Télécom Paris

Mastère Spécialisé Big Data : gestion et analyse des données massives

Projet Visualisation

IGR204 : Visualisation

Auteurs :
Morgan Fassier
Théo Lefièvre
Ernest Majdalani
Vincent Poquet
Adrien Senet

2020 - 2021

1 Présentation des données

1.1 Contexte

Les données utilisées dans le cadre de ce projet sont issues de Gapminder, une entreprise à but non lucratif, dont le but est la promotion du développement durable et des objectifs du millénaire pour le développement de l'ONU.

Nous avons choisi le jeu de données Income Inequalities de Gapminder : nous souhaitons créer des visualisations sur l'évolution des tendances sociales, économiques et environnementales liées aux inégalités de revenus dans le monde.

1.2 Dimensions et propriétés des données

Pour chaque ensemble de données, nous disposons de statistiques annuelles pour une majorité des pays dans le monde.

Nous avons quatre tableaux bivariés représentants quatre différents facteurs. Pour chacun de ces tableaux, une ligne correspond à un pays (nominal) et une colonne à une année (temporel, ordinal). Dans les cellules nous avons des variables numériques.

Les quatre facteurs sont :

- L'inégalité de revenus caractérisée par le coefficient de Gini entre 1800 et 2040 pour 195 pays
- Le niveau d'éducation caractérisé par le nombre d'années d'étude moyen entre 1870 et 2017 pour 188 pays
- L'espérance de vie entre 1870 et 2100 pour 188 pays
- La consommation énergétique par personne caractérisée par le nombre de kilogrammes d'équivalent pétrole consommé par une personne moyenne en une année entre 1960 et 2015 pour 170 pays.

1.3 Transformation des données

Afin de pouvoir exploiter les données dans notre application nous avons nettoyé et prétraité les données. En effet, comme décrit dans la partie précédente les données récupérées étaient sous forme de tableaux croisés (voir figure 1).

	country	1800	1801	1802	1803	1804	1805	1806	1807	1808	...	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040
0	Afghanistan	30.5	30.5	30.5	30.5	30.5	30.5	30.5	30.5	30.5	...	36.8	36.8	36.8	36.8	36.8	36.8	36.8	36.8	36.8	36.8
1	Albania	38.9	38.9	38.9	38.9	38.9	38.9	38.9	38.9	38.9	...	29.0	29.0	29.0	29.0	29.0	29.0	29.0	29.0	29.0	29.0
2	Algeria	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	...	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6	27.6
3	Andorra	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	...	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0
4	Angola	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	57.2	...	42.6	42.6	42.6	42.6	42.6	42.6	42.6	42.6	42.6	42.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
190	Venezuela	62.8	62.8	62.8	62.8	62.8	62.8	62.8	62.8	62.8	...	46.9	46.9	46.9	46.9	46.9	46.9	46.9	46.9	46.9	46.9
191	Vietnam	34.2	34.2	34.2	34.2	34.2	34.2	34.2	34.2	34.2	...	35.3	35.3	35.3	35.3	35.3	35.3	35.3	35.3	35.3	35.3
192	Yemen	50.1	50.1	50.1	50.1	50.1	50.1	50.1	50.1	50.1	...	36.7	36.7	36.7	36.7	36.7	36.7	36.7	36.7	36.7	36.7
193	Zambia	54.5	54.5	54.5	54.5	54.5	54.5	54.5	54.5	54.5	...	57.1	57.1	57.1	57.1	57.1	57.1	57.1	57.1	57.1	57.1
194	Zimbabwe	27.5	27.5	27.5	27.5	27.5	27.5	27.5	27.5	27.5	...	43.2	43.2	43.2	43.2	43.2	43.2	43.2	43.2	43.2	43.2

195 rows × 242 columns

FIGURE 1 – Extrait du jeu de données sur les inégalités de revenus

Pour notre projet, il est essentiel de garder de la continuité temporelle dans les données. Pour cela, nous avons retiré les pays dont les valeurs manquantes étaient nombreuses et les années dont la majorité des pays n'avaient de données correspondantes. Ceci dans le but d'avoir, pour les quatres jeux de données, les mêmes pays et les mêmes intervalles de temps. Finalement, cela nous a amené à garder les données de 101 pays entre 1971 et 2014 inclus.

