

Master 1 : Économétrie appliquée

Analyse de données et descriptive

Les déterminants du bonheur : Une Analyse en composantes principales

Le Roux Noa, Ammour Gabriel

09 Novembre 2022

Résumé

Dans cette étude, nous avons essayé de déterminer les différents facteurs influençant le bonheur d'un pays. À l'aide de l'analyse en composantes principales (ACP), nous avons synthétisé dix variables explicatives en deux grandes variables latentes. Nous en avons conclu que le niveau de développement économique et social ainsi que le niveau de liberté individuelle et collective étaient les deux meilleures composantes pour expliquer le niveau de bonheur d'un pays. Dans un second temps, nous nous sommes aidés de méthodes économétriques - Régression Linéaire Multiple et Régression Logistique Binaire - afin d'évaluer la qualité et l'efficacité de notre modèle.

Abstract

This study aims to determine the deciding factors of the country's happiness level. Using principal component analysis (PCA), we tried to synthesize ten dependent variables into two large latent variables. We arrived at the result that the level of economic and social development as well as the level of individual and collective freedom were the two best components to explain the level of happiness of a country. In a second phase, we used econometric methods - Multiple Linear Regression and Binary Logistic Regression - to evaluate the quality and efficiency of our model.

Sommaire

Sommaire	i
1 Introduction	1
2 Analyse détaillé du sujet	3
3 Analyse en composante principale	22
4 Corrélation entre la variable illustrative et les variables latentes	51
5 Conclusion et discussions des résultats	64
6 Annexes	66
Bibliographie	89
Table des matières	89

Chapitre 1

Introduction

Partout dans le monde, la notion de bonheur traverse les pensées depuis plusieurs siècles. Souvent confondue avec la notion de plaisir, le bonheur se définit comme un état émotionnel agréable, mais surtout équilibré et durable (contrairement au plaisir qui se maintient sur une durée bien plus courte).

A peine sortie d'une crise sanitaire et en plein désastre climatique, nous pensons que la notion de bonheur devient de plus en plus importante. En effet, le contexte dans lequel nous vivons nous donne une formidable opportunité de déterminer les facteurs les plus importants pour nous et notre bonheur. Étudié dans de nombreux domaines tels que la philosophie, l'anthropologie ou encore la biologie, nous utiliserons ici l'économétrie dans le but de dégager des axes qui pourraient, de façon la plus universelle, expliquer le niveau de bonheur d'une population. Cependant, le bonheur est une notion particulièrement subjective, il est donc très compliqué de déterminer précisément quels en sont ses déterminants. Par exemple, la simple mesure du PIB ne suffira pas à déterminer le bonheur (ou bien-être) d'un pays. En effet, ce dernier ne prend pas en compte plusieurs facteurs comme la répartition inégale des revenus, la pollution ou encore la santé de sa population.


Dès lors, tout au long de notre de l'histoire, sont apparus des indicateurs essayant de mesurer le bonheur. Par exemple, le concept de Bonheur National Brute (BNB) apparaît au Bhoutan en 1972. Un indicateur s'étendant sur 9 grandes dimensions telles que le bien-être psychologique, la santé, l'éducation, diversité écologique ou encore la vitalité de la communauté¹. Contrairement aux indicateurs occidentaux, ce dernier permet de mesurer le bonheur d'un

1. <https://www.cairn.info/revue-projet-2018-1-page-26.htm>pa9[consulté le 07/10/2022]

pays en mettant l'accent sur des valeurs de collectivité et de durabilité. Il permet même de faire la distinction entre le bonheur spirituel et le bonheur matériel. Si le BNB ne sera pas utilisé dans notre étude, il reste pertinent de le mentionner puisqu'il nous montre que dans certains pays, le bonheur peut devenir un indicateur à part entière, aidant même les politiques dans leurs prises de décisions.

Comme nous l'avons dit précédemment, nous allons tenter de quantifier le bonheur à l'échelle mondiale. Pour ce faire, nous commencerons dans une première partie à nous pencher en détail sur les variables présentes dans notre modèle. Ensuite, nous traiterons dans une deuxième partie du processus utilisé pour aboutir à des variables latentes. Enfin, nous verrons dans une dernière partie les méthodes économétriques mise à l'oeuvre pour déterminer la pertinence de notre travail.

1.0.1 Méthodologie

Tous les graphiques, tableaux ou résultats qui seront présentés dans ce dossier ont été obtenu via le logiciel . Néanmoins, pour le bien de la présentation, certains ont manuellement été convertis au format \LaTeX . Pour toutes questions supplémentaires, n'hésitez pas à contacter les auteurs.

Chapitre 2

Analyse détaillé du sujet

2.0.1 Pertinence des variables à expliquer

Pour calculer le bonheur, cette variable si subjective et difficile à définir, nous utiliserons le *World Happiness Report*¹ sur l'année 2016. Nous avons choisi cette année puisque c'était l'année pré-covid avec le plus de données disponibles. Le *World Happiness Report* ou Rapport Mondial sur le Bonheur est publié par les Nations Unies depuis 2012. Cet indicateur utilise des données d'enquêtes mondiales pour rendre compte de la façon dont les gens évaluent leur propre vie dans plus de 150 pays à travers le monde. La source principale des rapports sur le bonheur est le *Gallup World Poll*².

La méthode pour calculer le bonheur est très simple et porte même un nom : l'échelle de Cantril³. On demande aux interrogés d'imaginer la « meilleure » et la « pire » qualité de vie qu'ils puissent avoir, puis ils positionnent leur qualité de vie actuelle sur cette même échelle allant de 1 à 10. Plus l'on se rapproche de 10 et plus le niveau de bonheur est élevé. L'échantillon typique est constitué de 1000 personnes par pays, avec des intervalles de confiance à 95%. Autrement dit, la probabilité que la moyenne de la population se situe dans cette intervalle est de 95%.

Le bonheur est un indicateur de plus en plus mesuré, selon Ban Ki-moon, huitième Secrétaire général de l'Organisation des Nations Unies : « *Le monde a besoin d'un nouveau paradigme économique qui reconnaît la parité entre les trois piliers du développement durable. Les éléments*

1. <https://www.gallup.com/178667/gallup-world-poll-work.aspx>[Consulté le 06/11/2022]

2. <https://www.gallup.com/178667/gallup-world-poll-work.aspx>[Consulté le 06/11/2022]

3. <https://www.cairn.info/revue-francaise-d-economie-2012-3-page-35.htm>[consulté le 07/10/2022]

*de bien-être social, économique et environnemental sont indissociables. Ensemble, ils définissent le bonheur brut mondial*⁴ ». Il affirme également que « *le bonheur et le bien-être étaient des aspirations universelles à prendre en compte dans les objectifs politiques* »⁵ ». On comprend donc que le bonheur, plus qu'une émotion subjective, devient un indicateur indispensable à la mise en place de politiques économiques et sociales pour de nombreux pays dans le monde. Comprendre et déterminer cet indicateur devient depuis quelques années un enjeu crucial.

2.0.2 Rôle des variables explicatives

Initialement, nous avons intégré 15 variables explicatives dans notre modèle. Nous traitons dans cette partie uniquement des variables qui composent notre modèle final. En effet, toutes les variables n'étaient pas pertinentes pour expliquer le niveau de bonheur d'un pays. Nous avons donc décidé de retirer celles que nous trouvions les moins importantes. Nous reviendrons plus en détail sur ce processus au Chapitre 3 (Voir §3.0.4) du dossier où nous justifierons plus précisément les choix que nous avons effectués.

Liberte de la presse

La liberté de la presse se définit comme « *la possibilité effective pour les journalistes, en tant qu'individus et en tant que collectifs, de sélectionner, produire et diffuser des informations dans l'intérêt général, indépendamment des interférences politiques, économiques, légales et sociales, et sans menace pour leur sécurité physique et mentale* »⁶ ». Lorsque l'on regarde le classement mondial de la liberté de la presse 2016⁷, on s'aperçoit que les dix premiers pays de ce classement figurent, pour la plupart, parmi les 10 premiers du *World Happiness Report*. On se doute alors que la liberté de la presse a un rôle clé à jouer dans le niveau de bonheur d'une population. Ces données sont calculées des reports de restrictions d'accès et entraves à la couverture de l'actualité par *Reporters Sans Frontières* (RSF). Dans les pays où la liberté de la presse est faible, les journalistes sont donc privés d'information et d'accès aux terrains leur permettant de recueillir l'information.

Ce score, allant de 0 à 100, se base sur deux dimensions :

- une analyse quantitative : un report du nombre de privation et mauvais traitements subis par les journaliste du pays concerné.

4. <https://news.un.org/fr/story/2013/03/263422>[consulté le 08/10/2022]

5. *ibid*

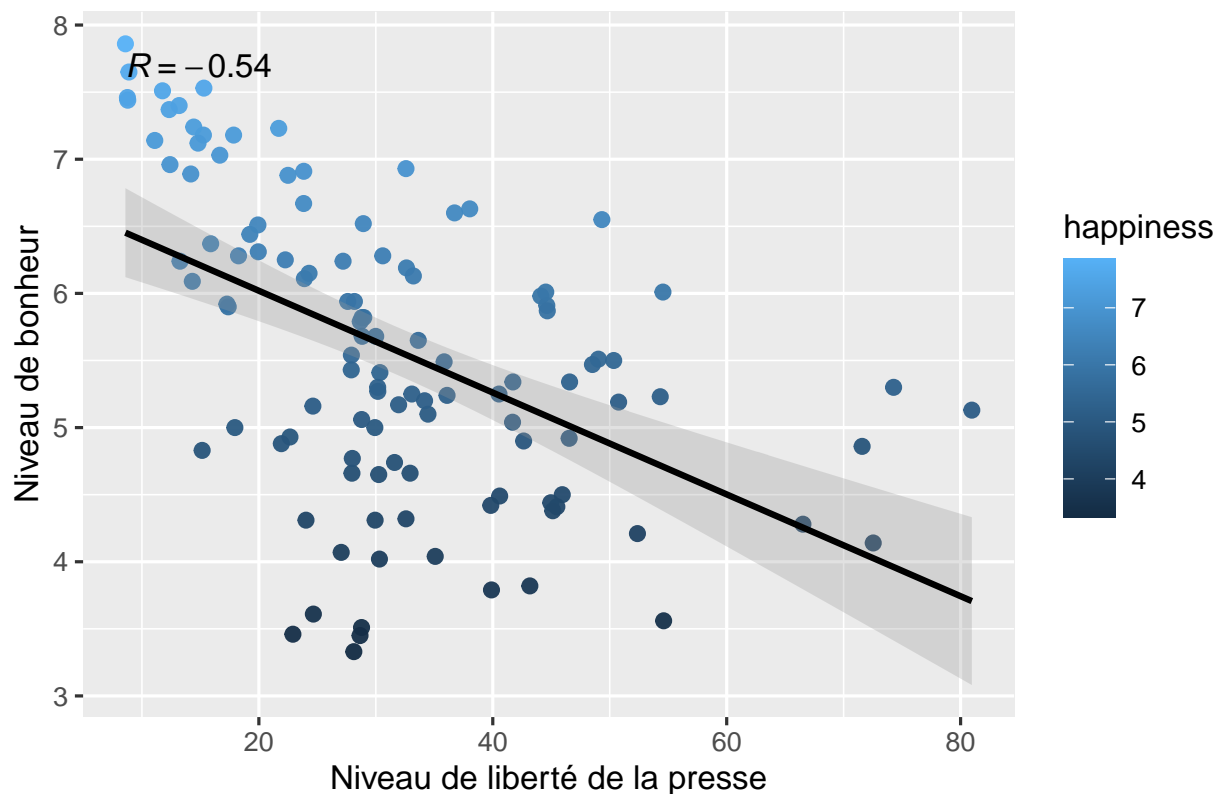
6. <https://rsf.org/fr/index-methodologie-2022>[consulté le 08/10/2022]

7. *ibid*

- une analyse qualitative : via un questionnaire adressé aux différents journalistes, chercheurs, universitaires, défenseurs des droits humains.

Attention, ici un score élevé (proche de 100) représente un pays dont la liberté de la presse n'est pas respectée, au contraire, un score faible (proche de 0) représente un pays dont la liberté de la presse est très respectée. Selon Anders Chydenius, un homme politique important du 18ème siècle, « *la démocratie, l'égalité et le respect des droits de l'homme sont les valeurs les plus importantes pour atteindre le progrès et le bonheur sur l'ensemble de la société* ⁸ ». Notre variable de liberté de la presse semble donc être pertinente pour déterminer le niveau de bonheur d'un pays.

FIGURE 2.1 – Corrélation entre la liberté de la presse et le niveau de bonheur



La Figure 2.1 nous permet d'observer une corrélation négative entre le niveau de bonheur et le niveau non-respect de la liberté de la presse. On peut voir le coefficient de corrélation en haut à gauche du graphique, il est de -0,54 en utilisant la méthode de Spearman. Ainsi, il

8. <https://finland.fi/fr/vie-amp-societe/la-finlande-tete-de-la-liberte-de-la-presse/> [consulté le 08/10/2022]

semblerait qu'au sein d'un même pays, plus le niveau de non-respect de la liberté de la presse est élevé, plus le niveau de bonheur est faible.

Néanmoins, on observe également que malgré une liberté de la presse relativement bonne, certains pays ont tout de même un niveau de bonheur relativement faible. Nous verrons quels sont les pays concernés plus loin dans notre analyse.

Liberte d'expression

La quantification de cette variable est basée sur les évaluations d'experts et la notation de l'indice de transformation Bertelsmann^{9 10}. Cet indicateur, mesuré tous les deux ans, permet d'attester la gestion politique d'un pays, et à quel point ce dernier prend des décisions politiques inscrites dans l'idéologie d'une démocratie. Concrètement, l'indice montre l'état de développement qu'un pays a atteint sur la voie de la démocratie et de l'économie de marché. Concernant la liberté d'expression, il indique dans quelle mesure les individus peuvent exprimer leurs opinions sans restrictions du gouvernement. Contrairement à la précédente variable, ce score va de 1 à 10, et ici, plus l'on se rapproche de 10, plus la liberté d'expression est bonne. Pour donner une idée de ce que chaque score peut représenter, en voici les plus importants.

- Score de 1 : « La liberté d'expression est refusée. Les médias indépendants n'existent pas ou sont interdits ».
- Score de 4 : « La liberté d'expression est souvent soumise à des interférences ou à des restrictions gouvernementales. La distorsion et la manipulation façonnent les sujets du débat public ».
- Score de 7 : « La liberté d'expression est occasionnellement soumise à des interférences ou à des restrictions gouvernementales, mais il n'y a généralement pas d'incidents d'intrusions flagrantes comme la censure d'État ou la fermeture des médias ».
- Score de 10 : « La liberté d'expression est garantie contre les interférences ou les restrictions gouvernementales. Les individus peuvent pleinement exercer leurs droits¹¹ ».

L'indice de transformation Bertelsmann suit une méthodologie précise. Tous les pays en développement et en transition de plus d'un million d'habitants sont examinés. Les pays en développement et en transition sont les pays qui ne sont pas considérés comme étant consolidés sur le plan démocratique et économique. Autrement dit, qui ne sont pas bien implémentés

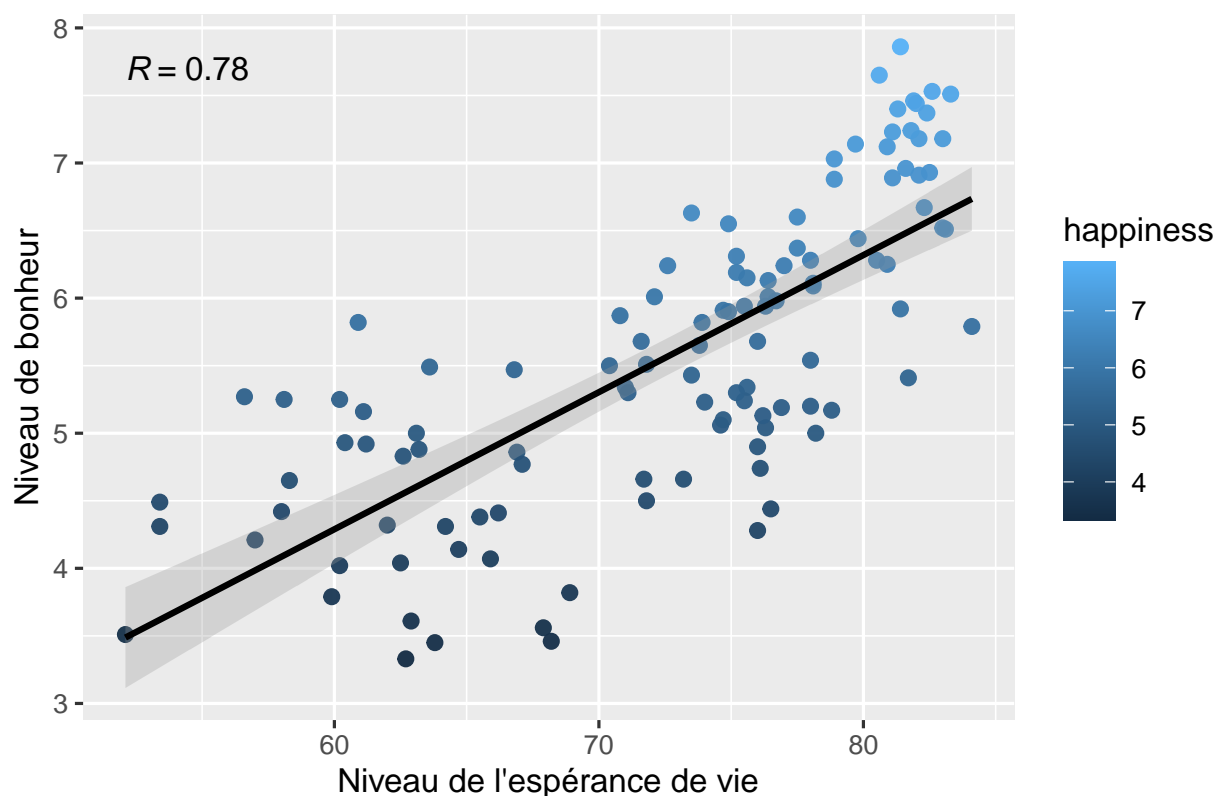
9. <https://bti-project.org/en/methodology>[consulté le 08/10/2022]

10. <https://core.ac.uk/download/pdf/36704556.pdf>[consulté le 08/10/2022]

11. <https://ourworldindata.org/grapher/freedom-of-expression-bti?tab=table>[consulté le 08/10/2022]

dans une économie de marché.

FIGURE 2.2 – Corrélation entre la liberté d'expression et le niveau de bonheur



Ainsi, on observe une corrélation positive entre le niveau de bonheur et le niveau de liberté d'expression (Figure 2.2). En utilisant la méthode de Spearman, on obtient un coefficient de corrélation de 0,6. Il semblerait donc qu'au sein d'un même pays, plus le niveau de liberté d'expression est élevé, plus le niveau de bonheur est élevé. Cependant, on observe une dispersion très importante des observations par rapport à la droite de corrélation.

Encore une fois, on remarque que certains pays ont un niveau de bonheur très faible malgré un niveau de liberté d'expression assez élevé. Cela nous indique que cet indicateur seul n'est pas suffisant pour déterminer le niveau de bonheur d'un pays.

Liberté d'association

Comme précédemment, cette variable est également basée sur les évaluations d'experts et la notation par le *Bertelsmann Transformation Index* (2022). Cette variable indique dans quelle mesure les individus et les groupes peuvent s'associer et se réunir sans oppressions et/ou

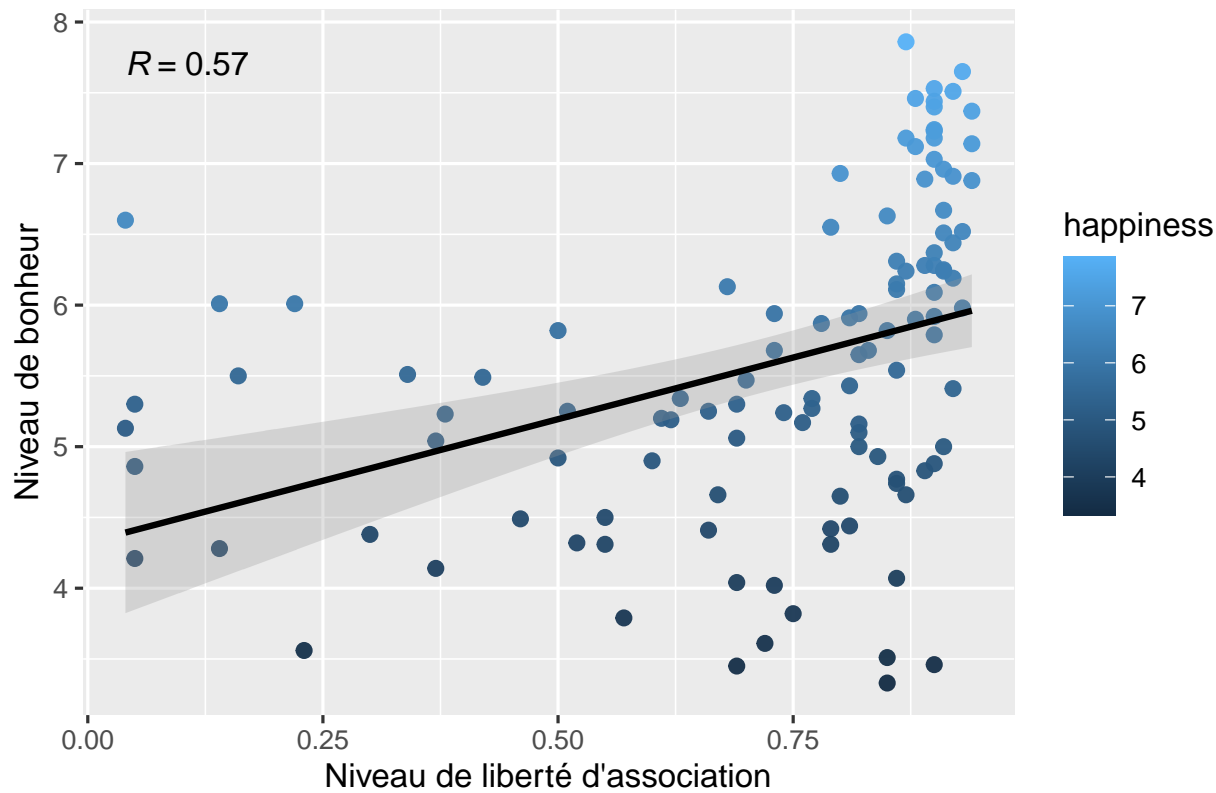
restrictions du gouvernement. De la même façon que pour la liberté d'expression, la liberté d'association adopte une notation allant de 1 à 10, plus on se rapproche de 10, plus la liberté est élevée.

- Score de 1 : « Les droits d'association et de réunion sont niés. Les groupes civiques indépendants n'existent pas ou sont interdits. »
- Score de 4 : « Les droits d'association et de réunion sont souvent soumis à des interférences ou à des restrictions gouvernementales. Les résidents et les groupes civiques qui ne soutiennent pas le gouvernement ne peuvent souvent pas exercer ces droits. »
- Score de 7 : « les droits d'association et de réunion sont occasionnellement soumis à des interférences ou à des restrictions gouvernementales, mais en général, il n'y a pas d'interdiction absolue des groupes politiques ou civiques indépendants. »
- Score de 10 : « les droits d'association et de réunion sont garantis contre toute interférence ou restriction gouvernementale. Les résidents et les groupes civiques peuvent pleinement exercer ces droits. » Les autres notes sont des catégories intermédiaires. »¹²

Comme pour la liberté d'expression, les catégories citées ci-dessus représentent les plus importantes, les autres sont seulement des sous-catégories. Cet indicateur nous permet, au même titre que les deux autres indicateurs de liberté de faire le lien entre la liberté que laisse le gouvernement à sa population et le niveau de bonheur de cette dernière.

12. <https://ourworldindata.org/grapher/freedom-of-association-bti?tab=table>[consulté le 09/10/2022]

FIGURE 2.3 – Corrélation entre la liberté d'association et le niveau de bonheur



Comme nous l'imaginons, dans un même pays, on observe une corrélation positive entre le niveau de bonheur et le niveau de liberté d'association (Figure 2.3). Le coefficient de corrélation de 0,57 nous indique une corrélation positive. Ainsi, il semblerait qu'au sein d'un même pays, plus le niveau de liberté d'association est élevé, plus le niveau de bonheur est élevé. Cependant, on remarquera que cette relation est plus fragile que les deux dernières.

Esperance de Vie

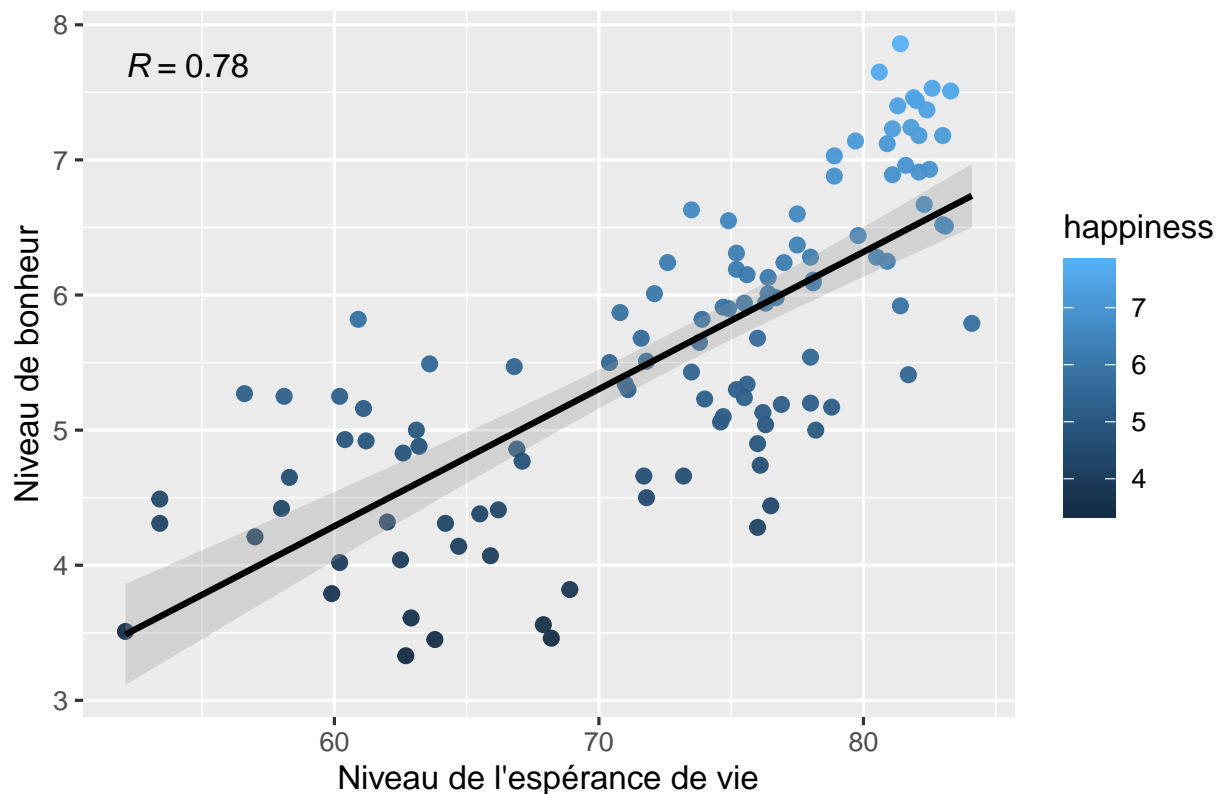
L'espérance de vie à la naissance est un indicateur statistique rendant compte de « l'âge moyen auquel seraient décédés les femmes et les hommes qui auraient eu tout au long de leur vie les taux de décès par sexe et par âge »¹³, pour une année donnée. Nous trouvons cet indicateur particulièrement pertinent pour évaluer l'état de santé d'une population. De plus, sa popularité fait que nous disposons d'un vaste échantillon de pays à étudier. S'il aurait été peut-être plus pertinent d'étudier l'espérance de vie en bonne santé, nous avons malheureusement fait

13. https://blog.insee.fr/wp-content/uploads/2022/01/blogInsee_2022_01-lesperance_de_vie.pdf [consulté le 08/10/2022]

face à l'indisponibilité des ces données à l'échelle mondiale.

Les liens entre bonheur et santé ont fait l'objet de nombreuses études au cours du temps. Dans un papier de 2014, l'économiste Fabio Sabatini relevait que, empiriquement, le bonheur était l'indicateur le plus efficace pour prédire la santé d'une population. C'est pourquoi, nous pensons qu'il existe un lien très fort entre l'espérance de vie et le niveau de bonheur d'une population¹⁴. On pourrait même s'attendre à ce que l'association de ces deux variables crée de l'endogénéité via un biais de simultanéité. La théorie voudrait donc que plus l'espérance de vie est élevée, plus le niveau de bonheur est élevé.

