Final Portfolio - INFO 526, Spring 2024, Sushma Anand Akoju

Reflection

In this final portfolio for the 7-week Spring semester INFO 526 course, I created plots and 1 tables to show analysis on Mineral Licks dataset over birds and mammals.

In addition to this I included an analysis of attempted **Generalized Linear Mixed Models** that I learnt from STATS 5010, Spring 2022, at University of colorado Boulder and how to analyse a time series from INFO VIZ 5602 also at University of Colorado Boulder.

Note: The Generalized Linear mixed models was selected based on the research work "Temporal patterns of visitation of birds and mammals at mineral licks in the Peruvian Amazon"

Definition of Moon Fraction Illuminated

Moon (Lunar) Phases: Crescent, Gibbous, New Moon, Full Moon and Quarter moon phases.

This data, specifically has poisson distribution or quasi-poisson distribution depending on the type of species. The plots 5,6,7,8,9 show that probability of visitations for variations in moon brightness is non-linear between two seasons. Specifically, each Type of Mammal/Bird has its own non-linear pattern. Since Moon brightness and seasons are events over time and are subject to time intervals such as how likely is it that a dove can visit during some moon brightness in a season. There may be some prediction interval which can be bootstrapped. However, this required additional calculations to convert date/moonphase information together to be combined into some form of duration for calculating the appropriate quantified duration.

My portfolio plots are designed to specifically show improvements in the following criteria:

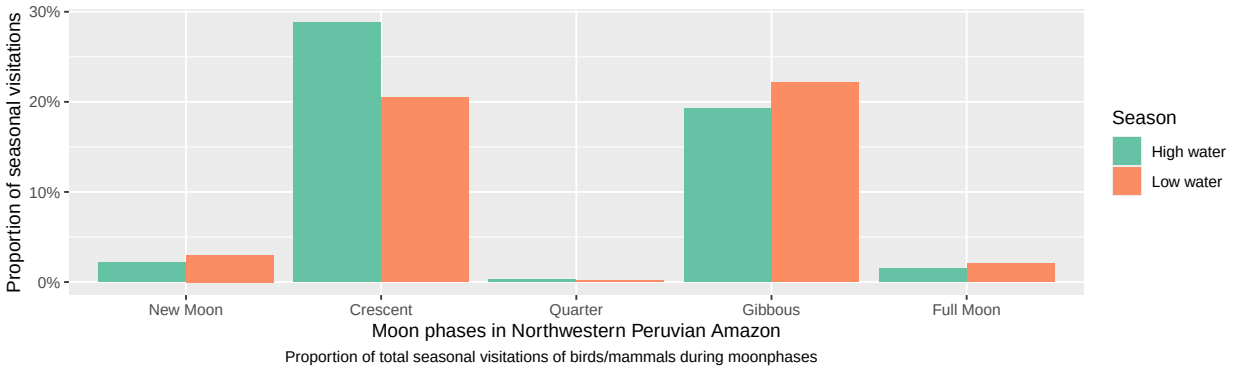
- Whole Portfolio - 1,2,3,4,5,6 and 3 plots (Generalized Linear and Linear Mixed Model analysis) - 7,8,9
- 1, plots - 1,2,3,4,5,6
- 2, a table - 1
- 3, a basic bar plot - 1
- 4, a multi-panel bar plot - 2
- 5, a multi-panel scatterplot with loess approx. - 3
- 6, a faceted bar plot - 4
- 7, a 2-panel, faceted, scatterplot with segments (linear regr.) - 5
- 8, a radial plot and a 3D scatter plot - 6
- 4, a multi-panel, scatterplot with segments, histogram,plot(GLM, GLMER) - 7,8,9
- 5, Linear regression, Generalized Linear Models, and Generalized Linear Mixed models analysis

