

Nama : M.Mahbubillah
NIM : 222011569
No. Absen : 16
Kelas : 3SI1
Mata Kuliah : Analisis Peubah Ganda

Tugas Pertemuan 10

Kerjakan penugasan pada modul "Analisis Korespondensi dan Analisis Biplot" (Catatan: variabelnya sesuaikan dengan datanya).

1. Berikut adalah data tentang kabupaten/kota Provinsi Papua dengan beberapa karakteristik:

- **Presentase penduduk miskin (Miskin)**
- **Indeks Pembangunan Manusia (IPM)**
- **Laju pertumbuhan ekonomi (Laju)**
- **Tingkat pengangguran terbuka (TPT)**
- **Rasio ketergantungan penduduk (Depend)**

Buatlah analisis deksriptif dengan menggunakan biplot.

Library

```
library(readxl)
library(psych)

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

library(GPArotation)
library(FactoMineR)

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

library(factoextra)

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

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.2.2

##
## 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(Rcpp)

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

```

Input Data

##No 1

```

dataPenugasanP10 <- read_excel("penugasanP10.xlsx")
dataPenugasanP10

```

```

## # A tibble: 29 × 5
##   Kab          Miskin   IPM   TPT   Gini
##   <chr>         <dbl> <dbl> <dbl> <dbl>
## 1 Merauke         10.0  70.1  2.61  0.38
## 2 Jayawijaya      37.2  58.0  2.39  0.342
## 3 Jayapura        12.4  71.7  9.68  0.432
## 4 Nabire          24.2  68.8  6.31  0.349
## 5 Kepulauan Yapen 26.3  67.7  5.78  0.4
## 6 Biak Numfor     24.6  72.2 10.4   0.401
## 7 Paniai          36.7  56.3  0.66  0.38
## 8 Puncak Jaya     34.7  48.4  1.78  0.376
## 9 Mimika          14.3  74.2  7.51  0.339
## 10 Boven Digoel    19.4  61.5  3.08  0.448
## # ... with 19 more rows

```

```

data <- dataPenugasanP10[, -1]
data

```

```

## # 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   0.401
## 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

```

Menghitung Factor Score Berdasarkan PA

#Factor Analysis

```
fa1 <- fa(data, nfactors = 2, fm = "pa", scores = T) #Factor Analysis
fa1
```

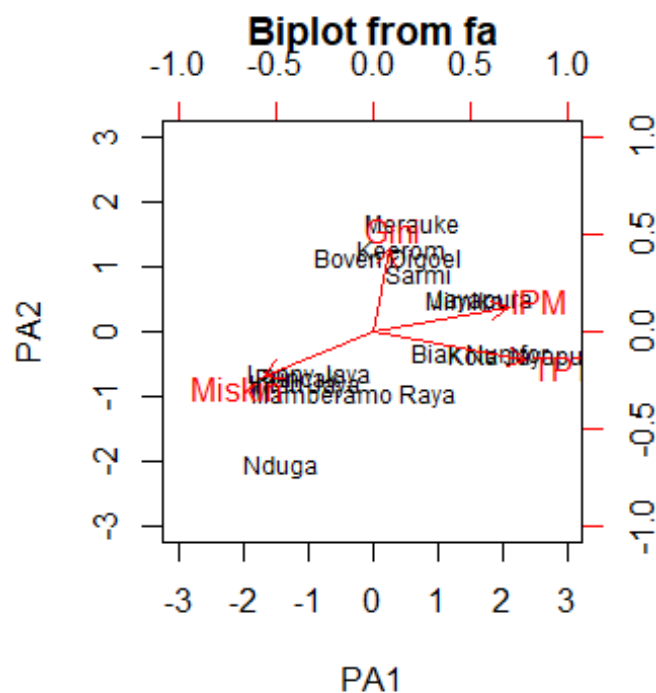
```
## Factor Analysis using method = pa
## Call: fa(r = data, nfactors = 2, scores = T, fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          PA1    PA2    h2    u2 com
## Miskin -0.70 -0.29 0.70 0.30 1.3
## IPM      0.87  0.16 0.87 0.13 1.1
## TPT      0.98 -0.18 0.88 0.12 1.1
## Gini     0.11  0.53 0.33 0.67 1.1
##
##                               PA1  PA2
## SS loadings                  2.30 0.49
## Proportion Var                0.57 0.12
## Cumulative Var                0.57 0.70
## Proportion Explained          0.82 0.18
## Cumulative Proportion         0.82 1.00
##
## With factor correlations of
##          PA1  PA2
## PA1 1.00 0.32
## PA2 0.32 1.00
##
## Mean item complexity = 1.1
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are 6 and the objective function was 2.27 with Chi Square of 58.61
## 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 = 1.121
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
##                               PA1  PA2
## Correlation of (regression) scores with factors 0.97 0.75
## Multiple R square of scores with factors         0.94 0.56
## Minimum correlation of possible factor scores    0.88 0.13
```

