2020 Seventh International Conference on Information Technology Trends (ITT). 25-26 November 2020

Visual Analysis of COVID-19 Trends

Jamal Alsakran
Computer Information Science
Higher Colleges of Technology
Fujairah, UAE
jalsakran@hct.ac.ae

Loai Alnemer
Computer Information Systems
The University of Jordan
Amman, Jordan
l.nemer@ju.edu.jo

2020 Seventh International Conference on Information Technology Trends (ITT) | 978-1-7281-8379-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/ITT51279.2020.9320891

***Abstract*—The declaration of COVID-19 as a pandemic has urged health and government authorities to investigate and control the outbreak. Among the various techniques that support the decision making, *visualization* represents a powerful tool to provide insightful information into every facet of the pandemic. In this effort, we provide a visual anaytic tool that facilitates the investigation of COVID-19 data including time series daily reports and mobility trends. The tool provides several visual layouts and enables users to gain insight into the impact of preventive and precautionary measures in combating the pandemic.**

***Index Terms*—COVID-19, visualization, mobility, treemap**

I. Introduction

The novel coronavirus COVID-19 is an infectious disease that was unknown before the outbreak began in Wuhan, China, in December 2019. COVID-19 is now a pandemic affecting most countries globally [1]. Countries have implemented various preventive and precautionary measures in an effort to contain the spread of the disease. Preventive and precautionary measures vary widely between countries and mostly include imposing lock down and movement restrictions, maintaining social distancing, wearing personal protective equipment, and schools and business closures.

After months of implementing these measures, the question remains as to whether they have been effective in slowing down and containing the spread [2], [3]. The answer to this question is debatable especially when taking into consideration the economic and social impact of such measures.

In this paper, we provide a visual analytic tool that incorporates several visual layouts and facilities investigative analysis of COVID-19 spread and mobility trends. The tool supports user interaction enabling user-in-the-loop exploratory analysis. The visual layouts include area plots for investigating the impact of preventive and precautionary measure on the COVID-19 spread. Furthermore, parallel coordinates, choropleth map, and treemap plots are provided to promote the understanding of how countries compare to each other.

Visualizing and tracking COVID-19 spread has been receiving increasing attention and various efforts have been initiated to visualize the spread and trends of the disease. The Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [4] provides an interactive dashboard tracking the global spread of COVID-19. Similarly, Zuo et al. [5] provide a dashboard with data mining and cloud computing capabilities for interactive data analytics and visualization to facilitate the understanding of the impact of the outbreak and corresponding policies such as social distancing on transportation systems

Among the early efforts to evaluate the effectiveness of some epidemic control measures, is the work of Chen at el. [6]. They collect and visualize publicly available data, then employ a mathematical model of disease transmission dynamics to offer a few tips on preventive measures. The authors in [7] propose a semantic visualization in order to enable exploration and discovery over large datasets of complex networks. They have applied their technique on COVID-10 dataset.

Our visual layout complements previous research by investigating the impact of mobility trends on COVID-19 spread and facilitating country wise comparison on various visual layouts.

II. Datasets

Health and research organizations are actively collecting and publishing up to date COVID-19 data from all over the world. Typically, the data is collected from figures published on official websites, in press releases and by social media accounts of national authorities, governments, ministries of health, or centers for disease controls. In this section, we describe the data that is used in our research.

A. Time Series

The Center for Systems Science and Engineering (CSSE) at John Hopkin University [8] collects and publishes up to date time series reports that track the global spread of COVID-19 cases. Table I describes the data fields that are processed in our visual analytic tool.

TABLE I: Time series data fields

Field	Description
Country/Region	Country, region or sovereignty name. The names of locations included on CSSE website [8]
Lat & Long	Latitude and Longitude coordinates
Confirmed	Counts include confirmed and probable (where reported).
Deaths	Counts include confirmed and probable (where reported)
Recovered	Recovered cases are estimates based on local media reports, and state and local reporting when available, and therefore may be substantially lower than the true number.
Observation Date	Date of observation from Jan 22, 2020 to present

B. Daily Report

CSSE also publishes reports on daily basis that includes most of the fields described in table I and few additional fields as described in table II

TABLE II: Daily report data fields

Field	Description
Active	Derivative field calculates as: Active cases = total confirmed - total recovered - total deaths
Incidence Rate	Incidence Rate = cases per 100,000 persons
Case-Fatality Ratio (%)	Case-Fatality Ratio (%) = Number recorded deaths / Number cases.

