

Analyse factorielle des correspondances des causes de décès par pays

Mkrtchyan, Toure

Table of contents

1	Introduction	2
2	Analyse factorielle des correspondances des causes de décès par pays	2
2.1	Préparation des données	2
3	Analyse exploratoire	4
3.1	Chargement des données	4
3.3	Visualisation des données par barplots	4
3.3.1	Arménie :	4
3.3.2	Mali :	5
3.3.3	France :	6
3.4	Comparaison de pays	6
3.5	Profils ligne et colonne	7
3.6	Comparaison des pays, graphiques de profils ligne et colonnes avec ggplot2	8
4	Analyses avancées (AFC et Tests)	9
4.1	Test du khi-deux et V de Cramer	9
4.2	Lancement de l'AFC	10
4.3	Critère du bâton brisé pour sélectionner les axes	14
4.4	Analyse des contributions des lignes et colonnes	16
4.5	Distances au centre de gravité	31
5	Conclusion	33
6	Références	34
6.1	Articles, manuels R, sources des données	34

1 Introduction

En 2019, en Afrique subsaharienne, les maladies non transmissibles représentent environ 35 à 40 % des décès chez les adultes, d'après les données mondiales de l'OMS. Intuitivement, les infections pourraient pourtant être considérées comme dominantes, compte tenu du niveau de développement des pays. Cela soulève plusieurs questions centrales pour l'analyse : *Existe-t-il un lien entre le niveau de développement des pays et différentes causes de décès ? Quels pays présentent des profils similaires de mortalité ?* Afin d'y répondre rigoureusement, il est pertinent de réaliser une Analyse Factorielle des Correspondances (AFC) sur les données enregistrées.

2 Analyse factorielle des correspondances des causes de décès par pays

2.1 Préparation des données

URL : <https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world>

Pour cette analyse, nous avons utilisé un jeu de données disponible sur Kaggle, portant sur les causes de mortalité dans le monde <https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world>. Le jeu de données original couvre la période 1990 à 2019 et recense les effectifs de décès pour 204 pays et territoires, organisés sous la forme d'un tableau pays × causes de décès.

Afin de concentrer l'étude sur l'année la plus récente, nous avons extrait uniquement les observations de 2019. Une vérification préalable a confirmé l'absence de valeurs manquantes dans les variables retenues. Les données extraites ont ensuite été enregistrées dans un nouveau fichier CSV, distinct du fichier initial.

La variable correspondant à l'année a été supprimée du fait que toutes les lignes concernent 2019. De même, parmi les deux colonnes descriptives des pays (nom complet et abréviation), seule la colonne des abréviations a été conservée, les correspondances avec les noms complets étant stockées dans un fichier séparé.

Pour améliorer la lisibilité et faciliter l'interprétation graphique dans l'Analyse Factorielle des Correspondances (AFC), nous avons procédé à un encodage des différentes causes de décès, chaque cause étant associée à un identifiant plus court. Cet encodage a été enregistré dans un document séparé, qui sera également fourni dans le rapport.

Le jeu de données final comporte ainsi 204 lignes correspondant aux abréviations des pays et territoires, et des colonnes représentant les différentes causes de décès encodées, avec pour

valeurs les effectifs observés. En étant deux variables qualitatives il convient de procéder à une AFC.

Table 1: Encodage des causes de décès

Meningitis	C1
Alzheimer's Disease and Other Dementias	C2
Parkinson's Disease	C3
Nutritional Deficiencies	C4
Malaria	C5
Drowning	C6
Interpersonal Violence	C7
Maternal Disorders	C8
HIV/AIDS	C9
Drug Use Disorders	C10
Tuberculosis	C11
Cardiovascular Diseases	C12
Lower Respiratory Infections	C13
Neonatal Disorders	C14
Alcohol Use Disorders	C15
Self-harm	C16
Exposure to Forces of Nature	C17
Diarrheal Diseases	C18
Environmental Heat and Cold Exposure	C19
Neoplasms	C20
Conflict and Terrorism	C21
Diabetes Mellitus	C22
Chronic Kidney Disease	C23
Poisonings	C24
Protein-Energy Malnutrition	C25
Road Injuries	C26
Chronic Respiratory Diseases	C27
Cirrhosis and Other Chronic Liver Diseases	C28
Digestive Diseases	C29
Fire, Heat, and Hot Substances	C30
Acute Hepatitis	C31

3 Analyse exploratoire

3.1 Chargement des données

La première étape de l'analyse consiste à charger les librairies nécessaires ainsi que le jeu de données à analyser dans le script R qui est le fichier CSV nettoyé. Après l'importation, une première visualisation globale du tableau est réalisée afin de vérifier la structure des données et la cohérence des variables.

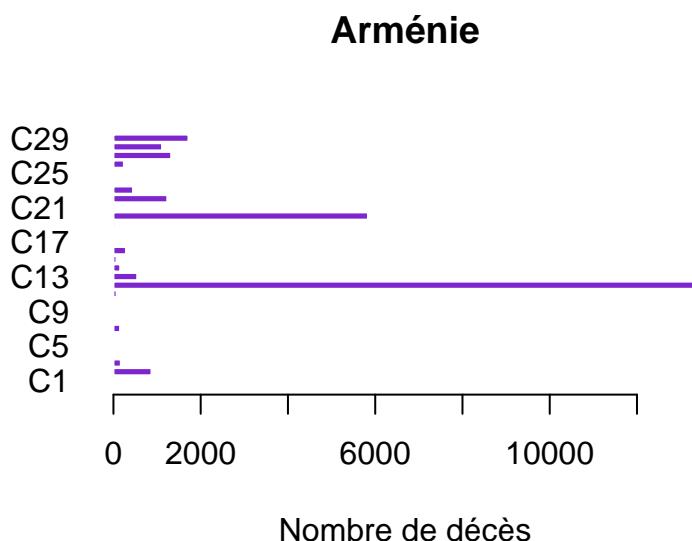
Au moment du chargement de données, la première colonne du jeu de données est désigné comme étant la colonne d'identification des lignes et elle est ensuite supprimée, afin de ne conserver que les causes de décès sous forme numérique. Cette étape permet de faciliter les traitements ultérieurs et l'accès direct aux informations par pays.

Le jeu de données obtenu est stocké dans la variable “morts” pour le reste de l'analyse. Une vérification du jeu de données à l'aide de l'affichage des premières lignes confirme la bonne manipulation des données et l'absence d'anomalies apparentes.

3.2

3.3 Visualisation des données par barplots

3.3.1 Arménie :



3.3.2 Mali :

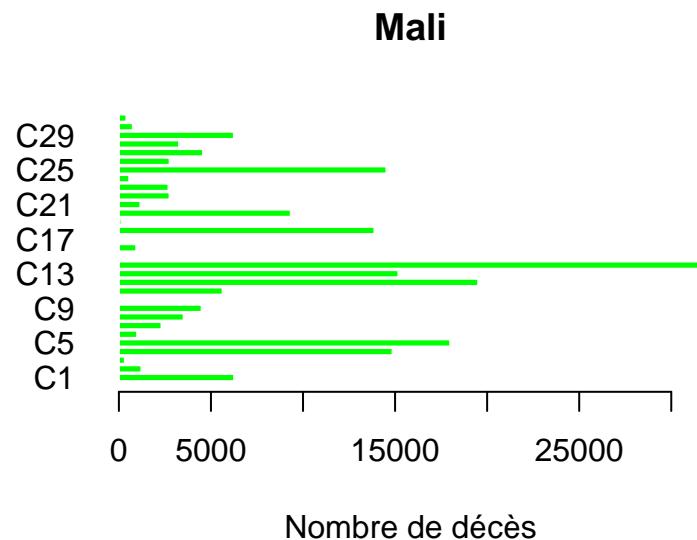
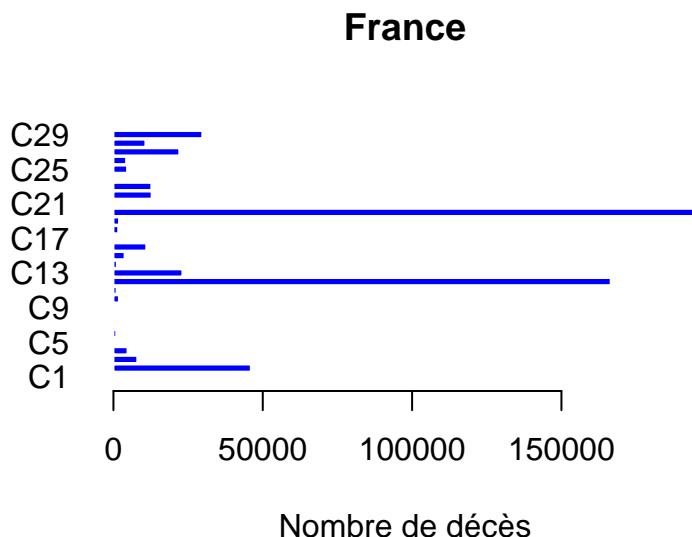


Figure 1: Répartition des causes de décès au Mali en 2019

3.3.3 France :



3.4 Comparaison de pays

Les Figures 1 à 3 présentent des diagrammes en barres horizontales illustrant le nombre de décès par cause respectivement en Arménie, au Mali et en France. Ces représentations graphiques ont pour objectif de comparer la structure de la mortalité selon les principales causes de décès dans chacun des pays étudiés.

Les causes de décès sont représentées sur l'axe des ordonnées, tandis que l'axe des abscisses indique le nombre de morts. Ce choix de visualisation permet une lecture claire et facilite la comparaison entre catégories.

L'analyse met en évidence des profils de mortalité contrastés. En Arménie, en 2019, les maladies cardiovasculaires (C12) sont responsables de plus de 12 000 décès, tandis que le néoplasme (C20) en représente environ 6 000. En France, ces mêmes causes figurent également parmi les principales causes de mortalité, avec plus de 15 000 décès pour le néoplasme et environ 20 000 pour la cause cardiovasculaires. À l'inverse, au Mali, les maladies transmissibles ainsi que certaines maladies infectueuses occupent une place plus importante dans la mortalité totale : les entraînent environ 32 000 décès, les maladies cardiovasculaires (C12) près de 20 000 décès, et l'automutilation et la malnutrition (C18 et C25) environ 14 000 décès chacune.

Cette analyse descriptive constitue une première étape exploratoire. Elle permet de dégager des tendances générales et justifie la poursuite de l'étude par des analyses comparatives plus approfondies entre pays et groupes de causes de décès.

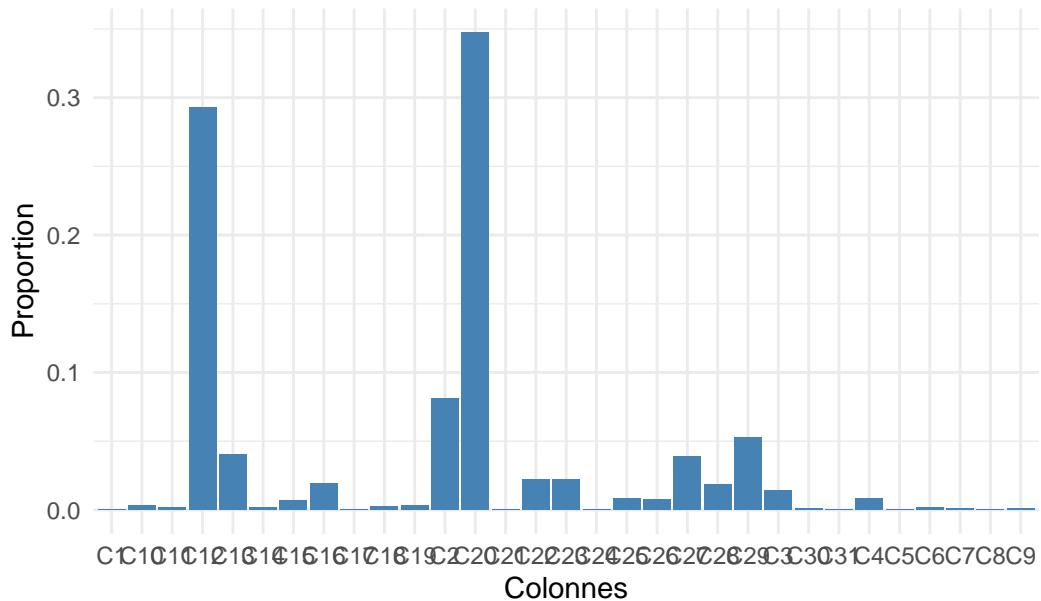
3.5 Profils ligne et colonne

Afin d'approfondir l'analyse descriptive, des profils lignes et des profils colonnes ont été construits à partir du tableau de contingence des décès par cause et par pays. Le profil ligne d'un pays correspond à la répartition relative des causes de décès à l'intérieur de ce pays : chaque valeur représente la proportion d'une cause donnée parmi l'ensemble des décès du pays considéré. Par exemple, en Afghanistan la meningite reperesente 0.7% des causes des décès en 2019 ($0.007 = 1563/(1563+1775+\dots+485 +1940)$). Ainsi, la somme des proportions sur une ligne est égale à 1, ce qui permet de comparer les structures de mortalité indépendamment du niveau absolu de mortalité.

