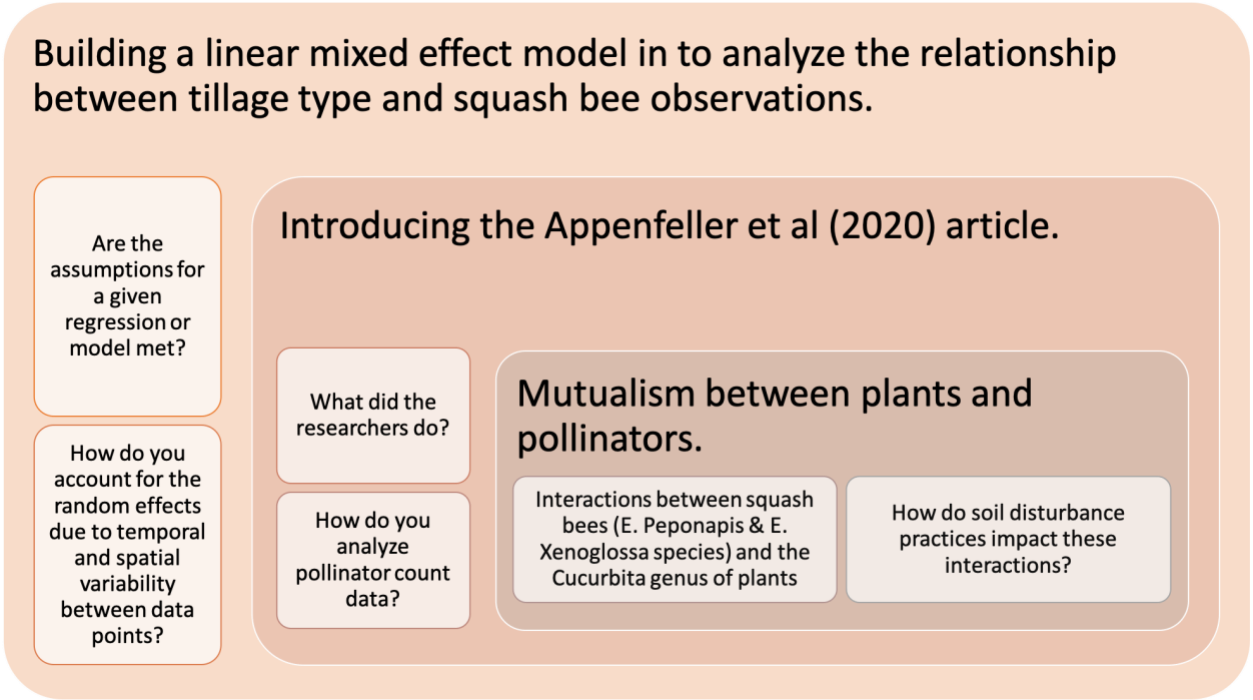
Save the Bees, Ditch the Plow:

A lesson exploring how squash bee abundance is affected by varying levels of soil disturbance.

By Grace Lumsden-Cook, *Environmental Research Methods* Spring 2022

The purpose of this lesson is to explore the effects that three varying degrees of soil disturbance has on the squash bee abundance that is observed. We will learn about the importance of plant-pollinator interactions, and the methods used to analyze pollinator count data. Using R Studio, an open-access software program, we will fit a linear mixed model using the lme4 package, an anova() test. Finally, we will learn about how community-based monitoring efforts can inform future pollinator conservation efforts.

