CS_ELEC_3C | Oct 28, 2025
LAB ACTIVITY - GEOSPATIAL

Baylon, Karyle Zhienelle V.
4CSD

Topic: Geospatial and Temporal Data Visualization
Dataset: All Weekly Excess Deaths (Global)
Software: Python or SAS Viya

After completing this lab, students will be able to:

1. Visualize temporal patterns of excess deaths using Python or SAS Viya.
2. Create geospatial maps showing the distribution of deaths across countries.
3. Combine spatial and temporal analysis to uncover global mortality trends.
4. Interpret insights from visual patterns and relate them to real-world events.

DATASET:
File Name: all_weekly_excess_deaths.csv
Download from Dataset drive
Note: "Excess deaths" refer to the difference between observed and expected deaths, often used to measure the impact of pandemics or disasters.

Task 1: Load and Explore the Dataset

Objective: Understand the data structure and content.

```
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px

# Load the dataset
df = pd.read_csv("all_weekly_excess_deaths.csv")

# Display the first few rows
df.head()
```

	country	region	region_code	start_date	end_date	days	year	week	population	total_deaths	covid_deaths	expected_deaths	excess_deaths	non_covid_deaths	covid_deaths_per_100k	excess_deaths_per_100k	excess_deaths_pct_change
0	Australia	Australia	0	2019-12-30	2020-01-05	7	2020	1	25921089	2926.0	0	2965.372893	-39.372893	2926.0	0.0	-0.151895	-0.013278
1	Australia	Australia	0	2020-01-06	2020-01-12	7	2020	2	25921089	2902.0	0	2956.539560	-54.539560	2902.0	0.0	-0.210406	-0.018447
2	Australia	Australia	0	2020-01-13	2020-01-19	7	2020	3	25921089	2914.0	0	2932.039560	-18.039560	2914.0	0.0	-0.069594	-0.006153
3	Australia	Australia	0	2020-01-20	2020-01-26	7	2020	4	25921089	3007.0	0	2940.372893	66.627107	3007.0	0.0	0.257038	0.022659
4	Australia	Australia	0	2020-01-27	2020-02-02	7	2020	5	25921089	2910.0	0	2900.039560	9.960440	2910.0	0.0	0.038426	0.003435

# Check data types and missing values
df.info()
df.isnull().sum()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5770 entries, 0 to 5769
Data columns (total 17 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   country                    5770 non-null   object
 1   region                     5770 non-null   object
 2   region_code                5770 non-null   int64
 3   start_date                 5770 non-null   object
 4   end_date                   5770 non-null   object
 5   days                       5770 non-null   int64
 6   year                       5770 non-null   int64
 7   week                       5770 non-null   int64
 8   population                 5770 non-null   int64
 9   total_deaths               5770 non-null   float64
 10  covid_deaths               5770 non-null   int64
 11  expected_deaths            5770 non-null   float64
 12  excess_deaths              5770 non-null   float64
 13  non_covid_deaths           5770 non-null   float64
 14  covid_deaths_per_100k      5770 non-null   float64
 15  excess_deaths_per_100k     5770 non-null   float64
 16  excess_deaths_pct_change   5770 non-null   float64
dtypes: float64(7), int64(6), object(4)
memory usage: 766.5+ KB
```

	0
country	0
region	0
region_code	0
start_date	0
end_date	0
days	0
year	0
week	0
population	0
total_deaths	0
covid_deaths	0
expected_deaths	0
excess_deaths	0
non_covid_deaths	0
covid_deaths_per_100k	0
excess_deaths_per_100k	0
excess_deaths_pct_change	0

dtype: int64

Question: Identify at least two important columns you will use for visualization.

Based on the dataset and the provided code, the two important columns that stand out for visualization purposes are:

- `excess_deaths`: This column represents the difference between the observed deaths and the expected deaths for a given period. It provides insights into the extra mortality that could be attributed to factors like COVID-19, making it a key metric for analysis.
- `country` or `region`: These columns provide geographical context, which is essential for creating visualizations like maps or comparing excess deaths across different countries or regions.

Task 2: Process and Prepare the Data

Objective: Ensure that the date and numerical values are ready for analysis.

```
# Convert year_week to a proper date format
df['date'] = pd.to_datetime(df['year_week'] + '-1', format='%G-W%V-%u', errors='coerce')

# Verify conversion
df[['year_week', 'date']].head()
```

	year_week	date
0	2020-W1	2019-12-30
1	2020-W2	2020-01-06
2	2020-W3	2020-01-13
3	2020-W4	2020-01-20
4	2020-W5	2020-01-27

Question: Why is converting the week to a date important for temporal visualization?

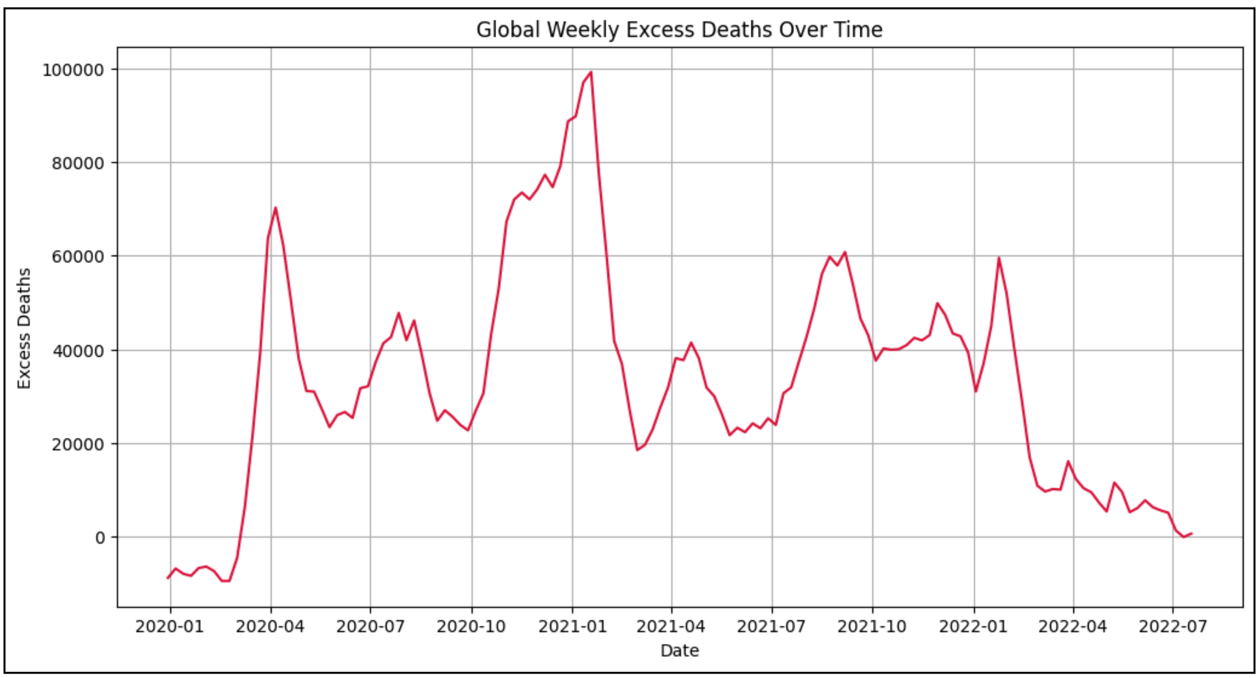
- **Accurate Time Representation**: Weeks are a time period, and visualizations (like line plots or time series graphs) often rely on continuous date or time data. Converting the week into a specific date (such as the start or end of the week) provides a clear and accurate timeline, which makes it easier to analyze trends over time.
- **Consistency in Temporal Analysis**: Without a proper date format, it would be difficult to compare and analyze data across different weeks, especially if we want to aggregate or group by month, quarter, or year. Converting weeks to dates enables consistent and easy handling of temporal data for operations like resampling, time-based grouping, or trend analysis.
- **Visualization Compatibility**: Many visualization libraries (like Plotly and Matplotlib) are optimized for working with date or datetime objects, allowing for smooth rendering of time-based data. Without converting weeks to dates, the analysis might not leverage the full power of these libraries' temporal plotting features.

Task 3: Temporal Visualization

Objective: Examine trends over time.

