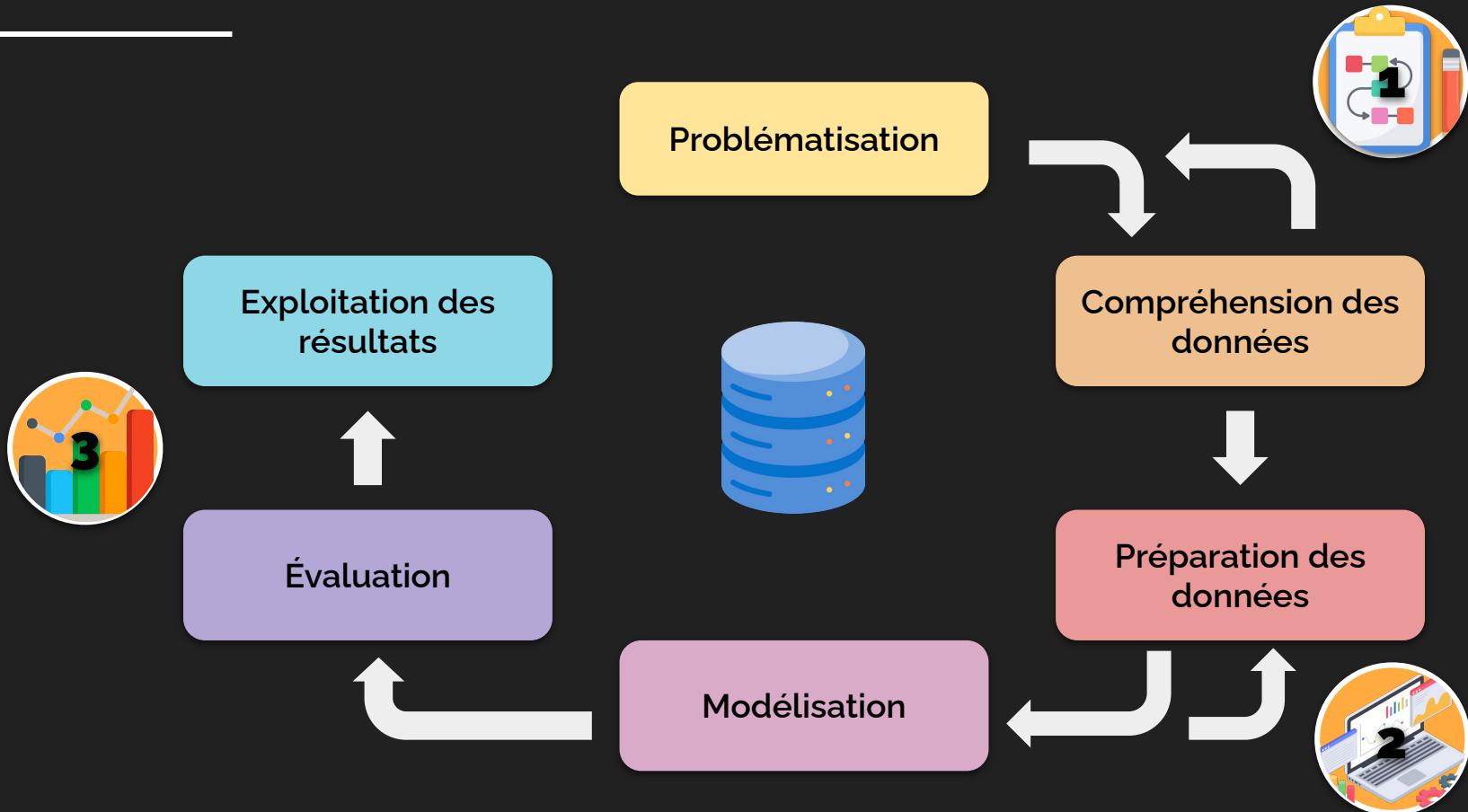


Life Expectancy

Data Mining

UE H: Épidémiologie statistique

Méthodologie : CRISP

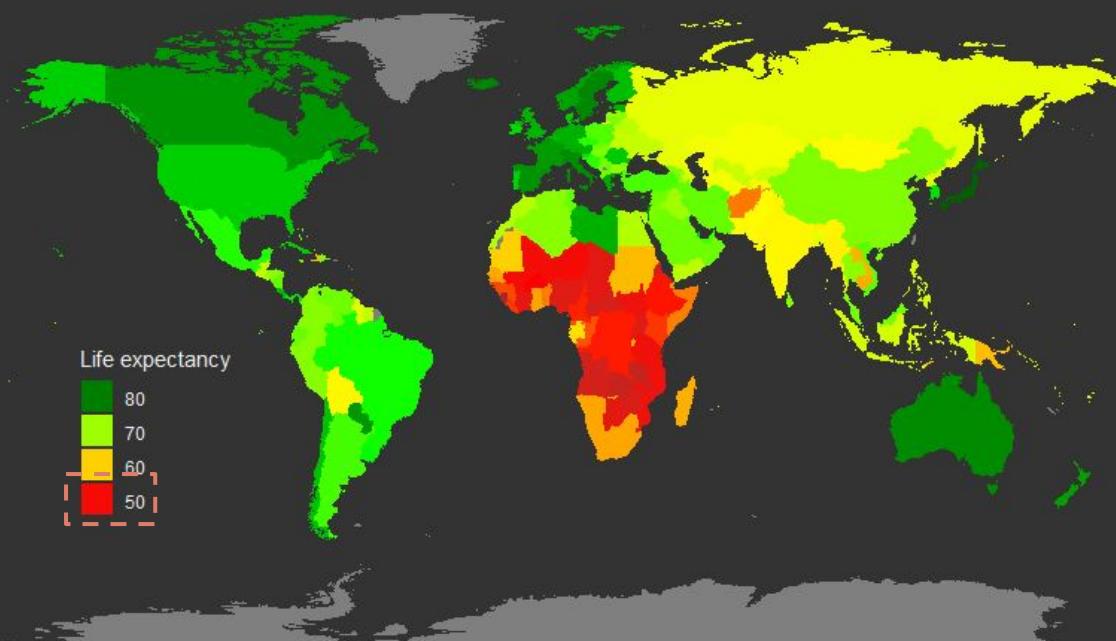


I. Exploration des données & Problématisation



Life expectancy by Country

2000



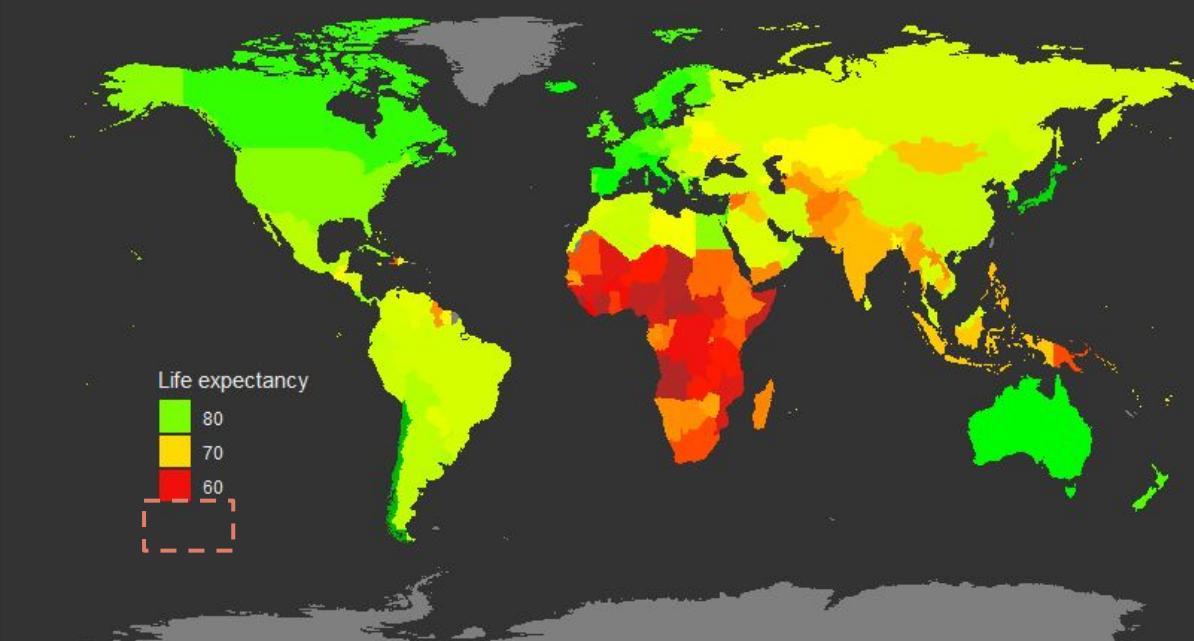
- **Source :** World Health Organization, [Kaggle](#)
- **Données :**
 - **Variable d'intérêt :** Life Expectancy
 - **Co-variables (features):**
 - ✓ Country
 - ✓ Year (2000 → 2015)
- Cartographie

I. Exploration des données & Problématisation



Life expectancy by Country

2015



