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

Julia Frederick

2019-11-08

# 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. The number one vector borne disease in the US is Lyme disease with 30,000 cases reported annually, and the Center for Disease Control and Prevention (CDC) estimating over 300,000 cases actually occurring (Vector-Borne Disease, 2019). Understanding how these pathogens are being moved throughout the environment gives insight into risk prevention efforts and effective targets. Especially when considering that a single species of tick is capable of transmitting multiple disease pathogens (Adelson et al., 2004). 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). For example, native species, *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). Multiple of these pathogens have been found in a single tick during field studies (co-infection) [REF]. Current research focusing on the dynamics of these co-infections has shown there to be correlations between specific pathogens occuring in higher frequencies [REF]. These co-infections can have an impact on human health and diagnostic tests used for detecting the specific pathogens [REF]. This isn’t only an issue within species, because many tick species are the main vector for specific pathogens and these species can feed on the same source which will mingle the pathogens between species. 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 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

## Data aquisition

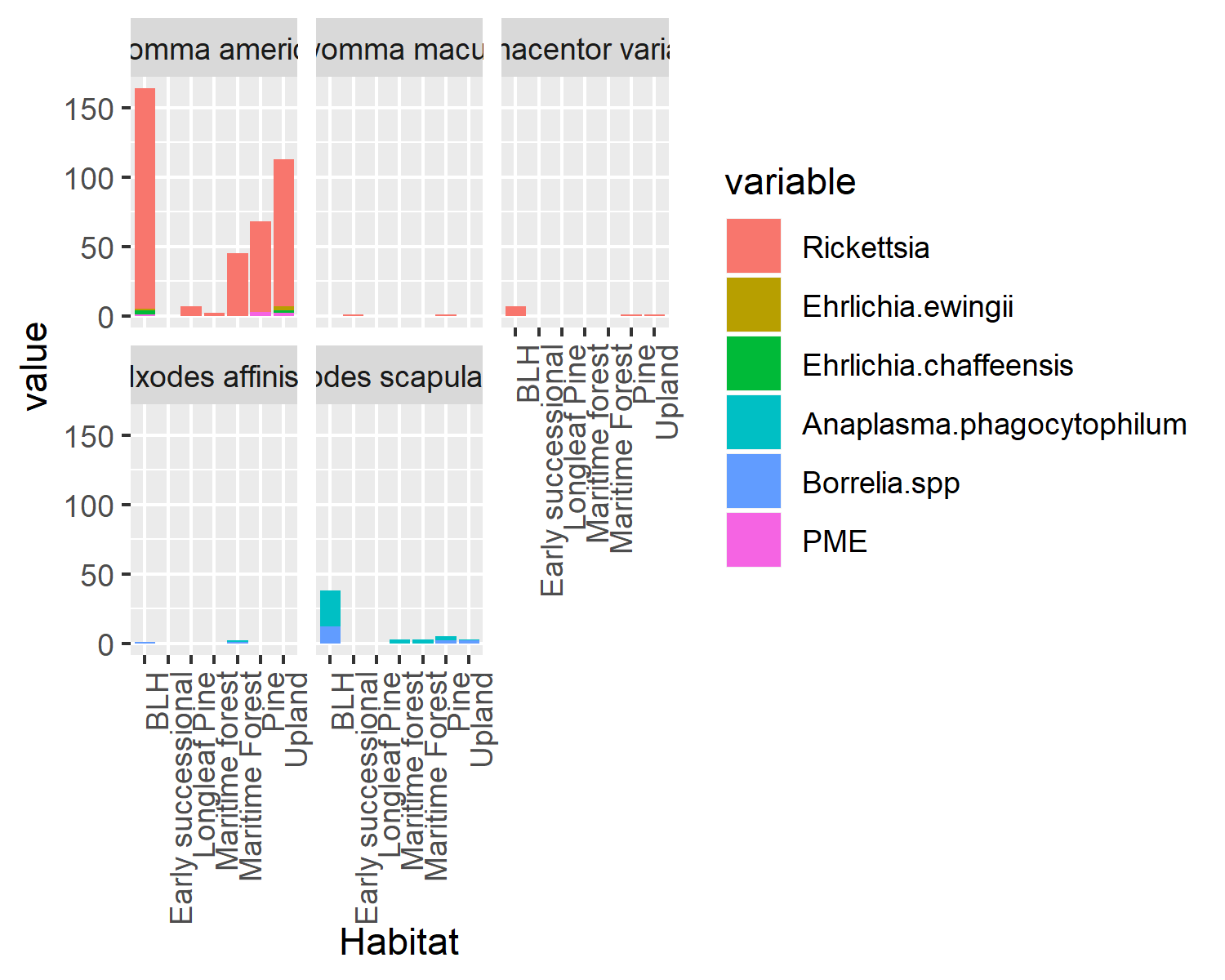
*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*  
Through the Southeastern Coopertative Wildlife Disease Study (SCWDS) ticks were collected from the eastern region of the United States over the 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.

