

Au-delà de l'analyse des correspondances multiples :

Illustration sur un exemple de diverses procédures  
graphiques et quantitatives.

Yannick Savina

Jean-Marc Bernard

Laboratoire de Psychologie Environnementale

Université Paris 5 & CNRS – UMR 8069

## Objectifs :

A partir d'un questionnaire

- Procéder à une *analyse des correspondances multiples* (ACM)
- En explorer les résultats
  - d'un point de vue qualitatif (*graphique*)
  - d'un point de vue quantitatif (*décomposition des variances*)

# Le questionnaire

- **Provenance** : Thèse de *Claudia Porto de Lima* (2005)
- **Problématique** : Étudier l'organisation du concept de satisfaction vis à vis de l'espace de travail et ses liens avec les caractéristiques de cet espace
- **Terrain** : Entreprises brésiliennes
- **Échantillon** : 523 personnes
- **Données analysées** :
  - 10 questions relatives à la manière dont les travailleurs ressentent leur espace de travail et s'en satisfont.

Réponses sur une échelle de 1 (tout à fait d'accord) à 4 (pas du tout d'accord)

# Les questions

## Espace

**RNG** : Vous avez suffisamment d'espace pour **ranger** votre matériel

**GRD** : L'espace qu'il y a entre vous et vos collègues est suffisamment **grand**

**EAV** : Dans l'entreprise vous avez un **espace** de travail qui est **à vous**

**PRI** : Votre espace de travail est un lieu **privé**

**CVS** : Vous vous sentez **chez vous** dans votre espace de travail

## Gêne

**BRU** : Il vous est difficile d'accomplir votre travail en raison du **bruit** qu'il y a autour

**ITR** : Pendant votre travail, vous êtes **interrompu** de manière indésirable, sans que vous puissiez en avoir le contrôle

**DIS** : Vous êtes **distrain** visuellement pendant la réalisation de votre travail

**PAR** : C'est vous qui choisissez quand vous voulez **parler** avec les autres

**COT** : Vos **conditions de travail** sont appropriées aux activités que vous réalisez

# Le tableau de données

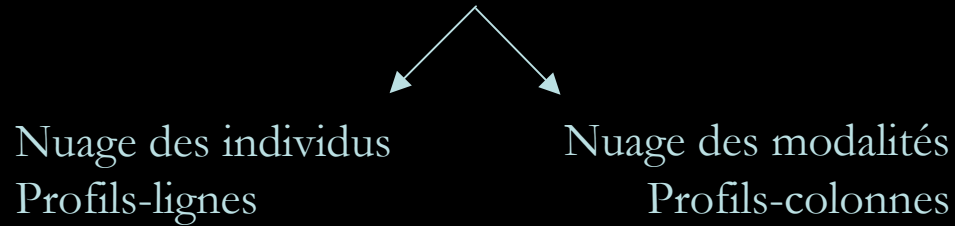
## Forme brute (codage condensé)

Index	RNG	GRD	...	COT
i001	2	2	...	1
i002	2	2	...	4
⋮	⋮	⋮	⋮	⋮
i523	1	1	...	1

## Forme disjonctive complète

Index	RNG				GRD				...				COT			
	1	2	3	4	1	2	3	4	.	.	.	.	1	2	3	4
i001	0	1	0	0	0	1	0	0	.	.	.	.	1	0	0	0
i002	0	1	0	0	0	1	0	0	.	.	.	.	0	0	0	1
⋮		⋮				⋮				⋮				⋮		
⋮		⋮				⋮				⋮				⋮		
⋮		⋮				⋮				⋮				⋮		
i523	1	0	0	0	1	0	0	0	.	.	.	.	1	0	0	0

## Tableau disjonctif complet



**Nuages de points dans un espace muni de la métrique du  $\chi^2$  (ou du  $\phi^2$ ) :**

- 2 individus sont d'autant plus éloignés qu'ils ont choisi des modalités différentes et encore plus si ce sont des modalités rares.
- 2 modalités sont d'autant plus éloignées que le nombre d'individus ayant choisi l'une et pas l'autre (et vice-versa) est grand.

**L'ACM consiste en la recherche simultanée des axes principaux des deux nuages, afin d'en obtenir une représentation simplifiée.**

# L'ACM du tableau disjonctif complet fournit les résultats suivants :

HISTOGRAMME DES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENTAGE	POURCENTAGE CUMULE	
1	0.3612	12.04	12.04	*****
2	0.2440	8.13	20.17	*****
3	0.1483	4.94	25.12	*****
4	0.1419	4.73	29.85	*****
5	0.1329	4.43	34.27	*****
6	0.1233	4.11	38.39	*****
7	0.1225	4.08	42.47	*****
8	0.1153	3.84	46.31	*****
9	0.1098	3.66	49.97	*****
10	0.1037	3.46	53.43	*****
11	0.1011	3.37	56.80	*****
12	0.1007	3.36	60.16	*****

On conservera les **2 premiers axes**.

Taux modifiés (Benzécri, 1979) pour ces axes : 75 % et 21 %.

# Le nuage des modalités



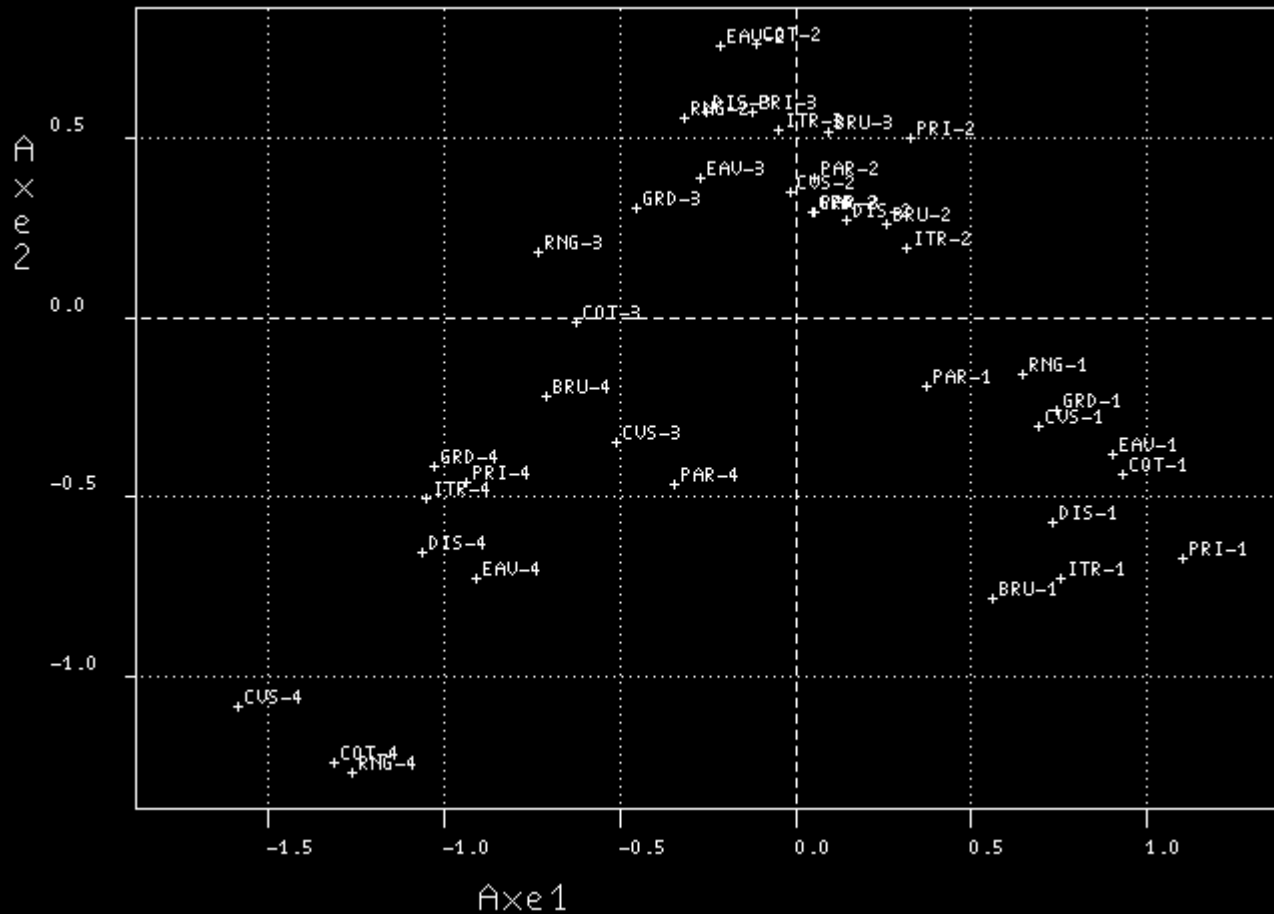
# Aides à l'interprétation du nuage des modalités

MODALITES			COORDONNEES		CONTRIBUTIONS		COSINUS CARRES	
- LIBELLE	P.REL	DISTO	1	2	1	2	1	2
1 . RNG								
- rng1	4.78	1.09	0.65	-0.16	5.5	0.5	0.38	0.02
- rng2	2.93	2.42	-0.32	0.56	0.8	3.7	0.04	0.13
- rng3	1.40	6.16	-0.73	0.18	2.1	0.2	0.09	0.01
- rng4	0.90	10.13	-1.26	-1.26	3.9	5.9	0.16	0.16
+CONTRIBUTION CUMULEE =					12.4	10.3	+-----	
2 . GRD								
- grd1	3.38	1.95	0.74	-0.26	5.2	0.9	0.28	0.03
- grd2	2.79	2.58	0.05	0.30	0.0	1.0	0.00	0.03
- grd3	2.26	3.43	-0.46	0.31	1.3	0.9	0.06	0.03
- grd4	1.57	5.38	-1.03	-0.41	4.6	1.1	0.20	0.03
+CONTRIBUTION CUMULEE =					11.1	3.9	+-----	
10 . COT								
- cot1	3.29	2.04	0.93	-0.44	7.8	2.6	0.42	0.09
- cot2	3.54	1.83	-0.11	0.76	0.1	8.5	0.01	0.32
- cot3	2.18	3.59	-0.62	-0.01	2.3	0.0	0.11	0.00
- cot4	0.99	9.06	-1.31	-1.24	4.7	6.3	0.19	0.17
+CONTRIBUTION CUMULEE =					15.0	17.4	+-----	

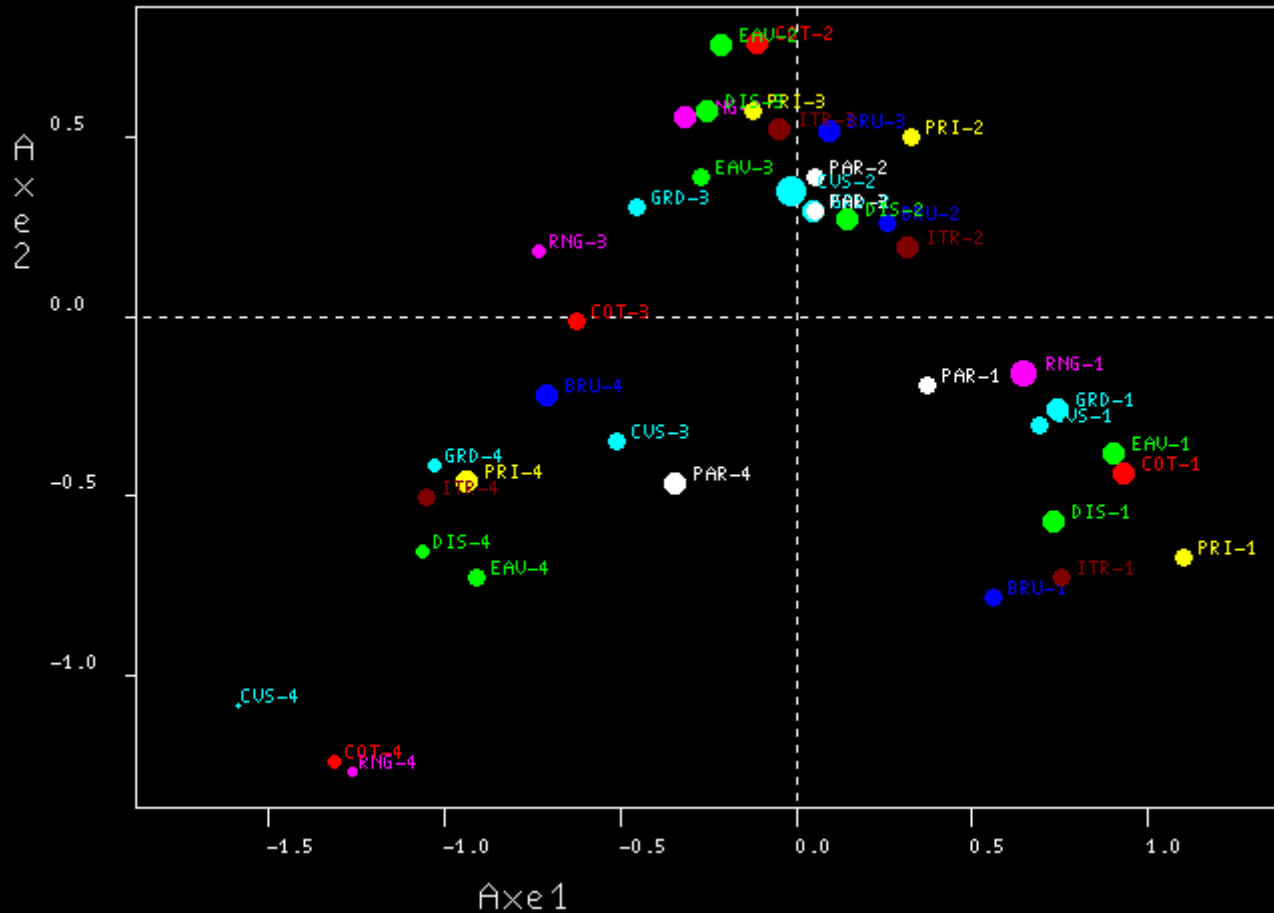
## Du fichier des modalités

Modalité	Poids	Axe 1	Axe 2	K	Q	R
rng1	250	0.65	-0.16	1	1	1
rng2	153	-0.32	0.56	2	1	2
rng3	73	-0.73	0.18	3	1	3
rng4	47	-1.26	-1.26	4	1	4
grd1	177	0.74	-0.26	5	2	1
grd2	146	0.05	0.30	6	2	2
grd3	118	-0.46	0.31	7	2	3
grd4	82	-1.03	-0.41	8	2	4
...						
cot1	172	0.93	-0.44	9	3	1
cot2	185	-0.11	0.76	10	3	2
cot3	114	-0.62	-0.01	11	3	3
cot4	52	-1.31	-1.24	12	3	4

vers son exploitation graphique ...

