Insert title of project here

https://github.com/rml41/EDA_2020_Project.git

Rachel Landman

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Ellerbe Creek runs into the Falls Lake Resovoir, through the city of Durham, North Carolina. Falls Lake serves as the source of drinking water for the City of Raleigh and does not meet North Carolina standards for *chlorophyll a*, which is found in algae (SOURCE: https://durhamnc.gov/716/Falls-Lake). Algal blooms generally come from excess nutrients such as phosphorus and nitrogen. Ellerbe Creek is one of the sources of excess nutrients and contaminents in Falls Lake. The Ellerbe Creek Watershed has the highest population density of Durham's watersheds, with an estimated 22% impervious surface (SOURCE: https://files.nc.gov/ncdeq/Water%20Quality/Planning/BPU/BPU/Neuse/Neuse%20Plans/2009%20Plan/Chapter%201.pdf). It is impacted by both point and nonpoint sources and was found to deliver the highest nutrient loads to Falls Lake (SOURCE: https://files.nc.gov/ncdeq/Water%20Quality/Planning/BPU/BPU/Neuse/Neuse%20Plans/2009%20Plan/Chapter%201.pdf). Ellerbe Creek and Falls Lake are both on the state's impaired water bodies list (303(d) list) (SOURCE https://durhamnc.gov/711/Ellerbe-Creek-Watershed, https://www.usgs.gov/centers/sa-water/science/groundwatersurface-water-interaction-near-ellerbe-creek-durham-nc?qt-science_center_objects=0#qt-science_center_objects). Ellerbe Creek was first listed on the 303(d) list in 1998 (SOURCE: https://files.nc.gov/ncdeq/Water%20Quality/Planning/BPU/BPU/Neuse/Neuse%20Plans/2009%20Plan/Chapter%201.pdf)

This analysis and report will aim to answer the following questions: 1. Are Nitrogen and Phosphorus levels in Ellerbe Creek above recommended levels? 2. Is there a relationship between flow and Nitrogen or Phosphorus levels? 3. How does location, upstream vs. downstream, impact nutrient levels?

2 Dataset Information

Nutrient data for this project were downloaded from the the Water Quality Portal, a coorperative service sponsered by the United States Geological Survey (USGS), the Environmental Protection Agency (EPA), and the National Water Quality Monitoring Council (NWQMC) on February 27, 2020. Discharge data were downloaded for two stream gages along Ellerbe Creek, HUC code 030202010403, from USGS using the data dataRetrieval package in R. The dataset analyzed contains 21 monitoring locations with measurments for nitrogen and phosphorus levels from 1982 to 2018 and daily discharge data from 2008 to 20202. Not all locations had data for each nutrient. Nitrogen and Phosphorus concentrations are recorded as mg/L of Nitrogen or Phosphorus in various compounds including, nitrate, nitrite, ammonia, ammonium, organic nitrogen, phosphate, and organic phosphorus. The USGS gage locations are Club Blvd (0208675010), upstream, and Gorman (02086849), downstream.

Variable	Units	Range	Mean	Median	Source
Nitrogen	mg/L N	0.37 - 33.00	7.18	2.82	NC DENR and USGS
Phosphorus	mg/L P	0.039 - 17.00	1.091	0.157	NC DENR and USGS
Discarge Club	ft^3/s	0.20 - 781.00	9.39	1.28	USGS
Discharge Gorman	ft^3/s	7.52 - 1750.00	48.84	20.50	USGS

3 Exploratory Analysis

3.1 Initial Exploration

Explored raw data from the water quality portal to determine potential variables for analysis and time period of data. Examined a summary of all the characteristics in the dataset to determine the count for each variable. Selected Nitrogen, mixed forms (NH3), (NH4), organic, (NO2) and (NO3) and Phosphorus as the two variables to analyze. Explored discharge data to determine date range of data.

Table 2. Sample of summary results from raw data

Variable	Count
Dissolved Oxygen	636
Nitrate	128
Nitrogen, mixed forms (NH3), (NH4), organic, (NO2) and (NO3)	209
Phosphorus	286
RBP Stream Width	14
Temperature, water	1146
Total Dissolved Solids	278

