

Rapport final du projet de visualisation

<http://vincent-gaillard.fr/igr204/>

code disponible à l'adresse : <https://github.com/Blacksocks/IGR204>

Le présent document présente le projet que nous avons mené durant la période 4 en *IGR204*. Ce rapport fait suite au design document que nous avons remis à mi-parcours. A travers les quatre parties qui structurent ce document, nous allons vous présenter notre travail, tant sur le plan du design que sur toute la réflexion en amont qui a abouti à notre prototype final. Dans une première partie, nous allons vous présenter les données sur lesquelles nous avons décidé de travailler. Ensuite, nous nous focaliserons sur les utilisateurs que nous visons à travers notre design. En lien direct avec la précédente partie, nous montrerons quelques tâches représentatives de celui-ci. Enfin, nous vous présenterons le design final, qui est la convergence des réflexions que nous avons menées dans les parties précédentes.

I) Le jeu de données

Nous sommes partis, initialement, avec un seul jeu de données sur les homicides aux États-Unis. Ces données sont extrêmement précises. Il y a notamment: le statut du crime (élucidé ou non), la ville dans laquelle a lieu le crime et l'état également, l'année, le mois, le type d'arme utilisée, le sexe de la victime, etc (voir ci-après pour une description exhaustive). Bref, nous avons des données de tout type. Nous allons voir dans la prochaine partie que nous avons effectué un tri sur celles-ci, pour ne garder uniquement celles qui ont une réelle importance vis-à-vis du public visé.

De manière exhaustive, notre jeu de données est :

Identification du crime, Code de l'agence, nom de l'agence, type de l'agence, ville, état, année, mois, incident, type de crime, crime résolu ou non, sexe, âge et ethnie de la victime, sexe, âge et ethnie du meurtrier, relation avec la victime, arme, nombre de victimes, nombre de meurtriers et enfin source d'enregistrement.

Dans l'intérêt de notre visualisation, et nous en parlerons expressément dans la suite de ce rapport, nous avons choisi d'utiliser un deuxième jeu de données, qui cette fois-ci, concerne la population aux Etats-Unis. De nouveau, certaines informations ne nous sont pas d'une très grande utilité et nous avons donc également effectué un tri sur celles-ci. Nous reviendrons sur ce tri également dans la suite de ce rapport.

Dans l'intérêt du rapport, voici de manière exhaustive à nouveau notre jeu de données :

Année, abréviation postale de l'état, code FIPS de l'état, code FIPS du comté, registre, ethnie, origine, sexe, âge, population.

Ce fichier est initialement très lourd à charger puisqu'il fait environ dix-huit millions de lignes. Pour réduire ce nombre drastiquement, nous avons directement effectué une fusion sur l'âge (chaque ligne correspond initialement à une tranche d'âge, et il y a en tout dix-huit tranches). De cette façon, notre fichier ne fait plus qu'environ un million de lignes, et a toujours autant d'intérêt pour notre représentation.

II) Les utilisateurs

Dans le design document que nous avons rendu précédemment, nous avons choisi comme utilisateurs cibles des utilisateurs non-professionnels, amateurs. Si le public que nous visons dorénavant reste globalement toujours le même, nous voulons en outre faire en sorte que notre design puisse servir comme support visuel à des fins éducatives, ou encore politiques. Dans cette partie, nous allons nous intéresser à ces différents utilisateurs, en précisant l'expérience qu'ils ont de nos données, des outils de visualisation. Nous allons également revenir sur ce que nous voulons leur faire comprendre au contact de notre visualisation.

Le premier public visé est donc un public amateur, qui n'a pas forcément une grande expérience à la fois de nos données et de divers outils de visualisation. Il nous appartient donc de leur présenter nos données de manière claire, et de faire en sorte que l'interaction soit la plus intuitive possible. Nous reviendrons en détail sur l'interaction avec l'utilisateur dans la dernière partie. Concernant la visualisation de nos données, nous voulons leur donner une idée générale du nombre d'homicides par état et au niveau national, soit en brut, soit en faisant un rapport avec la population. Nous voulons leur montrer quel état a le plus grand nombre d'homicides, ou encore le plus grand rapport homicides sur population. Nous voulons également leur présenter l'évolution de ces données

dans le temps, pour leur permettre de voir les tendances générales, les pics/creux des années précédentes. Nous faisons en sorte également de séparer les homicides et la population en fonction des ethnies, pour permettre une analyse plus en profondeur. Enfin, nous voulons donner la possibilité pour chaque état de voir les détails des homicides, en montrant notamment l'arme du crime, ou encore la relation que le meurtrier a avec la victime.

