



Data Discovery Challenge Using the COVID-19 Data Portal from New Zealand

Jon MacKay

To cite this article: Jon MacKay (2022) Data Discovery Challenge Using the COVID-19 Data Portal from New Zealand, Journal of Statistics and Data Science Education, 30:2, 187-190, DOI: [10.1080/26939169.2022.2058656](https://doi.org/10.1080/26939169.2022.2058656)

To link to this article: <https://doi.org/10.1080/26939169.2022.2058656>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 12 May 2022.



[Submit your article to this journal](#)



Article views: 658



[View related articles](#)

Data Discovery Challenge Using the COVID-19 Data Portal from New Zealand

Jon MacKay 

Department of Information Systems and Operations Management (ISOM), University of Auckland, Auckland, New Zealand

ABSTRACT

Students need to know how to discern patterns and make decisions using visual information in our modern economy. However, there are few sources of real-world information available to instructors that give students access to visualizations to help develop their skills in interpreting complex situations using diverse data sources. This article outlines a teaching exercise that uses the New Zealand government's data portal. This website contains detailed time series data and visualizations that span economic, social and health data derived from multiple government ministries and New Zealand businesses. The portal continues to be used by government decision-makers to make real-time decisions about the nation's economy and citizen well-being. Typically, statistical agencies carefully vet the data they supply. The data portal prioritizes the timeliness of the information for decision-makers working in a crisis. This brief communication outlines an exercise for students to explore and interpret data through visualizations.

KEYWORDS

COVID-19; Data visualization; Decision making; Exercise; Open data; Pandemic

1. Introduction

Considerable research argues that students learn statistical thinking by actively interacting with data to answer questions or make decisions (Carver et al., 2016; Snee, 1993; Wild, 1994). However, textbook examples and case studies tend to be framed narrowly and don't effectively challenge students to make inferences about complex systems. Such textbook exercises often allow students to infer the "correct" answer or method based on the chapter they are working on (Chance, 2002). In the real world, however, problems never arrive neatly packaged. Practitioners need to solve problems without the assurance of a single correct answer.

Few resources exist for teachers to expose students to real-world visualizations that allow students the freedom to consider how many variables are related to one another without getting bogged down in technical distractions. This teaching note fills that gap by outlining an exercise using a data portal produced by Statistics New Zealand (Stats NZ 2020). The country's top decision-makers continue to use the data visualizations on this web portal to make policy during the pandemic. This exercise exposes students to visual data interpretation and storytelling using actual, complex data.

The COVID-19 Data Portal URL is <https://www.stats.govt.nz/experimental/covid-19-data-portal>. The portal has been so successful that Stats NZ plans to release a permanent version for data sharing after the pandemic ends.

1.1. Interpreting Visualizations and Statistical Thinking



Although many students lack in-depth statistical training, they will eventually need to think critically about statistical issues. In


other words, they will need statistical thinking skills to effectively contribute to their organizations and even make effective decisions as citizens. The Guidelines for Assessment and Instruction in Statistics Education (GAISE) College report notes that statistical thinking means students should understand the need for data, think critically about issues using quantitative data, and explain variability (Carver et al. 2016, p. 12).

Chance (2002) reviews how researchers have formally defined statistical thinking. She points out that statistical thinking emphasizes seeing processes as part of a whole. Statistical thinking, then, is an iterative view of problem-solving (Wild 1994). Thinking statistically has been likened to a feedback loop (Box et al. 1978). Hypotheses are developed, deductions made, data is used to test hypotheses, and then the findings themselves are fed back to generate new, more accurate hypotheses. This paradigm emphasizes understanding the context of a problem, including the nonstatistical background and its relationship with statistical theory. Making statistics courses more activity-based has been an increasingly popular way to get students to think statistically (King 2000; Schaeffer et al. 2000). Activities allow students to interpret and analyze data to generate conclusions.

All modern organizations face situations in which they need to find and interpret patterns in data. One of the primary skills they need to develop is determining what data is relevant to the problems they face (Snee 1990). This process of determining the scope of the problem is what Mallows (1998) called the zeroth problem. Textbook exercises, by their nature, do not allow students to develop this essential skill.

For educators, the challenge is often finding data that will excite students about the subject. Textbook examples rarely contain the breadth or depth students are likely to face in the real

CONTACT Jon MacKay  jon.mackay@auckland.ac.nz  Department of Information Systems and Operations Management (ISOM), University of Auckland, 12 Grafton Road, Auckland, 1010 New Zealand.

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/ujse.

© 2022 The Author(s). Published with license by Taylor and Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

world. In my experience, overly simplified examples or stand-alone data sets quickly bore students who turn off because they cannot see the “real-world” purpose in an exercise.

Several techniques involve students in statistics-based project learning (Moore 2000). Such projects ideally allow students to define questions, generate hypotheses to answer those questions, design experiments or surveys and collect data. Students can then measure their data, analyze it, summarize it and communicate their findings.

