



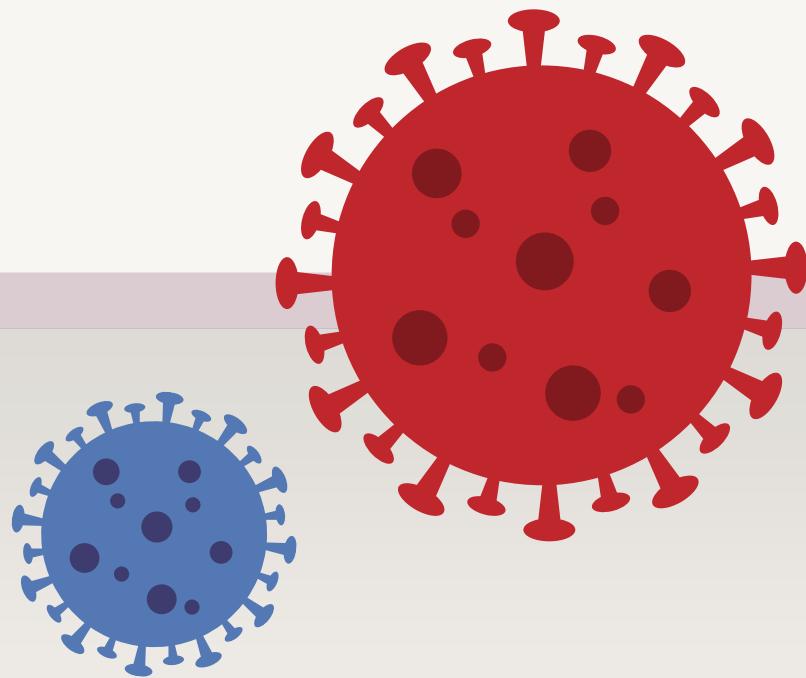
GROUP-4

# COVID-19 PANDEMIC

## United States(US)

### Counties

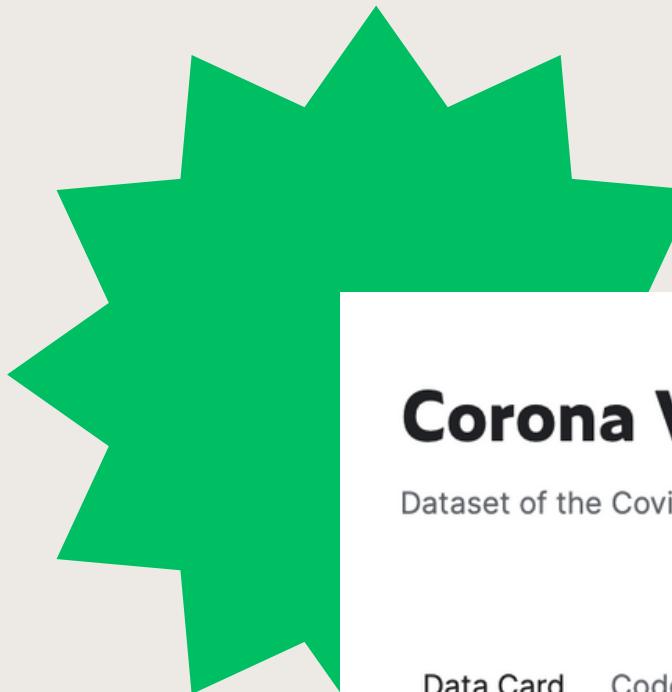
2020-2022



**B E F O R E  
W E  
S T A R T**

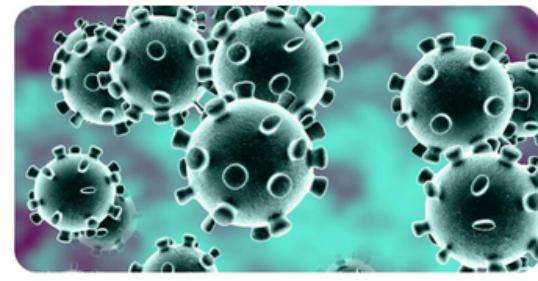


# ทำความรู้จักกับ dataset



## Corona Virus Covid-19 US Counties

Dataset of the Covid-19 pandemic United States(US) Counties



Data Card    Code (0)    Discussion (0)

### About Dataset

This dataset contains the covid-19 pandemic data of the years 2020, 2021 and 2022 of the United States. It has the following information;

1. The FIPS code
2. The State
3. City or Town
4. Date
5. Total Death
6. Total Confirmed Cases
7. The location (longitude and latitude)

**Usability** ⓘ  
9.41

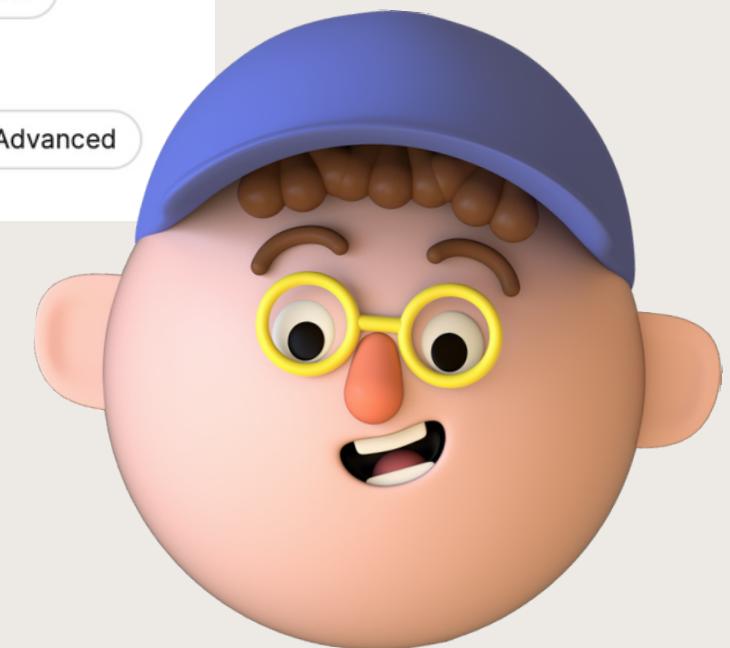
**License**  
[CC0: Public Domain](#)

**Expected update frequency**  
Never

**Tags**

[Text](#) [Intermediate](#)  
[Coronavirus](#)  
[United States](#) [Advanced](#)

- เป็นข้อมูลจำนวนผู้ติดเชื้อและผู้เสียชีวิตใน USA
- Open data ประเภท Public records
- มี Data types เป็น Quantitative data



# Dataset นี้บอกอะไร?



# Admin 2 FIPS Code	A Province/State	A Admin 2 Level (Ci...	Date	# Total Death	# Total Confirmed	A location
The FIPS code	The State	City or Town	Date	Total Deaths	Total Confirmed Cases	The location (longitude and latitude)
<b>0</b> total values	0.0 -111.4418764 Other (3044544)	2% 0% 98%	[null] 100%	<b>3099180</b> total values	<b>3099180</b> total values	<b>3099180</b> total values
12119	Florida	Sumter	2020-01-31	0	0	28.70181754, -82.0794267
13153	Georgia	Houston	2020-01-31	0	0	32.45802497, -83.66879087
13227	Georgia	Pickens	2020-01-31	0	0	34.46589159, -84.46406611
13303	Georgia	Washington	2020-01-31	0	0	32.96711864, -82.79357039
16003	Idaho	Adams	2020-01-31	0	0	44.89333571, -116.4545247
16023	Idaho	Butte	2020-01-31	0	0	43.72592209, -113.1735463



สถิติของผู้ติดเชื้อและผู้เสียชีวิตในแต่ละสถานที่ของ USA รายวัน



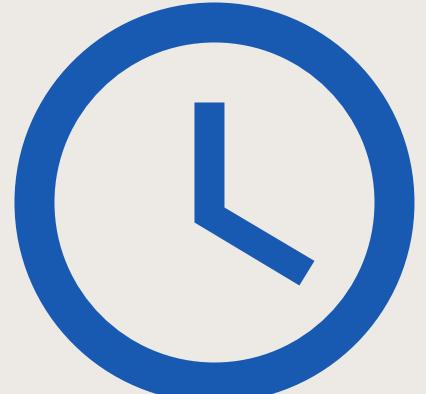
# Timestamp



# Admin 2 FIPS Code	A Province/State	A Admin 2 Level (Ci...	Date	# Total Death	# Total Confirmed	A location
The FIPS code	The State	City or Town	Date	Total Deaths	Total Confirmed Cases	The location (longitude and latitude)
<b>0</b> total values	0.0 -111.4418764 Other (3044544)	2% 0% 98%	[null] 100%	<b>3099180</b> total values	<b>3099180</b> total values	<b>3099180</b> total values
12119	Florida	Sumter	2020-01-31	0	0	28.70181754, -82.08794267
13153	Georgia	Houston	2020-01-31	0	0	32.45802497, -83.66879087
13227	Georgia	Pickens	2020-01-31	0	0	34.46589159, -84.46406611
13303	Georgia	Washington	2020-01-31	0	0	32.96711864, -82.79357039
16003	Idaho	Adams	2020-01-31	0	0	44.89333571, -116.4545247
16023	Idaho	Butte	2020-01-31	0	0	43.72592209, -113.1735463



ມີກາຣບັນທຶກຕົ້ງແຕ່ 2020-01-22 ຈະດຶງ 2022-08-20



# The State

ຈຳນວນຂອງຮັບຊອງສະຫະລຸດເມົາ

## 2. How many province/state do we have?

In 17

```
df["Province/State"].nunique()
```

Executed at 2023.10.04 02:17:51 in 40ms

Out 17

58

ມີບາງອໍານົງໄມ້ຖຸກຕ້ອນ?

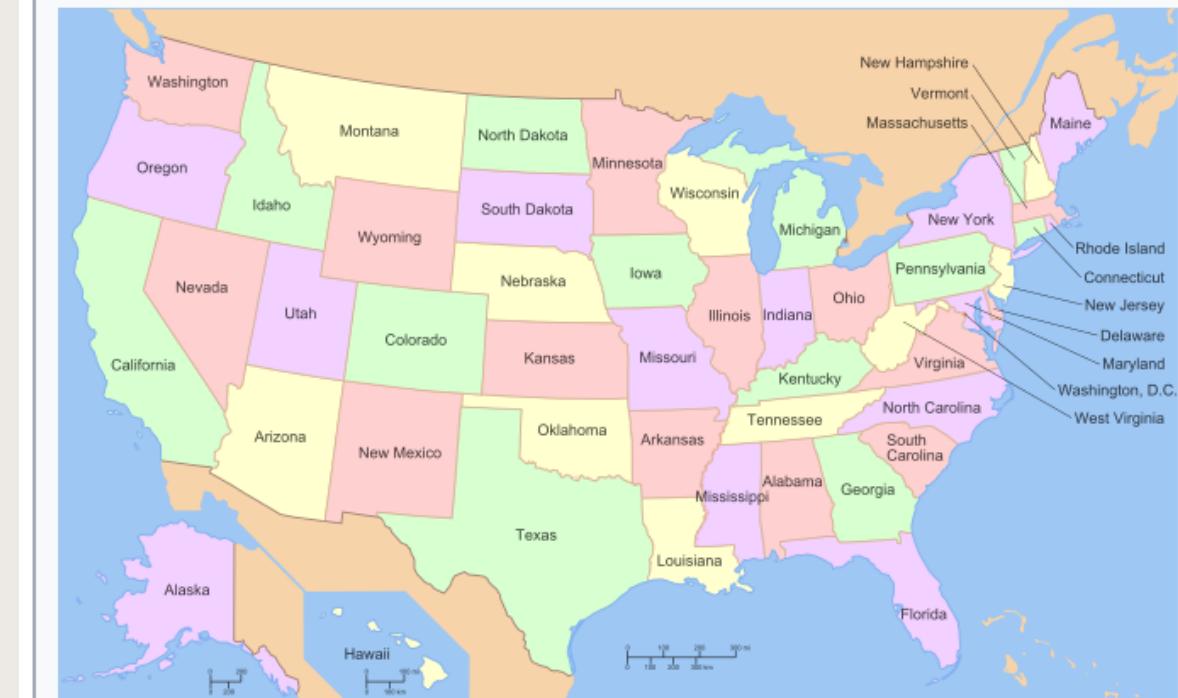


### State

Also known as:

Commonwealth

(the self-designation of four states)



### Category

Federated state

### Location

United States

### Number

50

### Populations

Smallest: [Wyoming](#), 576,851

Largest: [California](#), 39,538,223<sup>[1]</sup>

### Areas

Smallest: [Rhode Island](#), 1,545 square miles (4,000 km<sup>2</sup>)

Largest: [Alaska](#), 665,384 square miles (1,723,340 km<sup>2</sup>)<sup>[2]</sup>

### Government

State government

### Subdivisions

County (or equivalent)

# Data preparation

## เราจะจัดการกับรัฐที่เกินมาได้อย่างไร?

```
dt["CODE"] = dt["Province/State"].map(state_name_of_us)  
dt.head()
```

« ← 5 rows → » 5 rows × 8 columns pd.DataFrame

Ad...	Province/State	Admin 2 ...	Date	Total Death	Total Confirmed	location	CODE
0	12119.0	Florida	Sumter	2020-01-31	0	28.70181754,-82.0794267	FL
1	13153.0	Georgia	Houston	2020-01-31	0	32.45802497,-83.66879087	GA
2	13227.0	Georgia	Pickens	2020-01-31	0	34.46589159,-84.46406611	GA
3	13303.0	Georgia	Washington	2020-01-31	0	32.96711864,-82.79357039	GA
4	16003.0	Idaho	Adams	2020-01-31	0	44.89333571,-116.4545247	ID

เราทำการเพิ่ม column CODE แทนด้วยตัวอักษรย่อ 2 ตัวของชื่อรัฐนั้น ๆ  
หากเป็น **NULL** หมายความว่า row นั้นๆ ไม่ได้เป็นรัฐของอเมริกา

## Question

ทำไมเราไม่ลบ row ที่ไม่ใช่รัฐโดยการใช้ filter  
มีความจำเป็นอะไรที่ต้องเพิ่ม column CODE?

## Answer

เราจำเป็นต้องใช้ ชื่อย่อของรัฐในการ plot ลงบนแผนที่ของ  
อเมริกา จึงจำเป็นต้องมีการเพิ่มแอดวานซ์อยู่แล้วจึงใช้การหาค่า  
**NULL** ใน CODE แทน

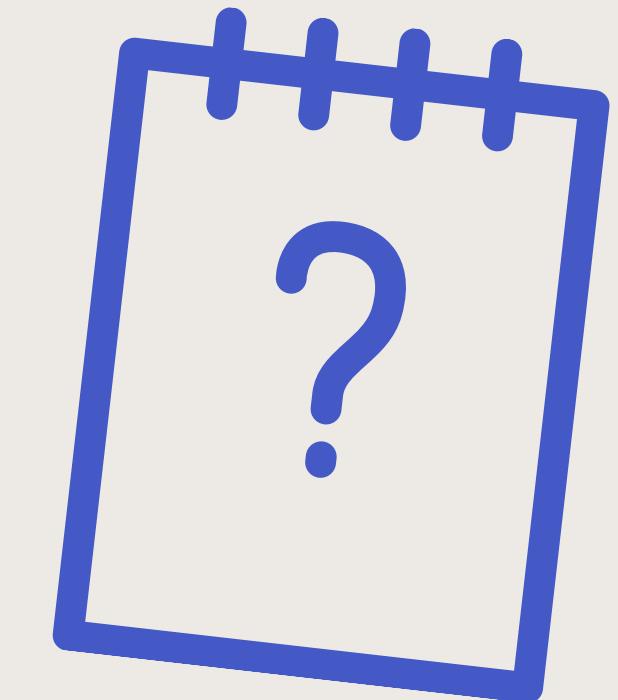
ตัวอย่างไฟล์ json

```
{  
    "Alabama": "AL",  
    "Alaska": "AK",  
    "Arizona": "AZ",  
    "Arkansas": "AR",  
    "California": "CA",  
    "Colorado": "CO",  
    "Connecticut": "CT",  
    "Delaware": "DE",  
    "Florida": "FL",  
    "Georgia": "GA",  
    "Hawaii": "HI",  
    "Idaho": "ID",  
    "Illinois": "IL",  
    "Indiana": "IN",  
    "Iowa": "IA",  
    "Kansas": "KS",  
    "Kentucky": "KY",  
    "Louisiana": "LA",  
    "Maine": "ME",  
    "Maryland": "MD",  
    "Massachusetts": "MA",  
    "Michigan": "MI",  
    "Minnesota": "MN",  
    "Mississippi": "MS",  
    "Missouri": "MO",  
    "Montana": "MT",  
    "Nebraska": "NE",  
    "Nevada": "NV",  
    "New Hampshire": "NH",  
    "New Jersey": "NJ",  
    "New Mexico": "NM",  
    "New York": "NY",  
    "North Carolina": "NC",  
    "North Dakota": "ND",  
    "Ohio": "OH",  
    "Oklahoma": "OK",  
    "Oregon": "OR",  
    "Pennsylvania": "PA",  
    "Rhode Island": "RI",  
    "South Carolina": "SC",  
    "South Dakota": "SD",  
    "Tennessee": "TN",  
    "Texas": "TX",  
    "Utah": "UT",  
    "Vermont": "VT",  
    "Virginia": "VA",  
    "Washington": "WA",  
    "West Virginia": "WV",  
    "Wisconsin": "WI",  
    "Wyoming": "WY"}  
}
```

# Missing Value

## ข้อมูลอื่นมีการสูญหายบ้างหรือไม่?

ใช้วิธีตรวจสอบข้อมูลที่ขาดหายไปด้วยตัว operation `isna()` เพื่อหา `Nan value`



#	FIPS	lat	lon	State	City/County/Borough/Region					
MI0000030	NaN	0.000000	0.000000	Michigan	Michigan Department of Corrections (MDO...					
UT0000089	NaN	37.854472	-111.441876	Utah	Southwest Utah					
UT00000202	NaN	37.854472	-111.441876	Utah	Southwest Utah					
M00000256	NaN	39.099700	-94.578600	Missouri	Kansas City					
MI0000349	NaN	0.000000	0.000000	Michigan	Michigan Department of Corrections (MDO...					
MI0000707	NaN	0.000000	0.000000	Michigan	Federal Correctional Institution (FCI)					
MI0000799	NaN	0.000000	0.000000	Michigan	Michigan Department of Corrections (MDO...					
UT0000810	NaN	40.124915	-109.517442	Utah	TriCounty					

#	FIPS	lat	lon	State	City/	infected total	death total
NaN	66.0	13.4443	144.7937	Guam	NaN	183	5
NaN	88888.0	0.0000	0.0000	Diamond Princess	NaN	49	0
NaN	60.0	-14.2710	-170.1320	American Samoa	NaN	0	0
NaN	78.0	18.3358	-64.8963	Virgin Islands	NaN	5093	41
NaN	66.0	13.4443	144.7937	Guam	NaN	8673	143
NaN	69.0	15.0979	145.6739	Northern Mariana Islands	NaN	198	2
NaN	99999.0	0.0000	0.0000	Grand Princess	NaN	103	3
NaN	78.0	18.3358	-64.8963	Virgin Islands	NaN	3608	28

แต่ในบางข้อมูลที่หายไปก็ไม่ใช่ข้อมูลที่สำคัญและจำเป็นต้องใช้ข้อมูลอื่นๆ ในการจัดการ ไม่จำเป็นที่จะต้องกำจัดไว้กับส่วนนี้  
เช่น FIPS ซึ่งเป็นเพียงมาตรฐานที่รัฐบาลสหรัฐฯ ประยุกต์เพื่อ ให้เป็นมาตรฐานที่เกี่ยวข้องกับระบบคอมพิวเตอร์  
สำหรับหน่วยงานภาครัฐที่ไม่ใช่ หน่วยงานทหาร เท่านั้น

# Clean Data

ขั้นตอนแรกหากเมืองที่เป็น Out of เพื่อตัดข้อมูลที่ไม่จำเป็น  
ออกจากข้อมูลที่จะนำไปใช้

# some City/County/Borough/Region are not in the US	data[data['City/County/Borough/Region'].str.contains('out of', case=False)]								
Date	FIPS	lat	lon	State	Abbr	City/County/Borough/Region	infected total	death total	
id									
MA0000102	2020-06-23	80025.0	0.0	0.0	Massachusetts	MA	Out of MA	0	0
CA0000180	2020-06-24	80006.0	0.0	0.0	California	CA	Out of CA	0	0
MS0000232	2020-06-24	80028.0	0.0	0.0	Mississippi	MS	Out of MS	0	0
OH0000256	2020-06-24	80039.0	0.0	0.0	Ohio	OH	Out of OH	0	0
ME0000354	2020-06-13	80023.0	0.0	0.0	Maine	ME	Out of ME	0	0
...	...	...	...	...	...	...	...	...	...
ND2401446	2022-02-20	80038.0	0.0	0.0	North Dakota	ND	Out of ND	0	0
DE2401493	2022-02-21	80010.0	0.0	0.0	Delaware	DE	Out of DE	0	0
LA2401637	2022-02-22	80022.0	0.0	0.0	Louisiana	LA	Out of LA	0	0
MA2401642	2022-02-22	80025.0	0.0	0.0	Massachusetts	MA	Out of MA	0	0
PA2401676	2022-02-22	80042.0	0.0	0.0	Pennsylvania	PA	Out of PA	0	0

โดยขั้นตอนนี้จะมีการตรวจสอบและนำข้อมูล  
เหล่านี้ออกไปเนื่องจากไม่จำเป็นในการนำไปใช้  
หากข้อมูลสกัดติดของสหรัฐ และข้อมูลเหล่านี้จะทำให้  
การประมวลผลซ้ำลังโดยไม่จำเป็น

ตัวอย่างไฟล์ json

```
{  
    "Alabama": "AL",  
    "Alaska": "AK",  
    "Arizona": "AZ",  
    "Arkansas": "AR",  
    "California": "CA",  
    "Colorado": "CO",  
    "Connecticut": "CT",  
    "Delaware": "DE",  
    "Florida": "FL",  
    "Georgia": "GA",  
    "Hawaii": "HI",  
    "Idaho": "ID",  
    ...}
```

ถัดมาตรวจสอบข้อมูลที่ไม่ถูกต้องหรือข้อมูลที่ไม่ต้องการ เช่น  
ประชากรผู้ติดเชื้อหรือเสียชีวิตติดลบ เพื่อป้องกันความผิดพลาด  
ที่อาจเกิดขึ้นได้ในภายหลัง

