

PROJET DATASANTÉ

Comment les disparités géographiques influencent elles
le taux de mortalité du cancer dans le monde ?

PRÉSENTÉ PAR ARTHUR BARDOT

Table des matières

1 | Choix du jeu de données

2 | Qualification du jeu de données

3 | Statistiques descriptives

4 | Modifications apportées au dataset

5 | Approches non supervisées

6 | Analyse et conclusion

1. Choix du jeu de données



Cancer and Deaths Dataset : 1990~2019
Globally

All Countries' Cancer and Deaths Dataset of last 29 years from 1990 to 2019.

 [kaggle.com](https://www.kaggle.com)

Nombre annuel de décès par cause pour chaque pays dans le monde

	Entity	Code	Year	Number of executions (Amnesty International)	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)
2071	France	FRA	1990	NA	466
2072	France	FRA	1991	NA	453
2073	France	FRA	1992	NA	393
2074	France	FRA	1993	NA	379
2075	France	FRA	1994	NA	364
2076	France	FRA	1995	NA	358

2. Qualification du jeu de données

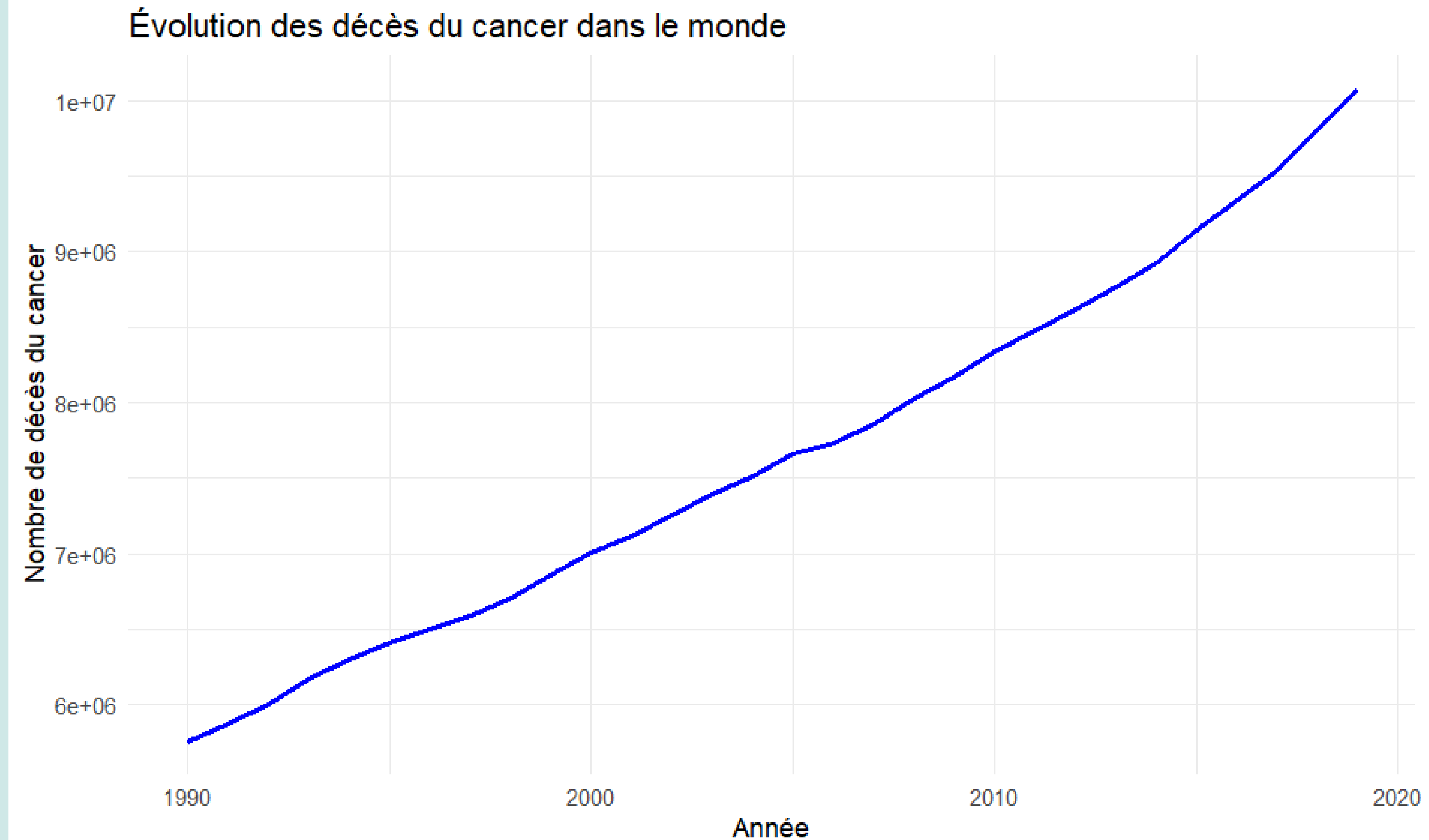
- dataset format cvs
- dimension : 6893 x 36
- peu de valeurs manquantes (et confondues avec valeurs 0 qui sont non notées donc NA)

