

Type de données et méthodes d'analyse

# STATISTIQUE MULTIVARIEE

ÉCOLE D'ÉTÉ INTERNATIONALE

## MÉTHODES ET OUTILS DES SCIENCES DES TERRITOIRES

UNE PERSPECTIVE NORD-SUD, SUD-NORD ET SUD-SUD

ÉTAPE 2 • IRSP, Ouidah (Bénin) 27 février - 10 mars 2023



## ANALYSE FACTORIELLE

Auteur.e.s

ANE LANDRY TANO, France GUERIN-PACE  
& Claude GRASLAND

*Contributeur.ice.s :*

# INTRODUCTION

## Objectifs du module

- ✓ Apprendre à résumer et à structurer l'information issue de grands tableaux numériques de données ;
- ✓ Réduire les dimensions du tableau de données en projetant dans des espaces de plus faibles dimensions;
- ✓ Établir un lien entre les types de tableaux de données (nature des variables) et le type d'analyse à appliquer afin de comprendre la particularité du type d'analyse;
- ✓ Connaître les types d'analyse multivariée ; savoir les réaliser et les interpréter ; savoir enchaîner les méthodes, repérer les limites des analyses et cartographier les informations.

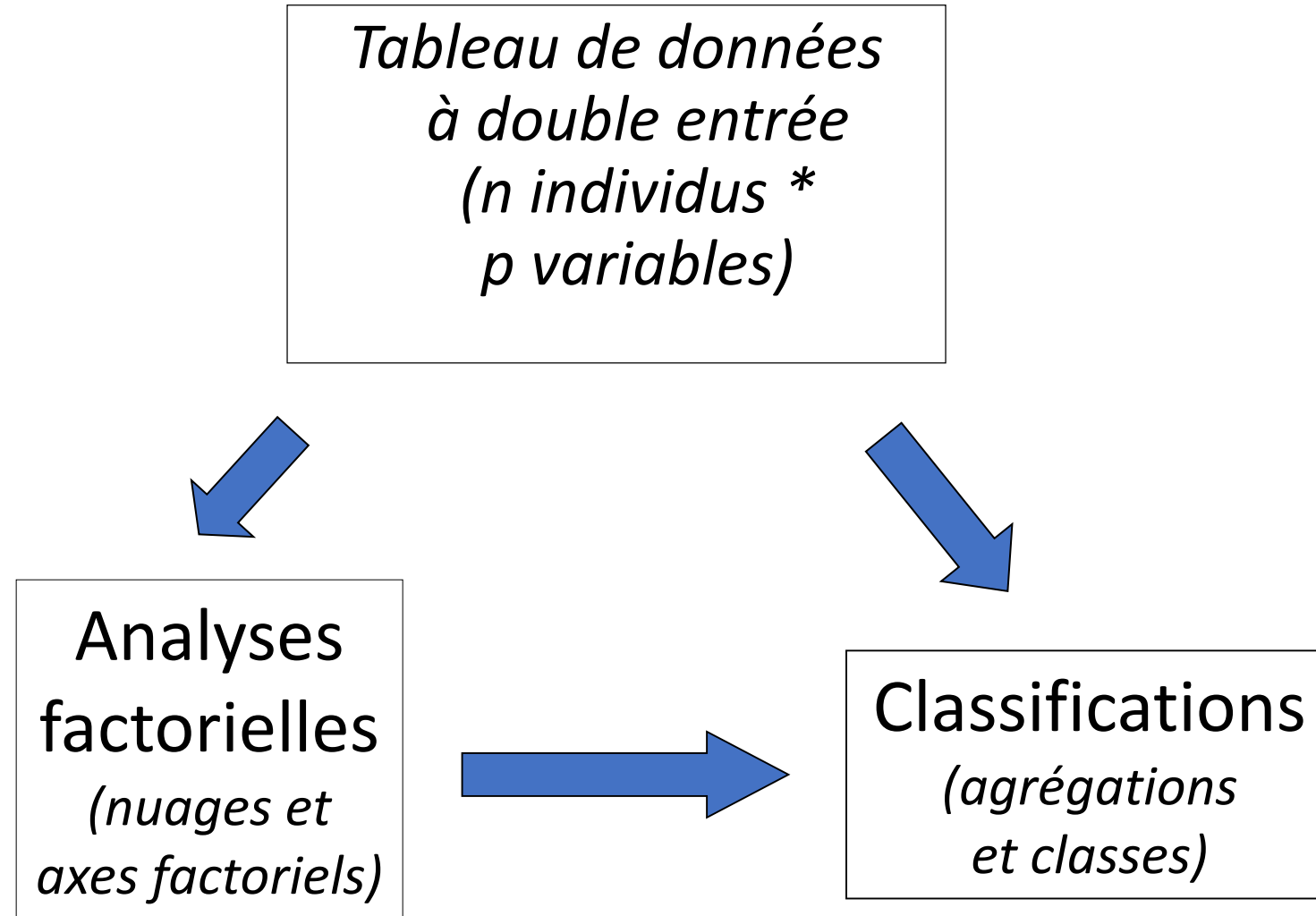
## Historique des analyses factorielles :

- ✓ Au début 20<sup>ème</sup> siècle, des psychomètres ont mis au point les premières analyses factorielles (Pearson, 1900) ;
- ✓ Avant la 2<sup>nde</sup> guerre mondiale, des statisticiens (Hotelling, Thurstone, 1934) ont repris ces travaux dans une perspective descriptive, mettant au point l'analyse en composantes principales (ACP), pour la synthèse de variables quantitatives ;
- ✓ Après la 2<sup>nde</sup> guerre mondiale, un statisticien français (J.P. Benzecri, 1957) a adapté, sous le nom l'analyse Factorielle des Correspondances (AFC), pour la synthèse de tableaux composés de variables qualitatives (tableau de contingence).

## Définition de l'analyse des données

- ✓ Ensemble de méthodes statistiques *multidimensionnelles* et *descriptives*
- ✓ Tableau à  $n$  individus (variables) et  $p$  variables
- ✓ Objectifs : Résumer et structurer l'information contenue dans un tableau de données en mettant en évidence des relations entre variables ou entre groupes de variables
- ✓ Principe : passer d'un nuage de points dans un espace de dimension  $p$  à un ou plusieurs sous-espaces de dimension 2. « On » cherche la meilleure « image approchée » du nuage en projection sur une droite. C'est celle qui respecte au mieux les distances entre tous les couples de points. La première droite ou *facteur*

# Deux grands types de méthodes



## Type de données et méthodes d'analyse

**Analyse Factorielle des Correspondances (AFC)** s'applique aux tableaux de contingence croisant deux variables qualitatives (mais pas seulement

➤ **Population par département et langues au Bénin** (source : RP 2002)

**Analyse Factorielle des Correspondances Multiples (AFCM)** s'applique aux tableaux de contingence multiples (plusieurs variables qualitatives juxtaposées).

➤ **Enquête sur la scolarisation en Côte d'Ivoire** (source : A.L. Tanoh)

**Analyse en Composantes Principales (ACP)** s'applique aux tableaux de mesure (variables quantitatives continues ou quasi continues, c'est-à-dire discrètes à valeurs nombreuses)

➤ **Ration alimentaire par pays ou régions du Monde** (source : FAO)

# Vocabulaire de l'analyse factorielle

- ✓ **Inertie** : d'un nuage de point mesure la dispersion des points autour du centre de gravité (somme des variances des différentes variables) ; information contenue dans le tableau de données ;
- ✓ **Axes factoriels = facteurs = composantes** : axes « virtuels » issus d'une synthèse entre les variables de l'analyse ; un axe est une combinaison linéaire de variables ;

Dans un tableau de dimension  $(n,p)$ ,  $n$  individus dans un espace à  $p$  dimensions. Un **plan factoriel** est de dimension 2. A chaque axe est associé une **valeur propre** ou part d'inertie (information) prise en compte par un axe.

- ✓ **Variables actives** : variables sélectionnées en lien avec la thématique qui vont participer à la détermination des axes factoriels ; **variables supplémentaires** ou **illustratives** ne participent pas à la formation des axes mais viennent illustrer celle-ci.



# Vocabulaire de l'analyse factorielle

- ✓ **Contribution** d'une variable à la formation d'un axe (somme des carrés des écarts, divisée par la variance de l'axe) ; part du rôle de la variable dans la formation de l'axe ; la somme des contributions (lignes ou colonnes) des variables à un axe est égale à 1 ou 100%;
- ✓ **Qualité de Représentation** d'une variable par un axe : somme des cosinus carré des angles induits par chaque axe factoriel ; elle indique la proximité ou l'éloignement du point à l'axe factoriel ; la somme des Q.R d'un point sur les axes retenus indique s'il est bien ou mal rendu par l'ensemble des axes.