Interpretasi:

Berdasarkan output di atas, dapat dilihat bahwa kedua faktor dapat menjelaskan keragaman data. Faktor 1 dapat menjelaskan data sebesar 66,2%. Faktor 2 dapat menjelaskan data sebesar 55,7%. Variabel miskin, IPM dan TPT merupakan variabel yang mempunyai korelasi cukup tinggi terhadap faktor 1. Sedangkan, Variabel gini rasio merupakan variabel yang mempunyai korelasi tinggi terhadap faktor 2.

Biplot

```
#Biplot  
biplot(fa1, labels = dataPenugasanP10$Kab)
```



Interpretasi:

Berdasarkan biplot tersebut, dapat diketahui bahwa variabel Tingkat Pengangguran Terbuka (TPT) mempunyai peranan terbesar dalam pembentukan factor 1, disusul oleh variabel Indeks Pembangunan Manusia (IPM) dan variabel rasio Kemiskinan (Miskin).

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 pengelompokan objek. Misalnya,

- Kelompok 1, terdiri dari kabupaten/kota Intan Jaya, Lanny Jaya, Puncak, dan Mamberamo Raya. Kelompok ini mempunyai karakteristik memiliki rasio penduduk miskin yang relative lebih tinggi dibanding kabupaten/kota lainnya.

- Kelompok 2, terdiri dari kabupaten/kota Jayapura dan Mimika. Sebaliknya kelompok ini mempunyai karakteristik memiliki Indeks Pembangunan Manusia yang relative lebih tinggi dibanding kabupaten/kota lainnya.
- Kelompok 3, terdiri dari kabupaten/kota Kota Jayapura dan Biak Numfor. kelompok ini mempunyai karakteristik memiliki Tingkat Pengangguran Terbuka yang relative lebih tinggi dibanding kabupaten/kota lainnya.
- Sementara itu ada juga kabupaten yang tidak ikut kedua kelompok tersebut, misalnya kabupaten/kota Merauke, Keerom, Sarmi, dan Boven Digoel. kelompok ini mempunyai karakteristik memiliki Tingkat Ketimpangan Pengeluaran (Gini Rasio) yang relative lebih tinggi dibanding kabupaten/kota lainnya. Sedangkan untuk variable lainnya cenderung rendah.

2. Di suatu kabupaten Proinsi Jawa Tengah diadakan survey untuk meneliti hubungan Pendidikan yang ditamatkan dan Lapangan usaha. Diperoleh table kontingensi Banyaknya penduduk yang bekerja menurut pendidikan dan jenis lapangan usaha sebagai berikut:

| | Tani_ | | Listrik_ | | | | | |
|------------|---------|----------|----------|------------|--------|-----------|-----------|------|
| Pendidikan | Tambang | Industri | Air_Gas | Konstruksi | Dagang | Transport | Akomodasi | Jasa |
| No_Ijazah | 127 | 38 | 2 | 8 | 28 | 5 | 5 | 17 |
| SD | 177 | 86 | 1 | 45 | 76 | 13 | 19 | 23 |
| SMP | 54 | 60 | 4 | 32 | 57 | 6 | 23 | 35 |
| SMA | 60 | 50 | 4 | 22 | 96 | 11 | 24 | 62 |
| PT | 6 | 0 | 0 | 2 | 10 | 1 | 1 | 67 |

Ket: Jasa di sini meliputi jasa Pendidikan, jasa Pemerintahan, Keuangan dan jasa lainnya.

Buatlah Analisis Korespondensi untuk menganalisis hubungan variable Pendidikan yang ditamatkan dan variable Lapangan usaha.

