Water Science in Agriscience: Promoting Data Literacy in Secondary Education through Stream Monitoring

Ian Curtis
Mentors: Amanda Buday, Bradford Dykes

Data Science & Analytics Master's Program Internship Report
Grand Valley State University
Social Science Lab
Winter 2025

Table of Contents

Internship Objectives

This internship will focus on building a data presentation tool for K-12 educators by building on essential skills and software introduced in DSA curriculum. The final product will assist teachers and students in high school environmental science courses in collecting, analyzing, and reporting on the water quality in their local streams and creeks. Classes will use scientific instruments to record data about the water and water quality and input these measurements into a tool developed during the internship. This tool will include a dashboard that produces human-readable and eye-appealing descriptive tables and plots for both numeric and non-numeric data (CIS 671, STA 623) that students can interpret in real-time (i.e., as data is added), with little latency. Given the time and other constraints many K-12 disciplinary teachers have, these instructors have shared with Dr. Amanda Buday (project lead, Sociology Department) that automating the creation of data summaries would allow them to focus more on interpretation and reporting and the importance of unbiased storytelling.

This project will be built in the Google Workplace for Education ecosystem (as most K-12 schools use Google products). Building an automated tool with Google will require a use of advanced functionalities and programming in R and inside of Google itself (STA 518). As the dashboard will rely heavily on Google Sheets, I will expand my data structures and systems and querying skills (CIS 660). I will also learn how to effectively work with Google's API to build an interface for K-12 teachers to create a Google Drive folder containing all items they need to implement the larger streams and creeks project for their unique classroom needs (number of student teams, number of collection sites, etc.). My project will also require that I visit these high school classes to observe how students use the tool, identify and correct any errors and bugs, and receive feedback from students and teachers on how to improve the user interface and presentation of results. Finally, I will regularly meet with project team members to assist in the creation of instructional materials and resource guides for K-12 teachers to integrate data literacy concepts into their lessons.

Introduction

Across industries and companies, mass amounts of data are being collected and stored each day. As a result, we are constantly being fed data analyses through the news, advertisements, internet searches, and more. How can we be so sure that the data was collected honestly or fairly and that the graphs, tables, and claims made from that data are unbiased and truthful?

Data literacy, or the concept of understanding how to collect, manipulate, chart, interpret, and synthesize data, is becoming more and more necessary but less and less prevalent. David Herzog, author of *Data Literacy: A User's Guide*, notes that although data surrounds us and is presented in many forms to us every day, "understanding and analyzing data is a dark art that we'd rather leave to the 'experts.'" (Herzog) Furthermore, this "dark art" is not one that is commonly taught in schools, forcing us to blindly accept whatever conclusions are reported. This can cause issues when we attempt to make conclusions from data (or make our own conclusions from conclusions that someone else made about data). How do we know the data source is trustworthy? How can we confirm that the conclusions we and others make are directly supported by the data? As the digital age continues to progress, we will not be able to escape data and it is imperative that we understand how to critically analyze reported data and its sources and to turn visualizations into a meaningful message.

The project was built and tested with Future Farmers of America (FFA) high school classes in rural Michigan schools. FFA students are exposed to several key environmental topics that play a role in effective farming, one of which pertains to safe reliance on natural resources. This project focuses on watersheds, a crucial aspect of farming and public health and includes students in each step of the data collection, analysis, and reporting process. Over the past few years, Amanda Buday (Sociology) and Bradford Dykes (Statistics) have worked to develop a hand-on classroom activity that can be run as a standalone unit or be inserted into a pre-existing FFA course. This activity requires students to explore a local watershed to collect water samples, basic stream chemistry readings, macroinvertebrate samples, and to read and interpret E. Coli results received from a GVSU lab. This project aims to aid students in interpreting the data that come from their samples and in converting those results into meaningful messages for them and for sharing with the broader community by removing the need for students to spend time cleaning data, creating graphs (and learning how to create accurate graphs), or working with advanced spreadsheet functions.

Description of Work

When we first sat down to discuss the internship and the project, we discussed several goals. What did we want students to understand? How do we want to convey that information? What role should high school instructors take in this process? In the end, we determined that we needed to cater to both students and instructors. Regardless of what the project looks like, we wanted to ensure that students could easily and quickly pull the information they need without needing to understand complex statistics or spending unnecessary time manually creating graphs and/or tables. At this point in their education, it is more important for students to be able to understand how to read graphs and understand their limitations, explain the results from a graph to a general audience, and connect the data they collected to the context of their class and the community.

