Computational Visual Media
https://doi.org/10.1007/s41095-023-0353-5 Vol. 10, No. 1, February 2024, 161–186

Research Article

A visual modeling method for spatiotemporal and multidimensional features in epidemiological analysis: Applied COVID-19 aggregated datasets

Yu Dong[1], Christy Jie Liang[1] (✉), Yi Chen[2] (✉), and Jie Hua[1]

Abstract The visual modeling method enables flexible interactions with rich graphical depictions of data and supports the exploration of the complexities of epidemiological analysis. However, most epidemiology visualizations do not support the combined analysis of objective factors that might influence the transmission situation, resulting in a lack of quantitative and qualitative evidence. To address this issue, we developed a portrait-based visual modeling method called *+msRNAer*. This method considers the spatiotemporal features of virus transmission patterns and multidimensional features of objective risk factors in communities, enabling portrait-based exploration and comparison in epidemiological analysis. We applied *+msRNAer* to aggregate COVID-19-related datasets in New South Wales, Australia, combining COVID-19 case number trends, geo-information, intervention events, and expert-supervised risk factors extracted from local government area-based censuses. We perfected the *+msRNAer* workflow with collaborative views and evaluated its feasibility, effectiveness, and usefulness through one user study and three subject-driven case studies. Positive feedback from experts indicates that *+msRNAer* provides a general understanding for analyzing comprehension that not only compares relationships between cases in time-varying and risk factors through portraits but also supports navigation in fundamental geographical, timeline, and other factor comparisons. By adopting interactions, experts discovered functional and practical implications for potential patterns of long-standing community factors regarding the vulnerability faced by the pandemic. Experts confirmed that *+msRNAer* is expected to deliver visual modeling benefits with spatiotemporal and multidimensional features in other epidemiological analysis scenarios.

Keywords visual modeling; epidemiological analysis; spatiotemporal; multidimensional; COVID-19

1 Introduction

The growing emphasis on the need for informatics and analytics in public health over the past 30 years has led to an increasing amount of investment in information systems [1, 2]. Visual modeling has played a significant role in big data analysis tasks in public health, especially in epidemiological analysis as connections between people grow. However, although many new tools and algorithms have already been developed to aid experts in analyzing and visualizing the complex data used in epidemiological analysis [3], numerous research challenges and gaps remain. As Carroll et al. [2] highlighted in a survey, complex epidemiological exploration requires novel visualization tools, and most existing visualizations applied in analysis tasks suffer from limited adoption. An increasing number of standard visualizations in epidemiology are evolving into visual modeling, which aims to combine multidimensional features, such as complexity, dynamism, and uncertainty, rather than simply spatiotemporal features [4, 5].

As a focal topic in epidemiology in the past three years, the global spread of COVID-19 has overwhelmed health systems and caused widespread

1 School of Computer Science, University of Technology Sydney, Sydney, NSW 2007, Australia. E-mail: Y. Dong, Yu.Dong-3@student.uts.edu.au; C. J. Liang, jie.liang@uts.edu.au (✉); J. Hua, jie.hua@uts.edu.au (✉).

2 Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China. E-mail: chenyi@th.btbu.edu.cn.

Manuscript received: 2022-12-05; accepted: 2023-04-20

social and economic disruption [6]. Because of the virus's high transmission ability, the responses of individual countries and communities to the unfolding pandemic will be a watershed moment in human history. Although a substantial body of work has been translated into COVID-19 data visualizations globally, there has been little prior work on visualization tools to conduct an analysis based on the characteristics of communities and risk factors of COVID-19. Such intricate analysis raises significant problems regarding pandemic-related uncertainty, which must be addressed [7].

To address these challenges, we optimized collaborative approaches to problem-solving using data analytics and subject matter expertise. We conducted interviews with domain experts in health and epidemiology to gather requirements for further analysis. With the assistance of experts, a visual modeling method, namely *+msRNAer*, was developed for spatiotemporal and multidimensional features in the majority of epidemiological analysis tasks. This visual modeling method adopts a portrait design inspired by viral anatomy and can be applied to visualize each community-based location, allowing for exploration and comparison of the complex relationships among communities' time-varying case numbers, objective risk factors that may affect pandemic, and epidemiological data patterns in fundamental geographic, timeline, and multidimensional visual designs. This creates a vivid understanding of how objective risk factors are interconnected and contributes more broadly to building resilient communities, particularly in light of the effects of pandemic transmission.

To validate the usability of *+msRNAer*, we collaborated with Australian Government experts to apply COVID-19 aggregated datasets with processes that (1) extracted evidence by each local government area (LGA) from the complete 2 years (742 days from January 1st, 2020, to January 11th, 2022) of COVID-19 case data compressed to 106 weeks or 53 fortnights for scalability; (2) added a marked timeline with intervention events extracted by natural language processing (NLP); and (3) connected the most recent census community profiles to each LGA, combining case data with LGA and postal area geolocations in New South Wales (NSW), Australia. We then identified risk factors related to LGAs under expert supervision, including demographic indicators (e.g., higher-risk populations), social indicators (e.g., relationships), economic indicators (e.g., rental and mortgage affordability), infrastructure indicators (e.g., housing), and resident travel behavior (e.g., using public transportation).

We adjusted the visual designs to include multiple views and interactions, which facilitated visual exploration in COVID-19-related community portraits. This was supported by an interactive control panel that included event timelines, a coordinated geographic view, and a multidimensional coordinate view with a filtering function. Our aim was not only to demonstrate the effectiveness and scalability of *+msRNAer*'s visual modeling but also to assist the government in investigating pre-existing community factors and discovering practical implications for potential patterns related to vulnerabilities against COVID-19. We conducted one user study and three subject-driven case studies using our COVID-19 aggregated datasets. These studies demonstrated how *+msRNAer* worked with a prototype among NSW LGAs and postal areas with spatiotemporal and multidimensional features. In the user study, we not only pre-evaluated the feasibility and effectiveness of the *+msRNAer* prototype but also made iterative improvements to its functions. Furthermore, the three case studies provided a selected, high-level picture of community resilience in infrastructure and explored the dimensions of resilience. We evaluated the COVID-19 exploration results and conducted interviews with domain experts to collect their feedback for future research.

The remainder of the paper is structured as follows. Section 2 reviews related work. We first discuss the design requirements in Section 3. Based on the extracted requirements for experts, we elaborate on the visual modeling with metaphor and design in Section 4. We describe the data preparation and prototype application with multiple views and interactions in Section 5. We describe our user study to pre-evaluate the feasibility and effectiveness of our visual modeling in Section 6. We further validate the capability of our visual modeling with applied COVID-19-related case studies in Section 7 and interviews with domain experts in Section 8. We discuss limitations and future work in Section 9. Finally, Section 10 concludes the paper.

2 Related work

In this section, we introduce related works concerning the visualization perspectives in epidemiological analysis research, and accordingly, discuss how visualization aids COVID-19 spread analysis as specific examples.

2.1 Visualization in epidemiological analysis

As early as 2020, a systemic review of COVID-19 epidemiology [8] proved high spreading speed when the epidemic broke out. Since then, an increasing number of visual representations were presented to aid COVID-19 analysis. A novel study [9] on computational modeling used visualization-centric and algorithm-assisted epidemiological modeling, proving that visualization plays a critical role in epidemiological analysis in 2022. Another ongoing collaboration [10] between epidemiological modelers and visualization researchers summarized concurrent challenges and solutions. They listed common visualization charts, such as heat maps [11, 12] and timelines [13], and further composite graphics with small multiple views, which could support epidemiological modeling. Wei et al. [14] surveyed and categorized geographic visual display techniques in epidemiology research into two categories of traditional cartography and geo-visualization.

From distinct angles of data analysis, certain risk features [15, 16] may influence the analysis of epidemiology and make the analysis tasks more challenging. Chui et al. [17] added human factors (age, gender, etc.) into the study of infectious diseases, showing that further analysis of visualization in epidemiology is beneficial and improves the precision of modeling and prediction algorithms [18]. According to our classification, there are two perspectives for visualization in epidemiological analysis depending on the type of data used: spatiotemporal-based and multidimensional-based.

Pandemics are geographical in nature and constitute spatiotemporal phenomena across large ranges of scales [19]. Improved geographic visualization plays an important role in pandemic research, as it offers an environment to represent multivariate data by cartographic means based on its geographical information effectively and attractively [20–22], and it is one of the top ten keywords of IEEE VIS [23] (top conference in the visualization field); also, 16% of existing related visualization works adopt maps [24].

Multidimensional-based data analysis makes it possible to connect other factors with epidemiological analysis. The Singapore epidemiology of eye disease research is a population-based study of 8697 adults of Malay, Indian, and Chinese ethnicity [25]. Steinger et al. [26] used generalized linear models to investigate the influence of key epidemiological factors on potato virus infection risk. A prototype has been deployed by Maciejewski et al. [27] to demonstrate the impact of social distancing strategies during the H1N1 (swine flu) outbreak. Trajkova et al. [28] analyzed relevant Twitter data and discussed facilitating data interpretation via visualization to avoid the spread of misconceptions and confusion on social media. Also, multidimensional data visualization, such as Parallel coordinates, is commonly employed for visualizing multidimensional geometry [29–32]. Visualization research on multidimensional attributes could be applied during a pandemic to promote the understanding of how data entries compare to each other.

2.2 Visualization of COVID-19 datasets

Since the COVID-19 pandemic's initial outbreak, a vast number of visual representations and models have been created to reflect the virus's global spread and the effects it has had on various nations and regions [11, 33–36]. A survey study [6] conducted 668 COVID-19 data visualizations to map the landscape of existing visual works. Another novel research [37] focused on investigating complex interplay based on COVID-19 datasets between design goals, tools and technologies, data information, emerging crisis contexts, and public engagement by a qualitative interview study among dashboard creators. There are two main types of data sources in common COVID-19 research: directly-linked data, such as infected cases, recovery, and mortality rates [38, 39], and indirectly-linked data, which contain community information, such as social impacts [40] and financial impacts [41, 42], which are not linked to the pandemic directly as objective factors.

In common COVID-19 visualization research, basic techniques include traditional line charts, bar charts, and maps. Kahn's report revealed that 38% of related works apply line/area charts, while 29% employ bar

charts [24]. We have collected 48 existing related research works to date in Australia; 10 of them deliver a similar dashboard view, and the others either offer graphs or are still ongoing works. The University of Melbourne conducts an online prototype that gives a 10-day forecast [43]. Seven projects import data from aspects such as financial and LGA details other than only the pandemic case details. Most works apply traditional bar, stacked bar, line, map. visualization methods, etc.

In our approach, visual modeling in COVID-19 datasets is separated into three aspects: data visualization, visual interactions, and intelligence-assisted visual analytics.

From the first aspect, raw data collected during the pandemic are applied as inputs to generate graphs. An interactive web-based dashboard [44] to track COVID-19 in real time was first presented by CSSE at Johns Hopkins University. This was followed by an online global dashboard by the WHO [45] to show COVID-19 statuses around the globe. Mathieu et al. [46] built 207 country profiles with aggregated cases, testings, vaccinations, etc., which allow users to explore the detailed statistics on the COVID-19 pandemic. In Australia, the State and Territory governments assist the public with recognizing current statuses by visual dashboards [39, 47–49].

From the second aspect, visual modeling methods can combine indirectly linked data and offer deeper insights by interactively integrating multiple attributes. Dashboards with adjustable parameters are a common feature from this perspective; they combine computational analysis techniques with interactive visualizations [50–57], which also emphasize analytical reasoning concerning the pandemic data and other factors that may affect infection cases or are affected by the pandemic through interaction techniques. Leung et al. [58] presented a big data visualization and visual prototype for analyzing COVID-19 epidemiological data. Besides, increasing research has included objective factors in the COVID-19 analysis. Lan et al. [59] reviewed the usefulness of GIS-based dashboards for mapping COVID-19 prevalence and proposed improved geo-visualization techniques to incorporate the temporal component in interactive animated maps to analyze pandemic transmission with COVID-19-related information. Wu et al. [60] developed a novel Joint Classification and Segmentation (JCS) system to perform real-time and explainable chest CT diagnosis of COVID-19 infections. Muto et al. [61] imported more facts about gender, age, marital status, poverty, and drinking/smoking habits into a matrix to address Japanese citizens' behavioral changes and preparedness against the outbreak.

From the third aspect, intelligent techniques are introduced to assist in the exploration and investigation of the COVID-19 pandemic. This takes into account more relevant yet not directly-linked complex data, which includes modeling [62–65], prediction [66–69], and other complex exploration strategies [70, 71] with AI algorithms [72]. Reinert et al. [73] developed a framework that enables effective and efficient visual exploration through interactive, human-guided analytical environments during the pandemic. Afzal et al. proposed a decision-making environment [74] for person-to-person contact modeling in the COVID-19 pandemic, which was based on their previous works [27, 75] in epidemiology. Bowe et al.'s research [76] indicated that the pandemic plays out differently across different scales; this is related to the global supply chain, local dynamics, neighborhood mutual aid networks, and personal geographies of mitigation and care. Preim and Lawonn [77] described visual analytical solutions aiming to provide preventive measures, where prevention aims at advocating behavior and policy changes likely to improve human health. Another research proposed a framework of four prediction methods that for anticipating pandemic viral attacks and how far they expand globally [78]. Guo's system discovered spatial interaction patterns, providing valuable insight into designing more effective pandemic mitigation strategies and supporting visual exploration in time-critical situations. An approach by Healey et al. [79] was conveyed through visual exploration by similarity comparison and predictions. Dong et al. [71] provided a user-centered visual explorer that applied COVID-19 datasets for in-process exploring and comparing spatiotemporal features in portrait-based perspectives.

