

Projet : Covid Situation

Projet de substitution au Stage / non effectué dans l'entreprise World Consulting (fictive)

TANG Kévin

L3 Informatique

Université Claude Bernard Lyon 1

Année 2021-2022

Tuteur : Alexandre MEYER

Sommaire

[**Introduction** 3](#_Toc111209734)

[**Environnement professionnel** 4](#_Toc111209735)

[**Mission** 5](#_Toc111209736)

[Objectif 5](#_Toc111209737)

[Outils 5](#_Toc111209738)

[Fichiers CSV 5](#_Toc111209739)

[Fichier GEOJSON 5](#_Toc111209740)

[D3JS 5](#_Toc111209741)

[Résultats 6](#_Toc111209742)

[Construction d’un tableau 6](#_Toc111209743)

[Construction d’une carte 6](#_Toc111209744)

[Actions de la souris 7](#_Toc111209745)

[Choix utilisateur 7](#_Toc111209746)

[**Retour sur expérience** 7](#_Toc111209747)

[**Conclusion et bilan** 8](#_Toc111209748)

[**Annexes** 9](#_Toc111209749)

[**Bibliographie** 11](#_Toc111209750)

# **Introduction**

N’ayant pas trouvé de stage durant cette année 2022, j’ai effectué un projet de substitution pour me permettre de valider ma 3e année de licence informatique à l’université Claude Bernard Lyon 1.

Etant intéressé par l’analyse et l’exploitation de données, j’ai décidé de mener un projet permettant une visualisation de données avec D3JS provenant d’un fichier spécifique en CSV.

Le jeu de données choisi répertorie les cas de COVID-19 du 15 février 2020 jusqu’au 14 mai 2022 dans le monde entier ainsi que d’autres données en lien avec le virus.

Voulant visualiser les données sur une carte interactive, il m’a aussi fallu ajouter un fichier GEOJSON me permettant de dessiner une carte du monde à partir des données de ce fichier.

Le contexte du projet nous amène donc dans l’entreprise World Consulting, un cabinet de conseil fictif imaginé pour l’occasion et qui aurait reçu pour mission par le gouvernement d’analyser la situation du COVID-19 dans le monde afin d’aider les autorités à prendre des décisions concernant l’arrêt ou la reprise des vols aériens vers certains pays.

Pour présenter mon expérience sur ce projet, nous allons tout d’abord expliciter l’environnement professionnel, bien que fictif, il n’en est pas moins intéressant de se pencher sur l’organisation d’une entreprise ayant des compétences aussi variées qu’un cabinet de conseil.

Puis nous aborderons le cœur même du projet qui est la mission effectuée, ses objectifs, la méthodologie, les difficultés rencontrées et les résultats obtenus durant ce projet.

Enfin, j’évoquerais mon ressenti sur cette expérience, les compétences que j’ai pu acquérir et ce que j’aimerais développer encore davantage.

# **Environnement professionnel**

Voici une entreprise type qui aurait pu m'accueillir pour réaliser ce projet, World Consulting, il s’agit d’une entreprise fictive que j’ai imaginé pendant la réalisation du projet et qui pourrait exploiter les données utilisées dans un scénario élaboré par mes soins.

Ce dernier présente l’entreprise comme un cabinet spécialisé dans les conseils à l’échelle international et missionné par l’Etat pour analyser la situation du COVID-19 dans le monde afin de rendre un avis concernant la nécessité d’arrêter certains vols internationaux vers et depuis des pays fortement touchés par la crise sanitaire.

Le cabinet de conseil est une organisation possédant plusieurs départements avec chacun son domaine d’expertise (Informatique, Finance, Juridique…) toute dirigée par une direction technique, elle-même sous la responsabilité de la direction générale du cabinet. Elle peut donc fournir une expertise technique dans des secteurs divers et variés.

Un organigramme du cabinet de conseil World Consulting se trouve en annexe (voir figure 1).

Le projet se déroule donc dans le département informatique du cabinet qui a reçu pour mission par la direction technique de réaliser une analyse de la situation sanitaire mondiale.

La première tâche revient au Data Scientist qui doit recueillir les données dont le cabinet aura besoin pour répondre à la demande du client et de mettre en forme ces données afin que celles-ci soient exploitables par la suite.

