**SKC Intro to R and K’avi Instruction- Adapted for the Geoscience Alliance Introduction to Forecasting Workshop**

**Written by Georgia Smies**

**Background**

I have worked with tribal governments off and on for 30+ years as a natural resource employee in water quality and as a private consultant. In both of those capacities, I was made aware that the EPA is the single biggest funder of tribal water quality programs and that tribes often hire consultants because they do not have staff trained in data analysis. As a result, funds that should stay on the Reservation often flow outward and at greatly inflated costs.

When I started working at SKC as a faculty member, I resolved to help bridge that gap by training my students to perform data analysis that the EPA requires. The positive outcomes are two-fold. My students can go to work prepared to do “in-house” data analysis and save their Tribe consultant fees. Additionally, students who have strong data analysis skills can expect to earn several thousand dollars more annually because of the value they add to natural resource programs.

**EPA Reporting Requirements**

The EPA has been coordinating with tribal programs since 1973 to strengthen environmental protection of tribal water, air, and land. Most tribes developed their natural resource programs by focusing upon water quality programs first. As a result, many tribes that I work with have at least 30 years of long-term monitoring data of tribal surface waters. Datasets include chemical, physical, and biological parameters of water quality.

The EPA has strict protocols that dictate how samples are collected, analyzed, and assessed. These protocols are summarized in a Quality Assurance Action Plan or QAPP that must be approved before funding is provided for annual monitoring efforts. The QAPP also specifies how data is to be stored, validated, and provides a reporting schedule that Tribes must follow. At a minimum, Tribes must submit annual assessment reports that describe the statistical trends of all monitored parameters. For instance, for each long-term monitoring location, the discharge, water temperature, dissolved oxygen, pH, specific conductance, turbidity, total suspended solids etc. must be analyzed to determine the minimum, mean (or median), and maximum values throughout the annual monitoring period. Chemical and biological data are similarly described. Typically, programs have small annual sample sizes (n<9) so more complex analysis is not possible. The EPA reviews the annual report as part of the funding criteria and makes suggestions to the Tribe as needed.

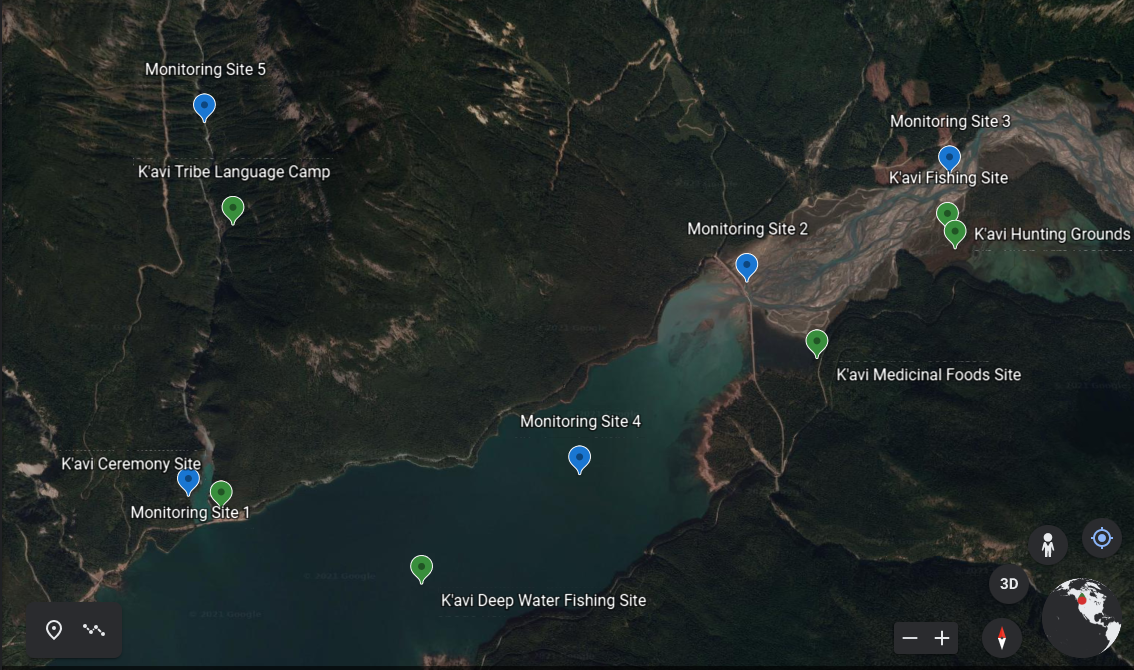
**Current Problems with Data Analysis and Reporting**

