Stroubles Creek Visualization

Deepika Rama Subramanian
dramasubramanian@vt.edu
Virginia Tech

Deepthi Peri
pdeepthi@vt.edu
Virginia Tech

ABSTRACT

The STREAM Lab at Virginia Tech is a stream rehabilitation program at Virginia Tech. It aims to remove Stroubles Creek from the Clean Water Act list of impaired waters. As a part of this project, the lab collects meteorological data, water quality conditions data such as dissolved oxygen (DO) levels, turbidity, conductivity, pH, temperature and the height of the stream. Visualizations are used and required to monitor the health of the stream. Current visualizations available to STREAMLab are outdated and static. This project aims in designing, implementing and evaluating visualizations to suit their needs.

KEYWORDS

Visualization, Hydrography, Line Charts

1 INTRODUCTION

1.1 Background

The STREAM Lab at Virginia Tech is a stream rehabilitation program at Virginia Tech. It aims to remove Stroubles Creek from the Clean Water Act list of impaired waters. Stroubles Creek is a 12-mile long stream that runs through Blacksburg and has been designated an impaired waterway since 2002. A lake, river or stream is considered "impaired" if it fails to meet specific water quality standards, according to its classification and intended use. Stroubles creek does not support a diverse community of aquatic insects and fish. The main pollutants are bacteria and sediments. The STREAM lab aims to create awareness and education among the residents as well as the stakeholders about this current scenario. They hope to improve the aquatic habitat within the creek, reduce sediment erosion and reduce bacteria loadings to the stream.

The goals of the Lab include implementing a developed monitoring plan and collecting data to a database for education and research,increasing the number of students utilizing their website and data for hands on learning across several scientific disciplines, developing outreach materials for K-12, the general public, and practicing professionals and providing a platform for launching proposals to better understand watershed sustainability.

In an attempt to do the same, the Lab has two setups to monitor the creek from two bridges, one upstream and the other downstream. The Stream lab website displays the following:

- Live data, including a live Stream Lab site image i.e, an hourly photo
- Hydro-graphs mapping the stream stage or depth at different time intervals for both the bridges
- Live Meteorological data including Wind conditions, Temperature/Relative Humidity, Barometric pressure and Rainfall Conditions at the site.
- Water Quality conditions at both bridges, including Dissolved Oxygen (DO) levels, Turbidity, Conductivity, pH and Temperature.

To effectively achieve their goals, the information, including the images and visualizations must be easy to understand, insightful, interactive and attractive. The current visualizations available on the website are simple but provide zero interactivity and look just as simple as their functionality, to say the least.

In this project we have attempted to reinvent these visualizations as well as generate some new ones for the STREAM Lab at Virginia Tech.

1.2 Problem statement

Here are the problem statements:

- Visualising what the creek looks like from the STREAM Lab at different heights of the creek (Stage)
- Visualising the Water quality data changing against the heights of the creek
- Improving the existing visualisations by making them interactive and beautiful

2 CLIENT INTERVIEW

To understand the project and elicit requirements, we interviewed Laura Lehmann, the Environmental/Agricultural research manager at the STREAM Lab. Here are some of the questions we posed to the client to obtain background information as well as an outline for the project :

- What are the objectives and goals of the STREAM Lab at VT?
  Client: *“ To sum up our work here, our goals are to remove the creek from the impaired waters category, prevent erosion, deposition of bacteria and sediments and create awareness about resuscitating the creek."*
- What are the parameters that you measure? What are variables involved?
  Client: *“ So at stream lab we have two main monitoring bridges and all those monitoring bridges, they are recording stream stage or stream depth. They are recording multiple parameters, the dissolved oxygen level in the water. The temperature and conductivity levels, and also pH of the water. We also have a full weather station out there which shows, half our values of total rainfall temperature relative humidity barometric pressure.”*
- What are the current visualizations telling you?
  Client: *“you’ll have a Hydro-graph which is going to be, like time. And then stage of depth. So you’re going to look at your base flow and then you get a storm of that, we usually get a small peak here which is all the runoff we get from Foxridge complex"*