Table 1: Number of visitations of birds and mammals for mineral licks by year

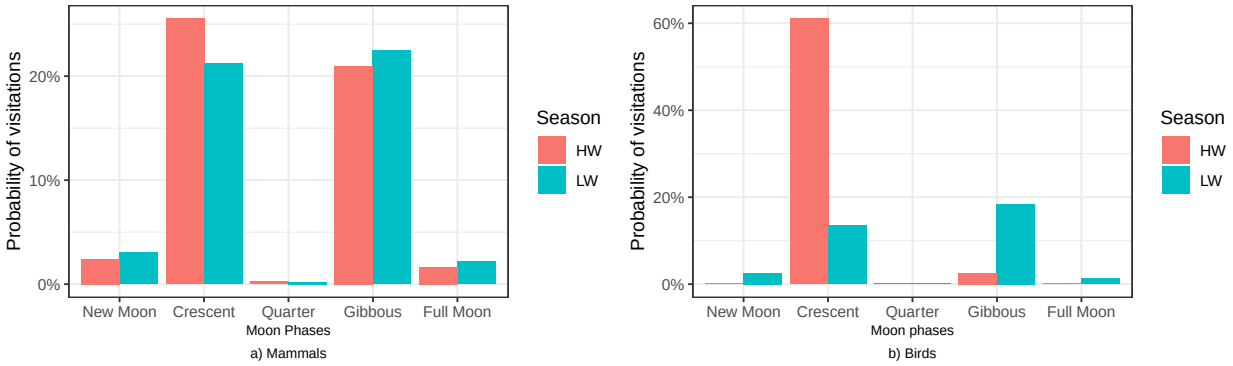
Year	Mammals	Birds
1970	0	333
2012	17	0
2017	156	0
2018	3067	205
2019	2517	42

Table 1 - The Mineral Lick data used for this assignment consists of number of visitations of birds and mammals, species, by date, by lick location, moon phase, lunar brightness, location coordinates, distance to - water, hunting access, hunting camp, community. The yearly number of visitations of birds and mammals is shown in Table 1.

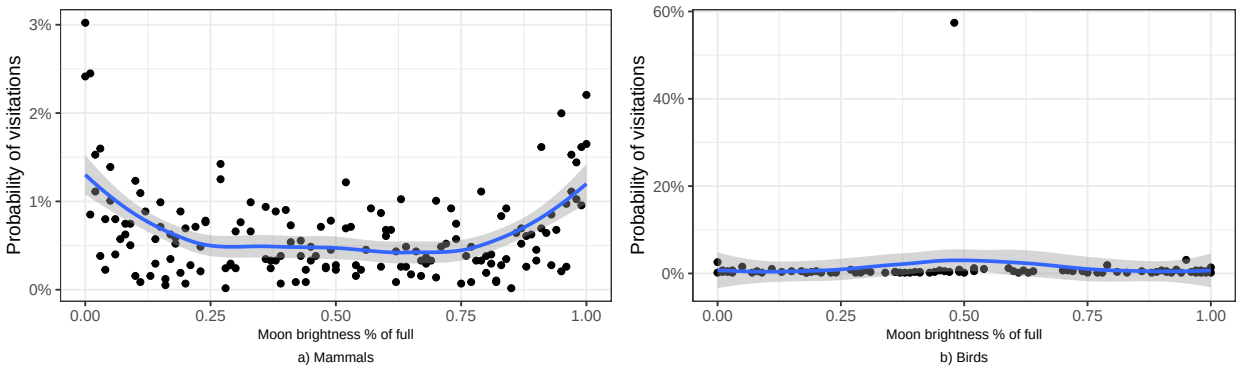
Table Notes: * Yearly visitations by the mammals and birds are listed for 2012, 2017, 2018, 2019.
* For the year 1970, the data is available only for birds, and not mammals. * For the year 2012 and 2017, the data is available only for mammals and not birds.