	Entity	Code	Year	Number of executions (Amnesty International)	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)
2071	France	FRA	1990	NA	466
2072	France	FRA	1991	NA	453
2073	France	FRA	1992	NA	393
2074	France	FRA	1993	NA	379
2075	France	FRA	1994	NA	364
2076	France	FRA	1995	NA	358

Types de données :

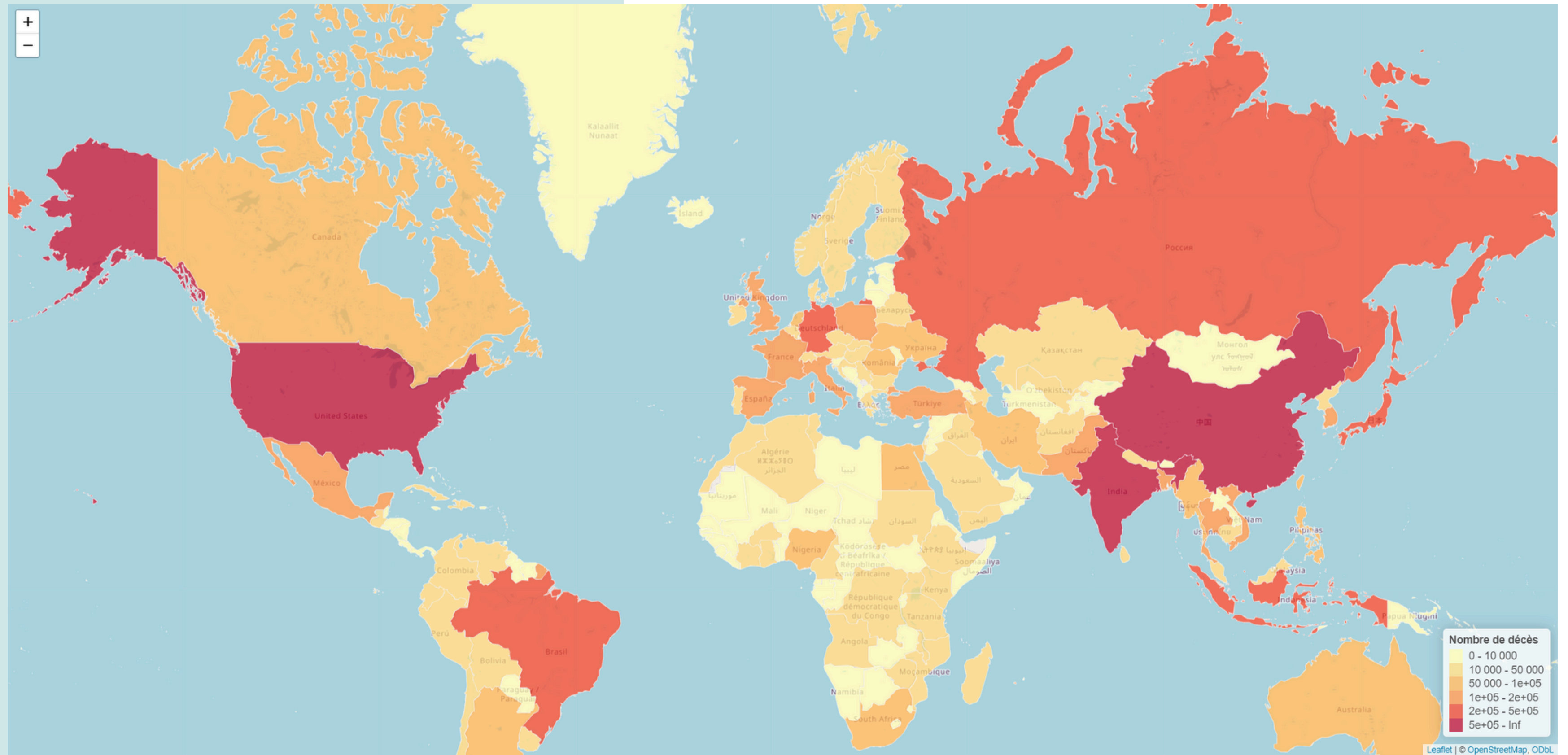
- **Entity** (tous les pays mais d'autres entités comme le monde, l'europe, le G20,...)
- **Code** : codes correspondants aux entités, pour les pays : ISO 3166-1 alpha-3 : codes sur trois lettres, permettant une association visuelle avec le nom usuel du pays
- **Year** : années de 1990 à 2019
- 33 colonnes sur les types de décès (dont le cancer)

