Island Planting Project - Exploratory Analysis

# Background

Restoration must balance limited resources towards the most effective outcomes, given specified targets. Removing and suppressing invasive vegetation can give introduced native species the opportunity to establish. There are a variety of techniques used to help native species establish by suppressing weeds. Initially, laying mulch can reduce the cover of weedy species. However, it is less clear how much this initial benefit affects long-term success and native cover, or whether un-mulched areas will “catch up” with mulched plots once species establish. Annual mowing is another technique that can be used to reduce invasive species cover, and therefore potentially benefit native species.

In addition to the suppression of weedy invasive species, native species must be reintroduced to the system. The nucleation model of succession may provide insight into a technique to successfully restore system with reduced resource inputs. By planting target species in clumps (“islands”) as opposed to throughout the plot, we may effectively restore areas with reduced inputs. After multiple years of growth and spread, the native islands may expand, resulting over the long term in similar levels of native species cover with a lower up-front resource investment.

### Questions

**This study addresses the following questions:**  
1. Will island-planted plots have the same native cover as full-planted plots after several years?, and Will island planted plots spread outside of their planted areas?  
2. Does annual mowing benefit native species cover for either grasses or forbs?  
3. Does mulching affect native cover in the long-term/how long does the mulching effect last?  
4. Of the species planted, which were most successful in the long term?

# Data Structure

Before I begin the analysis, I need to take a look at the data’s structure – the data is certainly zero-inflated, but is it over-dispersed? A look at the data structure will help me select the best model for the data. I’ll base this on the guide in Appendix A of **Mixed effects models and extensions in ecology with R: appendix walkthrough**.

The replicated sampling unit differs for different questions. For the first question (how do treatments vary?), the replicate is a single plot, so I will average the individual samples to the plot level. Additionally, I’ll need to separate forbs and grasses in the analysis, since they were planted in separate areas. I’ll use “F” to denote forbs, and “G” for grasses.

## Selecting Revelant Data

We have only 2 remaining datacleanup steps before we can examine the data structure: 1. we need to look at forbs only in the forb planted plots, and only grass in the grass planted plots. 2. we need to average the cover of each guild over the multiple samples from a single plot.

## Year Plot Cover Category   
## Min. :2012 1 CIM : 7 Min. : 0.000 Full Mulched :34   
## 1st Qu.:2013 12 CFM : 7 1st Qu.: 4.344 Full Not Mulched:33   
## Median :2015 17 CFNM: 7 Median :14.844 Island Mulched :33   
## Mean :2015 18 CIM : 7 Mean :20.105   
## 3rd Qu.:2017 2 CFM : 7 3rd Qu.:31.615   
## Max. :2018 22 CFM : 7 Max. :72.781   
## (Other):58   
## guild   
## Length:100   
## Class :character   
## Mode :character   
##   
##   
##   
##

## Classes 'tbl\_df', 'tbl' and 'data.frame': 100 obs. of 5 variables:  
## $ Year : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...  
## $ Plot : Factor w/ 15 levels "1 CIM","12 CFM",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Cover : num 5.62 8.12 4.38 3.44 1.88 ...  
## $ Category: Factor w/ 3 levels "Full Mulched",..: 3 1 2 3 2 3 3 1 1 2 ...  
## $ guild : chr "forb" "forb" "forb" "forb" ...

