This assignment should be submitted online through the course canvas page as a PDF. You are allowed and encouraged to work with others on this assignment, however you must do your own work and include the names of those that you worked with prior to your first problem. Solutions must contain sufficient explanatory details, be typed, and document any important assumptions. The Problem number in bold is followed by the number of points in parentheses.

Each problem should be answered on a separate page. Every figure or plot should be accompanied by a caption describing the figure in a matter-of-fact manner. Underlined text describes the task generically to aid in substituting your own data. Italicized text denotes what you should turn in.

1. (7) Determining an appropriate descriptive statistic of the central tendency and the modal value for a non-normal distribution. Write a script to load and plot the time series within file 'USGS10109000_DM.txt'. The Logan River above first dam. Determine an appropriate measure of central tendency for this time series and determine what is most likely the modal value. The most frequent value in a continuous dataset is not a useful descriptor of the mode, rather the most frequent value between a reasonable range of values is a better descriptor.

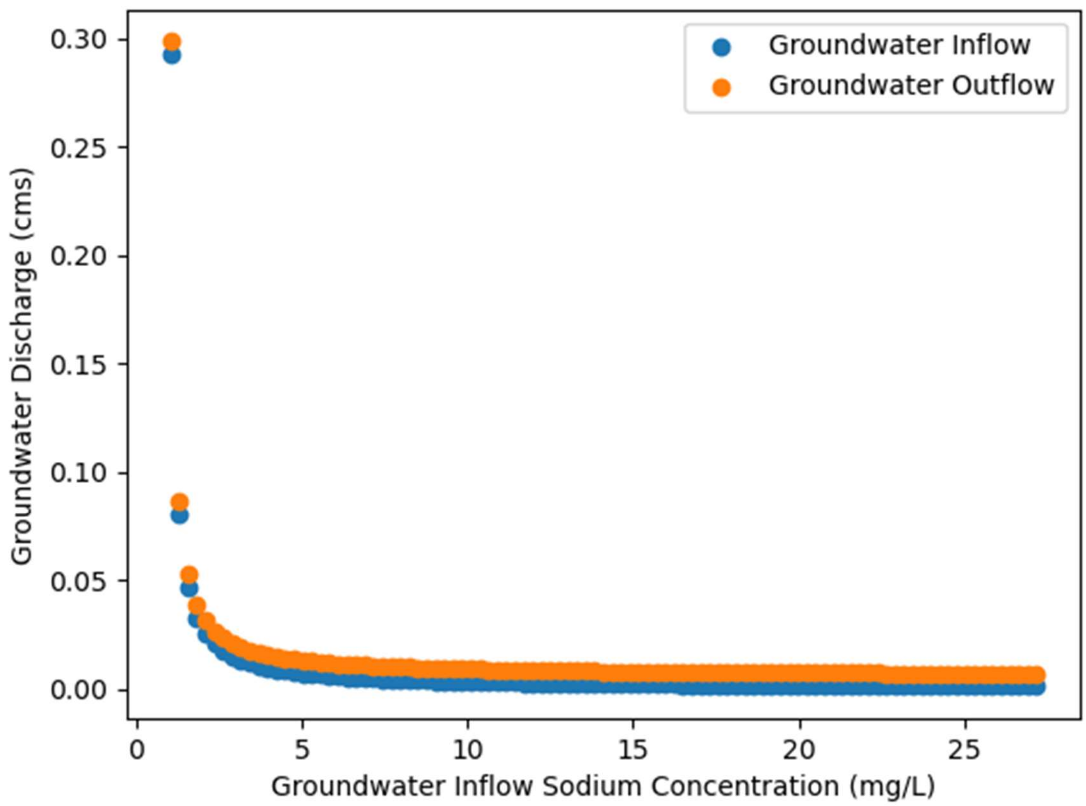

Figure 1. Solutions for a mass balance, iteratively solving for groundwater inflow rate and groundwater outflow rate by varying the groundwater inflow concentration

The relationship between groundwater inflow concentration and groundwater inflow rate is not linear (Figure 1). The relationship appears to be best approximated by log-log (Figure 2). I varied inflow concentration linearly to solve for groundwater discharge in and out. Perhaps I should vary groundwater inflow logarithmically? Even in log space, the distribution (Figure 3) is not normal, though log scale displays more than linear scale.

