An Analysis of Microplastics Bioaccumulation in Lake Mead

Andrea Del Guidice

Purdue University

GRAD 505: Foundations of Data Science

Dr. Timothy Keaton

March 1, 2025

An Analysis of Microplastics Bioaccumulation in Lake Mead

The U.S. Geological Survey (USGS) data release entitled “Microplastics in Lake Mead National Recreation Area, 2017-2018” (Baldwin, 2019) is a data set which measures concentrations of microplastics of different morphologies (i.e., shapes and sizes) in water, sediment, fish, and shellfish at nine locations within Lake Mead Recreational Area, NV/AZ. Samples of two species of fish (striped bass and common carp) and two species of shellfish (quagga mussels and Asian clams) were collected at the various locations to test levels of microplastic bioaccumulation. The differing feeding types of the collected species (benthic vs. pelagic), varied levels of anthropogenic (human) impact at collection locations, and assorted morphology of the bioaccumulated microplastics in this data set provide a number of different routes to statistically explore the biological uptake of microplastic.

Research Questions and Hypotheses

Although not an exhaustive list of all questions and hypotheses that can be posed against this data, in this research I focused on three questions and their accompanying hypotheses:

- **Q1:** Do the differing feeding styles of the two included fish species (i.e., striped bass = pelagic feeder; common carp = benthic feeder) have a statistically-relevant impact on their bioaccumulation of microplastic particles?
    - **Q1 Hypothesis**: Yes, the differing feeding styles (i.e., benthic vs. pelagic) of the two included fish species do have a statistically significant impact on their bioaccumulation of microplastic particles.
    - **Reasoning:** Since the benthic (bottom) feeders will be feeding closer to the more concentrated plastic particles in the lake floor sediment, it is reasonable

to expect the uptake of microplastic particles would be greater by the benthic feeders rather than the open water-feeding pelagic fish.

    - **Note:** I've also included an additional section (essentially, a Q1.b) which compares all benthic feeders, whether fish or shellfish. The hypothesis I tested is for this "Q1.b" is that all benthic feeders have a similar bioaccumulation of microplastic particles, since they feed near the same sediment.

- **Q2:** Does the shape of microplastic particles impact their biological uptake by fish and shellfish from the surrounding environment?
    - **Q2 Hypothesis**: Yes, the shape of microplastic particles impacts the biological uptake of those particles by both fish and shellfish.
    - **Reasoning:** Since there were a much greater number of fiber-shaped microplastic particles counted in these lake fish and shellfish, the fiber shape may allow for greater uptake than other microplastic particle shapes.
- **Q3:** Does a location with greater human (i.e., anthropogenic) impact statistically increase the biological uptake of microplastic particles by fish and shellfish?
    - **Q3 Hypothesis:** Yes, a location with greater anthropogenic impact will statistically increase the biological uptake of microplastic particles by fish and shellfish.
    - **Reasoning:** The areas of the lake most impacted by humans would likely have the highest concentration of microplastic particles, and that higher concentration of particles would lead to higher incidence of biological particle uptake.

Data Analysis and Visualizations

Q1: Impact of Feeding Type on Bioaccumulation of Microplastic

Pelagic vs. Benthic Fish

A key, yet implicit, dimension of this dataset is the feeding type of fish. The data includes one representative species from each feeding category: striped bass, a pelagic (open water) feeder, and common carp, a benthic (bottom) feeder. As seen in Table 1 and Figure 1 below, both fish species have relatively low overall bioaccumulation. For instance, median particle counts are only one to two particles per specimen observation. The data is also fairly dense (i.e., low spread), with a few upper outliers heavily skewing the mean values. In fact, the mean values for both the carp and the striped bass are so right-skewed that the means are very near the Q3 value in both cases.

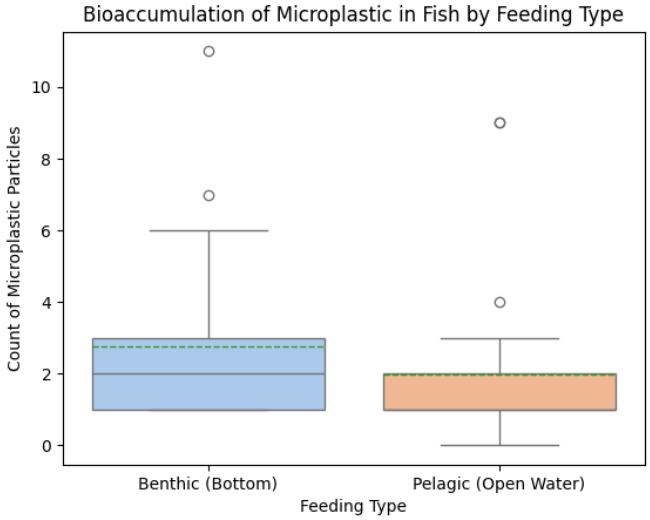
Table 1

Summary Statistics of Microplastic Bioaccumulation in Fish by Feeding Type

		Microplastic Particles							
		count	mean	std	min	25%	50%	75%	max
Feeding Type	Species Name								
Benthic (Bottom)	common carp	24.0	2.75000	2.506513	1.0	1.0	2.0	3.0	11.0
Pelagic (Open Water)	striped bass	32.0	1.96875	2.039677	0.0	1.0	1.0	2.0	9.0

Figure 1

Boxplot of Microplastic Bioaccumulation in Fish by Feeding Type



From the boxplot in Figure 1, we can also see that the pelagic striped bass has a Q3 of two particles, which is the same as the Q2 value for the benthic carp; Q3 of the carp data is at the upper fence of the striped bass data. Therefore, the counts of microplastic particles are statistically higher in the benthic carp data than the pelagic striped bass, suggesting that the bottom feeders may have small but meaningfully higher concentration of microplastic particles.

Statistical Testing. To test the hypothesis that this small difference in microplastic particles between the benthic carp and the pelagic striped bass is statistically meaningful, an unpaired 2-Sample t-test with hypotheses $H_0: \mu_p - \mu_b = 0;\ H_1: \mu_p - \mu_b \neq 0$ was conducted. This 2-Sample t-test returns a p-value of 0.2186, which is greater than significance level $\alpha = 0.05$ and we therefore fail to reject the null hypothesis.

Figure 2

2-Sample T-test of Microplastic Bioaccumulation in Fish by Feeding Type

```
> # Unpaired 2-sample t-based CI
> t.test(x=bass_vals, y=carp_vals, alternative = "two.sided", var.equal = FALSE)

        Welch Two Sample t-test

data:  bass_vals and carp_vals
t = -1.2481, df = 43.549, p-value = 0.2186
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.0430913  0.4805913
sample estimates:
mean of x mean of y
  1.96875   2.75000
```

An ANOVA test with hypotheses $H_0: \mu_p = \mu_b = 0$ and H_1: The μ_i differ was also run to test our benthic vs. pelagic hypothesis. This ANOVA test in Figure 3 resulted in p-value Pr(>F) = 0.2041, which is again greater than significance level $\alpha = 0.05$. Therefore, for both the ANOVA and 2-Sample tests we draw the same conclusion: we fail to reject the null hypotheses and determine there is not enough evidence to support a significant difference in means between the two feeding types.

Figure 3

ANOVA Test of Microplastic Bioaccumulation in Fish by Feeding Type

```
Analysis of Variance Table

Response: mead_df$Particles
                    Df  Sum Sq Mean Sq F value Pr(>F)
mead_df$Species_Name  1   8.371  8.3705  1.6529 0.2041
Residuals            54 273.469  5.0642
```

Validation of Assumptions. To use either the 2-Sample t-test or ANOVA test, the data samples must be independent, normally distributed, and with similar variances. Although there is n>30 samples for pelagic feeders, a normal assumption doesn't hold up for the sample data per either feeding types' Q-Q Plots (not a 45-degree angle line; see Figure 4 below). The linear model's residuals for particle count by feeding type also display this lack of normality (Figure 5). On p. 132 of our textbook, Montgomery (2020) states moderate departures from data normality do not adversely affect the procedure, but there may be a better procedure to test this data given its lack of normality; additional data observations would certainly help.