As a result of this thought process, it was determined that the output of this project would take the form of a data dashboard containing curated graphs of data that had been collected by students in their class periods. The general structure of the project involves students entering this in a spreadsheet (eventually, this would be a Google Form), feeding that data into an automated program that created a dashboard of charts, and presenting those charts back to the students online. The following were determined to be key features of the dashboard:

- **Automated.** The focus for students is to interpret charts; therefore, there is no need for students to be involved in either data manipulation and cleaning or the technicalities behind *how* to make accurate charts and the actual process of making them.
- **Easy**. Students, especially high school students, likely will not understand or appreciate the importance of good statistics. When observing classrooms, it was discovered that students find the data analysis portion of the project the least exciting (they would rather be doing hands-on activities). The easier the data entry and the less interaction with plain numbers and spreadsheets, the better. Moreover, high school instructors are not looking for complicated lesson plans or instructions, emphasizing the need for worksheets, guiding questions to ask students, and suggested topics of discussion. The dashboard should be curated to remove the burden on all involved in the project while still providing accurate and interpretable information.
- **Fast**. FFA classrooms have multiple projects and activities they participate in throughout the year, of which only one pertains to water quality. We need to be conscious of this and avoid wasting precious class periods with unnecessary explanations or lessons. Students should be able to enter their data and visualize the results in one or two class periods.
- **Concise**. As mentioned above, data analysis may not be the most exciting part of the process for students. The time spent handling data should be kept to a minimum and the time spent thinking about data and determining what the data means should be emphasized.
- **Communicable.** The output (i.e., charts) need to be presented in a manner that is easy to interpret. A piece of the project requires students to explain the patterns or trends (or lack thereof) in the plots and how those patterns relate to

what they expected to see and to what would be considered an extreme or concerning data value.

The files for instructors and students are located on Google Drive. See the Appendix for an example of a live dashboard. An instance of a stream monitoring project is created on a by-waterbody basis (so, in theory, if one school recorded data from two different waterbodies, they would receive two instances of a stream monitoring project). To aid in reducing the workload of instructors and to avoid the chance of error, each project is generated by a member of the GVSU community. This can be done by activating a Shiny app (written in the statistical programming language R) and walking through the steps in the app. This creates copies of a template project and saves the new copy inside of the user's Google Drive.

Inside the R Program

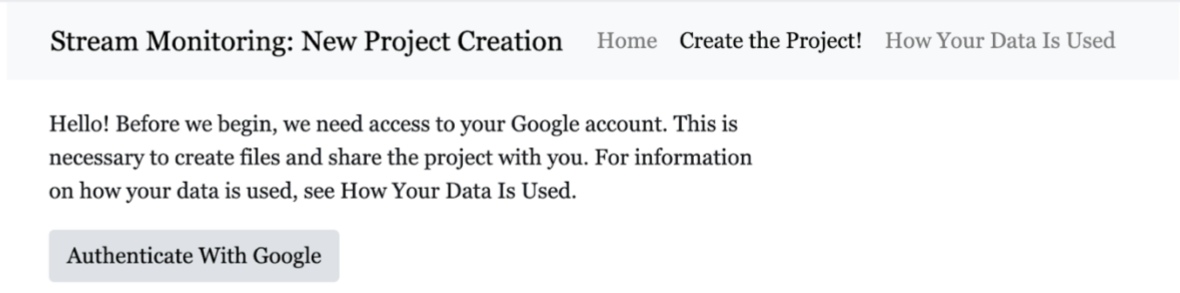

Figure 1: The Shiny App

The R program, corresponding Shiny app, and the template Google project files were all created as a part of this internship (Curtis, GitHub). The R program relies on the {googlesheets4} (Bryan 2023) and {googledrive} (D'Agostino McGown & Bryan 2023) packages (which utilize various APIs to connect to Google's servers) to manipulate Google files. The Shiny app works as follows:

1. The user opens the app and authenticates with their Google account. This gives the app permission to create a new project and save it to the user's Drive.
2. The user enters the name of the waterbody data will be collected from. The app then creates a new Google Drive folder using the name entered in the user's My Drive folder.
3. Once this is done, the app copies several files from a template project located on Google Drive. The user will have to pause several times to verify that the Google Sheets are properly linked together to enable the proper automation.

a. Note: It is up to the user to move the newly created folder to wherever they would like it in their Drive.

