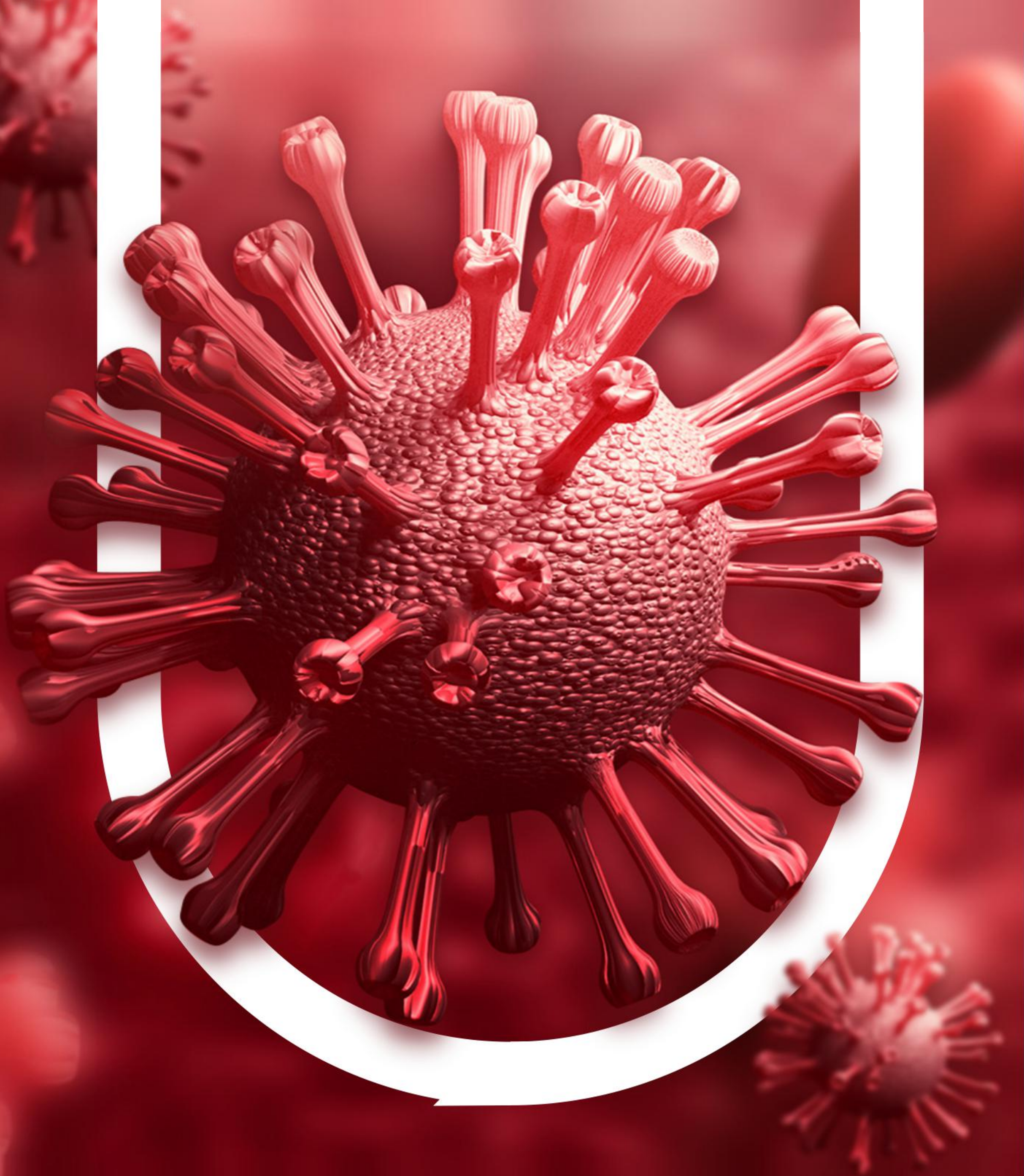
Understanding the Variances in COVID-19 Pandemic Outcome - Excess Mortality - with Social, Cultural, and Environmental Factors

Instructor: Professor Maryam Gooyabadi

Presenter: Jiner Zheng



01 INTRODUCTION

02 DATA & METHODS

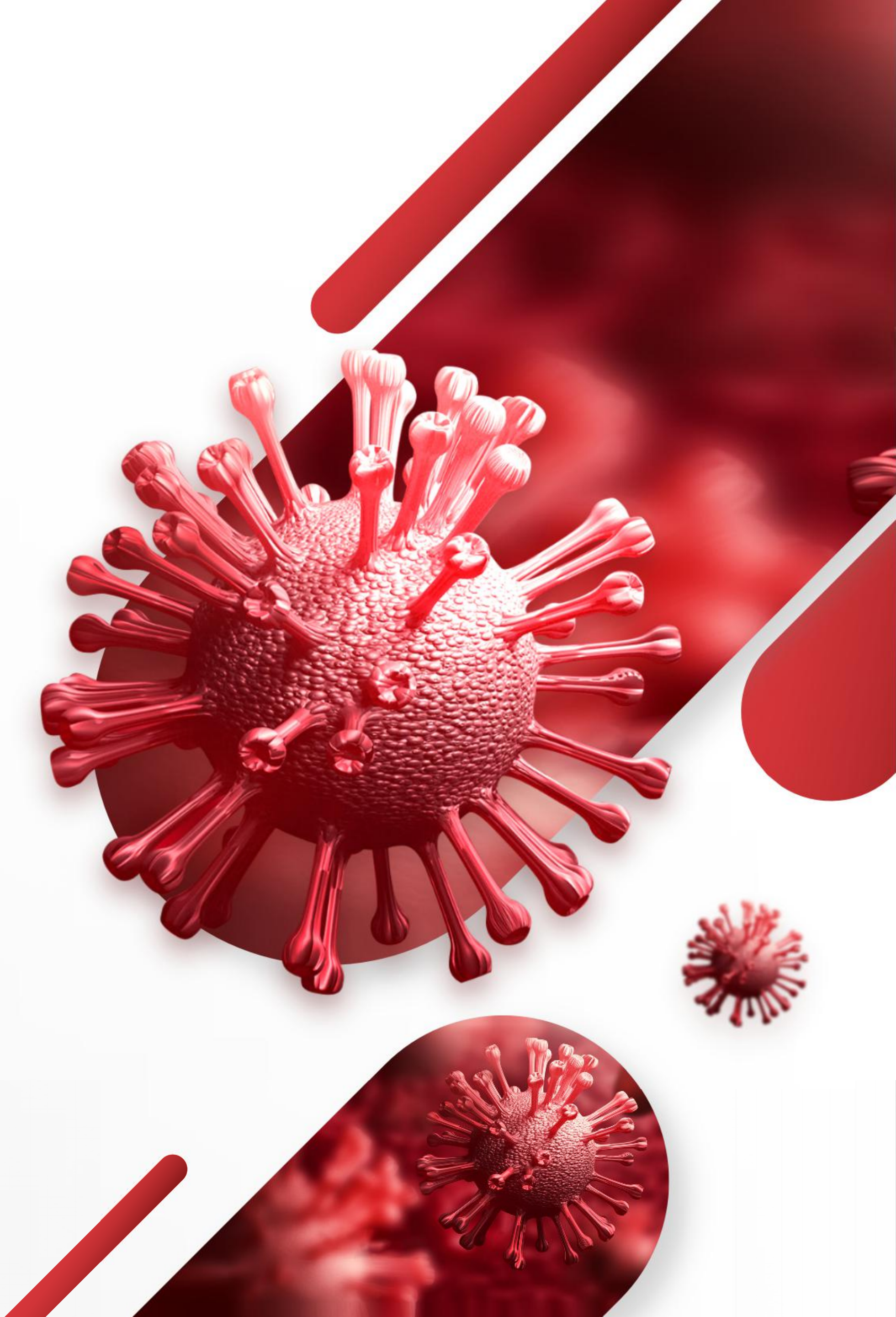
03 RESULTS

04 CONCLUSIONS

COVID-19

The importance of analyzing COVID-19 excess mortality

- The World Health Organization (WHO) has declared COVID-19 **a global pandemic** on March 11th, 2020, as it has infected billions of people across 213 countries and become a worldwide threat.
- Statistical analyses and forecasting methods are needed to monitor **the overall burden on mortality** and help governments allocate medical resources and hospitalizations.
- **Excess mortality** has been found as the most reliable statistics among different measurements related to the effects of COVID-19.



Social, economic, and cultural impacts on COVID-19

- Cultural values have been found to influence COVID-19 morbidity and mortality.
- Cultural tightness was found to seemingly help **reduce** Covid-19 **morbidity and mortality** (Cao et al., 2020; Gelfand et al., 2021).
- Kumar (2021) found that societal cultures affected the morbidity and mortality of COVID-19 through various cultural dimensions.
  - For example, cultural dimensions including **power distance, institutional collectivism, and performance orientation** were found **negatively** associated with COVID-19 outcomes.
- Cao, Li, and Liu (2020) used a hierarchical regression method to explore the impact of national cultures on the containment of COVID-19.
  - Findings suggested that **cultural looseness and individualism** resulted in **faster increases** in the response measures of COVID-19 outcomes (changes in the prevalence rate, crude mortality rate, and case fatality rate).

- GDP per capita can impact a country's capacity to detect, prevent, and respond to the COVID-19 outbreak (Kandel et al., 2020).
- Median age and the share of population older than 65 years were found positively associated with COVID-19 mortality rate (Cao et al., 2020; Liang et al., 2020; Upadhyaya et al., 2020).

Main Variables of Interests

01 COVID measure

- Excess Deaths
- P-score of Excess Deaths

03 Cultural Factor

- Government Effectiveness
- Preventative Interventions
- Personal Freedom
- Freedom of Assembly & Association
- Social Network
- Personal & Family Relationships
- Civic & Social Participation

02 Social/Economic Factors

- Policy Stringency Index
- Poplution & population density
- Median Age & Aged 65 older
- GDP per capita
- Life Expectancy

04 Enviornmental Factor

Climate Class

- Class A - Tropical
- Class B - Dry (Arid & Semiarid)
- Class C - Temperate (Mesothermal)
- Class D - Continental ((Microthermal)

Excess Deaths

World Mortality Dataset (Ariel Karlinsky & Dmitry Kobak, 2021)

Definition of Excess Deaths:

- the increase in deaths from all causes during a crisis relative to ‘normally’ expected deaths, mortality in the absence of exceptional events such as pandemics

$$\text{Excess Deaths} = \text{Reported Deaths} - \text{Expected Deaths}$$

Cross-country comparison metric

$$\text{P-score} = \frac{\text{Reported Deaths} - \text{Expected Deaths}}{\text{Expected Deaths}} * 100$$