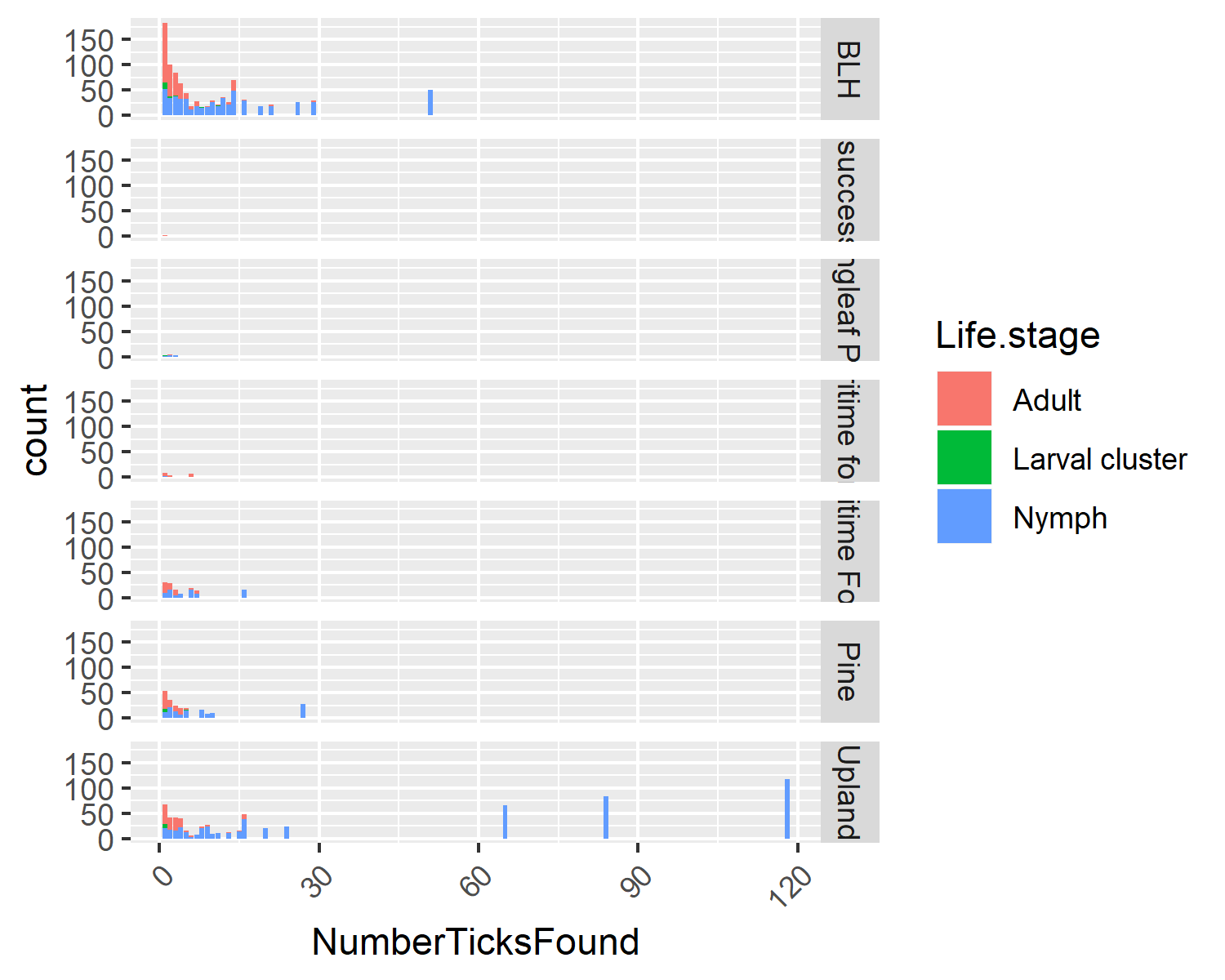
## Data import and cleaning

*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.*  
The data is originally imported from a .csv file which includes 16 variables and 1968 observations. These observations are from either a single nymph, a single adult, or a pool of larvae. The larvae need to be pooled as individually they do not contain enough DNA for detection. However, larvae could only be infected with pathogens that are transovarially passed to the offspring (like Rickettsia) and the two pathogens consistently tested for (*A. phagocytophilum* and *Borrelia spp.*) are not known to be transmitted in this way. In order to clean this data we first will import the data so all empty observations are coded as NA. Then we need to adjust a few levels within our variables because they have extra spaces or a spelled incorrectly. Once these are fixed we calculate the total number of pathogens found within single individuals. This is our main outcome of interest instead of looking at individual pathogens which weren’t consistently tested.

