Ecological Data Analysis

Introduction:

Aquatic habitats are distributed across the face of the earth, facing many challenges that are not found in terrestrial environments. It is without question that human culture has undoubtedly had a negative impact on biodiversity, especially since the Industrial Revolution. Overfishing and hunting, habitat loss due to agriculture and urban sprawl, pesticide and herbicide use, and the release of other toxic substances into the environment have all affected vertebrates. In urban and suburban areas, where they flourish in close proximity to humans, many animal and plant species have adapted to new stresses, food sources, predators, and threats. Their success gives researchers useful, and often surprising, insights into evolutionary and selection processes. This begs the question, is there a difference in the number of aquatic habitats in Natural areas compared to Human Settlements?

Method of Investigation:

In order to accurately determine a conclusion, we must conduct an aquatic habitat survey. Within this, we investigate two categories of locations: human settlements and natural settlements. Further within these locations lies two types of ecological locations. Within human settlements, we find oil palms and logged forests settlements, while in natural settlements, we find continuous forests and forest fragments. By studying and gathering data from each respective area, we are therefore able to provide reasonable evidence to support or fault the hypothesis.

The data collected ranged from simply figuring out the number of aquatic habitats distributed about a location, conducting a larval survey and determining larval presence and abundance, and the most painful of all, the adult population survey. The APS consisted of surveyors acting as bait to mosquitoes and collecting them by electric aspirators. This allows the mosquito abundance in a location to be determined, which would correlate to an abundance of aquatic habitation. All three methods were included for maximum accuracy and effect, to ensure that more than one outcome could be made to support our conclusion.

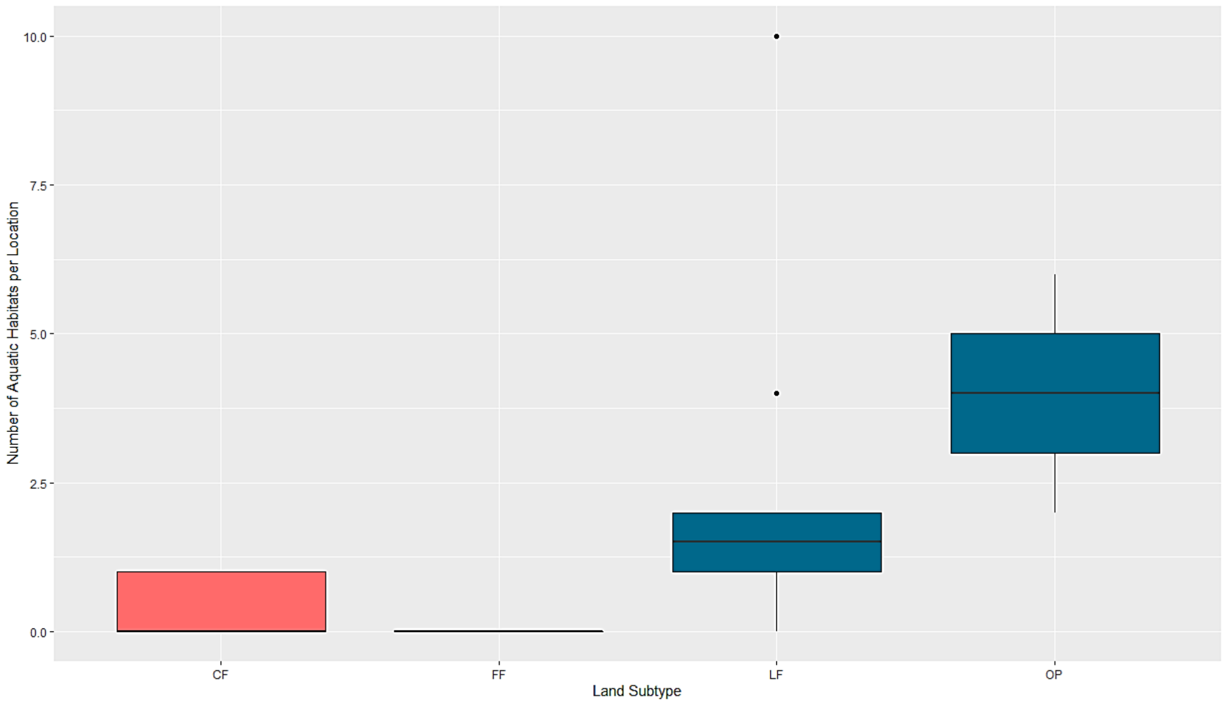
The data was collected, and visual representations of the data were generated, specifically in the form of box plots. The best possible representation of data in this context was a box plot, as we were able to compare relative medians to choose which area had more aquatic ecosystems on aggregate and interquartile ranges to see which consisted of a higher distribution than other areas. The graphical representation allows us to assess the differences at ease, without needing complicated models.

