Impact of Climbing Activity on Bird Populations at Climbing Areas

Aaron Shikh

December 2024

Contents

1 Introduction **1**

2 Understanding the Data **1**

3 Methodology **15**

3.1 Understanding Multicollinearity 15

3.2 Backward Selection 17

3.3 Random Forest 19

3.4 Ridge and Lasso Regression 21

4 Results and Further Exploration **24**

5 References **24**

1 Introduction

Rock Climbing is an ever more popular sport, that is constantly expanding in both the indoor gym climbing scene, as well as the outdoor scene. Climbing on real rock requires establishing climbing areas by drilling bolts into the rock, allowing for future climbers to use them for protection. This relatively unrestricted development of new climbing areas may cause some issues for wildlife however. Specifically, this paper will explore the effect of climbing development and popularity on endangered bird populations. A survey was done on 32 established and well known climbing cliffs in Boulder, Colorado, and various data on bird counts, climber presence, weather and cliff site metrics were collected. Specifically, data on the number birds and types of birds observed during the survey were compiled into one metric called the Community Conservation Value (CCV), which will be the response variable for this paper. Regression methods like linear regression, backward stepwise, lasso, and ridge will be used, as well as a random forest, to try to predict the CCV from surveyed data.

2 Understanding the Data

To collect the data, researchers went to 32 climbing cliffs in Colorado, and sat next to the cliff for about 1 hour collecting the following data:

- Species Richness: Number of different species of birds observed at all cliffs (S)
- Total number of species / Total number of birds observed at each cliff (RA_i)
- Conservation score: Weighted conservation score from 0 to 4 measuring how endangered the observed species of birds were at each cliff (w_i)

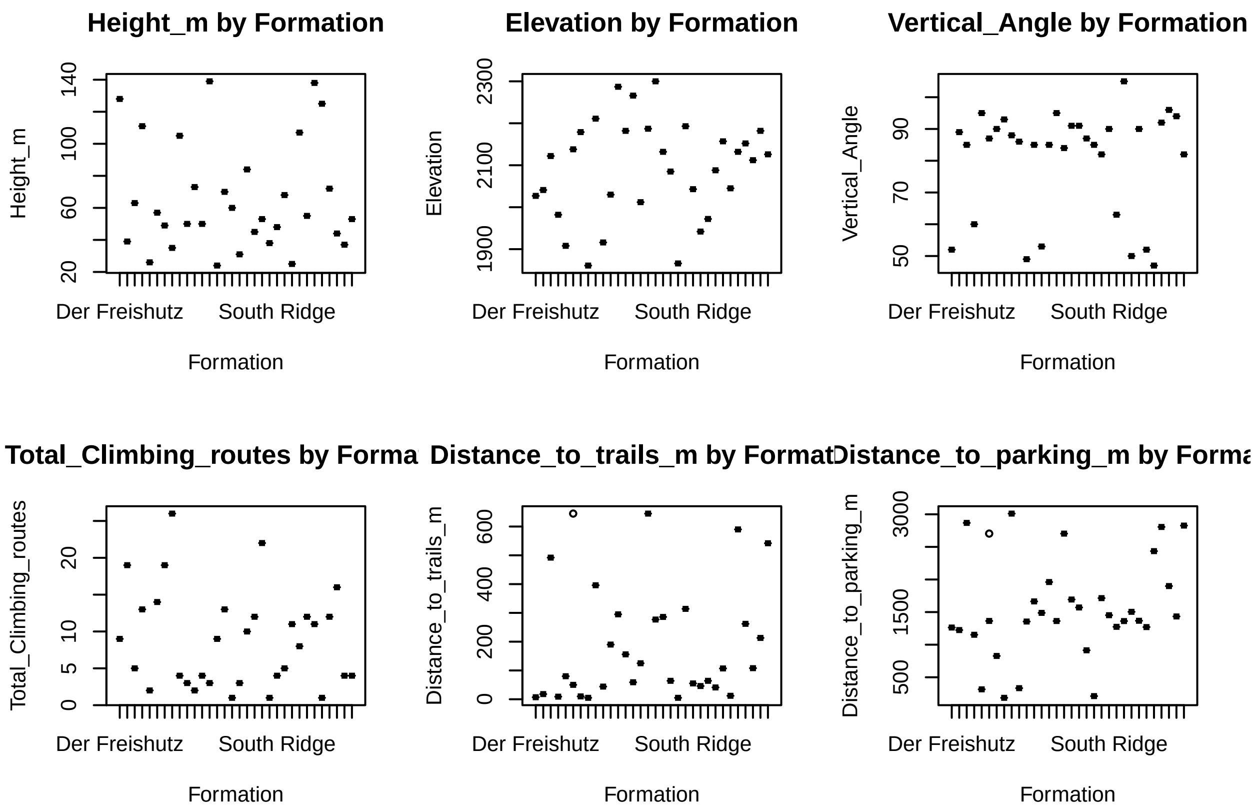
These three metrics were compiled into the one response variable called Community Conservation Value (CCV), which was calculated by the formula $S \sum (RA_i * w_i)$ which sums over each survey area (i). And the following were the predictor variables gathered at each survey area:

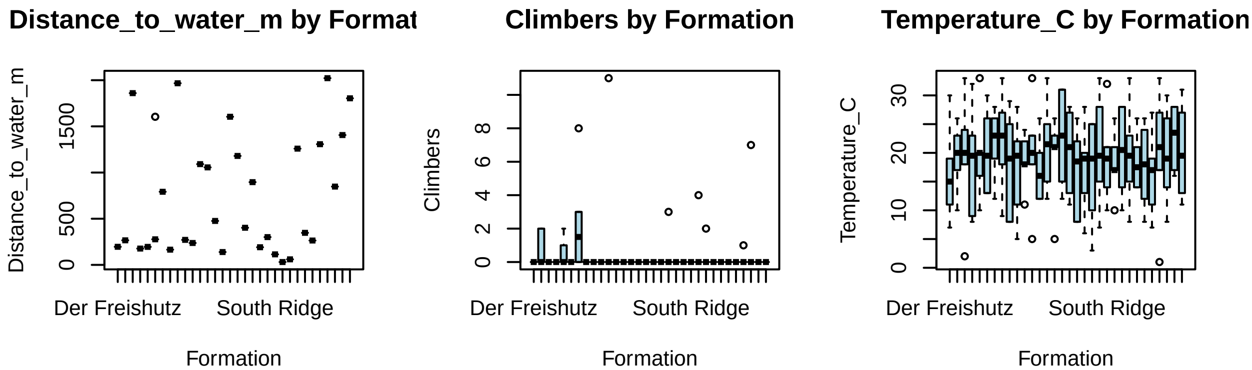
- Formation: Name of specific cliff, or survey site (Categorical, 32 levels)
- Aspect: Cardinal direction (N,S,E,W) that the formation is facing
- Total_Climbing_routes: Number of climbing routes at the surveyed cliff (Continuous)
- Climbing_use: "High" if popular area, "Low" if less popular area (Categorical, 2 levels)
- Climbers: Number of climbers present at the time of survey (Continuous)
- Survey_block: Time of day of the survey, eg. Sunrise, Midday, Evening (Categorical, 3 levels)
- Other cliff metrics (Height, Elevation, Temperature, Vertical_Angle)
- Distance metrics of the cliff site (Distance_to_trails, Distance_to_water, Distance_to_parking)