3 Design requirements

In 2022, a survey conducted by Dykens et al. [10] summarized the challenges, solutions, reflections, and recommendations of visualization for epidemiological modeling. The authors categorized the supporting

visual modeling in epidemiological analysis into three stages based on different time scales.

The initial stage involves the quick application of candidate templates with visualization tools to establish problems. Data are transferred to preset combinations of views for simple comparison.

In the short- to mid-term, ongoing research provides a redesign of the visualization prototype for iteratively redefining the problem, exploring potential patterns, and providing users insight into complex tasks.

In the long-term, a more stable visual system is developed for widespread application to common usage scenarios, as demonstrated through multiple cases in epidemiological analysis.

This summary underscores the motivation for our proposed visual modeling method in epidemiology to fulfill multiple objectives. The visual modeling method for epidemiological analysis should not only enable quick responses to basic information trends of a pandemic, such as the infection cases in time-varying trends, but also facilitate the discovery of knowledge concerning combined analysis tasks, such as identifying the factors that affect infection cases. Finally, the visual modeling method must be applied to a completed visual system and validated using real-world cases.

In consultation with experts from the epidemiology and health domains, we outlined the following user requirements: Experts required information to assess the profile of communities in terms of their resilience to virus attacks. Combined with the inspection of the number of infected cases, a few key factors might affect and help understand the impacts of different pandemic phases occurring concurrently. They must assess the impact of the government interventions and measure the resulting pandemic situations to investigate both the community profile and infection cases concurrently. They needed to investigate the link between community factors and the number of infection cases caused by outbreaks, intervention events, and responses. Usability studies conducted with early existing prototypes identified a variety of requirements. We distilled the desired features into three progressive categories of peer-to-peer design requirements for our approach: visual design, visual analytics, and modeling prototype application during several design iterations with subject matter expertise.

R1. Provide comparative visual portrait design of the numerical distribution of consecutively transmitted cases for each community. To perform effective epidemiological analysis, it is essential to describe consecutive infection cases in terms of timelines and locations. This involves investigating the geographic and temporal trends of the epidemic's spread, along with qualitative analyses utilizing spatiotemporal features. By providing comparative visualization portraits of these aspects for each community, we could better understand the patterns and trends of transmission.

R2. Offer visual exploration for analyzing transmission patterns with spatiotemporal and multidimensional features among each community. Epidemic outbreaks are often related to various objective factors in specific locations, such as socioeconomic or cultural factors. Therefore, it is crucial to support multidimensional feature exploration, including humanities and finance, among other fields, in addition to visualizing the spreading situation with spatiotemporal features. Combining both perspectives in interactive portraits of location-based risk factors and infection cases over time can provide a more comprehensive understanding of transmission patterns.

R3. Verify the effectiveness of visual modeling through a prototype system using actual epidemiological cases. To demonstrate the effectiveness of our visual modeling approach, the prototype system should offer multiple visual views with robust interactive functions, such as collaborative filtering and comparison. By applying this system to real user studies and case studies in epidemiological tasks, we can show how our approach can help researchers gain insights and make informed decisions in the field.

4 Visual modeling

To address the design requirements, this section presents our design rationale through visualization, design considerations, and guidelines for the components.

4.1 Design metaphor

Inspired by viral anatomy, our visual design primarily adopts the 2D genome structure of SARS-related

coronavirus particles. Coronaviruses, named for their "crown-like" shape observed under electron microscope, have particles packed with shells. This illustration became familiar to the public, as shown in Fig. 1, and is widely introduced by the news and media. Coronavirus particles are organized with *+sRNA* (positive single-stranded RNA) polymers packed inside, further surrounded by outer inserted proteins. These outer proteins are derived from the cells in which the virus is last assembled but are modified to contain specific viral proteins, including the Spike (S), Membrane (M), and Envelope (E) Proteins. The S Protein allows viruses to enter and infect other cells. After the virus enters the host cell, the genome is transcribed, and replication takes place involving coordinated processes of RNA synthesis. Our proposed visual modeling method is referred to as Positive Multiple Strands of RNA Encoder (*+msRNAer*). Inspired by this simulation and combined with our previous research [71, 80], we map different parts of the virus into multiple meanings and design *+msRNAer* to apply the aggregated datasets of its inner and outer parts.

4.2 Visual design

We depict a novel visualization that leverages the biological components of the coronavirus as metaphors to represent and compare communities' epidemiological characteristics. Specifically, we define the particle symbol for each community, which we call a "Portrait", and designed the outer proteins and inside RNAs to encode information related to cases (such as actual cases and cases per 10k/100k population) and transmission trends and cases of each community's unique risk factors, respectively. Figure 2 shows the detailed design for all components. Notably, we varied the RNA

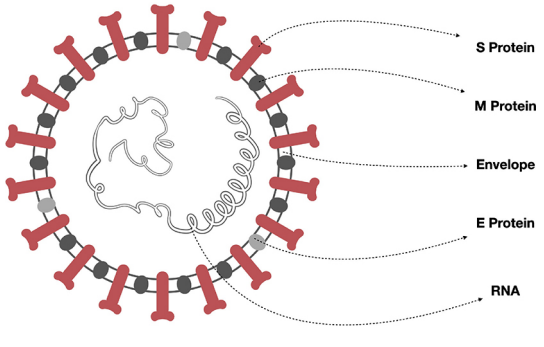

Fig. 1 Cross-sectional simulation of SARS-related coronavirus with its main components.

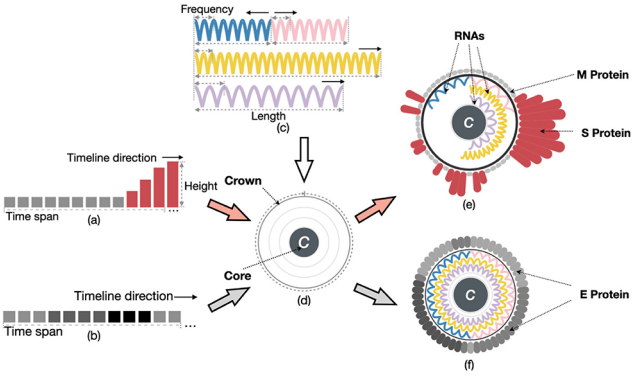

Fig. 2 Two types of visual portrait and their implementation processes: (e) Portrait with S and M Proteins and (f) Portrait with E Proteins. Both Portraits (e) and (f) include 4 strands of RNA. The red arrow path represents the process of (a) combining bars on the timeline and encoding them as S and M Proteins, while the light gray arrow path represents the portrait with (b) E Protein in grayscale, depicting different phases of the timeline.

from positive single strand to positive multiple strands to capture multiple key risk variables related to community characteristics. This design choice provides multidimensional insight into the epidemiological patterns of each community and aims to fulfill R1.

As domain experts suggested, the number of infection cases caused by the pandemic is related to multiple potential factors among communities, e.g., population and its percentage of higher-risk groups, including the aged, lower-income, and lone-person groups. As a result, we could emphasize community portraits based on the characteristics of the vulnerable population in high-risk areas. Hence, we assemble all visual elements as the portrait for each community, as shown in Fig. 2. This single aggregated crown-like portrait describes both the community's existing characteristics and also how the community reflects viral infection case number information. The visual portrait consists of three components: the crown, outer designs with proteins, and inner designs with RNA. The circle of a crown represents one unit of the whole timeline. On the outside, we use multi-segments as time spans, with the S Protein representing case number information, the M Protein representing zero cases, and multiple RNAs representing the highlights of selected factors in higher risk groups inside the crown. The E Protein, unlike the S and M Proteins embedded on the crown, can represent significant intervention events rather than case numbers.

Crown design. We define the central core with

radius R_c of the Crown to label the community C, as shown in Fig. 2(d). The circular loop surrounding the Crown easily depicts the entire timeline of any transmission taking place. The circle is evenly divided into continuous segments matching time spans for S or M Proteins, clockwise from the top point to the looped end. Further, we use E Proteins to denote the categorized intervention events within the timeline, colored in grayscale.

Outer design. The outer-growing S Proteins are densely packed together. For each community over each time span, we encode distinct bars with rounded corners, with their heights representing the case numbers shown in Fig. 2(a). We use the smallest-sized M Proteins to indicate zero cases in a certain time span. The S or M Protein heights of time span x are calculated as Eq. (1):

$$H_{\text{timespan}_x}=\begin{cases} h+R'_c*(a+\ln f(x))*b, & f(x)\in N^* \\ h, & f(x)=0 \end{cases} \tag{1}$$

A base height h appears when there are 0 cases in one time span, and growth height is proportional to the number of non-zero infection cases in a time span x, where $f(x)$ is the function of infection cases corresponding to this time span. Both a and b are customized parameters, and R'_c refers to the initial radius of the Crown.

Inner design. Although viruses normally only carry one strand of RNA, it is not ideal to employ only one strand to represent several important elements. We suggest splitting the RNA strand into numerous strands to show these parameters, which conserves space and reduces unnecessary requirements. Thus, we propose to distribute the four RNA strands across three channels rather than joining them head-to-tail. Each element's maximum value is set to occupy half of a channel when multiple data elements of the same type must be represented, such as visualizing the numbers of males and females (exclusive of transgender people). These two can then be combined and allocated to a single channel. Other whole channels may be assigned to data of a single type. Figure 2(c) demonstrates the potential of our visual design to share one channel with two RNAs. We implement three channels of RNA that grow from the middle point of the circle. Combining two related factors in one shared channel accommodates all four RNAs placed in three oriented channels within the Crown.

To fulfill the visual design, we employ a visual metaphor of spiral genomes with four cosine wave-shaped RNAs with the same amplitude in the three channels. As shown in Fig. 2, the length of RNA is encoded by the exact values (e) of its categories.

To denote the length of RNA L_{ij} in the arc, we define

$$L_{ij}=\left[R_c+(m-0.5)*\frac{R'_c-R_c}{3}\right]*\left(\frac{N_{ij}}{\max\{N_i\}}*\frac{2\pi}{n}+\frac{\gamma}{n}\right) \tag{2}$$

i.e., the arc angle θ_{ij} corresponding to the arc length L_{ij} is

$$\theta_{ij}=\left(\frac{N_{ij}}{\max\{N_i\}}*\frac{2\pi}{n}+\frac{\gamma}{n}\right)*\frac{2\pi}{2\pi+\gamma} \tag{3}$$

Equation (3) employs a rescaling that ensures the maximum will not exceed the current RNAs' located channel lengths. N_i represents the set of data from the independent variable RNA category i, and $\max\{N_i\}$ denotes the maximum value in N_i. N_{ij} refers to the value of data from an independent variable community j in an independent variable RNA category i. Parameter $m\in\{1,2,3\}$ allocates the exact location in the channel of the current RNA category i. Parameter γ is the minimum arc angle in RNAs, maintaining RNA even if the value of N_{ij} is very small. Parameter $n\in\{1,2\}$ indicates whether the maximum length of RNAs occupies a full or half channel. We define Θ_{ij} as the maximum arc angle with $\Theta_{ij}=\dfrac{2\pi}{n}$ when $N_{ij}=\max\{N_i\}$. Therefore, any $\theta_{ij}\in[0,\Theta_{ij}]$.

We further define the cosine wave function F_{ij} for drawing each RNA as

$$F_{ij}=\frac{R'_c-R_c}{3}*\left[|\cos(\theta_{ij}*\frac{N_{ij}}{N_j})|+(m-1)\right] \tag{4}$$

Equation (4) is simplified by the designed function $F_{ij}=\dfrac{R'_c-R_c}{3}*|\cos(\theta_{ij}*\dfrac{N_{ij}}{N_j})|+(m-1)*\dfrac{R'_c-R_c}{3}$. This equation is assembled from two parts: one draws the cosine shape with an absolute value function, while the other is used for radial translation. The parameter in Eq. (4) applies $\dfrac{R'_c-R_c}{3}$ as amplitude after the absolute value is calculated, N_j as the sum value of data from community j, and the frequency in cosine waves $\dfrac{N_{ij}}{N_j}$ represents the ratio of the current value from community j in RNA category i divided

by the sum of community j. The above RNA design allows for the comparison of different communities within the same RNA category from two distinct perspectives. The first involves comparing current categorical values among multiple communities based on RNA length, while the second relates to comparing the ratio of current values with the total value in the same community, as determined by RNA frequencies.

As an illustration, the RNA for selected factor *income* from community *Sydney* should be mapped when allocated in the second channel: $F_{\text{income,Sydney}} = \frac{R'_{\text{c}} - R_{\text{c}}}{3} * |\cos(\theta_{\text{income,Sydney}} * \frac{N_{\text{income,Sydney}}}{N_{\text{Sydney}}})| + \frac{R'_{\text{c}} - R_{\text{c}}}{3}$.

Filter trigger design. All the portraits are needed to motivate a filter trigger. We reset the Crown and created a sample portrait with E Proteins for filtering purposes, which can be used interactively to attach events to the timeline. On the outer circle, the grays indicate different events by timeline. In the inner Crown, three full-circled values of RNAs, as shown in Fig. 2(f), are the indicators for selected factor groups. During exploration, we intend to interact with the visualization by interacting with all these elements in the Control Panel.

