



# La pollution plastique



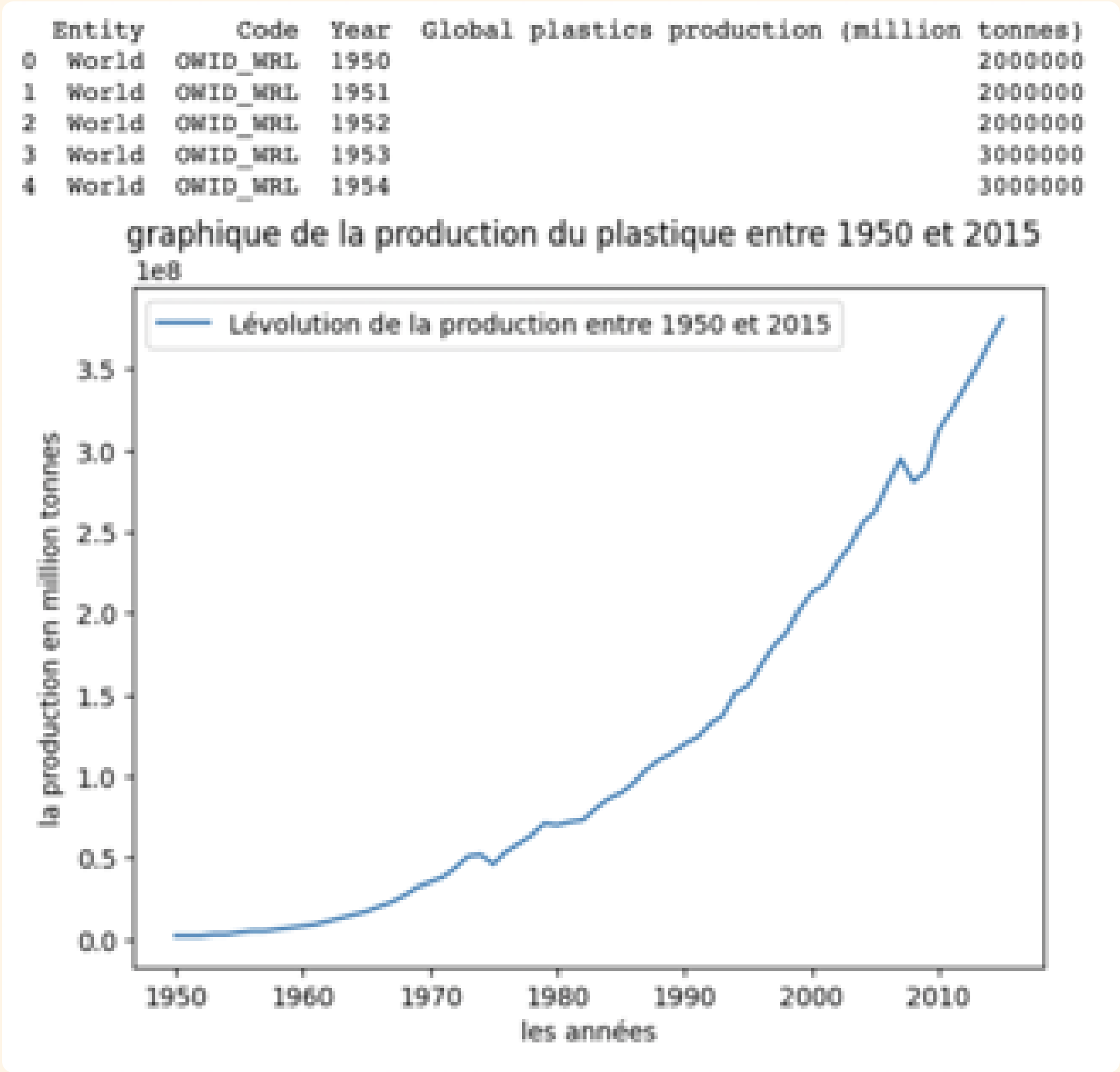
# Introduction :

Les Bases de Données

# Étude de Sefa

```
# -*- coding: utf-8 -*-  
import pandas as pd  
import matplotlib.pyplot as plt  
  
data = pd.read_csv('global-plastics-production.csv')  
  
print(data.head())  
  
plt.plot(data['Year'], data['Global plastics production (million tonnes)'],  
         label='L\'évolution de la production entre 1950 et 2015')  
  
plt.title('graphique de la production du plastique entre 1950 et 2015 ')  
plt.xlabel('les années')  
plt.ylabel('la production en million tonnes')  
  
plt.legend()  
  
plt.show()
```

# Titre : Production de plastique entre 1950 et 2015



# Interprétation :

## .Croissance exponentielle :

La production de plastique a connu une croissance exponentielle entre 1950 et 2015. On peut observer sur le graphique que la quantité de plastique produite a augmenté considérablement au fil du temps, passant de quelques millions de tonnes en 1950 à plus de 300 millions de tonnes en 2015. Cette augmentation soutenue montre l'ampleur du problème de la pollution plastique et souligne la nécessité d'agir rapidement pour trouver des solutions durables.