FIGURE 2.4 – Corrélation entre l'espérance de vie et le niveau de bonheur



Les données viennent confirmer la théorie, plus le niveau de l'espérance de vie est élevé, plus le niveau de bonheur est élevé. À l'aide d'une projection graphique et en utilisant la méthode de Spearman (Figure 2.4), nous observons une corrélation positive très forte (0,78) entre ces

14. https://www.sciencedirect.com/science/article/pii/S0277953614003189?casa_token=l_cVnCHico0AAAAA:k674cRYqF2K3pHUTHv6CDkjLGDiwKAQuyLDA-g4UMVzPRNs7T9J9cB5S4u1nfOCmv-aH1yLsUA[consulté le 10/10/2022]

deux variables. Plus la population d'un pays est en bonne santé, plus son niveau de bonheur est élevé.

PIB par habitant

Souhaitant intégrer une variable économique dans notre modèle, le PIB par habitant est apparu comme une évidence. Créé dans l'objectif de mesurer les impacts macroéconomiques de la grande dépression¹⁵, le Produit Intérieur Brut (PIB) s'est aujourd'hui imposé comme l'indicateur de référence pour évaluer la santé d'une économie.

Dans notre étude, nous avons fait le choix d'utiliser le PIB/habitant car il a l'avantage de prendre en compte variations démographiques qui peuvent exister entre pays. Par exemple, la Chine, qui a le PIB le plus important du monde, se classe seulement en soixante douzième place lorsque l'on prend en compte sa démographie nationale. De ce fait, la mesure du PIB/habitant nous a paru plus pertinente pour mesurer le niveau de développement économique d'un pays.

Empiriquement, il semblerait que les mouvements macroéconomiques aient une importance significative sur le bonheur des pays. Dans *The Macroeconomics of Happiness*, une forte corrélation positive a pu être établie entre le PIB/habitant et le niveau de bonheur¹⁶. Cependant, cet indicateur ne semble pas être le reflet parfait du bien-être. Comme l'indique le paradoxe d'Easterlin¹⁷, il semblerait qu'à un certain point, l'augmentation du revenu n'a plus d'impact significatif sur le niveau de bonheur¹⁸. On pourrait même aller plus loin en faisant référence au concept de croissance antiéconomique d'Herman Daly qui montre qu'à partir d'un certain point, la croissance du PIB aurait plus d'inconvénients que d'avantages. Cela serait notamment dû aux coûts sociaux et environnementaux qu'elle apporterait¹⁹.

Toutes ces contradictions rendent le PIB/habitant encore plus intéressant pour notre analyse. Il s'agit ici de comprendre la force du lien entre ces deux variables pour mieux déterminer les mécanismes déterminants du bien-être d'une société. Pour aller plus loin, il s'agit surtout de voir comment l'impact du PIB/Habitants sur le niveau de bonheur varie entre les différents

15. <https://www.cairn.info/revue-projet-2012-6-page-14.htmno4>[consulté le 08/10/2022]

16. <https://direct.mit.edu/rest/article-abstract/85/4/809/57422/The-Macroeconomics-of-Happiness>[consulté le 31/10/2022]

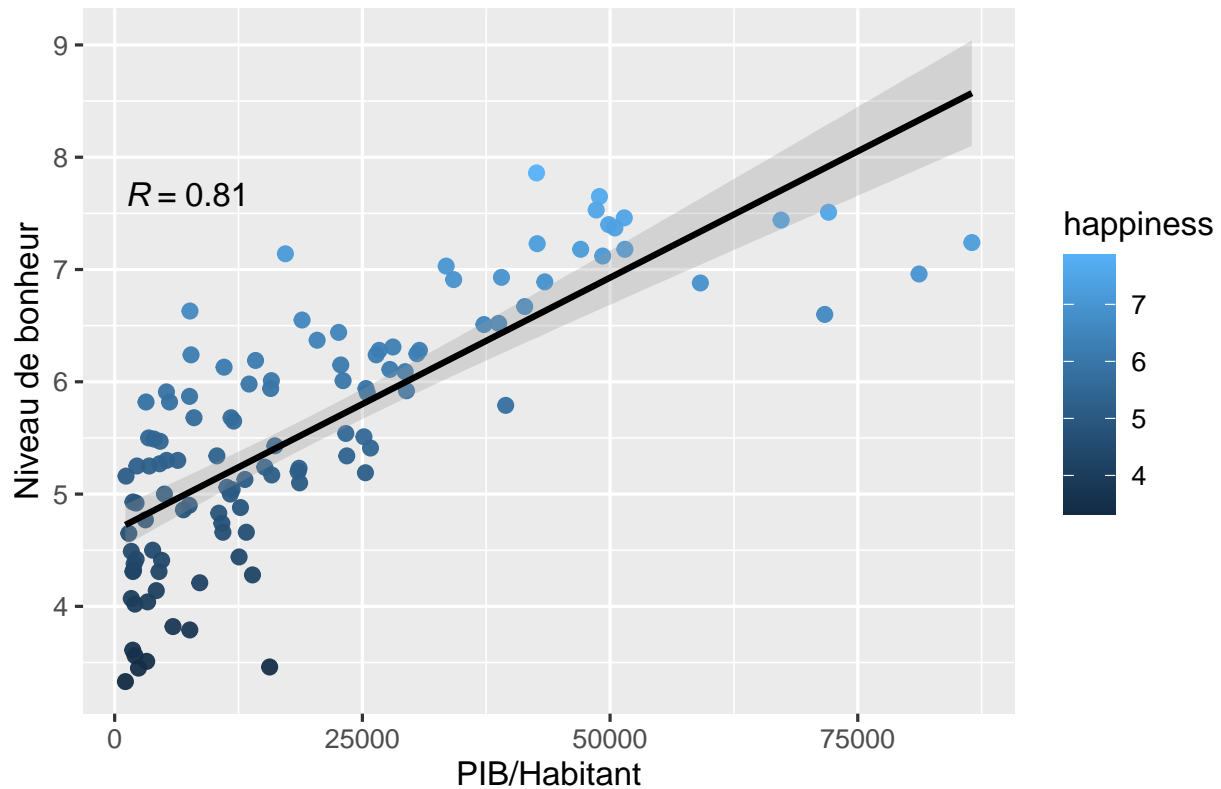
17. <http://graphics8.nytimes.com/images/2008/04/16/business/Easterlin1974.pdf>[consulté le 11/10/2022]

18. <https://www.sciencedirect.com/science/article/abs/pii/S0304387807000600>[consulté le 11/10/2022]

19. <http://pinguet.free.fr/daly1996.pdf>[consulté le 11/10/2022]

pays. Par exemple, entre des pays développés et des pays en voie de développement.

FIGURE 2.5 – Corrélation entre le PIB par habitant et le niveau de bonheur



Conformément aux observations de Rafael D. Tella mentionné plus haut, il existe une forte corrélation positive entre le PIB/habitant et le niveau de bonheur. On peut voir le coefficient de corrélation en haut à gauche de la Figure 2.5, il est de 0,81. On remarque également que certains pays ont un niveau de bonheur plus élevé malgré un PIB/habitant deux à trois fois plus faible, il sera donc intéressant de relever les facteurs (s' il y en a) ayant une influence négative sur le bonheur dans ces pays. Malgré tout, il semblerait qu'au sein d'un même pays, plus le PIB/Habitant est élevé, plus le niveau de bonheur est élevé.

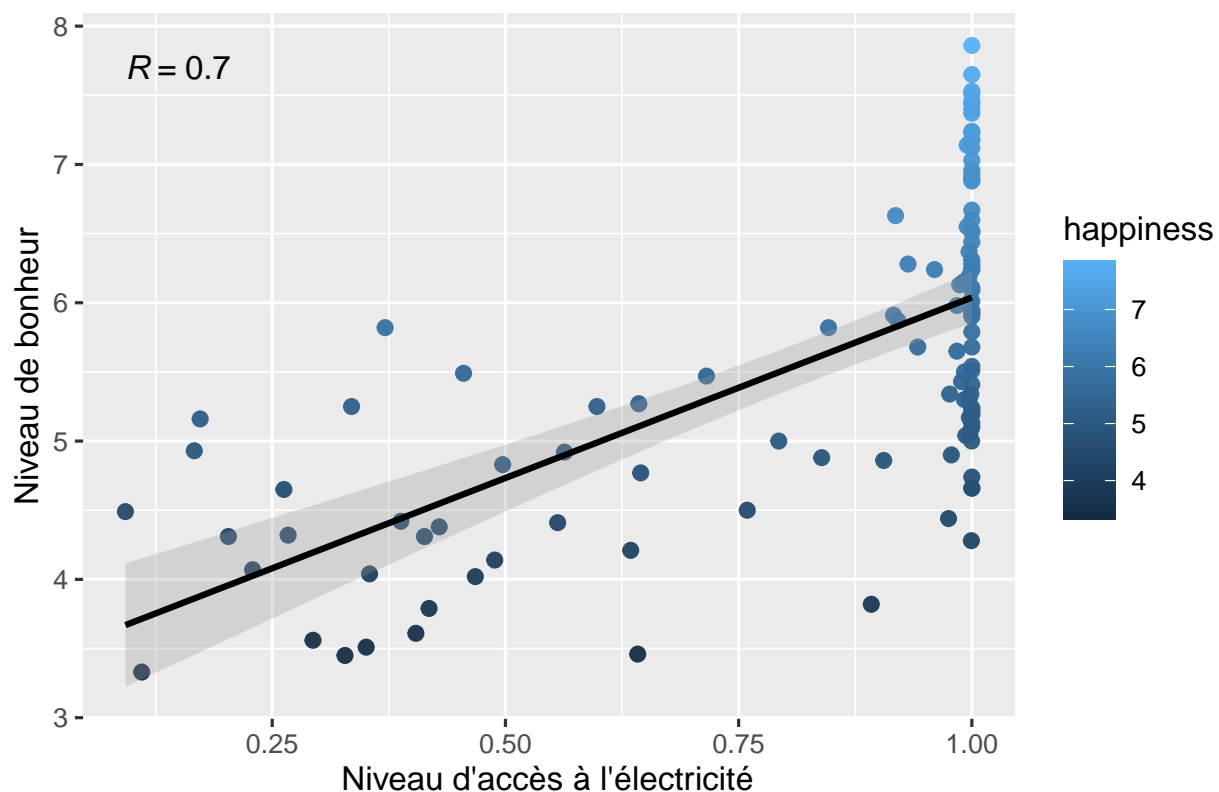
L'accès à l'électricité

L'accès à l'électricité se définit comme le pourcentage de la population d'un pays ayant accès à l'électricité. Cet indicateur est construit par la Banque Mondiale à l'aide de données collectées internationalement .

Dans nos économies développées l'accès à l'électricité apparaît comme une évidence, ce n'est pas le cas pour tout le monde. « *Energy Poverty : How to make modern energy access universal* ²⁰ », c'est le nom d'un rapport publié conjointement par les Nations Unies et l'Agence Internationale de l'énergie en 2010. Ce rapport souligne qu'aujourd'hui encore, plus d'un milliard de personnes ne disposent pas d'accès permanent à l'électricité. Or, nous savons à quel point cette énergie est importante pour s'éclairer, se nourrir ou encore se chauffer. Dès lors, l'accès à l'électricité est vital pour assurer le développement humain d'un pays ²¹.

C'est pourquoi, nous pensons que cet indicateur synonyme de développement et de confort est important dans la détermination du bonheur. Nous nous attendons donc à une corrélation positive entre l'accès à l'électricité et le niveau de bonheur.

FIGURE 2.6 – Corrélation entre l'accès à l'électricité et le niveau de bonheur



Il semble bien y avoir une corrélation positive entre l'accès à l'électricité et le niveau de bon-

20. https://www.undp.org/sites/g/files/zskgke326/files/publications/Energy_Poverty_Excerpt_WEO2010.pdf[consulté le 08/10/2022]

21. <https://www.sciencedirect.com/science/article/pii/S097308261300094X>[consulté le 08/10/2022]

heur. La Figure 2.6 indique un coefficient de corrélation de 0,7. Ainsi, il semblerait que plus le niveau d'accès à l'électricité est élevé, plus le niveau de bonheur est élevé. Néanmoins, la partie droite du graphique nous montre que la distribution du niveau de bonheur semble très hétérogène. Certains pays ont accès à l'électricité à 100% mais ont un niveau de bonheur relativement faible. Dès lors, l'accès à l'électricité ne semble pas une condition suffisante pour expliquer un niveau de bonheur élevé.

Le nombre d'années d'études

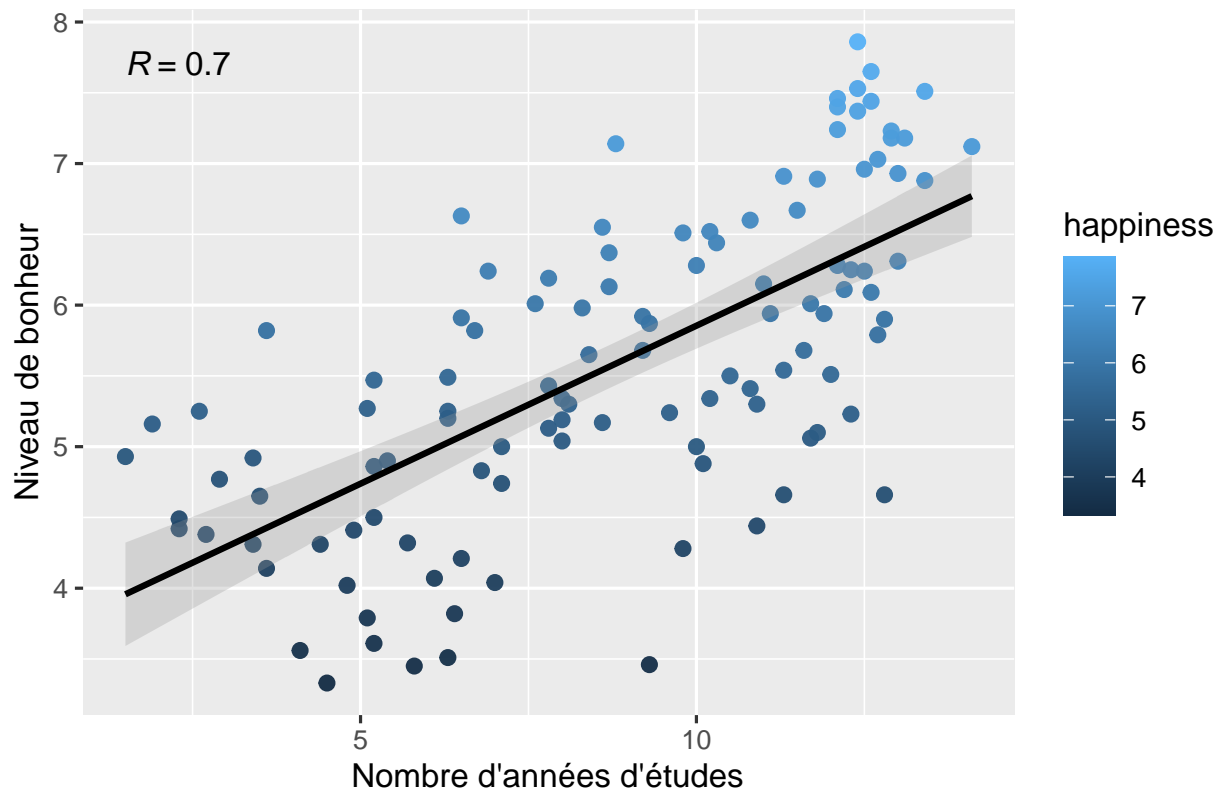
Il s'agit ici d'introduire une variable sociale dans notre étude. Nombreux sont les pays à la recherche d'une croissance forte et élevée, mais pour étudier le bien-être d'une société il ne faut pas négliger l'aspect émotionnel de l'humain, et surtout, le développement de ses facultés intellectuelles. La variable représentée correspond au nombre moyen d'années pendant lesquelles les personnes âgées de 25 ans et plus ont suivi un enseignement formel. Des recherches^{22 23} ont démontré que le nombre d'années d'étude avait un impact très significatif pour les pays les plus sous-développés et/ou en voie de développement. Autrement dit, l'impact de l'éducation est bien plus important sur des pays qui en ont un accès restreint que sur des pays déjà très développés. Il est intéressant de prendre pour seuil l'âge de 25 ans, puisqu'il semblerait que c'est à cet âge que le niveau de bonheur lié à l'éducation augmente²¹. Cependant il n'est pas certain que les pays au niveau d'éducation les plus élevés soit les pays les plus heureux, c'est ce que nous allons essayer de vérifier en introduisant cette variable dans notre étude.

22. https://www.jstor.org/stable/23261310?seq=3metadata_info_tab_contents[consulté le 09/10/2022]

23. <https://www.jcreview.com/admin/Uploads/Files/61c8411eba12d8.03312798.pdf>[consulté le 09/10/2022]

21. https://www.researchgate.net/publication/303698127_Education_and_happiness_an_alternative_hypothesis[consulté le 09/10/2022]

FIGURE 2.7 – Corrélation entre le nombre d'années d'études et le niveau de bonheur



Comme nous l'attendions, on observe une corrélation positive entre le niveau de bonheur et le nombre moyen d'années d'études des personnes âgées de 25 ans (Figure 2.7). En haut à gauche du graphique, on peut voir que le coefficient de corrélation est de 0,7 en utilisant la méthode de Spearman. Ainsi, il semblerait qu'au sein d'un même pays, plus le nombre moyen d'années d'études des personnes âgées de 25 ans est élevé, plus le niveau de bonheur est élevé.

La mortalité infantile

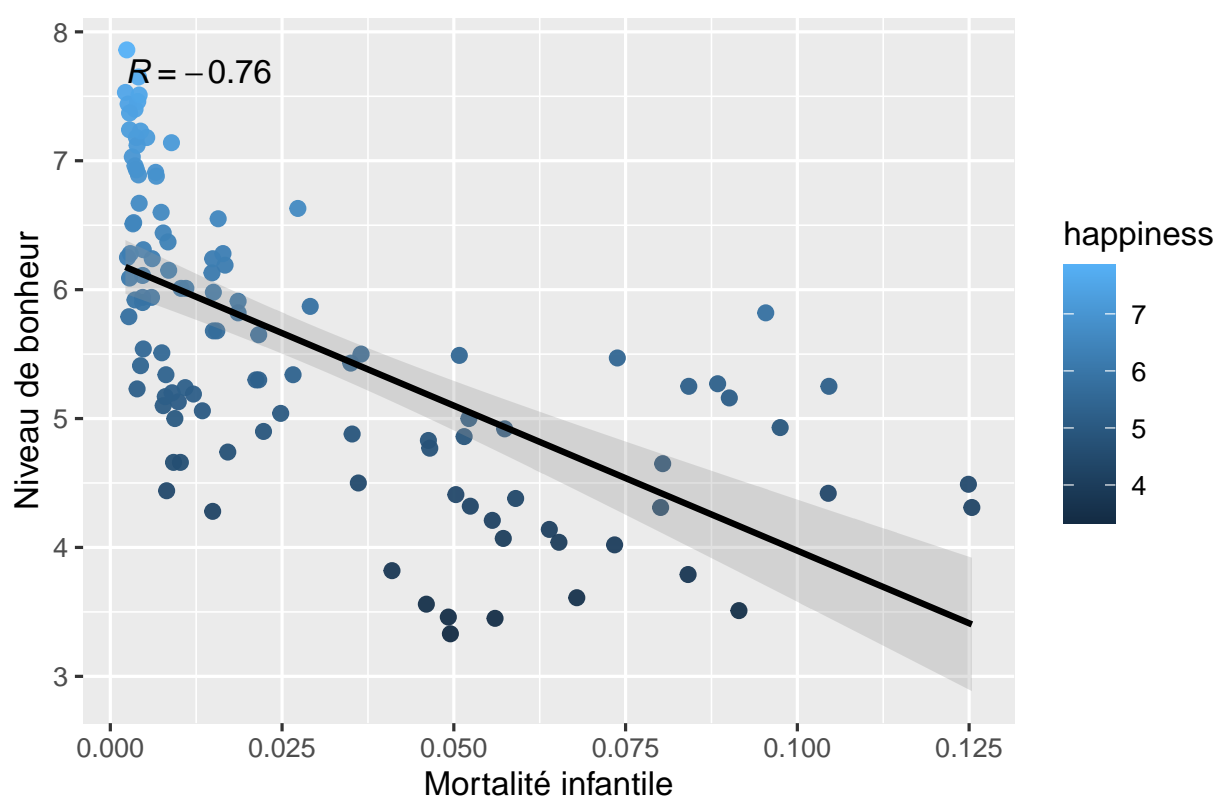
La mortalité infantile est un taux représentant le nombre (pour 1000 habitants) d'enfants mourant avant d'atteindre l'âge d'un an. Cette indicateur fait partie des nombreux indicateurs présents dans *World Development Report* ²⁴ de la Banque Mondiale.

Les pays pauvres sont les principaux touchés par la mortalité infantile. Si les causes sont mul-

24. <https://www.worldbank.org/en/publication/wdr/wdr-archive>[consulté le 08/10/2022]

tiples²⁵, la mortalité infantile est fortement accentuée par le manque ou la mauvaise qualité des infrastructures (santé et sanitaires). Cela tend à se confirmer puisque même au sein des pays très pauvres, des disparités très significatives de mortalités sont observées en fonction de ceux ayant accès à de bonnes ou mauvaises infrastructures²⁶. Comme nous l'avons souligné plus haut avec l'espérance de vie, la santé et le bonheur ont des liens très forts. Nous pensons donc qu'avoir un taux de mortalité infantile élevé influence négativement le niveau de bonheur. De plus, étant un piège à pauvreté²⁷, la mortalité infantile réduit grandement les probabilités pour une famille de se développer à tous les points, bonheur compris.

FIGURE 2.8 – Corrélation entre la mortalité infantile et le niveau de bonheur



Les données viennent confirmer la théorie, on observe une corrélation négative entre la mortalité infantile et le niveau de bonheur (Figure 2.8). Cette corrélation est relativement forte puisque le graphique nous indique un coefficient de -0,76 entre les deux variables. Ainsi, il

25. <https://urlz.fr/jFJk>[consulté le 09/10/2022]

26. <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.21.1.141>[consulté le 09/10/2022]

27. https://www.brookings.edu/wp-content/uploads/2004/01/2004a_bpea_sachs.pdf[consulté le 11/10/2022]

semblerait qu'au sein d'un même pays, plus la mortalité infantile est élevée, plus le niveau de bonheur est faible.

L'accès à l'eau

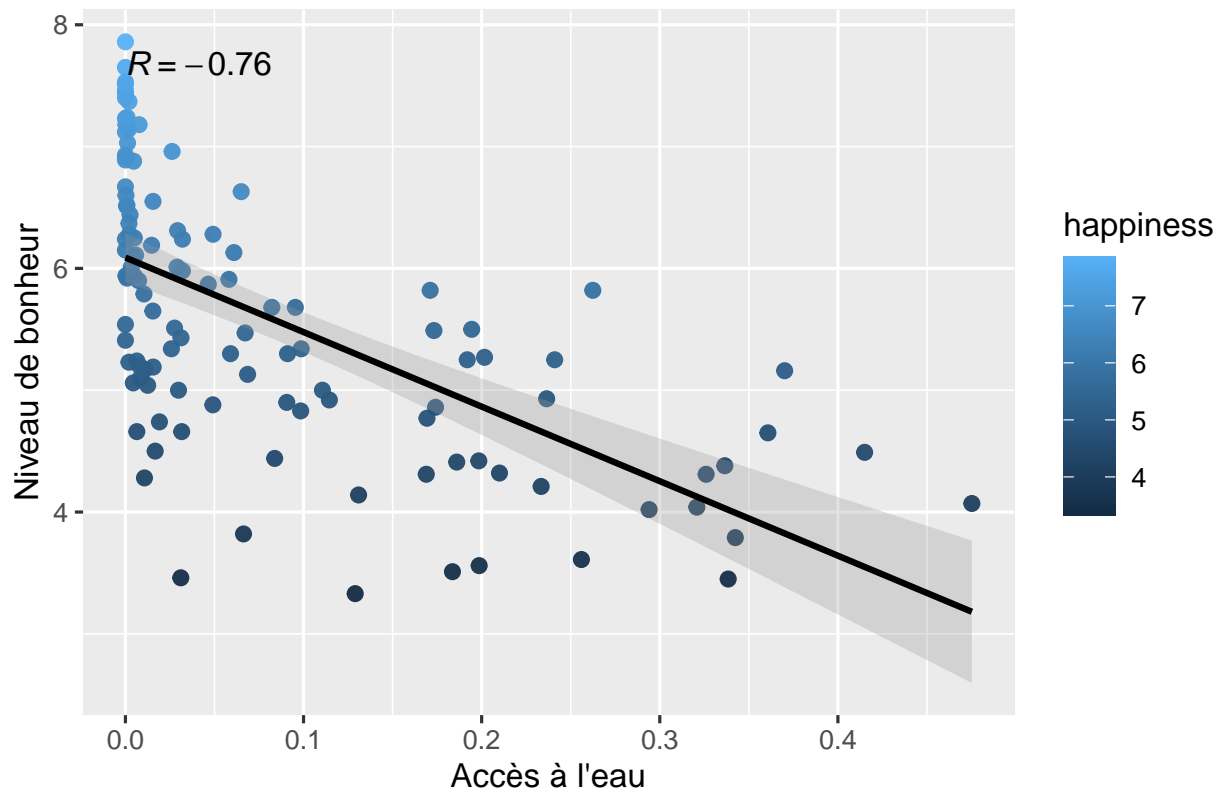
Dans notre étude, l'accès à l'eau renvoi à la part d'une population n'ayant pas accès à une source d'eau dite améliorée. Voici comment l'Unesco définit une source d'eau améliorée : « Une source d'eau potable améliorée est un point de distribution d'eau qui, par la nature de sa conception, protège l'eau de la contamination externe, en particulier d'origine fécale ²⁸ ». La valeur de la variable est comprise entre 0 et 1 et plus la valeur est proche de 1, plus la part de la population n'ayant pas accès à une source d'eau améliorée est grande. L'eau étant essentielle pour vivre, il nous a paru évident d'intégrer une variable à son sujet dans notre analyse.

Dans un papier portant sur la volonté des ménages marocain à disposer d'un accès à l'eau privé, des chercheurs ont montré qu'en plus d'augmenter les revenus et la participation au marché du travail, un accès à l'eau pouvait améliorer significativement le niveau de bonheur ²⁹. C'est pourquoi, nous pensons que l'accès à l'eau sera négativement corrélé avec notre variable explicative.

