Exploring the Effect of Revegetation Practices on Avian Biodiversity

Paper Replication

Isabel Arvelo and Gracie Guidotti

2022-11-30

1. Introduction:

This is a replication and extension of a paper by Martine Maron, *Threshold effect of eucalypt density on an aggressive avian competitor*, published in Biological Conservation Journal in 2007 (Maron).

A common method of landscape restoration is replanting and rebuilding the soil of land where vegetation is disturbed or absent to improve biodiversity. However, the success of this process may be affected by interspecific interactions that favor particular predators and negatively impact certain target species, which has an overall adverse effect on the entire assemblage. Typical woodland revegetation focuses on the establishment of tree species that grow quickly and blend in aesthetically. In the Wimmera region of western Victoria in Australia, *Eucalyptus* and *Acacia* species are common choices for plantings, even in locations originally dominated by the slower growing buloke. In New South Wales these plantings intended to recreate suitable habitat for threatened and declining native birds, have been dominated by the noiser miner (*manorina melanocephala*), a highly aggressive species that excludes other bird species. Noisy miners primarily inhabit eucalypt woodland and the increase in planted eucalypts, especially in buloke revegetation and degraded buloke woodland, has the potential to have a major impact on the birds that will inhabit the maturing revegetation.

Martine Maron aimed to identify the impact of current revegetation practices on bird species that inhabit mature revegetation and investigated the relationship between site-level factors, such as eucalypt density, and noisy miner invasion. She specifically attempted to determine whether there was a certain number of eucalypts in a buloke woodland that resulted in the presence of the Noisy Miner bird species, as well as the impact of noisy miners on surrounding bird species.

2. Methods:

Study Area

In order to measure the presence of the Noisy Miner species, the researchers studied sites in the Wimmera Plains region of western Victoria, Australia. The majority of the native vegetation in the area occurs in small patches of private land, public reserves, and roadsides, with the private land being heavily degraded from grazing activities. All sites were independent of one another.

31 Buloke woodland patches were studied, mainly on private land (with plant reserve areas) and one on unreserved Crown land (meaning that it is own landed by the Crown and has not been set aside for public use). Within each site, a 100m x 200m belt transect was haphazardly located with at least one edge of the belt within 50m of the edge of the woodland. This was done to increase likelihood of detecting Noisy Miners because Noisy Miners are considered an edge species.

Study Sites

Each transect was surveyed on 3 occasions:
1. 13-17 December, 2004
2. 28-31 March, 2005
3. 18-22 September, 2005

In order to survey for birds, researchers walked through the middle of the transect for 20 minutes and recorded all birds seen or heard.

Contributing Datasets:

nminer: Noisy miner abundance dataset in GLMSData

The publicly available dataset has 31 observations, each of which represents one of the 31 buloke woodland patches studied. The response variables of interest are Miners, the presence or absence of noisy miners, and Minerab, the number of noisy miners observed across the three surveys.

The available predictor variables are Eucs (the number of eucalypts in each 2 hectare transect), Area (the area in hectares of contiguous remnant patch of vegetation in which the transect was located), Grazed (whether the area was grazed or not), Shrubs (whether shrubs were present in the transect or not), Bulokes (the number of buloke trees in each 2 ha transect), and Timber (the number of pieces of fallen timber in the transect).

This varies slightly from the variables they use in their analysis because for each transect they recorded total numbers of buloke trees, shrubs, and pieces of fallen timber, and all eucalypts identified and counted. In four 3 m diameter circular quadrats randomly located within each transect, the percentage cover of bare ground, cryptogams, grass, fine litter and coarse litter were visually estimated, and the means for the transect calculated. However, we only have shrubs as a binary variable, and we do not have data on the percentage cover of each type of material for the transect. In their final model, they include shrubs as a continuous variable and fine litter, but we do not have access to either of these variables in our dataset.

Species Abundance data from Appendix 1 of paper

Appendix 1 in the paper includes a table that lists the number of sites with and without noisy miners in which each bird species was observed during the three surveys. A total of 56 species were recorded during surveys, including 11 declining woodland species (for details see Appendix 1).

3. Statistical Analysis:

Exploratory Data Analysis

Visualization

The authors did not do a lot of exploratory data analysis, but since we do not have all of the same explanatory variables at our disposal and do not have the same level of domain knowledge, we performed an initial investigation of the data using visualizations. These plots help summarize its main characteristics and uncover patterns or relationships within the data, primarily with bivariate visualizations of the explanatory variables against each other and against the response.

K-medioids clustering

Since the researchers are interested in site level factors, we also thought it would be interesting to cluster the data into non-hierarchical clusters to identify groups with similar site level factors that could potentially help researchers identify the best re vegetation practices for specific groups of ecosystems.

K-means clustering is a popular algorithm because it's a simple and fast classification method, but it uses Euclidian distance, which is only defined for numeric values and therefore not appropriate for binary variables like the ones in our data set (Kaufman). Instead, we opted to use Gower's distance (Gower), which uses a type of distance that is appropriate for each type of variable. The final dissimilarity between the ith and jth unit is obtained as a weighted sum of dissimilarities for each variable. For our analyses, we are assuming $w_k = 1$ for all k.

$$D_{Gower}(x_i, x_j) = \frac{\sum_k \delta_{ijk} d_{ijk} w_k}{\sum_k \delta_{ijk} w_k}$$

In particular, d_{ijk} represents the distance between the ith and jth unit computed considering the kth variable, while w_k is the weight assigned to variable k. For this analysis we are weighing all variables equally.

We use this to create distance matrix and then use a k-medoid method to minimizes the sum of dissimilarities between points labeled to be in a cluster and a point, referred to as the mediod, designated as the center of that cluster.

Bird Assemblages

The presence of noisy miners on bird assemblages is also a major focus of this paper.The two metrics the paper specifically considers are species richness and species abundance. However, without site-level species data, we can not directly address either of these questions because we don't know which or how many species were present at which sites. We only know how many sites each species was present in, with and without noisy miners.For example, the Galah bird was present in 6 out of the 17 sites where noisy miners were present (.35) and 12 out of the 14 sites where they were absent (.85). If there was no difference between sites with and without noisy miners, we would expect species to be present at the same relative frequency for each type of site. Comparing these values for each species could still give us some indication of whether each species is more or less likely to inhabit an ecosystem when noisy miners are present. In other words, the question we are asking is if the presence of noisy miners is associated with whether or not native species will inhabit a community.If we find that species are observed at lower frequencies among sites with noisy miners than without noisy miners, we would have evidence to conclude that noisy miners' aggressive behaviors do indeed exclude other birds from the territories they occupy.

The typical choice for this type of comparison would be a paired t-test, but the vector of the difference in proportions is not normally distributed. The paired-sample sign test is a non-parametric hypothesis test that makes no assumptions about the shape of the population distribution. In order to determine whether the median of a population is equal to a default value, the one sample sign test compares the number of observations greater than or less than the default value without accounting for the magnitude of the difference between each observation and the default value. In order to compare the pair of values for each bird, we will do a sign test on the difference between the sample pairs.

In our analysis, we rejected the null hypothesis. There is sufficient evidence that the relative frequency of species significantly varies between sites with and without noisy miners present. Therefore, we can conclude that noisy miners make native species less likely to inhabit an ecosystem. We cannot directly compare this result to the paper's findings because they had access to a lot more site-level information, but they found that the total abundance of birds did not differ significantly between sites with and without noisy miners, but species richness did.

Noisy Miner Presence

The collected data on the presence of Noisy Miners was converted into a binary presence/absence variable. We began by building a full model with all of the predictors and then used the backwards stepwise selection technique and likelihood ratio tests to determine the appropriate predictors of the model. For the likelihood ratio test, we used a significance value of $P<0.05$ in order to choose our variables for the final model.

Noisy Miner Counts

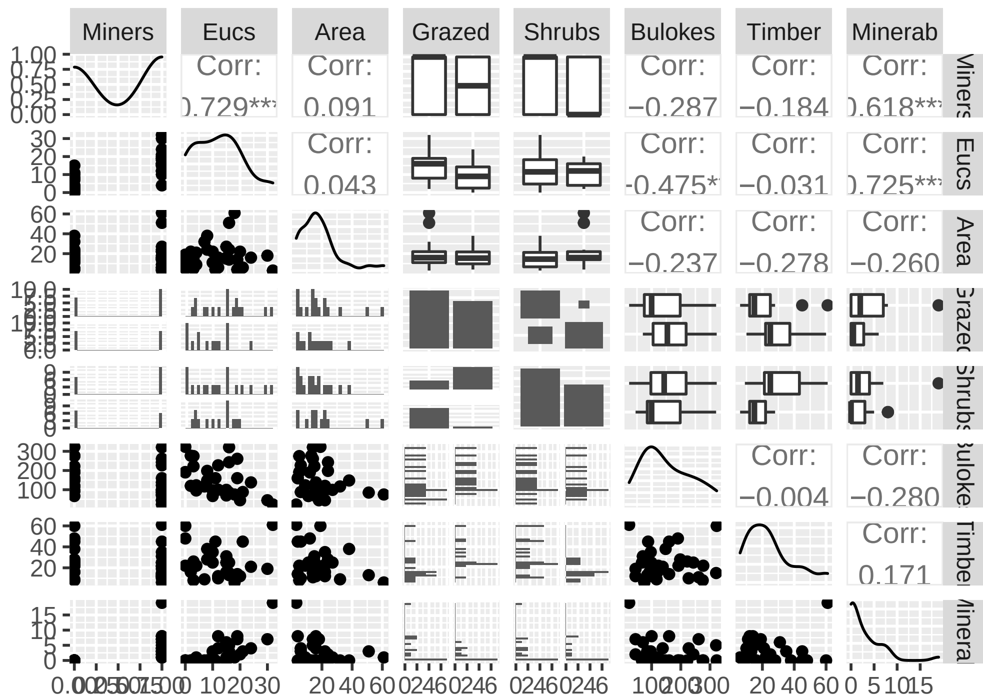
Next, we considered the *Minerab* variable that represents the number of noisy miners (abundance) observed in three 20 minute surveys. The number of noisy miners is highly skewed right (see Appendix) a Poisson model is an appropriate choice to model this response variable. We began by building a full model with all the variables included in the model and then found evidence of overdispersion so moved forward with both a quasi-poisson and negative binomial model.

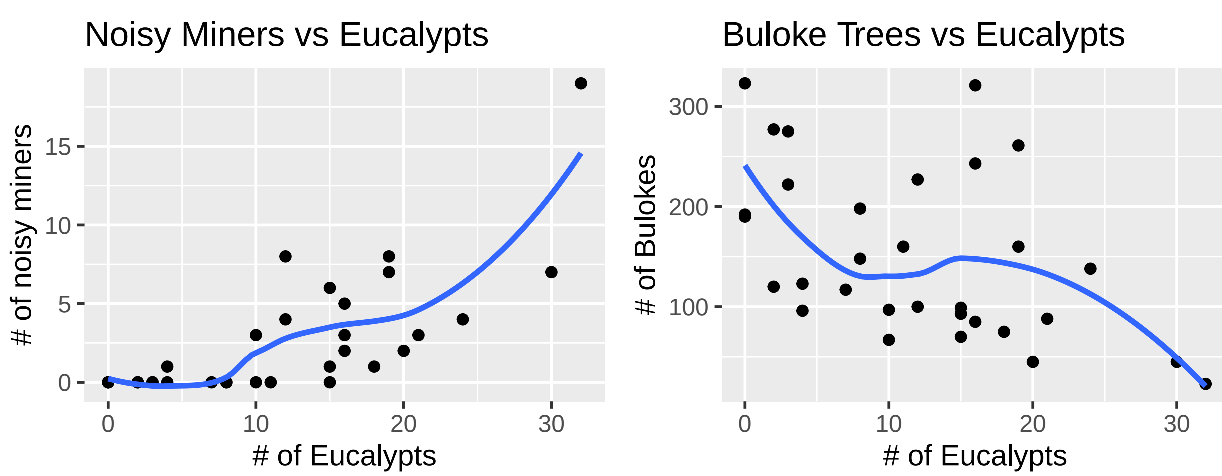
We used backward stepwise regression to identify which variables had a significant effect on the number of noisy miners. The likelihood ratio method was used to identify variables whose removal did not significantly reduce model fit (at a significance level of P $<$ 0.05) and these were removed from the model. The paper used a significance level of P<0.10, probably because of the small sample size, but we decided to test our hypotheses at $\alpha = 0.05$ to have more confidence in our inferences.

