

A Case Study on Chemical Flashiness of Streams and Canals in Florida

https://github.com/ytgong/ENV322_group_project

Ethan Ready, Gabi Richichi, Theo Cai, Yutao Gong

<Arrow brackets are used for annotating the RMarkdown files. Text within these brackets should not appear in the final version of the PDF document>

<**General Guidelines**> <1. Write in scientific style> <2. Global options for R chunks should be set so that only relevant output is displayed> <3. Make sure your final knitted PDF looks professional. Format tables appropriately, size figures appropriately, make sure bulleted and numbered lists appear as such, avoid awkwardly placed page breaks, etc.>

1 Rationale and Research Questions

<Write 1-2 paragraph(s) detailing the rationale for your study. This should include both the context of the topic as well as a rationale for your choice of dataset (reason for location, variables, etc.) A few citations should be included to give context for your topic. You may choose to configure autoreferencing for your citations or add these manually.>

<At the end of your rationale, introduce a numbered list of your questions (or an overarching question and sub-questions). Each question should be accompanied by one or more working hypotheses, inserted beneath each question.>

Each time a storm event occurs, it has the potential to influence the land and waterways with which it comes in contact. A storm can bring high levels of precipitation that can flood rivers and streams and temporarily alter their discharge levels. The precipitation can also alter the biogeochemistry of water bodies: eroding soil, increasing terrestrial inputs of organic matter, and transporting in contaminants such as pesticides and fertilizers. Alterations to the natural processes of streams impact the ecosystems present in the streams and can affect human activity such as hunting, fishing, recreation, and the utilization of river water for irrigation. Eutrophication is one example of a massive alteration to a river that can have negative consequences for people. Eutrophication is the induction of a dead zone devoid of oxygen in a body of water due to nutrient loading from a storm event.

The effects of storm events on bodies of water are compounded by human activity. The use of pesticides and fertilizers on agricultural land makes nutrients such as nitrogen and phosphorus accessible for storm transport in the first place. Furthermore, anthropogenically-induced climate change is intensifying natural precipitation events by making them wetter, longer, and more intense. As storms worsen, so do their effects.

This applies to the region we are studying in particular. Florida is a state that has a

It is important for us to understand the science behind the interaction of storms and st

Our research questions are:

What, if any, is the correlation between hydrologic flashiness and chemical flashiness in Florida?

What aspects of a stream influence chemical flashiness in Florida?

What aspects of a storm influence chemical flashiness in Florida?

2 Dataset Information

<Provide 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 metadata file for the dataset but formatted in a way that is more narrative.>

<Describe how your team wrangled your dataset in a format similar to a methods section of a journal article.>

<Add a table that summarizes your data structure (variables, units, ranges and/or central tendencies, data source if multiple are used, etc.). 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.>

For this analysis, our datasets were USGS NWIS high frequency and water use data. The data contained in our datasets for hypothesis 1 are the site number, date/time of measurement, instantaneous discharge, and instantaneous nitrate. Contained in the datasets for hypothesis 2 and 3 are site number, date/time of measurement, instantaneous discharge, nitrate, pH, dissolved oxygen, specific conductance; population of the county each site is in, and the amount of surface water used for thermoelectric, industrial, livestock and irrigation uses in each county. We wrangled all our data frames to include periods of data that included all of the variables, and that exhibited the most continuous measurements - all for easier, more standardized analysis and visualization. We chose sites from the state of Florida because it had sufficient sites with high frequency nitrate and discharge data.

Table 1: Variables used for analysis

Variable	Units	Type
Nitrate concentrations – high frequency	ng/L	Independent
Discharge – high frequency	m^3/s	Independent
Richards-Baker Index value		Independent
County population	k people	Independent
Irrigation water use	Bgal/day	Independent
Thermoelectric water use	Bgal/day	Independent
Chemostatic coefficient		Dependent

Table 2: USGS sites selected for analysis

Site No.	Station Name	County
2326526	WACISSA RIVER NR WACISSA FLA	Jefferson
2319302	MADISON BLUE SPRING NR BLUE SPRINGS, FL	Madison
2319950	BLUE SPRINGS NEAR DELL,FL	Lafayette
2323566	MANATEE SPRING NR CHIEFLAND FLA	Levy
2323502	FANNING SPRINGS NR WILCOX FLA	Levy