- **Source :** World Health Organization, [Kaggle](#)
- **Données :**
 - **Variable d'intérêt :** Life Expectancy
 - **Co-variables (features):**
 - ✓ Country
 - ✓ Year (2000 → 2015)

→ Cartographie

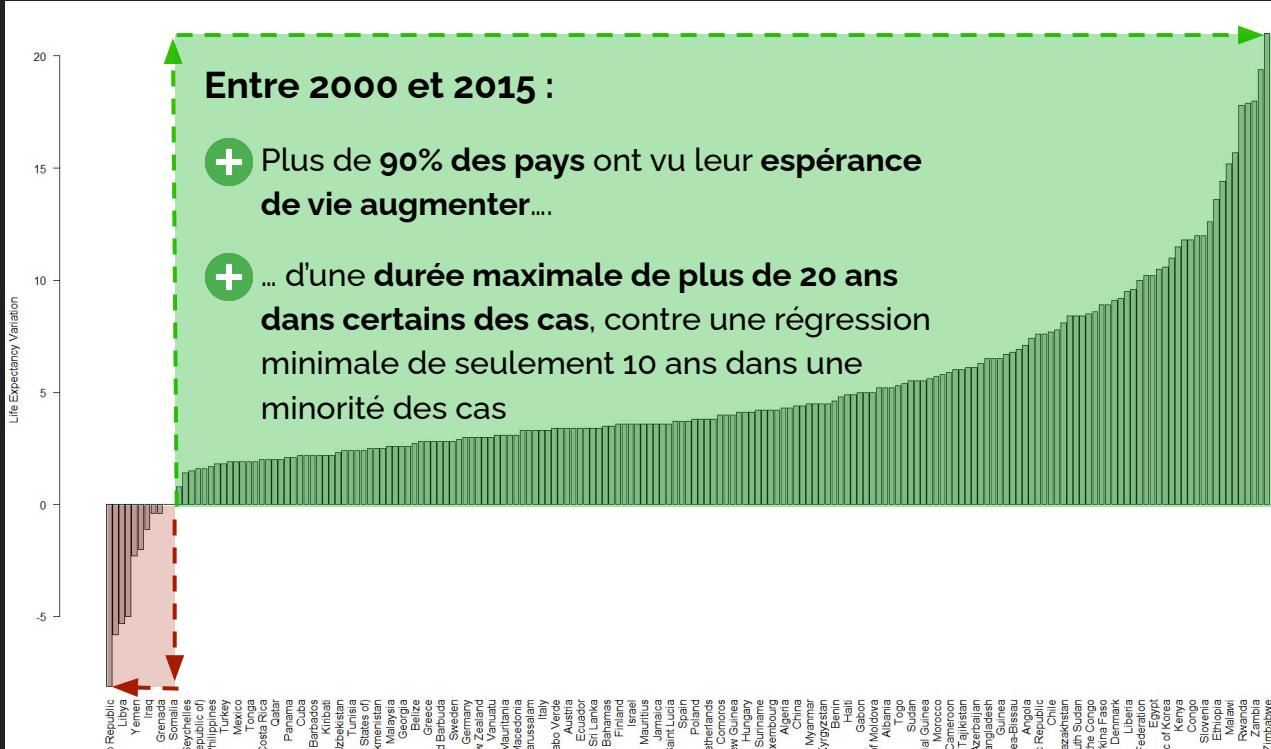


De 2000 à 2015,
amélioration visible,
notamment en Afrique

I. Exploration des données & Problématisation



Life Expectancy delta between 2000 and 2015, for all countries



- **Source :** World Health Organization, [Kaggle](#)

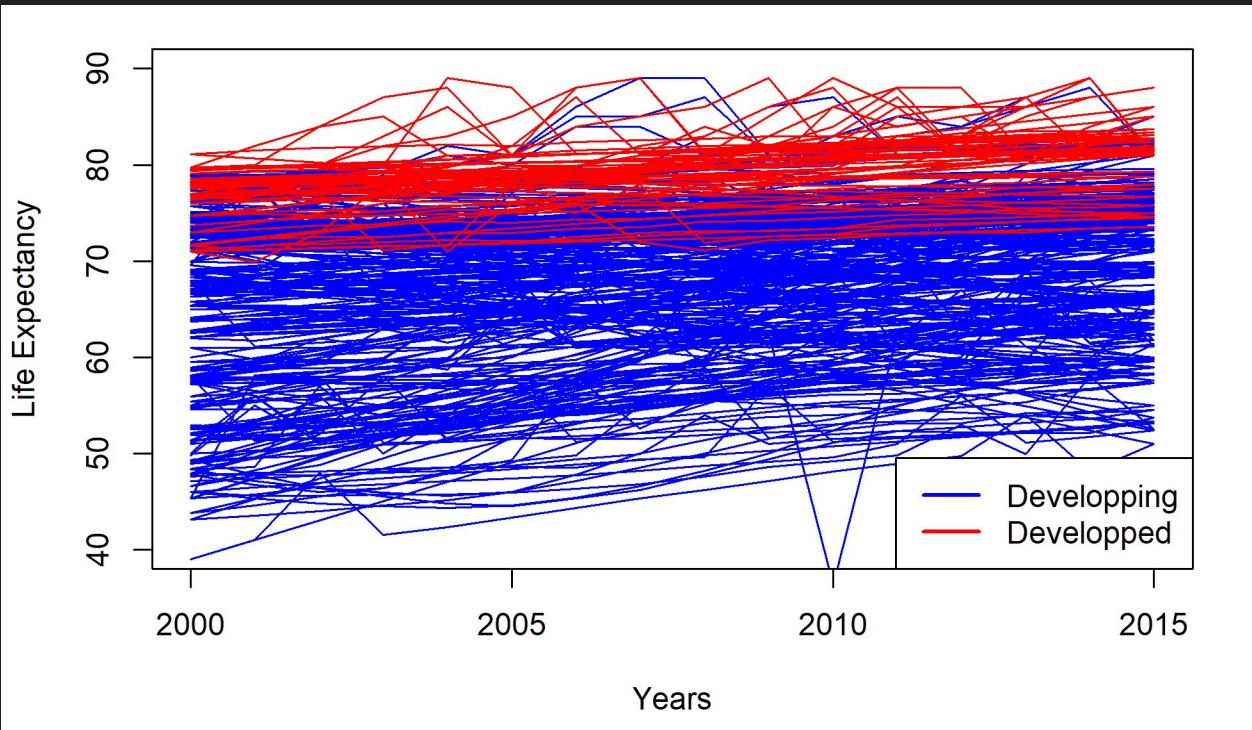
- **Données :**
 - 📌 **Variable d'intérêt :** Life Expectancy