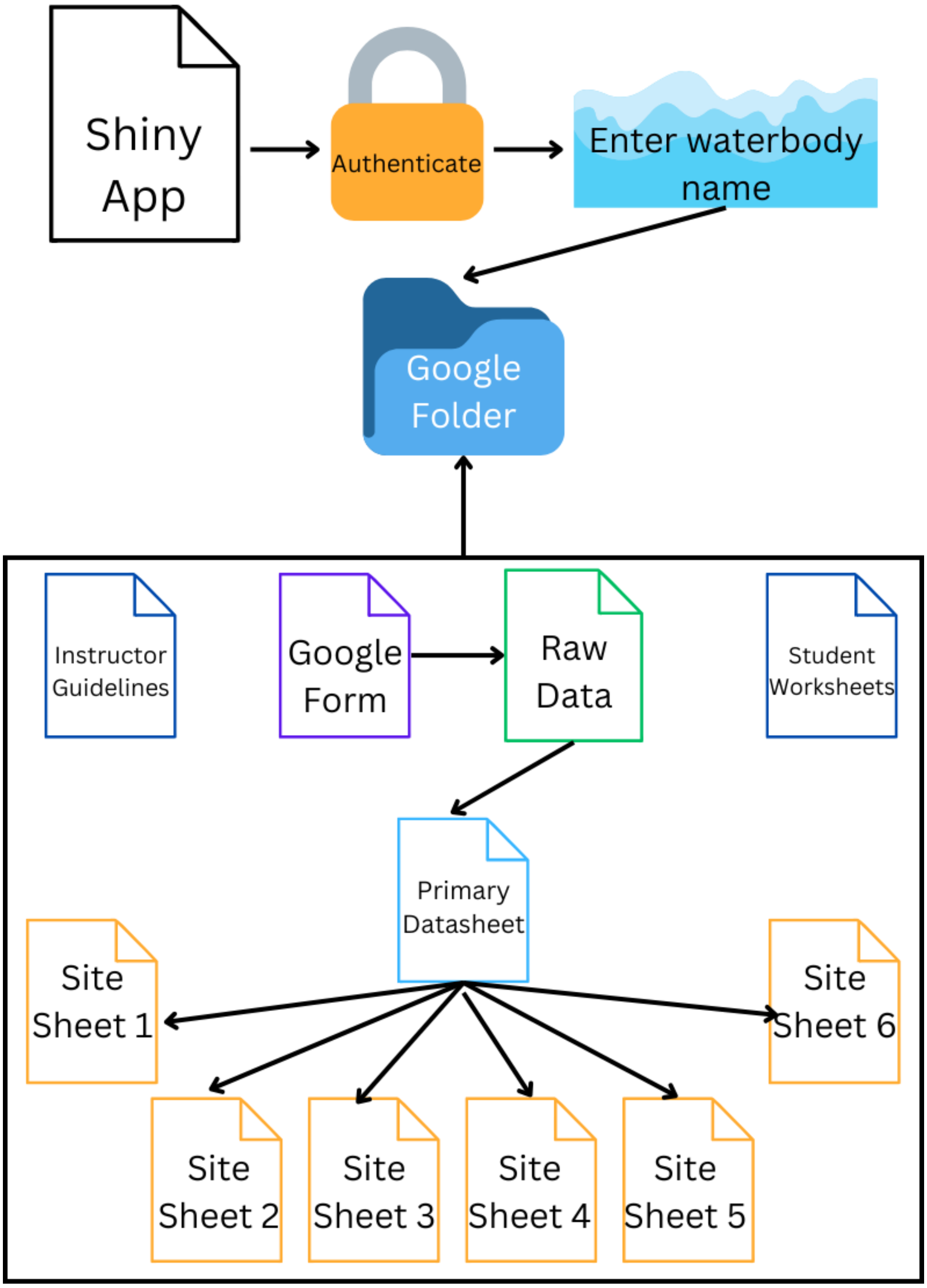

Figure 2: The Project Creation Process: from Shiny app to full Google folder

Inside the Google Project

Generated in each project is:

- A Google Form (for data entry)
- 8 Google Sheets
  - One for collecting the raw data from the Form
  - One that sorts and curates the data by category (E. Coli, Stream Chemistry, Macroinvertebrates, Water Quality)
  - 6 that correspond to six different sites / recording locations on the waterbody

- Student Worksheet (Making Predictions)*
- Student Worksheet (Data Interpretation and Synthesis)*
- Instructor Guidelines*

*In progress

Each project is set up to support up to six different sites to support instructor needs or project growth. There is no requirement to use all six sites at any given time. The Google project is set up such that students only need to enter their data using the Google Form (which is already set up to ask the correct questions). Once this is done, Google will automatically sort and send the data to the respective site sheet(s) and populate graphs for class use.

The Google Form

In the first test classrooms, students entered data themselves into a Google Sheet and the automation went from there. However, we noticed that several students had little experience using any sort of spreadsheet and were accidentally deleting formulas and key pieces needed to generate plots. This issue persisted even despite later additions of cell colors to indicate where data should be entered (green/red cell fill).

As a result, we decided to switch to a Google Form for accepting data and to use the resulting spreadsheet from the form for chart generation.

Raw Data and Primary Datasheet

There are two intermediate spreadsheets included in each project. These files give instructors the option to correct data entry errors and view sorted data tables without the risk of breaking the connections between the spreadsheets. The Raw Data file contains all data as imported from the Google Form and is where any errors can be corrected, simply by editing the data in the cells. The Primary Datasheet pulls from the Raw Data file but organizes the data by Site # and by data category (see below) allowing for quick identification of data issues or simple comparison between sites. The primary datasheet supplies the data that is then sent to the site sheets. See Figure 2 for a visualization of these connections.

