

Rapport projet UE3 Data Visualisation

Aviation incidents

Par: Nait Larbi Takfarinas et Benali Laetitia

Description des données:

Le jeu de donnée que nous avons utilisé a été pris sur Kaggle :

<https://www.kaggle.com/khsamaha/aviation-accident-database-synopses>

Le Dataset se compose de 33 colonnes et de 85977 lignes. On y retrouve des données d'accidents d'avions et d'hélicoptères depuis 1982 à 2021, suivant la localisation, le nombre de blessés et d'autres indicateurs.

Préparation des données:

Pour notre Dashboard nous avons pris un échantillon de données en fonction de celles qui nous intéressait le plus, à savoir:

- Garder que les données concernant les Etats unis (représentaient déjà la majorité)
- Garder les accidents ayant eu lieu après l'an 2000

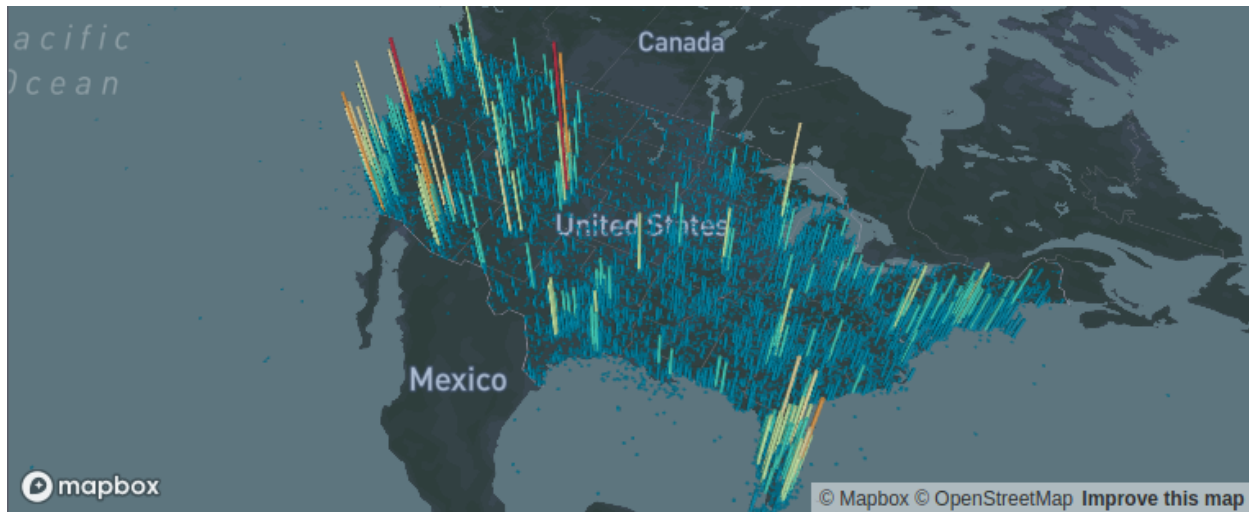
Cela nous réduit à 29704 lignes.

De plus, nous n'avons pas pu utiliser toutes les colonnes, plusieurs d'entre elles n'ont pas pu représenter des vrais cas d'usage, tels que les identifiants, code de l'aéroport, etc.

Description des visualisations:

Carte Géographique:

La carte représente la carte du monde mais centrée sur les états unis, en effet la majorité de nos données y faisaient référence. Nous avons donc exploité les données géographiques (latitude, longitude) de nos données afin de représenter le nombre d'accidents ayant eu lieu sur un lieu donné. La longueur du trait sortant d'un point représente le nombre d'accidents, le plus il est grand et plus il y a eu d'accidents. La colorimétrie change également et devient rouge si le nombre y est important.

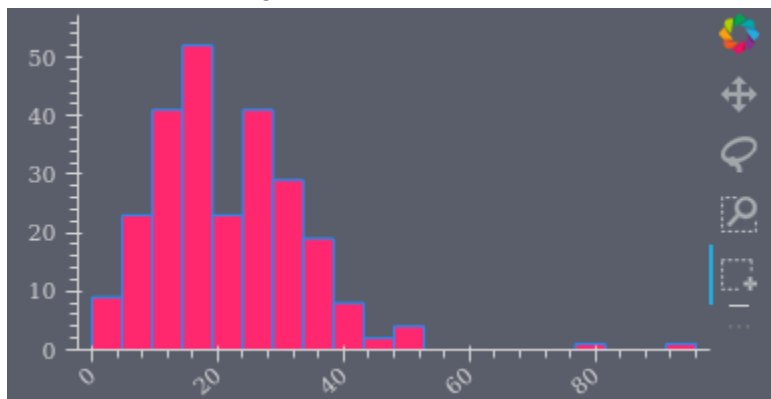


Il y a également la possibilité de choisir la rotation de notre carte sur une box en y mettant la rotation souhaitée.

