STAT 427 REPORT

Pollinator Habitat Restoration Tools

Spring 2018

Mengjie Wang
Renjie Wei
Tianxin Wu
Xinyan Yang

Table of Contents

1 Introduction

1.1 Background

In the long-term attempt of converting natural landscape into pollinator habitat, imazapic has been proved a highly effective method. Imazapic is a chemical used as an herbicide. It controls many broadleaf weeds and controls or suppresses some grasses in pasture, rangeland and certain types of turf. Imazapic and tolerant seed mixes are readily available, and are frequently used in practice, to transform local landscape into a strong community of desired flowers and grass. However, existing research in this field shows limit success in experiments and is far from satisfying in the real world. In this case, based on a field experiment carried out in Bond County, Illinois, our client and the team worked together to figure out factors that encourage the strongest community assembly of desired species as well as minimizing weed pressure.

Restoration failure is what usually occurs. Due to weed pressure and some other reasons, established desired species cannot out-compete native ones as we plan. In this case, artificial regular disturbances like controlled burning, seed mix preprocessing, cold stratification, etc. are introduced to achieve desired goals

1.2 Goals

To be specific, the study expects to solve three problems:

1.Whether cold stratification contributes to the success of native wildflower and native grass establishment.

2. How a traditional “pollinator habitat” seed mix fairs with and without imazapic application.

3. Whether application of imazapic immediately after seed broadcast (pre-emergent) or after germination has occurred (post-emergent), contributes to the success of native wildflower and native grass establishment

1.3 Experiment Design

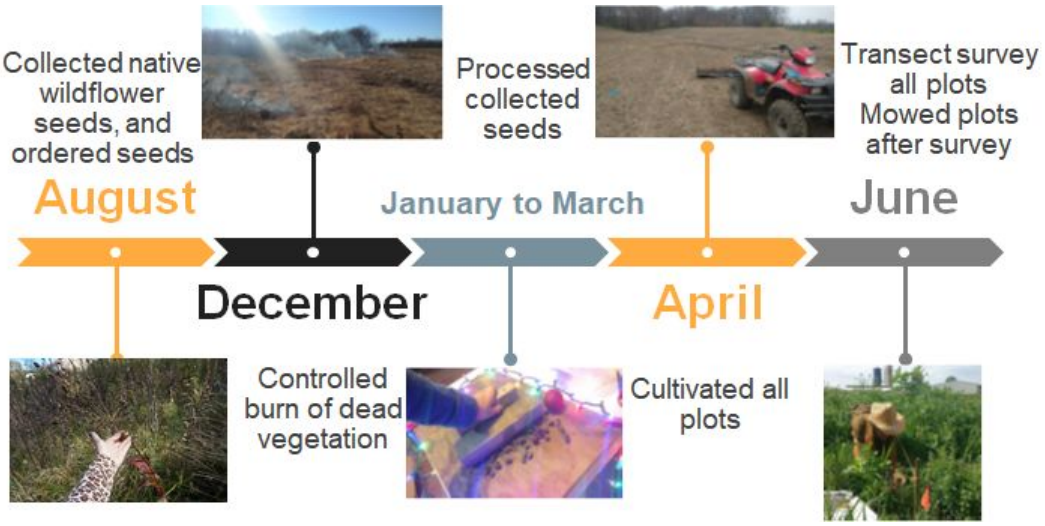
The study site is in Bond County, Illinois, Burgess Township, Section 34, NE Quarter. It is divided into 60 3-meter-squared plots with 3-meter-squared buffer zones. 12 different plot treatments are randomly assigned throughout the site for a total of five replications for each treatment.

Graph 1.1 The Aerial View

Twelve types of plots correspond to twelve different treatments. First, concerning cold stratification, seed mix in solid plots did not do this, while seed mix in striped plots did. Second, about herbicide application timing, red and yellow plots undertook pre-emergent herbicide, orange and green plots had post-emergent herbicide application, and blue and white plots had no herbicide application. Third, red, orange and blue plots share the same imazapic-tolerant seed mix, while yellow, green and white plots had pollinator traditional seed mix.

Graph 1.2 The Experimental Field

Before the start of sowing, lots of preparations have been done. In August of 2016, our client collected 30 species of native wildflower seeds, and ordered seeds. One month later, she controlled burning down existing vegetation. Then, from the end of 2016 to the second month of 2017, it was processing of collected seeds. In this part, firstly she created seed mixes with NRCS seed calculator, then mixed the seeds and applied cold stratification to half of the seeds. In March 2017, 60 study plots were measured and marked according to treatment color. All plots were cultivated in May. Just before that, herbicide was released in two ways (pre-emergent or post-emergent). Finally, in June, August and October of 2017, survey was conducted three times in all plots and research data was collected.



Graph 1.3 Time Schedule

1.4 Data Preprocessing

1.4.1 Data Cleaning

All the missing values in the dataset are replaced by 0 on the purpose of further analysis.

		Grass										Forbs
	June 17 - June 20, 2017	Big bluestem	Fall panicum	Giant foxtail	Green foxtail	Johnson grass	Large crab grass	Little bluestem	Reed canary grass	Tall fescue	Yellow foxtail	Bergamot
	Plot 1	3	9	39	13		9					
	Plot 2		34	22	7	2	5					
	Plot 3	2	9	10	3	1	6	2		6		
	Plot 4	3		24			4					
	Plot 5			1		6	1			3		
	Plot 1		34	78	49	2	12			16		
	Plot 2			16	2		31			3		
	Plot 3			9	2	10	18					
	Plot 4		3	5		6					3	
	Plot 5	5	2	32	8		36	3	34	9		
	Plot 1	5	65	5	4		17	10		1		
	Plot 2		24	17	9		2					
	Plot 3			9	2		12					
	Plot 4	3		12		3	2					
	Plot 5	4		6	2			1		3		
	Plot 1	2	11	23	6		5			1		1

Table 1.1 Raw Data

1.4.2 Data Integration

In this section, variables are created as follows:

The target variable is *Ratio*, which is calculated as belows:

$$Ratio = (Grass + Forbs) / (Grass + Forbs + Weeds)$$

This number shows how strong desired species grow in the field. Here *Grass* is defined as the add-up number of big bluestem and little bluestem - the only two desired grass. *Weeds* include weeds and all unwanted grass.

The predictors are created based on different treatments. All the predictor variables are categorical variables.

1	Plot	Stratified	Seedmix	Application
2	1	0	1	1
3	2	0	1	1
4	3	0	1	1
5	4	0	1	1
6	5	0	1	1
7	1	1	1	1
8	2	1	1	1
9	3	1	1	1
10	4	1	1	1
11	5	1	1	1
12	1	0	1	2
13	2	0	1	2
14	3	0	1	2
15	4	0	1	2
16	5	0	1	2
17	1	1	1	2
18	2	1	1	2
19	3	1	1	2
20	4	1	1	2
21	5	1	1	2
22	1	0	0	1
23	2	0	0	1

Table 1.2 Treatment Variables Encoding

The above table shows three variables representing three different treatments.

For variable "*Stratified*", there exists two levels "0" and "1"; "0" means it did not perform cold stratification, and "1" means it performed cold stratification.

For variable "*Seedmix*", there exists two levels "0" and "1"; "0" means it used traditional pollinator seed mix, and "1" means it used Plateau-tolerant pollinator seed mix.

For variable "*Application*", there exists three levels "0", "1" and "2"; "0" means there was no herbicide application, "1" means it performed pre-emergent application and "2" means it performed post-emergent application.

1	Plot	SoilHealth	SoilStructure	NutrientR	MoistureRegime
2	1	Healthy	Healthy		Well Drained
3	2	Healthy	Healthy	Mild	Well Drained
4	3	Less Healthy	Less Healthy	Mild	Well Drained
5	4	Less Healthy	Less Healthy		Well Drained
6	5	Healthy	Healthy		Well Drained
7	1	Healthy	Healthy		Well Drained
8	2	Disturbed	Compacted		Poorly Drained
9	3	Disturbed	Compacted	Mild	Well Drained
10	4	Healthy	Healthy		Well Drained
11	5	Disturbed	Hydric		Wet
12	1	Healthy	Healthy		Well Drained
13	2	Healthy	Healthy		Well Drained
14	3	Less Healthy	Less Healthy		Well Drained
15	4	Healthy	Healthy		Well Drained
16	5	Disturbed	Hydric		Wet
17	1	Healthy	Healthy		Well Drained
18	2	Less Healthy	Less Healthy		Well Drained
19	3	Disturbed	Compacted	Mild	Well Drained
20	4	Healthy	Healthy	Considera	Well Drained
21	5	Disturbed	Less Healthy	Mild	Poorly Drained
22	1	Healthy	Healthy		Well Drained
23	2	Healthy	Healthy		Well Drained