However, to determine the difference in the number of aquatic habitats in each respective settlement, we opted for a Generalized Linear Model. These models are based on Linear Mixed Models, which account for fixed and random effects. However, they assume the response variable is normally distributed, which GLMMs do not assume so. This is especially useful for ecological data, and an appropriate distribution for the response variable can be used instead.

In this case, the Poisson Distribution was used, which models the number of events occurring in a given time interval. This allowed an easier comparison of the number of aquatic habitats at different times of the day. The Poisson distribution helps us to make predictions on how many more aquatic habits there will be in the coming years, which is crucial in providing important ecological data in this case. As previously mentioned, the Poisson distribution is used for count data and is, therefore, more appropriate since it can take any value.

Results:

In order to compare the number of aquatic habitats, we can initially generate a box plot and visually compare the ranges and medians of each land subtype. While the initial prompt asked to show the comparison of land types, I have chosen to differentiate between subtypes instead, since this produced similar results and allows for a clearer comparison with reasoning.



We can see that the natural settlements land types (continuous forest and fragmented forest) consist of significantly lower numbers of aquatic habitats compared to human settlement land types (logged forest and oil palm).

Statistically, the oil palm consists of the largest interquartile range and median, while the lowest median and interquartile range are found in the natural settlements, specifically the fragmented forest. Since the median value of fragmented forest seems to be negligible, we can therefore assume that rarely any aquatic habitats are found in natural settlements, contrary to popular belief.

While this may quickly disprove the statement that natural settlements have more aquatic habitats than human settlements, it is important to provide further evidence despite the convincing evidence we see here.

To do so, a Generalized Linear Model was produced:

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod
]
 Family: poisson  ( log )
Formula: Number_aquatic_habitats ~ Land_type + (1 | Land_type/Land_subtype/Replicate)
   Data: AHS_abundance

     AIC      BIC   logLik deviance df.resid
   140.3    148.8    -65.1    130.3       36

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.5262 -0.6105 -0.3297  0.6715  4.0167

Random effects:
 Groups                              Name        Variance  Std.Dev.
 Replicate:(Land_subtype:Land_type) (Intercept) 3.153e-02 1.776e-01
 Land_subtype:Land_type             (Intercept) 2.619e-02 1.618e-01
 Land_type                          (Intercept) 2.578e-09 5.077e-05
Number of obs: 41, groups:
Replicate:(Land_subtype:Land_type), 12; Land_subtype:Land_type, 3; Land_type, 2