Site Sheets (6)

As mentioned above, the focus of this project is to present curated charts to students to allow them to focus on interpreting data they collected and on understanding the meaning of the data in the context of their own lives. The project is set up such that data that is added over time will also be plotted in the final charts. The charts themselves are located on a dashboard in each site-specific sheet, spread across multiple sub-sheets in the file:

Category	Sub-Sheet Name	Plot Type
Water Metrics	Temperature	Standard line chart with data points
Water Metrics	Conductivity	Bar chart with indicator lines of standard minimum and maximum values

Water Metrics	Salinity	Bar chart with indicator lines of standard minimum and maximum values
Water Metrics	pH	Bar chart with indicator lines of standard minimum and maximum values
Water Metrics	Turbidity	Standard line chart with data points
Water Metrics	Flow Rate	Standard line chart with data points
E. Coli	E. Coli	Bar chart with indicator lines of standard values for unsafe partial body contact and unsafe full body contact
Macroinvertebrates	Macroinvertebrates	Stacked bar chart, colored by semester
Macroinvertebrates	Water Quality	Line chart with data points, contains reference diagram

It is important that the charts given to students be easy to read and interpret but also be able to handle a growing amount of data to be added as semesters go on. For most metrics, a line chart was chosen, plotting data points in chronological order and connecting them together to show a trend (Figure 3). This is a rather simple plot; however, it conveys the message we want to get across to students: how certain values change over time, both within a semester and between semesters.

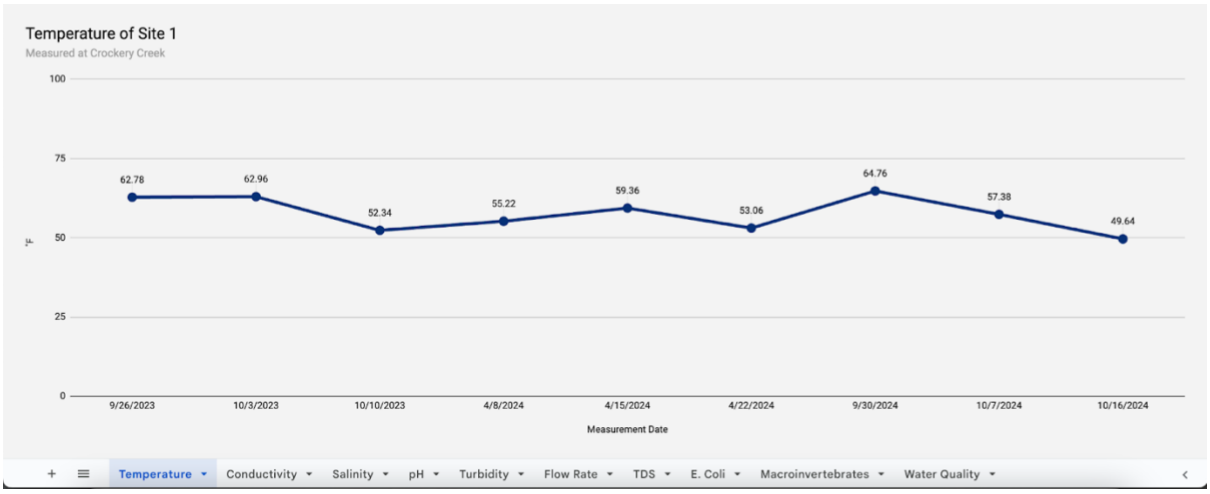

Figure 3: An example line chart. Depicts 3 semesters of Temperature readings from Crockery Creek.

Some of the metrics come with water quality "standard" recommendations for what is considered a high or a low value for freshwater streams. The plots for these metrics contain a solid horizontal line representing these high/low values whereas the observed data points are represented as bars (Figure 4). The design of this chart was built around the limitations of Google Sheets. Although Sheets has a variety of different built-in chart options, only a select few work with more than two continuous variables. We chose Google's Combo Chart here – despite the fact that a bar chart is not usually recommended for continuous data (Sackmann) – due to its ability to still show a trend over time while also providing visual labels of the "standard" minimums and maximums. The height of the bars represents observed data points and students are able to detect the change in the bar height over time to identify patterns. The addition of the static lines allows for further extension of interpretation value as students are now able to place their results in the context of what is “normal” or “expected” for waterbodies.

