

Data Visualization :

Rapport du projet:

Analyse et Visualisation de Données

Réalisé par :

Ghazouani Farah

1 MP DSB

2024-2025

Introduction

Dans un monde de plus en plus orienté vers les données, la capacité à explorer, analyser et représenter visuellement l'information est devenue essentielle pour produire du sens et faciliter la prise de décision. C'est dans ce cadre que s'inscrit le présent projet, réalisé dans le cadre du module **Data Visualisation**, qui vise à mobiliser des compétences analytiques et techniques autour de plusieurs jeux de données.

Le travail s'articule en deux volets complémentaires. Le premier consiste à analyser des jeux de données variés à l'aide de différents outils de visualisation (Python, Tableau Public, D3.js, Gephi), en concevant des représentations graphiques pertinentes et en interprétant les principales tendances observées. Ce volet inclut également une réflexion critique sur les mauvaises visualisations et leurs effets potentiels sur l'interprétation des données.

Le second volet porte sur la représentation cartographique de données thématiques relatives à la Tunisie. En s'appuyant sur des fichiers géographiques (GeoJSON) et des jeux de données régionaux et sectoriels, il s'agit de produire une carte thématique claire et significative, mettant en valeur des dynamiques territoriales comme la répartition des zones industrielles, des barrages ou des investissements.

Ce rapport détaille la démarche suivie, les outils mobilisés, les choix méthodologiques ainsi que les enseignements tirés de chaque visualisation.

1. Cartographie thématique de la Tunisie

1.1. L'utilité de l'exploration des données de la Tunisie avec Python

Dans cette étape, nous allons explorer un ensemble de jeux de données géographiques et statistiques liés à la Tunisie, à l'aide de l'outil Python, notamment : 'barrages_collinaires_beja.geojson', 'CollecteGraineBeja.geojson', 'lacs_collinaires_beja.geojson', 'man.xlsx', 'manpartgvt.xlsx', 'manpartsecteur.xlsx', 'mansecteur.xlsx', 'map3.geojson', 'ncirconscriptions.geojson', 'zonesindustrielles.geojson', 'z_interv_bizerte_pno4.geojson. Afin d'identifier les informations pertinentes et de mieux comprendre la structure des données, en vue de préparer efficacement l'étape suivante : la visualisation cartographique thématique de la Tunisie.

1.2. Analyse de fichier Excel man.xlsx

Le dataset man.xlsx comprend les colonnes suivantes : **Year**, **Adhésions**, **Approbations**, **Investissement (en DT)**, **Primes (en DT)**.

Nous avons choisi de tracer les courbes de l'évolution des **adhésions** et des **approbations** par année afin d'observer leur distribution temporelle. Ces deux indicateurs clés permettent d'évaluer la performance et la progression du système étudié.

Pour ce faire, nous avons nettoyé les données, notamment en convertissant la colonne **Year**, qui contient parfois des périodes (ex. : « 1996 - 2000 »), en années numériques exploitables.

```
# Traitement de La colonne "Year"
def convert_year(value):
    """ Prend la première année dans une période '1996 - 2000' ou retourne l'année entière si c'est un nombre. """
    if isinstance(value, str):
        return int(value.split(" - ")[0]) # Prendre La première année
    return value # Sinon, garder la valeur telle quelle

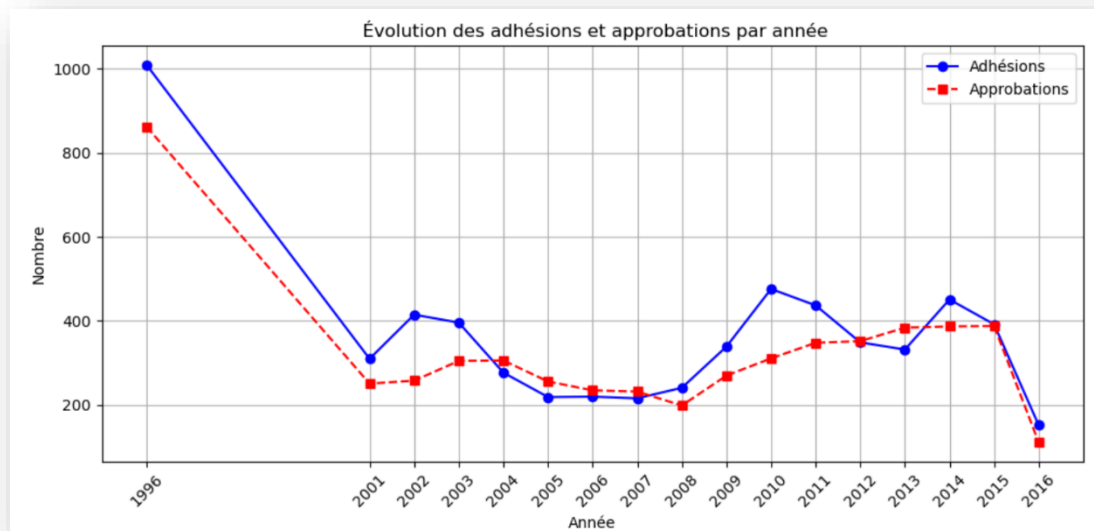
df["Year"] = df["Year"].apply(convert_year)
```

Par ailleurs, nous avons vérifié l'absence de valeurs manquantes avant de tracer les deux courbes de l'évolution des adhésions et approbations par année, à l'aide de Matplotlib pour visualiser les tendances temporelles de manière claire et comparative.

```
# Visualisation : évolution des adhésions et des approbations par année
plt.figure(figsize=(10, 5))
plt.plot(df["Year"], df["Adhésions"], marker="o", label="Adhésions", linestyle="-", color="blue")
plt.plot(df["Year"], df["Approbations"], marker="s", label="Approbations", linestyle="--", color="red")

plt.title("Évolution des adhésions et approbations par année")
plt.xlabel("Année")
plt.ylabel("Nombre")
plt.xticks(df["Year"].unique(), rotation=45) # Affiche seulement les années existantes
plt.legend()
plt.grid(True)
plt.tight_layout() # Ajuste automatiquement les marges du graphique
plt.show()
```

La courbe des adhésions (en bleu) montre un **pic majeur** durant la période 1996–2000 (près de 1 000 adhésions), suivi d’une **chute brutale** autour de 2001–2002, puis d’une remontée progressive jusqu’à un second pic vers 2010. Après un léger repli, on note un nouvel accroissement autour de 2014, avant une chute en 2016. Les approbations (en rouge, pointillé) suivent globalement la même tendance, mais avec un **décalage systématique à la baisse** : chaque pic d’adhésions est suivi d’un pic plus modéré d’approbations, ce qui peut indiquer un processus sélectif ou un délai entre le moment où de nouveaux membres rejoignent et celui où leurs dossiers sont validés.



1.3. Analyse du fichier manpartgvt.xlsx

Le dataset **manpartgvt** se compose des colonnes suivantes : **GVT** (correspondant aux différents gouvernorats), **dossiers approuvés**, **investissements exprimés en millions de dinars (MD)**, **primes également exprimées en millions de dinars (MD)**.

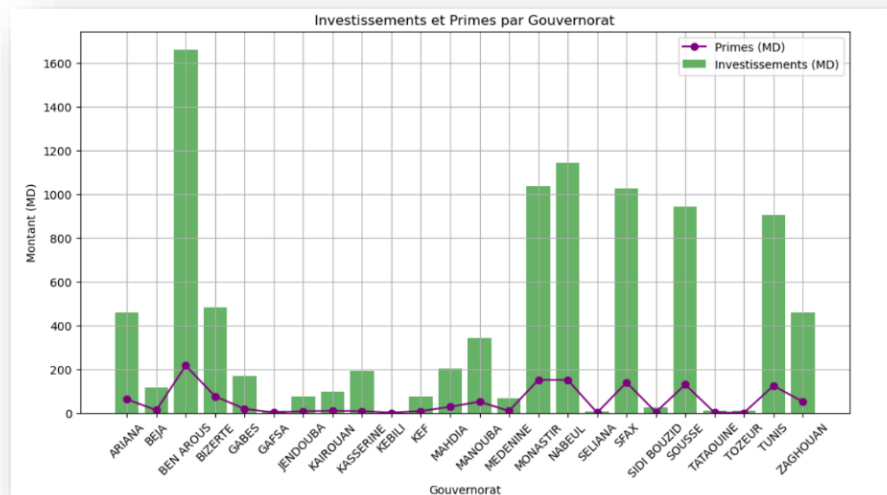
Après avoir vérifié l'absence de valeurs manquantes, nous avons généré une visualisation combinée : un histogramme représentant les investissements (en millions de dinars) par gouvernorat et une courbe ponctuée illustrant les primes allouées dans le but de comparer visuellement la répartition des financements, investissements et primes à travers les différentes régions de la Tunisie.

```
# 🇹🇳 Visualisation combinée : barres pour Les investissements et ligne pour Les primes
plt.figure(figsize=(12, 6))

# Histogramme des investissements par gouvernorat
plt.bar(
    df_gvt["GVT"], # Gouvernorats
    df_gvt["Investissements en MD"], # Montant des investissements
    color="green",
    alpha=0.6,
    label="Investissements (MD)"
)

# Courbe des primes sur Les mêmes gouvernorats
plt.plot(
    df_gvt["GVT"],
    df_gvt["Primes en MD"],
    marker="o",
    linestyle="--",
    color="purple",
    label="Primes (MD)"
)
```

On observe que la majorité des financements (en vert) se concentrent sur quelques gouvernorats clés : **Ben Arous**, **Monastir**, **Nabeul**, **Sfax** et **Tunis** dépassent les 1 000 MD d'investissements, tandis que la plupart des autres régions reçoivent moins de 500 MD. Les primes (en violet) restent globalement faibles en comparaison, avec un pic relatif à Béja et un deuxième petit pic à Sidi Bouzid, suggérant que l'octroi de primes est plus ciblé et moins volumineux que les investissements, probablement pour soutenir des projets spécifiques ou des secteurs prioritaires dans certaines régions.



