



Laporan Tugas 7

Multidimensional Scaling, Analisis Korespondensi (Sederhana dan Berganda), serta Analisis Biplot

Analisis Multivariat Kelas B

Kelompok L

Evan Haryowidyatna	2006485011
--------------------	------------

Muhammad Jauhar Hakim	2006463982
-----------------------	------------

Siskawati Simandalahi	2006572970
-----------------------	------------

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Indonesia

Depok

Mei 2022

A. Multidimensional Scaling

1. Data dan Sumber Data

Data yang kami gunakan merupakan data tabel 16.15 untuk *metric multidimensional scaling* dan tabel 16.16 untuk *nonmetric multidimensional scaling*. Data tersebut kami peroleh dari buku *Methods Of Multivariate Analysis Third Edition Wiley*.

Table 16.15 Road Distances Between Major UK Towns

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
B	431																					
C	611	185																				
D	515	88	170																			
E	535	108	205	47																		
F	232	198	378	282	302																	
G	595	206	78	210	245	398																
H	126	298	478	381	402	99	466															
I	159	407	587	491	511	209	608	133														
J	146	297	477	381	401	98	497	46	102													
K	461	153	333	237	205	228	353	328	438	327												
L	360	135	283	233	253	173	264	231	382	272	223											
M	106	456	636	540	560	258	657	158	65	172	488	431										
N	335	115	264	220	240	123	271	206	333	222	168	61	382									
O	358	102	282	185	205	125	302	225	335	224	105	130	383	75								
P	546	120	60	120	155	313	79	413	523	412	268	190	571	199	216							
Q	352	89	269	172	193	119	289	219	329	218	125	99	378	44	35	204						
R	237	202	350	300	320	57	352	108	241	152	268	142	268	93	175	285	144					
S	497	176	169	233	268	285	171	368	494	384	308	151	543	173	242	115	185	254				
T	402	55	196	146	166	190	217	273	400	289	177	93	449	73	109	131	70	160	120			
U	701	274	287	196	234	468	364	568	678	567	423	419	726	406	370	312	359	486	425	331		
V	630	203	216	125	162	397	292	496	607	496	352	348	655	335	300	241	287	414	354	260	78	
W	376	86	234	184	204	164	255	247	374	263	159	67	422	35	79	169	37	134	148	44	370	299

Table 16.16 Dissimilarity Matrix for World War II Politicians

Person	Hitler	Musso- lini	Churchill	Eisen- hower	Stalin	Attlee	Franco	De Gaulle	Mao Tse	Truman	Chamber- lain	Tito
Hitler	0	5	11	15	8	17	5	10	16	17	12	16
Mussolini	5	0	14	16	13	18	3	11	18	18	14	17
Churchill	11	14	0	7	11	11	12	5	16	8	10	8
Eisenhower	15	16	7	0	16	16	14	8	17	6	7	12
Stalin	8	13	11	16	0	15	13	11	12	14	16	12
Attlee	17	18	11	16	15	0	16	12	16	12	9	13
Franco	5	3	12	14	13	16	0	9	17	16	10	12
De Gaulle	10	11	5	8	11	12	9	0	13	9	11	7
Mao Tse	16	18	16	17	12	16	17	13	0	12	17	10
Truman	17	18	8	6	14	12	16	9	12	0	9	11
Chamberlain	12	14	10	7	16	9	10	11	17	9	0	15
Tito	16	17	8	12	12	13	12	7	10	11	15	0

Data pada tabel 16.15 merupakan tabel dengan data jarak (*distance matrix*) dari jalan besar yang ada pada kota di negara Inggris. Kota-kota tersebut adalah :

A = Aberdeen, B = Birmingham, C = Brighton, D = Bristol, E = Cardiff,
F = Carlisle, G = Dover, H = Edinburgh, I = Fort William, J = Glasgow,
K = Holyhead, L = Hull, M = Inverness, N = Leeds, O = Liverpool, P = London,
Q = Manchester, R = Newcastle, S = Norwich, T = Nottingham, U = Penzance,
V = Plymouth, W = Sheffield.

Data pada tabel 16.16 merupakan tabel dengan politik. Dua subjek menilai tingkat perbedaan antara politisi Perang Dunia II. Matriks data mewakili jumlah perbedaan antara dua subjek. Terdapat 12 politisi yang dinilai di antaranya adalah Hitler, Mussolini, Churchill, Eisenhower, Stalin, Attle, Franco, De Gaulle, Mao Tse, Truman, Chamberlain, dan Tito.

2. Tujuan

Multidimensional Scaling merupakan salah satu teknik peubah ganda yang dapat digunakan untuk menentukan posisi suatu objek lainnya berdasarkan penilaian kemiripannya. Multidimensional Scaling digunakan untuk mengetahui hubungan interdependensi atau saling ketergantungan antar variabel atau data. Multidimensional Scaling dibedakan atas multidimensional scaling berskala metrik dan berskala non-metrik. Tujuan dari multidimensional scaling adalah menemukan suatu konfigurasi sedemikian sehingga jarak antar titik sesuai dengan ketakmiripan antar objek.

3. Langkah-Langkah

- Memasukkan data ke R metric multi-dimensional scaling

```
data1 <- read.table("T16_15_UKROADDIST.csv", header=FALSE, sep=";", row.names  
="V1")
```

```
colnames(data1) <- rownames(data1)
```

```
head(data1)
```

- Membuat metric multi-dimensional scaling

```
data11 <- as.dist(data1)
```

```
mds1 <- data11 %>% cmdscale() %>% as_tibble()
```

```
colnames(mds1) <- c("Dim.1", "Dim.2")
```

```
mds11 <- mds(data11, ndim = 2, type = "ratio" )
```

```
mds11
```

-Membuat Plotting metric multi-dimensional scaling

```
options(repr.plot.width = 10, repr.plot.height = 7, repr.plot.res = 100)
# Plot MDS
ggscatter(mds1, x = "Dim.1", y = "Dim.2",
          label = rownames(data1),
          size = 1,
          repel = TRUE)
```

- Memasukkan data ke R non metric multi-dimensional scaling

```
data2 <- read.table("T16_16_WWIIPOLITICS.csv", header=FALSE, sep=";", row.names="V1")
colnames(data2) <- rownames(data2)
head(data2)
```

- Membuat non metric multi-dimensional scaling

```
# Compute MDS
data22 <- as.dist(t(data2))
mds2 <- data22 %>% isoMDS() %>% . $points %>% as_tibble()
colnames(mds2) <- c("Dim.1", "Dim.2")
```

- Membuat Plotting non metric multi-dimensional scaling

```
scree.plot = function(d, k) {
  stresses=sammon(d, k=k)$stress
  for(i in rev(seq(k-1)))
    stresses=append(stresses,sammon(d, k=i)$stress)
  plot(seq(k),rev(stresses), type="b", xaxp=c(1,k, k-1), ylab="Stress", xlab="Number of dimensions")
}
scree.plot(data22, k=5)
```

```
options(repr.plot.width = 10, repr.plot.height = 7, repr.plot.res = 100)
# Plot MDS
ggscatter(mds2, x = "Dim.1", y = "Dim.2",
          label = rownames(data2),
          size = 1,
          repel = TRUE)
```

4. Hasil

Metric Multi-dimensional Scaling

- Memasukkan data ke R metric multi-dimensional scaling

```
data1 <- read.table("T16_15_UKROADDIST.csv", header=FALSE, sep=";", row.names="V1")
colnames(data1) <- rownames(data1)
head(data1)
```

