Supplementary Materials: Plastic in the Urban Freshwater Environment

An Exploratory Study of Microplastics in the Upper Oconee Watershed via Community-Based Efforts

Amelia M. Foley

2021-12-08

# 1 Documentation for Supplementary Files

Code and data files required to reproduce this analysis are available on [GitHub](https://github.com/ameliafoley/AmeliaFoley-MADA-project)

* “**AmeliaFoley-MADA-project.Rproj**” Establishes relative file paths
* “**README.md**” Provides brief order of scripts for reproducing; summarizes project parts
* “**code**” folder contains all code for processing and analysis
  + “**analysis\_code**”
    - analysis.Rmd performs linear regression on single predictors and full predictor sets
    - exploration.Rmd contains the exploratory data analysis
    - land\_model.Rmd contains statistical models on land cover data
    - modeling.Rmd is the main analysis file containing final models (LASSO)
    - subset\_modeling.Rmd is an additional analysis file performing models on subsets of data; it is not used for the manuscript
  + “**processing\_code**”
    - processing\_code.Rmd is the master script for all processing. It combines all of the scripts from below, which have been moved to the ARCHIVE folder.
      * censusland\_code.Rmd combines land cover data with microplastic data
      * censusprocessing.Rmd imports census tract population data
      * mpprocessing.R processes the original microplastic data file and saves cleaned data
      * uownprocessing.Rmd imports additional water quality data from UOWN.org
      * wrfprocessing.Rmd imports distance for nearest water reclamation facility
    - map\_images.Rmd combines individual map images to produce a grid
  + “**readme.md**” provides an order and brief description of dependencies/outputs for processing code files
* “**data**” folder contains all raw and processed data
* “**products**” folder contains files to produce the Manuscript and Supplement
  + “**manuscript**” contains files for producing the manuscript
    - apa.csl optional format for citations
    - elsevier.csl optional format for citations
    - library.bib contains references
    - Manuscript.docx result of knitting the markdown
    - Manuscript.Rmd creates basic manuscript word document
    - style\_template.docx placeholder to add special formatting to manuscript
    - Supplement.Rmd creates basic supplemental material word document
    - Supplement.docx result of knitting markdown file
* “**results**” folder contains all plots, tables, figures produced in data analysis

# 2 Reproducing results

Reproducing this project requires R, RStudio, and Microsoft Word. Files should be run in the following order.

In the code > processing\_code folder:

1. processing\_code.Rmd

In the code > analysis\_code folder:

1. analysis.Rmd
2. modeling.Rmd
3. land\_model.Rmd

In the code > processing\_code folder:

1. map\_images.Rmd

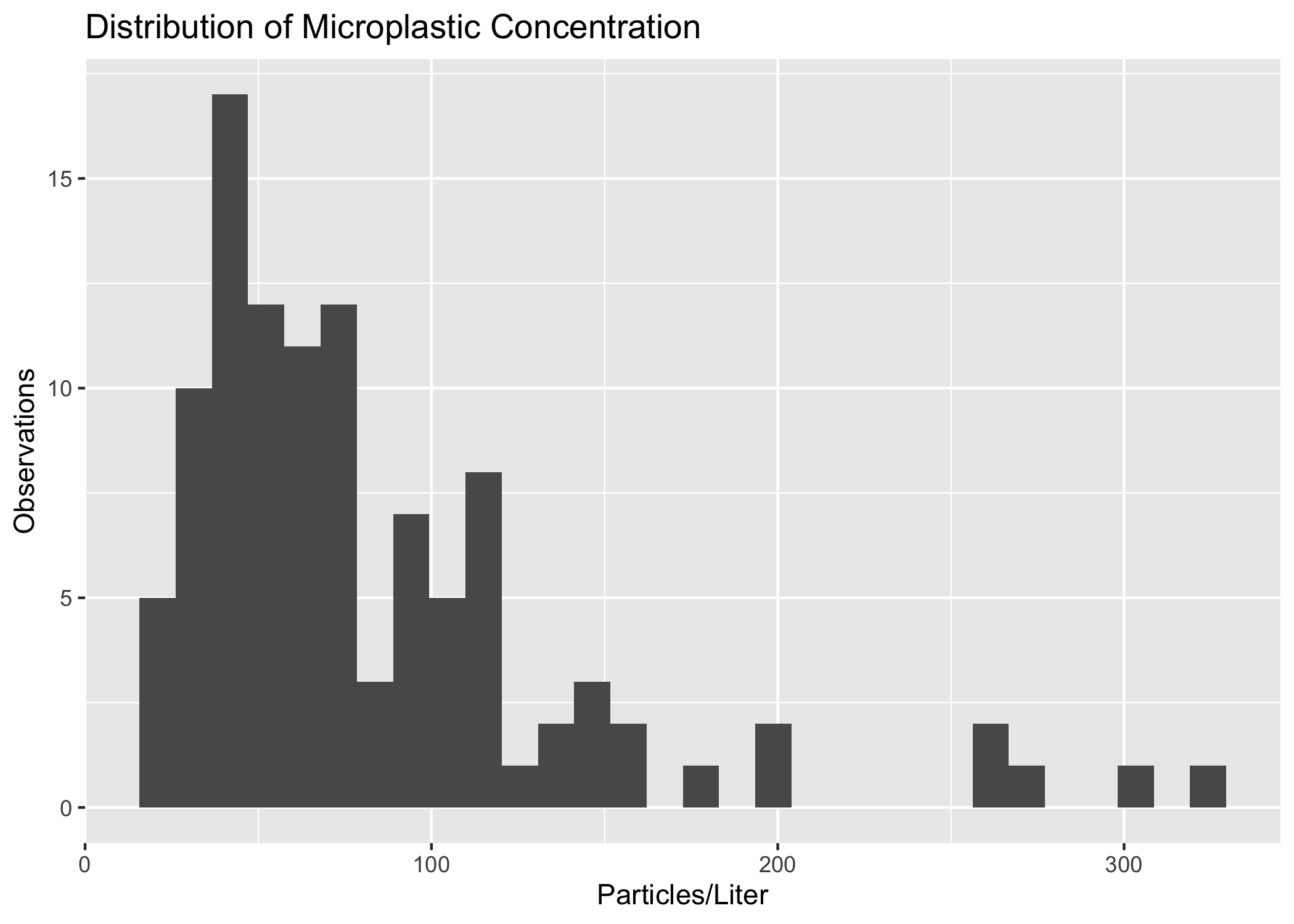
In the products folder

1. Manuscript.Rmd
2. Supplement.Rmd

# 3 Supplementary results

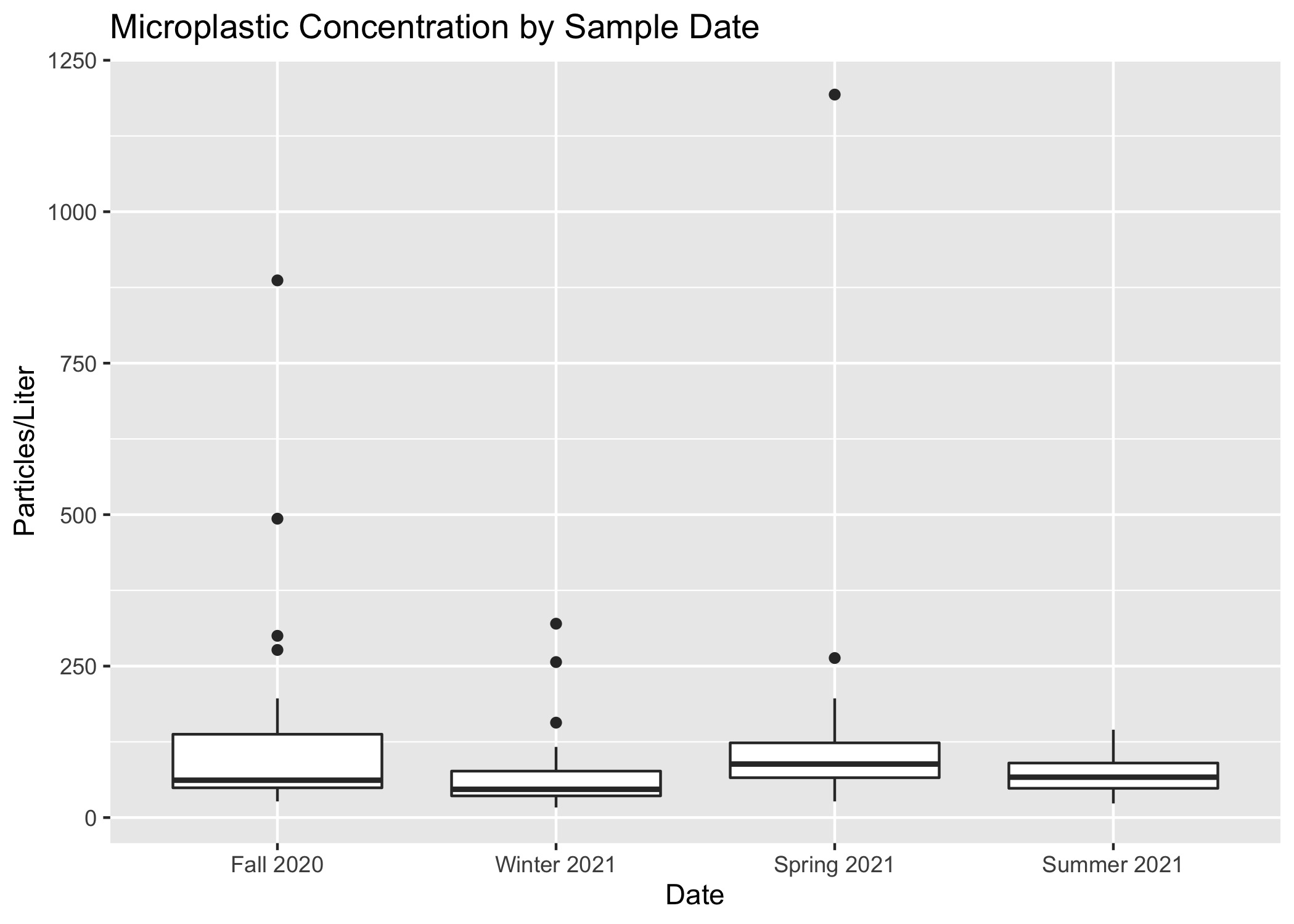
### 3.0.1 Distribution of Microplastic Concentration

Figure 1 shows a histogram of microplastic concentration observations. The minimum concentration is 16.67 particles/L and the maximum is 1193.33 particles/L. The mean concentration is 104.39 particles/L, and the median is 66.67 particles/L.



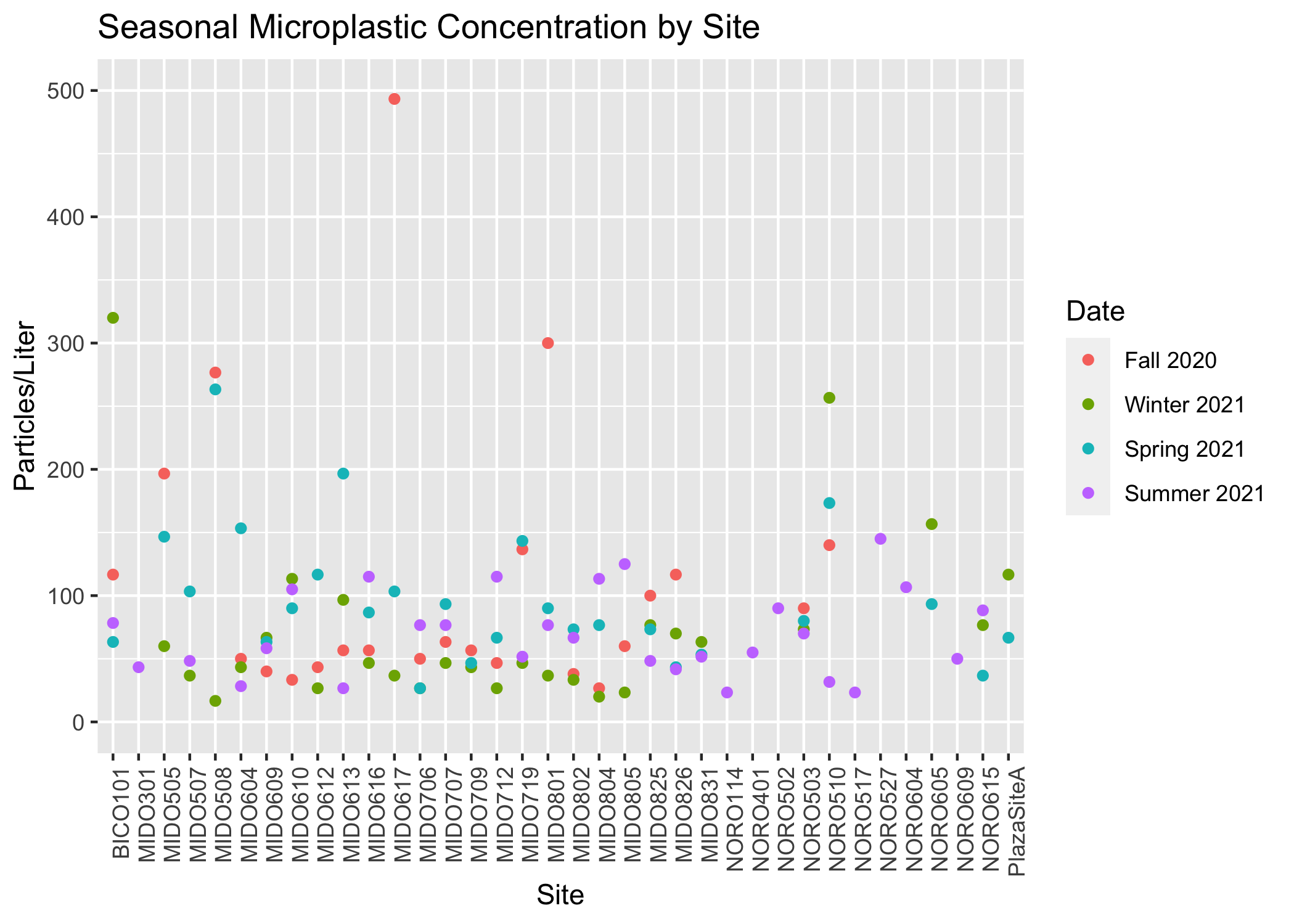
Distribution of Microplastic Concentration

Microplastic concentrations remained in similar ranges throughout the study period. Figure 2 shows a boxplot of concentrations by sample date.



