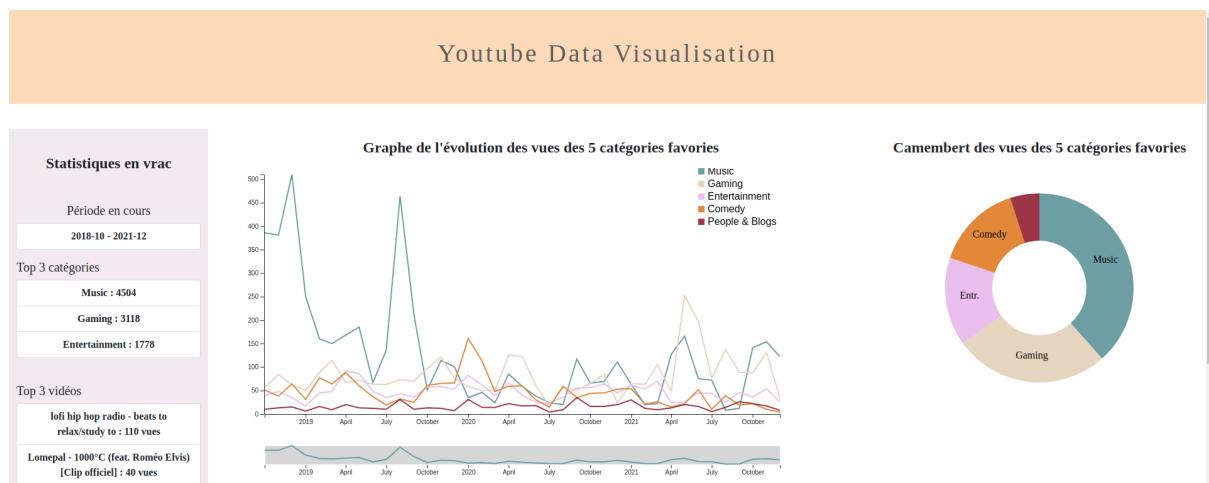


Document cadrage

Projet Youtube Data Visualisation



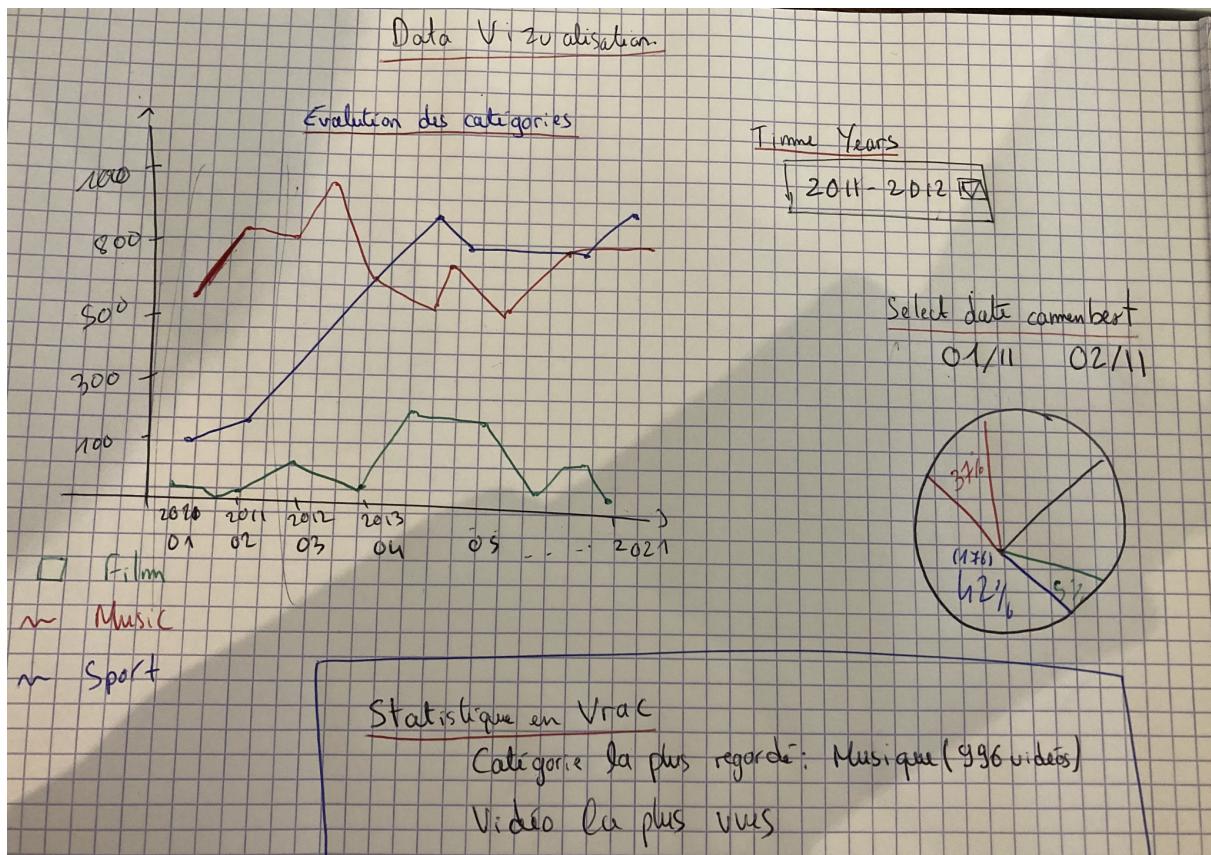
Semaine 09/12/2021 - Choix des données

Nous avons fait le choix de travailler sur nos données personnelles Youtube, à savoir l'analyse de notre historique. Pour ce faire, nous utiliserons l'API de Google.

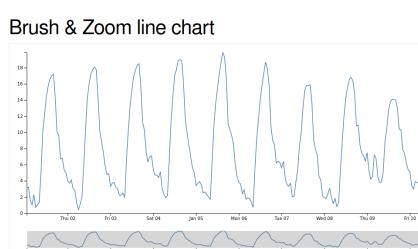
Semaine 13/12/2021 - Esquisse et récupération des premières données

Nous avons choisi de représenter nos catégories visionnées sous deux formes : une représentation sous forme de graphe et une représentation sous forme de camembert.

Esquisse de notre projet



Nous avions également prévu de faire un brush pour sélectionner les dates que l'on souhaite traiter. Nous comptons nous inspirer de ce graphe¹.



¹Efrat Vilenski. *Brush & Zoom line chart*. Efrat Vilenski, 2019, <https://bl.ocks.org/EfratVil/92f894ac0ba265192411e73f633a3e2f>.

Nous avons ensuite écrit un script JavaScript pour requêter l'API, seulement l'API de Google est limitée en nombre de requêtes (10000/jours) ce qui posait un problème avec la quantité de données à récupérer (plusieurs années d'historiques). Mais nous disposions tout de même de nos premières données, et nous allions pouvoir commencer la représentation.

Semaine 23/12/2021 - Mise en place squelette Bootstrap et première visualisation

Baptiste a mis la vue en place : choix des dispositions des éléments, des couleurs du site web, et du squelette Bootstrap.

Parallèlement, Timothée a mis en place la première visualisation : le graphe linéaire. Dans un premier temps statique affichant sur l'intégralité des dates. Pour cela, il a fallu également retravailler les données JSON pour être utilisable facilement lors de la création des courbes avec D3.

L'extraction des données à l'aide du script JS étant trop compliqué (Trop de requêtes à la seconde pour l'API Youtube qui refuse la connexion). Nous avons alors opté pour un script Python. Plus efficace et résistant aux erreurs (avec possibilité d'enregistrer les logs d'erreur lors de la récupération des données.)

