

Rapport de stage de recherche

Amélioration d'une interface de visualisation des cas de COVID-19 pour l'identification de patterns spatio-temporels

Joséphine Boyer - Master Carthagéo

Jacques Gautier & Sidonie Christophe - GEOVIS - LASTIG - IGN

Du 3 au 21 janvier 2022

Table des matières

<i>1. Introduction et mise en contexte.....</i>	<i>3</i>
a. L'équipe GEOVIS et les commanditaires.....	3
b. Le but du projet.....	3
c. Besoins.....	7
<i>2. État de l'art.....</i>	<i>7</i>
a. État de l'art sur les représentations spatio-temporelles de la pandémie.....	7
b. État de la connaissance sur les représentations spatio-temporelles	8
<i>3. Démarche et objectifs</i>	<i>10</i>
a. Présentation des données et de l'interface.....	10
c. Objectifs définis.....	11
<i>4. Réalisation.....</i>	<i>12</i>
a. La définition des cas d'usage	12
b. Premier cas d'usage : l'analyse des vagues épidémiques	12
c. Second cas d'usage : les conséquences de la mise en place de mesures sanitaires pour freiner l'épidémie	14
d. Troisième cas d'usage : analyse exploratoire des données	15
<i>5. Conclusion.....</i>	<i>20</i>
a. Réponse aux besoins	20
b. Limites et pistes d'amélioration	20
c. Bilan personnel	20
<i>Bibliographie.....</i>	<i>21</i>
<i>Table des figures.....</i>	<i>22</i>

1. Introduction et mise en contexte

a. L'équipe GEOVIS et les commanditaires

Ce stage est à l'initiative de Jacques Gautier (enseignant chercheur) et Sidonie Christophe, (directrice de recherche et responsable de l'équipe GEOVIS du LASTIG). Le LASTIG (LABoratoire en Sciences et Technologies de l'Information Géographique pour la ville intelligente et les territoires durables) est un laboratoire de recherche dépendant de l'IGN couvrant l'ensemble du cycle de vie de l'information géographique. Il regroupe quatre équipes, dont l'équipe GEOVIS, qui est spécialisée dans la géovisualisation. Elle permet à des utilisateurs variés de pouvoir interagir et raisonner spatialement et temporellement avec la donnée. Cette donnée géographique est diverse (cartes, imagerie, modèles 3D, modèles numériques de terrain, données externes spatialisées etc.) et parfois imprécise. L'équipe travaille donc également sur des problématiques d'information graphique, cartographie, interaction Homme-machine, visualisation d'information et traitement d'images. J'ai intégré cette équipe durant 3 semaines pour réaliser mon stage de recherche.

b. Le but du projet

Ce stage vise à améliorer une solution déjà développée par le laboratoire. Elle a pour objectif de visualiser et d'identifier les dynamiques spatio-temporelles de la pandémie de COVID-19, à travers une cartographie interactive qui représente par des figurés ponctuels une information spatiale mais aussi temporelle.

Cette solution est innovante dans la visualisation de données liée à la pandémie, car elle cherche à représenter les composantes spatiales et temporelles de ces données dans la même représentation graphique. Ce type de représentation permet de faire le lien entre ces composantes plus facilement et d'identifier les structures spatio-temporelles. En effet, si l'on observe les solutions mises en place depuis le début de la crise sanitaire pour visualiser ce type de données, beaucoup sont sous la forme de tableaux de bord contenant une carte liée à un graphique temporel. C'est l'exemple du site de l'université Johns Hopkins (<https://coronavirus.jhu.edu/map.html>), qui affiche la situation en temps réel via une carte du monde et des chiffres actualisés (Figure 1). Nous pouvons également citer les tableaux et graphiques de Germain Forestier (<https://germain-forestier.info/covid/>) qui montrent l'évolution nationale ou par région, sans carte (Figure 2 et Figure 3). D'autres proposent de visualiser une carte à plusieurs moments donnés via la sélection d'une date, ou une animation qui affiche les cartes par ordre chronologique. Toutes ces solutions ont des niveaux de complexité et d'interaction bien différents.

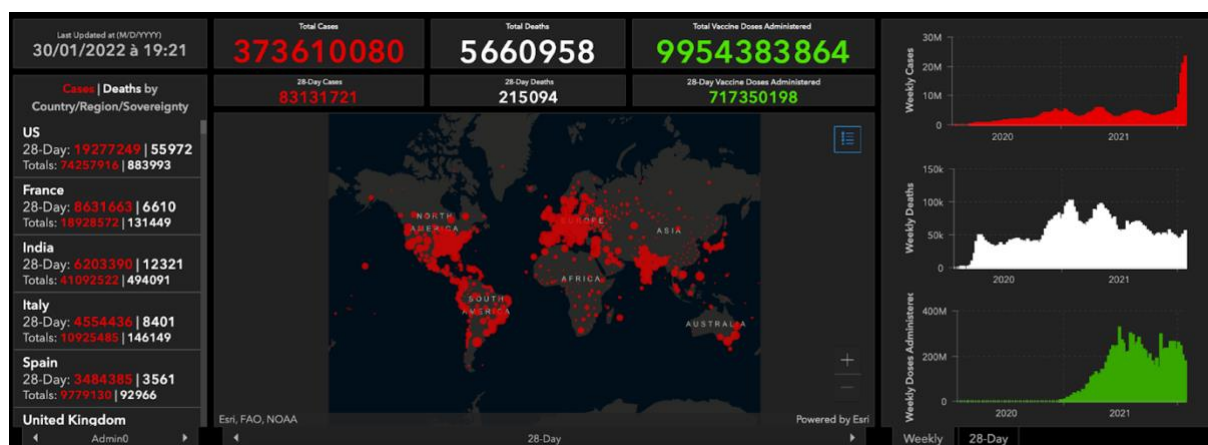


Figure 1 - Extrait du site de l'université Johns Hopkins pour la visualisation de la situation sanitaire dans le monde.

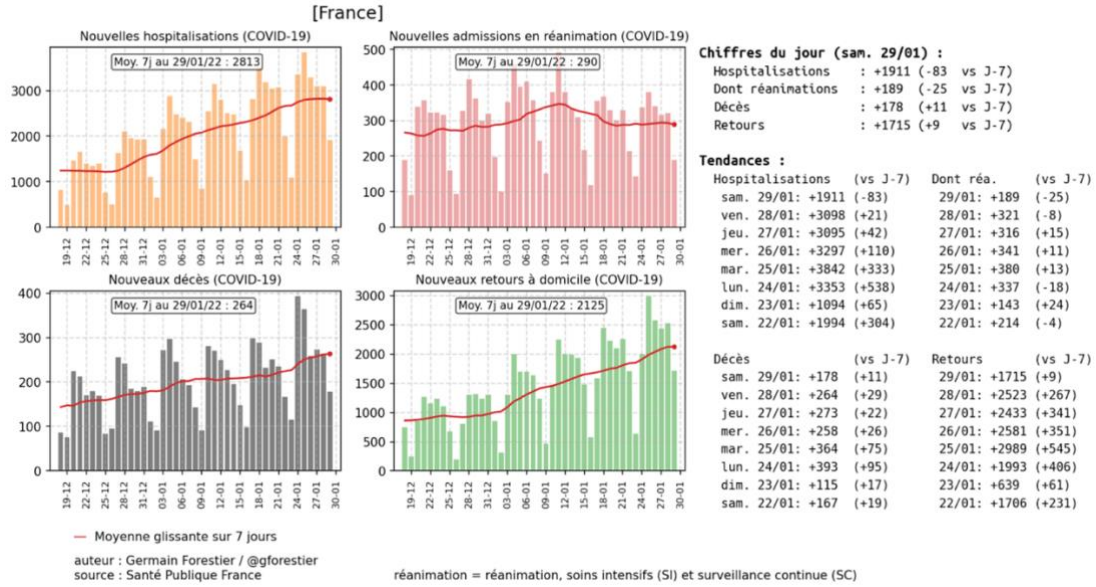


Figure 2 - Graphiques temporels de l'évolution sanitaire en France, d'après Germain Forestier.

