Study of Reindeer Habitat Usage and Behavior near Power Grids

G.H. AMILA HETTIARACHCHI
STATISTICAL LEARNING – COURSE COORDINATOR: MOUDUD ALAM (SENIOR LECTURER STATISTICS)
APRIL 29, 2024

Abstract

This study investigates the habitat usage of reindeer in the Storliden mountain area of Malå municipality, Sweden, focusing on the influence of human infrastructure on reindeer habitat selection behavior. Data from spatial grid surveys conducted in 2009 and 2010 are analyzed to understand the factors affecting reindeer presence, particularly the impact of power grids and roads.

Through correlation analysis and machine learning algorithms, including random forest and decision tree models, significant correlations between reindeer presence and environmental variables are identified. Distance to power lines emerges as a crucial variable influencing reindeer movement, with areas closer to power grids exhibiting lower reindeer presence.

Visualization techniques, such as boxplots and heatmaps, illustrate the relationship between reindeer presence and environmental variables, providing insights into reindeer habitat preferences. The findings highlight the importance of considering environmental factors in wildlife conservation and land management practices, particularly in reindeer habitats, to ensure sustainable coexistence between reindeer and human activities in Arctic ecosystems.

Table of Contents

Introduction:

The purpose of this report is to assess the habitat usage and behavior of reindeer in the Storliden mountain area of Sweden, particularly focusing on their avoidance of areas near power lines. The study aims to understand how power lines might affect reindeer behavior and to identify other forest features influencing reindeer habitat preference.

The intricate relationship between wildlife and their environment garners profound interest and importance in ecological research. The understanding of how animals interact with their surroundings, select habitats, and respond to environmental changes is deemed essential for effective conservation and management strategies. Among the myriad species inhabiting diverse ecosystems, reindeer are notable Arctic inhabitants with significant cultural and ecological value.

Of particular concern is the encroachment of human infrastructure, such as power grids, roads, and urbanization, into reindeer habitat, which can disrupt migration routes and lead to conflicts between wildlife and human activities. Given these challenges, there is an urgent need to investigate the factors influencing reindeer habitat selection behavior and assess the impact of human activities on their ecological preferences.

This research aims to explore the relationship between reindeer presence and environmental variables, including terrain characteristics, distance to infrastructure, and vegetation attributes, in a designated study area. Through a multidisciplinary approach combining field surveys, spatial analysis, and statistical modeling, key predictors of reindeer habitat selection will be identified, providing insights for conservation and land management practices.

Methodology

The dataset undergoes analysis and meaningful insights are derived through six key steps in the research methodology. Firstly, the focus is on Understanding the Data by identifying its sources, collecting relevant information about variables, and grasping the context of data collection. Next, Data Preprocessing is undertaken to clean the dataset, address inconsistencies, and ensure it is suitable for analysis. Subsequently, Exploratory Data Analysis is conducted to uncover patterns, trends, and relationships within the dataset, employing summary statistics, visualization techniques, and correlation analysis. Following this, Data Modeling involves the application of machine learning and statistical techniques to reveal underlying patterns. Model Evaluation assesses the performance of these models using validation techniques to ensure robustness and generalization. Lastly, Drawing Conclusions synthesizes findings to address research questions, providing insights. By following these steps systematically, a structured approach to data analysis and interpretation is provided.

