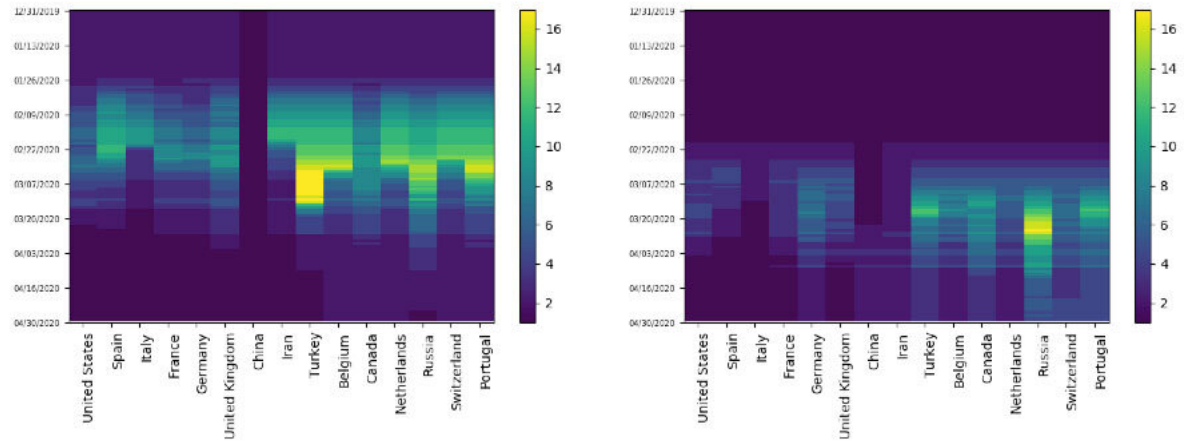
# Countries group into clusters as COVID-19 outbreak spreads

June 30 2020




LEFT: Heat maps track the changing cluster membership of the fifteen most severely impacted countries with respect to numbers of COVID-19 cases. Cluster membership depicts COVID-19 severity relative to the rest of the world. Clusters are ordered with 1 being the worst impacted at any time. Darker and lighter colors correspond to smaller and greater numbered cluster labels and represent worse and less affected clusters, respectively. RIGHT: Same as image on the left but for deaths. Credit: Nick James and Max Menzies

Mathematicians based in Australia and China have developed a method to analyze the large amount of data accumulated during the COVID-19 pandemic. The technique, described in the journal *Chaos*, can identify anomalous countries—those that are more successful than expected at

responding to the pandemic and those that are particularly unsuccessful.

The data comes from Our World in Data, a project of the Global Change Data Lab, a registered charity in England and Wales. This organization collected information from the European Centre for Disease Prevention and Control for cumulative daily case counts and deaths for 208 countries over a period of 122 days from Dec. 31, 2019, to April 30, 2020. The investigators analyzed the data with a variation of a statistical technique known as a cluster analysis.

In this approach, data points are grouped according to similarity. The countries form clusters as individual outbreaks become more similar.

For all of January, the investigators found only two clusters: China in one cluster, and all the other 207 countries in the other. As the virus spread, additional countries jumped into the China cluster. Italy was the first to join, followed by the U.S., Spain, France, Germany, Iran and the U.K.

By mid-March, case counts for countries around the world grouped into 16 clusters. By April, a similar grouping was seen in death counts. In mid-March, China moved out of the worst death cluster, while the U.S., Spain, Italy, France and the U.K. moved into it.

The investigators found a notable break in the cluster structure for cases between March 1 and March 2. This date is significant, because numerous countries reported their first COVID-19 cases at that time, mostly coming from Iran and Italy.

Another break in the cluster structure occurs between March 18 and March 19 for deaths, a 17-day difference from that of cases. This offset suggests a 17-day lag for deaths behind cases and agrees with medical data.

Once the investigators identified the 17-day offset between cases and deaths, they were able to compare countries' case and death numbers at the same point in time. This revealed countries with anomalous results.

"Anomalies may signify either disproportionately high or low number of deaths relative to the number of cases," said co-author Nick James.

Iran and Italy both had anomalously high death rates early in the pandemic, while Singapore was anomalously low, as were South Korea, Qatar and Australia.

"We also noticed a sort of critical mass effect in the progression of cases to deaths," said co-author Max Menzies. "Spain's death count as of March 28 was over twice that of its case count just 16 days earlier. This is an astonishing explosion of COVID-19. It also applies to the U.S. Its dramatic elevation in death count hit after the case count reached a critical mass in early March."

**More information:** "Cluster-based dual evolution for multivariate time series: Analyzing COVID-19," *Chaos* (2020). aip.scitation.org/doi/10.1063/5.0013156

Provided by American Institute of Physics