MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE RECHERCHE SCIENTIFIQUE

REPUBLIQUE DE COTE D'IVOIRE

UNION-DISCIPLINE-TRAVAIL

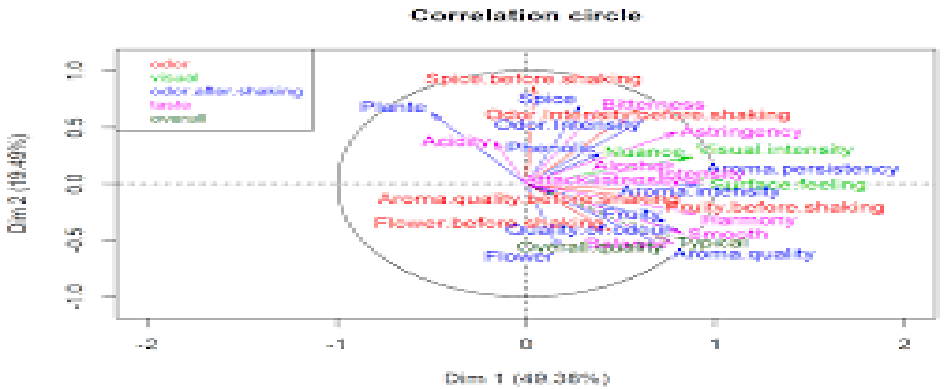
UNION - DISCIPLINE - TRAVAIL

Master 2

STATISTIQUE-ECONOMETRIE-DATA SCIENCE

MINI-PROJET
ANALYSE DES DONNEES MULTIDIMENSIONNELLES
CATEGORISER LES PAYS EN FONCTIONS DES FACTEURS SOCIO-ECONOMIQUES ET SANITAIRES QUI DETERMINENT LE DEVELOPPEMENT GLOBAL DU PAYS

2023-2024



ETUDIANTE :

WAWA LAURALIE MARIE-MICHELLE

ENSEIGNANT ENCADREUR :

AKPOSSO DIDIER MARTIAL

AVANT PROPOS

Cher lecteur,

C’est avec un grand enthousiasme que je vous présente ce mini-projet, fruit de Plusieurs semaines de réflexion, de recherche et de travail acharné. Ce mini-Projet est le résultat de mon engagement envers l’apprentissage, la créativité et la mise en pratique des connaissances acquises.

Ce mini-projet est de démontrer ma capacité à appliquer les compétences et les concepts que j’ai acquis dans le cadre de l’obtention du diplôme de Master Professionnel en Statistique-Econométrie- Machine Learning à l’institut Supérieur de Statistique d’Econométrie et de Data Science (INSSEDS). Le sujet abordé dans le cadre de ce mini-projet me semble à la fois stimulant et pertinent.

Dans un monde où les disparités socio-économiques persistent et où des populations entières luttent contre la pauvreté et les inégalités, l'action humanitaire revêt une importance cruciale. HELP International, en tant qu'organisation dévouée à l'amélioration des conditions de vie dans les pays en développement, se trouve à un carrefour stratégique. Avec un financement fraîchement acquis de 10 millions de dollars, il est impératif de prendre des décisions éclairées pour maximiser l'impact de chaque dollar dépensé.

Ce rapport s'inscrit dans cet objectif. En utilisant des données socio-économiques et sanitaires fournies pour un ensemble de pays, nous avons entrepris une analyse approfondie pour catégoriser ces nations en fonction de leur niveau de développement global. Ces catégories fourniront un cadre pour guider le PDG de HELP International dans la prise de décision sur la distribution des ressources.

Je tiens à remercier M. AKPOSSO Didier Martial, Directeur des études et Encadreur de ladit formation, les enseignants de l’INSSEDS, ma famille et mes amis qui m’ont soutenu tout au long de la rédaction de ce mini-projet en m’apportant leur expertise, leurs conseils et leur encouragement.

Enfin, je vous invite à parcourir ce document avec attention, en espérant qu’il vous apportera une vision claire et détaillé de mon travail, tout en suscitant votre intérêt et votre réflexion.

Merci de prendre le temps de découvrir ce mini-projet. Je vous souhaite une agréable lecture.

Cordialement

TABLES DE MATIERES

Contenu

SIGLES ET ABREVIATIONS

INSSEDS : Institut Supérieur de statistique d'Econométrie et de data science

Child-mort : Enfant mort

Export : Exportation

Health : Dépenses totales

Imports : Importation

Icome : Revenu

Inflation : Inflation

Life-expec : Espérance de vie

Total-fer : Taux de fécondité

Gdpp : PIB par habitant

PIB : Produit intérieur brut

SOURCE DU CODE R

INTRODUCTION GENERALE

Dans le paysage mondial actuel, les défis liés au développement socio-économique et sanitaire persistent, touchant particulièrement les populations des pays en développement. Face à ces réalités, les organisations humanitaires telles que HELP International jouent un rôle crucial dans la lutte contre la pauvreté et la promotion du bien-être dans ces régions.

Avec un financement récemment acquis de 10 millions de dollars, HELP International se trouve à un moment décisif. Le choix de l'allocation de ces ressources est une responsabilité majeure qui nécessite une approche stratégique et éclairée. Pour répondre à ce défi, une analyse approfondie des facteurs socio-économiques et sanitaires qui déterminent le développement global des pays est essentiel.

Ce rapport vise à fournir une feuille de route pour guider le PDG de HELP International dans la prise de décision sur la distribution des fonds disponibles. En utilisant des données fiables et des analyses approfondies, nous examinerons divers indicateurs tels que le taux de mortalité infantile, les dépenses de santé, le revenu par habitant, le taux d'inflation, l'espérance de vie, la fertilité et le PIB par habitant pour catégoriser les pays et identifier ceux qui ont le plus besoin d'aide.

L'objectif principal de cette analyse est d'optimiser l'impact des ressources limitées de HELP International en les ciblant là où elles peuvent avoir le plus grand effet positif sur les conditions de vie des populations les plus vulnérables. En fournissant une compréhension approfondie des besoins spécifiques de chaque pays, ce rapport aidera à orienter les efforts humanitaires vers une action plus efficace et efficiente.

Dans cette perspective, nous entamons une exploration approfondie des données disponibles, avec pour objectif ultime de fournir des recommandations stratégiques et éclairées pour orienter les initiatives de HELP International vers une contribution significative au développement global et à l'amélioration des conditions de vie dans les pays les plus nécessiteux.

Le présent document s’articulera autour de deux grandes parties :

➢ **PREMIERE PARTIE: PRETRAITEMENT DES DONNEES**

➢ **DEUXIEME PARTIE: ANALYSE EN COMPOSANTE PRINCIPALE**

CONTEXTE ET JUSTIFICATION DE L'ETUDE

Cette étude est fondamentale pour aider HELP International à réaliser sa mission humanitaire de manière efficace et efficiente, en travaillant pour un monde où chaque individu a accès aux ressources et aux opportunités nécessaires pour mener une vie digne et épanouissante.

PROBLEMATIQUE

Comment ces facteurs varient-ils d'un pays à l'autre, et quels sont les pays qui présentent les besoins les plus pressants en termes d'aide humanitaire et de développement ?

PRINCIPAUX RESULTATS ATTENDUS

Les principaux résultats attendus de cette étude sont de fournir à HELP International une feuille de route claire pour une action humanitaire efficace et ciblée, permettant à l'organisation de contribuer de manière significative à l'amélioration des conditions de vie dans les pays les plus vulnérables.