The modal range for the inflow is between approximately 0.001 and 0.002 cms (Figure 3). The mean is influenced by the outliers. The geometric mean or the median are most appropriate because the distribution is positively skewed.

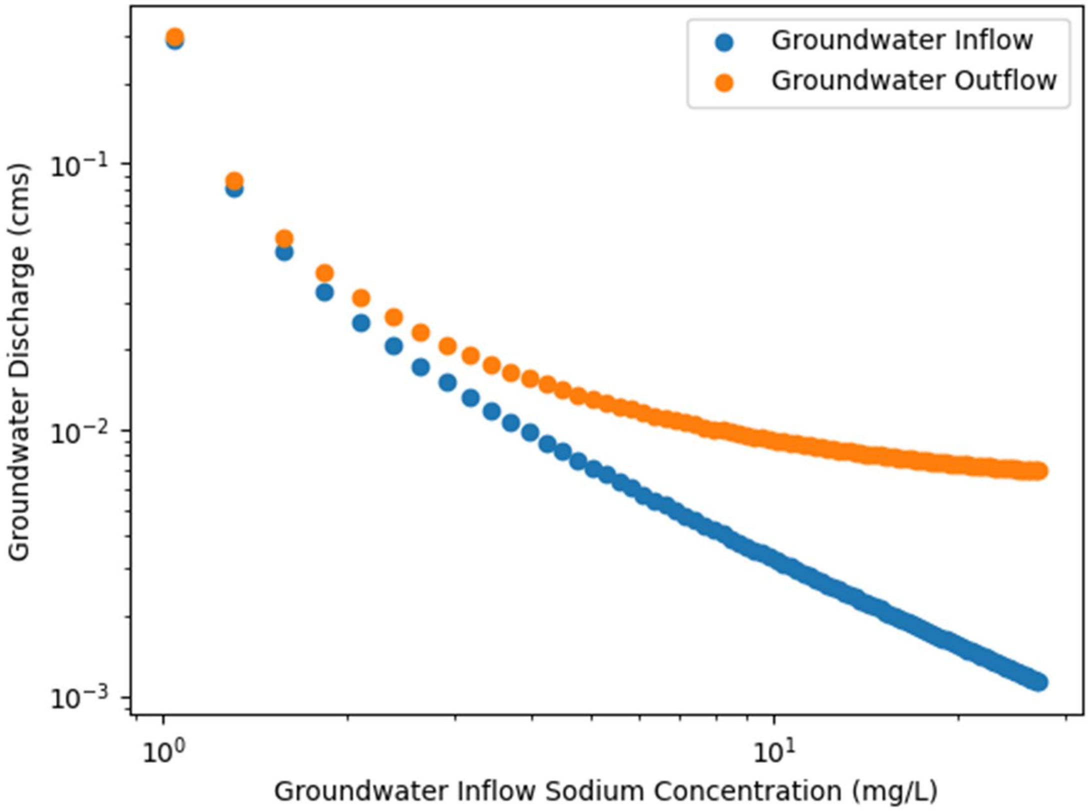

Figure 2. Solutions for a mass balance, iteratively solving for groundwater inflow rate and groundwater outflow rate by varying the groundwater inflow concentration (log-log)

