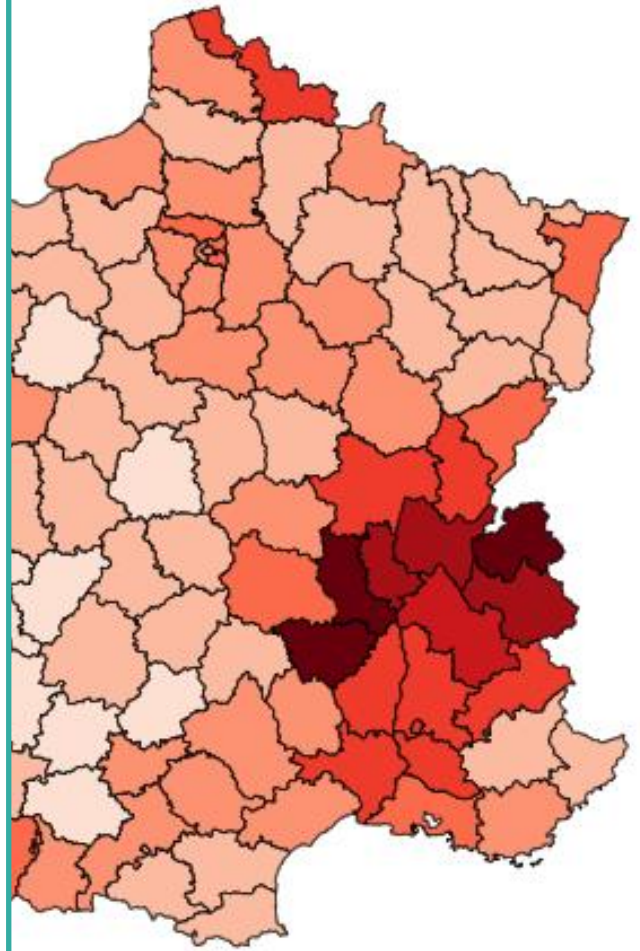


INF552

Projet de visualisation

Etude de la propagation du coronavirus en France



14 DECEMBRE

CALDAS Louis

TELECOM
SudParis



IP PARIS

Table des matières

I. Contexte de la visualisation.....	3
II. Fonctionnalités et choix esthétiques.....	3
III. Point sur les données.....	7
IV. Annexes.....	8

I. Contexte de la visualisation

L'objectif de la visualisation que j'ai réalisé dans le cadre du cours INF552 – Data Visualization vient d'un besoin simple : **tenter de mieux comprendre la propagation et le comportement de l'épidémie de la Covid-19 via l'exploitation de ces données statistiques**. Cette idée m'est notamment venu suite à une discussion avec un alumni de l'Institut Mines Télécom Business School et fondateur du site <https://covidtracker.fr/> Guillaume Rozier avec qui j'ai pu échanger au fur et à mesure de ma visualisation.

Après quelques recherches au niveau des bases de données disponibles en ligne à propos de la maladie, je me suis tourné vers les informations fournies par l'état. Plus particulièrement celle concernant le taux d'incidence quotidien de l'épidémie (à l'échelle nationale, régionale et départementale). Et j'ai essayé à travers ma visualisation de répondre à la question suivante :

Existe-t-il un réel lien entre densité de population et propagation du virus ?

J'ai également souhaité à travers ma visualisation tenter de montrer la réalité d'une seconde vague épidémique ainsi que l'intérêt de mesures restrictives pour freiner la propagation de la maladie.

II. Fonctionnalités et choix esthétiques

Tout d'abord, au niveau technique, la visualisation a été réalisé en javascript à l'aide des bibliothèques D3.js et Vega-Lite (étudiées lors des travaux pratiques du cours INF552). Une démonstration de la visualisation est disponible à l'adresse suivante : http://louiscaldas.fr/INF552/covid_map.html et une vidéo est jointe à ce rapport afin de se faire une idée des fonctionnalités.

Le travail final comprend 3 visualisations distinctes, l'une reposant sur des principes de Geovisualization (D3.js), l'une représentant une heatmap (Vega-Lite) et la dernière représentant un bar chart (Vega-Lite). Dans la suite, je vais détailler les différentes fonctionnalités de ces différentes visualisations ainsi que les choix esthétiques dans leur réalisation.

