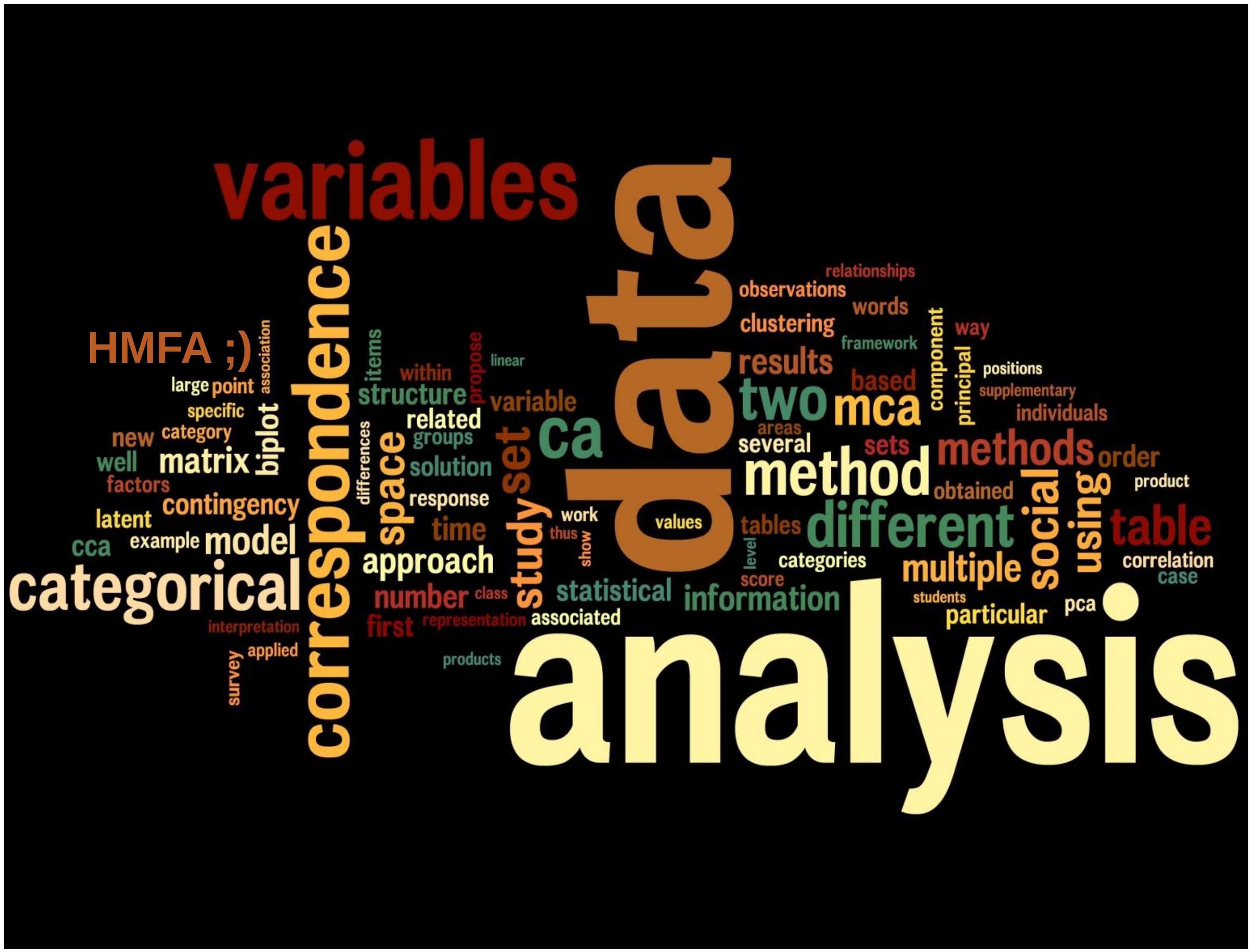
“THANK YOU FOR BEING HERE”

http://www.ted.com/talks/john_francis_walks_the_earth.html

Welcome to the 6th CARME conference



The applied mathematics department

- Jérôme PAGES : Professeur, Directeur du laboratoire
- Marine CADORET : Maître de conférences contractuelle
- David CAUSEUR : Professeur
- Thibaut DUTRION : Ingénieur d'étude
- Magalie HOUEE : Ingénieur d'étude
- François HUSSON : Maître de conférences
- Julie JOSSE : Maître de Conférences
- Sébastien LÊ : Maître de conférences
- Marie VERBANCK : Doctorante
- Elisabeth LENAULD, Karine BAGORY : Secrétaires

TODAY'S TUTORIALS

The tutorials of the day

Tutorials

- Lê Sébastien: *From one to multiple data tables with FactoMineR*
- Dray Stéphane: *Multivariate analysis of ecological data with ade4*
- Mair Patrick, de Leeuw Jan: *Multidimensional scaling using majorization with smacof*
- Nenadic Oleg, Greenacre Michael: *Correspondence analysis with ca*

What's is common to all those presentations?

The New York Times

I READ THE NEWS TODAY OH BOY…

The Beatles

A day in the life

Data Analysts Captivated by R's Power
January 6, 2009 from the New York Times

- "The popularity of R at universities could threaten SAS Institute (…)"

R You Ready for R?
January 8, 2009 from the New York Times

- "Intel Capital has placed the number of R users at 1 million, and Revolution kicks the estimate all the way up to 2 million."

What is R?

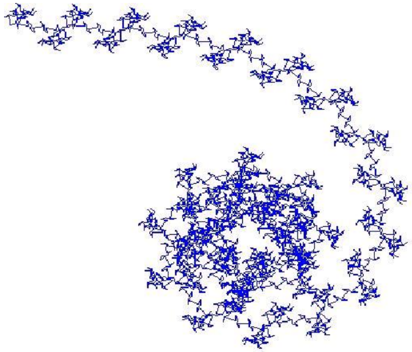
- R is a free software environment for statistical computing and graphics. (http://www.r-project.org/)
- R is a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. (http://cran.r-project.org/)
- R was designed by Ross Ihaka and Robert Gentleman

How R became a must in a decade?

- Its economic model: it's free.
- Mr. Ihaka said: "We could have chosen to be commercial, and we would have sold five copies of the software."

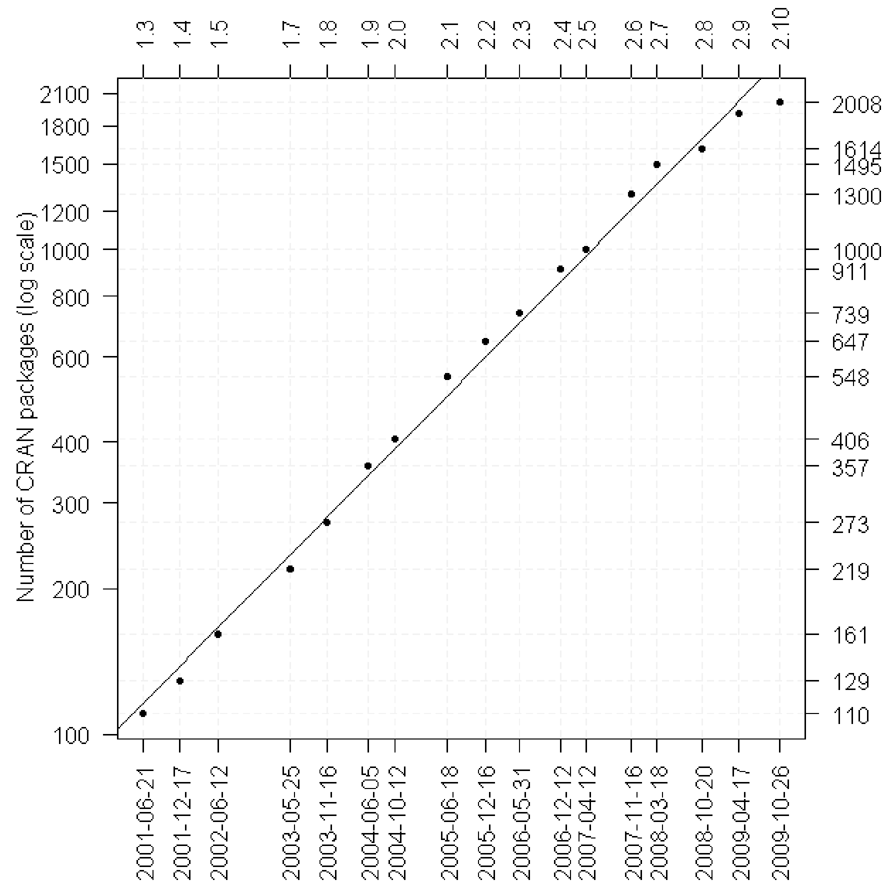
How R became a must in a decade?

- **The snowball effect:** a figurative term for a process that starts from an initial state of small significance and builds upon itself, becoming larger (graver, more serious), and perhaps potentially dangerous or disastrous (a vicious circle, a "spiral of decline"), though it might be beneficial instead (a virtuous circle) (wikipedia)



The snowball effect

Version	Packages	Date
1.3	110	21/06/2001
1.4	129	17/12/2001
1.5	161	12/06/2002
1.6	163	
1.7	219	25/05/2003
1.8	273	16/11/2003
1.9	357	05/06/2004
2.0	406	12/10/2004
2.1	548	18/06/2005
2.2	647	16/12/2005
2.3	739	31/05/2006
2.4	911	12/12/2006
2.5	1000	12/04/2007
2.6	1300	16/11/2007
2.7	1495	18/03/2008
2.8	1614	20/10/2008
2.9	1907	17/04/2009
2.10	2008	26/10/2009



Why create an R package? (P. Rossi)

There are three good reasons:

- Creating an R package forces you to document your code and provide test examples to insure that it actually works. It will also be much easier to use your code as documentation will only be a ? command away and all of your functions and shared libraries will be available for use.
- If your goal is disseminate your research, this is an ideal way of making sure others have access to your work. It will also increase the probability that eventually your work will be correct. You will also learn more about the properties of your research ideas through the experience of others.
- <u>Ease your sense of guilt by giving back something to this amazing community of volunteers!</u>

The R packages

The R packages

- There's an old bulgarian proverb that says « всеки проблем има R решение »
- In other words « each problem has its R package»

The « sudoku » package

Sudoku Puzzle Generator and Solver

Documentation for package 'sudoku' version 2.2

- DESCRIPTION file.

Help Pages

fetchSudokuUK	Fetch the daily sudoku puzzle from http://www.sudoku.org.uk/
generateSudoku	Randomly Generate a Sudoku Puzzle Grid
hintSudoku	Give a Hint for a Sudoku Cell
playSudoku	Interactively play a game of Sudoku
printSudoku	Print a Sudoku Grid to the Terminal.
readSudoku	Read a File Containing a Sudoku Grid
solveSudoku	Solve a Sudoku Puzzle
writeSudoku	Write a Sudoku Grid to a File

The « FactoMineR » package

http://factominer.free.fr

Journal of statistical software **FactoMineR: an R package for multivariate analysis**

FROM MULTIVARIATE TO MULTIPLE TABLES DATA ANALYSIS…AN OVERVIEW

Sébastien Lê and Jérôme Pagès

Objectives

- To understand what can be expected from multiway data methods
- To understand the motivations and the framework of Canonical Analysis (CA)
- To understand Generalized Canonical Analysis (GCA)
- To be able to place Multiple Factor Analysis *vs.* GCA

Outline

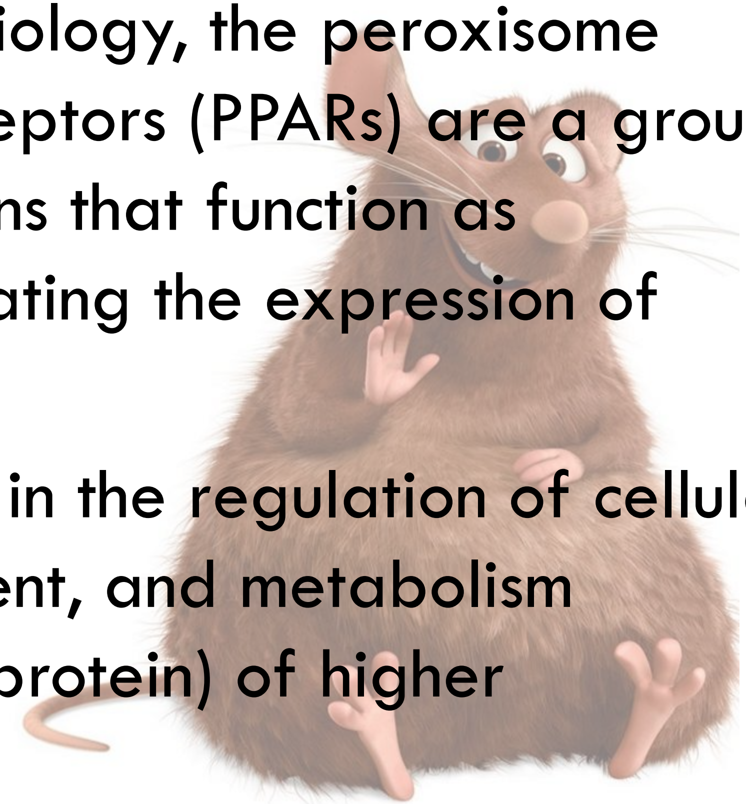
- What you know
- What you want to do and why you want to do it
- How to do it
- How to do it with R

The data

- 40 mice
- 2 genotypes (wild, PPARα-deficient)
- 5 diets (dha, efad, lin, ref, tsol)
- 120 genes (expression)
- 21 hepatic fatty acid concentration
- Thanks to Sandrine Lagarrigue, Genetics department-INRA Rennes, for the availability of the data

PPARα

- In the field of molecular biology, the peroxisome proliferator-activated receptors (PPARs) are a group of nuclear receptor proteins that function as transcription factors regulating the expression of genes.
- PPARs play essential roles in the regulation of cellular differentiation, development, and metabolism (carbohydrate, lipid, and protein) of higher organisms.
- PPARα are expressed in liver, kidney, heart, muscle, adipose tissue, and others



The diets

- dha: diet rich in fatty acids of the Omega 3 family and particularly docosahexaenoic acid (DHA), based on fish oil;
- efad (Essential Fatty Acid Deficient): diet based on of saturated fatty acids only, made from hydrogenated coconut oil;
- lin: diet rich in Omega 3, made from linseed oil;
- ref: regime whose contribution in Omega 6 and Omega 3 is adapted for the French population, seven times more Omega 6 than Omega 3;
- tsol: diet rich in Omega 6, based on sunflower oil.

Issues

MY FIRST PCA

with supplementary qualitative variables (and FactoMineR)

The dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Souris	Regime	Genotype	C14.0	C16.0	C18.0	C16.1n.9	C16.1n.7	C18.1n.9	X36b4	ACAT1	ACAT2	ACBP	ACC1	ACC2	ACO
2	67	lin	1	0.34	26.45	10.22	0.35	3.10	16.98	-0.42	-0.65	-0.84	-0.34	-1.29	-1.13	-0.93
3	68	tournesol	1	0.38	24.04	9.93	0.55	2.54	20.07	-0.44	-0.68	-0.91	-0.32	-1.23	-1.06	-0.99
4	69	tournesol	1	0.36	23.70	8.96	0.55	2.65	22.89	-0.48	-0.74	-1.1	-0.46	-1.3	-1.09	-1.06
5	75	dha	1	0.22	25.48	8.14	0.49	2.82	21.92	-0.45	-0.69	-0.65	-0.41	-1.26	-1.09	-0.93
6	88	ref	1	0.37	24.80	9.63	0.46	2.85	21.38	-0.42	-0.71	-0.54	-0.38	-1.21	-0.89	
7	89	efad	1	1.70	26.04	6.59	0.66	7.26	28.23	-0.43	-0.69	-0.8	-0.32	-1.13	-0.79	-0.93
8	90	lin	1	0.35	25.94	9.68	0.36	3.60	17.62	-0.53	-0.62	-1	-0.44	-1.22	-1	-0.94
9	92	lin	1	0.34	28.63	9.95	0.29	3.27	17.02	-0.49	-0.69	-0.91	-0.37	-1.29	-1.06	-1.05
10	96	dha	1	0.22	25.34	8.81	0.44	2.36	18.39	-0.36	-0.66	-0.74	-0.39	-1.15	-1.08	-0.88
11	98	efad	1	1.38	28.49	5.63	0.90	7.01	36.68	-0.5	-0.62	-0.79	-0.36	-1.21	-0.82	-0.92
12	99	dha	1	0.26	25.73	8.30	0.43	2.74	21.75	-0.4	-0.6	-0.55	-0.25	-1.22	-1.13	-0.81
13	101	ref	1	0.44	24.28	8.63	0.53	3.33	23.86	-0.52	-0.66	-0.66	-0.41	-1.28	-1.1	-0.95
14	102	tournesol	1	0.32	24.63	9.99	0.45	2.39	17.93	-0.52	-0.63	-0.99	-0.43	-1.24	-0.96	-0.96
15	111	ref	1	0.34	26.04	9.81	0.35	2.36	20.14	-0.47	-0.71	-0.44	-0.45	-1.44	-1.17	-1.02
16	118	tournesol	1	0.35	24.76	9.38	0.54	2.47	19.66	[illegible]	-0.66	-0.88	-0.33	-1.24	-1.03	-0.92
17	120	lin	1	0.24	26.46	10.97	0.31	2.81	14.69	[illegible]	-0.62	-0.85	-0.42	-1.33	-1.19	-0.91
18	121	efad	1	1.21	23.45	5.59	0.67	6.31	33.84	[illegible]	-0.69	-0.45	-0.32	-1.31	-0.93	-1.06
19	123	dha	1	0.30	29.72	8.95	0.45	2.86	17.79	-0.36	-0.58	-0.71	-0.34	-1.2	-0.98	-0.92
20	130	efad	1	1.30	27.00	5.72	0.81	7.86	33.50	-0.35	-0.62	-0.56	-0.24	-1.24	-0.88	-0.95
21	135	ref	1	0.38	24.09	8.22	0.60	3.89	24.61	-0.44	-0.75	-0.7	-0.33	-1.35	-1.04	-0.99
22	281	efad	2	3.24	23.59	2.68	1.11	13.09	35.61	-0.48	-0.71	-0.63	-0.55	-1.27	-1.03	-0.88
23	285	ref	2	0.60	19.95	3.18	1.21	4.89	35.91	-0.5	-0.69	-0.86	-0.58	-1.31	-1.06	-0.88
24	290	tournesol	2	0.38	17.64	6.99	0.74	2.58	21.23	-0.54	-0.69	-1.02	-0.51	-1.32	-1.2	-0.96
25	294	dha	2	0.44	22.73	4.71	0.75	2.27	25.10	-0.39	-0.7	-0.83	-0.47	-1.27	-1.09	-0.91
26	295	tournesol	2	0.47	14.65	4.29	0.66	2.88	23.15	-0.54	-0.63	-0.86	-0.56	-1.35	-1.17	-0.88
27	303	ref	2	0.64	20.49	2.71	1.09	4.05	38.32	-0.35	-0.52	-0.74	-0.4	-1.07	-0.83	-0.73
28	305	ref	2	0.52	18.44	5.21	0.87	3.37	32.04	-0.46	-0.69	-0.53	-0.51	-1.33	-0.99	-0.9
29	307	lin	2	0.49	17.72	6.02	0.58	3.77	21.04	-0.42	-0.56	-0.86	-0.57	-1.23	-1.04	-0.75
30	319	dha	2	0.40	21.70	6.12	0.78	2.07	22.49	-0.51	-0.7	-0.68	-0.46	-1.34	-1.13	-0.94
31	321	lin	2	0.61	16.25	4.55	0.61	5.42	21.43	-0.47	-0.74	-0.88	-0.58	-1.3	-1.03	-0.83
32	325	efad	2	3.19	22.91	3.60	0.99	13.90	33.55	-0.54	-0.64	-0.39	-0.46	-1.3	-1.03	-0.85

