

The background of the slide is a dark blue field filled with numerous 3D renderings of COVID-19 virus particles. These particles are spherical with a textured surface and are covered in many long, thin, spike-like protrusions (glycoprotein spikes) that give them a crown-like appearance. The particles are scattered across the frame, with some appearing larger and more detailed than others, creating a sense of depth and a microscopic environment.

# ANALYSE DES DONNÉES DE LA COVID 19

---

ESSOMBA LENJI Chris Manuel

26/05/2023



- Introduction
- Méthodologie
- Résultats
- Conclusion

---

## SOMMAIRE

# #Content



# INTRODUCTION

---

- Le présent projet vise à analyser les données relatives à la pandémie de COVID-19, une maladie infectieuse qui a profondément marqué le monde depuis son émergence en 2019. Face à cette crise sanitaire d'envergure mondiale, il est essentiel de comprendre les multiples facettes de cette pandémie et d'explorer les questions clés qui se posent :
- Quels sont les facteurs influençant la propagation et la gravité du COVID-19 ?
- Quel est l'effet de la vaccination sur la propagation et la mortalité liées au virus ?

# MÉTHODOLOGIE

Partie 0: Importation des données actualisées datant du 15 Mai 2023 depuis le site web OurWorldInData

Partie 1: Nettoyage des données :

- Identification et suppression des valeurs manquantes.
- Suppression des doublons.
- Prétraitement des données pour assurer leur cohérence et leur qualité.

Partie2 : Analyse des données sur PostgreSQL :

- Importation des données nettoyées dans une base de données PostgreSQL.
- Utilisation de requêtes SQL pour extraire des informations spécifiques, telles que le nombre de cas confirmés et de décès par année, le pourcentage de personnes infectées par région, etc.

Partie 3: Visualisations sur Tableau :

- Transfert des données analysées depuis PostgreSQL vers Tableau.
- Création de différentes visualisations, telles que des cartes interactives, des graphiques, et des comparaisons temporelles.
- Utilisation des fonctionnalités de Tableau pour explorer et présenter les tendances et les corrélations identifiées.



# ETAPE 1 : NETTOYAGE DES DONNÉES SUR JUPYTER NOTEBOOK ET POSTGRESQL

---

# SUPPRESSION DES VALEURS NULLES DANS LA BASE DE DONNÉES DES DÉCÈS

- Nous avons supprimé toutes les valeurs nulles de la colonne "continent" car leur localisation ne semblait pas pertinente.
- Nous avons choisi de remplacer les valeurs nulles dans les colonnes commençant par "new" par 0, ce qui indique l'absence de nouveaux cas enregistrés ce jour.
- Pour les autres colonnes, nous avons opté pour le remplacement par la valeur précédente du même pays, et si cette valeur n'est pas disponible, elle est remplacée par la valeur suivante.
- Les valeurs nulles pour toutes les lignes d'un pays ont été remplacées par -1, indiquant ainsi leur non-disponibilité.
- Le lien vers le notebook est le suivant  
<https://github.com/ChrisEssomba/Analyse-des-donnees-de-la-Covid-19/blob/main/covidDeathsCleaning.ipynb>

```
# Tri des données par pays et date
df.sort_values(['location', 'date'], inplace=True)

# Liste des colonnes numériques
numeric_columns = df.select_dtypes(include=['float64']).columns

# Groupe par pays
grouped = df.groupby('location')

# Boucle for pour remplacer les valeurs null
for col in numeric_columns:
    if col.startswith('new'):
        # Remplacement des valeurs null par 0 pour les colonnes commençant par 'new'
        df[col].fillna(0, inplace=True)
    else:
        # remplacement des valeurs null par la valeur precedente ou suivante du meme pays
        group_mask = df[col].isnull().groupby(df['location']).transform('any')
        df[col] = df.groupby('location')[col].ffill().bfill().where(group_mask, df[col])

# Affichage du DataFrame mis à jour
print(df.isnull().sum())
```

iso_code	0
continent	0
location	0
date	0
population	0
total_cases	0
new_cases_smoothed	0
total_deaths	0
new_deaths	0
new_deaths_smoothed	0
total_cases_per_million	0
new_cases_per_million	0