Learning Objectives



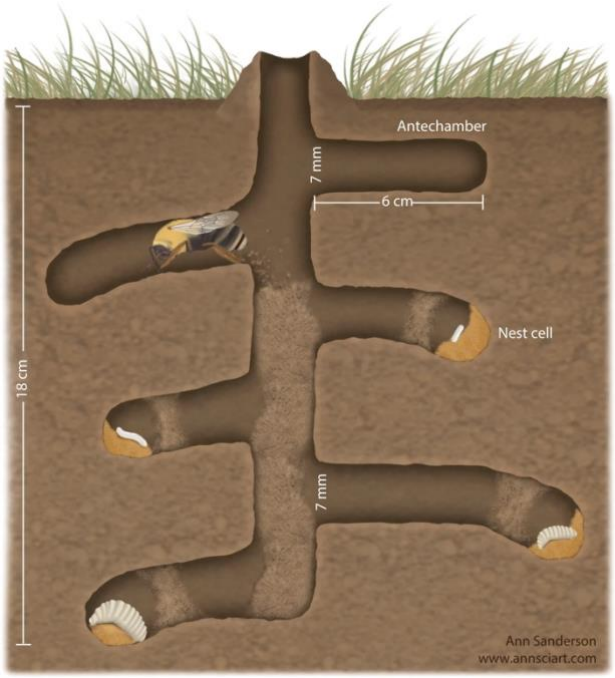
Knowledge Prerequisites

- Prior experience with linear regression and assumptions made regarding independence and distribution of data.
- Familiarity with multiple regression models (linear model with 2 or more independent variables) and interpreting their results.

Importance of plant-pollinator interactions

Animals such as bats, moths, and bees can be to thank for providing pollination services to over three quarters of all flowering plants, or angiosperms globally (Ollerton et al., 2011). As opposed to plants that are pollinated by abiotic processes, biotic pollinated species of plants mostly rely on bees (20,000 spp. globally) for cross-pollination (Lopez-Uribe, 2017). The bees transfer pollen from the stamen (male organ) to the stigma (female organ) so that fertilization and sexual reproduction can occur.

Plants like squash and gourds (Genus: *Cucurbita*) began to coevolve with bees about 10 million years ago (Dorchin et al., 2017). Insect-facilitated pollination is critical for their reproduction because these plants produce both male and female flowers separately. While other bees do enjoy collecting nectar from the female *Cucurbita* flowers, most bees do not intend to collect any pollen. In fact, its pollen has shown to negatively impact the reproduction, and increase mortality of bumbles bees that consumed it (Brochu et al., 2020).



Hoary squash bee (*Peponapis pruinose*) ground nest (Willis Chan et.al., 2019, Figure 1).

Squash bees (Eucera *Peponapis* & E. (*Xenoglossa*) are the only species of bees that are able use Cucurbita pollen as a resource (Lopez-Uribe, 2021). After a female squash bee collects pollen from a male *Cucurbita* flower, they bring it back to their ground nest. These nests house the next generation of squash bees, and the pollen provides nutrients for these developing larvae. Once an egg and enough pollen has been deposited in each cell, the nest it sealed and backfilled so that the larvae can overwinter and emerge as adults the next summer.

Squash bees are known to nest in areas that are in close proximity to *Cucurbita* plants. So, how could farm management practices near these nests impact the survival and success of these pollinators?

How these interactions are affected by farm management practices

We will be focusing on three types of tillage that have varying degrees of soil disturbance for this lesson.

- **No tillage** is characterized by a lack of soil disturbance between harvesting and planting crops, leaving crop stubble or residues.

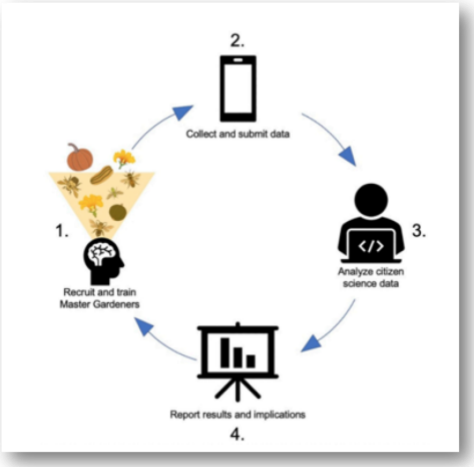
- **Reduced tillage (a.k.a. conservation tillage)** is defined by lower intensity of tillage leaving some crop residues on the soil surface.

- **Full tillage (a.k.a. conventional tillage)** uses cultivation (e.g. plowing, harrowing) as the primary means of weed control and seedbed preparation resulting in a loose soil surface and lack of plant residues on the soil surface.