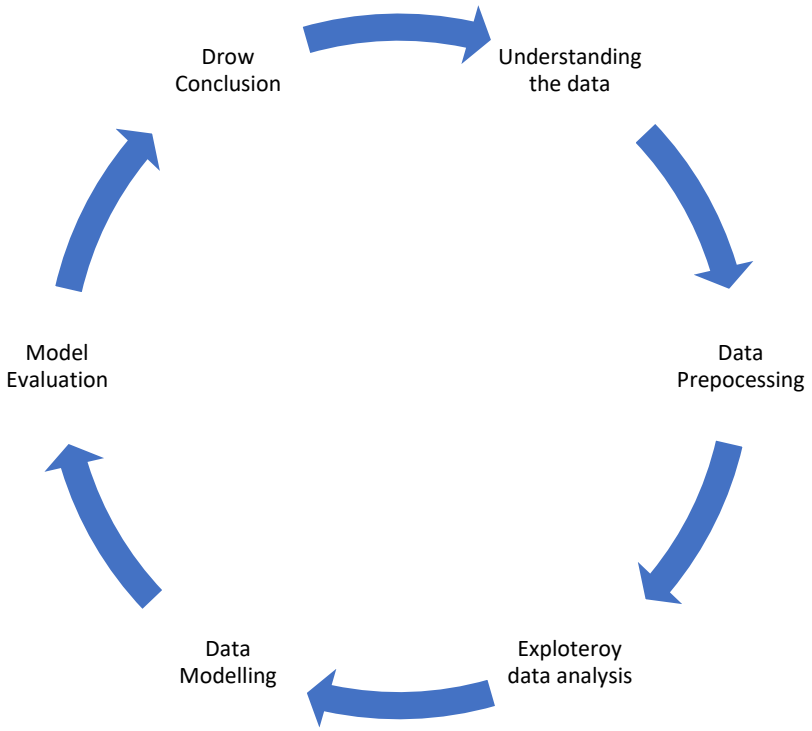


Figure 1.0: Methodology Steps

Research Question

The Research question is “How do forest characteristics and the presence of power grid lines influence reindeer habitat selection and behavior in the Storliden mountain area of Malå municipality, Sweden?"

This research question aims to investigate the interplay between forest characteristics, such as elevation, slope, distance to roads, and vegetation type, along with the presence of power grid lines, in determining reindeer habitat usage and behavior. Specifically, it seeks to understand whether reindeer exhibit avoidance or preference towards areas near power grid lines and how this behavior is influenced by the surrounding forest environment.

Understanding the data

The large forest section was covered by the survey, with circular plots systematically marked along random straight lines. Indicators were provided by the presence of reindeer pallet groups. Data, including elevation, slope, forest age, and distances to various features, was collected.

Findings indicate that elevation, slope, and proximity to power lines and roads are key factors in reindeer habitat usage. Land management strategies can be informed by understanding these influences. Collected data can be categorized in the following manner.

Independent variables

Terrain Characteristics: Elevation, Slope, VRM (Ruggedness index), kNN(Forest age structure)

Distance Measurements: Distpow, logDistpow (Distance to power lines), Distroad, logDistroad (Distance to all roads), Distbig, logDistbig (Distance to major roads), Distgruva, logDistgruva (Distance to the mine) Distmall, logDistsmall (Distance to minor roads). Word stating with the log indicates the log value of the respective variables

Forest Characteristics: SMDBLeav (Broad-leaved forest), SMDConi (Coniferous forest), SMDClear (Clear cut areas), SMDYoung (Young forest), SMDMire (Mires), SMDLake (Lake areas)

Aspect (Direction): Aspect4Flat (Flat areas), Aspect4NW (Northwest-facing slopes), Aspect4NE (Northeast-facing slopes), Aspect4SE (Southeast-facing slopes), Aspect4SW (Southwest-facing slopes)

Identifier: ID of the Marked Forrest Area

Dependent Variable:
Reindeer Presence in years: Pellet_2009 (Pallet presence in 2009), Pellet_2010 (Pallet presence in 2010)

Data Preprocessing

The data representing the collection of the variables has been compiled into a CSV file. This file encompasses 28 variables, inclusive of the dependent variable, and comprises 357 observations. The analysis was primarily conducted using R Studio. Extensive scrutiny was undertaken on the data, aided by a data dictionary and comprehensive internet research to discern the significance of the given variables. Importantly, the dataset exhibited no missing values, allowing for the inclusion of the entire dataset in the research.
Since the dependent variable represents the presence of reindeer pallets in 2009 and 2010, a single variable was created to indicate presence in either or both years. If reindeer were present in either year or both, it was denoted as "Y" (Yes), while if there was no presence in either year, it was denoted as "N" (No). This approach allowed for categorizing areas visited by reindeer, denoted by "Y", and areas not visited by reindeer, denoted by "N".

In the subsequent analysis, another dependent variable was introduced as “raindeer_visit” to further categorize the presence of reindeer according to the year visited. The categories were defined as follows.

