

Visualizing spatiotemporal epidemic clusters on a map-based dashboard: a case study of early COVID-19 cases in Singapore

Hui Zhang, Chenyu Zuo, Linfang Ding

Angaben zur Veröffentlichung / Publication details:

Zhang, Hui, Chenyu Zuo, and Linfang Ding. 2021. "Visualizing spatiotemporal epidemic clusters on a map-based dashboard: a case study of early COVID-19 cases in Singapore." Proceedings of the ICA 4: 117.
<https://doi.org/10.5194/ica-proc-4-117-2021>.

Nutzungsbedingungen / Terms of use:

CC BY 4.0

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

CC-BY 4.0: Creative Commons: Namensnennung

Weitere Informationen finden Sie unter: / For more information see:

<https://creativecommons.org/licenses/by/4.0/deed.de>



Visualizing Spatiotemporal Epidemic Clusters on a Map-based Dashboard: A case study of early COVID-19 cases in Singapore

Hui Zhang^a, Chenyu Zuo^{b*}, Linfang Ding^c

^a Department of Aerospace and Geodesy, Technical University of Munich, Germany, ge35qup@mytum.de

^b Chair of Cartography, Technical University of Munich, Germany, chenyu.zuo@tum.de

^c KRDB Research Centre for Knowledge and Data, Faculty of Computer Science, Free University of Bozen-Bolzano, 39100 Bozen-Bolzano, Italy, linfang.ding@unibz.it

* Corresponding Author

Abstract: Spatiotemporal distribution of the epidemic data plays an important role in its understanding and prediction. In order to understand the transmission patterns of infectious diseases in a more intuitive way, many works applied various visualizations to show the epidemic datasets. However, most of them focus on visualizing the epidemic information at the overall level such as the confirmed counts each country, while spending less effort on powering user to effectively understand and reason the very large and complex epidemic datasets through flexible interactions. In this paper, the authors proposed a novel map-based dashboard for visualizing and analyzing spatiotemporal clustering patterns and transmission chains of epidemic data. We used 102 confirmed cases officially reported by the Ministry of Health in Singapore as the test dataset. This experiment shown that the well-designed and interactive map-based dashboard is effective in shorten the time that users required to mine the spatiotemporal characteristics and transmission chains behind the textual and numerical epidemic data.

Keywords: COVID-19, geovisualization, map-based dashboard, visual analysis

1. Introduction

A map-based dashboard is an intuitive interface in informing and communicating multivariable information. It integrates multimedia presentation styles such as texts, charts, images, maps, videos, and gauges to allow users to instantly understand the data (Zuo et al., 2020). It is been increasingly used in various domains, such as business intelligence (Peters et al., 2016), online education (He et al., 2019), and urban planning (Wurstle et al., 2020).

Due to the rich spatiotemporal information contained in epidemic data, increasing studies adopt dashboards to monitor and analyze the changes in public health. Johns Hopkins University designed an interactive map-based dashboard for the ongoing COVID-19 epidemic (Johns Hopkins University, 2020). This real-time dashboard uses a bubble map with auxiliary texts and numbers to visualize the locations and numbers of confirmed COVID-19 cases, deaths and recoveries for all affected countries. It enables researchers, public health authorities and the general public to monitor and track the outbreak worldwide. However, sometimes just knowing the general news cannot meet people's demands. They are more eager to know local outbreak in detail around them. The Singapore news platform Zaobao uses various visualizations to unfold the local COVID-19 outbreak (Zaobao, 2020). It not only presents the development trend of COVID-19 using bar charts and stacked area charts, but also contains a dot map to show the spatial distribution of confirmed cases in different officially reported infection groups. Furthermore, some researches applies spatial analytical techniques to the epidemic data, such as using kernel smoothing and spatial cluster detection to investigate the clustering of HIV infections in South African (Tanser et al., 2009), or using sankey diagrams

to help discover the source of COVID-19 confirmed case in Henan Province, China (Liu et al., 2020). Visualizing the results of spatial analysis is efficient in helping users to understand the development trend, transmission chains and spatial distribution than just visualizing the locations of confirmed cases. However, these dashboards still can be improved in terms of interactions to meet different people's demands. For example, they can provide filtering function to enable people focus on the development trend of the epidemic in a certain period.