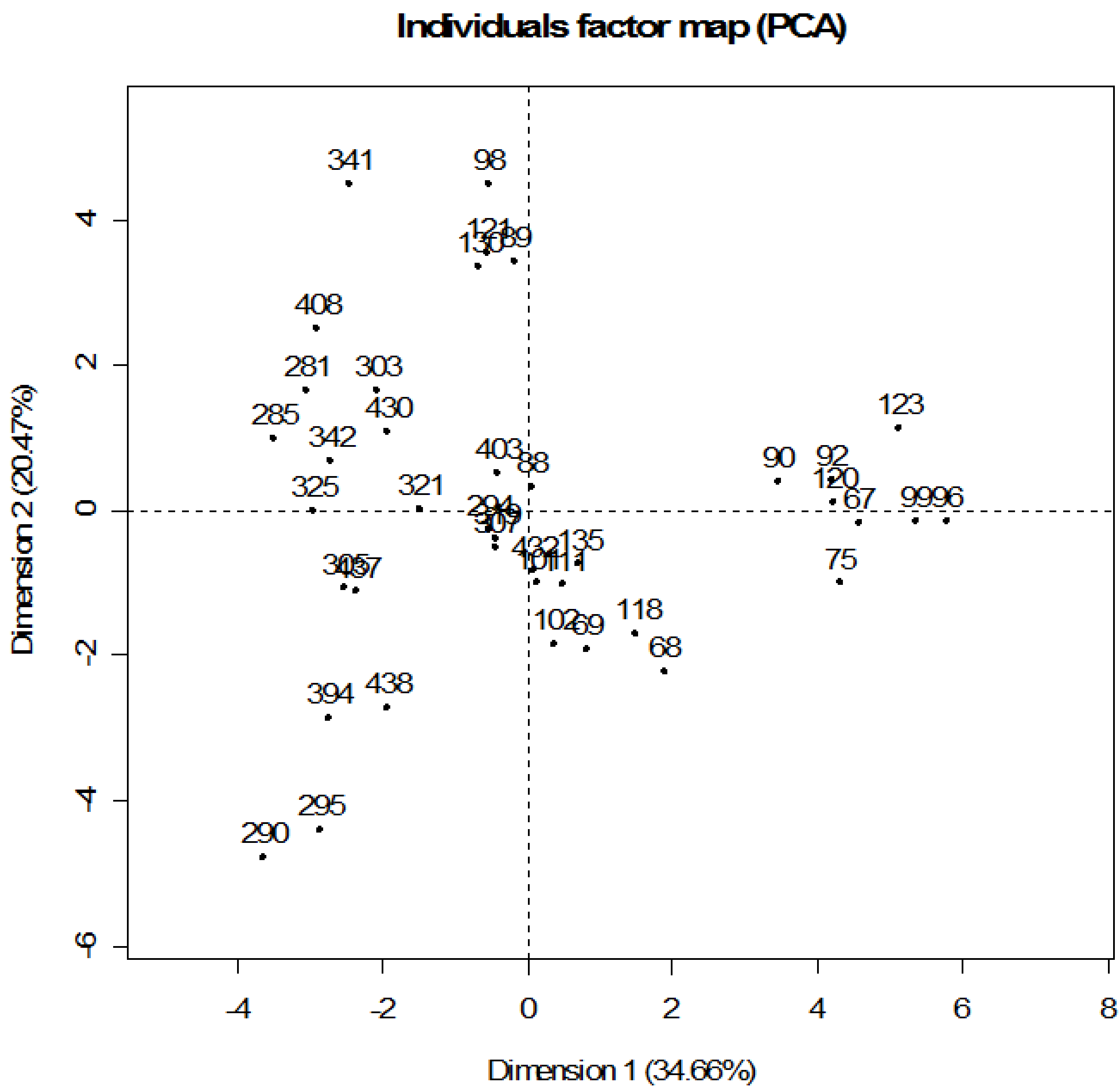
X

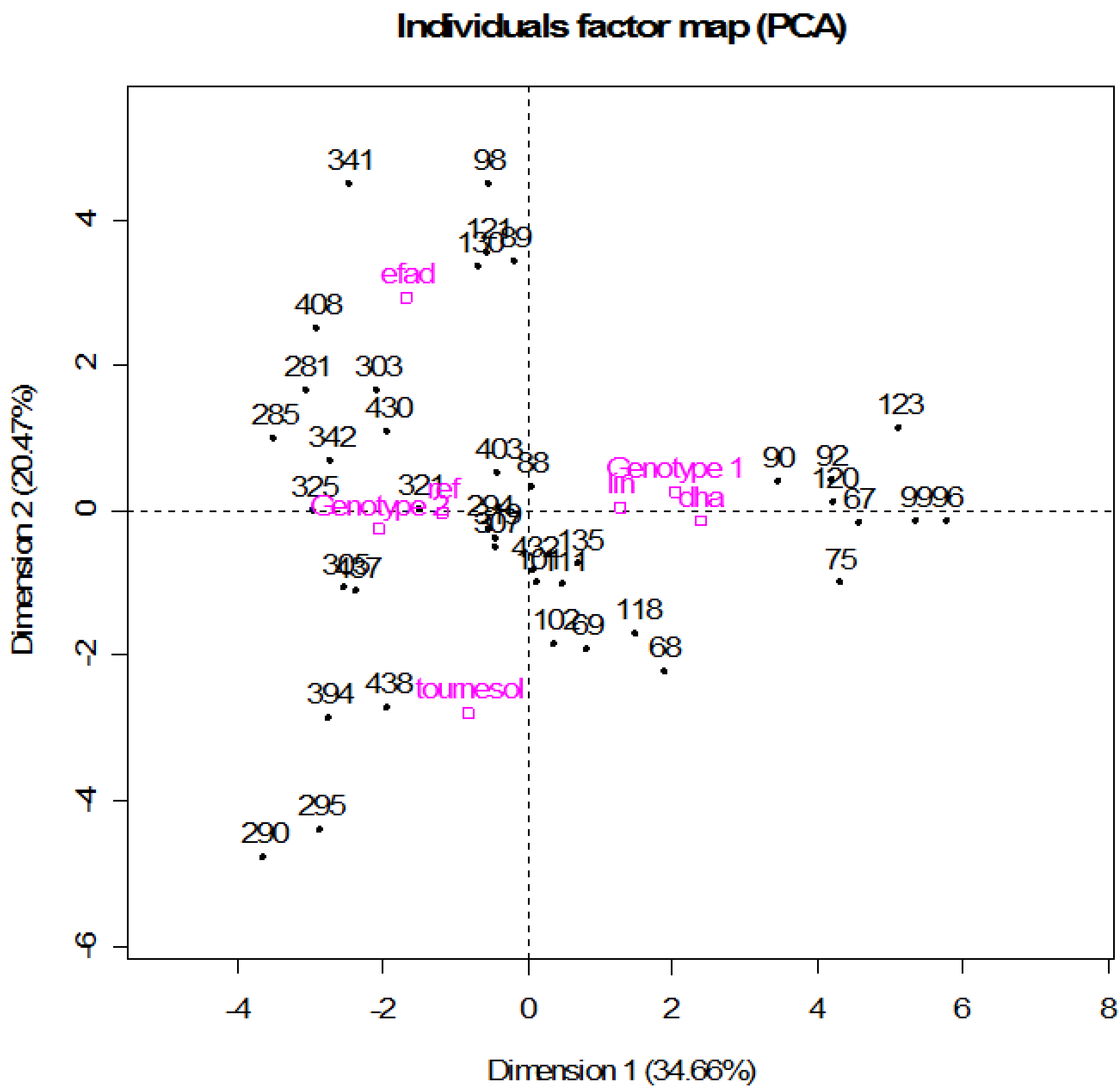
souris

The dataset

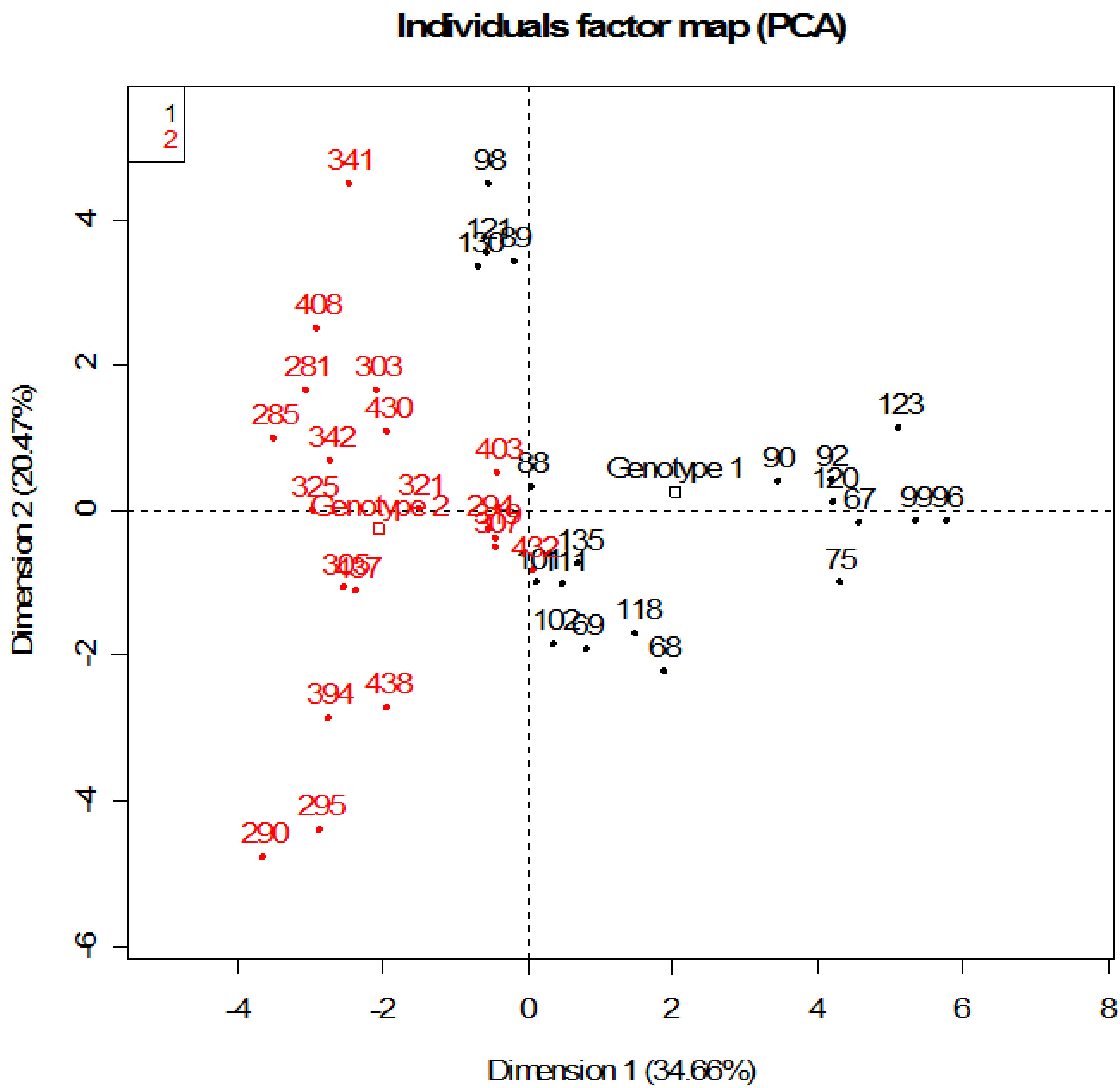
$$X \Rightarrow \Sigma = r(i,j)$$

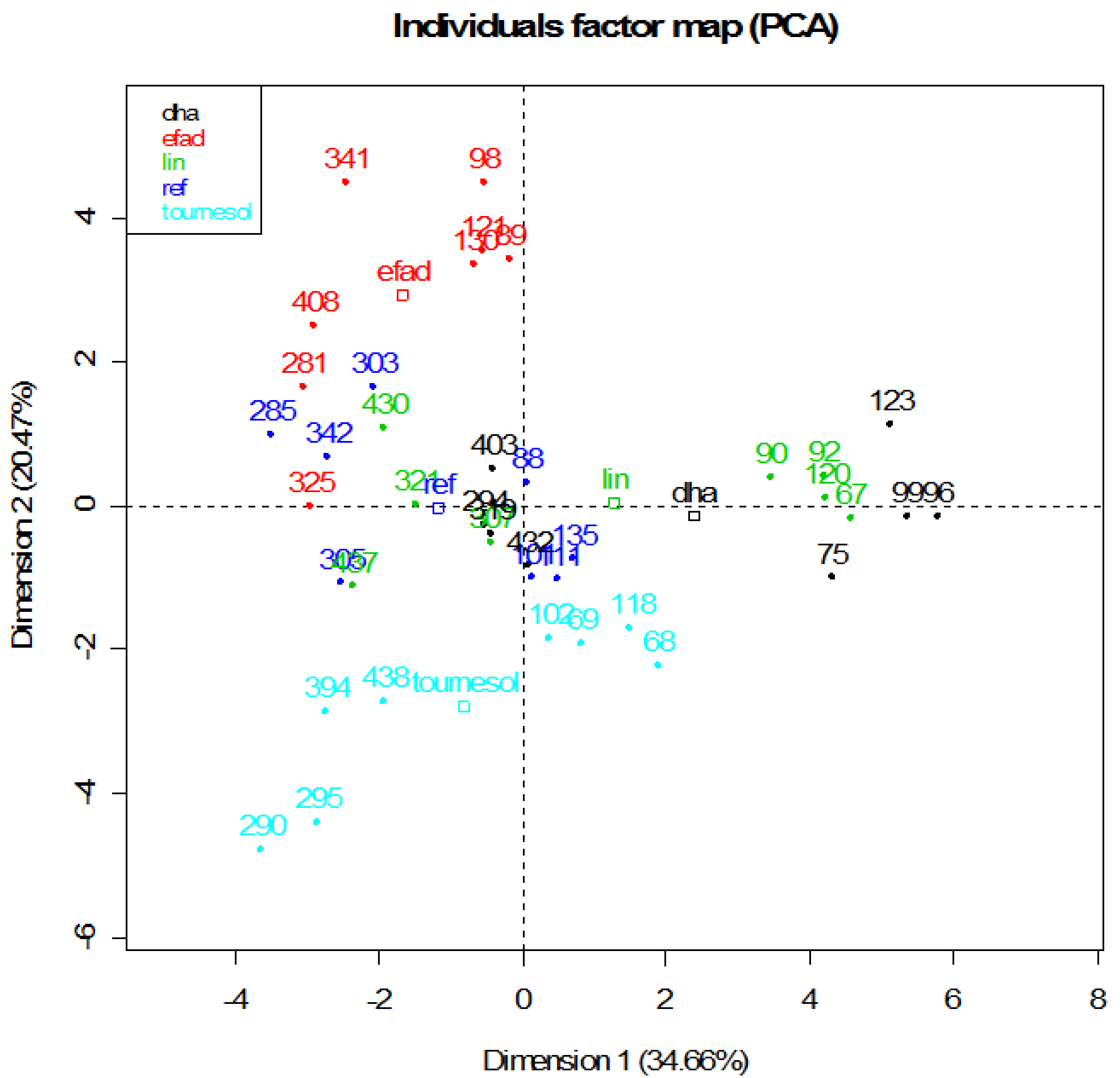
$M(n,p)$ $M(p,p)$


Individuals factor map (PCA)
Dimension 1 (34.66%)
Dimension 2 (20.47%)
-4
-2
0
2
4
6
8
-6
341
98
408
281
303
285
430
342
403
88
325
321
123
90
92
120
67
135
75
118
68
394
438
295
290

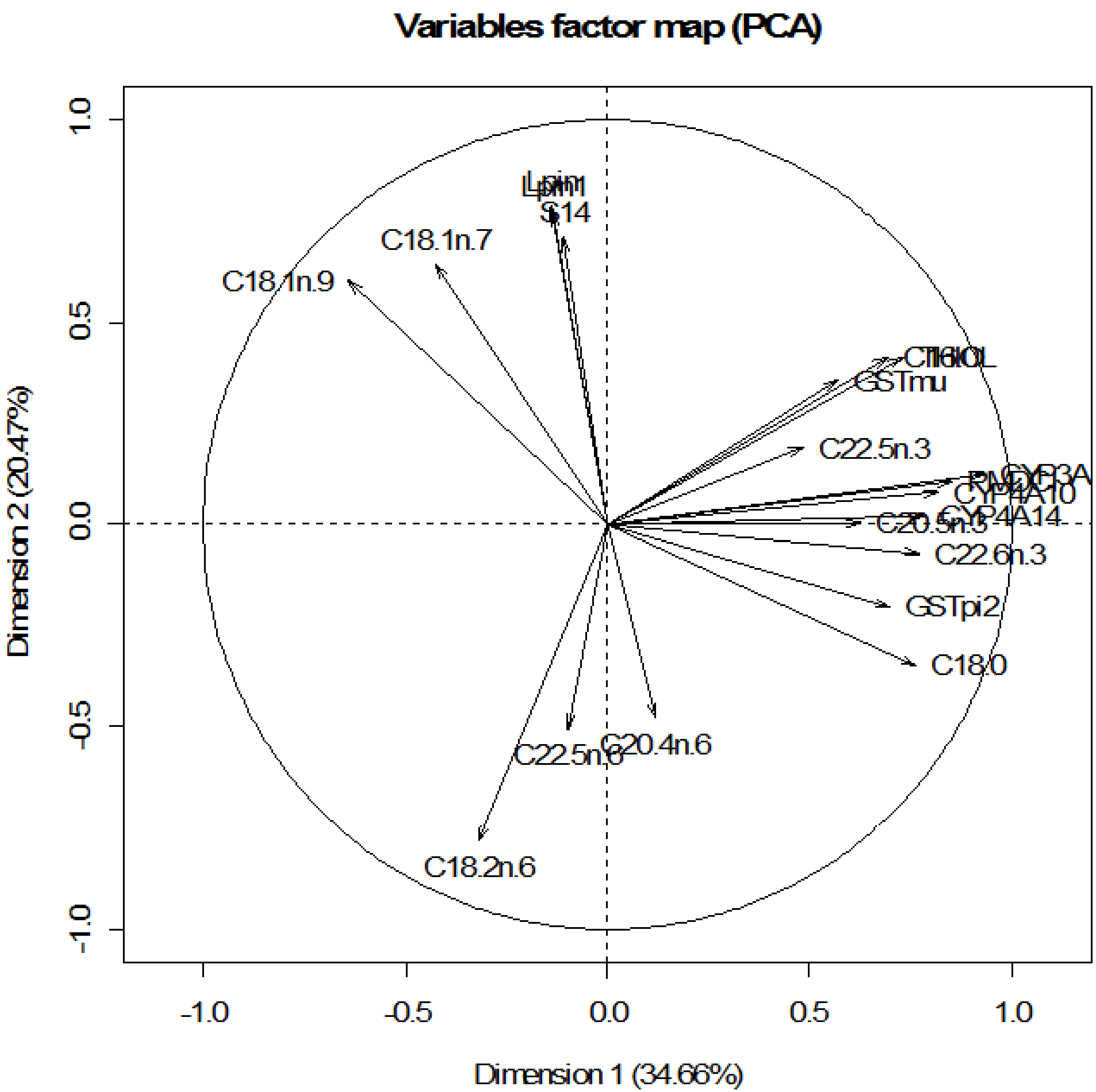


Individuals factor map (PCA)
Dimension 1 (34.66%)
Dimension 2 (20.47%)
-4
-2
0
2
4
6
8
-6
efad
Genotype 1
lin
dha
Genotype 2
ref
tournesol
341
98
121
130
408
281
303
285
430
342
403
88
325
321
90
92
120
67
123
75
135
118
102
68
394
438
295
290


Individuals factor map (PCA)
1
2
341
98
121
130
89
408
281
303
285
430
342
403
88
Genotype 1
90
92
120
123
67
99
96
325
321
Genotype 2
294
307
432
135
101
111
75
305
457
118
102
69
68
394
438
295
290
Dimension 2 (20.47%)
Dimension 1 (34.66%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8



Individuals factor map (PCA)
dha
efad
lin
ref
tournesol
Dimension 1 (34.66%)
Dimension 2 (20.47%)



Variables factor map (PCA)
C18.1n.7
C18.1n.9
S14
GSTmu
C22.5n.3
CYP4A10
CYP4A14
C22.6n.3
GSTpi2
C18.0
C22.5n.6
C20.4n.6
C18.2n.6
1.0
0.5
0.0
-0.5
-1.0
Dimension 1 (34.66%)
Dimension 2 (20.47%)

MY SECOND PCA

with supplementary qualitative and quantitative variables (and FactoMineR)

The dataset(s)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Souris	Regime	Genotype	C14.0	C16.0	C18.0	C16.1n.9	C16.1n.7	C18.1n.9	X36b4	ACAT1	ACAT2	ACBP	ACC1	ACC2	ACO
2	67	lin	1	0.34	26.45	10.22	0.35	3.10	16.98	-0.42	-0.65	-0.84	-0.34	-1.29	-1.13	-0.93
3	68	tournesol	1	0.38	24.04	9.93	0.55	2.54	20.07	-0.44	-0.68	-0.91	-0.32	-1.23	-1.06	-0.99
4	69	tournesol	1	0.36	23.70	8.96	0.55	2.65	22.89	-0.48	-0.74	-1.1	-0.46	-1.3	-1.09	-1.06
5	75	dha	1	0.22	25.48	8.14	0.49	2.82	21.92	-0.45	-0.69	-0.65	-0.41	-1.26	-1.09	-0.93
6	88	ref	1	0.37	24.80	9.63	0.46	2.85	21.38	-0.42	-0.71	-0.54	-0.38	-1.21	-0.89	
7	89	efad	1	1.70	26.04	6.59	0.66	7.26	28.23	-0.43	-0.69	-0.8	-0.32	-1.13	-0.79	-0.93
8	90	lin	1	0.35	25.94	9.68	0.36	3.60	17.62	-0.53	-0.62	-1	-0.44	-1.22	-1	-0.94
9	92	lin	1	0.34	28.63	9.95	0.29	3.27	17.02	-0.49	-0.69	-0.91	-0.37	-1.29	-1.06	-1.05
10	96	dha	1	0.22	25.34	8.81	0.44	2.36	18.39	-0.36	-0.66	-0.74	-0.39	-1.15	-1.08	-0.88
11	98	efad	1	1.38	28.49	5.63	0.90	7.01	36.68	-0.5	-0.62	-0.79	-0.36	-1.21	-0.82	-0.92
12	99	dha	1	0.26	25.73	8.30	0.43	2.74	21.75	-0.4	-0.6	-0.55	-0.25	-1.22	-1.13	-0.81
13	101	ref	1	0.44	24.28	8.63	0.53	3.33	23.86	-0.52	-0.66	-0.66	-0.41	-1.28	-1.1	-0.95
14	102	tournesol	1	0.32	24.63	9.99	0.45	2.39	17.93	-0.52	-0.63	-0.99	-0.43	-1.24	-0.96	-0.96
15	111	ref	1	0.34	26.04	9.81	0.35	2.36	20.14	-0.47	-0.71	-0.44	[illegible]	-1.44	-1.17	-1.02
16	118	tournesol	1	0.35	24.76	9.38	[illegible]	2.47	19.66	-0.42	-0.66	-0.88	[illegible]	-1.24	-1.03	-0.92
17	120	lin	1	0.24	26.46	10.97	[illegible]	2.81	14.69	-0.58	-0.62	-0.85	[illegible]	-1.33	-1.19	-0.91
18	121	efad	1	1.21	23.45	5.59	[illegible]	[illegible]	33.84	-0.46	-0.69	-0.45	[illegible]	[illegible]	-0.93	-1.06
19	123	dha	1	0.30	29.72	8.95	0.45	[illegible]	17.79	-0.36	-0.58	-0.71	-0.34	-1.2	-0.98	-0.92
20	130	efad	1	1.30	27.00	5.72	0.81	[illegible]	33.50	-0.35	-0.62	-0.56	-0.24	[illegible]	-0.88	-0.95
21	135	ref	1	0.38	24.09	8.22	0.60	[illegible]	24.61	-0.44	-0.75	-0.7	-0.33	[illegible]	-1.04	-0.99
22	281	efad	2	3.24	23.59	2.68	1.11	13.09	35.61	-0.48	-0.71	-0.63	-0.55	-1.27	-1.03	-0.88
23	285	ref	2	0.60	19.95	3.18	1.21	4.89	35.91	-0.5	-0.69	-0.86	-0.58	-1.31	-1.06	-0.88
24	290	tournesol	2	0.38	17.64	6.99	0.74	2.58	21.23	-0.54	-0.69	-1.02	-0.51	-1.32	-1.2	-0.96
25	294	dha	2	0.44	22.73	4.71	0.75	2.27	25.10	-0.39	-0.7	-0.83	-0.47	-1.27	-1.09	-0.91
26	295	tournesol	2	0.47	14.65	4.29	0.66	2.88	23.15	-0.54	-0.63	-0.86	-0.56	-1.35	-1.17	-0.88
27	303	ref	2	0.64	20.49	2.71	1.09	4.05	38.32	-0.35	-0.52	-0.74	-0.4	-1.07	-0.83	-0.73
28	305	ref	2	0.52	18.44	5.21	0.87	3.37	32.04	-0.46	-0.69	-0.53	-0.51	-1.33	-0.99	-0.9
29	307	lin	2	0.49	17.72	6.02	0.58	3.77	21.04	-0.42	-0.56	-0.86	-0.57	-1.23	-1.04	-0.75
30	319	dha	2	0.40	21.70	6.12	0.78	2.07	22.49	-0.51	-0.7	-0.68	-0.46	-1.34	-1.13	-0.94
31	321	lin	2	0.61	16.25	4.55	0.61	5.42	21.43	-0.47	-0.74	-0.88	-0.58	-1.3	-1.03	-0.83
32	325	efad	2	3.19	22.91	3.60	0.99	13.90	33.55	-0.54	-0.64	-0.39	-0.46	-1.3	-1.03	-0.85

X_1

X_2

souris

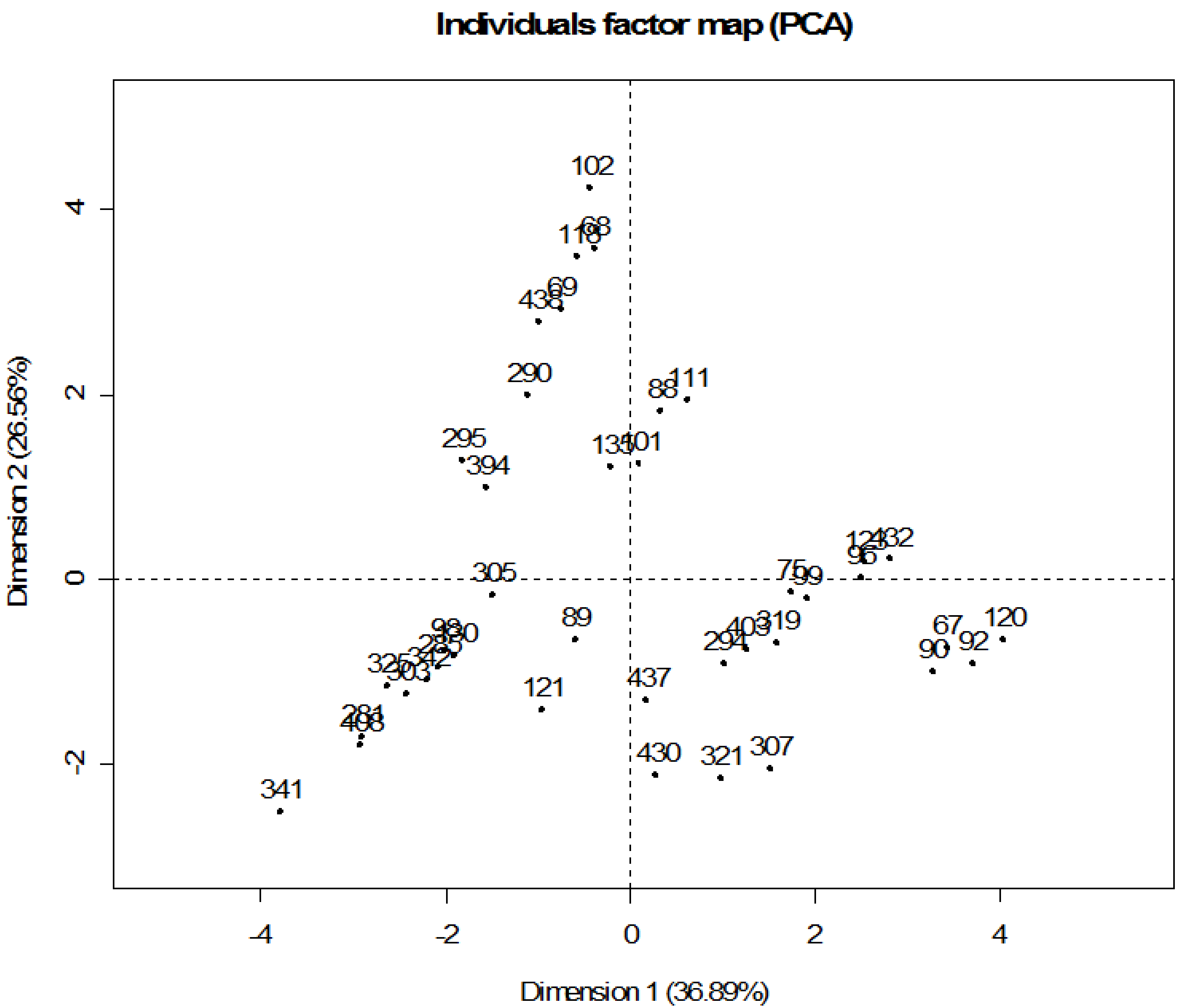
The dataset(s)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Souris	Regime	Genotype	C14.0	C16.0	C18.0	C16.1n.9	C16.1n.7	C18.1n.9	X36b4	ACAT1	ACAT2	ACBP	ACC1	ACC2	ACO
2	67	lin	1	0.34	26.45	10.22	0.35	3.10	16.98	-0.42	-0.65	-0.84	-0.34	-1.29	-1.13	-0.93
3	68	tournesol	1	0.38	24.04	9.93	0.55	2.54	20.07	-0.44	-0.68	-0.91	-0.32	-1.23	-1.06	-0.99
4	69	tournesol	1	0.36	23.70	8.96	0.55	2.65	22.89	-0.48	-0.74	-1.1	-0.46	-1.3	-1.09	-1.06
5	75	dha	1	0.22	25.48	8.14	0.49	2.82	21.92	-0.45	-0.69	-0.65	-0.41	-1.26	-1.09	-0.93
6	88	ref	1	0.37	24.80	9.63	0.46	2.85	21.38	-0.42	-0.71	-0.54	-0.38	-1.21	-0.89	
7	89	efad	1	1.70	26.04	6.59	0.66	7.26	28.23	-0.43	-0.69	-0.8	-0.32	-1.13	-0.79	-0.93
8	90	lin	1	0.35	25.94	9.68	0.36	3.60	17.62	-0.53	-0.62	-1	-0.44	-1.22	-1	-0.94
9	92	lin	1	0.34	28.63	9.95	0.29	3.27	17.02	-0.49	-0.69	-0.91	-0.37	-1.29	-1.06	-1.05
10	96	dha	1	0.22	25.34	8.81	0.44	2.36	18.39	-0.36	-0.66	-0.74	-0.39	-1.15	-1.08	-0.88
11	98	efad	1	1.38	28.49	5.63	0.90	7.01	36.68	-0.5	-0.62	-0.79	-0.36	-1.21	-0.82	-0.92
12	99	dha	1	0.26	25.73	8.30	0.43	2.74	21.75	-0.4	-0.6	-0.55	-0.25	-1.22	-1.13	-0.81
13	101	ref	1	0.44	24.28	8.63	0.53	3.33	23.86	-0.52	-0.66	-0.66	-0.41	-1.28	-1.1	-0.95
14	102	tournesol	1	0.32	24.63	9.99	0.45	2.39	17.93	-0.52	-0.63	-0.99	-0.43	-1.24	-0.96	-0.96
15	111	ref	1	0.34	26.04	9.81	0.35	2.36	20.14	-0.47	-0.71	-0.44	-0.45	-1.44	-1.17	-1.02
16	118	tournesol	1	0.35	24.76	9.38	[illegible]	2.47	19.66	-0.42	-0.66	-0.88	[illegible]	-1.24	-1.03	-0.92
17	120	lin	1	0.24	26.46	10.97	[illegible]	2.81	14.69	-0.58	-0.62	-0.85	[illegible]	-1.33	-1.19	-0.91
18	121	efad	1	1.21	23.45	5.59	[illegible]	[illegible]	33.84	-0.46	-0.69	-0.45	[illegible]	[illegible]	-0.93	-1.06
19	123	dha	1	0.30	29.72	8.95	0.45	[illegible]	17.79	-0.36	-0.58	-0.71	-0.34	-1.2	-0.98	-0.92
20	130	efad	1	1.30	27.00	5.72	0.81	[illegible]	33.50	-0.35	-0.62	-0.56	-0.24	[illegible]	-0.88	-0.95
21	135	ref	1	0.38	24.09	8.22	0.60	[illegible]	24.61	-0.44	-0.75	-0.7	-0.33	[illegible]	-1.04	-0.99
22	281	efad	2	3.24	23.59	2.68	1.11	13.09	35.61	-0.48	-0.71	-0.63	-0.55	-1.27	-1.03	-0.88
23	285	ref	2	0.60	19.95	3.18	1.21	4.89	35.91	-0.5	-0.69	-0.86	-0.58	-1.31	-1.06	-0.88
24	290	tournesol	2	0.38	17.64	6.99	0.74	2.58	21.23	-0.54	-0.69	-1.02	-0.51	-1.32	-1.2	-0.96
25	294	dha	2	0.44	22.73	4.71	0.75	2.27	25.10	-0.39	-0.7	-0.83	-0.47	-1.27	-1.09	-0.91
26	295	tournesol	2	0.47	14.65	4.29	0.66	2.88	23.15	-0.54	-0.63	-0.86	-0.56	-1.35	-1.17	-0.88
27	303	ref	2	0.64	20.49	2.71	1.09	4.05	38.32	-0.35	-0.52	-0.74	-0.4	-1.07	-0.83	-0.73
28	305	ref	2	0.52	18.44	5.21	0.87	3.37	32.04	-0.46	-0.69	-0.53	-0.51	-1.33	-0.99	-0.9
29	307	lin	2	0.49	17.72	6.02	0.58	3.77	21.04	-0.42	-0.56	-0.86	-0.57	-1.23	-1.04	-0.75
30	319	dha	2	0.40	21.70	6.12	0.78	2.07	22.49	-0.51	-0.7	-0.68	-0.46	-1.34	-1.13	-0.94
31	321	lin	2	0.61	16.25	4.55	0.61	5.42	21.43	-0.47	-0.74	-0.88	-0.58	-1.3	-1.03	-0.83
32	325	efad	2	3.19	22.91	3.60	0.99	13.90	33.55	-0.54	-0.64	-0.39	-0.46	-1.3	-1.03	-0.85