```
# Aggregate excess deaths globally
global_trend = df.groupby('date')['excess_deaths'].sum().reset_index()

# Plot the global trend
plt.figure(figsize=(12,6))
plt.plot(global_trend['date'], global_trend['excess_deaths'], color='crimson')
plt.title('Global Weekly Excess Deaths Over Time')
plt.xlabel('Date')
plt.ylabel('Excess Deaths')
plt.grid(True)
plt.show()
```



Questions:

a. Identify the periods with highest peaks in excess deaths.

- The highest peaks in excess deaths would correspond to the sharpest rises in the line graph. These spikes usually indicate weeks with unusually high excess deaths, which could be driven by significant events or health crises.

b. What global events could explain these spikes?

- **COVID-19 Pandemic**
  - Given the global nature of the data and the fact that excess deaths can be influenced by pandemics, it's likely that the sharpest peaks in excess deaths align with the height of the COVID-19 pandemic (starting in late 2019 and extending into 2020 and beyond).
- **Other Public Health Crises**
  - Spikes might also occur during other public health emergencies or waves of infection (such as a second or third wave of COVID-19).
- **Natural Disasters or Environmental Events**
  - Major events like hurricanes, floods, or extreme weather events could also lead to a spike in excess deaths.
- **Global Health Initiatives**
  - The beginning or end of large health interventions, such as vaccination campaigns, might have influenced mortality rates, particularly when COVID-19 vaccines were rolled out globally in 2021.

Task 4: Geospatial Visualization (Choropleth Map)

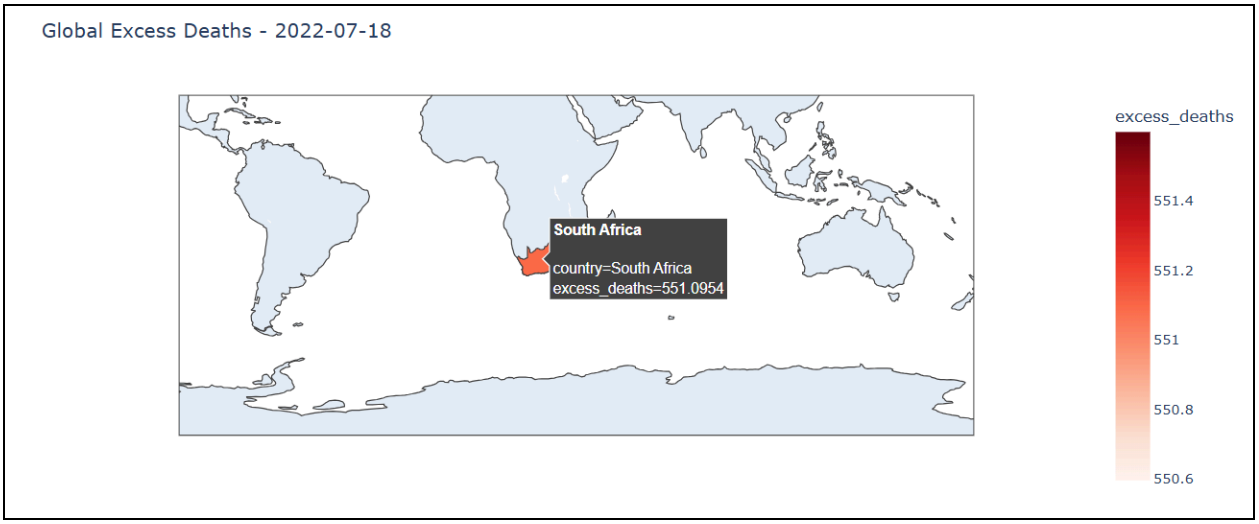
Objective: Observe geographic distribution of deaths.

```
# Get the latest week's data
latest_week = df['date'].max()
latest_data = df[df['date'] == latest_week]

