
UNIVERSITÉ PARIS 8 - VINCENNES À SAINT-DENIS

M1 MIASHS : Big Data et fouille de données

Projet : Visualisation de données via Tableau

PANCHALINGAMOORTHY GAJENTHRAN

Organisme d'accueil : Université Paris 8
Cours : Visualisation de masses de données

Table des matières

1	Introduction	1
1.1	Téléchargement et installation de l'outil	1
1.2	Chargement de données	2
1.3	Nettoyage et transformation de données	3
2	Développement	5
2.1	Chiffres clés	5
2.2	Nombre de morts à travers la carte du monde	5
2.3	Lecture de l'impact de la maladie grâce aux diagrammes	7
2.3.1	Barres côte à côte	7
2.3.2	Barres verticales	8
2.3.3	Lignes discrètes	10
2.4	Plan plus précis sur le nombre de morts	11
2.4.1	Diagramme	11
2.4.2	Tableau d'actualité	13
2.5	Dashboard	13
3	Conclusion	19

Chapitre 1

Introduction

Ce projet consiste à pouvoir utiliser l'outil de visualisation de données **Tableau** afin de pouvoir visualiser et comprendre les données provenant de <https://opendata.ecdc.europa.eu/covid19> traitant à propos des données liés au Covid19, plus précisément du nombre de morts et de cas infectés par le Coronavirus à travers le monde.

Depuis plusieurs mois maintenant, le monde est confronté à une sérieuse épidémie du nom de Covid19. Apparue en Chine en 2019, cette maladie a réussi à se propager tout autour du globe terrestre, causant ainsi de nombreux morts dans son passage.

C'est pourquoi à travers un dashboard, nous allons essayer de comprendre l'ampleur de ce virus, et l'impact qu'il a eu à travers le monde.

1.1 Téléchargement et installation de l'outil

Pour pouvoir télécharger l'outil **Tableau**, il faut se rendre sur leur site <https://www.tableau.com/fr-fr> et cliquer sur le bouton "Essayer" et entrer votre mail : le fichier pèse un peu plus de 500Mo. L'installation est tout aussi simple, il suffit de suivre les différentes instructions, une par une, il n'y a pas de piège. Vous pouvez utiliser la version d'évaluation pendant 15 jours.

Nous nous retrouverons ainsi sur la page suivante 1.1 ; sur la gauche, nous pourrions importer les données à partir d'un fichier que soit sous format Excel, texte, JSON, PDF... **Tableau** propose une large gamme de choix par rapport à cela, il est donc difficile de ne pas trouver son compte. Mais il est également possible de récupérer les données sur un serveur (MySQL, Oracle, Amazon Redshift ou encore Google Analytics).

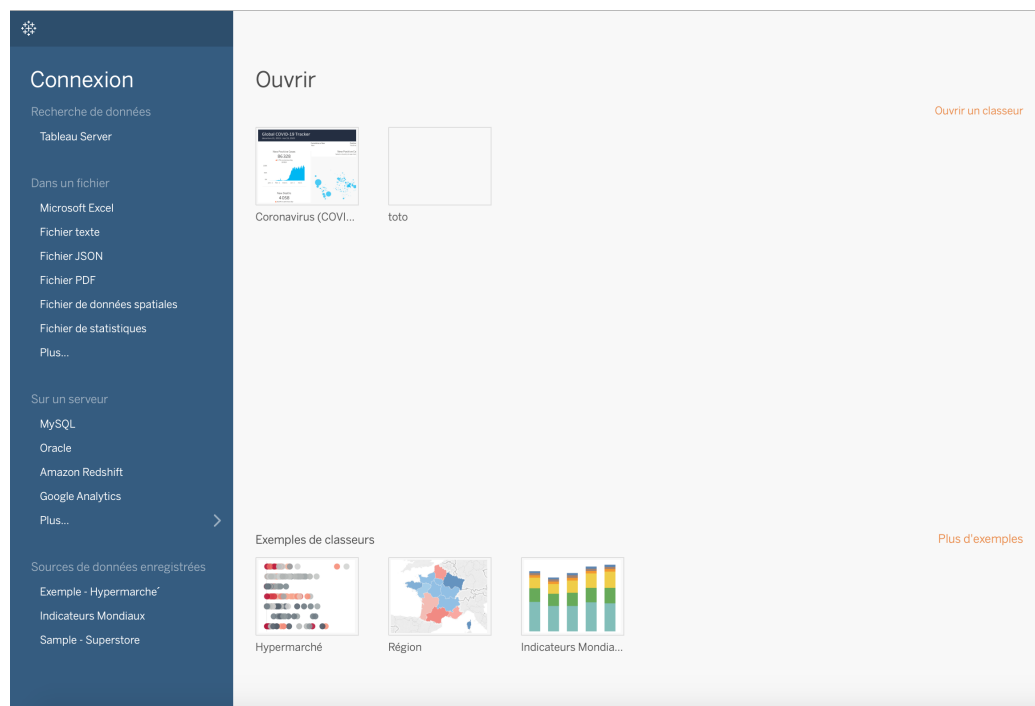


FIGURE 1.1 – Accueil du logiciel Tableau

1.2 Chargement de données

Après avoir trouver notre fichier, qui d'ailleurs est en format .csv, il suffira d'ouvrir ce dernier pour lire la source des données. A noter que les fichiers sous format CSV, sont par défaut délimités par des points virgules dans le logiciel, donc il faudra préciser que le séparateur de nos champs est une virgule 1.2