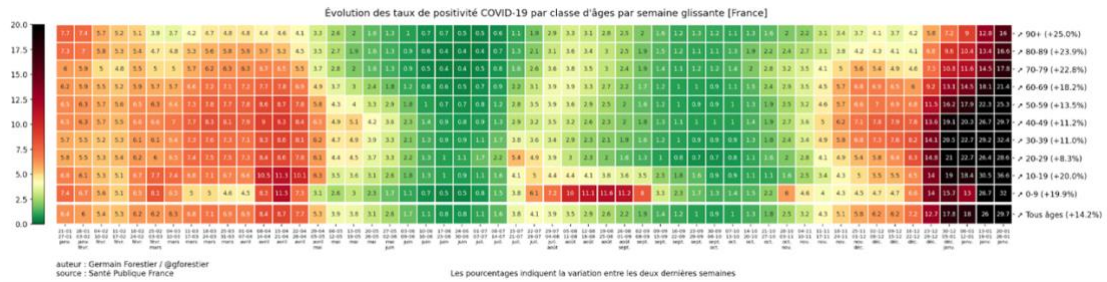


Figure 3 - Tableau illustrant l'évolution des taux de positivité par classe d'âge et par semaine, d'après Germain Forestier.

Pour avoir un aperçu de l'évolution de la donnée sans avoir recours à une animation, certaines solutions mettent en scène une carte 2D qui montre l'évolution à travers des procédés graphiques innovants. Ces procédés permettent de visualiser une troisième dimension temporelle sur des ponctuels (dont l'emplacement donne l'information spatiale). Une autre manière de faire est sous la forme d'une représentation 2D + 1, qui utilise la 3^{ème} dimension pour représenter l'évolution temporelle : c'est le principe du cube spatio-temporel visible sur la Figure 4.

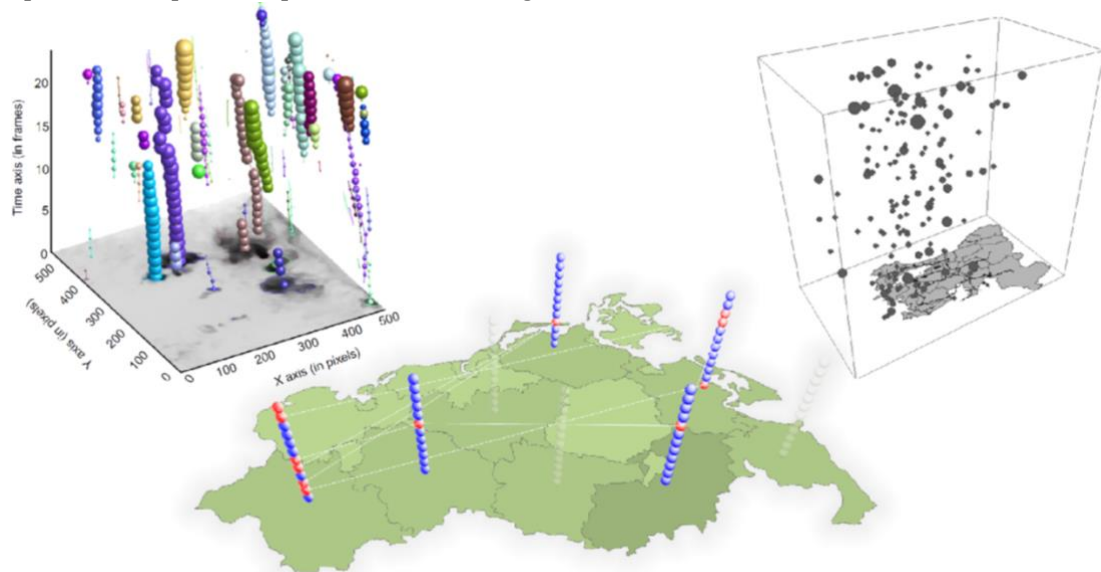
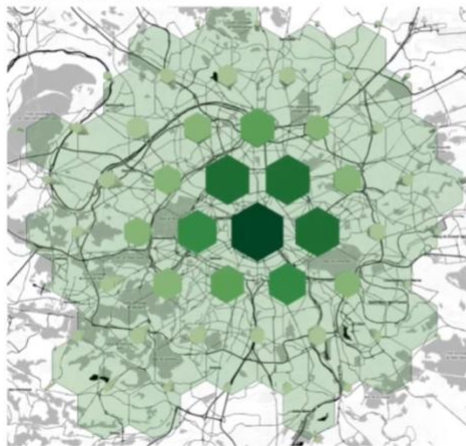


Figure 4 - Aperçu d'un cube spatio-temporel, d'après Turdukulov et al, the LandVis system, Gatalsky, P. et al (dans l'ordre).

Constatant tout cela, l'équipe GEOVIS s'est posé la question suivante : comment pourrions-nous identifier visuellement les structures spatio-temporelles de l'évolution de la pandémie ? En particulier, il serait intéressant de pouvoir discerner les phases successives de la pandémie (apparition de clusters successifs), ses axes de propagation, et ses tendances locales.

Pour ce faire, deux prototypes ont été développés par María Jesús Lobo (chercheuse dans l'équipe GEOVIS) et Jacques Gautier (Gautier, et al., 2021), chacun mettant en place une représentation différente de la composante temporelle des données. Le premier suit le principe du cube spatio-temporel, la troisième dimension est utilisée pour montrer le moment d'apparition des cas. Les points montrant les cas sont agrégés spatialement selon une grille hexagonale régulière, et temporellement selon le jour d'apparition du cas, afin de faciliter la visualisation. On obtient ainsi des blocs qui correspondent à un espace et une période temporelle bien précis, et dont l'aire est proportionnelle au nombre de cas à cet endroit et à cette date. La Figure 5 montre à quoi ressemble le prototype, qui se montre intéressant pour observer l'évolution d'une maille hexagonale en détail, en ayant la possibilité de l'isoler visuellement.

Vue du haut



Vue horizontale

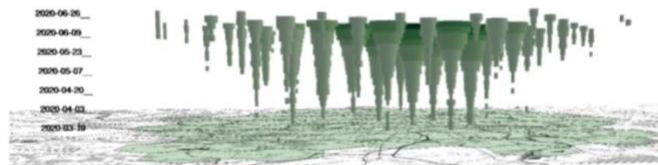


Figure 5 - Visualisation du prototype développé par María Jesús Lobo, et présenté durant la 30ème International Cartographic Conference (Gautier, et al., 2021).

Le second prototype s'appuie sur le concept des Growth Ring Maps (Gautier, et al., 2021), c'est celui auquel se rattache le sujet de ce stage. On a cette fois une représentation en deux dimensions. L'idée est que les cas spatialement proches sont agrégés, et la représentation de l'agrégat obtenu dépend du nombre de points qui le composent et de leur dimension temporelle. On commence par choisir une échelle de couleurs qui permet de discrétiser le temps en périodes (régulières ou non), chaque période ayant une

couleur différente (teinte et/ou valeur uniques). L'application propose plusieurs échelles de couleurs séquentielles et divergentes (Figure 6).

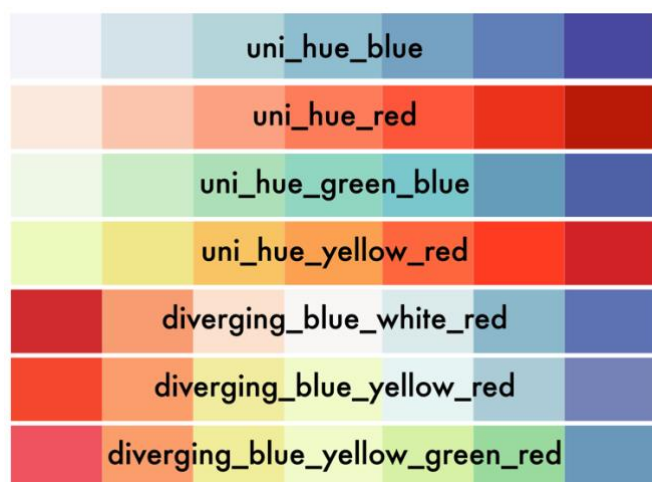


Figure 6 - Palettes proposées dans l'application et leur nom.

Avec l'appui de cette échelle, on attribue une couleur à chaque point selon sa date (ici, on prend la date du début de contamination). L'agrégat est ensuite construit suivant l'ordre de l'échelle de couleur. Pour chaque classe de l'échelle, on compte le nombre de point qui la compose, et on crée un anneau (ou un cercle) dont la taille est proportionnelle au nombre de points de la classe. Chaque anneau (ou le cercle central) a la même couleur que la classe de l'échelle à laquelle il correspond. Une explication de la construction des agrégats en image est proposée en Figure 7.

Avec le visuel obtenu, les clusters infectés par le virus sont identifiables dans les agrégats qui comportent un large anneau d'une seule couleur.

