

le cnam



Sparse Non-Symmetrical Correspondence Analysis

Ruiping Liu¹, Ndeye Niang², Gilbert Saporta²

¹ Beijing Information Sciences & Technology University

² CEDRIC, Conservatoire National des Arts et Métiers, Paris

A joint work with:

- Liu Ruiping

School of Applied Science, Beijing
Information Science and
Technology University



- Ndeye Niang

Cedric- CNAM, Paris



Outline

1. Introduction
2. Sparse CA
3. Non-symmetrical correspondence analysis (NSCA)
4. Sparse NSCA
5. Examples
6. Conclusion and perspectives

1. Introduction

- Sparse PCA (since 20 years)
 - Following sparse regression (Lasso)
 - Linear combinations of a few input variables
 - Simple structure: an alternative to rotations (Trendafilov, 2014)
 - Sparse weights vs sparse loadings
 - Many criteria and algorithms
- Sparse MCA (Bernard et al. , 2012 and Mori et al., 2016)
- Sparse CA (Liu et al., 2023)
 - Sparse weights (or null contributions)
 - Both-sides (rows and columns) sparseness or only one-side
- **Non-symmetrical CA**

Notations (Balbi, 1998)

\mathbf{P} bivariate distribution p_{ij}

$p_{i.}$ $p_{.j}$ rows and columns margins. $\mathbf{D}_I = \text{diag}(p_{i.})$ $\mathbf{D}_J = \text{diag}(p_{.j})$

row profiles: $\mathbf{D}_I^{-1}\mathbf{P}$, column profiles: $\mathbf{P}\mathbf{D}_J^{-1}$

$$\tilde{\mathbf{P}} = \begin{pmatrix} p_{ij} - p_{i.}p_{.j} \end{pmatrix}$$

$$\mathbf{\Pi} = \begin{pmatrix} \frac{p_{ij}}{p_{.j}} - p_{i.} \end{pmatrix} = \tilde{\mathbf{P}}\mathbf{D}_J^{-1} \text{ matrix of centered column profiles}$$

2. Sparse CA

- CA is a simultaneous PCA of rows and columns weighted profiles matrices with chi-square metrics
 - Triplets $(\mathbf{D}_I^{-1}\tilde{\mathbf{P}}, \mathbf{D}_J^{-1}, \mathbf{D}_I)$ and $(\mathbf{D}_J^{-1}\tilde{\mathbf{P}}', \mathbf{D}_I^{-1}, \mathbf{D}_J)$
 - Plain SVD of $\mathbf{D}_I^{-1/2}\tilde{\mathbf{P}}\mathbf{D}_J^{-1/2}$ or a GSVD of $\tilde{\mathbf{P}}$
- Hence the idea of applying some variant of sparse PCA or of sparse SVD, namely PMD for *Penalized Matrix Decomposition* (Witten et al., 2009) by Liu et al., 2023

- PMD applied to $\mathbf{X} = \mathbf{D}_I^{-1/2} \tilde{\mathbf{P}} \mathbf{D}_J^{-1/2}$

$$\max \mathbf{u}' \mathbf{X} \mathbf{v}, \text{ s.t. } \|\mathbf{u}\|^2 = \|\mathbf{v}\|^2 = 1 \text{ and } \|\mathbf{u}\|_1 \leq c_1, \|\mathbf{v}\|_1 \leq c_2$$

- Pseudo singular value $\alpha = \sqrt{\lambda} \mathbf{u}' \mathbf{X} \mathbf{v}$ and pseudo eigenvalue λ
- Sparse rows and columns weights $\mathbf{f} = \alpha \mathbf{D}_I^{-1/2} \mathbf{u}$ and $\mathbf{g} = \alpha \mathbf{D}_J^{-1/2} \mathbf{v}$
 - The level of sparsity for rows and columns weights is driven by c_1 and c_2
- Rows and columns coordinates \mathbf{a} and \mathbf{b} are given by pseudo-barycentric equations

$$\mathbf{a} \propto \mathbf{D}_I^{-1} \mathbf{P} \mathbf{g} \quad \text{with} \quad \mathbf{a}' \mathbf{D}_I \mathbf{a} = \lambda,$$

$$\mathbf{b} \propto \mathbf{D}_J^{-1} \mathbf{P}' \mathbf{f} \quad \text{with} \quad \mathbf{b}' \mathbf{D}_J \mathbf{b} = \lambda.$$

- Weights are sparse but coordinates are not.

Lost properties

- Transition equations are no longer true: \mathbf{g} is not proportional to \mathbf{b} and \mathbf{f} is not proportional to \mathbf{a} . The weights differ from the coordinates.
- However maps should use coordinates \mathbf{a} and \mathbf{b} and not weights. Otherwise there is a risk of having too many points stuck to the axes, those with zero weights, and thus uninterpretable maps.

Issues

How to choose the right level of sparseness?

- An index derived from Trendafilov (2014):

$$IS_u = \frac{V_s}{V_0} \cdot \frac{V_a}{V_0} \cdot \frac{\#0(\mathbf{u})}{I}$$

$$IS_v = \frac{V_s}{V_0} \cdot \frac{V_a}{V_0} \cdot \frac{\#0(\mathbf{v})}{J}$$

$$IS = IS_u + IS_v$$

Which deflation method to use to get higher order solutions?

- A new deflation algorithm BiOPD (Liu et al., 2023) :

Bilateral Orthogonalized Projected Deflation based upon an iteration of:

$$\left(\mathbf{I}_I - \mathbf{u}_1 \mathbf{u}_1'\right) \mathbf{X} \left(\mathbf{I}_J - \mathbf{v}_1 \mathbf{v}_1'\right)$$

Preserves orthogonality of weights, but not of components (however almost uncorrelated)