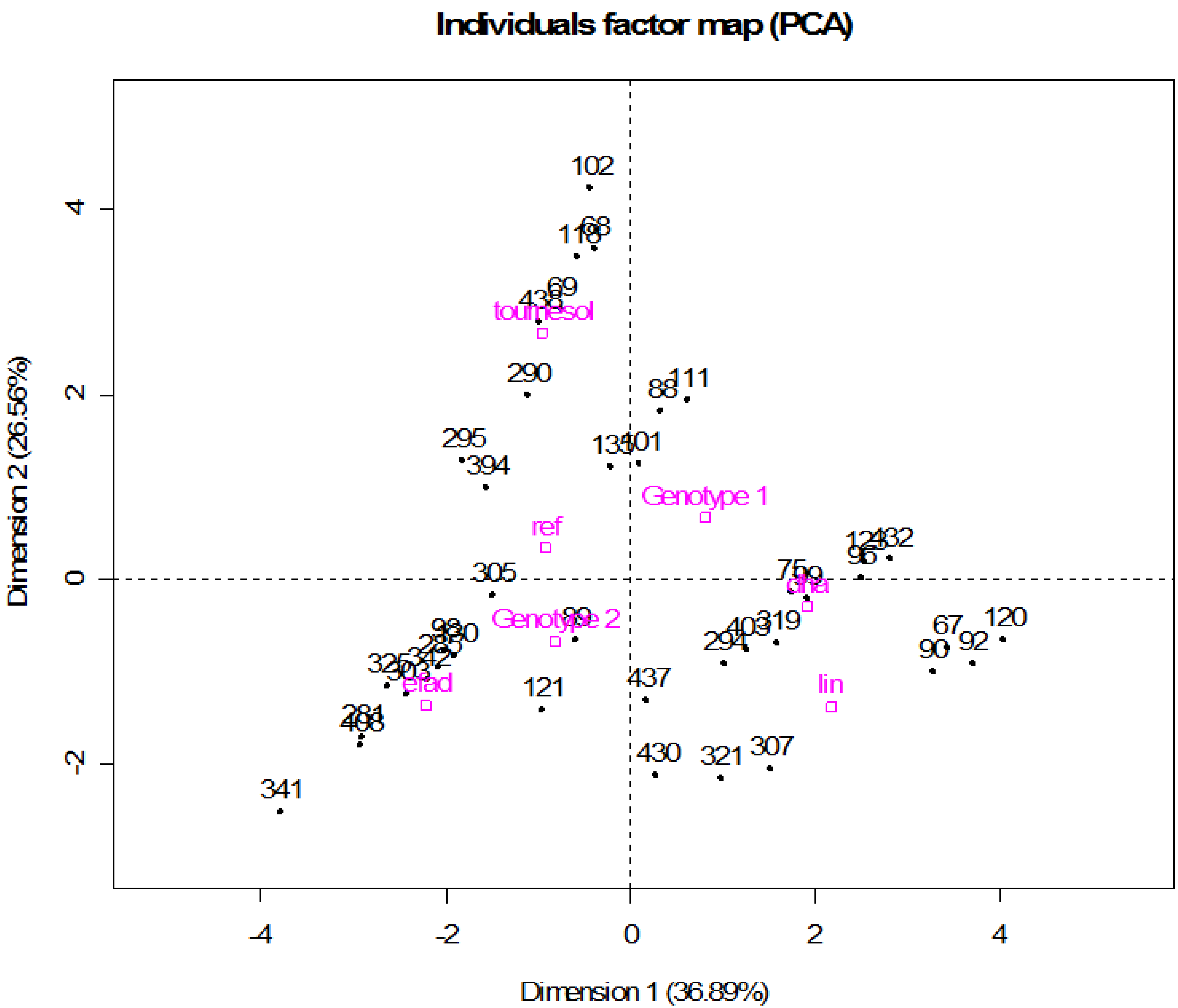
souris

X_1

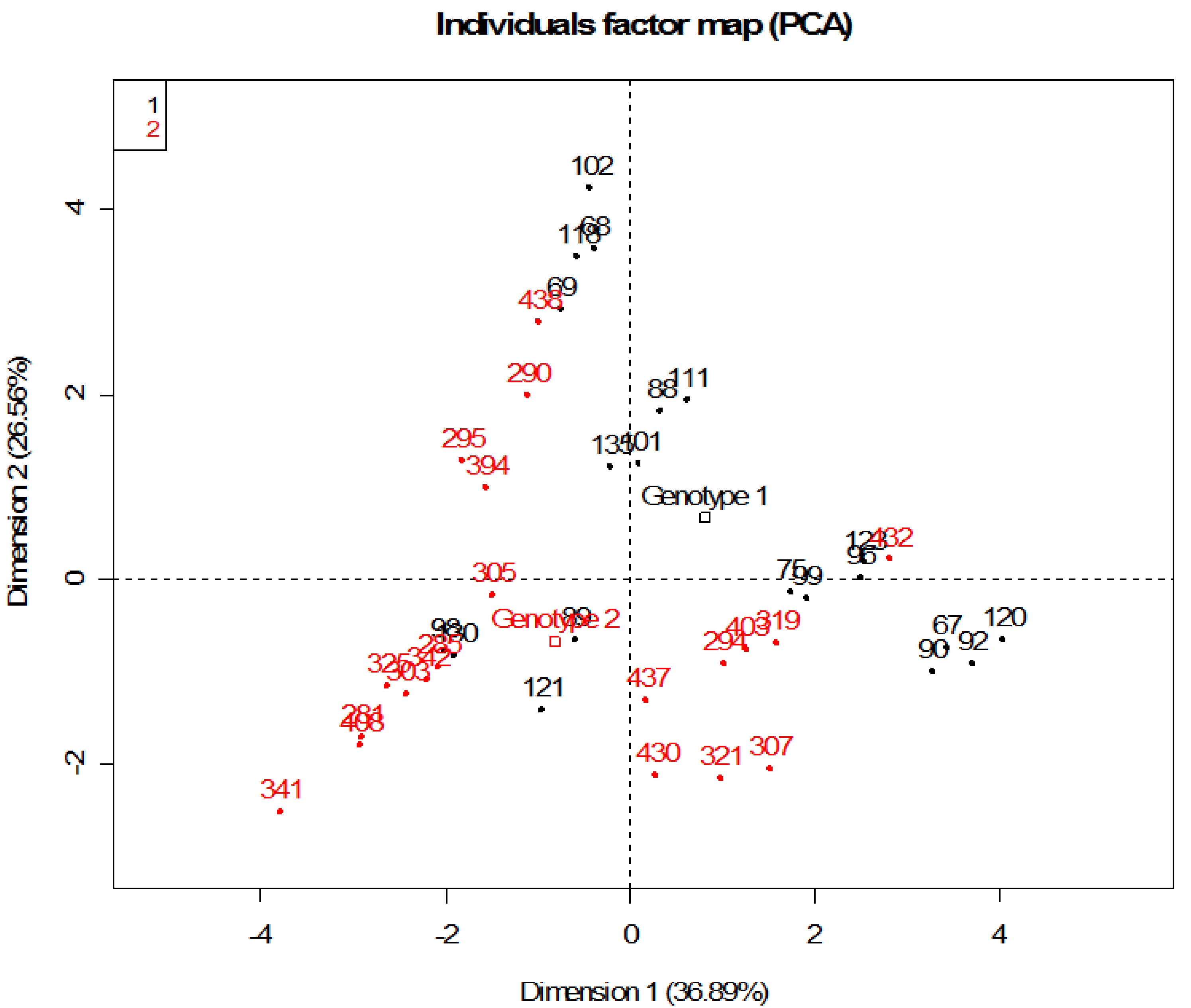
X_2

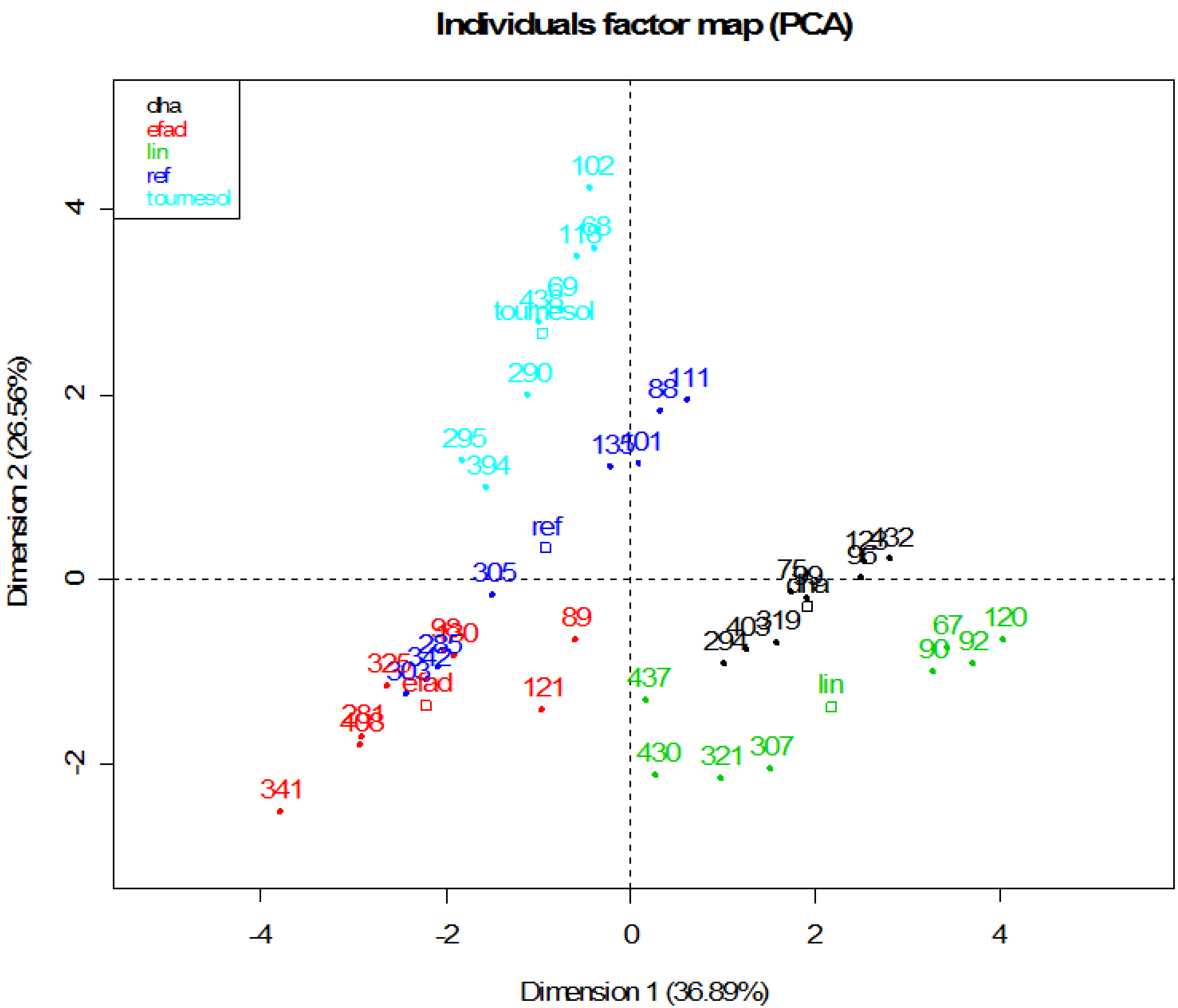


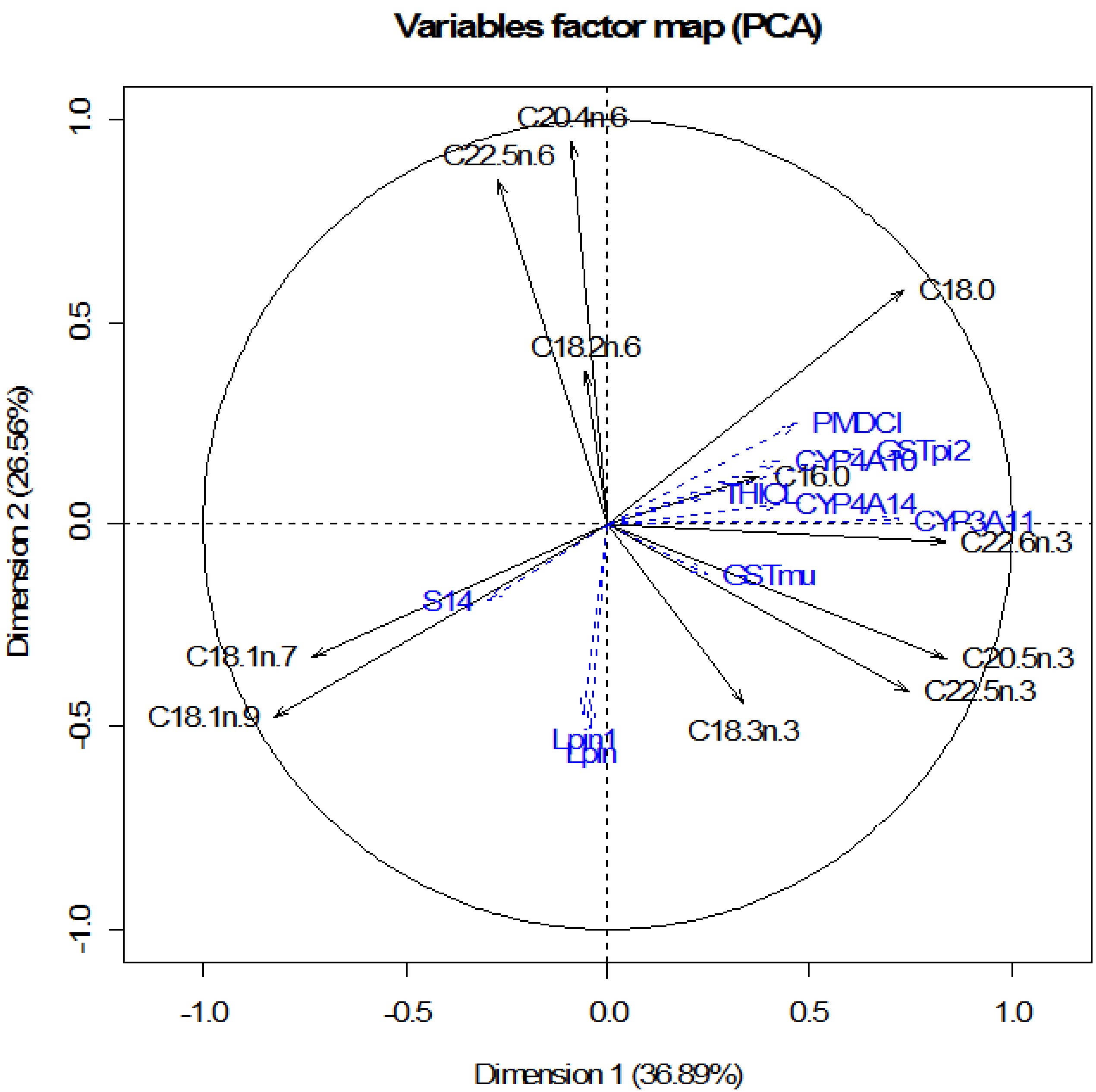
Individuals factor map (PCA)
Dimension 1 (36.89%)
Dimension 2 (26.56%)
-4
-2
0
2
4
102
68
118
69
438
290
111
88
295
135
101
394
124
432
96
305
75
99
89
98
380
285
342
325
303
403
319
294
67
120
92
90
121
437
281
408
430
321
307
341



Individuals factor map (PCA)
Dimension 1 (36.89%)
Dimension 2 (26.56%)
tournesol
ref
Genotype 1
Genotype 2
efad
dha
lin


Individuals factor map (PCA)
1
2
102
118
68
69
438
290
88
111
295
394
135
101
Genotype 1
124
432
96
75
99
305
98
130
Genotype 2
89
403
319
294
67
120
92
90
285
342
325
303
121
437
281
408
430
321
307
341
Dimension 1 (36.89%)
Dimension 2 (26.56%)
-4
-2
0
2
4


Individuals factor map (PCA)
dha
efad
lin
ref
tournesol
Dimension 1 (36.89%)
Dimension 2 (26.56%)
-4
-2
0
2
4
102
118
68
438
69
tournesol
290
295
394
88
111
135
101
ref
305
124
132
96
75
dha
403
319
294
67
120
90
92
437
lin
430
321
307
89
98
130
235
325
342
303
efad
121
281
408
341



Variables factor map (PCA)
C20.4n.6
C22.5n.6
C18.0
C18.2n.6
PMDCI
GSTpi2
CYP4A10
C16.0
THIOL
CYP4A14
CYP3A11
C22.6n.3
GSTmu
S14
C18.1n.7
C20.5n.3
C22.5n.3
C18.1n.9
C18.3n.3
Lpin1
Lpin
Dimension 1 (36.89%)
Dimension 2 (26.56%)
-1.0
-0.5
0.0
0.5
1.0

The dataset(s)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Souris	Regime	Genotype	C14.0	C16.0	C18.0	C16.1n.9	C16.1n.7	C18.1n.9	X36b4	ACAT1	ACAT2	ACBP	ACC1	ACC2	ACO
2	67	lin	1	0.34	26.45	10.22	0.35	3.10	16.98	-0.42	-0.65	-0.84	-0.34	-1.29	-1.13	-0.93
3	68	tournesol	1	0.38	24.04	9.93	0.55	2.54	20.07	-0.44	-0.68	-0.91	-0.32	-1.23	-1.06	-0.99
4	69	tournesol	1	0.36	23.70	8.96	0.55	2.65	22.89	-0.48	-0.74	-1.1	-0.46	-1.3	-1.09	-1.06
5	75	dha	1	0.22	25.48	8.14	0.49	2.82	21.92	-0.45	-0.69	-0.65	-0.41	-1.26	-1.09	-0.93
6	88	ref	1	0.37	24.80	9.63	0.46	2.85	21.38	-0.42	-0.71	-0.54	-0.38	-1.21	-0.89	
7	89	efad	1	1.70	26.04	6.59	0.66	7.26	28.23	-0.43	-0.69	-0.8	-0.32	-1.13	-0.79	-0.93
8	90	lin	1	0.35	25.94	9.68	0.36	3.60	17.62	-0.53	-0.62	-1	-0.44	-1.22	-1	-0.94
9	92	lin	1	0.34	28.63	9.95	0.29	3.27	17.02	-0.49	-0.69	-0.91	-0.37	-1.29	-1.06	-1.05
10	96	dha	1	0.22	25.34	8.81	0.44	2.36	18.39	-0.36	-0.66	-0.74	-0.39	-1.15	-1.08	-0.88
11	98	efad	1	1.38	28.49	5.63	0.90	7.01	36.68	-0.5	-0.62	-0.79	-0.36	-1.21	-0.82	-0.92
12	99	dha	1	0.26	25.73	8.30	0.43	2.74	21.75	-0.4	-0.6	-0.55	-0.25	-1.22	-1.13	-0.81
13	101	ref	1	0.44	24.28	8.63	0.53	3.33	23.86	-0.52	-0.66	-0.66	-0.41	-1.28	-1.1	-0.95
14	102	tournesol	1	0.32	24.63	9.99	0.45	2.39	17.93	-0.52	-0.63	-0.99	-0.43	-1.24	-0.96	-0.96
15	111	ref	1	0.34	26.04	9.81	0.35	2.36	20.14	-0.47	-0.71	-0.44	-0.45	-1.44	-1.17	-1.02
16	118	tournesol	1	0.35	24.76	9.38	[illegible]	2.47	19.66	-0.42	-0.66	-0.88	[illegible]	-1.24	-1.03	-0.92
17	120	lin	1	0.24	26.46	10.97	[illegible]	2.81	14.69	-0.58	-0.62	-0.85	[illegible]	-1.33	-1.19	-0.91
18	121	efad	1	1.21	23.45	5.59	[illegible]	[illegible]	33.84	-0.46	-0.69	-0.45	[illegible]	[illegible]	-0.93	-1.06
19	123	dha	1	0.30	29.72	8.95	0.45	[illegible]	17.79	-0.36	-0.58	-0.71	-0.34	-1.2	-0.98	-0.92
20	130	efad	1	1.30	27.00	5.72	0.81	[illegible]	33.50	-0.35	-0.62	-0.56	-0.24	[illegible]	-0.88	-0.95
21	135	ref	1	0.38	24.09	8.22	0.60	[illegible]	24.61	-0.44	-0.75	-0.7	-0.33	[illegible]	-1.04	-0.99
22	281	efad	2	3.24	23.59	2.68	1.11	13.09	35.61	-0.48	-0.71	-0.63	-0.55	-1.27	-1.03	-0.88
23	285	ref	2	0.60	19.95	3.18	1.21	4.89	35.91	-0.5	-0.69	-0.86	-0.58	-1.31	-1.06	-0.88
24	290	tournesol	2	0.38	17.64	6.99	0.74	2.58	21.23	-0.54	-0.69	-1.02	-0.51	-1.32	-1.2	-0.96
25	294	dha	2	0.44	22.73	4.71	0.75	2.27	25.10	-0.39	-0.7	-0.83	-0.47	-1.27	-1.09	-0.91
26	295	tournesol	2	0.47	14.65	4.29	0.66	2.88	23.15	-0.54	-0.63	-0.86	-0.56	-1.35	-1.17	-0.88
27	303	ref	2	0.64	20.49	2.71	1.09	4.05	38.32	-0.35	-0.52	-0.74	-0.4	-1.07	-0.83	-0.73
28	305	ref	2	0.52	18.44	5.21	0.87	3.37	32.04	-0.46	-0.69	-0.53	-0.51	-1.33	-0.99	-0.9
29	307	lin	2	0.49	17.72	6.02	0.58	3.77	21.04	-0.42	-0.56	-0.86	-0.57	-1.23	-1.04	-0.75
30	319	dha	2	0.40	21.70	6.12	0.78	2.07	22.49	-0.51	-0.7	-0.68	-0.46	-1.34	-1.13	-0.94
31	321	lin	2	0.61	16.25	4.55	0.61	5.42	21.43	-0.47	-0.74	-0.88	-0.58	-1.3	-1.03	-0.83
32	325	efad	2	3.19	22.91	3.60	0.99	13.90	33.55	-0.54	-0.64	-0.39	-0.46	-1.3	-1.03	-0.85

X_1

X_2

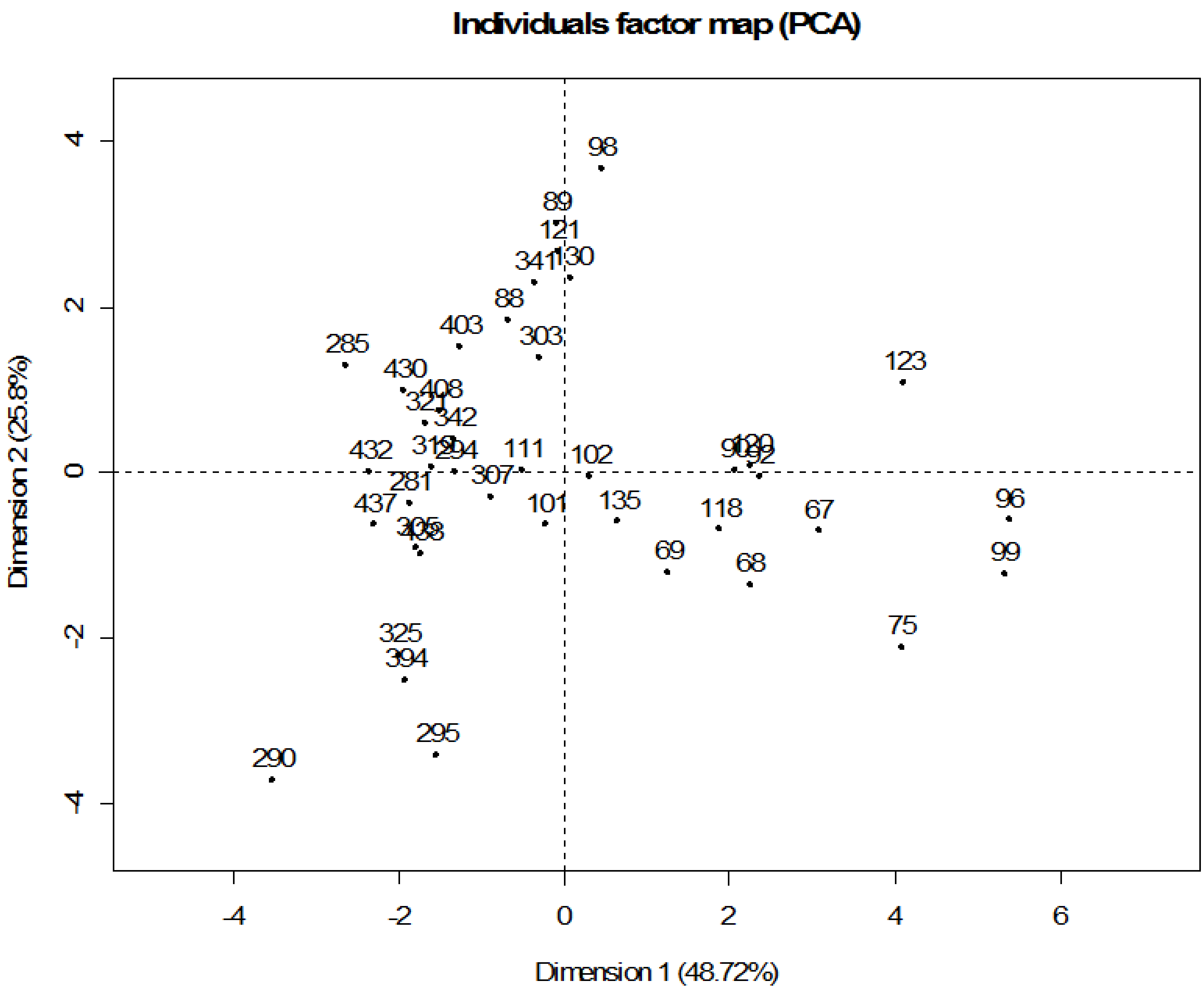
souris

The dataset(s)