Site No.	Station Name	County
2322800	SANTA FE RIVER NR HILDRETH FLA	Gilchrist
2322700	ICHE TUCKNEE R @ HWY27 NR HILDRETH, FL	Columbia
2322688	BLUE HOLE SPRING NR HILDRETH, FL	Columbia
2310743	HUNTER SPR RUN AT BEACH LANE AT CRYSTAL RIVER FL	Citrus
2310678	HOMOSASSA SPRINGS AT HOMOSASSA SPRINGS FL	Citrus
2310650	CHASSAHOWITZKA RIVER NEAR HOMOSASSA FL	Citrus
2313100	RAINBOW RIVER AT DUNNELLON, FL	Marion
2313098	RAINBOW RIVER NEAR DUNNELLON, FL	Marion
2292900	CALOOSAHATCHEE RIVER AT S-79, NR.OLGA, FLA	Lee
2289035	THREE MILE CANAL BELOW G409 NEAR CLEWISTON, FL	Hendry
2248350	TURNBULL CREEK NR OAK HILL, FL	Volusia
2248600	DRAINAGE CANAL AT PLAZA PKWY AT COCOA, FL	Brevard
2249500	CRANE CREEK AT MELBOURNE, FL	Brevard
2250030	TURKEY CREEK AT PALM BAY, FL	Brevard
2251767	FELLSMERE CANAL NEAR MICCO, FL	Brevard
2253000	MAIN CANAL AT VERO BEACH, FL	Indian River

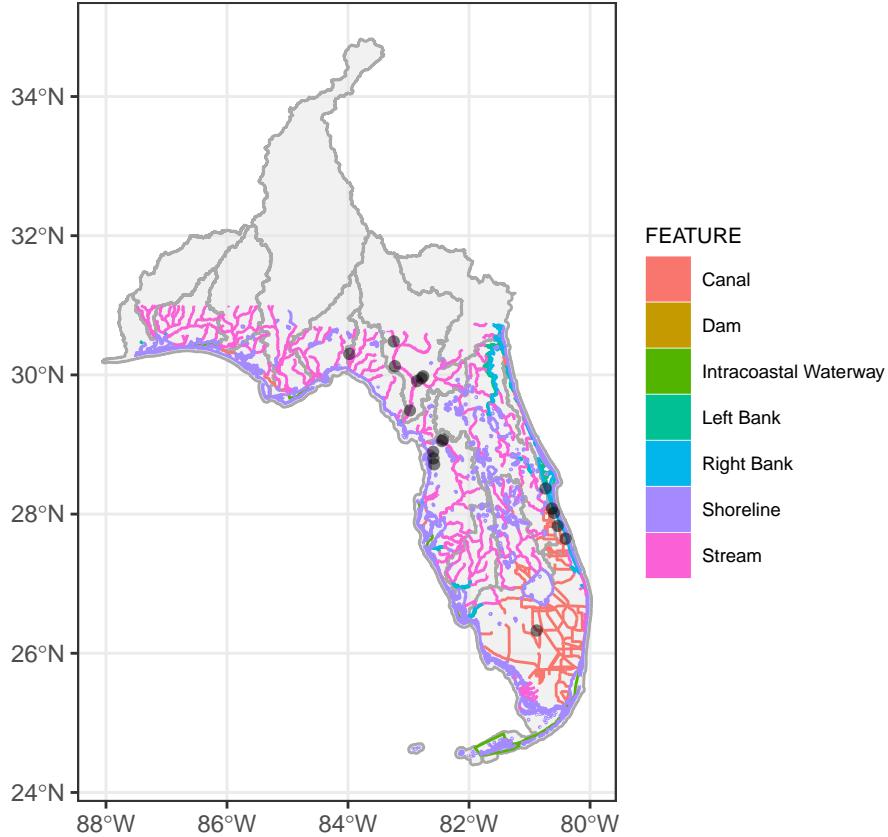


Figure 1: Location of sites analyzed on watersheds

3 Exploratory Analysis

<Insert exploratory visualizations of your dataset. This may include, but is not limited to, graphs illustrating the distributions of variables of interest and/or maps of the spatial context of your dataset. Format your R chunks so that graphs are displayed but code is not displayed. Accompany these graphs with text sections that describe the visualizations and provide context for further analyses.>

<Each figure should be accompanied by a caption, and each figure should be referenced within the text>

```
## Warning: Ignoring unknown aesthetics: show.legend
```

4 Analysis

<Insert visualizations and text describing your main analyses. Format your R chunks so that graphs are displayed but code and other output is not displayed. Instead, describe the results of any statistical tests in the main text (e.g., “Variable x was significantly different among y groups (ANOVA; df = 300, F = 5.55, p < 0.0001)”). Each paragraph, accompanied by one or more visualizations, should describe the major findings and how they relate to the question and hypotheses. Divide this section into subsections, one for each research question.>

<Each figure should be accompanied by a caption, and each figure should be referenced within the text>

4.1 Question 1: Are sites with high hydrologic flashiness more or less likely to have high chemical flashiness?

We chose to analyze our first question, on the relationship between hydrologic flashiness and chemical flashiness, using sites with high frequency nitrate and discharge measurements. At each site, we calculated the Richards-Baker Index to quantify the hydrologic flashiness.

We had to exclude catchment size in the formula because 17/21 sites did not have a recorded catchment size.

To quantify chemical flashiness, we ran a linear regression between mean daily nitrate and mean daily discharge values and took the coefficient as our chemostaticity value. We excluded 3 sites with poor linear regression fits, 1 of which had a large outlier for chemostaticity coefficient.