Ensuite, afin de pouvoir combiner les quatres différentes sources de données nous avons mis les tableaux sous forme de tableaux plats comme sur la figure 2.

	country	Year	value	theme
0	Albania	1971	0.221	education
1	Albania	1972	0.221	education
2	Albania	1973	0.221	education
3	Albania	1974	0.221	education
4	Albania	1975	0.265	education
...	...	...	...	...
4439	Zimbabwe	2010	0.487	education
4440	Zimbabwe	2011	0.487	education
4441	Zimbabwe	2012	0.527	education
4442	Zimbabwe	2013	0.533	education
4443	Zimbabwe	2014	0.547	education

4444 rows × 4 columns

FIGURE 2 – Jeu de données sur l'éducation après prétraitement

Enfin, on a combiné les quatres tableaux pour avoir notre jeu de données final que l'on peut voir sur la figure 3.

	country	year	health	economy	energy	education
0	Albania	1971	67.5	26.8	785.0	0.221
1	Albania	1972	68.1	26.8	866.0	0.221
2	Albania	1973	68.7	26.8	763.0	0.221
3	Albania	1974	69.3	26.8	777.0	0.221
4	Albania	1975	69.8	26.8	827.0	0.265
...	...	...	...	...	...	...
4439	Zimbabwe	2010	49.7	43.2	745.0	0.487
4440	Zimbabwe	2011	52.4	43.2	795.0	0.487
4441	Zimbabwe	2012	54.9	43.2	824.0	0.527
4442	Zimbabwe	2013	56.8	43.2	845.0	0.533
4443	Zimbabwe	2014	58.5	43.2	845.0	0.547

4444 rows × 6 columns

FIGURE 3 – Jeu de données final

1.4 Augmentation du jeu de données

Dans le but de pouvoir construire une carte choroplèthe, on a ajouté à notre jeu de données les données géographique des pays provenant d'un fichier geojson externe [1] (voir figure 17) permettant d'avoir les polygones de chaque pays avec un précision de 110 mètres (faible résolution, ce qui a l'avantage d'être léger, 600 kB, et donc d'être fluide dans les animations). La carte et les animations seront détaillées dans la partie sur le choix du design.

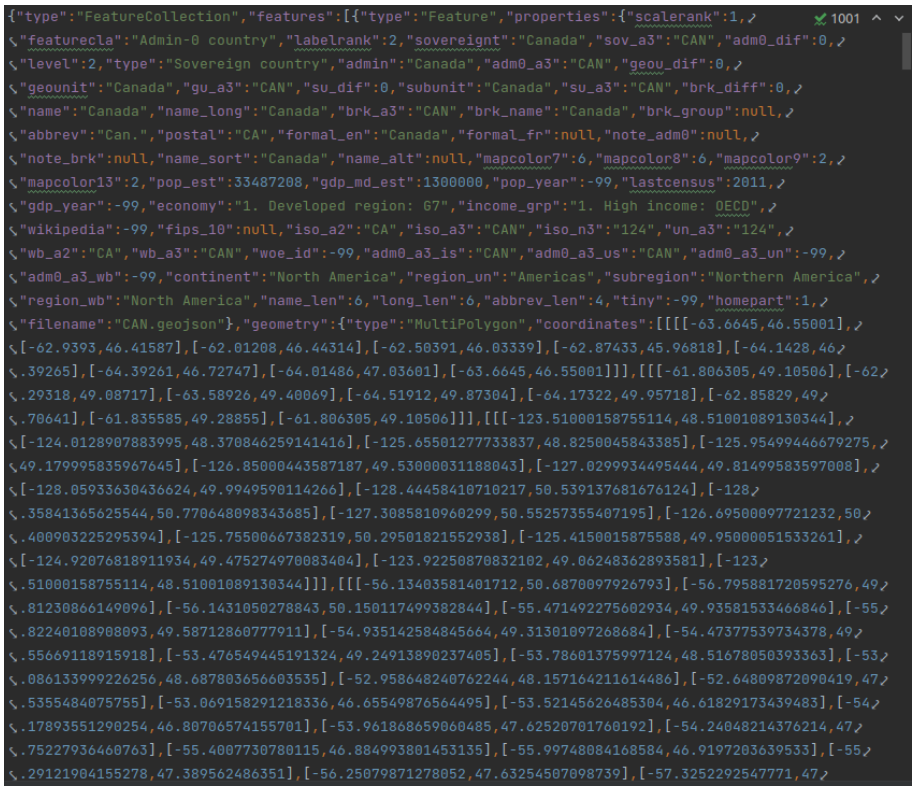

FIGURE 4 – Extrait du jeux de données du geojson

2 Utilisateurs cibles

Les utilisateurs cibles de l'application développée peuvent être divers. Les données utilisées étant relatives aux objectifs du millénaire pour le développement, les visualisations proposées s'adressent à l'ensemble des acteurs du développement visant à identifier et réduire les inégalités, notamment économiques.

