



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
Environnement professionnel.....	4
Mission.....	6
Objectif.....	6
Outils.....	6
Fichiers CSV.....	6
Fichier GEOJSON	6
D3JS.....	7
Résultats	7
Construction d'un tableau	7
Construction d'une carte.....	8
Actions de la souris.....	10
Construction d'un graphique	10
Retour sur expérience.....	12
Conclusion et bilan.....	13
Table des illustrations.....	14
Annexes.....	15

Introduction

N'ayant pas trouvé de stage durant cette année 2022, j'ai effectué un projet de substitution pour me permettre de valider ma 3^e 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 CSV spécifique.

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, j'ai eu l'idée d'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 internationale 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.

Voici l'organigramme du cabinet de conseil World Consulting :

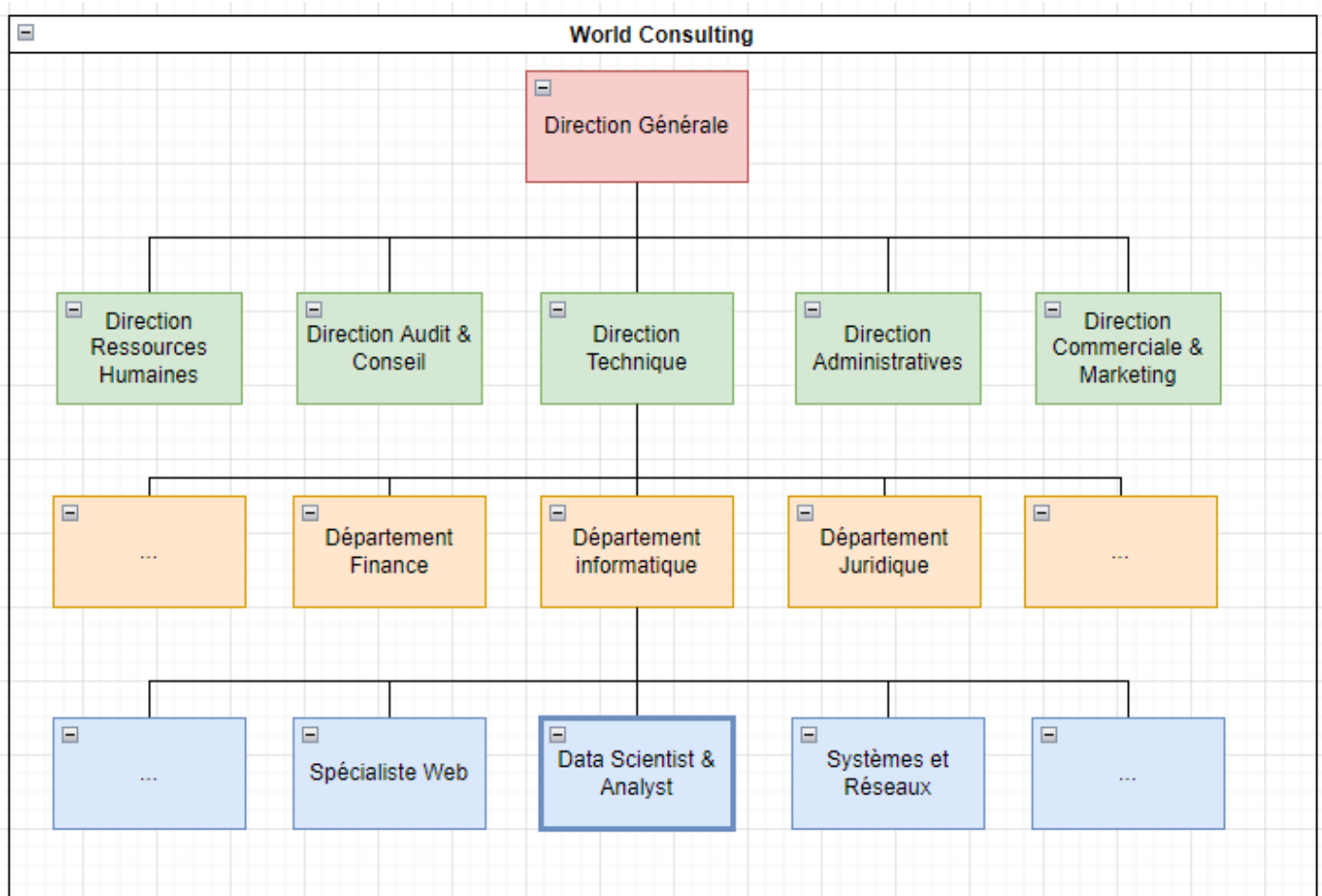


Figure 1 : Organigramme du cabinet World Consulting

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 et qui remettra enfin 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 1 : 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.

Le fichier, trouvé sur le site datahub.io, possède pour source le site Natural Earth Data qui contient des données géographiques à différentes échelles dans le but de fournir une base de données pour des projets de construction de carte, ce qui est idéal pour notre projet.

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, j'ai parcouru plusieurs sites internet tutoriaux expliquant les bases de la librairie avant de commencer à développer le projet sur une page HTML.

Résultats

Construction d'un tableau

Pour débiter, 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

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.

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. En effet, c'était déjà une difficulté pour moi lors de mon parcours en licence de comprendre et prévoir l'ordre d'exécution avec des fonctions asynchrones.

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.

La figure ci-dessous montre le tableau qui en résulte. Comme on peut le voir, il s'agit seulement du début du tableau, les pays sont rangés par ordre alphabétique.

Pays	Continent	Cas confirmés	Décès	Rétablis	Malades actifs	Malades critiques	Taux malades sur 1M	Taux mortalité sur 1M	Nb tests effectués	Taux tests sur 1M	Population