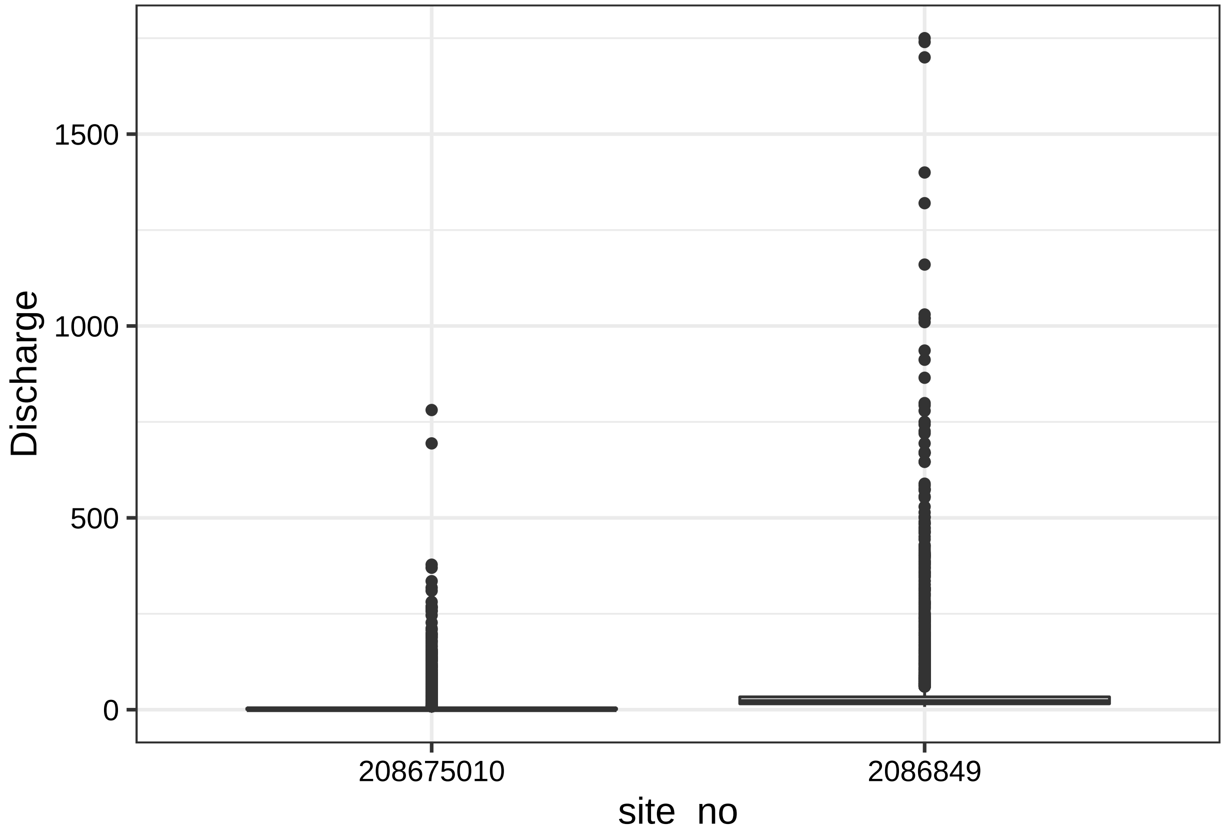
3.2 Discharge Data Wrangling

Flow data from the two USGS stream gages were combined in two data sets, one as a long format with all discharge in one column and one in wide format, with two seperate columns for discharge based on location.