	A	B	C	D	E	F	G	H	I	J	...	N	O
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	...	<int>	<ir
A	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	...	NA	NA
B	431	0	NA	NA	NA	NA	NA	NA	NA	NA	...	NA	NA
C	611	185	0	NA	NA	NA	NA	NA	NA	NA	...	NA	NA
D	515	88	170	0	NA	NA	NA	NA	NA	NA	...	NA	NA
E	535	108	205	47	0	NA	NA	NA	NA	NA	...	NA	NA
F	232	198	378	282	302	0	NA	NA	NA	NA	...	NA	NA

A data.frame: 6 × 23

- Membuat metric multi-dimensional scaling

Call:

```
mds(delta = data11, ndim = 2, type = "ratio")
```

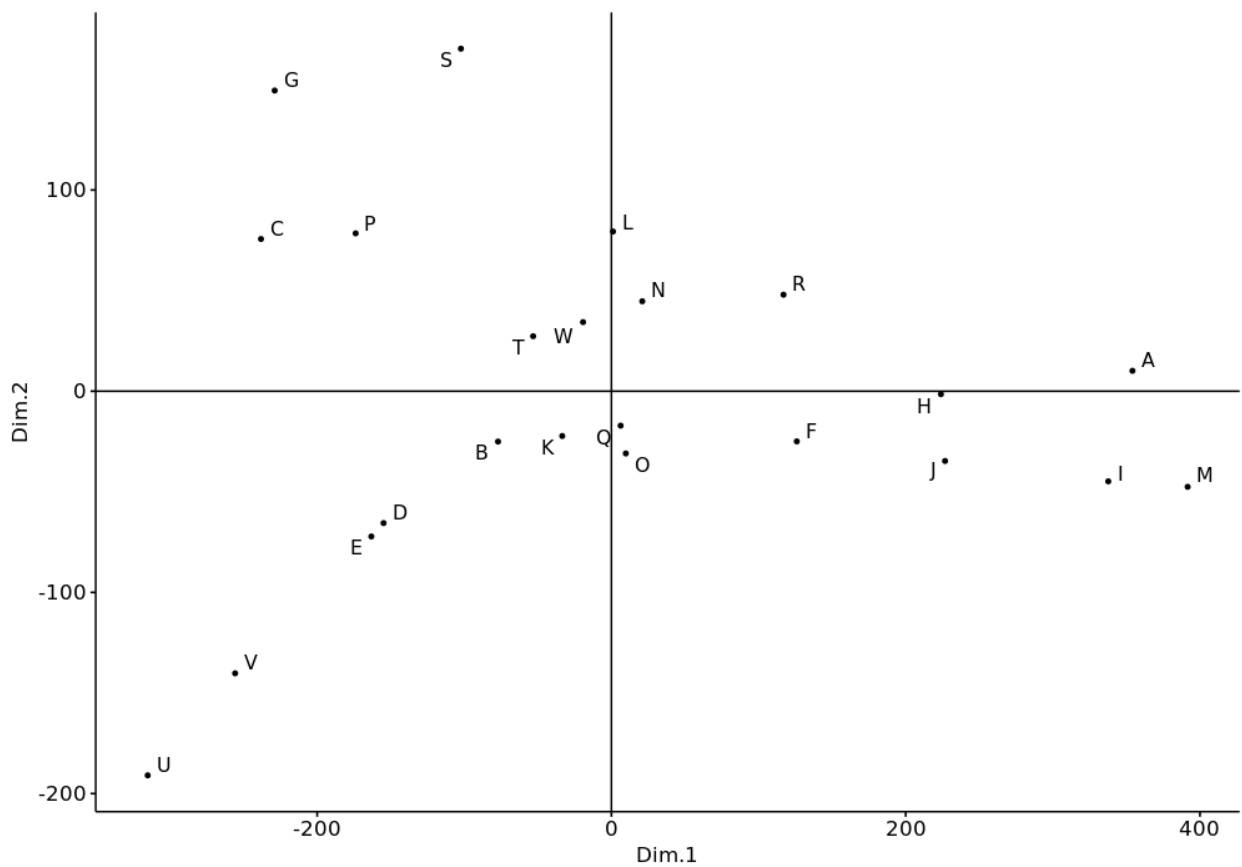
Model: Symmetric SMACOF

Number of objects: 23

Stress-1 value: 0.055

Number of iterations: 31

- Membuat Plotting metric multi-dimensional scaling



Non Metric Multi-dimensional Scaling

- Memasukkan data ke R non metric multi-dimensional scaling

```
data2 <- read.table("T16_16_WWIIPOLITICS.csv", header=FALSE, sep=";", row.names="V1")
colnames(data2) <- rownames(data2)
head(data2)
```

	Hitler	Mussolin	Churchil	Eisenhow	Stalin	Attlee	Franco	DeGaulle	Mao Tse	Tr
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<j
Hitler	0	5	11	15	8	17	5	10	16	17
Mussolin	NA	0	14	16	13	18	3	11	18	18
Churchil	NA	NA	0	7	11	11	12	5	16	8
Eisenhow	NA	NA	NA	0	16	16	14	8	17	6
Stalin	NA	NA	NA	NA	0	15	13	11	12	14
Attlee	NA	NA	NA	NA	NA	0	16	12	16	12