We wrangled the RBI values and Chemostaticity Coefficients to create a dataframe with columns site, site type(canal, river, etc.), chemostaticity coefficient, chemostaticity coefficient absolute value, and RBI value. We included the absolute value of the Chemostaticity Coefficient to investigate the flashiness, as absolute value would account for large chemostaticity in either a flushing or diluting system.

We created ggplot visualizations of Chemostaticity Coefficient vs. RBI and |Chemostaticity Coefficient| vs. RBI separating site type by color. The visualization without absolute value shows a distribution that follows the x and y axis. Sites with large RBI values have low Chemostatic Coefficients, and sites with low RBI values have Chemostatic Coefficients that are far above 0 or far below 0. The absolute value visualization reinforces this finding, as large absolute Chemostatic Coefficients correspond with small RBI values. Similar trends were viewed between site type, except that the largest outlier for both Chemostatic Coefficient and RBI are both canals.

We ran a linear regression on log(Absolute Chemostaticity Coefficient) vs. RBI, which was the closest fit that we could find. The regression produced a coefficient of -2.6847 , a P-value of 0.473, and a multiple r-squared of 0.03264. The model does not adequately predict the data.

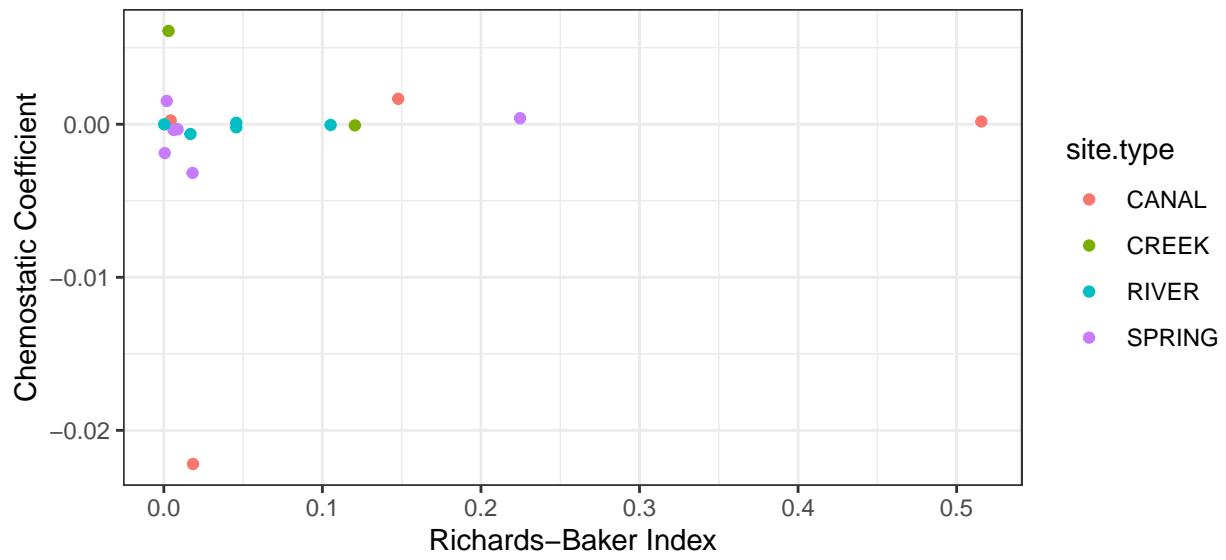


Figure 2: Chemostatic coefficient values and Richards-Baker Index of 18 sites

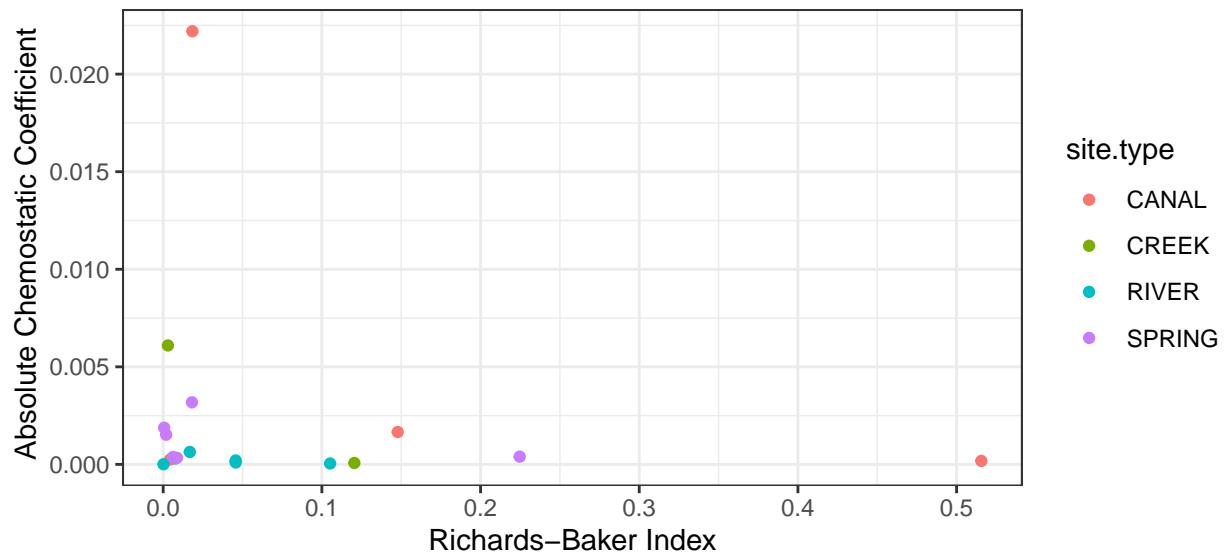


Figure 3: Absolute Chemostatic coefficient values and Richards-Baker Index of 18 sites

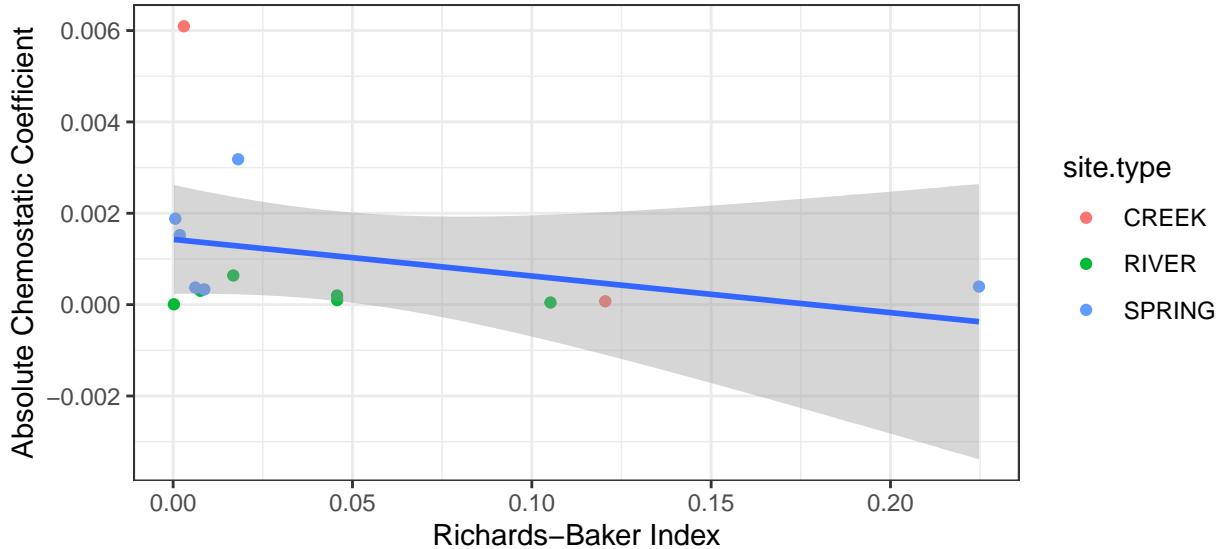


Figure 4: Absolute Chemostatic coefficient values and Richards-Baker Index of 14 non-canal sites

```
##
## Call:
## lm(formula = log(chemo.coef.abs) ~ RBI_value_ind_noncatch, data = RBI_CHEMO_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -4.2881 -0.8347 -0.0628  1.2037  3.7717 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -7.5299     0.5232 -14.392 1.42e-10 ***
## RBI_value_ind_noncatch -2.6847     3.6538  -0.735    0.473    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.921 on 16 degrees of freedom
## Multiple R-squared:  0.03264,    Adjusted R-squared:  -0.02782 
## F-statistic: 0.5399 on 1 and 16 DF,  p-value: 0.4731
```

Since canals often act differently than natural waterways, and the largest outlier for both RBI and Chemostaticity Coefficient are canals, we also visualized the data without canals. The visualizations are more clear due to smaller axis ranges, and show similar results to the first visualizations. We ran the same linear regression on the Absolute Chemostatic Coefficient without canals.

```
##
```

```

## Call:
## lm(formula = log(chemo.coef.abs) ~ RBI_value_ind_noncatch, data = RBI_nocanals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -4.1233 -0.8833 -0.2023  1.3591  2.6141 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -7.694     0.587  -13.106 1.8e-08 ***
## RBI_value_ind_noncatch -6.889     7.716   -0.893    0.39    
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 1.808 on 12 degrees of freedom
## Multiple R-squared:  0.06228,    Adjusted R-squared:  -0.01586 
## F-statistic: 0.797 on 1 and 12 DF,  p-value: 0.3895

```

The linear regression for Absolute Chemostatic Coefficient vs. RBI provided a Coefficient of -6.889, a P-value of 0.39, and a multiple r-squared of 0.06228. While the fit doesn't sufficiently match the data, it is better than the data with canals and does provide some idea of the relationship between the variables.