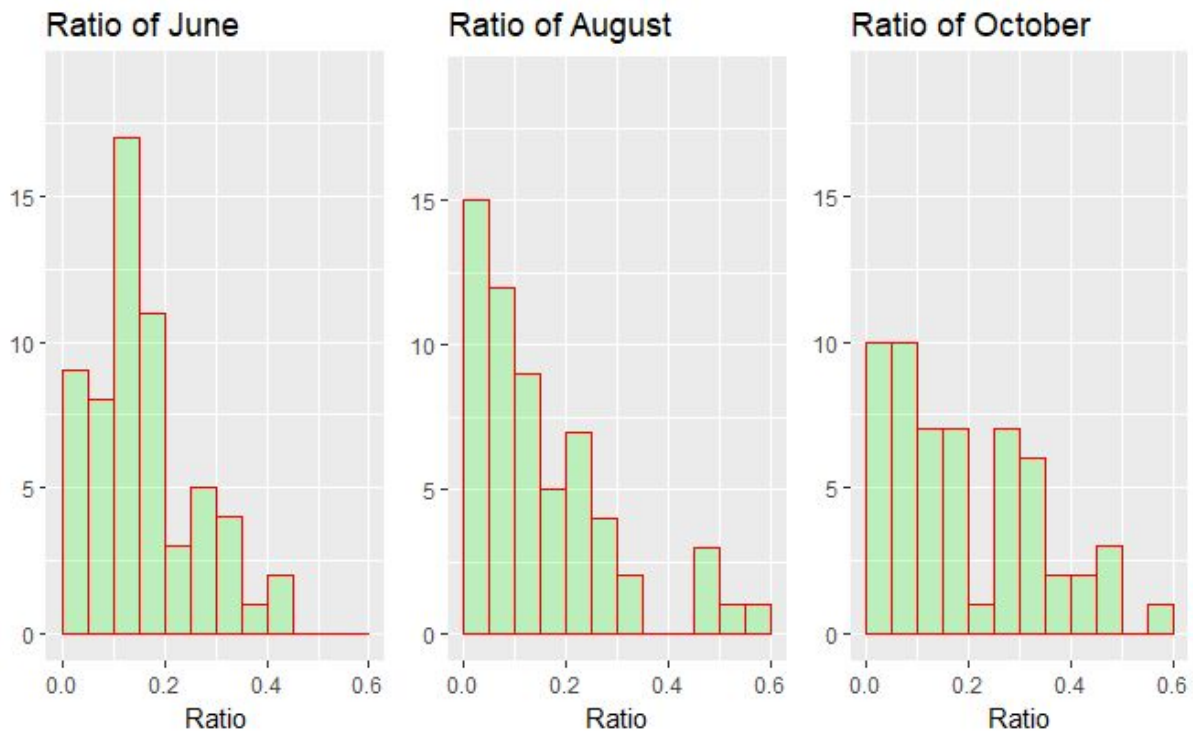
Table 1.3 Soil Conditions Variables Encoding

From the table above, we can see four variables which reveal the soil conditions. The variable "*SoilHealth*" has three levels, which are "healthy", "less healthy" and "disturbed"; The variable "*SoilStructure*" has three levels, which are "Healthy", "Less Healthy" and "Compacted"; The variable "NutrientRunoff"has three levels, which are "Mild", "Considerable" and "Normal"(Converting the missing values into "Normal"); The variable "*MoistureRegime*" has three levels, which are "Well Drained", "Poorly Drained" and "Wet".

Also, there exists high correlation between "*SoilHealth*" and "*SoilStructure*". More information about checking correlation between the four variables is included in Appendix. And there exist many missing values in "*MoistureRegime*". Thus, only two variables "*SoilHealth*" and "*NutrientRunoff*" are considered in the following analysis.

2 Visualization

2.1 Check the distribution of response variables



Plot 2.1 The distribution of Ratio

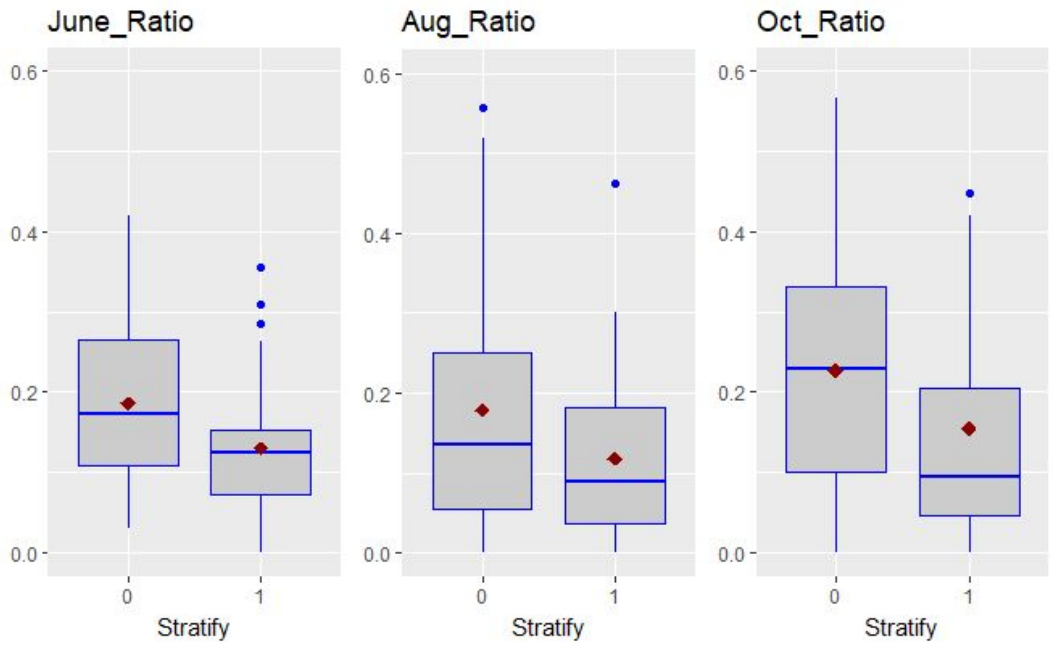
At first, the distribution of response variables (the ratio value of three different months) was checked. In the plot above, we could find that all three months' ratios fall in a range of 0 to 0.6 and many cases have a ratio value of 0. In June, we have the most frequent ratio value of 0.1 while in August it becomes 0.0. When it comes to October, more observations have a larger ratio value, which is the same as what we expected. The reason is that as time goes by, more expected species would grow.

2.2 Check the effects of three different factors

Next, the effects of three interested factors were checked one by one through box plot.

2.2.1 Check the effect of seed treatment

First the effect of seed treatment was checked. In the following three plot, 0 indicates no stratification and 1 indicates cold stratification. The effects of this factor on three different months were put together for comparing purpose.



Plot 2.2 The effect of Stratification

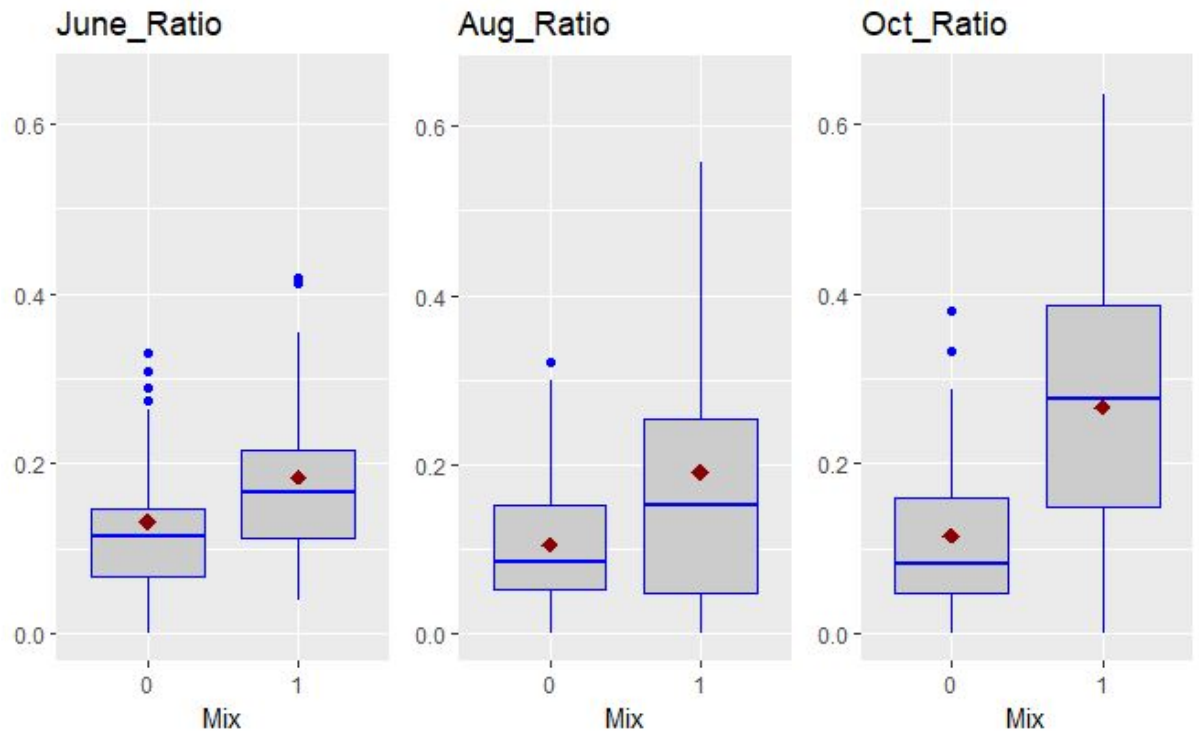
From the above three plots, we notice a similarance that level 0 always has a slightly higher mean and median ratio value than the level 1, which means that no stratification would have a better influence on the ratio than the cold stratification does. This result is different from what we expected before and our group discussed the potential reasons of this phenomenon.

It could be that the time of the cold stratification was not appropriate or that there was any other unknown factors which may have a mixed effect with the cold stratification treatment. In addition, the sample size of this experiment is very small, which could

result in a inaccurate measurement of the effect of this treatment on the species. Therefore more experiments may need to explore the effect of this factor.

2.2.2 Check the effect of seed mixes

The second factor is the seed mix. In the following three box plots, 0 indicates traditional seed mixes and 1 indicates plateau-tolerant pollinator seed mixes. Again the effects of this factor on three different months were put together for comparing purpose.