# La quantité de plastique mal utilisé et gaspillé

```
import pandas as pd
import matplotlib.pyplot as plt

# Charger le fichier CSV dans un DataFrame
fichier_csv= 'fichierproj.csv'
donnees = pd.read_csv(fichier_csv,sep=";")
# Trier le DataFrame par quantité de gaspillage de plastique
donnees_triees = donnees.sort_values(by='quant de plastique gaspille(kg/person/day)',ascending=False)

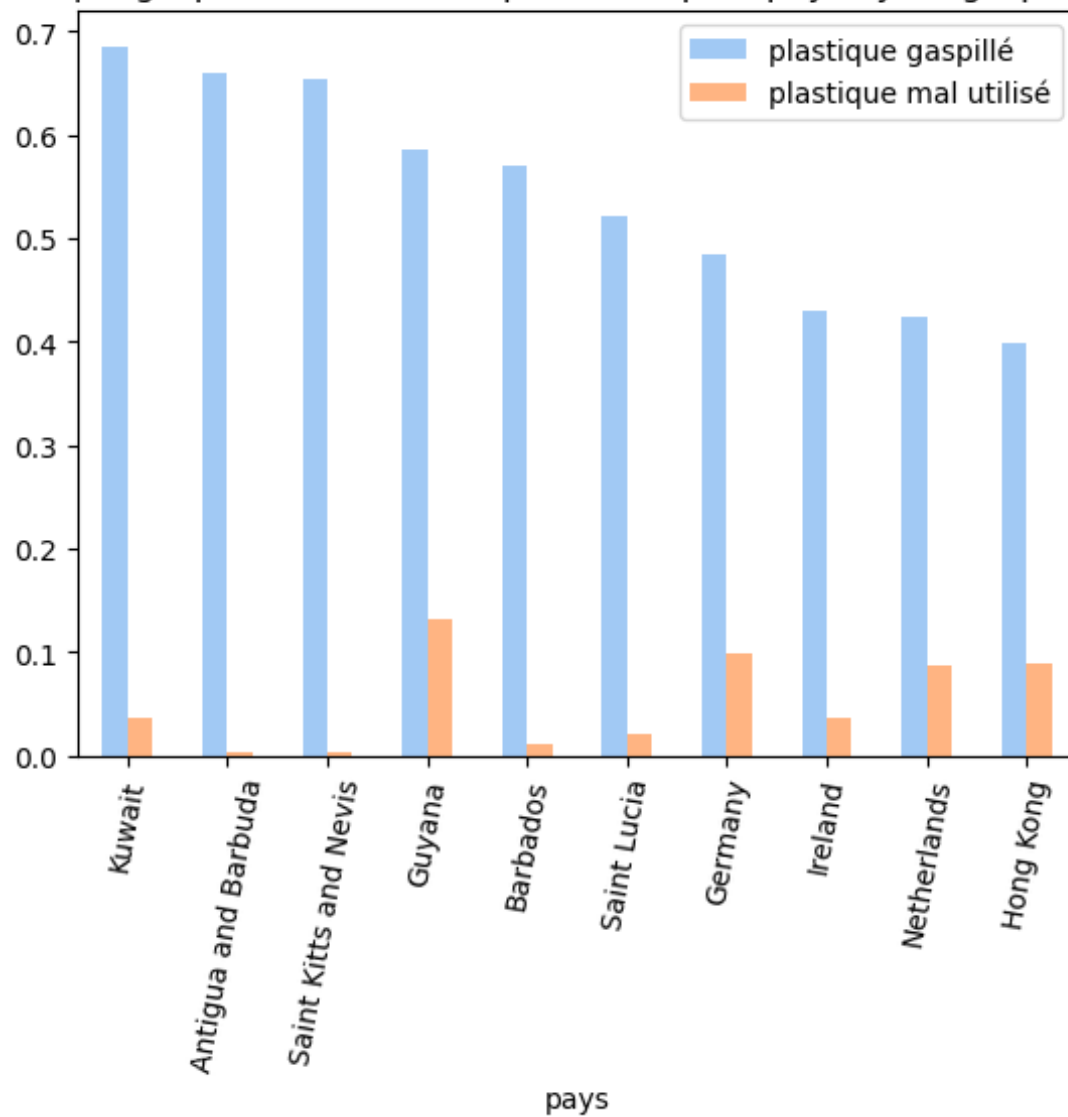
# Sélectionner les 10 premières lignes
top_10_pays = donnees_triees.head(10)

# Afficher les résultats
print("Top 10 des pays ayant gaspillé le plus de plastique en 2010 :\n", top_10_pays)

mydata = pd.DataFrame({"plastique gaspillé":top_10_pays['quant de plastique gaspille(kg/person/day)'],
                      "plastique mal utilisé":top_10_pays['plastique mal utilise (% global total)']})
mydata.index=top_10_pays['pays']
from pandas import plotting
mydata.plot(kind="bar",rot=80)
plt.title('la quantité de plastique gaspillé et mal utilisé pour les top 10 pays ayant gaspillé le plus de plastique')

plt.show()
```

la quantité de plastique gaspillé et mal utilisé pour les top 10 pays ayant gaspillé le plus de plastique





# Étude de Chafik



# Code utilisé pour le premier graphe :

```
Entrée [8]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

J'ouvre le fichier puis je trie les données du plus grand au plus petit.