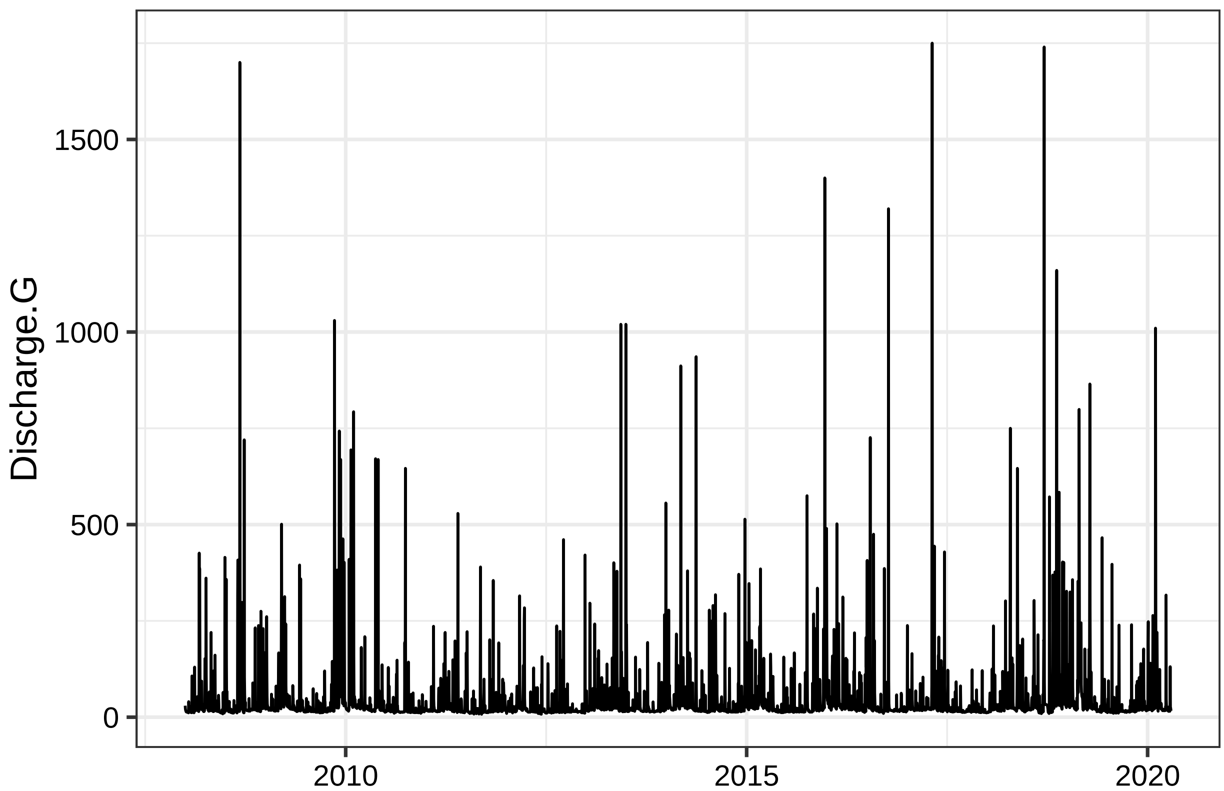
3.3 Discharge Data Exploration

A boxplot was made to visualize the range of discharge at each site

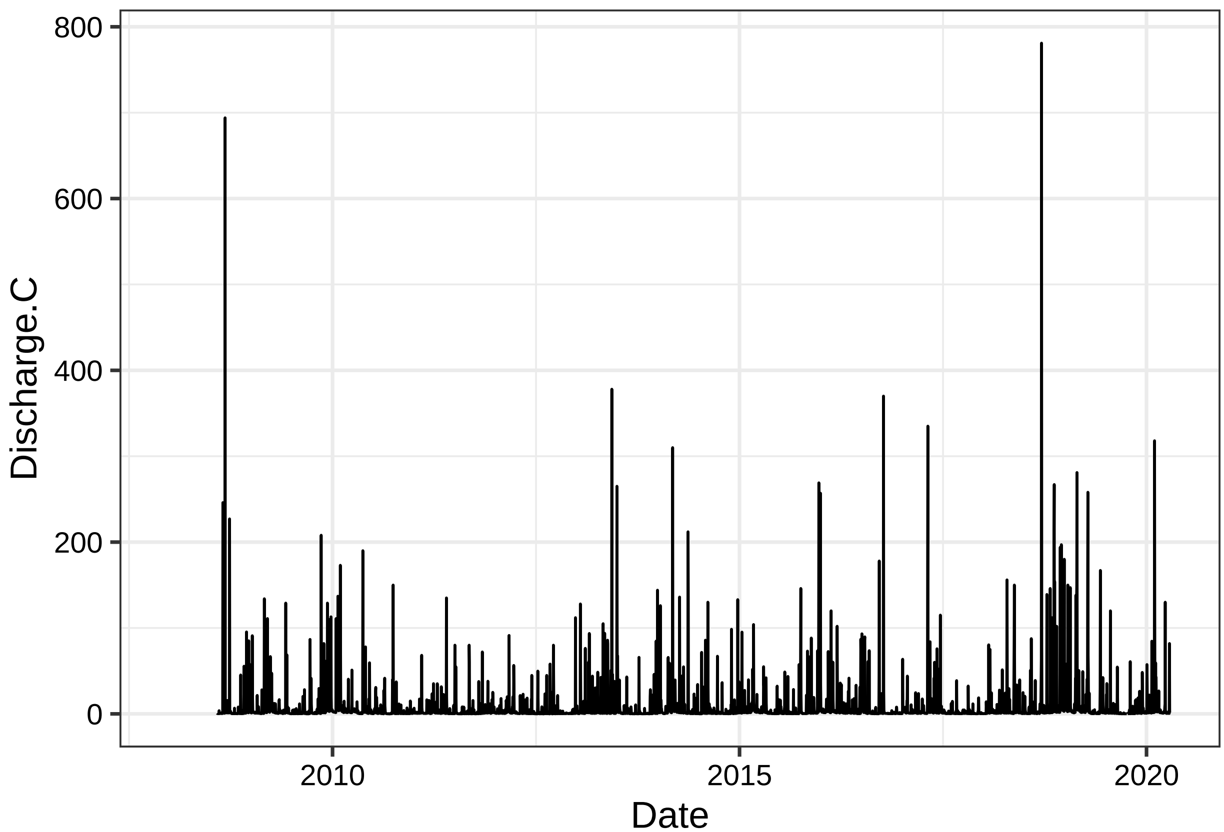
```
## Warning: Removed 9 rows containing non-finite values (stat_boxplot).
```