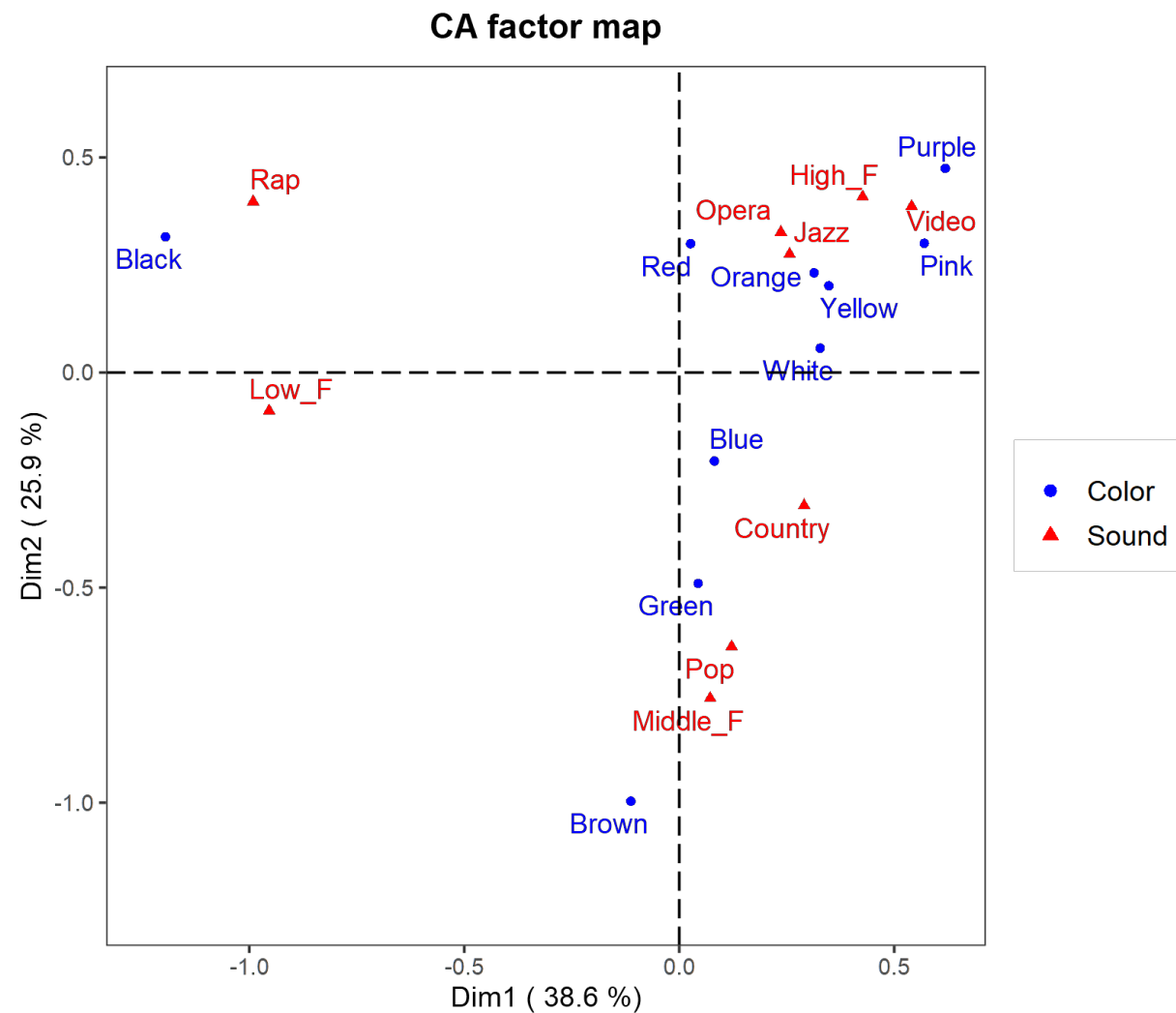
A toy example: colours of sound

Color	Video	Jazz	Country	Rap	Pop	Opera	Low F	High F	Middle F	x_{i+}	\mathbf{r}
Red	4	2	4	4	1	2	2	4	1	24	0.121
Orange	3	4	2	2	1	1	0	3	2	18	0.091
Yellow	6	4	5	2	3	1	1	3	0	25	0.126
Green	2	0	5	1	3	3	3	1	5	23	0.116
Blue	2	5	0	1	4	1	2	1	3	19	0.096
Purple	3	3	1	0	0	3	0	2	1	13	0.066
White	0	0	0	0	1	4	1	5	3	14	0.071
Black	0	2	0	11	1	3	10	1	1	29	0.146
Pink	2	1	1	0	2	4	0	2	0	12	0.061
Brown	0	1	4	1	6	0	3	0	6	21	0.106
x_{+j}	22	22	22	22	22	22	22	22	22	$N = 198$	1.000
\mathbf{c}^T	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11		

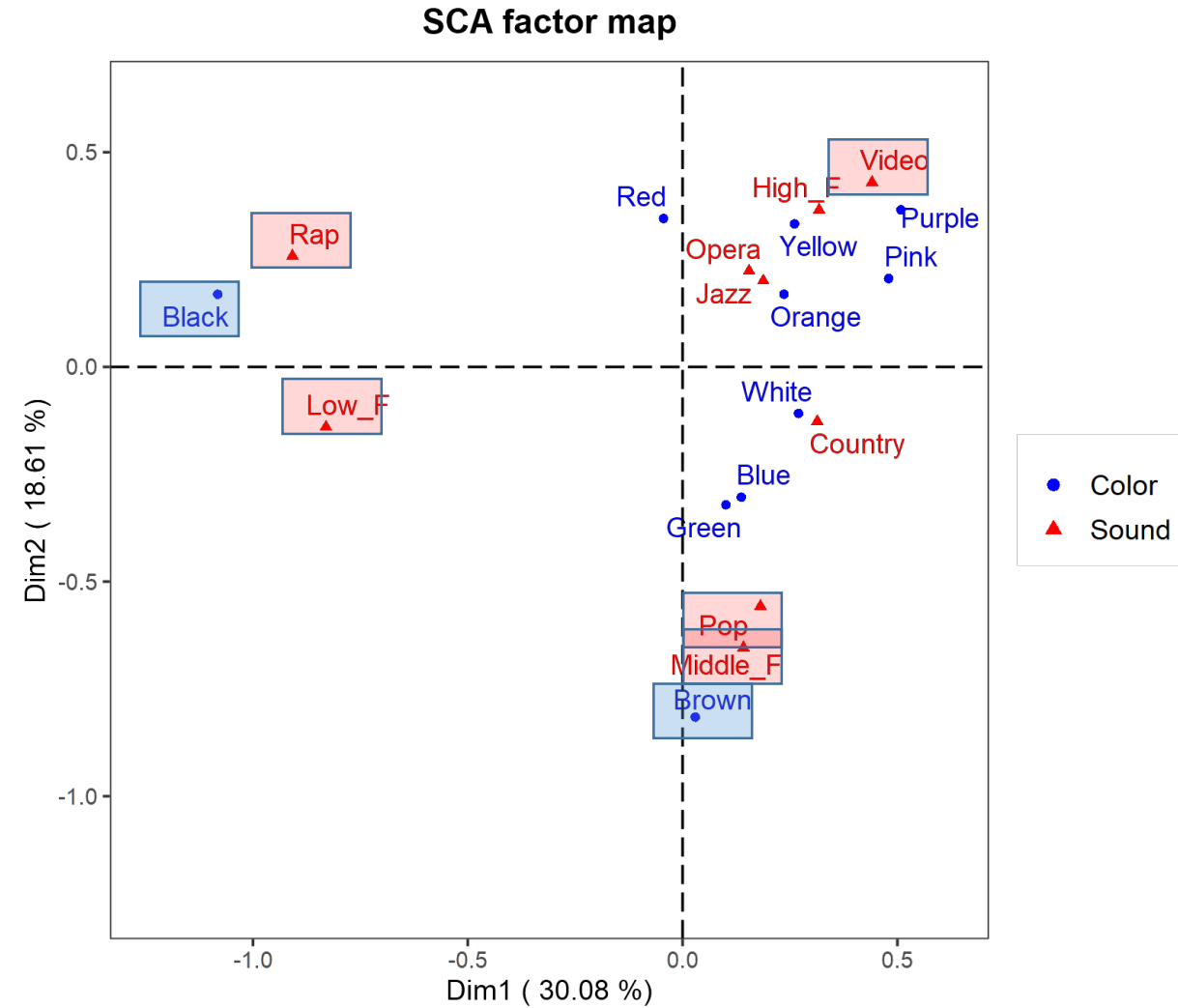
Abdi H., Béra M. (2017) *Correspondence Analysis*

