Math 543 Project 2 Report: Pine Ridge Reservation Water Chemistry

Charles Jason Tinant and Trevor Nicholas

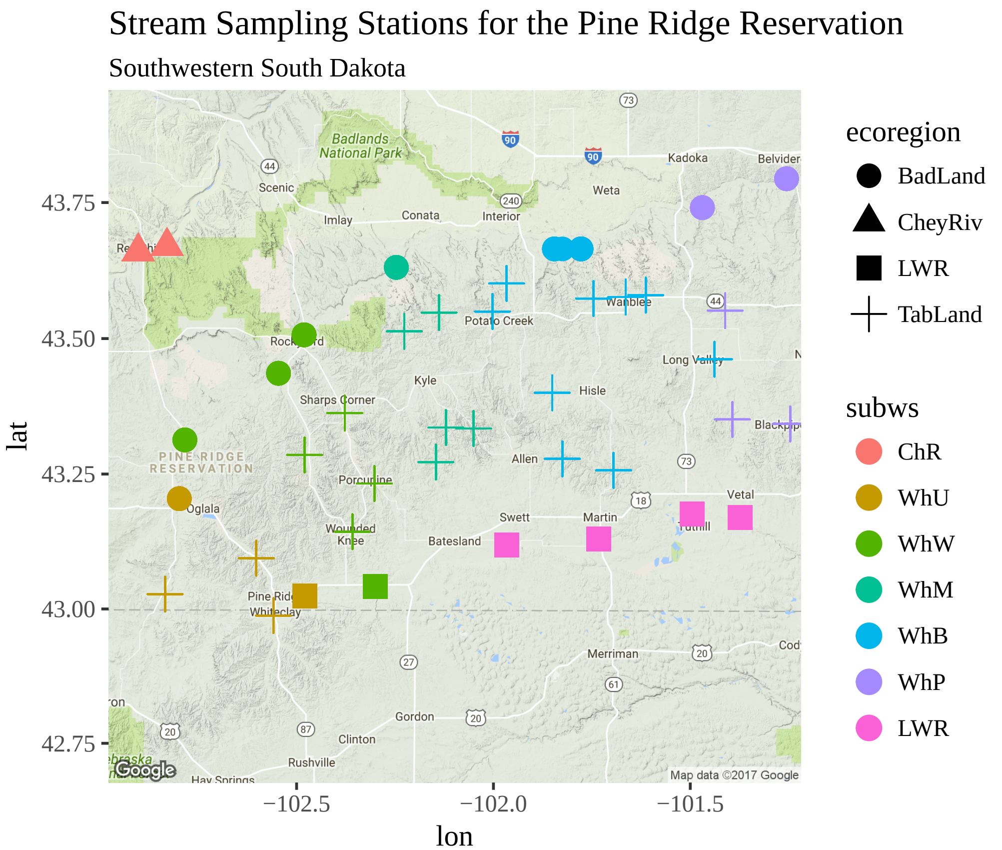
4/28/2017

Overview

States and Tribal Nations are mandated to protect and restore the Nation's waters under the Clean Water Act. Increasing nutrient concentrations are the greatest source of impairment to the Nation's waters. Nitrogen and phosphorus are the nutrients limiting the growth of algal biomass and the likelihood of low dissolved oxygen event occurrence. Increased nutrient concentration over time may explain the substantial change in Pine Ridge reservation (PRR) stream biological integrity in that was first observed in the mid-2000s. Challenges in evaluating changes in nutrient concentration in PRR streams are regional physiographic heterogeneity, possible transcription errors, incomplete sampling data, and a mixed population of samples taken during events and during base flows. A first step in testing a hypothesis that nutrient concentration is increasing over time is to estimate missing nitrogen and phosphorus data based on an informed prior. We assume that geochemically similar stations, in other words stations with similar major anion and cation concentrations and ratios, will have similar land uses that can be used to as an informed prior to estimate nutrient concentrations. The following report outlines our method for estimating missing data, identifying functional groups, and testing for significant differences among the groups identified in the analysis using classical statistical methods. We discuss our findings and the next steps in the latter sections of the report.