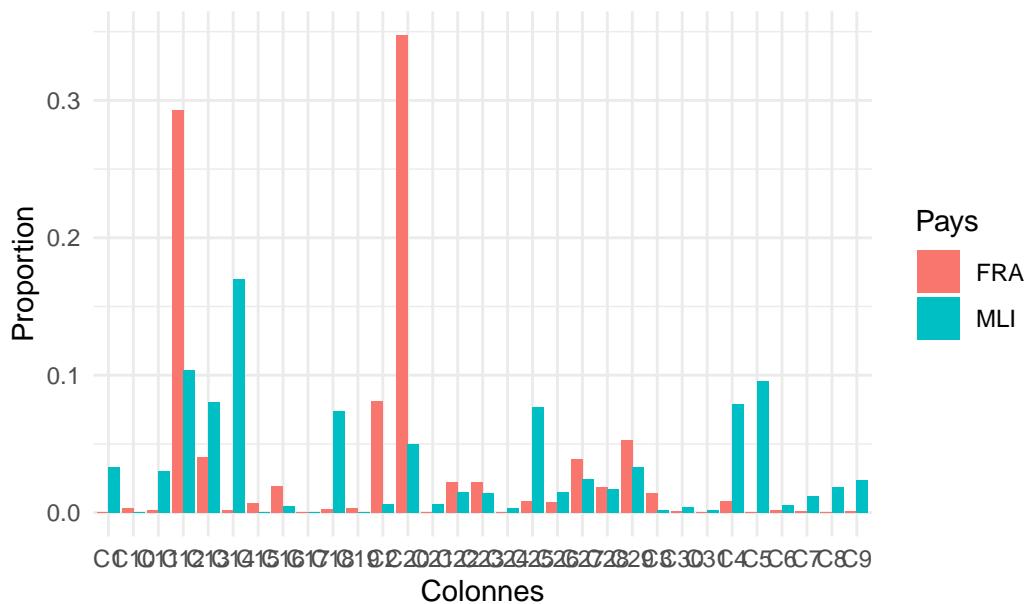
De manière complémentaire, les profils colonnes décrivent, pour chaque cause de décès, la répartition relative des pays. Chaque valeur indique la contribution d'un pays donné au total des décès observés pour une cause spécifique . En 2019, l'Afghanistan représentait 0,66 % ($0.0066= 1563/(1563 +13+\dots+2065+1450)$)de la mortalité mondiale due à la méningite. Là encore, la normalisation par colonne permet de comparer les pays entre eux pour une cause donnée, sans être influencé par les différences globales de population ou de mortalité.

3.6 Comparaison des pays, graphiques de profils ligne et colonnes avec ggplot2

Profil ligne de la France (FRA)



Comparaison des profils ligne : France vs Mali



Les profils lignes sont ensuite représentés graphiquement afin de faciliter l'interprétation. Le profil ligne de la France met en évidence la prédominance des maladies non infectueuses dans

la structure de la mortalité, confirmant les observations issues de l'analyse par diagrammes en barres. Une comparaison des profils lignes de la France et du Mali souligne des contrastes marqués : la France présente une concentration plus forte des décès sur les maladies chroniques, tandis que le Mali affiche une part relativement plus importante de maladies infectueuses et de causes évitables. Ces différences traduisent des stades distincts de transition épidémiologique et fournissent une base pertinente pour des analyses multivariées ultérieures, telles que l'analyse des correspondances.

3.7

4 Analyses avancées (AFC et Tests)

4.1 Test du khi-deux et V de Cramer

Afin d'évaluer statistiquement l'existence d'un lien entre les pays et les causes de décès, un test du khi-deux d'indépendance est réalisé à partir du tableau de contingence. L'hypothèse nulle H₀ suppose l'indépendance entre le pays et la cause de décès, tandis que l'hypothèse alternative H₁ postule l'existence d'une dépendance entre ces deux variables.

Le test met en évidence une statistique du khi-deux très élevée (36 396 969) associée à une p-value nulle, conduisant au rejet de l'hypothèse d'indépendance au seuil usuel. Cela suggère que la répartition des causes de décès diffère significativement selon les pays considérés. Toutefois, l'interprétation de ce résultat doit être nuancée, car la structure du tableau, notamment la présence de nombreux effectifs élevés et de fortes disparités entre pays limite la pertinence stricte du test du khi-deux.

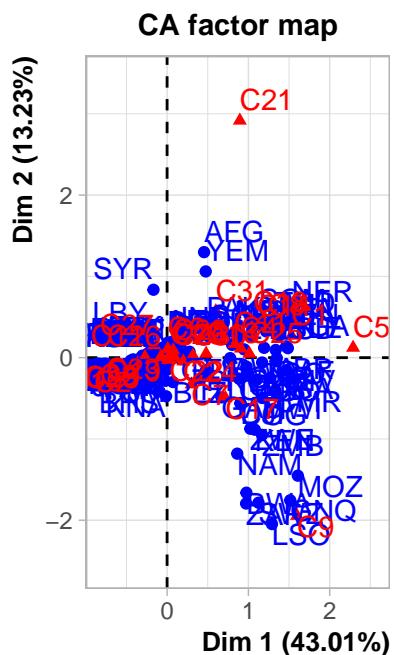
Afin de quantifier l'intensité de la liaison entre les deux variables, le coefficient de V de Cramér est calculé. La valeur obtenue étant 0.149 indique une liaison faible entre le pays et la cause de décès, malgré le rejet de l'hypothèse d'indépendance. Ce résultat souligne que la dépendance statistique observée est réelle mais modérée.

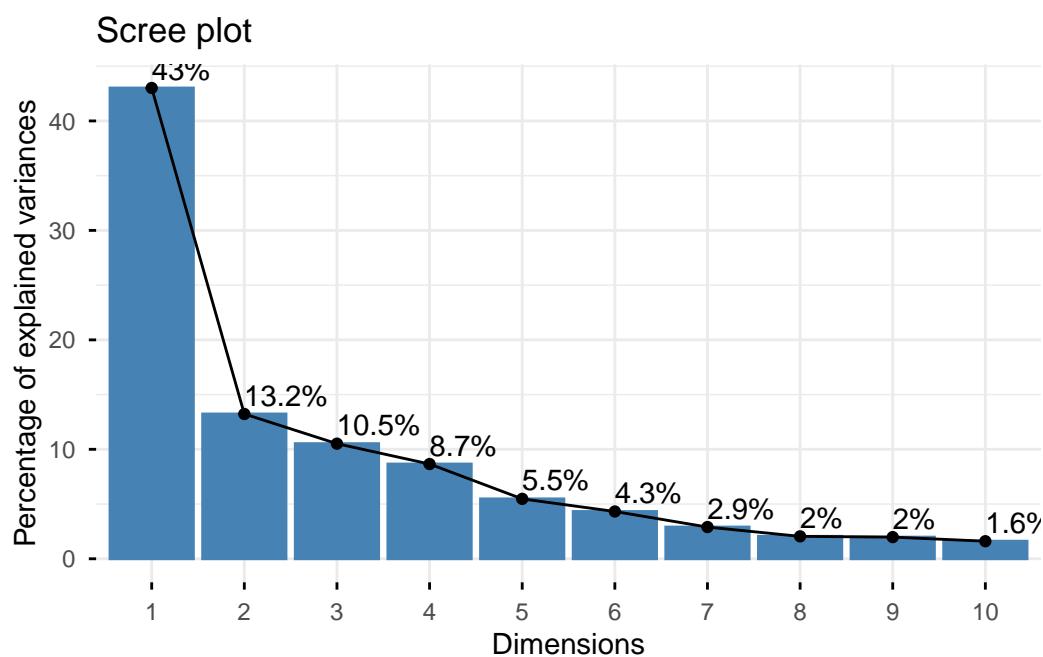
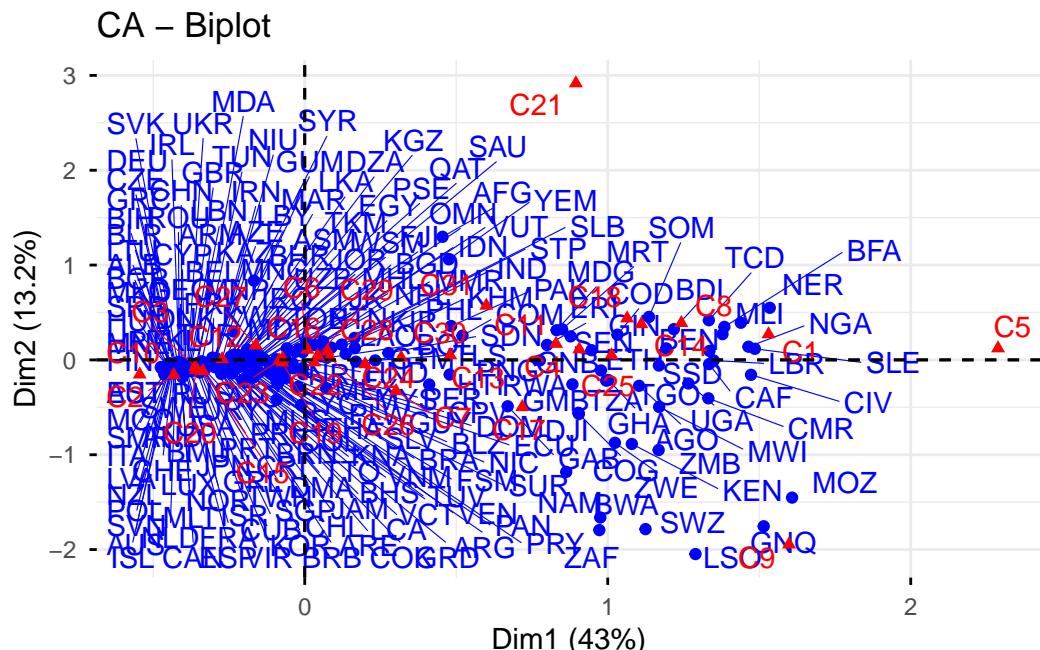
L'examen des effectifs attendus sous l'hypothèse d'indépendance ainsi que des résidus standardisés permet d'identifier les principales contributions au khi-deux. Ces écarts traduisent des phénomènes d'attraction et de répulsion : un résidu standardisé positif indique qu'un pays et une cause se rencontrent plus fréquemment que prévu (attraction), tandis qu'un résidu standardisé négatif montre une occurrence moindre que prévu (répulsion). Un résidu supérieur à 2 ou inférieur à -2 est généralement considéré comme significatif.

À titre d'illustration, le Bangladesh (BGD) présente un résidu standardisé de -99,09 pour la malaria (C5), ce qui traduit une répulsion extrêmement marquée : cette cause est très fortement sous-représentée dans ce pays par rapport à ce qui serait attendu sous l'hypothèse d'indépendance. À l'inverse, l'Afghanistan (AFG) affiche un résidu standardisé très élevé

(116,45) pour les troubles maternels (C8), indiquant une attirance très forte entre ce pays et cette cause de décès, largement surreprésentée par rapport au modèle d'indépendance.