Figure 4

Lack of Normality in Feeding Type Observations

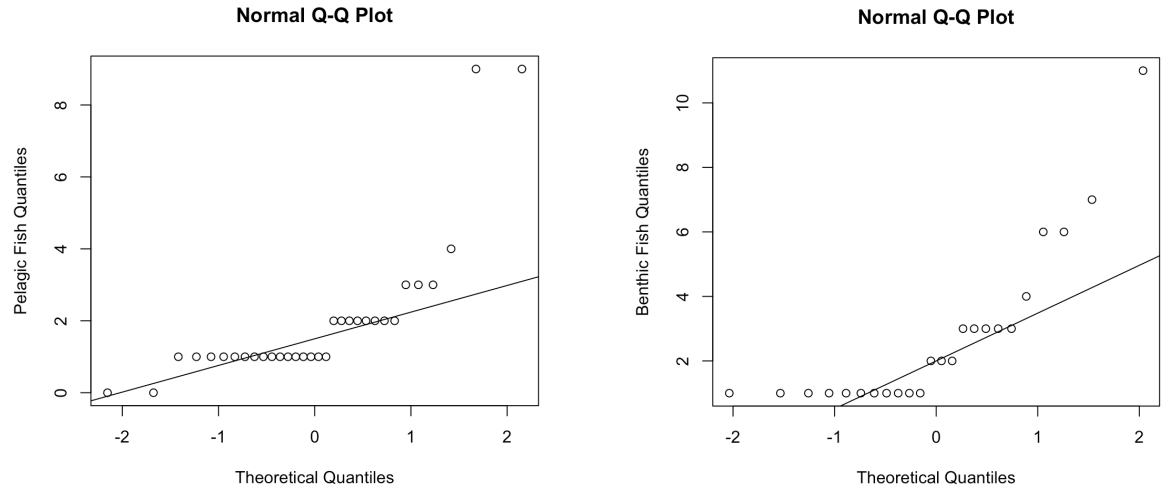
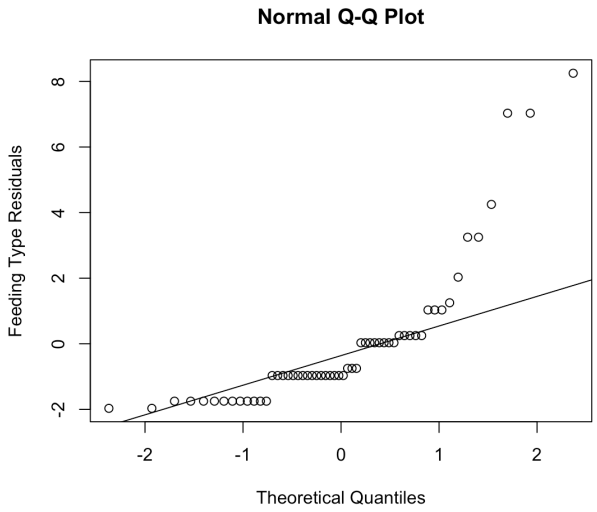


Figure 5

Lack of Normality in Feeding Type Residuals



Regarding variance, while homogeneity of variance was not assumed across feeding types for this test, running an F-test for variance shows we do not have enough evidence to *reject* their variances being equal (p-value = 0.2821 > significance level $\alpha = 0.05$ = don't reject H_0). Therefore, homogeneity of variance is not a concern for these Q1 tests.

Benthic Fish vs. Shellfish

Shellfish like clams and mussels are also benthic, so a comparison of shellfish and benthic fish particle counts is also included below. Table 2 and the boxplots in Figure 6 show the microplastic particle count statistics for all benthic feeders in this data:

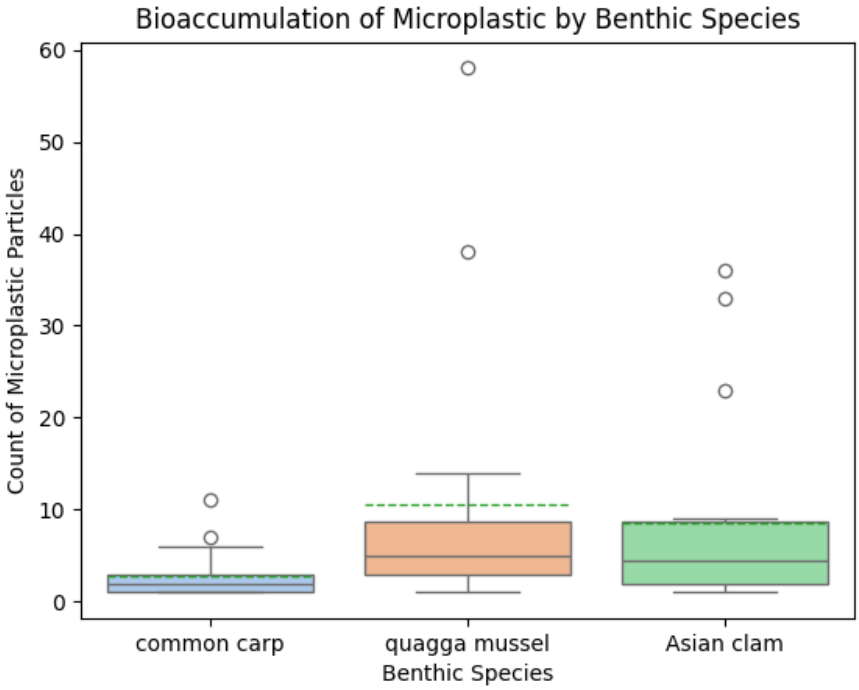
Table 2

Summary Statistics of Microplastic Bioaccumulation in Benthic Feeders

			Microplastic Particles							
			count	mean	std	min	25%	50%	75%	max
Feeding Type	Specimen Type	Species Name								
Benthic (bottom)	shellfish	quagga mussel	16.0	10.625000	15.426708	1.0	3.0	5.0	8.75	58.0
		Asian clam	18.0	8.611111	10.738271	1.0	2.0	4.5	8.75	36.0
	fish	common carp	24.0	2.750000	2.506513	1.0	1.0	2.0	3.00	11.0

Figure 6

Boxplot of Microplastic Bioaccumulation in Benthic Feeders



Looking at the Figure 5 boxplots, we can see that the Asian clam and quagga mussel data are comparable to one another, with close median values, the same Q3 values, similar spreads, and upper outliers heavily right-skewing the means (the mean is even above Q3 for the quagga mussels). Importantly, with the carp particle data's Q3 value well below the median of both shellfish particle data, we can also see there exists a much greater microplastic uptake in the both shellfish than the common carp fish, even if all three are bottom feeders.

Statistical Testing. To test the hypothesis that the mean particle count of at least one of these benthic species (presumably the carp, given our boxplots) is meaningfully different, we can run an ANOVA test with hypotheses H_0: $\mu_{b1} = \mu_{b2} = \mu_{b3} = 0$ and H_1: At least one μ_i differs. The results of this ANOVA test (Figure 7) show a p-value Pr(>F) = 0.04324 that is smaller than $\alpha = 0.05$. Therefore, the null hypothesis H_0 can be rejected and we accept the alternative hypothesis that at least one of the benthic species' particle count means differ.