1.4. Analyse du fichier manpartsecteur.xlsx

Le fichier **manpartsecteur** comprend les colonnes suivantes : **Secteur** (type d'activité économique), **total adhésion** (nombre d'adhésions par secteur), **dossiers approuvés** (nombre de dossiers validés), **investissements en MD** (montants investis en millions de dinars).

Après le chargement du fichier et la vérification de l'absence de valeurs manquantes, une visualisation combinée est générée : un histogramme représente les **investissements** (en millions de dinars) par secteur, tandis qu'une ligne ponctuée illustre les **primes octroyées** pour comparer visuellement la répartition des financements selon les secteurs d'activité en Tunisie.

```
# Visualisation : investissements et primes par secteur
plt.figure(figsize=(12, 6))

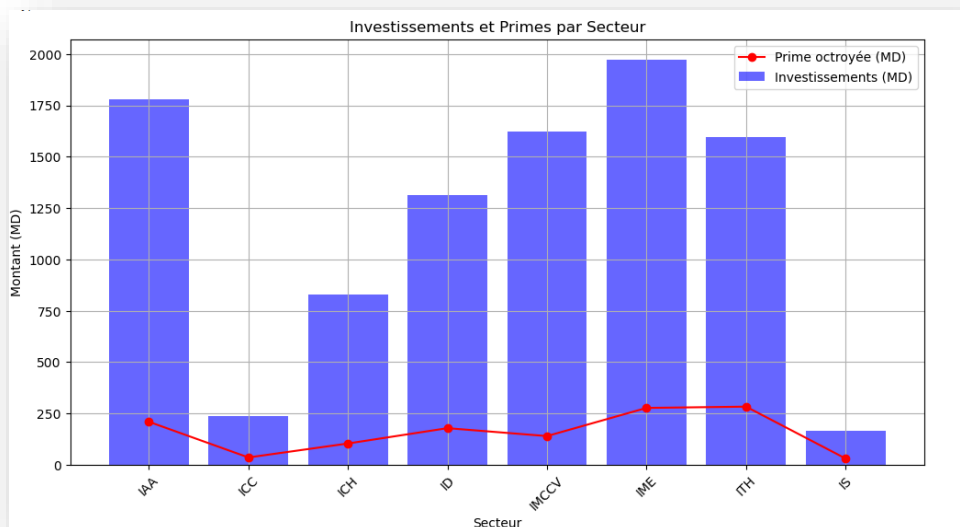
# Histogramme des investissements
plt.bar(
    df_secteur["Secteur"],
    df_secteur["Investissements en MD"],
    color="blue",
    alpha=0.6,
    label="Investissements (MD)"
)

# Courbe des primes
plt.plot(
    df_secteur["Secteur"],
    df_secteur["Prime octroyée en MD"],
    marker="o",
    linestyle="-",
    color="red",
    label="Prime octroyée (MD)"
)
```

Ce graphe montre que le secteur **IME** (Industries Mécaniques et Électriques) arrive en tête avec près de **2000 MD**. Il est suivi par **IAA** (Industries Agro-Alimentaires), **IMCCV** (Matériaux de Construction, Céramique et Verre) et **ITH** (Industries Textiles et Habillement). Ce sont les secteurs **les plus investis**.

Le secteur **IS** (Services) et **ICC** (Commerce) enregistrent les montants les plus faibles **Secteurs avec moins d'investissements**.

Pour les primes, ils suivent une tendance moins marquée mais montrent un soutien important dans les secteurs les plus investis, notamment **IME** et **IAA**. Toutefois, il existe un écart entre l'investissement et la prime, indiquant une part de financement privé importante.



⇒ Les investissements sont principalement orientés vers les secteurs industriels lourds, particulièrement les secteurs mécaniques, agroalimentaires et matériaux. Ces secteurs reçoivent également une part significative de primes, ce qui reflète une politique d'encouragement ciblée.

1.5. Analyse du fichier mansecteur.xlsx

Ce fichier contient les informations suivantes : **Secteur** (désigne le domaine d'activité économique), **total adhésion** : (nombre d'adhésions enregistrées dans chaque secteur), **dossiers approuvés** (nombre de dossiers ayant reçu un avis favorable), **investissements en MD** (montants investis dans chaque secteur), **prime octroyée en MD** (montants des primes accordées). Après le chargement des données, nous avons nettoyé les noms de colonnes (suppression des espaces) et convertit la colonne **Year** en chaîne de caractères pour éviter d'éventuelles erreurs.

```
# 🧹 Nettoyer les noms des colonnes (supprimer espaces)
df_secteur.columns = df_secteur.columns.str.strip()

# 📅 Convertir la colonne "Year" en chaîne de caractères (pour éviter erreurs)
df_secteur["Year"] = df_secteur["Year"].astype(str)
```

Une fois les valeurs manquantes vérifiées, nous avons généré un graphique montrant l'évolution des investissements par secteur au fil des années. Chaque courbe représente un secteur, ce qui permet de comparer visuellement la dynamique d'investissement dans le temps.

Ce graphique montre l'évolution annuelle des montants investis par secteur entre **1996 et 2016**. On remarque que les investissements fluctuent fortement d'une année à l'autre dans la majorité des secteurs, ce qui peut être dû à des facteurs économiques, politiques ou stratégiques.

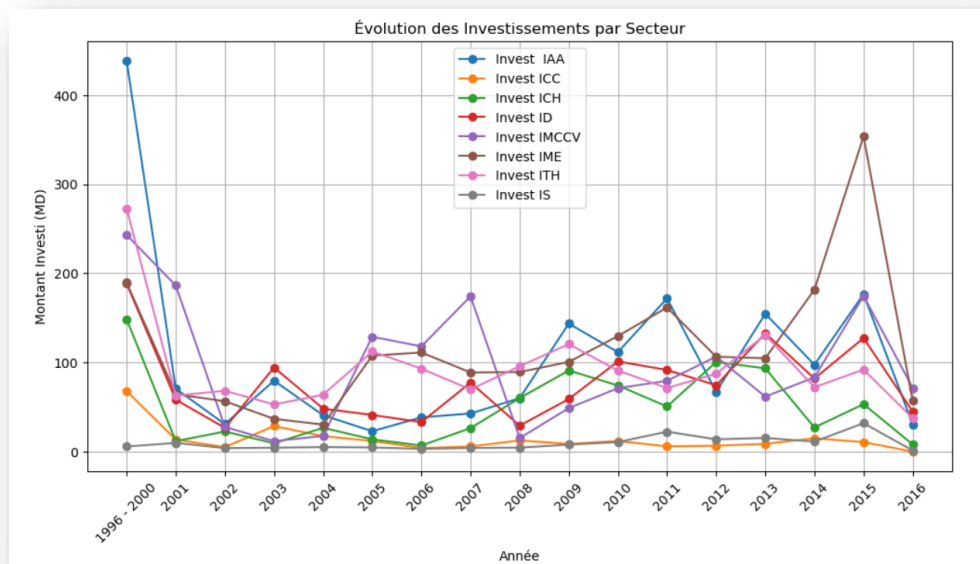
L'**IAA** connaît un pic important en **1996–2000**, suivi d'une baisse, puis d'une reprise stable.

L'**IME** se démarque fortement en **2015**, indiquant une année exceptionnelle pour ce secteur.

L'**IMCCV**, l'**ITH** et l'**ID** ont également connu plusieurs hausses notables, notamment en **2007** et **2015**.

Pour les investissements dans **IS** (Services) et **ICC**, ils restent faibles et relativement constants.

Et finalement, pour le secteur **ICH**, il présente une légère croissance mais sans pics majeurs.



⇒ L'analyse temporelle confirme l'intérêt soutenu pour les secteurs industriels, notamment les industries mécaniques et agroalimentaires. Les pics ponctuels peuvent être attribués à des projets structurants ou à des incitations gouvernementales spécifiques. À l'inverse, les services et le commerce semblent bénéficier de moins d'attention en termes d'investissements.

1.6. Analyse du fichier lacs_collinaires_beja.geojson

Nous avons analysé le fichier **GeoJSON** contenant les lacs collinaires de Béja (avec les colonnes suivantes **name** (nom ou code du barrage), **X_COORD** / **Y_COORD** (coordonnées géographiques du barrage), **Nom_ar** / **Nom_fr** (nom du barrage en arabe et en français, **Annee_de_creation** (année de mise en service du barrage), **Apport_mille_m3** (capacité d'apport en milliers de mètres cubes), **Cout_mille_DT** (coût de réalisation en milliers de dinars tunisiens), **Secteur** / **Delegation** (localisation administrative du barrage), **geometry** (géométrie spatiale du barrage).

Nous avons affiché le **nombre total de lacs** présents et vérifié l'existence de deux colonnes : **IHPLC_ANN** (année de création) pour en extraire la **première et dernière année de construction** et **IHPLC_CAP** (capacité en mille m³) afin de calculer et afficher la **capacité moyenne** des lacs.

```
# Étape 1 : Statistiques générales sur les données
# Nombre total de lacs collinaires présents dans le fichier GeoJSON
print("\n📌 Nombre total de lacs collinaires :", len(gdf))

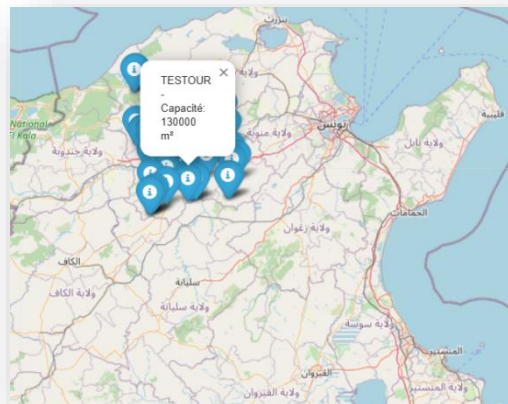
# Étape 2 : Affichage des statistiques sur l'année de création des lacs (colonne 'IHPLC_ANN')
print("📌 Année de création des lacs : min =", gdf["IHPLC_ANN"].min(), ", max =", gdf["IHPLC_ANN"].max())

# Étape 3 : Affichage des statistiques sur la capacité des lacs (colonne 'IHPLC_CAP' - capacité en milliers de m³)
print("📌 Capacité des lacs (en milliers de m³) - Moyenne :", gdf["IHPLC_CAP"].mean())
```

Une première exploration rapide des caractéristiques des lacs collinaires.

