

# DATA VIZUALISATION

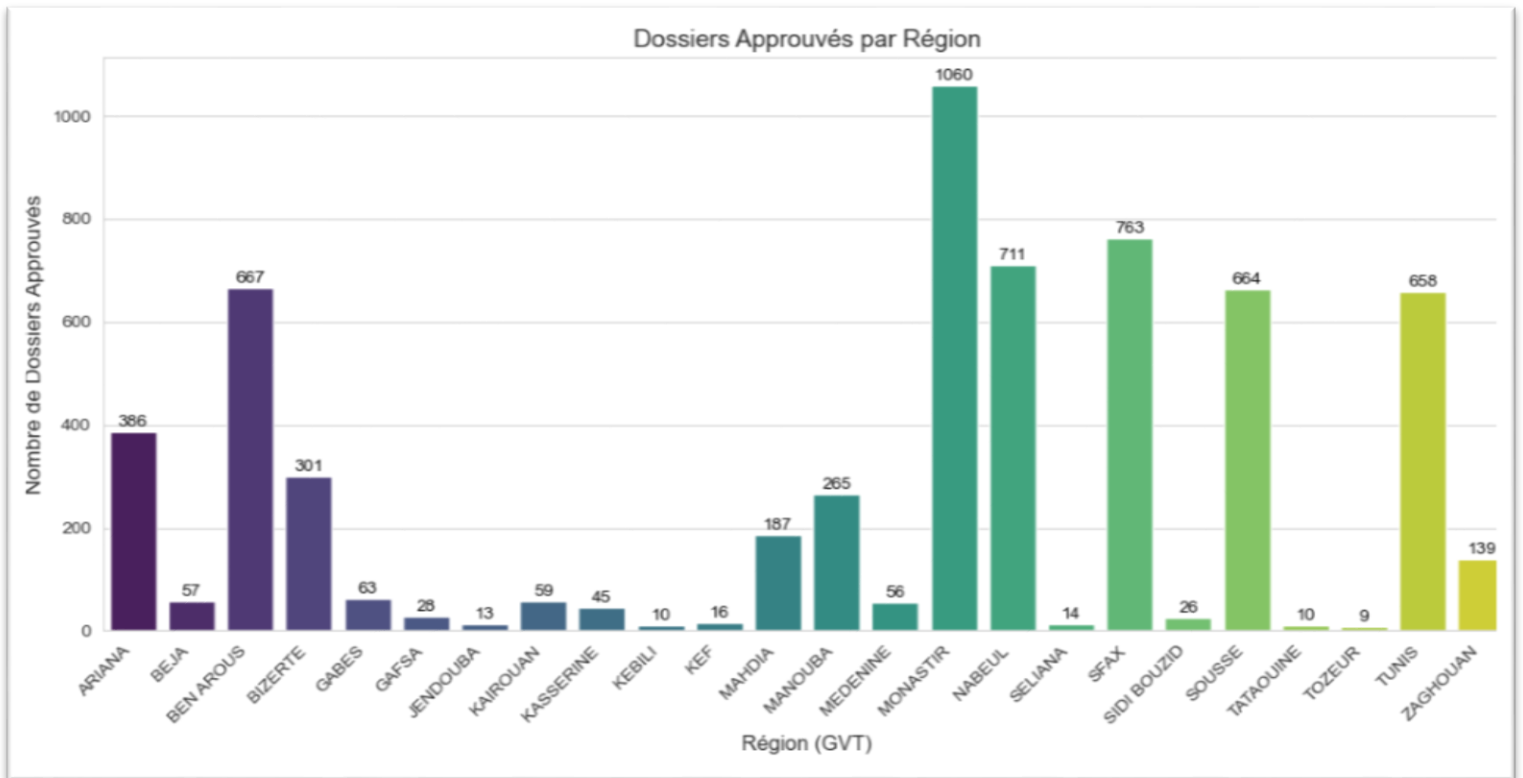
Travail à rendre

**Elaboré par :**  
**Ben rbiha Yassine**

# Partie 1 : Analyse et Visualisation de Données

## Python :

Utilisation du fichier Excel 'manpargvt' (Graphique a barre)



### Interprétation du graphique "Dossiers Approuvés par Région" :

#### ✓ Régions avec le plus de dossiers approuvés

1. **Monastir** arrive en tête avec **1060 dossiers approuvés**, ce qui montre une forte dynamique dans cette région.
2. **Sfax (763 dossiers)** et **Nabeul (711 dossiers)** suivent de près, indiquant aussi une activité administrative ou économique significative.
3. **Ben Arous (667)**, **Sousse (664)** et **Tunis (658)** sont également bien positionnées.