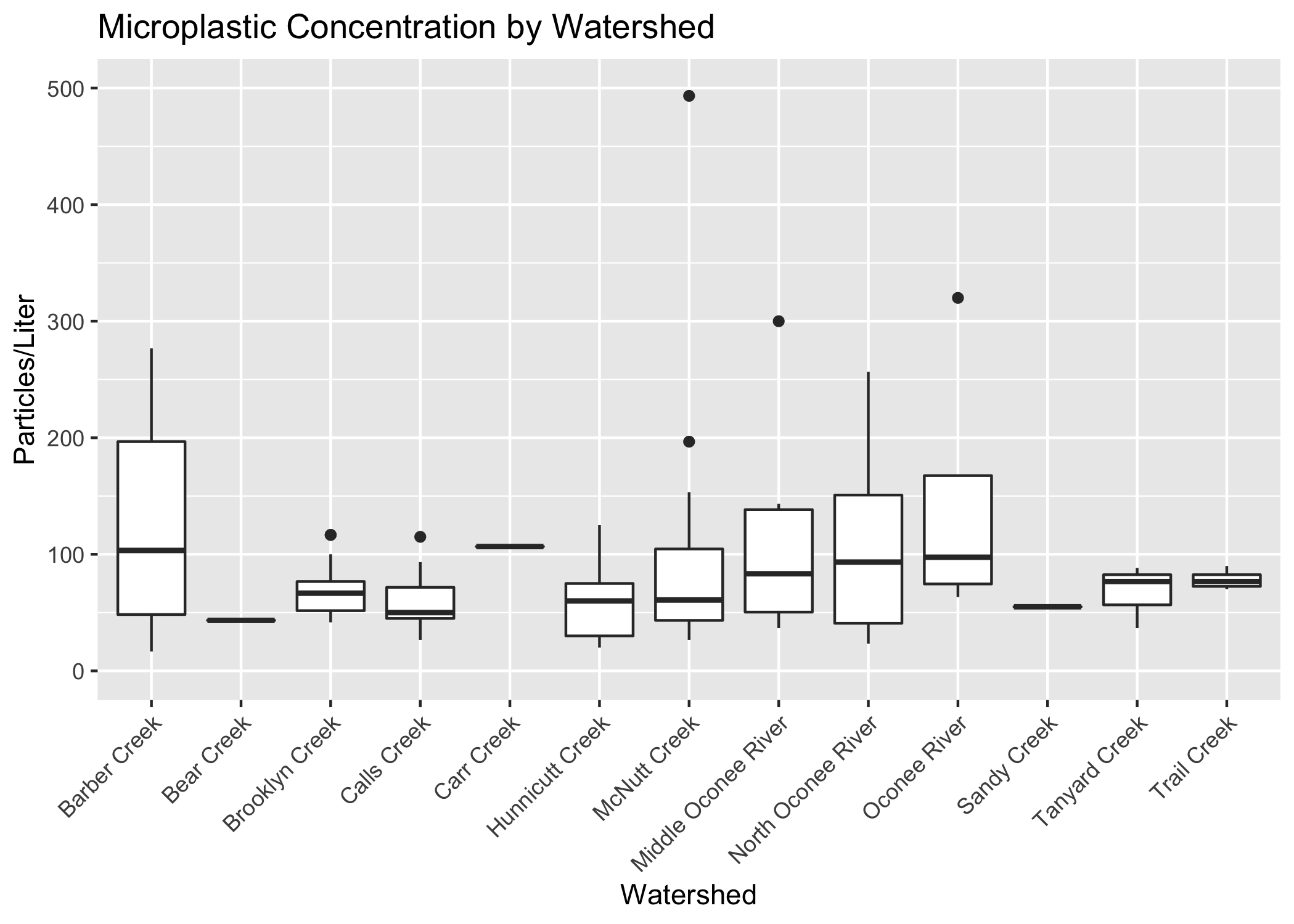
Particles/L by Sample Date

There is some seasonal variation in concentration at each individual site. Figure 3 shows a plot of concentrations at each site.



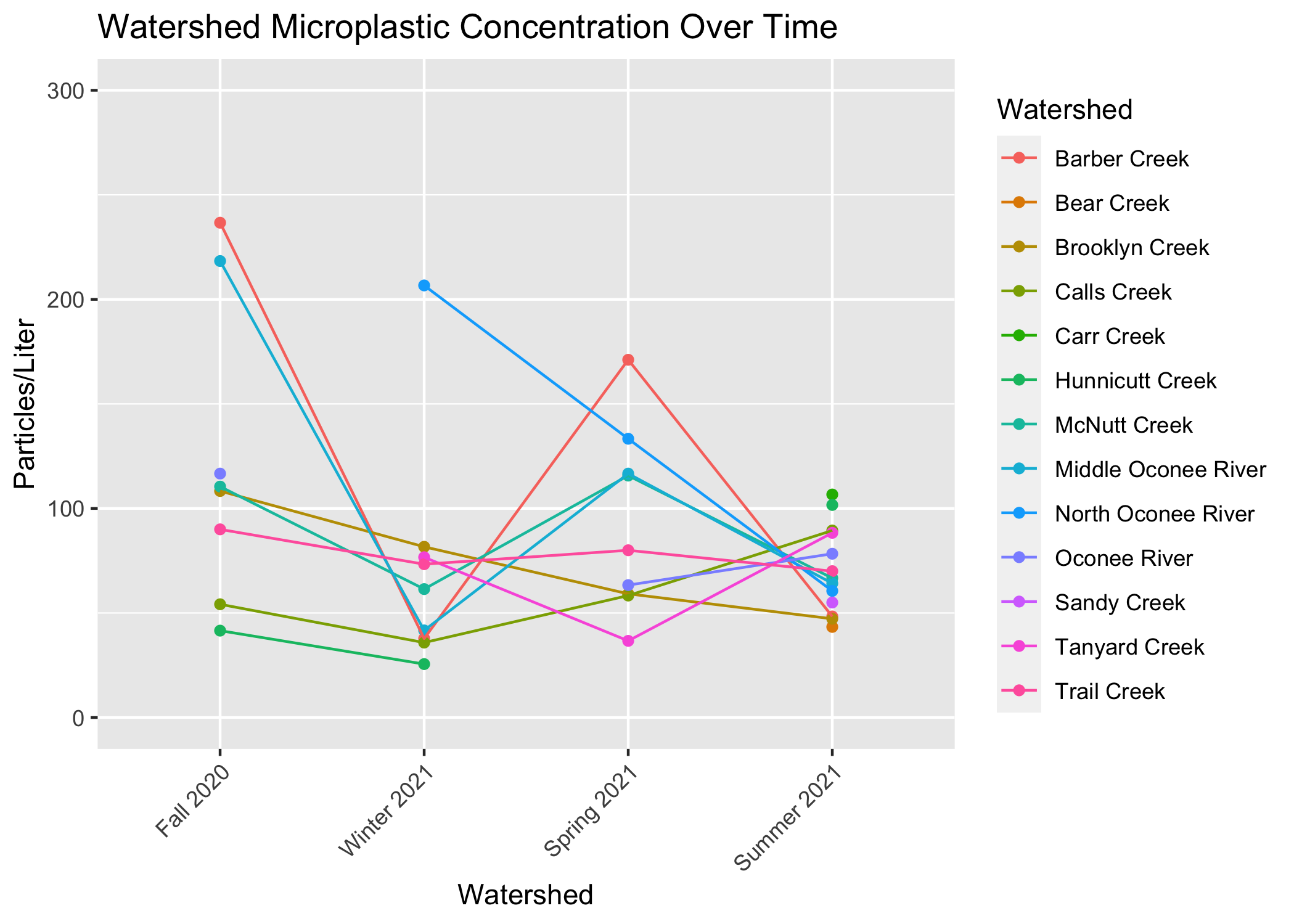
Seasonal Variation in Particles/L

There are similar microplastic levels throughout the watersheds within the Upper Oconee. Some watersheds experienced greater variation in microplastic levels than other watersheds. Figure 4 shows the microplastic concentrations by watershed.