Figure 1: Stroubles Creek Watershed

- What kind of data are you in possession of? How are the files organized?
  Client: *" Data files from both bridges with stage levels, water quality data, all hourly. They've the timestamps attached. We also have the tower cam videos which look out over the, the stream lab, and be able to say okay well you have the high draft looks like this or the flood and see how the floodplain. Typically, the hourly images of the creek we have and I could share you on the Google Drive. They are named according to the date and time. "*
- In an event like yesterday's storm, what kind of changes did you notice? What would have made it more informative?
  Client: *" we usually get a small peak here which is all the runoff we get from Foxridge complex because it's so close and there's no storm water controls. we get this initial peak, which just. And then we'll get this larger peak. That was like last night. It's like slowly trailing off today. But then, every hour we should have like a picture. So you'd have a picture here and then another hour you'd have another picture here (pointing to current graphs)"*
- If you had all the time in the world, what visualizations would you design for the website?
  Client: *" Maybe better hydro-graphs, time-lapse videos. "*

3 REQUIREMENTS ELICITED

From the detailed interview and discussion, these were the initial requirements that were elicited for the project :

- **To map a given image at a particular time with respect to water levels (stage data). This will enable the client to see the creek drain into the floodplains around it during storm events and view the corresponding peak in the level of the river.**
  Client quote: *" The picture is like seeing the stage come up in the movie, you would have the hydro-graph paired with that some way so you could see the horizon on the graph so people could visualize what 1.2 meters stage at stream lab really looks like from the picture."*
- **Visualizing stream level (stage data) to rainfall, turbidity and conductivity data**

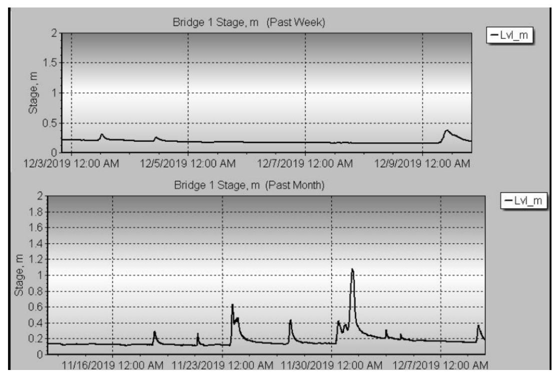

Figure 2: Current Visualization for Stage vs Time data

  Client quote: *"Stage versus turbid or stage versus conductivity, those are the interesting relationships are stage versus rainfall, even to show that, because that's a lot of the graphs that we do from looking at a storm is going to be this."*
- **To Highlight specific important/relevant events in this visualization. Eg : Storms and Flash Floods.**
  Client quote: *" And I don't know for you guys what your requirements are for your project specifically, but Dr.Hession was thinking, even if you just targeted some of the larger storms that we have these videos for maybe or that we have this data for, and just helping people to visualize how that looks out there. When that storm events happening what what what does this look like"*
- **To improve existing visualization**
  Client quote: *(Q: Do you think these graphs could be revamped?) "Sure, Interesting and well better. I'm sure."*

4 DATA DESCRIPTION

This was the data available to us through a shared Google Drive file :

