

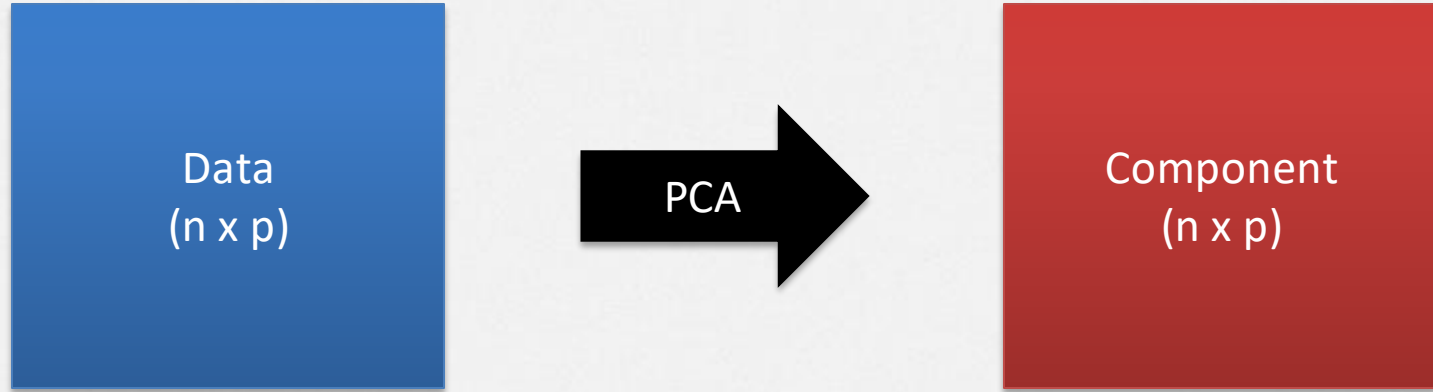
Analisis Komponen Utama (Principal Component Analysis)

Farit Mochamad Afendi
08128592194 – fmafendi@apps.ipb.ac.id

	A	B	C	D	E	F	G	H
1	respondent	brand	convenience	large_choice	low_prices	preference_score	product_quality	service_quality
2	antoine	OfficeStar	4	6	2	4	2	5
3	antoine	PaperNCo	1	5	4	3	4	1
4	antoine	OfficeEquipment	5	5	1	2	2	4
5	antoine	Supermarket	5	1	2	2	3	3
6	elisabeth	OfficeStar	2	6	2	2	4	3
7	elisabeth	PaperNCo	2	2	5	5	2	4
8	elisabeth	OfficeEquipment	5	5	4	4	2	4
9	elisabeth	Supermarket	5	2	4	3	3	3
10	hubert	OfficeStar	3	5	2	1	4	6
11	hubert	PaperNCo	2	5	4	2	2	1
12	hubert	OfficeEquipment	4	2	2	3	5	3
13	hubert	Supermarket	6	3	2	5	3	1

- ✓ Several respondents were asked to rate four brands of office equipment on six dimensions.
- ✓ How can we understand the perception of consumer on the brand?
- ✓ Reducing the data dimension may help to achieve this goal easily → principal component analysis

PCA: How it works?



- Complex structure
- Information is not easy to obtain

- Information is structured into a more meaningful way
- Usually used to reduced the dimension without losing lot of information
- Some keywords: eigen value, eigen vector, component

Basic Idea

Student	Height (cm)	Weight (kg)
A	180	80
B	175	75
C	170	60
D	160	80
E	150	45
F	155	48
G	160	60
H	162	59
I	165	60
J	150	80

Univariate

- ✓ Who's the tallest?
- ✓ Who's the shortest?
- ✓ Who's the most weight?

