Examining the Hydrologic Properties of the Missouri River Basin

https://github.com/cwatson1013/Hydrologic_Data_Analysis_Final_Proj

Rachel Bash, Keqi He, Caroline Watson, and Haoyu Zhang

Abstract

The Missouri River provides critical water resources that drives the region's agriculture, industry, and ecosystems. This is a region that experiences surface water variability, characterized by damaging floods and severe droughts, greatly impacting the agricultural production of the area. It is reported that a serious flood disaster occurred in the lower Missouri River in the spring of 2019 and the Missouri River experienced severe drought in 2012-2013. This project highlights the changes in stream flow and water quality over time, and identifies key characteristics of the river. Twenty two sites across the lower Missouri River Basin were examined in order to get a fuller picture of the Missouri River and its tributaries over time. By analyzing the trend of the Missouri River discharge, we can predict future changes in the Missouri River flow to provide a reference for water resources management. In addition, we focus on the stream flow and water quality of Missouri River during March to July, 2019 to see how discharge influence the water quality and what can be done to keep the water in the Missouri River in good quality.

<Information in these brackets are used for annotating the RMarkdown file. They will not appear in the final version of the PDF document>

Contents

List of Tables

List of Figures

<Note: set up autoreferencing for figures and tables in your document>

1 Research Question and Rationale

<Paragraph detailing the rationale for your analysis. What is the significant application and/or interest in this topic? Connect to environmental topic(s)/challenge(s).>

<Paragraph detailing your research question(s) and goals. What do you want to find out? Include a sentence (or a few) on the dataset you are using to answer this question - just enough to give your reader an idea of where you are going with the analysis.>

The Missouri River is the largest river in North America (2,540 miles) and has the second largest watershed (529,000 mi^2/339 acres, U.S.-Canada). Its watershed covers portions of ten states, which account for approximately one-sixth of the continental United States, as well as a small part of Canada. The headwater is located in the Bitterroot Mountains River of northwestern Wyoming and southwestern Montana. The watershed is home to around 12 million people in 1990, and has been inhabited by indigenous people for millennia. Demands for managing the river for the benefit of human livelihood has resulted in drastic modification in the river and the floodplains. Numerous reservoirs and dams have been constructed, of which six major dams were built on the mainstream, following the Pick-Sloan Plan in 1944. Now, the river is used intensively in multiple ways, including municipal, agricultural, hydropower, recreation, flood control etc.
Within the 328 million acres of the basin's total area in the United States, 95% is related to agricultural uses, while the rest dedicated for recreation, fish and wildlife, and urban. More than half of the total is pasture and range grassland primarily for grazing, and cropland consists of almost 104 million acres, which is 32% of the whole basin. Irrigated land comprises 7.4 million acres, and 6.9 million acres are intensively cropped. Water bodies, on the other hand, cover 3.9 million acres. In spite of the low proportion of water areas (1.2%), they are the pivotal foundation for agricultural or other usages, and thus critical to the whole region's economy.
Along with the agricultural, urban, and industrial development in the region is nutrient loading and enrichment in water bodies, especially for nitrogen (N) and phosphorus (P). Unlike other regions, agricultural input through fertilizer is the predominant anthropogenic source for nutrient in water bodies in the whole basin. Regardless of the major anthropogenic source, nutrient enrichment is considered nationally as one of the leading factors for water quality impairment. According to USEPA 303(d) lists, more than 160 stream reaches, lakes, or reservoirs were reported by USEPA to be nutrient-related impaired in 2006.
In addition to change in nutrient concentration, discharge appears to be highly variable in the basin, and both severe drought and flooding events occurred in the basin in the past. For example, in the spring and summer of 2011, an unprecedented flooding event caused over $2 billion damage FEMA disaster declaration was made in all states along the Missouri River. Subsequently, in 2012, a drought even struck the Central Great Plain, including the basin, and inflicted at least $12 billion of loss before July, 2012. Recently, another flooding event occurred in the spring of 2019. Given all the background information above, we would like to

know the current state of Missouri River and its tributaries, with a focus on the changing pattern in discharge and nutrient. We are interested in how the dramatic change in discharge (i.e. water quantity) could potentially interact with nutrient enrichment (i.e. water quality). Also, we examined a few specific flooding events, during which changes in both water quality and quantity were well recorded, so that we could make concrete inference on the interplay between quantity and quality. Finally, based on the pattern in the past and the best model we could fit, we attempted to predict the likely future conditions and trends in the Missouri River Basin. To achieve these goals, we retrieved data on water quantity, water quality (N, P concentrations), pH, coliform concentrations from USGS National Water Information System, using the package `dataRetrieval`.