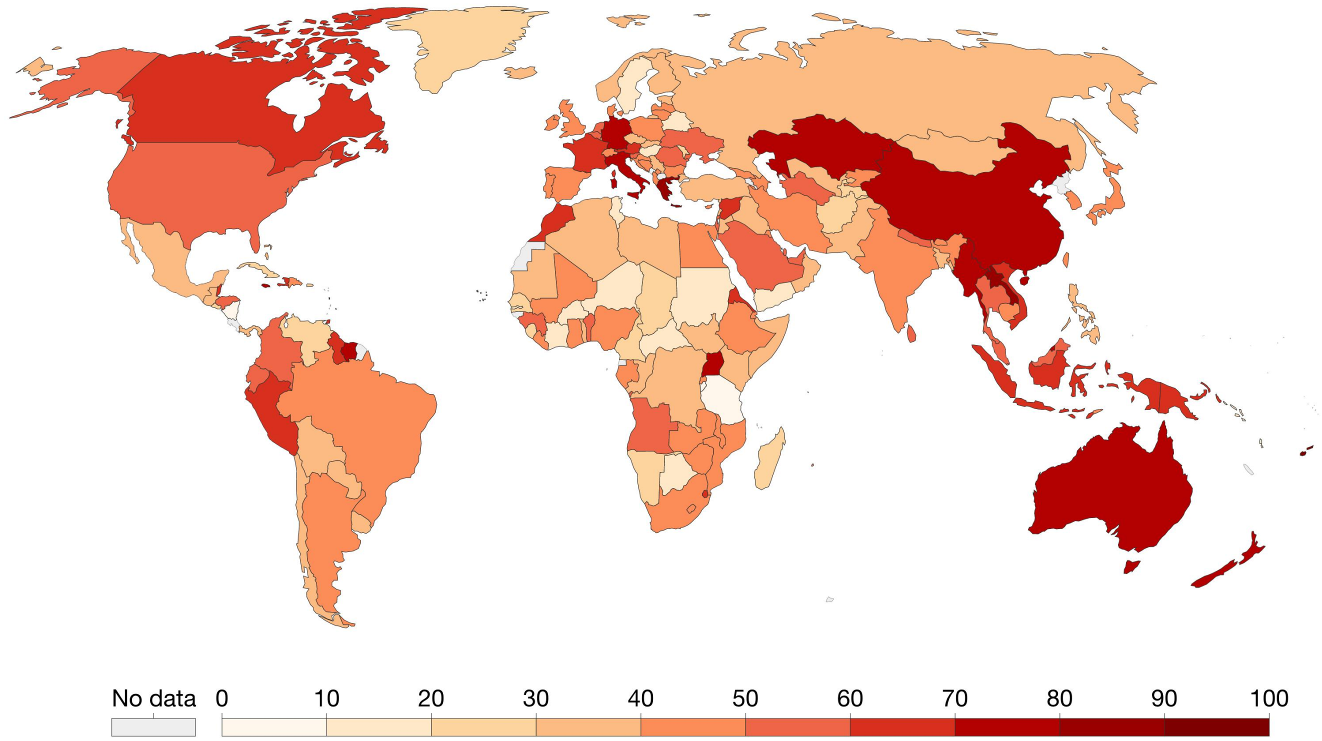
Policy Stringency Index

The Oxford Coronavirus Government Response Tracker (OxCGRT) project calculated the Policy Stringency Index as a composite measure of nine of the response metrics.

COVID-19 Stringency Index

Our World in Data

The stringency index is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest). If policies vary at the subnational level, the index shows the response level of the strictest subregion.



Source: Oxford COVID-19 Government Response Tracker, Blavatnik School of Government, University of Oxford – Last updated 5 December 2021, 07:50 (London time)
OurWorldInData.org/coronavirus •

Cultural Factors

Legatum Institute's 2019 Prosperity Index

- **Government Effectiveness:** the quality of public health provision, the competence of officials, and the quality of the bureaucracy
- **Health Preventative interventions:** the extent to which the health system prevents diseases and other medical complications from occurring
- **Freedom of Assembly and Association:** the degree to which citizens have the freedom to assemble with others in public spaces, or to express their opinions
- **Personal Freedom:** the degree to which citizens are free from restriction and are free to move or act independently
- **Social Network:** the strength and opportunities of the individual's relationships with the wider social network
    - e.g. opportunity to make friends or helping another household
- **Personal and Family Relationships:** the strength of the closest personal relationships and family ties
- **Civic and Social Participation:** the amount to which citizens participate within the society

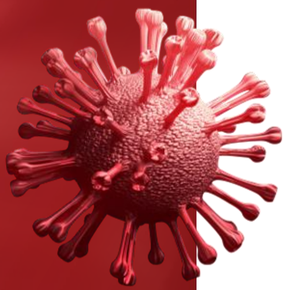
Other Covariates

Köppen-Geiger updated climate zones by Kottek et al. (2006), using the downscaling algorithms of Rubel et al. (2017)

- **Class A - Tropical climates:** geographic regions with tropical monsoon, rainforests, or savannas
- **Class B - Dry climates:** hot or cold deserts, or semiarid climates
- **Class C - Temperate climates:** subtropical, mediterranean, and oceanic climates
- **Class D - Continental climates:** continental and subarctic climates

Our World in Data

- GDP per capita / Median Age / Aged 65 older / Population & Population Density / Life Expectancy

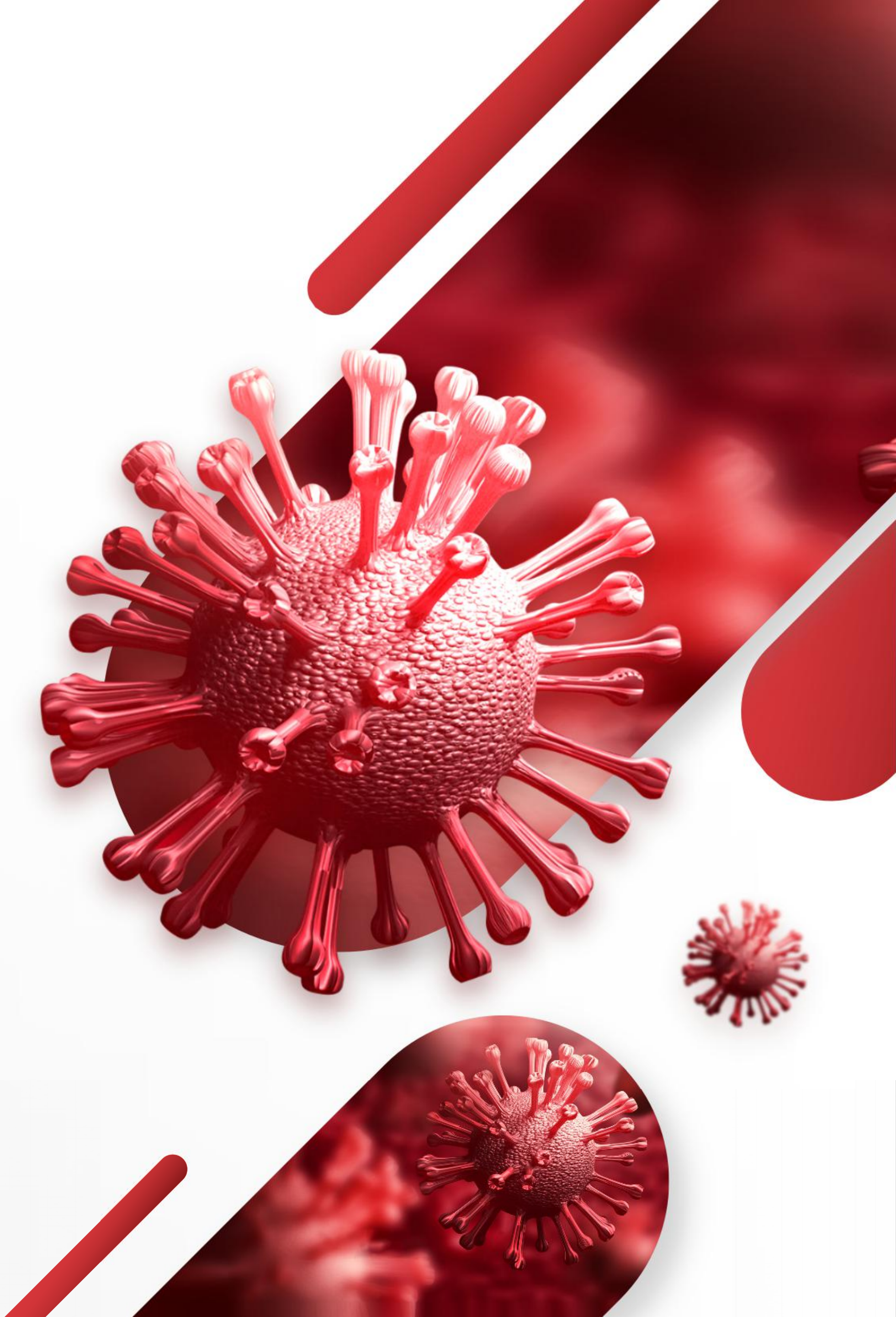


Data Sample

70 Countries/Regions

51% in Europe (n=36), 23% in Asia (n=16), 11% in South America (n=8), 6% in Africa (n=4), 6% in North America (n=4), 3% in Oceania (n=2)

- **Mean GDP per capita: 30,394**
- **Mean Median Age: 38 years old**
- **Mean population density: 354.48**



02

Analytical Methods

Part 1 — Time Series Data Modeling (To compute Excess Deaths) -- ARIMA/GARCH

Part 2 — Clustering Time Series Variables (Excess Deaths & Policy Stringency Index) -- Dynamic Time Warping (DTW)

Part 3 — Clustering All Mix Type Variables -- Gower Distance & K-medoids

Computing Excess Deaths (ARIMA/GARCH)

- A hybrid of **ARIMA** and **GARCH** models were used to estimate the number of **expected deaths** using historical data on deaths from all causes in 2015–2019.

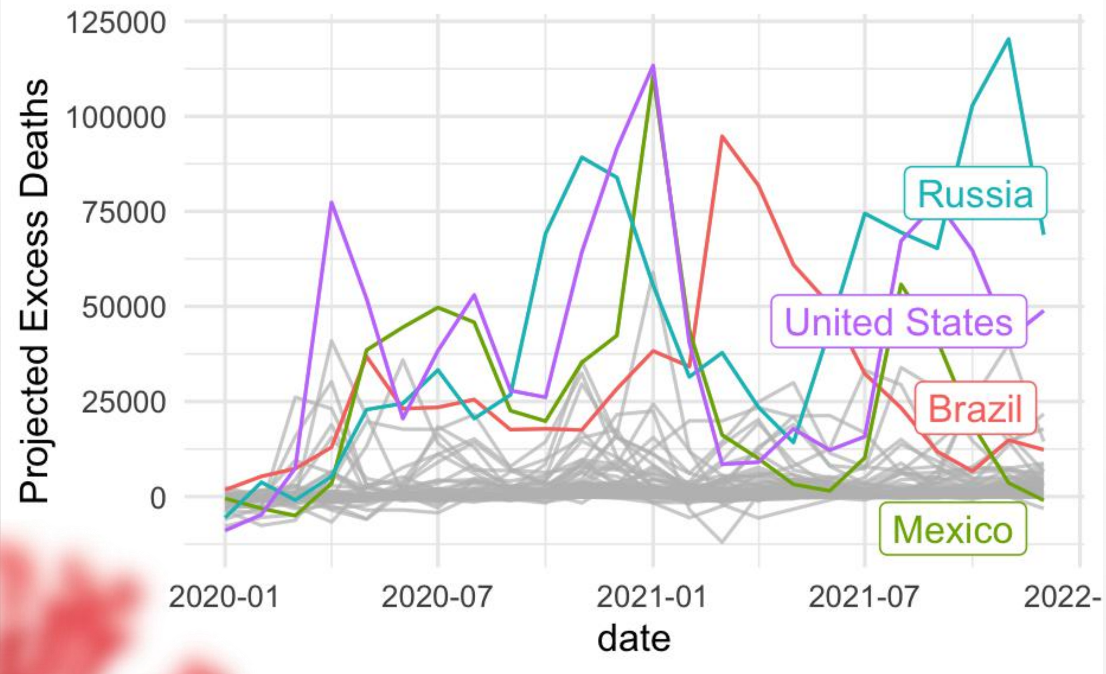
- An **ARIMA** model can be understood by three major components:
  - **Autoregression (AR) --> ** dependency between current observations and its lagged values (i.e. values in prior time periods)
  - **Integrated (I):** the degree of differencing between values at current time and previous valus to make time series data become stationary
  - **Moving Average (MA):** the moving average process apllied to lagged values when current observations also depend on lagged residual errors

- **GARCH** (Generalized Autoregressive Conditional Heteroskedasticity): models the changes in variance of time series or volatility that are time-dependent.