- Several .dat files containing stage data with Timestamps

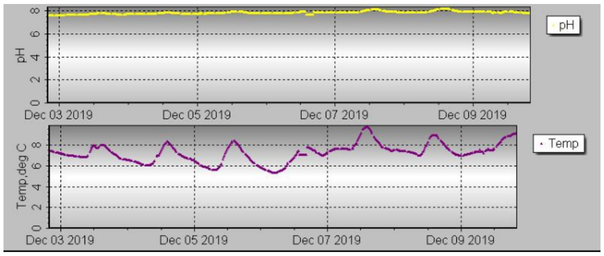

Figure 3: Current Visualization for Water quality vs Time data

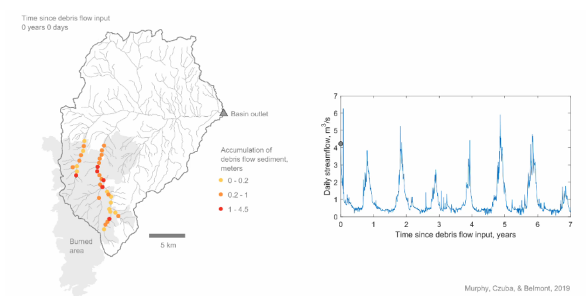

Figure 4: Murphy et al.'s Visualization

- Several .dat files containing values for Dissolved Oxygen levels, Turbidity, Conductivity, pH and Temperature at the bridges with Timestamps
- Several .jpeg images from the observation camera setup with hourly images from the bridges

We had to fetch data from these files for the visualizations every 15 minutes through a Cron job to populate the datasets.

5 RELATED WORK AND SURVEY

5.1 Timelapse Hydrographs

Murphy et al.,[2] in their paper, Post-wildfire sediment cascades: a modeling framework linking debris flow generation and network-scale sediment routing, present a visualization as shown in Figure 4 that juxtaposes a dynamic time-series line chart of the Daily stream flow against a moving image of the basin draining.

Peek et al. [3]from the Centre of Watershed sciences at UC Davis have created timelapse hydrographs of the North Fork American River and the South Fork Yuba River.5 They pulled in river stage data from USGS flow gauges at nearby, but not exact, locations. Using Python and matplotlib, they created a hydrograph for each time. Then, using Perl and ImageMagick, they overlaid the hydrographs with images based on the timestamps and then paired images together based upon closest timestamps.

6 METHODOLOGY

The following methodology was followed to take this project to completion is as follows:

- Analyzing available data manually to understand what we require

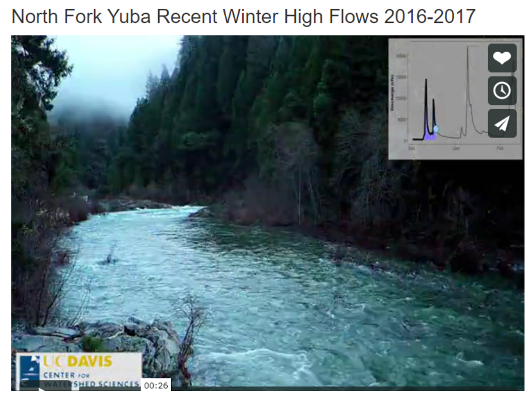

Figure 5: Peek et al.'s Visualization

- Understand the Google Drive API and set up the scripts to fetch data from the .dat file
- Preprocess the data in the .dat file to prepare it for visualizing and create a .csv file
- Use the csv file to create our visualization using D3 and Javascript

7 DATA COLLECTING AND PRE-PROCESSING

The data made available to us from STREAMLab was through a Google Drive folder. This data was being updated at time intervals of 10-15 minutes. In order to obtain this data, the project team required to set up the Google Drive API, create a token and download the necessary information. We further used pandas to preprocess the data and bring it to a form that is appropriate for the visualization. These scripts have been included with the submission. In addition, we had set up a cron job that would fetch data from the server every 15 minutes and pre-process this.

8 IMPLEMENTATION

We implemented three main line charts for our clients to satisfy their various requirements. Detailed information of our design choices, screenshots of our visualization and links to our actual visualization are also provided.

8.1 Timeseries line-graph with dots

This graph is a time-series line graph that has also the data points distinctly visible. This chart is a graph of the Stage Height, i.e., height of Stroubles Creek versus time. In other words, the members of STREAMLab are able to visualize the height of the stream with respect to time for large intervals of time. Please refer to figure ??) as it has been labelled to show each of the components that we mentioned. In our case, the first data point was available in June 2019. Since the amount of data was large, we provided the users with a context graph that can be used to zoom into times of interest. 'E' in figure 7 shows the context graph and 'F' shows the area that is currently being zoomed into. Also, since the frequency of data was very high, we chose to provide the users the height of the stream and the exact time the reading was taken when they

