Analyse factorielle des correspondances des causes de décès par pays

Mkrtchyan, Toure

Table of contents

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 $\times$ 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.

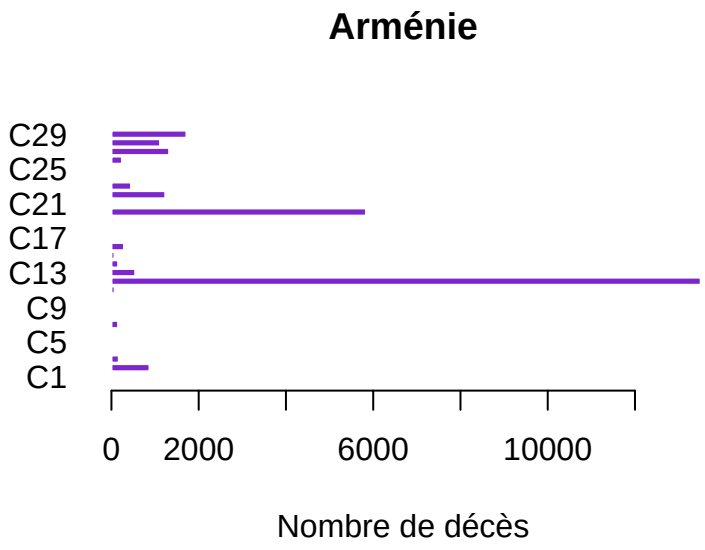
Au moment du chargement de données, la première colonne du jeu de données est désigné comme étant la colonne d'identification des lignes et elle est ensuite supprimée, afin de ne conserver que les causes de décès sous forme numérique. Cette étape permet de faciliter les traitements ultérieurs et l'accès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

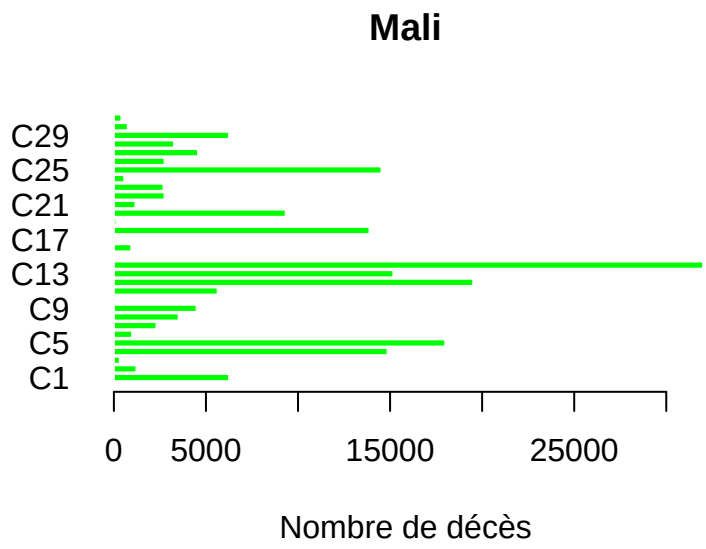
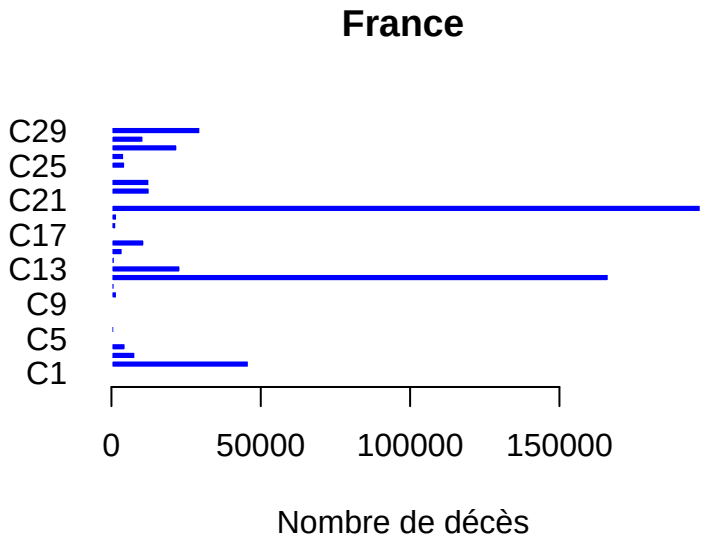


Figure 1: 2) Répartition des causes de décès au Mali en 2019

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 entraîné 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+…+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.

Table 2: Profils lignes pour quelques pays et causes sélectionnés

	C1	C2	C3	C4	C5	C6	C7	C8
ARM	0.0001788	0.0312187	0.0060435	0.0001430	0.0000000	0.0015735	0.0054713	0.0002861
MLI	0.0332092	0.0065039	0.0017241	0.0788586	0.0954897	0.0052678	0.0122810	0.0186735
FRA	0.0003979	0.0810118	0.0140813	0.0083357	0.0000000	0.0016622	0.0009473	0.0000810
MOZ	0.0109682	0.0061453	0.0013705	0.0141875	0.0825047	0.0028777	0.0072786	0.0083276
AUS	0.0002859	0.0698592	0.0149352	0.0010193	0.0000000	0.0012182	0.0020013	0.0000808
ARG	0.0013440	0.0357363	0.0090283	0.0039460	0.0000000	0.0016793	0.0085476	0.0011185
USA	0.0004042	0.0507657	0.0113620	0.0021482	0.0000000	0.0012751	0.0062466	0.0003492

Table 3: Profils colonnes pour quelques pays et causes sélectionnés

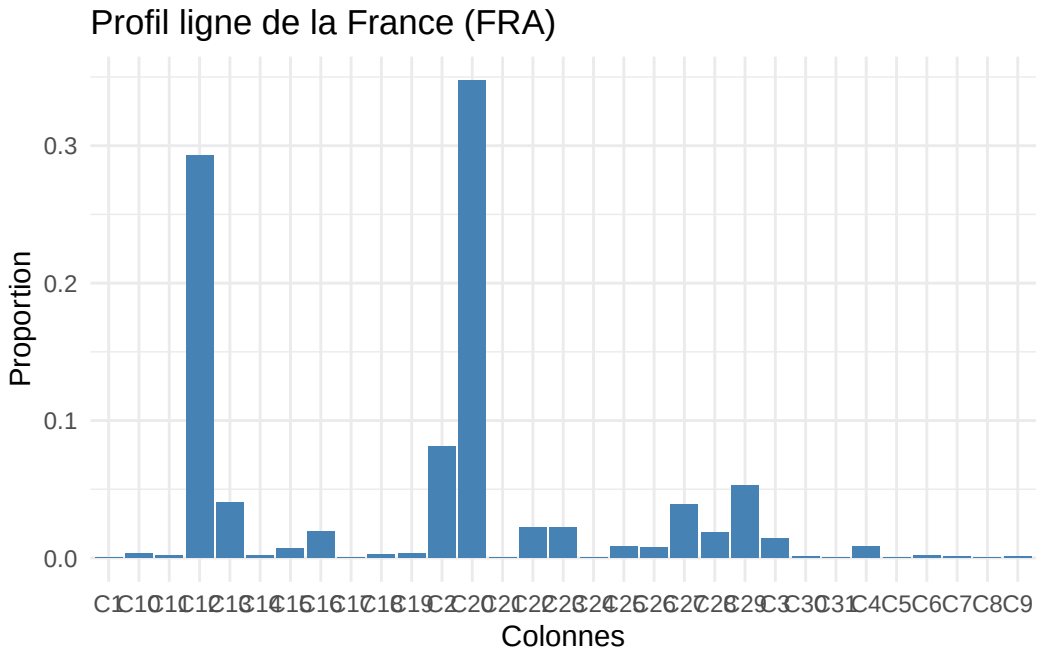
	C1	C2	C3	C4	C5	C6	C7	C8
ARM	0.0000212	0.0005381	0.0004659	0.0000159	0.0000000	0.0001856	0.0003694	0.0000408
MLI	0.0265160	0.0007557	0.0008961	0.0591263	0.0279850	0.0041887	0.0055897	0.0179312
FRA	0.0009573	0.0283575	0.0220484	0.0188297	0.0000000	0.0039820	0.0012990	0.0002343
MOZ	0.0115594	0.0009424	0.0009402	0.0140408	0.0319154	0.0030202	0.0043727	0.0105549
AUS	0.0001948	0.0069279	0.0066253	0.0006523	0.0000000	0.0008268	0.0007775	0.0000662
ARG	0.0019188	0.0074241	0.0083898	0.0052901	0.0000000	0.0023875	0.0069563	0.0019205
USA	0.0048542	0.0887060	0.0888084	0.0242233	0.0000000	0.0152487	0.0427591	0.0050431

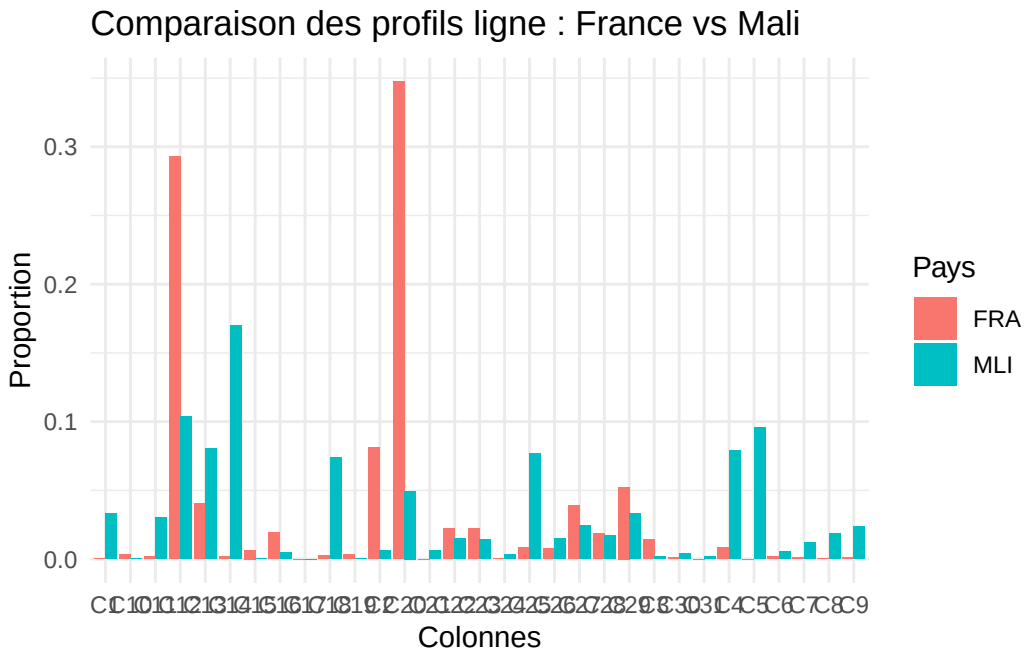
De manière complémentaire, les profils colonnes décrivent, pour chaque cause de décès, la répartition relative des pays. Chaque valeur indique la contribution d'un pays donné au total des décès observés pour une cause spécifique . En 2019, l'Afghanistan représentait 0,66 %

(0.0066= 1563/(1563 +13+…+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`

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.