PREMIERE PARTIE : PRETAITEMENT DES DONNEES

I-PRETAITEMENT DES DONNEES

Le prétraitement des données est une étape cruciale dans le processus d'analyse des données, visant à préparer les données brutes en vue de les rendre appropriées pour une analyse ultérieure. Cette partie consiste, dans un premier temps, à importer le jeu de données dans un logiciel statistique. Ensuite, on procèdera à la visualisation du jeu de données afin d'en dégager la structure. Enfin, la dernière étape consistera à apurer le jeu de données. L'apurement de données, souvent appelé "data cleansing" en anglais, est le processus de nettoyage et de correction des données stockées dans une base de données ou un ensemble de données. L'objectif de l'apurement de données est d'identifier et de corriger les incohérences, les erreurs, les doublons et les données obsolètes ou incorrectes afin d'assurer la qualité et la précision des données.

1-DESCRIPTION DU JEU DE DONNEE : DICTIONNAIRE DES DONNEES

Le dictionnaire de données est une documentation détaillée qui répertorie et décrit chacune des variables du jeu de données. Il sert de référence pour comprendre la signification, la structure et les propriétés des données stockées, ce qui facilite la gestion, la maintenance, l'analyse et l'utilisation des données. Le jeu de données de notre étude se décrit comme suit :

Tableau 1: Dictionnaire des données

VARIABLE	NATURE	DESCRIPTION	MODALITES
Pays	Qualitative	Nom du pays	Modalité économique sociale , politique….
Child-mort	Quantitative	Décès d'enfants de moins de 5 ans pour 1000 naissances vivantes	Numérique
Exports	Quantitative	Exportations des biens et services par habitant	Numérique
Health	Quantitative	Dépenses totales de santé par habitant	Numérique
Imports	Quantitative	Importations des biens et services par habitant	Numérique
Income	Quantitative	Revenu net par personne	Numérique
Inflation	Quantitative	Mesure du taux de croissance annuel	Numérique
Life-expec	Quantitative	Nombre moyen d'années qu'un niveau née vivait si les schémas de mortalité actuels devaient rester les mêmes	Numérique
Total-fer	Quantitative	Le nombre d'enfants qui naitraient si le taux de fécondité restaient les mêmes	Numérique
Gdpp	Quantitative	Le PIB par habitant	Entier

2-Structure des données

Notre jeu de données est constitué de 167 observations et de 9 variables.

Tableau 2: Structure du jeu de donnée

```
'data.frame':   167 obs. of  9 variables:
 $ child_mort: num  90.2 16.6 27.3 119 10.3 14.5 18.1 4.8 4.3 39.2 ...
 $ exports   : num  10 28 38.4 62.3 45.5 18.9 20.8 19.8 51.3 54.3 ...
 $ health    : num  7.58 6.55 4.17 2.85 6.03 8.1 4.4 8.73 11 5.88 ...
 $ imports   : num  44.9 48.6 31.4 42.9 58.9 16 45.3 20.9 47.8 20.7 ...
 $ income    : int  1610 9930 12900 5900 19100 18700 6700 41400 43200 16000 ...
 $ inflation : num  9.44 4.49 16.1 22.4 1.44 20.9 7.77 1.16 0.873 13.8 ...
 $ life_expec: num  56.2 76.3 76.5 60.1 76.8 75.8 73.3 82 80.5 69.1 ...
```

```
 $ total_fer : num  5.82 1.65 2.89 6.16 2.13 2.37 1.69 1.93 1.44 1.92 ...
 $ gdpp      : int  553 4090 4460 3530 12200 10300 3220 51900
46900 5840 ...
```

3-Visualisation du jeu de donnée

Le tableau ci-dessous présente les cinq (5) premières et dernières observations de notre jeu de données.

Tableau 3 : Visualisation des 5 premières observations

	child_mort	exports	health	imports	income	inflation	life_expec	total_fer
Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82
Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65
Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89
Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16
Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13
	gdpp							
Afghanistan	553							
Albania	4090							
Algeria	4460							
Angola	3530							
Antigua and Barbuda	12200							

Tableau 4: Visualisation des 5 dernières observations

	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
Vanuatu	29.2	46.6	5.25	52.7	2950	2.62	63.0	3.50	2970
Venezuela	17.1	28.5	4.91	17.6	16500	45.90	75.4	2.47	13500
Vietnam	23.3	72.0	6.84	80.2	4490	12.10	73.1	1.95	1310
Yemen	56.3	30.0	5.18	34.4	4480	23.60	67.5	4.67	1310
Zambia	83.1	37.0	5.89	30.9	3280	14.00	52.0	5.40	1460

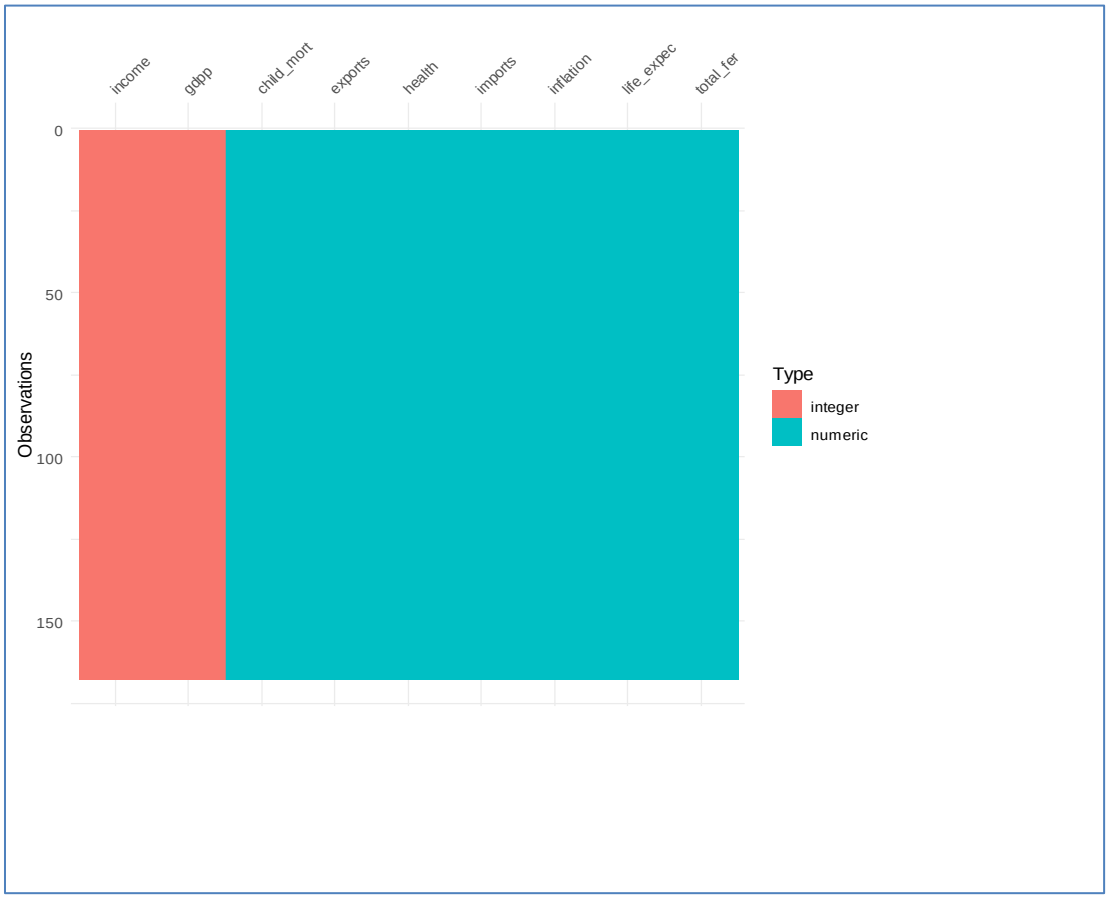
II-TRAITEMENT DES VALEURS MANQUANTES

En statistique, une valeur manquante, également appelée donnée manquante ou observation manquante, fait référence à l'absence d'une valeur pour une variable particulière dans un ensemble de données ou un échantillon.

