Data Visualization for the Understanding of COVID-19

João L. D. Comba
Universidade Federal do Rio Grande do Sul

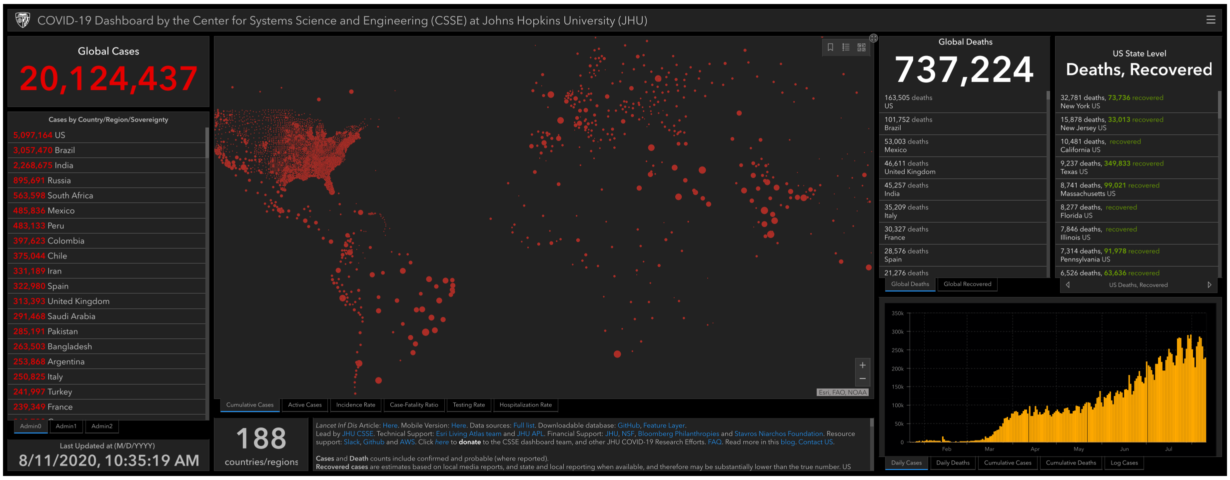
***Abstract*—Visualization techniques have been front-and-center in the efforts to communicate the science around COVID-19 to the very broad audience of policy makers, scientists, healthcare providers, and the general public. In this article, I summarize and illustrate with examples how visualization can help understand different aspects of the pandemic.**

THE DEADLY IMPACT of COVID-19 is driving a massive amount of research that aims at understanding the various characteristics of the pandemic. While there is no vaccine, considerable effort has been devoted to understanding the spread of the disease in different places in the world. The speed with which the disease has spread throughout the world demands agile solutions to understand and estimate the disease progression.

