0.63306980  
## [4,] -0.2647120 -0.90884823 -0.3076358 0.09634666

• Interpretasi cari eigen value >=1, maka disini hanya 2 faktor

scree(cor(data10.1\_new))

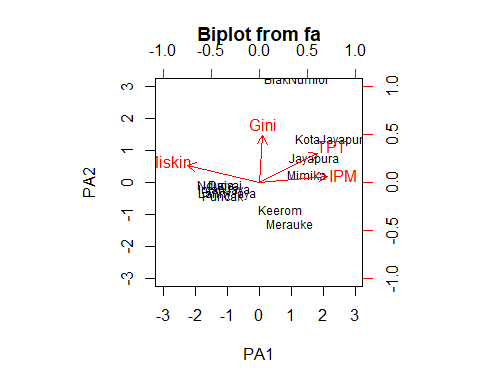


fa1 <- fa(data10.1\_new,nfactors=2,fm="pa",scores=TRUE)  
fa1

## Factor Analysis using method = pa  
## Call: fa(r = data10.1\_new, nfactors = 2, scores = TRUE, fm = "pa")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## PA1 PA2 h2 u2 com  
## Miskin -0.93 0.22 0.79 0.21 1.1  
## IPM 0.88 0.08 0.84 0.16 1.0  
## TPT 0.76 0.38 0.90 0.10 1.5  
## Gini 0.05 0.61 0.39 0.61 1.0  
##   
## PA1 PA2  
## SS loadings 2.29 0.62  
## Proportion Var 0.57 0.16  
## Cumulative Var 0.57 0.73  
## Proportion Explained 0.79 0.21  
## Cumulative Proportion 0.79 1.00  
##   
## With factor correlations of   
## PA1 PA2  
## PA1 1.00 0.32  
## PA2 0.32 1.00  
##   
## Mean item complexity = 1.2  
## Test of the hypothesis that 2 factors are sufficient.  
##   
## The degrees of freedom for the null model are 6 and the objective function was 2.36 with Chi Square of 60.87  
## The degrees of freedom for the model are -1 and the objective function was 0   
##   
## The root mean square of the residuals (RMSR) is 0   
## The df corrected root mean square of the residuals is NA   
##   
## The harmonic number of observations is 29 with the empirical chi square 0 with prob < NA   
## The total number of observations was 29 with Likelihood Chi Square = 0 with prob < NA   
##   
## Tucker Lewis Index of factoring reliability = 1.116  
## Fit based upon off diagonal values = 1  
## Measures of factor score adequacy   
## PA1 PA2  
## Correlation of (regression) scores with factors 0.97 0.83  
## Multiple R square of scores with factors 0.93 0.68  
## Minimum correlation of possible factor scores 0.86 0.36

Berdasarkan analisis faktor tersebut, kedua faktor dapat menjelaskan keragaman data sebesar 73 persen, dengan factor 1 (57 persen) dan factor 2 (16 persen). Variabel- variable yang berkorelasi tinggi dengan factor 1 adalah: presentase penduduk miskin (miskin), indeks pembangunan manusia (IPM), dan tingkat pengangguran terbuka (TPT). Sementara variable yang berkorelasi dengan factor 2 adalah: gini ratio. Skor faktor pada output tersebut, tidak dapat dilihat langsung. Namun system di R sudah menghitungnya, sehingga dapat dibuat biplot nya

biplot(fa1,labels=data10.1$Kab)



Garis merah melambangkan vektor ciri (eigen vector). Semakin panjang eigen vector menunjukkan bahwa semakin besar kontribusi variable tersebut terhadap pembentukan factor. Berdasarkan biplot tersebut, dapat diketahui bahwa variable persentase penduduk miskin (miskin) mempunyai peranan terbesar dalam pembentukan factor 1.

Garis-garis eigen vector yang membentuk sudut semakin lancip menunjukkan bahwa ada korelasi yang tinggi diantara variabel. Pada Gambar biplot tersebut dapat dilihat bahwa variable tingkat pengangguran terbuka (TPT) dengan indeks pembangunan manusia (IPM) berkorelasi sangat tinggi .