Pour le deuxième public visé, averti sur nos données, mais moins sur l'interaction avec des outils de visualisation, leur regard va être beaucoup plus critique. Le fait, par exemple, que le nombre d'homicides diminue au cours du temps peut être utilisé à des fins politiques. Si les homicides diminuent, pourquoi, par exemple, faudrait-il revenir sur des lois concernant le port d'armes ? En contre-argument, le fait de voir que les armes à feu sont les principales armes utilisées pour commettre ces crimes peut clairement être utilisé par les partis opposés. Notre visualisation peut également servir à des fins pédagogiques, pour faire un historique en quelque sorte des flux d'homicides par exemple. Enfin, pour les crimes non-élucidés, notre visualisation apporte un intérêt très pratique. En effet, il suffit de regarder, pour les crimes élucidés, les relations entre la victime et la meurtrier. En fonction, s'il y a une relation majoritaire (par exemple, mari ou femme), il y a une plus grande probabilité de trouver le meurtrier en se référant à cette relation.

Après ces réflexions, nous savons donc maintenant ce que nous voulons montrer et faire comprendre à nos utilisateurs. Ainsi, quels que soient les utilisateurs que nous ciblons, comme rapidement mentionné dans la partie précédente, nous faisons un tri sur nos deux jeux de données pour ne garder que les informations pertinentes. Cette étape a le mérite de faire diminuer fortement le temps de chargement de nos données, tout en gardant tout ce qui est pertinent pour notre visualisation.

- Sur le jeu de données des homicides, nous supprimons les informations sur les agences, nous ne gardons que les années et les états (pas les mois ni les villes, car trop précis) et nous enlevons la source d'enregistrement.
- Sur le jeu de données de la population, nous ne gardons que les états et ne considérons plus les comtés ni les registres, et nous supprimons l'origine (seul l'ethnie véritable nous intéresse).

III) Tâches représentatives

La première tâche qui nous semble importante est la sélection de ce que l'on veut visualiser: le nombre brut des homicides, le nombre brut de la population ou le ratio entre les deux. C'est évidemment une tâche de filtrage des données. Nous voulons faire en sorte que cette action soit la plus simple possible, donc juste en cochant la case correspondante. Nous voulons également faire une sélection/filtrage au niveau des années. Il faut aussi que la sélection soit aisée, par exemple par simple scroll de la souris ou par un clic sur la timeline. Nous souhaitons en outre que l'utilisateur puisse avoir les informations pour chaque état, et nous devons donc une nouvelle fois faire une sélection/filtrage, mais cette fois-ci sur l'état en question. Nous pensons qu'effectuer cette action par simple survol de la souris est pertinent et particulièrement adapté. En fait, cela permet d'avoir très rapidement les informations sur chaque état. De plus, nous réservons le clic sur l'état pour rentrer dans le détail des homicides. Nous voulons notamment rendre accessible les données sur l'arme utilisée, la relation avec la victime et le statut du crime (résolu ou non), des informations qui peuvent s'avérer être intéressantes.

IV) Le design final

Nous connaissons nos utilisateurs, et nous savons quelles tâches notre visualisation doit accomplir. Notre design s'est donc construit autour de ces réflexions, que nous vous avons présentées dans les parties précédentes.

Nous avons deux designs distincts: le premier permet de voir les statistiques générales au niveau national et les tendances générales au niveau des états; le second est une vue plus détaillée des homicides dans l'état considéré.

Revenons d'abord sur le premier design :