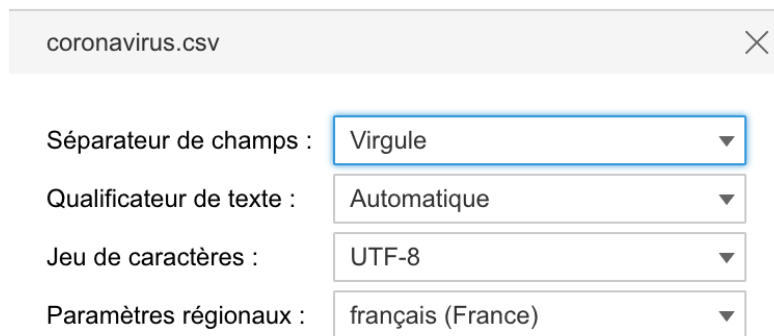


FIGURE 1.2 – Propriété du fichier .csv

1.3 Nettoyage et transformation de données

Avant de pouvoir réaliser notre dashboard, on va tout d'abord penser à nettoyer et transformer les données ; nous allons commencer par vérifier le type des champs 1.3, puis modifier le nom de nos variables. Le seul soucis que pouvons connaître sur Tableau concerne les enregistrements contenant des valeurs nulles qui ne peuvent pas être supprimer. Ainsi, nous allons donc procéder à une autre manipulations pour filtrer les enregistrements 1.5. Par ailleurs, nous supprimerons les colonnes inutiles comme "geoId" 1.4.



FIGURE 1.3 – Changement du type de chaque colonne

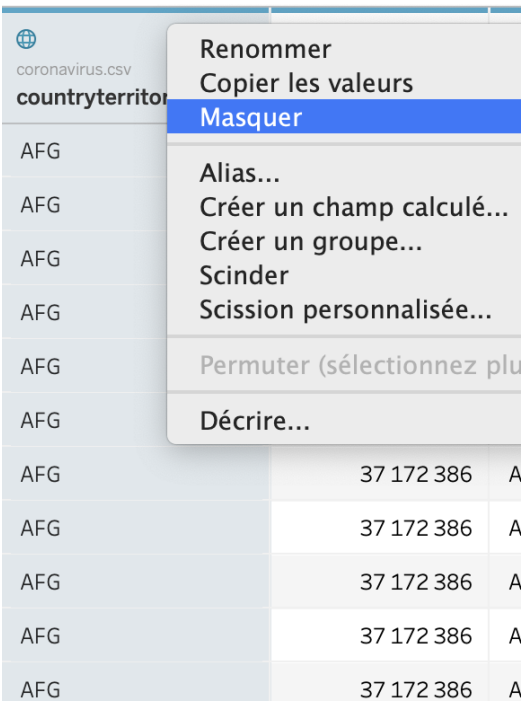


FIGURE 1.4 – Suppression de la colonne "geoId"

 coronavirus.csv date	 coronavirus.csv day	 coronavirus.csv month	 coronavirus.csv year	 coronavirus.csv cases	 coronavirus.csv deaths	 coronavirus.csv countries	 coronavirus.csv geold	 coronavirus.csv population	 coronavirus.csv continent
---	--	--	---	--	---	--	--	--	--

FIGURE 1.5 – Colonnes de notre sources de données

Chapitre 2

Développement

Pour commencer, nous allons essayer de détailler la manière dont nos informations seront structurées :

- Dans un premier temps, on retrouvera le titre de ce Dashboard ainsi que les chiffres clés de ce dernier, pour essayer de captiver le plus possible l'utilisateur
- Puis, nous continuerons notre dashboard, avec un élément qui englobe toutes les informations, en montrant la carte du monde et le nombre de morts selon les pays.
- Ensuite, arrivera la partie avec les diagrammes, dans laquelle, nous retrouverons des informations plus ou moins précises des cas et des morts liés au Covid19, avec notamment des barres verticales, côte à côte ou encore des lignes.
- Enfin, on terminera avec des informations beaucoup plus précises pour essayer de satisfaire le plus de gens possible.