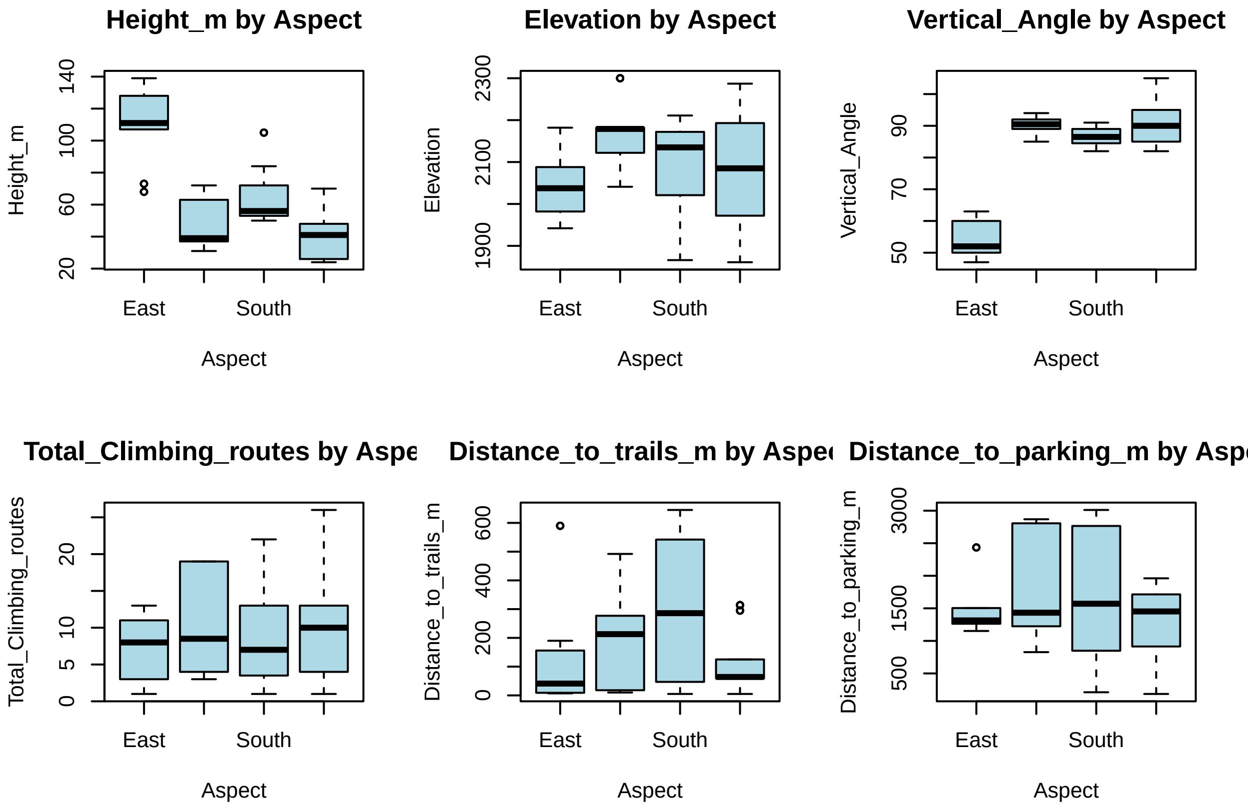
```
data <- read.csv("avian_cliff_communities_dataset.csv")

colnames(data) <- gsub("^X", "", colnames(data))
colnames(data) <- gsub("\\.", "_", colnames(data))
colnames(data) <- gsub("_+", "_", colnames(data))
colnames(data) <- gsub("^_+|_+$", "", colnames(data))

data <- data[, !(names(data) %in% c("Date", 'Climbing_routes_in_survey_area', "Surveyor", "trees", "Veg

data <- data %>% mutate(across(.cols = where(is.character),as.factor))
data <- data %>% rename(CCV = CCV_full_survey_area)

data <- na.omit(data)
attach(data)
```

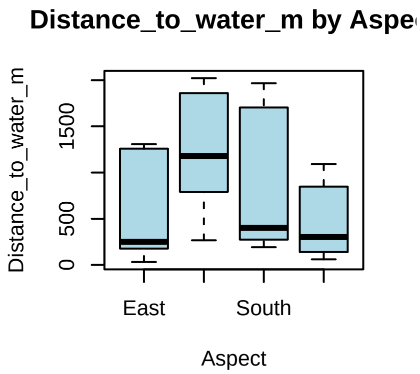
Here we show all of the relationships between the predictors, which can be used for reference whenever a relationship between two predictors is discussed.

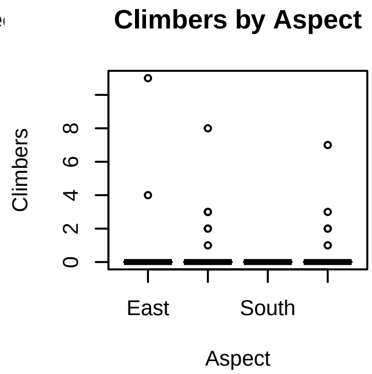
```
numeric_data <- data %>% select_if(is.numeric)
numeric_predictors <- setdiff(colnames(numeric_data), "CCV")
categorical_predictors <- c("Formation", "Aspect", 'Climbing_Use', 'Survey_block', 'Climbers_present')


for (cat in categorical_predictors) {
  par(mfrow = c(2, 3))
  for (num in numeric_predictors) {
    boxplot(data[[num]] ~ data[[cat]],
            main = paste(num, "by", cat),
            xlab = cat, ylab = num,
            col = "lightblue")
  }
}
```

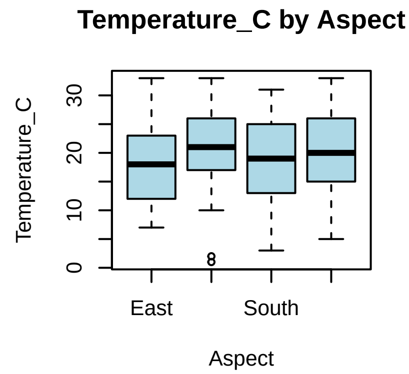

Height_m by Formation
Height_m
20
60
100
140
Der Freishutz
South Ridge
Formation
Elevation by Formation
Elevation
1900
2100
2300
Der Freishutz
South Ridge
Formation
Vertical_Angle by Formation
Vertical_Angle
50
70
90
Der Freishutz
South Ridge
Formation
Total_Climbing_routes by Forma
Total_Climbing_routes
0
5
10
20
Der Freishutz
South Ridge
Formation
Distance_to_trails_m by Format
Distance_to_trails_m
0
200
400
600
Der Freishutz
South Ridge
Formation
Distance_to_parking_m by Forma
Distance_to_parking_m
500
1500
3000
Der Freishutz
South Ridge
Formation