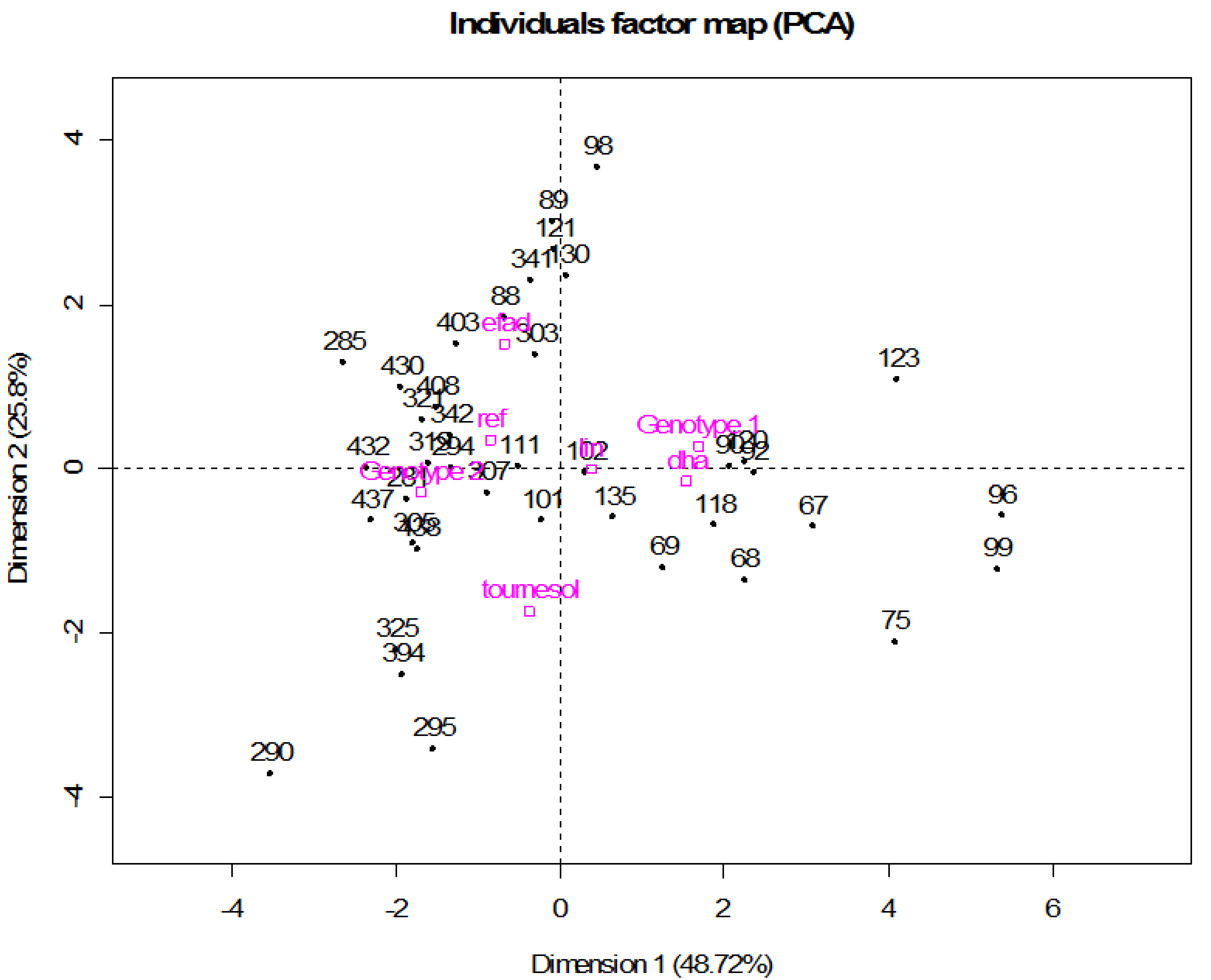
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Souris	Regime	Genotype	C14.0	C16.0	C18.0	C16.1n.9	C16.1n.7	C18.1n.9	X36b4	ACAT1	ACAT2	ACBP	ACC1	ACC2	ACO
2	67	lin	1	0.34	26.45	10.22	0.35	3.10	16.98	-0.42	-0.65	-0.84	-0.34	-1.29	-1.13	-0.93
3	68	tournesol	1	0.38	24.04	9.93	0.55	2.54	20.07	-0.44	-0.68	-0.91	-0.32	-1.23	-1.06	-0.99
4	69	tournesol	1	0.36	23.70	8.96	0.55	2.65	22.89	-0.48	-0.74	-1.1	-0.46	-1.3	-1.09	-1.06
5	75	dha	1	0.22	25.48	8.14	0.49	2.82	21.92	-0.45	-0.69	-0.65	-0.41	-1.26	-1.09	-0.93
6	88	ref	1	0.37	24.80	9.63	0.46	2.85	21.38	-0.42	-0.71	-0.54	-0.38	-1.21	-0.89	
7	89	efad	1	1.70	26.04	6.59	0.66	7.26	28.23	-0.43	-0.69	-0.8	-0.32	-1.13	-0.79	-0.93
8	90	lin	1	0.35	25.94	9.68	0.36	3.60	17.62	-0.53	-0.62	-1	-0.44	-1.22	-1	-0.94
9	92	lin	1	0.34	28.63	9.95	0.29	3.27	17.02	-0.49	-0.69	-0.91	-0.37	-1.29	-1.06	-1.05
10	96	dha	1	0.22	25.34	8.81	0.44	2.36	18.39	-0.36	-0.66	-0.74	-0.39	-1.15	-1.08	-0.88
11	98	efad	1	1.38	28.49	5.63	0.90	7.01	36.68	-0.5	-0.62	-0.79	-0.36	-1.21	-0.82	-0.92
12	99	dha	1	0.26	25.73	8.30	0.43	2.74	21.75	-0.4	-0.6	-0.55	-0.25	-1.22	-1.13	-0.81
13	101	ref	1	0.44	24.28	8.63	0.53	3.33	23.86	-0.52	-0.66	-0.66	-0.41	-1.28	-1.1	-0.95
14	102	tournesol	1	0.32	24.63	9.99	0.45	2.39	17.93	-0.52	-0.63	-0.99	-0.43	-1.24	-0.96	-0.96
15	111	ref	1	0.34	26.04	9.81	0.35	2.36	20.14	-0.47	-0.71	-0.44	[illegible]	-1.44	-1.17	-1.02
16	118	tournesol	1	0.35	24.76	9.38	[illegible]	2.47	19.66	-0.42	-0.66	-0.88	[illegible]	-1.24	-1.03	-0.92
17	120	lin	1	0.24	26.46	10.97	[illegible]	2.81	14.69	-0.58	-0.62	-0.85	[illegible]	-1.33	-1.19	-0.91
18	121	efad	1	1.21	23.45	5.59	[illegible]	[illegible]	33.84	-0.46	-0.69	-0.45	[illegible]	[illegible]	-0.93	-1.06
19	123	dha	1	0.30	29.72	8.95	0.45	[illegible]	17.79	-0.36	-0.58	-0.71	-0.34	-1.2	-0.98	-0.92
20	130	efad	1	1.30	27.00	5.72	0.81	[illegible]	33.50	-0.35	-0.62	-0.56	-0.24	[illegible]	-0.88	-0.95
21	135	ref	1	0.38	24.09	8.22	0.60	[illegible]	24.61	-0.44	-0.75	-0.7	-0.33	[illegible]	-1.04	-0.99
22	281	efad	2	3.24	23.59	2.68	1.11	13.09	35.61	-0.48	-0.71	-0.63	-0.55	-1.27	-1.03	-0.88
23	285	ref	2	0.60	19.95	3.18	1.21	4.89	35.91	-0.5	-0.69	-0.86	-0.58	-1.31	-1.06	-0.88
24	290	tournesol	2	0.38	17.64	6.99	0.74	2.58	21.23	-0.54	-0.69	-1.02	-0.51	-1.32	-1.2	-0.96
25	294	dha	2	0.44	22.73	4.71	0.75	2.27	25.10	-0.39	-0.7	-0.83	-0.47	-1.27	-1.09	-0.91
26	295	tournesol	2	0.47	14.65	4.29	0.66	2.88	23.15	-0.54	-0.63	-0.86	-0.56	-1.35	-1.17	-0.88
27	303	ref	2	0.64	20.49	2.71	1.09	4.05	38.32	-0.35	-0.52	-0.74	-0.4	-1.07	-0.83	-0.73
28	305	ref	2	0.52	18.44	5.21	0.87	3.37	32.04	-0.46	-0.69	-0.53	-0.51	-1.33	-0.99	-0.9
29	307	lin	2	0.49	17.72	6.02	0.58	3.77	21.04	-0.42	-0.56	-0.86	-0.57	-1.23	-1.04	-0.75
30	319	dha	2	0.40	21.70	6.12	0.78	2.07	22.49	-0.51	-0.7	-0.68	-0.46	-1.34	-1.13	-0.94
31	321	lin	2	0.61	16.25	4.55	0.61	5.42	21.43	-0.47	-0.74	-0.88	-0.58	-1.3	-1.03	-0.83
32	325	efad	2	3.19	22.91	3.60	0.99	13.90	33.55	-0.54	-0.64	-0.39	-0.46	-1.3	-1.03	-0.85

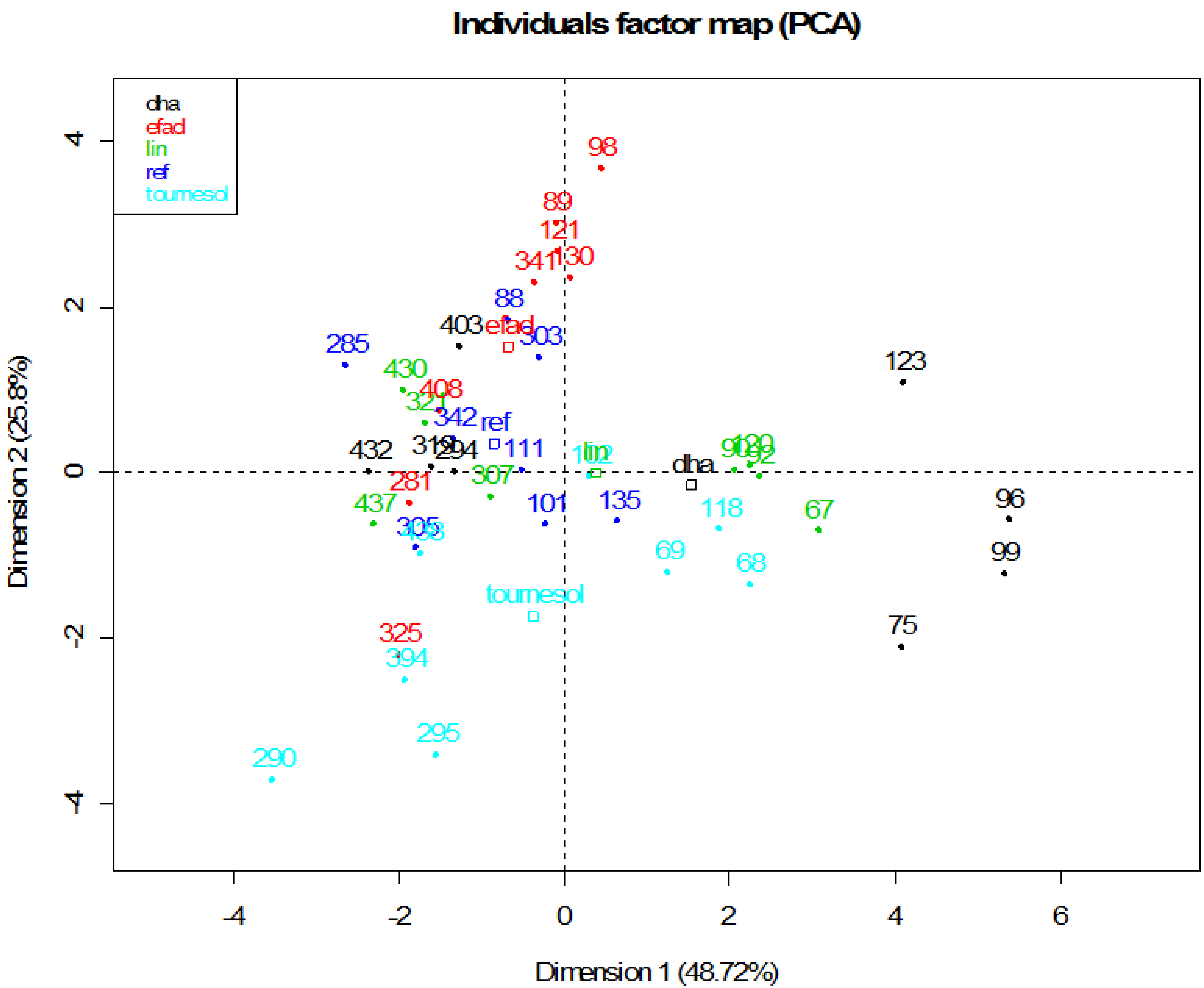
X_1 X_2

souris

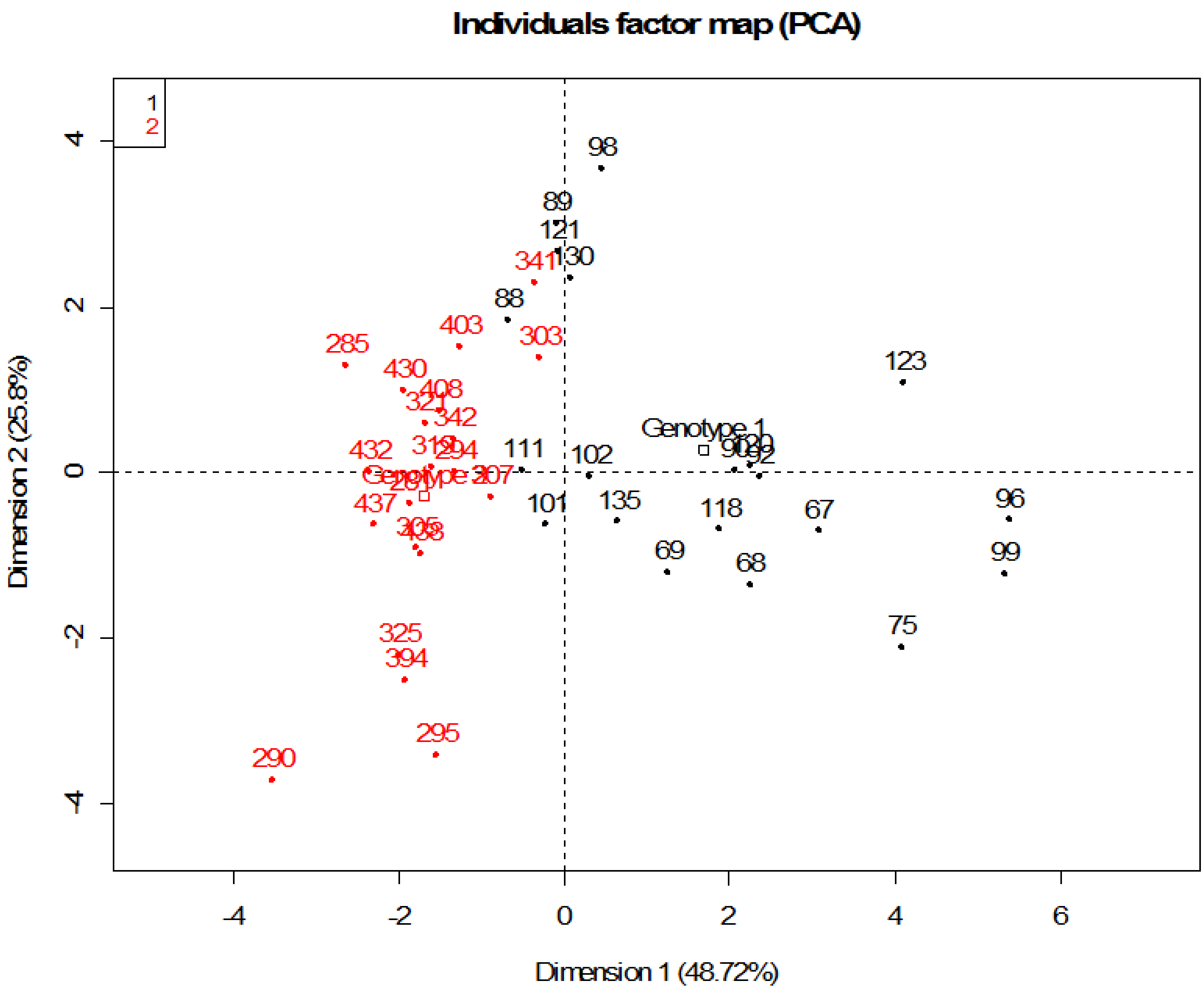


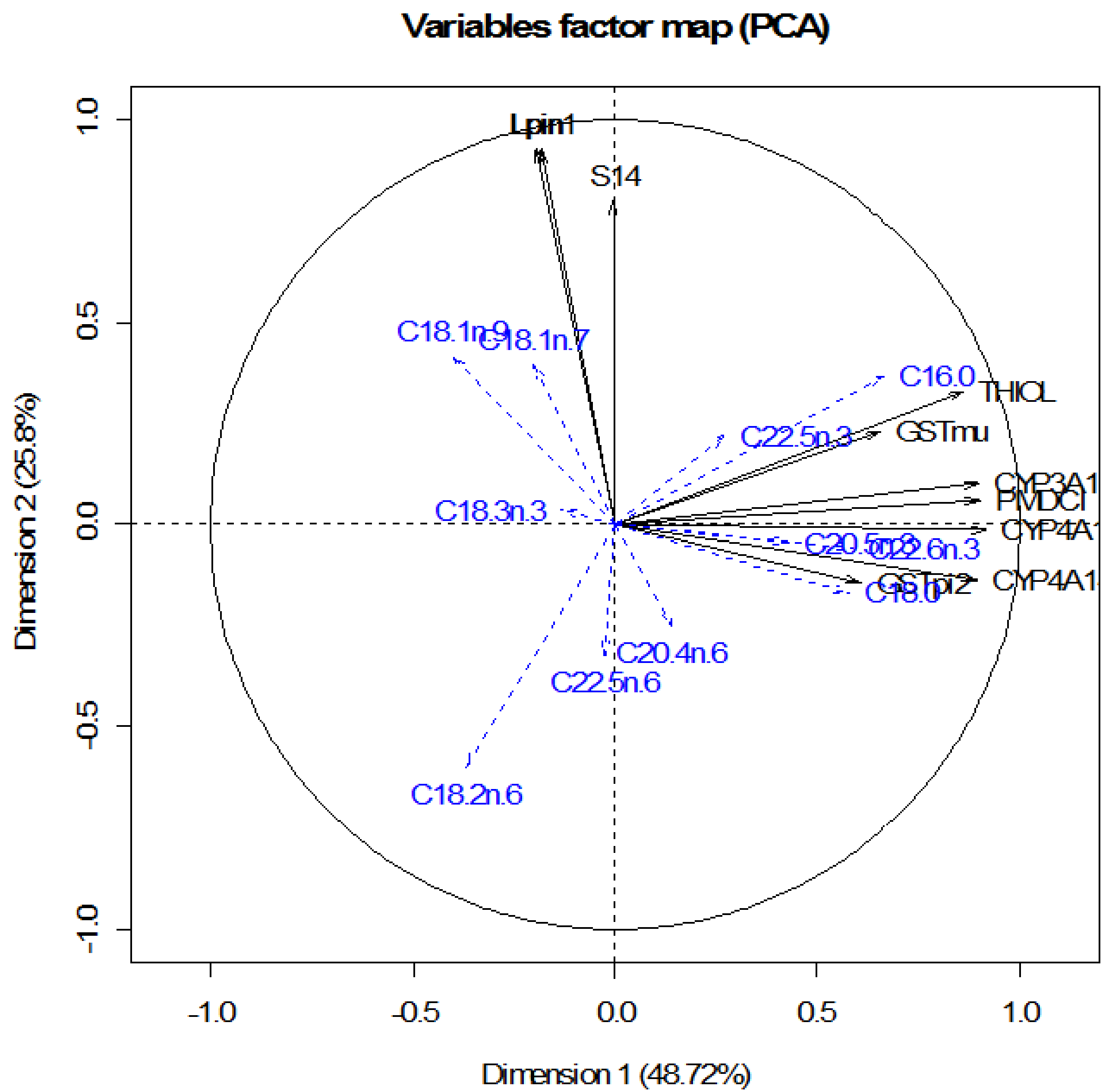
Individuals factor map (PCA)
Dimension 1 (48.72%)
Dimension 2 (25.8%)
-4
-2
0
2
4
6
98
89
121
341
130
88
403
303
285
430
408
321
342
123
432
319
294
111
102
90
120
92
281
307
437
101
135
118
67
96
305
438
69
68
99
325
394
75
295
290


Individuals factor map (PCA)
Dimension 1 (48.72%)
Dimension 2 (25.8%)
-4
-2
0
2
4
6
98
89
121
341
130
88
403
efad
303
285
430
408
342
ref
111
432
Genotype 2
Genotype 1
dha
123
437
101
135
118
67
96
69
68
99
tournesol
75
325
394
295
290

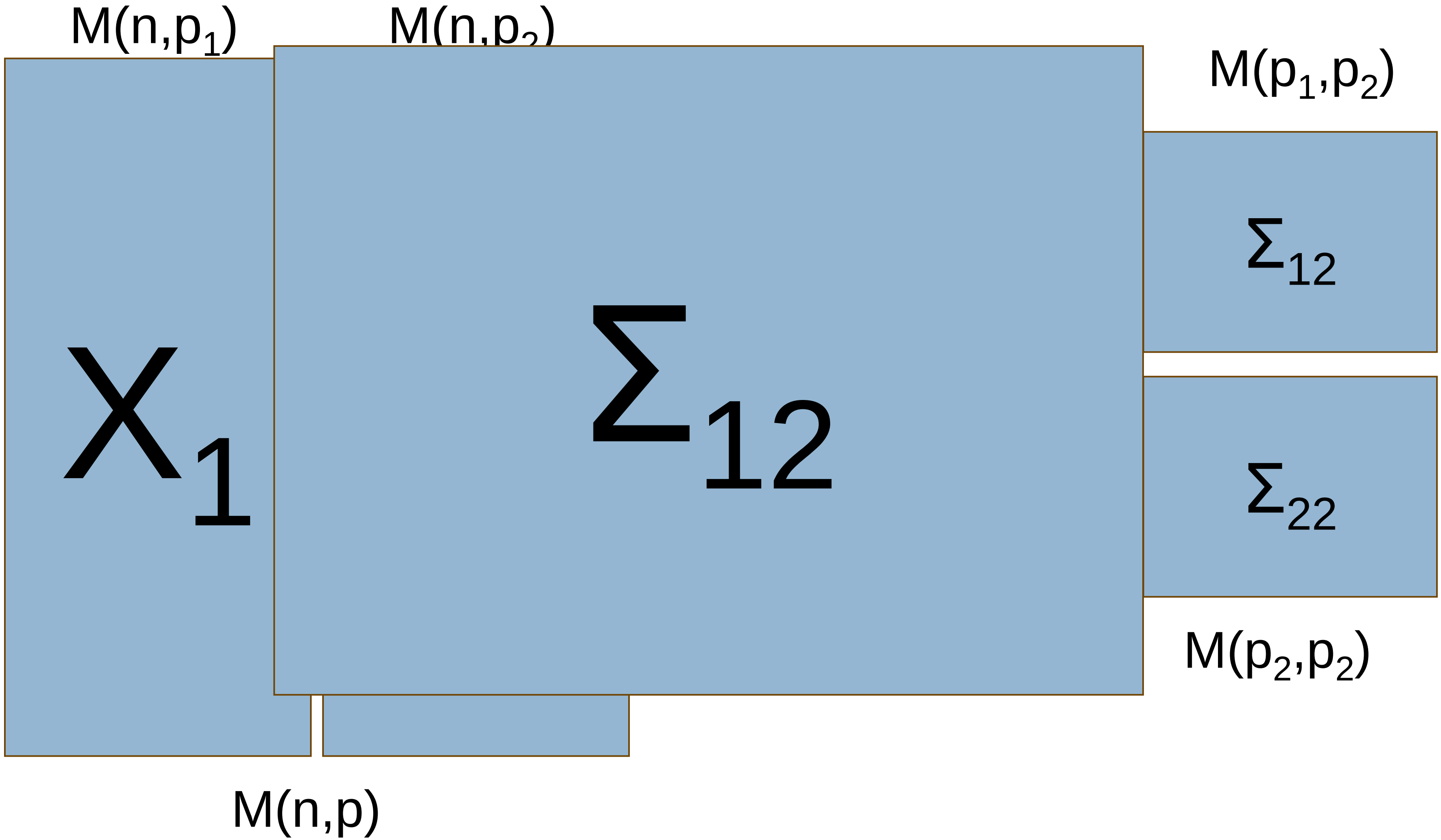


Individuals factor map (PCA)
dha
efad
lin
ref
tournesol
98
89
121
341
130
88
403
efad
303
285
123
430
408
321
342
ref
432
111
lin
dha
307
281
437
101
135
118
67
96
305
69
68
99
tournesol
325
75
394
295
290
Dimension 2 (25.8%)
Dimension 1 (48.72%)
-4
-2
0
2
4
6

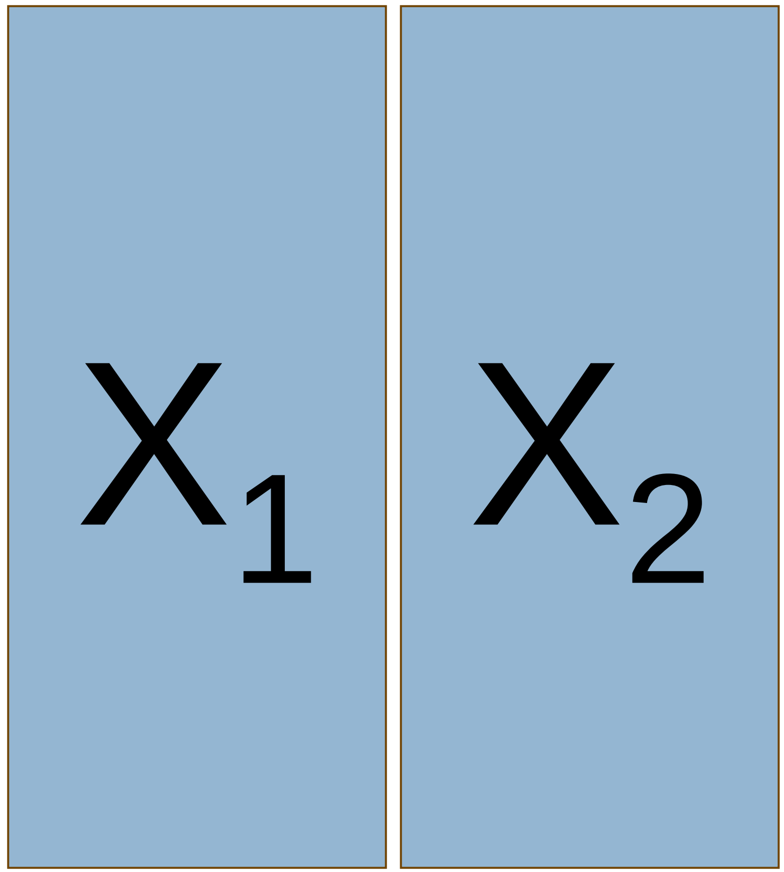

Individuals factor map (PCA)
Dimension 1 (48.72%)
Dimension 2 (25.8%)
1
2
-4
-2
0
2
4
6
Genotype 1
Genotype 2
98
89
121
130
341
88
403
303
285
430
408
321
342
123
432
111
102
437
101
135
118
67
96
69
68
99
325
394
75
295
290


Variables factor map (PCA)
Lpin1
S14
C18.1n.9
C18.1n.7
C16.0
THIOL
C22.5n.3
GSTmu
CYP3A1
PMDCI
C18.3n.3
CYP4A1
C20.5n.3
C22.6n.3
CYP4A1
GSTpi2
C18.0
C20.4n.6
C22.5n.6
C18.2n.6
Dimension 2 (25.8%)
Dimension 1 (48.72%)
1.0
0.5
0.0
-0.5
-1.0

The dataset(s)



Close Encounters of the Third Kind



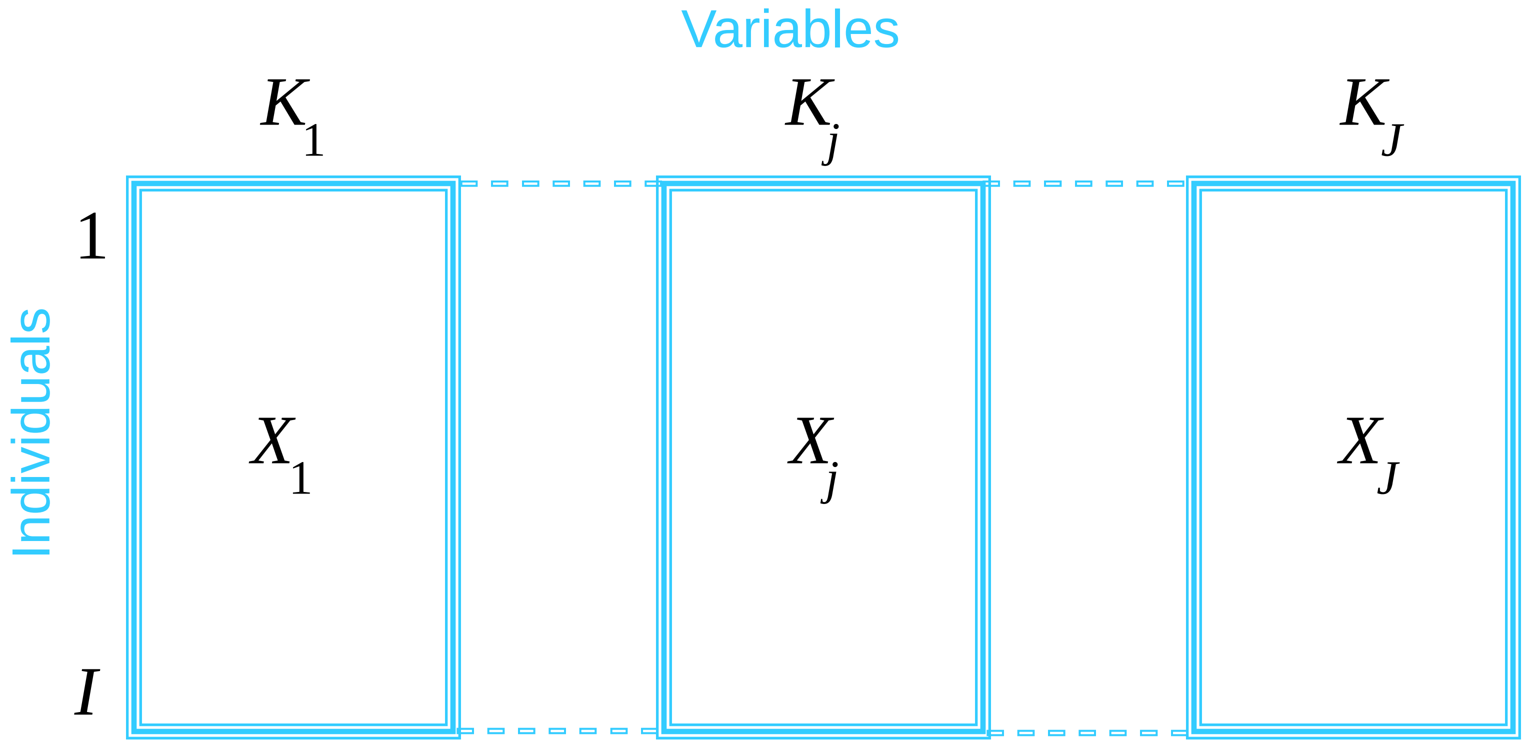
FROM MULTIVARIATE TO MULTIPLE TABLES DATA ANALYSIS

Rashomon (1950, Kurosawa)

Rashomon (1950, Kurosawa)

- The film depicts the rape of a woman and the apparent murder of her husband through the widely differing accounts of four witnesses, including the rapist and, through a medium, the dead man.
- The stories are mutually contradictory, leaving the viewer to determine which, if any, is the truth.

