MOSS Final Report

Krisztian Magori

1/18/2021

During the Summer of 2019, we identified 5 different sites across Spokane with mosquito activity in stormwater catch basins (see Figure 1 below). For 10 weeks, between July 9th 2019 and Sept 13th 2019, the graduate student (Sawyer Volyn) and the undergraduate research assistant (Charles Johnson) collected 3 water samples from 6 locations each within these 5 breeding sites once a week, along with environmental variables such as water temperature and depth. They took these 3x300 samples to the lab, and extracted and counted any mosquito larvae found in them. They grew some mosquito larvae to adults for species identification, and cultured some others for bacterial isolation. Water samples were stored frozen in -80C.

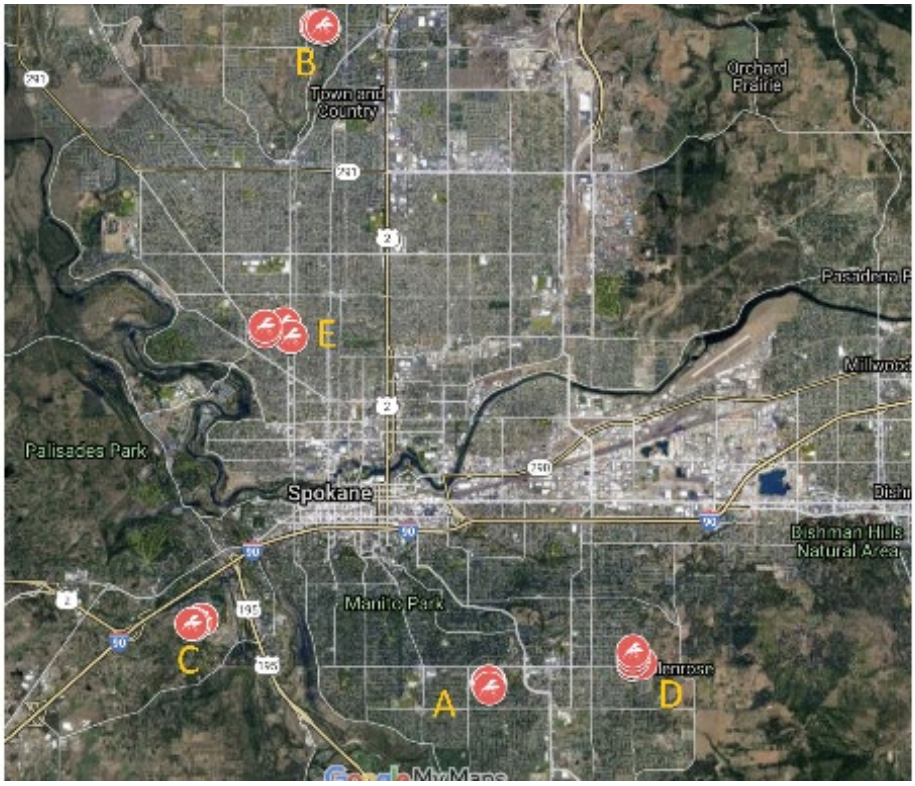

Figure 1. Locations of catch basins and culverts where water and mosquito samples were collected in 2019.

Exploratory data analysis

First, we observed that we were able to collect mosquito larvae at each of the 5 locations consistently. The overall number of mosquito larvae collected at the 5 locations is below:

The total number of larvae collected at the 5 locations across 10 weeks.

Location	Larvae
A	1709
B	344
C	1408
D	393
E	108

However, location A had 9 collection sites, while all other locations had only 6 collection sites. Therefore, looking at the distribution of the number of larvae collected per week is more appropriate:

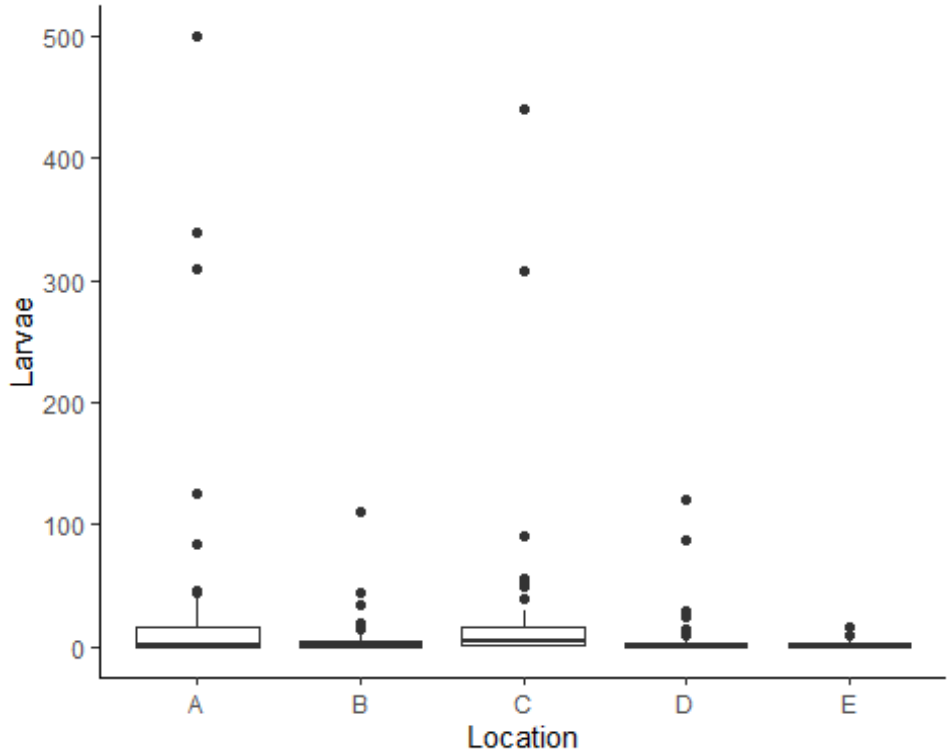

Figure2. Distribution of the number of larvae collected at each location across the study

This graph suggests that there might have been more larvae collected on average at locations A and C, relative to the other locations.

We also plotted the number of larvae collected per date (Figure 3) and per week (Figure 4) at each of the locations below.

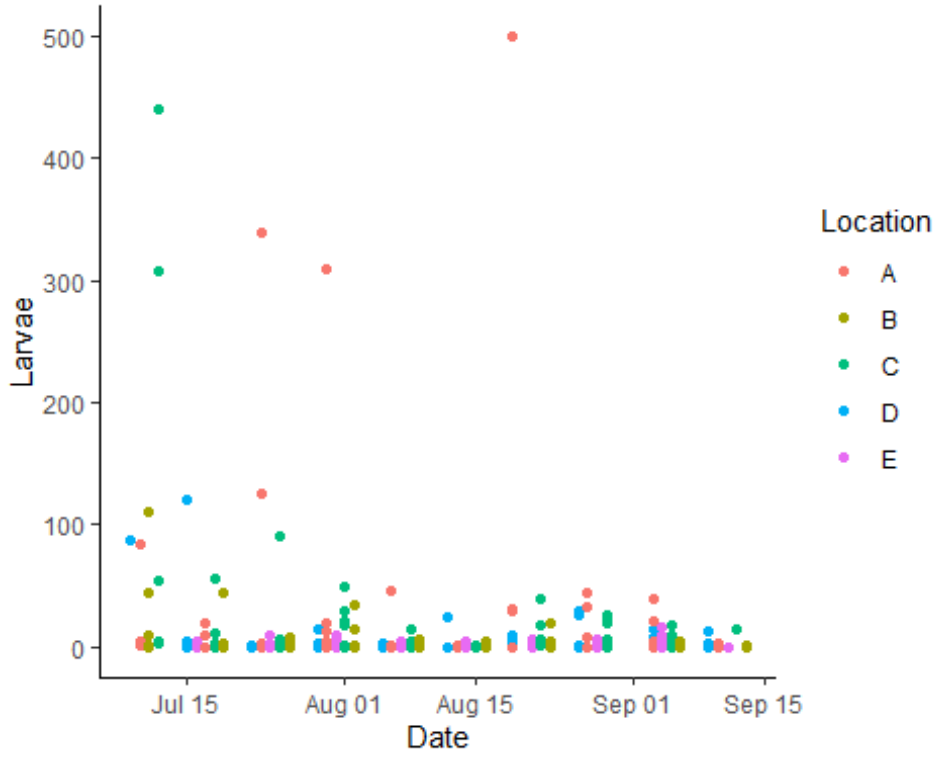

Figure 3. The number of larvae collected by date and location

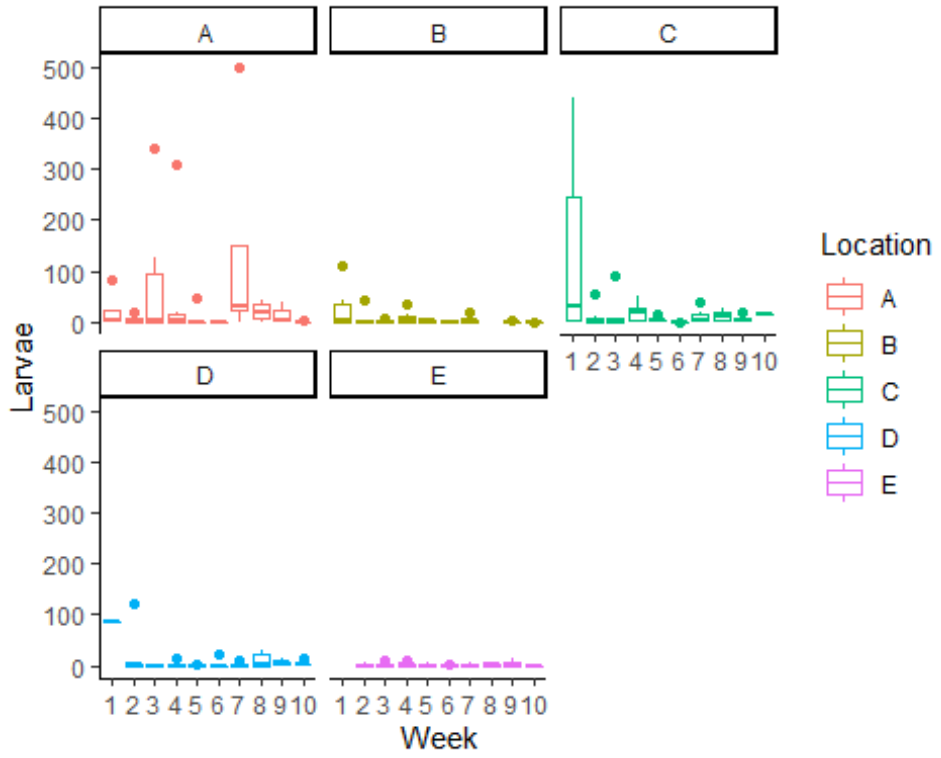

Figure 4. The number of larvae collected by collection week and location

Both of these graphs suggest that the highest number of larvae were collected right at the beginning of the collection, and numbers decreased at all locations across the rest of the season. There were a few remarkably high collections, e.g vials containing 500 larvae at location A on August 19th at collection site A5.

