

# Visualisation

Alban FALCOZ, Alexandre GRIPARI, Jilian LUBRAT, Valentin QUIOT, Theo CHOLEWA

2 Novembre 2025

## Table des matières

Introduction . . . . .	1
Préparation des datasets . . . . .	1
Type des datasets . . . . .	2
Type des données . . . . .	2
Persona . . . . .	3
Le créateur de contenu . . . . .	3
Le journaliste / analyste spécialisé . . . . .	4
Pipeline . . . . .	4
Fonctionnement étape par étape . . . . .	4
Types de visualisations . . . . .	5
Tree visualisation . . . . .	5
Bubble chart . . . . .	6
Histogram . . . . .	7
Carte interactive . . . . .	7
Pie chart . . . . .	9
Démarrage du projet . . . . .	9

## Introduction

Ce rapport présente le projet de visualisation des chaînes YouTube 2025, un site web permettant d'explorer et d'analyser les données de plus de 15 000 chaînes YouTube à travers le monde. L'objectif principal est d'offrir une compréhension visuelle et intuitive des tendances du paysage YouTube : quels sont les pays les plus représentés, quelles catégories dominent, comment les abonnés se répartissent géographiquement, et quelles vidéos génèrent le plus d'engagement.

Le projet s'articule autour de plusieurs visualisations complémentaires (Treemap, Bubble chart, Carte géographique, Histogramme, Pie chart) qui permettent d'explorer les données selon différentes perspectives et niveaux de détail.

L'architecture du projet repose sur une pipeline de traitement des données flexible et extensible, qui facilite le nettoyage, la transformation et l'agrégation des datasets avant leur visualisation. Cette approche modulaire garantit la cohérence des filtres appliqués à travers toutes les vues et permet d'ajouter facilement de nouvelles opérations de traitement ou de nouvelles visualisations.

Ce document détaille les choix méthodologiques, la préparation des datasets, le fonctionnement de la pipeline de données, ainsi que les caractéristiques et interactions de chaque type de visualisation implémentée.

## Préparation des datasets

Pour cette visualisation, nous nous sommes basés sur les deux datasets suivants:

- Youtube 2025 channels : Ce dataset fournit des informations détaillées sur un grand nombre de chaînes YouTube (plus de 15 000), combinant à la fois des métadonnées statiques et des mesures d'engagement. Chaque chaîne comprend des attributs tels que le nom, l'ID, la catégorie, le pays, la date de création, le

nombre d'abonnés, le nombre total de vues, la langue par défaut, la description, l'URL personnalisée, la vignette et le nombre total de vidéos. C'est donc sur lui que se basent quasiment toutes nos visualisations.

- YouTube Top Channel Videos (2025): Ce dataset contient des informations sur les 10 dernières vidéos mises en ligne par chaque chaîne YouTube, capturées début octobre 2025. Chaque entrée correspond à une seule vidéo et comprend des informations telles que le titre de la vidéo, son id, le nombre de vues au moment de la collecte des données, la chaîne à laquelle elle appartient, etc. Ce dataset est utilisé uniquement dans le bubble chart.

Nous avons ensuite nettoyé les deux datasets en retirant les données nulles. Toutefois, les chaînes avec un pays non défini par l'utilisateur ont été conservées car elles constituent une source d'information pertinente : l'absence de pays indique simplement que le créateur n'a pas renseigné cette information, ce qui est fréquent pour certains types de chaînes comme les chaînes de musique internationales ou les labels, qui représentent souvent des volumes d'abonnés significatifs.

Pour le bubble chart, nous avons effectué une jointure sur le champ `channel_id` afin d'associer chaque vidéo à sa chaîne YouTube correspondante. Cette opération permet d'afficher les 10 dernières vidéos publiées par chaque YouTubeur lorsque l'utilisateur clique sur une chaîne spécifique dans la visualisation.

Nous avons également utilisé un jeu de données de coordonnées géographiques pour la visualisation sur une carte. Ce jeu de données, au format GeoJSON, contient les coordonnées des contours de tous les pays, associées à leur code ISO-2, ISO-3, ainsi qu'au nom du pays.