- ✦ Nombre total de barrages : 59
- ✦ Année de création : min = 1988 , max = 2011
- ✦ Capacité (mille m³) - Moyenne : 125796.86440677966

Finalement, nous avons généré une **carte interactive** avec Folium, où chaque lac est représenté par un **marqueur** indiquant sa délégation et sa **capacité de stockage (en m³)**.



Cette visualisation permet de mieux comprendre la répartition géographique et les capacités des lacs dans la région.

1.7. Analyse du fichier CollecteGraineBeja.geojson

Ce fichier **GeoJSON** représente les points de collecte de graines à Béja. Les colonnes clés sont : (**Name** (nom de la coopérative ou du centre de collecte, généralement en arabe ou bilingue), **description**, **timestamp**, **begin**, **end** (métadonnées temporelles), **altitudeMode**, **tessellate**, **extrude**, **visibility**, **drawOrder**, **icon** (paramètres de visualisation), **geometry** (coordonnées géographiques du point de collecte).

Nous avons commencé par **remplacer les valeurs manquantes** dans plusieurs colonnes (textuelles ou numériques) par leur **valeur la plus fréquente** (ou "Inconnu"/0 si non disponible).

✓ Valeurs manquantes avant traitement :

Name	0
description	8
timestamp	8
begin	8
end	8
altitudeMode	8
tessellate	0
extrude	0
visibility	0
drawOrder	8
icon	8
geometry	0
dtype: int64	

✓ Valeurs manquantes après traitement

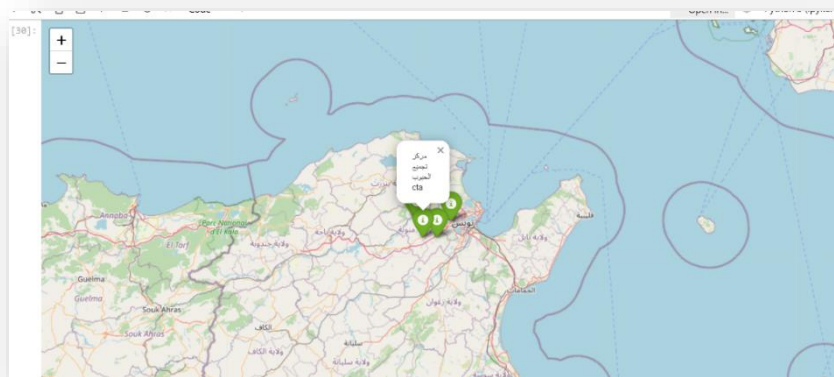
Name	0
description	0
timestamp	0
begin	0
end	0
altitudeMode	0
tessellate	0
extrude	0
visibility	0
drawOrder	0
icon	0
geometry	0
dtype: int64	

Ensuite, nous avons affiché les **statistiques générales** : le nombre total de points de collecte, les entreprises impliquées et le nombre des entreprises qui sont différentes.

```

✦ Nombre total de points de collecte : 8
✦ Entreprises présentes : ['الشركة التعاونية المركزية للبنور' COSEM'
'الشركة التعاونية المركزية للبنور' COSEM'
'الشركة التعاونية المركزية للزراعات الكبرى' CCGC'
'الشركة التعاونية المركزية للبنور' COSEM'
'الشركة التعاونية المركزية للقمح' COCEBLE'
'مصرف الخدمات الفلاحية المتحدة' CMA' 'مركز تجميع الجنوب' cta']
✦ Nombre d'entreprises différentes : 7
  
```

Enfin, nous avons généré une **carte interactive** avec Folium où chaque point de collecte est représenté par un **marqueur cliquable** indiquant le nom de l'entreprise.



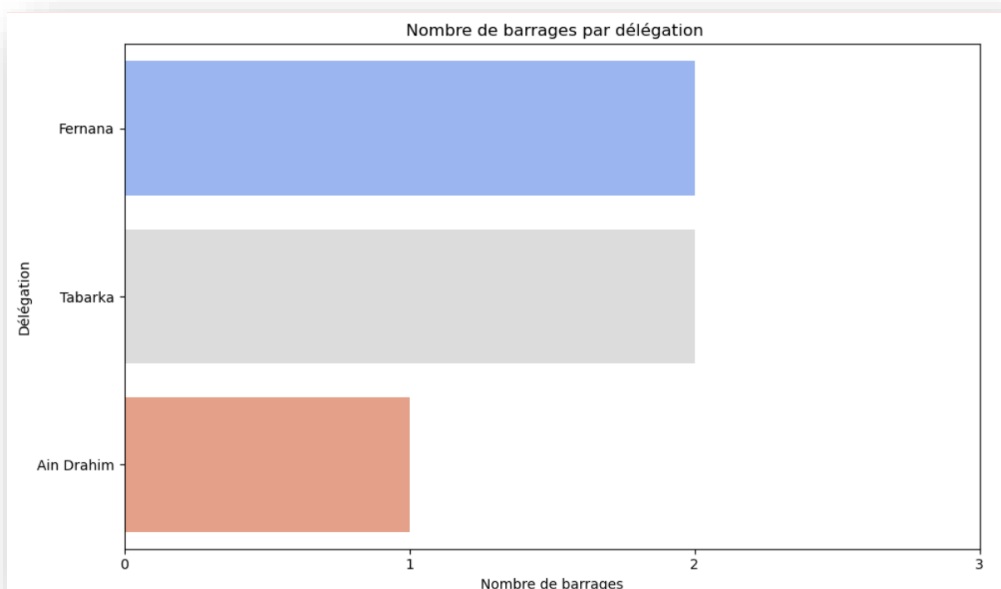
1.8. Analyse du fichier map3.geojson

Le fichier intitulé map3.geojson recense les **barrages** en Tunisie. Les colonnes principales sont : **marker-color** / **marker-size** / **marker-symbol** (paramètres de style pour la visualisation), **Nom-ar** / **Nom-fr** (noms des barrages en arabe et en français), **capacité-mille-litre** (capacité de stockage exprimée en milliers de litres), **Date-creation** (année de mise en service du barrage), **Delegation** (la délégation administrative où se situe le barrage), **geometry** (coordonnées géographiques du barrage).

Nous avons affiché des informations générales, notamment le **nombre total de barrages**, la **liste des délégations concernées** et la **répartition des barrages par délégation avec un graphique en barres**.

```
✦ Nombre total de barrages : 5
✦ Délégations concernées : ['Fernana' 'Ain Drahim' 'Tabarka']
✦ Nombre de délégations uniques : 3
```

On remarque que **Fernana** et **Tabarka** comptent chacune **2 barrages**, ce qui en fait les délégations les mieux équipées en termes de retenues d'eau. **Aïn Draham** suit avec **1 seul barrage**, ce qui peut s'expliquer par ses caractéristiques géographiques plus montagneuses et forestières, rendant potentiellement plus difficile la construction de barrages.



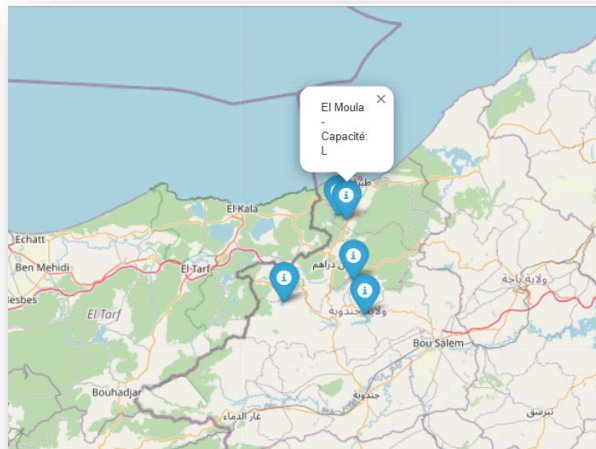
⇒ La répartition des barrages n'est pas homogène sur le territoire. Fernana et Tabarka concentrent la majorité des infrastructures de stockage d'eau, ce qui pourrait influencer localement l'accès à l'irrigation, le développement agricole et la gestion des ressources hydriques.

Puis nous avons créé une **carte interactive** à l'aide de **Folium**, où chaque barrage est représenté par un marqueur affichant son **nom** et sa **capacité** de stockage.

```
# Carte interactive avec Folium
m = folium.Map(location=[gdf.geometry.y.mean(), gdf.geometry.x.mean()], zoom_start=10)

for _, row in gdf.iterrows():
    folium.Marker(
        location=[row.geometry.y, row.geometry.x],
        popup=f"{row['Nom-fr']} - Capacité: {row['capacité-mille-litre']} L",
        icon=folium.Icon(color="blue", icon="info-sign")
    ).add_to(m)

# 🗺️ Afficher la carte interactive
m
```



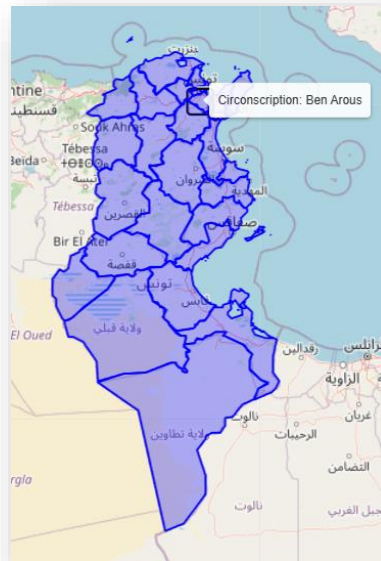
1.9. Analyse du fichier tncirconscriptions.geojson

Ce fichier contient les **circonscriptions électorales tunisiennes**. Il inclut les colonnes suivantes : **circo_id** (identifiant unique de chaque circonscription), **circo_na_1** (nom de la circonscription), **geometry** (forme géographique de la circonscription, utilisée pour la représentation cartographique).

Nous avons affiché le **nombre total de circonscriptions** et listé leurs **noms**.

