



Note Méthodologique – Data Visualisation des données de la DGAC

Principes de la solution de data visualisation proposée

Une [étude publiée récemment](#) dans Nature Climate Change met évidence la contribution massive du secteur du tourisme globalisé aux émissions de gaz à effet de serre. Au sein du secteur du tourisme, le trafic aérien contribue particulièrement à l'augmentation de ces émissions. La compréhension du transport aérien, qu'il soit à visée touristique ou non, est donc primordiale dans un contexte de transition vers des comportements plus sobres en émission de carbone. C'est tout l'objet de la visualisation que nous vous proposons.

Dans le cadre du concours de data visualisation de la DGAC, nous avons choisi de nous intéresser aux grandes dynamiques du secteur aérien depuis 1990. Notre objectif est de mettre en évidence ces phénomènes et leur impact en termes d'émissions de CO₂.

Augmentation du nombre de passagers, optimisation du remplissage des appareils, impact des crises géopolitiques, ouvertures de lignes, pattern saisonnier du trafic ou développement du low-cost, nous proposons à l'utilisateur de comprendre l'évolution du trafic aérien depuis 1990 grâce à un ensemble de figures et graphiques interconnectés.

Les différentes fenêtres de notre application permettent une exploration plus fine du jeu de données. Croisement de dimensions, sélection temporelle ou géographique, analyse ciblée d'une destination ou d'un aéroport, notre visualisation interactive permet également une utilisation experte, tout en restant simple et intuitive, du jeu de données.

Les visualisations développées par Données Brutes sont structurées autour de la notion d'interactivité pour permettre à l'utilisateur d'être acteur de l'exploration de données. Son utilisation des données peut se décliner sur plusieurs niveaux de lecture. Nous sommes convaincus qu'une bonne visualisation doit répondre à une problématique précise, permettre plusieurs couches d'analyse tout en présentant des informations cohérentes. C'est le challenge que nous nous sommes fixés pour cette visualisation.

Traitements des données réalisés

Choix des données

Pour concilier l'utilisation de données temporelles avec les besoins de réactivité et d'interaction en temps réel de l'application, nous avons décidé d'opérer une sélection dans les fichiers proposés. Notre objectif? Faire apparaître les tendances de long terme du secteur aérien tout en conservant une base de données légère et maniable. Nous nous sommes également montrés attentifs à



sélectionner les données les plus complètes pour proposer une visualisation fiable qualitativement comme quantitativement. Notre choix s'est porté sur les fichiers « trafic », qui fournissent les informations les plus complètes sur le long terme.

L'analyse de ces données de trafic nous a appris que plus de 95% du trafic aérien français transite par les dix principaux aéroports du pays (classement établi sur la volume de passagers qui transitent par ces aéroports). Nous avons donc choisi de concentrer notre analyse sur ce top ten. De même, nous avons constaté une forte symétrie entre les vols au départ de ces aéroports et les vols à l'arrivée. Nous avons donc décidé de filtrer les données en ne gardant que les paires de villes pour lesquelles un aéroport du top ten français était mentionné dans la colonne « DEP ».

En ce qui concerne l'impact environnemental du trafic aérien, nous avons renoncé à utiliser les fichiers « émissions » fournis. Malgré la richesse des informations, puisqu'ils déclinent l'impact environnemental en plusieurs types d'émissions, nous ne pouvions pas en déduire les émissions de CO₂ associées à chaque paire d'aéroports. En effet, les émissions y sont données de manière agrégée pour chaque aéroport du top ten en ne distinguant que le faisceau, la zone du vol et le type d'appareil. De plus, les données des fichiers "émissions" étaient limitées à 2016 pour les aéroports du top ten (Orly et Charles de Gaulle exclus). Nous avons donc mobilisé [l'éco-calculateur de l'aviation civile](#) pour collecter cette information.

Les informations disponibles ont été récupérées en utilisant des méthodes de collecte automatisée. Ces méthodes visent à contrôler un navigateur internet via un langage de programmation, ici Python, et extraire les informations de chaque page. C'est ainsi que nous avons construit la base de données d'émissions.

Les données d'émissions de l'éco-calculateur de la DGAC fournies en 2018 ont été appliquées à toutes les années de vol. Nous faisons donc l'hypothèse que les émissions par passager sont constantes dans le temps, ce qui est sûrement réducteur du fait de l'évolution des types d'appareils, du taux de chargement des avions etc.

Traitement des données collectées

A partir des informations sur les émissions de carbone associées à chaque paire d'aéroports collectée, nous avons constitué la base de données. Elle inclut, pour chaque mois entre janvier 1990 et décembre 2016 et chaque paire d'aéroports intégrant un départ depuis l'un des aéroports du top ten, les informations relatives au nombre de vols, de passagers, les émissions de CO₂, les km parcourus par les appareils sur le trajet concerné et des données techniques sur les aéroports de destination (zone, faisceau, dénominations, coordonnées GPS).

Cette base de données permet l'agrégation des données par pays, par mois et par aéroport de départ pour la construction de l'application de visualisation. Il convient de noter que les données d'émission de carbone disponibles dans l'application sont sous-évaluées. Certaines paires d'aéroports ne figuraient pas dans l'éco-calculateur de l'aviation civile. Les émissions de CO₂ associées à ces

trajets ne sont donc pas prises en compte. Nous avons pris le parti de ne pas chercher à les estimer afin de rester représentatifs de l'information brute contenue dans les données du concours.