Tout d'abord, les principaux utilisateurs seront les organisations non gouvernementales. Ces organisations pourront, au travers de l'application, bénéficier d'outils visuels afin d'identifier et d'agir sur les pays ou secteurs les plus inégalitaires. Pour permettre un développement plus juste et plus équitable, il est nécessaire de déployer une large palette d'actions touchant des secteurs tels que l'éducation ou encore l'accès au soin et à l'énergie. Les visualisations proposées au sein de l'application permettent d'obtenir rapidement un état des lieux de ces thématiques au sein d'un ou plusieurs pays.

La possibilité de visualiser plusieurs pays en même temps peut, par exemple, permettre d'identifier des similarités ou différences au sein de larges régions et rendre ainsi plus efficientes les missions engagées. Cet outil se veut également être un outil de communication pour ses utilisateurs. En effet, les ONG pourraient, par exemple, se servir de l'application comme d'un moyen d'alerter les pouvoirs publics de certains pays sur la nécessité d'améliorer l'accès à l'énergie des tranches les plus pauvres de la population, et ce, afin de réduire les inégalités de revenus.

Ensuite, les organisations gouvernementales sont également des utilisateurs cibles. La comparaison entre différents pays pourrait leur permettre de situer le pays dans le monde sur les thématiques abordées. La visualisation de l'évolution des données sur plusieurs décennies peut également donner des indications sur l'effet des politiques publiques mises en place par le passé. Enfin, dans une moindre mesure, cette application peut être utile à des particuliers souhaitant s'engager sur le plan sociétal en leur permettant d'identifier les secteurs défaillants au sein de leur pays ou à l'étranger.

Qu'il soit un particulier, un bénévole ou membre d'une organisation gouvernementale, l'utilisateur sera en mesure d'interagir de manière intuitive avec les données. L'application développée se veut être la plus accessible possible à des personnes non initiées aux outils de visualisation. Nous avons pour cela choisi de restituer les données sous forme de graphiques permettant d'appréhender rapidement des ordres de grandeur et de comparer facilement plusieurs entités (graphique en radar, diagramme en barres). Ces choix ont été motivés par notre volonté de créer un outil qui puisse convoyer efficacement l'information dans le but d'obtenir une vision globale de la situation d'un territoire ou encore convaincre un auditoire.

Aucune connaissance préalable des données n'est nécessaire. L'utilisateur peut rapidement identifier l'espace des données disponible au travers d'une carte, de boutons radios ainsi que d'une barre de défilement. Il ne lui est jamais demandé d'apporter de l'information au travers d'un quelconque champ de saisie.

3 Designs choisis

L'utilisateur est amené à un dashboard constitué de plusieurs composantes dont les visualisations et interactions sont élaborées ci-après.