La visualisation dans sa version finale permet notamment la visualisation de certaines caractéristiques des régions/départements française :

- Visualisation via une carte choroplèthe de la population française par région ou par département.
- Visualisation via une carte choroplèthe du taux d'incidences par région ou par département. Ce taux pouvant être sous la forme de la moyenne mensuel ou de la somme du nombre de cas sur les 7 derniers jours.
- Visualisation via une carte choroplèthe de la densité de population par région ou par département.

Le choix des cartes choroplèthes vient de la nécessité d'encoder une caractéristique par zone géographique (qui sont plus ou moins toute de même taille en France ce qui permet de ne pas avoir de fausse intuition quant à la variable quantitative utilisée). Ces 3 visualisations sont réalisées à l'aide d'un fichier geojson de la France provenant du site (<https://france-geojson.gregoiredavid.fr/>). Ces fichiers m'ont permis de designer facilement l'espace géographique à l'aide de l'outil path de D3.JS via la projection conique conforme de Lambert centré sur la France. Il restait un problème à régler : celui de la légende de ces différents paramètres (taux d'incidences, densité de population et population), en effet leur étendu (maximum – minimum) est grande et leur valeur est singulière, c'est pourquoi pour rendre la visualisation plus signifiante j'ai choisi d'utiliser une échelle séparée en 9 quantiles enfin de capter au premier regard les zones géographiques disposant des mêmes caractéristiques. Le choix des couleurs des différentes légendes est arbitraire mais basé sur des nuances de couleurs (pour assurer une perceptibilité par le maximum de personnes), bleu pour la population des régions, vert pour la population des départements (ne pas garder la même couleur permet d'éviter d'assimiler la population d'une région à celle d'un département) et rouge pour le taux d'incidence (couleur utilisé par les conférences de presses gouvernementales depuis le début de la crise et donc associée à cet indicateur ainsi qu'à quelque chose de négatif).

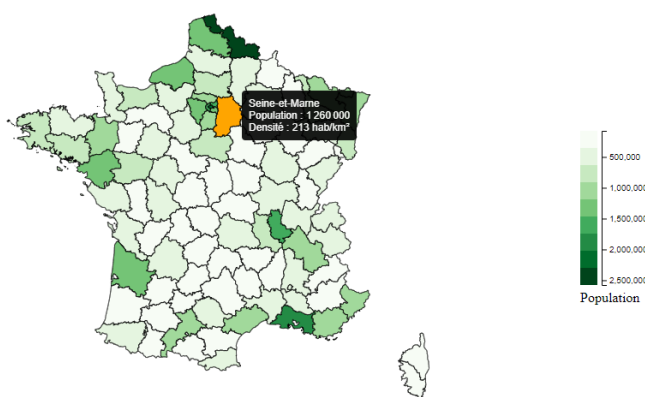


Figure 1 : Carte choroplète représentant la population des différents départements français (métropole)

La visualisation propose également une interaction avec les cartes choroplèthes : lorsque l'utilisateur clique sur une région, un zoom est déclenché pour afficher en « gros » les départements de la région choisie et la variable associée.

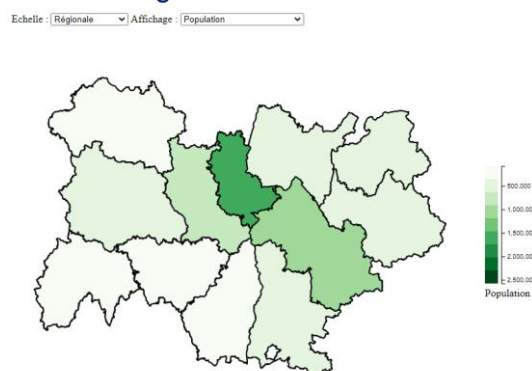


Figure 2 : Visualisation après zoom sur les départements d'une région

Puis dans un second temps, j'ai tenté au sein de ma visualisation de décorréler les paramètres d'agréations comme la population ou la densité de population et le taux d'incidence. Pour cela, je devais trouver comment sur une seule carte réussir à dissocier ces paramètres : les deux étant quantitatifs. J'ai donc naturellement commencé à chercher deux « magnitude channels » me permettant de transmettre de l'information sans perdre l'utilisateur. J'ai alors décidé de quantifier la variable d'agrégation par des cercles dont la taille serait associée à la valeur de cette dernière et à les coloriser en fonction du taux d'incidence (en faisant varier les nuances de rouges), en effet ce choix me semblait le plus intuitif pour l'utilisateur. On associe facilement (dans le cas par exemple de la densité de population) un cercle à gros rayon à une grande quantité de personnes et comme expliqué précédemment la couleur rouge est dans l'imaginaire collectif depuis quelques mois associé aux taux d'incidences (via les médias, les conférences de presse...).