- 📌 **Co-variables (features):**
 - ✓ Country
 - ✓ Year (2000 → 2015)

I. Exploration des données & Problématisation



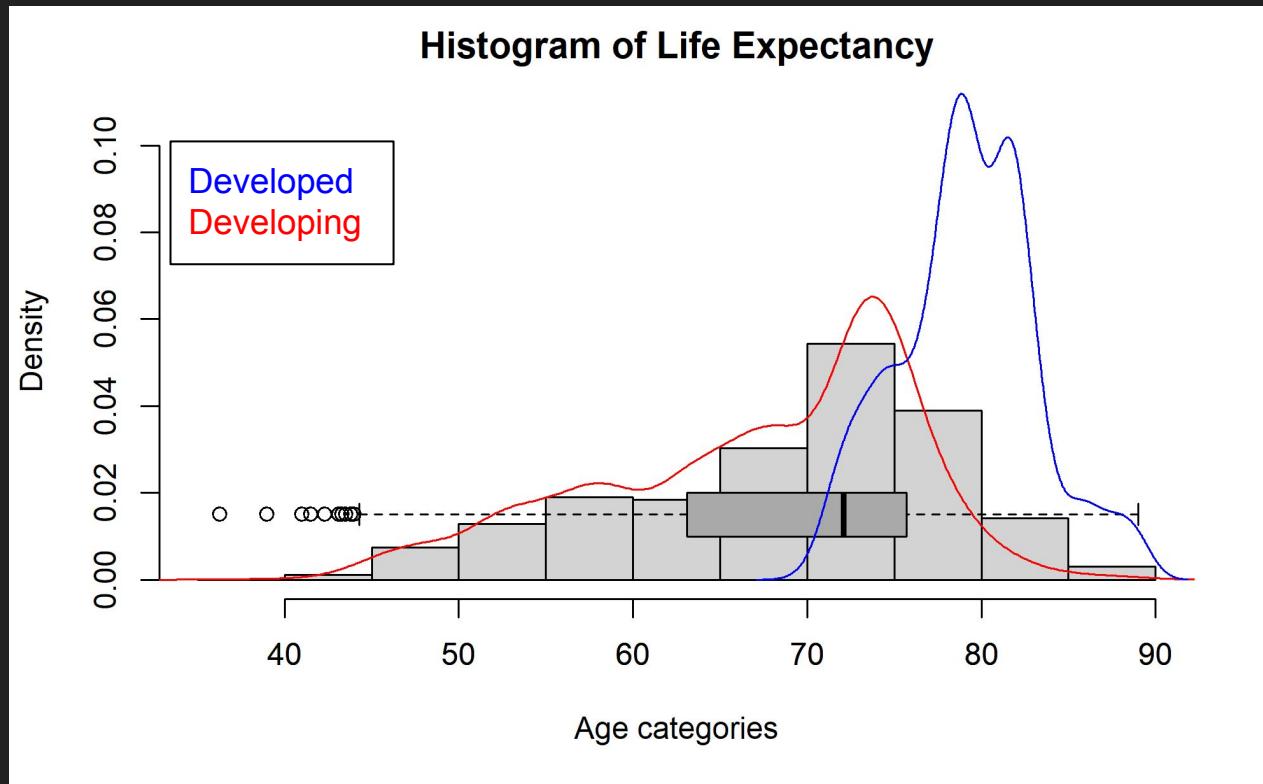
Life Expectancy evolution over 2000 - 2015, for all countries



Co-variables (*features*):

- ✓ Country
 - ✓ Year (2000 → 2015)
 - ✓ Status :
 - Developed
 - Developing
- ↓
- Minorité de pays "développés"
 - L'espérance de vie ne semble **pas être un facteur discriminant** pour les différencier
 - L'âge maximal semble **continuer d'augmenter**

I. Exploration des données & Problématisation



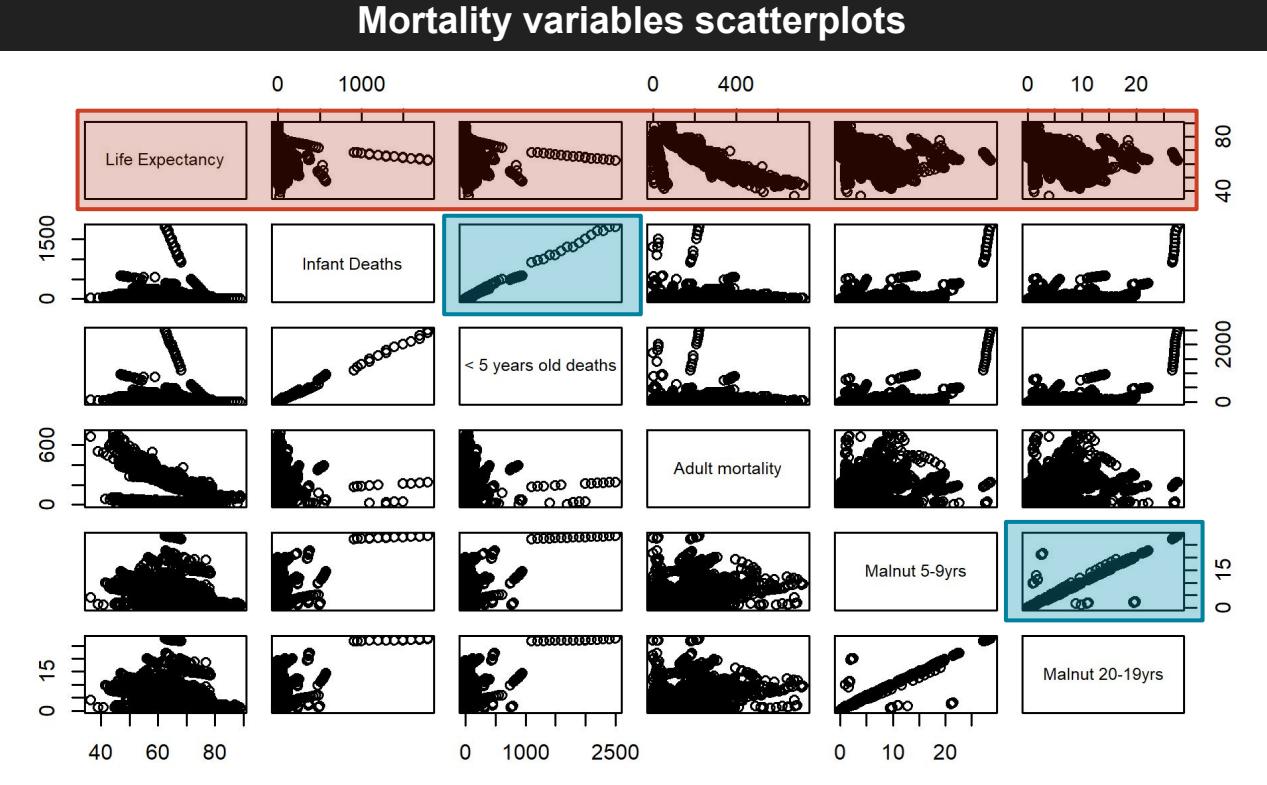
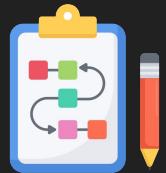
Co-variables (*features*):