4.2 Question 2: How do features of a site impact its chemical flashiness?

4.2.1 Qualitative Approach: How do geographical features impact the site's chemical flashiness?

We made observations by first sorting sites into three categories - Chemostatic, Flushing, and Diluting - based on the slopes of the regression lines of their C-Q plots - values near 0, positive values, and negative values, respectively. Then, we went site by site and searched their coordinates into Google Maps and ArcGIS, and noted the general surrounding geography - is it mostly urban or rural? Is it in a state park, adjacent to it, or nowhere near? Where, generally, in Florida is this site located? We also paid special attention to the site type - river, spring, or canal - as each most likely have differing management strategies and hydrologies.

When we were done listing the qualitative traits of all the sites in each hydrologic behavioral categorical, we noticed some trends had emerged. For the most flushing sites, or, the ones with the steepest regression line slopes, we noticed that all are located near airports. (Main Canal, Fellsmere Canal, Crane Creek, and Madison Blue Spring) There does not seem to be a correlation with the urbanity of the location, however, as two out of four of these sites are mostly rural and two are mostly urban - Fellsmere, Madison Blue Spring and Main Canal, Crane Creek, respectively. We hesitate to draw any concrete conclusions about possible

relationships between flushing systems and proximity to airports, as there may be many other factors involved that are not within the scope of this project. One important factor could be the management style of the site; three out of four of the canals analyzed in this project tend towards flushing. This, too, could be a coincidence, and we would need to do a deeper analysis with more sites to draw conclusions. Our list of diluting sites show more convincing trends - seven out of nine of the sites are located in very rural locations, most in or on the edge of a Florida state park. (Ichetucknee, Santa Fe, Blue Spring, Blue Hole Spring, Manatee Spring, Homosassa Spring, Turkey Creek) However, this number may be slightly inflated because three of the four most diluting sites are directly interconnected, increasing the likelihood of similar hydrologic tendencies between them. (Ichetucknee, Santa Fe, Blue Hole Spring) There seem to be no other notable qualitative trends for diluting sites. Chemostatic or mostly chemostatic systems do not seem to display any particular kind of trend. They are located in all kinds of environments - from hyper-urban to deep in state parks.

```
## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Removed 3263 rows containing non-finite values (stat_smooth).

## Warning: Removed 3015 rows containing missing values (geom_point).

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Removed 1488 rows containing non-finite values (stat_smooth).

## Warning: Removed 1488 rows containing missing values (geom_point).

## TableGrob (1 x 1) "arrange": 1 grobs
##   z      cells    name        grob
## 1 1 (1-1,1-1) arrange gtable[arrange]
```