Background

The Oglala Sioux Tribe (OST) Water Pollution Control Program is responsible for assessing non-point source impacts to streams and developing watershed protection plans to characterize impacts to water quality and to identify and implement best management practices to restore water quality. The Tribe has established approximately forty surface-water quality sampling stations on Pine Ridge Reservation lands. The sampling stations have been grouped by location into subwatersheds. Water quality field data and water samples are collected at monthly intervals, typically from May to October, by OST Environmental Protection Program staff for one to two subwatersheds per year. The water samples are taken to an Environmental Protection Agency (EPA) certified laboratory for water quality testing and the results are returned to the Tribe. The data represent observational rather than experimental data. Some degree of randomization exists in terms of water quality concentrations because of: 1) individual water quality parameter concentrations are interdependent and concentrations may change as a result of calcite or fluorapatite precipitation, and 2) increased suspended sediment loads following precipitation events.



The Oglala Lakota College (OLC) Science Technology Engineering and Mathematics (STEM) Department staff began partnering with the Tribe in 2011 to: 1) analyze and compare water quality data collected by the Oglala Sioux Tribe (OST) Water Quality Program historically from 1993-2011, and for 2012-2013, 2) integrate macroinvertebrate sampling data collected historically from 1993-2011 are for 2012-2013, and to 3) to refine recommendations for future monitoring and implementation of best management practices (BMPs).

The regional climate is typical of the Northern Great Plains with a median temperature of 50F in January and 73 degrees F in July (SDSU climate website, accessed March 6, 2013). The average annual precipitation between the years 1971 and 2000 at the Manderson 3 NE station is 19.3 inches with January as the driest month (0.39 inches) and June as the wettest month (3.18 inches). The Manderson station is slightly drier than the rest of the south-central region of South Dakota, which has an average precipitation of 21.6 +/- 4.5 inches (standard deviation) (SDSU climate website, accessed March 6, 2013). The Keya Paha Tablelands receives a mean annual precipitation of 16-20 inches while the Pine Ridge Escarpment receives 16-17 inches (Bryce et al, 1998). The Great Plains has an historical record of major drought (Meyer et al. 1999). The 2012 drought resulted in zero flow conditions in the White River and a majority of White River tributary streams (Tinant personal observation).

The amount and frequency of hydrological exchange between the floodplain, groundwater and the stream depends on geology and alluvial composition. Upper White River tributaries begin in central Nebraska as springs. Middle White River tributaries begin in Arikaree Group sandstones and siltstones, and then flow across White River Group volcaniclastic claystones, before draining into the White River (Bryce et al. 1998, Heaken 1999). Sandier units of the Arikaree group form water table aquifers that sustain base flow in the upper reaches of White River tributaries (Heaken 1999). The hydrologic regime shifts from a mixed flow regime to an event-dominated regime as streams cross the Arikaree Group – White River Group contact as infiltration rates decrease and the volume of overland flow increases (Foreman 2006). Infiltration, percolation and ground water flow dominates Sandhills ecoregion hydrology in the Little White River watershed.

