

EDS Capstone Presentation

QAQC Project

Alex Turse, Claire Porter, Jackson Boran

2/5/2025

Introduction

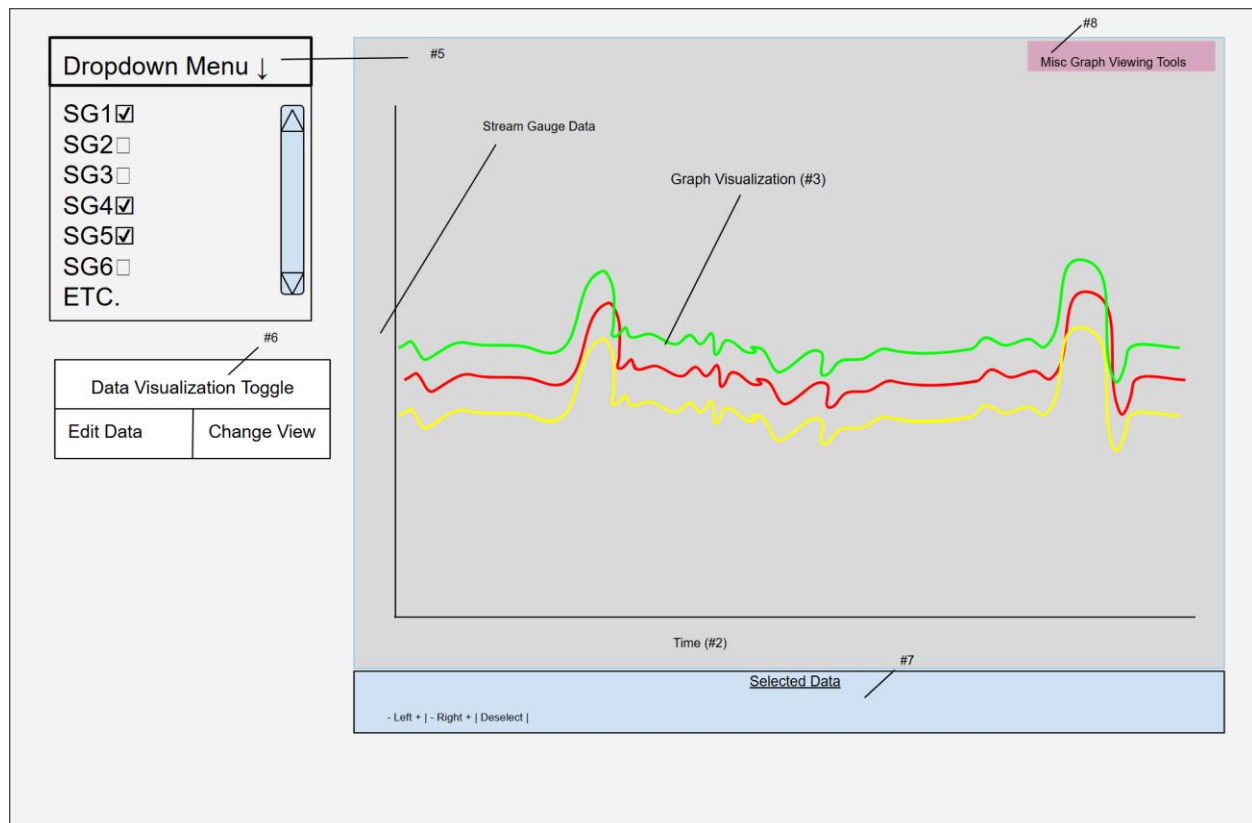
The Hubbard Brook Experimental Forest, located in West Thornton, New Hampshire, is an ecological research site established to study forest and watershed dynamics. Researchers have utilized the forest to understand hydrological, biogeochemical, and ecological processes within paired watersheds. The primary objective of the study is to assess the inflow and outflow of water within these watersheds, with a specific focus on streamflow measurements. The data collection process relies on weirs and stream gauging stations, which provide continuous streamflow measurements, commonly referred to as outputs. These measurements are recorded at five-minute intervals, allowing for a detailed temporal analysis of water movement. However, certain challenges exist within the dataset, including anomalies such as water output spikes caused by natural obstructions like twigs and leaves in the measurement equipment.

