

Panongbene Jean Mohamed Sawadogo

M2 Data Science

Visual Analytics

Assignment 3 : Exploratory Analysis

I. Data Descriptions

Notre ensemble de données est constitué de données que nous avons scrapées depuis le site web <https://www.populationdata.net/pays/>.

Ces données ont été nettoyées et sont représentées sous forme d'un tableau de 253 lignes et de 41 colonnes contenant différentes informations concernant les pays du monde. Chaque ligne représente un pays et chaque colonne représente une information concernant les pays.

Column	Nom officiel	Nom propre	Continent	Sous-continent	Population(habitants)	Date du recensement de la population	Croissance démographique(% / an)	Superficie(km²)
Abkhazie	République d'Abkhazie	Аҧсны Аҳәынҭқарра (ab) აფხაზეთი (ka) Республ...	Asie	Asie de l'Ouest	246 313	2018	0,49	8 653
Açores	Région autonome des Açores	Região Autónoma dos Açores (pt)	Europe	Union Européenne	243 862	2018	-0,39	2 322
Afghanistan	République islamique d'Afghanistan	د افغانستان اسلامي جمهوریت (ps) جمهوری اسلامی...	Asie	Asie centrale	34 233 212	2020	6,23	652 864
Afrique du Sud	République d'Afrique du Sud	Republiek van Suid-Afrika (af) Republic of So...	Afrique	Afrique Australe	57 725 606	2018	2,13	1 219 912
Albanie	République d'Albanie	Republika e Shqipërisë (sq)	Europe	Europe (hors Union Européenne)	2 870 324	2018	-0,22	28 748
Algérie	République algérienne démocratique et populaire	الجمهورية الجزائرية الديمقراطية الشعبية (ar)	Afrique	Afrique du Nord	43 820 839	2020	1,90	2 381 741
Allemagne	République Fédérale d'Allemagne	Bundesrepublik Deutschland (de)	Europe	Union Européenne	83 042 235	2019	0,30	357 386
Andorre	Principauté d'Andorre	Principat d'Andorra (ca)	Europe	Europe (hors Union Européenne)	76 522	2018	2,31	468

Nous allons utiliser ces données pour essayer de répondre aux questions ci-dessous.

II. Initial Analysis Questions :

Les questions auxquelles nous allons essayer de répondre sont les suivantes :

Quels sont les pays et les régions du monde où le taux d'alphabétisation est le plus élevé ?

Quels sont les facteurs qui expliquent le mieux le taux de mortalité infantile dans les pays du monde ?

Quels sont les facteurs qui expliquent le mieux la croissance démographique par an dans les pays du monde ?

III. Exploratory Analysis

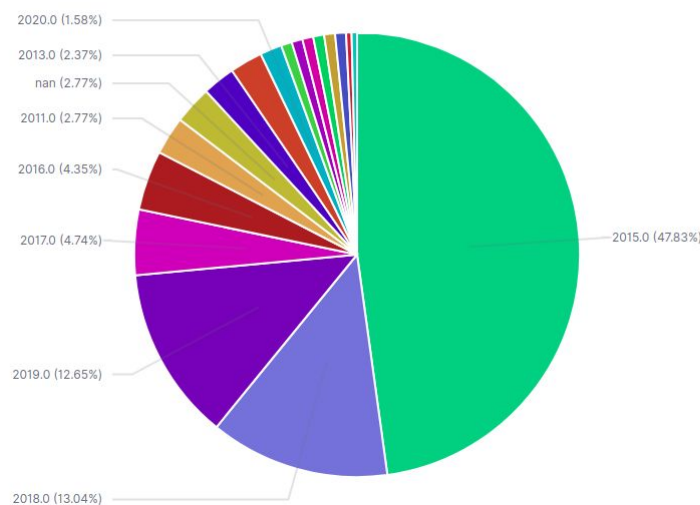
1. Quels sont les pays et les régions du monde ou le taux d'alphabétisation est le plus élevé ?

Le taux d'alphabétisation dans un pays est défini comme étant le pourcentage de la population qui sait lire et écrire dans une langue donnée. La population considérée est la population adulte et adolescente âgée de 15 ans au moins.

L'analyse qui a été effectuée dans ce travail est basée sur les données statistiques tirées du site web <https://www.populationdata.net/pays/>.

a) Qualité des données :

● 2015.0 ● 2018.0 ● 2019.0 ● 2017.0 ● 2016.0 ● 2011.0 ● nan ● 2013.0 ● 2014.0 ● 2020.0 ● 2001.0 ● 2003.0 ● 2007.0 ● 2008.0 ● 2009.0 ● 2010.0 ● 2002.0 ● 2005.0



Date de prélèvement des statistiques sur le taux d'alphabétisation
(image réalisée avec kibana et elasticsearch)

Les données qui ont été scrapées depuis le site web <https://www.populationdata.net/pays/> ont été prélevées entre 2011 et 2020 mais la majorité des données ont été prélevées en 2015 (environ 47.83% des données) comme on peut le constater sur l'image ci-dessus.

Donc ces statistiques ne sont pas uniformes dans le temps.

On constate aussi que 2.77% de nos lignes ne disposent pas de l'attribut taux d'alphabétisation.