28. <http://uis.unesco.org/fr/glossary-term/eau-potable-0>[consulté le 09/10/2022]

29. <https://web.stanford.edu/~pdupas/MoroccoWaterConnections.pdf> [consulté le 09/10/2022]

FIGURE 2.9 – Corrélation entre l'accès à l'eau et le niveau de bonheur



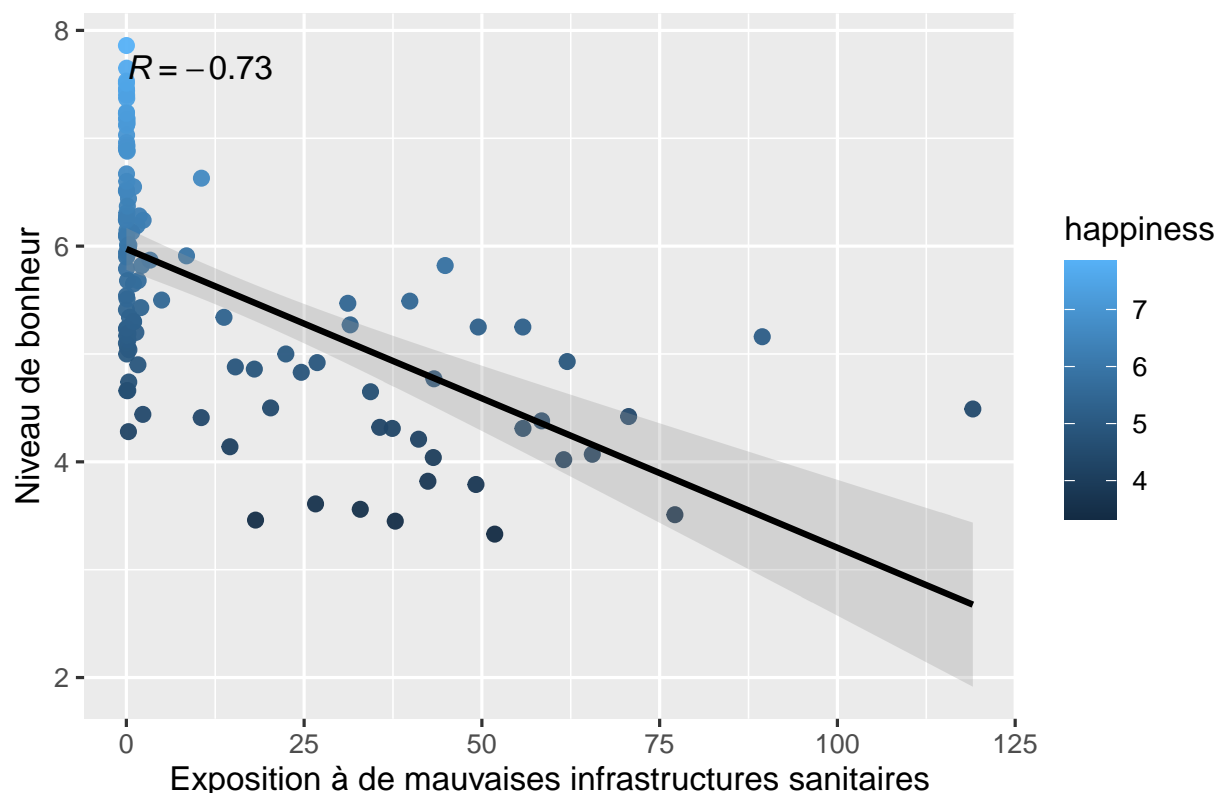
La figure 2.9 met en évidence la relation négative que nous avons prévu. On observe une corrélation négative (-0,76) entre le niveau de bonheur et la part de la population n'ayant pas accès à une source d'eau améliorée. On remarquera que la forme du graphique et les coefficients de corrélations entre cette variable et la mortalité infantile sont quasiment identiques. Cela nous montre à quel point ces deux variables explicatives sont liées. Cette observation n'est pas surprenante lorsque l'on sait à quelle point l'accès à une eau propre joue dans la mortalité infantile. Ainsi, il semblerait qu'au sein d'un même pays, plus la part de la population n'ayant pas accès à une source d'eau améliorée est élevée, plus le niveau de bonheur est faible.

Exposition à de mauvaises infrastructures sanitaires

L'exposition à de mauvaises infrastructures sanitaires renvoi au nombre de morts (pour 100 000 habitants) causé par l'exposition à de mauvaises infrastructures sanitaires. L'exposition à de mauvaises infrastructures sanitaires peut conduire à des maladies telles que la diarrhée,

l'ascaridiose ou l'ankylostomiase³⁰. Souvent associé à une eau de mauvaise qualité et un système de santé peu développé, ces maladies comptent parmi les causes de morts les plus importantes dans les pays sous-développés.

FIGURE 2.10 – Corrélation entre l'exposition à de mauvaises infrastructures sanitaires et le niveau de bonheur.



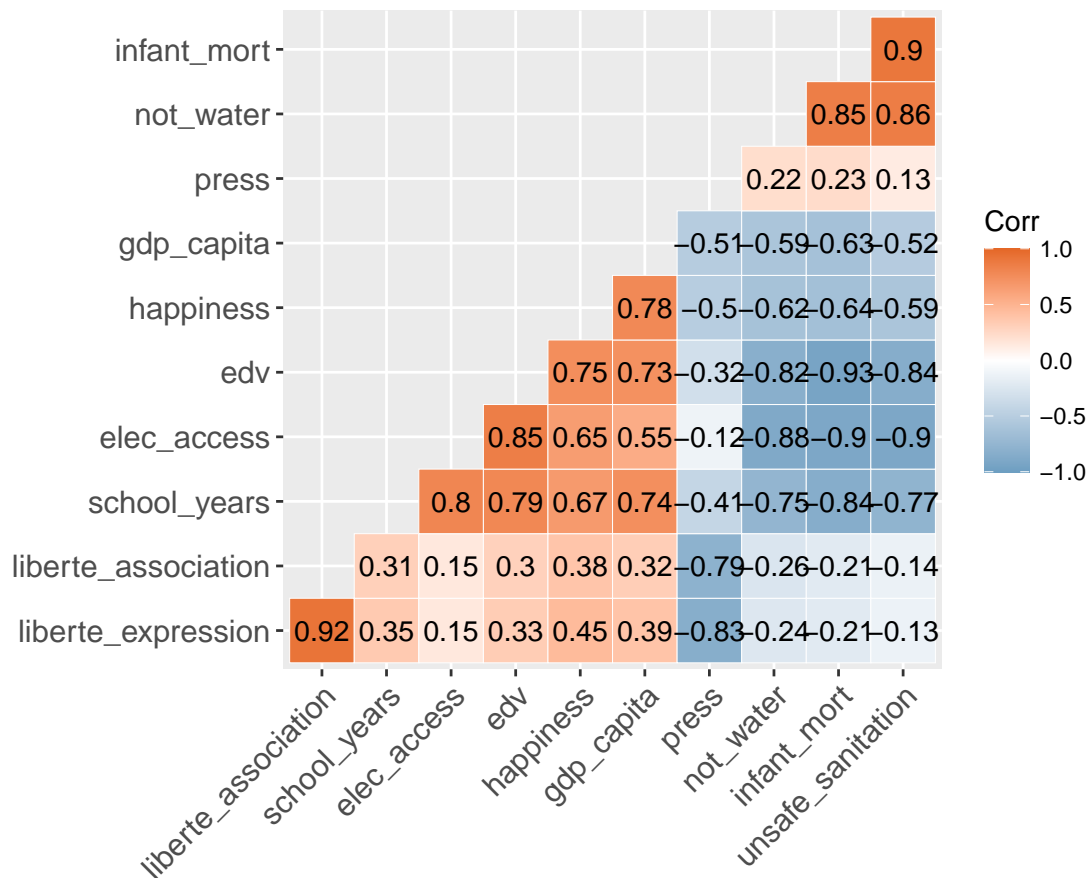
Comme pour les deux variables présentées au-dessus (accès à l'eau et mortalité infantile), on observe une corrélation négative entre notre variable explicative et notre variable illustrative. Si le coefficient de corrélation observé de -0,73 (Figure 2.10) est légèrement plus faible que pour les deux variables précédentes, les relations restent encore une fois très équivalentes. Ainsi, il semblerait qu'au sein d'un même pays, plus le taux de mortalité dû à l'insalubrité des installations sanitaires est élevé, plus le niveau de bonheur est faible.

30. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6593152/> [consulté le 11/10/2022]

2.0.3 Corrélation entre la variable illustrative et les variables actives

Matrice des corrélations

FIGURE 2.11 – Matrice des corrélations



La matrice des corrélations (Figure 2.11) nous permet d’observer l’ensemble des corrélations entre nos variables explicatives et notre variable illustrative. Dans un premier temps, nous nous pencherons sur notre variable illustrative. Les variables les plus corrélées positivement avec notre indicateur de bonheur sont le PIB par habitant et l’espérance de vie avec respectivement 0,78 et 0,75 comme coefficient de corrélation. Celles les plus corrélées, mais négativement, sont l’accès à l’eau et la mortalité infantile avec -0,64 et -0,62. Comme nous l’avons déjà deviné plus tôt, une forte corrélation positive peut être observée entre les variables : mortalité infantile (infant_mort), mauvaises infrastructures sanitaires (unsafe_sanitation) et accès à l’eau (not_water). Les coefficients de corrélations observés entre ces trois variables sont

tous proches de 0,9. Ces trois variables sont très corrélées puisque la mortalité infantile est largement due à des maladies causées par l'absence d'infrastructures sanitaires décentes ou un accès à l'eau limité. Le coefficient de corrélation entre la liberté d'expression et la liberté d'association est très proche de 1, les variables sont donc fortement corrélées. Cela semble logique quand l'on sait que la liberté d'association consiste en la possibilité de former des groupes ou des associations, ce qui représente déjà une forme d'expression en soi. Globalement, les hypothèses que nous avons formulées dans la partie économique semblent être respectées.

Chapitre 3

Analyse en composante principale

3.0.1 Statistiques descriptives

Résumé de nos variables

TABLE 3.1 – Statistiques descriptives de nos variables

Pays	happiness	gdp_capita	elec_access
Length:114	Min. :3.330	Min. : 1085	Min. :0.0926
Class :character	1st Qu.:4.838	1st Qu. : 4811	1st Qu.:0.6435
Mode :character	Median :5.495	Median :13427	Median :0.9970
	Mean :5.579	Mean :20046	Mean :0.8240
	3rd Qu.:6.303	3rd Qu.:28010	3rd Qu.:1.0000
	Max. :7.860	Max. :86516	Max. :1.0000
school_years	liberte_expression	liberte_association	press
Min. : 1.500	Min. :0.0200	Min. :0.0400	Min. : 8.59
1st Qu. : 6.300	1st Qu.:0.6050	1st Qu.:0.6600	1st Qu.:22.00
Median : 9.200	Median :0.8050	Median :0.8200	Median :29.45
Mean : 8.765	Mean :0.7207	Mean :0.7209	Mean :31.62
3rd Qu.:11.875	3rd Qu.:0.9275	3rd Qu.:0.9000	3rd Qu.:40.58
Max. :14.100	Max. :0.9900	Max. :0.9400	Max. :80.96
edv	unsafe_sanitation	infant_mort	not_water
Min. :52.10	Min. : 0.0100	Min. :0.002200	Min. :0.00000
1st Qu.:66.35	1st Qu. : 0.0425	1st Qu.:0.004725	1st Qu.:0.00205
Median :75.20	Median : 0.3550	Median :0.014850	Median :0.02830
Mean :72.71	Mean : 14.2700	Mean :0.028739	Mean :0.08339
3rd Qu.:78.88	3rd Qu. : 24.0575	3rd Qu.:0.049425	3rd Qu.:0.13043
Max. :84.10	Max. :119.0500	Max. :0.125400	Max. :0.47520

Dans cette partie, nous allons nous aider de la Table 3.1 dans le but d'effectuer une première analyse descriptive de notre base.

Pour notre variable illustrative, nous remarquons que la médiane est quasiment identique à notre moyenne. Une moyenne proche de la médiane nous donne l'indication que la distribution de notre variable illustrative est relativement symétrique, la skewness devrait donc être très proche de 0. Nous pouvons dire la même chose de l'espérance de vie et du nombre d'années d'études.

Pour l'accès à l'électricité, la médiane est à 99,7%, cela signifie que la moitié des pays de notre échantillon ont un accès à l'électricité quasiment parfait. Néanmoins, le premier quartile nous indique qu'un quart des pays de notre échantillon ont un accès à l'électricité entre 50% et 66%. On peut donc déjà identifier de fortes disparités de développement entre certains pays de notre échantillon.

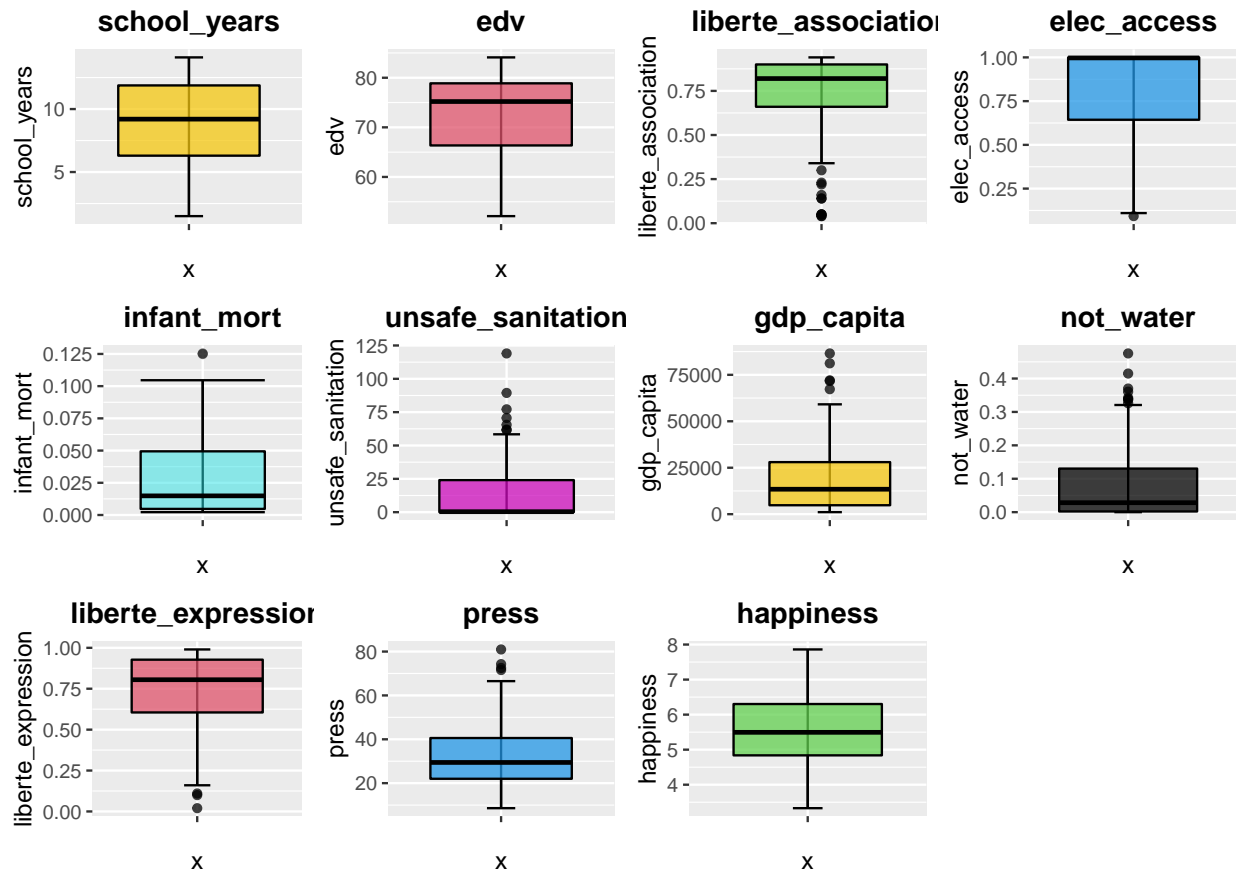
En ce qui concerne le PIB/habitant, on remarque que les trois premiers quarts de notre échantillon ont un PIB/habitant variant de 1 000\$ à 28 000\$ et que le dernier quart a un PIB/habitant compris entre 28 000 et 86 000\$. On a donc une minorité de pays qui tirent probablement artificiellement la moyenne des revenus vers le haut. Il est aussi fort probable qu'on détecte la présence de Pays atypiques dans notre échantillon pour cette variable. Nous vérifierons ça dans la partie suivante où nous traiterons des boîtes à moustaches.

Pour les indicateurs de liberté d'expression et d'association, nous observons un écart très important entre la valeur minimum et le premier quartile. De plus, leurs médianes sont supérieures à leur moyenne. Ainsi, nous pouvons supposer une asymétrie dans la distribution de ces deux variables.

Pour quasiment l'ensemble des variables, on peut noter des disparités importantes entre les valeurs minimums/maximums et les moyennes/quartiles. Cela laisse supposer la présence de valeurs atypiques.

Boîtes à moustaches

FIGURE 3.1 – Boîtes à moustaches



Grâce aux boîtes à moustache, nous pouvons vérifier l'existence (ou non) de valeurs atypiques (Figure 3.1). Nous constatons que pour le nombre d'années d'étude (*school_years*), l'espérance de vie (*edv*) et notre niveau de bonheur (*happiness*), il n'y a aucune valeur atypique. En revanche, toutes les autres variables présentent des valeurs atypiques. Pour vérifier si ces valeurs sont réellement atypiques, nous utiliserons 3 tests différents. Dans le cas où il n'y a qu'une seule valeur supposée atypique présente sur le diagramme (c'est le cas pour *elec_access* et *infant_mort*), nous utilisons le test de Grubbs. Pour ces deux variables, la p-valeur du test est supérieure à 0.05, autrement dit, nous ne rejetons pas l'hypothèse nulle H_0 qui confirme qu'il n'y a pas d'observations atypiques dans notre échantillon au seuil de risque de 5% (Annexe Table 6.1).

Dans le second cas, où il y a entre 2 et 10 valeurs supposées atypiques qui apparaissent dans

le diagramme (c'est le cas pour les variables *liberte_association*, *unsafe_sanitation*, *gdp_capita*, *not_water*, *liberte_expression* et *press*) nous utilisons le test de Rosner. Pour les variables *press*, *liberte_expression* et *liberte_association* le test nous indique qu'il n'y a aucune valeurs atypiques. En revanche, les autres variables testées présentent des valeurs atypiques. En effet, la variable *unsafe_sanitation* admet deux valeurs atypiques : 119.05 et 89.43 correspondant (respectivement) au taux de mortalité pour 100 000 individus dû aux mauvaises infrastructures sanitaires du Tchad et du Niger. Le PIB/habitant admet 1 valeur atypique : 86516.09 correspondant au PIB/Habitant du Luxembourg. La variable *not_water* 1 valeur atypique : 0.4752 correspondant à la part de la population n'ayant pas accès à une source d'eau améliorée à Madagascar. Ce résultat n'est pas surprenant quand on sait que le taux de la population malgache ayant accès à l'eau potable est d'environ 24%. En effet, d'après le rapport de mise à jour du Plan sectoriel de l'Eau, de l'Assainissement et de l'Hygiène du mois de juillet 2020 ¹, on peut y lire : « Lors de l'élaboration de ce dernier, des collectes des données relatives aux taux d'accès en Eau potable et Assainissement de base ont sorti un taux d'accès en Eau potable de 24% et un taux d'accès aux infrastructures d'assainissement de base de 30% ».

Grâce aux tests de normalité de Shapiro, on peut tester les hypothèses de normalité des variables possédant des valeurs atypiques. L'hypothèse nul (H_0) signifie que la variable suit une loi normale, tandis que H_1 signifie l'inverse. Les résultats du test (Annexe Table 6.2) démontrent que toutes les variables présentant des valeurs atypiques ne suivent pas une loi normale.

3.0.2 Valeurs propres et nombre d'axes

Nous allons également consacrer une partie à l'analyse de nos valeurs propres. Les valeurs propres permettent de nous donner l'inertie de chaque axe.

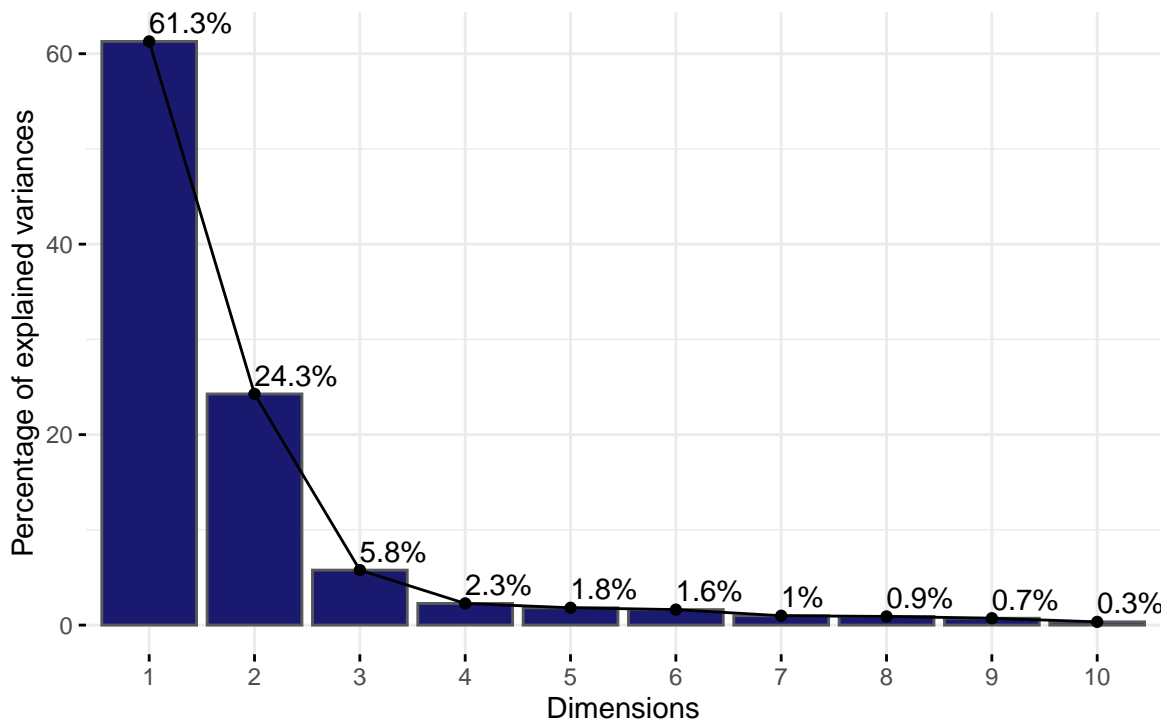
1. https://www.pseau.org/outils/ouvrages/unicef_meah_phase_5_rapport_de_mise_a_jour_revision_du_pseah_2021.pdf [consulté le 28/10/2022]

TABLE 3.2 – Tableau des valeurs propres

Axes	Valeurs propres	Variance (%)	Variance cumulée (%)
1	6.13	61.29	61.29
2	2.43	24.27	85.56
3	0.58	5.77	91.33
4	0.23	2.27	93.61
5	0.18	1.82	95.42
6	0.16	1.63	97.05
7	0.10	0.98	98.03
8	0.09	0.91	98.94
9	0.07	0.72	99.66
10	0.03	0.34	100.00

Grâce à la Table 3.2, nous remarquons que nous pouvons représenter notre base de données sur 10 dimensions, l'objectif ici est de réduire ce nombre de dimensions pour nous focaliser sur deux, ou trois dimensions seulement. On constate que nos deux premiers axes nous fournissent déjà 85,6% d'inertie. Si la différence entre les deux premiers axes est très grande, elle a tendance à se réduire au fur et à mesure (Figure 3.2) L'inertie étant très importante sur ces deux axes, nous décidons de nous concentrer majoritairement sur ces derniers. Nous verrons plus tard dans l'analyse si l'Axe 3 peut nous être utile.

FIGURE 3.2 – Histogramme des valeurs propres

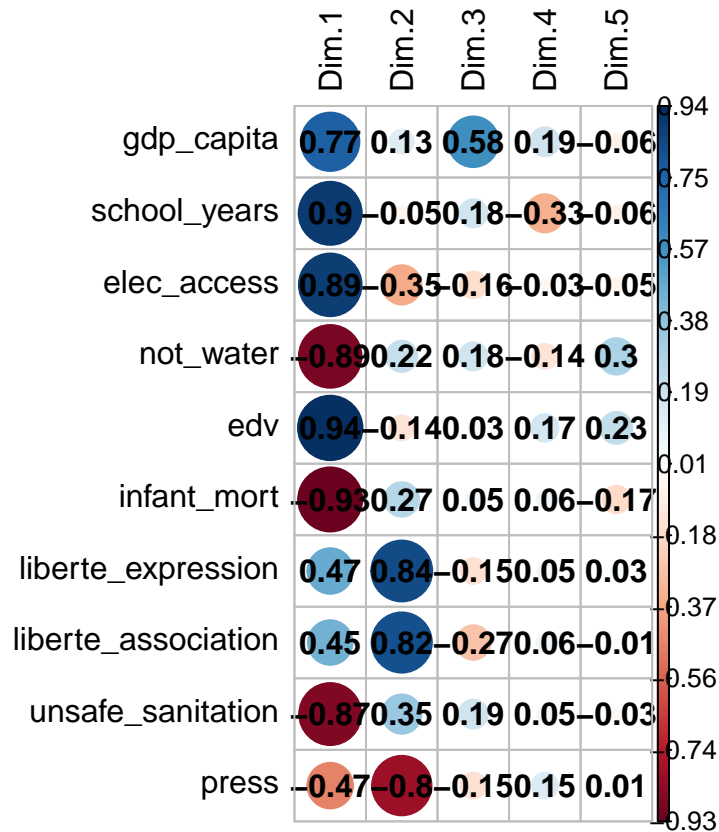


3.0.3 Variables, contribution, corrélations, cosinus carrés

Les coordonnées/corrélations

Pour commencer, nous allons analyser les coordonnées /corrélations de nos différentes variables. En effet, les coordonnées permettent d'indiquer la position de chaque variable le long des axes, et ainsi de les positionner dans les différents cadrans.

FIGURE 3.3 – Coordonnées de nos variables



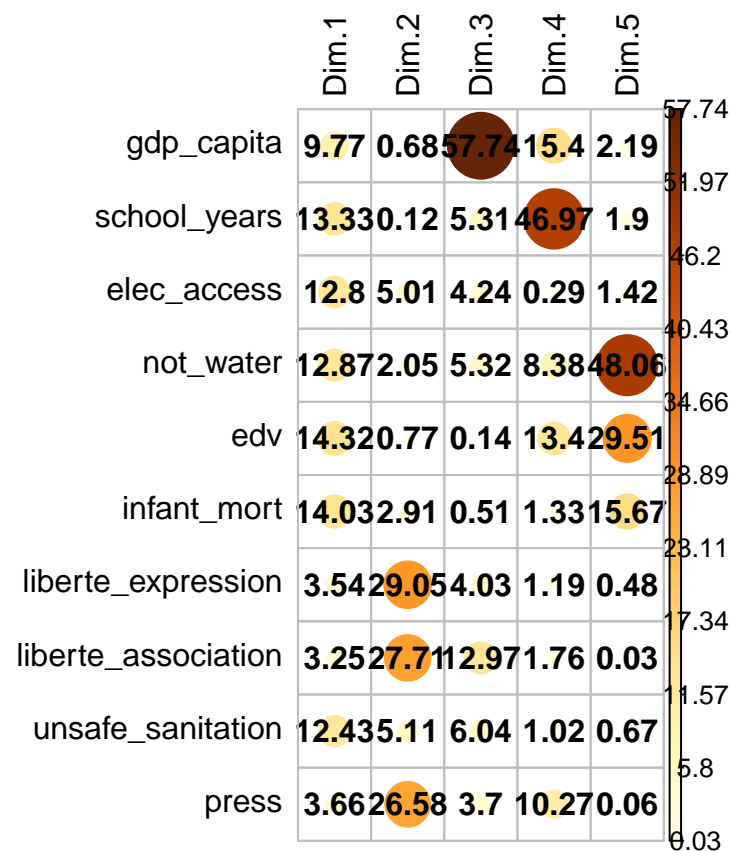
On remarque directement les variables qui représentent le mieux les différents axes. Dans notre graphique (Figure 3.3), la force de la corrélation est déterminée à la fois par la taille du cercle et par sa couleur. Plus le diamètre du cercle est grand, plus la corrélation est forte. En ce qui concerne les couleurs, le bleu nous indique une force positive et le rouge une force négative. Plus ces couleurs seront foncées, plus la force sera grande.

Concernant l'axe 1, nous remarquons que presque toutes les variables (hormis *liberte_expression*, *liberte_association* et *press*) ont des coordonnées / corrélations très fortes. Autrement dit, nous pouvons en déduire que ces différentes variables seront positionnées très loin sur l'axe 1 (loin du centre). À l'aide du code couleur définie plus haut, nous pouvons également définir dans quelle cadran chaque variable va se situer. Pour l'axe 1, les variables dont la corrélation est positivement élevée se situeront dans les cadrans I et IV (à droite), celles dont la corrélation est négativement élevée se situeront dans les cadrans II et III (à gauche).

Concernant l'axe n°2, les variables les plus contributives sont toutes celles qui ne l'étaient pas sur l'axe 1, c'est-à-dire *liberte_expression*, *liberte_association* ainsi que *press*. Les deux premières se situeront dans cadran I (en haut à droite) car elles sont positivement corrélées à l'axe 1 et 2. La dernière, négativement corrélée aux deux axes, se situera dans le cadran III (en bas à gauche). Notre objectif ici est de voir si des groupes de variables se forment, par opposition ou par regroupement. Nous y reviendrons plus tard dans l'analyse.