Statistical comparison between locations and weeks

Next, we needed to investigate if these differences between locations and weeks are statistically significant. First, we needed to see if the number of larvae per collection, our response variable, has a normal distribution. For that, we plotted a histogram:

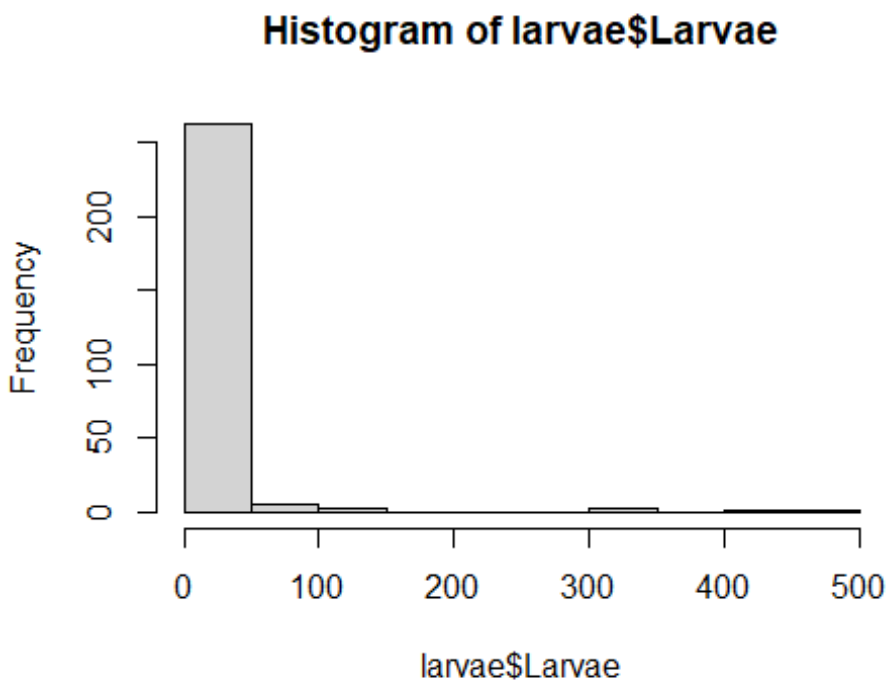

Figure 5. The distribution of the number of larvae collected per collection session throughout the study

The histogram indicated that the number of larvae collected was not normally distributed. Since, it is a count variable, we decided to use a quasi-Poisson regression. First, we tested the null hypothesis that the mean number of larvae collected was the same across all locations, estimating across the study period.

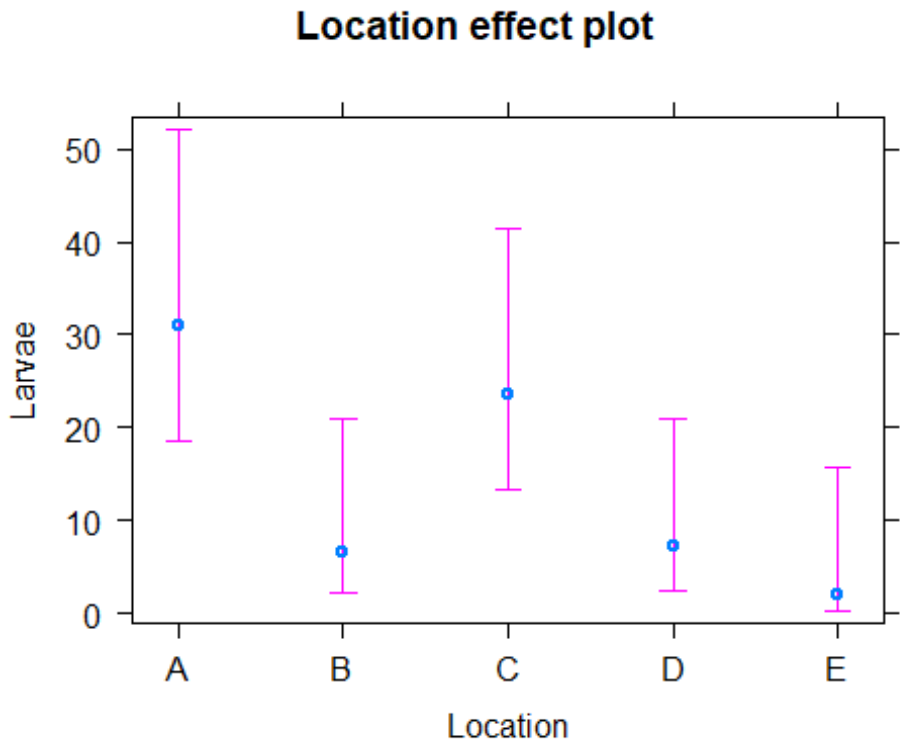

Figure 6. The estimated mean number of larvae per week at each location, with the 95% confidence interval

The p-value for this null hypothesis, based on a Chi-square test, was 2.680065810^{-4}, allowing us to reject this null hypothesis with a significance level of 0.05. Figure 6 suggests that Locations A and C have significantly higher mean number of larvae collected per week, relative to locations B, D and E, while there is not a significant difference between Locations A and C, and Locations B, D and E, respectively.

In terms of seasonality, we can also test the null hypothesis that there is no difference in the mean number of mosquito larvae collected in different weeks across the different locations.

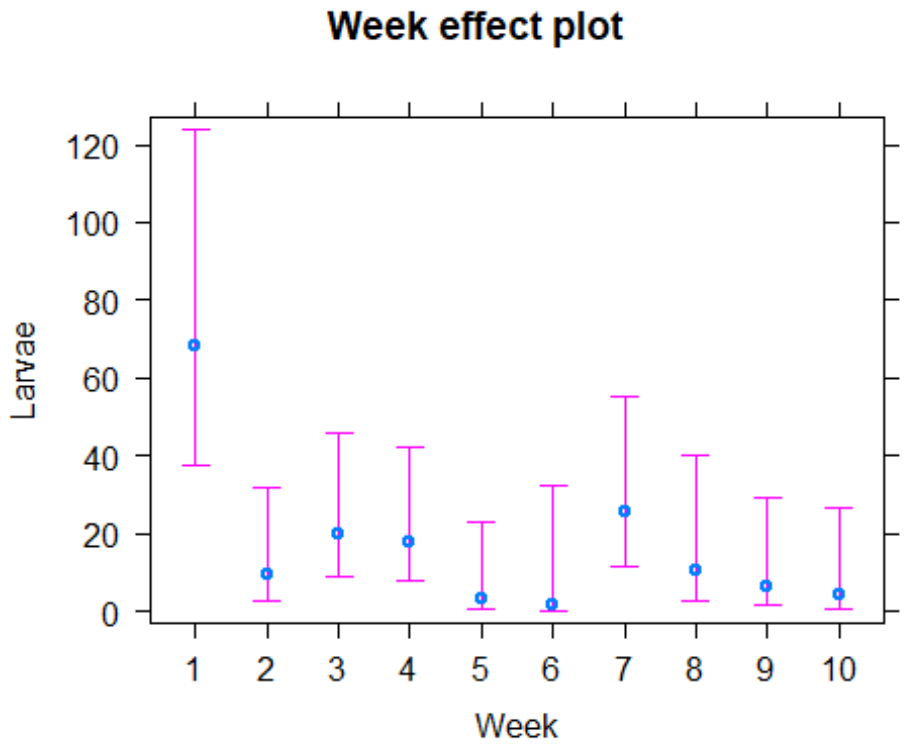

Figure 7. The mean number of mosquito larvae collected in different weeks across locations, and the 95% confidence interval of those

We were also able to reject the null hypothesis that there was no difference between the mean number of mosquito larvae collected in different weeks across the locations, with a p-value of 1.121754810^{-4}. Based on Figure 7, the estimated mean number of mosquito larvae collected was significantly higher in Week 1 relative to all other weeks, with smaller differences between other weeks.