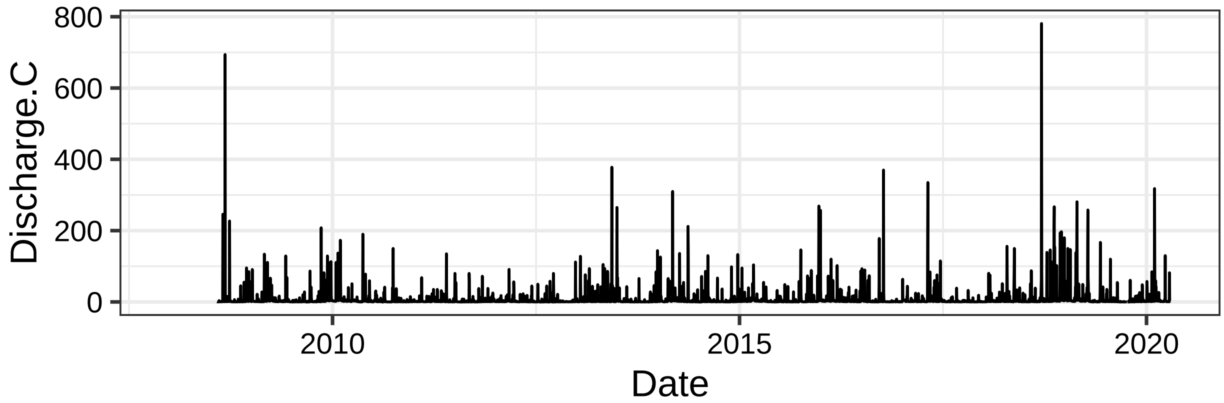
The discharge over time at each location does not show any obvious seasonal or annual trends.

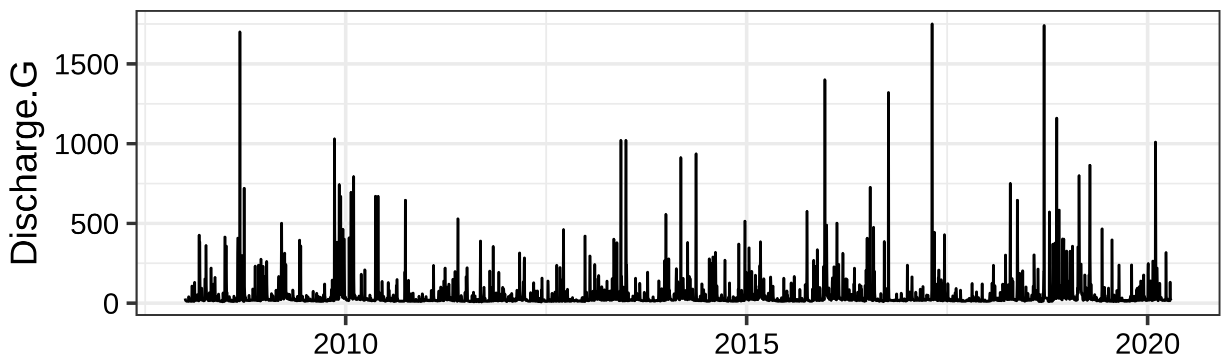



```
## Warning: Removed 213 rows containing missing values (geom_path).
```




```
## Warning: Removed 213 rows containing missing values (geom_path).
```