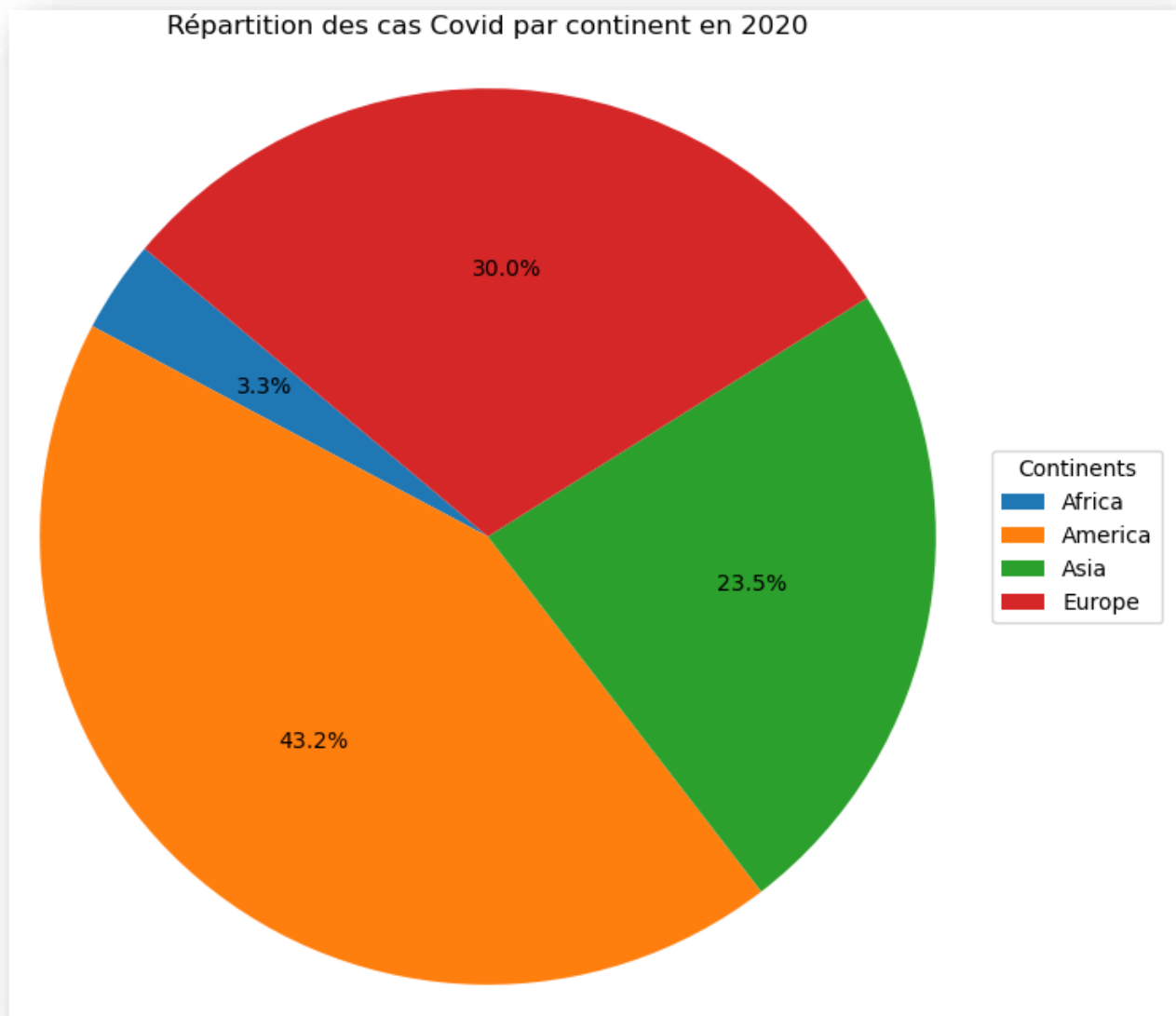
#### ⚠ Régions avec peu de dossiers approuvés

- **Tozeur (9 dossiers)**, **Kebili (10)** et **Jendouba (13)** ... ont les chiffres les plus bas. Cela pourrait refléter :

- Une faible demande
- Un manque d'accès aux services
- Des ressources limitées pour l'approbation des dossiers

### Python :

Utilisation du fichier Excel 'DATA\_covid' (**Diagramme En secteur**)



## Interprétation de la Visualisation :

### Amérique (~43%)

→ C'est le continent qui exporte le plus de cas dans les données. Près de la moitié des cas viennent d'Amérique.

### Europe (30%)

→ L'Europe est le deuxième plus grand contributeur. Elle représente environ un tiers des cas.