3. Results

Exploratory Data Analysis

Bivariate Relationships

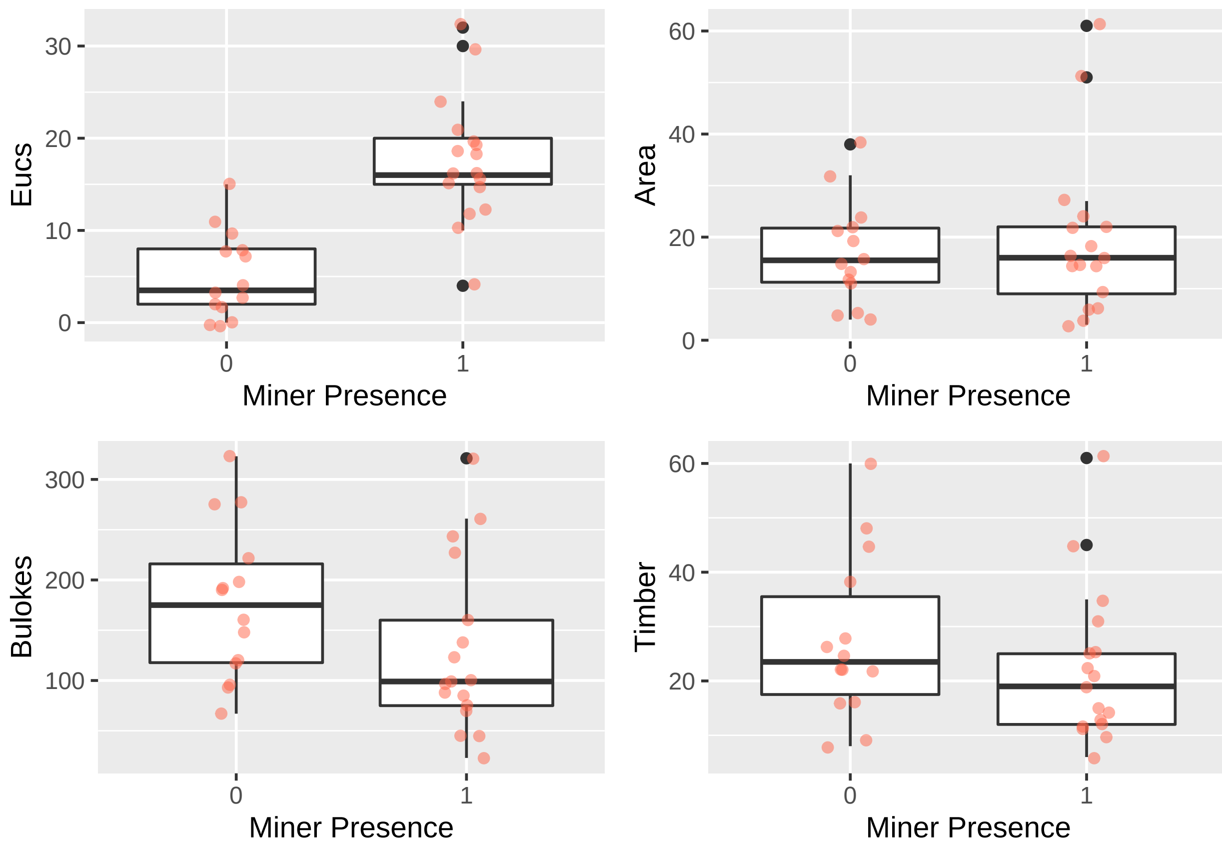




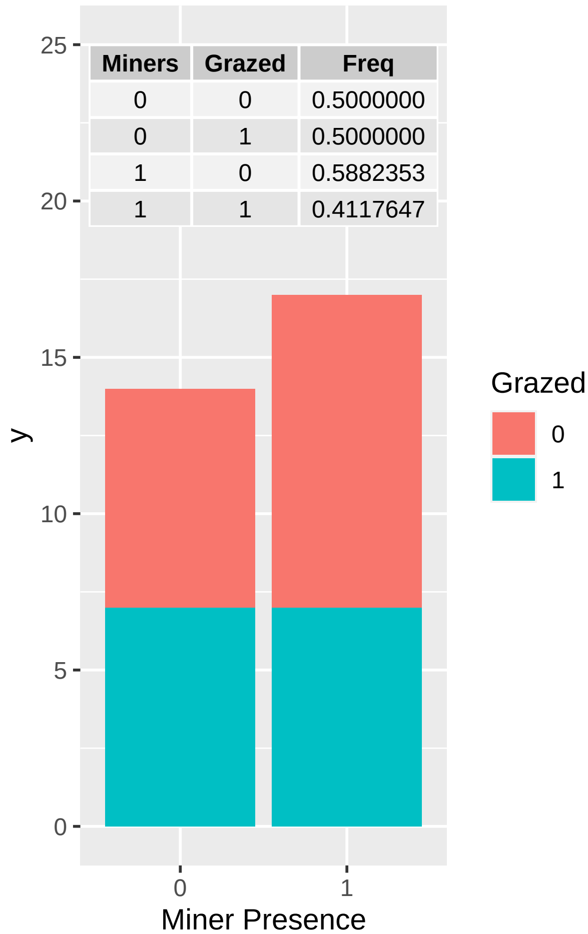
We began with a matrix of plots that illustrate the relationship between all the pairwise combinations of variables. The two most noteworthy relationships above seem to be between Eucs and Minerab ($r = 0.725$)

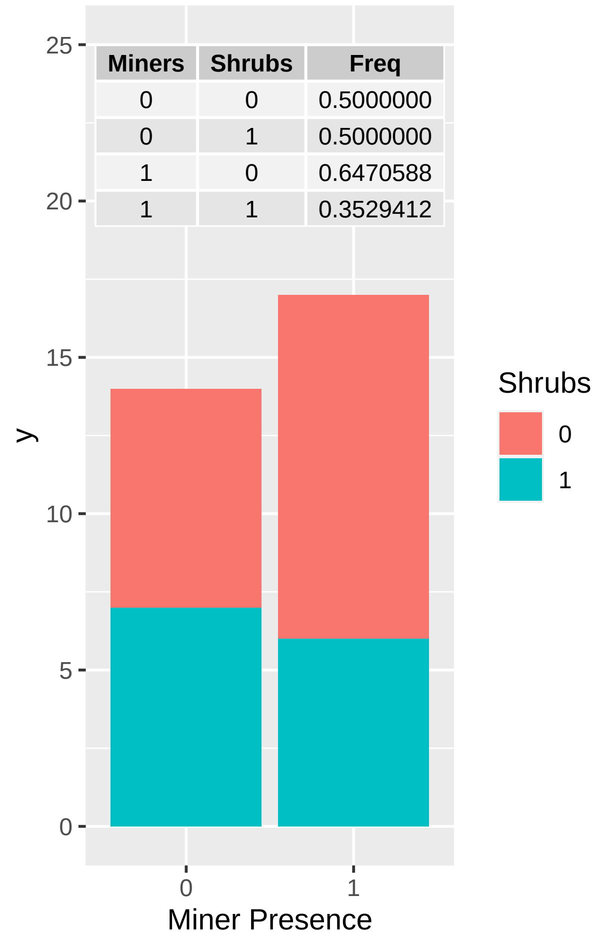
and Eucs and Bulokes ($r = -0.475$). Looking at these relationships closer, we found that transects with a greater number of eucalypts tend to be inhabited by more noisy miners and less Buloke trees.

We then specifically wanted to look at the distribution of the response variables of interest, Miners and Minerab, against the different levels of the explanatory variables.

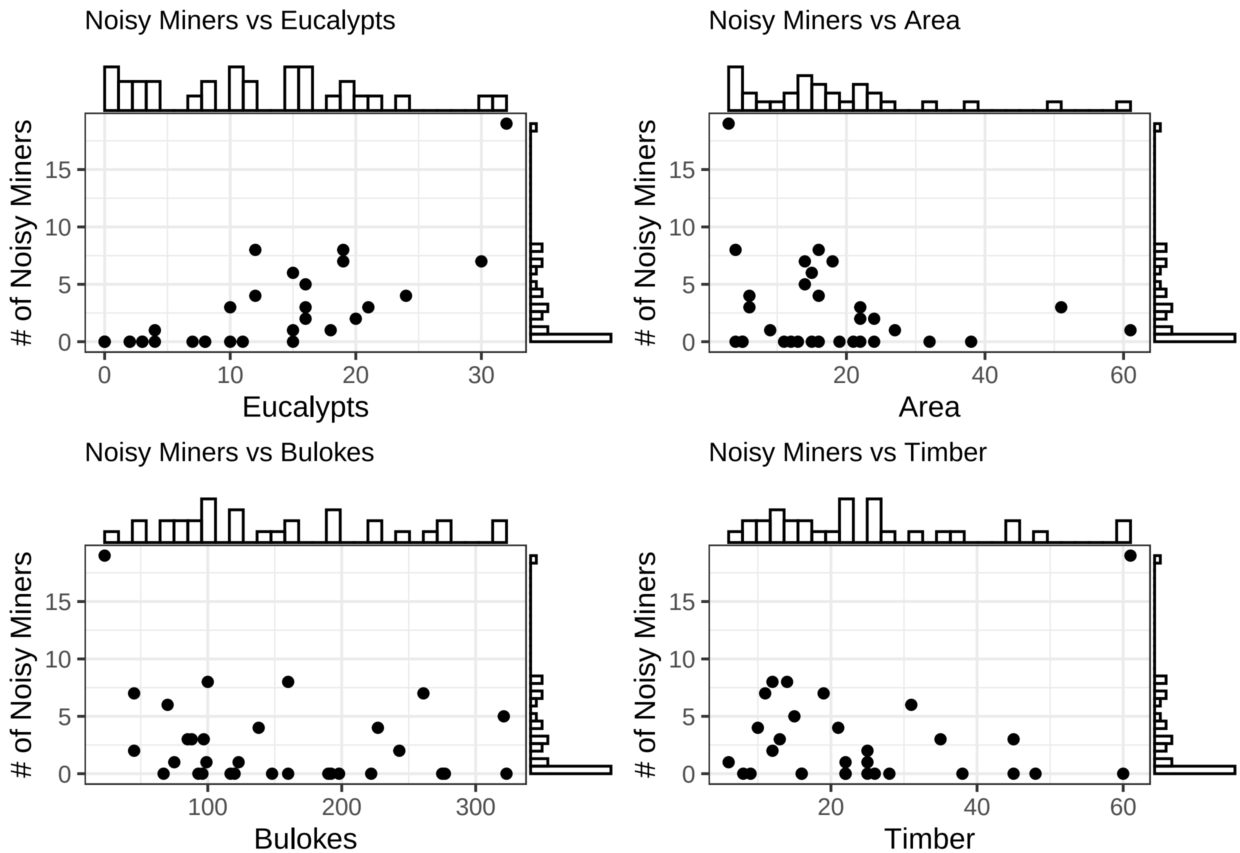


These plots are further evidence that sites with miners present tend to have more eucalypts and less bulokes than sites without noisy miners. Sites without miners have a wider distribution for the number of piece of fallen timber. Area does not vary by miner presence.

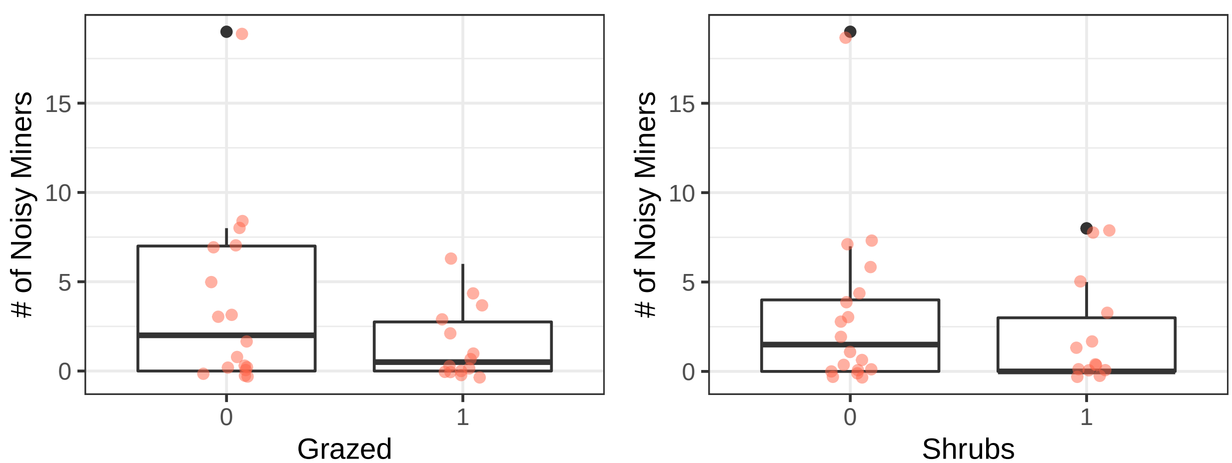




It appears that sites with miners are less likely to be grazed or have shrubs than sites without miners.



Here, we see a positive relationship between number of eucalyptus trees and number of noisy miner birds and potentially a negative curved relationship between areas and number of noisy miner birds which could potentially indicate that a transformation may be necessary.

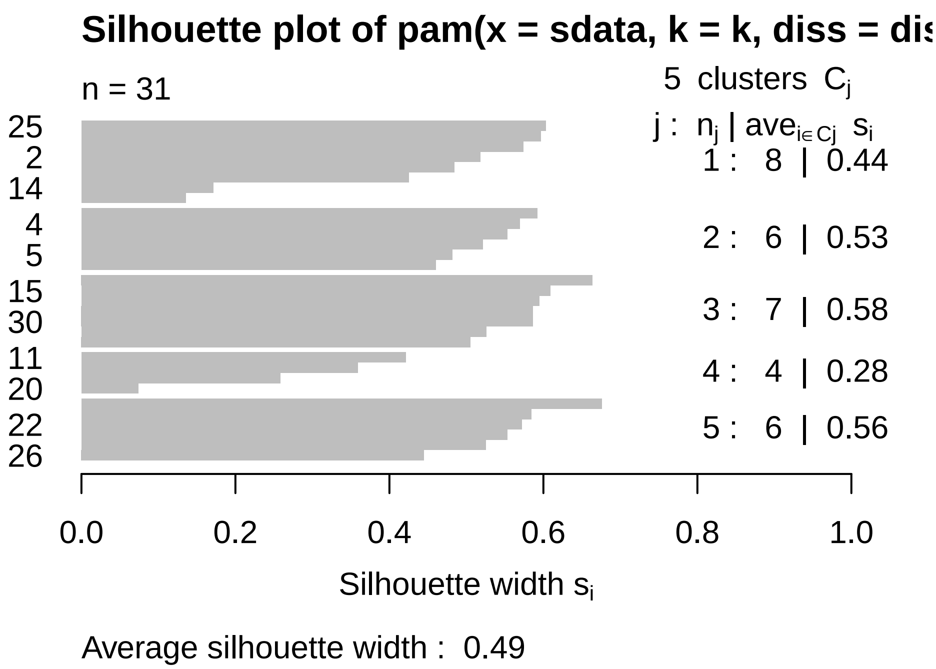


Sites that have been grazed appear to have a lower median and narrower distribution of number of noisy miners than those have not been grazed. Sites with shrubs have a lower median number of noisy miners, but its unclear whether noisy miner counts are significantly different between the two levels of this factor.

Clustering

In order to cluster the mixed types of data (both the explanatory variables and response), we first compute all the pairwise dissimilarities (using a generalization of Gower's formula) between observations in the data set and then perform a partitioning around medoids clustering with the number of clusters estimated by optimum average silhouette width.

A silhouette plot is a graphical tool that represents if the clustering configuration is appropriate. Values range from -1 to 1; a high value indicates that the object is cohesive with its own cluster and poorly matched to neighboring clusters. The number of clusters is estimated by optimum average silhouette width.



The optimal number of clusters is 5, 3 of the clusters include sites with noisy miners and 2 of them do not.

Table 1: Clusters with Noisy Miners

Variable	Cluster2	Cluster3	Cluster4
Miners	0:0, 1:6	0:0, 1:7	0:0, 1:4
Eucs	17.00	15.00	25.5
Area	19.00	16	10.00
Grazed	0:6, 1:0	0:0, 1:7	0:4, 1:0
Shrubs	0:0, 1:6	0:7, 1:0	0:4, 1:0
Bulokes	92.5	123.0	66.5
Timber	12.50	25.00	32
Minerab	4.00	3.0	7.00

Table 2: Clusters without Noisy Miners

Variable	Cluster1	Cluster5
Miners	0:8 , 1:0	0:6, 1: 0
Eucs	5.5	1.0
Area	18.50	12.50
Grazed	0:7, 1:1	0:0, 1:6
Shrubs	0:1, 1:7	0:6, 1:0
Bulokes	66.5	191.0
Timber	118.50	41.50
Minerab	0	0

Unsurprisingly, as the median number of Eucs increases across the three groups where noisy miners are present, the median number of noisy miners increase and the number of Bulokes decreases inversely. We also find that groups with less Eucs and more Bulokes are more likely to be grazed and than those with less Eucs and more Bulokes.