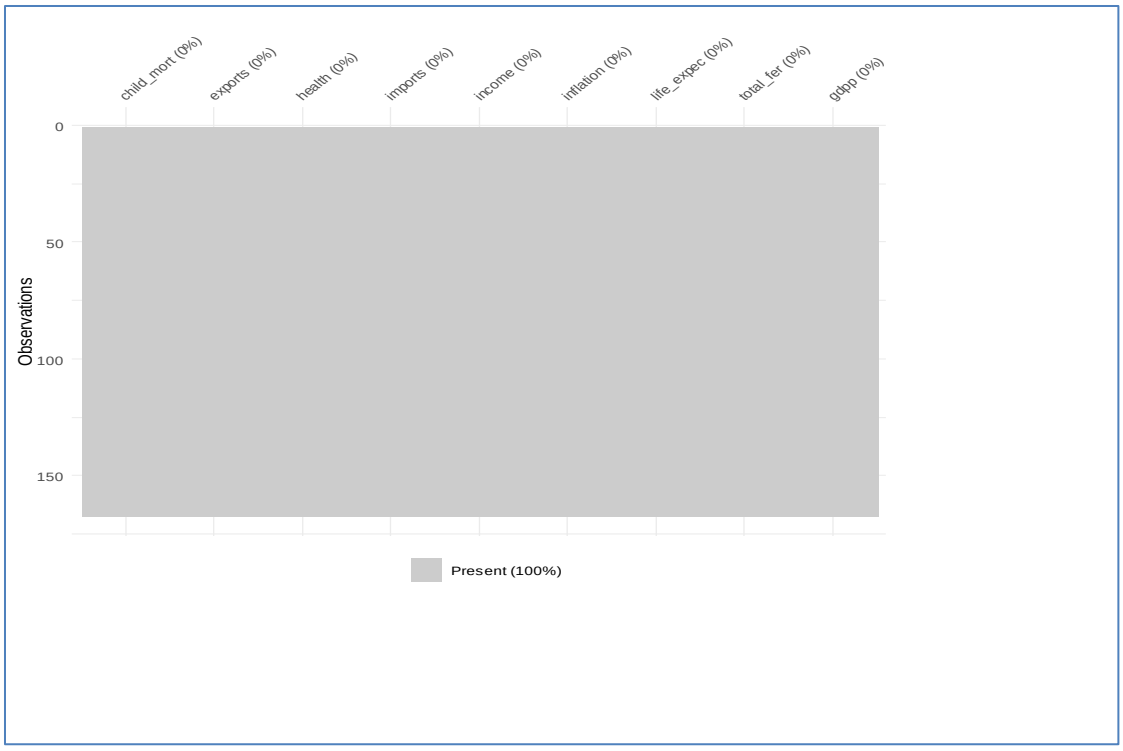
a-VISUALISATION DES DONNEES MANQUANTES

La visualisation du graphique 1 montre bien que le jeu de données ne contient pas des valeurs manquantes.

Graphique 1: Visualisation des valeurs manquantes



Graphique 2: Autre visualisation des valeurs manquantes



Le jeu de donnée ne contient pas de valeur manquante.

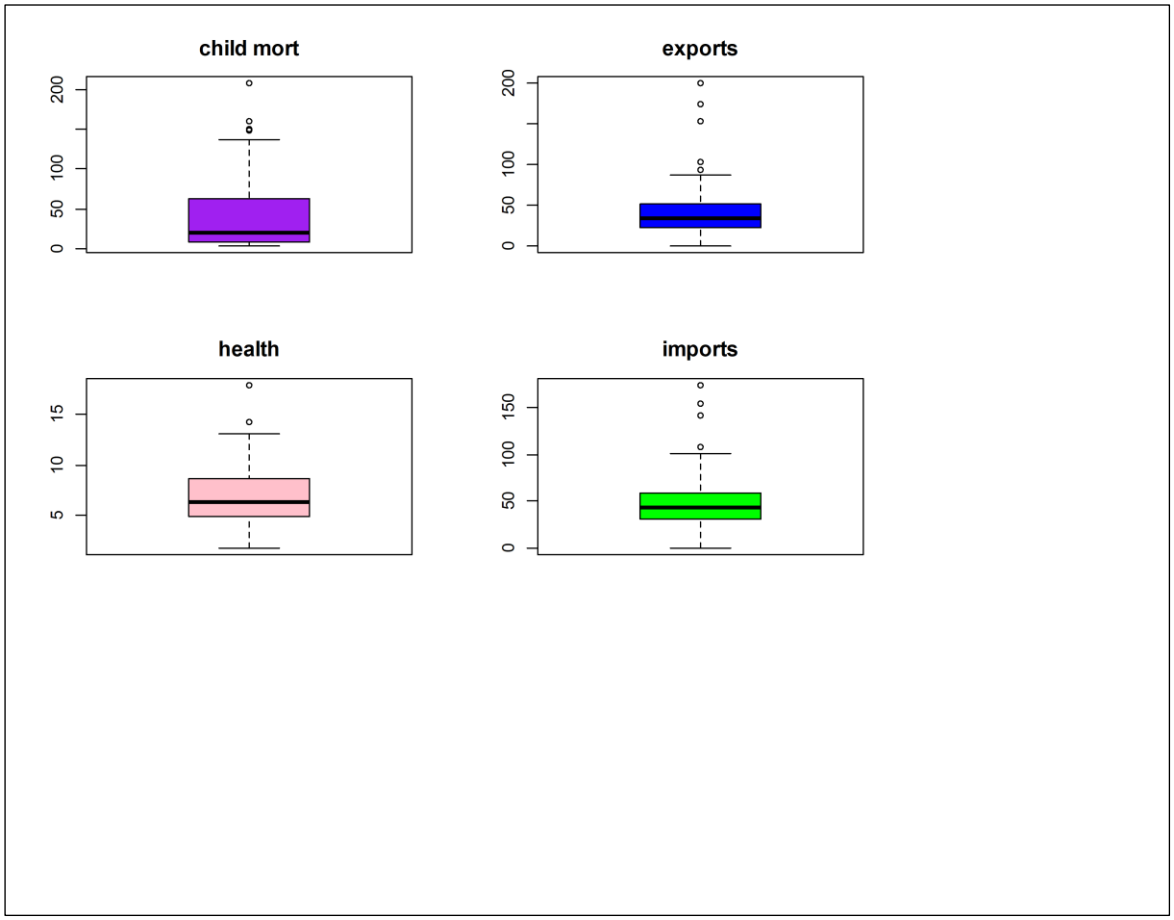
III-TRAITEMENT DES VALEURS ABERRRANTES ET EXTREMES