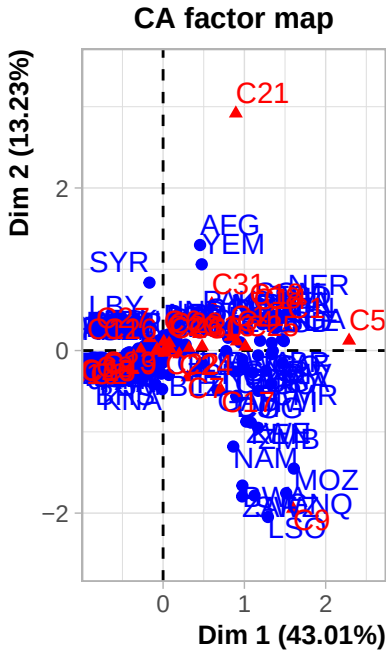
Afin de quantifier l'intensité de la liaison entre les deux variables, le coefficient de V de Cramér est calculé. La valeur obtenue étant 0.149 indique une liaison faible entre le pays et la cause de décès, malgré le rejet de l'hypothèse d'indépendance. Ce résultat souligne que la dépendance statistique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

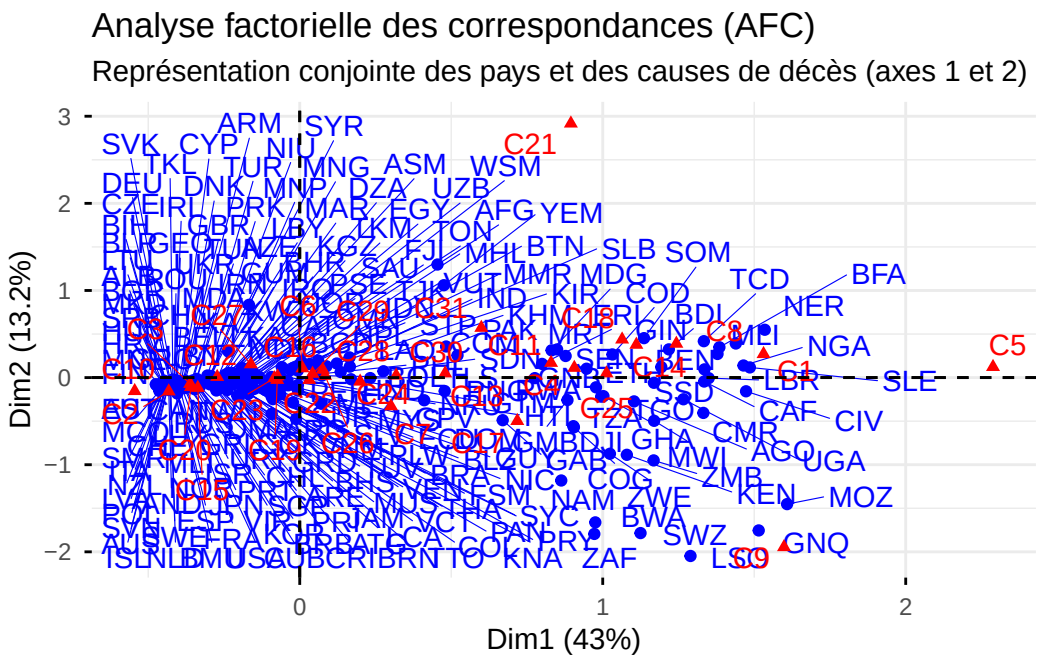
L'examen des effectifs attendus sous l'hypothèse d'indépendance ainsi que des résidus standardisés permet d'identifier les principales contributions au khi-deux. Ces écarts traduisent des phénomènes d'attirance et de répulsion : un résidu standardisé positif indique qu'un pays et une cause se rencontrent plus fréquemment que prévu (attirance), 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.

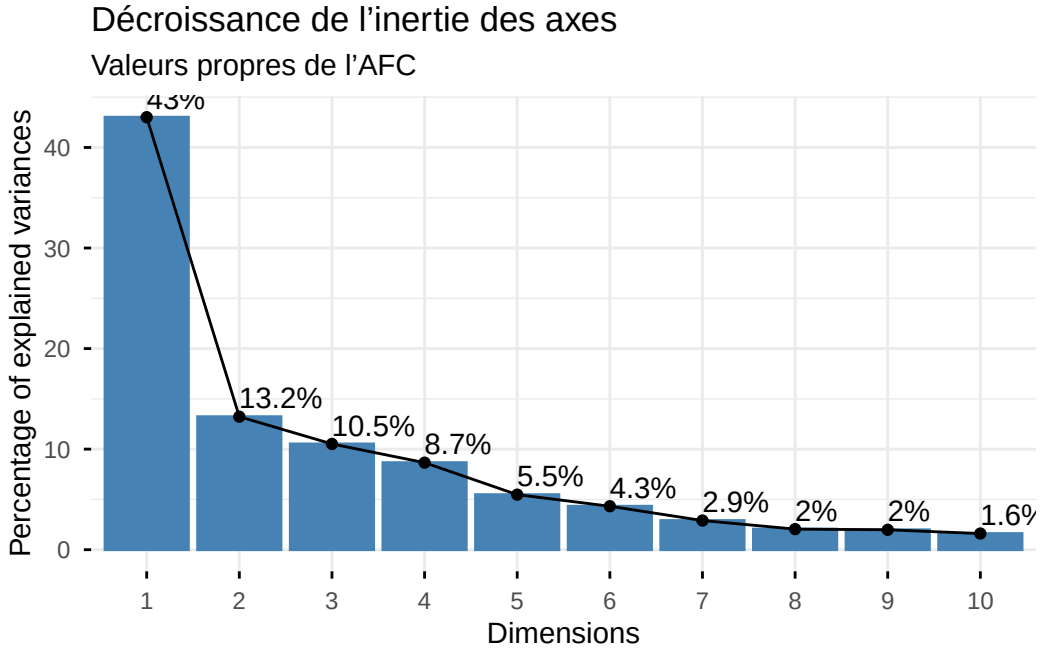
À titre d'illustration, le Bangladesh (BGD) présente un résidu standardisé de $-99{,}09$ pour la malaria (C5), ce qui traduit une répulsion extrêmement marquée : cette cause est très fortement sous-représentée dans ce pays par rapport à ce qui serait attendu sous l'hypothèse d'indépendance. À l'inverse,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

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.



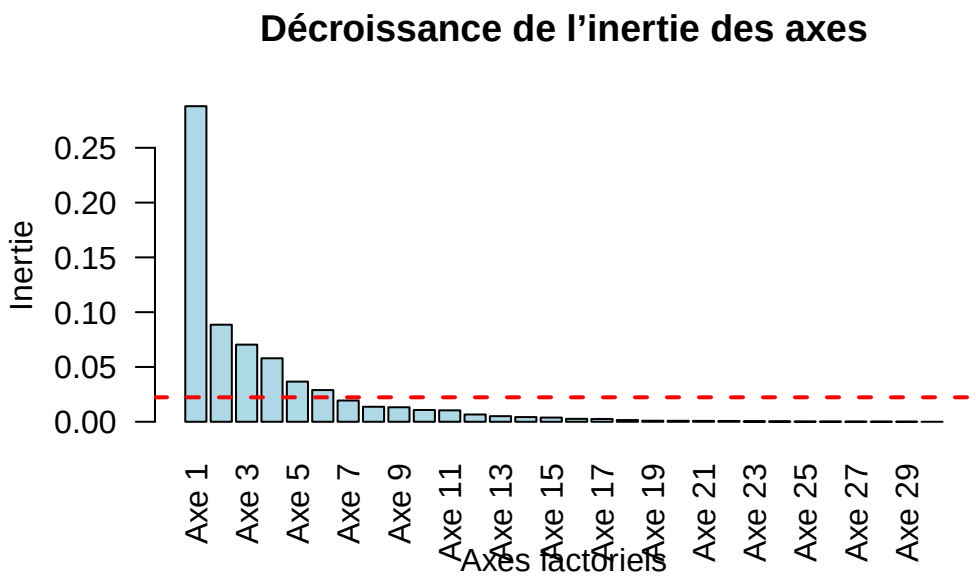


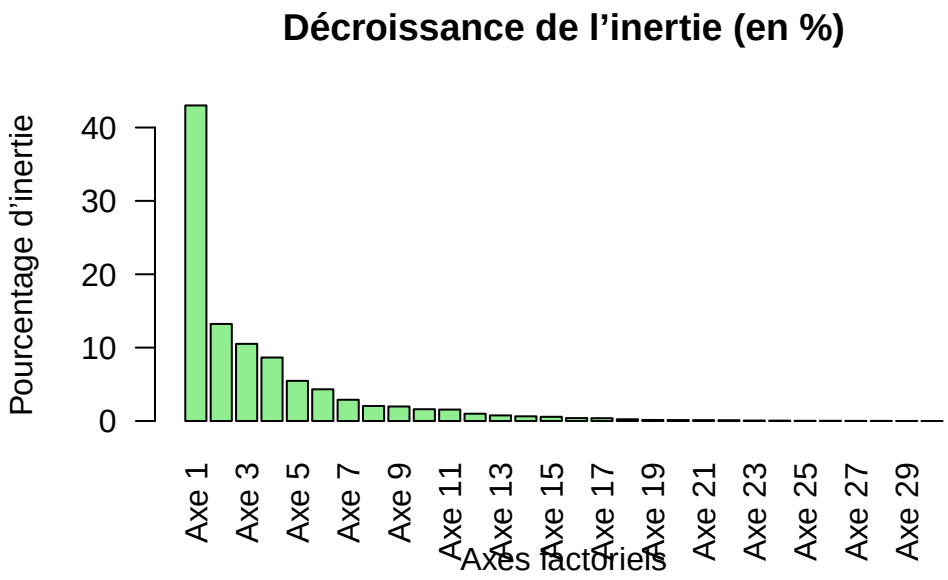


L'AFC est appliquée aux 204 pays du tableau, considérés comme lignes actives. Les valeurs propres associées aux axes factoriels sont examinées afin de déterminer la part d'inertie expliquée par chaque axe et d'identifier le nombre d'axes significatifs à retenir. La décroissance des valeurs propres est 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.