```
Entrée [9]: malutilise = pd.read_csv('mismanaged-waste-global-total.csv')
malutilise= malutilise.sort_values(by='Mismanaged waste (% global total)', ascending=False)
```

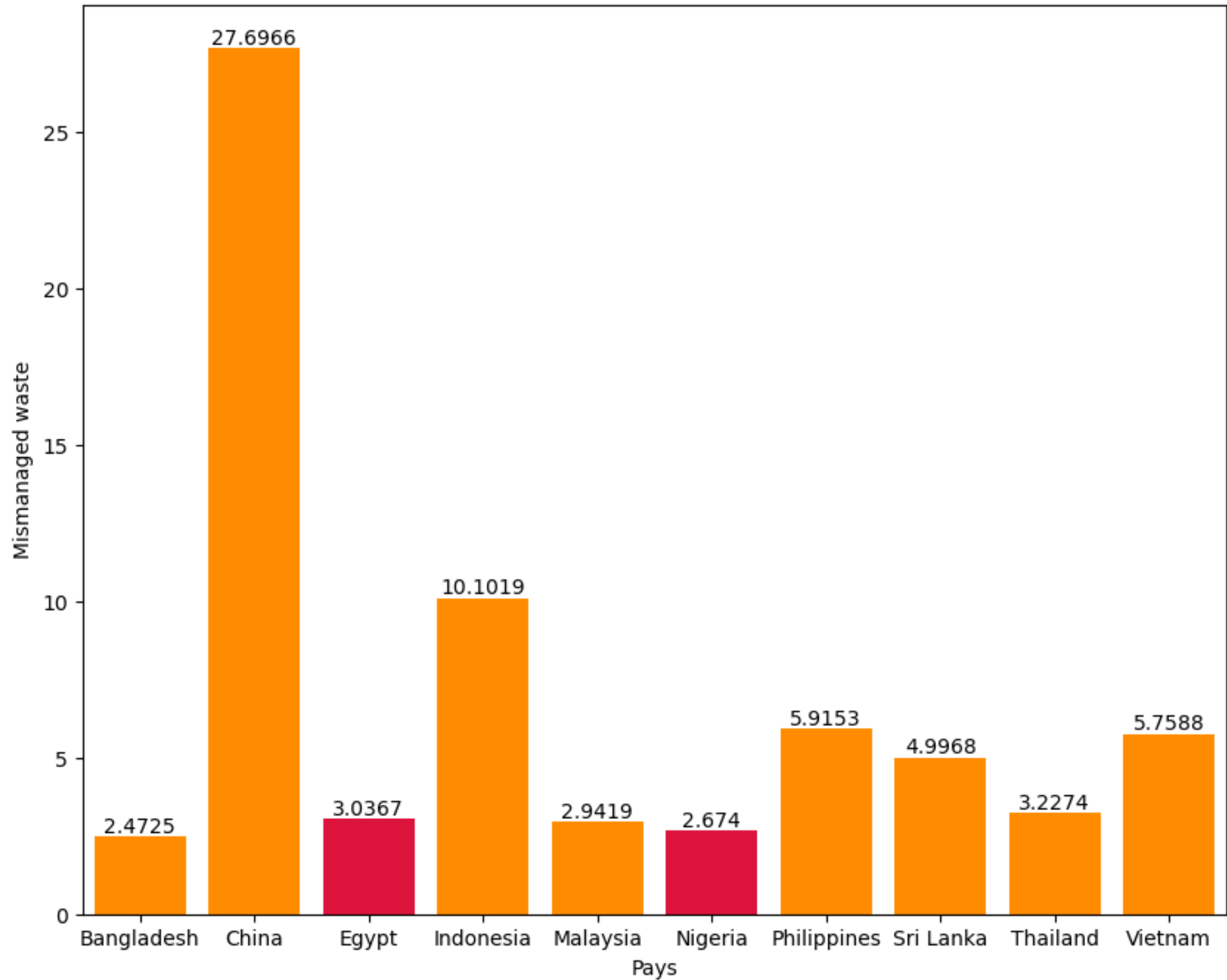
Je garde les 10 premiers pays. Je renomme deux colonnes.