Within the groups in which noisy miners are absent, one cluster has mostly sites that are not grazed with shrubs , while the other has mostly sites that are grazed without shrubs. Interestingly, the latter group has a considerable lower value for Eucs and higher value for Bulokes. This supports the finding among clusters with noisy miners that grazed areas tend to have less Eucs and more Bulokes than non-grazed areas.

Bird Assemblage

$H_0 : p_{present} - p_{absent} = 0$

$H_0 : p_{present} - p_{absent} \neq 0$

where $p_{present}$ represents the probability of inhabiting in a site with noisy miners present and p_{absent} respresents the probability of inhabiting a site with noisy miners absent

In our analysis (See Appendix 7), we rejected the null hypothesis ($p < 0.005$). There is sufficient evidence that the relative frequency of species significantly varies between sites with and without noisy miners present. Therefore we can conclude that noisy miners make native species less likely to inhabit an ecosystem. We cannot directly compare this result to the paper's findings because they had access to a lot more site-level information, but they found that the total abundance of birds did not differ significantly between sites with and without noisy miners but species richness did.

Binomial GLM

First, we considered the *Miners* variable that represents whether or not noisy miners are present (1 or 0).

Examining the full model, the table below shows the estimates and standard errors.

Table 3: Coefficients of Full Model

	exp(Estimate)	Estimate	SE
(Intercept)	2.954500e-02	-3.5218424	5.0214679
Eucs	5.072093e+00	1.6237535	0.9079007
Area	9.171657e-01	-0.0864671	0.1109307
Grazed1	1.123565e+06	13.9320175	8.2096853
Shrubs1	8.441000e-04	-7.0771904	5.4016515
Bulokes	9.924435e-01	-0.0075852	0.0159880
Timber	5.370502e-01	-0.6216637	0.3869457

In order to analyze the percentage of variation explained by the current model, the researchers chose to use the Nagelkerke R-Squared value. This value is an adjusted version of the Cox and Snell R-squared value. It is adjusted by dividing by the maximum value in order to keep the R-Squared value between 0 and 1.

The Cox-Snell R^2 formula is:

$R^2 = 1 - [\frac{L(0)}{L(\hat{\beta})}]^{2/n}$ where $L(0)$ is the likelihood of the null model and $L(\hat{\beta})$ is the likelihood of the full model

The formula for the Nagelkerke R^2 is:

$= \frac{R^2}{max(R^2)}$ where $max(R^2) = 1 - exp(2n^{-1}l(0)) = L(0)^{2/n}$ where $l(0)$ is the log likelihood of the null model

Although this is a pseudo R-squared estimate it can still be interpreted as the approximate percentage of total variation explained by the model because its values are between 0 and 1.

Therefore, approximately 89% of the variation in the miners variable can be described by the full model including all variables.

In order to extend this analysis further, it would also be interested to compare this value to our 'pseudo' R-squared estimate explored in class defined as:

$$R^2_{pseudo} = 1 - \frac{D(Y,\hat{\mu})}{D(Y,\bar{Y})}$$

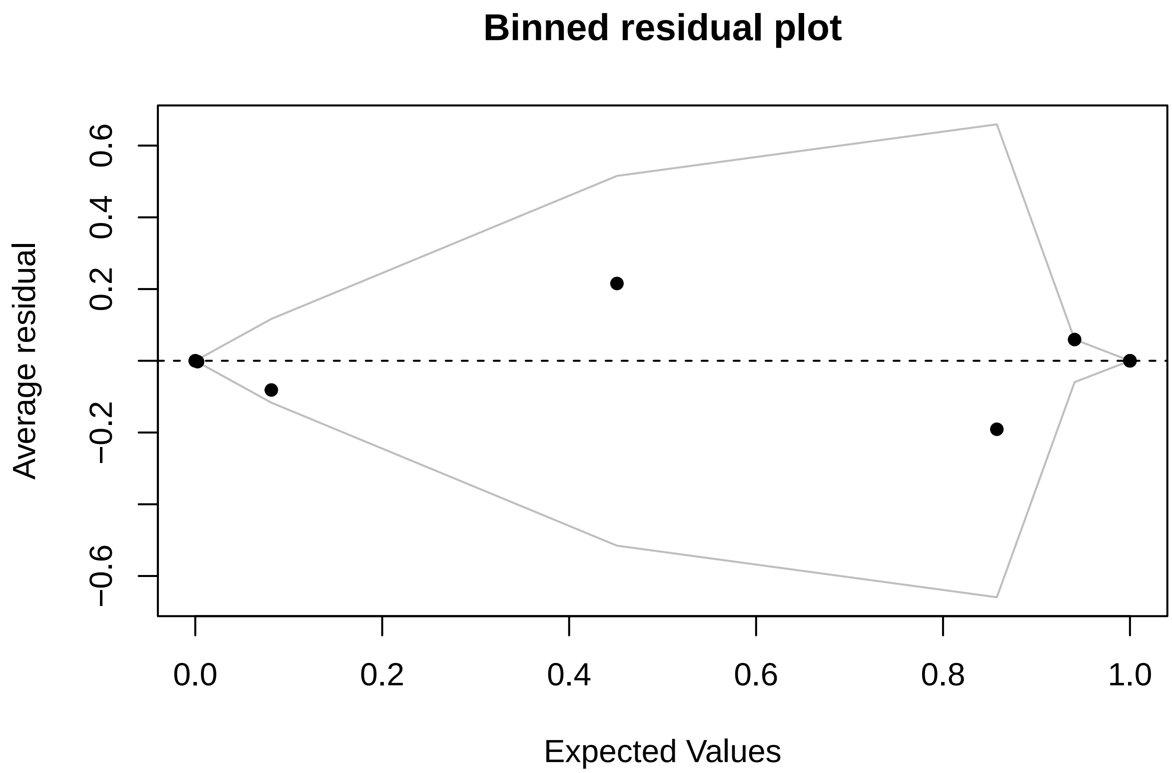
where $D(Y, \hat{\mu}$ is the deviance of our model and $D(Y, \bar{Y})$ is the deviance for the null model.

Using this method, the pseudo R-squared value is 0.81. Because this is a pseudo value we cannot interpret it in terms of percentage of variation explained.

Additionally, using this model the error rate is 6% using a threshold of 0.5. Therefore, using this model, 94% of transects were correctly predicted. Alternatively, for the null model, the null error rate is 45%. Meaning that the model is strong at predicting values compared to just the null model.

Lastly, the model has a dispersion parameter of 0.37 (<1) so it is not overdispersed and therefore we do not need to explore quasi models.

The figure below shows the binned residual plot for the full model. The binned residual plot for the full model shows that all but two points are within the range of 2 standard errors. Additionally, there is no major pattern for the residuals other than a slight bunching around the average. Examining this plot, there are no major concerns about the residuals.



The paper analyzed a model just including the Eucs variable, so we felt it would be important to explore the value of this model to determine a threshold effect. The table below shows the coefficients and standard errors for the model.

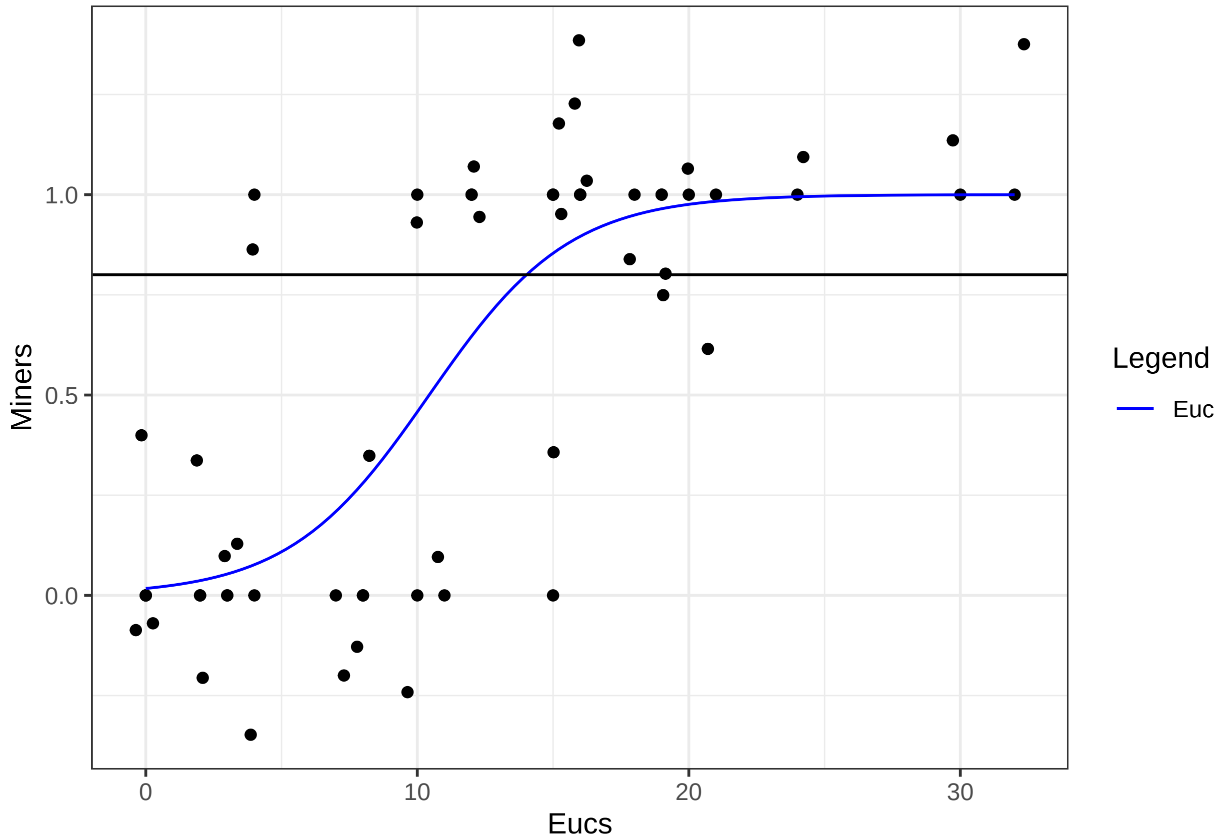
The coefficents for the model using just eucalpyt number can be interpreted as

- Intercept: For a plot with zero eucalyptus trees, the odds of a noisy miner being present is 0.017.
- Eucalypts: For each additional Eucalyptus tree in the plot of land, the odds of a noisy miner being present increases by a factor of 1.47.

Table 4: Coefficients of Full Model

	exp(Estimate)	Estimate	SE	P-Value
(Intercept)	0.0176391	-4.037636	1.5433366	0.0088921
Eucs	1.4723488	0.386859	0.1326968	0.0035528

The plot below shows the plot of the probability of presence of noisy miners with number of Eucalypt trees. The plot shows that at a threshold of 14 trees there is approximately a probability of 0.79 of the presence of Noisy Miners and at 15 trees there is an approximate probability of 0.85. Because our threshold is 0.8 we can conclude that the threshold is 15 Eucalypt trees.



From the model only including eucalyptus trees, the Nagelkerke R-squared value shows that approximately 71% of variation in miner presence can be predicted with the eucalpyt only model. Using the Pseudo R-squared estimate method, the R-squared value is only 0.55.

Additionally, using this model the error rate is 13% using a threshold of 0.5. Therefore, using this model, 87% of transects were correctly predicted. Alternatively, for the null model, the null error rate is 45%. Meaning that the model is strong at predicting values compared to just the null model.

These findings mirror the findings of the paper as they also found a Nagelkerke R-Squared value of 0.71 and a 13% error rate.

Backwards AIC stepwise selection recommends that the appropriate predictors to include in the model are Euc, Grazed, Shrubs, and Timber. Additionally, using a likelihood ratio test, we tested whether any variables should be dropped from the full model. The likelihood ratio test recommended that in addition to the recommendations of the backwards selection model, Shrubs should also be dropped from the model due to a p-value of 0.075 (>0.05). Both of these tests can be found in Appendix 5.

Our variable selection results were different than the variable selection results in the paper due to the difference in variables. Some variables were binary in the data set made available to us and continuous in the original paper's data set. This may have led to differences in final model selection.

The table below shows the exponentiated estimates, estimates and standard errors for the final model recommended by the likelihood ratio tests. Examining this model, none of the predictors are significant in the final model. Additionally, similar to the full model it is also under-dispersed with a dispersion parameter of 0.50.

