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 : 13 Min. : 0.000 Full Mulched :62   
## 1st Qu.:2014 12 CFM : 13 1st Qu.: 3.333 Full Not Mulched:61   
## Median :2015 17 CFNM: 13 Median :13.933 Island Mulched :61   
## Mean :2015 18 CIM : 13 Mean :21.625   
## 3rd Qu.:2017 2 CFM : 13 3rd Qu.:34.703   
## Max. :2018 22 CFM : 13 Max. :76.875   
## (Other):106   
## Mowing.treatment guild   
## Mowed :84 Length:184   
## None :15 Class :character   
## Unmowed:85 Mode :character   
##   
##   
##   
##

## Classes 'tbl\_df', 'tbl' and 'data.frame': 184 obs. of 6 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 ...  
## $ Mowing.treatment: Factor w/ 3 levels "Mowed","None",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ guild : chr "forb" "forb" "forb" "forb" ...

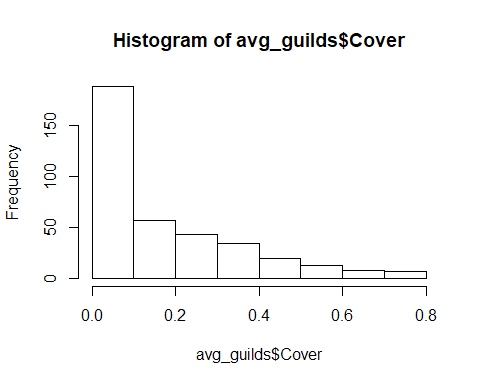
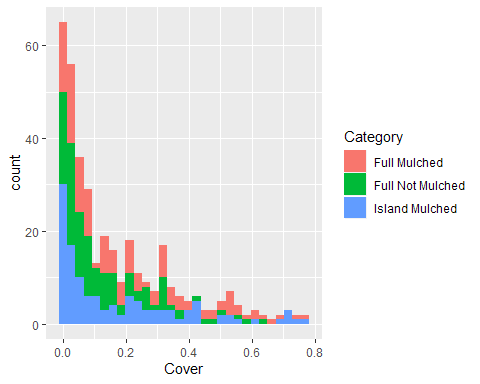
## Classes 'tbl\_df', 'tbl' and 'data.frame': 369 obs. of 8 variables:  
## $ Year : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...  
## $ Plot : Factor w/ 15 levels "1 CIM","12 CFM",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Cover : num 0.0563 0.0813 0.0437 0.0344 0.0187 ...  
## $ Category : Factor w/ 3 levels "Full Mulched",..: 3 1 2 3 2 3 3 1 1 2 ...  
## $ Mowing.treatment: Factor w/ 3 levels "Mowed","None",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ guild : chr "forb" "forb" "forb" "forb" ...  
## $ Planting : chr "Island" "Full" "Full" "Island" ...  
## $ Mulch : chr "Mulch" "Mulch" "No Mulch" "Mulch" ...

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, dispersion, outliers

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`.



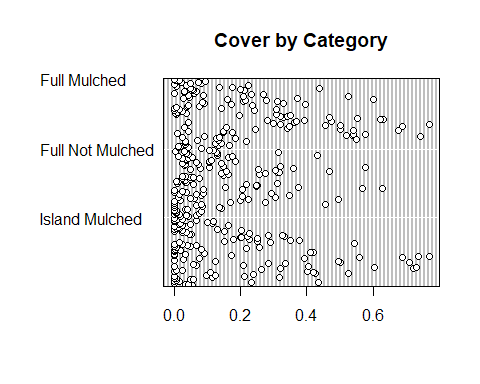
Our data isn’t normal. Here’s a table of the variance and dispersion index of our data:

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

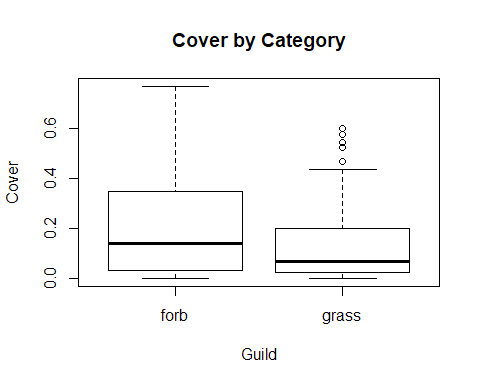
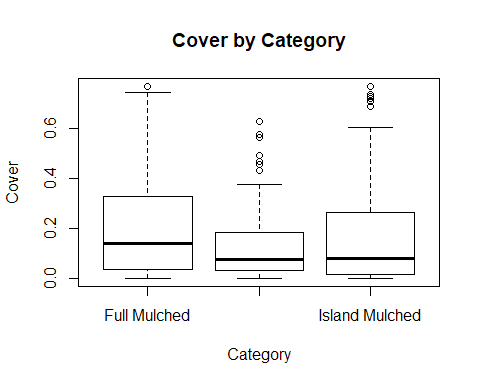
## # A tibble: 1 x 3  
## mean var d  
## <dbl> <dbl> <dbl>  
## 1 0.169 0.0339 0.201