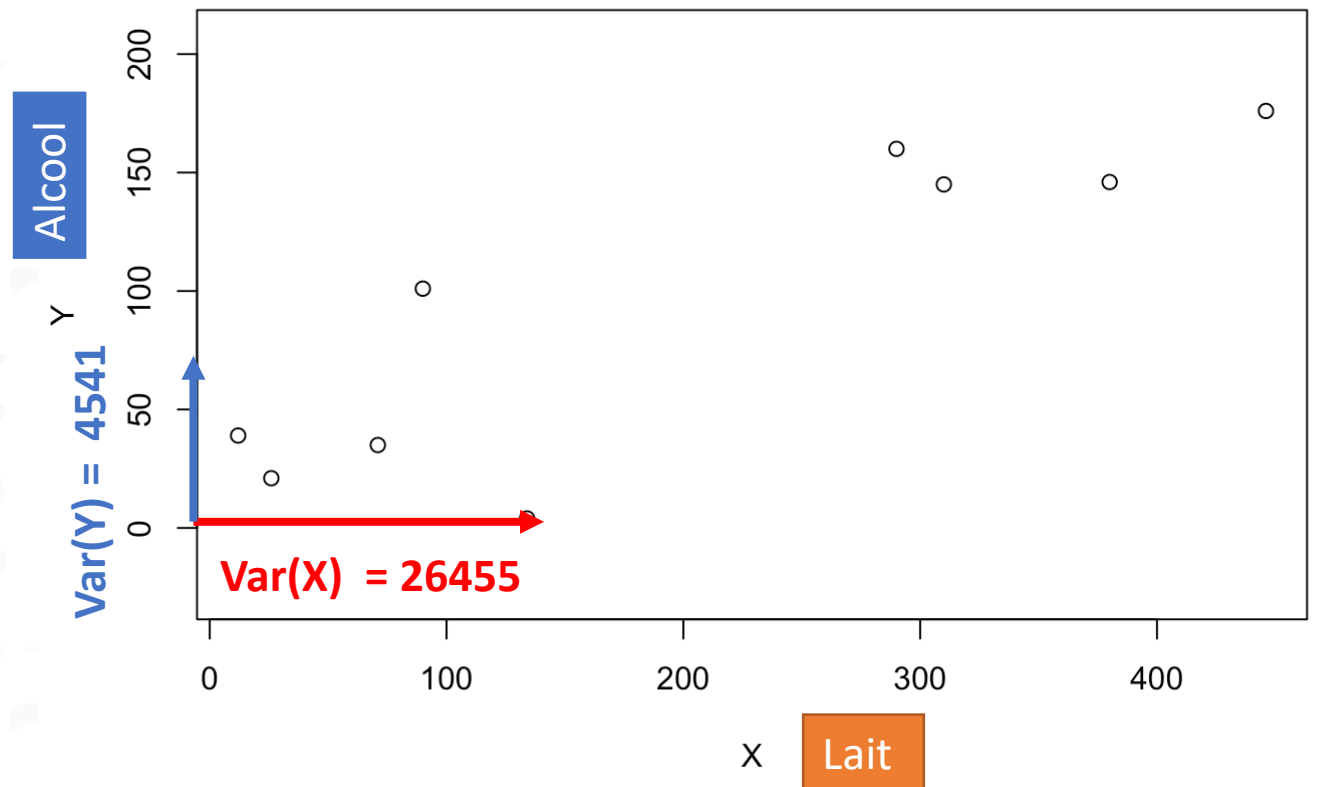
# ANALYSE EN COMPOSANTES PRINCIPALES (ACP)

<https://ee-cist.github.io/EXP3-Multivar/Cours-ACP.html>

# ACP : Cas de 2 variables quantitatives

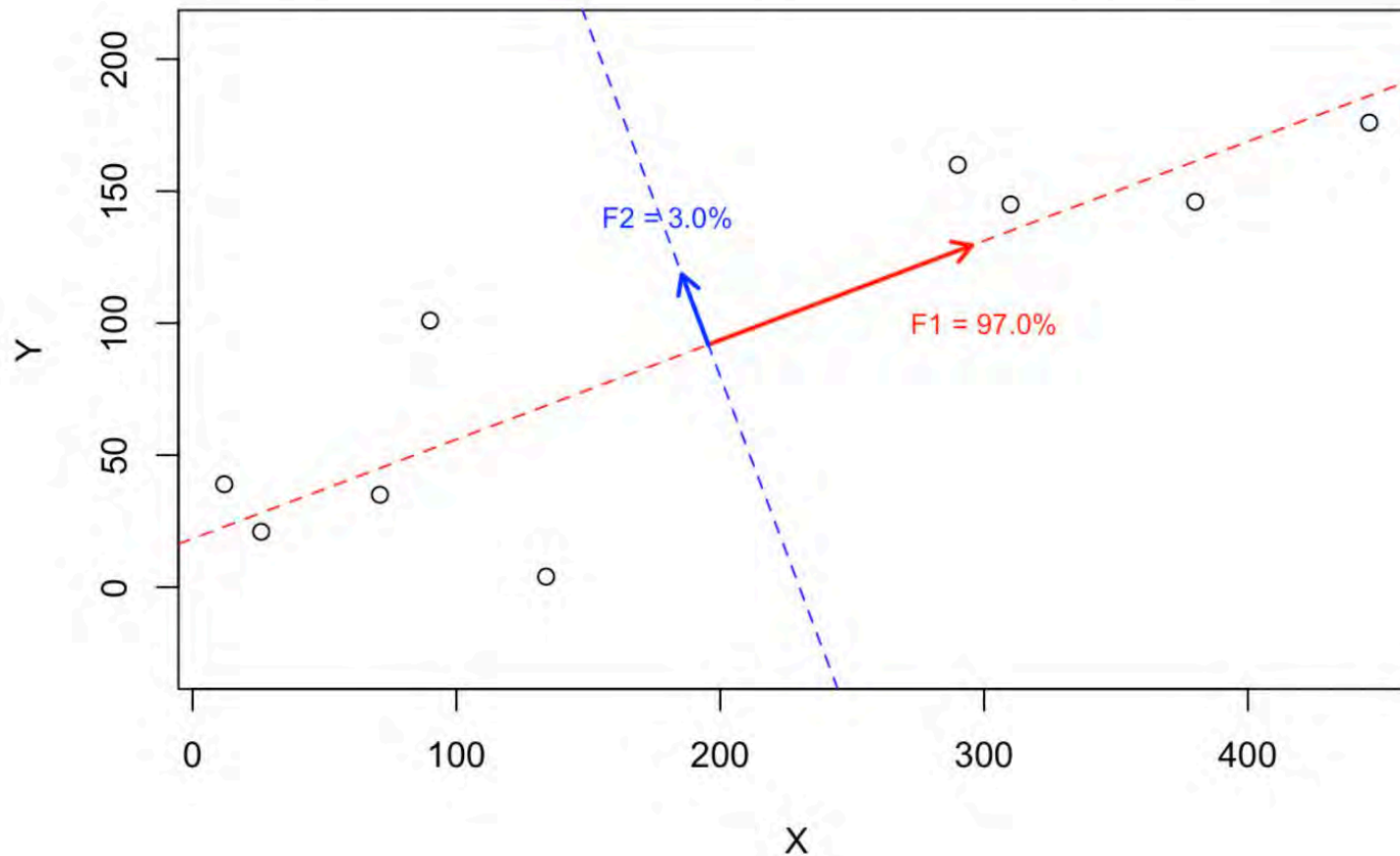
Consommation moyenne de lait et d'alcool en Europe et en Afrique en 2020 (en kCal/pers /jour)

	region	Lait	Alcool
AFR_Sud	Afrique australe	90	101
AFR_Centre	Afrique centrale	12	39
AFR_Ouest	Afrique occidentale	26	21
AFR_Est	Afrique orientale	71	35
AFR_Nord	Afrique septentrionale	134	4
EUR_Sud	Europe méridionale	310	145
EUR_Ouest	Europe occidentale	446	176
EUR_Est	Europe orientale	290	160
EUR_Nord	Europe septentrionale	380	146



# ACP : Cas de 2 variables quantitatives

Centrage et rotation des axes effectuée par l'ACP



L'ACP opère un **centrage** et une **rotation** de façon à ce que le premier axe résume le maximum de la variance totale du nuage de points

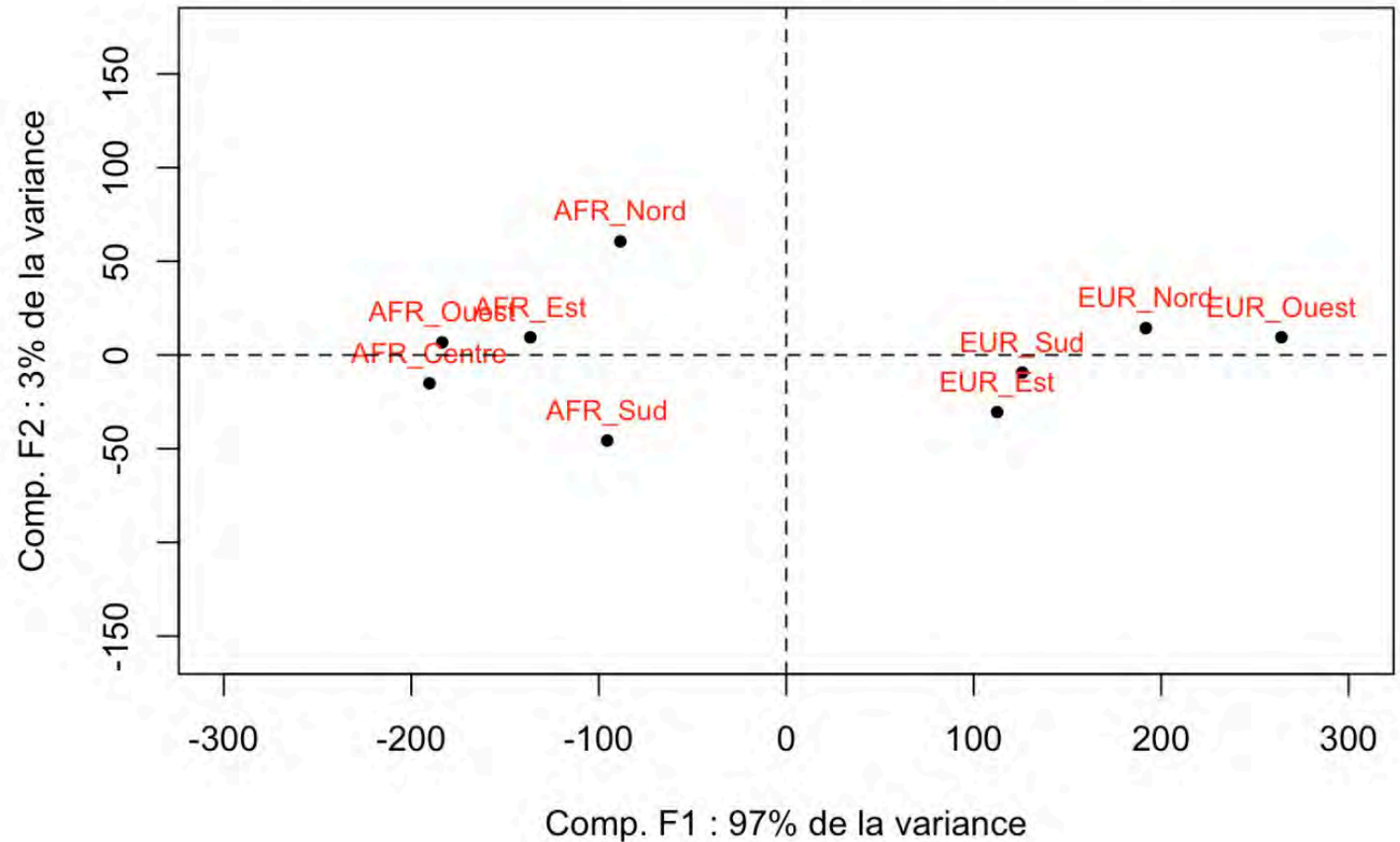
$$\begin{cases} F_i^1 = a_1(X_i - \bar{X}) + b_1(Y_i - \bar{Y}) \\ F_i^2 = a_2(X_i - \bar{X}) + b_2(Y_i - \bar{Y}) \end{cases}$$

