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>

Table 1: Summary of Daily Value Data at 22 sites in the Missouri River Basin

X	agency_cd	site_no	Date	Discharge	Approval.Code	parm_cd	sample_tm	sample_end_dt	sample_end_tm	sample_start_time_datum_cd	tm_datum_rlbty_cd
Min. : 1	USGS:658650	Min. :6609500	2009-10-20: 36	Min. : 0	A :618881	Min. : 60.00	11:00 : 1108	Mode:logical	14:15 : 8	CDT : 8481	K : 5636
1st Qu.:164663	NA	1st Qu.:6800000	2009-11-18: 36	1st Qu.: 154	A e : 21858	1st Qu.: 60.00	10:30 : 883	NA's:658650	11:15 : 6	CST : 4747	T : 7592
Median :329326	NA	Median :6856600	1982-03-09: 35	Median : 756	P : 4304	Median : 60.00	10:00 : 831	NA	04:15 : 4	NA's:645422	NA's:645422
Mean :329326	NA	Mean :6828931	1980-08-12: 34	Mean : 10226	P e : 246	Mean : 71.54	12:00 : 750	NA	05:00 : 4	NA	NA
3rd Qu.:493988	NA	3rd Qu.:6892350	2017-07-25: 34	3rd Qu.: 4670	P Ice : 79	3rd Qu.: 60.00	11:30 : 747	NA	06:15 : 4	NA	NA
Max. :658650	NA	Max. :6934500	1979-06-13: 33	Max. :739000	(Other): 54	Max. :665.00	(Other): 8903	NA	(Other): 98	NA	NA
NA	NA	NA	(Other) :658442	NA's :13360	NA's : 13228	NA	NA's :645428	NA	NA's :658526	NA	NA

X	coll_ent_cd	medium_cd	project_cd	aqfr_cd	tu_id	body_part_id	hyd_cond_cd	samp_type_cd	hyd_event_cd	sample_lab_cm_tx	remark_cd
Min. : 1	USGS-WRD: 9272	WS : 13164	NAWQA-SWS: 879	Mode:logical	Mode:logical	Mode:logical	9 : 5662	9 : 11792	9 : 11673	tech samples : 90	< : 470
1st Qu.:164663	USGS-NEL: 1051	WSQ : 64	462900300: 689	NA's:658650	NA's:658650	NA's:658650	A : 3373	A : 813	Z : 863	tech sample : 75	E : 71
Median :329326	USGSKSWC: 596	NA's:645422	463106100: 649	NA	NA	NA	5 : 1397	7 : 515	J : 292	tech sample;no sampling method given : 25	NA's:658109
Mean :329326	USGS : 517	NA	861100399: 542	NA	NA	NA	4 : 1235	H : 94	7 : 151	tech samples;cross section from churn: 22	NA
3rd Qu.:493988	USGSMOLS: 186	NA	463100300: 498	NA	NA	NA	8 : 692	5 : 8	B : 149	BOTTLES OK : 18	NA
Max. :658650	(Other) : 188	NA	(Other) : 6154	NA	NA	NA	(Other): 869	(Other): 6	(Other): 100	(Other) : 3975	NA
NA	NA's :646840	NA	NA's :649239	NA	NA	NA	NA's :645422	NA's :645422	NA's :645422	NA's :654445	NA

X	result_va	val_qual_tx	meth_cd	dqi_cd	rpt_lev_va	rpt_lev_cd	lab_std_va	prep_set_no	prep_dt	anl_set_no	anl_dt	result_lab_cm_tx	anl_ent_cd	startDateTime	parameter
Min. : 1	Min. : 0.0	d : 426	ALGOR : 4992	A : 6906	Min. :0.0	DLDQC : 614	Mode:logical	Mode:logical	Mode:logical	KJNT254A: 12	Min. :20010918	The parameter 00665 was swapped from labcode 2333 to labcode 2759 : 160	USGSNWQL: 2213	2018-05-08 10:30:00: 6	Discharge :645422
1st Qu.:164663	1st Qu.: 0.3	oc : 266	KJ009 : 2579	R : 5574	1st Qu.:0.0	LRL : 1226	NA's:658650	NA's:658650	NA's:658650	KJNT135A: 11	1st Qu.:20070427	The parameter 00665 was swapped from labcode 2333 to labcode 1984 because the result from labcode 2333 exceeded the calibration range.: 79	NA's :656437	1975-01-15 11:20:00: 4	Total Nitrogen : 6207
Median :329326	Median : 0.7	n : 56	CL084 : 701	S : 748	Median :0.0	LT-MDL: 1059	NA	NA	NA	KJNT200A: 10	Median :20120366	The parameter 00665 was swapped from labcode 2333 to labcode 2759 because the result from labcode 2333 exceeded the calibration range.: 77	NA	1975-02-12 10:30:00: 4	Total Phosphorus: 7021
Mean :329326	Mean : 1.7	doc : 47	CL021 : 609	NA's:645422	Mean :0.0	MRL : 21	NA	NA	NA	KJNT021A: 9	Mean :20111394	Report level code updated Oct., Nov. 2015. Reference: NWQL TM 2015.02 (RLC: LT-MDL => DLDQC) : 23	NA	1975-03-11 10:50:00: 4	NA
3rd Qu.:493988	3rd Qu.: 2.2	@d : 11	AKP01 : 411	NA	3rd Qu.:0.0	NA's :655730	NA	NA	NA	KJNT023A: 9	3rd Qu.:20151101	The holding time for the processing of this sample has been exceeded : 12	NA	1975-07-22 10:00:00: 4	NA
Max. :658650	Max. :100.0	(Other): 62	(Other): 392	NA	Max. :0.8	NA	NA	NA	NA	(Other) : 2869	Max. :20191023	(Other) : 28	NA	(Other) : 13200	NA
NA	NA's :645422	NA's :657782	NA's :648966	NA	NA's :655730	NA	NA	NA	NA	NA's :655730	NA's :655730	NA's :658271	NA	NA's :645428	NA