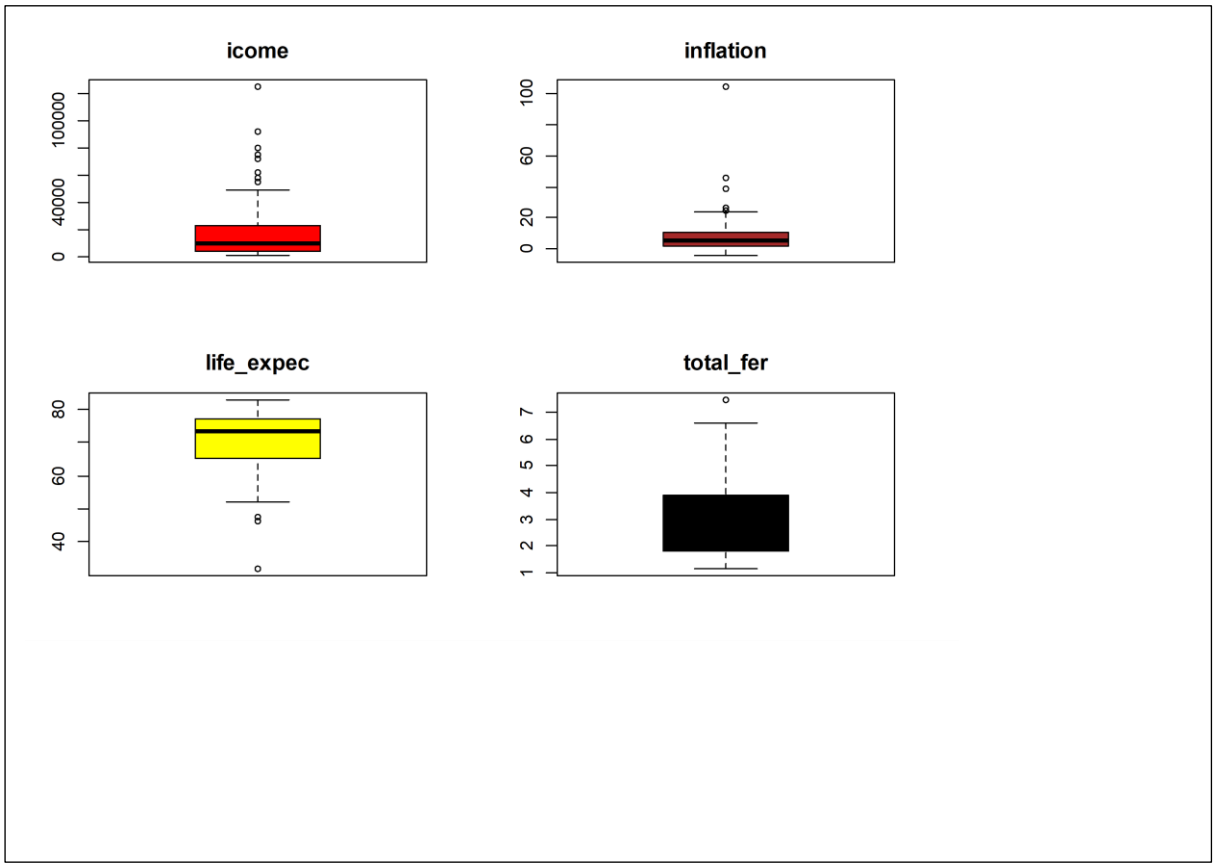
Les valeurs aberrantes (ou outliers) Ce sont des observations qui sont très différentes de la majorité des autres observations dans un ensemble de données. Elles peuvent résulter d'erreurs de mesure, de saisie de données incorrecte, ou représenter des événements rares ou inhabituels.

Les valeurs aberrantes peuvent fausser les analyses statistiques et doivent souvent être identifiées et traitées de manière appropriée pour éviter des conclusions erronées. Les valeurs extrêmes, Ces valeurs se trouvent à l'extrémité de la distribution des données et représentent les valeurs les plus élevées ou les plus basses dans un ensemble de données.

Elles peuvent être importantes pour comprendre la variabilité des données et les tendances générales, mais elles peuvent également être sources de biais si elles ne sont pas correctement gérées dans l'analyse. Elles se font uniquement avec des variables quantitatives. La visualisation de ces valeurs se fait à l'aide des boîtes à moustache. Une valeur est considérée comme aberrante ou extrême lorsqu'elle est située au-delà des moustaches.

Graphique 3: Visualisation des valeurs aberrantes et extrêmes


child mort
0
50
100
200
exports
0
50
100
200
health
5
10
15
imports
0
50
100
150

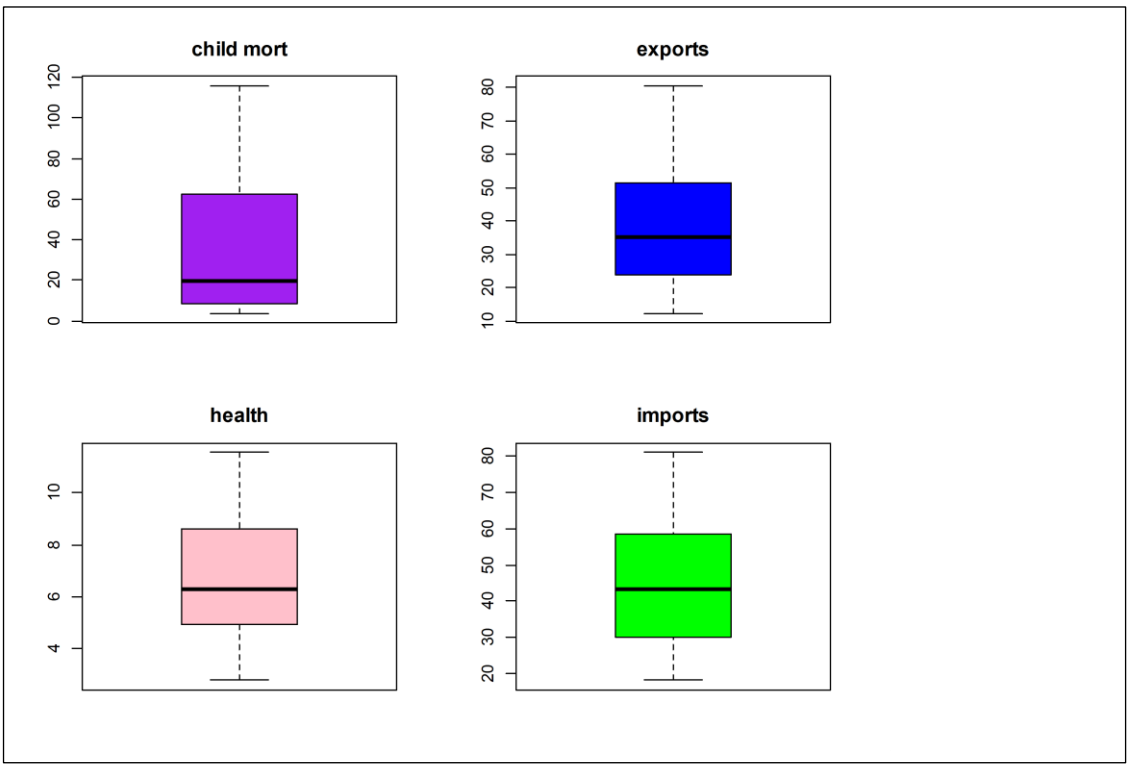

icome
0
40000
100000
inflation
0
20
60
100
life_expec
40
60
80
total_fer
1
2
3
4
5
6
7