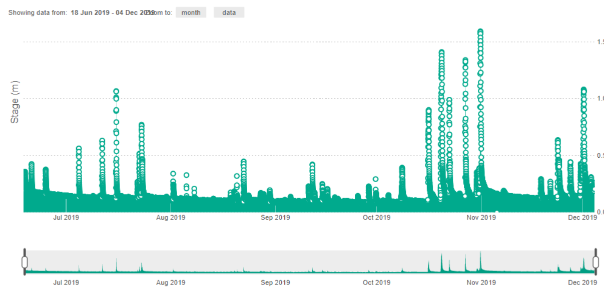

Figure 6: Stream Height (Stage) vs Time graph with dots depicting each data point

hover over a data point (See B in figure 7). Above the chart, we provide the users the ability to zoom into the whole month they are currently looking at and then pan it back to the entire data by using the Zoom to feature's 'Month' and 'Data' buttons respectively (See A in figure 7). An example of this graph (without hover, zoom, etc.) can be seen in figure 6. We also satisfied the requirement of visualizing the physical stream itself alongside the data with this visualization. When the user clicks on any of the data points, the image obtained by one of the STREAMLab cameras (Bridge1, Bridge2 or Tower) will be displayed if available as depicted as G in figure 7. For our demonstration, we chose the pictures taken from the Tower Camera as they were more clear and were the most recent data points available (data available until September). As we can observe, the picture was taken at a time when the graph peaked indicating that the stream was clearly flowing at a greater height than usual. The image shows droplets of water on the camera indicating that it was raining. You can find this interactive visualization here: http://metagrid2.sv.vt.edu/~dramasubramanian/timeseries_dots%20_trial_finalfinal.html.

8.2 Multi-line toggle graph for several parameters

Our clients at the STREAMLab told us that they look at primarily 6 important parameters: Stage Height (Stream Height), Temperature (Temp), Conductivity (Cond), Dissolved Oxygen (domgl), pH and Turbidity (turb) of the water. In addition to visualizing just these parameters as a time-series, they also find it useful to visualize the change in any one of the parameters against the stage height (or the stream's height). To accommodate this, we created a chart that could do both. This graph allows the user to select the parameters they want to see on the screen and displays them. In this, we allow a multi-select. A sample of the graph depicting a single parameter, two parameters and then four parameters respectively in figures 10, 8 and 9 respectively. This visualization uses 3 main components, the main timeseries that shows the current time selected which is selected from the context chart given below. This design is similar to the design presented in Subsection 8.1. The third component is little boxes that toggle the lines (parameters) shown in the visualization. Each parameter has a little box and a corresponding colour that the line in the graph will reflect. This will also serve as a key to understand which line on our graph belongs to which parameter. Next to these boxes are numbers that show the values of each parameter at the point where the cursor is as depicted in figure 8. In this figure, the arrow represents where the cursor is and the values of the parameters like the stage at 1.59 meters and the temperature at 17.56 degrees. These values are available irrespective of whether the parameter is on the graph or not. The interactive visualization can be found here: http://metagrid2.sv.vt.edu/~dramasubramanian/MultiLineToggle_inprogress.html. As is obvious, the Y axis is shared by all the parameters making scaling the graph very difficult. When parameters like temperature that scales upwards of 30 degrees are visualized against parameters like the stage that doesn't exceed 2m in our case, the stage value becomes almost invisible on our graph as you can see in figure 8. Therefore two tackle this, we created pairwise graphs of parameters visualised against stage presented in the next section.

8.3 A vs B Line Graphs

Our rationale behind creating these line graphs are pretty straightforward. Our clients at STREAMLab needed a way to view various water quality parameters against the height of the stream. One of the reasons could be because with an increase in rain, there will also be an increase in turbidity which indicates an eroding basin around the creek. We also made sure with the client that there are no other significant requirements in the other water quality parameters. The resulting plots can be found in figure 11. Once again, we used the entirety of the dataset available to us in 2019 to make this visualization as this would be better to visualize the trends. As in section 8.1, we provided a context graph that will help with zooming in timelines of interest for the user. One issue that we ran into while coming up with this visualization is to keep the second Y axis in its place. The interactive visualizations are available at: http://metagrid2.sv.vt.edu/~dramasubramanian/stagevsCond.html, http://metagrid2.sv.vt.edu/~dramasubramanian/stagevsDomg.html, http://metagrid2.sv.vt.edu/~dramasubramanian/stagevsTemp.html, http://metagrid2.sv.vt.edu/~dramasubramanian/stagevsTurb.html and http://metagrid2.sv.vt.edu/~dramasubramanian/stagevspH.html.