Table 2: Summary of Water Quality Data in the Missouri River Basin

X	agency_cd	site_no	Date	X_00060_00003	X_00060_00003_cd	parm_cd	sample_tm	sample_end_dt	sample_end_tm	sample_start_time_datum_cd	tm_datum_rlbty_cd	coll_ent_cd	medium_cd	project_cd	aqfr_cd	tu_id	body_part_id	hyd_cond_cd	samp_type_cd	hyd_event_cd	sample_lab_cm_tx	remark_cd	result_va	val_qual_tx	meth_cd	dqi_cd	rpt_lev_va	rpt_lev_cd	lab_std_va	prep_set_no	prep_dt	anl_set_no	anl_dt	result_lab_cm_tx	anl_ent_cd	startDateTime
Min. : 1	USGS:668800	Min. :6609500	1981-09-01: 56	Min. : 0	A :618881	Min. : 60.00	11:00 : 1819	Mode:logical	14:15 : 12	CDT : 14622	K : 8908	USGS-WRD: 15396	BI : 11	NAWQA-SWS: 1411	Mode:logical	Min. :0	Min. :59.0	9 : 9726	9 : 20163	9 : 19933	tech samples : 149	< : 472	Min. : 0.0	d : 426	ALGOR : 4994	A : 12661	Min. :0.0	DLDQC : 616	Mode:logical	Mode:logical	Mode:logical	KJNT254A: 12	Min. :20010918	The parameter 00665 was swapped from labcode 2333 to labcode 2759 : 160	USGS-WRD: 2206	2018-05-08 10:30:00: 9
1st Qu.:167201	NA	1st Qu.:6800000	1981-09-02: 50	1st Qu.: 154	A e : 21858	1st Qu.: 60.00	10:30 : 1402	NA's:668800	11:15 : 10	CST : 8672	T : 14386	USGS-NEL: 1690	BP : 41	462900300: 1089	NA's:668800	1st Qu.:0	1st Qu.:59.0	A : 6963	A : 2127	Z : 2203	tech sample : 127	> : 1	1st Qu.: 0.6	oc : 266	KJ009 : 2581	R : 9500	1st Qu.:0.0	LRL : 1226	NA's:668800	NA's:668800	NA's:668800	KJNT135A: 11	1st Qu.:20070428	The parameter 00665 was swapped from labcode 2333 to labcode 1984 because the result from labcode 2333 exceeded the calibration range.: 79	USGSKSWC: 392	1979-06-13 11:00:00: 7
Median :334401	NA	Median :6856600	2009-10-20: 43	Median : 756	P : 4388	Median : 60.00	10:00 : 1390	NA	05:00 : 7	NA's:645506	NA's:645506	USGSKSWC: 988	ON : 26	463106100: 991	NA	Median :0	Median :94.0	5 : 2201	7 : 831	J : 530	tech sample;no sampling method given : 50	A : 21	Median : 3.8	n : 56	PROBE : 2013	S : 1133	Median :0.0	LT-MDL: 1059	NA	NA	NA	KJNT200A: 10	Median :20120404	The parameter 00665 was swapped from labcode 2333 to labcode 2759 because the result from labcode 2333 exceeded the calibration range.: 77	USGSNWQL: 2215	1975-01-15 11:20:00: 6
Mean :334401	NA	Mean :6829150	2009-11-18: 43	Mean : 10227	P e : 246	Mean : 84.53	12:00 : 1277	NA	04:15 : 6	NA	NA	USGS : 805	SB : 8	463100300: 853	NA	Mean :0	Mean :81.2	4 : 1939	H : 149	7 : 248	Billed FY19. : 48	E : 113	Mean : 88.9	doc : 47	CL084 : 701	NA's:645506	Mean :0.0	MRL : 21	NA	NA	NA	KJNT021A: 9	Mean :20111448	Report level code updated Oct., Nov. 2015. Reference: NWQL TM 2015.02 (RLC: LT-MDL => DLDQC) : 23	NA's :663987	1975-02-12 10:30:00: 6
3rd Qu.:501600	NA	3rd Qu.:6892350	1980-08-12: 42	3rd Qu.: 4670	P Ice : 79	3rd Qu.: 60.00	11:30 : 1232	NA	06:15 : 6	NA	NA	USGSMOLS: 290	WS : 23094	861100399: 813	NA	3rd Qu.:0	3rd Qu.:94.0	8 : 1089	5 : 15	B : 224	tech samples;cross section from churn: 33	NA's:668193	3rd Qu.: 8.0	@d : 11	EL003 : 678	NA	3rd Qu.:0.0	NA's :665878	NA	NA	NA	KJNT023A: 9	3rd Qu.:20151102	The holding time for the processing of this sample has been exceeded : 12	NA	1975-03-11 10:50:00: 6
Max. :668800	NA	Max. :6934500	1982-03-09: 42	Max. :739000	(Other): 54	Max. :31501.00	(Other): 15541	NA	(Other): 150	NA	NA	(Other) : 282	WSQ : 114	(Other) : 10401	NA	Max. :0	Max. :94.0	(Other): 1376	(Other): 9	(Other): 156	(Other) : 6153	NA	Max. :400000.0	(Other): 67	(Other): 1566	NA	Max. :0.8	NA	NA	NA	NA	(Other) : 2871	Max. :20191023	(Other) : 29	NA	(Other) : 22627
NA	NA	NA	(Other) :668524	NA's :23426	NA's : 23294	NA	NA's :646139	NA	NA's :668609	NA	NA	NA's :649349	NA's:645506	NA's :653242	NA	NA's :668759	NA's :668759	NA's :645506	NA's :645506	NA's :645506	NA's :662240	NA	NA's :645506	NA's :667927	NA's :656267	NA	NA's :665878	NA	NA	NA	NA	NA's :665878	NA's :665878	NA's :668420	NA	NA's :646139