Today, most of the Tribes I work with have not analyzed their long-term datasets for trends because they lack the analysis capacity. This is a serious limitation. The Tribes that have hired me to look at their data as a consultant are often not aware of the environmental changes occurring within their streams. Often, water quality program employees have frequent staff turnovers and long-term familiarity with streams and rivers is not “baked” into the program. Additionally, even the programs like the Northern Cheyenne, who do have staff who have been in the same position for greater 20 years lack the ability to communicate what they intrinsically understand because they lack the data analysis training. For instance, I did work for the CSKT and evaluated 20 years of water quality data on valley bottom streams. I found that water temperature had increased by an average 6 degrees Fahrenheit since 2000. I then coupled my finding with tribal fisheries data that indicated changes in the local fishery. This provided powerful information for the tribes to use for management decisions.

Further, tribes do not have a tool(s) that allows them to evaluate their surface waters based upon what is important to them culturally and spiritually. ***The EPA asks only for quantitative analysis of measured parameters. Sites that have “good” water quality based upon these parameters are given priority from a management perspective.*** This analysis does not address what many consider to be the 5 pillars of TEK (native language preservation, traditional food gathering, traditional medicine, cultural practices, spiritual practices, and some would add food sovereignty). As a result, Tribes do not have a means to communicate to the EPA what is important to them and sites that are important culturally may not be given the same protection as those that are assessed using only quantitative assessment of physical variables. In short, western scientific approaches fail to embrace the native world view.

**My Solution to this Problem**

In 2019, I developed an Intro to R class at SKC to help students in our natural resources program become familiar with R programming. Students were taught the basics of coding. They were given the basics in evaluating and cleaning data, importing data, and finally how to write code to statistically analyze and visualize data. In 2020, I refined the class to focus on water quality data sets and EPA reporting requirements. However, I felt my approach was still lacking since it did not address TEK and the specific needs of Tribes. In 2021 I decided to create a new framework for my R class. I created an imaginary tribe called the K’avi (my daughter thought this tribe should have a powerful medicinal source that gave them energy and cleared their thinking….what you and I call coffee and thus the name was born). I needed a fictional tribe because I did not want to focus on any indigenous group that I teach. I also needed an imaginary reservation with ample surface water resources. I used Google Maps to zoom in on a remote portion of Banff National Park that had high elevation streams fed by glaciers. The watershed in my “Reservation” supported a rich complex of streams, rivers, wetlands, and a large glacial lake.



I created a 10 year database of physical water quality parameters and built in some temperature, dissolved oxygen, and specific conductance trends to mirror what I have seen in datasets of tribes I have worked with in western Montana.

To incorporate the TEK element, I added a layer of sites adjacent to or near the long-term water quality monitoring sites that were important to the K’avi. Specifically, there is a language camp, lake and river fishing sites, spiritual ceremony site, hunting area, and a site for gathering medicinal foods. The tricky bit was going back through my 10 year database and ensuring that sites had varying degrees of water quality to mirror real life management issues that I see Tribes face. For example, Tribes may have sites for sweat lodges along river margins that have been important to many families for sites are rarely monitored and….even if they are….tribes have no way to place priority on these sites for management and protection. They often get overlooked.

I started the quarter teaching my students the basics of R. Once they were comfortable with evaluating and importing data, I taught them the EPA reporting requirements and had them write codes that would enable them to evaluate annual trends as well as long-term trends. These are “stock” codes that they can take with them into their first job and I am very excited about this element of my class. All of my students this year were women and two were mothers. This training will be very helpful to them financially I hope.

Next, I taught them how to visualize this data in a very basic fashion.