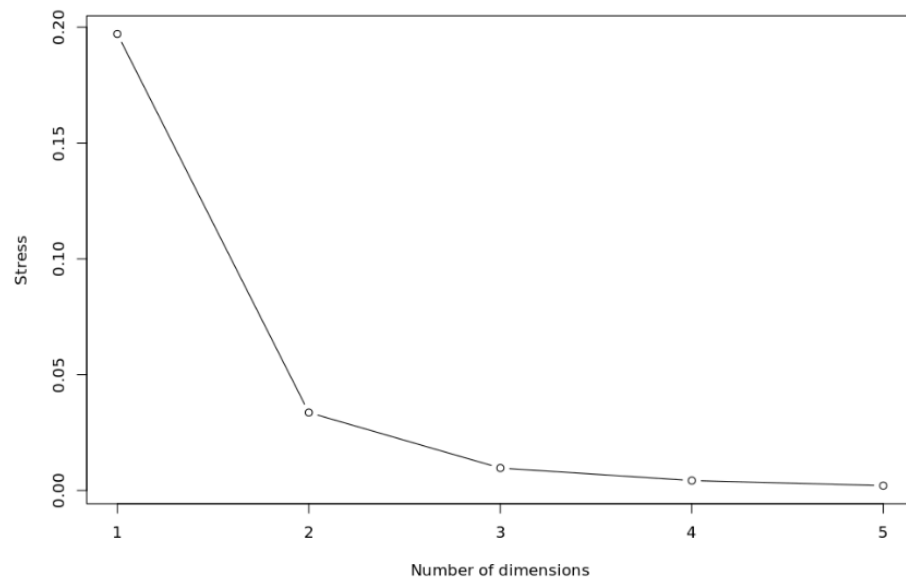
A data.frame: 6 × 12

- Membuat non metric multi-dimensional scaling

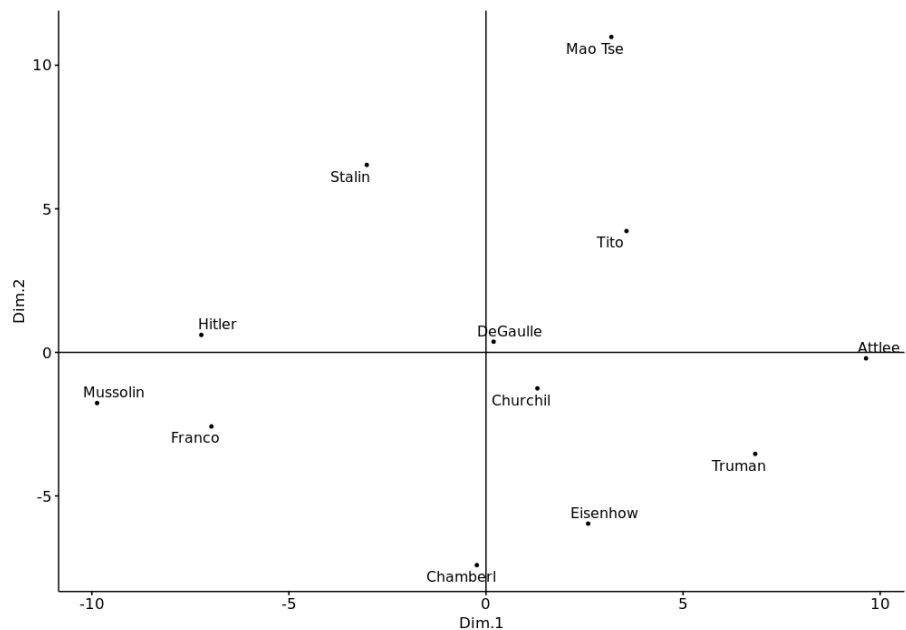
```
# Compute MDS
data22 <- as.dist(t(data2))
mds2 <- data22 %>% isoMDS() %>% .$points %>% as_tibble()
colnames(mds2) <- c("Dim.1", "Dim.2")
```

```
initial value 18.887607
iter 5 value 14.915153
iter 10 value 12.972441
final value 12.927660
converged
```

- Membuat scree plot untuk nilai STRESS



- Membuat Plotting non metric multi-dimensional scaling



5. Interpretasi

Untuk *metric multidimensional scaling* dapat dilihat bahwa jalan C (Brighton), P (London), G (Dover), dan S (Norwich) berdekatan karena dalam kuadran yang sama (kuadran ketiga), yang artinya keempat jalan tersebut memiliki kemiripan karena jaraknya dekat.

Lalu selanjutnya ada jalan B (Birmingham), K (Holyhead), L (Hull), N (Leeds), O (Liverpool), Q (Manchester), T (Nottingham), dan W (Sheffield) yang walau di kuadran yang berbeda tetapi memiliki jarak yang berdekatan di dalam plot, yaitu memusat di tengah sumbu (0,0), yang artinya juga memiliki jarak berdekatan.

Berikutnya ada kelompok yang berada di sekitar sumbu x positif yaitu jalan A (Aberdeen), F (Carlisle), H (Edinburgh), I (Hull), J (Glasgow), M(Inverness), dan R (Newcastle) yang memiliki jarak berdekatan.

Dan yang terakhir ada kelompok yang berada di kuadran 4 yaitu jalan D (Bristol), E (Cardiff), U (Penzance), dan V (Plymouth). Walau dalam satu kuadran jarak jalan Cardiff dan Plymouth lumayan jauh dikarenakan pada plot juga tidak dekat kemiripannya.

Untuk *nonmetric multidimensional scaling*, dapat dilihat dari *scree plot* nilai STRESS nya dengan menggunakan 2 dimensi sudah dapat menurunkan hingga kurang dari 0.05, dan juga dapat dilihat dari plot MDS nya, Hitler, Musolini, dan Franco memiliki nilai yang mirip karena berdekatan. Selanjutnya ada Stalin yang tidak dekat dengan siapa pun. Eisenhower, Truman, Churchill, dan Chamberlain memiliki kemiripan karena berada pada atau dekat dengan kuadran 1. Dan yang tersisa ada De Gaulle, Tito, dan Attle yang berada di kuadran 2.

B. Analisis Korespondensi Sederhana

1. Data dan Sumber Data

Data yang kamu gunakan merupakan data tabel 16.17. Data tersebut kami peroleh dari buku *Methods Of Multivariate Analysis Third Edition Wiley*.

Table 16.17 Birth and Death Months of 1281 People

Death	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Birth													
Jan	9	14	12	14	9	11	10	15	9	11	11	13	138
Feb	13	7	6	8	9	5	4	5	7	11	1	13	89
Mar	12	14	9	7	17	10	9	10	9	2	13	9	121
Apr	7	11	10	7	9	11	4	11	4	9	8	12	103
May	8	6	11	7	4	12	9	9	11	6	9	4	96
Jun	14	5	5	7	13	5	7	4	8	9	7	5	89
Jul	12	10	13	4	5	11	7	6	4	10	4	8	94
Aug	7	7	11	13	10	9	3	8	8	8	8	12	104
Sep	7	12	11	10	10	11	4	8	12	6	8	7	106
Oct	16	8	11	9	7	10	12	8	9	8	7	7	112
Nov	7	9	15	10	11	10	7	10	7	10	9	6	111
Dec	6	18	9	11	14	8	9	9	10	6	10	8	118
Total	118	121	123	107	118	113	85	103	98	96	95	104	1281

Data tersebut merupakan data bulan kelahiran dan kematian dari 1281 orang yang berasal dari 4 kelompok dan dibuat tabel kontingensi frekuensinya seperti yang ada pada tabel 16.17 tersebut.

2. Tujuan

Analisis Korespondensi merupakan metode untuk mereduksi dimensi variable dan menggambarkan profil vector baris dan vector kolom suatu matriks data dari tabel kontingensi. Analisis Korespondensi Sederhana adalah metode yang digunakan untuk menganalisis tabel kontingensi dengan dua klasifikasi

3. Langkah-Langkah

- Memasukkan data pada R

```
data3 <- read.csv("T16_17_BIRTHDEATH.csv", header=TRUE, sep=";")
data3
data33 <- data3 %>% group_by(Birth,Death) %>% summarise(count = sum(count))
data33$Birth = factor(data33$Birth, levels = month.name)
data33$Death = factor(data33$Death, levels = month.name)
data33<-with(data33, data33[order(Birth, Death, count),])
head(data33)
```

- Membuat tabel

```
column_names <- c("D-Jan", "D-Feb", "D-Mar", "D-Apr", "D-May", "D-Jun", "D-Jul", "D-
Aug", "D-Sep", "D-Oct", "D-Nov", "D-Dec")
row_names <- c("B-Jan", "B-Feb", "B-Mar", "B-Apr", "B-May", "B-Jun", "B-Jul", "B-
Aug", "B-Sep", "B-Oct", "B-Nov", "B-Dec")
tabel3 <- matrix(data33$count, ncol=12, byrow=TRUE)
colnames(tabel3) <- column_names
rownames(tabel3) <- row_names
tabel3
```

```
#Matriks P atau Matriks Korespondensi
prop.table(tabel3)
```

```
#The value of each cell is divided by the sum of the row cells.
#Matriks R (Row)
prop.table(prop.table(tabel3), 1)
```

```
#The value of each cell is divided by the sum of the column cells.
#Matriks C (Column)
prop.table(prop.table(tabel3), 2)
```

- Membuat analisis korespondensi

```
fit <- ca(tabel3)
print(fit) # basic results
summary(fit) # extended results
```

- Uji Independensi antara kategori baris bulan kelahiran dan kolom bulan kematian

```
#H0 : baris bulan kelahiran dan kolom bulan kematian saling bebas
CA(tabel3)
```

- Membuat plot analisis korespondensi

```
plot(fit) # symmetric map
plot(fit, mass = TRUE, contrib = "absolute", map =
  "rowgreen", arrows = c(FALSE, TRUE)) # asymmetric map
```

4. Hasil

- Memasukkan data pada R

```
data33 <- data3 %>% group_by(Birth,Death) %>% summarise(count = sum(count))
data33$Birth = factor(data33$Birth, levels = month.name)
data33$Death = factor(data33$Death, levels = month.name)
data33<-with(data33, data33[order(Birth, Death, count),])
head(data33)
```

`summarise()` has grouped output by 'Birth'. You can override using the `.groups` argument.