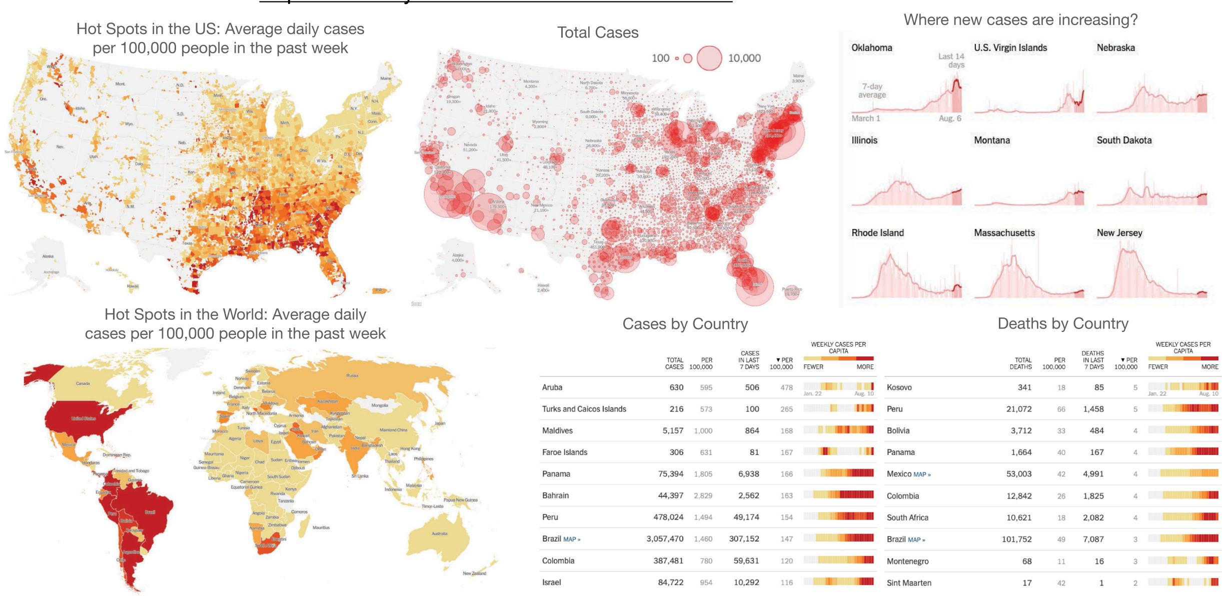
Interactive *dashboards* with several charts surfaced in different formats to offer concise ways to express the pandemic's growth. Figure 1 illustrates some examples. The dashboard developed by Johns Hopkins University (JHU)[1] was the first to track and display information on cases and death totals for different countries and states in the United States. Along with lists of total counts and histograms, a *bubble map* composed of circles of different radii allows a visual inspection of how serious the pandemic is around the world. Interesting plots were created on news outlets such as the New York Times (NYT)[2,3] and the Washington Post.[3] For example, NYT used *Choropleth maps*, a representation where geographical regions (countries or states) are mapped to colors associated with a measurement for that region (e.g., number of cases). This representation is useful to communicate trends, such as the average daily counts for the past week. Similarly, they display the time series evolution for each region using *line heatmaps*, where daily values are mapped to colors and displayed in a row. Our work (described in more detail in the following) used both Choropleth maps and

Digital Object Identifier 10.1109/MCSE.2020.3019834
Date of current version 9 October 2020.

JHU COVID-19 Dashboard https://coronavirus.jhu.edu/map.html (copyright 2020 Johns Hopkins University, all rights reserved)



The New York Times https://www.nytimes.com/interactive/2020/us/ (used with permission)



Brazil Dashboard https://covid19.ufrgs.dev/dashboard-int

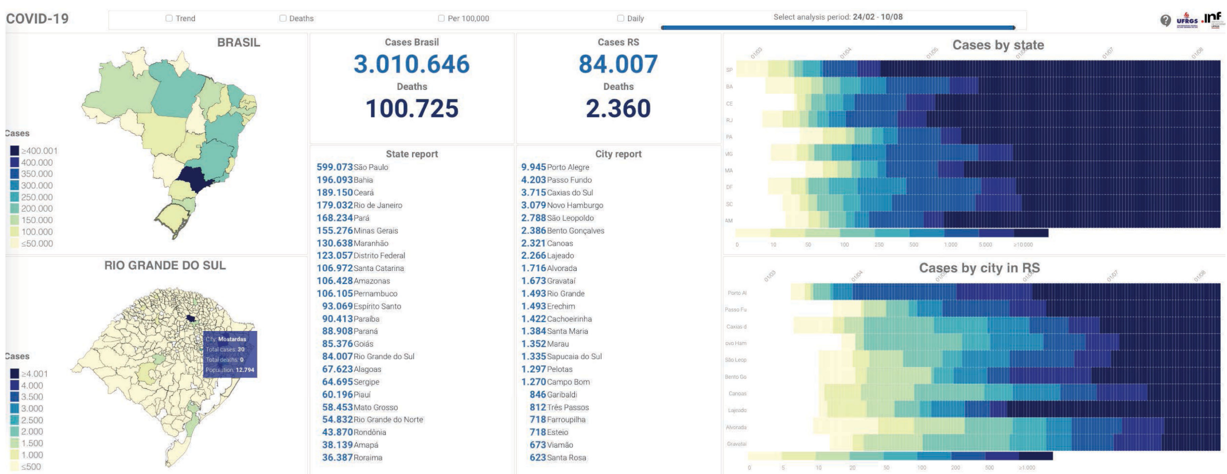


Figure 1. Dashboards and interactive tools to analyze COVID-19 data: Johns Hopkins University Dashboard, The New York Times charts (used with permission), and INF-UFRGS Brazil Dashboard.