C. Google Mobility

Google publishes daily Mobility reports [9] that show how visits to places, such as grocery stores and parks, are changing due to the impact of the outbreak of COVID-19. The reports include categories that are useful to observe social distancing efforts as well as access to essential services. These categories are: Grocery & pharmacy, Parks, Transit stations, Retail & recreation, Residential, and Workplaces.

Changes for each day are compared to a baseline value for that day of the week. The baseline is the median value, for the corresponding day of the week, during the 5-week period Jan 3–Feb 6, 2020. The datasets show trends over several months with the most recent data representing approximately 2-3 days ago[1]. The data is reported starting Feb 15, 2020.

D. Apple Mobility

Similarly, Apple publishes daily Mobility reports [10] that are generated by counting the number of requests made to Apple Maps for directions in select countries/regions, sub-regions, and cities. The reports include the following categories: Driving, Walking, and Transit. Changes for each day are compared to a baseline value for that day of the week. The data is reported starting Jan 13, 2020 with a baseline at 100. For visual consistency, we have transformed the data to a baseline of 0.

E. Missing Data

Despite the tremendous effort taken, by the organizations, to collect accurate and up to date COVID-19 data, we observe missing data in most of the reports. Some of the missing data is attributed to national authorities decision not to report their COVID-19 cases, for example, Sweden authorities do not publish number of Recovered cases.

In other cases, the data is missing due to unavailability from the sources, for example, Apple Mobility data for May 11-12, 2020 is not available and appears as blank columns in the report. We handled the missing Apple Mobility data by substituting the missing values with the average of May 10 and May 13.

[1]This is how long it takes to produce the datasets

III. Tracking Spread of COVID-19

Understanding how COVID-19 is spreading over time is crucial in evaluating the effectiveness of the various preventive measures to slow down and contain the spread. We investigate the spread of COVID-19 cases on time series data in our initial effort to study the trends and rates of spread and identify critical regions.

Fig. 1[2] (a) shows the global increase of *Confirmed*, *Recovered*, and *Deaths* cases over the period between Jan 22, 2020 and Aug 30, 2020. The figure clearly shows that the number of *Confirmed* cases are increasing despite the various measures that are implemented globally to contain or slow down the spread. Fortunately, the number of *Recovered* cases is also increasing and *Deaths* cases seem to level off which indicates no increase over time.

As we discussed in section II, time series data is reported per Country/Region and it would make more sense to investigate each country separately due to the different combating measures implemented. In Fig. 1 (b) - (e), we plot the spread of cases in United States (US), Italy, Sweden, and United Arab Emirates (UAE), respectively. These countries are selected as representatives of how countries reacted differently to combat the disease.

Fig. 1 (b) shows the spread of COVID-19 in US that represents an example of countries that enforce moderate regional lock down and preventive measures. As of Aug 2020, US is the most effected country and as the figure conveys the cases are rapidly increasing which suggests that their combating measure have not be effective.

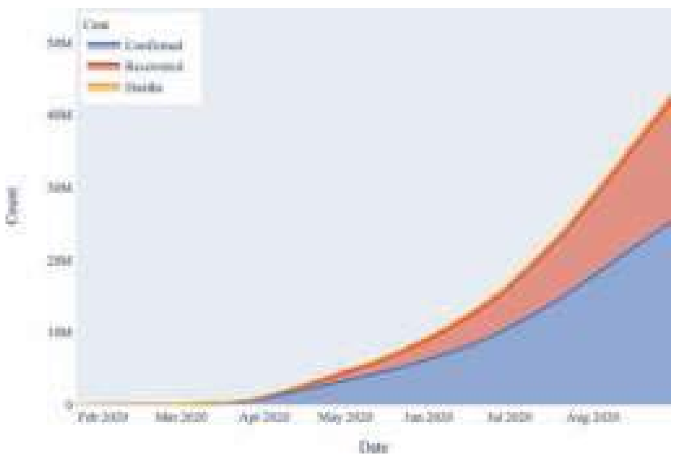
Fig. 1 (c) shows the spread of COVID-19 in Italy that is among the countries that had an early surge in cases which had led to a total lock down of the country for months starting March 2020. The figure shows the number of cases levels as the time progresses which suggest that the lock down has been effective in containing the spread.