Afghanistan	Asia	179267	7690.0	162202.0	9375.0	1124.0	4420	190.0	951337.0	23455.0	40560636
Albania	Europe	275574	3497.0	271826.0	251.0	2.0	95954	1218.0	1817530.0	632857.0	2871945
Algeria	Africa	265816	6875.0	178371.0	80570.0	6.0	5865	152.0	230861.0	5093.0	45325517
Andorra	Europe	42156	153.0	41021.0	982.0	14.0	543983	1974.0	249838.0	3223924.0	77495
Angola	Africa	99194	1900.0	97149.0	145.0		2853	55.0	1499795.0	43136.0	34769277
Anguilla	North America	2984	9.0	2916.0	59.0	4.0	195646	590.0	51382.0	3368870.0	15252
Antigua And Barbuda	North America	7721	137.0	7511.0	73.0	1.0	77646	1378.0	18901.0	190076.0	99439
Argentina	South America	9101319	128729.0	8895999.0	76591.0	372.0	197992	2800.0	35716069.0	776974.0	45968174
Armenia	Asia	422896	8623.0	412048.0	2225.0		142219	2900.0	3068217.0	1031834.0	2973558
Aruba	North America	35693	213.0	35199.0	281.0		331689	1979.0	177885.0	1653053.0	107610
Australia	Australia/Oceania	6593795	7794.0	6199822.0	386179.0	129.0	253112	299.0	70885598.0	2721042.0	26050899
Austria	Europe	4212492	18303.0	4135885.0	58304.0	58.0	462804	2011.0	185034905.0	20328801.0	9102106
Azerbaijan	Asia	792638	9709.0	782869.0	60.0		76885	942.0	6838458.0	663324.0	10309383
Bahamas	North America	33871	801.0	32488.0	582.0		84626	2001.0	233473.0	583330.0	400242
Bahrain	Asia	576997	1479.0	569758.0	5760.0	4.0	318491	816.0	9775981.0	5396149.0	1811659
Bangladesh	Asia	1953012	29127.0	1899419.0	24466.0	1273.0	11643	174.0	14051455.0	83767.0	167745162
Barbados	North America	76261	433.0	71497.0	4331.0		264768	1503.0	669421.0	2324136.0	288030
Belarus	Europe	982867	6978.0				104078	739.0	13220483.0	1399951.0	9443535
Belgium	Europe	4116397	31613.0	3941350.0	143434.0	117.0	352324	2706.0	33846023.0	2896893.0	11683561
Belize	North America	57896	676.0	56761.0	459.0	6.0	140843	1644.0	549217.0	1336070.0	411069
Benin	Africa	26952	163.0	25506.0	1283.0	5.0	2120	13.0	604310.0	47530.0	12714341
Bermuda	North America	14135	135.0	13664.0	336.0		228537	2183.0	893980.0	14454002.0	61850
Bhutan	Asia	59570	21.0	59491.0	58.0		75651	27.0	2303690.0	2925573.0	787432
Bolivia	South America	906146	21943.0	863665.0	20538.0	220.0	75704	1833.0	2705422.0	226025.0	11969554
Bosnia And Herzegovina	Europe	377511	15778.0				116414	4866.0	1767199.0	544957.0	3242823
Botswana	Africa	306324	2690.0	303026.0	608.0	1.0	125541	1102.0	2026898.0	830683.0	2440037
Brazil	South America	30682094	664920.0	29718402.0	298772.0	8318.0	142460	3087.0	63776166.0	296119.0	215373503