Plot 2.3 The effect of Seed Mixes

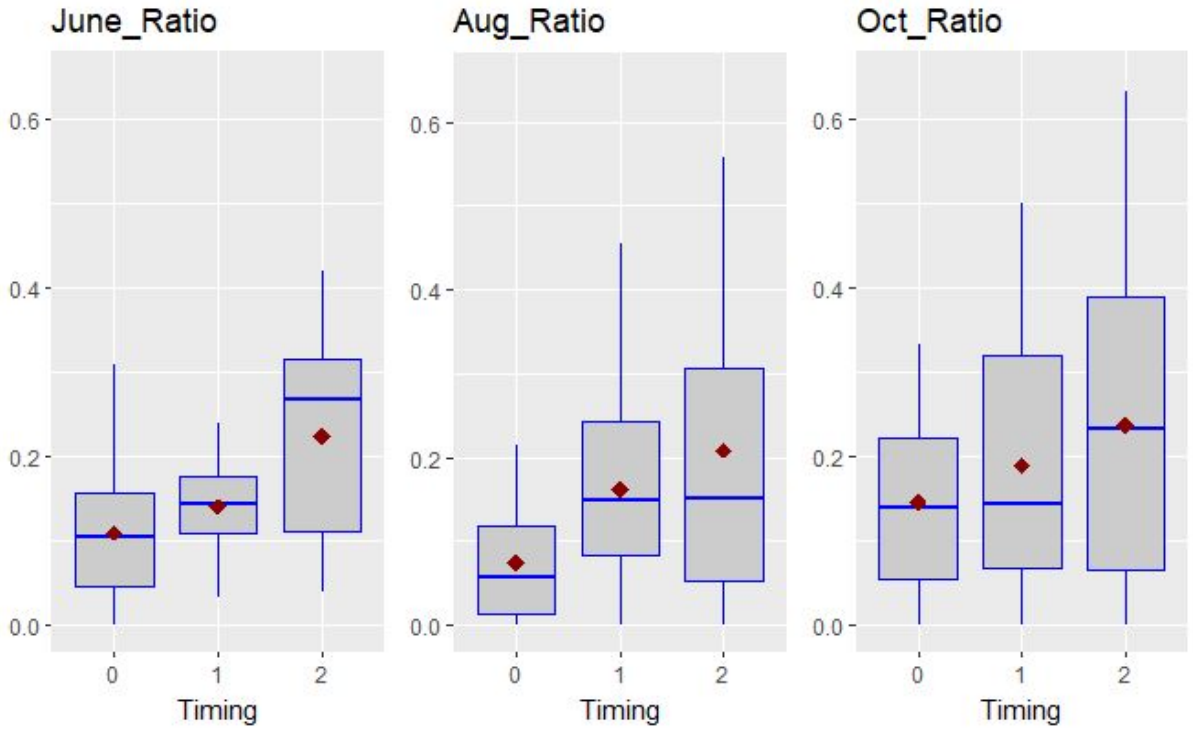
In the above three plots, although the mean and median ratio value of three different months on the same treatment level may be different from each other, they all got higher value at level 1 than at level 0, which means that the plateau-tolerant seed

mixes have a positive effect on the average ratio value compared to the traditional seed mixes. This result is the same as what we expected.

Another phenomena is that the difference of mean ratio between level 0 and 1 become larger as time goes by. In June, the difference is about 0.05 and in August, it grows to near 0.1, while in October it become 0.2. This suggests that the effect of plateau-tolerant seed mixes would become more obvious as time goes by.

2.2.3 Check the effect of timing

The third factor is the timing of application. In the following three box plots, 0 indicates no application, 1 indicates pre-emergent application and 2 indicates post-emergent application. Again the effects of this factor on three different months were put together for comparing purpose.



Plot 2.4 The effect of Timing

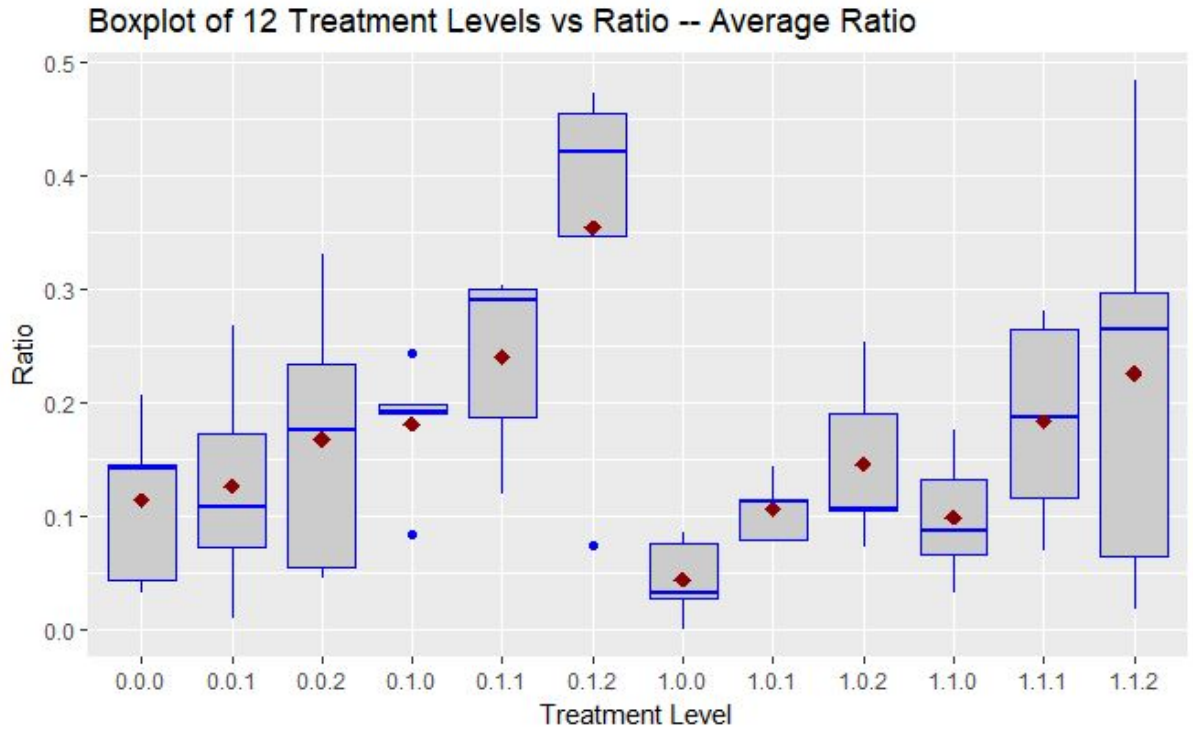
In the above three plots, we can easily notice that there's an increasing trend of ratio value from level 0 to level 1 and then level 2, which means that post-emergent applications has the best performance among three levels of timing factor.

2.3 Check the overall effect of three factors

To check the overall effect of three factors, the average ratio of June, August and October is calculated and is used to measure the performance of different treatment levels combination. The following plot shows the boxplot of 12 treatment levels combination and their meanings are explained in the table.

Treatment	**Stratification**	**Seed Mixes**	**Timing**
0.0.0	No stratification	Traditional	No application
0.0.1	No stratification	Traditional	Pre-emergent
0.0.2	No stratification	Traditional	Post-emergent
0.1.0	No stratification	Plateau-Tolerant	No application
0.1.1	No stratification	Plateau-Tolerant	Pre-emergent
0.1.2	No stratification	Plateau-Tolerant	Post-emergent
1.0.0	Cold stratification	Traditional	No application
1.0.1	Cold stratification	Traditional	Pre-emergent
1.0.2	Cold stratification	Traditional	Post-emergent
1.1.0	Cold stratification	Plateau-Tolerant	No application
1.1.1	Cold stratification	Plateau-Tolerant	Pre-emergent
1.1.2	Cold stratification	Plateau-Tolerant	Post-emergent

Table 2.1 The Explanation of Treatment level combinations



Plot 2.5 Boxplot of 12 Treatment Levels

In general, the level combination of 0.1.2 has a significantly higher ratio value than the other combinations, which means that no stratification, plateau-tolerant seed mixes and post-emergent application work together can generate more expecting species.

Considering that our original expectation was that the combination of 1.1.2 would have the highest ratio value, let's take a look at the 1.1.2 group. Although it has a lower mean ratio than the 0.1.2, we can also notice that the length of this box is much longer than the other groups, which mean that it has a large variance with some observations end up with really high ratio value while some don't have a good performance under this circumstance. This suggests that we may need more samples to further verify the effect of this treatment level combination.

2.4 Check the effect of soilhealth

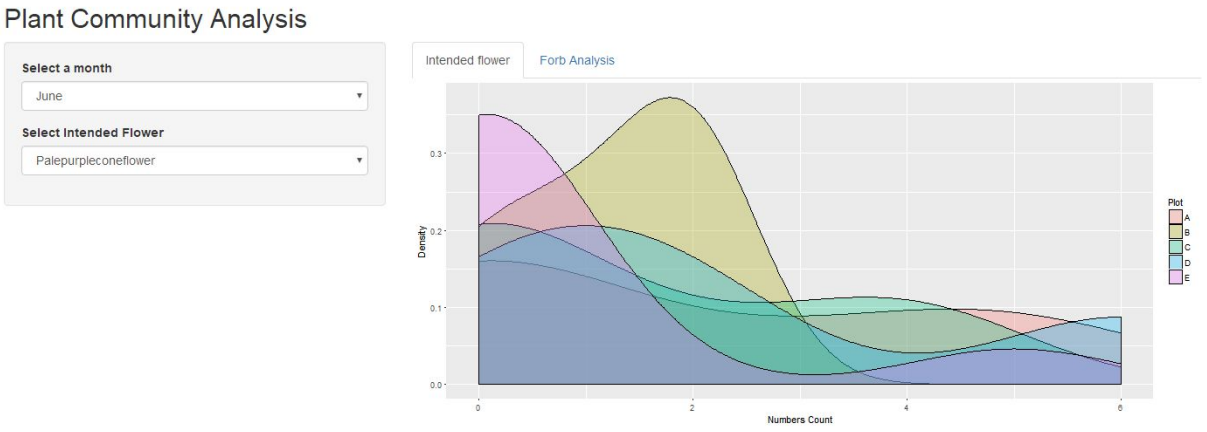
Since we already find that the variables related to the soil are correlated with each other, here we only choose one soil variable which is soilhealth to visualize the effect of it.

Plot 2.6 The effect of Soilhealth

The above plot clearly shows that the healthy group has the highest mean ratio while the disturbed group has the lowest mean ratio, which is not surprising at all since plants definitely prefer healthy soil than unhealthy ones.