Birth	Death	count
<fct>	<fct>	<int>
January	January	9
January	February	14
January	March	12
January	April	14
January	May	9
January	June	11

A grouped_df: 6 × 3

- Membuat tabel kontingensi 2 arah (baris untuk bulan kelahiran dan kolom untuk bulan kematian) masih merupakan matriks frekuensi

	D-Jan	D-Feb	D-Mar	D-Apr	D-May	D-Jun	D-Jul	D-Aug	D-Sep	D-Oct	D-Nov	D-Dec
B-Jan	9	14	12	14	9	11	10	15	9	11	11	13
B-Feb	13	7	6	8	9	5	4	5	7	11	1	13
B-Mar	12	14	9	7	17	10	9	10	9	2	13	9
B-Apr	7	11	10	7	9	11	4	11	4	9	8	12
B-May	8	6	11	7	4	12	9	9	11	6	9	4
B-Jun	14	5	5	7	13	5	7	4	8	9	7	5
B-Jul	12	10	13	4	5	11	7	6	4	10	4	8
B-Aug	7	7	11	13	10	9	3	8	8	8	8	12
B-Sep	7	12	11	10	10	11	4	8	12	6	8	7
B-Oct	16	8	11	9	7	10	12	8	9	8	7	7
B-Nov	7	9	15	10	11	10	7	10	7	10	9	6
B-Dec	6	18	9	11	14	8	9	9	10	6	10	8

A matrix: 12 × 12 of type int

- Matriks P atau Matriks Korespondensi

	D-Jan	D-Feb	D-Mar	D-Apr	D-May	D-Jun	D-Jul
B-Jan	0.007025761	0.010928962	0.009367681	0.010928962	0.007025761	0.008587041	0.007806401
B-Feb	0.010148322	0.005464481	0.004683841	0.006245121	0.007025761	0.003903201	0.003122560
B-Mar	0.009367681	0.010928962	0.007025761	0.005464481	0.013270882	0.007806401	0.007025761
B-Apr	0.005464481	0.008587041	0.007806401	0.005464481	0.007025761	0.008587041	0.003122560
B-May	0.006245121	0.004683841	0.008587041	0.005464481	0.003122560	0.009367681	0.007025761
B-Jun	0.010928962	0.003903201	0.003903201	0.005464481	0.010148322	0.003903201	0.005464481
B-Jul	0.009367681	0.007806401	0.010148322	0.003122560	0.003903201	0.008587041	0.005464481
B-Aug	0.005464481	0.005464481	0.008587041	0.010148322	0.007806401	0.007025761	0.002341920
B-Sep	0.005464481	0.009367681	0.008587041	0.007806401	0.007806401	0.008587041	0.003122560
B-Oct	0.012490242	0.006245121	0.008587041	0.007025761	0.005464481	0.007806401	0.009367681
B-Nov	0.005464481	0.007025761	0.011709602	0.007806401	0.008587041	0.007806401	0.005464481
B-Dec	0.004683841	0.014051522	0.007025761	0.008587041	0.010928962	0.006245121	0.007025761

A matrix: 12 × 12 of type dbl

- Matriks R

<pre> #The value of each cell is divided by the sum of the row cells. #Matriks R (Row) prop.table(prop.table(tabel3), 1) #The value of each cell is divided by the sum of the column cells. #Matriks C (Column) prop.table(prop.table(tabel3), 2) </pre>								[29]
	D-Jan	D-Feb	D-Mar	D-Apr	D-May	D-Jun	D-Jul	D-Aug
B-Jan	0.06521739	0.10144928	0.08695652	0.10144928	0.06521739	0.07971014	0.07246377	0.1086
B-Feb	0.14606742	0.07865169	0.06741573	0.08908764	0.10112360	0.05617978	0.04494382	0.0561
B-Mar	0.09917355	0.11570248	0.07438017	0.05785124	0.14049587	0.08264463	0.07438017	0.0826
B-Apr	0.06796117	0.10679612	0.09708738	0.06796117	0.08737864	0.10679612	0.03883495	0.1067
B-May	0.08333333	0.06250000	0.11458333	0.07291667	0.04166667	0.12500000	0.09375000	0.0937
B-Jun	0.15730337	0.05617978	0.05617978	0.07865169	0.14606742	0.05617978	0.07865169	0.0449
B-Jul	0.12765957	0.10630298	0.13829787	0.04255319	0.05319149	0.11702128	0.07446809	0.0638
B-Aug	0.06730769	0.06730769	0.10576923	0.12500000	0.09615385	0.08653846	0.02884615	0.0769
B-Sep	0.06603774	0.11320755	0.10377358	0.09433962	0.09433962	0.10377358	0.03773585	0.0754
B-Oct	0.14285714	0.07142857	0.09821429	0.08035714	0.06250000	0.08928571	0.10714286	0.0714
B-Nov	0.06306306	0.08108108	0.13513514	0.09009009	0.09909910	0.09009009	0.06306306	0.0900
B-Dec	0.05084746	0.15254237	0.07627119	0.09322034	0.11864407	0.06779661	0.07627119	0.0762

- Matriks C

	D-Jan	D-Feb	D-Mar	D-Apr	D-May	D-Jun	D-Jul	D-Aug
B-Jan	0.07627119	0.11570248	0.09756098	0.13084112	0.07627119	0.09734513	0.11764706	0.1456
B-Feb	0.11016949	0.05785124	0.04878049	0.07476636	0.07627119	0.04424779	0.04705882	0.0485
B-Mar	0.10169492	0.11570248	0.07317073	0.06542056	0.14406780	0.08849558	0.10588235	0.0976
B-Apr	0.05932203	0.09090909	0.08130081	0.06542056	0.07627119	0.09734513	0.04705882	0.1067
B-May	0.06779661	0.04958678	0.08943089	0.06542056	0.03389831	0.10619469	0.10588235	0.0873
B-Jun	0.11864407	0.04132231	0.04065041	0.06542056	0.11016949	0.04424779	0.08235294	0.0388
B-Jul	0.10169492	0.08264463	0.10569106	0.03738318	0.04237288	0.09734513	0.08235294	0.0582
B-Aug	0.05932203	0.05785124	0.08943089	0.12149533	0.08474576	0.07964602	0.03529412	0.0776
B-Sep	0.05932203	0.09917355	0.08943089	0.09345794	0.08474576	0.09734513	0.04705882	0.0776
B-Oct	0.13559322	0.06611570	0.08943089	0.08411215	0.05932203	0.08849558	0.14117647	0.0776
B-Nov	0.05932203	0.07438017	0.12195122	0.09345794	0.09322034	0.08849558	0.08235294	0.0976
B-Dec	0.05084746	0.14876033	0.07317073	0.10280374	0.11864407	0.07079646	0.10588235	0.0873