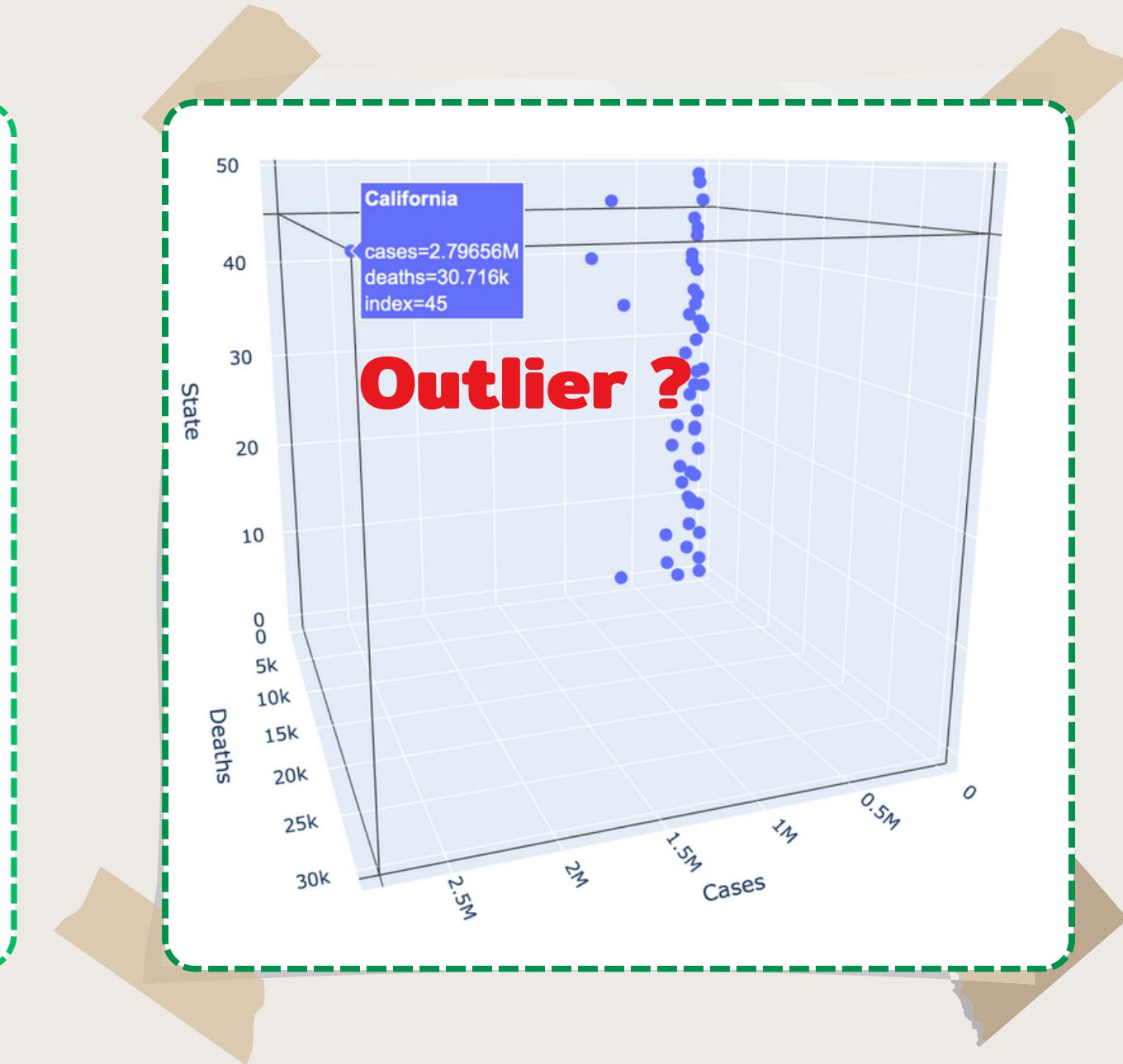
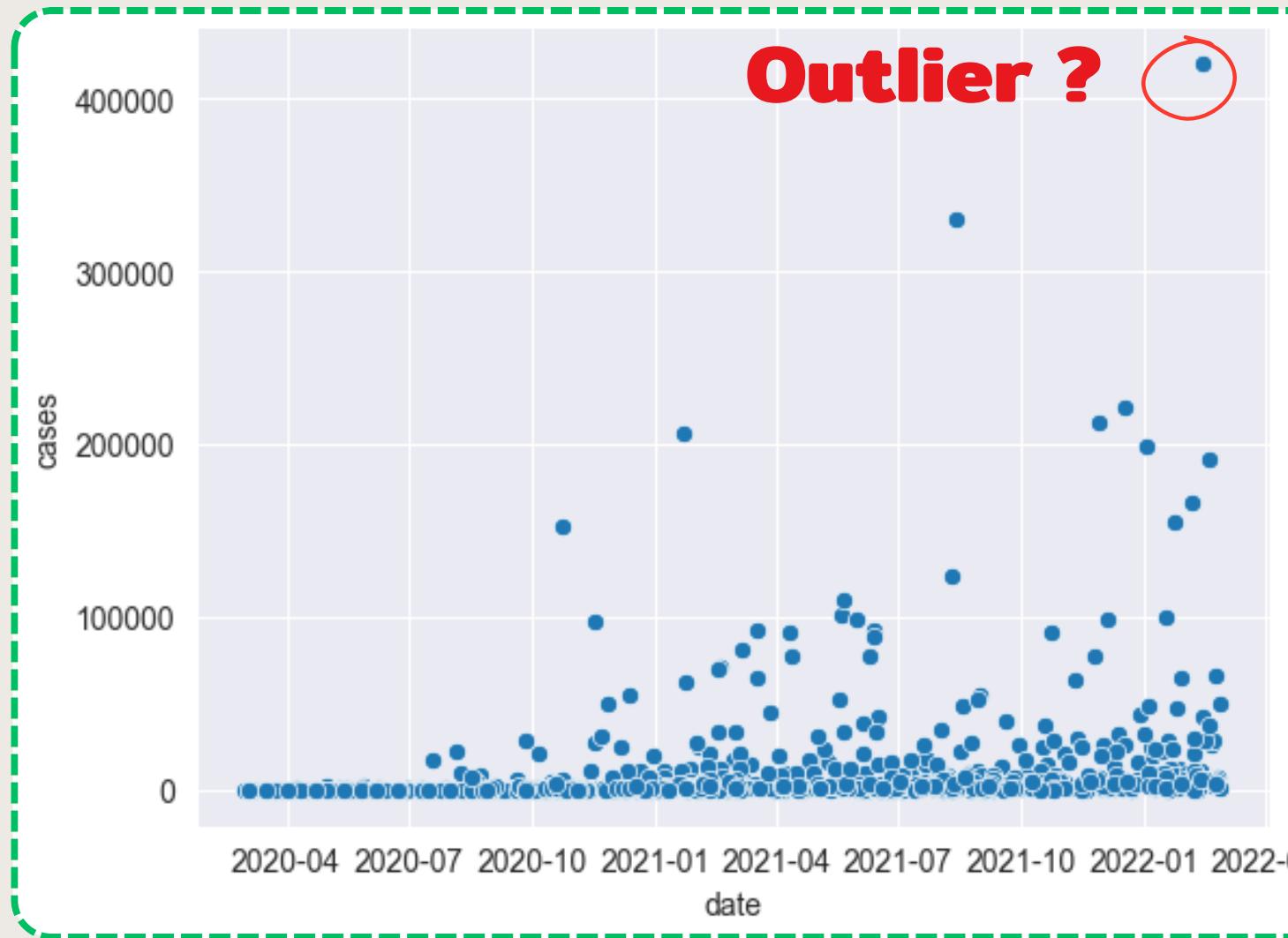
# check infected total	data[data['infected total'] < 0] # ok, not have negative values							
Date	FIPS	lat	lon	State	Abbr	City/County/Borough/Region	infected total	death total
id								
# check death total	data[data['death total'] < 0] # ok, not have negative values							
Date	FIPS	lat	lon	State	Abbr	City/County/Borough/Region	infected total	death total
id								

วิธีการหนันสามารถหาได้  
โดยตรงจากข้อมูลที่ต้องการหา  
ใน column นั้นๆนำมาเทียบกับ 0

Date	FIPS	lat	lon	State	Abbr	City/County/Borough/Region	infected total	death total	
id									
MI0002598	2021-07-16	NaN	0.0	0.0	Michigan	MI	Michigan Department of Corrections (MDOC)	26760	153
MI0005325	2020-05-14	NaN	0.0	0.0	Michigan	MI	Michigan Department of Corrections (MDOC)	2257	56

