Engaging students with COVID-19 data

Maria Tackett and Mine Çetinkaya-Rundel

July 2020

COVID-19 has undoubtedly generated a plethora of data, from daily counts on tests, positive results, hospitalisations, deaths, etc. to data from scientific studies on the virus and treatment to data on the impact of the pandemic on the economy, education, mental health, etc. Engaging statistics and data science students with these data seems obvious, but the decision of whether to bring COVID-19 data into the statistics and data science classrooms is not simple. On one hand, the answer might seem like an obvious *yes* – what better way to engage students with real data than data about the pandemic that has turned all of our lives upside down. On the other hand, at least at the time of writing this column, we're still in the midst of a pandemic that has taken many lives and, indeed, turned all of our lives upside down, which might be the very reason to to say *no* to bringing these data into the classroom.

In this column we showcase a few approaches for bringing COVID-19 data into our teaching as well as highlight resources that might help you decide whether to do so in the first place.

DataFest: COVID-19 Virtual Data Challenge

The American Statistical Association (ASA) DataFest is a data analysis competition where teams of up to five students analyze a real and complex dataset over the course of one weekend. In 2019, DataFest was held at over 40 locations in the United States and internationally with more than 2000 students participating in the event. The surprise data set is revealed to participants at the kick-off event on Friday afternoon, and students work throughout the weekend analyzing the data and deriving insights. On Sunday afternoon, the groups present their work to a panel of judges made up of instructors and data science professionals in industry. By the end of the DataFest weekend, students have not only gained experience analyzing real data, they have also practiced presentation skills, all while connecting with other students, faculty members, and industry professionals.

In March 2020 as many colleges and universities transitioned to a remote format, the ASA DataFest steering committee considered alternatives for this year's competition. The goal was to adapt ASA DataFest to the new environment, yet still include the parts of the event that make it an inviting and valuable experience for students all students with a wide range of statistics and data analysis experience. The 2020 ASA DataFest ran as a virtual data challenge where students worked in teams to explore an impact of the COVID-19 global pandemic. Given the variety of potential topics, part of what made this year's challenge unique was that it involved participants finding a data set for their analysis.

DataFest events were held in April through June, a time when data and modeling about the direct health outcomes of the pandemic was rapidly changing and unreliable (see *Why It's So Freaking Hard To Make A Good COVID-19 Model*[1]). Building models and drawing reliable conclusions about infection, mortality, or recovery rates require understanding the nuances and limitations of the COVID-19 health data at a level that would likely not be feasible in the short span of the DataFest competition. Therefore, participants were advised to "tell us about something affected by the COVID-19 pandemic other than its direct health outcomes", to discourage them from presenting conclusions that could be potentially misleading or harmful.

[1]Koerth, M., Bronner, L. and Mithani, J., 2020. Why It'S So Freaking Hard To Make A Good COVID-19 Model. [online] FiveThirtyEight. Available at: https://fivethirtyeight.com/features/why-its-so-freaking-hard-to-make-a-good-covid-19-model/ [Accessed 8 July 2020].

A few suggested analysis questions included

- How has the pandemic affected the airline industry and what are some potential downstream effects of this other than economic strain on the industry?
- As a student, how would you quantify the effect of the pandemic on your education?
- With shelter in place / lockdown in place, many workers have started working from home, which requires internet access. How prepared was the nation / your local area for this shift?
- How has the spread of the pandemic affected people's opinion on government tracking and privacy?
- What is the effect of the social distancing / shelter in place / lockdown recommendations and policies on pollution?
- How can we quantify the potential effects on nutrition and general health of the public, outside of those affected by the virus? How are refugees affected by COVID-19?
- Examples of students' data sets and analysis questions?
- Description of some of the winning projects?
  - *Ten Considerations Before You Create Another Chart About COVID-19*

Using COVID-19 data in the classroom

At the May 2020 Electronic Conference on Teaching Statistics (eCOTS), Laura Le, Kari Lock Morgan, and Lucy McGowan presented Engaging Students during the COVID-19 Health Crisis about how to incorporate data related to the COVID-19 pandemic in the classroom. The primary focus was that the pedagogy should be "trauma-informed" due to the direct impact on students. By taking this trauma-informed approach instructors can create a classroom environment where students feel safe to discuss the subject and reduce risk of retraumatizing students impacted by the pandemic[2]

The panelists shared practical ways instructors can use a trauma-informed approach when discussing data related to the pandemic in class:

- Anonymously poll students about whether or not they want to talk about the data related to COVID-19 in class. It is a good idea to poll multiple times to get point-in-time feedback, since students' feelings may change as the situation around the pandemic evolves.
- Indicate in the syllabus when COVID-19 data will be used, so students know when the topic will come up in class.
- Create an alternative assignment or discussion prompt for students who do not wish to discuss the pandemic.
- If the course is designed to a more specialized audience, such as a biostatistics or graduate-level course, the instructor can address the fact that the topic is sensitive but is also an important area of research in biostatistics. This is also an opportunity for the instructor to talk about strategies for maintaining a healthy relationship with emotions when doing research on sensitive topics.
- As with this year's DataFest, the analysis examples can focus on societal impacts of the pandemic other than direct health outcomes.

Example activities

One way to use COVID-19 data in the classroom is by having students think about effective data visualizations. Since the beginning of the pandemic, there have been a variety of data visualizations used by government organizations, news media, and the public outlets to help the general public better understand the pandemic

[2]Abuelezam, N, (2020). *Teaching Public Health Will Never Be the Same.* American Journal of PUblic Health, 110(7), 976-977. https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2020.305710

(e.g. the "flattening the curve" visualizations). [Insert why it's important for students to understand how to interpret and critique these visualizations.]

Activity outline:

1. Critique existing visualizations using *Ten Considerations Before You Create Another Chart About COVID-19* [and OTHER RESOURCES?] as a guide. The critiques should address what makes the visualizatin effectiv and what is confusing or even misleading.
2. Students can create their own visualization taking into account the tips and critiques they discussed. This gives students an opportunity to hone their computing skills as they wrangle the data and create visualizations using statistical software. Students can design their own visualizations or replicate existing ones as done in the I Eat Data Science for Breakfast: pandemic 2020 edition workshop.

Resources for teaching

1. Teaching Statistics During the COVID-19 Health Crisis
2. covid-r: Collection of analyses, packages, visualisations of COVID19 data in R
3. I Eat Data Science for Breakfast: pandemic 2020 edition

Further reading

- Koerth, M., Bronner, L. and Mithani, J., 2020. Why It'S So Freaking Hard To Make A Good COVID-19 Model. [online] FiveThirtyEight. Available at: https://fivethirtyeight.com/features/why-its-so-freaking-hard-to-make-a-good-covid-19-model/ [Accessed 8 July 2020].
-