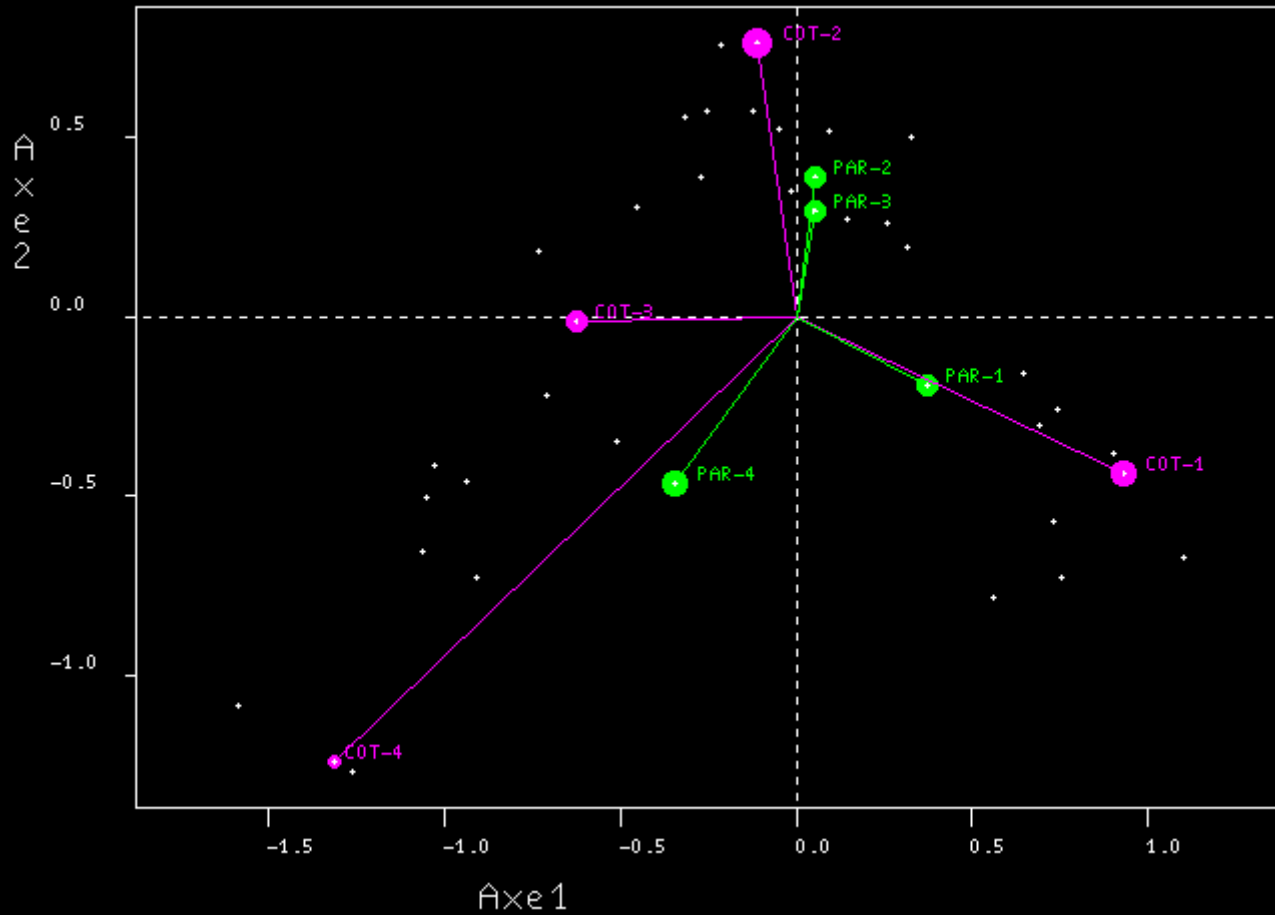


**Premier graphique avec axes solidarisés puis quelques habillages supplémentaires...**

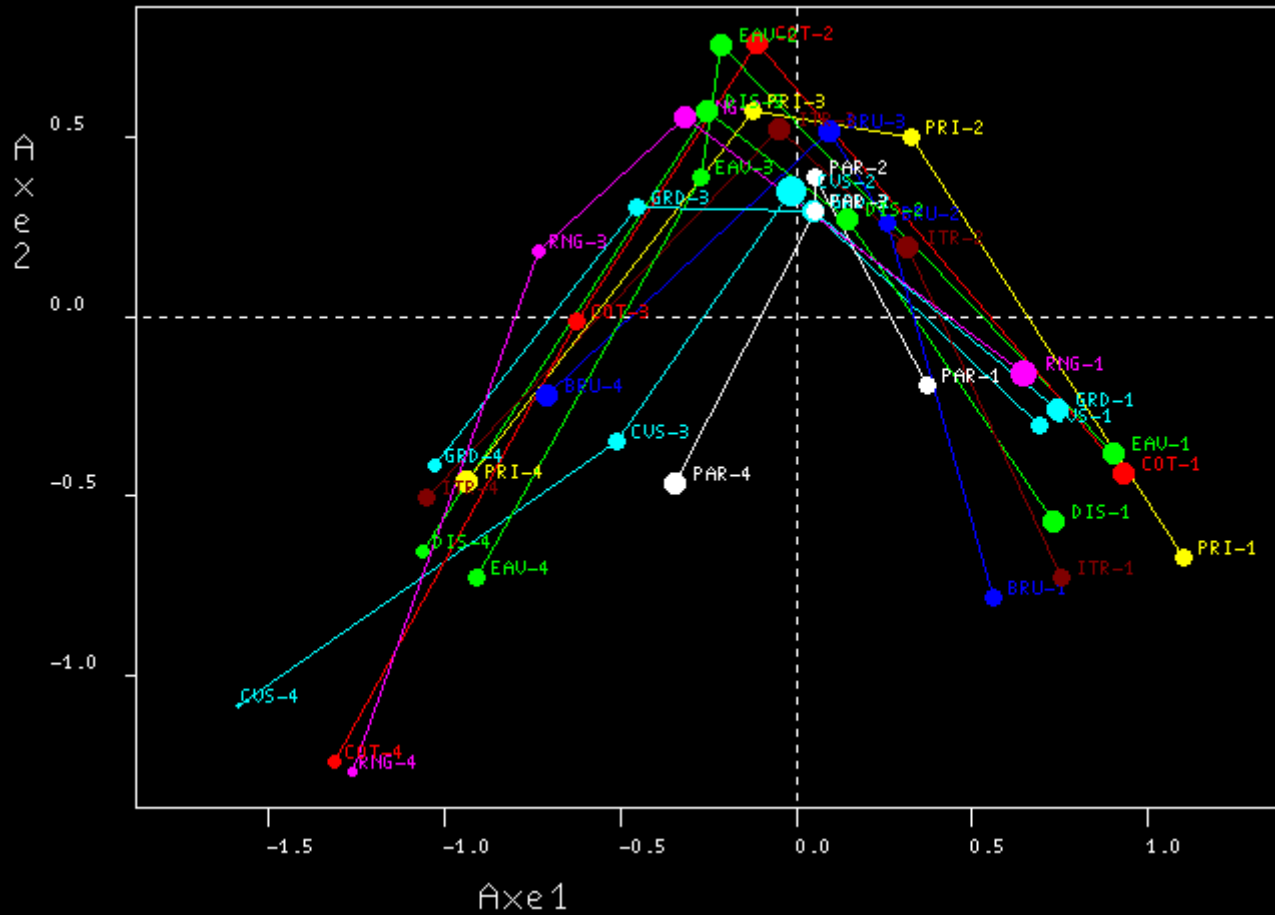


## Pondération et coloriage des questions (facteur Q)

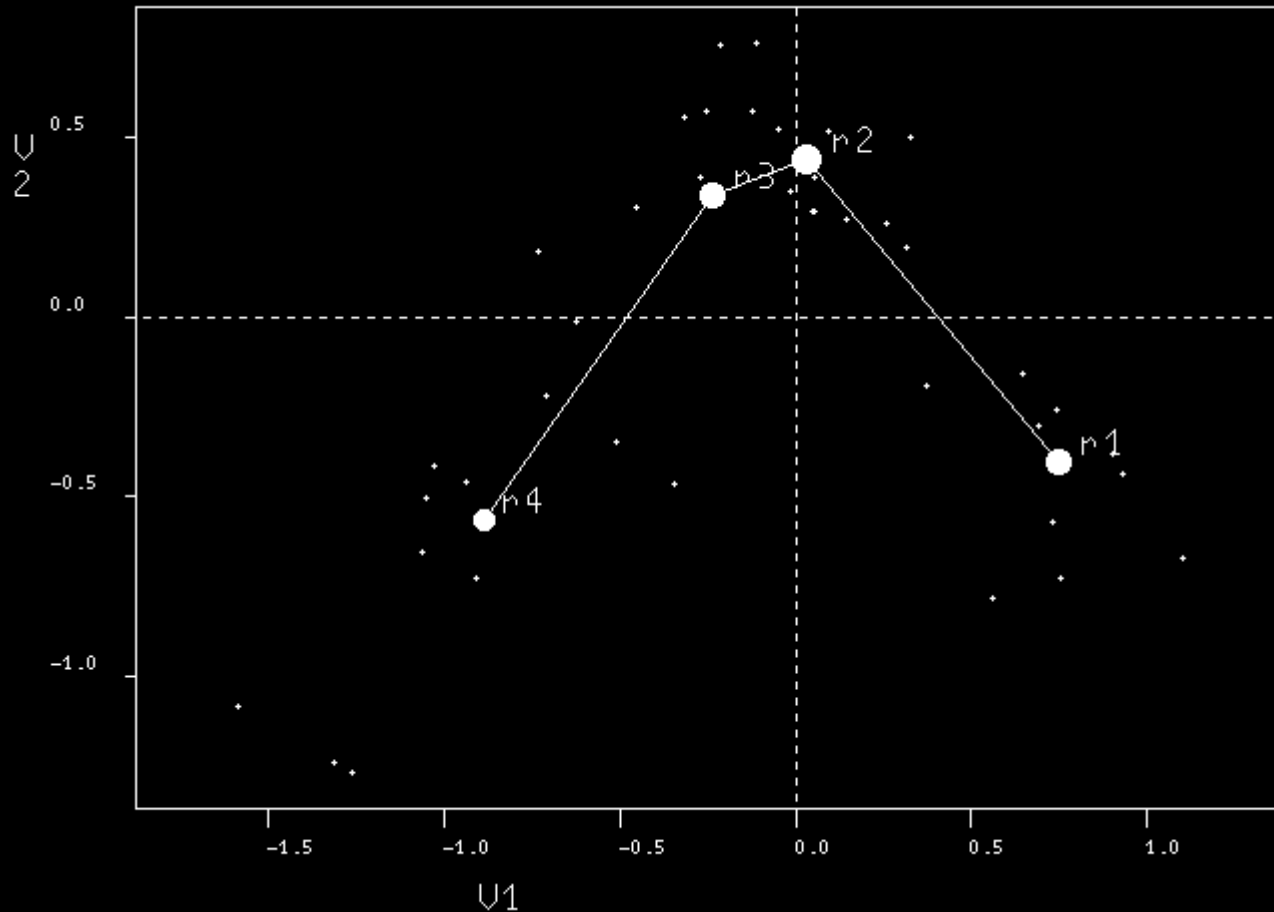
2 Du mal à isoler une question, à comparer des questions ?



## Sous-nuages des questions COT et PAR

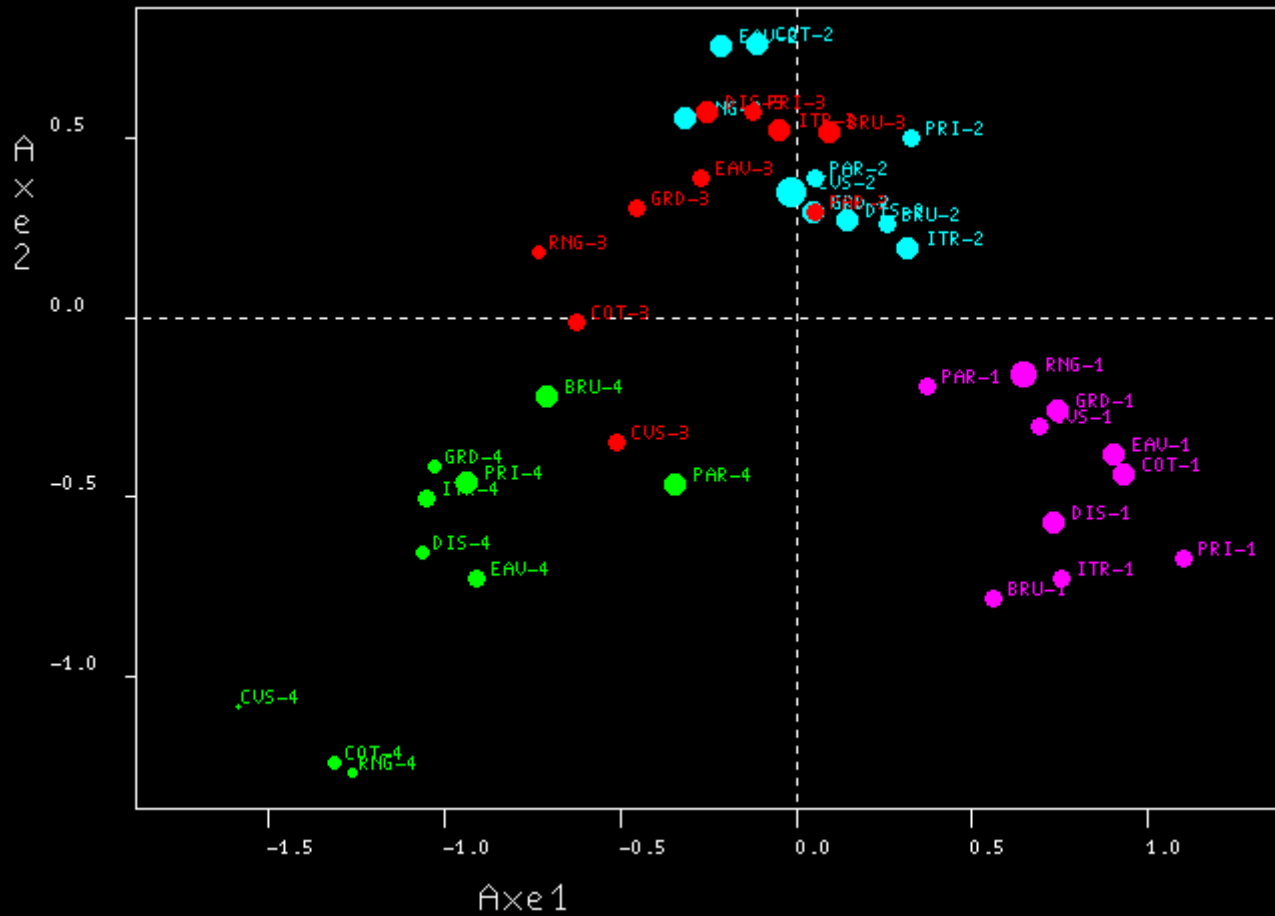


De l'ordre dans l'ordre grâce au facteur R (réponses)



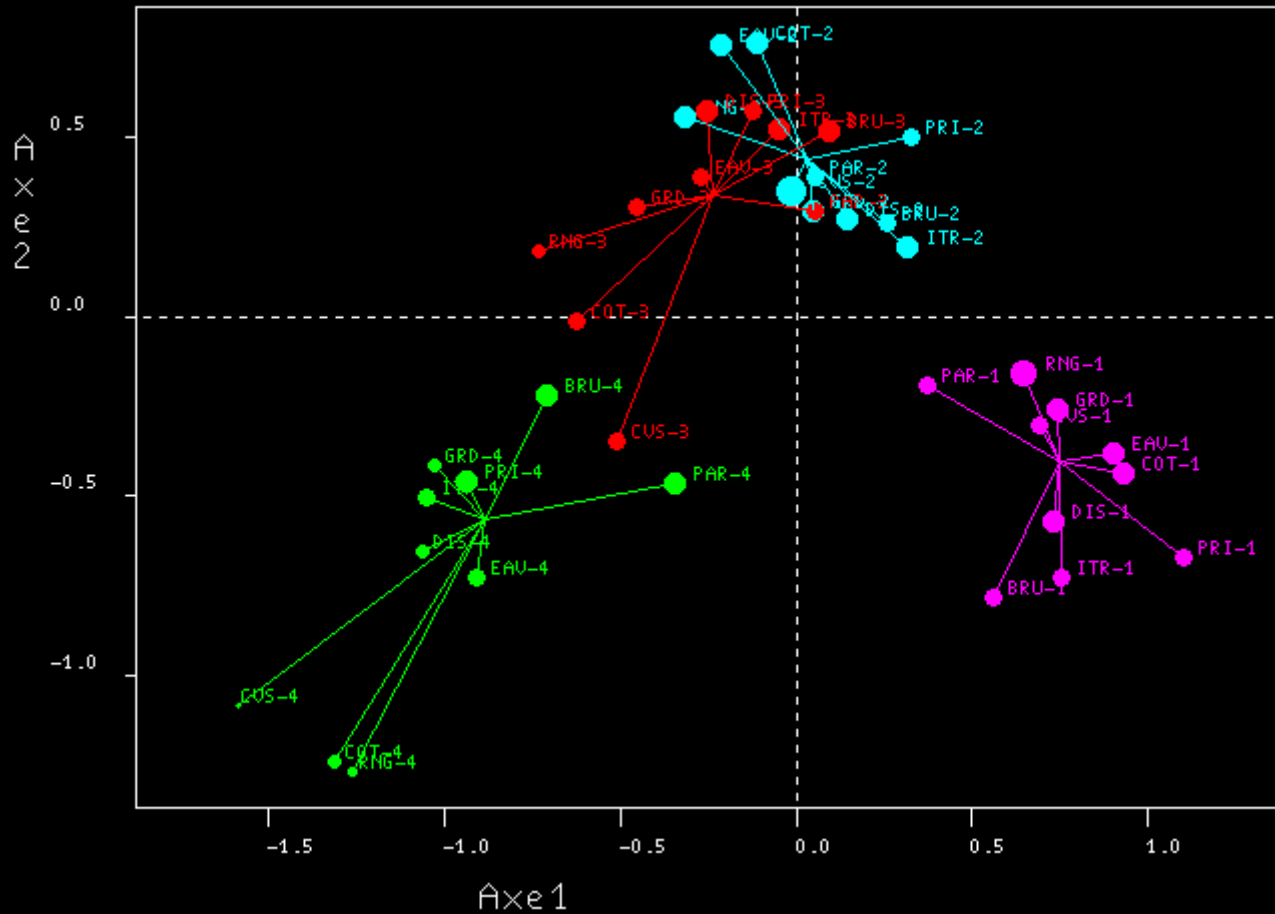
**L'ordre sous-jacent est celui du facteur Réponses**