2.5 Shiny

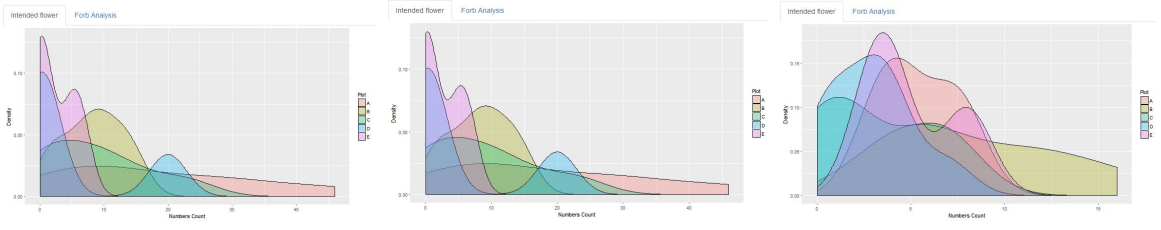
Apart from graphs above, we intend to have more direct knowledge about the database we have and potential problems that we may encounter. Thus, a shiny app was built up to realize these functions. Also, by sharing this app with our client, we are aimed to bring more convenience to her research.



Plot 2.7 Plant Community Analysis

As is shown in this graph, the ui was designed to realize two functions.It demonstrates density plot of different species in a specified month and overall distribution in different months and different plots. Several conclusions could be reached.

First, numbers of desired species change a lot in different months. Below are density plots of black-eyed susan in three months and they differ a lot. The same case happens to most of other exotic flowers and grass.



Plot 2.8 The Density Plot of Species

Second, it is unfortunate that most species did not grow so well as expected. Seen from result, we conclude that most intended flowers even did not germinate in most plots, or they might not have grown large enough to be noticed.

Third, three species make up the main part of desired species. They are black-eyed susan, purple coneflower and purple prairie clover. They consist of nearly 60% of the entire desired community.

Plot 2.9 Plant Community Anakysis

3 ANOVA Analysis

3.1 Steps of Multiple ANOVA analysis

The full model is built with main factors and all other possible interactions. Use the three-way factor effects model as an example. The full model is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

In this formula, μ is the overall mean; α_i, β_j, γ_k are the main effects on factors A, B and C; $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$ are the two-way (first order) interactions; $(\alpha\beta\gamma)_{ijk}$ is the three-way (second order) interactions.

Based on the full model, use Levene test to check the homogeneity of variance assumption and Shapiro-Wilk test on the ANOVA residuals to check the normality assumption.

If the assumptions are satisfied, according to the output ANOVA table, select significant items and rebuild a simpler model. Check the new output ANOVA table to repeat the previous step until all the remained items are significant. If there are multiple choices of models, use F-test to select the final model.
Based on the final model, use Tukey multiple pairwise-comparisons to compare the differences between the levels for each factor.

3.2 Multiple ANOVA in seperate months

3.2.1 ANOVA considering three designed factors

In this section, only the original three factors (*Stratified, Seedmix and Application*) are considered, leaving the extra information of soil for next section. Follow the steps described previously, the data in each separate month satisfies homogeneity and normality assumptions. More information of checking assumptions is included in Appendix. The final models are shown as follow:
June:

$$Ratio \sim Stratified + Seedmix + Application$$

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Stratified   1 0.0476 0.04762   6.998 0.010618 *
Seedmix      1 0.0419 0.04189   6.155 0.016189 *
Application  2 0.1405 0.07024  10.323 0.000156 ***
Residuals   55 0.3743 0.00680
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 3.1 The ANOVA Output of June Data

August:

Ratio ~ Stratified + Seedmix + Application + Stratified:Seedmix

```
                  Df Sum Sq Mean Sq F value  Pr(>F)
Stratified         1 0.0546 0.05456   4.027 0.04981 *
Seedmix            1 0.1096 0.10958   8.087 0.00628 **
Application        2 0.1872 0.09359   6.907 0.00213 **
Stratified:Seedmix 1 0.0706 0.07058   5.209 0.02644 *
Residuals         54 0.7317 0.01355
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 3.2 The ANOVA Output of August Data

October:

Ratio ~ Stratified + Seedmix

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Stratified   1 0.0804  0.0804   4.239   0.0441 *
Seedmix      1 0.3428  0.3428  18.063 8.01e-05 ***
Residuals   57 1.0817  0.0190
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 3.3 The ANOVA Output of October Data

From tables above, *Stratified* and *Seedmix* exist in all three models, meaning they have statistically significant main effects for every month. The Tukey multiple

pairwise-comparisons is used to check the differences between each level. The outputs are summarized as follow. The original Tukey multiple pairwise-comparisons tables are included in Appendix.

Factor Level	June		August		October	
	diff	p-value	diff	p-value	diff	p-value
Stratified 1-0	-0.0563	0.0106	-0.0603	0.0498	-0.0732	0.0398
Seedmix 1-0	0.0528	0.0161	0.0855	0.0063	0.1512	0.0001

Table 3.4 The Tukey Multiple Pairwise-comparisons Table

All the p-values in previous table are smaller than 0.05, which means the differences are statistically significant in all three months. The cold stratification has a negative effect on the response and the Plateau-Tolerant seedmix has a positive effect on the response. The effect of cold stratification does not match the expectation, as the cold stratification is thought to be helpful for seeds sprouting. The potential reason might be the timing of cold stratification matters in different periods of seeds. In this experiment, the timing cannot be controlled.

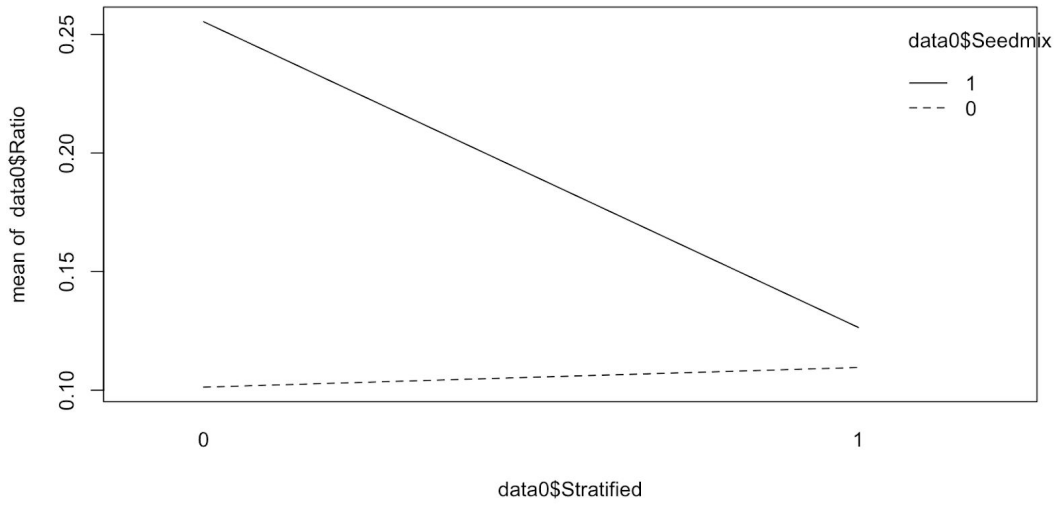
Another designed factor, Application, exist in models for June and August, but not in model for October. The potential reason might be the effects of seasons or other unknown factors. The Tukey multiple pairwise-comparisons outputs are summarized as follow.

Application Level	June		August		October	
	diff	p-value	diff	p-value	diff	p-value
1-0	0.0311	0.4623	0.0881	0.0520	0.0436	0.5653
2-0	0.1146	0.0002	0.1347	0.0016	0.0917	0.0886
2-1	0.0835	0.0064	0.0466	0.4204	0.0481	0.5000

Table 3.5 The Tukey Multiple Pairwise-comparisons Table

According to the table above, the differences between pre-emergent application and no application are not statistically significant under 0.05 significance level for all three months. The differences between post-emergent application and no application are statistically significant under 0.05 significance level for June and August. The differences between post-emergent application and pre-emergent application are statistically significant under 0.05 significance level only for June. The result varies in different month. The potential reasons might be the effects of seasons and the period of life cycle for seeds. The p-value of the difference between post-emergent application and no application is 0.0886, which we can consider as significant under 0.1 significance level. As a result, in general, the post-emergent application has a positive effect on ratio than no application, while there are not significant differences between other levels.

The model for August contains three main effects and an interaction between Stratified and Seedmix, while other models just contain main effects. If there were no interaction term, the response would be interpreted as the unique effect of each factor. But the interaction means that the effect of one factor on the response is different for different values of the other factor. So in August, the unique effect of Seedmix on Ratio also depends on the values of Stratified. The interaction plot gives a visualization of this effect.



Plot 3.1 The Interaction Plot of Seedmix and Stratified

In the above plot, the solid line represents the group with Plateau-Tolerant seedmix (group 1). The dot line represents the group with traditional seedmix (group 2). The solid line lays above the dot line, indicating the the mean ratio in group 1 is higher than group 2, no matter the levels in Stratified. If there were no interaction, these two lines should be parallel, meaning the difference in ratio is only depends on the level in *Seedmix*. The average of increase in ratio should be relatively same in both group without cold stratification and the group applied cold stratification. While in the plot above, these two lines are not parallel, indicating an interaction between *Stratified and Seedmix*. In group 1, the average increase in those without cold stratification is higher than those applied cold stratified, which means the cold stratification may limit the increase effect given by Plateau-Tolerant seedmix.

3.2.2 ANOVA considering three designed factors and two extra soil factors

In this section, not only the original three factors (*Stratified, Seedmix* and *Application)*, but also the two extra soil factors (*SoilHealth* and *NutrientRunoff*) are considered. Follow the steps described previously, only the final model in June changes, models for August and October remain same.

June:

Ratio ~ Stratified + Seedmix + Application + SoilHealth + NutrientRunoff + Application:NutrientRunoff

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Stratified	1	0.04762	0.04762	10.513	0.00218	**
Seedmix	1	0.04189	0.04189	9.247	0.00385	**
Application	2	0.14049	0.07024	15.508	6.73e-06	***
SoilHealth	2	0.05474	0.02737	6.043	0.00462	**
NutrientRunoff	2	0.04537	0.02269	5.008	0.01067	*
Application:NutrientRunoff	4	0.06126	0.01531	3.381	0.01641	*
Residuals	47	0.21290	0.00453			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 3.6 The ANOVA Output of June Data

From the table above, in June, both extra soil factors have significant main effects on the response, and the model contains an interaction between *Application* and *NutrientRunoff* under 0.05 significance level.

Even the information provided by the subjective soil factors can explain the variations in June in some degree. However, as the soil factors are not selected into models for other two months, in general, these two factors are not convinced to fully explained the variations among each treatment. The potential reason is the very small size, 5 for each treatment. Even in the same designed treatment, the actual treatment is not the same, other important factors related to the species, for example the soil nutrition and soil moisture, are not been controlled. As a result, a more controlled experiment with larger sample size is recommended.

3.3 Multiple ANOVA in entire data

In this section, the original three factors (*Stratified, Seedmix* and *Application*) and two soil factors(*SoilHealth* and *NutrientRunoff*) from three month are combined and considered. Follow the steps described previously, the combined dataset satisfies

homogeneity and normality assumptions. More information of checking assumptions is included in Appendix. The final models are shown as follow:

Ratio ~ Stratified + Seedmix + Application + SoilHealth + NutrientRunoff + Seedmix:Application + Seedmix:SoilHealth + SoilHealth:NutrientRunoff + Stratified:Application:SoilHealth + Stratified:Application:NutrientRunoff + Seedmix:Application:NutrientRunoff

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Stratified	1	0.1803	0.1803	21.107	8.90e-06	***
Seedmix	1	0.4190	0.4190	49.056	6.99e-11	***
Application	2	0.3879	0.1939	22.707	2.21e-09	***
SoilHealth	1	0.0227	0.0227	2.660	0.104936	
NutrientRunoff	1	0.1182	0.1182	13.840	0.000277	***
Seedmix:Application	2	0.0550	0.0275	3.220	0.042641	*
Seedmix:SoilHealth	1	0.0553	0.0553	6.477	0.011896	*
SoilHealth:NutrientRunoff	1	0.0836	0.0836	9.790	0.002094	**
Stratified:Application:SoilHealth	5	0.1120	0.0224	2.622	0.026273	*
Stratified:Application:NutrientRunoff	5	0.4638	0.0928	10.861	5.59e-09	***
Seedmix:Application:NutrientRunoff	3	0.0924	0.0308	3.607	0.014800	*
Residuals	156	1.3324	0.0085			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 3.6 The ANOVA Output of the Entire Data

From the table above, in the combined dataset, all the three original factors (Stratified, *Seedmix* and *Application*) and *NutrientRunoff* have significant main effects on the response under 0.05 significance level. Also, the final model includes 6 interaction terms, which are *Seedmix:Application, Seedmix:SoilHealth, SoilHealth : NutrientRunoff, Stratified:Application:SoilHealth,Stratified : Application : NutrientRunoff, Seedmix:Application:NutrientRunoff*. All of the 6 interaction terms are significant under 0.05 significance level.

4 Conclusion

In this section, we give short answers to the questions that this experiment study tries to understand.

1. Two treatments: Does cold stratification contribute to the success of native plant establishment?

Answer: The cold stratification did not performs well compared to no stratification.

2. Two seed mixes: How would a traditional “pollinator habitat” seed mix perform with imazapic application?

Answer: The plateau-tolerant pollinator seed mixes perform better than the traditional seed mixes.

3. Timing: How will plants respond to pre-emergent application of imazapic versus post-emergent application?

Answer: The post-emergent application outperforms the pre-emergent application and the two perform better than no application.

5 Appendix

5.1 More tables of results

5.1.1 Checking the assumptions of ANOVA for each month

June:

```
> leveneTest(Ratio ~ Stratified * Seedmix * Application, data = data0)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 11  1.1439 0.3502
      48
> aov_residuals <- residuals(object = m02)
> shapiro.test(x = aov_residuals)

        Shapiro-Wilk normality test

data:  aov_residuals
W = 0.98912, p-value = 0.8712
```

August:

```
> leveneTest(Ratio ~ Stratified * Seedmix * Application, data = data0)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 11   0.929 0.5213
      48
> aov_residuals <- residuals(object = m02)
> shapiro.test(x = aov_residuals)

        Shapiro-Wilk normality test

data:  aov_residuals
W = 0.98542, p-value = 0.6916
```

October:

```
> leveneTest(Ratio ~ Stratified * Seedmix * Application, data = data0)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 11  1.6283 0.1209
      48
> aov_residuals <- residuals(object = m02)
> shapiro.test(x = aov_residuals)

        Shapiro-Wilk normality test

data:  aov_residuals
W = 0.97194, p-value = 0.1814
```

As p-values above of both tests for each month are greater than 0.05, the null hypothesis cannot be rejected. It's reasonable to assume the homogeneity of variances in the different treatment groups and the normality of the data.

5.1.2 Tukey multiple pairwise-comparisons for each month

June:

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Ratio ~ Stratified + Seedmix + Application, data = data0)

$Stratified
           diff         lwr         upr     p adj
1-0 -0.05634412 -0.09902871 -0.01365952 0.0106176

$Seedmix
          diff        lwr       upr     p adj
1-0 0.05284281 0.01015821 0.0955274 0.0161893

$Application
          diff         lwr        upr     p adj
1-0 0.03112160 -0.03171336 0.09395655 0.4623355
2-0 0.11460750  0.05177255 0.17744246 0.0001501
2-1 0.08348591  0.02065095 0.14632086 0.0063535
```

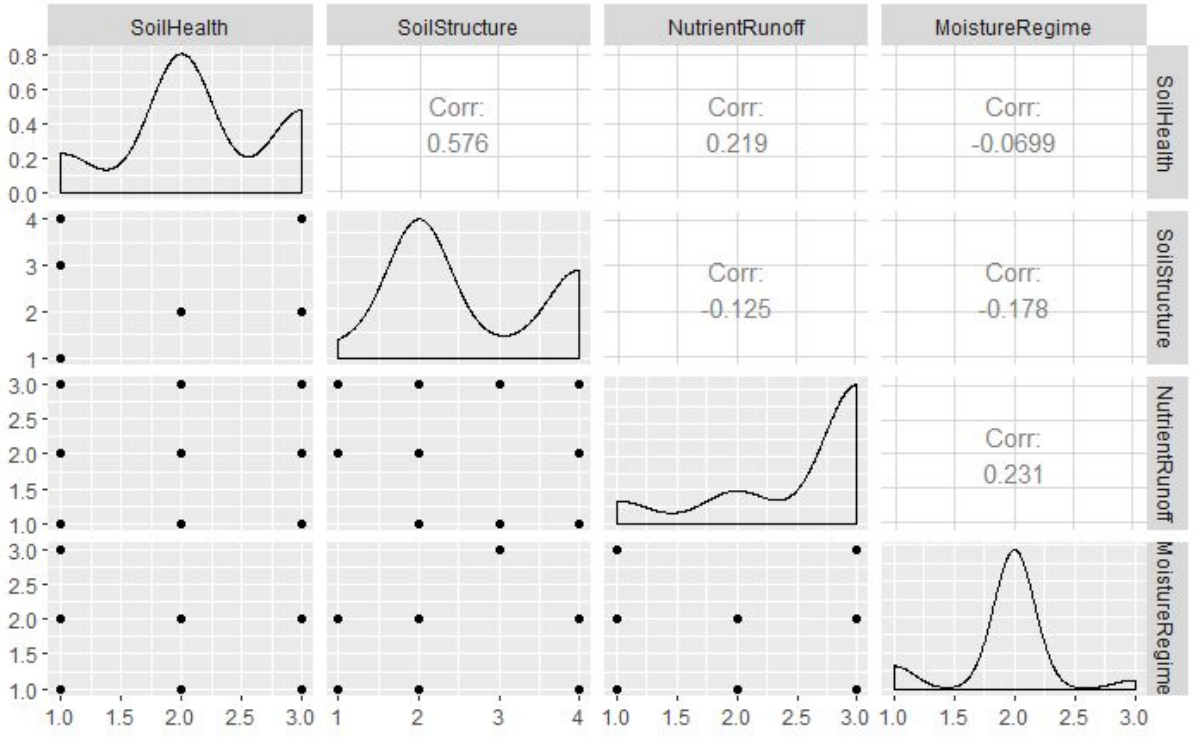
August:

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Ratio ~ Stratified + Seedmix + Application + Stratified:Seedmix, data = d
ata0)

$Stratified
           diff        lwr           upr     p adj
1-0 -0.06030905 -0.1205653 -5.278164e-05 0.0498067

$Seedmix
          diff        lwr       upr     p adj
1-0 0.08547045 0.02521418 0.1457267 0.0062824

$Application
          diff           lwr       upr     p adj
1-0 0.08809792 -0.0006123282 0.1768082 0.0519727
2-0 0.13469933  0.0459890831 0.2234096 0.0016427
2-1 0.04660141 -0.0421088380 0.1353117 0.4203506
```

October:

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Ratio ~ Stratified + Seedmix + Application, data = data0)

$Stratified
            diff        lwr           upr     p adj
1-0 -0.07322991 -0.1429155 -0.003544302 0.0397852

$Seedmix
          diff        lwr       upr p adj
1-0 0.1511696 0.08148403 0.2208552 6e-05

$Application
           diff         lwr       upr     p adj
1-0 0.04358933 -0.05899316 0.1461718 0.5653335
2-0 0.09170687 -0.01087562 0.1942894 0.0885589
2-1 0.04811754 -0.05446495 0.1507000 0.5000166
```

5.1.3 Checking the correlation of the four soil condition variables



From the table above, the high correlation between “SoilHealth” and “SoilStructure” can be seen.

