Examining Tickborne Disease Prevalence and the Effects Habitat and the Microbial Community Play

Julia Frederick

2019-10-22

# Summary/Abstract

*Write a summary of your project.*

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# Introduction (required for part 1)

## General Background Information

*Provide enough background on your topic that others can understand the why and how of your analysis*  
Tickborne disease prevalence is on the rise in the United States with 50,000 cases being confirmed by the CDC each year. Lyme disease alone is the number one vector borne disease reported in the U.S. each year with over 30,000 cases, and the Center for Disease Control and Prevention (CDC) estimating over 300,000 cases actually occurring (Vector-Borne Disease, 2019). A single species of tick is capable of transmitting multiple disease pathogens (Adelson et al., 2004). For example, Ixodes scapularis has a range covering the entire eastern half of the U.S. and is capable of transmitting pathogens that cause Lyme disease, anaplasmosis, ehrlichiosis, babesiosis, and Powassan virus (Adelson et al., 2004; Dantas-Torres, Chomel, & Otranto, 2012). There are multiple tick species native to the U.S. that can spread at least one human disease, and the consequences of increased globalization and livestock trade are becoming apparent with invasive species making landfall (Jongejan & Uilenberg, 2004). When investing multiple tick species, it becomes clear that each species has its own variation in lifecycle and disease prevalence that comes along with it. Knowing what these differences are give researchers the power to create control and prevention efforts for human disease. Current vector research is focusing on understanding the biology of these vectors in order to put future range expansions (Ginsberg et al., 2017), and increased disease threat (Arsnoe, Hickling, Ginsberg, McElreath, & Tsao, 2015; Eisen & Eisen, 2018) into a clearer context. Our research aims to define trends over time in tick species looking into pathogen prevalence, and microbiome shifts. This information will inform future efforts to model disease expansion across the U.S. and into Canada.

## Description of data and data source

*Describe what the data is, what it contains, where it is from, etc.*  
Through the Southeastern Coopertative Wildlife Disease Study (SCWDS) ticks were collected from the eastern region of the United States over the course of a year, the ticks used in this study were all from Georgia. When collected the habitat type and location was recorded for future reference. These ticks were then IDed, and used for 16S and PCR in order to determine the presence of pathogens and the microbial community within each specimen. The pathogen and microbial community will not be investigated genetically, instead we will be determining changes in prevalence over time and space. The data used for this analysis includes the pathogens *Rickettsia*, *Ehrlichia ewingii*, *Ehrlichia chaffeensis*, *Anaplasma phagocytophilum*, *Borrelia spp*, and Panola Mountain Ehrlichia (PME). There are multiple tick species that were identified through morphology, and the life stage the tick was collected. The habitat information includes: season, region, site, and general habitat.

## Questions/Hypotheses to be addressed

*State the research questions you plan to answer with this analysis*  
**1. How the habitat effects the prevalence of pathogens in ticks?**  
The habitats are broken into a few different locations, some more rural than others. The effect of habitat on pathogen prevalence is likely due to the different community make-up in each location. Doing this will give us an idea of how the community can sway pathogen prevalence in the select environment. In order to answer this question we will use an ANOVA, looking at habitat type and specific pathogen.  
**2. How the microbial composition changes with the prevalence of pathogens?**  
This data will give insight into how the microbial community and pathogen prevalence shift with or against each other over a period of time. It will show future intereactions that can be on interest within the microbiome of a tick. We will be using Chi squared tests to look at these changes and the interactions.  
**3. How pathogen prevalence shifts over time within different habitats?**  
This will show how changes in habitat, likely due to temperature and humidity changes, effects the pathogen prevalence in a single area. We will be comparing habitats to each other through ANOVA tests.

# Methods and Results

*In most research papers, results and methods are separate. You can combine them here if you find it easier. You are also welcome to structure things such that those are separate sections.*

## Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*  
Through the Southeastern Coopertative Wildlife Disease Study (SCWDS) ticks were collected from the eastern region of the United States over the course of a year, the ticks used in this study were all from Georgia. In order to collect the ticks typical drag sampling was done in each transect of the area, with checks being done of the drag every 10m. When collected the habitat type and location was recorded for future reference. These ticks were then IDed through morphology, and used for 16S and PCR in order to determine the presence of pathogens and the microbial community within each specimen. The pathogen and microbial community will not be investigated genetically, instead we will be determining changes in prevalence over time and space. The data used for this analysis includes the pathogens *Rickettsia*, *Ehrlichia ewingii*, *Ehrlichia chaffeensis*, *Anaplasma phagocytophilum*, *Borrelia spp*, and Panola Mountain Ehrlichia (PME). The habitat information includes: season, region, site, and general habitat. The raw data can be found in the project file ./data/raw\_data/RawDataTickPathLoc.csv. This data is used for the analysis.

## Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what each file does. The files themselves should be commented well so everyone can follow along.*

#load needed packages. make sure they are installed.  
library(readxl)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(DataCombine)  
library(ggplot2)  
library(stats)  
library(reshape2)  
  
#load data. path is relative to project directory.  
#plateres <- read.csv("./data/raw\_data/CompiledPlateResults.csv")  
tickpath <- read.csv("../../data/raw\_data/RawDataTickPathLoc.csv",na.strings = c("", "NA"))  
 #Add NA to observations where pathogen species was not tested (all observations without neg/pos were not tested)  
 #Where sex is blank it is a nymph or larval cluster which do not currently phenotypically sex -> add NA to those columns  
  
#take a look at the data  
dplyr::glimpse(tickpath)

## Observations: 1,968  
## Variables: 15  
## $ ï..Season <fct> Summer, Summer, Summer, Summer, Summ...  
## $ Region <fct> Lower Coastal, Lower Coastal, Lower ...  
## $ Site <fct> Wando, Wando, Wando, Wando, Wando, W...  
## $ Habitat <fct> BLH, BLH, BLH, BLH, BLH, BLH, BLH, B...  
## $ Transect.. <int> 13, 13, 7, 5, 5, 17, 17, 6, 6, 9, 16...  
## $ Tick...On.transect <int> 1, 2, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, ...  
## $ Species <fct> Amblyomma americanum, Ixodes affinis...  
## $ Sex <fct> Female, Female, Male, Male, NA, NA, ...  
## $ Life.stage <fct> Adult, Adult, Adult, Adult, Nymph, N...  
## $ Rickettsia <fct> PosRick, NA, PosRick, NA, NegRick, P...  
## $ Ehrlichia.ewingii <fct> NegEew, NA, NegEew, NA, NegEew, NegE...  
## $ Ehrlichia.chaffeensis <fct> NegEch, NA, NegEch, NA, NegEch, NegE...  
## $ Anaplasma.phagocytophilum <fct> NA, NegAna, NA, NegAna, NA, NA, NA, ...  
## $ Borrelia.spp. <fct> NA, NegBor, NA, NegBor, NA, NA, NA, ...  
## $ PME <fct> NegPME, NA, PosPME, NA, NA, NA, NA, ...

#Note to self PME = Panola Mountain Ehrlichia  
  
####Processing data####  
  
#Renaming columns that don't convert from the .csv file  
names(tickpath)[1]<-"Season"  
names(tickpath)[5]<-"TransectNum"  
names(tickpath)[6]<-"TickID"  
  
#Renaming the "Borrelia.spp." column to not have the . at the end  
names(tickpath)[14]<-"Borrelia.spp"  
   
#Checking that all factors are coded correctly  
levels(tickpath$Site)

## [1] "Battery Warren" "Buck Hall" "DiLane" "FM McConnell"   
## [5] "Halfway Creek" "McConnell" "Ocmulgee" "Rock Eagle"   
## [9] "Scull Shoals" "Sewee" "Sewee " "South Tibwin"   
## [13] "SRS N" "SRS North" "SRS S" "SRS S "   
## [17] "Steed Creek" "Tom Yawkey" "Wando" "Wando "

##Some sites have spaces at the end that shouldn't be there,   
 #or have two names for the same place  
tickpath$Site <- replace(tickpath$Site, tickpath$Site == "Sewee ", "Sewee")  
tickpath$Site <- replace(tickpath$Site, tickpath$Site == "SRS North", "SRS N")  
tickpath$Site <- replace(tickpath$Site, tickpath$Site == "SRS S ", "SRS S")  
tickpath$Site <- replace(tickpath$Site, tickpath$Site == "Wando ", "Wando")  
tickpath$Site <- replace(tickpath$Site, tickpath$Site == "FM McConnell", "McConnell")  
tickpath$Site <- factor(tickpath$Site)   
levels(tickpath$Site)