# Create a world map
fig = px.choropleth(
    latest_data,
    locations="country",
    locationmode="country names",
    color="excess_deaths",
    hover_name="country",
    color_continuous_scale="Reds",
    title=f"Global Excess Deaths - {latest_week.strftime('%Y-%m-%d')}"
)
fig.show()
```

CS_ELEC_3C | Oct 28, 2025

LAB ACTIVITY - GEOSPATIAL



Questions:

a. Which regions appear darkest (highest excess deaths)?

- The map highlights **South Africa** as the darkest region, indicating that it experienced the highest excess deaths for the week shown (2022-07-18). The deep red color signifies that this region had notably higher excess deaths compared to others.

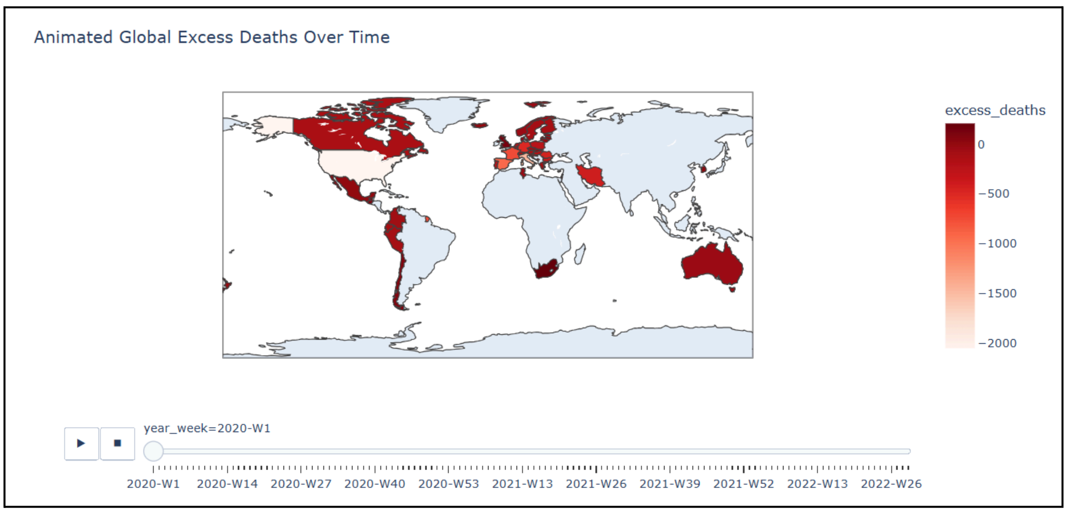
b. What possible socio-economic or health factors might explain this?

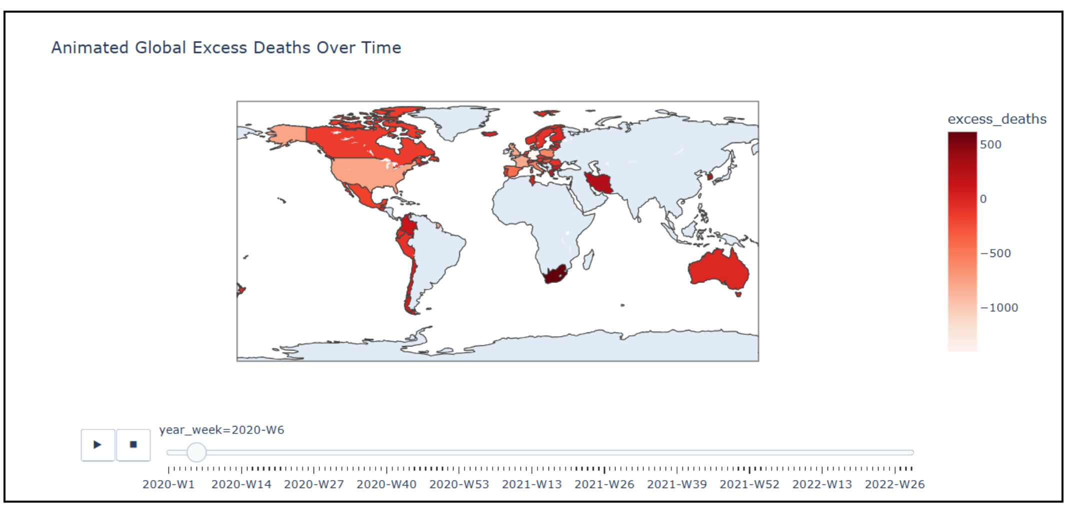
- **COVID-19 Impact**: South Africa endured significant waves of COVID-19, including the **Omicron variant**, which resulted in a higher-than-expected death toll. The country's healthcare system was overwhelmed in some areas, contributing to increased mortality rates.
- **Healthcare System Limitations**: South Africa's healthcare system has faced ongoing struggles, including limited access to care, particularly in rural regions. This likely worsened excess death numbers, especially during health crises like the COVID-19 pandemic.
- **Social Determinants of Health**: High levels of **poverty**, poor access to quality healthcare, and widespread underlying health conditions, such as **HIV/AIDS** and **tuberculosis**, intensified the pandemic's impact, leading to more excess deaths than anticipated.
- **Inequality**: As one of the most unequal countries globally, South Africa has significant disparities in healthcare, education, and income. These inequalities contribute to poorer health outcomes and higher death rates in marginalized populations, which may have driven the excess mortality.