```
Entrée [10]: malutilise=malutilise.head(10)
malutilise.rename(columns={'Entity' : 'Pays'}, inplace=True)
malutilise.rename(columns={'Mismanaged waste (% global total)' : 'Mismanaged waste'}, inplace=True)
```

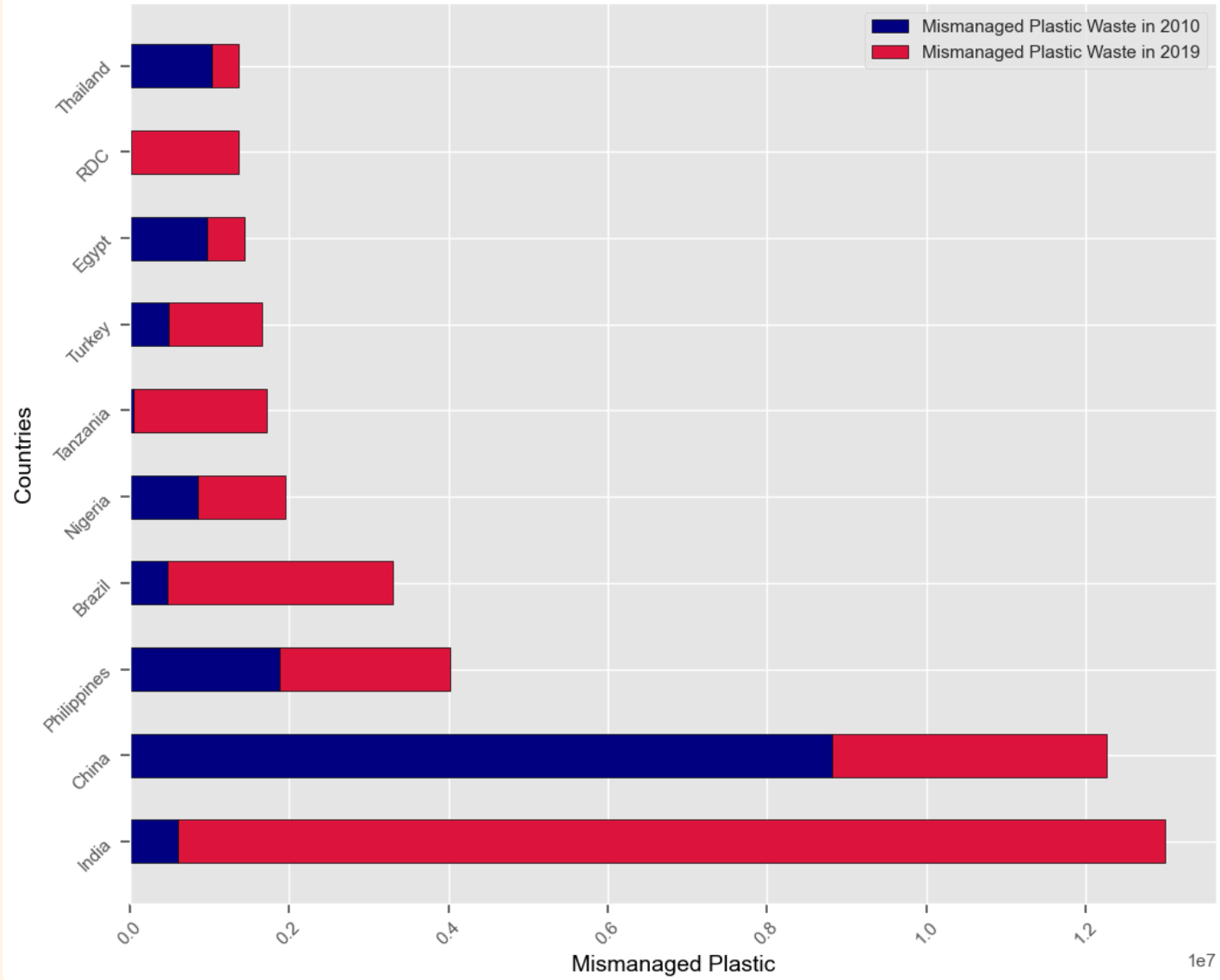
Je crée le graph avec seaborn et je change de couleur pour les pays africains.

```
Entrée [11]: plt.figure(figsize=(10, 8))
plt.title('Le plastique mal géré selon les pays en 2010')
ax = sns.barplot(x=malutilise['Pays'], y=malutilise['Mismanaged waste'], color='crimson', estimator="sum", errorbar=None)
ax.bar_label(ax.containers[0], fontsize=10);
barres = plt.gca().patches
barres[6].set_facecolor('orange')
barres[8].set_facecolor('orange')
```

Le plastique mal utilisé selon les pays en 2010



Plastique mal utilisé entre 2010 et 2019



# Code pour le deuxième graphique :

```
Entrée [14]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib import style
```

Je lis le fichier puis je trie pour avoir le Top 10 des pays qui ont le moins bien géré leurs plastique en 2019.

```
Entrée [15]: df = pd.read_csv('mismanaged_plasticwaste.csv')
df2 = df.sort_values(by='Total_MismanagedPlasticWaste_2019 (millionT)', ascending=False)
```

Je garde que les 10 premiers pays :

```
Entrée [16]: df2 = df_sorted.head(10)
```

Je renomme les deux colonnes et une ligne dans la colonne des pays :

```
Entrée [17]: df2.rename(columns={'Total_MismanagedPlasticWaste_2010 (millionT)': 'Mismanaged Plastic Waste in 2010'}, inplace=True)
df2.rename(columns={'Total_MismanagedPlasticWaste_2019 (millionT)': 'Mismanaged Plastic Waste in 2019'}, inplace=True)
df2['Country'] = df2['Country'].replace('Democratic Republic of Congo', 'RDC')
```

Je crée un graphique à barres superposées, et pour la deuxième bar je fais une soustraction ("df2['Mismanaged Plastic Waste in 2019']-df2['Mismanaged Plastic Waste in 2010']") pour que les valeurs ne s'additionnent pas.

Je rajoute des informations, j'ajuste la taille de la police et je fais pivoter les valeurs de l'axe des abscisses et des ordonnées.