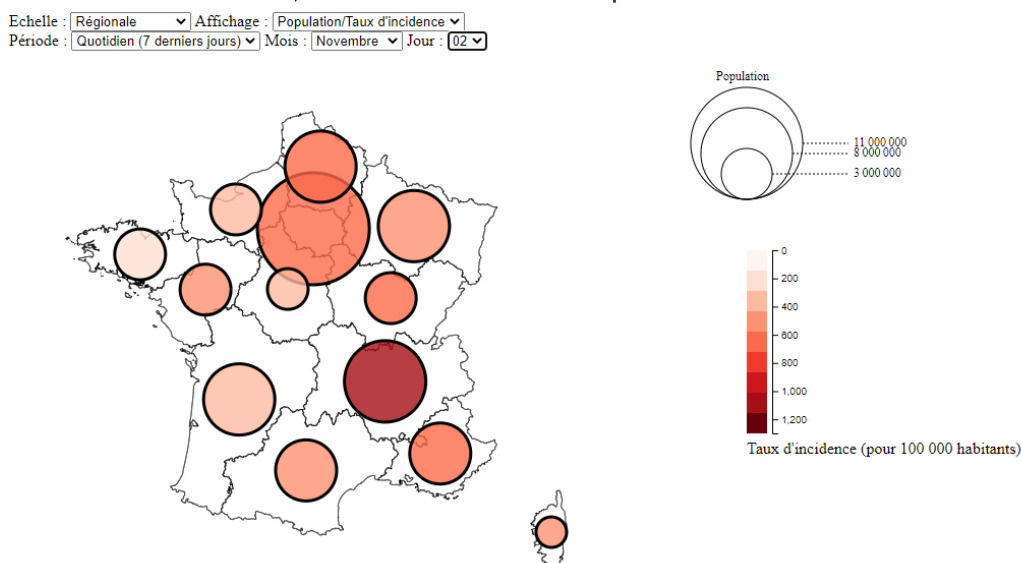


Figure 3 : Affichage du temps d'incidence en fonction de la population

J'aurais également pu choisir le fait de réaliser un cartogramme cependant, je trouvais cette visualisation moins claire, une partie de notre regard était attiré par essayer de reconstruire l'image de la France basique que nous avons en tête (sans déformation des zones géographiques), exemple de l'utilisation de ce type de visualisation sur la Suisse en annexe.

Cependant, les cercles se chevauchant, et la présence de la carte en arrière-plan étaient des sources de distractions du regard et ne permettait pas de se rendre réellement compte du lien entre les deux variables étudiées. C'est pourquoi j'ai décidé d'utiliser l'outil de visualisation de graph « Force directed layout » afin de rassembler les cercles entre eux (circle packing). Quitte à perdre l'information de la zone géographique (qui dans le cadre de l'étude n'est pas essentielle), la nouvelle visualisation permet de bien se faire une idée du lien entre variable d'agrégation et taux d'incidence (cf annexe pour une visualisation à l'échelle départementale). Cependant, à chaque changement de temporalité pour le taux d'incidence, on repart de la carte de la France pour permettre à l'utilisateur de se faire une idée de la provenance des cercles sans utiliser le survol.