2 Dataset Information

<Information on how the dataset for this analysis were collected, the data contained in the dataset, and any important pieces of information that are relevant to your analyses. This section should contain much of same information as the README file for the dataset but formatted in a way that is more narrative.>

The data we are analyzing comes from the United States Geological Survey (USGS) database called the National Water Information System interface, or NWIS. We pulled data from the interface using the R package `dataRetrieval`. Because we are interested in the lower Missouri River basin, we pulled sites from each HUC4 subbasin from 1020 to 1030 (see Figure below). We chose these subbasins because they had a variety of tributaries that all flowed into the Missouri River, and we wanted a variety of river sizes and lengths. We filtered these subbasin queries to only show us sites that had discharge, nitrogen, and phosphorus data. Once we found the sites with all of this data, we chose 2 sites from each HUC sub basin as our 22 "best sites". Our best sites had the overall best time period range for all of our "must have" variables.

Only seven sites within our HUC subbasin boundary contained any high frequency discharge and nitrogen data. Therefore, we also looked at these 7 sites in order to do analyses and answer our research question about flooding.

After doing initial data wrangling and analysis on our 22 "best sites", we decided to pare it down further and only do subsequent analyses on **10** sites. While we initially wanted to look at many sites that were varied in size and location, we determined that it was too many to look at and draw relevant conclusions from.

We have three main datasets:

- The daily values dataset with our 22 "best sites"
- The water quality dataset with our 22 "best sites", with only six sites that had total coliform data.
- The high freqency dataset with 7 sites that contain both high freqency discharge and high frequency nitrogen data.

<Add a table that summarizes your data structure. This table can be made in markdown text or inserted as a `kable` function in an R chunk. If the latter, do not include the code used to generate your table.>

<C will do data table for water quality and daily values, R will do for high freq>

Variable	Units	TypeOfVariable	Hypothesis
discharge	cfs or m3/s	both	all
time	UTC	independent	all
nitrogen	mg/L	dependent	2 and 3
pH	1	dependent	2 and 3
total coliform	cfu/100mL	dependent	2
O2 Concentration	mg/L	dependent	2 and 3
Phosphorus	mg/L	dependent	2 and 3

3 Exploratory Data Analysis and Wrangling

<Include R chunks for 5+ lines of summary code (display code and output), 3+ exploratory graphs (display graphs only), and any wrangling you do to your dataset(s).>

<Include text sections to accompany these R chunks to explain the reasoning behind your workflow, and the rationale for your approach.>

3.0.0.1 Water Quality Data

There were 6 sites in the Missiour River Basin that had total coliform data in addition to total nitrogen, total phosphorus, pH, and discharge data. Out of the 6 sites, only *3* of the sites were chosen for water quality analysis because they had the most data for all of the parameters for water quality. The water quality data was wrangeled to include certain columns necessary for the analysis. The sites that were looked at in depth for water quality analysis are:

```
- 06810000
- 06856600
- 06934500
```

```r
#filtering water quality dataset to include only 3 sites
bestsites.wq.skinny <- bestsites.wq %>%
    select(Site = site_no,
           Date = Date,
           Parameter = parm_cd,
           Value = result_va,
           Discharge = X_00060_00003) %>%
    group_by(Date, Parameter, Site) %>%
    summarize(Value = mean(Value),
           Discharge = mean(Discharge)) %>%
    spread(key = Parameter, value = Value) %>%
    rename(pH = '00400', total.coliform = '31501',
        Discharge2 = '00060', total.nitrogen = '00600',
        total.phosphorus = '00665') %>%
    mutate(Year = year(Date)) %>%
    select(-Discharge2) %>%
    filter(Site == "06810000" | Site == "06856600" |
        Site == "06934500")
```

Graphs were made to look at total phosphorus, total nitrogen, pH, and total coliform over time. These figures were also faceted by site to see whether there were trends in specific sites.

As seen by Figure 1, total phosphorus values have a slight positive trend. From Figure 1, site 06934500 has the most total phosphorus data.