In: Alhajj R., Rokne J. (eds) *Encyclopedia of Social Network Analysis and Mining*, Springer, New York,

	f_1	f_2	ctr1	ctr2
Red	0.026	0.299	0	56
Orange	0.314	0.232	31	25
Yellow	0.348	0.202	53	27
Green	0.044	-0.49	1	144
Blue	0.082	-0.206	2	21
Purple	0.619	0.475	87	77
White	0.328	0.057	26	1
Black	-1.195	0.315	726	75
Pink	0.57	0.3	68	28
Brown	-0.113	-0.997	5	545
	g_1	g_2	ctr1	ctr2
Video	0.541	0.386	113	86
Jazz	0.257	0.275	25	44
Country	0.291	-0.309	33	55
Rap	-0.991	0.397	379	91
Pop	0.122	-0.637	6	234
Opera	0.236	0.326	22	61
LowF	-0.954	-0.089	351	5
HighF	0.427	0.408	70	96
MiddleF	0.072	-0.757	2	330



	f_1	f_2	ctr1	ctr2	a_1	a_2
Red	0	0.291	0	74	-0.044	0.346
Orange	0.114	-0.007	5	0	0.236	0.169
Yellow	0.169	0.274	16	68	0.261	0.333
Green	0	-0.249	0	52	0.101	-0.321
Blue	0	-0.197	0	27	0.137	-0.303
Purple	0.411	0.232	49	25	0.508	0.366
White	0.132	-0.008	5	0	0.270	-0.108
Black	-1.166	0.067	887	5	-1.082	0.169
Pink	0.369	-0.021	37	0	0.480	0.206
Brown	0	-0.99	0	749	0.030	-0.816
	g_1	g_2	ctr1	ctr2	b_1	b_2
Video	0.215	0.386	23	119	0.441	0.430
Jazz	0	0	0	0	0.188	0.201
Country	0	0	0	0	0.314	-0.127
Rap	-1.062	0.049	558	2	-0.908	0.258
Pop	0	-0.621	0	309	0.181	-0.557
Opera	0	0	0	0	0.155	0.223
LowF	-0.92	0.033	419	1	-0.830	-0.140
HighF	0	0.264	0	56	0.318	0.366
MiddleF	0	-0.801	0	513	0.142	-0.654



- Some decrease of inertia: 48.7 % instead of 64.5 %
- But
 - Similar displays
 - Important contributions are highlighted, small ones set to zero
 - Easier to interpret

Both-sides or one-side sparseness?

- Similar levels of sparseness were sought for rows and columns
- Relaxing constraints on $\|\mathbf{u}\|_1$ or $\|\mathbf{v}\|_1$ lead to column-sparse or row-sparse CA.
- Useful for *eg* document-terms data where one likes to keep all documents but not all words

Symmetry or not?

- CA and sparse CA consider the rows and columns of \mathbf{P} in the same way
- Not always a good fit if the relationship is not symmetrical:
 - In the toy example we first listen to the music before associating colors with it, and not the other way around.

3. Non symmetrical correspondence analysis

- NSCA was introduced by Lauro & d'Ambra (1984) for the case *where one variable is treated as a predictor variable and the other variable is treated as a response variable* (Beh & Lombardo, 2021). See also Lauro & Balbi (1999), Kroonenberg & Lombardo (1999).
- When rows are responses and columns are predictors, NSCA is a PCA of the triplet:

$$\left(\tilde{\mathbf{P}}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{I}_I \right)$$

- A PCA of the weighted centered columns profiles, **where columns are weighted but not rows,**
- or a generalized SVD of $\tilde{\mathbf{P}}\mathbf{D}_J^{-1}$

- In NSCA, inertia is equal to the numerator of Goodman-Kruskal's τ_b unlike in CA where inertia is equal to ϕ^2
- NSCA provides a decomposition of τ_b according to orthogonal dimensions
- NSCA allows to create simultaneous graphical displays, in particular isometric biplots in two dimensions.

4. Sparse NSCA

- A sparse GSVD of $\mathbf{X} = \tilde{\mathbf{P}}\mathbf{D}_J^{-1}$
or a sparse SVD of $\mathbf{X} = \tilde{\mathbf{P}}\mathbf{D}_J^{-1/2}$

$$\max \mathbf{u}'\mathbf{X}\mathbf{v} , \text{ s.t. } \|\mathbf{u}\|^2 = \|\mathbf{v}\|^2 = 1 \text{ and } \|\mathbf{u}\|_1 \leq c_1 , \|\mathbf{v}\|_1 \leq c_2$$

- Sparseness for rows only: no constraints on $\|\mathbf{v}\|_1$
- Sparseness for columns only: no constraints on $\|\mathbf{u}\|_1$

5. Examples

- 5.1 Food data

Beh, E. J., Lombardo, R., & Simonetti, B. (2011).

A European perception of food using two methods of correspondence analysis.

Food quality and preference, 22(2), 226-231.

Beh, E.J., Lombardo, R. (2021). Non-Symmetrical Correspondence Analysis. In: *An Introduction to Correspondence Analysis*, chapter 3, John Wiley, 71-97