Figure 7

ANOVA Test of Microplastic Bioaccumulation in Benthic Feeders

```
> anova(particles_by_benthic_species)
Analysis of Variance Table

Response: mead_benthic_df$Particles
          Df Sum Sq Mean Sq F value  Pr(>F)
species    2  686.6  343.30  3.3274 0.04324 *
Residuals 55 5674.5  103.17
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Validation of Assumptions. As mentioned in the benthic vs. pelagic testing above, to use an ANOVA test, data samples must be independent, normally distributed, and with similar variances. Independence is not an issue because we are comparing three distinct species (no overlap of samples). Although there is n>30 samples for shellfish, normality again does not exist for either group of benthic feeders nor their residuals, per Q-Q Plots; lack of a 45-degree angle line for benthic feeder residuals shown below in Figure 8. More data observations would help normality. As for homogeneity of variance, per Levene's Test, p-value is Pr(>F) = 0.0965, small but not < α =0.05, i.e., not small enough to reject a null hypothesis of equal variance (Figure 9).

Figure 8

Lack of Normality in Benthic Feeder Residuals

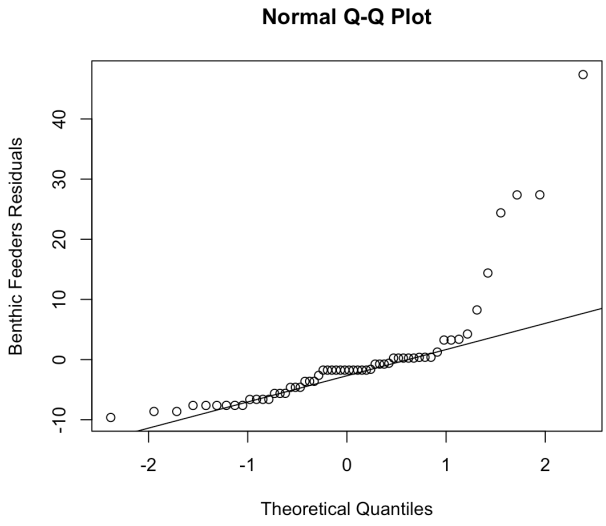


Figure 9

Levene's Test for Homogeneity of Variance for Benthic Feeders

```
> # Perform Levene's test for variance
> levene_test <- leveneTest(mead_benthic_df$Particles~species)
> print(levene_test)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  2.4405 0.0965 .
      55
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Q2: Impact of Particle Shape on Bioaccumulation of Microplastic

Another factor considered in this microplastic bioaccumulation study is if the shape of the microplastic particles influences biological uptake. From Table 3 below, we can see that while fiber- and fragment-shaped particles both appear to have greater intake rates for fish than film or foam, the impact of shape influencing microplastic uptake is much more pronounced for shellfish.

Table 3

Summary Statistics of Microplastic Bioaccumulation by Shape

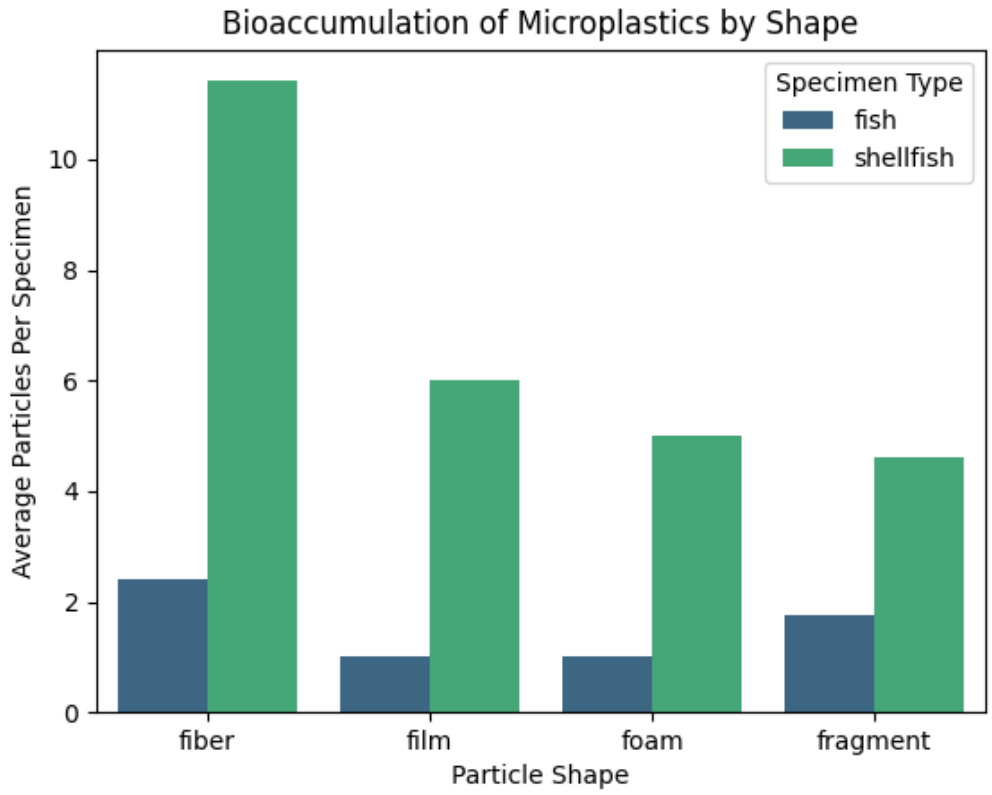
		Microplastic Particles							
		count	mean	std	min	25%	50%	75%	max
Specimen	Shape								
fish	fiber	50.0	2.400000	2.347382	0.0	1.00	2.0	3.00	11.0
	film	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
	foam	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
	fragment	4.0	1.750000	1.500000	1.0	1.00	1.0	1.75	4.0
shellfish	fiber	24.0	11.416667	14.943129	1.0	2.75	5.0	9.50	58.0
	film	3.0	6.000000	7.000000	1.0	2.00	3.0	8.50	14.0
	foam	2.0	5.000000	4.242641	2.0	3.50	5.0	6.50	8.0
	fragment	5.0	4.600000	2.607681	1.0	3.00	5.0	7.00	7.0

From Figure 10 below, which plots microplastic particle shapes by their average (mean) accumulation per specimen, we see that fiber-shaped particles are the shape of microplastic most accumulated by shellfish. This visualization also shows us that while fibers are the most

commonly accumulated shape in both fish and shellfish, *all* observed shapes of microplastic particles are accumulated at a higher rate in shellfish.

Figure 10

Bar Plot of Microplastic Bioaccumulation by Shape



Statistical Testing. To test the hypothesis that the shape of microplastic particles has a statistically-meaningful impact on the bioaccumulation of those particles, an ANOVA test with hypotheses H_0: $\mu_{s1} = \mu_{s2} = \mu_{s3} = \mu_{s4} = 0$ and H_1: At least one μ_i differs was conducted on all 90 fish and shellfish samples together. Surprisingly, the p-value Pr(>F) = 0.9241 was *much* greater than significance level α =0.05, so we therefore fail to reject the null hypothesis H_0. Based on these results, the shape of microplastic particles does not have a significant effect on the number of bioaccumulated particles.