- This hybrid approach can capture seasonal variations in deaths, yearly trends over recent years due to population growths and other social, economic factors, short-term serial correlations, and the potential conditional heteroscedasticity in mortality time series.

1. **Model equation: ARIMA(p,d,q)(P,D,Q)[m]+GARCH(p',q')**

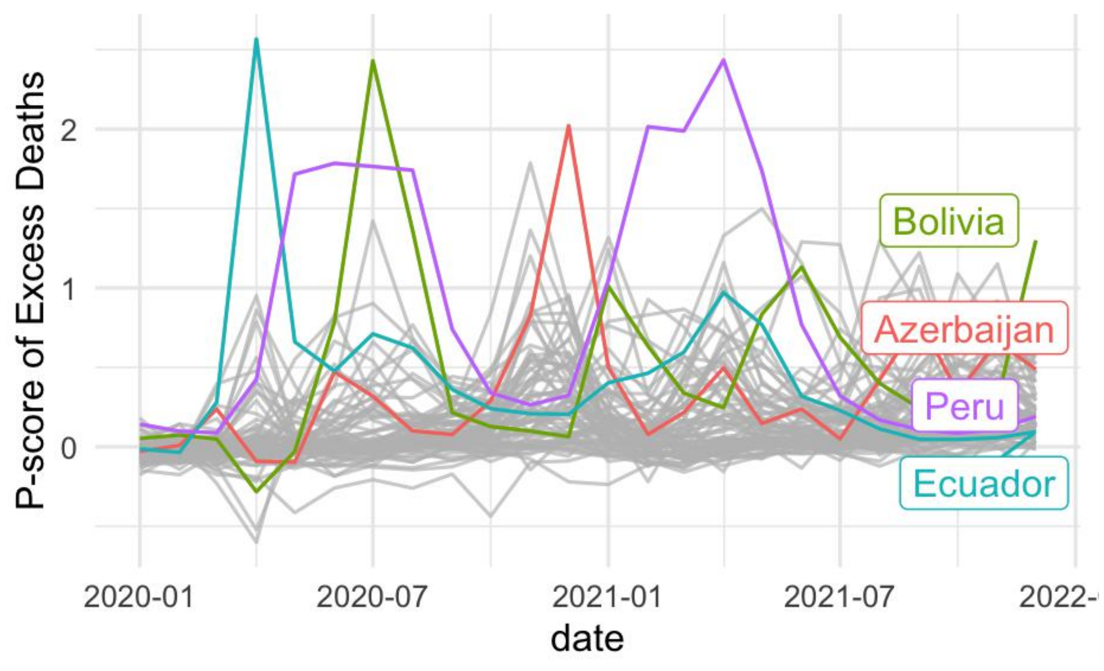
$$D_{t,Y} = c + \sum_{i=1}^{p} \alpha_i D_{t-i,Y} + D_{t,Y} - D_{t-d,Y} + Z_t + \sum_{j=1}^{q} \lambda_j Z_{t-j,Y}$$
$$+ \sum_{k=1}^{P} \alpha_k D_{t-k\cdot m,Y} + D_{t,Y} - D_{t-D\cdot m,Y} + \sum_{l=1}^{Q} \lambda_l Z_{t-Q\cdot m,Y}$$
$$+ \delta^2_{t-q_2,Y}\, \epsilon_{t-p_2,Y}.$$



2. **Final estimate of excess mortality: summation across all time intervals**

$$\hat{E} = \sum_{t \geq t_1} (D_{t,2020} - \hat{D}_t) + \sum_{t} (D_{t,2021} - \hat{D}_t).$$

$$\text{P-score} = \frac{\text{Reported Deaths} - \text{Expected Deaths}}{\text{Expected Deaths}} * 100$$



Clustering Time Series Data

- Most popular methods: Euclidean (most efficient) vs. Dynamic Time Warping - DTW (most accurate)
  - Euclidean distance takes pairs of datapoints and compares them to each other.
  - DTW calculates the smallest distance between all points --> enables a one-to-many match.
- **Why is DTW better?**
- If datapoints are shifted between each other (lag) & we want to look rather at their shapes or patterns.
- Two time series don't have to be of the same length

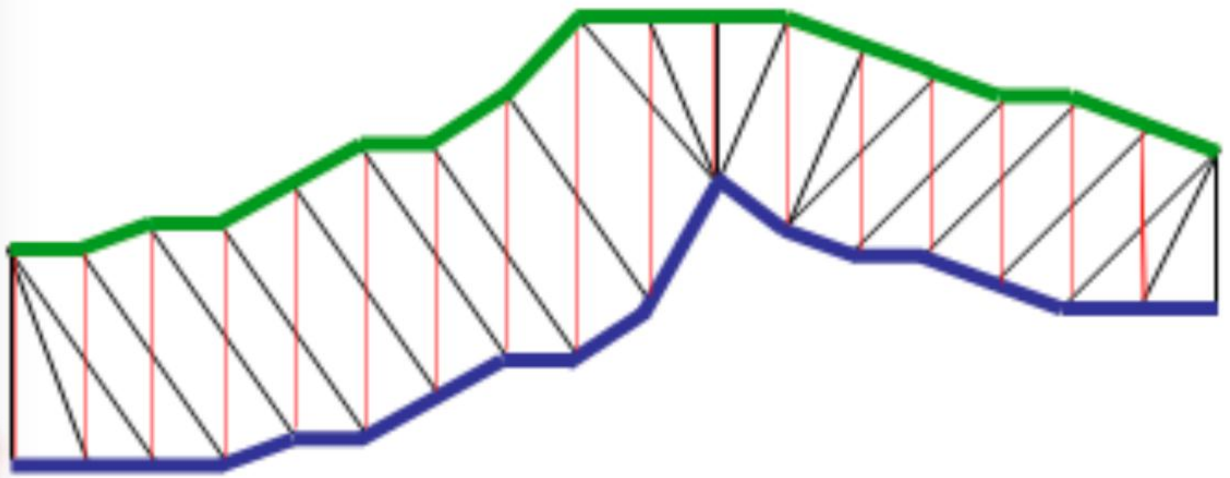


Fig. 1: Visual comparison of matched points based on DTW (black) and Euclidean (red) distance

Using dtw package and hclust() for clustering Excess Deaths and Policy Stringency Index Time Series.

- Both excess deaths and policy stringency time series data were first aggregated into monthly values, and time periods in all countries were restricted to be from January 2020 to December 2021.

- Normalizing data with **BBmisc::normalize()** using the "standardize" method **(center & scale**).

- DTW distances were calculated via **dtw::dtwDist()** function respectively, which computes the dissimilarity matrix based on DTW distance between single-variate timeseries.

- **Hierarchical clustering** analyses were performed via **hclust()**. In this project, I used hierarchical clustering using **Ward's method** as proposed by Gabriel Dance, Tom Meagher, & Emily Hopkins, 2016).

    - **Ward's Minimum Variance Method**: "minimizes the total within-cluster variance. At each step of clustering, the pair of clusters with minimum between-cluster distance are merged" (according to Bradley Boehmke of the University of Cincinnati).

Clustering Mix Data Types: Gower Distance

- **Gower distance:** the average of partial dissimilarities across individuals depending on the type of variable being evaluated.

$$d(i,j) = \frac{1}{p}\sum_{i=1}^{p} d_{ij}^{(f)}$$

Gower distance's formula

- For a **numerical** feature f (e.g. population density), partial dissimilarity is the ratio between 1) absolute difference of two observations and 2) maximum range observed from all individuals.

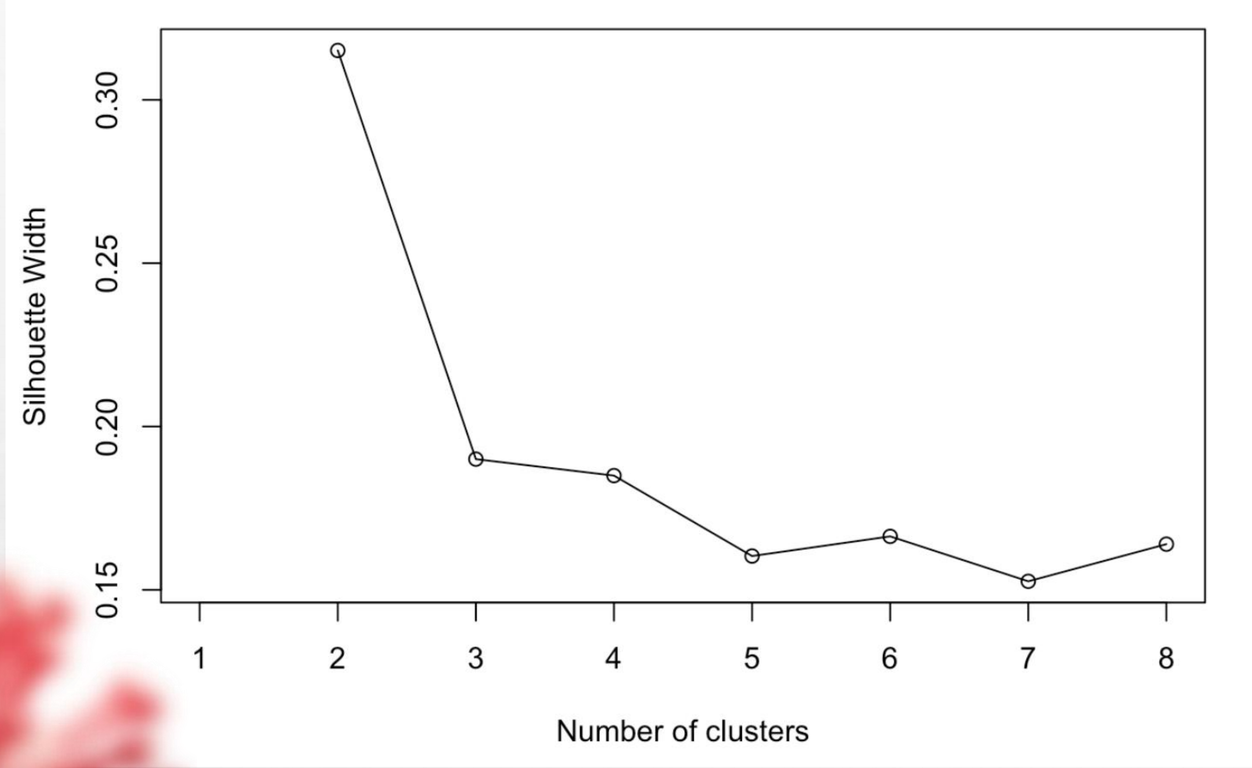
$$d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{R_f}$$

Partial dissimilarity computation for numerical features (R_f = maximal range observed)

- For a **qualitative** feature f (e.g. climate class, excess deaths class), partial dissimilarity equals 1 only if two observations have the same value.
    - cluster::daisy() function were used to compute gower distance, where numeric features were first automatically standardized.

