Drivers of Cyanobacterial Blooms

**Explanation of Research Topic**

I focused on environmental data. Specifically, I analyzed water quality data pertaining to 20 reservoirs over the course of 31 years (1987-2018) published by the US EPA and authored by Smucker et. al.

The response variable that I focused on was cyanobacteria density (cells/mL) and the regressors that I focused on were nutrient density (units vary depending on the nutrient) and surface temperature (°C).

Cyanobacteria blooms can be very harmful to local wildlife, the surrounding ecosystem, and even human health, which is why I am interested in studying them. By analyzing this dataset, I hope to glean insights into what the major factors are in causing cyanobacteria blooms so that they can be combatted effectively.

**Data Collection/Data Source**

The data was collected by the United States Army Corps of Engineers (USACE) from 20 reservoirs which were built by USACE between 1953 and 1983. USACE monitored these reservoirs intermittently starting in 1987 by collecting water samples and phytoplankton samples using Kemmerer samplers and Van Dorn samplers. American Public Health Association standard methods were used to determine cyanobacteria cell densities.

In addition to the data that I analyzed (cyanobacteria density, total phosphorous, dissolved phosphorous, total Kjeldahl nitrogen, ammonia, nitrates and nitrites, total organic carbon, May surface temperature, June surface temperature, July surface temperature, and August surface temperature), Smucker et. al. worked with cyanobacteria species, deep water temperatures, surrounding landcover type, and more.

**Method of Analysis**

I believe that cyanobacteria density is positively correlated to both nutrient density (of all the nutrient data collected) and surface temperature (from all months collected). This is indirectly supported by Smucker et. al.’s conclusions, which were that cyanobacteria density increased more severely in watersheds surrounded by extensive agriculture (and consequently had higher nutrient runoff) and that cyanobacteria density increases coincided with earlier and longer summers, meaning higher surface temperatures. Therefore, I included nutrient densities and surface temperatures in my analyses.

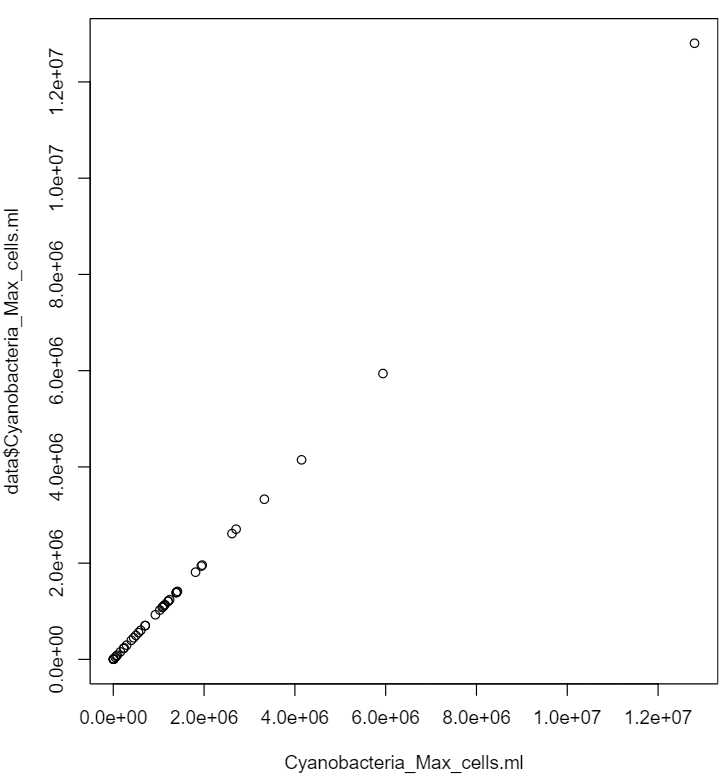
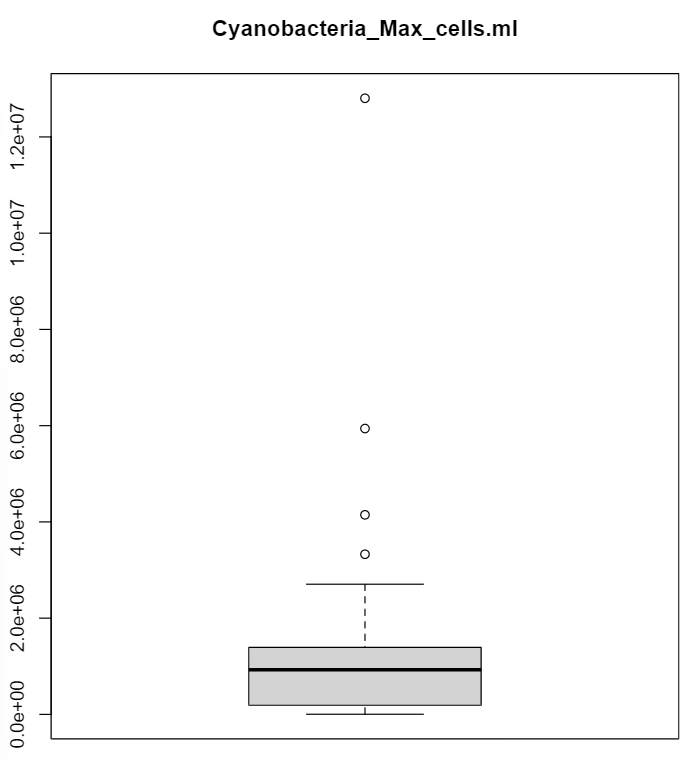
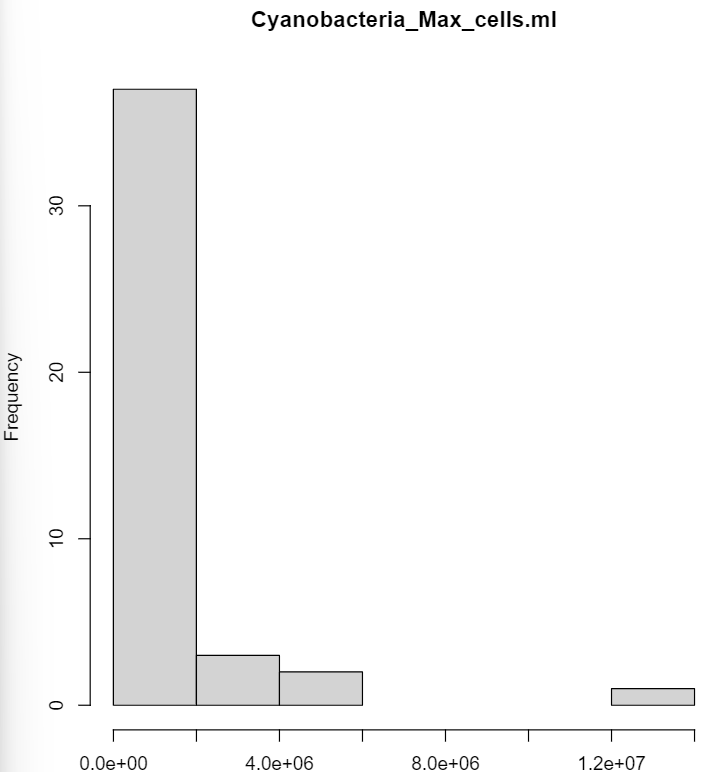
In order to check my model’s validity, I analyzed its residual plots, checked for multicollinearity, outliers, and influential observations.