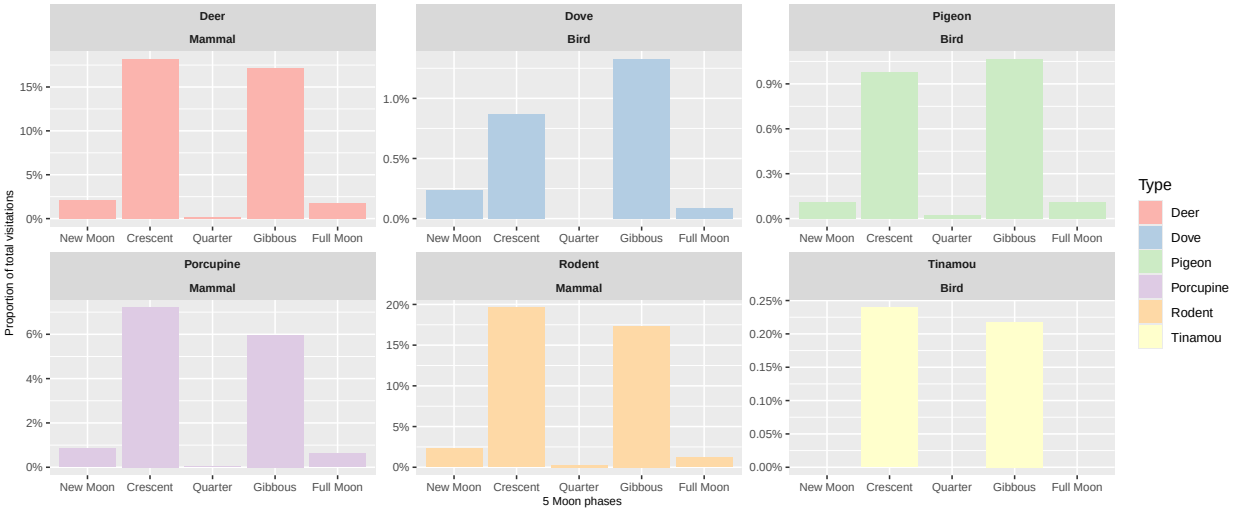
Plot 1: The bar plot shows proportion of total seasonal visitations of birds/mammals together, during 5 moon phases (Crescent, Gibbous, New Moon, Full Moon and Quarter moon phases). Total number of visitations across all seasons, moonphases together is: 6337. Of 6337 total visitations, 3124 were during Crescent, 2623 were during Gibbous, 329 during, 231 during Full moon phases. This data was collected from visitations of birds and mammals around moon phases at Northwestern Peruvian Amazon forest to study the mineral licks at 57 lick locations, that were collected in the years 1970, 2012, 2017, 2018 and 2019. There are only two seasons at the Northwestern Peruvian Amazon: low water (Jun-Nov) and high water seasons (Dec-May). Mineral Licks are ecological resources, that provide essential dietary nutrients and clays to birds and mammals. Plot shows more visitations during highwater season than low water. The Crescent and Gibbous moon phases have greater number of visitations, as an overall trend.