5.2 R code

Data Preprocessing

```
library(readxl)
data_Jun = read_excel("Fulldata.xlsx")
data_Aug = read_excel("Fulldata.xlsx", sheet = 2)
data_Oct = read_excel("Fulldata.xlsx", sheet = 3)
ratio_Jun <- data.frame((data_Jun$Forbs+data_Jun$Grass)/(data_Jun$Forbs+data_Jun$Grass+data_Jun$Weeds))
ratio_Aug <- (data_Aug$Forbs+data_Aug$Grass)/(data_Aug$Forbs+data_Aug$Grass+data_Aug$Weeds)
ratio_Oct <- (data_Oct$Forbs+data_Oct$Grass)/(data_Oct$Forbs+data_Oct$Grass+data_Oct$Weeds)

# check the na, decided to set to 0
ratio_Jun[is.na(ratio_Jun)] <- 0
ratio_Aug[is.na(ratio_Aug)] <- 0
ratio_Oct[is.na(ratio_Oct)] <- 0

# set to different treatments to factors
data_Jun$Stratified = as.factor(data_Jun$Stratified)
data_Jun$Seedmix = as.factor(data_Jun$Seedmix)
data_Jun$Application = as.factor(data_Jun$Application)
data_Jun$SoilHealth=as.numeric(as.factor(data_Jun$SoilHealth))
data_Jun$NutrientRunoff=as.numeric(as.factor(data_Jun$NutrientRunoff))
#data_Jun$Date = as.factor(matrix(0,nrow=60,ncol=1))
data_Aug$Stratified = as.factor(data_Aug$Stratified)
data_Aug$Seedmix = as.factor(data_Aug$Seedmix)
data_Aug$Application = as.factor(data_Aug$Application)
data_Aug$SoilHealth=as.numeric(as.factor(data_Aug$SoilHealth))
data_Aug$NutrientRunoff=as.numeric(as.factor(data_Aug$NutrientRunoff))
#data_Aug$Date = as.factor(matrix(1,nrow=60,ncol=1))
data_Oct$Stratified = as.factor(data_Oct$Stratified)
data_Oct$Seedmix = as.factor(data_Oct$Seedmix)
data_Oct$Application = as.factor(data_Oct$Application)
#data_Oct$Date = as.factor(matrix(2,nrow=60,ncol=1))
data_Oct$SoilHealth=as.numeric(as.factor(data_Oct$SoilHealth))
#data_Oct$SoilStructure=as.numeric(as.factor(data_Oct$SoilStructure))
data_Oct$NutrientRunoff=as.numeric(as.factor(data_Oct$NutrientRunoff))
#data_Oct$MoistureRegime=as.numeric(as.factor(data_Oct$MoistureRegime))

DATA_Jun <- data.frame(ratio_Jun,data_Jun$Stratified,data_Jun$Seedmix,data_Jun$Application,data_Jun$Soil
DATA_Aug <- data.frame(ratio_Aug,data_Aug$Stratified,data_Aug$Seedmix,data_Aug$Application,data_Aug$Soil
DATA_Oct <- data.frame(ratio_Oct,data_Oct$Stratified,data_Oct$Seedmix,data_Oct$Application,data_Oct$Soil
colnames(DATA_Jun) <- c("ratio","Stratified",'Seedmix',"Application","SoilHealth","NutrientRunoff")
colnames(DATA_Aug) <- c("ratio","Stratified",'Seedmix',"Application","SoilHealth","NutrientRunoff")
colnames(DATA_Oct) <- c("ratio","Stratified",'Seedmix',"Application","SoilHealth","NutrientRunoff")
DATA_TOTAL <- rbind(DATA_Jun,DATA_Aug,DATA_Oct)
```

Visualization

```
#read data
library(readxl)
library(ggplot2)
library("ggpubr")
data_Jun = read_excel("Fulldata.xlsx")
data_Aug = read_excel("Fulldata.xlsx", sheet = 2)
```

```
data_Oct = read_excel("Fulldata.xlsx", sheet = 3)
```

```
#compute the ratio
ratio_Jun = (data_Jun$Forbs + data_Jun$Grass) / (data_Jun$Forbs + data_Jun$Grass + data_Jun$Weeds)
ratio_Aug = (data_Aug$Forbs + data_Aug$Grass) / (data_Aug$Forbs + data_Aug$Grass +
data_Aug$Weeds)
ratio_Oct = (data_Oct$Forbs + data_Oct$Grass) / (data_Oct$Forbs + data_Oct$Grass +
data_Oct$Weeds)
par(mfrow = c(1, 3))
hist(ratio_Jun, col = 'darkorange', ylim = c(0, 22), breaks = 7)
hist(ratio_Aug, col = 'darkorange', ylim = c(0, 22), breaks = 7)
hist(ratio_Oct, col = 'darkorange', ylim = c(0, 22), breaks = 7)
```

```
#generate the new dataset
Jun_data = data.frame(data_Jun$Plot, data_Jun$Stratify, data_Jun$Mix, data_Jun$Timing, data_Jun$SoilHeal
Aug_data = data.frame(data_Aug$Plot, data_Aug$Stratify, data_Aug$Mix, data_Aug$Timing, data_Jun$SoilHeal
Oct_data = data.frame(data_Oct$Plot, data_Oct$Stratify, data_Oct$Mix, data_Oct$Timing, data_Jun$SoilHeal
```

```
colnames(Jun_data)  = c("Plot", "Stratify", "Mix", "Timing", "Soilhealth", "Jun_ratio")
colnames(Aug_data)  = c("Plot", "Stratify", "Mix", "Timing", "Soilhealth", "Aug_ratio")
colnames(Oct_data)  = c("Plot", "Stratify", "Mix", "Timing", "Soilhealth", "Oct_ratio")
```

```
A = ggplot(data = Jun_data, aes(Jun_data$Jun_ratio)) +
  geom_histogram(breaks = seq(0, 0.6, by = 0.05),
                 col = "red",
                 fill = "green",
                 alpha = .2) +
  labs(title = "Ratio of June") +
  labs(x = "Ratio", y = "Frequency") +
  xlim(c(0,0.6)) +
  ylim(c(0,19)) + theme(axis.title.y = element_blank())

B = ggplot(data = Aug_data, aes(Aug_data$Aug_ratio)) +
  geom_histogram(breaks = seq(0, 0.6, by = 0.05),
                 col = "red",
                 fill = "green",
                 alpha = .2) +
  labs(title = "Ratio of August") +
  labs(x = "Ratio", y = "Frequency") +
  xlim(c(0,0.6)) +
  ylim(c(0,19)) + theme(axis.title.y = element_blank())

C = ggplot(data = Oct_data, aes(Oct_data$Oct_ratio)) +
  geom_histogram(breaks = seq(0, 0.6, by = 0.05),
                 col = "red",
                 fill = "green",
                 alpha = .2) +
  labs(title = "Ratio of October") +
  labs(x = "Ratio", y = "Frequency") +
  xlim(c(0,0.6)) +
  ylim(c(0,19)) + theme(axis.title.y = element_blank())

figure = ggarrange(A, B, C, ncol = 3, nrow = 1)
```

```
figure
```

```
total1 = merge(Jun_data, Aug_data, by = c("Plot", "Stratify", "Mix", "Timing", "Soilhealth"))
total = merge(total1, Oct_data, by = c("Plot", "Stratify", "Mix", "Timing", "Soilhealth"))
```

```
#convert the data type of first 4 columns into factor
total$Plot = as.factor(total$Plot)
total$Stratify = as.factor(total$Stratify)
total$Mix = as.factor(total$Mix)
total$Timing = as.factor(total$Timing)
total$lev_name = paste(as.character(total$Stratify), as.character(total$Mix), as.character(total$Timing
```

```
#replace all NAs with 0
total[is.na(total)] <- 0
```

```
#compute the correlation between the ratio of three different months
cor_ratio = cor(total[, 6:8], total[, 6:8], method = "pearson")
cor_ratio
```

```
total$avg_ratio = rowMeans(total[, 6:8])
```

```
ggplot(total, aes(x = Soilhealth, y = avg_ratio)) +
    geom_boxplot(fill = "grey80", colour = "blue") +
    scale_x_discrete() + xlab("Soil Condition") +
    labs(title = "The Effect of Soilhealth") +
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18, size = 3,show_guide = FA
```

```
A = ggplot(total, aes(x = Stratify, y = Jun_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + s

B = ggplot(total, aes(x = Stratify, y = Aug_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + s


C = ggplot(total, aes(x = Stratify, y = Oct_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + s

figure = ggarrange(A, B, C, ncol = 3, nrow = 1)

figure
```

```
A = ggplot(total, aes(x = Mix, y = Jun_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + scale_

B = ggplot(total, aes(x = Mix, y = Aug_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + scale_


C = ggplot(total, aes(x = Mix, y = Oct_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + scale_

figure = ggarrange(A, B, C, ncol = 3, nrow = 1)

figure
```

```
A = ggplot(total, aes(x = Timing, y = Jun_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + sca
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18, size = 3,show_guide = FA

B = ggplot(total, aes(x = Timing, y = Aug_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + sca
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18, size = 3,show_guide = FA
```

```
C = ggplot(total, aes(x = Timing, y = Oct_ratio)) + geom_boxplot(fill = "grey80", colour = "blue") + sca
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18, size = 3,show_guide = FAl

figure = ggarrange(A, B, C, ncol = 3, nrow = 1)

figure
```

```
#plot the ratio of 12 different treatment level
ggplot(total, aes(x = lev_name, y = Jun_ratio)) +
    geom_boxplot(fill = "grey80", colour = "blue") +
    scale_x_discrete() + xlab("Treatment Level") + ylab("Ratio") +
    labs(title = "Boxplot of 12 Treatment Levels vs Ratio -- June") +
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18, size = 3,show_guide = FAl

ggplot(total, aes(x = lev_name, y = Aug_ratio)) +
    geom_boxplot(fill = "grey80", colour = "blue") +
    scale_x_discrete() + xlab("Treatment Level") + ylab("Ratio") +
    labs(title = "Boxplot of 12 Treatment Levels vs Ratio -- August") +
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18, size = 3,show_guide = FAl

ggplot(total, aes(x = lev_name, y = Oct_ratio)) +
    geom_boxplot(fill = "grey80", colour = "blue") +
    scale_x_discrete() + xlab("Treatment Level") + ylab("Ratio") +
    labs(title = "Boxplot of 12 Treatment Levels vs Ratio -- October") +
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18, size = 3,show_guide = FAl

ggplot(total, aes(x = lev_name, y = avg_ratio)) +
    geom_boxplot(fill = "grey80", colour = "blue") +
    scale_x_discrete() + xlab("Treatment Level") + ylab("Ratio") +
    labs(title = "Boxplot of 12 Treatment Levels vs Ratio -- Average Ratio") +
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18, size = 3,show_guide = FAl
```

```
#Plot the interaction plot- use June as an example
with(total, {
    interaction.plot( Stratify, Mix, avg_ratio, type = "b", legend = "T", ylab = "ratio", main = "Intera
    interaction.plot( Stratify, Timing, avg_ratio, type = "b", legend = "T", ylab = "ratio", main = "In
    interaction.plot( Mix, Timing, avg_ratio, type = "b", legend = "T", ylab = "ratio", main = "Interac
)
```