```
Entrée [20]: style.use('ggplot')
plt.figure(figsize=(12,10))
w = 0.5
plt.barh(df2['Country'], df2['Mismanaged Plastic Waste in 2010'],w, color='navy',edgecolor='black', label="Mismanaged Plastic
plt.barh(df2['Country'], df2['Mismanaged Plastic Waste in 2019']-df2['Mismanaged Plastic Waste in 2010'],w,edgecolor='black',l
plt.legend()
plt.xlabel('Mismanaged Plastic',color = 'black', fontsize=14)
plt.ylabel('Countries',color = 'black', fontsize=14)
plt.title('Plastique mal utilisé entre 2010 et 2019', fontsize=14)
plt.xticks(rotation=45)
plt.yticks(rotation=45)
plt.show()
```

# Étude de Mohamed-Yacine Lassouani

Mais qu'est qu'on en fait de tout ce plastique ?

# Analyse de l'utilisation de plastique en Chine en Inde et aux États-Unies

Pour ce faire j'ai décidé de créer des camemberts. C'est le graphique le plus adapté à ce genre d'étude.

# Code utilisé :

```
[5] #Importation du jeu de données
df = pd.read_csv('pupc.csv')

[7] #Sélection des Pays intéressants :
selected_countries = df[df['Pays'].isin(['United States', 'China', 'India'])]
print(selected_countries.head())

[8] # Ici j'ai enlevé les colonnes non pertinentes, ici la colonne Pays et la colonne Total
selected_data = selected_countries.iloc[:, 1:-1]

[9] #Dans cette ligne de code j'applique la fonction "lambda", qui remplacera les , par des . et convertira le tout en float
selected_data = selected_data.apply(lambda x: x.str.replace(',', '.').astype(float), axis=1)

[10] #Je souhaite réaliser un graphe sur le total, pour ce faire j'applique la fonction sum à toutes les colonnes du dataframe :
total_data = selected_data.iloc[:, 1:-1].sum()
```

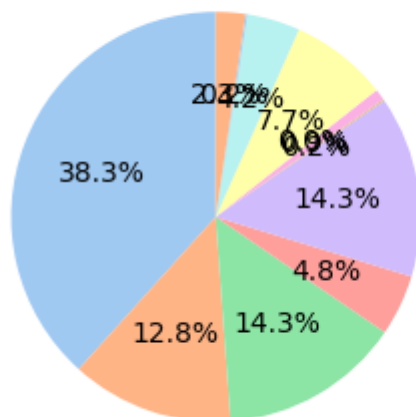
Création du premier graphe :

+ Code

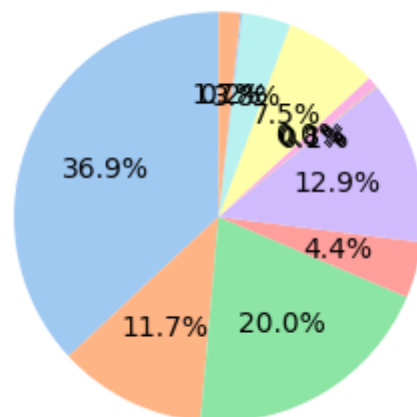
+ Texte

```
[29] fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(8, 8))
#une boucle for pour créer les 2 premiers diagrammes (j'aurais pu copié collé c'est vrai)
for i in range(2):
    ax = axes[0, i]
    ax.pie(selected_data.iloc[i, 1:], autopct='%1.1f%%', startangle=90) #ici le autopct='%1.1f%%' détermine le nombre de chiffre après la virgule des mes données affichées.
    ax.set_title(selected_countries.iloc[i, 0])
ax = axes[1, 0]
ax.pie(selected_data.iloc[2, 1:], autopct='%1.1f%%', startangle=90)
ax.set_title(selected_countries.iloc[2, 0])
ax = axes[1, 1]
sns.set_palette("pastel")
ax.pie(total_data, autopct='%1.1f%%', startangle=90)
ax.set_title('Somme des colonnes')
ax.legend(total_data.index, title="Légende", loc="upper left", bbox_to_anchor=(1, 0.5), fontsize='small') #je détermine la position de la légende
plt.tight_layout() # cette fonction matplotlib me permet d'éviter un chevauchement des graphiques
plt.show()
```

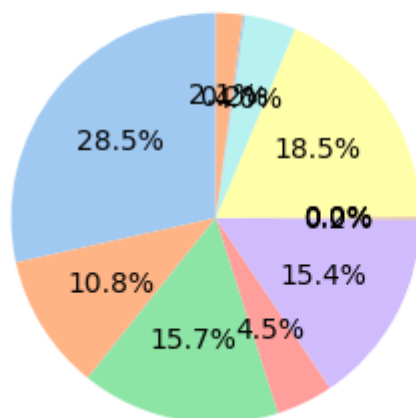
United States



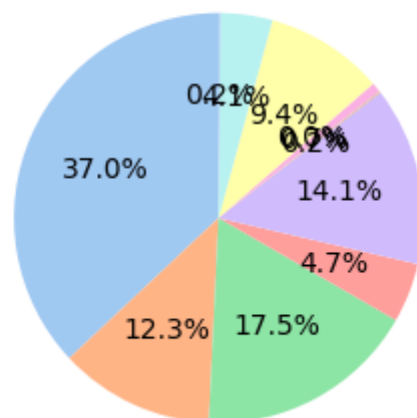
China



India



Somme des colonnes



#### Légende

- Packaging
- Consumer & institutional Products
- Building & construction
- Electrical/electronics
- Transportation - other
- Marine coatings
- Industrial/machinery
- Personal care products
- Textile sector - clothing
- Textile sector - others
- Road marking



# Interprétation :

.Utilisation excessive du plastique dans l'emballage des produits.

similitudes frappantes avec la politique économique de ces pays.

.Chine/USA des chiffres qui choquent.

# Kmeans

Études des Kmoyens pour essayer de trouver une ressemblance entre les pays et le plastique jeté

# Premier problème...

Jeu de données non compatible.

# La solution :

Changement du jeu de donnée pour prendre le même que celui de Chafik.

# Le code :