Figure 11

ANOVA Test of Microplastic Bioaccumulation by Shape in All Species

```
> anova(particles_by_shape)
Analysis of Variance Table

Response: mead_df$Particles
           Df Sum Sq Mean Sq F value Pr(>F)
shapes      3   38.2  12.730  0.1582 0.9241
Residuals 86 6921.6  80.484
```

To see if particle shape is any more statistically impactful on either shellfish or fish, the data for fish and shellfish was separated and the same ANOVA test was rerun on each set. Particle shape remained an insignificant impact on bioaccumulation in both fish and shellfish,

Figure 12

ANOVA Test of Microplastic Bioaccumulation by Shape in both Fish and Shellfish

```
> anova(fish_particles_by_shape)
Analysis of Variance Table

Response: mead_fish_df$Particles
           Df  Sum Sq Mean Sq F value Pr(>F)
shapes      3   5.089  1.6964  0.3188 0.8118
Residuals 52 276.750  5.3221
```

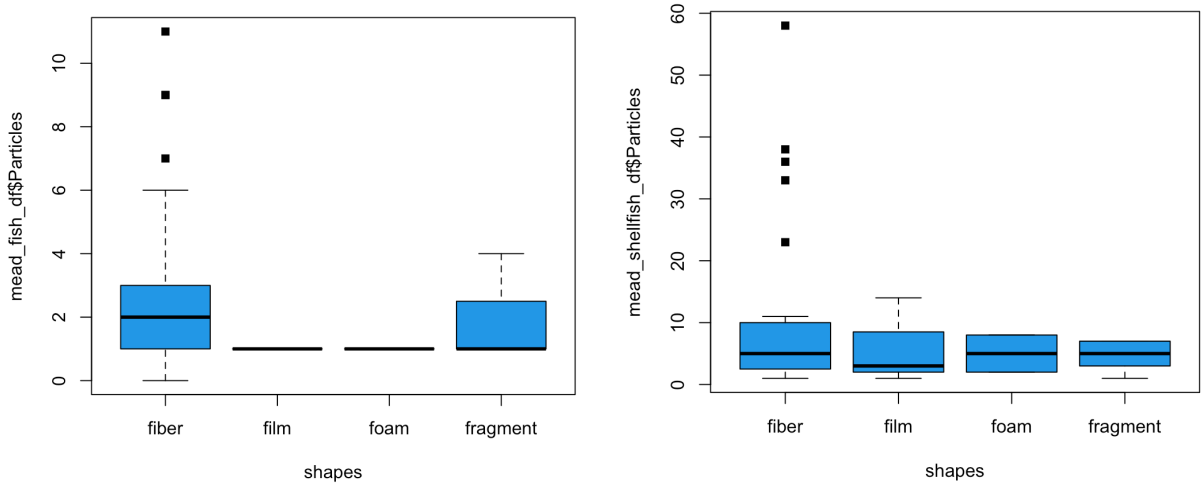
```
> anova(shellfish_particles_by_shape)
Analysis of Variance Table

Response: mead_shellfish_df$Particles
           Df Sum Sq Mean Sq F value Pr(>F)
shapes      3  285.3  95.116  0.5405 0.6582
Residuals 30 5279.0 175.968
```

as we see in Figure 12 that p-value Pr(>F) = 0.8118 for the 56 fish samples and Pr(>F) = 0.6582 for the 34 shellfish samples. Such high p-values for both datasets means that we are still unable to reject H_0 for both fish and shellfish and that particle shape is not a meaningful predictor of the count of microplastic particles in a sample. Also, as we can see in the relatively similar microplastic bioaccumulation medians across all shapes within both fish and shellfish below (Figures 13), it makes sense that particle shape is not a significant predictor of microplastic bioaccumulation.

Figure 13

Boxplot of Microplastic Bioaccumulation by Shape in Fish (left) and Shellfish (right)



Validation of Assumptions. As stated above, to use ANOVA, data groups must be normally distributed and independent, with homogeneity of variance. Using Q-Q plots to validate normality, we again do not see very normal distribution in the residuals, whether for all species together (Figure 14) or for just fish or shellfish (Figure 15):

Figure 14

Lack of Normality in Bioaccumulation by Shape Residuals in All Species

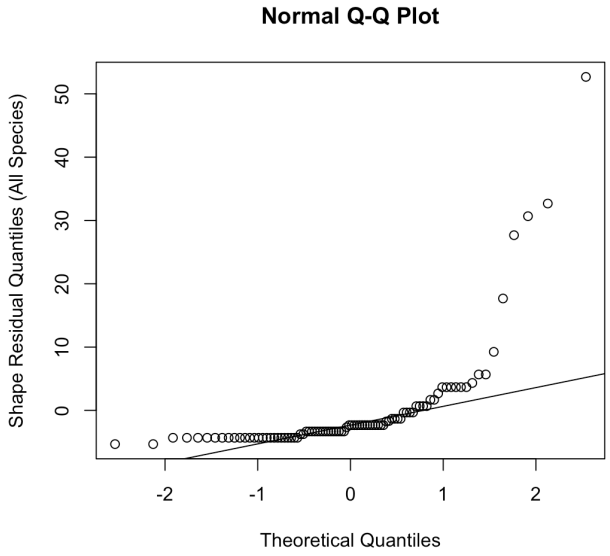
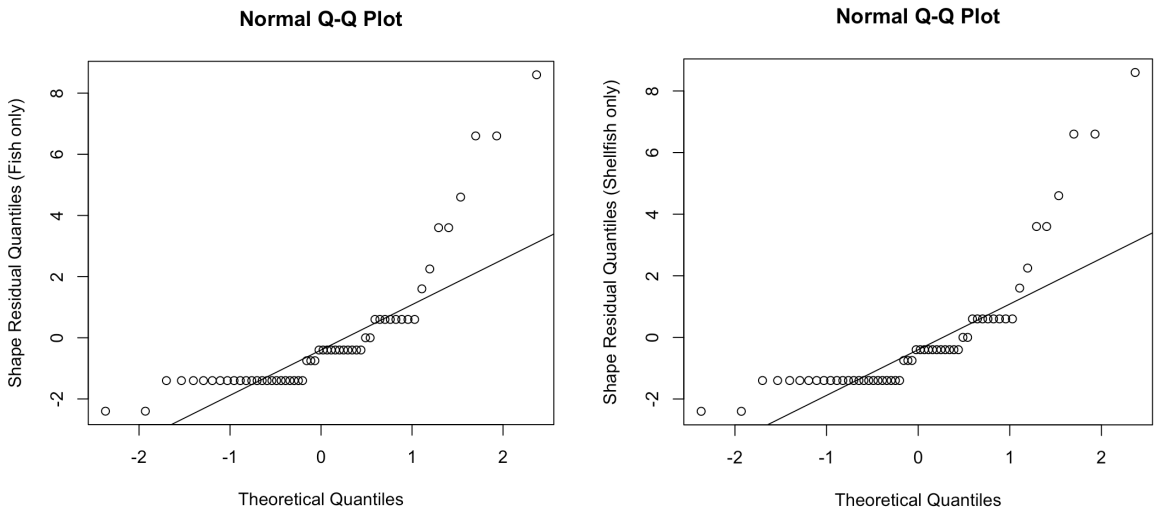


Figure 15

Lack of Normality in Bioaccumulation by Shape Residuals in Fish (left) and Shellfish (right)



To validate homogeneity of variance, if we only had two groups (levels) in our factor shape we could've used an F-test, but instead we can run a Levene Test. Because the Levene's Test p-value is a very large 0.9169, we can't reject our null hypothesis and we conclude we have homogeneity of variance (Figure 16):

Figure 16

Levene's Test for Homogeneity of Variance for Bioaccumulation by Shape

```
> # Perform Levene's test
> levene_test <- leveneTest(mead_df$Particles~shapes)
> print(levene_test)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.1692 0.9169
      86
```

Q3: Anthropogenic Impact on Bioaccumulation of Microplastic by Location

Of the nine Lake Mead locations that this data set's observations were taken at, the study's authors state in the accompanying Mead_microplastics4.xml metadata file (Baldwin, 2019) that the location with the most anthropogenic (i.e. human) impact is Las Vegas Bay and the

location with the least anthropogenic impact is Overton Arm. Therefore, I filtered the microplastics data set down to just these two locations and the fish and shellfish that were observed at both sites. (Note, the data has both Las Vegas Bay Shallow and Las Vegas Bay Deep locations, but no fish or shellfish observations were taken at Las Vegas Bay Deep so that site has been excluded from this particular analysis.)