Les contributions

FIGURE 3.4 – Contributions de nos variables



Ce tableau des contributions (Figure 3.4) nous montre quelles variables contribuent le plus aux différents axes, autrement dit, quelles variables participent le plus à la formation des différents axes ou bien encore quel poids ont ces variables dans la formation des axes. Nous pouvons utiliser un seuil pour sélectionner les variables les plus pertinentes dans la définition de

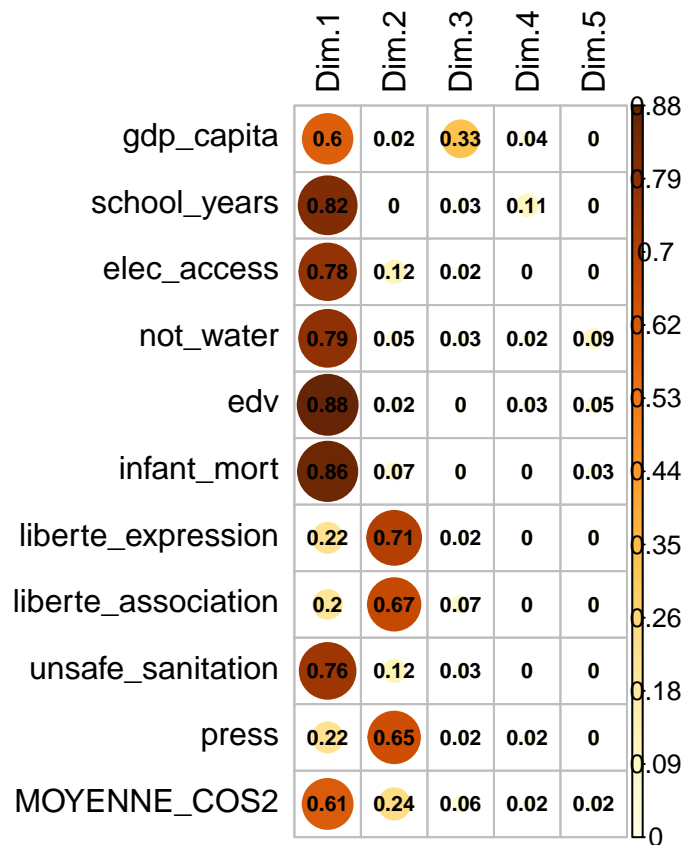
nos variables latentes, tel que :

$$seuildecontribution(\%) = \frac{100}{nombredevariables} = \frac{100}{10} = 10$$

En ce qui concerne l'axe 1, la plupart des variables ont environ la même contribution, entre 10% et 14% (Figure 3.4), ce qui est supérieur à notre seuil. Elles participent presque toutes à la formation de l'axe de façon équivalente. On notera que la variable PIB/habitant, qui on le rappelle était la plus corrélée avec le bonheur, est celle qui contribue le moins parmi les variables ayant un impact relativement important sur l'axe 1, cette dernière étant légèrement inférieure à notre seuil de 10% (Figure 3.4). Nous la garderons parmi nos variables dites contributives, puisqu'elle n'est que très légèrement en dessous du seuil. De plus, on remarque que cette variable contribue de façon très importante à l'axe 3. Les seules variables qui participent moins à l'axe 1 sont justement celles qui participent à la formation de l'axe 2, à savoir les libertés d'expressions, d'associations et des presses. Ces dernières contribuent à peu près toutes de la même façon à la formation de l'axe 2, entre 26,58% et 29,05% (Figure 3.4) Enfin, on note une grande disparité au niveau de l'axe 3 avec le PIB/habitant qui, comme nous l'avons mentionné plus haut, participe à la formation de l'axe à hauteur de 57% (Figure 3.4), une valeur bien plus élevée que le poids de la deuxième variable la plus contributive, la liberté d'association, qui n'est que de 12,97% (Figure 3.4).

Les cosinus carrés

FIGURE 3.5 – Cosinus carrés de nos variables



Nous pouvons désormais passer à l'interprétation des cosinus carrés, ces derniers servent à décrire la qualité de projection des variables vis-à-vis des axes. Cette qualité de représentation va de 0 à 1, et permet également d'identifier la proximité de la variable avec les axes du nuage. Une variable proche d'un axe est une variable au cosinus carré élevé (proche de 1), autrement dit, une variable avec une qualité de représentation (par rapport à l'axe) élevée. À l'inverse, une variable loin d'un axe est une variable au cosinus carré faible (proche de 0), autrement dit, une variable avec une qualité de représentation (par rapport à l'axe concerné) faible.

Encore une fois, nous pouvons utiliser un seuil qui nous permettrait de choisir quelles variables sont les plus pertinentes. Ici il s'agira de la moyenne des cosinus carrés. Une variable dépassant ce seuil sera une variable bien représentée par l'axe. Pour l'axe 1, hormis les va-

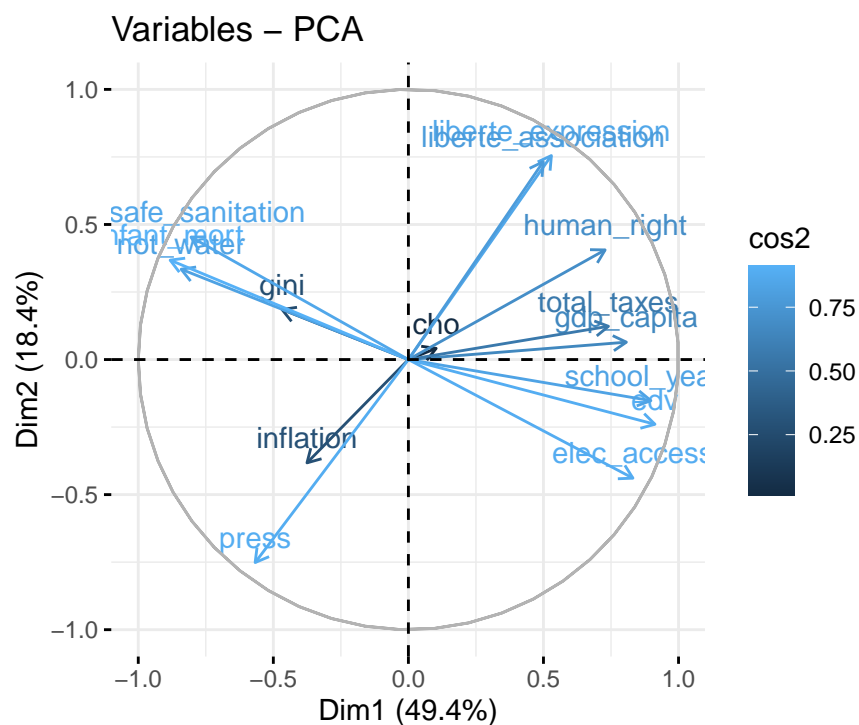
riables de libertés, toutes les variables peuvent être considérées comme bien représentées par l'axe 1 (Figure 3.5) (La variable PIB/habitant est quasiment égale à la moyenne des cosinus carrés à 0.01 près, elle sera donc considérée comme pertinente).

Concernant l'axe 2, les variables pertinentes sont les 3 variables restantes. Autrement dit *liberte_expression*, *liberte_association* et *press* qui sont toutes les trois supérieures à la moyenne des cosinus carrés de l'axe 2 (0,24), et donc bien représentées par ce dernier.

3.0.4 Cercle de corrélation

Comme nous l'avons déjà mentionné, le bonheur est multifactoriel. C'est pourquoi, nous avons dans un premier temps construit un modèle à partir de 15 variables. Notre objectif était d'incorporer le maximum de variables dans notre analyse afin qu'elle soit la plus complète possible.

FIGURE 3.6 – Cercle de corrélation : Modèle initial



La Figure 3.6 représente le cercle de corrélation de nos variables sur les axes 1 et 2. À première vue, le modèle est de bonne qualité. En effet, nous constatons une inertie cumulée de 68% avec les deux premiers axes. De plus, le troisième axe a une inertie non négligeable de 8%

TABLE 3.3 – Contributions de toutes nos variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
gdp_capita	0.65	0.00	0.05	0.01	0.03
school_years	0.81	0.02	0.00	0.00	0.00
elec_access	0.69	0.19	0.01	0.04	0.00
gini	0.22	0.03	0.33	0.17	0.00
not_water	0.71	0.11	0.03	0.01	0.01
edv	0.83	0.06	0.00	0.01	0.00
cho	0.01	0.00	0.75	0.11	0.06
total_taxes	0.55	0.02	0.00	0.17	0.04
inflation	0.14	0.15	0.01	0.32	0.32
infant_mort	0.78	0.14	0.01	0.01	0.00
human_right	0.53	0.16	0.03	0.04	0.02
liberte_expression	0.28	0.57	0.00	0.01	0.08
liberte_association	0.25	0.54	0.03	0.01	0.11
unsafe_sanitation	0.65	0.21	0.02	0.03	0.02
press	0.32	0.56	0.00	0.01	0.00

(Annexe Figure 6.1). Cependant, à ce moment-là, nous n'étions pas satisfait de la définition de nos axes. En effet, nous souhaitons un axe 2 plus identifiable pour nous aider à construire des variables latentes. C'est pourquoi, nous avons décidé de supprimer certaines variables en nous appuyant sur les tableaux de contributions et cosinus carrés.

Dans la première projection, beaucoup de variables sont projetées autour de l'axe 1 dans les cadrans I et IV (Figure 3.6). C'est pourquoi, nous avons d'abord décidé de supprimer les variables contribuant le moins à cette zone. À l'aide de la Table de contribution aux axes (Table 3.3), nous avons pris la décision de supprimer les variables : Droits Humains (*human_right*) et Taxes (*Total_Taxes*). Ces variables étaient les deux variables contribuant le moins aux trois axes (Figure 3.3). De plus, ces deux variables sont les variables ayant le cosinus carré le plus faible parmi toutes les variables présentes dans le cadran I (Figure 3.4).

La suppression de ces deux variables a légèrement fait augmenter l'inertie de notre deuxième axe (Annexe Figure 6.2). De plus, à l'aide des cosinus carré, nous nous sommes rendus compte que la position de nos variables entre les deux projections avait très peu changé (Annexe Figure 6.3). C'est pourquoi, nous avons décidé de les supprimer définitivement. Au départ, nous souhaitons également retirer les variables suivantes : Taux de Chômage (*cho*), Indice de Gini (*gini*) et Inflation. Effectivement, elles contribuent très faiblement aux deux premiers axes. Le taux de chômage par exemple, a une contribution à l'axe 1 de 0.14 (Table 3.3). Cependant, en regardant le tableau des contributions, on peut apercevoir que ces variables contribuent

TABLE 3.4 – Cosinus carrés de toutes nos variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
gdp_capita	8.78	0.15	4.06	1.53	4.04
school_years	10.87	0.84	0.05	0.52	0.09
elec_access	9.32	6.98	0.53	3.98	0.09
gini	2.95	1.24	25.41	18.13	0.07
not_water	9.54	4.05	2.66	0.55	1.79
edv	11.24	2.08	0.38	1.43	0.49
cho	0.14	0.06	58.64	11.05	9.14
total_taxes	7.40	0.55	0.27	17.84	6.05
inflation	1.90	5.29	1.17	33.73	45.65
infant_mort	10.49	4.90	0.53	1.05	0.52
human_right	7.15	5.96	2.11	3.94	3.15
liberte_expression	3.78	20.63	0.20	1.03	11.23
liberte_association	3.36	19.40	2.32	1.39	15.34
unsafe_sanitation	8.74	7.44	1.67	3.15	2.31
press	4.34	20.43	0.00	0.67	0.05

énormément aux axes trois et quatre (Figure 3.3).

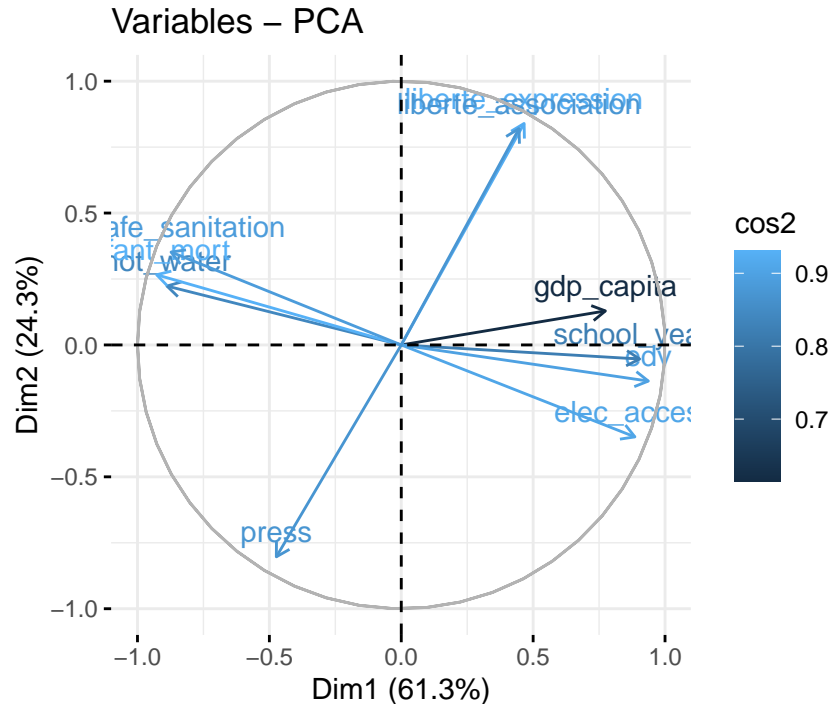
Comme nous n'étions pas encore satisfait de la définition de nos axes, nous avons tenté de retirer ces trois variables. En retirant ces trois variables, l'inertie cumulée des trois axes a fortement augmenté passant de $\approx 80\%$ (Annexe Table 6.4) à $\approx 91\%$ (Annexe Table 6.5). De la même manière que nous l'avons vu plus tôt, la position de nos variables a très peu changé (Annexe Table 6.3). Le fait que la position de nos autres variables n'ait quasiment pas changé et que l'inertie de notre modèle ait augmenté nous a confortés dans le choix de retirer ces trois variables.

TABLE 3.5 – Tableau récapitulatif : Modèle initial

Signe Axe	Positif	Negatif
Axe 1	Espérance de vie Nombre d'années d'études PIB/habitant Accès à l'électricité Droit humains Taxes	Mortalité infantile Accès à l'eau Mauvaise infrastructures sanitaires
Axe 2	Liberté d'expression Liberté d'association	Liberté de la presse Accès à l'électricité Inflation
Axe 3	Taux de chômage Indice de Gini	

Ainsi, nous obtenons enfin notre modèle contenant les dix variables dont nous avons parlé dans le Chapitre 2 économique. Le cercle de corrélation (Figure 3.7) synthétise graphiquement les éléments vus dans les parties précédentes, à savoir la contribution aux axes et les cosinus carrés. En ne gardant que 10 variables, on observe une inertie cumulée de près de 85%. Cette inertie nous signale que a priori, la qualité de l'information retenue dans nos deux axes est très bonne.

FIGURE 3.7 – Cercle de corrélation : Modèle final



Concernant l'axe 3, le cercle de corrélation nous indique une inertie d'environ 6 % (Annexe Figure 6.5). Cette faible inertie s'explique par le fait qu'une grande partie de la variance est déjà concentrée dans les deux premiers axes. Étant donné son inertie relativement faible, nous décidons de ne pas prendre en considération l'axe 3 pour le moment dans notre analyse. Le long de notre axe 1, on observe deux groupes bien distincts. D'un côté, nous avons le PIB/habitant, le nombre d'années d'études, l'espérance de vie et l'accès à l'électricité qui contribuent positivement à l'axe 1. De l'autre, nous avons la mortalité infantile, l'accès à l'eau et l'exposition à de mauvaises infrastructures sanitaires qui contribuent négativement à l'axe 1. Pour l'axe 2, nous retrouvons uniquement les variables liées à la liberté. Nous pouvons remarquer que la liberté de la presse est à l'opposé des libertés d'expression et d'association, elle contribue donc négativement à la construction de l'axe 2. Cela s'explique par la façon dont l'indicateur est construit, en effet plus ce dernier est grand, moins la liberté associée est élevée. C'est donc tout à fait normal qu'elle soit quasiment symétrique avec les deux autres variables de liberté. On remarque que lorsqu'on se déplace le long de l'axe 1 : la mortalité infantile, les difficultés d'accès à l'eau ainsi que l'exposition à de mauvaises infrastructures sanitaires diminuent alors que le PIB/habitant, l'espérance de vie, l'accès à l'électricité ainsi que le nombre d'années d'études augmentent. Quand on se déplace le long de l'axe 2, la liberté de la presse, la

liberté d'association et d'expression augmentent. À partir de ces conclusions, il nous reste donc à définir deux variables latentes.

3.0.5 Définition des variables latentes

Nous allons maintenant définir les variables latentes via le tableau ci-dessous (Table 3.6) qui nous servira à représenter les regroupements et oppositions de ces dernières.

TABLE 3.6 – Tableau récapitulatif : Modèle final

Signe Axe	Positif	Négatif
Axe 1	Espérance de vie Nombre d'années d'études Accès à l'électricité PIB/habitant	Mortalité infantile Accès à l'eau Mauvaise infrastructures sanitaires
Axe 2	Liberté d'expression Liberté d'association	Liberté de la presse

On remarque une opposition sur l'axe 1. Du côté positif se placent les variables indiquant un **développement économique et social**. En effet, on observe un regroupement entre le PIB/Habitant, le nombre d'années d'études, l'espérance de vie et l'accès à l'électricité. On remarque que ces trois variables s'apparentent beaucoup à l'*Indice de Développement Humain* (IDH) qui prend en compte des variables similaires (le PIB par habitant, l'espérance de vie à la naissance et le niveau d'éducation des enfants de 17 ans et plus ²). Un regroupement de variables vient s'opposer au côté positif avec, l'exposition à de mauvaises infrastructures sanitaires, le taux de mortalité infantile, et le manque d'accès à l'eau. L'opposition entre ces deux groupes de variables nous conforte dans la définition de notre première variable latente, le long de l'axe 1, à savoir le **développement économique et social**. En effet, du côté positif de l'axe 1 nous pourrions retrouver les pays développés économiquement et socialement, dont le PIB et l'IDH sont élevés. À l'inverse, du côté négatif de l'axe 1 nous pourrions retrouver les pays peu ou sous-développés dont l'IDH et le PIB sont faibles. Autrement dit, en longeant l'axe 1 (de gauche à droite) nous passons de pays peu ou sous-développés à des pays développés. L'opposition entre ces deux groupes de variables nous conforte dans la définition de notre première variable latente, le long de l'axe 1, à savoir le **développement économique**

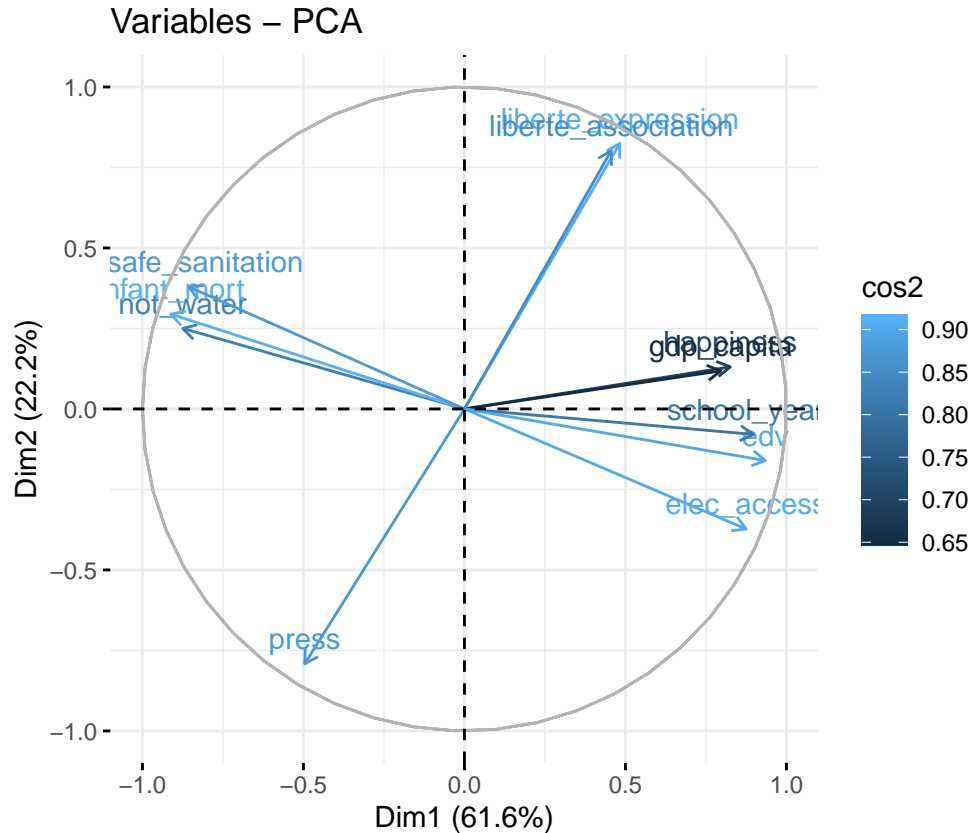
2. <https://hdr.undp.org/system/files/documents//2018humandevlopmentstatisticalupdatefrpdf.pdf>. pdf[consulté le 24/10/2022]

et social. En effet, du côté positif de l'axe 1 nous pourrions retrouver les pays développés économiquement et socialement, dont le PIB et l'IDH sont élevés. À l'inverse, du côté négatif de l'axe 1 nous pourrions retrouver les pays peu ou sous-développés dont l'IDH et le PIB sont faibles. Autrement dit, en longeant l'axe 1 (de gauche à droite) nous passons de pays peu ou sous-développés à des pays développés.

Concernant l'axe 2, il n'existe pas vraiment d'opposition dans la mesure où les 3 variables de libertés expriment globalement la même chose mais différemment. Stratégiquement, l'indicateur de la liberté de la presse aurait plutôt dû s'appeler « le non-respect de la liberté de la presse ». En effet, quand les indicateurs de la liberté d'expression et d'association sont élevés (un score proche de 100) , les libertés sont bien respectées sans restrictions ou oppressions du gouvernement. En revanche, l'indicateur de liberté de la presse fonctionne différemment puisque, plus il est élevé, plus la liberté de la presse est opprimée et restreinte par le gouvernement. C'est pour ces raisons qu'il n'existe pas d'opposition sur l'axe 2, puisqu'en longeant ce dernier (du bas vers le haut), le niveau des différentes libertés augmente. À l'inverse, en longeant l'axe 2 du haut vers le bas, le niveau des différentes liberté diminue. Un pays situé du côté négatif de l'axe 2 est un pays dont le niveau de liberté est restreint et/ou opprimé par le gouvernement, contrairement aux pays situés du côté positif de l'axe 2 où le niveau de liberté est globalement bien respecté. Ainsi, il convient de définir cette variable latente comme les **libertés individuelles et collectives**.

Désormais, nous souhaiterions avoir une idée de l'impact de nos variables latentes sur le niveau de bonheur d'un pays. Pour ce faire, nous pouvons projeter le niveau de bonheur sur un cercle de corrélation avec les variables explicatives qui forment nos deux axes. Nous obtenons le cercle suivant :

FIGURE 3.8 – Cercle de corrélation du modèle final avec la variable illustrative



On observe que notre variable illustrative (*happiness*) se situe dans le cadran I de la Figure 3.8. Cette dernière se superpose presque parfaitement avec le PIB/habitant. La projection confirme notre intuition. Effectivement, lorsque l'on se déplace le long de l'axe 1, de la gauche vers la droite, le **développement économique et social** s'améliore et le niveau de bonheur s'améliore également. Si ce constat est moins prononcé pour l'axe 2, nous observons quand même que plus on se déplace vers le haut, le long de l'axe 2, plus les **libertés individuelles et collectives** sont importantes et plus le niveau de bonheur est élevé. Nous étudierons plus en détail cette relation au chapitre suivant à l'aide de la régression linéaire multiple (§4.0.1).

3.0.6 Projection des individus sur le plan factoriel, interprétation

Nous pouvons désormais projeter les individus (pays) de notre modèle sur le plan factoriel, cela nous permettra de confirmer (ou non) le choix de nos variables latentes. Avant d'afficher la projection, nous regarderons les coordonnées, les contributions puis les cosinus carrés des différents pays.

Coordonnées des individus

Dans un premier temps, nous regarderons les coordonnées des individus, elles nous permettront de voir dans quel cadran chaque individu se situe.

TABLE 3.7 – Coordonnées de nos dix premiers individus

	Pays	Dim.1	Dim.2
1	Albania	1.197	-0.289
2	Algeria	0.161	-1.853
3	Angola	-4.154	0.283
4	Armenia	1.096	-0.693
5	Australia	2.954	0.852
6	Austria	2.832	1.056
7	Bangladesh	-1.277	-1.441
8	Belarus	0.558	-3.029
9	Belgium	2.664	1.001
10	Benin	-3.582	2.041

Dans un souci de lisibilité, plutôt que de regarder les 114 différents pays, nous regarderons uniquement les 10 premiers, cela nous permettra de vérifier s’ils se situent effectivement dans les cadrans attendus. Dans ce tableau (Table 3.7), nous pouvons voir les coordonnées des 10 premiers individus en fonction des 5 dimensions. Nous nous intéresserons uniquement aux dimensions 1 et 2.

Concernant l’axe 1, plus les valeurs positives sont élevées, plus le pays se situe loin du barycentre, du côté positif de l’axe 1 (tout à droite), autrement dit, plus le pays est sujet à un **développement économique et social** important. L’inverse est donc vrai, une valeur négativement importante représente un pays loin du barycentre du côté négatif de l’axe 1 (tout à gauche) ce qui représente un pays au **développement économique et social** faible. On voit que l’Australie, l’Autriche, et la Belgique ont les valeurs les plus importantes parmi nos 10 pays représentés dans le tableau, ce qui signifie qu’ils se situent loin du barycentre du côté positif de l’axe 1 (tout à droite). Ce sont effectivement tous des pays développés, autant sur le plan économique que social. On peut noter, que plus on longe l’axe 1 vers la droite (en s’éloignant ainsi du barycentre), plus le **développement économique et social** du pays est important. L’Australie est donc censée être située plus à droite le long de l’axe 1 que la Belgique et l’Autriche.

Concernant l'axe 2, plus les valeurs positives sont élevées, plus le pays se situe loin du barycentre du côté positif de l'axe 2 (tout en haut du graphique), autrement dit, plus le gouvernement du pays respecte et ne restreint pas les libertés collectives et individuelles. À l'inverse, plus la valeur négative est élevée, plus le pays se situe loin du barycentre du côté négatif de l'axe 2 (tout en bas du graphique), autrement dit, plus le gouvernement du pays restreint les **libertés individuelles et collectives**. On voit que la Biélorussie, l'Algérie et le Bangladesh ont les valeurs les plus négativement importantes de notre tableau. Il est bon de rappeler que ce tableau représente uniquement les 10 premiers pays (par ordre alphabétique) de notre base de données. Cela étant dit, d'après ces valeurs les 3 pays cités précédemment devront se situer en bas de notre axe 2, car ce sont des pays où les libertés collectives et individuelles ne sont pas ou peu respectées.

Pour connaître la position exacte de chaque pays, il faut prendre en compte les coordonnées sur l'axe 1 et 2. Parmi les pays interprétés précédemment, nous distinguons deux groupes. Ceux aux coordonnées positives autant sur l'axe 1 que l'axe 2 (Australie, Autriche et Belgique), qui devraient donc se situer dans le cadran 1. Puis, ceux aux coordonnées positives sur l'axe 1 mais négatives sur l'axe 2 (Biélorussie, Algérie). Ces pays devraient donc se situer dans le cadran 4. Concernant le Bangladesh, ses coordonnées sont négatives autant sur l'axe 1 que l'axe 2, ce pays devrait donc se situer dans le cadran 3. Nous vérifierons un peu plus loin si cette hypothèse est vraie.

Cet échantillon de notre base de données sert uniquement d'exemple, puisqu'il ne prend que les dix premiers pays par ordre alphabétique. Pour avoir une analyse complète il faudrait utiliser le tableau avec les 114 pays différents.

Contributions des individus

Dans un second temps, nous regarderons les contributions de chaque pays (Table 3.8). Les contributions illustrent le poids de chaque pays sur chaque axe, c'est-à-dire à quel degré il participe à la formation de l'axe. Une valeur élevée représente un pays qui participe de manière importante à la formation de l'axe. Encore une fois, nous nous focalisons sur l'axe 1 et 2. Les pays les plus contributifs sont ceux dépassant la contribution moyenne, soit :

$$\frac{100}{\text{Nombreindividus}} = \frac{100}{114} = 0.87$$

Concernant l'axe 1, de nombreux pays dépassent ce seuil, notamment l'Algérie, l'Autriche, la

TABLE 3.8 – Contributions de nos dix premiers individus

	Pays	Dim.1	Dim.2
1	Albania	0.205	0.030
2	Algeria	0.004	1.241
3	Angola	2.470	0.029
4	Armenia	0.172	0.174
5	Australia	1.249	0.262
6	Austria	1.148	0.403
7	Bangladesh	0.233	0.751
8	Belarus	0.045	3.316
9	Belgium	1.016	0.362
10	Benin	1.837	1.506

Belgique et le Bénin. Concernant l'axe 2, quelques pays dépassent ce seuil notamment l'Algérie, la Biélorussie et le Bénin. Les pays cités précédemment participent donc de manière importante à la formation de leurs axes respectifs.