Partitioning Around Medoids (PAM)

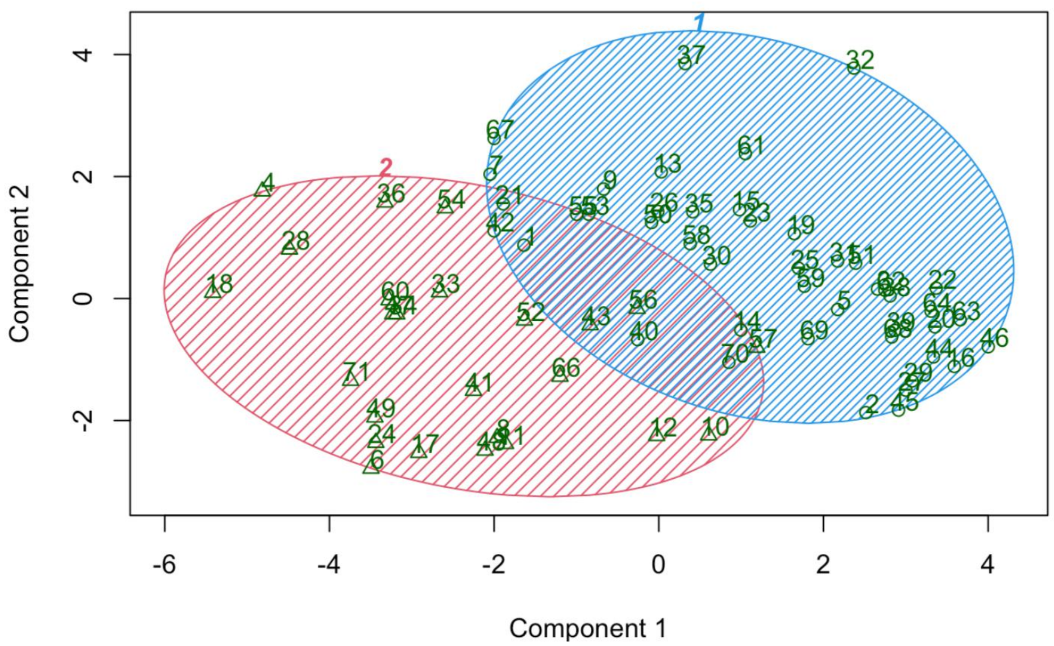
- **Gower distance** fits well with the **k-medoids** algorithm -- a classical partitioning technique of clustering that clusters data into k clusters (a more robust version of K-means)
- **Why is k-medoids better?**
- Compared to k-means clustering, it is more robust to noise and outliers.



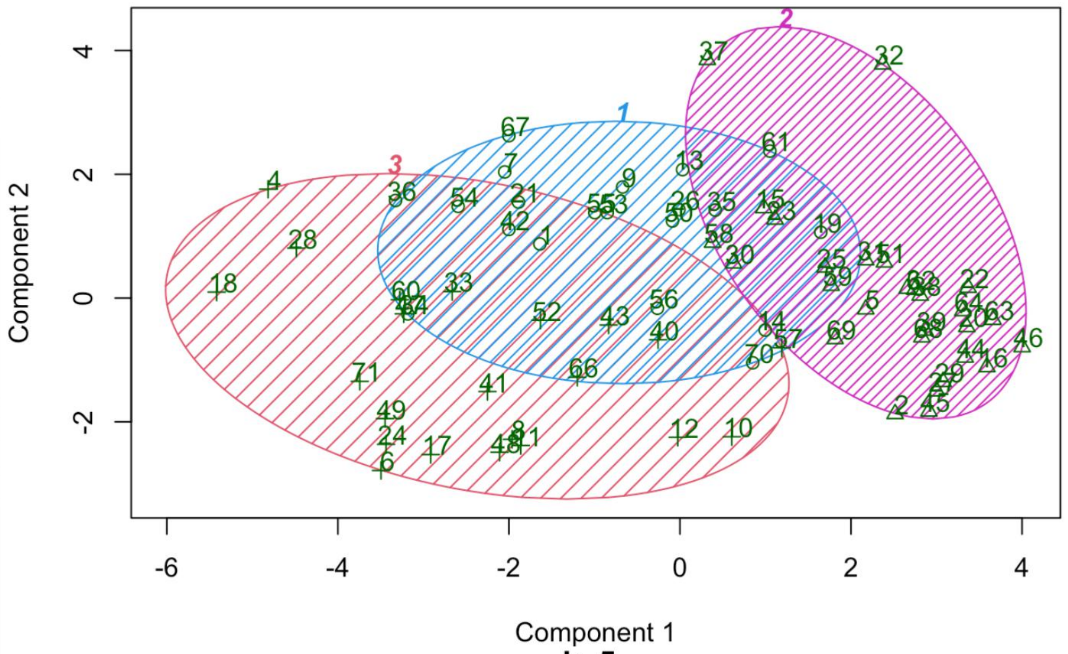
Silhouette coefficent was used to determine the optimal number of clusters

- contrasts the average distance to elements in the same cluster with the average distance to elements in other clusters (High silhouette values --> well-clustered, less outliers)