Watershed Microplastic Concentrations

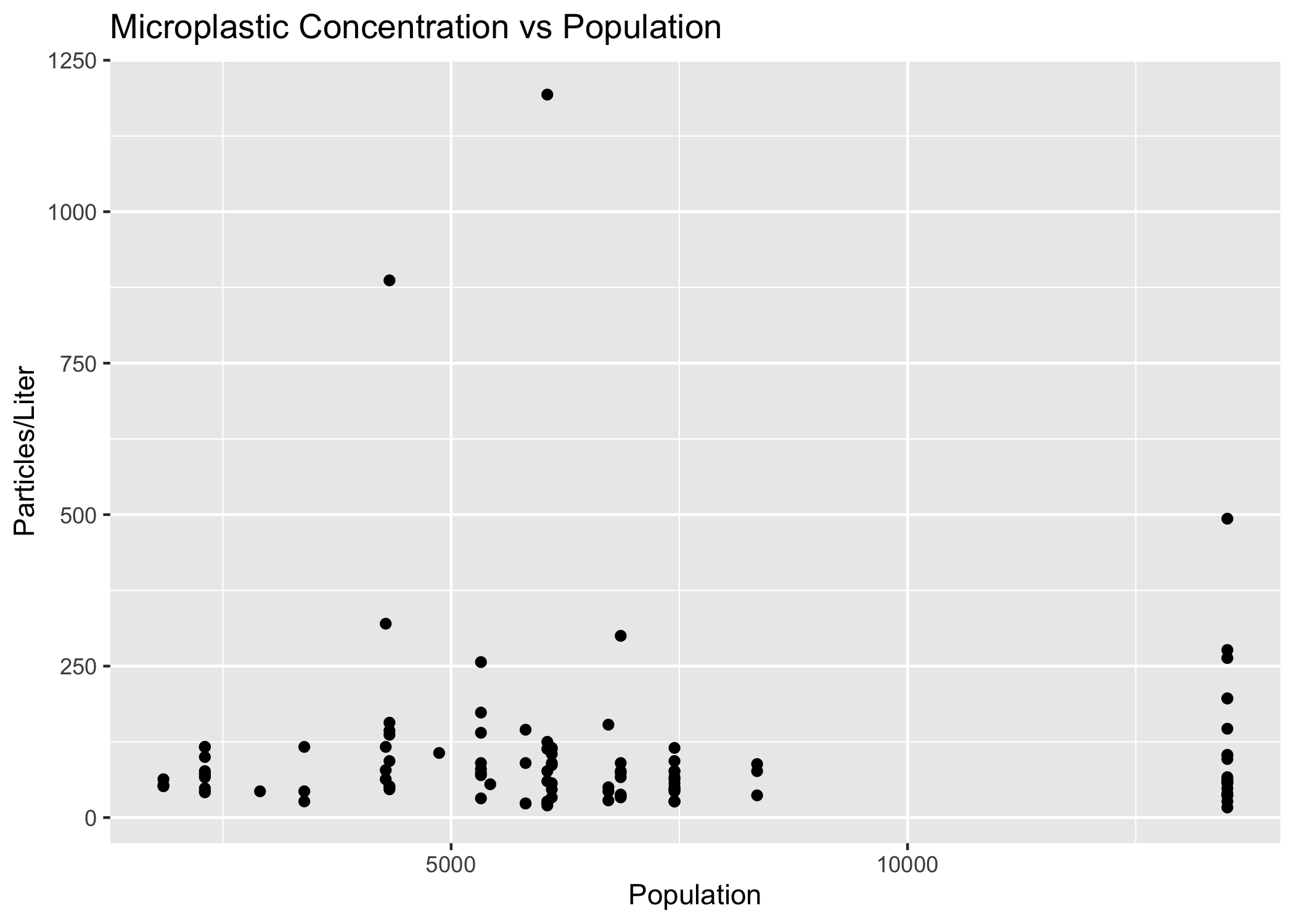
Figure 5 shows a line graph of the mean watershed microplastic concentrations at each seasonal sampling date.



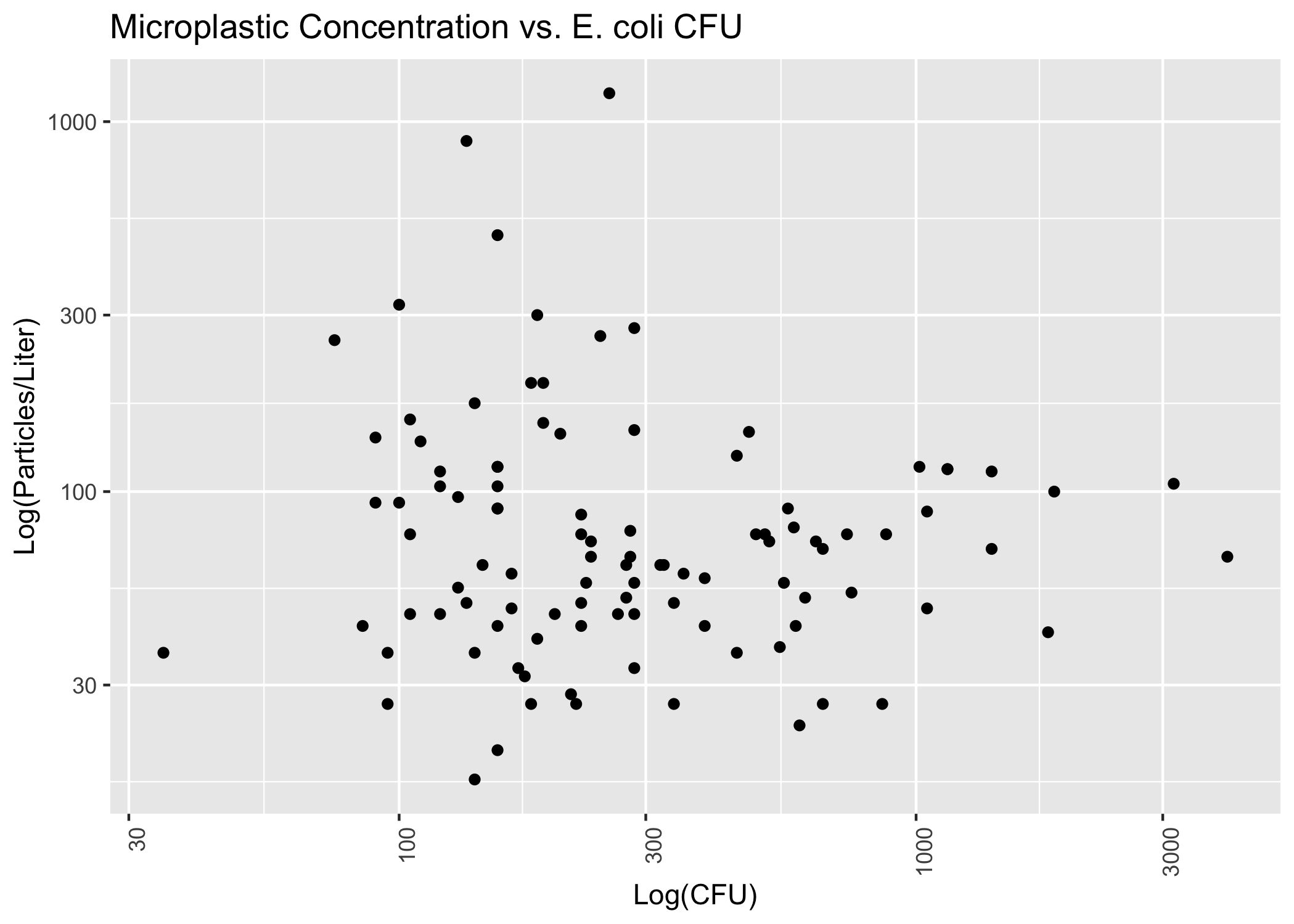
Microplastic Concentration Over Time

### 3.0.2 Predictors

Population, land cover/use, and bacteria levels are hypothesized predictors of microplastic concentration. Figure 6 and Figure 7 demonstrate the relationship between microplastic concentration and population and microplastic concentration and bacteria levels (CFU/100mL), respectively.

