Assessing the Woonasquatucket River Watershed Water Quality: A Comprehensive Exploratory Analysis

Quinn Chamblin, Katy Cummings, Jeniah Franco, Nina Hernandez

About us!

Quinn: Statistics & Data Science and Biology double major, post-Smith moving back to California to apply to Biostatistics or Data Analytics fields!

Katy: Statistics & Data Science and Psychology double major, accepted an offering with Teach for America, so post-graduation I am moving to Philly!

Jeniah: Statistics and & Data Science, Spanish minor, post-grad moving to Boston!

Nina: Statistics & Data Science and Bio-Chem double major, post-Smith attending post-bacc program at Brown University in Biotech!

SDS Capstone 410

The Goal of the Capstone:

- Leverage students’ previous coursework to address a real-world data analysis problem

What does it entail?

- Students collaborate in teams on projects sponsored by industry, government, or academia.
- Students are required to use skills obtained from previous coursework such as ethics, project management, collaborative software development and consulting.

Introduction

- WRWC has accumulated data from over 30 years that has not been efficiently assessed and organized
- No way to produce **quality analysis** with previous set up
- Lead to undiscovered discrepancies in data and overall lack of understanding
- Previous approaches have focused on incremental data collection with limited analysis, due to limited resources

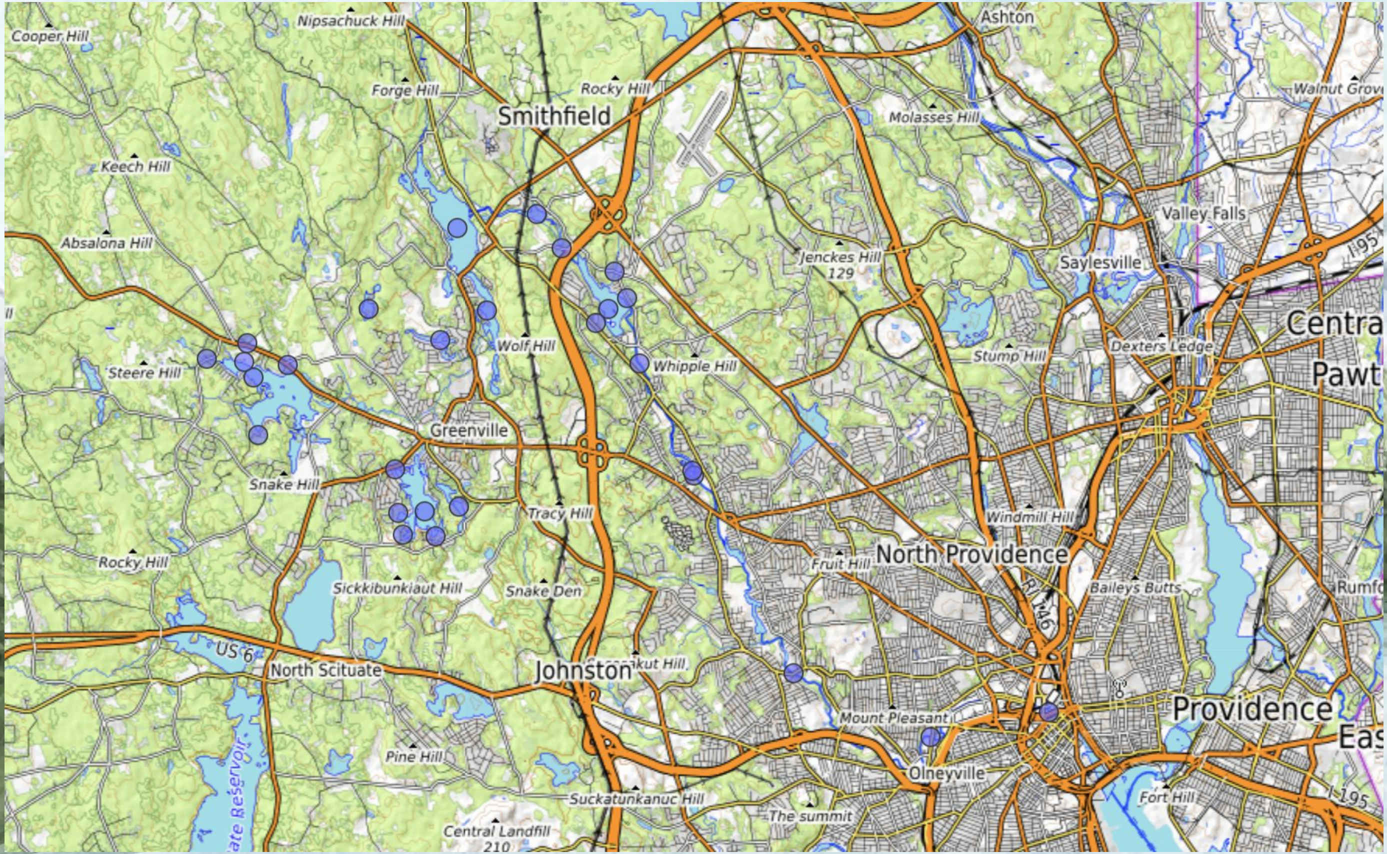
Research Question/Purpose

- Overall, we want to **close the gaps** between the potential the data has and WRWC’s understanding of their data by:
    - **Attaining** a holistic understanding of the watershed's health
    - **Identifying** specific problem areas effectively
    - **Addressing** challenges in engaging the community due to the absence of accessible and engaging data representations.