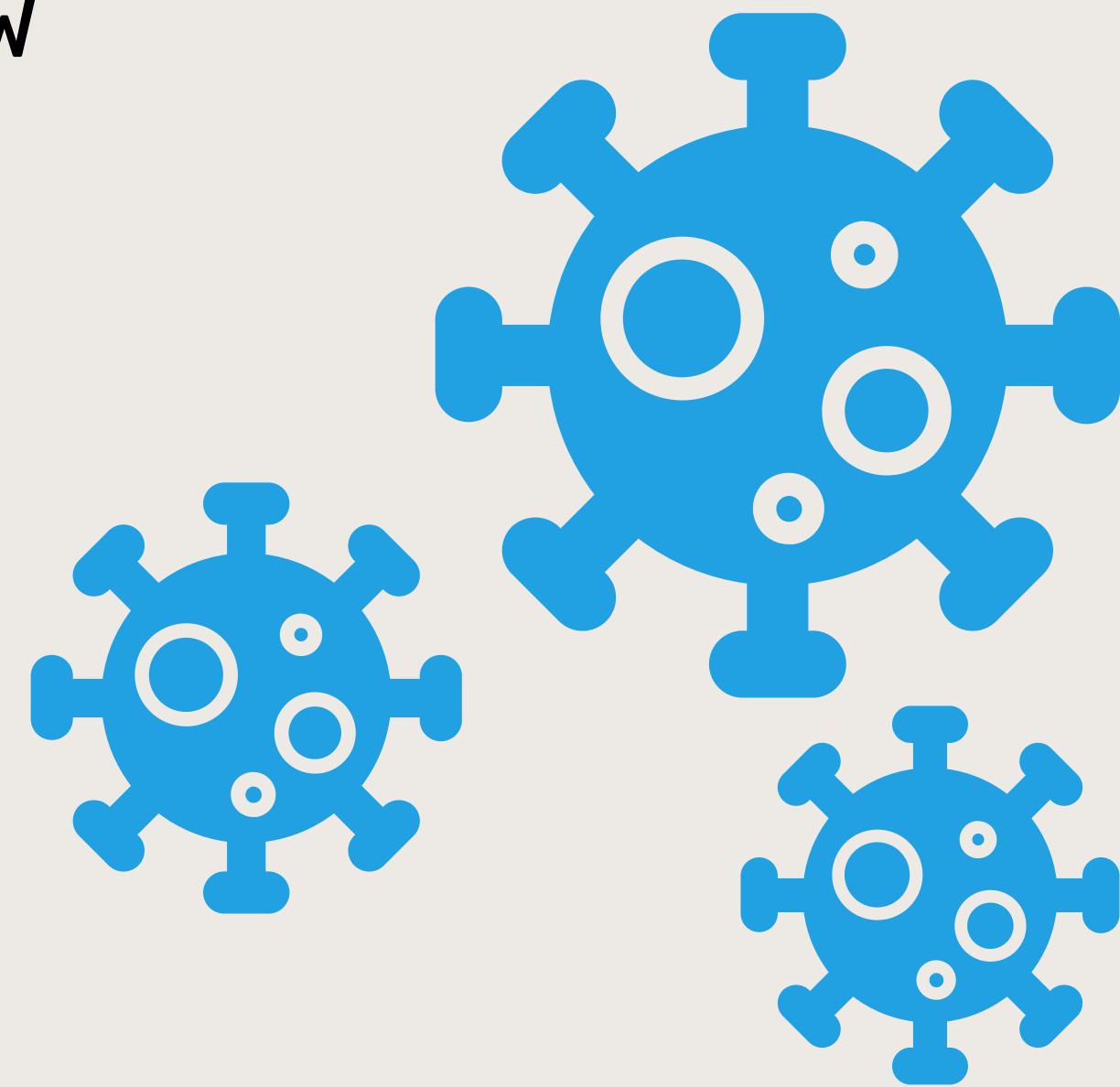
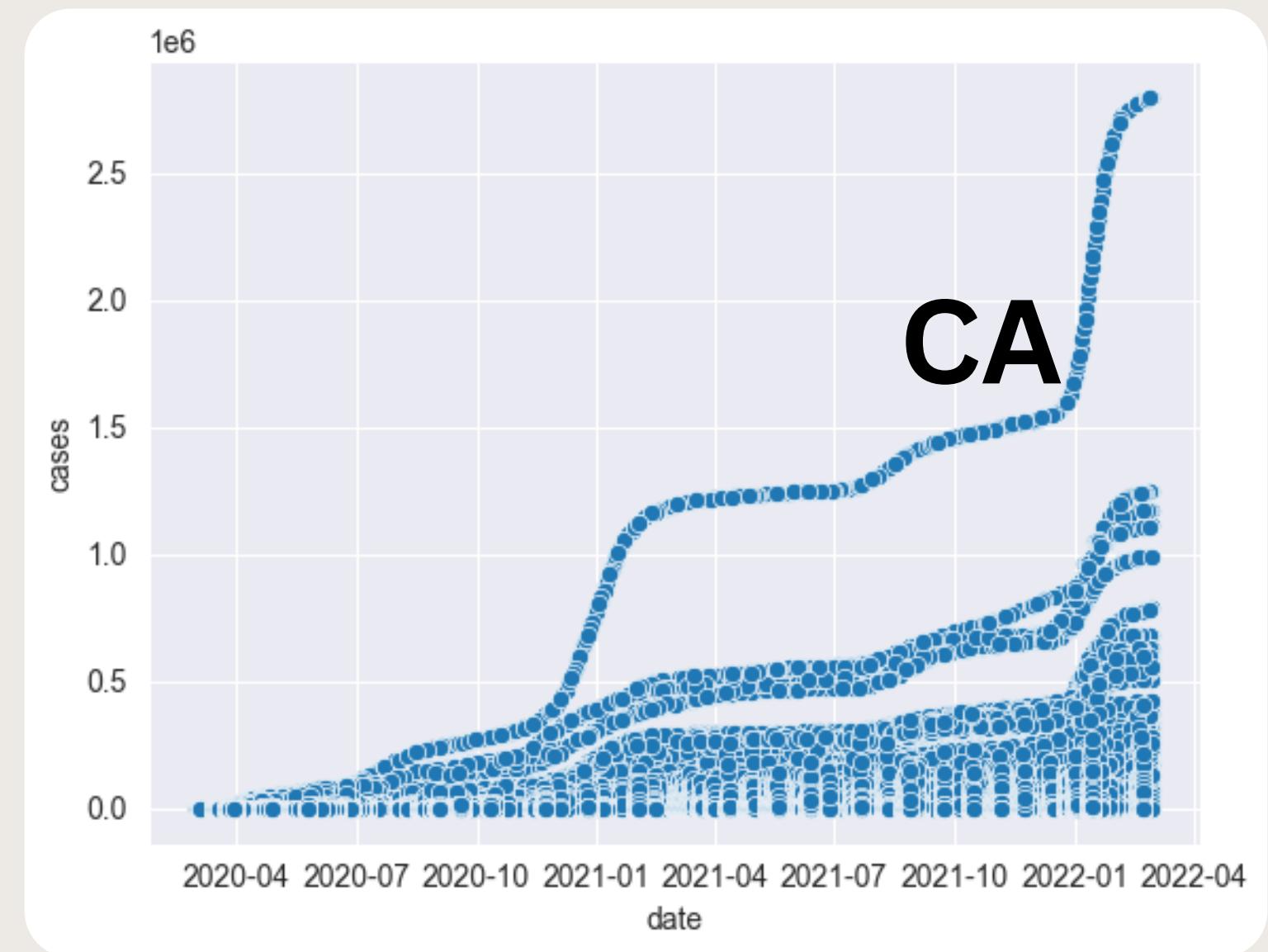
ข้อมูลที่ไม่ถูกต้อง

ลองนำ sample มา plot ดูว่าข้อมูลของเรามีลักษณะอย่างไร



เรายังไม่สามารถสรุปได้ว่า นี่คือ Outlier  
จนกว่าจะทราบข้อมูลทั้งหมด

# ต่อมาใช้ข้อมูลเคสผู้ติดเชื้อและเสียชีวิตกั้งหมดมาplotกราฟ



จากส่วนนี้จะถือว่าข้อมูลไม่มี outlier  
เนื่องจากไม่มีข้อมูลที่โดดออกจากอย่างผิดปกติ เนื่องจากกราฟมี  
การโต้ย่างต่อเนื่องไม่ได้เป็นการมีค่าที่กระโดนออกจาก

ភោជន៍  
ជាសក

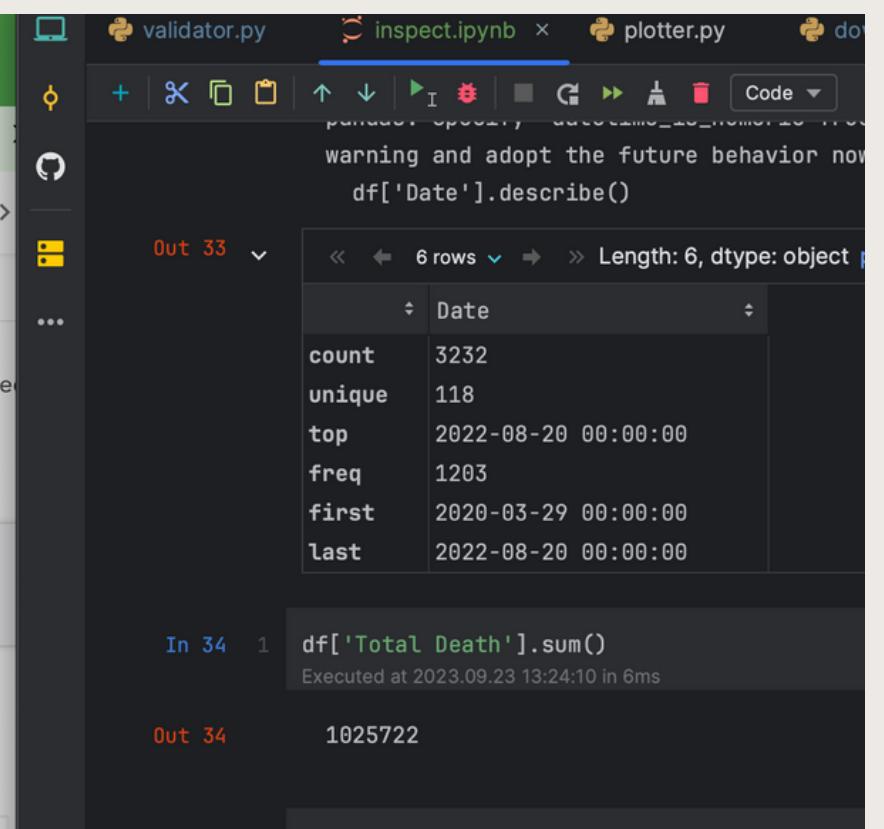
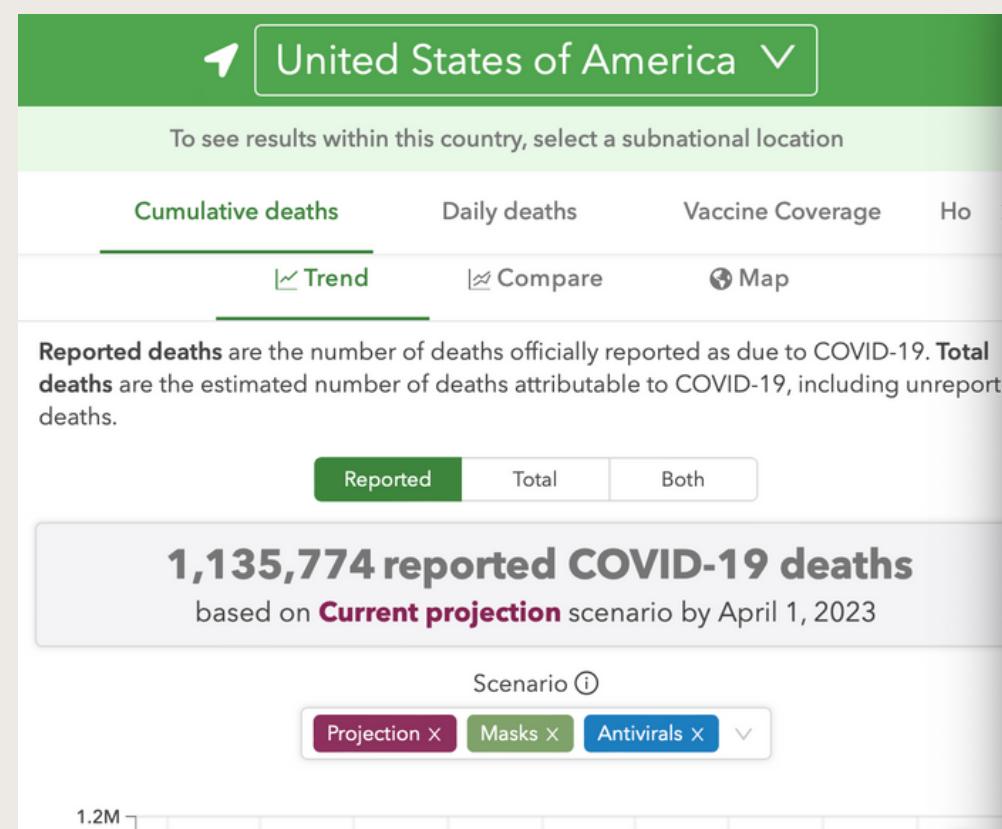


## การตรวจสอบความถูกต้องกับข้อมูลภายนอก

เพื่อยืนยันความถูกต้องและความสะอาดของชุดข้อมูล เราจึงตรวจสอบโดยนำไปเปรียบเทียบกับข้อมูลจากชุดข้อมูล "Reported Covid-19" ที่ <https://www.healthdata.org> ชุดข้อมูลนี้ได้รับการดูแลอย่างดีโดย The Institute for Health Metrics and Evaluation (IHME) แห่งมหาวิทยาลัยวอชิงตัน และได้รับการยกย่องว่าเป็นหนึ่งในแหล่งข้อมูลเกี่ยวกับโควิด-19 ที่น่าเชื่อถือที่สุด

## ผลการตรวจสอบข้อมูล

ด้านซ้ายแสดงข้อมูลจาก <https://www.healthdata.org> ด้านขวาแสดงข้อมูลของเรา :