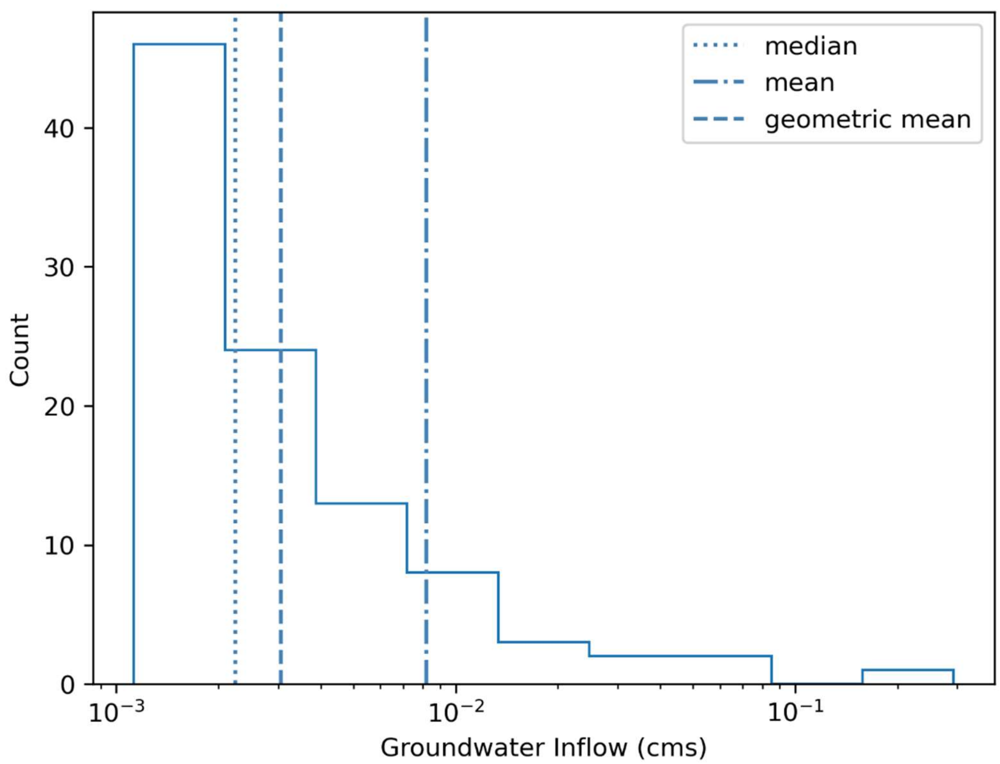

Figure 3. Groundwater inflow rate distribution and measures of central tendency

2. (6) Comparing two distributions from different time periods. Using the dataset from problem 1, compute the probability density function (histogram) or an appropriate cumulative distribution function for two different non-overlapping 10-year periods. Turn in a plot showing your two distributions and appropriately label them or denote which is which within the caption.

