EDS Capstone Final Proposal

QAQC Project

Alex Turse, Claire Porter, Jackson Boran

2/14/2025

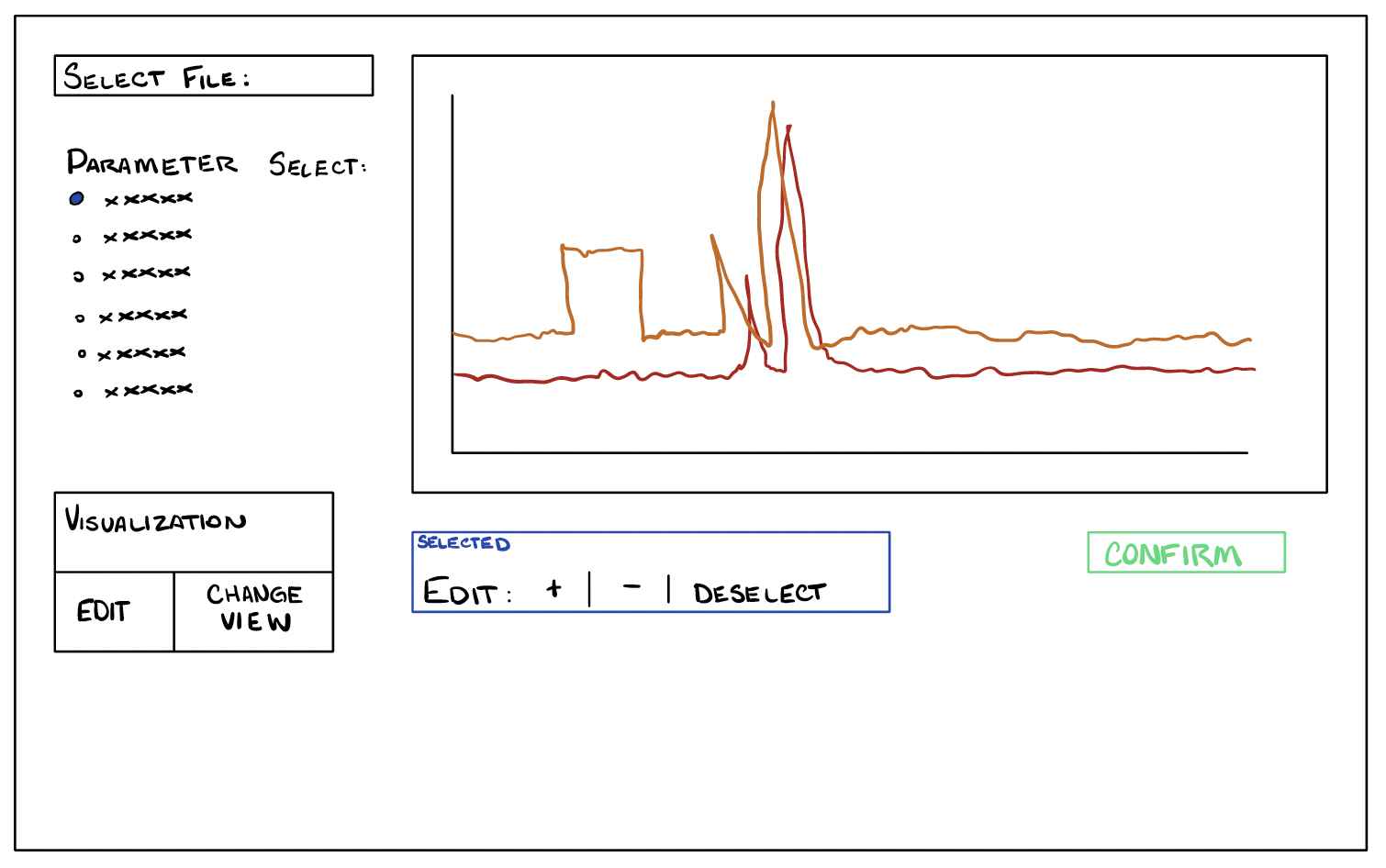
**Introduction**

This project aims to develop an interactive Shiny application to assist researchers in the Hubbard Brook Experimental Forest with processing and cleaning stream gauge data. The application will provide a user-friendly interface for exploring datasets from multiple watersheds, identifying anomalies, and managing erroneous values. By streamlining data cleaning, the tool will reduce manual workload, improve data quality, and enhance the efficiency of hydrological research.

Data collection in the Hubbard Brook Experimental Forest relies on weirs and stream height gauges that record continuous water level measurements at five-minute intervals. These high-frequency recordings allow for detailed temporal analysis of streamflow dynamics, aiding in hydrological and ecological studies. However, the measurement process is susceptible to interference from natural debris such as twigs, leaves, and ice, which can obstruct the gauges and produce inaccurate readings. These anomalies must be manually reviewed and corrected to ensure the reliability of the dataset. Currently, this process is time-consuming and often requires researchers to work with raw data files using spreadsheets or coding environments, making it inefficient and prone to inconsistencies.

To address these challenges, the proposed application will integrate interactive data visualization tools, enabling users to examine trends, compare measurements across sites, and detect inconsistencies with greater ease. The system will allow researchers to select specific watersheds, flag erroneous values, and assign non-applicable (N/A) values where necessary. Additionally, it will preserve the integrity of the original dataset while maintaining a transparent record of all modifications. This ensures that data cleaning processes remain reproducible and well-documented, supporting long-term research efforts at Hubbard Brook and beyond.

**App Mockup**

  
 We will develop a shiny app that includes several key features for data visualization and interaction. At the top left corner, we will implement a file explorer for selecting user datasets. Beneath this the user will be able to select parameters (columns) from the data set. Selected parameters will be viewed in the main panel visualizer. The app will also include tools allowing users to edit data and customize their view. The editing functionality will enable users to remove specific points or sections from the graph as needed.

Additionally, we will ensure that the graph is dynamic, allowing users to interact with the data seamlessly. This includes implementing a feature where users can click and drag a selection box to zoom into specific data points. We will also design a dynamic control box, either directly on the graph or positioned at the bottom. This control panel will provide buttons for panning left and right, zooming in, and potentially integrating keyboard or mouse shortcuts for smooth navigation.

**Existing apps with same functionality**

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. 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 do not seem to be a lot of applications that directly support QAQC work as a vehicle. A lot of the applications, such as SQLlite allow QC to work in their functionality, but do not 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 key challenge in this project is balancing interactive data exploration with accuracy. Users should be able to zoom, filter, and manipulate the data without altering the original dataset, which requires careful coding, design, and testing. Coding will be the most time-consuming portion due to the number of functions needed to make sure all features work. We have already run into some challenges with coding involving being able to switch back and forth between editing and the interactive portions.

Another priority is ensuring the system is portable and compatible across platforms such as windows, Mac, and PC to provide access to more than one user. The app should be user-friendly, even for those who do not have coding experience or technical knowledge. We are confident that the product will be a productive and well used data cleaning tool.