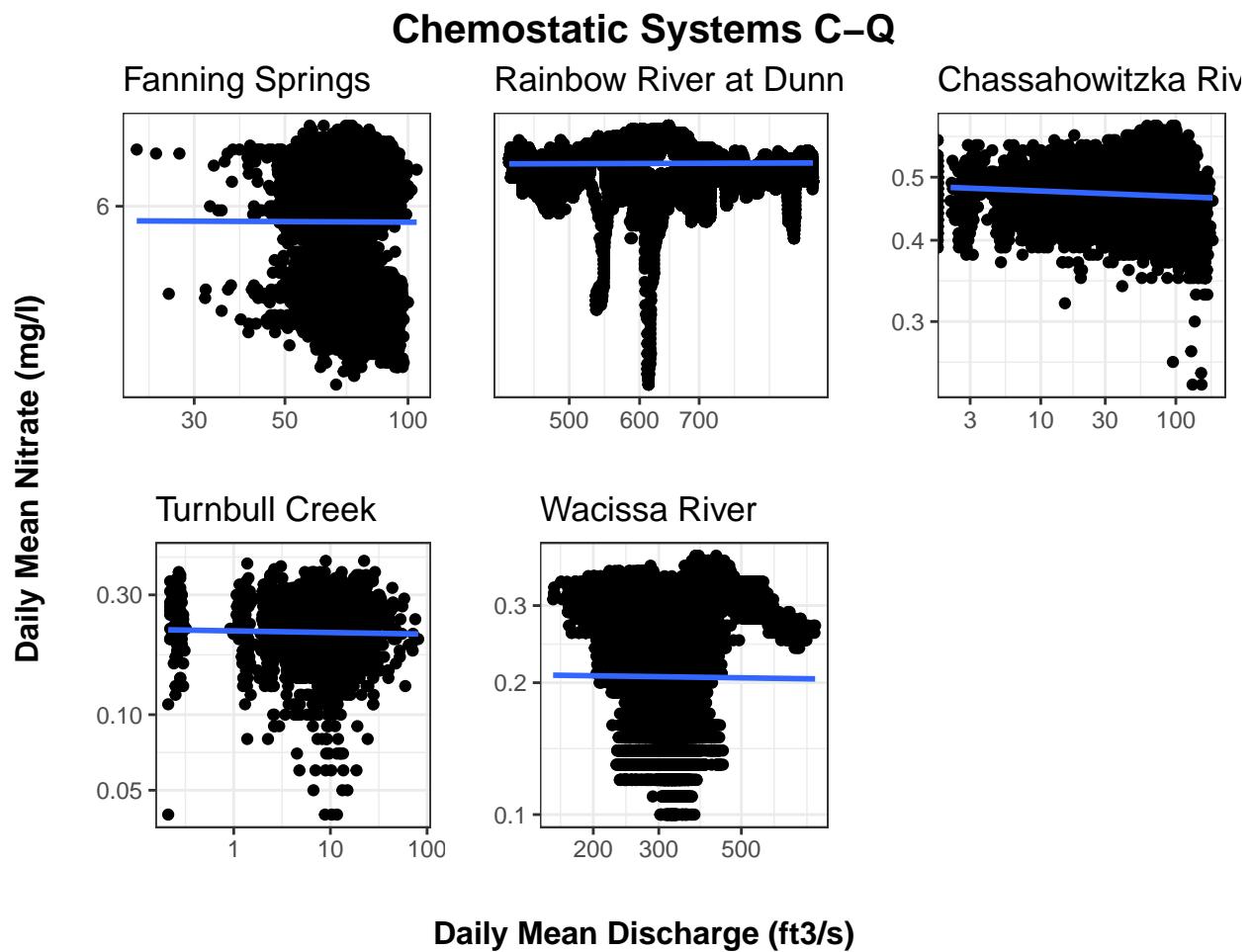


Figure 5: C-Q Plots of Chemostatic Sites

```
## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Removed 603 rows containing non-finite values (stat_smooth).

## Warning: Removed 603 rows containing missing values (geom_point).

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 35646 rows containing non-finite values (stat_smooth).

## Warning: Removed 21788 rows containing missing values (geom_point).

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Removed 3749 rows containing non-finite values (stat_smooth).

## Warning: Removed 2 rows containing missing values (geom_point).
```