Ensuite, le Data Analyst, comme l’indique son rôle, va analyser, exploiter et interpréter les données extraites pour en tirer des observations utiles et essayer de prévoir à plus ou moins long terme le développement de la situation. Réaliser une visualisation graphique des données peut faire partie de son rôle afin de rendre un rapport plus clair de ses conclusions.

Ce rapport sera remis à d’autres départements afin de faire des analyses complémentaires dans d’autres domaines tels que l’évaluation des risques, des conséquences ou des coûts avant de rendre un avis complet au chef de projet en charge de la mission globale qui remettra l’avis final au client.

Ici, mon rôle se situe entre ces deux experts. Je dois réaliser une visualisation graphique des données extraites par le Data Scientist afin que le Data Analyst puisse fournir une analyse plus rapide en facilitant sa compréhension de la situation sanitaire.

# **Mission**

## Objectif

L’objectif du projet est de parvenir à créer une visualisation graphique de données contenues dans un fichier CSV avec D3JS.

## Outils

### Fichiers CSV

Pour cela, j’ai cherché un jeu de données accessible et surtout rapidement opérationnel, ce qui m’a conduit à consulter Kaggle, site célèbre qui accueille la plus grande communauté de Data Science au monde et regroupe de nombreuses données exploitables et libre d’accès provenant de domaines très divers tels que des données scientifiques, financières, géographiques ou culturelles.

J’ai voulu opter pour un sujet connu de tous, ce qui m’a amené à faire le choix de la crise sanitaire que nous subissons/avons subi ces dernières années.

J’ai donc découvert un jeu de données possédant deux fichiers CSV : le premier contient un résumé par pays des conséquences qu’a causé le virus tandis que le second est un journal quotidien de la situation sanitaire dans chaque pays.

Voici l’intitulé des données recueillies dans chaque fichier :

|  |  |
| --- | --- |
| Fichier 1 : « worldometer\_coronavirus\_summary\_data » | Fichier 2 : « worldometer\_coronavirus\_daily\_data » |
| Country  continent,total\_confirmed  total\_deaths,total\_recovered  active\_cases,serious\_or\_critical  total\_cases\_per\_1m\_population  total\_deaths\_per\_1m\_population  total\_tests  total\_tests\_per\_1m\_population  population | Date  Country  cumulative\_total\_cases  daily\_new\_cases  active\_cases  cumulative\_total\_deaths  daily\_new\_deaths |

Tableau : Tableau des intitulés des fichiers CSV

### Fichier GEOJSON

Voyant ces données, il m’est venu à l’idée de construire une carte du monde pour visualiser plus facilement ces données, il m’a donc aussi fallu me procurer un fichier GEOJSON qui est un fichier contenant des données géospatiales au format JSON et donc exploitable en Javascript tout comme D3JS.

### D3JS

Recommandé par mon tuteur académique, D3JS est une librairie permettant une manipulation et une visualisation de données de manière simple et efficace provenant de fichiers lourds.

Ne connaissant pas D3JS avant ce projet, cela a été une belle découverte et un joli défi à entreprendre qui m’a donc permis d’afficher les résultats du projet sur une page HTML. Je ne me suis pas attardé sur la partie CSS qui n’était pas un objectif prioritaire, ce qui explique pourquoi elle ne se trouve pas dans un fichier séparé.

## Résultats

### Construction d’un tableau

Pour débuter, je me suis intéressé sur la manière d’extraire les données d’un fichier CSV, c’est pourquoi la première tâche que je me suis fixé était de construire un tableau affichant l’entièreté des données du premier fichier sur une page HTML. Un protocole empêchant l’accès aux fichiers locaux, il m’a fallu télécharger et utiliser l’extension *Live Server* sur Visual Studio Code simulant ainsi un serveur local et me permettant de construire une page dynamique.

Pour me faciliter la tâche, J’ai commencé par créer deux variables qui contiennent respectivement l’intitulé des colonnes (*columns*) et un tableau contenant les données brutes (*dataset*). La première est codée de manière grossière en écrivant chaque intitulé à la main car je me suis concentré sur l’extraction des données pour la deuxième variable.