## [1] "Battery Warren" "Buck Hall" "DiLane" "Halfway Creek"   
## [5] "McConnell" "Ocmulgee" "Rock Eagle" "Scull Shoals"   
## [9] "Sewee" "South Tibwin" "SRS N" "SRS S"   
## [13] "Steed Creek" "Tom Yawkey" "Wando"

levels(tickpath$Habitat)

## [1] "BLH" "Early successional" "Longleaf Pine"   
## [4] "Maritime forest" "Maritime Forest" "Pine"   
## [7] "Upland" "Upland "

##Subset of "Upland" has an extra space at the end  
tickpath$Habitat <- replace(tickpath$Habitat, tickpath$Habitat == "Upland ", "Upland")  
tickpath$Habitat <- factor(tickpath$Habitat)   
levels(tickpath$Habitat)

## [1] "BLH" "Early successional" "Longleaf Pine"   
## [4] "Maritime forest" "Maritime Forest" "Pine"   
## [7] "Upland"

levels(tickpath$Species)

## [1] "Amblyomma americanum" "Amblyomma maculatum"   
## [3] "Dermacentor variabilis" "Ixodes affinis"   
## [5] "Ixodes scapularis" "Ixodes scapularis "

##Species has duplicated, one has an extra space at the end  
 #"Ixodes scapularis" "Ixodes scapularis " Want to remove the space from the second  
tickpath$Species <- replace(tickpath$Species, tickpath$Species == "Ixodes scapularis ", "Ixodes scapularis")  
tickpath$Species <- factor(tickpath$Species)   
levels(tickpath$Species)

## [1] "Amblyomma americanum" "Amblyomma maculatum"   
## [3] "Dermacentor variabilis" "Ixodes affinis"   
## [5] "Ixodes scapularis"

#Each pathogen species that is tested is listed at "Pos\_\_\_" or "Neg\_\_\_" with 3 letters corresponding to the pathogen  
 #need to transform to 0 for neg, 1 for positive  
positiveNames <- c("PosRick","PosEew", "PosEch", "PosAna", "PosBor", "PosPME")  
negNames <- c("NegRick", "negRick", "NegEew", "NegEch", "NegAna", "NegBor", "NegPME", "NegPMe")  
columnNames <- c("Rickettsia", "Ehrlichia.ewingii", "Ehrlichia.chaffeensis", "Anaplasma.phagocytophilum", "Borrelia.spp", "PME")  
  
#creating a dataframe that will have what needs to be changed and to what  
preplacedf <- data.frame("Current" = positiveNames, "New" = 1)  
nreplacedf <- data.frame("Current" = negNames, "New" = 0)  
replacedf <- rbind(preplacedf, nreplacedf)  
  
#FindReplace is from the DataCombine package  
for (i in columnNames){  
 tickpath <- FindReplace(tickpath, i, replacedf, from = "Current", to = "New", exact = TRUE)  
}

## Only exact matches will be replaced.

## Only exact matches will be replaced.  
## Only exact matches will be replaced.  
## Only exact matches will be replaced.  
## Only exact matches will be replaced.  
## Only exact matches will be replaced.

#Change from character strings to numeric  
sapply(tickpath, class)

## Season Region   
## "factor" "factor"   
## Site Habitat   
## "factor" "factor"   
## TransectNum TickID   
## "integer" "integer"   
## Species Sex   
## "factor" "factor"   
## Life.stage Rickettsia   
## "factor" "character"   
## Ehrlichia.ewingii Ehrlichia.chaffeensis   
## "character" "character"   
## Anaplasma.phagocytophilum Borrelia.spp   
## "character" "character"   
## PME   
## "character"

tickpath[columnNames] <- sapply(tickpath[columnNames], as.numeric)  
sapply(tickpath, class)

## Season Region   
## "factor" "factor"   
## Site Habitat   
## "factor" "factor"   
## TransectNum TickID   
## "integer" "integer"   
## Species Sex   
## "factor" "factor"   
## Life.stage Rickettsia   
## "factor" "numeric"   
## Ehrlichia.ewingii Ehrlichia.chaffeensis   
## "numeric" "numeric"   
## Anaplasma.phagocytophilum Borrelia.spp   
## "numeric" "numeric"   
## PME   
## "numeric"