Objectives underlying the study of several groups of variables



Objectives underlying the study of several groups of variables

- Weighting of the variables
- Looking for common factors
- Comparison of factors
- Overall representation of groups
- Superimposed representation

Objectives underlying the study of several groups of variables

- Weighting of the variables
- Looking for common factors
- Comparison of factors
- Overall representation of groups
- Superimposed representation

To balance the influence of each group in a simultaneous analysis

Objectives underlying the study of several groups of variables

- Weighting of the variables
- Looking for common factors
- Comparison of factors
- Overall representation of groups
- Superimposed representation

To search for factors that are common to the group of variables

Objectives underlying the study of several groups of variables

- Weighting of the variables
- Looking for common factors
- Comparison of factors
- Overall representation of groups
- Superimposed representation

To compare the factors of several groups of variables

Objectives underlying the study of several groups of variables

- Weighting of the variables
- Looking for common factors
- Comparison of factors
- Overall representation of groups
- Superimposed representation

Two groups are all the more close that they induce the same structure

Objectives underlying the study of several groups of variables

- Weighting of the variables
- Looking for common factors
- Comparison of factors
- Overall representation of groups
- Superimposed representation

An individual is all the more "homogenous" that its superimposed representations are close

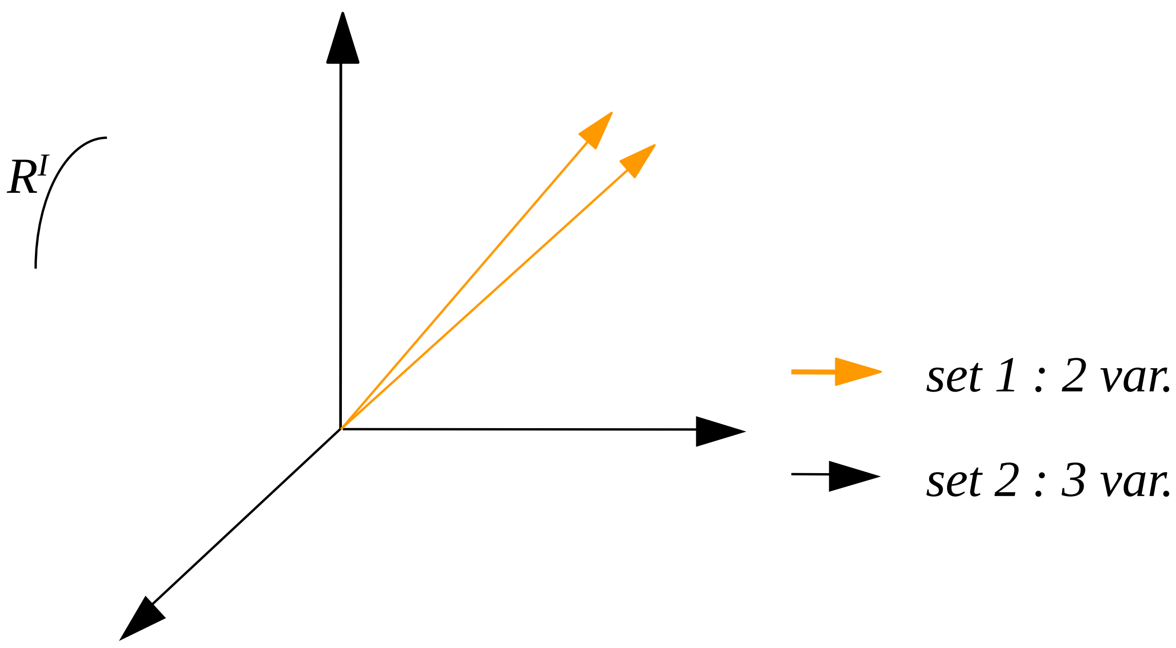
WEIGHTING OF THE VARIABLES

On the interest of balancing the influence of each group of variables

- By analogy with the individuals: weighting sample surveys, balanced data
- Same weight for each variable of a given group
  - Number of variables
  - Structure of each group

On the interest of balancing the influence of each group of variables

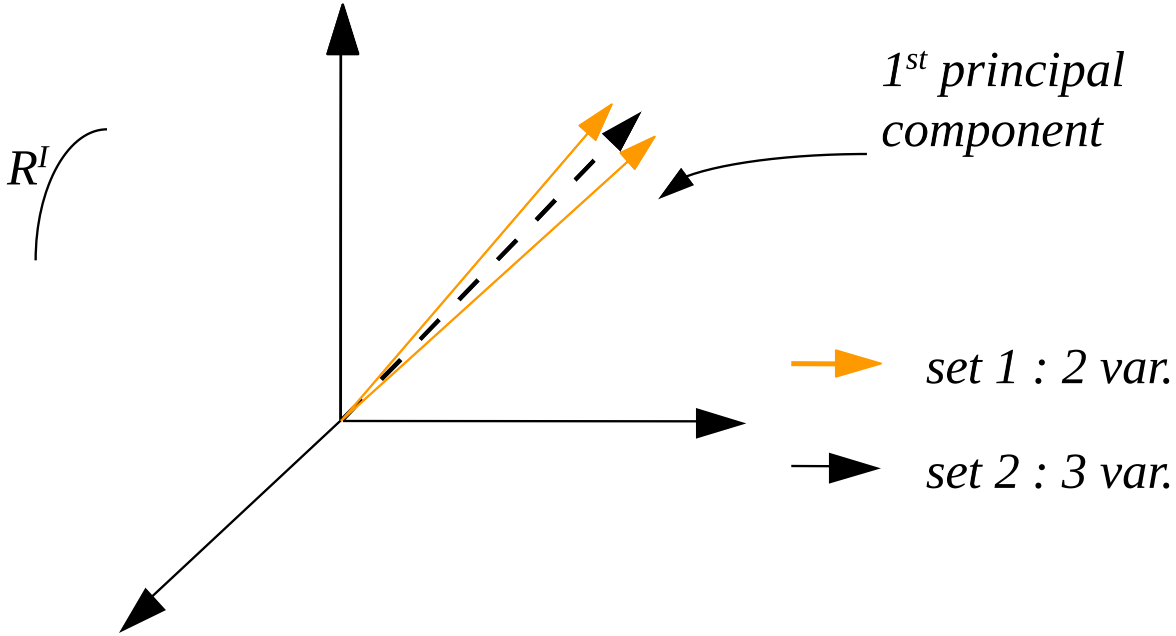
Reference example



On the interest of balancing the influence of each group of variables

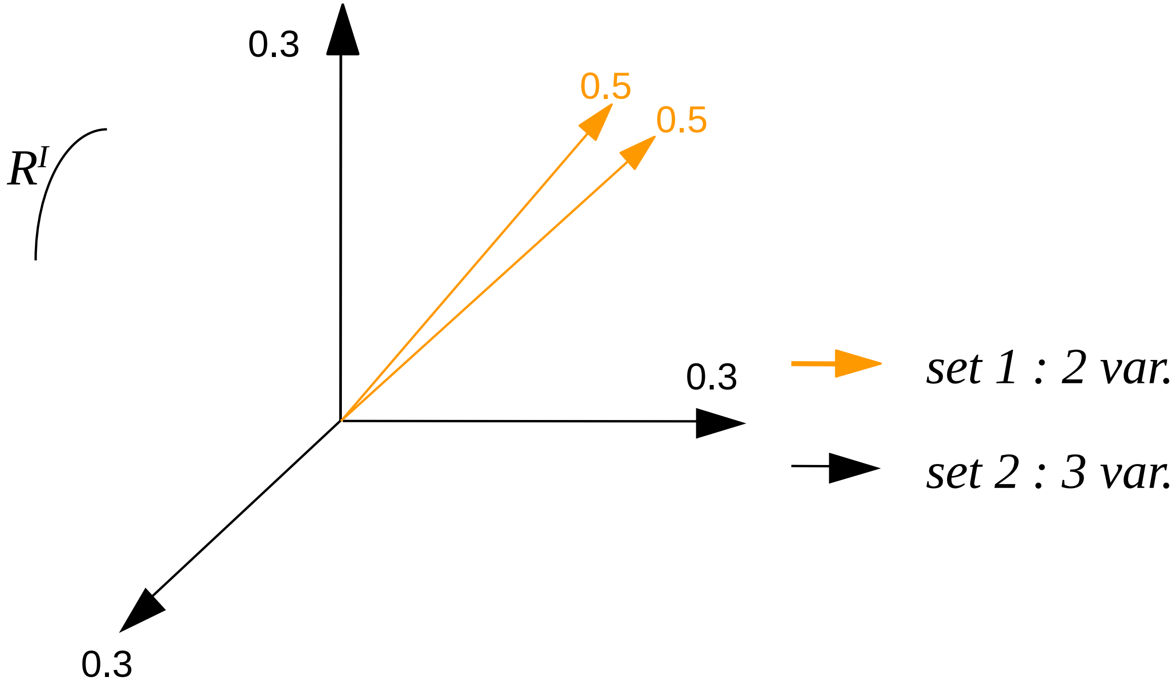
Reference example

PCA of the 5 variables, without considering the sets



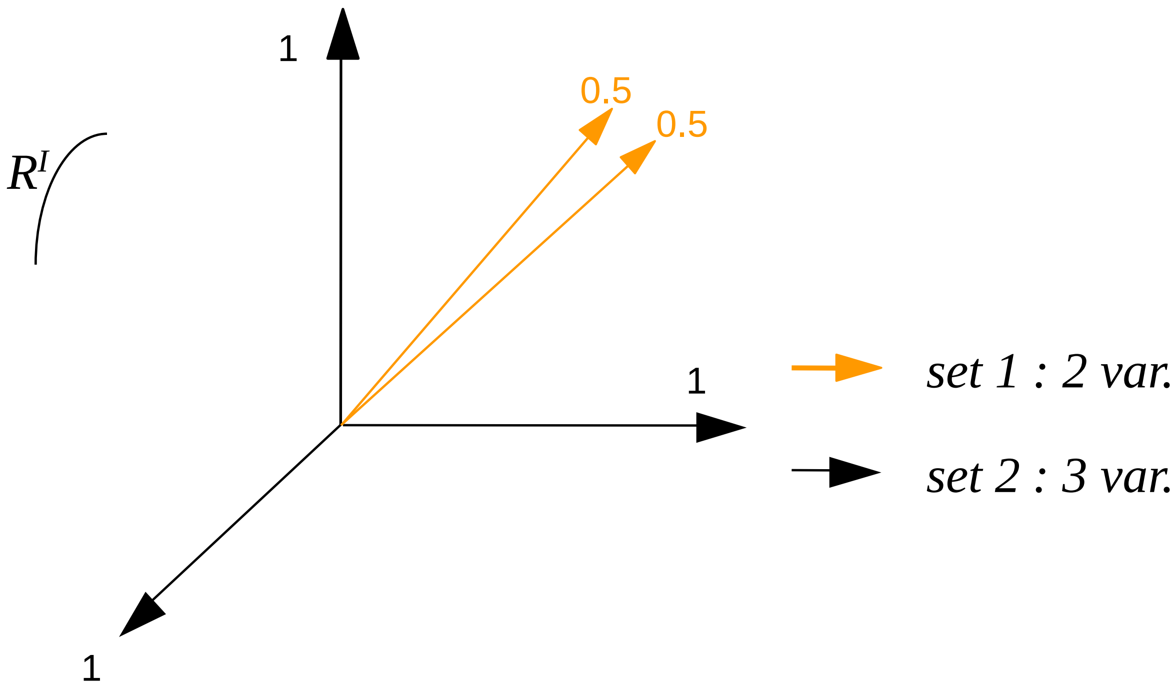
On the interest of balancing the influence of each group of variables

Reference example
Balancing the sets by the total “inertia”



On the interest of balancing the influence of each group of variables

Reference example
Balancing the sets of variables in MFA



Each variable of the set j is weighted by $1/\lambda_1^{\,j}$

$\lambda_1^{\,j}$: 1st eigenvalue of PCA applied to set j.

On the interest of balancing the influence of each group of variables

- For each group the variance of the main axis of variability is equal to 1
- No group can generate all by itself the first global axis
- A “multidimensional” group will contribute to the construction of more axes than a “one-dimensional” group
- This weighting is a specific characteristic of MFA; it induces many properties described later

Weighted factorial analysis

MFA is based on a "factorial analysis" applied to all active sets of variables

Weighted factorial analysis

MFA is based on a "factorial analysis" applied to all active sets of variables

De facto: MFA beneficiates from the transition formulae and from the duality between *individuals* and *variables*.

Weighted factorial analysis

MFA is based on a "factorial analysis" applied to all active sets of variables

Quantitative variables: MFA is based on a weighted PCA

standardized variables

unstandardized variables

mixed

Equivalence

When each set is composed by 1 quantitative variable: MFA=PCA

Weighted factorial analysis

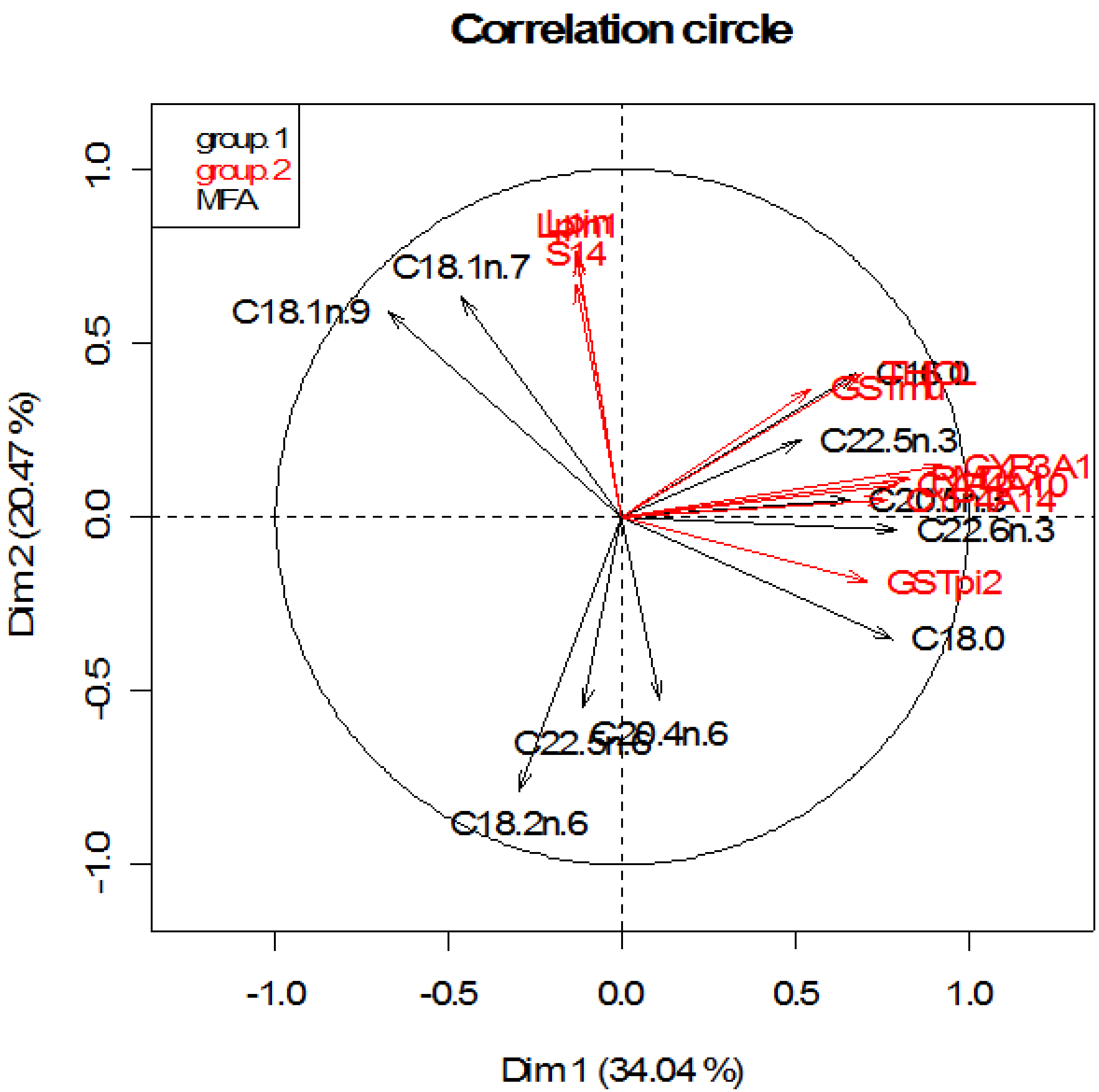
MFA is based on a "factorial analysis" applied to all active sets of variables

MFA provides:

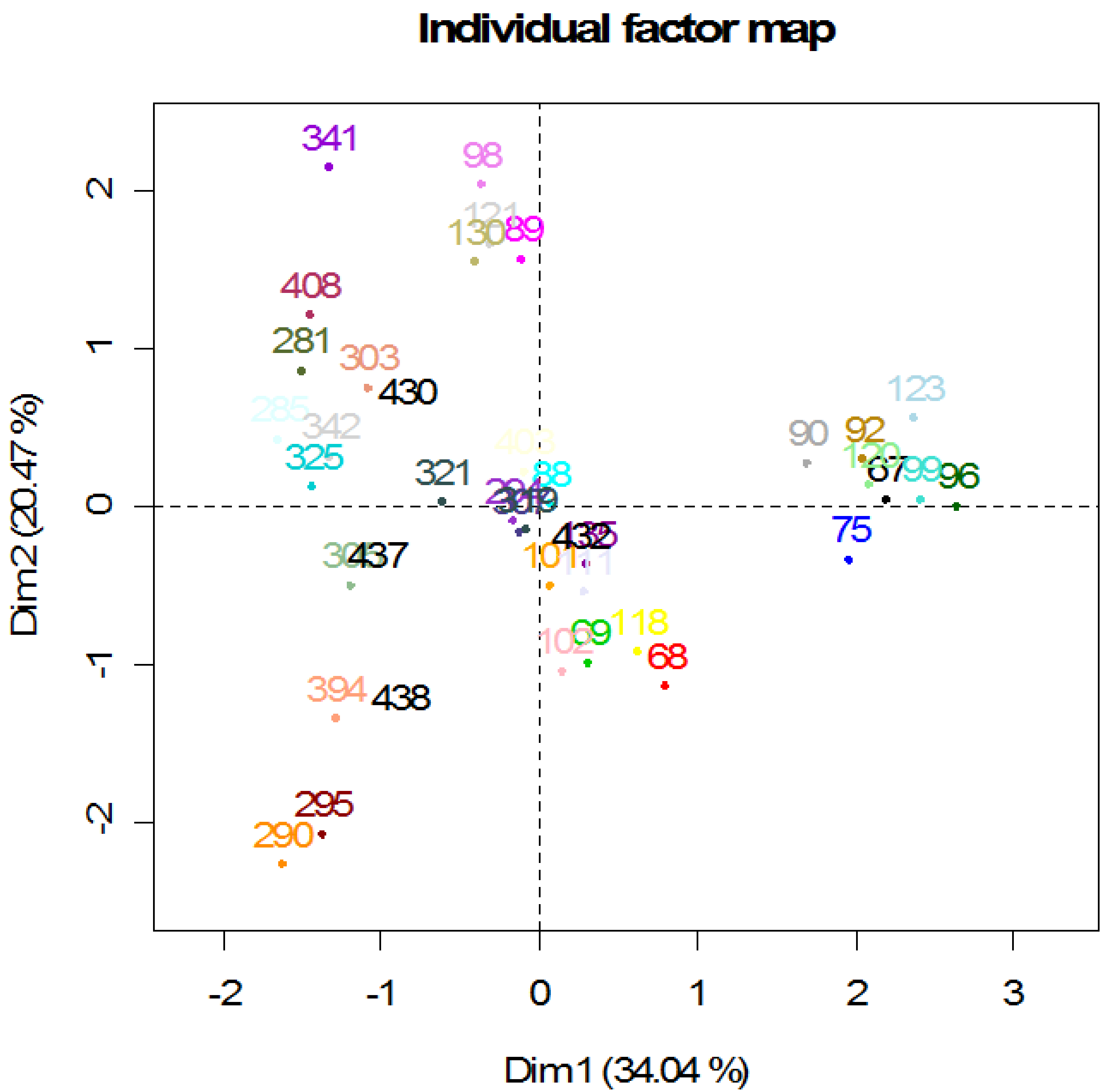
Firstly: classical results of factorial analysis

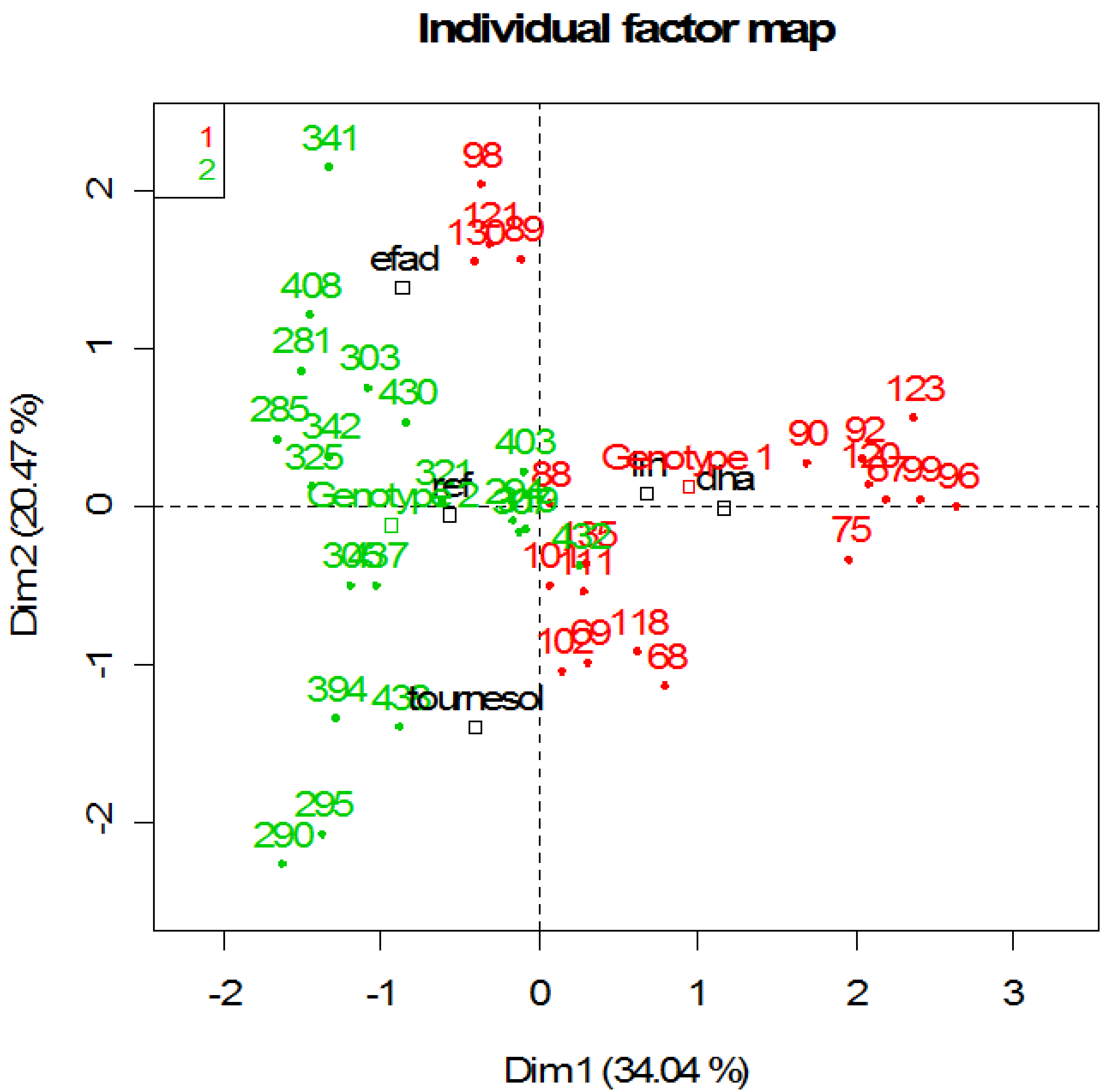
For each axis:

- Co-ordinates, contributions and squared cosines of individuals
- Correlation coefficients between factors and continuous variables

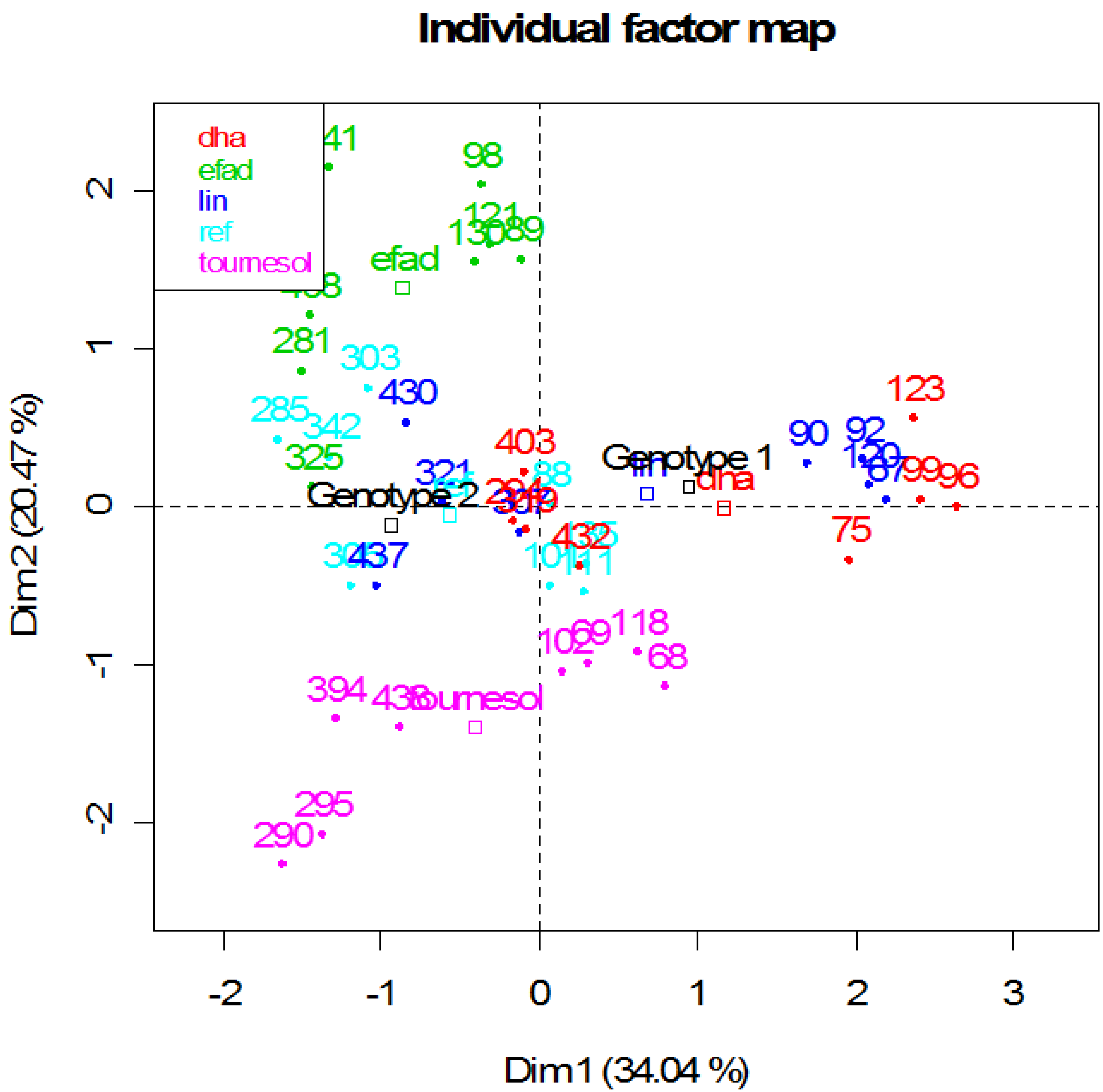


Correlation circle
group.1
group.2
MFA
C18.1n.7
C18.1n.9
S14
C22.5n.3
GSTpi2
C18.0
C22.6n.3
C18.2n.6
Dim 1 (34.04 %)
Dim 2 (20.47 %)


Individual factor map
Dim1 (34.04 %)
Dim2 (20.47 %)
-2
-1
0
1
2
3
341
98
121
130
89
408
281
303
430
285
342
325
321
88
90
92
123
67
99
96
75
437
101
118
102
68
394
438
295
290



Individual factor map
Dim1 (34.04 %)
Dim2 (20.47 %)
1
2
efad
dna
ref
toumesol
Genotype 1
341
98
408
281
303
430
285
342
325
403
88
123
90
92
75
118
68
394
295
290
-2
-1
0
1
2
3


Individual factor map
dha
efad
lin
ref
tournesol
Dim1 (34.04 %)
Dim2 (20.47 %)
-2
-1
0
1
2
3
41
98
121
130
89
efad
281
303
430
285
342
325
321
403
88
Genotype 1
lin
dha
90
92
123
99
96
75
Genotype 2
432
437
101
111
118
69
102
68
394
438
tournesol
295
290

SETTING UP COMMON FACTORS

Looking for factors common to TWO sets of variables

X_1

X_2

Reference method: Canonical Analysis

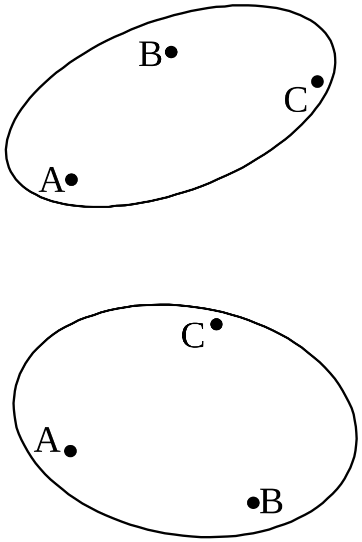
Hotelling, 1936

Looking for factors common to TWO sets of variables

- The word "**canonical**" comes from the Greek κανών / *kanôn* that means "ruler"
- The purpose of Canonical analysis is to find the relationship between two groups of variables
- It works by finding two linear combinations of variables, one for each group, which are most highly correlated
- Hotelling, H. (1936) Relations between two sets of variables. *Biometrika*, 28, 321-377

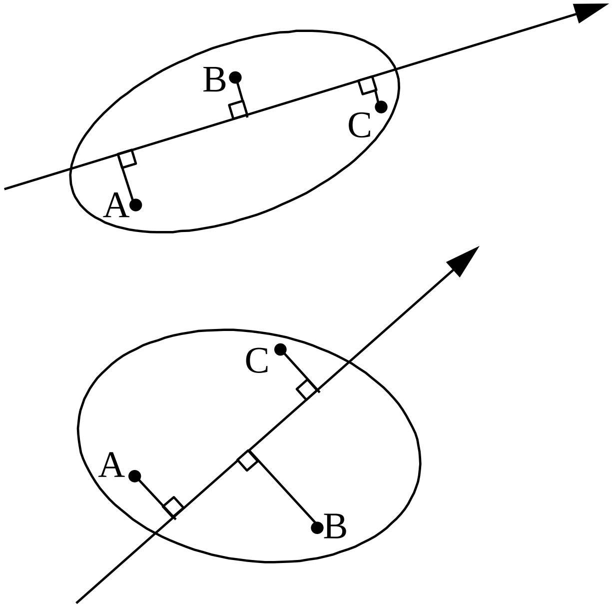
Looking for factors common to TWO sets of variables

A factor common to two clouds?