Methods

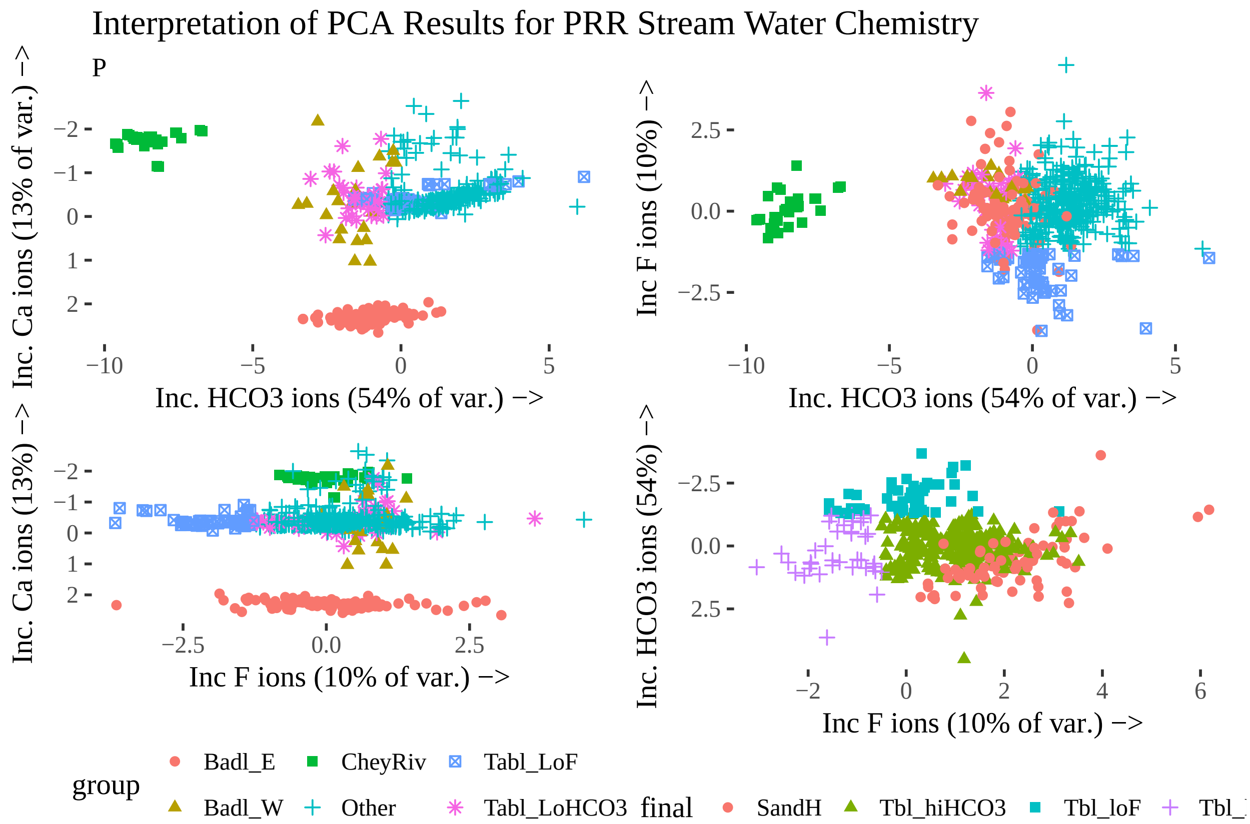
We prepared and validated the raw data as a first step in the analysis. The primary author compiled raw water quality data collected by OST staff from the period of 1993 to 2013 for approximately forty long-term surface water sampling locations across the reservations in an Excel spreadsheet. The spreadsheet was converted in Excel to a 'csv' format and imported into an R statistical software package running under R Studio. The format of the raw data was standardized, empty columns removed, grouping variables assigned, and data with values below the laboratory instrument detection limit assigned a value of half the detection limit using the 'janitor' and 'dplyr' packages. Date and time variables were encoded as time-series variables using the 'lubridate' package. Trace metals concentrations, which were available for approximately half of the observations, and biological variables were removed from the analysis. Field observations of pH and conductivity were also removed as validation of these parameters indicated instrument calibration was a major factor in the variance of these data.

We prepared the cleaned data for multivariate analysis by estimating missing non-nutrient major cation and anion concentrations. We filled in missing data using group means from level four ecoregions, sensu Obernick et al, using the dplyr::group_by function. Milliequvilants per liter ionic concentrations were derived by multiplying the absolute value of charge by mass concentration and dividing by the atomic weight. Bicarbonate concentrations for each of the observations were calculated by subtracting the sum of cations from the sum of cations.

We identified water quality groups following a principal component analysis (PCA) of prepared data that was transformed prior to analysis. We plotted plots of untransformed and transformed non-nutrient ion data using the 'ggpairs' package as a first step to validating the assumption of multivariate normality. We applied transformations using the 'dplyr' package after estimating Box-Cox transformation parameters using the 'AID' package. We calculated an evaluated a PCA using the 'prcomp' function to calculate principal components, and the 'cor' and 'eigen' functions to calculate a correlation matrix and calculate eigenvectors. We plotted the resulting eigenvectors as a scree-plot using the 'ggplot2' package. We tidied the significant PCA axes into a data frame using the broom::tidy, broom::augment functions, and tidyr::spread functions. We joined the results with factor variables using the dplyr:: left_join function. We plotted the PCA loadings using the ggfortify::autoplot function. We identified initial water quality groups from natural breaks along the PCA-1, PCA-2, and PCA-3 axes. We identified final water quality groups after plotting the initial water quality groups on a map using the 'ggmap' package.