Cosinus carrés des individus

TABLE 3.9 – Cosinus carrés de nos dix premiers individus

	Pays	Dim.1	Dim.2
1	Albania	0.606	0.035
2	Algeria	0.006	0.773
3	Angola	0.925	0.004
4	Armenia	0.444	0.178
5	Australia	0.855	0.071
6	Austria	0.827	0.115
7	Bangladesh	0.317	0.404
8	Belarus	0.030	0.887
9	Belgium	0.849	0.120
10	Benin	0.734	0.238

Enfin, nous pouvons à présent analyser les cosinus carrés des individus (Table 3.9). Ces derniers nous indiquent la qualité de représentation des pays sur chaque axe. Nous pouvons utiliser un seuil qui nous permettrait de choisir quelles variables ont la meilleure qualité de représentation. Encore une fois, ce seuil sera la moyenne des cosinus carrés, qui est de 0.55

pour l'axe 1 et de 0.21 pour l'axe 2. Les variables dépassant ce seuil seront des variables dont la qualité de représentation sur l'axe concernée est élevée.

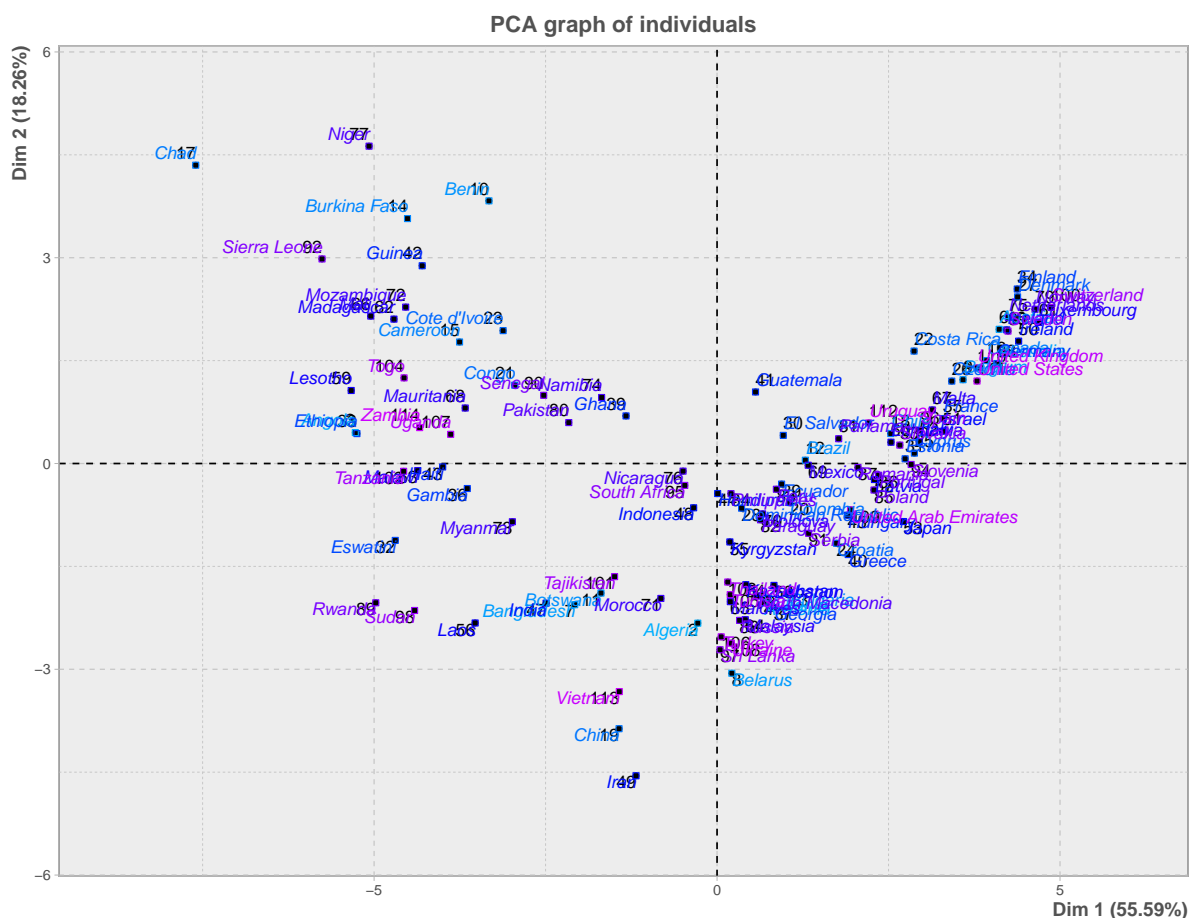
Concernant l'axe 1, de nombreux pays dépassent ce seuil de 0.55 notamment l'Albanie, l'Angola, l'Australie, l'Autriche, la Belgique et le Bénin. Ces pays ont donc une bonne qualité de représentation sur l'axe 1. On peut tout de même noter que le pays le mieux représenté sur l'axe 1, l'Angola, est très mal représenté sur l'axe 2.

Concernant l'axe 2, quelques pays dépassent ce seuil de 0.21 notamment l'Algérie, le Bangladesh, la Biélorussie et le Bénin. Ces pays ont donc une bonne qualité de représentation sur l'axe 2. Encore une fois, on peut noter que le pays le mieux représenté sur l'axe 2, Biélorussie, est très mal représenté sur l'axe 1.

Projection des individus sur les axes 1 et 2

Après avoir analysé les coordonnées, contributions et cosinus carrés des pays, nous pouvons projeter ces derniers sur les différents axes.

FIGURE 3.9 – Projection de tous nos individus sur les axes 1 et 2 : Noms des pays



Dans un premier temps, nous pouvons présenter un graphique de projection avec les noms des différents pays plutôt que des numéros d'observations (Figure 3.9). Ce dernier n'étant pas très lisible, il nous permet tout de même d'avoir une idée globale de la position de chaque pays. Cela nous permet de regarder les pays proches graphiquement, et voir, si oui ou non ils sont globalement assez similaires. Ce graphique va également nous permettre de valider le choix de nos variables latentes. Même si la Chine est le pays qui dispose de la classe moyenne la plus importante au monde, le niveau de développement au sein même du pays est encore très inégal. C'est pourquoi, tout comme le Vietnam, elle dispose d'un PIB/habitant plutôt faible ³¹. Concernant les libertés d'expression, d'association et de la presse, elles sont peu respectées (autrement dit le gouvernement oppose des restrictions / oppressions à ces libertés.)

31. <https://www.sciencedirect.com/science/article/pii/S2213020915000518>

Individuals – PCA

Dim2 (24.3%)

Dim1 (61.3%)

cos2

0.75

0.50

0.25

The plot shows a clear separation between two groups of individuals along the Dim1 axis. The group with higher cos2 values (orange/red) is clustered on the left (negative Dim1), while the group with lower cos2 values (green/blue) is clustered on the right (positive Dim1). The y-axis (Dim2) shows some overlap between the groups. A vertical dashed line is drawn at Dim1 = 0.0, and a horizontal dashed line is drawn at Dim2 = 0.0.

Cette projection va nous permettre de valider le choix de nos variables latentes. Concernant les coordonnées, nous avons déterminé la position des 10 premiers pays du tableau dans les différents cadrans. Nous pouvons désormais vérifier directement sur la projection. Conformément à ce que nous avons observé plus tôt, on voit bien que l'Australie, l'Autriche et la Belgique que nous avons placés dans le cadran 1 se situent effectivement dans ce dernier (en haut à droite), de plus, la Biélorussie et l'Algérie se situent bien dans le cadran 4 (en bas à droite) et enfin, le Bangladesh se situe bien dans le cadran 3 (en bas à gauche).

Concernant les contributions, si on regarde dans notre tableau des contributions (Table 3.8), le pays contribuant le moins (axe 1 et axe 2), à savoir l'Albanie, est assez proche du barycentre puisqu'il participe peu à la formation de l'axe 1 et 2. Nous pouvons également regarder les valeurs du Guatemala (41), qui se positionne quasi-parfaitement au barycentre de notre projection. Les valeurs de ce pays pour l'axe 1 et l'axe 2 sont de 0 (Annexe Table 6.6). Autrement dit, le Guatemala ne contribue absolument pas ni à la formation de l'axe 1 ni à la formation de l'axe 2, il est donc logique qu'il se situe au barycentre de notre projection. Concernant les cosinus carrés, nous avons dit que l'Albanie, l'Angola, l'Australie, l'Autriche, la Belgique et le Bénin avaient une bonne qualité de représentation sur l'axe 1, ce qui est le cas puisqu'ils sont relativement proche de l'axe 1. Nous avons également souligné que sur le pays le plus contributif à l'axe 1, l'Angola, était très mal représenté sur l'axe 2, nous pouvons effectivement voir qu'il est très proche de l'axe 1 et très éloigné de l'axe 2 (Figure 3.10).

Pour illustrer le rôle des cosinus carrés, nous pouvons nous intéresser à une valeur extrême : L'Iran (41). Si on regarde les coordonnées de ce pays sur l'axe 1 et 2 (Annexe Table 6.7), on remarque que sa qualité de représentation est quasi-nulle sur l'axe 1 mais très importante sur l'axe 2. Nous pouvons le confirmer graphiquement puisque ce pays est très proche de l'axe 2 mais très éloigné de l'axe 1 (Figure 3.10).

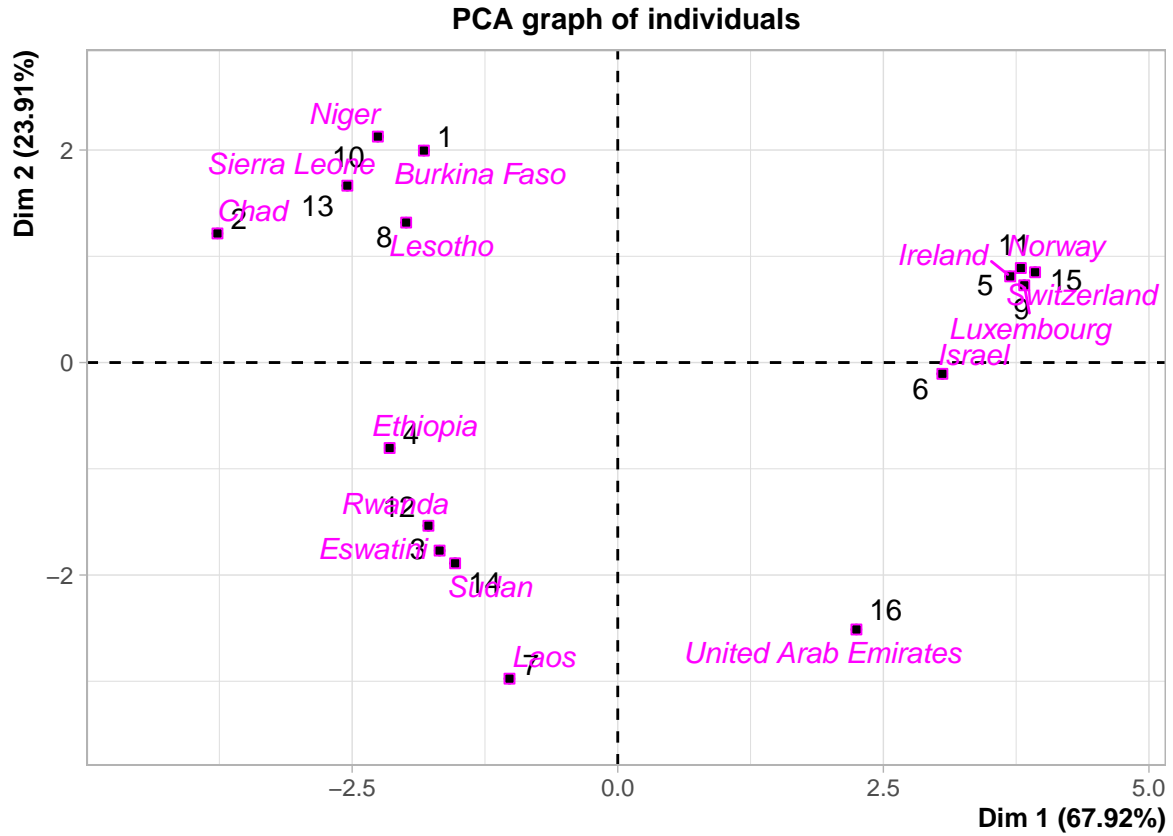
Pour confirmer pleinement le choix de nos variables latentes, il serait intéressant d'opposer les individus dits « extrêmes » dans un tableau. Nous choisissons donc des groupes de valeurs extrêmes pour chaque cadran de notre représentation, comme dans le graphique ci-dessous (Figure 3.11).

FIGURE 3.11 – Projections des individus extrêmes : Numéros



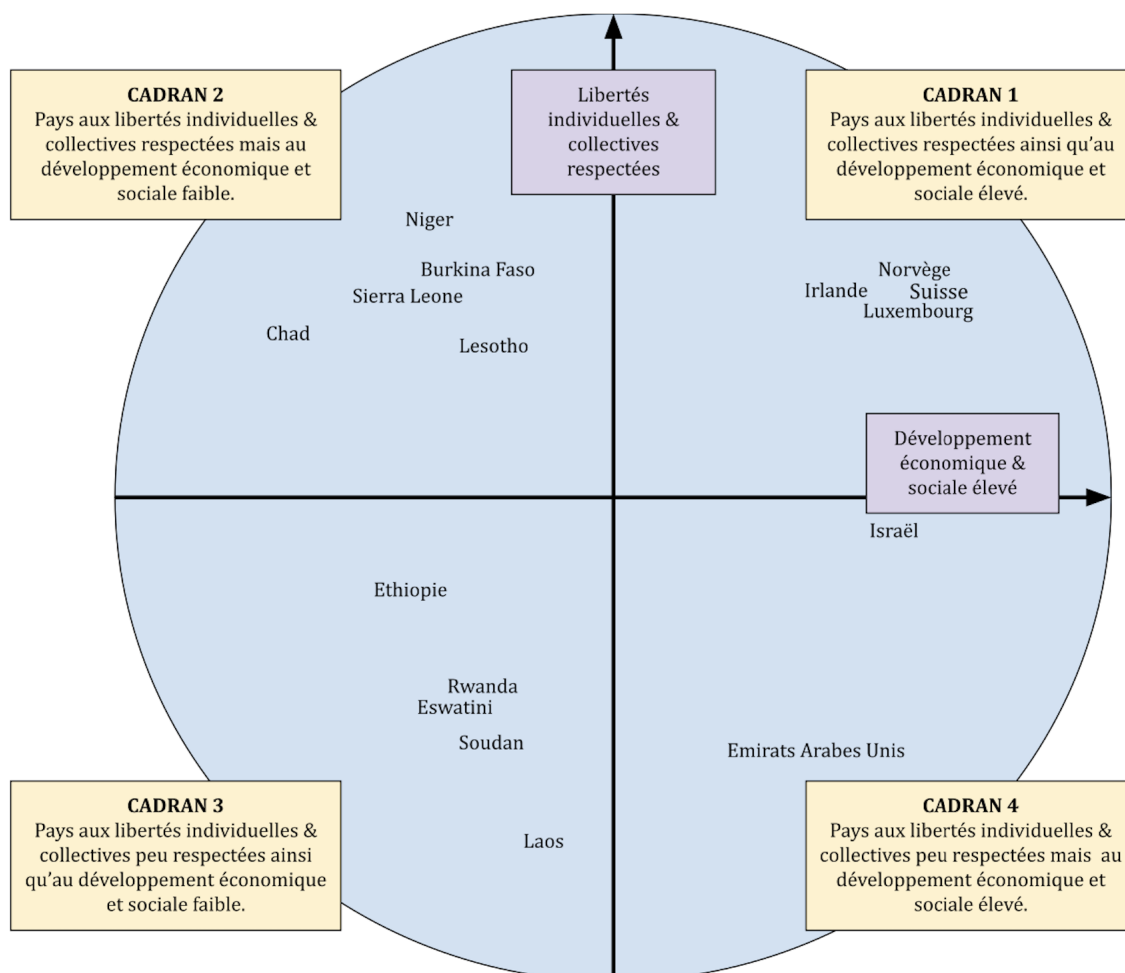
Nous pouvons représenter le choix de ces valeurs extrêmes dans un graphique qui laisse apparaître le nom des différents pays (Figure 3.12).

FIGURE 3.12 – Projections des individus extrêmes : Noms des pays



Il est important de noter que dans le graphique ci-dessus, les numéros apparaissant sur chaque nom de pays ne correspondent pas à ceux de notre base de données, nous n’y prêtons donc aucune attention. Nous pouvons dès à présent opposer ces différents pays dans un graphique avec différents cadrons.

FIGURE 3.13 – Graphique récapitulatif : Individus



En plaçant les différents pays dits « extrêmes » dans les différents cadrans, nous pouvons dégager 4 groupes distincts (Figure 3.13).

Le cadran n°1 représente les pays au **développement économique et social** élevé dont les libertés collectives et individuelles sont respectées. Dans ce groupe figure des pays tels que la Norvège, la Suisse, l'Irlande et le Luxembourg. Économiquement, ces pays sont effectivement développés, avec un PIB/Habitant très important. Ils sont également développés socialement puisqu'ils figurent parmi les pays où le nombre d'années d'étude est le plus élevé. La mortalité infantile y est très faible et l'accès à l'eau et l'électricité est parfait ou quasi-parfait. Ces pays sont très respectueux des libertés d'expression, d'association et de la presse. On peut également noter que l'intégralité des pays présents dans ce cadran sont des pays d'Europe.

Dans le cadran n°2 les **libertés individuelles et collectives** sont également respectées, cependant le **développement économique et social** y est faible. Dans ce groupe figure des pays tels que le Niger, le Burkina Faso, Sierra Leone, le Tchad, et enfin Lesotho. Ici aussi, les pays du même cadran semble avoir approximativement la même position géographique, autrement dit ce sont uniquement des pays du continent Africain (Afrique de l'Ouest, Central et Australe). On peut noter que le Lesotho fait partie des pays où le taux de prévalence du VIH est le plus élevé, à hauteur de 27% de la population.

Dans le cadran n°3 figurent les pays dont le **développement économique et social** ainsi que faible ainsi que les **libertés individuelles et collectives** sont faibles. Dans ce groupe figurent des pays d'Afrique mais également d'Asie. Ces pays figurent parmi les pays considérés comme sous-développés, avec des PIB/Habitants parmi les plus faibles, un accès à l'eau et à l'électricité très restreint, une espérance de vie faible et un gouvernement qui oppresse / restreint les différentes **libertés individuelles et collectives**.

Enfin, dans le cadran n°4 figurent les pays dont les **libertés individuelles et collectives** sont restreintes et oppressées par le gouvernement mais dont le **développement économique et social** y est important. Il n'est absolument pas surprenant d'y retrouver les Emirats Arabe Unis, seul pays figurant parmi les cas « extrêmes » du cadran n°4. En effet, il n'existe pas beaucoup de pays développés économiquement et socialement avec des **libertés individuelles et collectives** très restreintes / oppressées. *In fine*, les pays apparaissant dans chaque cadran semblent confirmer le choix de nos axes.

Chapitre 4

Corrélation entre la variable illustrative et les variables latentes

4.0.1 Régression linéaire multiple

Dans ce dernier chapitre, nous allons nous aider des méthodes économétriques afin d'évaluer la qualité et l'efficacité de nos deux variables latentes. Dans un premier temps, nous allons nous aider d'une régression linéaire multiple (RLM) dans le but d'estimer la qualité de prédiction de nos deux axes. Pour ce faire, on exprimera l'équation suivante :

$$happiness = \beta_0 + \beta_1 Dvlp_eco_et_social + \beta_2 Lib_ind_et_coll + \varepsilon_i \quad (4.1)$$

Dans cette expression, *happiness*, notre variable à expliquer, représente le niveau de bonheur par pays. Nous avons *Dvlp_eco_et_social* et *Lib_ind_et_coll* qui représentent nos deux variables explicatives. Ensuite, nous avons β_1 et β_2 représentant les deux coefficients associés aux deux variables latentes. Enfin, ε_i représentant le terme d'erreur de notre régression. Dans un premier temps, nous estimerons une régression sur les axes 1 et 2.

Régression linéaire multiple sur les axes 1 et 2

TABLE 4.1 – RLM sur Axes 1 et 2

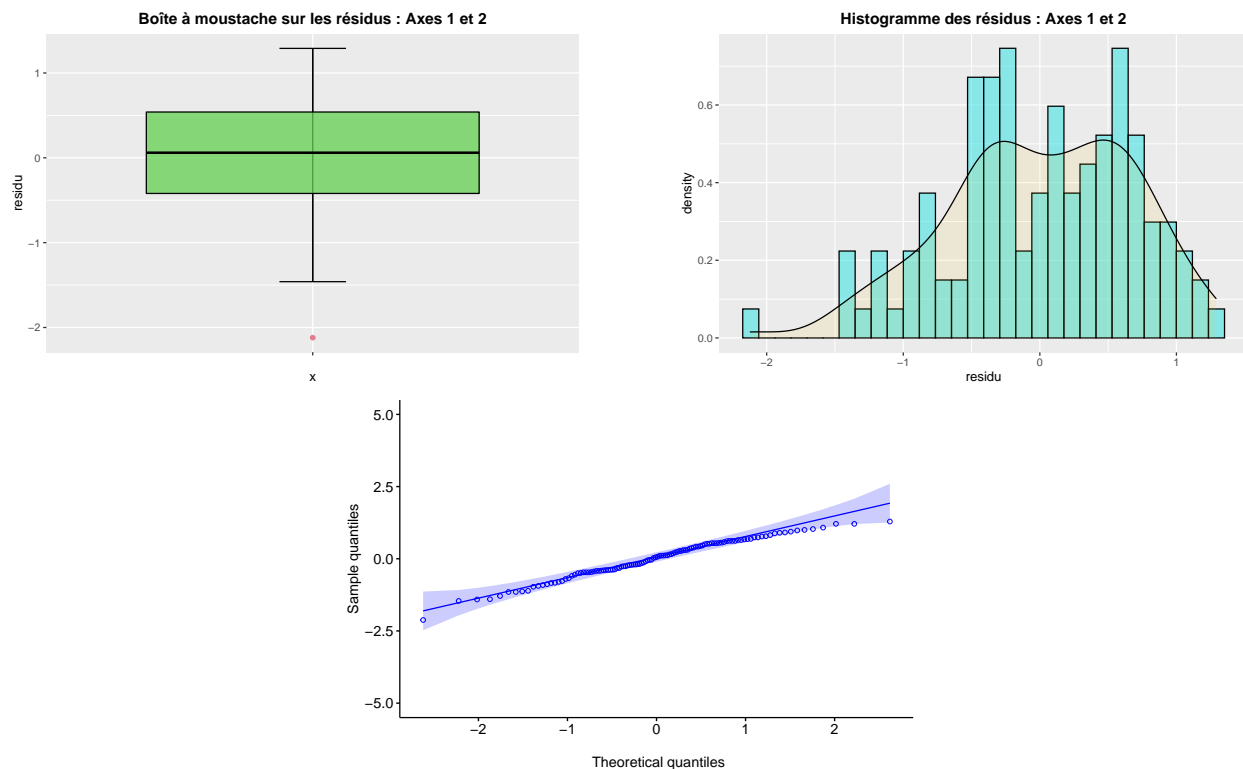
	<i>Dependent variable :</i>
	happiness
Dvlp_economique_et_social	0.346*** (0.026)
Libertes_ind_et_coll	0.093** (0.041)
Constant	5.579*** (0.064)
Observations	114
R ²	0.623
Adjusted R ²	0.616
Residual Std. Error	0.685 (df = 111)
F Statistic	91.760*** (df = 2; 111)
<i>Note :</i>	*p<0.1; **p<0.05; ***p<0.01


A l'aide du test de Fisher présent dans la Table 4.1, nous remarquons qu'au seuil de 1%, notre modèle est globalement significatif, au moins un des coefficient est non nul. Ensuite, nous regardons le test de significativité des coefficients estimés de Student. La p-value du test est inférieure à 0.05 (Annexe Table 6.8) et nous indique donc que notre modèle à un intérêt. On se penche maintenant sur la qualité d'ajustement de notre régression. Le R^2 observé est de 0.623 (Table 4.1), le modèle est donc de bonne qualité. Cela signifie que 62,3% de la variance de notre variable *happiness* est expliquée par nos deux variables explicatives. Ensuite, nous pouvons analyser les coefficients de nos deux variables explicatives. À l'aide du tableau ci-dessus, on peut voir qu'au seuil de 5%, nos deux variables latentes ont un impact significatif sur le niveau de bonheur d'un pays. Néanmoins, avant d'interpréter les résultats, nous avons besoin d'effectuer plusieurs tests afin de valider la significativité de nos coefficients.

Nous pouvons dès à présent analyser les résidus via trois graphiques différents. Dans un premier temps, l'histogramme nous montre une légère asymétrie vers le droite, ainsi qu'une forte

concentration entre -0.5 et 0 puis en +0.5. La normalité de la distribution n'est pas flagrante sur l'histogramme, la distribution ne semble pas parfaitement symétrique. Nous compléterons avec un test de Shapiro par la suite. Nous pouvons également analyser la boîte à moustache des résidus (Figure 4.1). Cette dernière nous montre une potentielle valeur atypique aux alentours de -2. Nous pouvons ensuite regarder le diagramme Q-Q qui nous permet de voir que une dispersion des résidus par rapport à la droite (Figure 4.1). Néanmoins la distribution observée semble relativement proche de la distribution théorique.

FIGURE 4.1 – Graphiques des résidus des axes 1 et 2 : Boîte à moustaches, Histogramme, Diagramme quantile-quantile



Cependant, pour s'assurer que nos résidus suivent une loi normale, nous allons effectuer le test de Shapiro-Wilk. Si la p-value du test est supérieur à 0.05, nous ne rejetons pas l'hypothèse H_0 de normalité des résidus. En effectuant le test sur le logiciel , on obtient une p-value de 0.09 (Table 4.2). Étant supérieur à 0.05, nous pouvons dire qu'au seuil de 5% nous ne rejetons donc pas l'hypothèse de normalité des résidus.

Même si le test de Shapiro-Wilk nous indique que le résidus suivent une loi normale, nous pouvons compléter notre analyse avec le test de Kolmogorov-Smirnov. Ce test non-paramétrique peut être ici appliqué puisque nous avons plus de 50 individus, le non rejet de l'hypothèse

TABLE 4.2 – Test de Shapiro

Shapiro-Wilk normality test	
W	0.98
p-value	0.087

nulle H_0 signifie que les résidus suivent une loi normale au seuil de risque de 5%.

TABLE 4.3 – Test de Kolmogorov-Smirnov

Asymptotic one-sample Kolmogorov-Smirnov test	
D	0.068
p-value	0.67
alternative hypothesis	two-sided

Ici la p-value est largement supérieur à 0.05 (Table 4.3), autrement dit, on ne rejette pas H_0 ce qui signifie que les résidus suivent bel et bien une loi normale au seuil de risque de 5%.


Nous allons à présent vérifier si nos résidus sont homoscedastiques, autrement dit, si la variance des résidus de la régression est la même pour chaque observations. Pour vérifier cette hypothèse, nous faisons appel au test de Breusch-Pagan. La règle de décision pour le test de Breusch-Pagan est la suivante : Si la p-valeur du test est supérieur à 0.05, alors l'hypothèse H_0 d'homoscedasticité des résidus n'est pas rejetée. Grâce au logiciel , nous trouvons une p-valeur de 0.13 (Table 4.4). Nous pouvons donc dire qu'au seuil de 5%, l'hypothèse nulle H_0 n'est pas rejetée et la variance de nos résidus est constante.

TABLE 4.4 – Test de Breush-Pagan

Studentized Breush-Pagan Test	
BP	4.1
p-value	0.13

Les hypothèses de normalité des résidus et d'homoscedasticité étant vérifiées, nous allons à présent chercher à savoir si la forme fonctionnelle de notre modèle est la bonne. Une meilleure forme fonctionnelle pourrait améliorer la significativité globale de notre modèle. Pour ce faire,

nous utiliserons le test de Ramsey (RESET test). Ce test nous permettra de vérifier que notre modèle est bien linéaire. Si la p-value du test est supérieure à 0.05, nous ne rejetons pas la forme linéaire de notre modèle au seuil de 5%.

TABLE 4.5 – Test de Ramsey

RESET Test	
RESET	14.75
df1	2
df2	109
p-value	2.142e-06

D’après la Table 4.5, la p-value du test est largement inférieure à 0.05. Ainsi, la linéarité de notre modèle ne peut pas être vérifiée.

Enfin, pour confirmer nos tests (et plus particulièrement le résultat du test de Ramsey), nous pouvons utiliser le test de Global Validation of Linear Model Assumptions (GVLMA).