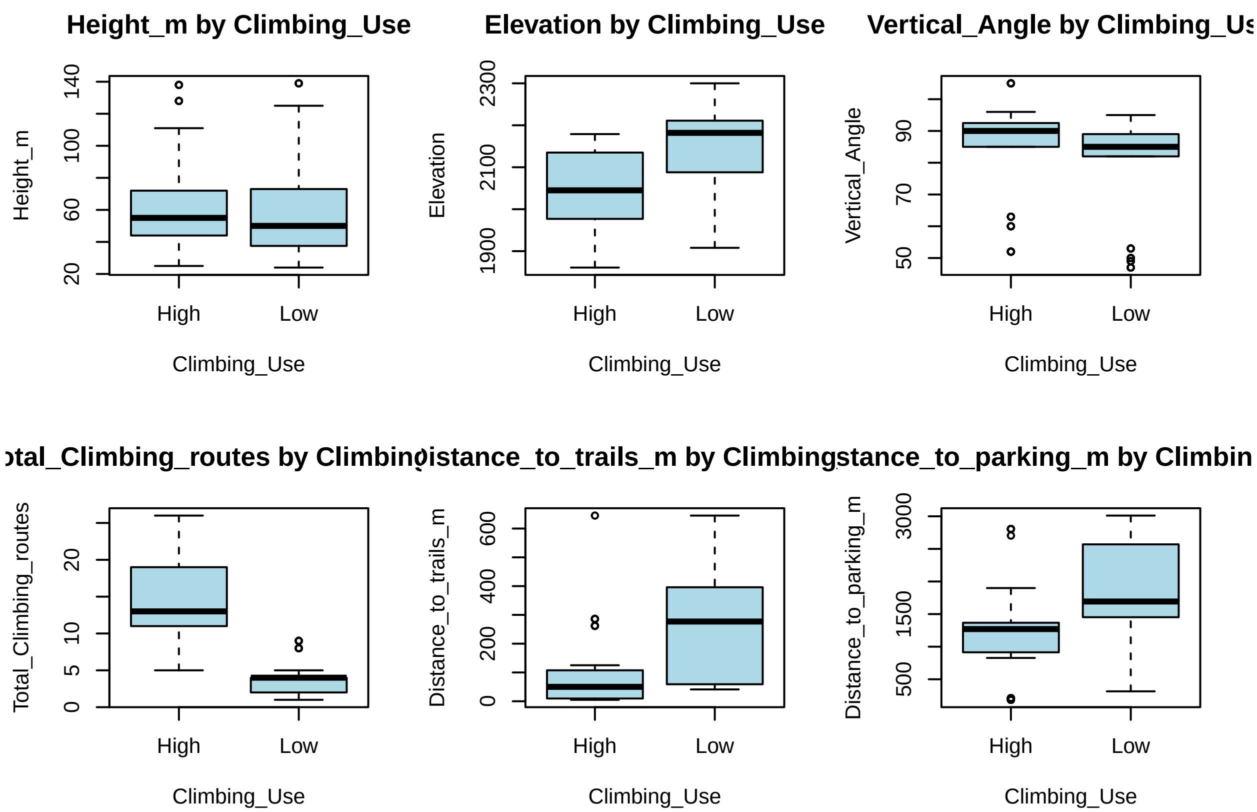

Distance_to_water_m by Format
Distance_to_water_m
0
500
1500
Der Freishutz
South Ridge
Formation
Climbers by Formation
Climbers
0
2
4
6
8
Der Freishutz
South Ridge
Formation
Temperature_C by Formation
Temperature_C
0
10
20
30
Der Freishutz
South Ridge
Formation

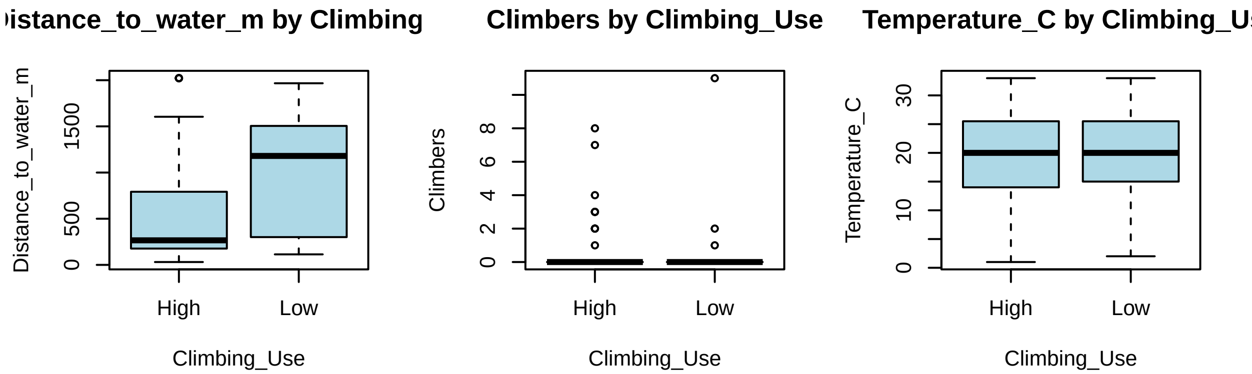

Height_m by Aspect
Height_m
20
60
100
140
East
South
Aspect
Elevation by Aspect
Elevation
1900
2100
2300
East
South
Aspect
Vertical_Angle by Aspect
Vertical_Angle
50
70
90
East
South
Aspect
Total_Climbing_routes by Aspe
Total_Climbing_routes
0
5
10
20
East
South
Aspect
Distance_to_trails_m by Aspe
Distance_to_trails_m
0
200
400
600
East
South
Aspect
Distance_to_parking_m by Asp
Distance_to_parking_m
500
1500
3000
East
South
Aspect