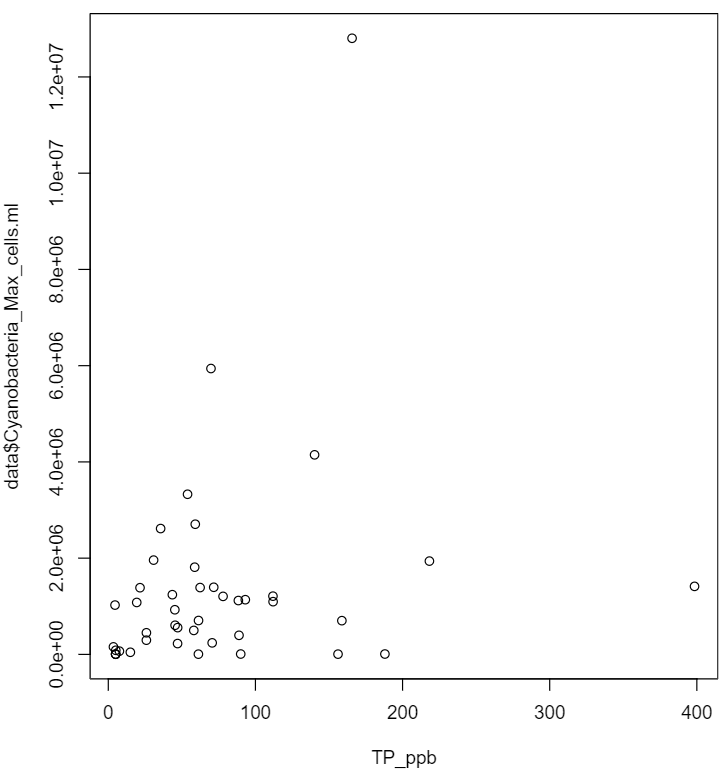
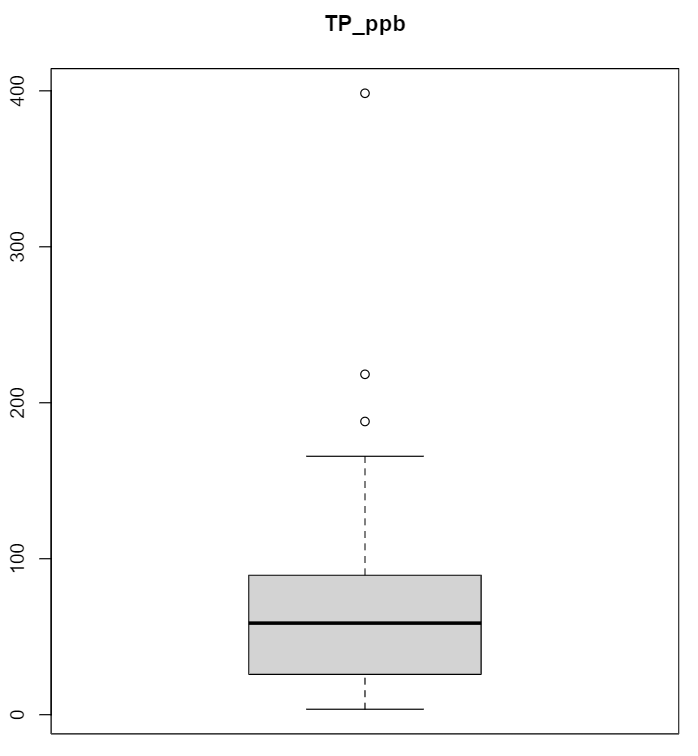
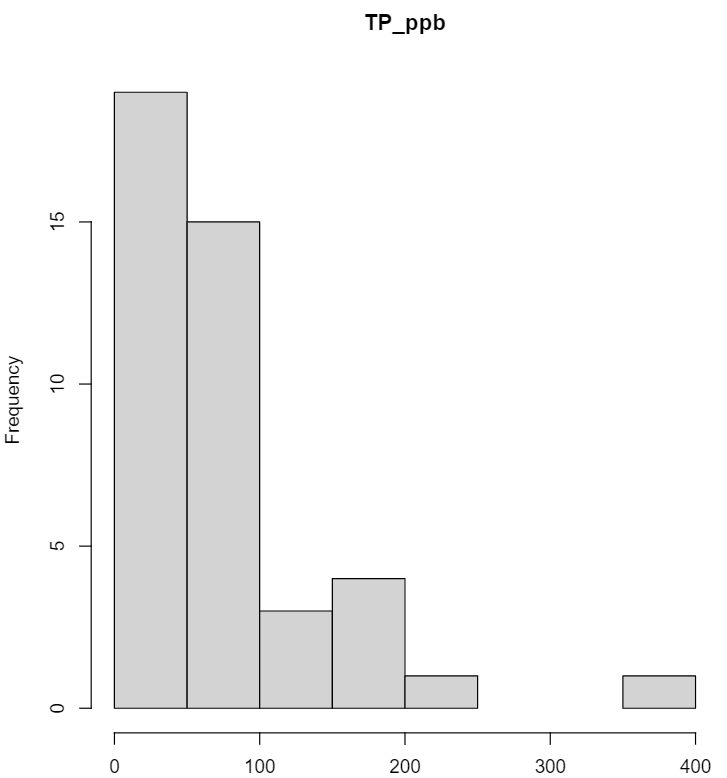
**Results**

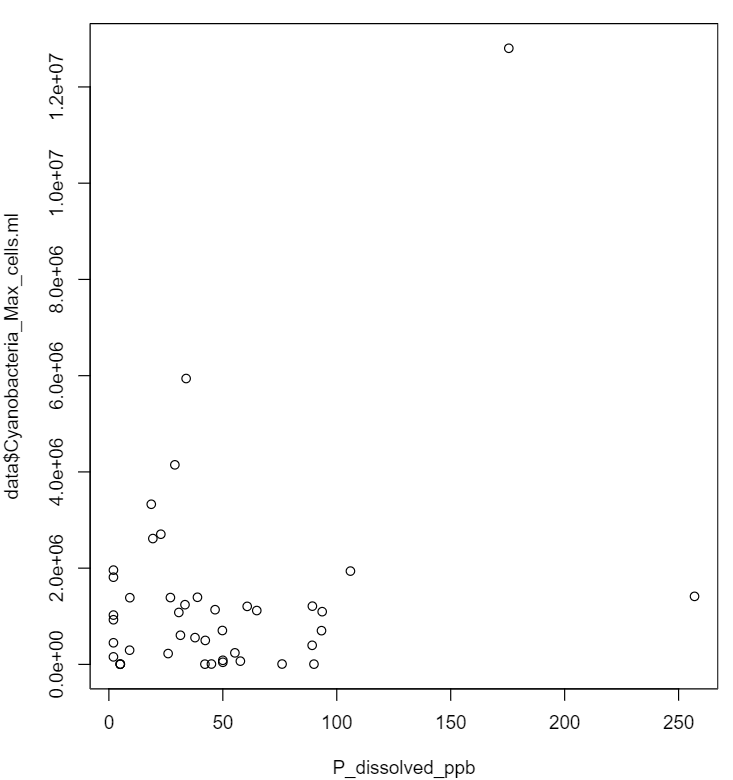
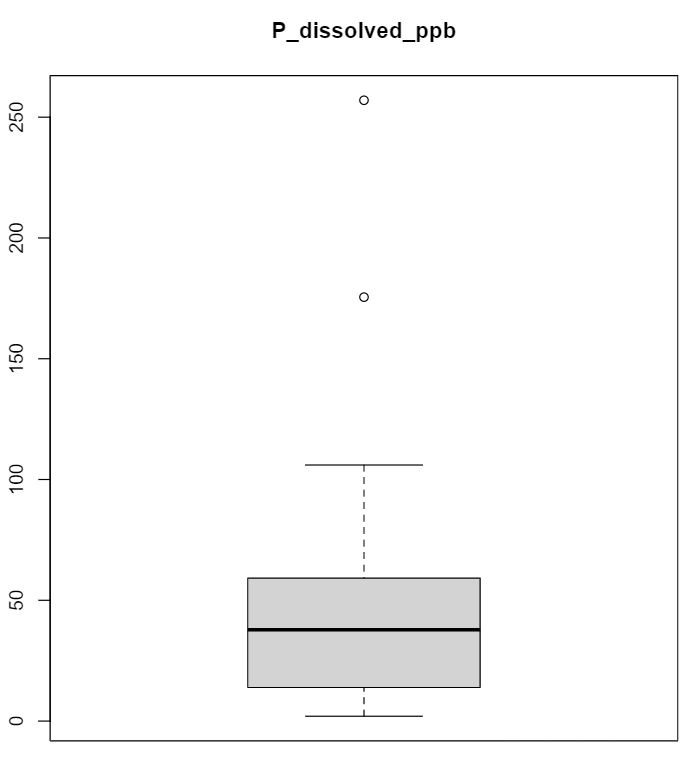
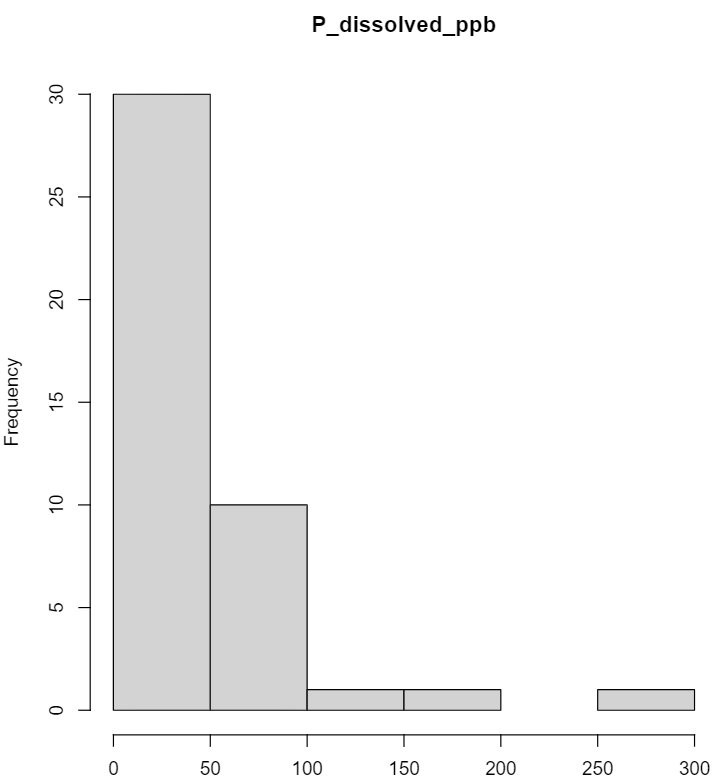
After standardizing the regressors that I was working with, I found the appropriate model to be:

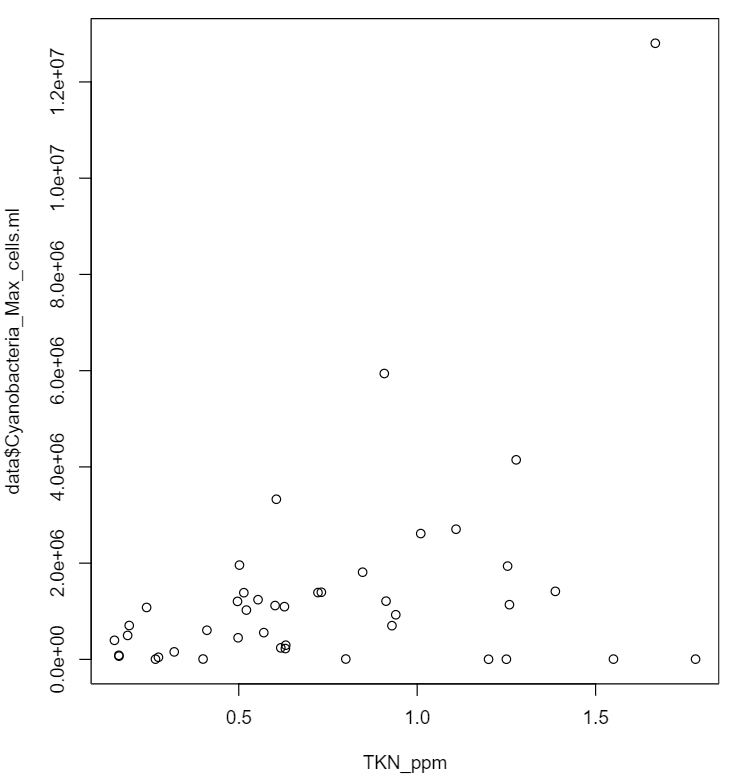
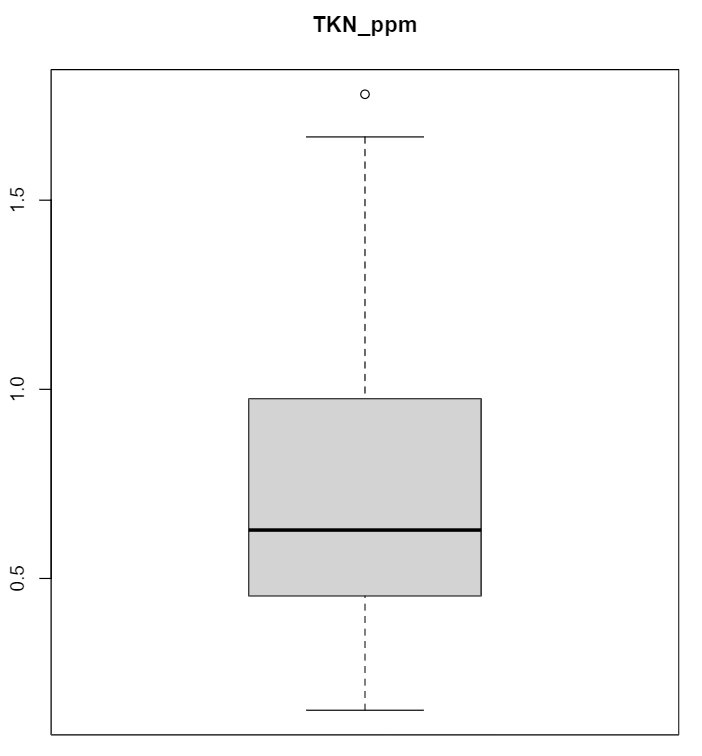
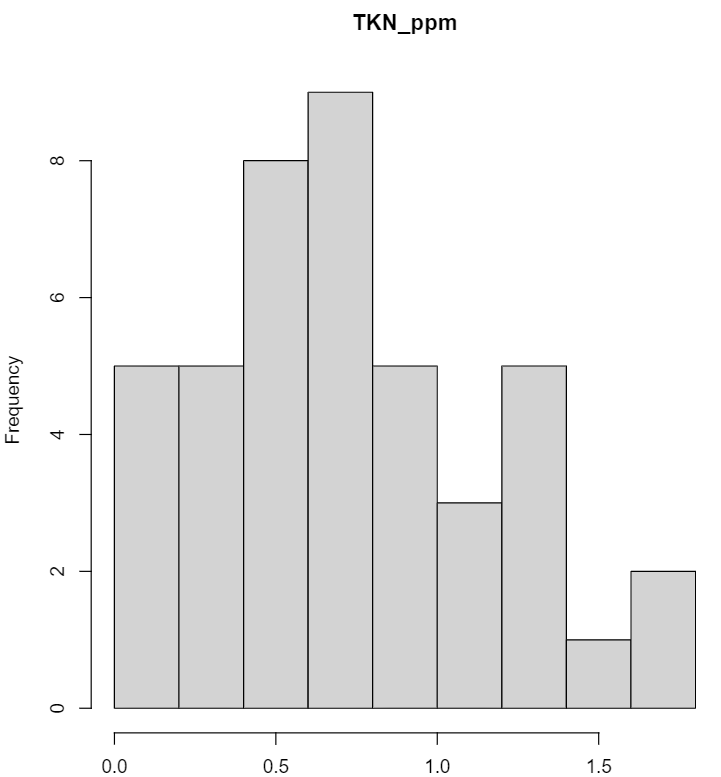
Where

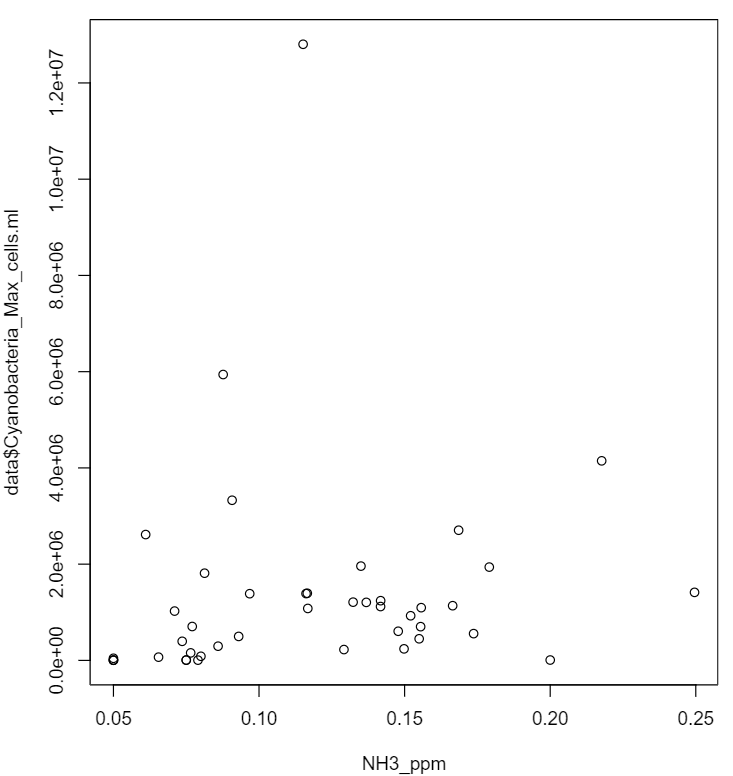
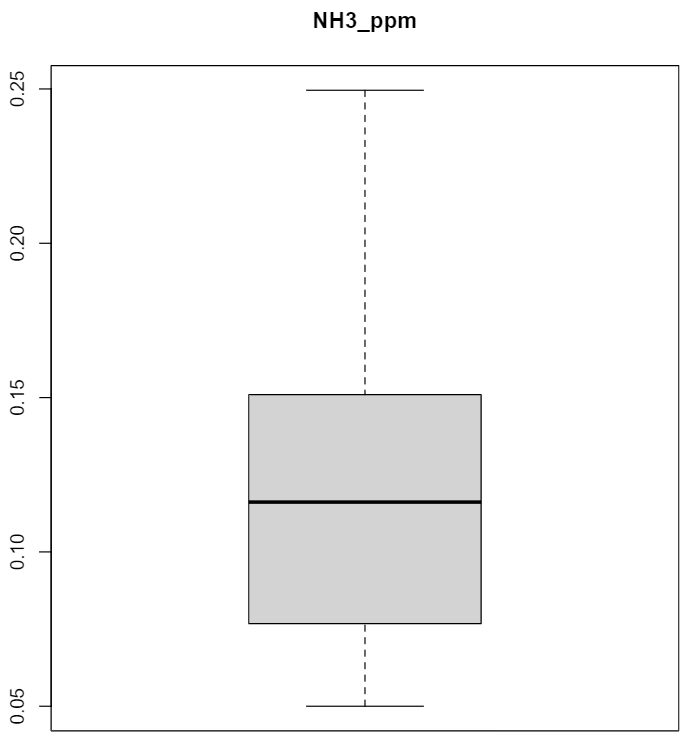
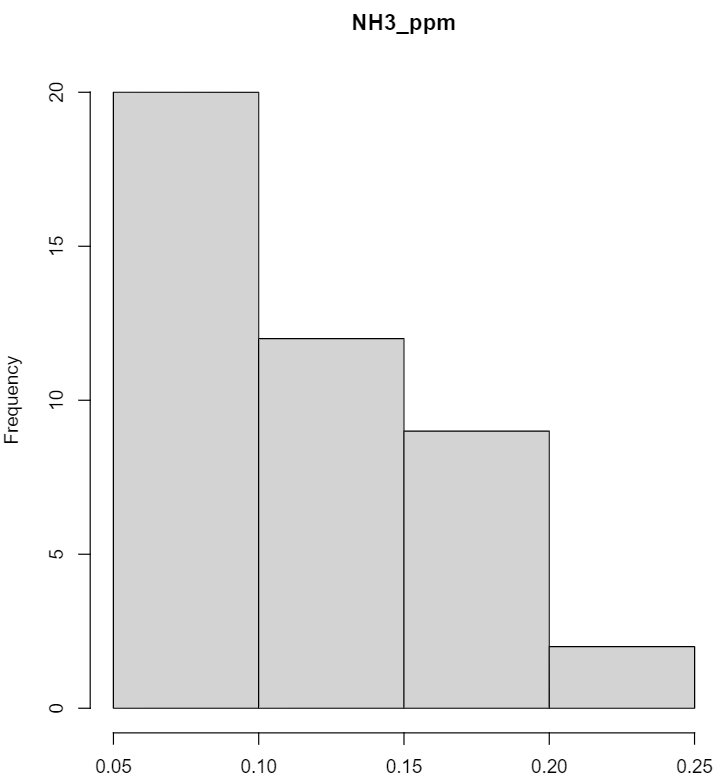
Descriptive graphs of each regressor below:

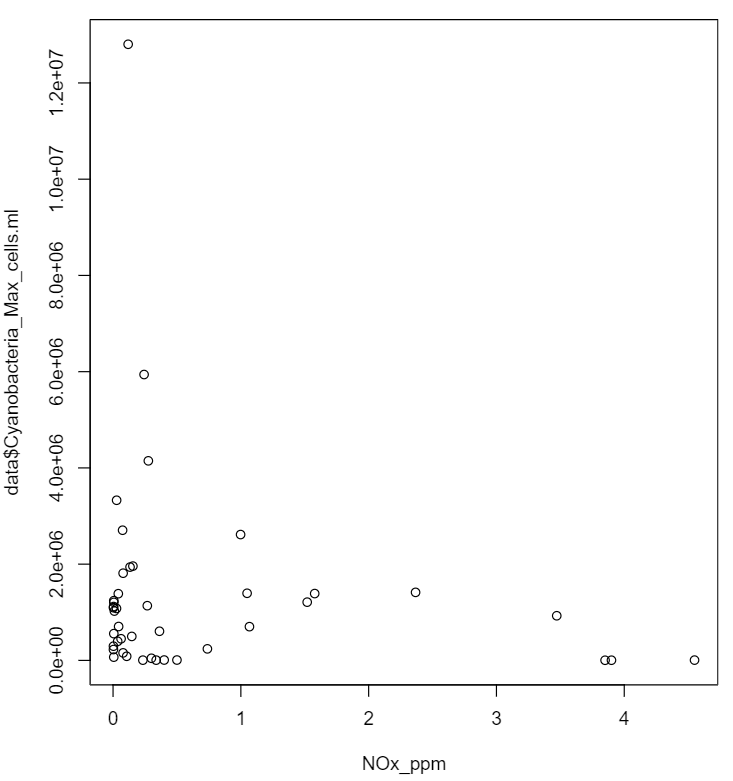
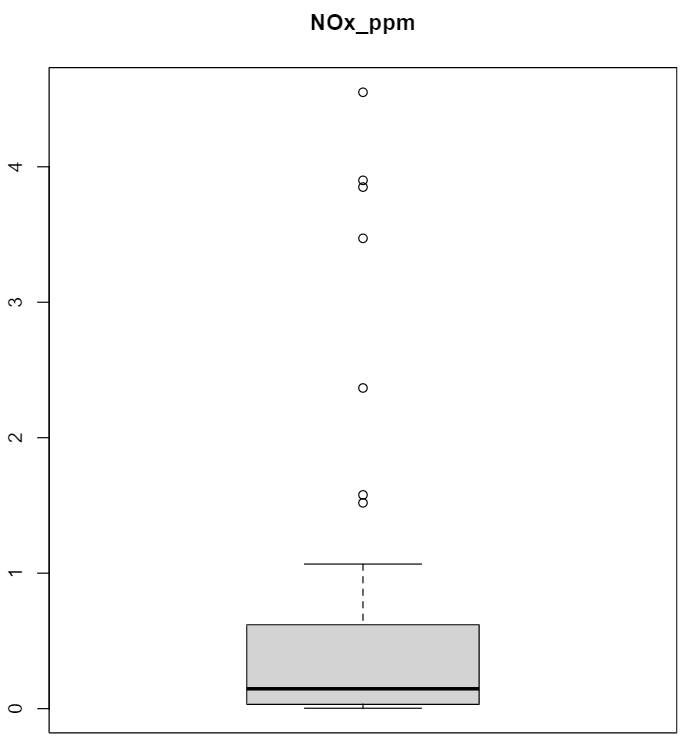
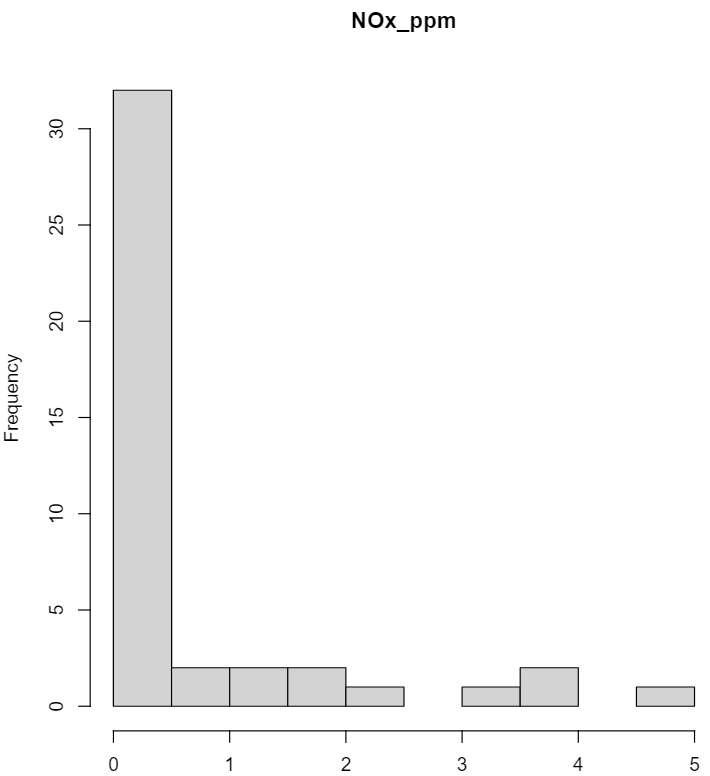