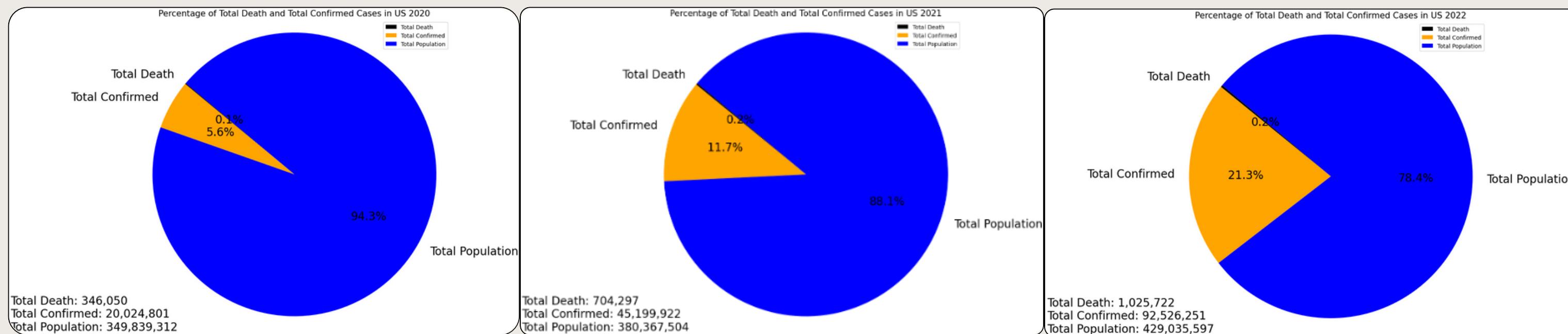
เมื่อเทียบชุดข้อมูลของเรา กับ ข้อมูลจาก <https://www.healthdata.org> มีความใกล้เคียงกันมาก (ความต่างของจำนวนเกิดจากจำนวนวันที่ทำการบันทึกนับไม่เท่ากัน) ทำให้ดูว่ามี **ความแม่นยำและความสะอาดสูง**



# Data Exploration

## EDA (Exploratory Data Analysis)

การวิเคราะห์ข้อมูลจะเริ่มจากการแบ่งข้อมูลเป็นส่วนๆ แบ่งตามถูกต้องหรือแต่ละปี เพื่อให้สามารถวิเคราะห์ข้อมูลและจัดการข้อมูลได้ง่ายขึ้น



ข้อมูล pie chart นี้แสดงถึงจำนวนของประชากรผู้ติดเชื้อ, ผู้เสียชีวิตและประชากรกังวล  
เพื่อดูสัดส่วนของผู้ติดเชื้อและไม่ติดเชื้ออีกด้วย  
จากข้อมูลในแต่ละปีจะสังเกตุเห็นว่าจำนวนของผู้ติดเชื้อจะเพิ่มขึ้นราวกับเท่ากันทุกๆ ปี

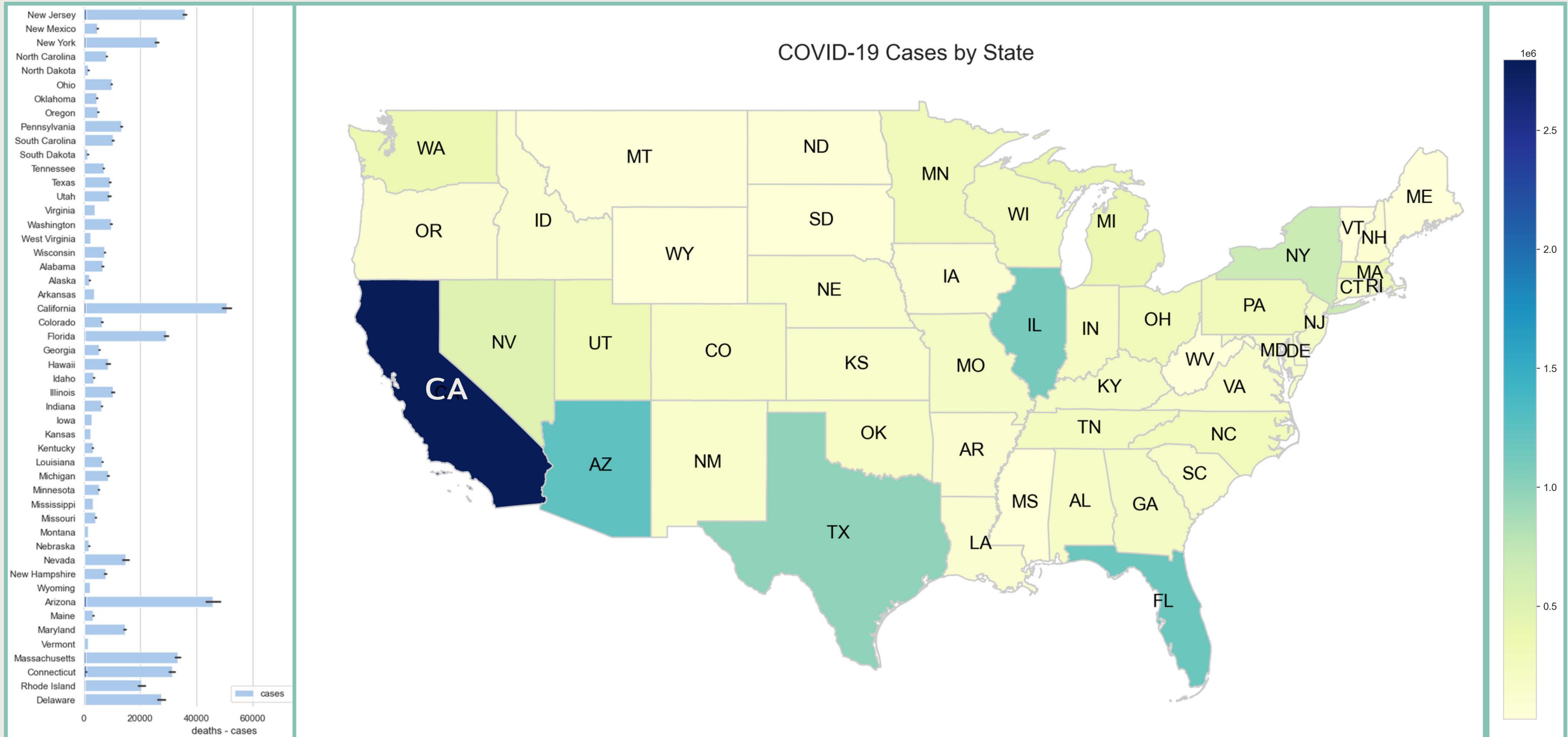
การหา % ของ pie chart นี้

ผู้เสียชีวิต = จำนวนผู้เสียชีวิต

ผู้ติดเชื้อ = จำนวนผู้ติดเชื้อ - จำนวนผู้เสียชีวิต

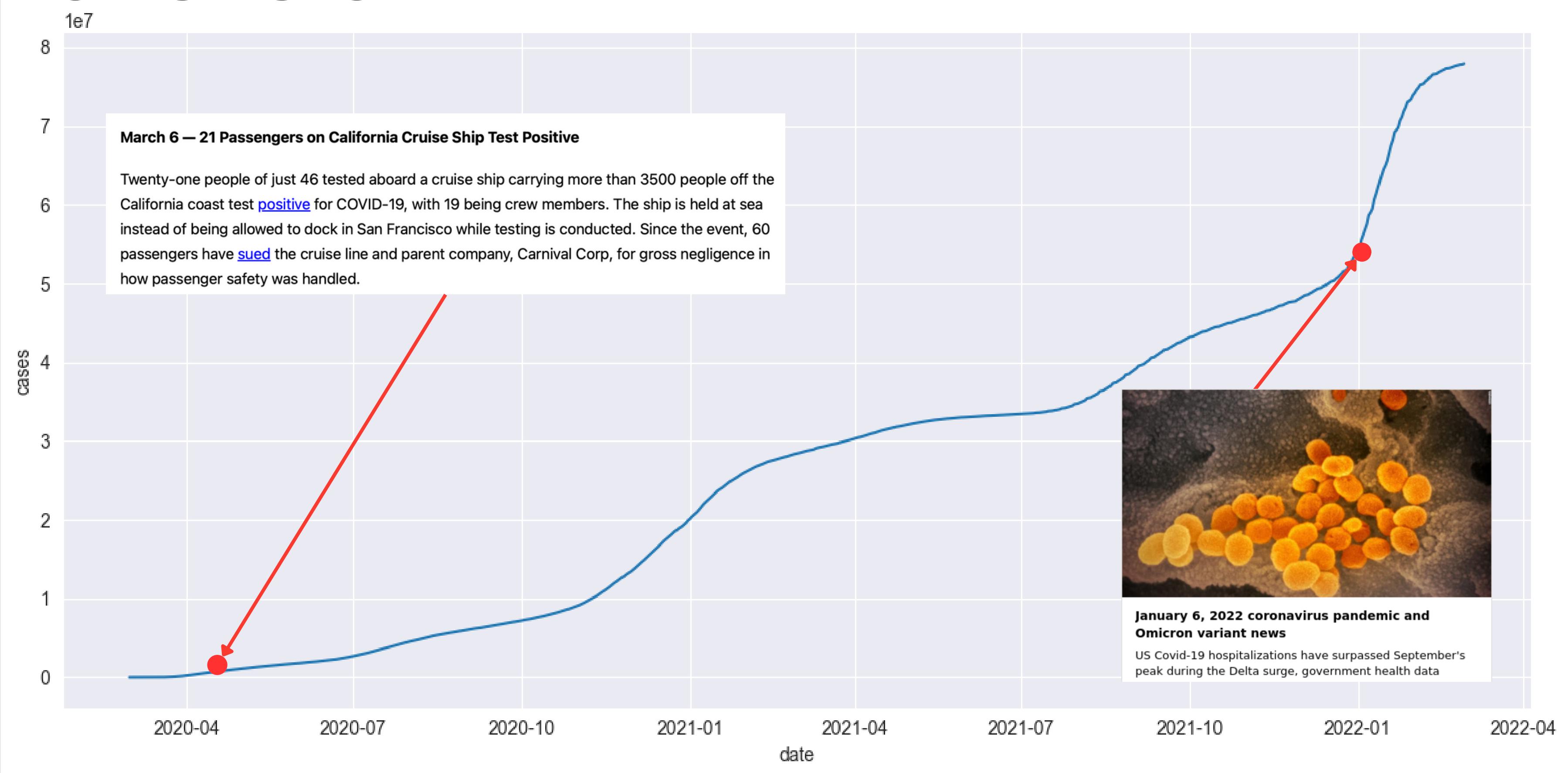
ประชากรกี่เหลือ = จำนวนประชากรกังวล - จำนวนผู้ติดเชื้อ - จำนวนผู้เสียชีวิต





จากภาพเราจะเห็นได้ว่าช่วงปี 2020 - 2022 California มีจำนวนประชากรติดเชื้อมากที่สุด เราจะมาดูกันว่าเกิดเหตุการณ์อะไรขึ้นบ้าง