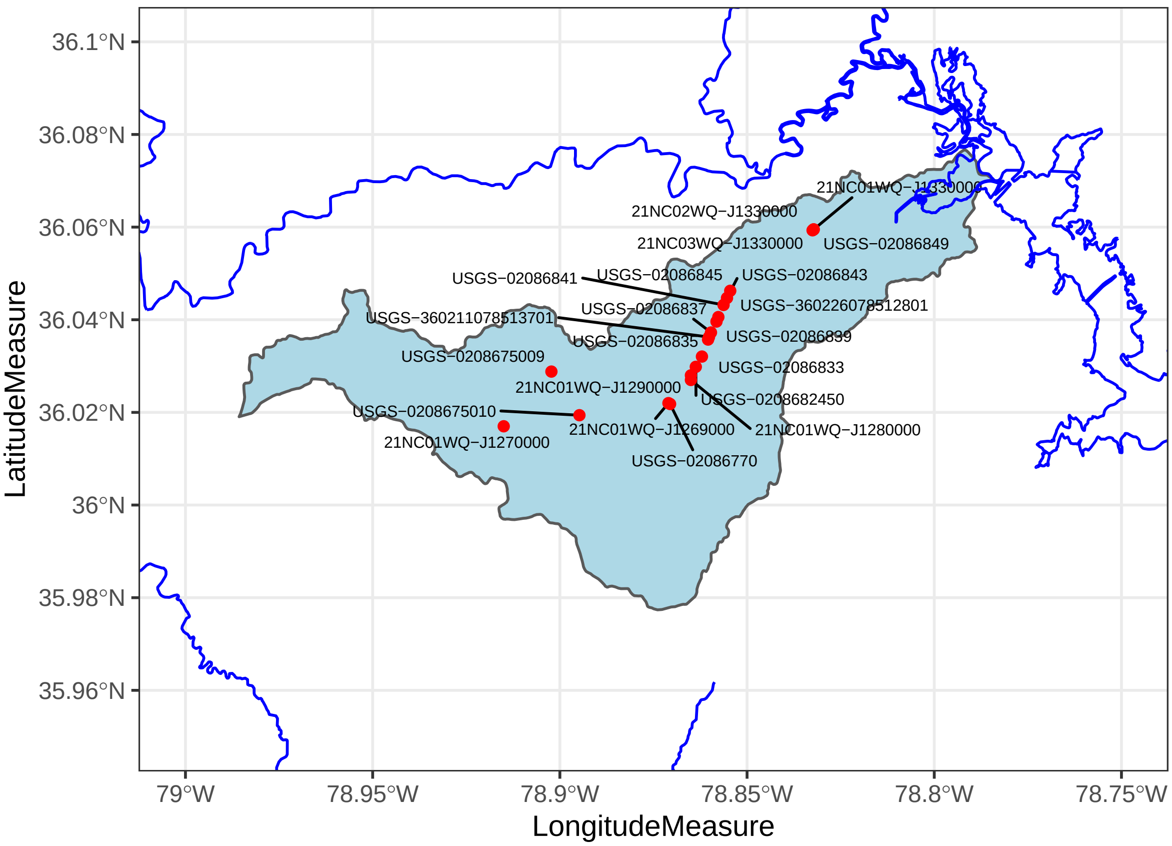


3.4 Nutrient Data Wrangling

The nutrient data set from the water quality portal was cleans to remove all irrelevant information and retain just characterstics of interest, Nitrogen and Phosphorus. Nitrogen and Phosphorus values for many samples were recorded as both mg/L of N and P, and of NO3 and PO4 respectively. Data were downloaded in long format and were converted to wide format in order to convert Nitrogen and Phosphorus values to mg/L of N or P. Relevant columns such as data, location, hydrologic event, variable name, measured value, and units were selected and processed data were saved as both long and wide format.