# SUPPRESSION DES VALEURS NULLES DANS LA BASE DE DONNÉES DES VACCINATIONS

- Le même principe de nettoyage expliqué précédemment a été appliqué dans ce cas.
- Le lien vers le Notebook :  
<https://github.com/ChrisEssomba/Analyse-des-donnees-de-la-Covid-19/blob/main/covidVaccinationCleaning.ipynb>

```
[7]: # Tri des données par pays et date
df.sort_values(['location', 'date'], inplace=True)

# Liste des colonnes numériques
numeric_columns = df.select_dtypes(include=['float64']).columns

# Groupe par pays
grouped = df.groupby('location')

# Boucle for pour remplacer les valeurs null
for col in df:
    if col.startswith('new'):
        # Remplacement des valeurs null par 0 pour les colonnes commençant par 'new'
        df[col].fillna(0, inplace=True)
    else:
        #remplacement des valeurs null par la valeur precedente ou suivante du meme pays
        group_mask = df[col].isnull().groupby(df['location']).transform('any')
        df[col] = df.groupby('location')[col].ffill().bfill().where(group_mask, df[col])

# Affichage du DataFrame mis à jour
print(df.isnull().sum())
```

iso_code	0
continent	0
location	0
date	0
total_tests	0
new_tests	0
total_tests_per_thousand	0
new_tests_per_thousand	0
new_tests_smoothed	0
new_tests_smoothed_per_thousand	0
positive_rate	0
tests_per_case	0
tests_units	0
total_vaccinations	0
people_vaccinated	0
people_fully_vaccinated	0
new_vaccinations	0
new_vaccinations_smoothed	0
total_vaccinations_per_hundred	0
people_vaccinated_per_hundred	0
people_fully_vaccinated_per_hundred	0
new_vaccinations_smoothed_per_million	0
new_people_vaccinated_smoothed	0
new people vaccinated smoothed per hundred	0

# ÉLIMINATION DES DOUBLONS DANS LES TABLES DES VACCINATIONS ET DES DÉCÈS

- Nous avons entrepris d'éliminer les valeurs en double de notre table en suivant les étapes suivantes :
- Nous avons créé une copie temporaire de notre table qui contient uniquement les valeurs uniques des champs de la table.
- Ensuite, nous avons supprimé notre table d'origine.
- Nous avons donné à la copie temporaire le même nom que notre table d'origine pour la remplacer.
- Le lien vers le notebook est le suivant :  
<https://github.com/ChrisEssomba/Analyse-des-donnees-de-la-Covid-19/blob/main/AnalysisRequests.sql>

```
--Etapes pour supprimer les doublons de date dans la table des vaccinations
-- Créer une copie temporaire de la table
CREATE TABLE temp_table AS
SELECT DISTINCT *
FROM covidvaccinations;
-- Supprimer l'ancienne table
DROP TABLE covidvaccinations;
-- Renommer la copie temporaire avec le nom original
ALTER TABLE temp_table RENAME TO covidvaccinations;

--Etapes pour supprimer les doublons de date dans la table des deces
-- Créer une copie temporaire de la table
CREATE TABLE temp_table AS
SELECT DISTINCT *
FROM coviddeaths;
-- Supprimer l'ancienne table
DROP TABLE coviddeaths;
-- Renommer la copie temporaire avec le nom original
ALTER TABLE temp_table RENAME TO coviddeaths;
```



The image features a magnifying glass held over a bar chart. The chart displays two data series, one in blue and one in green, across four quarters labeled Q1, Q2, Q3, and Q4. The magnifying glass is positioned over the Q2 and Q3 bars, making them more prominent. The text 'ETAPE 2: ANALYSE DES DONNÉES SUR POSTGRESQL' is overlaid in the center of the image, with a horizontal line underneath it. A '1,000' label is visible on the right side of the chart, indicating a scale.

