Disease Outbreak Radar: A Tool for Epidemiologists

Cloris Feng (clorisf@student.ubc.ca), Tae Yoon Lee (taeyoon.harry.lee@gmail.com), Derek Tam (derek.tam.94@gmail.com)

1. Introduction

Disease monitoring is an important function of the public health system globally. With the ongoing COVID-19 pandemic, public interest in disease monitoring has risen dramatically in the past months, and various dashboards visualizing disease count data have been made available by government agencies, public health organizations, universities, and hobbyists alike. The dashboards have been developed for multiple purposes: didactic purposes for general public and specific purposes to help healthcare workers effectively and efficiently detect trends in the data and make decisions.

In particular, a critical task for epidemiologists is to detect disease outbreaks, where a disease outbreak is defined as "the occurrence of disease cases in excess normal expectancy" [1]. COVID-19 is one of more than 60 notifiable diseases or disease categories, and these diseases exhibit a variety of trends, such as seasonality in meningococcal disease and a gradual decline in acute hepatitis A. Consequently, it creates a great burden on epidemiologists, who need to consider these complex disease characteristics and trends when deciding whether a reported disease count is higher than expected. Inevitably, the detection process is both fallible and costly. To alleviate the burden on epidemiologists, an automated method has been developed in collaboration with the British Columbia Centre for Disease Control (BC CDC) [2].

The goal of this project is to develop a visualization dashboard specifically designed for epidemiologists to help them visualize the output of the automated method for the task of detecting disease outbreaks. We will first analyze existing dashboards and papers for visualizing disease count data. We will also review guidelines and recommendations that have been published by professionals and academics for both design and ethics considerations. Then we will integrate a set of guidelines, recommendations, and idioms based on our previous analysis and apply them to our dashboard implementation. The dashboard will be developed in R and ShinyApp, given most epidemiologists are familiar with those platforms.

1.1 Personal expertise

Harry is a PhD student in health outcomes with a MSc in Statistics. For his MSc thesis, he worked with the British Columbia Centre for Disease Control to develop an automated method for detecting disease outbreaks. He analyzed their in-house algorithm for detecting disease outbreaks and found areas of improvement, which he incorporated in his automated method. His work was invited for presentation at COMPSTAT 2018. He can program in R, python, Julia, and SPSS. He has made several ShinyApp dashboards to present results for his MSc thesis and to explain the underlying mechanism of a statistical method for didactic purposes.

Cloris Feng is a masters student at the Electrical and Computer Engineering department, with research experiences in applied ML on medical data. She is interested in conducting a project on dashboards regarding to COVID-19 as various COVID dashboards are available online for public interest due to the popularity of the pandemic.

Derek Tam is a masters student in the Bioinformatics department. During his undergrad, he performed field and lab work at the BC CDC studying Lyme disease in the department of Zoonotics and Emerging Pathogens. He has experience programming in R and python, with a particular focus in data science and applied machine learning.

2. Relevant work, analysis and recommendations

There are case studies on existing dashboards designed as disease trackers and for finding correlation between diseases and potential factors related to them. In part 2.1, Kamadjeu et al. conducted a case study for the Somalia Polio Room Dashboard and McLean et al. used a dashboard to evaluate correlations between reported asthma cases and forest fire smoke.

2.1 Relevant work

Kamadjeu et al. designed the Somalia Polio Room Dashboard to track polio outbreak in Somalia, Kenya, and Ethiopia during 2013-2014 [3]. The design aims to address the needs of timely information on cases and performance indicators of the polio outbreak and to provide decision makers with a graphical display of the needed information. The authors utilized composite graphs, maps, and tables to show the trend and geographical information of polio cases, the key factors of polio cases, and essential performance indicators for Acute Flaccid Paralysis surveillance. The authors indicate that the Somalia Polio Room Dashboard is able to provide users with efficient information display and integration.