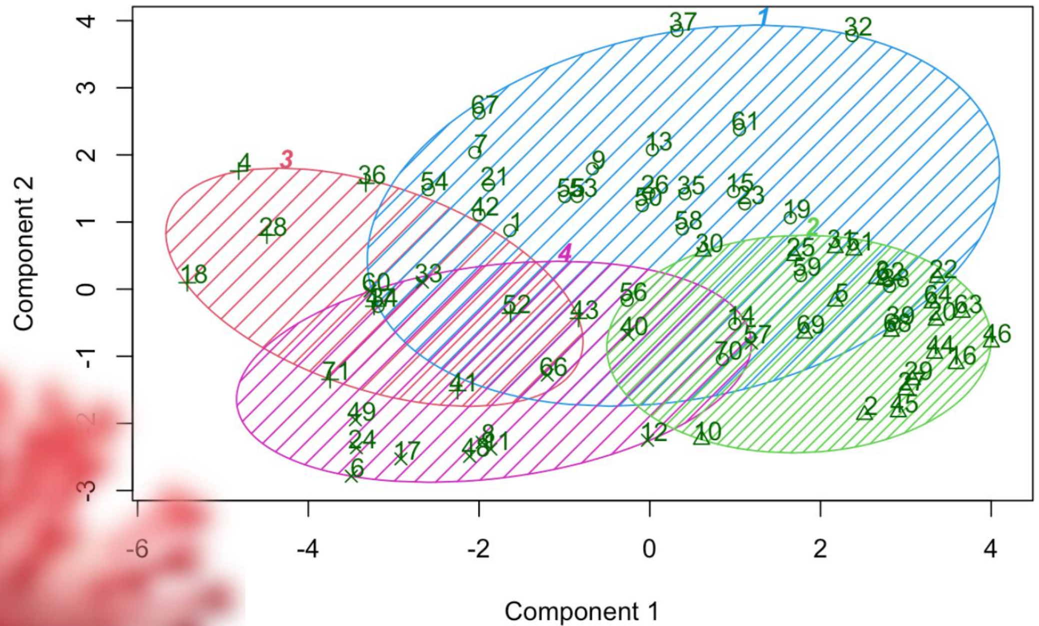
k=2


Component 2
Component 1

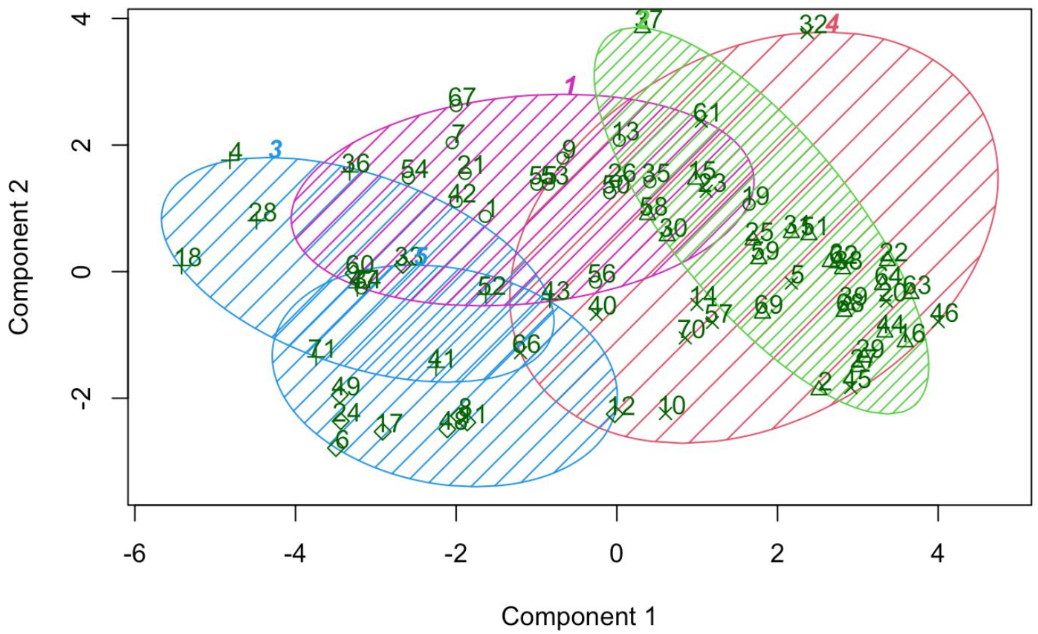
k=3


Component 2
Component 1

k=4


Component 2
Component 1

k=5


Component 2
Component 1

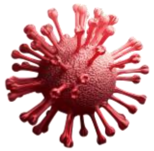
03

Results

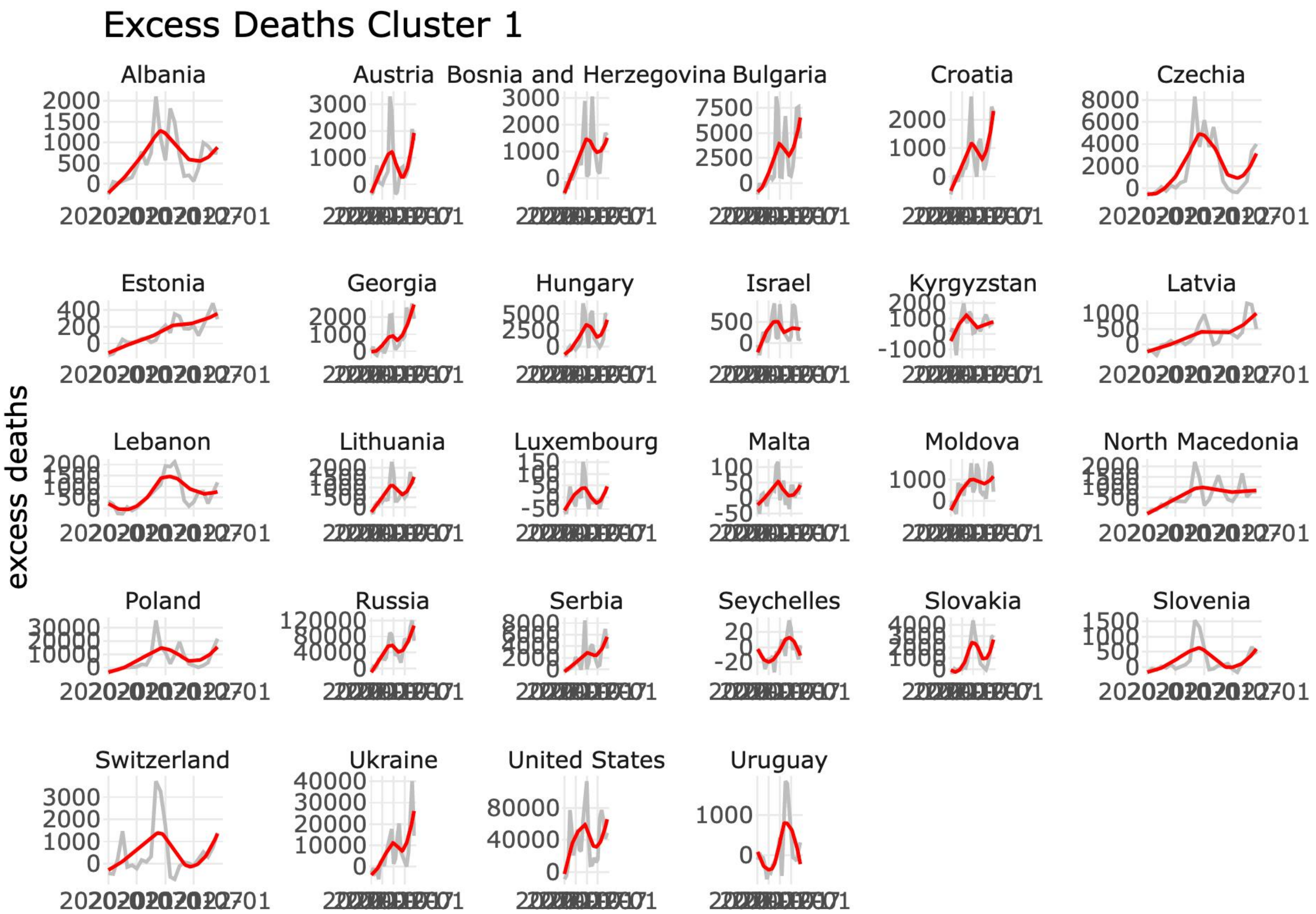
Part 1 Excess Deaths Class Profiles

Part 2 Policy Stringency Class Profiles

Part 3 Clustered Country Profiles

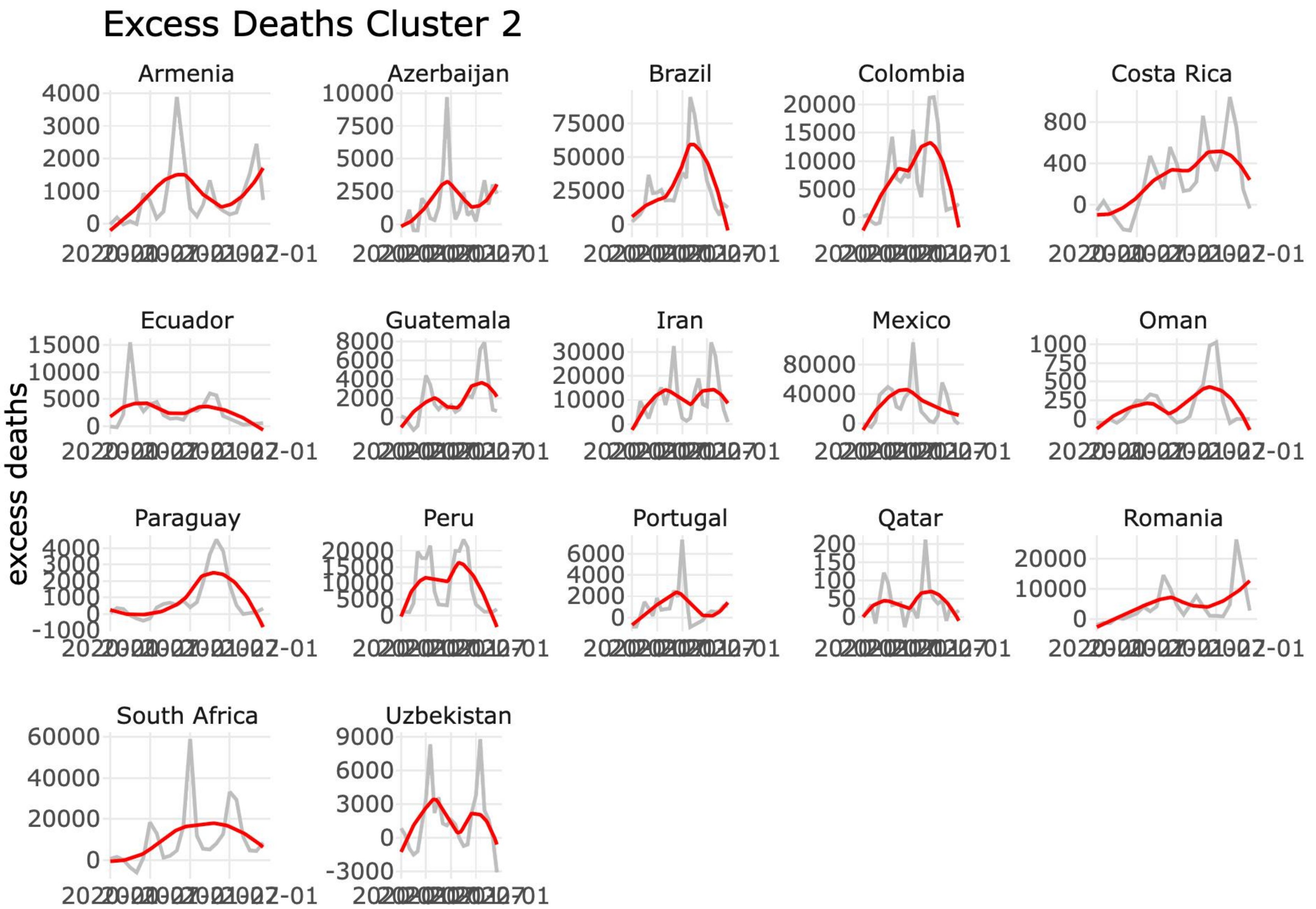


Excess Deaths Cluster 1

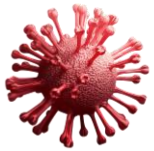


- **28 countries in Cluster 1**
- **Most recent trend of excess deaths seems to be slightly upwards.**
- **Most severe outbreaks have occurred by the end of 2020.**

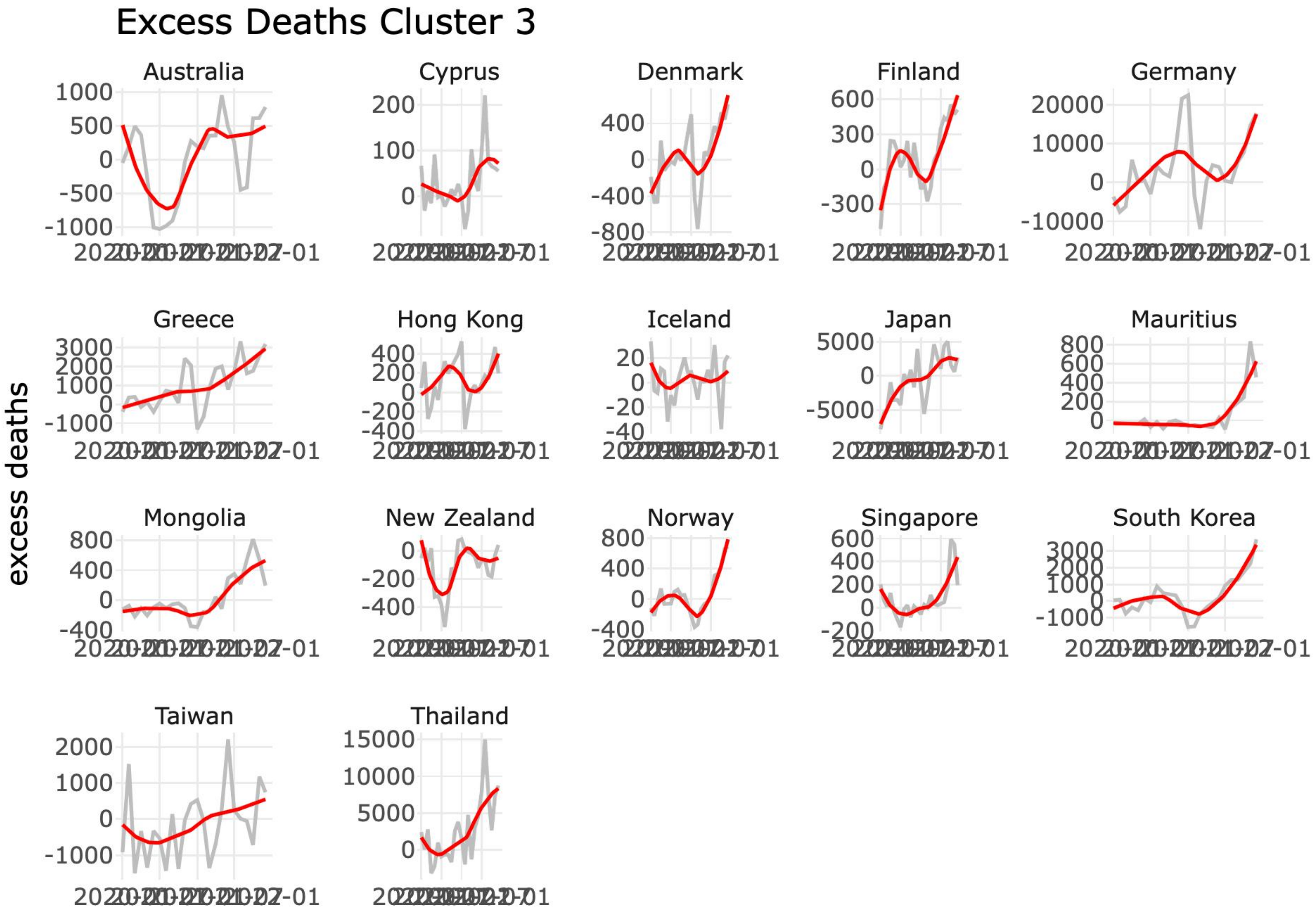
Excess Deaths Cluster 2



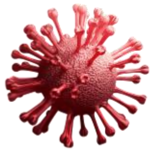
- 17 countries in Cluster 2
- Most recent trend of excess deaths seems to be downwards.
- Most severe outbreaks do not seem to have a common pattern across countries within cluster 2.