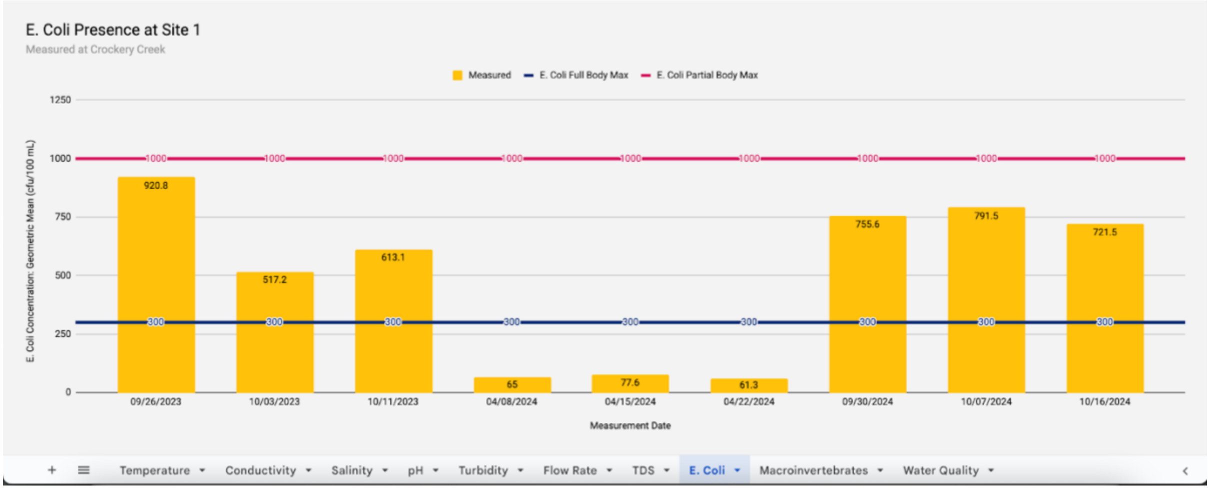

Figure 4: An example bar chart with reference lines. Depicts 3 semesters of E. Coli readings from Crockery Creek.

The only chart that does not follow the above two patterns depicts macroinvertebrate counts over time. Macroinvertebrate counts (and, by extension, Water Quality) are reported on a by-collection-date basis meaning that it is not logical to aggregate these findings across multiple years or seasons. For instance, we would not want to create a plot with macroinvertebrate totals across all instances data was recorded in fall or for the entire 2024 year. Since there are many factors that go into which macroinvertebrates are present at any given time, these findings have to be reported independently from other collection times. As a result, we group all macroinvertebrate and water quality results by semester collected (e.g., Spring 2024, Fall 2025, etc.). To best display this data, we chose to use a horizontally stacked bar chart (Figure 5). The essence of this chart is to allow students to see their individual results and compare the counts across semesters. Labels have been added directly on the plot to provide counts of each bug type by semester and bars are colored by semester to visually separate the collection times, per Gestalt principles (Wagemans, 2025).

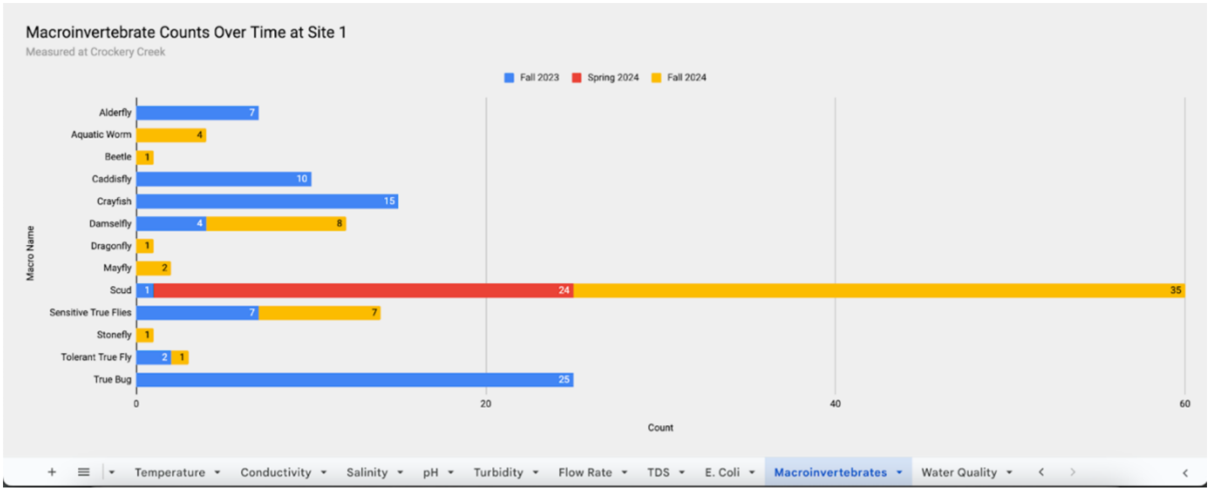

Figure 5: An example stacked bar chart. Depicts 3 semesters of macroinvertebrate quantities from Crockery Creek.

We would also like to note that the water quality plot (derived from macroinvertebrate abundance and diversity) is paired with a reference diagram placing the water quality scores in a spectrum of values indicating how poor the water quality is (Figure 6). This score is calculated off screen (although students are given the formula in their workbooks) and organized into a plot for students. The calculation can be found in the Appendix.