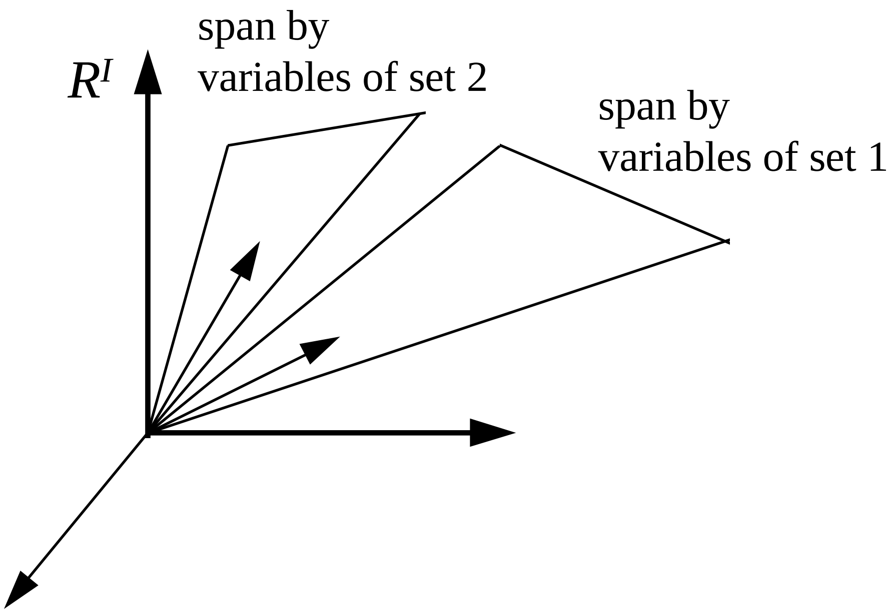
Looking for factors common to TWO sets of variables

A factor common to two clouds!



Looking for factors common to TWO sets of variables

Looking for jointly linear combinations of variables of sets 1 and 2



cancor(lip,gen)

R Console

Fichier Edition Misc Packages Aide

```
> cancor(lip,gen)
$cor
 [1] 0.96297273 0.93199442 0.91149643 0.85708380 0.78971594 0.71861893 0.60593887 0.40912187 0.24543555 0.04250940

$xcoef
                [,1]         [,2]         [,3]        [,4]         [,5]         [,6]         [,7]         [,8]        [,9]
C18.1n.9  0.028997123  0.0171655617 -0.032635011 -0.01450857 -0.018555194 -0.041183319 -0.013174087  0.040586602  0.08710232
C18.1n.7  0.032505732 -0.0                                                                                                   3
C18.2n.6  0.011143241 -0.0                                                                                                   0
C20.4n.6  0.016316003  0.0                                                                                                   9
C22.5n.6  0.013895923 -0.1994208277 -0.237792213  0.04828274 -0.294048565 -0.034069099 -0.278425025 -0.097912648 -0.01606464
C22.6n.3 -0.009625662  0.0180690151 -0.057893604 -0.01258168  0.010209536 -0.016871470 -0.040695824  0.018336510  0.07227701
C20.5n.3  0.001693435  0.0287109724 -0.052755724  0.11695136  0.028633355 -0.033946554  0.011660799 -0.006932460  0.11639351
C22.5n.3  0.080104739  0.0189714987 -0.078025488 -0.09597613  0.004368761  0.114483628  0.243505319  0.029838355  0.28063148
C16.0     0.025879989  0.0002305252 -0.014961055 -0.01606147 -0.087258413 -0.021124677  0.052003685  0.075732380  0.05434408
C18.0     0.053045462  0.0023552992  0.060777399 -0.15890538 -0.105767029 -0.003898920 -0.111955143  0.100838476  0.11642331
C18.3n.3  0.012796276  0.0                                                 .031400092 -0.011095415  0.033605112  0.09007832

$ycoef
              [,1]       [,2]       [,3]        [,4]        [,5]        [,6]        [,7]        [,8]       [,9]      [,10]
PMDCI    0.51640610  0.6523181  1.0410813 -1.17646805 -0.23554143  0.96947711 -0.52053547 -0.16848881  0.1407219 -0.5615899
THIOL    0.63929635  0.3485394 -0.5578365  1.21794331  1.00912511 -0.40821774  0.12353419 -0.25882855  0.2587056  0.0627558
CYP3A11 -0.60991186  0.0407698 -0.2250082  0.26683169 -0.40367530 -0.01861773  0.53348063 -0.90329962 -0.4644879  0.3434860
CYP4A10 -0.05180081 -0.3360889 -0.1978164  0.19342476 -0.91569517 -0.05456395 -0.49015959  1.02698256  0.7043031  1.4347729
Lpin     0.26195541  0.2501206 -0.6088044 -0.13511715 -1.10231985 -0.27382711 -0.09176426  0.13063574  3.9408262 -1.7372312
Lpin1   -0.27203283  0.335                                                                        113  2.1529151
GSTmu    0.24643161 -0.565                                                                        318  0.3644918
GSTpi2  -0.25550623  0.374                                                                        683 -0.6715756
S14      0.16336084 -0.3291935  0.1997579  0.06817508 -0.64894459 -0.01595339  0.56824837  0.05062431 -0.5990380 -0.6348794
CYP4A14 -0.30233566  0.1143903 -0.1328348 -0.33811760  0.30070528 -0.69227204 -0.18733677 -0.10210937 -0.5592082 -1.3047358

$xcenter
C18.1n.9 C18.1n.7 C18.2n.6 C20.4n.6 C22.5n.6 C22.6n.3 C20.5n.3 C22.5n.3    C16.0    C18.0 C18.3n.3
25.27325  4.42600 15.27750  5.27925  0.43700  5.91400  1.78950  0.87175 23.02600  6.74700  2.88800

$ycenter
   PMDCI    THIOL  CYP3A11  CYP4A10     Lpin    Lpin1    GSTmu   GSTpi2      S14  CYP4A14
-0.76725 -0.41100 -0.50825 -0.97975 -0.75325 -0.76475 -0.11900  0.22975 -0.80675 -0.99300
```

L_1 = 0.028c18.1.n-9+0.032c18.1.n-7+…+0.012c18.3.n-3

Beware: canonical variables

G_1 = 0.51PMDCI+0.63THIOL+…-0.30CYP4A14

cancor(lip,gen)

$G_1 = 0.51PMDCI + 0.63THIOL + \ldots + 0.26LPIN - 0.27LPIN1 + \ldots - 0.30CYP4A14$

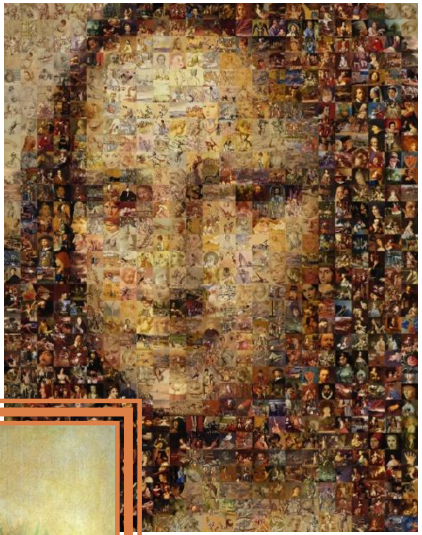
	PMDCI	THIOL	CYP3A11	CYP4A10	Lpin	Lpin1	GSTmu	GSTpi2	S14	CYP4A14
PMDCI	1.0000000	0.84318213	0.78800310	0.85333860	-0.15426020	-0.12132735	0.42444476	0.4422014	0.09117950	0.7509967
THIOL	0.8431821	1.00000000	0.73554530	0.79996380	0.07321187	0.09424437	0.56878570	0.3596621	0.33377479	0.6987517
CYP3A11	0.7880031	0.73554530	1.00000000	0.75508008	-0.01894254	-0.03396576	0.62005000	0.6041323	-0.03192741	0.7368760
CYP4A10	0.8533386	0.79996380	0.75508008	1.00000000	-0.22755111	-0.19530806	0.43563040	0.4159180	0.07638834	0.8853495
Lpin	-0.1542602	0.07321187	-0.01894254	-0.22755111	1.00000000	0.97166955	0.10963198	-0.1489248	0.57820562	-0.2886763
Lpin1	-0.1213273	0.09424437	-0.03396576	-0.19530806	0.97166955	1.00000000	0.06449885	-0.1234486	0.58591046	-0.2782659
GSTmu	0.4244448	0.56878570	0.62005000	0.43563040	0.10963198	0.06449885	1.00000000	0.4505130	0.08559840	0.5285176
GSTpi2	0.4422014	0.35966214	0.60413234	0.41591801	-0.14892482	-0.12344856	0.45051300	1.0000000	-0.27370381	0.4171278
S14	0.0911795	0.33377479	-0.03192741	0.07638834	0.57820562	0.58591046	0.08559840	-0.2737038	1.00000000	-0.1123210
CYP4A14	0.7509967	0.69875167	0.73687596	0.88534950	-0.28867628	-0.27826594	0.52851762	0.4171278	-0.11232101	1.0000000

>

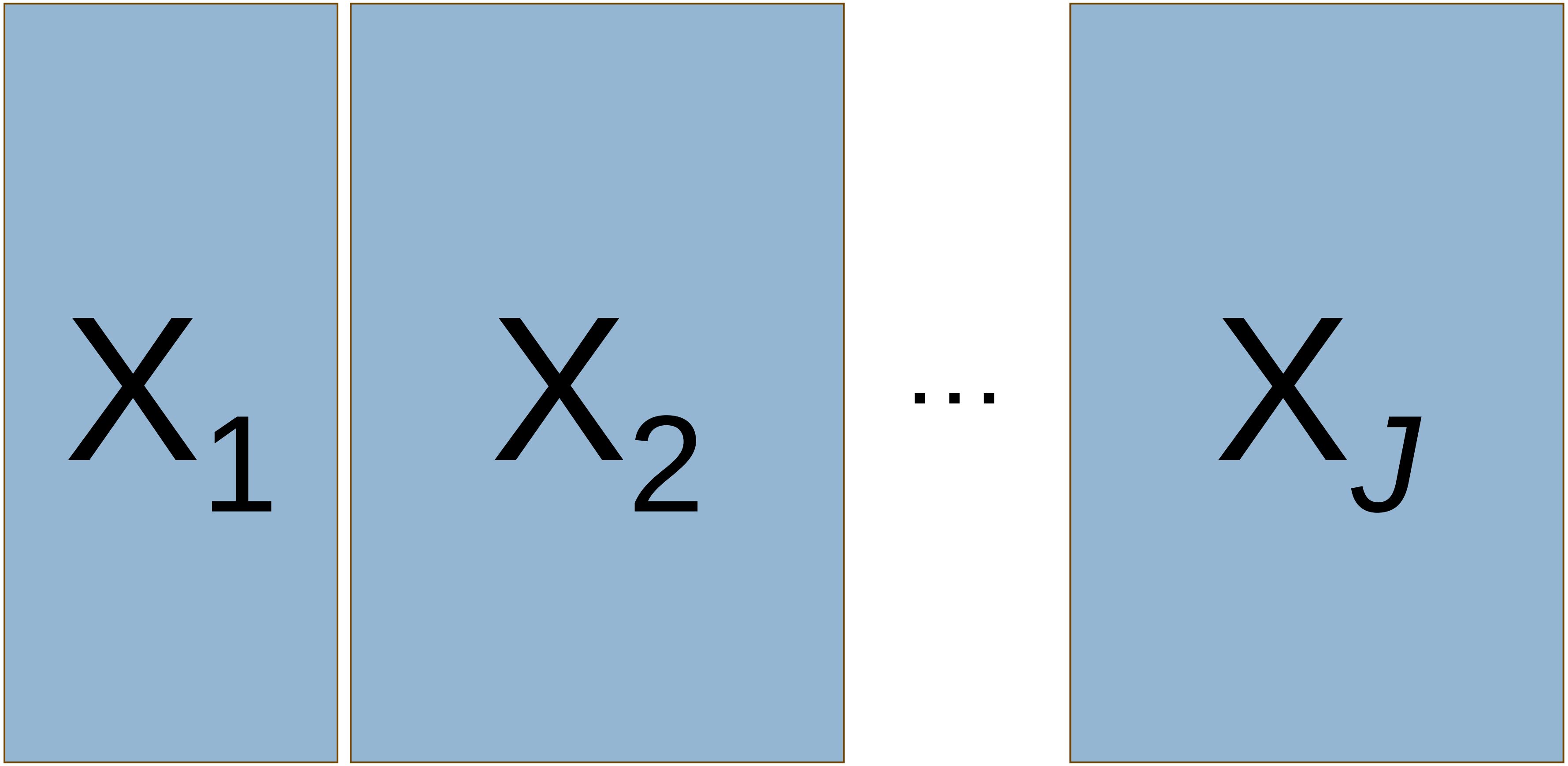
r(LPIN,LPIN1) = 0.97

Looking for factors common to <u>SEVERAL</u> sets of variables





Generalized Canonical Analysis



Generalized Canonical Analysis

X_1 X_2 … X_J

$$z_s \,/ \sum_j R^2(z_s, K_j) \text{ is max}$$

$$\mathrm{Var}(z_s) = 1 \text{ and } \mathrm{Cor}(z_s, z_t) = 0, \forall t < s$$

Generalized Canonical Analysis

For each step *s*:

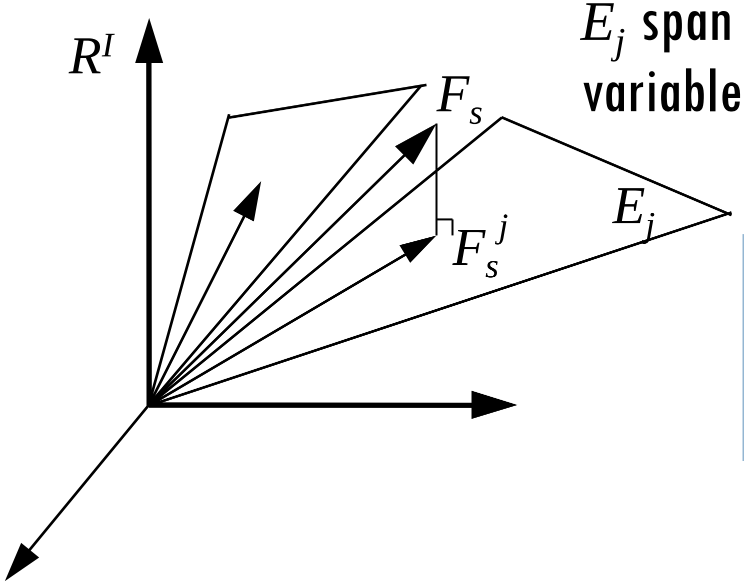
Firstly: general variable (related to all the sets of variables)

Secondly: canonical variables

(linear combination of variables of sets *j* related to general variable)

E_j span by variables of set j

$R^2(z, K_j)$: determination coefficient



F_s maximises $\sum_j R^2(z, K_j)$

$F_s^{\,j}$ projection of F_s on E_j

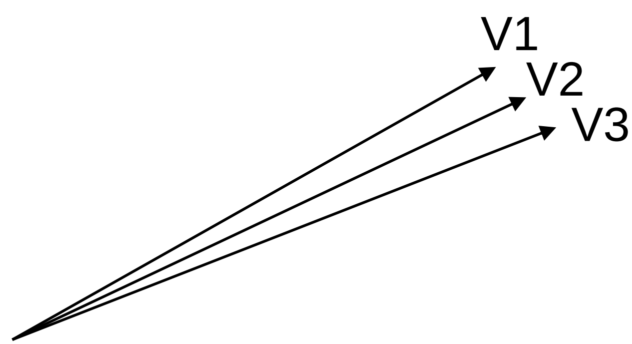
Generalized canonical analysis (Carroll, 1968)

Generalized Canonical Analysis

K_1 K_2 … K_J

$$z_s \,/\, \sum_j R^2(z_s, K_j) \text{ is max}$$

$$\mathrm{Var}(z_s) = 1 \text{ and } \mathrm{Cor}(z_s, z_t) = 0, \forall t < s$$

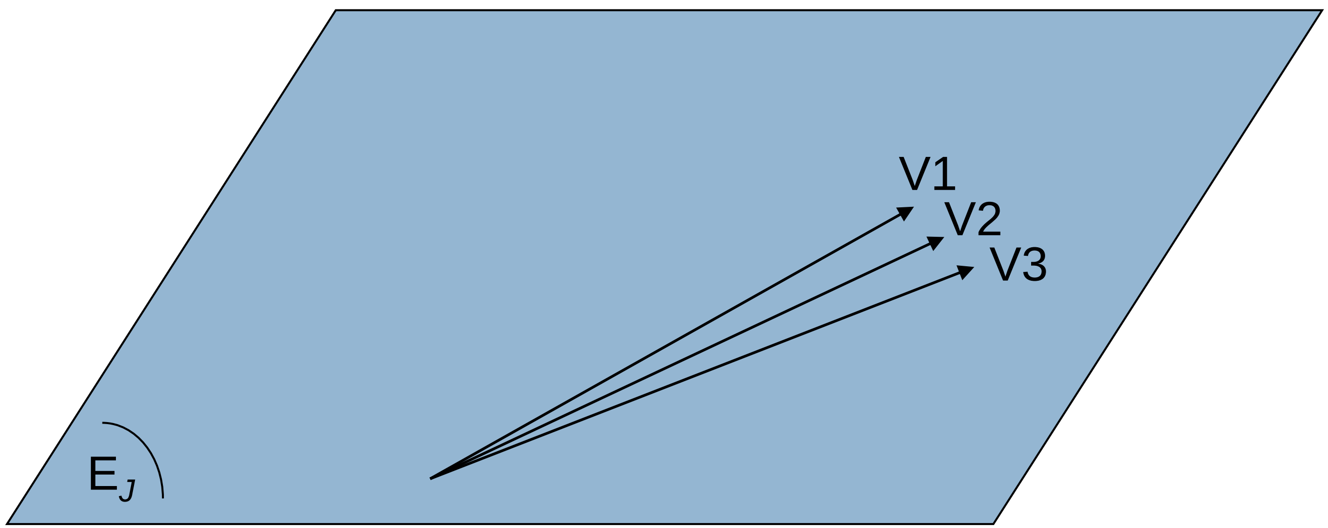


Generalized Canonical Analysis

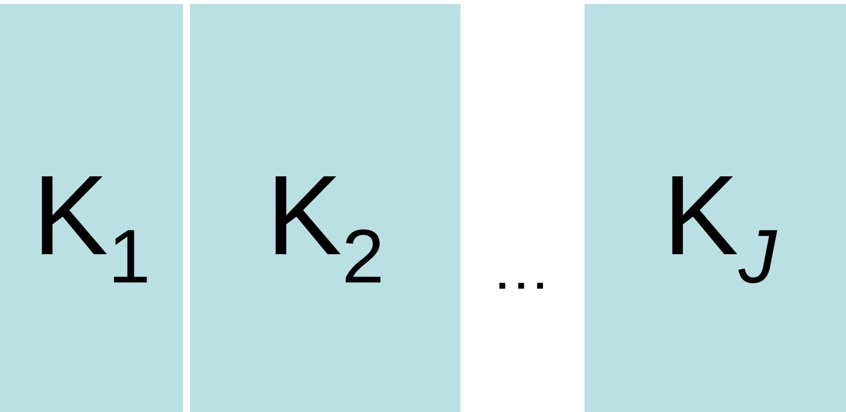
K_1 | K_2 | … | K_J

$$z_s / \sum_j R^2(z_s, K_j) \text{ is max}$$

$$\mathrm{Var}(z_s) = 1 \text{ and } \mathrm{Cor}(z_s, z_t) = 0, \forall t < s$$

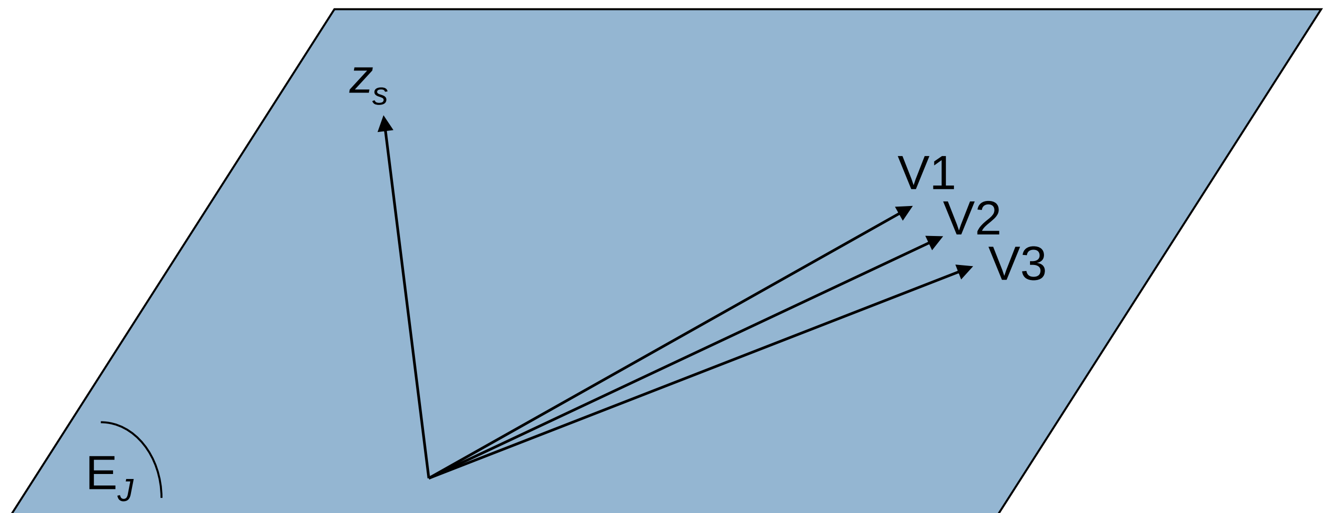


Generalized Canonical Analysis



$$z_s / \sum_j R^2(z_s, K_j) \text{ is max}$$

$$\mathrm{Var}(z_s) = 1 \text{ and } \mathrm{Cor}(z_s, z_t) = 0, \forall t < s$$

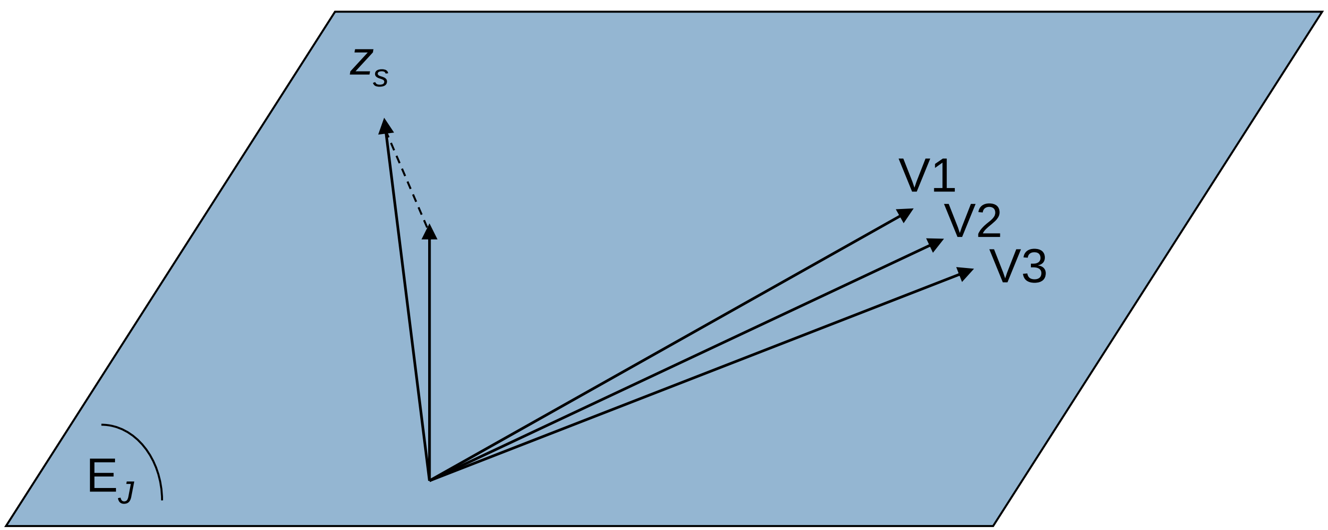


Generalized Canonical Analysis

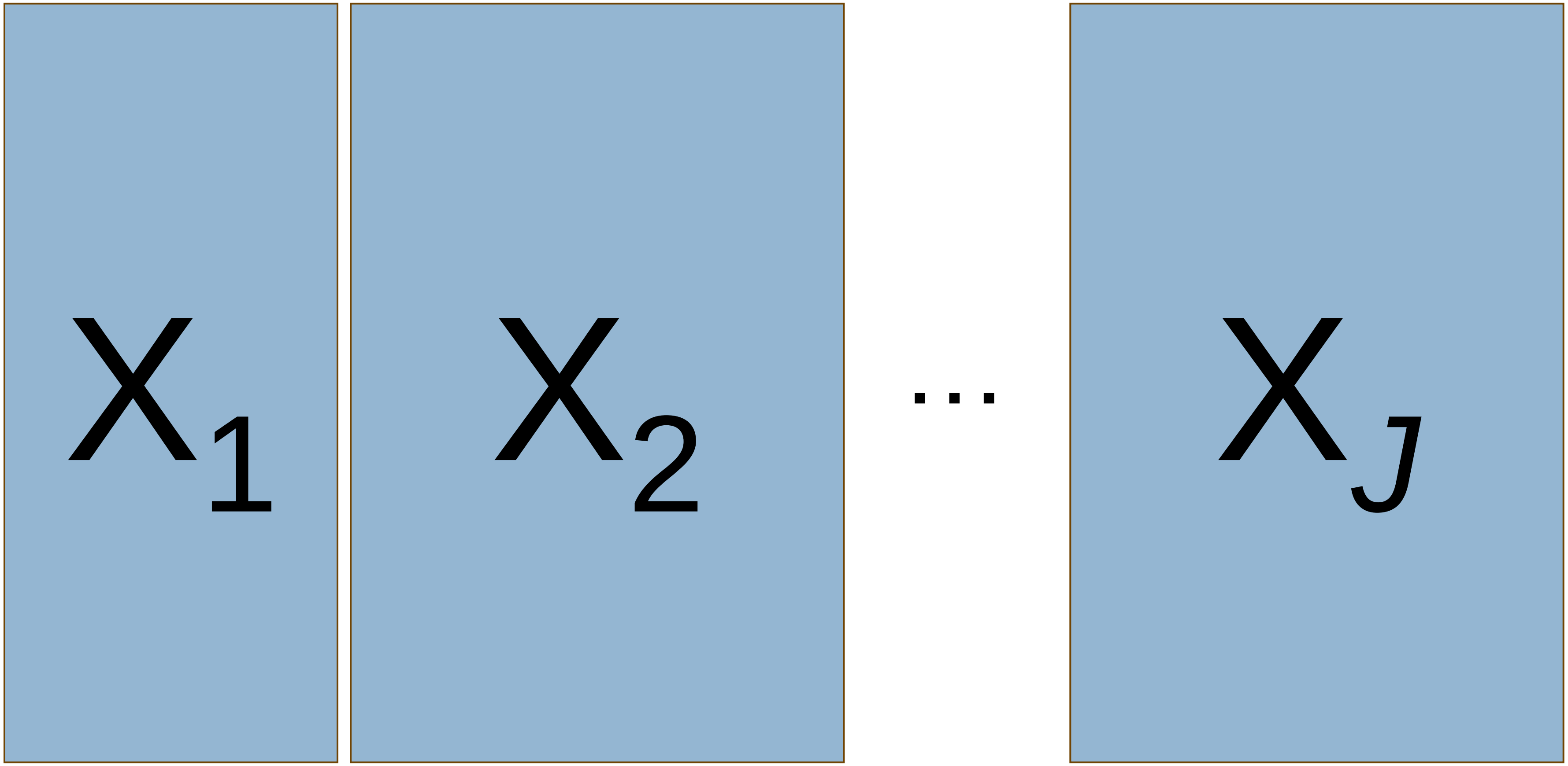
K_1 K_2 … K_J

$$z_s \,/ \sum_j R^2(z_s, K_j) \text{ is max}$$

$$\text{Var}(z_s) = 1 \text{ and } \text{Cor}(z_s, z_t) = 0, \forall t < s$$



Multiple Factor Analysis



Multiple Factor Analysis

X_1 X_2 … X_J

$$z_s \,/ \sum_j Lg(z_s, K_j) \text{ is max}$$

$$\mathrm{Var}(z_s) = 1 \text{ and } \mathrm{Cor}(z_s, z_t) = 0, \forall t < s$$

Multiple Factor Analysis

A measure of relationship between

one variable z

a set of variables $K_j = \{v_k \, ; \, k = 1, K_j\}$

$Lg(z, K_j)$ = projected inertia of the whole set of the variables v_k onto z

Case of standardized variables weighted in a MFA

$$Lg(z, K_j) = \frac{1}{\lambda_1^j} \sum_{k \in K_j} r^2(z, v_k)$$

In every case, owing to the weighting of MFA:

$$0 \le Lg(z, K_j) \le 1$$

Multiple Factor Analysis

K_1	K_2	$\cdots$	K_J

$$z_s \,/\, \sum_j \mathit{Inertia\ of}\ K_j\ \mathit{projected\ on}\ z_s \text{ is max}$$