Some of these farm management techniques have the potential to negatively alter the habitats of ground nesting organisms depending on the level of disturbance that the soil is undergoing. Not only can this be harmful for animals like the squash bee, it also has the potential of limiting the ecosystem services that farmers rely on to make a living. It is extremely important to understand how these bees behave and what ecosystem services that they

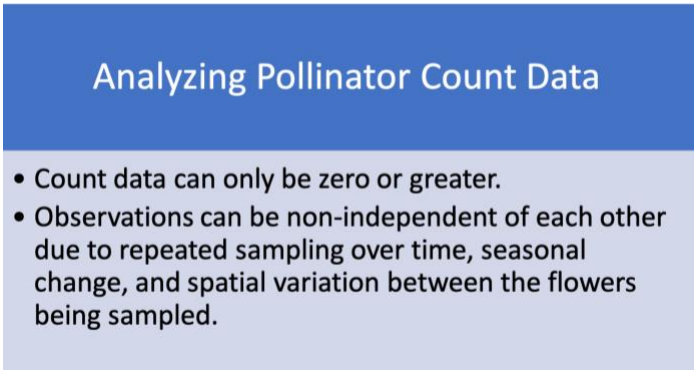
provide locally. Having that knowledge can help to better target conservation efforts for pollinator species, some of which are experiencing dramatic declines in numbers.

Introducing the focal paper



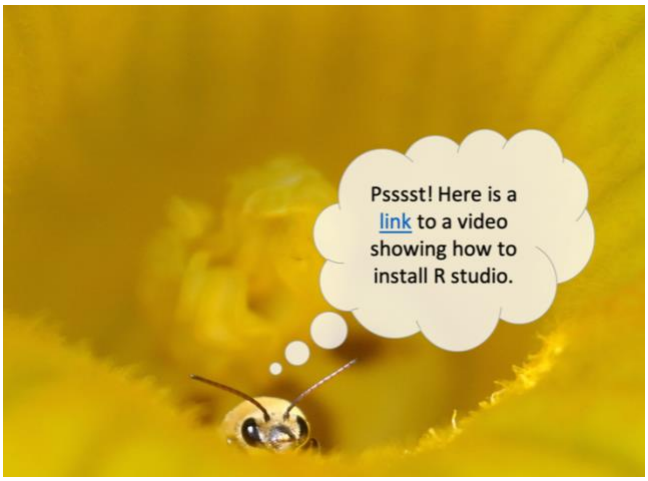
This lesson is based on a paper published in March of 2020 titled, “Citizen science improves our understanding of the impact of soil management on wild pollinator abundance in agroecosystems” by Appenfeller, Lloyd, and Szendrei. The researchers facilitated the collection of pollinator count data by creating an easy- to-use, smartphone interface which they piloted in Michigan, USA. The state’s Master Gardeners were recruited and trained on how to visually identify pollinators, and the method for submitting observations to the research team for data analysis. With this experimental setup, the researchers were able to get a huge amount of data across spatial and temporal scales without having to shell out the resources that a survey like that would need.

How can count data be analyzed?



Before we open R Studio, let's talk about how count data is collected and analyzed. These are discrete variables that can tell us if there are any pollinators present, and how many are observed. Therefore, an analysis that assumes continuous variables would not fit the data. A linear model is not going to work in this case.

We also have to ask ourselves, “are the variables that I am studying independent of each other?”. In situations where, visual sampling is done over long periods of time, or over large areas, the data that you collect may be correlated due to random and fixed effects. These effects can contribute to misinformed variability in the dataset due to ill-fitting statistical models.