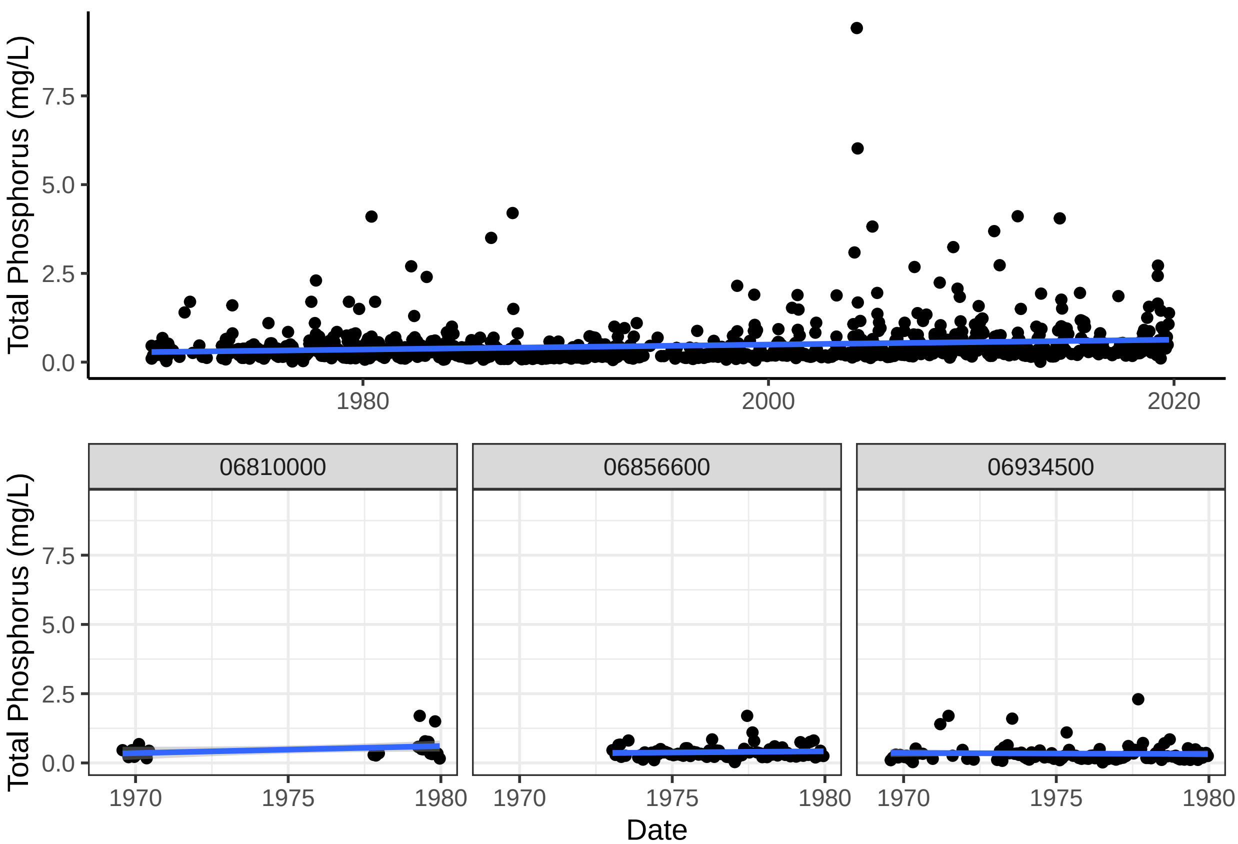

Figure 1: Total Phosphorus Over Time

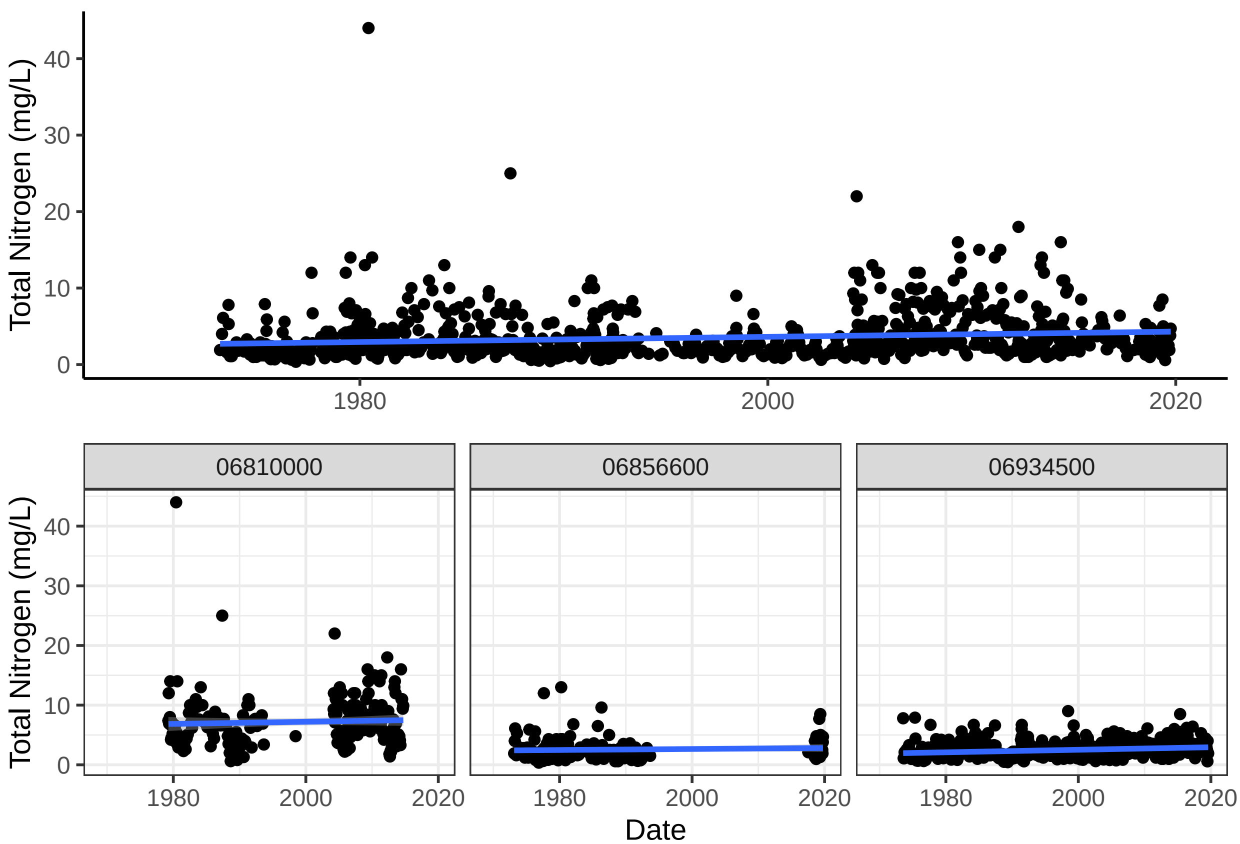

Figure 2: Total Nitrogen Over Time

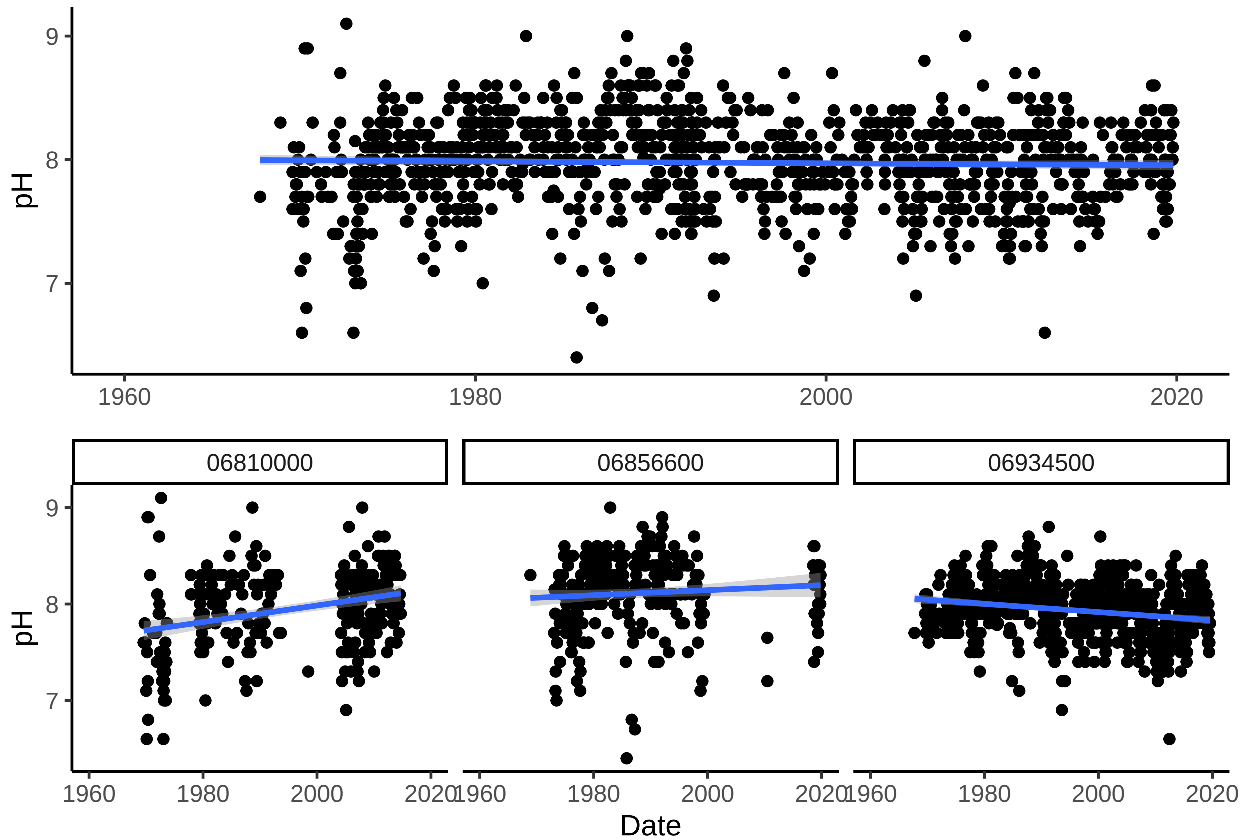

Figure 3: pH Over Time

From Figure 2, total nitrogen looks as though there is a slight positive trend from 1980 to 2019. Again, site 06934500 has the most total nitrogen data, as evidenced in Figure 2.