A matrix: 12 x 12 of type dbl

- Membuat analisis korespondensi

```
fit <- ca(tabel3)
print(fit) # basic results
```

```
Principal inertias (eigenvalues):
      1      2      3      4      5      6      7
Value  0.027658 0.02008 0.019242 0.009205 0.005122 0.004521 0.004022
Percentage 30.08% 21.84% 20.93% 10.01% 5.57% 4.92% 4.37%
      8      9      10      11
Value  0.001449 0.000381 0.000225 3.4e-05
Percentage 1.58% 0.41% 0.24% 0.04%

Rows:
      B-Jan  B-Feb  B-Mar  B-Apr  B-May  B-Jun  B-Jul
Mass  0.107728 0.069477 0.094457 0.080406 0.074941 0.069477 0.073380
ChiDist 0.191656 0.454817 0.325537 0.266895 0.344261 0.398838 0.355135
Inertia 0.003957 0.014372 0.010010 0.005728 0.008882 0.011052 0.009255
Dim. 1 -0.325884 2.404140 -1.052097 -0.071642 -0.653475 1.200487 1.157301
Dim. 2 -0.788072 -0.617658 1.148960 -1.514511 0.773082 1.939944 -0.304897
      B-Aug  B-Sep  B-Oct  B-Nov  B-Dec
Mass  0.081187 0.082748 0.087432 0.086651 0.092116
ChiDist 0.268682 0.231349 0.275389 0.211054 0.292926
Inertia 0.005861 0.004429 0.006631 0.003860 0.007904
Dim. 1 0.046875 -0.811163 0.725437 -0.386559 -1.224087
Dim. 2 -1.263058 -0.231053 1.245813 -0.372932 0.171183

Columns:
      D-Jan  D-Feb  D-Mar  D-Apr  D-May  D-Jun  D-Jul
Mass  0.092116 0.094457 0.096019 0.083528 0.092116 0.088212 0.066354
ChiDist 0.389437 0.285619 0.246212 0.243897 0.336024 0.227419 0.340968
Inertia 0.013970 0.007706 0.005821 0.004969 0.010401 0.004562 0.007714
```

```
summary(fit) # extended results
```

Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.027658	30.1	30.1	*****
2	0.020080	21.8	51.9	*****
3	0.019242	20.9	72.9	*****
4	0.009205	10.0	82.9	***
5	0.005122	5.6	88.4	*
6	0.004521	4.9	93.4	*
7	0.004022	4.4	97.7	*
8	0.001449	1.6	99.3	
9	0.000381	0.4	99.7	
10	0.000225	0.2	100.0	
11	3.4e-050	0.0	100.0	

Total: 0.091939 100.0

Rows:

	name	mass	qlt	inr	k=1 cor	ctr	k=2 cor	ctr
1	BJan	108	419	43	-54	80	11	-112 340 67
2	BFeb	69	810	156	400	773	402	-88 37 27
3	BMar	94	539	109	-175	289	105	163 250 125
4	BApr	80	649	62	-12	2	0	-215 647 184
5	BMay	75	201	97	-109	100	32	110 101 45
6	BJun	69	726	120	200	251	100	275 475 261
7	BJul	73	309	101	192	294	98	-43 15 7
8	BAug	81	445	64	8	1	0	-179 444 130
9	BSep	83	360	48	-135	340	54	-33 20 4
10	BOct	87	603	72	121	192	46	177 411 136
11	BNov	87	155	42	-64	93	13	-53 63 12

- Uji Independensi antara kategori baris bulan kelahiran dan kolom bulan kematian

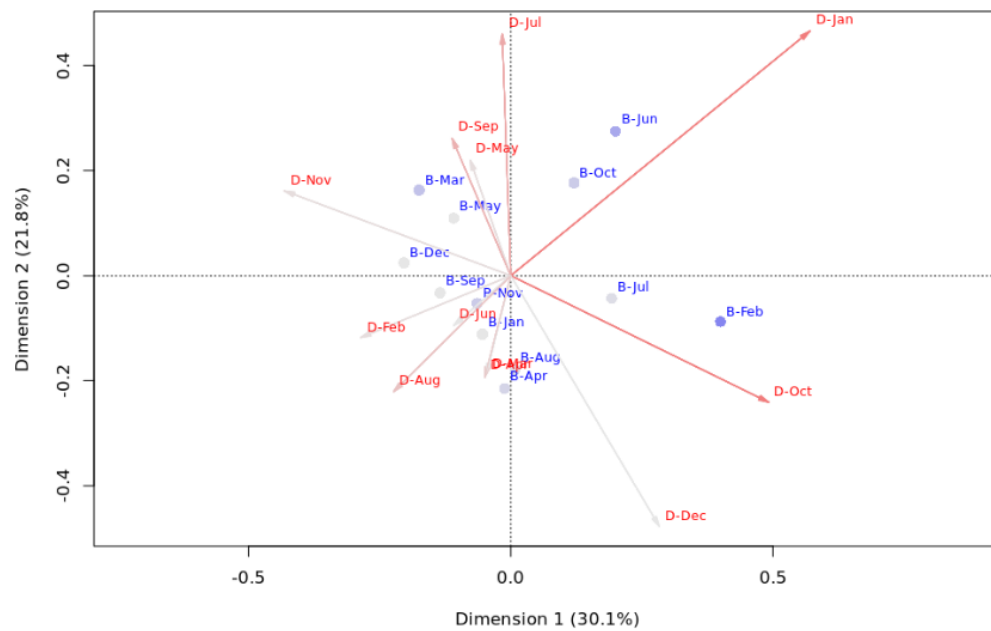
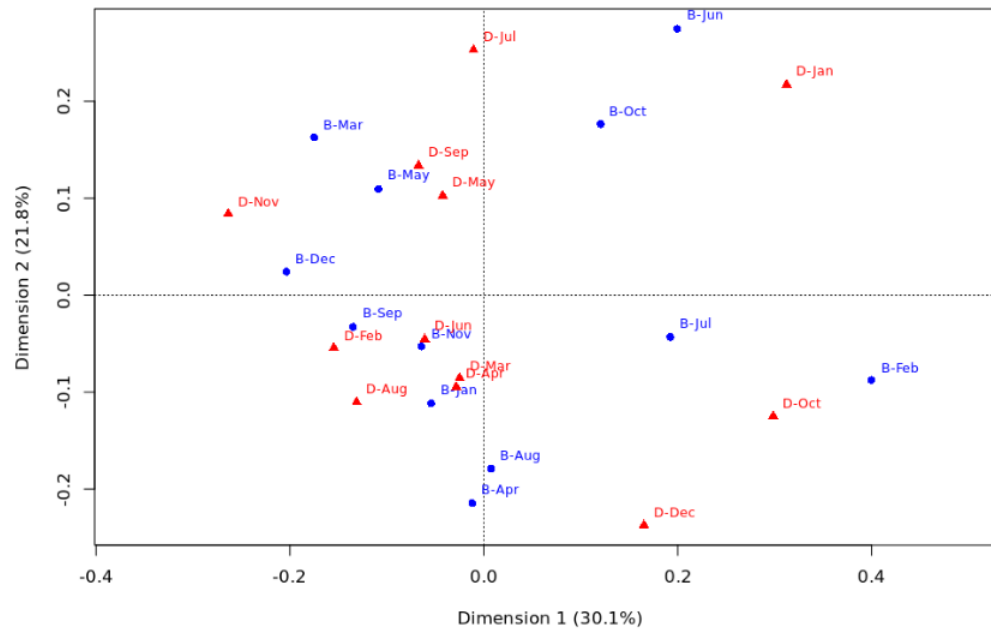
****Results of the Correspondence Analysis (CA)****

The row variable has 12 categories; the column variable has 12 categories

The chi square of independence between the two variables is equal to 117.7742 (p-value = 0.5660272).

