Island Planting Project - Exploratory Analysis

Josie C. Lesage

# 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.

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.

### Variables and values

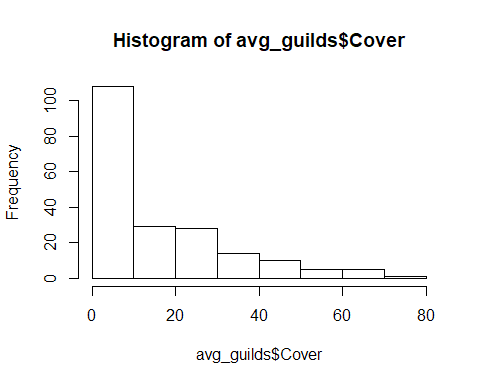
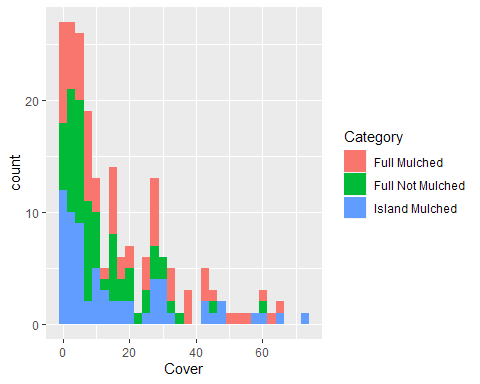
## Year Plot Cover Category   
## Min. :2012 1 CIM : 7 Min. : 0.000 Full Mulched :34   
## 1st Qu.:2013 12 CFM : 7 1st Qu.: 4.344 Full Not Mulched:33   
## Median :2015 17 CFNM: 7 Median :14.844 Island Mulched :33   
## Mean :2015 18 CIM : 7 Mean :20.105   
## 3rd Qu.:2017 2 CFM : 7 3rd Qu.:31.615  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

## Classes 'tbl\_df', 'tbl' and 'data.frame': 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 – *ask Karen about it, perhaps there is still hard copy data somewhere on file? Alternatively, we may not care, if most analyses are interested in comparisons in 2018.*

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. Let’s check the dispersion.

### Dispersion and Variance

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

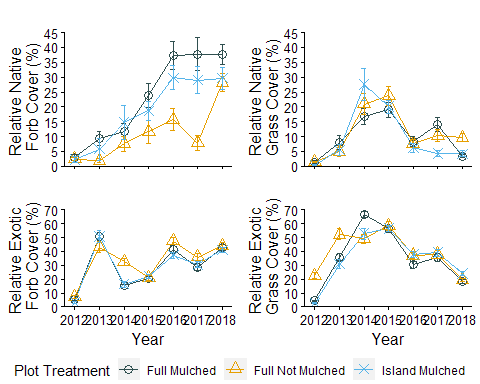
## # A tibble: 1 x 3  
## mean var d  
## <dbl> <dbl> <dbl>  
## 1 15.7 266. 16.9

Very clearly, we need to use a model that best handles a large number of zeroes and overdispersed data: a zero-inflated negative bionomial distribution.

## 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 and exotic species was fairly similar across all three treatments over time.



To test whether there is a significant difference between the treatments in 2018, we will likely use a zero-inflated negative binomial model.

## 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?