## ETAPE 2: ANALYSE DES DONNÉES SUR POSTGRESQL

---

# ÉVOLUTION ANNUELLE DES CAS CONFIRMÉS ET DES DÉCÈS LIÉS À LA COVID-19

- Globalement, nous observons une diminution du taux de mortalité au fil des années pour la pandémie de COVID-19. Cependant, une observation intéressante est que le nombre de cas enregistrés en 2020 est nettement inférieur à ceux de 2021 et 2022, mais le pourcentage de décès est significativement plus élevé. Cette constatation soulève une fois de plus la question des facteurs qui ont influencé cette régression du taux de mortalité.

```
143 SELECT EXTRACT(YEAR FROM date) AS year, SUM(new_cases) AS total_cases,  
144 SUM(new_deaths) AS total_deaths, sum(new_deaths)/sum(new_cases)*100 as deathsPercentage  
145 FROM coviddeaths  
146 ---WHERE EXTRACT(YEAR FROM date) >= 2020  
147 WHERE continent IS NOT null AND new_cases <> 0  
148 GROUP BY EXTRACT(YEAR FROM date);  
149  
150  
151
```

	year numeric	total_cases double precision	total_deaths double precision	deathpercentage double precision
1	2020	82936738	1941634	2.3411024436480727
2	2021	204215925	3527777	1.7274739959677483
3	2022	445708066	1217362	0.27312990113129343
4	2023	33058821	211493	0.6397475578454538

# ANALYSE COMPARATIVE DES CAS CONFIRMÉS ET DU POURCENTAGE DE PERSONNES INFECTÉES PAR RÉGION DANS LE MONDE

- Lorsqu'on analyse la corrélation entre le nombre de cas confirmés et le nombre de personnes testées, une observation frappante se dégage : le pourcentage d'infection par la Covid-19 est de 40% pour les Sud-Américains et de 21% pour les Africains. Ces chiffres laissent supposer que la génétique pourrait jouer un rôle dans la susceptibilité à la contamination par la Covid-19.

```
152 select continent, SUM(new_cases) AS total_cases,  
153 sum(new_cases)/sum(new_tests)*100 as Infection_percentage  
154 from coviddeaths  
155 where continent is not null  
156 group by continent  
157 order by 3 desc  
158
```

Data Output Messages Notifications			
	continent text	total_cases double precision	infection_percentage double precision
1	South America	68541063	41.908630428243676
2	Africa	13085807	21.26638631654435
3	Oceania	14033973	17.898313585102493
4	Asia	296997904	16.515137229903534
5	Europe	249194673	12.749569733497731
6	North America	124066130	12.197314786958733

# IMPACT DE LA DENSITÉ DE POPULATION SUR LE TAUX DE CONTAGION DE LA COVID-19

- Une observation intéressante émerge de nos résultats : les continents ayant une densité de population élevée présentent en réalité un taux de contagion plus faible par rapport à ceux ayant une densité plus faible. Cela suggère que la densité de la population ne semble pas être un facteur déterminant influençant le taux de contagion de la Covid-19.

```
159 --Decouvrons la densité de la population et le taux de contamination
160 select continent, AVG(population_density) AS population_density,
161 sum(new_cases)/sum(new_tests)*100 as InfectionsPercentage
162 from coviddeaths
163 where continent is not null and new_cases <> 0
164 group by continent
165 order by 2, 3 desc
166
```

Data Output Messages Notifications

	continent text	population_density double precision	infectionspercentage double precision
1	South America	24.923395264598522	42.188208873943445
2	Oceania	90.79992433577225	46.31219083484138
3	Africa	107.21987189860397	21.731258551536616
4	North America	246.9993029063091	12.22087107245021
5	Asia	481.8924413517921	16.653113954756485
6	Europe	567.6139105306086	12.75385773495558



# ANALYSE DE LA CORRÉLATION ENTRE LE POURCENTAGE DE PERSONNES ÂGÉES ET LE TAUX DE MORTALITÉ

- Lorsque l'on compare le pourcentage de la population âgée de 65 ans ou plus dans les 10 pays présentant les taux de décès les plus élevés liés à la COVID-19 (au-dessus) avec les 10 pays ayant le pourcentage le plus élevé de personnes âgées de 65 ans ou plus (en dessous), on constate que les pays avec un taux de décès très élevé ont un pourcentage de personnes âgées assez faible. De même, les pays présentant un pourcentage élevé de personnes âgées ont un taux de décès beaucoup plus faible. Par conséquent, on peut conclure que l'âge avancé ne semble pas être un facteur prédominant influençant la mortalité liée à la COVID-19.

```
187 select location, MIN(aged_65_older) as max,  
188 sum(new_deaths)/sum(new_cases)*100 as deathsPercentage  
189 from coviddeaths  
190 where new_cases <>0  
191 group by location  
192 order by 3 desc  
193
```