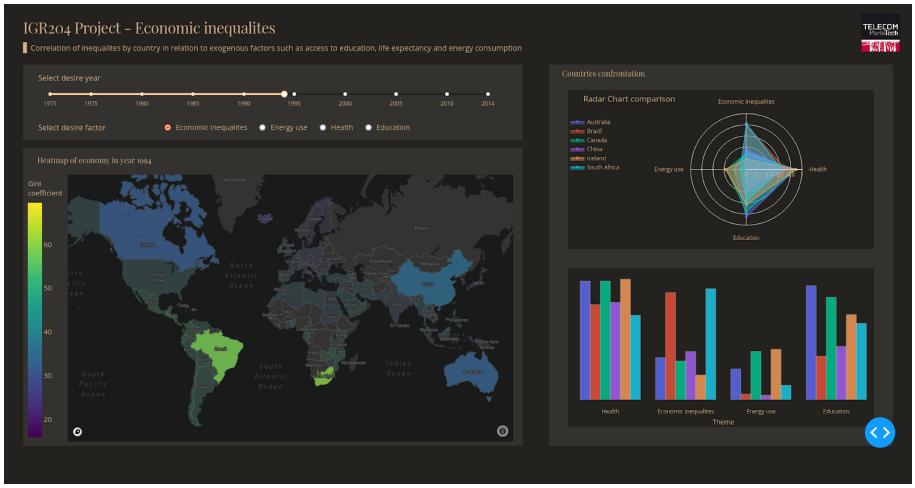

FIGURE 5 – Aperçu de l'application

3.1 Les représentations choisies

3.1.1 Slider

Le slider permet de choisir facilement le filtre temporel souhaité, tant en ayant connaissance des valeurs extrêmes.

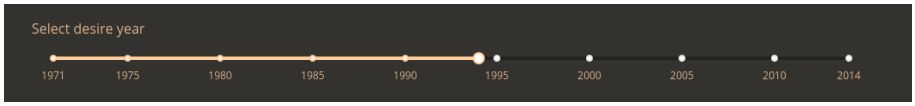

FIGURE 6 – Slider

3.1.2 Boutons radio

Les boutons radio permettent de permuter d'un indicateur à l'autre rapidement.

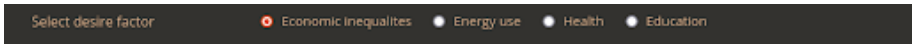

FIGURE 7 – Boutons radio

3.1.3 Carte choroplèthe

La carte choroplèthe permet d'avoir une vue globale de la disparité de la métrique choisie entre les pays en s'appuyant sur la légende des couleurs.

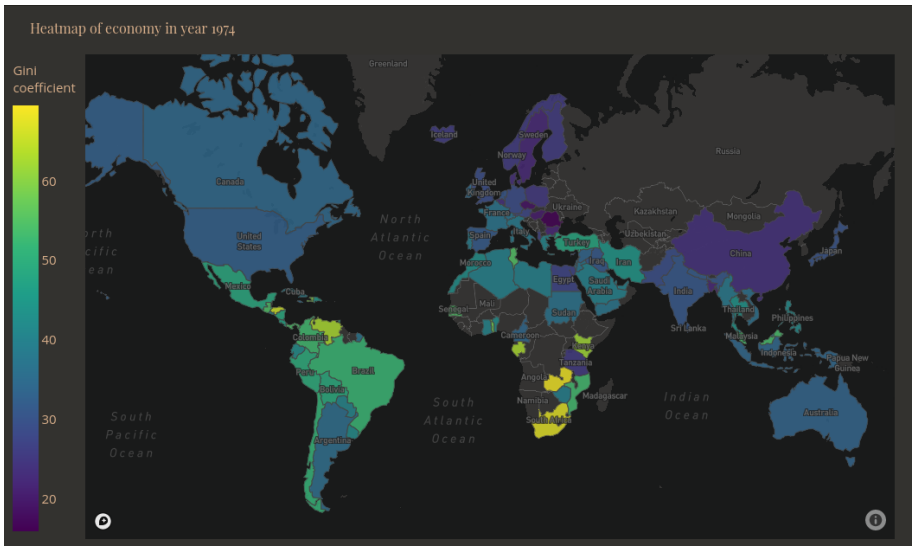

FIGURE 8 – Carte choroplèthe

3.1.4 Diagramme Radar

Ce graphique permet d'obtenir un aperçu rapide des valeurs catégorielles de façon dynamique.