Table 5: Coefficients of Full Model

	exp(Estimate)	Estimate	SE	P-Value
(Intercept)	0.0017558	-6.3448218	3.8788237	0.1018900
Eucs	2.1273579	0.7548808	0.3902220	0.0530528
Grazed1	865.9839719	6.7638664	4.4999227	0.1328114
Timber	0.8177917	-0.2011477	0.1266389	0.1122058

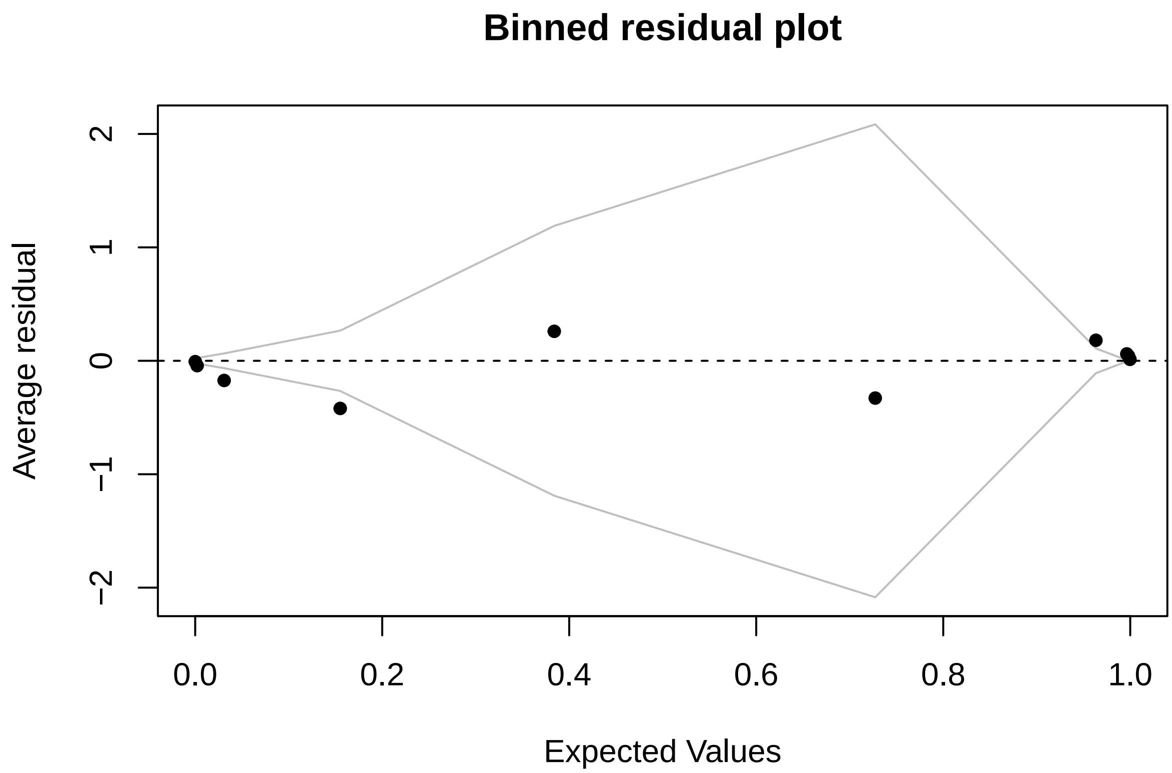
We can interpret the coefficients for the model as:

- Intercept: For a plot of land with no timber, not grazed, and zero eucalpytus trees, the odds of a noisy miner being present is 1.76e-03.
- Eucs: Holding amount of Timber and Grazed status constant, for each additional eucalyptus tree on the plot of land there is an increase in the odds of a noisy miner being present by a factor of 2.12.
- Timber: Holding number of eucalpytus trees and whether the land was grazed constant, on a plot of land for each additional piece of timber on the ground, there is an increase in the odds of a noisy miner being present by a factor of 0.82.
- Grazed: Holding number of eucalpytus trees and timber constant, for a plot of land that is grazed, there is a 866-factor increase in the odds of noisy miners being present compared to a plot of land that is not grazed.

Using this model and the Nagelkerke R-Squared estimate, approximately 84% of variation can be predicted. Alternatively, using the Pseudo R-squared method, the r-squared value is 0.71.

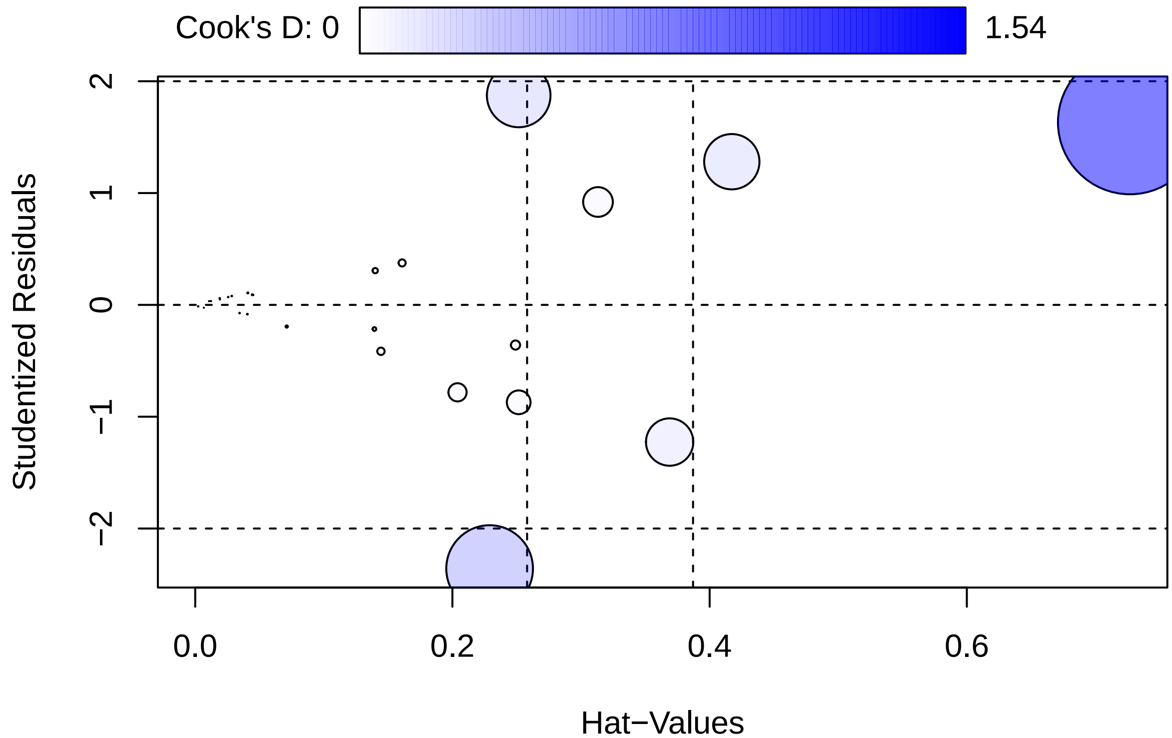
Additionally, using this model, the model correctly predicted the presence of miners 93% of the time (using a threshold of 0.5) compared to the null error rate of 45%.

The chart below shows the binned residual plot for the final model. Similar to the residual plot for the full model, there are no concerns raised when examining this residual plot. Almost all of the points are within 2 standard errors and show little patterns other than slight bunching around the average.



Next will look into the influence plot of the model to identify if there are any influential points.

Exploring the points identified on the influence plot, points 18, 29, 7, and 28 are of higher influence.

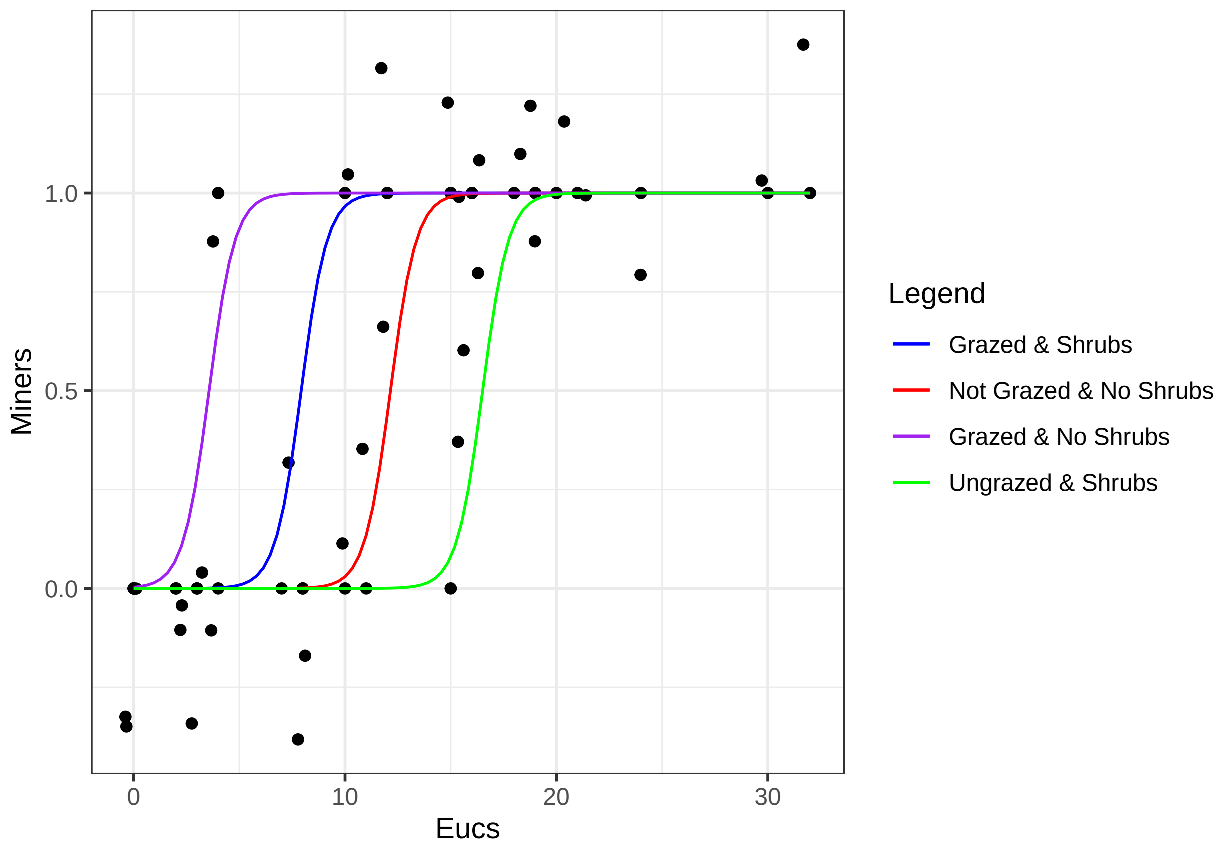


Examining the impact of each case on each of the coefficient values, the 28th datapoint has a considerable influence on the Eucs and Timber coefficients, decreasing the eucs coefficeint by 1.08 and increasing the timber coefficient by 1.3. Additionally, the 29th data point changes the Timber coefficient by 2.6 and the grazed coefficient by -2.098670. This is likely due to an above average level of timber value. See Appendix 5 for further information on influence of each plot.

Additionally, exploring the Cook's distance plot of the final model, point 29 has the largest Cook's distance value. Upon further examination of this plot does not cause any concern as being an outlier.

Although these points prove to be technically influential in the model they do not provide any reason for removal or concern.

Laslty, below is a plot demonstrating how the probability of a Noisy Miner being present changes with number of Eucalpytus trees for the full model.



Poisson Regression

We began by building a full model using all the data available by including all the variables in the model. It is important to note that the assumed mean variance for a Poisson distribution is that the variance is equal to the mean, so we assume $\phi = 1$. However, the full model has residual deviance of 54.254 on 24 degrees of freedom indicating overdispersion ($\phi > 1$). The Pearson estimate of ϕ results in $\hat{\phi} = 2.4202$, confirming that the variance of the response appears to exceed what we expect. When $\phi > 1$, we underestimate our standard errors, so coefficients may look significant when they are indeed not. In order to address the overdispersion, we tried fitting a quasi-posson model to retain some parts of the Poisson model without involving the full likelihood. This allowed us vary the usual variance function, by assuming a value for the dispersion ϕ greater than one. For the quasi poison model with all of the explanatory variables, the only coefficent deemed significant is Eucs.

	Est.Pois	Est.Quasi	SE.Pois	SE.Quasi	ratio
(Intercept)	-0.8863455	-0.8863455	0.8757365	1.3623585	1.555672
Eucs	0.1293090	0.1293090	0.0217570	0.0338467	1.555672
Grazed1	0.1408309	0.1408309	0.3646217	0.5672317	1.555672
Shrubs1	0.3358275	0.3358275	0.3750588	0.5834683	1.555672
Bulokes	0.0014688	0.0014688	0.0017735	0.0027589	1.555672
Timber	-0.0067814	-0.0067814	0.0090738	0.0141159	1.555672
Area	-0.0287357	-0.0287357	0.0132406	0.0205980	1.555672

Table 6 illustrates that coefficent estimates are the same for the GLM and Quasi Poisson model, but the standard errors for each variable increase when we relax the assumption about the mean-variance relationship. This explains how Area was significant in the Poisson model, but is no longer significant in the Quasi-Poisson model.

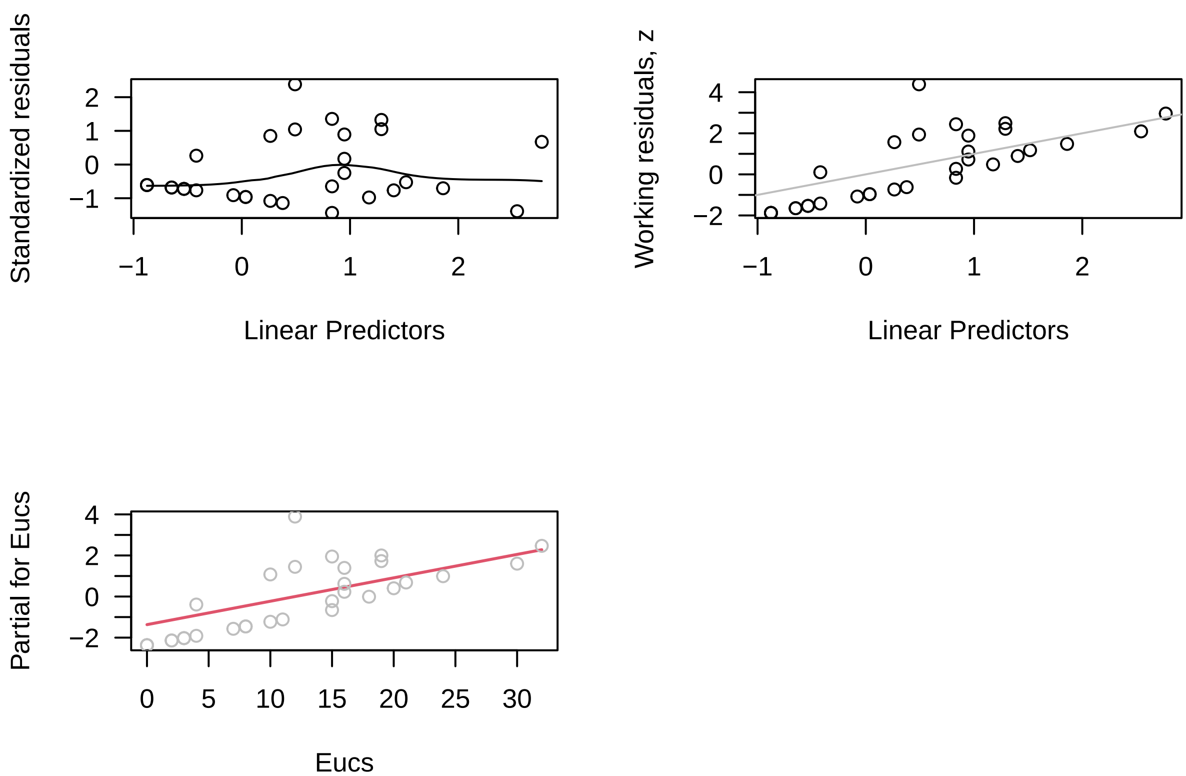
We suspect that not all of the variables are important or significant explanatory variables so we want to do variable selection to create a more parsimonious model. The quasi poisson behaves like a log-likelihood function, but does not correspond to any probability function. As a result, the AIC and related statistics are not defined for quasi-models. Analysis of deviance tests are based on the F-tests since ϕ is estimated for the quasi-models. The F-test comparing the models with single term deletions indicates that when dropping single terms from the model, the only model that is significantly different from the current model is the model in which Eucs is dropped.