Sweden has pursued a policy for containing the virus that avoided strict government measures such as school or restaurant closures. Fig. 1 (d) shows the number of cases is rapidly increasing and does not seem to show any signs of slowing down. It is worth noting that Sweden authorities do not report number of *Recovered* cases.

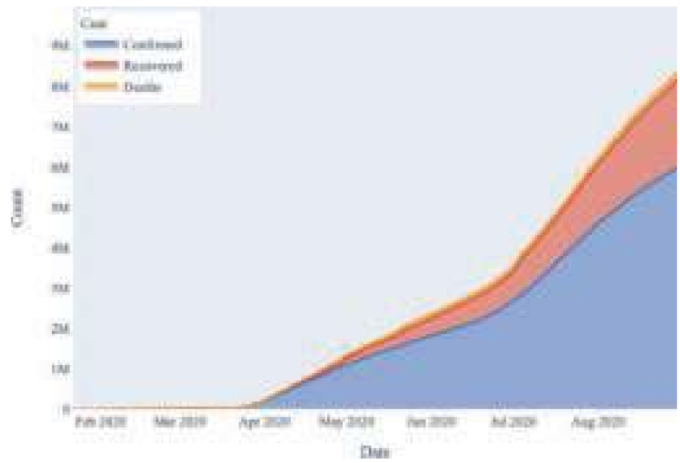
UAE had implemented various combating measures to contain the spread of the disease including a national sterilizing and disinfection program between March and June 2020. Fig. 1 (e) shows that the number of cases are increasing at a relatively low rate.

The time series data reports cumulative number of cases and as shown in Fig. 1 (a) - (e) the graph will either rise indicating new cases discovered or level off indicating no new cases. Such graph can be misinterpreted or confusing to viewers, we therefore plot the daily increase of global *Confirmed* cases in Fig. 1 (f). The figure depicts an increase in newly discovered cases with average around 250k new cases per day in July and August of 2020. The figure suggests that globally the

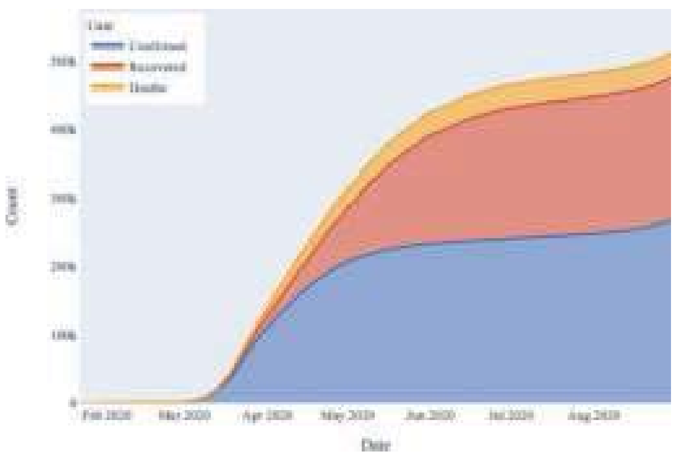
[2]Note that the figures are on different y-axis scales