- Membuat plot analisis korespondensi

```
plot(fit) # symmetric map
plot(fit, mass = TRUE, contrib = "absolute", map =
     "rowgreen", arrows = c(FALSE, TRUE)) # asymmetric map
```



5. Interpretasi

Pada pengujian independensi kategori baris bulan kelahiran dan kategori kolom bulan kematian dengan H_0 : Data saling bebas didapatkan statistik chi-square sebesar $\chi^2 = 117.7742$ dan didapatkan $p - value = 0.5660272$ yang artinya H_0 tidak ditolak dikarenakan $p - value > 0.05 = \alpha$. Karena H_0 tidak ditolak artinya kategori baris bulan kelahiran dan kategori kolom bulan kematian saling bebas.

Dan dilihat dari nilai principal inertia atau eigen value nya dengan 2 dimensi sudah dapat menjelaskan 51.9% di mana dimensi 1 menjelaskan 30.1% dan dimensi 2 menjelaskan 21.8%. Pada plot korespondensi, didapatkan asosiasinya sebagai berikut :

{D-Jun, B-Nov}, {D-Mar, D-Apr, B-Jan}, {D-Feb, B-Sep}. {B-Apr, B-Aug}, {D-May, D-Sep, B-May}

Asosiasi tersebut dilihat dari kedekatannya dalam plot korespondensi.

C. Analisis Korespondensi Berganda

1. Data dan Sumber Data

Data yang kami gunakan merupakan data tabel 16.19. Data tersebut kami peroleh dari buku *Methods Of Multivariate Analysis Third Edition Wiley*.

Table 16.19 Byssinosis Data

Smoking	Gender	Years in job	Suffer Byssi- nosis	Race					
				Other			White		
				High dust	Low dust	Med dust	High dust	Low dust	Med dust
Non- smoker	Female	≤ 10	No	24	301	142	4	169	54
			Yes	1	4	3	0	2	1
		≥ 20	No	0	3	2	2	340	187
			Yes	0	0	0	0	2	3
		10–20	No	0	4	4	0	90	30
			Yes	0	0	0	0	1	0
	Male	≤ 10	No	75	122	47	16	134	35
			Yes	6	1	1	0	0	0
		≥ 20	No	15	23	1	47	182	39
			Yes	3	0	0	5	3	0
		10–20	No	9	7	0	8	58	16
			Yes	1	0	0	2	0	1
Smoker	Female	≤ 10	No	22	260	145	5	180	93
			Yes	2	3	2	0	3	1
		≥ 20	No	1	2	0	1	176	91
			Yes	0	0	0	0	3	3
		10–20	No	0	3	4	0	94	33
			Yes	0	0	0	0	2	1
	Male	≤ 10	No	139	242	88	37	258	74
			Yes	25	3	0	3	2	0
		≥ 20	No	31	45	1	77	495	141
			Yes	10	0	0	31	12	1
		10–20	No	30	33	5	21	187	50
			Yes	8	0	0	8	1	1

Data yang digunakan merupakan data dari 5419 pekerja tekstil katun yang dijelaskan dalam 5 faktor yaitu tipe tempat kerja, tahun kerja, perokok, jenis kelamin, dan ras. Data tersebut diambil dari orang yang terkena penyakit bisinosis dan juga orang yang tidak terkena penyakit bisinosis.

2. Tujuan

Analisis Korespondensi merupakan metode untuk mereduksi dimensi variable dan menggambarkan profil vector baris dan vector kolom suatu matriks data dari tabel kontingensi. Analisis Korespondensi Berganda adalah metode yang digunakan untuk menganalisis tabel kontingensi dengan lebih dari dua klasifikasi.

3. Langkah-Langkah

- Memasukkan data pada R

```
data4 <- read.delim("T16_19_BYSSINOSIS.dat", header=TRUE)
head(data4)
```

-Membuat multiple correspondence analysis

```
mca <- MCA(data4[,3:7], graph = FALSE)
mca
```

-Membuat tabel nilai eigen, column coordinates, row coordinates, number of categories per variable

```
# table of eigenvalues
mca$eig
# column coordinates
head(mca$var$coord)
# row coordinates
head(mca$ind$coord)
# number of categories per variable
cats = apply(data4[,3:7], 2, function(x) nlevels(as.factor(x)))
cats
```

-Membuat plot

```
# data frames for ggplot
mca_vars_df = data.frame(mca$var$coord, Variable = rep(names(cats),
  cats))
mca_obs_df = data.frame(mca$ind$coord)

# plot of variable categories
ggplot(data = mca_vars_df, aes(x = Dim.1, y = Dim.2, label = rownames(mca_var
s_df))) +
  geom_hline(yintercept = 0, colour = "gray70") + geom_vline(xintercept = 0
,
  colour = "gray70") + geom_text(aes(colour = Variable)) + ggtitle("MCA plo
t of variables using R package FactoMineR")
# MCA plot of observations and categories
ggplot(data = mca_obs_df, aes(x = Dim.1, y = Dim.2)) + geom_hline(yintercept
= 0,
  colour = "gray70") + geom_vline(xintercept = 0, colour = "gray70") + geom
_point(colour = "gray50",
  alpha = 0.7) + geom_density2d(colour = "gray80") + geom_text(data = mca_v
ars_df,
  aes(x = Dim.1, y = Dim.2, label = rownames(mca_vars_df), colour = Variabl
e)) +
  ggtitle("MCA plot of variables using R package FactoMineR") + scale_colou
r_discrete(name = "Variable")
# default biplot in FactoMineR
plot(mca)
```

4. Hasil

- Memasukkan data pada R

```
data4 <- read.delim("T16_19_BYSSINOSIS.dat", header=TRUE)
head(data4)
```

	Yes	No	Dust	Race	Sex	Smoking	Work
	<int>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
1	3	37	Dust_High	White	Male	Smoker	<10
2	0	74	Dust_Med	White	Male	Smoker	<10
3	2	258	Dust_Low	White	Male	Smoker	<10
4	25	139	Dust_High	Race-Other	Male	Smoker	<10
5	0	88	Dust_Med	Race-Other	Male	Smoker	<10
6	3	242	Dust_Low	Race-Other	Male	Smoker	<10

A data.frame: 6 × 7

- Membuat multiple correspondence analysis

```
mca <- MCA(data4[3:7], graph = FALSE)
mca
```

```
**Results of the Multiple Correspondence Analysis (MCA)**
The analysis was performed on 72 individuals, described by 5 variables
*The results are available in the following objects:
```

	name	description
1	"\$eig"	"eigenvalues"
2	"\$var"	"results for the variables"
3	"\$var\$coord"	"coord. of the categories"
4	"\$var\$cos2"	"cos2 for the categories"
5	"\$var\$contrib"	"contributions of the categories"
6	"\$var\$sv.test"	"v-test for the categories"
7	"\$ind"	"results for the individuals"
8	"\$ind\$coord"	"coord. for the individuals"
9	"\$ind\$cos2"	"cos2 for the individuals"
10	"\$ind\$contrib"	"contributions of the individuals"
11	"\$call"	"intermediate results"
12	"\$call\$marge.col"	"weights of columns"
13	"\$call\$marge.li"	"weights of rows"

- Membuat tabel nilai eigen, column coordinates, row coordinates, number of categories per variable

```
# table of eigenvalues
mca$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.2	14.28571	14.28571
dim 2	0.2	14.28571	28.57143
dim 3	0.2	14.28571	42.85714
dim 4	0.2	14.28571	57.14286
dim 5	0.2	14.28571	71.42857
dim 6	0.2	14.28571	85.71429
dim 7	0.2	14.28571	100.00000

A matrix: 7 × 3 of type dbl