**Mais ces points moyens sont-ils  
un bon résumé du nuage dans son ensemble ?**



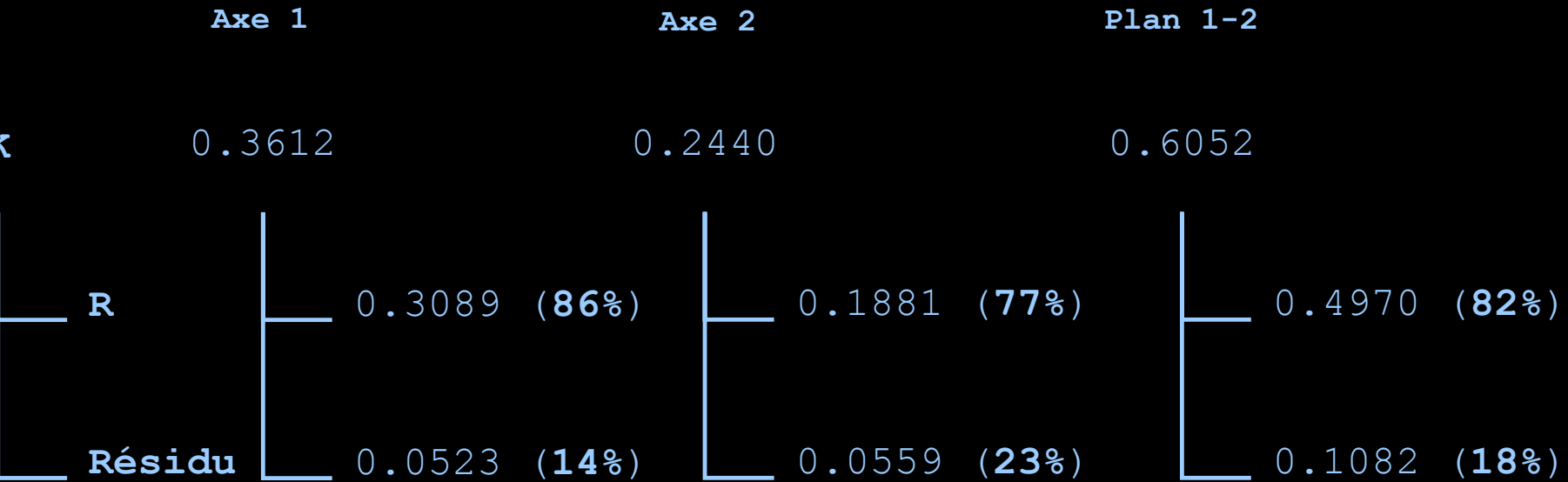
## Sous-nuages des modalités du facteur Réponses





Sous-nuages des modalités du facteur Réponses  
 Et  
 Points moyens

# Quantifier le phénomène



# Quantifier le phénomène



# Le nuage des individus

# Aides à l'interprétation du nuage des individus

INDIVIDUS			COORDONNEES		CONTRIBUTIONS		COSINUS CARRES	
IDEN	P.REL	DISTO	1	2	1	2	1	2
1	0.19	2.75	-0.03	0.62	0.0	0.3	0.00	0.14
2	0.19	3.87	-0.37	-0.05	0.1	0.0	0.04	0.00
3	0.19	2.92	0.68	-0.27	0.2	0.1	0.16	0.02

etc durant encore 520 lignes

**Conséquence : Nuage rarement exploité**

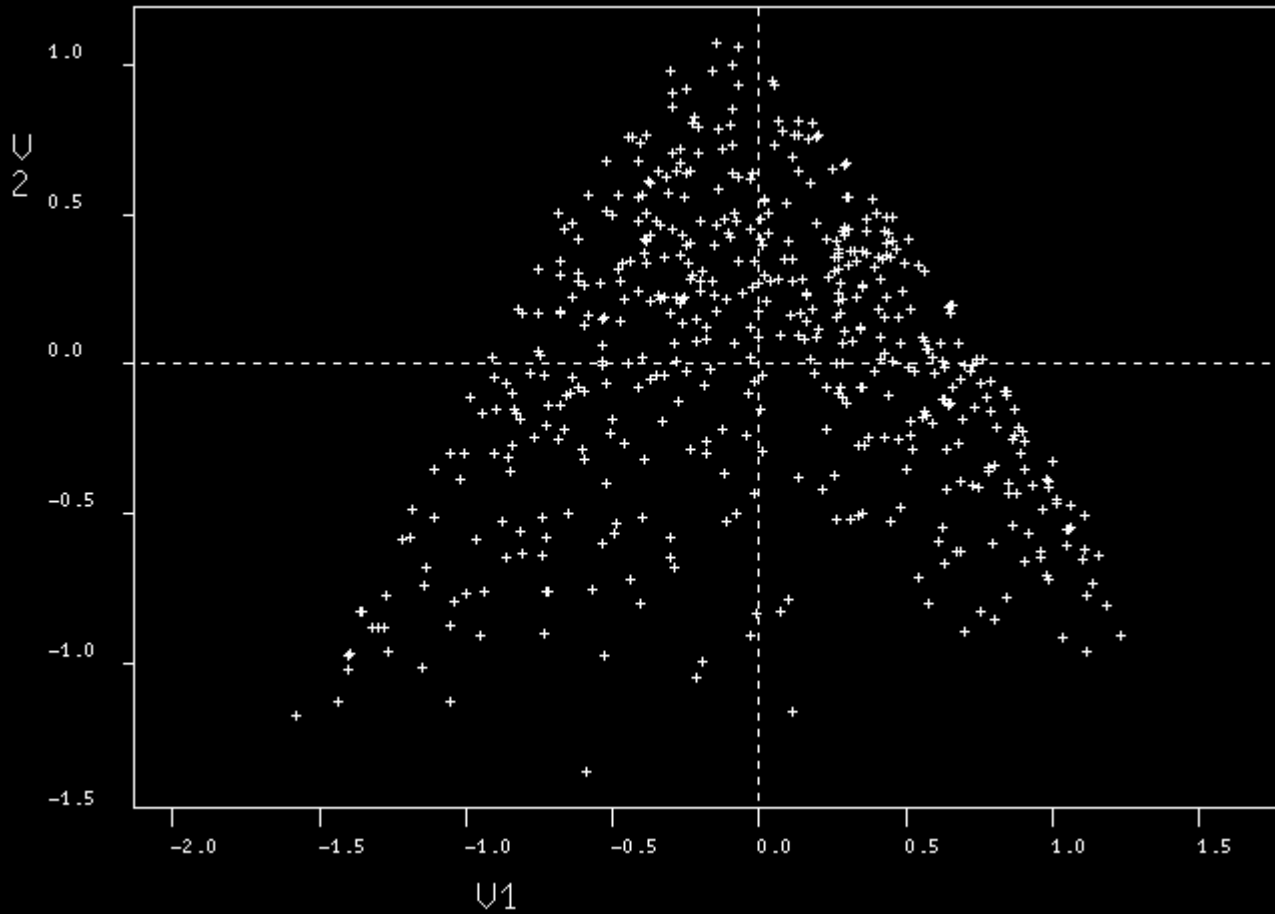
**Et pourtant ce nuage est porteur de toute l'information.**

**Il permet d'aller beaucoup plus loin  
que les analyses sur le nuage des modalités.**

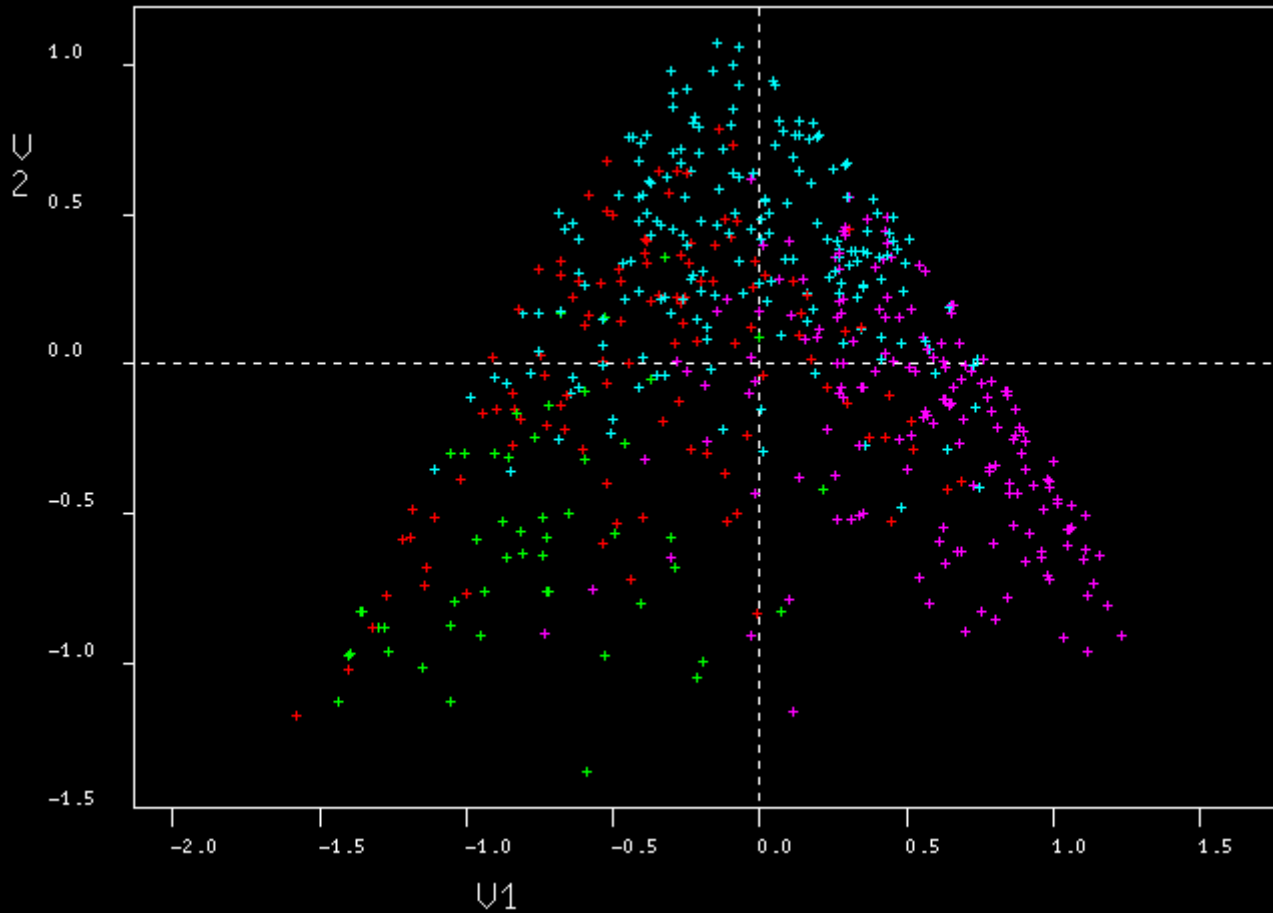
# Du fichier des individus

Ind	Poids	Axe 1	Axe 2	I	T	RNG	GRD	COT	EAV	PRI	BRU	PAR	ITR	DIS	CVS	A	B	C	D
i001	1	-0.03	0.62	1	1	2	2	1	2	3	2	3	3	3	3	2	2	3	3
i002	1	-0.37	-0.05	2	1	2	2	4	3	1	3	2	3	4	3	2	3	4	4
i003	1	0.68	-0.27	3	1	1	2	1	3	2	1	1	1	1	2	2	2	1	1
i004	1	-0.06	0.23	4	1	1	2	2	4	3	4	1	2	2	2	3	3	2	2
i005	1	0.29	0.44	5	1	2	1	1	3	3	2	3	2	2	2	3	3	1	1
i006	1	0.45	0.35	6	1	2	2	1	1	2	3	2	3	1	2	1	1	3	3
i007	1	0.40	0.50	7	1	1	1	2	2	1	3	2	3	2	2	1	1	3	3
...																			
i523	1	0.88	-0.43	523	1	1	1	1	1	1	2	4	2	1	2	1	1	1	1
+																			
i600	1	1.23	-0.91	524	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
i601	1	0.08	0.88	525	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1
i602	1	-0.48	0.62	526	2	3	3	3	3	3	3	3	3	3	3	5	6	4	4
i603	1	-1.70	-1.42	527	2	4	4	4	4	4	4	4	4	4	4	5	6	4	4

vers son exploitation graphique ...