3.5 Nutrient Data Exploration

3.6 Geographic Exploration



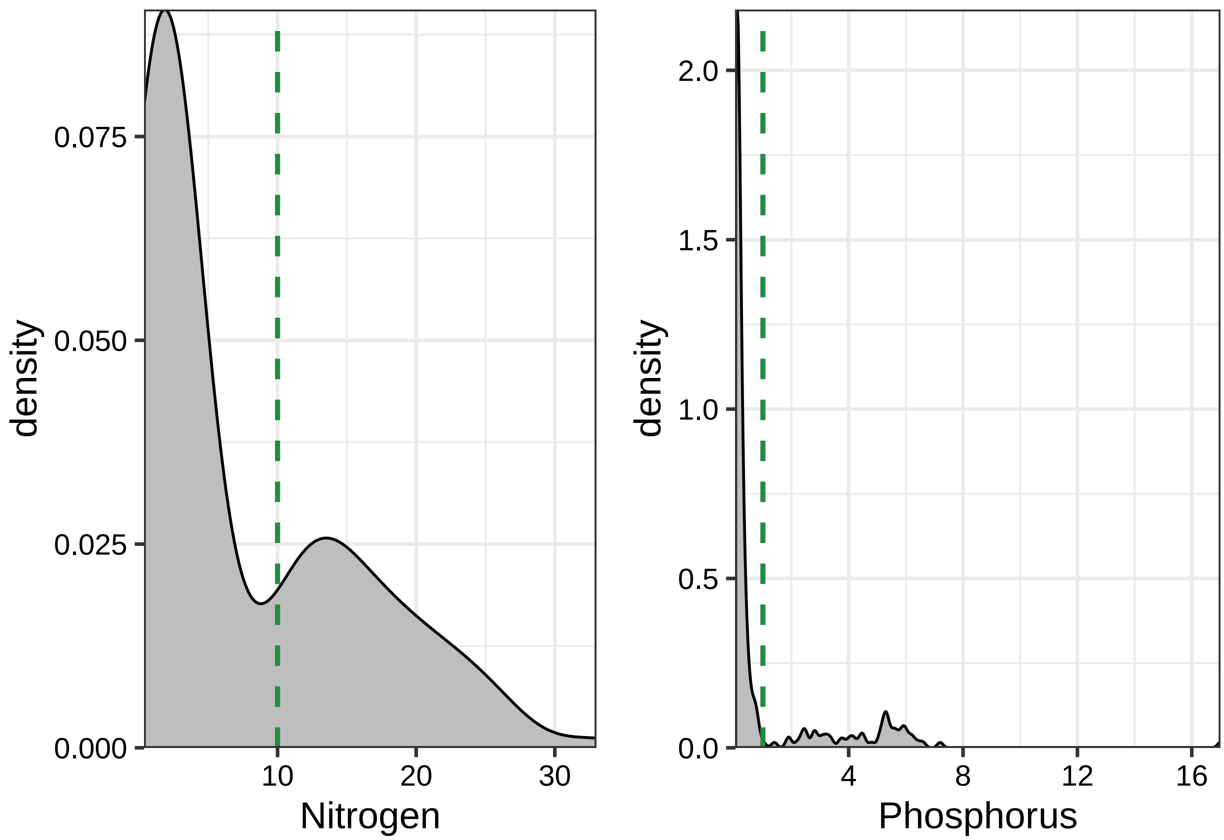
4 Analysis

4.1 Question 1: Are Nitrogen and Phosphorus levels in Ellerbe Creek above recommended levels?

Table 3: Total Nitrogen and Phosphorus Conentrations

mean.N	min.N	max.N	Standard.dev.N	mean.P	min.P	max.P	Standard.dev.P
7.183095	0.37	33	7.773407	1.091398	0.039	17	2.082384

```
##
##  Shapiro-Wilk normality test
##
## data:  (EC_Flow.Nutrients_Wide$Nitrogen)
## W = 0.79284, p-value = 8.616e-13

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  EC_Flow.Nutrients_Wide$Nitrogen
## V = 2773.5, p-value = 1.735e-05
## alternative hypothesis: true location is not equal to 10

##
##  Shapiro-Wilk normality test
##
## data:  (EC_Flow.Nutrients_Wide$Phosphorus)
## W = 0.54427, p-value < 2.2e-16

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  EC_Flow.Nutrients_Wide$Phosphorus
## V = 10319, p-value = 5.817e-05
## alternative hypothesis: true location is not equal to 1

## Warning: Removed 5428 rows containing non-finite values (stat_density).

## Warning: Removed 5325 rows containing non-finite values (stat_density).
```



4.2 Question 2: Is there a relationship between flow and Nitrogen or Phosphorus levels?

```
##
## Call:
## lm(formula = log(Nitrogen) ~ log(Discharge.C), data = EC_Flow.Nutrients_Wide)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09532 -0.42011 -0.04936  0.34030  1.27602
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.23863    0.11547   2.067    0.043 *
## log(Discharge.C)  0.05712    0.03698   1.545    0.127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5865 on 62 degrees of freedom
##   (5504 observations deleted due to missingness)
## Multiple R-squared:  0.03707,    Adjusted R-squared:  0.02154
```

```
## F-statistic: 2.387 on 1 and 62 DF,  p-value: 0.1275
```

```
##
## Call:
## lm(formula = log(Nitrogen) ~ log(Discharge.G), data = EC_Flow.Nutrients_Wide)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19343 -0.41534 -0.06062  0.36557  1.22521
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.0004363  0.2646138   0.002    0.999
## log(Discharge.G) 0.0885713  0.0598832   1.479    0.144
##
## Residual standard error: 0.5874 on 62 degrees of freedom
##   (5504 observations deleted due to missingness)
## Multiple R-squared:  0.03408,    Adjusted R-squared:  0.0185
## F-statistic: 2.188 on 1 and 62 DF,  p-value: 0.1442
```