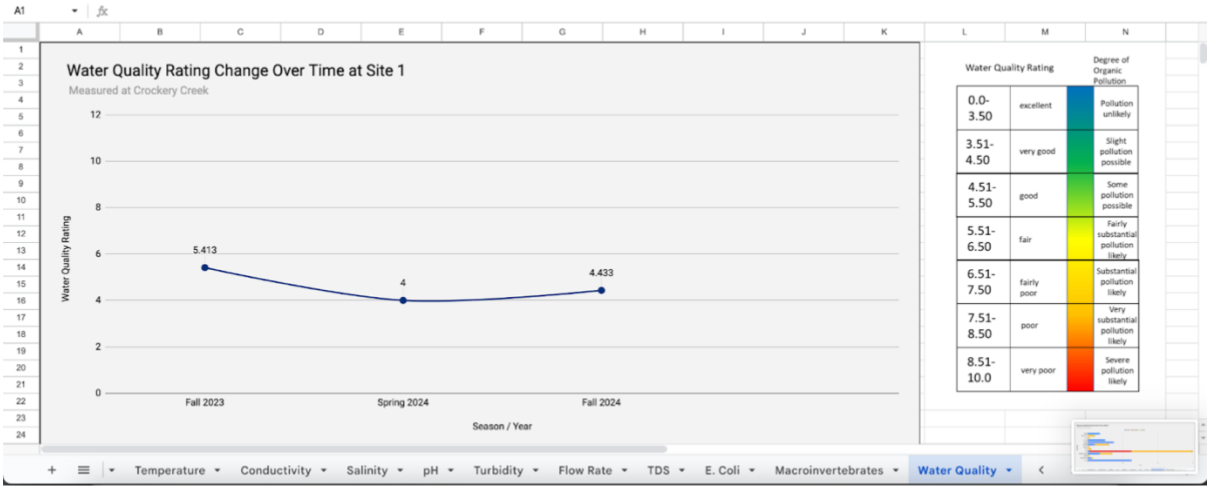

Figure 6: An example Water Quality Rating plot (and corresponding reference diagram). Depicts three semesters of results from Crockery Creek.

Student Worksheet 1

A data preparation worksheet asking students to make predictions about their data was drafted during this internship. This is intended to be filled out after data collection but before data entry and analysis of charts. The goal with this worksheet is to get students thinking about the data they collected before they actually see the results. We are in the process of collaborating with the appropriate teams to ensure that the worksheet meets the State of Michigan education criteria for science courses.

Student Worksheet 2

A data interpretation worksheet containing guiding questions for students was drafted during this internship. This is intended to be filled out after data collection and entry and walks students through interpreting the charts and connecting the results to the real world. We are in the process of collaborating with the appropriate teams to ensure that the worksheet meets the State of Michigan education criteria for science courses.

Instructor Guidelines

An instructor guidelines file was drafted during this internship. At present, this file will not be included in projects generated for schools since GVSU is actively involved. As the project continues to expand and schools are using the project without direct GVSU involvement, this file will be revised and added into new projects to give instructors guidance on how to use the dashboard, worksheets, and course plan in their classes.

See the Appendix for links to the drafts of the Student Worksheets and Instructor Guidelines.

Testing and Implementation

Three high schools were visited over the course of this project to test the functionality and ease of use of the dashboard and interpretability of the resulting charts.

School Name	Waterbody Name	Seasons Recorded
Ravenna High School	Crockery Creek	Fall 2023 Spring 2024 Fall 2024 *Spring 2025**
Coopersville High School	Deer Creek	Spring 2024 Fall 2024 *Spring 2025**
Montague High School	Little Flower Creek	Fall 2024 *Spring 2025**

**In progress at the time of writing this report*

Over the course of 3 weeks, following a course plan created by Amanda Buday, students travel to a local creek or river to collect their data, which they record in a paper logbook. Once all data is collected and E. Coli results are received from the lab, students use the Google Form to submit their data into the dashboard ecosystem. Since all of the Google files are interconnected, charts are generated as data is entered with very little latency. As a result, instructors may begin guiding students in the interpretation of the charts as soon as data entry has ended.

Feedback and direct observation during the school visits helped shape the dashboard and Google files. Both the R program (and corresponding Shiny app) and the set of Google files evolved through several iterations of the project with each classroom visit providing insight into how to cater the files to the instructor and student preferences and needs. The following pieces of the dashboard were edited or changed as a result of testing:

- Data entry is now a Google Form rather than a Google Sheet
- The project supports 6 sites for flexibility rather than exactly how many are needed at present
- Sites are numbered instead of given specific names
- The Water Metrics charts are on their own sheets rather than having all six on one sheet
    - This works against our Concise goal but ultimately supports the longevity of the project making it easier to visualize the charts are more data is added

- Total Dissolved Solids were added as a reported metrics (students were already collecting these data but a chart was not in the original dashboard)
- Solid lines were added to charts that have standard recommended values
- Macroinvertebrates are plotted as a stacked bar graph, grouped by semester
    - Previously, multiple charts existed, separated by year and by season; however, it is not standard to report macroinvertebrates in this way as they are to be interpreted on a by-measurement basis.
- Several backend edits (in the R code and the Google sheets) were added and adjusted over time to reduce and eliminate the need to make changes in the future (i.e., extend the lifetime of the project).