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.1523     0.1804   6.387 1.69e-10 ***
Land_typeN   -2.1585     0.4819  -4.479 7.50e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
           (Intr)
Land_typeN -0.363
optimizer (Nelder_Mead) convergence code: 0 (OK)
boundary (singular) fit: see ?isSingular
```

Using the results from the generalized linear model, we can identify once more that the number of aquatic habitats in human settlements is much greater than those in natural settlements. This is proven by the p values and the z values.

Discussion:

While one may initially predict that natural settlements have more aquatic habitats due to the actions and destruction of nature in human settlements, the investigation proves to not be as confident with assumptions. One may argue about the consequences that human settlements have to wildlife, but disregard the benefits that they also bring.

Green roofs and other green spaces within cities form ecological networks that provide a versatile habitat for birds, insects, and some plants on a small total surface area. Green buildings may also be beneficial outside of cities by reducing the effects of obstacles to animal and plant movement, such as roads and railways.

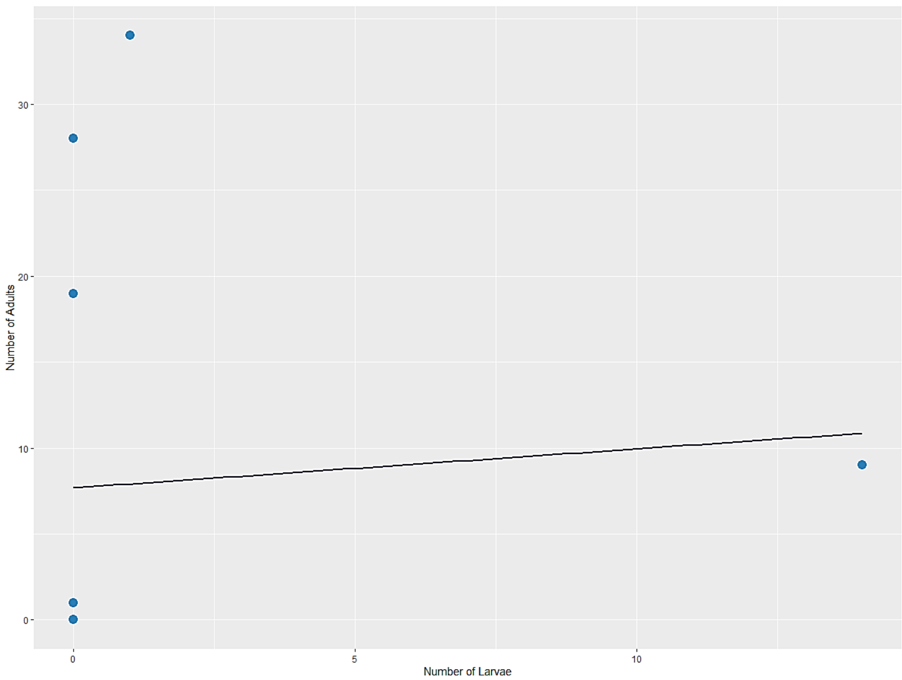
In comparison to natural sites, human settlements will have:

1. More larval habitats
2. A greater abundance of larvae
3. A higher density of adults than natural forest during an extended period of drought

We can resolve some misconceptions by looking at the effect of mosquitoes on aquatic habitats. As we may know, part of the mosquito life cycle is aquatic, meaning that one target for control methods is the aquatic stage. This is due to the availability of aquatic habitats being an unstable abiotic factor, influencing adult density. Albopictus, for example, prefers artificial larval habitats as opposed to natural habitats like tree holes and bamboo stumps. This simple preference can have a great deal of consequence to others in the ecosystem. As we see later with the data provided, it is no surprise that the albopictus has a direct correlation to aquatic habitat population.

Extension 1:

A further investigation of importance would involve discovering the number of adult albopictus in locations with albopictus larvae. The question gives, is the more adult albopictus in locations with more albopictus larvae, and is there a correlation between the two? To determine, we can plot a scatter graph to visually describe a correlation between larvae and adults on each axis.



As seen from the figure above, there is a weak correlation between the number of larvae and the number of adult albopictus. Although a regression line can be seen in the scatter plot below, nearly half of the data points are anomalous. As a result, it may be argued that there is no real correlation between the number of albopictus and larvae, but this is only concerning a limited sample size.

In order to determine if there is a significant correlation between the two statistically, we can run a Spearmann's Rank Correlation Coefficient test. This measures the strength and direction of the monotonic relationship between two ranked variables, in this case, larvae and adult albopictus population. This is useful since the data is non-parametric, data does not have to follow a specific distribution, and the test can be used broadly since the relationship does not need to be linear.