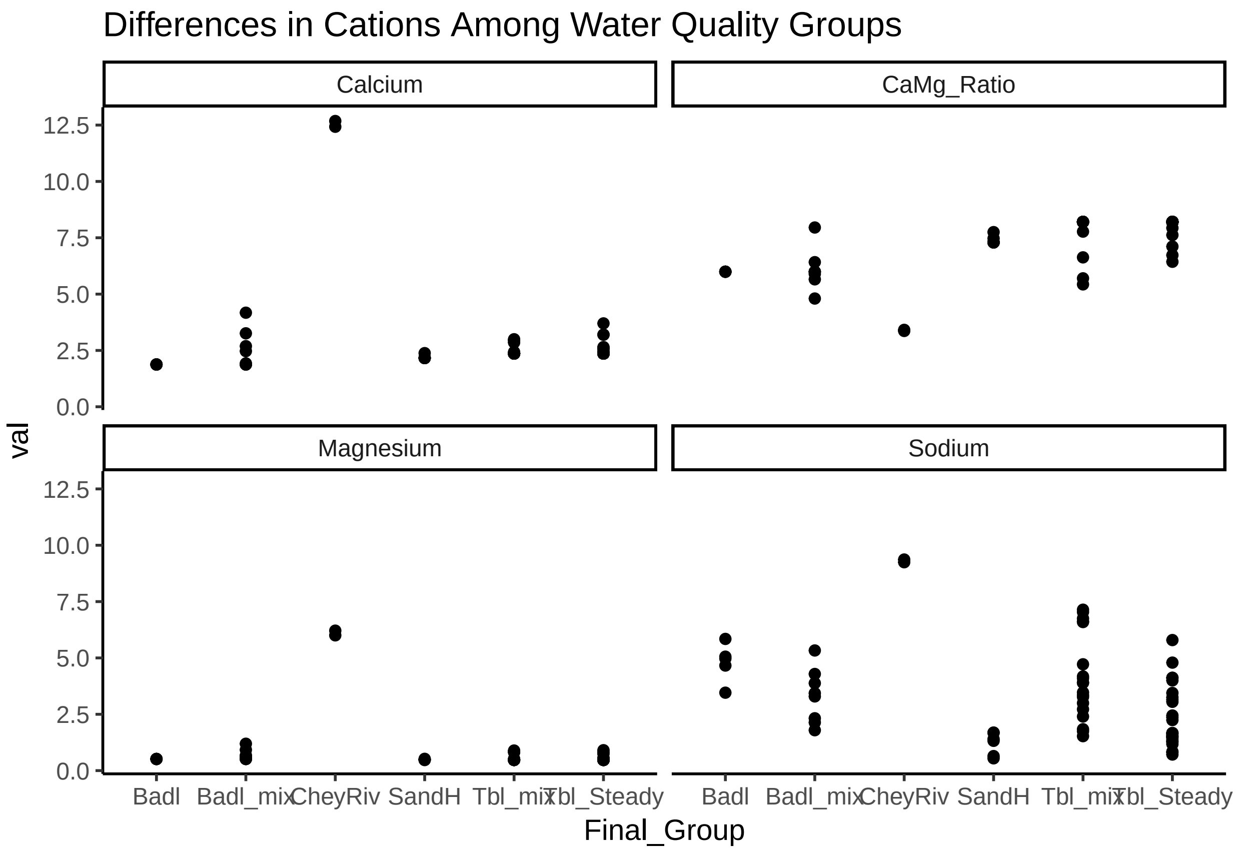
Results

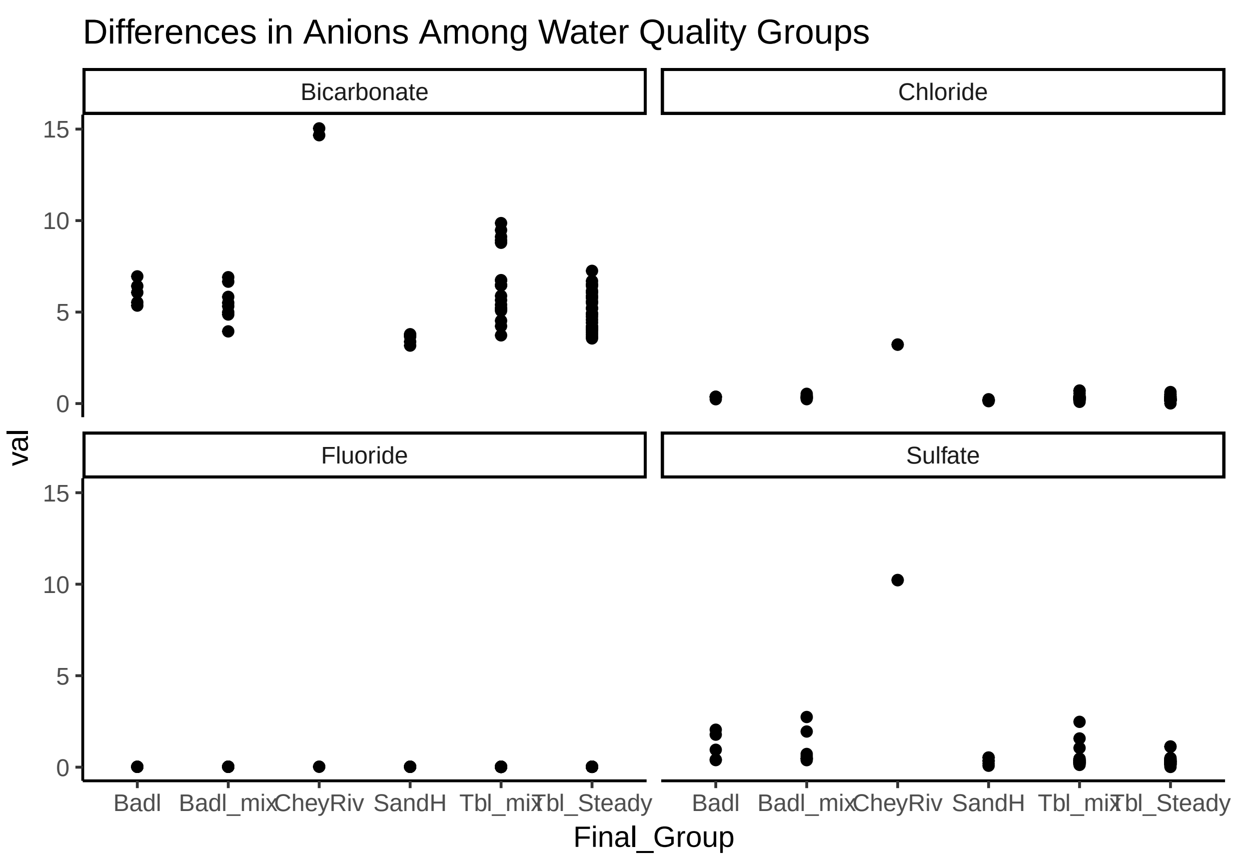
We found that the PCA analysis of stream water samples explained about 77% of the variance in major non-nutrient ions along three significant axes. The first PCA-axis explained 54% of variance and is interpreted as describing a gradient of higher conductivity and lower carbonate waters to low conductivity high carbonate waters. The second PCA-axis explained 13% of variance and is interpreted as describing a gradient of low to high calcium waters. The third PCA axis explained 10% of variance and is interpreted as describing a gradient of low to high fluorite waters. However, we feel somewhat cautious of our analysis as the assumption of normality was not met, as shown in pairs plots located in the appendices of this report. The appendices also include information on the PCA loadings.