- ✓ Country
- ✓ Year (2000 → 2015)
- ✓ Status :
 - Developed
 - Developing

Table 1

Life Expectancy	Mean	Median [min, max]
Developing	67.1	69 [36.3, 89]
Developed	79.2	79.3 [69, 89]
Overall	69.2	72.1

I. Exploration des données & Problématisation



Co-variables (*features*):

- ✓ Country
- ✓ Year (2000 → 2015)
- ✓ Status

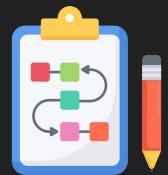
x 5 ✓ Facteurs de mortalité :

- Mortalité infantile
- Mortalité adulte
- Décès < 5 ans
- Malnutrition 5-9 ans
- Malnutrition 10-19 ans

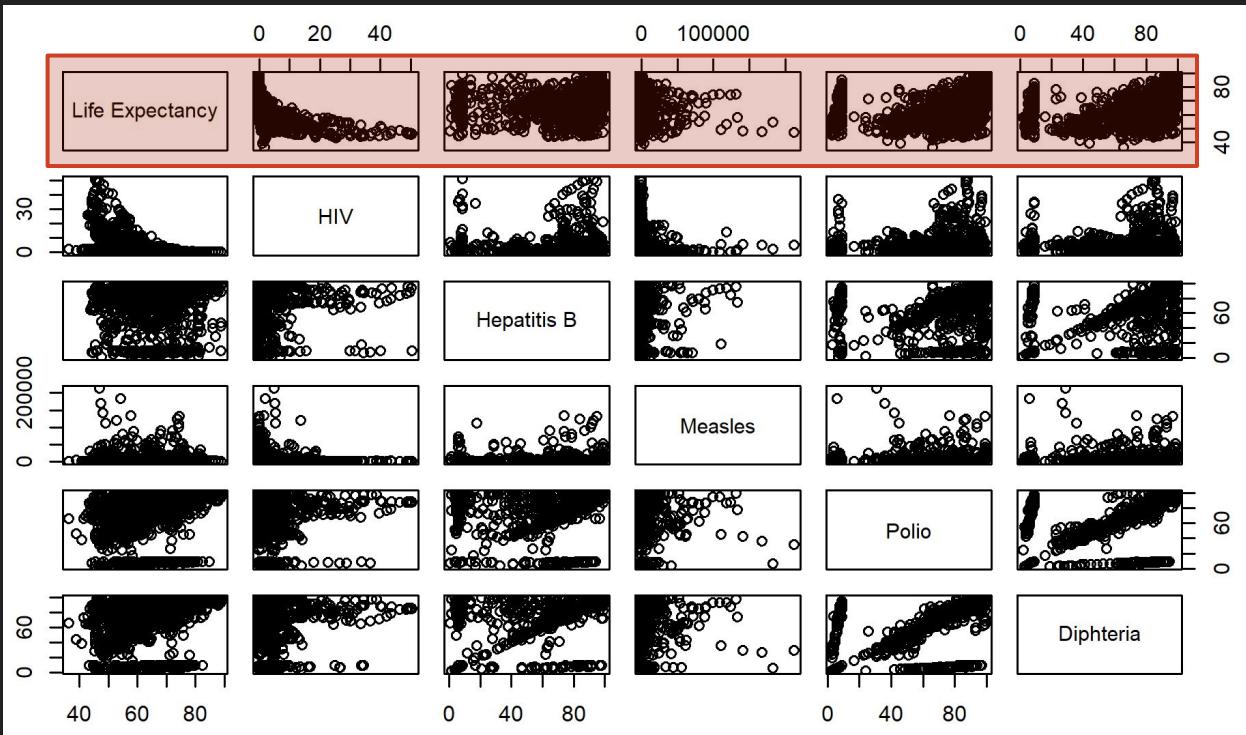


Forte corrélation négative entre la mortalité à l'âge adulte et l'espérance de vie

I. Exploration des données & Problématisation



Health variables scatterplots



Co-variables (*features*):

- ✓ Country
 - ✓ Year (2000 → 2015)
 - ✓ Status
 - x 5 ✓ Facteurs de mortalité
 - x 5 ✓ **Facteurs sanitaires**

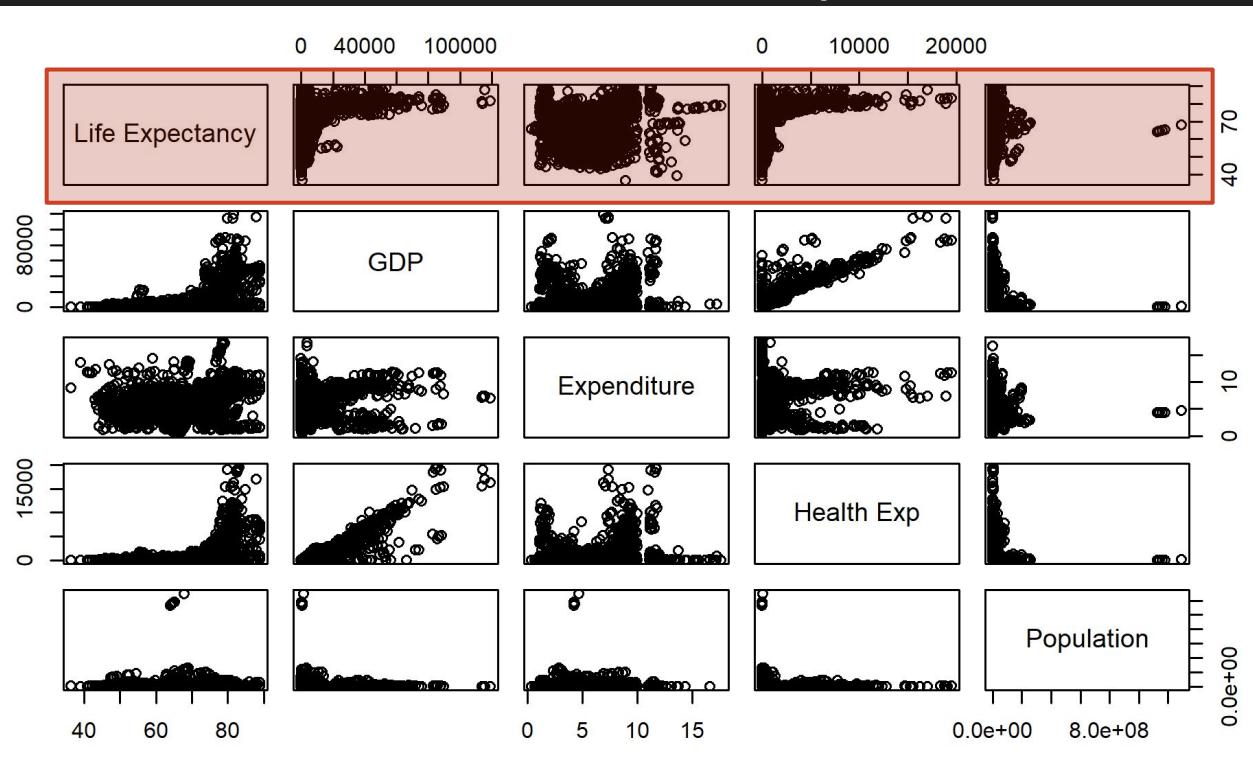
- HIV
 - Hépatite B
 - Measles
 - Poliomyélia
 - Diphtéria