Bivariate – among objects

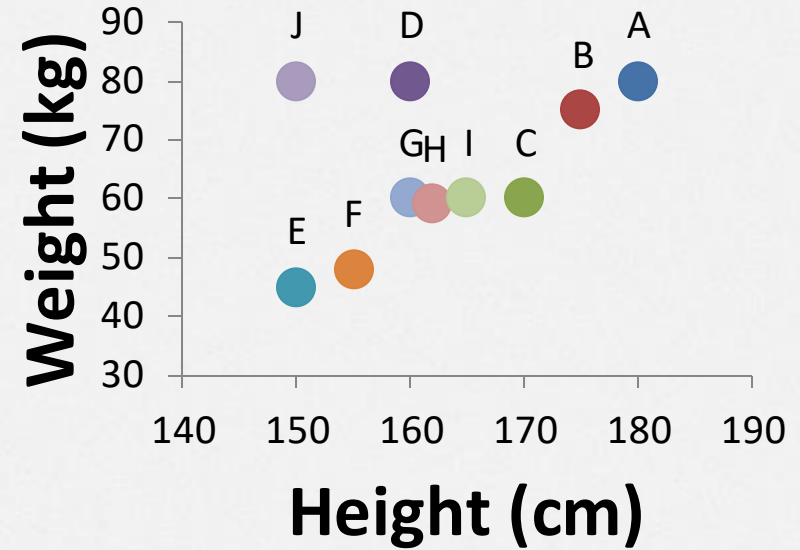
- ✓ Who has similar posture with student B?
- ✓ If we want to divide students into groups having similar posture:
 - ✓ how many groups we have?
 - ✓ How do we allocate students into that group?

Bivariate – among variables

- ✓ How close correlation between height and weight?

Basic Idea

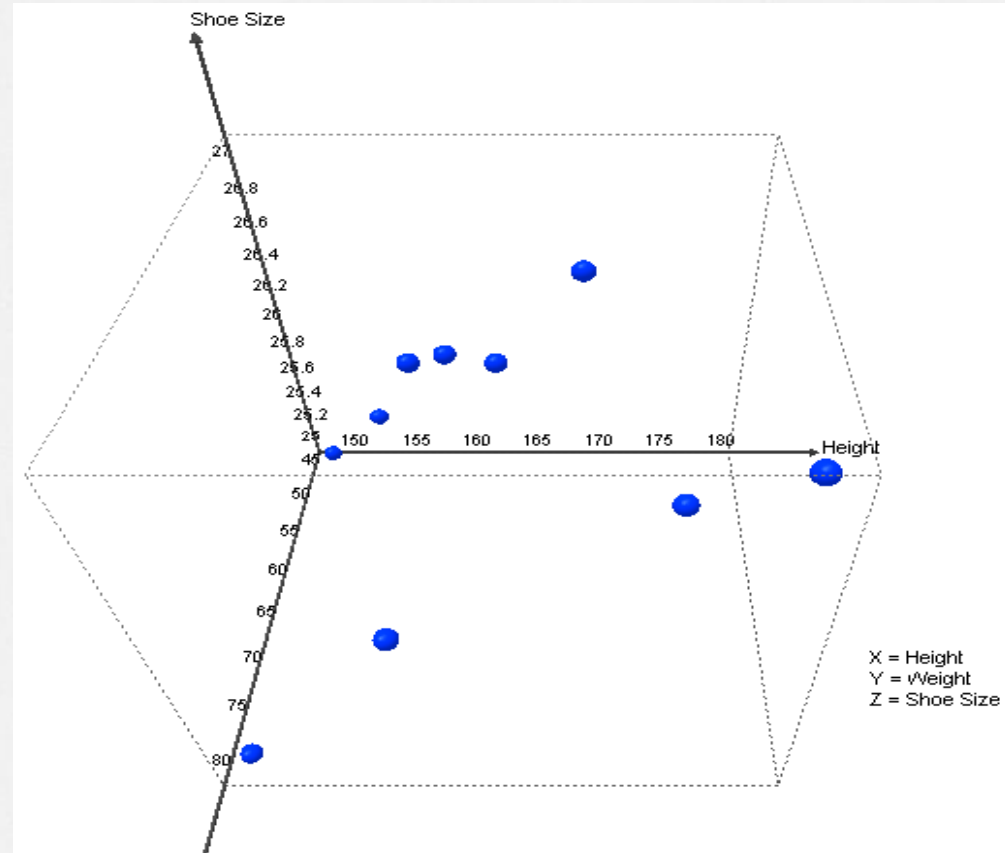
Student	Height (cm)	Weight (kg)
A	180	80
B	175	75
C	170	60
D	160	80
E	150	45
F	155	48
G	160	60
H	162	59
I	165	60
J	150	80



- ✓ Visual representation is more informative
- ✓ Variable act as dimension
- ✓ Similarity among objects is easier to detect
- ✓ Relative position of object to variable is easier to obtained