```
# column coordinates
head(mca$var$coord)
```

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Dust_High	2.475568e-16	0.00000000	4.315845e-16	8.791211e-16	0.00000000
Dust_Low	-2.146114e-01	0.40720020	-8.911164e-01	3.750029e-01	0.53598958
Dust_Med	2.146114e-01	-0.40720020	8.911164e-01	-3.750029e-01	-0.53598958
Race-Other	9.347689e-01	-0.01046484	-4.197691e-02	1.772536e-01	0.04369256
White	-9.347689e-01	0.01046484	4.197691e-02	-1.772536e-01	-0.04369256
Female	-3.172536e-02	-0.78024344	1.602544e-02	-2.211975e-02	0.62406106

A matrix: 6 × 5 of type dbl

```
# row coordinates
head(mca$ind$coord)
```

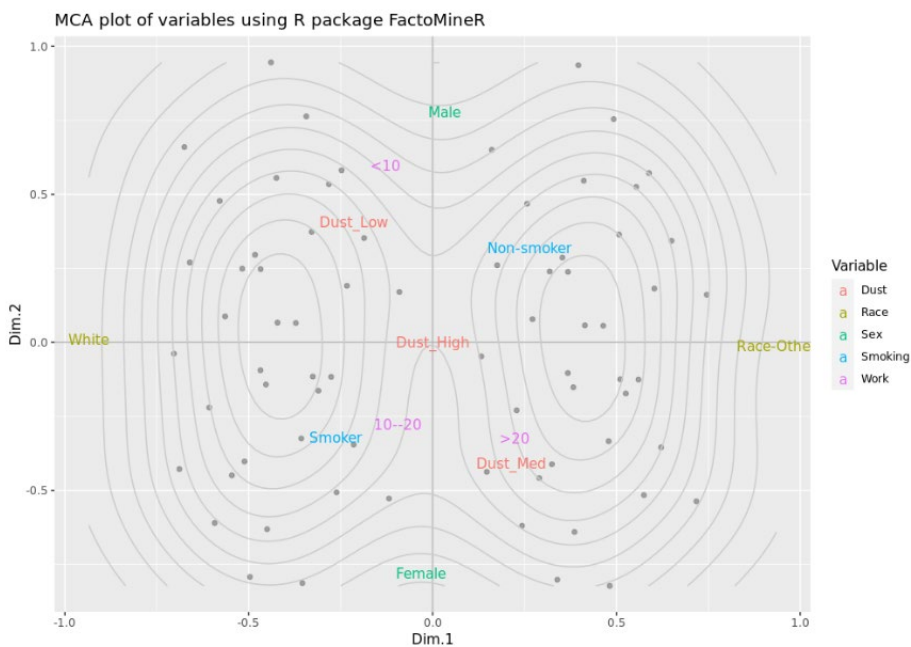
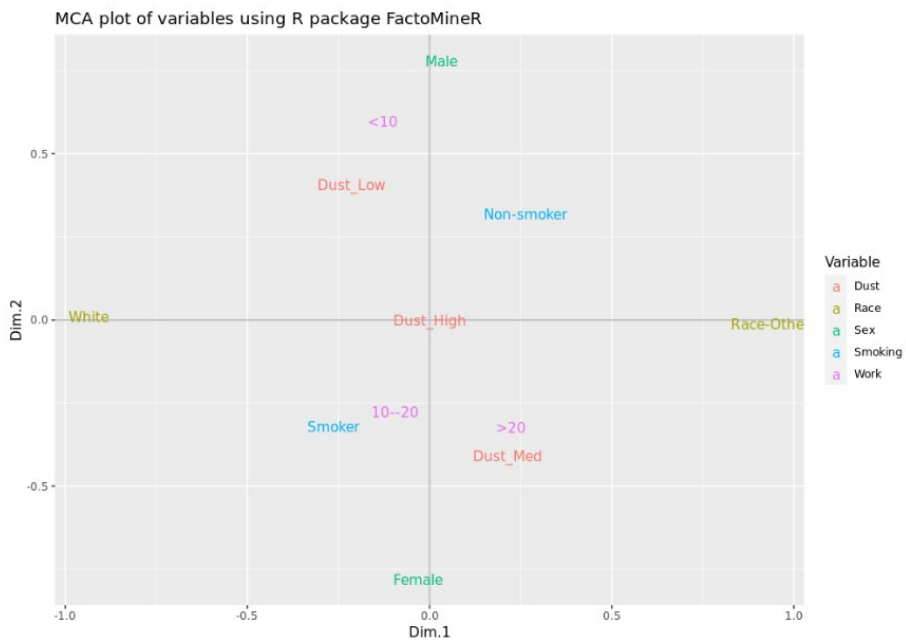
	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
1	-0.5789628	0.4778202	0.48461360	0.3768534	-0.1465078
2	-0.4829856	0.2957147	0.88313299	0.2091470	-0.3862097
3	-0.6749399	0.6599257	0.08609421	0.5445598	0.0931940
4	0.2571199	0.4684602	0.44706830	0.5353938	-0.1074280
5	0.3530971	0.2863547	0.84558769	0.3676874	-0.3471298
6	0.1611428	0.6505656	0.04854891	0.7031002	0.1322738

A matrix: 6 × 5 of type dbl