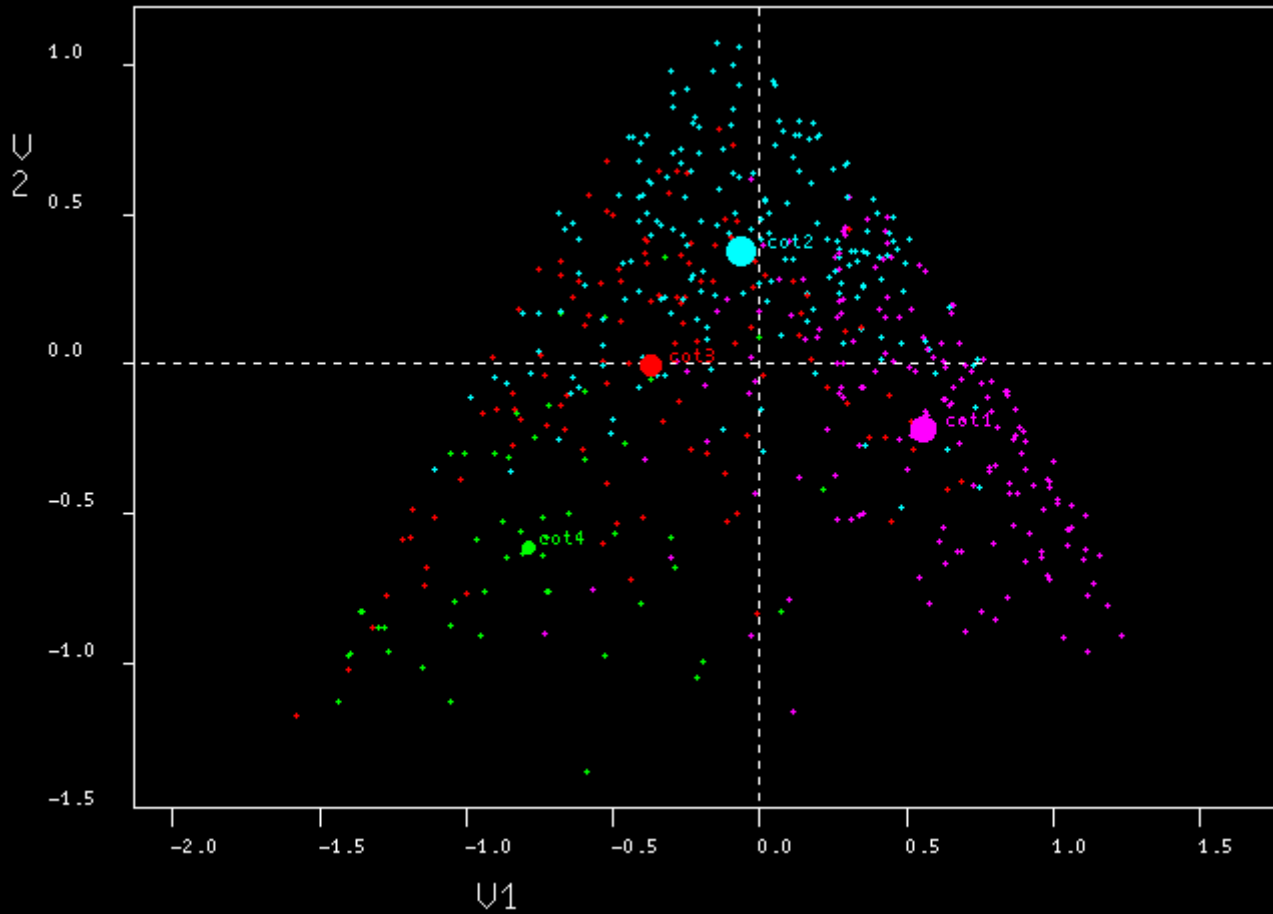


**La base pour aller au-delà**

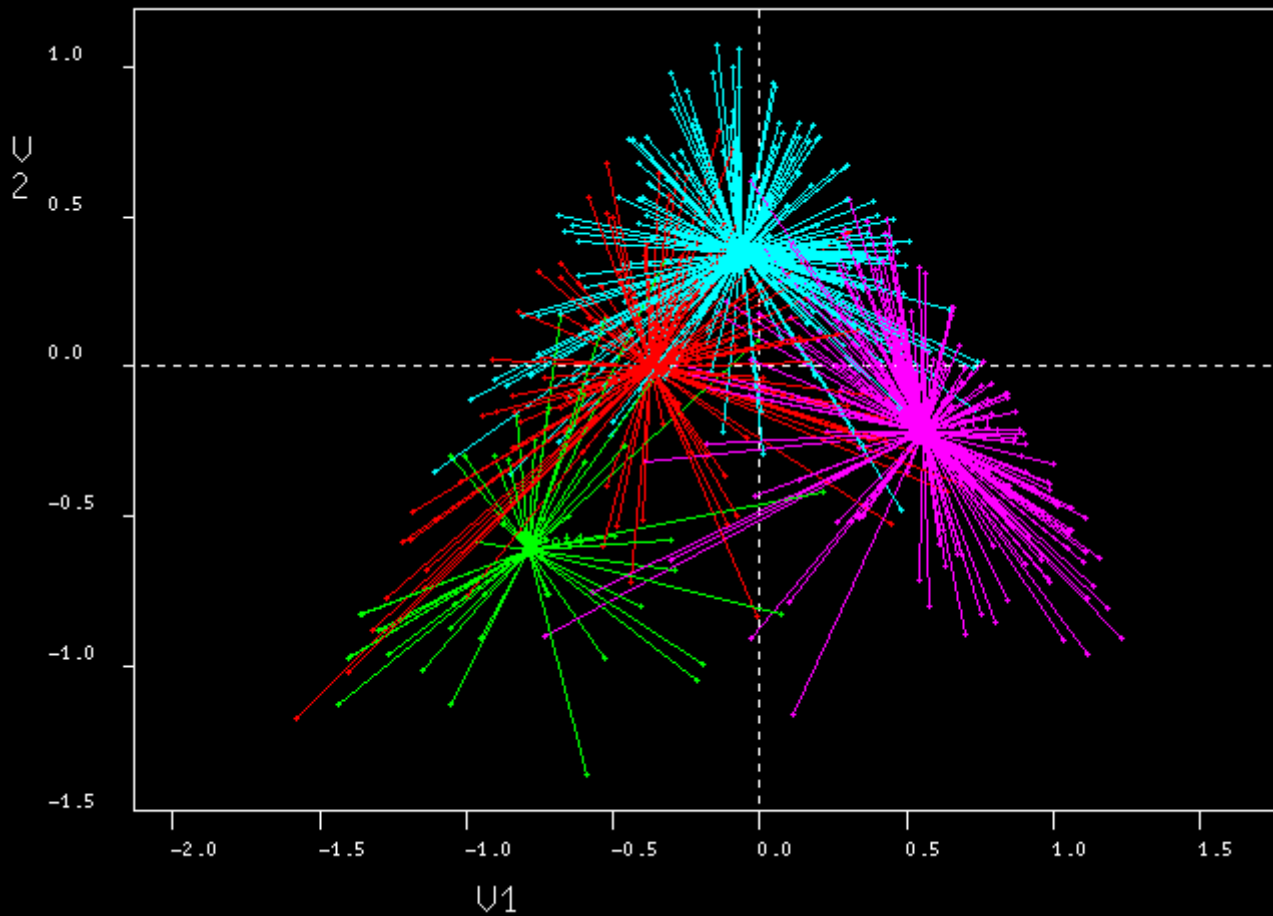


**Exploration du nuage à la lumière d'un facteur :**  
**COT**

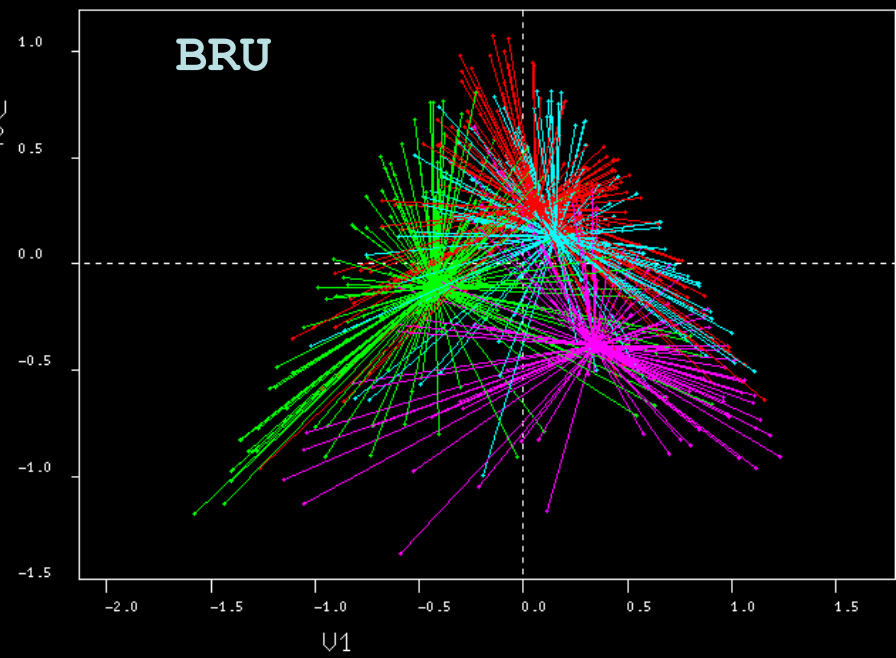
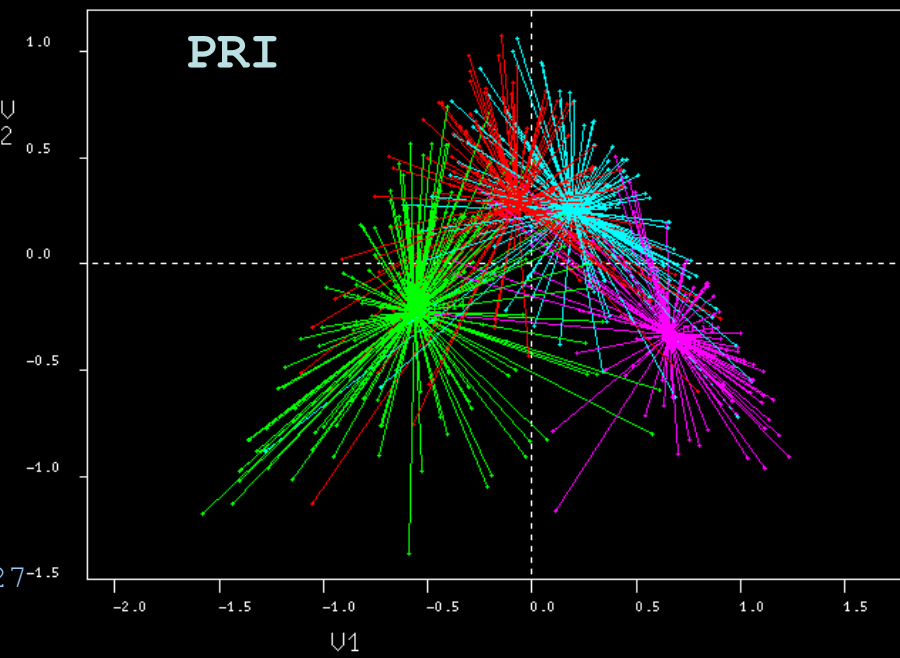
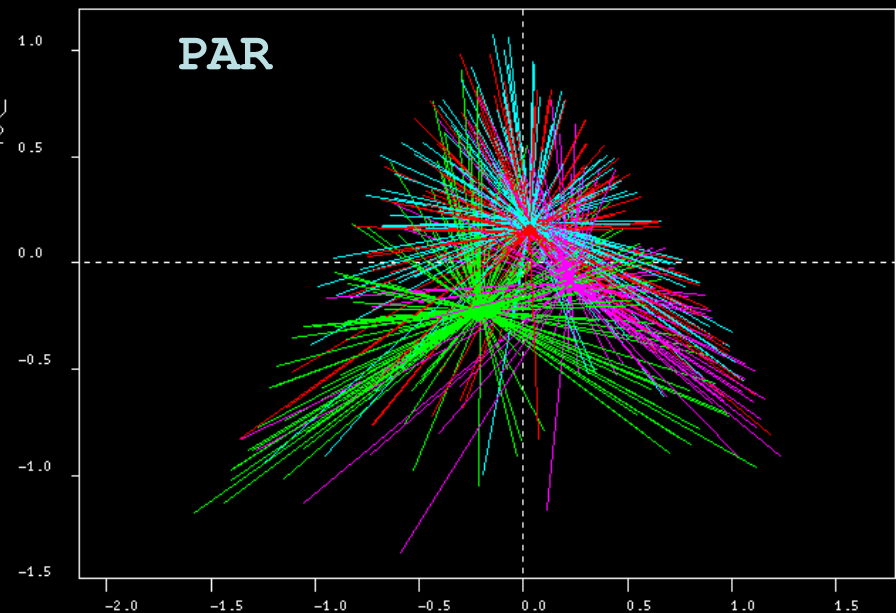
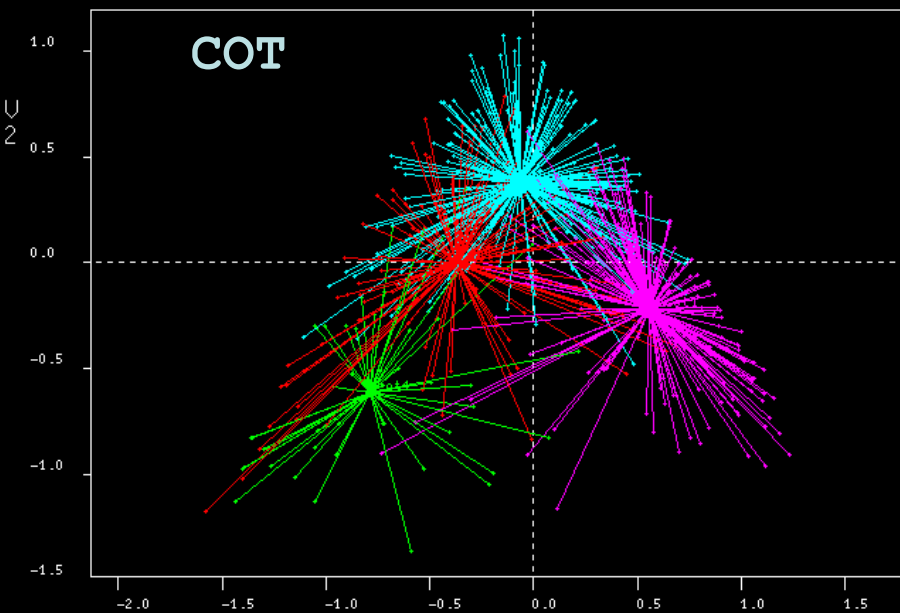