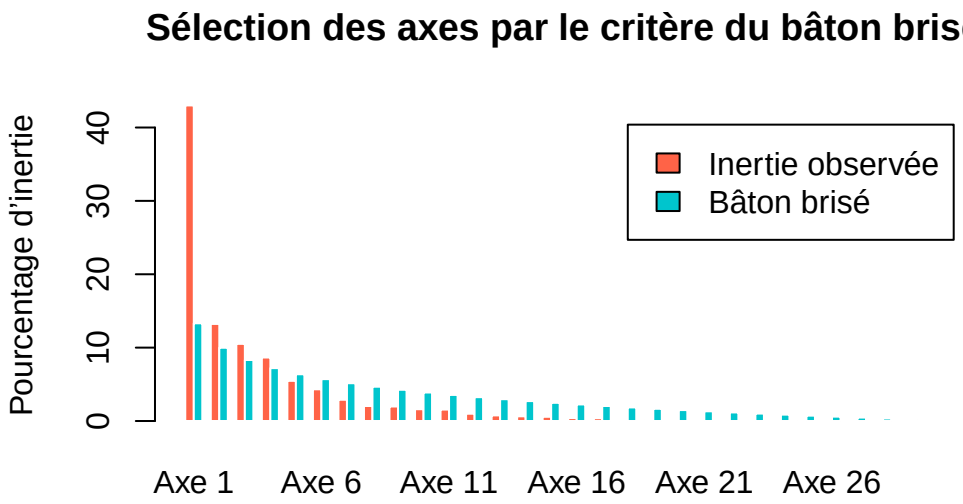


Les diagrammes de l'inertie totale et du pourcentage d'inertie permettent de visualiser la contribution relative de chaque axe à la structure globale des données, préparant ainsi l'interprétation des positions des pays et des causes sur les plans factoriels. Cette étape constitue un prérequis essentiel 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

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.

```
 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  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  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
```



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.

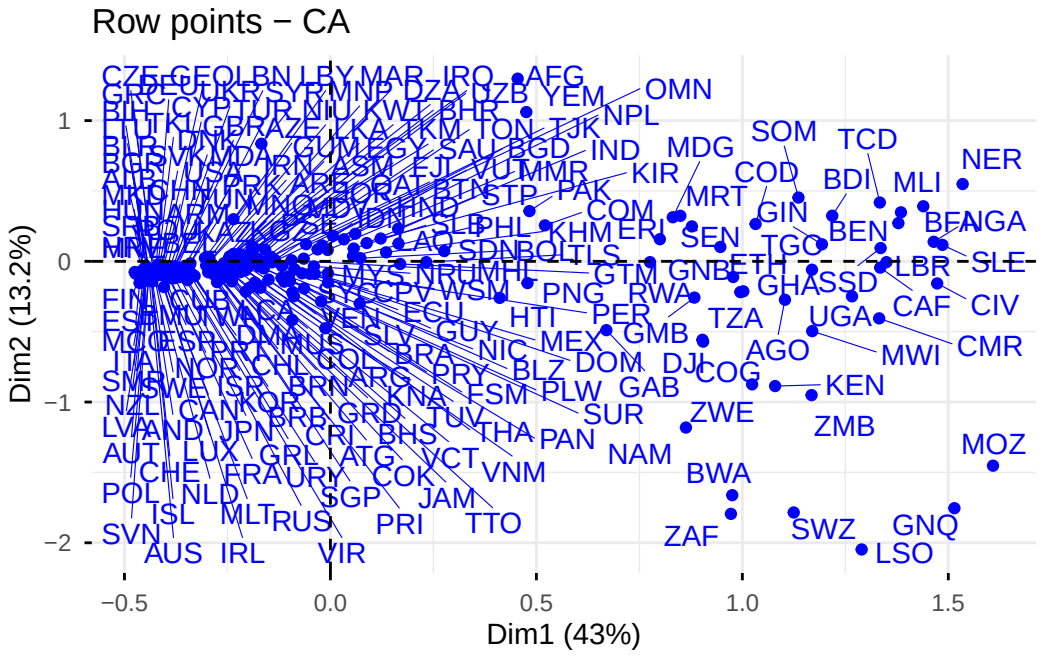
Un diagramme comparatif illustre visuellement la sélection des axes : les barres rouges représentant l'inertie observée sont mises côte à côte avec celles du bâton brisé, représentées en bleu. Cette représentation permet de visualiser aisément quels axes dépassent le seuil et confirment leur pertinence 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.

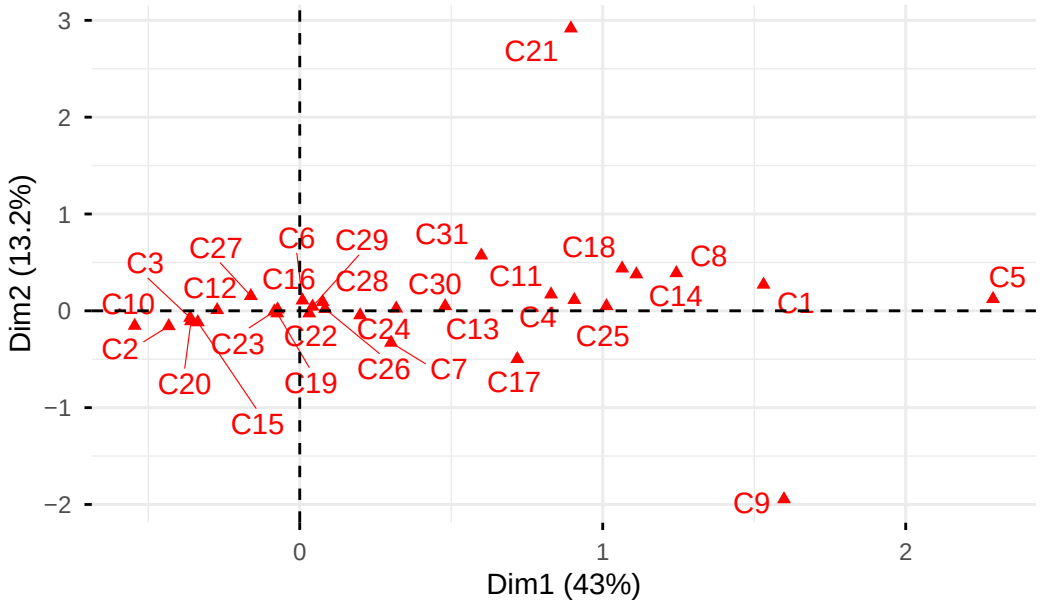
4.4 Analyse des contributions des lignes et colonnes (il faut adapter le texte à partir d'ici)

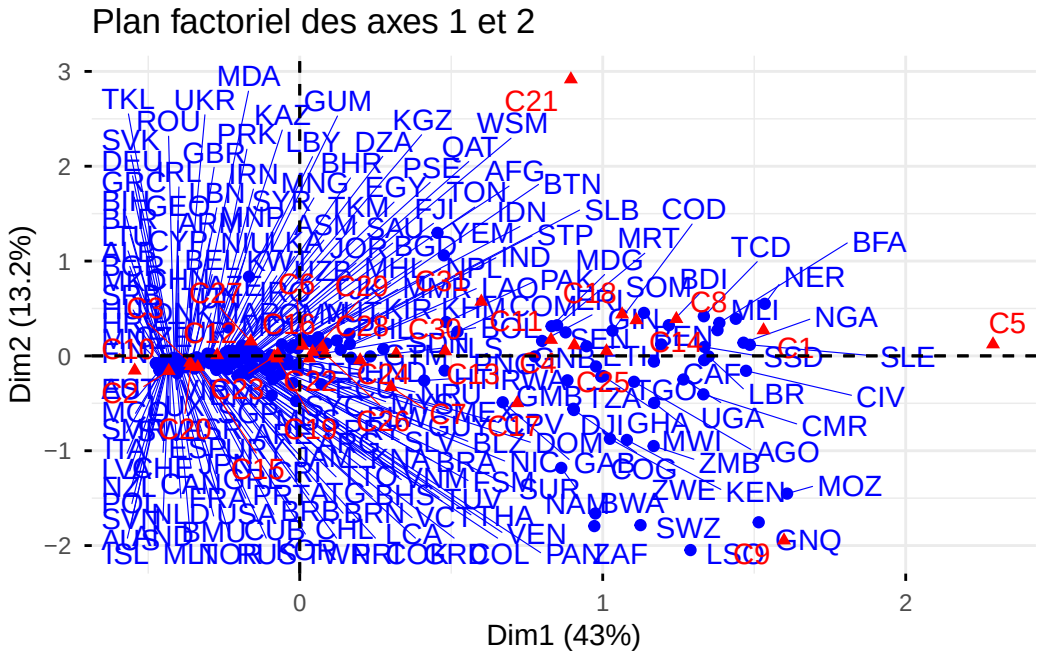
Après sélection des axes significatifs, l'AFC permet de représenter les modalités des lignes (pays) et des colonnes (causes) dans l'espace factoriel. Les diagrammes factoriels des lignes et des colonnes mettent en évidence les similitudes et les différences structurelles : les pays proches sur le plan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

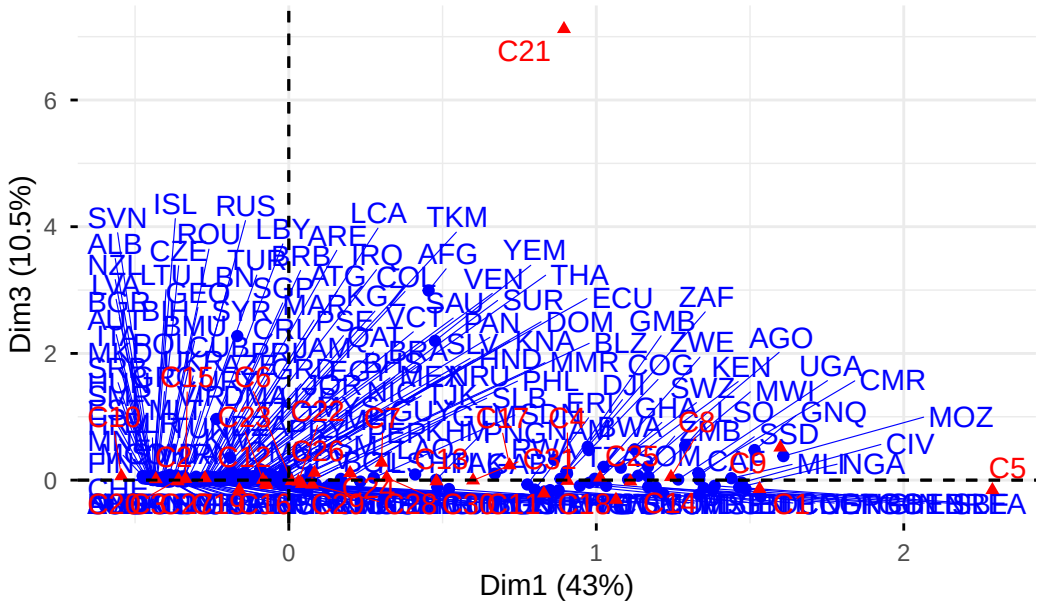
Les biplots combinant lignes et colonnes sur les premiers plans factoriels fournissent une vue d'ensemble des relations entre pays et causes. Le premier plan factoriel (axes 1 et 2) est