4.2 Lancement de l'AFC

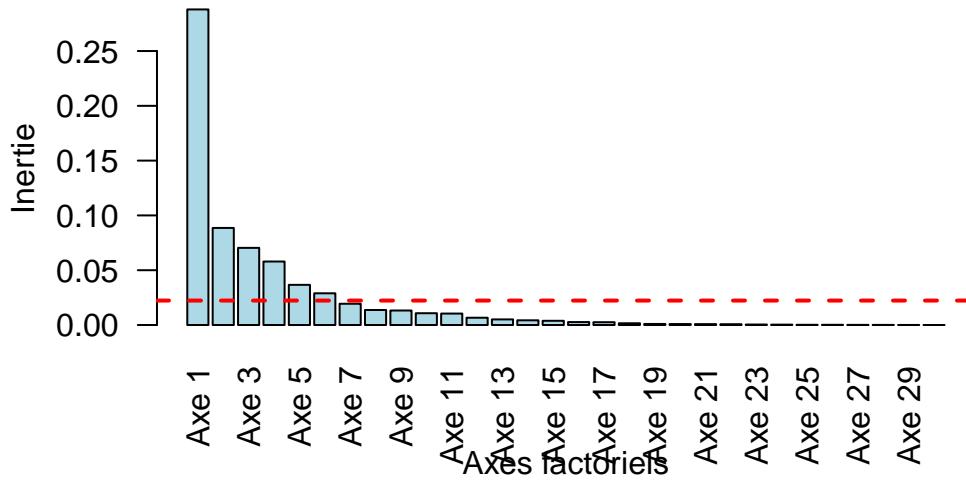




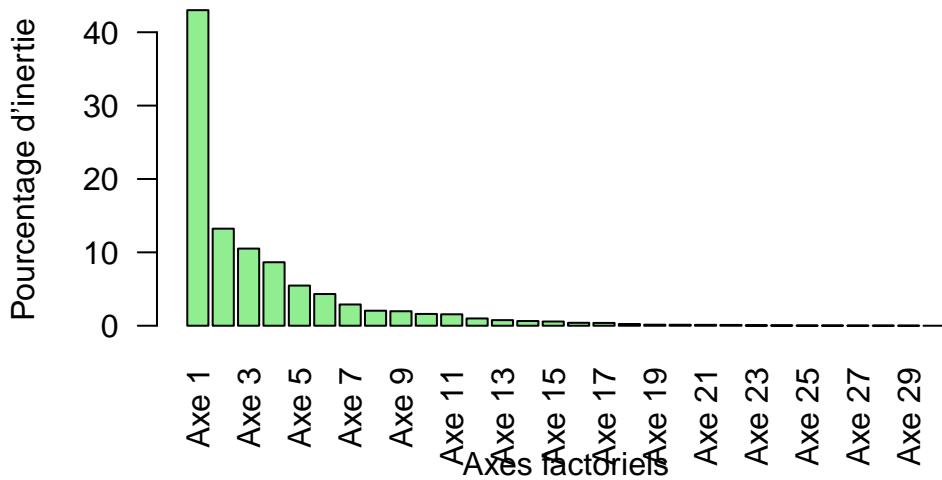
	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	2.879531e-01	43.008992105	43.00899
dim 2	8.856327e-02	13.227908630	56.23690

dim 3	7.040802e-02	10.516220550	66.75312
dim 4	5.794988e-02	8.655459441	75.40858
dim 5	3.663964e-02	5.472536961	80.88112
dim 6	2.892649e-02	4.320493435	85.20161
dim 7	1.940163e-02	2.897849533	88.09946
dim 8	1.371023e-02	2.047774987	90.14724
dim 9	1.324817e-02	1.978761606	92.12600
dim 10	1.074718e-02	1.605210155	93.73121
dim 11	1.043798e-02	1.559028905	95.29024
dim 12	6.646055e-03	0.992662261	96.28290
dim 13	5.118583e-03	0.764517287	97.04742
dim 14	4.287202e-03	0.640341314	97.68776
dim 15	3.855132e-03	0.575806791	98.26356
dim 16	2.672064e-03	0.399102428	98.66267
dim 17	2.554353e-03	0.381520976	99.04419
dim 18	1.589897e-03	0.237468757	99.28166
dim 19	9.481969e-04	0.141623744	99.42328
dim 20	8.720465e-04	0.130249843	99.55353
dim 21	8.078866e-04	0.120666853	99.67420
dim 22	7.325013e-04	0.109407220	99.78360
dim 23	4.320702e-04	0.064534491	99.84814
dim 24	3.419578e-04	0.051075205	99.89921
dim 25	1.996364e-04	0.029817912	99.92903
dim 26	1.932988e-04	0.028871332	99.95790
dim 27	1.205122e-04	0.017999839	99.97590
dim 28	1.086474e-04	0.016227699	99.99213
dim 29	4.418177e-05	0.006599038	99.99873
dim 30	8.507578e-06	0.001270701	100.00000

Décroissance de l'inertie des axes



Décroissance de l'inertie (en %)



Pour explorer les relations entre les pays et les causes de décès, une analyse factorielle des correspondances (AFC) est réalisée. Cette méthode multivariée est particulièrement adaptée aux tableaux de contingence et permet de visualiser les dépendances entre lignes (pays) et colonnes (causes) tout en réduisant la dimensionnalité des données.

L’AFC est appliquée aux 204 pays du tableau, considérés comme lignes actives. Les valeurs propres associées aux axes factoriels sont examinées afin de déterminer la part d’inertie expliquée par chaque axe et d’identifier le nombre d’axes significatifs à retenir. La décroissance des valeurs propres est représentée graphiquement pour faciliter l’interprétation. La somme des valeurs propres correspond à l’inertie totale du tableau, qui peut également être calculée à partir de la statistique du khi-deux normalisée par l’effectif total (Phi2). Dans les deux cas la valeur trouvée est 0.6695183.

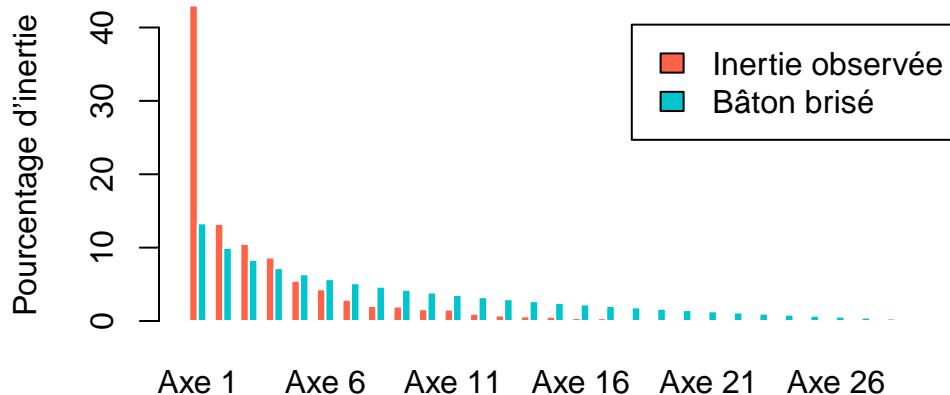
Le critère de Kaiser est utilisé pour identifier les axes factoriels dont l’inertie dépasse la moyenne qui est 0.02231728, indiquant qu’ils expliquent une proportion d’information supérieure à celle attendue par hasard. Dans cette analyse, six axes présentent une inertie supérieure à la moyenne et sont donc retenus pour l’interprétation.

Les diagrammes de l’inertie totale et du pourcentage d’inertie permettent de visualiser la contribution relative de chaque axe à la structure globale des données, préparant ainsi l’interprétation des positions des pays et des causes sur les plans factoriels. Cette étape constitue un prérequis essentiel avant de produire les représentations graphiques des individus et des variables dans l’espace factoriel.

4.3 Critère du bâton brisé pour sélectionner les axes

```
dim 1  dim 2  dim 3  dim 4  dim 5  dim 6  dim 7  dim 8  dim 9 dim 10 dim 11
  TRUE    TRUE    TRUE    TRUE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
dim 12 dim 13 dim 14 dim 15 dim 16 dim 17 dim 18 dim 19 dim 20 dim 21 dim 22
 FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
dim 23 dim 24 dim 25 dim 26 dim 27 dim 28 dim 29 dim 30
 FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
```

Sélection des axes par le critère du bâton brisé



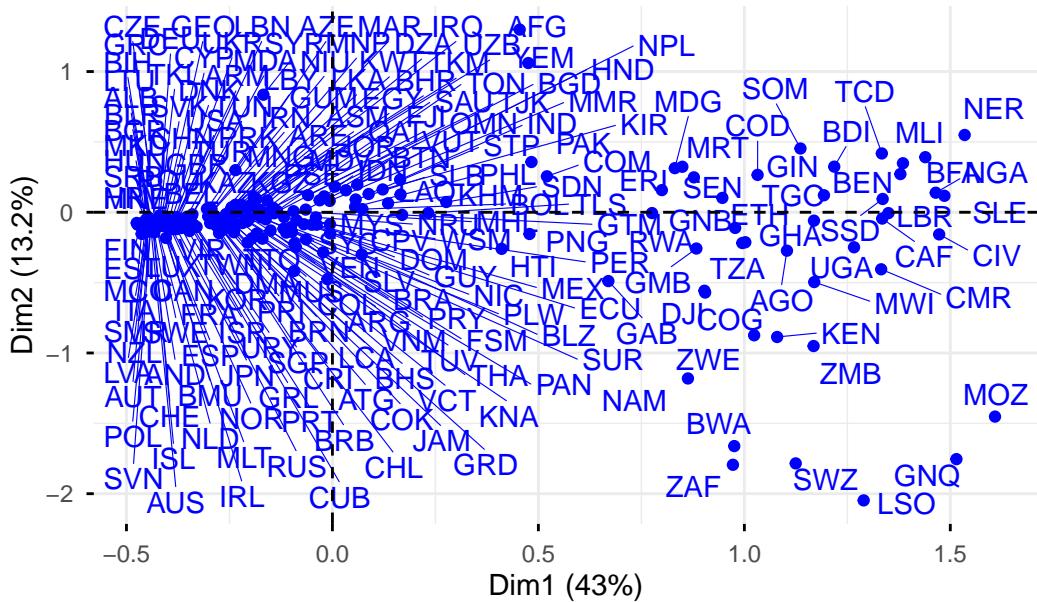
Pour compléter l'identification des axes significatifs, le critère du bâton brisé est utilisé. Les valeurs propres observées sont ainsi mises en relation avec les proportions théoriques du bâton brisé, permettant de distinguer les axes apportant une information réelle de ceux qui ne contiennent qu'un bruit statistique.

La comparaison numérique indique que les axes dont l'inertie observée dépasse celle du bâton brisé sont considérés comme pertinents. Dans cette analyse, quatre axes présentent une inertie supérieure à la valeur attendue par le bâton brisé et sont donc retenus pour l'interprétation factorielle.

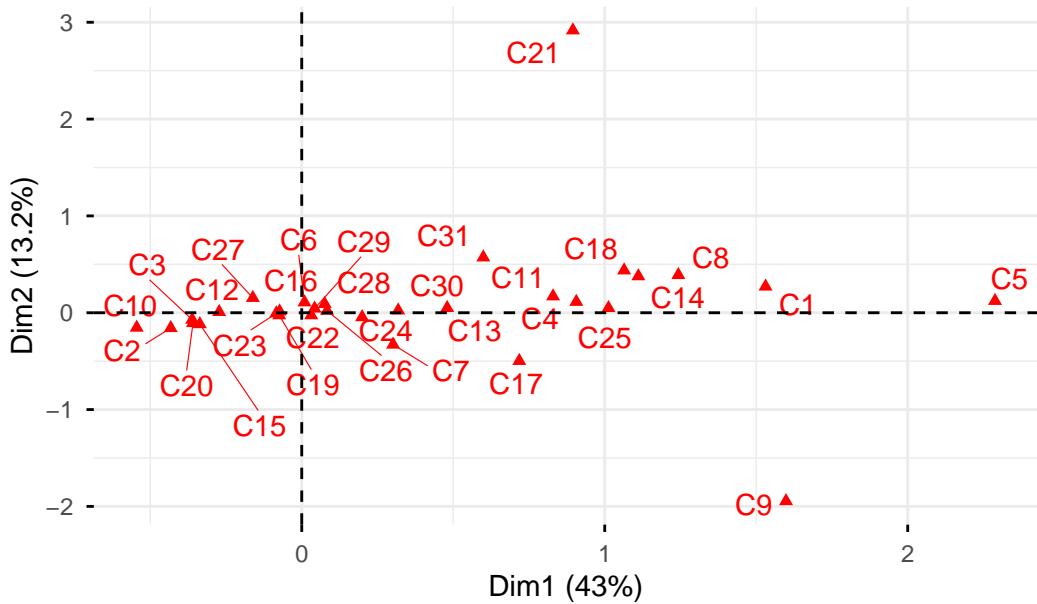
Un diagramme comparatif illustre visuellement la sélection des axes : les barres rouges représentant l'inertie observée sont mises côte à côte avec celles du bâton brisé, représentées en bleu. Cette représentation permet de visualiser aisément quels axes dépassent le seuil et confirment leur pertinence pour l'étude des relations entre les pays et les causes de décès. Ce critère identifiant moins d'axes pertinents que le critère de Kaiser, son résultat sera privilégié pour la suite de l'analyse.