```
#je charge les données
df = pd.read_csv('mismanaged_plasticwaste.csv')

#Dans mon fichier csv je n'ai pas de code Pays, j'ai donc ajouté cette fonction pour rajouter ce code.
def findCountry(country_code):
    try:
        return pycountry.countries.get(name=country_code).alpha_3
    except:
        return 'NaN'

df['country_code'] = df.apply(lambda row: findCountry(row.Country), axis=1) #Je rajoute les codes pays.

#Clustering

#J'enlève les valeurs non numérique de mon dataframe
df1 = df.drop(['Country', 'country_code'], axis=1)

#
kmeans_model = KMeans(n_clusters=5, random_state=10).fit(df1)
labels = kmeans_model.labels_
df1['cluster'] = labels #Ici je créer un 2nd dataframe avec les countrycode pour pouvoir afficher ma carte correctement.
df1['Country'] = df['Country']
df1['country_code'] = df['country_code']

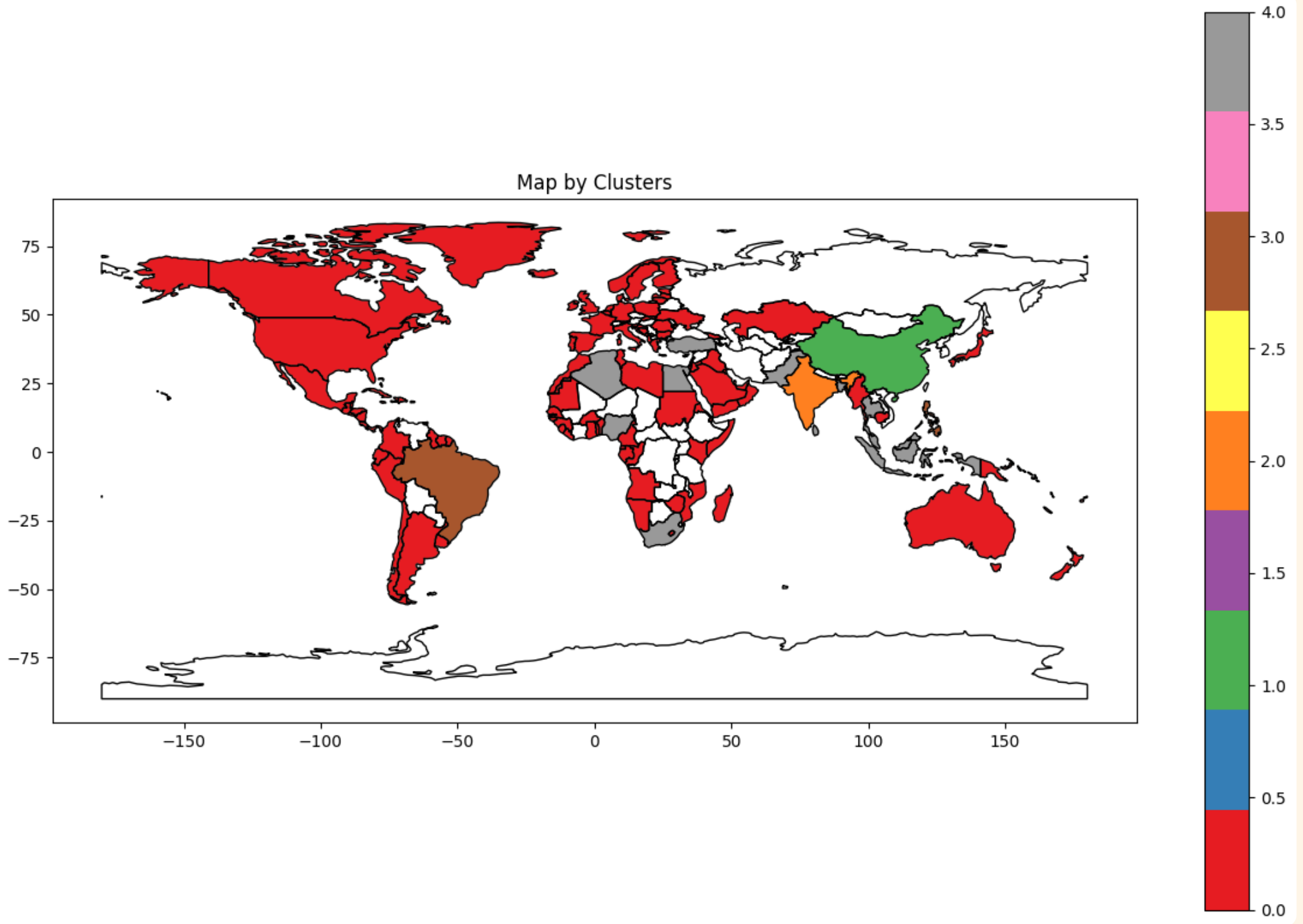
# Créer un GeoDataFrame sans coordonnées géographiques
gdf = gpd.GeoDataFrame(df1)