Table 1

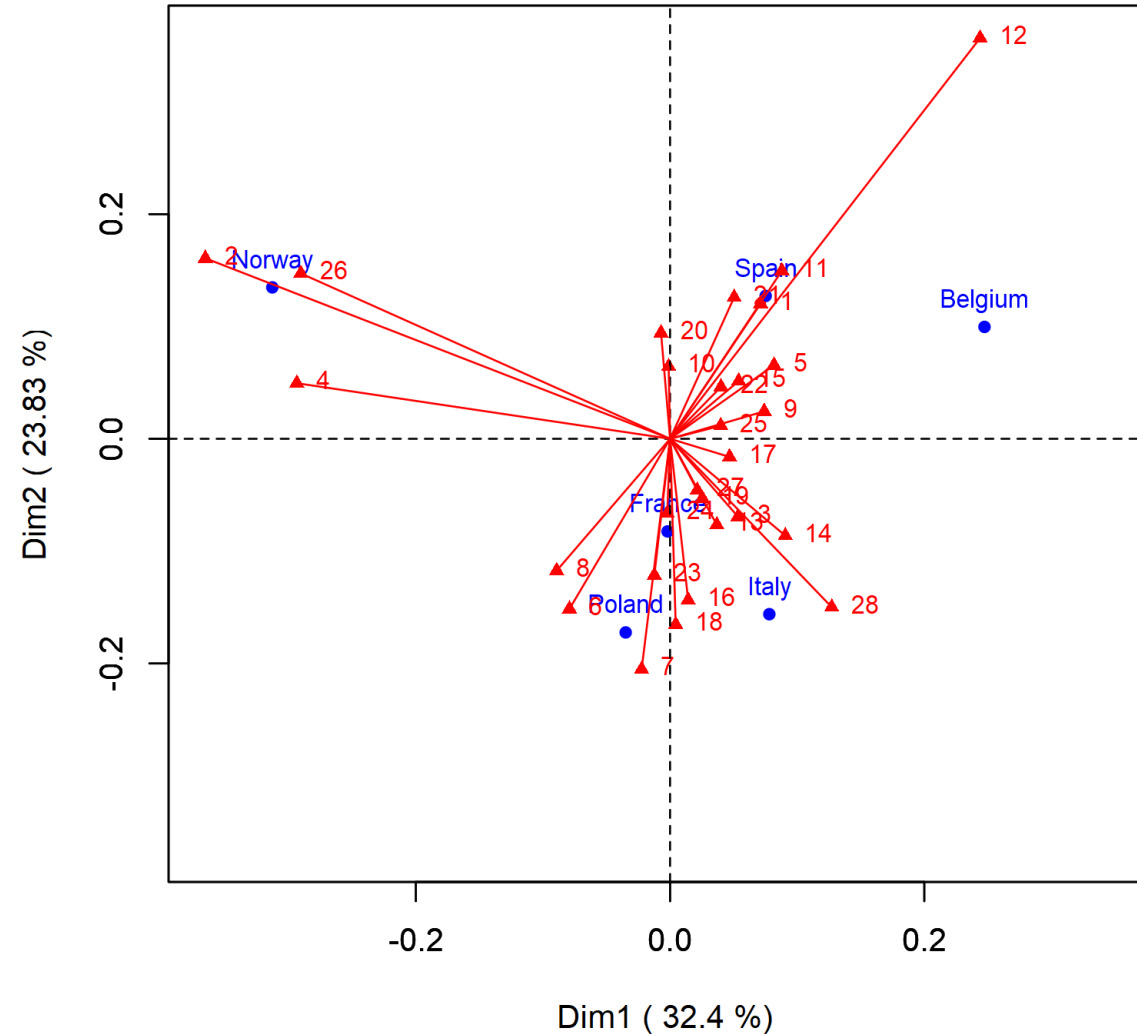
Cross-classification of individuals according to nationality and the 28 most common free-word associations of “Traditional” when describing food.

Code	Word Association	Country					
		Belgium	France	Italy	Norway	Poland	Spain
1	Ancient	0	0	2	0	0	15
2	Christmas	0	1	0	22	11	9
3	Cooking	2	17	0	0	5	9
4	Country	0	4	0	17	12	6
5	Culture	4	3	0	0	1	7
6	Dinner	0	0	0	3	14	0
7	Dish	0	10	0	0	11	0
8	Family	3	16	5	10	16	13
9	Feast	8	9	0	2	0	0
10	Good	1	2	6	5	2	12
11	Grandmother	10	4	1	5	2	9
12	Habit	14	3	0	3	0	25
13	Healthy	4	1	3	2	11	6
14	Holidays	8	0	0	0	12	1
15	Home	1	1	0	0	5	12
16	Home-made	0	0	8	2	12	9
17	Kitchen	4	0	2	1	4	2
18	Meal	1	6	1	0	11	2
19	Natural	2	2	5	2	5	5
20	Old	2	8	0	6	11	22
21	Old-fashioned	10	0	0	5	0	0
22	Quality	3	4	0	1	0	4
23	Recipe	0	7	0	0	6	0
24	Regional	1	7	4	2	1	1
25	Restaurant	5	5	1	2	1	1
26	Rural	0	12	1	19	5	11
27	Simple	2	1	3	1	3	1
28	Tasty	8	3	4	0	17	8

	u1	u2	ctr.f1	ctr.f2
Belgium	1,588	0,746	316	70
France	-0,013	-0,628	0	67
Italy	0,501	-1,161	16	83
Norway	-2,009	1,01	597	151
Poland	-0,224	-1,287	12	397
Spain	0,483	0,953	60	232
	v1	v2	ctr.g1	ctr.g2
Ancient	0,107	0,181	11	33
Christmas	-0,548	0,242	301	59
Cooking	0,08	-0,106	6	11
Country	-0,44	0,075	194	6
Culture	0,123	0,098	15	10
Dinner	-0,119	-0,226	14	51
Fish	-0,034	-0,309	1	95
Family	-0,134	-0,177	18	31
Feast	0,111	0,035	12	1
Good	-0,002	0,097	0	9
Grandmother	0,132	0,224	17	50
Habit	0,366	0,535	134	286
Healthy	0,055	-0,114	3	13
Holidays	0,136	-0,129	18	17
Home	0,081	0,077	7	6
Home-made	0,021	-0,214	0	46
Kitchen	0,07	-0,024	5	1
Meal	0,007	-0,248	0	62
Natural	0,038	-0,079	1	6
Old	-0,011	0,141	0	20
Old-fashioned	0,076	0,189	6	36
Quality	0,06	0,069	4	5
Recipe	-0,019	-0,183	0	34
Reginal	-0,003	-0,1	0	10
Restaurant	0,06	0,017	4	0
Rural	-0,436	0,22	190	48
Simple	0,032	-0,068	1	5
Tasty	0,191	-0,224	36	50