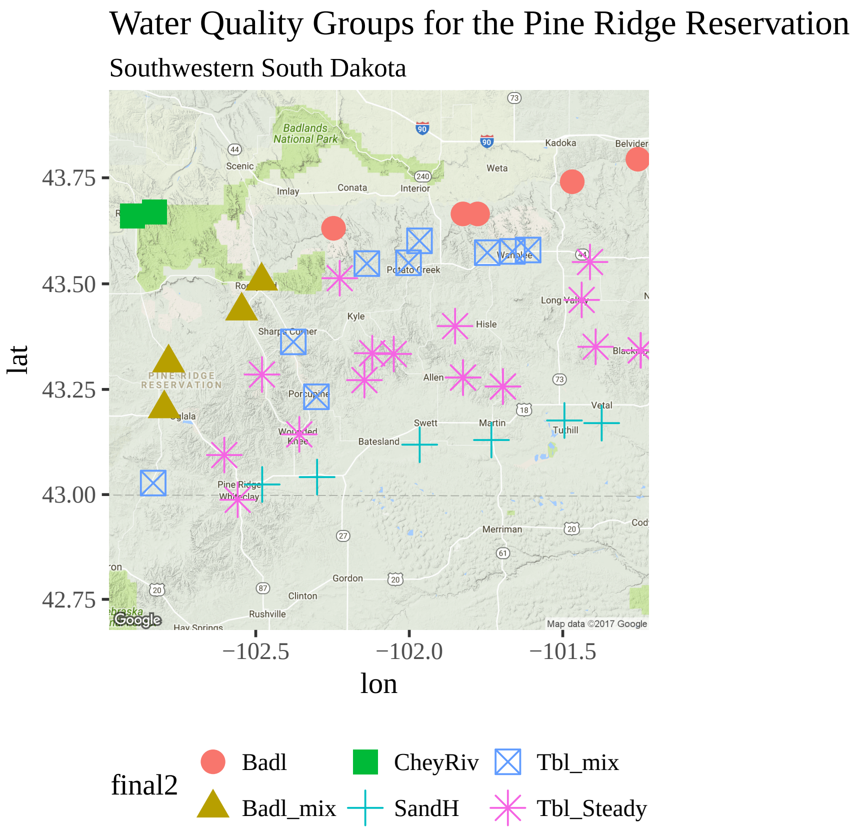


Discussion

We identify two water quality groups in addition to the water quality groups observed prior to PCA analysis. We interpret the Badlands ecoregion as consisting of two water quality groups: a group with relatively steady cation concentrations and a mixed group with varying calcium, sodium, and bicarbonate. The Cheyenne ecoregion consists of high ionic concentration waters with a low calcium magnesium ratio and high sulfates. The Sandhills ecoregion consists of low total ionic concentration and sodium waters with varying calcium concentrations. We interpret the Tablelands ecoregion as consisting of two water quality groups: a group with relatively steady ionic concentrations and a mixed group made up of waters with varying calcium, sodium, bicarbonate, sulfate, and fluoride concentrations. We interpret the differences in the Badlands and Tablelands steady and mixed groups to be the result of calcite and/or fluorapatite precipitation.


Differences in Cations Among Water Quality Groups
Calcium
CaMg_Ratio
Magnesium
Sodium
val
12.5
10.0
7.5
5.0
2.5
0.0
Badl
Badl_mix
CheyRiv
SandH
Tbl_mix
Tbl_Steady
Final_Group