2.1 Chiffres clés

2.2 Nombre de morts à travers la carte du monde

Pour donner un premier aperçu du nombre de morts, on va représenter les dégâts causés par le Covid19 grâce à une carte. La carte représentera le nombre de mort sous forme de bulles ; on remarque que les pays les plus impactés sont les États-Unis, la France, l'Italie, le Royaume-Uni et l'Espagne. De plus, il est également possible d'avoir plus de précision sur le nombre de morts d'un pays, en dirigeant notre curseur sur un pays 2.3.



FIGURE 2.1 – Représentation du nombre de décès sur la carte du monde

De plus, on remarque également en bas à droite de notre carte qu'il existe 2 inconnues sur notre carte 2.2. En cliquant sur ce bouton, on remarque que les valeurs nulles n'ont pas été traitées ; pour gérer le cas des valeurs nulles, vu que nous pouvons pas les traiter dans la phase précédente, il suffit de filter certaines données de la variable "codeCountry", en l'occurrence ici les valeurs nulles, en glissant la variable sur la fenêtre "Filter" 2.2.

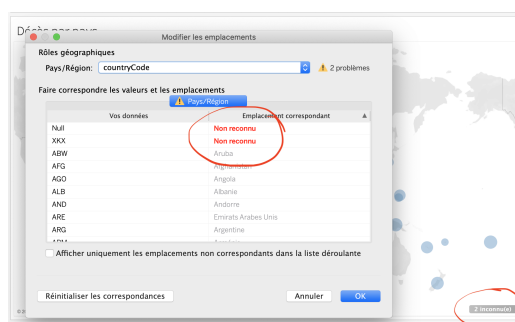


FIGURE 2.2 – Les valeurs non reconnues par le logiciel

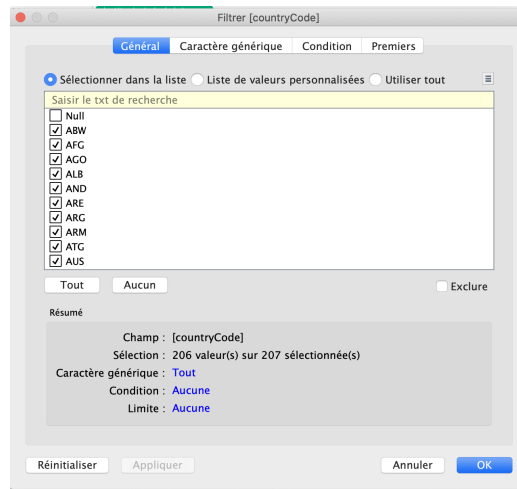


FIGURE 2.3 – Filtrer la variable "codeCountry"

2.3 Lecture de l'impact de la maladie grâce aux diagrammes

2.3.1 Barres côte à côte

Les barres côte à côte vont nous permettre de suivre l'évolution de la maladie depuis le début de l'année. En effet, ces dernières indiqueront le nombre de cas ou le nombre de morts suite au coronavirus. Pour cela, il nous suffit simplement de sélectionner "Barre côte à côte", de mettre dans le champs colonne la date (qui est séparé 3 groupes : quartiles, mois et jour) et le nombre de morts/cas dans le champs ligne 2.4. On peut observer des grands pics lors des mois de Mars/Mai.

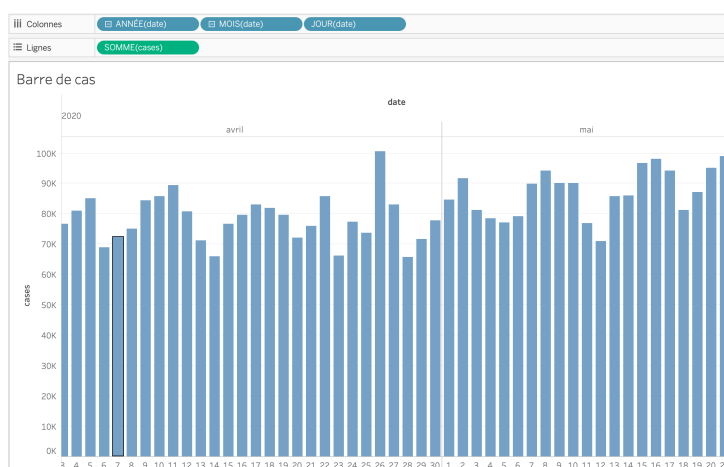


FIGURE 2.4 – Représentation du nombre de décès sur une barre côte à côte

2.3.2 Barres verticales

Les barres verticales vont nous aider à apporter plus de précision en ciblant plutôt des régions. Dans la première barre verticale, on apercevra le nombre de morts(/cas)/continent qui seront ensuite classés dans l'ordre décroissant. On observe notamment que l'Europe et l'Amérique sont les continents les plus touchés par ce virus, suivi de l'Asie qui est le foyer de la maladie.