3. Statistiques descriptives



3. Statistiques descriptives

Carte du nombre de décès du cancer par pays en 2019



4. Modifications apportées au dataset

Problèmes :

- Le dataset ne contient pas seulement des pays mais d'autres entités comme le monde, l'Europe, le G20,...
- Difficile de comparer des pays par leur nombre de décès du cancer car cela dépend forcément du nombre d'habitants

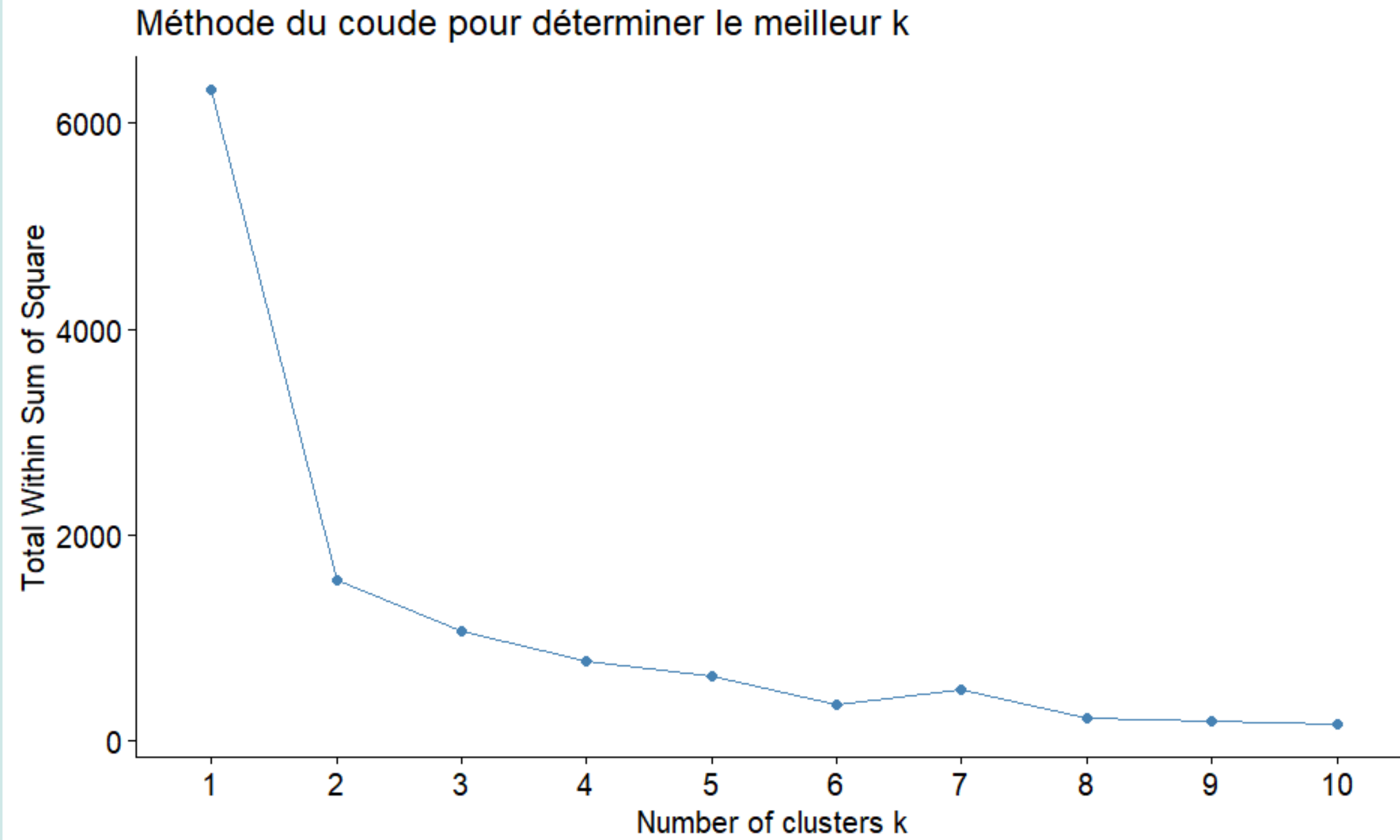
Solutions :