Le jeu de données contenant les chaînes YouTube associe le code ISO-2 des pays aux chaînes, ce qui permet de joindre les deux jeux de données. Une zone rectangulaire a été ajoutée dans le sud de l'océan Pacifique afin de pouvoir afficher sur la carte les chaînes qui n'ont pas de pays défini, permettant ainsi leur visualisation malgré tout.

Enfin, un dernier jeu de données a été associé aux autres. Il place un point géographique (latitude/longitude) pour chaque chaîne YouTube dans son pays. Ainsi, pour celles dont le pays n'est pas défini, elles sont placées à l'intérieur du rectangle présent dans le Pacifique définit plus tôt.

### Type des datasets

Les datasets `Youtube 2025 channels` et `YouTube Top Channel Videos (2025)` sont des tables.

### Type des données

#### Youtube 2025 channels

Attribute	Class of data
<code>channel_name</code>	Categorical
<code>channel_id</code>	Categorical
<code>view_count</code>	Quantitative
<code>category</code>	Categorical
<code>country</code>	Categorical
<code>defaultLanguage</code>	Categorical
<code>subscriber_count</code>	Quantitative
<code>created_date</code>	Categorical
<code>description</code>	Categorical
<code>custom_url</code>	Categorical
<code>thumbnail</code>	Categorical
<code>video_count</code>	Quantitative
<code>videos_last_30_days</code>	Quantitative
<code>views_last_30_days</code>	Quantitative
<code>uploads_playlist_id</code>	Categorical
<code>last_10_video_ids</code>	Categorical

## YouTube Top Channel Videos

Attribute	Class of data
video_id	Categorical
title	Categorical
channel_name	Categorical
channel_id	Categorical
view_count	Quantitative
like_count	Quantitative
comment_count	Quantitative
published_date	Categorical
thumbnail	Categorical

## Persona

Pour ce projet, les deux personas principaux sont les profils cibles suivants : les créateurs de contenu et les journalistes / analystes spécialisés dans l'univers de YouTube. Ces personas ont guidé les choix de visualisation, les types d'interactions proposées et les indicateurs mis en avant dans l'interface.

### Le créateur de contenu

Le créateur de contenu est un utilisateur actif de YouTube, gérant une ou plusieurs chaînes. Il est attentif à ses statistiques et souhaite comprendre son positionnement dans un écosystème de plus en plus compétitif. Son objectif principal est de comparer ses performances à celles d'autres créateurs, d'identifier les catégories les plus populaires et de repérer des opportunités de croissance.

Ses objectifs sont:

- Identifier les catégories les plus populaires dans son pays ou à l'échelle mondiale.
- Connaître le pourcentage de vues et la répartition des abonnés dans sa catégorie.
- Observer des métriques clés sur sa propre chaîne : nombre d'abonnés, vues totales, fréquence de publication, productivité.
- Se comparer à d'autres créateurs de sa catégorie ou de sa région.
- Visualiser sa position géographique et celle de ses concurrents potentiels via la carte interactive.
- Suivre les tendances émergentes afin d'adapter sa stratégie éditoriale.

Pour cela, il pourra utiliser les visualisations suivantes:

- Treemap : pour explorer la hiérarchie pays → catégorie → chaîne et situer sa propre chaîne parmi ses pairs.
- Bubble chart : pour comparer les abonnés, vues et vidéos à différentes échelles.
- Pie chart : pour comprendre la répartition des catégories et leur poids relatif.
- Histogramme : pour analyser le rapport entre abonnés et vues moyennes, et identifier les chaînes les plus performantes en termes d'engagement.
- Carte interactive : pour visualiser sa position géographique et celle de ses concurrents potentiels, et se comparer à d'autres créateurs de son pays via le classement du nombre d'abonnés.

## Le journaliste / analyste spécialisé

Le journaliste ou analyste est un professionnel des médias ou de la data, intéressé par l'évolution du paysage numérique et la mesure de l'influence des créateurs sur YouTube. Il utilise la plateforme comme un outil d'analyse et de storytelling, afin de produire des articles, des classements ou des rapports sur les tendances globales.

Ses objectifs sont:

- Obtenir une vue d'ensemble du paysage YouTube mondial.
- Étudier les dynamiques par catégorie ou par région.
- Identifier les pays concentrant le plus de grands créateurs.
- Produire des classements ou des comparaisons entre pays, catégories ou types de chaînes.
- Observer les corrélations entre abonnés, vues et productivité.
- Exporter des insights visuels pour enrichir des articles ou présentations.