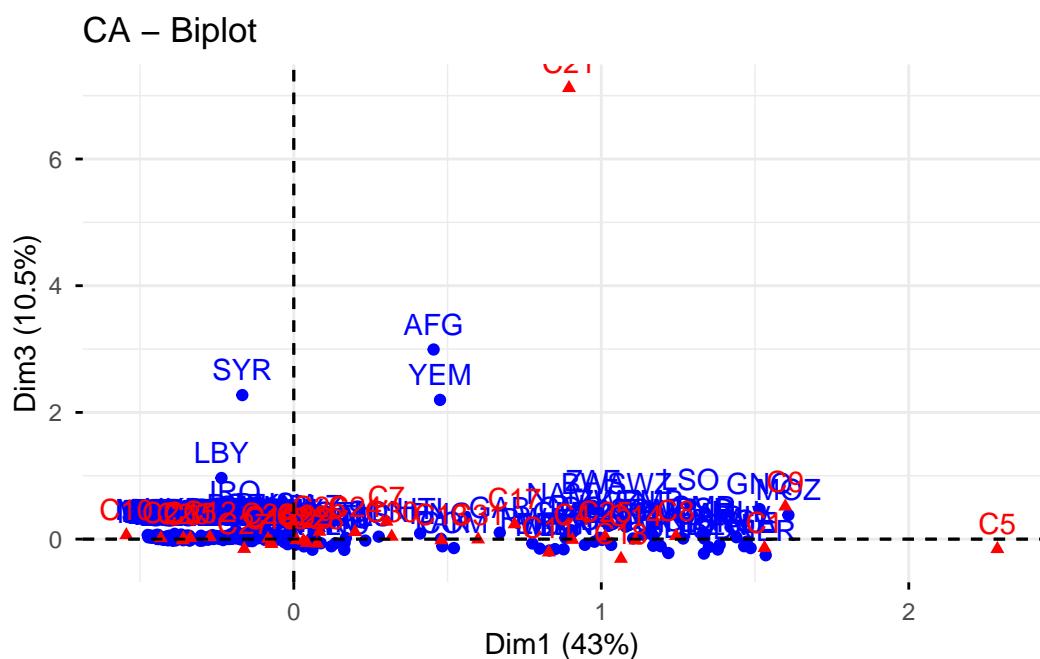
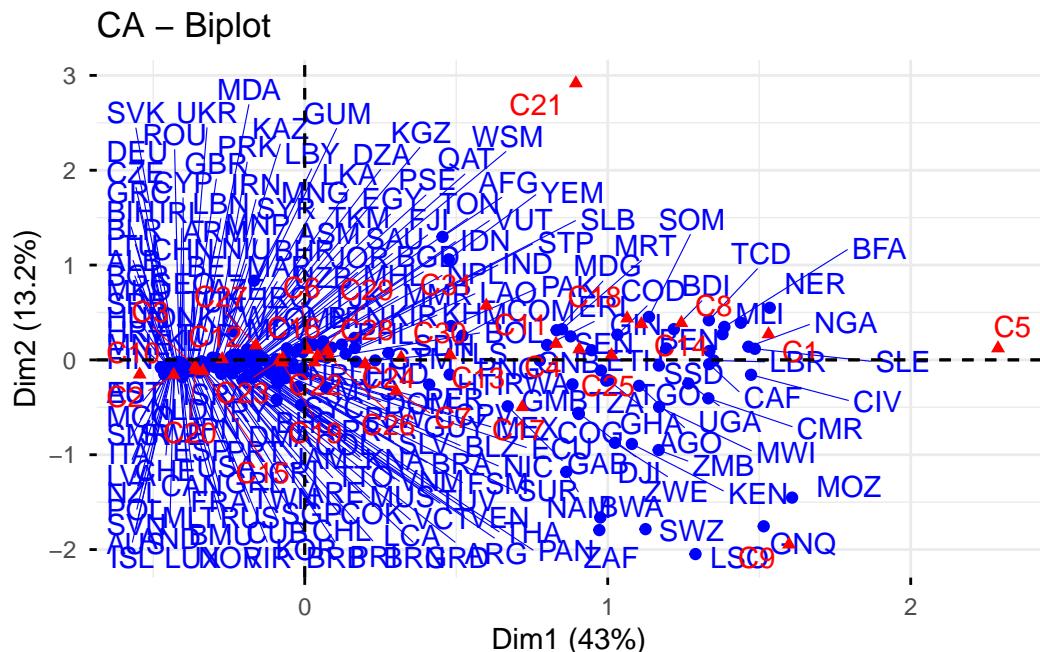
4.4 Analyse des contributions des lignes et colonnes

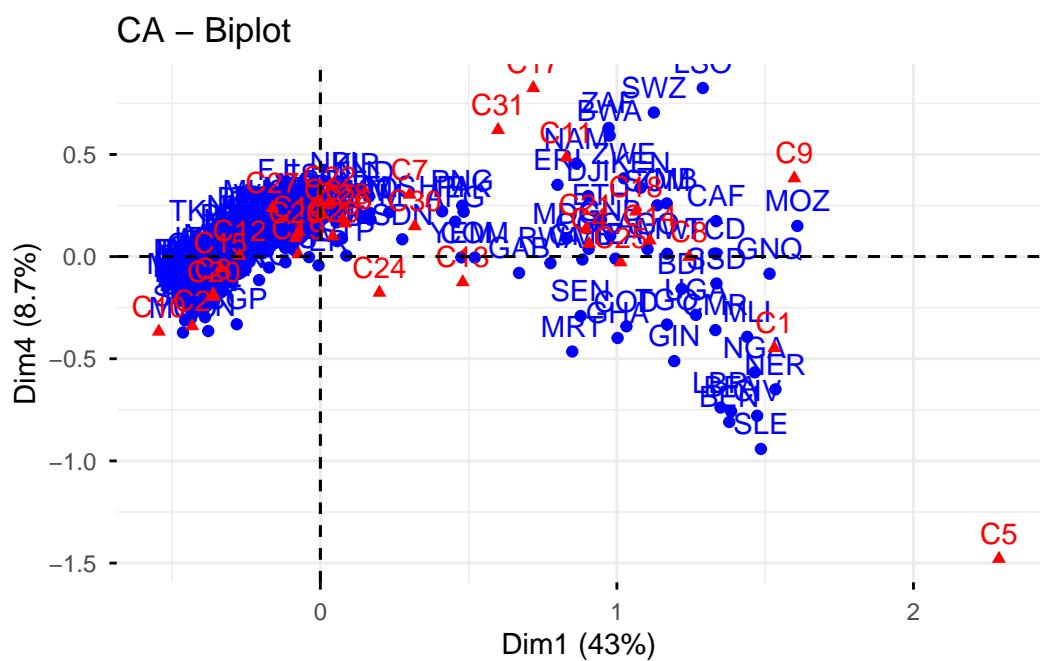
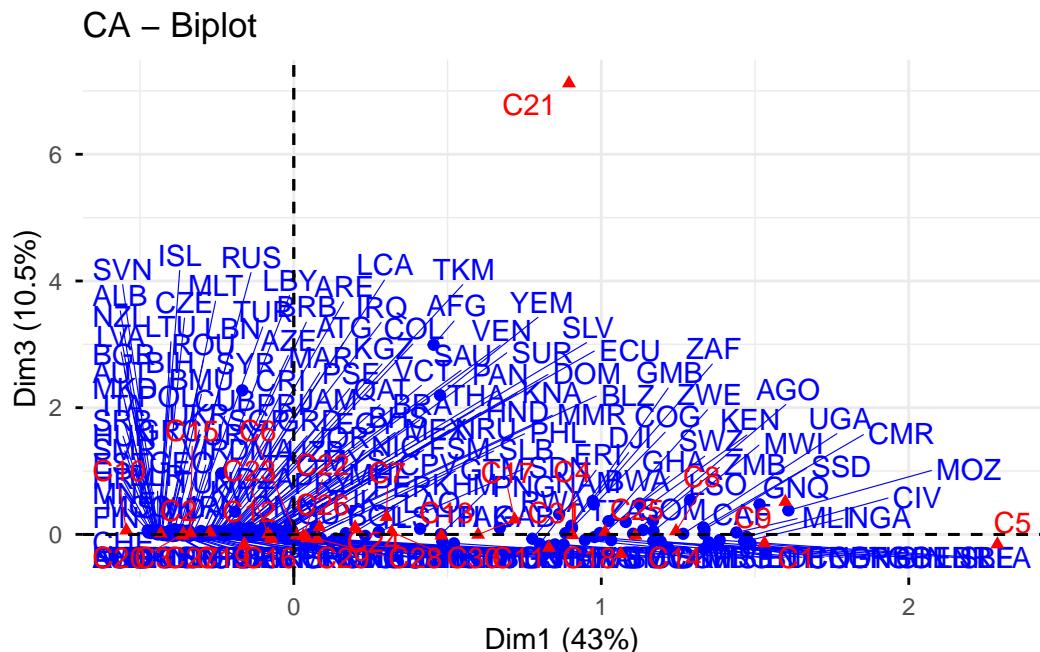
Row points – CA