# ACP : Cas de 2 variables quantitatives

Coordonnées des individus

code	F1	F2
1 AFR_Sud	-95.5	-45.7
2 AFR_Centre	-190.3	-15.1
3 AFR_Ouest	-183.6	6.7
4 AFR_Est	-136.5	9.4
5 AFR_Nord	-88.5	60.6
6 EUR_Sud	125.9	-9.4
7 EUR_Ouest	264.1	9.5
8 EUR_Est	112.5	-30.5
9 EUR_Nord	191.8	14.3

Coordonnées des individus sur les composantes



# ACP : Cas de 2 variables quantitatives

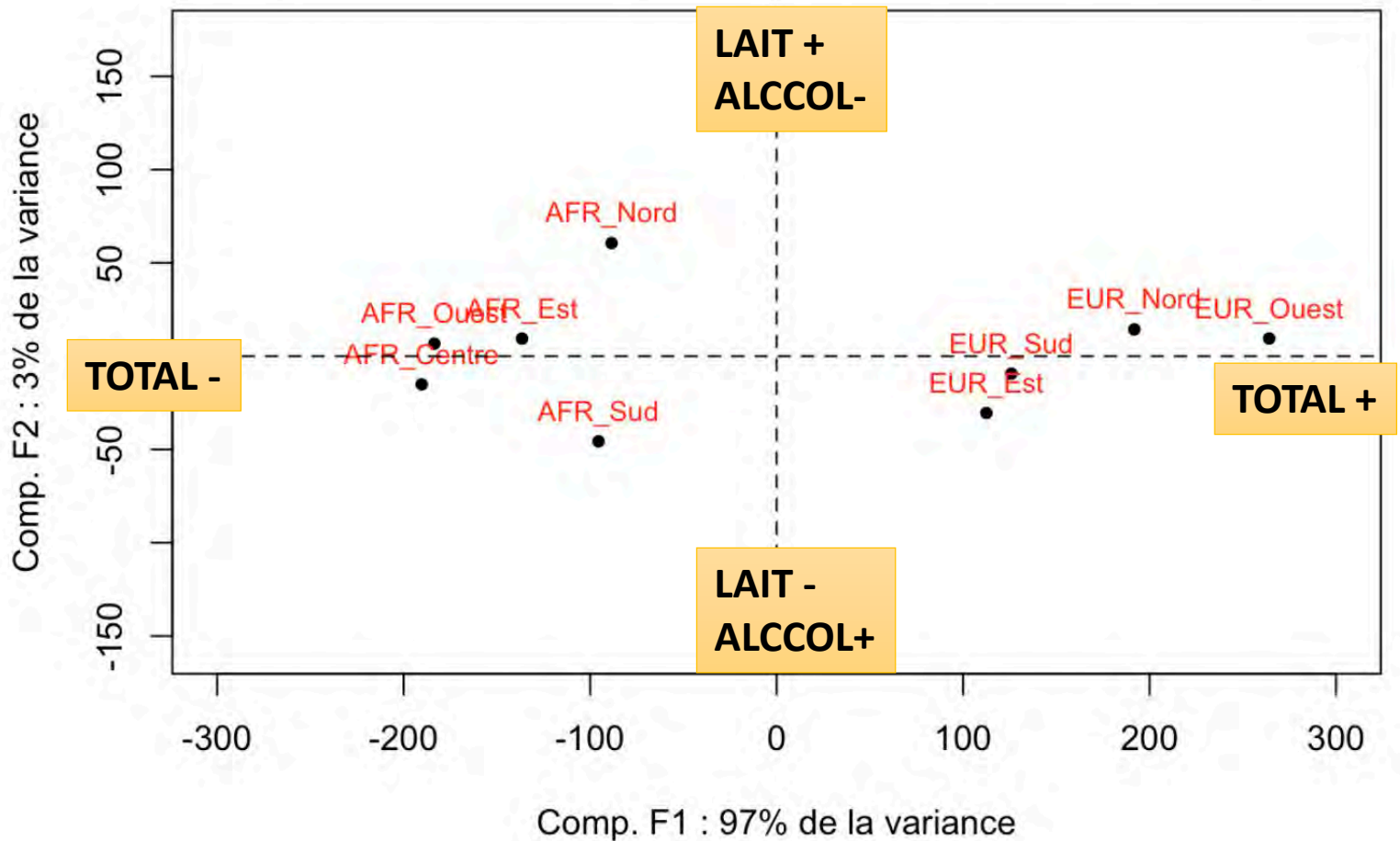
- **la composante 1 exprime le niveau global de consommation de lait ET d'alcool** puisque les variables lait et d'alcool sont positivement corrélées entre elles et avec cette composante. Elle oppose donc les pays à forte consommation de **lait ET d'alcool** et les pays à faible consommation.
- **la composante 2 exprime une préférence relative pour le lait OU l'alcool.** Comme elle est **orthogonale** à la première composante, elle exprime donc un choix *toutes choses égales quant au niveau global de consommation de ces deux boissons*. La mention "*toutes choses égales*" est essentielle pour souligner que ces excédents ou déficits sont relatifs au niveau global de consommation des deux boissons.

# ACP : Cas de 2 variables quantitatives

Contribution des individus aux composantes

	Dim.1	Dim.2
Afrique australe	3.8	28.0
Afrique centrale	15.1	3.1
Afrique occidentale	14.0	0.6
Afrique orientale	7.7	1.2
Afrique septentrionale	3.3	49.5
Europe méridionale	6.6	1.2
Europe occidentale	29.0	1.2
Europe orientale	5.3	12.5
Europe septentrionale	15.3	2.8
Sum	100.0	100.0

Coordonnées des individus sur les composantes



# ACP : Cas de k variables quantitatives

Consommation alimentaire en kCal/pers/jou (tableau brut)

	<b>Animaux</b>	<b>Céréales</b>	<b>Légumes</b>	<b>Huiles</b>	<b>Sucres</b>	<b>Divers</b>
AFR_Sud	448	1358	114	421	300	161
AFR_Centre	113	633	1048	249	76	138
AFR_Ouest	120	1209	692	359	108	214
AFR_Est	165	1178	380	197	107	229
AFR_Nord	326	1819	243	363	304	183
EUR_Sud	876	970	297	650	305	311
EUR_Ouest	1202	882	294	506	408	300
EUR_Est	897	1138	272	385	370	281
EUR_Nord	1051	971	271	463	298	341

- **Viandes** : abats, viandes, poissons, lait, graisses animales,
- **Céréales** : blé, maïs, riz, sorgho...
- **Huiles** : cultures oléagineuses, huiles végétales
- **Sucres** : cultures sucrières, sucres et édulcorants
- **Légumes**: légumes, fruits, ignames, pommes de terres, ...
- **Divers** : épices, stimulants, boissons alcooliques

Paramètres principaux des variables brutes

	<b>Animaux</b>	<b>Céréales</b>	<b>Légumes</b>	<b>Huiles</b>	<b>Sucres</b>	<b>Divers</b>
Moyenne	577.6	1128.7	401.22	399	252.9	239.8
écart-type	430.2	334.4	288.72	135	122.9	71.8
coeff. de variation	0.7	0.3	0.72	0	0.5	0.3
variance	185078.8	111836.5	83358.69	18178	15115.4	5151.7
variance (%)	44.2	26.7	19.91	4	3.6	1.2



# ACP normée : variables standardisées

Consommation alimentaire en kCal/pers/jou (tableau standardisé)

	<b>Animaux</b>	<b>Céréales</b>	<b>Légumes</b>	<b>Huiles</b>	<b>Sucres</b>	<b>Divers</b>
AFR_Sud	-0.30	0.69	-0.99	0.16	0.38	-1.10
AFR_Centre	-1.08	-1.48	2.24	-1.11	-1.44	-1.42
AFR_Ouest	-1.06	0.24	1.01	-0.30	-1.18	-0.36
AFR_Est	-0.96	0.15	-0.07	-1.50	-1.19	-0.15
AFR_Nord	-0.58	2.06	-0.55	-0.27	0.42	-0.79
EUR_Sud	0.69	-0.47	-0.36	1.86	0.42	0.99
EUR_Ouest	1.45	-0.74	-0.37	0.79	1.26	0.84
EUR_Est	0.74	0.03	-0.45	-0.11	0.95	0.57
EUR_Nord	1.10	-0.47	-0.45	0.47	0.37	1.41

- **Viandes** : abats, viandes, poissons, lait, graisses animales,
- **Céréales** : blé, maïs, riz, sorgho...
- **Huiles** : cultures oléagineuses, huiles végétales
- **Sucres** : cultures sucrières, sucres et édulcorants
- **Légumes**: légumes, fruits, ignames, pommes de terres, ...
- **Divers** : épices, stimulants, boissons alcooliques

Paramètres principaux des variables standardisées

	<b>Animaux</b>	<b>Céréales</b>	<b>Légumes</b>	<b>Huiles</b>	<b>Sucres</b>	<b>Divers</b>
moyenne	0.0	0.0	0.0	0.0	0.0	0.0
écart-type	1.0	1.0	1.0	1.0	1.0	1.0
variance	1.0	1.0	1.0	1.0	1.0	1.0
variance (%)	16.7	16.7	16.7	16.7	16.7	16.7