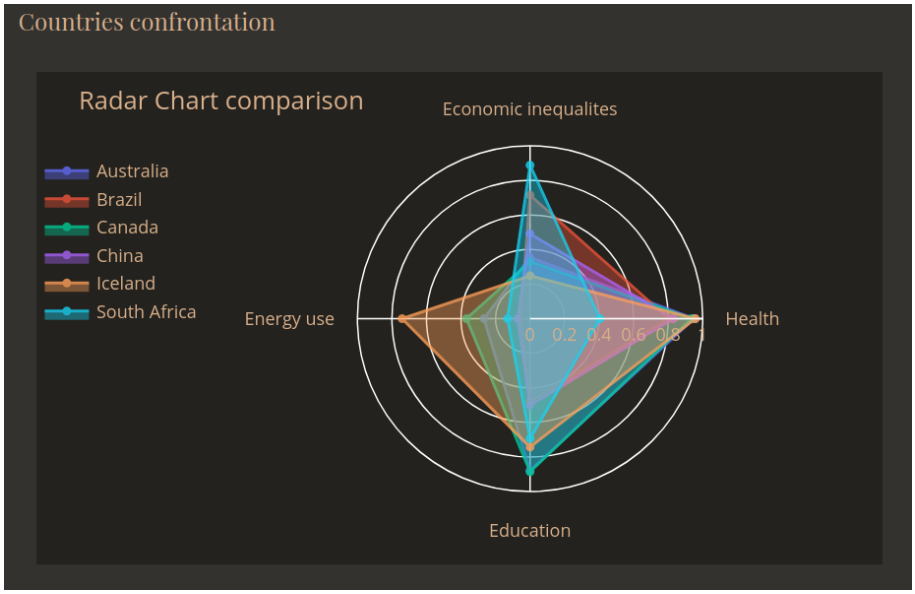

FIGURE 9 – Graphique en radar

3.1.5 Diagramme en barre

Ce graphique permet de voir la répartition de chaque métrique entre les pays sélectionnés.

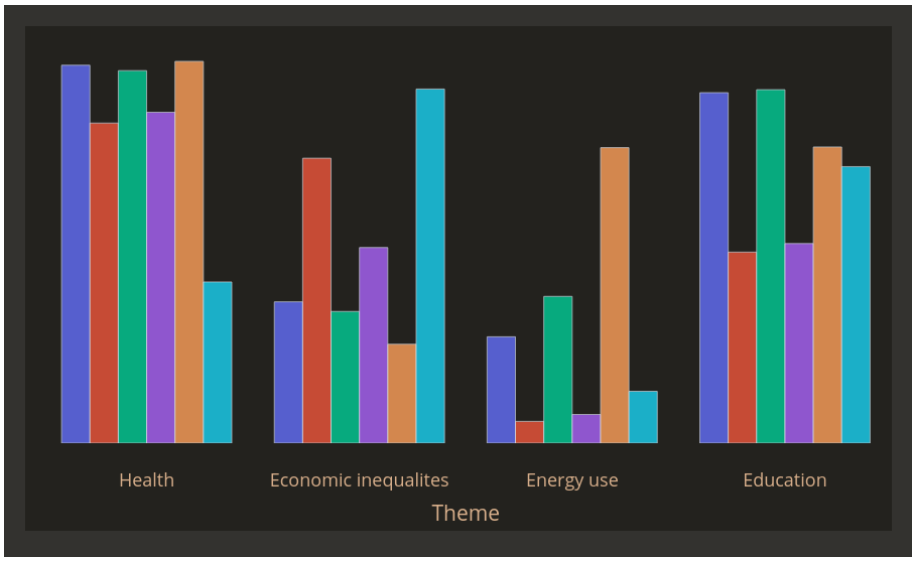

FIGURE 10 – Diagramme en barre

3.2 Interactions

3.2.1 Slider et boutons radio

Lors du choix de l'année sur le slider, ou de l'indicateur sur les boutons radio, la mise à jour des valeurs sur la carte est instantanée.

3.2.2 Carte choroplèthe

Lorsque l'on survole les pays coloriés sur la carte, une étiquette permet d'avoir les valeurs des différentes dimensions au croisement de l'indicateur choisi.

FIGURE 11 – Interactions carte

Il est également possible d'effectuer une sélection des pays :

- en cliquant dessus (et en maintenant la touche shift pour une multi-sélection)
- en utilisant une box ou un lasso pour sélectionner plusieurs pays côte à côte.

Cela a pour impact de baisser l'opacité des autres pays pour mieux voir ceux sélectionnés.

D'autre part, les pays sélectionnées déclenchent l'affichage d'informations plus détaillées sur les autres graphes, décrits ci-après.