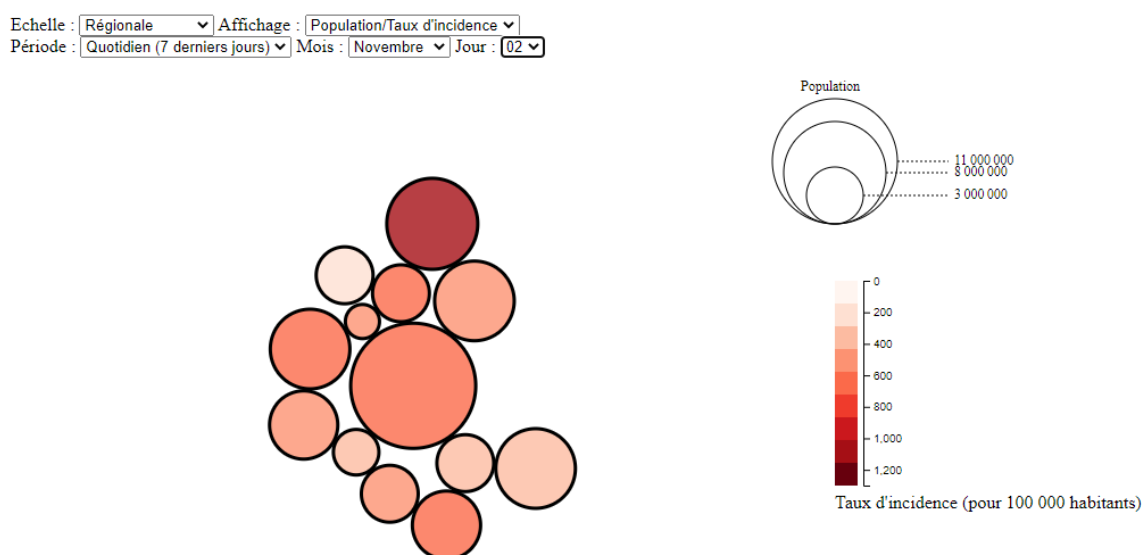


Figure 4 : Affichage du lien entre population et taux d'incidence sous la forme de circle packing

Cette dernière visualisation permet de remarquer qu'il existe effectivement un lien entre population/densité de population et taux d'incidence durant les périodes de circulation du virus (les grands cercles sont en général plus colorés que les petits) mais qu'il n'est cependant pas systématique (cela se voit d'autant plus dans le cadre de l'étude sur la densité de population, où la région Ile de France écrase les autres mais son taux d'incidence n'est pas plus élevé de façon proportionnel). **Cependant, il faut être honnête et dire que le lien densité/taux d'incidence ne saute pas aux yeux. Tout du moins avec cette visualisation, l'objectif n'est pas complètement atteint.**

Enfin, dans le but d'appuyer la visualisation globale, je lui ai ajouté deux graphiques réalisés à l'aide de la librairie Vega-Lite.

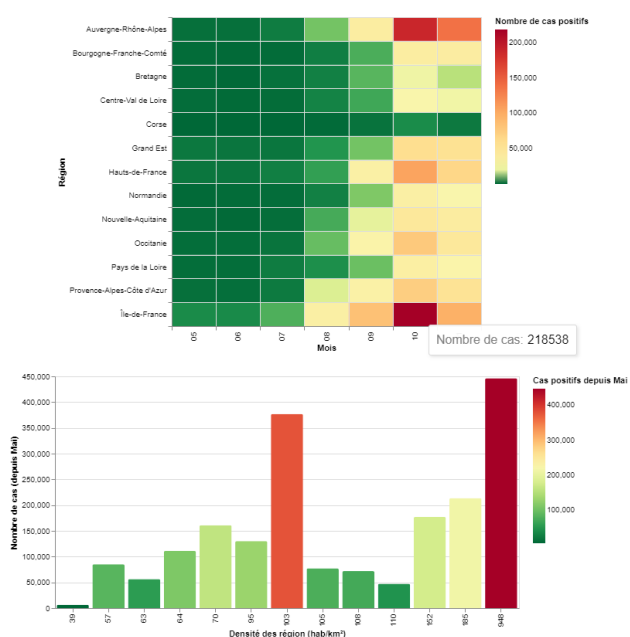


Figure 5 : Visualisations complémentaires réalisées à l'aide de Vega-Lite

Le premier représentant une heatmap liant le nombre de cas positifs mensuels par région (ou département en fonction de la visualisation en cours cf Annexes pour un exemple à l'échelle départementale) et le second est un histogramme qui affiche le nombre de cas positifs depuis mai 2020 (post première vague) par rapport à la densité (là encore par région ou département en fonction du zoom de la visualisation). Les légendes de ces deux visualisations sont identiques pour garder une cohérence au niveau de la visualisation globale et se basant sur le principe basique : vert=positif, rouge=négatif. Au niveau de l'histogramme, on garde

l'utilisation de la couleur pour faire le lien avec le graphique du haut et rendre plus facile la lecture et la compréhension.