##No 2

```
dataPenugasanP10_2 <- read_excel("penugasanP10.xlsx", sheet = "no2")
dataPenugasanP10_2
```

```
## # A tibble: 5 × 9
## Pendidikan Tani_\r\nTam...1 Indus...2 Listr...3 Konst...4 Dagang Trans...5 Akom
o...6 Jasa
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <d
bl> <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`Tani_\r\nTambang`, 2Industri,
## # 3`Listrik_\r\nAir_Gas`, 4Konstruksi, 5Transport, 6Akomodasi

data2 <- dataPenugasanP10_2[,-1]
data2

## # A tibble: 5 × 8
##   `Tani_\r\nTambang` Industri Listrik_\r\n...1 Konst...2 Dagang Trans...3 Akom
o...4 Jasa
##           <dbl>      <dbl>          <dbl>  <dbl>  <dbl>  <dbl>  <d
bl> <dbl>
## 1           127        38             2      8    28      5
5    17
## 2           177        86             1     45    76     13
19   23
## 3            54        60             4     32    57      6
23   35
## 4            60        50             4     22    96     11
24   62
## 5             6         0             0      2    10      1
1    67
## # ... with abbreviated variable names 1`Listrik_\r\nAir_Gas`, 2Konstruksi
,
## # 3Transport, 4Akomodasi

# 1. convert data menjadi sebuah tabel
data2 <- as.table(as.matrix(data2))
# 2. membuat plot
balloonplot(t(data2), main = "Tingkat Pendidikan", xlab = "", ylab = "", label
= FALSE, show.margins = FALSE, dotsize=4, text.size=0.5)
```

Tingkat Pendidikan



Interpretasi:

Berdasarkan output (plot) di atas, dapat dilihat bahwa:

- Penduduk yang tidak memiliki ijazah dan lulusan SD banyak bekerja di bidang tani dan tambang.
- Penduduk lulusan SMP banyak bekerja di bidang tani dan tambang, industri, serta perdagangan.
- Penduduk lulusan SMA banyak bekerja di bidang perdagangan.
- Penduduk lulusan perguruan tinggi banyak bekerja di bidang jasa.

Analisis Korespondensi

```
an.ca <- CA(data2, 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"
```

Interpretasi:

Berdasarkan output di atas, didapatkan nilai Chi-Square yang menunjukkan statistik uji untuk independence test antar kategori baris dan kolom adalah sebesar 449,44 (p-value mendekati 0). Nilai Chi-Square yang tinggi tersebut menunjukkan adanya hubungan/ketergantungan antar kategori baris dan kolom.

Summary dari Analisis Korespondensi

```
summary(an.ca)

##
## Call:
## CA(X = data2, 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   c
tr
## A          |   50.319 | -0.328   7.505   0.363 | -0.421  41.4
61
## B          |   35.072 | -0.308  12.640   0.878 | -0.067   2.0
21
## C          |   20.338 | -0.006   0.003   0.000 |   0.284  22.2
66
## D          |   30.472 |   0.197   3.867   0.309 |   0.252  21.1
92
## E          |  195.000 |   1.699  75.986   0.949 | -0.384  13.0
60
##          cos2      Dim.3      ctr      cos2
## A          0.597 |  -0.081   9.548   0.022 |
## B          0.042 |   0.063  11.223   0.037 |
## C          0.793 |   0.124  26.425   0.150 |
## D          0.504 |  -0.152  48.796   0.185 |
## E          0.049 |   0.085   4.007   0.002 |
##
## Columns
##          Iner*1000   Dim.1   ctr   cos2   Dim.2
ctr
## Tani_\r\nTambang      |   73.028 | -0.356  16.278   0.543 | -0.326  45
.860
## Industri            |   15.468 | -0.255   4.610   0.726 |   0.125   3
.705
## Listrik_\r\nAir_Gas  |    3.470 | -0.037   0.004   0.003 |   0.417   1
.944
## Konstruksi          |   13.036 | -0.166   0.911   0.170 |   0.255   7
.189
## Dagang              |   19.297 |   0.023   0.041   0.005 |   0.273  20
.245
## Transport           |    1.385 | -0.102   0.114   0.200 |   0.115   0
.481
## Akomodasi           |   11.434 | -0.034   0.025   0.005 |   0.455  15
.136
## Jasa                |  194.083 |   1.124  78.017   0.979 | -0.162   5
.440
##          cos2      Dim.3      ctr      cos2
## Tani_\r\nTambang      0.455 |  -0.022   1.345   0.002 |
## Industri            0.174 |   0.086  11.093   0.083 |
## Listrik_\r\nAir_Gas  0.406 |  -0.181   2.291   0.076 |
## Konstruksi          0.400 |   0.254  44.801   0.397 |
## Dagang              0.760 |  -0.146  36.429   0.218 |
## Transport           0.252 |  -0.112   2.864   0.239 |
## Akomodasi           0.959 |   0.009   0.035   0.000 |
## Jasa                0.020 |   0.030   1.143   0.001 |