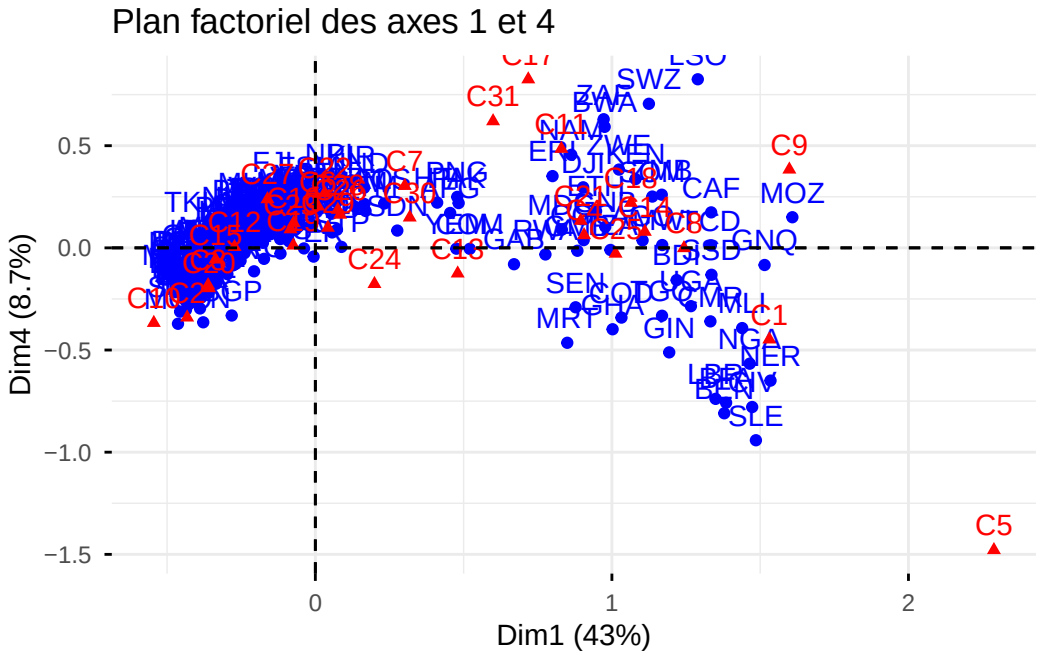
généralement le plus informatif, mais d'autres combinaisons d'axes peuvent également révéler des dimensions secondaires de variation.

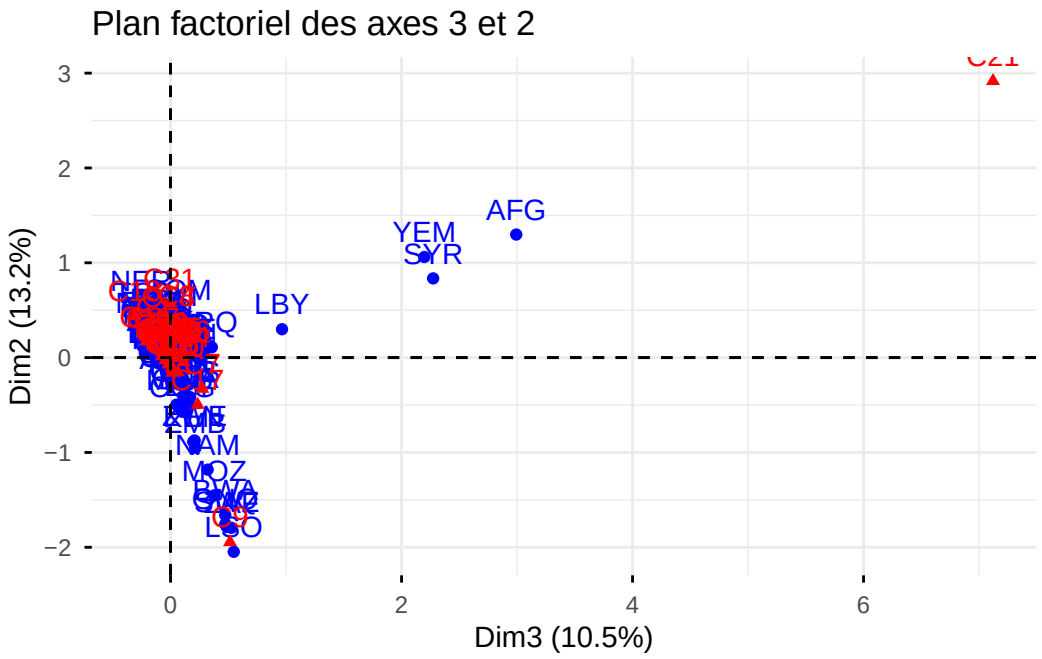


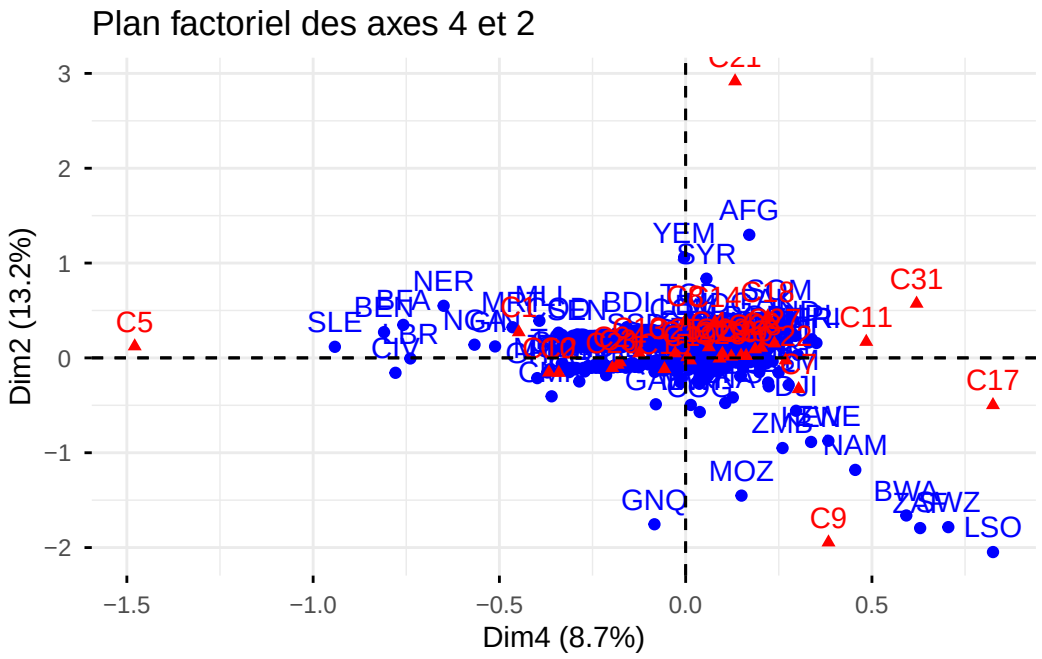


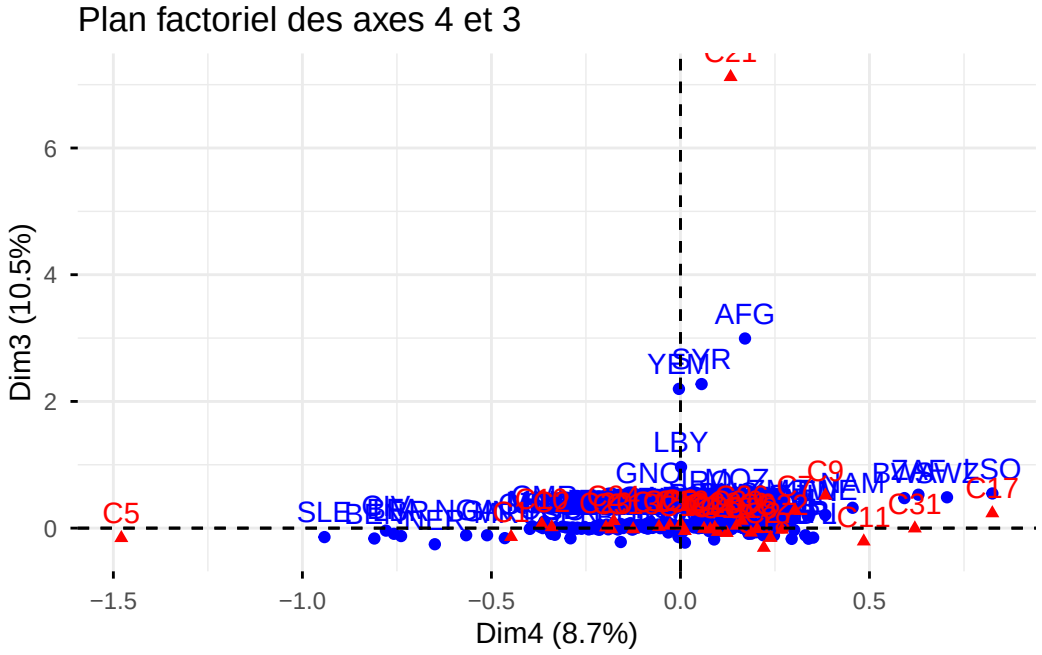

Plan factoriel des axes 1 et 2
Dim2 (13.2%)
Dim1 (43%)
3
2
1
0
−1
−2
0
1
2
MDA
TKL
UKR
GUM
C21
ROU
KAZ
KGZ
WSM
SVK
PRK
LBY
DZA
OAT
DEU
GBR
BHR
PSE
AFG
GRC
IRN
MNG
EGY
TON
BTN
FJI
IDN
SLB
COD
YEM
STP
MRT
TCD
BFA
IND
MDG
BDI
NER
SOM
C8
MLI
NGA
C5
C1
SSD
SLE
CAF
LBR
CIV
GMB
ZAF
TGO
UGA
CMR
DJI
GHA
MWI
AGO
ZMB
BRA
NIC
GAB
COG
ZWE
KEN
MOZ
NAM
BWA
SWZ
GNQ
C9
ISL
RUS
PAN
ZAF