# ACP normée : matrice de corrélation

Coefficient de corrélation de Pearson

## **Animaux Céréales Légumes Huiles Sucres Divers**

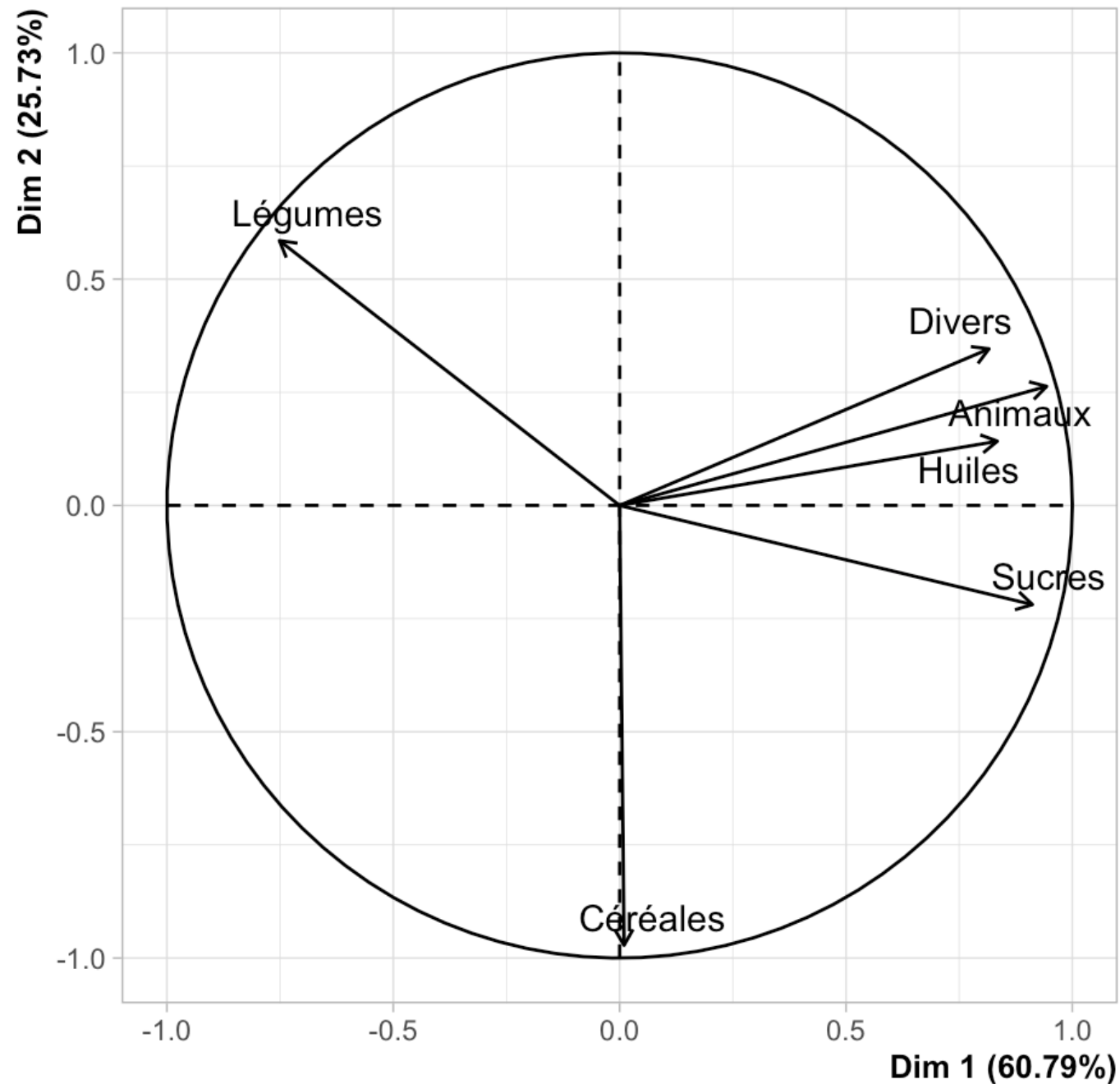
Animaux	1.00	-0.26	-0.55	0.74	0.85	0.84
Céréales	-0.26	1.00	-0.53	-0.10	0.19	-0.27
Légumes	-0.55	-0.53	1.00	-0.48	-0.77	-0.45
Huiles	0.74	-0.10	-0.48	1.00	0.69	0.63
Sucres	0.85	0.19	-0.77	0.69	1.00	0.55
Divers	0.84	-0.27	-0.45	0.63	0.55	1.00

# ACP normée

Contribution des variables aux facteurs

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Animaux	24.39	4.48	1.33	10.18	2.40
Céréales	0.00	61.18	0.01	8.53	24.75
Légumes	15.48	22.17	6.67	0.00	55.66
Huiles	19.08	1.31	55.48	23.12	0.98
Sucres	22.79	3.11	2.88	32.36	9.67
Divers	18.25	7.75	33.62	25.81	6.53

PCA graph of variables

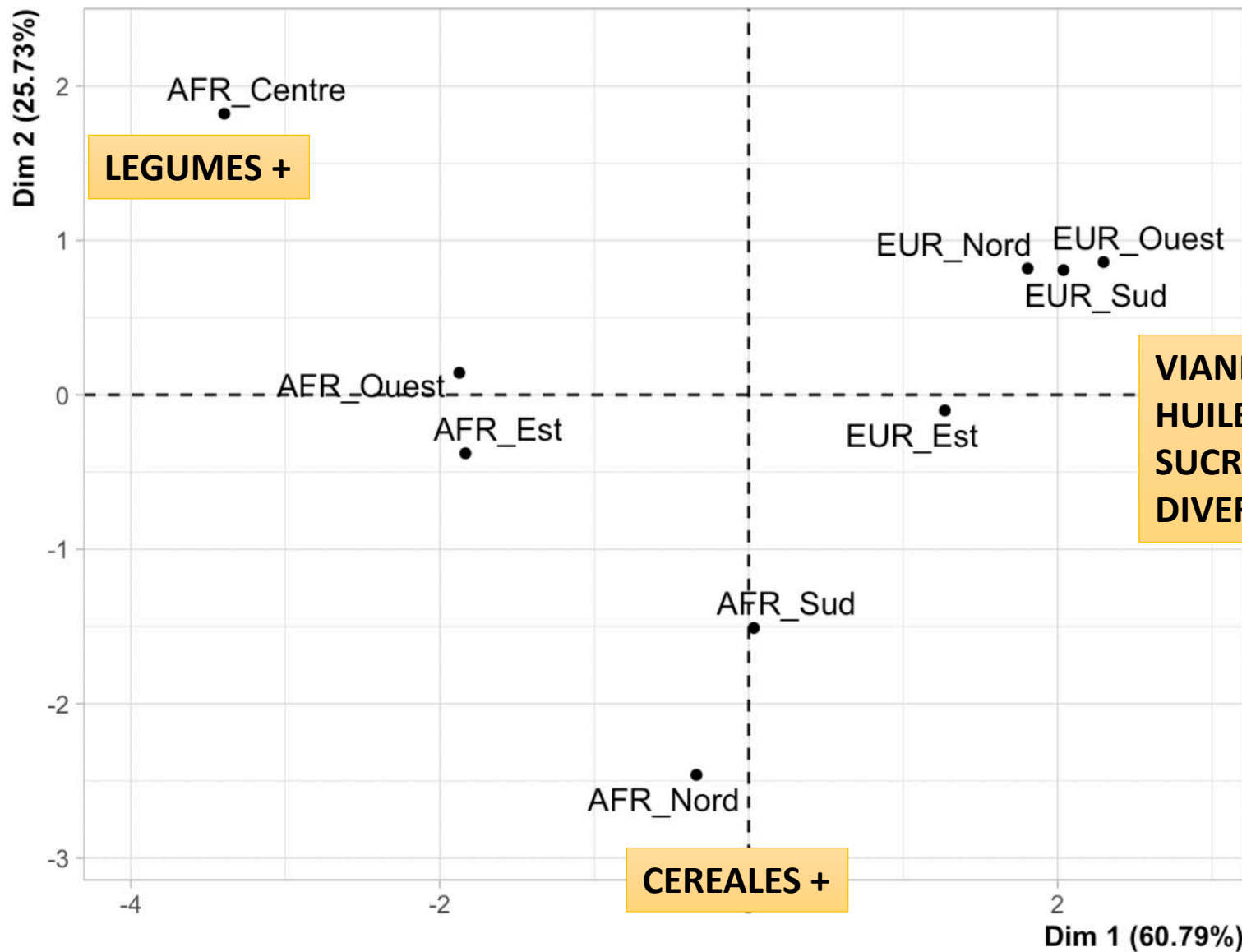


# ACP normée

Contribution des individus aux facteurs

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
AFR_Sud	0.00	16.41	11.35	6.72	40.35
AFR_Centre	35.13	23.87	7.13	11.77	0.46
AFR_Ouest	10.70	0.15	0.90	23.76	7.94
AFR_Est	10.25	1.03	42.27	2.16	15.64
AFR_Nord	0.35	43.62	1.73	0.02	23.93
EUR_Sud	12.65	4.70	16.55	26.73	3.99
EUR_Ouest	16.08	5.33	0.13	14.54	1.24
EUR_Est	4.91	0.07	6.61	11.74	6.38
EUR_Nord	9.94	4.82	13.34	2.56	0.08