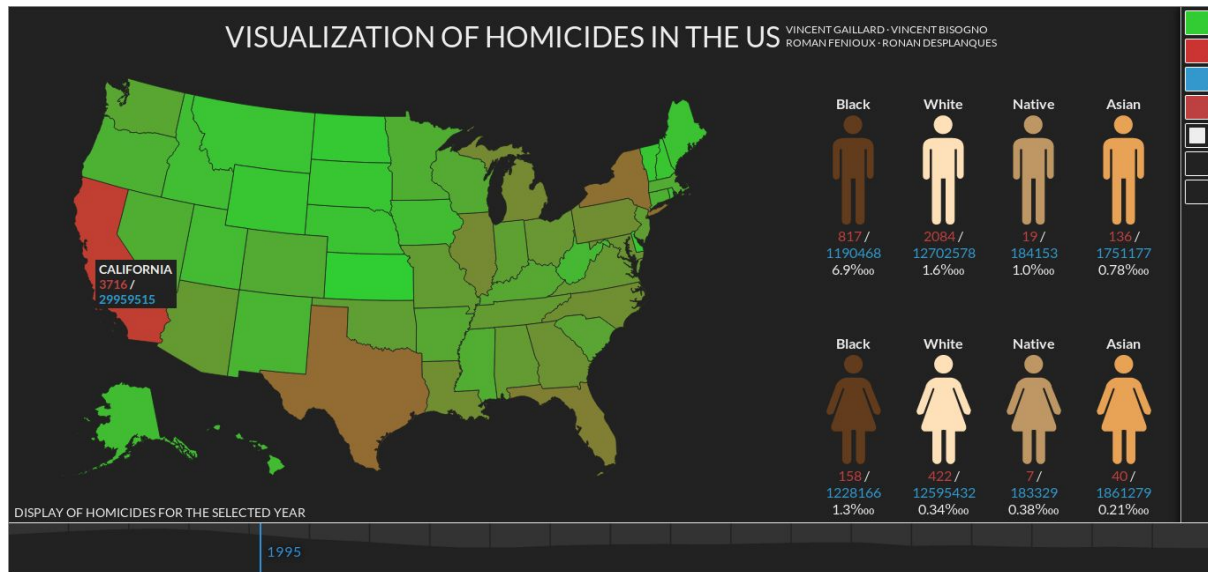


Figure 1 : Premier design

Comme vous pouvez le voir ci-dessus, ce design est constitué de quatre parties.

La première partie est la carte des Etats-Unis. Chaque état est coloré selon un code simple, mais efficace: le rouge signifie une valeur élevée, le vert une valeur faible. La couleur est déterminée en fonction de la valeur minimale et maximale du critère choisi. Cela permet aux utilisateurs de voir directement les tendances, et ce de manière très aisée.

La deuxième partie est constituée de représentations d'hommes et de femmes, triées selon leur ethnie : blanche, noire, asiatique ou native. Dessous, nous affichons le nombre d'homicides appartenant à cette ethnie, la population et un rapport entre les deux (x pour 10000). Cette dernière information est particulièrement intéressante, puisqu'elle permet de déterminer si une ethnie est ciblée en majorité. Pour faire suite au design document, nous avons choisi cette représentation plutôt que les diagrammes circulaires pour sa clarté : l'utilisateur comprend directement l'information en fonction de l'ethnie.

La troisième partie est la timeline. Nous montrons sur celle-ci la courbe correspondant au critère choisi, mais au niveau national. Les évolutions en fonction du temps sont donc directement visibles et exploitables, aux fins souhaitées par les utilisateurs (fins éducatives comme nous l'énonçons plus haut, ou fins politiques).

Enfin la quatrième partie est la légende, où l'utilisateur peut sélectionner le critère qu'il veut visualiser (cf ci-dessous).

L'interaction de l'utilisateur avec ce design est triple, mais reste très intuitive. C'est un point crucial pour notre représentation qui est avant tout, nous le rappelons, orienté vers un public amateur. La première interaction est par simple survol: soit sur les états pour que les informations s'affichent à droite au niveau des bonhommes, soit sur la légende pour que celle-ci se déplie. La deuxième

interaction est par clic : soit sur les états à nouveau, pour arriver sur l'autre design, soit pour sélectionner le critère une fois que la légende est dépliée, soit enfin pour changer d'année au niveau de la timeline. Enfin, nous avons introduit une autre option au niveau de cette dernière. En effet, l'utilisateur peut également utiliser le scroll de la souris pour se déplacer au sein de celle-ci. Il résulte de tout ceci que ce design est à la fois intuitif et ludique pour l'utilisateur.

STATES	
MIN. NUMBER OF HOMICIDES	■
MAX. NUMBER OF HOMICIDES	■
POPULATION AND DEATH	
POPULATION IN A STATE	■
NUMBER OF DEATH	■
DATA DISPLAYED ON MAP	
DISPLAY NUMBER OF HOMICIDES	<input type="checkbox"/>
DISPLAY ONLY POPULATION	<input type="checkbox"/>
DISPLAY NB. OF HOMICIDES / POP.	<input type="checkbox"/>

Figure 2 : Légende et sélection du critère

Une fois que l'utilisateur clique sur un état, nous arrivons donc à ce design, un "radial tree map".