* Nombre total de circonscriptions : 27
 * Liste des circonscriptions : ['Tunis 1' 'Tunis 2' 'Ariana' 'Ben Arous' 'Manubah' 'Nabeul 1' 'Nabeul 2'
 'Zaghouan' 'Bizerte' 'Béja' 'Jendouba' 'Le Kef' 'Siliana' 'Sousse'
 'Monastir' 'Mahdia' 'Sfax 1' 'Sfax 2' 'Kairouan' 'Kassérine'
 'Sidi Bou Zid' 'Gabès' 'Médenine' 'Tataouine' 'Gafsa' 'Tozeur' 'Kebili']

Ensuite, nous avons généré une **carte interactive** à l'aide de **Folium**, centrée sur la moyenne des centroïdes des géométries. Chaque circonscription est représentée par un **polygone bleu semi-transparent** avec un **tooltip interactif** affichant son nom. Cette carte permet une visualisation géographique claire de la répartition des circonscriptions à l'échelle nationale.



1.10. Analyse du fichier `z_interv_bizerte_pno4.geojson`

Le fichier `z_interv_bizerte_pno4.geojson` contenant les zones d'intervention à Bizerte. Les colonnes principales sont les suivantes : (**Name** (souvent vide ou non défini dans cet extrait), **description** (contient des métadonnées encodées en HTML, dont le gouvernorat (**GOUVER = BIZERTE**) et possiblement d'autres informations techniques ou d'identification), **geometry** (définit la forme de chaque zone d'intervention sous forme de polygones).

Après avoir remplacé les valeurs manquantes par le mode de chaque colonne,

```

✓ Valeurs manquantes après traitement :
  Name          16
description      0
geometry         0
dtype: int64
✓ Valeurs manquantes après traitement :
  Name          0
description      0
geometry         0
dtype: int64

```

Nous avons affiché le nombre total de zones.

✦ Nombre total de zones : 16

Une carte interactive est ensuite générée en plaçant des marqueurs au centre de chaque zone. Chaque point affiche une description, facilitant la visualisation de la répartition des interventions sur le territoire.



1.11. Analyse du fichier zonesindustrielles.geojson

Ce fichier comprend les colonnes suivantes : (**name** (nom de la zone industrielle), **styleUrl** / **styleHash** (attributs de style pour la cartographie), **description** (informations sur la source des

données, souvent des institutions publiques ou privées), **المصدر** : (source), **الولاية / الجهة** : (gouvernorat et région administrative), superficie de la zone en hectares, **lon / lat** (coordonnées géographiques), **geometry**).

Nous avons fourni des statistiques sur le nombre total de zones, les gouvernorats concernés et la distribution des superficies.

```

♦ Nombre total de zones industrielles : 152
♦ Gouvernorats uniques : [ 'الكاف' 'جندوبة' 'القيروان' 'زغوان' 'أريانة' 'بن عروس' 'أريانة' 'بن عروس' 'منوبة' 'تونس'
'باجة' 'صفافس' 'سليانة' 'بنزرت' 'توزر' 'قفصة' 'القصرين'
'سوسة' 'مدنين' 'كلمة' 'المهديّة' 'سيدي بوزيد'
'تطاوين' 'قابس' 'نايل' 'المنتسير' ]
♦ Nombre de gouvernorats uniques : 25
♦ Nombre de zones industrielles par gouvernorat :
الولاية
20      بن عروس
16      صفاقس
10      نابل

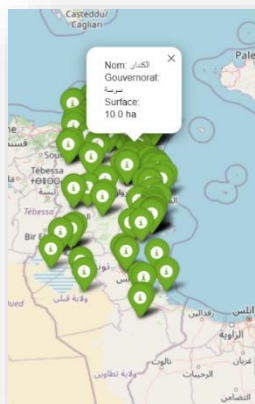
```

```

♦ Statistiques sur la surface des zones industrielles :
count      152
unique      59
top         10.0
freq        17
Name: المساحة (هـ), dtype: object

```

Une carte interactive avec **Folium** permet ensuite de visualiser chaque zone, en affichant son nom, son gouvernorat et sa surface.



2. Cartographie de la Tunisie avec D3.js

2.1. Choix des jeux de données

Après avoir visualisé, exploré et analysé les tendances des différentes données disponibles, nous avons sélectionné les jeux de données suivants : `tncirconscriptions.geojson`, `barrages_collinaires_beja.geojson`, `map3.geojson`, `zonesindustrielles.geojson`, `CollecteGraineBeja.geojson` et `lacs_collinaires_beja.geojson`.

Ces jeux de données seront utilisés pour tracer une carte interactive et attrayante de la Tunisie, affichant des informations pertinentes sur plusieurs thématiques et

intégrant des données géographiques spécifiques, telles que les barrages, les zones industrielles ou les points de collecte de graines.

2.2. Préparation de la carte

2.2.1. Partie Html

La bibliothèque D3.js est chargée à l'aide d'un lien vers sa version 7 en ligne : cela permet de manipuler facilement des données géographiques.

```
<script src="https://d3js.org/d3.v7.min.js"></script>
```

Le bloc `<style>` contient les styles de la page et le body utilise Flexbox pour placer la carte SVG à gauche et un panneau d'information à droite. Des styles sont aussi définis pour les path SVG (traits, couleurs, survol), ainsi que pour chaque type de donnée (lacs, graines, zones industrielles...).

À la fin, le fichier JavaScript `script.js` est chargé pour dessiner la carte, charger les fichiers GeoJSON, ajouter les événements interactifs et mettre à jour le contenu du panneau d'information.

2.2.2. Partie CSS

Le code CSS définit le style d'une carte interactive et d'un panneau d'informations à côté. Le corps de la page utilise un fond gris clair et une disposition flexible pour organiser la carte (SVG) à gauche et les informations à droite. La carte a une couleur de fond bleu pâle, des coins arrondis et une ombre douce. Les régions, barrages, zones industrielles et autres éléments sont colorés différemment pour les différencier.

Ainsi un panneau d'informations présenté en haut à droite pour afficher les détails avec une boîte blanche et une légende en bas à gauche pour expliquer les couleurs.

2.2.3. Partie Js

Le code commence par définir la projection de la carte à l'aide de `geoConicConformal` pour optimiser l'affichage de la Tunisie. La projection est centrée sur la Tunisie et utilise une échelle et une translation adaptées à la taille de l'écran.

```
const svg = d3.select("svg");
const width = +svg.attr("width") || 1000; // Augmenté
pour mieux voir
const height = +svg.attr("height") || 433;

// Projection OPTIMISÉE pour la Tunisie
const projection = d3.geoConicConformal()
  .center([12, 34])
  .parallels([33, 38])
  .scale(3800)
  .translate([width / 2, height / 2]);
```

Pour charger plusieurs jeux de données GeoJSON à partir de GitHub nous avons utilisé `Promise.all`.

```
// Stockage des données
let allData = {};

Promise.all([
  d3.json("https://raw.githubusercontent.com/farahghazouani/carte-tunisie/main/tncirconscriptions.geojson"),
  d3.json("https://raw.githubusercontent.com/farahghazouani/carte-tunisie/main/barrages_collinaires_beja.geojson"),
  d3.json("https://raw.githubusercontent.com/farahghazouani/carte-tunisie/main/map3.geojson"),
  d3.json("https://raw.githubusercontent.com/farahghazouani/carte-tunisie/main/zonesindustrielles.geojson"),
  d3.json("https://raw.githubusercontent.com/farahghazouani/carte-tunisie/main/CollecteGraineBeja.geojson"),
  d3.json("https://raw.githubusercontent.com/farahghazouani/carte-tunisie/main/lacs_collinaires_beja.geojson")
]).then(([circonscriptions, barragesCollinaires, barragesNord, zonesIndustrielles, collecteGraine, lacsBeja]) => {
```

Nous avons affiché **des éléments sur la carte** :

- **Régions (circonscriptions)** : Les régions sont dessinées sous forme de chemins (`<path>`) avec un style de base et un effet au survol.
- **Barrage, Industrie, Collecte** : Ces éléments sont représentés par des cercles colorés (**bleu** pour les barrages, **vert** pour les industries, **orange** pour les sites de collecte). Chaque élément est interactif et lorsqu'on clique sur un élément, une fonction affiche les détails correspondants dans un panneau d'information à droite.

```
// Dessiner les régions - FOND DE CARTE
svg.selectAll(".region")
  .data(allData.regions)
  .enter().append("path")
  .attr("class", "region")
  .attr("d", path)
  .attr("fill", "#4e79a7")
  .on("click", function(event, d) {
    showRegionDetails(d);
  });
```