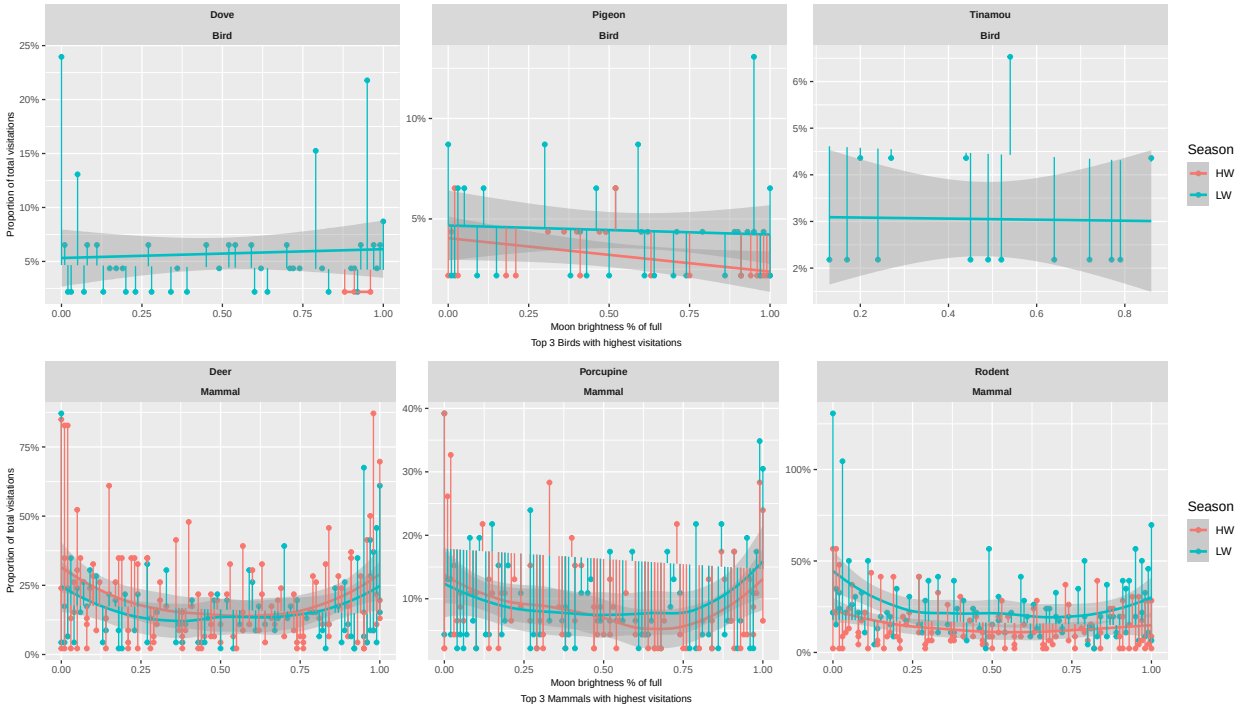
Plot 2: This plot shows the visitations among mammals and birds during the 5 moon phases. This plot shows visitations as proportion of all visitations during 5 moon Phases. Most visitations seems to occur around Crescent and Gibbous moon phases. Among the moon phases, the plot shows proportion of total visitations by month based on the two different classes/categories of animals i.e. birds and mammals with total of 25 types of mammals and birds.



Plot 3: This is a multi-panel scatterplot with loess, that shows the likelihood of visitations among mammals and birds based on moon brightness. This plot shows visitations as proportion of all visitations during varying moon brightness. Most visitations seems to occur when moon brightness was closer to 0.0 and 1.0 (synonymous to new moon and full moon phases). Among the various values of moon brightness, the plot shows proportion of total visitations by month based on the class of animals (birds and mammals - with a total of 25 types of birds/mammals and a total of 42 different species).



Plot 4: This faceted bar plot shows proportion of all visitations during 5 moon Phases and the proportion of visitations per month among - top 3 mammals: deer, porcupine and rodent and top 3 birds: dove, pigeon and tinamou, among each one of the moon phases. Most visitations seems to occur when moon phases were Crescent and Gibbous phases. The proportion of total visitations by month based on the class of animals with top 3 visitors among birds and top 3 visitors among the mammals).



Plot 5: This is 2-row, faceted, scatterplot each one of the plots with linear regression method, for analysis of highest visitations. There are two rows. So first row contains linear regression fit with faceted scatter plot for top 3 birds - doves, pigeons and tinamou as they had the highest number of visitations over the moon brightness variations. And the second row presents the top 3 mammals: deers, porcupines and rodents that had the highest number of visitations over moon brightness variations.

I added a linear regression analysis to show that the frequency of visitations i.e. the fitted values do not really fit the measured values. Moon fraction illumination is a temporal event that repeats over 30-day time period. During the moon brightness < 0.48 (Crescent moon phase) and moon brightness >0.50 (Gibbous moon phase) moon phases, there were highest visitations from birds from earlier plots. So given the Moon fraction illumination (which a measure of moon brightness (% of full), 0,0.25, 0.5, 0.75 and 1 respectively), the probability of visitations of each one of 6 animals during High water (HW) and Low water (LW) seasons were plotted.