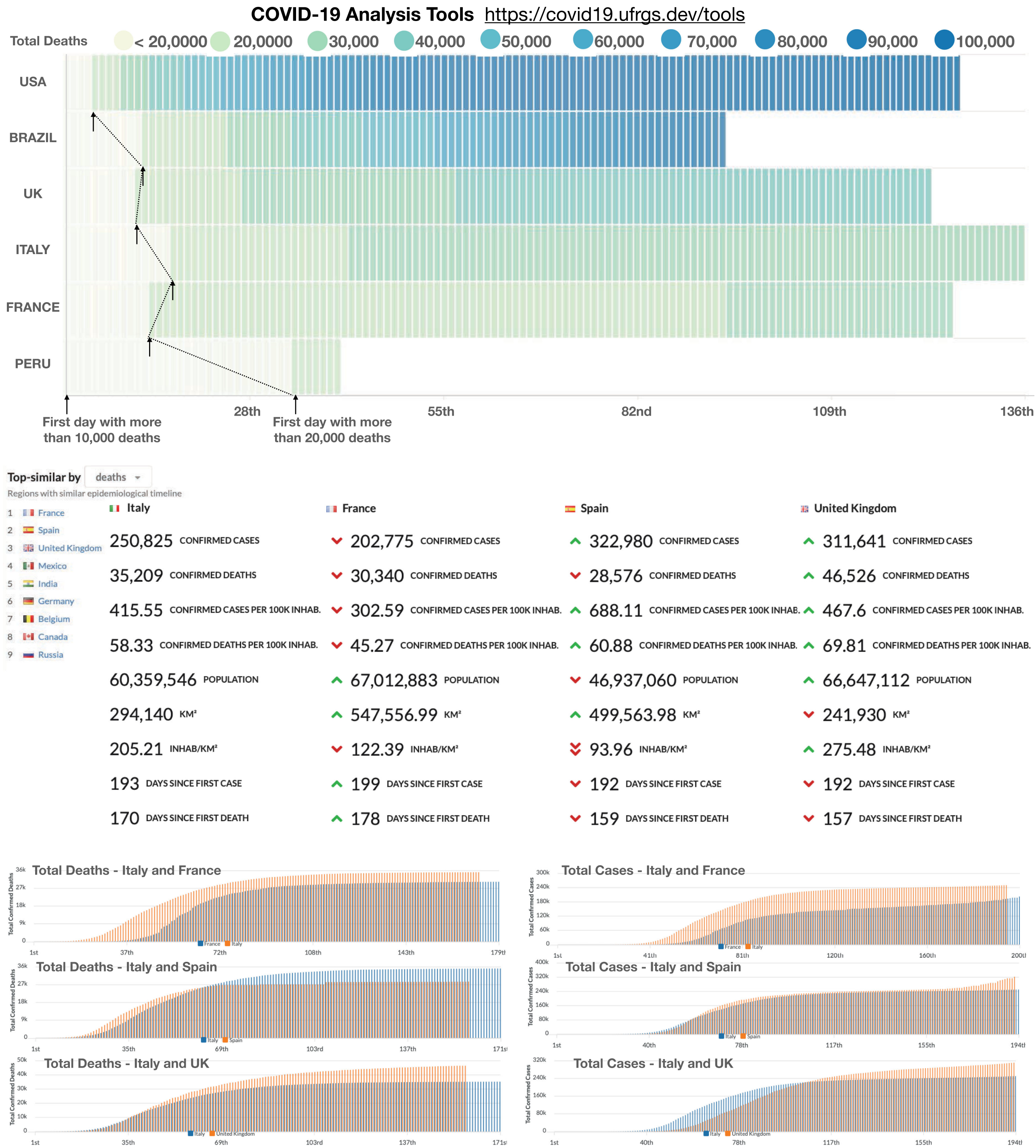


Figure 2. (Top) Heatmap matrices are useful for comparing time series such as the total deaths for different countries. Columns can be aligned by the first date after reaching a certain threshold, which allows us to compare when countries passed through specific checkpoints. (Bottom) Searching for places with similar timelines of deaths to Italy.