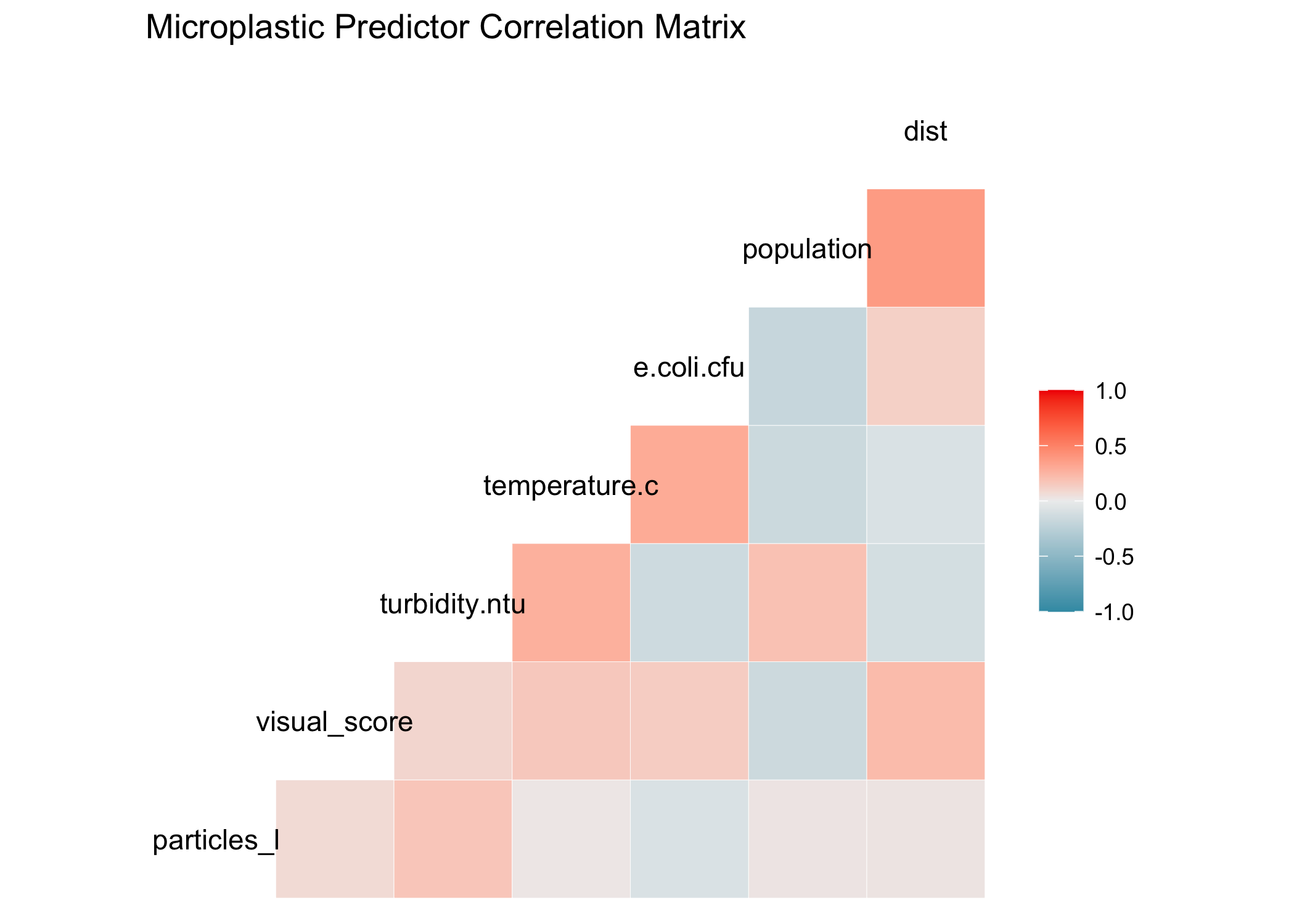


Particles/L vs Population

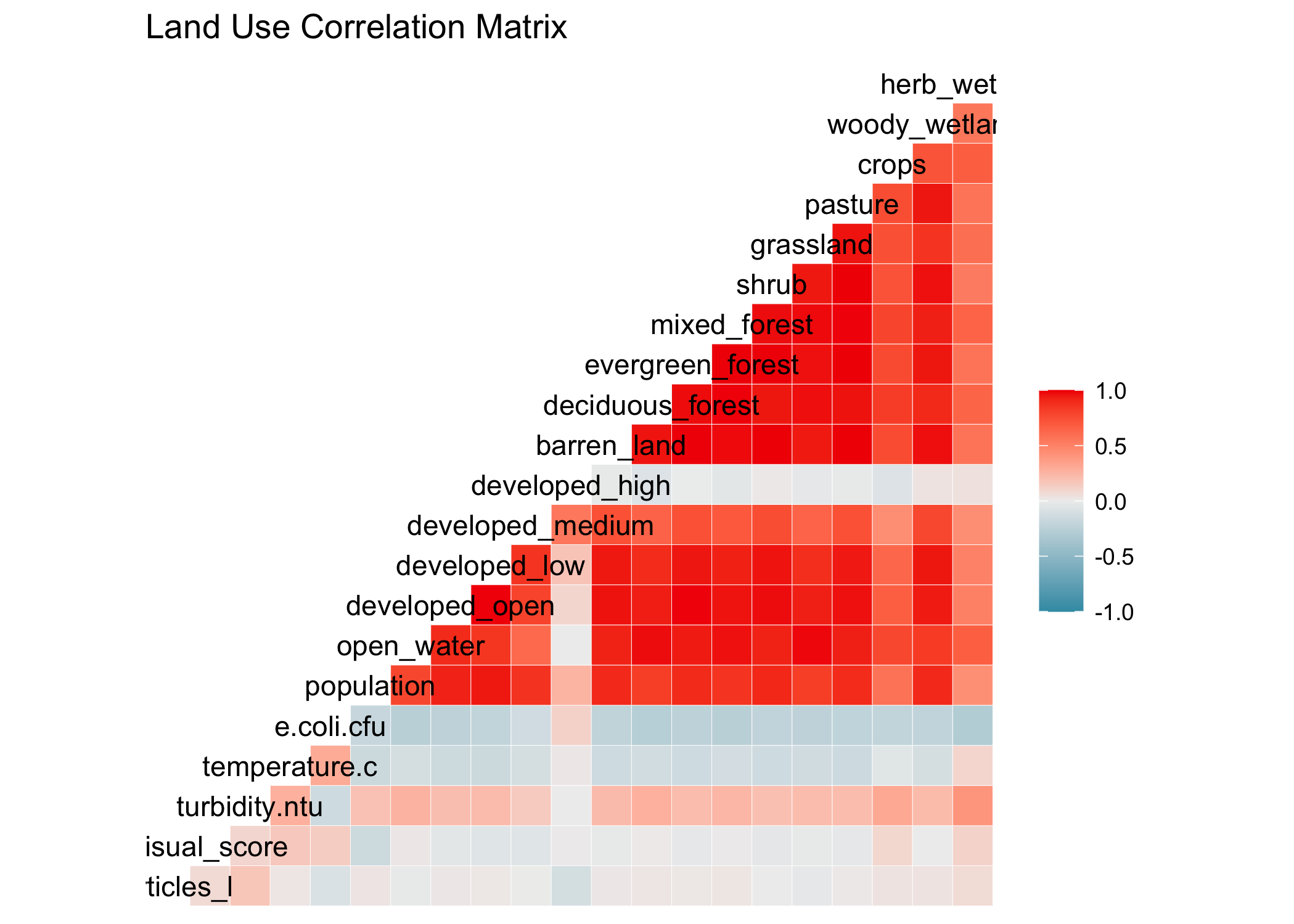


Log particles/L vs CFU

Figure 8 and Figure 9 show correlation matrices for the hypothesized predictor and for the different categories of land use.