Task 5: Spatiotemporal Animation

Objective: Show how excess deaths evolved across countries over time.

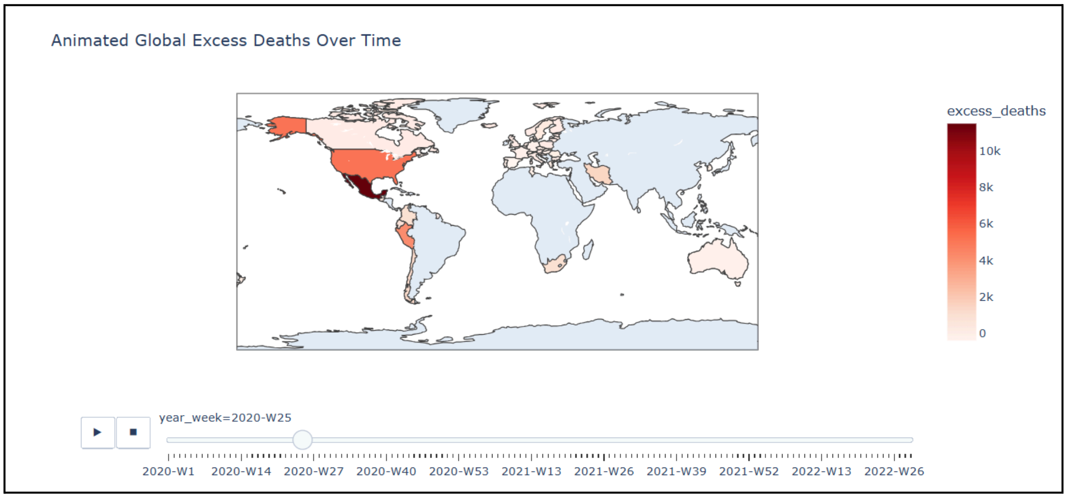
```
fig = px.choropleth(
    df,
    locations="country",
    locationmode="country names",
    color="excess_deaths",
    animation_frame="year_week",
    title="Animated Global Excess Deaths Over Time",
    color_continuous_scale="Reds"
)
fig.show()
```

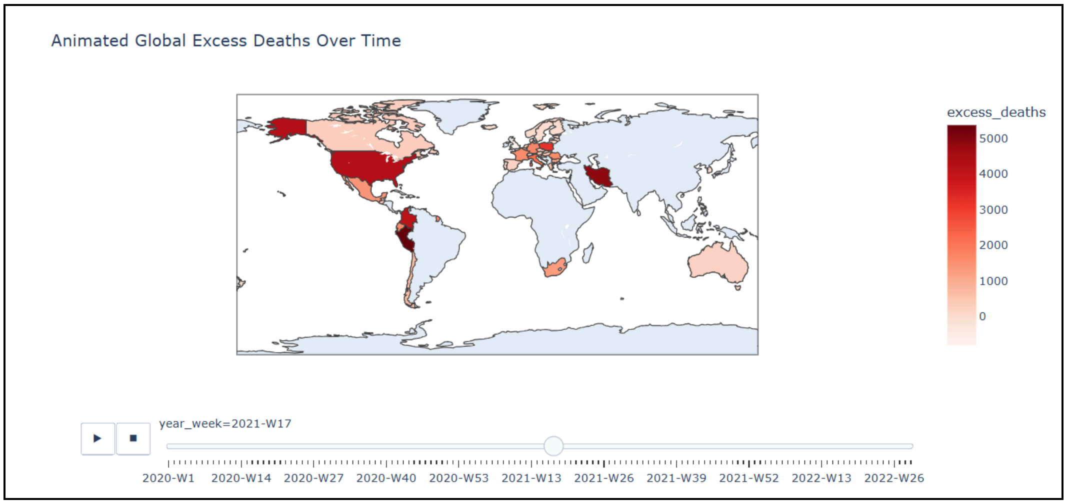


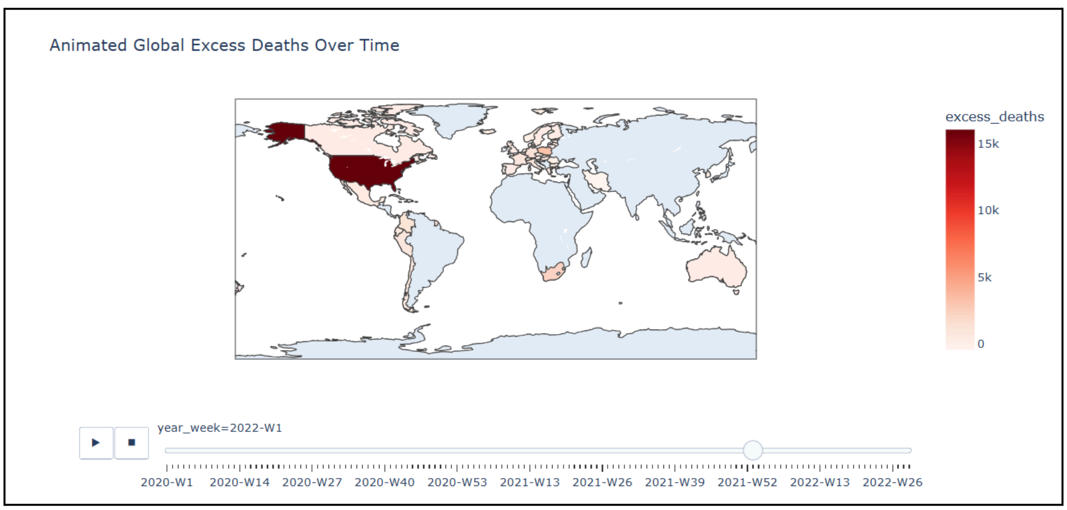


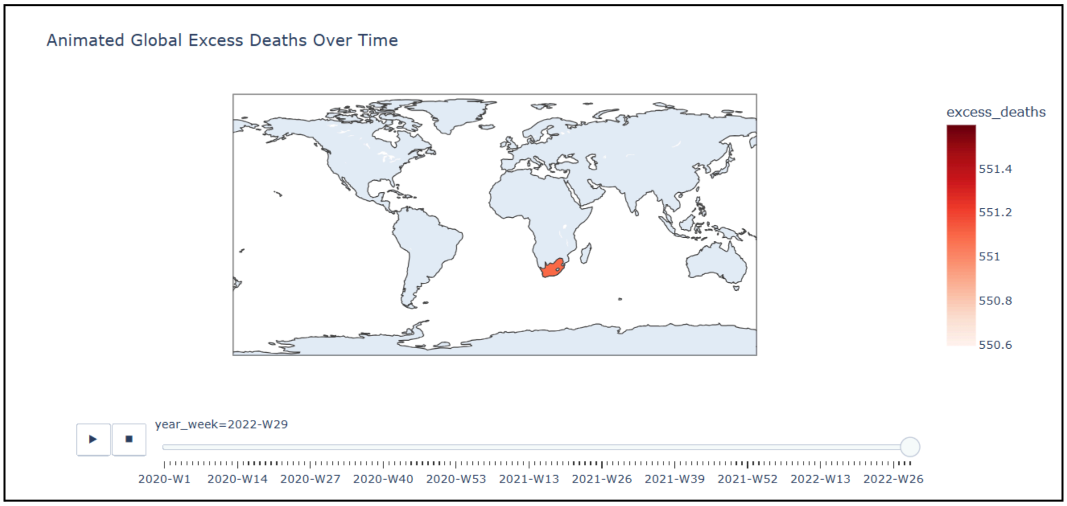
CS_ELEC_3C | Oct 28, 2025

LAB ACTIVITY - GEOSPATIAL









***Question:* Observe the animation and describe how the spread of high-excess-death regions changes through the timeline.**

- **Early 2020 (2020-W1 to 2020-W5)**
  - At the beginning of 2020, **countries in Europe**, such as **Italy**, **Spain**, and **France**, along with the **United States**, began to show signs of high excess deaths. These countries were the first to be heavily impacted by the COVID-19 pandemic. The **dark red regions** on the map during these weeks reflect the overwhelming burden placed on healthcare systems by the initial surge in COVID-19 cases. The virus spread rapidly, especially in European countries, leading to significant excess mortality as healthcare systems struggled to cope.
- **Mid 2020 (2020-W6 to 2020-W24)**
  - As the pandemic continued to spread, excess deaths were particularly high in **North America** (especially the **United States**) and **Brazil**, alongside continued outbreaks in **Europe**. Countries like the **United Kingdom**, **Italy**, **Spain**, and **France** experienced some of the highest excess death rates. This period saw the first wave of COVID-19 reach its peak, with deaths attributed to both the virus itself and overwhelmed healthcare systems, exacerbated by limited testing and early stages of medical responses. The **dark red** regions in these areas reflected a high mortality burden.