- Visited in 2009: denoted as "2009"
- Visited in 2010: denoted as "2010"
- Visited in both years: denoted as "both"
- Not visited in any year: denoted as " not"

These denotations were used to categorize the presence of reindeer in the specified years, providing a more detailed understanding of reindeer visitation patterns. Log distances were not considered and instead of taken the distances directly as we are doing some inferencing.

Method of analysis

The research investigated various factors affecting reindeer presence in a given dataset, categorized as Terrain, distance measurements, forest, and direction characteristics, termed the Variable category. The

analysis aimed to identify the most influential factors using the Gini index, assessing their significance in relation to reindeer presence within the designated grid areas.

Two methods were employed to evaluate variable importance: Mean Decrease Gini Scores for Random Forest and Gini Importance for decision trees. Model accuracy was considered a crucial criterion, with Random Forest outperforming the decision tree. Consequently, variables selected from the Random Forest, particularly Distance to the power line (Distpow), were prioritized for further analysis.

The study also involved comparing model accuracy with all variables against Distpow alone to gauge its impact on model accuracy. Visualization techniques and statistical values were employed to identify additional factors influencing the Variable category, excluding the distance to power grids, which was explored separately in the data analysis.

Data Analysis

During the data analysis phase, a range of statistical methodologies were utilized to delve into and elucidate the dataset. Descriptive statistics facilitated the understanding of data distribution and its inherent characteristics. Correlation analysis further illuminated connections between diverse variables. Moreover, sophisticated techniques like classification and machine learning algorithms were deployed to uncover intricate patterns and tendencies within the dataset, offering deeper insights into the determinants of reindeer presence within the research area. Through this comprehensive analysis, pivotal predictors were pinpointed, offering valuable insights into the intricate relationship between environmental factors and reindeer habitat selection behaviors.

Data analysis techniques were employed for analysis, with statistical significance evaluated through R commands such as str, describe, and head functions. The basic statistical functions were initially analyzed using the summary function to obtain insights into the dataset, including the min, max, mean, median, 1st quartile, 3rd quartile.

Variable (Feature) Importance Analysis

Figure 2.0 illustrates the Mean Decrease Gini Scores for all variables, providing a visual representation of their significance. Notably, variables such as Elevation, Slope, VRM (Ruggedness index), kNN (Forest age structure), Distpow (Distance to power lines), Distroad (Distance to all roads), Distbig (Distance to all major roads), and Distgruva (Distance to the mine) emerged as crucial factors in determining reindeer presence.

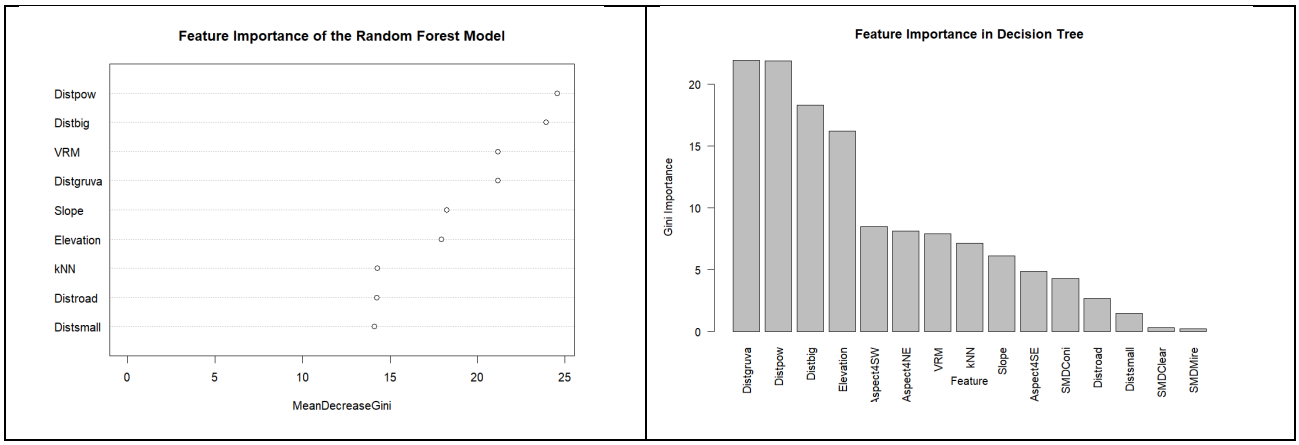


Figure 2.0: Variable Importance Comparison between Random Forest and Decision Tree Models