- Filtrer les entités en ne gardant que celles qui ont pour Code un ISO 3 (et qui sont donc des pays)
- A l'aide d'un dataset contenant le nombre d'habitants par pays, création d'un nouveau dataset contenant le **taux de mortalité** du cancer par pays avec le simple calcul :

Taux de mortalité = nb de morts/population totale x 100

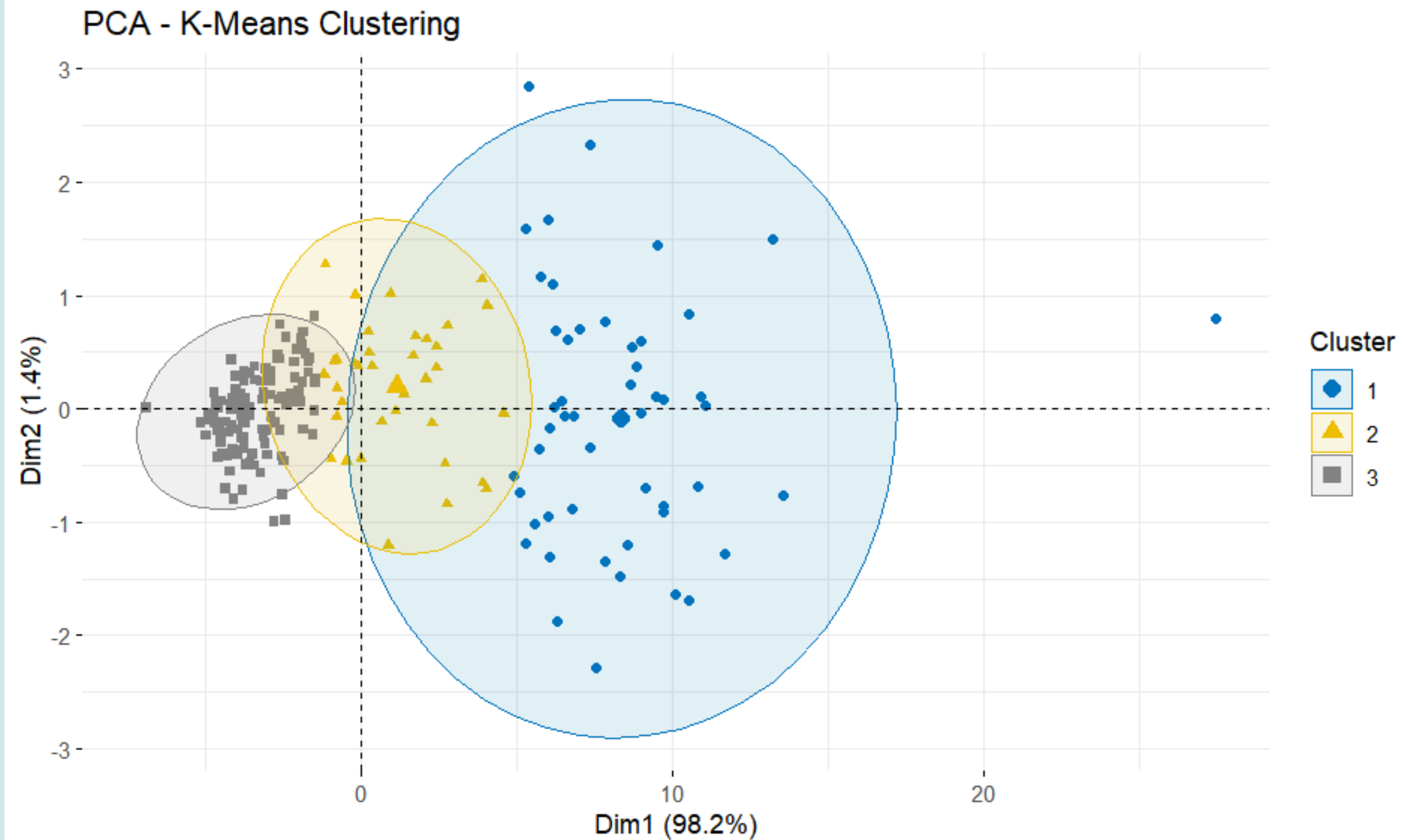
5. Approches non supervisées :