Figure 2 : Début du tableau chiffrant la situation sanitaire des pays

Construction d'une carte

On arrive ici au cœur même du rôle de D3JS.

Le site Datavis qui contient de nombreux exemples de cartes différentes et complexes m'a beaucoup aidé à comprendre le fonctionnement d'une telle structure. Il possède des tutoriels très détaillés expliquant chaque étape de la construction, ce qui m'a déjà permis de bâtir un modèle simple de mappemonde, ceux du site étant bien plus complexes.

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 organisée de telle sorte qu'elle montre une couleur de plus en plus sombre dans une teinte de rouge lorsque sa valeur augmente. Les pays dont nous n'avons aucune donnée sont affichés en blanc.

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 puis écrire une fonction qui puisse trouver l'indice d'un pays avec son nom afin de pouvoir faire le lien entre les deux documents, c'est ce dont je me suis servi pour colorier chaque pays selon la donnée mise en valeur et la légende.

Nous pouvons donc observer ci-dessous la carte obtenue à partir des données grâce à D3JS et au fichier GEOJSON.

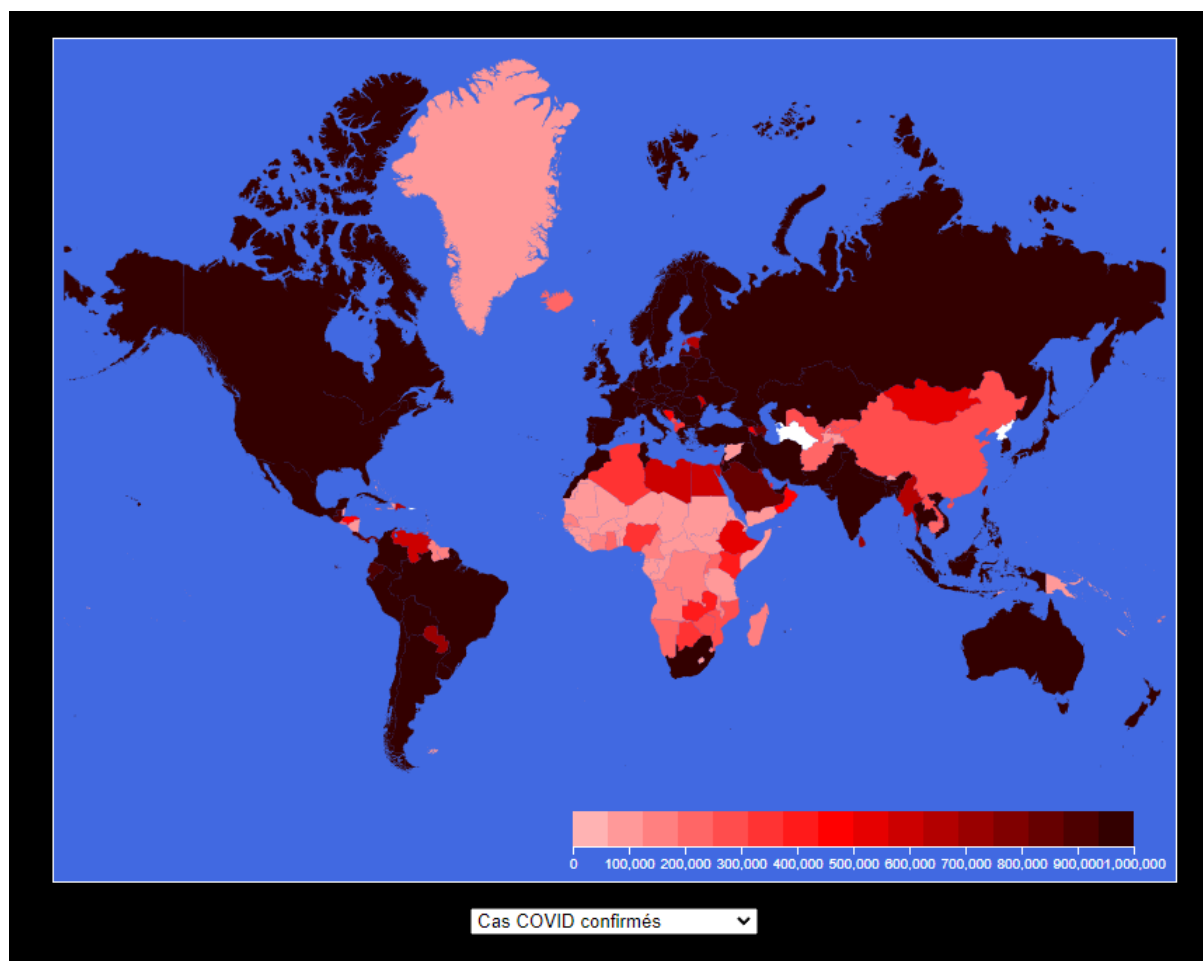


Figure 3 : Carte de la situation sanitaire mondiale

Les choix concernant les options d’affichage permettent de visualiser les autres données du fichier sur la carte. Pour permettre cela, il était nécessaire de réorganiser pour planifier la suppression et la reconstruction des éléments. Ainsi, lorsque l’utilisateur appuie sur une option de la liste déroulante, une fonction est appelée changeant la variable possédant l’option sélectionnée initialement et appelant elle-même la fonction de construction afin d’afficher la nouvelle carte en supprimant la précédente.

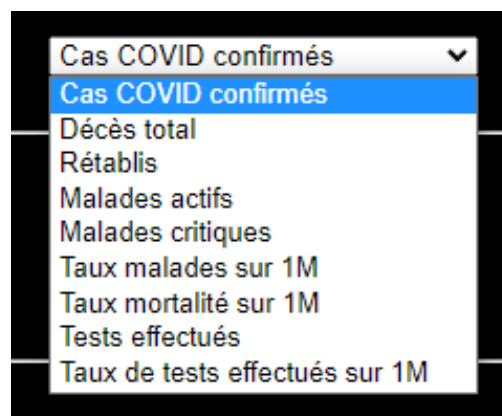


Figure 4 : Liste déroulante de la carte

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 quitte ce territoire.
- Quand on clique sur un pays, cela fait apparaître un graphique montrant l'évolution du virus dans le temps pour le pays concerné.

Ces actions sont facilement réalisables grâce à D3JS, la difficulté réside surtout dans la construction du graphique que nous voyons ensuite.