Pour cela, il utilisera les visualisations suivantes:

- Carte interactive : pour localiser les grands créateurs et observer leur répartition géographique.
- Treemap : pour hiérarchiser les pays, catégories et chaînes selon le nombre d'abonnés.
- Histogramme : pour étudier la relation entre abonnés et vues moyennes.
- Pie chart : pour visualiser la distribution des catégories au sein d'un pays ou d'une région.

## Pipeline

La pipeline de données a été conçue pour faciliter la préparation et le nettoyage des données avant leur visualisation, pour que les filtres soient communs à toutes les visualisations.

L'idée principale est de pouvoir enchaîner dynamiquement des opérations de traitement (filtrage, tri, agrégation, jointure géographique, etc.) tout en conservant un code clair, flexible et extensible.

La pipeline repose sur deux éléments principaux :

- Les données sources (data) : un tableau d'objets initial représentant les observations ou entités à traiter.
- Les opérations (operations) : une collection ordonnée d'opérations stockées dans une Map, chacune identifiée par un nom unique.

## Fonctionnement étape par étape

**1. Chargement des données** La méthode `load(path, type)` permet de charger un dataset depuis un fichier (par exemple un `.csv`) en utilisant `D3.js`.

Les données sont alors stockées dans `data` pour être traitées ensuite.

**2. Ajout d'opérations** Les différentes méthodes de la classe ajoutent des étapes de traitement à la pipeline via `addOperation(name, op)` :

- `filter(name, fn)` : filtre les lignes selon une fonction conditionnelle.
- `limit(name, n)` : limite le nombre d'éléments.

- `map(name, fn)` : applique une transformation à chaque élément.
- `sortBy(name, key, ascending)` : trie les données selon une clé donnée.
- `groupBy(name, key)` : regroupe les données selon une clé.
- `aggregate(name, reducer)` : agrège les groupes créés avec une fonction de réduction (moyenne, somme, etc.).
- `convertMap(name, keyField)` : permet de convertir un tableau d'objets en `Map`, en utilisant comme **clé** la valeur du champ `keyField`.

Chaque opération est nommée pour pouvoir être supprimée ou ignorée plus tard.

**3. Exécution** La méthode `run()` applique toutes les opérations ajoutées, dans leur ordre d'insertion. Il est possible d'exclure certaines opérations en précisant leurs noms.

Le résultat final est un ensemble de données prêtes à être utilisées dans les visualisations (tree map, bubble charts, etc.).

## Types de visualisations

### Tree visualisation

La Treemap permet de représenter les chaînes YouTube sous forme de rectangles imbriqués, offrant une vue hiérarchique et proportionnelle des données.

Il s'agit d'une visualisation à plusieurs niveaux d'exploration :

Pays → Catégories → Chaînes YouTube.

Chaque rectangle correspond à une entité, et sa surface est proportionnelle à une métrique donnée :

- Au niveau pays, la surface reflète le nombre total d'abonnés cumulés de toutes les chaînes du pays.
- Au niveau catégorie, elle représente le nombre total d'abonnés des chaînes appartenant à cette catégorie dans le pays sélectionné.
- Au niveau chaîne, elle correspond au nombre d'abonnés de la chaîne individuelle.

Cette visualisation intègre plusieurs mécanismes d'interaction permettant à l'utilisateur d'explorer les données de manière fluide et intuitive :

- **Navigation hiérarchique** : L'utilisateur peut cliquer sur un rectangle pour zoomer et passer à un niveau de détail inférieur :
  - En cliquant sur un pays, on découvre les catégories de chaînes présentes dans ce pays.
  - En cliquant sur une catégorie, on accède aux chaînes YouTube individuelles appartenant à cette catégorie.
  - En cliquant sur une chaîne, on est redirigé vers sa page YouTube.
  - Le bouton “← Retour” permet de remonter d'un niveau à tout moment.
- **Informations contextuelles** : En survolant un rectangle, un tooltip dynamique affiche des informations détaillées sur l'élément sélectionné, sans surcharger la vue principale:
  - Pour un pays : le nom complet du pays, le nombre total de YouTubeurs et le nombre total d'abonnés.
  - Pour une catégorie : le nom de la catégorie, le nombre de chaînes et le nombre total d'abonnés.