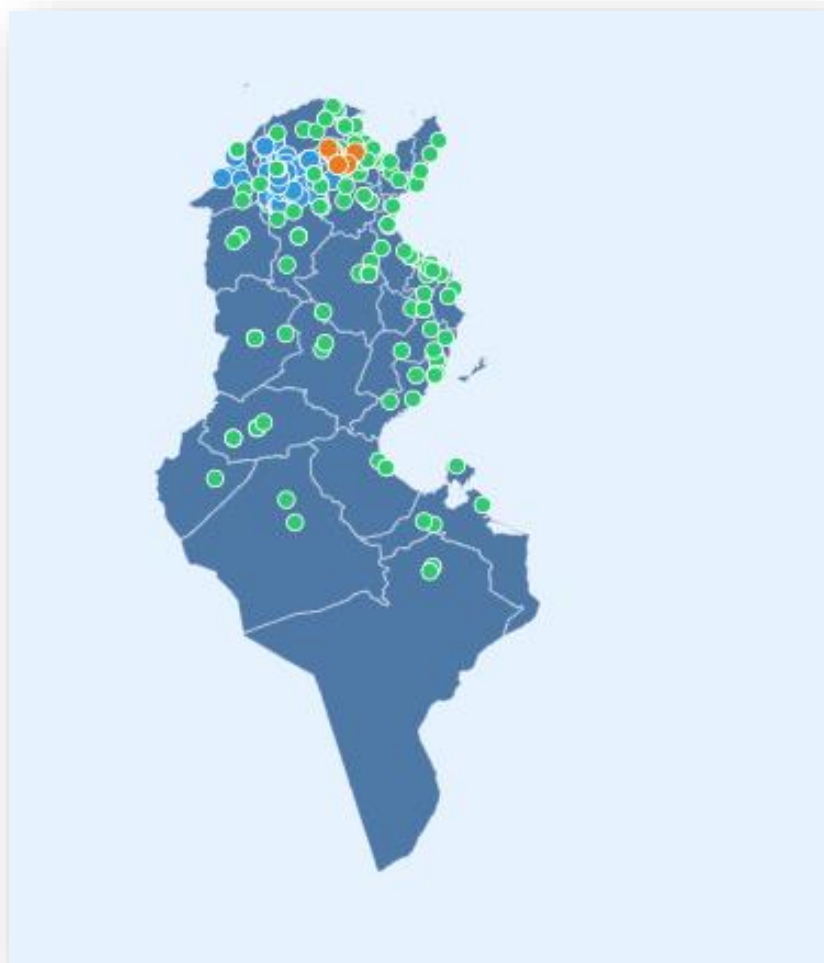
```
// Dessiner les barrages - POINTS BLEUS
svg.selectAll(".barrage")
  .data(allData.barrages)
  .enter().append("circle")
  .attr("class", "barrage")
  .attr("cx", d => projection(d.geometry.coordinates)[0])
  .attr("cy", d => projection(d.geometry.coordinates)[1])
  .attr("r", 6)
  .attr("fill", "#3498db")
  .on("click", function(event, d) {
    event.stopPropagation();
    showBarrageDetails(d);
  });
```

```
// Dessiner les industries - POINTS VERTS
svg.selectAll(".industrie")
  .data(allData.industries)
  .enter().append("circle")
  .attr("class", "industrie")
  .attr("cx", d => projection(d.geometry.coordinates)[0])
  .attr("cy", d => projection(d.geometry.coordinates)[1])
  .attr("r", 6)
  .attr("fill", "#2ecc71")
  .on("click", function(event, d) {
    event.stopPropagation();
    showIndustryDetails(d);
  });
```

```
// Dessiner les collectes - POINTS ORANGES
svg.selectAll(".collecte")
  .data(allData.collectes)
  .enter().append("circle")
  .attr("class", "collecte")
  .attr("cx", d => projection(d.geometry.coordinates)[0])
  .attr("cy", d => projection(d.geometry.coordinates)[1])
  .attr("r", 6)
  .attr("fill", "#e67e22")
  .on("click", function(event, d) {
    event.stopPropagation();
    showCollecteDetails(d);
  });
```

Lorsque l'utilisateur clique sur un élément (région, barrage, zone industrielle ou site de collecte), les détails pertinents (nom, capacité, localisation, etc.) sont affichés dans un panneau d'informations.

2.2.4. Résultat



O. Zitoun

Type: Barrage

Capacité: 100,000 m³

Année: 2002

Localisation: N/A

Barrage

Type: Barrage

Capacité: 112,000 litres

Année: N/A

Localisation: Fernana

cta مركز تجميع الحبوب

Localisation: N/A

Nabeul 2

Circonscription

Code: 152

3. Visualisation du réseau électrique du Qatar avec D3.js

3.1. Objectif de la visualisation

L'objectif de cette visualisation est de cartographier le réseau électrique de transmission du Qatar, en mettant en évidence :

Les lignes de transmission électrique (et leurs caractéristiques comme la puissance ou le type),

Les points d'interconnexion (villes ou nœuds techniques),

Les connexions transfrontalières (notamment avec l'Arabie Saoudite).

Ce type de visualisation permet de mieux comprendre l'organisation spatiale du réseau, d'identifier les zones stratégiques et de détecter d'éventuels points critiques ou isolés.

3.2. Choix techniques et explication du code

Le fichier utilisé est un GeoJSON représentant le réseau :

- Les lignes (type LineString) ont des attributs comme la puissance (transmissionPower) et le type de ligne (lineType).
- Les nœuds (type Point) correspondent à des villes ou à des points techniques avec un nom.

3.3. Technologie choisie : D3.js

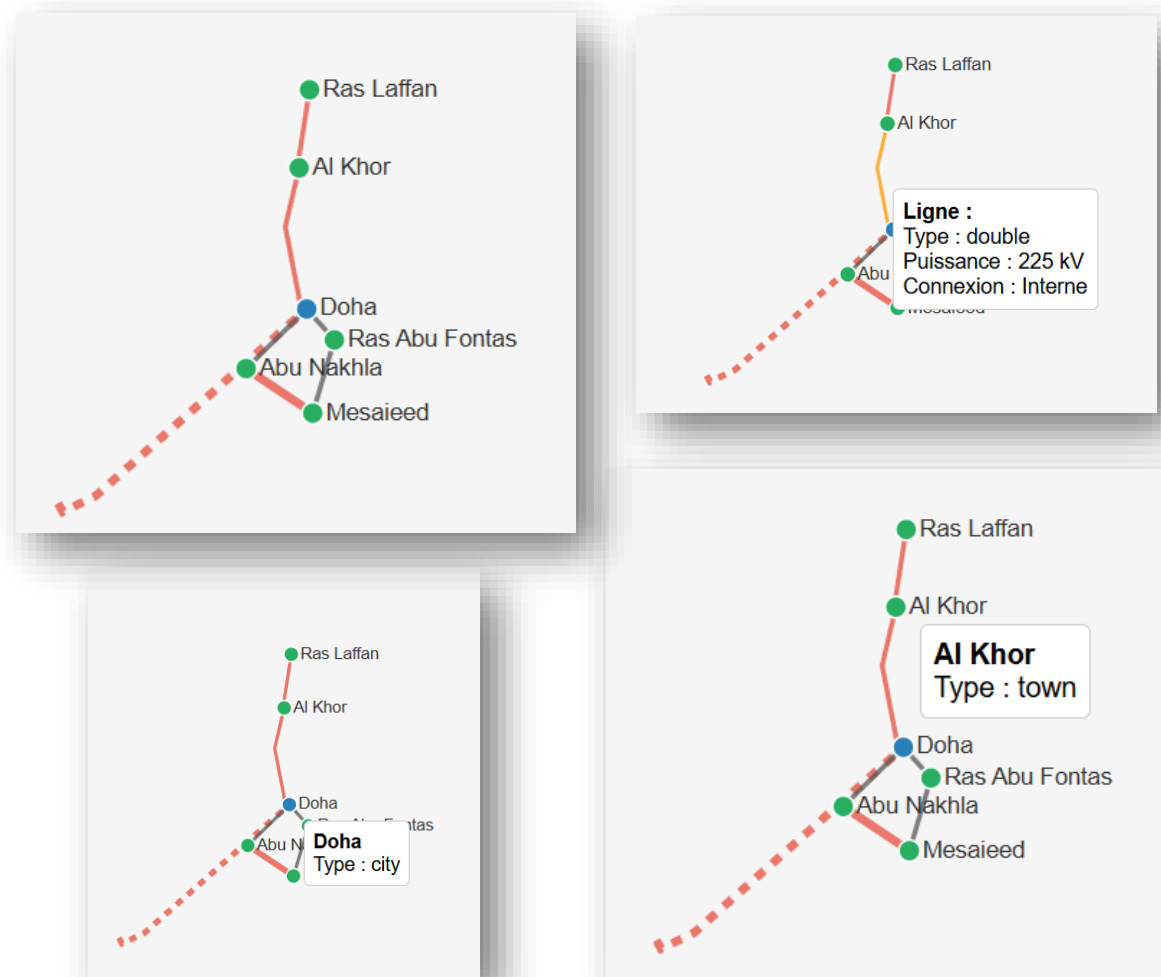
Dans ce projet, la bibliothèque D3.js (version 7) a été utilisée pour visualiser le réseau électrique du Qatar à partir de données géographiques au format GeoJSON.

Structure du code

- **Projection utilisée** : geoMercator, centrée sur le Qatar ([51.2, 25.3]) et zoomée avec un scale de 8500.
- **Chargement des données** : les données GeoJSON sont chargées depuis GitHub (<https://raw.githubusercontent.com/farahghazouani/carte-tunisie/main/electric-network-qatar.geojson>), ce qui permet une mise à jour facile et un hébergement gratuit.
- **Lignes électriques** :
 - Représentées comme des path SVG.
 - Leur **épaisseur** varie selon la puissance (transmissionPower) et leur **couleur** selon le type (lineType).
 - Les **interconnexions internationales** (ex : Qatar – Arabie Saoudite) sont visualisées avec un trait pointillé (stroke-dasharray).
- **Nœuds du réseau** :
 - Les **villes** ou **sous-stations** sont représentées par des cercles.

- Les couleurs différencient les types de nœuds : bleu pour les villes (city), vert pour les nœuds techniques.
- **Étiquettes :**
 - Chaque nœud nommé affiche un petit texte en SVG à côté du point.
- **Interactions (tooltips) :**
 - Survol d'un élément affiche des informations détaillées : puissance, type, connexion, nom.

3.4. Résultat



Cette visualisation permet de:

- **Comprendre la structure du réseau électrique du Qatar**, notamment la hiérarchie des lignes selon leur puissance.

- **Identifier les villes et points de connexion majeurs** (ex. Doha, Salwa).
- **Mettre en évidence les interconnexions internationales** comme celles avec l'Arabie Saoudite, ce qui est important dans un contexte de coopération énergétique régionale.
- **Faciliter l'analyse visuelle** pour des décideurs, chercheurs ou ingénieurs réseau.

4. Analyse Réseau avec Gephi : Étude du Dataset WebAtlas_EuroSiS

4.1. Introduction et Objectifs

L'analyse des réseaux permet de mieux comprendre les dynamiques relationnelles entre acteurs, institutions ou entités. Dans ce rapport, nous explorons la structure et les sous-communautés d'un réseau institutionnel à l'aide de **Gephi**, un outil de visualisation puissant pour l'analyse de graphes complexes.

Notre objectif est d'extraire des connaissances invisibles à l'œil nu à partir de données brutes, telles que :

- L'identification des communautés (clusters),
- L'analyse des rôles structurants,
- La mise en évidence de partenariats stratégiques.