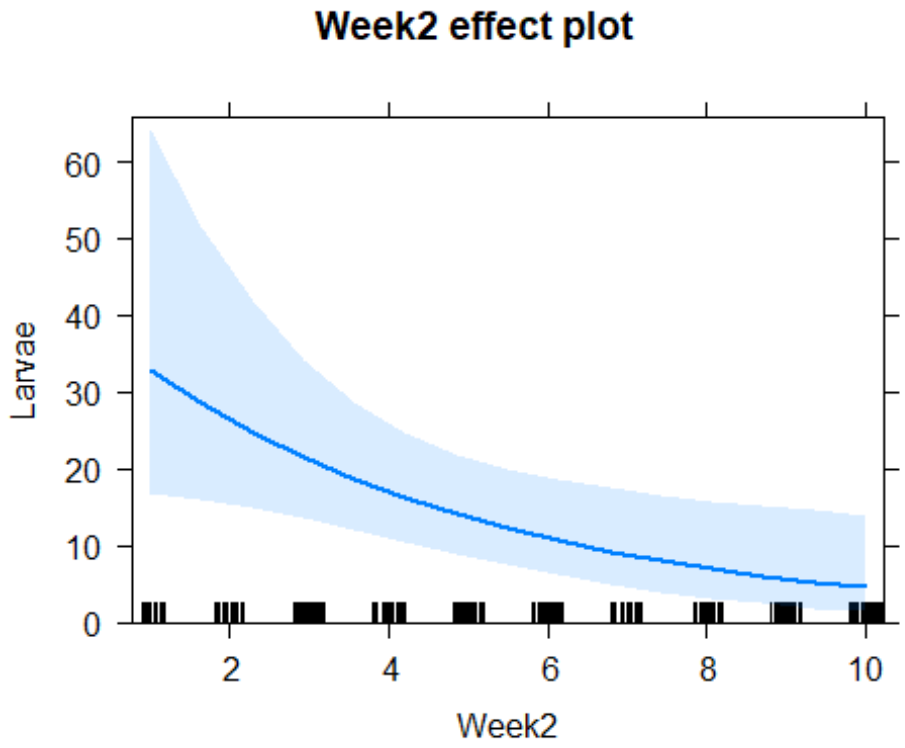

Figure 8. The estimated relationship between the mean number of larvae collected and the week of collection

There was also a significant negative relationship between the mean number of mosquito larvae collected and the number of week as a numerical variable (p=0.0128832), with the mean number of mosquito larvae collected decreasing by a factor of 0.8032886.

Since, based on the above results, there seem to be significant differences both between locations and between the weeks, we further investigate if different locations have a different seasonal pattern in terms of the mean number of mosquito larvae collected. Using the weeks as a categorical variable leads to an unwieldy model with two many combinations between the 5 locations and 10 weeks to compare with the available data. Instead, we use the number of weeks as a numerical variable in a multiple regression with quasi-Poisson distribution.

```
## Analysis of Deviance Table (Type III tests)
##
## Response: Larvae
##                LR Chisq Df Pr(>Chisq)
## Location         7.7872  4    0.09969 .
## Week2            0.3495  1    0.55441
## Location:Week2   7.0769  4    0.13188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

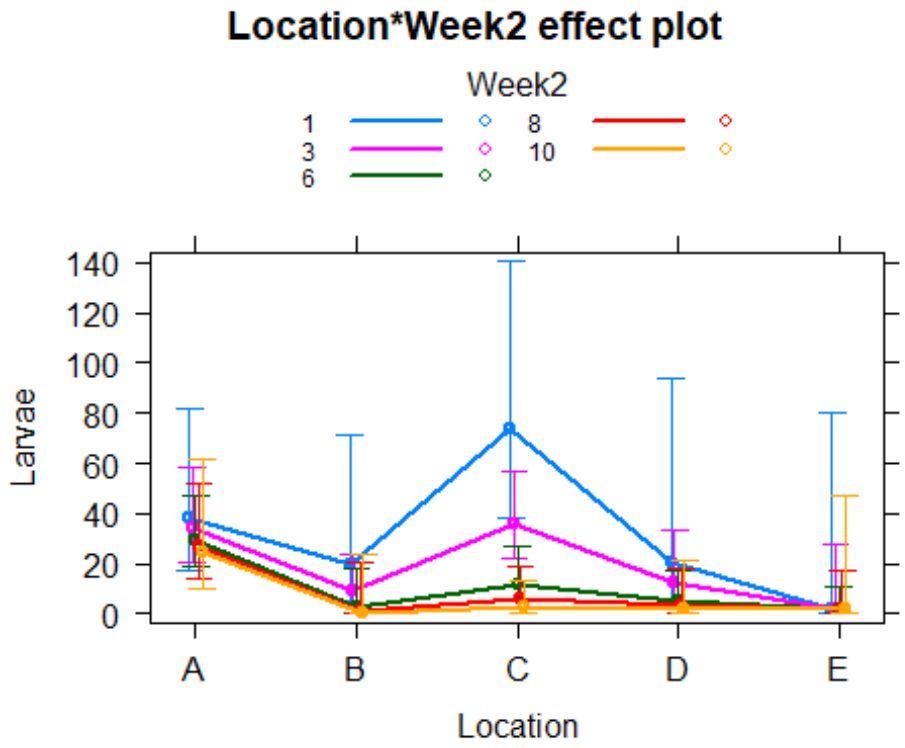

Figure 9. The estimated mean number of mosquito larvae collected in different weeks and different locations

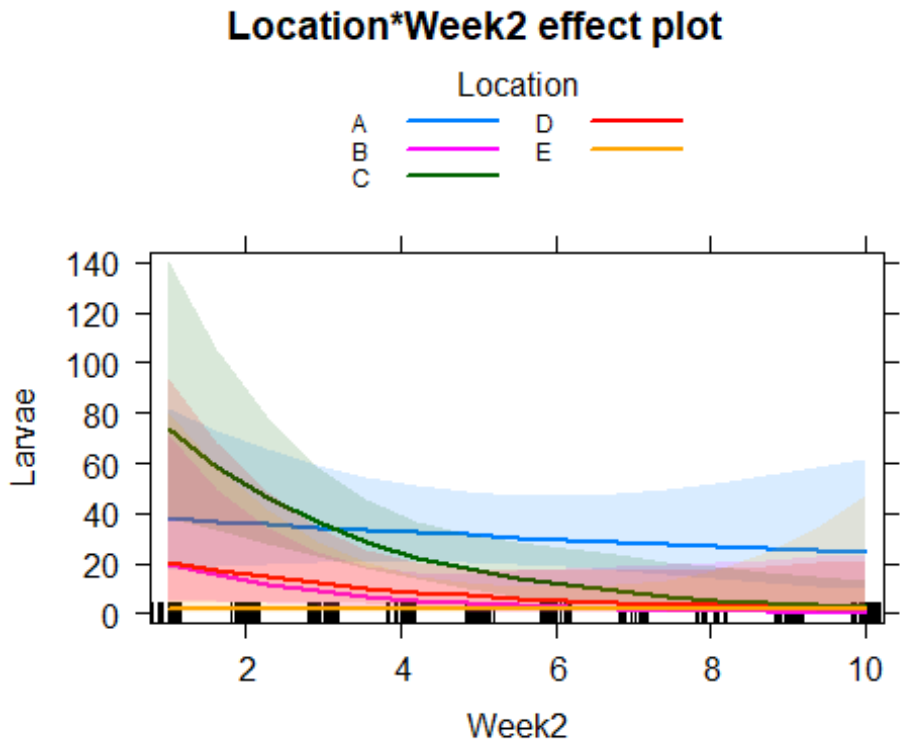

Figure 10. The estimated mean number of mosquito larvae collected in different weeks and different locations

The table and Figure 9 and Figure 10 shows that when trying to compare both locations and different weeks, we don't find a significant difference either between locations, or between weeks, or their interactions, which is most likely due to insufficient data. The overall model explains 26.6641521% of the variation in larval density.

Environmental variables

We hypothesize that there are specific environmental conditions that make Locations A and C more favorable to mosquito breeding in Week 1, relative to other locations throughout the study period.

The graduate student and the undergraduate student measured the depth of each catch basin and culvert when they were collecting mosquito and water samples. Depth was measured as the distance from the top of the grate and culvert to the top of the water surface, so it does not measure the depth of the water body itself. Therefore, the lower the depth value is, the shorter the distance is from the top of the water surface to the grate, i.e. the more water there is in the catch basin or culvert. First, we explore Depth between locations in different weeks.

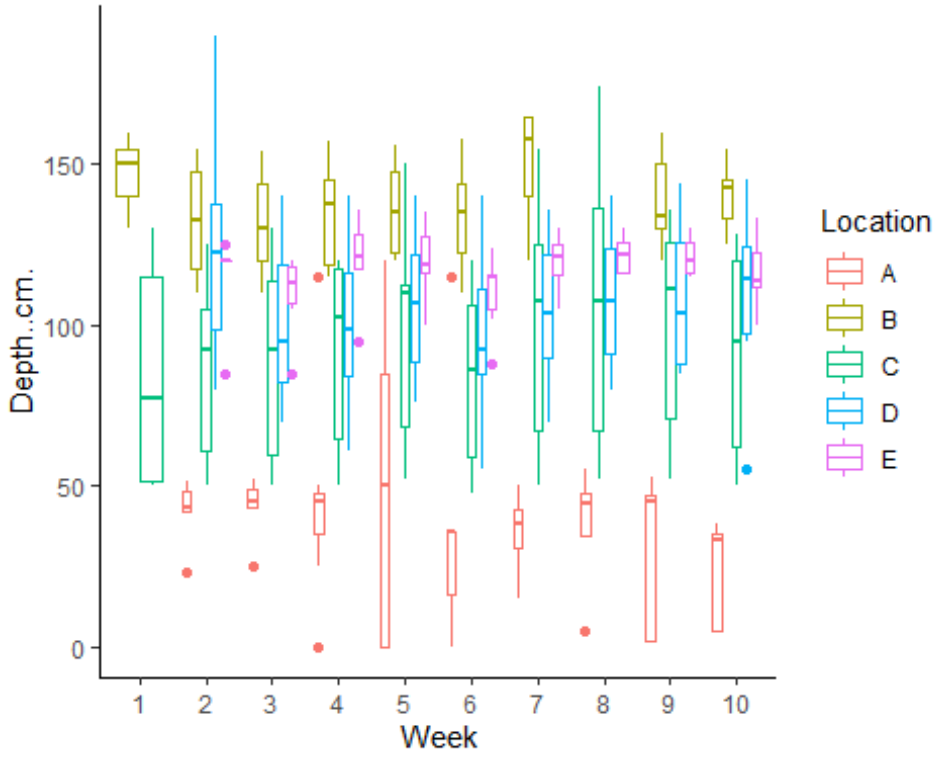

Figure 11. The distribution of depth in different weeks and locations

These figures suggest that Location A and C have lower depth, i.e. they have water closer to the surface, meaning more water. The figures also suggest that Depth did not change over the weeks appreciably. In order to test the null hypothesis that the mean depth of the surface is the same across weeks and across locations, we run a multiple regression with Location and the numerical version of week as predictor variables, assuming a normal distribution for Depth. Week was not a significant predictor in this model. Therefore we ran a simple one-way ANOVA to compare Depth between the different locations:

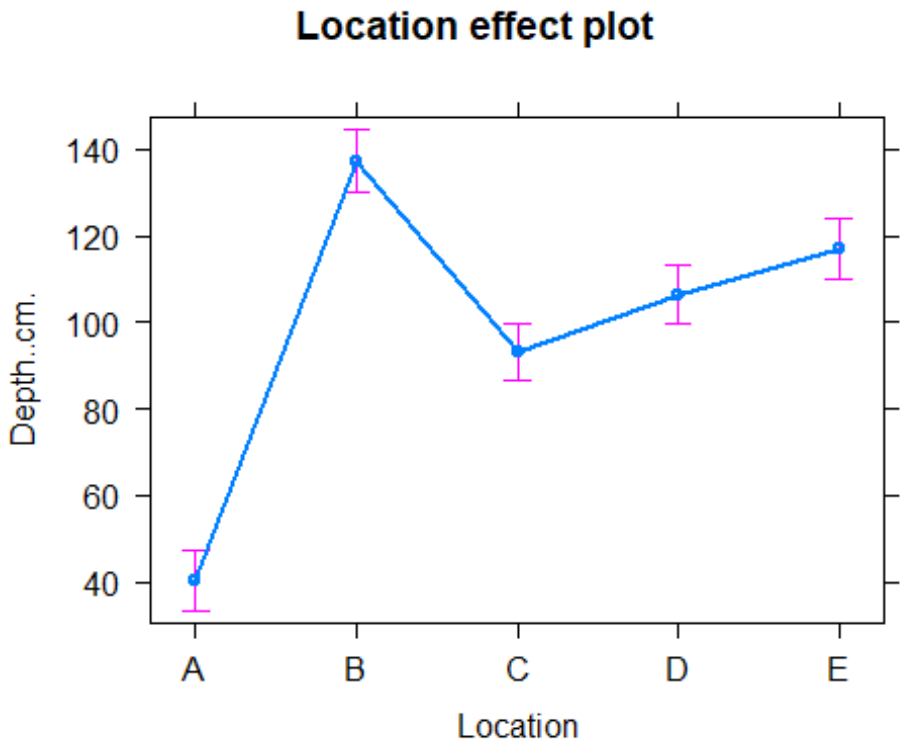

Figure 12. The estimated mean depth at different locations across the weeks

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Depth..cm. ~ Location, data = larvae)
##
## $Location
##          diff         lwr        upr     p adj
## B-A  96.97181  82.9064261 111.037201 0.0000000
## C-A  52.86765  39.5465317  66.188762 0.0000000
## D-A  66.09913  52.4422170  79.756040 0.0000000
## E-A  76.67320  63.0162911  90.330114 0.0000000
## C-B -44.10417 -57.6484318 -30.559901 0.0000000
## D-B -30.87269 -44.7473468 -16.998024 0.0000000
## E-B -20.29861 -34.1732728  -6.423949 0.0007253
## D-C  13.23148   0.1119074  26.351056 0.0469390
## E-C  23.80556  10.6859815  36.925130 0.0000112
## E-D  10.57407  -2.8863249  24.034473 0.1992074
```