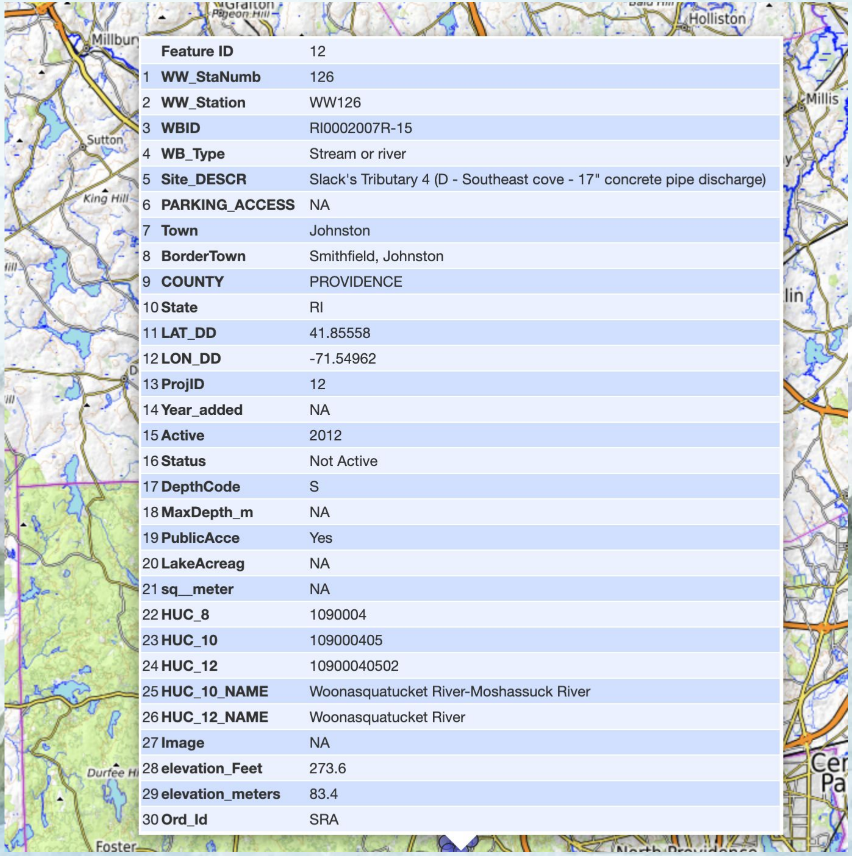
Methodology Overview

- **Data Collection:** Description of the 28 sampling sites along the watershed.
- **Analytical Tools Used:** R programming, with packages like ggplot2, dplyr, and tidyverse.
- **Data Preprocessing:** Steps included data cleaning, filtering for specific parameters, and handling of missing values.
- **Statistical Analysis:** Use of linear regression models to understand trends.
- **Exploratory Data Analysis:** Employed to identify patterns, trends, and outliers in the data.

28 Location sites
7 Parameters



Interactive Map of Sites



	Feature ID	12
1	**WW_StaNumb**	126
2	**WW_Station**	WW126
3	**WBID**	RI0002007R-15
4	**WB_Type**	Stream or river
5	**Site_DESCR**	Slack's Tributary 4 (D - Southeast cove - 17" concrete pipe discharge)
6	**PARKING_ACCESS**	NA
7	**Town**	Johnston
8	**BorderTown**	Smithfield, Johnston
9	**COUNTY**	PROVIDENCE
10	**State**	RI
11	**LAT_DD**	41.85558
12	**LON_DD**	-71.54962
13	**ProjID**	12
14	**Year_added**	NA
15	**Active**	2012
16	**Status**	Not Active
17	**DepthCode**	S
18	**MaxDepth_m**	NA
19	**PublicAcce**	Yes
20	**LakeAcreag**	NA
21	**sq__meter**	NA
22	**HUC_8**	1090004
23	**HUC_10**	109000405
24	**HUC_12**	10900040502
25	**HUC_10_NAME**	Woonasquatucket River-Moshassuck River
26	**HUC_12_NAME**	Woonasquatucket River
27	**Image**	NA
28	**elevation_Feet**	273.6
29	**elevation_meters**	83.4
30	**Ord_Id**	SRA

Interactive Map of Sites

Upstream / Downstream (part 1)

Table: Watershed Locations in Descending Order by Site Latitude (1-7)

Site Location Number
52
65
113
114
680
144
16

Table: Watershed Locations in Descending Order by Site Latitude (8-14)

Site Location Number
153
679
24
518
239
240
635

Table 1 & 2: Site Locations in Descending order (first-fourteen)

Upstream / Downstream (part 2)

Table: Watershed Locations in Descending Order by Site Latitude (15-21)

Site Location Number
238
61
241
123
437
226
201

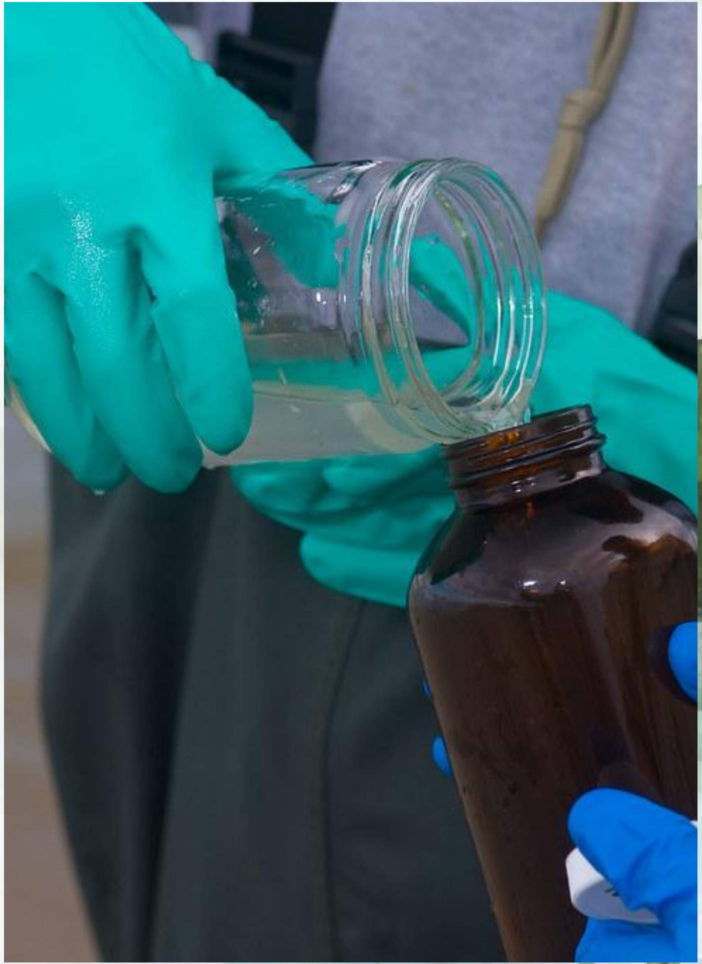
Table: Watershed Locations in Descending Order by Site Latitude (22-28)