McLean et al. evaluated the British Columbia Asthma Monitoring System (BCAMS) during the 2014 Forest Fire Season [4]. The BCAMS is to identify abnormal rises of asthma cases due to forest fire smoke by tracking forest fire smoke exposure and the number of asthma cases. The visualization of BCAMS utilizes a tool called Public Health Intelligence for Disease Outbreak (PHIDO), which is a disease outbreak detection algorithm developed by the BCCDC. The authors use composite graphs and maps to display information regarding forest fire smoke exposure and the number of asthma cases. Through the analysis, the authors demonstrate that excursions of asthma cases are related to high concentration of $PM_{2.5}$ in the air which has correlation with the forest fire.

2.2 Analysis

Due to the COVID-19 pandemic, many diverse visualizations and public-facing dashboards have been published in recent months. This provides a unique opportunity to assess and compare the effectiveness and accessibility of these different dashboards. We will analyze the COVID-19 dashboards (shown in Table 1) that have different features as well as a couple of dashboards designed for epidemiologists, PHIDO dashboard [5] and SpatialEpiApp [6]. For the COVID-19 dashboards, we will go over pros and cons of each idiom under various scenarios. Next, we will conduct in-depth analysis of the PHIDO dashboard and SpatialEpiApp by examining the functionality of each feature and idiom. For the final step, we will pick out a combination of idioms and design features that would be best suitable for tasks related to detecting disease outbreaks (see Section 3).

Name	**Idiom**
WHO [7]	Choropleths & bubble maps,barcharts,stacked barcharts
JHU [8]	Tables, bubble maps, bar charts
Bing [9]	Tables, treemaps, colored stacked bar charts, line charts
NORS [10]	Line/bar charts
Show Me Strong [2]	Line over bar charts, map charts

Table 1. A brief overview of dashboards and their corresponding idioms used in each case.

2.3 Guidelines and recommendations

Meta-commentary has also increased as a result of the spike in COVID-19 related dashboards. Numerous guidelines and recommendations have been published by professionals and academics in response, addressing issues with both design and ethics. Here we have compiled a shortlist of formal recommendations as well as informal commentary on the subject.

#	**Name**
1	Amanda Makulec on Medium [11]
2	Evan Peck, Bucknell University on Twitter [12]
3	Lisa Charlotte Rost on Chartable [13]
4	WHO Dashboard Update 2020-04-14 [7]
5	FAIR Guiding Principles [14]
6	Amanda Makulec on FastCompany [15]

7	COVID Tracking Project by The Atlantic [16]
8	Visualizations That Really Work on Harvard Business Review [17]
9	Mapping Coronavirus, Responsibly on ESRI [18]
10	John Burn-Murdoch, Financial Times on Twitter [19]

Table 2. A brief overview of meta-commentary with guidelines and recommendations on visualizing epidemiologic data.

An initial step in this project will be to produce a list of guidelines and recommendations for a generalized outbreak tracker. To do this, we will compile and summarise recommendations made in response to COVID-19-specific visualizations, and assess their practical application in our own brief survey of publicly available dashboards. Combining the two, our team will modify and improve our proposed outbreak detection dashboard, integrating relevant changes and improvements. As the intent of the tool we are implementing is adjacent but not identical to the COVID infographic-style visualizations popularized by the current pandemic, part of the task will involve determining which recommendations are relevant to our tool and how we can use these recommendations to improve a dashboard targeted at epidemiologists rather than laypeople.

3. Data and task abstraction

We will use a publicly available disease count (spatio-temporal) database in the US from Project Tycho [20]. After preprocessing the database, we have a subset of the database that contains no missing values. Description of the finalized database, including derived attributes, is provided in Table 3. The number of cases of each disease is provided on a weekly basis. The derived attributes, week time, and trigonometric basis, are used as the input variables for the automated method to compute a *robust* estimate of the number of cases, an outbreak score, and its corresponding alert level. Finally, a clustering algorithm can be used to detect any spatial correlation for the most recent outbreak scores across the states by proximity.