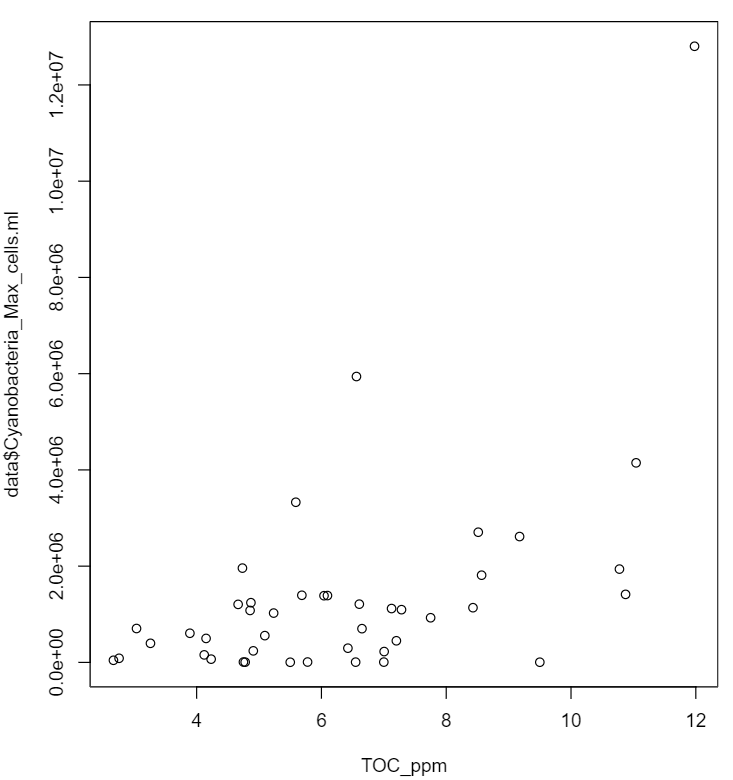
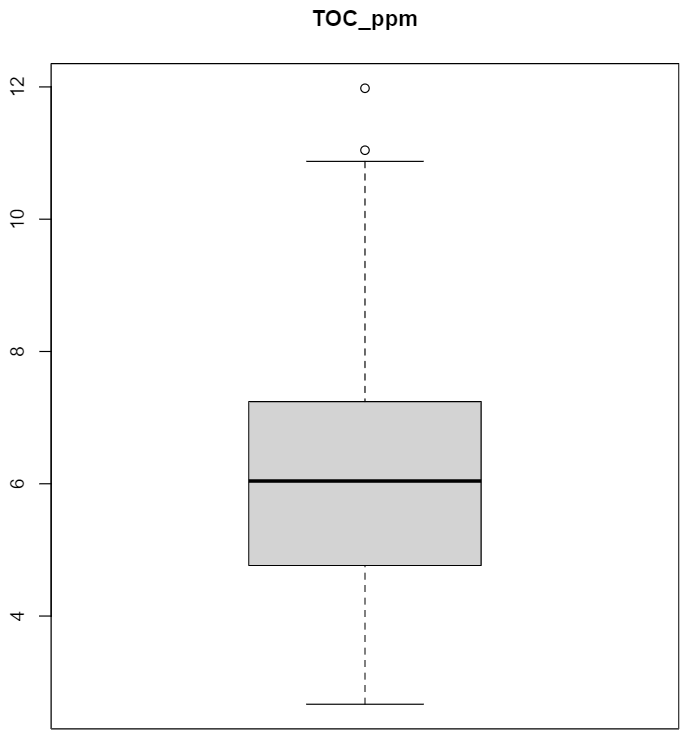
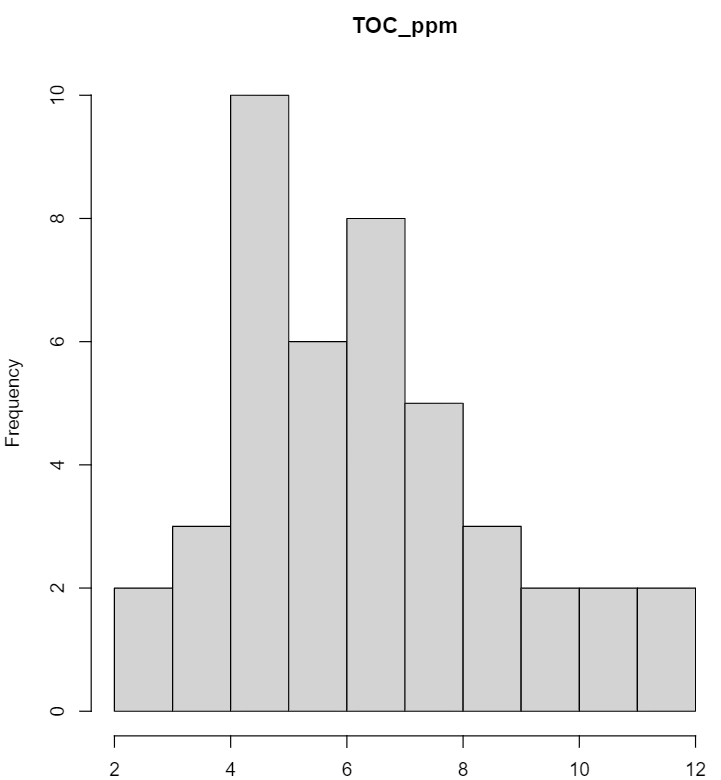


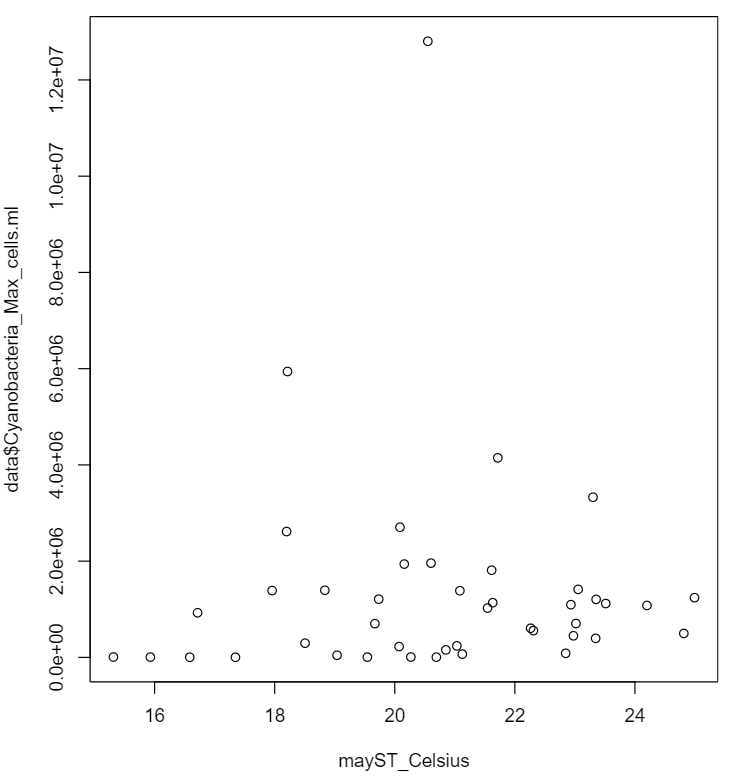
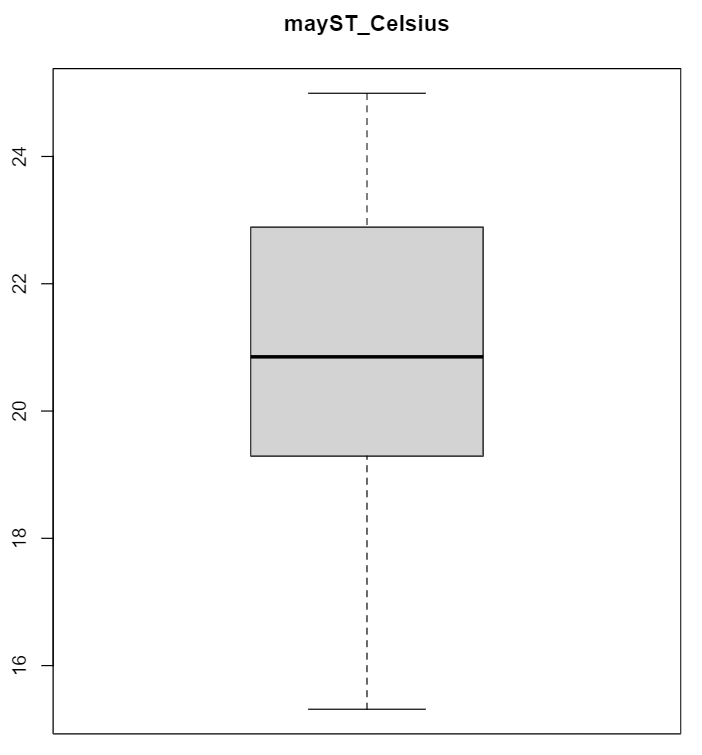
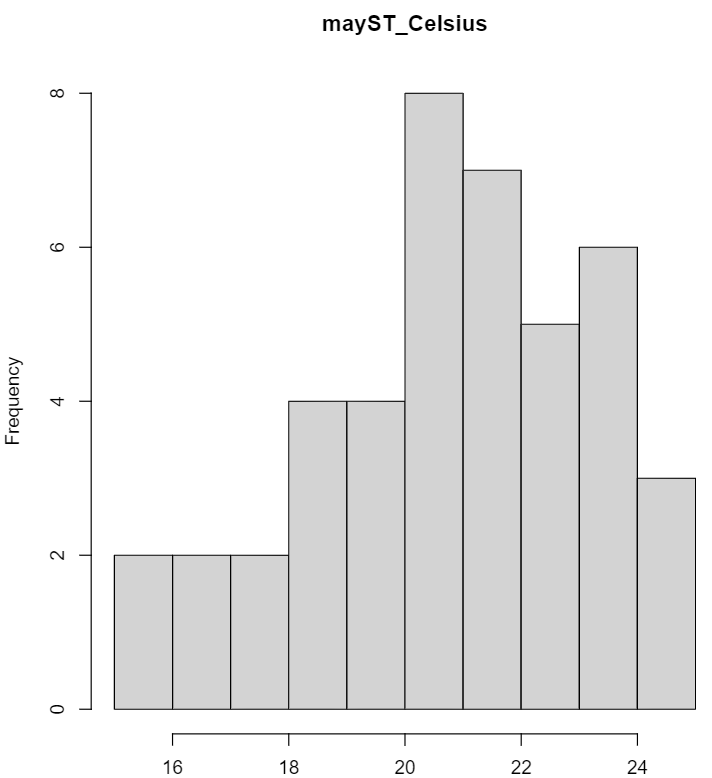