Very clearly, we need to use a model that can handle the quirks of our data: proportions, skewed, and an unbalanced comparison. A zero-inflated negative binomial is probably not the best choice. A normal ANOVA is not robust against the overdispersion in our data. We’ll settle for a GLM.

### A.2.1: Outliers

**Mixed effects models and extensions in ecology with R** suggests using dot plots to check for outliers:  
 There is an outlier or two for exotic species cover, but because we’ll be analyzing native species cover, we can leave those in.

### Boxplots



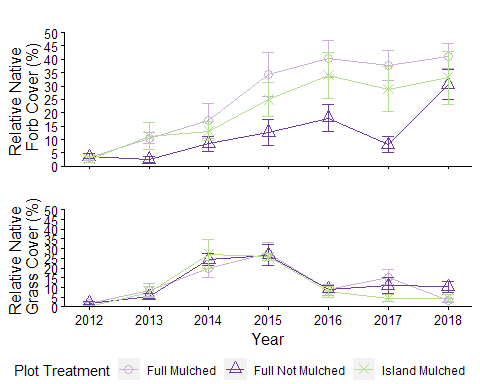
From the first boxplot of categories, it is fairly clear that the heterogeneity rules are being broken.

# Results

## Questions 1, 2, & 3:

**1. Do island-planted plots have the same native cover as full-planted plots after 6 years?**  
**2. Does mulching affect native cover in the long-term/how long does the mulching effect last?**  
**3. Does annual mowing benefit native species cover for either grasses or forbs?**

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 GLM on only the 2018 data.

The zero-inflated model is not the correct model to use because it is based on count data - which our data are not.  
I am running a generalized linear model with a quasibinomial structure, given that our data are proportion data and slightly overdispersed.

We also suspect that our data do not have homogeneous variance, so we will allow flexible variance.

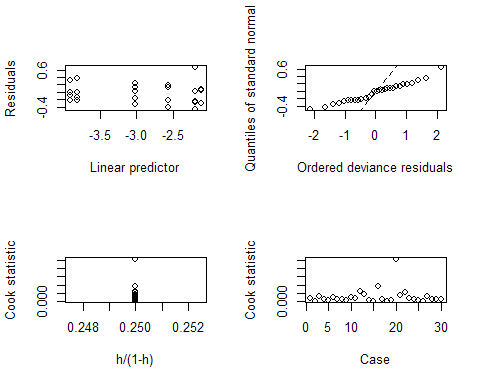
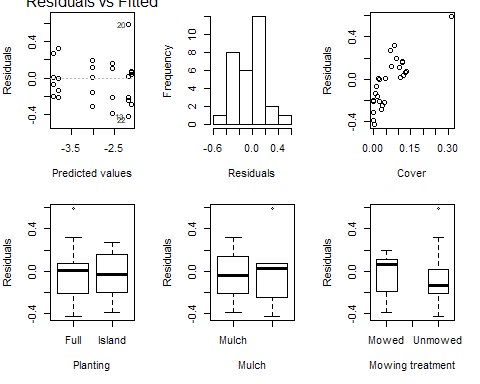
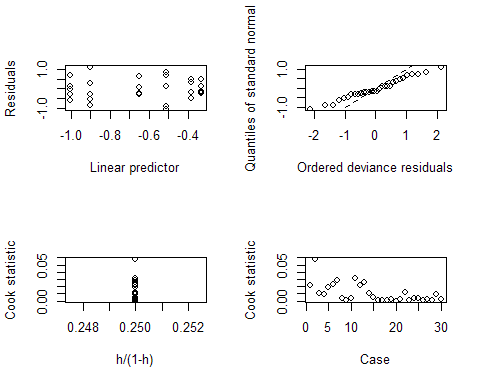
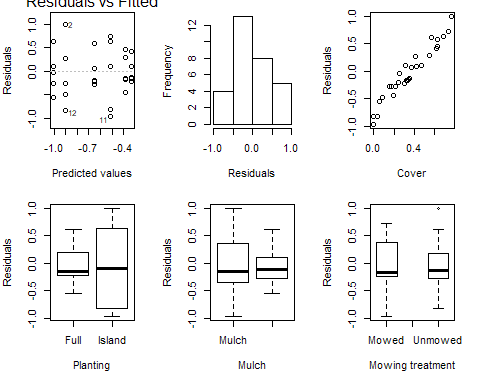
##   
## Call:  
## glm(formula = Cover ~ Planting \* Mulch \* Mowing.treatment, family = quasibinomial,   
## data = F2018)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9695 -0.2710 -0.1485 0.2885 0.9922   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error  
## (Intercept) -0.38208 0.45572  
## PlantingIsland -0.12874 0.64909  
## MulchNo Mulch -0.26676 0.65559  
## Mowing.treatmentUnmowed 0.04904 0.64310  
## PlantingIsland:MulchNo Mulch NA NA  
## PlantingIsland:Mowing.treatmentUnmowed -0.44089 0.93335  
## MulchNo Mulch:Mowing.treatmentUnmowed -0.40244 0.94373  
## PlantingIsland:MulchNo Mulch:Mowing.treatmentUnmowed NA NA  
## t value Pr(>|t|)  
## (Intercept) -0.838 0.410  
## PlantingIsland -0.198 0.844  
## MulchNo Mulch -0.407 0.688  
## Mowing.treatmentUnmowed 0.076 0.940  
## PlantingIsland:MulchNo Mulch NA NA  
## PlantingIsland:Mowing.treatmentUnmowed -0.472 0.641  
## MulchNo Mulch:Mowing.treatmentUnmowed -0.426 0.674  
## PlantingIsland:MulchNo Mulch:Mowing.treatmentUnmowed NA NA  
##   
## (Dispersion parameter for quasibinomial family taken to be 0.2503557)  
##   
## Null deviance: 7.1184 on 29 degrees of freedom  
## Residual deviance: 6.7016 on 24 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 4