TABLE 4.6 – Global Validation of Linear Model Assumptions

Test	Value	p-value	Assumptions
Global Stat	28.8	8.535e-06	NOT satisfied
Skewness	3.1	0.077	Acceptable
Kurtosis	0.18	0.67	Acceptable
Link Function	22.75	1.842e-06	NOT satisfied
Heteroscedasticity	2.7	0.098	Acceptable

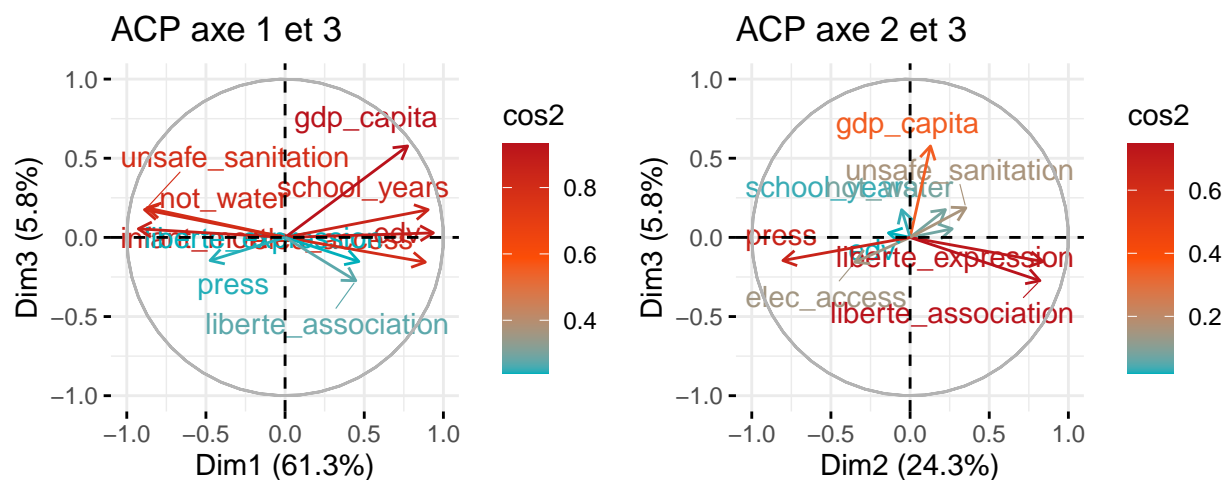
D’après la Table 4.6, notre modèle respecte bien les hypothèses de normalité des résidus (*Skewness* et *Kurtosis*) ainsi que celle d’homoscédasticité. Le terme *Link Function* nous indique que nos variables sont discrètes. Nous verrons donc dans la dernière partie si une régression logistique binaire peut nous permettre d’obtenir une meilleure estimation pour notre modèle. Le fait que *Global Stat* ne soit pas validé nous indique que la relation entre notre variable illustrative et nos variable latentes n’est pas linéaire, cela confirme les résultats du test de Ramsey.

Ainsi, si la forme de notre modèle avait été linéaire, nous aurions pu interpréter les résultats comme suit : Toutes choses égales par ailleurs, quand le **développement économique et social** augmente de 1, le niveau de bonheur augmente de 0.346. Quand ce sont les **libertés individuelles et collectives** qui augmentent de 1, le niveau de bonheur augmente de 0,093.

Régression Linéaire multiple sur les Axes 1/3 et 2/3

Dans l'analyse réalisée au Chapitre 3 §3.0.4, nous avons décidé de prendre en compte uniquement les axes 1 et 2. En effet, ces deux axes offrent déjà une contribution très significative. Cependant, pour confirmer notre intuition, nous avons pris la décision de tout de même effectuer des régressions sur l'axe 3. Si notre intuition était la bonne, ces régressions devraient être de moins bonnes qualité que la première. Pour que l'analyse soit moins redondante et plus agréable, nous allons réaliser l'analyse de régression de axes 1&3 et 2&3 en même temps. Nous afficherons les graphiques résultant des deux régressions à côté pour pouvoir les comparer, puis nous comparerons les résultats des tests en les affichant dans un tableau. Cela nous permettra une meilleure lisibilité, pour comparer les deux régressions. Nous incluons également les résultats de la première régression dans ces derniers.

FIGURE 4.2 – Projection des cercles des cercles de corrélations supplémentaires



Voici les représentations graphiques de nos ACP entre les axes 1&3 ainsi que 2&3 (Figure 4.2). Nous utilisons les mêmes graphiques et tests que précédemment à savoir Histogramme,

boîte à moustache, diagramme Q-Q des résidus, ainsi que les tests de Shapiro, Breusch-Pagan, Kolmogorov et GVLMA.

Les boîtes à moustache (Annexe Figure 6.3) nous montrent des potentiels valeurs atypiques, notamment pour la régression entre l'axe 1 et 3.

Les deux histogrammes (Annexe Figure 6.4) nous indiquent une légère asymétrie vers la gauche pour les résidus entre l'axe 1&3, et à l'inverse, une légère asymétrie vers la droite pour les résidus entre l'axe 2&3. Encore une fois, si normalité de la distribution il y a, elle n'est pas flagrante graphiquement. C'est pourquoi nous allons le vérifier plus loin via un test de Shapiro.

Les diagrammes Q-Q (Annexe Figure 6.5) nous indiquent que la distribution observée semble relativement proche de la distribution théorique pour l'axe 1&3, c'est un peu moins le cas pour l'axe 2&3.

TABLE 4.7 – Tableau comparatif

Modèle \ Test	R^2_{adj}	Student	Shapiro	Kolmogorov-Smirnov	Breusch-Pagan
1	0.61	$\beta_1 \rightarrow 2e-16$ $\beta_2 \rightarrow 0.0266$	0.08 Validé	0.67 Validé	0.13 Validé
2	0.65	$\beta_1 \rightarrow 2e-16$ $\beta_2 \rightarrow 6.56e-05$	0.89 Validé	0.99 Validé	0.03 Pas validé
3	0.05	$\beta_1 \rightarrow 0.16$ $\beta_2 \rightarrow 0.013$	0.01 Pas validé	0.78 Validé	6.351e-06 Pas validé

La Table 4.7 nous résume les résultats des différents tests sur nos 3 modèles. On peut voir que le R^2 ajusté (R^2_{adj}) le plus élevé est celui du modèle 2 puisqu'il atteint 0,652. Autrement dit, le modèle 2 explique 65% de la variation de la variable à expliquer. Le R^2 ajusté du premier modèle est presque aussi élevé, avec une valeur de 61%. En revanche, le modèle 3 semble très peu significatif puisque que son R^2 ajusté atteint seulement 5%.

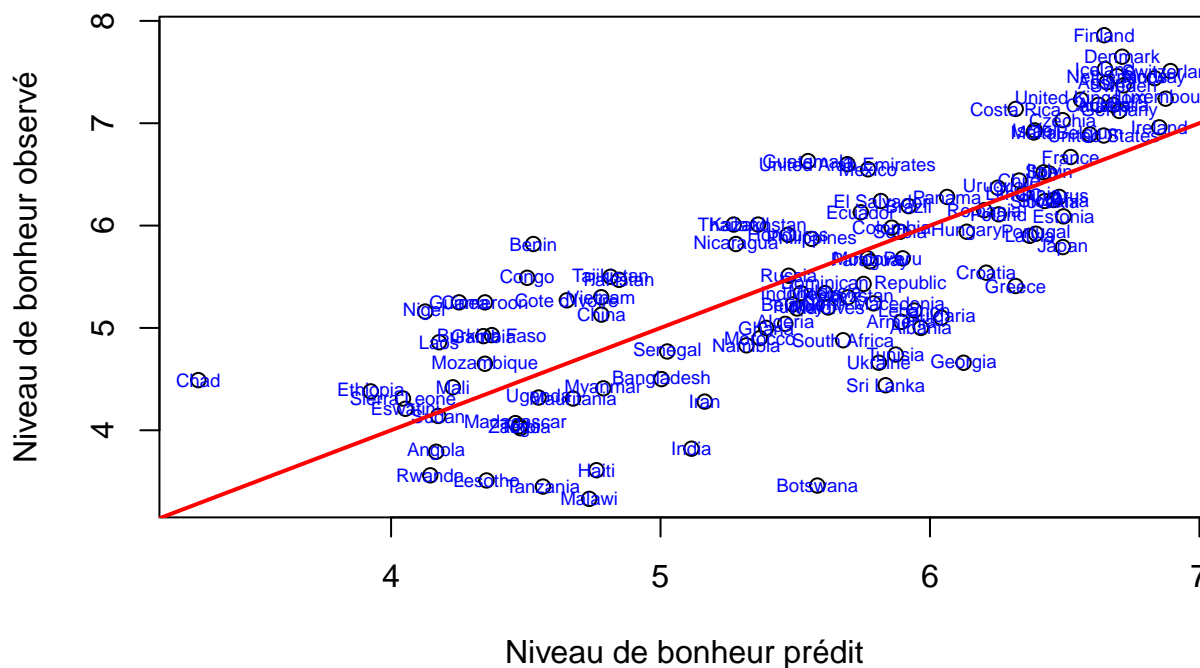
Les tests de normalité de Shapiro valide la l'hypothèse H0 indiquant une normalité des résidus pour les modèle 1 et 2, mais pas pour le modèle 3. On constate un résultat différent avec le test non-paramétrique de Kolmogorov Smirnov où les 3 modèles valident les hypothèses de normalité des résidus. Ce résultat n'est pas particulièrement surprenant quand on sait que les tests non-paramétriques sont moins puissants et plus restrictifs que les tests paramétriques. En ce qui concerne le test d'homoscédasticité de Breusch-Pagan, seul le modèle valide l'hypothèse H0 indiquant une homoscédasticité des résidus. Les modèles 2 et 3 présentent une homoscédasticité des résidus étant donné que leur p-value est inférieur à 0,05.

Au vu des résultats présentés dans la Table 4.7, nous pouvons conclure que le modèle 1 reste le meilleur, ce dernier présente un R^2 ajusté élevé, valide les deux tests de normalité ainsi que le test d'homoscédasticité des résidus. Cela nous conforte dans l'idée de ne pas prendre l'axe 3 dans notre analyse.

4.0.2 Relation entre la variable illustrative observée et prédite par la régression linéaire multiple

Au vu des résultats précédents, il semblerait que le modèle 1 soit le meilleur. C'est donc celui-ci que nous utiliserons pour faire une comparaisons entre les résultats observés et prédits.

FIGURE 4.3 – Projection de nos valeurs observées par rapport aux valeurs prédites



La Figure 4.3 permet de visualiser la différence entre le bonheur réellement observé au travers de notre base de données et les valeurs prédites par le modèle 1. Les pays situés proche de la ligne bleue issue de notre régression ont un niveau de bonheur presque parfaitement prédit par notre modèle. À l'inverse, les pays éloignés de notre droite de régression sont des

pays où le niveau de bonheur est moins bien prédit par le modèle. On remarque que des pays tels que : le Botswana, le Tchad, le Malawi ou encore l'Inde sont très éloignés de la droite bleue, notre modèle prédit assez mal leur niveau de bonheur réel. En revanche, les niveaux de bonheur des pays tels que la Russie, la Serbie, l'Espagne, l'Uruguay, la Lituanie, ou encore la France, sont très bien prédits par notre modèle. Leur niveau de bonheur réel est proche de celui prédit.

TABLE 4.8 – Statistiques descriptives du niveau de bonheur prédit entre l'axe 1 et 2

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.285	4.782	5.748	5.579	6.364	6.893

TABLE 4.9 – Statistiques descriptives du niveau de bonheur observé

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.330	4.838	5.495	5.579	6.303	7.860

Les Tables 4.8 et 4.9 nous permettent de comparer notre niveau de bonheur prédit et observé. Si dans l'ensemble, notre projection arrive bien à reproduire les variations de la variable illustrative, on remarque qu'elle a plus de mal à capter les individus du haut de la distribution. En effet, notre régression n'arrive pas à bien répliquer le niveau de bonheur des pays au-dessus du seuil de 7. Cette observation est potentiellement due au fait que la forme fonctionnelle de notre modèle ne soit pas linéaire. Nous allons dans la prochaine partie essayer de savoir si une autre forme fonctionnelle pourrait être plus adéquate.

4.0.3 Régression logistique binaire

Comme nous l'avons vu dans la §4.0.1, la forme fonctionnelle de notre modèle n'est pas linéaire. De plus, notre variable illustrative est quantitative discrète. Cet environnement fait qu'il est tout à fait pertinent pour nous d'utiliser la régression logistique binaire. Le but de la régression logistique binaire (RLB) est de nous permettre de définir un seuil à partir duquel un pays aurait ou n'aurait pas un certain niveau de bonheur. Pour pouvoir réaliser cette régression, nous allons créer une variable dichotomique prenant la valeur 1 si au-dessus d'un certain seuil de bonheur et 0 sinon. L'enjeu est donc de savoir quel seuil utiliser. Pour ce faire,

nous allons commencer par établir un seuil au niveau de la médiane de notre variable à expliquer. Nous estimerons donc le modèle suivant :

$$happiness = \beta_0 + \beta_1 Dvlp_eco_et_social + \beta_2 Lib_ind_et_coll + \varepsilon_i \quad (4.2)$$

Avec :

$$happiness = \begin{cases} 1 & \text{si } happiness > 5.5 \\ 0 & \text{sinon} \end{cases}$$

En rentrant cette expression dans le logiciel , nous obtenons les résultats suivants :

TABLE 4.10 – RLB au seuil médian de 5.5

	<i>Dependent variable :</i>
	happiness
Dvlp_economique_et_social	1.104*** (0.221)
Libertes_ind_et_coll	0.235 (0.205)
Constant	−0.395 (0.332)
Observations	114
Log Likelihood	−40.838
Akaike Inf. Crit.	87.675
<i>Note :</i> *p<0.1; **p<0.05; ***p<0.01	

En prenant la médiane comme seuil, nous remarquons que seul un des coefficient est significatif au seuil de 5% (Table 4.10). La constante β_0 et notre variable latente β_2 ne sont pas significatives.

Nous allons maintenant estimer notre modèle avec des seuils différents pour essayer d’avoir une régression logistique significative au seuil de 5%. En prenant le seuil moyen (Annexe Figure 6.9), ainsi que celui du troisième quartile (Annexe Figure 6.10), nous ne trouvons aucun résultat satisfaisant.

Nous avons donc dû chercher le seuil par tâtonnement. Au bout de quelques essais, nous avons réussi à trouver un seuil à partir duquel nos variables latentes étaient significatives. Ce seuil est également celui qui minimise le critère d'information d'Akaike (AIC). Le critère AIC mesure la qualité d'un modèle, c'est pourquoi il faut avoir un modèle qui minimise l'AIC. Puisque notre seuil de 6.93 est celui qui à l'AIC le plus faible, nous pouvons dire qu'il représente le meilleur modèle possible pour notre échantillon.

Nous estimons donc le modèle suivant :

$$happiness = \beta_0 + \beta_1 Dvlp_eco_et_social + \beta_2 Lib_ind_et_coll + \varepsilon_i \quad (4.3)$$

Avec :

$$happiness = \begin{cases} 1 & \text{si } happiness > 6.93 \\ 0 & \text{sinon} \end{cases}$$

En rentrant cette expression dans le logiciel , nous obtenons les résultats suivants :

TABLE 4.11 – RLB au seuil médian de 6.93

	Dependent variable :
	happiness
Dvlp_economique_et_social	3.753** (1.585)
Libertes_ind_et_coll	7.220** (3.605)
Constant	-15.851*** (5.259)
Observations	114
Log Likelihood	-10.131
Akaike Inf. Crit.	26.262
Note : *p<0.1; **p<0.05; ***p<0.01	

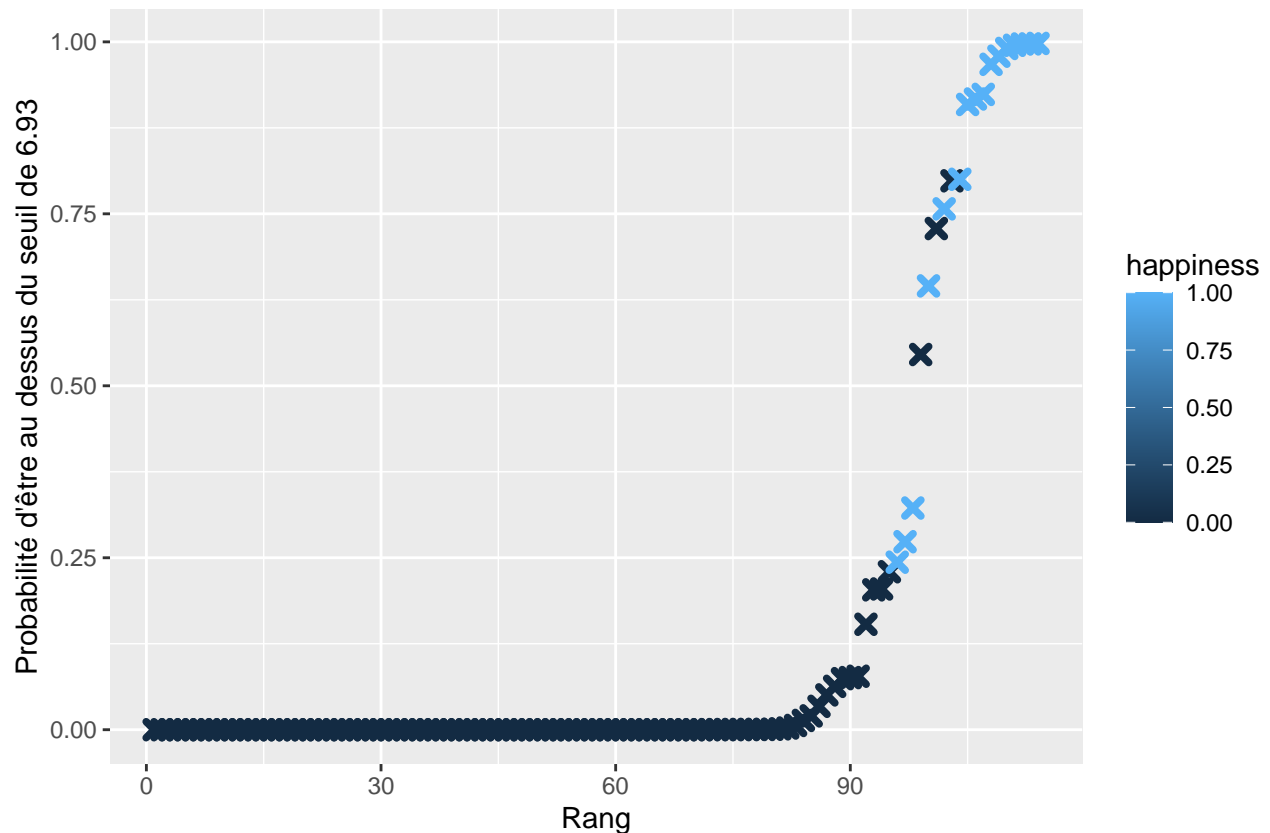
Nos résultats sont désormais tous significatifs au seuil de 5%(Table 4.11), nous pouvons interpréter les coefficients. Pour interpréter les coefficients d'un modèle Logit, nous devons les passer à l'exponentiel. Ainsi, nous obtenons :

$$\beta_1 = e^{3.753} = 42.65 \text{ et } \beta_2 = e^{7.22} = 1366.49$$

Nous pouvons donc dire que quand le **développement économique et social** augmente d'une unité, la probabilité qu'un pays ait un niveau de bonheur supérieur au seuil de 6.93 augmente de 42,65. Lorsque les **libertés individuelles et collectives** augmentent d'une unité, un pays a 1366,5 fois plus de probabilité d'avoir un niveau de bonheur supérieur au seuil de 6.93.

Enfin, nous pouvons projeter notre régression au seuil de 6.93 sur le graphique suivant :

FIGURE 4.4 – Projection graphique de notre RLB



Sur l'abscisse de la Figure 4.4, est représentée la probabilité qu'un pays est un seuil supérieur à 6.93 en fonction des résultats de notre régression logistique binaire. Sur l'ordonnée, nous avons le rang de ces pays en fonction de la probabilité précédente. Plus la probabilité qu'un pays ait un seuil de bonheur supérieur à 6.93 est grande, plus son rang est élevé. De plus, un code couleur nous indique si le pays à effectivement un niveau de bonheur supérieur au seuil ou non. On remarque qu'avec notre estimation, la Belgique, l'Estonie et les Etats-Unis ont une

probabilité supérieure à 50% d'avoir un niveau de bonheur supérieur à 6.93 alors qu'en réalité, ces trois pays ont un niveau inférieur à 6.93. À l'inverse, Le Royaume-Uni, le Costa-Rica et la République Tchèque qui ont des probabilités relativement faible (25%) d'avoir un niveau de bonheur supérieur à notre seuil se retrouvent avec un niveau de bonheur réel supérieur à 6.93 ¹.

Même si la régression logistique binaire entre l'axe 1 et 2 nous donne de bons résultats, nous souhaitons obtenir un seuil plus bas pour rendre l'analyse plus intéressante. En effet, seulement 16 pays de notre échantillon ont un niveau de bonheur supérieur à 6.93. C'est pourquoi, nous avons décidé de faire une régression logistique binaire avec les axes 1 et 3. De la même façon que pour les deux premiers axes, ni la médiane, ni la moyenne ne nous ont donné des résultats satisfaisants. En essayant plusieurs combinaisons, nous sommes arrivés à un seuil de 6.44. Même si le seuil ne concerne que 10 pays en plus, nous trouvons des résultats intéressants. En effet, en projetant les résultats de notre régression logistique binaire des axes 1 et 3 (Annexe Figure 6.6), nous remarquons que le Mexique et le Guatemala ont une probabilité proche de 0 d'avoir un seuil supérieur à 6.44. Cependant, comme pour le Costa-Rica dans notre première régression, ces deux pays ont un niveau de bonheur réel supérieur à 6.44. Ces trois pays ont la particularité d'être tous les trois en Amérique centrale. Cela nous indique peut-être que nos variables ne sont pas suffisantes pour expliquer le niveau de bonheur dans cette région du monde ou alors qu'ils ont naturellement une population plus heureuse.

1. Malheureusement, nous n'avons pas réussi à afficher les pays sur le graphique.

Chapitre 5

Conclusion et discussions des résultats

5.0.1 Conclusion

Dans ce dossier, nous avons essayé de déterminer les variables expliquant le niveau de bonheur d'un pays. À l'aide de l'analyse en composantes principales, nous avons réussi à transformer dix variables explicatives en deux variables latentes bien distinctes : Le niveau de **développement économique et social** et le niveau de **liberté individuelle et collective**. A partir de ces deux composantes, nous avons estimé un modèle de régression linéaire multiple entre introduisant le bonheur comme variable illustrative. Les résultats de cette régression se sont avérés plutôt concluants puisque d'une part, le modèle s'est avéré être significatif au seuil de 5%, et de l'autre, ses prédictions sont relativement bonnes. Enfin, puisque notre modèle n'était pas linéaire, nous avons décidé de compléter notre analyse en effectuant une régression logistique binaire. Cette régression a abouti à la détermination d'un seuil de bonheur à 6.93 sur l'échelle de Cantril.

5.0.2 Discussion

Essayer de trouver les déterminants du bonheur est un exercice à la fois difficile et passionnant. Comme nous l'avons mentionné dans notre introduction, le bonheur est une notion particulièrement subjective. Il est donc difficile d'avoir un indicateur parfait pour le calculer. De plus, le choix de nos variables explicatives pourrait être contesté, le bonheur étant largement multifactoriel. Concernant nos variables explicatives, nous sommes aussi conscients du fait qu'il pourrait y avoir un certain biais de multicollinéarité. Cela pourrait être le cas pour nos variables liées aux libertés ainsi que celles liées à la mortalité infantile. Peut-être qu'en garder que certaines d'entre elles auraient été plus pertinent pour notre analyse. Cela nous au-

rait peut-être permis d'intégrer des variables de thématiques diverses. Néanmoins, on pourrait aussi avancer l'argument que ce choix nous a permis de dégager deux variables latentes aussi distinctes l'une de l'autre. Nous pourrions peut-être aussi critiquer notre sélection de variables. En effet, l'envie de déterminer ces deux grands axes nous a peut-être poussé à retirer trop de variables de notre modèle. Néanmoins, nous assumons pleinement ce choix et sommes satisfaits des résultats que nous obtenons. En effet, malgré le fait que le bonheur soit subjectif, nous pensons qu'à l'échelle d'un pays, certaines conditions doivent être remplies afin de bénéficier d'un niveau de bonheur décent. C'est pourquoi, nous sommes satisfait des deux variables latentes que nous obtenons car nous pensons qu'elles représentent deux conditions *sine qua non* pour qu'un pays puisse bénéficier d'un certain niveau de bonheur.

En espérant qu'il puisse devenir un jour un indicateur à part entière guidant les politiques des pays, nous aimerions terminer ce travail avec Aristote qui, dans son livre *La métaphysique* expliquait que *tous les hommes cherchent à être heureux, et qu'un bon gouvernement est celui qui permettra de donner aux citoyens cette possibilité*¹.