Après avoir posé les bases du tableau sur la page HTML, je me suis servi de D3JS pour faire appel au fichier contenant les données et m’initier à ce langage nouveau pour moi tout en revoyant les bases du Javascript et notamment la notion de promesse.

Une fois les données extraites et facilement accessibles via les deux variables, il me suffisait donc d’écrire une fonction pour construire le tableau avec ces deux paramètres. Le tableau est composé de deux éléments distincts : la balise *thead* qui représente l’ensemble des intitulés de chaque colonne et la balise *tbody* qui contient les données. J’ai pour cela fait le choix d’utiliser deux méthodes différentes selon la complexité du code et ma compréhension : pour construire le *thead* j’ai choisi d’utiliser la méthode de D3JS qui construit une nouvelle case pour chaque élément dans la variable *columns* alors que pour le *tbody*, j’ai réalisé des boucles et construit les cellules avec du simple Javascript.

Résultat en annexe (figure 2).

### Construction d’une carte

Arrivé ici, pas de choix possible, on en arrive au cœur même du rôle de D3JS.

Comme pour la partie précédente, j’ai commencé par définir des variables qui me seront utiles par la suite tel que les dimensions de la carte et les couleurs.

Avant de construire la carte du monde, j’ai mis en place le seul élément qui restera immuable dans la représentation graphique du monde : la légende. Cette dernière a été organisé de tel sorte qu’elle montre une couleur de plus en plus sombre dans une teinte de rouge lorsque sa valeur augmente.

Vient ensuite la fonction qui dessine la carte avec les données du GEOJSON, ceci se fait étonnamment facilement une fois les données extraites, on remarque toute la puissance de la librairie D3JS qui nous simplifie grandement la tâche.

La difficulté rencontrée ici était d’accorder les données provenant des deux fichiers. En effet, provenant de sources différentes, ces derniers ne possédaient pas de moyen d’identification stricte autre que le nom pour chaque pays. Des différences dans la syntaxe comme les majuscules, les tirets ou les accents ne permettaient pas le rapprochement entre les deux fichiers. Il m’a donc fallu parcourir les documents et d’harmoniser les noms des pays un par un.

Résultat en annexe (figure 3).

### Actions de la souris

Dans la fonction de construction de la carte, j’ai ajouté plusieurs actions qui peuvent être réalisées avec la souris sur cette carte :

* Au survol d’un pays, ce dernier se colore en vert et le nom du pays s’affiche dans un coin de la carte. Il revient évidemment à son état initial lorsque la souris ce territoire.
* Quand on clique sur un pays, une nouvelle partie de la page se découvre. On y aperçoit un graphique utilisant les données journalières du deuxième fichier CSV (voir figure 4).

### Choix utilisateur

Que ce soit pour les données représentées sur la carte du monde ou sur le graphique journalier, il est possible pour l’utilisateur de basculer sur les différents intitulés grâce à des listes déroulantes et modifier ainsi les données représentées sans réinitialiser la page.

# **Retour sur expérience**

Ayant eu le choix du projet et étant très intéressé par la science des données, j’étais très enthousiaste à l’idée de prendre en main D3JS que je ne connaissais pas du tout, même de nom, avant de commencer ce projet.

Effectuer des recherches a donc occupé la plus grande majorité du temps consacré dans ce projet. En effet, débuter dans un environnement inconnu avec un langage avec lequel j’ai peu d’affinité a été un vrai défi pour moi. J’ai dû me remettre dans le bain du Javascript avec notamment la gestion des fonctions asynchrones et des promesses, ce qui ne m’avait pas vraiment manqué.

Il est toujours difficile de commencer un projet seul, on ne sait pas si ce que l’on fait suit les bonnes méthodes, si on comprend bien les choses ou si on est en bonne voie. Devoir prendre des décisions importantes seul qui peuvent influencer la suite du projet est une grande responsabilité et engendre beaucoup de stress, ce fut le cas lorsque je n’arrivais pas à avancer comme je le voulais durant ce projet. J’imagine que cette responsabilité doit encore plus grande lorsque ce choix implique de nombreuses personnes comme c’est le cas en entreprise.