9 EVALUATION

Following the Visualization Validation technique by Tamara Munzner [1], we perform the What, Why, How analysis to arrive at our Visualizations. Figure 12 shows the steps to arrive at a visualization choice. Based on this, our Domain situation lies roughly in Biological systems. Data abstraction is based abstracting the data from the .dat files we have into a generic representation.

Visual encoding is where we choose our visualization and its specifics. According to Munzner, Time-series data should be represented either as scatter plots or as line charts, with the horizontal axis showing time and the vertical axis showing value. The user interacts with this overview to select a time period of interest to show in the scatter plot/ Line chart detail view by changing the position or width of the time slider. We incorporate interactivity by allowing zoom and toggle. Finally, the algorithm part where we design the block in d3 and JavaScript.

Next, we present the evaluation points mentioned to us by Laura, our client.

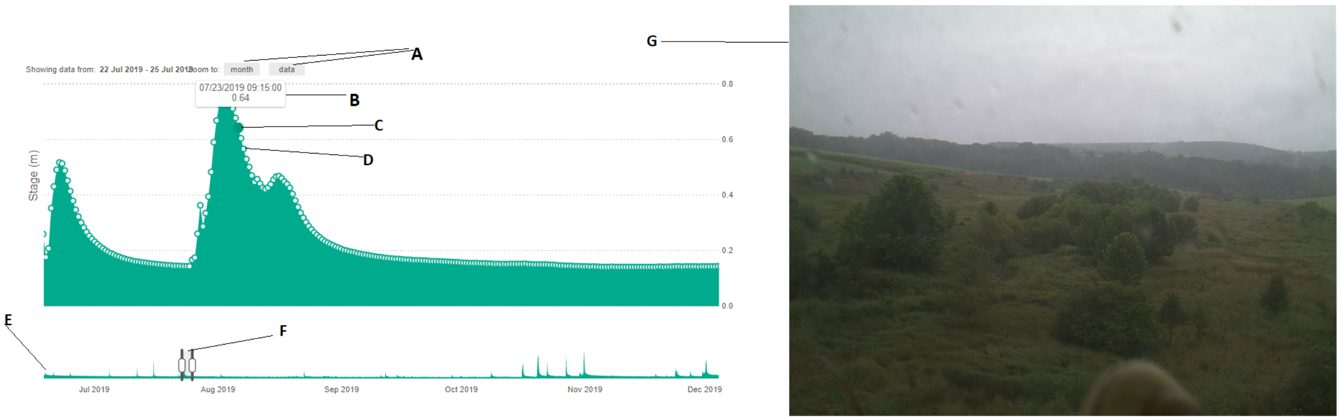

Figure 7: Stream Height (Stage) vs Time graph with dots depicting each data point alongside the image that appears on-click.

Figure 8: Multi-line plot showing stage vs temperature and depicting the hover feature of the viz.

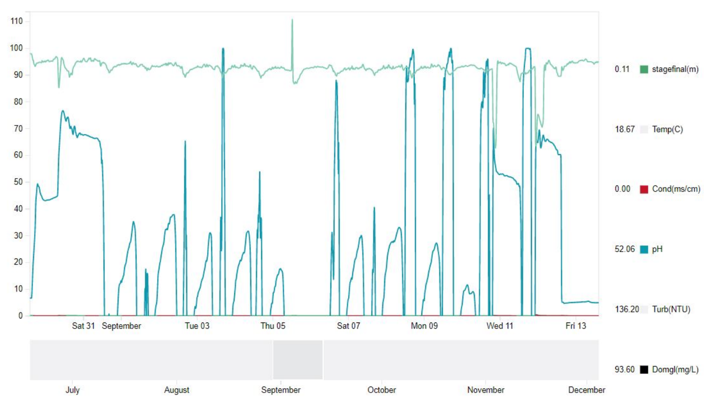

Figure 9: Multi-line plot showing four parameters of stagefinal, conductivity, pH and dissolved oxygen against one another.