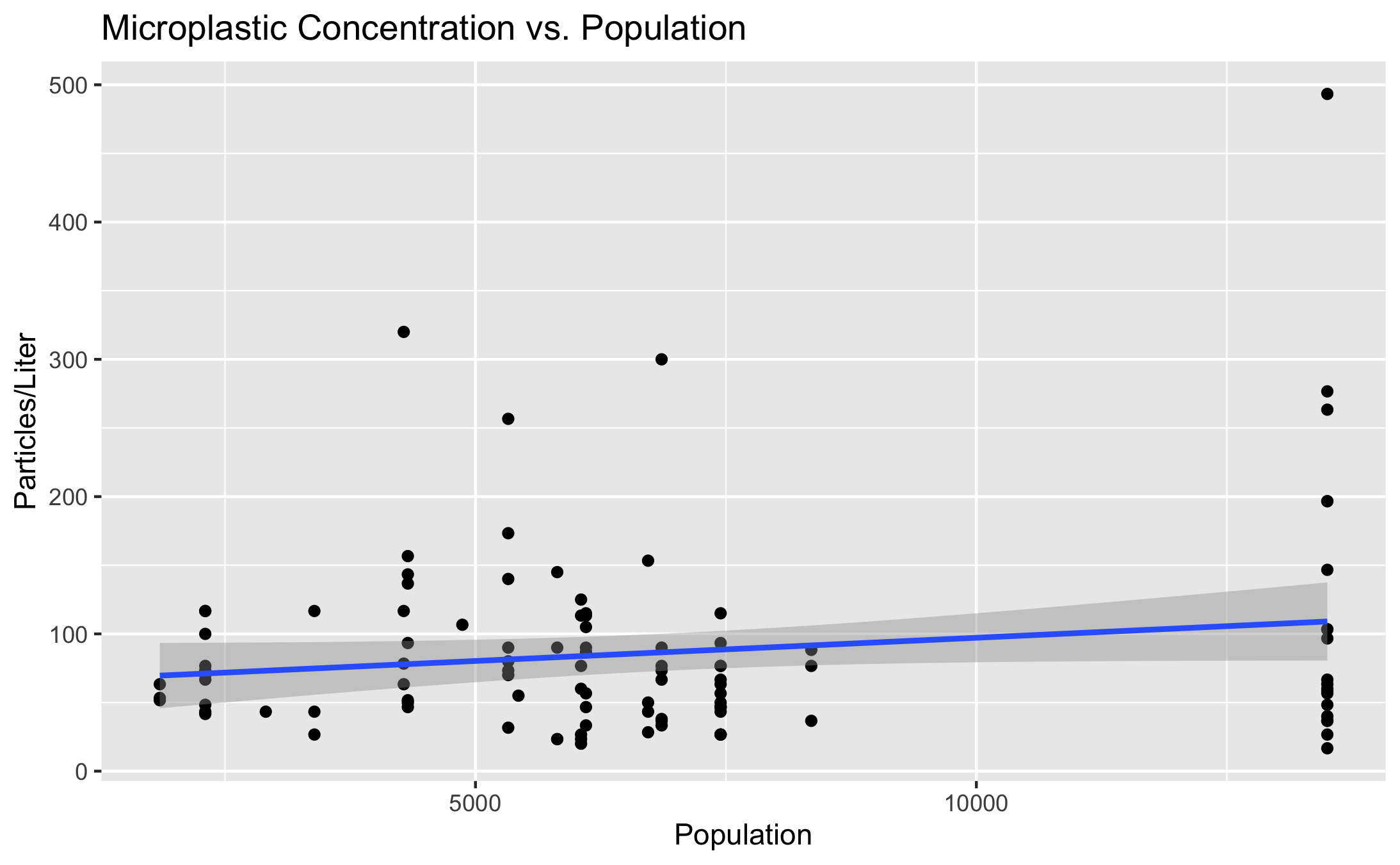
Predictor matrix



Land cover matrix

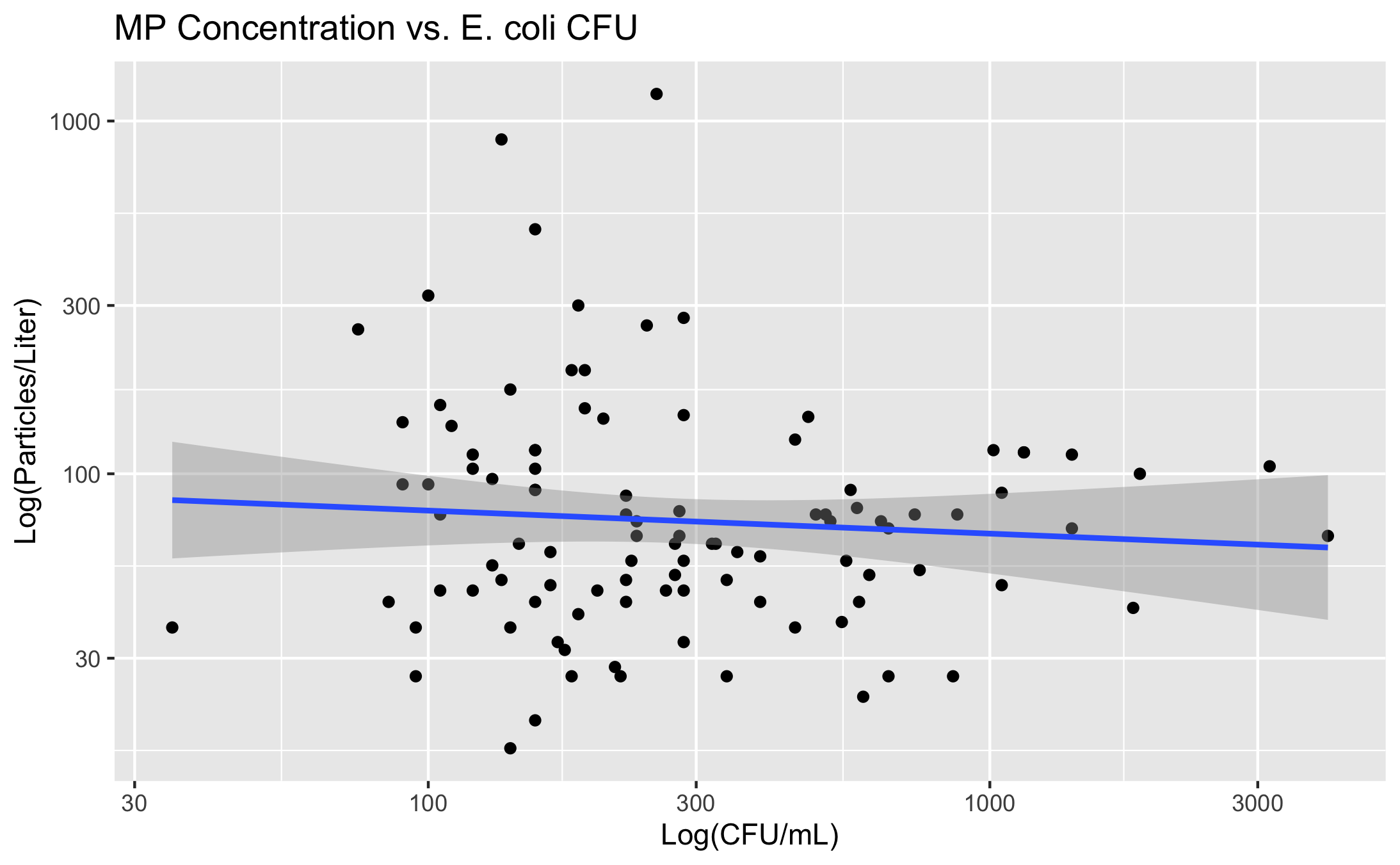
## 3.1 Full analysis

Preliminary modeling reveals that there is not a strong relationship between microplastic concentration and population level. Figure 10 demonstrates a linear model fit.