Au fur à mesure que le projet se mettait en place, j’ai compris qu’il était important de définir dès le départ les résultats ce que l’on souhaitait obtenir car avancer selon ses envies conduit à des doutes et n’amène pas forcément à des résultats probants et aboutis.

De plus, parvenir à atteindre un objectif amène toujours des questions sur la suite du projet : « que faire ensuite ? ». C’est le type de question qu’on se pose tout au long du développement et même maintenant : « ai-je vraiment terminé ? », « Que pourrais-je ajouter de plus pour compléter le projet ? ».

Finalement, D3JS s’est avéré être facile à utiliser et très pratique bien que cela ne soit pas encore très intuitif à prendre en main pour moi. C’est un outil qui s’avère très puissant pour ceux qui en comprenne toutes les subtilités et cette petite expérience m’a permis d’entrevoir une grande variété de possibilités d’utilisation. Je suis content d’avoir découvert cette librairie car cela faisait partie de mes questionnements pendant mes études de savoir comment réaliser facilement ce genre de graphique à partir de bases de données aussi grandes, cela me sera sans doute très utile à l’avenir.

# **Conclusion et bilan**

L’objectif de ce projet était de réaliser une visualisation graphique de données provenant d’un fichier CSV, ce que j’ai réussi à réaliser avec deux fichiers en permettant une présentation des informations à la fois sous forme d’une carte mondiale et d’un graphique et en liant ces deux constructions par des actions de l’utilisateur avec un choix d’options et des actions de la souris.

Finalement, le but du projet a été atteint bien que j’aurais aussi voulu y ajouter une touche plus analytique. Cela m’a aussi permis de revoir des notions vues au cours de ma formation en HTML, JavaScript et CSS dans une moindre mesure.

Ce projet a été une expérience enrichissante pour moi qui ne connaissait pas D3JS avant et qui n’avait jamais réalisé de la visualisation de données, je pense que cette librairie me suivra encore un moment dans ma carrière et me sera très utile à l’avenir même si je ressens le besoin de grandement m’améliorer avec ce nouvel outil. J’espère continuer dans cette voie et me faire une place dans la Data Science.

