

16/11 Choix de thèmes

Brainstorming

Nous avons commencé par donner chacune plusieurs idées en vrac, sans se préoccuper des données dans un premier temps. Voici les différentes idées qui sont ressorties :

- Destination week-end en Europe
- Utilisation des transports lyonnais en fonction de la météo (zoom vélov)
- Réseau de covoiturage en France sur une date particulière (ex. 24 décembre)
- Réseau de covoiturage maison-travail dans la région lyonnaise
- Choix écologique ou non des transports maison-travail en Europe occidentale
- Comparatif de la desserte des villes (transports en commun)
- Comparatif des types de transports en terme d'accident
- Carte de disparition des avions
- Utilisation autoroute/RN en France
- Réseau de transports de cannabis dans le monde
- Distraction homme/femme en voiture

Nous nous sommes ensuite attribuées cinq points chacune à dispatcher sur les différents sujets, dans la limite de deux points pour un sujet.

Elagage

Nous avons gardé une liste réduite et ordonnées de sujets en fonction des points que nous avons attribués :

- 1) Réseau de covoiturage en France sur une date particulière (ex. 24 décembre)
- 2) Comparatif des types de transports en terme d'accident
- 3) Distraction homme/femme en voiture

Puis nous avons éliminé progressivement les sujets en se concentrant sur ceux ayant un potentiel de données intéressant. Finalement nous n'avons gardé que le deuxième :

Comparatif des types de transports en terme d'accident.

Affinage de l'idée

Le sujet étant un peu large, nous avons décidé de l'affiner pour faire une représentation géographique des accidents de voiture.

Nous avons ensuite recherché les données dont nous pouvions disposer librement, et nous avons trouvé sur le site [data.gouv.fr](https://www.data.gouv.fr) une base de donnée des accidents de la circulation en France :

<https://www.data.gouv.fr/fr/datasets/base-de-donnees-accidents-corporels-de-la-circulation>

Une description de ces données est disponible :

https://www.data.gouv.fr/s/resources/base-de-donnees-accidents-corporels-de-la-circulation/20170915-155209/Description_des_bases_de_donnees_ONISR_-Annees_2005_a_2016.pdf

De nombreux paramètres sont à disposition, ce qui permet d'imaginer plusieurs représentations assez différentes.

20/11 Choix de représentations

Pour proposer plusieurs représentations possibles, nous avons réfléchi à ce que nous pouvions montrer.

1) Représentation géographique

Avec une carte, nous pouvons représenter les points chauds géographiques des lieux d'accidents en France, avec un découpage au niveau départemental. Le graphique sera interactif et un zoom sera possible sur un département, pour afficher au survol une répartition des types de véhicules impactés.

2) Représentation par sexe

Nous pouvons représenter les données d'un autre point de vue, celui du sexe. Nous aurions la répartition des accidents pour les hommes et les femmes, avec différents stades de gravité des accidents, et avec un focus sur la répartition des accidents par âge.

3) Représentation par type de véhicule

Nous pouvons représenter la répartition des accidents par véhicules, par exemple voiture/voiture, voiture/piéton, voiture/moto, moto/bus, etc..., puis faire un focus sur la répartition des gravités d'accident, sexe, etc.

Nous avons esquissé ces trois types de représentation pour avoir une idée générale des différentes informations que nous voulons montrer. Voir les scans dans le dépôt git.

04/12 Peer review

Nous avons reçu le numéro du groupe à reviewer. Le point commun de leur projet avec le notre est qu'il consiste en l'affichage de données sur une map. Nous avons chacune reviewé de notre côté, puis nous avons mis nos remarques en commun pour synthétiser tout ça.

07/12 Session développement (leaflet)

Nous cherchons une librairie qui nous permettrait d'afficher la carte de la France, et sur laquelle nous aurions la possibilité de zoomer et de se déplacer sur la map.

Nous décidons d'utiliser la librairie javascript leaflet, qui peut se brancher facilement à la librairie D3.

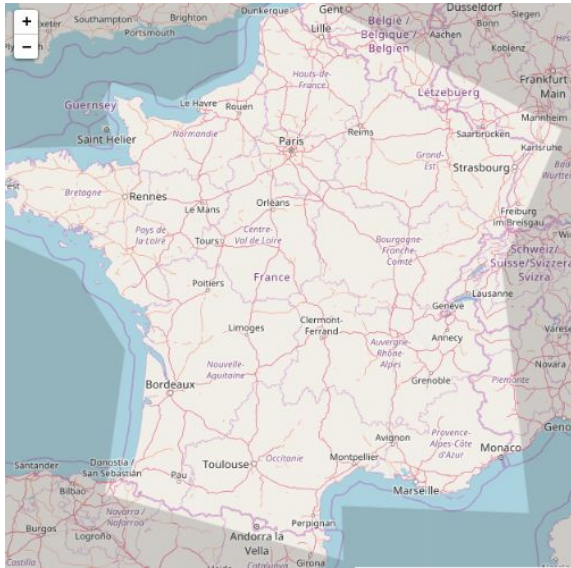
A cette date, nous avons réussi à afficher et cadrer la carte de la France.