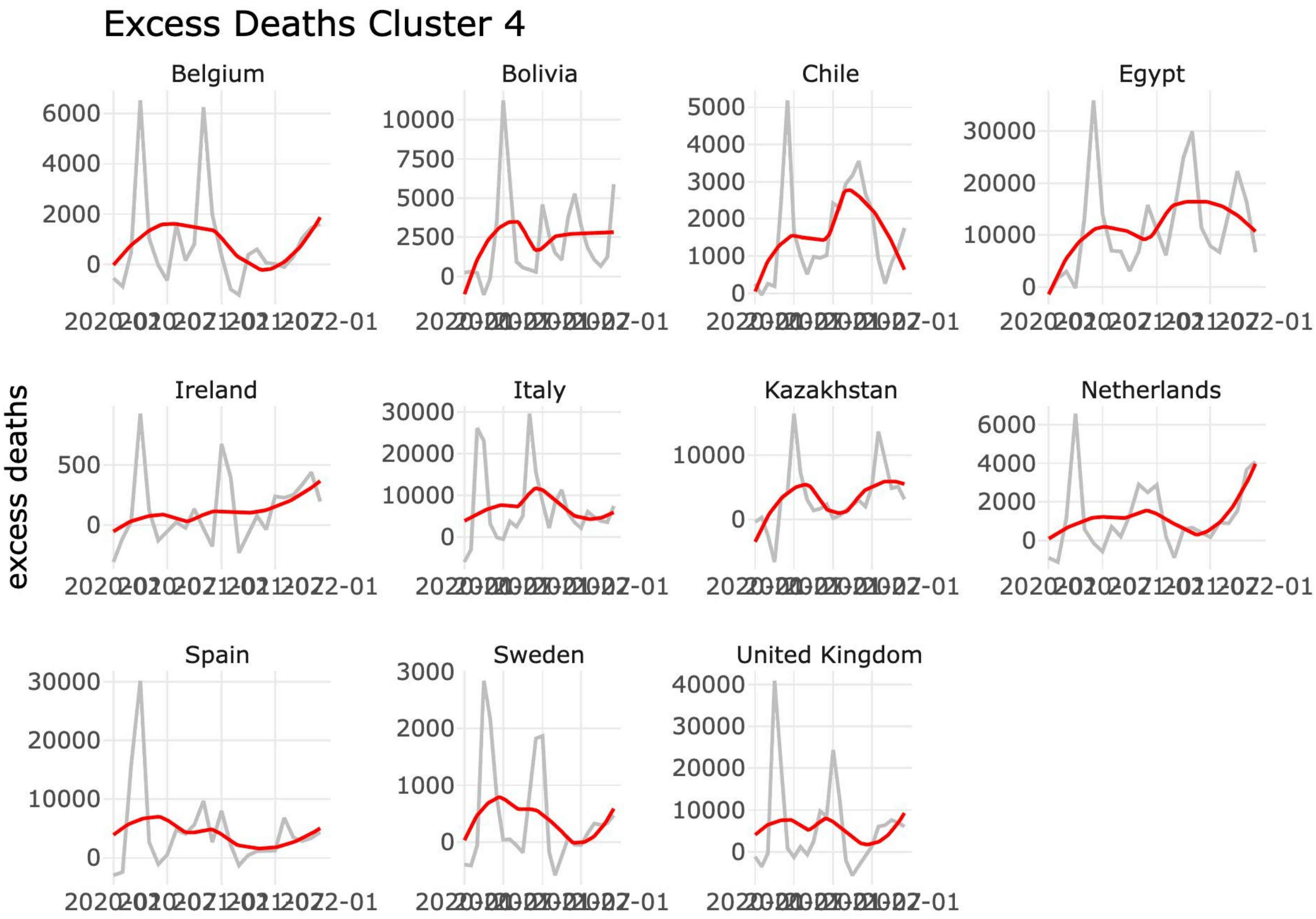
Excess Deaths Cluster 3



- 17 countries in Cluster 3
- Most recent excess deaths seem to be increasing dramatically.
- Most severe outbreaks have occurred by the end of 2021.



Excess Deaths Cluster 4



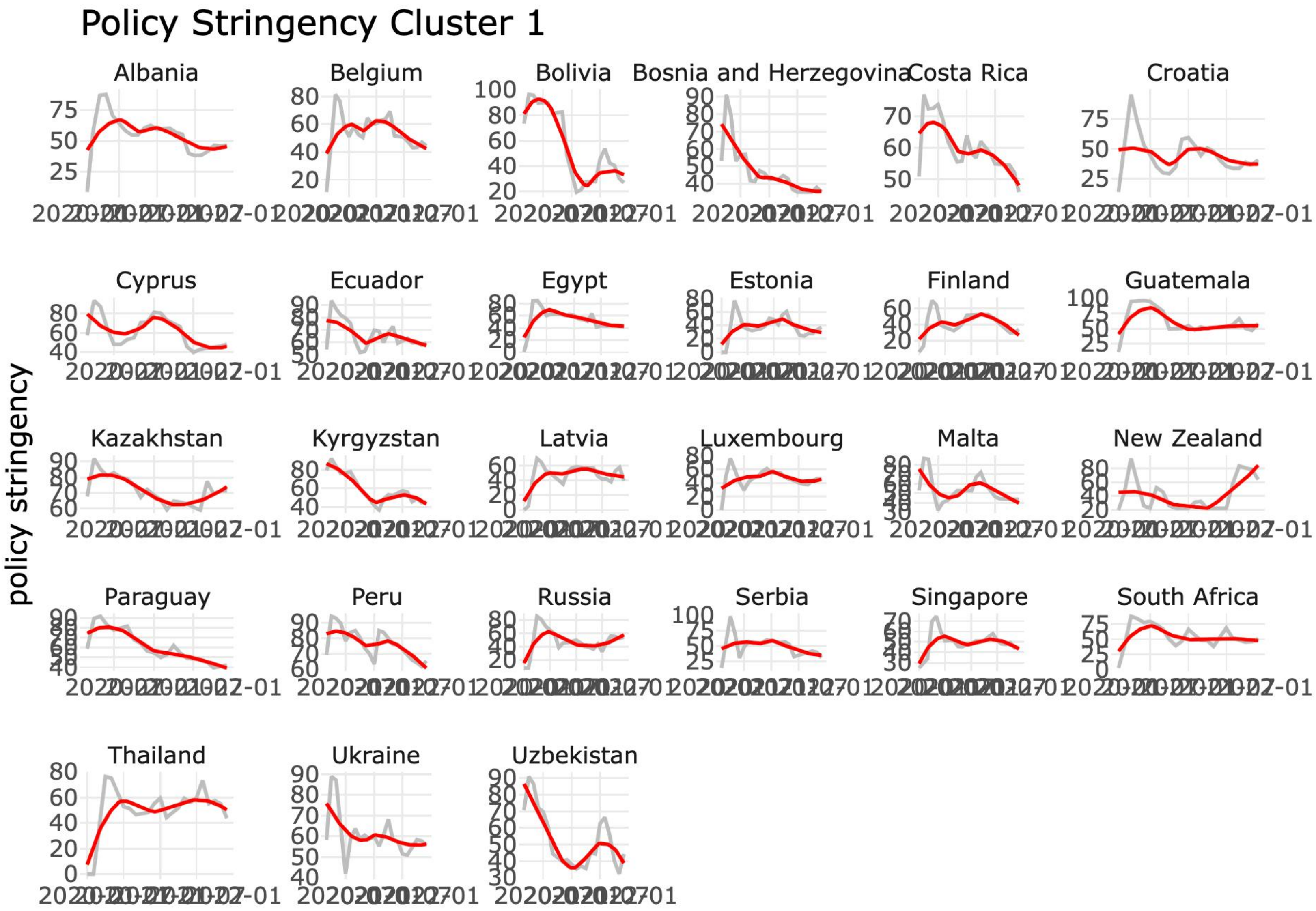
- **11 countries in Cluster 4**
- **Most recent excess deaths seem to be going downwards in general with some slight cyclic increases in certain countries.**
- **Most severe outbreaks have occurred at the beginning of pandemic (March 2020).**

Policy Stringency Cluster 1

- 27 countries in Cluster 1

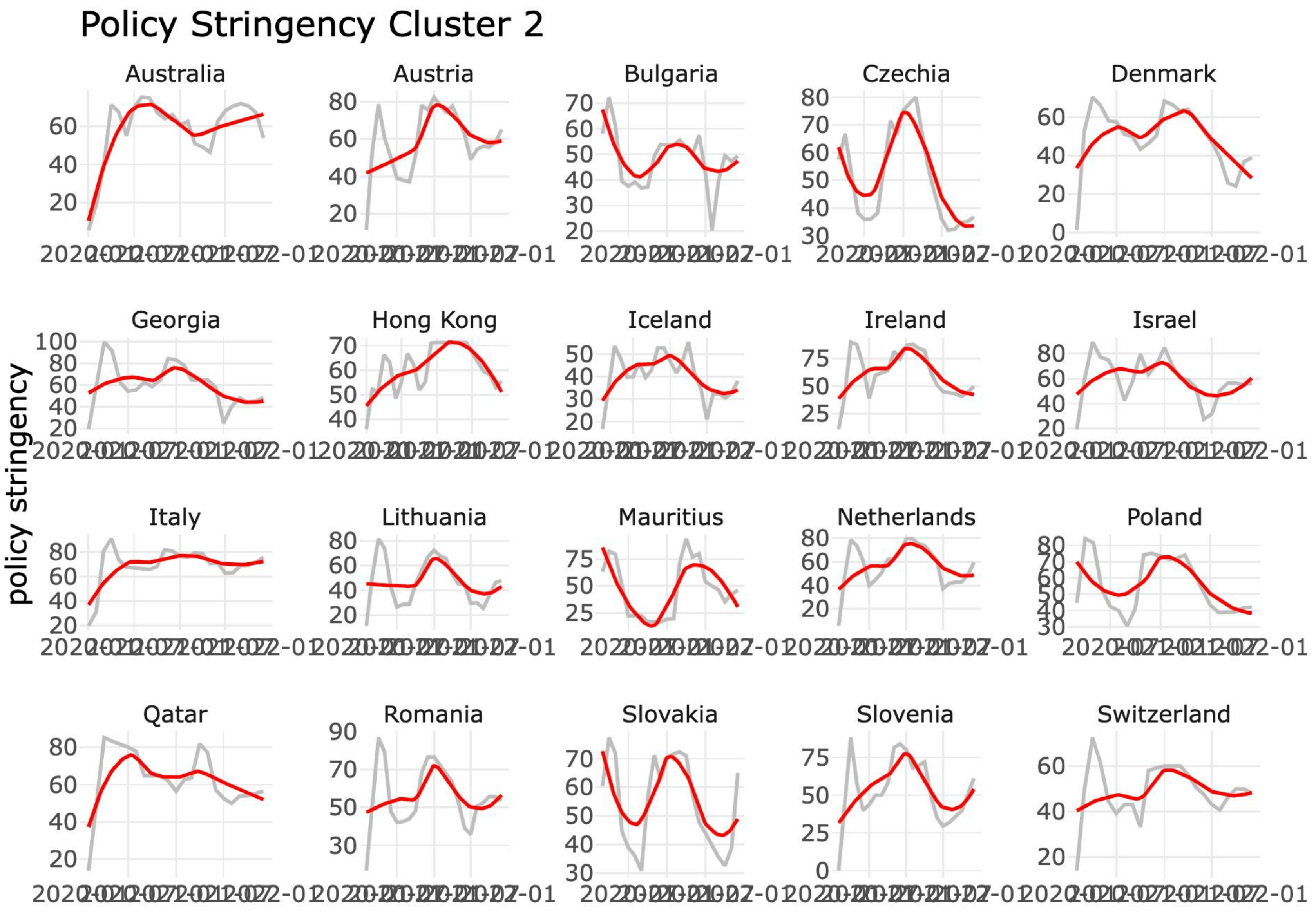
- Policy stringency index was the highest at the beginning of pandemic (March 2020) and then went down dramatically.