Data Output Messages Notifications			
	location text	max double precision	deathspercentage double precision
1	Yemen	2.922	17.67266638760988
2	Sudan	3.548	7.883674776928727
3	Syria	3.466	5.504762899883322
4	Somalia	2.731	4.949879271237287
5	Peru	7.151	4.880549970665448
6	Egypt	5.159	4.8118010243729445
7	Mexico	6.857	4.3967704880802305
8	Bosnia and Herzegovina	16.569	4.0554493782422
9	Afghanistan	2.581	3.6345066502270416
10	Liberia	3.057	3.6336670374490176

```
195 select location, MAX(aged_65_older) as max,  
196 sum(new_deaths)/sum(new_cases)*100 as deathsPercentage  
197 from coviddeaths  
198 where new_cases <>0 and new_deaths <>0  
199 group by location  
200 order by 2 desc  
201
```

Data Output Messages Notifications			
	location text	max double precision	deathspercentage double precision
1	Japan	27.049	0.22143128435938952
2	Italy	23.021	0.7357998939215104
3	Portugal	21.502	0.47753370419822744
4	Germany	21.453	0.4516533580459585
5	Finland	21.228	0.647150652225396
6	Bulgaria	20.801	2.958773491500886
7	Greece	20.396	0.6119758135640813
8	Gibraltar	20.396	4.2084942084942085
9	Sweden	19.985	0.8961871048571266
10	Latvia	19.754	0.6874180098972608

# ANALYSE DES TAUX DE MORTALITÉ DANS LES PAYS À FORTE PRÉVALENCE DU DIABÈTE

- Lors de l'analyse des taux de mortalité dans les 10 pays présentant les taux de prévalence les plus élevés du diabète, nous avons constaté que ces taux étaient généralement similaires ou inférieurs à la moyenne mondiale (0,90 %). Cette observation suggère que le diabète ne semble pas être un facteur prépondérant de la mortalité liée à la COVID-19.

```
203 select location, MAX(diabetes_prevalence) as diabetes_prevalence,  
204 sum(new_deaths)/sum(new_cases)*100 as deathsPercentage  
205 from coviddeaths  
206 where new_cases <>0 and new_deaths <>0  
207 group by location  
208 order by 2 desc  
209
```

Data Output Messages Notifications

	location text	diabetes_prevalence double precision	deathpercentage double precision
1	Marshall Islands	30.53	0.2047452727929664
2	Turks and Caicos Islands	27.25	3.2753326509723646
3	Nauru	24.07	0.5649717514124294
4	New Caledonia	23.36	1.5067368522721785
5	Kiribati	22.66	1.6826923076923077
6	French Polynesia	22.63	0.8304783381759834
7	French Guiana	22.63	0.7737981229649493
8	Mauritius	22.02	1.195616074393889
9	Guam	21.52	1.2368365255495088
10	Guadeloupe	21.52	0.5094878563027442

# ANALYSE DE LA CORRÉLATION ENTRE LA CAPACITÉ DE PRISE EN CHARGE ET LE TAUX DE MORTALITÉ

- L'observation des taux de décès liés à la COVID-19 dans les 10 pays présentant les pourcentages les plus bas de lits d'hôpital par tranche de 1000 personnes révèle des taux de mortalité très élevés, atteignant jusqu'à 8,5 %. Ce chiffre représente plus de neuf fois la moyenne mondiale. Ces résultats indiquent clairement que la capacité de prise en charge des personnes infectées dans un pays a une influence considérable sur le taux de mortalité à la COVID-19.

```
211 select location, MAX(hospital_beds_per_thousand) as hospital_beds_per_thousand,
212 sum(new_deaths)/sum(new_cases)*100 as deathsPercentage
213 from coviddeaths
214 where new_cases <>0 and new_deaths <>0
215 group by location
216 order by 2
217
218
```

	location	hospital_beds_per_thousand	deathspercentage
	text	double precision	double precision
1	Mali	0.1	3.4957020057306587
2	Maldives	0.1	0.33973859603566176
3	Madagascar	0.2	2.2577039757304806
4	Niger	0.3	8.558807288790723
5	Guernsey	0.3	0.6111320045129748
6	Nepal	0.3	1.2338626639372843
7	Ethiopia	0.3	1.5898604936250247
8	Guinea	0.3	2.3696440426944485
9	Burkina Faso	0.4	3.9336636667003737
10	Turks and Caicos Islands	0.5	3.2753326509723646