Hépatite B
Measles
Polio ~ Diphtheria
HIV

I. Exploration des données & Problématisation



Economic variables scatterplots



Co-variables (features):

- ✓ Country
- ✓ Year (2000 → 2015)
- ✓ Status
- x 5 ✓ Facteurs de mortalité
- x 5 ✓ Facteurs sanitaires
- x 4 ✓ **Facteurs économiques :**
 - GDP
 - Expenditure
 - Health % expenditure
 - Population

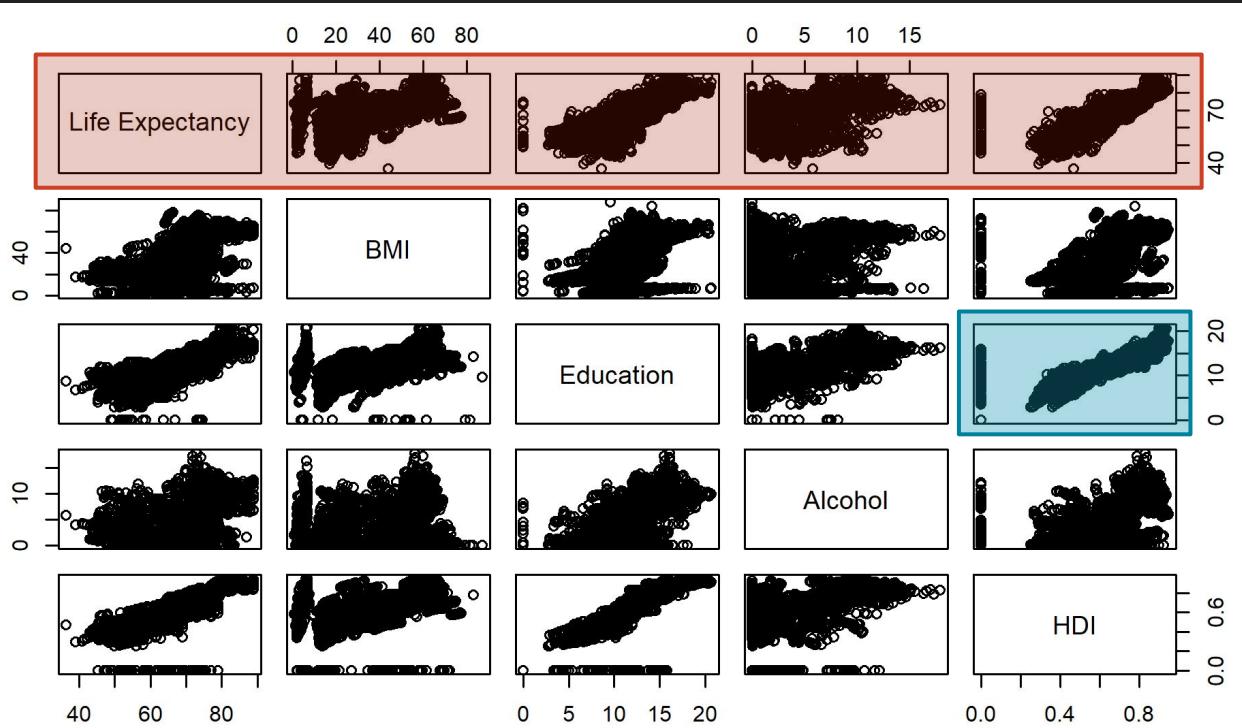


High GDP and Health
Exp seem to be an
indicators of ↗ LE

I. Exploration des données & Problématisation



Social variables scatterplots



Co-variables (*features*):

- ✓ Country
- ✓ Year (2000 → 2015)
- ✓ Status
- x 5 ✓ Facteurs de mortalité
- x 5 ✓ Facteurs sanitaires
- x 4 ✓ Facteurs économiques
- x 4 ✓ **Facteurs sociaux :**
 - HDI
 - Education
 - BMI
 - Alcohol

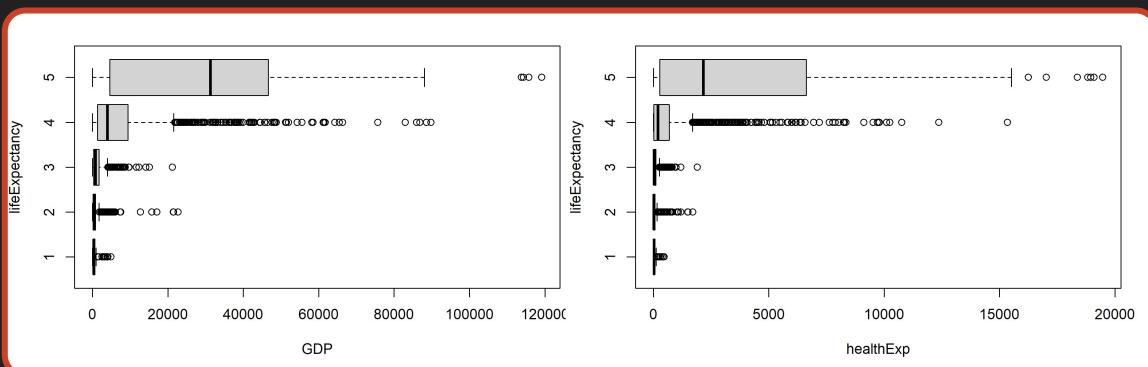
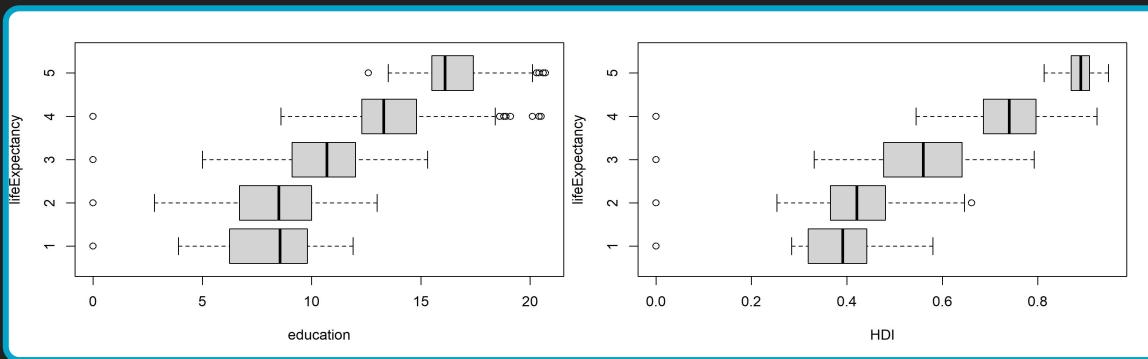