ANOVA Analysis

June Data

```
raw_data = read_excel("Fulldata.xlsx")

# data0 just consider 3 factor, leave the soil(plot) condition
data0 = raw_data[, c('Forbs', 'Grass', 'Weeds', 'Stratified', 'Seedmix', 'Application')]
# use the ratio of forbs + grass to weeds as the target
# data0[, "Ratio"] = (data0[, "Forbs"] + data0[, "Grass"]) / data0[, "Weeds"]
data0[, "Ratio"] = (data0[, "Forbs"] + data0[, "Grass"]) /
  (data0[, "Forbs"] + data0[, "Grass"] + data0[, "Weeds"])

# check the na, decided to set to 0 (denominator, Weeds, is 0)
```

```r
data0[, "Ratio"][is.na(data0[, "Ratio"])] <- 0
# set to different treatments to factors
data0$Stratified = as.factor(data0$Stratified)
data0$Seedmix = as.factor(data0$Seedmix)
data0$Application = as.factor(data0$Application )

par(mfrow=c(1,1))
# Plot the mean of Ratio for the different factors levels
plot.design(Ratio ~ Stratified + Seedmix + Application, data = data0)
par(mfrow=c(3,1))
# Plot the mean of Ratio for two-way combinations of factors
interaction.plot(data0$Stratified, data0$Seedmix, data0$Ratio)
interaction.plot(data0$Stratified, data0$Application, data0$Ratio)
interaction.plot(data0$Application, data0$Seedmix, data0$Ratio)

library(dplyr)
group_by(data0, Stratified, Seedmix, Application) %>%
  summarise(
    count = n(),
    mean = mean(Ratio, na.rm = TRUE),
    sd = sd(Ratio, na.rm = TRUE)
  )

# Generate the full model
m01 <- aov(Ratio ~ Stratified * Seedmix * Application, data = data0)
summary(m01)
# Just keep significant items
m02 <- aov(Ratio ~ Stratified + Seedmix + Application, data = data0)
summary(m02)
# Comparison between nested models
anova(m01, m02)
# Residual plots
par(mfrow=c(2,4))
plot(m01)
plot(m02)

library(car)
leveneTest(Ratio ~ Stratified * Seedmix * Application, data = data0)
# From the output above we can see that the p-value is not less than the significance level of 0.05.
# This means that there is no evidence to suggest that the variance across groups is statistically sign
# Therefore, we can assume the homogeneity of variances in the different treatment groups.

# the Shapiro-Wilk test on the ANOVA residuals
# Extractthe residuals
aov_residuals <- residuals(object = m02)
# Run Shapiro-Wilk test
shapiro.test(x = aov_residuals)
# No indication that normality is violated.

# Tukey multiple pairwise-comparisons
TukeyHSD(m02, which = c("Stratified", "Seedmix", "Application"))

data1 = data0
```

```
data1[, c('SoilHealth', 'SoilStructure', 'NutrientRunoff', 'MoistureRegime')] =
  raw_data[, c('SoilHealth', 'SoilStructure', 'NutrientRunoff', 'MoistureRegime')]
summary(data1)
levels(data1$NutrientRunoff)[1] = 'Nor'
summary(data1)

par(mfrow=c(1,1))
plot.design(Ratio ~ Stratified + Seedmix + Application + SoilHealth + SoilStructure + NutrientRunoff + M

m11 <- aov(Ratio ~ Stratified * Seedmix * Application
           * SoilHealth * NutrientRunoff, data = data1)
summary(m11)

m12 <- aov(Ratio ~ Stratified + Seedmix + Application + SoilHealth + NutrientRunoff + Application:Nutri
summary(m12)


chisq.test(data1$SoilStructure, data1$SoilHealth) # high correlation
chisq.test(data1$NutrientRunoff, data1$SoilHealth)
chisq.test(data1$MoistureRegime, data1$SoilHealth) # high correlation
chisq.test(data1$MoistureRegime, data1$SoilStructure) # high

m12 <- aov(Ratio ~ Stratified * Seedmix * Application
           * SoilHealth * NutrientRunoff, data = data1)
summary(m12)

m13 <- aov(Ratio ~ Stratified + Seedmix + Application
           + SoilHealth + NutrientRunoff + Application:NutrientRunoff, data = data1)
summary(m13) # all sig.

m14 <- aov(Ratio ~ Stratified * Seedmix * Application, data = data1)
summary(m14)

m15 <- aov(Ratio ~ Stratified + Seedmix + Application, data = data1)
summary(m15) # all sig.

anova(m13, m15)
# 0.0004554, cannot ignore those intems, m13 is better.
```

August Data

```
raw_data = read_excel("Fulldata.xlsx", sheet = 2)

# data0 just consider 3 factor, leave the soil(plot) condition
data0 = raw_data[, c('Forbs', 'Grass', 'Weeds', 'Stratified', 'Seedmix', 'Application')]
# use the ratio of forbs + grass to weeds as the target
# data0[, "Ratio"] = (data0[, "Forbs"] + data0[, "Grass"]) / data0[, "Weeds"]
data0[, "Ratio"] = (data0[, "Forbs"] + data0[, "Grass"]) /
  (data0[, "Forbs"] + data0[, "Grass"] + data0[, "Weeds"])

# check the na, decided to set to 0 (denominator, Weeds, is 0)
data0[, "Ratio"][is.na(data0[, "Ratio"])] <- 0
```

```r
# set to different treatments to factors
data0$Stratified = as.factor(data0$Stratified)
data0$Seedmix = as.factor(data0$Seedmix)
data0$Application = as.factor(data0$Application )

par(mfrow=c(1,1))
# Plot the mean of Ratio for the different factors levels
plot.design(Ratio ~ Stratified + Seedmix + Application, data = data0)
par(mfrow=c(1,1))
# Plot the mean of Ratio for two-way combinations of factors
interaction.plot(data0$Stratified, data0$Seedmix, data0$Ratio)
interaction.plot(data0$Stratified, data0$Application, data0$Ratio)
interaction.plot(data0$Application, data0$Seedmix, data0$Ratio)

# Summary Statistics
library(dplyr)
group_by(data0, Stratified, Seedmix, Application) %>%
  summarise(
    count = n(),
    mean = mean(Ratio, na.rm = TRUE),
    sd = sd(Ratio, na.rm = TRUE)
  )

# Generate the full model
m01 <- aov(Ratio ~ Stratified * Seedmix * Application, data = data0)
summary(m01)
# Just keep significant items
m02 <- aov(Ratio ~ Stratified + Seedmix + Application + Stratified:Seedmix, data = data0)
summary(m02)

m03 <- aov(Ratio ~ Stratified, data = data0)
summary(m03)

# Comparison between nested models
anova(m01, m02)
# Residual plots
par(mfrow=c(2,4))
plot(m01)
plot(m02)

## Problem! ANOVA assumes that the data are normally distributed
## and the variance across groups are homogeneous...
## Remove outliers??

# Check the homogeneity of variance assumption
library(car)
leveneTest(Ratio ~ Stratified * Seedmix * Application, data = data0)
# From the output above we can see that the p-value is not less than the significance level of 0.05.
# This means that there is no evidence to suggest that the variance across groups is statistically sign
# Therefore, we can assume the homogeneity of variances in the different treatment groups.

# the Shapiro-Wilk test on the ANOVA residuals
# Extractthe residuals
```

```r
aov_residuals <- residuals(object = m02)
# Run Shapiro-Wilk test
shapiro.test(x = aov_residuals)
# No indication that normality is violated.

# Tukey multiple pairwise-comparisons
TukeyHSD(m02, which = c("Stratified", "Seedmix", "Application"))

# Add information of soil condition
data1 = data0
data1[, c('SoilHealth', 'SoilStructure', 'NutrientRunoff', 'MoistureRegime')] =
  raw_data[, c('SoilHealth', 'SoilStructure', 'NutrientRunoff', 'MoistureRegime')]
summary(data1)
levels(data1$NutrientRunoff)[1] = 'Nor'
summary(data1)

par(mfrow=c(1,1))
plot.design(Ratio ~ Stratified + Seedmix + Application + SoilHealth + SoilStructure + NutrientRunoff + M


m11 <- aov(Ratio ~ Stratified * Seedmix * Application
           * SoilHealth * NutrientRunoff , data = data1)
summary(m11)

m12 <- aov(Ratio ~ Stratified + Seedmix + Application + SoilHealth
           + Stratified:Seedmix, data = data1)
summary(m12)

m13 <- aov(Ratio ~ Stratified + Seedmix + Application + Stratified:Seedmix, data = data1)
summary(m13)
anova(m12, m13)


m15 <- aov(Ratio ~ Seedmix * Application, data = data1)
summary(m15)
m16 <- aov(Ratio ~ Seedmix + Application, data = data1)
summary(m16)
m17 <- aov(Ratio ~ Stratified + Seedmix + Application, data = data1)
summary(m17)
# Stratified not sig.

m19 <- aov(Ratio ~ Stratified + Seedmix + Application
         + SoilHealth + Stratified:Seedmix, data = data1)
summary(m19)

anova(m02, m16)
# m02 better


par(mfrow=c(2,2))
plot(m11)
m12 <- aov(Ratio ~ Stratified + Seedmix + Application + SoilHealth
           +Stratified:Seedmix + Stratified:Application, data = data1)
```

```r
summary(m12)
plot(m12)
# here main 3 are sig.

chisq.test(data1$Stratified, data1$SoilHealth)
chisq.test(data1$Seedmix, data1$SoilHealth)
chisq.test(data1$Application, data1$SoilHealth)

chisq.test(data1$Stratified, data1$Seedmix)
chisq.test(data1$Stratified, data1$Application)
chisq.test(data1$Seedmix, data1$Application)

chisq.test(data1$SoilStructure, data1$SoilHealth)
# high correlation
chisq.test(data1$NutrientRunoff, data1$SoilHealth) # 0.07
chisq.test(data1$MoistureRegime, data1$SoilHealth)
# high corrlation



m18 <- aov(Ratio ~ Stratified * Seedmix * Application
         * SoilHealth * NutrientRunoff , data = data1)
summary(m18)

m13 <- aov(Ratio ~ Stratified * Seedmix * Application * SoilHealth, data = data1)
summary(m13)

m14 <- aov(Ratio ~ Stratified * Seedmix * Application, data = data1)
summary(m14)


dataSeed1 = data1[which(data1$Seedmix == 1) ,]
m20 = aov(Ratio ~ Stratified * Application
          * SoilHealth * NutrientRunoff, data = data1)
summary(m20)

dataSeed1 = data1[which(data1$Seedmix == 0) ,]
m21 = aov(Ratio ~ Stratified + Application, data = data1)
summary(m21)
```