In order to not only present the results of spatial analysis but also provide effective interactions which can allow users to navigate to selected data and display it at various levels of detail and in various formats, we proposed a map-based interactive dashboard to communicate the spatiotemporal distribution characteristics and transmission chains of large epidemic data to the public.

2. Methodologies

The aim of this study is to help users with diverse backgrounds to understand the features of epidemic data according to their demands more quickly compared with only reading the original textual data. The map-based dashboard serves as the interface to help users view the epidemic data intuitively and interactively. In this section, we introduce principles for spatial clustering, several visualization methods and the dashboard design.

2.1 Spatial clustering

Spatial phenomena have the tendency to co-occur spatially according to Tobler's First Law of Geography (Tobler, 1970). The spatial clustering, which represents a general tendency of events occurring closer to each other, is one of

the most common spatial patterns of point events (Tao and Thill, 2016). Objects within a cluster show a high degree of similarity, whereas the clusters are as much dissimilar as possible. Spatial clustering methods can roughly be divided into five categories: the partition clustering, the hierarchical clustering, the density-based clustering, the fuzzy clustering and the model-based clustering (Pattnaik, 2020). The density-based clustering groups dense regions in the data space, separated by regions of lower object density. One well-known density-based clustering algorithm is the DBSCAN algorithm (Density-based spatial clustering with the application of noise). For example, a team from University of Liverpool used DBSCAN to identify the local retail agglomerations within Great Britain (Pavlis et al., 2018). Also, it can be used in the public healthy domain, like determining the potential areas of dengue fever outbreaks based on daily surveillance information in India (Nandana et al., 2019). As it works well in planar space and in against noisy data points, we used this algorithm in our experiment.

2.2 Visualization methods

Visualizations play an important role in data analytics and help explore complex data. Different visualization methods can be used to show the spatial distribution or some other perspectives of the data, such as dot map, heat map, flow map, and various charts. A dot map is efficient in helping detecting spatial the distribution of data in an area by placing equally sized points (The Data Visualisation Catalogue, 2021). For example, a team in Europe used the dot map to detect the Q fever outbreak in the Netherlands and the pertussis outbreak in Germany (Soetens et al., 2017). A heat map shows magnitude of a phenomenon by using variations in color. It provides obvious visual cues to help investigate how the phenomenon varies over space. For instance, the heat map was used to help learn spatial-temporal characteristics of urban population aggregation in central areas of Wuhan, China (Lucang, 2018). A flow map combines a map and a flow chart. It uses lines to show the movement of objects from one locations to another, such as the number of people in a migration, the amount of goods being traded, or the number of packets in a network (Doantam Phan et al., 2005). Charts are often used in combination with maps. For example, a team from China created a map-based dashboard which contained a doughnut pie chart and a bar chart to help to illustrate the percentage of different types of COVID-19 cases and the numbers of new confirmed cases daily in Henan (Liu et al., 2020).

2.3 Map-based dashboard

Map-based dashboard has its specific design guidelines. The visualization, layout, function of panels, color, font, and interactive tools are important elements to be considered in dashboard design process. There are many design principles need to be considered. (1) Identify what information to be visualized and choose the suitable visualizations (Janes et al., 2013). For example, a pie chart can be used to visualize a part to whole relationship; (2) Pay attention to the compactness and the modularity of the content. Presenting information compactly helps users gather more information faster and separating information into blocks using borders makes the dashboard clearer; (3) Ensure that each panel has only one function, which makes the content

in the dashboard much easier to be understood; (4) Choose suitable colors and fonts (Minhas, 2019). Usually, neutral colors such as white, black or gray can be used as the background color, which can make text or data visualization elements more prominent. The font styles including size, family and color in the dashboard should be selected according to the importance of the text; (5) Offer different interactions such as selecting, filtering, or drilling down, to enable users to view and explore information much more effectively and flexibility.