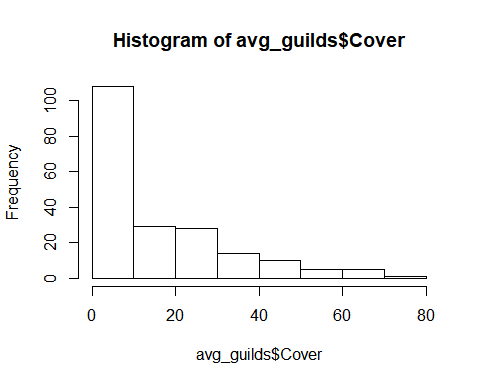
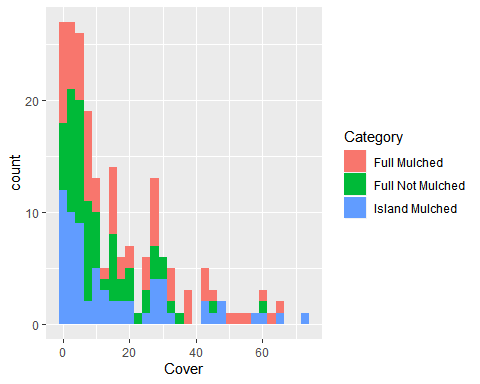
## Classes 'tbl\_df', 'tbl' and 'data.frame': 200 obs. of 5 variables:  
## $ Year : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...  
## $ Plot : Factor w/ 15 levels "1 CIM","12 CFM",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Cover : num 5.62 8.12 4.38 3.44 1.88 ...  
## $ Category: Factor w/ 3 levels "Full Mulched",..: 3 1 2 3 2 3 3 1 1 2 ...  
## $ guild : chr "forb" "forb" "forb" "forb" ...

Clearly, some of the data are missing (there should be 15 plots, 5 per category x 7 years = 35 plots per category). All of the missing plots are due to the 2014 dataset missing values. We are missing 1 FNM plot (#9), 2 FNM plots (#14, 25), and 2 IM plots (#15, 19), all from the 2014 dataset.  
There’s unfortunately nothing I can do about this – *Perhaps there is still hard copy data somewhere on file? Alternatively, we may not care, if most analyses are interested in comparisons in 2018.*

## Histogram and dispersion

Let’s take a look at the histogram of the data, since we’re fairly certain it’s nowhere near normal:

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



Yep, about as “normal” as we expected. Clearly, we have a zero-heavy dataset.  
In order to know what type of statistical model to use for our data, we need to know how overdispersed the data is. Generally, data is considered “overdispersed” when *d* (the ratio of the variance over the mean of the data). We can find the dispersion index for this data pretty easily:

disp <- avg\_guilds %>%  
 summarise(mean = mean(Cover),  
 var = var(Cover)) %>%  
 mutate(d = var/mean)  
  
disp

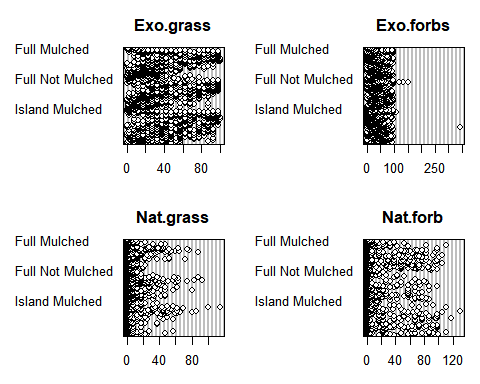
## # A tibble: 1 x 3  
## mean var d  
## <dbl> <dbl> <dbl>  
## 1 15.7 266. 16.9

Very clearly, we need to use a model that can handles a large number of zeroes and overdispersed data, but we’re also working with proportion data – not counts! So a zero-inflated negative binomial is probably not the best choice. A normal ANOVA is not robust against the overdispersion and zero-inflation in our data. .

# Appendix A guidelines

**Mixed effects models and extensions in ecology with R** has an appendix with a handy data-structure guide, so it’s valuable to take a look at our data through their guide.

### A.2.1: Outliers



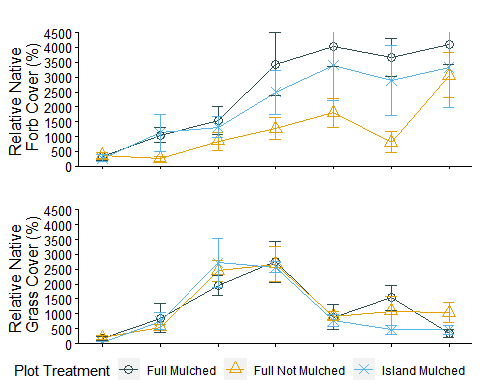
Clearly, there’s one outlier for exotic forb cover. Since our analyses might not deal with that, we can probably leave it in for now.