A little more complex data set

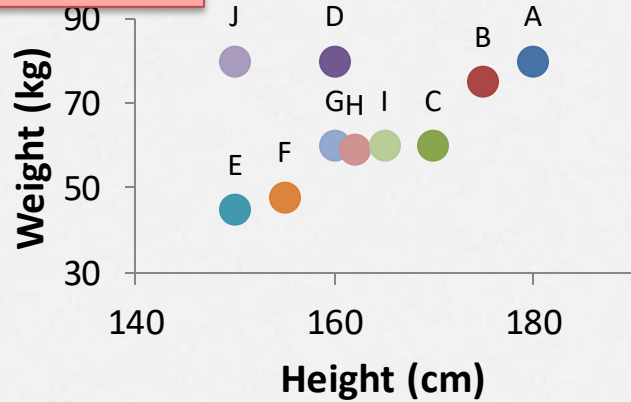
Student	Height (cm)	Weight (kg)	Shoe Size (cm)
A	180	80	27
B	175	75	26.5
C	170	60	27
D	160	80	26
E	150	45	25
F	155	48	25.5
G	160	60	26.5
H	162	59	26.5
I	165	60	26.5
J	150	80	25



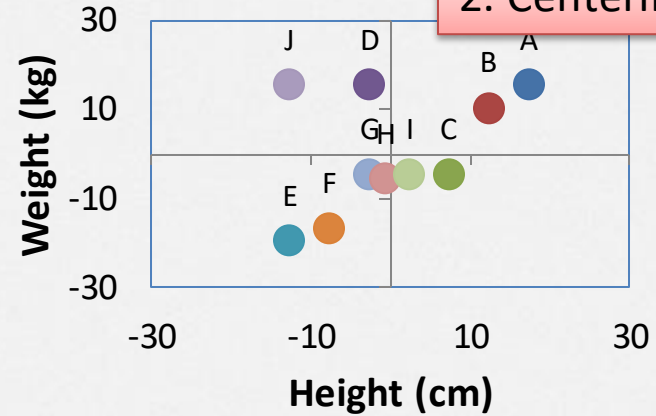
- ✓ Adding one more variable means adding one more dimension in the plot
- ✓ Difficult to observe the plot
- ✓ What should we do?

Rotating the coordinates (1)