3.2.3 Diagramme radar et histogramme

Lorsqu'un ou plusieurs pays sont sélectionnés sur la carte, le radar et l'histogramme s'affichent pour montrer des informations détaillées. Un survol des aires ou des points du radar affichent les données de ceux-ci :

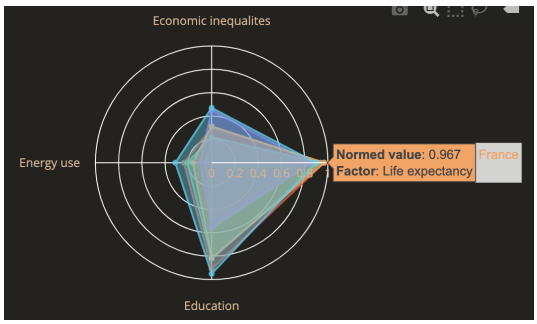

FIGURE 12 – Interactions graphique en radar

De même pour l'histogramme :

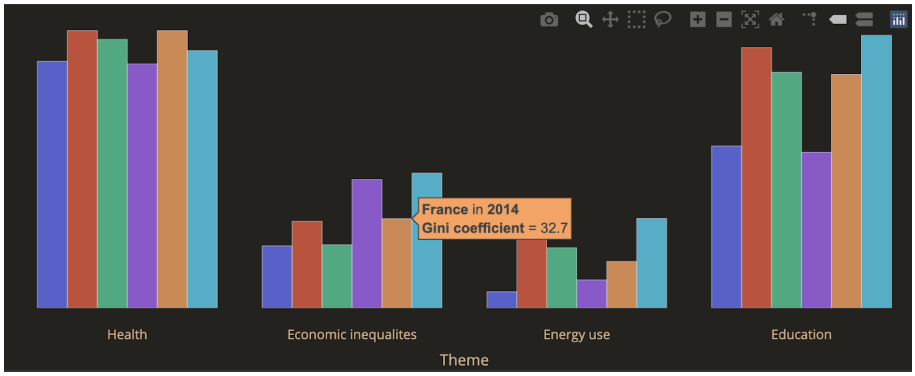

FIGURE 13 – Interactions diagrammes en barre

Enfin, si, en visualisant les données des deux graphiques, l'utilisateur souhaite de nouveau filtrer sur des pays en particulier, il peut le faire sur la légende, sans avoir besoin de retourner sur la carte.

FIGURE 14 – Filtrage légende

3.3 Décisions de conception associées aux utilisateurs, données et tâches

L'information commune la plus importante de nos quatre sources de données étant le pays, il nous a semblé naturel de choisir une carte comme figure principale. Celle-ci a l'avantage de montrer d'un rapide coup d'œil des phénomènes marquants. Par exemple, pour l'éducation, on dénote une cassure entre les deux hémisphères.

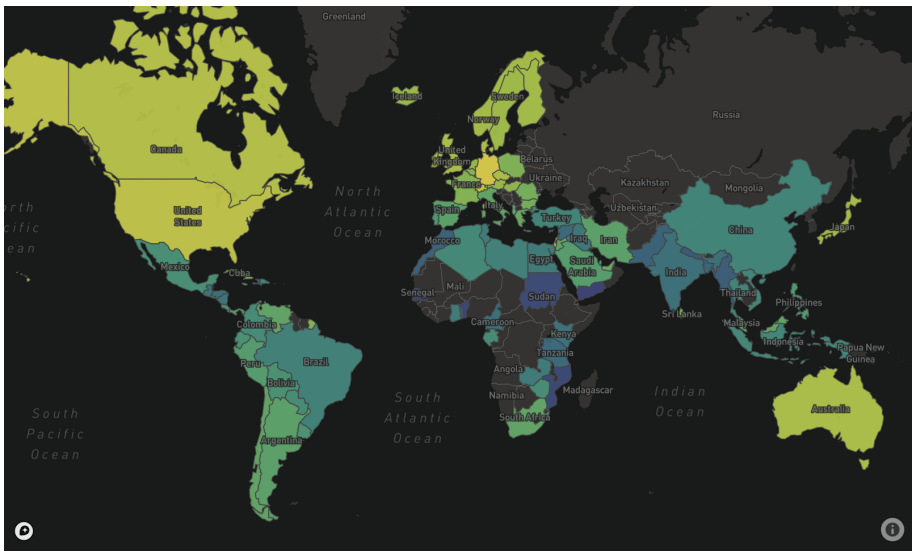

FIGURE 15 – Éducation

Ou encore pour l'utilisation de l'énergie, où certains pays, comme l'Islande et le Qatar, sortent du lot.