Berdasarkan biplot tersebut, dapat dilihat bahwa terdapat pengelompokkan objek. Misalnya,

• Kelompok 1, terdiri dari kota jayapura, jayapura, mimika. Kelompok ini mempunyai karakteristik memiliki TPT dan IPM yang relative lebih tinggi dibanding kabupaten/kota lainnya. Kelompok ini bisa dikatakan kelompok perkotaan maju. • Kelompok 2, terdiri dari kabupaten/kota selain kota jayapura, jayapura, mimika, keerom, merauke. Kelompok ini mempunyai tingkat kemiskinan relative lebih tinggi. Kelompok ini bisa dikatakan kelompok perkotaan dengan tingkat kemiskinan yang tinggi. • Sementara itu ada juga keerom dan merauke yang tidak ikut kedua kelompok tersebut dengan ciri rasio guni rasio yang rendah.

#No2

library(readxl)  
data10.2 <- read\_excel("C:/Users/lili/Downloads/Data10Penugasan.xlsx",sheet="Sheet2")  
data10.2

## # A tibble: 5 × 9  
## Pendidikan Tani\_Tambang Industri Listri…¹ Konst…² Dagang Trans…³ Akomo…⁴ Jasa  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 No\_Ijazah 127 38 2 8 28 5 5 17  
## 2 SD 177 86 1 45 76 13 19 23  
## 3 SMP 54 60 4 32 57 6 23 35  
## 4 SMA 60 50 4 22 96 11 24 62  
## 5 PT 6 0 0 2 10 1 1 67  
## # … with abbreviated variable names ¹​Listrik\_Air\_Gas, ²​Konstruksi, ³​Transport,  
## # ⁴​Akomodasi

library(FactoMineR)  
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.2.2

## Loading required package: ggplot2

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(ggplot2)  
library(gplots)

## Warning: package 'gplots' was built under R version 4.2.2

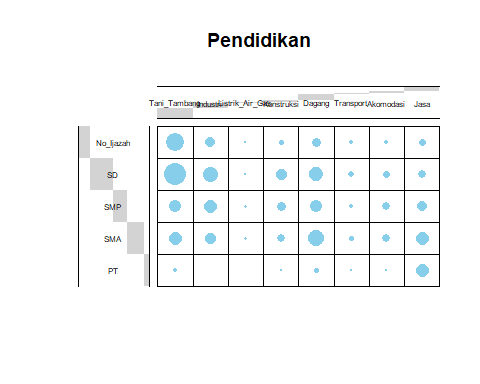
##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library(gplots)  
data <- as.table(as.matrix(data10.2[,-1]))  
rownames(data) <- data10.2$Pendidikan  
data

## Tani\_Tambang Industri Listrik\_Air\_Gas Konstruksi Dagang Transport  
## No\_Ijazah 127 38 2 8 28 5  
## SD 177 86 1 45 76 13  
## SMP 54 60 4 32 57 6  
## SMA 60 50 4 22 96 11  
## PT 6 0 0 2 10 1  
## Akomodasi Jasa  
## No\_Ijazah 5 17  
## SD 19 23  
## SMP 23 35  
## SMA 24 62  
## PT 1 67

balloonplot(t(data), main ="Pendidikan", xlab ="", ylab="",label = FALSE, show.margins = FALSE,dotsize=4, text.size=0.5)

 - Pekerjaan tani tambang banyak dilakukan oleh orang berpendidikan rendah (tanpa ijazah atau SD) - pendidikan SMA lebih banyak bekerja dibidang dagang dan jasa. - pendidikan SMP lebih banyak bekerja di bidang industri atau dagang - Orang orang yang melakukan pekerjaan dibidang jasa, semakin besar seiring dengan semakin tingginya pendidikannya

an.ca <- CA(data, graph = FALSE)  
print(an.ca)

## \*\*Results of the Correspondence Analysis (CA)\*\*  
## The row variable has 5 categories; the column variable has 8 categories  
## The chi square of independence between the two variables is equal to 449.4398 (p-value = 1.615109e-77 ).  
## \*The results are available in the following objects:  
##   
## name description   
## 1 "$eig" "eigenvalues"   
## 2 "$col" "results for the columns"   
## 3 "$col$coord" "coord. for the columns"   
## 4 "$col$cos2" "cos2 for the columns"   
## 5 "$col$contrib" "contributions of the columns"  
## 6 "$row" "results for the rows"   
## 7 "$row$coord" "coord. for the rows"   
## 8 "$row$cos2" "cos2 for the rows"   
## 9 "$row$contrib" "contributions of the rows"   
## 10 "$call" "summary called parameters"   
## 11 "$call$marge.col" "weights of the columns"   
## 12 "$call$marge.row" "weights of the rows"

summary(an.ca)