Plan factoriel des axes 1 et 3
Dim3 (10.5%)
Dim1 (43%)
6
4
2
0
0
1
2
C21
ISL
RUS
SVN
LCA
TKM
ROU
LBY
ARE
ALB
CZE
TUR
SRB
IRQ
AFG
YEM
NZL
LTU
LBN
ATG
COL
VEN
THA
LVA
SGP
GEO
KGZ
SAU
SUR
ECU
ZAF
BGR
BIH
SYR
MAR
PSE
VCT
PAN
DOM
GMB
ARG
BMU
CRI
BRA
SLV
KNA
BLZ
ZWE
AGO
HND
MMR
COG
KEN
UGA
CMR
PHL
DJI
SWZ
MWI
GHA
LSO
GNQ
MOZ
SSD
CIV
MLI
NGA
C5


Plan factoriel des axes 1 et 4
Dim4 (8.7%)
Dim1 (43%)
0.5
0.0
−0.5
−1.0
−1.5
0
1
2
C31
C9
C24
C5
C1
SWZ
MOZ
CAF
SEN
MRT
GIN
SLE
NER
AFG
YEM
SYR
LBY
NAM
LSO

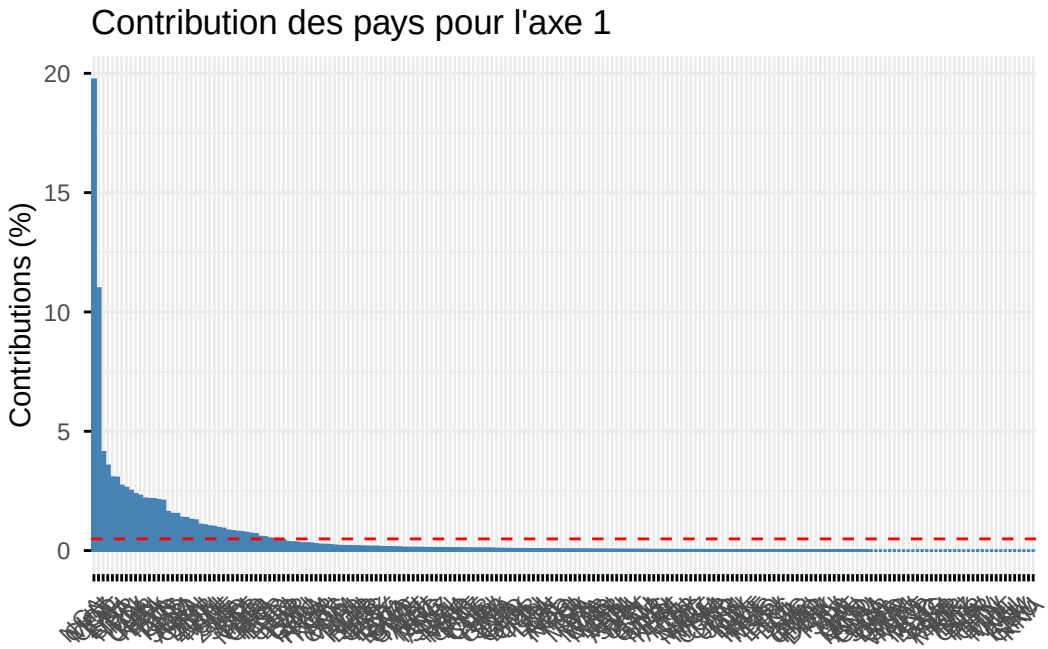

Plan factoriel des axes 3 et 2
Dim2 (13.2%)
Dim3 (10.5%)
3
2
1
0
−1
−2
0
2
4
6
AFG
YEM
SYR
LBY
NAM
MOZ
LSO

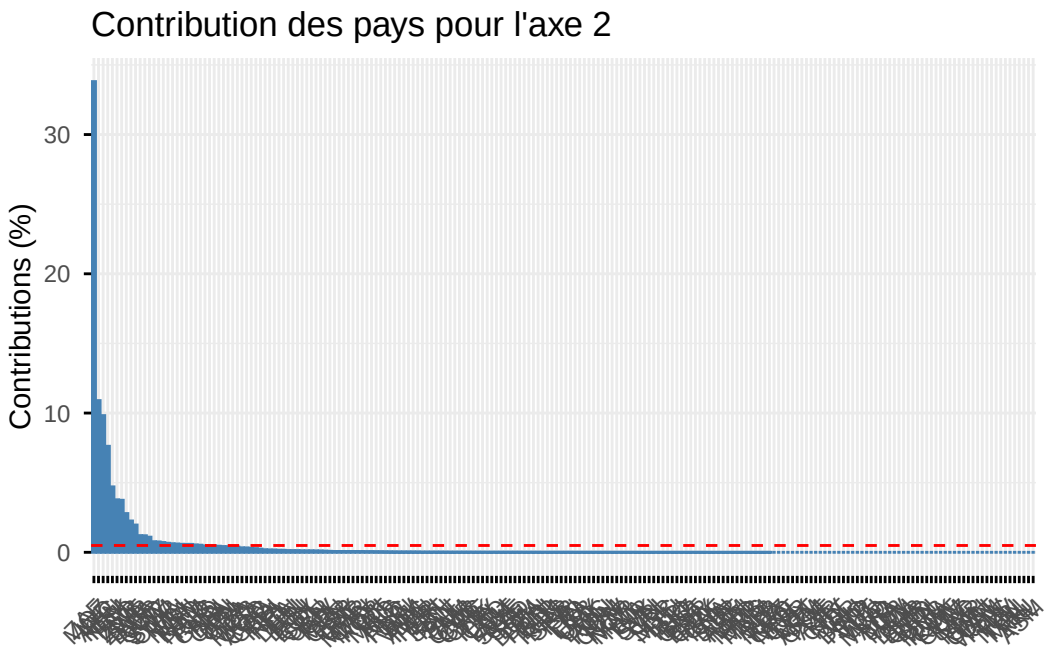


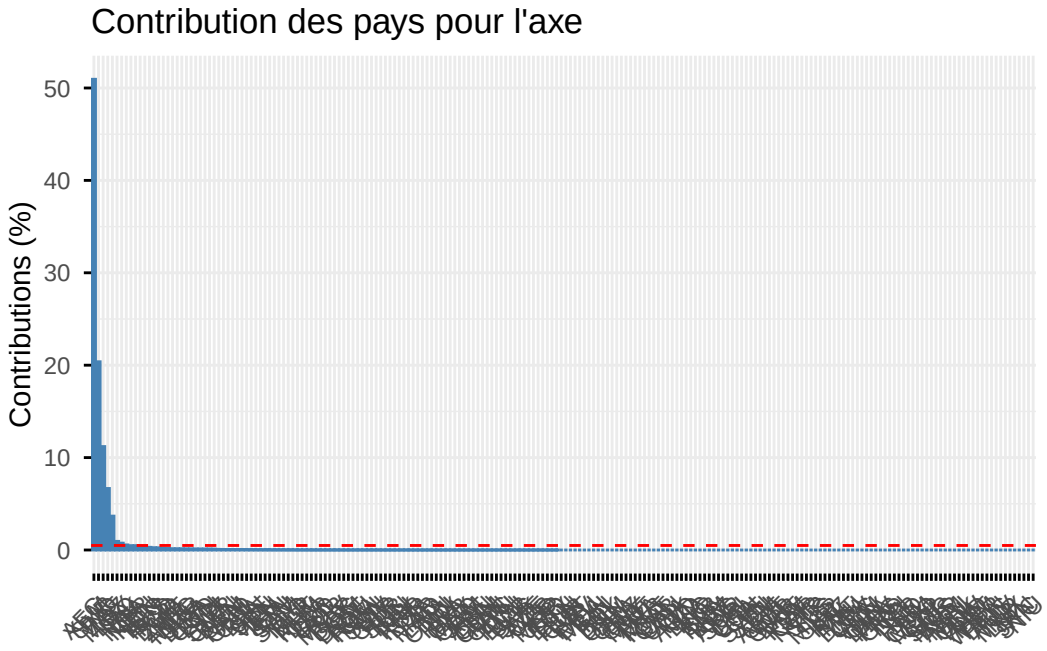


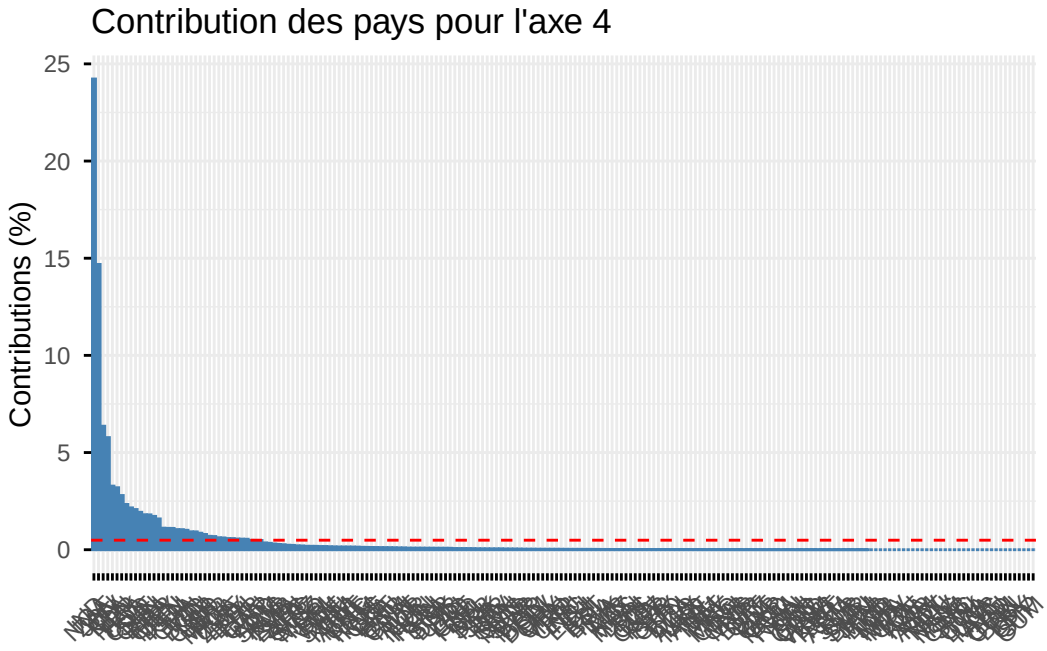
Pour approfondir l'interprétation, les contributions des lignes et des colonnes aux axes factoriels sont examinées. Les modalités les plus contributives sont identifiées, ce qui permet de se concentrer sur les pays et causes qui expliquent le mieux la variance sur chaque axe. Étant donné le grand nombre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