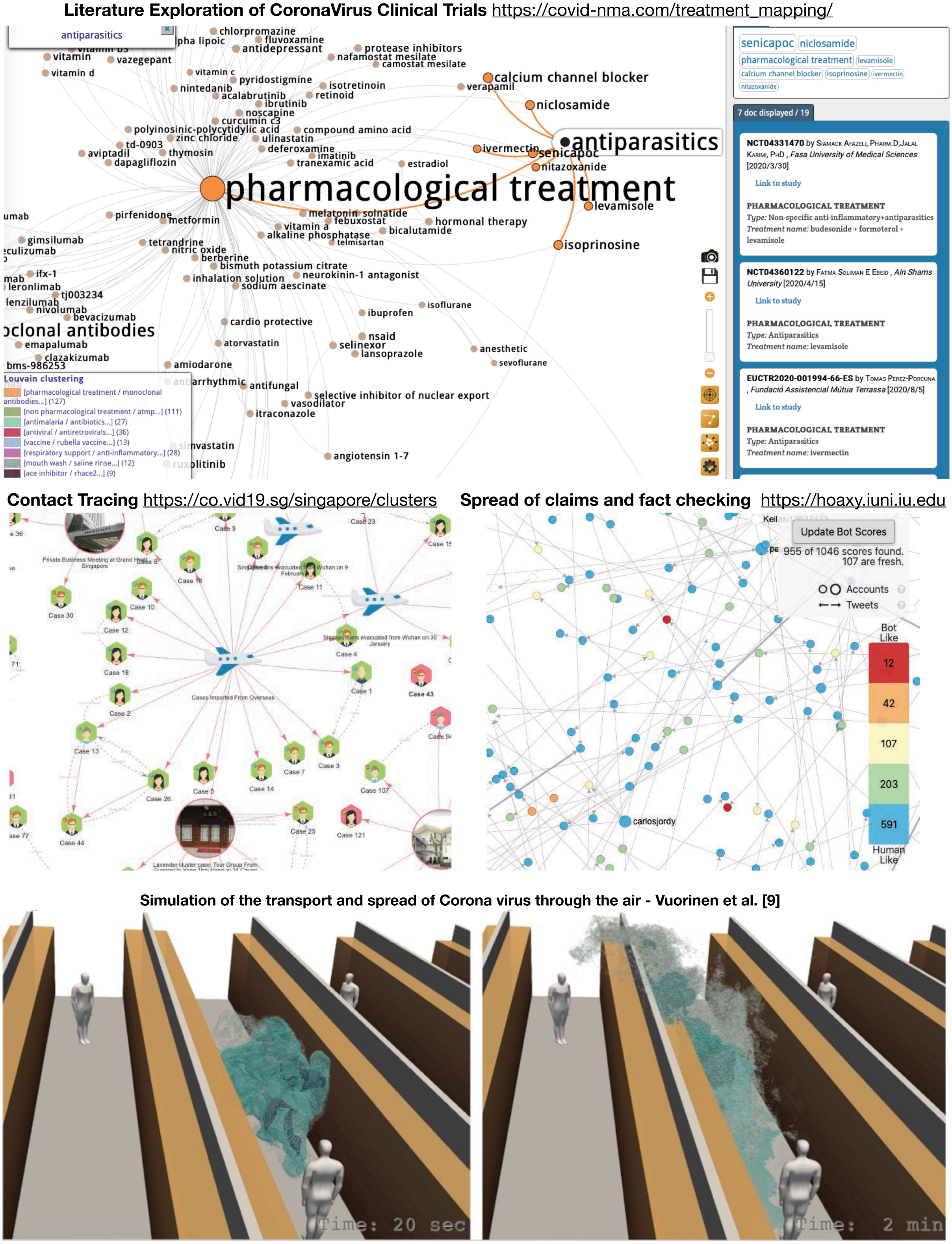


Figure 3. Applications: Literature exploration, contact tracing, spread of claims and fact checking, and simulation of the transport and spread of the novel coronavirus.

line heatmaps in the development of dashboard specially designed for Brazil. Using a focus-plus-context interface with coordinated views, the user can inspect the context using two Choropleth maps (one for the states and a second for a state selected), as well as details using two *matrix heatmaps* (collection of stacked line heatmaps) for states and cities. The tool displays daily or cumulative data, absolute or relative to population, and supports filtering by a time interval.