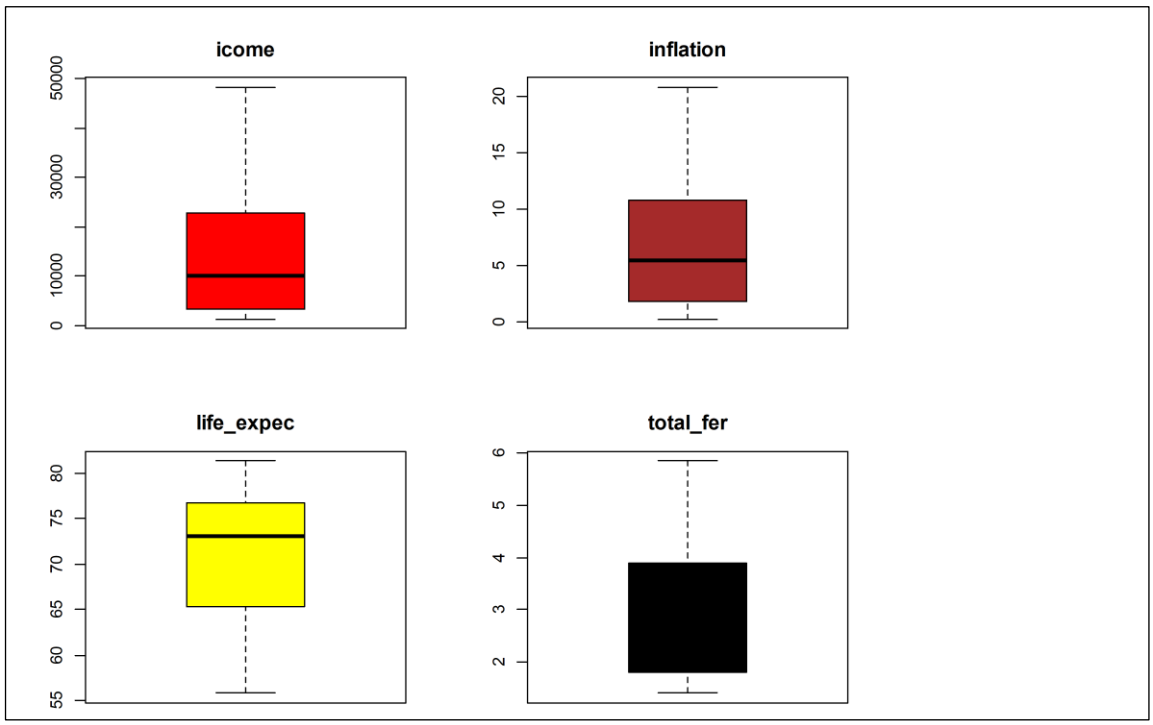
On voit bien sur le graphique 3 que les variables child-mort, exports, health, imports, icome, inflation, life-expec, total-fer et gdpp contiennent des valeurs aberrantes et extrêmes.

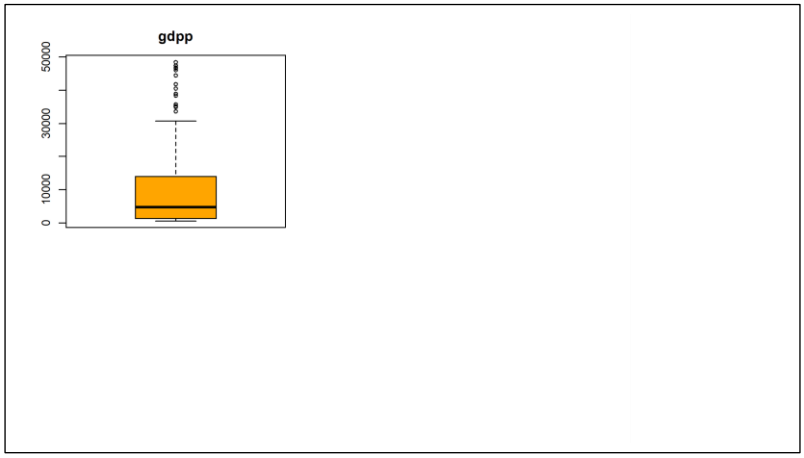
Pour traiter les valeurs aberrantes observées sur le graphique, nous allons utiliser la winsorisation. C'est une technique de traitement des valeurs aberrantes dans un jeu de données qui consiste à ramener ces valeurs dans la limite des bornes (inférieure et supérieure). Cette méthode doit son nom à son concepteur, Charles P. Winsor, un statisticien américain.

Après avoir traité les valeurs aberrantes, la visualisation nous donne ci-dessous (graphique 4).

Graphique 4: Visualisation des valeurs aberrantes et extrêmes après traitement







DEUXIEME PARTIE: ANALYSE EN COMPOSASANTE PRINCIPALE

L'Analyse en Composantes Principales (ACP) est une technique statistique puissante utilisée pour réduire la dimensionnalité des données tout en préservant au mieux leur variance. Elle permet de transformer un ensemble de variables corrélées en un ensemble de variables non

corrélées, appelées composantes principales, qui capturent l'essentiel de l'information contenue dans les données d'origine.

L'ACP est souvent utilisée pour explorer la structure sous-jacente des données, identifier les tendances et les relations entre les variables, et faciliter la visualisation des données dans un espace de dimensions réduites. Elle est particulièrement utile dans les cas où les données sont complexes et comportent de nombreuses variables, ce qui rend difficile l'interprétation et l'analyse directe.

Le processus de l'ACP implique plusieurs étapes, notamment le calcul des vecteurs propres et des valeurs propres de la matrice de covariance ou de corrélation des données, la sélection des composantes principales en fonction de leur contribution à la variance totale, et enfin la projection des données originales dans l'espace des composantes principales.