The results show that there is a significant difference between locations in terms of the distance to the water surface, with Location A having water significantly closer to the surface compared to all other locations, and significant differences between all locations except between locations D and E.

Nutrient analysis

In the Fall of 2019, water samples were transported to EWU. Starting at the end of Fall quarter, the graduate student, with the help of Dr. McNeely, worked extremely hard to analyze the levels of nitrate, ammonium and phosphate in a subset of 179 water samples collected. He also extracted the DNA from the same water samples, and stored them in the freezer.

In Spring 2020, the plan was to complete the nutrient analysis with Dr. McNeely, and process the DNA extracted from the water samples for next-generation sequencing, with the help of Dr. Walke. Of course, COVID hit, and we weren't allowed to enter our labs to complete these analyses, and then the graduate student left our program.

Five nutrient measurements were taken in the lab: (1) SRP; (2) Total Phosphorus; (3) NH4; (4) Nitrate; and (5) Total nitrogen. First, we conducted exploratory data analysis to see if we can discover differences between locations and between weeks in the nutrients.

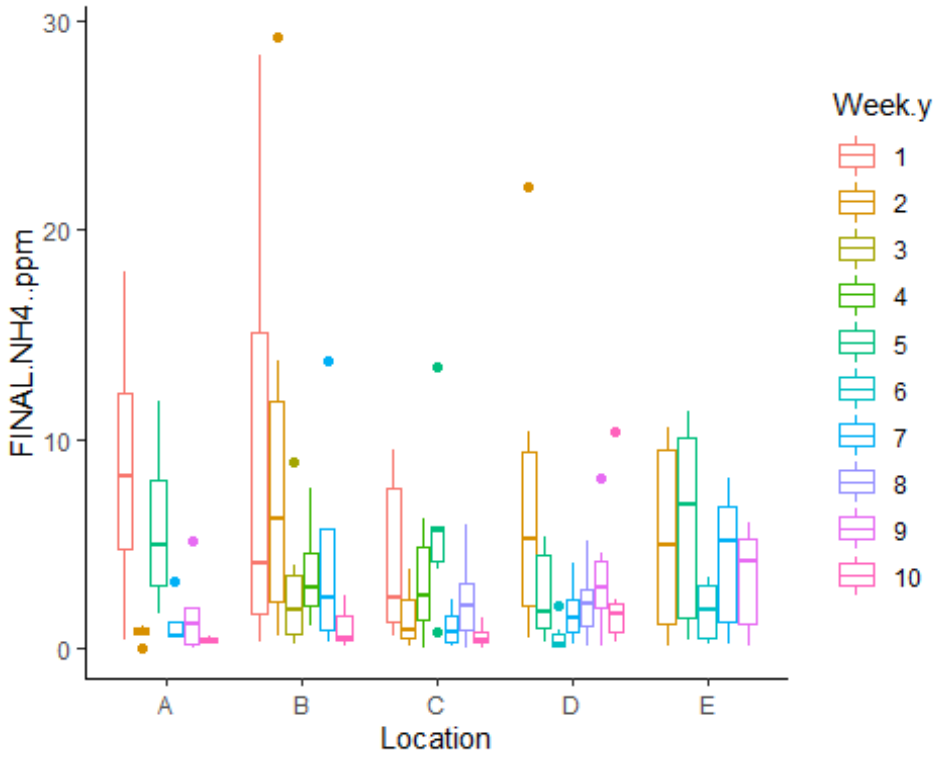

Figure 13. The distribution of NH4 concentration in water samples across weeks and locations.

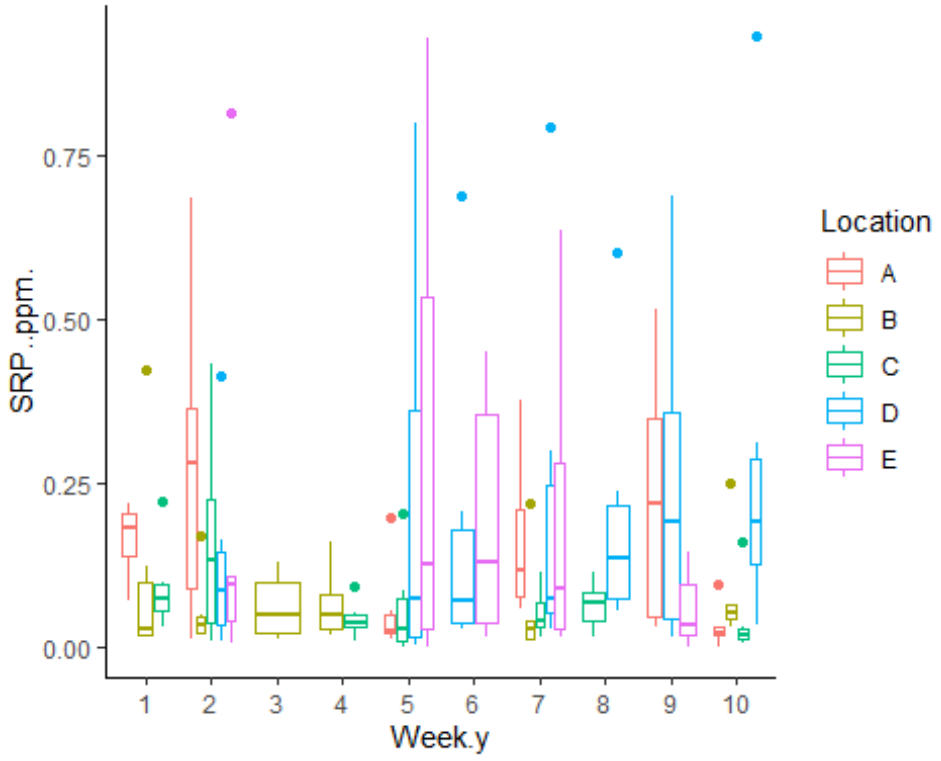

Figure 14. The distribution of SRP concentration in water samples across weeks and locations.