Design for all colors. Two sets of color scales for Control Panel and Portrait View are used in our visual design. To raise awareness of the threat, we encode bright red for the S Protein design, encoding the infection case numbers, and light grey for the M Protein, indicating there were no cases this week. Inside the Crown of the portrait, inner color scales are used to representatively paint community portraits on RNAs, which include azure blue, mint pink, gold yellow, and pale purple. In the Control Panel, a pre-defined grayscale is designed for the E Protein to differentiate the types of intervention events along the timeline, initially with normal gray, silver gray, and dark gray. A darker shade of gray indicates a higher level of restriction for the events. For interaction, charcoal gray is used in both the Control Panel and portrait design for selection interaction.

5 Data preparation and prototype application

We further developed a web-based visual prototype based on *+msRNAer*, which aims to assist in investigating the pre-existing community factors and discovering practical implications for potential patterns of established community characters against the vulnerability faced by epidemiological analysis.

Although *+msRNAer* should be suitable for most epidemiological analyses with community factors, we introduced the aggregated COVID-19 datasets in NSW as hot topics from epidemiological analysis to better present the design ideas in Section 4. We then applied *+msRNAer* to the prototype implementation by each view and introduced the interaction designs in this application.

5.1 Data sources

We investigated whether there are any patterns or potential relationships between NSW COVID-19 cases and demographics, geo-information, infrastructure information, or other factors in the census fields, as clarified by the requirement of analyzing COVID-19 situations. Multiple datasets were aggregated to demonstrate the effectiveness of our approach, which includes COVID-19 case data with event timelines and selected columns of NSW census data.

The COVID-19 case data in NSW were collected by the government, and the pandemic data program led by the Data Analytics Centre (DAC) provides digital information to improve the coordination of the government's COVID-19 response. The datasets provided applicable information about infection cases based on the location of usual residence since the first infection case; they excluded 189 cases in crew members who tested positive while onboard a ship docked in NSW at the time of diagnosis; and case aspects include confirmed, tested, recovered, and deaths by their notification date, location, age-range, and likely source of infection. Some of them were no longer released due to privacy. Plus, the GIS map data and event information were extracted from media releases [81] and NSW Property Web Service [82] authorized by NSW government websites, respectively.

We decided to identify the high-risk factors that may contribute to COVID-19 infection from the census data, which involves millions of people and households and is conducted by the Australian Bureau of Statistics (ABS) every four years. The data provide a rich snapshot of the nation and inform

the government, communities, and businesses. It contains essential concepts, such as populations, rents, mortgages, incomes, religions, languages, and housing. Besides, based on the definition from ABS, a lone person is classified as the only person aged 15 or over who lives in a private dwelling. Also, people in NSW earning more than 50% but less than 80% of the NSW or Sydney median income are described as earning a lower income [83]. The next wave of census data was released in stages since June 2022, but detailed information will not be released until mid-2023. Thus, the latest census data from 2016 were applied in this paper.

5.2 Variable consideration

The COVID-19 case dataset delivers every case recorded by NSW Health and contains multiple attributes concerning cases by notification date and postcode, local health district, LGA, and likely source of infection. Considering the risk of leaking information that could directly identify individuals, only personal age, gender, and location of their usual residence are included in this dataset, which is assessed to measure the risk of identifying an individual and to measure the information gained if it is known that an individual is in the dataset. LGA is an official spatial unit that contains multiple postal areas that represent the whole geographical area, and there are 128 LGAs in NSW in Australia (Bayside Council was formed on September 9th, 2016 after the 2016 census, and relevant LGA data are merged from the City of Botany Bay and Rockdale City Councils). Moreover. As data journalism may affect the COVID-19 pandemic [84], relevant news articles, alerts, and ministerial media releases issued by the NSW government about COVID-19 are attached as events to combine with the notified date of infection cases.

The timeline selection was intercepted during the COVID-19 pandemic, from January 2020 to January 2022. We used NLP [85] to extract textual information from media data and considered marking the timeline related to keywords in interventions and social restrictions as three types of phases: uncontrolled, eased (e.g., keeping social distance, masks required), and restricted control (e.g., curfew, bubble restriction, lockdown), and the marked timeline was attached to the COVID-19 case dataset for two data periods: the long period used 53 fortnights of COVID-19 case data as a biweekly time span grouped by 106 weeks of data from January 1st, 2020, to January 11th, 2022; the short period used 53 fortnights of case data as a weekly time span from January 1st, 2020, to January 5th, 2021. Further imported JSON files combined a single LGA layer for the long period and an LGA layer with postal areas for the short period of COVID-19 case data.

We further consulted domain experts and selected the four top key factors from a suite of factors by their supervision that may cause infection in their communities (LGAs and postal areas) as directed: males and females aged 70+, lower income groups, and living alone groups. Experts extended four additional categories of indicators from censuses that may affect the COVID-19 pandemic: LGA-based demographic indicators, social indicators, economic indicators, infrastructure indicators, and resident travel behavior. Each category contained over 30 indicators and descriptions. Some indicators contained complex hybrid patterns that may influence transmission. For instance, apart from living alone, the indicator of household is defined as different from a family, which refers to at least one person over the age of 15 who lives in the same private dwelling. In this situation, it is necessary to deliberate on the comprehension of demographic, social, and economic indicators. We extracted the data on median age, population, and area size from LGA demographic information and further calculated population density; median rent expense, median mortgage, median personal income, median family income, and median household income from social and economic indicators; average bedrooms per person and average bedroom size per household as dwelling factors from the infrastructure indicator; and public transportation rate for traveling from resident travel behavior.

Finally, we had 96,460 rows of COVID-19 case data up until January 11th, 2022, and aggregated both datasets with intervention events and spatiotemporal features as well as simplified objective factors from the NSW census data tables of size 449 MB.

5.3 Prototype implementation

The prototype was implemented to meet requirements that offer extra visual explorations of interactive community portraits. For common epidemiological analysis tasks, we assembled the proposed *+msRNAer* prototype with collaborative visualization views, as

shown in Fig. 3, including the Control Panel (D), Portrait View (E), Geographic View (F), and Multidimensional Coordinates View (G), implemented based on D3.js [86] and Mapbox.js [87] as a base map.

For better illustration, we applied our data preparation to the NSW aggregated COVID-19 dataset. The entire workflow of the applied *+msRNAer* prototype is demonstrated in Fig. 3. In this section, the *+msRNAer* prototype is introduced as being integrated into the pipeline workflow with summarized human-participated actions and data flow directions by steps, which include collecting and dealing with raw datasets, processing specific data with NLP, extracting information under expert supervision to the aggregated COVID-19 datasets, and applying *+msRNAer* to the prototype system with several functions and interactions finalized.

Control Panel and Portrait View. The Portrait View is built containing a Control Panel based on our visual modeling method, with a force-directed layout to avoid overlapping and improve readability. First, we can select the applied datasets and inspection of modes—either actual cases or cases per 10k population—in the Control Panel. Then, we can hover over or click on each time span on the sample portrait to interactively observe the corresponding filter result on each portrait and inspect all color tooltips among each view.

We can further inspect, explore, and compare infection cases and community key factors for each in the Portrait View, with aggregated case numbers by time span on the Crown and overviews of selected risk factors in the inner. We fix the Portrait View as the main view and set each portrait with the first channel with azure blue for older males and mint pink for older females, the second with gold yellow for lower-income groups, and the third inner channel with pale purple for lone people groups.

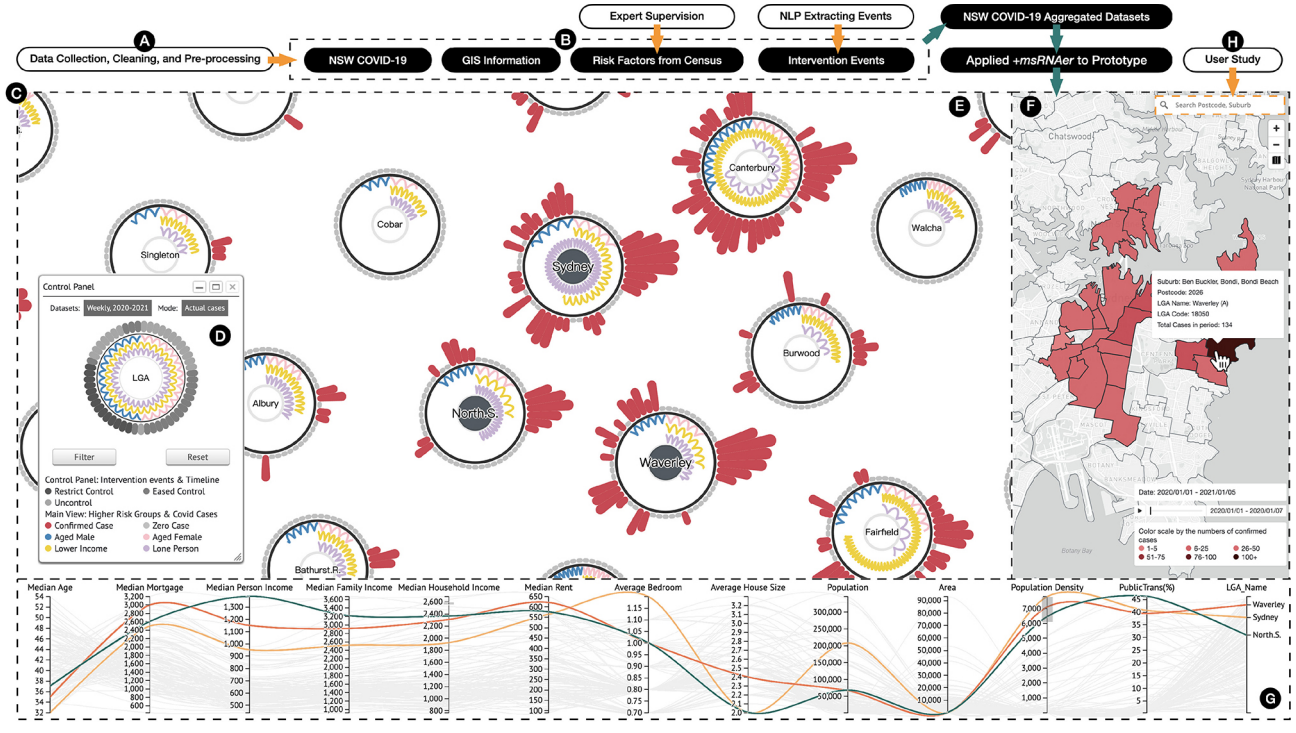

Fig. 3 Entire workflow of the visual modeling method *+msRNAer* applied to NSW COVID-19 aggregated datasets, from raw data collection to prototype system, which embodies one completed pathway. It consists of two types of arrows, where the orange color connects the human-participated actions, and the cyan color represents the data flow direction. The pathway starts from (A) data collection and processing, (B) dealing with datasets of summarized COVID-19 cases, GIS information, LGA-based censuses under expert supervision, and intervention events extracted by NLP, and assembling the aggregated datasets for application in *+msRNAer*. (C) The prototype interface with (E) Portrait View depicts four high-risk factors and COVID-19 cases corresponding to the selected timeline in (D) Control Panel and interacts with (F) and (G). (F) The GIS View with highlighted polygons of LGAs and postal areas shows their COVID-19 distribution. (G) The MDC View is a high-level overview of other risk factors from the census indicators. This *+msRNAer* prototype is finally completed by a conducted user study by adding a search function (H) in GIS View.

GIS View. We added extra visual polygonal layers to the Mapbox-drawn landscape to divide different LGAs and postal areas in NSW. When an LGA or postal area is selected, it will be highlighted with a charcoal gray polygonal boundary and will display detailed geographical information and infection cases. Using the GIS View, we can explore geographical information by zooming in on postal areas within the LGA and zooming out for an overview. The GIS View also provides a playable timeline window to enable the inspection of the spreading situations in selected time periods. The visual design of the GIS View utilizes a red gradient scale. The darker the red, the more infection cases are represented in each LGA.

MDC View. We utilized parallel coordinate-based visualization for multidimensional variables. The MDC View aims to show a high-level overview of other factors related to community resilience and allows users to explore the dimensions of resilience. Each LGA polyline is distinguished by coordinates, different colors, and different styles. Polyline uses different dash styles or colors according to the positive proportion of infection cases.

5.4 Visual interactions

We can interact with multiple views, including Portrait, GIS, and MDC Views, triggered by Control Panel. We offer flexible interactions within and among multiple views collaboratively on this visual prototype system for each visual exploration from multiple perspectives, summarized as follows.

By combining the Control Panel and Portrait View, we implemented several interactions that enable efficient navigation through the LGA portraits and visual cues within the core for further exploration and comparison.

Filtering by time spans. The interactive Control Panel serves as a starter to filter LGAs in all visual views and by time spans based on intervention events and to address all color legends in the Portrait View. Users can independently inspect the COVID-19 cases on the S or M Protein among each LGA portrait based on the gray-scale mapped on the E Protein, a time period of intervention events, or a time period before and after the restriction events.

Zoom and pan. An interactive portrait based on SVG supports zooming and panning for exploring overviews or detailed visual elements.

Drag and reposition. The force-directed layout ensures that each portrait does not overlap. Users can drag and lock any portrait to a new location to allocate any selected LGA for comparison. In addition, we enhance interactions for further functionality by left-clicking to mark the cores and canceling their current location by double-clicking.