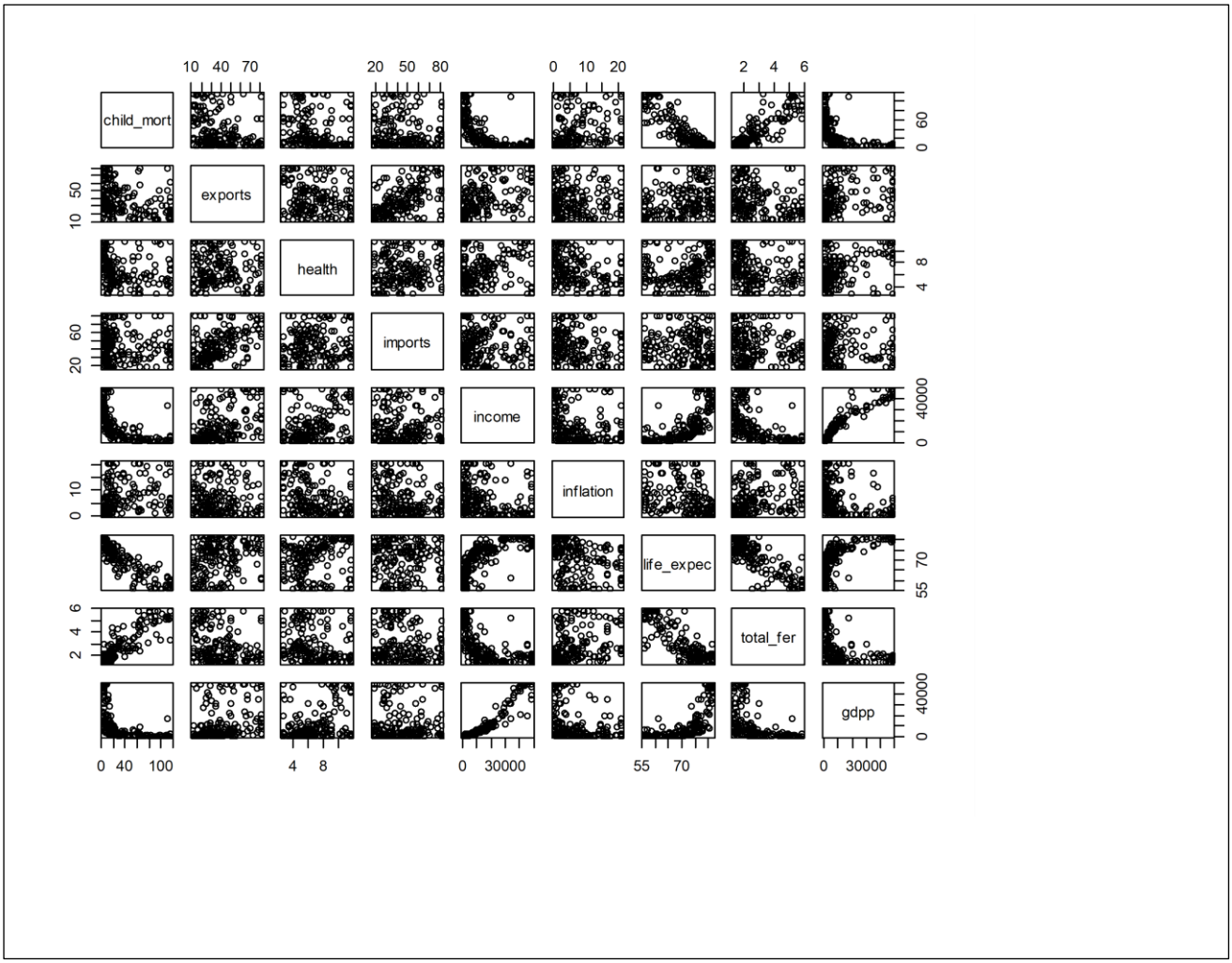
I-CORRELATION ENTRE LES VARIABLES

La corrélation entre les variables est une mesure statistique qui évalue la relation linéaire entre deux variables. Elle indique dans quelle mesure les variations d'une variable sont associées aux variations de l'autre. Une corrélation positive signifie que les deux variables évoluent dans le même sens, tandis qu'une corrélation négative indique une relation inverse.

La corrélation est souvent mesurée par le coefficient de corrélation de Pearson, qui varie de -1 à +1. Un coefficient proche de +1 indique une corrélation positive parfaite, tandis qu'un coefficient proche de -1 indique une corrélation négative parfaite. Un coefficient proche de 0 indique une faible corrélation ou l'absence de corrélation.

L'analyse de la corrélation entre les variables est importante pour plusieurs raisons. Elle permet de comprendre les relations entre les différentes variables d'un ensemble de données, ce qui peut aider à identifier les facteurs influençant un phénomène donné. De plus, elle permet de détecter des associations potentielles qui peuvent être explorées plus en détail dans des analyses ultérieures.

Graphique 5: Visualisation des variables



Ces données représentent une matrice de corrélation entre différentes variables socio-économiques et sanitaires pour un ensemble de pays. Chaque cellule de la matrice indique le degré de corrélation entre deux variables, avec des valeurs allant de -1 à 1.

Tableau 5: Visualisation des variables

	child_mort	exports	health	imports	income	inflation	life_expec
child_mort	1.0000000	-0.35023412	-0.2367269	-0.12450936	-0.63759365	0.32508231	-0.89020097
exports	-0.3502341	1.00000000	-0.1281939	0.57781652	0.46918521	-0.06568619	0.31402644
health	-0.2367269	-0.12819391	1.0000000	0.13649667	0.25628092	-0.36205685	0.25093646
imports	-0.1245094	0.57781652	0.1364967	1.00000000	-0.01282769	-0.29107800	0.01964844
income	-0.6375936	0.46918521	0.2562809	-0.01282769	1.00000000	-0.23485327	0.72925262
inflation	0.3250823	-0.06568619	-0.3620569	-0.29107800	-0.23485327	1.00000000	-0.33655119
life_expec	-0.8902010	0.31402644	0.2509365	0.01964844	0.72925262	-0.33655119	1.00000000
total_fer	0.8927821	-0.33737134	-0.2168567	-0.13766307	-0.59065401	0.36832455	-0.82106459
gdpp	-0.5579115	0.33675818	0.3871823	-0.03469369	0.94151387	-0.33290216	0.68182897

	total_fer	gdpp
child_mort	0.8927821	-0.55791152
exports	-0.3373713	0.33675818
health	-0.2168567	0.38718230
imports	-0.1376631	-0.03469369
income	-0.5906540	0.94151387
inflation	0.3683245	-0.33290216
life_expec	-0.8210646	0.68182897
total_fer	1.0000000	-0.50836159
gdpp	-0.5083616	1.00000000

Mortalité infantile (child_mort) et espérance de vie (life_expec) :

- Corrélation négative très forte (-0.89).
- Cela signifie qu'une augmentation de la mortalité infantile est fortement associée à une diminution de l'espérance de vie. Les pays avec une mortalité infantile élevée ont tendance à avoir une espérance de vie plus courte.

Revenu par habitant (income) et espérance de vie (life_expec) :

- Corrélation positive forte (0.73).
- Cela indique qu'une augmentation du revenu par habitant est associée à une augmentation de l'espérance de vie. Les pays avec un revenu par habitant plus élevé ont tendance à avoir une espérance de vie plus longue.

Revenu par habitant (income) et PIB par habitant (gdpp) :

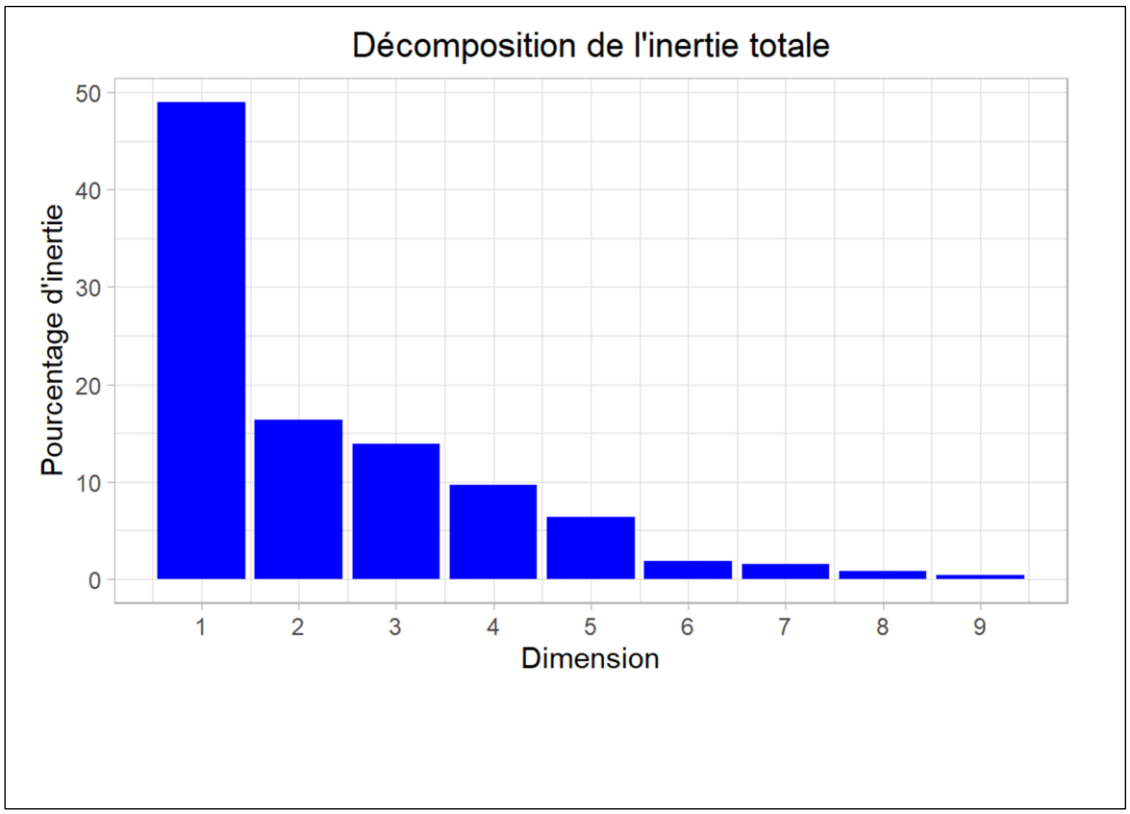
- Corrélation positive très forte (0.94).
- Cela signifie qu'il existe une forte relation entre le revenu par habitant et le PIB par habitant. Les pays avec un revenu par habitant plus élevé ont tendance à avoir un PIB par habitant plus élevé, et vice versa.