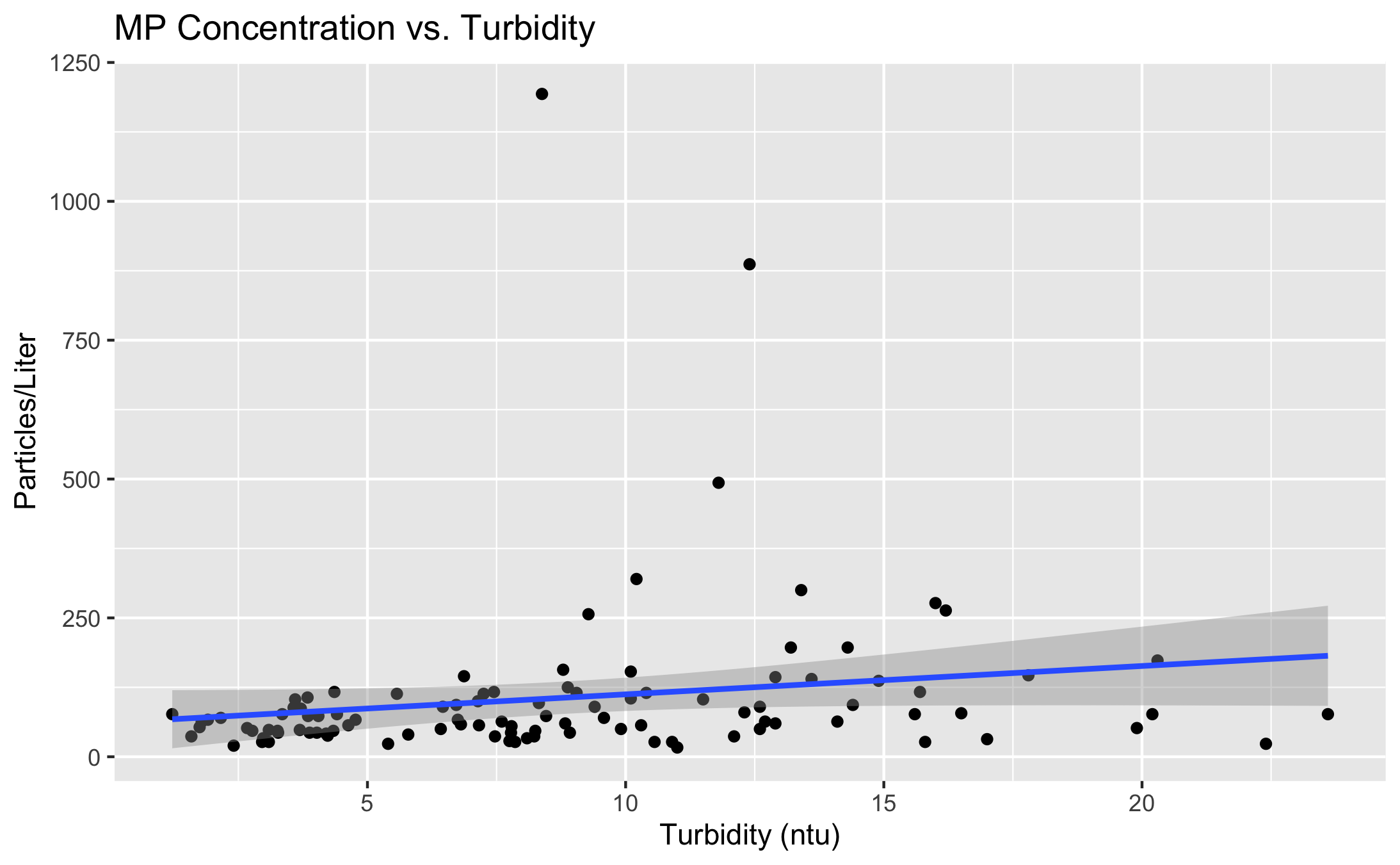
Concentration vs Population Linear Model

Figure 11 shows a linear model of microplastic concentration vs CFU (both variables log-transformed).



Concentration vs CFU Linear Model

Figure 12 demonstrates a linear model of particles/L vs turbidity.