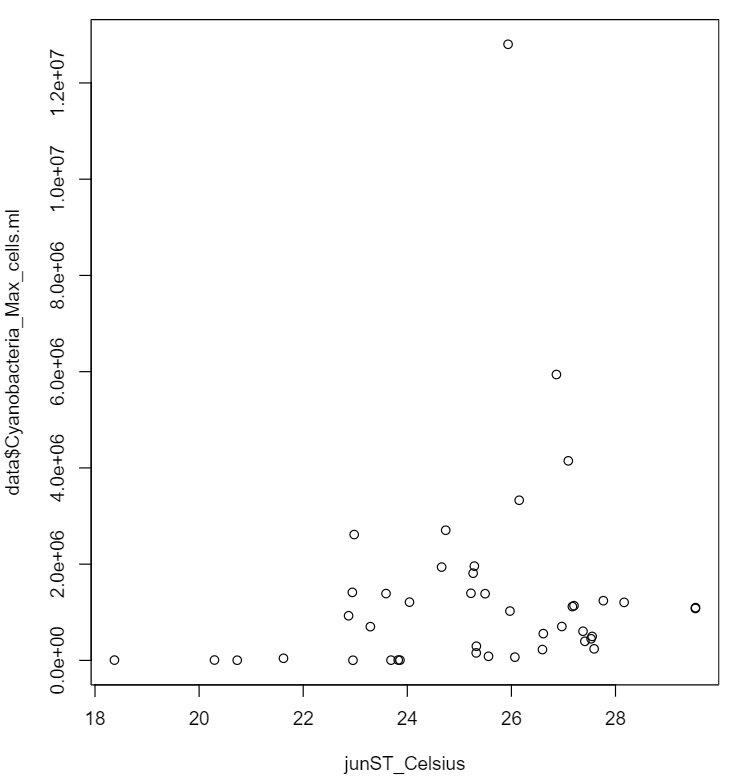
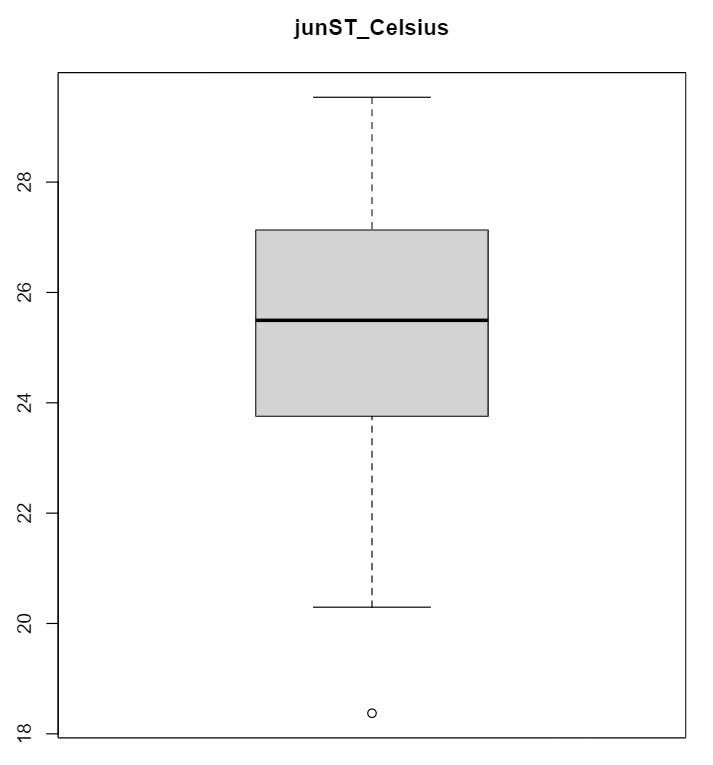
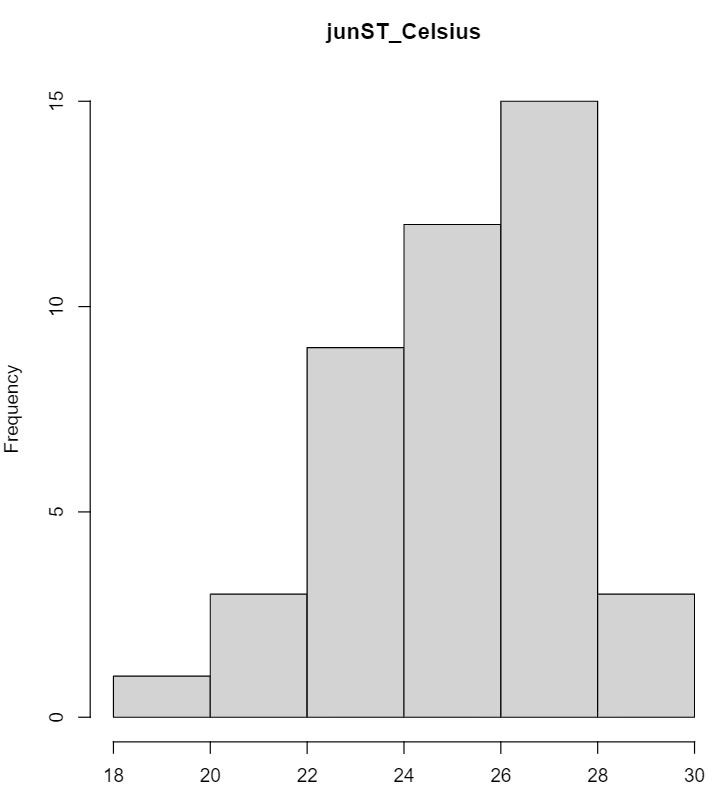


