Visualizing early Social, Economic, and Financial impacts of the COVID-19 pandemic

Team 83: Kunting Chua, Zhi Kang Lin, Vincent Quek, Tian Yong Woon
Georgia Institute of Technology

1 INTRODUCTION

The COVID-19 pandemic, which originated from China in late 2019, has resulted in widespread repercussions across the globe. Strong measures meant to contain the spread of the virus have been enacted at the expense of economic activity, resulting in a disruption of trade and a rise in unemployment worldwide [2].

Currently, available visualization tools focus on health metrics such as the number of cases, hospitalizations, and deaths. In this project, we propose a web-based tool for visualizing and analyzing the social, economic, and financial market impacts of COVID-19 worldwide. Our project aims to provide interactive visualizations, allowing users to compare countries and draw cross-border insights.

2 LITERATURE SURVEY

Macroeconomic indicators are commonly used to analyze and visualize the impacts of the COVID-19 pandemic. For example, the Gross Domestic Product (GDP) has been extensively studied and found to be negatively impacted following the onset of the pandemic [7, 11]. [5] analyzed export data and found both negative effects on supply chains and consequently trade. [3] studied domestic and foreigner flows into and out of countries and found additional impacts on capital flows from the pandemic. Although the economic situations seem to be captured by these metrics, we do note however that their utilities (e.g. GDP) can be quite limited, since some countries showed economic growth despite having significant health impacts [e.g. 7].

Another important aspect is the global financial market, which has been adversely affected by the pandemic. For example, Canada and US stock market returns were observed to correlate negatively with their number of reported cases [15], while currencies weakened against the US dollar early in the pandemic [14]. The financial market also affected policy decisions, as found by [10], who noted that countries with a greater stock market decline also had bigger COVID-19 stimulus packages.

We note that many economic and financial analyses are narrow and niche, be it on a country or region level: they do not adequately demonstrate the downstream reverberations on societies and the people living in them. Repercussions of the pandemic can manifest directly in the social welfare of individuals as well. For instance, mobility has been found to be affected: mobility, as measured with smartphone data, was significantly reduced due to voluntary social distancing by individuals as well as involuntary social distancing imposed by governments [1, 11]. However, mobility data may not be entirely representative of a country, since they are collected from smartphone owners and thus neglect rural areas.

Other important social indicators include the birth rates [12] and the mental health of populations [6]. Both factors have been found to be adversely affected by COVID-19 and can have potential longer-term implications on countries. Mental health can be assessed as the prevalence of depressive and anxiety disorders in population survey data. We note that it can be hard to attribute some of the observed effects to the pandemic since these factors are interwoven and cannot be looked at in isolation. For instance, mental health and access to birth control can also influence future birth rates.

There have also been attempts to use Machine Learning (ML) to evaluate the effectiveness of individual COVID-19 measures [4], as well as to forecast economic impacts [8]. We note that these studies are not directly applicable to our project, since our goal is not to evaluate specific measures nor to provide forecasts. However, they do highlight the potential benefits of using ML algorithms to draw insights from existing data.

3 METHOD

COVID-19 has affected a large number of non-health sectors: its full impact cannot be summarized by health statistics alone and must be substantiated with social, economic, and financial indicators. In this project, we improve on past visualizations by focusing on diverse factors demonstrating the pandemic impacts on countries around the globe. Our first innovation involves the integration of multiple indicators, allowing users to visualize all of them in a single tool. Secondly, for a more quantitative evaluation, we integrate a regression tool to help users understand the most impact factors globally. For our final innovation, a clustering analysis can be performed to help identify groups of countries that have experienced similar impacts. We further elaborate upon these aspects in the following sections.

This multi-layered visualization tool is targeted both at the public and governmental institutions: users will be able to gain a deeper understanding of the wider impact of the pandemic while governments and institutions can learn from other countries and avoid inefficient measures. Our tool is designed as a dynamic web application with interactive visualizations implemented using the JavaScript D3.js library. Using the Flask framework, we create Application Programming Interfaces (APIs) to retrieve and process data in Python.