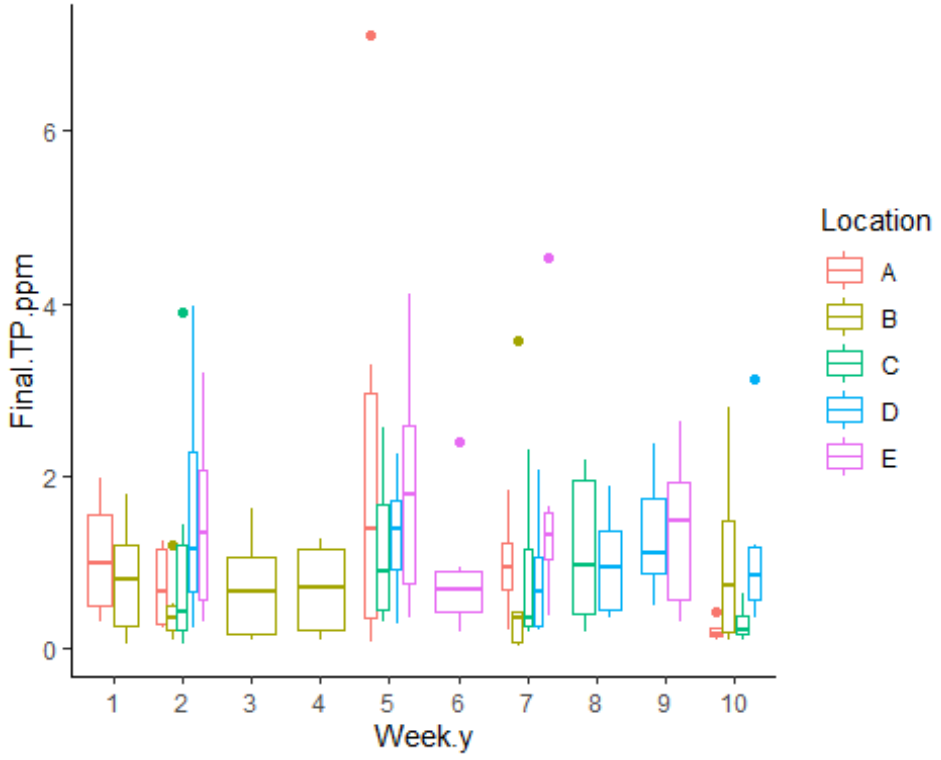

Figure 15. The distribution of TP concentration in water samples across weeks and locations.

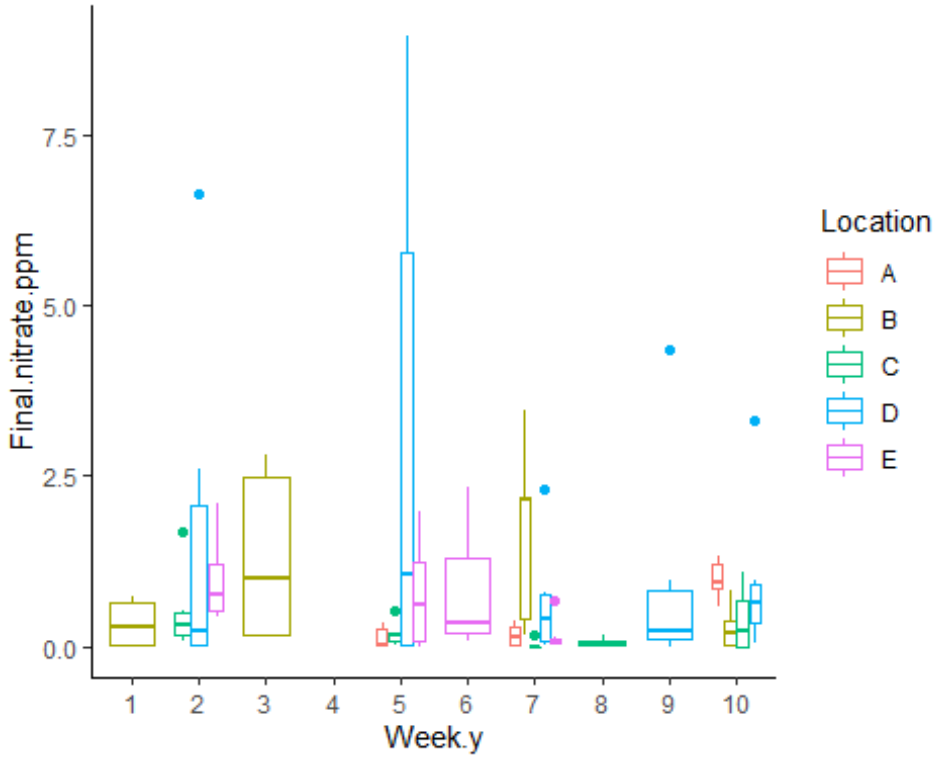

Figure 16. The distribution of NO3 concentration in water samples across weeks and locations.

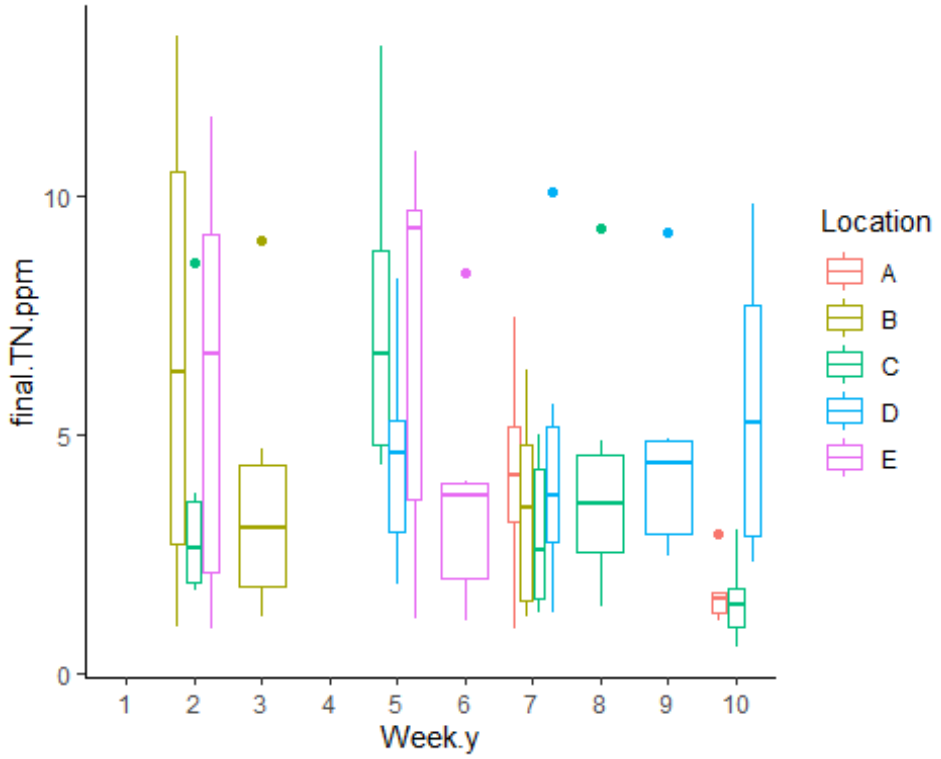

Figure 17. The distribution of TN concentration in water samples across weeks and locations.

The exploratory graph do not suggest a clear difference in any of the nutrient concentrations between the locations or weeks. In order to confirm, we will test the null hypothesis that there is no difference in the mean concentration of these nutrients between locations and weeks, using multiple regression assuming normality for the response variable. Because all of the nutrients have a skewed distribution, we will log-transform each of them, and add 1 before the log-transformation, to improve normality.

NH4

```
## Anova Table (Type III tests)
##
## Response: logNH4
##                  Sum Sq  Df F value    Pr(>F)
## (Intercept)      17.099   1 26.8726 6.181e-07 ***
## Location          1.269   4  0.4986    0.7368
## Week2             2.698   1  4.2409    0.0410 *
## Location:Week2    0.771   4  0.3031    0.8756
## Residuals       106.898 168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

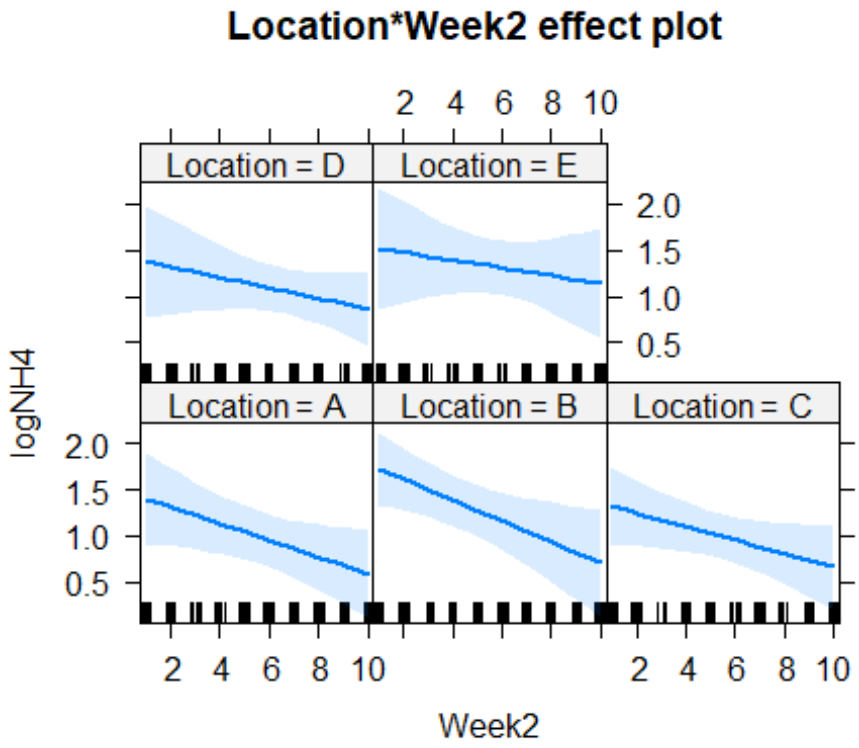

Figure 18. The relationship between NH4 concentration in water samples and the week of the study across the different locations.

Based on these results, there does not seem to be a significant difference in NH4 concentration between different locations. NH4 concentration does seem to significantly decrease with increasing weeks. However, location and week of the study only explains 6.9662368% of the variation in NH4 concentration.

NO3