PCA graph of individuals



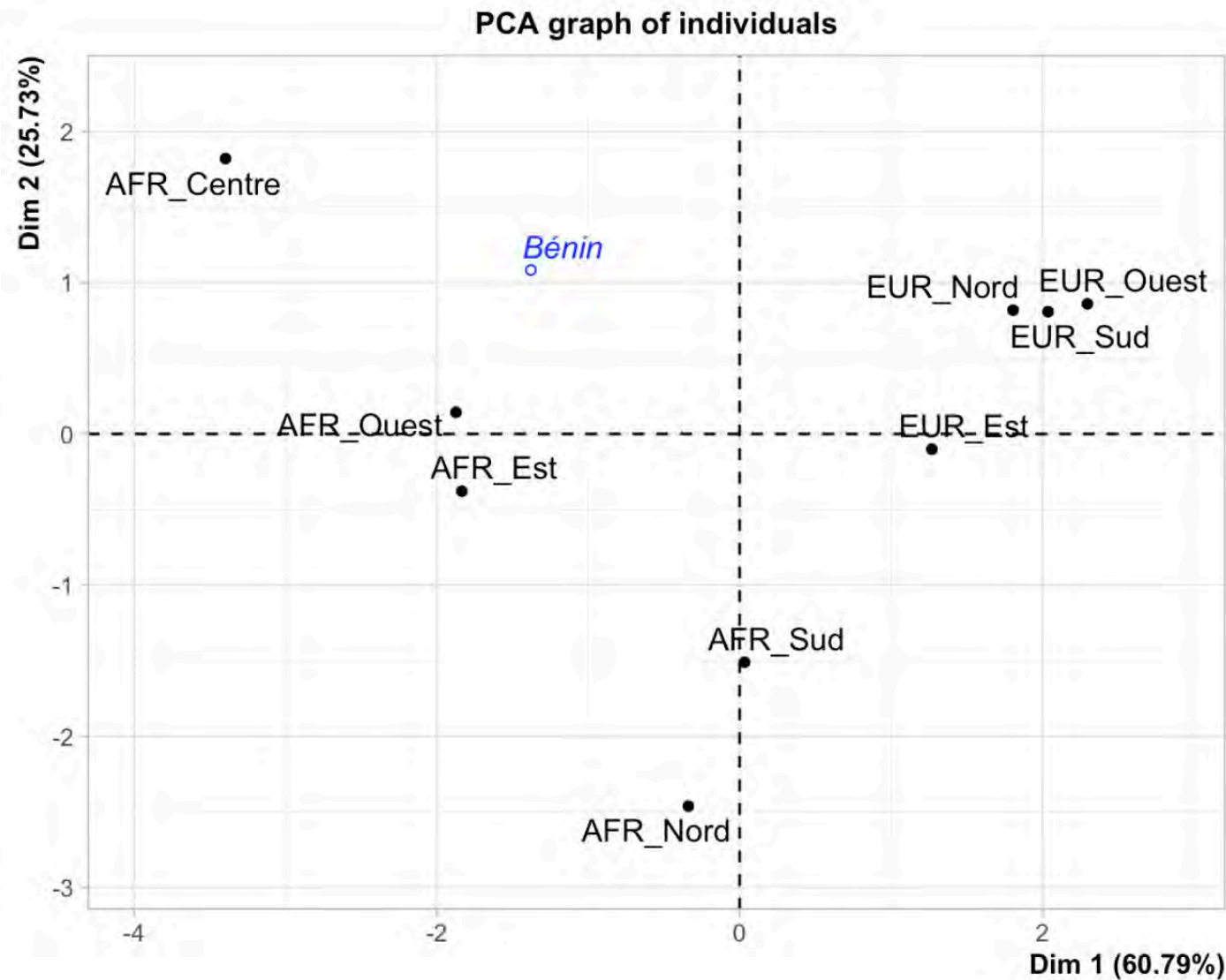
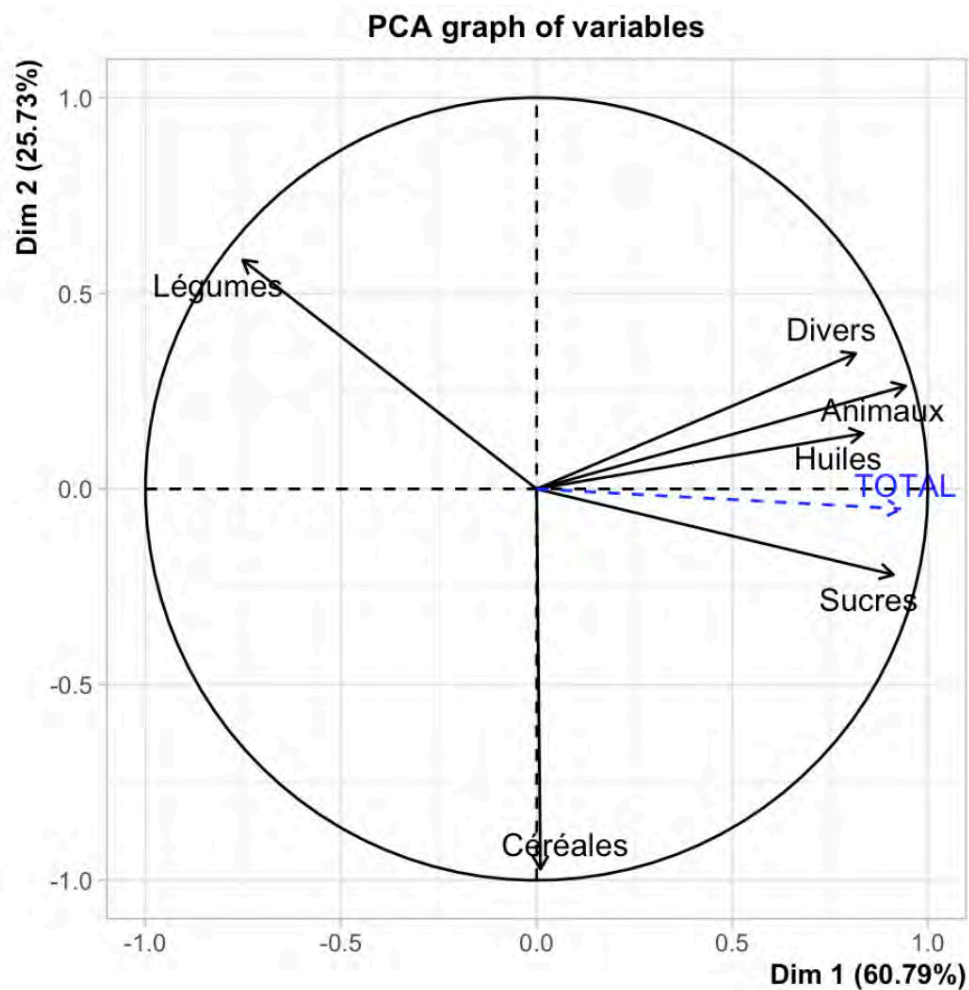
# ACP : Individus et variables supplémentaires

Ajout d'un individu et d'une variable supplémentaires

	<b>Animaux</b>	<b>Céréales</b>	<b>Légumes</b>	<b>Huiles</b>	<b>Sucres</b>	<b>Divers</b>	<b>TOTAL</b>
AFR_Sud	448	1358	114	421	300	161	<i>2802</i>
AFR_Centre	113	633	1048	249	76	138	<i>2257</i>
AFR_Ouest	120	1209	692	359	108	214	<i>2702</i>
AFR_Est	165	1178	380	197	107	229	<i>2256</i>
AFR_Nord	326	1819	243	363	304	183	<i>3238</i>
EUR_Sud	876	970	297	650	305	311	<i>3409</i>
EUR_Ouest	1202	882	294	506	408	300	<i>3592</i>
EUR_Est	897	1138	272	385	370	281	<i>3343</i>
EUR_Nord	1051	971	271	463	298	341	<i>3395</i>
<i>Bénin</i>	<i>113</i>	<i>1092</i>	<i>832</i>	<i>402</i>	<i>106</i>	<i>304</i>	<i>2849</i>

Individus et variables supplémentaires sont **projetés sur les axes sans les modifier.**

# ACP : Individus et variables supplémentaires



# ANALYSE FACTORIELLE DES CORRESPONDANCES (AFC)

# Analyse Factorielle des Correspondances

- ✓ Pourquoi correspondance ? Correspondance entre lignes et colonnes qui jouent un rôle symétrique dans l'analyse
- ✓ Sur un tableau d'effectifs dit *tableau de contingence* croisant deux variables qualitatives
- ✓ On analyse les deux tableaux de profils.
- ✓ La distance est celle du  $\chi^2$  (« chi-deux »).



# AFC : deux variables qualitatives

- ✓ Deux variables qualitatives V1 et V2 mesurées sur n individus (exemple : 12 départements du Bénin et langues parlées au Bénin)
- ✓ La variable « départements » a 12 modalités (lignes du tableau de contingence)
- ✓ La variable « langues » a 9 modalités (colonnes du tableau de données)
- ✓  $n_{\text{Zou Fgn}}$  = nombre d'individus qui résident dans le département Zou et qui déclarent être locuteur de la langue Fond, soit 546 personnes
- ✓  $N = \sum \sum n_{jk}$

# Tableau de contingence

	Fond	Adja	Yoruba	Bariba	Peul	Ottamari	Yoa	Dendi	Autres	
Alibori	16	0	10	177	94	5	0	214	5	521
Atacora	27	0	16	280	38	143	5	33	5	547
Atlantique	489	56	80	40	32	24	16	32	32	801
Borgou	51	7	58	275	210	29	14	65	14	723
Collines	150	11	273	0	32	27	27	0	16	536
Couffo	52	441	26	0	5	0	0	0	0	524
Donga	18	0	11	21	35	0	112	151	4	352
Littoral	319	40	120	53	20	27	7	33	47	666
Mono	76	227	22	4	18	4	0	4	7	362
Oueme	519	22	110	0	29	22	0	7	22	731
Plateau	85	28	273	0	12	8	0	0	0	406
Zou	546	24	18	0	6	0	0	0	6	600
	2348	856	1017	850	531	289	181	539	158	6769

# Nuage des profils-ligne

- ✓ Pour comparer les lignes entre elles, en transformant les fréquences en fréquences relatives
- ✓ Division de chaque ligne par la somme de la ligne
  - ✓  $n_{ij} / n_i$ .
- ✓ Nombre de personnes qui parlent Yoruba dans le département Atlantique = 227
- ✓ Fréquence de la langue Yoruba dans le département Atlantique =  $80/801 = 9,99 \%$  « pourcentages en ligne »)