3. Data and dashboard design

We designed an map-based dashboard prototype incorporate with the proposed methods. We adopted the public COVID-19 epidemic data in Singapore as a test data. In this section, we introduce the test area, the test data and the dashboard interface.

3.1 Test area and data

Singapore is an island city-state located in Southeast Asia. It consists of the diamond-shaped Singapore Island and some 60 small islets. It is heavily urbanized and has a high population density (Leinbach et al., 2021). The earliest COVID-19 case in Singapore was reported in January 2020 and then this infectious disease quickly spread in local. The epidemic reached its most serious stage in April 2020. In April 22, more than 1000 new cases were reported. But then the epidemic was gradually controlled by the government through effective policies. By February 2021, the number of confirmed new cases per day has been less than 10.

We collected the daily COVID-19 reports from the Ministry of Health (MOH) of Singapore (<https://www.moh.gov.sg/covid-19>). We selected the first 102 confirmed cases in this study, because they included the geoinformation of the cases. The time coverage is from January 24th, 2020 to February 29th, 2020.

The data preprocessing including mainly data cleaning, structuring, and geocoding. The MOH updates the COVID-19 local situation in text description every day. We cleaned and summarized the raw data based on keywords such as "citizen", "is linked to", "worked at" and "stayed at home at". The selected information for each case are the spatial information, temporal information, and the infection information: the case source, the confirmed date, the home or the work address of the domestic cases, the hotel address of the imported cases, the infection group, and the infection linkage between the cases. In order to protect the patients' privacy, the MOH only offered the street names without street numbers of the home and work addresses. Moreover, We converted the street names into latitudes and longitudes using the geocoding toolbox in ArcGIS Pro.

Take Case 88 for an example (<https://www.moh.gov.sg/news-highlights/details/two-more-cases-discharged-three-new-cases-of-covid-19-infection-confirmed>), the textual description related to this case offered by the MOH on 22 February is:

"

Links between previous cases found

.....

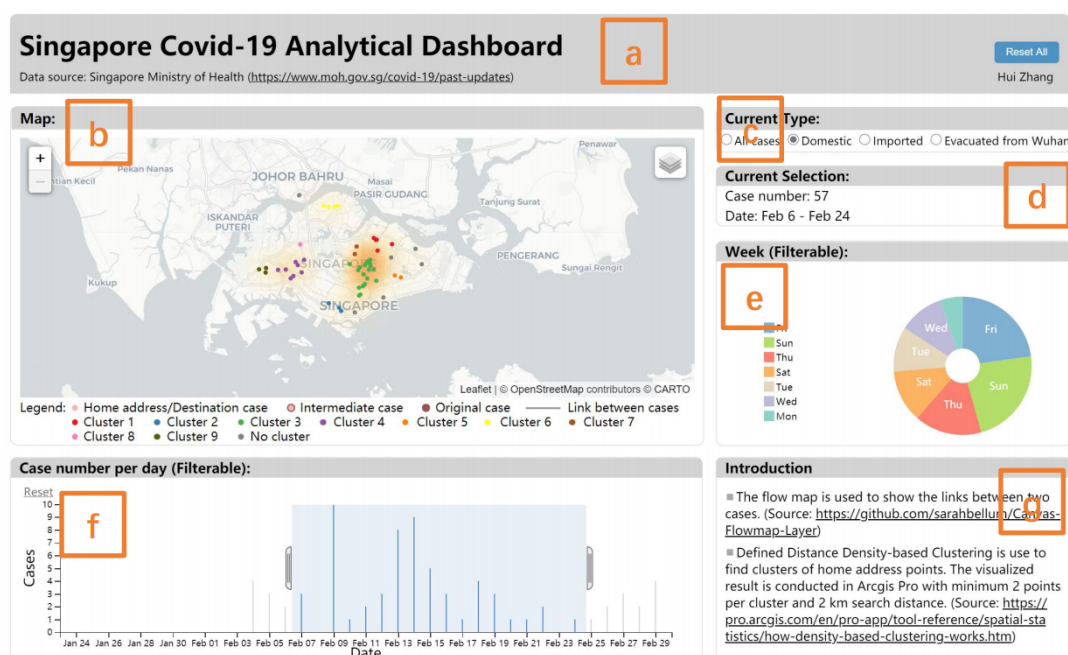


Figure 1. The map-based dashboard consists of seven panels showing the COVID-19 cases epidemic data in Singapore (a) a title panel; (b) a spatial information panel; (c) a case source selection panel; (d) a selected information panel; (e) a weekly information panel; (f) a daily information panel; (g) an background information panel.