Utilisation de la méthode de clustering des k-means



5. Approches non supervisées :

Utilisation de la méthode de clustering des k-means



5. Approches non supervisées :

Utilisation de la méthode de clustering des k-means

Les pays ont été divisés dans 3 clusters et on remarque qu’ils correspondent plutôt bien aux niveaux de développement des pays :

Cluster 1 ~ pays développés

Cluster 2 ~ pays en développement

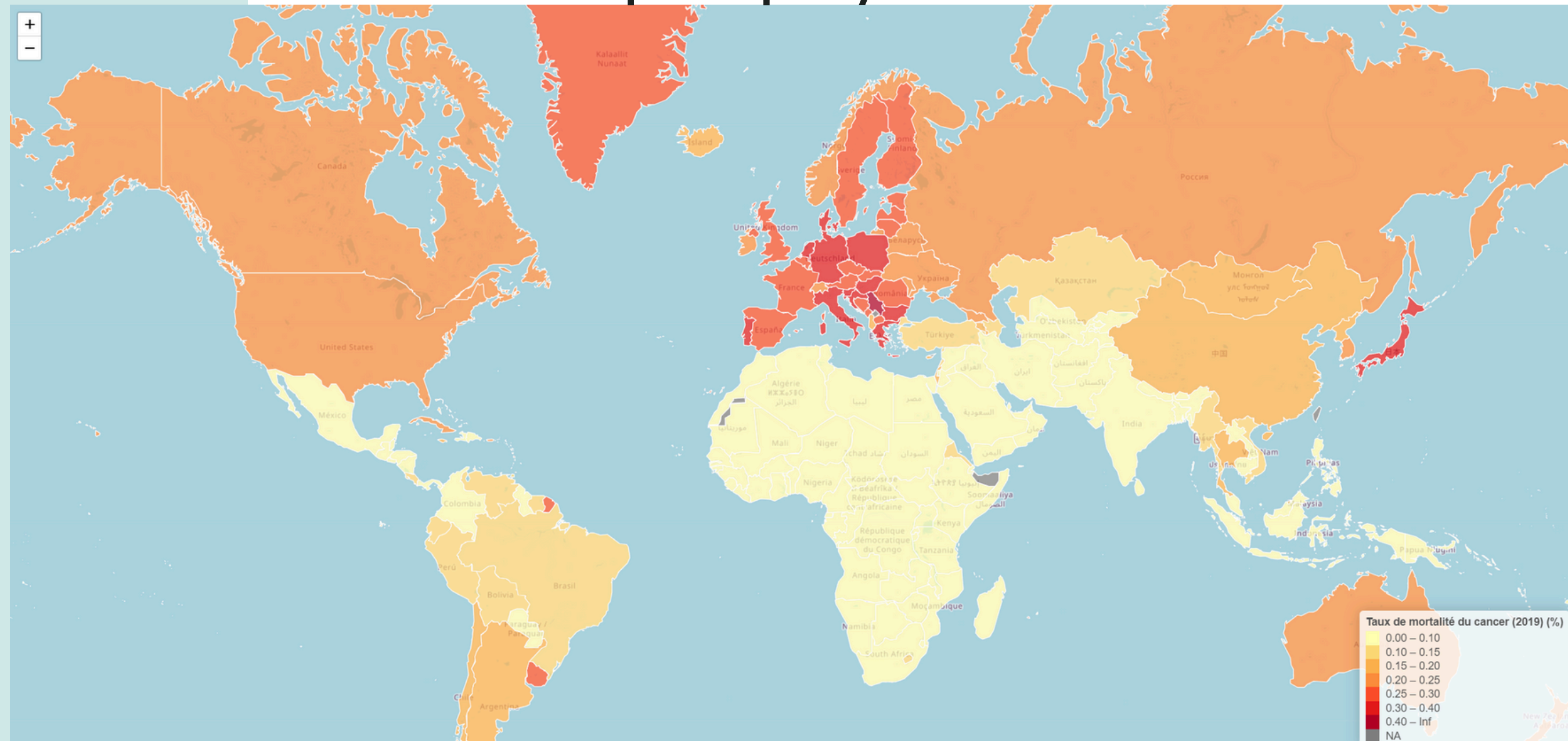
Cluster 3 ~ pays moins avancés

Pays dans le Cluster 3 :			
[1]	"Afghanistan"	"Algeria"	"American Samoa"
[5]	"Bahrain"	"Bangladesh"	"Belize"
[9]	"Bhutan"	"Botswana"	"Brunei"
[13]	"Burundi"	"Cambodia"	"Cameroon"
[17]	"Central African Republic"	"Chad"	"Colombia"
[21]	"Congo"	"Cook Islands"	"Costa Rica"
[25]	"Democratic Republic of Congo"	"Djibouti"	"Dominican Republic"
[29]	"Egypt"	"El Salvador"	"Equatorial Guinea"
[33]	"Eswatini"	"Ethiopia"	"Fiji"
[37]	"French Polynesia"	"Gabon"	"Gambia"
[41]	"Guadeloupe"	"Guam"	"Guatemala"
[45]	"Guinea-Bissau"	"Guyana"	"Haiti"
[49]	"Hong Kong"	"India"	"Indonesia"
[53]	"Iraq"	"Jordan"	"Kenya"
[57]	"Kyrgyzstan"	"Laos"	"Lesotho"
[61]	"Libya"	"Madagascar"	"Malawi"
[65]	"Maldives"	"Mali"	"Marshall Islands"
[69]	"Mauritania"	"Mauritius"	"Mexico"
[73]	"Morocco"	"Mozambique"	"Myanmar"
[77]	"Nauru"	"Nepal"	"New Caledonia"
[81]	"Niger"	"Nigeria"	"Niue"
[85]	"Pakistan"	"Palestine"	"Panama"
[89]	"Paraguay"	"Peru"	"Philippines"
[93]	"Rwanda"	"Samoa"	"Sao Tome and Principe"
[97]	"Senegal"	"Sierra Leone"	"Solomon Islands"
[101]	"South Africa"	"South Sudan"	"Sri Lanka"
[105]	"Suriname"	"Syria"	"Taiwan"
[109]	"Tanzania"	"Timor"	"Togo"
[113]	"Tunisia"	"Turkmenistan"	"Uganda"
[117]	"Uzbekistan"	"Vanuatu"	"Venezuela"
[121]	"Wallis and Futuna"	"Western Sahara"	"Yemen"