```
## Anova Table (Type III tests)
##
## Response: logNitrate
##                 Sum Sq  Df F value  Pr(>F)
## (Intercept)     0.3045   1  1.3861 0.24158
## Location        2.1223   4  2.4153 0.05301 .
## Week2           0.8779   1  3.9965 0.04804 *
## Location:Week2  1.5837   4  1.8023 0.13342
## Residuals      24.3841 111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

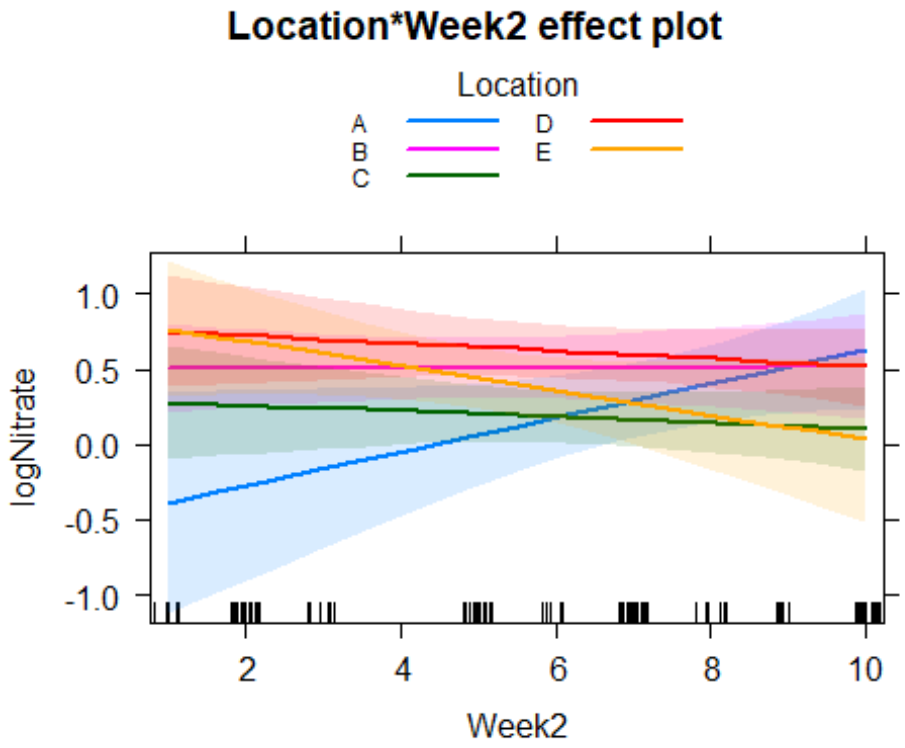

Figure 19. The relationship of NO3 concentration in water samples across weeks and locations.

Based on these results, there does not seem to be a significant difference in NO3 concentration between different locations. NO3 concentration does seem to significantly increase with increasing weeks for location A. However, location and week of the study only explains 9.7925552% of the variation in NO3 concentration.

TN

```
## Anova Table (Type III tests)
##
## Response: logTN
##                 Sum Sq Df F value  Pr(>F)
## (Intercept)     2.0081  1  6.2948 0.01396 *
## Location        0.4393  4  0.3443 0.84733
## Week2           0.6545  1  2.0517 0.15562
## Location:Week2  0.9547  4  0.7482 0.56188
## Residuals      27.7546 87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

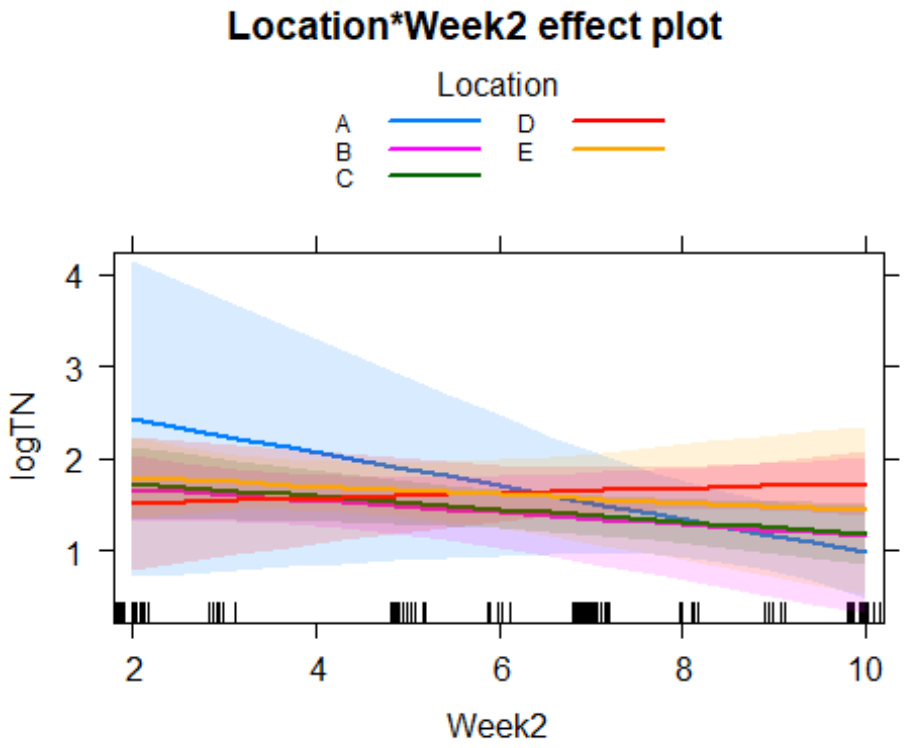

Figure 20. The relationship of TN concentration in water samples across weeks and locations.

Based on these results, there does not seem to be a significant difference in TN concentration between different locations. TN concentration does not seem to significantly change with increasing weeks. However, location and week of the study only explains 4.4555695% of the variation in TN concentration.

SRP

```
## Anova Table (Type III tests)
##
## Response: logSRP
##                Sum Sq  Df F value    Pr(>F)
## (Intercept)    0.2772   1 14.1743 0.0002301 ***
## Location       0.1286   4  1.6442 0.1654531
## Week2          0.0312   1  1.5935 0.2085694
## Location:Week2 0.1079   4  1.3799 0.2429971
## Residuals      3.2853 168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

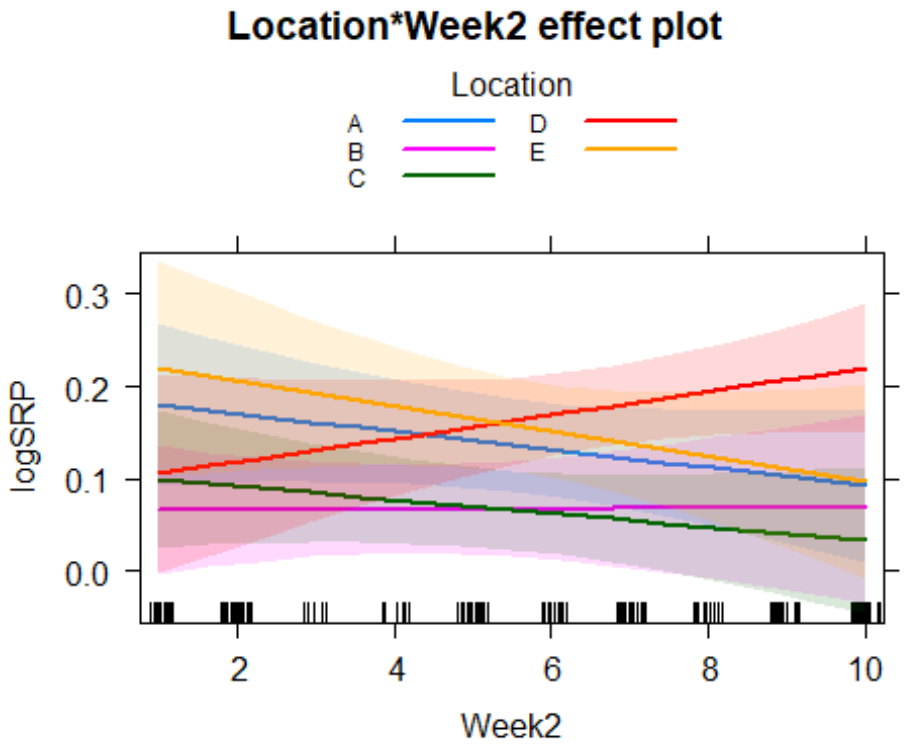

Figure 21. The relationship of SRP concentration in water samples across weeks and locations.

Based on these results, there does not seem to be a significant difference in SRP concentration between different locations. SRP concentration does not seem to significantly change with increasing weeks. However, location and week of the study only explains 9.0011391% of the variation in SRP concentration.