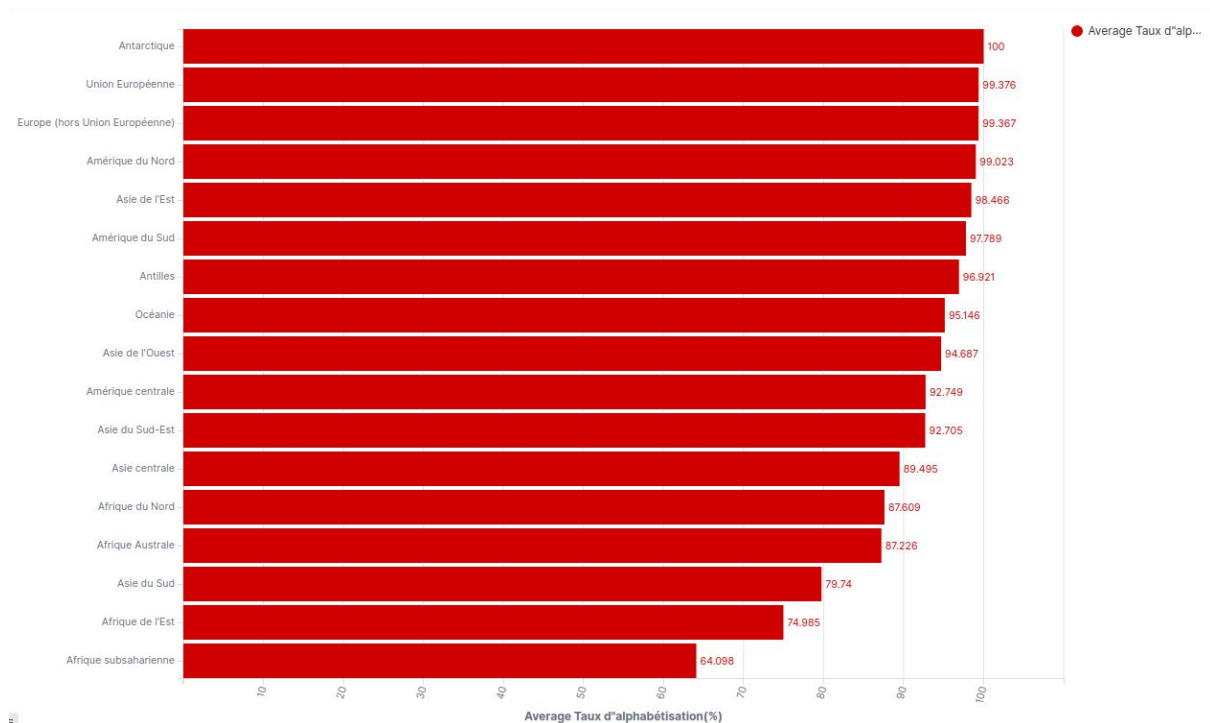
7 hits			
Nom propre		Taux d'alphabétisation(%)	Column
> Аҟсны Аҳәынҭқарра (ab) აფხაზეთი (ka) Республика Абхазия (ru)		NaN	Abkhazie
> Kurdewari (ku) Kurdistan (ku) کوردستان (ku)		NaN	Kurdistan
> Республика Хуссар Ирыстон (os) бӕддӕтӕ нӕгъон ӕрӕмбӕрӕтӕ (ka) Республика Южная Осетия (ru)		NaN	Ossétie du Sud-Alanie
> Caribisch Nederland BES-eilanden (nl)		NaN	Pays-Bas caribéens
> ارم البنط (ar) Ārd al-Bunṭ (ar)		NaN	Pount
> الجمهورية العربية الصحراوية الديمقراطية (ar) República árabe saharawi democrática (es)		NaN	Sahara occidental
> Република Молдовеняскэ Нистрянэ (mo) Придністровська Молдавська Республіка (uk) Приднестровская Молдавская Республика (ru)		NaN	Transnistrie

Pays dont le taux d'alphabétisation est manquante
(image réalisée avec kibana et elasticsearch).

Ces données manquantes sont en effet au nombre de 7 et représentent le taux d'alphabétisation des pays de l'Abkhazie, du Kurdistan, de l'Ossétie du Sud-Alante, des Pays-Bas Caribéens, de Pount, du Sahara occidental et de la Transnistrie comme on peut le voir dans l'image ci-dessus.

Nous pouvons ajouter comme commentaire qu'en faisant un peu de recherche sur internet, on se rend compte que ces pays sont en effet des pays ayant une reconnaissance internationale ambiguë.

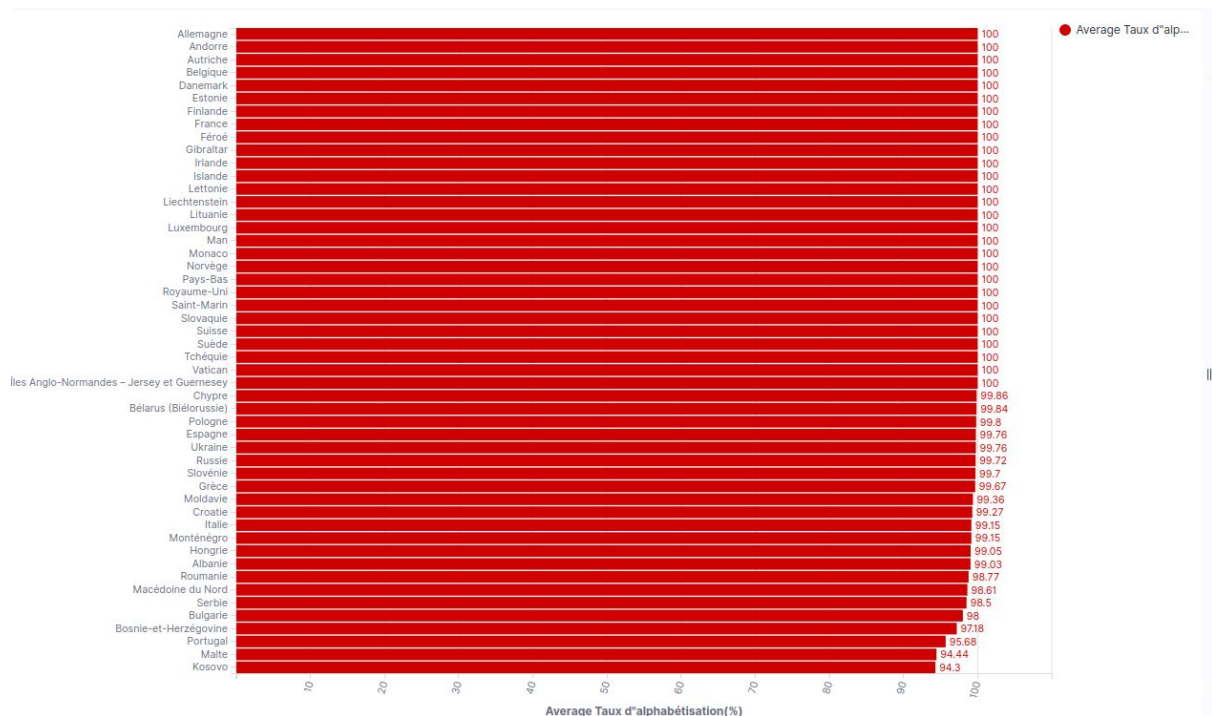
b) Le taux d'alphabétisation dans les sous-continent



Taux d'alphabétisation par sous-continent
(image réalisée avec kibana et elasticsearch).