Semaine 27/12/2021 - Les statistiques

Statistiques en vrac

Période en cours

2018-10 - 2021-12

Top 3 catégories

Music : 4504

Gaming : 3118

Entertainment : 1778

Top 3 vidéos

lofi hip hop radio - beats to relax/study to : 110 vues

Lomepal - 1000°C (feat. Roméo Elvis)
[Clip officiel] : 40 vues

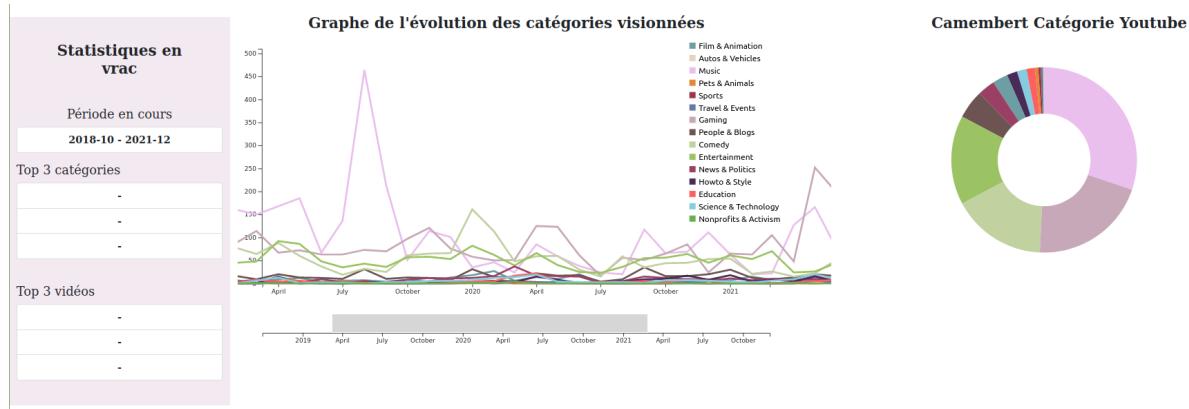
Get Abs in 2 WEEKS | Abs Workout Challenge : 35 vues

Baptiste a créé une fonction qui permet de calculer en temps réel des statistiques sur les données avec les dates sélectionnées. Nous avions alors 2 statistiques en place : les 3 catégories les plus visionnées, ainsi que les 3 vidéos les plus regardées.

Timothée a continué d'avancer le graphe linéaire. Il a notamment mis en place le brush pour sélectionner les dates à afficher. Au vu du nombre de lignes à afficher et de l'écart entre les courbes des catégories les plus et les moins vues, Timothée a ajouté la possibilité d'afficher ou de retirer des lignes en cliquant sur la légende ou en cliquant directement sur une courbe pour n'afficher qu'elle.

Semaine 03/01/2022 - Deuxième visualisation et première soutenance

Baptiste a mis en place la deuxième visualisation : le camembert. Proposant ainsi une autre façon de représenter les données et de les explorer.



Bilan soutenance 06/01

- Il y a trop de catégories pour nos visualisations, ce qui rend le visuel un peu brouillon.

Solution apportée : nous séparerons en 4 visualisations : un graphe et un camembert pour les 5 premières catégories favorites, et un graphe et un camembert pour les 10 autres catégories.

- D'autres statistiques seraient intéressantes à explorer.

Solution apportée : nous ajouterons le nombre de vidéos total vues sur la période ainsi que les 3 chaînes les plus visionnées.

- Ajouter une page de référence pour les sources que nous avons utilisées pour le projet.

Codage des solutions

Nous avons encapsulé le code des créations de nos visualisations dans des fonctions. Afin de pouvoir appeler une fonction création de camembert et de graphe sur des jeux de données distincts. Et ainsi créer nos quatre visualisations sans redondances. Nous pourrions alors facilement dupliquer avec d'autres graphes supplémentaires si besoin.

Deux nouvelles statiques apparaissent sur notre site web, ainsi qu'une page de référence.



Semaine 10/01/2022 - Refactorisation du code et seconde soutenance

Actuellement, nous avons un code d3.js codé dans une version 4, ainsi que deux fonctions de créations : une pour le graphe et une pour le camembert, et deux fonctions update pour les mêmes éléments.

Nous avons décidé de re-coder nos visualisons en version 6 de D3 en utilisant la fonction join permettant à nos fonctions de créations de servir également d'update.

Bilan soutenance 12/01 :

- Certaines couleurs de catégories se ressemblaient et étaient donc moins dissociables.

Solution apportée : choisir une meilleure échelle de couleur

- Pour le graphe : la légende des catégories n'était pas ordonnée, et les courbes dépassaient des axes.

Solution apportée : ordonner la légende des catégories, régler le problème de dépassement

Nous avons corrigé les éléments du bilan de la soutenance et réglé les derniers détails du projet. Ce qui clôture le projet.

Erreurs, difficultés et pistes non suivies

- 1) Une de nos premières erreurs a été d'ajouter manuellement les différentes courbes du graphe linéaire, alors qu'il aurait été nettement plus facile d'utiliser dès le début la fonction Join proposée par la bibliothèque D3.
- 2) Lors de nos premières visualisations, nous avions fait le choix de garder l'intégralité des données que nous renvoyait l'API Youtube. Le fichier JSON était alors très lourd et prenait par conséquent du temps à charger. Nous avons par la suite tronçonné les données afin de ne garder que l'essentiel.
- 3) Au départ, nous aurions aimé travailler plus sur les tags des vidéos (mot clef de chaque vidéo) mais cela aurait nécessité d'utiliser une autre visualisation, car cela aurait surchargé les données déjà présentes sur ces graphes.