1. <https://www.cairn.info/revue-projet-2018-1-page-26.htm>[consulté le 09/10/2022]

Chapitre 6

Annexes

6.0.1 Figures et Tables supplémentaires

TABLE 6.1 – Tests de Grubbs supplémentaires

Variable	G	U	p-value	alternative hypothesis
elec_access	2.65	0.94	0.81	highest value 0.0926 is an outlier
infant_mort	3.09	0.91	0.19	highest value 0.1254 is an outlier

TABLE 6.2 – Tests de Shapiro-Wilk supplémentaires

Variables	W	p-value
gdp_capita	0.85	2.16e-09
elec_access	0.68	1.738e-14
not_water	0.78	1.03e-11
infant_mort	0.80	3.281e-11
liberte_expression	0.87	9.828e-09
liberte_association	0.78	1.03e-11
unsafe_sanitation	0.67	1.37e-14
press	0.943	0.00011

FIGURE 6.1 – Cercle de corrélation initial entre les Axes 1 et 3

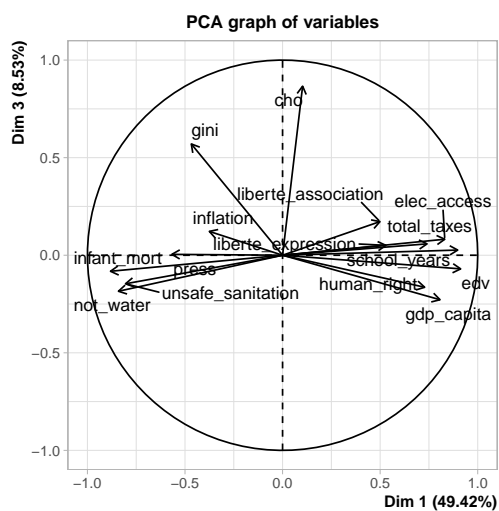


FIGURE 6.2 – Cercle de corrélation intermédiaire 1 entre Axes 1 et 2

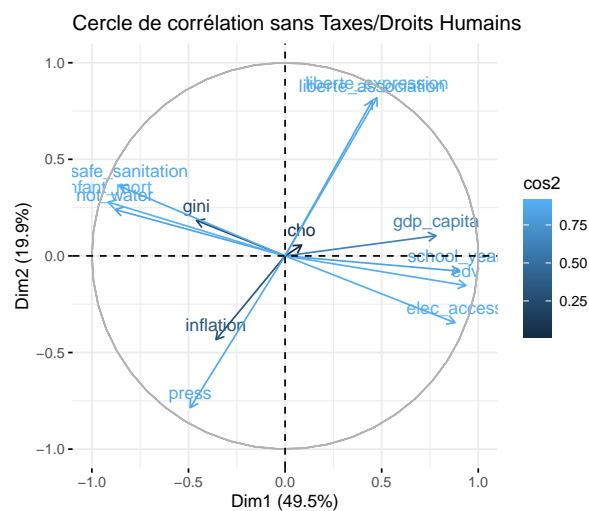


TABLE 6.6 – Coordonnées des individus supplémentaires : Guatemala

Pays	Dim. 1	Dim. 2
Guatemala	0	0

TABLE 6.7 – Cosinus carrés des individus supplémentaires : Iran

Pays	Dim. 1	Dim. 2
Iran	0	0.95

TABLE 6.3 – Cosinus carrés supplémentaires : sans Taxes et Droits Humains

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
gdp_capita	0.61	0.01	0.05	0.01	0.04
school_years	0.81	0.01	0.00	0.01	0.01
elec_access	0.77	0.12	0.01	0.02	0.00
gini	0.21	0.03	0.32	0.19	0.12
not_water	0.77	0.06	0.03	0.00	0.00
edv	0.88	0.02	0.01	0.00	0.01
cho	0.01	0.00	0.77	0.03	0.18
inflation	0.13	0.19	0.02	0.51	0.13
infant_mort	0.84	0.08	0.01	0.00	0.01
liberte_expression	0.23	0.67	0.00	0.01	0.03
liberte_association	0.20	0.64	0.01	0.01	0.06
unsafe_sanitation	0.73	0.13	0.02	0.01	0.03
press	0.24	0.62	0.00	0.02	0.01

TABLE 6.4 – Valeurs propres supplémentaires : Modèle intermédiaire

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	6	50	50
comp 2	3	20	69
comp 3	1	10	79

TABLE 6.5 – Valeurs propres supplémentaires : Modèle final

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	6	61	61
comp 2	2	24	86
comp 3	1	6	91

TABLE 6.8 – Résultats supplémentaires : RLM sur Axes 1 et 2

pvalue	< 2.2e-16
--------	-----------

FIGURE 6.3 – Boîte à moustaches : Régression Axes 1/3 et 2/3

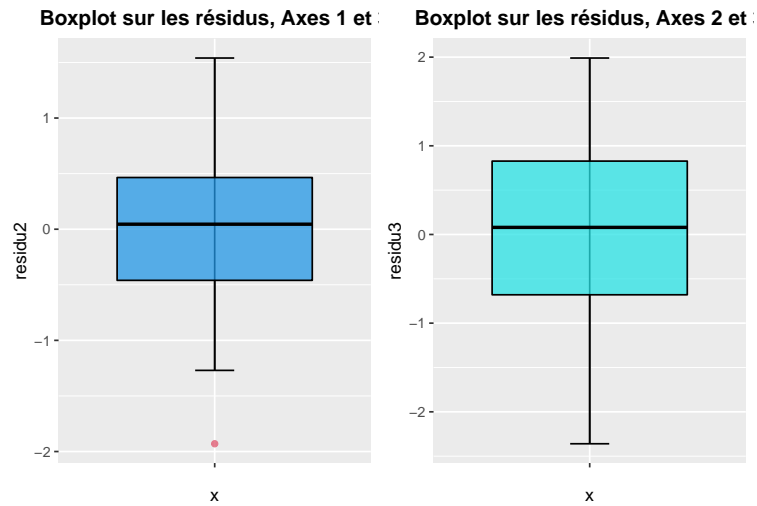


FIGURE 6.4 – Histogrammes : Régression Axes 1/3 et 2/3

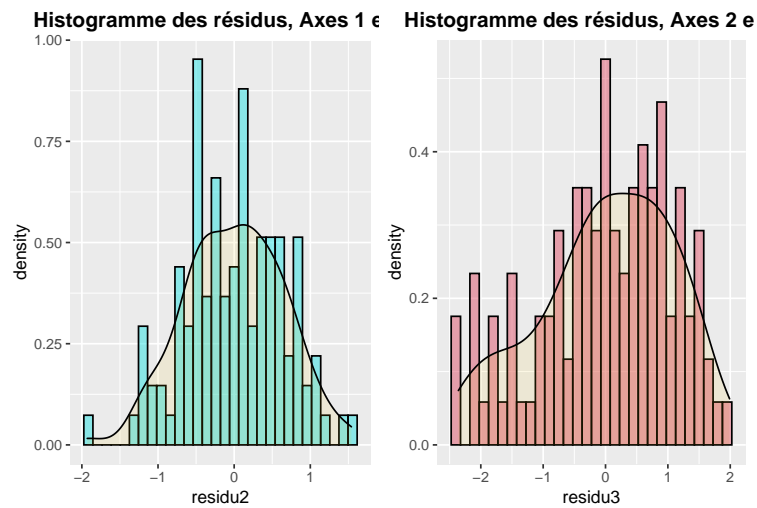


FIGURE 6.5 – QQplots : Régression Axes 1/3 et 2/3

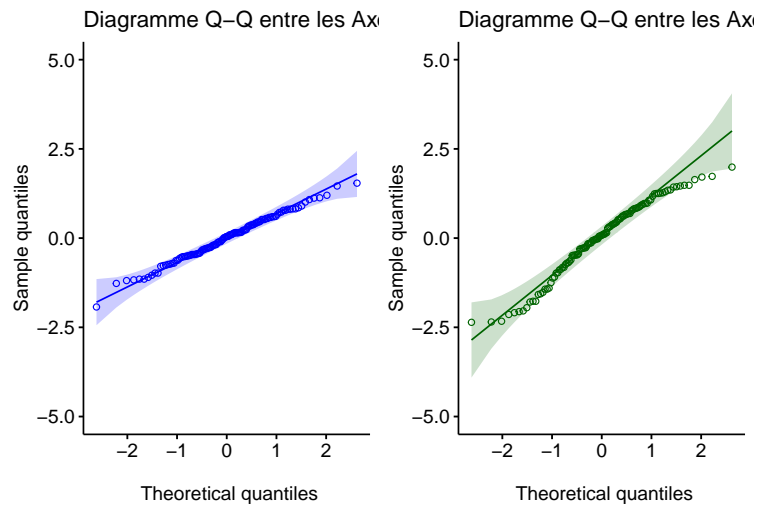


FIGURE 6.6 – Représentation graphique supplémentaire : RLB

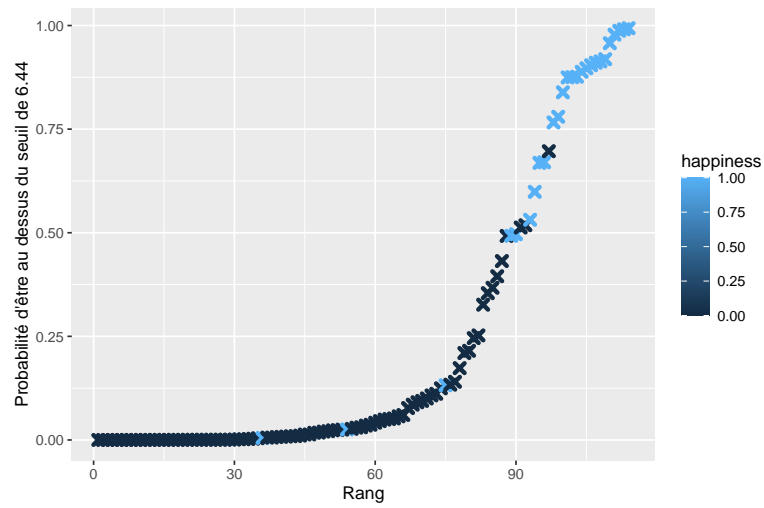


TABLE 6.9 – RLB au seuil de 5.8

	<i>Dependent variable :</i>
	happiness
Dvlp_economique_et_social	0.996*** (0.204)
Libertes_ind_et_coll	0.288 (0.215)
Constant	−0.737** (0.331)
Observations	114
Log Likelihood	−43.986
Akaike Inf. Crit.	93.973
<i>Note :</i> *p<0.1; **p<0.05; ***p<0.01	

TABLE 6.10 – RLB au seuil de 6.3 (Q3)

	<i>Dependent variable :</i>
	happiness
Dvlp_economique_et_social	2.143*** (0.614)
Libertes_ind_et_coll	−0.138 (0.487)
Constant	−4.279*** (1.159)
Observations	114
Log Likelihood	−29.579
Akaike Inf. Crit.	65.159
<i>Note :</i> *p<0.1; **p<0.05; ***p<0.01	

6.0.2 Script

Pour plus de visibilité, le script que vous trouverez à la page suivante a été converti au format Markdown.

Importation des library

```
library(readxl)
library(tidyverse)
library(dplyr)
library(readxl)
library(WriteXLS)
library(FactoMineR)
library(factoextra)
library(ggplot2)
library(ggpubr)
library(ggcorrplot)
library(gridExtra)
library(cowplot)
library(corrplot)
library(outliers)
library(EnvStats)
library(tseries)
library(ggpubr)
library(stats)
library(lmtest)
library(lattice)
library(leaps)
library(ggrepel)
```

Model initial

```
base = read_xlsx("/Users/noa/Desktop/M1/ADD/BDD_definitive2_ACP.xlsx")
base2 = na.omit(base)
dim(base2)

base3 = select(base2, -c(gini,
                          cho,
                          total_taxes,
                          inflation,
                          human_right))

res.pca = PCA(base2[3:17], axes=c(1,2))
```

```
fviz_pca_var(res.pca, col.var="cos2")
```

Corrélation entre la liberté de la presse et le niveau de bonheur

```
ggplot(base3, aes(press, happiness, color=happiness))+
  geom_point(size=2, alpha=0.95)+
  stat_cor(method="spearman", aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3, color="black", aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et de liberté de la presse")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("Niveau de liberté de la presse")+
  ylab("Niveau de bonheur")
```

Corrélation entre la liberté d'expression et le niveau de bonheur

```
ggplot(base3, aes(liberte_expression, happiness, color=happiness))+
  geom_point(size=2, alpha=0.95)+
  stat_cor(method="spearman", aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3, color="black", aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et de liberté d'expression")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("Niveau de liberté d'expression")+
  ylab("Niveau de bonheur")
```

Corrélation entre la liberté d'association et le niveau de bonheur

```
ggplot(base3, aes(liberte_association,happiness,color=happiness))+
  geom_point(size=2,alpha=0.95)+
  stat_cor(method="spearman",aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3,color="black",aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et de liberté d'association")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("Niveau de liberté d'association")+
  ylab("Niveau de bonheur")
```

Corrélation entre l'esperance de vie et le niveau de bonheur

```
ggplot(base3, aes(edv,happiness,color=happiness))+
  geom_point(size=2,alpha=0.95)+
  stat_cor(method="spearman",aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3,color="black",aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur l'esperance de vie")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("Niveau de l'esperance de vie")+
  ylab("Niveau de bonheur")
```

Corrélation entre le PIB/HAB et le niveau de bonheur

```
ggplot(base3, aes(gdp_capita,happiness,color=happiness))+
  geom_point(size=2,alpha=0.95)+
  stat_cor(method="spearman",aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3,color="black",aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et le PIB/Habitant")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("Niveau du PIB/Habitant")+
  ylab("Niveau de bonheur")
```

Corrélation entre le niveau d'accès à l'électricité et le niveau de bonheur

```
ggplot(base3, aes(elec_access,happiness,color=happiness))+
  geom_point(size=2,alpha=0.95)+
  stat_cor(method="spearman",aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3,color="black",aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et le Niveau d'accès à l'électricité")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("Niveau d'accès à l'électricité")+
  ylab("Niveau de bonheur")
```

Corrélation entre le niveau d'étude et le niveau de bonheur

```
ggplot(base3, aes(school_years,happiness,color=happiness))+
  geom_point(size=2,alpha=0.95)+
  stat_cor(method="spearman",aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3,color="black",aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et le nombre d'années d'étude moyen")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("nombre d'années d'étude moyen")+
  ylab("Niveau de bonheur")
```

Corrélation entre la mortalité infantile et le niveau de bonheur

```
ggplot(base3, aes(infant_mort,happiness,color=happiness))+
  geom_point(size=2,alpha=0.95)+
  stat_cor(method="spearman",aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3,color="black",aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et la
  Part des nouveau-nés qui meurent avant d'avoir atteint l'âge de cinq ans.")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("part des nouveau-nés qui meurent avant d'avoir atteint l'âge de cinq ans.")+
  ylab("Niveau de bonheur")
```

Corrélation entre l'accès à l'eau et le niveau de bonheur

```
ggplot(base3, aes(not_water, happiness, color=happiness))+
  geom_point(size=2, alpha=0.95)+
  stat_cor(method="spearman", aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3, color="black", aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et la
          part de la population n'ayant pas accès à une source d'eau améliorée")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("Part de la population n'ayant pas accès à une source d'eau améliorée.")+
  ylab("Niveau de bonheur")
```

Corrélation entre les mauvaises conditions sanitaires et le niveau de bonheur

```
ggplot(base3, aes(unsafe_sanitation, happiness, color=happiness))+
  geom_point(size=2, alpha=0.95)+
  stat_cor(method="spearman", aes(label = ..r.label..))+
  geom_smooth(method=lm, alpha=0.3, color="black", aes(fill=happiness))+
  ggtitle("Corrélation entre le niveau de bonheur et le
          taux de mortalité dû à l'insalubrité des installations sanitaires")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))+
  xlab("Taux de mortalité dû à l'insalubrité des installations sanitaires")+
  ylab("Niveau de bonheur")
```

Matrice des corrélations

```
Tableau_Corrélation = round(cor(base3[,c(2:12)]), use="complete.obs"), 2)

ggcorrplot(Tableau_Corrélation, hc.order = TRUE, lab = TRUE, type = "lower",
            outline.col = "white",
            ggtheme = ggplot2::theme_gray,
            colors = c("#6D9EC1", "white", "#E46726"))+
  labs(title = "Matrice des corrélations")
```

Cercles ACP

```
#1ER CERCLE ACP
res.pca1 = PCA(base2[3:17], axes=c(1,2))
```

```
fviz_pca_var(res.pca1, col.var="cos2")
```

```
#CERCLE ACP AVEC NOS VARIABLES PERTINENTES -- AXE 1 & 2
res.pca2 = PCA(base3[3:12], axes=c(1,2))
```

```
fviz_pca_var(res.pca2, col.var="cos2")
```

```
#CERCLE ACP AVEC NOS VARIABLES PERTINENTES -- AXE 1 & 3
res.pca3 = PCA(base3[3:12], axes=c(1,3))
```

```
fviz_pca_var(axes=c(1,3), res.pca3, col.var="cos2")
```

Histogramme des valeurs propres

```
fviz_eig(res.pca2, choice = c("variance", "eigenvalue"),
        geom = c("bar", "line"), barfill = "royalblue1",
        barcolor = "grey35", linecolor = "black",
        ncp = 16, addlabels = TRUE,
        main="Histogramme_des_valeurs_propres")
```

```
eig_ACP1<-round(res.pca$eig,2)
```

Histogrammes des niveaux de libertés

```
histo_expression=ggplot(base3, aes(x=liberte_expression)) +  
  geom_histogram(aes(y=..density..), colour="black", fill=5,alpha=0.5)+  
  geom_density(alpha=0.10, fill=7)+  
  ggtitle("Histogramme de la liberté d'expression")+  
  theme(plot.title = element_text(hjust = 0.5, face="bold"))  
  
histo_association=ggplot(base3, aes(x=liberte_association)) +  
  geom_histogram(aes(y=..density..), colour="black", fill=4,alpha=0.5)+  
  geom_density(alpha=0.10, fill=7)+  
  ggtitle("Histogramme de la liberté d'association")+  
  theme(plot.title = element_text(hjust = 0.5, face="bold"))  
  
plot_grid(histo_expression,  
          histo_association,  
          ncol = 2, nrow = 1)
```

Boxplots de toutes les variables

```

#BOXPLOT school_years
boxplot1 = ggplot(base3, aes(x="",y=school_years))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 7,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("school_years")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT edv
boxplot2 = ggplot(base3, aes(x="",y=edv))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 2,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("edv")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT liberte_association
boxplot3 = ggplot(base3, aes(x="",y=liberte_association))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 3,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("liberte_association")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT elec_access
boxplot4 = ggplot(base3, aes(x="",y=elec_access))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 4,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("elec_access")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT infant_mort
boxplot5 = ggplot(base3, aes(x="",y=infant_mort))+
  stat_boxplot(geom = "errorbar",
    width = 0.75,
    color = 1)+
  geom_boxplot(fill = 5,
    alpha = 0.5,
    color = 1,
    outlier.colour = 1)+
  ggtitle("infant_mort")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT unsafe_sanitation
boxplot6 = ggplot(base3, aes(x="",y=unsafe_sanitation))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 6,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("unsafe_sanitation")+

```



```

    theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT gdp_capita
boxplot7 = ggplot(base3, aes(x="",y=gdp_capita))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 7,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("gdp_capita")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT not_water
boxplot8 = ggplot(base3, aes(x="",y=not_water))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 1,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("not_water")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT liberte_expression
boxplot9 = ggplot(base3, aes(x="",y=liberte_expression))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 2,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("liberte_expression")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT press
boxplot10 = ggplot(base3, aes(x="",y=press))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 4,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("press")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#BOXPLOT HAPPY
boxplot11 = ggplot(base3, aes(x="",y=happiness))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 3,
    alpha = 0.75,
    color = 1,
    outlier.colour = 1)+
  ggtitle("happiness")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

plot_grid(boxplot1,
  boxplot2,
  boxplot3,
  boxplot4,
  boxplot5,
  boxplot6,
  boxplot7,
  boxplot8,
  boxplot9,
  boxplot10,

```

```
boxplot11,
ncol = 4, nrow = 3)
```

Test d'atypicité

```
grubbs.test(base3$elec_access,type=10, two.sided = TRUE)
grubbs.test(base3$infant_mort,type=10, two.sided = TRUE)

rosnerTest(base3$liberte_association, k = 7, alpha = 0.05)
#pas de valeurs atypiques

rosnerTest(base3$unsafe_sanitation, k = 7, alpha = 0.05)
#2 valeurs atypiques: 119.05 et 89.43 correspondant (respectivement)
#au Chad et au Niger.

rosnerTest(base3$gdp_capita, k = 5, alpha = 0.05)
#1 valeur atypique : 86516.09 correspondant au gdp du Luxembourg.

rosnerTest(base3$not_water, k = 8, alpha = 0.05)
#1 valeur atypique : 0.4752 correspondant à la valeur de Madagascar

rosnerTest(base3$liberte_expression, k = 3, alpha = 0.05)
#pas de valeurs atypiques

rosnerTest(base3$press, k = 4, alpha = 0.05)
#pas de valeurs atypiques
```

Test de Rosner

```
rosnerTest(base3$liberte_association, k = 7, alpha = 0.05)
#pas de valeurs atypiques

rosnerTest(base3$unsafe_sanitation, k = 7, alpha = 0.05)
#2 valeurs atypiques: 119.05 et 89.43 correspondant (respectivement)
#au Chad et au Niger.

rosnerTest(base3$gdp_capita, k = 5, alpha = 0.05)
#1 valeur atypique : 86516.09 correspondant au gdp du Luxembourg.

rosnerTest(base3$not_water, k = 8, alpha = 0.05)
#1 valeur atypique : 0.4752 correspondant à la valeur de Madagascar

rosnerTest(base3$liberte_expression, k = 3, alpha = 0.05)
#pas de valeurs atypiques

rosnerTest(base3$press, k = 4, alpha = 0.05)
#pas de valeurs atypiques
```

Test de normalité

```
#H0: suit une loi normale

shapiro.test(base3$elec_access)
#p<0.05 donc la variable ne suit PAS une loi normale.

shapiro.test(base3$infant_mort)
#p<0.05 donc la variable ne suit PAS une loi normale.

shapiro.test(base3$press)
#p<0.05 donc la variable ne suit PAS une loi normale.

shapiro.test(base3$liberte_expression)
#p<0.05 donc la variable ne suit PAS une loi normale.

shapiro.test(base3$liberte_association)
#p<0.05 donc la variable ne suit PAS une loi normale.

shapiro.test(base3$unsafe_sanitation)
#ici, p<0.05 donc la variable ne suit PAS une loi normale.

shapiro.test(base3$gdp_capita)
#ici, p<0.05 donc la variable ne suit PAS une loi normale.

shapiro.test(base3$not_water)
#ici, p<0.05 donc la variable ne suit PAS une loi normale.
```

Coordonnées

```
round(res.pca2$var$coord, 2)

coordonnees<-round(res.pca$var$coord, 2)
corrplot(coordonnees, is.corr=FALSE,method='circle',color=palette,
         addCoef.col=TRUE,tl.col= "black",number.cax=0.7 )
```

Contributions

```
round(res.pca2$var$contrib, 2)

contributions<-round(res.pca$var$contrib, 2)
corrplot(contributions, is.corr=FALSE,method='circle',color=palette,
         addCoef.col=TRUE,tl.col= "black",number.cax=0.7 )
```

Cosinus carrés

```
#COSINUS CARRÉS AVEC VARIABLES INITIALS
round(res.pca$var$cos2, 2)

cosinus_initial=round(res.pca$var$cos2, 2)
corrplot(cosinus_initial, is.corr=FALSE,method='circle',color=palette,
         addCoef.col=TRUE,tl.col= "black",number.cax=0.7)

rbind(cosinus_initial,MOYENNE_COS2_=colSums(cosinus_initial/dim(cosinus_initial)[1]))
round(rbind(cosinus_initial,MOYENNE_COS2_=colSums(cosinus_initial/dim(cosinus_initial)[1])),2)

cos2_moyenne_ = round(rbind(cosinus_initial,MOYENNE_COS2_=colSums(cosinus_initial/dim(cosinus_initial)[1]))

corrplot(cos2_moyenne_,
         is.corr=FALSE,
         method='circle',
         color=palette,
         title="Tableau des cosinus carrés",
         addCoef.col=TRUE,
         tl.col="black",
         number.cex=0.7)
```

```
#COSINUS CARRES AVEC VARIABLES FINALES
round(res.pca2$var$cos2, 2)
```

```
cosinus=round(res.pca2$var$cos2, 2)
corrplot(cosinus, is.corr=FALSE,method='circle',color=palette,
          addCoef.col=TRUE,tl.col= "black",number.cax=0.7)
```

```
rbind(cosinus,MOYENNE_COS2=colSums(cosinus/dim(cosinus)[1]))
round(rbind(cosinus,MOYENNE_COS2=colSums(cosinus/dim(cosinus)[1])),2)

cos2_moyenne<-round(rbind(cosinus,MOYENNE_COS2=colSums(cosinus/dim(cosinus)[1])),2)

corrplot(cos2_moyenne,
          is.corr=FALSE,
          method='circle',
          color=palette,
          title="Tableau des cosinus carrés",
          addCoef.col=TRUE,
          tl.col="black",
          number.cex=0.7)
```

Projection des individus

```
coord_individus<- round(res.pca2$ind$coord, 2)
contrib_individus<- round(res.pca2$ind$contrib, 2)
cos2_individus <- round(res.pca2$ind$cos2, 2)

fviz_pca_ind(res.pca2,
             col.ind="cos2",
             gradient.cols=c("#00AFBB", "#E7B800", "#FC4E07"),
             repel=TRUE,
             title="projection des pays sur l'axe 1 et 2")
```

Projection des individus avec les pays affichés

```
ok2 = PCA(base3,
           scale.unit=TRUE,
           ncp=2,
           quali.sup=1, graph=T)
```

```
gr = plot.PCA(ok2, axes=c(1, 2), choix="ind",
              col.quali = rainbow(114, start = 0, end = 1))
gr + theme(panel.background = element_rect(fill = 'grey93', color = 'gray27'),
           panel.grid = element_line(color = "gray78",
                                     size = 0.75,
                                     linetype = 2),
           plot.title=element_text(size=14, color="grey32"),
           axis.title = element_text(size=12, color="grey32"))
```

Coordonnées,contributions et cos2 des individus

```
COORD_IND<- round(res.pca2$ind$coord, 2)
COORD_IND

CONTRIB_IND<- round(res.pca2$ind$contrib, 2)
CONTRIB_IND
CONTRIB_IND[41,1:2] #permet de voir le Guatemala, au barycentre.

COS2_IND<- round(res.pca2$ind$cos2, 2)
COS2_IND
COS2_IND[c(49),1:2] #permet de voir valeur extreme

rbind(COS2_IND,MOYENNE_COS2=colSums(COS2_IND/dim(COS2_IND)[1]))
round(rbind(COS2_IND,MOYENNE_COS2=colSums(COS2_IND/dim(COS2_IND)[1])),2)
```

Oppositions des valeurs extrêmes

```
ok3 = PCA(base3[c(14,17,32,33,50,51,56,59,61,77,79,89,92,98,100,109)],,
  scale.unit=TRUE,
  ncp=2,
  quali.sup=1, graph=T)
```

```
gr = plot.PCA(ok2, axes=c(1, 2), choix="ind",
  col.quali = rainbow(114, start = 0.55, end = 0.85))
gr + theme(panel.background = element_rect(fill = 'grey93', color = 'gray27'),
  panel.grid = element_line(color = "gray78",
    size = 0.75,
    linetype = 2),
  plot.title=element_text(size=14, color="grey32"),
  axis.title = element_text(size=12, color="grey32"))
```

Projection axes 1&3 puis 2&3

```
#Projection axe 1 & 3
ACP_1_3 = fviz_pca_var(res.pca2,
  axes=c(1,3),
  col.var = "cos2",
  gradient.cols = c("#00AFBB", "#FC4E07", "#B9121B"),
  repel = TRUE,
  title= "ACP axe 1 et 3")

#Projection axe 2 & 3
ACP_2_3 = fviz_pca_var(res.pca2,
  axes=c(2,3),
  col.var = "cos2",
  gradient.cols = c("#00AFBB", "#FC4E07", "#B9121B"),
  repel = TRUE,
  title= "ACP axe 2 et 3")

plot_grid(ACP_1_3,
  ACP_2_3,
  ncol = 2, nrow = 1)
```

Equations de régression

```
Dvlp_economique_et_social = res.pca2$ind$coord[,1]
Libertes_ind_et_coll = res.pca2$ind$coord[,2]

reg1 = lm(base3$happiness~ Dvlp_economique_et_social + Libertes_ind_et_coll)
summary(reg1)

reg2=lm(base3$happiness ~ res.pca2$ind$coord[,1]+res.pca2$ind$coord[,3])
summary(reg2)

reg3=lm(base3$happiness ~ res.pca2$ind$coord[,2]+res.pca2$ind$coord[,3])
summary(reg3)
```

Création des variables predict et résidu

```
base3$predict = round(predict(reg1), 2)
base3$residu = round(residuals(reg1), 2)

base3$predict2 = round(predict(reg2), 2)
base3$residu2 = round(residuals(reg2), 2)

base3$predict3 = round(predict(reg3), 2)
base3$residu3 = round(residuals(reg3), 2)
```

Graphique des résidus

```

#GRAPHIQUE AXE 1-2
boxplot_residu_1 = ggplot(base3, aes(x="", y=residu))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 3,
    alpha = 0.75,
    color = 1,
    outlier.colour = 2)+
  ggtitle("boxplot sur les résidus, AXE 1 et 2")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#GRAPHIQUE AXE 1-3
boxplot_residu_2 = ggplot(base3, aes(x="", y=residu2))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 4,
    alpha = 0.75,
    color = 1,
    outlier.colour = 2)+
  ggtitle("boxplot sur les résidus, AXE 1 et 3")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#GRAPHIQUE AXE 2-3
boxplot_residu_3 = ggplot(base3, aes(x="", y=residu3))+
  stat_boxplot(geom = "errorbar",
    width = 0.15,
    color = 1)+
  geom_boxplot(fill = 5,
    alpha = 0.75,
    color = 1,
    outlier.colour = 2)+
  ggtitle("boxplot sur les résidus, AXE 2 et 3")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

plot_grid(boxplot_residu_2,
  boxplot_residu_3,
  ncol = 2, nrow = 1)

```

Histogramme des résidus

```

#Histogramme axe 1 & 2
histo_residu=ggplot(base3, aes(x=residu)) +
  geom_histogram(aes(y=..density..), colour="black", fill=5,alpha=0.5)+
  geom_density(alpha=0.10, fill=7)+
  ggtitle("Histogramme des résidus")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#Histogramme axe 1 & 3
hist_1_3 = histo_residu=ggplot(base3, aes(x=residu2)) +
  geom_histogram(aes(y=..density..), colour="black", fill=5,alpha=0.5)+
  geom_density(alpha=0.10, fill=7)+
  ggtitle("Histogramme des résidus, axe 1 & 3")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

#Histogramme axe 2 & 3
hist_2_3 = histo_residu=ggplot(base3, aes(x=residu3)) +
  geom_histogram(aes(y=..density..), colour="black", fill=2,alpha=0.5)+
  geom_density(alpha=0.10, fill=7)+
  ggtitle("Histogramme des résidus, axe 2 & 3")+
  theme(plot.title = element_text(hjust = 0.5, face="bold"))

plot_grid(hist_1_3,
  hist_2_3,
  ncol = 2, nrow = 1)

```

QQ plot des résidus

```
#QQ PLOT AXE 1 & 2
ggqqplot(base3$residu, add="qqline", xlab="\n Theoretical quantiles",
          ylab="Sample quantiles", shape=1, ylim=c(-5,5), col="blue")

#QQ PLOT AXE 1 & 3
qq_1_3 = ggqqplot(base3$residu2, add="qqline", xlab="\n Theoretical quantiles",
                  ylab="Sample quantiles", shape=1, ylim=c(-5,5), col="blue",
                  main="Diagramme Q-Q entre les axes 1 & 3")

#QQ PLOT AXE 2 & 3
qq_2_3 = ggqqplot(base3$residu3, add="qqline", xlab="\n Theoretical quantiles",
                  ylab="Sample quantiles", shape=1, ylim=c(-5,5), col="dark green",
                  main="Diagramme Q-Q entre les axes 2 & 3")

plot_grid(qq_1_3,
          qq_2_3,
          ncol = 2, nrow = 1)
```

Test de normalité sur les résidus

```
shapiro.test(base3$residu)
shapiro.test(base3$residu2)
shapiro.test(base3$residu3)
```

Test de KS

```
ks.test(base3$residu, "pnorm", mean(base3$residu), sd(base3$residu))

ks.test(base3$residu2, "pnorm", mean(base3$residu2), sd(base3$residu2))

ks.test(base3$residu3, "pnorm", mean(base3$residu3), sd(base3$residu3))
```

Regression logistique binaire

```

#On garde uniquement les variables du modèle final
datastat = select(data,-c(total_taxes, human_right,gini,cho,inflation))
#Construction de trois set différents pour avoir des tableaux utilisables dans Latex
datastat1 = select(datastat, c(Pays,happiness,gdp_capita,elec_access))
datastat2 = select(datastat, c(school_years, liberte_expression, liberte_association, press))
datastat3 = select(datastat, c(edv, unsafe_sanitation, infant_mort, not_water))
#Exportation Latex
print(xtable(summary(datastat1), type = "latex"), file = "summary1.txt", include.rownames = F)
print(xtable(summary(datastat2), type = "latex"), file = "summary2.txt", include.rownames = F)
print(xtable(summary(datastat3), type = "latex"), file = "summary3.txt", include.rownames = F)

#REGRESSION LOGISTIQUE BINAIRE
databin = data3
databin$happiness[data3$happiness>6.93]="1"
databin$happiness[data3$happiness<=6.93]="0"
databin$happiness = as.numeric(databin$happiness)
logit = glm(databin$happiness~Dvlp_economique_et_social + Libertes_ind_et_coll, family = "binomial")
summary(logit)
stargazer(logit)
#GRAPHIQUE RLB
predictdata = data.frame(prob5.5 = logit$fitted.values, happiness=databin$happiness)
predictdata = predictdata[order(predictdata$prob5.5, decreasing = F),]
predictdata$rank =1:nrow(predictdata)
ggplot(data=predictdata, aes(x=rank, y=prob5.5))+ geom_point(aes(color = happiness), alpha = 1, shape = 4,

#Regression Logistique Binaire 1/3

databin = data3
databin$happiness[data3$happiness>6.44]="1"
databin$happiness[data3$happiness<=6.44]="0"
databin$happiness = as.numeric(databin$happiness)
logit2 = glm(databin$happiness~Dvlp_economique_et_social + Libertes_ind_et_coll_3, family = "binomial")
summary(logit2)
stargazer(logit2)
#GRAPHIQUE RLB
predictdata = data.frame(prob5.5 = logit2$fitted.values, happiness=databin$happiness)
predictdata = predictdata[order(predictdata$prob5.5, decreasing = F),]
predictdata$rank =1:nrow(predictdata)
ggplot(data=predictdata, aes(x=rank, y=prob5.5))+ geom_point(aes(color = happiness), alpha = 1, shape = 4,

```


6.0.3 Base de donnée

Toutes nos données ont été obtenues via le site *Our World in Data*. Ci-dessous les liens pour chaque variables.