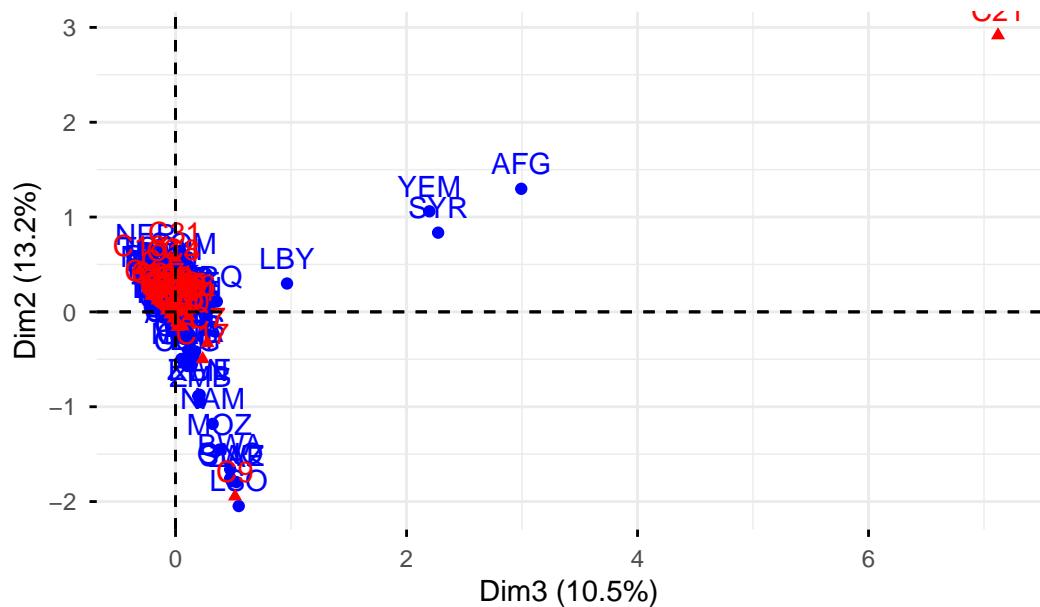
Column points – CA



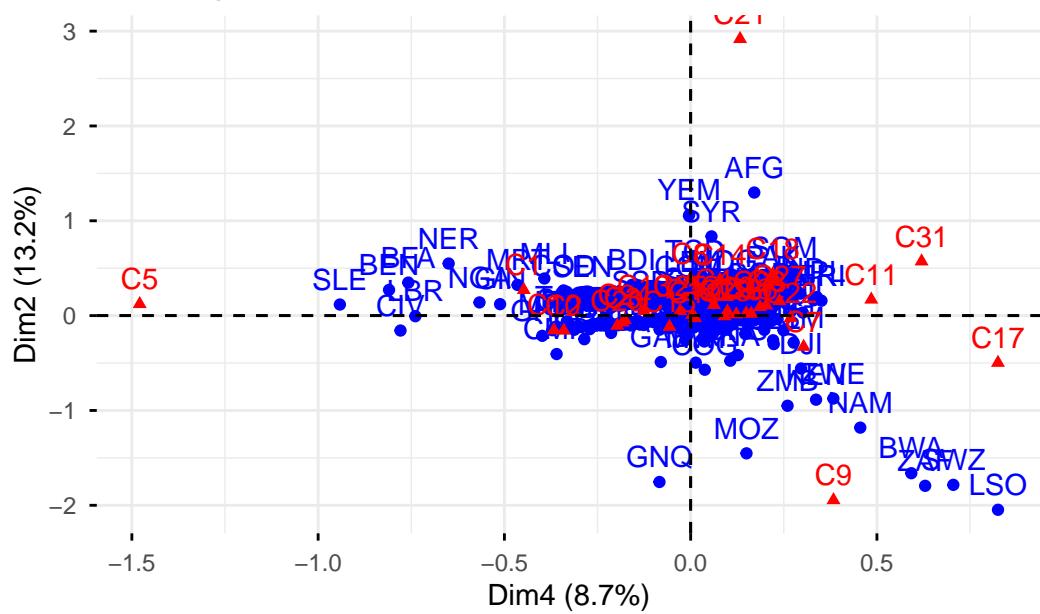




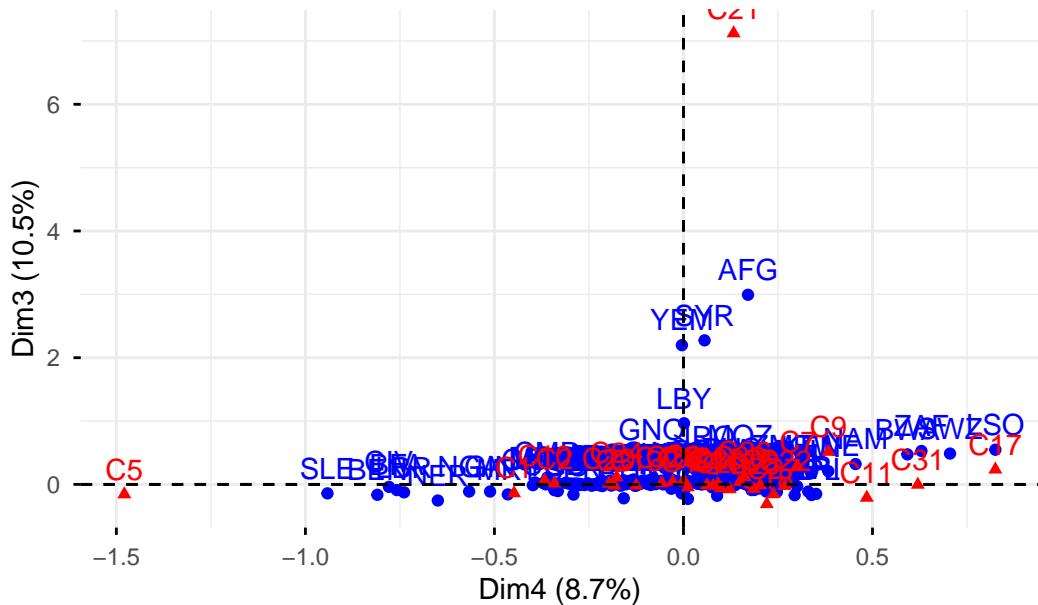
CA – Biplot



CA – Biplot



CA – Biplot



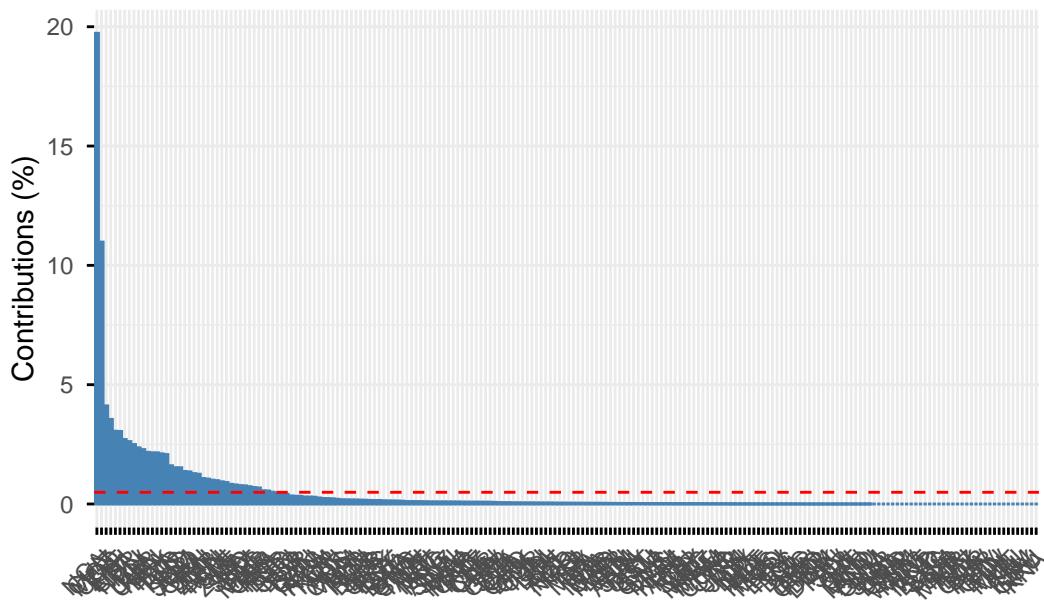
NGA	CHN	MOZ	COD	ZAF	ETH
1.972600e+01	1.097690e+01	4.112398e+00	3.546350e+00	3.051948e+00	3.040174e+00
USA	NER	MLI	CIV	UGA	CMR
2.704453e+00	2.617140e+00	2.494568e+00	2.356005e+00	2.281417e+00	2.168424e+00
KEN	BFA	PAK	TZA	TCD	RUS
2.145580e+00	2.144070e+00	2.105711e+00	2.074027e+00	1.605238e+00	1.522055e+00
IND	SOM	AGO	GHA	JPN	DEU
1.521142e+00	1.363663e+00	1.348174e+00	1.276305e+00	1.250440e+00	1.071690e+00
ZMB	BEN	MWI	GIN	SLE	ZWE
1.045949e+00	1.002340e+00	9.801908e-01	9.321365e-01	9.024169e-01	8.271409e-01
ITA	SSD	BDI	CAF	MDG	UKR
8.022629e-01	7.737101e-01	7.623462e-01	7.281017e-01	6.927550e-01	6.746617e-01
GBR	FRA	POL	ESP	TGO	SEN
5.600329e-01	5.487953e-01	4.976482e-01	4.402356e-01	4.386891e-01	4.110469e-01
LSO	TUR	LBR	CAN	ROU	AFG
3.398167e-01	3.264306e-01	3.244937e-01	2.944403e-01	2.937560e-01	2.874319e-01
RWA	YEM	KOR	IRN	AUS	COG
2.558301e-01	2.333072e-01	2.290476e-01	2.185308e-01	1.998801e-01	1.806127e-01
VNM	ERI	NLD	HUN	BGR	SRB
1.769272e-01	1.739556e-01	1.706016e-01	1.669552e-01	1.593717e-01	1.523494e-01
BLR	GRC	PRK	EGY	CZE	BWA
1.500816e-01	1.498720e-01	1.386534e-01	1.318928e-01	1.302167e-01	1.262932e-01
BRA	GNQ	SWE	TWN	AUT	HTI

1.205327e-01	1.052378e-01	1.049020e-01	1.026277e-01	1.024692e-01	9.880697e-02
MRT	PNG	ARG	SWZ	SDN	BEL
9.306157e-02	9.252984e-02	9.198700e-02	9.118239e-02	8.937911e-02	8.931218e-02
NAM	MAR	GNB	PRT	KAZ	CHE
8.818762e-02	8.496812e-02	8.120315e-02	8.049203e-02	7.980266e-02	7.953637e-02
CUB	FIN	HRV	GMB	SVK	DZA
7.788342e-02	7.745852e-02	7.002104e-02	6.269620e-02	6.044403e-02	5.944796e-02
CHL	GEO	DNK	COL	UZB	TUN
5.528741e-02	5.299551e-02	5.258174e-02	5.105293e-02	4.992190e-02	4.854097e-02
LTU	LKA	BIH	AZE	NZL	NOR
4.727728e-02	4.659072e-02	4.602689e-02	4.174454e-02	4.124504e-02	3.918434e-02
IRQ	DJI	ISR	MEX	LVA	IRL
3.836830e-02	3.763060e-02	3.679920e-02	3.548623e-02	3.418809e-02	3.395906e-02
MDA	GAB	MKD	LBN	ALB	VEN
3.391938e-02	3.279853e-02	3.130896e-02	2.943907e-02	2.800547e-02	2.610080e-02
ARM	SVN	THA	URY	EST	SAU
2.553221e-02	2.512450e-02	2.433797e-02	2.229884e-02	2.168277e-02	1.923721e-02
GTM	SYR	MYS	PRI	KHM	ARE
1.721468e-02	1.486119e-02	1.478922e-02	1.427864e-02	1.304220e-02	1.212781e-02
PHL	CRI	SGP	LBY	MNG	KGZ
1.176433e-02	1.159761e-02	1.137921e-02	1.080026e-02	1.077230e-02	9.770222e-03
MNE	TKM	COM	CYP	LAO	JOR
9.692841e-03	9.054842e-03	8.433604e-03	8.285450e-03	7.424738e-03	6.006270e-03
JAM	IDN	LUX	NPL	MLT	KWT
5.041502e-03	4.391072e-03	4.368215e-03	4.356792e-03	3.865886e-03	3.864648e-03
PSE	MUS	TTO	ECU	PRY	TLS
3.596396e-03	3.232662e-03	2.759043e-03	2.752724e-03	2.624600e-03	2.583735e-03
BOL	ISL	NIC	BHR	OMN	HND
2.481448e-03	2.445309e-03	1.898439e-03	1.540244e-03	1.506583e-03	1.382452e-03
BRB	DOM	TJK	PAN	SLV	VIR
1.379973e-03	1.114681e-03	1.051818e-03	1.048171e-03	1.044695e-03	9.565579e-04
FJI	QAT	BRN	MCO	AND	BMU
9.514172e-04	8.134107e-04	7.705617e-04	7.029966e-04	7.009521e-04	5.812986e-04
GUM	LCA	SMR	GRL	DMA	GRD
5.142038e-04	4.890383e-04	3.944932e-04	3.364986e-04	2.964065e-04	2.806769e-04
SUR	WSM	MDV	ATG	VCT	MNP
2.567409e-04	2.515622e-04	2.201939e-04	2.108838e-04	1.809470e-04	1.689041e-04
BGD	SYC	BHS	ASM	SLB	MMR
1.684047e-04	1.499977e-04	1.402641e-04	1.215069e-04	1.189319e-04	1.129359e-04
KIR	TON	BLZ	COK	STP	PLW
1.070237e-04	7.399901e-05	6.159382e-05	6.016454e-05	4.760063e-05	3.378867e-05
PER	CPV	BTN	TUV	MHL	VUT
3.361327e-05	3.206319e-05	2.746204e-05	2.657485e-05	2.318264e-05	1.325692e-05

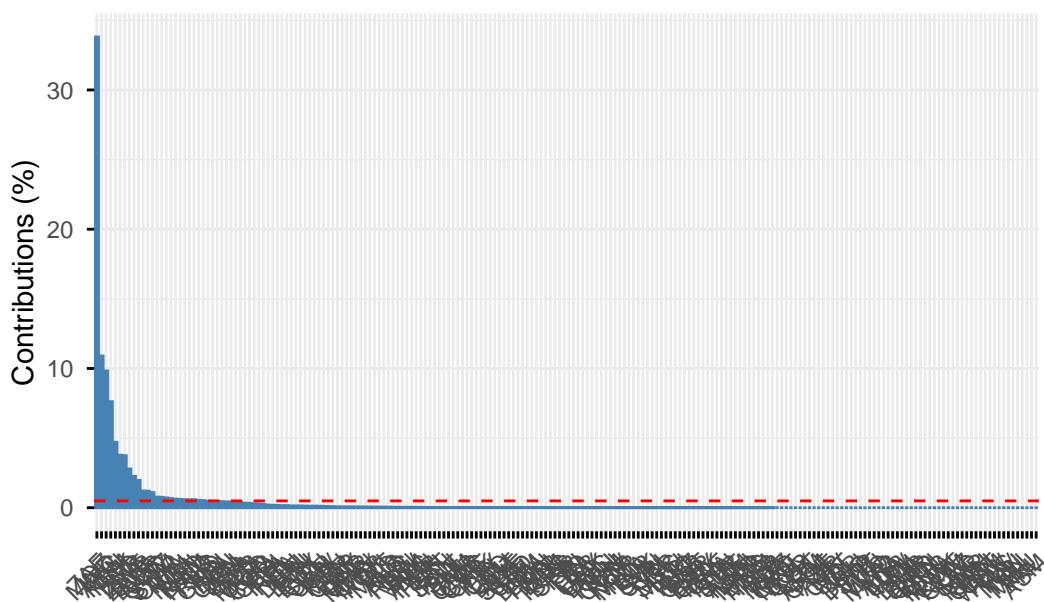
NIU	TKL	FSM	GUY	NRU	KNA
1.026228e-05	1.007530e-05	2.968293e-06	1.328026e-06	6.461912e-07	4.598659e-07
ZAF	MOZ	IND	AFG	KEN	YEM
3.380379e+01	1.089643e+01	9.817344e+00	7.616354e+00	4.701756e+00	3.772577e+00
PAK	LSO	ZMB	ZWE	SYR	BWA
3.740590e+00	2.784479e+00	2.252237e+00	1.960896e+00	1.201585e+00	1.191412e+00
NER	COD	SWZ	SOM	CMR	THA
1.089745e+00	7.631764e-01	7.463981e-01	7.050948e-01	6.524266e-01	6.178494e-01
MLI	NGA	MWI	BGD	NAM	TCD
5.995435e-01	5.780372e-01	5.730353e-01	5.722458e-01	5.368862e-01	5.125519e-01
RUS	GNQ	IDN	BFA	CHN	JPN
4.645845e-01	4.589166e-01	4.541263e-01	4.404130e-01	4.175677e-01	4.070878e-01
USA	BRA	TZA	MDG	UGA	AGO
4.041022e-01	3.791721e-01	3.254830e-01	3.198426e-01	2.861206e-01	2.684575e-01
COG	GHA	BDI	FRA	NPL	ITA
2.334712e-01	1.877122e-01	1.753963e-01	1.712221e-01	1.488765e-01	1.487567e-01
DEU	HTI	BEN	ETH	EGY	COL
1.289423e-01	1.286297e-01	1.263394e-01	1.146381e-01	1.131120e-01	1.129409e-01
SEN	UKR	CIV	POL	VEN	GBR
1.068105e-01	1.008010e-01	8.718282e-02	7.470358e-02	6.328228e-02	6.092560e-02
ESP	LBY	GAB	MMR	CAN	MEX
5.870263e-02	5.684699e-02	5.680507e-02	5.502210e-02	5.410704e-02	5.032048e-02
KOR	DJI	MRT	IRQ	AUS	NLD
4.904470e-02	4.663925e-02	4.385548e-02	4.112529e-02	3.634264e-02	3.448693e-02
PNG	GIN	SLV	VNM	ARG	TWN
3.132475e-02	3.115526e-02	3.076661e-02	2.995529e-02	2.901653e-02	2.281793e-02
PER	ERI	PRT	SDN	PAN	SLE
2.253401e-02	2.187421e-02	2.084460e-02	2.021068e-02	1.972072e-02	1.816747e-02
PHL	JAM	GMB	CUB	GRC	CHE
1.808511e-02	1.787402e-02	1.732716e-02	1.631812e-02	1.558014e-02	1.450955e-02
TJK	PRY	CHL	LAO	ECU	AUT
1.447046e-02	1.411759e-02	1.403946e-02	1.400368e-02	1.358298e-02	1.351448e-02
SWE	CZE	SSD	BEL	DOM	DZA
1.315300e-02	1.291130e-02	1.257719e-02	1.142510e-02	1.132050e-02	1.067409e-02
HUN	PRI	BLR	KHM	BHS	SRB
1.061724e-02	1.029819e-02	9.752588e-03	9.500895e-03	9.236902e-03	9.175424e-03
ISR	FIN	TTO	LKA	UZB	URY
8.713699e-03	8.652433e-03	8.118078e-03	7.684947e-03	7.227331e-03	7.206456e-03
DNK	NIC	ROU	IRN	COM	HRV
7.008283e-03	6.834110e-03	6.752645e-03	6.730465e-03	6.604777e-03	6.337705e-03
NOR	MAR	NZL	LVA	TUR	SVK