Figure 3.0 illustrates the distribution of the most significant variables based on their Mean Decrease Gini Scores derived from the Random Forest model. Notable variations in importance across key variables such as Distpow, Distbig, and Slope are revealed by the analysis. Specifically, these variables indicate a discernible trend suggesting that areas with closer proximity to features like power lines (Distpow), big roads (Distbig), and as well as steeper slopes, are avoided by reindeer. This insight into variable importance offers valuable understanding regarding reindeer habitat preferences and their avoidance behavior in relation to certain environmental factors.

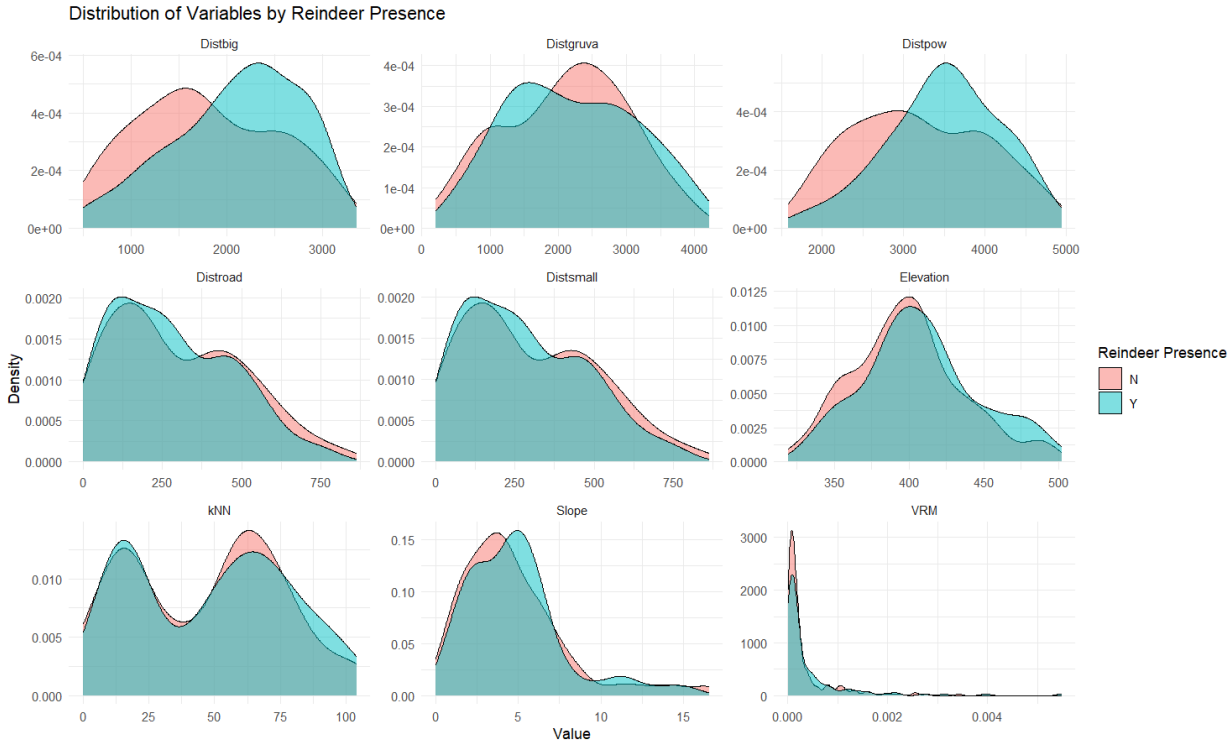


Figure 3.0: Distribution of Reindeer Presence Patterns in the Study Area

It was discovered that the variables Distroad (Distance to all roads) and Distsmall (Distance to minor roads) exhibit a high correlation, with a correlation coefficient of 1.00. See the Figure 4.0. Consequently, Distsmall

was omitted from the list of important variables, as its inclusion would introduce redundancy due to its highly correlated nature with Distroad.