Table 4

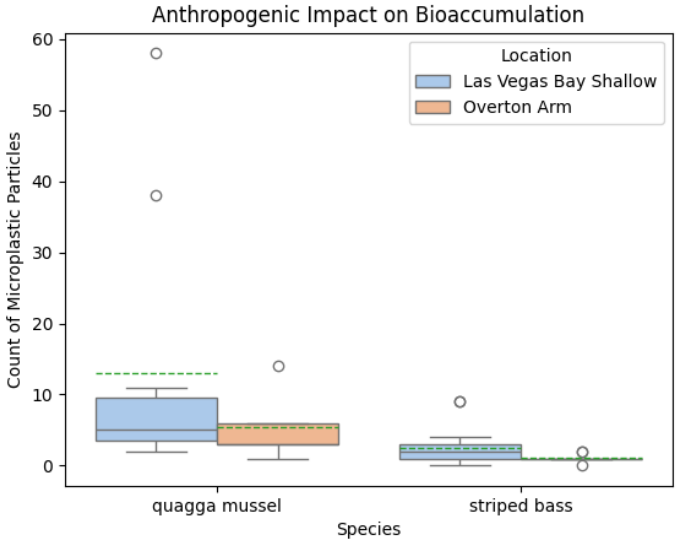
Summary Statistics of Anthropogenic Impact on Microplastic Bioaccumulation

		Microplastic Particles							
		count	mean	std	min	25%	50%	75%	max
Species Name	Location								
quagga mussel	Las Vegas Bay Shallow	11.0	13.000000	18.072078	2.0	3.5	5.0	9.5	58.0
	Overton Arm	5.0	5.400000	5.128353	1.0	3.0	3.0	6.0	14.0
striped bass	Las Vegas Bay Shallow	20.0	2.500000	2.417045	0.0	1.0	2.0	3.0	9.0
	Overton Arm	12.0	1.083333	0.514929	0.0	1.0	1.0	1.0	2.0

Unfortunately, only data for the striped bass (pelagic fish) and the quagga mussel (benthic shellfish) had been taken at both the Las Vegas Bay Shallow and Overton Arm locations, so it made sense to structure the analysis and visualization so each specimen type is only compared against itself at both locations. Accordingly, the Figure 17 boxplot below shows the count of accumulated microplastic particles at both locations grouped by species type.

Figure 17

Boxplot of Anthropogenic Impact on Microplastic Bioaccumulation

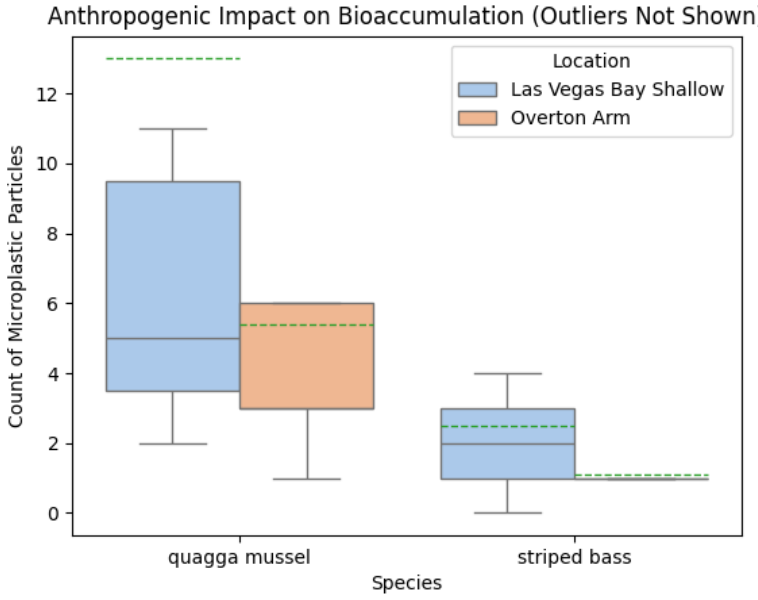


From this boxplot, we can see a definite increase in bioaccumulation at the more heavily human-impacted Las Vegas Bay for both species' specimens.

To 'zoom in' and better see the impact of the location and its implicit human influence on these species, in Figure 18 below, we simply hide the outliers from view (without affecting the mean or other statistics). Zoomed in, it is now clearer that the specimens in Las Vegas Bay are accumulating both more microplastic particles and greater outliers. The mean = 13 for the Las Vegas Bay mussels, way above the upper fence, is quite striking. For the spread, we can calculate from Table 5 that the IQR of the Overton Arm striped bass is 0 (Q3-Q1 = 0-0 = 0) whereas the IQR of the Las Vegas Bay bass has increased to 2 (Q3-Q1 = 3-1 = 2); for the Overton Arm mussels, IQR is 6-3 = 3, and for the Las Vegas Bay mussels, the IQR is doubled to 9.5-3.5 = 6! Visually, we can also see there is little to no microplastic bioaccumulation for the striped bass at Overton Arm but measurable microplastic bioaccumulation for that species in Las Vegas Bay. Finally, this figure again suggests that microplastic bioaccumulation appears to be affecting shellfish to a greater level than fish, especially pelagic fish.

Figure 18

Boxplot of Anthropogenic Impact on Microplastic Bioaccumulation, Outliers Not Shown



Statistical Testing. To test our hypothesis that higher human impact to the lake environment results in an increase in biological uptake of microplastic particles, a one-sided, unpaired 2-Sample t-test was performed twice, once for striped bass and once for quagga mussels. A t-test was utilized since the population variances are unknown. The variances were also assumed equal within each test, since we are testing within one species (i.e., striped bass to striped bass and quagga mussels to quagga mussels). The hypotheses for both species' one-sided, unpaired 2-Sample t-test are $H_0: \mu_{lvb} - \mu_{oa} = 0$ and $H_1: \mu_{lvb} - \mu_{oa} > 0$.

In test 1, a group of twenty striped bass samples from Las Vegas Bay were tested against a group of twelve striped bass from Overton Arm. This test in Figure 19 returned a p-value of 0.02783, which is smaller than significance level $\alpha = 0.05$. Because the p-value is low, we can reject the null hypothesis H_0 and accept H_1, concluding that the mean particle count is statistically greater in Las Vegas Bay than in Overton Arm for striped bass.

Figure 19

2-Sample T-test of Anthropogenic Impact on Microplastic Bioaccumulation in Striped Bass

```
> # Unpaired 2-sample t-based CI; var.equal because comparing species to themselves
> t.test(x=lvb_bass_vals, y=oarm_bass_vals, alternative = "greater", var.equal = TRUE, conf.level=0.95)

        Two Sample t-test

data:  lvb_bass_vals and oarm_bass_vals
t = 1.991, df = 30, p-value = 0.02783
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.2089892       Inf
sample estimates:
mean of x mean of y
 2.500000  1.083333
```

Next, we again use the same one-sided, unpaired 2-Sample t-test to test differences in particle count means, this time within the quagga mussel population. Eleven samples of quagga mussels from Overton Arm were compared to just five quagga mussels from Las Vegas Bay. Interestingly, running this one-sided, unpaired 2-sample t-test for quagga mussels

(Figure 20) returns a p-value = 0.1896, which is greater than significance level = 0.05. Therefore, we conclude we don't have enough data to reject null hypothesis H_0 for the quagga mussels. More samples are needed to perform a more accurate test.

Figure 20

2-Sample T-test of Anthropogenic Impact on Microplastic Bioaccumulation in Quagga Mussels

```
        Two Sample t-test

data:  lvb_quagga_vals and oarm_quagga_vals
t = 0.90804, df = 14, p-value = 0.1896
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -7.141535       Inf
sample estimates:
mean of x mean of y
     13.0       5.4
```

Validation of Assumptions. To use this 2-Sample t-test, we assumed variance were equal in each test group (same species) and that the data was roughly normal. Our Q-Q plots below in Figures 21 & 22 reinforce that there is not enough data for this test, with twenty striped bass samples and eleven quagga mussel samples at Las Vegas Bay and only twelve striped bass samples and five quagga mussel samples at Overton Arm. The Las Vegas Bay striped bass and Overton Arm mussel data looks slightly more normal than the other two plots, but additional data from these locations should be gathered to improve normality.