### Asie (~24%)

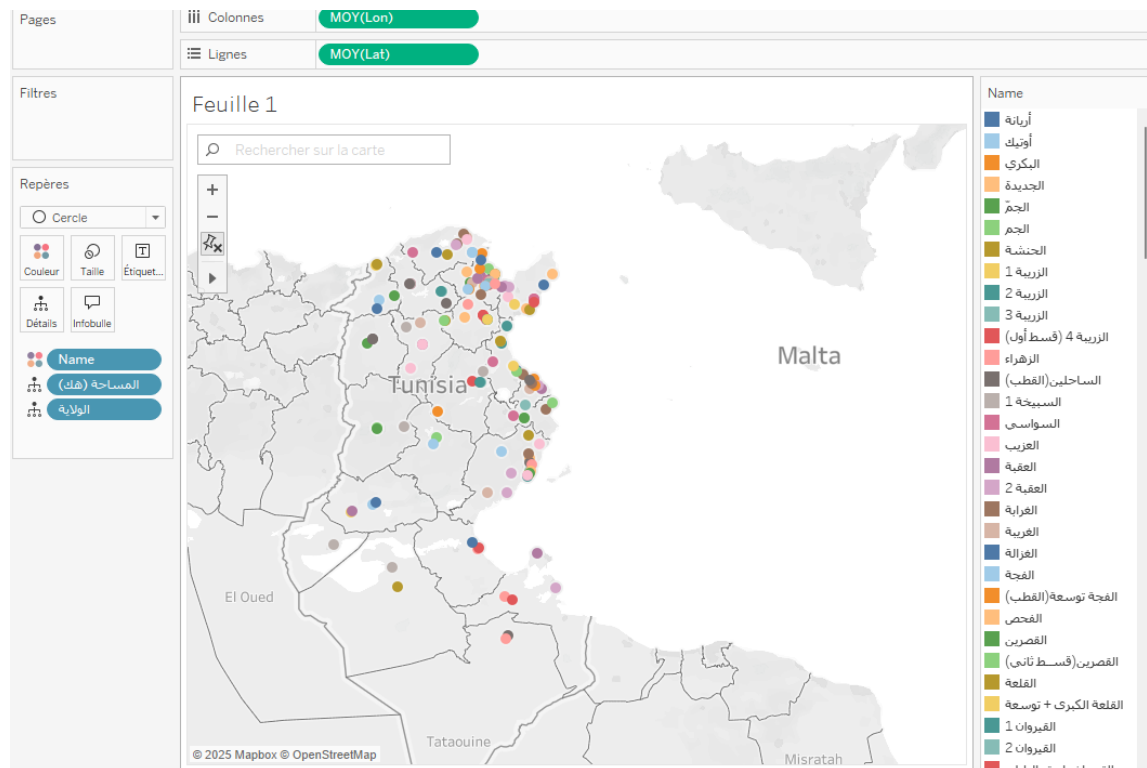
→ L'Asie contribue de manière significative mais moins que l'Europe ou l'Amérique.

### Afrique (~3%)

→ Une contribution faible comparée aux autres continents, mais toujours visible.

## Tableau :

Utilisation du fichier du type geojson 'zoneindustrielles'



Cette visualisation représente une carte de répartition géographique des différentes régions en Tunisie issues d'un fichier GeoJSON.

- Chaque cercle correspond à une localité ou une zone géographique, positionnée à l'aide des coordonnées géographiques (latitude et longitude).
- Les couleurs différencient les noms des localités (variable "Name").
- Les cercles sont placés selon les coordonnées (longitude et latitude)
- Des infos supplémentaires comme la superficie (hectares) ou la gouvernorat (الولاية) sont disponibles en survolant les points (info-bulle).

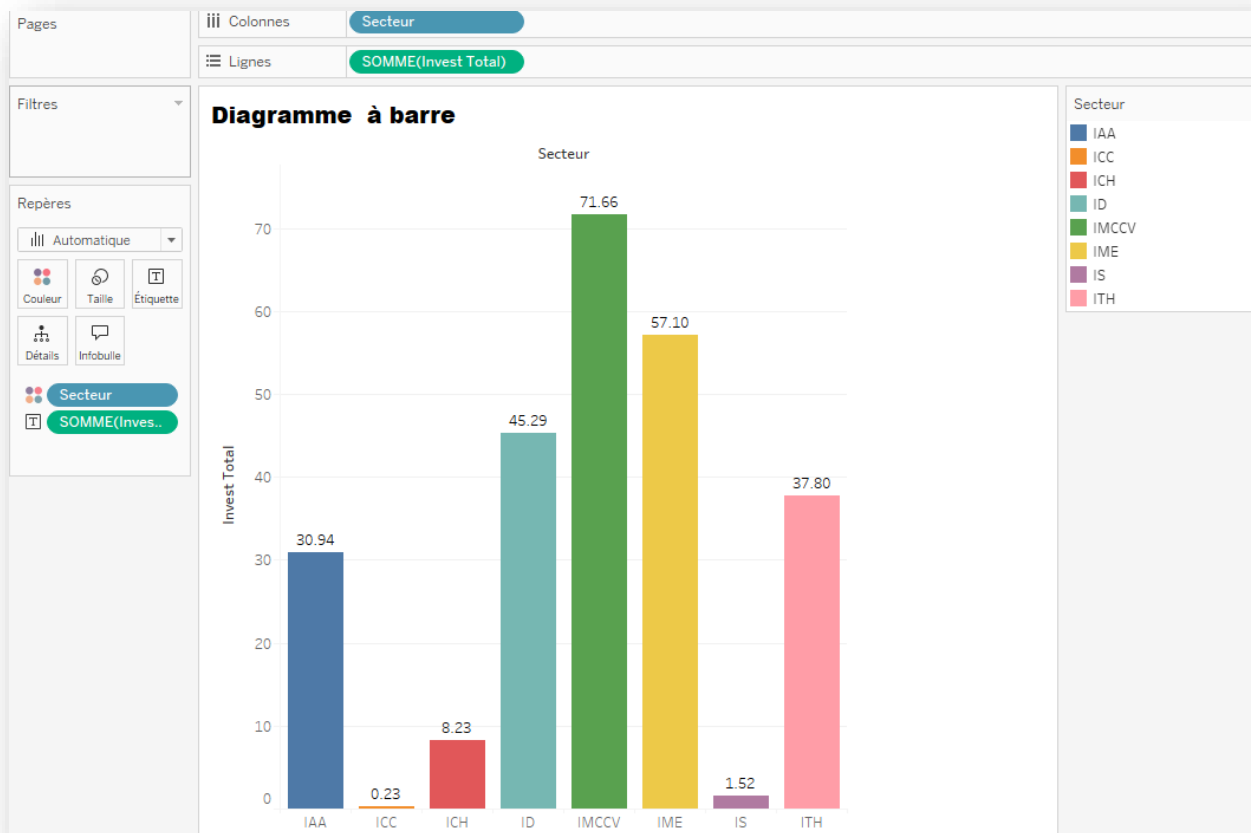
On observe une concentration plus importante de zones renseignées dans la partie nord et nord-est de la Tunisie, notamment autour du Grand Tunis, du Sahel et de la région de Nabeul.