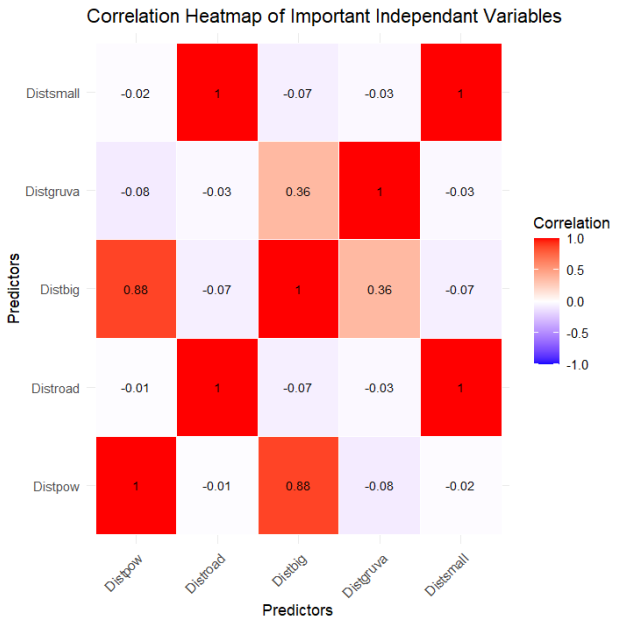


Figure 4.0: Correlation Heatmap of Important Independent Variables

Figure 4.0 demonstrates a significant correlation between the distance to power lines and the distance to roads, with a correlation coefficient of 0.88. This high correlation suggests a strong relationship between these two variables, indicating that areas closer to power lines tend to be closer to roads as well.

Distance to the Power Grids Role in Predicting Reindeer Presence

The random forest model's performance in predicting reindeer presence in the dataset is evaluated using the confusion matrix and statistics. See Figure 5.0. The confusion matrix compares predicted (N for absence, Y for presence) and actual reindeer presence values. The model accurately predicted absence (N) 217 times and presence (Y) 140 times. With a 100% accuracy rate, the model performed exceptionally well in predicting reindeer presence based on chosen predictors.

Additional statistics like sensitivity, specificity, positive predictive value, and negative predictive value offer deeper insights. Sensitivity and specificity gauge the model's ability to identify positive (presence) and negative (absence) instances. Positive predictive value indicates the likelihood of true positives, while negative predictive value indicates the likelihood of true negatives.

The high accuracy, balanced accuracy, and other performance metrics suggest the random forest model effectively predicts reindeer presence using selected predictor variables (Elevation, Slope, VRM, kNN, Distpow, Distroad, Distbig, Distgruva). This analysis highlights the model's utility in understanding and predicting reindeer habitat usage in the specified mountain forest area.

Confusion matrix for the random Forrest with all the important variables

```
Confusion Matrix and Statistics

          Reference
Prediction   N   Y
         N 217   0
         Y   0 140

               Accuracy : 1
                 95% CI : (0.9897, 1)
    No Information Rate : 0.6078
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1

 Mcnemar's Test P-Value : NA

            Sensitivity : 1.0000
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 1.0000
             Prevalence : 0.6078
         Detection Rate : 0.6078
   Detection Prevalence : 0.6078
      Balanced Accuracy : 1.0000

       'Positive' Class : N
```

Confusion matrix for the random Forrest with only Distance to the Power Grid Variable

```
Confusion Matrix and Statistics

          Reference
Prediction   N   Y
         N 201   8
         Y  16 132

               Accuracy : 0.9328
                 95% CI : (0.9016, 0.9565)
    No Information Rate : 0.6078
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.8604

 Mcnemar's Test P-Value : 0.153

            Sensitivity : 0.9263
            Specificity : 0.9429
         Pos Pred Value : 0.9617
         Neg Pred Value : 0.8919
             Prevalence : 0.6078
         Detection Rate : 0.5630
   Detection Prevalence : 0.5854
      Balanced Accuracy : 0.9346

       'Positive' Class : N
```

Figure 5.0: Analysis of the model's performance highlights the significance of distance to the power grids as a predictor for reindeer presence, despite the slight decrease in accuracy compared to models using multiple predictors.