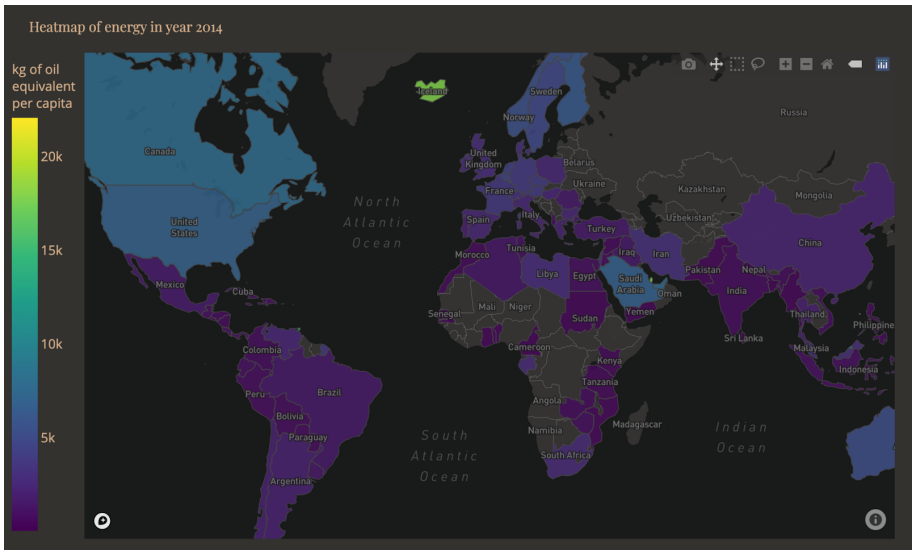

FIGURE 16 – Énergie

La seconde information commune à toutes nos sources était le temps, même si leurs valeurs extrêmes différaient. Ainsi, en faisant varier l'année de visualisation sur le slider, l'évolution des indicateurs dans le temps est très simple à voir.

FIGURE 17 – Évolution indicateur

Cette représentation permet donc à l'utilisateur de voir des phénomènes marquants, dont il voudrait connaître l'origine et analyser plus en détail. Ceci va justement être possible grâce aux deux autres graphiques : le radar et l'histogramme.

En effet, une fois les pays concernés par le phénomène sélectionnés, les informations détaillées de ceux-ci apparaissent et permettent une analyse approfondie entre pays, et au niveau des différentes métriques. L'utilisateur pourrait donc trouver des corrélations permettant d'alerter ou de prendre des décisions.

3.4 Points forts de la conception

Tout d'abord, nous avons choisi de développer notre application en tant que page web, pour l'ergonomie, et lui permettre de s'adapter à différentes tailles ou résolutions d'écran.

Ensuite, nous avons choisi un template d'affichage plutôt sombre. Si un fond blanc et une écriture noire rend la lecture d'un texte plus aisée, un fond noir et une écriture claire repose les yeux. Dans notre cas, les textes n'apparaissent qu'à la demande de l'utilisateur (hovering) et nous avons choisi de les mettre dans une taille de police suffisamment grande pour faciliter leur lecture. Ainsi, nous essayons de prévenir la fatigue de l'utilisateur tout en lui permettant d'obtenir aisément les informations souhaitées lorsqu'il en a besoin.

L'utilisation de la carte "cliquable" et le choix des paramètres via les boutons radios ou le slider est instinctive et ne demande aucun "mode d'emploi". de plus, cela nous permet d'ajouter des données (nouveaux pays, nouvelles métriques, nouvelles valeurs de dimension) sans reprendre le développement, et de façon totalement transparente pour l'utilisateur

3.5 Points faibles

L'évolution dans le temps des valeurs des indicateurs sur la carte pourrait être encore plus visuelle avec une animation, comme sur une vidéo. Nous avons pensé, sans avoir eu le temps de la concevoir, à un line chart pour voir l'évolution d'une métrique au fil du temps pour des pays sélectionnés. Un inconvénient de la carte est que les petits pays sont évidemment plus difficiles à voir, et cela peut empêcher l'utilisateur de voir des phénomènes marquants (exemple avec le Qatar pour la métrique énergie).

4 Dépôt

Url du dépôt github : https://github.com/ernestmajdalani/data-visualization-project

Références

[1] URL : https://dash-gallery.plotly.host/dash-opioid-epidemic/.
[2] URL : https://github.com/plotly/dash-sample-apps/tree/main/apps/dash-opioid-epidemic.