[125]	"Zimbabwe"		
Pays dans le Cluster 2 :			
[1]	"Albania"	"Antigua and Barbuda"	"Argentina"
[5]	"Azerbaijan"	"Bahamas"	"Bolivia"
[9]	"Chile"	"China"	"Cuba"
[13]	"Georgia"	"Grenada"	"Iceland"
[17]	"Jamaica"	"Kazakhstan"	"Kiribati"
[21]	"Malta"	"Mongolia"	"North Korea"
[25]	"Palau"	"Puerto Rico"	"Saint Kitts and Nevis"
[29]	"Saint Vincent and the Grenadines"	"Seychelles"	"Singapore"
[33]	"Thailand"	"Tonga"	"Trinidad and Tobago"
[37]	"Tuvalu"		
Pays dans le Cluster 1 :			
[1]	"Andorra"	"Australia"	"Austria"
[5]	"Belarus"	"Belgium"	"Bermuda"
[9]	"Bulgaria"	"Canada"	"Croatia"
[13]	"Denmark"	"Dominica"	"Estonia"
[17]	"France"	"Germany"	"Greece"
[21]	"Hungary"	"Ireland"	"Italy"
[25]	"Latvia"	"Lithuania"	"Luxembourg"
[29]	"Monaco"	"Montenegro"	"Netherlands"
[33]	"North Macedonia"	"Norway"	"Poland"
[37]	"Romania"	"Russia"	"San Marino"
[41]	"Slovakia"	"Slovenia"	"Spain"
[45]	"Switzerland"	"Ukraine"	"United Kingdom"
[49]	"United States Virgin Islands"	"Uruguay"	
			"Angola"
			"Benin"
			"Burkina Faso"
			"Cape Verde"
			"Comoros"
			"Cote d'Ivoire"
			"Ecuador"
			"Eritrea"
			"French Guiana"
			"Ghana"
			"Guinea"
			"Honduras"
			"Iran"
			"Kuwait"
			"Liberia"
			"Malaysia"
			"Martinique"
			"Micronesia (country)"
			"Namibia"
			"Nicaragua"
			"Oman"
			"Papua New Guinea"
			"Qatar"
			"Saudi Arabia"
			"Somalia"
			"Sudan"
			"Tajikistan"
			"Tokelau"
			"United Arab Emirates"
			"Vietnam"
			"Zambia"
			"Armenia"
			"Brazil"
			"Cyprus"
			"Israel"
			"Lebanon"
			"Northern Mariana Islands"
			"Saint Lucia"
			"South Korea"
			"Turkey"
			"Barbados"
			"Bosnia and Herzegovina"
			"Czechia"
			"Finland"
			"Greenland"
			"Japan"
			"Moldova"
			"New Zealand"
			"Portugal"
			"Serbia"
			"Sweden"
			"United States"