TP

```
## Anova Table (Type III tests)
##
## Response: logTP
##                 Sum Sq  Df F value    Pr(>F)
## (Intercept)     4.0793   1 22.6377 4.706e-06 ***
## Location        1.0296   4  1.4284    0.2276
## Week2           0.2382   1  1.3217    0.2522
## Location:Week2  0.4136   4  0.5738    0.6821
## Residuals      25.9487 144
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

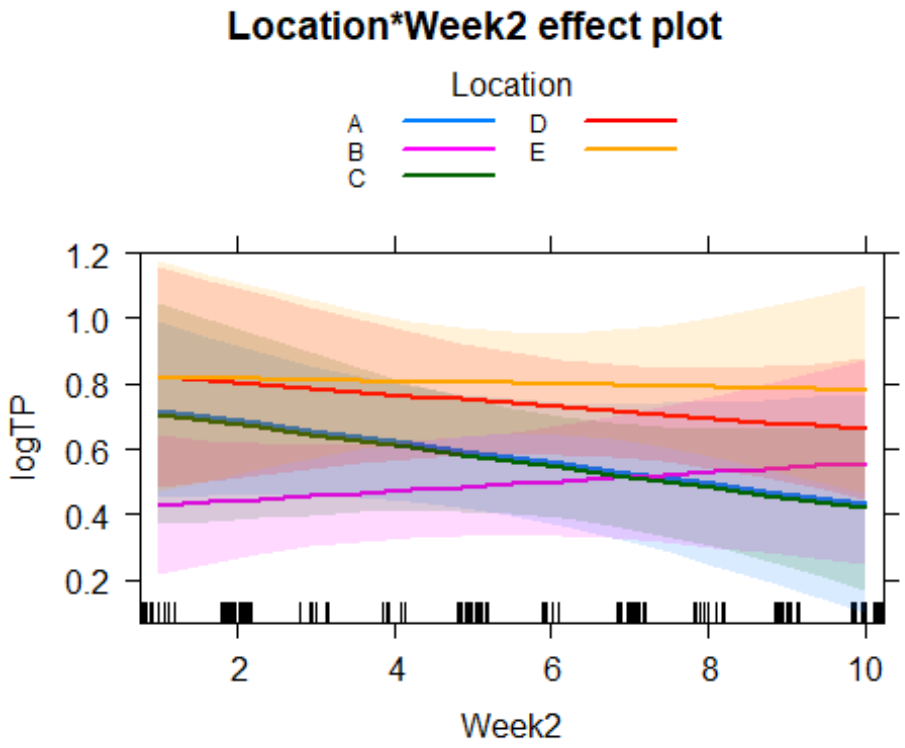

Figure 22. The relationship of TP concentration in water samples across weeks and locations.

Based on these results, there does not seem to be a significant difference in TP concentration between different locations. TP concentration does not seem to significantly change with increasing weeks. However, location and week of the study only explains 4.2524899% of the variation in TP concentration.

Can we explain the number of larvae based on Depth and nutrients alone?

The objective of our study was to develop a statistical model that can predict the number of mosquito larvae to be found in a catch-basin or culvert based on the characteristics of the storm-water structure, such as the amount of water in it and the level of nutrients. In order to achieve this, we searched for the best model to explain the number of mosquito larvae collected, using depth and the available nutrient measurements. We deliberately did not include Location and Week in the model, as we wanted to see how well we could predict mean mosquito larval density without them. We used the logarithm of NH4, NO3, and SRP concentration, as well as Depth, as predictor variables. We restricted our dataset for 118 sampling events where we had all of these information available. We used a Poisson regression as the number of larvae is a count variable. We conducted all subsets model selection on this complete dataset, which compared all possible subsets of models in terms of AICc. The best model was the full model with all four predictors and all their interactions, which we then re-ran as a quasi-Poisson model.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	117	5012.099	NA
Depth..cm.	1	1060.473030	116	3951.626	0.0000000
logNH4	1	10.422687	115	3941.203	0.5730198
logNitrate	1	24.161478	114	3917.041	0.3908237
logSRP	1	310.383555	113	3606.658	0.0021004
Depth..cm.:logNH4	1	167.555814	112	3439.102	0.0238342
Depth..cm.:logNitrate	1	465.250080	111	2973.852	0.0001662
Depth..cm.:logSRP	1	1045.787150	110	1928.065	0.0000000
logNH4:logNitrate	1	1.043669	109	1927.021	0.8584491
logNH4:logSRP	1	74.492230	108	1852.529	0.1318722
logNitrate:logSRP	1	54.448516	107	1798.080	0.1976781
Depth..cm.:logNH4:logNitrate	1	18.708159	106	1779.372	0.4501893
Depth..cm.:logNH4:logSRP	1	6.482031	105	1772.890	0.6567015
Depth..cm.:logNitrate:logSRP	1	15.942691	104	1756.948	0.4857660
logNH4:logNitrate:logSRP	1	166.631103	103	1590.316	0.0242247
Depth..cm.:logNH4:logNitrate:logSRP	1	3.281370	102	1587.035	0.7518207

As you see in the above table, both Depth and SRP are significant predictors, as well as interactions of Depth with all other predictors, and a three-way interaction of all predictors except for Depth. However, any of the terms of the model, except for the 4-way interaction, results in a model with much higher AICc values. This model explains 68.3359168% of the variation in the mean number of larvae collected per collection session.

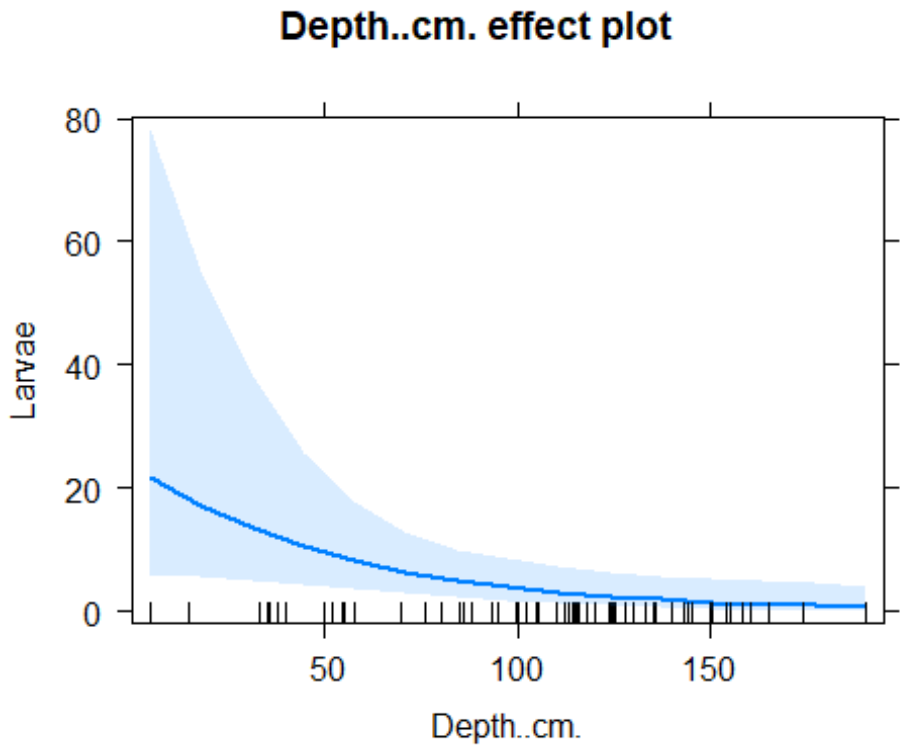

Figure 23. The relationship between the the number of larvae collected per collection session and the Depth to the surface of the water body.

The model suggests that the mean number of mosquito larvae collected significantly decreases as Depth increases, meaning when the distance is larger to the water surface, and there's presumably less water in the stormwater structure. This agrees with the finding that Location A and C had significantly more mean larval density, as Location A and C also had significantly lower Depth values.

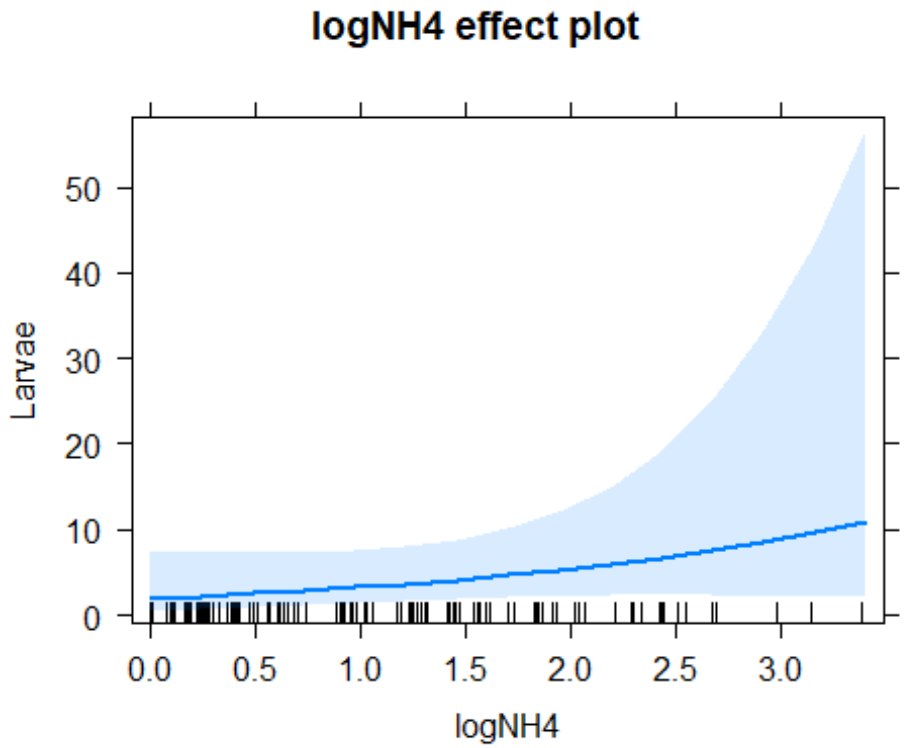

Figure 24. The relationship between the the number of larvae collected per collection session and the logarithm of NH4 concentration in the water body.

The model suggest that there is not a significant relationship between the logarithm of NH4 and mean larval density. There is a non-significant positive trend, which would make sense as NH4 concentrations significantly dropped across the weeks, while the larval density also dropped throughout the study. However, this relationship is potentially confounded with interaction with Depth.

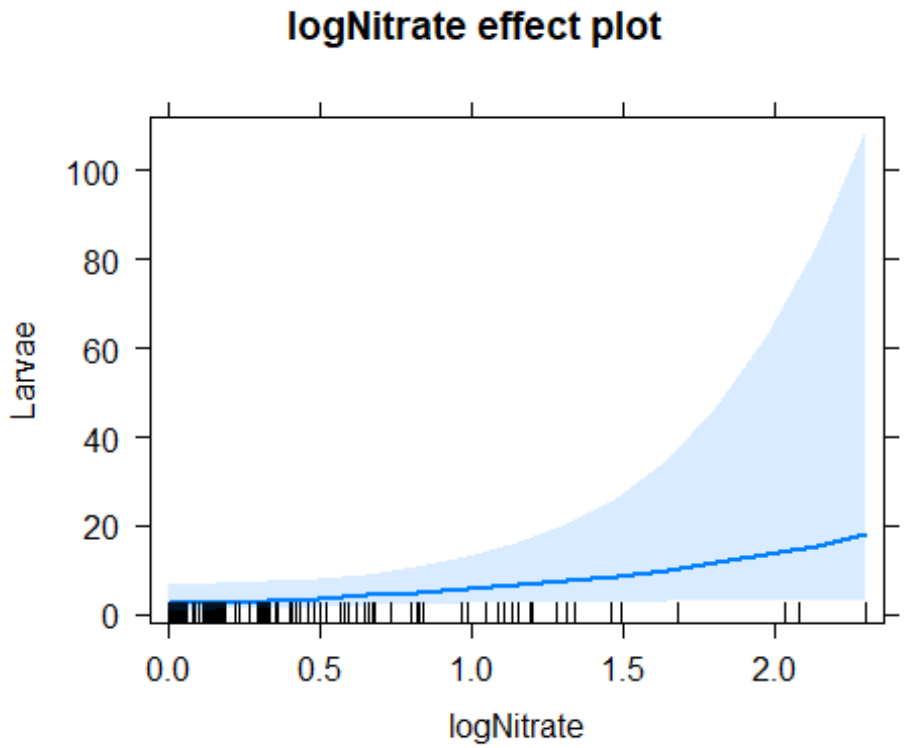

Figure 25. The relationship between the the number of larvae collected per collection session and the logarithm of the NO3 concentration of the water body.

The model suggest that there is not a significant relationship between the logarithm of NO3 and mean larval density. There is a non-significant positive trend, which does not make sense as NO3 concentrations slightly increased across the weeks at Location A, while the larval density also dropped throughout the study. However, this relationship is potentially confounded with interaction with Depth.

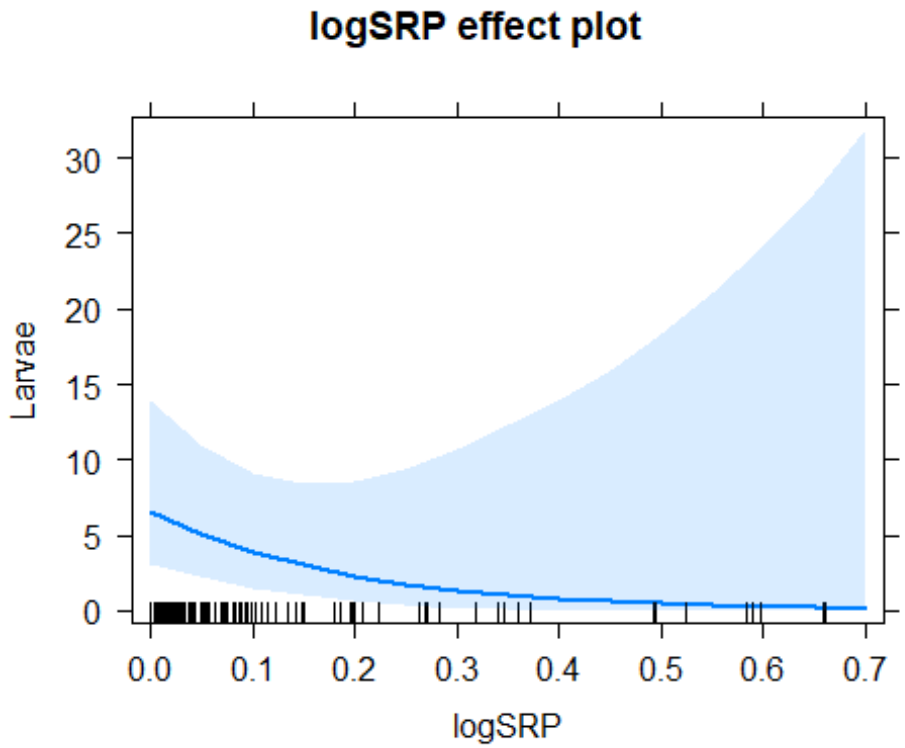

Figure 26. The relationship between the the number of larvae collected per collection session and the logarithm of the SRP concentration of the water body.

The model suggest that there is a significant relationship between the logarithm of SRP and mean larval density. There seems to be a significant negative relationship, which is difficult to explain as SRP concentrations did not significantly differ between locations or weeks. However, this relationship is potentially confounded with interaction with Depth.