- A significant downwards trend in most recent time periods available.



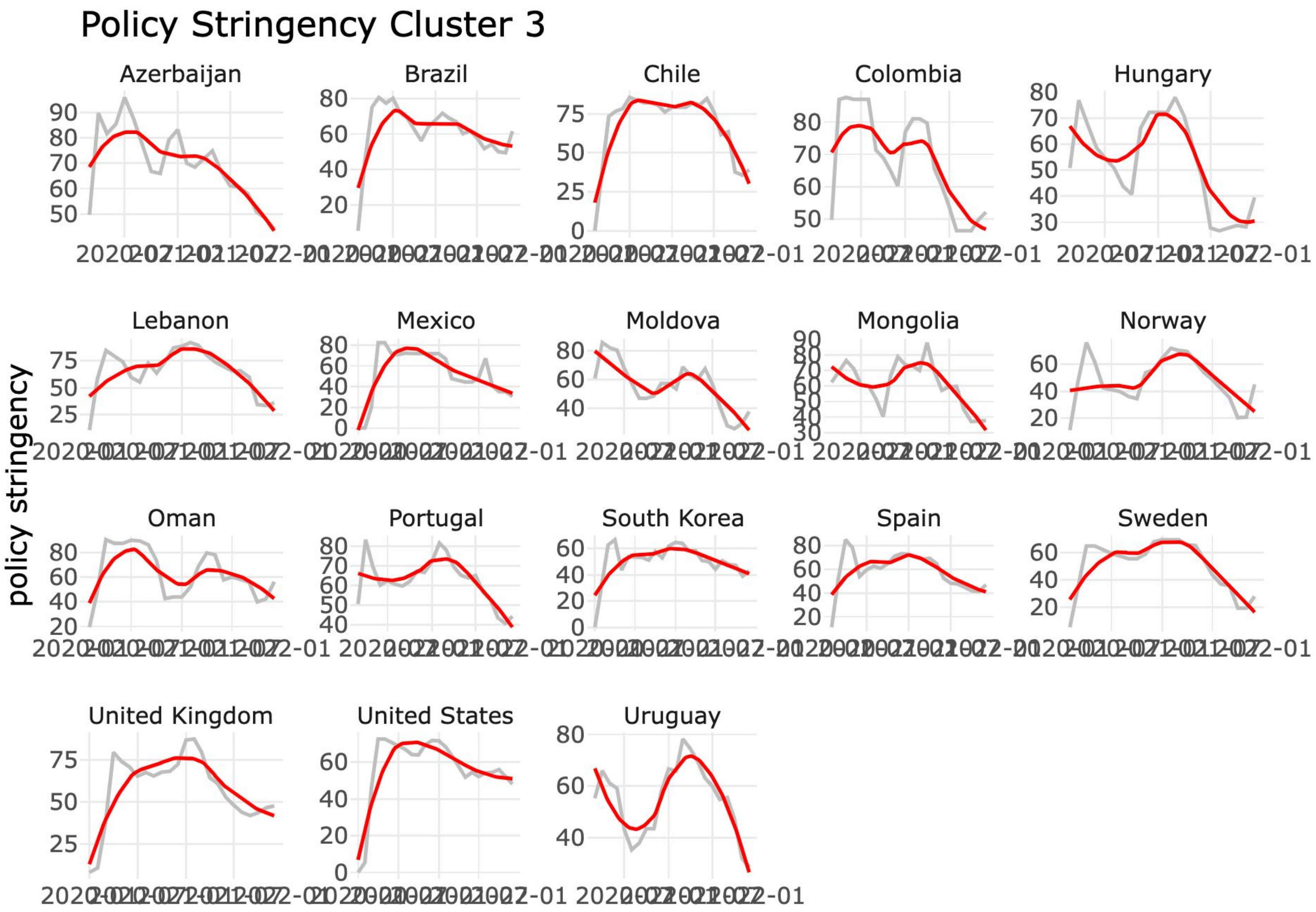
Policy Stringency Cluster 2

- 20 countries in Cluster 2
- The strictest policy responses were mainly on March 2020, but the level of stringency has been relatively stable with a slightly decreasing trend.



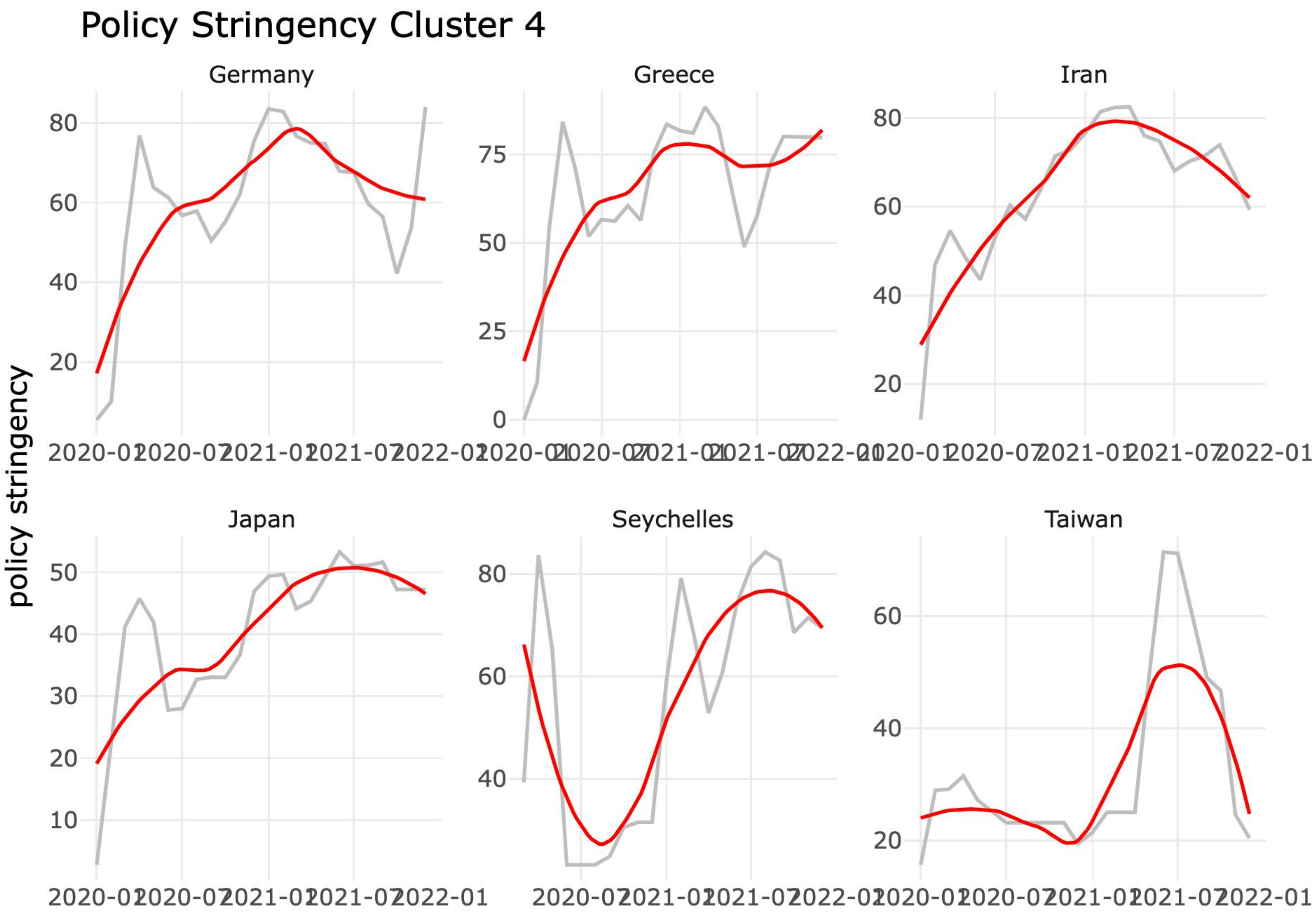
Policy Stringency Cluster 3

- 18 countries in Cluster 3
- Similar to Cluster 2, policy stringency level has been relatively stable after the highest value at the beginning of the pandemic, but with a more **significant downwards** trend.



Policy Stringency Cluster 4

- 6 countries in Cluster 4
- The highest policy response level seems to appear in the year of 2021.
- A **significant increasing** trend in 2020–2021, followed by a relatively **smoother decreasing** trend by the end of 2021.

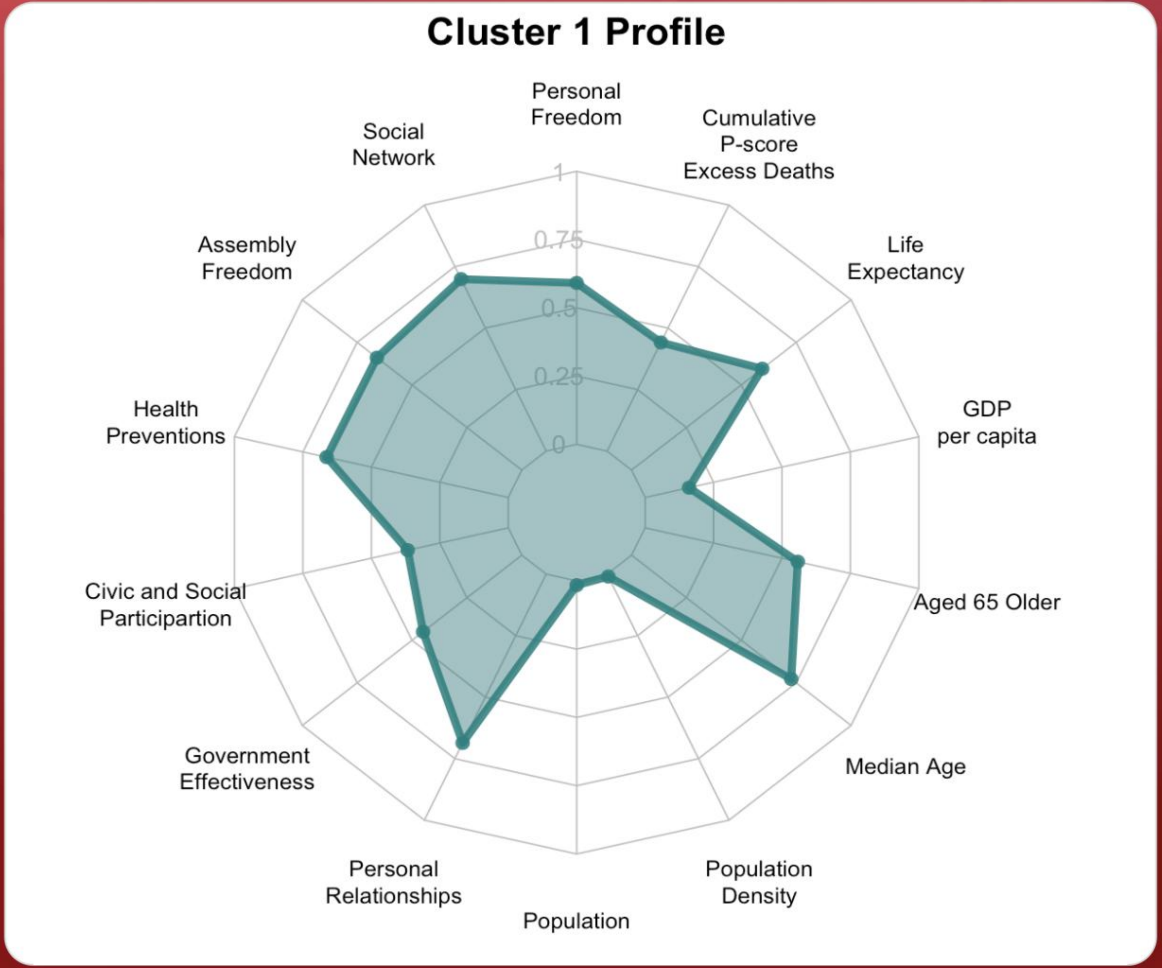


20 Countries in Cluster 1

e.g. South Korea, Albania, Russia

- **Moderate** personal freedom & freedom of assembly, social Interactions with individuals and society, government effectiveness
- **Significantly downward** trend in policy responses to COVID