Since we do not have evidence that any of the other variables have a significant effect on the number of noisy miners we can drop them from the model. The final quasi-poisson model is below.

Table 7: Coefficients of Quasi Poisson Model

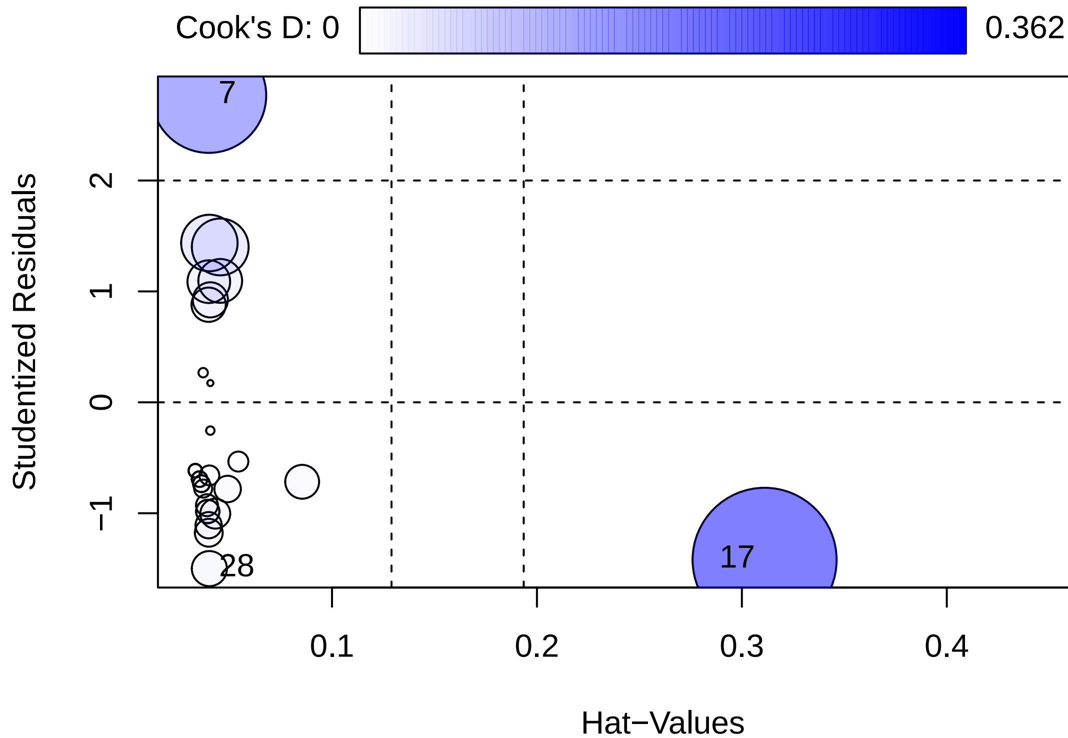
	exp(Estimate)	Estimate	SE	Wald Coef
(Intercept)	0.4163573	-0.8762114	0.4314856	0.0515452
Eucs	1.1207312	0.1139813	0.0189673	0.0000016

On average, the expected number of noisy miners in a transect with 0 eucalypt trees is 0.416. Each additional tree in a transect is associated with a 12.07% increase in the expected number of noisy miners.



The diagnostic plots look good, the model appears to be a good fit for the data and there are no clear patterns in any of the residual plots. The working residuals look roughly linear and the standard deviance residuals don't look completely random, but there are not any strong distinguishable patterns that raise major concerns.

We also decided to identify outliers and influential points to determine whether or not the points actually have an inordinate influence on our model.




```
##       StudRes        Hat     CookD
## 7   2.7671345 0.03987337 0.2302135
## 11  0.6992208 0.48961774 0.2307877
## 17 -1.4204496 0.31108992 0.3621766
## 28 -1.4994110 0.04014949 0.0215395
```

Observations 7, 11 and 17 have relatively high values for Cook's Distance, indicating that they have high influence on the fitted response values. We looked at these influential points to assess whether there is anything unusual about them that may be introducing bias to our model.

```
##    Miners Eucs Area Grazed Shrubs Bulokes Timber Minerab
## 7       1   12   16      0      1     100     12       8
## 11      1   32    3      0      0      23     61      19
## 17      1   30   18      0      0      45     19       7
```

Observation 11 is an outlier because it has an unusually high number of noisy miners, with a value (18) more than twice as large as the next highest value in the distribution (9). Both of these observations have unusually high number of eucalyptus trees. The range of this variable is 32, and both of these values are at least 6 a greater than the next highest Eucs value. Observation 7 has a high number of noisy miners present, despite only having the median number of eucalypts, leading to a large residual. We have no information to lead us to believe these observations are impossible or data entry errors so we will keep them in the analysis.

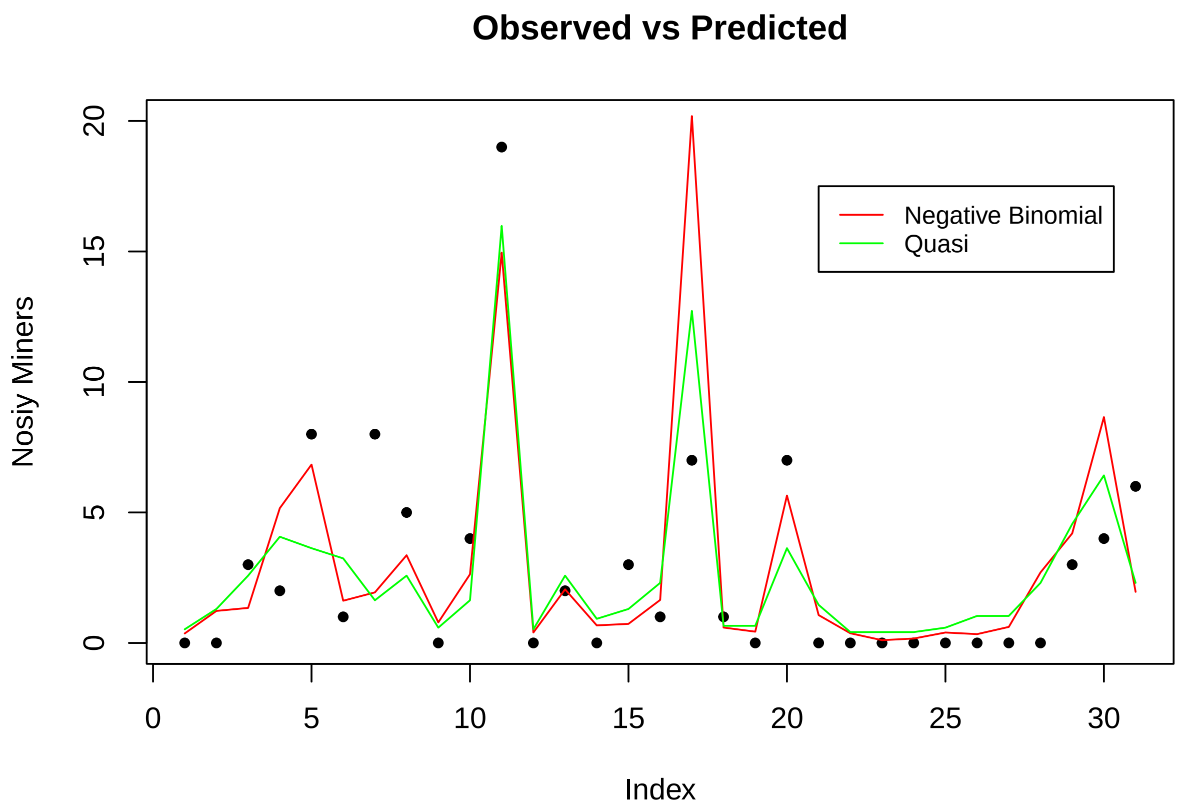
Negative Binomial Regression

Another option when a Poisson distribution is overdispersed is to fit a negative binomial that starts with Poisson regression model and add a multiplicative random effect θ to represent the unobserved heterogeneity with a gamma distribution.

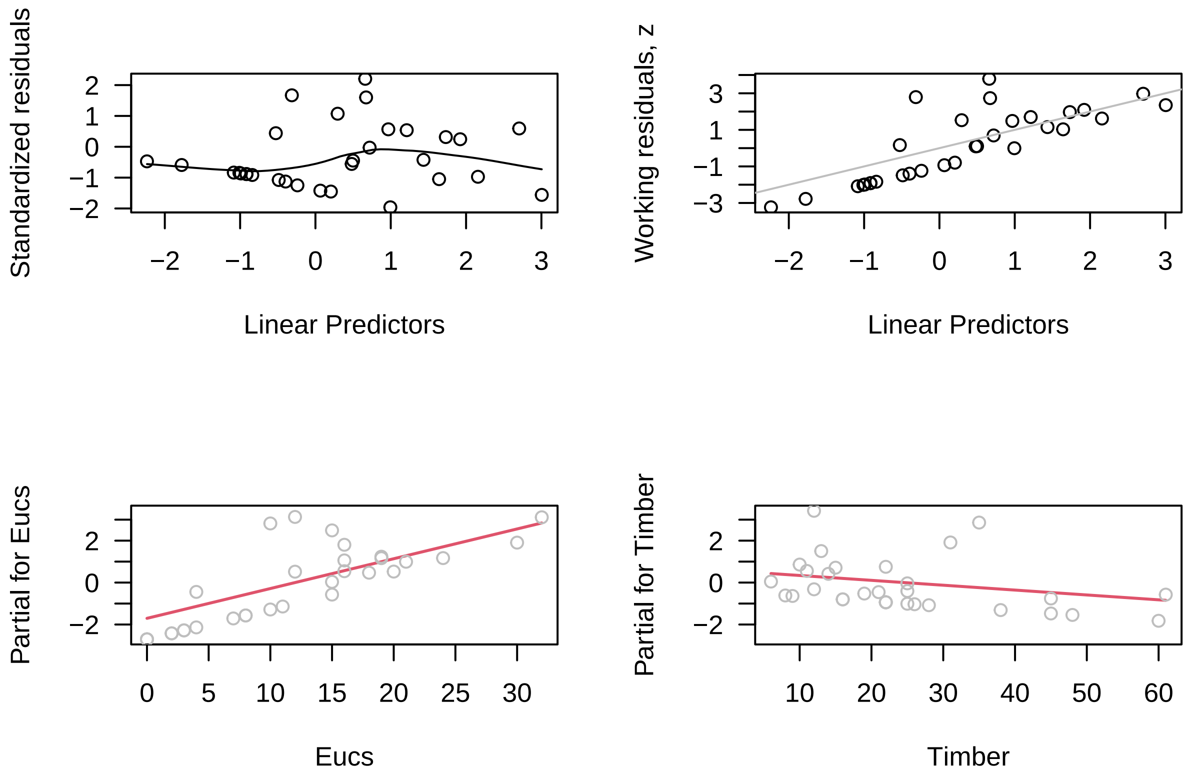
We built a model with all of the explanatory variables and Eucs is the only coefficient that is deemed significant. Using a backwards stepwise algorithm by AIC, the final model includes Timber, Area, and Eucs.

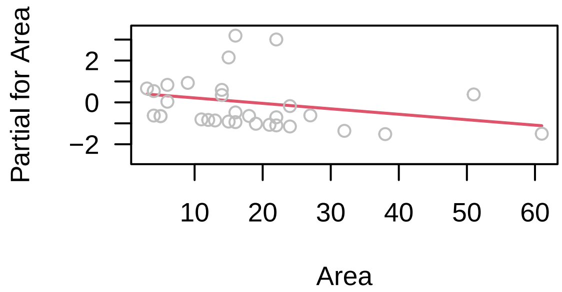
Table 8: Coefficients of Negative Binomial Model

	exp(Estimate)	Estimate	SE	Wald Coef
(Intercept)	0.7049263	-0.3496620	0.6038619	0.0515452
Eucs	1.1527421	0.1421435	0.0240935	0.0000016
Timber	0.9770593	-0.0232079	0.0143010	0.0515452
Area	0.9742986	-0.0260375	0.0157337	0.0000016



As we can see in the chart above, the predicted values for both models are fairly similar, but we do some considerable discrepancies for certain observations. The negative binomial model seems to slightly better fit the observed counts and even if you add all the explanatory variables from the final negative binomial model to the Quasi Poisson model, the negative binomial still has lower residual deviance, so it appears to better capture the variation in the response.


Standardized residuals
Linear Predictors
Working residuals, z
Linear Predictors
Partial for Eucs
Eucs
Partial for Timber
Timber

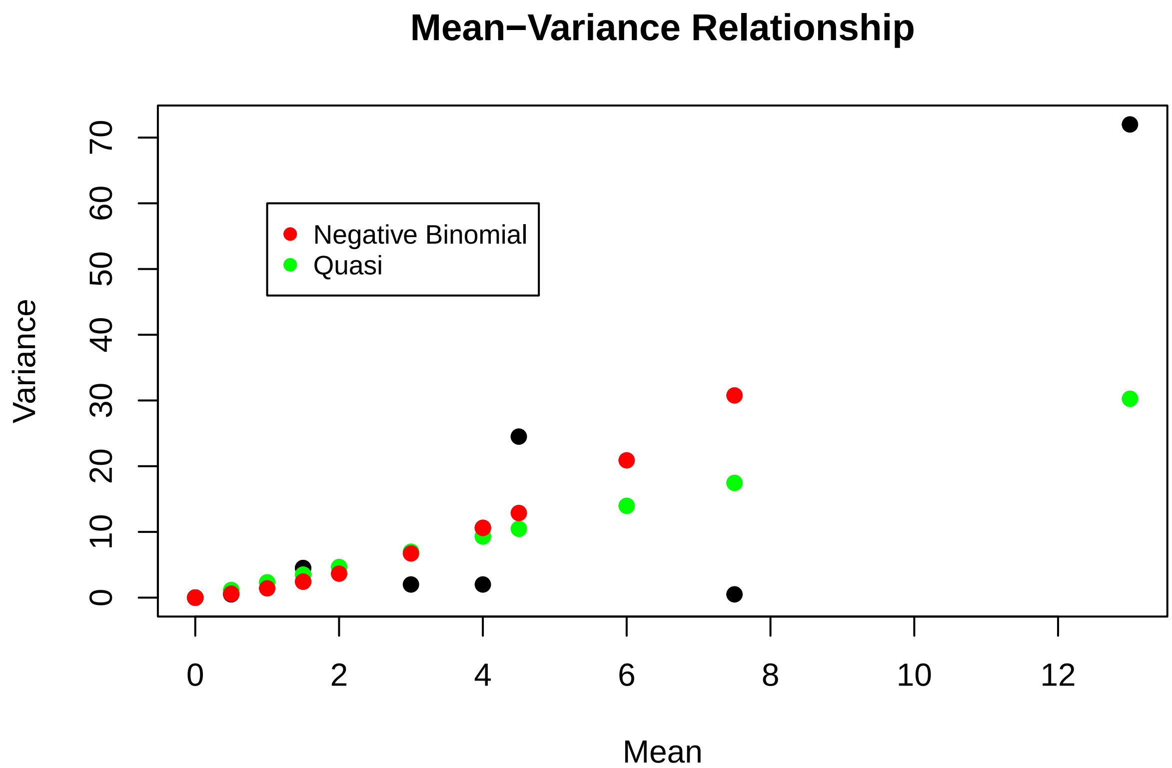


The residual diagnostic plots look okay. There are no concerning patterns or major violations of the assumptions.