- Pour une chaîne : le nom, la catégorie et le nombre d'abonnés.
- **Effets visuels** : Au survol, les rectangles changent d'opacité pour améliorer la lisibilité et l'interactivité.
- **Transitions animées**: Les changements de niveau utilisent des transitions fluides qui morphent les rectangles de leur position précédente à leur nouvelle position, offrant une continuité visuelle et une meilleure compréhension de la navigation.
- **Adaptation du contenu**: Les labels et les informations textuelles (nom, nombre d'abonnés) s'adaptent automatiquement à la taille des rectangles, avec troncature si nécessaire pour éviter les débordements.
- **Représentation visuelle des chaînes** : Au niveau le plus détaillé (chaînes individuelles), la miniature de chaque chaîne YouTube est affichée comme fond du rectangle, avec un overlay semi-transparent pour conserver la lisibilité.

## Bubble chart

Le Bubble chart permet de représenter les chaînes YouTube et leurs vidéos sous forme de cercles proportionnels à une métrique donnée.

Il s'agit d'une visualisation hiérarchique et interactive qui offre plusieurs niveaux d'exploration : Pays → Chaînes YouTube → Vidéos.

Chaque bulle correspond à une entité, et sa taille dépend d'une mesure quantitative :

- Au niveau pays, la taille reflète le nombre total de chaînes provenant de ce pays.
- Au niveau chaîne, elle correspond au nombre d'abonnés.
- Au niveau vidéo, elle représente le nombre de vues.

Cette visualisation intègre plusieurs mécanismes d'interaction permettant à l'utilisateur d'explorer les données de manière fluide et intuitive :

- L'utilisateur peut cliquer sur une bulle pour zoomer et passer à un niveau de détail inférieur :
  - En cliquant sur un pays, on découvre les chaînes YouTube associées.
  - En cliquant sur une chaîne, on accède à ses vidéos récentes.
  - Le bouton “Retour” permet de remonter d'un niveau à tout moment.
- En survolant une bulle, un tooltip dynamique affiche des informations détaillées sur l'élément sélectionné, sans surcharger la vue principale :
  - Pour un pays : le nombre total de chaînes.
  - Pour une chaîne : le nom, la catégorie et le nombre d'abonnés.
  - Pour une vidéo : le titre, le nombre de vues, de likes et de commentaires.

Au niveau vidéo, un clic sur une bulle ouvre directement la vidéo correspondante sur YouTube dans un nouvel onglet.

Les chaînes et vidéos sont affichées à l'aide de leurs miniatures (thumbnails) intégrées au centre des bulles.

## Traitement des données

Pour cette visualisation, les deux jeux de données ont été utilisés conjointement :

- Le dataset “YouTube 2025 channels”, contenant les métadonnées des chaînes (pays, catégorie, abonnés, etc.).
- Le dataset “YouTube Top Channel Videos (2025)”, qui recense les 10 dernières vidéos par chaîne.

Une jointure sur le champ channel\_id a été réalisée pour relier chaque vidéo à sa chaîne. Les données ont ensuite été agrégées et filtrées dynamiquement, permettant de mettre à jour le Bubble chart en fonction des filtres sélectionnés (pays, catégories, etc.).

## Histogram

L’Histogramme permet de représenter l’engagement des chaînes YouTube sous forme de barres verticales, offrant une comparaison visuelle directe des performances relatives des différentes chaînes.

La métrique utilisée est le ratio entre les vues totales et le nombre d’abonnés (view\_count / subscriber\_count). Ce ratio permet d’évaluer l’efficacité d’une chaîne à générer des vues par rapport à sa base d’abonnés : un ratio élevé indique que la chaîne génère beaucoup de vues relativement à son nombre d’abonnés, ce qui peut refléter un contenu viral, une forte fidélisation, ou une audience très active.

La visualisation se décline en deux vues complémentaires :

**Vue principale** : Affiche les 50 chaînes avec le meilleur ratio d’engagement, classées par ordre décroissant. Chaque barre représente une chaîne, et sa hauteur correspond à son ratio d’engagement.