# CASES OVER TIME

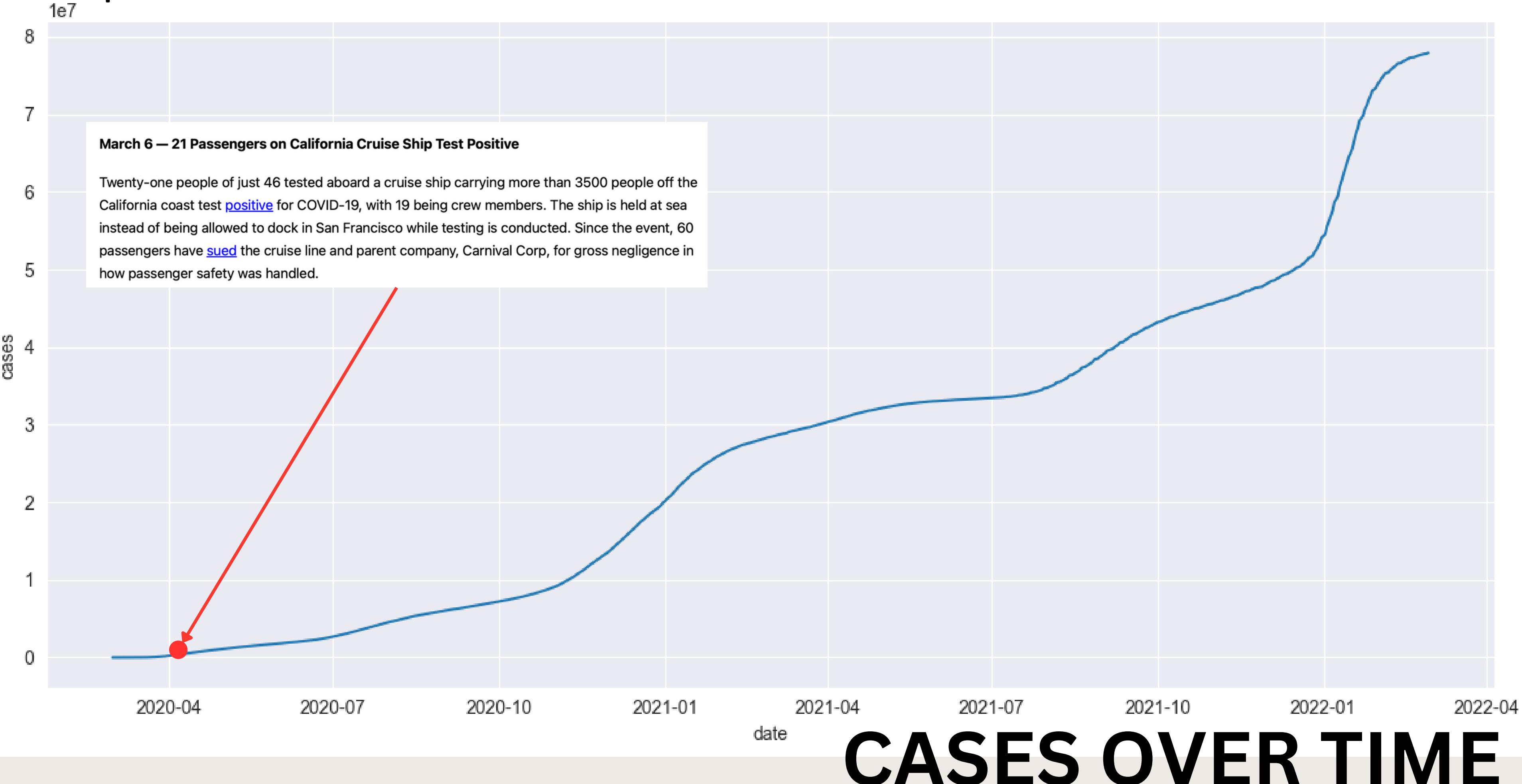


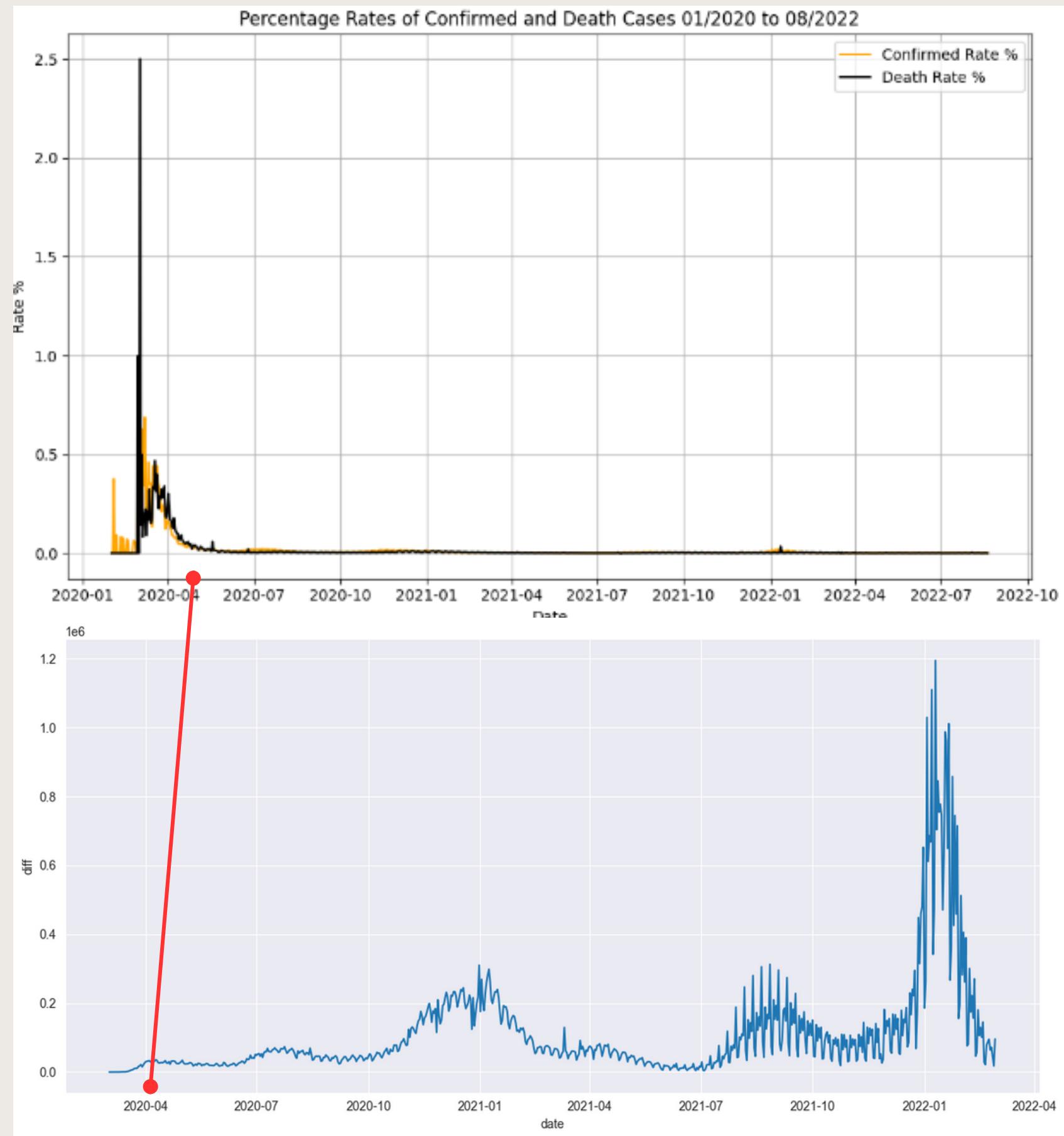


စာချုပ်

နံ ၂ အောက်ပါ

# ຈຸດເຮີມຕົ້ນຂອງ covid-19 ໃນ CA

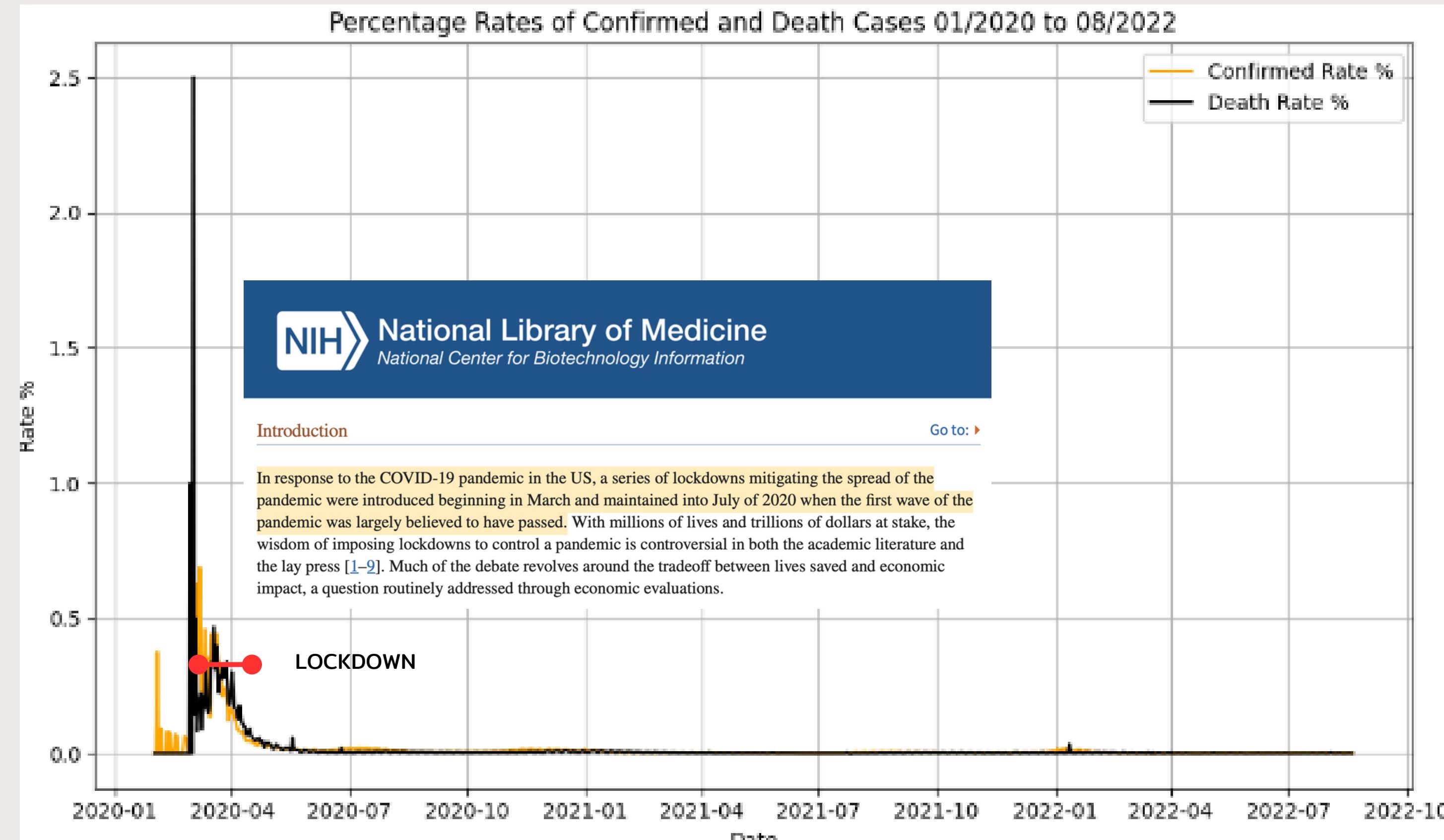




จำนวนของผู้ติดเชื้อที่เพิ่มขึ้นต่อวัน

จากรูปเราจะเห็นได้ว่าเมื่อเทียบกับช่วงเวลาอื่นๆ ช่วงต้นของการแพร่ระบาด มีอัตราการเพิ่มขึ้นที่ต่ำมากแต่กลับมีอัตราการเสียชีวิตที่สูงมาก