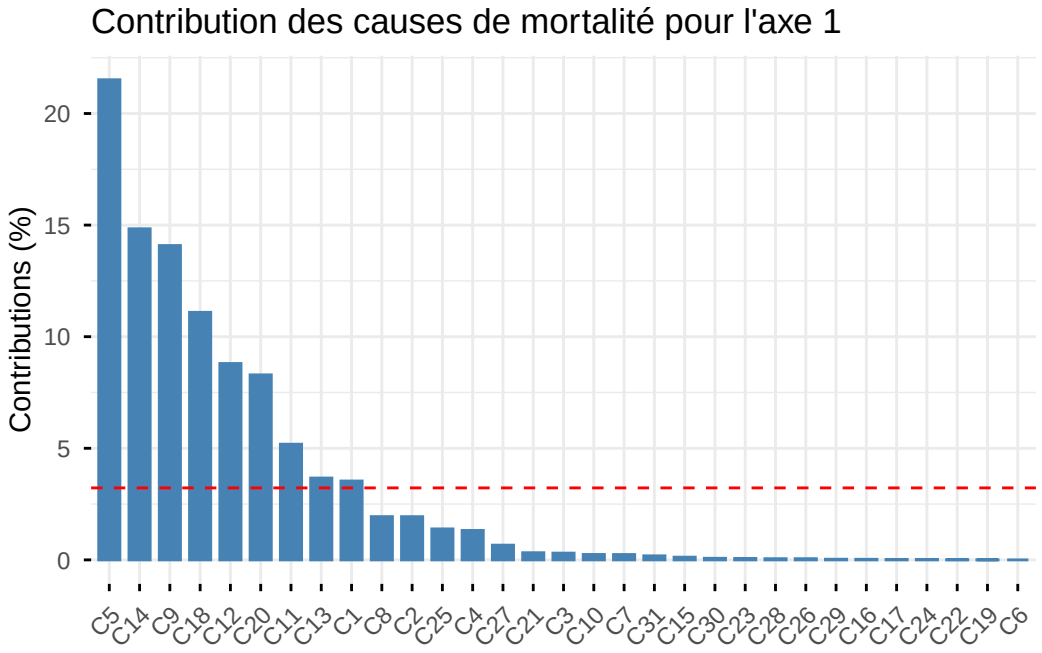
chaque axe facilite la lecture et l'interprétation des résultats.

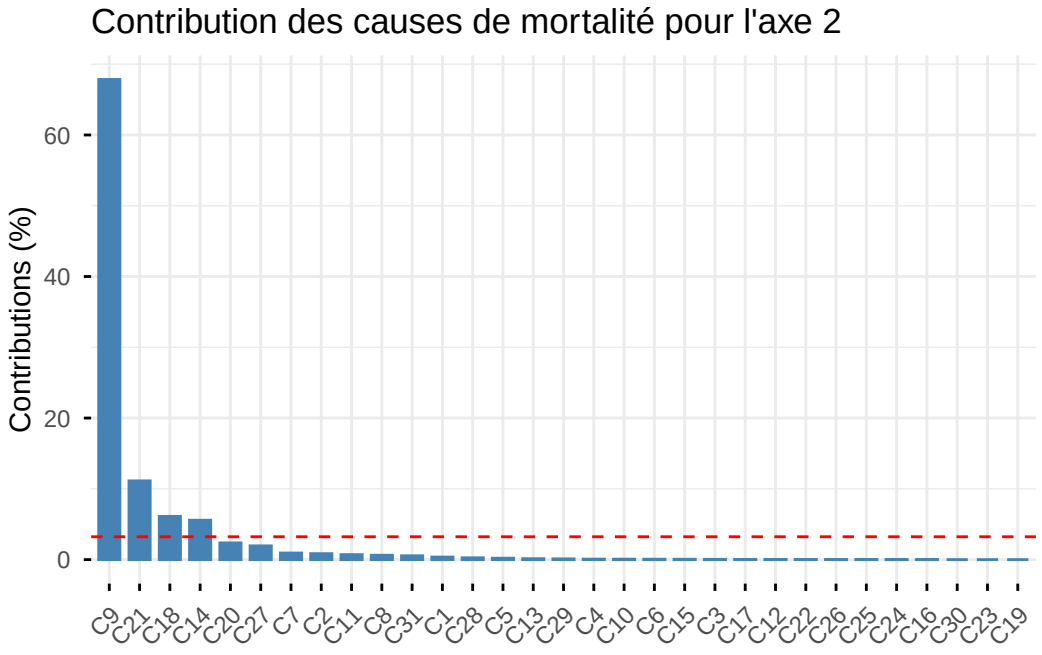


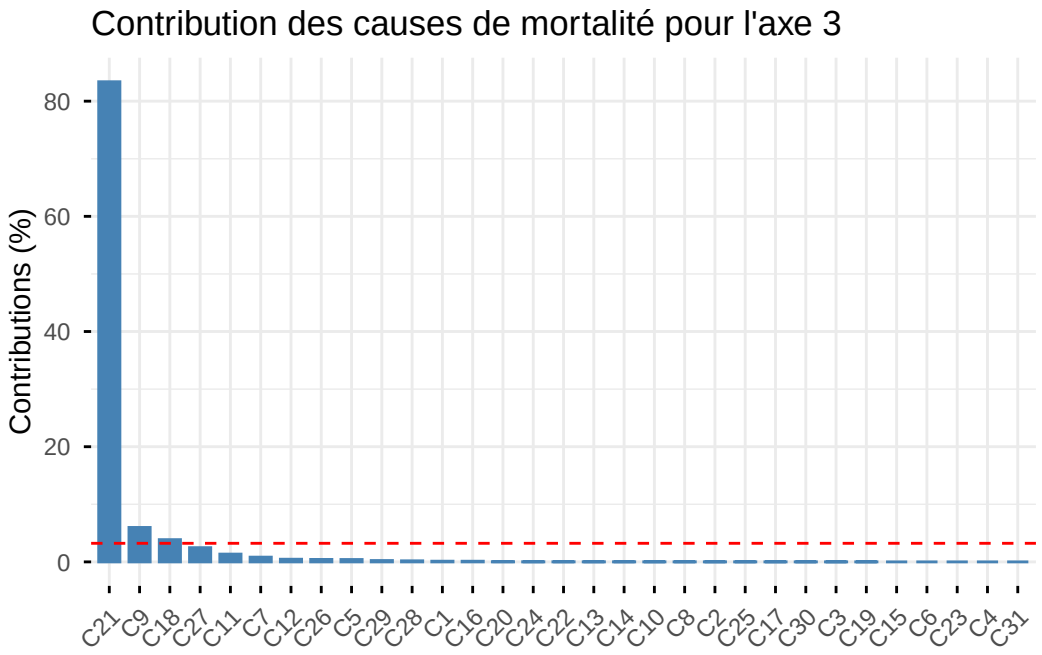


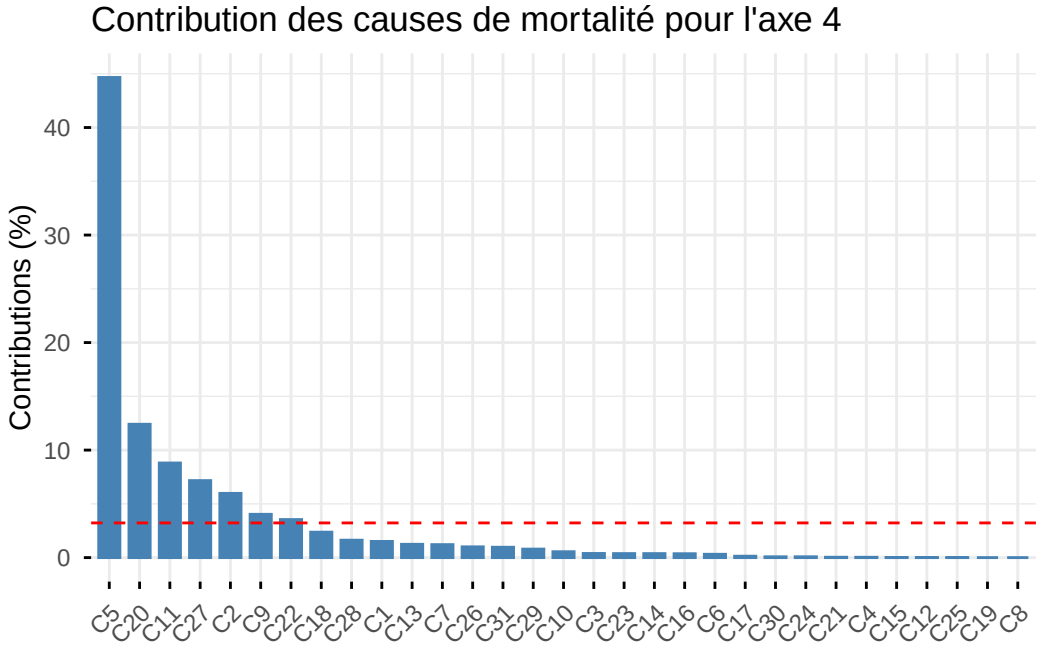

Contribution des pays pour l'axe
Contributions (%)
50
40
30
20
10
0


Contribution des pays pour l'axe 4
Contributions (%)
25
20
15
10
5
0


Contribution des causes de mortalité pour l'axe 1
Contributions (%)
0
5
10
15
20
C5 C14 C9 C18 C12 C20 C11 C13 C1 C8 C2 C25 C4 C27 C21 C3 C10 C7 C31 C15 C30 C23 C28 C26 C29 C16 C17 C24 C22 C19 C6