Site Location Number
46
124
125
126
508
308
227

Table 3 & 4: Site Locations in Descending order (Fifteen-Twenty-Eight)

Parameters of Interest

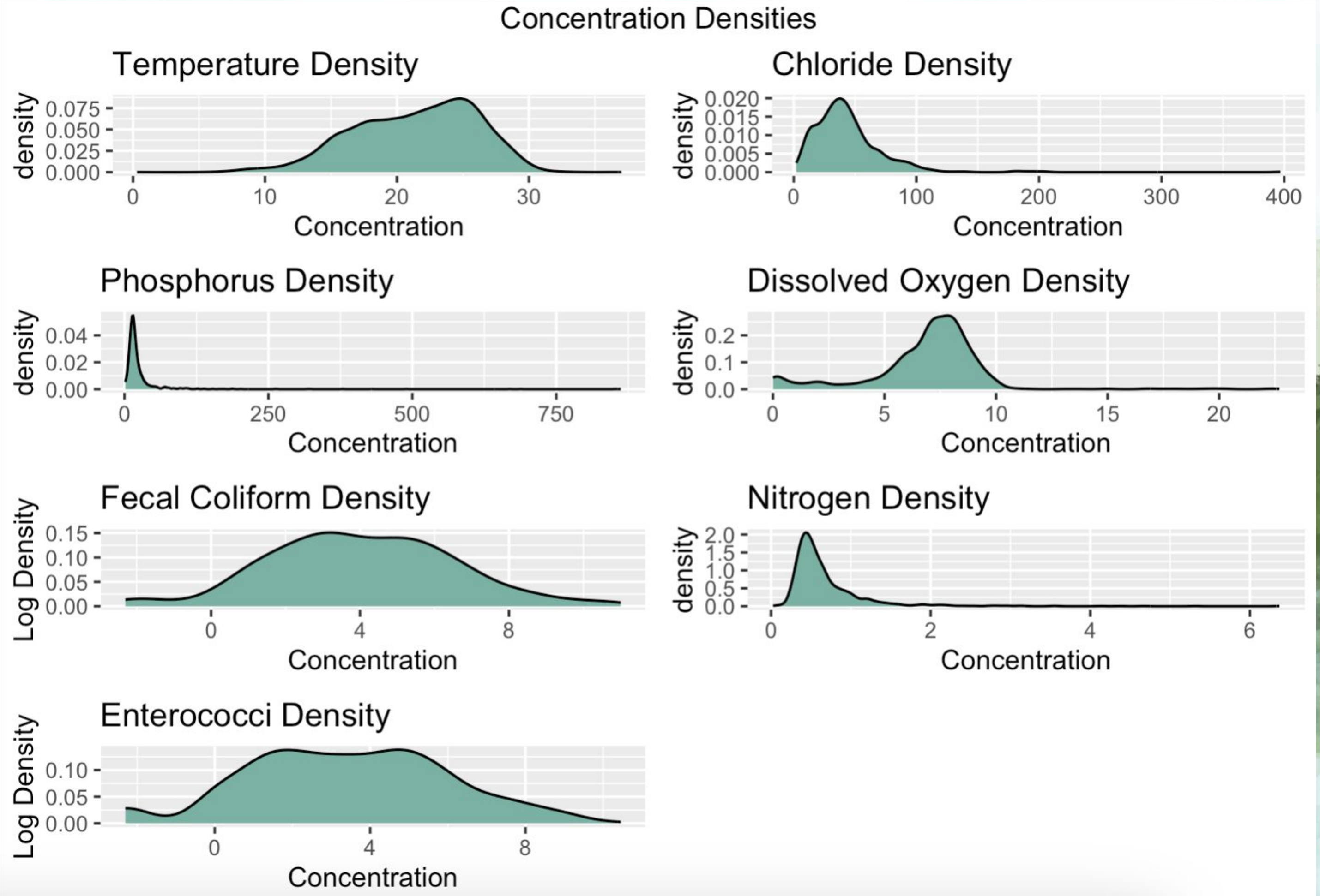
- **Chloride** - impacts on aquatic life
- **Phosphorus** - eutrophication risk
- **Dissolved Oxygen** - health of aquatic organisms
- **Fecal Coliform** - indicator of water quality and contamination
- **Nitrogen** - nutrient loading and algal blooms
- **Enterococci** - beach closures and human health risks
- **Temperature** - affects dissolved oxygen levels and species composition



Summary Statistics of the Seven Parameters

	Mean	Median	Standard_deviation	Sample_size
Chloride (mg/l)	43.4674419	39.500	29.2530227	860
Dissolved Oxygen (mg/l)	6.8215494	7.300	2.5855086	1791
Enterococci (MPN/100)	506.2722160	30.600	2006.6218433	898
Fecal Coliform (CFU/100ml)	1227.8516129	72.000	5238.2804001	403
Nitrogen (mg/l)	0.6748992	0.535	0.4882856	1532
Phosphorus (ug/l)	25.3616822	16.000	48.6313124	1605
Temperature (C)	21.4370765	22.000	4.5889805	4965

* 2,698 Enterococci (MPN/100)
*18,853 Fecal Coliform (CFU/100ml)



Density Plots for All Seven Parameters (normality condition is violated)