(e) 23 of the confirmed cases (Cases 48, 49, 51, 53, 54, 57, 58, 60, 61, 62, 63, 66, 67, 68, 70, 71, 73, 74, 78, 80, 81, 84 and 88) are linked to the Grace Assembly of God.

.....

About the confirmed cases

.....

Case 88

1. Case 88 is a 30 year-old female Singapore Citizen who has no recent travel history to China. She is currently warded in an isolation room at NCID. She is linked to Case 66.

2. She reported onset of symptoms on 16 February and had sought treatment at a general practitioner (GP) clinic on 17 February. As she had been identified as a contact of Case 66, she was referred by MOH to NCID on 21 February.

Subsequent test results confirmed COVID-19 infection on 21 February afternoon. 3. Prior to hospital admission, she had mostly stayed at her home at Hougang Street 91."

Based on this report, the information we summarized are the type (a domestic case), confirmed data (22.02.2020), home address (Hougang Street 91), the infection group it belongs (Grace Assembly of God), linked cases (48, 49, 51, 53, 54, 57, 58, 60, 61, 62, 63, 66, 67, 68, 70, 71, 73, 74, 78, 80, 81, 84, 88). And then we geocoded her home address to the latitude(1.377N) and the longitude(103.882E).

3.2 Dashboard interface design

This subsection presents the style and the structure of the dashboard and introduces the interactions designed for exploration of the epidemic data.

We adopt a user-friendly design styles for this dashboard interface. The main color scheme of the map-based dashboard is white and gray. Each panel is bounded by a gray

border, while the white space is used to separate them from each other. The title of each panel has the gray background, to facilitate the user to distinguish the titles from other information. Moreover, these panels are linked with each other. If users changed the content in one panel, the content in other panels will change accordingly.

The dashboard consists of seven panels as shown in Figure 1: (a) a title panel containing the title, the data source, the producer, and a reset button; (b) a spatial information panel showing the locations of confirmed cases and a legend of the shown maps; (c) a selection panel that allow users to switch among different sources; (d) a information panel providing the aggregated information of total case number and the time range according to the user selection; (e) a weekly information panel showing the aggregated number and its proportion in a week; (f) a daily information panel showing the number of confirmed new cases on a daily basis; (g) a background information panel providing information about how to interpret the dashboard and the analysis methods we have applied.

Various and flexible interactions are provided to enable users to explore the data. (1) Select different spatial analysis results shown on maps (Figure 1b). Dot maps are used to present the locations of the home address, the work address and the hotel address and also the spatial clustering results (based on DBSCAN algorithm). The heat map is responsible for showing how the COVID-19 epidemic phenomenon varies over space and the flow map shows the chains of infection among confirmed cases. (2) Select the source of the cases: all cases, domestic cases, imported cases, or evacuated cases from Wuhan (Figure 1c). (3) Select the time range by clicking and dragging on the bar chart (Figure 1f). (4) Filter the cases according to the week days by clicking in the pie chart (Figure 2e).