**Vue comparative (zoom)** : En cliquant sur une barre, l’utilisateur accède à une vue comparative qui présente trois barres côté à côté :

- La barre bleue représente le ratio d’engagement de la chaîne sélectionnée.
- La barre verte représente le ratio moyen de toutes les chaînes du même pays.
- La barre orange représente le ratio moyen de toutes les chaînes appartenant à la même catégorie principale.

Cette comparaison contextuelle permet d’évaluer la performance d’une chaîne non seulement en valeur absolue, mais aussi relativement à son contexte géographique et thématique.

Cette visualisation intègre plusieurs mécanismes d’interaction permettant à l’utilisateur d’explorer les données de manière fluide et intuitive :

- **Informations contextuelles** : En survolant une barre, un tooltip dynamique affiche des informations détaillées sur l’élément sélectionné.
- **Transitions animées** : Les changements de vue (passage de la vue principale à la vue comparative et vice-versa) utilisent des transitions fluides qui animent les barres de leur position précédente à leur nouvelle position, offrant une continuité visuelle et une meilleure compréhension de la navigation.
- **Adaptation au filtrage** : Lorsque l’utilisateur applique ou réinitialise des filtres alors qu’il se trouve dans la vue comparative, la visualisation retourne automatiquement à la vue principale pour refléter les nouvelles données filtrées. Cette logique garantit la cohérence entre les filtres appliqués et les données affichées.
- **Encodage visuel** : Les trois barres de la vue comparative utilisent des couleurs distinctes (bleu, vert, orange) pour faciliter l’identification immédiate de chaque type de comparaison. Les valeurs numériques sont affichées au-dessus de chaque barre pour une lecture précise des ratios.
- **Échelle dynamique** : L’axe vertical s’ajuste automatiquement en fonction des données affichées, garantissant une utilisation optimale de l’espace de visualisation et une meilleure lisibilité des différences entre les valeurs.

## Carte interactive

La carte permet de visualiser la répartition géographique des chaînes YouTube à travers le monde, tout en comparant les pays d’un seul coup d’œil selon différents indicateurs, tels que le nombre maximal d’abonnés

pour une chaîne, le nombre moyen d'abonnés, le nombre de chaînes ou encore le total de vidéos.

#### Traitement des données :

Pour la carte, il a été nécessaire de combiner trois sources de données : Le dataset “YouTube 2025 channels”, contenant les métadonnées des chaînes (pays, catégorie, abonnés, etc.), un GeoJSON contenant la forme des pays, et un dernier dataset avec la position géographique des chaînes.

En utilisant le pipeline de données commun, garantissant la cohérence avec les autres visualisations (mêmes filtres), les données sont **agrégées par pays** à partir du dataset “YouTube 2025 channels” afin d'obtenir les métriques pour chaque pays et pour le monde :

- Nombre total de chaînes
- Moyenne et total d'abonnés
- Moyenne et total de vidéos
- Plus grande chaîne du pays (en nombre d'abonnés)

Les bornes maximales et minimales de chaque métrique sont ensuite récupérées afin de mettre à jour dynamiquement la légende.

#### Fonctionnalités et interactions :

- L'utilisateur peut choisir la métrique à afficher (par exemple : *abonnés moyens par chaîne, total d'abonnés, total de vidéos, plus grosse chaîne*, etc.) via un sélecteur. Le filtre des pays est désactivé car il n'est pas nécessaire sur la carte.
- Chaque pays est coloré en fonction de la valeur de la métrique sélectionnée, selon une échelle allant du blanc (valeur faible) au rouge foncé (valeur élevée). Une légende dynamique en bas à droite indique la correspondance entre les couleurs et les intervalles de valeurs. L'utilisateur peut cliquer sur les plages de couleurs pour masquer ou afficher certaines, afin de se concentrer sur celles qu'il souhaite analyser.
- Un panneau latéral s'affiche lorsqu'on clique sur un pays et présente le classement des chaînes principales sous forme d'histogramme vertical.
- Au survol d'un pays, un tooltip dynamique affiche les statistiques principales :
  - Nombre total de chaînes
  - Moyenne et total d'abonnés
  - Plus grande chaîne du pays
  - Moyenne et total de vidéos
- En l'absence de chaînes pour un pays, un hachurage diagonal apparaît pour signaler visuellement l'absence de données.
- En zoomant sur la carte au-delà d'un certain seuil, des marqueurs apparaissent pour indiquer la position géographique des chaînes YouTube. Au survol d'un marqueur, un tooltip affiche le nom de la chaîne, ses abonnés, ses vidéos, ses catégories et son pays.
- Lorsque l'utilisateur survole une zone sans interaction directe avec un pays, le tooltip affiche les statistiques mondiales. En cas de clic, le panneau latéral présente le classement global des chaînes.