We highlight that, based on our research, there are no free and readily available online tools that allow for the multi-faceted analysis of global COVID-19 data against social, economic, and financial variables that our visualization tool provides. Our interactive aspects are envisioned to enhance usability compared to static figures found in current research papers and tools. Different user types also benefit as the tool is highly configurable to visualize different aspects of the multivariate data.

3.1 Indicators and Datasets

We obtain our datasets from four sources, with a total of 14 indicators for up to 220 countries. The different economic, social, and financial indicators are described in Table 1, along with their sources. In addition, we also consider the traditional health indicators using the number of new COVID cases and deaths. We note that the exchange rates are benchmarked against the US, i.e. the US exchange rate remains at a constant value of one. Hence, it fails to correctly measure the financial impact on the US. However, we retain this indicator due to its general applicability to other countries.

As Table 1 shows, the datasets have different time frequencies. In particular, the Google mobility dataset consists of 9 million rows in total, and aggregates daily movement trends by region worldwide. In contrast, the economic and financial data are only available at monthly intervals. In order to compare all the available variables simultaneously, we process data with higher temporal and spatial resolutions and down-sample them to a common monthly baseline suitable for visualization and analysis. Firstly, we handle the daily quantities as follows: (i) for the Google mobility dataset, monthly averages are obtained by taking the mean of all values within a month, and (ii) for the health indicators (cases and deaths), daily values are summed to provide a monthly value. Secondly, we aggregate regional data into country-level data for the mobility data.

3.2 Global Visualization

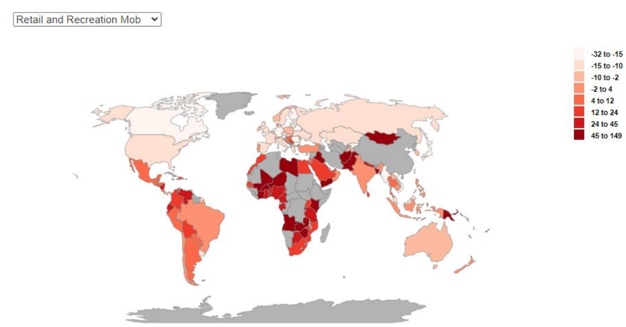

Figure 1: Choropleth showing Retail and Recreation Mobility for Feb 2022

.

The first aspect of our application serves to provide a global outlook of the indicators at different time points. On this page, two choropleths are displayed: the first showing the number of COVID-19 cases or deaths normalized by population, and the second showing a user-selected metric (Table 1). This enables users to get a quick overview of both the traditional health aspects of the pandemic and the less traditional socio-economic metrics.

As shown in Fig. 1, each variable is divided into eight quantiles, with colors for countries assigned accordingly. The drop-down bar for each choropleth allows users to toggle between different variables of interest. A single slider bar allows the user to adjust the date for both choropleths jointly and visualize changes over time globally.

3.3 Linear Regression Analysis

Here, we explore how linear relationships between variable pairs (e.g, COVID-19 cases vs. mobility) evolve across all available countries, over time.

	Indicator	Description
Economic	1. Retail sales[1]	Consumer demand for finished goods (percentage change; monthly)
	2. Industrial production[1]	Output of the industrial sector (percentage change; monthly)
	3. Inflation[1]	Consumer Price Index (percentage change; monthly)
	4. Unemployment rate[1]	Percentage of jobless out of the labor force actively seeking work (monthly)
Financial	5. Stock price[1]	Main equity index return (percentage change; monthly)
	6. Exchange rate[1]	Against US dollar (percentage change; monthly)
Social	(7-12) Google mobility[2]	Movement trends in the following 6 categories: (percentage change to baseline; daily)
		7. Grocery and pharmacy, 8. Parks, 9. Transit Stations
		10. Retail and Recreation, 11. Residential, 12. Workplaces
Health	13. New COVID-19 cases[3]	Number of confirmed cases per 100000 (daily)
	14. New COVID-19 deaths[3]	Number of deaths per 100000 (daily)

Table 1: Indicators and datasets used in this project, separated by categories. The unit of measurement and the time frequencies are given in parenthesis. All daily quantities are aggregated into monthly intervals to obtain a common baseline. Sources: [1]World Bank DataBank, [2]Google Mobility Trends, [3]World Health Organization.