Future Work and Limitations

The project and files as they currently stand are functional and can be used properly in classes. However, the project depends on a GVSU member's involvement and there are a few pieces that are still in development. The main limitation is that the R Shiny App can only be used by those who know how to use R, have R installed, and have the code for the app. The app is not currently able to be posted online as Google authentication is required and the app does not support multiple people signed on at the same time. Future work may involve looking for and/or working on a service to make this happen. For now, someone on the GVSU Social Science Lab team will have to generate a new project for each school that wishes to participate in the project. Future work will also involve continuing to develop the instructor guidelines document mentioned above as well as documents containing guiding questions for students to fill out before data entry (to aid with making predictions) and after data entry (to aid with interpretation and analysis).

In order to use this stream monitoring project as a part of approved curriculum, it is necessary to include some form of formal assessment. Another goal is to connect with the GVSU education department as well as the GVSU Annis Water Resources Institute (AWRI) to align the project with Michigan education standards and to build assessments to help demonstrate student learning.

Internship Discussion

All in all, my internship experience was a great opportunity to put some of the skills learned about in courses into practice. Since the DSA degree is half computer science and half statistics, I felt that it was great to use the internship to see how these two sides can work together. I do think that there should be some specific courses in the DSA program that discuss how to integrate statistics with computer science topics as the program currently implies that there are two separate, distinct areas of focus. In practice, data scientists blend statistics and computer science topics and skills, and I think the coursework should reflect this. That being said, this internship experience did offer a chance at exploring the open world of data science and I pulled from my coursework to do so.

Data science is an open concept and there are still multiple interpretations of what it means to be a data scientist. To me, data science includes more than just sharing a plot with someone. It involves looking at data to try and identify a trend or helping others to identify trends. It may also involve educating others on how to properly gain insight from a given table or chart. The data scientist is responsible for cleaning raw data and preparing it for use. They then may use that data to create models, run machine learning algorithms, and build charts. Then, critically, the data scientist must be able to present their findings and what they mean in the context of the problem being studied to an audience who does not understand detailed statistical information. Through my coursework at GVSU, I did gain an understanding of the skills necessary for each of these parts and, through the internship experience, put them all together into a single project.

The major challenge that arose in this internship related to an extension of the technical skills I had learned in courses. In my courses, I had done my work in major programming languages (R, SAS, and Python) and learned how to work with data to produce models, charts, and other output there, including guidelines to make them readable and accessible. However, this internship required the use of the free tools that were accessible to those who are not data scientists. Hence, we decided on the Google ecosystem for the project which gave the added benefit of fitting within many schools' use of Google Classroom. Other than basic Google Sheets functions, I had not done anything with automation or coding in the Google ecosystem. The internship gave mea a new experience of extending my knowledge of R into using the Google API and my knowledge of SQL and data querying into Google Sheets automation. The internship project went through 3 different iterations while we were modifying what the dashboard should look like to students and in each new iteration, I made slight adjustments and improvements to the backend code for faster automation, simpler code, and lower chances of error.

Another challenge was creating data literacy material that was interesting to students and kept them engaged throughout the class period. The first iteration of the project had students enter data directly into a spreadsheet. This took up too much time, was confusing to students, and bored them. The second iteration of the project asked too many questions in the Google Form and made too many assumptions about what students knew about data entry. Both of these instances led to the data entry portion lasting longer than one class period which was an unnecessary use of class time. To correct this, I took out all questions that were redundant and added in instructions to help students fill the form out faster. For instance, I did not need to collect specific date data for macroinvertebrates since this data is reported in aggregate by semester. Students also struggled with entering dates in the correct format, so I added in a note in the form instructing students how to get the correct date format. I wouldn't say that this challenge was fully accounted for but we took the necessary steps to mitigate the annoyance or boredom of the students during this portion of the project.

Outside of technical skills, I learned how to better collaborate with others not in my discipline. In this internship, I was acting mostly as a consultant, but I did not make final decisions in terms of dashboard appearance or student objectives. I made recommendations throughout the course of the project but had to collaborate with people from other disciplines and with high school students and instructors in order to bring this project to life. It was necessary to take feedback from all of these groups during the development of the dashboard. Working with a diverse group of people helped me refine my recommendations and decisions, especially as I was thinking about the future of the project after the internship and who was going to be using it and maintaining it. Had I not participated in high school classroom use of the dashboard, I would not have fully understood how students might make errors in data entry/interpretation nor would I have fully understood that students aren't as passionate about graphs as I am (they would rather be doing hands on work). Having seen this, I was able to make adjustments to the data entry form and enter discussions with my collaborators about fine-tuning the data entry/analysis phase of the project to keep students engaged and to help them understand the big picture ideas: what the data indicates and why it is important to what they're studying.

Relatedly, I also think that the information from the PSM curriculum on sending emails and other workplace communication was valuable for this internship and the majority of the work was done asynchronously with online meetings (other than on-site meetings at high schools). It was essential to be able to communicate my progress, ideas, and suggestions over email in a coherent manner to avoid unnecessary confusion. Over the course of the internship, I made several recommendations in regard to the layout of the dashboards, the data interpretation worksheets, and the key pieces of information students should take away from the project. These all had to be made with a general audience in mind as not all collaborators on the project were as deeply involved in statistics and data science as I was. Furthermore, I also had to explain the details of the R Shiny App and some of the backend functionalities of the Google ecosystem to analyze its limitations and to justify my decisions and the time spent on those pieces of the project.