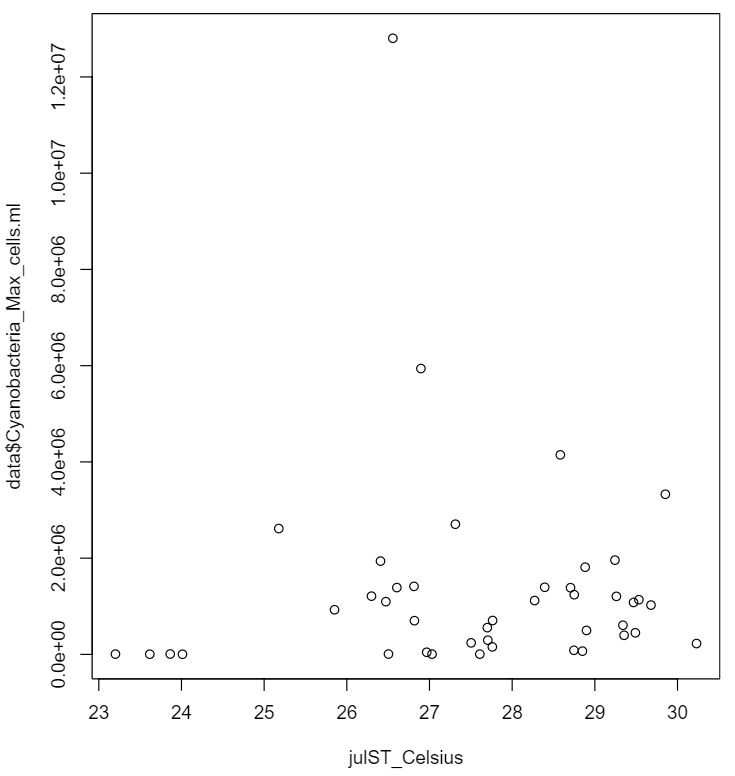
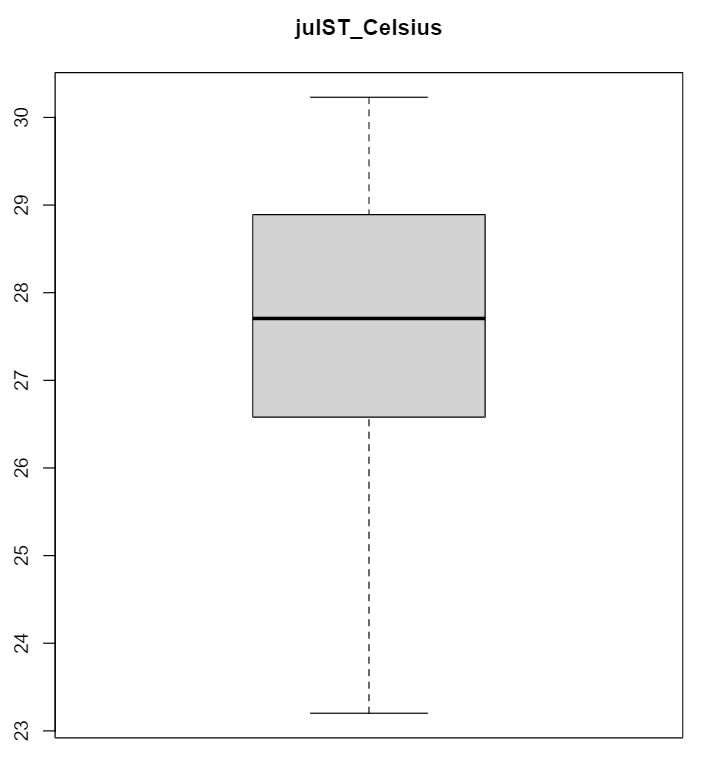
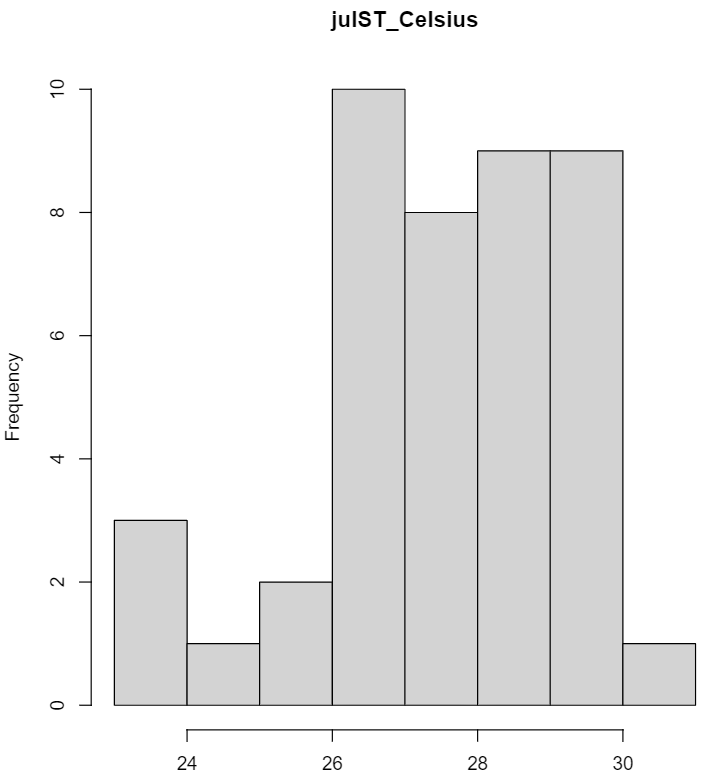


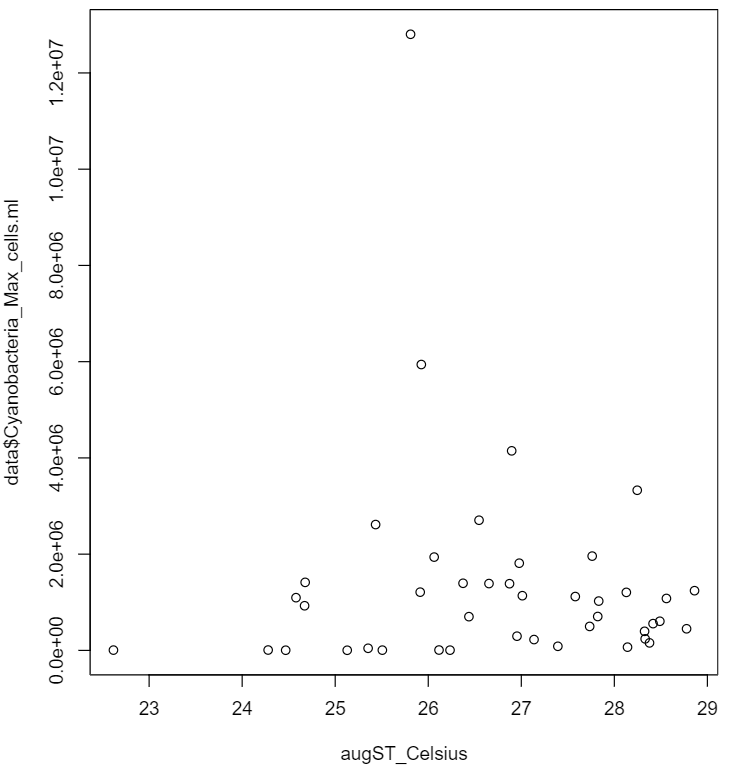
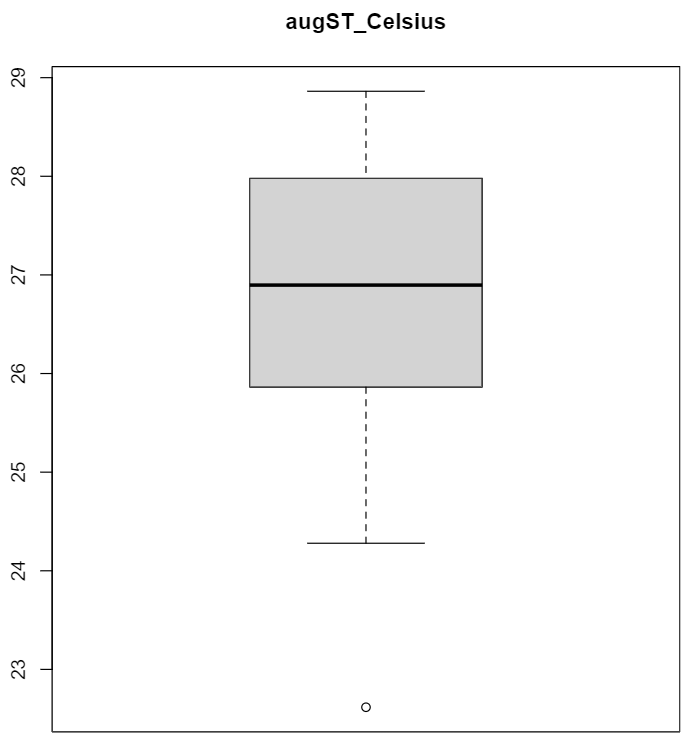
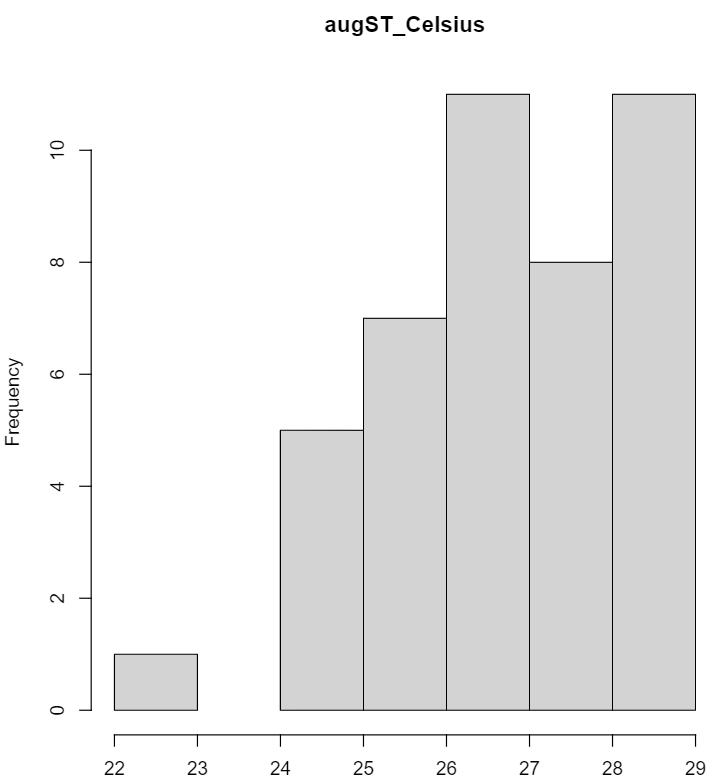