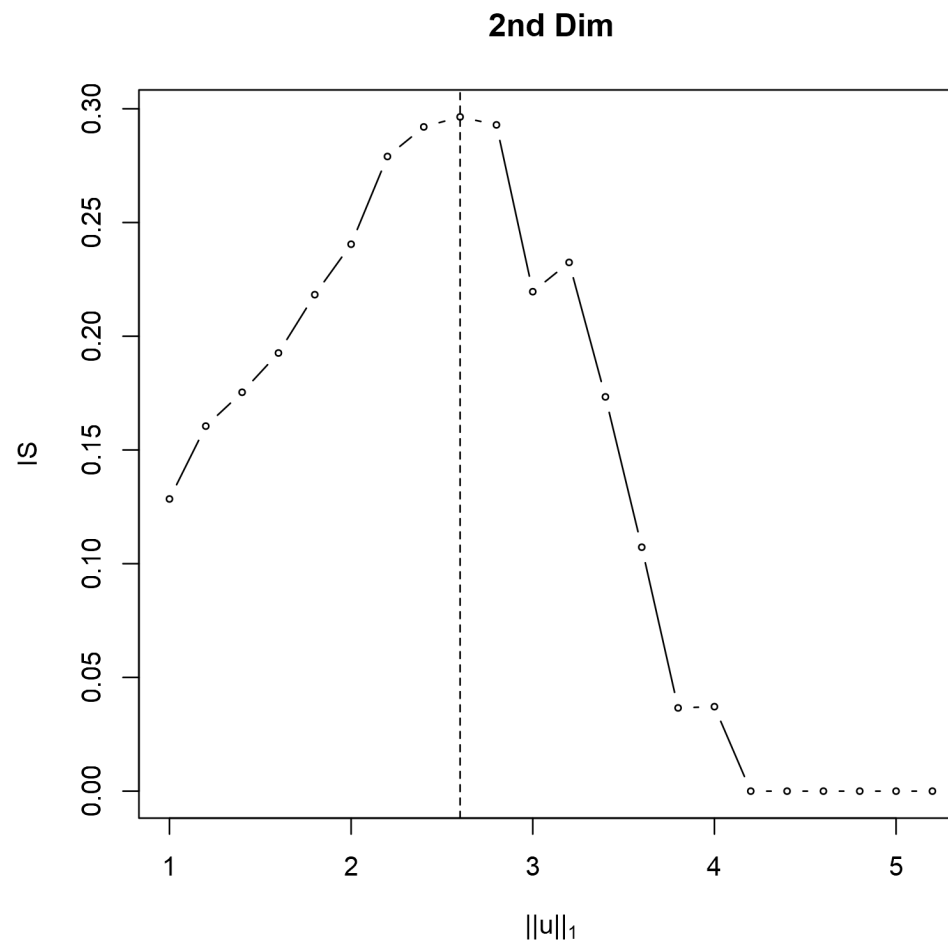
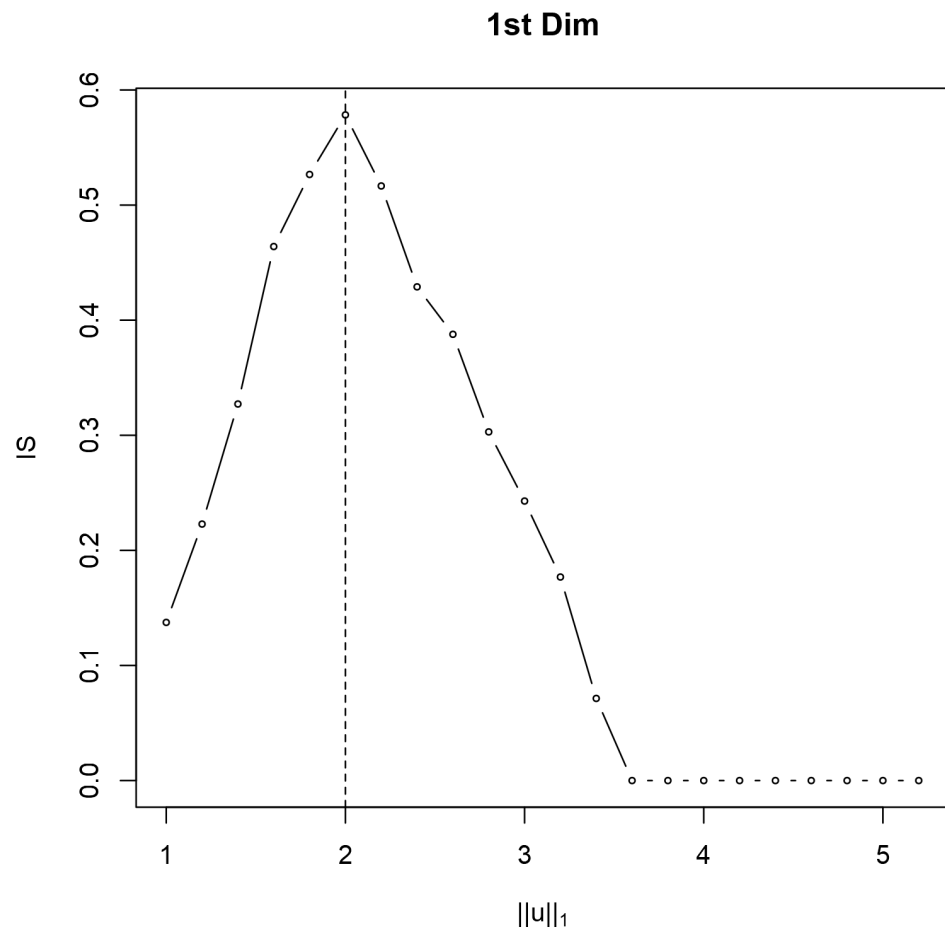
NonSymmetrical Correspondence analysis