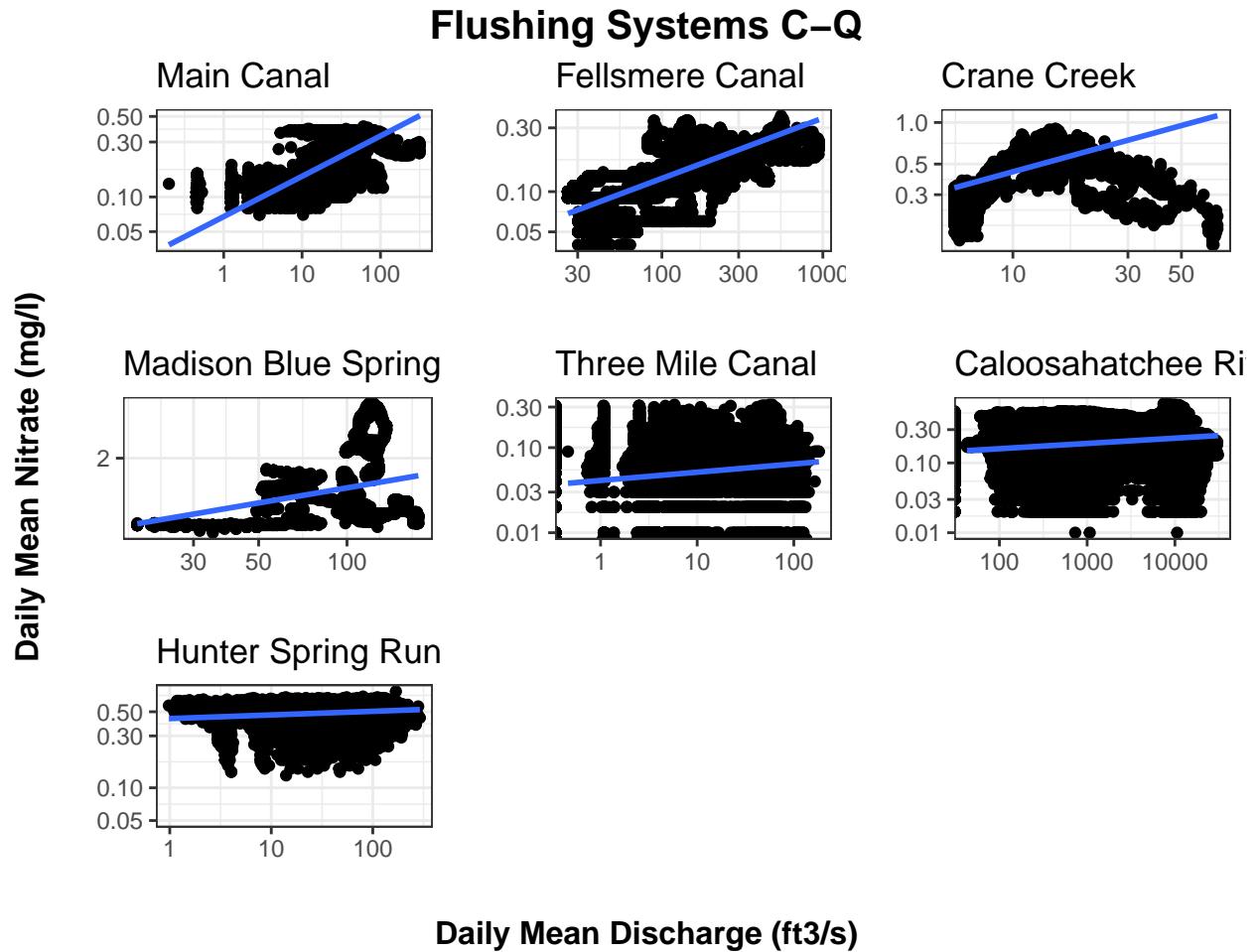


Figure 6: C-Q Plots of Flushing Sites

```
## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Removed 19144 rows containing non-finite values (stat_smooth).

## Warning: Removed 19144 rows containing missing values (geom_point).

## TableGrob (1 x 1) "arrange": 1 grobs
##   z   cells   name      grob
## 1 1 (1-1,1-1) arrange gtable[arrange]

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning in self$trans$transform(x): NaNs produced