Although guan was the most frequent visitor, but the data is only available for 1970. Besides some data is available for 2012 for some birds. Therefore I chose dove, pigeon and tinamou for an assessment of analysis that were the top most frequent visitors for which data was available.

Doves have higher visitations during Low water season.

Pigeons are diurnal birds. Pigeons seem to have higher visitations during Low water season, although their frequency of visitations is relatively lower in comparison to others. Source 1 credits and and source 2

Tinamou belong to terrestrial group of birds and are also diurnal. Tinamou only visit during Low water season.

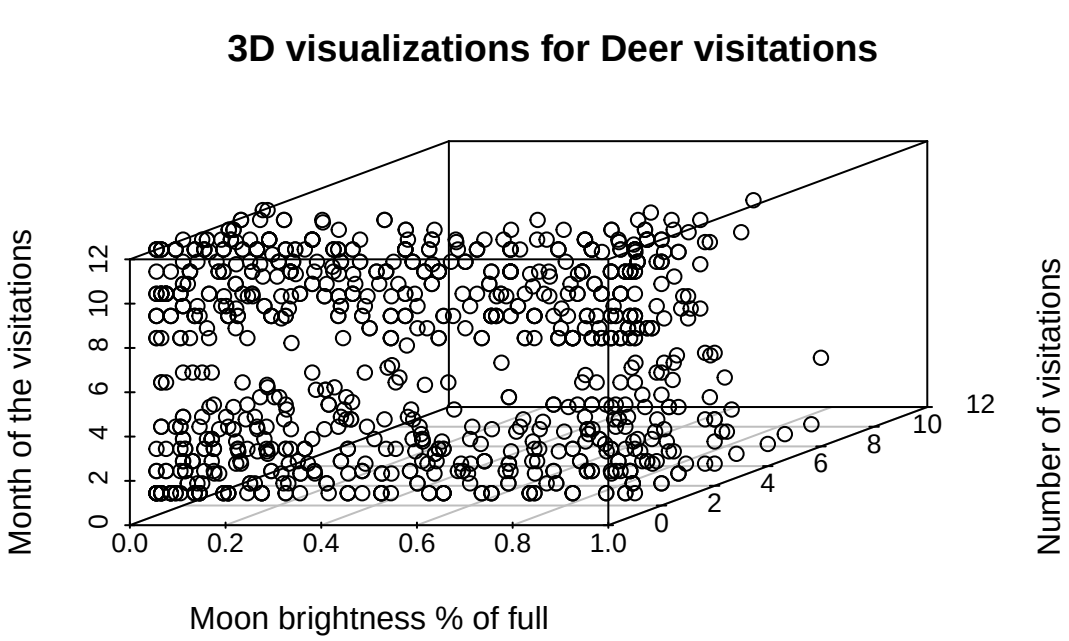
Deers seem to have several times better night vision than humans, even though they are not nocturnal. As per the analysis, deers seem to navigate relatively similar between new moon to full moon but they seem to have higher visitations during High water season than Low water. Some species of deer are crepuscular (active during twilight moon).

Porcupines (type of rodents with spines) are nocturnal but their eyesight is poor but possess a good hearing. Porcupine's quills protects them from being an easy prey, and they have sharper sense of smell than other mammals. So it seems to suggest porcupines seem to visit for mineral licks more often during Low water season as they are also terrestrial.

Similar to porcupines, rodents are also terrestrial. But rodents seem to visit more frequently during Low water season. They have better vision and can detect UV light.