À travers l'image ci-dessous, nous pouvons constater que le taux d'alphabétisation dans les sous-continent du monde est hyper élevé dans certaines parties du monde comme les pays de l'Union européens et très bas dans d'autres.

c) Taux D'alphabétisation en Europe

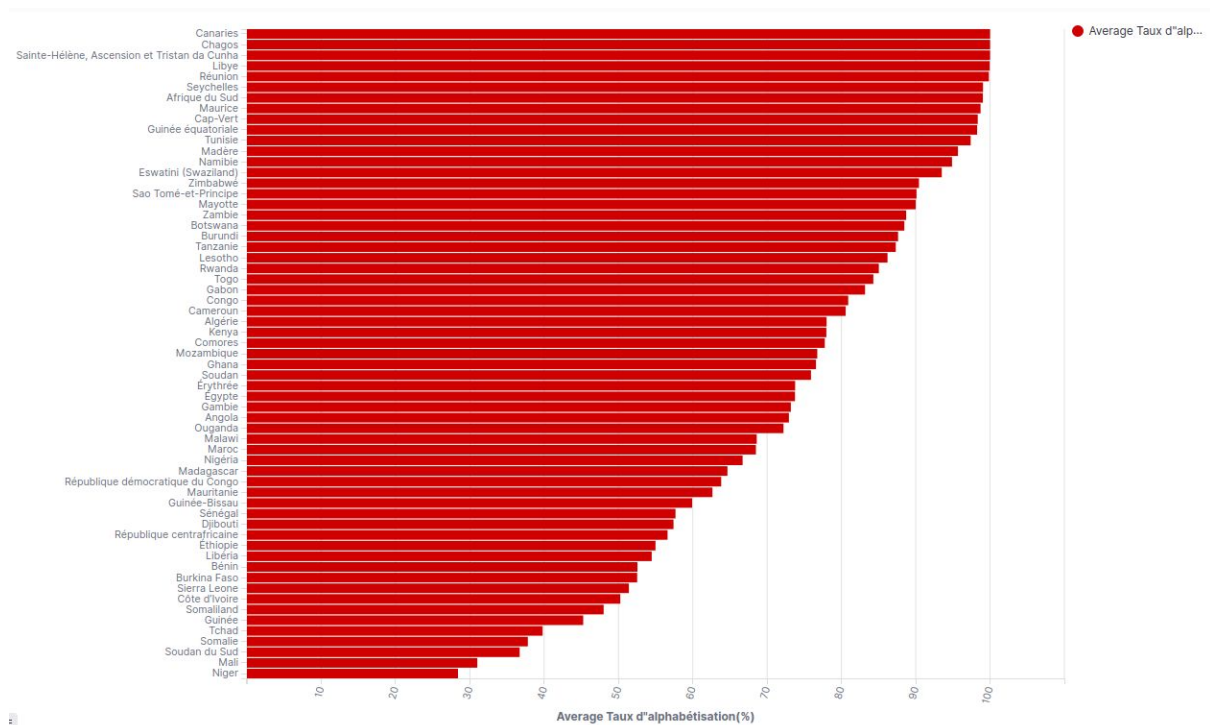


Taux d'alphabétisation au sein du continent Européen
(image réalisée avec kibana et elasticsearch).

Au sein du continent européen (constitué des pays appartenant aux sous-continent Union Européenne et hors Union Européenne), on peut dire que le taux d'alphabétisation est presque le même pour tous les pays. En effet ce taux d'alphabétisation avoisine les 100% pour presque tous les pays à l'exception de certains pays d'Europe de l'est où le taux d'alphabétisation est au-dessous de 98%.

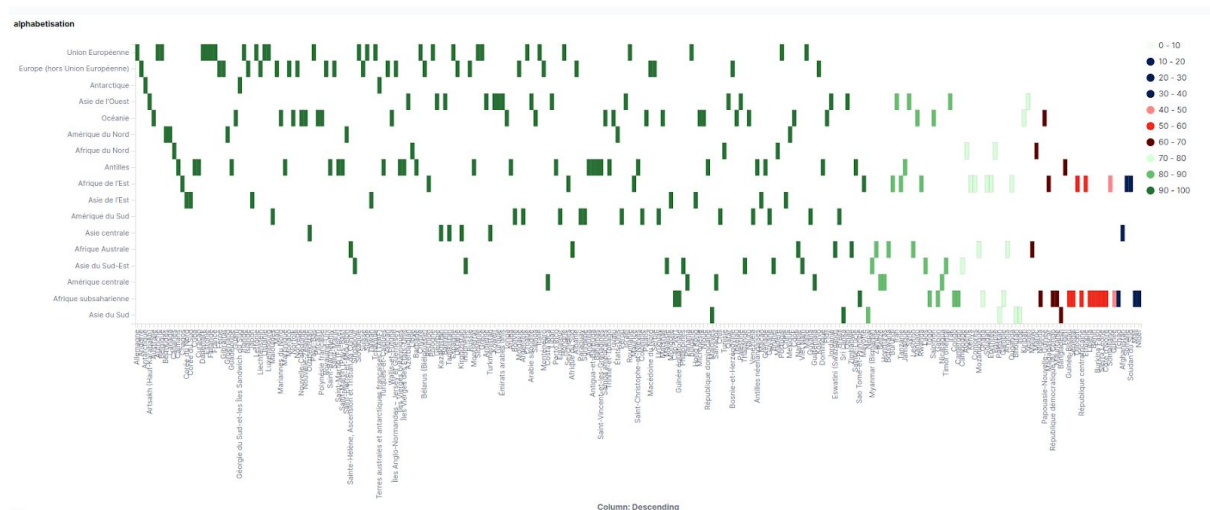
d) Taux D'alphabétisation en Afrique

Contrairement à l'Europe, on voit de forte variation au sein du taux d'alphabétisation en Afrique avec certains pays qui ont des taux de 100% et d'autres qui sont au-dessous de 60% comme le montre l'image ci-dessous.