### A.2.2: Collinearity

## Question 1:

**1. Do island-planted plots have the same native cover as full-planted plots after 6 years?, and Will island planted plots spread outside of their planted areas?**

The cover of native species was fairly similar across all three treatments over time.



To test whether there is a signficiant difference in treatments after 7 years, we will use a zero-inflated negative binomial model. As I understand it, this type of model has two components - first, it will tell us whether the factors affect the likelihood of a zero occurring in the data, and as well as whether categories differ after accounting for differences in the likelihood of zeros.

### Old work – Zero inflated models?

avg\_F\_2018 <- avg\_F %>%  
 filter(Year == "2018")  
  
FZINB\_Cover <- zeroinfl(Cover ~ Category | Category,  
 data = avg\_F\_2018, dist="negbin")  
summary(FZINB\_Cover)

##   
## Call:  
## zeroinfl(formula = Cover ~ Category | Category, data = avg\_F\_2018,   
## dist = "negbin")  
##   
## Pearson residuals:  
## Min 1Q Median 3Q Max   
## -1.3150 -0.5696 -0.2558 0.4884 1.8547   
##   
## Count model coefficients (negbin with log link):  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.708683 0.229266 16.176 < 2e-16 \*\*\*  
## CategoryFull Not Mulched -0.307486 0.326941 -0.940 0.346965   
## CategoryIsland Mulched 0.004844 0.343869 0.014 0.988760   
## Log(theta) 1.434204 0.421740 3.401 0.000672 \*\*\*  
##   
## Zero-inflation model coefficients (binomial with logit link):  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -2.057e+01 1.307e+04 -0.002 0.999  
## CategoryFull Not Mulched -3.321e-08 1.849e+04 0.000 1.000  
## CategoryIsland Mulched 1.918e+01 1.307e+04 0.001 0.999  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1   
##   
## Theta = 4.1963   
## Number of iterations in BFGS optimization: 11   
## Log-likelihood: -62.38 on 7 Df

According to our zero-inflated model, there is no significant effect of planting treatment on the likelihood of there being non-0 cover, or an effect on the total quantity of native forbs in 2018. However, we may want to compare this with a non-zero inflated model and compare their fits – we’ll use a Vuong test to make the comparison.

FNB <- glm.nb(Cover ~ Category,  
 data = avg\_F\_2018)  
  
summary(FNB)

##   
## Call:  
## glm.nb(formula = Cover ~ Category, data = avg\_F\_2018, init.theta = 1.852007787,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.2938 -0.4419 -0.1843 0.3760 1.2039   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.7087 0.3360 11.038 <2e-16 \*\*\*  
## CategoryFull Not Mulched -0.3075 0.4770 -0.645 0.519   
## CategoryIsland Mulched -0.2183 0.4764 -0.458 0.647   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.852) family taken to be 1)  
##   
## Null deviance: 18.767 on 14 degrees of freedom  
## Residual deviance: 18.319 on 12 degrees of freedom  
## AIC: 142.34  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.852   
## Std. Err.: 0.755   
##   
## 2 x log-likelihood: -134.343

vuong(FZINB\_Cover, FNB)

## Vuong Non-Nested Hypothesis Test-Statistic:   
## (test-statistic is asymptotically distributed N(0,1) under the  
## null that the models are indistinguishible)  
## -------------------------------------------------------------  
## Vuong z-statistic H\_A p-value  
## Raw 1.0750153 model1 > model2 0.14118  
## AIC-corrected 0.4015369 model1 > model2 0.34401  
## BIC-corrected 0.1631087 model1 > model2 0.43522

#### Parameter Estimation through Bootstrapping

### Grasses

.  
.

## Question 2: Does annual mowing benefit native species cover for either grasses or forbs?

## Question 3: Does mulching affect native cover in the long-term/how long does the mulching effect last?

## Question 4: Of the species planted, which were most successful in the long term?