Pour réaliser les barres verticales, nous avons suivi le même procédé que pour les barres côte à côte. Sauf que nous allons également indexer les valeurs afin de pouvoir les ranger dans l'ordre décroissant. Pour cela, on va d'abord créer un index (Analyse > Modifier le champs calculé 2.5) en utilisant la fonction `INDEX()`, et l'insérer devant la variable que l'on souhaite, en l'occurrence ici "continents". Cela étant fait, on pourra maintenant, ranger dans l'ordre décroissant, en modifiant le calcul de la table (2.6) ; cela ouvrira une fenêtre où l'on pourra choisir l'organisation de nos valeur (dans notre cas : il faudra cliquer sur Dimension spécifique > "continents" > Tri personnalisé > Ordre décroissant).

Pour finir cette partie, afin de mieux visualiser les continents, on rajoutera de la couleur en glissant "continents" dans les couleurs (2.7).



FIGURE 2.5 – Fenêtre pour l’indexation des variables

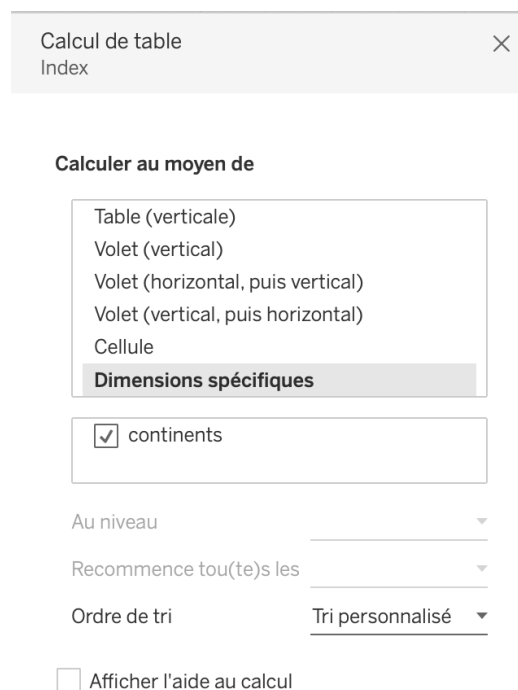


FIGURE 2.6 – Fenêtre pour la manipulation des index

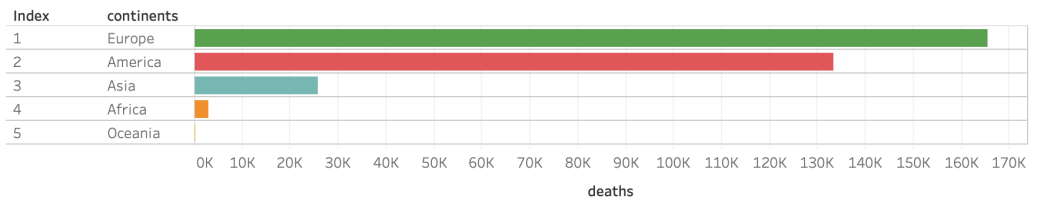


FIGURE 2.7 – Représentation du nombre de décès/continent sur une barre verticale

Nous réaliserons les mêmes étapes pour le nombre de cas positifs au Covid19.

2.3.3 Lignes discrètes

A côté de cette barre verticales, nous rajouterons l'évolution du nombre de mort/cas de chaque continent à travers un graphe. Nous conserverons les mêmes couleurs que pour les barres verticales 2.8, histoire de garder une cohérence. Le graphe suivra le nombre de morts en fonction des jours, de chaque continent. C'est quasiment les mêmes manipulations que pour la barre verticale, sauf qu'il n'y aura pas d'indexation et le champs colonne aura à la place les dates 2.9.