Taux d'alphabétisation en Africain
(image réalisée avec kibana et elasticsearch).

e) Conclusion



Taux d'alphabétisation dans le monde
(image réalisée avec kibana et elasticsearch).

L'analyse effectuée ci-dessous nous montre que le taux d'alphabétisation à travers le monde est globalement élevé en moyenne.

Mais la répartition spatiale présente de grosses inégalités : en effet dans certaines parties du monde le taux d'alphabétisation avoisine les 100% alors que dans d'autres pays le taux d'alphabétisation est au-dessous de 60% comme on peut le voir sur la figure ci-dessus.

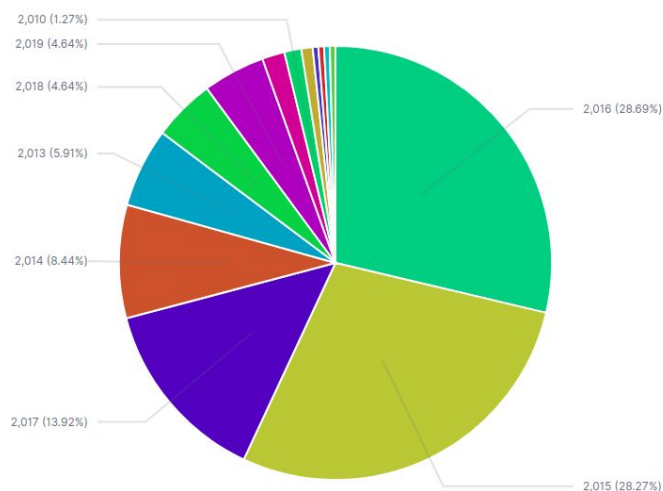
2. Quels sont les facteurs qui expliquent le mieux le taux de mortalité infantile dans les pays du monde ?

Le taux de mortalité infantile est défini comme le rapport entre le nombre d'enfants morts avants l'âge de 1 an sur le nombre total d'enfants nés vivants.

L'analyse qui a été effectuée dans ce travail est basée sur les données statistiques tirées du site web <https://www.populationdata.net/pays/>.

a) Qualité des données:

2,016 2,015 2,017 2,014 2,013 2,018 2,019 2,012 2,010 2,011 2,005 2,008 2,009 2,020



Date de prélèvement du taux de mortalité infantile dans le monde
(image réalisée avec kibana et elasticsearch).

Les données qui ont été scrapées depuis le site web <https://www.populationdata.net/pays/> ont été prélevées entre 2005 et 2019 mais la majorité des données ont été prélevées en 2015, 2016 et en 2017 (environ 70.88% des données) comme on peut le constater sur l'image ci-dessus. **Donc ces statistiques ne sont pas uniformes dans le temps.**

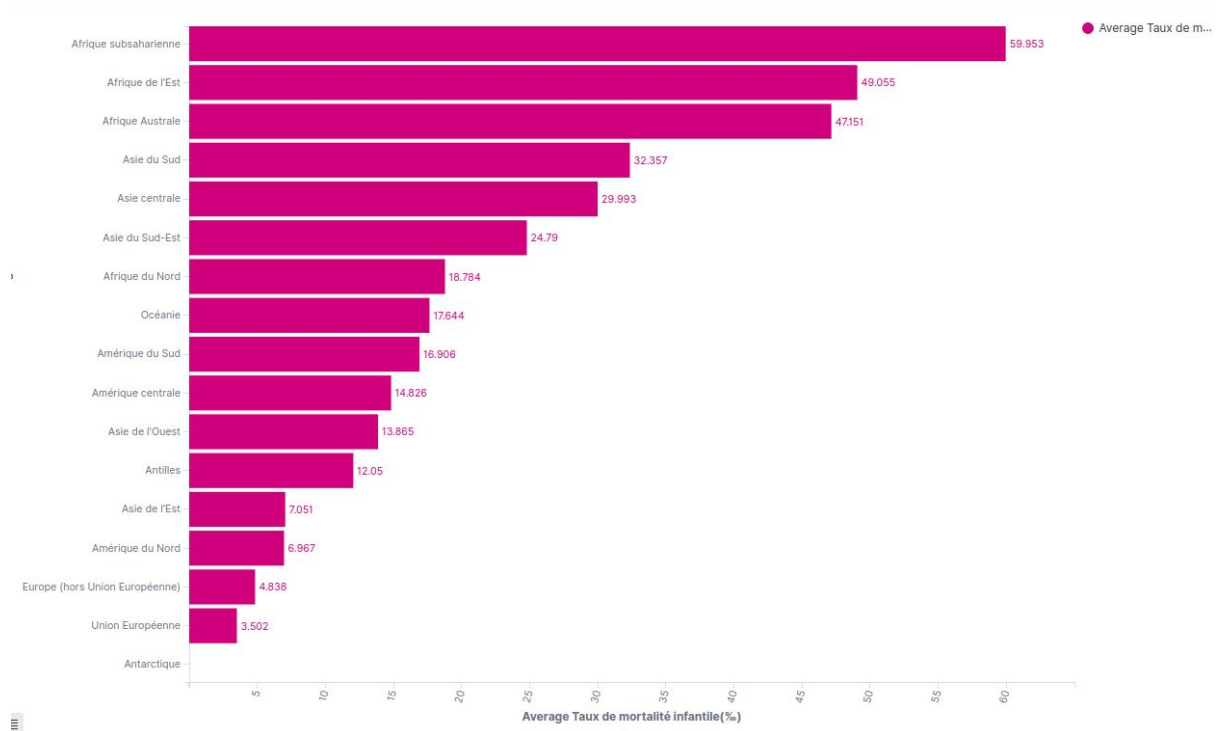
16 hits

Column
> Abkhazie
> Antarctique
> Chagos
> Cook
> Géorgie du Sud-et-les îles Sandwich du Sud
> Kurdistan
> Malouines
> Montserrat
> Norfolk
> Ossétie du Sud-Alanie
> Pays-Bas caribéens
> Pitcairn
> Saint-Barthélemy
> Terres australes et antarctiques françaises (TAAF)
> Transnistrie
> Vatican

Les pays ne disposant pas de statistique sur le taux de mortalité infantile dans le monde (image réalisée avec kibana et elasticsearch).