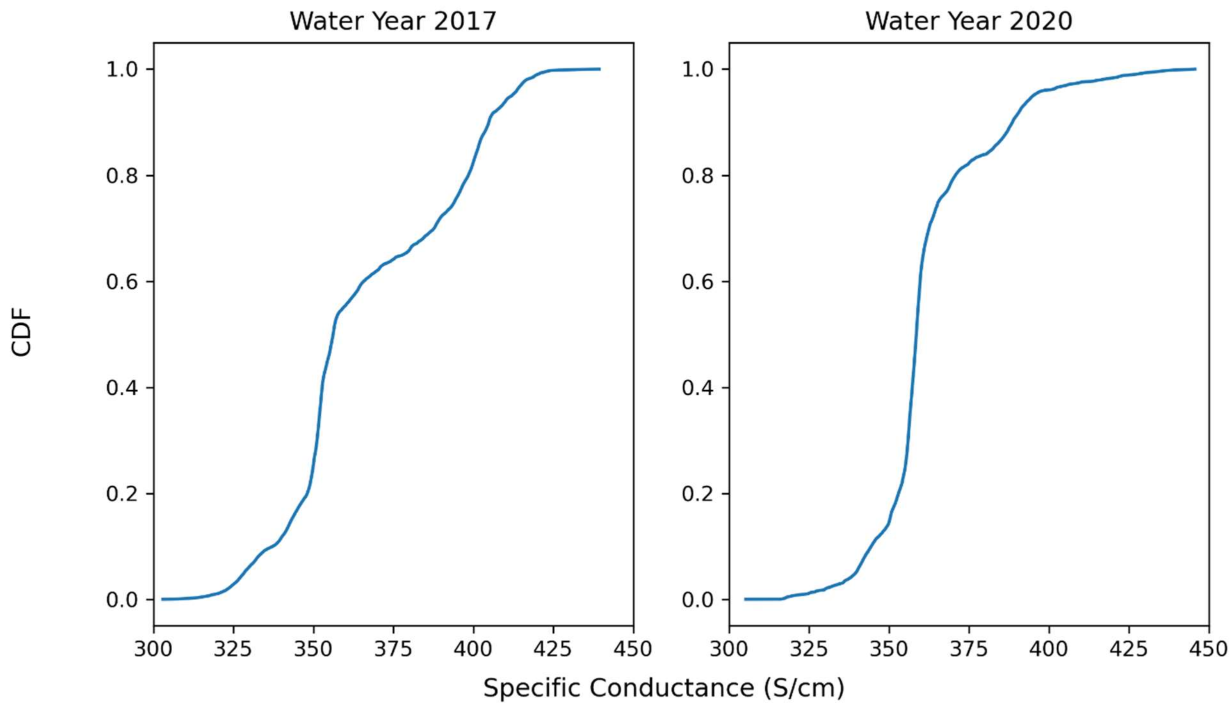

Figure 4. Specific conductance CDFs in a wet year (2017) and a dry year (2020) at Ricks Spring in Logan Canyon

The main difference in the graphs is a large bump in the middle of the cumulative distribution function (CDF). I interpret this has more frequent lower concentrations of specific conductance in 2017 than 2020. Also, since the slope is steeper in 2020, there is less of a steady distribution, where as in 2017 the distribution is more steady.

3. (6) Comparing two distributions with similar distributional shapes. Write a script to load in river grain size data from two the sites within the Logan_river_grainsize folder. Which section (dataset) of the river has a larger grainsize? What is an appropriate transform/scaling/standardization of these two datasets? Support your answers with appropriate descriptive statistics and plots.