Users select two variables of interest, on which simple linear regression will be performed, i.e. $Y = aX + b$. The regression is computed in Python across all available time periods and passed to the browser to be displayed

The regression results are visualized with the following four main elements:

(1) the two-dimensional scatter plot,
(2) the fitted linear least-squares line,
(3) the regression equation, and
(4) the R^2 of the fit, which quantifies the proportion of the variation in the dependent variable explained by the independent variable.

Within the scatter plot, a tooltip displaying numerical variable values appears when hovering over each point, as shown in Fig. 2. Users can also toggle through time points using the interactive slider mechanism provided, with the selection interface shown in Fig. 3. This allows users to observe changes in variable distributions, pair-wise linear relationships, and the R^2 of fitted linear regression models over time.

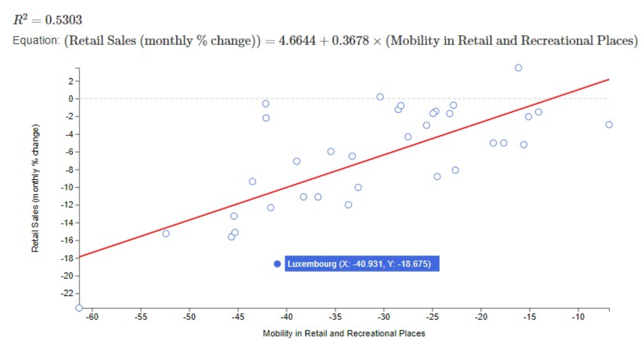

Figure 2: Example showing the main result elements of the linear regression interface.

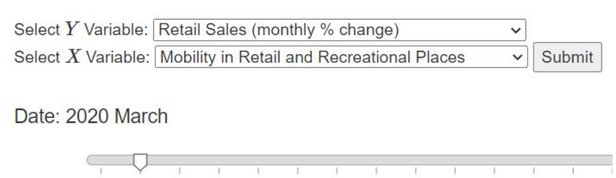

Figure 3: Interactive dropdown and slider for date selection for linear regression analysis.

3.4 Clustering Analysis

We further explore the similarities across countries in terms of the various impacts using a combination of t-SNE visualization and k-means clustering. Since both these algorithms are based on pair-wise distances between points, each feature is normalized between 0 and 1 to obtain comparable scales.

To display the high-dimensional data in lower dimensions, we make use of the statistical technique t-SNE (t-distributed Stochastic Neighbor Embedding) [13]. t-SNE non-linearly projects high-dimensional data onto a smaller number of dimensions, keeping nearby points close to one another. Hence, similar points (countries) will form clusters in the resulting embedding. This is implemented in Javascript using the tSNEJS library.

The t-SNE algorithm involves three main parameters: i) learning rate, ii) perplexity and iii) the number of iterations. For simplicity, we fix the maximum number of iterations to 5000, with buttons to pause the algorithm before convergence. We supply default values of 5 and 10 for the learning rate and the perplexity, respectively, which have been tested to provide good results for the current data. Experienced users are allowed to change both parameters using the interface.

A downside of t-SNE is that it is primarily a visualization tool and lacks quantitative information. To address this limitation, we rely on k-means clustering, which partitions the data into k clusters in an unsupervised manner. Since k-means achieves its objective by minimizing the variance within each cluster, points within a cluster will in general be considered more similar.