Attribute name	**Attribute type**	**levels/range**	**Description**
Disease type	Categorical	15	Disease type
Week range	Sequential/ ordered	2009-01-04 - 2014-07-27	Start and end dates of a week in which the number of cases is collected (~234 observations)
US state	Categorical	52	US states
Number of cases	Quantitative	0 - 3140 (integers)	Number of cases of a disease
		Derived attributes	
GPS of state	Spatial	GPS in the US	Centroid of a US state
Week time	Sequential	2009-01-11 - 2014-07-27	The end date of the week range
Trigonometric basis	Cyclic	cos(2 pi 7* t/365.25), sin(2 pi t/365.25) for t=0, …, 365	Trigonometric basis based on the week time
Temperature	Quantitative	Non-negative real valued (in Fahrenheit)	Min and max temperature values for the week range
Estimated number of cases	Quantitative	Nonnegative real valued	The number of cases estimated by the automated method
Outbreak score	Quantitative	0-1 (real-valued)	Probability of observing the number of cases based on the automated method
Alert levels	Categorial	Low, medium, high	Levels of alert specified by epidemiologists based on the outbreak score. For example, an alert level is low, medium, or high if the outbreak score is greater than 0.05, 0.01 - 0.05, or less than 0.01, respectively.
Spatial cluster	Categorial	1-52	Cluster assignment of the states.

Table 3. List of attributes and derived attributes.

The dashboard will be designed specifically for epidemiologists. The main task is to **analyze** whether the current week's reported number of cases of a specific disease is abnormally high. Epidemiologists will **search and query** for a specific disease and carry an evaluation of the current week's reported case based on the estimated trend that is robust to abnormal cases in the data and an outbreak score of the current week's number at the state level. To examine the data at the national level, the cases will be aggregated by the state and the automated method will be used to compute the outbreak score. The secondary task is to **examine** any spatial pattern in the outbreak score across adjacent states.

4. Proposed solution

The final goal of this project is to produce a dashboard for internal use at public healthcare systems, such as the BC CDC, to determine outbreak status for diseases under monitoring. Based on the compiled and summarised recommendations from our review combined with our observations on publicly available dashboards, we will adjust design decisions for the dashboard proposed below.

We plan to implement in R and deploy the dashboard online using the Shiny package.