As seen in Figure 3, pH values range from below pH of 7 to above pH of 9 from 1970 - 2019. From Figure 3, site 06810000 has a positive increase in pH over time, whereas site 06934500 has a decreasing trend in pH over time.

Figure 4 shows the total coliform measurements in the 3 chosen sites over time. From Figure 4, it is evident that there is not much data on total coliform in the Missouri River Basin and monitoring for total coliform occured at these 3 sites from late 1960s through 1975.

Figure 5 was created to determine whether high amounts of total coliform coincided with an increase in discharge. Because there were limited total coliform measurements taken, as evidenced in Figure 4, there is not a great conclusion from this data. However, Figure 5 shows a spike in discharge events between 1972 - 1974, which also happens to be a time when total coliform was sampled. Figure 5 also shows increases in total coliform between 1972 - 1975.

```
#high frequency data wrangling
highfreqsite2019 <- highfreqsiteinfo %>%
  filter(end_date > "2019-03-31"); head(highfreqsite2019)
```

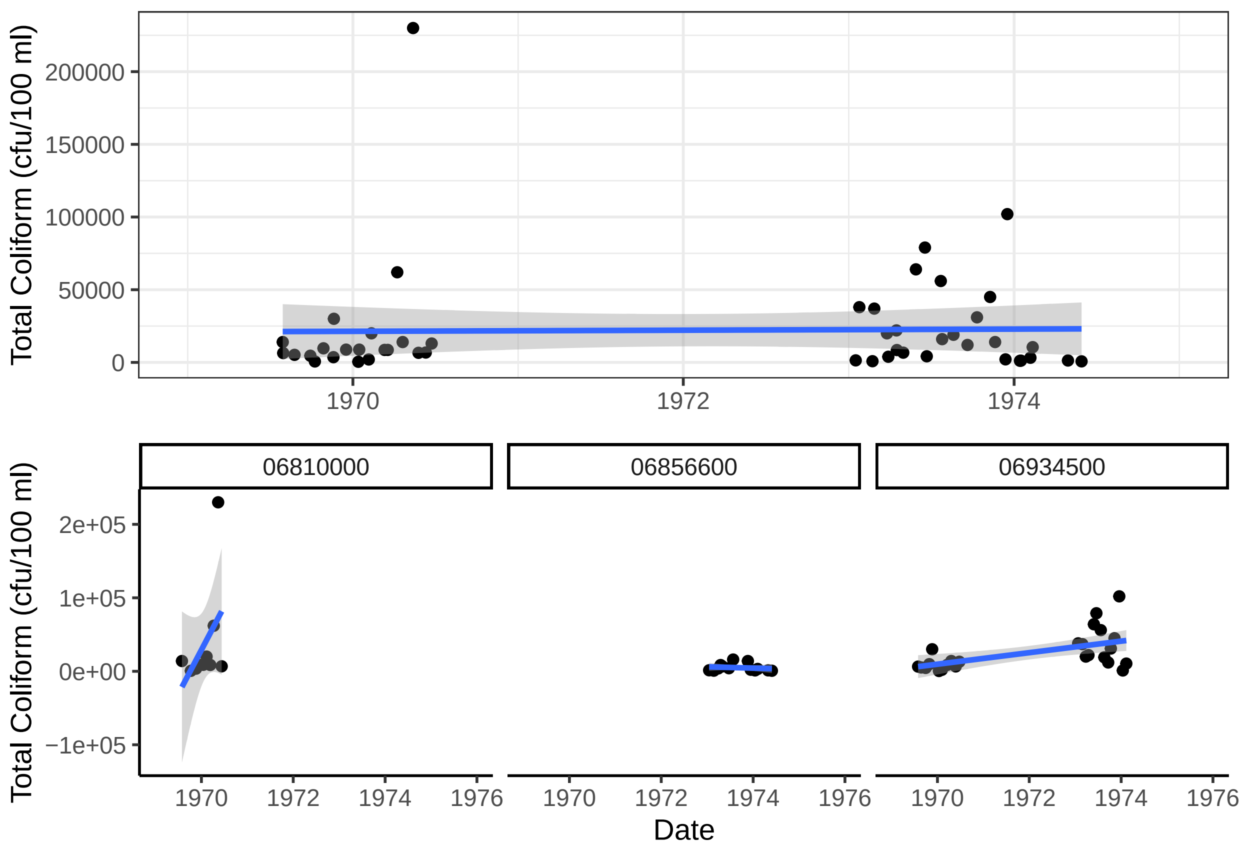

Figure 4: Total Coliform Over Time

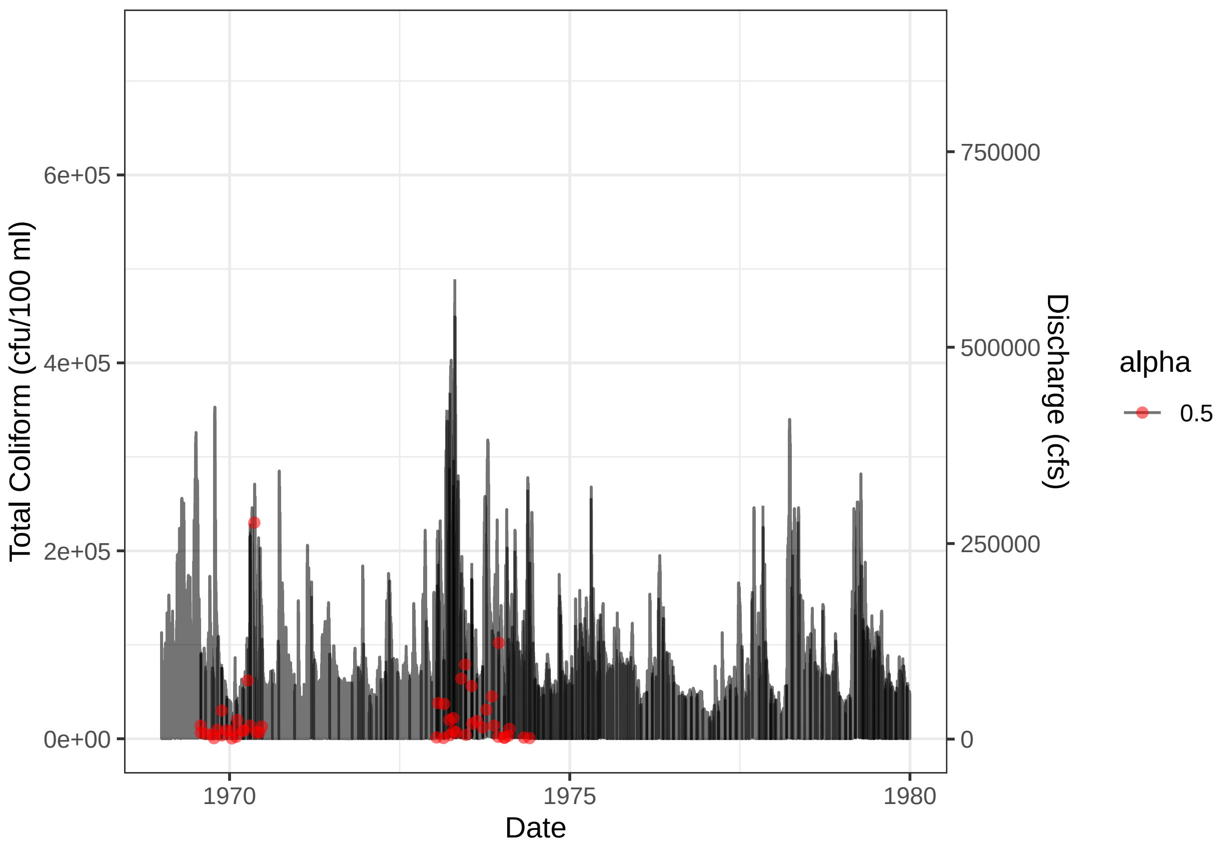

Figure 5: Total Coliform over time and Discharge over time

```