Les variables visuelles principalement utilisées dans cette visualisation sont la couleur, la position et les hachures. L'utilisateur peut interagir avec la carte via le survol, le clic et le zoom.

Cette visualisation illustre parfaitement les principes de Shneiderman :

- Vue d'ensemble : coloration des pays selon différents indicateurs.
- Zoom et exploration : localisation détaillée des chaînes au sein de chaque pays.
- Détails à la demande : tooltips affichant les statistiques précises.
- Filtrage et sélection : possibilité de sélectionner et filtrer les données pour affiner l'analyse.

### Pie chart

Le Pie chart permet de visualiser la répartition des catégories de chaînes YouTube, soit à l'échelle mondiale, soit pour un pays donné. Il offre une vue synthétique du poids relatif de chaque catégorie suivant l'échelle choisie, en fonction du nombre de vues cumulées.

#### Fonctionnalités et interactions :

- L'utilisateur peut filtrer par pays et/ou par catégorie via le panneau de filtres commun à toutes les visualisations.
- Si un seul pays est sélectionné, le pie chart affiche la répartition des catégories pour ce pays. Si plusieurs pays sont sélectionnés ou aucun, la vue « Monde » est affichée.
- Si l'utilisateur sélectionne une ou plusieurs catégories, le diagramme s'adapte :
  - Si une seule catégorie est sélectionnée, le pie chart affiche la répartition des chaînes principales de cette catégorie (par nombre de vues).
  - Si plusieurs catégories sont sélectionnées, le pie chart affiche successivement la répartition des chaînes pour chaque catégorie choisie.
- Un message d'erreur s'affiche si aucun résultat ne correspond aux filtres appliqués.
- Un fil d'Ariane permet de naviguer facilement entre les niveaux (Monde → Pays → Catégorie).
- Au survol d'une portion du pie chart, un tooltip dynamique affiche le nom de la catégorie, le pourcentage de vues et le nombre total de vues.
- Le pie chart s'adapte dynamiquement à la taille de l'écran et au nombre de catégories (regroupement automatique des plus petites catégories sous « Autres »). Il ne peut avoir que 10 valeurs différentes à la fois pour éviter une confusion avec les couleurs.
- Si on clique sur une catégorie (que l'on soit en vue Monde ou pays), cela affiche les chaînes youtube correspondantes à la vue et la catégorie.

#### Traitement des données :

- Le Pie chart utilise le pipeline de données commun, garantissant que tous les filtres (pays, catégories, abonnés, vidéos, dates) sont appliqués de façon cohérente avec les autres visualisations.
- L'agrégation principale se fait sur le nombre de vues par catégorie, puis par chaîne lors du drilldown.

Cette visualisation permet d'identifier rapidement les catégories dominantes, de comparer leur poids relatif, et d'explorer le détail des chaînes qui composent chaque catégorie, à l'échelle mondiale ou nationale.

## Démarrage du projet

Pour lancer le projet:

1. Récupérez le projet en clonant le repository sur votre machine locale.

```
git clone https://github.com/LubratJilian/Project-Information-Visualisation.git
```

2. Installez les dépendances nécessaires à partir des fichiers package.json du frontend et du backend.

```
cd frontend  
npm install  
  
cd backend  
npm install
```

3. Rendez-vous dans le dossier backend puis exécutez la commande :

```
node index.js
```

4. Ouvrez votre navigateur à l'adresse suivante : <http://localhost:3000>

Si vous souhaitez remplacer les jeux de données, ils se trouvent dans le dossier **frontend/data/** : le fichier **youtube.csv** correspond aux chaînes, et **youtube\_videos.csv** aux vidéos.