On constate aussi que 16 pays (lignes) ne disposent pas de l'attribut taux mortalité infantile. Mais comme pour la précédente analyse, ces pays ont des dispositions de reconnaissance internationale particulières et donc n'influent pas beaucoup sur notre analyse.

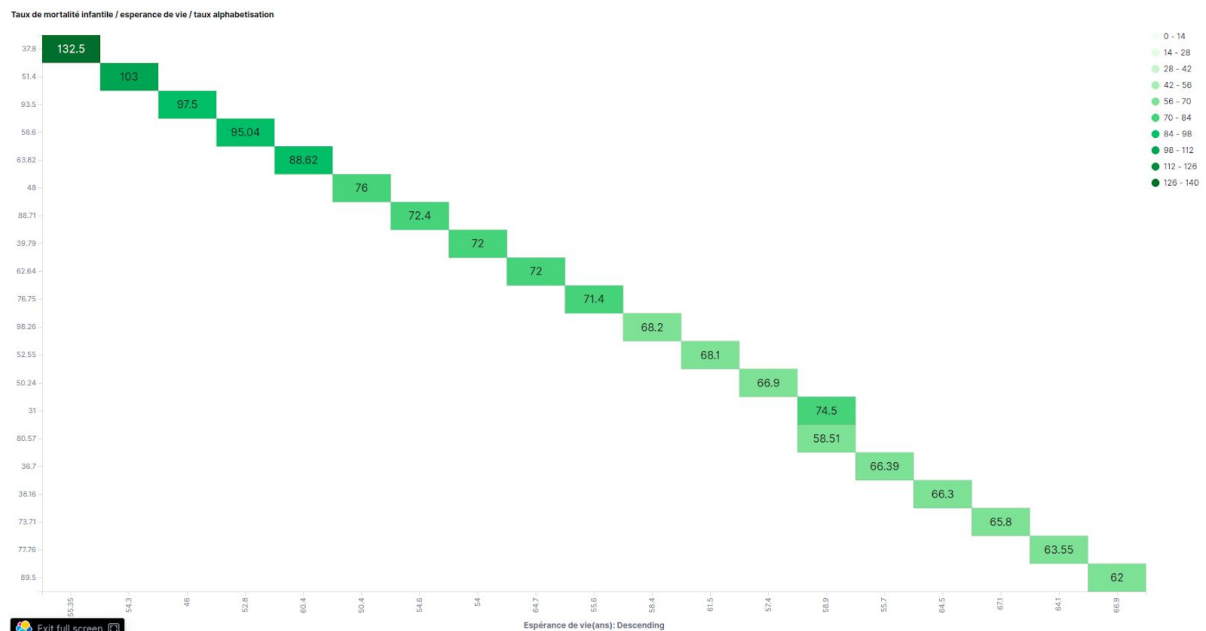
b) Répartition Géographique du Taux de mortalité Infantile dans les sous-continent



Répartition du taux de mortalité infantile dans les sous-continent (image réalisée avec kibana et elasticsearch)

Sur l'image ci-dessus, on peut constater que le taux de mortalité infantile à travers les régions du monde présente de fortes disparités. On peut constater qu'il est juste de 3.502 pour les milles dans les pays de l'union européenne alors qu'il est de 59.953 pour 1000 dans les pays d'Afrique Subsaharienne.

c) Taux de mortalité de mortalité infantile / espérance de vie et Taux d'alphabétisation



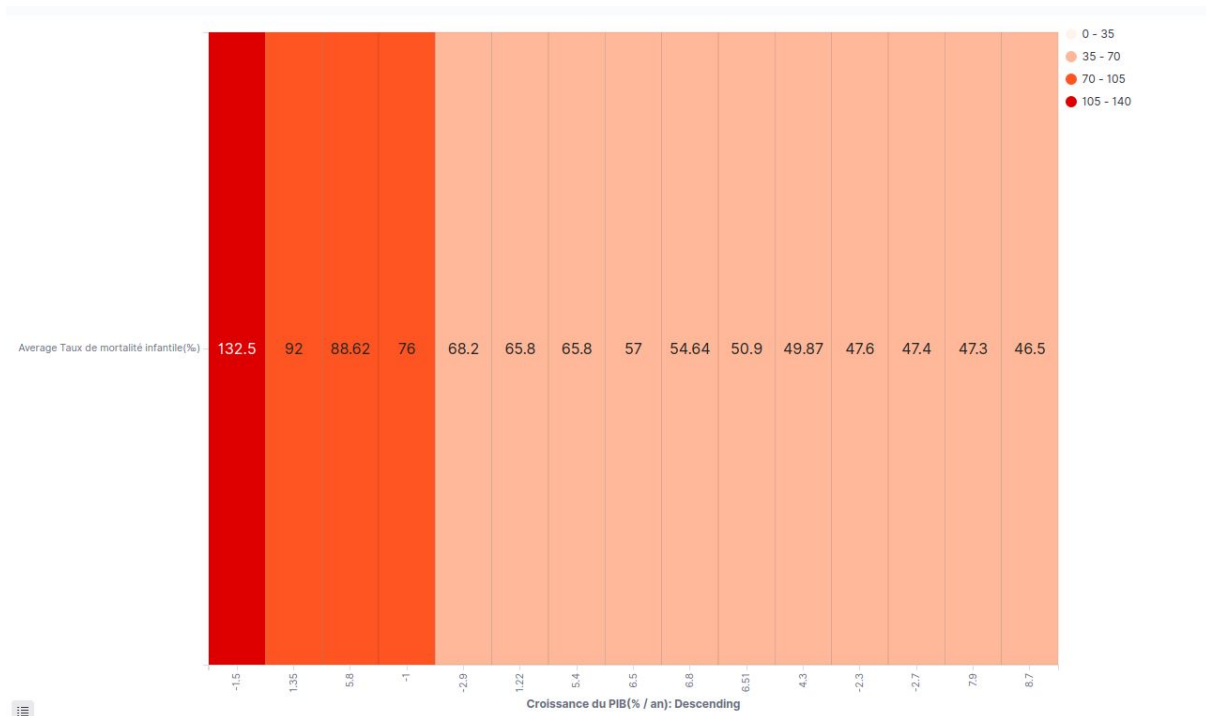
Relation entre le taux de mortalité infantile, l'espérance de vie et le taux d'alphabétisation (image réalisée avec kibana et elasticsearch).