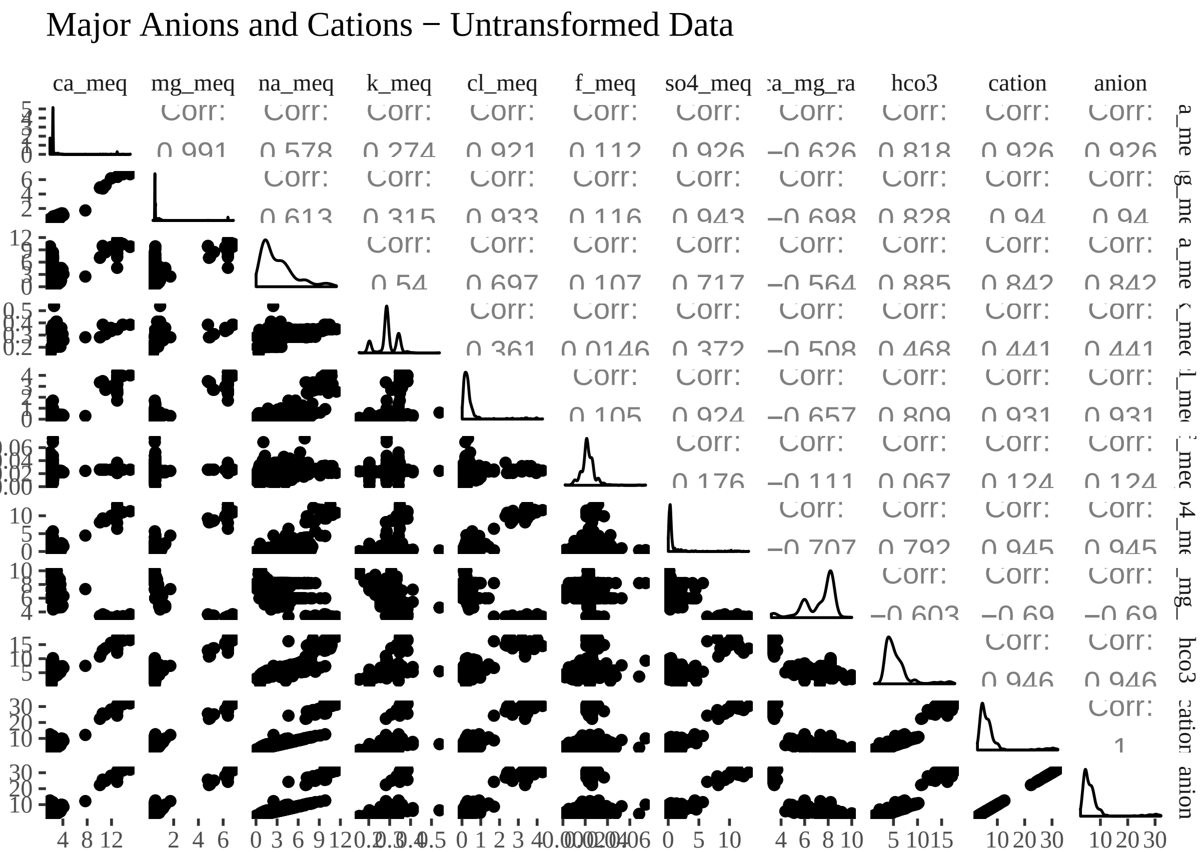

Differences in Anions Among Water Quality Groups
Bicarbonate
Chloride
Fluoride
Sulfate
val
15
10
5
0
Badl
Badl_mix
CheyRiv
SandH
Tbl_mix
Tbl_Steady
Final_Group