# ANALYSE DE L'IMPACT DE LA VACCINATION SUR LE TAUX DE CONTAMINATION ET LE TAUX DE DÉCÈS

- L'analyse des 10 pays affichant les taux de vaccination les plus élevés révèle des tendances encourageantes. Malgré le fait que seulement environ 10% des personnes testées se révèlent positives à la COVID-19, le taux de décès parmi les personnes testées est d'environ 0,2%. Ces résultats indiquent clairement que la vaccination joue un rôle majeur dans la prévention des contaminations et des décès liés à la COVID-19.
- En d'autres termes, la vaccination a un impact significatif sur la réduction de la propagation du virus et la diminution des conséquences graves de la maladie.

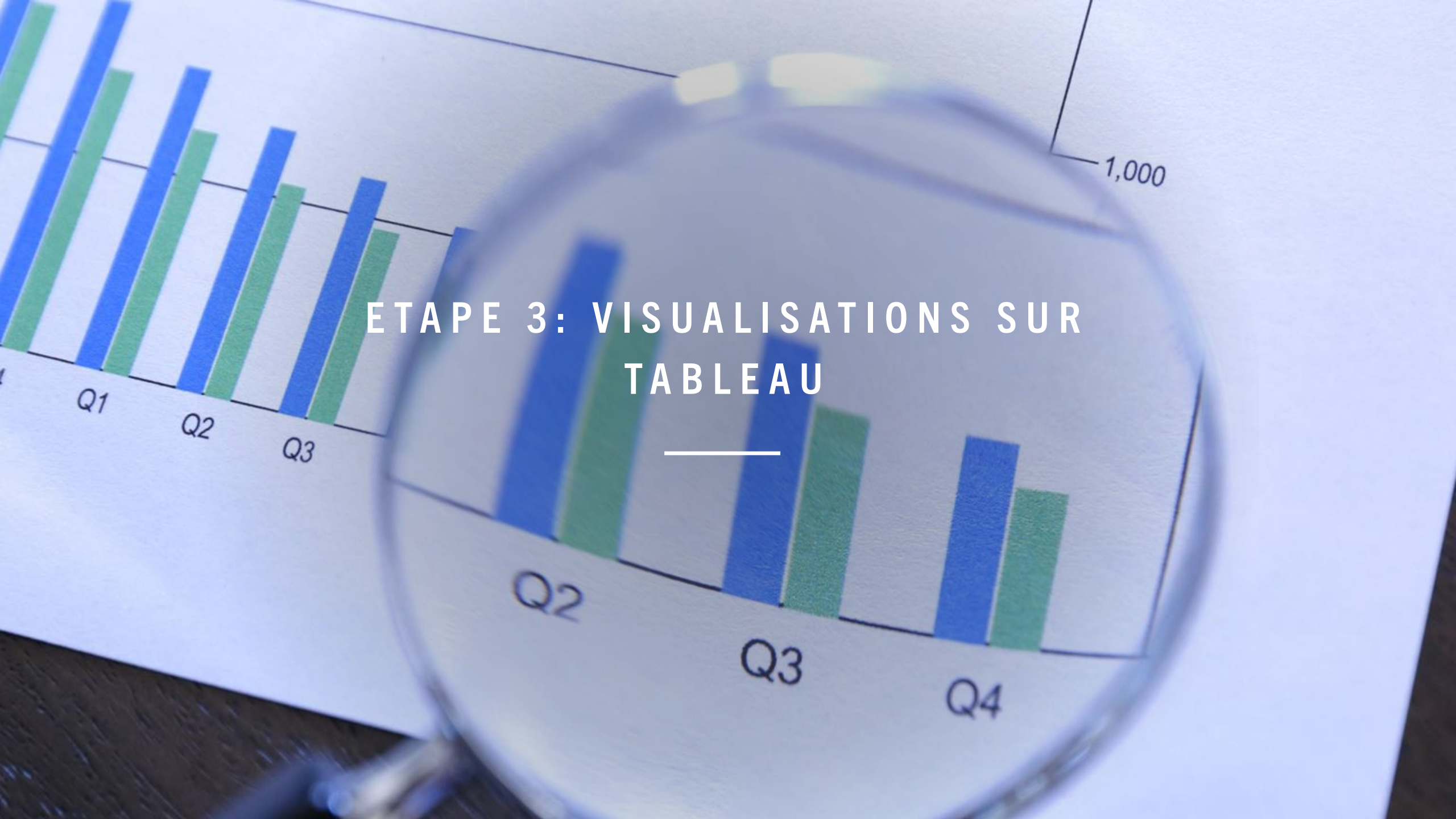
```
select d.location, max(v.people_vaccinated)/max(d.population)*100 as vaccinationPercentage,
sum(d.new_cases)/sum(d.new_tests)*100 as InfectionsPercentage,
sum(d.new_deaths)/sum(d.new_tests)*100 as deathsPercentageAlmonstPeopleTested
from coviddeaths d
inner join covidvaccinations v
on d.location = v.location
where d.new_cases <>0 and new_deaths <>0 and d.new_tests <>0
group by d.location
order by 2 desc
```