##   
## Call:  
## CA(X = data, graph = FALSE)   
##   
## The chi square of independence between the two variables is equal to 449.4398 (p-value = 1.615109e-77 ).  
##   
## Eigenvalues  
## Dim.1 Dim.2 Dim.3 Dim.4  
## Variance 0.244 0.072 0.012 0.004  
## % of var. 73.532 21.877 3.483 1.108  
## Cumulative % of var. 73.532 95.409 98.892 100.000  
##   
## Rows  
## Iner\*1000 Dim.1 ctr cos2 Dim.2 ctr cos2  
## No\_Ijazah | 50.319 | -0.328 7.505 0.363 | -0.421 41.461 0.597  
## SD | 35.072 | -0.308 12.640 0.878 | -0.067 2.021 0.042  
## SMP | 20.338 | -0.006 0.003 0.000 | 0.284 22.266 0.793  
## SMA | 30.472 | 0.197 3.867 0.309 | 0.252 21.192 0.504  
## PT | 195.000 | 1.699 75.986 0.949 | -0.384 13.060 0.049  
## Dim.3 ctr cos2   
## No\_Ijazah | -0.081 9.548 0.022 |  
## SD | 0.063 11.223 0.037 |  
## SMP | 0.124 26.425 0.150 |  
## SMA | -0.152 48.796 0.185 |  
## PT | 0.085 4.007 0.002 |  
##   
## Columns  
## Iner\*1000 Dim.1 ctr cos2 Dim.2 ctr cos2  
## Tani\_Tambang | 73.028 | -0.356 16.278 0.543 | -0.326 45.860 0.455  
## Industri | 15.468 | -0.255 4.610 0.726 | 0.125 3.705 0.174  
## Listrik\_Air\_Gas | 3.470 | -0.037 0.004 0.003 | 0.417 1.944 0.406  
## Konstruksi | 13.036 | -0.166 0.911 0.170 | 0.255 7.189 0.400  
## Dagang | 19.297 | 0.023 0.041 0.005 | 0.273 20.245 0.760  
## Transport | 1.385 | -0.102 0.114 0.200 | 0.115 0.481 0.252  
## Akomodasi | 11.434 | -0.034 0.025 0.005 | 0.455 15.136 0.959  
## Jasa | 194.083 | 1.124 78.017 0.979 | -0.162 5.440 0.020  
## Dim.3 ctr cos2   
## Tani\_Tambang | -0.022 1.345 0.002 |  
## Industri | 0.086 11.093 0.083 |  
## Listrik\_Air\_Gas | -0.181 2.291 0.076 |  
## Konstruksi | 0.254 44.801 0.397 |  
## Dagang | -0.146 36.429 0.218 |  
## Transport | -0.112 2.864 0.239 |  
## Akomodasi | 0.009 0.035 0.000 |  
## Jasa | 0.030 1.143 0.001 |

Dari output tersebut didapatkan: nilai Chi square yang menunjukkan statistik uji untuk independence test antar kategori baris dan kolom adalah sebesar 449.4398 (p-value =0 ). Nilai Chi square yang tinggi tersebut menunjukkan ada hubungan/ketergantunganantar kategori baris dan kolom.

Terbentuk 4 dimensi, dimensi 1 menjelaskan 73.532% dari total inertia (variasi data), dimensi 2 sebesar 21.877%,dimensi 3 sebesar 3.483% dan dimensi 4 sebesar 1.108%. Namun jika kita menggunakan dua dimensi pertama (dimensi 1 dan 2) saja, total inertia (varians) yang bisa dijelaskan sudah cukup tinggi yaitu sebesar 95.409%.

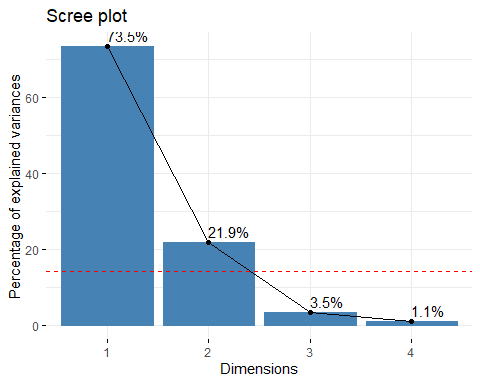
Output juga menghasilkan skor setiap dimensi untuk kategori baris dan kolom(ini tidak lain adalah titik-titik koordinat). Namun cukup sulit diinterpretasikan. Oleh karena itu perlu digunakan visualisasi grafik.