6. Analyse et Conclusion :

Cependant on remarque que les pays développés ont un taux de mortalités plus importants que les pays en développement et les pays moins avancés. Pourquoi ?

Carte du taux de mortalité du cancer par pays en 2019



6. Analyse et Conclusion :

Différents facteurs :

- Facteurs de risques liés au mode de vie (alimentation transformée, sédentarité et obésité, tabac et alcool, exposition à certains polluants,...)
- Espérance de vie plus longue (cancer : maladie liée à l'âge)
- D'autres causes de mortalité prédominent dans les pays moins avancés et en développement
- Un meilleur dépistage et suivi médical (on enregistre mieux les décès dus au cancer)

Elément media : GitHub



Mordred6 / datasante_cancer_dataset

Q Type / to search

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

datasante_cancer_dataset

Public

Pin

Unwatch 1

Fork 0

Star 0

main

1 Branch

0 Tags

Go to file

Add file

<> Code

About

Mordred6

Add files via upload

8f064c5 · 1 minute ago

3 Commits

01 annual-number-of-deaths-by-cause.csv	Add files via upload	2 minutes ago
BARDOT_new_dataset_cancer.R	Add files via upload	1 minute ago
README.md	Initial commit	3 minutes ago
carte_cancer_01.R	Add files via upload	1 minute ago
lecture_cancer_01.R	Add files via upload	1 minute ago

README

datasante_cancer_dataset

Projet Datasanté - Exploration d'un dataset sur le cancer

Projet Datasanté - Exploration d'un dataset sur le cancer

Readme

Activity

0 stars

1 watching

0 forks

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

© 2025 GitHub, Inc.

Terms

Privacy

Security

Status

Docs

Contact

Manage cookies

Do not share my personal information

Merci pour votre attention !

PRÉSENTÉ PAR ARTHUR BARDOT

Annexe

Model	Silhouette
:-----:	-----:
K-Means	0.174
Hierarchical	0.150
GMM	0.128
DBSCAN	0.036