## Analysis of Deviance Table  
##   
## Model: quasibinomial, link: logit  
##   
## Response: Cover  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)  
## NULL 29 7.1184   
## Planting 1 0.021411 28 7.0970 0.7700  
## Mulch 1 0.243448 27 6.8535 0.3241  
## Mowing.treatment 1 0.082549 26 6.7710 0.5658  
## Planting:Mulch 0 0.000000 26 6.7710   
## Planting:Mowing.treatment 1 0.023758 25 6.7472 0.7580  
## Mulch:Mowing.treatment 1 0.045622 24 6.7016 0.6695  
## Planting:Mulch:Mowing.treatment 0 0.000000 24 6.7016

##   
## Call:  
## glm(formula = Cover ~ Planting \* Mulch \* Mowing.treatment, family = quasibinomial,   
## data = G2018)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.42295 -0.20910 0.00228 0.10121 0.58564   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error  
## (Intercept) -3.0235 0.5272  
## PlantingIsland 0.4653 0.6798  
## MulchNo Mulch 0.9096 0.6370  
## Mowing.treatmentUnmowed -0.7885 0.9267  
## PlantingIsland:MulchNo Mulch NA NA  
## PlantingIsland:Mowing.treatmentUnmowed -0.5644 1.2968  
## MulchNo Mulch:Mowing.treatmentUnmowed 0.7108 1.0595  
## PlantingIsland:MulchNo Mulch:Mowing.treatmentUnmowed NA NA  
## t value Pr(>|t|)   
## (Intercept) -5.735 6.56e-06 \*\*\*  
## PlantingIsland 0.684 0.500   
## MulchNo Mulch 1.428 0.166   
## Mowing.treatmentUnmowed -0.851 0.403   
## PlantingIsland:MulchNo Mulch NA NA   
## PlantingIsland:Mowing.treatmentUnmowed -0.435 0.667   
## MulchNo Mulch:Mowing.treatmentUnmowed 0.671 0.509   
## PlantingIsland:MulchNo Mulch:Mowing.treatmentUnmowed NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for quasibinomial family taken to be 0.06145429)  
##   
## Null deviance: 2.1998 on 29 degrees of freedom  
## Residual deviance: 1.5194 on 24 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 6