III. Point sur les données

La visualisation a été pensée de telle sorte que les données gouvernementales puissent être mise à jour sans nécessité des changements en termes de codage. En effet, pour mettre à jour la visualisation (au niveau du taux d'incidence), il suffit de télécharger les nouvelles données publiées par le gouvernement (<https://www.data.gouv.fr/fr/datasets/taux-dincidence-de-lepidemie-de-covid-19/>) et les ajouter dans le dossier adapté au niveau de la visualisation (/csv/daily/) pour pouvoir observer la suite de la propagation du virus.

Les seules données ayant nécessité une légère tâche de pre-processing sont celles des variables d'agrégations afin de concaténer plusieurs sources pour les obtenir au sein d'un même fichier de base (population, densité, commerces) ainsi que lier les différents départements aux régions afin de pouvoir réaliser le zoom notamment. On pourrait imaginer dans la suite intégrer d'autres variables d'agrégations telles que le nombre de commerces par département/région (c'est implémenté), le nombre d'associations, le nombre d'écoles... Le code est fait de façon à ce que l'ajout de nouveaux éléments soient simples (~15 lignes à ajouter/changer).

IV. Annexes



Figure 6 : Représentation de la Suisse sans échelle particulière

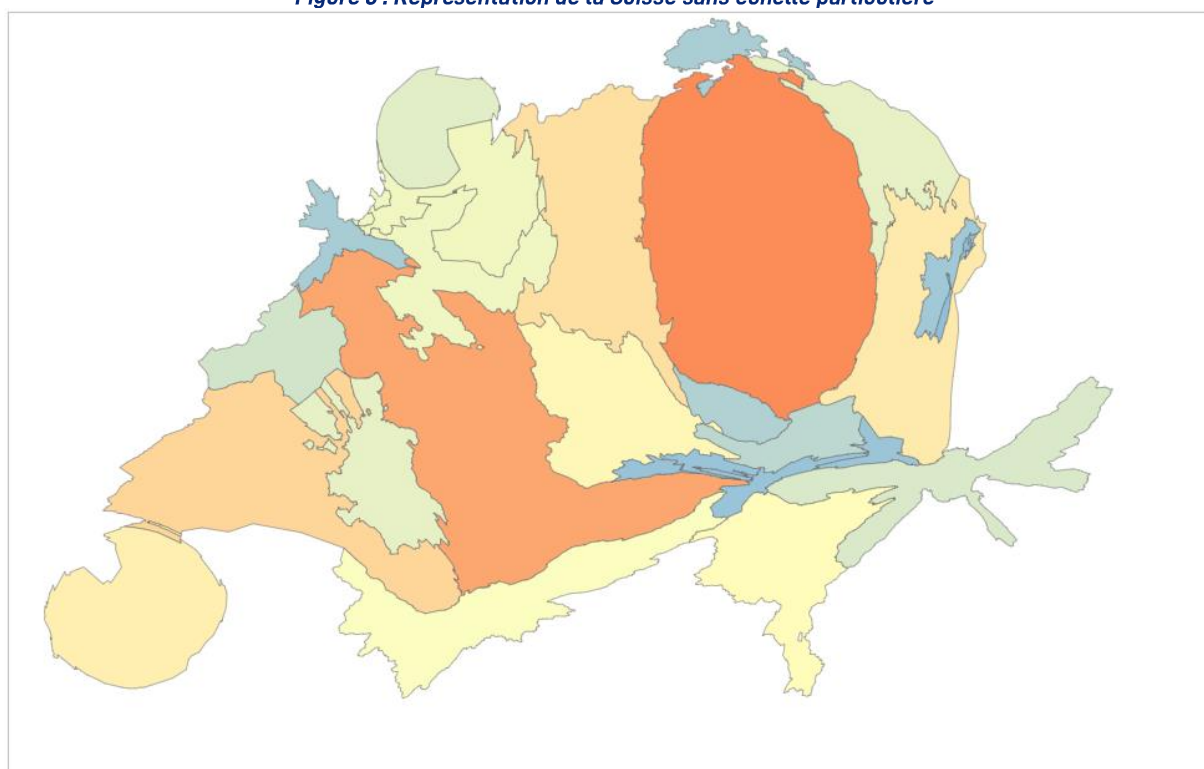
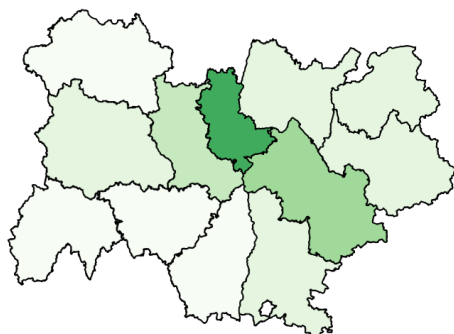