Exploratory Findings - Bar plots of mean concentration per year

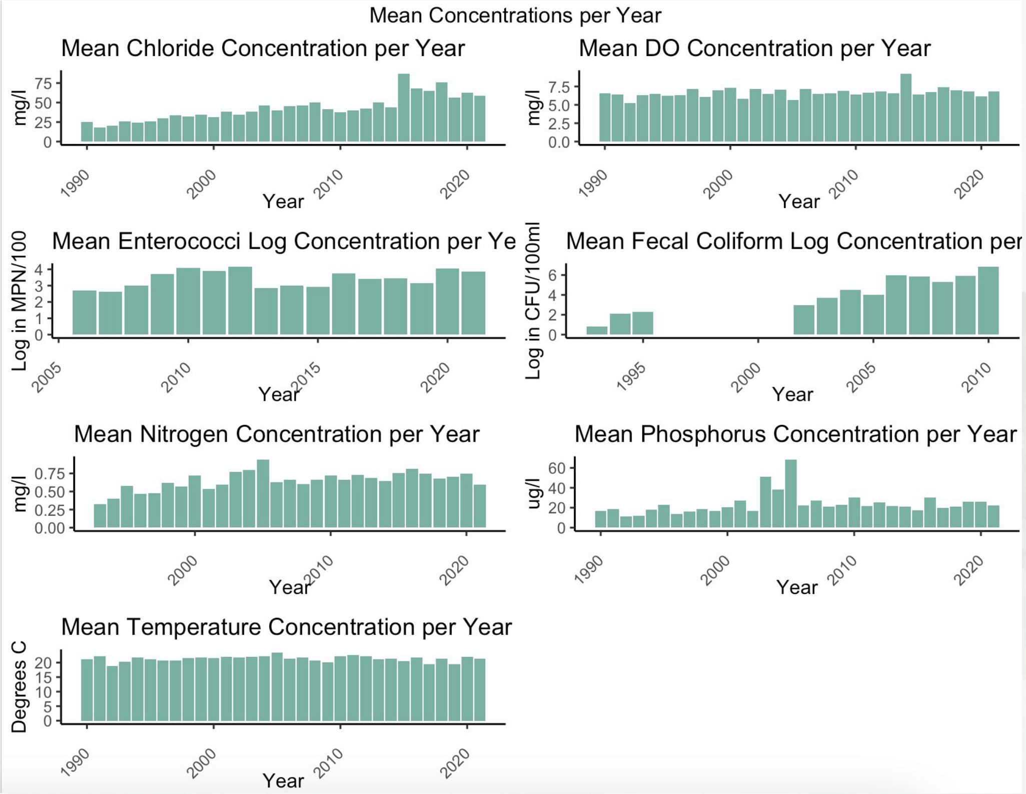


Figure 1. Bar plot distribution of mean parameter concentration for 7 parameters from 1990 to 2021

Exploratory Findings - Bar plots of mean concentration per site

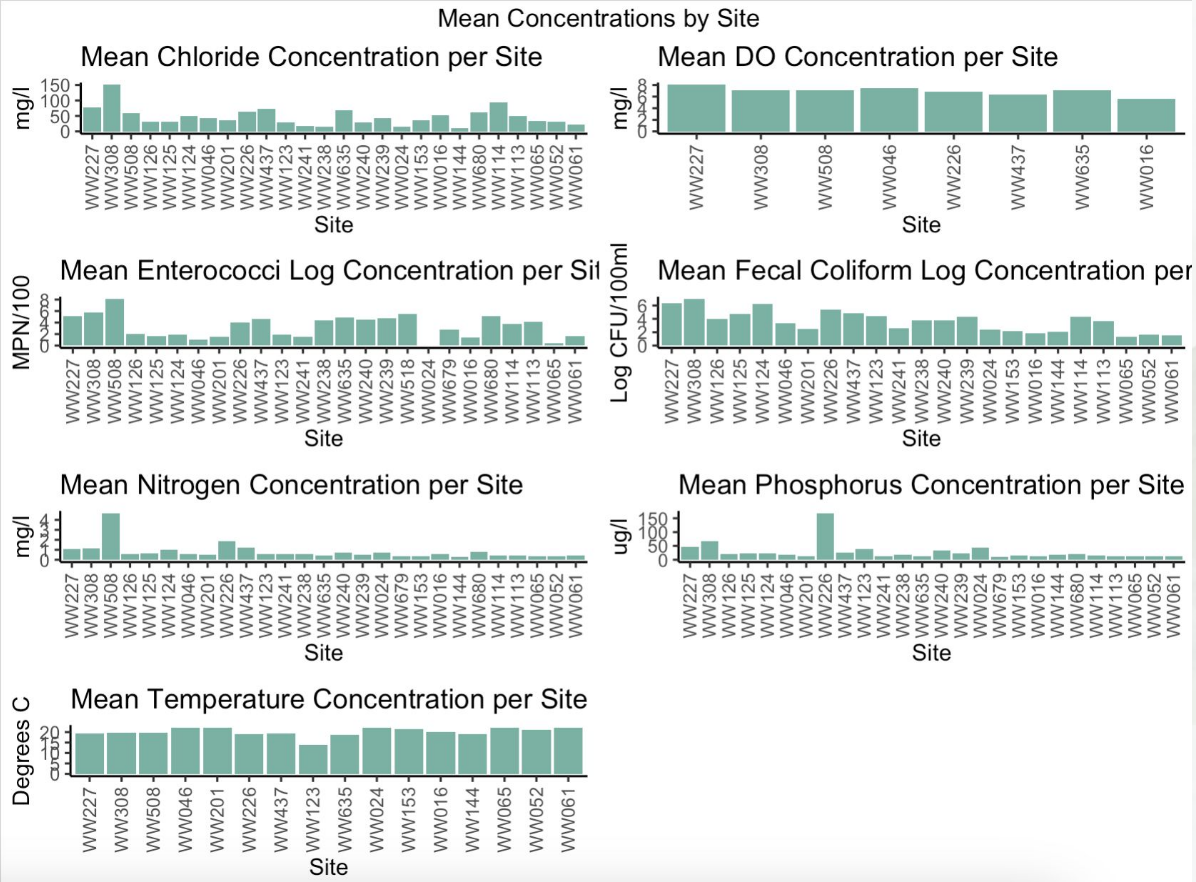


Figure 2. Bar plot distribution of mean parameter concentration for 7 parameters per site