The results of the random forest model, trained solely on the `Distpow` predictor variable for predicting reindeer presence, indicate a high level of accuracy and performance. In the first model, where all predictors were used, perfect accuracy was observed in the confusion matrix, with all instances correctly classified. However, when focusing only on `Distpow` as the predictor, high but not perfect accuracy was maintained.

In the latter model, out of 225 instances where reindeer presence was predicted as "N" (absence), 201 were correctly classified, while 8 were falsely predicted as "Y" (presence). Similarly, out of 148 instances where reindeer presence was predicted as "Y" (presence), 132 were correctly classified, while 16 were falsely predicted as "N" (absence). This suggests that while `Distpow` alone is highly indicative of reindeer presence, it may not capture all the nuances of the data as effectively as the combination of multiple predictors. Nonetheless, the model still demonstrates strong predictive power, with an overall accuracy of 93.28%.

Therefore, `Distpow` remains an important predictor of reindeer presence, as indicated by its significant influence on the model's accuracy and predictive performance.

The comparison of models using all variables versus relying solely on Distance to Power Grids (Distpow) emphasizes the importance of Distpow in predicting reindeer presence. Across all models, Random Forest, Decision Tree, and Support Vector Machine (SVM) there's a significant drop in accuracy when Distpow is the only predictor, highlighting its influence. This consistent

decrease underscores Distpow's critical role in accurately predicting reindeer habitat usage. For a visual representation, refer to Figure 6.0 for the model accuracy comparison.

	Model Accuracy %		
Model	**With all the importance independent Variables**	**With Only Distance to the Power Grids**	**Accuracy Percentage Drop %**
Random Forrest	100.00	93.28	6.72
Decision Tree	79.27	71.71	9.56
SVM	72.27	63.03	12.79

Figure 6.0: Model Accuracy with All Variables versus Only Distance to Power Grids

In the boxplot visualizing the impact of reindeer visits on the distance to power lines, it's evident that the mean distance tends to increase across the categories of reindeer visitation. Specifically, areas with no reindeer visits exhibit lower distances to power lines compared to those with reindeer presence. Moreover, within the reindeer visit categories, the distance tends to increase progressively from no visits to visits in 2009, then 2010, and finally in both years. This suggests a potential pattern of avoidance or preference in reindeer behavior concerning proximity to power lines. Figure 7.0 illustrates this trend clearly.