Their seem to be a clear positive correlation between LE and BMI, Education, and HDI

I. Exploration des données & Problématisation



Social variables Vs. Economic variables



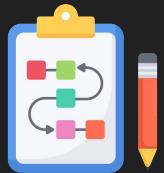
Co-variables (*features*):

- ✓ Country
- ✓ Year (2000 → 2015)
- ✓ Status
- ✗ 5 ✓ Facteurs de mortalité
- ✗ 5 ✓ Facteurs sanitaires
- ✗ 4 ✓ Facteurs économiques
- ✗ 4 ✓ Facteurs sociaux

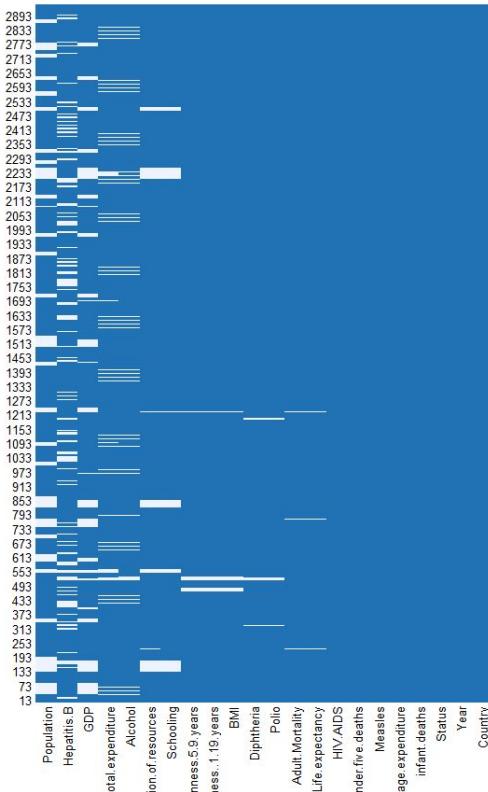
“Effet de seuil” : Marqueurs de la catégorie d’âge la plus importante.

Évolution plus progressive : Permet de caractériser chaque catégorie de manière quasi unique

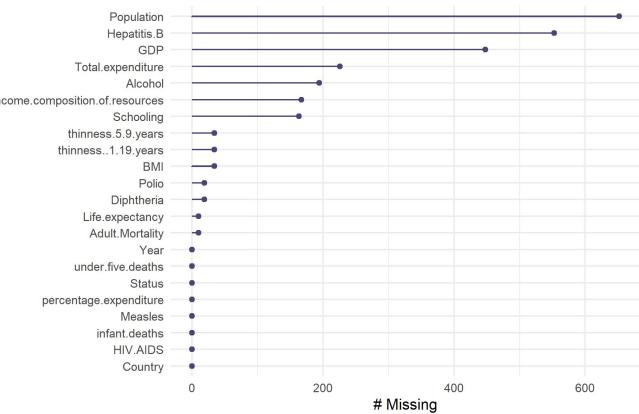
I. Exploration des données & Problématisation



Missingness Map



Variables



Data size :

Missing values
⚠ 43% de lignes incomplètes

Co-variables (*features*):

- ✓ Country
- ✓ Year (2000 → 2015)
- ✓ Status
- x 5 ✓ Facteurs de mortalité
- x 5 ✓ Facteurs sanitaires
- x 4 ✓ Facteurs économiques
- x 4 ✓ Facteurs sociaux

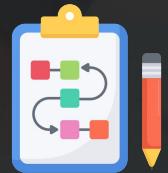
21 features

- 20 continuous
- 1 discrete

2938 lines

Une large base de données

I. Exploration des données & Problématisation



Sur **quels paramètres jouer pour conseiller un pays** qui chercherait à augmenter son espérance de vie ?

Ces variables sont-elles **identiques selon que le pays est considéré comme développé ou non** ?

Quelles variables influent le plus sur l'espérance de vie d'un pays à un autre ?

Est-ce que la catégorisation "**développés** "/"**en développement**" pour caractériser les pays est **encore pertinente d'après ces indicateurs** ?

L'espérance de vie va-t-elle **continuer à augmenter** en moyenne dans les années à venir ?

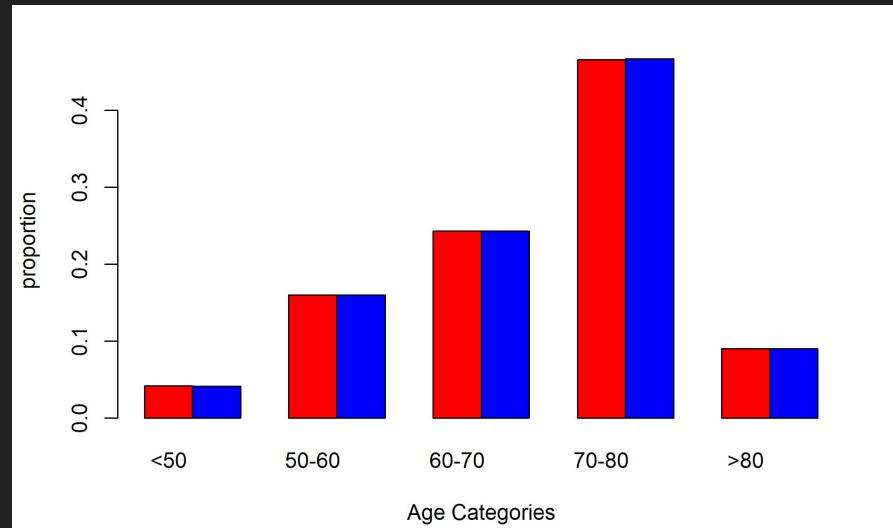
II. Préparation des données et modélisation



- 📌 Classification des pays en fonction de leur espérance de vie

- ✓ **Modèle N° 1** : Générique (tous les pays)
- ✓ **Data** :
 - Prise en compte de **toutes les co-variables**, dont l'année et le statut.
 - Nécessité de **discrétiser la variable d'espérance de vie**.
 - 70 % **train** / 30% **test**