Distance_to_water_m by Aspe
Distance_to_water_m
0
500
1500
East
South
Aspect

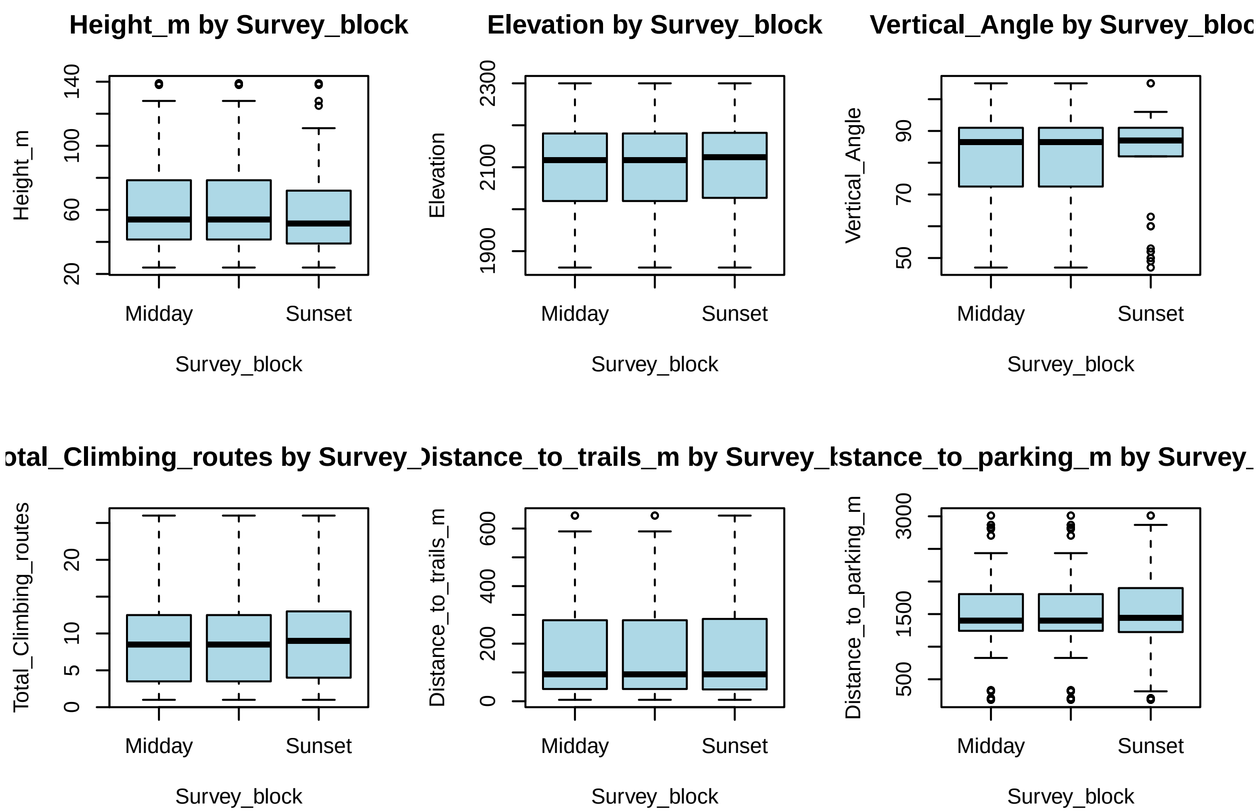

Climbers by Aspect
Climbers
0
2
4
6
8
East
South
Aspect


Temperature_C by Aspect
Temperature_C
0
10
20
30
East
South
Aspect

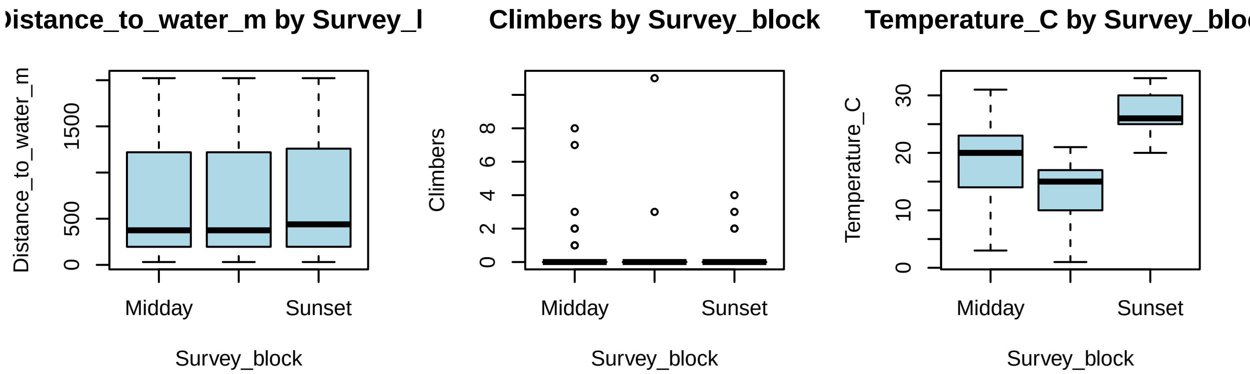

Height_m by Climbing_Use
Height_m
20
60
100
140
High
Low
Climbing_Use
Elevation by Climbing_Use
Elevation
1900
2100
2300
High
Low
Climbing_Use
Vertical_Angle by Climbing_Us
Vertical_Angle
50
70
90
High
Low
Climbing_Use
otal_Climbing_routes by Climbin
Total_Climbing_routes
0
5
10
20
High
Low
Climbing_Use
istance_to_trails_m by Climbing
Distance_to_trails_m
0
200
400
600
High
Low
Climbing_Use
stance_to_parking_m by Climbin
Distance_to_parking_m
500
1500
3000
High
Low
Climbing_Use

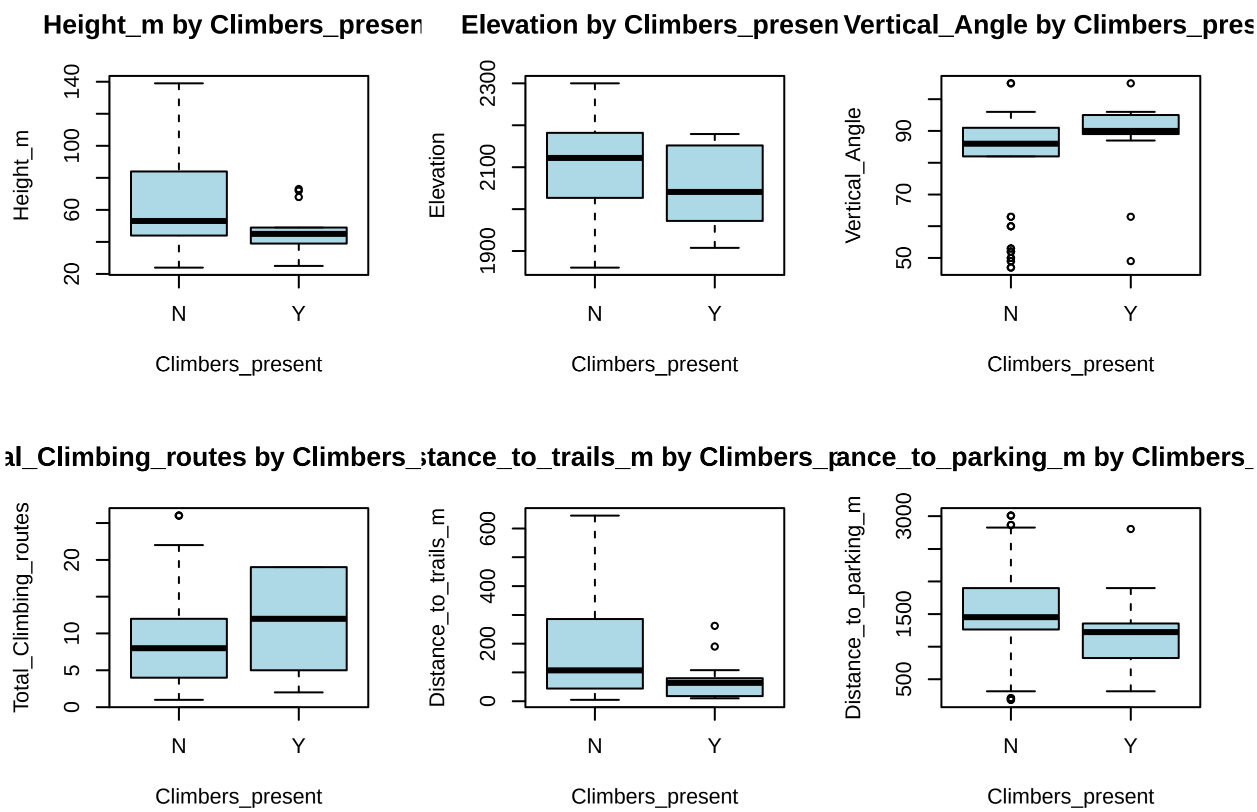


istance_to_water_m by Climbing
Distance_to_water_m
0
500
1500
High
Low
Climbing_Use
Climbers by Climbing_Use
Climbers
0
2
4
6
8
High
Low
Climbing_Use
Temperature_C by Climbing_Us
Temperature_C
0
10
20
30
High
Low
Climbing_Use

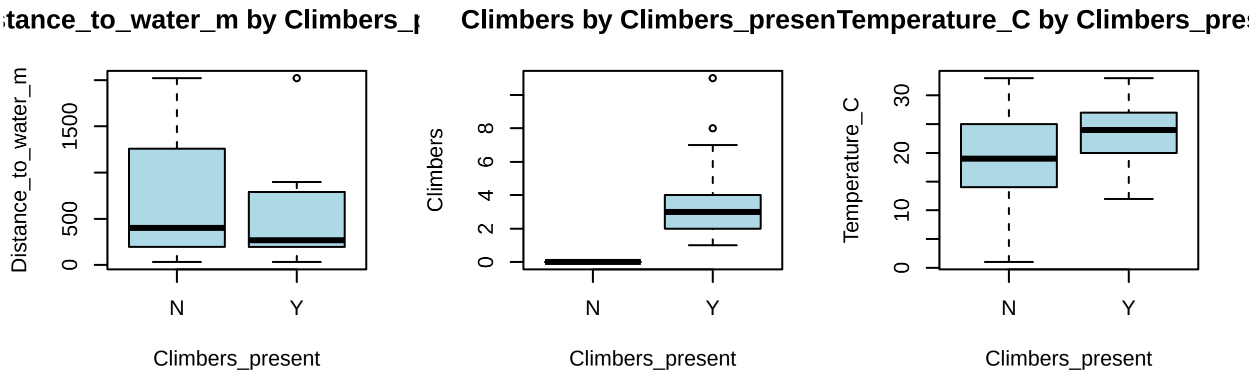


Height_m by Survey_block
Height_m
20
60
100
140
Midday
Sunset
Survey_block
Elevation by Survey_block
Elevation
1900
2100
2300
Midday
Sunset
Survey_block
Vertical_Angle by Survey_bloc
Vertical_Angle
50
70
90
Midday
Sunset
Survey_block
otal_Climbing_routes by Survey_
Total_Climbing_routes
0
5
10
20
Midday
Sunset
Survey_block
Distance_to_trails_m by Survey_
Distance_to_trails_m
0
200
400
600
Midday
Sunset
Survey_block
stance_to_parking_m by Survey_
Distance_to_parking_m
500
1500
3000
Midday
Sunset
Survey_block


istance_to_water_m by Survey_l
Distance_to_water_m
0
500
1500
Midday
Sunset
Survey_block
Climbers by Survey_block
Climbers
0
2
4
6
8
Midday
Sunset
Survey_block
Temperature_C by Survey_blo
Temperature_C
0
10
20
30
Midday
Sunset
Survey_block