The l^2-norm is chosen as the dissimilarity function, which leaves the number of clusters k as the sole parameter. While k can be chosen using the elbow technique or the silhouette score, we leave this as user-defined for simplicity. Rather, we encourage users to choose a suitable value of k in conjunction with the t-SNE visualization.

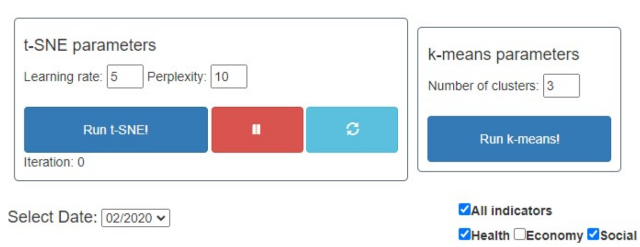

Figure 4: User input interface for clustering.

Fig. 4 shows the relevant input user interface. For clarity, the t-SNE and k-means inputs are separated. Users can select a specific date from the dropdown list. Furthermore, they are also able to select from the check-boxes different groups of indicators to be used for the analysis. Overall, we have designed the input for flexibility while keeping the complexity low.

An example of the visualized results is shown in Fig. 5 Points are colored according to their assigned k-means

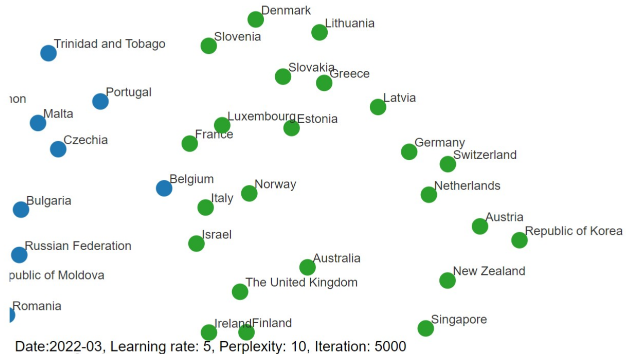

Figure 5: Example of clustering output.

clusters while the spatial arrangement reflects the t-SNE embedding. This combination of both methods allows users to easily identify trends across countries at a single time point.

4 EVALUATION

4.1 Observations

In this section, we describe some observations users can make using our tool, showing how each part is able to address pertinent questions about the pandemic. These questions include:

(1) From the choropleths, are there any patterns for the different indicators?
(2) Are there any pair of indicators with high correlation?
(3) Can we accurately group countries with similar impacts?

For the first question, we observed in Fig. 1 with our choropleth map that during February-March this year (2022) when the Omicron wave was circulating, many countries saw sharp declines in retail and recreational mobility, with the exception of African nations where mobility increased instead.

Apart from the well-known relationship between COVID-19 cases and deaths, our linear regression analysis tool has also picked up, for the early months of the pandemic, strong and positive correlations between countries' monthly percentage change in retail sales and their mobility readings for retail and recreational places (shown in Fig. 2). This is in line with strict lockdown measures enacted in many parts of the world during the early months of the pandemic, which adversely affected overall mobility readings and most brick and mortar business revenues.

Lastly, our clustering tool showed that Singapore, a country in Southeast Asia, is assigned in the same cluster as a number of European nations such as Germany and Denmark. This is shown in Fig. 5 and may come as a surprise to some but considering that Singapore is Southeast Asia's most developed country, it would share many similar characteristics with developed nations in Europe. Also within the same cluster are developed nations New Zealand and South Korea.

4.2 Scaling

In this section, we analyze how the three main algorithms - linear regression, t-SNE, and k-Means - scale with the number of points N and dimensions d. Although we focused on 14 indicators, other factors can also be included when data become available in the future. Instead of performing our analysis at a particular time, it is also possible to include time as a variable, e.g. by treating each time point as a separate data point (row). Such an analysis would increase N from the current $O(10^2)$ to $O(10^4)$.