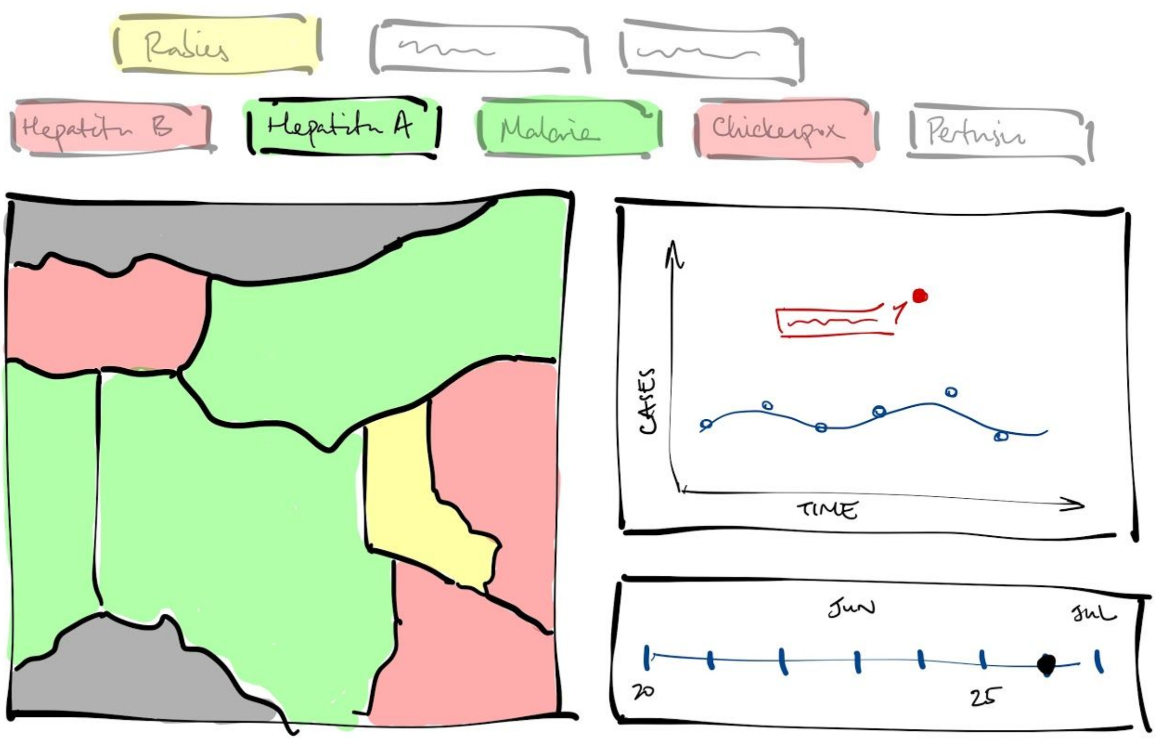

Figure 1. Sketch of proposed dashboard prior to aggregation and integration of recommendations.

4.1 Potential idioms and reasoning

Choropleth map:

- Provides a high level overview for the current outbreak risk of a disease. Doubles as a selection tool for displaying the data on the scatterplot.
- While there are known shortcomings with choropleth maps, this is currently the best compromise in terms of providing a broad overview and as more of a navigational interface.

Scatterplot with fitted curve:

- Provides the detailed case numbers and trends for the inspected disease, and highlights any past outliers as well as the status of the most recent case numbers.
- Hovering over a data point could provide exact case numbers and statistical likelihood of this case number being an outbreak.

Timeline:

- Provides the ability to “scrub” through time, where the choropleth map will update to reflect the date of interest and the scatter plot of case numbers will slide to match the time window of interest.

4.2 Proposed use case

An epidemiologist at the PHAC would like to know if a case number this week is higher than expected in BC and whether this disease is experiencing an outbreak, so as to recommend an adequate response. The epidemiologist can open the dashboard to select BC among the provinces on the choropleth map. The choropleth map will be colored by the most recent case numbers based on outbreak likelihood. Selecting a province will display the time-series data for that province on the right, with the robustly fitted curve indicating the trend for the past *n* weeks, and outbreaks will be highlighted with annotations for case number and significance.

5. Milestones and schedule

Task	Deadline	Time invested (hrs)	Description
Pre-proposal meetings	Oct 15	16	Preprocessing data (HARRY), conducting literature review for recommendation and for vis tools (Cloris and Derek), and preparing a short report for a proposal meeting with Tamara (ALL)

Proposal	Oct 23	16	Meeting with Tamara, working on proposal (ALL)
Analysis	Nov 10	40	Finish and write up the analysis & recommendation sections (Cloris and Derek)
Data analysis (preparatory work)	Nov 10	20	Obtain and compute all the derived attributes (Harry)
Implementation	Nov 14	20	Develop a dashboard following the guidelines and recommendations (ALL)
Project Update	Nov 17	10	Write up (ALL)
Peer Project Review	Nov 19	6	Meeting with other groups (ALL)
Post-update Meeting	Nov 24	5	Meet with Tamara and group discussion (ALL)
Dashboard	Dec 5	60	Develop a final version of the dashboard (ALL)
Final presentation	Dec 10	30	Work on slides, present in class
Final paper	Dec 14	30	Write up paper (ALL)

References

[1] "WHO | Disease outbreaks," *WHO*. https://www.who.int/environmental_health_emergencies/disease_outbreaks/en/ (accessed Oct. 22, 2020).

[2] T. Y. (Harry) Lee, "Robust methods for generalized partial linear partial additive models with an application to detection of disease outbreaks," University of British Columbia, 2019.

[3] R. Kamadjeu and C. Gathenji, "Designing and implementing an electronic dashboard for disease outbreaks response - Case study of the 2013-2014 Somalia Polio outbreak response dashboard," *Pan Afr Med J*, vol. 27, no. Suppl 3, p. 22, 2017, doi: 10.11604/pamj.supp.2017.27.3.11062.