# **Annexes**

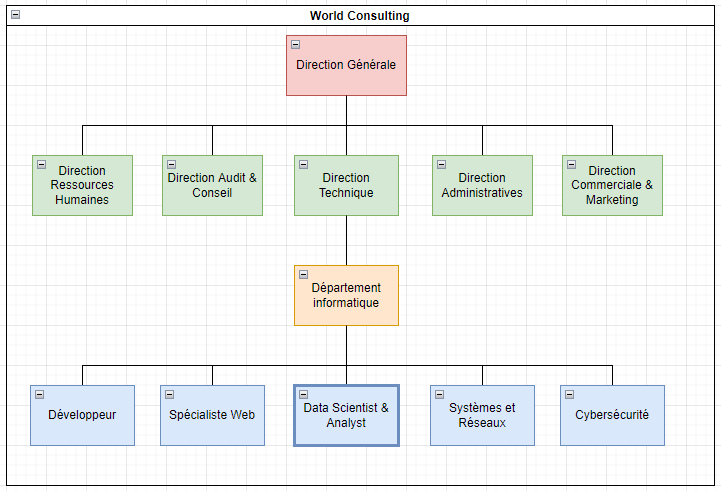


Figure 1 : Organigramme du cabinet World Consulting

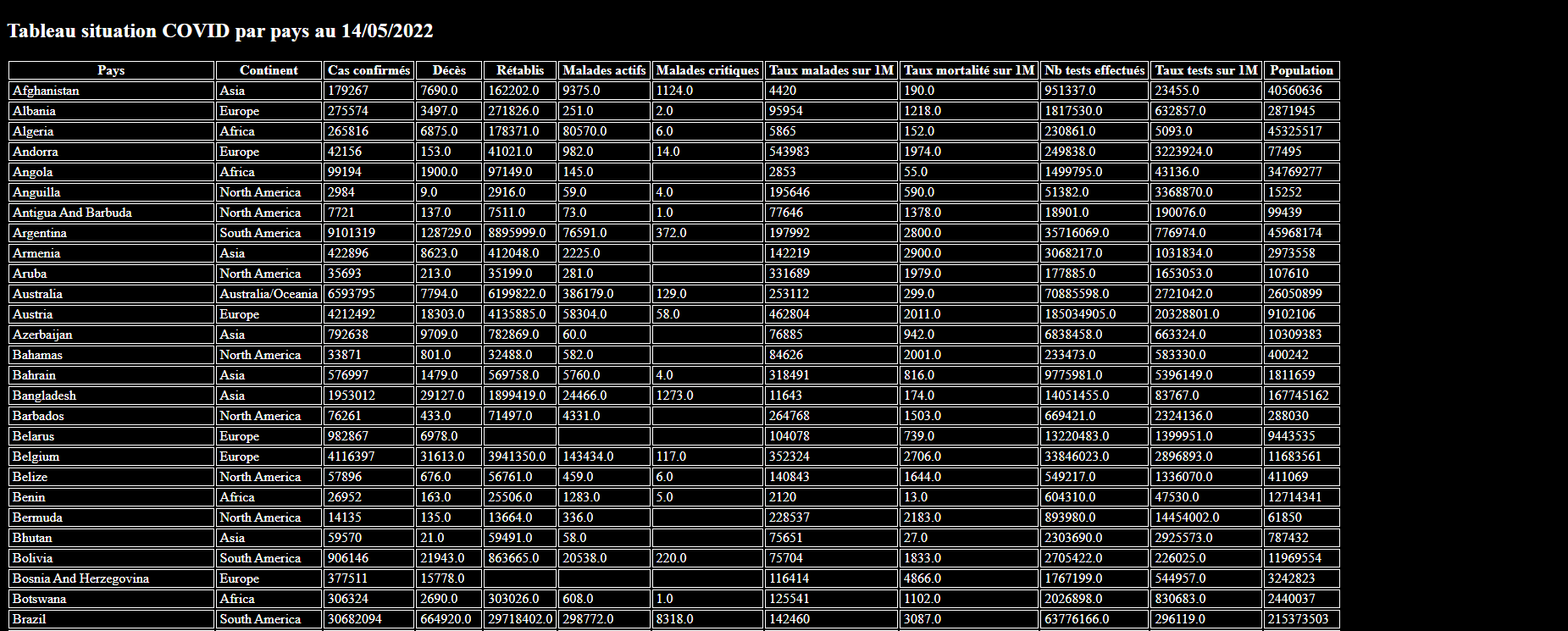


Figure : Début du tableau chiffrant la situation sanitaire des pays

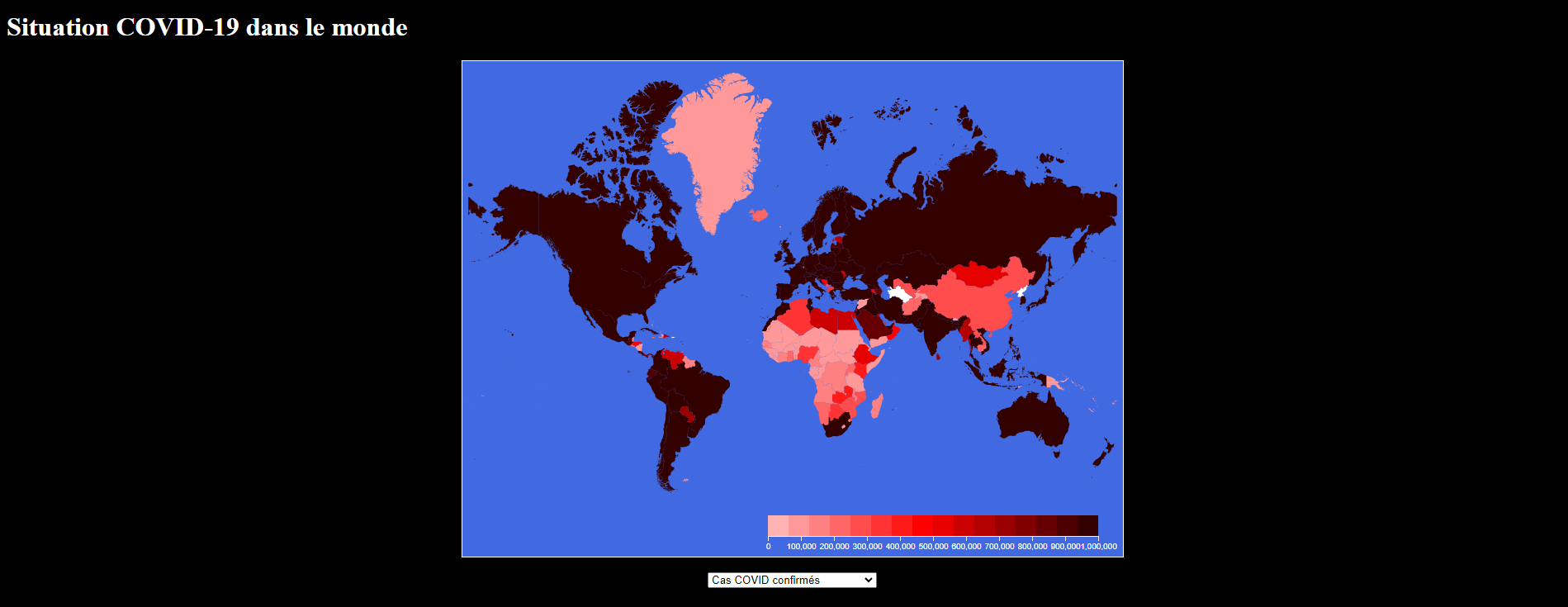


Figure : Carte de la situation sanitaire mondiale

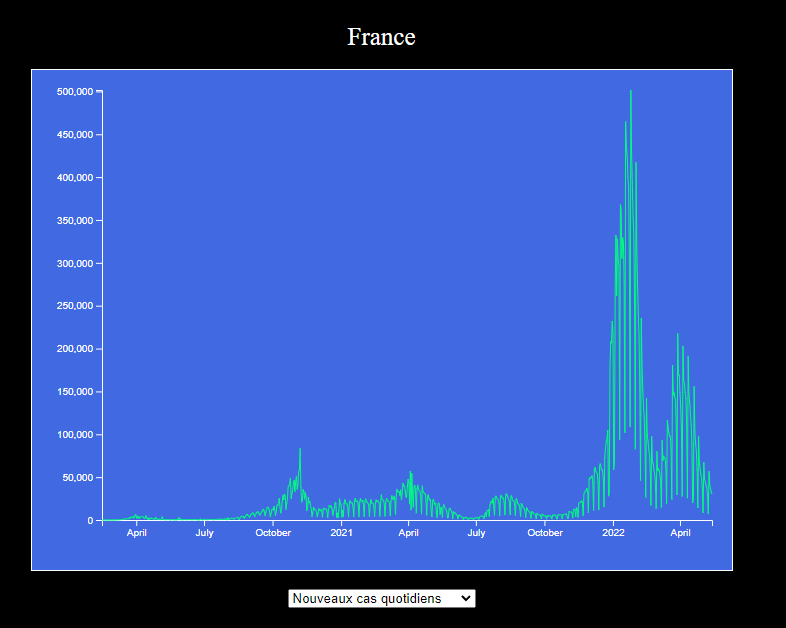


Figure : Graphique des nouveaux cas quotidiens de COVID-19 en France

# **Bibliographie**

* Source des fichiers CSV

ASSAKER Joseph. *COVID-19 Global Dataset*, 2020.

Adresse URL : <https://www.kaggle.com/datasets/josephassaker/covid19-global-dataset>

* Source du fichier GEOJSON

Open Data Commons Public Domain Dedication and Licence, Natural Earth, Lexman et Open Knowledge Foundation. *Country Polygons as GeoJSON*, 2018.

Adresse URL : <https://datahub.io/core/geo-countries>

* Librairie D3JS

D3Js.org : <https://d3js.org/>

* Ressources pédagogiques

TutorialsTeacher : <https://www.tutorialsteacher.com/d3js>

Le Blog de Lulu : <https://lucidar.me/fr/d3.js/d3-js-summary/>

Datavis : <https://www.datavis.fr/>

The D3.js Graph Gallery : <https://d3-graph-gallery.com/index.html>