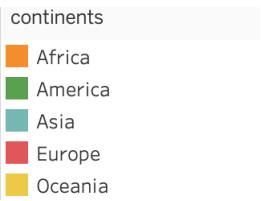


FIGURE 2.8 – Représentation du nombre de décès/continent sur une barre verticale

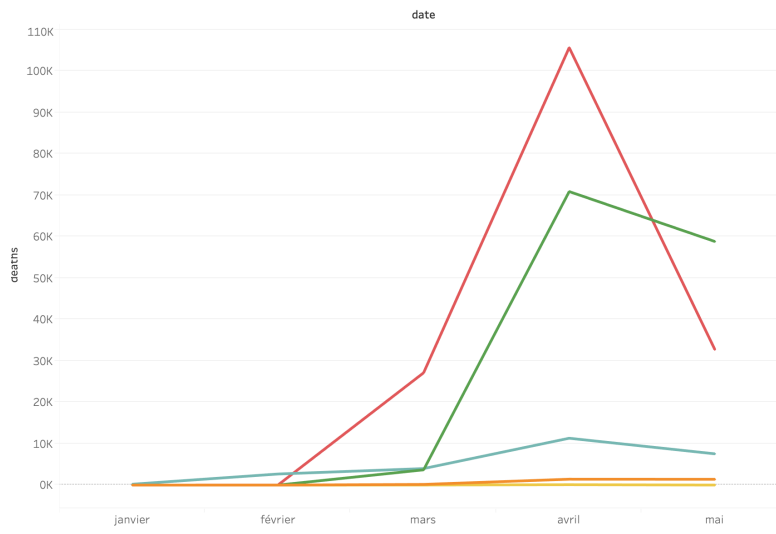


FIGURE 2.9 – Représentation du nombre de décès/continent sur un graphe

2.4 Plan plus précis sur le nombre de morts

2.4.1 Diagramme

On indiquera maintenant, dans la prochaine section, le nombre de morts et de cas dans chaque pays. La section sera coupée en deux parties : à gauche une barre de verticales, et à droite un graphe sur l'évolution du virus dans le pays en question. Afin de ne pas trop surcharger la page, on se contentera seulement du minimum d'informations sur le graphe en retirant notamment les légendes, et les axes ; nous conserverons seulement la première valeur et la dernière valeur (2.11).

Vu qu'il s'agit d'une barre verticale (2.10), on va une fois de plus, créer un index afin de classer les pays par ordre décroissant par rapport au nombre de morts/cas, et surtout permettre de limiter le nombre de pays, en l'occurrence ici, nous afficherons seulement 20 pays (il suffit de modifier les filtres et de cocher seulement le nombre de pays que nous souhaitons afficher).

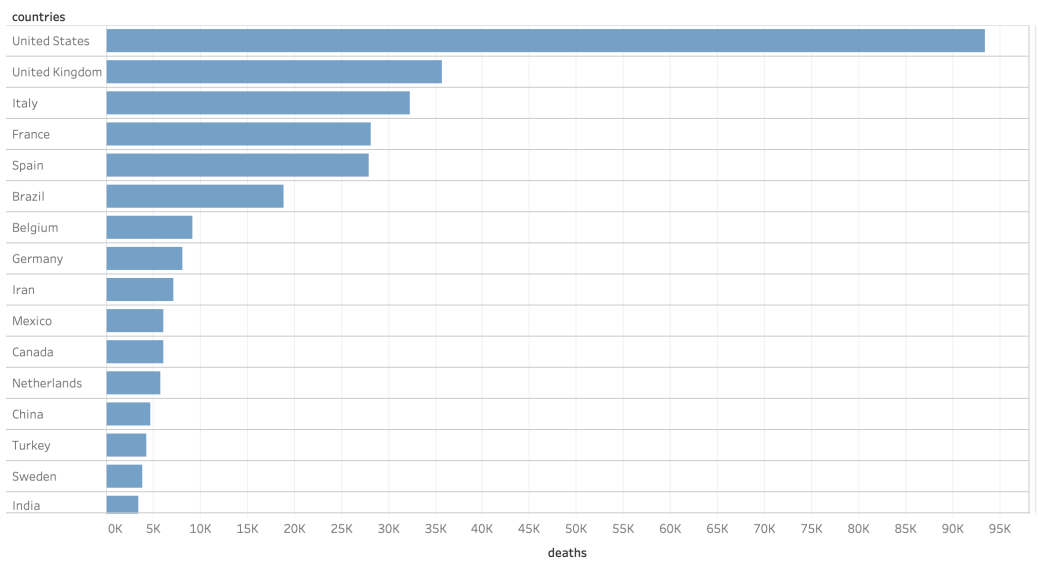


FIGURE 2.10 – Représentation du nombre de décès/continent sur une barre verticale