df1 <- (nrow(data) - 1) \* (ncol(data) - 1)  
qchisq(1-0.05,df=df1)

## [1] 41.33714

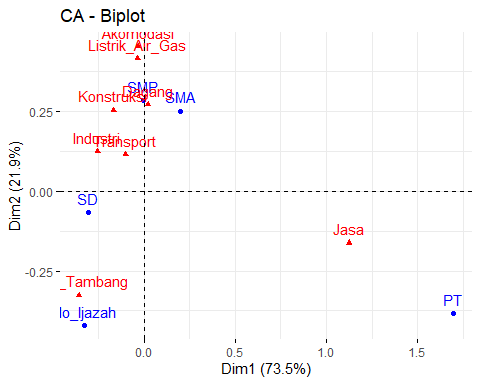
Nilai Chi-square tabel pada saat alpha=0.05% dan derajat bebas df=(nr-1)\*(nc-1) adalah sebesar 41.33714. Kita lihat nilai Chi-square hitung = 449.4398 > Chi-square tabel. Hal ini menghasilkan keputusan “tolak H0” (antar kategori baris dan kolom independent), artinya ada hubungan antar kategori baris dan kolom

fviz\_screeplot(an.ca,addlabels = TRUE) +  
geom\_hline(yintercept=14.29, linetype=2, color="red")

 Data berisi 8 baris dan 5 kolom. Nilai yang diharapkan dari nilai eigen (eigen rata-rata) untuk setiap sumbu adalah 1/(nrow (data) -1) = 1/4 = 25% dalam bentuk baris. 1/(ncol (data) -1) = 1/7 = 14.29% dalam bentuk kolom. Setiap sumbu dengan kontribusi yang maksimum dari dua persentase ini harus dianggap penting dan termasuk dalam solusi untuk interpretasi data. Oleh karena itu untuk menentukan banyaknya dimensi dapat dilihat dari screeplot dengan cut off =14.29%

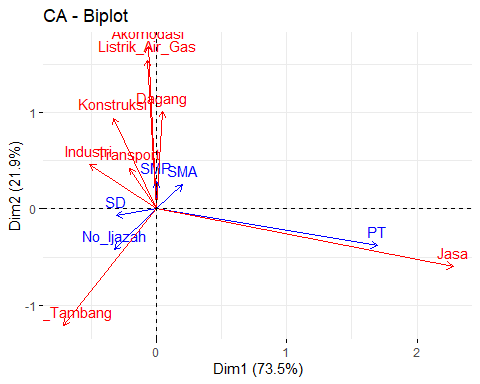
Menurut grafik di atas, cukup digunakan dimensi 1 dan 2. Dimensi 3 dan 4 menjelaskan hanya= 3.5% dan 1.1% dari total inertia yang berada di bawah nilai eigen rata-rata (14.29%) dan terlalu sedikit untuk disimpan dalam analisis lebih lanjut.

library(Rcpp)  
# repel= TRUE untuk menghindari text tumpang tindih(running lambat jika titiknya banyak)  
fviz\_ca\_biplot(an.ca, repel = F)

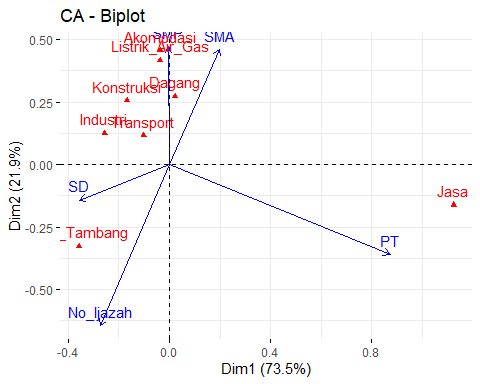
 Plot simetris mewakili profil baris dan kolom secara bersamaan dalam ruang bersama. Biplot simetris ini menunjukkan pola umum data. Warna biru adalah kategori untuk baris dan merah untuk kolom. Dalam hal ini, hanya jarak antar titik dalam baris atau jarak antar titik dalam kolom yang dapat diinterpretasikan. Sementara itu, jarak antara item baris dan kolom tidak memiliki makna. Jarak antara titik baris atau titik kolom memberikan ukuran similarity (atau disimilarity). Titik yang berdekatan menunjukkan adanya kesamaan profil