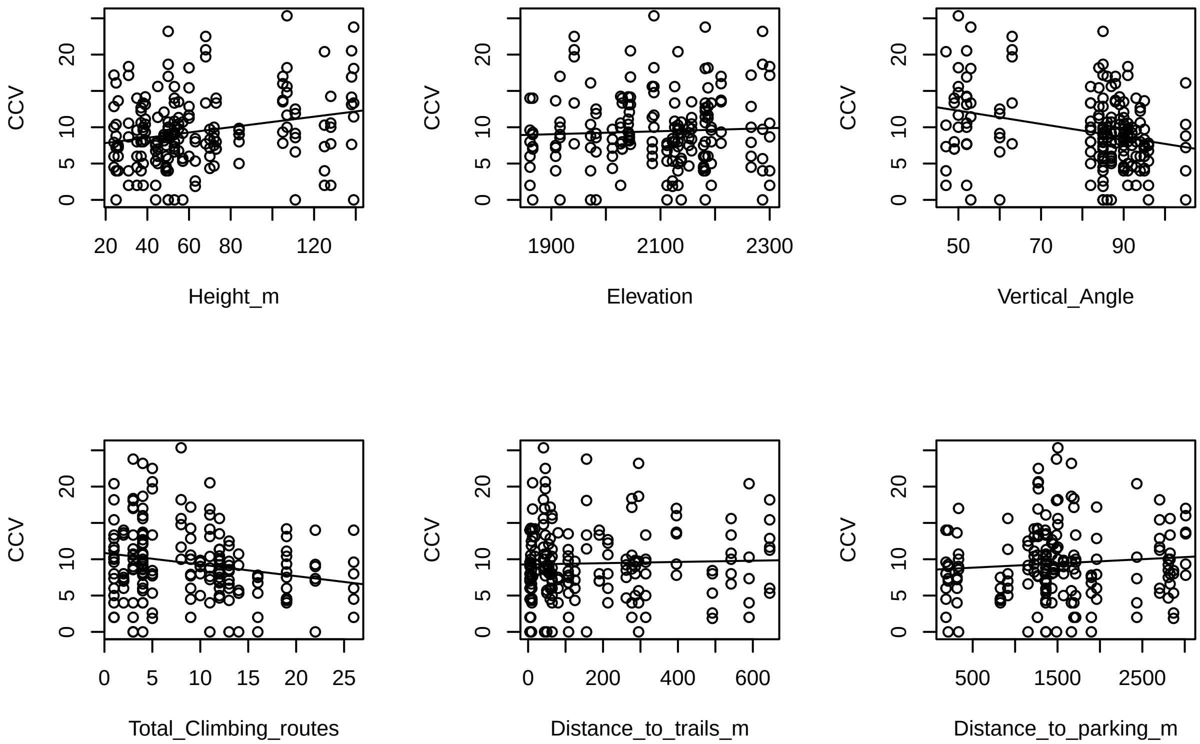



```r
par(mfrow = c(1, 1))
```

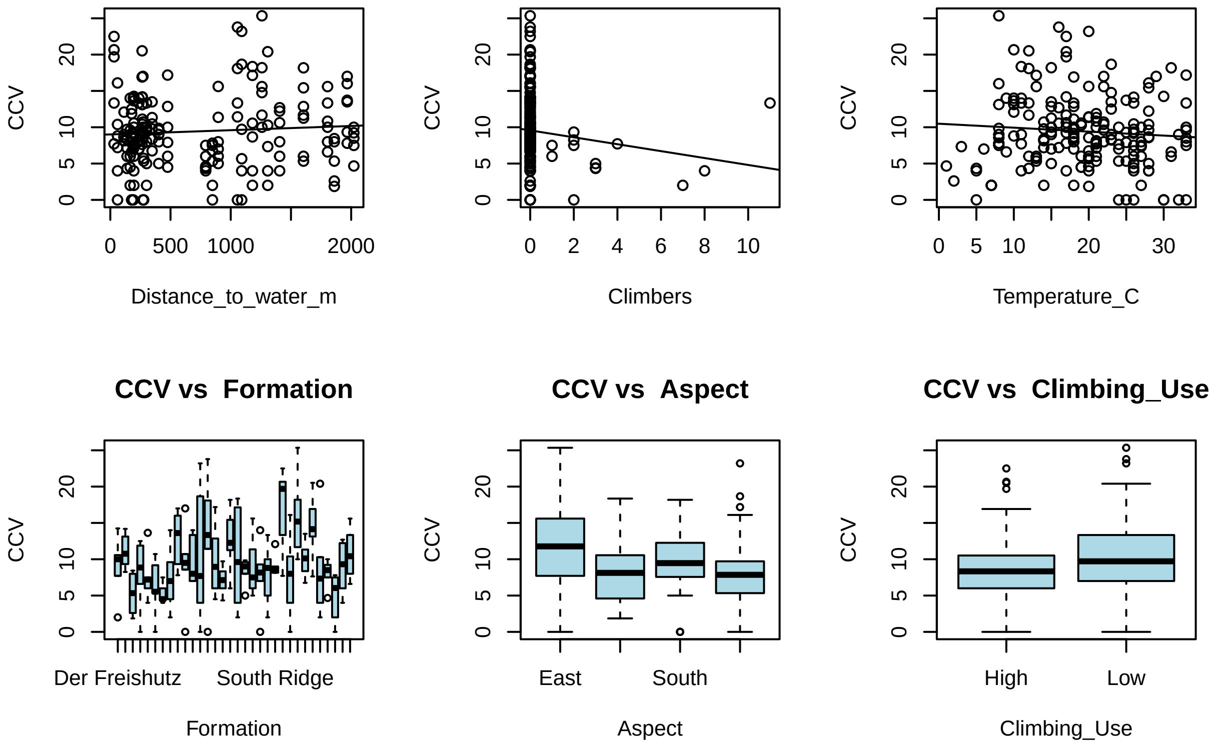


And the following are all the relationships between CCV and the predictors

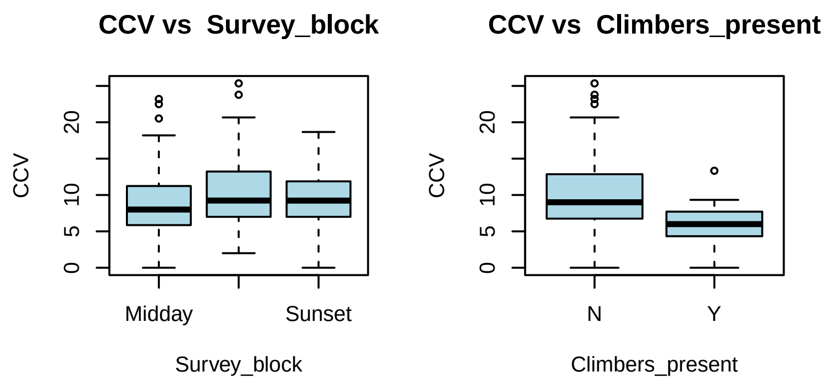
```
par(mfrow = c(2, 3))
for (predictor in numeric_predictors) {
  formula <- as.formula(paste('CCV', "~", predictor))
  model <- lm(formula, data = data)

  plot(CCV ~ data[[predictor]], ylab = 'CCV', xlab = predictor)
  abline(model)
}
```




```
for (predictor in categorical_predictors) {
  boxplot(CCV ~ data[[predictor]],
          main = paste('CCV vs ', predictor),
          xlab = predictor,
          ylab = 'CCV',
          col = "lightblue")
}
```




```r
par(mfrow = c(1, 1))
```