Les zones du sud et du centre sont moins représentées ou comportent moins de points.

## Tableau :

### Utilisation du fichier du type Excel 'manpartsecteur4mois'

Ce graphique à barres montre la somme des investissements totaux (Invest Total) pour chaque secteur d'activité. Chaque barre représente un secteur, identifié par son



abréviation (IAA, ICC, ICH, etc.), et la hauteur de la barre reflète le montant total investi.

Axe des abscisses (X) : Secteurs (catégorie)

Axe des ordonnées (Y) : Montant total investi (SOMME(Invest Total))

Les étiquettes numériques au-dessus des barres affichent les valeurs précises.

Les couleurs différencient chaque secteur pour une meilleure lisibilité.

### **Secteurs avec les plus hauts investissements :**

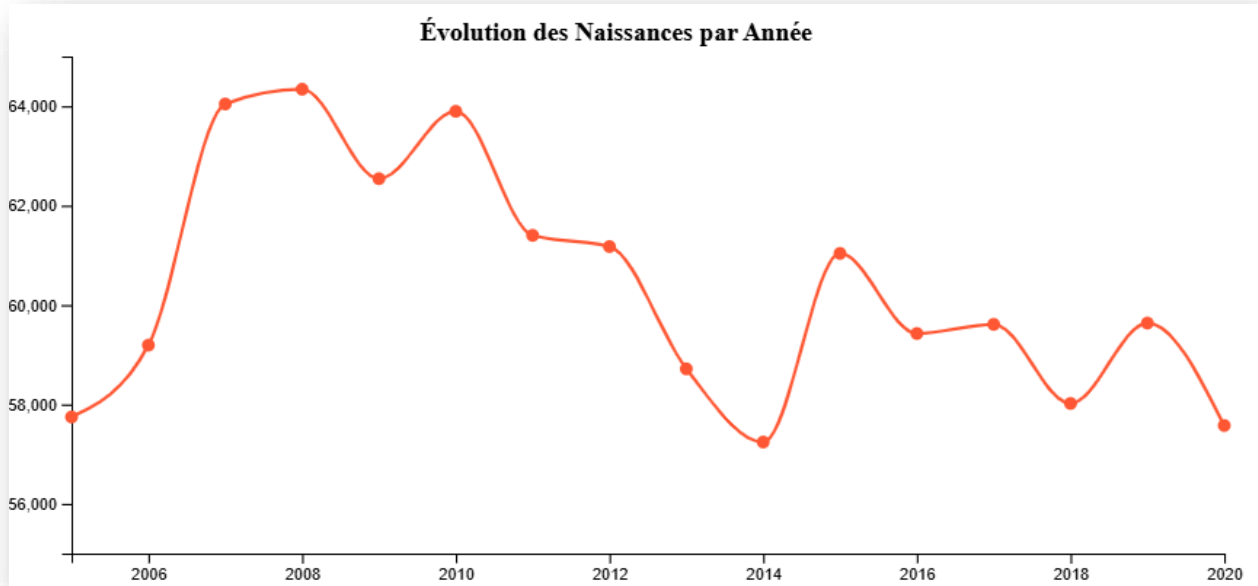
- IMCCV (Industries des Matériaux de Construction, Céramique et Verre) : 71.66MD
  - IME(Industries Mécaniques et Électriques) : 57.10MD
  - ID(Industries Diverses) : 45.29MD
- Ces secteurs reçoivent les investissements les plus élevés, ce qui suggère une priorité