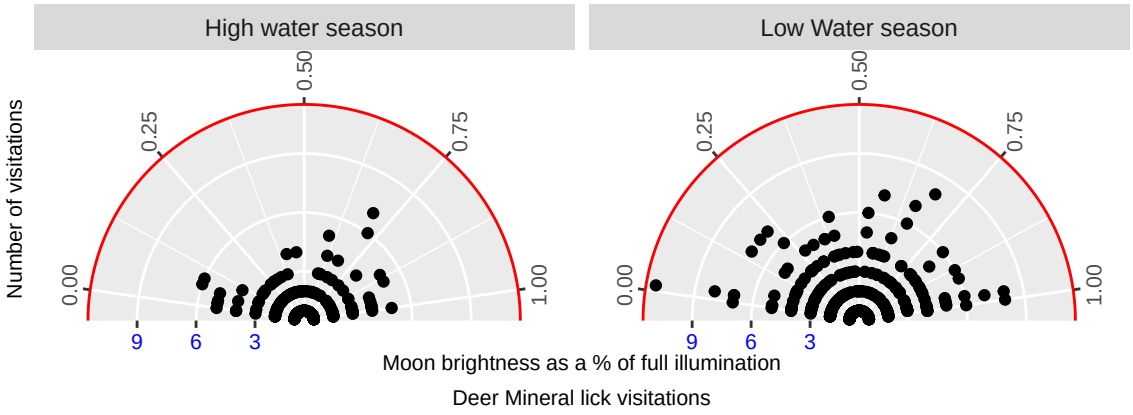
I conducted a Generalized Linear Model (GLM) and Generalized Linear Mixed Model (GLMM) analyses for the deers, as a single species of mammals. The following analysis is for understanding the temporal visitation patterns and its effects from moon brightness, based on the month, for deers as one of the most frequent visitor among mammals.

Note: Given the Moon fraction illumination (which is a measure of moon brightness (as a percentage of full illumination), 0,0.25, 0.5, 0.75 and 1 respectively), the probability of number of visitations of an animal over High water versus Low water seasons was analyzed. The trends in visitations over the two seasons, are influenced by the moon brightness. Moon brightness itself is a variable that changes across regular intervals of time. So the trends also showcase some sort of poisson or quasi-poisson randomized effects (to record the non-normal distributions which potentially suggest non-linearity, which is evident in these plots).