**The Really Exciting Part of Class**

Helena Kleiner wrote codes for me that enabled my students to do two very exciting things. First, she developed codes that enabled them to visualize their data. Next, she created codes that enabled students to rank the water quality of all sample sites using only quantitative means (e.g. the way the EPA and Tribes do monitoring today). She also wrote codes that provided a visualization of these site rankings.

Next, Helena wrote code to “rank” TEK values on the K’avi Reservation by taking a survey of tribal members. My students were asked to rank TEK values as part of an in-class discussion activity and it was a wonderful (albeit thorny) discussion. For instance, Tribal members from Arizona who have retained their native language placed a higher value on traditional food gathering than Montana tribal members who are struggling to retain their language. This was a very intentional exercise in my class because this is a very real-life challenge that native natural resource managers contend with when working with their membership. The class did come to a consensus and I gave those values to Helena.

Helena then produced codes that allowed students to visualize the sites based upon their TEK rankings. We compared the quantitative rankings to the TEK rankings and were surprised at the new depth in our analysis.

The culminating activity in the class was to combine the quantitative and TEK values into a final data visualization that sorted out all of the sites based upon an equal weight being placed on “white” science and TEK.

Helena’s code added the element that I had long wanted for my class but simply did not know how to achieve. **It was a major breakthrough.**

**K’avi Exercise Instructions for the 2022 GA Forecasting Workshop**

**Written by Cazimir Kowalski**

This exercise has been adapted from the above classroom activity. Our goals today are to get hands-on experience with data management and analysis in R, to explore the relationship between Western scientific priorities and non-Western community values, and to create meaningful data visualizations combining water quality data and community values.

All of the resources necessary for this activity are stored in [this Drive folder](https://drive.google.com/drive/folders/16R6Bfzr5pA5nztGt-sYUzYwSij5tZjyJ?usp=sharing).

The first step in this exercise is to make sure that R and RStudio are installed on at least one laptop in your group. In your file browser, navigate to your documents folder, and create a folder called “K'avi Tribe”. Download the contents of the above Google Drive to this folder, and make sure they are unzipped and not in any other folder. Then, launch RStudio and open the file “1-Kavi\_Tribe\_Water\_Quality.R”. Make your way through the code in the file, and make sure to take your time with your group to read the comments and understand what each line of code does. If you run into difficulties, flag down the TA/put a red sticky note on your laptop. Once you have successfully completed this first code file, place a green sticky note on your laptop to indicate you are finished. Once everyone is finished, we will reconvene for a discussion.

With your group, please discuss the five community sites present in the exercise: the language camp, the hunting grounds, the fishing site, the medicinal food gathering site, and the ceremony site. Once your group has thoroughly discussed the options and come to a consensus on the ranking, please fill out [this survey](https://forms.gle/xMGPVjvPKdrLViMSA) ranking the sites in the order decided by your group. Please only fill out the survey once per group.

Once we return from lunch, we will start part two of this exercise. First, download [this file](https://docs.google.com/spreadsheets/d/1i6jsJsRcqjdWMR1i9lTZMSb5gJNcNF6cFMx8MAP3tzk/edit?usp=sharing) (Important: do not download the file until you return from lunch). Then, open the file labeled “2-Visualizing\_TEK.R” (which should be in ~/Documents/K’avi Tribe/) in RStudio. Follow the instructions in the code file, and make your way through the code, formatting and visualizing the survey data. Once complete, move on to the third exercise, “3-K\_avi\_water\_rank.R”. In this exercise you will combine the analyzed survey data with the water quality data to determine which sites have the highest quality and how that correlates with the cultural sites that were identified as most important.

Finally, time permitting, start exercise 4, “4-Bacteria\_Nutrient\_Code\_Integration.R”, and work your way through the file. This process is similar to the last exercise, except instead of abiotic water quality metrics, it is specifically biotic metrics.

Once we have completed all four exercises, we will end up with several different kinds of visualizations, and will have experience with basic R data management, analysis, and visualization. We also will have attempted to integrate community values into data science, something that is generally ignored by Western scientists. This puts us in a good place for our closing discussion surrounding ethics in forecasting, community engagement in data science, and the accessibility of forecasting.