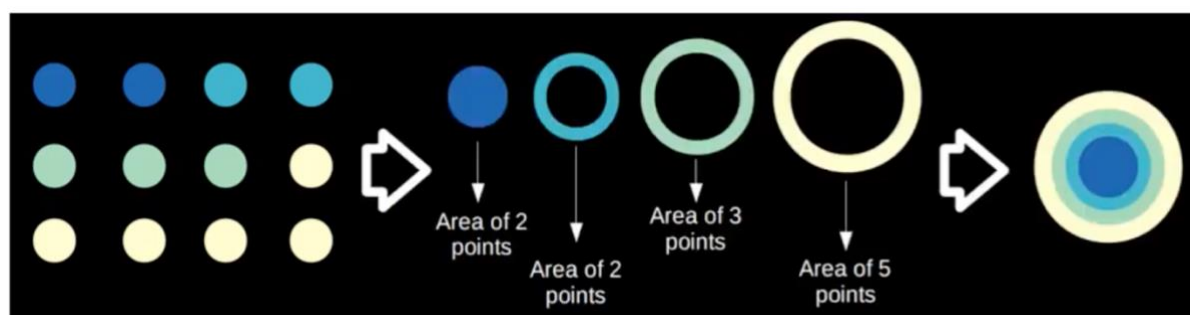


Figure 7 - Construction d'un glyphe selon le principe des Growth Ring Maps (d'après Jacques Gautier durant la 30ème International Cartographic Conference) (Gautier, et al., 2021).

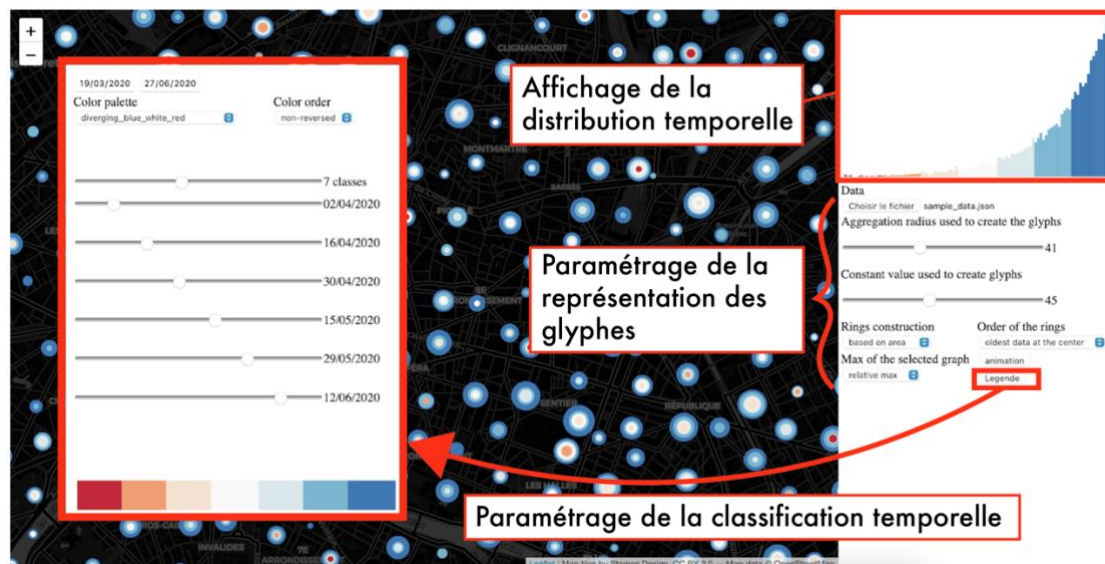


Figure 8 - Application de visualisation en Growth Rings Map pour la visualisation cas de COVID-19 pour l'identification de patterns spatio-temporels développée par Jacques Gautier (Gautier, et al., 2021).

Ainsi, le stage ne consiste pas en la création, mais en l'adaptation d'un outil déjà développé par les commanditaires (dont une capture d'écran est illustrée sur la **Error! Reference source not found.**). La démarche est donc d'ajouter une ou plusieurs fonctionnalités à la suite d'un travail de recherche, et à des essais sur l'outil. Étant donné la durée réduite du stage, il fallait choisir un axe de travail et s'y tenir pour l'implémenter. Ce travail devra permettre d'améliorer la visualisation de l'évolution du phénomène bien particulier qu'est la propagation du SARS-CoV-2.

c. Besoins

Les besoins définis dans le cadre de ce stage sont liés aux utilisateurs finaux comme le voudrait tout travail de développement, mais ils dépendent aussi fortement de l'état de développement de l'application et des fonctionnalités à ajouter ou améliorer. Ainsi, avec l'aide des commanditaires, nous avons défini un axe de travail qui guide toutes les recherches et le développement à effectuer durant ce stage. Cet axe de travail vise à réviser le mode de représentation des données pour offrir à l'utilisateur des axes d'analyse prédéfinis. Pour cela, il a fallu répertorier des cas d'usage, et les mettre en place dans l'application, via l'ajout de fonctionnalités ou la modification de l'affichage des données.

2. État de l'art

Avant d'entreprendre le développement de l'application, il a été nécessaire de lire des articles divers afin de s'inspirer et de trouver une solution peu répandue qui pourrait répondre à nos besoins.

a. État de l'art sur les représentations spatio-temporelles de la pandémie

La propagation du SARS-CoV-2 a mis la cartographie en avant, cette dernière est incontournable pour mieux comprendre l'évolution d'un virus qui vient d'apparaître (Lan, Desjardins, Hohl, & Delmelle, 2021). Cette évolution donne donc lieu à de nombreuses représentations d'une part pour renseigner le grand public et d'autre part à des fins de recherche, afin de prévoir l'évolution de la situation sanitaire et d'aider les décideurs aux moments critiques. Prenons justement l'exemple des visualisations réalisées par le gouvernement français, qui propose un suivi de la pandémie sur son site internet (<https://www.gouvernement.fr/info-coronavirus/carte-et-donnees>) à travers des graphiques temporels et des cartes choroplètes, la Figure 9 en donne un aperçu. Nous y retrouvons une carte et un graphique pour

chaque indicateur de la situation sanitaire (nombre de nouveaux cas, taux d'incidence, occupations des hôpitaux etc.). Ces visualisations ne permettent pas de voir l'évolution spatialement, mais seulement à travers des graphiques qui prennent en compte le pays entier ou certaines régions une à une. L'information spatiale n'a pas d'aspect temporel, car chaque carte correspond à une date fixe. Ils ont l'avantage d'être mis à jour très régulièrement et d'être facilement compréhensibles par un grand public.



Figure 9 - Extrait d'une page du site du gouvernement consacrée à la visualisation de la situation sanitaire.

En plus de cela, et comme mentionné dans l'introduction, un nombre important de tableaux de bords a vu le jour car ils permettent de visualiser la situation en temps réel et facilitent le partage d'informations. Un tableau de bord est, en data visualisation, l'affichage de plusieurs graphiques interactifs basés sur un même jeu de données (Lan, Desjardins, Hohl, & Delmelle, 2021). La combinaison de plusieurs modes de représentation (graphes linéaires, cartes, nuages de point etc.) combinée avec l'interactivité permet d'avoir une bonne analyse des données représentées. Cela facilite le processus de décision dans de nombreux domaines tels que l'éducation, la santé, le commerce ou la gestion de catastrophes naturelles. Malgré ces avantages, la visualisation sous forme de tableau de bord laisse souvent de côté la dimension temporelle dans les cartes, empêchant de bien analyser les variations spatio-temporelles des contaminations, qui sont difficiles à analyser. Aussi, il arrive que la réalisation des tableaux de bords ne soit pas appuyée par un cartographe expérimenté, donnant lieu à des cartes qui déforment la réalité à cause de leurs choix cartographiques (Juergens & Carsten, 2020).

b. État de la connaissance sur les représentations spatio-temporelles

Les procédés de représentation spatio-temporelles pour le cas particulier d'une pandémie ont été définis dans la partie précédente. De ce fait, cette partie s'intéresse aux procédés cartographiques qui visent à représenter le temps sur une carte en deux dimensions de manière générale.