## Sous-nuages et points moyens

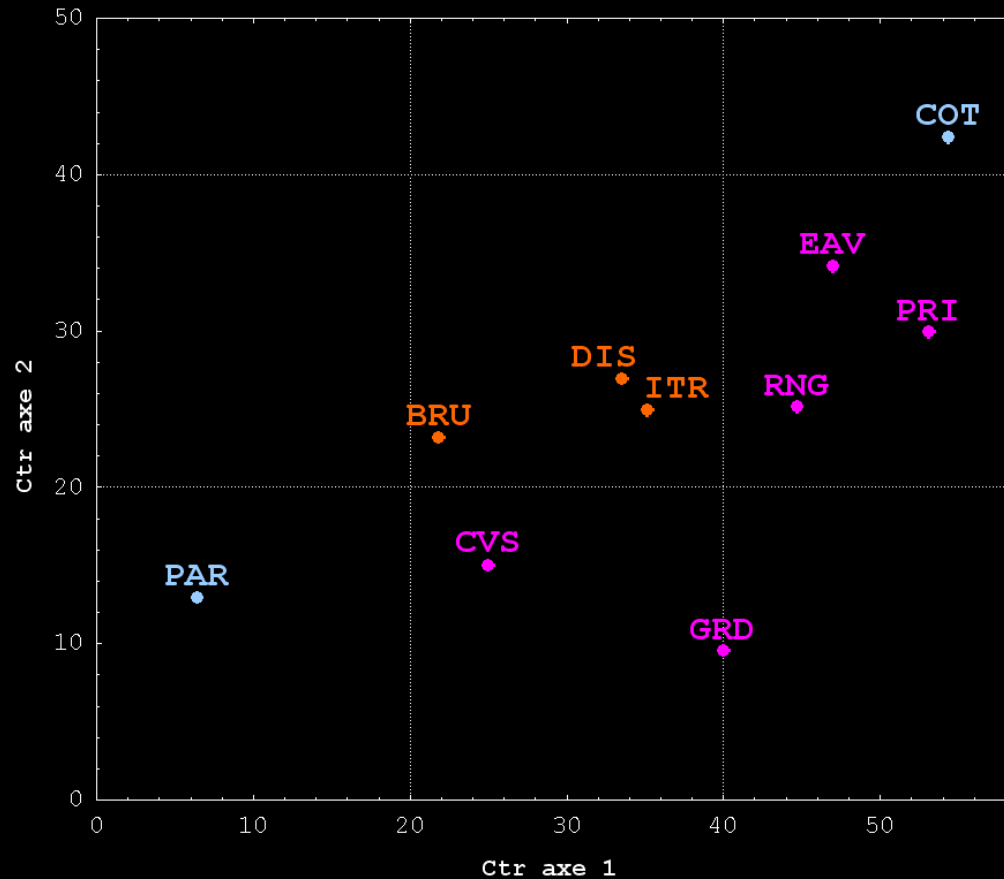


**Sous-nuages et points moyens,  
le tout relié**



# Contributions des facteurs aux axes

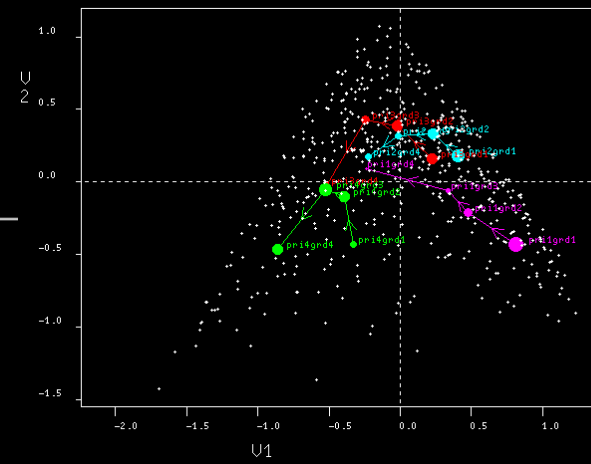
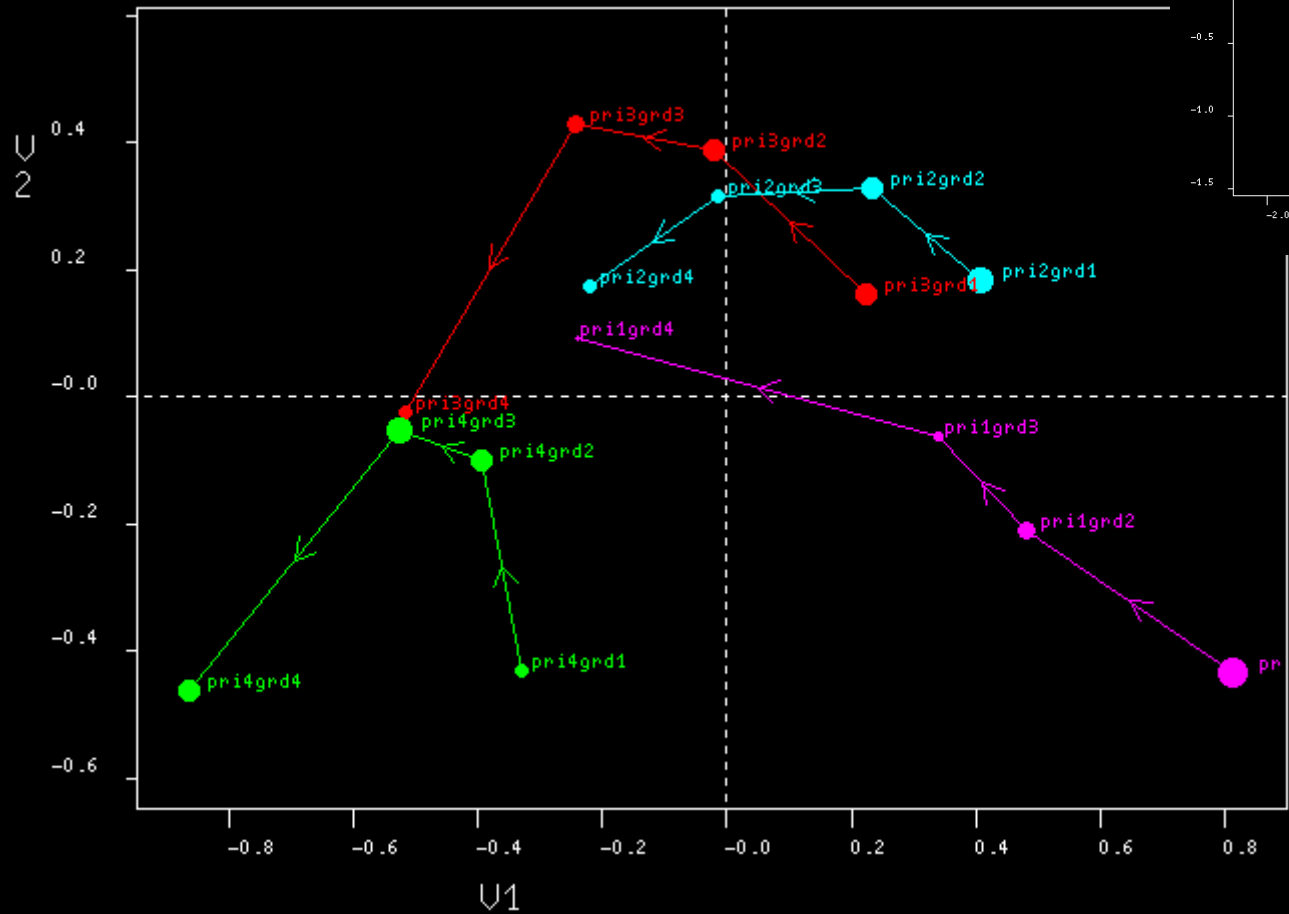
Source	Axe 1		Axe 2	
	Cta	Ctr	Cta	Ctr
	0.3612		0.2440	
ANG	0.1615	44.7%	0.0612	25.1%
ARD	0.1446	40.0%	0.0232	9.5%
AV	0.1695	46.9%	0.0832	34.1%
PRI	0.1920	53.2%	0.0729	29.9%
VS	0.0905	25.1%	0.0365	15.0%
BRU	0.0790	21.9%	0.0566	23.2%
TR	0.1272	35.2%	0.0608	24.9%
DIS	0.1212	33.6%	0.0658	27.0%
COT	0.1962	54.3%	0.1033	42.3%
PAR	0.0231	6.4%	0.0316	13.0%



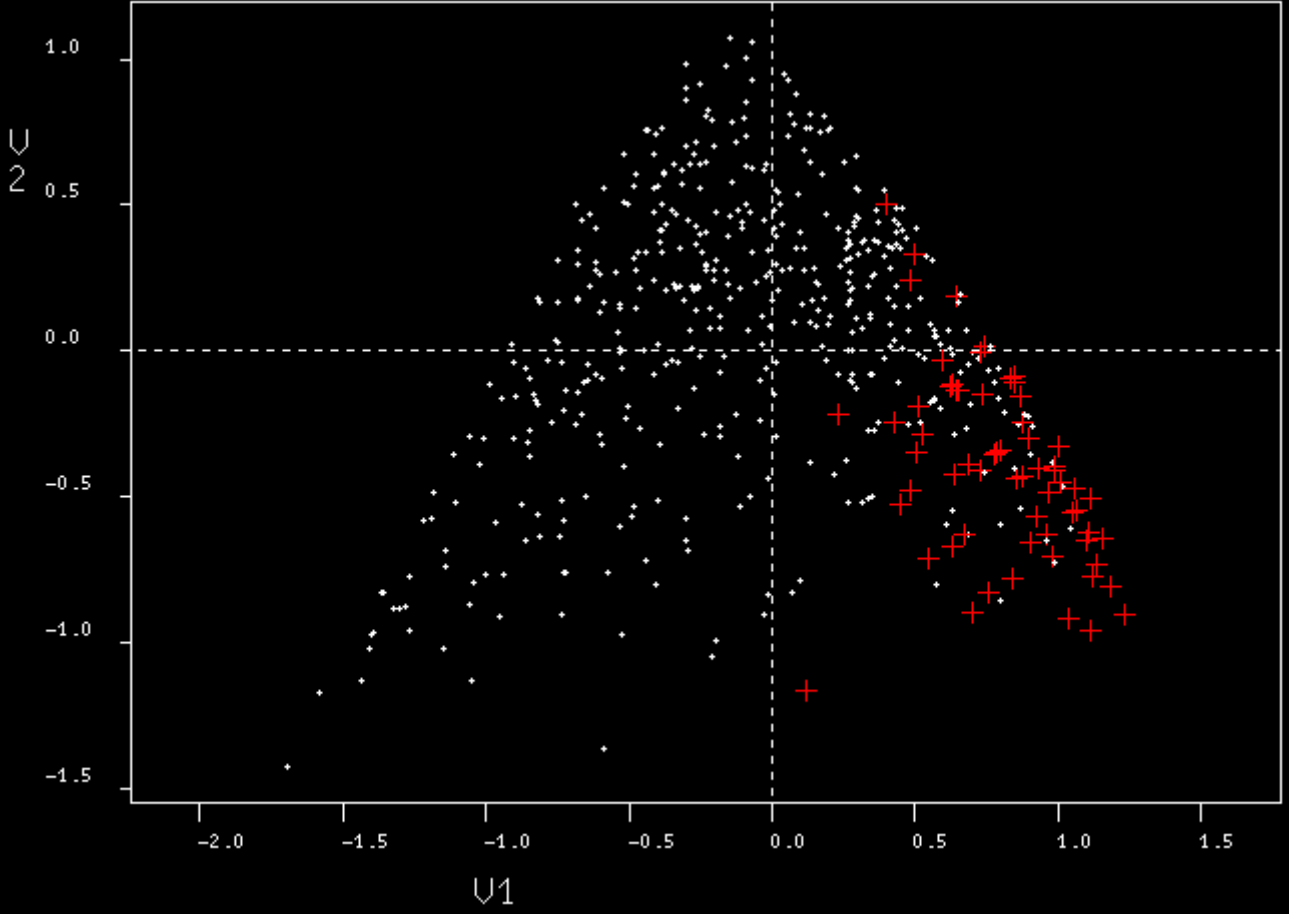
# Décomposition de la variance



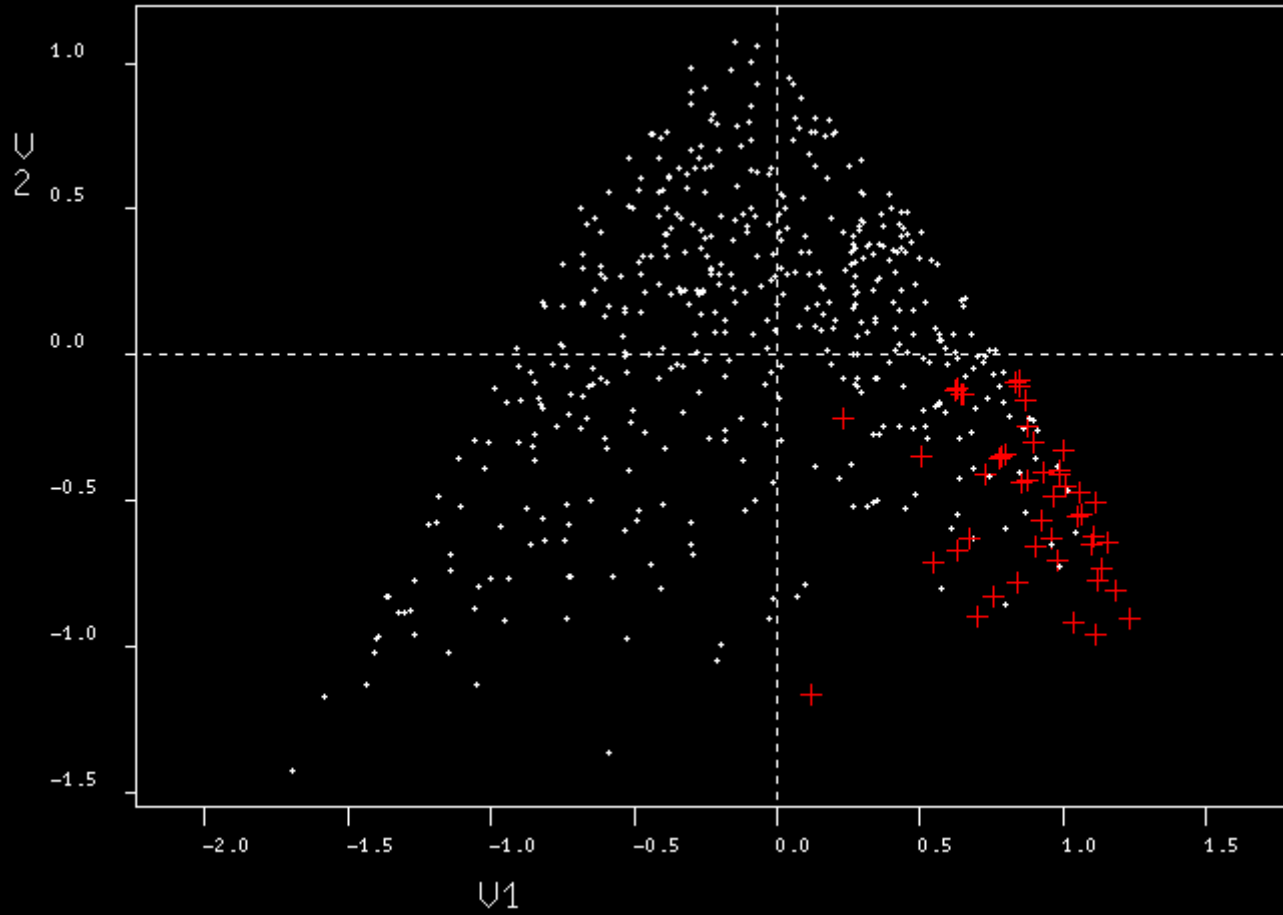
# Étudier la conjonction de questions : PRI et GRD