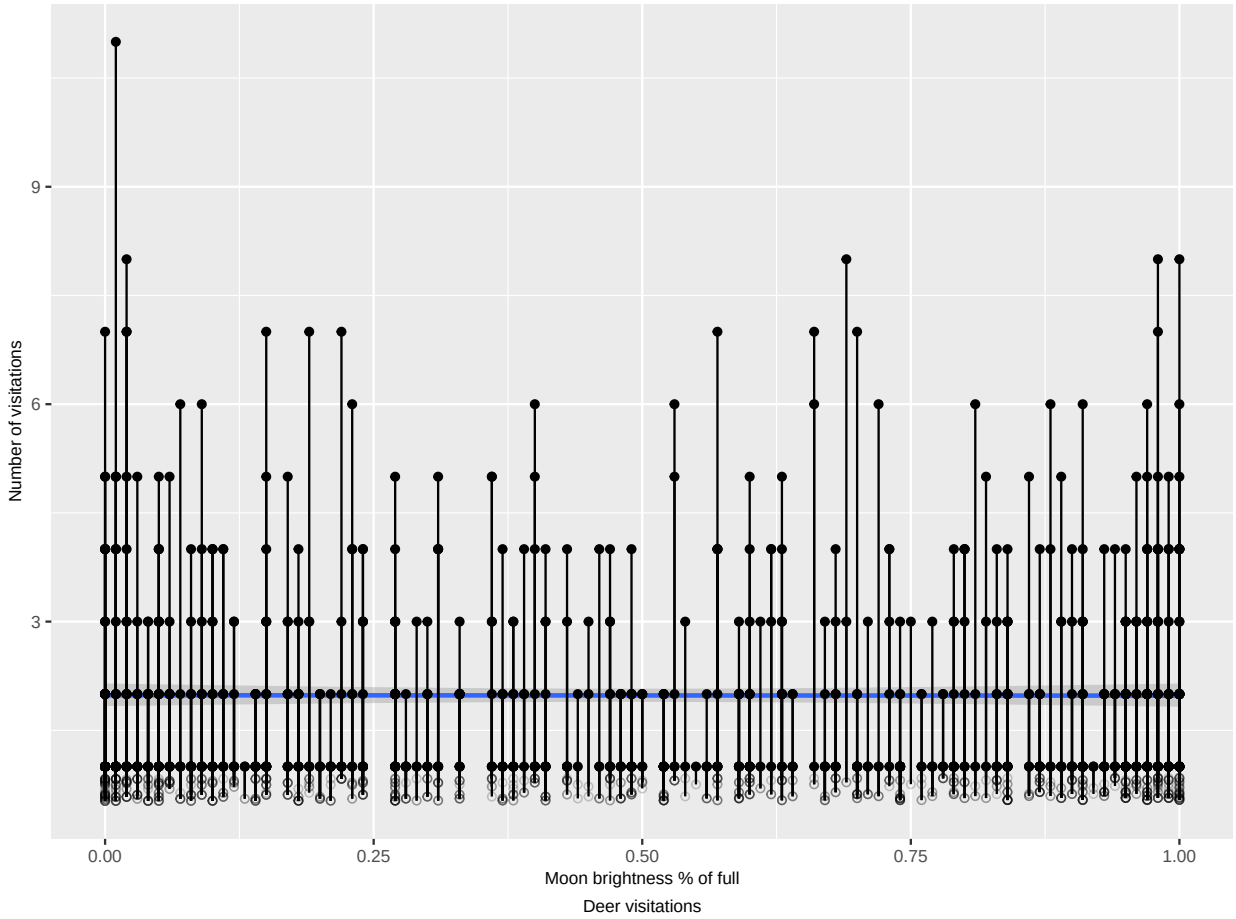
Plot

6a: This plot shows the 3D visualization of MoonFactorIllumination, Month and Number of visitations of the deer.



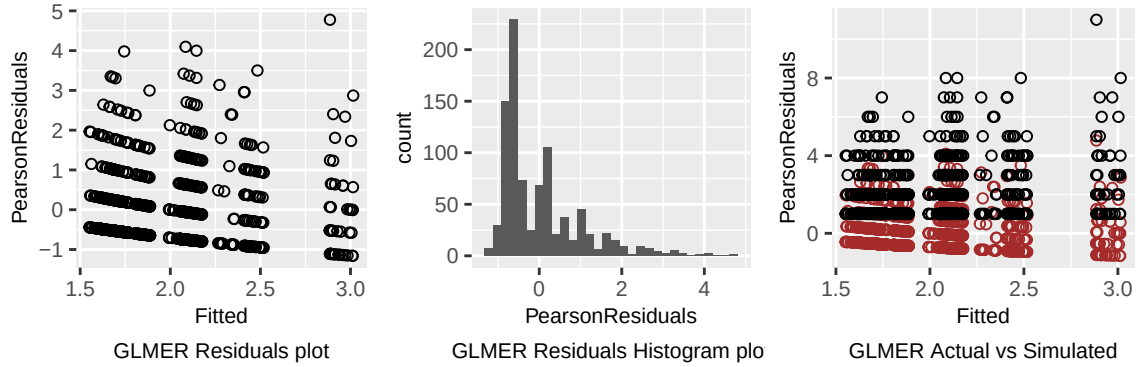
Plot

6b This is a radial plot that shows the number of visitations of the deers during High water and Low water seasons with periodic, repeating variations in moon brightness. The numbers along the red colored line are the moon phases are the moon phases (New Moon is 0.00, at First and Last Quarter is 0.50, and at Full Moon is 1.00). at an angle along the radial axis (theta). The density of numbers between each one of the two moon phases is number of visitations when the percentage of full moon illumination falls between the phases (an illumination of 0.73 falls between 0.50 and 0.75). The blue colored numbers are total number of visitations per moon phase. The points inside the radial plot represent the number of visitations per moon fraction illumination (the number of visitations for an illumination of 0.73 fall within 6 to 9 total visitations).