Construction d'un graphique

Comme dit précédemment, il est possible de cliquer sur chaque pays de la carte pour faire apparaître une nouvelle partie de la page. On y aperçoit un graphique utilisant les données journalières du deuxième fichier CSV.

Sa construction n'est finalement pas si difficile si l'on procède un élément à la fois, en dissociant la courbe et les axes mais cela requiert tout de même de connaître les différentes fonctions de construction d'un graphique pour utiliser au mieux les données.

Voici un exemple des graphiques qui peuvent s'afficher, l'image ci-dessous montre l'évolution du nombre de cas positifs au COVID-19 quotidien en France.

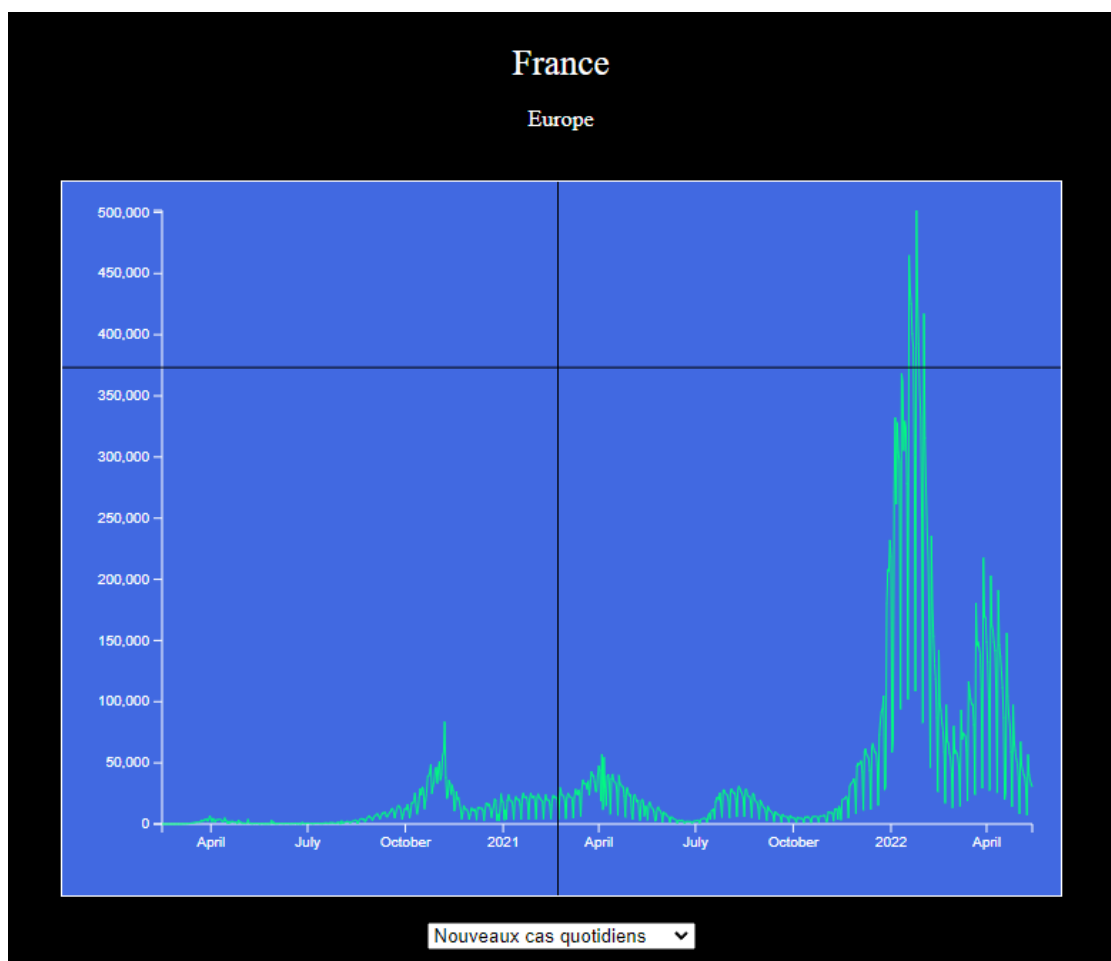


Figure 5 : Graphique des nouveaux cas quotidiens de COVID-19 en France

Ce graphique montre par défaut l'évolution du nombre de cas de personnes infectées par le virus au cours du temps, mais l'utilisateur a la possibilité de choisir de faire figurer d'autres données parmi celles du fichier 2.