# Charger une carte du monde avec GeoPandas
world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))

# Fusionner votre DataFrame avec la carte du monde
merged = world.set_index('iso_a3').join(gdf.set_index('country_code'))

# Tracer la carte
fig, ax = plt.subplots(1, 1, figsize=(15, 10))
world.boundary.plot(ax=ax, linewidth=1, color='black')
merged.plot(column='cluster', cmap='Set1', linewidth=0.8, ax=ax, edgecolor='0.8', legend=True)
ax.set_title('Map by Clusters')

# Afficher la carte
plt.show()
```

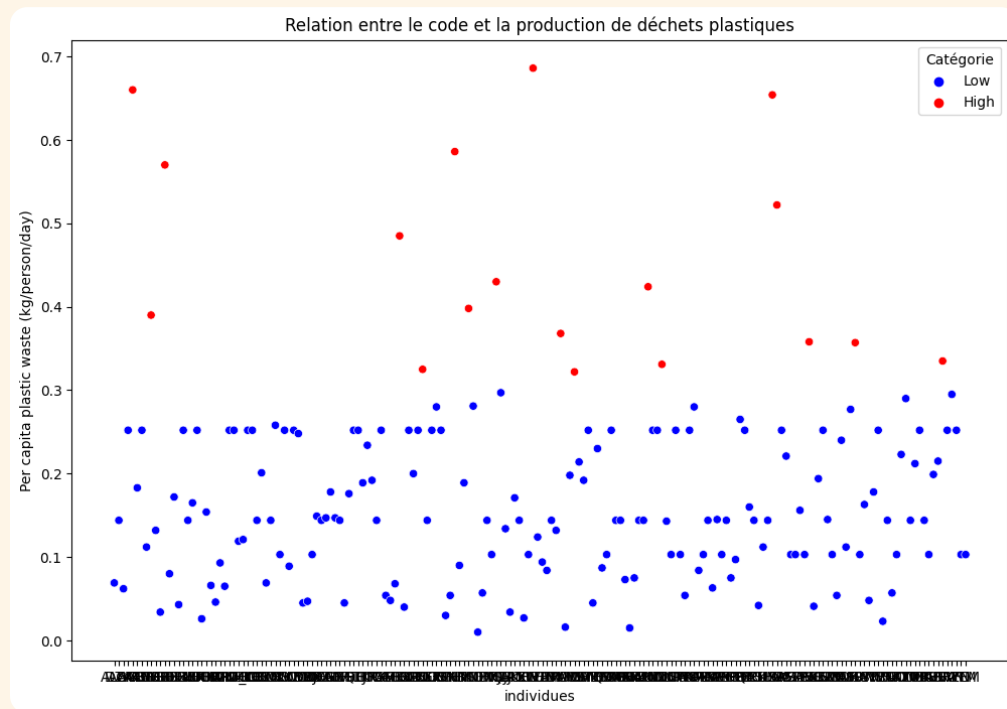


# Interprétation :

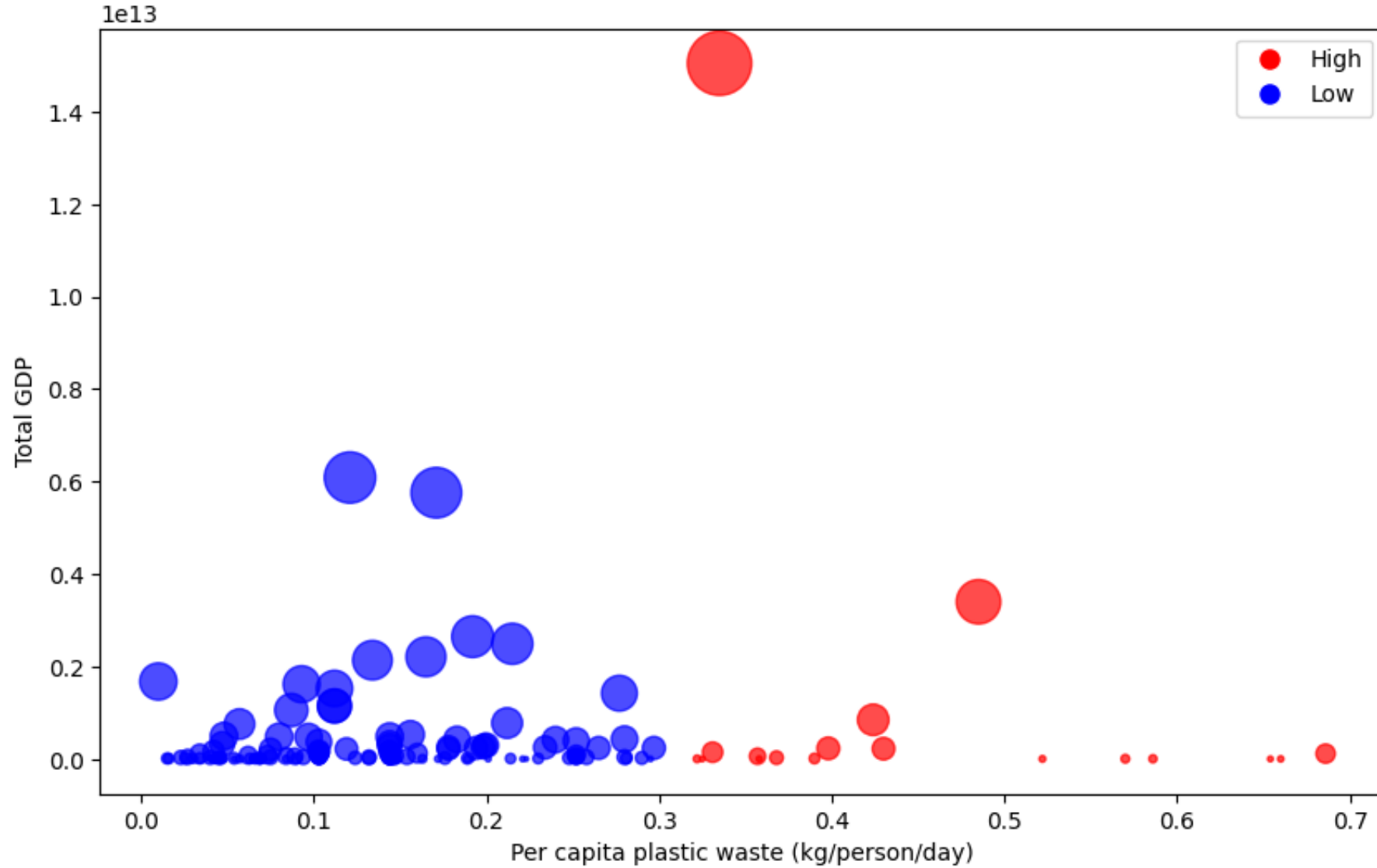
- .les pays d'Asie du Sud se retrouvent regroupés
- .Pays d'Afrique Subsaharienne dans le même cluster que les pays européens
- .Les limites de mon jeu de données.

# ETUDE de Chamss





Pictogramme en fonction de la production de déchets plastiques, du PIB total et de la catégorie



Le projet de science des données présenté vise à explorer et représenter graphiquement des données cruciales concernant la quantité de déchets plastiques générés par pays à l'échelle mondiale. La problématique croissante des déchets plastiques constitue un défi environnemental majeur, suscitant un intérêt accru pour la recherche de solutions durables. Dans cette démarche, l'analyse des données offre une perspective précieuse pour comprendre la répartition géographique, les tendances temporelles et les corrélations potentielles entre la production de déchets plastiques et d'autres indicateurs économiques.

L'absence apparente de corrélation entre la production de déchets plastiques par habitant et le Produit Intérieur Brut (PIB) dans la figure peut indiquer que d'autres facteurs non économiques jouent également un rôle significatif. Les données pourraient être influencées par des variables environnementales, politiques ou sociales qui ne sont pas directement reflétées dans le PIB.