Figure 10: Single parameter charts of each of the parameters. From top left to bottom right, the parameters are: Conductivity, Dissolved Oxygen, pH, Temperature, Stage and Turbidity.

- *"So this is looking great to me I mean as far as what we're trying to do to make it more accessible to people where they can see some, especially images where you can click on it and see during the storm events and that's really cool."*
- *"If there are ways to incorporate different types of lines, because I know some people are colorblind so if you were to graph, two parameters much red one that's green they might be hard to separate for someone who's colorblind but sometimes what we'll do is like dash one and dot the other symbol Yes."*

For a more detailed evaluation with our client, please visit this link - https://otter.ai/s/rkrXOTu9TF2Ld2jqNnGhvw.

10 PROBLEMS AND CHALLENGES

- Data is brought in at different intervals, e.g stage at 30 min intervals and images at Random time intervals. Hence while syncing the data, we lose many data points.
- Many Data points did not have the corresponding Image due to equipment failure. Due to this, clicking on some points might bring up the message that No Image is available.
- Some images in the night look dark/completely black. For this very reason, we could not create a Time -lapse video of the images alongside the hydrograph, apart from the fact that the the image dataset is sparsely populated. .
- Missing Stage and Water quality data for long periods of time due to equipment failure
- The data has not currently been passed through a quality check resulting in several anomalous data points obviously visible in the visualization.
- The storage of the data is currently on Google Drive. Authentication of the Google Drive API would be required periodically. This means that the automated the script (through a cron job) would eventually fail at some point.
- The images are available through three cameras producing images of varying qualities. No camera consistently produces the best quality images. The project group had to go through the pictures manually to determine which set to use for this project. We also advised the STREAMLab for the purpose of the project to not name their files differently per camera.