location text	vaccinationpercentage double precision	infectionspercentage double precision	deathspercentagealmonstpeopletested double precision
Cuba	95.76686926149537	4.597343554363606	0.03293501522135116
Portugal	95.2627224777835	11.093088905420895	0.0548038317787052
Chile	92.27076410333214	9.821206530858328	0.14715361211181402
Vietnam	91.92018634347555	12.373410351062386	0.06645184312888325
Singapore	91.54746956815141	7.70726063977306	0.004437365789494537
Argentina	91.1685489208998	25.78832899872782	0.3609628710828055
Cambodia	91.16155671946274	11.066136446599495	0.20803923025484805
Nepal	90.98742532388647	15.51342080433344	0.19525055039420713
Canada	90.40125210353436	6.25776155893889	0.0666066593007816
Malta	89.77728940751145	6.5441167032170595	0.0716855079747659
Costa Rica	89.42908055765517	26.79131985875166	0.25700493497881033

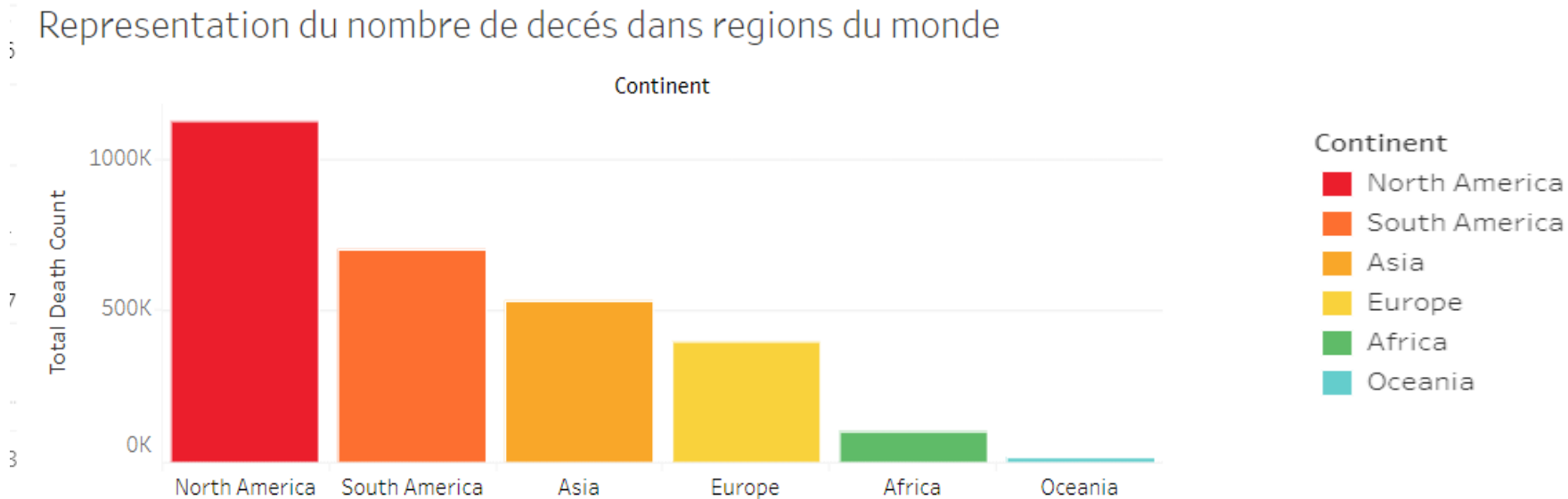


## ETAPE 3: VISUALISATIONS SUR TABLEAU

---



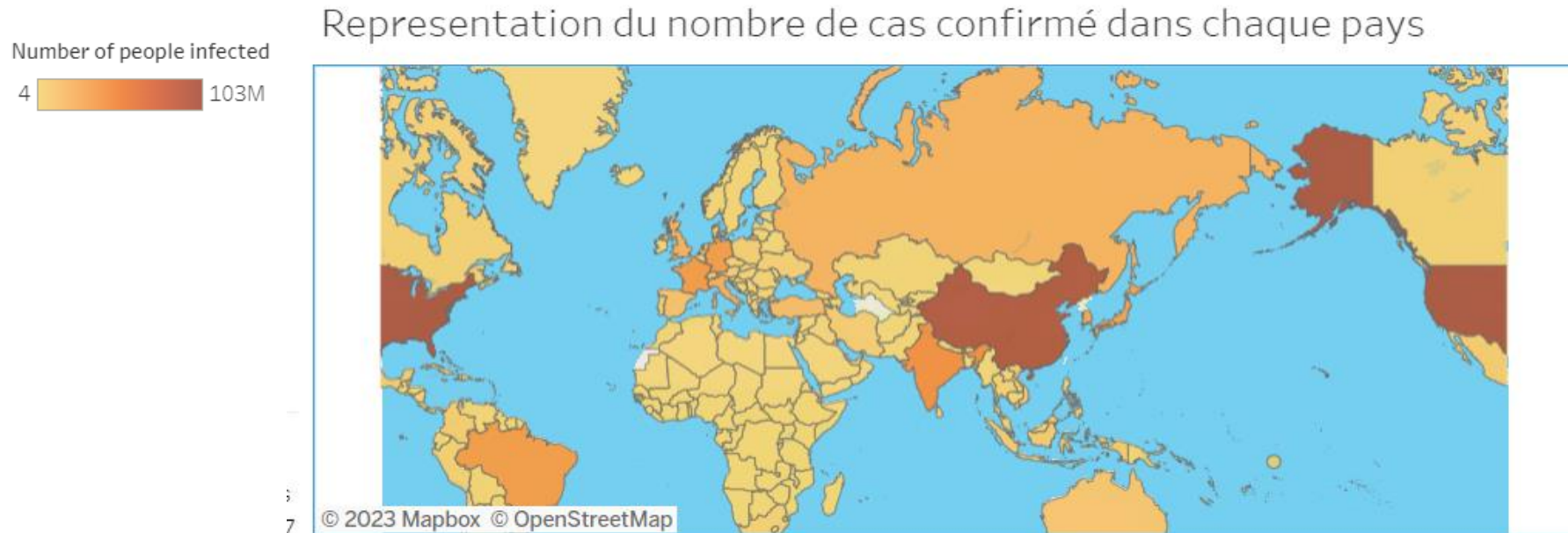
# RÉPARTITION MONDIALE DES CAS CONFIRMÉS DE COVID-19



- Ce graphique illustre la répartition inégale des décès par régions du monde, mettant en évidence les disparités dans les taux de mortalité liés à la COVID-19. Ces disparités peuvent être influencées par divers facteurs précédemment identifiés. Cela souligne l'importance de prendre en compte ces différences régionales dans la gestion de la pandémie et d'adapter les stratégies en conséquence pour réduire les effets négatifs de la maladie.