In addition to streamflow data, the study incorporates various environmental parameters such as air temperature, precipitation, and overall climate conditions. These factors play a role in understanding seasonal variations in water movement. For instance, summer months typically experience lower water flow, which must be accounted for in data analysis. The collected data must be visualized effectively to identify patterns and anomalies, necessitating an interactive approach where users can select specific sites and manage data integrity by assigning non-applicable (N/A) values when necessary. The system must ensure that the underlying data remains preserved even after such modifications. Challenges in working with this dataset are the

management of multiple axes corresponding to different variables, as well as handling discrepancies in data scales. The ability to zoom into individual data lines while maintaining an overview of the entire dataset is essential for comprehensive analysis.

A requirement of the project is the seamless integration of all necessary dependencies, ensuring that users do not need to manually install required packages. The program should automatically download all dependencies and packages needed so the user experiences efficient data analysis. By incorporating these functionalities, the system aims to provide researchers with an intuitive and reliable tool for examining watershed dynamics within the Hubbard Brook Experimental Forest.

App mockup



Mockup made using Google Draw (<https://docs.google.com/drawings/d/1-wOHdJrbIcp0JhkNDFYBh2p5Xby1ekMn8FA9TEzLkgA/edit?usp=sharing>)

This app as a whole is still a work in progress however, here are a couple ideas from the design and productivity. At the top left corner will be our drop-down menu for our data sets, each one will contain its own box to be checked or unchecked to be shown on the graph. Below it, there will be a data box, with data visualization. Which will give the user the ability to edit the data and change the view. For the editing data button, it will likely be a function where the user can delete a certain point on the graph or section.

Another thing that should be done is making sure that the graph showing the data is dynamic, which entails being able to click and drag a box around the certain data that the user wants to

zoom into. There will be some kind of dynamic box whether it's on the graph that the user can move around or at the bottom of the graph, which is labeled as number seven and eight on our mockup screenshot. Those include buttons that are made for moving left, right or zooming in, along with the keys on a keyboard or mouse.

Existing apps with same functionality

Before this capstone project, the primary tool used for data visualization and editing was the GCE Data Toolbox, a MATLAB-based software. While it provides essential functionality for working with environmental datasets, our client has highlighted some key issues, mainly that it is difficult to use and has an outdated interface. These challenges make it less accessible, especially for those who are not experienced with MATLAB. That said, the GCE Data Toolbox does some things well such as handling large datasets, automating certain tasks, or offering specific statistical tools. Its functionality relies on MATLAB's built-in features or specific libraries.

There are also other tools out there that can perform this role for quality assurance and quality control. One example is Plotly, an open-source graphing library known for its interactive visualization tools. Plotly is a strong tool for creating dynamic graphs, but it does not currently offer an easy way to edit data directly within the plot. In order to work around this, there are a couple of alternatives that may be available. There's a plotly library that allows you to convert GGplot created graphs into interactive plotly maps called ggplotly. The other issue with plotly is the filesize limitations, plotly does not support as large visualizations as ggplot does.

There don't seem to be a lot of applications that directly support QAQC work as a vehicle. A lot of the applications, such as SQLite allow QC work in their functionality, but don't offer direct support for our goals with this project. This leaves a lot of space open for an app in R/Shiny that can quickly and simply perform QAQC.

Expected issues

A challenge that is thought-out in this project is achieving an optimal balance between providing interactive features for data exploration and maintaining the accuracy of the displayed data. The users should be able to zoom, filter, or manipulate data without distorting the underlying dataset. Making sure data stays accurate while enabling flexible, dynamic analysis will require careful design and testing. Additionally, making the system easily transportable and compatible across various platforms is another consideration. It should be a smooth process for users that do not have experience using code or technology. There are a lot of different variables that need to be thought out when considering app development.