# Situer les individus qui ont répondu **prilgrd1**

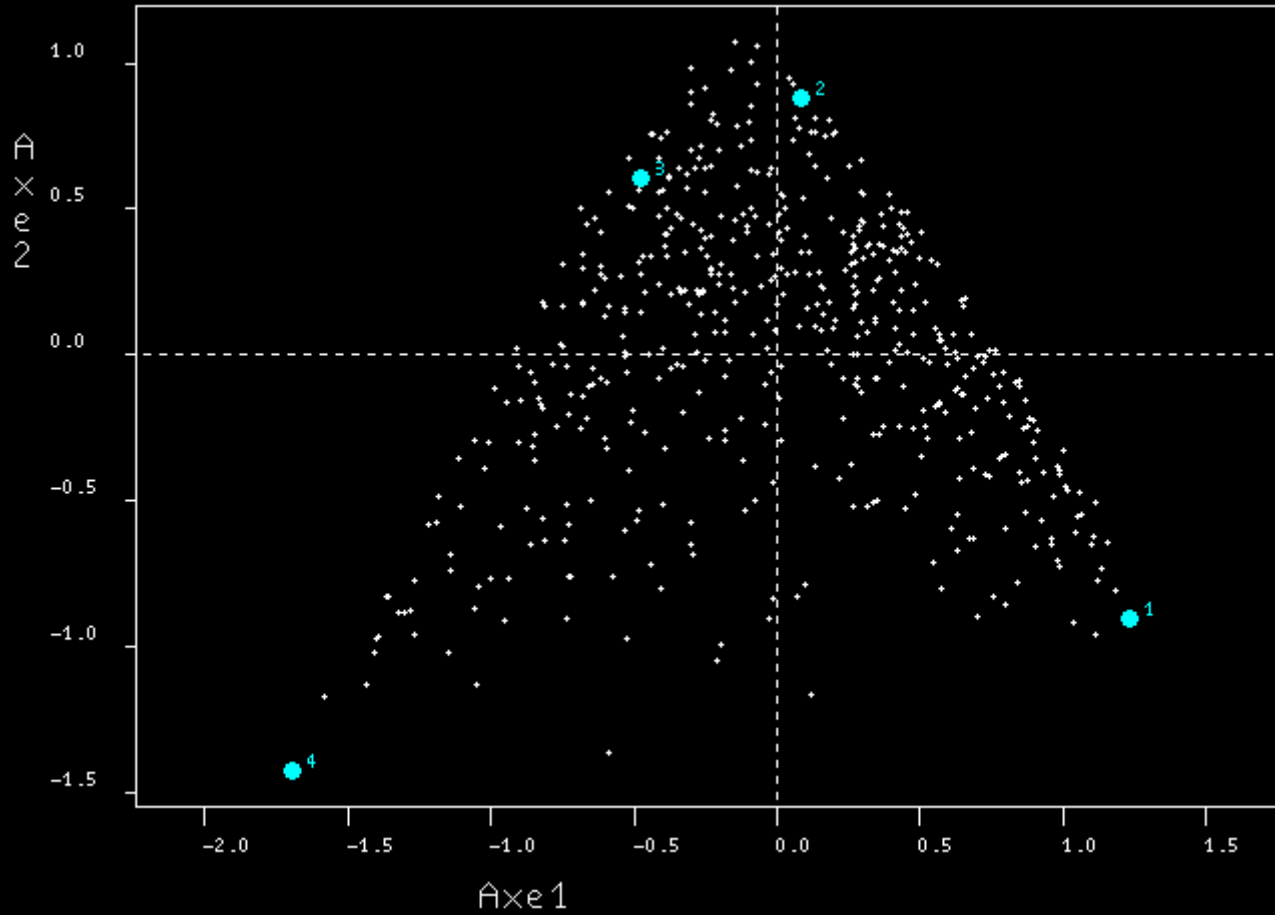


# pr1grd1cot1





# Jusqu'à jalonner le nuage de patrons-repères

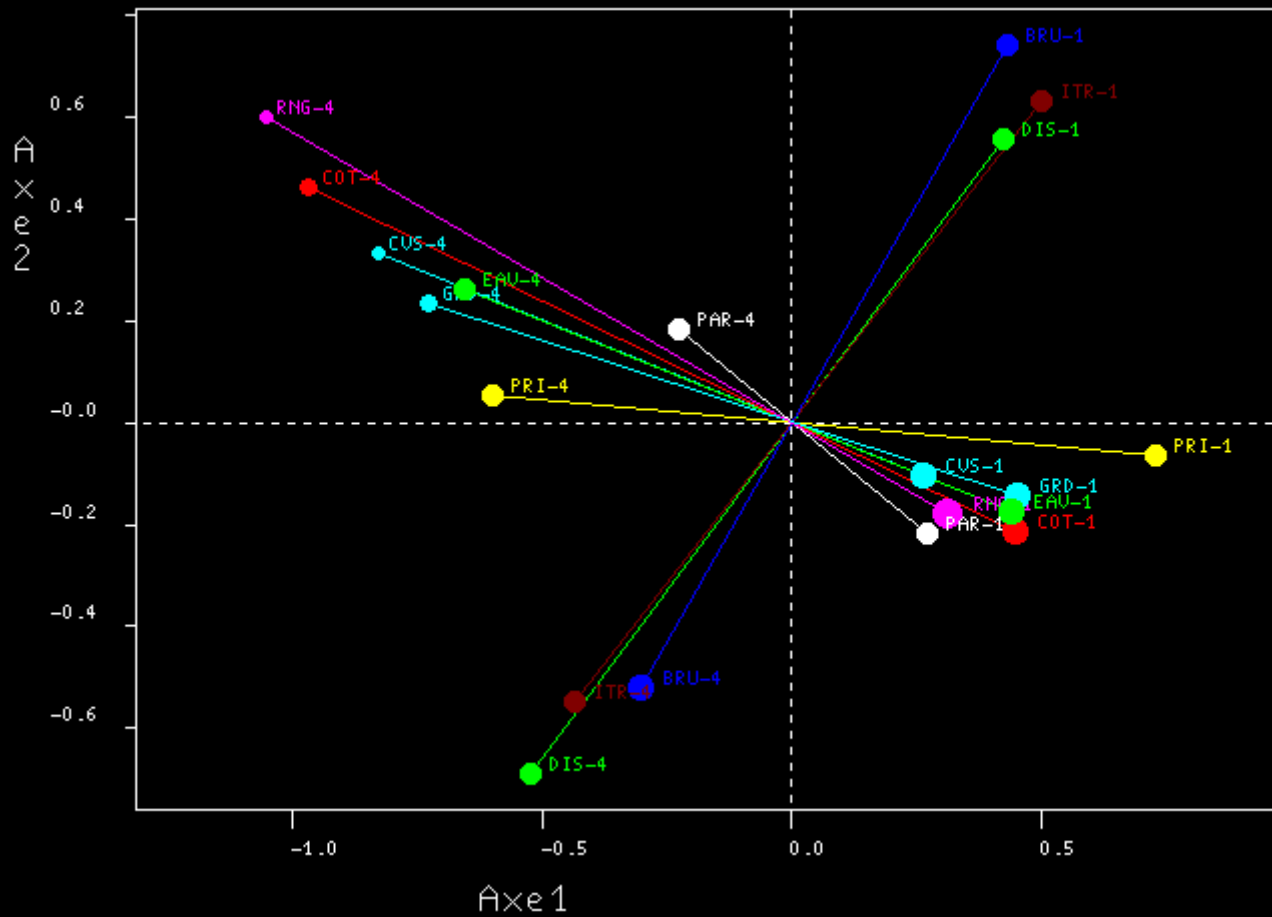


## Deuxième ACM : Données recodées

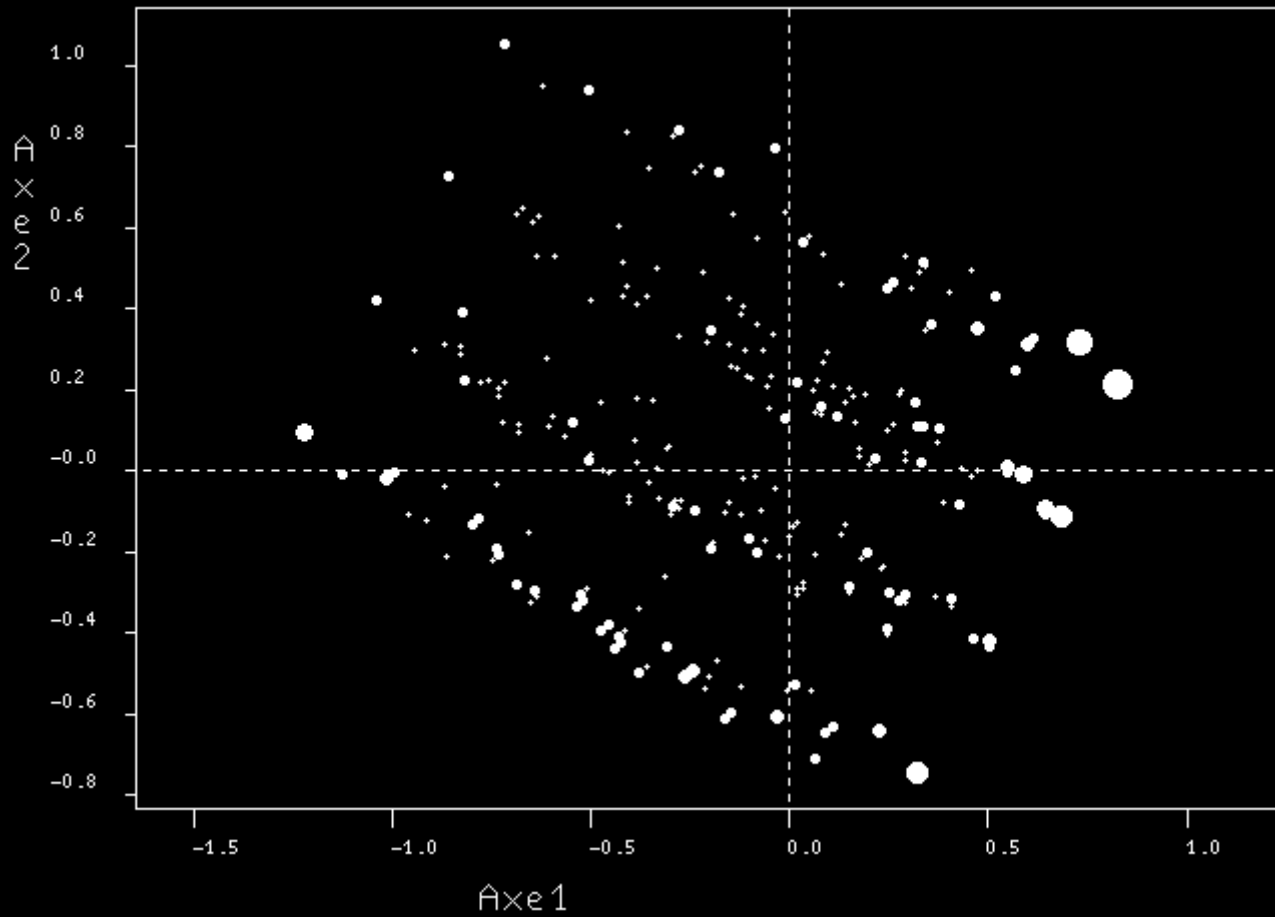
1 2 -> 1 = Satisfait

3 4 -> 4 = Insatisfait

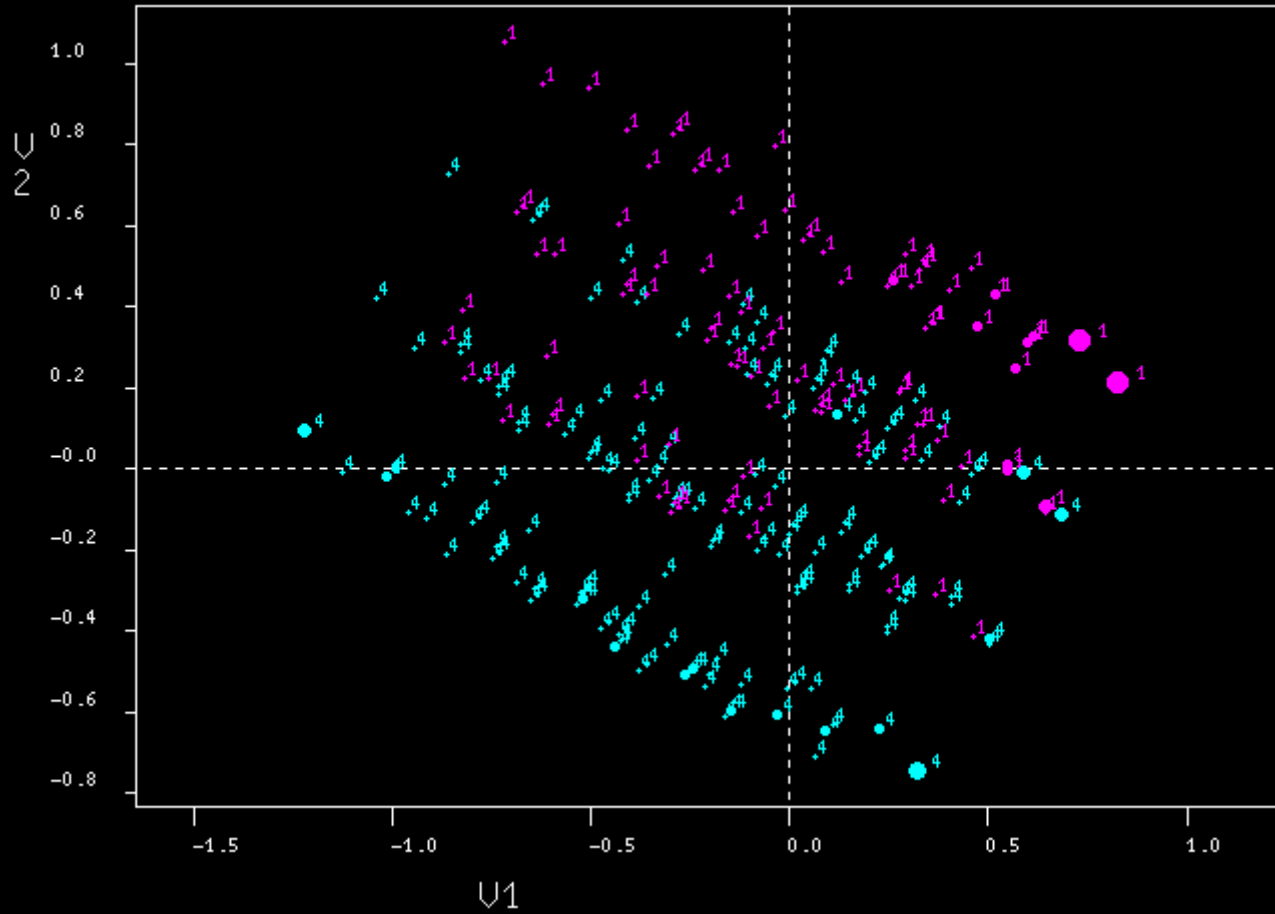
# Nuage des modalités après recodage



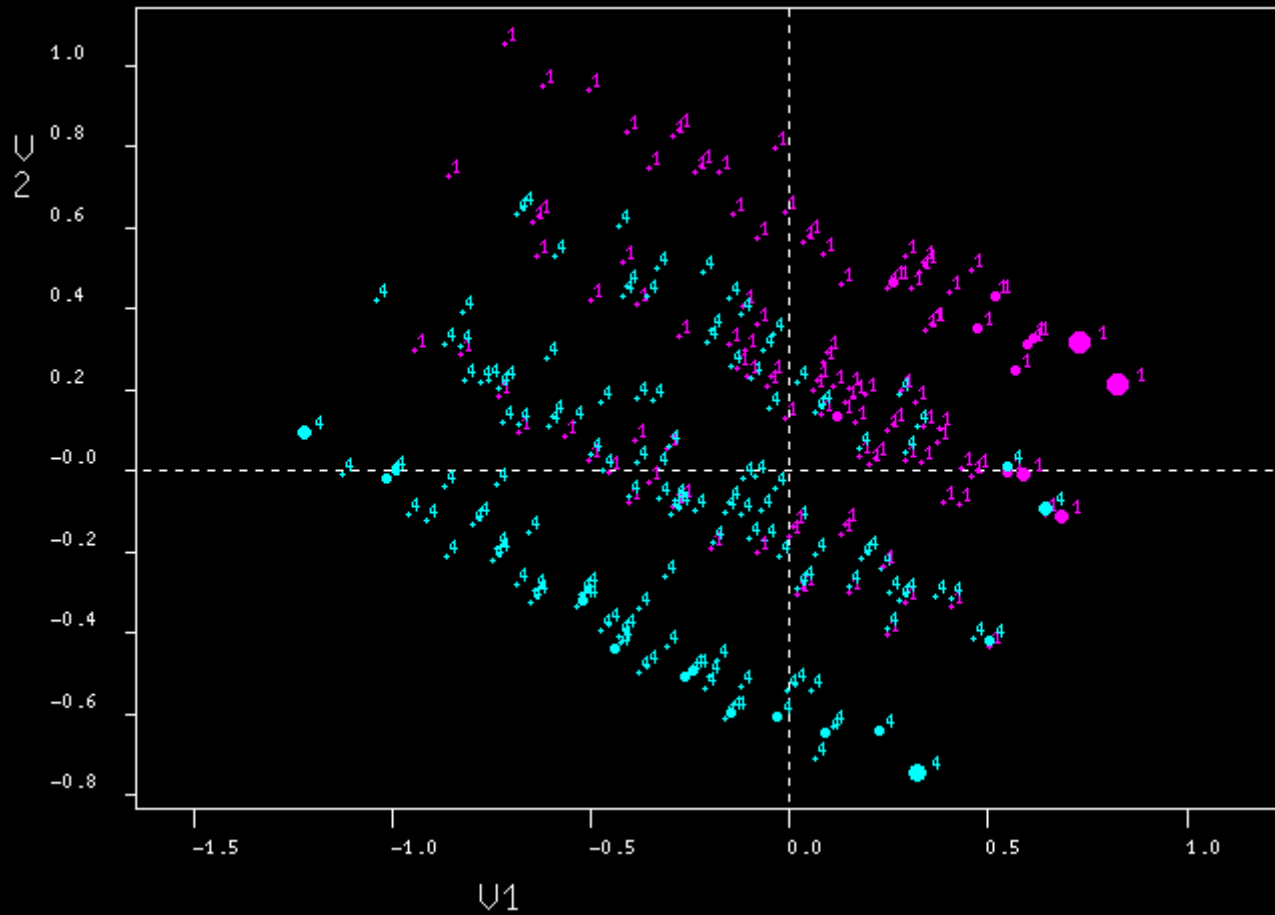