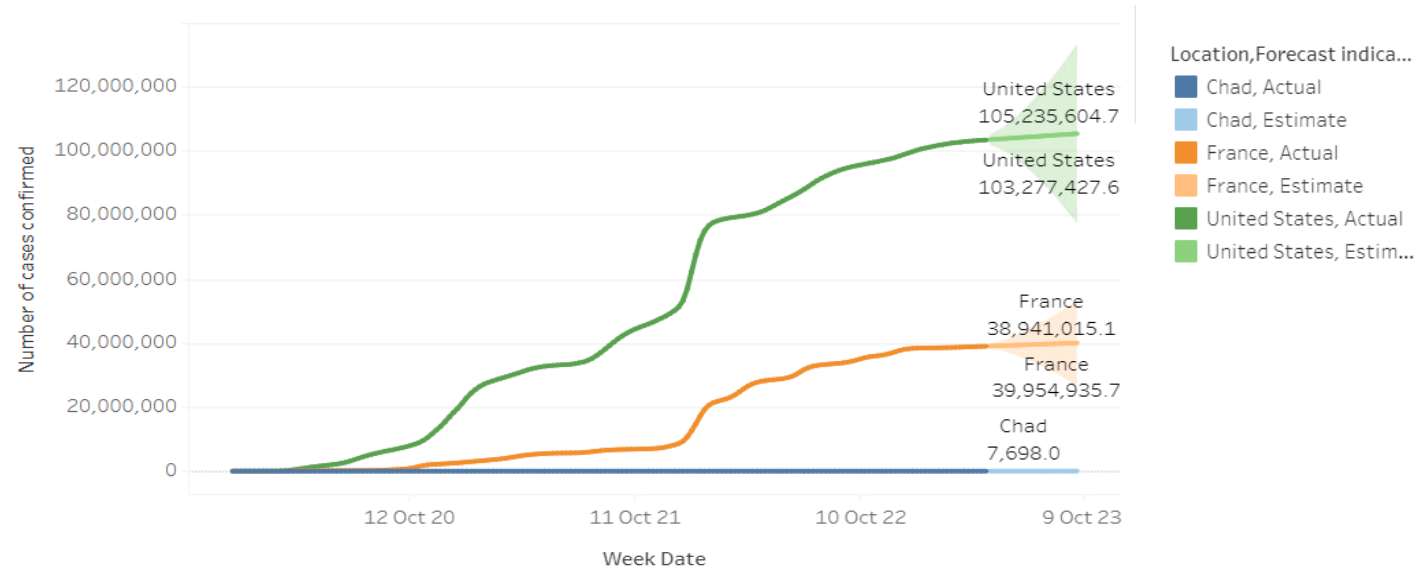
# CARTE INTERACTIVE DE LA COVID-19 : IMPACT MONDIAL ET STATISTIQUES PAR PAYS

- Cette carte du monde interactive facilite la compréhension de l'impact de la Covid-19 dans différents pays en utilisant des couleurs pour représenter les niveaux faibles, moyens et élevés de contamination. En pointant le curseur sur un pays, on peut également connaître le nombre de personnes infectées.
- Lien Tableau Public  
: [https://public.tableau.com/app/profile/chris.essomba/viz/CovidDeathsVisualization\\_16845886904710/Tableaudebord1](https://public.tableau.com/app/profile/chris.essomba/viz/CovidDeathsVisualization_16845886904710/Tableaudebord1)



# COMPARAISON DE L'ÉVOLUTION DES CAS DE COVID-19 ENTRE LA FRANCE, LES ÉTATS-UNIS ET LE TCHAD

- Nous avons analysé l'évolution du nombre total de personnes infectées en France par rapport aux États-Unis (le pays ayant enregistré le plus de cas confirmés) et au Tchad (le pays ayant enregistré le moins de cas). Nous observons une augmentation constante de la courbe entre octobre 2020 et octobre 2022, qui semble maintenant se stabiliser à une certaine valeur, indiquant une très faible incidence de nouveaux cas de contamination. Cependant, les prévisions indiquent qu'il pourrait y avoir environ 1000 nouveaux cas d'ici la fin de cette année.





# CONCLUSION

---

- Cette analyse approfondie des données liées à la pandémie de COVID-19 a permis de répondre à des questions clés. Nous avons identifié les facteurs qui influencent la propagation et la gravité de la maladie, tels que la densité de population, l'âge moyen et la capacité de prise en charge des patients infectés. De plus, nous avons examiné l'impact de la vaccination sur la propagation et la mortalité, en évaluant l'efficacité des campagnes de vaccination.
- En somme, cette étude approfondie a contribué à une meilleure compréhension des dynamiques de la pandémie de COVID-19 et des facteurs qui influencent significativement sa propagation et son taux de mortalité. Ces informations fournissent des bases solides pour orienter les mesures de prévention et les décisions stratégiques visant à atténuer l'impact de la maladie. Il est essentiel de continuer à suivre et à analyser les données afin de rester informé et de s'adapter aux défis évolutifs posés par la COVID-19.

# RÉFÉRENCE

---

- Nom du site web : OurWorldInData
- URL du site web : <https://ourworldindata.org/covid-deaths>
- Date de récupération des données : 15 mai 2023



JE VOUS REMERCIE POUR VOTRE  
ATTENTION !

---