21/12 Session développement (geojson)

Nous trouvons dérangeant qu'il n'y ait pas vraiment de délimitation entre la France et les pays qui l'entourent, donc nous faisons des recherches pour trouver comment n'afficher que la France.

Nous retenons une solution qui consiste à tracer un polygone, puis jouer sur l'opacité à l'extérieur de ce polygone.

Pour tracer notre polygone, nous utilisons l'outil web geojson (<http://geojson.io>), puis nous intégrons les coordonnées dans notre map et nous fixons l'opacité sur l'extérieur du polygone à 0.2 pour que la coupure ne soit pas trop nette :



En cherchant comment désactiver le dragging si la map n'est pas zoomée, nous tombons sur cette visualisation, dont nous souhaitons peut-être nous inspirer :

<http://leafletjs.com/examples/choropleth/example.html>

Pistes :

penser à normaliser (diviser par pop locale)

k-medoids => variante qui force les centres à faire partie des données existantes

22/12 Session développement (D3/Leaflet)

Nous cherchons comment utiliser D3 avec Leaflet.

Dans un premier temps, pendant qu'Elisa se penche sur le clustering, nous voudrions afficher des points sur la map plutôt que des bulles pour voir ce que ça donne. Un point serait un accident, la couleur indiquerait la gravité (ou le sexe ?). Nous nous attendons à ce que visuellement le clustering se fasse de lui-même.

Nous rajouterons cette démarche et cette visualisation dans l'article pour montrer par quelles étapes nous sommes passés.

En faisant quelques recherches, nous trouvons cet exemple qui fait ce que nous voulons faire (sur le fond) :

<http://bl.ocks.org/Sumbera/10463358>

Nettoyage de données :

En parcourant nos données sur les accidents de 2016, nous nous sommes rendues compte que pour certaines lignes du CSV, les données spatiales étaient absentes, ce qui rend la ligne inexploitable.

Nous décidons d'écrire un script qui va nettoyer ces données.

wc -l data/caracteristiques_2016.csv :

03/01 Session développement (Leaflet / Clustering)

Nous réussissons à clusteriser nos données à l'aide de Leaflet (il le fait tout seul).

Au zoom, les clusters se mettent à jour tout seuls.

Il restera à faire des graphiques dynamiquement sur chaque cluster.

05/01 Session développement (Changement de plan)

Nous récupérons dynamiquement les données attribuées à chaque cluster pour les utiliser pour faire des visualisations découpées par cluster.

=> les données sont trop volumineuses, ça fait planter la viz.

On ne peut pas pré-traiter les données car il faudrait le faire pour tous les niveaux de zoom.

On décide de chercher une nouvelle stratégie : notre visualisation ne sera pas une carte, mais finalement on décide de faire plusieurs visualisations pour représenter un bilan de l'année sous différents angles.

06/01 Session développement (Différents types de graphiques & innovation)

Nous avons réfléchi entre temps à ce que nous voulons représenter, et nous répartissons les graphiques en fonction des envies de chacune.

On a entre-temps l'idée de faire un graphique représentant les différents points d'impact d'un véhicule. Ça permet d'exploiter un peu d3 sans partir d'un graphique déjà fait. On met une voiture vue du dessus en fond pour rendre ça un peu plus joli.

En parallèle, on fait un calendrier pour rajouter les accidents répartis dans les jours de l'année.

On fait une répartition par sexe, puis par âge (total, homme, femme), puis par gravité.

07/01 Session développement (Intégration page web)

Nous créons un compte github pour héberger notre page.

Nous créons une page HTML en mettant un peu les formes (on utilise bootstrap4).

Nous intégrons les différents graphiques que nous avons faits dans la page web. Nous ajoutons un menu avec des ancres, nous peaufinons un peu tout ça.

08/01 Session développement (Dynamisation)

Nous retravaillons les pie charts que nous avons fait pour ajouter un peu de dynamisme.

Nous retravaillons le line chart pour ajouter du dynamisme.

09/01 Session développement (Peaufinage)

Nous continuons le travail sur les pie charts et le line chart.
Nous commençons à préparer l'oral.