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Removed 28 rows containing non-finite values (stat_smooth).

## Warning: Removed 28 rows containing missing values (geom_point).

## TableGrob (1 x 1) "arrange": 1 grobs
##   z   cells   name      grob
## 1 1 (1-1,1-1) arrange gtable[arrange]
```

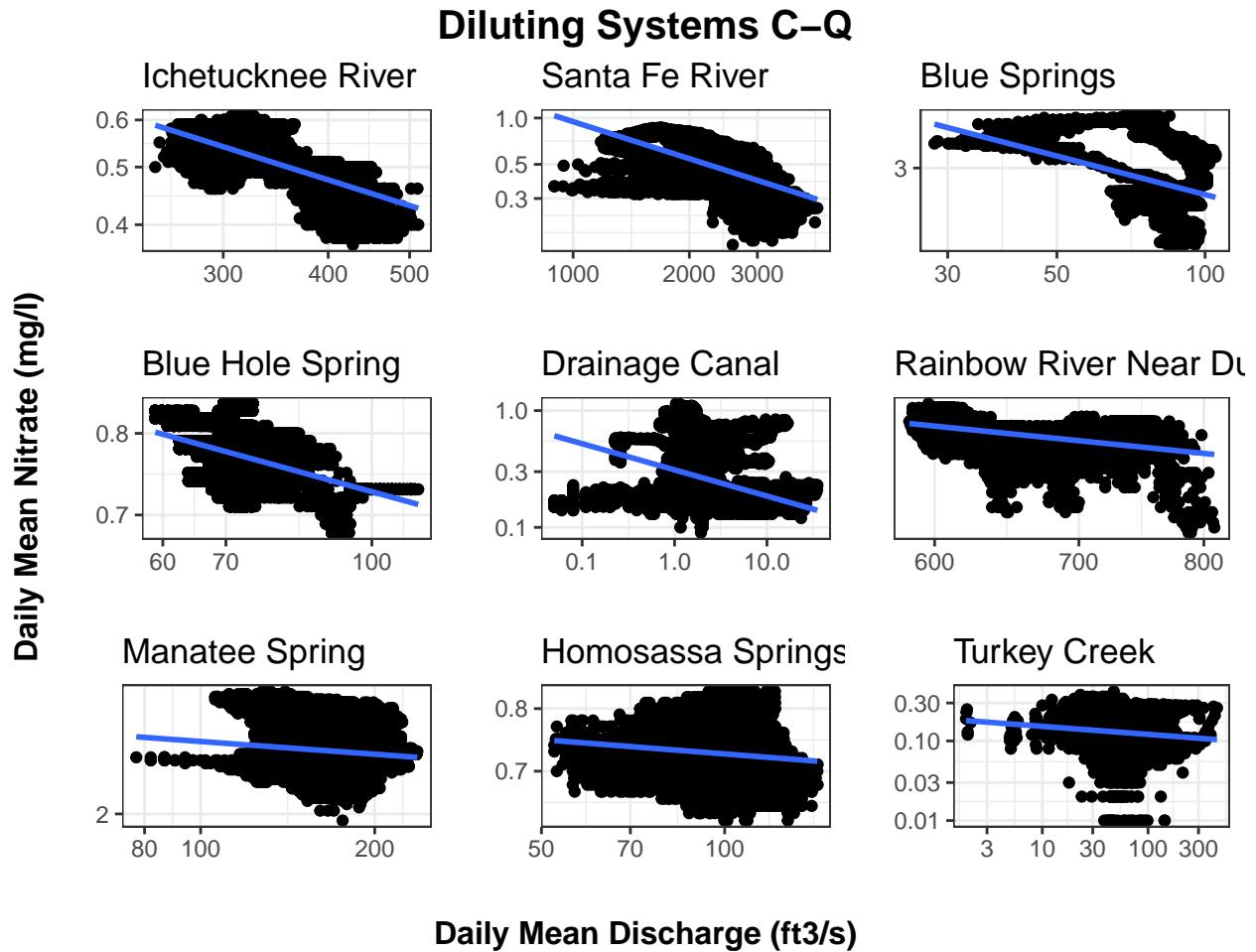
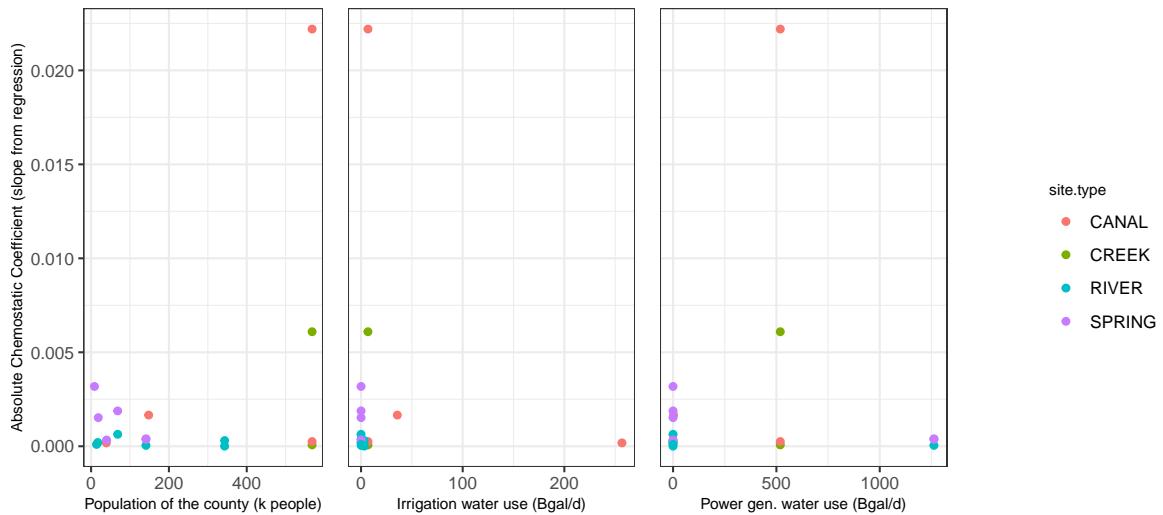