Dans l'image ci-dessus, on peut constater que le taux de mortalité infantile (valeur à l'intérieur des cases vertes) est d'autant plus élevé que l'espérance de vie est faible (axe des abscisses) et le taux d'alphabétisation est faible (axe des ordonnées).

d) Taux de mortalité Infantile / Croissance du PIB

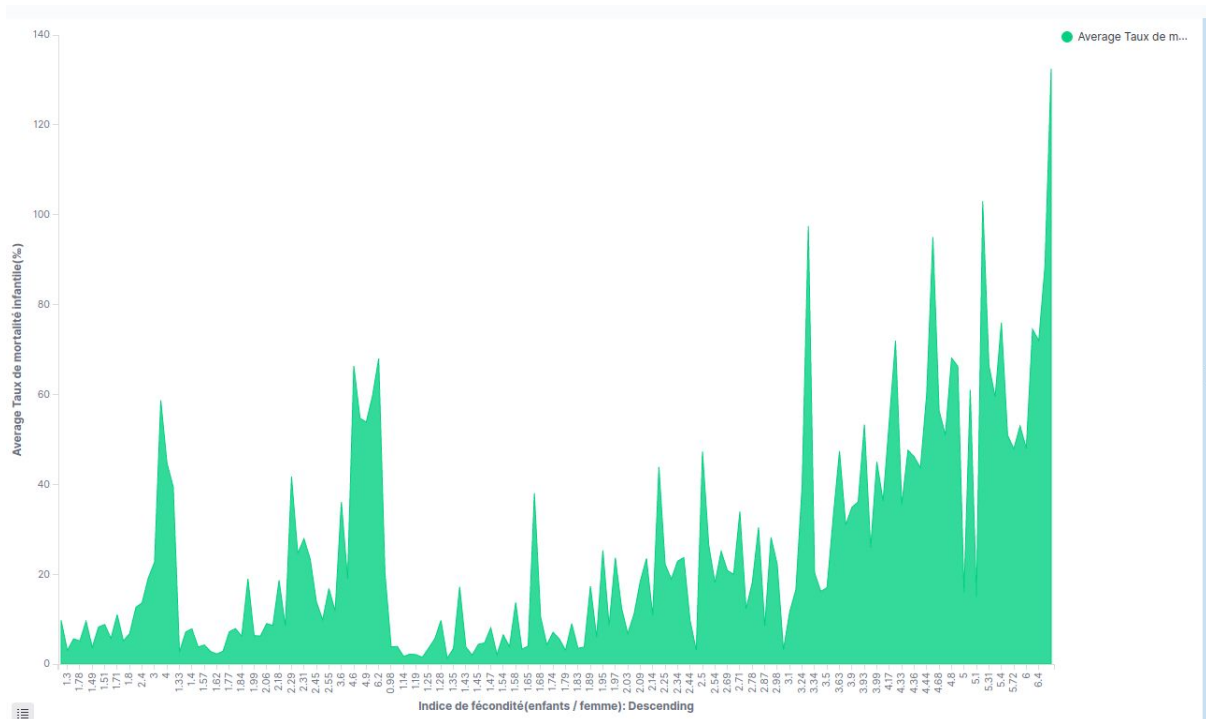
L'image ci-dessous nous montre le taux de mortalité infantile en fonction de la croissance du PIB.

On constate ainsi qu'il n'y a pas de corrélation entre le taux de mortalité infantile et la croissance du PIB. En effet certains pays ont une croissance du PIB très faible et un taux de mortalité infantile très faible aussi et d'autres une croissance du PIB très élevé et un taux de mortalité infantile élevé aussi donc on ne peut pas tirer de conclusion sur le taux de mortalité infantile en se basant sur la croissance du PIB.



Relation entre le taux de mortalité infantile, et la croissance du PIB
(image réalisée avec kibana et elasticsearch).

e) Taux de mortalité Infantile / Indice de fécondité (enfants / femme)



Relation entre le taux de mortalité infantile, et l'indice de fécondité
(image réalisée avec kibana et elasticsearch).

La courbe ci-dessus nous montre la relation entre le taux de mortalité infantile en fonction de l'indice de fécondité dans les pays du monde. Malgré quelques irrégularités, cette courbe nous montre que plus le taux de fécondité est élevé, plus le taux de mortalité infantile est élevé.

f) Conclusion

La visualisation effectuée ci-dessus, nous permet de constater sans faire une analyse statistique poussée que le taux de mortalité infantile est plus élevé dans certaines régions du monde que dans d'autres.

Aussi cette visualisation nous permet de dire que le taux de mortalité infantile est en corrélation avec d'autres facteurs comme par exemple le taux de fécondité..