### **Secteurs avec les plus faibles investissements :**

- ICC (Industries du Cuir et de la Chaussure) : 0.23MD
  - IS (Industries du Secteur des Services) : 1.52MD
- Ces faibles valeurs peuvent indiquer des secteurs moins développés, moins prioritaires

## D3.js :

Utilisation du courbe temporelle (Line Chart) avec le fichier Excel 'bd-dec20-births-by mothers-age'



### Détails du graphique :

- Type de graphique : Courbe avec points marqués — utile pour montrer des tendances dans le temps.
- Axe des X (horizontal) : Les années (de 2005 à 2019).
- Axe des Y (vertical) : Le nombre de naissances.

### Interpretation :

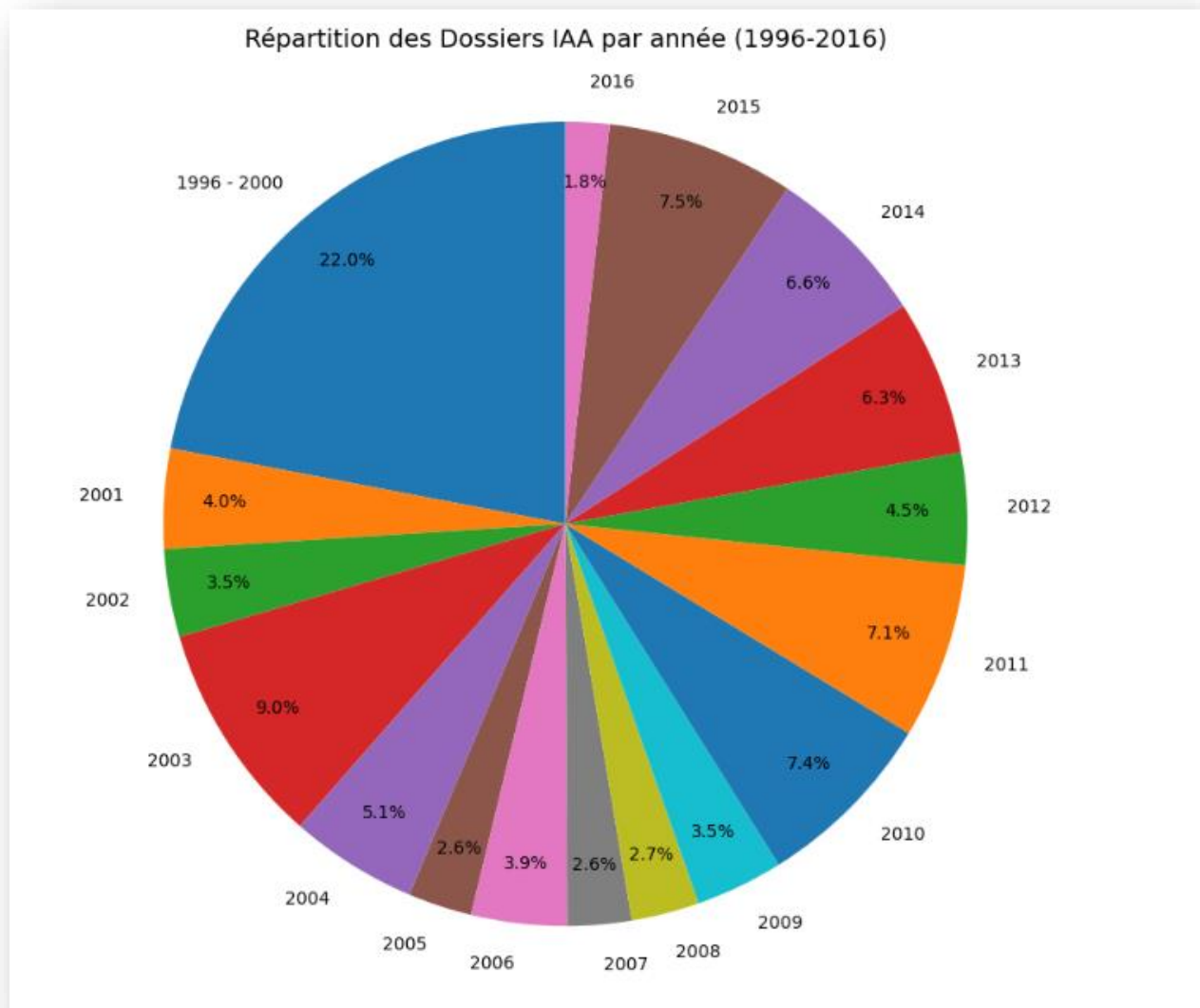
- Période de hausse : Entre 2005 et 2008, forte augmentation du nombre de naissances.
- Période de stabilité : Autour de 2009–2011, des légères variations sans grande baisse.
- Tendance à la baisse : À partir de 2012, on observe une diminution progressive avec quelques reprises ponctuelles (pics en 2015 et 2019).