# Tableau des profils-ligne

	Fond	Adja	Yoruba	Bariba	Peul	Ottamari	Yoa	Dendi	Autres	
Alibori	3,07	0,00	1,92	33,97	18,04	0,96	0,00	41,07	0,96	100,00
Atacora	4,94	0,00	2,93	51,19	6,95	26,14	0,91	6,03	0,91	100,00
Atlantique	61,05	6,99	9,99	4,99	4,00	3,00	2,00	4,00	4,00	100,00
Borgou	7,05	0,97	8,02	38,04	29,05	4,01	1,94	8,99	1,94	100,00
Collines	27,99	2,05	50,93	0,00	5,97	5,04	5,04	0,00	2,99	100,00
Couffo	9,92	84,16	4,96	0,00	0,95	0,00	0,00	0,00	0,00	100,00
Donga	5,11	0,00	3,13	5,97	9,94	0,00	31,82	42,90	1,14	100,00
Littoral	47,90	6,01	18,02	7,96	3,00	4,05	1,05	4,95	7,06	100,00
Mono	20,99	62,71	6,08	1,10	4,97	1,10	0,00	1,10	1,93	100,00
Oueme	71,00	3,01	15,05	0,00	3,97	3,01	0,00	0,96	3,01	100,00
Plateau	20,94	6,90	67,24	0,00	2,96	1,97	0,00	0,00	0,00	100,00
Zou	91,00	4,00	3,00	0,00	1,00	0,00	0,00	0,00	1,00	100,00

# Nuage des profils-colonne

- Pour comparer les colonnes entre elles en transformant les fréquences en fréquences relatives
- Division de chaque colonne par la somme de la colonne

$$n_{ij} / n_{.j}$$

- Nombre de personnes qui parlent Adja dans le Mono = 227
- Fréquence de la langue Yoruba dans le département Atlantique =  $227/856 = 26,52 \%$  « pourcentages en colonne »)

# Tableau des profils-colonne

	Fond	Adja	Yoruba	Bariba	Peul	Ottamari	Yoa	Dendi	Autres
Alibori	0,68	0,00	0,98	20,82	17,70	1,73	0,00	39,70	3,16
Atacora	1,15	0,00	1,57	32,94	7,16	49,48	2,76	6,12	3,16
Atlantique	20,83	6,54	7,87	4,71	6,03	8,30	8,84	5,94	20,25
Borgou	2,17	0,82	5,70	32,35	39,55	10,03	7,73	12,06	8,86
Collines	6,39	1,29	26,84	0,00	6,03	9,34	14,92	0,00	10,13
Couffo	2,21	51,52	2,56	0,00	0,94	0,00	0,00	0,00	0,00
Donga	0,77	0,00	1,08	2,47	6,59	0,00	61,88	28,01	2,53
Littoral	13,59	4,67	11,80	6,24	3,77	9,34	3,87	6,12	29,75
Mono	3,24	26,52	2,16	0,47	3,39	1,38	0,00	0,74	4,43
Oueme	22,10	2,57	10,82	0,00	5,46	7,61	0,00	1,30	13,92
Plateau	3,62	3,27	26,84	0,00	2,26	2,77	0,00	0,00	0,00
Zou	23,25	2,80	1,77	0,00	1,13	0,00	0,00	0,00	3,80
	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

# Démarche de l'AFC

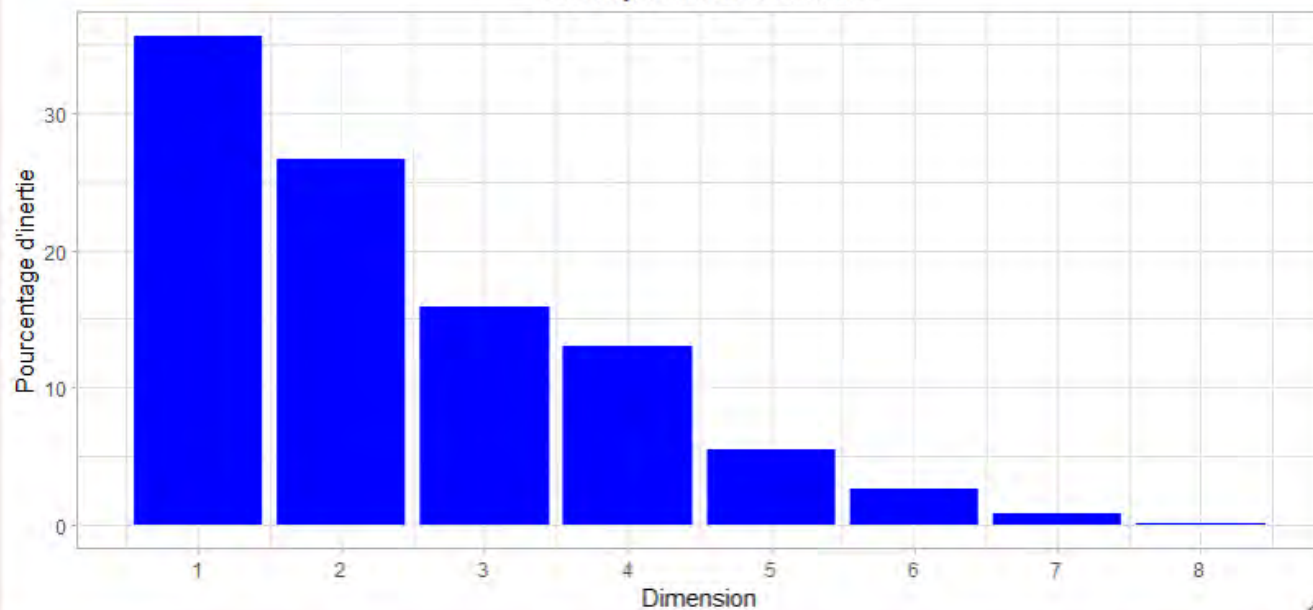
A partir d'un tableau de contingence :

- ✓ On fait une ACP sur le nuage des profils-lignes ( $N_l$ ) avec la distance du « chi-deux ».

$$d_{\chi^2}^2(\mathbf{i}, \mathbf{l}) = \sum_{j=1}^q \frac{1}{f_{.j}} \left( \frac{f_{ij}}{f_{i.}} - \frac{f_{lj}}{f_{l.}} \right)^2$$

- ✓ L'inertie totale du nuage de points lignes est égale au Chi2 (différence avec l'ACP normée où l'inertie du nuage est égale au nombre de variables = dépend du format des données et non des données elles-mêmes)
- ✓ On fait une ACP des profils-colonnes pondérés chacun par  $n_j / n$  et avec la distance du « chi-deux ».
- ✓ On procède ensuite pour les deux nuages comme pour l'ajustement du nuage des individus en ACP
- ✓ équivalent à une double ACP avec un distance du CHI2
- ✓ Principe d'équivalence distributionnelle = on peut déduire les caractéristiques lignes des caractéristiques des colonnes à un coefficient près  $1/\sqrt{\lambda}$

Décomposition de l'inertie

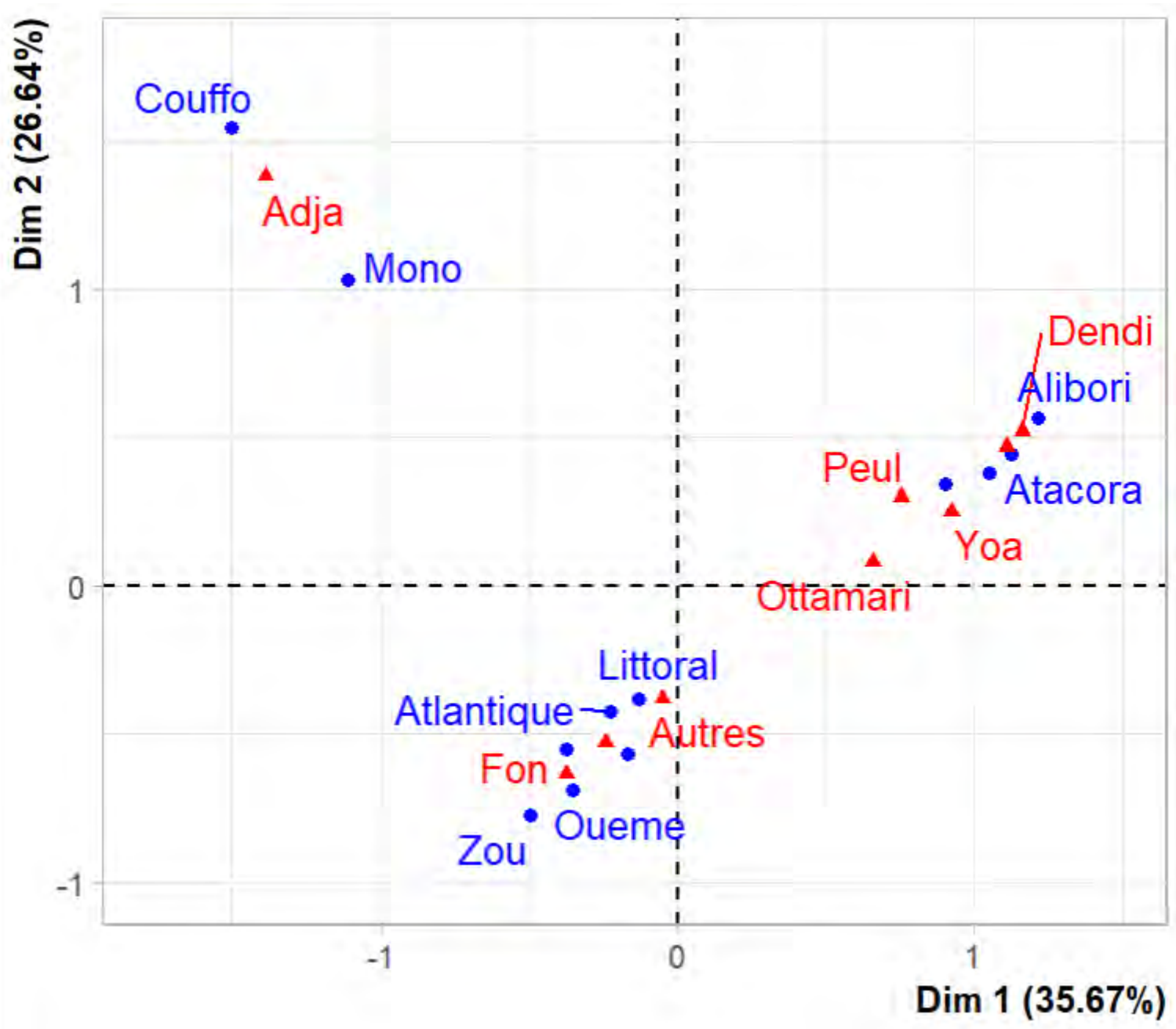


# Résultats de l'AFC

**eigenvalue    percentage of variance    cumulative percentage of variance**

dim 1	0.65	35.67	35.67
dim 2	0.49	26.64	62.32
dim 3	0.29	15.86	78.18
dim 4	0.24	12.94	91.12
dim 5	0.10	5.47	96.59
dim 6	0.05	2.58	99.18
dim 7	0.01	0.74	99.91
dim 8	0.00	0.09	100.00