- **Late 2020 to Early 2021 (2020-W25 to 2021-W16)**
  - The **United States** and **Europe** (especially **Italy**, **Spain**, **United Kingdom**) continued to see very high excess deaths as the second wave of COVID-19 emerged in the fall of 2020. The second wave was aggravated by the winter months, when the virus spreads more easily. By early 2021, the **Indian**

subcontinent, particularly **India**, began showing noticeable spikes in excess deaths, largely driven by the rise of the **Delta variant**. This period also saw **Brazil** facing high mortality rates due to both the virus and systemic health challenges.

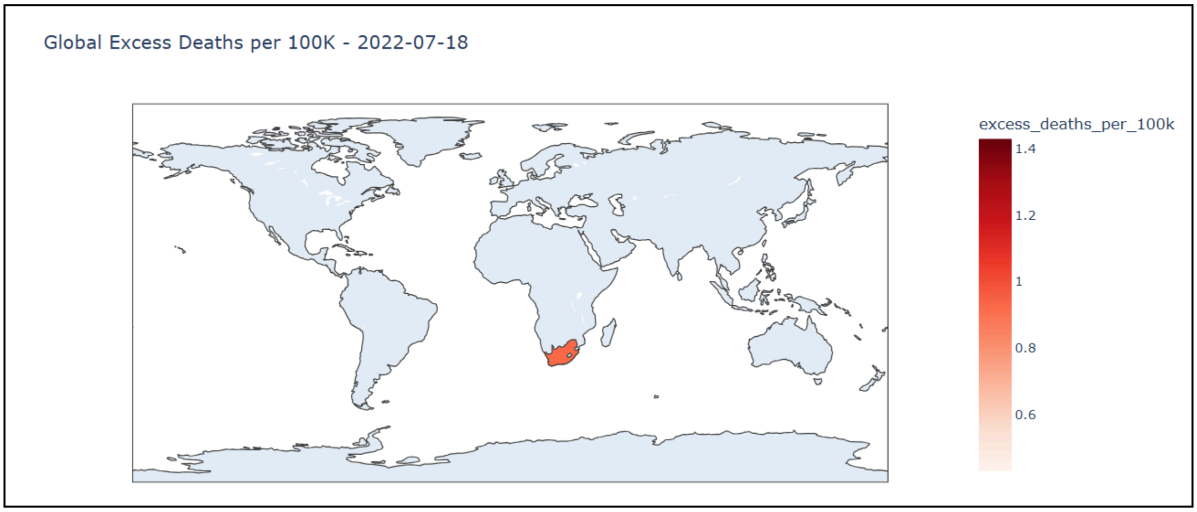
- **Mid 2021 (2021-W17 to 2021-W28)**
  - During this period, **North America** and **South America** continued experiencing significant excess death rates due to ongoing COVID-19 transmission, while some European countries (**France**, **Italy**, **Germany**) remained affected. **Africa** had a relatively lower impact compared to other regions, though countries like **South Africa** still faced some significant increases in excess deaths. The **United States** and parts of **Latin America** (such as **Brazil**) continued to struggle with high mortality rates, despite the development of vaccines.
- **2022 and Beyond (2022-W1 to 2022-W24):**
  - By 2022, the number of regions with extreme excess deaths began to decrease, with fewer areas showing dark red regions of excess mortality. However, **North America** and **Europe**, particularly the **United Kingdom** and **Italy**, still experienced higher-than-usual mortality rates, albeit at lower levels than during the earlier waves. In **South Africa**, there was a significant uptick in excess deaths following the **Omicron** variant wave. The high death toll during this period was influenced not only by the virus itself but also by **socio-economic factors**, including limited access to healthcare and vaccine hesitancy.