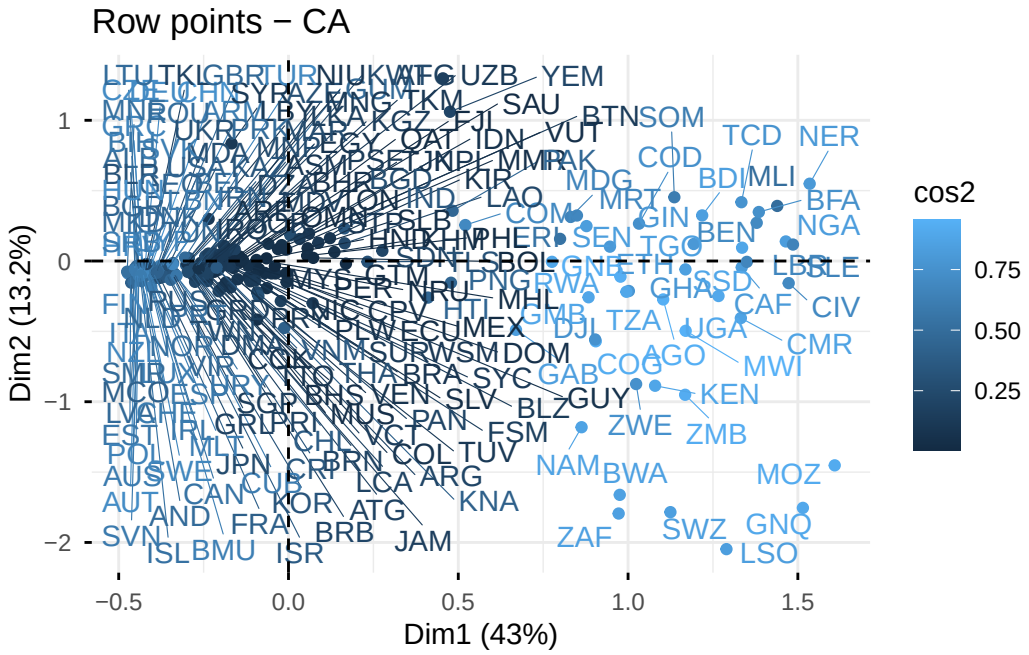

Contribution des causes de mortalité pour l'axe 2
Contributions (%)
0
20
40
60
C9 C21 C18 C14 C20 C27 C7 C2 C11 C8 C31 C1 C28 C5 C13 C29 C4 C10 C6 C15 C3 C17 C12 C22 C26 C25 C24 C16 C30 C23 C19

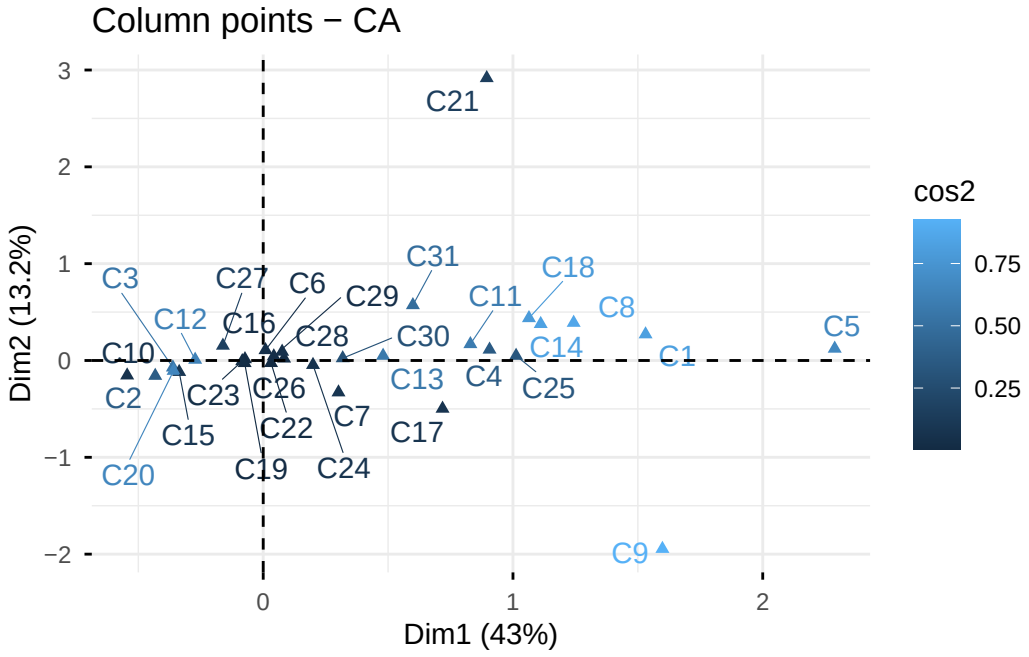


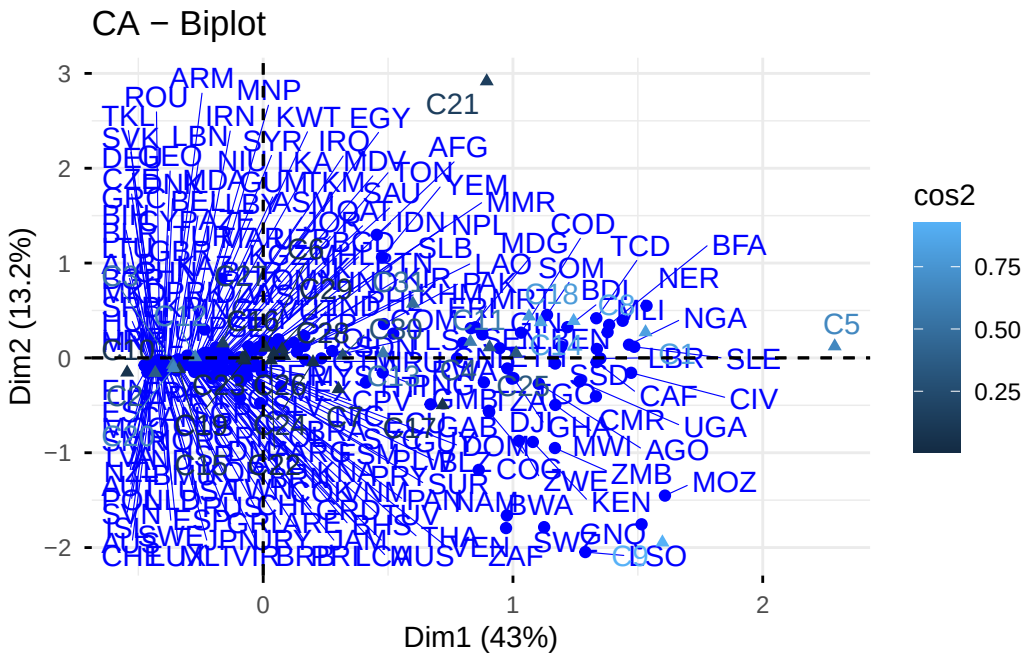


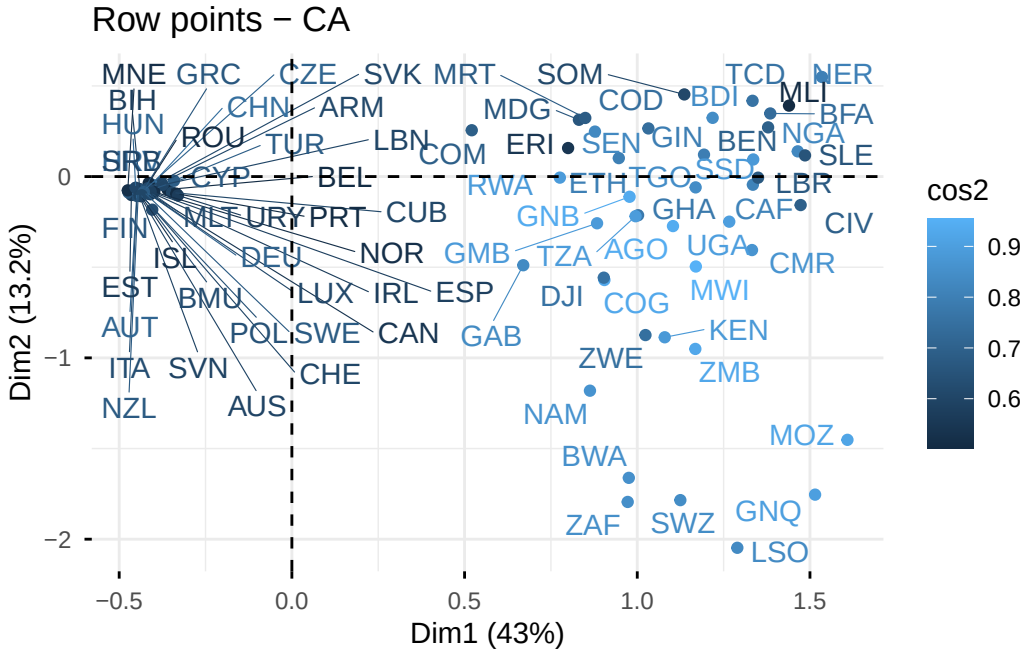
La qualité de représentation de chaque modalité est évaluée à l'aide du cos2, indiquant la proportion de la variance de la modalité expliquée par les axes retenus. Les modalités avec un cos2 élevé sont mieux représentées sur le plan factoriel et constituent des points de référence fiables pour l'interprétation.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

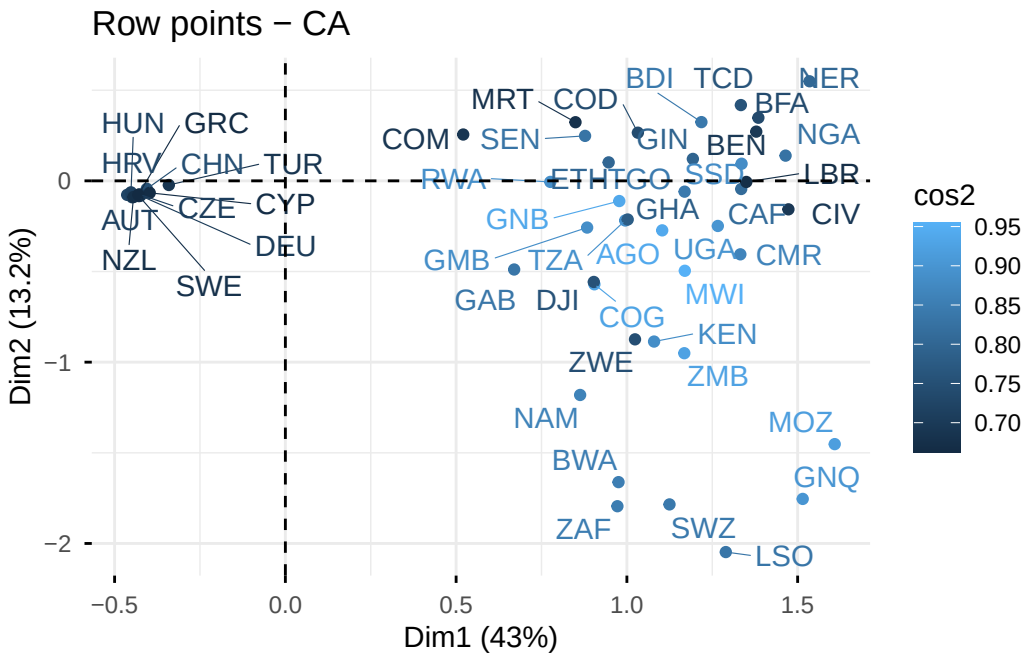
ment les pays et causes les mieux représentés, tout en réduisant l'encombrement visuel des graphiques.