A vast collection of community-developed dashboards and interactive tools about COVID-19 are available. Good starting places to look are the data hub hosted by Tableau and the top 100 R-resources organized by Soetewey.[4] In-depth analysis is available at sites, such as *Our World in Data*,[5] *Bing*,[6] and the *COVID Tracking Project*,[7] among others. After developing the Brazilian dashboard, we devoted our efforts to create a set of tools to compare the spread of COVID-19 data in different regions of the world.[8] We collected data from more than 6 000 locations in the world, and our interface has different charts that support visualizing multiple locations in a single chart. Since the pandemic is at different stages in the world, we allow the user to align the time-series of data by a certain data the series passes a given threshold (e.g., after 100 cases). This representation is useful to observe when different locations passed through specific checkpoints (top of Figure 2).

While our initial charts support the comparison of different regions, the tool required the user to drive the process of choosing the regions to compare. From the beginning, we felt the need to have an automatic tool that could, given a region of interest, return the closest regions given a similarity function. Looking at the disease's spread in distinct places, but with similar growth patterns, can be useful to predict behaviors. Thus, we developed a search engine to support queries using different similarity functions. For example, at the bottom of Figure 2, we show the results of searching for regions similar to Italy concerning the number of deaths. The results are listed in a ranking, with pairwise comparisons that detail attributes of the different locations and evolution charts, aligned by the day of the first death. We observe the similarities in the evolution in the number of deaths and cases for Italy, France, Spain, and the United Kingdom. While using the tool, we also saw similarities among cities from Brazil and the United States, both countries with large COVID-19 numbers.

The examples so far give a glimpse of how data visualization can help in the understanding of COVID-19. Figure 3 illustrates other applications where data visualization can help. The first example shows how *multidimensional projections* and *network visualizations* can help the *literature exploration* of papers that describe novel coronavirus research. As new research about COVID-19 is published, there is a great need to review up-to-date literature and treatments conveniently. *Contact tracing* is another application that relies on graph visualization to trace the network of people who may have been in contact with a COVID-19 patient, an activity essential to control the dissemination of the disease and essential for directing social distancing regulations. Graph visualization is also important in for *social media spread and fact checking*. With many people at home, social networks are playing a significant role in people's lives these days. Unfortunately, the dissemination of fake news and automated posting from robots is also rising. Fact-checking over the propagation network can help identify misleading information and patterns of dissemination. The fourth and final example highlights the importance of data visualization in the analysis of scientific simulations. It shows the visualization of simulating the transport and spread of novel coronavirus in closed spaces,[9] which shows how an infected person can disseminate the virus indoors.

Many other examples include data visualization in the analysis of COVID-19 data, and many more is surfacing every day. We hope this summary highlights interesting examples, give pointers to other references, and motivate people to pursue other applications.

REFERENCES

1. E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infectious Diseases*, vol. 20, pp. 533–534, 2020, doi: 10.1016/S1473-3099(20)30120-1.
2. The New York Times, "Coronavirus map: Tracking the global outbreak," 2020. [Online]. Available: https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html

3. The Washington Post, "Mapping the worldwide spread of the coronavirus," 2020. [Online]. Available: https://www.washingtonpost.com/graphics/2020/world/mapping-spread-new-c oronavirus/
4. A. Soetewey, "Top 100 R resources on novel COVID-19 coronavirus," 2020. [Online]. Available: https://www.statsandr.com/blog/top-r-resources-on-covid-19-coronavirus
5. D. Kennedy, A. Seale, D. Bausch, H. Ritchie, and M. Roser, "How experts use data to identify emerging COVID-19 success stories," 2020. [Online]. Available: https://ourworldindata.org/identify-covid-exemplars
6. Microsoft BING, "COVID-19 live map tracker by Microsoft," 2020. [Online]. Available: https://www.bing.com/covid/
7. The Atlantics, "COVID tracking project," 2020. [Online]. Available: https://covidtracking.com
8. F. Pontes, F. Pinto, W. Leuschner, and J. L. D. Comba, "COVID-19 analysis tools at Instituto de Informática-ufrgs," 2020. [Online]. Available: https://covid19.ufrgs.dev
9. V. Vuorinen *et al.*, "Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors," *Safety Sci.*, vol. 130, 2020, Art. no. 104866. doi: 10.1016/j.ssci.2020.104866.

João L. D. Comba is currently a full professor in computer science with Instituto de Informática—UFRGS, Porto Alegre, Brazil. His research interests include visualization, visual analytics, data science, geometric algorithms, and spatial data structures. He received the Ph.D. degree in computer science from Stanford University, Stanford, CA, USA. Contact him at comba@inf.ufrgs.br.