3 Methodology

3.1 Understanding Multicollinearity

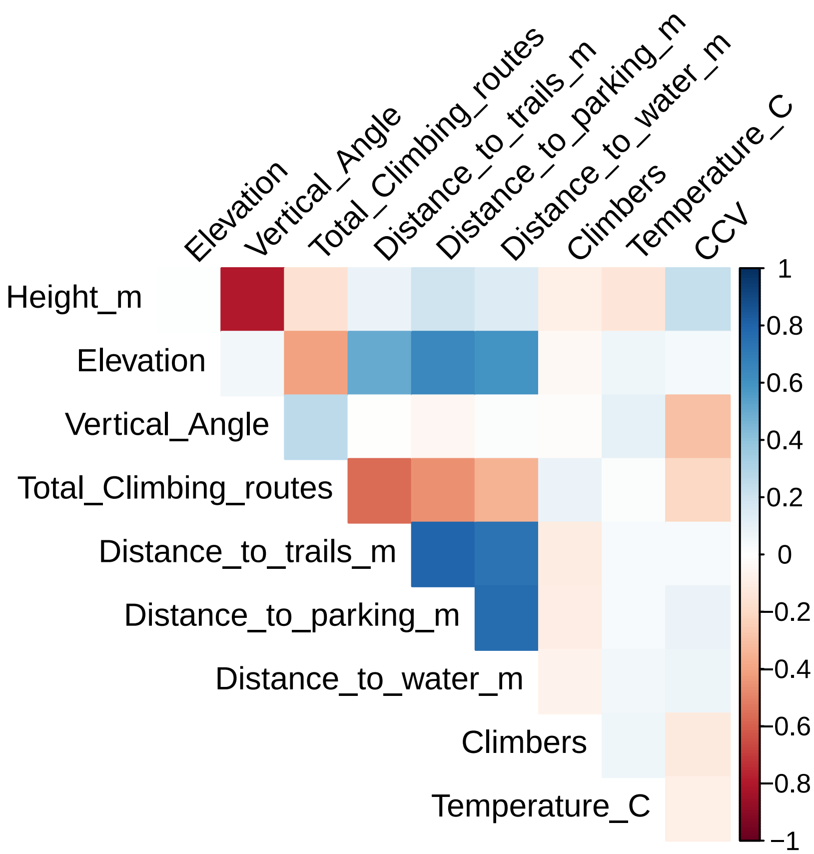
Something we will see in this analysis is that in the end, multicollinearity is present in all of the predictors. In the end, the regression results will essentially be used for feature selection, and to get a better understanding of what factors may effect bird populations, whether it be some of predictors, or some confounding variables not present in the dataset. To explore multicollinearity, we can analyze the VIFs and the correlation matrix.

```
vif_values <- vif(fit <- lm(CCV ~ . -Formation, data = data))
print(vif_values)
```

```
##                            GVIF Df GVIF^(1/(2*Df))
## Height_m               4.769115  1        2.183830
## Climbing_Use           5.635834  1        2.373991
## Aspect                47.838574  3        1.905299
## Elevation              2.515012  1        1.585879
## Vertical_Angle        21.045840  1        4.587574
## Total_Climbing_routes  5.386756  1        2.320939
## Distance_to_trails_m   4.851259  1        2.202557
## Distance_to_parking_m  4.883187  1        2.209793
## Distance_to_water_m    4.770571  1        2.184164
## Survey_block           2.287314  2        1.229791
```

```
## Climbers_present       2.926781  1        1.710784
## Climbers               2.727424  1        1.651492
## Temperature_C          2.327601  1        1.525648
```

```r
cor_mat <- cor(data %>% select_if(is.numeric), use = "complete.obs")
corrplot(cor_mat, method = "color", type = "upper", tl.col = "black", tl.srt = 45, diag = FALSE)
```



From these figures, we can see that some of the predictors form collinear "groups." For example, all of the distance metrics are correlated, as well as the cliff metrics like vertical_angle, elevation, height, and aspect. Formation is also directly correlated with every predictor and hence had to be removed from the VIF chart. Before simply removing predictors, it is also important to remember that confounding variables, relating to bird behavior not recorded in this survey, may be present. These behaviors may be present at different cliffs, and the behaviors could be as simple as birds prefer certain areas over others for nesting, and the locations they choose may not necessarily be related to climbing use.

Next we can attempt to remove certain predictors in this collinear groups, and make our first linear model.

```r
linear_model <- lm(CCV ~ . -Formation - Height_m - Vertical_Angle - Distance_to_trails_m - Distance_to_
summary(linear_model)
```

```
##
## Call:
## lm(formula = CCV ~ . - Formation - Height_m - Vertical_Angle -
##     Distance_to_trails_m - Distance_to_water_m, data = data)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max 
## -12.9432  -2.6693  -0.2578   2.7549  14.3318 
## 
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            7.3191143  8.9033256   0.822 0.412198    
## Climbing_UseLow        0.7240132  1.2638189   0.573 0.567489    
## AspectNorth           -3.5885422  1.3214330  -2.716 0.007302 ** 
## AspectSouth           -2.3864733  1.0670872  -2.236 0.026632 *  
## AspectWest            -3.7784014  1.0577511  -3.572 0.000462 ***
## Elevation              0.0022695  0.0044463   0.510 0.610429    
## Total_Climbing_routes -0.0493494  0.1035691  -0.476 0.634342    
## Distance_to_parking_m -0.0001532  0.0006710  -0.228 0.819688    
## Survey_blockSunrise    0.9516624  0.9425926   1.010 0.314120    
## Survey_blockSunset     0.4989724  1.0692443   0.467 0.641345    
## Climbers_presentY     -2.0517465  2.3523886  -0.872 0.384338    
## Climbers              -0.0679357  0.4680683  -0.145 0.884773    
## Temperature_C         -0.0067247  0.0724501  -0.093 0.926158    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 4.802 on 169 degrees of freedom
## Multiple R-squared:  0.1472, Adjusted R-squared:  0.08663 
## F-statistic: 2.431 on 12 and 169 DF,  p-value: 0.006144
```

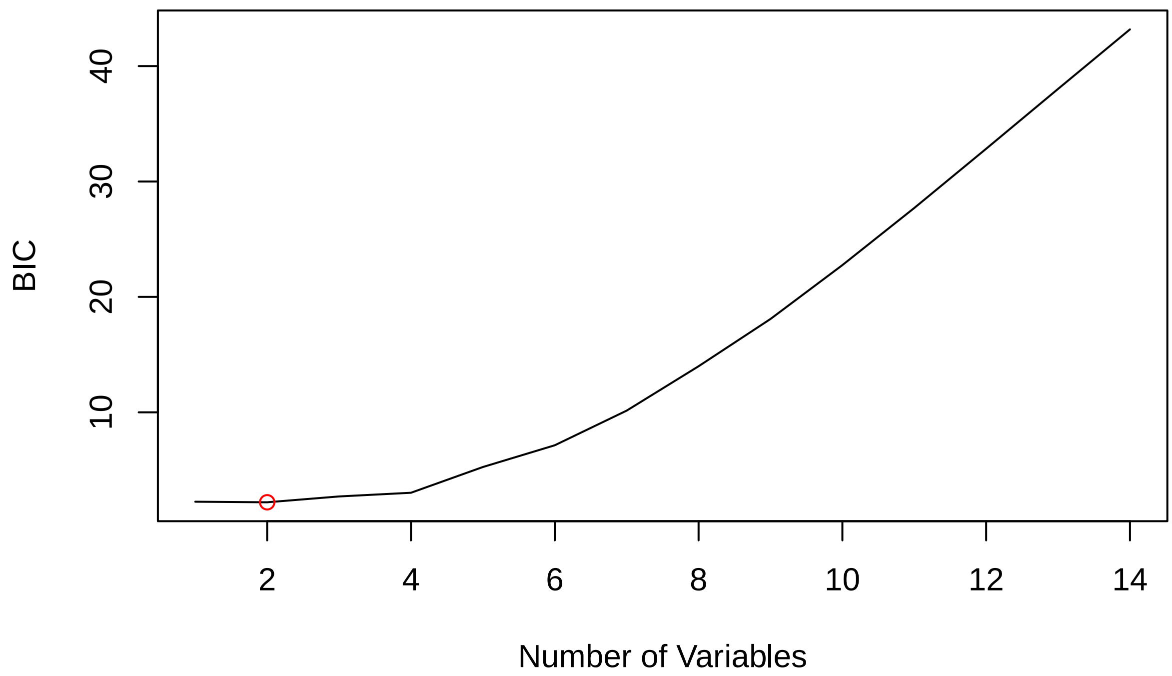
This model shows that when all predictors are used, only the cliff Aspect is significant in predicting the CCV.