**แต่ในขณะเดียวกัน  
กลับลดลงอย่างรวดเร็ว  
เกิดอะไรขึ้น?**

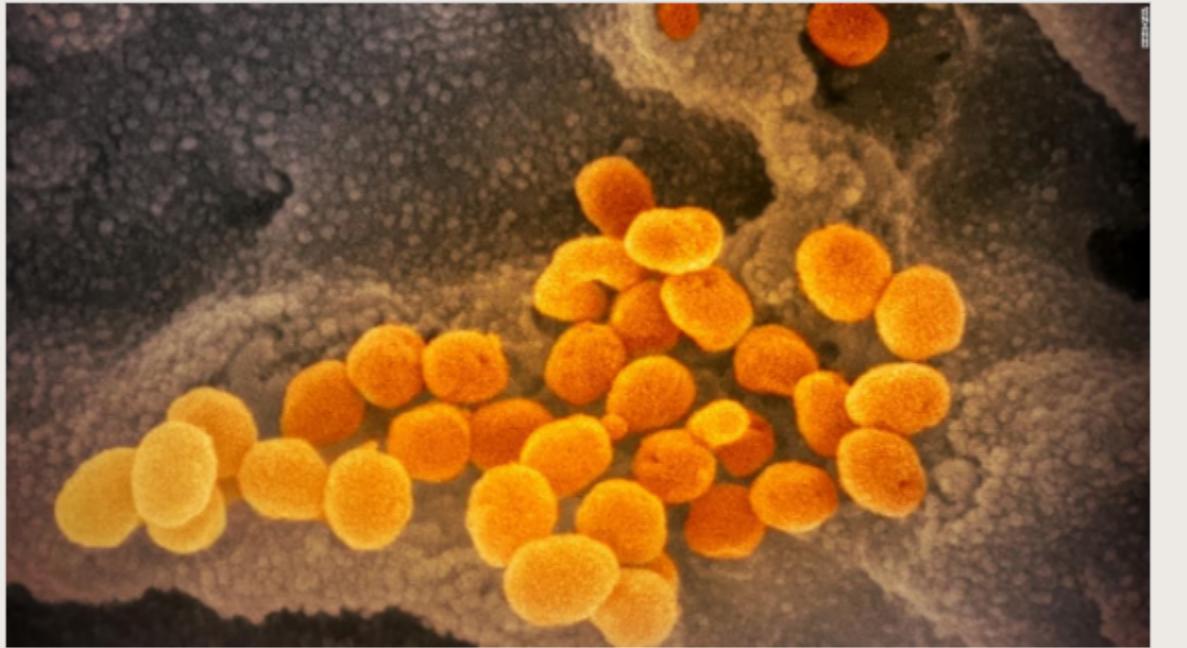


เราจะเห็นได้ว่า การรับมือของภาครัฐโดยการสั่ง **lockdown** ส่งผลอย่างมากต่อจำนวนผู้ติดเชื้อลดลงอย่างมีนัยสำคัญ

# CASES OVER TIME

ວະໄໄທມາໃຫ້ເກີດກາຣເພີ່ມຂຶ້ນ  
ຂອງຜູ້ຕິດເສື້ອວ່າຍາງ  
ຮວດເຮົວ?

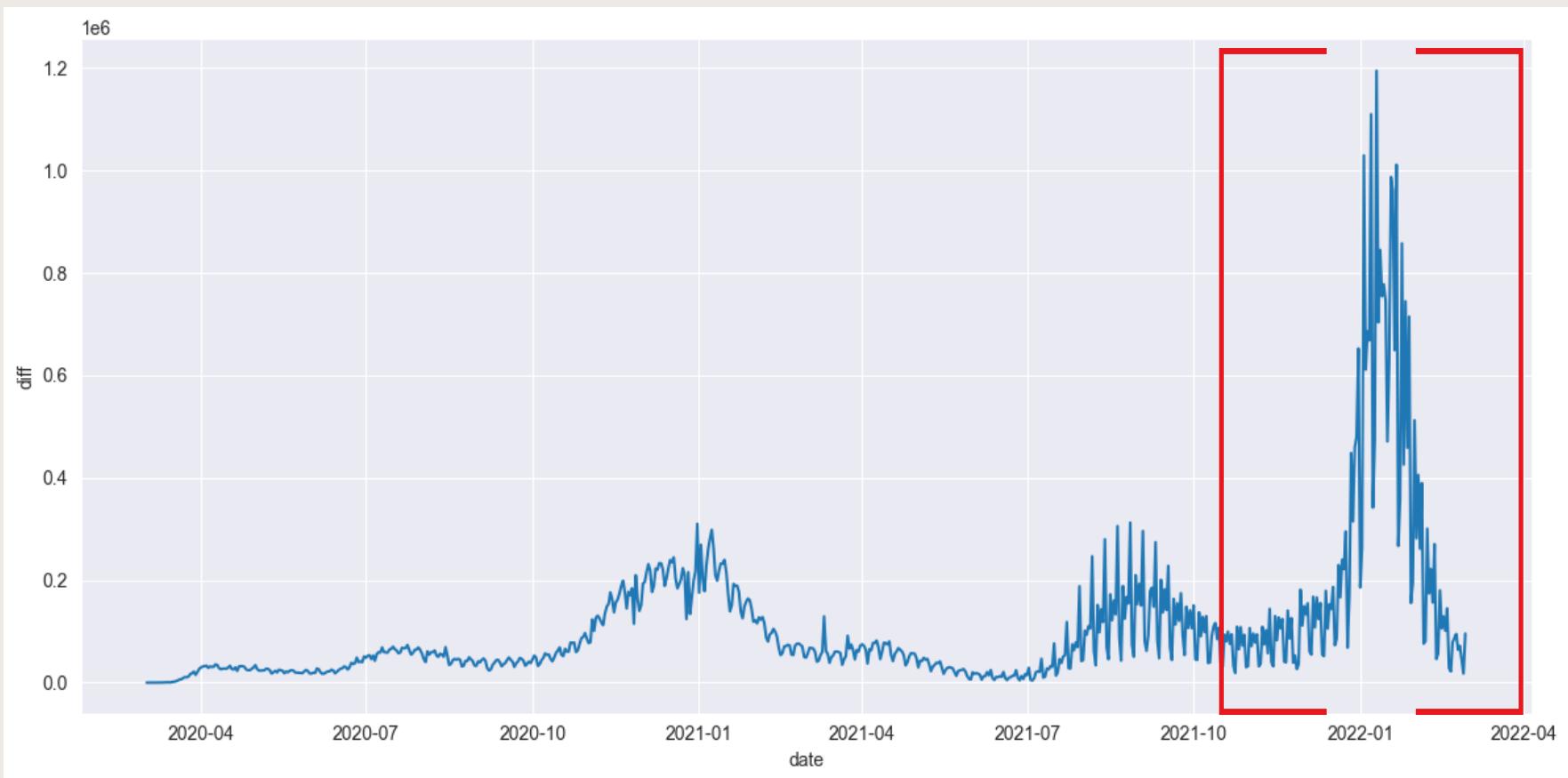
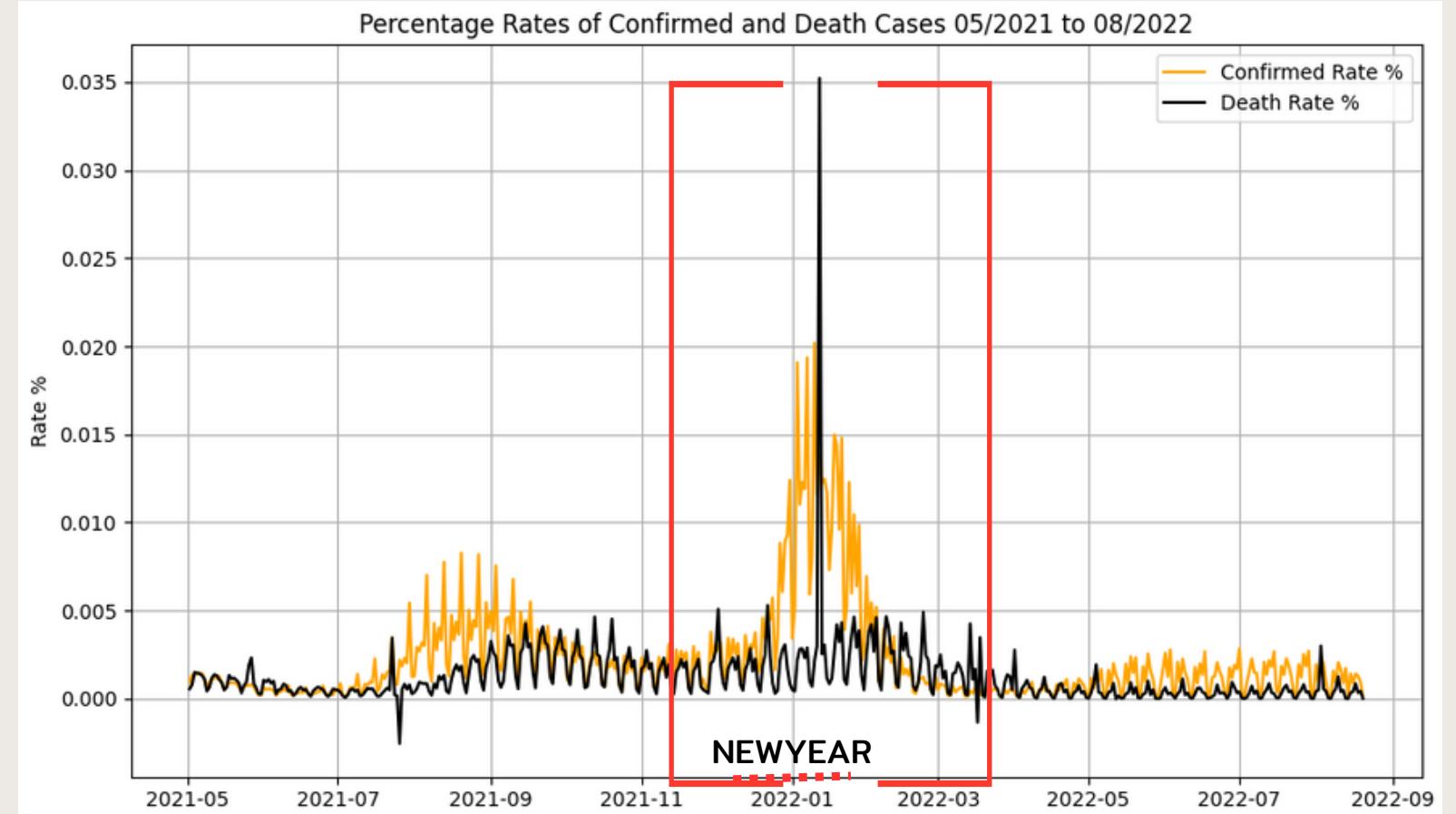