For most people, interpreting data visualizations is an essential skill. Visualizations are often the first exposure people have to new information. More information is available now than at any other time in human history. Interpreting visualizations allows people to explore new information, generate and test simple hypotheses (Berinato 2016; Unwin 2020). The interactive visualizations are one reason why the COVID-19 Data Portal became so frequently used by New Zealand’s government decision-makers.

1.2. Introduction to the Data Portal

New Zealand is a small country in the South Pacific with just over 5 million people, spread across two large islands. One-third of the population lives in the city of Auckland on the North Island. The country exports mainly resource-based products such as dairy, meat, lumber and fruits. The top industries that contributed to the country’s GDP in 2018 were professional, scientific and technical services; real estate and construction and property-related businesses; and financial and insurance services (see supplementary materials for more resources about New Zealand.)

New Zealand’s initial response to the COVID-19 pandemic was to close its borders and begin a lockdown to eliminate the virus. The strict lockdown that began in late March 2020 eventually eliminated the virus but also largely stopped major sectors of the economy. Once the lockdown began, policymakers were left without any reliable national statistics to judge how the economy continued to perform. The early success of the elimination strategy meant that the country was initially spared from a prolonged shutdown of its economy. Periodic outbreaks resulted in a series of smaller lockdowns. However, by October 2021, the government lost control of the virus and abandoned the elimination strategy.

In response to the needs of several different policy departments, Stats NZ implemented a web portal containing data and visualizations from government and private sector sources that agreed to make their data available for public use. It posted the information as open data to serve as many different stakeholders as possible. The portal’s goal was to make its data readily interpretable so that users did not have to take the time to visualize the data themselves. To this end, analysts at Stats NZ made a web application to show visualizations of the data listed on the portal.

Initially, just a few time series were posted on the open data website. However, the website’s popularity encouraged Stats NZ to continue adding to and updating the website even as the pandemic subsided in the country. There are currently over 100 data files and visualizations available.

The data portal has five top-level thematic groups containing economic, health, income support, social and environmental data visualizations. The economic theme tends to have the most up-to-date data from various public and private sources. The other themes can be helpful for students who are studying in related areas.

The dynamic web interface allows users to move a slider to change the date range of the data visualization displayed. This interface allows students to focus tightly on events such as lockdown periods or take a longer view of a given time series.

2. Exercise: Data Discovery Challenge

The data discovery challenge is an introduction to data visualization interpretation and analysis. I’ve used this exercise successfully both with professionals returning to school and students with more statistical training. Members of the former group often need to hone their data analysis skills for their work, where they often need to interpret data visualizations. This exercise is also helpful for more statistically advanced students because it allows them to think about the underlying relationships between variables and their human implications.

2.1. An Outline of the Exercise

For this exercise, I recommend dividing a class into teams. Each team is assigned to look at one broad category of data. (Alternatively, the student teams can represent different interest groups and explain how their sector performs. This second approach is in a separate exercise sheet in the Appendix, supplementary materials.) Each team must try to discern important trends within that data category and explain the practical impact in the data they see. When the students are done, groups can present their findings to the class. This final class presentation encourages students to think about how patterns within the data can be brought together to create coherent and interesting narratives.

I usually ask student groups to focus on the economic theme in the data portal (the default tab on the website) and assign one of the major categories: Activity, Employment, Financial and Visa, although any set of categories can be chosen (see exercise outline in the Appendix, supplementary materials for other suggestions). Within these different categories are multiple time series visualizations. Looking at the different data visualizations allows students to develop their interpretation and analysis skills. For student teams to tell a coherent story, they need to collectively make sense of the changing data patterns and construct a compelling narrative around them.

When groups present to the class, I ask them to justify their interpretations based on the visual information. The process of explaining the visualization is helpful because students can see how their peers interpret data and contextualize it. Less sophisticated answers may interpret the data visualization directly and repeat the pattern. Better answers will consider the follow-on implications and connect them with other insights. In other words, they will begin to see the myriad of data visualizations of different representations as representations of an interrelated system.

Some students see themselves as weaker at quantitative subjects and better at more humanistic pursuits. Interpreting data visualizations and constructing narratives requires multiple different skill sets. I've found that challenging students to think about who would be interested in their results and why their insights are essential to those stakeholders helps make the exercise more realistic.

Of course, narratives can also be misleading. Instructors need to be prepared to challenge interpretations or offer counter-explanations for the same phenomenon. Sometimes students who have examined other aspects of the data portal will be able to point out inconsistencies with the interpretations of their peers.

To help ensure that student interpretations do not stray too far from reality, instructors can use the background materials supplied in the Appendix, supplementary materials to give students an overview of New Zealand's economy. Instructors can also ask student groups to verify their findings using a basic internet search or provide more detailed background materials. The supplementary materials section of this article contains a brief write-up about New Zealand's economy and links to additional resources, including handouts and PowerPoint slides.