# Nuage des individus après recodage



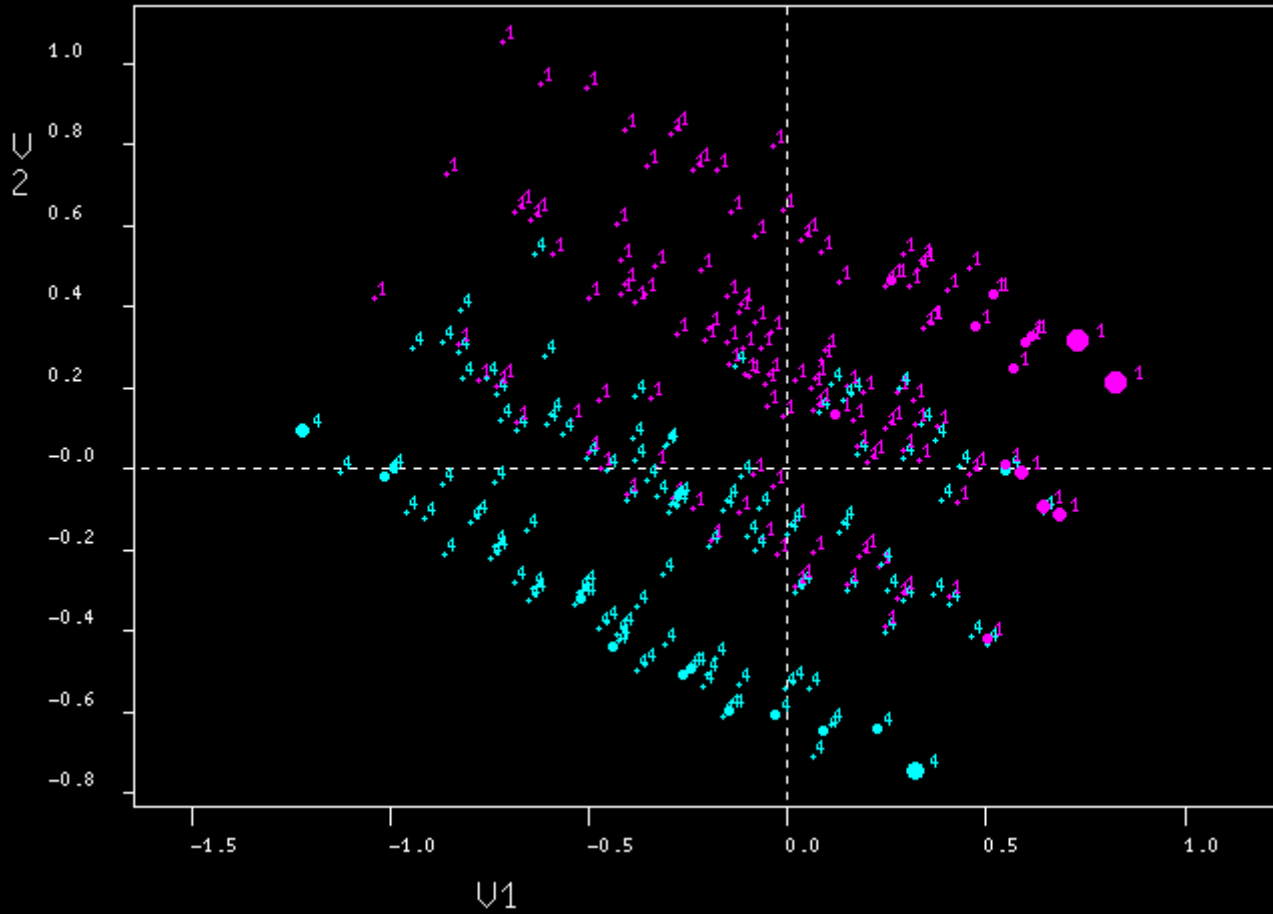
# BRU et les bandes



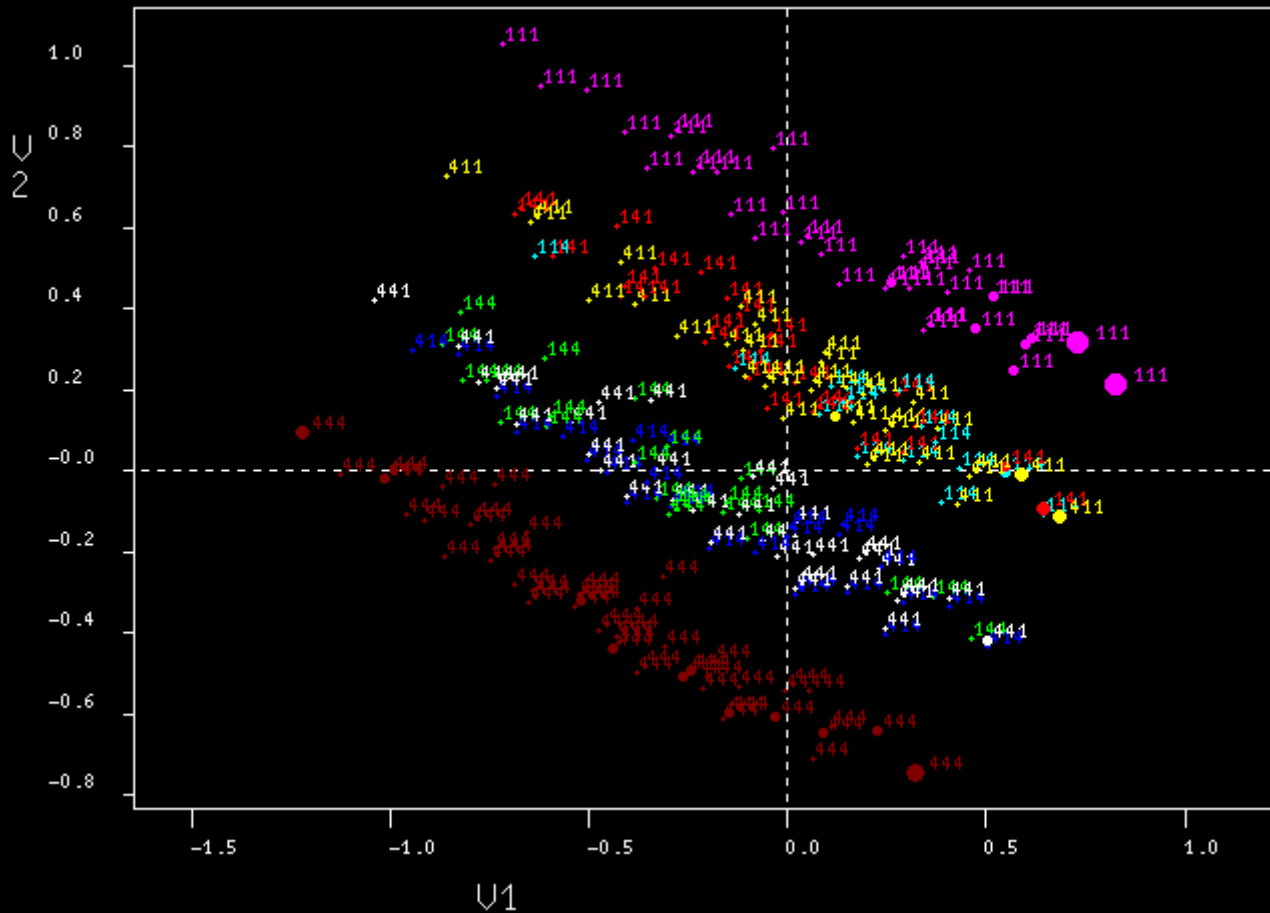
# ITR et les bandes



# DIS et les bandes

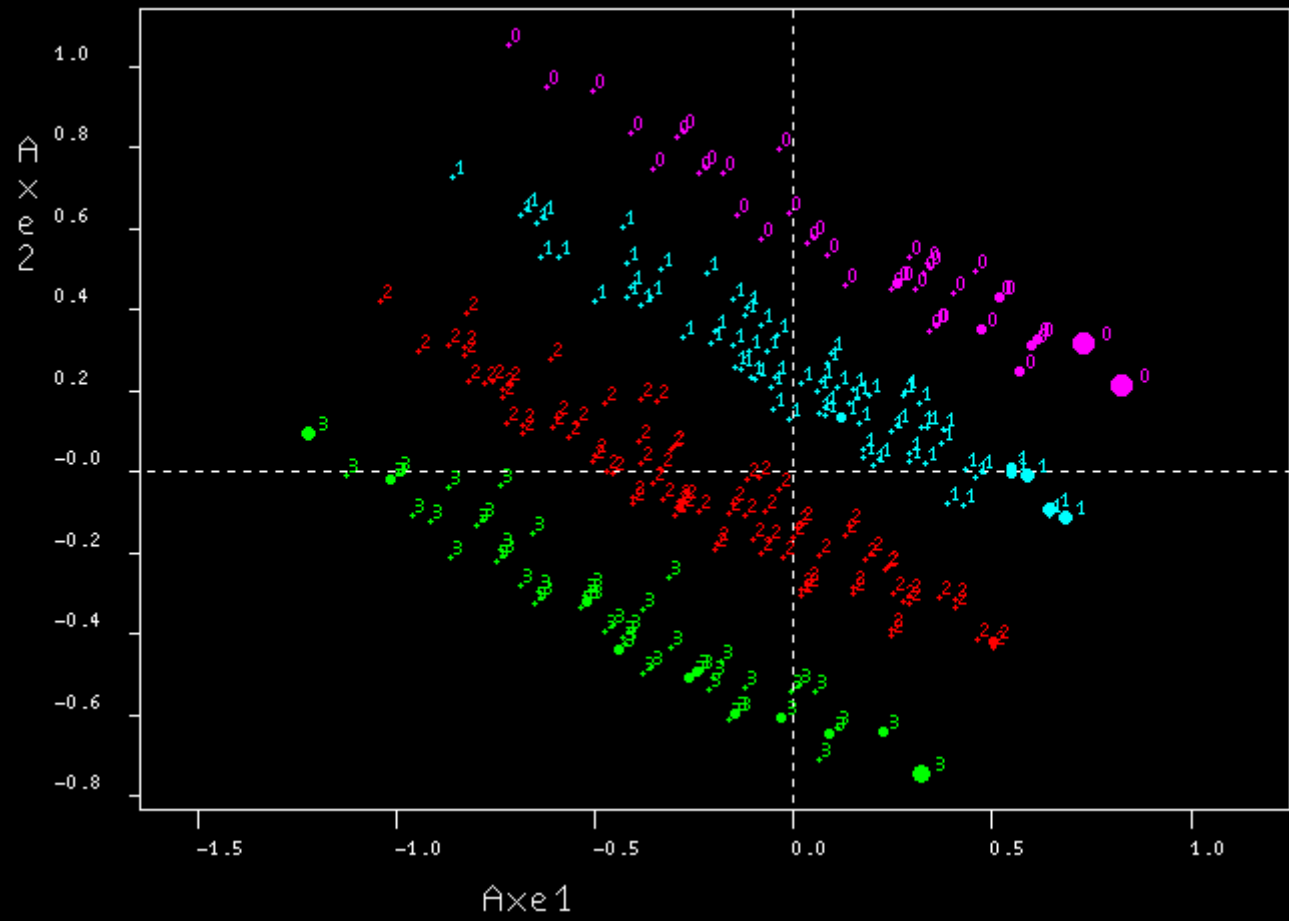


# Facteur croisant BRU, ITR et DIS

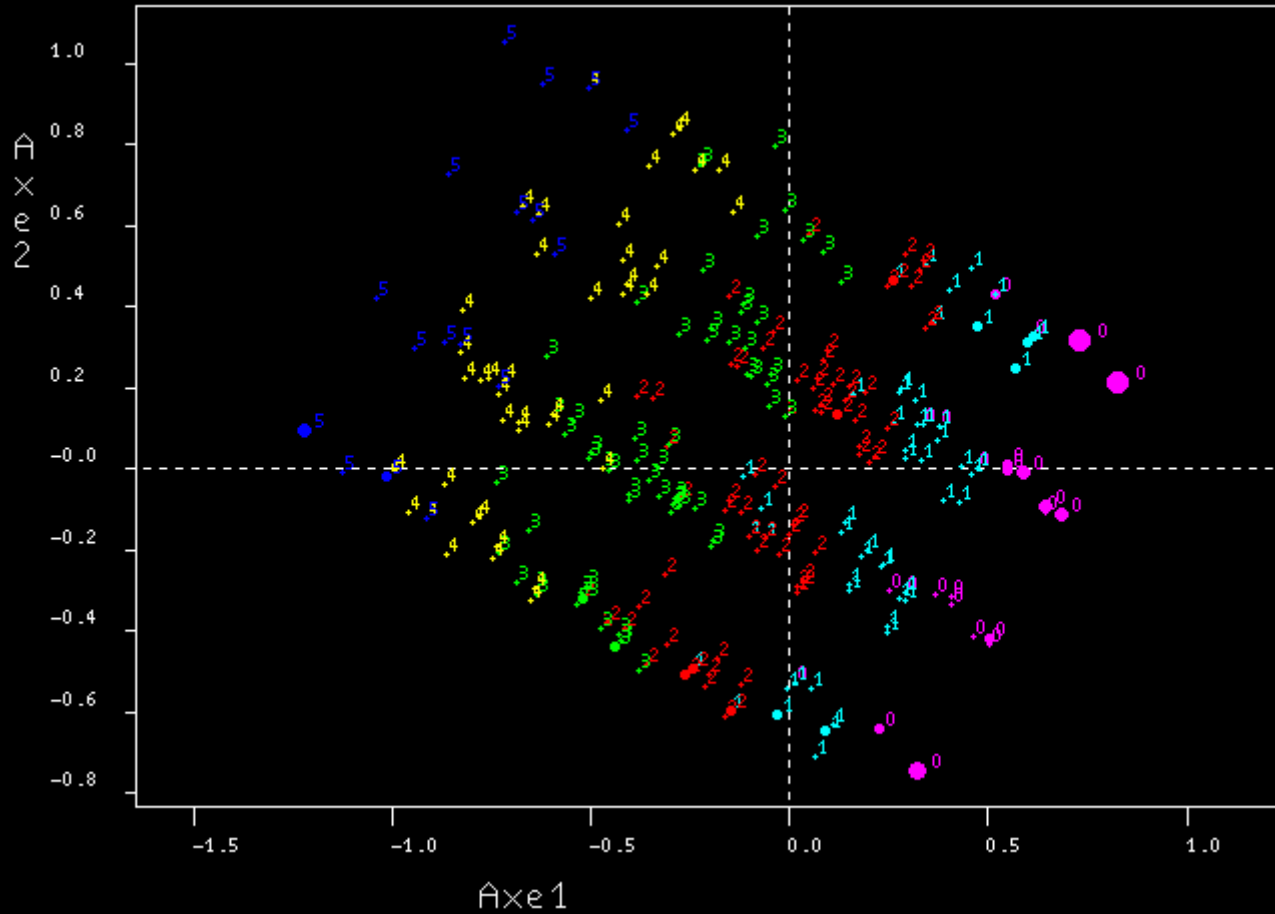




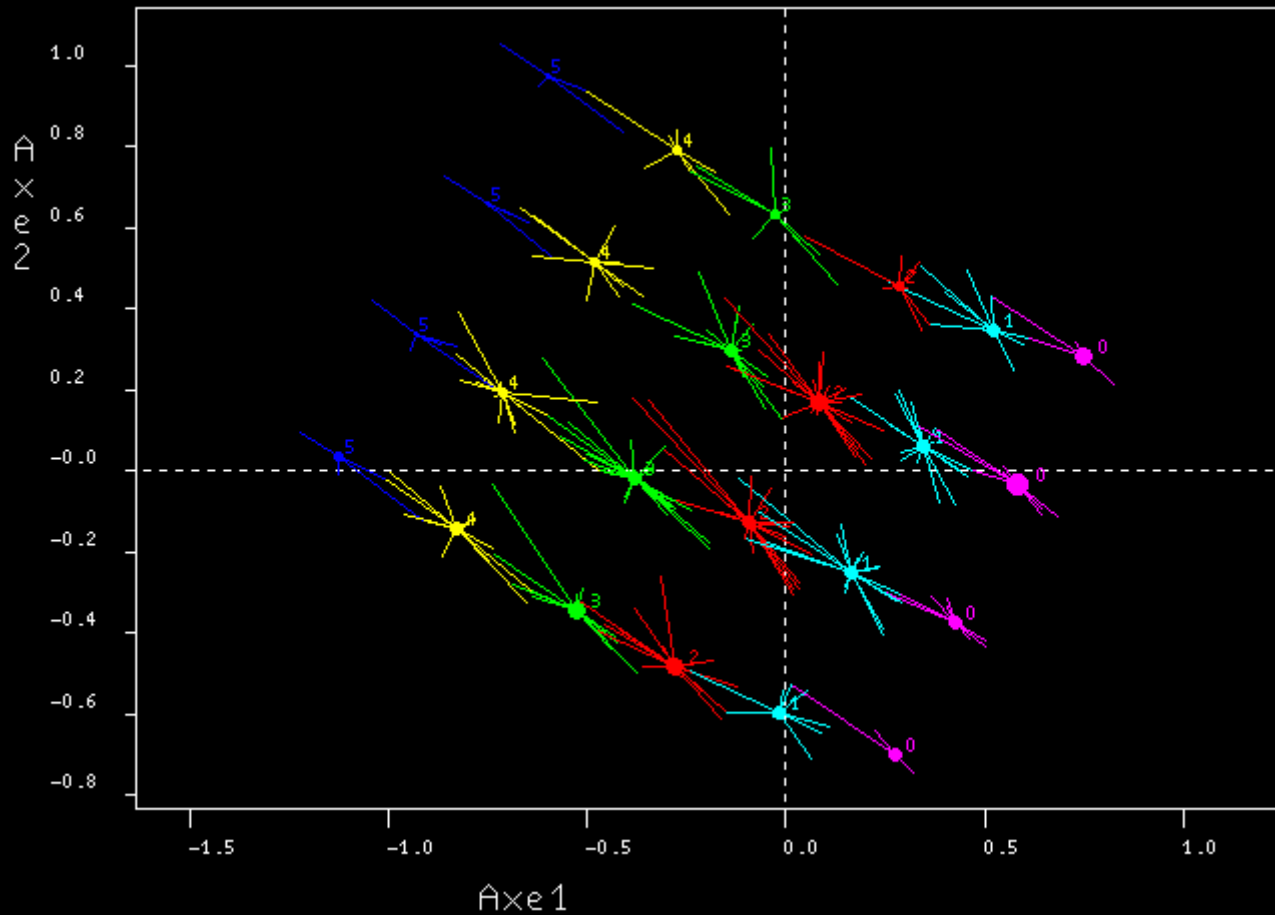
# Facteur **GENE** comptant le nombre de réponses « Insatisfaction » à **BRU**, **ITR** et **DIS**