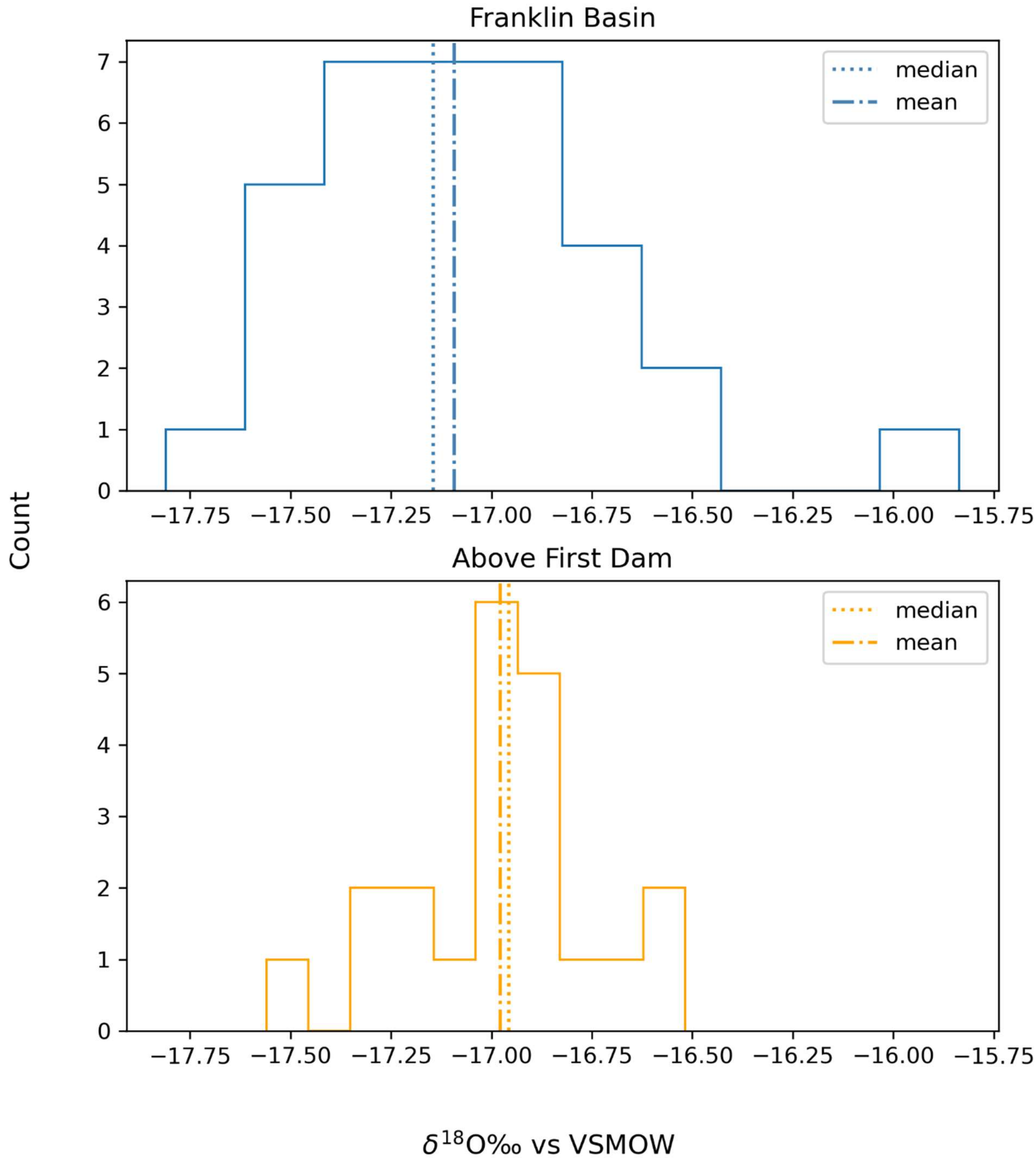

Figure 5. PDFs for oxygen isotopes in the Logan River at Franklin Basin and above First Dam

The oxygen isotope ratio at First Dam is slightly closer to the standard ratio because the value is less negative. These datasets are close to normally distributed with linear scales because the mean, median, and mode for each distribution are very close (Figure 5). Additionally, the plots visually fit the general description for normally distributed: peaking in the middle, with even tails on either side.

4. (7) Examining different examples and finding some context for these methods. Find a scientific peer reviewed paper relevant to your research, work, or interests that utilizes one of the methods that we have discussed in class, practice, or within an assignment. (Worry less about getting the terminology correct and focus on describing things in your own words. A few sentences will suffice for each answer.) a.) What method did they use? b.) What other methods do they use throughout the paper? (Provide a list) c.) Why is this paper relevant to you and what do you like about this paper? d.) Provide the title of the paper and either load the paper as a PDF along with your homework submission or provide the doi here.