4.2. Présentation du Dataset et Justification de l'Outil Gephi

Gephi est l'outil idéal pour prendre en charge les fichiers GEXF, Il offre des algorithmes de visualisation puissants et permet une analyse intuitive et visuelle de structures complexes (communautés, centralité, etc.). Cet outil nous a permis d'analyser le fichier **WebAtlas_EuroSiS.gexf**. En fait, ce dernier contient :

- **1285 nœuds**, représentant des institutions ou entités,
- **6462 arêtes**, représentant les collaborations entre elles.

Autrement dit, chaque nœud est une entité impliquée dans une relation et chaque arête correspond à une collaboration ou interaction. Le graphe est **non orienté**, ce qui signifie que les relations sont bidirectionnelles (si A est lié à B, alors B est aussi lié à A).

4.3. Analyse Globale du Réseau

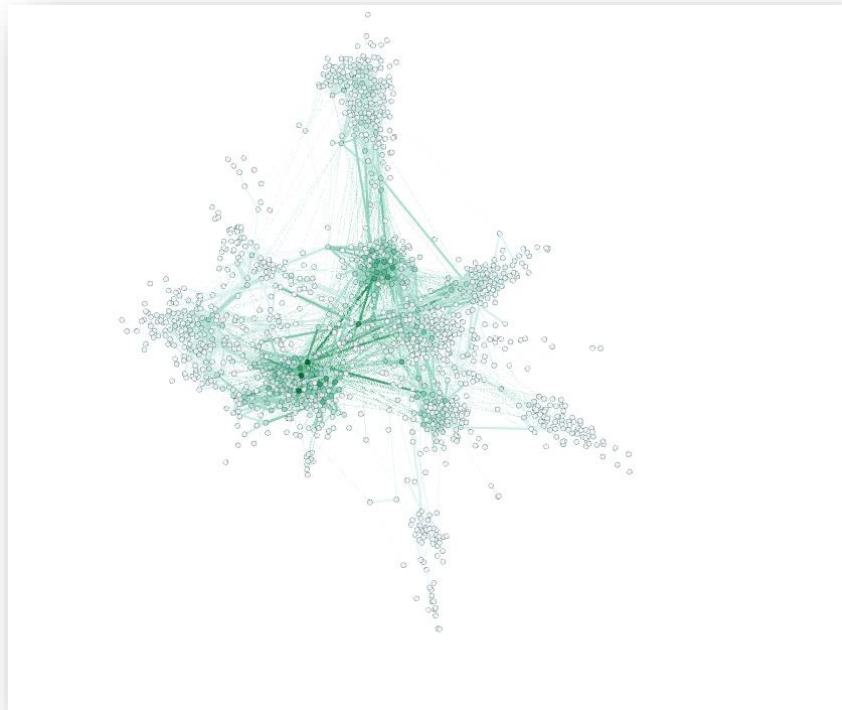
4.3.1. Choix de l'algorithme

On va choisir l'algorithme **ForceAtlas2** qui s'impose naturellement car il est particulièrement adapté aux réseaux denses comme **WebAtlas_EuroSiS** (1285 nœuds, 6462 arêtes). Il permet de positionner les **nœuds fortement connectés à proximité**, facilitant ainsi la lecture des **communautés** sous forme de clusters bien distincts.

Contrairement à d'autres algorithmes de disposition, **ForceAtlas2** valorise visuellement la **structure modulaire** du réseau grâce à son principe de **forces d'attraction et de répulsion**, rendant l'interprétation plus intuitive. Enfin, ses **paramètres ajustables** (comme *scaling* ou *gravity*) offrent une grande souplesse pour optimiser la **clarté visuelle** du graphe selon les besoins de l'analyse.

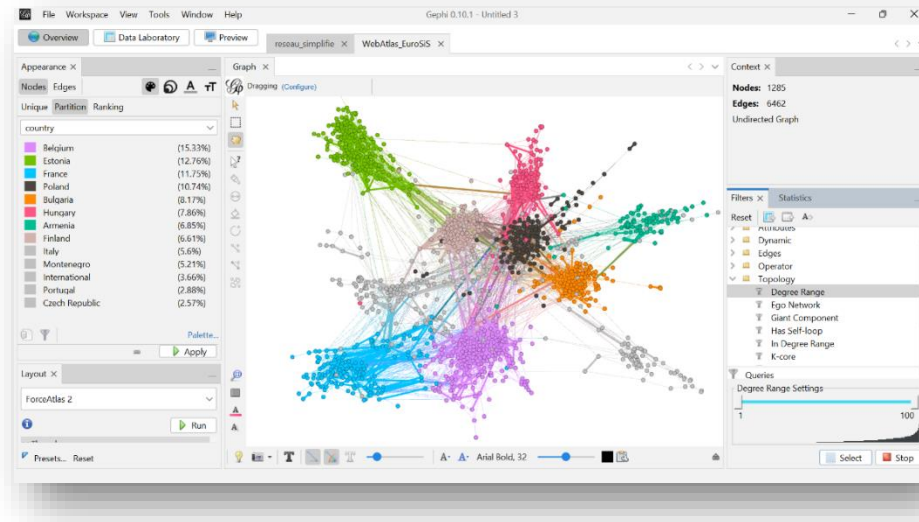
4.3.2. Carte du réseau

Cette représentation met en lumière les **acteurs centraux du réseau**, souvent responsables de la coordination entre plusieurs communautés ou de la circulation de l'information à grande échelle. Ces nœuds possèdent une **centralité élevée** (souvent de degré ou harmonique) et sont visuellement identifiables par leur **taille** et leur **position centrale**. On remarque aussi la présence de **sous-communautés fortement interconnectées**.



4.3.3. Partition par Pays

Une première analyse par pays révèle des **pôles géographiques dominants**: Belgique (15.33 %), Estonie (12.76 %), France (11.75 %) et l'Arménie (6.85 %). Ces chiffres suggèrent une forte centralité ou influence géographique. Des acteurs transnationaux, regroupés sous l'étiquette *International* (3.66 %), comme **Counterpart International Inc.**, soulignent aussi l'envergure du réseau.

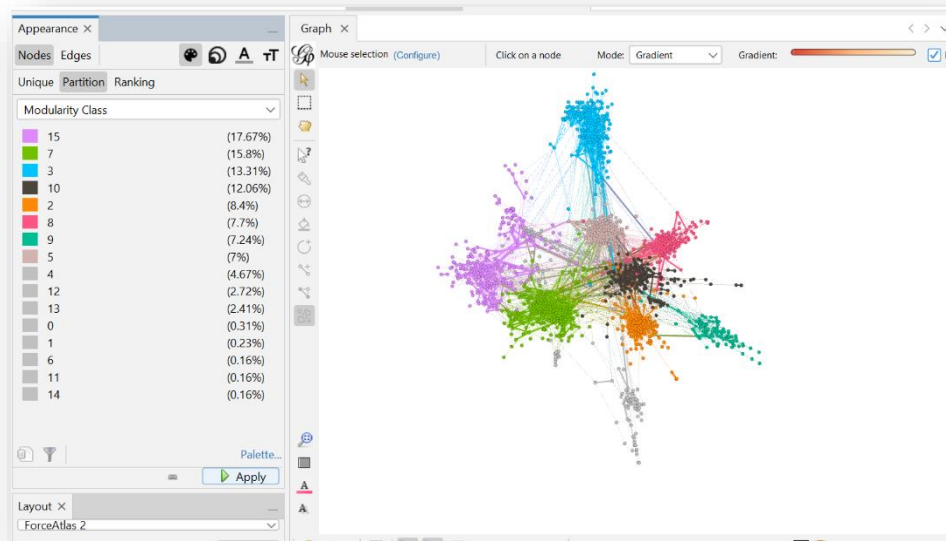


4.3.4. Détection de Communautés : Analyse de la Modularité

- Visualisation des Communautés

L'algorithme de modularité appliqué sur le réseau a permis d'identifier **16 communautés** distinctes. Ces groupes sont colorés différemment selon leur **modularity class**. Les plus importantes en volume sont : Classe 15 (17.67 % des nœuds), Classe 7 (15.8 %), Classe 3 (13.31 %), Classe 10 (12.06 %). À l'inverse, les classes comme 0, 1, 6, 11 ou 14 (< 0.5 %) représentent des **groupes marginaux ou isolés**, souvent en périphérie du graphe.

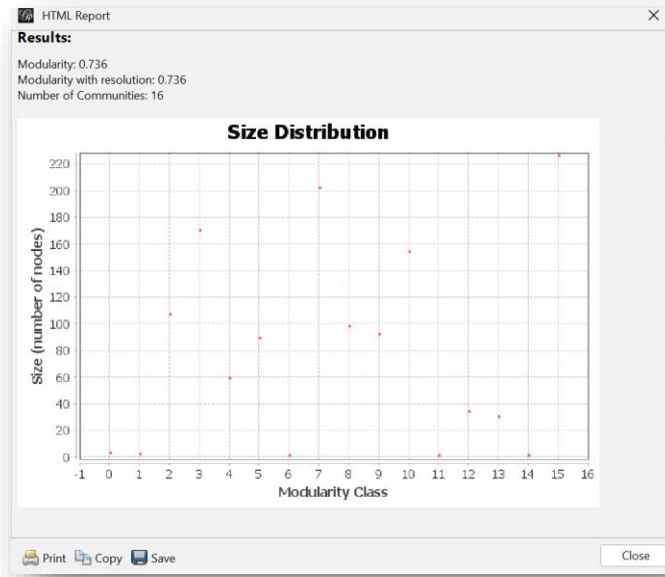
Ces clusters peuvent refléter des dynamiques **thématiques** (groupes de recherche, secteurs d'activité) ou **géographiques**.



- **Interprétation Numérique de la modularité**

La valeur de modularité globale est de **0.736**, ce qui indique une structure communautaire **fortement marquée**. Cette modularité élevée signifie que les entités sont **plus connectées entre elles à l'intérieur des groupes qu'avec l'extérieur**, ce qui reflète des sous-structures stratégiques et cohérentes.