#Need total number of ticks found in each transect  
Totaltick<-tickpath %>%   
 group\_by(Season, Region, Site, Habitat, TransectNum, Species) %>%  
 count(Species)  
names(Totaltick)[7]<-"NumberTicksFound"  
  
tickpath<-merge(x=tickpath,y=Totaltick, by=c("Season","Region","Site","Habitat","TransectNum","Species"))  
  
##Need to reorder the seasons to put them in chronological order  
levels(tickpath$Season)

## [1] "Fall" "Spring" "Summer" "Winter"

tickpath$Season <- factor(tickpath$Season, levels = c("Spring", "Summer", "Fall", "Winter"))  
levels(tickpath$Season)

## [1] "Spring" "Summer" "Fall" "Winter"

saveRDS(tickpath, file = "../../data/processed\_data/processeddata\_tickpath.rds")  
saveRDS(Totaltick, file = "../../data/processed\_data/processeddata\_Totaltick.rds")

#load needed packages. make sure they are installed.  
library(ggplot2)  
library(dplyr)  
library(broom)  
library(reshape2)  
  
#load data. path is relative to project directory.  
tickpath <- readRDS("../../data/processed\_data/processeddata\_tickpath.rds")  
  
#How the habitat effects the prevalence of pathogens in ticks?  
 #Total number of positive tests found in a habitat broken down by species.  
 #Creating the number of total pathogens  
tickpath$TotalPath <- rowSums(tickpath[,c("Rickettsia", "Ehrlichia.ewingii",  
 "Ehrlichia.chaffeensis", "Anaplasma.phagocytophilum",   
 "Borrelia.spp", "PME")], na.rm=TRUE)  
#Associating all pathogen tests on a single tick with that tick  
meltpath <- melt(tickpath, id.vars=c("Season","Region","Site","Habitat","TransectNum","TickID",  
 "Species","Sex","Life.stage","TotalPath","NumberTicksFound"))  
  
speciesLand<-meltpath %>%  
 ggplot(aes(x=Habitat,y=value, fill=variable)) +  
 geom\_bar(stat = "identity") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 facet\_wrap(.~Species)  
  
ggsave(filename="../../results/speciesLandfigure.png",plot=speciesLand)

## Saving 5 x 4 in image

## Warning: Removed 9059 rows containing missing values (position\_stack).

# fit linear model  
lmfitSpecies <- lm(Habitat ~ Species, meltpath)

## Warning in model.response(mf, "numeric"): using type = "numeric" with a  
## factor response will be ignored

## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors

lmfitSpecies

##   
## Call:  
## lm(formula = Habitat ~ Species, data = meltpath)  
##   
## Coefficients:  
## (Intercept) SpeciesAmblyomma maculatum   
## 4.1159 -0.1159   
## SpeciesDermacentor variabilis SpeciesIxodes affinis   
## -1.8937 -1.3034   
## SpeciesIxodes scapularis   
## -0.8744

# place results from fit into a data frame with the tidy function  
lmtableSpecies <- broom::tidy(lmfitSpecies)

## Warning in Ops.factor(r, 2): '^' not meaningful for factors

lmtableSpecies

## # A tibble: 5 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 4.12 NA NA NA  
## 2 SpeciesAmblyomma maculatum -0.116 NA NA NA  
## 3 SpeciesDermacentor variabilis -1.89 NA NA NA  
## 4 SpeciesIxodes affinis -1.30 NA NA NA  
## 5 SpeciesIxodes scapularis -0.874 NA NA NA

# save table   
saveRDS(lmtableSpecies, file = "../../results/resulttableSpecies.rds")  
  
#How the tick lifestage prevelance is changed within habitat?  
lifeLand<-tickpath %>%  
 ggplot(aes(x=NumberTicksFound, fill=Life.stage)) +  
 geom\_bar() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +  
 facet\_grid(Habitat~.)  
#Nymphal Amblyomma americanum is by far the most prevelant  
#Ixodes scapularis adults peak in the Fall, and A.A. peaks in summer  
  
ggsave(filename="../../results/lifeLandfigure.png",plot=lifeLand)