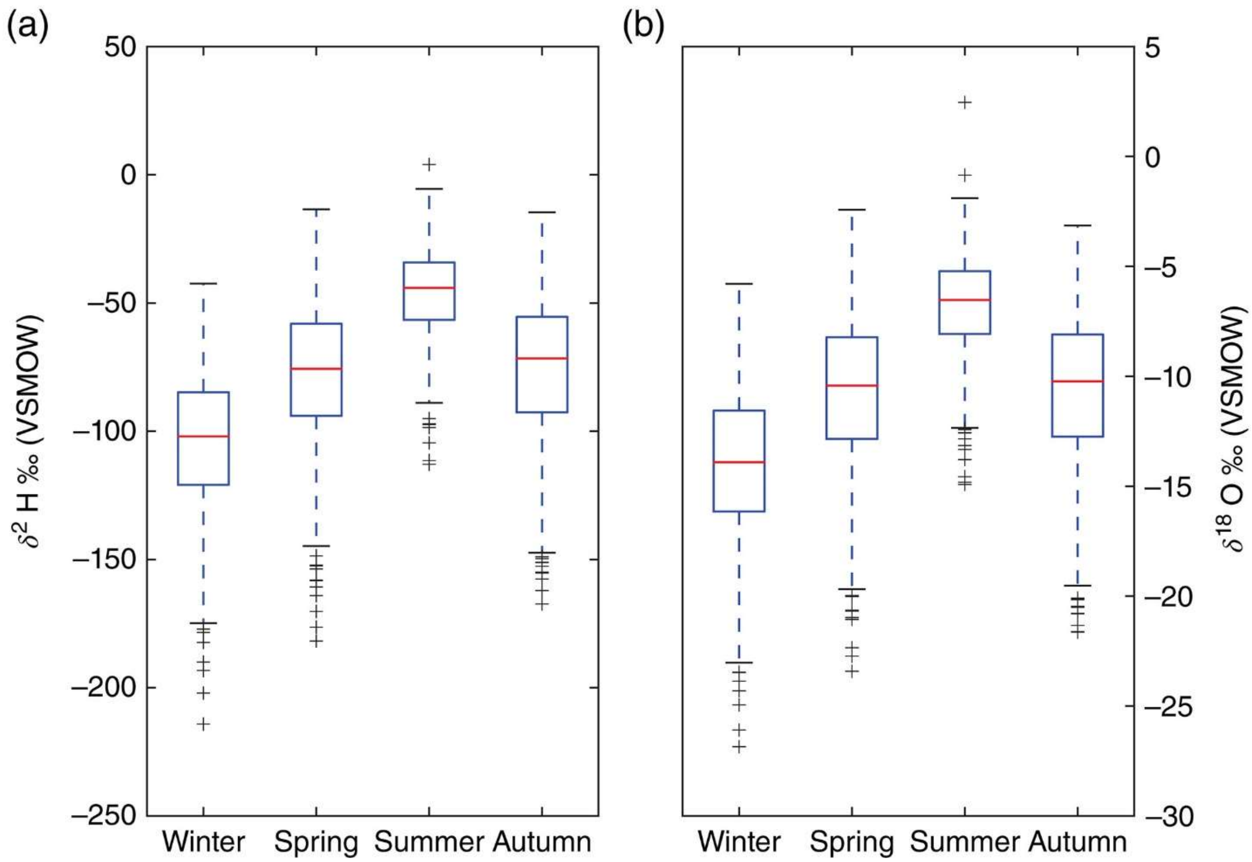

Figure 6. Seasonal variation of δ2H and δ18O in precipitation samples collected by the GNIP (Global Network of Isotopes in Precipitation) network of gauging stations in Switzerland (data from 1966 to 2014) reproduced from https://doi.org/10.1002/wat2.1311

a. Beria et al. used box and whisker plots. The box and whisker plots represent the seasonal distribution of hydrogen and oxygen isotopes relative to the VSMOW standard. They used resistant measures of central tendency and variability (median and interquartile range).
b. Beria et al. used other plotting and statistical methods as well:
    - Scatter plots to display latitude, elevation, and annual snow fall depth or annual snowfall to total precipitation ratio
    - Linear regression to find a relationship between elevation and oxygen or hydrogen isotope ratios

c. This paper is relevant to me because there are case studies on the utility of stable isotopes to estimate seasonality in the groundwater recharge. In the Logan watershed, I am trying to understand the karst flow paths, so understanding the recharge mechanisms is very helpful.
d. Beria H, Larsen JR, Ceperley NC, Michelon A, Vennemann T, Schaefli B. Understanding snow hydrological processes through the lens of stable water isotopes. *WIREs Water*. 2018;5:e1311. https://doi.org/10.1002/wat2.1311

5. (7) <u>Loading multiple datasets, computing statistics, and saving the new data.</u> Write script to compute the median from the 10 files within the Logan_river_grainsize folder. Save the medians as a two-column dataset with one column as a site ID or site name and the other column for the medians. Copy and paste contents of that file as your submission for this problem. (Feel free to substitute other datasets for the grainsize data, just try to find a minimum of four files so that the iteration loop is worth writing.)

SiteCode	2H	18O
BSS_CONF	-126.3	-17.0
DS_CONF_A	-125.9	-16.9
LCS_CONF	-126.4	-17.0
LR_FB_BA	-126.3	-17.2
LR_TG_BA	-126.7	-17.1
RHF_CONF_A	-130.4	-17.4
RS_CONF_A	-127.2	-17.1
SHC_CONF	-125.8	-16.9
TF_CONF_A	-129.1	-17.3
WCS_CONF	-126.0	-17.0

6. (7) Working with two datasets and exploring correlations between two variables. Using data from the Logan River Observatory (data page linked below), download the discharge and a water quality parameter (temperature, specific conductance, DO, pH, Turbidity, etc) for a single gage. Explore the relation between these two variables. At a minimum, turn in a scatter plot of the two variables and any descriptive statistics that seem relevant to the analysis. Additionally describe your process for loading/prepping the dataset for python. https://lrodata.usu.edu/ (Feel free to use a climate station if it is more appropriate to your research or interests. Acquiring and cleaning/wrangling data for analysis is usually about 80% of the work.)