In the end, we accomplished many of our original objectives. A solid, fully functional final draft of a Google dashboard was developed and tested on several occurrences. In addition, documentation was left to instruct future users of the project how to use the R program to generate new projects and how to advise instructors on how to best utilize the dashboard in their classrooms. We didn't quite complete the supplemental documents (Instructor Guidelines, Student Worksheets 1 and 2). Nevertheless, a developed draft of each was created and feedback was received from education experts to help guide the next iteration of the files. The Student Worksheets were tested in two classrooms during the course of the internship which gave valuable insight into how to adjust the questions within to aid students in understanding the data.

Appendix

Calculating Water Quality Rating

There are nineteen types of macroinvertebrate groups, or families, that students can report. Each of these groups are assigned a weight, also known as a "sensitivity rating". In general, bugs can be "tolerant" or "sensitive" where the more tolerant a macroinvertebrate is, the more able it is to survive in polluted environments. For example, finding 10 Aquatic Worms (sensitivity rating of 10 out of 10) and not many other bugs would be problematic as this would likely indicate that the water is polluted enough to kill off more sensitive bugs leaving Aquatic Worms behind to thrive. Conversely, finding 10 Sensitive True Flies (sensitivity rating of 1 out of 10) might not be too much of a concern depending on the other bugs found.

Water Quality Rating is calculated first by multiplying the quantity of each bug found by the respective sensitivity rating and then adding up each of those quantities. The final value reported is the sum of the products divided by the total quantity of all bugs found. This creates a scale ranging from 0 to 10 where 0 indicates the water is very clean and 10 indicates the water is very polluted.

See the Michigan Clean Water Corps webpage for more information on these calculations. See also the guides put together by Chadde and the Government of South Australia for more information. Although these resources use slightly different calculations and scales for Water Quality Rating and are many years dated, they share the same thought process.

Links to Relevant Project Files

See below for view-only links to an example project and draft files:

- Crockery Creek (Ravenna) Stream Monitoring Google folder
- DRAFT Instructor Guidelines
- DRAFT Student Data Worksheets

References

Bryan, Jennifer. *googlesheets4: Access Google Sheets Using the Sheets API V4*. tidyverse, 2023. *GitHub*, https://github.com/tidyverse/googlesheets4.

Chadde, Joan Schumaker. *Macroinvertebrates as Bioindicators of Stream Health*. 2006, https://orange.wateratlas.usf.edu/upload/documents/Macroinvertebrate_id.pdf

Curtis, Ian. *Crockery Creek Stream Monitoring Dashboard*. 2025, https://drive.google.com/drive/folders/12vdMhaJMymzKrK24rXM4EdzXIHO5eVZm?usp=sharing.

---. *(Draft) Instructor Guidelines*. 2025, https://docs.google.com/document/d/1RW-VPwzQrBcCtpJE7F7KZB-4nhWQcmQFCUAtvtYdOt0/edit?usp=drive_link.

---. *(Draft) Student Data Worksheets All Together*. 2025, https://docs.google.com/document/d/1wdybpljTm7Xsaz3joDuaKxKBcJlUEp_KSHpk3HrStMc/edit?usp=drive_link.

---. *ian-curtis/stream-monitoring*. 2023. 21 Mar. 2025. *GitHub*, https://github.com/ian-curtis/stream-monitoring.

D'Agostino McGowan, Lucy, and Jennifer Bryan. *Googledrive: An Interface to Google Drive*. tidyverse, 2023. *GitHub*, https://github.com/tidyverse/googledrive.

Government of South Australia. *Aquatic Macroinvertebrate Identification Guide*. 2009, https://data.environment.sa.gov.au/Content/Publications/macroinvertebrate_guide.pdf.

Herzog, David. *Data Literacy: A User's Guide*. SAGE Publications, Inc., 2016. *methods.sagepub.com*, https://doi.org/10.4135/9781483399966.

Michigan Clean Water Corps. *MiCorps Macroinvertebrate Datasheet*. Sept. 2024, https://www.micorps.net/wp-content/uploads/2024/10/2024-VSMP-Macro-OrderLevel-Datasheet.pdf.

Sackmann, Anna. *Library Guides: Data Visualization: About*. https://guides.lib.berkeley.edu/data-visualization/about. Accessed 20 Mar. 2025.

Wagemans, Johan, et al. "A Century of Gestalt Psychology in Visual Perception I. Perceptual Grouping and Figure-Ground Organization." *Psychological Bulletin*, vol. 138, no. 6, Nov. 2012, pp. 1172–217. *PubMed Central*, https://doi.org/10.1037/a0029333.

Wickham, Hadley, et al. *R for Data Science (2e)*. 2nd ed., https://r4ds.hadley.nz/. Accessed 24 Mar. 2025.