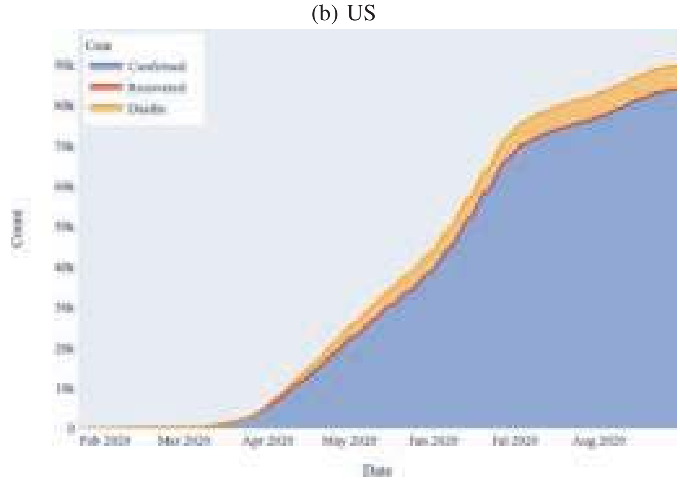
(a) Global



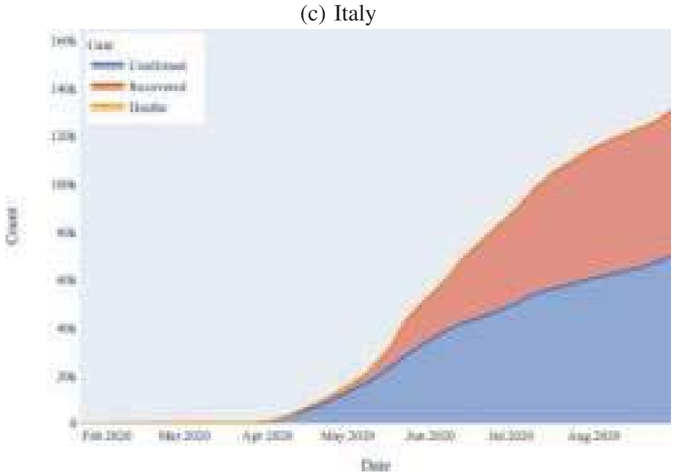
(b) US



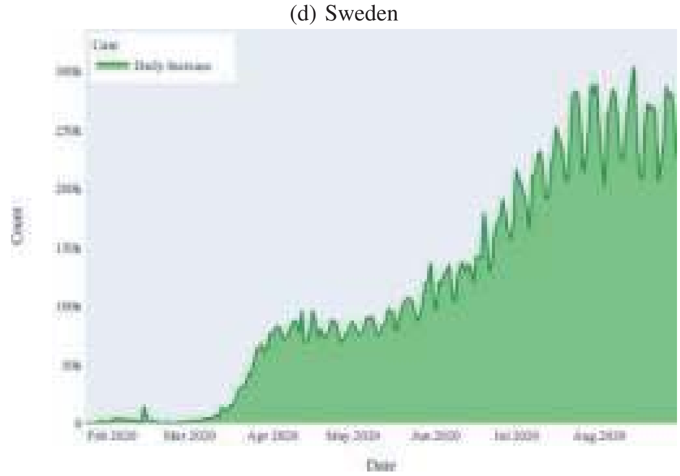
(c) Italy



(d) Sweden



(e) UAE



(f) Global - Daily increase

Fig. 1: Cases between Jan 22, 2020 and Aug 30, 2020

combating measures are not effective in containing the spread and we speculate that the number will stay one the rise for the foreseeable future. Nonetheless, the increase is not at an alarming rate and does not show a sudden surge in new cases.

IV. Mobility Trends

Mobility trends reflect the changes of public movement and social habits due to the preventive measures to combat and mitigate the effect of COVID-19. Among the various measures, mobility trends are mostly impacted by lock down, movement restriction and business closures.

In our effort to investigate the effectivenes of lock down and movement restrictions, we incorporate mobility data into our visual analytic tool. Mobility data is collected and reported on a daily basis by Google [9] and Apple [10] as described in section II.

Fig. 2: Daily Increase vs Apple Mobility in UAE

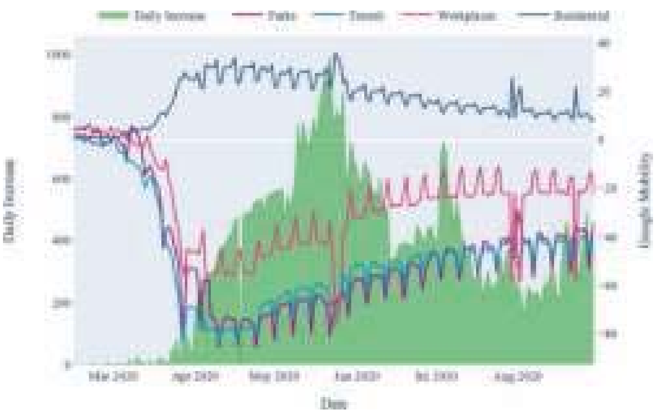


Fig. 3: Daily Increase vs Google Mobility in UAE

Fig. 2 plots the daily increase of *Confirmed* cases versus Apple Mobility that features Driving and Walking trends in UAE. Mobility trends are transformed to a baseline of 0 as shown on the right y-axis. The figure clearly shows that both Driving and Walking trends dipped below the baseline during the national serialization program and have gradually increased as the restrictions were loosened since June 2020. Furthermore, there is a sharp spike of movement around mid August that coincides with the start of the new academic year.