Highlight visual cues and switch contexts. When hovering the mouse over each type of visual cue in the Portrait View, corresponding visual cues and associated data information tooltips will be highlighted for further inspection.

Conjunction with filtering interaction. As shown in Fig. 4, dynamic highlighting of ranking or case numbers in the core is supported based on the cases or prevalence rate with the average cases per 10k population in the previous filtered timeline added with the hovering S Protein's height. Alternatively, users can hover only on RNA to highlight its ranking among all LGAs.

The GIS and MDC Views collaborate with the Portrait View by highlighting both the selected LGAs and postal areas. Some other interactions are also supported.

Playable timeline window. The GIS View provides an interactive timeline window that allows users to investigate the transmission situation over different time spans. It includes several functions, such as auto-play and pause, enabling users to customize their explorations.

Boundaries highlighted. The GIS View supports clicking to highlight boundaries and reflect

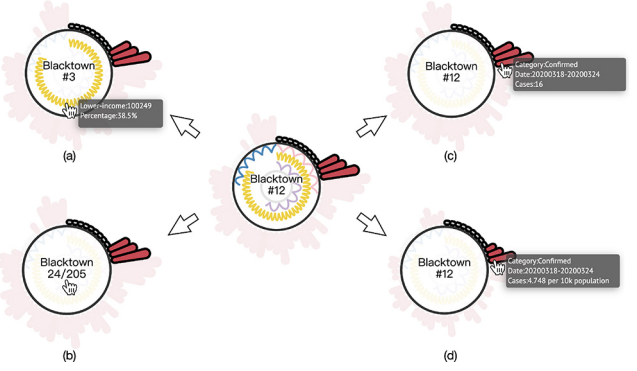

Fig. 4 Four highlighting strategies are displayed when hovering over the (a) ranking and tooltip of the selected RNA among all LGA portraits, (b) sum of case numbers in the filtered phase with the total cases in the entire period, (c) ranking and tooltip of case numbers summed by hovering over any time span, and (d) time span context is switched to case numbers per 10k population.

LGA portraits in the Portrait View. It also supports connecting with the cores in the LGA selection.

Heatmap highlighted. The GIS View offers a heatmap layer in LGA or postal areas by timeline filtering in the Control Panel. The colors reflect the number of COVID-19 cases in the selected time period.

Brush on MDC View. In the MDC View, the brush function is used to filter the portrait factors in multiple dimensions and reflect them to other views.

6 User study

To ascertain the viability and effectiveness of the *+msRNAer* prototype, we conducted a user study before conducting the case studies. In this section, we present the details of the study setup and analyze the obtained results.

6.1 Participants and apparatus

In our user study, we endeavored to recruit a diverse group of participants with varied backgrounds and levels of research experience in the field of computer-related disciplines. Ultimately, we successfully recruited 16 volunteers from our campus, comprising an equal representation of 8 male and 8 female individuals. Notably, the age range of our participants was wide, ranging from 19 to 30 years old, with a mean age of 24.06 years old. Additionally, we verified that 5 of our participants were enrolled college students without any prior research experience, while the remaining 11 were postgraduate candidates with research experience ranging from one to eight years, resulting in a mean of 1.88 years of research experience. Despite their shared interest in visualization, all participants reported being unfamiliar with visualization methodologies.

All user studies were planned to be presented in the campus study pods. Our *+msRNAer* prototype was supported by an Apple MacBook Pro (15in, 2018) equipped with 16 GB of memory, i7 processors, and a Radeon Pro 555X Graphics Card, allowing participants to visualize and interact with the *+msRNAer* prototype clearly and effectively on a 60-inch LED external monitor with 1920 × 1080 resolution.

6.2 Tasks

The tasks included in our user study incorporated quantitative and qualitative analysis. The quantitative analysis provided objective numerical data, while the qualitative analysis allowed us to gain subjective insights into participants' interactions, aiming to verify the feasibility and effectiveness of *+msRNAer* prototype.

To enhance clarity in the quantitative analysis, we established specific exploration tasks for each view as well as collaborative tasks among the views to test the exploration capabilities of the *+msRNAer* prototype. By recording the completion time of each task, we conducted analyses to determine whether the *+msRNAer* prototype improved the exploration capabilities for COVID-19 case trend and risk factor portraits and whether they had positive impacts from a visual metaphor perspective. Additionally, we planned to gather participants' feedback on their interactions with the prototype as a whole and each view, using a set of six well-designed questions to assess their personal experience. Figure 5 lists

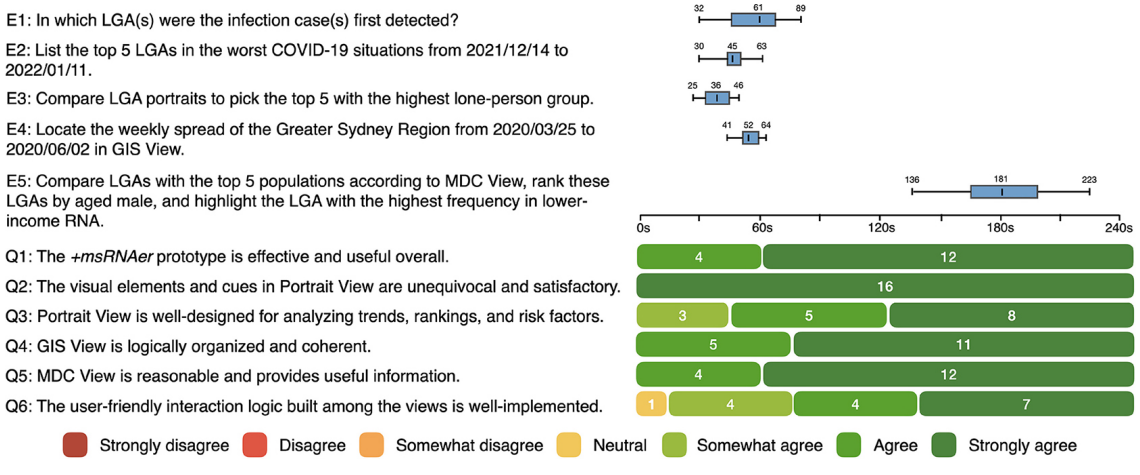

Fig. 5 Quantitative analysis results of the exploration task list (E1–E5) and questionnaire (Q1–Q6), where exploration tasks record the completed time in a box plot and the questionnaire counts participants' choices in stacked bars.

specific exploration tasks and detailed questionnaires. Besides, we planned to record their feedback in transcripts via interview for qualitative analysis.

6.3 Procedure

After setting up the exploration tasks and questionnaires, we rehearsed a tutorial on how to use *+msRNAer* for exploration, repeated the five exploration tasks, and recorded completion times. We found that each exploration task could be completed within four minutes. Although we conducted the design of *+msRNAer*, the randomness of a force-directed layout in Portrait View did not provide significant benefits in terms of reducing completion time. In other words, every participant could finish each exploration task within four minutes after tutorials. Therefore, we decided to allocate an approximately 60-min face-to-face session for each participant, consisting of a 10-min tutorial, an approximate 20-min exploration, a fixed 10-min preset questionnaire, and a 20-min open-ended interview. Participants were instructed on how to apply the *+msRNAer* prototype and were required to complete five specific exploration tasks, with completion time recorded. They were also asked to complete the questionnaire based on their subjective experiences during prototype usage. Additionally, their complementary feedback was recorded and used for subsequent qualitative analysis.

6.4 Results

In this section, we discuss the quantitative and qualitative results of our user study.

Quantitative results. The quantitative results of our user study were reflected in two aspects. The first aspect was task completion time. We observed that the fluctuations in completion time across these five tasks were within reasonable ranges. The median time to complete the first four exploration tasks varied from 36 to 61 s, while the last, more complicated task took a median of 181 s to complete. This consequence could be easily explained by the fact that the completion time of tasks E2–E4 was slightly reduced as participants became more familiar with the prototype during the first exploration task. Additionally, the first four exploration tasks focused on simple tasks that could be completed within one or two views. In contrast, the last exploration task required participants to engage with all collaborative views and gradually decipher three interactive results. Consequently, this task demanded more time and attention, leading to a higher median completion time. These results are in line with our expectations, as all tasks were completed within four minutes.

According to the results of the questionnaire, the majority of participants expressed satisfaction with the *+msRNAer* prototype. Specifically, 12 out of 16 participants strongly agreed with its performance, while the remaining four were satisfied with it overall. All participants were highly impressed with the prototype's visual design in Portrait View and expressed varying degrees of satisfaction with its analysis functions, including trends, rankings, and risk factors. Two later questions further highlighted the exceptional integration of both GIS View and MDC View in the prototype. The final question evaluated the implementation of interaction logic, with one participant responding neutrally and the other 15 participants expressing varying degrees of agreement that the interaction implementation was user-friendly.

The quantitative results from these two aspects indicate that participants rated the feasibility and effectiveness of the applied exploration tasks and prototype designs relatively highly. They also found the visualization and interaction designs to be intuitive and impressive.

Qualitative results. As an extension of the quantitative results, we further employed qualitative analysis during the last 20-min open-ended discussion. In particular, we conducted opinion research for individuals who did not provide the most satisfactory options on the questionnaire. To derive the qualitative results, we repeatedly reviewed the interview recording, summarized their critical thinking, and achieved a consensus based on their complementary feedback during the interview.

Their feedback was mostly positive with high marks, with only a few critical comments. These comments can be mainly divided into two aspects:

(1) A few participants felt that the current force-directed layout of Portrait View was a "double-edged sword". On the one hand, it provided dynamicity to avoid overlapping and allowed for easy dragging and relocating, but on the other hand, its randomness may increase the workload of recognizing specific portraits in multiple explorations.

(2) Some participants who marked Neutral or Somewhat Agree in Q6 expected to see more interaction logic among views, such as enabling the inclusion of more risk factors in Portrait View or providing solutions for swapping key risk factors with other factors in MDC for higher levels of comparison.

We felt grateful for their feedback as it served as a pre-evaluation before case study on domain experts and helped us improve this applied prototype. Regarding the comments on the applied force-directed layout, we believe that its advantages outweigh its disadvantages, but we still consider it necessary to improve it. As a result, we implemented the search function in GIS View to filter any LGAs or postal areas, which also strengthened the interaction between GIS View and Portrait View to a certain extent.

Because we are proposing a visual modeling method for epidemiological analysis, any other related applications using *+msRNAer* in prototypes may vary in detail. In this situation, with the prior requirements that we consulted with domain experts, we deemed that locking the highest risk factors in each portrait is acceptable because aged, lone, and lower-income groups have been proven by domain experts to be high-risk factors, while other objective factors selected from the census were considered to be indirectly related categories. Thus, we explained the reasoning to the participants and derived their understanding and acceptance.

Overall, our user study showed that the design of *+msRNAer* is creative, and the application of *+msRNAer* with the COVID-19 aggregated dataset was also proven to be feasible and effective, although there were some imperfections in certain details. According to the user study, we not only improved *+msRNAer* with search functions in GIS View but also inspired some interesting cases that were demonstrated in future subject-driven case studies.

7 Case studies

Coastal areas are considered densely populated areas prone to cluster infections for transmission, in contrast to the vast land and sparsely populated areas of NSW, Australia. For other incidents, intermittent spreading cycles of virus variants, corresponding policies or restrictions, densely populated residential areas, areas where older people gather, more impoverished areas, or areas with relatively poor public infrastructure may affect the infection situation among different LGAs. In this section, we finalized our *+msRNAer* prototype from two different data aspects based on the COVID-19 aggregated dataset in NSW, which consists of a weekly case summary of each postal area within an LGA in one year and a bi-weekly case summary of LGAs in two years, to provide three prominent cases: overview-driven, event-driven, and portrait-driven cases, which are based on highlighted driving aspects to compare and explore the significant connections between COVID-19 issues and detailed factors in each LGA census, even attempting to discover the relationships and potential patterns that were really affecting the COVID-19 pandemic behind each LGA portrait. We also combined our findings with facts and news for verification and analysis.

7.1 Overview-driven cases

We first set up a visual representation for the LGA portraits within the period from January 1st to January 11th, 2022. Connected to the GIS View, it showed the severity distribution of COVID-19 cases in geo-polygons. With a quick glance in Fig. 6(a), we could distinguish whether the situations are severe or not with roughly two types of portrait appearances. The results are also correspondingly reflected on the map, showing the contrasting perspectives between coastal and inland areas in case numbers.

When comparing one LGA portrait to another, one can examine variations in the COVID-19 cases spreading situations among LGAs. The S Proteins of each LGA portrait showed the changing trends of COVID-19 cases by height. The only two LGA portrait types that can be distinguished based on appearance are those with COVID-19 cases in both 2020 and 2021 and others with cases only in 2021 but none in 2020.