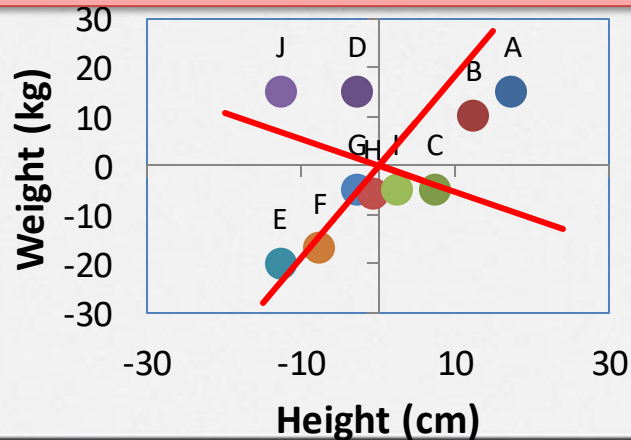
1. Original plot



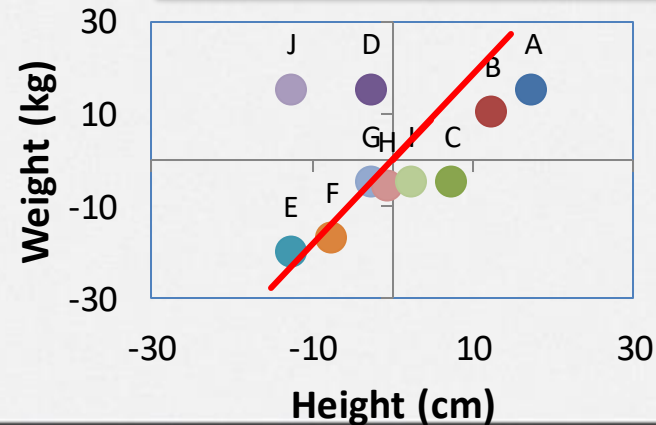
2. Centering the plot



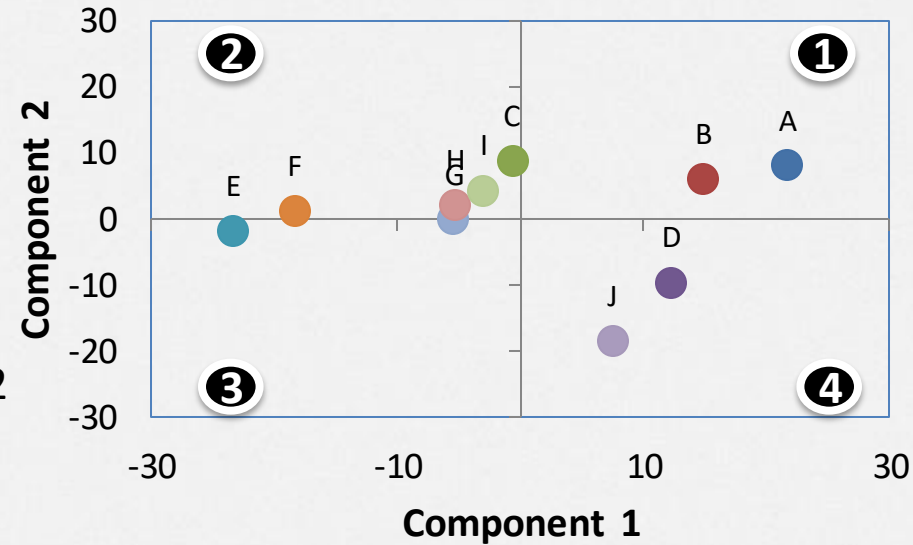
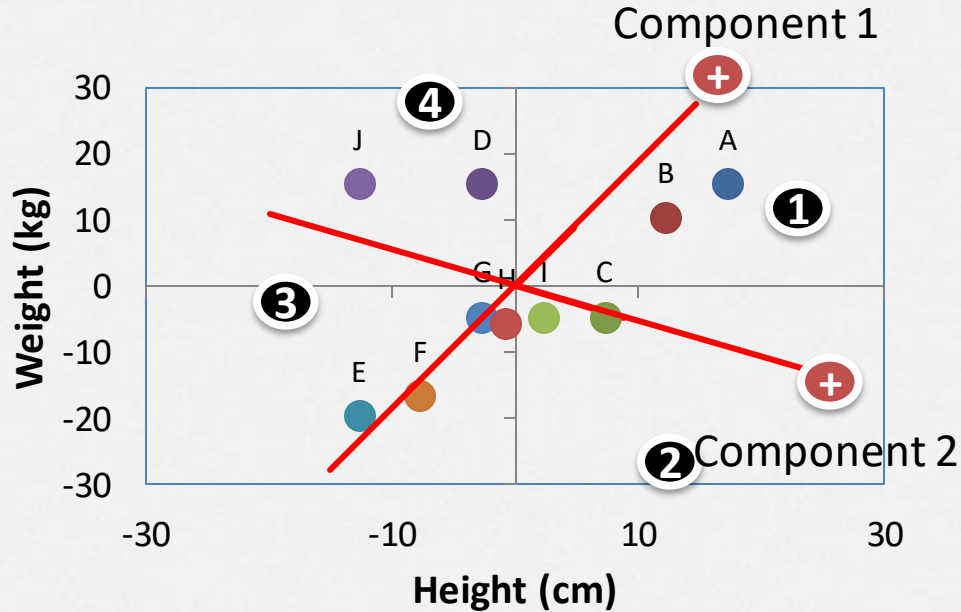
4. 2nd axis, maximizing the residual variance, perpendicular to previous axis



3. 1st axis, maximizing the variance



Rotating the coordinates (2)



Covariance matrix

	Height	Weight
Height	100.68	56.23
Weight	56.23	174.9

Eigen analysis

		Comp 1	Comp 2
Eigen value		205.16	70.41
Eigen vector	Height	0.474	0.881
	Weight	0.881	-0.474

Variance of the comp.

Rotation matrix

Rotating the coordinates (3)

Student	Height (cm)	Weight (kg)
A	180	80
B	175	75
C	170	60
D	160	80
E	150	45
F	155	48
G	160	60
H	162	59
I	165	60
J	150	80

1.Centering
2.Rotating

Student	Comp 1	Comp 2
A	21.67	7.98
B	14.90	5.95
C	-0.68	8.66
D	12.19	-9.63
E	-23.37	-1.85
F	-18.35	1.13
G	-5.42	-0.15
H	-5.35	2.08
I	-3.05	4.25
J	7.45	-18.43

Scores of the components

Rotation matrix

	Comp 1	Comp 2
Height	0.474	0.881
Weight	0.881	-0.474

Comp 1 = $0.474 \cdot \text{height} + 0.881 \cdot \text{weight}$

Comp 2 = $0.881 \cdot \text{height} - 0.474 \cdot \text{weight}$

Principal Component Analysis

Student	Height (cm)	Weight (kg)	Shoe Size (cm)
A	180	80	27
B	175	75	26.5
C	170	60	27
D	160	80	26
E	150	45	25
F	155	48	25.5
G	160	60	26.5
H	162	59	26.5
I	165	60	26.5
J	150	80	25