5.805441e-03	5.514215e-03	5.398731e-03	5.019548e-03	4.764243e-03	4.711707e-03
SVN	CRI	IRL	SGP	MYS	LTU
4.568470e-03	4.367124e-03	4.360892e-03	4.321286e-03	4.030292e-03	3.760756e-03
BIH	TGO	BLZ	GNB	EST	BGR
3.713909e-03	3.705276e-03	3.561045e-03	3.426878e-03	3.368532e-03	3.342762e-03
GUY	PRK	CAF	KNA	BTN	SUR
3.054865e-03	2.857844e-03	2.681053e-03	2.268013e-03	2.033365e-03	1.675969e-03
ARE	FSM	LBN	MKD	BRB	MUS
1.571773e-03	1.547415e-03	1.450304e-03	1.334426e-03	1.295328e-03	1.293708e-03
TKM	ALB	SLB	MDA	KGZ	GEO
1.179008e-03	1.171106e-03	9.984614e-04	9.517872e-04	8.597197e-04	8.176726e-04
MNE	BOL	CYP	LUX	VCT	ARM
8.055884e-04	7.570388e-04	7.456412e-04	7.292559e-04	7.183579e-04	6.911436e-04
GTM	PSE	AZE	KIR	CPV	MLT
6.845214e-04	6.657086e-04	6.450861e-04	6.128108e-04	5.581204e-04	5.498042e-04
ISL	VIR	BMU	HND	STP	VUT
4.905277e-04	4.605647e-04	3.857341e-04	3.804279e-04	3.376216e-04	3.217539e-04
TUN	MCO	AND	ATG	LCA	JOR
2.890280e-04	2.521852e-04	2.494163e-04	2.428497e-04	2.176977e-04	2.116852e-04
MDV	FJI	DMA	BHR	SMR	GRD
1.961164e-04	1.560342e-04	1.497239e-04	1.489058e-04	1.370959e-04	1.249801e-04
BRN	GRL	MNG	RWA	COK	KWT
1.180379e-04	9.220963e-05	8.537299e-05	4.697491e-05	4.374485e-05	4.109254e-05
GUM	KAZ	SYC	LBR	MNP	SAU
3.800437e-05	2.991143e-05	2.144863e-05	2.122536e-05	1.570164e-05	1.394703e-05
OMN	QAT	WSM	PLW	TLS	TUV
1.234579e-05	1.201471e-05	1.129047e-05	1.051295e-05	4.680961e-06	3.554718e-06
NRU	TON	MHL	NIU	TKL	ASM
1.723373e-06	1.210184e-06	4.838918e-07	1.045754e-07	8.753926e-08	4.719367e-10