## Warning in Ops.factor(end_date, "2019-03-31"): '>' not meaningful for
## factors

##  [1] X                  agency_cd          site_no
##  [4] station_nm         site_tp_cd         dec_lat_va
##  [7] dec_long_va        coord_acy_cd       dec_coord_datum_cd
## [10] alt_va             alt_acy_va         alt_datum_cd
## [13] huc_cd             data_type_cd       parm_cd
## [16] stat_cd            ts_id              loc_web_ds
## [19] medium_grp_cd      parm_grp_cd        srs_id
## [22] access_cd          begin_date         end_date
## [25] count_nu
## <0 rows> (or 0-length row.names)
```

```
highfreqsites.DN <- readNWISuv(site = c("06808500", "06817000", "06892350", "06934500"),
                               parameterCd = c("00060", "99133"),
                               # Discharge in cfs & Nitrate in mg/l NO3-N
                               startDate = "2019-01-01",
                               endDate = "2019-11-01") %>%
                               renameNWISColumns() %>%
                               rename(Nitrate_mgl = 6)

#individual sites
Hermann <- highfreqsites.DN %>%
           filter(site_no=="06934500")
Desoto <- highfreqsites.DN %>%
          filter(site_no=="06892350")
Clarinda <- highfreqsites.DN %>%
            filter(site_no=="06817000")
Randolph <- highfreqsites.DN %>%
            filter(site_no=="06808500")
```