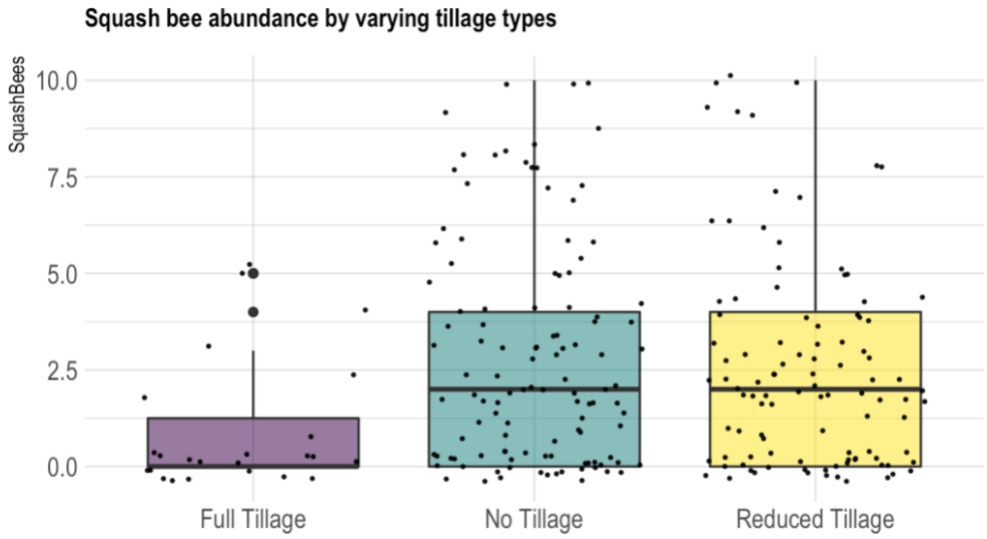
R Studio Tutorial: Exploring How Types of Soil Tillage Effects Squash Bee Abundance

After loading the dataset into R Studio, compute some summary statistics for the number of squash bee visits (n =242) for each type of tillage.

TillageType	N	mean	median	sum	sd	se
<chr>	<int>	<dbl>	<dbl>	<int>	<dbl>	<dbl>
Full Tillage	24	0.9166667	0	22	1.665942	0.3400590
No Tillage	114	2.8596491	2	326	2.865347	0.2683644
Reduced Tillage	104	2.5480769	2	265	2.661775	0.2610085

Question 1: Which tillage type resulted in the highest number of squash bee observations total? Lowest number of total observations?

No Tillage areas had the highest total number of squash bee observations with 326. Full Tillage had the least number of total squash bee observations with 22.



Question 2: Does there look to be a best and worse tillage type for promoting squash bee abundance?

It is hard to tell whether No Tillage and Reduced Tillage just looking at the box and whisker plot. However, the mean number of observations for No Tillage is 2.86, versus 2.5 for Reduced Tillage. So, No Tillage looks to be the best, and Full Tillage looks to be the worst for promoting squash bee abundance.

Linear Mixed Models (aka Mixed-Effect Model)

A mixed effect model is a type of linear model that considers both the fixed and random effects that can lead to variability in your dataset. These models can help to prevent pseudoreplication (Hurlbert, 1984). This can happen when correlated variables are assumed to be independent of each other, but in reality, they have some degree of an effect on each other. This could be due to random or other fixed effects that are not reflected in the analysis or experimental design.

These mixed-effect models allow us to account for the influences that these random effects (also called factors when using categorical data) have on how well your response variable is explained by the chosen predictor variables.

We are trying to understand how well the predictor variable (tillage type) explains the response variable (squash bee abundance). However, due to the data collection methods there is some randomness that must be built into the model in order to account for the "noise", or natural variability of the data. The random effects in this model are county and date.

```
SB_mixed_spat <- lmer(SquashBees ~ tilltype.sb + (1|county.sb), data=SB_resp)
summary(SB_mixed_spat)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: SquashBees ~ tilltype.sb + (1 | county.sb)
   Data: SB_resp

REML criterion at convergence: 1147.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.5916 -0.7227 -0.2273  0.4610  2.8302

Random effects:
 Groups     Name        Variance Std.Dev.
 county.sb (Intercept) 1.657    1.287
 Residual              6.222    2.494
Number of obs: 242, groups:  county.sb, 22

Fixed effects:
                          Estimate Std. Error t value
(Intercept)                1.0019     0.7063   1.418
tilltype.sbNo Tillage      1.0245     0.7268   1.410
tilltype.sbReduced Tillage 1.2810     0.7447   1.720

Correlation of Fixed Effects:
            (Intr) tll.NT
tlltyp.sbNT -0.798
tlltyp.sbRT -0.822  0.820
```