Singular values	Eigen values	Cumulative % of Eigenvalues
42.98	1847.66	74.35
25.23	636.56	99.96
1.00	1.00	100.00

Variables	Eigen vectors		
	PC1	PC2	PC3
Height	0.475	0.877	0.071
Weight	0.880	-0.475	-0.010
Shoe Size	0.025	0.067	-0.997

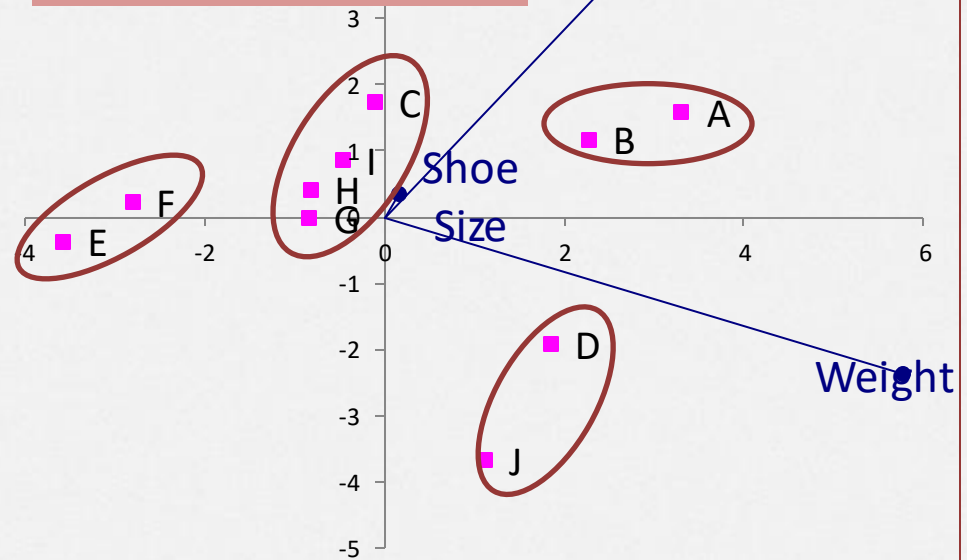
Students	Score component		
	PC1	PC2	PC3
A	0.505	0.316	0.232
B	0.347	0.235	0.424
C	-0.015	0.345	-0.286
D	0.283	-0.382	-0.187
E	-0.544	-0.074	0.434
F	-0.427	0.045	0.261
G	-0.126	-0.004	-0.495
H	-0.124	0.084	-0.344
I	-0.071	0.169	-0.141
J	0.172	-0.733	0.102

Biplot

Student	Height (cm)	Weight (kg)	Shoe Size (cm)
A	180	80	27
B	175	75	26.5
C	170	60	27
D	160	80	26
E	150	45	25
F	155	48	25.5
G	160	60	26.5
H	162	59	26.5
I	165	60	26.5
J	150	80	25

Variable	s	Correlation		
		Height	Weight	Shoe Size
Height	10.03	1		
Weight	13.22	0.424	1	
Shoe Size	0.75	0.955	0.359	1

Score and loading plot (Biplot) of PC1 vs PC2



- ❑ Circles showing cluster among students
- ❑ Magnitude of variable line represents its variance
- ❑ Cosines value of angle between two variables shows their correlation
- ❑ Relative position of objects to variables showing value of those objects on the variables

Illustration

- How can we understand the perception of consumer on the brand

	A	B	C	D	E	F	G
1	brand	large_choice	low_prices	service_quality	product_quality	convenience	preference_score
2	OfficeStar	5.2	2.1	4.2	3.7	2.7	5
3	PaperNCo	4.4	4.5	2.3	2.6	1.4	3
4	OfficeEquipment	3.9	2.6	3.1	3.1	4.7	3
5	Supermarket	2.3	4.1	1.8	2.9	5.1	1