## Saving 5 x 4 in image

# fit linear model  
lmfitLife <- lm(Life.stage ~ Habitat, tickpath)

## Warning in model.response(mf, "numeric"): using type = "numeric" with a  
## factor response will be ignored  
  
## Warning in model.response(mf, "numeric"): '-' not meaningful for factors

lmfitLife

##   
## Call:  
## lm(formula = Life.stage ~ Habitat, data = tickpath)  
##   
## Coefficients:  
## (Intercept) HabitatEarly successional   
## 2.27560 -1.27560   
## HabitatLongleaf Pine HabitatMaritime forest   
## 0.39107 -1.05338   
## HabitatMaritime Forest HabitatPine   
## -0.08180 -0.05234   
## HabitatUpland   
## 0.31880

# place results from fit into a data frame with the tidy function  
lmtableLife <- broom::tidy(lmfitLife)

## Warning in Ops.factor(r, 2): '^' not meaningful for factors

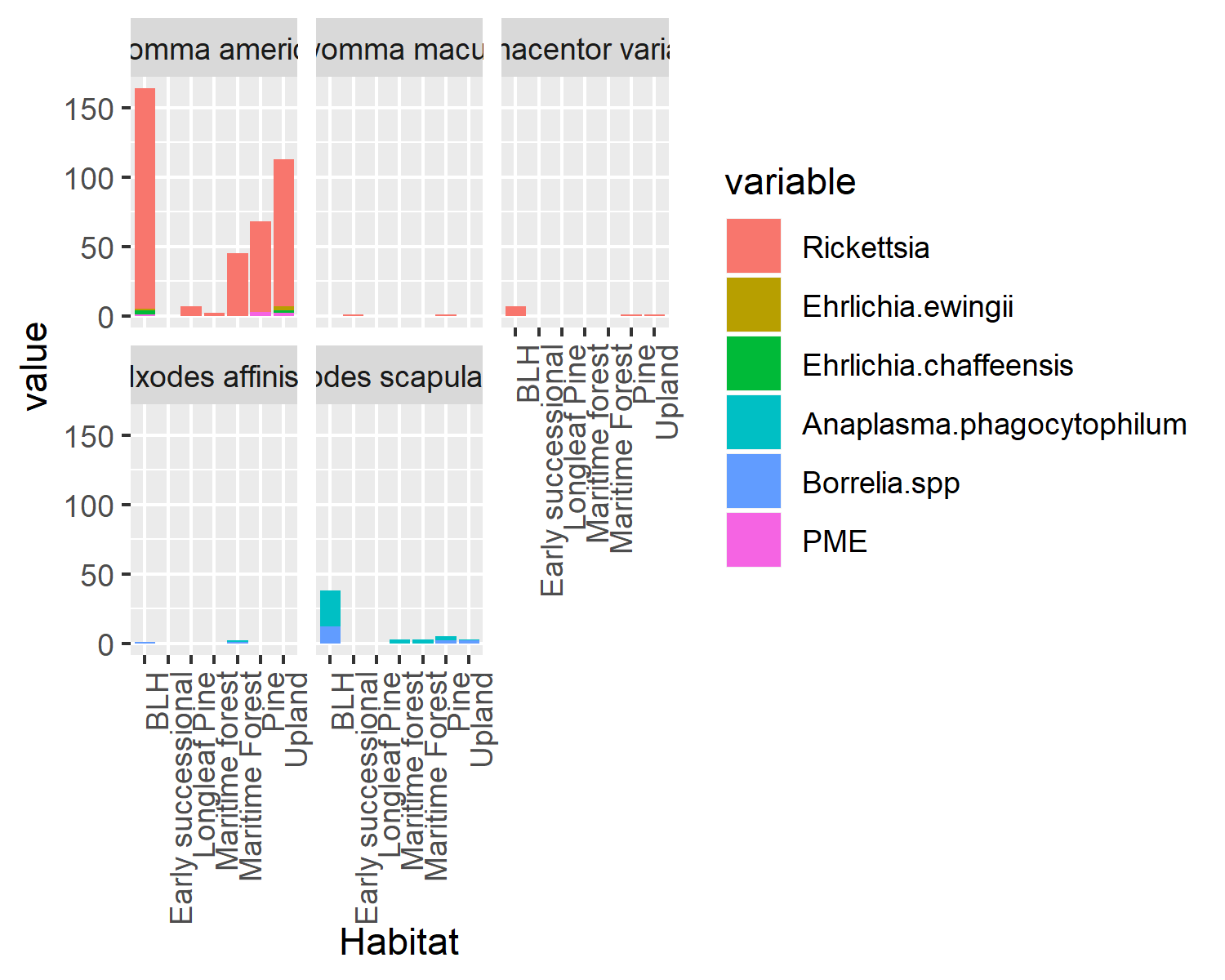
lmtableLife

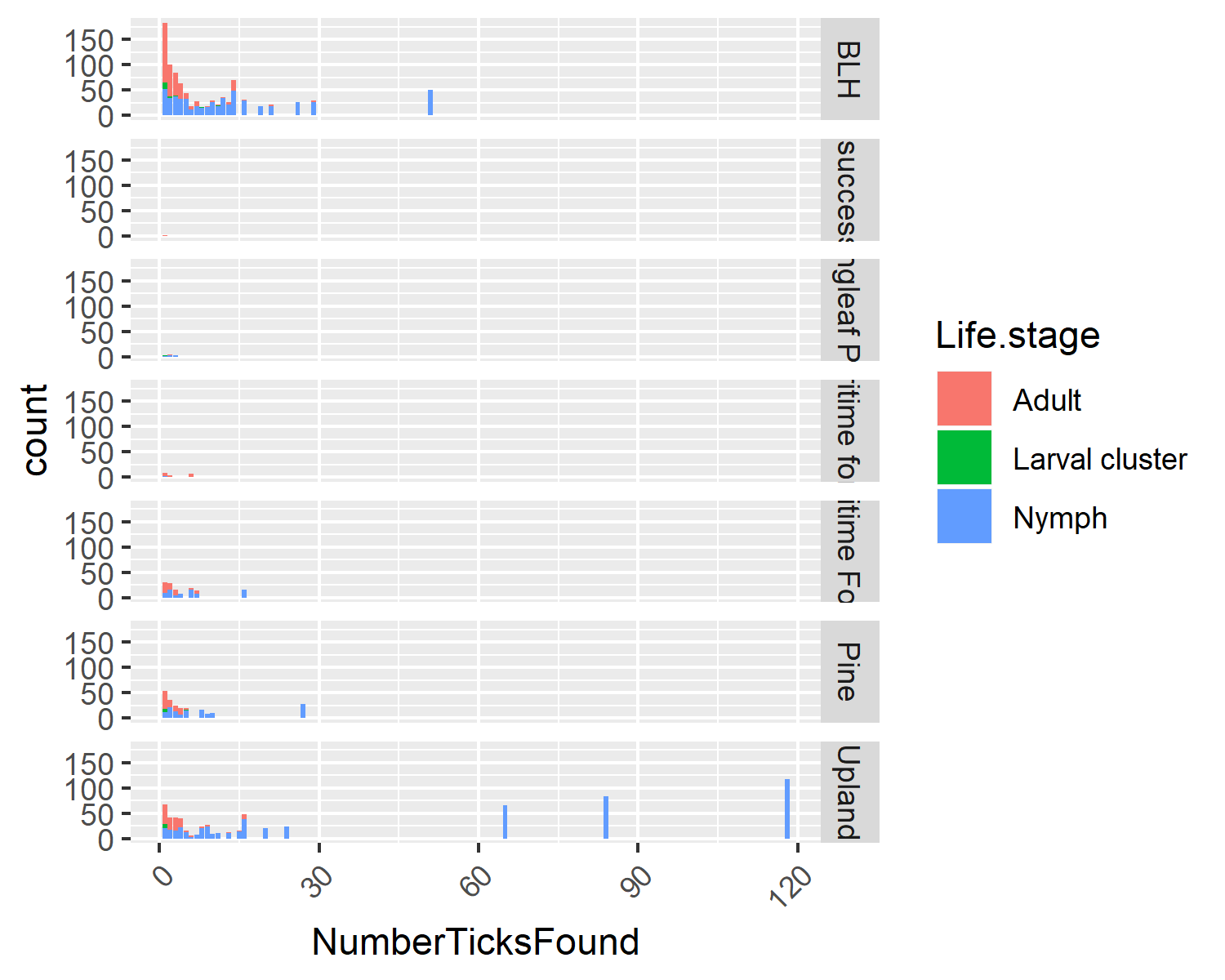
## # A tibble: 7 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 2.28 NA NA NA  
## 2 HabitatEarly successional -1.28 NA NA NA  
## 3 HabitatLongleaf Pine 0.391 NA NA NA  
## 4 HabitatMaritime forest -1.05 NA NA NA  
## 5 HabitatMaritime Forest -0.0818 NA NA NA  
## 6 HabitatPine -0.0523 NA NA NA  
## 7 HabitatUpland 0.319 NA NA NA