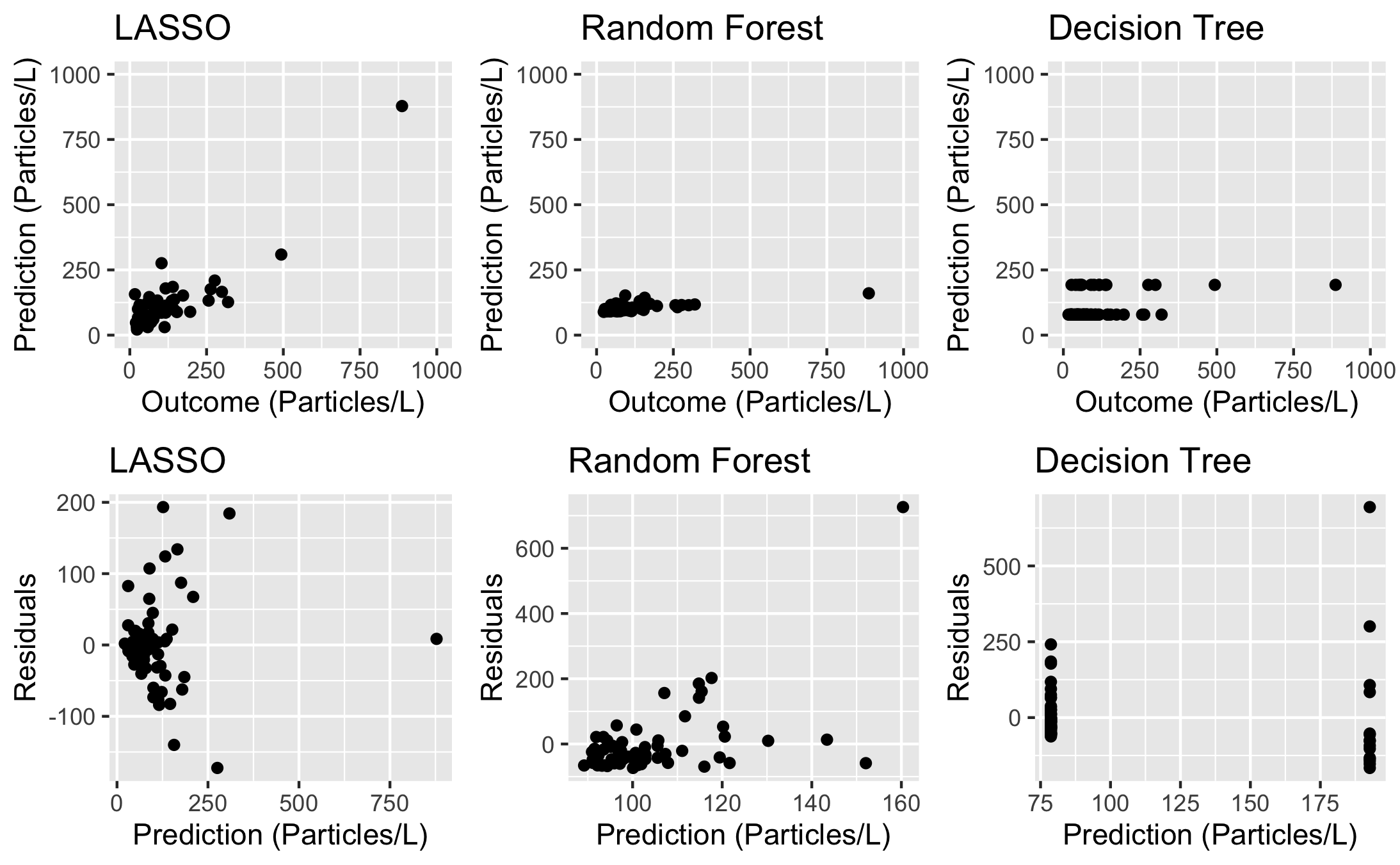
Concentration vs Turbidity Linear Model

Table 1 shows a table summarizing a linear model fit predicting particles/L with 6 predictors.

Linear model fit table.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -85.4155522 | 200.1731751 | -0.4267083 | 0.6708307 |
| visual\_score | 3.1501966 | 3.5503662 | 0.8872878 | 0.3777987 |
| turbidity.ntu | 8.1849262 | 5.1982368 | 1.5745582 | 0.1196244 |
| temperature.c | -0.9153529 | 3.7770358 | -0.2423469 | 0.8091818 |
| e.coli.cfu | -0.0269844 | 0.0352879 | -0.7646913 | 0.4468871 |
| population | -0.0019187 | 0.0108574 | -0.1767177 | 0.8602129 |
| dist | 0.0033504 | 0.0079782 | 0.4199472 | 0.6757408 |
| watershedBear Creek | -96.7602059 | 210.3266462 | -0.4600473 | 0.6468312 |
| watershedBrooklyn Creek | -4.7845357 | 140.7570975 | -0.0339914 | 0.9729755 |
| watershedCalls Creek | -17.2263069 | 97.7382122 | -0.1762495 | 0.8605794 |
| watershedHunnicutt Creek | 67.1138455 | 115.7405074 | 0.5798648 | 0.5637672 |
| watershedMcNutt Creek | -5.3292443 | 87.4009356 | -0.0609747 | 0.9515438 |
| watershedMiddle Oconee River | -49.3718448 | 114.0909789 | -0.4327410 | 0.6664615 |
| watershedNorth Oconee River | 104.2539248 | 116.1557777 | 0.8975354 | 0.3723442 |
| watershedOconee River | -0.7654564 | 137.9000000 | -0.0055508 | 0.9955861 |
| watershedSandy Creek | -60.1850641 | 203.0550068 | -0.2963978 | 0.7677565 |
| watershedTanyard Creek | 77.6196199 | 158.0304559 | 0.4911687 | 0.6247607 |
| watershedTrail Creek | 13.0256205 | 132.2887015 | 0.0984636 | 0.9218304 |

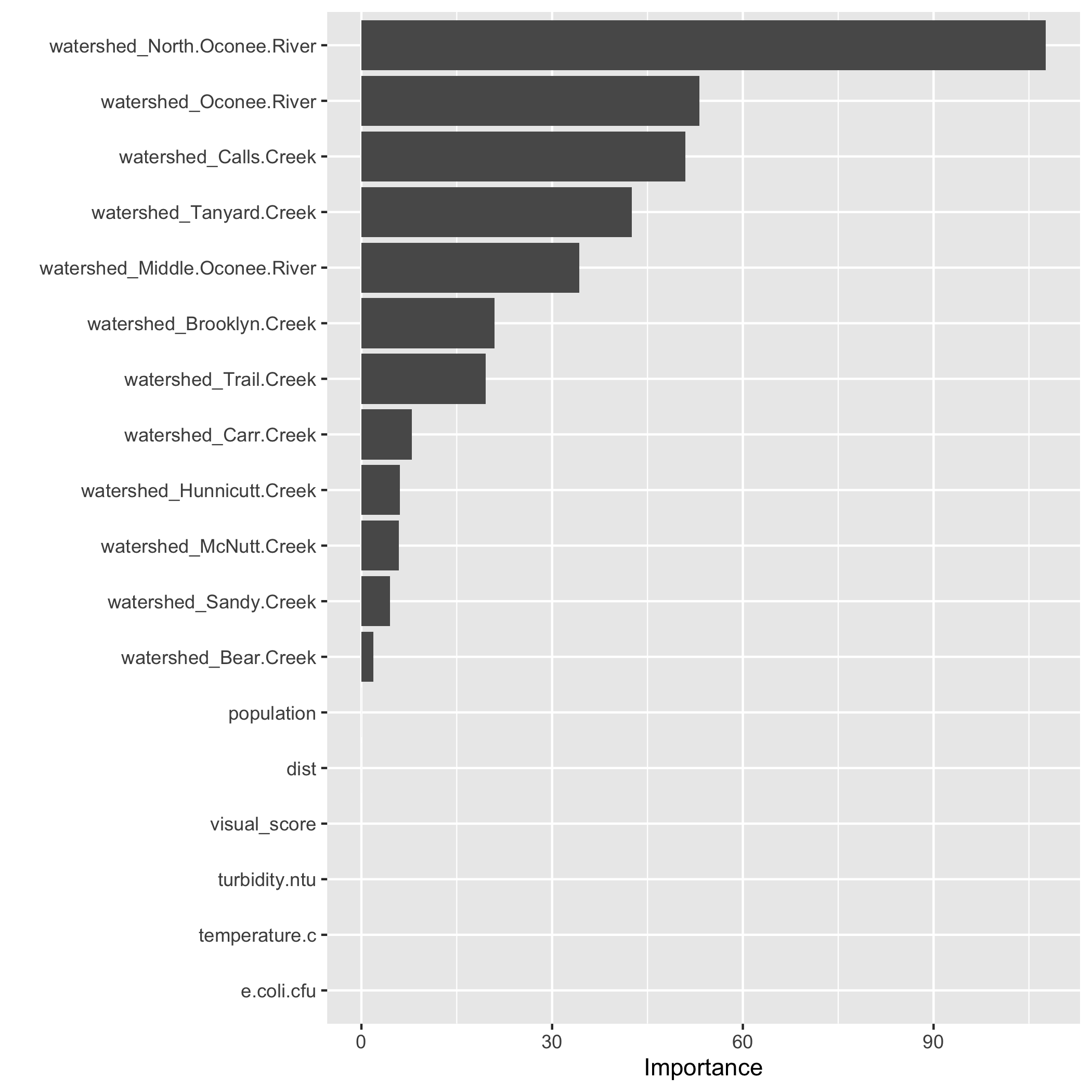
Beyond the basic linear model, we have applied additional methods to improve model performance, including LASSO regularization and building decision trees and random forests for model comparison. The predictions, outcomes, and residuals resulting from each type of plot are demonstrated in Figure 13



Model Quality

Based on the results of the three different models, the LASSO model is the best option for this dataset, though the minimal difference in RMSE when compared to the null model suggests that even though LASSO is the better model method compared to others, it still does not produce a great model for predicting microplastic concentration.

Figure 14 demonstrates variable importance in the final selected LASSO model. None of the hypothesized predictors appear as important variables in this model.



Variable importance