Exploratory Findings - Bar plots of mean concentration per decade

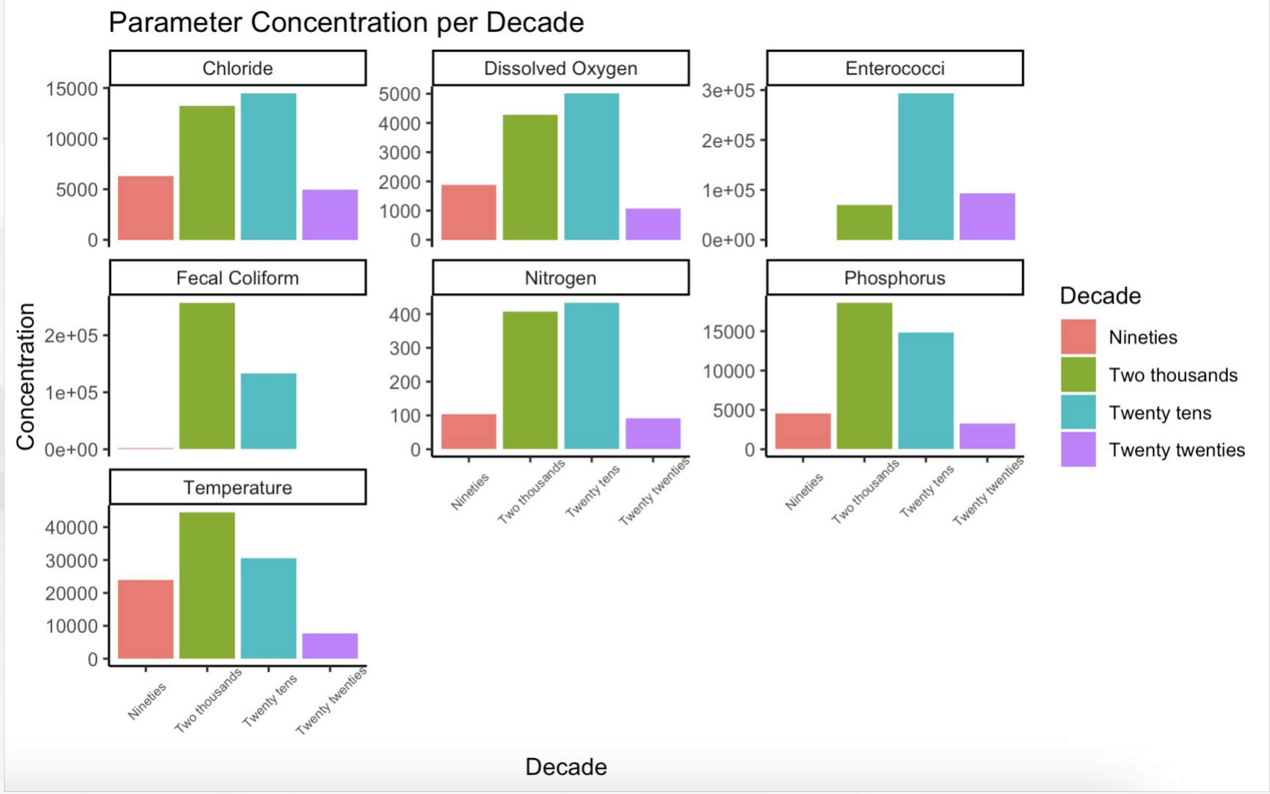


Figure 2. Bar plot distribution of mean parameter concentration for 7 parameters per decade

Kruskal-Wallis Test by Year

Table: Statistical Significance Table for Parameters by Year

Parameter	P_value	Significance
Chloride	0.000	Significant
Dissolved Oxygen	0.000	Significant
Enterococci	0.003	Significant
Fecal Coliform	0.000	Significant
Nitrogen	0.000	Significant
Phosphorus	0.000	Significant
Temperature	0.000	Significant

p-value less than or equal to 0.05 (by alpha number) is considered to be statistically significant

Kruskal-Wallis Test by Site

Table: Statistical Significance Table for Parameters by Site

Parameter	P_value	Significance
Chloride	0	Significant
Dissolved Oxygen	0	Significant
Enterococci	0	Significant
Fecal Coliform	0	Significant
Nitrogen	0	Significant
Phosphorus	0	Significant
Temperature	0	Significant

p-value less than or equal to 0.05 (by alpha number) is considered to be statistically significant