For simplicity, we fix input parameters (for t-SNE and k-means) at their default values. We randomly generate 5 sets of synthetic data and measure the wall-clock time for each algorithm to complete. Data animation is disabled to isolate the algorithm speeds. For scaling with N, we keep $d = 100$. For scaling with d, we keep $N = 250$.

Fig. 6 shows the mean time taken t as a function of N (top) and d (bottom), with error bars denoting the standard deviations. Note that t is measured in seconds (s) for t-SNE, since it is much slower compared to both k-means and linear regression. Both k-means ($t \propto N^{0.76}$) and linear regression ($t \propto N^{0.43}$) scale sub-linearly with N, while t-SNE scales almost quadratically ($t \propto N^{1.83}$). This means that k-means and linear regression scale well to large N while t-SNE is restricted to relatively small N. We note that the t-SNE JavaScript library used is single-threaded and improvements could be obtained through parallel implementations.

One feature of t-SNE is the constant scaling with respect to d, which is reflected in the bottom panel of Fig. 6. This allows t-SNE to scale easily to high-dimensional data. k-means scales more poorly with d, thus other forms of dimension reduction such as Principal Components Analysis are often performed to retain

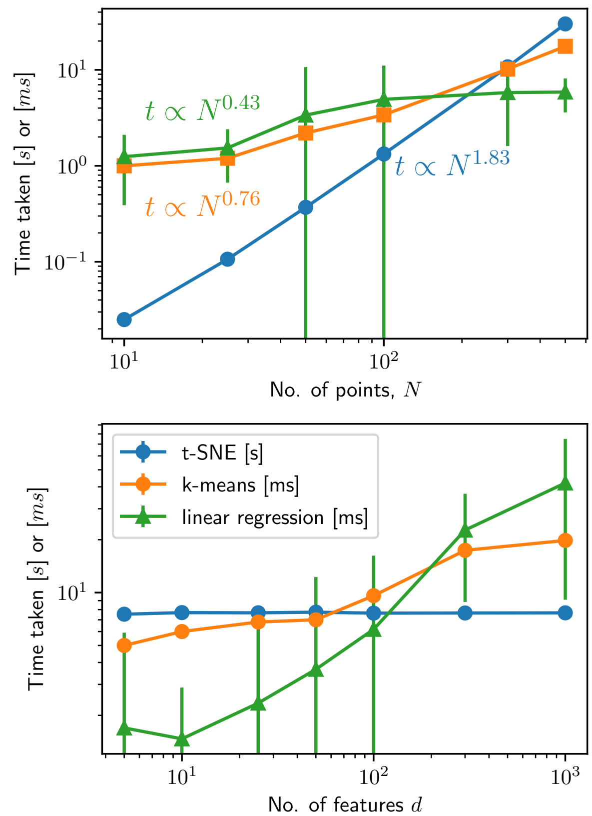

Figure 6: Scaling of algorithms with number of points (top) and data dimensions (bottom). Error bars denote standard deviation across the runs.

only the important features. Similarly, linear regression also scales more poorly with d relative to t-SNE.

Finally, we note that although processing of the raw data (reading, cleaning etc.) is slow due to the large data size, this only takes place once during initialization of the tool and does not affect subsequent usage. Hence, scaling tests for data processing have not been included.

4.3 User Survey

We further gauge the success of our tool in terms of:

(1) **Informativeness** - Were users better informed about the far-reaching effects of COVID-19 after using the tool?

(2) **Impact** - Were users prompted to reconsider their perspectives, habits, and possibly lifestyles after using the tool?