L'idée première est d'utiliser les variables visuelles (Bertin, 1973), car la mise en œuvre de ce type de représentation ne demande pas de compétences techniques poussées. Le travail de (Saint-Marc, et al.) propose d'utiliser la couleur de deux manières en utilisant la valeur et la teinte (la Figure 10 en donne un aperçu). La valeur représentant des variables quantitatives, elle peut représenter le temps de façon linéaire. Quant à la teinte, elle s'applique habituellement aux variables qualitatives, et peut permettre de discerner des périodes bien définies et discernables (comme un changement d'année, de mois, de saison). Cette utilisation de la couleur doit toutefois être maniée avec prudence, car son rôle premier n'est pas de quantifier le temps, mais de faire ressortir une relation d'ordre ou d'association entre séquences temporelle, il faut donc veiller à ce que le lecteur ait une bonne idée de ce qu'il est en train d'observer. Cette représentation visuelle du temps est celle qui est mise en place dans l'outil amélioré pendant ce stage.



Figure 10 - Adaptation de l'échelle de Peter Bak pour sa visualisation de Growth Ring Map.

En outre, il existe d'autres représentations du temps sur une carte. Par exemple, en indiquant les lieux du phénomène liés des symboles ponctuels (Huang, et al., 2008). Ces symboles prennent la forme d'un diagramme temporel, et sont placés autour de la carte en forme d'anneau (donnant le nom de *Ring Map*). Un exemple d'une visualisation en *ring map* est représenté Figure 11. Cette représentation a l'avantage de simplifier la représentation, car la carte n'est pas surchargée par une donnée temporelle, mais le lien à effectuer entre chaque diagramme et l'élément correspondant sur la carte est difficile à visualiser. De plus, l'intégration de diagrammes directement dans la carte peut poser des difficultés de lecture.

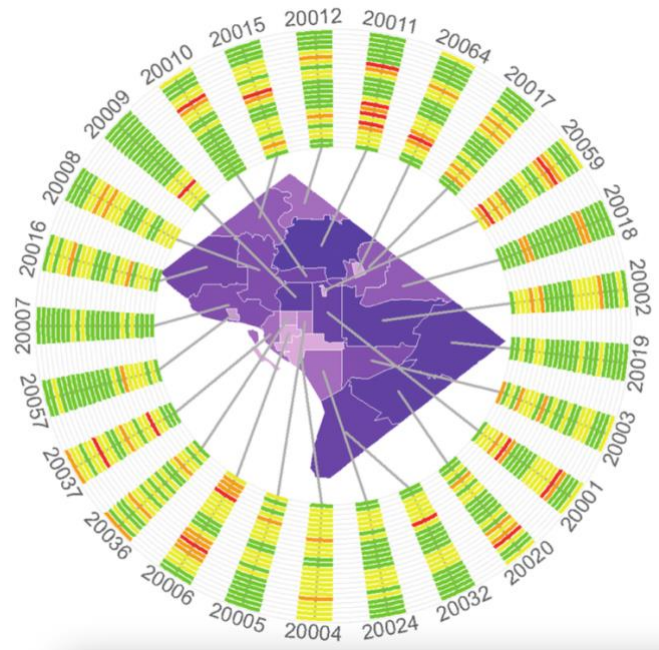


Figure 11 - Ring map affichant le statut de l'alerte sanitaire pour 24 semaines pour chaque code postal.

La solution optimale pour visualiser le temps est l'animation de cartes, les cartes sans dimension temporelles se basent sur "l'éternel présent" et négligent l'importance de la progression d'un phénomène, au contraire des animations (Harrower, Fabrikant, & Dodge, 2008). L'animation de carte permet de bien visualiser les modèles de diffusion et les clusters épidémiques. Mais comme toute forme de visualisation, l'animation de cartes a ses limites car ce procédé laisse peu de temps au lecteur pour analyser chaque carte et mémoriser l'information. L'animation doit donc être compréhensible et équivoque.

Enfin, une dernière approche approfondie les modalités et normes de représentations cartographiques pour proposer une visualisation du temps sur une carte, c'est celle des Growth Ring Maps. Cette méthode de représentation a été introduite en 2009 pour l'étude du comportement de souris (Bak, Mansmann, Janetzko, & Keim, 2009; Brewer, 1994). Elle a depuis été étudiée dans le cas de l'étude simultanée de la localisation, de la date et de la quantité des photos prises par des touristes (Andrienko, et al., 2011) (Figure 12).

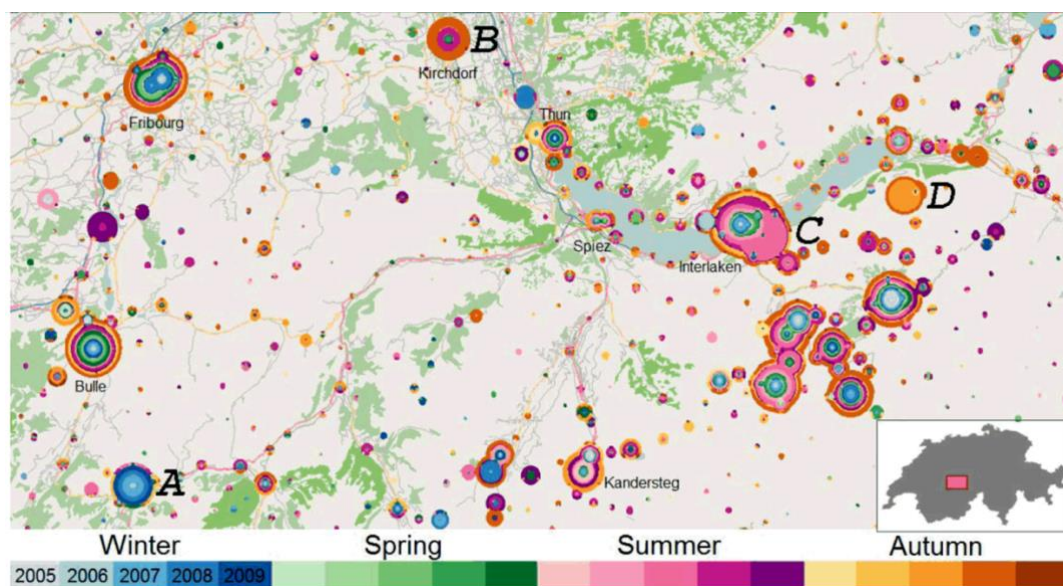


Figure 12 - Exemple de représentation en Growth Ring Map, d'après Peter Bak et al.

3. Démarche et objectifs

a. Présentation des données et de l'interface

Le prototype développé par Jacques Gautier et María Jesús Lobo (Gautier, et al., 2021) est développé en Javascript et s'appuie sur la bibliothèque graphique D3.js. Cette bibliothèque s'appuie sur les technologies Javascript, SVG et CSS pour proposer la lecture et la visualisation graphique de données numériques. Dans notre cas, D3.js permet d'afficher des glyphs sur la carte, et des histogrammes montrant la distribution temporelle des données.

L'interface comprend une carte interactive sous la forme d'un objet Leaflet, dont les propriétés ont été définies antérieurement. Leaflet est une bibliothèque de cartographie qui permet de créer des cartes interactives (zoom, déplacement, placement d'objets à des coordonnées données notamment) tout en accédant à des fonds de carte de diverses sources (WMS, GeoJSON, vectoriels, etc.).

La Figure 13 montre l'application dans l'état où elle était au début de ce stage. On y trouve la carte qui prend la majeure partie de l'écran, ainsi que l'interface de paramétrage de l'affichage sur la gauche. La Figure 8 complète cette description en montrant plus en détail les fonctionnalités de l'application.