We randomly sorted a few portraits of each type for analysis, as shown in Fig. 6(b). The second visual result depicts the trends of COVID-19 cases within an 8-LGA portrait collection. We gained insights demonstrating that they all suffered relatively worse situations and caused two typical waves of cases from mid-year to the end of 2021. Compared to the height of S Proteins in each portrait, they reached the peak of the first wave in September, dropped back to low levels in November, and rose sharply to the highest

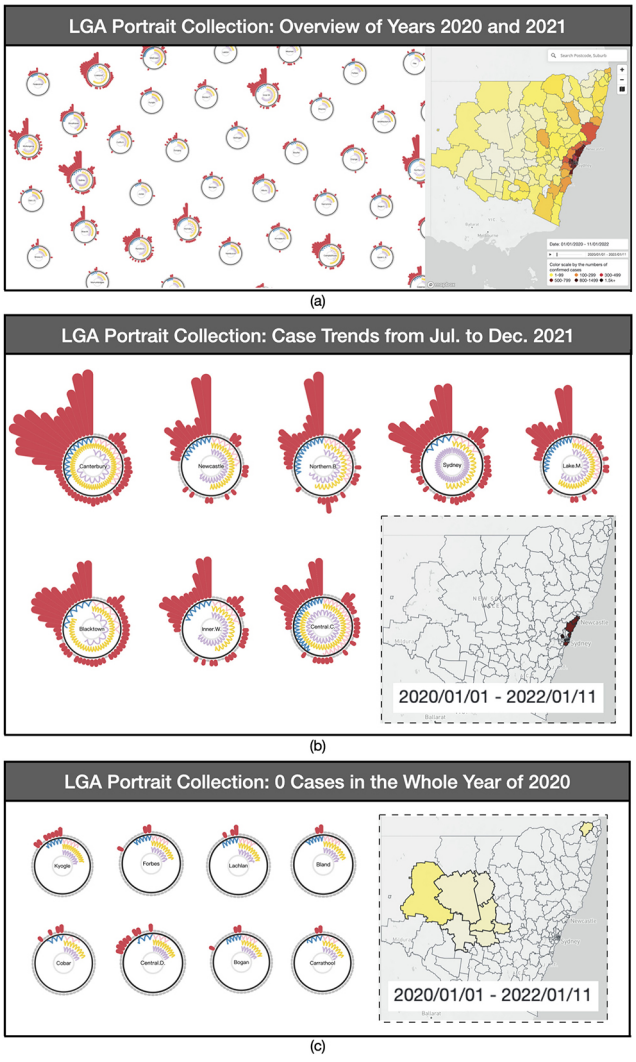

Fig. 6 Three cohorts of LGA Portrait Collections. (a) A partial overview of LGA Portraits and geo-locations for 2020 and 2021. (b) Selected LGA Portrait with cases appearing continuously from July to December 2021. (c) Selected LGA Portraits with no cases for the full year of 2020.

point in two years. We corroborated the visual result with the actual situation as the NSW government released Delta concerns on July 30th, 2021 [88], and the information released on November 28th, 2021, of the first Omicron variant case in NSW [89]. From GIS View, we found they are all located in coastal areas surrounding Sydney City.

The following visual output in Fig. 6(c) showed all LGAs with portraits had COVID-19 cases during 2021, although they did not have any cases in 2020. We could further observe that there were 7 LGAs located in inland areas despite the presence of Kyogle LGA near the coastal regions of these 8 LGAs. Compared with cases appearing frequently in other developed LGAs, this consequence was most likely related to the sparse population of these LGAs.

7.2 Event-driven cases

We observed that the first wave of COVID-19 spread across NSW from approximately January to October 2020. Based on the events on the timeline, we divided this period into three phases: uncontrolled, eased, and restricted control. Therefore, we further filtered each LGA portrait into weekly time spans from January 1st to October 20th, 2020. This enabled us to inspect the COVID-19 spreading situations among LGAs during this period, as tracked with the details of these events.

Uncontrolled phase. After interacting on the Control Panel, we hovered the mouse over each E Protein to spot LGAs, which detected the first case occurring in week 4 in Randwick, Paramatta, Kurring-gai, and Burwood.

Concurrently, we noticed the first intervention event started in week 13. Thus, we began by selecting the phase from January 1st, 2020 to March 24th, 2020, to identify the uncontrolled phase. We further utilized the playable timeline window in GIS View to explore the spreading pattern in the Greater Sydney Region and surrounding places. As shown in Fig. 7(a), besides the case that occurred in week 4, there were no other cases during the period from January 1st, 2020 to February 25th, 2020. However, starting from week 9 to the end of week 12, cases began spreading in the surrounding places of the Greater Sydney Region, and in a short period, the cases spread severely and finally caused outbreaks in almost all postal areas, especially in the Waverley LGA in the darker polygon color.

Drawing an overview of this uncontrolled phase, the visualization results in Fig. 7(b) show that all LGAs were affected by the first wave of weekly increasing COVID-19 cases, with no restricted events yet. We focused on LGAs with the highest number of COVID-19 infections and interacted with them in GIS View. We located them around the main cities, including Sydney. During this period, Bondi Beach, in the Waverley LGA, had more cases than other LGAs, peaking in only two weeks. We noted that Bondi Beach was crowded with massive gatherings of people, as reported in the news, and health officials announced a crowd ban on March 21st, 2020. NSW residents were facing physical and psychological

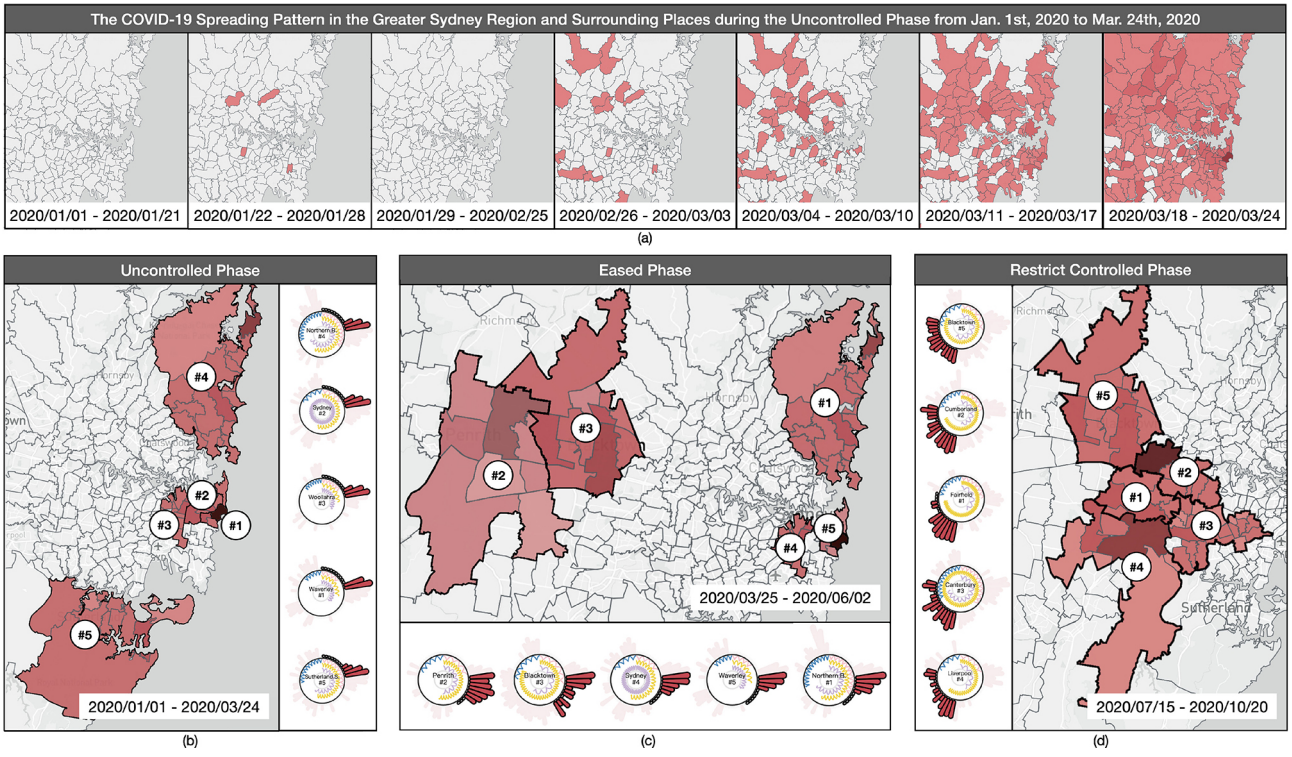

Fig. 7 The event-driven cases encompass explorations of (a) the COVID-19 spreading pattern in the uncontrolled phase and (b)–(d) three phases of overviews. (a) provides a detailed breakdown by time spans of the COVID-19 spreading pattern in the Greater Sydney Region and surrounding places during the uncontrolled phase from 2020/01/01 to 2020/03/24. In (b)–(d), each marked ranking number is arranged according to geo-location, where (b) contains the top 5 LGA portraits in total infection cases with non-controlled events applied to the same phase as (a); (c) displays the top 5 LGA portraits in total in the eased phase from 2020/03/25 to 2020/06/02; and (d) includes the top 5 LGA portraits in total case numbers during the restricted controlled phase from 2020/07/15 to 2020/10/20.

pressures due to the spread of COVID-19 and bushfires across the state. These factors indirectly contributed to overcrowding in these tourist hotspots. This situation was also reflected in the Northern Beaches LGA, which saw a rapid increase in COVID-19 cases in the first wave of the pandemic. We further examined the other two LGA portraits, Sydney and Woollahra, and concluded that the consequences were likely due to the connection to Waverley in GIS View, resulting from the movement of people in the adjacent LGA. The fifth infection case in the LGA occurred in Sutherland Shire, located in the southern region of Sydney. We could assume that the COVID-19 pandemic had spread to the outer LGAs in this phase.

Eased phase. We adjusted the timeline for the first NSW lockdown events from March 25th to June 30th, 2020. The visual results of the top 5 LGAs in Fig. 7(c) show that Northern Beaches had the highest number of cases, followed by Penrith and Blacktown, with Sydney LGA dropping to fourth place. Waverley finished fifth. With the special ban in place at Bondi Beach, the number of infection cases in Waverley and its adjacent LGA, Sydney, significantly decreased. We also noted that the NSW lockdown event had a positive impact on LGA cases in the Greater Sydney Region, as reflected by 0 infection cases in Northern Beaches, Sydney, and Waverley for several weeks in the second half of the selected timeline. Penrith, Blacktown, and Blackburn all had a larger number of lower-income residents and are located in the Greater Western Sydney Region, which had more than half the cosine arc of the lower-income RNA and an intense frequency. As a result, we anticipated that the pandemic would spread quickly to other LGAs in the Greater Western Sydney Region.

Restrict controlled phase. To test our hypothesis, we further filtered the timeline from July 15th to October 20th, 2020, which was when a new lockdown and tighter restrictions were implemented by the NSW government. The visual results, as shown in Fig. 7(d), reveal that all of the LGAs are connected in the Greater Western Sydney Region. We observed that the most strictly controlled events showed stable

increases in cases, even in Cumberland and Liverpool, which had continuous growth in cases over the next few weeks. This confirms our conjecture about the pandemic situation, with a large number of infection cases in the Greater Western Sydney Region.

We also noticed potential patterns among LGA portraits, particularly in the lower-income RNA, demonstrating that all five LGAs have significant lower-income population groups and high population density. In other words, during the implementation of the tightened lockdown, most people worked from home or were self-isolated, and the number of infection cases in the CBD and tourist hotspots markedly decreased. Thus, the factors in the LGA portraits will dominate and reflect the number of infection cases. Lower-income groups or other population factors in the census may be the most crucial factors affecting the spread of the pandemic.

7.3 Portrait-driven cases

The highlighted interaction in RNAs facilitates the comparison of four key factors within each portrait and multiple attached attributes in MDC View. We selected 12 alternative factors in the census indicators to enable a more comprehensive comparison of the COVID-19 situations based on their portrait factors. To conduct these comparisons, we analyzed portrait-driven cases with different factor influences in collaboration among views.

Prevalence rate influence. We began by exploring LGA portraits for the entire time period from January 1st, 2020 to January 11th, 2022, aimed at comparing LGA portraits of coastal areas with inland areas across different case modes. We randomly selected two coastal LGA polygons (Sydney and Mid-Coast) and two inland LGA polygons (Cobar and Hay) as representatives and compared them using actual case mode and prevalence rate mode, as depicted in Fig. 8.

The consequences conveyed that all S Proteins in Sydney and Mid-Coast LGAs decreased in prevalence rate mode, while all S Proteins in inland LGAs increased in height. The opposite occurred in actual case mode. We were able to make comparisons among all LGA portraits in both modes because they had been standardized using the same calculation methods as for the S Protein heights.

These findings suggest that almost all LGAs in inland areas had insufficient 10k population, resulting

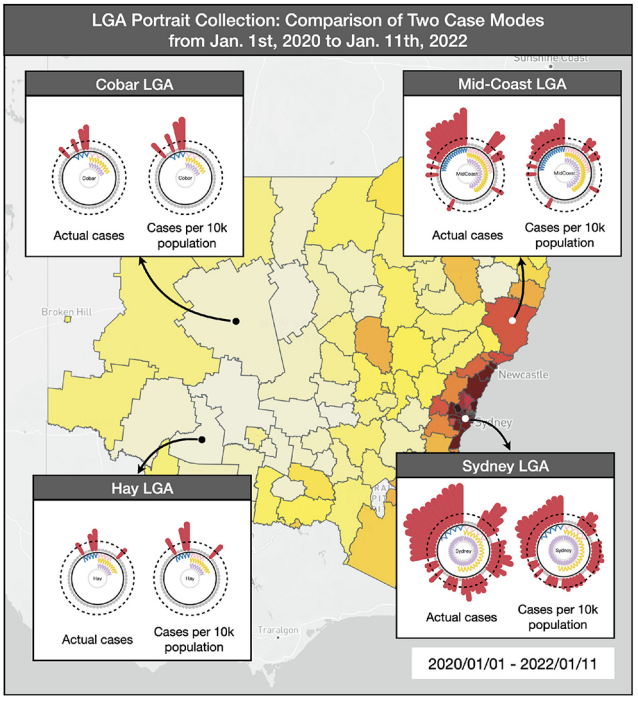

Fig. 8 Four cohorts of LGA portraits in different modes (actual cases and cases per 10k population) were selected using LGA polygons in GIS View. Two coastal LGAs (Sydney and Mid-Coast) are marked with white pins, while two inland LGAs (Cobar and Hay) are marked with black pins.

in the height of S Protein in their portraits growing. We also concluded that the cases per 10k population with standardization were similar in both coastal and inland areas, which validated the strong infectiousness of COVID-19.