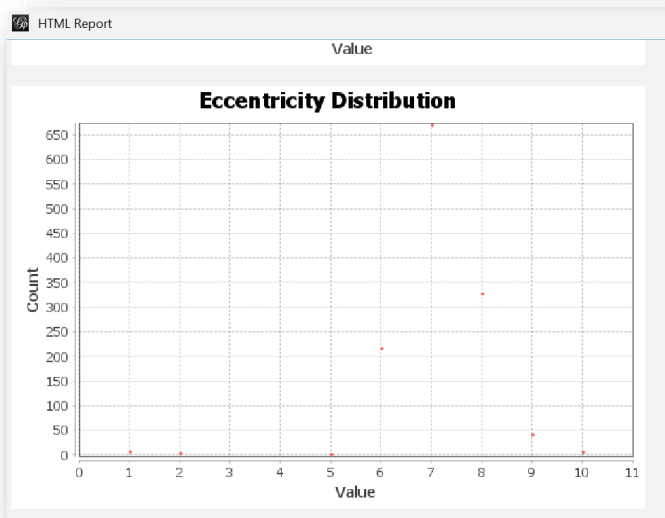
Certaines classes regroupent **un grand nombre de nœuds**, comme la **classe 15 (≈225 nœuds)**, la **classe 7 (≈200)**, la **classe 3 (≈170)** ou la **classe 10 (≈150)** : ces classes sont **structurées et centrales** dans le réseau. D'autres classes, comme **les classes 0, 1, 6, 11, 14**, ont une **taille très faible (proche de 0)**, ce qui peut indiquer des **acteurs marginaux**, des **sous-groupes très spécifiques** ou du **bruit dans les données**.



4.3.5. Accessibilité du Réseau : Distribution de l'Excentricité

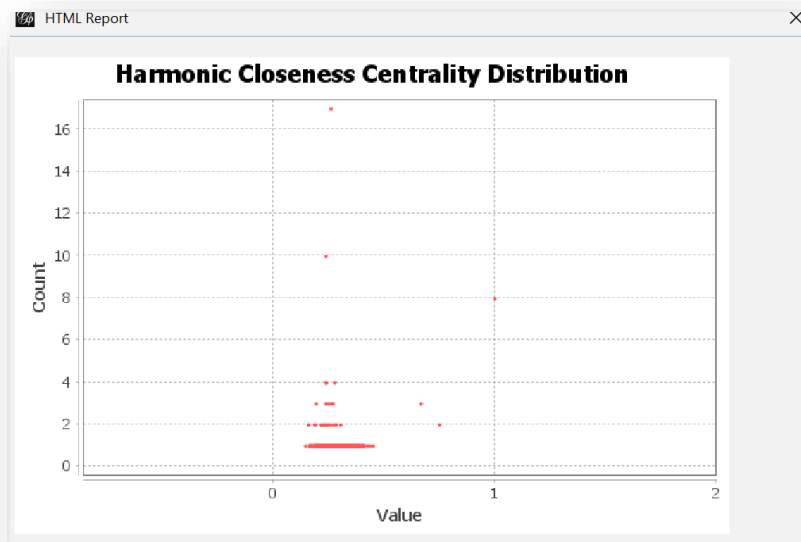
L'excentricité mesure la distance maximale d'un nœud à tous les autres. Dans notre graphe, la majorité des nœuds sont à **7 ou 8 sauts** les uns des autres. Cela indique un réseau **compact**, où **l'information circule efficacement**, même entre acteurs éloignés.

Cette faible profondeur est typique des réseaux institutionnels denses, comme ceux observés dans l'espace européen de la recherche et de l'innovation.



4.3.6. Influence et Rôle Central : Analyse de la Centralité

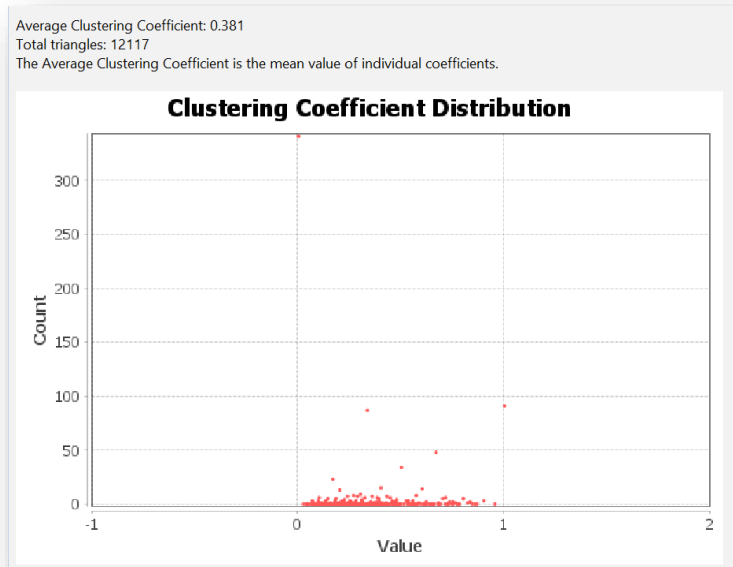
La plupart des nœuds ont une **centralité faible** (valeurs proches de 0) et quelques-uns se démarquent avec une centralité élevée : ce sont les **acteurs stratégiques, pivot** ou **nœuds passerelles**. Ces institutions influentes facilitent la circulation d'informations ou de ressources entre les différentes parties du réseau. Elles agissent comme des **intermédiaires-clés** et des hubs de coordination.



4.3.7. Cohésion Locale : Analyse du Clustering

Le **clustering moyen** est de **0.381**. Certains nœuds ont un **clustering parfait (1.0)** : ils appartiennent à des **mini-communautés très soudées**. On observe **12117 triangles**, ce qui suggère une **coopération modérée mais significative** entre les voisins.

Ces clusters denses peuvent représenter des **consortiums**, des **groupes de travail stables** ou des **partenariats structurés**.

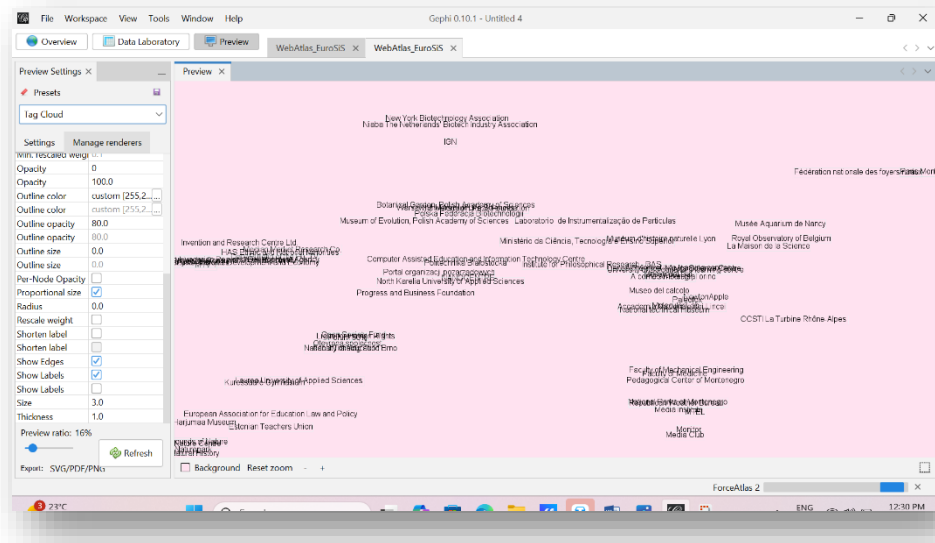


4.3.8. Analyse Visuelle du Réseau Institutionnel (Extrait – Couverture Partielle : 16%)

Cette représentation, issue d'un extrait couvrant **16 % du réseau total**, met en évidence les **logiques d'organisation** sous-jacentes du graphe. Les **nœuds centraux**, comme *Royal Observatory of Belgium* ou *Museum of Evolution*, se démarquent par leur **taille et position stratégique**, soulignant leur **rôle clé dans la circulation de l'information**.

Les **communautés** se forment naturellement en **amas denses et cohérents**, à l'image de *La Maison de la Science* proche de *Facultad de Educación y Tecnologías*, illustrant des **collaborations étroites**.

Bien que partielle, cette vue permet déjà d'**identifier visuellement les hubs, les sous-groupes actifs et les dynamiques relationnelles**.



Conclusion

Ce réseau est **fortement structuré** en sous-communautés cohérentes, il est **compact** avec une circulation rapide de l'information. Il s'agit également des **acteurs clés** qui structurent le graphe, jouant un rôle d'**interface** entre les groupes.

5. Tableau

Dans le cadre de ce petit projet, nous avons exploré l'évolution de la pandémie de COVID-19 à travers un jeu de données fourni par l'**ECDC (European Centre for Disease Prevention and Control)**. Le dataset, intitulé data_COVID.csv, regroupe des données quotidiennes sur les cas et les décès liés à la COVID-19, réparties par pays, date et continent.

Ce jeu de données comprend plusieurs variables clés, parmi lesquelles :

- **dateRep** : date de déclaration des cas ;
- **Cases** et **deaths** : nombre de cas et de décès quotidiens ;
- **countriesAndTerritories** et **geoId** : identifiants des pays ;
- **PopData2019** : population estimée en 2019 pour chaque pays ;
- **continentExp** : continent correspondant ;
- **Cumulative_number_for_14_days_of_COVID-19_cases_per_100000** : indicateur de l'incidence cumulée sur 14 jours.

5.1. Choix du thème : la pandémie de COVID-19

Le sujet du coronavirus s'impose comme un choix pertinent et d'actualité. L'impact sanitaire, économique et social de cette pandémie mondiale en fait un sujet d'analyse incontournable. À

travers ce rapport, nous avons souhaité **visualiser les dynamiques spatio-temporelles de la propagation du virus**, à la fois au niveau mondial (comparaison entre continents) et national (évolution mensuelle en Tunisie).