II-DISTRIBUTION DE L'INERTIE

Une distribution de l'inertie, également connue sous le nom de "scree plot" en anglais, est un graphique qui montre la quantité d'inertie expliquée par chaque composante principale dans une analyse en composantes principales (ACP) ou une autre technique de réduction de dimensionnalité.

L'inertie, dans ce contexte, fait référence à la quantité de variance expliquée par chaque composante principale. Plus une composante principale explique de variance, plus elle est importante pour décrire la structure des données.

Figure 1: Visualisation de la distribution l'inertie



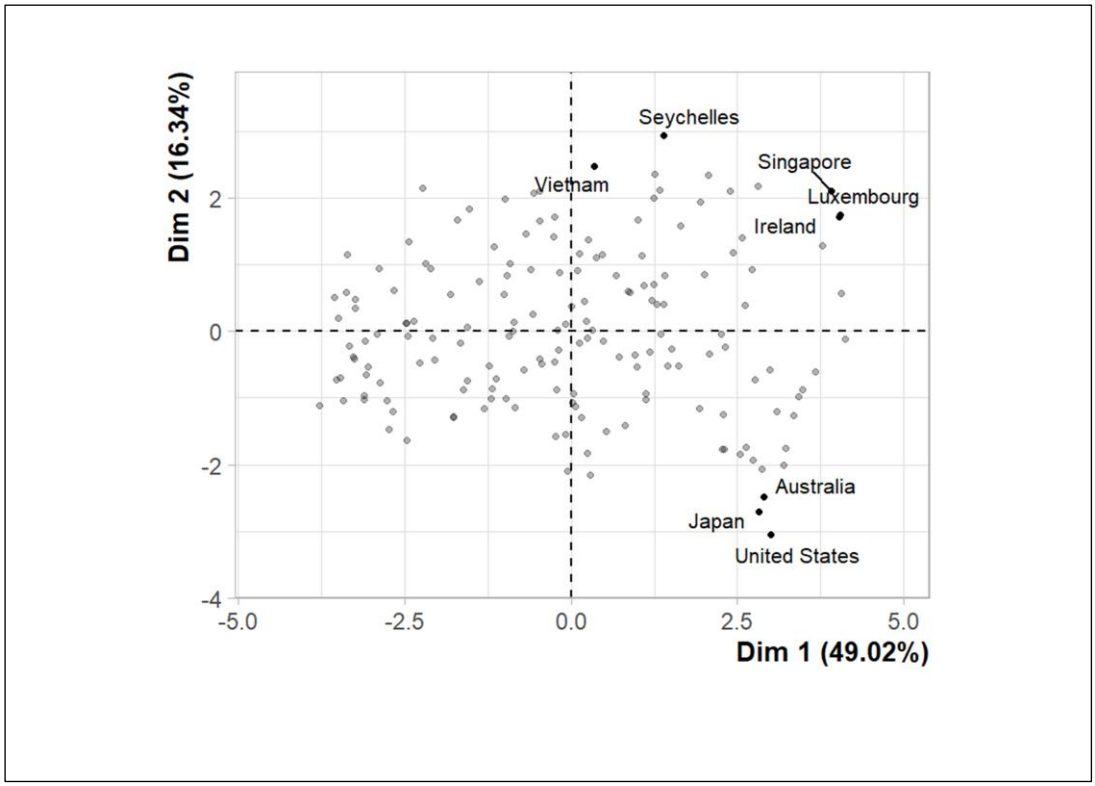
Les deux premiers axes de l'analyse en composantes principales capturent une proportion significative de la variabilité présente dans le jeu de données. Avec 65.36% de l'inertie totale représentée dans ce plan, cela indique que la structure sous-jacente des données est relativement bien capturée par ces deux axes principaux.

Cette proportion élevée de variabilité expliquée par les deux premiers axes suggère que les variables incluses dans l'analyse ont des relations cohérentes et significatives qui peuvent être résumées efficacement dans un espace de dimensions réduit. La valeur de 30.72% mentionnée comme référence représente un seuil significatif pour évaluer l'importance de l'inertie expliquée.

1-Graphe des individus

Un graphe des individus est un outil puissant pour visualiser et interpréter la structure des données multidimensionnelles, en permettant une compréhension intuitive des relations entre les observations et les variables d'un jeu de données.

Figure 2: Graphe des individus



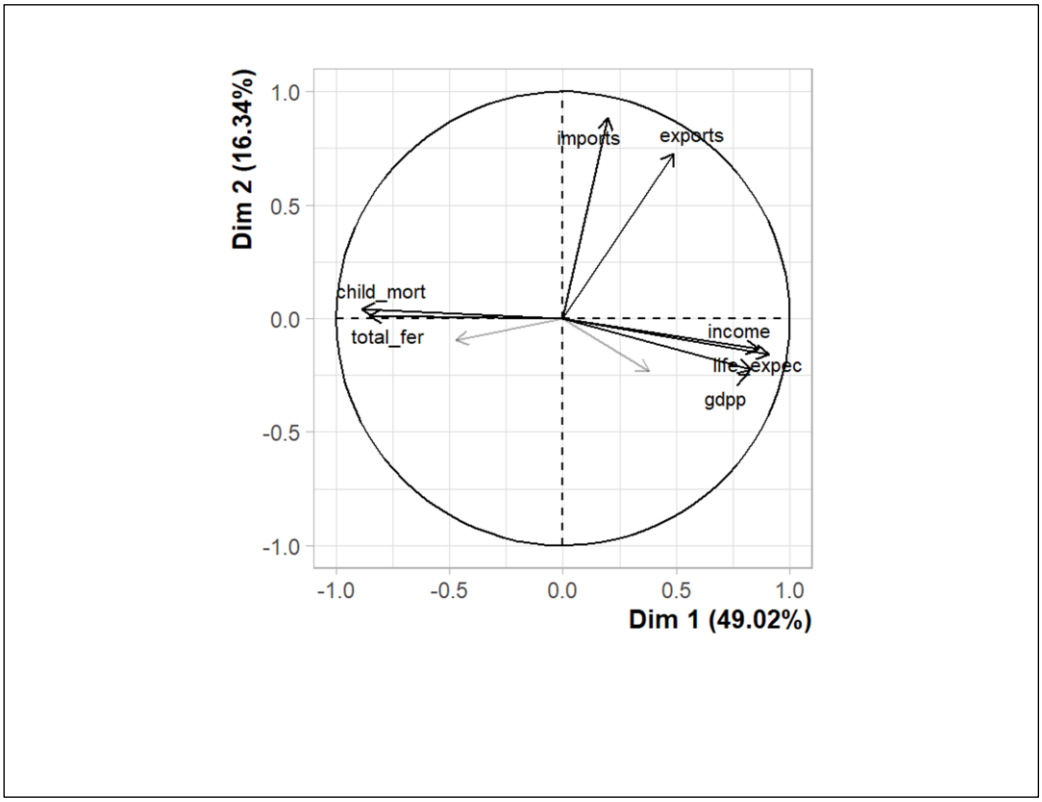
Les pays tels que Singapore, Luxembourg, Seychelles, Vietnam et Ireland semblent être des économies plus développées et ouvertes sur le plan commercial, avec des conditions de vie et de santé généralement meilleures.