CA − Biplot
Dim2 (13.2%)
Dim1 (43%)
cos2
0.75
0.50
0.25
C21
C5
ARM
MNP
ROU
TKL
IRN
KWT
EGY
SVK
LBN
SYR
IRQ
AFG
DEU
GEO
NIU
LKA
MDV
TON
YEM
MMR
IDN
NPL
COD
SLB
MDG
TCD
BFA
SOM
NER
LAO
PAK
BDI
MLI
NGA
LBR
SLE
SSD
CAF
CIV
CMR
UGA
GAB
DJI
GHA
MWI
AGO
COG
ZWE
ZMB
MOZ
BWA
KEN
GNQ
SWZ
LSO
ZAF
THA
VEN
MUS


Row points − CA
Dim2 (13.2%)
Dim1 (43%)
cos2
0.9
0.8
0.7
0.6
MNE
GRC
CZE
SVK
MRT
SOM
TCD
NER
BIH
CHN
ARM
MDG
COD
BDI
MLI
HUN
ROU
TUR
LBN
ERI
SEN
GIN
BEN
BFA
NGA
SRB
COM
SSD
SLE
CYP
BEL
RWA
ETH
TGO
LBR
MLT
URY
PRT
CUB
GNB
GHA
CAF
CIV
FIN
NOR
GMB
TZA
AGO
UGA
CMR
ISL
DEU
EST
BMU
LUX
IRL
ESP
DJI
COG
MWI
AUT
POL
SWE
CAN
GAB
KEN
ZWE
ITA
SVN
CHE
ZMB
NZL
AUS
NAM
MOZ
BWA
ZAF
SWZ
GNQ
LSO

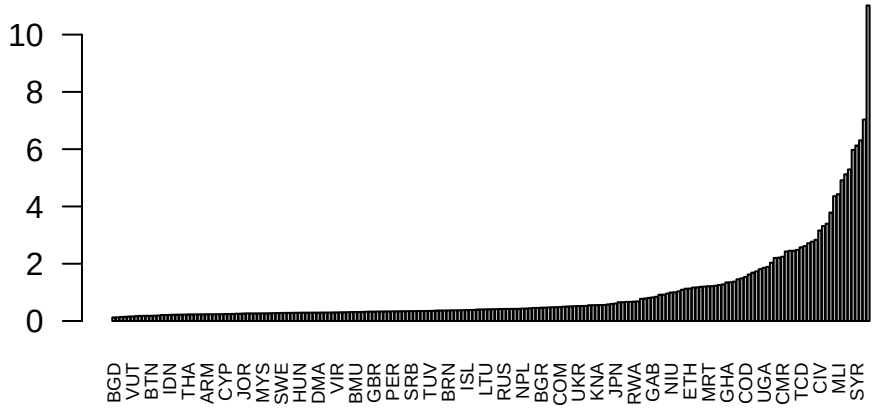


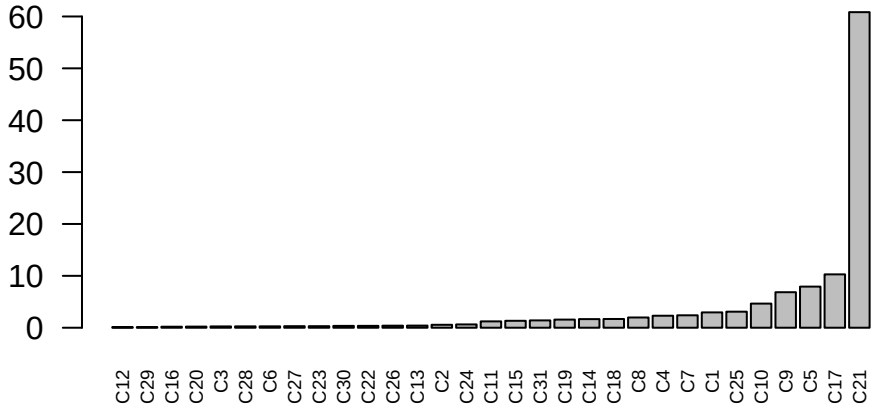
Cette analyse graphique et quantitative des contributions et de la qualité de représentation fournit une base solide pour tirer des conclusions sur les profils de mortalité et les relations entre pays et causes dans l'ensemble du jeu de données.

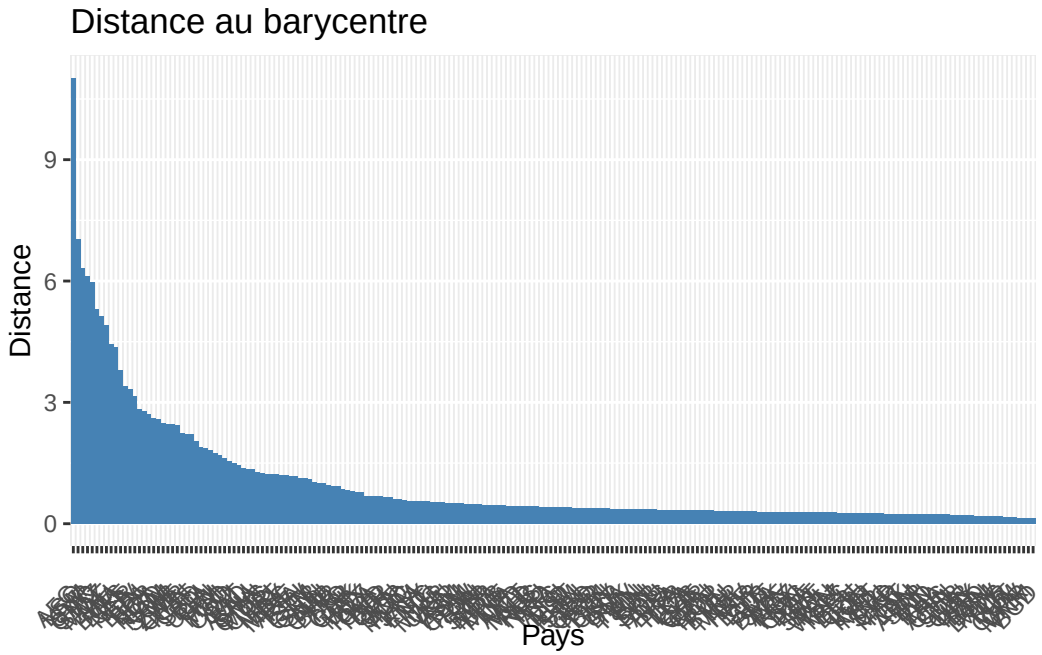
4.5 Distances au centre de gravité

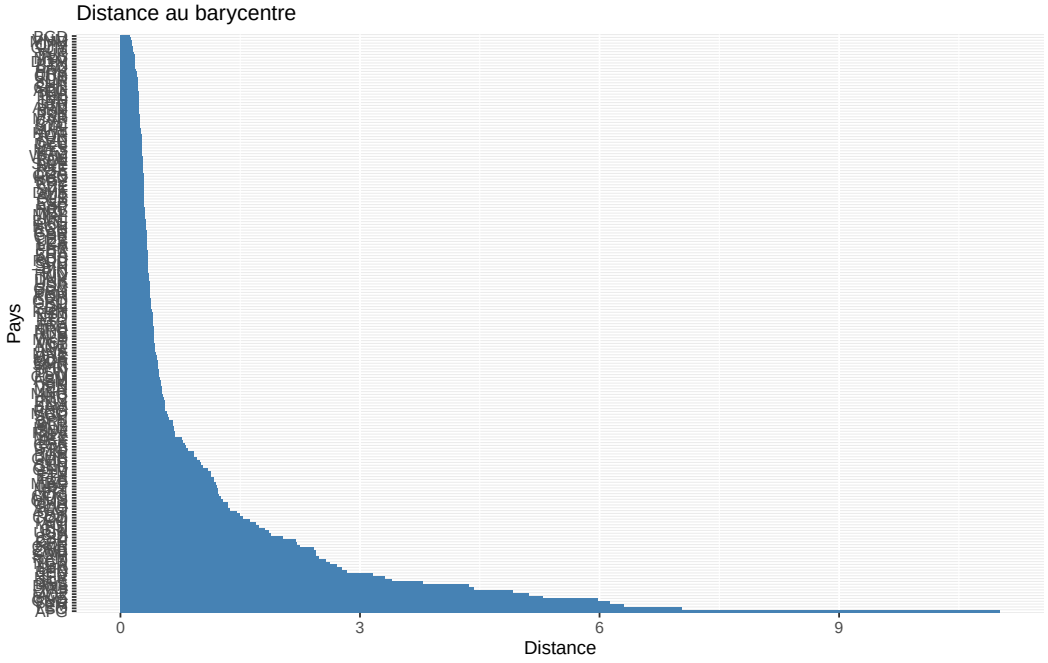
L'AFC permet également de calculer la distance de chaque pays et de chaque cause au centre de gravité (barycentre) de l'ensemble des données. Cette distance est obtenue en rapportant l'inertie d'une modalité à sa masse, ce qui fournit une mesure de l'écart par rapport au profil moyen. Les pays dont la 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.


10
8
6
4
2
0
BGD
VUT
BTN
IDN
THA
ARM
CYP
JOR
MYS
SWE
HUN
DMA
VIR
BMU
GBR
PER
SRB
TUV
BRN
ISL
LTU
RUS
NPL
BGR
COM
UKR
KNA
JPN
RWA
GAB
NIU
ETH
MRT
GHA
COD
UGA
CMR
TCD
CIV
MLI
SYR


60
50
40
30
20
10
0
C12
C29
C16
C20
C3
C28
C6
C27
C23
C30
C22
C26
C13
C2
C24
C11
C15
C31
C19
C14
C18
C8
C4
C7
C1
C25
C10
C9
C5
C17
C21


Distance au barycentre
Distance
9
6
3
0
Pays


Distance au barycentre
Pays
0
3
6
9
Distance

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