

Teaching Note

DATA ANALYTICS FOR EMERGENCY PREPAREDNESS AND RESPONSE DURING THE COVID-19 PANDEMIC IN MINNESOTA

Jeffrey Clement wrote this case solely to provide material for class discussion. The author does not intend to illustrate either effective or ineffective handling of a public health situation. The "problems" with the data set were deliberately introduced for the purpose of identifying and fixing them. Development of this case was supported by an Emergency Preparedness and Response Curriculum Development Project Grant from the Minnesota Private College Fund in partnership with the Minnesota Department of Health.

This case is released under the [Creative Commons CC BY-NC-SA 4.0 license](https://creativecommons.org/licenses/by-nc-sa/4.0/). You may share, copy, redistribute and adapt the material but must give appropriate credit and may not use the material for commercial purposes.

Version: 2024-08-12

SYNOPSIS

Data literacy is a powerful tool, especially for those responding to public health emergencies. Data analysis helps agencies deploy resources appropriately, maximizing their impact, and detecting disparities (e.g., disproportionate impacts along socioeconomic lines). Data literacy and analysis are tightly linked to at least two of the CDC's fifteen key capabilities (surveillance and informatics).

This case asks students to evaluate a fairly straightforward question: Was the prevalence of Covid-19 over time similar or different between urban (metropolitan and micropolitan) versus rural areas in Minnesota? We use this question as an opportunity to introduce how data analytics can support public health and emergency preparedness and response (EPR) and some key pitfalls around data analytics.

INCLUDED FILES

- Data Analysis for EPR Teaching Case.pdf: Student Case
- Data Analysis for EPR Teaching Case - Teaching Note.pdf: Teaching Note (Do not provide this to students)
- MN DHS County Case Data - Student Version.xlsx: Dataset for students to complete the demo in the case
- MN DHS County Case Data – Demo.xlsx: Dataset with the issues in the case corrected (Do not provide this to students)
- Covid-19 Case - Sample Tableau Workbook.twbx: Sample Tableau Workbook built from the demo instructions in the case (Do not provide this to students)

LEARNING OBJECTIVES

This case discusses the alignment between organizational structure, operational management, and technology, particularly how that may change over time. The case also opens opportunities to discuss organizational fragility and resilience, as well as public relations in technology-driven crises.

Gain an appreciation for how data analysis can inform emergency preparation and response

This case provides an opportunity to look at how data is used to allocate scarce resources. Depending on the student makeup, instructors can discuss specific examples and experiences relevant to their specialty.

Discuss data quality issues like missing data

Data analysis requires quality data, which is notoriously hard to collect during public health emergencies. Understanding the limitations of data—and communicating those limitations to people in leadership—is an important part of being a data analyst.

Introduce key considerations around data analysis like cybersecurity

Cybersecurity and data backups are often considered the domain of the IT team, but everyone has a role in protecting the organization, especially when it comes to protected health information (PHI). These topics do not fit “neatly” in many other professional discussions, so we introduce them here.

Apply Tableau Data Visualization Software to Answer an Analytics Question

Tableau is one of the most common data visualization suites in use (along with PowerBI and more complicated tools that require coding like R).

TECHNOLOGY REQUIREMENTS

The examples in the case are presented with Tableau version 2024.1. Tableau offers a free license for students (<https://www.tableau.com/academic/students>) as well as free licenses for academic courses for computer labs and students (<https://www.tableau.com/academic/teaching>).

Several steps require spreadsheet software. Examples are presented in Microsoft Excel but Apple Numbers, Google Sheets, LibreOffice Calc (free) would work fine.

POSITION IN COURSE

This case can be used for discussion in the following courses:

- Introductory undergraduate courses on public health, nursing, or epidemiology, where the goal is to introduce public health decisionmaking and functions
- Undergraduate courses on data analytics (non-health related fields); this case can introduce students who may not have thought about EPR or healthcare fields as a career option
- Executive education courses introducing data-driven managerial decision making.

DISCUSSION QUESTIONS

Discussion Question 1: Can AI tools like ChatGPT or Google Gemini replace public health professionals? Can they augment or assist with EPR tasks?