Task 6: Normalization

Objective: Normalize data for fair comparison.

df['excess_deaths_per_100k'] = (df['excess_deaths'] / df['population']) * 100000
Then repeat the choropleth using the new variable.

	country	excess_deaths	population	excess_deaths_per_100k
0	Australia	-39.372893	25921089	-0.151895
1	Australia	-54.539560	25921089	-0.210406
2	Australia	-18.039560	25921089	-0.069594
3	Australia	66.627107	25921089	0.257038
4	Australia	9.960440	25921089	0.038426



Question*: *How does normalization change your interpretation?

Normalization in this context involves adjusting the excess death figures by considering the population size of each country, specifically by calculating excess deaths per 100,000 people. This approach helps provide a clearer understanding of the impact on mortality, especially when comparing countries with varying population sizes.

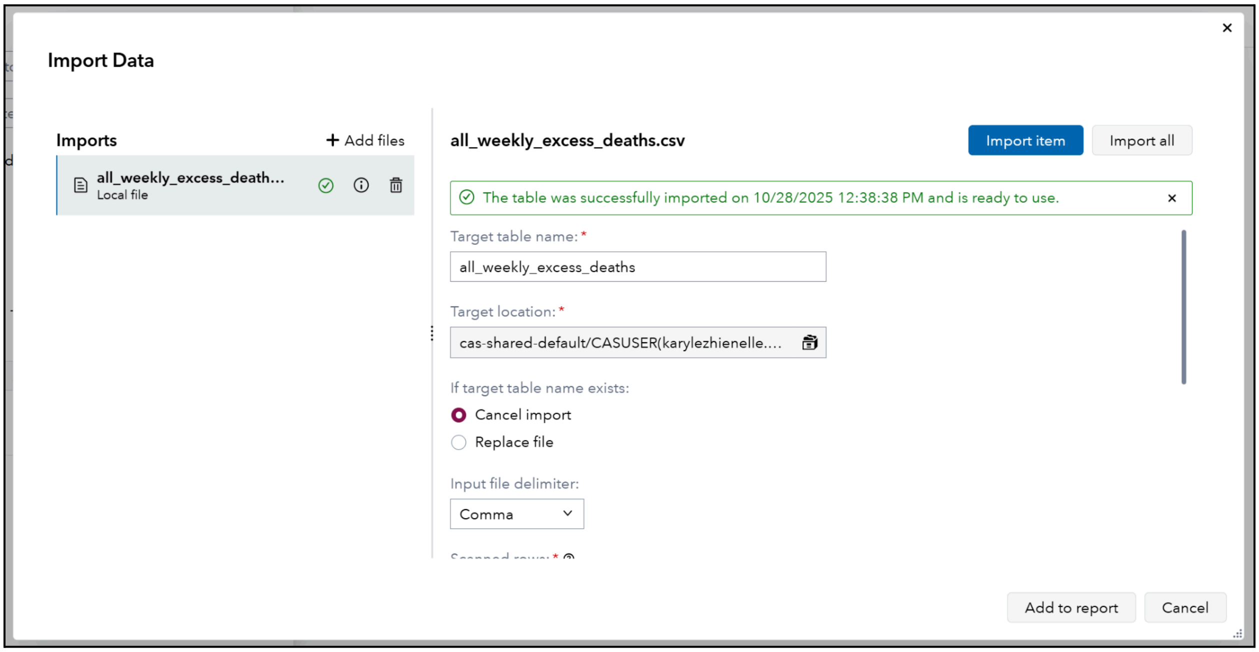
- **Comparison Across Countries of Different Sizes**
  - Without normalization, countries with larger populations, such as **China** or **India**, would naturally report higher total excess deaths, even if the rates of excess mortality aren't as significant. Normalizing the data by using a per 100,000 scale allows for more accurate comparisons between countries of different sizes.
- **Improved Visualization of Regional Impact**
  - Normalization provides a more accurate visualization of the regional impact of excess deaths. Countries or regions that have a higher number of deaths per 100,000 will stand out, highlighting areas where the pandemic has disproportionately affected the population. Without normalization, larger countries could dominate the visualization, even if their per capita mortality rates are lower.