La structure du code pour la déconstruction et la reconstruction du graphique selon l'option choisie est identique à celle pour la mappemonde.

A noter que lorsque l'utilisateur change d'option sur le graphique, celle-ci est conservée lorsqu'il sélectionne un autre pays sur la carte par la suite.

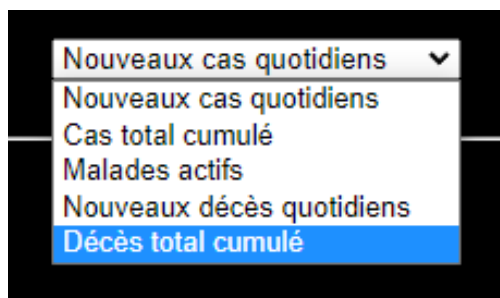


Figure 6 : Liste déroulante du graphique

Dans les derniers instants du projet, j'ai pensé faire afficher sur ce graphique la valeur représentée pour chaque point de la courbe lorsque la souris survole celle-ci mais par manque de temps, je n'ai pas pu identifier un moyen de récupérer la donnée voulue. Je me suis contenté d'y dessiner une ligne horizontale et une ligne verticale se croisant sur la position de la souris pour permettre à l'utilisateur de lire plus facilement le graphique.

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, débiter 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.

La documentation sur D3JS laisse entrevoir des possibilités de création innombrables mais les sites internet et vidéos exploitant la librairie s'avèrent peu nombreuses, exhaustives et ne montre que des bases de son utilisation. Je n'ai pas trouvé non plus à travers le web de cas concret d'utilisation de D3JS dans un environnement professionnel, c'est sans doute ce qu'il m'a fait défaut en n'ayant pas pu réaliser mon stage en entreprise.

De plus, le scénario élaboré au début du rapport ainsi que la structure de l'entreprise d'accueil sont entièrement fictifs et je ne sais pas si le déroulement des événements est pertinent dans le cadre d'un réel projet de ce genre.

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.

En commençant le projet, j'avais quelques vagues idées concernant ce dont je voulais réaliser avec les données recueillies, la plupart me sont venues pendant la réalisation et j'étais très inquiet à l'idée que le contenu soit insuffisant. A chaque fois qu'un objectif que je m'étais fixé était atteint, je ne savais pas comment continuer et ce que je devais faire par la suite.

Après être parvenu au but du projet, j'ai pensé y ajouter une partie plus analytique des données mais n'ayant aucune idée de ce je pourrais montrer, je n'ai pas vraiment approfondi le sujet.

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 comprennent 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, 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.

Avoir une expérience réelle en entreprise est évidemment ce qui a manqué car j'aurais voulu voir comment se déroule le développement et l'organisation des équipes pour de tels projets.

Pour finir, 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. C'est donc une porte d'entrée intéressante dans le monde de la Data Science, j'espère pouvoir continuer dans cette voie.

Table des illustrations

Figure 1 : Organigramme du cabinet World Consulting	4
Tableau 1 : Tableau des intitulés des fichiers CSV	6
Figure 2 : Début du tableau chiffrant la situation sanitaire des pays	8
Figure 3 : Carte de la situation sanitaire mondiale.....	9
Figure 4 : Liste déroulante de la carte	9
Figure 5 : Graphique des nouveaux cas quotidiens de COVID-19 en France	10
Figure 6 : Liste déroulante du graphique.....	11

Annexes

- 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>

Youtube : <https://www.youtube.com/>