3.2 Backward Selection

Because we may have one or two predictors that may be the most significant in the model, we will perform backward selection instead of forward to better see the difference in the models.

```r
#backwards


backward_model <- regsubsets(CCV ~ . -Formation, data = data, nvmax = ncol(data) - 1, method = "backwar
model_summary <- summary(backward_model)

results <- data.frame(
  Model_Size = 1:length(model_summary$adjr2),
  R2 = model_summary$rsq,
  Adjusted_R2 = model_summary$adjr2,
  CP =  model_summary$cp,
  BIC = model_summary$bic
)


results$Formula <- apply(model_summary$which, 1, function(row) {
  included_predictors <- names(row)[row]
  paste(included_predictors[-1], collapse = " + ")
})
```

```
lowest_bic <- which.min(model_summary$bic)
plot(model_summary$bic, xlab = "Number of Variables", ylab = "BIC", type = "l")
points(lowest_bic, model_summary$bic[lowest_bic], col = "red")
```




```
print(head(results[order(results$BIC), ]))
```

```
##   Model_Size         R2 Adjusted_R2        CP      BIC
## 2          2 0.07101894  0.06063926  9.698347 2.204640
## 1          1 0.04381910  0.03850698 13.135450 2.252929
## 3          3 0.09469733  0.07943942  6.965164 2.709585
## 4          4 0.11867574  0.09875881  4.172007 3.028034
## 5          5 0.13296251  0.10833076  3.316159 5.257543
## 6          6 0.14861952  0.11942933  2.186406 7.144947
##
## 2                                                                                    AspectWest + Total_C
## 1                                                                                                 Total_C
## 3                                                                AspectWest + Total_Climbing_routes + Cli
## 4                                           AspectWest + Total_Climbing_routes + Distance_to_trails_m + Cli
## 5                              AspectNorth + AspectWest + Total_Climbing_routes + Distance_to_trails_m + Cli
## 6 AspectNorth + AspectWest + Total_Climbing_routes + Distance_to_trails_m + Distance_to_water_m + Cli
```

From these results we can see that predictors like aspect, Climbers_present, and total_climbing_routes, which is correlated to climbers and Climbers_present, are the most significant predictors of CCV and result in the lowest BIC, although the R^2 is very low, indicating a very poor explanation of the variance of CCV.

3.3 Random Forest

```
#random forest

set.seed(198)
train <- sample(1:nrow(data), nrow(data) * 0.7)
test_data <- data[-train, "CCV"]

mtry <- floor(ncol(data) / 3)
forest_model <- randomForest(CCV ~ . -Formation, data = data, subset = train, mtry = mtry, ntree = 100,
yhat.rf <- predict(forest_model, newdata = data[-train, ])

MSE <- mean((yhat.rf - test_data)^2)
print("Test set MSE:")
```

```
## [1] "Test set MSE:"
```

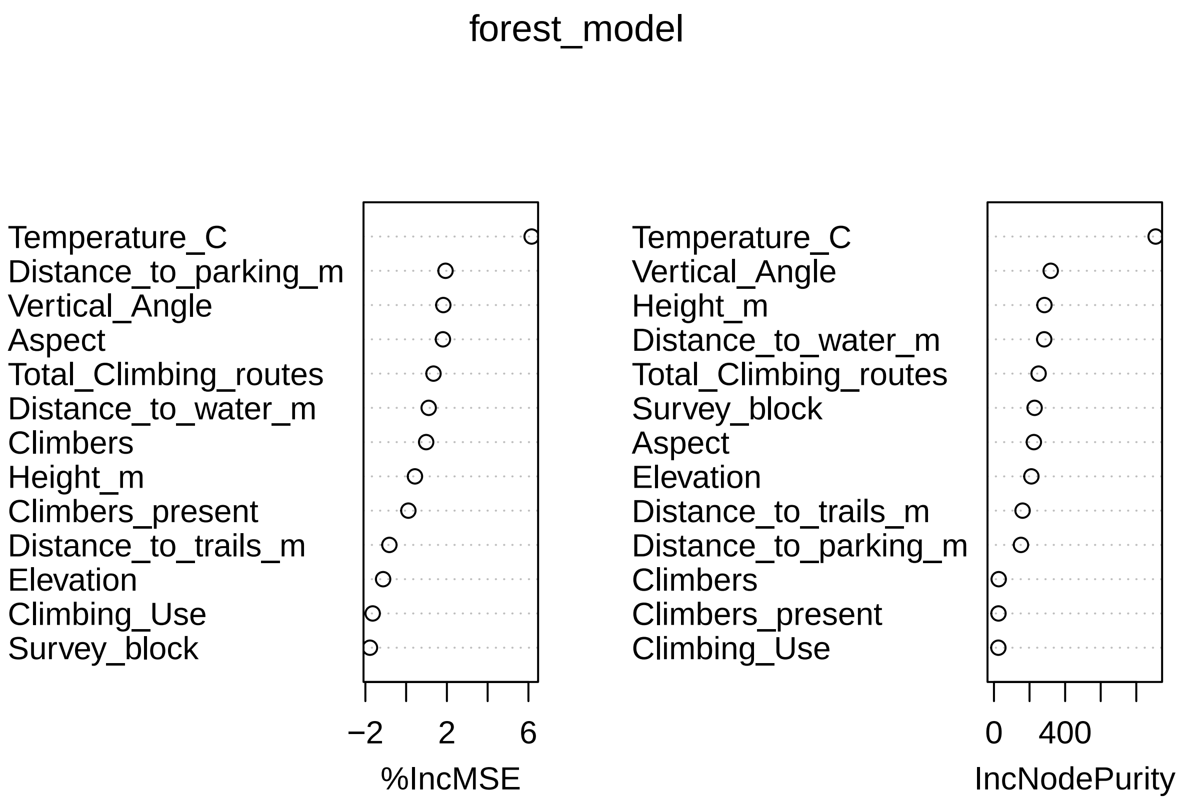
```
print(MSE)
```

```
## [1] 20.84419
```

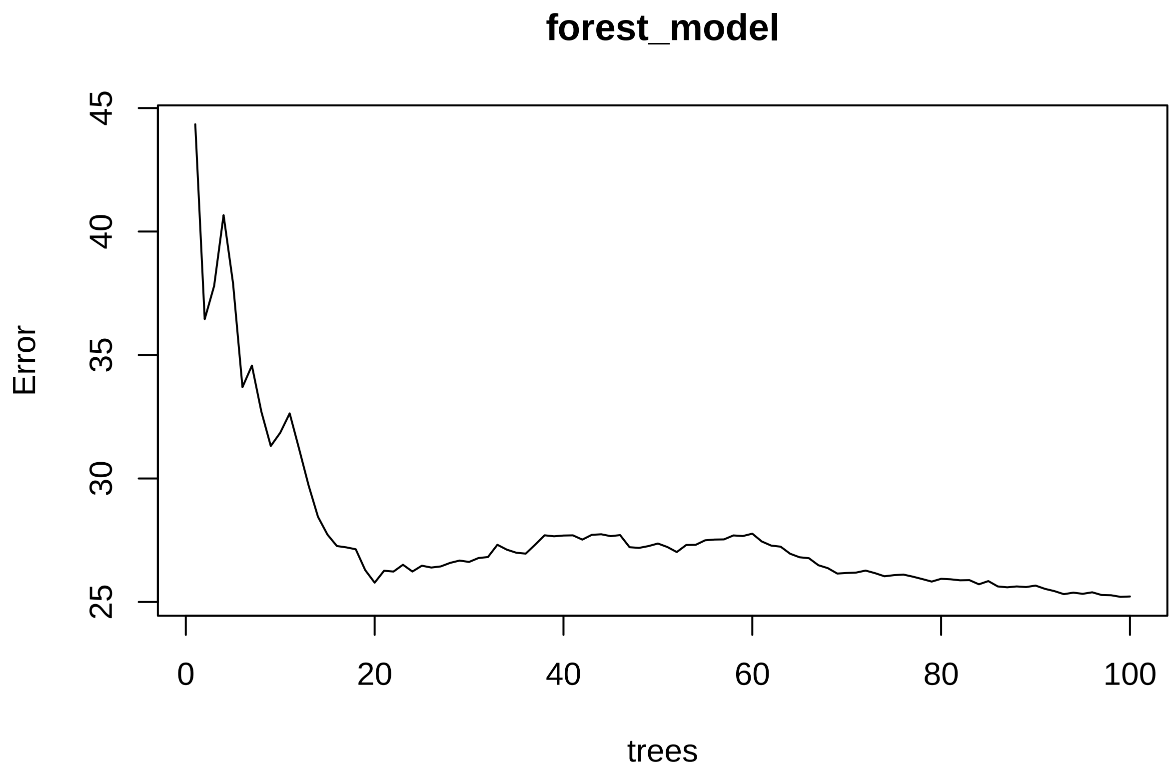
```
importance(forest_model)
```

```
##                          %IncMSE IncNodePurity
## Height_m               0.4323646     283.50106
## Climbing_Use          -1.6416519      24.74851
## Aspect                 1.8057819     224.15803
## Elevation             -1.1300448     210.25118
## Vertical_Angle         1.8231433     318.95556
## Total_Climbing_routes  1.3423247     250.92776
## Distance_to_trails_m  -0.8216514     160.75926
## Distance_to_parking_m  1.9319074     151.71240
## Distance_to_water_m    1.1065115     282.35652
## Survey_block          -1.7744987     228.36133
## Climbers_present       0.1091684      25.31828
## Climbers               0.9827647      26.64622
## Temperature_C          6.1579090     908.74746
```

```
varImpPlot(forest_model)
```




```
plot(forest_model)
```