```
# number of categories per variable
cats = apply(data4[3:7], 2, function(x) nlevels(as.factor(x)))
cats
```

Dust: 3 Race: 2 Sex: 2 Smoking: 2 Work: 3

- Membuat plot



5. Interpretasi

Dari tabel nilai eigen, didapatkan dengan 2 dimensi hanya dapat menjelaskan 28% dari total inersia-nya yang artinya masih kurang bagus karena bisa saja masih terdapat beberapa informasi tersembunyi jika hanya dilihat dari 2 dimensi.

Dapat dilihat dari plot nya variabel jenis kelamin pria berasosiasi dengan lama pekerjaan kurang dari 10 tahun, tidak merokok, dan juga terkena paparan debu yang rendah. Sedangkan variabel jenis kelamin wanita berada di antara kuadran 1 dan 4 dan berasosiasi dengan variabel perokok, paparan debu sedang, dan lama bekerja 10 - 20 tahun dan lebih dari 20 tahun.

Untuk variabel terkena paparan debu yang tinggi, ras kulit putih, dan ras kulit lainnya terlihat tidak dekat dengan variabel apa pun, yang artinya mungkin saja kecil hubungan antara warna kulit ras dengan jumlah paparan debu yang di dapat.

D. Analisis Biplot

1. Data dan Sumber Data

Data yang kamu gunakan merupakan data tabel 16.17. Data tersebut kami peroleh dari buku Methods Of Multivariate Analysis Third Edition Wiley.

Table 16.17 Birth and Death Months of 1281 People

Death	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Birth													
Jan	9	14	12	14	9	11	10	15	9	11	11	13	138
Feb	13	7	6	8	9	5	4	5	7	11	1	13	89
Mar	12	14	9	7	17	10	9	10	9	2	13	9	121
Apr	7	11	10	7	9	11	4	11	4	9	8	12	103
May	8	6	11	7	4	12	9	9	11	6	9	4	96
Jun	14	5	5	7	13	5	7	4	8	9	7	5	89
Jul	12	10	13	4	5	11	7	6	4	10	4	8	94
Aug	7	7	11	13	10	9	3	8	8	8	8	12	104
Sep	7	12	11	10	10	11	4	8	12	6	8	7	106
Oct	16	8	11	9	7	10	12	8	9	8	7	7	112
Nov	7	9	15	10	11	10	7	10	7	10	9	6	111
Dec	6	18	9	11	14	8	9	9	10	6	10	8	118
Total	118	121	123	107	118	113	85	103	98	96	95	104	1281

Data tersebut merupakan data bulan kelahiran dan kematian dari 1281 orang yang berasal dari 4 kelompok dan dibuat tabel kontingensi frekuensinya seperti yang ada pada tabel 16.17 tersebut.

2. Tujuan

Analisis biplot adalah suatu metode multivariat yang menggunakan baris dan kolom dalam suatu grafik. Metode ini digunakan untuk menampilkan objek dan variabel-variabel dengan objek yang diteliti. Bilpot merupakan teknik statistik deskriptif dimensi ganda yang dapat menyajikan secara simultan segugus objek pengamatan dan variabel dalam suatu grafik pada suatu bidang datar sehingga ciri-ciri variabel dan objek pengamatan serta posisi relatif antara objek pengamatan serta posisi relatif antara objek pengamatan dengan variabel dapat dianalisis.

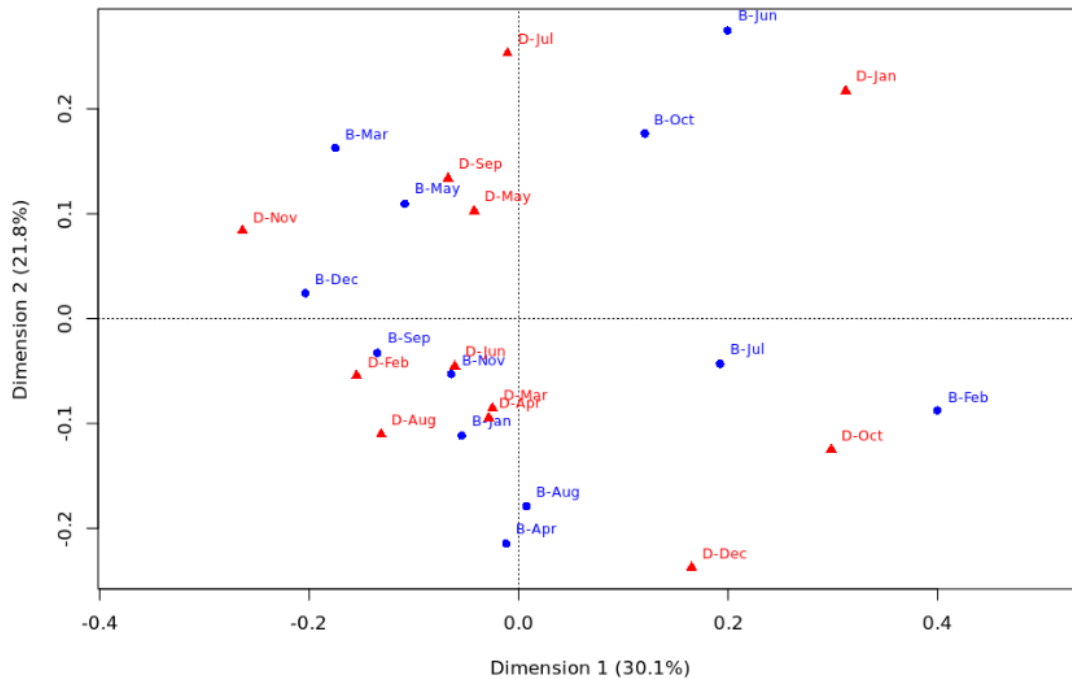
3. Langkah-Langkah

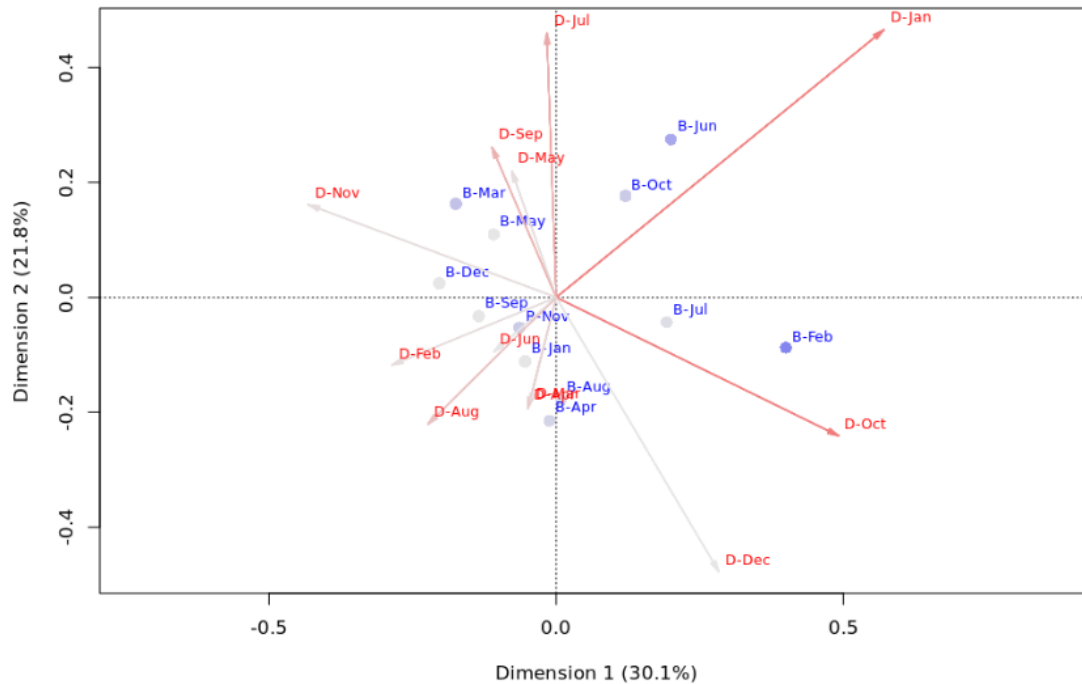
Dengan melanjutkan pada analisis korespondensi sederhana sebelumnya, maka akan dilakukan komputasi berikut ini

```
plot(fit) # symmetric map
plot(fit, mass = TRUE, contrib = "absolute", map =
     "rowgreen", arrows = c(FALSE, TRUE)) # asymmetric map
```

4. Hasil

```
plot(fit) # symmetric map
plot(fit, mass = TRUE, contrib = "absolute", map =
     "rowgreen", arrows = c(FALSE, TRUE)) # asymmetric map
```





5. Interpretasi

Selain kedekatan/kemiripan yang sudah dijelaskan dalam interpretasi analisis korepondensi sederhana, dapat dilihat juga dari tanda panah pada biplot di atas. Dilihat dari tanda panah pada plot peubah D-Jan, D-Oct, D-Dec, dan D-Jul memiliki variansi yang besar sedangkan peubah D-Mar memiliki variansi yang terkecil.

Dua tanda panah yang membentuk sudut siku atau 90° artinya antara kedua perubah tersebut tidak memiliki korelasi seperti peubah D-Jan dan juga peubah D-Dec, sedangkan dua tanda panah yang membentuk sudut lancip memiliki korelasi yang besar seperti D-Feb dan D-Aug.

Nilai relatif peubah suatu objek, di mana objek yang terletak searah dengan arah vektor peubah dapat dinyatakan bahwa objek tersebut mempunyai nilai di atas rata-rata. Contohnya dapat dilihat dari B-Jun dan B-Oct yang terletak searah dengan peubah D-Jan atau B-Dec yang searah dengan peubah D-Nov.