Figure 21

Lack of Normality in Anthropogenic Impact on Bioaccumulation Data in Striped Bass

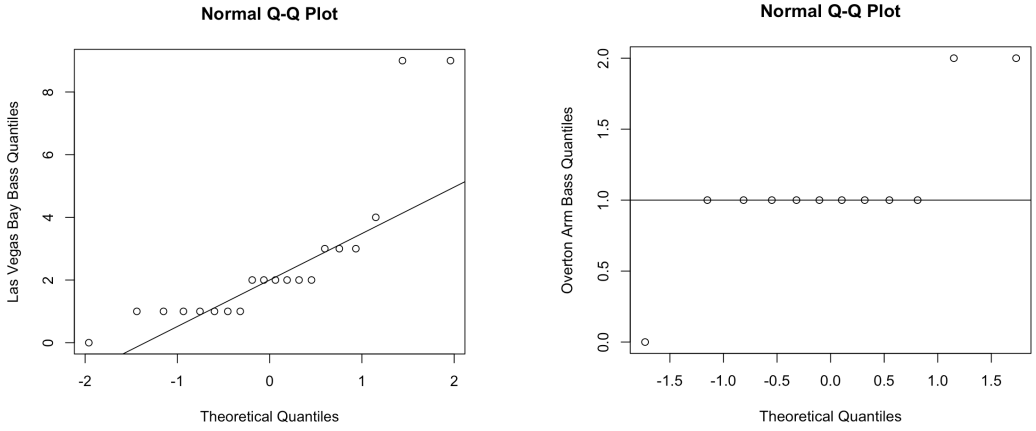
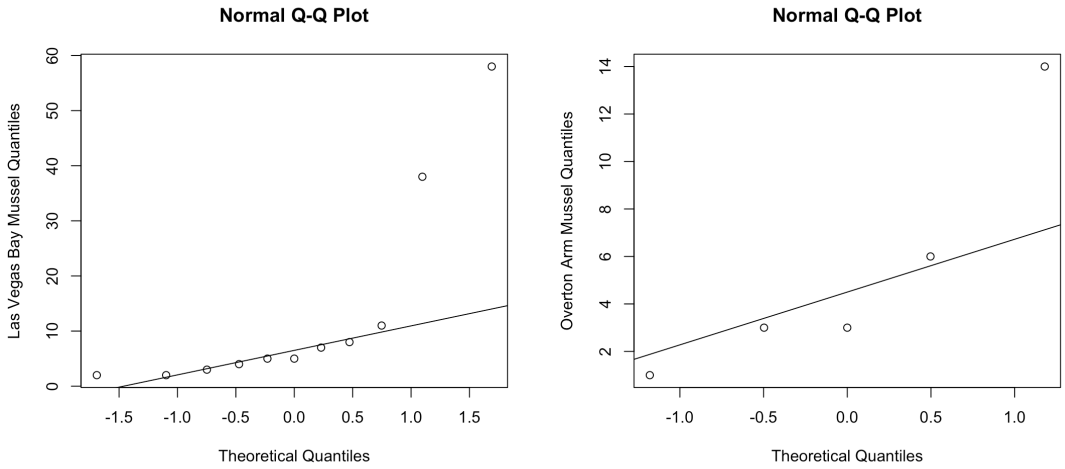


Figure 22

Lack of Normality in Anthropogenic Impact on Bioaccumulation Data in Quagga Mussels



Conclusions

Limits to Collected Data

Although I found this Lake Mead microplastics dataset sufficient to provide initial insights into my proposed research questions and exploratory data analysis, I was later hampered during statistical analysis by the dataset's small sample sizes and somewhat incomplete information. For a more complete and accurate analysis, additional data would be need to be collected, including a greater number of sample observations, observations of *all* species at each location site, and information such as the weight of the fish and shellfish (to support calculation of microplastic particles per unit of weight).

Preliminary Conclusions

In most cases, I was unable to reliably prove or disprove the three proposed hypotheses, as there was not enough (or barely enough) data available for data normality (i.e., requiring n >= 30 samples and Q-Q plots with closer to 45 degree plot lines). With 90 samples for testing, the strongest statistical conclusion we can draw from the analysis of

these three hypotheses is that for Q2, the shape of microplastic particles do not have a statistically-relevant impact on microplastic bioaccumulation.

Regarding our other two hypotheses, low sample counts make the statistical findings somewhat inconclusive. For Q1, comparing 32 observations for striped bass to 24 observations for common carp, we do not have enough evidence to determine a significant difference in microplastic bioaccumulation between benthic and pelagic feeding styles. However, we do have sufficient evidence to support an alternative hypothesis stating the mean particle count of one of the benthic feeders differs significantly from the others. Based on the included boxplot of all benthic feeders, we infer it is the common carp which differs from the shellfish.

For Q3, although our dataset only includes twenty striped bass samples from Las Vegas Bay and twelve from Overton Arm, the low t-test p-value and relatively normal Q-Q plot for Las Vegas Bay suggest that our Q3 alternative hypothesis—that greater anthropogenic impact on a lake leads to higher biological uptake of microplastic—may be valid for striped bass. However, there is not enough data available to adequately test any species group except for striped bass for Q3, with no data for Asian clams and common carp and only sixteen total quagga mussel observations.

Path for Future Investigations

An interesting path for continued investigation of this Lake Mead microplastic bioaccumulation data, especially if additional fish and shellfish observations are added, is if there are any statistically-relevant correlations between the ratio of microplastic particles in each Lake Mead species and their primary feeding substrate. During my exploratory analysis, I had included the water and surficial sediment observations to the microplastic particle data by shape, resulting in the summary table and bar chart below (Table 5 & Figure 23):

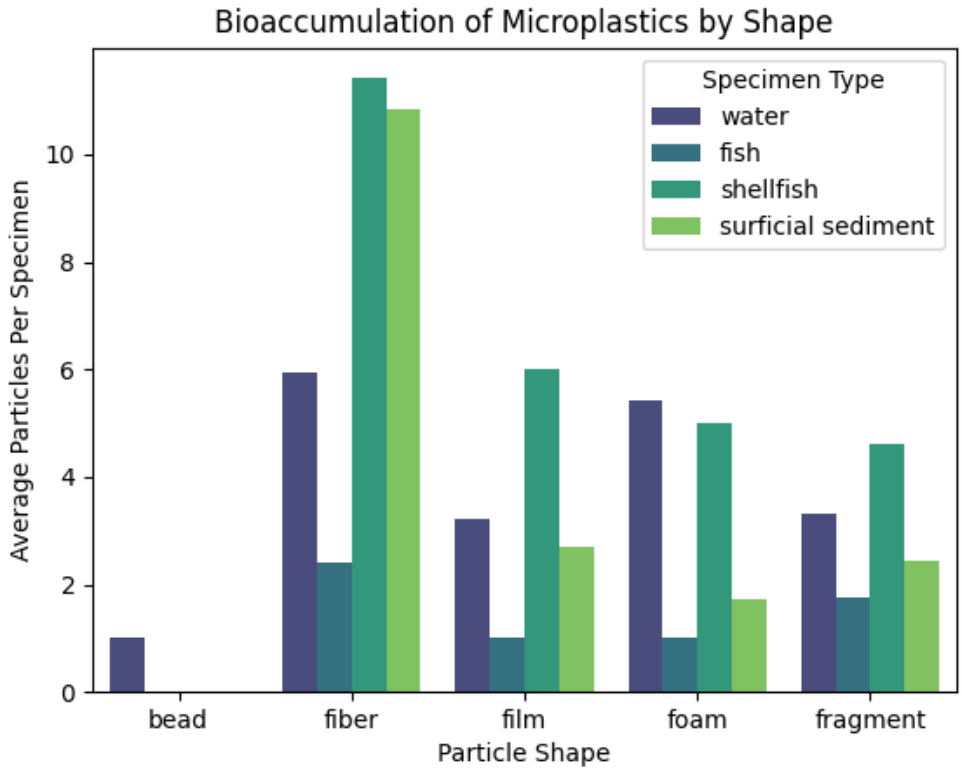
Table 5

Summary Statistics of Microplastic Bioaccumulation by Shape (Feeding Environments Included)