* Pekerjaan akomodasi, kontruksi, dagang, industri, transportasi, listrik air gas banyak dilakukan orang berpendidikan SMP dilanjutkan dengan pendidikan SMA
* Pekerjaan dibidang jasa banyak dilakukan lulusan pergurusan tinggi

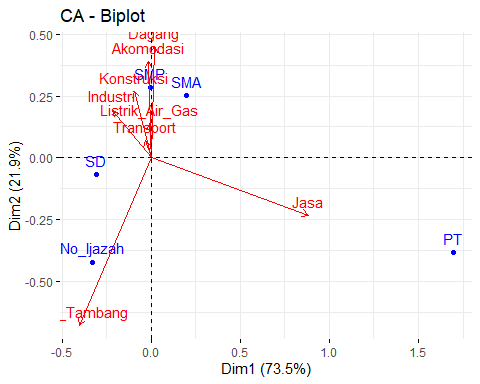
fviz\_ca\_biplot(an.ca,map ="rowprincipal", arrow = c(TRUE, TRUE),  
repel = F)

 Jika sudut antara dua anak panah (biru dan merah) lancip, maka itu adalah hubungan yang kuat antara baris dan kolom yang bersesuaian. Pada gambar tersebut menunjukkan bahwa - Pekerjaan dibidang jasa berhubungan erat dengan lulusan perguruan tinggi - Sementara pekerjaan tani tambang berhubungan erat dengan masyarakat yang tidak memiliki ijazah

#Biplot kontribusi setiap kategori Baris terhadap pembentukan Dimensi  
fviz\_ca\_biplot(an.ca, map ="colgreen", arrow = c(TRUE, FALSE),  
repel = F)

 • Kategori “PT dan SD” berkontribusi penting dalam pembentukan Dimensi pertama • Kategori “SMP,SMA,No Ijazah” berkontribusi penting dalam Dimensi 2.

#Biplot kontribusi setiap kategori Kolom terhadap Dimensi  
fviz\_ca\_biplot(an.ca, map ="rowgreen", arrow = c(FALSE, TRUE), repel  
= F)



Semakin dekat panah (dalam hal jarak sudut) ke sumbu dan panjang vektor /panah semakin panjang, maka semakin besar kontribusi kategori baris pada sumbu itu relatif terhadap sumbu lainnya. Jika panah berada di tengah antara keduanya, kategori barisnya berkontribusi pada dua sumbu pada tingkat yang sama. Pada Gambar tersebut dapat kita lihat bahwa: - Kategori kolom “tani tambang” memiliki kontribusi penting terhadap dimensi 2 secara negatif, sedangkan kolom “air gas listrik” serta akomodasi berkontribusi penting terhadap dimensi 2 secara positif - Kategori kolom “jasa” memiliki kontribusi penting terhadap dimensi 1

deskripsi.dimensi <- dimdesc(an.ca, axes = c(1,2))  
deskripsi.dimensi

## $`Dim 1`  
## $`Dim 1`$row  
## coord  
## No\_Ijazah -0.328378314  
## SD -0.308121933  
## SMP -0.005726466  
## SMA 0.197093697  
## PT 1.698950677  
##   
## $`Dim 1`$col  
## coord  
## Tani\_Tambang -0.35619612  
## Industri -0.25516251  
## Konstruksi -0.16623151  
## Transport -0.10215373  
## Listrik\_Air\_Gas -0.03673402  
## Akomodasi -0.03373645  
## Dagang 0.02255728  
## Jasa 1.12422676  
##   
##   
## $`Dim 2`  
## $`Dim 2`$row  
## coord  
## No\_Ijazah -0.42100051  
## PT -0.38417847  
## SD -0.06719954  
## SMA 0.25166018  
## SMP 0.28422598  
##   
## $`Dim 2`$col  
## coord  
## Tani\_Tambang -0.3261061  
## Jasa -0.1619271  
## Transport 0.1146342  
## Industri 0.1247777  
## Konstruksi 0.2546482  
## Dagang 0.2730391  
## Listrik\_Air\_Gas 0.4168769  
## Akomodasi 0.4546324

Dengan demikian kita dapat melabeli Dimensi 1 sebagai “Pekerjaan yang didominasi lulusan perguruan tinggi”, sementara Dimensi 2 sebagai “Pekerjaan yang didominasi masyarakat yang tidak menyenyam perguruan tinggi”.