Les pays tels que Australie, Japon et United States semblent être confrontés à des défis économiques et sociaux plus importants, avec des niveaux de vie et de santé plus faibles.

2-Graphe des variables

Un graphe des variables est un outil utile pour explorer et interpréter la structure des données multidimensionnelles, en mettant en évidence les relations entre les variables et en identifiant les variables les plus importantes dans la description des composantes principales.

Figure 3: Graphe des variables



Le groupe caractérisé par des coordonnées positives sur la dimension 1 semble être composé de pays qui ont des performances économiques relativement élevées, avec de fortes valeurs pour les variables d'exportations, d'importations, de revenu, de PIB par habitant et d'espérance de vie. Ces pays ont également des niveaux relativement bas de mortalité infantile, de fécondité totale et d'inflation.

En revanche, le groupe caractérisé par des coordonnées négatives sur la dimension 1 semble être constitué de pays avec des performances économiques plus faibles, présentant des taux plus élevés de mortalité infantile, une fécondité totale plus élevée et une inflation plus élevée, ainsi que des valeurs plus basses pour le revenu, le PIB par habitant et l'espérance de vie.

La dimension 2 semble quant à elle mettre en évidence des différences dans les niveaux de développement économique et social. Les pays situés en haut de l'axe semblent avoir de meilleures performances en termes d'importations, d'exportations et d'espérance de vie, tandis que ceux situés en bas de l'axe ont des niveaux plus élevés de revenu, de PIB par habitant, de santé et d'éducation.

CONCLUSION

À la lumière des observations effectuées sur les groupes de pays dans l'analyse, il est recommandé que le PDG de HELP International se concentre principalement sur les pays appartenant aux groupes caractérisés par des coordonnées positives sur les axes de l'analyse en composantes principales (ACP). Ces pays présentent des caractéristiques socio-économiques et sanitaires plus favorables, ce qui suggère qu'ils pourraient bénéficier d'avantage des interventions et de l'assistance de l'organisation. Plus spécifiquement, les pays regroupés avec Seychelles, Luxembourg, Ireland, Singapore et Vietnam présentent des performances économiques relativement élevées, avec des niveaux élevés d'exportations et d'importations, ainsi qu'une espérance de vie élevée. Cependant, ces pays pourraient bénéficier d'une assistance supplémentaire pour réduire la fécondité totale et la mortalité infantile, ainsi que pour stabiliser les taux d'inflation. Les pays associés au japon, Australie et United States présentent également des indicateurs économiques et sanitaires positifs, tels que des niveaux élevés de revenu, de PIB par habitant, d'espérance de vie et de santé. Cependant, ces pays pourraient nécessiter une attention particulière pour maintenir ces niveaux élevés tout en gérant efficacement les défis liés à la fécondité totale et à la mortalité infantile. En concentrant les efforts de l'organisation sur ces pays, le PDG de HELP International peut maximiser l'impact de ses interventions en fournissant une assistance là où elle est le plus nécessaire et en contribuant à améliorer durablement les conditions de vie et de santé dans ces régions.

SOURCE DU CODE R

```
# IMPORTATION DES DONNEES

Country.data <- read.csv("C:/Users/hp/Downloads/Country-data.csv",
stringsAsFactors=TRUE,row.names=1)


# EXPLORATION DES DONNES

head(Country.data, 5)

tail(Country.data, 5)

str(Country.data)

summary(Country.data)


# VISUALISATION DES DONNEES MANQUANTES

library(visdat)

vis_dat(Country.data)

vis_miss(Country.data)


#PRETRAITEMENT DES DONNEES

# Traitement des doublons

Country.data_1 = unique(Country.data)

nrow(Country.data) - nrow(Country.data_1)


# identifier le nombre d'individus ayant des donnees manquantes

Country.data[!complete.cases(Country.data),]

nrow(Country.data[!complete.cases(Country.data),])
```

```
# Traitement des valeurs aberrantes et extrêmes
par(mfrow=c(2,2), mar=c(3,3,3,3))
boxplot(Country.data$child_mort, main = "child mort", col ="purple")
boxplot(Country.data$exports, main = "exports", col = "blue")
boxplot(Country.data$health, main = "health", col = "pink")
boxplot(Country.data$imports, main = "imports", col = "green")
boxplot(Country.data$income, main = "icome", col ="red")
boxplot(Country.data$inflation, main = "inflation", col = "brown")
boxplot(Country.data$life_expec, main = "life_expec", col = "yellow")
boxplot(Country.data$total_fer, main = "total_fer", col = "black")
boxplot(Country.data$gdpp, main = "gdpp", col = "orange")
par(mfrow=c(2,2), mar=c(3,3,3,3))

library(DescTools)
Country.data$child_mort <- Winsorize(Country.data$child_mort)
Country.data$exports <- Winsorize(Country.data$exports)
Country.data$health <- Winsorize(Country.data$health)
Country.data$imports<- Winsorize(Country.data$imports)
Country.data$income <- Winsorize(Country.data$income)
Country.data$inflation <- Winsorize(Country.data$inflation)
Country.data$life_expec<- Winsorize(Country.data$life_expec)
Country.data$total_fer <- Winsorize(Country.data$total_fer)
Country.data$gdpp <- Winsorize(Country.data$gdpp)
boxplot(Country.data$child_mort, main = "child mort", col ="purple")
boxplot(Country.data$exports, main = "exports", col = "blue")
boxplot(Country.data$health, main = "health", col = "pink")
boxplot(Country.data$imports, main = "imports", col = "green")
```

```
boxplot(Country.data$income, main = "icome", col ="red")
boxplot(Country.data$inflation, main = "inflation", col = "brown")
boxplot(Country.data$life_expec, main = "life_expec", col = "yellow")
boxplot(Country.data$total_fer, main = "total_fer", col = "black")
boxplot(Country.data$gdpp, main = "gdpp", col = "orange")
#VISUALISATION DES CORRELATIONS
pairs(Country.data[1:9])
cor(Country.data[1:9])
#OUTILS POUR ANALYSE DE DONNEES INTERACTIFS AVEC UN INTERFACE
library(Factoshiny)
res<-PCAshiny(Country.data)
```