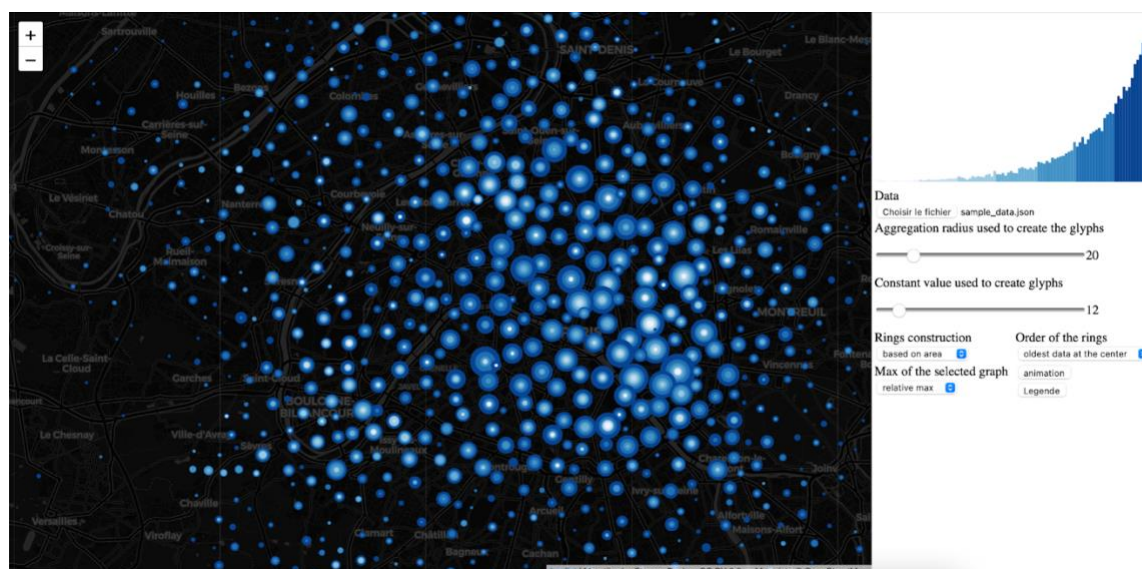


Figure 13 - Capture d'écran de l'application au début du stage.

```
{
  "id": 1,
  "date": "2020/03/19",
  "lon": 2.39466625447853,
  "lat": 48.87439312961423,
  "X": 655600,
  "Y": 6863989,
  "incubation": "2020/03/22",
  "fin_infect": "2020/03/30"
},
{
  "id": 33,
  "date": "2020/03/19",
  "lon": 2.273684197869761,
  "lat": 48.84951855429459,
  "X": 646701,
  "Y": 6861298,
  "incubation": "2020/03/22",
  "fin_infect": "2020/03/30"
},
```

Concrètement, une première étape sera d'identifier clairement ces cas d'usage, et de les mettre en place dans un second temps. La partie suivante décrit l'implémentation de chaque cas d'usage et les choix généraux qui ont été établis au fur et à mesure de la réflexion.

4. Réalisation

a. La définition des cas d'usage

Avant de commencer l'implémentation, il convenait de se renseigner sur les cas d'usages, liés à l'utilisation qui sera faite l'outil. Un cas d'usage correspond à une manière d'utiliser l'application. Pour cela, nous devons nous intéresser aux besoins des utilisateurs, et deviner quels pourraient être leur demande s'ils voulaient une analyse particulière. Ce travail a été réalisé durant la première journée dédiée au développement de l'application. Les cas d'usages retenus après concertation avec les commanditaires sont les suivants :

- Dans le cas de données couvrant une vague épidémique, il faudrait proposer une classification temporelle qui distingue les différentes phases de la vague. En outre, si les données couvrent plusieurs vagues épidémiques, il serait intéressant de pouvoir comparer la distribution spatio-temporelle de plusieurs vagues pour repérer des patterns liés aux structures de distributions différentes et mieux prévoir l'évolution des prochaines.
- Si les données intersectent une période de confinement ou d'application stricte des gestes barrières, il conviendrait de voir si cette mise en application de mesures sanitaires a eu une conséquence sur la distribution spatio-temporelle des cas positifs.
- Il a également été suggéré de pouvoir permettre une analyse exploratoire de la distribution spatio-temporelle des cas.

b. Premier cas d'usage : l'analyse des vagues épidémiques

L'implémentation de ce cas d'usage doit se réaliser en étant éclairée sur la définition d'une vague épidémique, qui n'est malheureusement pas formelle. L'office québécois de la langue française (2020) est l'un des seuls organismes à en proposer une définition : "Brusque intensification d'une épidémie pendant laquelle le nombre de cas d'une affection contagieuse évolue rapidement vers un pic épidémique avant de redescendre tranquillement, et qui se traduit par une courbe épidémique en forme d'une vague.". Donc, le seul repère temporel fixe d'une vague serait le pic épidémique. À partir de cette constatation, plusieurs modélisations ont été envisagées, séparant la vague en 2 ou 3 classes : les différents scénarios sont représentés sur la Figure 15. Chaque limite de classe doit correspondre à une date du jeu de données, et ce sont ces limites que nous cherchons à déterminer.

Nombre de classes dans une vague	Début de la vague (calcul de l'augmentation d'un jour à l'autre)	Délimitation du centre de la vague	Fin de la vague
2	Augmentation (relative ou absolue ?) supérieure à un seuil.	Jour recensant le nombre maximal de cas positifs.	Retour en dessous du nombre de cas marquant le début de la vague.
3		Période durant laquelle le nombre de cas vaut au moins 85% du maximal.	

Figure 15 - Choix possibles pour classer une vague épidémique.

Concernant l'établissement de la date de début de la vague, il a été décidé de la définir selon l'augmentation quotidienne du nombre de cas. Il faut ainsi calculer l'augmentation d'un jour à l'autre chronologiquement pour chaque date, et dès qu'elle est supérieure à un seuil, on sélectionne la date qui correspond à cette augmentation limite pour symboliser le début de la vague. Nous n'avons pas considéré d'écart absolu entre deux journées consécutives, car le nombre de cas peut augmenter fortement sans que l'augmentation relative soit forte, dans le cas où un grand nombre de personnes sont contaminées. On peut ici prendre l'exemple de la cinquième vague, qui voit énormément de nouveaux cas comparé aux précédentes. Ainsi,

il a été décidé de calculer l'écart relatif entre les nombres de cas entre deux jours consécutifs. Le seuil paramétré dans l'application correspond à une augmentation de 100%, i.e. le nombre de contaminations double d'un jour à l'autre. Ce paramétrage a des limites, car il ne correspond à aucune définition formelle, et n'a été testé que sur un jeu de données. On ne sait pas si ce paramétrage permettra de distinguer les phases d'une vague légèrement différente, mais il est aisément modifiable dans le code.

La fin de la vague correspond au jour à partir duquel le nombre de cas tombe en dessous de celui qui annonçait le début de la vague. Cela correspond à un retour à la normale.

Pour la détermination du pic, deux options s'offraient :

- Soit on détermine une date qui correspond au pic (nombre de cas maximal, ou jour médian de la durée totale de la vague), le pic est alors une limite de classe, et on distingue deux parties : la période où les cas augmentent globalement, et celle lors de laquelle les cas diminuent globalement.
- Soit on divise la vague en trois parties, en prenant en compte une partie centrale qui symboliserait le pic avec une période de temps, et non une date unique assimilée à une limite de classe. Cette classe centrale correspondrait à la période lors de laquelle le nombre de cas est au moins égal à 85% du nombre de cas maximal.

Nous avons choisi de garder la classification en trois parties de la vague (Figure 16), car la définition du pic est assez floue, on ne peut pas définir le pic par un seul jour, car il y a une légère stagnation des cas. Il est préférable de désigner une période lors de laquelle le nombre de cas est extrêmement haut. Les deux autres périodes avant et après correspondent respectivement à une augmentation et à une diminution du nombre de cas.

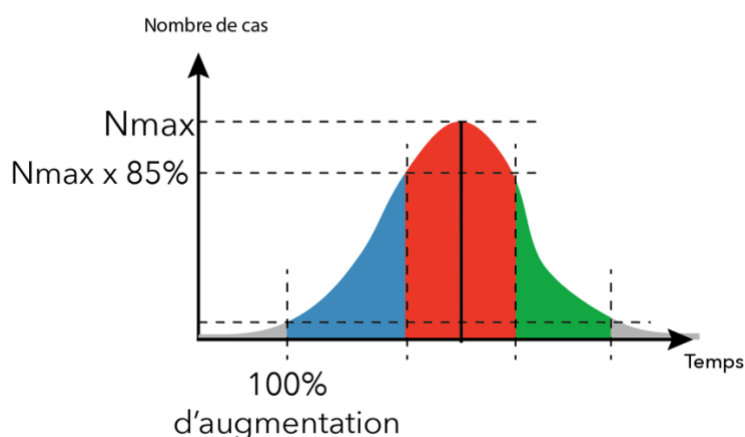


Figure 16 - Classification d'une vague calculée par l'algorithme de l'application.

Le mode de détermination des phases de la vague choisi a été implémenté dans l'application sous la forme d'une échelle de couleurs dont les bornes sont prédéfinies pour l'utilisateur (et que l'application calcule grâce à un algorithme au moment d'afficher cette échelle, en fonction du jeu de données entré en paramètre), visible sur la **Error! Reference source not found.** L'utilisateur peut donc analyser la distribution spatio-temporelle de chaque phase, sans pouvoir gérer la définition des classes.