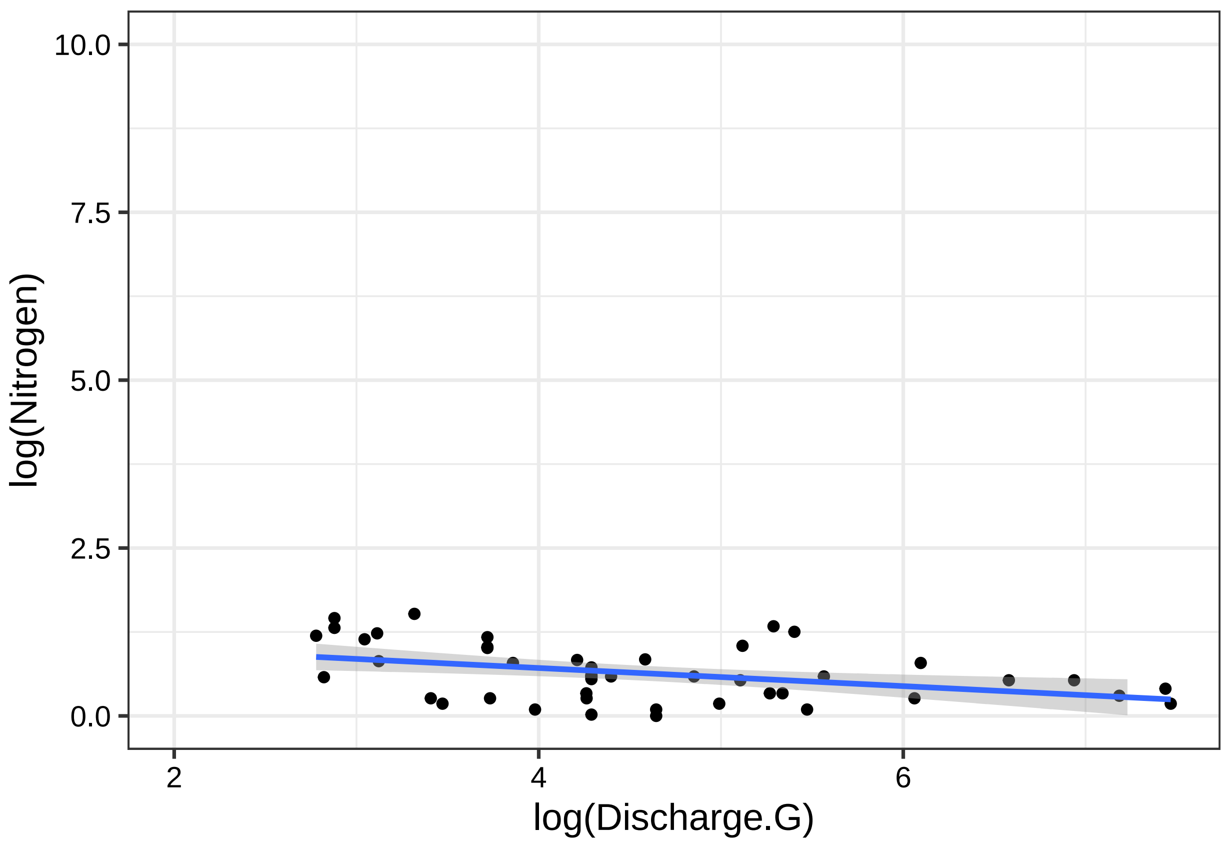
```
##
## Call:
## lm(formula = log(Phosphorus) ~ log(Discharge.C), data = EC_Flow.Nutrients_Wide)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.07959 -0.40661 -0.03792  0.36471  1.39270
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.28888    0.04884 -46.864  < 2e-16 ***
## log(Discharge.C)  0.18248    0.02132   8.557 9.91e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5479 on 157 degrees of freedom
##   (5409 observations deleted due to missingness)
## Multiple R-squared:  0.3181, Adjusted R-squared:  0.3137
## F-statistic: 73.22 on 1 and 157 DF,  p-value: 9.906e-15
```

```
##
## Call:
## lm(formula = log(Phosphorus) ~ log(Discharge.G), data = EC_Flow.Nutrients_Wide)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.06245 -0.41659 -0.08926  0.37264  2.19712
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -3.38679    0.15130 -22.385  < 2e-16 ***
## log(Discharge.G)  0.36805    0.04104   8.967 7.12e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5482 on 162 degrees of freedom
##   (5404 observations deleted due to missingness)
## Multiple R-squared:  0.3317, Adjusted R-squared:  0.3276
## F-statistic: 80.41 on 1 and 162 DF,  p-value: 7.124e-16
```