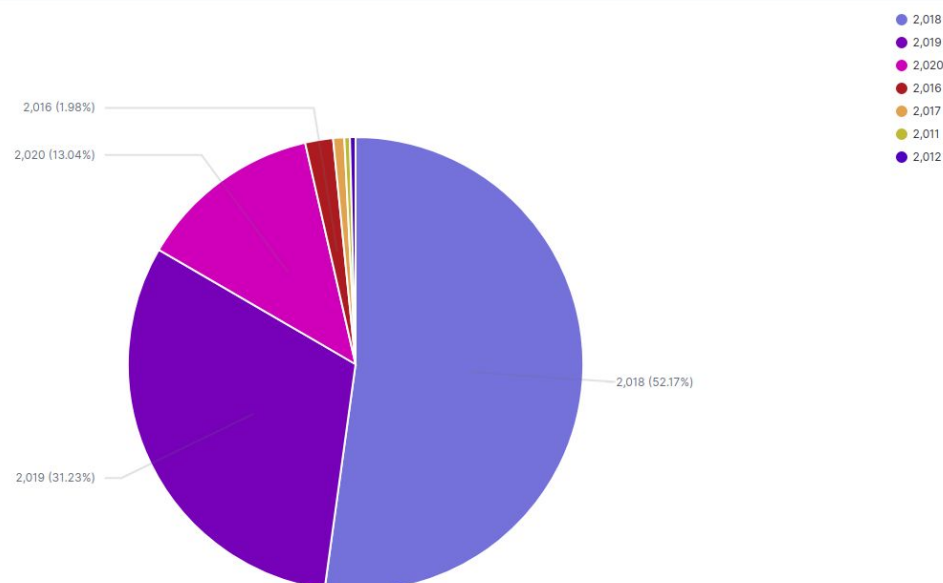
3. Quels sont les facteurs qui expliquent le mieux la croissance démographique par an dans les pays du monde ?

La croissance démographique est définie comme étant la différence entre la population au début d'une année et à la fin de l'année dans un pays donnée.

En général pour une meilleure normalisation, (comme c'est le cas dans nos données) la croissance démographique est représentée sous forme de pourcentage représentant ainsi le pourcentage de la population qui a augmenté ou diminué au cours d'une année.

L'analyse qui a été effectuée dans ce travail est basée sur les données statistiques tirées du site web <https://www.populationdata.net/pays/>.

a) Qualité des données:



Date de Recensement de la population
(image réalisée avec kibana et elasticsearch).

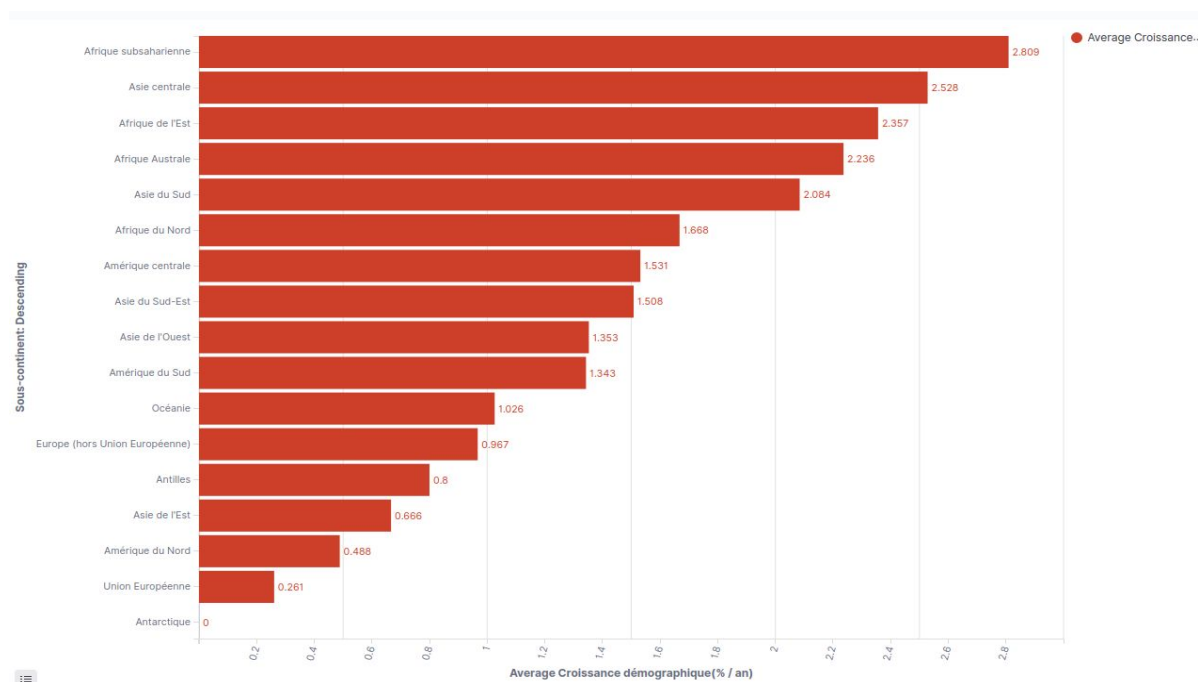
Les données relatives à la croissance de la population ont été prélevées en même temps que les données concernant la population et cela a été fait entre 2011 et 2019. Mais la majorité des données environ 52.17% ont été prélevées en 2018 et 31.23 en 2019 comme on peut le constater dans l'image ci-dessus. **Donc ces statistiques ne sont pas uniformes dans le temps.**

Croissance démographique(% / an)	10 hits
> 0	Nom officiel
> 0	Antarctique
> 0	Archipel des Chagos
> 0	Géorgie du Sud-et-les Îles Sandwich du Sud
> 0	Jamaïque
> 0	Kurdistan
> 0	Monténégro
> 0	Sultanat d'Oman
> 0	Pitcairn, Henderson, Ducie et Oeno
> 0	Terres australes et antarctiques françaises
> 0	République de Trinité-et-Tobago

Ligne présentant une croissance de la population égale à 0.
(image réalisée avec kibana et elasticsearch).

Nos données présentent 10 lignes dont la croissance de la population est égale à 0. En effet ce 0 représente l'absence de données et concerne des pays à statut particulier à l'exception de la Jamaïque du Kurdistan, du sultanat d'Oman et de la république de Trinité-et-Tobago.