The trend of daily increase in Fig. 2 suggests little evidence to corroborate that the restrictions on movement is effective in containing the spread. Though, the sudden increase starting mid August could be attributed to that sharp spike of movement.

Fig. 3 plots the daily increase of *Confirmed* cases versus Google Mobility that features trends visits at Parks, Transit, Workplaces, and Residential places in UAE. Similarly, the trends are compared to a baseline of 0 as shown on the right y-axis. The figure readily shows Residential trend above the baseline due to lock down while Parks, Transit, and Workplaces below the baseline. Furthermore, all trends show repeated patterns that coincide with weekends, for example, Residential curve dips on weekends while Parks visits, on the other hand, increase on weekends.

Google Mobility, as Fig. 3 depicts, does not agree with the trend of daily increase of *Confirmed* cases and seems to dispute the effectiveness of restrictions on movement. In fact, the highest increase of newly cases, mid May, corresponds to low presence at Parks, Transit, and Workplaces.

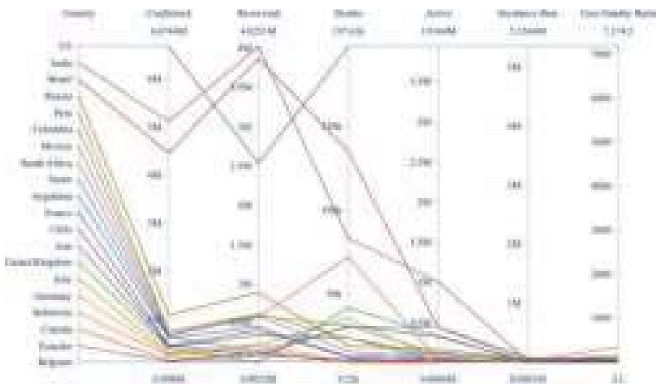


Fig. 4: Parallel coordinates of top 20 counties on Aug 30, 2020

V. Countries on Map

The rates of the spread vary widely between countries depending on the preventive and precautionary measures. In our visual analytic tool, we provide several visual layouts to investigate how countries compare to each other. The visual layouts process the daily report data described in section II.

Visualizing how countries compare to each other is quite challenging due to the higher number of countries and multiple data fields. In fact, this is a well known challenge in visualization literature referred to as ”The Curse of Dimensionality”. Commonly, this issue is tackled by provide a visual layout that supports user interaction. We provide Parallel Coordinates [12] layout that is commonly used in visualizing multidimensional data and perceiving their patterns.

Fig. 4 shows parallel coordinate layout depicting the top 20 most affected countries as reported on Aug 30, 2020. The figure readily shows that US, India, and Brazil are the most