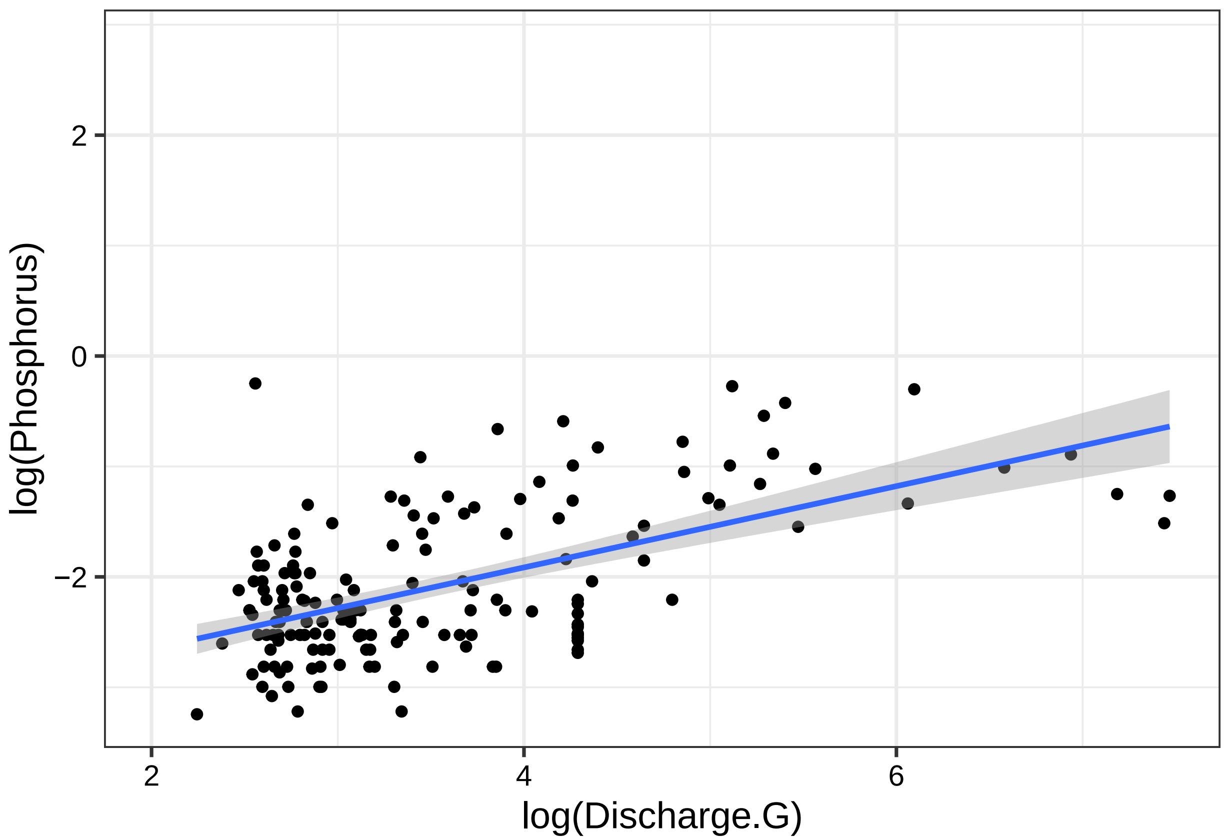
```
## Warning: Removed 5520 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 5520 rows containing missing values (geom_point).
```




```
## Warning: Removed 5404 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 5404 rows containing missing values (geom_point).
```



4.3 Question 3: How does location, upstream vs. downstream, impact nutrient levels?

Table 4: Nitrogen and Phosphorus conentrations by Location

mean.N	min.N	max.N	Standard.dev.N	mean.P	min.P	max.P	Standard.dev.P
7.740577	0.37	33	7.952157	1.134506	0.039	17	2.116099

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Club.Gorman_Flow.Nutrients_Wide$Nitrogen by Club.Gorman_Flow.Nutrients_Wide$Lo
## W = 2938, p-value = 2.446e-14
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Club.Gorman_Flow.Nutrients_Wide$Phosphorus by Club.Gorman_Flow.Nutrients_Wide$
```

```
## W = 4062.5, p-value = 0.1651
## alternative hypothesis: true location shift is not equal to 0
```

5 Summary and Conclusions

6 References

<add references here if relevant, otherwise delete this section>