To quantify these aspects, we reached out to contacts from various walks of life (e.g., age, occupation, educational level, have you contracted COVID before?) to try and rate our visualization tool. Participants were asked to fill out a Google Forms survey with the questions:

(1) **Informativeness** - On a scale of 1 (no new knowledge gained) to 10 (significant knowledge gained), how would you describe your knowledge of the various effects of the COVID-19 pandemic after using our visualization tool?
(2) **Impact** - On a scale of 1 (not at all likely) to 10 (extremely likely), and relating to the COVID-19 pandemic, how likely are you to **consider** changes to any of the following after using our visualization tool?
    - Perspectives and views
    - Day-to-day habits and behavior
    - Lifestyle choices and direction

    If you answered 6 or higher for any of the above, please specify the changes you are considering.

We obtained a total of 20 responses from our survey, with results shown in Fig. 7 and 8. On the question around informativeness, all scores were 7 or higher, with an average rating of 8.55. As to the question around impact, 15 out of 20 participants rated 6 and above, with an average rating of 7.35, that they will consider changing their perspectives and habits, with all 15 of them willing to change their perspectives and views on the impact of COVID-19.

In general, the participants found that the new tool helped them gain new knowledge and insights on the effects of COVID-19 and they are likely to consider changing some of their perspectives and behaviors to help to curb the spread of COVID-19. More participants also found that our tool was more informative, insightful, impactful, and novel than the reference visualization from Our World in Data [9]. This preliminary evaluation showed that our tool has achieved its goal, although larger sample size is needed for stronger conclusions.

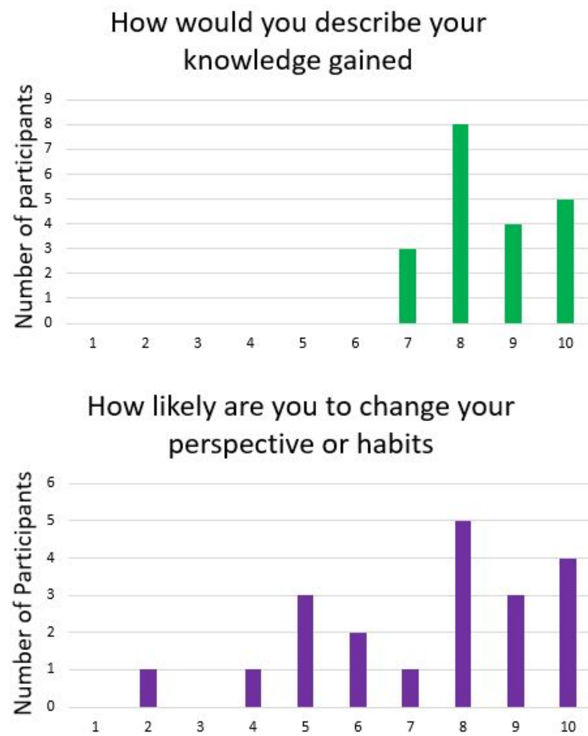

Figure 7: Results from questions 1 and 2 from our survey.

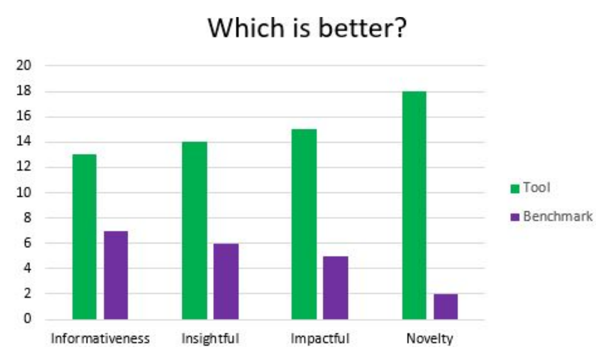

Figure 8: Results from comparison with external tool.

5 CONCLUSION AND DISCUSSION

In this project, we have developed a novel visualization tool for the COVID-19 pandemic relying on non-traditional but important social, economic and financial indicators. Through the three aspects - global visualization, linear regression, and clustering - we have found that our tool is able to have a large positive impact on users in our trial compared to other online visualizations.

Our tool provides high reusability, as the underlying data can be replaced for a variety of situations, including another pandemic.