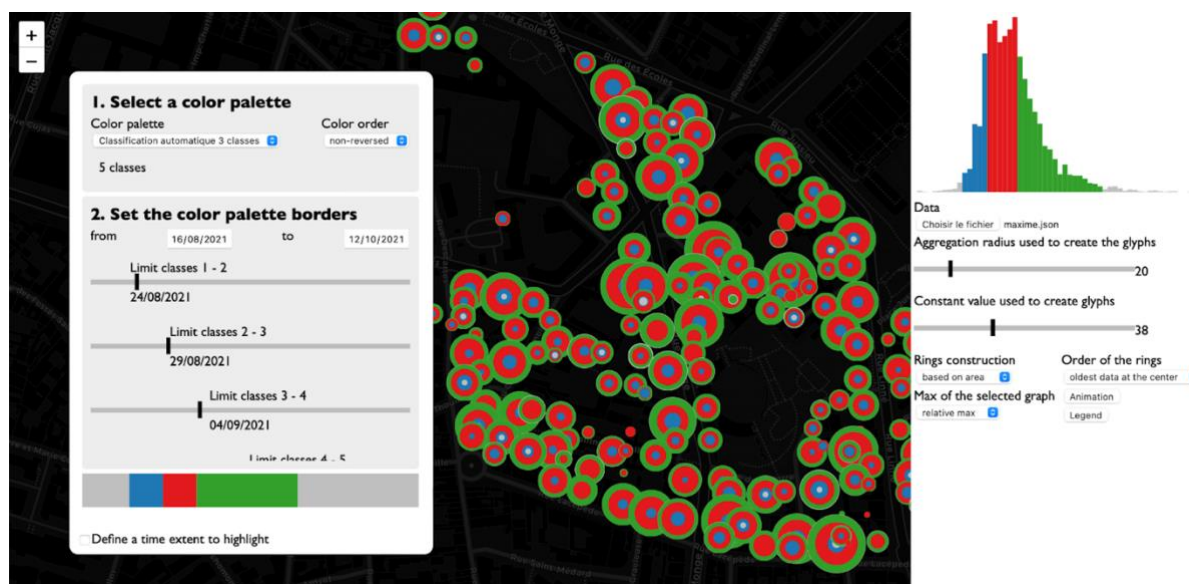


Figure 17 - Visualisation de la classification automatique en 3 classes d'une vague épidémique sur l'application.

Concernant les couleurs attribuées aux différentes périodes temporelles, il a été décidé après réflexion, essais sur l'application, et discussion avec les commanditaires de représenter les périodes temporelles hors de la vague en gris, pour diminuer leur intensité visuellement et ainsi mettre en valeur les classes qui font partie de la vague. Les différentes phases de la vague ont une teinte différente chacune, cela permet de distinguer les trois phases plutôt que de voir l'ordre temporel des trois phases.

c. Second cas d'usage : les conséquences de la mise en place de mesures sanitaires pour freiner l'épidémie

Ce cas d'usage a été paramétré très sommairement, car je n'avais pas de jeu de données adéquat qui montre un tel phénomène. Il a été mis en place dans l'application sous la forme d'une échelle prédéfinie qui tient compte d'une date fictive comme début de confinement. Cette date fictive est fixée au 29/08/2021 pour l'implémentation de la fonctionnalité.

La réflexion autour de ce cas d'usage s'est surtout axée autour de la création d'une classification temporelle et du choix des couleurs à attribuer aux classes définies.

La classification qui nous a paru la plus vraisemblable pour mieux rendre compte de l'évolution du nombre de cas est la division en périodes régulières contenant un nombre de jours constant (des semaines dans le cas de notre jeu de données, qui couvre un mois et demi). Pour signifier la mise en pratique du confinement (ou d'un couvre-feu par exemple), la date d'entrée en vigueur des restrictions correspond à une limite de classe. Les autres classes sont définies de façon que cette date soit une limite de classe, et ont toutes une étendue de 7 jours. Ainsi nous obtenons 4 classes : une du début des données à J-7, une de J-7 à J, une de J à J+7, et une de J+7 à la fin des données. Cette étendue de 7 jours reflète la manière dont l'épidémie est suivie, car les données ne sont pas récoltées le week-end, et il est préférable de faire une moyenne par semaine lorsqu'on calcule un indicateur de l'évolution de l'épidémie. Ce phénomène est visible sur la Figure 2, alors que l'évolution jour par jour des histogrammes est rythmée par des creux le week-end, la moyenne par semaine des courbes rouges est plus lissée et montre l'évolution plus fidèlement. De plus, comme décrit ci-dessus, l'étendue de notre jeu de données est propice à une classification par semaine et non par mois pour avoir un nombre de classes suffisant pour l'analyse des données.

Concernant le choix de la teinte et de la valeur de la couleur de chaque classe, voici ce qui a été retenu :

- Il y a deux teintes distinctes, séparées par la date d'entrée en vigueur des restrictions,
- La valeur de chaque classe augmente avec le temps.

Ces choix reprennent les considérations faites précédemment, i.e. un changement de valeur montre une relation d'ordre entre les périodes temporelles (et permet de les classer par ancienneté), et un changement de teinte révèle un changement de nature de la représentation (on distingue deux périodes lors desquelles

la lutte contre la pandémie est abordée différemment). La Figure 18 montre ce à quoi ressemble ce paramétrage dans l'application.

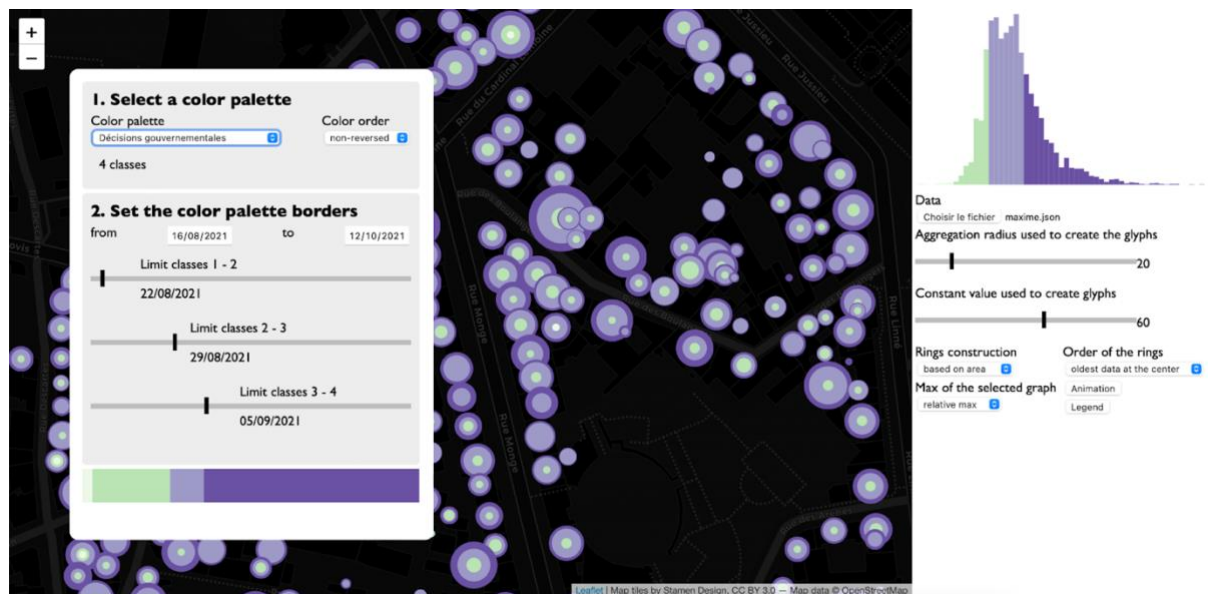


Figure 18 - Visualisation des conséquences d'un confinement sur l'évolution de l'épidémie (les données sont fictives).

d. Troisième cas d'usage : analyse exploratoire des données

Ce cas d'usage propose une nouvelle manière d'aborder la classification temporelle. Pour cela, une réflexion sur l'échelle de couleur a été menée. En effet, deux variables sont considérées pour la représentation des agrégats : les dates des événements et le nombre d'événements. Elles sont représentées respectivement par l'ordre des anneaux du glyphe et l'aire des mêmes anneaux. La couleur permet de différencier les anneaux, et est initialement prévue pour définir le temps (d'après les échelles implémentées dans l'application). Cependant, les échelles jouent sur une variation de teinte, qui est une variable visuelle représentant l'ordre entre les valeurs. Dès lors, deux solutions s'offrent à nous, représentées en Figure 19 et Figure 20 :

- faire figurer en plus foncé les anneaux ayant le plus de cas (et donc la plus grande aire), pour symboliser la relation d'ordre entre les périodes où les contaminations sont plus ou moins fortes (Figure 20),
- ou représenter le temps de façon ordonnée en augmentant la valeur avec le temps qui passe (Figure 19).