Exploratory Findings

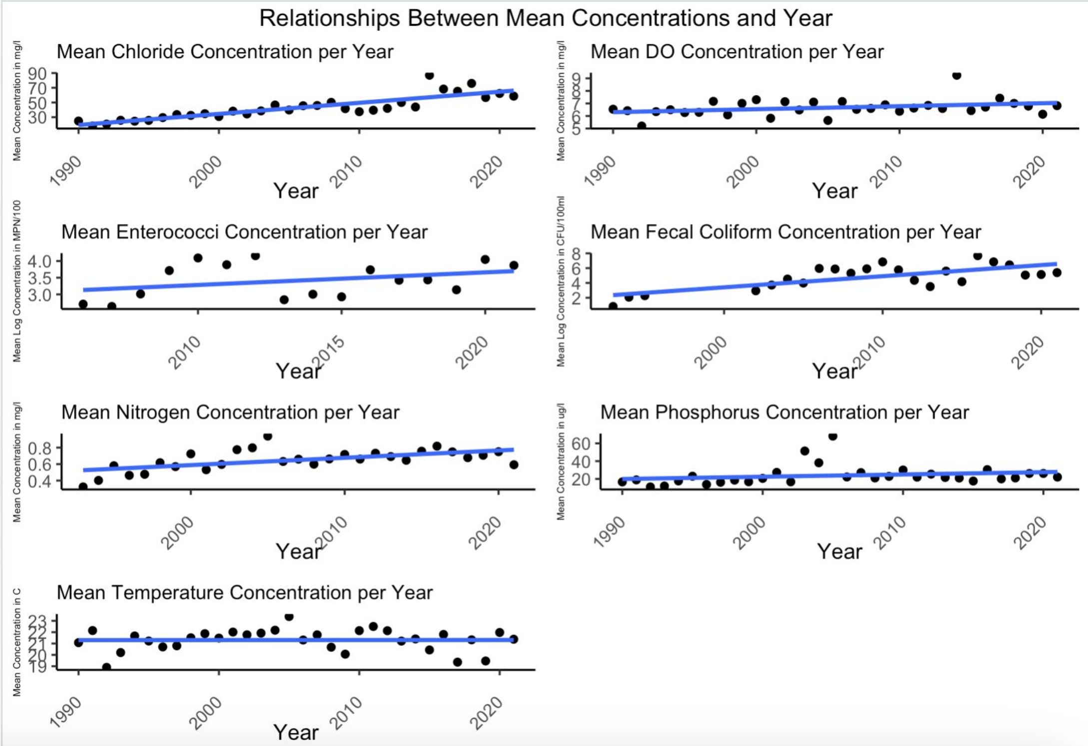
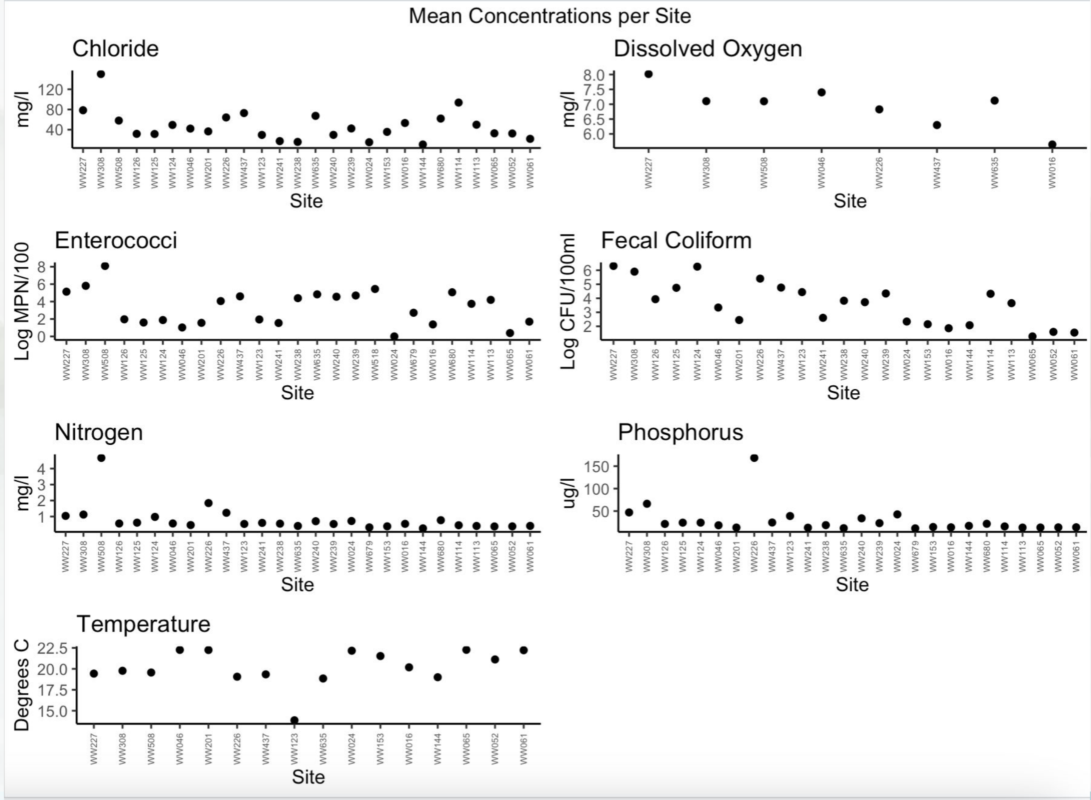
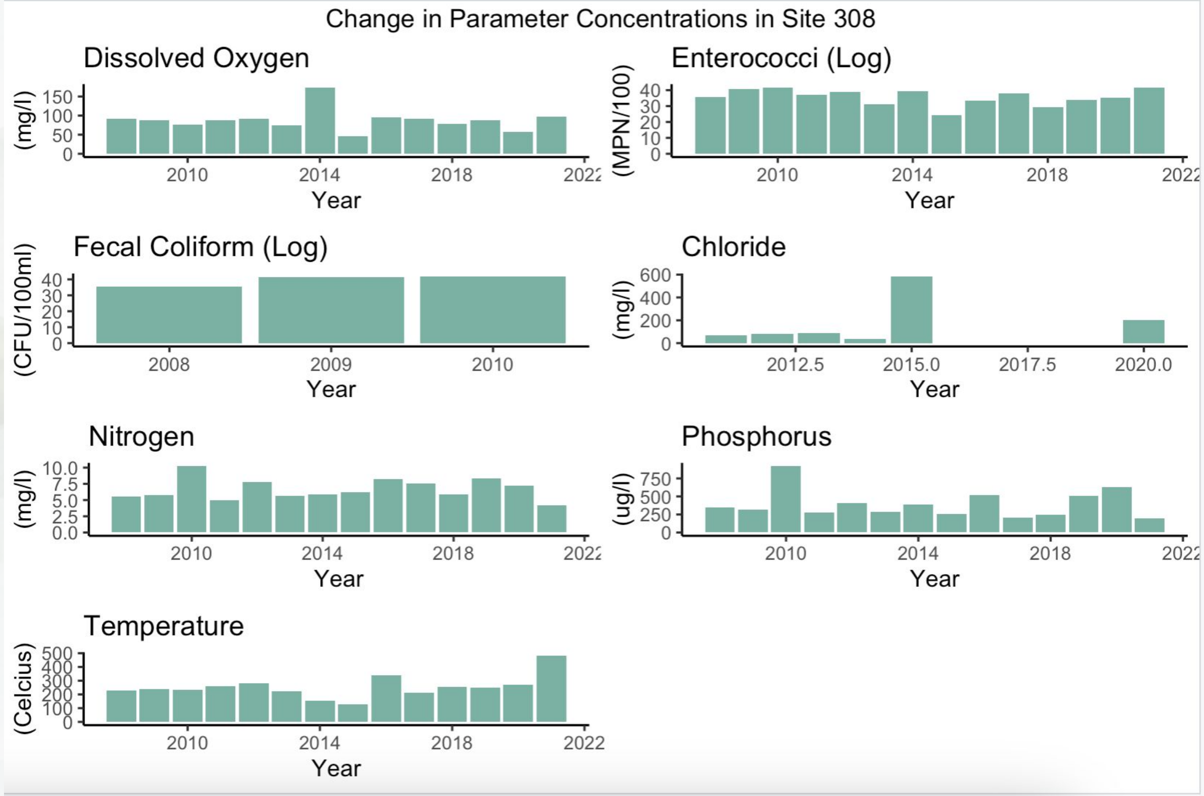