Cluster 1	
Cumulative P–score of Excess Deaths	16.36%
Median Age	41.30
Aged 65 Older	15.66%
GDP per capita	21,611
Population Density	78.84
Life Expectancy	76.47
Continent	70% in Europe (n=14)
Climate Class	80% Temperate (n=16)
Stringency Class	50% Class 1 (n=10)
Excess Deaths Class	85% Class 1 (n=17)

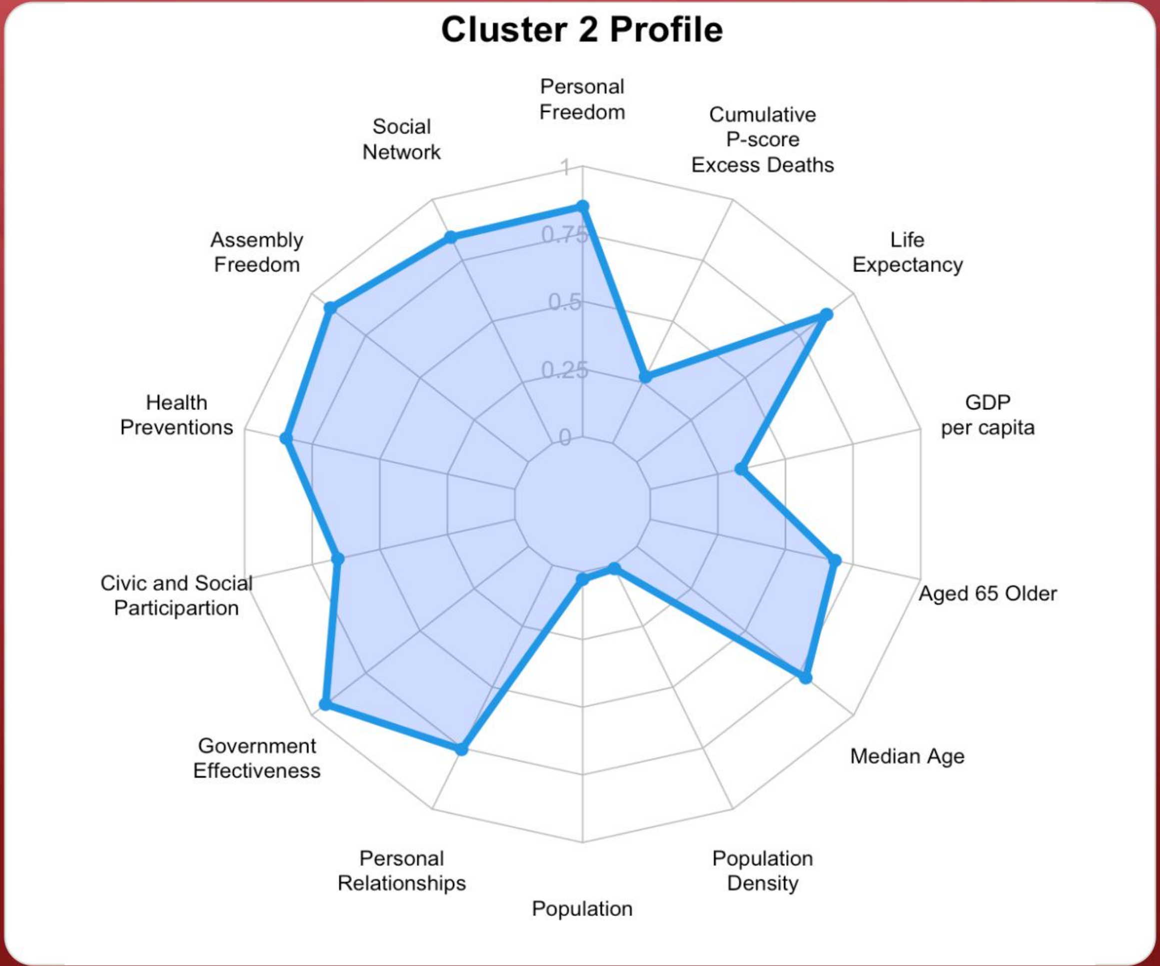


28 Countries in Cluster 2

e.g. US, UK, Germany

- High personal freedom & freedom of assembly, Social Interactions with individuals and society, government effectiveness
- Stable and slightly decreasing trend in policy responses to COVID

Cluster 2	
Cumulative P–score of Excess Deaths	7.47%
Median Age	42.60
Aged 65 Older	18.89%
GDP per capita	41,622
Population Density	112.75
Life Expectancy	82.27
Continent	79% in Europe (n=22)
Climate Class	79% Temperate (n=22)
Stringency Class	50% Class 2 (n=14)
Excess Deaths Class	71% Class 1 & 3 (n=10,10)

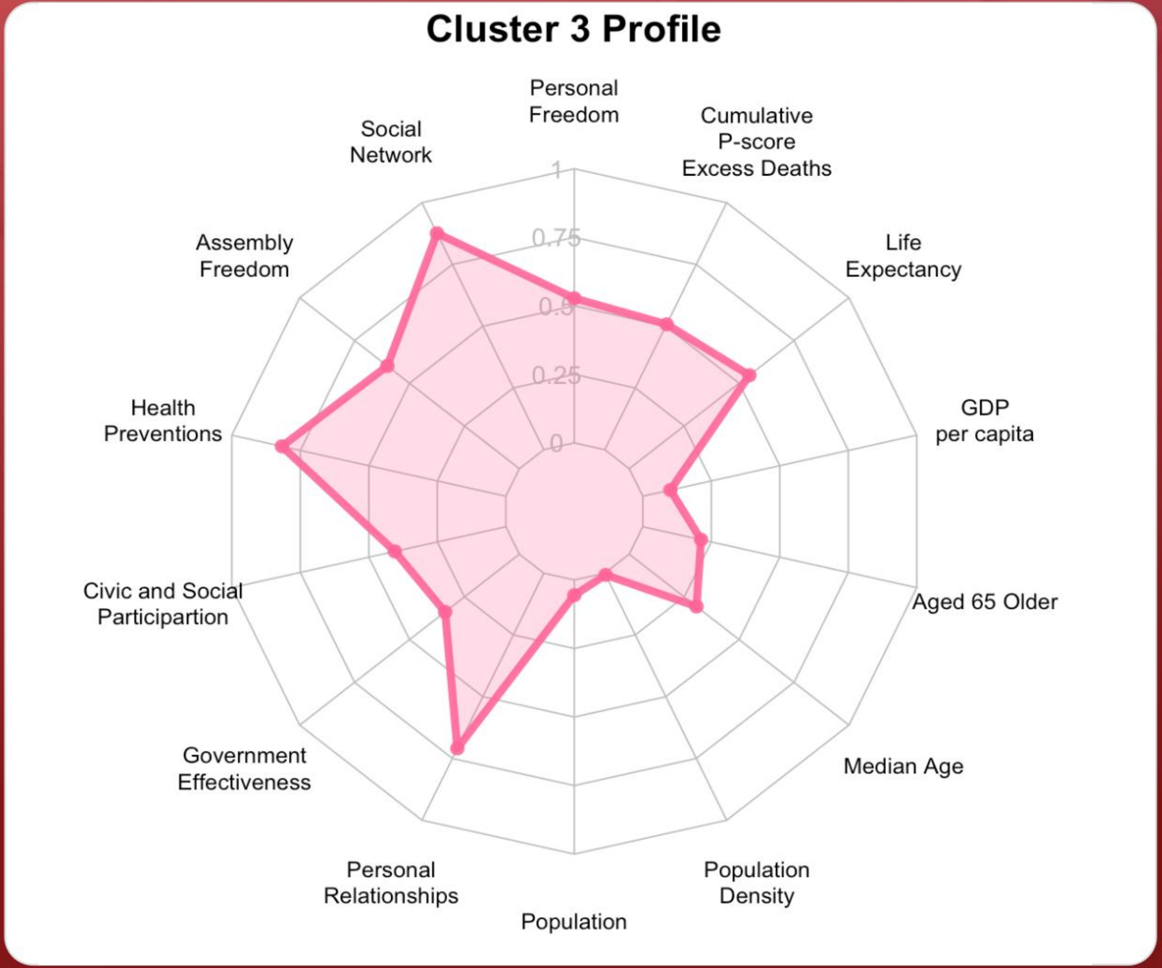


22 Countries in Cluster 3

e.g. Brazil, Mexico, Bolivia

- Low personal freedom & freedom of assembly, high social Interactions with individuals and society, low government effectiveness
- Significantly downward trend in policy responses to COVID

Cluster 3	
Cumulative P–score of Excess Deaths	19.83%
Median Age	30.65
Aged 65 Older	6.78%
GDP per capita	14,814
Population Density	58.14
Life Expectancy	75.47
Continent	73% in Asia / South America (n=9, 7 respectively)
Climate Class	86% in Tropical / Dry (n=10, 9 respectively)
Stringency Class	55% Class 1 (n=12) 32% Class 3 (n=7)
Excess Deaths Class	85% Class 2 (n=14)



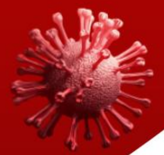
Implications & Limitations

- This project primarily focused on 1) producing **estimates of COVID-19 excess deaths** using a hybrid of **ARIMA/GARCH** models; 2) creating **clustered cross-country profiles** using indicators within social, economic, cultural, and environmental dimensions as covariates.

- Clustering of country profiles was only for **descriptive purpose,** but we were still able to see some characteristics of the three clusters -- low, moderate, and high risk:
  - for example, the **low risk** cluster (with lowest cumulative p-score of excess deaths) was categorized with more **liberal and assertive cultures**, temperate climates, highest GDP per capita, and a **stable, slightly downward trend in policy stringency level**, regardless of its highest share of older population and highest population density.

- Data sample was **small and not representative** enough (51% in Europe and only 6% in Africa or North America).

- Clustering of time series objects (excess deaths and policy stringency level) can be improved with better algorithm such as **model-based clustering**.
  - Model-based clustering uses a **probability-based approach** (uses an **Expectation-Maximization (EM)** algorithm to find the most likely model components and the number of clusters).
  - It assumes that the data is generated by an **underlying probability distribution** and tries to recover the distribution from the data.



THANK YOU FOR WATCHING

Presenter：Jiner Zheng