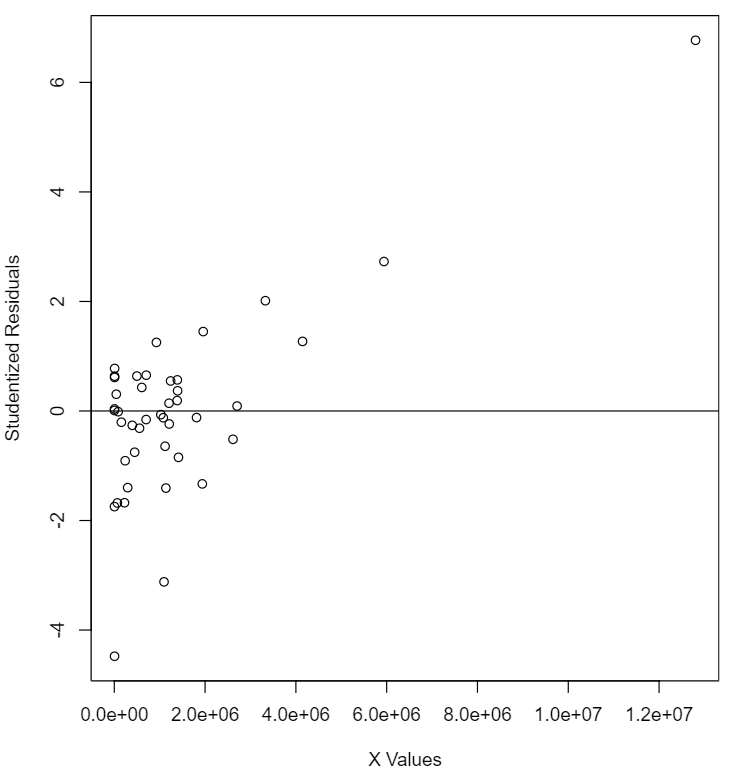
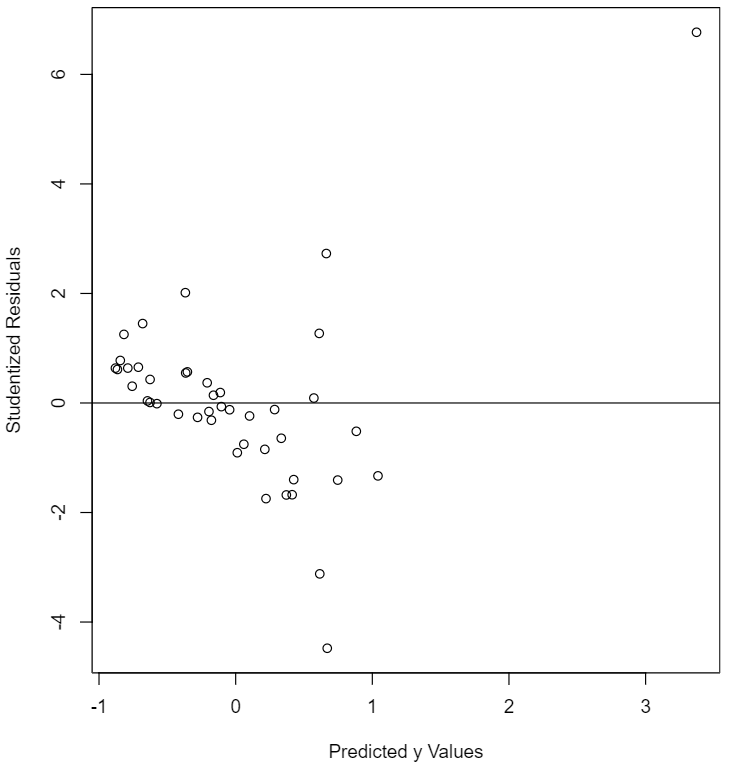
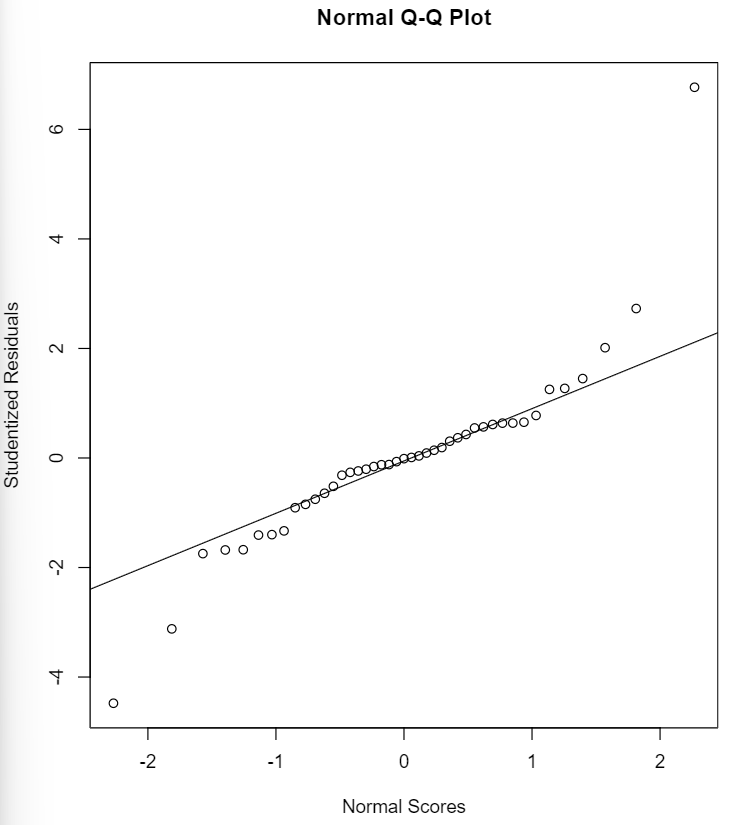






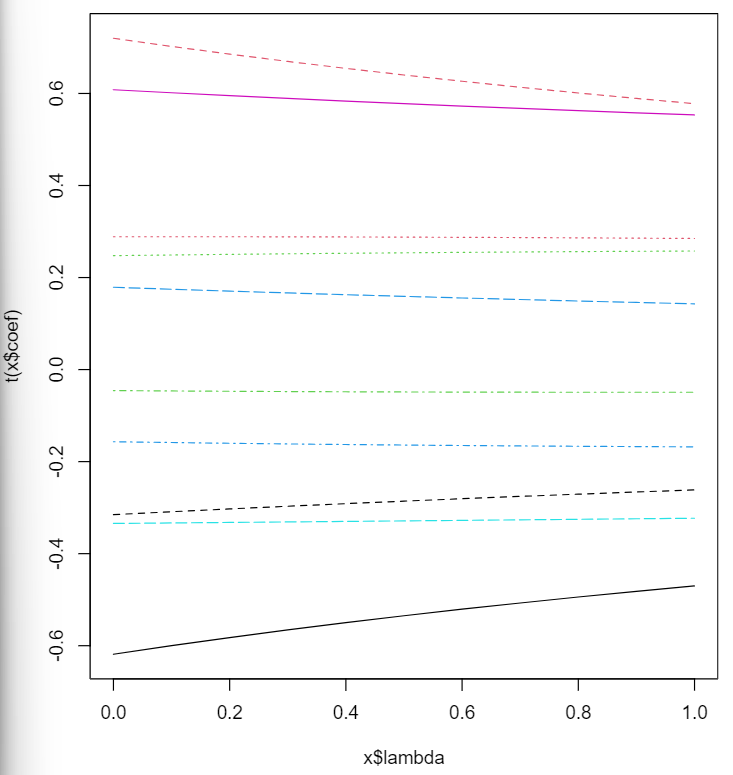


Residual plots of the model shown below:



None of the residual plots suggest issues, so no remedial measures were taken here.

Furthermore, the model was analyzed for multicollinearity and it was unclear whether it was an issue. However, the ridge trace (shown below) appears stable, so no ridge regression was performed in remediation.



**Conclusions**

It appears that cyanobacteria density is positively correlated to dissolved phosphorous, total Kjeldahl nitrogen, total organic carbon, June surface temperature, and August surface temperature. On the other hand, cyanobacteria density appears to be negatively correlated to the other regressors, which is somewhat surprising.

**References**

“1987-2018 Cyanobacteria and Water Quality Data for 20 Reservoirs.” *Catalog*, Publisher U.S. EPA Office of Research and Development (ORD), 22 May 2021, catalog.data.gov/dataset/1987-2018-cyanobacteria-and-water-quality-data-for-20-reservoirs.

Smucker, Nathan J., et al. “Increasingly severe cyanobacterial blooms and deep water hypoxia coincide with warming water temperatures in reservoirs.” *Global Change Biology*, vol. 27, no. 11, 2021, pp. 2507–2519, https://doi.org/10.1111/gcb.15618.

**Appendices**

Data set attached.

R program attached.