Case trends and lower-income influences. Based on our previous findings, there may be potential relationships between COVID-19 cases and lower-income groups in each LGA. Therefore, we kept the COVID-19 outbreak timeline from the first year (January 1st, 2020 to January 5th, 2021) in the Control Panel and selected the top 10 LGAs with the worst COVID-19 pandemic situations (i.e., highest total case numbers).

From the top 10 rankings displayed on each portrait in the first year, as shown in Fig. 9(a), we identified that areas with RNAs longer than half the range of lengths, or LGAs with the busiest frequency of RNA in one factor, have comparatively worse COVID-19 situations. For example, Fairfield ranked the 8th highest in the number of cases with 143 total cases. This area contains the 4th longest lower-income RNA length but the highest tightness frequency. Additionally, the RNA lengths of aged and lone

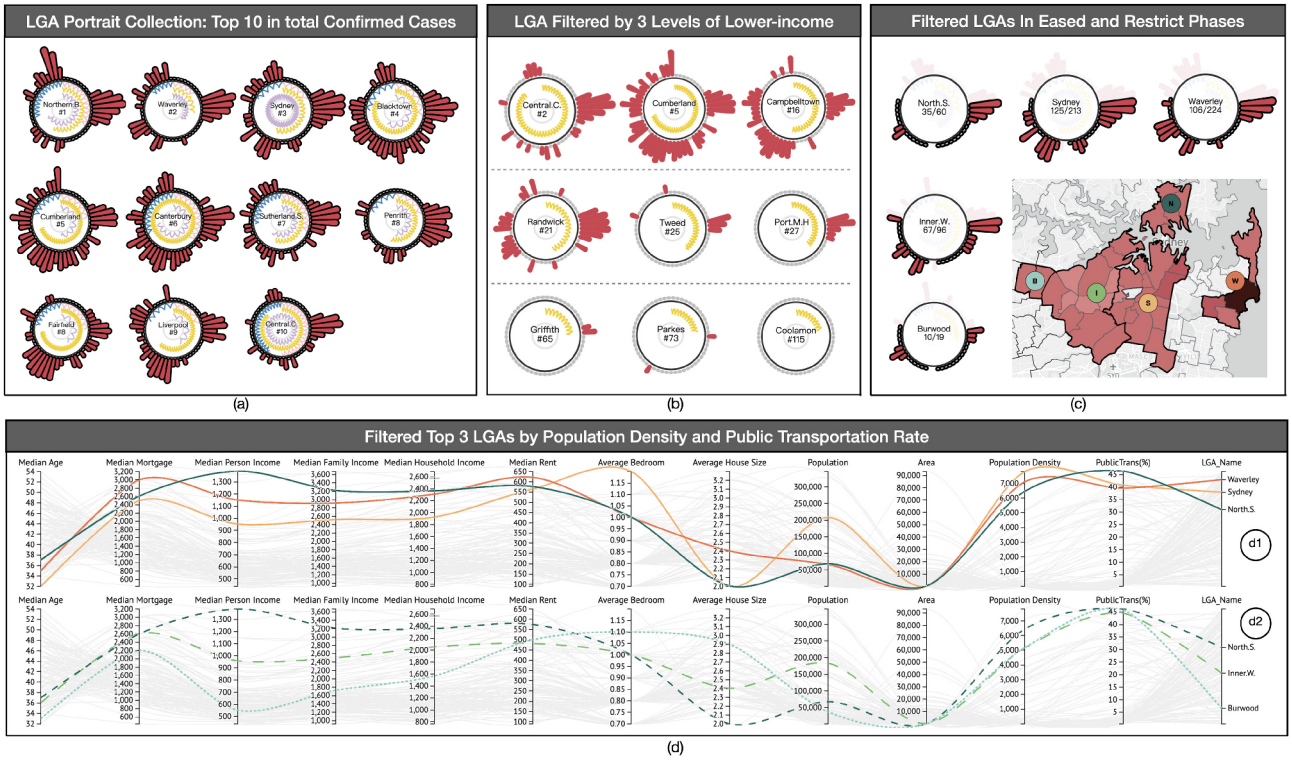

Fig. 9 The compiled cohorts of cases are organized according to LGA portrait factors. Specifically, (a) an LGA portrait collection showcases the top ten areas with the highest infection case counts across the entire timeline; (b) a visualization shelf of LGAs is presented in three levels, filtered by lower-income factors; (c) detailed portraits of filtered LGAs are provided, and the colors of map pin-points correspond to the colors used in (d); and (d) two portrait cohorts are presented after being filtered by population density and public transportation rate, displayed in multidimensional coordinates and distinguished by colored lines to differentiate LGAs.

persons in this area do not approach half the length of the full path.

Given that all lower-income RNAs are highlighted in the visual overview, we assume that the lower-income RNA in the LGA portrait is the most influential factor in COVID-19. We selected 9 LGA portraits only based on the appearances of low-income RNA and categorized them into three layers whose lengths are longer than one-half, longer than one-quarter but shorter than one-half, and shorter than one-quarter. As shown in Fig. 9(b), significant patterns are revealed, showing that the COVID-19 situation improves by layers, where the length of lower-income RNAs achieves shorter from left to right and top to bottom.

Combined influences on LGA portraits, geo-locations, and events. We reset the time period to encompass both the eased and restricted controlled phases. Upon analyzing the Greater Sydney Region using GIS View, we observed that adjacent LGAs exhibited similar heatmap polygons. Using an iterative approach, we selected the top 5 LGA portraits with the highest case sums during this period and evaluated their case sums, as well as the total case sums throughout the year. These five LGAs are highlighted in Fig. 9(c), whose pin-points correspond to the colors used in Fig. 9(d). With the exception of the Waverley LGA, which had been analyzed in previous cases, the other four LGAs were adjacent. While all five LGAs had case ratios exceeding 50% during the selected period, given the duration of our chosen timeframe, which spanned more than half a year, these eased and restricted events effectively suppressed the spread of COVID-19 in this densely populated region to a certain extent.

Household and dwelling influence. The LGA portraits were reverse-selected in the MDC View to discover the consequences of COVID-19 spread in conjunction with other highlighted census attributes from residents' perspectives. In each column, the median age of each LGA is roughly inversely proportional to financial factors, including the median mortgage, personal income, family income, household income, and rent cost. In most living environments

with more than two bedrooms in one household, the majority of people do not meet the standard of one bedroom per person. This indicates that most people still live with others or families, which must also be considered in the analysis of the consequences of the pandemic's spread.

Population density influence. The three LGAs with the highest population densities in the Greater Sydney Region are Sydney, Waverley, and Northern Sydney. These areas have all been severely affected by COVID-19. Upon analyzing their trends on coordinates, we discovered that they share similarities such as a young median age, similar financial profiles, high living costs, and a propensity for using public transit.

8 Discussion and interview with domain experts

The *+msRNAer* prototype applied with our aggregated COVID-19 dataset was deployed and used in the workplaces of three domain experts from the Australian government. Through interaction with domain experts, we observed the following.

8.1 Influence of key risk factors

The combined influence of key risk factors has a notable negative relationship with the community's resilience profile against the virus spread. According to infection cases, LGAs with a higher representation in four risk groups are all at the top of the list. For example, among the entire list of factors in NSW, the Central Coast has the highest-ranked key factors, ranking #1 in aged group, #1 in lone person group, and #2 in lower-income group; Canterbury also has the highest-ranked key factors, ranking #2 in aged group, #1 in lower-income group, and #3 in lone person group.

The larger the population in the high-risk group, the higher the COVID-19 cases.

Key factor—aged group. The areas with smaller aged male or female groups among LGAs in NSW have significantly reduced infection cases. In the ranking according to infection cases, the top LGAs generally have older age groups in community representation. There is not much difference in the effect caused by aged male or female groups.

Key factor—lower-income. Among four key risk factors, the lower-income group has the highest impact on COVID-19 risk. The visualization results show that infection cases significantly increase for LGAs in the lower-income group, roughly occupying one-quarter of the maximum of the lower-income group population in NSW. For example, Cumberland, Canterbury, Sutherland, Sydney, Fairfield, Penrith, Blacktown, Liverpool, and Central Coast are the LGAs with the worst COVID-19 situation, which have larger population sizes in the lower-income group from small to large.

Key factor—lone-person. The lone-person group has a certain impact as well. For example, Sydney CBD ranked #3 according to infection cases and has key risk factor rankings: lone person (#1), low-income (#12), and age group (male #25, female #29). Compared with other factors across all LGAs, its influence seems the weakest. However, it would also affect the LGA's resilience if taking into account other risk factors.

Lower risk factors, higher resilience. According to the domain expert, if LGAs have a lower presentation of most risk factors, their resilience appears to be higher. This also showed in the situation where several higher-resilience LGAs would respond quickly and positively to flatten the curve and reduce the risk, even when they were the areas where these early and severe infections happened during outbreaks.

Domain experts pointed out that these implications largely support the guidelines developed by the Australian government during the pandemic. However, they would also include representation of other under-resourced people in future work, who may also be more at risk of exposure, including Aboriginal and Torres Strait Islander people, people living in aged care facilities, and people with disabilities. They plan to share with us more datasets to increase the diversity of higher-risk groups. For the lone-person risk group, domain experts explained that an individual or group's social relationships should be significantly affected and explained that a lone person would have much fewer resources and social support compared to people living together. Australian and most Western societies encourage young adults to move out and live alone. However, this situation may not be beneficial during a pandemic. Apart from living conditions, they would also like to explore diverse relationship statuses as they believe that residents' relationships might have a greater impact

on social support and mental well-being.

8.2 Influence of other factors

The higher population does attract more risk for infection. All LGAs with larger resident populations should be given more attention during all phases of the pandemic. For example, Canterbury, Blacktown, and Central Coast ranked top 3 in population and have higher infection cases than other outbreak LGAs.

Travel restrictions could effectively control the risk caused by activities related to public transportation. For instance, North Sydney, Burwood, and Inner West are the top 3 in terms of the number of people who frequently use public transit. These LGAs have the earliest cases during outbreaks before the controlled phase due to the high volume of public transit. However, they respond positively and quickly when the government increases travel restrictions.

Geographic factor. From the overview, the total number of infection cases in coastal areas is generally higher than in inland areas for the first 12 months. In particular, those areas with famous beaches had severe infections, for example, the Northern Beaches and Waverley (with Bondi Beach). Analyzing the spreading pattern in the Greater Sydney Region during the uncontrolled phase, domain experts observed that the outbreak initially appeared in scattered locations throughout the LGA and subsequently spread to areas adjacent to existing cases. During the early stage of the outbreak without control, spreading patterns were more apparent in residential areas. However, in the latter half of the phase, the pattern became more prevalent in tourist attractions, ultimately resulting in the peak of infection cases in Waverley.

Domain experts have also discovered that the combination of key risk factors and other factors can have a more significant impact. Taking Northern Beaches as an example, they had the most infection cases in the first year of the COVID-19 outbreak, with a greater effect on geographic factors and the highest presentation in lone-person groups, despite a reasonable economic indicator and low presentation in the aged group. Although most of these implications seem obvious, they would consider strengthening governance based on these factors, as these implications also prove the effectiveness of government intervention. The experts also emphasized the crucial need for clearer regulations about beaches and recreational usage.

After experimenting with experts, we also conducted a follow-up interview. Overall, experts were impressed by the "actionable insights" that *+msRNAer* obtained. They felt excited about the interface design inspired by the viral 2D structure. They expressed that it supported their memory of each data object and their representation because of the familiarity of the metaphor adopted. They commented, "this would help us reduce the effort and time involved in training other staff." They found that "interaction with humans and interface-in-the-loop is intuitive and assists them in manipulating exploration along a timeline and geographic map." They also used the prototype system to make several valid assumptions and discover unexpected implications. For assumption validation, they found that the prototype's capability to help them target higher-risk groups is "specifically useful", and the interactive portrait was "the most useful treasure" considering the magnitude of the data they have.

Experts consistently agreed that the visual modeling of *+msRNAer* can be easily applied in most epidemiological analyses due to the similarity of virus transmission, which is often caused by spatiotemporal and objective factors. The prototype will be reusable because the integrated portrait designs helped them augment awareness by porting most data features. They were swiftly able to identify higher-risk regions with "constrained factors and higher presentation in risk groups using the realistic". The experts mentioned that they expect the prototype to enable them to understand how to build profiles and predictions for communities in the future based on "reasonable and meaningful information provided".