```


Interpretasi:

Berdasarkan output di atas, dapat diketahui bahwa telah terbentuk 4 dimensi. Dimensi 1 menjelaskan 73,5% dari total inertia (variasi data), dimensi 2 sebesar 21,9%, dimensi 3 sebesar 3.5%, dan dimensi 4 sebesar 0.004%. Jika kita menggunakan dua dimensi pertama (dimensi 1 dan 2) saja, total inertia (varians) yang bisa dijelaskan sudah cukup tinggi yaitu sebesar 95,4%.

Nilai Kritis

```
df1 <- (nrow(dataPenugasanP10_2) - 1) * (ncol(dataPenugasanP10_2) - 1)
qchisq(1-0.05,df=df1)

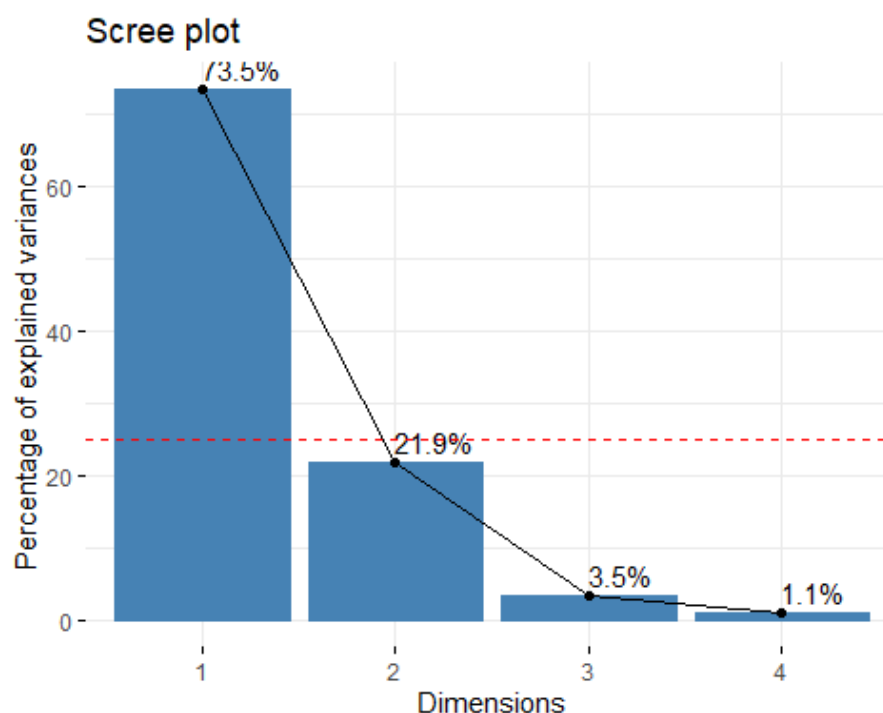
## [1] 46.19426
```

Interpretasi:

Berdasarkan output di atas dengan tingkat kepercayaan 5%, dapat disimpulkan bahwa **ada hubungan antar kategori baris dan kolom**. Hal ini dikarenakan nilai kritis Chi-Square dengan derajat bebas 32 bernilai sebesar 46,194. Sedangkan, nilai Chi-Square hitung bernilai sebesar 449,4398. Sehingga, menghasilkan keputusan **tolak H0** (Chi-Square Hitung > 46,194).