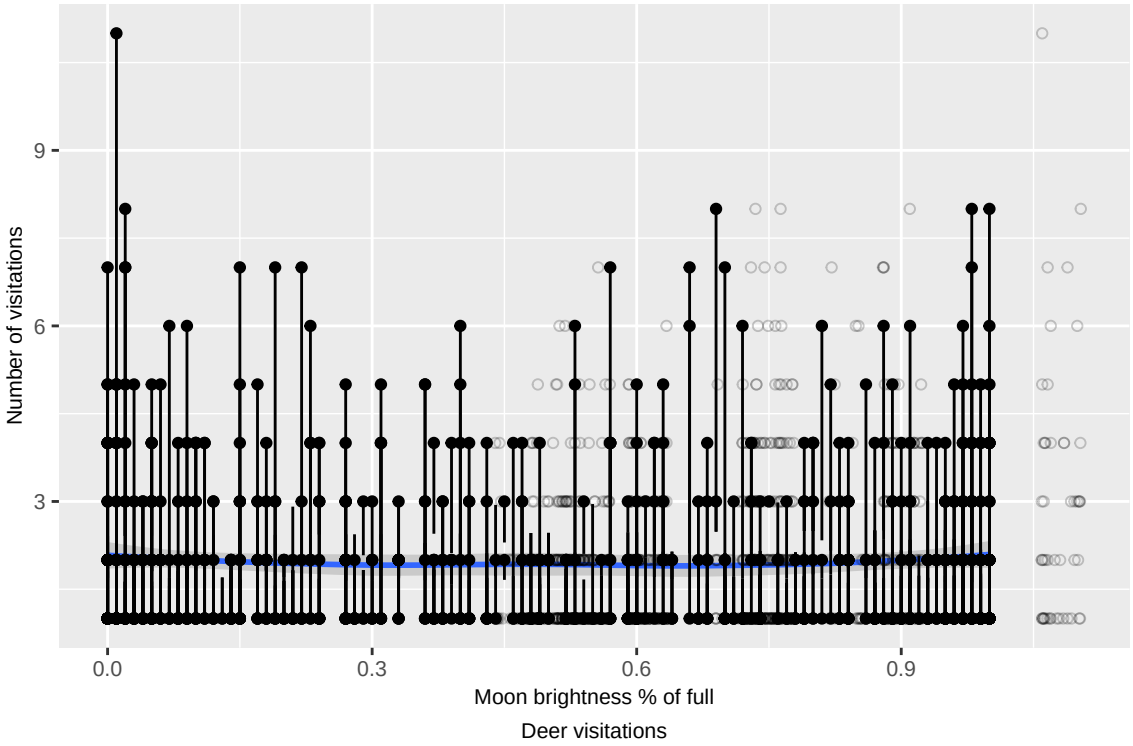
Plot 7: This is a scatterplot with segmented lines that shows the number of visitations of the deers during periodic variations in moon brightness compared with that of predictions of a Generalized Linear model. The response variable is the number of visitations and predictors are moon brightness and month. Clearly all predicted values vary. However, there seems to be some overfitting (due to smaller density of actual values that are < 0.5 probability of frequency of visitations) i.e. the predicted values are > 0.5.

GLM Model Summary: Since p-value for Moon illumination is very high, which is a predictor, there is no sufficient evidence that the effect of moon illumination exists within the population. Also we note that the sample size is small i.e. moon brightness is sparse. We fail to reject null hypothesis that there are no effects of moon illuminations on the response variable based on Generalized Linear Model analysis, given that the Moon Illumination is an influential variable over visitations.



GLMM model analysis Plot 8: This multi-panel plot is a visualization of results containing 3

subplots: a) Residuals plot i.e. between residuals and slope of coefficients of each predictor, b) histogram plot of the residuals plot and c) Actual vs simulated values. Note that histogram does not follow a bell shape as it is not a normal distribution. This model is not a true model, given that some of the data are missing for some of the years and was modified for the species (deers).



Plot 9: This plot shows that the predicted values are too close to actual values which is an indicator of overfitting, here we use the Generalized Linear Mixed Model analysis.