	Est.NB	Est.Quasi	SE.NB	SE.Quasi
(Intercept)	-0.350	-0.876211420742833	0.604	0.431485613298561
Eucs	0.142	0.113981349205161	0.024	0.0189672874355309
Timber	-0.023	-	0.014	-
Area	-0.026	-	0.016	-

In comparing the two models we see similar coefficients and standard errors for eucalypts. In both models, the Wald Test indicates that Eucs is the only significant explanatory variable. We include the coefficients of timber and area in the negative binomial model but these effects are not large in comparison to Eucs.

The quasi Poisson and negative binomial models have different variance functions so we can compute the mean and variance of the response for each group to assess which regression captures the mean-variance relationship more accurately.



This plot would be more informative if we had more observations, but it seems that the negative binomial distribution better captures the observed mean-variance relationship for the number of noisy miners.

4. Discussion

Confirmed Paper Conclusions

Through our analysis we were able to confirm two major conclusions from the paper.

1. "The results demonstrate that small differences in floristics can result in large differences in the local avifauna by providing habitat for undesirable species"

Through our analysis using a paired-sample sign test we found that there was sufficent evidence to reject the null hypothesis and conclude that the relative frequency of species significantly varies between sites with and without noisy miners present

2. "Eucalypt density alone was a predictor of noisy miner presence in buloke woodland, with a clear threshold effect apparent at relatively low eucalypt densities. A density of 15 eucalypts in a two-hectare transect resulted in a 0.8 probability of noisy miner presence, but miners were recorded on only one transect with fewer than 10 eucalypts"

Through the use of our binomial models and threshold analysis we were able to confirm the conclusion that the threshold of 15 eucalypts led to above a 0.8 probability of noisy miners being present.

New Conclusions

In addition to confirming the main conclusions of Maron's paper, we were also able to further extend our analysis to include a clustering model and examine the abundance of noisy miners in each woodland patch. Through our analysis of a quasi binomial model we found:

- Eucalypt density, number of pieces of fallen timber, and area were significant predictors of Noisy Miner abundance in Buloke woodlands
- Grazed areas tend to have less Eucalypts and more Bulokes than non-grazed areas
- Evidence that the number of birds for each species varies between sites with and without noisy miners present

Further Extensions

Given that our data set included many counts of 0 for the Minerab variable it would be interesting to apply the Zero-Inflated Poisson model or the Zero-Inflated Negative Binomial Model to our dataset to see how our findings might change.

Applications

By understanding which site-level factors influence the presence of Noisy Miners, we can help conservationists more successfully revegetate Buloke Woodlands.
Given that research has found that the presence of Noisy Miners negatively effects the diversity of bird species within a woodland, identifying the factors that discourage Noisy Miners is important. Our results show that when revegetating areas, conservationists should keep the number of Eucalypts within a 2 hectare patch of vegetation below 15. However for land that is grazed that threshold is much lower. Ultimately, these results will be integral to making decisions about conservation practices to best conserve woodland bird species.

References

Gower, J. C. (1971), “A general coefficient of similarity and some of its properties”. Biometrics, 27, 623–637.

Kaufman, L. and Rousseeuw, P.J. (1990), Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.

Maron, Martine. “Threshold Effect of Eucalypt Density on an Aggressive Avian Competitor.” Biological Conservation, vol. 136, no. 1, Nov. 2007, pp. 100–107., https://doi.org/10.1016/j.biocon.2006.11.007.

Appendix 1

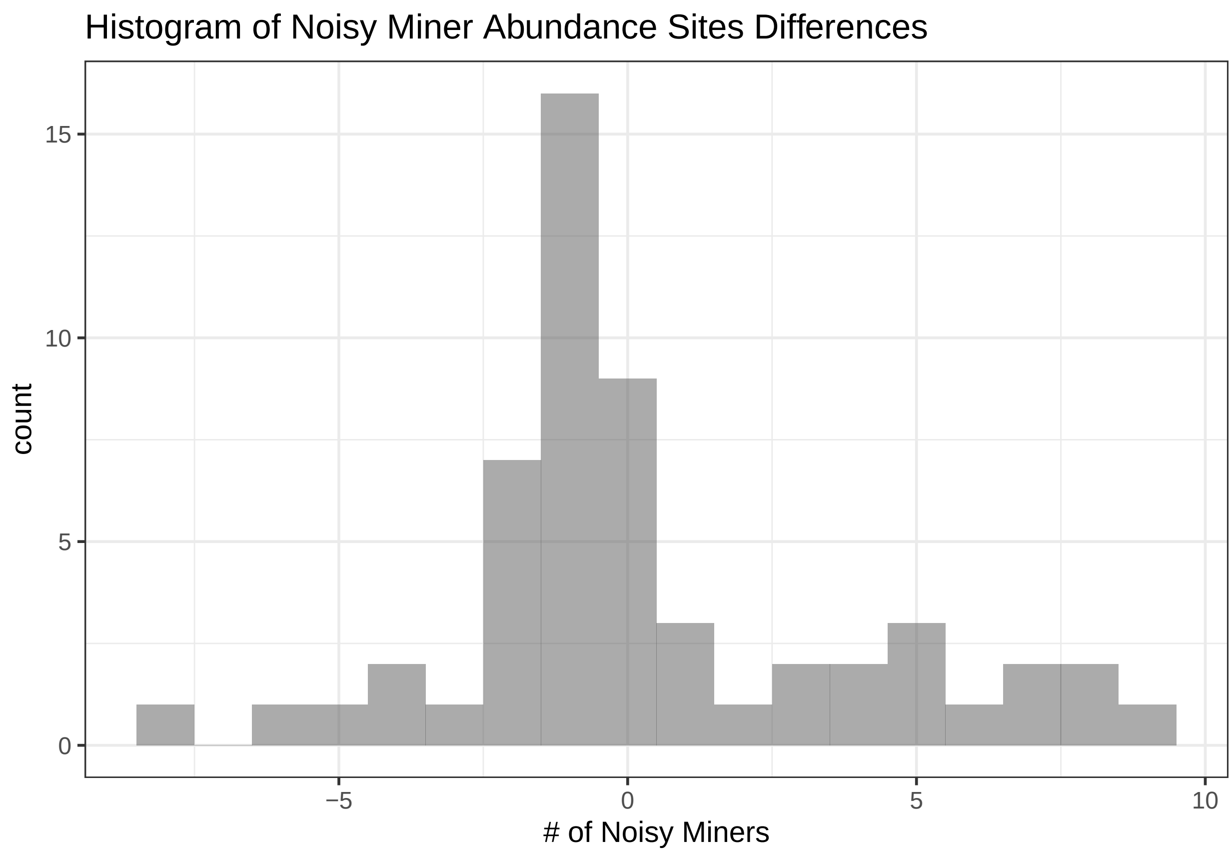
```
species_df
```

```
##                          Species Rminer_Present Rminer_Absent Declining_Woodland
## 1        Australian wood duck              1             2                  0
## 2       Black-shouldered kite              0             1                  0
## 3                Brown falcon              2             2                  0
## 4             Nankeen kestrel              1             1                  0
## 5            Australian hobby              0             1                  0
## 6              Crested pigeon             11            12                  0
## 7                       Galah              6            12                  0
## 8          Long-billed corella              0             1                  0
## 9                   Cockatiel              0             2                  0
## 10     Purple-crowned lorikeet              1             3                  0
## 11               Musk lorikeet              0             2                  0
## 12             Eastern rosella              5            13                  0
## 13                 Blue bonnet              2             3                  0
## 14            Red-rumped parrot             12            15                  0
## 15 Horsfield's bronze-cuckoo               0             1                  0
## 16             Tawny frogmouth              0             1                  0
## 17          Laughing kookaburra              1             2                  0
## 18           Brown treecreeper              8            10                  1
## 19       Variegated fairy-wren              1             1                  0
## 20           Striated pardalote              5            10                  0
## 21 Chestnut-rumped thornbill               6             1                  1
## 22     Yellow-rumped thornbill             11             5                  0
## 23             Yellow thornbill              8             0                  0
## 24                    Weebill               2             0                  0
## 25          Southern whiteface              7             0                  1
## 26               Red wattlebird              1             1                  0
## 27    Spiny-cheeked honeyeater              1             2                  0
## 28       Yellow-throated miner              0             1                  0
## 29          Singing honeyeater              5             0                  0
## 30     White-plumed honeyeater              8            10                  0
## 31     Brown-headed honeyeater              1             1                  0
## 32                Jacky winter              5             1                  1
## 33             Red-capped robin              7             0                  1
## 34                Hooded robin             11             3                  1
## 35             Varied sittella              9             0                  1
## 36             Rufous whistler              4             0                  1
## 37                 Magpie-lark              0             2                  0
## 38              Willie wagtail             10             7                  0
## 39                Grey fantail              5             0                  0
## 40 Black-faced cuckoo-shrike               1             1                  0
## 41    White-browed woodswallow              1             1                  1
## 42     Black-faced woodswallow              0             1                  0
## 43           Dusky woodswallow              6             3                  1
## 44            Pied butcherbird              0             1                  0
## 45           Australian magpie             10            14                  0
## 46            Australian raven              1             2                  0
## 47                Little raven              1             0                  0
```

```
## 48       White-winged chough              0             2                   0
## 49        European goldfinch              0             1                   0
## 50             House sparrow             10             9                   0
## 51             Mistletoebird              1             1                   0
## 52           Welcome swallow              2             3                   0
## 53               Tree martin              2             2                   0
## 54           Rufous songlark              1             0                   0
## 55           Common starling              6            10                   0
##    Difference
## 1          -1
## 2          -1
## 3           0
## 4           0
## 5          -1
## 6          -1
## 7          -6
## 8          -1
## 9          -2
## 10         -2
## 11         -2
## 12         -8
## 13         -1
## 14         -3
## 15         -1
## 16         -1
## 17         -1
## 18         -2
## 19          0
## 20         -5
## 21          5
## 22          6
## 23          8
## 24          2
## 25          7
## 26          0
## 27         -1
## 28         -1
## 29          5
## 30         -2
## 31          0
## 32          4
## 33          7
## 34          8
## 35          9
## 36          4
## 37         -2
## 38          3
## 39          5
## 40          0
## 41          0
## 42         -1
## 43          3
## 44         -1
## 45         -4
```

```
## 46         -1
## 47          1
## 48         -2
## 49         -1
## 50          1
## 51          0
## 52         -1
## 53          0
## 54          1
## 55         -4
```

```
ggplot(species_df, aes(x=Difference)) +
  geom_histogram(alpha=0.5, position="identity", binwidth = 1) + ggtitle("Histogram of Noisy Miner Abund
```



As shown in the plot above, the distribution of the differences is not normal so a paired t-test would not be appropriate.

Appendix 2 : Noisy Miner Count Modelling

Poisson Full Model:

```
##
## Call:
## glm(formula = Minerab ~ Eucs + Grazed + Shrubs + Bulokes + Timber +
```

```
##     Area, family = poisson, data = nminer)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2826  -1.1220  -0.8011   0.4159   3.3511
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.886345   0.875737  -1.012    0.311
## Eucs         0.129309   0.021757   5.943 2.79e-09 ***
## Grazed1      0.140831   0.364622   0.386    0.699
## Shrubs1      0.335828   0.375059   0.895    0.371
## Bulokes      0.001469   0.001773   0.828    0.408
## Timber      -0.006781   0.009074  -0.747    0.455
## Area        -0.028736   0.013241  -2.170    0.030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 150.545  on 30  degrees of freedom
## Residual deviance:  54.254  on 24  degrees of freedom
## AIC: 122.41
##
## Number of Fisher Scoring iterations: 6
```

Pearson estimate of ϕ:

```r
sum(residuals(poisson_mod1, type = "pearson")^2)/24
```

```
## [1] 2.420114
```

Full Quasi Poisson Model:

```
##
## Call:
## glm(formula = Minerab ~ Eucs + Grazed + Shrubs + Bulokes + Timber +
##     Area, family = "quasipoisson", data = nminer)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2826  -1.1220  -0.8011   0.4159   3.3511
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.886345   1.362358  -0.651 0.521487
## Eucs         0.129309   0.033847   3.820 0.000828 ***
## Grazed1      0.140831   0.567232   0.248 0.806032
## Shrubs1      0.335828   0.583468   0.576 0.570263
## Bulokes      0.001469   0.002759   0.532 0.599343
## Timber      -0.006781   0.014116  -0.480 0.635284
## Area        -0.028736   0.020598  -1.395 0.175771
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 2.420114)
##
##     Null deviance: 150.545  on 30  degrees of freedom
## Residual deviance:  54.254  on 24  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

Quasi Poisson Model Variable Selection:

```
## Single term deletions
##
## Model:
## Minerab ~ Eucs + Grazed + Shrubs + Bulokes + Timber + Area
##         Df Deviance F value    Pr(>F)
## <none>       54.254
## Eucs     1   95.513 18.2515 0.0002643 ***
## Grazed   1   54.403  0.0659 0.7995968
## Shrubs   1   55.071  0.3614 0.5533751
## Bulokes  1   54.920  0.2945 0.5923648
## Timber   1   54.818  0.2493 0.6221143
## Area     1   59.765  2.4378 0.1315333
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Single term deletions
##
## Model:
## Minerab ~ Eucs + Shrubs + Bulokes + Timber + Area
##         Df Deviance F value    Pr(>F)
## <none>       54.403
## Eucs     1  112.615 26.7501 2.389e-05 ***
## Shrubs   1   55.128  0.3331    0.5690
## Bulokes  1   55.073  0.3077    0.5840
## Timber   1   55.017  0.2818    0.6002
## Area     1   59.785  2.4732    0.1284
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Single term deletions
##
## Model:
## Minerab ~ Eucs + Shrubs + Timber + Area
##        Df Deviance F value    Pr(>F)
## <none>      55.073
## Eucs    1  133.115 36.8437 2.056e-06 ***
## Shrubs  1   55.574  0.2366   0.63077
## Timber  1   56.668  0.7531   0.39345
## Area    1   62.954  3.7208   0.06472 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Single term deletions
##
## Model:
## Minerab ~ Eucs + Timber + Area
##        Df Deviance F value    Pr(>F)
## <none>      55.574
## Eucs    1  133.516 37.8674 1.413e-06 ***
## Timber  1   58.081  1.2179   0.27952
## Area    1   63.281  3.7444   0.06353 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Single term deletions
##
## Model:
## Minerab ~ Eucs + Area
##        Df Deviance F value   Pr(>F)
## <none>      58.081
## Eucs    1  135.061 37.1111 1.43e-06 ***
## Area    1   63.318  2.5249   0.1233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Full negative binomial model:

```
nbGLM2 <- glm.nb(Minerab ~ Eucs + Grazed + Shrubs + Bulokes + Timber + Area, data=nminer)
summary(nbGLM2)
```

```
##
## Call:
## glm.nb(formula = Minerab ~ Eucs + Grazed + Shrubs + Bulokes +
##     Timber + Area, data = nminer, init.theta = 2.264283882, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8918  -0.9825  -0.5552   0.3783   2.1299
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.6936645  1.3471584  -0.515    0.607
## Eucs         0.1559905  0.0336134   4.641 3.47e-06 ***
## Grazed1      0.4152207  0.5923117   0.701    0.483
## Shrubs1      0.2382180  0.6203954   0.384    0.701
## Bulokes     -0.0003395  0.0028601  -0.119    0.905
## Timber      -0.0252807  0.0175026  -1.444    0.149
## Area        -0.0278414  0.0172383  -1.615    0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.2643) family taken to be 1)
##
##     Null deviance: 75.270  on 30  degrees of freedom
## Residual deviance: 30.012  on 24  degrees of freedom
```

```
## AIC: 118.38
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  2.26
##           Std. Err.:  1.32
## Warning while fitting theta: alternation limit reached
##
##  2 x log-likelihood:  -102.383
```

Negative Binomial Model Variable Selection:

```
step(nbGLM2)
```

```
## Start:  AIC=116.38
## Minerab ~ Eucs + Grazed + Shrubs + Bulokes + Timber + Area
##
##           Df Deviance    AIC
## - Bulokes  1   30.027 114.40
## - Shrubs   1   30.152 114.52
## - Grazed   1   30.446 114.82
## <none>         30.012 116.38
## - Timber   1   32.077 116.45
## - Area     1   32.834 117.20
## - Eucs     1   52.810 137.18
##
## Step:  AIC=114.4
## Minerab ~ Eucs + Grazed + Shrubs + Timber + Area
##
##          Df Deviance    AIC
## - Shrubs  1   30.377 112.57
## - Grazed  1   30.661 112.85
## <none>        30.209 114.40
## - Timber  1   32.627 114.82
## - Area    1   33.278 115.47
## - Eucs    1   56.851 139.04
##
## Step:  AIC=112.57
## Minerab ~ Eucs + Grazed + Timber + Area
##
##          Df Deviance    AIC
## - Grazed  1   30.563 110.86
## <none>        30.270 112.57
## - Area    1   33.171 113.47
## - Timber  1   33.315 113.61
## - Eucs    1   62.237 142.53
##
## Step:  AIC=110.85
## Minerab ~ Eucs + Timber + Area
##
##          Df Deviance    AIC
## <none>        31.205 110.85
```

```
## - Timber  1   34.001 111.65
## - Area    1   34.198 111.84
## - Eucs    1   70.665 148.31
```

```
##
## Call:  glm.nb(formula = Minerab ~ Eucs + Timber + Area, data = nminer,
##     init.theta = 2.417669611, link = log)
##
## Coefficients:
## (Intercept)         Eucs       Timber         Area
##    -0.34966      0.14214     -0.02321     -0.02604
##
## Degrees of Freedom: 30 Total (i.e. Null);  27 Residual
## Null Deviance:       77.33
## Residual Deviance: 31.21     AIC: 112.9
```

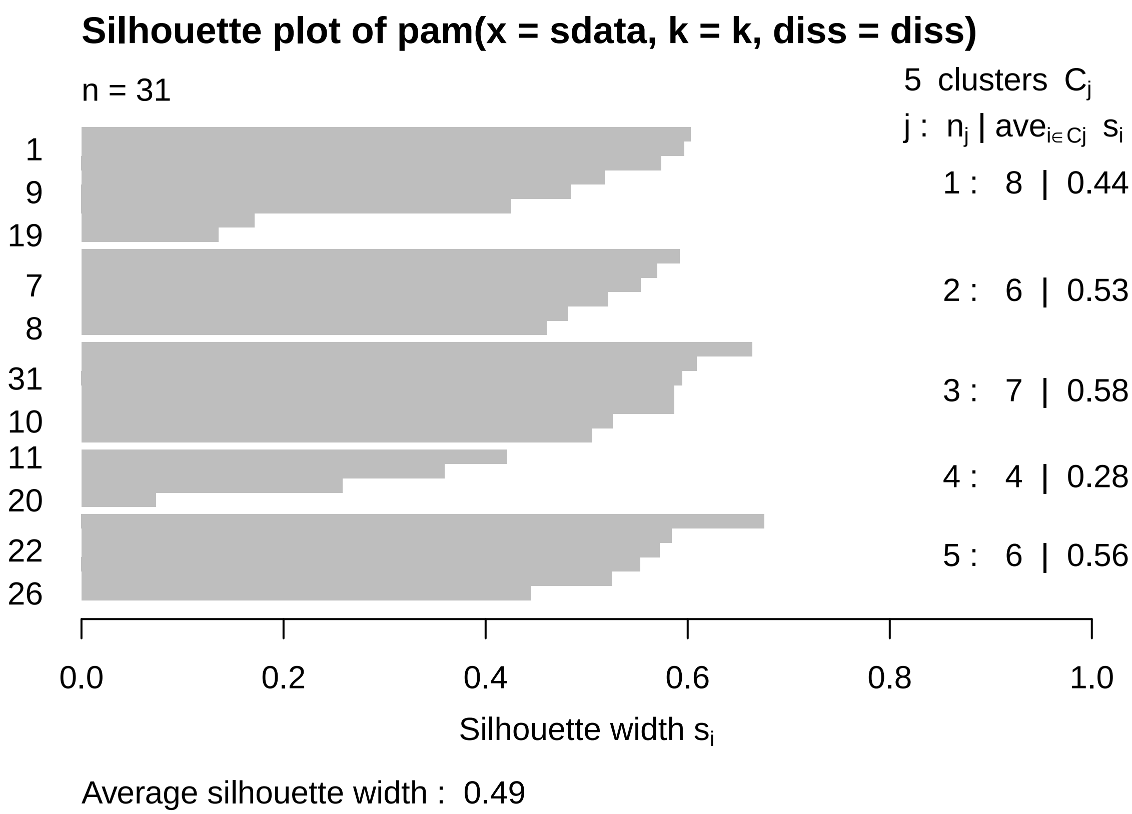
```
summary(final.NB)
```

```
##
## Call:
## glm.nb(formula = Minerab ~ Eucs + Timber + Area, data = nminer,
##     init.theta = 2.417669611, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9051  -0.9488  -0.5721   0.3270   2.1261
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.34966    0.60386  -0.579   0.5626
## Eucs         0.14214    0.02409   5.900 3.64e-09 ***
## Timber      -0.02321    0.01430  -1.623   0.1046
## Area        -0.02604    0.01573  -1.655   0.0979 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.4177) family taken to be 1)
##
##     Null deviance: 77.327  on 30  degrees of freedom
## Residual deviance: 31.205  on 27  degrees of freedom
## AIC: 112.85
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  2.42
##           Std. Err.:  1.50
##
##  2 x log-likelihood:  -102.851
```

Appendix 3: Clustering

```
g.dist = daisy(nminer, metric="gower")
pc = pamk(g.dist, criterion="asw")
pc_object = pc$pamobject

plot(pc_object)
```




```
cluster1 <- c()
cluster2 <- c()
cluster3 <- c()
cluster4 <- c()
cluster5 <- c()

for (i in 1: length(pc_object$clustering) ) {

    if (pc_object$clustering[i] == 1) {
      cluster1 <- c(cluster1, names(pc_object$clustering)[i])
  } else if (pc_object$clustering[i] ==2) {
    cluster2 <- c(cluster2, names(pc_object$clustering)[i])
  } else if (pc_object$clustering[i] ==3) {
    cluster3 <- c(cluster3, names(pc_object$clustering)[i])
  } else if (pc_object$clustering[i] ==4) {
    cluster4 <- c(cluster4, names(pc_object$clustering)[i])
```

```r
  } else {
    cluster5 <- c(cluster5, names(pc_object$clustering)[i])
  }

}

cluster1_df <- nminer[cluster1, ]
cluster2_df <- nminer[cluster2, ]
cluster3_df <- nminer[cluster3, ]
cluster4_df <- nminer[cluster4, ]
cluster5_df <- nminer[cluster5, ]
```

Appendix 5: Negative Binomial

```r
summary(final.NB)
```

```
##
## Call:
## glm.nb(formula = Minerab ~ Eucs + Timber + Area, data = nminer,
##     init.theta = 2.417669611, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9051  -0.9488  -0.5721   0.3270   2.1261
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.34966    0.60386  -0.579   0.5626
## Eucs         0.14214    0.02409   5.900 3.64e-09 ***
## Timber      -0.02321    0.01430  -1.623   0.1046
## Area        -0.02604    0.01573  -1.655   0.0979 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.4177) family taken to be 1)
##
##     Null deviance: 77.327  on 30  degrees of freedom
## Residual deviance: 31.205  on 27  degrees of freedom
## AIC: 112.85
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  2.42
##           Std. Err.:  1.50
##
##  2 x log-likelihood:  -102.851
```

Appendix 5: Binomial GLM

Backwards variable selection for binomial GLM:

```r
bw_mod <- step(bin_mod1, scope = formula(bin_mod1), direction = c("backward"))
```

```
## Start:  AIC=22.45
## Miners ~ Eucs + Area + Grazed + Shrubs + Bulokes + Timber
##
##           Df Deviance    AIC
## - Bulokes  1    8.646 20.646
## - Area     1    9.152 21.152
## <none>          8.453 22.453
## - Shrubs   1   11.633 23.633
## - Grazed   1   14.761 26.761
## - Timber   1   16.856 28.856
## - Eucs     1   34.332 46.332
##
## Step:  AIC=20.65
## Miners ~ Eucs + Area + Grazed + Shrubs + Timber
##
##          Df Deviance    AIC
## - Area    1    9.272 19.272
## <none>         8.646 20.646
## - Shrubs  1   11.790 21.790
## - Grazed  1   14.796 24.796
## - Timber  1   16.857 26.857
## - Eucs    1   36.861 46.861
##
## Step:  AIC=19.27
## Miners ~ Eucs + Grazed + Shrubs + Timber
##
##          Df Deviance    AIC
## <none>         9.272 19.272
## - Shrubs  1   12.075 20.075
## - Grazed  1   14.986 22.986
## - Timber  1   16.865 24.865
## - Eucs    1   37.161 45.161
```

Likelihood ratio test:

```r
#test to see if dropping bulokes and area is correct
test <-bw_mod$deviance -bin_mod1$deviance
1-pchisq(test, 1) #drop bulokes and area
```

```
## [1] 0.3654317
```

```r
#from reduced model drop shrubs
mod_remove <- glm(Miners~ Eucs + Grazed + Timber, data=nminer, family= "binomial")
test <-mod_remove$deviance -bin_mod1$deviance
1-pchisq(test, 1) #keep eucs
```

```
## [1] 0.05702892
```

```r
#remove Area
mod_remove <- glm(Miners~ Eucs + Grazed + Shrubs + Bulokes + Timber, data=nminer, family= "binomial")
test <-mod_remove$deviance -bin_mod1$deviance
1-pchisq(test, 1) #remove area
```

```
## [1] 0.403031
```

```r
#remove  grazed
mod_remove <- glm(Miners~ Eucs +Area+ Shrubs + Bulokes + Timber, data=nminer, family= "binomial")
test <-mod_remove$deviance -bin_mod1$deviance
1-pchisq(test, 1)
```

```
## [1] 0.01201956
```

```r
#remove shrubs
mod_remove <- glm(Miners~ Eucs + Area + Grazed + Bulokes + Timber, data=nminer, family= "binomial")
test <-mod_remove$deviance -bin_mod1$deviance
1-pchisq(test, 1)
```

```
## [1] 0.07453822
```

```r
#remove Bulokes
mod_remove <- glm(Miners~ Eucs + Area + Grazed + Shrubs + Timber, data=nminer, family= "binomial")
test <-mod_remove$deviance -bin_mod1$deviance
1-pchisq(test, 1)
```

```
## [1] 0.6602756
```

```r
#remove Timber
mod_remove <- glm(Miners~ Eucs + Area + Grazed + Shrubs + Bulokes, data=nminer, family= "binomial")
test <-mod_remove$deviance -bin_mod1$deviance
1-pchisq(test, 1)
```

```
## [1] 0.003746181
```

Summmary of Backwards Selection Model:

```r
summary(bw_mod)
```

```
##
## Call:
## glm(formula = Miners ~ Eucs + Grazed + Shrubs + Timber, family = "binomial",
##     data = nminer)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.75996  -0.00447   0.00000   0.01082   1.65178
##
```

```
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.7086     4.3062  -1.558   0.1193
## Eucs          1.4941     0.8688   1.720   0.0855 .
## Grazed1      12.7870     8.1296   1.573   0.1157
## Shrubs1      -5.9792     4.5601  -1.311   0.1898
## Timber       -0.5258     0.3294  -1.596   0.1105
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 42.6843  on 30  degrees of freedom
## Residual deviance:  9.2721  on 26  degrees of freedom
## AIC: 19.272
##
## Number of Fisher Scoring iterations: 9
```

Summary of Full Model

```
summary(bin_mod1)
```

```
##
## Call:
## glm(formula = Miners ~ Eucs + Area + Grazed + Shrubs + Bulokes +
##     Timber, family = "binomial", data = nminer)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.94227  -0.00336   0.00000   0.00936   1.49256
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.521842   5.021468  -0.701   0.4831
## Eucs         1.623754   0.907901   1.788   0.0737 .
## Area        -0.086467   0.110931  -0.779   0.4357
## Grazed1     13.932017   8.209685   1.697   0.0897 .
## Shrubs1     -7.077190   5.401651  -1.310   0.1901
## Bulokes     -0.007585   0.015988  -0.474   0.6352
## Timber      -0.621664   0.386946  -1.607   0.1081
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 42.684  on 30  degrees of freedom
## Residual deviance:  8.453  on 24  degrees of freedom
## AIC: 22.453
##
## Number of Fisher Scoring iterations: 9
```