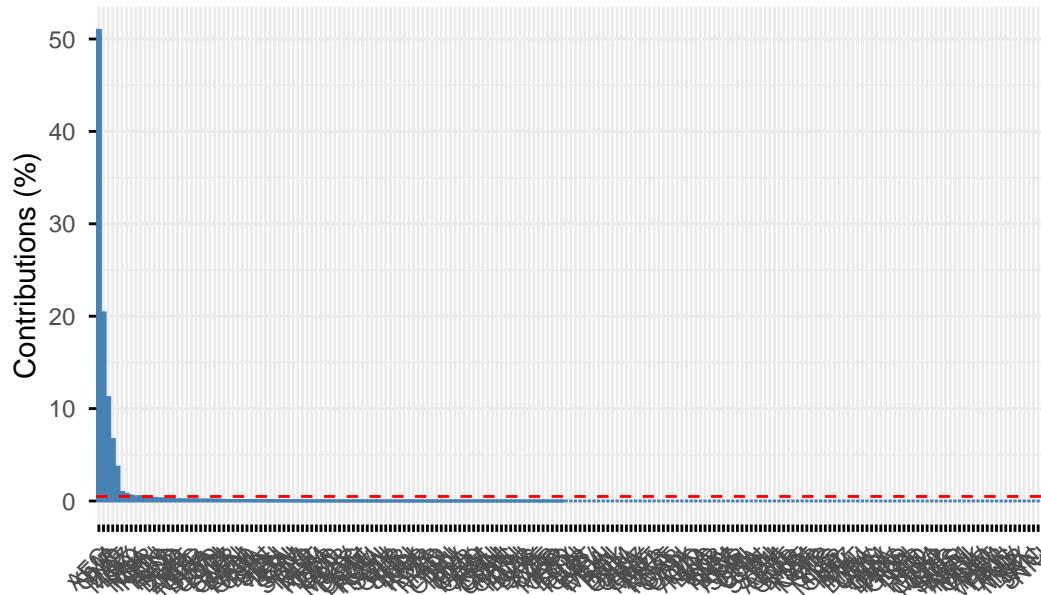
Contribution of rows to Dim-1



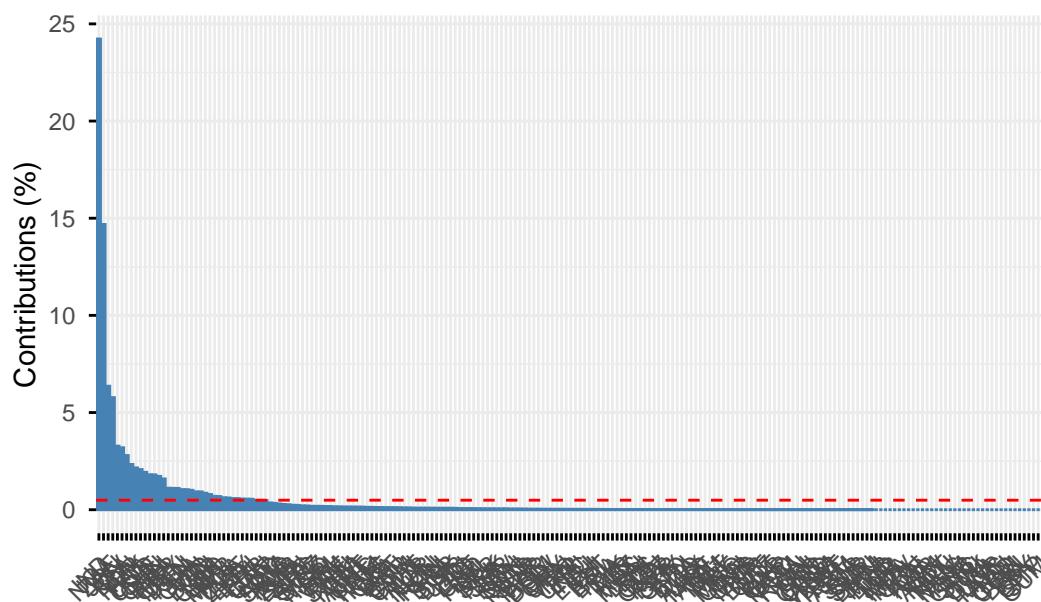
Contribution of rows to Dim-2



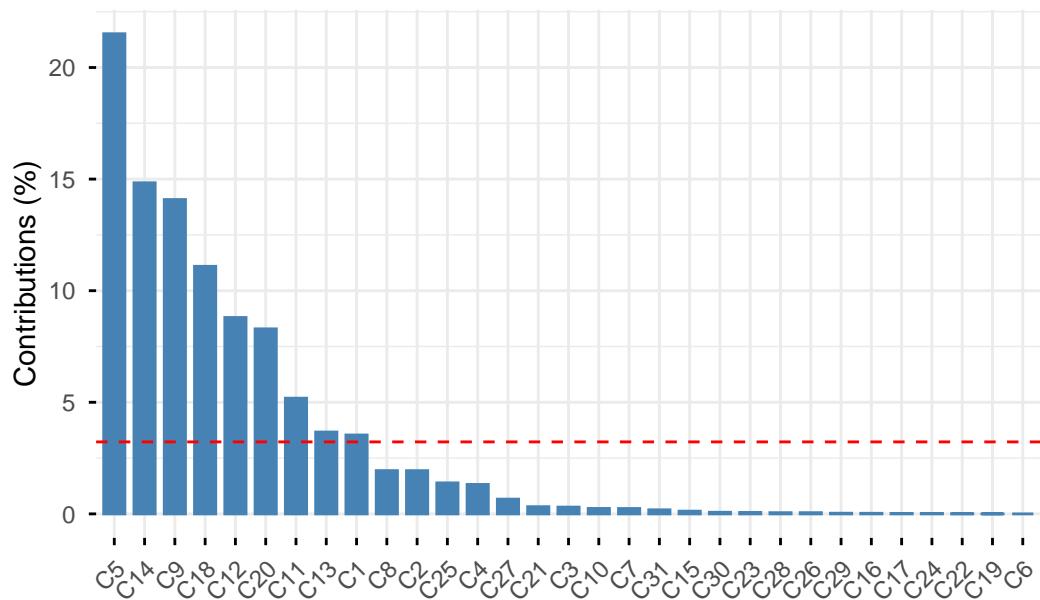
Contribution of rows to Dim-3



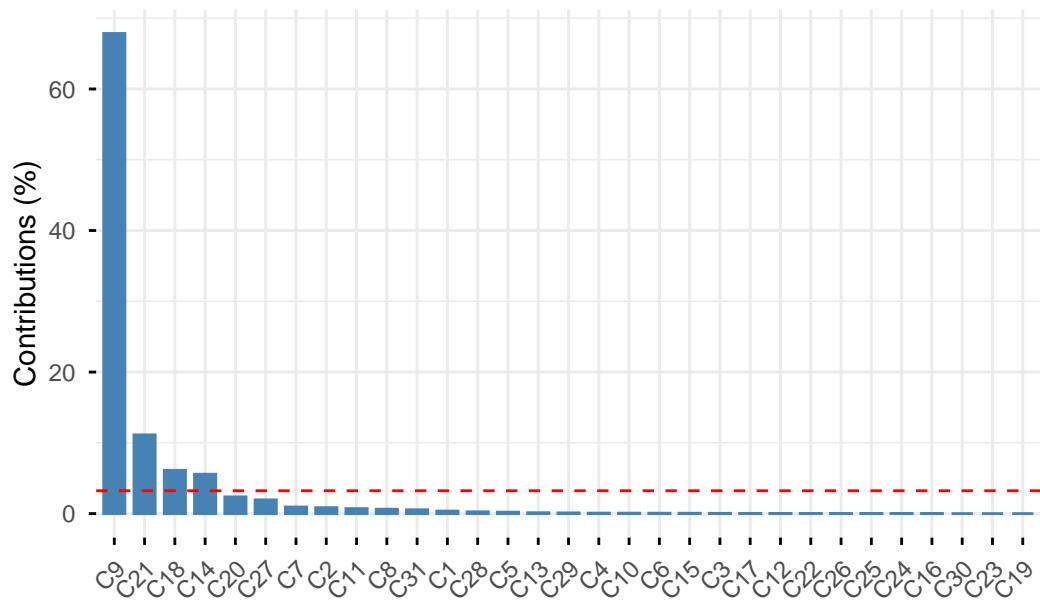
Contribution of rows to Dim-4



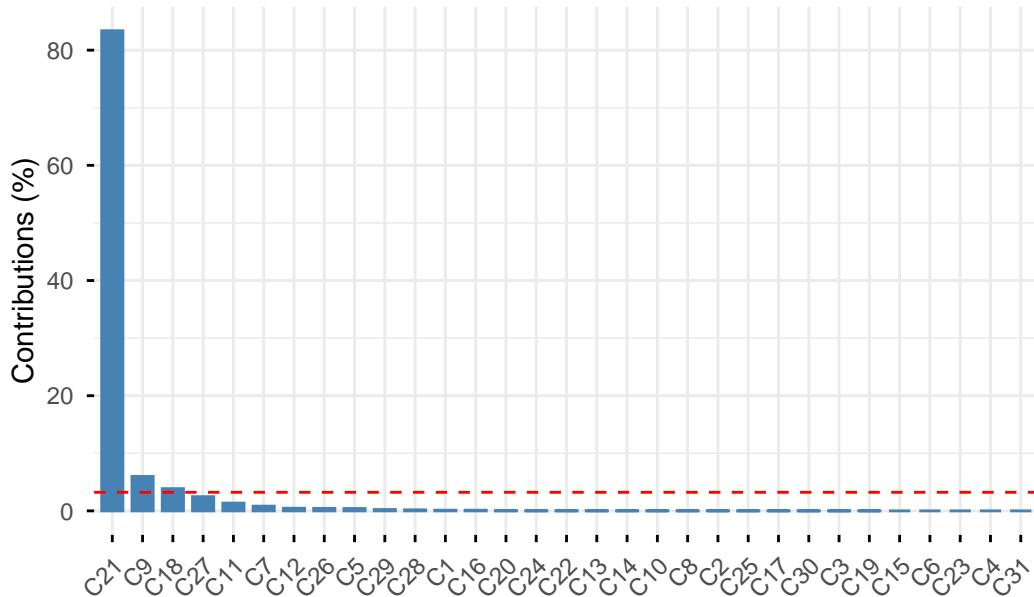
Contribution of columns to Dim-1



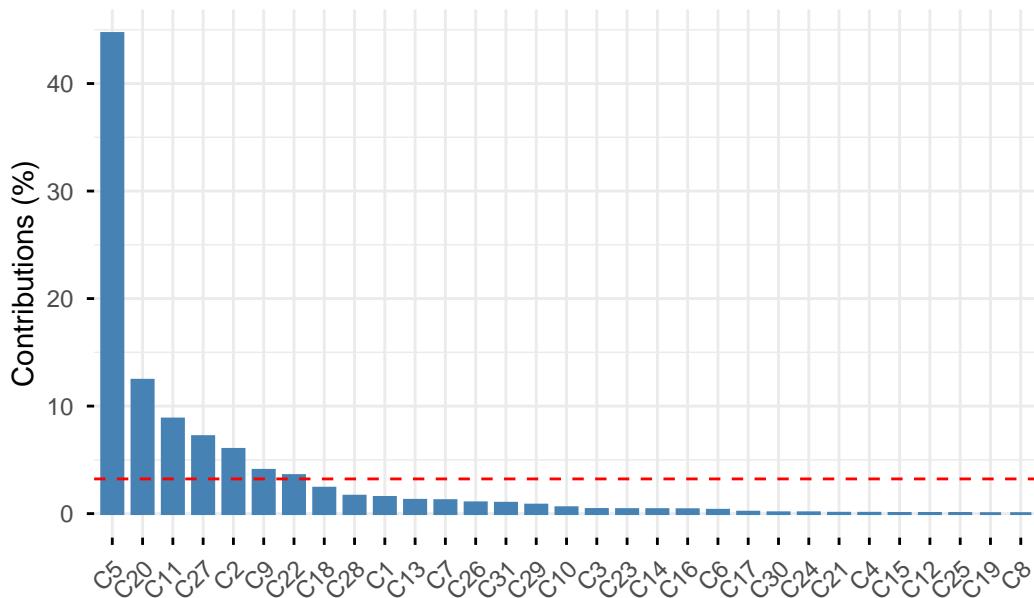
Contribution of columns to Dim-2



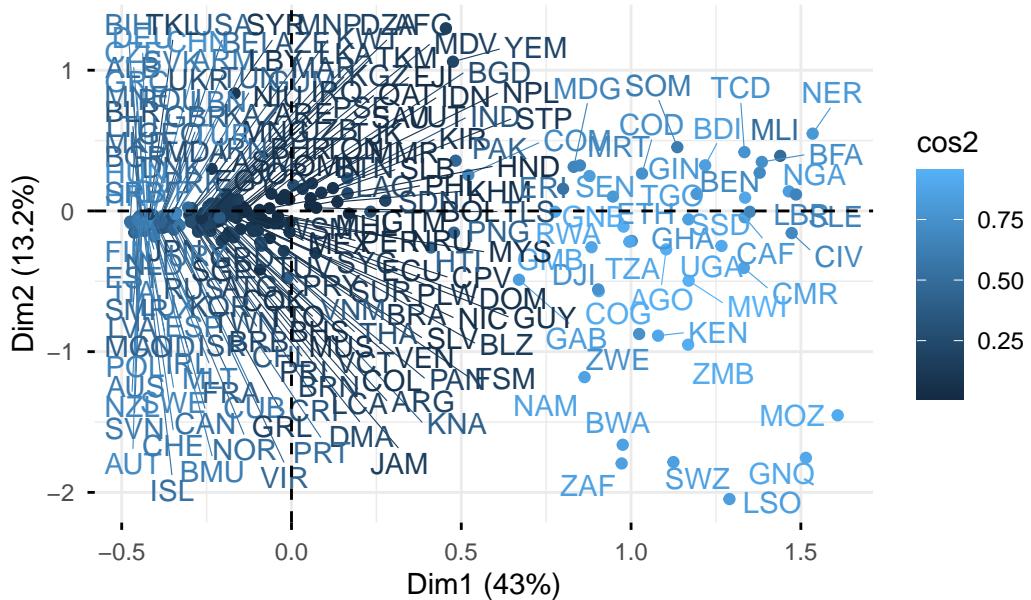
Contribution of columns to Dim–3



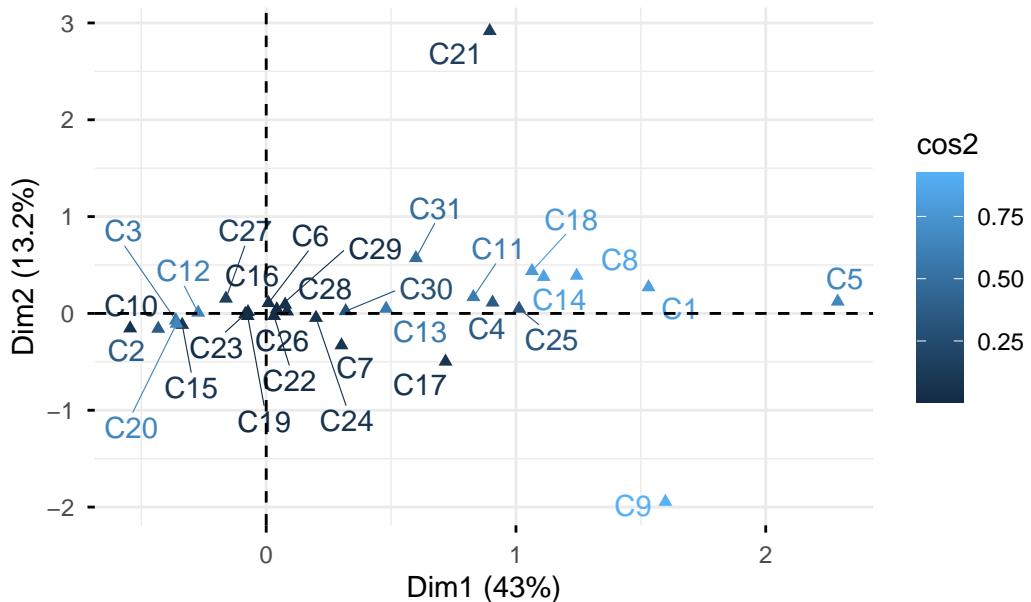
Contribution of columns to Dim–4



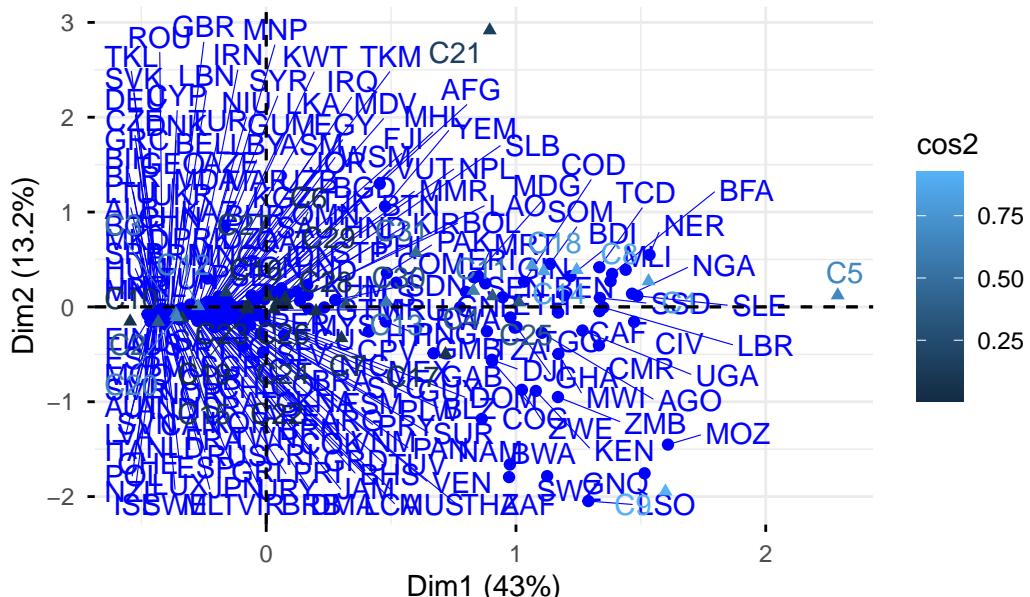
Row points – CA



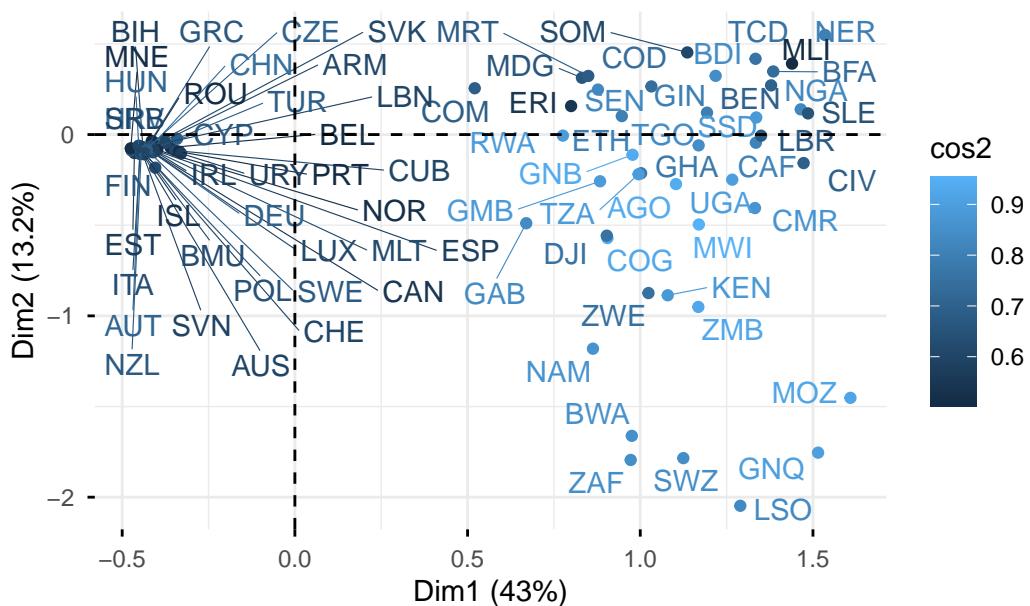
Column points – CA



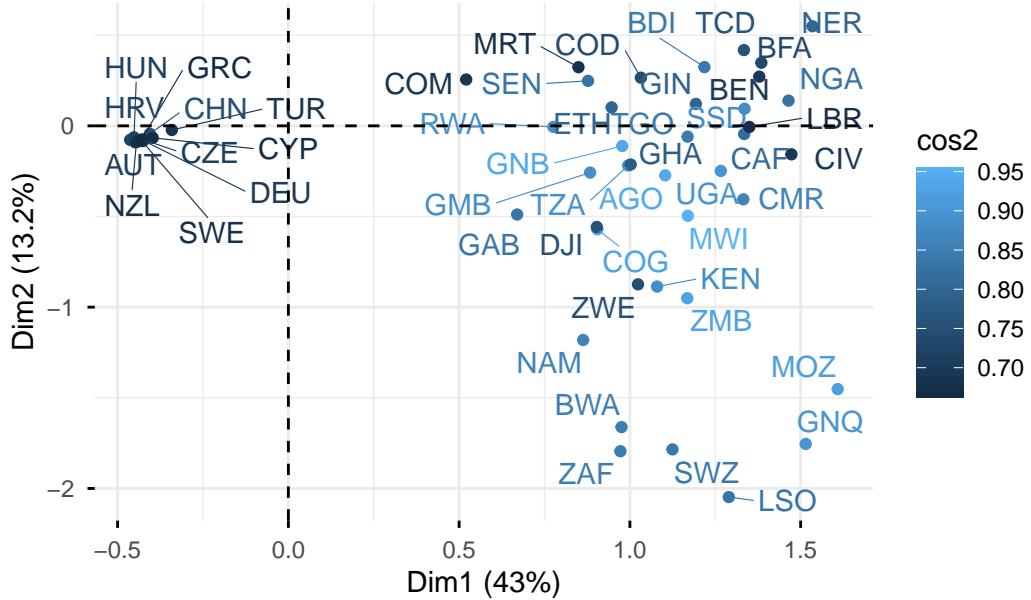
CA – Biplot



Row points – CA



Row points – CA



Après sélection des axes significatifs, l'AFC permet de représenter les modalités des lignes (pays) et des colonnes (causes) dans l'espace factoriel. Les diagrammes factoriels des lignes et des colonnes mettent en évidence les similitudes et les différences structurelles : les pays proches sur le plan factoriel présentent des profils de mortalité similaires, tandis que les causes de décès rapprochées indiquent une co-occurrence relative dans différents pays.

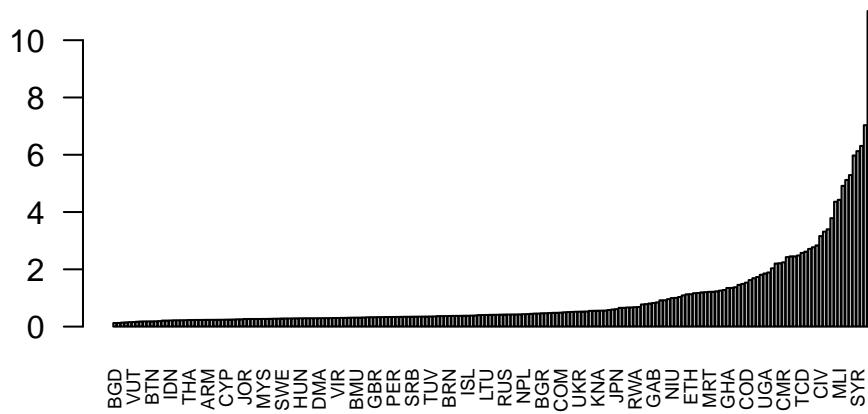
Les biplots combinant lignes et colonnes sur les premiers plans factoriels fournissent une vue d'ensemble des relations entre pays et causes. Le premier plan factoriel (axes 1 et 2) est généralement le plus informatif, mais d'autres combinaisons d'axes peuvent également révéler des dimensions secondaires de variation.

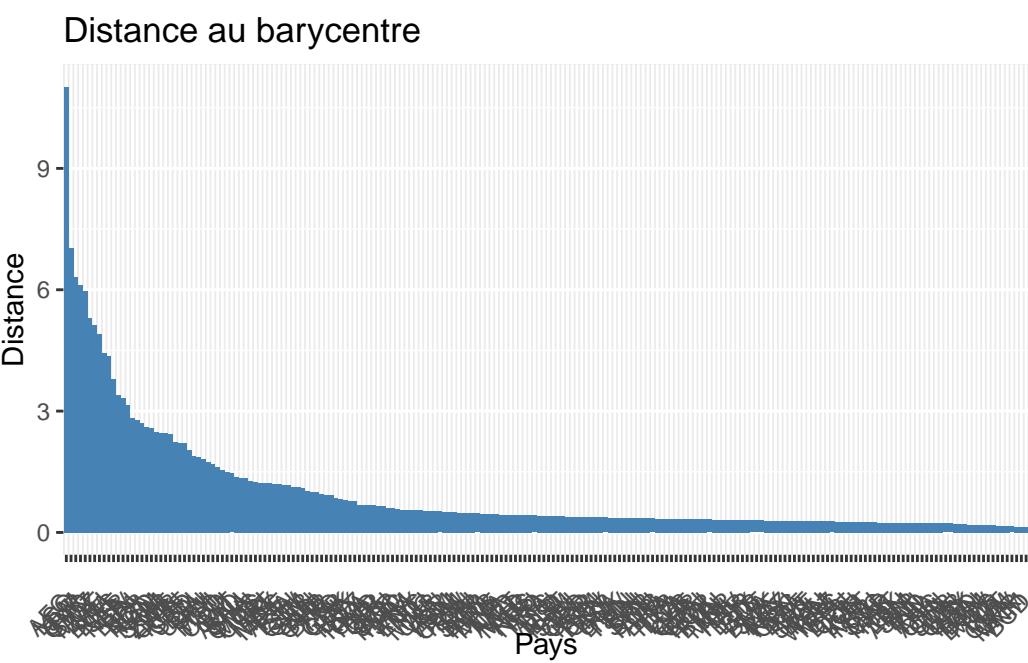
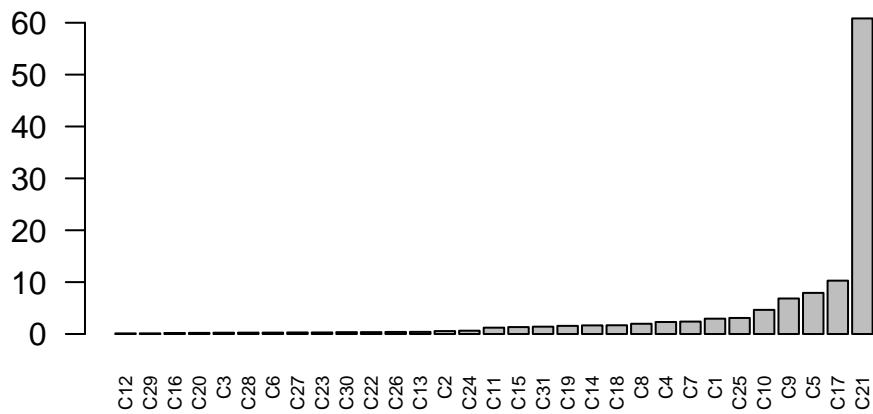
Pour approfondir l'interprétation, les contributions des lignes et des colonnes aux axes factoriels sont examinées. Les modalités les plus contributives sont identifiées, ce qui permet de se concentrer sur les pays et causes qui expliquent le mieux la variance sur chaque axe. Étant donné le grand nombre de pays (204), une sélection des dix premières contributions pour chaque axe facilite la lecture et l'interprétation des résultats.

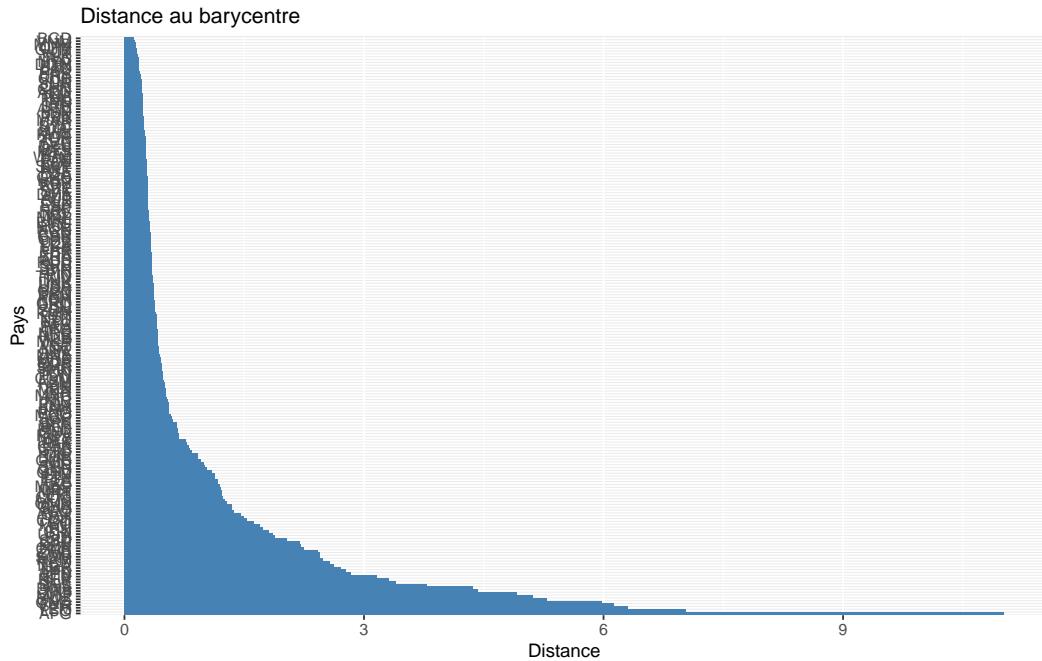
La qualité de représentation de chaque modalité est évaluée à l'aide du cos2, indiquant la proportion de la variance de la modalité expliquée par les axes retenus. Les modalités avec un cos2 élevé sont mieux représentées sur le plan factoriel et constituent des points de référence fiables pour l'interprétation. Des diagrammes filtrés sur le cos2 permettent de visualiser clairement les pays et causes les mieux représentés, tout en réduisant l'encombrement visuel des graphiques.