- ✓ **Méthode** :
Random Forest

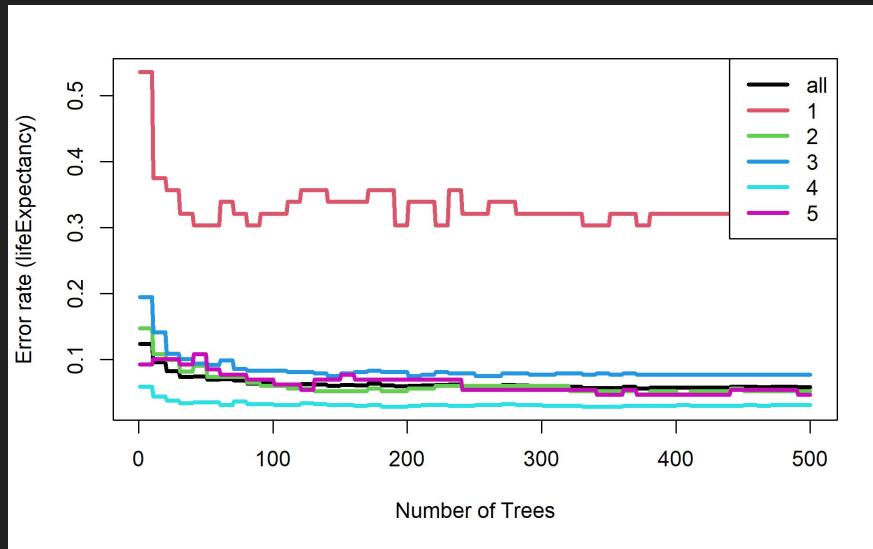


III. Evaluation et exploitation des résultats

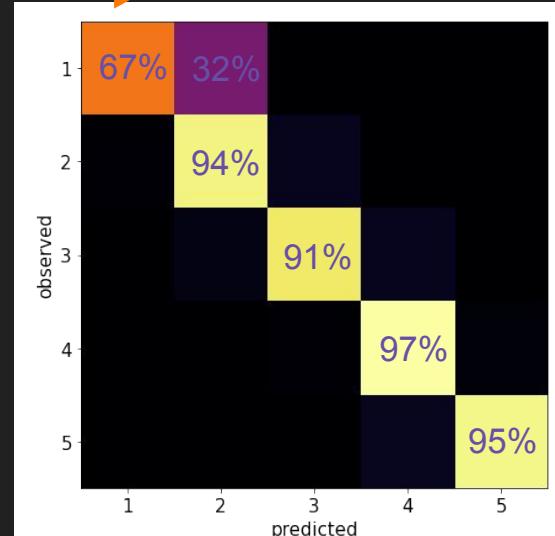


📌 Classification des pays en fonction de leur espérance de vie

✓ **Modèle N° 1** : Générique (tous les pays)



Unbalanced classes issue

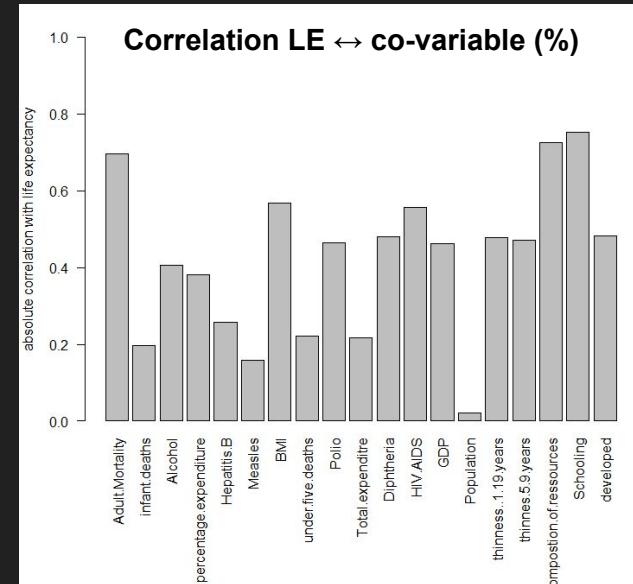
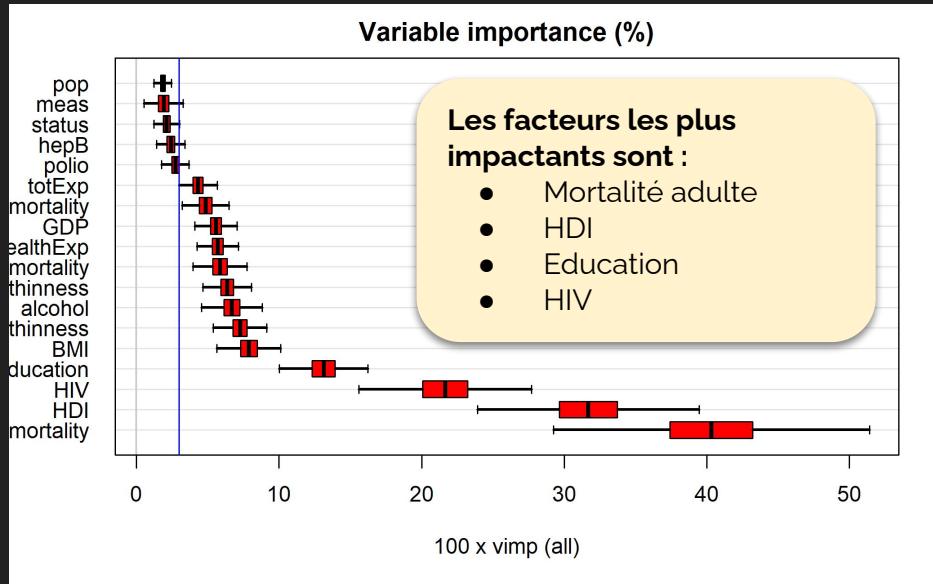


III. Evaluation et exploitation des résultats



Classification des pays en fonction de leur espérance de vie

✓ Modèle N° 1 : Générique : tous les pays

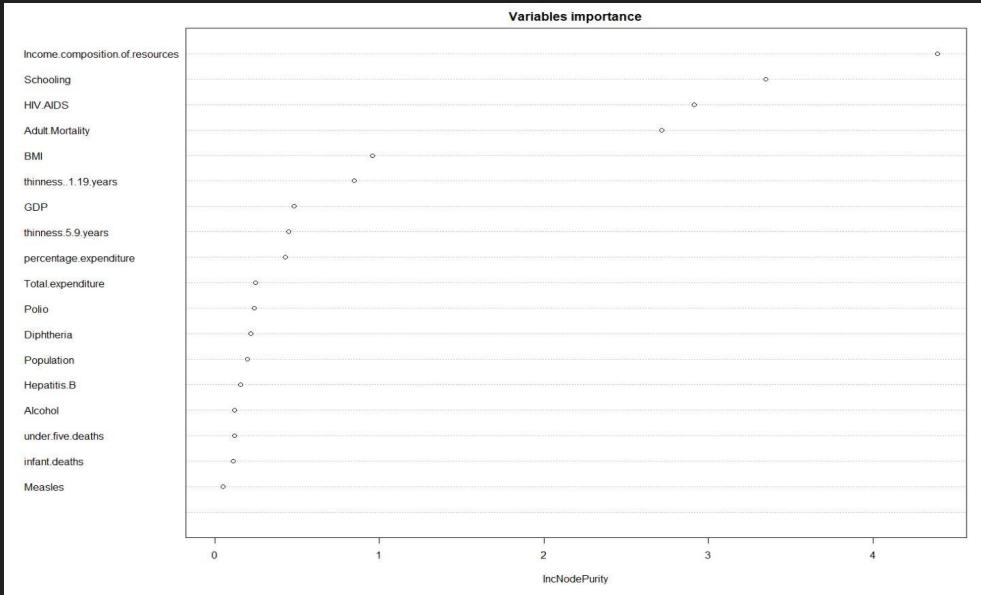


III. Evaluation et exploitation des résultats



📌 Classification des pays en fonction de leur espérance de vie

✓ **Modèle N° 2** : Spécifique : les pays en voie de développement



Les facteurs les plus impactants sont :