Globalement : Le nombre de naissances a diminué entre 2008 (le pic) et 2019.

### 4. Étudier une « mauvaise visualisation » (Bad Visualization) :

Utilisation du fichier Excel 'mansecteur'

**Trop de catégories (années) :**



Il y a 17 segments dans ce graphique, ce qui surcharge visuellement le camembert.

L'œil humain a du mal à comparer les tailles d'angles, surtout quand elles sont nombreuses et proches.

Les années comme 2006, 2007, 2008, etc., ont des parts très proches et difficiles à distinguer.

**Inadapté pour une variable temporelle :**



Les années sont ordonnées aléatoirement autour du cercle, ce qui brise la logique temporelle.

Un graphique temporel devrait montrer une évolution dans le temps, ce qui est impossible ici.

### **Effet trompeur sur la perception :**

Des petites différences de valeur peuvent paraître plus grandes ou plus petites selon leur couleur, position ou orientation

Exemple : la différence entre 2015 (7.5%) et 2011 (7.1%) semble plus importante qu'elle ne l'est vraiment.

### **Biais cognitifs induits :**

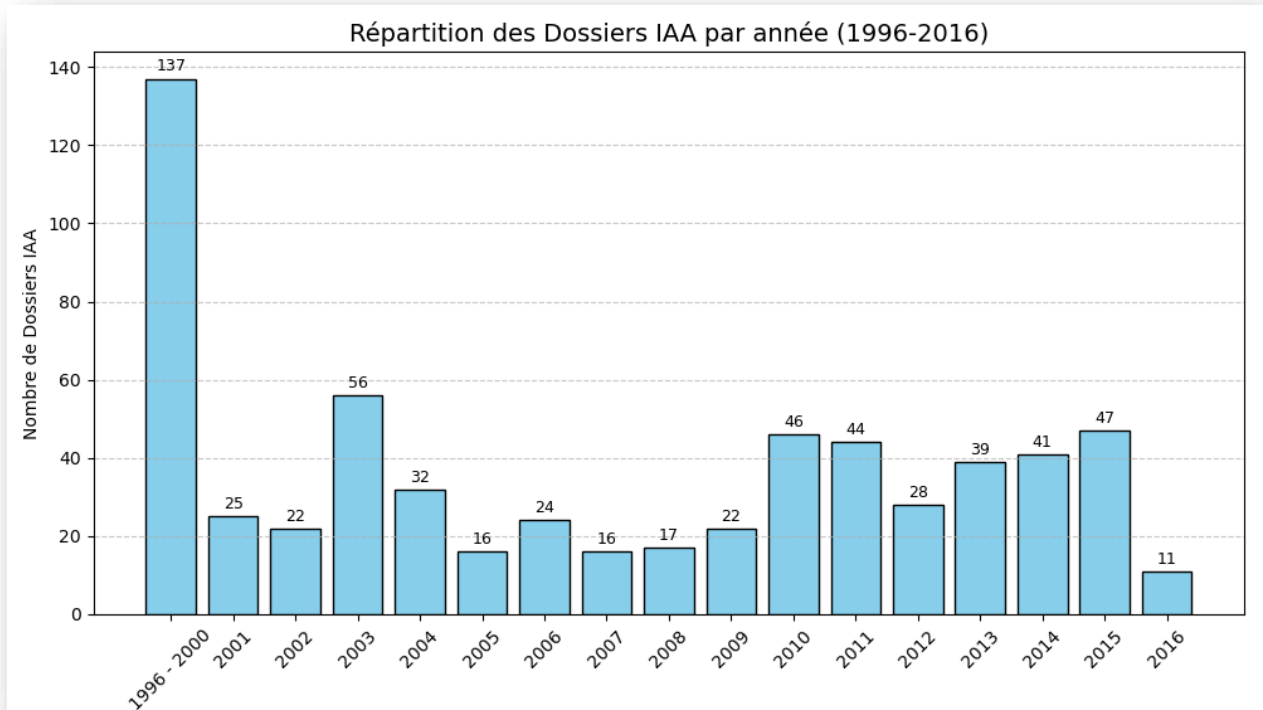
Illusion de proportion : les utilisateurs peuvent surestimer ou sous-estimer certaines parts.

Charge mentale : trop d'informations à traiter d'un coup, ce qui peut décourager l'analyse correcte.

### **Un graphe en barres (bar chart) serait bien plus lisible :**

- Affiche chaque année en ordre chronologique.
- Permet de comparer facilement les valeurs.
- Idéal pour montrer des tendances ou variations au fil des ans.

**Le diagramme en secteur a été remplacé par un diagramme à barres pour une meilleure lisibilité et une interprétation plus claire des données, tout en conservant le même jeu de données :**



- Dans un **diagramme à barres** : - Il est facile de comparer visuellement les hauteurs des barres.