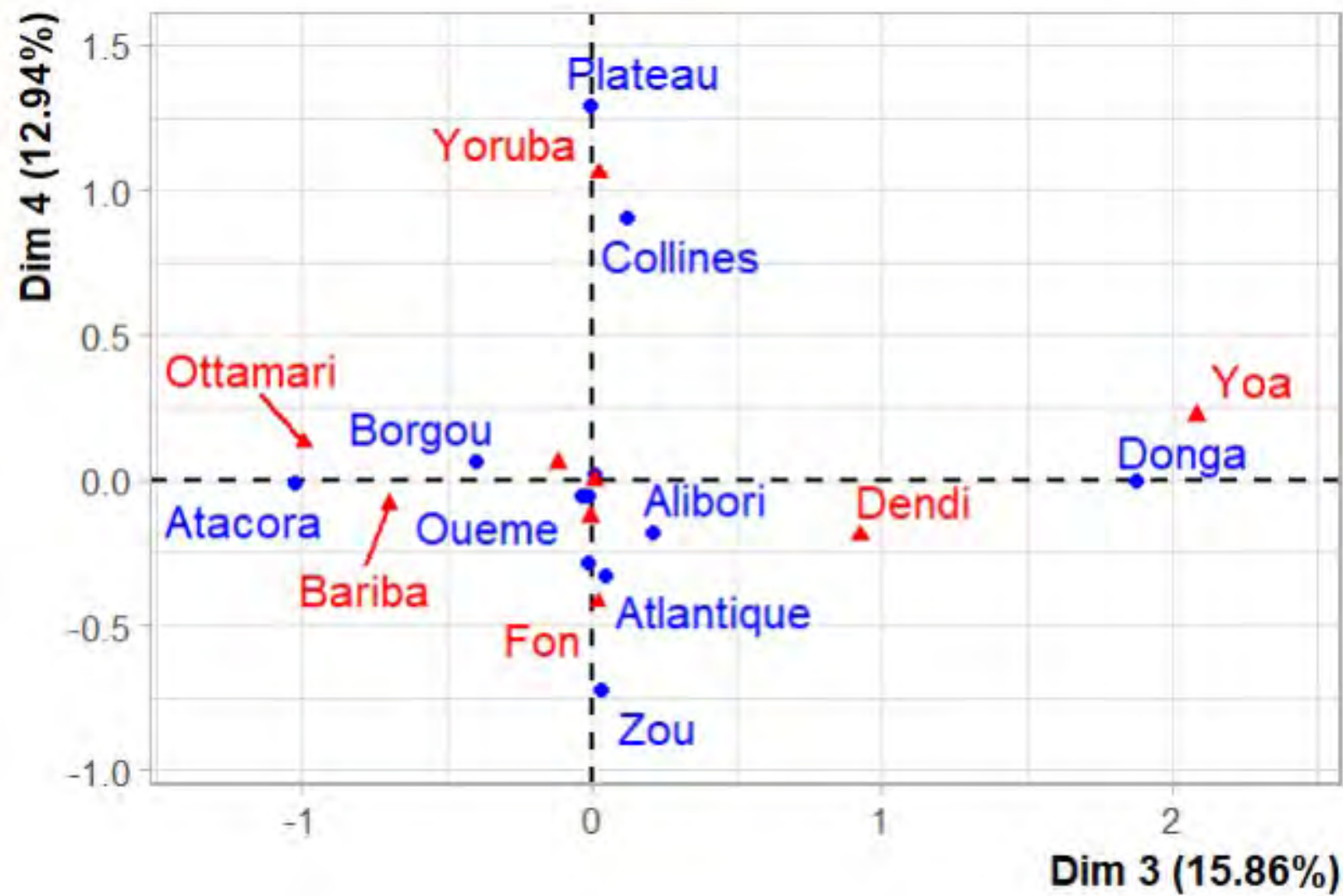
Premier plan factoriel

# Contribution des langues à la formation des axes

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Fon	7.56	28.81	0.05	26.12	0.00
Adja	37.51	49.75	0.00	0.00	0.08
Yoruba	1.37	8.72	0.02	71.16	1.19
Bariba	23.81	5.69	21.35	0.36	0.06
Peul	6.87	1.46	0.38	0.11	20.08
Ottamari	2.85	0.05	14.52	0.31	41.07
Yoa	3.50	0.34	40.18	0.56	31.02
Dendi	16.52	4.48	23.50	1.22	6.50
Autres	0.01	0.70	0.00	0.16	0.00

# Qualité de représentation des différentes langues sur les facteurs

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Fon	0.20	0.56	0.00	0.25	0.00
Adja	0.50	0.50	0.00	0.00	0.00
Yoruba	0.04	0.19	0.00	0.76	0.01
Bariba	0.63	0.11	0.25	0.00	0.00
Peul	0.49	0.08	0.01	0.00	0.22
Ottamari	0.18	0.00	0.40	0.01	0.39
Yoa	0.13	0.01	0.64	0.01	0.17
Dendi	0.48	0.10	0.30	0.01	0.03
Autres	0.00	0.20	0.00	0.02	0.00



# AFC : proximités entre modalités ligne et colonne

Deux modalités de la même variable sont proches, si leurs profils sont similaires. Deux points ligne départements sont proches signifie que le profil de répartition des langues est proche au sein des différents départements. Deux points colonne (langues) sont proches si la distribution des locuteurs selon les départements est proche.

# Éléments supplémentaires

- ✓ Supplémentaire = inactif = n'ayant pas participé à la détermination des axes factoriels du nuage
- ✓ Replacé a posteriori dans l'espace
- ✓ Permet d'éclairer certains aspects de l'analyse
- ✓ Dans une démarche exploratoire de données d'enquête, on peut faire le choix de mettre un ensemble de réponses sur une même thématique en actif et les variables sociodémo comme illustratives

# ANALYSE FACTORIELLE DES CORRESPONDANCES MULTIPLES (ACM)

# ACM

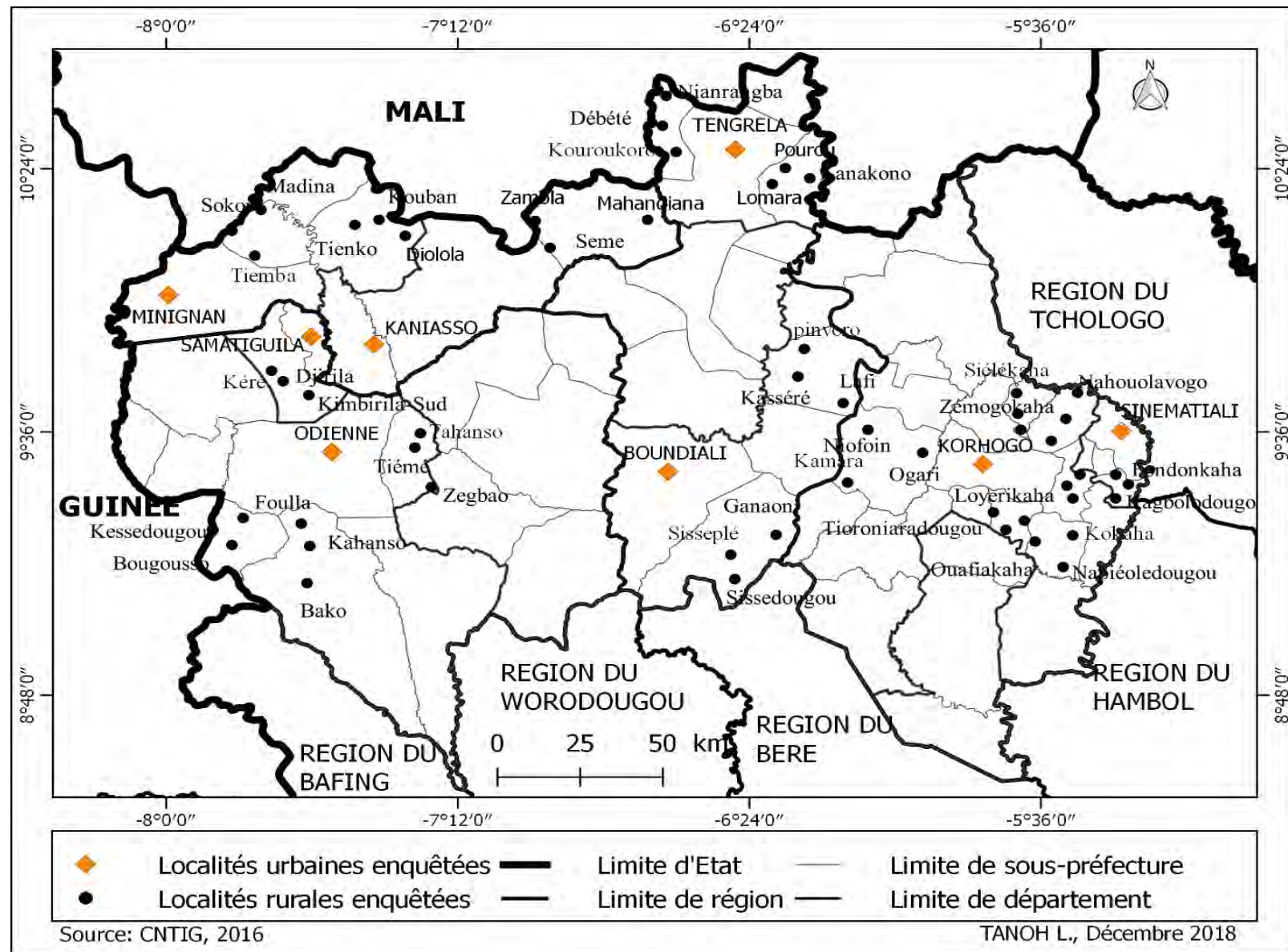
- Généralisation de l'AFC  
à plus de deux variables qualitatives
- Tableau de BURT  
= généralisation du tableau de contingence