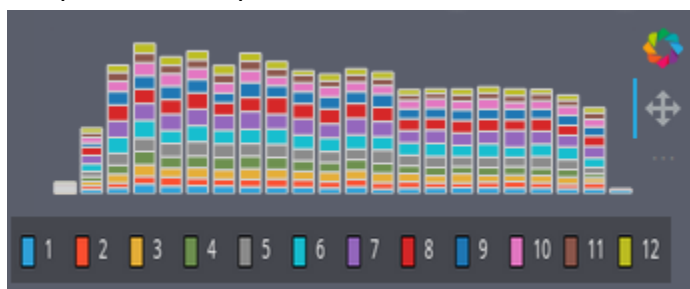
Histogrammes:

Nous retrouvons sur notre Dashboard 2 histogrammes.

Le premier étant interactif avec d'autres graphiques représente le nombre de mois par rapport au nombre de blessés graves

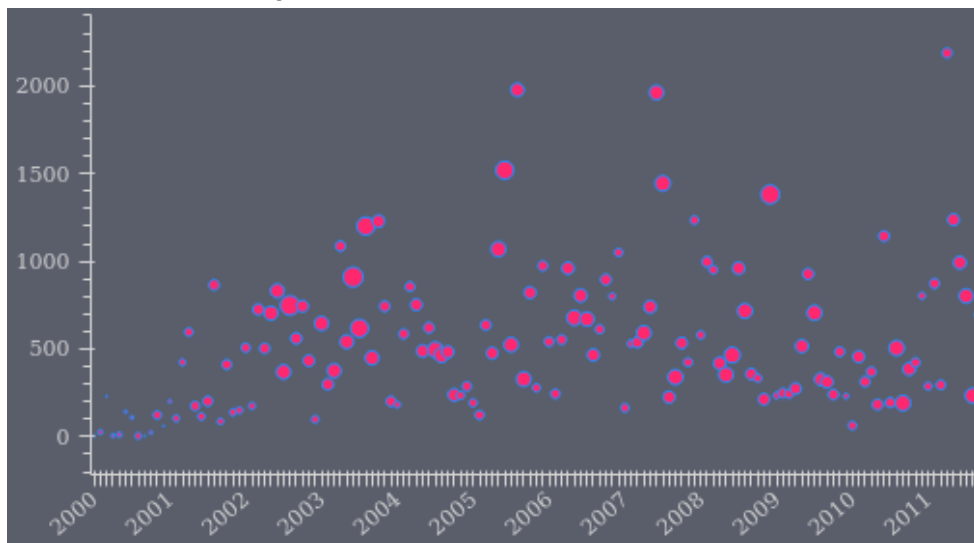


Le deuxième est un histogramme empilé qui représente la distribution des incidents par année, où chaque couleur représente un mois de l'année.



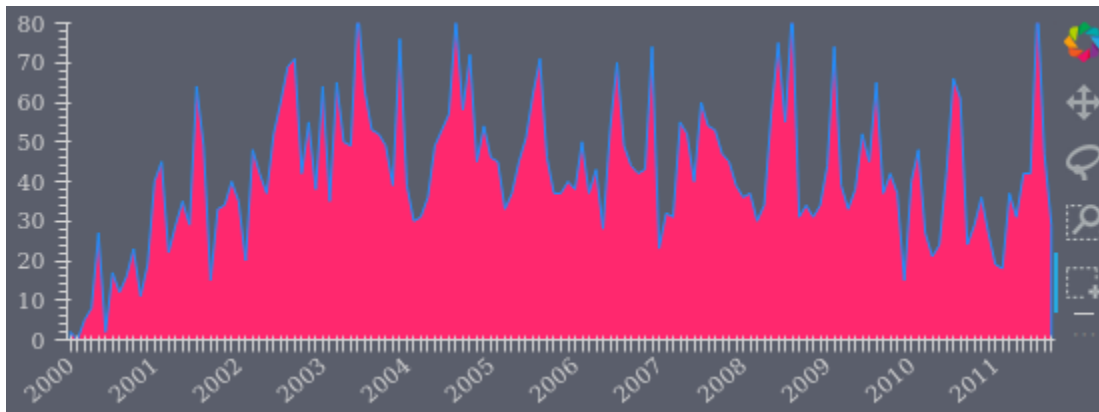
Scatter plot:

Le scatter que nous avons représenté illustre 2 mesures différentes, en premier lieu nous avons le nombre de non blessés en axe y, en fonction du mois. De plus, la taille des points compte aussi en termes de blessés légers.



Courbe:

Ce dernier graphique représente une courbe du nombre de décès en fonction par mois depuis 2000. La spécificité de cette courbe est que nous pouvons lui appliquer un lissage gaussien selon le degré souhaité, ce paramètre est modifiable grâce à une zone d'input en bas de la page.



Lien entre les graphiques:

Nous avons évidemment rendu nos graphs interactifs. Ces interactions sont appliquées sur l'histogramme, le scatterplot et la courbe, et les fonctionnalités disponibles sont:

- La sélection d'une zone (rectangle) sur l'un des graphiques agit sur les autres en zoomant sur ces derniers
- Un slide bar où nous pouvons choisir un intervalle d'années qui filtre sur tous les graphiques cités, mais également la carte géographique.

Conclusion:

Tout le travail effectué a été réalisé en python à l'aide de bibliothèques spécifiques tel que Panel, cela nous a permis d'apprendre à réaliser de vrais Dashboard from scratch, sans compter le fait d'avoir acquis ce nouvel œil sur la dataviz et son importance.