9 Limitations and future work

9.1 Limitations

Our work was constrained by the quality and availability of the data provided, including the related datasets and redefined risk factors with domain experts. These data issues could have led to inaccuracies in our final results. For example, COVID-19 case data collected by the NSW government is based on usual residence, but

some records did not track location accurately, which could have compromised the precision of our analysis. Concurrently, the case data we had access to were recorded by LGAs or postal areas, which limited our ability to examine more granular infection trends in communities. Additionally, our use of a weekly time frame, as dictated by the aggregated datasets, may have limited our ability to uncover more nuanced patterns in the pandemic's progression. The Australian Bureau of Statistics conducted the latest census surveys on communities in 2020, but the complete summary will not be released until mid-2023. Thus, we utilized the 2016 census in aggregated datasets, which may pose bias in accurately assessing the impact of the pandemic on population demographics and social indicators during this ongoing period. Moreover, the LGA profiles we used contained a vast amount of data, with over 15,000 variables each. We could not concentrate on all variables, necessitating a reduction to only thirteen based on expert supervision. We considered social and economic indicators and indicators of vulnerability to COVID-19, such as rental and mortgage affordability. However, variables in other categories may have directly or indirectly influenced our results. Furthermore, the intervention events in the COVID-19 dataset were constrained by the Australian government system, with different levels of government responsible for administering public policies and programs, which may have resulted in varying intervention measures implemented across LGAs in NSW.

9.2 Future work

Based on the feedback from participants in the user study and discussions with domain experts, we have targeted four aspects to focus on in our future work.

Firstly, we plan to update our prototype by importing refined COVID-19 case data and new census data later next year, when community datasets are collected during the pandemic.

Secondly, to increase the diversity of community profiles, we will include more variables, such as relationships, ethnic groups mentioned by domain experts, and education level. We will continue to apply machine learning algorithms, such as Principal Components Analysis (PCA) for dimensionality reduction to add more categorical indicators and time series predictions for infectious cases affected by risk factors. We also expect to propose a rating system for all variables with assistance and evaluation from domain experts, which would be able to quantify the characteristics of communities. By designing an appropriate measurement matrix for indicators, we will be able to create index metrics for future use in other epidemiological analysis scenarios. These metrics will not only assist decision-makers in making pandemic prevention measurements but will also educate the public on personal influence, and eventually, how to work together to tackle this great challenge through their own efforts.

Thirdly, with the improved approach, we intend to conduct a systematic review with more government staff. As suggested, we will offer an automated reporting function for storytelling purposes. This function would help disseminate the right messages to the public.

Finally, we will consider applying our prototype to more transmission datasets in epidemiological analyses to validate the scalability of *+msRNAer*.

10 Conclusions

With epidemiological analysis as a backdrop, this paper proposes a new visual modeling method called *+msRNAer* for exploring and comparing spatiotemporal and multidimensional features based on requirements in epidemiology. The method employs a metaphor to assemble portraits that can be used for visualizing each community by combining the visual encoding of time-varying case numbers with objective risk factors that may affect transmissions. The method integrates multiple views, including a Control Panel, GIS, and MDC Views, to provide wide-scope observations on filtering events at different severity levels, geo-based spreading distributions, and multidimensional risk factors for each community.

To evaluate the feasibility and effectiveness of *+msRNAer*, we deployed and applied a two-year-scale aggregated dataset by integrating COVID-19 cases with geo-information, NLP-extracted event division on timelines, and risk factors from NSW census data based on expert supervision. After applying *+msRNAer* to this COVID-19 aggregated dataset, we progressively validated the feasibility and effectiveness of *+msRNAer* by conducting one user study to iteratively improve the applied prototype and comparing visual portraits from

profuse perspectives in three subject-driven case studies. We summarized how geography, phases of intervention events, and objective risk factors affected COVID-19 spreading situations during the pandemic.

In further interviews with domain experts, we identified additional objective factors that may be influential in the Australian context. We examined pre-existing community factors and discovered practical implications for potential patterns of established community characteristics against the vulnerability facing this pandemic. Despite some limitations and future work, feedback from domain experts suggests that the *+msRNAer* can be considered a common visual modeling method for exploring community-based spatiotemporal and multidimensional features. This method is expected to be applied to abundant epidemiological analyses, such as investigating case trends and comparisons, geo-distribution and transmission, risk factors explorations and rankings, and other related tasks.

Availability of data and materials

We confirm that all datasets used in this work are open-source and accessible. The source code is available on GitHub: https://github.com/YuDong5018/msRNAers.

Funding

This work is supported by National Natural Science Foundation of China (NSFC) under Grant No. 61972010 and UTS–CSC Scholarship by the University of Technology Sydney and China Scholarship Council under Agreement No. 201908200009.

Author contributions

All authors participated in conceiving, discussing, designing, and writing this work. Yu Dong and Jie Hua contributed to collecting and processing data. Yu Dong and Christy Jie Liang performed the programming for this visual modeling. Yu Dong, Christy Jie Liang, and Yi Chen contributed to proofreading. Christy Jie Liang and Yi Chen behaved under the corresponding author's direction.

Acknowledgements

The authors would like to thank all domain experts from the Australian government who provided valuable feedback and comments based on their expertise.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Deodhar, S.; Bisset, K. R.; Chen, J. Z.; Ma, Y. F.; Marathe, M. V. An interactive, web-based high performance modeling environment for computational epidemiology. *ACM Transactions on Management Information Systems* Vol. 5, No. 2, Article No. 7, 2014.

[2] Carroll, L. N.; Au, A. P.; Detwiler, L. T.; Fu, T. C.; Painter, I. S.; Abernethy, N. F. Visualization and analytics tools for infectious disease epidemiology: A systematic review. *Journal of Biomedical Informatics* Vol. 51, 287–298, 2014.

[3] Christakis, N. A.; Fowler, J. H. Social network visualization in epidemiology. *Norwegian Journal of Epidemiology* Vol. 19, No. 1, 5–16, 2009.

[4] Andrienko, G.; Andrienko, N.; Demsar, U.; Dransch, D.; Dykes, J.; Fabrikant, S. I.; Jern, M.; Kraak, M. J.; Schumann, H.; Tominski, C. Space, time and visual analytics. *International Journal of Geographical Information Science* Vol. 24, No. 10, 1577–1600, 2010.

[5] Angelini, M.; Cazzetta, G. Progressive visualization of epidemiological models for COVID-19 visual analysis. In: *Advanced Visual Interfaces. Supporting Artificial Intelligence and Big Data Applications. Lecture Notes in Computer Science, Vol. 12585*. Reis, T.; Bornschlegl, M. X.; Angelini, M.; Hemmje, M. L. Eds. Springer Cham, 163–173, 2021.

[6] Zhang, Y. X.; Sun, Y. F.; Padilla, L.; Barua, S.; Bertini, E.; Parker, A. G. Mapping the landscape of COVID-19 crisis visualizations. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Article No. 608, 2021.

[7] Bras, P. L.; Gharavi, A.; Robb, D. A.; Vidal, A. F.; Padilla, S.; Chantler, M. J. Visualising COVID-19 research. *arXiv preprint* arXiv:2005.06380, 2020.

[8] Park, M.; Cook, A. R.; Lim, J. T.; Sun, Y.; Dickens, B. L. A systematic review of COVID-19 epidemiology based on current evidence. *Journal of Clinical Medicine* Vol. 9, No. 4, 967, 2020.

[9] Rydow, E.; Borgo, R.; Fang, H.; Torsney-Weir, T.; Swallow, B.; Porphyre, T.; Turkay, C.; Chen, M. Development and evaluation of two approaches of visual sensitivity analysis to support epidemiological modeling. *IEEE Transactions on Visualization and Computer Graphics* Vol. 29, No. 1, 1255–1265, 2023.

[10] Dykes, J.; Abdul-Rahman, A.; Archambault, D.; Bach, B.; Borgo, R.; Chen, M.; Enright, J.; Fang, H.; Firat, E.

E.; Freeman, E.; et al. Visualization for epidemiological modelling: Challenges, solutions, reflections and recommendations. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* Vol. 380, No. 2233, 20210299, 2022.

[11] Sanz-Leon, P.; Hamilton, L. H. W.; Raison, S. J.; Pan, A. J. X.; Stevenson, N. J.; Stuart, R. M.; Abeysuriya, R. G.; Kerr, C. C.; Lambert, S. B.; Roberts, J. A. Modelling herd immunity requirements in Queensland: Impact of vaccination effectiveness, hesitancy and variants of SARS-CoV-2. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* Vol. 380, No. 2233, 20210311, 2022.

[12] Pooley, C. M.; Doeschl-Wilson, A. B.; Marion, G. Estimation of age-stratified contact rates during the COVID-19 pandemic using a novel inference algorithm. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* Vol. 380, No. 2233, 20210298, 2022.

[13] Panovska-Griffiths, J.; Swallow, B.; Hinch, R.; Cohen, J.; Rosenfeld, K.; Stuart, R. M.; Ferretti, L.; Di Lauro, F.; Wymant, C.; Izzo, A.; et al. Statistical and agent-based modelling of the transmissibility of different SARS-CoV-2 variants in England and impact of different interventions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* Vol. 380, No. 2233, 20210315, 2022.

[14] Wei, L. L. Y.; Ibrahim, A. A. A.; Nisar, K.; Ismail, Z. I. A.; Welch, I. Survey on geographic visual display techniques in epidemiology: Taxonomy and characterization. *Journal of Industrial Information Integration* Vol. 18, 100139, 2020.

[15] Lord, S. R.; Dayhew, J. Visual risk factors for falls in older people. *Journal of the American Geriatrics Society* Vol. 49, No. 5, 508–515, 2001.

[16] DelCourt, C.; Moreau, G.; Cougnard-Gregoire, A. The potential of cardiovascular risk factors for reducing visual impairment: A pooled analysis of European epidemiological studies. *Investigative Ophthalmology & Visual Science* Vol. 58, No. 8, 2209–2209, 2017.

[17] Chui, K. K. H.; Wenger, J. B.; Cohen, S. A.; Naumova, E. N. Visual analytics for epidemiologists: Understanding the interactions between age, time, and disease with multi-panel graphs. *PLoS One* Vol. 6, No. 2, e14683, 2011.

[18] Muhammad, L. J.; Algehyne, E. A.; Usman, S. S.; Ahmad, A.; Chakraborty, C.; Mohammed, I. A. Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. *SN Computer Science* Vol. 2, No. 1, 11, 2020.

[19] Mocnik, F. B.; Raposo, P.; Feringa, W.; Kraak, M. J.; Köbben, B. Epidemics and pandemics in maps—the case of COVID-19. *Journal of Maps* Vol. 16, No. 1, 144–152, 2020.

[20] Thöny, M.; Schnürer, R.; Sieber, R.; Hurni, L.; Pajarola, R. Storytelling in interactive 3D geographic visualization systems. *ISPRS International Journal of Geo-Information* Vol. 7, No. 3, 123, 2018.

[21] Goodwin, S.; Dykes, J.; Slingsby, A.; Turkay, C. Visualizing multiple variables across scale and geography. *IEEE Transactions on Visualization and Computer Graphics* Vol. 22, No. 1, 599–608, 2016.

[22] Shapiro, B. R.; Pearman, F. A. Using the interaction geography slicer to visualize New York City Stop & Frisk. In: Proceedings of the IEEE VIS Arts Program, 1–8, 2017.

[23] Isenberg, P.; Isenberg, T.; Sedlmair, M.; Chen, J.; Möller, T. Visualization as seen through its research paper keywords. *IEEE Transactions on Visualization and Computer Graphics* Vol. 23, No. 1, 771–780, 2017.

[24] Kahn, P. COVID-19 online visualization collection (COVIC). Available at https://mprove.de/script/20/covic/_media/COVICProjectSummary011621.pdf.

[25] Chua, J.; Lim, B.; Fenwick, E. K.; Gan, A. T.; Tan, A. G.; Lamoureux, E.; Mitchell, P.; Wang, J. J.; Wong, T. Y.; Cheng, C. Y. Prevalence, risk factors, and impact of undiagnosed visually significant cataract: The Singapore epidemiology of eye diseases study. *PLoS One* Vol. 12, No. 1, e0170804, 2017.

[26] Steinger, T.; Gilliand, H.; Hebeisen, T. Epidemiological analysis of risk factors for the spread of potato viruses in Switzerland. *Annals of Applied Biology* Vol. 164, No. 2, 200–207, 2014.

[27] Maciejewski, R.; Livengood, P.; Rudolph, S.; Collins, T. F.; Ebert, D. S.; Brigantic, R. T.; Corley, C. D.; Muller, G. A.; Sanders, S. W. A pandemic influenza modeling and visualization tool. *Journal of Visual Languages & Computing* Vol. 22, No. 4, 268–278, 2011.

[28] Trajkova, M.; Alhakamy, A.; Cafaro, F.; Vedak, S.; Mallappa, R.; Kankara, S. R. Exploring casual COVID-19 data visualizations on Twitter: Topics and challenges. *Informatics* Vol. 7, No. 3, 35, 2020.

[29] Inselberg, A.; Dimsdale, B. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In: Proceedings of the 1st IEEE Conference on Visualization: Visualization, 361–378, 1990.

[30] Matute, J.; Linsen, L. Visual stratification for epidemiological analysis. In: Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Posters, 81–83, 2017.

[31] Akram Hassan, K.; Rönnberg, N.; Forsell, C.; Cooper, M.; Johansson, J. A study on 2D and 3D

parallel coordinates for pattern identification in temporal multivariate data. In: Proceedings of the 23rd International Conference Information Visualisation, 145–150, 2019.