## Analysis of Deviance Table  
##   
## Model: quasibinomial, link: logit  
##   
## Response: Cover  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)  
## NULL 29 2.1998   
## Planting 1 0.06601 28 2.1338 0.30002  
## Mulch 1 0.39957 27 1.7342 0.01078  
## Mowing.treatment 1 0.10539 26 1.6288 0.19036  
## Planting:Mulch 0 0.00000 26 1.6288   
## Planting:Mowing.treatment 1 0.08069 25 1.5482 0.25185  
## Mulch:Mowing.treatment 1 0.02875 24 1.5194 0.49400  
## Planting:Mulch:Mowing.treatment 0 0.00000 24 1.5194   
##   
## NULL   
## Planting   
## Mulch \*  
## Mowing.treatment   
## Planting:Mulch   
## Planting:Mowing.treatment   
## Mulch:Mowing.treatment   
## Planting:Mulch:Mowing.treatment   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Checking the residuals

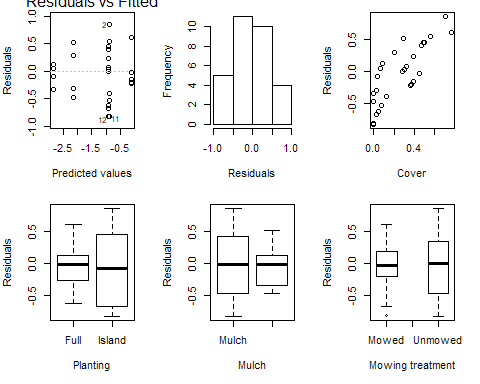


### Curiosity about 2017

I’m curious what happens for the 2017 data, since it seems much clearer that there’s an effect of mulching/planting style.

##   
## Call:  
## glm(formula = Cover ~ Planting \* Mowing.treatment \* Mulch, family = quasibinomial,   
## data = F2017)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.83330 -0.38569 -0.01656 0.28059 0.85599   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error  
## (Intercept) -0.1201 0.4096  
## PlantingIsland -0.8078 0.6111  
## Mowing.treatmentUnmowed -0.8058 0.6110  
## MulchNo Mulch -2.6999 0.9771  
## PlantingIsland:Mowing.treatmentUnmowed 0.8545 0.8835  
## PlantingIsland:MulchNo Mulch NA NA  
## Mowing.treatmentUnmowed:MulchNo Mulch 1.4830 1.2669  
## PlantingIsland:Mowing.treatmentUnmowed:MulchNo Mulch NA NA  
## t value Pr(>|t|)   
## (Intercept) -0.293 0.7718   
## PlantingIsland -1.322 0.1987   
## Mowing.treatmentUnmowed -1.319 0.1997   
## MulchNo Mulch -2.763 0.0108 \*  
## PlantingIsland:Mowing.treatmentUnmowed 0.967 0.3431   
## PlantingIsland:MulchNo Mulch NA NA   
## Mowing.treatmentUnmowed:MulchNo Mulch 1.171 0.2532   
## PlantingIsland:Mowing.treatmentUnmowed:MulchNo Mulch NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for quasibinomial family taken to be 0.2089098)  
##   
## Null deviance: 9.1045 on 29 degrees of freedom  
## Residual deviance: 5.8707 on 24 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 5

## Analysis of Deviance Table  
##   
## Model: quasibinomial, link: logit  
##   
## Response: Cover  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)  
## NULL 29 9.1045   
## Planting 1 0.12473 28 8.9797 0.4396964  
## Mowing.treatment 1 0.07290 27 8.9068 0.5547148  
## Mulch 1 2.66055 26 6.2463 0.0003588  
## Planting:Mowing.treatment 1 0.07560 25 6.1707 0.5474689  
## Planting:Mulch 0 0.00000 25 6.1707   
## Mowing.treatment:Mulch 1 0.30002 24 5.8707 0.2307647  
## Planting:Mowing.treatment:Mulch 0 0.00000 24 5.8707   
##   
## NULL   
## Planting   
## Mowing.treatment   
## Mulch \*\*\*  
## Planting:Mowing.treatment   
## Planting:Mulch   
## Mowing.treatment:Mulch   
## Planting:Mowing.treatment:Mulch   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



## Question 4: Did the cover in island planted plots spread outside of the planted areas?

First, we need to create a dataframe of the island planted sections vs. the full sections for only 2018.