Puisque le temps est représenté de façon ordonnée dans les glyphes, la couleur sera associée à l'échelle temporelle et non au nombre de contaminations. Cela a deux avantages : l'utilisateur sera moins perdu, il

distinguera mieux les variations de valeur si celles-ci sont disposées par ordre croissant dans l'histogramme et dans l'échelle de couleurs.



Figure 19 - La valeur représente la dimension temporelle.

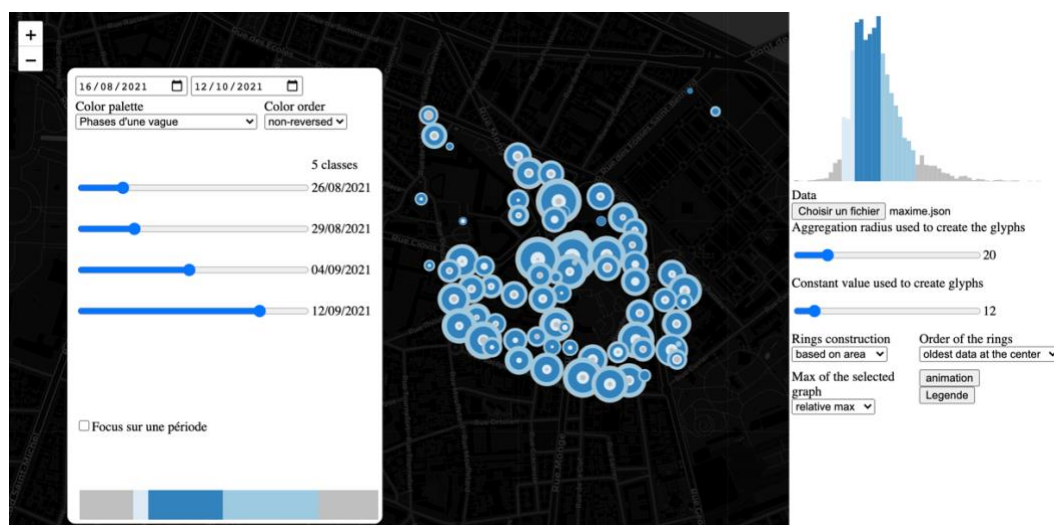


Figure 20 - La valeur représente le nombre de cas.

En outre, cette analyse exploratoire des données cas d'usage doit permettre à l'utilisateur de choisir une période temporelle à mettre en évidence. Cette période temporelle est délimitée par deux dates, qu'il pourra choisir grâce à un *slider*. Ce *slider* est configuré comme indiqué sur la Figure 21, il comporte 3

poignées : date de début, date de fin et poignée au milieu qui permet de faire coulisser la période choisie en gardant la même étendue temporelle.

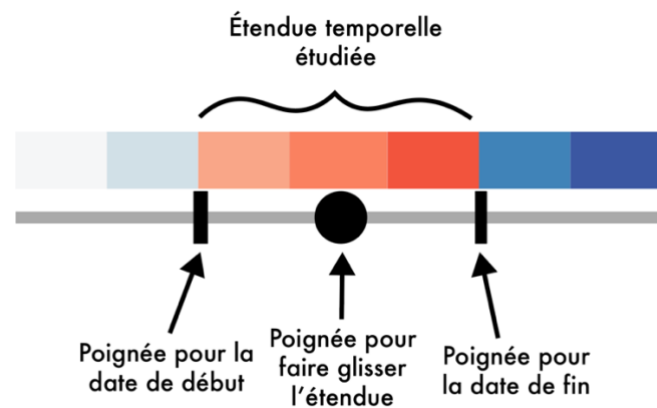


Figure 21 - Description des poignées du slider.

Concernant la symbologie, l'étendue choisie est mise en valeur car un changement de teinte par rapport à celle(s) de l'échelle choisie par l'utilisateur. Le paramétrage de cette étendue temporelle doit donc être proche de l'affichage de cette échelle de couleur, pour mieux indiquer à l'utilisateur qu'il agit sur cette dernière, et non sur la classification temporelle. La Figure 23 montre l'affichage dans l'application du slider de paramétrage en question dans ce paragraphe.

Pour définir les teintes à appliquer à la période de temps mise en avant, nous avons sélectionné des couleurs qui tranchent bien avec la ou les teintes initiales, en conservant le nombre de teintes (une pour les palettes séquentielles et deux pour les palettes divergentes).

Nous avons fait le choix de désactiver cette fonctionnalité pour les classifications et palettes de couleur prédéfinies (telles que celle correspondant au cas d'usage précédant sur les décisions gouvernementales). En effet, ces palettes mènent déjà à une analyse, et il ne serait pas judicieux de proposer une analyse supplémentaire à ce type de classification, car c'est inutile à une meilleure compréhension du phénomène étudié (la palette prédéfinie est là pour ça).

Enfin, puisque ce cas d'usage invite l'utilisateur à manier la classification temporelle, j'ai effectué un travail qui vise à améliorer l'expérience utilisateur afin qu'il utilise tout le potentiel de l'application. Pour cela j'ai notamment réfléchi à la mise en page du panneau de configuration de la légende, visible en son état initial sur la Figure 22.



Figure 22 - Évolution de la mise en page du panneau de paramétrage de la légende avant le stage (à gauche) et à la fin du stage (à droite).

Premièrement, j'ai déplacé les éléments pour guider l'utilisateur lorsqu'il construit sa légende. La construction se réalise en deux étapes, et l'affichage de la carte et de l'histogramme est actualisé à chaque fois qu'il définit un paramètre :

- D'abord, il sélectionne une palette de couleur et la paramètre (nombre de classes et ordre des couleurs).
- Ensuite, il se concentre sur la classification temporelle des données. Il choisit la date de début et de fin de l'affichage de données (qui sont définies par défaut sur les dates limites du jeu de données). Puis il peut changer les bornes entre deux classes selon des dates précises (le paramétrage par défaut correspond à des périodes régulières). L'affichage des dates qui correspondent aux limites de classes suivent le curseur afin de mieux relier cet aspect temporel à la définition d'une borne de classe, et de mieux faire correspondre la position du curseur à la position de la borne sur l'aperçu de l'échelle en bas de la fenêtre.

Enfin, l'emplacement sous la barre d'aperçu de l'échelle est réservé à la fonctionnalité de mise en évidence d'une partie de l'étendue temporelle, car elle est indépendante de la construction de l'échelle (la Figure 23 en donne un aperçu). La barre s'affiche à la demande de l'utilisateur (s'il coche la fonctionnalité), et est automatiquement retirée dans le cas où l'utilisateur choisit une palette qui analyse un phénomène donné n'ayant pas besoin de cette fonctionnalité.

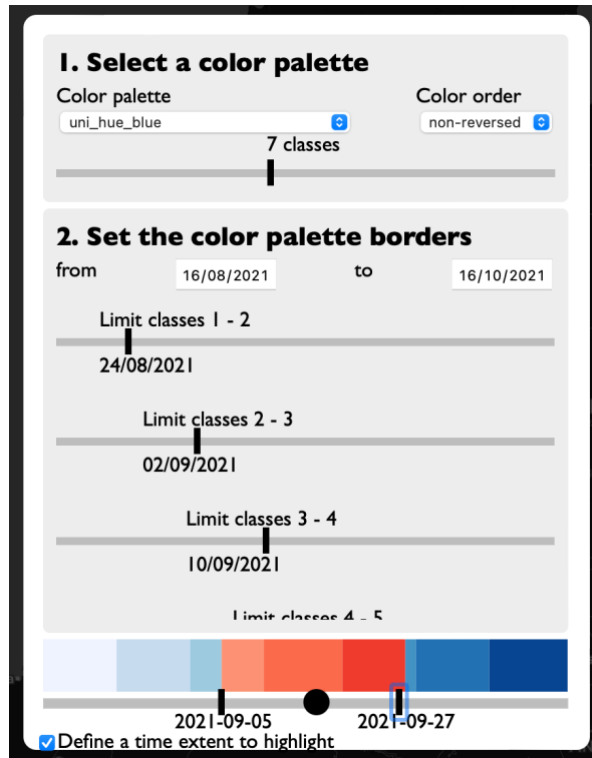


Figure 23 - Panneau de légende avec la mise en évidence d'une étendue temporelle.