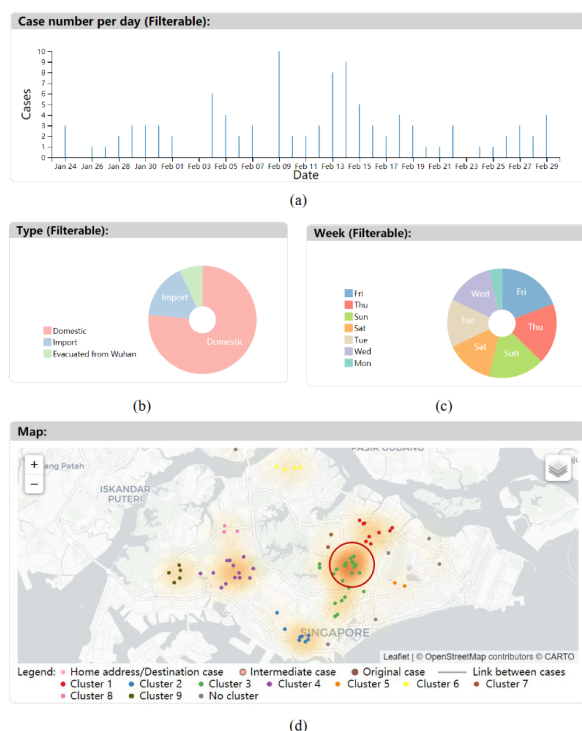


Figure 2. Analyzing the spatiotemporal distribution using the proposed dashboard: (a) the daily information of all confirmed cases; (b) the proportion of different sources; (c) the weekly pattern of domestic cases; (d) the spatial clustering analysis of domestic cases.

4. Use cases

In this section, we demonstrate how the dashboard can help users to gain insights by applying it to two use cases. We focus more on the spatial-related analysis, therefore we describe the spatiotemporal distribution and the spatial patterns of the transmissions.

4.1 Spatiotemporal distribution analysis

The spatiotemporal distribution analyses help users to understand the spatial distribution and the development trend of COVID-19 in Singapore. In combination of a dot map showing the spatial clusters, a bar chart and a pie chart showing the temporal information as shown in Figure 2, users can have an in-depth understanding.

In the dashboard, we can use the heat map and the dot map, which presents the results of DBSCAN spatial clustering algorithm, to find the spatial clustering patterns. Figure 2(d) presents the heat map of locations of domestic cases. We can see that there is an area, which is indicated by a red circle, with the highest density of the confirmed cases in central Singapore. As for spatial clustering results, we can see that data points were classified into 9 clusters. Among these clusters, cluster 1, 3 and 4 are the three with the largest numbers of people.

The bar chart and the pie chart help to investigate the development trend and the temporal clustering of the epidemic. From the bar chart in Figure 2(a), we can see that there were less than 10 confirmed new cases every day at the early stage of the epidemic in Singapore. After interacting with the pie chart (selecting only one type of cases) in Figure 2(b), we notice that the first confirmed 17 cases were all imported cases from Wuhan and then the COVID-19

quickly spread among the locals. For the domestic cases, we can notice from the pie chart in Figure 2(c) that the count of confirmed cases was lower on Mondays, while counts of confirmed cases were more equal on the other six days in a week.

4.2 Transmission analysis

The transmission analyses help users figuring out how the disease spread among people. The flow map presents the transmissions between domestic cases as shown on Figure 3. There are three important source cases, case 19 from Yong Thai Hang cluster, 66 from Grace Assembly of God cluster, and case 93 from Wizlearn Technologies Cluster as noted by three red circles. After these three cases were confirmed, some people who had been contacted with them were later confirmed to have COVID-19.

Note that the date of diagnosis is usually not the date of being infected. Therefore, it is very hasty to draw the conclusions that people diagnosed earlier are then the source of infection. In order to determine the transmission patterns more accurately, more confirmed cases should be studied and further detailed information like the specific contact tracking are needed.

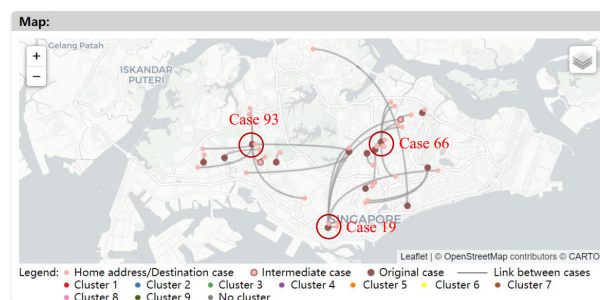


Figure 3. Analyzing the transmission using the flow map. The red circles indicates three most dominated transmission groups.