		Microplastic Particles							
		count	mean	std	min	25%	50%	75%	max
Specimen Type	Shape								
fish	fiber	50.0	2.400000	2.347382	0.0	1.00	2.0	3.00	11.0
	film	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
	foam	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
	fragment	4.0	1.750000	1.500000	1.0	1.00	1.0	1.75	4.0
shellfish	fiber	24.0	11.416667	14.943129	1.0	2.75	5.0	9.50	58.0
	film	3.0	6.000000	7.000000	1.0	2.00	3.0	8.50	14.0
	foam	2.0	5.000000	4.242641	2.0	3.50	5.0	6.50	8.0
	fragment	5.0	4.600000	2.607681	1.0	3.00	5.0	7.00	7.0
surficial sediment	fiber	97.0	10.824742	18.096391	1.0	2.00	5.0	10.00	125.0
	film	26.0	2.692308	2.908872	1.0	1.00	1.0	3.00	12.0
	foam	11.0	1.727273	1.793929	1.0	1.00	1.0	1.50	7.0
	fragment	31.0	2.451613	2.527696	1.0	1.00	1.0	3.00	11.0
water	bead	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
	fiber	127.0	5.929134	7.876206	1.0	1.00	3.0	7.00	48.0
	film	39.0	3.230769	3.652408	1.0	1.00	2.0	3.50	18.0
	foam	16.0	5.437500	7.483036	1.0	2.00	3.0	5.25	30.0
	fragment	72.0	3.305556	4.751987	1.0	1.00	2.0	3.25	28.0

Figure 23

Bar Plot of Microplastic Bioaccumulation by Shape (Feeding Environments Included)



These visualizations suggest, for instance, that the concentration of fiber-shaped particles may be strongly correlated for shellfish and their primary ‘surficial sediment’ (i.e., lakebed) feeding environment. There is also categorical data on particle size in this dataset that could be introduced into these analyses. Overall, deeper statistical evaluation on specimen feeding type and the particle concentration, shape, and size at their primary feeding locations should be considered.

Appendix

Note: For data and metadata files, please see accompanying uploaded .xlsx and .xml files.

Visualizations from Project Deliverable 2 (in Python)

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

# Set the Pandas display width to 1000 characters and display all columns
pd.set_option('display.width', 1000)
# Show all columns
pd.set_option('display.max_columns', None)

file_path = 'Mead_microplastics.xlsx'
mead_df = pd.read_excel(file_path)
# remove laboratory & field blanks (assume there for quality control) and sediment
core (not needed)
mead_df = mead_df[~mead_df['Environmental compartment'].isin(['laboratory
blank','field blank','sediment core'])]

# column heading clean-up
mead_df.rename(columns={'Location name':'Location'}, inplace=True)
mead_df.rename(columns={'Environmental compartment':'Specimen Type'}, inplace=True)
mead_df.rename(columns={'Number of microplastic particles': 'Microplastic Particles'},
inplace=True)
mead_df.rename(columns={'Species common name':'Species Name'}, inplace=True)
mead_df.drop(columns={'Comment','Sample name'}, inplace=True)

# add in Feeding Type column
feeding_types = {'Asian clam':'Benthic (bottom)','quagga mussel':'Benthic (bottom)',
                 'common carp':'Benthic (bottom)', 'striped bass':'Pelagic (open
water)', 'not applicable':'not applicable'}
mead_df['Feeding Type'] = mead_df['Species Name'].map(feeding_types)

###################
# Q1 Feeding Type analysis

# Create boxplot
particles_by_specimen = mead_df[~(mead_df['Feeding Type']=='not applicable')]
particles_by_fish = particles_by_specimen[particles_by_specimen['Specimen Type'] ==
'fish'].sort_values(by='Species Name')
particles_by_fish = particles_by_fish[['Microplastic Particles','Species Name',
'Feeding Type']].reset_index(drop=True)

print(particles_by_fish.groupby(['Feeding Type','Species Name']).describe())
print("\n")

# Make boxplots of microplastic particle count separated by fish type in a single
graph.
sns.boxplot(data=particles_by_fish, x="Feeding Type", y="Microplastic Particles",
meanline=True,
            showmeans=True, palette="pastel")
plt.title('Bioaccumulation of Microplastic in Fish by Feeding Type')
plt.xlabel('Feeding Type')
plt.ylabel('Count of Microplastic Particles')
plt.show()

# Make boxplots of microplastic particle count separated by specimen type in a single
```

```
graph.
particles_by_benthic_specimen = (particles_by_specimen[particles_by_specimen['Feeding
Type'] == 'Benthic (bottom)'].sort_values(by='Specimen Type'))
particles_by_benthic_table = particles_by_benthic_specimen.groupby(['Feeding
Type','Specimen Type','Species Name']).describe()
print(particles_by_benthic_table)

sns.boxplot(data=particles_by_benthic_specimen, x="Species Name", y="Microplastic
Particles",
            meanline=True, showmeans=True, palette="pastel")
plt.title('Bioaccumulation of Microplastic by Benthic Species')
plt.xlabel('Benthic Species')
plt.ylabel('Count of Microplastic Particles')
plt.show()

###################
# Q2 - Particle uptake by specimen type
shape_by_specimen = mead_df[~(mead_df['Shape'] == 'other')]
print(shape_by_specimen.groupby(['Specimen Type','Shape']).describe())

# data for particles by shape bar chart
shape_by_specimen = shape_by_specimen[['Shape','Microplastic Particles', 'Specimen
Type']]

### Uncomment this line when you want to exclude sediment and water data ####
#shape_by_specimen = shape_by_specimen[shape_by_specimen['Specimen
Type'].isin(['shellfish','fish'])].reset_index(
# drop=True)
avg_parts_shape_by_specimen = shape_by_specimen.groupby(['Shape','Specimen
Type']).mean()

sns.barplot(x='Shape', y='Microplastic Particles', hue='Specimen Type',
data=avg_parts_shape_by_specimen, palette="viridis")
plt.title("Bioaccumulation of Microplastics by Shape")
plt.xlabel("Particle Shape")
plt.ylabel("Average Particles Per Specimen")
plt.show()
###################

# Q3 - Anthropogenic Impact analysis
specimen_by_location = mead_df[mead_df['Species Name'].isin(['striped bass','quagga
mussel'])]

sns.boxplot(data=specimen_by_location, x="Species Name", y="Microplastic Particles",
hue="Location", meanline=True,
            showmeans=True, palette="pastel")#, showfliers=False)
plt.title('Anthropogenic Impact on Bioaccumulation') #(Outliers Not Shown)')
plt.xlabel('Species')
plt.ylabel('Count of Microplastic Particles')
plt.show()
```

Statistical Analysis and Additional Visualizations from Project Deliverable 3 (in R)

```
# Install readxl package
#install.packages("readxl")
library(readxl)

# Load dplyr package
#install.packages("dplyr")
library(dplyr)

# Install the car package if not already installed
#install.packages("car")
library(car)

#install.packages('EnvStats')
library(EnvStats)

#Q1-----------
mead_df=read_excel('/Users/adelguidice/Documents/education/Purdue/courses/GRAD505/Proj
ect/Mead_microplastics.xlsx')