# Facteur **ESPA** comptant le nombre de réponses « Insatisfaction » à **RNG, GRD, EAV, PRI et COT**



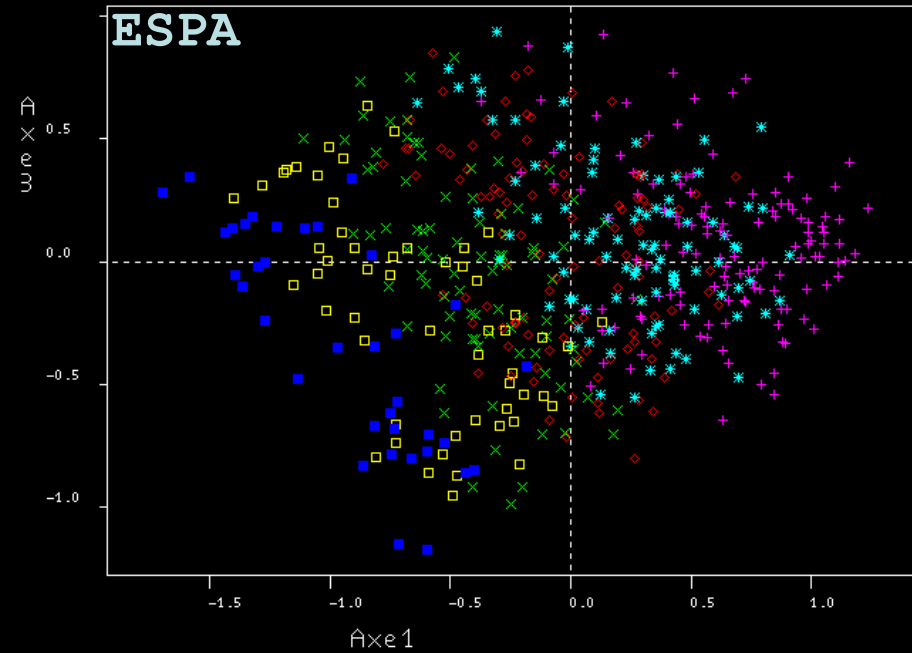
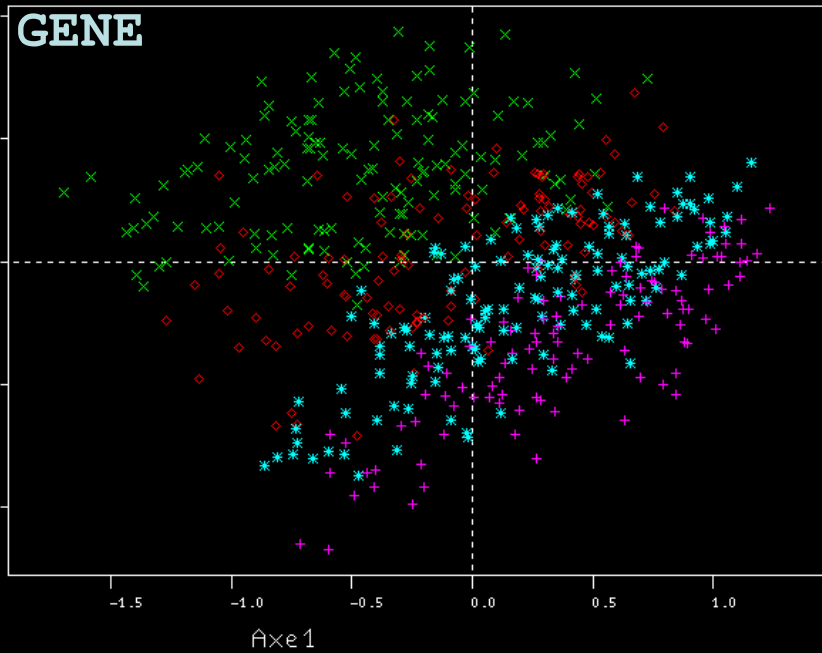
# Visualisation de la variance inter et intra liée au croisement des 2 facteurs construits : **ESPA** et **GENE**



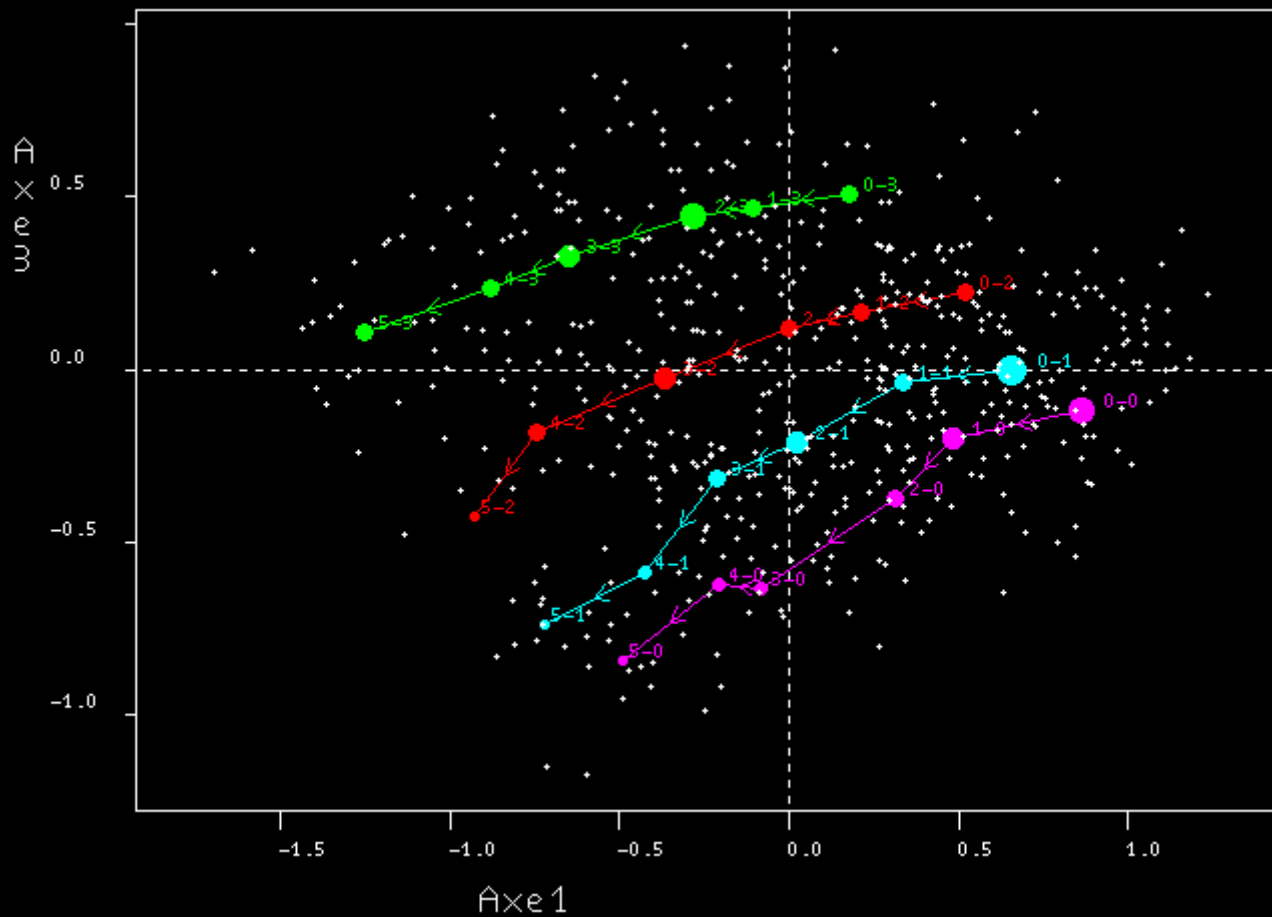
96% de variance du plan expliquée

**Ce phénomène était-il visible dans le nuage initial ?**

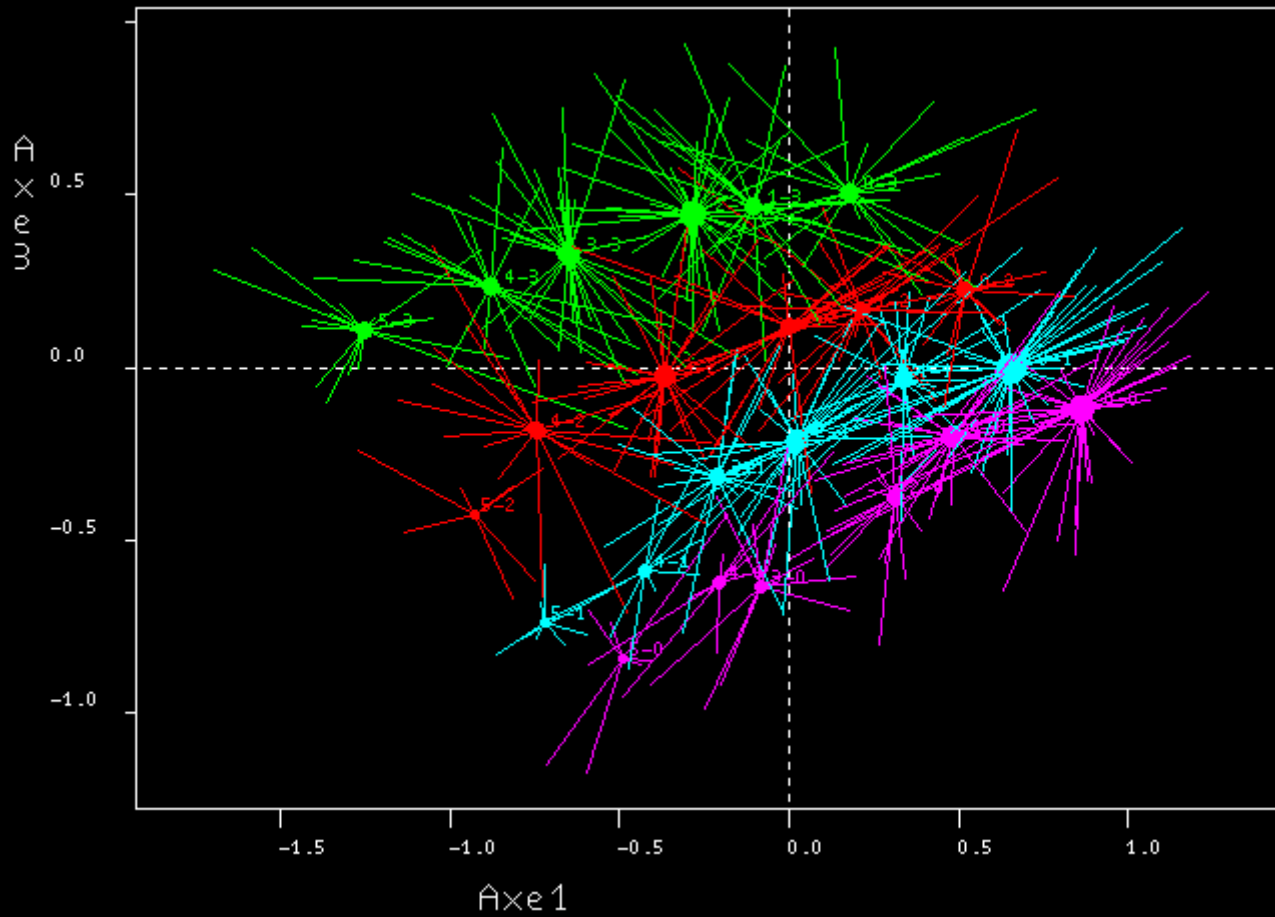
# Nuage des individus dans le plan 1-3 à la lumière des facteurs **GENE** et **ESPA**



# Points moyens du facteur croisé GENE avec ESPA



# 79% de variance du plan 1-3 expliquée



# Synthèse succincte des résultats obtenus

- **Sur les données initiales**
  - **Dimension générale de Satisfaction, commune à toutes les questions**
- **Sur les données recodées**
  - **Deux dimensions presque orthogonales de la Satisfaction :**
    - ✓ Une dimension liée à l'espace, à son aménagement
    - ✓ Une dimension liée à la gêne
  - **La satisfaction vis-à-vis des conditions de travail (COT) apparaît liée à l'aménagement de l'espace, et beaucoup moins à la gêne.**



## Les logiciels utilisés

- Analyse des correspondances multiples : **ADDAD** et **SPAD**
- Exploration des nuages
  - Visualisation des nuages : **EyeLID-2**
  - Décomposition des variances : **EyeLID-2**
  - Avantage : facteurs structurants

## D'autres logiciels ?

- ACM : très nombreux (**SAS, SPSS, Statistica, R,...**)
- Pour l'exploration des nuages (visualisation et décomposition) :
  - Nuage des modalités par défaut
  - Nuage des individus pas toujours accessible, et si accessible, pas d'exploitation simple.
  - Nécessité de passer par l'**ANOVA** pour les décompositions

## Conclusions

- De manière générale, dès que l'on dispose d'un nuage sur lequel on a défini des facteurs structurants, on peut :
  - L'explorer visuellement : sous-nuages, points moyens, nuages dérivés
  - L'explorer numériquement : décomposition des variances
- Dans le cas de l'ACM :
  - Mise en avant du nuage des individus

Cette méthodologie relève de  
**l'Analyse géométrique des données**  
(Benzécri; Rouanet & Le Roux)

## Etudes de cas :

**Bonnet P., Le Roux B. & Lemaine G.** (1996). Analyse géométrique des données : une enquête sur le racisme. *Mathématiques et Sciences Humaines*, 136, 5-24.

**Le Roux B. & Rouanet H.** (2004). Geometric data analysis of individuals differences among gifted students in the third grade.

<http://epgy.stanford.edu/research/GeometricDataAnalysis.pdf>

## Ouvrages, Articles généraux

**Bernard, J-M., Le Roux B., Rouanet H. & Schiltz M-A.** (1989). L'analyse des données multidimensionnelles par le langage d'interrogation de données (LID) : au-delà de l'analyse des correspondances. *Bulletin de Méthodologie Sociologique*, 23, 3-46.

**Bernard, J-M.** (1994). L'analyse descriptive des données planifiées. *Mathématiques, Informatique et Sciences Humaines*, 126, 7-96.

**Rouanet H. & Le Roux B.** (1993). *Analyse des données multidimensionnelles*. Paris : Dunod.

**Le Roux B. & Rouanet H.** (2004). *Geometric Data Analysis: from Correspondence Analysis to Structured Data*. Dordrecht : Kluwer.