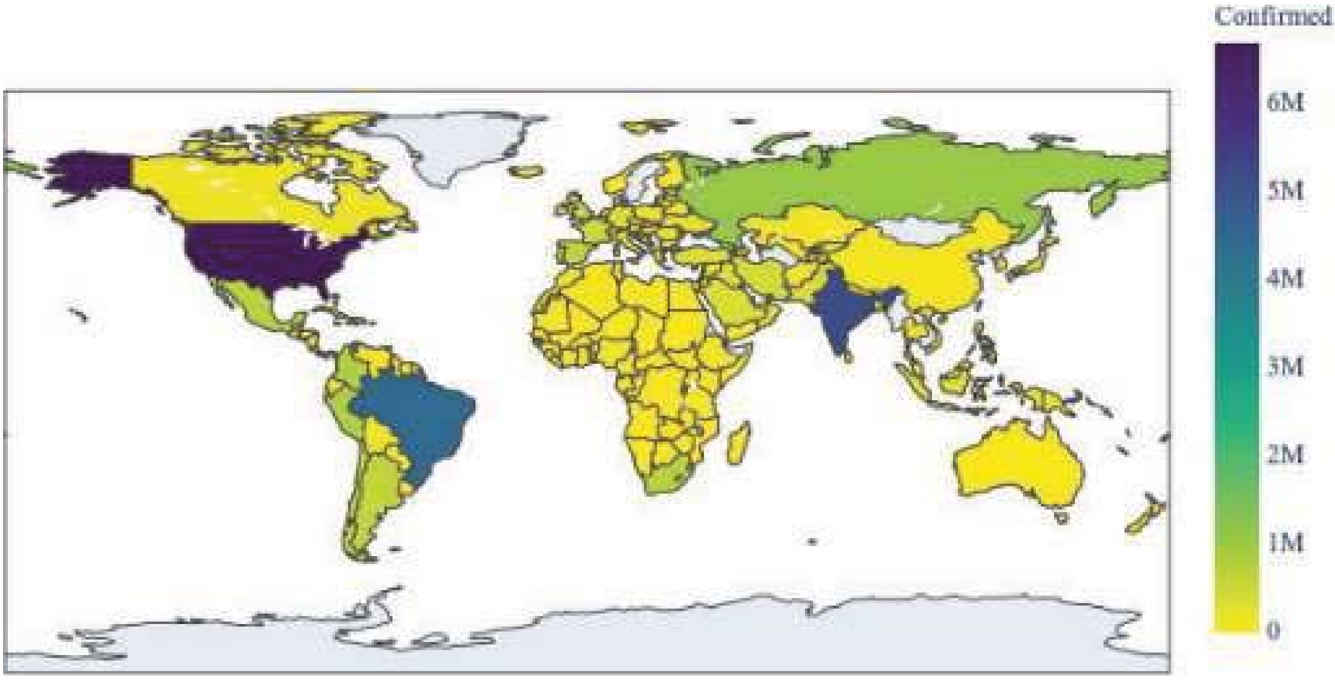


Fig. 5: Choropleth map of *Confirmed* cases on Aug 30, 2020

affected countries. Moreover, US situation is the most severe because *Incidence Rate* and *Case Fatality Rate* are extremely higher than any other countries including India and Brazil. Apart from top three, rates in other countries seem comparable.

Furthermore, our visual layout provides a choropleth map to visualize how the spread varies across geographic regions. The visual layout allows user to choose which field is encoded in color. In Fig. 5, the color is mapped to number of *Confirmed* cases which shows that US, India, and Brazil are among the most affected countries. China, for example, is among the lowest in number of *Confirmed* cases on Aug 30, 2020.

Choropleth maps are ubiquitous, well understood, easy to navigate, and invaluable in visualizing information across geographic regions. Nonetheless, information shown on a choropleth map is typically limited to one statistical value at a time. For example, the color in Fig. 5 represents *Confirmed* cases and user would need to interact with the layout to choose another field. An alternative layout that supports visualizing multiple features encoded in color, size, and hierarchy, namely Treemap [13], is also provided in our visual analytic tool.

Treemaps display hierarchical data as a set of nested rectangles. Each rectangle has an area proportional to a specified dimension of the data. Often the rectangles are colored to show a separate dimension of the data. When the color and size dimensions are correlated, one can often easily see patterns that would be difficult to spot in other ways. A second advantage of treemaps is that, by construction, they make efficient use of space

Fig. 6 (a) shows a treemap layout visualizing *Confirmed* cases on Aug 30, 2020. Both size and color are mapped to *Confirmed* and hierarchical is not utilized here, however, the visual layout allows users to include sub region and continent information to reflect hierarchical structures. As shown in the figure, US, India, and Brazil occupy most of the space followed by Russia that ranks fourth. In comparison with Fig. 5 where Russia occupies the most space, one could argue the benefit of treemap over choropleth map in encoding multiple information.