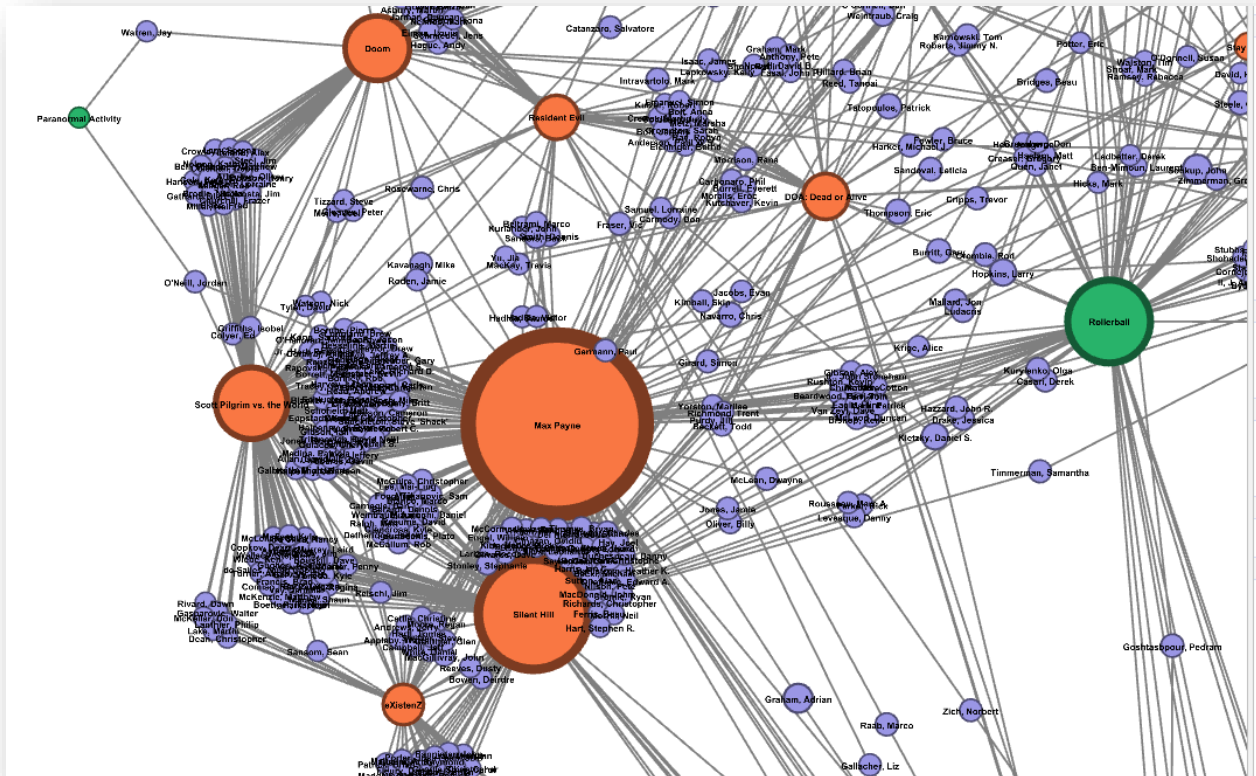
- Il permet de voir facilement une évolution dans le temps ou une hiérarchie.

- Dans un **pie chart** : - les différences d'angles sont plus difficiles à estimer, surtout si les parts sont proches ou nombreuses.

- Le pie chart ne montre pas d'ordre naturel → **pas adapté pour suivre une progression.**

# Gephi :

## Utilisation du fichier Crossed\_Chroma (GEFX graph file)



### Interprétation du graphe :

#### Deux types de nœuds :

- Les cercles orange représentent des films (ex : Silent Hill, eXistenZ).
- Les cercles violets représentent des personnes (acteurs, réalisateurs, etc.).

#### Les arêtes (lignes) :

- Chaque ligne relie une personne à un film.
- Cela signifie que la personne a contribué d'une manière ou d'une autre à ce film.

#### La taille des nœuds :

Les plus gros nœuds sont ceux qui ont plus de connexions.

**Exemple : Silent Hill a une taille plus grande → il est lié à plus de personnes dans ce graphe.**

## Application de l'algorithme ForceAtlas2 :

L'algorithme ForceAtlas2 dans Gephi est un algorithme de disposition spatiale (layout algorithm) conçu pour visualiser les graphes de manière claire et intuitive, en positionnant automatiquement les nœuds (sommets) selon des forces physiques simulées.