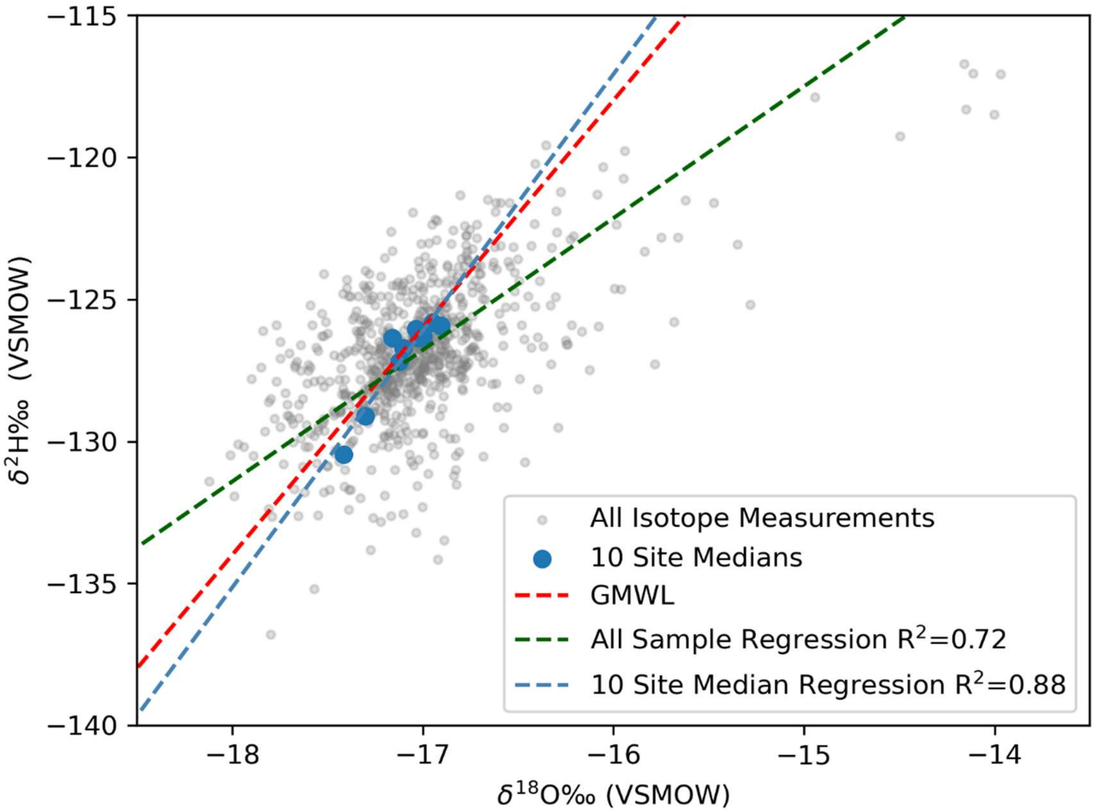

Figure 7. Logan River stable isotope ratios in comparison to the global meteoric water line (GMWL)

Dr. Neilson's research team collected this data over the past 8 years. The data is on the database her team uses. I extracted the data from a compilation of data from sampling events. I just loaded the compilation as a csv file and then used pandas to select the data I am interested in. Next, I used the medians I calculated in problem 5 to populate the median points. Then I plotted all the data we have. Finally, I plotted linear regressions for the 10 points and for all the data. I compared the linear regressions to the global meteoric water line (GMWL).

The correlation coefficient for the two regression lines are given in the plot. Interestingly, the median linear regression slope is steeper than the GMWL while the linear regression for all data yields a shallower slope. I think the shallow slope is a result of the seasonality of samples.