5. Conclusion

This paper proposed a novel map-base dashboard for analyzing and visualizing the infectious disease transmission data. Our proposed dashboard demonstrated the calculated spatial clusters, daily and weekly temporal trend, and the background information. The interactive visualizations allow users to learn and reason the patterns and transmission visually. By using this interface, we could easily read the spatiotemporal clustering patterns and the transmission chains among the 102 cases.

In the future, we plan to conduct an usability evaluation to improve the proposed map-based dashboard design. The evaluation will focus on how well the users can use the map-base dashboard to get their needed information and refers to how satisfied users are with the interactions. Besides, we plan to integrate other information, such as social media data into our map-based dashboard. This will allow us to study the impact of media reports about the epidemic or measures the effectiveness of the launched policies.

References

Doantam Phan, Ling Xiao, Yeh, R. and Hanrahan, P., 2005. Flow map layout. In: *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 219–224.

- Janes, A., Sillitti, A. and Succi, G., 2013. Effective dashboard design. *Cutter IT Journal* 26(1), pp. 17–24.
- Johns Hopkins University, 2020. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). <https://coronavirus.jhu.edu/map.html>. Assessed January 19, 2020.
- Leinbach, T. R., Ho, R., Winstedt, R. O. and Kennard, A., 2021. Singapore. <https://www.britannica.com/place/Singapore>. Assessed February 9, 2021.
- Liu, J., Liu, H., Chen, X., Li, J., Kang, L. and Zhao, Q., 2020. Visual analysis method for covid-19 epidemic situation. *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics* pp. 1617–1627.
- Lucang, W., 2018. Spatial-temporal characteristics of urban population aggregation based on baidu heat map in central areas of wuhan city. *J. Hum. Settl. West China* 33, pp. 52–56.
- Minhas, S., 2019. 10 Rules of Dashboard Design. <https://medium.muz.li/10-rules-of-dashboard-design-f1a4123028a2>. Assessed January 21, 2021.
- Ministry of Health in Singapore, 2020. Two More Cases Discharged; Three New Cases of COVID-19 Infection Confirmed. <https://www.moh.gov.sg/news-highlights/details/two-more-cases-discharged-three-new-cases-of-covid-19-infection-confirmed>. Assessed February 6, 2021.
- Nandana, G. M., Mala, S. and Rawat, A., 2019. Hotspot detection of dengue fever outbreaks using dbscan algorithm. In: *2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pp. 158–161.
- Pattnaik, A., 2020. Geospatial Clustering: Kinds and Uses. <https://towardsdatascience.com/geospatial-clustering-kinds-and-uses-9aef7601f386>. Assessed December 29, 2020.
- Pavlis, M., Dolega, L. and Singleton, A., 2018. A modified dbscan clustering method to estimate retail center extent. *Geographical Analysis* 50(2), pp. 141–161.
- Soetens, L., Hahné, S. and Wallinga, J., 2017. Dot map cartograms for detection of infectious disease outbreaks: an application to q fever, the netherlands and pertussis, germany. *Eurosurveillance* 22(26), pp. 30562.
- Tanser, F., Bärnighausen, T., Cooke, G. S. and Newell, M.-L., 2009. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *International Journal of Epidemiology* 38(4), pp. 1008–1016.
- Tao, R. and Thill, J.-C., 2016. Spatial cluster detection in spatial flow data. *Geographical Analysis* 48(4), pp. 355–372.
- The Data Visualisation Catalogue, 2021. Dot Map. https://datavizcatalogue.com/methods/dot_map.html. Assessed February 9, 2021.
- Tobler, W. R., 1970. A computer movie simulating urban growth in the detroit region. *Economic Geography* 46(sup1), pp. 234–240.
- Zaobao, 2020. Novel Coronavirus Cases in Singapore. <https://interactive.zaobao.com/2020/novel-coronavirus-cases-in-singapore/>. Assessed December 14, 2020.