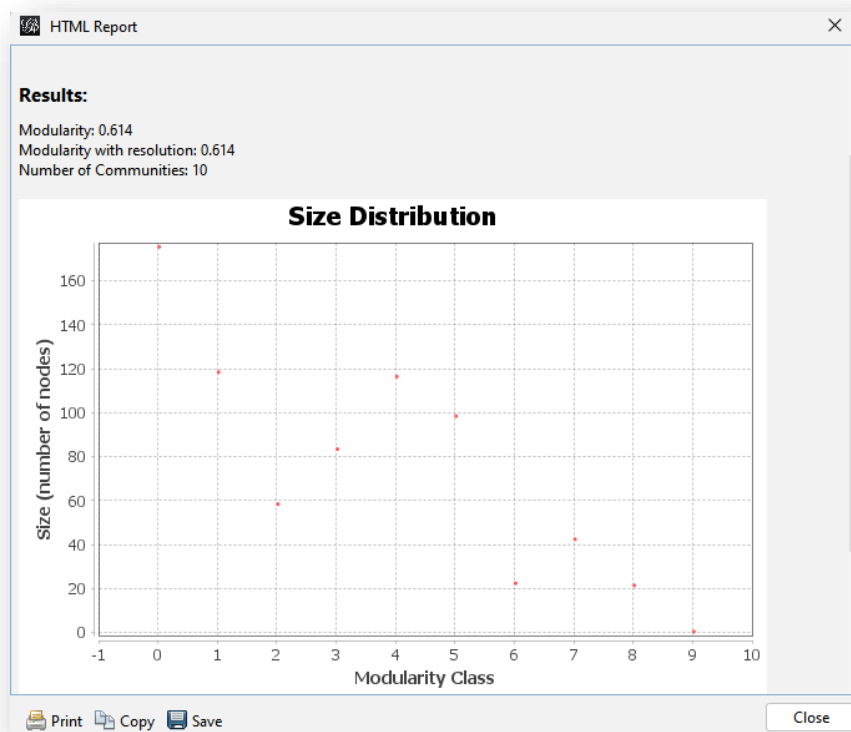
### 1. Attraction et répulsion

- Les nœuds se repoussent entre eux (comme des charges électriques).
- Les arêtes (liens entre les nœuds) agissent comme des ressorts, tirant les nœuds reliés les uns vers les autres.

### 2. Objectif

- Regrouper naturellement les communautés ou clusters (les groupes fortement connectés sont proches).
- Espacer les nœuds peu connectés pour éviter le chevauchement.
- Créer une visualisation intuitive et lisible du graphe.

## Modularity



Il existe 10 grandes zones (cluster) où les sommets sont beaucoup plus connectés entre eux qu'avec les autres.

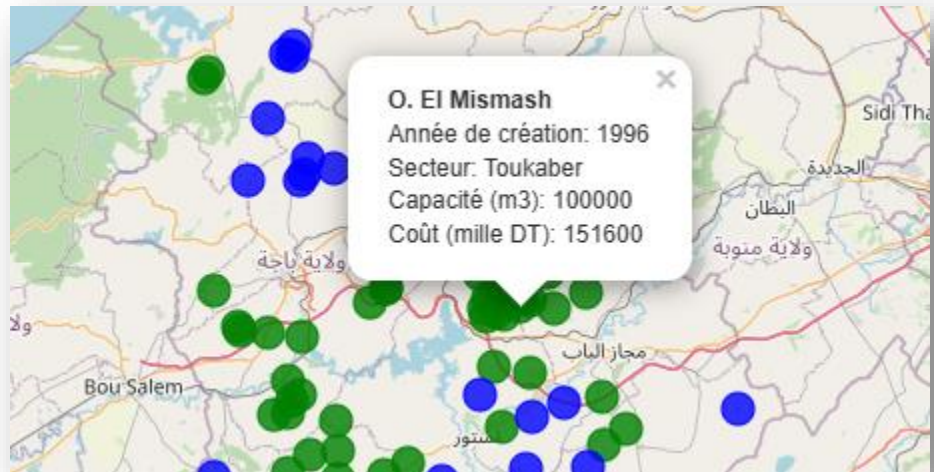
Modularity 0.614 → C'est un signe de structure communautaire significative, mais pas extrême non plus (on considère  $>0.3$  comme significatif généralement).

## Partie 2 : Cartographie thématique de la Tunisie

Les données géographiques utilisées proviennent de fichiers GeoJSON, format adapté aux représentations spatiales sur le web.

- ✓ **Les deux jeux de données concernent :**
  - **Les barrages collinaires (barrages\_collinaires\_beja.geojson)**
  - **Les lacs collinaires (lacs\_collinaires\_beja.geojson)**

Chaque point contient des attributs riches : **nom, année de création, secteur, capacité, coût** .



La carte a été centrée sur la Tunisie avec un zoom adapté (zoom 6) pour englober tous les points.

- ✓ **Utilisation de Leaflet.js, bibliothèque légère**

**et interactive, idéale pour :**

- L'affichage dynamique des points géolocalisés
- L'intégration de popups informatives
- La gestion de couches multiples (barrages en bleu, lacs en vert)

- ✓ **Interactivité et clarté de lecture :**

- **Popups interactifs** : clic sur chaque point → accès aux informations détaillées (année, secteur, coût, etc.).

- **Codage couleur clair** : les deux types d'infrastructures sont distingués facilement.