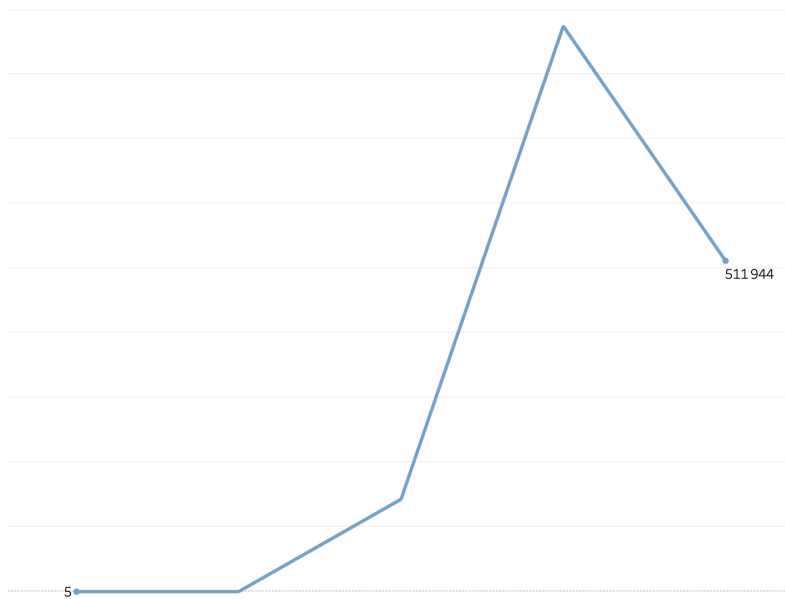


FIGURE 2.11 – Représentation du nombre de décès/continent sur un graphe

2.4.2 Tableau d'actualité

Le tableau d'actualité permet d'avoir un renseignement sur tous les pays que ça soit le nombre de cas, ou le nombre de mort, mais également le taux de mortalité dû au Coronavirus, et la population du pays en question. On mettra un tableau complet de tous les pays, avec la possibilité de scroller. Pour cela, on va devoir "Créer un champs calculé" (2.12) dans lequel nous calculerons le taux de mortalité (*taux de décès / taux de cas*), en précisant qu'il s'agit d'un pourcentage dans les réglages. Comme nous pouvons le voir sur le tableau ci-dessous (2.13), les enregistrements sont indexés et ranger selon le taux de mortalité.



FIGURE 2.12 – Représentation du nombre de décès/continent sur un graphe

countries	cases	deaths	population	mortality rate %
Nicaragua	25	8	413 792 832	32,0%
France	143 845	28 132	9 579 175 892	19,6%
Sint_Maarten	77	15	236 470	19,5%
Belgium	55 983	10 000	11 589 621	16,3%
Yemen	180	10 000	29 825 964	16,1%

FIGURE 2.13 – Représentation du nombre de décès/continent sur un graphe

2.5 Dashboard

Une fois les composants finis, il nous reste plus qu'à les rassembler pour former notre dashboard ; je précise tout de même que notre dashboard est

reparti en 4 pages par manque de place.

Tout d'abord, nous devons fixer les titres pour chaque partie, et ainsi créer de nouveaux composants. Pour le titre principal, je me suis dit qu'il était intéressant de mettre la date du début de la propagation de la maladie jusqu'à Mai, et de faire une progression au fur et à mesure des représentations selon la date donnée. Pour cela, nous allons rajouter la variable "date" dans la fenêtre "Page" et mettre le repère à "Tout" pour pouvoir afficher toutes les dates. Cela va nous permettre de suivre de chronologiquement les dégâts du virus et de mieux visualiser la propagation de la maladie (2.14).

The image shows a user interface for displaying historical data. It consists of several sections:

- Pages:** A section with a blue button labeled "date".
- Repères pour lesquels afficher l'historique:** A section with two tabs: "Sélectionné" and "Surligné". Under "Sélectionné", there are two buttons: "Manuel" and "Tout".
- Longueur:** A section with two radio buttons: "Tout" (selected) and "Dernier(-ière)". Next to "Dernier(-ière)" is a numeric input field with the value "5".
- Afficher:** A section with three buttons: "Repères", "Pistes", and "Les deux".
- Repères:** A section with a "Format:" label and a dropdown menu showing "Aucun(e)". Below it is a checkbox labeled "Estomper:" which is checked, followed by a slider control.
- Pistes:** A section with a "Format:" label and a dropdown menu showing "Aucun(e)".

FIGURE 2.14 – Représentation du nombre de décès/continent sur un graphe

Pour aider l'utilisateur à voir l'évolution d'un pays, j'ai décidé de mettre dans la 3ème page, la possibilité de filtrer les graphes selon le pays que l'utilisateur va cliquer. Le graphe va donner une indication évidemment sur la croissance du nombre de mort et de cas positifs dans le pays en question. Pour que les composants soit reliés entre eux, il faudra rajouter une action de filtre, et lier les composants que l'on souhaite associer comme nous pouvons

le voir dans l'image suivante (2.15).