# Cadre de l'étude et localités enquêtées



4 Régions administratives;  
 14 départements ;  
 63 sous-préfectures;  
 1 134 villages et  
 209 746 ménages;  
 635 ménages enquêtés



# Sélection de variables

Les variables retenues sont organisées en **variables actives** et **variables supplémentaires**

## A. Scolarisation

TRA : les enfants travaillent (O/N)

MMQ : la famille manque de moyens financiers (O/N) –

AID : les enfants aident la famille (O/N)

ECO : il y a une école dans la localité (O/N)

PAR : les enfants sont scolarisés de façon paritaire (O/N)

**VARIABLES ACTIVES**

## B. Contexte familial

SEX : sexe du chef de ménage (H/F)

EDU : niveau d'éducation du chef de ménage (AUC = aucun, PRI = primaire, SEC = secondaire ou supérieur)

## C. Contexte spatial

MIL : milieu urbain ou rural

DEP : Département

**VARIABLES SUPPLEMENTAIRES**

# Données de l'enquête

## Tableau de données original

TRA	MMF	AID	ECO	PAR	SEX	EDU	REG	MIL
O	O	N	Non	Oui	F	PRI	Poro	Urb
N	O	O	Oui	Non	H	PRI	Poro	Urb
N	O	O	Oui	Non	H	AUC	Poro	Urb
N	O	O	Oui	Non	F	AUC	Poro	Urb
N	O	O	Oui	Non	F	AUC	Poro	Urb
N	O	O	Oui	Non	H	PRI	Poro	Urb

## Tableau disjonctif complet

On doit transformer ce tableau initial en tableau disjonctif complet pour réaliser une analyse des correspondances multiples. L'opération se fait de façon implicite dans FactoMineR mais on peut la rendre visible.

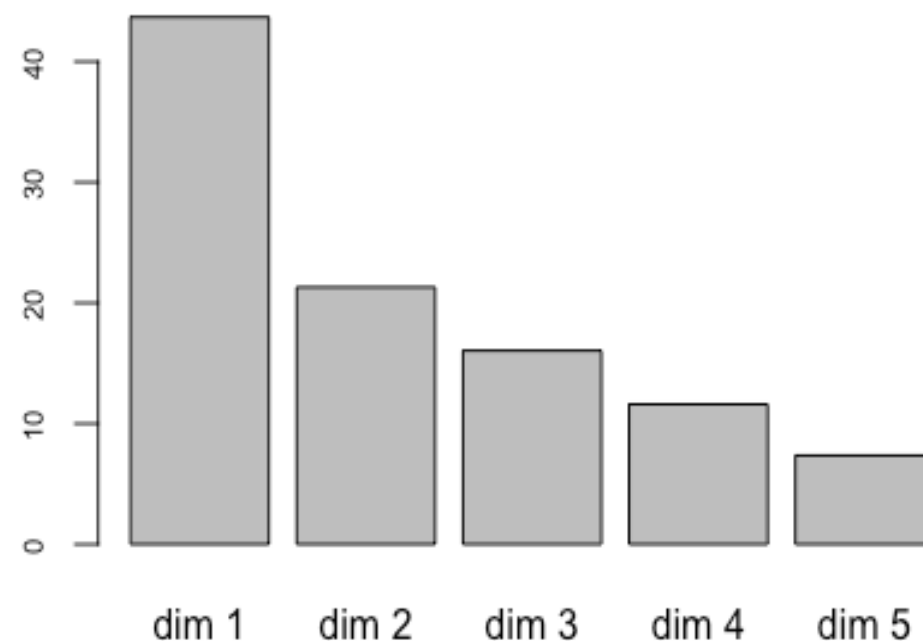
##	TRA_O	TRA_N	MMF_O	MMF_N	AID_O	AID_N	ECO_Oui	ECO_Non	PAR_Oui	PAR_Non	H	F	AUC
## 1	1	0	1	0	0	1	0	1	1	0	0	1	0
## 2	0	1	1	0	1	0	1	0	0	1	1	0	0
## 3	0	1	1	0	1	0	1	0	0	1	1	0	1
## 4	0	1	1	0	1	0	1	0	0	1	0	1	1
## 5	0	1	1	0	1	0	1	0	0	1	0	1	1
## 6	0	1	1	0	1	0	1	0	0	1	1	0	0
##	PRI	SEC	Bagoué	Folon	Kabadougou	Porro	Urb	Rur					
## 1	1	0	0	0	0	1	1	0					
## 2	1	0	0	0	0	1	1	0					
## 3	0	0	0	0	0	1	1	0					
## 4	0	0	0	0	0	1	1	0					
## 5	0	0	0	0	0	1	1	0					
## 6	1	0	0	0	0	1	1	0					

## Valeurs propres

	Valeur propre	% variance	% variance cumulée
dim 1	0.44	43.69	43.69
dim 2	0.21	21.30	64.99
dim 3	0.16	16.05	81.04
dim 4	0.12	11.60	92.64
dim 5	0.07	7.36	100.00

% information

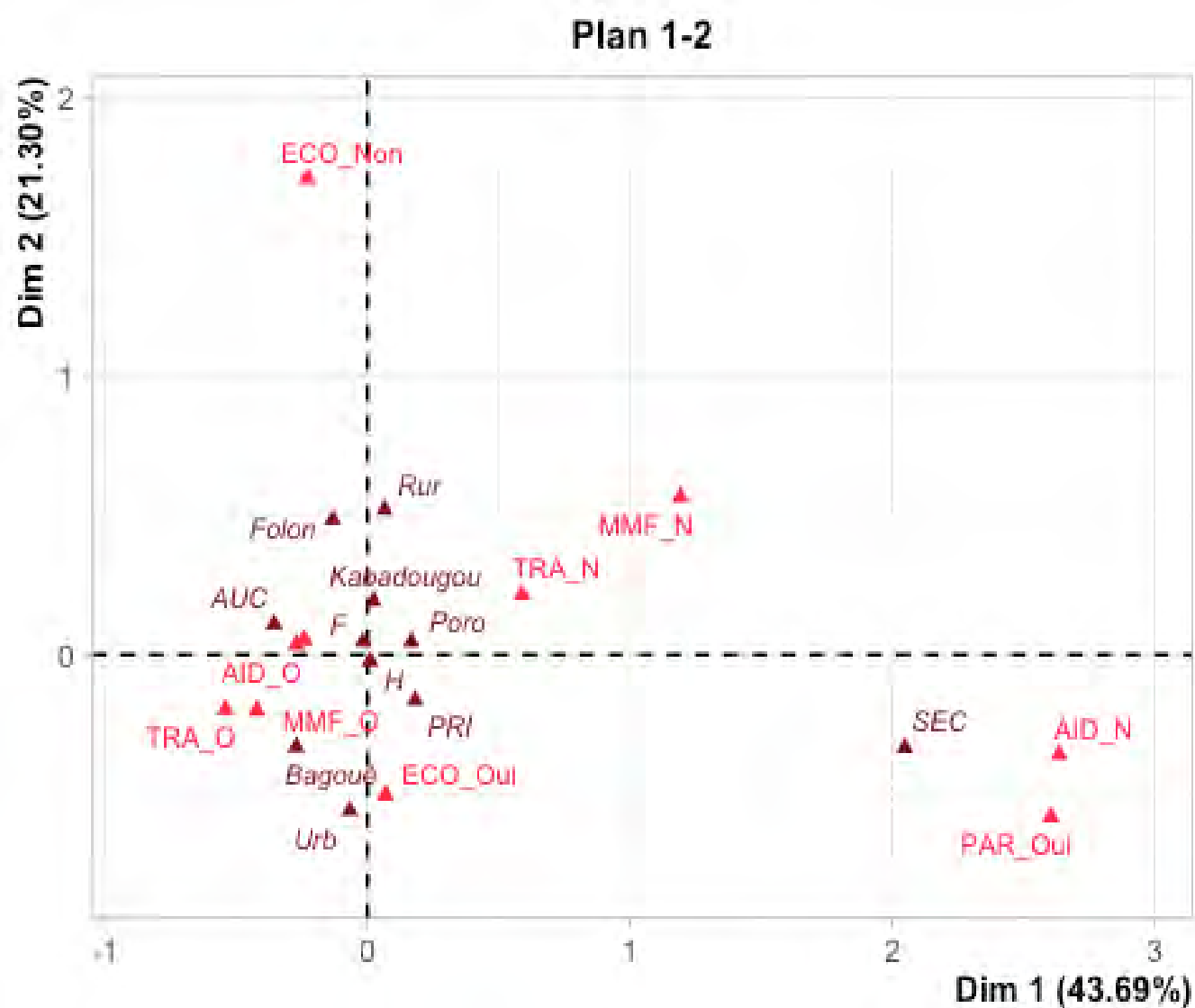
## Hierarchie des Facteurs



## Coordonnées et contribution des variables aux axes

	coo_1	ctr_1	coo_2	ctr_2
TRA_O	-0.54	7.02	-0.20	1.90
TRA_N	0.59	7.60	0.21	2.06
MMF_O	-0.42	6.04	-0.20	2.78
MMF_N	1.19	17.04	0.56	7.84
AID_O	-0.27	3.06	0.04	0.12
AID_N	2.64	29.69	-0.36	1.13
ECO_Oui	0.07	0.16	-0.51	18.54
ECO_Non	-0.23	0.55	1.71	62.67
PAR_Oui	2.60	26.38	-0.58	2.71
PAR_Non	-0.24	2.45	0.05	0.25

## Graphiques des variables



# ACM : propriétés

- A un coefficient près, une modalité est le centre de gravité des individus qui correspondent à cette modalité.
- Les modalités d'une même variable forment un sous-nuage, dont le centre de gravité est l'origine.
- La part d'inertie due à une modalité est d'autant plus grande que son effectif est faible (modalité rare)
- La part d'inertie due à une variable est d'autant plus grande que le nombre de modalités de la variable est grand !

# EXERCICES D'APPLICATION

Clé USB : MODULES / EXP3\_StatMulti