Une analyse plus approfondie pourrait inclure l'examen d'autres caractéristiques, telles que les politiques de gestion des déchets, les comportements de consommation, les initiatives environnementales, etc. De plus, des analyses statistiques plus avancées, telles que des corrélations plus spécifiques ou des modèles de régression, pourraient être utilisées pour identifier des relations subtiles entre les variables.

En fin de compte, cette observation initiale sert de point de départ pour une exploration plus approfondie, encourageant une approche holistique de l'analyse des données afin de saisir la complexité des facteurs qui influent sur la production de déchets plastiques à l'échelle mondiale.

# Conclusion de notre Étude :

Cette analyse m'a fait réalisé les points suivants quant à la production et aux gaspillage de plastiques dans le monde :

Complexité de la question : La gestion des déchets plastiques est une question complexe qui va au-delà des seuls volumes de production. Les pratiques de gestion des déchets varient selon les économies et les régions, ajoutant une dimension supplémentaire à la complexité du problème.

Approche multifacette : Pour mener à bien cette étude et pour qu'elle soit la plus pertinente possible il y a nécessité d'une approche multifacette pour lutter contre la pollution plastique, impliquant à la fois la production et les pratiques d'élimination. Cela indique une compréhension approfondie de la diversité des facteurs contribuant au problème.

Portée mondiale : la pollution plastique est un problème mondial qui nécessite des solutions coordonnées à l'échelle internationale.