L'objectif de cette analyse est double :

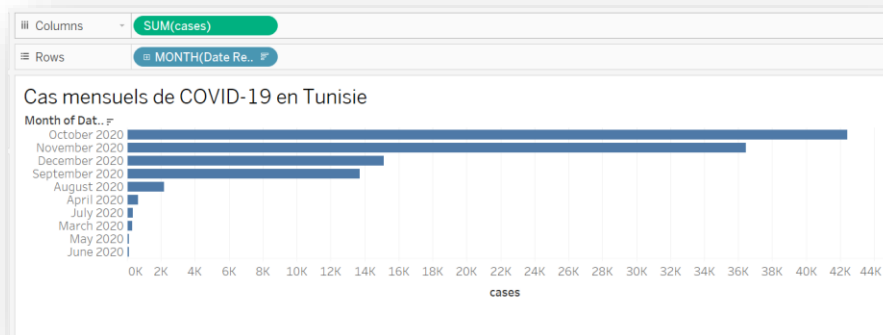
- **Mettre en évidence les disparités régionales** dans la propagation du virus ;
- **Comprendre les tendances temporelles** liées aux cas et aux décès afin d'en tirer des enseignements sur l'efficacité des mesures de contrôle mises en place.

5.2. Les visualisations fournies

5.2.1. Cas mensuels de COVID-19 en Tunisie

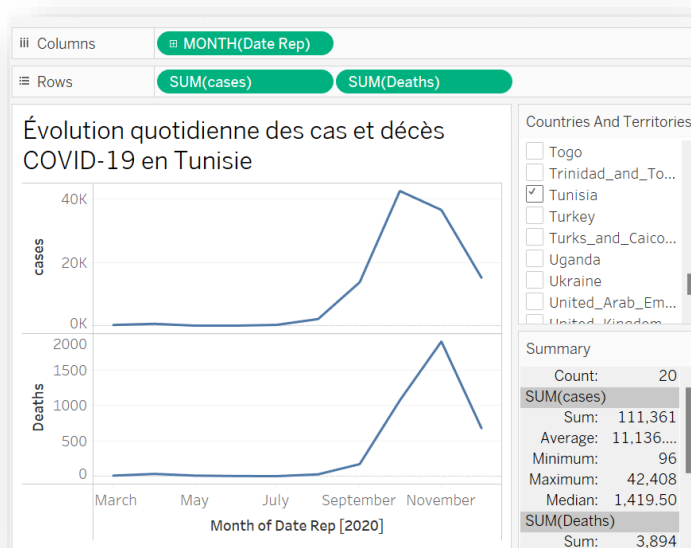
La première visualisation montre l'évolution mensuelle des cas de COVID-19 en Tunisie entre Octobre 2020 et Juin 2020, sous forme de barres horizontales. L'axe vertical représente les mois, tandis que l'axe horizontal montre le nombre total de cas enregistrés pour chaque mois. On observe une nette augmentation des infections à partir d'août 2020, avec un pic en octobre (plus de 42 000 cas). Avant cette période, les cas étaient peu nombreux, ce qui reflète soit une faible propagation initiale, soit un dépistage limité. Cette hausse rapide peut être liée à l'assouplissement des mesures sanitaires pendant l'été.

⇒ Ces tendances peuvent être mises en relation avec l'**allègement des mesures sanitaires** en été 2020, suivi d'une forte **deuxième vague**.



5.2.2. Évolution mensuelle des cas et des décès en Tunisie

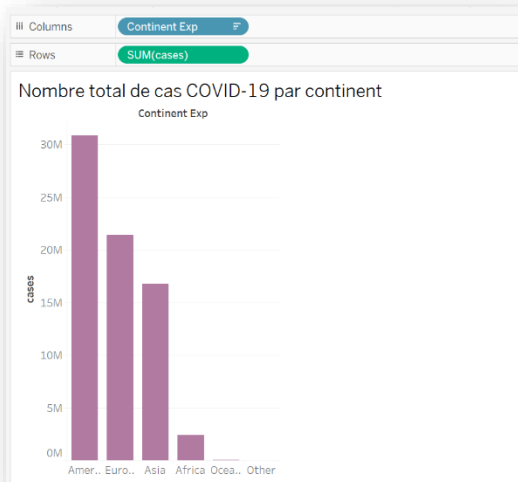
La deuxième visualisation compare l'évolution des cas et des décès sur la même période. Les deux courbes suivent une tendance similaire, avec un pic marqué en novembre : environ 40 000 cas et près de 2 000 décès. Cette synchronisation met en évidence l'impact significatif des vagues épidémiques sur la santé publique tunisienne.



5.2.3. Cas mondiaux de COVID-19 par continent

Cette figure montre que le **nombre total de cas de COVID-19 par continent**. On observe une **disparité importante** entre les continents : L'**Amérique** avec plus de 30 millions de cas, ce qui reflète une forte propagation, notamment aux États-Unis et au Brésil. L'**Europe** est également très touchée, avec plus de 20 millions de cas. L'**Asie**, bien que plus peuplée, elle affiche un total inférieur à l'Europe, ce qui pourrait s'expliquer par des politiques de santé publique plus strictes dans certains pays asiatiques et pour l'**Afrique**, l'**Océanie** et les **Autres**, ils présentent des nombres nettement plus faibles. Cela peut s'expliquer par une combinaison de facteurs : moins de tests, population plus jeune, ou une propagation initiale plus lente.

Ces chiffres offrent un **contexte comparatif utile** pour analyser la situation tunisienne. En effet, bien que la Tunisie ait connu une forte augmentation des cas à partir d'août 2020, son volume absolu reste bien inférieur à celui des grandes zones épidémiques.



5.2.4. Évolution mensuelle des décès COVID-19 dans le monde

Cette visualisation présente l'évolution mensuelle des décès liés à la COVID-19 à l'échelle mondiale. On y distingue plusieurs pics successifs :

- Une **hausse rapide à partir de mars-avril 2020**, en lien avec la première vague mondiale.
- Une **relative stabilité pendant l'été**, suivie d'un **fort rebond en novembre 2020**, avec un pic de mortalité atteignant plus de 250 000 décès ce mois-là.
- Une baisse nette en décembre, probablement liée à des confinements ou débuts de vaccination.

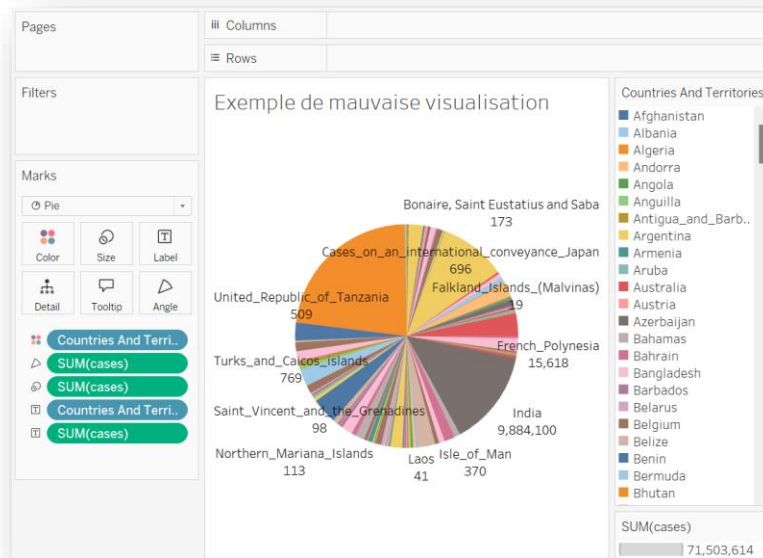
Cette tendance globale **fait écho à la courbe tunisienne** : une première hausse au printemps, une accalmie estivale, puis une **seconde vague marquée à l'automne**. En Tunisie comme ailleurs, on constate un **décalage temporel entre les pics de contamination et les décès**, ce qui souligne l'importance des mesures de prévention rapides.



6. Exemple de mauvaise visualisation

Nous avons volontairement créé un **camembert** représentant le **nombre total de cas COVID-19 par pays**. Ce type de graphique est ici inadapté pour plusieurs raisons :

- **Trop de catégories** : Le graphique contient près de 200 pays, ce qui rend la lecture impossible. Les parts sont trop petites pour être visibles ou distinguées clairement.
- **Couleurs redondantes** : Tableau recycle les couleurs après quelques valeurs, créant une confusion visuelle importante.
- **Étiquettes illisibles** : La plupart des noms de pays sont coupés ou absents, ce qui empêche toute identification précise.
- **Aucune hiérarchisation visuelle** : Tous les pays sont présentés au même niveau, qu'ils aient quelques centaines ou plusieurs millions de cas.



Impacts sur la compréhension

Cette visualisation est **esthétiquement confuse et informativement inefficace**. Elle provoque plusieurs biais cognitifs :

- **Biais de perception des tailles** : L'œil humain interprète mal les angles et les aires. Les différences importantes entre les pays sont peu visibles.
- **Surcharge cognitive** : Le trop grand nombre de parts empêche l'utilisateur de se concentrer sur une tendance ou de tirer une conclusion.
- **Fausse impression d'équivalence** : Certains pays très touchés (comme les États-Unis, l'Inde ou le Brésil) semblent aussi importants que de petits pays, ce qui est trompeur.

⇒ Ce graphique illustre parfaitement les erreurs à éviter : **ne pas utiliser de pie chart pour de nombreuses catégories** et préférer un **bar chart classé** pour comparer des valeurs aussi hétérogènes.

Conclusion

Ce projet a permis d'explorer concrètement les multiples facettes de la data visualisation en mobilisant une diversité d'outils adaptés à différents types de données : tabulaires, relationnelles et géospatiales. À travers les différentes visualisations développées, il a été possible de dégager des **tendances significatives**, de **comparer des phénomènes** à différentes échelles et de **mettre en lumière des structures cachées** au sein des jeux de données.

L'utilisation de **Gephi** a apporté une dimension relationnelle précieuse, notamment dans l'analyse de réseaux complexes. **Python** et **D3.js** ont permis une approche fine et interactive de la cartographie et de l'exploration des données territoriales, tandis que **Tableau Public** a facilité la création de tableaux de bord clairs et accessibles. L'analyse d'une visualisation erronée a quant à elle souligné l'importance de la rigueur dans le choix des représentations graphiques, tant en termes d'éthique que d'efficacité communicationnelle.

Enfin, la cartographie thématique de la Tunisie a permis de combiner données statistiques et géographiques dans une interface visuelle dynamique, offrant ainsi une lecture territorialisée pertinente et informative.

Ce travail a renforcé notre capacité à choisir, concevoir et interpréter des visualisations de données pertinentes, tout en développant un regard critique sur la manière dont l'information est perçue et transmise.