Possibilities for future improvements include adding more indicators or using different sources for current indicators such as unemployment rate or inflation. However, we have found that obtaining up-to-date extensive economic data can be difficult due to limited availability.

All team members have contributed a similar amount of effort throughout the course of the project.

REFERENCES

[1] Selin Atalay and Gaye Solmazer. 2021. The relationship between cultural value orientations and the changes in mobility during the Covid-19 pandemic: a national-level analysis. *Frontiers in Psychology* 12 (2021).

[2] Martin Feldkircher, Florian Huber, and Michael Pfarrhofer. 2021. Measuring the effectiveness of US monetary policy during the COVID-19 recession. *Scottish journal of political economy* 68, 3 (2021), 287–297.

[3] Kristin J. Forbes and Francis E. Warnock. 2021. Capital flow waves—or ripples? Extreme capital flow movements since the crisis. *Journal of International Money and Finance* 116 (2021), 102394. https://doi.org/10.1016/j.jimonfin.2021.102394

[4] Nils Haug, Lukas Geyrhofer, Alessandro Londei, Elma Dervic, Amélie Desvars-Larrive, Vittorio Loreto, Beate Pinior, Stefan Thurner, and Peter Klimek. 2020. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature human behaviour* 4, 12 (2020), 1303–1312.

[5] Kazunobu Hayakawa and Hiroshi Mukunoki. 2021. Impacts of COVID-19 on global value chains. *The Developing Economies* 59, 2 (2021), 154–177.

[6] Carmen Moreno, Til Wykes, Silvana Galderisi, Merete Nordentoft, Nicolas Crossley, Nev Jones, Mary Cannon, Christoph U Correll, Louise Byrne, Sarah Carr, et al. 2020. How mental health care should change as a consequence of the COVID-19 pandemic. *The Lancet Psychiatry* 7, 9 (2020), 813–824.

[7] Marcos Raúl Ochoa and Carlos Maximiliano Albornoz. 2022. Economic and health impact of the COVID 19 Pandemic in Argentina and the world. *Estudios económicos* 39, 78 (2022), 223–242.

[8] Ahmed Rakha, Hansi Hettiarachchi, Dina Rady, Mohamed Medhat Gaber, Emad Rakha, and Mohammed M Abdelsamea. 2021. Predicting the economic impact of the COVID-19 pandemic in the United Kingdom using time-series mining. *Economies* 9, 4 (2021), 137.

[9] Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, and Max Roser. 2020. Coronavirus Pandemic (COVID-19). *Our World in Data* (2020). https://ourworldindata.org/coronavirus.

[10] Muhammad Shafiullah, Usman Khalid, and Sajid M. Chaudhry. 2022. Do stock markets play a role in determining COVID-19 economic stimulus? A cross-country analysis. *The World Economy* 45, 2 (2022), 386–408.

[11] Satoshi Tanaka. 2022. Economic impacts of SARS/MERS/COVID-19 in Asian countries. *Asian Economic Policy Review* 17, 1 (2022), 41–61.

[12] Md Ullah, Abu Tayab Moin, Yusha Araf, Atiqur Rahman Bhuiyan, Mark D Griffiths, David Gozal, et al. 2020. Potential effects of the COVID-19 pandemic on future birth rate. *Frontiers in Public Health* 8 (2020), 893.

[13] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9 (Nov. 2008), 2579–2605. http://www.jmlr.org/papers/v9/vandermaaten08a.html

[14] Dariusz Wójcik and Stefanos Ioannou. 2020. COVID-19 and Finance: Market Developments So Far and Potential Impacts on the Financial Sector and Centres. *Journal of Economic and Human Geography* 111, 3 (2020), 387–400.

[15] Libo Xu. 2021. Stock Return and the COVID-19 pandemic: Evidence from Canada and the US. *Finance Research Letters* 38 (2021), 101872.