## 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*

readRDS("../../results/resulttablePathHab.rds")

## # A tibble: 5 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.160 0.0217 7.37 9.25e-13  
## 2 HabitatMaritime forest 0.0277 0.0875 0.316 7.52e- 1  
## 3 HabitatMaritime Forest 0.00145 0.0646 0.0225 9.82e- 1  
## 4 HabitatPine -0.0689 0.0506 -1.36 1.74e- 1  
## 5 HabitatUpland -0.124 0.0429 -2.89 3.99e- 3

readRDS("../../results/resulttablePathSea.rds")

## # A tibble: 4 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.0476 0.0744 0.640 0.523   
## 2 SeasonSummer 0.286 0.158 1.81 0.0711  
## 3 SeasonFall 0.0774 0.0792 0.977 0.329   
## 4 SeasonWinter 0.0841 0.0776 1.08 0.279

readRDS("../../results/resulttablePathSpe.rds")

## # A tibble: 5 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.261 0.0110 23.7 2.46e-109  
## 2 SpeciesAmblyomma maculatum 0.739 0.305 2.42 1.55e- 2  
## 3 SpeciesDermacentor variabilis 0.739 0.144 5.13 3.24e- 7  
## 4 SpeciesIxodes affinis -0.0738 0.108 -0.681 4.96e- 1  
## 5 SpeciesIxodes scapularis -0.136 0.0239 -5.68 1.53e- 8

## 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*

When seeing which variables influence total pathogen’s in an individual we get a decision tree that covers a few of the main key indicators in an environment (figure 3). This tree is showing that tick density in a region is the largest predictor and then is further defined by species, season, and habitat. All of which make biological sense.

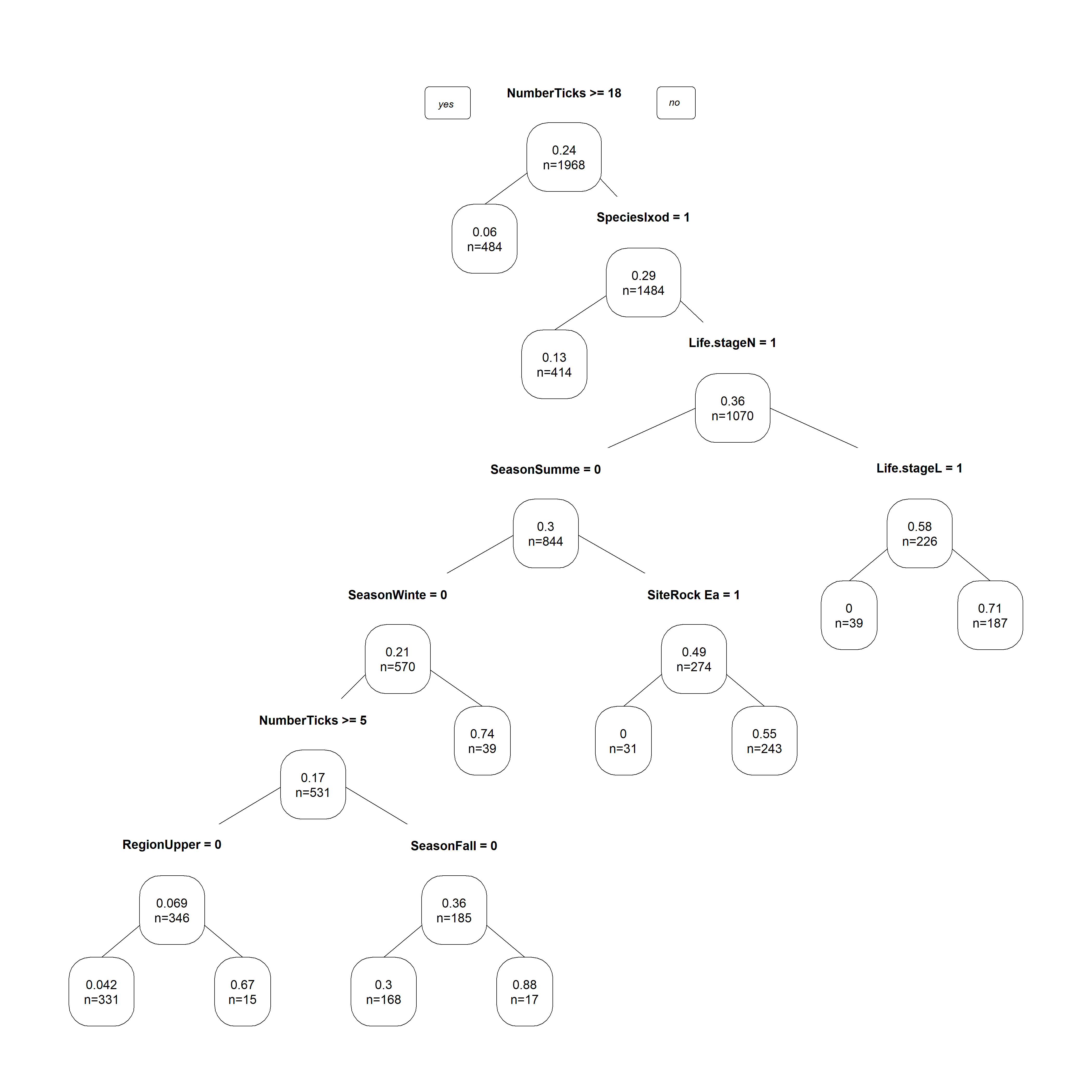


Figure 3. Decision Tree for Predictors of Total Pathogen

# 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>