3.0.1 High Frequency Nitrogen and Discharge

There were 7 sites in our region of interest that had high freq N data, and only 4 sites had high freq N data during the floods of 2019. The sites looked at in depth are:

```
- West Nishnabotna River in Randolph, IA
- Nodaway River at Clarinda, IA
- Kansas River in Desoto, KS
- Missouri River at Hermann, MO
```

The Missouri River is the biggest river, with an average of 214693 cfs discharge rate during the year 2019, and the Nodaway River is the smallest river, with an average of 1185 cfs discharge rate for 2019.

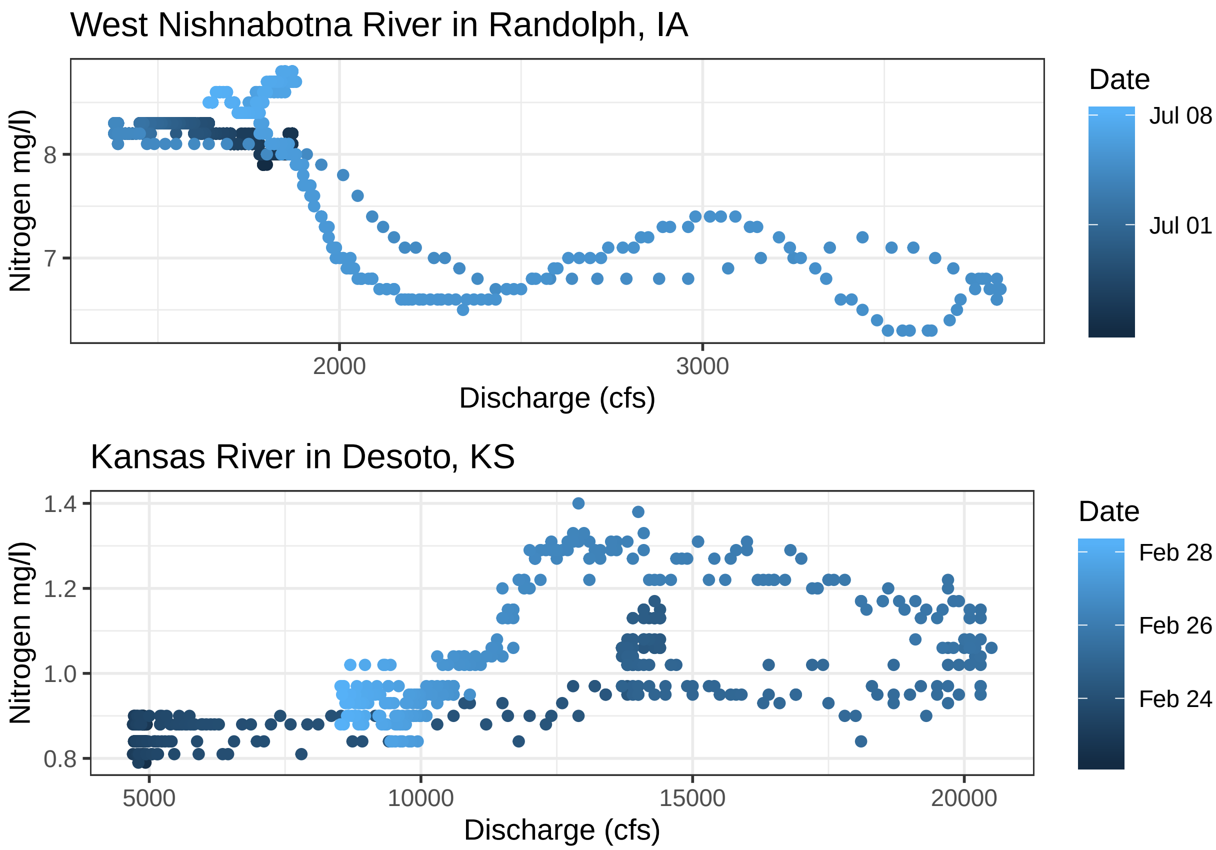

Figure 6: Hysteresis plots

In March of 2019, a bomb cyclone hit the midwest. Our initial research question, what effect did the March 2019 storm have on water quality, attempted to look into the behavior of nitrogen in the discharge of the rivers. Unfortunately, instantaneous Nitrogen values stopped recording during the peak of the storm events in March, so it was hard to create hysteresis plots that exhibited the type of storm and its effects on nitrogen concentration.