Figure 7 : Représentation de la Suisse sous la forme de cartogramme basée sur la population par canton

Source : https://schuelerzahlen-zuerich.opendata.iwi.unibe.ch/App2/d3geo-2/ch_var/index.html

Echelle : Régionale Affichage : Population



Population

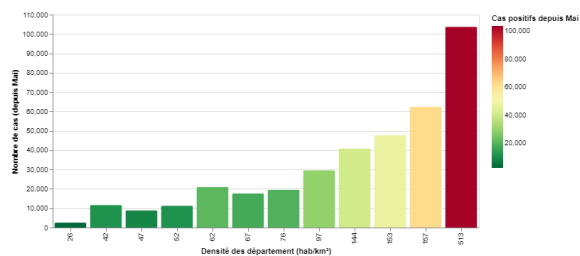
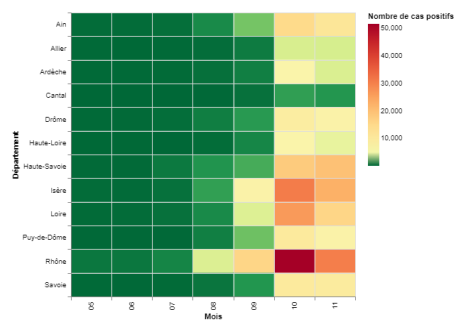
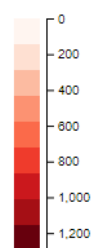
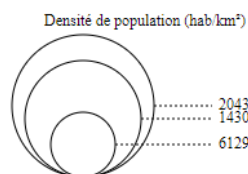
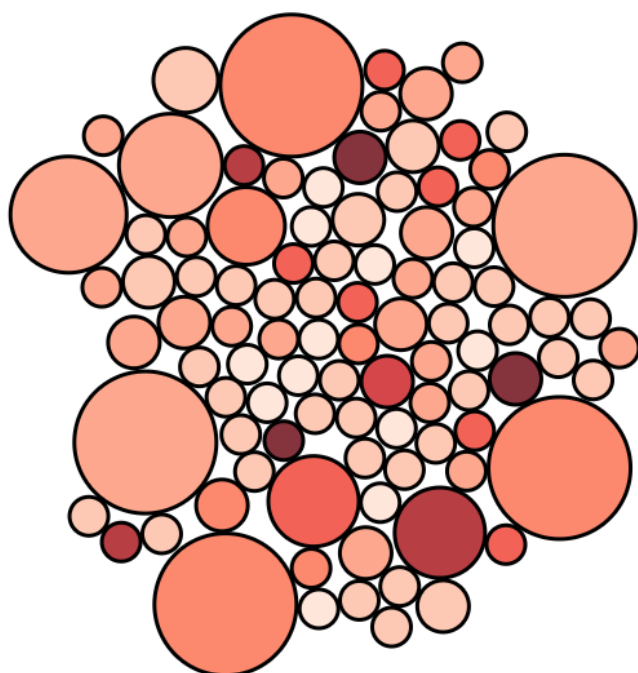


Figure 8 : Exemple de la visualisation complète à l'échelle d'un zoom départemental (région Auvergne Rhône Alpes)

Echelle : Départementale Affichage : Densité/Taux d'incidence
Période : Quotidien (7 derniers jours) Mois : Octobre Jour : 30



Taux d'incidence (pour 100 000 habitants)

Figure 9 : Exemple de visualisation à l'échelle départementale de la corrélation entre Densité & Taux d'incidence