Scree Plot

```
fviz_screplot(an.ca,addlabels = TRUE) +
  geom_hline(yintercept=25, linetype=2, color="red")
```

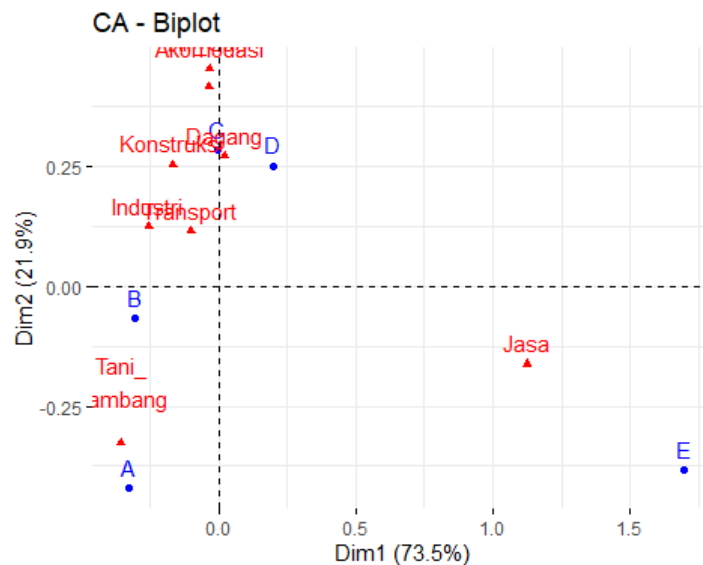


Interpretasi:

Berdasarkan output (grafik) di atas, dapat disimpulkan bahwa hanya cukup gunakan dimensi 1 saja. Dimensi 2, 3, dan 4 masing-masing menjelaskan total inertia lebih rendah dari nilai eigen rata-rata (25%) dan terlalu sedikit untuk disimpan dalam analisis lebih lanjut.

Biplot Simetris

```
library(Rcpp)
fviz_ca_biplot(an.ca, repel = F)
```

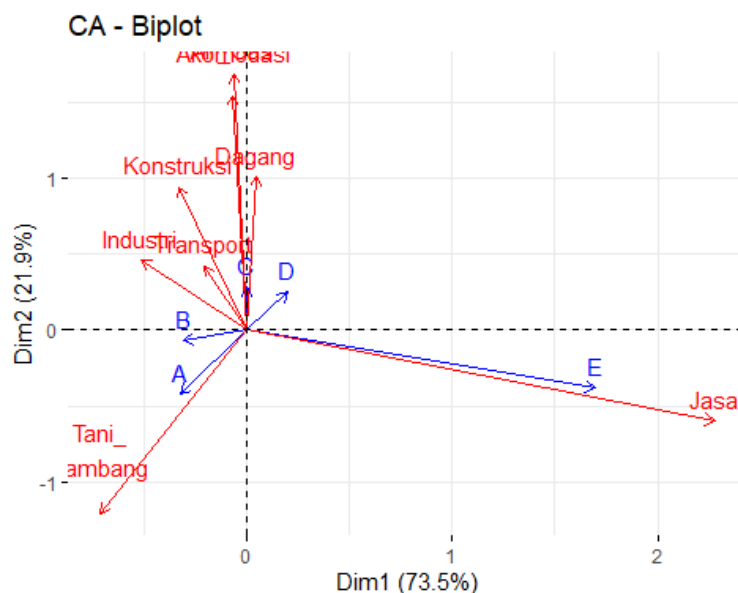


Interpretasi:

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. Grafik tersebut menunjukkan bahwa lulusan SD, SMP, dan SMA memiliki profil yang sama. Mereka banyak menggeluti pekerjaan di bidang selain pertanian dan pertambangan dan jasa.

Biplot Asimetris

```
fviz_ca_biplot(an.ca,
  map = "rowprincipal", arrow = c(TRUE, TRUE), repel = F)
```



