Dataset Analysis: Skill Practice for Graduate Students

Rebecca G. Lawver
Department Head & Professor
Applied Sciences, Technology, and Education
Utah State University
1498 N 800 E | ASTE 101C
UMC 2300
Logan, UT 84321
(435) 797-1254
rebecca.lawver@usu.edu

Michael Spiess
Professor Emeritus
California State University, Chico
College of Agriculture
Chico, CA 95929-0310
(530) 521-6866
mspiess@csuchico.edu

Ashley Rogers
Program Manager
Professional Development & Teacher Recruitment and Retention
National Association of Agricultural Educators
859-967-2895 (office)
arogers.NAAE@uky.edu

Amy R. Smith
Associate Professor and Program Leader
University of Minnesota
146 Ruttan Hall
St. Paul, MN 55108
(612) 624-6590
arsmith@umn.edu

Daniel D. Foster
Associate Professor
The Pennsylvania State University
212 Ferguson Building
University Park, PA 16802
(814) 863-0192
foster@psu.edu

Dataset Analysis: Skill Practice for Graduate Students

Introduction

Education as an industry generates huge amounts of data (Fergison, 1017). A wide range of public data is available from federal agencies (ex. National Center for Education Statistics), state agencies (ex. Texas Education Agency), and other studies (Food and Agricultural Education Information System). Table 1 lists a few of the many examples of available datasets.

Table 1
Common Educational Datasets

Dataset	Link
National Teacher and Principal Survey (NTPS)	https://nces.ed.gov/datalab/
Common Core of Data (CCD)	https://nces.ed.gov/ccd/
High School Longitudinal Study	https://nces.ed.gov/surveys/hsls09/
Education Longitudinal Study (ELS:2002)	https://nces.ed.gov/surveys/els2002/
Integrated Postsecondary Education Data System (IPEDS}	https://nces.ed.gov/ipeds/use-the-data
California DataQuest	https://dq.cde.ca.gov/
Texas Education Agency (TEA) Reports	https://tea.texas.gov/reports-and-data

Guiding learners to understand the tools used with open data is crucial because it enables them to move from simply viewing data to critically analyzing it, engaging in real-world problem-solving, and developing essential data literacy skills. This guidance helps students become active participants rather than passive users, fostering deeper understanding and more authentic learning experiences (Coughlan, 2020).

Data driven decision making (DDDM) is important for educators. Policymakers demand educators use data to inform practice. Data literacy is at the center of DDDM (Mandinach and Gummer, 2016). When we push for more DDDM we need to remember that people are what makes the data useful and accessible (Fergison, 1017). Koltay (2017) states that data literacy is vital for modern researchers.

Often these public datasets can support a researcher's own studies. Graduate students have a need to learn how to process and visualize these data; commonly considered data literacy skills. Most agricultural educations programs will include a research methods course. However, these courses do not commonly include the analysis and visualization of larger datasets. Kjelvik and Schultheis (2019) suggest that the strongest learning experiences surrounding data literacy may arise when students are given opportunities to work with authentic data. Varlamis (2025) found that using real world data to practice using digital literacy (DL) tools is an effective strategy for student learning. With the proliferation of online datasets and advanced features in Excel students can easily process large data sets on their desktops. These activities were developed around a publicly accessible dataset related to agricultural education and focus on the use of tools and techniques used in dataset analysis.

How it works

The National Supply and Demand dataset (Foster et al. 2025) provides a dataset for data exploration with data that is relevant to agricultural education. DL skills include downloading datasets, working with data in different formats (e.g. CSV), manipulating data using pivot tables, and visualizing data. Six activities were developed that guide students in accessing the data on the web site: https://nsd.aaaeline.org. The activities are short and designed to be blended into existing courses. There is a Lesson Guide which provides an overview of the activities, a Pivot Table Guide which illustrates how to create a pivot table in Excel, and a Quick Guide to the NSD Website. Each activity is introduced with a brief background and research questions to be answered. A data and analysis section describes how to access the data set and procedure to perform the analysis in Excel. Questions are posed about the process. Deliverables are typically charts and tables. A discussion section concludes the activity with questions designed to stimulate thinking about the data and the process and ties the activity back to the research questions.

Results to Date

Ph.D. students completed the activities independently before a face-to-face session and successfully navigated the NSD website, downloaded data, created pivot tables, and produced charts. Students reported increased confidence working with large datasets and clearer connections between data and agricultural education research. In-person discussions focused on data interpretation, limitations, and effective visualization. Students also practiced presenting their findings and designing posters, strengthening their communication and data-storytelling skills. Despite varied Excel experience levels, nearly all completed the tasks, and peer support emerged naturally. Activities work well for individual or group-based learning.

Future Plans

Planned refinements include additional scaffolding for students new to Excel, expanded opportunities for deeper data interpretation, and new advanced activities such as integrating related datasets (e.g., NCES), addressing data quality issues, joining datasets, learning introductory Structured Query Language **(**SQL**)**, and exploring AI tools for data analysis. Enhanced instruction in poster design and data communication will further support students' visualization and presentation skills. These updates aim to improve data literacy and strengthen students' analytical, communication, and visualization competencies.

Resources Needed/Cost

No costs are incurred to access the data. Activities are available at no charge. Dissemination of the activities to students is done electronically, but they could be printed for face-to-face class meetings. Most students have access to Excel however Google Sheets can be used as well. Google Sheets is not as robust a tool as Excel, but all activities can be adapted for Google Sheets. A separate guide is provided for Google Sheets. Materials were developed by the NSD team.

References

Coughlan, T. (2020). The use of open data as a material for learning. *Educational Technology Research and Development, 68*, 383–411. https://doi.org/10.1007/s11423-019-09706-y

Ferguson, M. (2017). Good time to be a data nerd. *Phi Delta Kappan*, *98*(7), 74–75.

Foster, D. D., Smith, A. R., Rogers, A., Lawver, R. G., & Spiess, M. (2025). *National Agricultural Education Supply and Demand Project* [data set]. American Association for Agricultural Education. http://nsd.aaaeonline.org

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting Messy with Authentic Data: Exploring the Potential of Using Data from Scientific Research to Support Student Data Literacy. *CBE Life Sciences Education*, *18*(2), es2. https://doi.org/10.1187/cbe.18-02-0023

Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science, 49*(1), 3–14. https://doi.org/10.1177/0961000615616450

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate? Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education, 60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011

Varlamis, I. (2025). Messy data in education: Enhancing data science literacy through real-world datasets in a master's program. *Education Sciences, 15*(4), 500. https://doi.org/10.3390/educsci15040500