Even though Nitrogen concentrations were not recorded in March, they were recorded in other times of the year. 2019 was a wet year and many large storm events occurred.

```
## Warning: Removed 9 rows containing missing values (geom_point).
```

The Figure 6 shows Hysteresis plots for two storm events in the Missouri River Basin. The storm event on the West Nishnabotna River exhibits an oddly-shaped plot that has a negative slope, indicating it is a diluting storm. The Kansas River experienced a storm in late February that has a counter-clockwise motion and a positive slope, indicating a flushing storm. These two plots illustrate that two rivers near each other can have very different behaviors.

4 Analysis

<Include R chunks for 3+ statistical tests (display code and output) and 3+ final visualization graphs (display graphs only).>

<Include text sections to accompany these R chunks to explain the reasoning behind your workflow, rationale for your approach, and the justification of meeting or failing to meet assumptions of tests.>

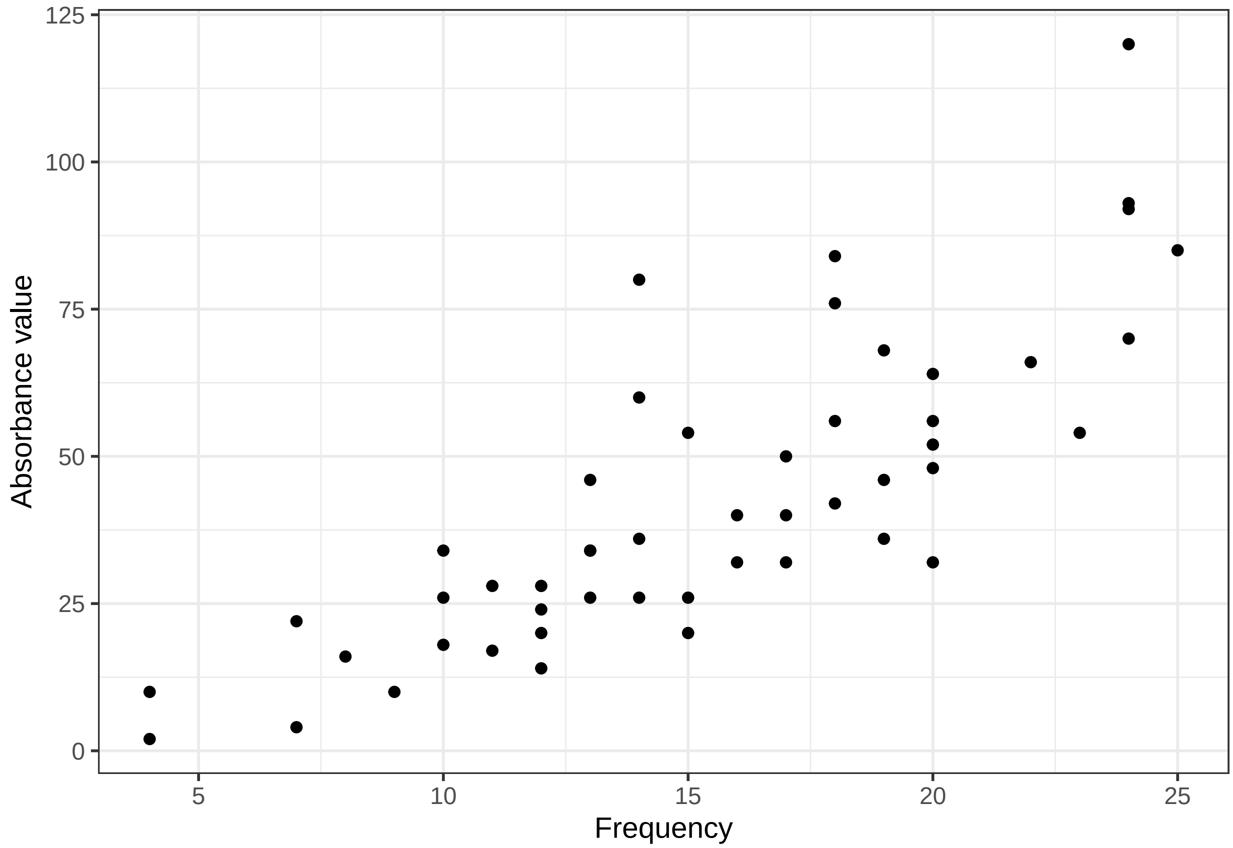

Figure 7: Absorbance frequency

5 Summary and Conclusions

<Summarize your major findings from your analyses. What conclusions do you draw from your findings? Make sure to apply this to a broader application for the research question you have answered.>

5.1 Example for autoreferencing

As seen by Figure 7, Absorbance values are not normally distributed. This is expected, as we are dealing with ecological data.