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.>

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

```
## 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)
```

```r
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.

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 1 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.

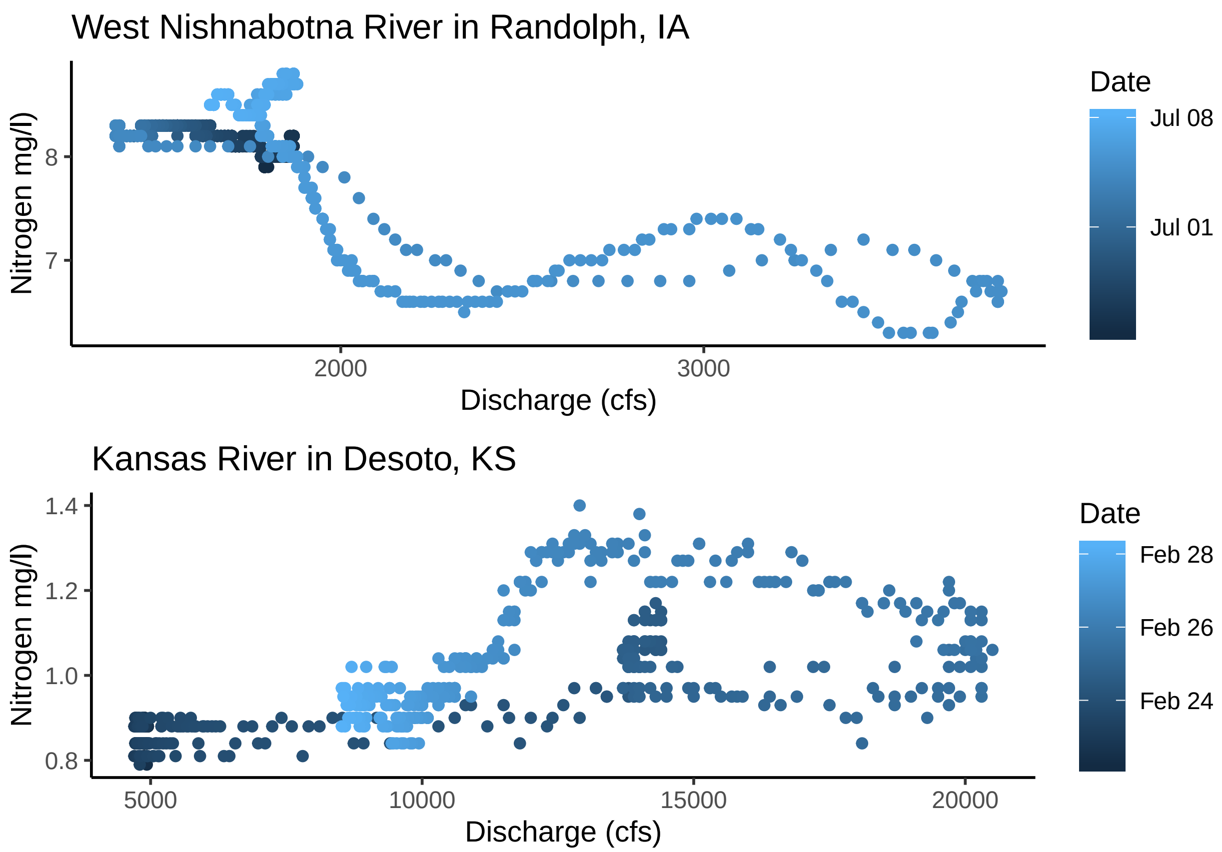

Figure 1: Hysteresis plots

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.>

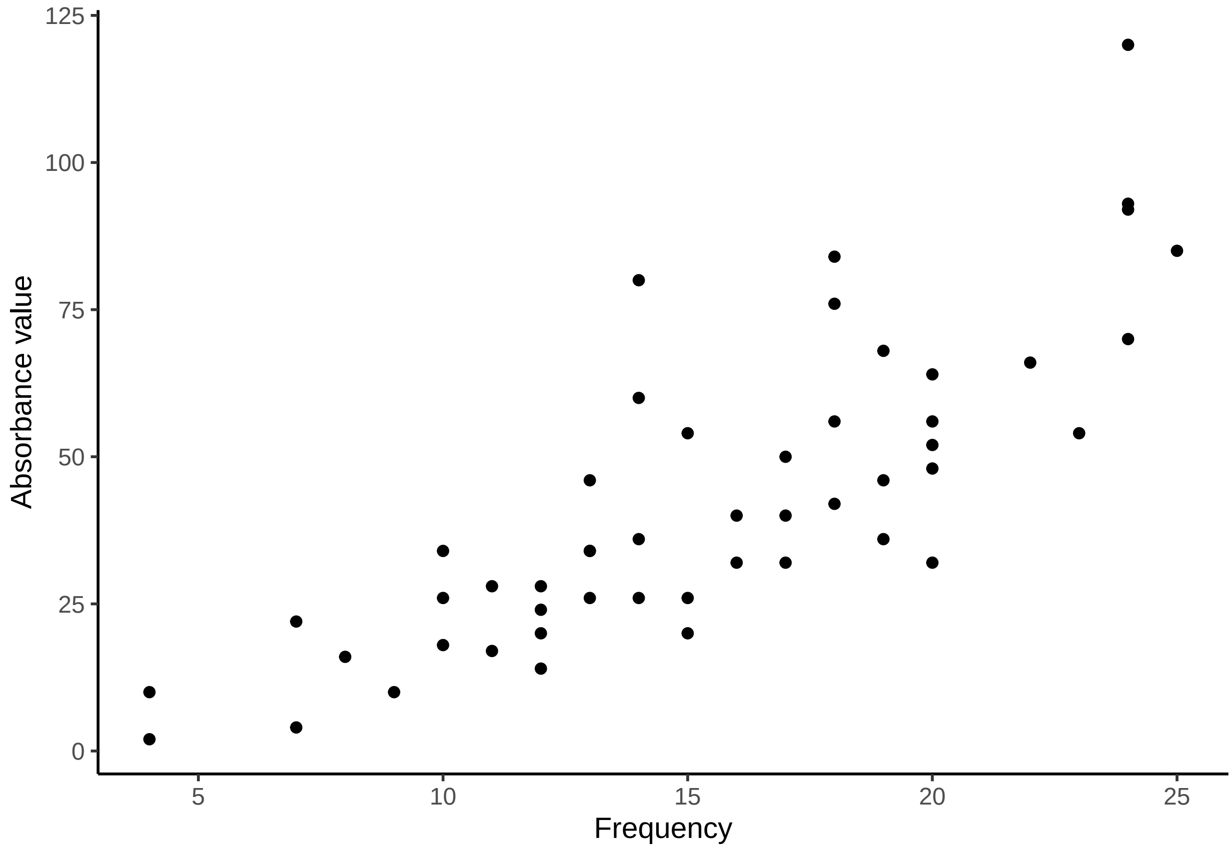

Figure 2: Absorbance frequency

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 2, Absorbance values are not normally distributed. This is expected, as we are dealing with ecological data.