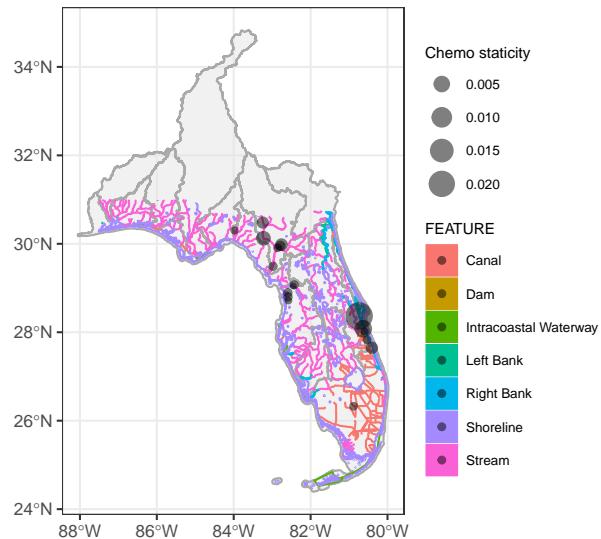
Figure 7: C-Q Plots of Diluting Sites

4.2.2 Quantitative/Statistical Approach: How do counties' population and water usage impact the chemical flashiness of sites within the county?

Impact of Population and Water Use on Abs. Chemostatic Coeff.



```
## Warning: Ignoring unknown aesthetics: show.legend
```



5 Summary and Conclusions

<Summarize your major findings from your analyses in a few paragraphs. What conclusions do you draw from your findings? Relate your findings back to the original research questions and rationale.>

6 References

- Chislock, M. F., Doster, E., Zitomer, R. A. & Wilson, A. E. (2013) Eutrophication: Causes, Consequences, and Controls in Aquatic Ecosystems. *Nature Education Knowledge* 4(4):10
- Smith, V. H. & Schindler, D. W. Eutrophication science: where do we go from here? *Trends in Ecology and Evolution* 24, 201-207 (2009).
- Stephenson, C. "Addressing Eutrophication in Florida, one watershed at a time." University of Florida: IFAS Extension. July 23, 2018. Web: <https://nwdistrict.ifas.ufl.edu/nat/2018/07/23/addressing-eutrophication-in-florida-one-watershed-at-a-time/>.