Column isometric biplot



Row Sparse NonSymmetrical Correspondence analysis

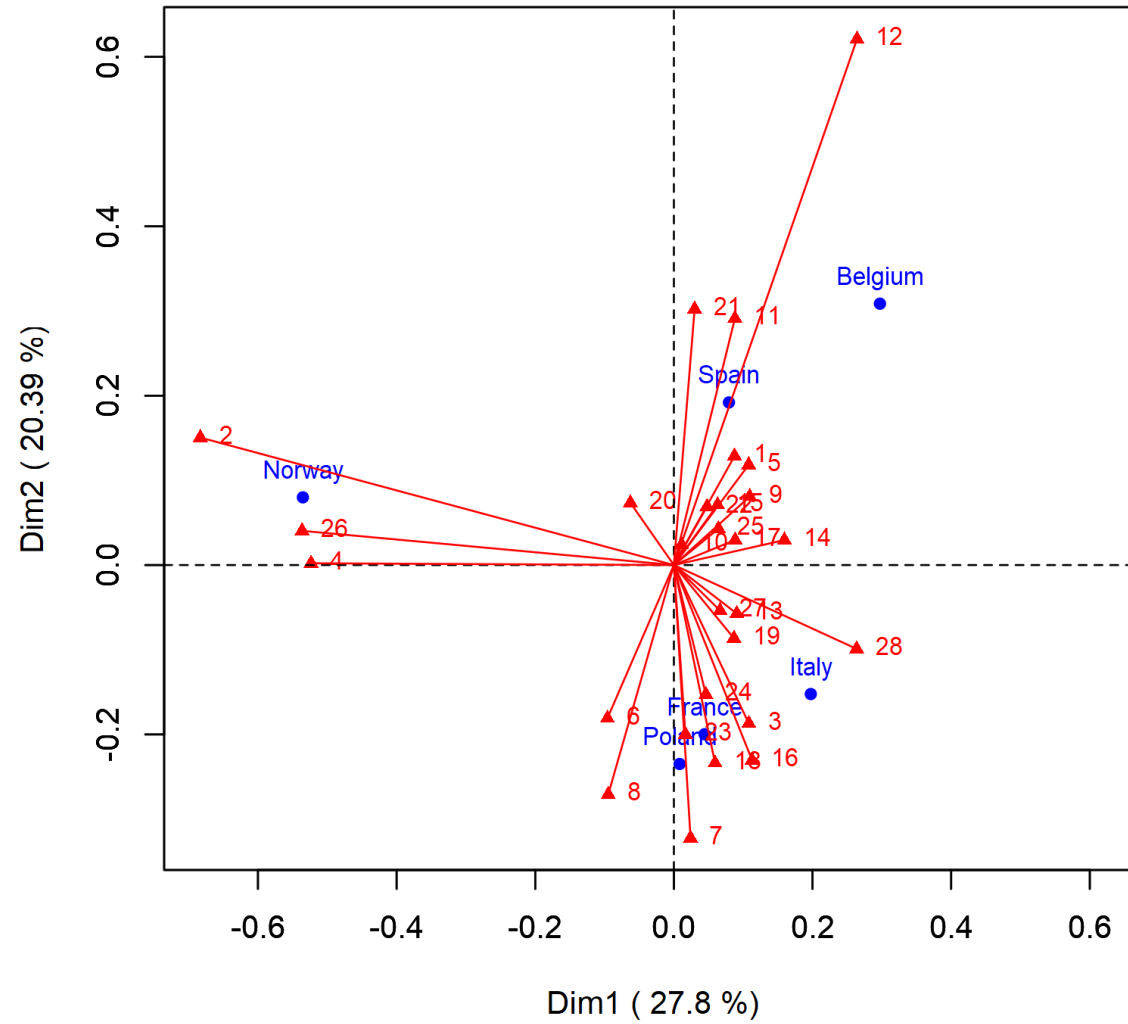
Finding the sparseness parameters



	u1	u2	ctr.f1	ctr.f2
Belgium	1,236	1,355	191	230
France	0,184	-0,978	6	162
Italy	0,821	-0,82	42	42
Norway	-2,225	0,617	733	56
Poland	0,034	-1,133	0	308
Spain	0,33	0,889	28	202
	v1	v2	ctr.g1	ctr.g2
Ancient	0	0,1	0	10
Christmas	0,687	0,108	472	12
Cooking	0	-0,097	0	9
Country	0,48	0,066	230	4
Culture	0	0	0	0
Dinner	0	-0,156	0	24
Fish	0	-0,357	0	127
Family	0	-0,207	0	43
Feast	0	0	0	0
Good	0	0	0	0
Grandmother	0	0,244	0	60
Habit	-0,186	0,766	34	586
Healthy	0	0	0	0
Holidays	-0,016	-0,002	0	0
Home	0	0	0	0
Home-made	0	-0,081	0	7
Kitchen	0	0	0	0
Meal	0	-0,205	0	42
Natural	0	0	0	0
Old	0	0	0	0
Old-fashioned	0	0,22	0	48
Quality	0	0	0	0
Recipe	0	-0,147	0	22
Reginal	0	-0,005	0	0
Restaurant	0	0	0	0
Rural	0,494	0,068	244	5
Simple	0	0	0	0
Tasty	-0,138	-0,019	19	0

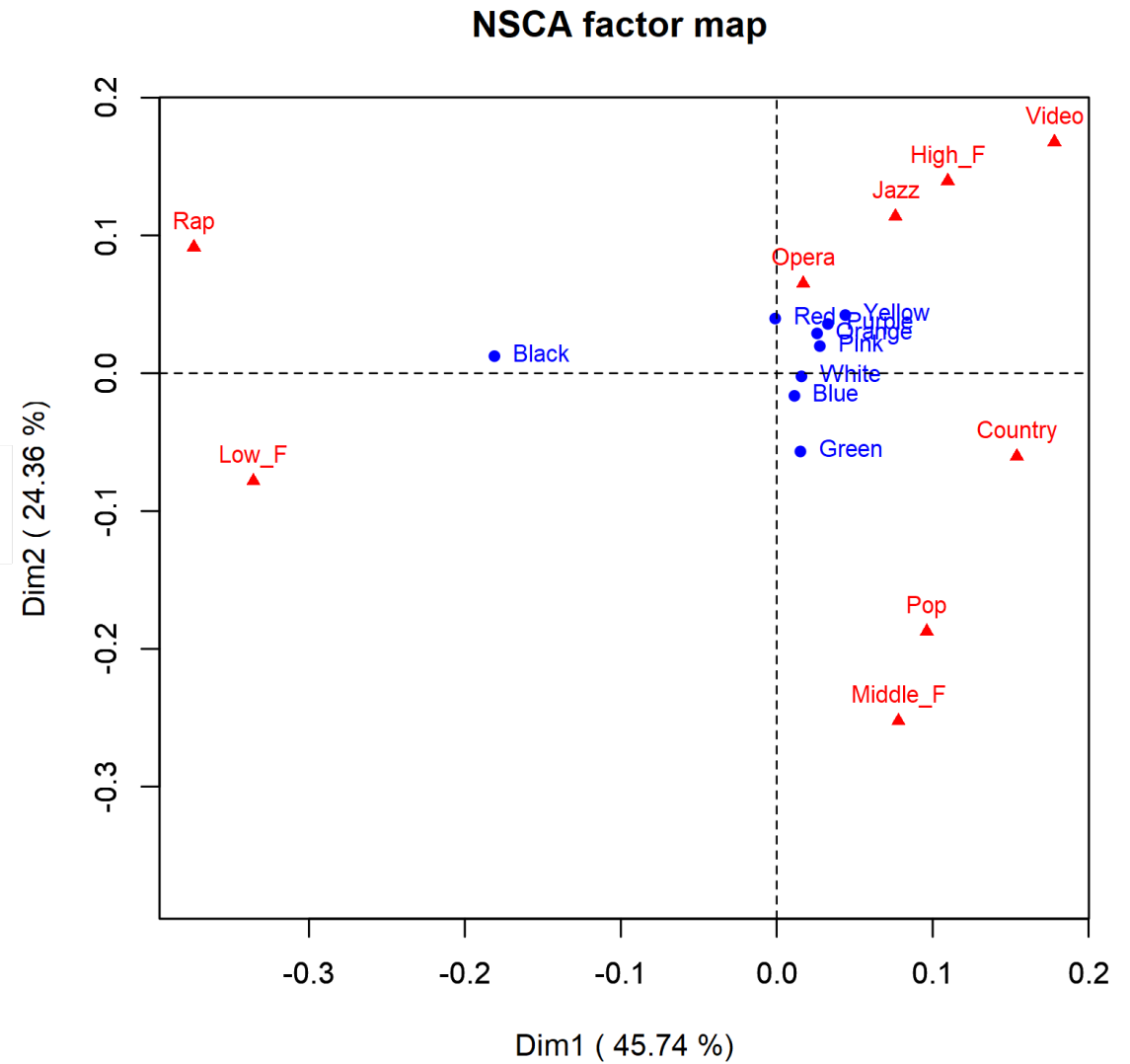
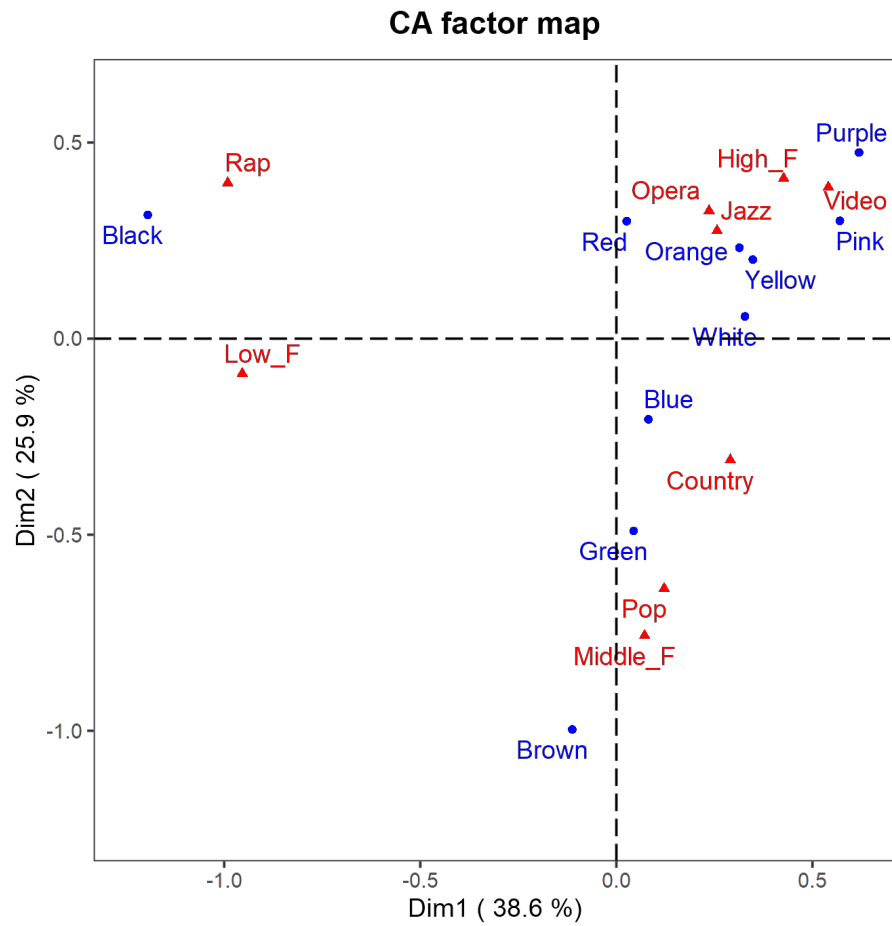
Row Sparse NonSymmetrical Correspondence analysis