[4] K. E. McLean, J. Yao, and S. B. Henderson, "An Evaluation of the British Columbia Asthma Monitoring System (BCAMS) and PM2.5 Exposure Metrics during the 2014 Forest Fire Season," *Int J Environ Res Public Health*, vol. 12, no. 6, pp. 6710–6724, Jun. 2015, doi: 10.3390/ijerph120606710.

[5] BC CDC, "PHIDO (public health intelligence for disease outbreaks) for Windows: User manual," 2010.

[6] P. Moraga, "SpatialEpiApp : A Shiny web application for the analysis of spatial and spatio-temporal disease data," *Spatial and Spatio-temporal Epidemiology*, vol. 23, pp. 47–57, Nov. 2017, doi: 10.1016/j.sste.2017.08.001.

[7] "WHO updates COVID-19 dashboard with better data visualization." https://www.who.int/news-room/feature-stories/detail/who-updates-covid-19-dashboard-with-better

-data-visualization (accessed Oct. 22, 2020).
[8] "COVID-19 Map," *Johns Hopkins Coronavirus Resource Center*. https://coronavirus.jhu.edu/map.html (accessed Oct. 22, 2020).
[9] "Microsoft Bing COVID-19 Tracker." http://bing.com/covid/local/maine_unitedstates?dynamicSharing=true&shtp=Facebook&shwth=900&shh=800&shtk=Y29yb25hdmlydXMgdHJhY2tlciB1cGRhdGVz&shdk=dGVzdA%3D%3D&shth=OSH.Mmq%2BwuM5WWl/TcdViNGxBA&redirect_uri=http%3A//veeraux%3A81/covid/local/unitedstates%3FdynamicSharing%3Dtrue&ref=Coronavirus&al (accessed Oct. 22, 2020).
[10] "National Outbreak Reporting System (NORS) Dashboard | CDC." https://wwwn.cdc.gov/norsdashboard/ (accessed Oct. 22, 2020).
[11] A. Makulec, "Ten Considerations Before you Create another Chart about COVID-19," *Medium*, Apr. 27, 2020. https://medium.com/nightingale/ten-considerations-before-you-create-another-chart-about-covid-19-27d3bd691be8 (accessed Oct. 22, 2020).
[12] "EvanMPeck on Twitter," *Twitter*. https://twitter.com/EvanMPeck/status/1235568532840120321 (accessed Oct. 22, 2020).
[13] D. GmbH, "17 (or so) responsible live visualizations about the coronavirus, for you to use," *Chartable*, Mar. 06, 2020. https://blog.datawrapper.de/coronaviruscharts/index.html (accessed Oct. 22, 2020).
[14] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci Data*, vol. 3, no. 1, p. 160018, Dec. 2016, doi: 10.1038/sdata.2016.18.
[15] "A complete guide to coronavirus charts: Be informed, not terrified." https://www.fastcompany.com/90477393/a-complete-guide-to-coronavirus-charts-be-informed-not-terrified (accessed Oct. 22, 2020).
[16] "Visualization Guide," *The COVID Tracking Project*. https://covidtracking.com/about-data/visualization-guide (accessed Oct. 22, 2020).
[17] S. Berinato, "Visualizations That Really Work," *Harvard Business Review*, no. June 2016, Jun. 01, 2016.
[18] "Mapping coronavirus, responsibly." https://www.esri.com/arcgis-blog/products/product/mapping/mapping-coronavirus-responsibly/ (accessed Oct. 22, 2020).
[19] J. Burn-Murdoch, "John Burn-Murdoch (@jburnmurdoch) / Twitter," *Twitter*. https://twitter.com/jburnmurdoch (accessed Oct. 22, 2020).
[20] W. G. van Panhuis *et al.*, "Contagious diseases in the United States from 1888 to the present," *N Engl J Med*, vol. 369, no. 22, pp. 2152–2158, Nov. 2013, doi: 10.1056/NEJMms1215400.