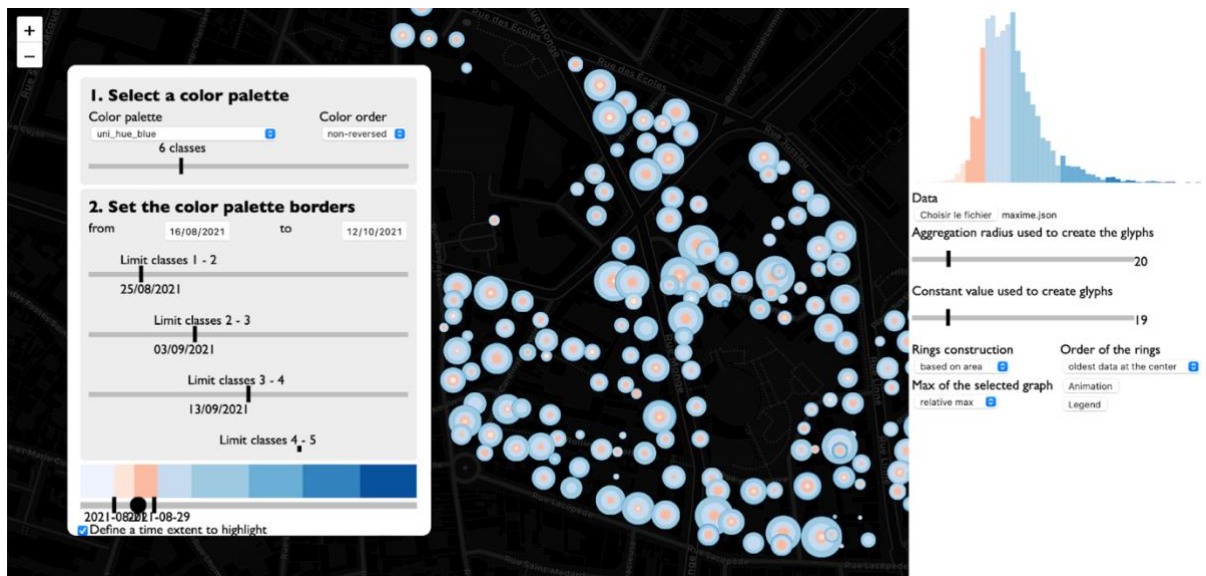


Figure 24 - Exemple d'utilisation de cette fonctionnalité : observation des jours lors desquels le nombre de cas augmentait très fortement.

5. Conclusion

a. Réponse aux besoins

Pour rappel, le principal objectif posé pendant le stage est le suivant : améliorer et guider l'analyse tirée de la visualisation en Growth Ring Map.

L'objectif est globalement atteint, l'application propose trois fonctionnalités supplémentaires qui reflètent trois moyens d'analyse des données :

- l'analyse d'une ou plusieurs vague(s) épidémique(s) grâce à un algorithme de classification automatique
- l'analyse de l'impact de politiques publiques pour lutter contre la hausse des cas
- une analyse exploratoire des données qui guide l'utilisateur dans l'appréhension de ses données.

L'application offre à l'utilisateur de nombreuses possibilités de paramétrer la carte et de représenter les données qu'il entre. En plus de cela, la visualisation de données sous la forme d'une Growth Ring Map n'est pas très répandue, et il faut pour cela "former" l'utilisateur à cette représentation.

b. Limites et pistes d'amélioration

La limite principale concerne les données. Les visualisations ont été réalisées sur un unique jeu de données qui simule une vague épidémique. Le prototype actuel fonctionne sur ce jeu de données, et est développé pour fonctionner sur des données différentes. Malheureusement, il n'a pas été possible de vérifier cette opérabilité, notamment concernant les échelles de couleurs développées pour les cas 1 et 2. De plus, il est impossible à l'heure actuelle de disposer de données événementielles sur la pandémie, pour respecter le secret médical.

Les données correspondant aux cas de figures spécifiques étudiés lors de ce travail n'étaient pas disponibles.

c. Bilan personnel

Il a été très intéressant de pouvoir s'immerger pendant ces quelques semaines dans le monde de la recherche. Malgré la modalité du distanciel, j'ai été très bien accompagnée et conseillée lors de mon travail. J'en reviens avec plus de compétences, notamment sur l'utilisation de la librairie D3.js. J'ai également une meilleure appréhension de la rigueur dont il faut faire preuve lors du développement, car il faut savoir justifier tous les choix réalisés.

De plus, il est toujours intéressant de pouvoir contribuer à un projet existant, et d'avoir l'occasion d'y contribuer même peu longtemps, car cela pousse à produire une solution rapidement, et donc d'organiser le travail en conséquence. J'espère que mon travail sera utile et repris malgré la courte échéance.

Bibliographie

- Andrienko, Gennady, Natalia, Bak, Peter, Keim, . . . Stefan. (2011). *A conceptual framework and taxonomy of techniques for analyzing movement*.
- Bak, Mansmann, Janetko, & Keim. (2009). *Spatiotemporal analysis of sensor logs using growth ring maps*. IEEE transactions on visualization and computer graphics.
- Bertin, J. (1973). *Sémiologie graphique*. Flammarion Paris.
- Brewer, C. A. (1994). *Color use guidelines for mapping. Visualization in modern cartography*.
- Gautier, J., Lobo, M. J., Fau, B., Drugeon, A., Christophe, S., & Touya, G. (2021). *COVID-19 Geoviz For Spatio-temporal Structures Detection*.
- Harrower, M., Fabrikant, S., & Dodge. (2008). *The role of map animation in geographic visualization*.
- Huang, Guilan, Govoni, Sergio, Choi, Jae, . . . James. (2008). *Improves comprehension when mapping many variables with Ring Maps Geovisualizing Data*.
- Juergens, & Carsten. (2020). *Trustworthy COVID-19 Mapping: Geo-spatial Data Literacy Aspects of Choropleth Maps*. KN-Journal of Cartography and Geographic Information.
- Lan, Y., Desjardins, M., Hohl, A., & Delmelle, E. (2021). *Geovisualization of COVID-19: state-of-the-art and opportunities*. Accepted for Cartographica.
- Saint-Marc, C., Davoine, P.-A., Dimuro, A., Staudacher, T., Villanova-Oliver, M., & Boissier, P. (2012). *Constructions cartographiques pour la caractérisation de l'aléa volcanologique du Piton de la Fournaise*. Colloque SAGEO.

Table des figures

Figure 1 - Extrait du site de l'université Johns Hopkins pour la visualisation de la situation sanitaire dans le monde.....	3
Figure 2 - Graphiques temporels de l'évolution sanitaire en France, d'après Germain Forestier.....	4
Figure 3 - Tableau illustrant l'évolution des taux de positivité par classe d'âge et par semaine, d'après Germain Forestier.....	4
Figure 4 - Aperçu d'un cube spatio-temporel, d'après Turdukulov et al, the LandVis system, Gatafsky, P. et al (dans l'ordre).	4
Figure 5 - Visualisation du prototype développé par María Jesús Lobo, et présenté durant la 30ème International Cartographic Conference.....	5
Figure 6 - Construction d'un glyphe selon le principe des Growth Ring Maps (d'après Jacques Gautier durant la 30ème International Cartographic Conference).....	6
Figure 7 - Application de visualisation en Growth Rings Map pour la visualisation cas de COVID-19 pour l'identification de patterns spatio-temporels développée par Jacques Gautier.	5
Figure 8 - Extrait d'une page du site du gouvernement consacrée à la visualisation de la situation sanitaire.	8
Figure 9 - Exemple de représentation en Growth Ring Map, d'après Peter Bak et al.....	10
Figure 10 - Adaptation de l'échelle de Peter Bak pour sa visualisation de Growth Ring Map.	8
Figure 11 - Ring map affichant le statut de l'alerte sanitaire pour 24 semaines pour chaque code postal..	9
Figure 12 - Capture d'écran de l'application au début du stage.....	10
Figure 13 - Aperçu des données lues par l'application.....	11
Figure 14 - Visualisation de la classification automatique en 3 classes d'une vague épidémique sur l'application.....	12
Figure 15 - Visualisation des conséquences d'un confinement sur l'évolution de l'épidémie (les données sont fictives).	15
Figure 16 - Exemple d'utilisation de cette fonctionnalité : observation des jours lors desquels le nombre de cas augmentait très fortement.....	19