# Using the subset() function to create a new dataframe
mead_df <- subset(mead_df, select = c('Species common name', 'Environmental
compartment', 'Number of microplastic particles', 'Shape'))

mead_df = mead_df %>%
  rename(
    Environment = 'Environmental compartment',
    Particles = 'Number of microplastic particles',
    Species_Name = 'Species common name'
  )

mead_fish_df = filter(mead_df, Species_Name %in% c("striped bass","common carp"))
print(mead_fish_df, n = Inf)

bass_vals = mead_fish_df$Particles[mead_fish_df$Species_Name=='striped bass']
length(bass_vals)

carp_vals = mead_fish_df$Particles[mead_fish_df$Species_Name=='common carp']
length(carp_vals)

#are these datasets normal?
qqnorm(bass_vals, ylab = "Pelagic Fish Quantiles")
qqline(bass_vals)
qqnorm(carp_vals, ylab = "Benthic Fish Quantiles")
qqline(carp_vals)

# Unpaired 2-sample t-based CI of pelagic vs. benthic fish
t.test(x=bass_vals, y=carp_vals, alternative = "two.sided", var.equal = FALSE)

#anova for pelagic vs. benthic fish
feeding_lm=lm(mead_fish_df$Particles~mead_fish_df$Species_Name)

feeding_type = factor(mead_fish_df$Species_Name)

#F-test for homogenity of variance
result = var.test(mead_fish_df$Particles~mead_fish_df$Species_Name)
print(result)

plot(mead_fish_df$Particles~feeding_type, pch=16, col='lightblue')

anova(feeding_lm)
```

```
#normal QQ plot of residuals
qqnorm(feeding_lm$residuals, ylab = "Feeding Type Residuals")
qqline(feeding_lm$residuals)

# to perform anova test on all benthic feeders, shellfish and fish; reload excel file
first
mead_benthic_df = filter(mead_df, Environment %in% c("fish","shellfish"))
mead_benthic_df = filter(mead_benthic_df, Species_Name != "striped bass" )

species = factor(mead_benthic_df$Species_Name)

#linear model = particles in benthic feeders by species
particles_by_benthic_species_lm=lm(mead_benthic_df$Particles~species)

summary(particles_by_benthic_species_lm)
anova(particles_by_benthic_species_lm)

#normal QQ plot of residuals
qqnorm(particles_by_benthic_species_lm$residuals, ylab = "Benthic Feeders Residuals")
qqline(particles_by_benthic_species_lm$residuals)

# Perform Levene's test for variance
levene_test <- leveneTest(mead_benthic_df$Particles~species)
print(levene_test)

#Q2------------
#refresh dataset
mead_df=read_excel('/Users/adelguidice/Documents/education/Purdue/courses/GRAD505/Proj
ect/Mead_microplastics.xlsx')

# Using the subset() function to create a new dataframe
mead_df <- subset(mead_df, select = c('Environmental compartment', 'Number of
microplastic particles', 'Shape'))

mead_df = mead_df %>%
  rename(
    Environment = 'Environmental compartment',
    Particles = 'Number of microplastic particles',
  )

mead_df = filter(mead_df, Environment %in% c("fish","shellfish"))
print(mead_df, n = Inf)
length(mead_df)

shapes = factor(mead_df$Shape)
#print(shapes)

#linear model = particles by shape
particles_by_shape=lm(mead_df$Particles~shapes)

summary(particles_by_shape)
anova(particles_by_shape)

plot(particles_by_shape,
     pch=as.numeric(shapes)+14, col=as.numeric(shapes)+3)

qqnorm(particles_by_shape$residuals, ylab = "Shape Residual Quantiles (All Species)")
qqline(particles_by_shape$residuals)

# Perform Levene's test for variance
levene_test <- leveneTest(mead_df$Particles~shapes)
print(levene_test)
```

```
#fish
mead_fish_df = filter(mead_df, Environment == "fish")
#print(mead_fish_df, n = Inf) #56 samples

shapes = factor(mead_fish_df$Shape)
fish_particles_by_shape=lm(mead_fish_df$Particles~shapes)

plot(mead_fish_df$Particles~shapes,
     pch=as.numeric(shapes)+14, col=as.numeric(shapes)+3)

summary(fish_particles_by_shape)
anova(fish_particles_by_shape)

qqnorm(fish_particles_by_shape$residuals, ylab = "Shape Residual Quantiles (Fish
only)")
qqline(fish_particles_by_shape$residuals)

plot(fish_particles_by_shape,
     pch=as.numeric(shapes)+14, col=as.numeric(shapes)+3)

#shellfish
mead_shellfish_df = filter(mead_df, Environment == "shellfish")
print(mead_shellfish_df, n = Inf)

shapes = factor(mead_shellfish_df$Shape)
shellfish_particles_by_shape=lm(mead_shellfish_df$Particles~shapes)

plot(mead_shellfish_df$Particles~shapes,
     pch=as.numeric(shapes)+14, col=as.numeric(shapes)+3)

summary(shellfish_particles_by_shape)
anova(shellfish_particles_by_shape)

qqnorm(fish_particles_by_shape$residuals, ylab = "Shape Residual Quantiles (Shellfish
only)")
qqline(fish_particles_by_shape$residuals)

#Q3---------
mead_df=read_excel('/Users/andrea/Documents/education/Purdue/courses/GRAD505/Project/M
ead_microplastics.xlsx')

# Using the subset() function to create a new dataframe
mead_df <- subset(mead_df, select = c('Species common name',
                                      'Environmental compartment',
                                      'Number of microplastic particles',
                                      'Location name'))

mead_df = mead_df %>%
  rename(
    Environment = 'Environmental compartment',
    Particles = 'Number of microplastic particles',
    Species_Name = 'Species common name',
    Location = 'Location name'
  )

mead_df = filter(mead_df, Species_Name %in% c("striped bass","quagga mussel"))
#print(mead_df, n = Inf)

#test LV Bay location samples vs Overton Arm samples for same species
#striped bass first
lvb_bass_vals = mead_df$Particles[mead_df$Location =='Las Vegas Bay Shallow' &
mead_df$Species_Name =='striped bass']
```

```
lvb_bass_vals
length(lvb_bass_vals)

oarm_bass_vals = mead_df$Particles[mead_df$Location =='Overton Arm' &
mead_df$Species_Name =='striped bass']
oarm_bass_vals
length(oarm_bass_vals)

# Unpaired 2-sample t-based CI; var.equal because comparing species to themselves
t.test(x=lvb_bass_vals, y=oarm_bass_vals, alternative = "greater", var.equal = TRUE,
conf.level=0.95)

#are these datasets normal?
qqnorm(oarm_vals, ylab = "Overton Arm Bass Quantiles")
qqline(oarm_vals)
qqnorm(lvb_vals, ylab = "Las Vegas Bay Bass Quantiles")
qqline(lvb_vals)

#quagga mussels next
lvb_quagga_vals = mead_df$Particles[mead_df$Location =='Las Vegas Bay Shallow' &
mead_df$Species_Name =='quagga mussel']
lvb_quagga_vals
length(lvb_quagga_vals)

oarm_quagga_vals = mead_df$Particles[mead_df$Location =='Overton Arm' &
mead_df$Species_Name =='quagga mussel']
oarm_quagga_vals
length(oarm_quagga_vals)

# Unpaired 2-sample t-based CI; var.equal because comparing species to themselves
t.test(x=lvb_quagga_vals, y=oarm_quagga_vals, alternative = "greater", var.equal =
TRUE, conf.level=0.95)

#are these datasets normal?
qqnorm(oarm_quagga_vals, ylab = "Overton Arm Mussel Quantiles")
qqline(oarm_quagga_vals)
qqnorm(lvb_quagga_vals, ylab = "Las Vegas Bay Mussel Quantiles")
qqline(lvb_quagga_vals)
```

References

Baldwin, A.K., Spanjer, A.R., Rosen, M.R., and Thom, T. (2019). Microplastics in Lake Mead National Recreation Area, 2017-2018: U.S. Geological Survey data release [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P9V1MNHH

Montgomery, D. C. (2020). *Introduction to Statistical Quality Control* (8th ed.). John Wiley & Sons, Inc.