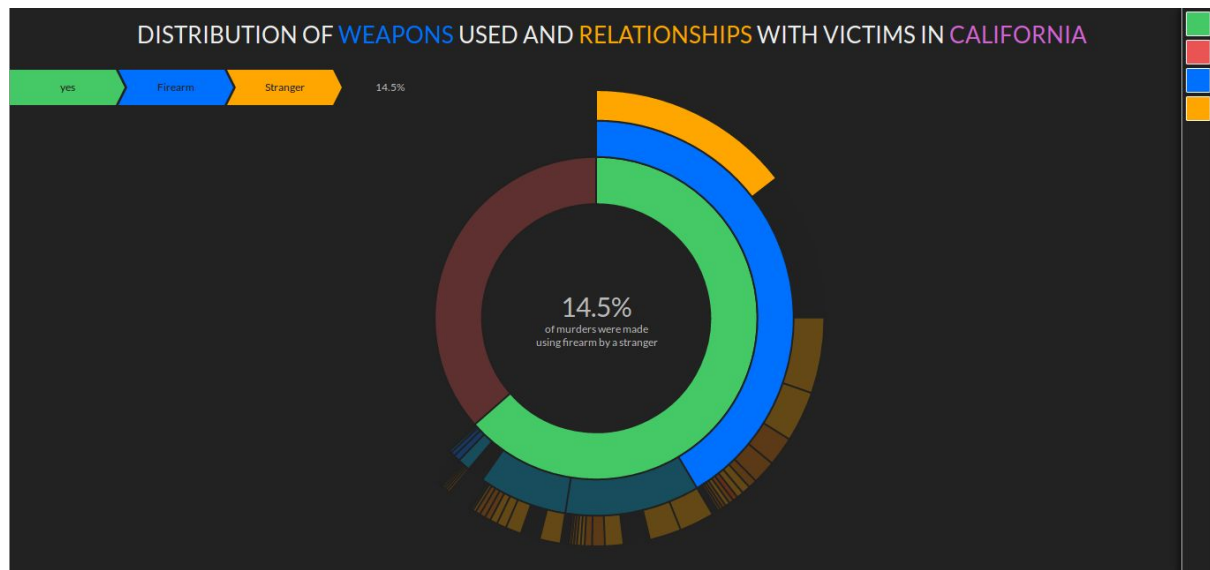


Figure 3 : Second design

Ce design permet de montrer les détails liés aux homicides. A nouveau, la légende sur la droite de la page (cf figure 4) ne se déplie que par survol de la souris. Nous commençons par montrer le statut du crime (élucidé, en vert; non élucidé, en rouge), puis le type d'arme utilisée : armes à feu, couteaux, armes contondantes notamment. Enfin, nous affichons la relation que le meurtrier entretient

avec la victime. L'interaction avec l'utilisateur est simple et efficace : on part du centre pour aller vers l'extérieur.

Ce design permet de bien mettre en oeuvre un de nos objectifs, qui est de se référer aux meurtres élucidés pour essayer de résoudre ceux qui ne le sont pas. L'utilisateur voit très facilement la proportion associée à la relation entre victime et meurtrier en fonction de l'arme utilisée. Il peut donc déterminer vers quel groupe de personnes il y a la plus grande probabilité de trouver le tueur pour les homicides non-élucidés. Ce design montre aussi clairement quelles armes sont majoritairement utilisées par les meurtriers.

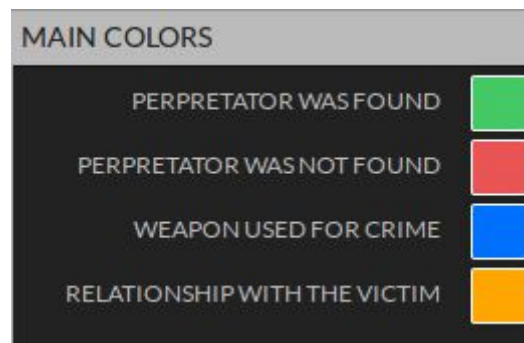


Figure 4 : Second design, légende

V) Conclusion et ouverture

Le design final que nous avons implémenté permet d'effectuer les tâches que nous avons citées plus haut, telles que la sélection d'un état, ou la sélection d'une année particulière, et de s'adresser à notre public cible qui est un public avant tout amateur, mais qui peut avoir une certaine expérience avec nos données, en particulier pour le public qui veut se servir de notre visualisation à des fins pédagogiques ou politiques. Nous le rappelons, notre prototype final est en fait constitué de deux designs, positionnés à des niveaux de précision différents, mais dont les objectifs sont parfaitement complémentaires.

Un des retours majoritaires que nous avons eu pendant la séance des posters concerne notre second design, le "radial tree map". Il serait intéressant en fait de pouvoir changer l'ordre des attributs que nous affichons (statut, armes et relations) et pourquoi pas en afficher d'autres selon les données disponibles (comme par exemple, l'âge des victimes et des meurtriers, une donnée que nous n'utilisons pas en effet). L'idée est de donner à l'utilisateur tout d'abord le choix des données à afficher, puis l'ordre dans lequel les afficher. Il est certain que c'est une idée particulièrement intéressante, et nous la laissons en ouverture de notre compte rendu de rapport.