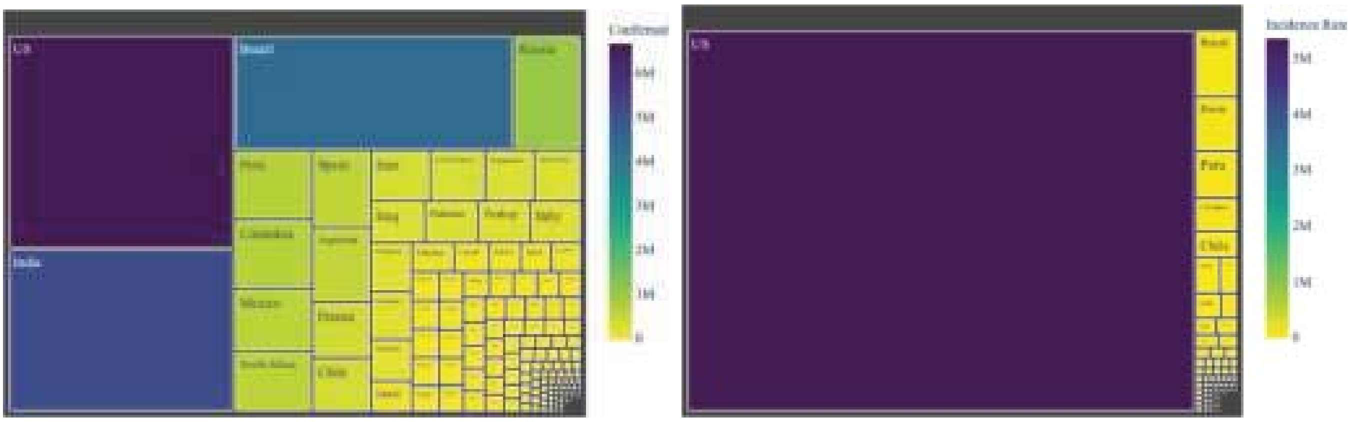
Furthermore, the visual layout supports user interactivity and allows users to choose among the various data fields. In Fig. 6 (b), we plot treemap with size and color being mapped to *Incidence Rate* which represents cases per $100,000$ persons. The figure provides insight into the severity of the situation in US as it dwarfs reported cases in all other countries. Moreover, the figure conveys that some countries with high *Confirmed* cases do not seem to be in critical situation, for example, India has a low *Incidence Rate* despite the high number of *Confirmed* cases.

VI. Conclusion

COVID-19 pandemic has claimed the life of millions and urged countries to implement preventive and precautionary measures to contain the spread. The effectiveness of those measures is not conclusively clear though and our findings suggest that movement restriction and business closure do not seem to be effective in slowing down the spread. Our visual analytic tool facilitates investigative analysis in order to gain insight into COVID-19 data.

References

[1] World Health Organization. Coronavirus disease 2019 (COVID-19). Situation report – 45. Geneva, Switzerland: World



(a) *Confirmed* (b) *Incidence Rate*

Fig. 6: Treemap of cases on Aug 30, 2020

Health Organization; 2020. https www.who.intdocsdefault-sourcecoronavirusesituationreports20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4

[2] Ng Y, Li Z, Chua YX et al. "Evaluation of the effectiveness of surveillance and response measures for the first 100 patients with COVID-19 in Singapore". MMWR Morb Mortal Wkly Rep 2020; 69:307–311.

[3] Saez, M., Tobias, A., Varga, D., Barceló, M.A., 2020. "Effectiveness of the measures to flatten the epidemic curve of COVID-19. The case of spain". Science ofThe Total Environment 727.

[4] COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https www.arcgis.comappsopsdashboardindex.html# bda7594740fd40299423467b48e9ecf6

[5] F. Zuo and Jingxing Wang and Jingqin Gao and Kaan Ozbay and Xuegang Jeff Ban and Yubin Shen and Hong Yang and Shri Iyer. "An Interactive Data Visualization and Analytics Tool to Evaluate Mobility and Sociability Trends During COVID-19", 2020, 2006.14882, cs.HC

[6] B. Chen, M. Shi, X. Ni, L. Ruan, H. Jiang, H. Yao, M. Wang, Z. Song, Q. Zhou, T. Ge. "Visual Data Analysis and Simulation Prediction for COVID-19", 2020 2002.07096, arXiv, physics.med-ph

[7] J. Tu, M. Verhagen, B. Cochran, J. Pustejovsky, "Exploration and Discovery of the COVID-19 Literature through Semantic Visualization", 2020, 2007.01800, arXiv, cs.CL

[8] CSSE at JHU, https github.comCSSEGISandDataCOVID-19treemaste̊csse_covid_19_data

[9] Google Mobility Report, https www.google.comcovid19mobility

[10] Apple Mobility Report, https covid19.apple.commobility

[11] E. Keogh, A. Mueen. "Curse of Dimensionality". In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA, 2017.

[12] A. Inselberg. "Parallel Coordinates: Visual Multidimensional Geometry and Its Applications". Springer-Verlag New York, 2009.

[13] M. Bruls, K. Huizing, J.J. van Wijk."Squarified Treemaps". In: de Leeuw W.C., van Liere R. (eds) Data Visualization 2000. Eurographics. Springer, Vienna. 2000.