- **Spotting Outliers**
  - By normalizing the excess death data to a per 100,000 basis, it's easier to identify regions or countries that have been disproportionately impacted. For instance, **Australia** or certain **European countries** might show up with high excess deaths per 100,000, even if their total death counts aren't among the highest.

CS_ELEC_3C | Oct 28, 2025

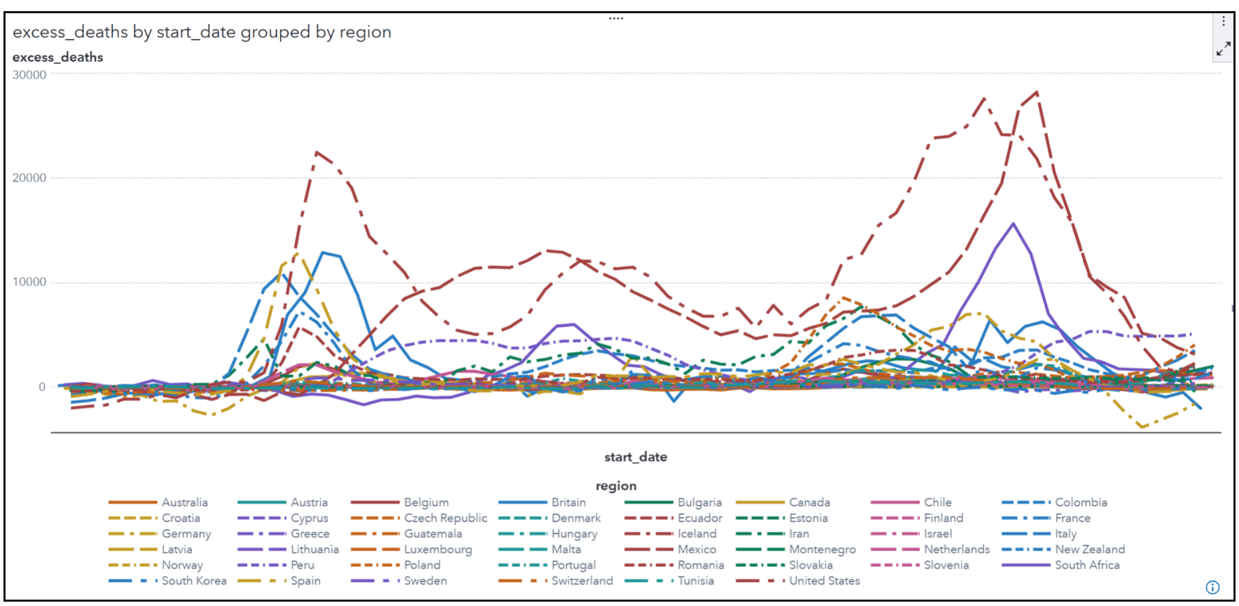
LAB ACTIVITY - GEOSPATIAL

Using SAS Viya:

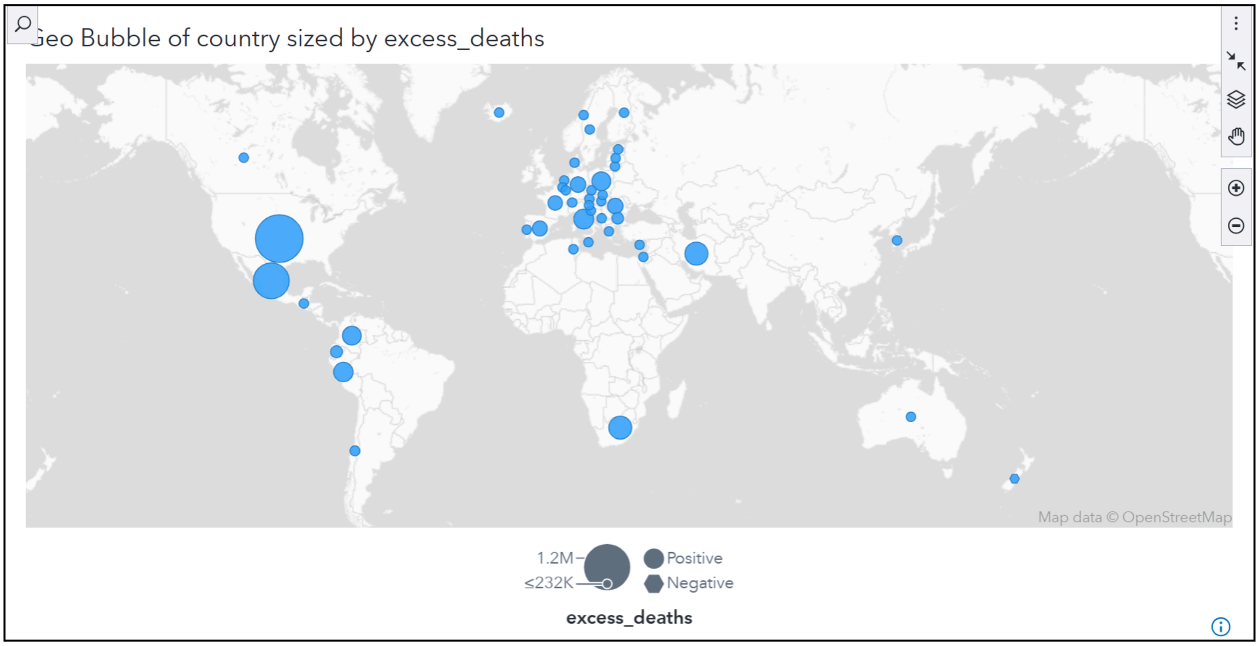
1. Upload the CSV file to SAS Viya.



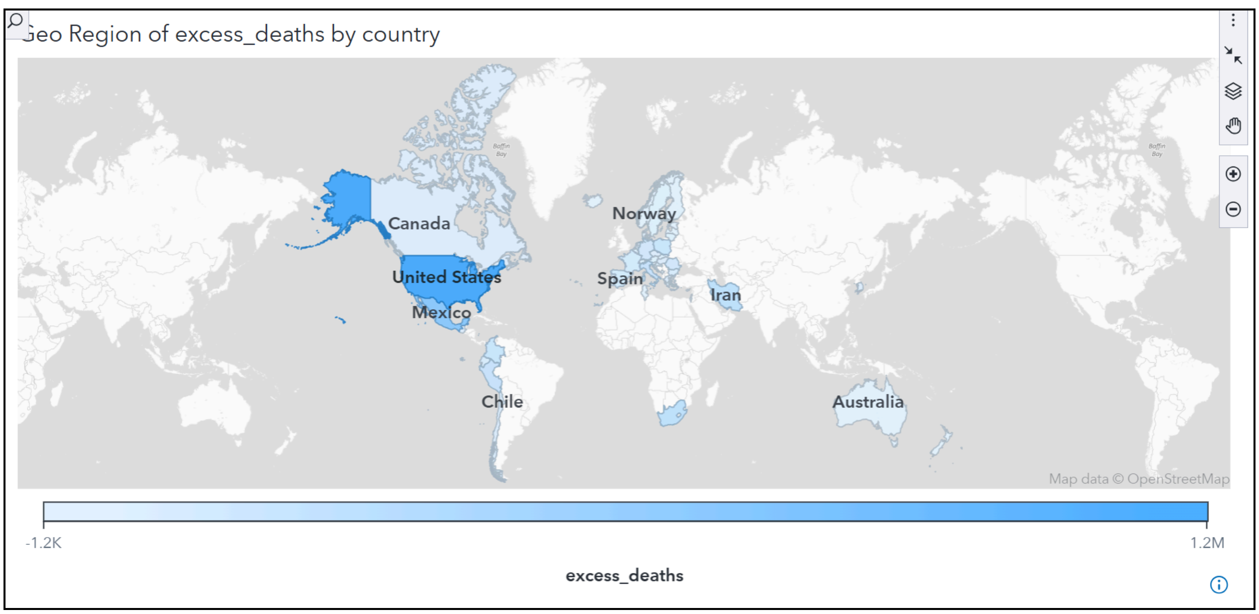
2. Open Visual Analytics → New Report.
3. Add a Line Chart:
   - X-axis: date
   - Y-axis: excess_deaths
   - Group by: region



4. Add a Geo Map:
    - Geography role: country
    - Color/Size role: excess_deaths

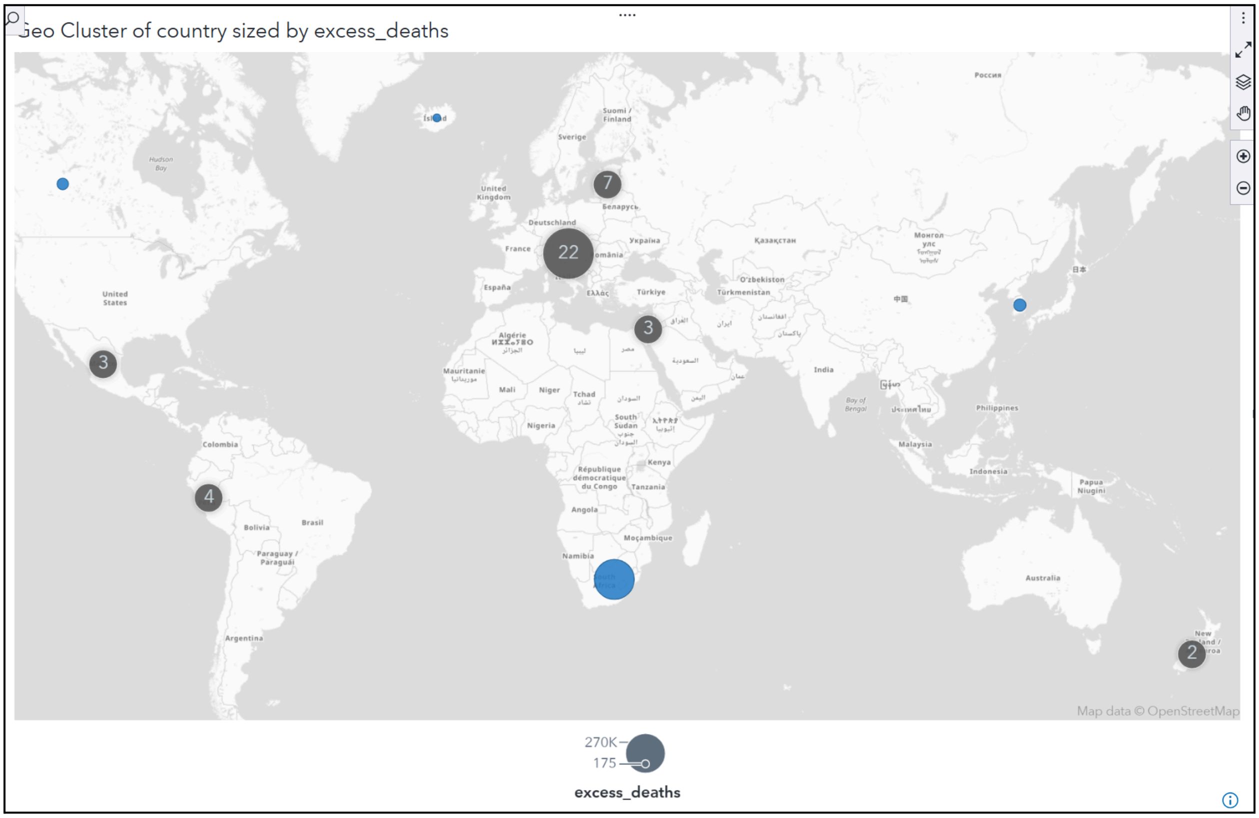


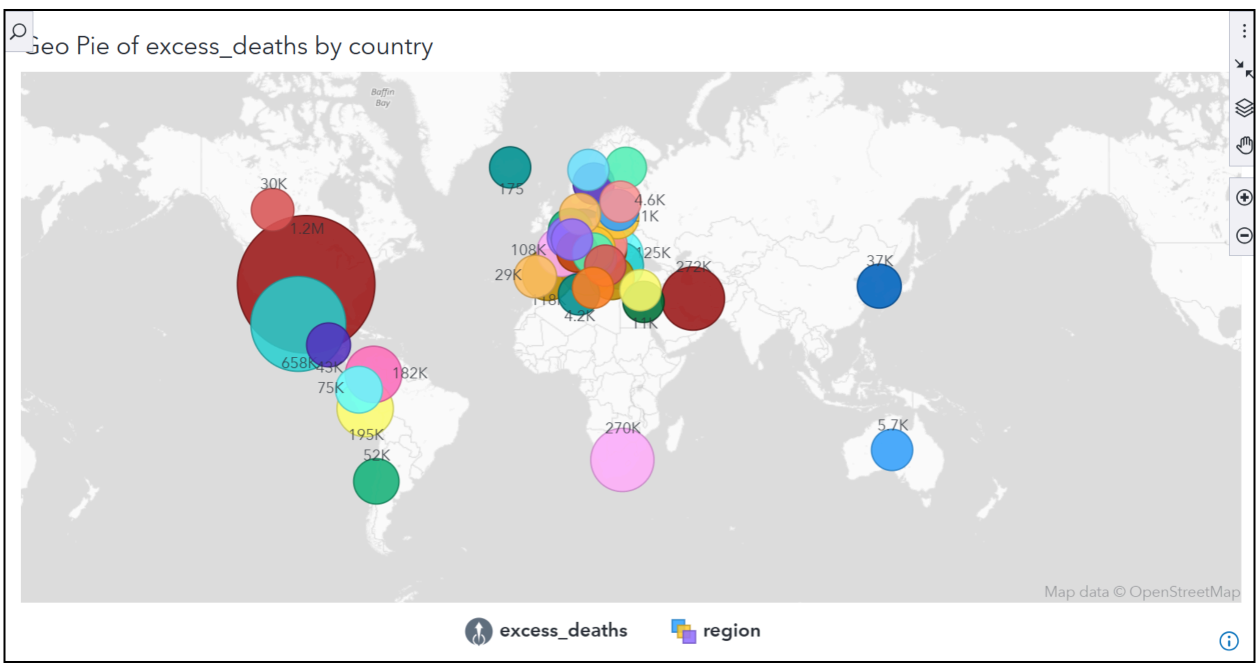
5. Try different geo-spatial charts and answer the questions below.



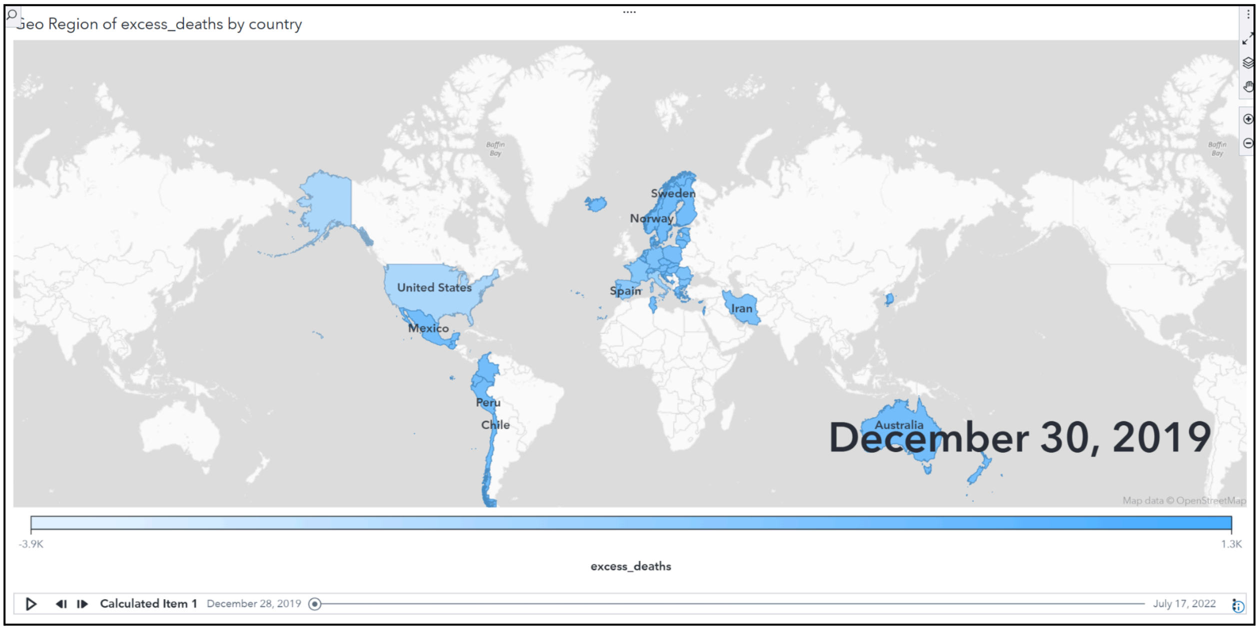
CS_ELEC_3C | Oct 28, 2025

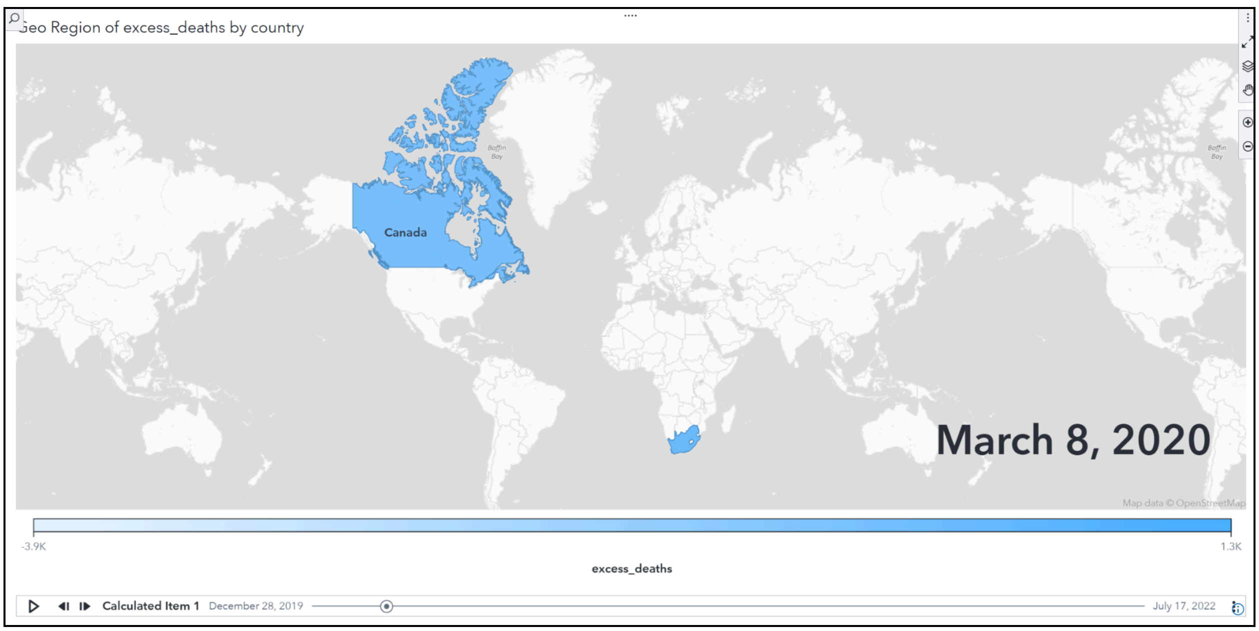
LAB ACTIVITY - GEOSPATIAL





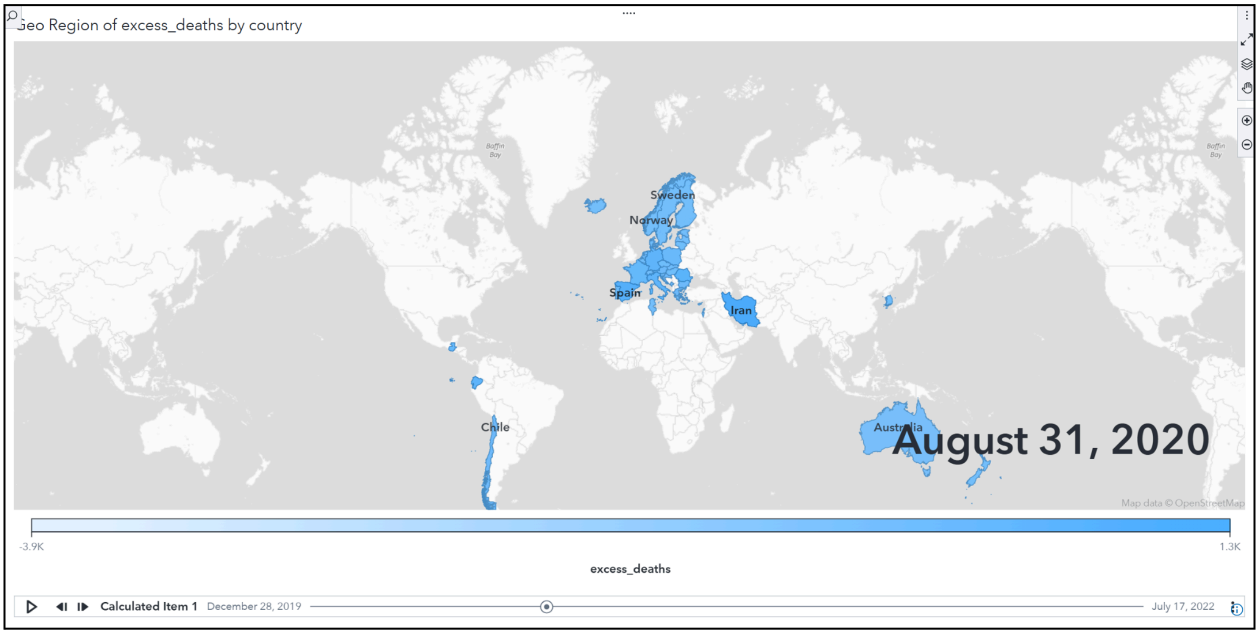
6. Add an Animation role using date to show temporal change.

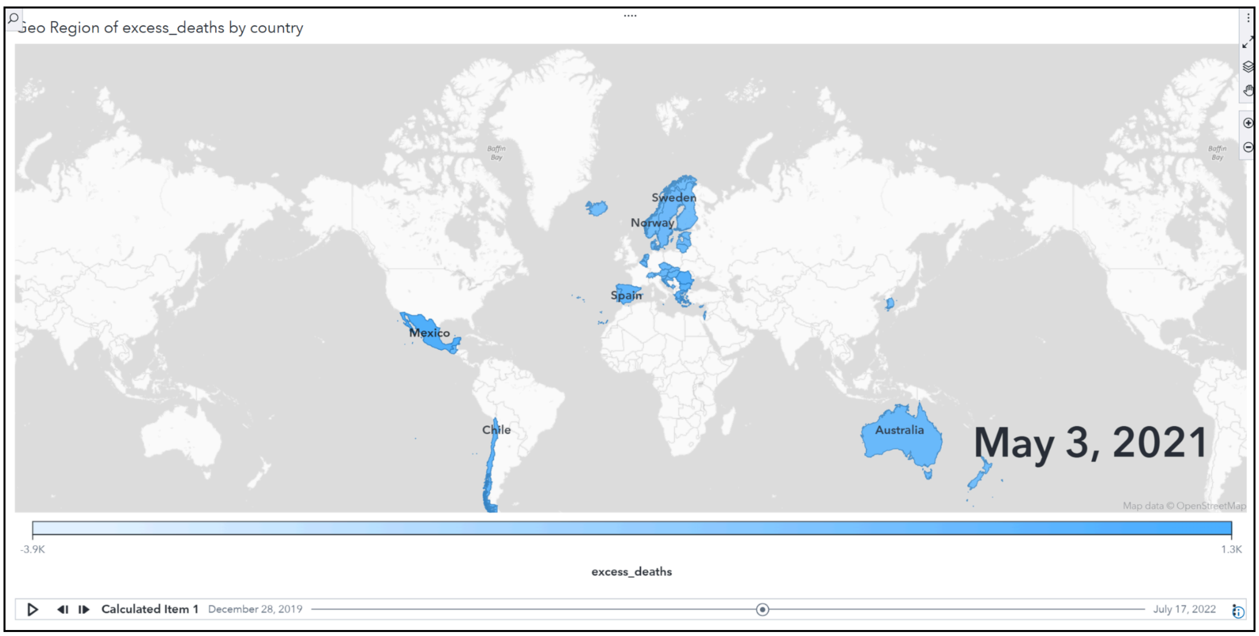




CS_ELEC_3C | Oct 28, 2025

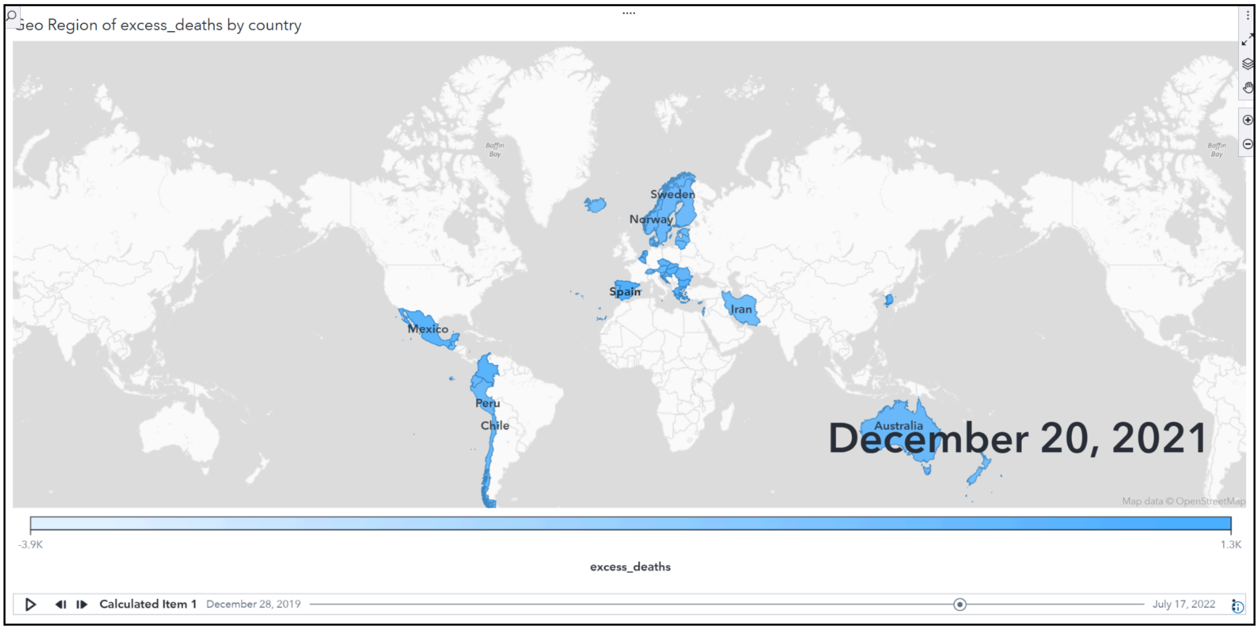
LAB ACTIVITY - GEOSPATIAL





CS_ELEC_3C | Oct 28, 2025

LAB ACTIVITY - GEOSPATIAL



LAB ACTIVITY - GEOSPATIAL

INSIGHTS AND ANALYSIS:

1. ***Which countries show consistent excess mortality across time?***
   - **Australia** and **Chile** show relatively consistent fluctuations in excess deaths, with periods of both negative and positive excess deaths.
   - **South Africa** has more variability in the excess deaths, with some weeks showing significant positive excess deaths.