Summary of Final Model

```r
summary(final_mod)
```

```
##
## Call:
## glm(formula = Miners ~ Eucs + Grazed + Timber, family = "binomial",
##     data = nminer)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.95897  -0.13606   0.00276   0.08924   1.61465
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.3448     3.8788  -1.636   0.1019
## Eucs          0.7549     0.3902   1.934   0.0531 .
## Grazed1       6.7639     4.4999   1.503   0.1328
## Timber       -0.2011     0.1266  -1.588   0.1122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 42.684  on 30  degrees of freedom
## Residual deviance: 12.075  on 27  degrees of freedom
## AIC: 20.075
##
## Number of Fisher Scoring iterations: 8
```

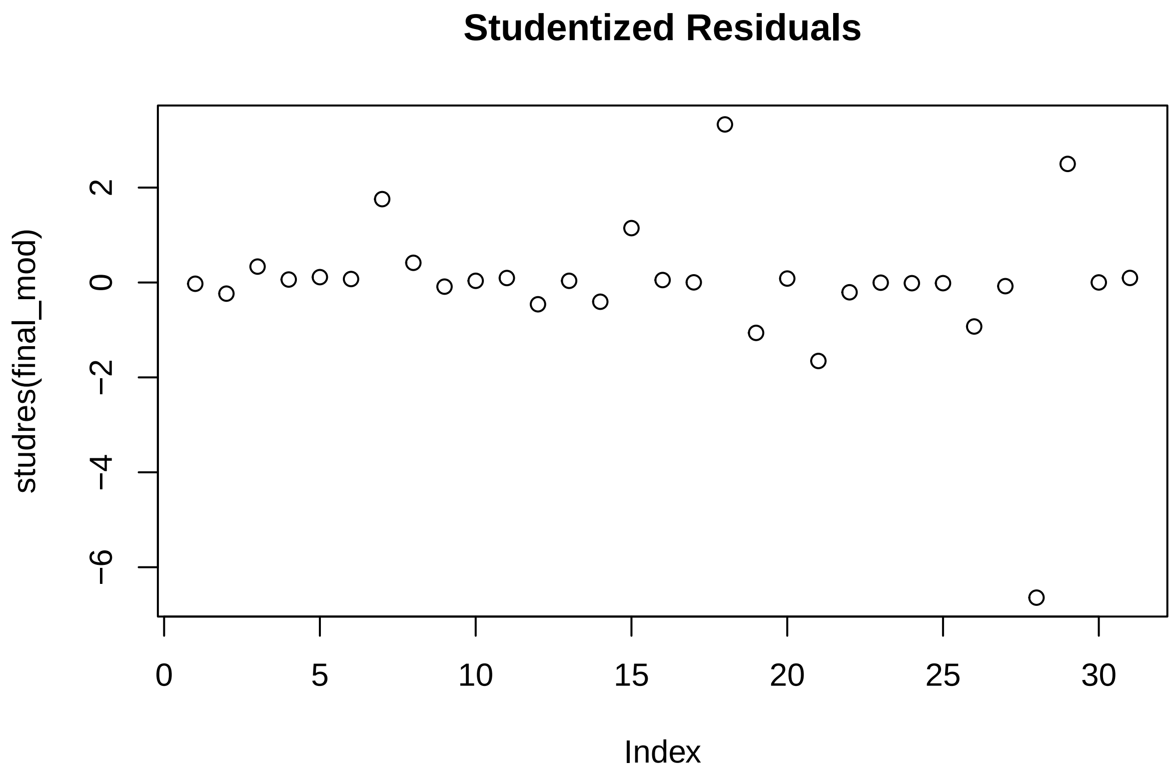
Influence plots and matrices for Final Model:

```r
im <- influence.measures(final_mod)
#understand how each case affects coefficient values
im$infmat
```

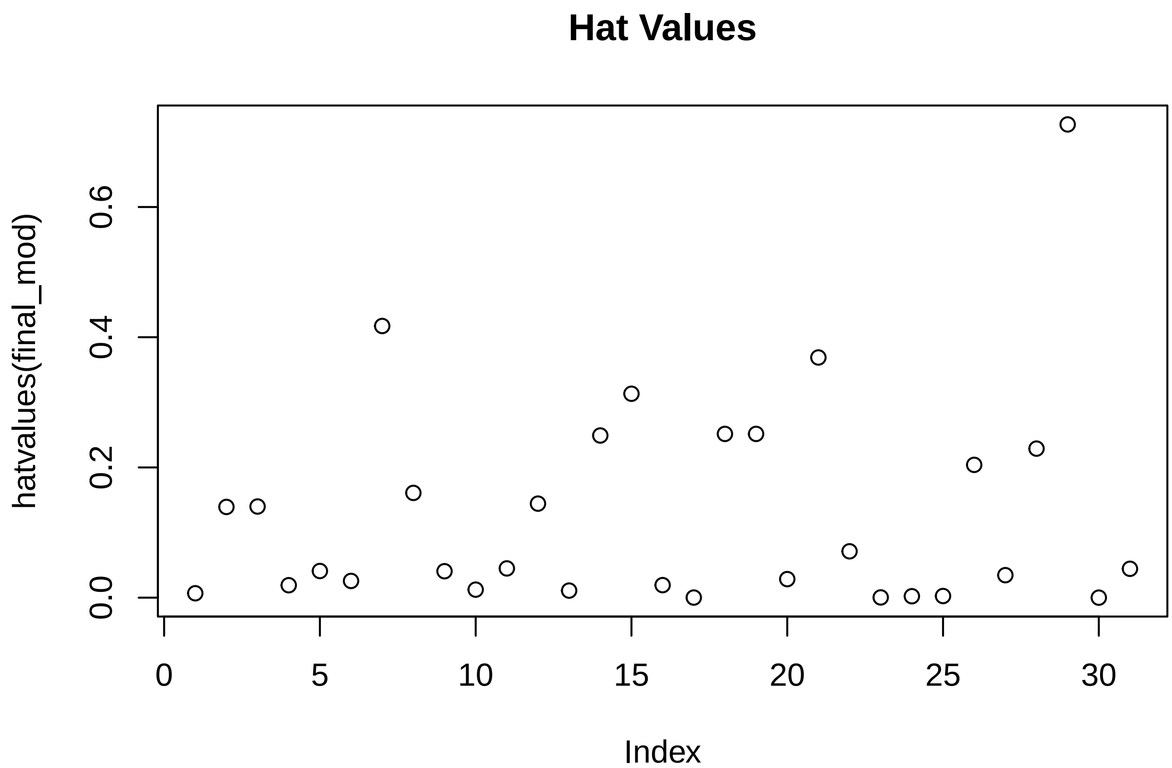
```
##          dfb.1_      dfb.Eucs      dfb.Grz1      dfb.Tmbr         dffit
## 1  -2.936211e-03  2.951754e-03  2.919249e-03 -2.459912e-03 -3.056928e-03
## 2  -1.203312e-01  1.193193e-01  1.268714e-01 -1.077158e-01 -1.325245e-01
## 3  -6.701118e-02  1.373595e-01  1.080305e-01 -1.502396e-01  1.885979e-01
## 4  -7.914810e-03  1.127969e-02  9.842833e-03 -1.110713e-02  1.228130e-02
## 5  -1.958828e-02  2.913180e-02  2.476328e-02 -2.877254e-02  3.278393e-02
## 6  -9.612795e-03  1.479164e-02  1.308905e-02 -1.550087e-02  1.668653e-02
## 7   9.259593e-01 -1.379504e-01 -4.119433e-01 -3.342184e-01  1.865947e+00
## 8  -7.683070e-02  1.717189e-01  1.277432e-01 -1.863482e-01  2.526495e-01
## 9  -2.518877e-02  2.304756e-02  2.285075e-02 -1.710405e-02 -2.532885e-02
## 10 -4.063168e-03  5.222253e-03  5.327596e-03 -5.350711e-03  5.577098e-03
## 11 -2.118579e-02  1.909740e-02  1.324939e-02 -7.556627e-03  2.881852e-02
## 12 -1.837515e-01  1.625468e-01  1.030604e-01 -9.059487e-02 -2.613380e-01
## 13 -4.320607e-03  5.071607e-03  5.055671e-03 -4.699613e-03  5.221052e-03
## 14 -3.158631e-01  2.519089e-01  2.622041e-01 -1.619440e-01 -3.257053e-01
## 15 -6.230881e-01  6.416770e-01  7.474120e-01 -4.932725e-01  1.017879e+00
## 16 -8.374329e-03  9.871932e-03  9.899835e-03 -9.195275e-03  1.022376e-02
## 17 -3.982959e-05  4.852433e-05  4.355783e-05 -4.359620e-05  4.935825e-05
```

```
## 18  5.255538e-01 -1.851154e-01  2.786043e-01 -3.161941e-01  1.882149e+00
## 19 -2.266583e-01  7.983567e-02 -1.201552e-01  1.363667e-01 -8.117241e-01
## 20 -1.212106e-02  1.795943e-02  1.559941e-02 -1.806482e-02  1.996854e-02
## 21  4.318588e-01 -7.684456e-02 -2.083693e-01 -4.145918e-01 -1.587894e+00
## 22 -6.771963e-02  6.658072e-02  5.166258e-02 -4.711474e-02 -8.018892e-02
## 23 -8.387203e-05  1.068673e-04  1.002806e-04 -1.085360e-04 -1.124726e-04
## 24 -7.513187e-04  9.256389e-04  8.512072e-04 -9.071375e-04 -9.653761e-04
## 25 -9.125000e-04  9.730086e-04  9.688811e-04 -8.712293e-04 -9.960510e-04
## 26 -9.623743e-02  1.945000e-01  9.096301e-02 -2.723320e-01 -6.256038e-01
## 27 -1.847392e-02  1.943760e-02  2.000053e-02 -1.788663e-02 -2.050643e-02
## 28  1.413523e-01 -1.081802e+00 -6.248723e-01  1.302248e+00 -2.326405e+00
## 29  2.400120e-01 -8.021461e-01 -2.098670e+00  2.652759e+00  5.415442e+00
## 30 -8.449954e-06  9.858247e-06  9.598418e-06 -9.066864e-06  9.965729e-06
## 31 -2.457377e-02  2.789557e-02  2.797610e-02 -2.477689e-02  2.911071e-02
##        cov.r       cook.d          hat
## 1  1.1705861 5.425382e-07 6.732554e-03
## 2  1.3289108 1.026348e-03 1.393377e-01
## 3  1.3077045 2.092590e-03 1.399746e-01
## 4  1.1841757 8.761049e-06 1.909299e-02
## 5  1.2079076 6.250757e-05 4.093855e-02
## 6  1.1916816 1.617665e-05 2.569245e-02
## 7  1.0051208 2.275489e-01 4.172424e-01
## 8  1.3171904 3.778385e-03 1.609015e-01
## 9  1.2093329 3.728420e-05 4.059278e-02
## 10 1.1770028 1.805980e-06 1.231715e-02
## 11 1.2143451 4.827389e-05 4.491925e-02
## 12 1.2778153 4.062700e-03 1.444100e-01
## 13 1.1752609 1.582750e-06 1.085175e-02
## 14 1.4747353 6.254212e-03 2.490573e-01
## 15 1.2111270 6.564317e-02 3.131911e-01
## 16 1.1847865 6.070291e-06 1.923377e-02
## 17 1.1631217 1.414276e-10 1.487207e-04
## 18 0.3982490 3.010913e-01 2.515325e-01
## 19 1.1617007 4.184832e-02 2.515325e-01
## 20 1.1945326 2.317104e-05 2.847306e-02
## 21 0.9972996 1.666591e-01 3.689083e-01
## 22 1.2360418 3.754622e-04 7.118053e-02
## 23 1.1633365 7.343591e-10 3.366097e-04
## 24 1.1654528 5.410298e-08 2.210658e-03
## 25 1.1658341 5.759572e-08 2.532431e-03
## 26 1.1627136 2.459294e-02 2.039787e-01
## 27 1.2023791 2.443261e-05 3.453835e-02
## 28 0.1800824 5.594792e-01 2.289164e-01
## 29 1.0358376 1.543698e+00 7.268582e-01
## 30 1.1629953 5.765436e-12 3.788938e-05
## 31 1.2134815 4.926000e-05 4.432932e-02
```

```r
plot(studres(final_mod), main ="Studentized Residuals")
```



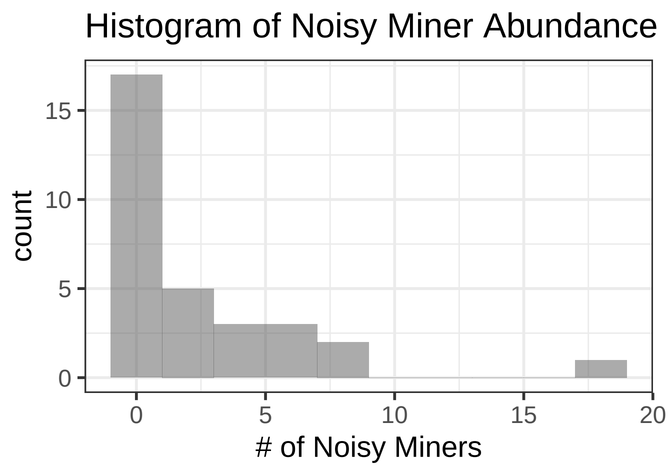

```r
plot(hatvalues(final_mod), main = 'Hat Values')
```



Appendix 7: One Sample Sign Test

```
##
##  One-sample Sign-Test
##
## data:  species_df$Difference
## s = 16, p-value = 0.002667
## alternative hypothesis: true median is not equal to 0
## 95 percent confidence interval:
##  -0.07505433 -0.01260504
## sample estimates:
## median of x
## -0.07142857
##
## Achieved and Interpolated Confidence Intervals:
##
##                    Conf.Level  L.E.pt  U.E.pt
## Lower Achieved CI     0.9419 -0.0714 -0.0126
## Interpolated CI       0.9500 -0.0751 -0.0126
## Upper Achieved CI     0.9700 -0.0840 -0.0126
```

Appendix 8: Nosiy Miner Count Modeling



As we can see from the plot above, the distribution of noisy miner abundance is highly skewed right.