Discussion

Given the lack of significant differences in nutrient levels between locations and weeks, it was pleasantly surprising how much percentage of the variation in larval density was explained by differences in nutrient levels and water level (about 68%). This is about 40% higher than the variation that is explained by simply the location and the week of collection. Given this accuracy, the statistical model could be used to predict mean larval density at combinations of depth and different nutrients. However, there is a remaining 32% of the variation that is unexplained by nutrient levels and depth, and might have to do with characteristics we haven't incorporated into this analysis.

One of those characteristics is the microbiome of the storm-water catch basin. There was considerable uncertainty during the Summer of 2020 on when we can re-enter our labs, and both Dr. Walke and the undergraduate research assistant were unable to complete the laboratory procedures. In the meantime, Dr. Andrade, our collaborator at Gonzaga, was able to complete DNA extraction from mosquito larvae from 2019, and transferred those to

us for the PCR step before sending it off for next-generation sequencing. In Fall 2020, we regrouped to complete the project to process the DNA samples for the next-generation sequencing starting the middle of November. The undergraduate student under Dr. Walke's guidance completed the processing the samples and sent them off for next-generation sequencing by the end of 2020, and sent us the results of the microbiome testing.

In this study, we measured larval density in storm-water infrastructure, such as catch basins and culverts. However, larval density does not necessarily correlate well with adult mosquito productivity. It is possible for a breeding site to host large numbers of first instar larvae, and then fail to produce proportional number of adults. For example, our collection site A5 have produced 500 first instar larvae on August 19, 2019. While this is a large number, there is no guarantee that they turn into 500 adult female mosquitoes. The number of juvenile larvae at any given time is a function of the attractiveness of the breeding site for ovipositing female mosquitoes, as well as the survival of the larvae due to nutrients, other abiotic characteristics as well as density itself. Our collections were basically constituting a weekly snapshot at each of the locations, and measured a combination of the preference of female ovipositing mosquitoes to breeding sites, as well as the survival of juvenile mosquito larvae. In the future, building on this study, we will place emergence traps made out of wire cones on top of each breeding site, in order to measure their adult productivity. That will allow us to determine the relationship between larval density and adult productivity, as well as productivity and abiotic conditions, such as nutrients, specifically.

In this study, we assumed that mosquito larvae breeding in each of the breeding sites comprised a single species, such as Culex pipiens. While this is most likely the case, we did not grow up every single mosquito larvae that we collected. Therefore, it is possible that some mosquitoes we collected belonged to different species. Different species have different environmental constraints, which would complicate the task of identifying the ecological niche of a single mosquito species.

At the outset of this study, we started with the assumption that Culex pipiens are limited by nutrients in potential breeding sites in stormwater structures. One of the main results of or study is that we documented the presence of mosquito larvae in every single catch basin and culvert inspected, demonstrating their ubiquitous presence across the City of Spokane. These locations have been selected based on advice from the Public Works Stormwater Department of the City of Spokane, including locations with previous history of mosquito collections in the past. We can hypothesize that these locations are not particularly special, and that mosquito larvae are breeding in every stormwater structure across the City. We will test this hypothesis both in the City of Spokane and the City of Cheney by sampling stormwater catch basins and culverts for mosquitoes in a follow-up study.

One potential conclusion of our study is that, as opposed to our assumption of a limiting nutrient level with a threshold above which mosquitoes can survive, mosquito larval density is instead determined as a complex function of the combination of a range of environmental conditions, such as water level and nutrient levels. It is possible that there is no combination below which mosquito larval density will be absolutely zero, as ovipositing female mosquitoes may lay their eggs into conditions that are not optimal for the survival

and development of their offspring. It is also possible that there is not a single optimal combination of environmental conditions that will maximize mosquito larval density (and productivity), but rather that there is a range of combinations that provide similar conditions. Chance events, such as a female mosquito ovipositing into a potential breeding site, or rain event flushing it, can introduce environmental stochasticity that could then increase the variability in the quantity of larval density, making it look like their might be more variability in environmental suitability. Separating such environmental stochasticity from environmental suitability will require a different conceptual framework and study design, potentially including experiments, which will build on the results of this seed grant.