```
        Spearman's rank correlation rho

data:  ALAb$Total.Albo.Larvae and ALAb$Total.Adults
S = 240.69, p-value = 0.08914
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4710003
```

From running the test, we gather a p-value of 0.08914 and a rho value of 0.471. The rho value being almost +1 gives an association between the larvae and albopictus population, while the p-value being above 0 allows us to conclude that there is some correlation between the two variables.

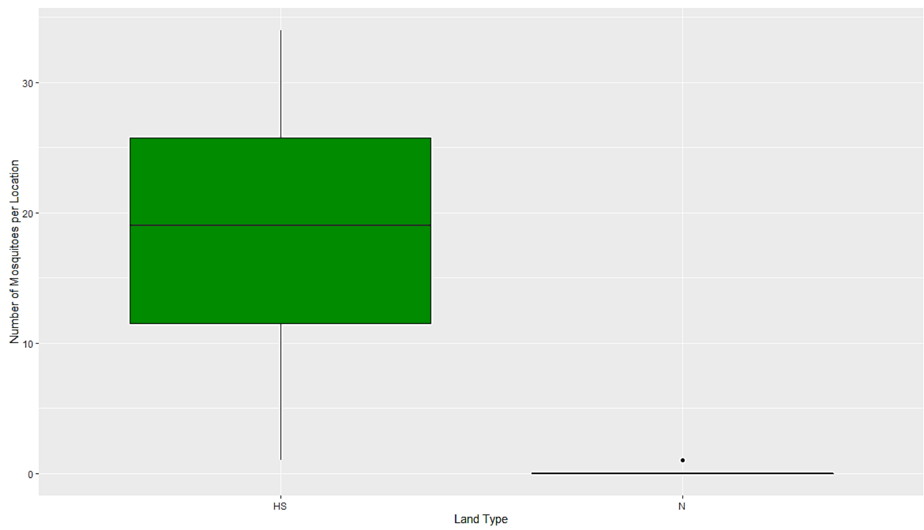
Out of interest, a test with axes being reverse can be performed.

```
data:  ALAb$Total.Adults and ALAb$Total.Albo.Larvae
S = 240.69, p-value = 0.08914
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4710003
```

The values are identical and prove that the placement of variables does not affect the overall outcome. As a result, we can safely finalize the judgement that the number of albopictus larvae correlates with the population of albopictus.

Extension 2:

A further investigation on the mosquitoes can also be done to further support our overall conclusion. We can use the results obtained from the surveys to explore whether there are more adult mosquitoes in human settlements as opposed to natural settlements. We can represent the data from each land type in a box plot to initially visualize the differences between the two.



As seen, we see a significantly greater number of mosquitoes in human settlements as opposed to natural settlements. We can also analyse the model using the summary() function and concluding p-values and a GLMM:

```
Call:
glm(formula = Total.Adults ~ Land.Type, family = poisson(link = "log"),
    data = ALAb)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.3711  -0.5000  -0.5000   0.1548   3.2659

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.90872    0.09535  30.507  < 2e-16 ***
Land.TypeN  -4.98816    1.00434  -4.967 6.81e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 230.078  on 13  degrees of freedom
Residual deviance:  53.964  on 12  degrees of freedom
AIC: 86.145