October Data

```r
raw_data = read_excel("Fulldata.xlsx", sheet = 3)

# data0 just consider 3 factor, leave the soil(plot) condition
data0 = raw_data[, c('Forbs', 'Grass', 'Weeds', 'Stratified', 'Seedmix', 'Application')]
# use the ratio of forbs + grass to weeds as the target
# data0[, "Ratio"] = (data0[, "Forbs"] + data0[, "Grass"]) / data0[, "Weeds"]
data0[, "Ratio"] = (data0[, "Forbs"] + data0[, "Grass"]) /
  (data0[, "Forbs"] + data0[, "Grass"] + data0[, "Weeds"])

# check the na, decided to set to 0 (denominator, Weeds, is 0)
```

```r
data0[, "Ratio"][is.na(data0[, "Ratio"])] <- 0
# set to different treatments to factors
data0$Stratified = as.factor(data0$Stratified)
data0$Seedmix = as.factor(data0$Seedmix)
data0$Application = as.factor(data0$Application )

par(mfrow=c(1,1))
# Plot the mean of Ratio for the different factors levels
plot.design(Ratio ~ Stratified + Seedmix + Application, data = data0)
par(mfrow=c(3,1))
# Plot the mean of Ratio for two-way combinations of factors
interaction.plot(data0$Stratified, data0$Seedmix, data0$Ratio)
interaction.plot(data0$Stratified, data0$Application, data0$Ratio)
interaction.plot(data0$Application, data0$Seedmix, data0$Ratio)

-library(dplyr)
group_by(data0, Stratified, Seedmix, Application) %>%
  summarise(
    count = n(),
    mean = mean(Ratio, na.rm = TRUE),
    sd = sd(Ratio, na.rm = TRUE)
  )

# Generate the full model
m01 <- aov(Ratio ~ Stratified * Seedmix * Application, data = data0)
summary(m01)
# Just keep significant items
m02 <- aov(Ratio ~ Stratified + Seedmix, data = data0)
summary(m02)
anova(m01, m02)

m03 <- aov(Ratio ~ Stratified + Seedmix + Application, data = data0)
summary(m03)
TukeyHSD(m03, which = c("Stratified", "Seedmix", "Application"))
# Comparison between nested models
anova(m03, m02)
# Residual plots
par(mfrow=c(2,4))
plot(m01)
plot(m02)

library(car)
leveneTest(Ratio ~ Stratified * Seedmix * Application, data = data0)
# From the output above we can see that the p-value is not less than the significance level of 0.05.
# This means that there is no evidence to suggest that the variance across groups is statistically sign
# Therefore, we can assume the homogeneity of variances in the different treatment groups.

# the Shapiro-Wilk test on the ANOVA residuals
# Extractthe residuals
aov_residuals <- residuals(object = m02)
# Run Shapiro-Wilk test
shapiro.test(x = aov_residuals)
# No indication that normality is violated.
```

```r
# Tukey multiple pairwise-comparisons
TukeyHSD(m02, which = c("Stratified", "Seedmix"))
TukeyHSD(m03, which = c("Stratified", "Seedmix", "Application"))

data1 = data0
data1[, c('SoilHealth', 'SoilStructure', 'NutrientRunoff', 'MoistureRegime')] =
  raw_data[, c('SoilHealth', 'SoilStructure', 'NutrientRunoff', 'MoistureRegime')]
summary(data1)
levels(data1$NutrientRunoff)[1] = 'Nor'
summary(data1)

par(mfrow=c(1,1))
plot.design(Ratio ~ Stratified + Seedmix + Application + SoilHealth + SoilStructure + NutrientRunoff + I




m11 <- aov(Ratio ~ Stratified * Seedmix * Application
           * SoilHealth * NutrientRunoff, data = data1)
summary(m11)

m12 <- aov(Ratio ~ Stratified + Seedmix + Application
           + SoilHealth + Seedmix:SoilHealth, data = data1)
summary(m12)

m13 <- aov(Ratio ~ Stratified + Seedmix + SoilHealth, data = data1)
summary(m13)

m14 <- aov(Ratio ~ Stratified + Seedmix, data = data1)
summary(m14)


dataSeed1 = data1[which(data1$Seedmix == 1) ,]
m20 = aov(Ratio ~ Stratified * Application
          * SoilHealth * NutrientRunoff, data = data1)
summary(m20)

dataSeed1 = data1[which(data1$Seedmix == 0) ,]
m21 = aov(Ratio ~ Stratified + Application, data = data1)
summary(m21)
```

Full Data

```r
# Generate the full model for three month
m01 <- aov(ratio ~ Stratified * Seedmix * Application * SoilHealth*NutrientRunoff, data = DATA_TOTAL)
summary(m01)
```

```r
# Just keep significant items(Stratified, Seedmix,Application,SoilHealth,NutrientRunoff,Seedmix:Applica
m02 <- aov(ratio ~ Stratified+Seedmix+Application+SoilHealth+NutrientRunoff+Seedmix:Application+Seedmix
summary(m02)

m03 <- aov(ratio ~ Stratified+Seedmix+Application+SoilHealth+NutrientRunoff+Seedmix:Application+Seedmix
```

```r
summary(m03)
```

```r
# Comparison between nested models
anova(m01, m03)
```

Shiny

```r
library(shiny)
library(ggplot2)
library(grid)
library(gridExtra)
library(reshape2)

r_colors <- rgb(t(col2rgb(colors()) / 255))
names(r_colors) <- colors()

# Define UI for data upload app ----
ui <- fluidPage(

  # App title ----
  titlePanel("Plant Community Analysis"),

  # Sidebar layout with input and output definitions ----
  sidebarLayout(

    # Sidebar panel for inputs ----
    sidebarPanel(

      selectInput(
        'month', 'Select a month',
        choices = c("June"=6,
                    "August"=8,
                    "October"=10),
        selectize = FALSE
      ),

      selectInput(
        'flo', 'Select Intended Flower',
        choices = c("Palepurpleconeflower"=12,
                    "Blackeyedsusan"=4,
                    "Bergamot"=3,
                    "Goldenalexanders"=5,
                    "GreyhcFlo"=6,
                    "Illinoisbundleflower"=7,
                    "Lanceleafedcoreopsis"=8,
                    "Leadplant  Lupine"=9,
                    "Mountainmint"=10,
                    "NewEnglandaster"=11,
                    "Partridgepea"=13,
                    "Purpleconeflower"=14,
                    "Purpleprairieclover"=15,
                    "Roundheadedbushclover"=16,
                    "Royalcatchfly"=17,
```

```
                    "Showyticktrefoil"=18,
                    "Stiffgoldenrod"=19,
                    "Sweetconeflower"=20,
                    "Whiteprairieclover"=21
      ),
         selectize = FALSE
      )
    ),

    # Main panel for displaying outputs ----
    mainPanel(

      # Output: Tabset w/ summary, plot, and map ----
      tabsetPanel(type = "tabs",
                  tabPanel("Intended flower",

                           plotOutput("plot1")

                  ),

                  tabPanel("Forb Analysis",

                           plotOutput("plot2"),
                           plotOutput("plot3")
                  )
      ))

  )
)

# Define server logic to read selected file ----
server <- function(input, output, session) {

  output$plot1 <- renderPlot({
    data <- read.csv("data21.csv",
                     header = T,
                     sep = ",",
                     quote = '  " '
    )
    data <- data[data$Month ==input$month,]
    aa <- as.numeric(input$flo)
    ab <- c(1,2,aa,70,71,72)
    data <- data[,ab]
    p <- ggplot(data, aes(x=data[[3]], fill=Plot)) +
       geom_density(alpha=.3)
    p <- p + labs(x = "Numbers Count",
                             y = "Density")
    p
  })

  runReg <- reactive({
    data <- read.csv("data21.csv",
                      header = T,
```

```
                    sep = ",",
                    quote = '  " '
    )
    data1 <- as.data.frame(data1)
    data1$Stratified = as.factor(data1$Stratified)
    data1$Seedmix = as.factor(data1$Seedmix)
    data1$Application = as.factor(data1n$Application)
    lm(as.formula(paste("Ratio ~ Stratified + Seedmix   + Application")), data=data1)
  })

  output$reg <- renderTable({
    summary(runReg())$coefficients
  })

  output$summary <- renderPrint({
    data <- read.csv("data22.csv",
                    header = T,
                    sep = ",",
                    quote = '  " '
    )
    data <- data[data$Month==input$month,]
    summary(data)
  })

  output$plot2 <- renderPlot({
    data <- read.csv("data22.csv",
                    header = T,
                    sep = ",",
                    quote = '  " '
    )
    ggplot(data = data, mapping = aes(x = Plot, y = Count, fill = Name)) + geom_bar(stat= 'identity', p
  })

  output$plot3 <- renderPlot({
    data <- read.csv("data23.csv",
                    header = T,
                    sep = ",",
                    quote = '  " '
    )
    ggplot(data = data, mapping = aes(x = Month, y = Count, fill = Name)) + geom_bar(stat= 'identity',
  })

}

# Create Shiny app ----
shinyApp(ui, server)
```