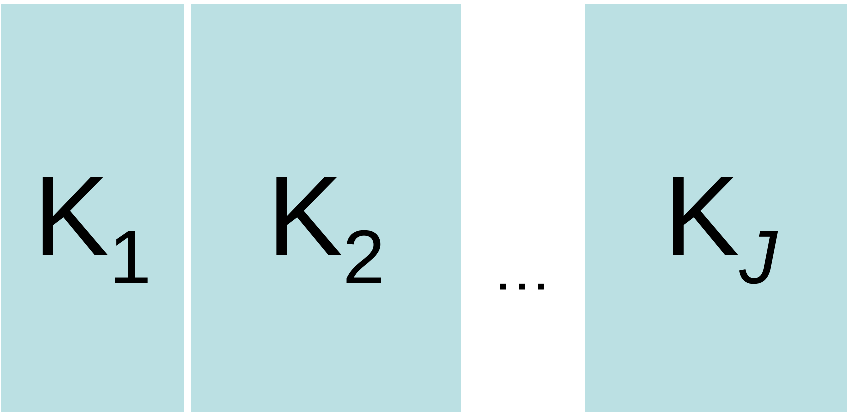
$$\mathrm{Var}(z_s) = 1 \text{ and } \mathrm{Cor}(z_s, z_t) = 0, \forall t < s$$

$$Lg(z, K_j) = \mathit{Inertia\ of}\ K_j\ \mathit{projected\ on}\ z$$

$$= \frac{1}{\lambda_1^j} \sum_{k \in K_j} Cor^2(z, k)$$

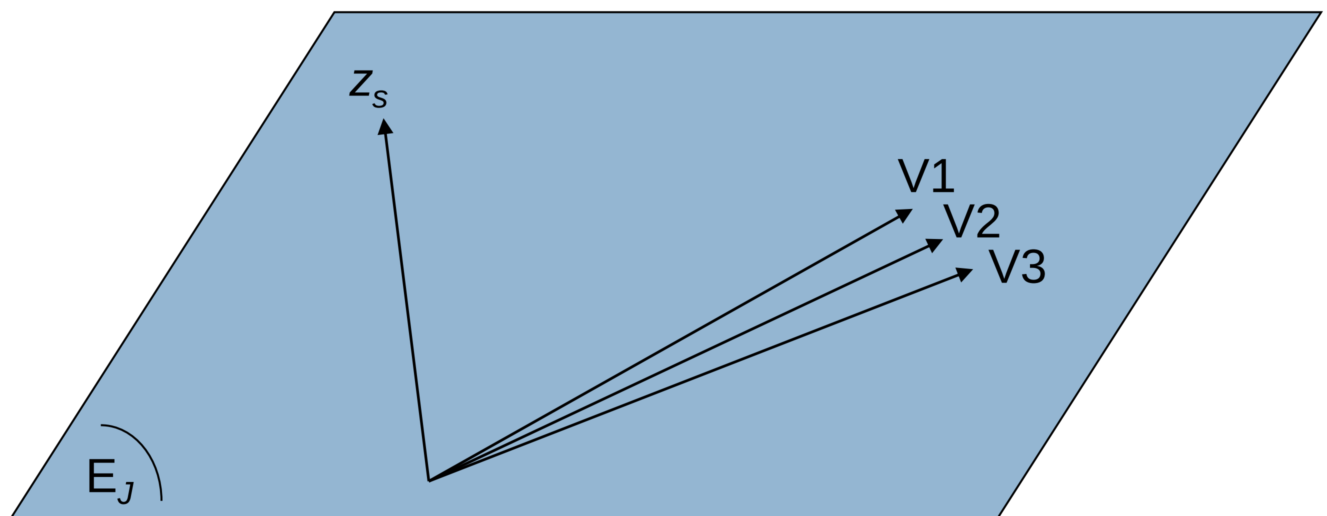
$$Lg(z, K_j) = 1 \Leftrightarrow z \text{ is the } 1^{\text{st}} \text{ princ. comp. of } K_j$$

Multiple Factor Analysis



$$z_s / \sum_j \mathit{Inertia\ of\ } K_j \mathit{\ projected\ on\ } z_s \text{ is max}$$

$$\mathrm{Var}(z_s) = 1 \text{ and } \mathrm{Cor}(z_s, z_t) = 0, \forall t < s$$

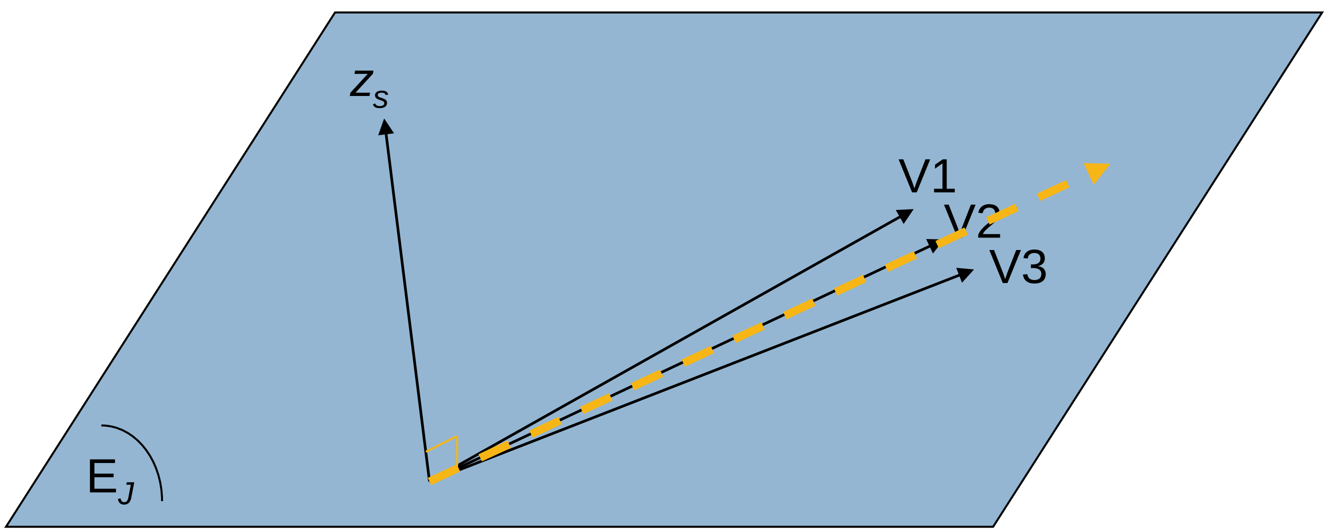


Multiple Factor Analysis

K_1 K_2 … K_J

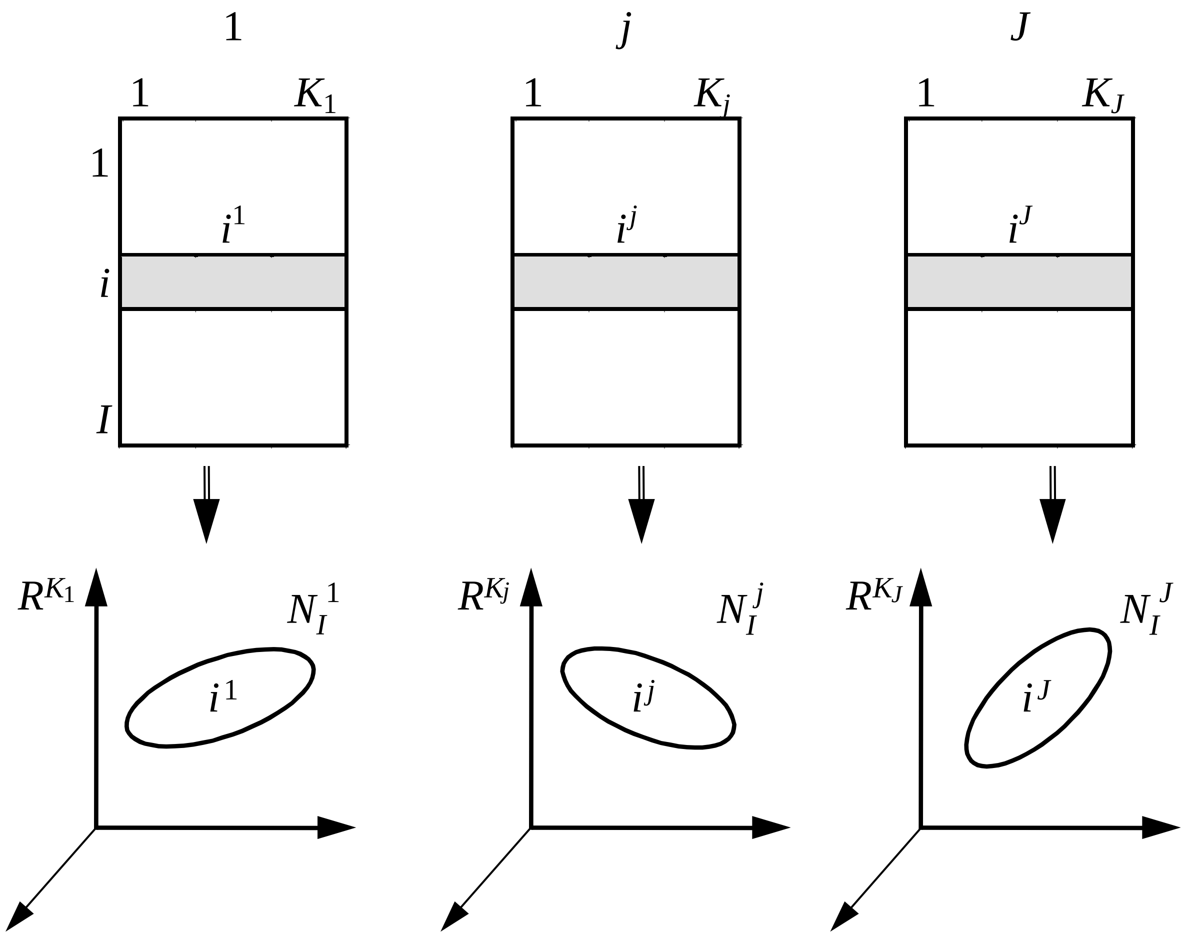
$$z_s / \sum_j \textit{Inertia of } K_j \textit{ projected on } z_s \text{ is max}$$

$$\mathrm{Var}(z_s) = 1 \text{ and } \mathrm{Cor}(z_s, z_t) = 0, \forall t < s$$



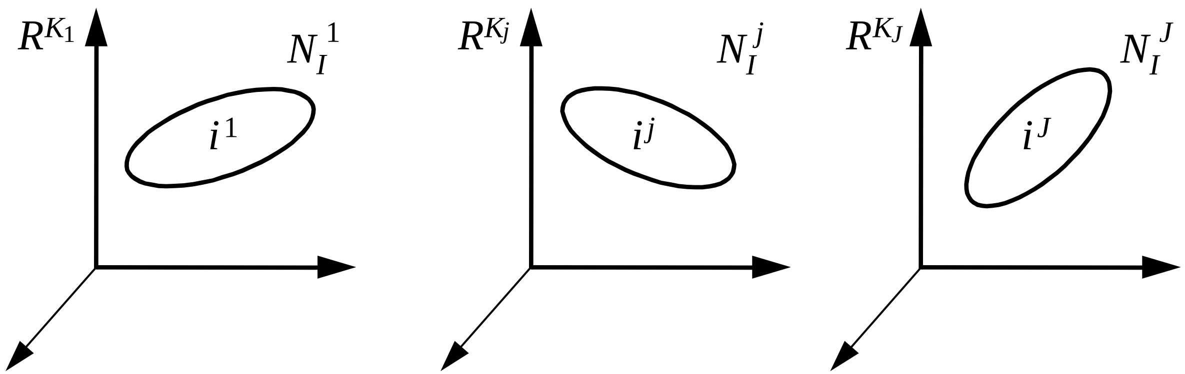
SUPERIMPOSED REPRESENTATION OF THE *J* CLOUDS OF INDIVIDUALS

Superimposed representation of the J clouds of individuals



N_I^j : partial cloud (of individuals; relatively to the set j)

Superimposed representation of the *J* clouds of individuals



How to compare clouds representing the same objects but in different spaces ?

Reference method: Procrustes analysis (Green, 1952; Gower, 1975)

Superimposed representation of the *J* clouds of individuals

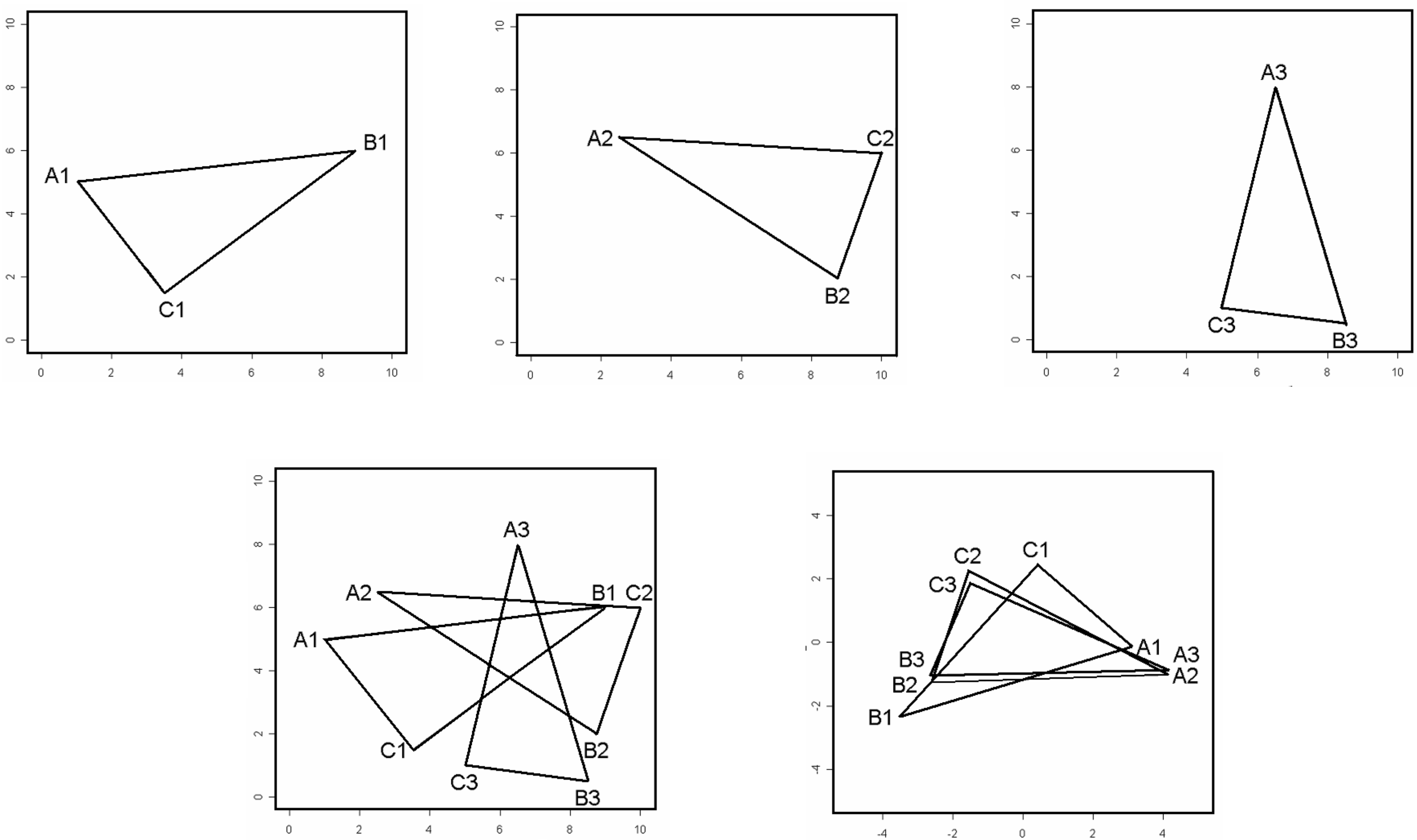
- Procrustes was a character of Greek myth. An innkeeper who plied his trade in Attica, he put his victims on an iron bed. If they were longer than the bed, he cut off their feet. If they were shorter, he stretched them…

Superimposed representation of the *J* clouds of individuals

Make the configurations fit each other

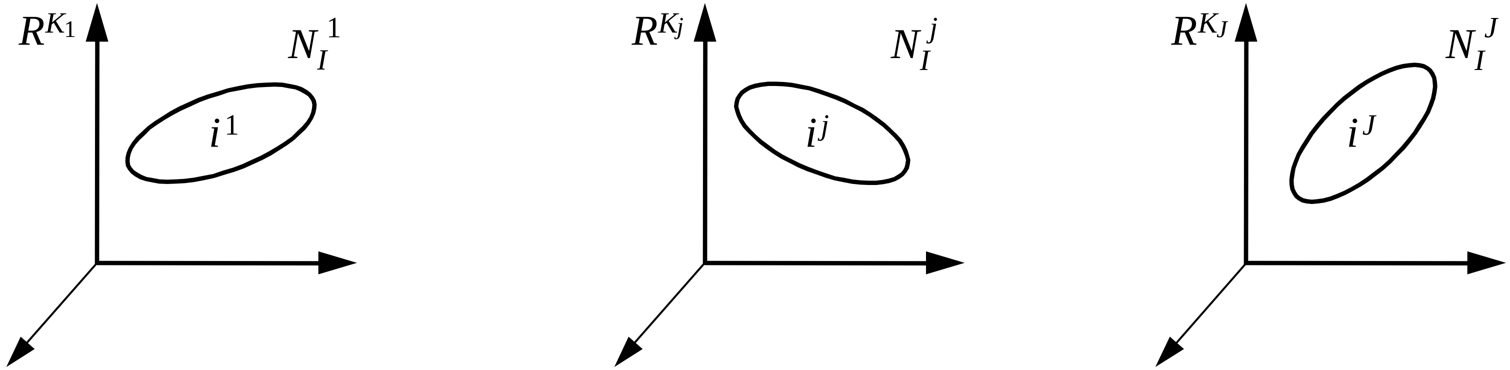
- do this by moving them to a common origin
- stretch or shrink each configuration in order to make it fit as good as possible
- if needed, flip them around

Superimposed representation of the *J* clouds of individuals



Superimposed representation of the *J* clouds of individuals

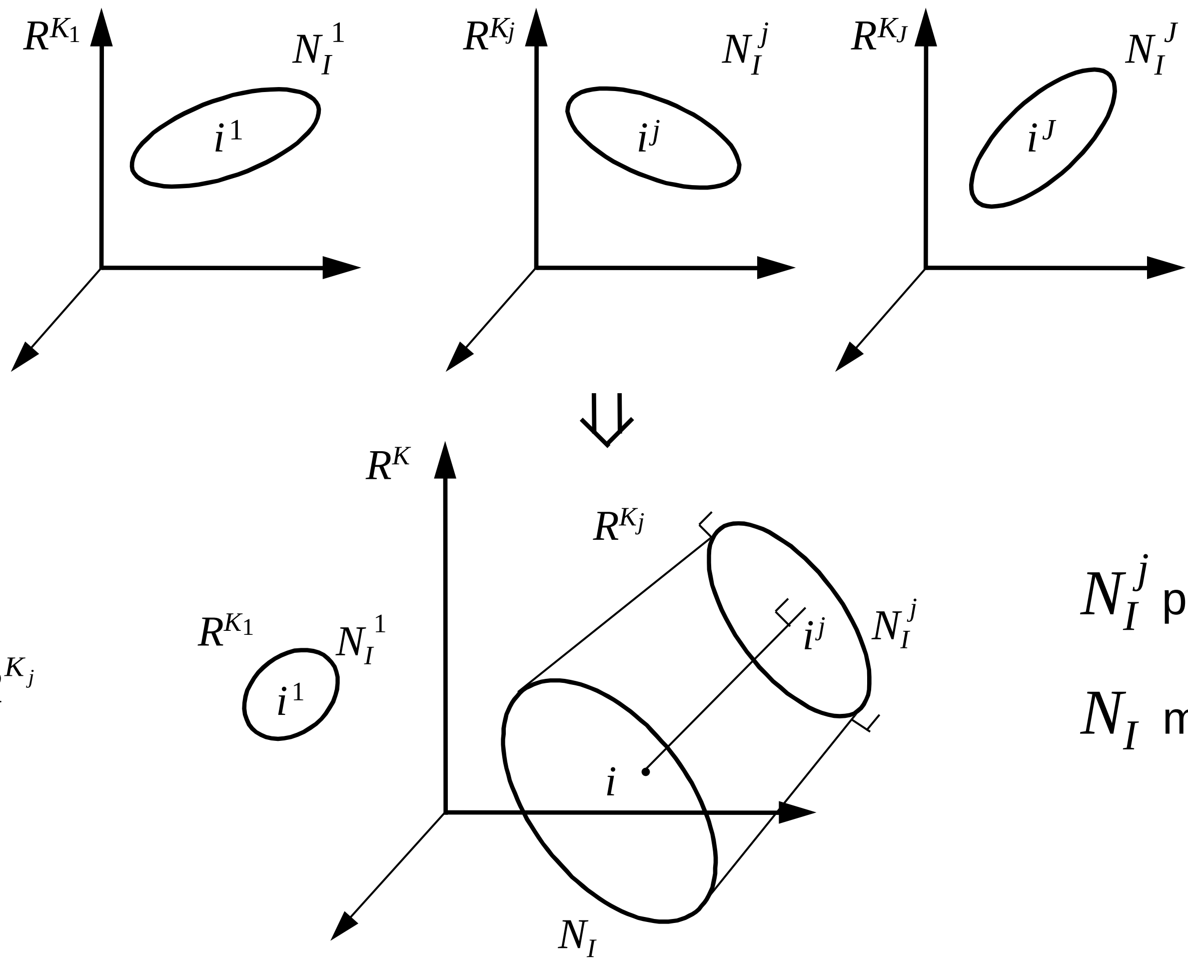
Geometrical framework



1) N_I^j must be well represented
2) The *J* points representing the same individual must be close to one another

Superimposed representation of the *J* clouds of individuals

Geometrical framework



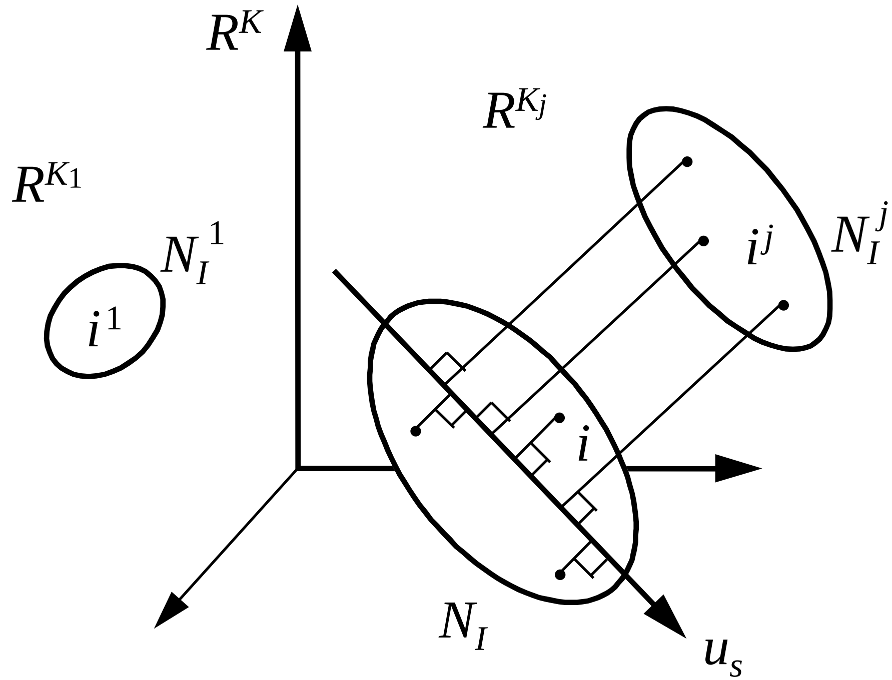
$$R^K = \oplus R^{K_j}$$

N_I^j partial cloud

N_I mean cloud

Superimposed representation of the *J* clouds of individuals

Principle



The partial clouds are projected onto the principal components of the mean cloud

Superimposed representation of the *J* clouds of individuals

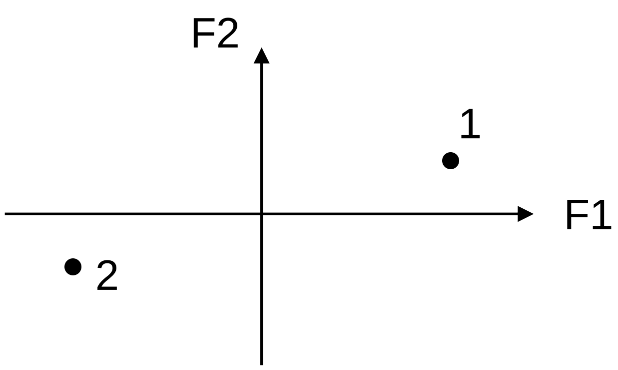
- The superimposed representation and the canonical variables provided by MFA express a very same problematic since they both correspond to the same solution to two apparently different problems