Column isometric biplot

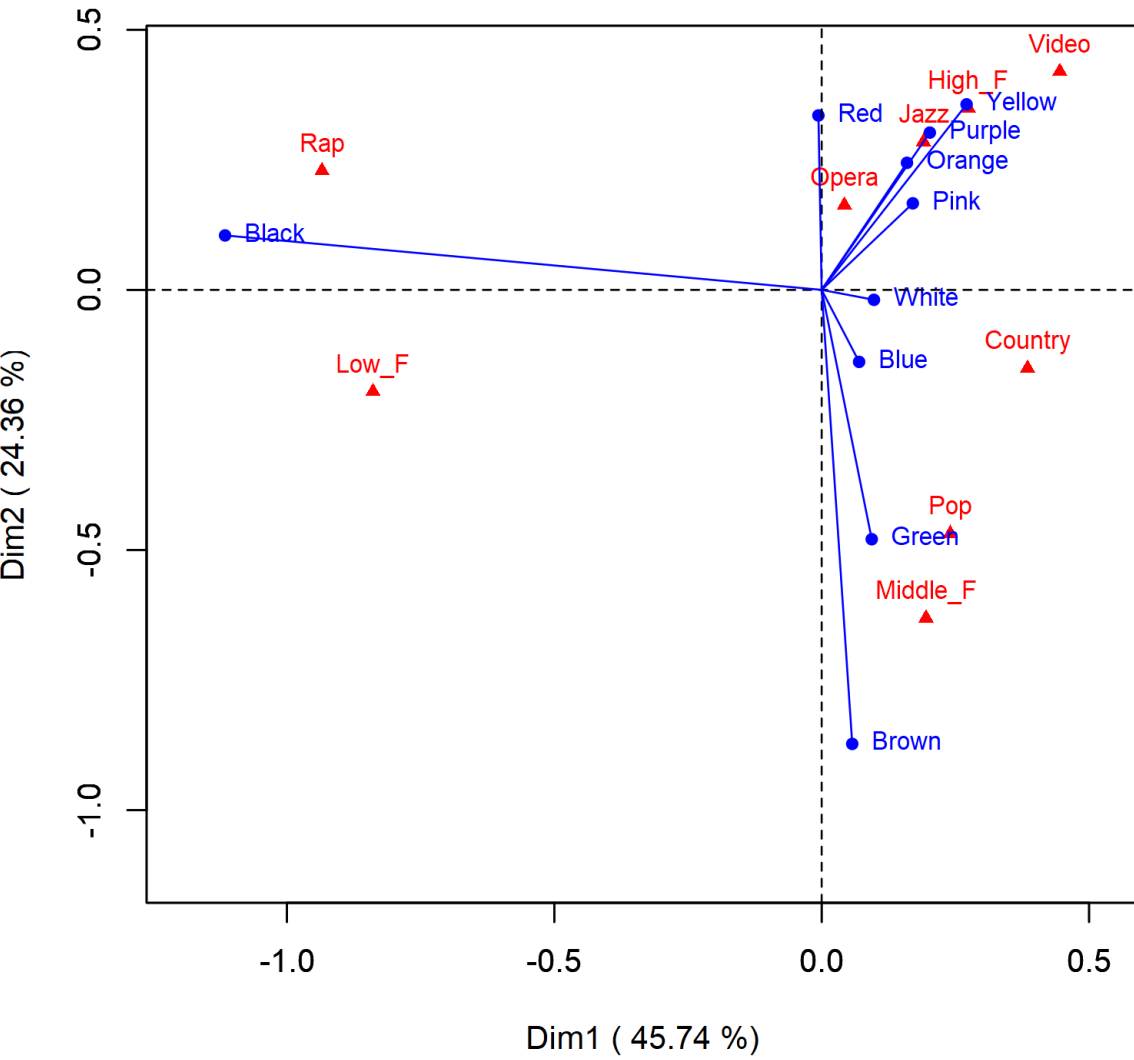


- 48.2 % of τ_b explained instead of 56.2 %
 - Similar display of countries
 - Five significant words highlighted
 - 37 coefficients out of 56 have been put to zero