### January 6, 2022 coronavirus pandemic and Omicron variant news

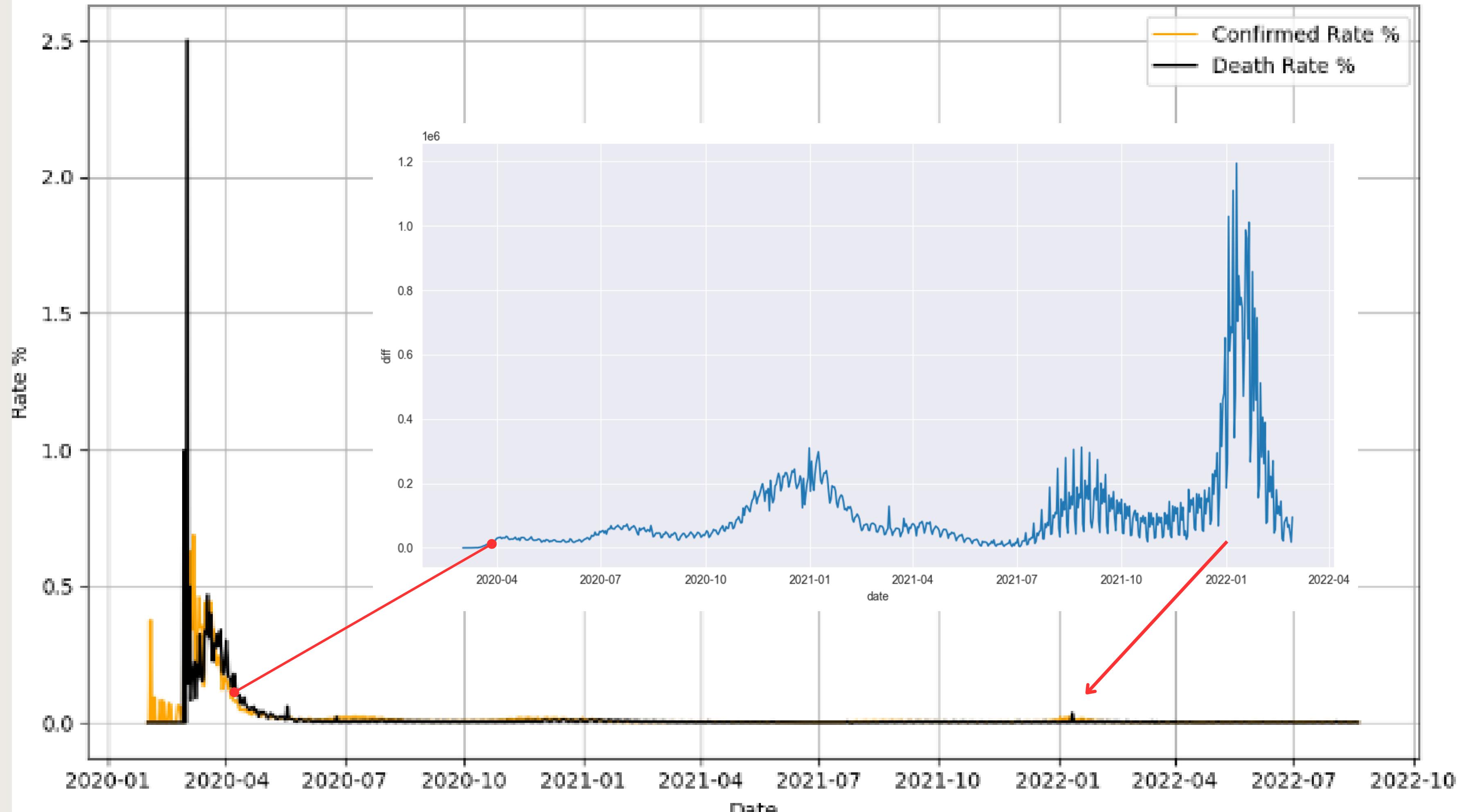
US Covid-19 hospitalizations have surpassed September's peak during the Delta surge, government health data shows. Follow here for the latest.

CNN / Jan 6, 2022

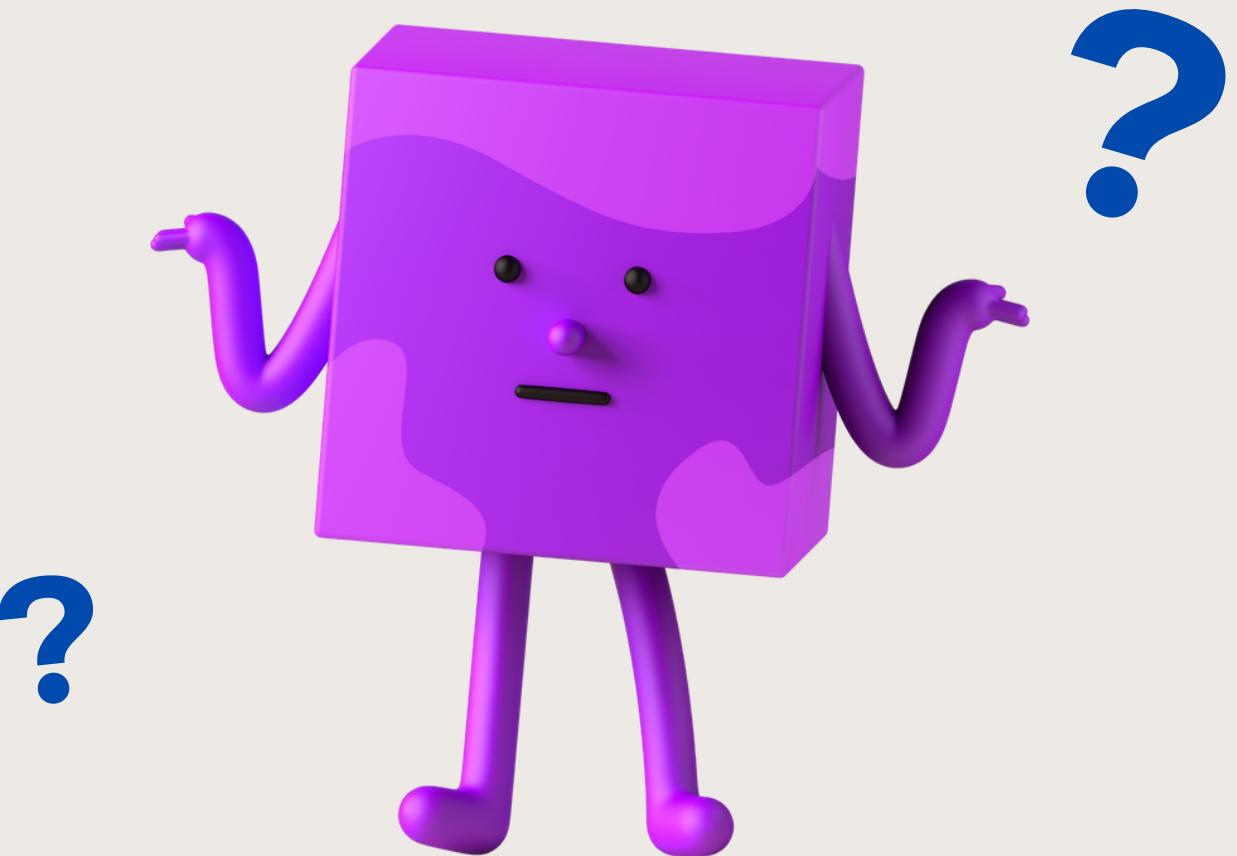


การเกิดขึ้นของเชื้อ  
COVID19 สายพันธุ์ใหม่  
ทำให้เกิดการแพร่ระบาด  
อย่างรวดเร็ว

# Percentage Rates of Confirmed and Death Cases 01/2020 to 08/2022



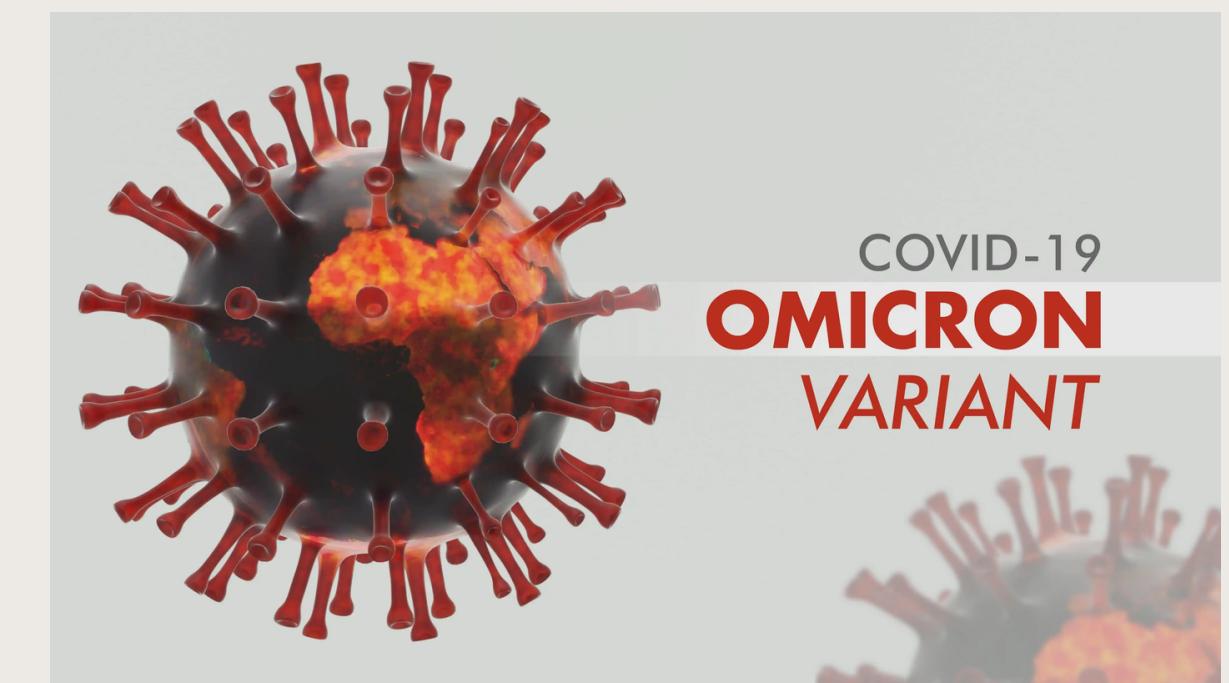
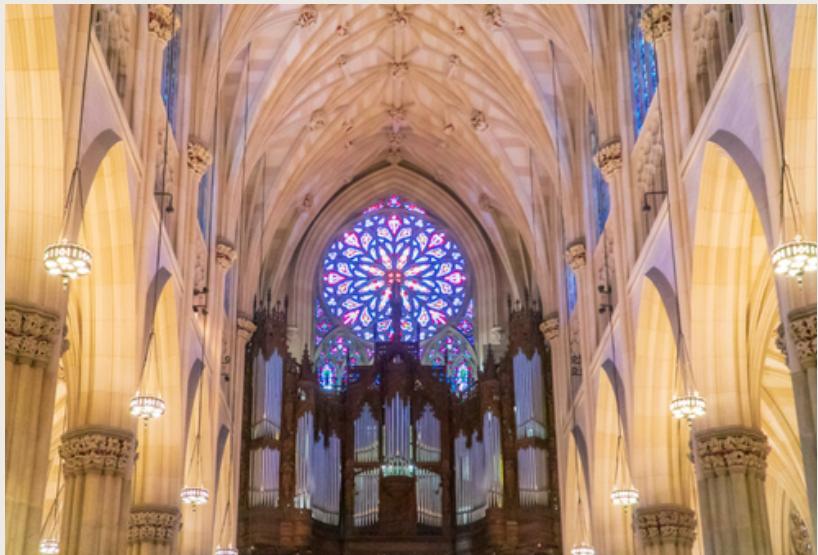
# ทำไมเราถึงต้องรวบรวมข้อมูลเหล่านี้



# Conclusion



ข้อมูลผู้ติดเชื้อและเสียชีวิตจากเชื้อ Covid-19 ของสหรัฐ ในช่วงปี 2020-2022 สามารถสรุปได้ว่า จะมีการเปลี่ยนแปลงได้จากหลายปัจจัย เช่น สถานที่, ภูมิภาค, ฤดูกาล, ความหนาแน่นของประชากร และปัจจัยอื่นๆ ในช่วงเวลาที่มักจะมีผู้ติดเชื้อเพิ่มสูงจะเป็นช่วงต้นปี



# **What do we learn from this?**

จากการภาพ เรายังเห็นได้ว่ารัฐมีการรับมืออย่างไรกับการแพร่ระบาดของเชื้อ และมาตรการการรับมือไหนบ้าง ที่ส่งผลอย่างเห็นได้ชัดว่ามีผลต่อจำนวนของผู้ติดเชื้อ และจำนวนของผู้เสียชีวิต ทำให้สามารถนำไปเป็นแนวทางในการจัดเตรียมมาตรการ การรับมือของเชื้อที่อาจเกิดการระบาดอีกได้ในอนาคต



# References



- Coronavirus pandemic and Omicron variant news
- A Timeline of COVID-19 Developments in 2020
- Lives saved and lost in the first six month of the US COVID-19 pandemic: A retrospective cost-benefit analysis

**THANK YO.U.**