The analysis showed that environmental and physical factors of the cliffs, such as Temperature_C, Vertical_Angle, and Height_m, are the most significant predictors of CCV, while Climbing_Use has minimal impact. With a negative %IncMSE and low IncNodePurity, Climbing_Use does not appear to strongly influence CCV, which means that CCV is mainly predicted by environmental factors rather than climbing activity. However, climbing related variables like Climbers and Climbers_present show slight importance, suggesting that climbers may have a slight impact, potentially influencing CCV.

3.4 Ridge and Lasso Regression

```
#ridge + lasso

train_indices <- sample(1:nrow(data), size = 0.7 * nrow(data))
train_data <- data[train_indices, ]
test_data <- data[-train_indices, ]


X_train <- model.matrix(CCV ~ . - Formation, data = train_data)[,-1]
y_train <- train_data$CCV
X_test <- model.matrix(CCV ~ . - Formation, data = test_data)[,-1]
y_test <- test_data$CCV

set.seed(1278)
lasso_model <- cv.glmnet(X_train, y_train, alpha = 1, nfolds = 10)
```

```r
lasso_lambda_min <- lasso_model$lambda.min
print("Best Lambda for Lasso: ")
```

```
## [1] "Best Lambda for Lasso: "
```

```r
print(lasso_lambda_min)
```

```
## [1] 0.8172761
```

```r
y_pred_lasso <- predict(lasso_model, s = "lambda.min", newx = X_test)
mse_lasso <- mean((y_test - y_pred_lasso)^2)
print("MSE for Lasso on Test Set:")
```

```
## [1] "MSE for Lasso on Test Set:"
```

```r
print(mse_lasso)
```

```
## [1] 26.2441
```

```r
set.seed(3345)
ridge_model <- cv.glmnet(X_train, y_train, alpha = 0, nfolds = 10)

ridge_lambda_min <- ridge_model$lambda.min
print("Best Lambda for Ridge:")
```

```
## [1] "Best Lambda for Ridge:"
```

```r
print(ridge_lambda_min)
```

```
## [1] 37.9346
```

```r
y_pred_ridge <- predict(ridge_model, s = "lambda.min", newx = X_test)
mse_ridge <- mean((y_test - y_pred_ridge)^2)
print("MSE for Ridge on Test Set:")
```

```
## [1] "MSE for Ridge on Test Set:"
```

```r
print(mse_ridge)
```

```
## [1] 24.70008
```

```r
lasso_coefs <- coef(lasso_model, s = "lambda.min")
print("Lasso Coefficients:")
```

```
## [1] "Lasso Coefficients:"
```

```r
print(lasso_coefs)
```

```
## 17 x 1 sparse Matrix of class "dgCMatrix"
##                             s1
## (Intercept)           9.93549
## Height_m              .
## Climbing_UseLow       .
## AspectNorth           .
## AspectSouth           .
## AspectWest            .
## Elevation             .
## Vertical_Angle        .
## Total_Climbing_routes .
## Distance_to_trails_m  .
## Distance_to_parking_m .
## Distance_to_water_m   .
## Survey_blockSunrise   .
## Survey_blockSunset    .
## Climbers_presentY     .
## Climbers              .
## Temperature_C         .
```

```r
ridge_coefs <- coef(ridge_model, s = "lambda.min")
print("Ridge Coefficients:")
```

```
## [1] "Ridge Coefficients:"
```

```r
print(ridge_coefs)
```

```
## 17 x 1 sparse Matrix of class "dgCMatrix"
##                                   s1
## (Intercept)            9.289331e+00
## Height_m               1.817768e-03
## Climbing_UseLow        1.238287e-01
## AspectNorth           -6.247643e-02
## AspectSouth           -1.252899e-03
## AspectWest            -1.333257e-01
## Elevation              4.493631e-04
## Vertical_Angle        -4.851003e-03
## Total_Climbing_routes -1.242278e-02
## Distance_to_trails_m  -1.596866e-05
## Distance_to_parking_m  5.948211e-05
## Distance_to_water_m    9.896589e-05
## Survey_blockSunrise   -3.013860e-04
## Survey_blockSunset     5.539128e-02
## Climbers_presentY     -2.927789e-01
## Climbers              -1.525685e-02
## Temperature_C         -4.645001e-03
```

The lasso and ridge analysis showed that environmental factors, such as Temperature_C, Height_m, and Vertical_Angle, are the most significant predictors of CCV, while climbing-related variables like Climbing_Use and Climbers_present have limited significance. Ridge regression achieved a lower MSE (24.7) compared to Lasso (26.24), as Ridge retains small contributions from all variables, including correlated ones, while Lasso excluded climbing related predictors entirely by shrinking their coefficients to zero. The Ridge model retained small coefficients for environmental factors, show their that they are important in predicting CCV.

4 Results and Further Exploration

Overall, we can see that the linear model, random forest, lasso and ridge regression, all show that that climbing predictors have a relatively low impact on CCV values. In particular, we find that cliffs facing west may have slightly larger and more diverse bird populations. And and predictors like temperature, which may correlate to the time of day the survey was taken, are more significant than predictors like Climbing_use or Climbers_present. It is also worth noting that in the original diagnostic plots, the box plot of Climbers_present vs CCV seemed to show that the presence of climbers had a lower CCV score, but if we look at climbers vs CCV, and the distribution of climbers during the survey times, we see that almost all the surveys taken had 0 climbers. This means that in the few instances climbers were present, there were slightly fewer birds probably because birds may not want to be around people. This isn't enough evidence to say that climbers effect bird populations however. If we look at the box plots of formation vs CCV, we see that every formation has a different CCV value, which proves that Formation, which is correlated with almost every predictor including external confounding variables about bird behavior, is the main cause of CCV variance.

Some limitations of this study are that there was never a control group taken, on cliffs that are not used for climbing. In theory, we can account for the external variables about bird behavior by also surveying more cliffs without any established climbing routes, and then compare both groups. This would however, also require more cliffs to be surveyed, to better account for and understand the environmental factors at play. Additionally, it is simply very hard to account for the many external environmental and bird behavior variables, that it would take an enormousness amount of effort to better understand the impact of climbing on the CCV of different areas.

5 References

Covy N, Benedict L, Keeley WH (2019) Rock climbing activity and physical habitat attributes impact avian community diversity in cliff environments. PLoS ONE 14(1): e0209557. https://doi.org/10.1371/journal.pone.0209557