Number of Fisher Scoring iterations: 5
```

From this, we see that there are more mosquitoes in human settlements from the p values and z values. A possible reason for this is the tendencies of mosquitoes. Mosquitoes prefer stagnant and places with an unpleasant odour. This is the sort of aquatic area found in human settlements. Since aquatic areas are a result of mass destruction of an area and consists of remains and industrial waste, it provides the ideal habitat for mosquitos to survive. The findings also indicate that human settlements have a much higher risk of disease transmission than natural areas, owing to the higher number of mosquitos and unsanitary conditions present in natural areas due to human involvement.

As a result of the direct correlations and links we can draw from mosquitoes and aquatic habitats, we can understand and explain how and why aquatic habitats are more frequently found in natural settlements compared to human settlements. It is our biological knowledge and use of data that allows us to appreciate and fully understand why and how ecosystems function and differ from each other. By using software like R, and recording data in Google Sheets, we can correlate and compare to create these differences.

Final Thoughts:

Through completing this task, I began to appreciate the importance of data analysis and how it can be used in the wider context. I used what I had learned from many seminars such as those on Population Dynamics, to compare and justify what I had seen from the code I had produced. I should also take the time to thank Maisie Vollans for the introduction to R, and the easy step by step guide to appreciating the task she laid out for us. While the task was relatively challenging, it opened my eyes to a whole world of biological data I would have never appreciated before.

Final Code: (Attached to Email)

```r
# Clear the environment

rm(list = ls())

# Installing and calling all functions/packages


library(gsheet)
library(lme4)
library(dplyr)
library(stats)
library(ggplot2)

# Importing data for analysis

AHS_abundance<- gsheet2tbl('https://docs.google.com/spreadsheets/d/1dqcFrvIfF2WVPrKA83ExjGnxAiVKlloP/edit#gid=228026339')
View(AHS_abundance)

AMD<- gsheet2tbl('https://docs.google.com/spreadsheets/d/1Rqp9Bvi9K5-asZhSynNrUxYsNVSW4IWm/edit#gid=2027687899')
View(AMD)

ALAb<- gsheet2tbl('https://docs.google.com/spreadsheets/d/1BHoLClHQ46arFSaRhfIAw9yJq9UDDsI6/edit#gid=2034489775')
View(ALAb)

# Creating the boxplot

AHS_abundance_boxplot <- ggplot(AHS_abundance, aes(x=Land_subtype, y=Number_aquatic_habitats, fill=Land_type)) +
  geom_boxplot() +
  xlab ("Land Subtype") +
  ylab ("Number of Aquatic Habitats per Location") +
  scale_fill_manual(values=c("deepskyblue4", "indianred1")) +
  theme(legend.position = "none")

AHS_abundance_boxplot

AHS_abundance<- filter (AHS_abundance, Land_subtype != "FF")

# Creating the model

model1<- glmer(Number_aquatic_habitats ~ Land_type + (1 | Land_type/Land_subtype/ Replicate), AHS_abundance, family = poisson(link = "log"))

# Outputting the model

summary(model1)

#### Extensions ####

# Extension 1; Correlation between Albo Larvae and Adult Albo

ggplot(ALAb, aes(x=Total.Albo.Larvae, y=Total.Adults)) +
  xlab ("Number of Larvae") +
  ylab ("Number of Adults") +
  geom_point(color='#2980B9', size = 4) +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE, color='#2C3E50')

# SPRC

cor.test(y=ALAb$Total.Albo.Larvae, x=ALAb$Total.Adults, method = 'spearman', exact = FALSE)


# Extension 2; Adult Mosquitoes within Landtypes

ALAb_boxplot <- ggplot(ALAb, aes(x=Land.Type, y=Total.Adults, fill=Land.Type)) +
  geom_boxplot() +
  xlab ("Land Type") +
  ylab ("Number of Mosquitoes per Location") +
  scale_fill_manual(values=c("green4", "cadetblue4")) +
  theme(legend.position = "none")

ALAb_boxplot

model2<- glm (Total.Adults ~ Land.Type, ALAb, family = poisson(link = "log"))

summary(model2)
```

Download at: https://tinyurl.com/umarsiddique