Discussion Question 2: What could happen if a public health organization suffered a ransomware attack and was unable to access any of their data? How could they mitigate this risk?

Discussion Question 3: What are some things that could cause some data points to be completely missing (like this county that missed one of the weekly case counts)? As far as you can tell, is this data point more likely to be Missing Completely at Random (MCAR) or Missing Not at Random (MNAR)?

Discussion Question 4: Even though we have the same number of data points for each county (weekly reports) this dataset almost certainly is missing data because the case counts are too low. Why are the case counts probably too low? Is that data more likely to MCAR or MNAR?

Discussion Question 5: The original dataset includes each county as well as a row labeled “Total” that is the sum of each individual county’s cases. What is the effect of this row on our analysis up to this point? In other words, when this row is included, there are approximately 700k cases in Jan 2022, but what should the right number be?

Discussion Question 6: Why does Hennepin County probably have so many more cases than the rest of the state?

Discussion Question 7: Are trends in Covid-19 cases different between rural and urban areas? Does it make more sense to talk about case count or prevalence (count per 10k people) or are both important?

TEACHING PLAN

This case is appropriate for a 120-minute session depending on student skill level. This case also lends itself to much shorter discussions if the scope of discussion is appropriately limited—for example, the instructor can lay out the groundwork and background that sets up the discussion of one or more the Discussion questions and then assign the entire Tableau portion of the assignment to students as homework or group work.

Discussion Point	Time (Minutes)
Introduction	5 Min
Role of Analytics in Supporting EPR Functions	15 Min
Analytics Tools and Infrastructure	15 Min
Tableau Demo Part 1 (through Fig 8)	30 Min
Missing Data and Difficulty in Collecting Data	15 Min
Tableau Demo Part 2	30 Min
Wrap-up and Final Answer	10 Min

ANALYSIS AND DISCUSSION GUIDE

One way to approach teaching this case is to go from the specific to the general—start with how data analytics can support public health and EPR, and then move into the specifics of the case and how to use Tableau to analyze the data.

Analytics for Public Health and EPR

This case offers an opportunity to discuss how data analytics can support EPR and public health. Especially for students in non-healthcare fields where one goal is to introduce the idea of a healthcare career, A possible opening/cold call question could be:

(Evaluation Question)

How can data analytics help in EPR?

Based on the specialty of the students in question (e.g., MPH vs. School Nurses vs. Epidemiology vs. Non-healthcare Students), you can focus on different topics or draw on your own experiences. The key is to focus on the goals of EPR, emphasizing what must be done. From there, you can get at one of the key learnings from the case, which is that data analytics can *inform and guide* EPR but cannot actually accomplish any of the functions. This will lead to Discussion Question #1.

(Discussion Question #1 – Evaluation Question)

Can AI tools like ChatGPT or Google Gemini replace public health professionals? Can they augment or assist with EPR tasks?

As you discuss these questions, it might be effective to step back and look at the goals of EPR and the types of tasks that are done in a given EPR role.

Analytics Tools and Infrastructure

After discussing data analytics in general, you can start to talk about how it is implemented. The data analysis cycle is one approach, but feel free to introduce your own experiences—particularly cases where there have been missteps or problems arising from data collection. Talk about the data tools that you have used, either as an analyst or as a data user. While the case uses Tableau, the best data analytics tool is the one that gets the job done that the organization uses, and most people will use multiple tools throughout the course of their career.

Then you can pivot to how organizations structure data resources. This is often considered the domain of IT or “someone is taking care of it.” A relevant joke is, “Unless the organization employs someone whose first name is ‘Someone,’ then ‘Someone is taking care of this,’ means ‘Nobody is taking care of this.’” Everyone is responsible for cybersecurity and ensuring their data is backed up. That might mean just verifying things, but it could also mean taking personal action. And nearly every organization has a story about people bypassing IT controls (e.g., by storing data on a personal hard drive) resulting in either a security breach or data not being backed up. This information does not often fit elsewhere in many curricula, so we include it in this case as Discussion Question 2:

(Discussion Question 2 – Evaluation Question)

What could happen if a public health organization suffered a ransomware attack and was unable to access any of their data? How could they mitigate this risk?