Solution technique pour la visualisation

Pour ce concours, nous souhaitons développer une application interactive, consultable facilement et, a fortiori, sans mobiliser de licence. La visualisation devait permettre de représenter plusieurs informations selon le pays de destination, l'aéroport de départ, l'année et le mois.

Tous les codes de traitement de données ont été écrits en Python. Un jeu de données au format json a été construit avec toutes les données nécessaires. Un serveur web a été mis en place avec la technologie Node.js. Ce serveur réalise des traitements à la volée, met à disposition une API et la page de l'application.

La visualisation a été réalisée à partir de la librairie open source dc.js. Cette librairie permet de combiner la puissance de visualisation de d3.js et la possibilité de filtrer les données à la volée de crossfilter. Cela en fait un outil particulièrement efficace pour développer des visualisations fluides qui permettent plusieurs niveaux de lectures.

Dans la mesure où la visualisation de la page est une application web, elle a été structurée en html, les éléments de design sont en css et la logique applicative est codée en javascript.

Captures d'écran

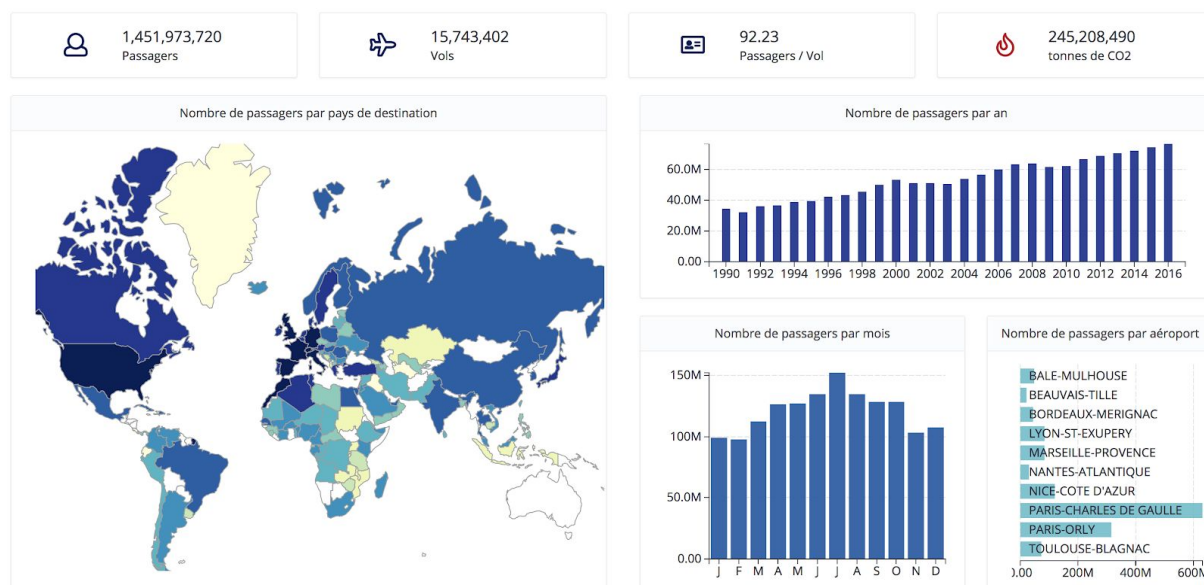


Figure 1 - Aperçu de l'application de visualisation. Les logos du haut permettent de sélectionner la variable d'intérêt. Tous les graphiques sont interactifs, cliquer sur un élément permet de filtrer les données.



Figures 2 et 3 - Sélection d'une autre variable (à gauche) et filtres sur les données (à droite). Dans tous les cas, les données s'actualisent en fonction des sélections.

Pistes de recherche et d'amélioration de l'outil

Les données fournies dans le cadre du concours sont très riches et couvrent de nombreuses problématiques. Certaines variables ne sont pas disponibles pour l'ensemble de la période, des aéroports etc. Pour fournir une vision plus exhaustive des évolutions du transport aérien, et notamment de son impact environnemental, un travail d'extraction d'information à partir des données disponibles est nécessaire.

Pour disposer d'une information plus précise sur les différents types d'émissions associés à chaque destination (agrégation par pays ou par aéroport), on pourrait mener une étude statistique sur la corrélation entre les types d'émissions. Cela nous permettrait de convertir l'impact des vols, non seulement en termes de CO₂ mais également de méthane ou de particules fines.

Pour mieux saisir l'impact des choix des voyageurs, on peut également imaginer une étude économétrique pour évaluer la manière dont les prix des vols influent sur les choix de destination des voyageurs.

Nous espérons que ce travail saura vous convaincre et que vous aurez autant de plaisir à l'explorer que nous à le réaliser !

Nina Ganchou & Mathieu Anderhalt
pour Données Brutes

Vidéo :

- google : <https://drive.google.com/file/d/1xw07O9xxSgnW3y5BnKl5jvCZNXFdPI7/view?usp=sharing>
- sync : <https://ln.sync.com/dl/1a715d530/2tubpbzt-68g4phja-ncjxu692-shb7a39z>

Application : <http://www.donneesbrutes.com/dgac>