To find the variance due to the random effect of the factor "county.sp":

1.657 ÷ (6.222 + 1.657)
=
~20% of the total variance

```
SB_mixed_date <- lmer(SquashBees ~ tilltype.sb + (1|date.sb), data=SB_resp)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: SquashBees ~ tilltype.sb + (1 | date.sb)
   Data: SB_resp

REML criterion at convergence: 1142.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.0951 -0.6928 -0.2011  0.5096  2.7568

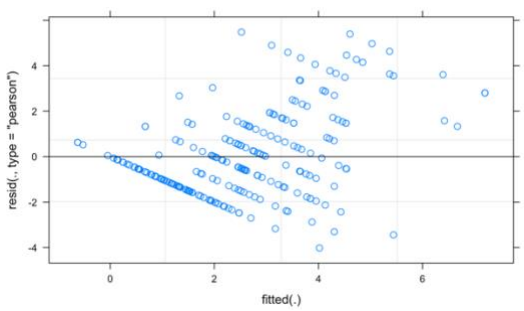
Random effects:
 Groups   Name        Variance Std.Dev.
 date.sb  (Intercept) 2.255    1.502
 Residual             4.947    2.224
Number of obs: 242, groups:  date.sb, 110

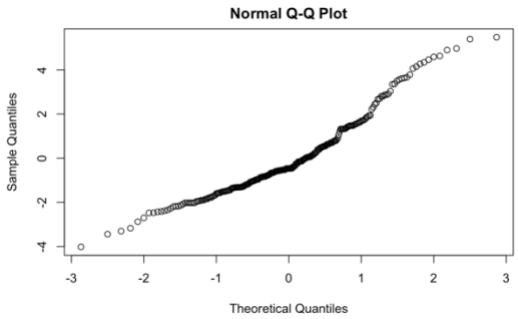
Fixed effects:
                          Estimate Std. Error t value
(Intercept)                0.8577     0.5259   1.631
tilltype.sbNo Tillage      1.9085     0.5602   3.407
tilltype.sbReduced Tillage 1.5137     0.5627   2.690

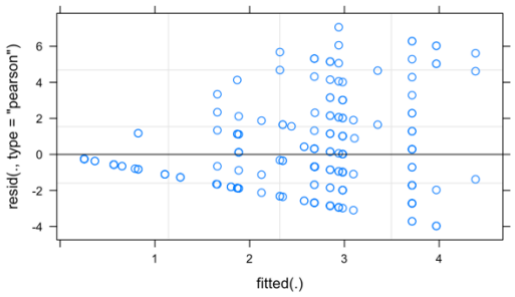
Correlation of Fixed Effects:
            (Intr) tll.NT
tlltyp.sbNT -0.870
tlltyp.sbRT -0.874  0.815
```

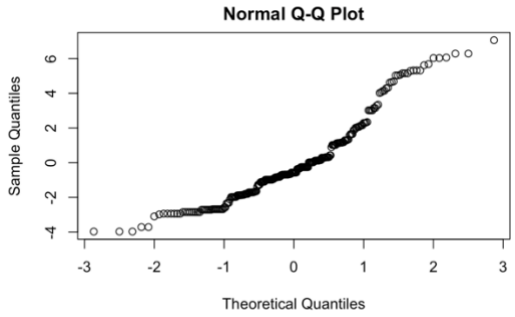
Question 3: How much variance is due to the temporal random effect?

2.255 / (4.947 + 2.255) = ~31%









Question 4: Looking at the previous plots, does either model appropriately fit the data? Why? (Hint: are there any patterns or clumping of points in the first, and do the points follow a straight line in the second plot)

No, they both fit the data poorly. The model with spatial random effects had a better dispersion of data points when it was plotted. This means that it better explained the variance of the data. However, the model with the temporal random effects looks to follow a straight line better than the spatial model, so it follows a more normal distribution of data.