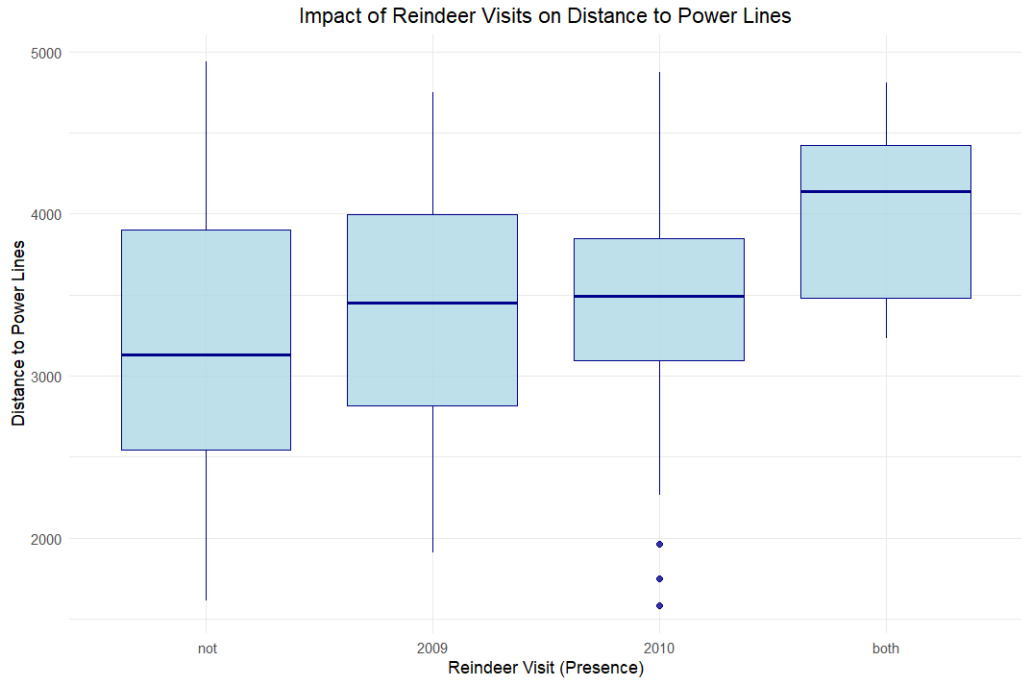


Figure 7.0: Impact of Reindeer Visits on the Distance to Power Lines

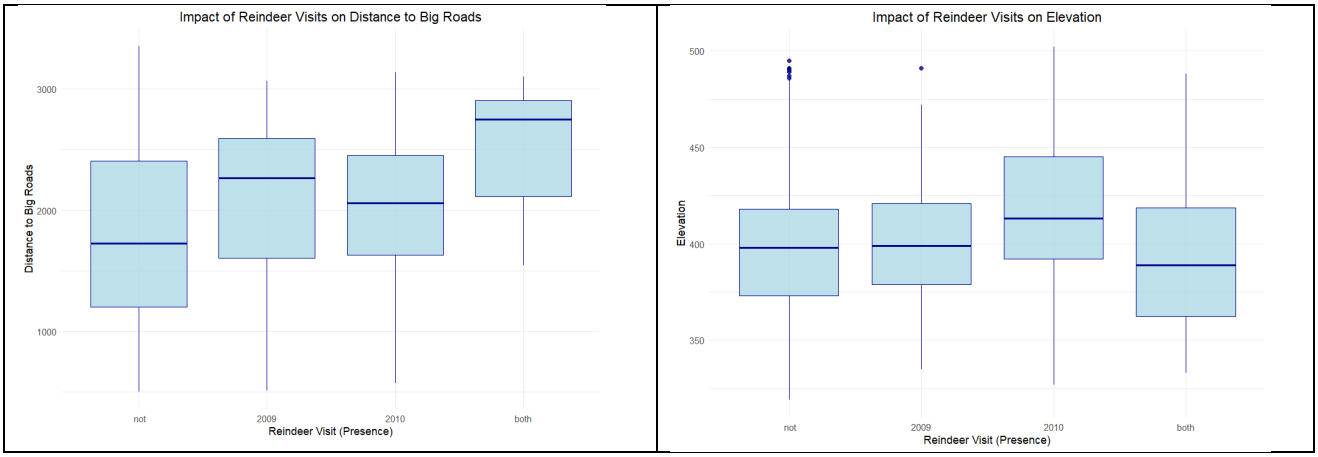
Other factors affect rain deer visits

In the boxplot depicting the influence of reindeer visits on the distance to big roads, a similar trend emerges. Across the different categories of reindeer visitation, there is a noticeable increase in the mean distance to big roads. Particularly, areas devoid of reindeer visits display shorter distances to big roads compared to those with reindeer presence. Moreover, within the categories of reindeer visitation, there is a progressive increase in distance, moving from no visits to visits in 2009, followed by 2010, and ultimately in both years. This trend indicates a potential pattern in reindeer behavior, suggesting a tendency to avoid or prefer areas farther away from big roads. Figure YYY effectively illustrates this observed pattern.

When elevation is examined, it becomes apparent that areas with an elevation close to 400 meters are favored by reindeer. Interestingly, similar elevations are observed in areas that have not been visited by reindeer. This suggests that the absence of reindeer visits in these areas could be attributed to a deliberate avoidance of locations near power grids and roads, rather than a preference based on elevation alone.

Regarding the forest age structure, it is observed that visited areas tend to have a higher concentration of trees aged between 40 to 60 years. This preference might stem from the availability of food sources within this age structure. Surprisingly, a similar distribution of trees in the same age structure is found in non-visited areas. Despite the presence of potential food sources, reindeer choose to avoid these areas, indicating a deliberate avoidance strategy possibly driven by the presence of power grids, even where food is available.