Monitoring Covid-19 Case Counts in Minnesota

Next, we pivot to the case at hand and walk through the demo in Tableau. It is important to discuss the data source with students. Depending on the student population, you may want to take a deep dive with students into the Data Dictionary. You could talk in some detail about the `spec_date_mmwr`, the MMWR week in the dataset, and discuss the importance of standardizing that nationwide:

(Descriptive Question)

What is MMWR date and how is it specified? Does it perfectly align with calendar years and dates?

At any rate, it is worth discussing where data comes from, and how messy data could cause problems.

Tableau Demo

Then, begin the Tableau tutorial, walking through the steps in the case starting on Page 5.

The first few steps focus on loading the data and looking at the number of data points (MMWR Weeks) reported by each county. There should be 169, but one county has 168 and one has 170. This introduces the idea of missing data.

Missing Data

Having identified that one county is missing a data point (i.e., not reporting for a particular MMWR Week), you can discuss why that might happen. As noted in the case, this issue was introduced by the authors for the purpose of the case and was not a problem caused by the Minnesota Department of Health!

(Discussion Question 3 – Evaluation Question)

What are some things that could cause some data points to be completely missing (like this county that missed one of the weekly case counts)? As far as you can tell, is this data point more likely to be Missing Completely at Random (MCAR) or Missing Not at Random (MNAR)?

This type of error, if infrequent, is probably MCAR due to, e.g., someone going on vacation or a random technical issue.

Then go through and fix the issue with duplicate reporting by Anoka County.

After that, discuss the idea of missing data more generally. While there is data for each MMWR week for each county, there are likely many Covid-19 *cases* that are unreported and not collected in the data, which leads to Discussion Question 4:

(Discussion Question 4 – Evaluation Question)

Even though we have the same number of data points for each county (weekly reports) this dataset almost certainly is missing data because the case counts are too low. Why are the case counts probably too low? Is that data more likely to MCAR or MNAR?

These may be MCAR or MNAR (or potentially MAR, but that tends to be a more complicated topic for beginners). There is no set answer, but there is a strong case to be made that case underreporting is MNAR—people who are better off financially and in better health are probably in a better position to seek testing, and potentially better able to take time off work (or be working from home and not needing to avoid a positive test to avoid missing work).

Tableau Demo, continued.

After addressing the missing data issues, continue with the Tableau demo. The next major issue that pops up is the original dataset (as distributed by MN Department of Health) includes a “county” called “Total” which adds up all of the other counties. That leads to Discussion Question 5:

(Discussion Question 5 – Evaluation Question)

The original dataset includes each county as well as a row labeled “Total” that is the sum of each individual county’s cases. What is the effect of this row on our analysis up to this point?

In other words, when this row is included, there are approximately 700k cases in Jan 2022, but what should the right number be?

The issue here is that Total *already* adds up all of the other counties, so the number reported is twice the true number. After filtering out the county called “Total,” the number will be correct.

Beginning to analyze things by county, the graph looks like many flat lines because the Y-axis all have the same range. Hennepin County (and a few others) look like we might expect but the others are flat lines at the bottom of the range because Hennepin has more cases.

(Discussion Question 6 – Evaluation Question)

Why does Hennepin County probably have so many more cases than the rest of the state?

This is, most likely, a population effect. Hennepin has more residents so there are more cases of Covid-19. This is perhaps the first part of the main question we have to answer, but we need to go deeper.

The Tableau demo continues, joining the demographic data, and analyzing the data by population density (according to US federal OMB criteria).

Looking at Figure 22, we can answer the final discussion question:

(Discussion Question 7 – Evaluation Question)

**Are trends in Covid-19 cases different between rural and urban areas?
Does it make more sense to talk about case count or prevalence (count per 10k people) or are both important?**

Overall, the trends were similar over time in terms of prevalence (cases per 10k residents). However, the larger counties had more cases, so both metrics are important to consider.