- HDI
- HIV
- Education
- Mortalité adulte

III. Evaluation et exploitation des résultats



Classification des pays en fonction de leur espérance de vie

✓ Modèle N° 2 : Validation, le cas de Zimbabwe



↑20

Les facteurs les plus impactants sont :

- HDI
- HIV
- Education
- Mortalité adulte

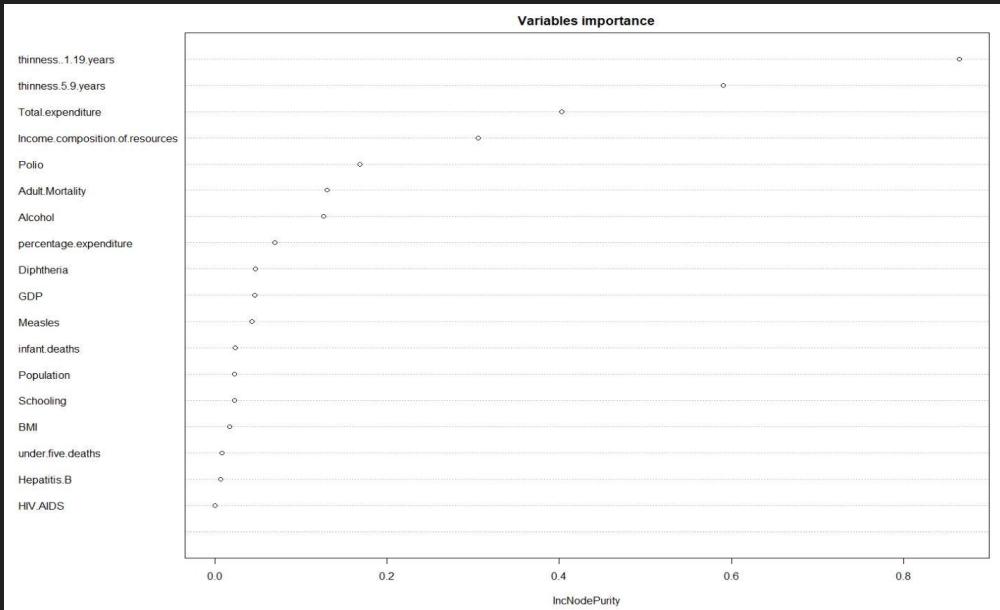


III. Evaluation et exploitation des résultats



📌 Classification des pays en fonction de leur espérance de vie

✓ **Modèle N° 3** : Spécifique : les pays développés



Les facteurs les plus impactants sont :

- Maigreurs entre 1 et 19 ans
- Maigreurs entre 5 et 9 ans
- Dépenses totales
- HDI

IV. Conclusion

Sur **quels paramètres jouer pour conseiller un pays** qui chercherait à augmenter son espérance de vie ?

Ces variables sont-elles **identiques selon que le pays est considéré comme développé ou non** ?

Quelles variables influent le plus sur l'espérance de vie d'un pays à un autre ?

Est-ce que la catégorisation "**développés** "/"**en développement**" pour caractériser les pays est **encore pertinente d'après ces indicateurs** ?

L'espérance de vie va-t-elle **continuer à augmenter** en moyenne dans les années à venir ?

IV. Conclusion

Sur **quels paramètres jouer pour conseiller un pays** qui chercherait à augmenter son espérance de vie ?

Ces variables sont-elles **identiques selon que le pays est considéré comme développé ou non** ?

D'après le **modèle générique** : le taux de **mortalité à l'âge adulte**, l'**IDH**, l'**éducation**, et le **HIV**,

Est-ce que la catégorisation "**développés** "/"**en développement**" pour caractériser les pays est **encore pertinente d'après ces indicateurs** ?

L'espérance de vie va-t-elle **continuer à augmenter** en moyenne dans les années à venir ?

IV. Conclusion

Sur **quels paramètres jouer pour conseiller un pays** qui chercherait à augmenter son espérance de vie ?

Ces **variables sont différentes** si le pays est considéré comme **développé ou non**

D'après le **modèle générique** : le taux de **mortalité à l'âge adulte**, l'**IDH**, l'**éducation**, et le **HIV**,

Est-ce que la catégorisation "**développés** "/"**en développement**" pour caractériser les pays est **encore pertinente d'après ces indicateurs** ?

L'espérance de vie va-t-elle **continuer à augmenter** en moyenne dans les années à venir ?

IV. Conclusion

Pour les pays développés:

- Maigreur entre 1 et 19 ans
- Maigreur entre 5 et 9 ans
- Dépenses totales
- HDI

Pour les pays en voie de développement:

- HDI
- HIV
- Education
- Mortalité adulte

Est-ce que la catégorisation
“développés” / “en développement” pour
caractériser les pays est **encore**
pertinente d'après ces indicateurs ?

Ces **variables sont différentes** si le pays
est considéré comme **développé ou non**

D'après le **modèle générique** : le taux de **mortalité**
à l'âge adulte , l'**IDH** , l'**éducation** , et le **HIV**,

L'espérance de vie va-t-elle **continuer à**
augmenter en moyenne dans les années à
venir ?

IV. Conclusion

Pour les pays développés:

- Maigreur entre 1 et 19 ans
- Maigreur entre 5 et 9 ans
- Dépenses totales
- HDI

Pour les pays en voie de développement:

- HDI
- HIV
- Education
- Mortalité adulte

Valable mais incomplet :

Des autres divisions sont nécessaires. Par exemple, le SIDA est un problème présent dans les pays africains en développement mais pas dans les pays d'Amérique du Sud.

Ces **variables sont différentes** si le pays est considéré comme **développé ou non**

D'après le **modèle générique** : le taux de **mortalité à l'âge adulte**, l'**IDH**, l'**éducation**, et le **HIV**,

L'espérance de vie va-t-elle **continuer à augmenter** en moyenne dans les années à venir ?

IV. Conclusion

Pour les pays développés:

- Maigreur entre 1 et 19 ans
- Maigreur entre 5 et 9 ans
- Dépenses totales
- HDI

Pour les pays en voie de développement:

- HDI
- HIV
- Education
- Mortalité adulte

Valable mais incomplet :

Des autres divisions sont nécessaires. Par exemple, le SIDA est un problème présent dans les pays africains en développement mais pas dans les pays d'Amérique du Sud.

Ces **variables sont différentes** si le pays est considéré comme **développé ou non**

D'après le **modèle générique** : le taux de **mortalité à l'âge adulte**, l'**IDH**, l'**éducation**, et le **HIV**,

OUI

Pour les pays en voie de développement:

L'exemple présenté a démontré que est possible d'augmenter si les variables plus influentes sont traitées.

Pour les pays développées:

Il devrait être aussi possible si leurs variables sont traités