Niveau de bonheur

<https://ourworldindata.org/grapher/happiness-cantril-ladder?tab=map>

Liberte de la presse

<https://ourworldindata.org/grapher/press-freedom-rsf>

Liberte d'expression

<https://ourworldindata.org/grapher/freedom-of-expression-bti>

Liberte d'association

<https://ourworldindata.org/grapher/freedom-of-association-bti>

Esperance de Vie

<https://ourworldindata.org/grapher/life-expectancy?time=2016>

PIB par habitant

<https://ourworldindata.org/grapher/gdp-per-capita-in-us-dollar-world-bank>

L'accès à l'électricité

<https://ourworldindata.org/grapher/share-of-the-population-with-access-to-electricity>

Le nombre d'années d'études

<https://ourworldindata.org/grapher/mean-years-of-schooling-long-run>

La mortalité infantile

<https://ourworldindata.org/grapher/infant-mortality>

L'accès à l'eau

<https://ourworldindata.org/water-access>

Exposition à de mauvaises infrastructures sanitaires

<https://ourworldindata.org/grapher/death-rate-from-unsafe-sanitation>

Droits Humains

<https://ourworldindata.org/grapher/human-rights-vdem>

Chômage

<https://ourworldindata.org/grapher/unemployment-rate>

Indice de Gini

<https://ourworldindata.org/grapher/economic-inequality-gini-index>

Taxes

<https://ourworldindata.org/grapher/total-tax-revenues-gdp>

Présentation de la base de donnée

La Table 6.11 reproduit notre base de donnée au format \LaTeX . Si cette dernière est de petite taille, sa définition fait que l'on peut facilement zoomer dessus. Pour plus de clarté, la base de donnée sera également disponible au format Excel.

TABLE 6.11 – Base de donnée

Pays	highways	gdp_capita	school_years	elect_access	gini	sea_water	edu	cha	total_bares	inflation	infant_mort	human_right	liberty_expression	liberty_association	sanitary_sanitation	press
Albania	5	11674.59	10	0.9998	0.34	0.0298	78.2	0.1542	0.2354	112.7302286	0.0094	1.09	0.72	0.02	0.05	29.92
Algeria	5.04	11854.6	8	0.9935	0.28	0.0125	76.3	0.1402	0.1417	134.8448762	0.0248	0.48	0.39	0.37	0.36	41.69
Angola	3.79	7581.55	5.1	0.4181	0.51	0.3424	59.9	0.0741	0.1467	288.335491	0.0841	-0.15	0.48	0.05	49.17	39.89
Argentina	5.06	11325.96	11.7	0.999	0.33	0.0044	74.6	0.1792	0.2176	123.0434194	0.0134	0.38	0.64	0.09	0.19	28.79
Australia	7.18	51497.21	12.9	1	0.34	0.0003	83	0.0571	0.2861	113.4751643	0.0036	2.02	0.96	0.9	0.02	17.84
Austria	7.4	49065.79	12.1	1	0.31	0	81.3	0.0601	0.4236	111.4754721	0.0036	2.42	0.97	0.9	0.04	13.18
Bangladesh	4.5	3033.73	5.2	0.7392	0.32	0.0167	71.8	0.0435	0.0877	152.5291483	0.0261	-1.19	0.44	0.53	0.29	45.94
Belarus	5.23	18609.34	12.3	1	0.25	0.002	74	0.0594	0.3575	452.9126326	0.0039	0.38	0.3	0.38	0.03	54.32
Belgium	6.09	43413.09	11.8	1	0.28	0	81.1	0.0703	0.4271	110.7738822	0.0041	3.07	0.98	0.98	0.06	14.18
Benin	5.02	3162.43	3.6	0.371	0.48	0.2624	68.9	0.0183	0.09	180.8655189	0.0954	1.08	0.86	0.05	44.84	28.97
Bhutan	3.46	15642.1	9.3	0.6418	0.55	0.0011	68.7	0.2101	0.2807	136.6215564	0.0402	1.64	0.9	0.87	0.21	22.91
Brazil	6.19	14219.4	7.8	0.997	0.53	0.0347	75.2	0.116	0.3379	150.4825942	0.0167	-1.03	0.85	0.92	1.48	32.62
Bulgaria	5.1	18665.13	11.8	1	0.41	0.0084	74.7	0.0737	0.2893	102.7558881	0.0077	1.11	0.78	0.82	0.01	34.46
Burkina Faso	4.93	1844.69	1.5	0.1644	0.47	0.2365	60.7	0.0453	0.1352	188.2362246	0.0975	0.16	0.92	0.84	0.42	22.66
Cambodia	5.25	3684.3	6.3	0.5951	0.47	0.241	58.1	0.0358	0.1338	113.8454985	0.0451	-1.48	0.5	0.51	49.49	48.53
Canada	7.18	47032.89	13.1	1	0.33	0.0077	82.1	0.07	0.3326	110.2246709	0.0053	3.11	0.93	0.87	0.1	15.26
Chad	4.49	1461.3	2.3	0.0926	0.37	0.415	53.4	0.011	0.07	155.768479	0.1249	-0.32	0.6	0.46	119.05	40.59
Chile	6.44	23086.17	10.3	1	0.44	0.0036	79.8	0.0748	0.1884	122.894326	0.0077	1.28	0.94	0.92	0.3	19.23
China	5.13	13147.85	7.9	1	0.38	0.0089	76.2	0.0453	0.2378	117.2305674	0.0099	-1.15	0.1	0.04	0.22	88.96
Colombia	5.08	13566.16	8.3	0.984	0.51	0.032	76.7	0.0809	0.1921	126.4256182	0.015	-0.66	0.84	0.93	0.28	44.11
Congo	5.49	3968.59	6.3	0.4551	0.49	0.1732	63.6	0.2056	0.1342	128.1185166	0.0598	-0.92	0.38	0.42	39.87	35.84
Costa Rica	7.14	17331.13	8.8	0.995	0.49	0.0016	79.7	0.0466	0.1391	121.4833689	0.0089	1.82	0.96	0.94	0.11	11.3
Cote d'Ivoire	5.27	4546.25	5.1	0.643	0.41	0.2016	56.6	0.026	0.118	111.4857049	0.0884	-0.33	0.7	0.77	31.48	30.17
Croatia	5.54	23327.59	11.3	1	0.31	0	78	0.131	0.3766	106.1649964	0.0048	2.5	0.82	0.86	0.02	27.91
Cyprus	6.28	30758.47	12.1	1	0.33	0.0022	80.5	0.1295	0.2562	106.274597	0.0029	1.76	0.95	0.9	0.03	18.26
Czechia	7.03	23560.09	12.7	1	0.25	0.0002	78.3	0.0409	0.2486	103.2551662	0.003	2.29	0.95	0.93	0.02	16.93
Denmark	7.05	48930.09	12.6	1	0.28	0	80.6	0.0599	0.4637	107.4011249	0.0041	3.31	0.99	0.93	0.04	8.89
Dominican Republic	5.43	16169.96	7.8	0.969	0.46	0.0112	73.5	0.0728	0.1291	121.378365	0.035	-0.69	0.74	0.81	2.03	27.9
Ecuador	6.13	11897.75	8.7	0.907	0.45	0.0089	76.4	0.0443	0.1462	123.5734784	0.0148	0.81	0.72	0.68	0.76	33.23
El Salvador	6.24	7707.88	6.9	0.96	0.4	0.0092	72.8	0.0494	0.1736	108.8484795	0.0149	0.42	0.82	0.85	0.26	27.2
Estonia	6.09	29319.8	12.6	1	0.31	0.0046	78.1	0.0676	0.3348	111.4483124	0.0028	3.03	0.98	0.9	0.02	14.31
Everest	4.21	8586.8	6.5	0.843	0.55	0.2314	57	0.2272	0.2332	146.842108	0.0556	0.7	0.22	0.05	41.1	52.37
Finland	6.38	19332.32	12.7	0.929	0.45	0.0065	83.1	0.0239	0.4387	123.6273463	0.0037	2	0.97	0.97	0.2	28.78
France	7.06	42577.58	12.4	1	0.27	0	81.4	0.0862	0.4387	109.1775646	0.0024	2.87	0.96	0.87	0.01	8.59
Germany	6.67	41886.22	11.5	1	0.32	0	82.3	0.1005	0.473	105.7726889	0.0042	1.9	0.98	0.91	0.02	23.83
Gambia	4.92	2137.3	3.4	0.5633	0.36	0.1145	63.2	0.095	0.1095	148.1261532	0.0574	0.34	0.18	0.5	26.83	46.53
Ghana	6.06	11387.48	11.8	0.97	0.37	0.0116	73.2	0.1466	0.2452	112.1121733	0.0102	1.88	0.91	0.82	0.18	29.12
Guatemala	7.12	49255.84	14.1	1	0.32	0	80.7	0.0412	0.3908	107.7261381	0.0039	3.25	0.88	0.88	0.03	14.8
Greece	5	5822.62	7.1	0.793	0.44	0.1106	63.1	0.0553	0.1238	214.5219104	0.0522	0.45	0.89	0.91	22.43	17.95
Guinea	5.41	25485.57	10.8	1	0.45	0	84.87	0.0544	0.4687	99.9192926	0.0044	0.92	0.92	0.92	0.02	30.85
Haiti	6.43	7099.51	6.5	0.9382	0.48	0.005	73.5	0.0258	0.1121	122.2086824	0.0273	0.63	0.79	0.85	0.05	18.84
Honduras	5.25	2241.38	5.6	0.335	0.3	0.1919	60.2	0.0488	0.1418	208.8802914	0.1046	0.29	0.69	0.66	55.76	33.88
Hungary	3.61	18626.07	5.2	0.404	0.41	0.256	62.9	0.1393	0.0798	144.0152485	0.0679	0.7	0.82	0.72	26.63	24.66
Iceland	5.95	23225.99	16.5	0.956	0.54	0.0003	76.7	0.0673	0.225	132.9273286	0.0096	-0.34	0.84	0.87	0.06	44.62
Indonesia	5.94	25366.87	11.9	1	0.33	0.0083	76.3	0.0511	0.3916	113.8245134	0.0047	1.42	0.89	0.82	0.22	28.37
Ireland	7.53	48601.99	12.4	1	0.27	0	82.6	0.0298	0.37	119.8471527	0.0022	5.16	0.93	0.9	0.01	15.3
India	3.02	5886.46	6.4	0.8922	0.35	0.0663	68.9	0.0542	0.1737	154.8548331	0.041	-0.93	0.65	0.75	42.41	43.17
Indonesia	5.14	10112.03	8	0.39	0.0087	0	75	0.043	0.0866	126.7653666	0.007	-0.48	0.86	0.82	0.17	41.72
Iran	4.28	13911.05	9.8	0.9995	0.4	0.1057	76	0.1262	0.0771	388.0281378	0.0149	-1.17	0.34	0.14	0.29	66.52
Ireland	6.96	81188.49	12.5	1	0.33	0.0062	81.6	0.0837	0.2452	194.7229712	0.0036	3.03	0.97	0.91	0.03	12.4
Israel	6.93	36939.24	13	1	0.39	0	82.3	0.048	0.1593	186.1412946	0.0038	-0.77	0.85	0.8	0.09	32.58
Italy	6.52	38755.73	10.2	0.35	0.0808	0	83	0.049	0.4252	103.295318	0.003	1.6	0.92	0.94	0.02	16.99
Japan	5.79	39467.09	12.7	1	0.33	0.0184	84.1	0.0331	0.3301	103.462823	0.0027	2.7	0.89	0.9	0.02	28.67
Kazakhstan	6.01	23054.08	11.7	1	0.27	0.029	72.1	0.0496	0.1323	157.2517616	0.011	-0.02	0.35	0.22	0.17	54.55
Kyrgyzstan	4.3	5444.41	6.69	0.9973	0.27	0.091	71.1	0.0711	0.251	146.4271997	0.012	0.42	0.75	0.61	1.8	30.19
Laos	4.06	6938.69	5.2	0.8055	0.39	0.1741	64.9	0.0079	0.1243	127.833692	0.0515	1.16	0.62	0.05	0	71.58
Latvia	5.9	25508.61	12.8	1	0.34	0.0074	74.9	0.0964	0.3115	107.4961144	0.0047	2.37	0.94	0.88	0.01	17.28
Lebanon	5.17	15853.27	8.6	0.9973	0.32	0.021	78.8	0.0978	0.1337	114.8699432	0.008	0.88	0.75	0.76	0.06	31.95
Lesotho	3.55	2332.32	2.3	0.45	0.3066	0.53	62.84	0.0238	0.29	125.6271942	0.0915	0.4	0.87	0.87	0.17	28.78
Lithuania	6.31	28886.76	13	1	0.38	0.0293	75.2	0.0786	0.2999	108.5994531	0.0048	2.35	0.94	0.86	0.06	19.95
Luxembourg	7.24	86316.09	12.1	1	0.32	0.0009	81.8	0.0629	0.3819	109.5176995	0.0028	5.22	0.96	0.9	0.02	14.43
Madagascar	4.07	1676.65	6.1	0.229	0.43	0.4752	65.9	0.118	0.0924	147.9719218	0.0572	-0.34	0.79	0.86	46.53	27.84
Malawi	4.13	1084.65	6.5	0.43	0.129	0.427	63.7	0.0502	0.1332	300.8117463	0.049	0.56	0.82	0.82	0.18	29.12
Malaysia	5.34	23426.95	10.2	0.999	0.41	0.0258	75.6	0.0344	0.1355	115.1474754	0.0083	0.1	0.32	0.63	0.45	46.57
Maldives	5.2	18115.88	6.3	1	0.31	0.0079	78	0.0612	0.1964	132.7868791	0.009	0.81	0.83	0.61	1.34	34.17
Mali	4.42	2165.88	2.1	0.3079	0.36	0.1894	58	0.0757	0.1491	108.1221931	0.1405	-1.14	0.83	0.79	79.64	39.83
Malta	6.91	34217	11.3	0.29	0	0	82.1	0.0409	0.3111	108.856824	0.0066	2.71	0.88	0.92	0.01	16.99
Mauritania	6.31	4473.63	4.4	0.4132	0.33	0.1689	64.2	0.1026	0.1206	125.2291689	0.0803	0.31	0.78	0.55	37.4	24.83
Mexico	6.55	18913.81	8.6	0.995	0.48	0.0155	74.9	0.0386	0.1353	122.7889955	0.0157	-1.26	0.81	0.79	1.01	49.33
Moldova	5.48	8965.57	11.6	0.997	0.36	0.0054	83.1	0.0682	0.2483	144.9250728	0.0037	0.43	0.73	0.87	0.03	28.83
Morocco	4.9	7520.17	6.4	0.978	0.4	0.0966	76	0.093	0.2146	107.7548895	0.0225	0.5	0.82	0.6	1.03	

Bibliographie

- BANERJEE, Abhijit V. et Esther DUFLO (mars 2007). "The Economic Lives of the Poor". In : *Journal of Economic Perspectives* 21.1, p. 141-168. URL : <https://www.aeaweb.org/articles?id=10.1257/jep.21.1.141>.
- Burden of Disease from Inadequate Water, Sanitation and Hygiene for Selected Adverse Health Outcomes : An Updated Analysis with a Focus on Low- and Middle-Income Countries - PMC* (s. d.). URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6593152/>.
- Classement / RSF* (s. d.). URL : <https://rsf.org/fr/classement>.
- CUÑADO, Juncal et Fernando PÉREZ DE GRACIA (1^{er} août 2012). "Does Education Affect Happiness? Evidence for Spain". In : *Social Indicators Research* 108.1, p. 185-196. URL : <https://doi.org/10.1007/s11205-011-9874-x>.
- DALY, Herman E. (30 sept. 2014). *Beyond Growth : The Economics of Sustainable Development*. Beacon Press. 268 p. ISBN : 978-0-8070-4706-4.
- DEVOTO, Florencia et al. (mai 2012). "Happiness on Tap : Piped Water Adoption in Urban Morocco". In : *American Economic Journal : Economic Policy* 4.4, p. 68-99. URL : <https://www.aeaweb.org/articles?id=10.1257/pol.4.4.68>.
- DI TELLA, Rafael et Robert MACCULLOCH (1^{er} avr. 2008). "Gross National Happiness as an Answer to the Easterlin Paradox?" In : *Journal of Development Economics* 86.1, p. 22-42. URL : <https://www.sciencedirect.com/science/article/pii/S0304387807000600>.
- Eau Potable* (21 sept. 2020). URL : <http://uis.unesco.org/fr/glossary-term/eau-potable-0>.
- Enfants : améliorer leur survie et leur bien-être* (s. d.). URL : <https://www.who.int/fr/news-room/fact-sheets/detail/children-reducing-mortality>.
- HAN, Jin, Qingxia ZHAO et Mengnan ZHANG (1^{er} mars 2016). "China's Income Inequality in the Global Context". In : *Perspectives in Science*. 1st Czech-China Scientific Conference 2015 7, p. 24-29. URL : <https://www.sciencedirect.com/science/article/pii/S2213020915000518>.

- HILLENBRAND, Olaf et Peter THIERY (s. d.). "The Bertelsmann Transformation Index 2006 : On the Way to Democracy and Market-Economy Strategic Insights, Volume IV, Issue 12 (December 2005)". In : (), p. 16.
- Les Différentes Mesures Du Bien-Être Subjectif* | Cairn.Info (s. d.). URL : <https://www.cairn.info/revue-francaise-d-economie-2012-3-page-35.htm>.
- Les Trois Piliers Du Développement, Condition Essentielles Pour Réaliser Le Bien-Être Des Peuples, Affirme Ban* || 10NU Info (s. d.). URL : <https://news.un.org/fr/story/2013/03/263422>.
- MÉDA, Dominique (2012). "Comment Le Pib a Pris Le Pouvoir". In : *Revue Projet* 331.6, p. 14-21. URL : <https://www.cairn.info/revue-projet-2012-6-page-14.htm>.
- Méthodologie Détaillée Du Classement Mondial de La Liberté de La Presse 2022* | RSF (s. d.). URL : <https://rsf.org/fr/index-methodologie-2022>.
- Methodology* (s. d.). BTI 2022. URL : <https://bti-project.org/en/methodology>.
- NATIONS, United (s. d.). *Human Development Index*. United Nations. URL : <https://hdr.undp.org/data-center/human-development-index>.
- NIKOLAEV, Boris et Pavel RUSAKOV (12 août 2016). "Education and Happiness : An Alternative Hypothesis". In : *Applied Economics Letters* 23.12, p. 827-830. DOI : 10.1080/13504851.2015.1111982. URL : <https://doi.org/10.1080/13504851.2015.1111982>.
- Que Nous Apprend Le « Bonheur National Brut » ?* | Cairn.Info (s. d.). URL : <https://www.cairn.info/revue-projet-2018-1-page-26.htm#pa9>.
- RAISKILA, Erja (29 avr. 2016). *La Finlande En Tête de La Liberté de La Presse*. URL : <https://finland.fi/fr/vie-amp-societe/la-finlande-tete-de-la-liberte-de-la-presse/>.
- SABATINI, Fabio (1^{er} août 2014). "The Relationship between Happiness and Health : Evidence from Italy". In : *Social Science & Medicine* 114, p. 178-187. URL : <https://www.sciencedirect.com/science/article/pii/S0277953614003189>.
- SHYU, Chian-Woei (1^{er} avr. 2014). "Ensuring Access to Electricity and Minimum Basic Electricity Needs as a Goal for the Post-MDG Development Agenda after 2015". In : *Energy for Sustainable Development* 19, p. 29-38. URL : <https://www.sciencedirect.com/science/article/pii/S097308261300094X>.
- TELLA, Rafael Di, Robert J. MACCULLOCH et Andrew J. OSWALD (1^{er} nov. 2003). "The Macroeconomics of Happiness". In : *The Review of Economics and Statistics* 85.4, p. 809-827. URL : <https://doi.org/10.1162/003465303772815745>.
- WDR Reports* (s. d.). World Bank. URL : <https://www.worldbank.org/en/publication/wdr/wdr-archive>.

- WHITAKER, Celina (2018). "Que Nous Apprend Le « Bonheur National Brut » ?" In : *Revue Projet* 362.1, p. 26-30. URL : <https://www.cairn.info/revue-projet-2018-1-page-26.htm>.
- ZEIDAN, Jinan (2012). "Les Différentes Mesures Du Bien-Être Subjectif". In : *Revue française d'économie* XXVII.3, p. 35-70. URL : <https://www.cairn.info/revue-francaise-d-economie-2012-3-page-35.htm>.

Table des figures

2.1	Corrélation entre la liberté de la presse et le niveau de bonheur	5
2.2	Corrélation entre la liberté d'expression et le niveau de bonheur	7
2.3	Corrélation entre la liberté d'association et le niveau de bonheur	9
2.4	Corrélation entre l'espérance de vie et le niveau de bonheur	10
2.5	Corrélation entre le PIB par habitant et le niveau de bonheur	12
2.6	Corrélation entre l'accès à l'électricité et le niveau de bonheur	13
2.7	Corrélation entre le nombre d'années d'études et le niveau de bonheur	15
2.8	Corrélation entre la mortalité infantile et le niveau de bonheur	16
2.9	Corrélation entre l'accès à l'eau et le niveau de bonheur	18
2.10	Corrélation entre l'exposition à de mauvaises infrastructures sanitaires et le niveau de bonheur.	19
2.11	Matrice des corrélations	20
3.1	Boîtes à moustaches	24
3.2	Histogramme des valeurs propres	27
3.3	Coordonnées de nos variables	28
3.4	Contributions de nos variables	29
3.5	Cosinus carrés de nos variables	31
3.6	Cercle de corrélation : Modèle initial	32
3.7	Cercle de corrélation : Modèle final	36
3.8	Cercle de corrélation du modèle final avec la variable illustrative	39
3.9	Projection de tous nos individus sur les axes 1 et 2 : Noms des pays	44
3.10	Projection de tous nos individus sur les axes 1 et 2	45
3.11	Projections des individus extrêmes : Numéros	47
3.12	Projections des individus extrêmes : Noms des pays	48
3.13	Graphique récapitulatif : Individus	49

4.1	Graphiques des résidus des axes 1 et 2 : Boîte à moustaches, Histogramme, Diagramme quantile-quantile	53
4.2	Projection des cercles des cercles de corrélations supplémentaires	56
4.3	Projection de nos valeurs observées par rapport aux valeurs prédites	58
4.4	Projection graphique de notre RLB	62
6.1	Cercle de corrélation initial entre les Axes 1 et 3	67
6.2	Cercle de corrélation intermédiaire 1 entre Axes 1 et 2	67
6.3	Boîte à moustaches : Régression Axes 1/3 et 2/3	69
6.4	Histogrammes : Régression Axes 1/3 et 2/3	69
6.5	QQplots : Régression Axes 1/3 et 2/3	70
6.6	Représentation graphique supplémentaire : RLB	70


Liste des tableaux

3.1	Statistiques descriptives de nos variables	22
3.2	Tableau des valeurs propres	26
3.3	Contributions de toutes nos variables	33
3.4	Cosinus carrés de toutes nos variables	34
3.5	Tableau récapitulatif : Modèle initial	35
3.6	Tableau récapitulatif : Modèle final	37
3.7	Coordonnées de nos dix premiers individus	40
3.8	Contributions de nos dix premiers individus	42
3.9	Cosinus carrés de nos dix premiers individus	42
4.1	RLM sur Axes 1 et 2	52
4.2	Test de Shapiro	54
4.3	Test de Kolmogorov-Smirnov	54
4.4	Test de Breush-Pagan	54
4.5	Test de Ramsey	55
4.6	Global Validation of Linear Model Assumptions	55
4.7	Tableau comparatif	57
4.8	Statistiques descriptives du niveau de bonheur prédit entre l'axe 1 et 2	59
4.9	Statistiques descriptives du niveau de bonheur observé	59
4.10	RLB au seuil médian de 5.5	60
4.11	RLB au seuil médian de 6.93	61
6.1	Tests de Grubbs supplémentaires	66
6.2	Tests de Shapiro-Wilk supplémentaires	66
6.6	Coordonnées des individus supplémentaires : Guatemala	67
6.7	Cosinus carrés des individus supplémentaires : Iran	67
6.3	Cosinus carrés supplémentaires : sans Taxes et Droits Humains	68

6.4	Valeurs propres supplémentaires : Modèle intermédiaire	68
6.5	Valeurs propres supplémentaires : Modèle final	68
6.8	Résultats supplémentaires : RLM sur Axes 1 et 2	68
6.9	RLB au seuil de 5.8	71
6.10	RLB au seuil de 6.3 (Q3)	71
6.11	Base de donnée	88

Table des matières

Sommaire	i
1 Introduction	1
1.0.1 Méthodologie	2
2 Analyse détaillé du sujet	3
2.0.1 Pertinence des variables à expliquer	3
2.0.2 Rôle des variables explicatives	4
2.0.3 Corrélation entre la variable illustrative et les variables actives	20
3 Analyse en composante principale	22
3.0.1 Statistiques descriptives	22
3.0.2 Valeurs propres et nombre d'axes	25
3.0.3 Variables, contribution, corrélations, cosinus carrés	27
3.0.4 Cercle de corrélation	32
3.0.5 Définition des variables latentes	37
3.0.6 Projection des individus sur le plan factoriel, interprétation	39
4 Corrélation entre la variable illustrative et les variables latentes	51
4.0.1 Régression linéaire multiple	51
4.0.2 Relation entre la variable illustrative observée et prédite par la régression linéaire multiple	58
4.0.3 Régression logistique binaire	59
5 Conclusion et discussions des résultats	64
5.0.1 Conclusion	64
5.0.2 Discussion	64

6	Annexes	66
6.0.1	Figures et Tables supplémentaires	66
6.0.2	Script 	72
6.0.3	Base de donnée	86
	Bibliographie	89
	Table des matières	89