# save table   
saveRDS(lmtableLife, file = "../../results/resulttableLife.rds")

## Univariate analysis

*Use a combination of text/tables/figures to explore and describe your data. You should produce plots or tables or other summary quantities for most of your variables. You definitely need to do it for the important variables, i.e. if you have main exposure or outcome variables, those need to be explored. Depending on the total number of variables in your dataset, explore all or some of the others.*

When comparing the prevalence of pathogens in all tick species in different habitats, trends begin to emerge. Here we look at how the total number of positive individuals are broken down by species (fig 1). From here we can see that *Rickettsia* is the most common pathogen in *A. americanum*. However, for *I. scapularis* we see that Borrelia species and *Anaplasma* are the two most common. These results trend with what is in current literature.  
  
Figure 1. Pathogen prevalence per species. This figure shows the number of individuals per species that tested positive for a particular pathogen.

We can also use this dataset to see if lifestage has any relationship to habitat (fig 2). Overall, we can see that BHL and Upland have the most ticks of any lifestage found throughout it. However, there are pockets in Upland that have large numbers of nymphs found in a single transect which is an interesting phenomenom.  
  
Figure 2. Tick lifestage versus the habitat. This figure shows the number of individuals in each lifestage found in the different habitats.

## Bivariate analysis

*Create plots or tables and compute simple statistics (e.g. t-tests, simple regression model with 1 predictor, etc.) to look for associations between your outcome(s) and each individual predictor variable*

## Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You can then load the results produced by this code*

# Discussion

## Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

# References

Adelson, M. E., Rao, R. V., Tilton, R. C., Cabets, K., Eskow, E., Fein, L., … Mordechai, E. (2004). Prevalence of borrelia burgdorferi, bartonella spp., babesia microti, and anaplasma phagocytophila in ixodes scapularis ticks collected in northern new jersey [Journal Article]. *J Clin Microbiol*, *42*(6), 2799–2801. <https://doi.org/10.1128/JCM.42.6.2799-2801.2004>

Arsnoe, I. M., Hickling, G. J., Ginsberg, H. S., McElreath, R., & Tsao, J. I. (2015). Different populations of blacklegged tick nymphs exhibit differences in questing behavior that have implications for human lyme disease risk [Journal Article]. *PLoS One*, *10*(5), e0127450. <https://doi.org/10.1371/journal.pone.0127450>

Dantas-Torres, F., Chomel, B. B., & Otranto, D. (2012). Ticks and tick-borne diseases: A one health perspective [Journal Article]. *Trends Parasitol*, *28*(10), 437–446. <https://doi.org/10.1016/j.pt.2012.07.003>

Eisen, R. J., & Eisen, L. (2018). The blacklegged tick, ixodes scapularis: An increasing public health concern [Journal Article]. *Trends Parasitol*, *34*(4), 295–309. <https://doi.org/10.1016/j.pt.2017.12.006>

Ginsberg, H. S., Albert, M., Acevedo, L., Dyer, M. C., Arsnoe, I. M., Tsao, J. I., … LeBrun, R. A. (2017). Environmental factors affecting survival of immature ixodes scapularis and implications for geographical distribution of lyme disease: The climate/behavior hypothesis [Journal Article]. *PLoS One*, *12*(1), e0168723. <https://doi.org/10.1371/journal.pone.0168723>

Jongejan, F., & Uilenberg, G. (2004). The global importance of ticks [Journal Article]. *Parasitology*, *129 Suppl*, S3–14. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/15938502>

Vector-Borne Disease, D. of. (2019). [Web Page]. Retrieved from <https://www.cdc.gov/lyme/datasurveillance/index.html>