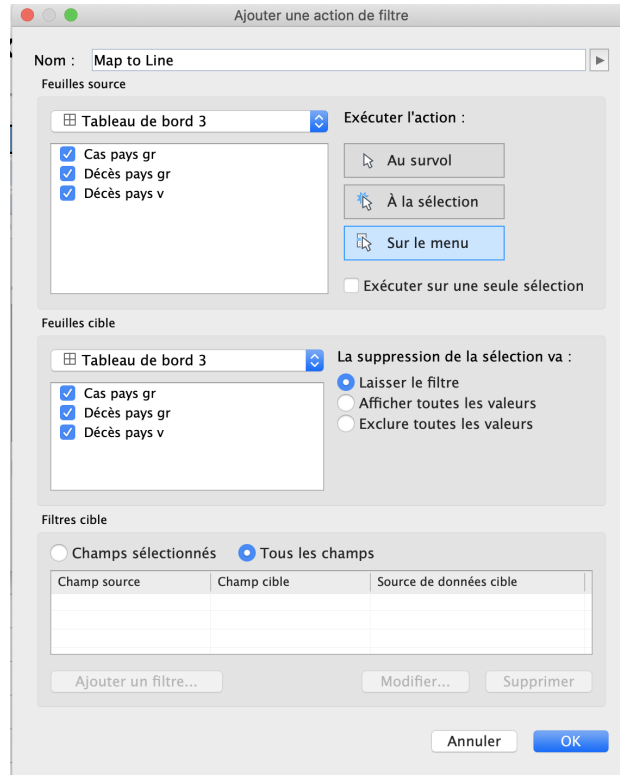


FIGURE 2.15 – Mapping des données selon le pays cliqué

Un autre ajout intéressant serait des boutons permettant de switcher entre plusieurs options, comme la possibilité d’alterner entre une carte représentant le nombre de cas ou le nombre de morts. De même pour les barres horizontales, il serait bien de choisir entre "obtenir des valeurs journalières" ou "des valeurs cumulées". Pour cela, on va créer un nouvel objet dans lequel nous allons rajouter rajouter un de nos composants. L’objet créé se superposera sur l’autre objet qui va être switché (et donc aura les mêmes coordonnées que celui-ci), puis, nous ajouterons un bouton (2.16) pour lier les deux objets à chaque clics.

FIGURE 2.16 – Bouton pour switcher entre plusieurs paramètres

Ainsi, voici un petit recapitulatif des différents pages :

- **Page 1** : Page d'accueil, avec la carte, le nombre de morts et de cas suite au Coronavirus. J'ai décidé de mettre des informations large sur la première page (??).
- **Page 2** : Informations sur les continents, pour essayer d'apporter plus de précision, et également l'évolution de la maladie dans chaque continent (??).
- **Page 3** : Encore plus précis, on s'occupe désormais des pays les plus touchés par le coronavirus (??).
- **Page 4** : Dans la continuité de la page 3, nous avons également le taux de mortalité de chaque pays dû au Coronavirus (??).

A noter que je me suis inspiré du dashboard provenant de <https://covid19.who.int/region/amro/country/us> et de <https://ncov2019.live/> pour réaliser le mien, c'est pourquoi nous pouvons noter quelques similitudes.