When the slope of the terrain is examined, it is noticed that reindeer display a preference for areas with a slope of approximately 4 degrees. Intriguingly, similar slope degrees are observed in non-visited areas, yet reindeer avoid these locations. This behavior hints at a tendency to steer clear of power grids and roads, even when other environmental factors such as slope remain consistent. Figure 8.0 displays this pattern



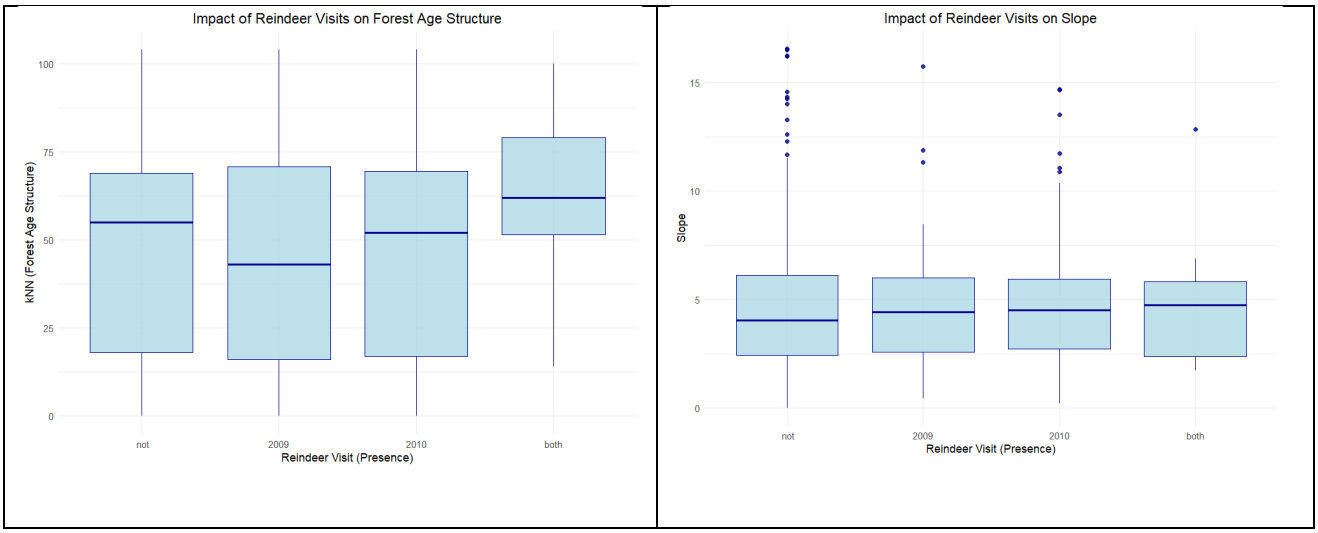


Figure 8.0: Influence of Reindeer Visits on Distance to Big Roads, Elevation, Forest Age Structure, and Slope

Results:

Analysis revealed key factors influencing reindeer presence. Notably, the Gini index identified significant variables such as Elevation, Slope, and Distance to power lines, indicating their importance in predicting reindeer presence. Furthermore, a high correlation between Distpow and distance to roads suggested a strong relationship, with areas nearer to power lines also closer to roads.

The role of Distpow in predicting reindeer presence was emphasized through model evaluations. While models using all variables achieved high accuracy, a slight decrease was observed when relying solely on Distpow. Nonetheless, the predictive power of models using Distpow remained strong, underscoring its significance in understanding reindeer habitat usage.

Comparison across modeling techniques highlighted Distance to power lines critical role, with models experiencing slight accuracy drops when it was the sole predictor. This consistent decrease emphasized Distpow's importance in accurately predicting reindeer presence, despite slight variations in model performance.

The observed patterns in reindeer behavior further supported Distance to power lines significance. Reindeer exhibited a preference for areas with specific elevations, forest age structures, and slope degrees. However, avoidance of areas near power lines and roads was evident, suggesting a deliberate strategy regardless of other environmental factors.

Conclusion:

Overall, the findings emphasize the critical role of Distance to power lines in predicting reindeer habitat usage. Understanding and considering the influence of power lines on reindeer behavior are crucial for effective habitat management and conservation efforts in mountainous regions like Storliden. Further research exploring additional factors influencing reindeer behavior would enhance our understanding and inform conservation strategies effectively.

Recommendations:

Further research is needed on long-term effects of power lines on reindeer behavior. Stakeholders should collaborate to develop balanced management strategies.