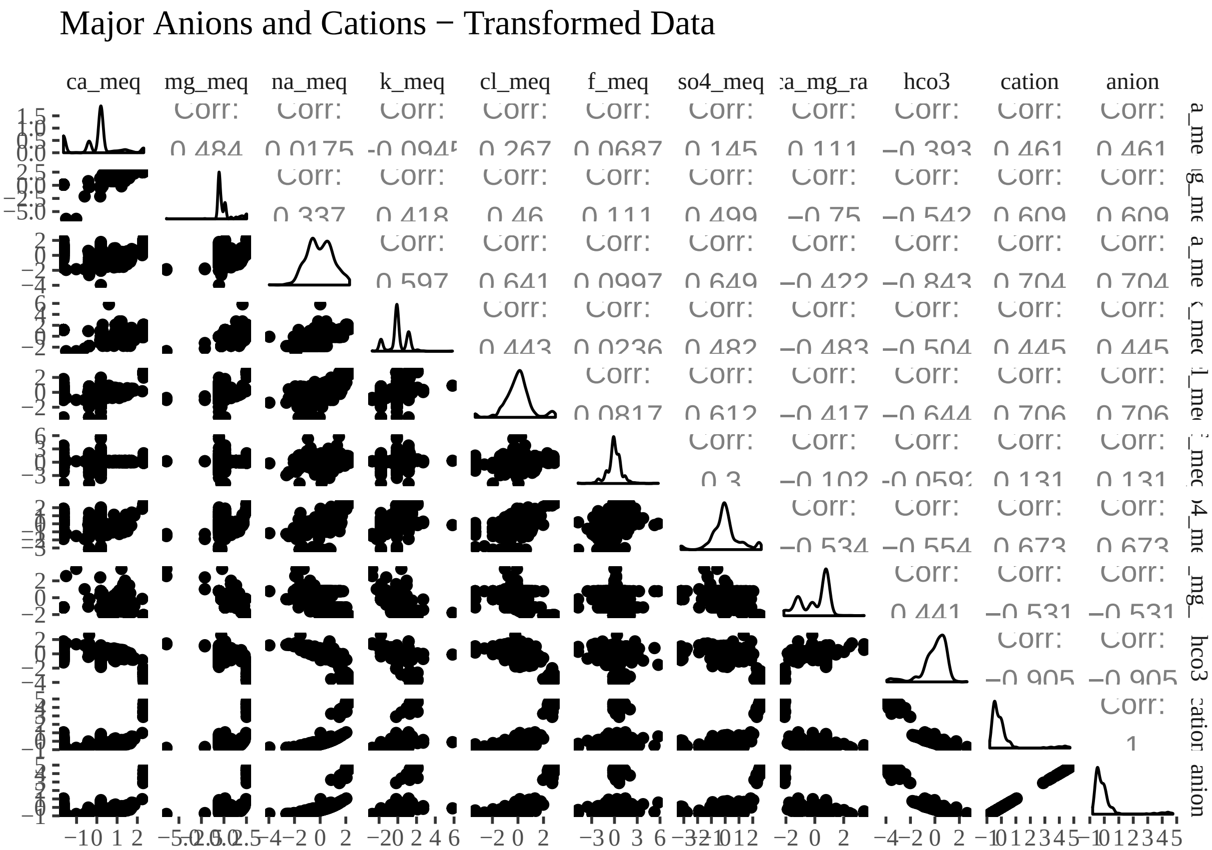


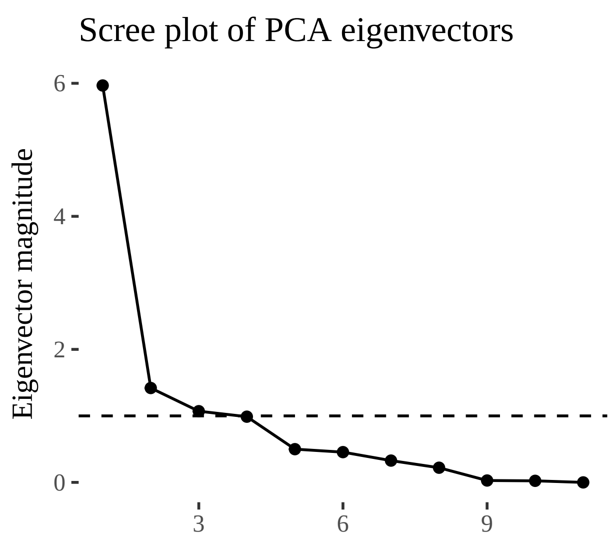
Conclusions and Next Steps

We identified differences among major non-nutrient cations and anions in Pine Ridge Reservation water samples and classified the samples into water quality groups. Time limitations prevented MANOVA or MRPP analysis of the groups to determine if the differences among group are significant. The next step in the analysis are: 1) to use the identified groups as priors for a quadratic discriminant analysis to identify misclassified observations following the exploratory principal component analysis, and 2) to estimate nitrate and phosphorus concentrations from the groups.

Appendices


Major Anions and Cations − Untransformed Data
ca_meq
mg_meq
na_meq
k_meq
cl_meq
f_meq
so4_meq
hco3
cation
anion
Corr:
0.991
0.578
0.274
0.921
0.112
0.926
−0.626
0.818
0.926
0.926
0.613
0.315
0.933
0.116
0.943
−0.698
0.828
0.94
0.54
0.697
0.107
0.717
−0.564
0.885
0.842
0.361
0.0146
0.372
−0.508
0.468
0.441
0.105
0.924
−0.657
0.809
0.931
0.176
−0.111
0.067
0.124
−0.707
0.792
0.945
−0.603
−0.69
0.946
1





Variable	PC1	PC2	PC3
anion	−0.38	−0.17	−0.08
ca_meq	−0.15	−0.75	0.03
ca_mg_rat	0.27	−0.32	−0.26
cation	−0.38	−0.17	−0.08
cl_meq	−0.32	−0.01	−0.11
f_meq	−0.07	−0.02	0.82
hco3	0.37	0.10	0.23
k_meq	−0.25	0.42	−0.11
mg_meq	−0.30	−0.13	0.26
na_meq	−0.33	0.23	−0.22
so4_meq	−0.32	0.15	0.23