Cette analyse graphique et quantitative des contributions et de la qualité de représentation fournit une base solide pour tirer des conclusions sur les profils de mortalité et les relations entre pays et causes dans l'ensemble du jeu de données.

4.5 Distances au centre de gravité







L'AFC permet également de calculer la distance de chaque pays et de chaque cause au centre de gravité (barycentre) de l'ensemble des données. Cette distance est obtenue en rapportant l'inertie d'une modalité à sa masse, ce qui fournit une mesure de l'écart par rapport au profil moyen. Les pays dont la distance est faible possèdent un profil proche de la moyenne, tandis que ceux ayant une distance élevée présentent des profils atypiques ou fortement différenciés.

La visualisation des distances au barycentre, par diagrammes en barres verticales ou horizontales, met en évidence les pays dont les profils sont les plus représentatifs et ceux qui s'écartent le plus du centre de gravité. Cette information est utile pour identifier les pays ou causes présentant des caractéristiques particulières et guide l'interprétation des plans factoriels précédemment étudiés.

5 Conclusion

Fatoumata

Résumer les principaux résultats :Combien d'axes retenus, quelles causes/pays se distinguent
Interprétation des distances et profils. Limites de l'analyse. Perspectives éventuelles

Yester

L'analyse des données de mortalité par pays et par cause a permis d'illustrer la richesse et la diversité des profils de décès dans le monde. L'exploration initiale par diagrammes en barres a montré des différences marquées entre pays, notamment entre pays à revenu élevé et pays à

revenu faible. Les profils lignes et colonnes ont permis de comparer les structures de mortalité de manière normalisée et d'identifier les causes et pays les plus contributifs à chaque axe factoriel.

Le test du khi-deux a confirmé l'existence d'une dépendance statistique entre pays et causes de décès, bien que la force de cette liaison soit modérée. L'AFC a permis de synthétiser cette information dans un espace de faible dimension, en sélectionnant les axes les plus pertinents selon les critères de Kaiser et du bâton brisé. L'examen des plans factoriels, des contributions et des distances au centre de gravité a fourni une compréhension détaillée des profils atypiques et des similarités entre pays et causes.

En conclusion, cette approche méthodologique a permis d'identifier les tendances globales et les particularités locales de la mortalité, fournissant une base solide pour des analyses comparatives et pour des prises de décision en santé publique. Les résultats mettent en évidence à la fois les déterminants communs et les spécificités nationales des causes de décès, illustrant l'intérêt des méthodes multivariées pour l'analyse de données de grande dimension.

6 Références

6.1 Articles, manuels R, sources des données