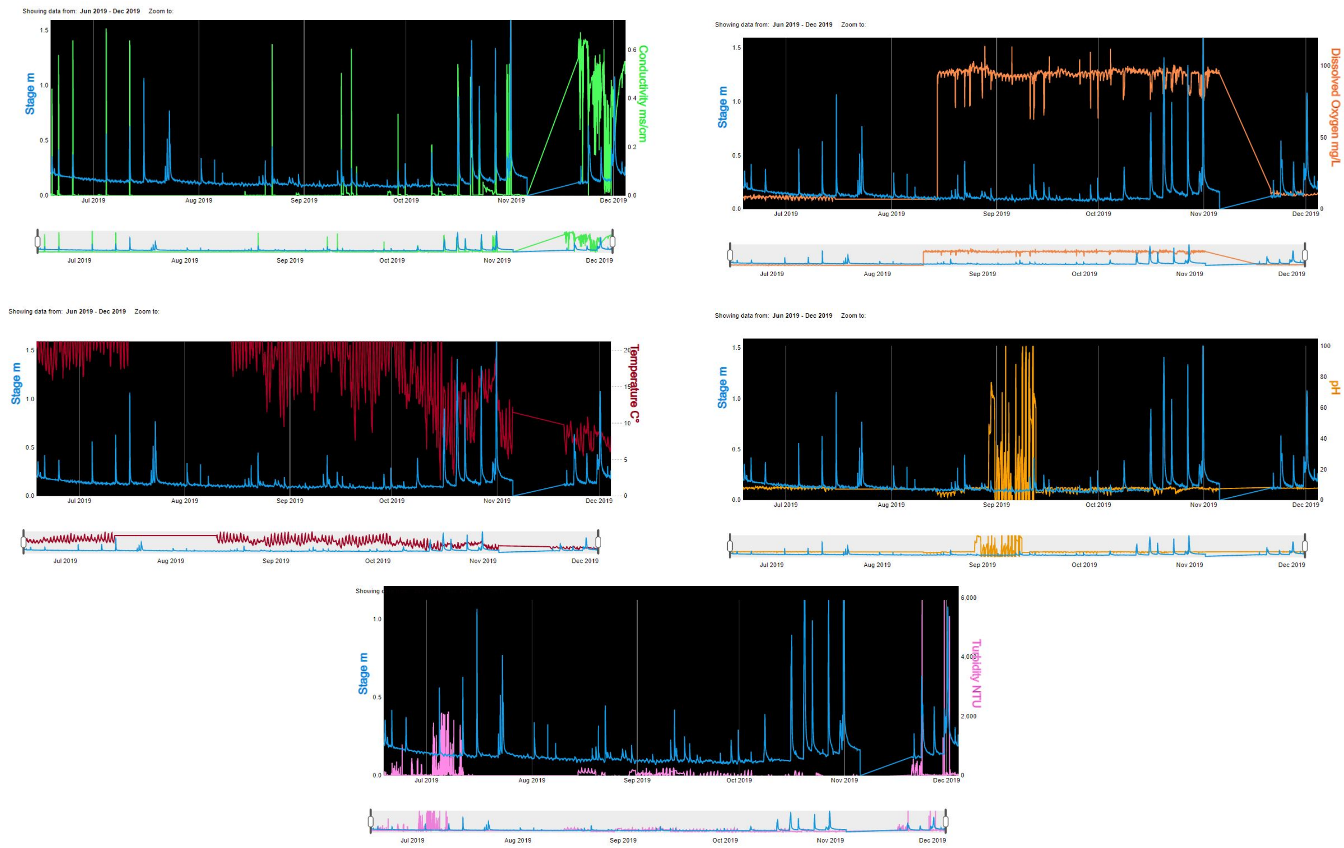

Figure 11: Stage vs x charts of each of the parameters. From top left to bottom right, x is the parameters: Conductivity, Dissolved Oxygen, Temperature, pH, and Turbidity.

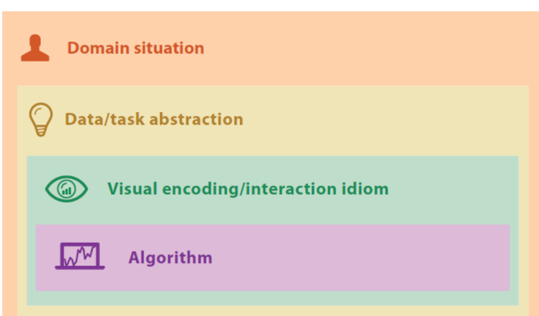

Figure 12: Munzner's Visualization Validation technique

That way, separate scripts may not be required to fetch each one.

11 CONCLUSION AND FUTURE WORK

11.1 Conclusion

To summarize, our project involved designing, implementing and evaluating a visualization for Stroubles Creek's STREAMLab. After a final demo and discussion with Laura, we have successfully fulfilled all the requirements that they provided to us. Additionally, the STREAM lab representatives were happy with the visualizations we provided to them. They acknowledged the problem with the lack of images and the multiple sources of data. We will work with them to improve the data situation as well as our visualizations after more images are available.

11.2 Future Work and Improvements

- With the help and input of the client, merge the various image sources in order to have more images throughout the day
- Put the data into a Database instead of working with .dat files
- After obtaining more consistent images, we can create Time-lapse video of images moving in tandem with the points on the hydrograph

- Improve interactivity further by optimising the code to allow more data at great speed, thereby eliminating the 10 second stutter
- Make the visualizations more friendly to color blind people

REFERENCES

[1] T. Munzner. *Visualization Analysis and Design.* AK Peters Visualization Series. CRC Press, 2015.

[2] B. Murphy, J. Czuba, and P. Belmont. Post-wildfire sediment cascades: A modeling framework linking debris flow generation and network-scale sediment routing. *Earth Surface Processes and Landforms*, 04 2019.

[3] N. S. Ryan Peek, Eric Holmes. Time lapse hydrography- center for watershed sciences, uc davis. 2017.