## Year Plot Cover QuadLocation  
## Min. :2013 1 CIM :12 Min. : 0.000 Edge :20   
## 1st Qu.:2015 15 CIM :12 1st Qu.: 5.703 Full :50   
## Median :2016 18 CIM :12 Median :21.375 Island :20   
## Mean :2016 19 CIM :12 Mean :27.161 Not recorded: 0   
## 3rd Qu.:2017 4 CIM :12 3rd Qu.:43.719 Out :20   
## Max. :2018 12 CFM : 5 Max. :97.250   
## (Other):45   
## guild   
## Length:110   
## Class :character   
## Mode :character   
##   
##   
##   
##

## Classes 'tbl\_df', 'tbl' and 'data.frame': 110 obs. of 5 variables:  
## $ Year : int 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 ...  
## $ Plot : Factor w/ 15 levels "1 CIM","12 CFM",..: 2 3 5 8 9 10 11 13 14 15 ...  
## $ Cover : num 15 3.75 0.833 8.75 7.917 ...  
## $ QuadLocation: Factor w/ 5 levels "Edge","Full",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ guild : chr "forb" "forb" "forb" "forb" ...

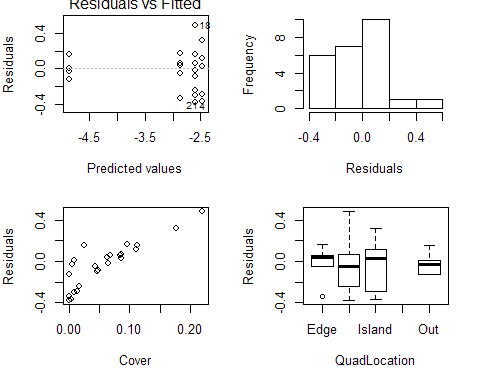
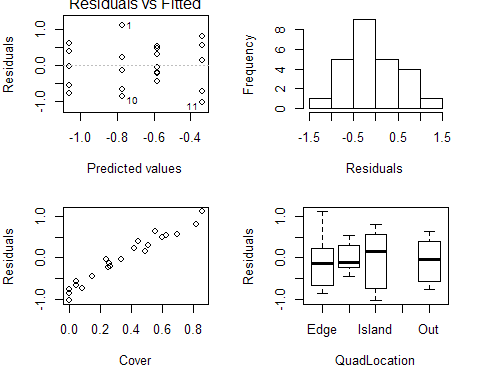
## Joining, by = c("Year", "Plot", "Cover", "QuadLocation", "guild")

##   
## Call:  
## glm(formula = Cover ~ QuadLocation, family = quasibinomial, data = F\_spread18)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.03937 -0.46090 -0.05102 0.40025 1.12131   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.7711 0.5395 -1.429 0.168  
## QuadLocationFull 0.1901 0.6541 0.291 0.774  
## QuadLocationIsland 0.4373 0.7416 0.590 0.562  
## QuadLocationOut -0.2932 0.7881 -0.372 0.714  
##   
## (Dispersion parameter for quasibinomial family taken to be 0.3147187)  
##   
## Null deviance: 7.9710 on 24 degrees of freedom  
## Residual deviance: 7.6511 on 21 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 4

## Analysis of Deviance Table  
##   
## Model: quasibinomial, link: logit  
##   
## Response: Cover  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)  
## NULL 24 7.9710   
## QuadLocation 3 0.31995 21 7.6511 0.7972

##   
## Call:  
## glm(formula = Cover ~ QuadLocation, family = quasibinomial, data = G\_spread18)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.37832 -0.12407 -0.01492 0.06542 0.48442   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.8584 0.4351 -6.569 1.66e-06 \*\*\*  
## QuadLocationFull 0.2572 0.5147 0.500 0.622   
## QuadLocationIsland 0.3909 0.5693 0.687 0.500   
## QuadLocationOut -2.0048 1.2108 -1.656 0.113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for quasibinomial family taken to be 0.04856866)  
##   
## Null deviance: 1.5267 on 24 degrees of freedom  
## Residual deviance: 1.1117 on 21 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 7

## Analysis of Deviance Table  
##   
## Model: quasibinomial, link: logit  
##   
## Response: Cover  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 24 1.5267   
## QuadLocation 3 0.41508 21 1.1117 0.03597 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



# Means and Std. Err reports

These are the calculations for the means and std. errors, to be reported in the final paper.

## Warning: Removed 3 rows containing missing values (geom\_path).

## Warning: Removed 3 rows containing missing values (geom\_point).

## Warning: Removed 3 rows containing missing values (geom\_path).

## Warning: Removed 3 rows containing missing values (geom\_point).

## Warning: Removed 3 rows containing missing values (geom\_path).

## Warning: Removed 3 rows containing missing values (geom\_point).

## Warning: Graphs cannot be vertically aligned unless the axis parameter is  
## set. Placing graphs unaligned.