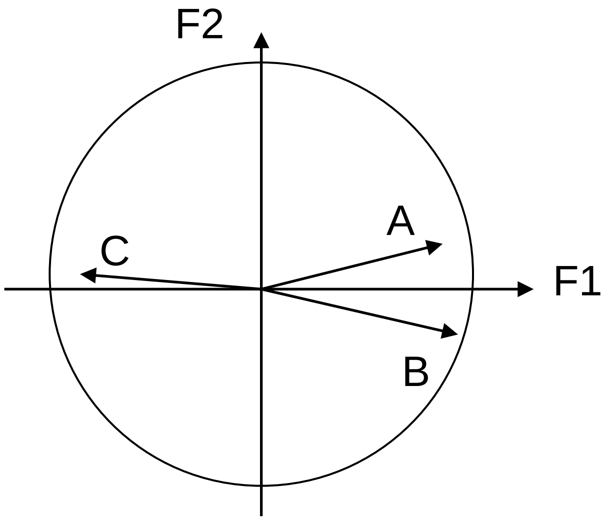
Superimposed representation of the J clouds of individuals

Usual transition relationship in PCA

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} x_{ik} G_s(k)$$

$F_s(i)$ coordinate of i along the axis s
$G_s(k)$ coordinate of variable k along the axis s
λ_s eigenvalue associated to the axis s
x_{ik} data (value of k for i)





Superimposed representation of the *J* clouds of individuals

Usual transition relationship in PCA

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} x_{ik} G_s(k)$$

If the variable *k* has the weight m_k

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} x_{ik} m_k G_s(k)$$

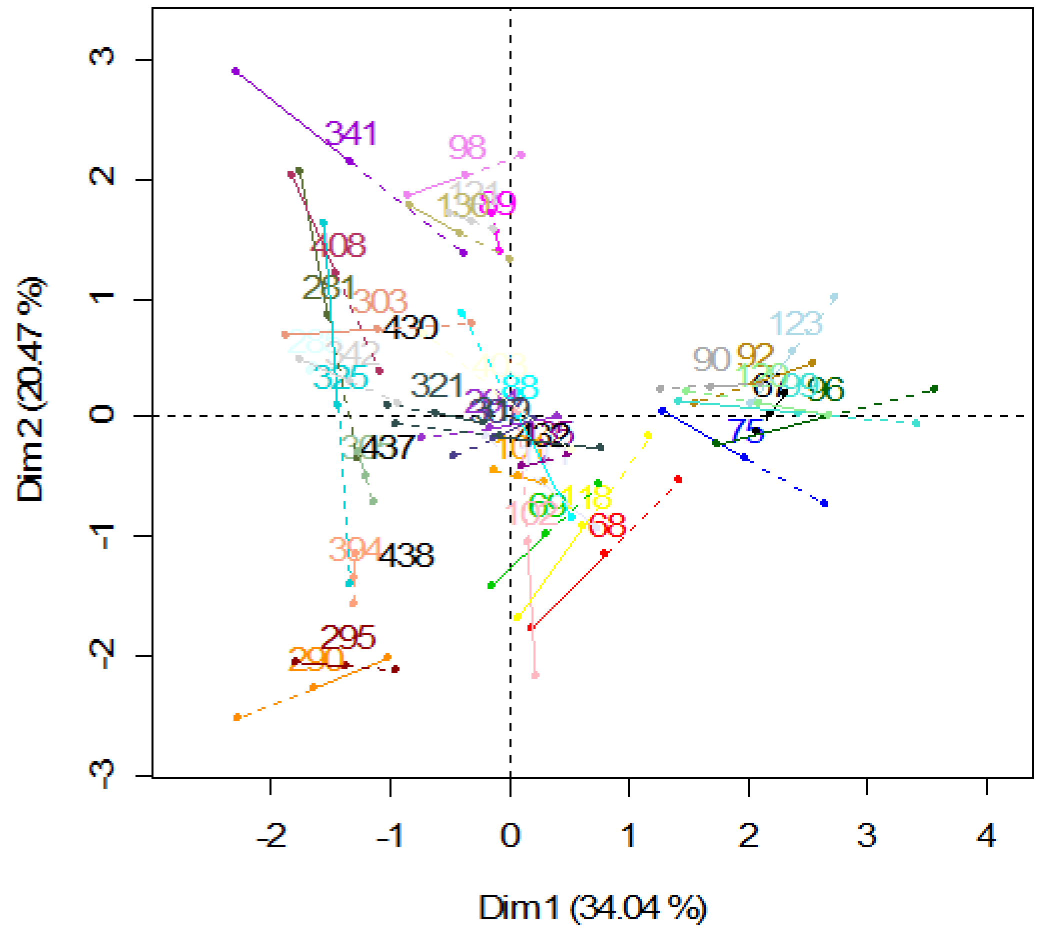
Usual transition relationship applied to the mean cloud in MFA

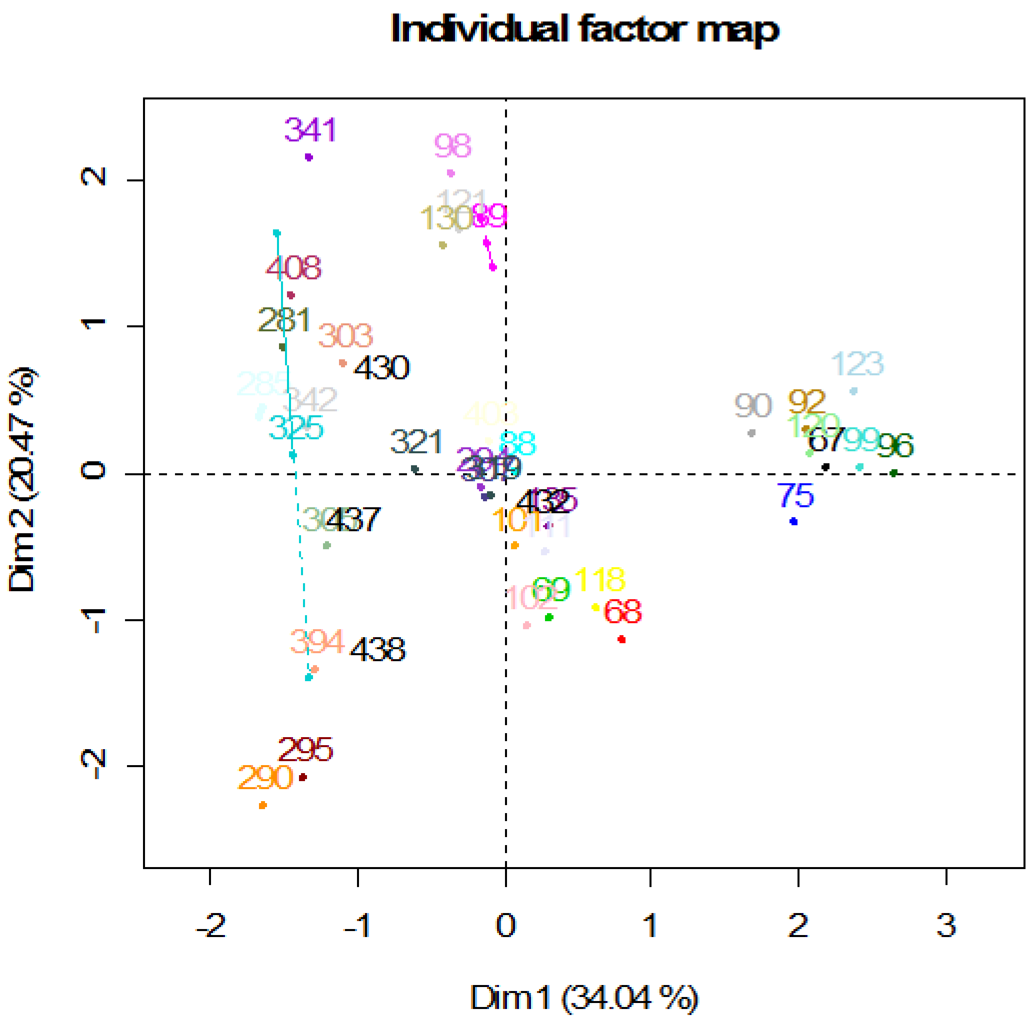
$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j \in J} \frac{1}{\lambda_1^j} \sum_{k \in K_j} x_{ik} G_s(k)$$

Partial transition relationship

$$F_s(i^j) = \frac{1}{\sqrt{\lambda_s}} \frac{J}{\lambda_1^j} \sum_{k \in K_j} x_{ik} G_s(k) \qquad F_s(i) = \frac{1}{J} \sum_{j \in J} F_s(i^j)$$

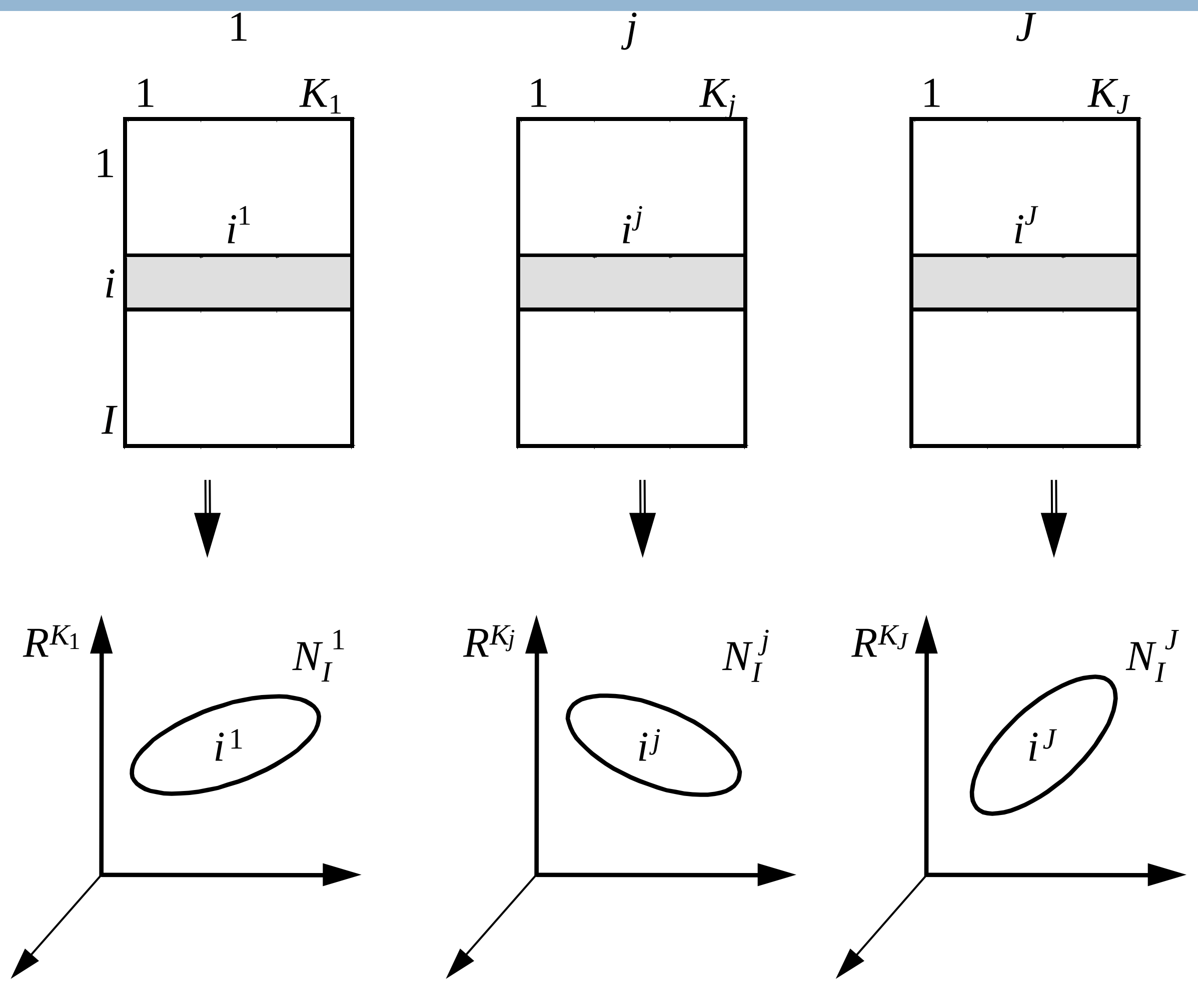
Individual factor map




Individual factor map
Dim 1 (34.04 %)
Dim 2 (20.47 %)
-2
-1
0
1
2
3
341
98
130
89
408
281
303
430
342
325
321
88
437
101
118
69
68
394
438
295
290
90
92
123
67
99
96
75

GLOBAL REPRESENTATION OF SETS OF VARIABLES

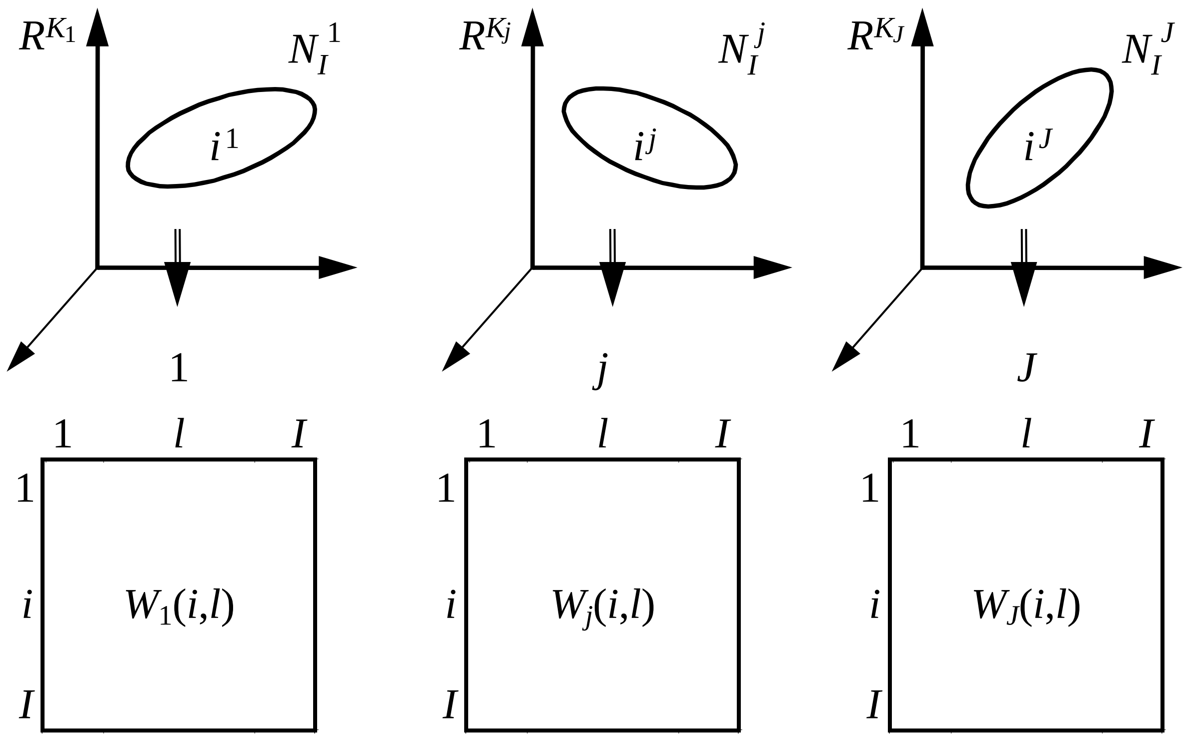
Global representation of sets of variables



N_I^j : partial cloud (of individuals ; associated to the set j)

Global representation of sets of variables

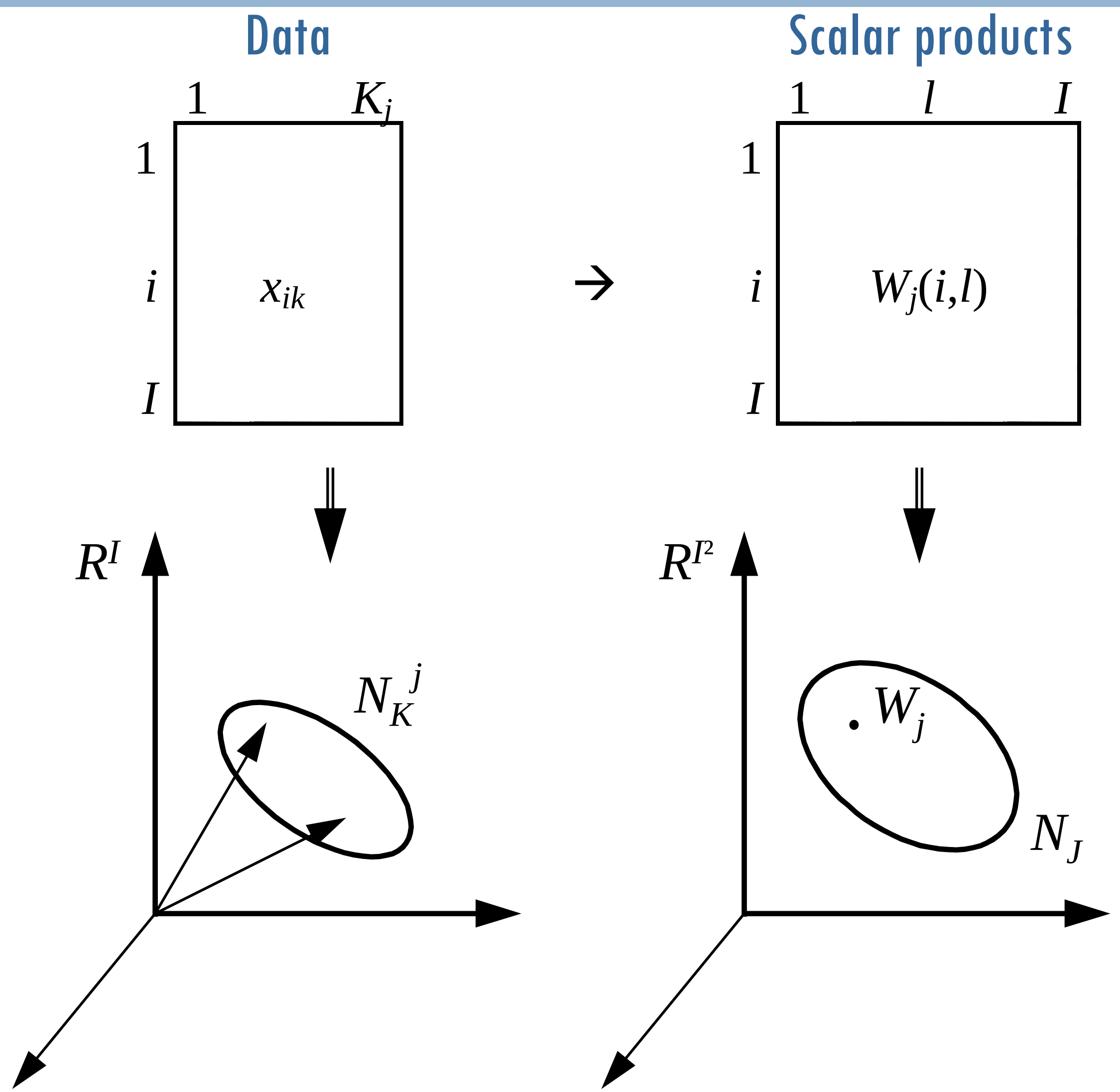
How to measure the global resemblance of the N_I^j ?



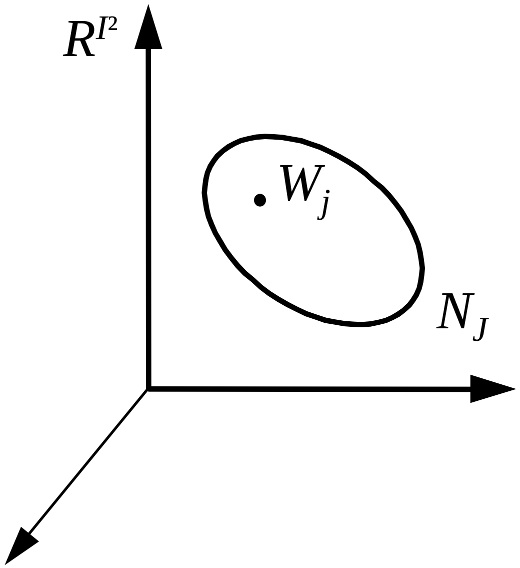
Matrices of scalar products between individuals for each set of variables

$$W_j = X_j X_j'$$

Global representation of sets of variables



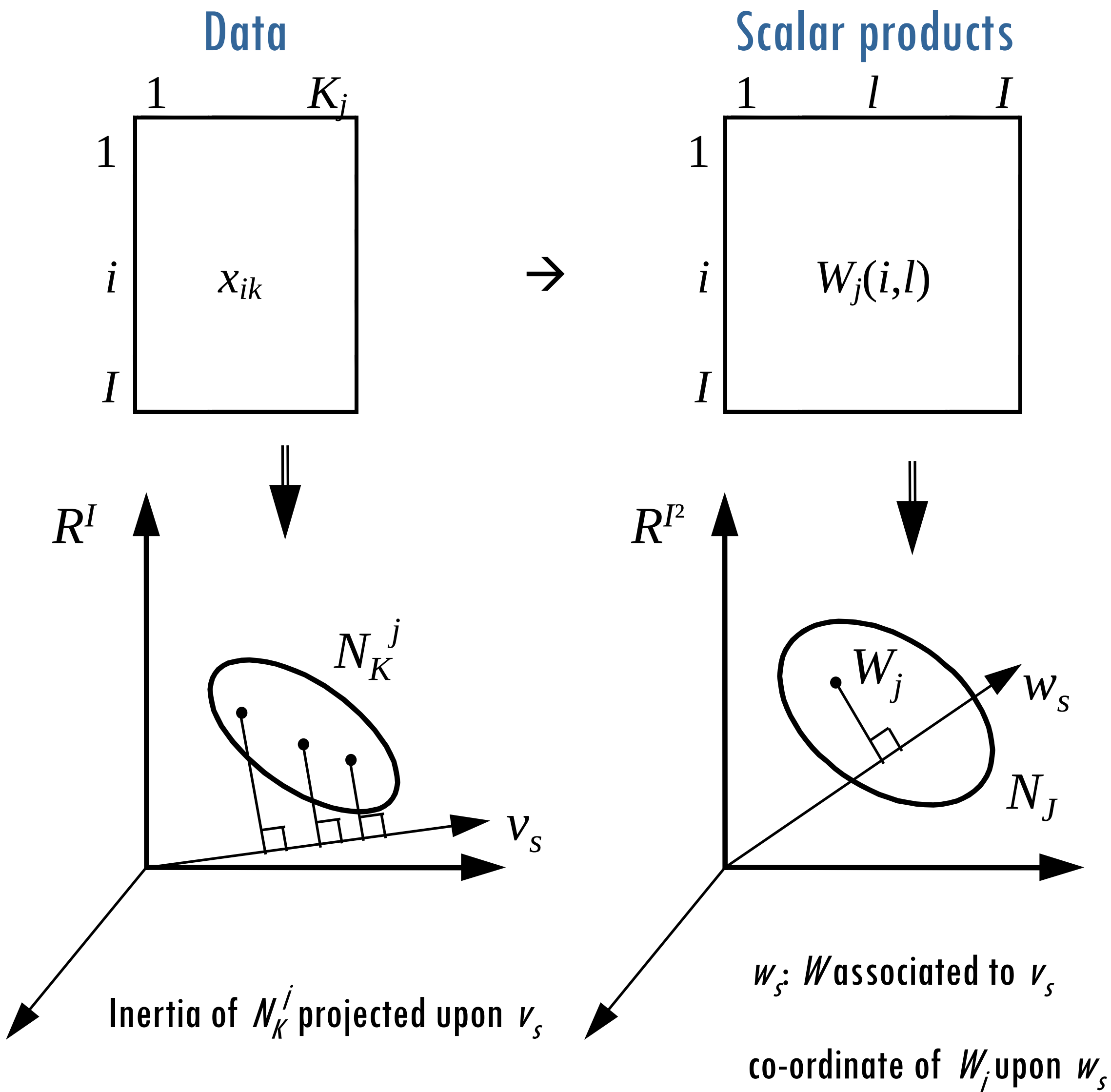
Global representation of sets of variables

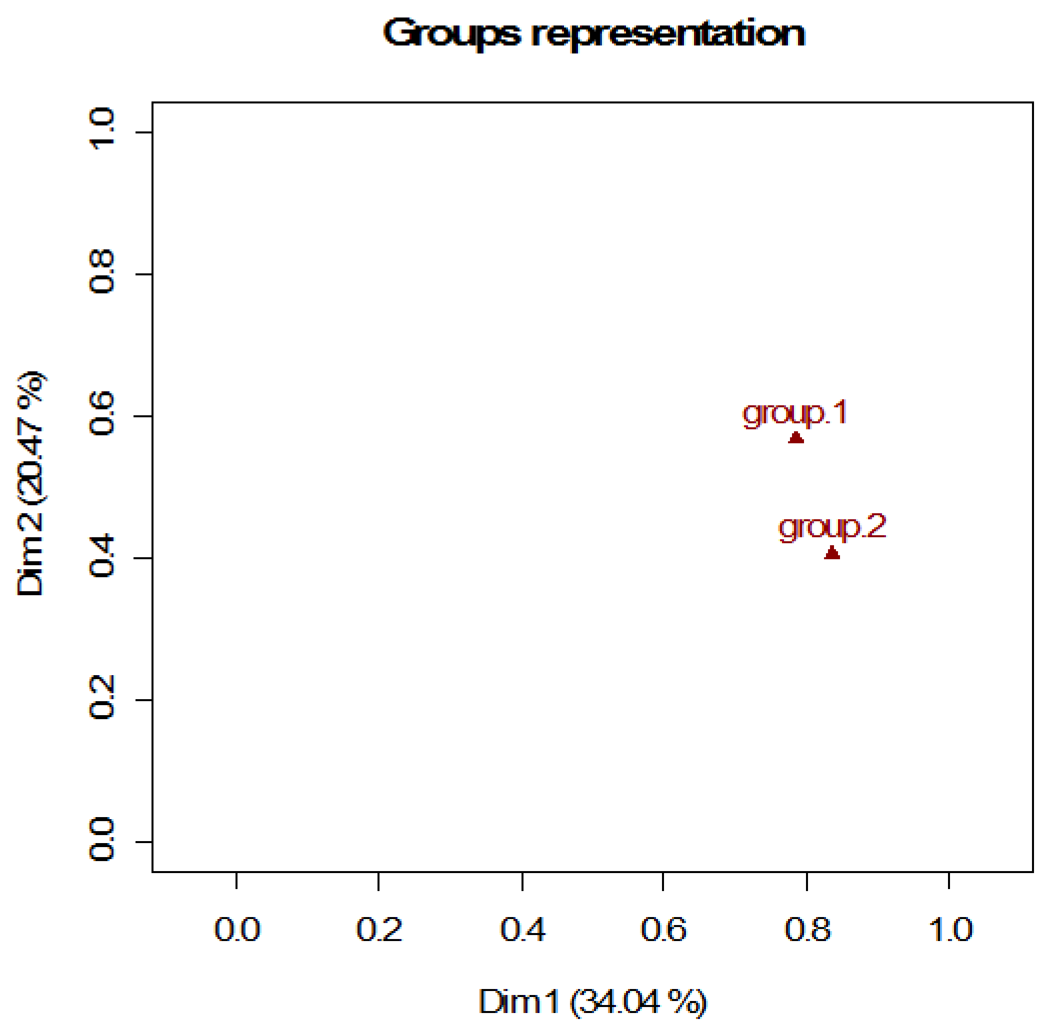


Studying the cloud N_J

Reference method: STATIS (Escoufier Y., Lavit C.)

Global representation of sets of variables




Groups representation
1.0
0.8
0.6
0.4
0.2
0.0
Dim2 (20.47 %)
group.1
group.2
0.0
0.2
0.4
0.6
0.8
1.0
Dim1 (34.04 %)

"THANK YOU FOR BEING HERE"

http://www.ted.com/talks/john_francis_walks_the_earth.html