2.2. Time Series

Depending on the experience level of participants, I give them more or less structured assignments. More experienced graduate students who have already taken introductory statistics, finance or economics courses will interpret multiple different time series, only needing assistance on more nuanced interpretation. Even so, I recommend that the instructor works through an example, covering what a time series is, how data is visualized in a time series with the variable of interest on the vertical axis and time along the horizontal axis. Basic concepts like the trend, seasonal patterns, cyclical patterns and random noise can all be explained to students. Breaking down the components of a time series will give students a better sense of precisely what they are looking at in a time series plot.

Some instructors may be interested in the sharp changes in the time series that occur when lockdowns begin and end. For those who prefer to focus on those structural breaks in the series, detailed information about the lockdown dates and regions affected are in the Appendix, supplementary materials.

A good time series to use as an example is monthly card transaction spending for the retail industry. This is an intuitive visualization of increasing consumer expenditures over time that contains a striking structural break due to the first 2020 lockdown. This visualization is also helpful because it shows a clear upward trend and seasonal patterns, making it convenient to discuss some reasonably typical time series patterns.

2.3. Learning Goals

The open-ended nature of the exercise encourages students to explore the data portal. The learning goals of this exercise are consistent with the outcomes proposed in the GAISE report (Carver et al. 2016, p. 8). The exercise encourages students to become critical information consumers by selecting and interpreting multiple relevant data visualizations. In each case,

students need to appropriately determine what is relevant and what is not by investigating the data using visualizations. The interpretation of the visualizations represents a preliminary investigation of the causes and relationships between different time series. Students will need to develop hypotheses about multiple variables and their relationships, then use the visualizations to help support or contradict those hypotheses or consider other potential steps in their analysis.

3. Conclusion

In this article, I've explained how the data and visualizations from the New Zealand data portal can be used as part of a teaching exercise for students at various levels. The portal provides a broad selection of information and visualizations of the underlying data. The website shows information based on the user's choice from a few drop-down menus. There is no hierarchy to the information presented beyond classifying measures under a few broad themes.

The unstructured data presentations make it ideal for students to explore, interpret, and eventually report to the class on what they have found. In my experience, this type of unstructured learning environment is rare in academic contexts. The exercise provides students with an open-ended data analysis experience not unlike what they are likely to experience in business or government agencies.

The large amounts of unstructured data make it an ideal setting to use to challenge students to interpret data visualizations to understand better the complex economic and social implications of the global pandemic on a small country's economy.

ORCID

Jon MacKay  <http://orcid.org/0000-0003-1544-4168>

References

- Berinato, S. (2016), "Visualizations That Really Work," *Harvard Business Review*, 94, 93–100.
- Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York: Wiley.
- Carver, R., College, S., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Rowell, G. H., Velleman, P., Witmer, J., and Wood, B. (2016), Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016, 141. [https://www.amstat.org/education/guidelines-for-assessment-and-instruction-in-statistics-education-\(gaise\)-reports](https://www.amstat.org/education/guidelines-for-assessment-and-instruction-in-statistics-education-(gaise)-reports)
- Chance, B. L. (2002), "Components of Statistical Thinking and Implications for Instruction and Assessment," *Journal of Statistics Education*, 10. <https://doi.org/10.1080/10691898.2002.11910677>
- King, B. (2000), "A Companion Piece to Activity-Based Statistics," in *Teaching Statistics: Resources for Undergraduate Instructors*, eds. T. L. Moore, pp. 71–76, Washington DC: Mathematical Association of America: American Statistical Association.
- Mallows, C. (1998), "The Zeroth Problem," *The American Statistician*, 52, 1–9. <https://doi.org/10.1080/00031305.1998.10480528>
- Moore, T. L., ed. (2000), *Teaching Statistics: Resources for Undergraduate Instructors* (Vol. 52), Washington DC: Mathematical Association of America: American Statistical Association.
- Schaeffer, R. L., Gnanadesikan, M., Watkins, A., and Witmer, J. A. (2000), "Excerpts from Activity-Based Statistics," in *Teaching Statistics: Resources for Undergraduate Instructors*, eds. T. L. Moore, pp. 61–70, Washington DC: Mathematical Association of America: American Statistical Association.

- Snee, R. D. (1990), "Statistical Thinking and Its Contribution to Total Quality," *The American Statistician*, 44, 116–121.
- Snee, R. D. (1993), "What's Missing in Statistical Education?" *The American Statistician*, 47, 149–154. <https://doi.org/10.2307/2685201>
- Stats NZ. (2020), "COVID-19 Data Portal," COVID-19 Data Portal. Available at <https://www.stats.govt.nz/experimental/covid-19-data-portal>
- Unwin, A. (2020), "Why is Data Visualization Important? What is Important in Data Visualization?" *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.8ae4d525>
- Wild, C. J. (1994), "Embracing the "Wider View" of Statistics," *The American Statistician*, 48, 163–171. <https://doi.org/10.1080/00031305.1994.10476049>