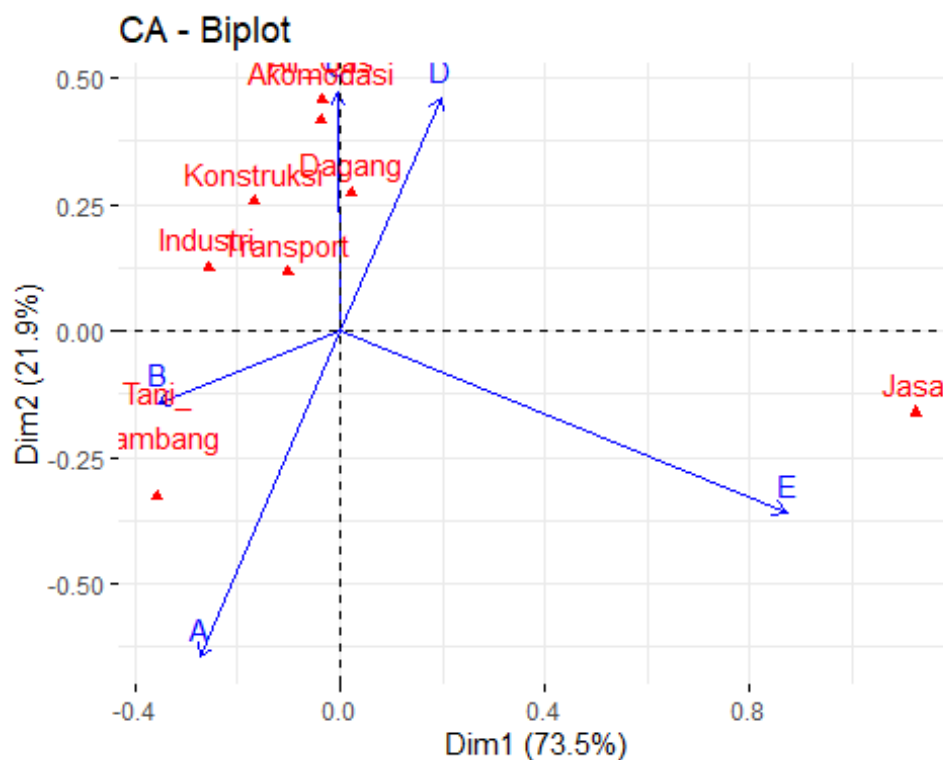
Interpretasi:

Jika sudut antara dua anak panah (biru dan merah) lancip, maka itu adalah hubungan yang kuat antara baris dan kolom yang bersesuaian.

- Pekerjaan di bidang jasa sangat erat hubungannya dengan penduduk lulusan perguruan tinggi.
- Di bidang pertanian dan pertambangan sangat erat berhubungan dengan penduduk yang tidak memiliki ijazah.
- Penduduk lulusan SD cukup berhubungan erat dengan pekerjaan di bidang pertanian dan pertambangan.
- Pekerjaan di bidang perdagangan, akomodasi, serta listrik, air, dan gas sangat erat berhubungan dengan penduduk lulusan SMP.
- Penduduk lulusan SMA cukup berhubungan erat dengan pekerjaan di bidang perdagangan.

Biplot Kontribusi Kategori Kolom

```
fviz_ca_biplot(an.ca, map = "colgreen", arrow = c(TRUE, FALSE), repel = F)
```



Interpretasi:

Berdasarkan output di atas, dapat diketahui bahwa:

- Kutub negatif dimensi 1 terutama ditentukan oleh kategori baris "SD", sedangkan kutub positifnya terutama ditentukan oleh kategori baris "PT".
- Dimensi 2 terutama ditentukan oleh kategori baris "SMP". - Kategori baris "No_Ijazah" dan "SMA" berkontribusi pada dua sumbu pada tingkat yang relatif sama.

Nilai Kontribusi tiap Kategori

```
dim.desc <- dimdesc(an.ca, axes = c(1,2))
dim.desc

## $`Dim 1`
## $`Dim 1`$row
##      coord
## A -0.328378314
## B -0.308121933
## C -0.005726466
## D  0.197093697
## E  1.698950677
##
## $`Dim 1`$col
##      coord
## Tani_\r\nTambang -0.35619612
## Industri -0.25516251
## Konstruksi -0.16623151
## Transport -0.10215373
## Listrik_\r\nAir_Gas -0.03673402
## Akomodasi -0.03373645
## Dagang 0.02255728
## Jasa 1.12422676
```

Interpretasi:

Berdasarkan output di atas, dimensi 1 dapat dilabel sebagai **Bidang Pekerjaan Yang Banyak Digeluti Lulusan Perguruan**. Sedangkan, dimensi 2 dapat dilabel sebagai **Bidang Pekerjaan Yang Banyak Digeluti Penduduk Yang Tidak Memiliki Ijazah**.