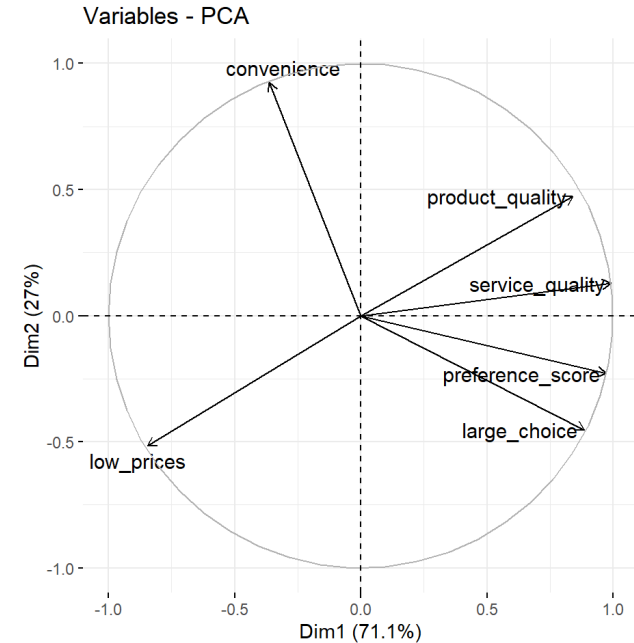
Eigen value

##	eigenvalue	percentage of variance	cumulative percentage of variance
## comp 1	4.2656310	71.093850	71.09385
## comp 2	1.6197932	26.996554	98.09040
## comp 3	0.1145758	1.909596	100.00000

- ❑ two components explain 98.1 percent of the variance in the ratings

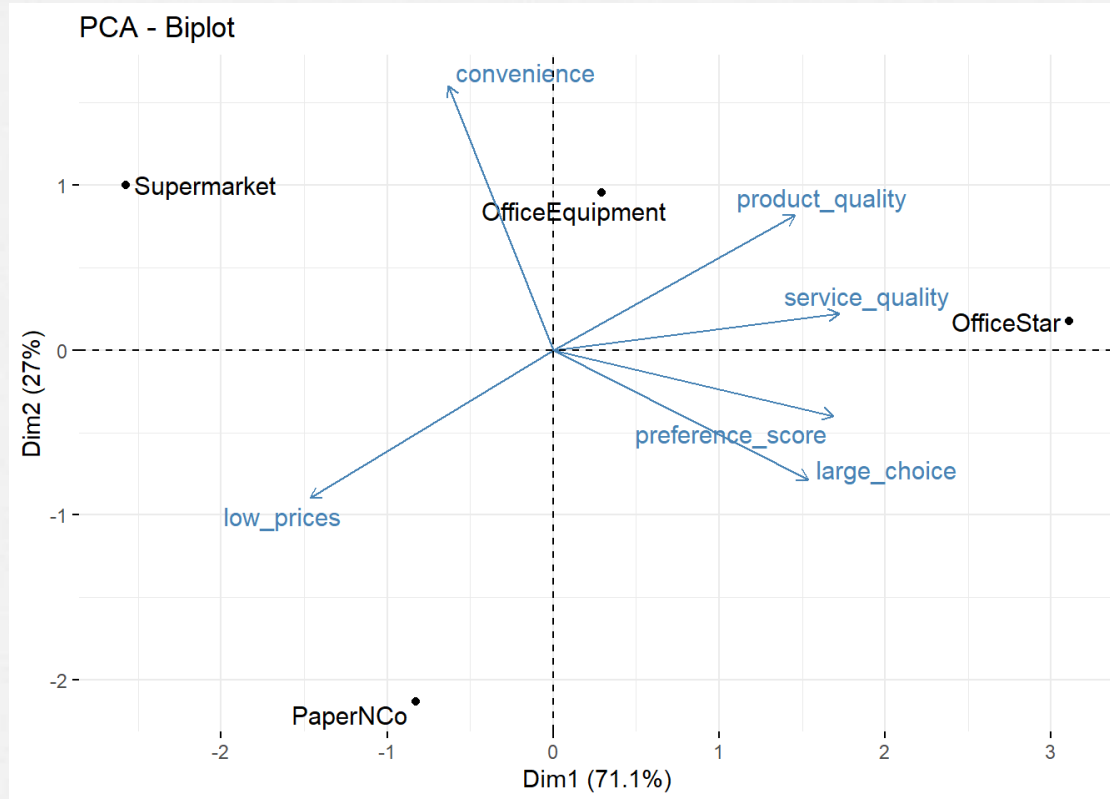
Loading

##	Dim.1	Dim.2
## large_choice	0.8851204	-0.4527205
## low_prices	-0.8448204	-0.5161931
## service_quality	0.9913907	0.1283554
## product_quality	0.8406022	0.4730741
## convenience	-0.3639026	0.9247757
## preference_score	0.9729227	-0.2299953



- ❑ the first factor describes the price and quality of the brand
- ❑ the second factor describes the convenience of the brand's stores.

Biplot



TERIMA KASIH