[32] Zeng, W.; Zhang, Y. G.; Hu, K.; Jiang, Y. X. Multi-dimensional visualization and simulation analysis of COVID-19 outbreak. In: *Advances in Artificial Intelligence and Security. Communications in Computer and Information Science, Vol. 1423*. Sun, X.; Zhang, X.; Xia, Z.; Bertino, E. Eds. Springer Cham, 541–553, 2021.

[33] Liu, Q.; Zheng, Z. Q.; Zheng, J. B.; Chen, Q. Y.; Liu, G.; Chen, S. H.; Chu, B. J.; Zhu, H. Y.; Akinwunmi, B.; Huang, J.; et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: Digital topic modeling approach. *Journal of Medical Internet Research* Vol. 22, No. 4, e19118, 2020.

[34] Badr, H. S.; Du, H. R.; Marshall, M.; Dong, E. S.; Squire, M. M.; Gardner, L. M. Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases* Vol. 20, No. 11, 1247–1254, 2020.

[35] Ulahannan, J. P.; Narayanan, N.; Thalhath, N.; Prabhakaran, P.; Chaliyeduth, S.; Suresh, S. P.; Mohammed, M.; Rajeevan, E.; Joseph, S.; Balakrishnan, A.; et al. A citizen science initiative for open data and visualization of COVID-19 outbreak in Kerala, India. *Journal of the American Medical Informatics Association* Vol. 27, No. 12, 1913–1920, 2020.

[36] Dey, S. K.; Rahman, M. M.; Siddiqi, U. R.; Howlader, A. Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. *Journal of Medical Virology* Vol. 92, No. 6, 632–638, 2020.

[37] Zhang, Y. X.; Sun, Y. F.; Gaggiano, J. D.; Kumar, N.; Andris, C.; Parker, A. G. Visualization design practices in a crisis: Behind the scenes with COVID-19 dashboard creators. *IEEE Transactions on Visualization and Computer Graphics* Vol. 29, No. 1, 1037–1047, 2023.

[38] World Health Organization. WHO coronavirus (COVID-19) dashboard. Available at https://covid19.who.int/.

[39] Data.NSW. NSW COVID-19 cases data. Available at https://data.nsw.gov.au/nsw-covid-19-data/cases.

[40] Renton, S. Australians post COVID-19. 2020. Available at https://mccrindle.com.au/wp-content/uploads/reports/COVID19-Phase3-Report-2020.pdf.

[41] Sweet, A.; Davies, J. Fear down, job-seeking up as Australians feel the financial impact of COVID-19. 2021. Available at https://www.coredata.com.au/blog/fear-down-job-seeking-activity-up-as-australians-feel-the-financial-impact-of-covid-19.

[42] AIHW. Health expenditure Australia 2018-19. Available at https://www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure-australia-2018-19/contents/data-visualisation.

[43] University of Melbourne. Coronavirus 10-day forecast. Available at https://covid19forecast.science.unimelb.edu.au.

[44] Dong, E. S.; Du, H. R.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* Vol. 20, No. 5, 533–534, 2020.

[45] World Health Organization. WHO Coronavirus (COVID-19) dashboard. Available at https://covid19.who.int/.

[46] Mathieu, E.; Ritchie, H.; Rodés-Guirao, L.; Appel, C.; Giattino, C.; Hasell, J.; Macdonald, B.; Dattani, S.; Beltekian, D.; Ortiz-Ospina, E.; Roser, M. Coronavirus pandemic (COVID-19). Our World in Data, 2020. Available at https://ourworldindata.org/coronavirus.

[47] Northern Territory Government of Australia. COVID-19 data. Available at https://coronavirus.nt.gov.au/current-status.

[48] Queensland Government. Queensland COVID-19 statistics. Available at https://www.qld.gov.au/health/conditions/health-alerts/coronavirus-covid-19/current-status/statistics.

[49] Li, R. Visualizing COVID-19 information for public: Designs, effectiveness, and preference of thematic maps. *Human Behavior and Emerging Technologies* Vol. 3, No. 1, 97–106, 2021.

[50] Australian Government Department of Health and Aged Care. Weekly COVID-19 reporting. Available at https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/coronavirus-covid-19-current-situation-and-case-numbers.

[51] Covid-19-au. COVID-19 in Australia real-time report. Available at https://covid-19-au.com.

[52] The University of Sydney. NSW COVID-19 cases and community profile by The University of Sydney. Available at https://covid19-data.sydney.edu.au.

[53] Australian National University. New data visualisation tool to help track COVID-19. Available at https://nceph.anu.edu.au/news-events/news/new-data-visualisation-tool-help-track-covid-19.

[54] Government of South Australia. New data visualisation tool to help track COVID-19. Available at https://www.covid-19.sa.gov.au/home/dashboard.

[55] State Government of Victoria. Victorian COVID-19 data. Available at https://www.dhhs.vic.gov.au/victorian-coronavirus-covid-19-data.

[56] Bing. Conronavirus Australia - live map tracker from Microsoft Bing. Available at https://www.bing.com/covid/local/australia.

[57] Yu, X. M.; Ferreira, M. D.; Paulovich, F. V. Senti-COVID19: An interactive visual analytics system for detecting public sentiment and insights regarding COVID-19 from social media. *IEEE Access* Vol. 9, 126684–126697, 2021.

[58] Leung, C. K.; Chen, Y. B.; Hoi, C. S. H.; Shang, S. Y.; Wen, Y.; Cuzzocrea, A. Big data visualization and visual analytics of COVID-19 data. In: Proceedings of the 24th International Conference Information Visualisation, 415–420, 2020.

[59] Lan, Y.; Desjardins, M. R.; Hohl, A.; Delmelle, E. Geovisualization of COVID-19: State of the art and opportunities. *Cartographica*: *the International Journal for Geographic Information and Geovisualization* Vol. 56, No. 1, 2–13, 2021.

[60] Wu, Y. H.; Gao, S. H.; Mei, J.; Xu, J.; Fan, D. P.; Zhang, R. G.; Cheng, M. M. JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation. *IEEE Transactions on Image Processing* Vol. 30, 3113–3126, 2021.

[61] Muto, K.; Yamamoto, I.; Nagasu, M.; Tanaka, M.; Wada, K. Japanese citizens' behavioral changes and preparedness against COVID-19: An online surveyduring the early phase of the pandemic. *PLoS One* Vol. 15, No. 6, e0234292, 2020.

[62] Giordano, G.; Blanchini, F.; Bruno, R.; Colaneri, P.; di Filippo, A.; di Matteo, A.; Colaneri, M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* Vol. 26, No. 6, 855–860, 2020.

[63] Bachtiger, P.; Peters, N. S.; Walsh, S. L. Machine learning for COVID-19—Asking the right questions. *The Lancet Digital Health* Vol. 2, No. 8, e391–e392, 2020.

[64] Regulski, P.; Wendykier, P.; Kantiem, K.; Murdzek, W. Advanced methods of visual analysis and visualization of various aspects of the COVID-19 outbreak in Poland. *Procedia Computer Science* Vol. 192, 4194–4199, 2021.

[65] Hemied, O. S.; Gadelrab, M. S.; Sharara, E. A.; Soliman, T. H. A.; Tsuji, A.; Terada, K. A COVID-19 visual diagnosis model based on deep learning and GradCAM. *IEEJ Transactions on Electrical and Electronic Engineering* Vol. 17, No. 7, 1038–1047, 2022.

[66] Leite, R. A.; Schetinger, V.; Ceneda, D.; Henz, B.; Miksch, S. COVIs: Supporting temporal visual analysis of COVID-19 events usable in data-driven journalism. In: Proceedings of the IEEE Visualization Conference, 56–60, 2020.

[67] Chen, B.; Shi, M.; Ni, X.; Ruan, L.; Jiang, H.; Yao, H.; Wang, M.; Song, Z.; Zhou, Q.; Ge, T. Visual data analysis and simulation prediction for COVID-19. *arXiv preprint* arXiv:2002.07096, 2020.

[68] Antweiler, D.; Sessler, D.; Ginzel, S.; Kohlhammer, J. Towards the detection and visual analysis of COVID-19 infection clusters. In: Proceedings of the EuroVis Workshop on Visual Analytics, 43–47, 2021.

[69] Xu, H. W.; Berres, A.; Thakur, G.; Sanyal, J.; Chinthavali, S. EPIsembleVis: A geo-visual analysis and comparison of the prediction ensembles of multiple COVID-19 models. *Journal of Biomedical Informatics* Vol. 124, 103941, 2021.

[70] Zhou, Y.; He, H.; Rong, J. Q.; Cheng, Y.; Li, Y. C.; Zhong, W.; Jiang, F. Visual analysis and exploration of COVID-19 based on multi-source heterogeneous data. In: Proceedings of the International Conferences on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data and IEEE Congress on Cybermatics, 62–69, 2020.

[71] Dong, Y.; Ian, O.; Liang, J.; Yuan, X. R.; Vinh, N. Q. User-centered visual explorer of in-process comparison in spatiotemporal space. *Journal of Visualization* Vol. 26, No. 2, 403–421, 2023.

[72] Cao, L. B. AI in combating the COVID-19 pandemic. *IEEE Intelligent Systems* Vol. 37, No. 2, 3–13, 2022.

[73] Reinert, A.; Snyder, L. S.; Zhao, J. Q.; Fox, A. S.; Hougen, D. F.; Nicholson, C.; Ebert, D. S. Visual analytics for decision-making during pandemics. *Computing in Science & Engineering* Vol. 22, No. 6, 48–59, 2020.

[74] Afzal, S.; Ghani, S.; Jenkins-Smith, H. C.; Ebert, D. S.; Hadwiger, M.; Hoteit, I. A visual analytics based decision making environment for COVID-19 modeling and visualization. In: Proceedings of the IEEE Visualization Conference, 86–90, 2020.

[75] Afzal, S.; Maciejewski, R.; Ebert, D. S. Visual analytics decision support environment for epidemic modeling and response evaluation. In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology, 191–200, 2011.

[76] Bowe, E.; Simmons, E.; Mattern, S. Learning from lines: Critical COVID data visualizations and the quarantine quotidian. *Big Data & Society* Vol. 7, No. 2, https://doi.org/10.1177/2053951720939236, 2020.

[77] Preim, B.; Lawonn, K. A survey of visual analytics for public health. *Computer Graphics Forum* Vol. 39, No. 1, 543–580, 2020.

[78] Hassan, A. H. M.; Ali Mohammed Qasem, A.; Abdalla, W. F. M.; Elhassan, O. H. Visualization & prediction of COVID-19 future outbreak by using machine learning. *International Journal of Information Technology and Computer Science* Vol. 13, No. 3, 16–32, 2021.

[79] Healey, C. G.; Simmons, S. J.; Manivannan, C.; Ro, Y. Visual analytics for the coronavirus COVID-19 pandemic. *Big Data* Vol. 10, No. 2, 95–114, 2022.

[80] Dong, Y.; Fauth, A.; Huang, M. L.; Chen, Y.; Liang, J. PansyTree: Merging multiple hierarchies. In: Proceedings of the IEEE Pacific Visualization Symposium, 131–135, 2020.

[81] NSW Health. Latest media releases from NSW Health. Available at https://www.health.nsw.gov.au/news/Pages/default.aspx.

[82] NSW Government. NSW administrative boundaries. Available at https://datasets.seed.nsw.gov.au/dataset/nsw-administrative-boundaries.

[83] NSW Government. Who are very low to moderate income earners? Available at https://www.facs.nsw.gov.au/providers/housing/affordable/about/chapters/who-are-very-low-to-moderate-income-earners.

[84] Desai, A.; Nouvellet, P.; Bhatia, S.; Cori, A.; Lassmann, B. Data journalism and the COVID-19 pandemic: Opportunities and challenges. *The Lancet Digital Health* Vol. 3, No. 10, e619–e621, 2021.

[85] Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv preprint* arXiv:1301.3781, 2013.

[86] Bostock, M.; Ogievetsky, V.; Heer, J. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* Vol. 17, No. 12, 2301–2309, 2011.

[87] Mapbox. Maps and location for developers. Available at https://www.mapbox.com/.

[88] NSW Health. Fighting the Delta outbreak with new restrictions for local government areas (LGAs) of concern. Available at https://www.health.nsw.gov.au/news/Pages/20210730_01.aspx.

[89] NSW Health. Omicron variant in confirmed NSW cases. Available at https://www.health.nsw.gov.au/news/Pages/20211128_02.aspx.

Yu Dong received his M.S. degree from Beijing Technology and Business University in 2017. He is currently a Ph.D. candidate at the University of Technology, Sydney (UTS). His research interests include data visualization and visual analytics, machine learning, and human–computer interaction.

Christy Jie Liang received her doctorate degree in data visualization from the University of Technology, Sydney, and now leads the Data Visualization Research Lab in the Visualization Institute at UTS. Her research interests focus on data visualization and visual analytics.

Yi Chen is a professor and director at Beijing Technology and Business University. She received her Ph.D. degree in computer application technology from Beijing Institute of Technology in 2002. Her research interests include visualization, machine learning, and big data technology for food safety.

Jie Hua received his Ph.D. degree in software engineering from the University of Technology, Sydney in 2014. He is currently working as a researcher at UTS and professor at Shaoyang University. His research interests include graph drawing, visualization, big data, and deep learning.