• 5.2 Colors of music

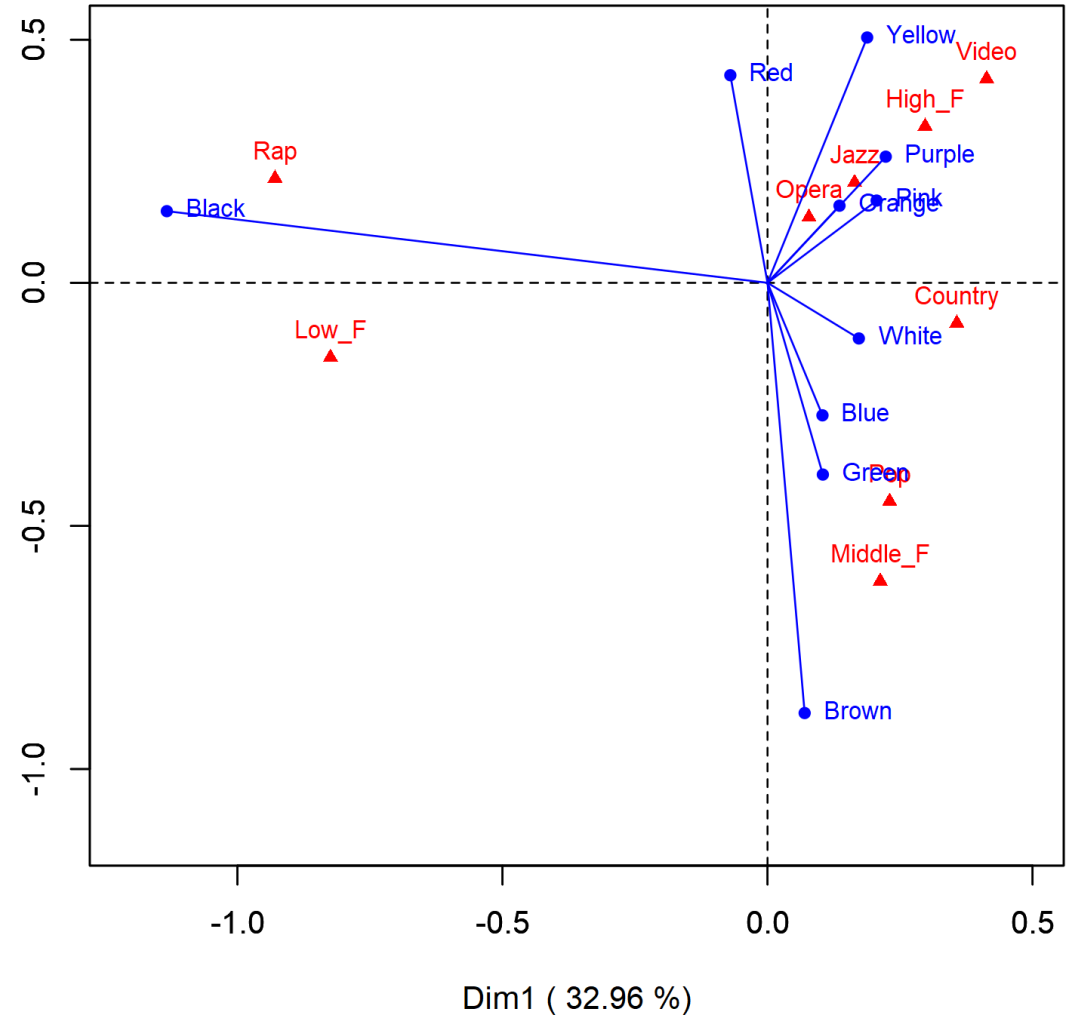


Column isometric biplot



NSCA

Column isometric biplot



Sparse NSCA

	u1	u2	ctr.f1	ctr.f2
Red	0	0,283	0	80
Orange	0,048	-0,002	2	0
Yellow	0,096	0,379	9	144
Green	0,019	-0,243	0	59
Blue	0,018	-0,096	0	9
Purple	0,129	0,084	17	7
White	0,082	-0,003	7	0
Black	-0,976	0,041	952	2
Pink	0,113	-0,005	13	0
Brown	0	-0,836	0	699
	v1	v2	ctr.g1	ctr.g2
Video	0	1,189	0	157
Jazz	0	0	0	0
Country	0	0	0	0
Rap	-2,36	0	619	0
Pop	0	-1,361	0	206
Opera	0	0	0	0
Low_F	-1,852	0	381	0
High_F	0	0,623	0	43
Middle_F	0	-2,312	0	594

Once again:

- Loss of inertia
but
- similar plots
- simpler structure

6. Conclusion and perspectives

- Adding sparsity constraints makes interpretation of NSCA easier, by highlighting the most significant rows or columns
- Choosing the right level of sparsity by IS criterium is generally satisfactory for small size tables. For large tables, prespecifying the number of zeros is often a better solution
- Issues: robustness and stability?

References

- Balbi, S. (1998). Graphical displays in non-symmetrical correspondence analysis. In J. Blasius and M. Greenacre (eds), *Visualization of Categorical Data*. Academic Press, San Diego, 297–309
- Bernard, A., Guinot, C., Saporta, G. (2012). Sparse principal component analysis for multiblock data and its extension to sparse multiple correspondence analysis. In: Colubi, A. et al (eds) *Proceedings of the 20th International Conference on Computational Statistics (COMPSTAT 2012)*. International Association for Statistical Computing, 99–106
- Beh, E.J., Lombardo, R. (2021). Non-Symmetrical Correspondence Analysis. In: *An Introduction to Correspondence Analysis, chapter 3*, John Wiley, 71-97
- Kroonenberg, P. M., Lombardo, R. (1999). Nonsymmetric correspondence analysis: A tool for analysing contingency tables with a dependence structure. *Multivariate Behavioral Research*, 34(3), 367-396.
- Lauro N.C., d'Ambra L. (1984). L'Analyse non symétrique des correspondances, in *Data Analysis and Informatics 3*, E. Diday et al. (eds.), North-Holland, 433-446.
- Lauro, C., Balbi, S. (1999). The analysis of structured qualitative data. *Applied Stochastic Models and Data Analysis*, 15(1), 1–27
- Liu, R., Niang, N., Saporta, G., & Wang, H. (2023). Sparse correspondence analysis for large contingency tables. *Advances in Data Analysis and Classification*, <https://doi.org/10.1007/s11634-022-00531-5>
- Mori, Y., Kuroda, M., Makino, N. (2016). Sparse Multiple Correspondence Analysis. In: Mori, Y., Kuroda, M., Makino, N. (eds) *Nonlinear Principal Component Analysis and Its Applications*. Springer-Verlag, 47–56
- Trendafilov, N. T. (2014). From simple structure to sparse components: a review. *Computational Statistics*, 29, 431-454.
- Witten, D.M., Tibshirani, R., Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation. *Biostatistics*, 10, 515–534