Figure 2. Linear relationships between Mean Concentration of Parameters by Year
****relationships between mean concentration and site do not have enough data to show a line of best fit**

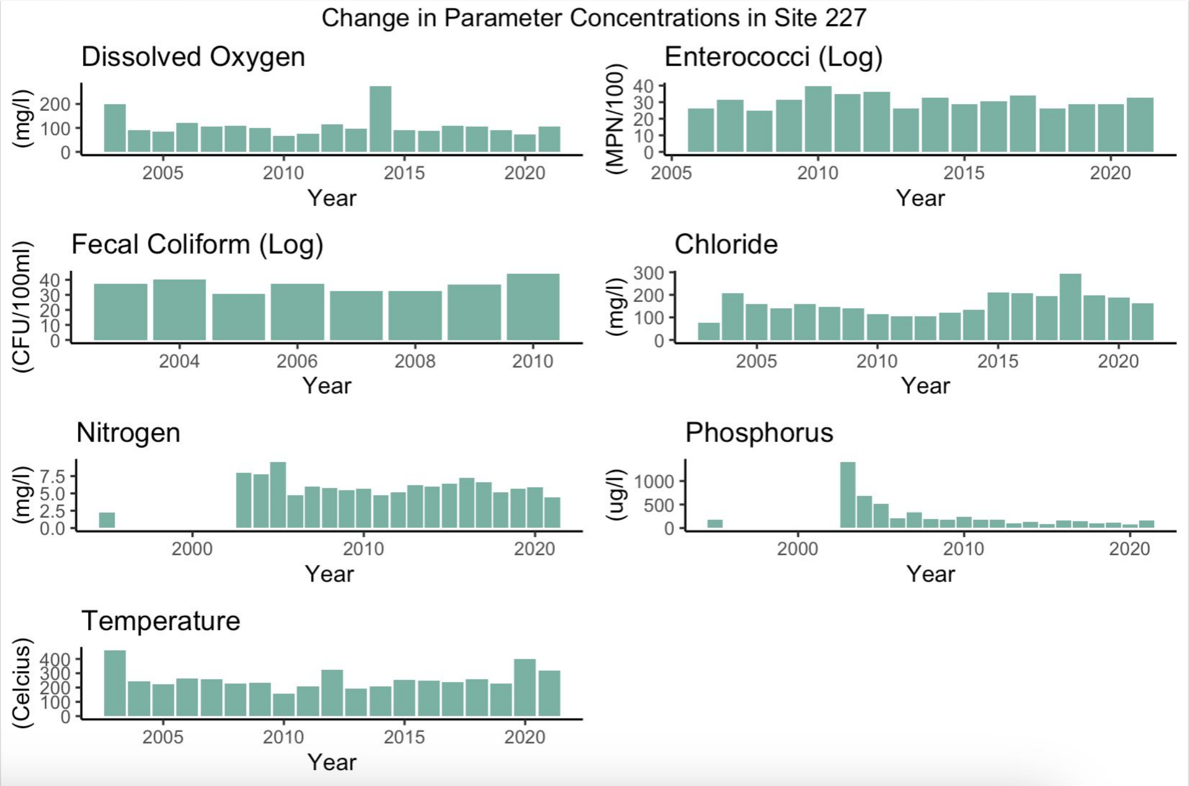
Exploratory Findings



Changes in Site 308 over Time

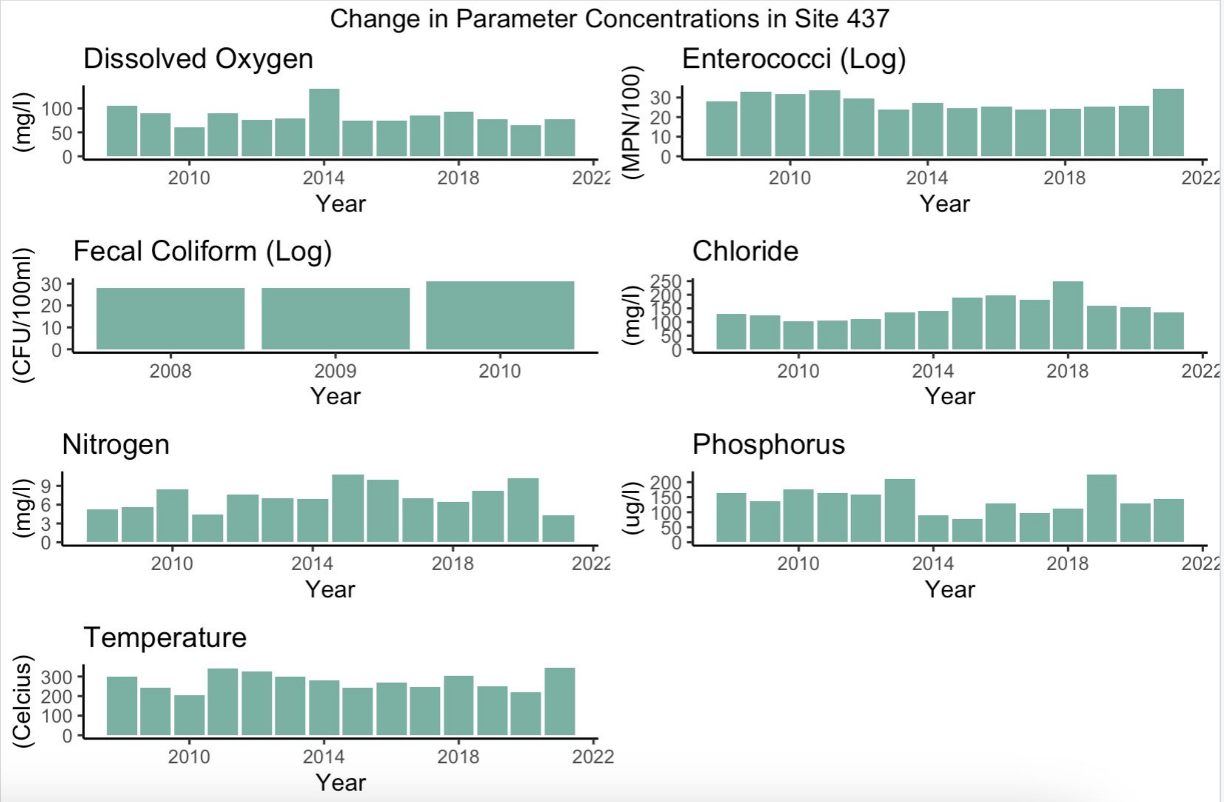


Changes in Site 227* Over Time

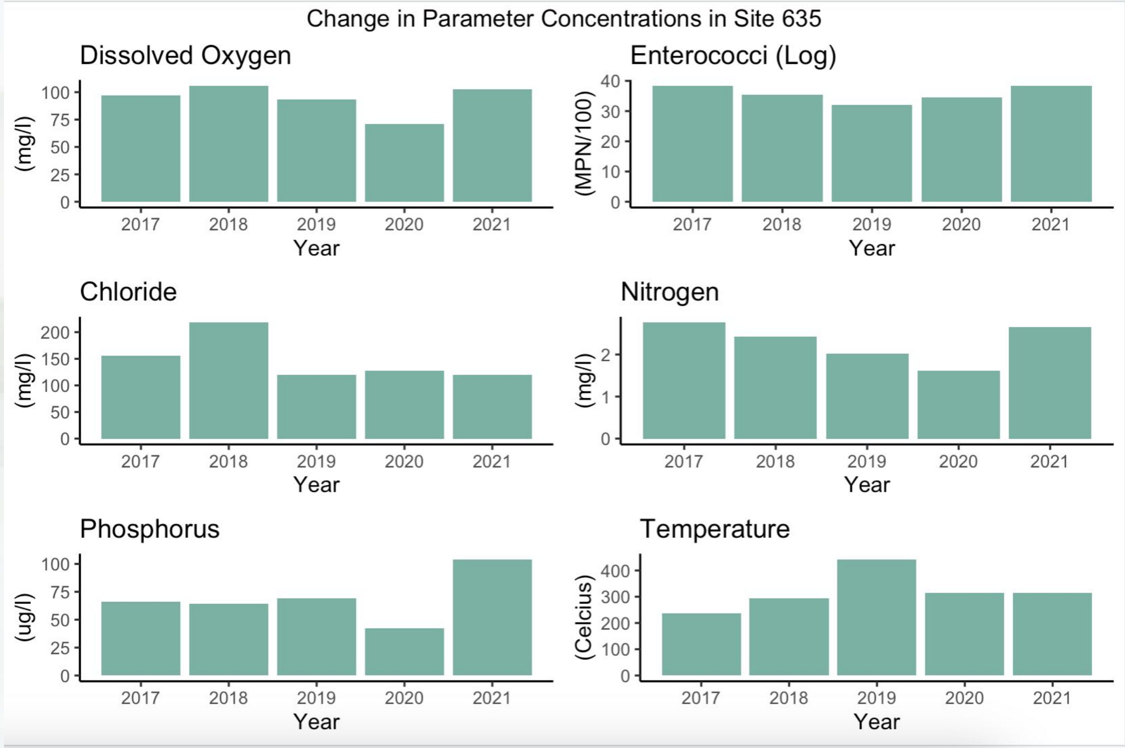


*This is the most downstream site

Changes in Site 437 over Time



Changes in Site 635 Over Time



- No data for Fecal Coliform at this site

Implications of Findings

- Reveals significant correlations between the year and changes in water quality parameters such as chloride and fecal coliform

- The high R^2 values for chloride and fecal coliform suggest these parameters are more predictable over time, which is crucial for managing public health risks associated with water quality.

Implications cont.

- Lower R^2 values for parameters like dissolved oxygen and phosphorus highlight the complexity of ecological dynamics in the watershed, suggesting that current management strategies may need reevaluation to effectively address these variables.

- The shift in EPA standards for measuring fecal contamination highlights the need for updated and consistent methods in monitoring water quality to align with the latest public health guidelines.

Recommendations

- Explore non-linear models such as logistic or exponential regressions to better fit the relationships between the year and water quality parameters.

- Investigate the effects of different sites within the watershed to understand spatial variations in water quality.

- Transition to testing parameters like E. coli as recommended by EPA to ensure that future water quality assessments align with current public health standards.

- Standardize and improve data collection methods to reduce inconsistencies and improve the reliability of future studies. Training on ethical and organized data collection is essential.

- Expand the scope of research to include environmental factors such as climate change, human activity, and land use changes, which likely impact water quality.

Conclusion

- Our research has laid a foundational framework by establishing a comprehensive dataset that highlights key correlations and trends in the Woonasquatucket River Watershed water quality over the years.
- We urge environmental scientists and policymakers to utilize our findings as a basis for further detailed studies. Advanced analytical techniques are necessary to uncover deeper insights and effectively guide environmental policy.
- By advancing our understanding of the watershed's dynamics and improving monitoring practices, we can better manage and preserve water quality, ensuring the health and safety of the community and the environment.

Acknowledgements

Alicia Leher
WRWC, Executive Director

Shiya Cao
My My Tran

Thanks!

Do you have any questions? Email us!

alehrer@wrwc.org
qchamblin@smith.edu
kcummings@smith.edu
jfranco@smith.edu
nhernandez@smith.edu

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**