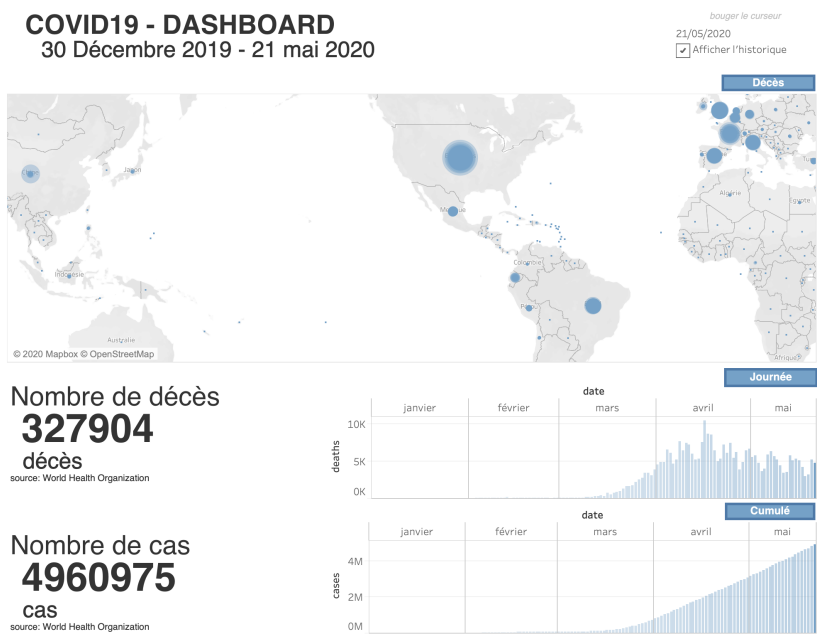


FIGURE 2.17 – Page 1 de notre dashboard

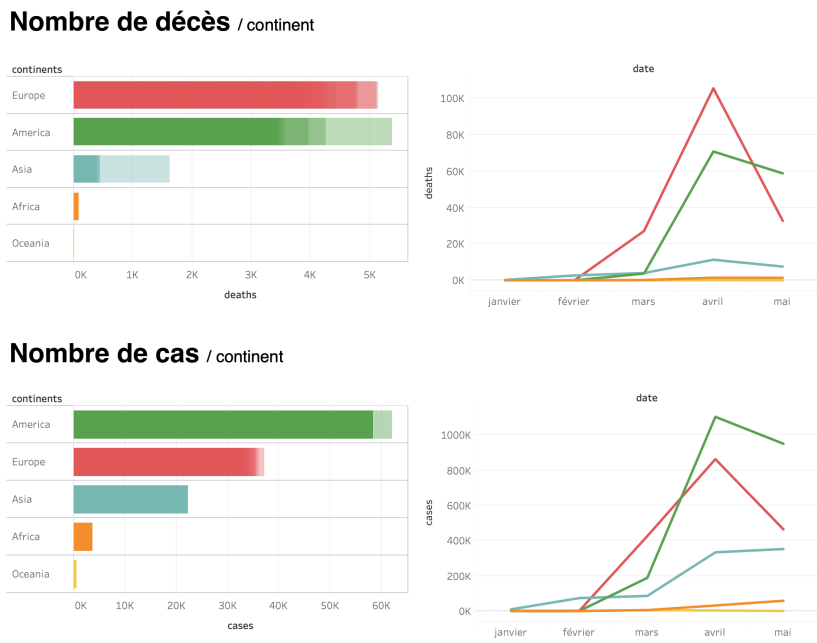


FIGURE 2.18 – Page 2 de notre dashboard

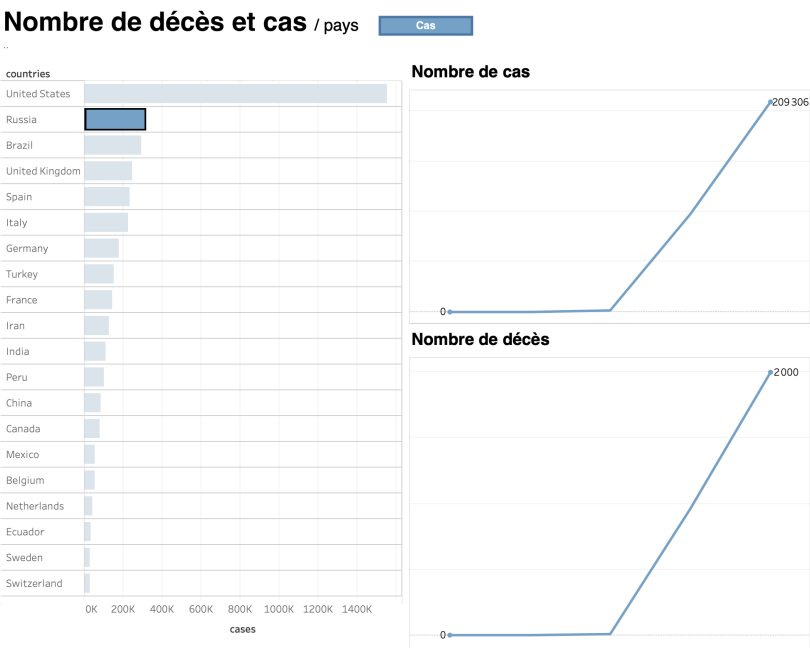


FIGURE 2.19 – Page 3 de notre dashboard

countries	cases	deaths	population	mortality rate %
Nicaragua	25	8	413 792 832	32,0%
France	143 845	28 132	9 579 175 892	19,6%
Sint_Maarten	77	15	2 364 702	19,5%
Belgium	55 983	9 150	1 633 355 724	16,3%
Yemen	180	29	1 196 944 854	16,1%
United Kingdom	248 293	35 704	9 507 925 713	14,4%
Italy	227 364	32 330	8 641 673 469	14,2%
Hungary	3 641	473	752 196 445	13,0%
Netherlands	44 447	5 748	2 464 035 431	12,9%
British_Virgin_Islands	8	1	1 668 912	12,5%
Sweden	31 523	3 831	1 456 194 025	12,2%
Antigua_and_Barbuda	25	3	6 066 018	12,0%
Spain	232 555	27 888	6 634 772 358	12,0%
Bahamas	97	11	25 066 600	11,3%
Belize	18	2	22 601 189	11,1%
Mexico	56 594	6 090	17 035 756 380	10,8%
Chad	565	57	975 098 313	10,1%
Liberia	238	23	318 052 482	9,7%
Jersey	306	29	6 728 400	9,5%
Montserrat	11	1	359 900	9,1%
Northern_Mariana_Islands	22	2	2 957 864	9,1%
Suriname	11	1	36 287 433	9,1%
United_States_Virgin_Isl...	69	6	6 204 666	8,7%

FIGURE 2.20 – Page 4 de notre dashboard

Chapitre 3

Conclusion

Pour conclure, le sujet du coronavirus et son impact sur le monde, nous a permis d'apprendre à utiliser l'outil de visualisation **Tableau**. **Tableau** semble être un excellent outil de visualisation, il est d'ailleurs reconnu, dans la plupart des sites que j'ai pu rencontrés, comme étant une des références en la matière. Mais je trouve parfois que nous sommes limités dans certaines actions, ou que certaines actions ne sont pas forcément évidentes/intuitives. Là où, par exemple, Google Data Studio a l'air d'être un peu plus intuitifs et surtout qu'il offre plus de possibilités dans la mise en forme de notre dashboard, des quelques exemples que j'ai pu voir. Néanmoins, **Tableau** reste un outil de grande qualité.