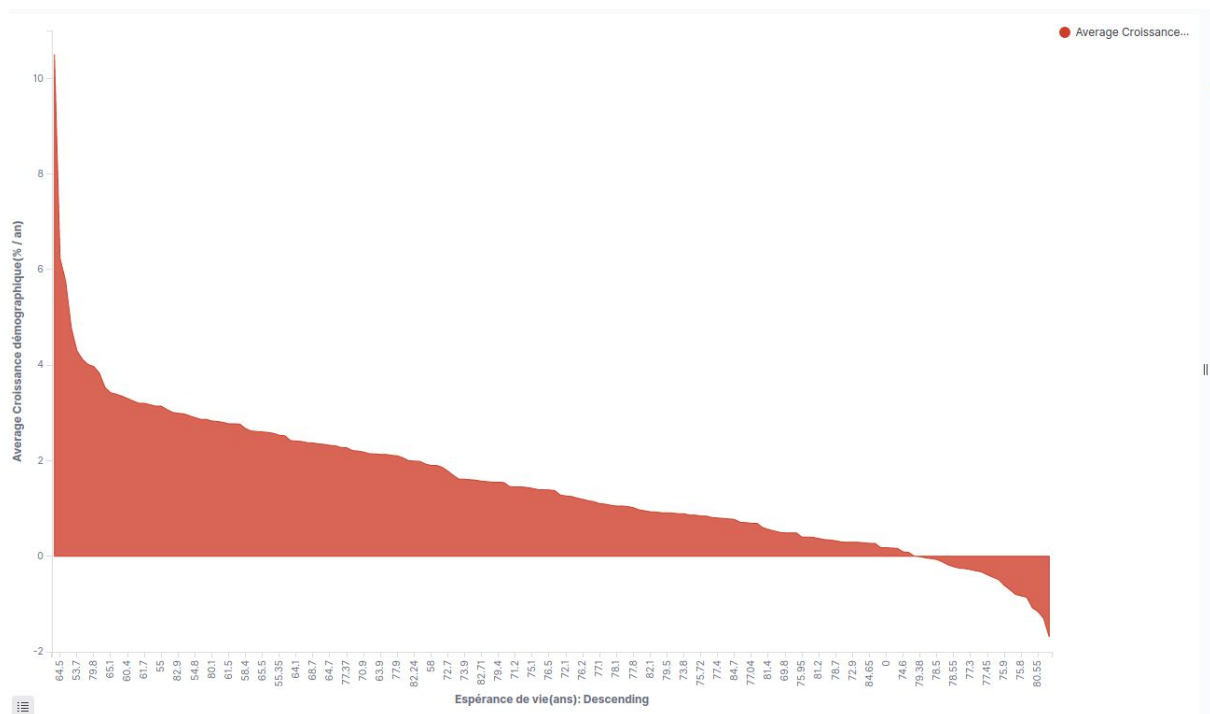
b) Répartition Géographique de la croissance démographique dans les sous-continent



Répartition de la croissance de la population dans les sous-continent
(image réalisée avec kibana et elasticsearch)

Dans cette image on constate que le taux de croissance de la population est très fort dans certaines zones du monde et très faible dans d'autres parties du monde. On peut ainsi dire qu'il y a une corrélation spatiale avec le taux de croissance de la population dans le monde.

c) Croissance de la population / Espérance de vie(ans)

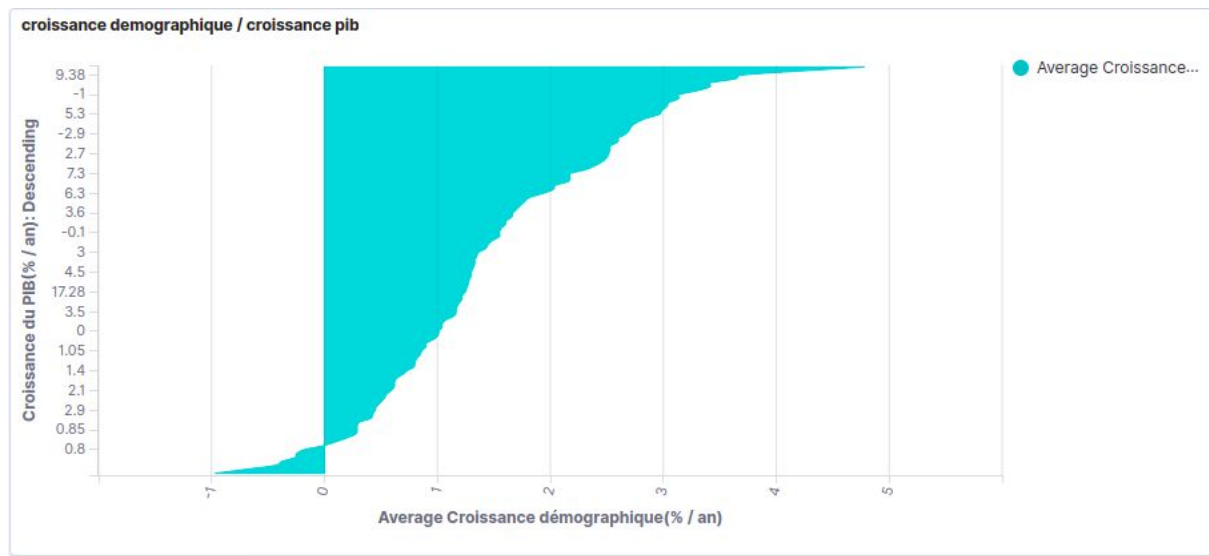


Croissance de la population en fonction de l'espérance de vie
(image réalisée avec kibana et elasticsearch)

Dans cette image on constate que le taux de croissance de la population est d'autant plus grand que l'espérance de vie est faible. Même s'il y a quelques exceptions, on constate plutôt une relation décroissante entre le taux de croissance de la population et l'espérance de vie.

d) Croissance de la population / Croissance PIB

Dans cette image, on constate empiriquement que le taux de croissance de la population n'a pas une très grande corrélation avec la croissance du PIB. En effet, il y a certains pays qui ont une forte croissance démographique avec une forte croissance du PIB et vice-versa.



Croissance de la population en fonction de la croissance du PIB
(image réalisée avec kibana et elasticsearch)

e) Conclusion

On constate empiriquement dans les images ci-dessus que le taux de croissance de la population est un facteur de la position géographique et il n'y a pas de conclusion particulière qu'on peut faire entre le taux de croissance de la population et le taux de croissance du PIB.

IV. ConClusion

La visualisation qui a été effectuée dans ce travail grâce à l'outil Elasticsearch/Kibana nous permet de tirer des conclusions sur un ensemble de fait lié au pays du monde sans faire une analyse statistique.