   - **United States** seems to experience sharp fluctuations, particularly with negative excess deaths followed by more pronounced positive values.
2. ***What might explain sharp spikes or sudden declines in certain weeks?***
   - The **United States** and **South Africa** show sharp declines or spikes in excess deaths, which may align with specific events like the peak of the COVID-19 pandemic or other significant public health events.
   - The **declines** could reflect periods when fewer deaths occurred than expected, possibly due to health interventions or other factors (e.g., fewer deaths from non-COVID causes during certain periods).
   - Conversely, **sharp spikes** typically align with the height of pandemics, natural disasters, or other health crises that cause more deaths than expected.

3. ***Give additional insights you got from your visualizations.***
   - **COVID-19 Surge Impact**: The United States and South Africa experienced significant spikes in excess deaths, particularly in 2020, indicating the heavy toll of the COVID-19 pandemic.
   - **Stable Trends in Australia and Chile**: These countries show relatively stable trends in excess deaths, suggesting effective public health measures and less severe impacts from COVID-19.
   - **South Africa's Volatility**: South Africa displays greater fluctuations in excess deaths, which may reflect a combination of COVID-19 waves, other health crises, and healthcare system challenges.
   - **Excess Deaths Reflecting Health System Strain**: Countries like the U.S., with major fluctuations in excess deaths, may indicate periods of healthcare strain or system overload, particularly during major health crises.
   - **Possible Seasonal Effects**: Smaller fluctuations in countries like Australia might align with seasonal factors, such as flu seasons, rather than being driven by major public health events like the pandemic.