Fitting the mixed effect model with the nested random effects

```
SB_nested_mixed <- lmer(SquashBees ~ tilltype.sb + (1|date.sb/county.sb), data=SB_resp)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: SquashBees ~ tilltype.sb + (1 | date.sb/county.sb)
   Data: SB_resp

REML criterion at convergence: 1138.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.7536 -0.6139 -0.1695  0.4274  2.2734

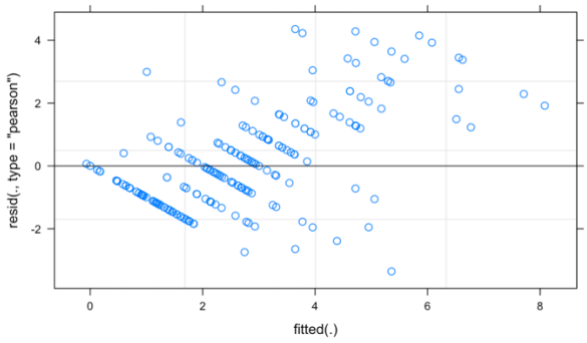
Random effects:
 Groups            Name        Variance Std.Dev.
 county.sb:date.sb (Intercept) 1.618    1.272
 date.sb           (Intercept) 1.844    1.358
 Residual                      3.667    1.915
Number of obs: 242, groups:  county.sb:date.sb, 201; date.sb, 110

Fixed effects:
                       Estimate Std. Error t value
(Intercept)              0.9381     0.5235   1.792
tilltype.sbNo Tillage    1.8088     0.5592   3.235
tilltype.sbReduced Tillage   1.4364     0.5624   2.554

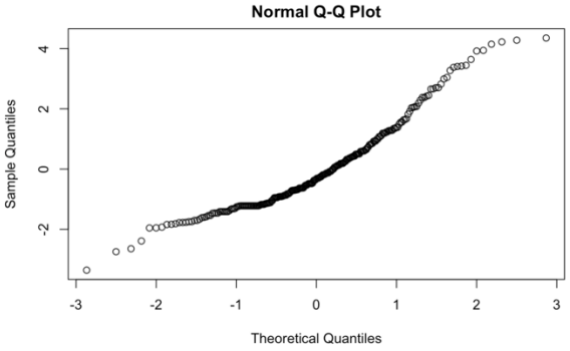
Correlation of Fixed Effects:
            (Intr) tll.NT
tlltyp.sbNT -0.869
tlltyp.sbRT -0.874  0.808
```

Question 5: What is the explained variance attributed to the nested random effect?

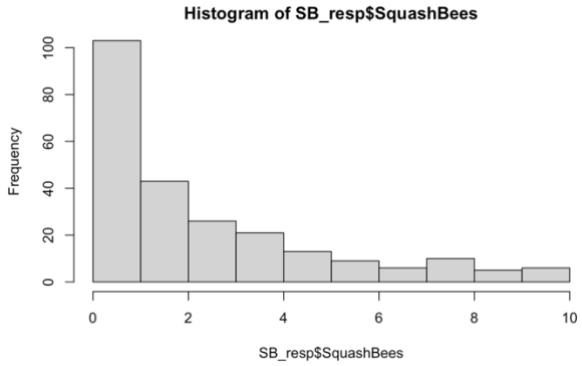
Explained variance from the nested random effect: 1.618 / (1.618 + 3.667) = ~30%



Notice how there is still this clumping pattern when the full LMM is plotted. If there was an even distribution in the data, there would be no evident patterns expressed in the distribution of the data points.



When a the full LMM is plotted to show how normally distributed it is (qqplot()), there looks to be a curve in the middle of the line. If the data were normally distributed, the points would follow a straight line



Due to the non-normal (negative binomial) distribution of this data, the researchers chose to use a Generalized Linear Mixed model (GLMM) to account for the large number of observations resulting in zero observations. See Bolker et al (2009), "Generalized linear mixed models: a practical guide for ecology and evolution", to learn more about GLMMs.

Assessing the fit of the model using an anova() test

```
full_SB_mixed <- lmer(SquashBees ~ tilltype.sb + (1|date.sb/county.sb), data=SB_resp, REML = FALSE)
.
reduced_SB_mixed <- lmer(SquashBees ~ 1 + (1|date.sb/county.sb), data=SB_resp, REML = FALSE)
```

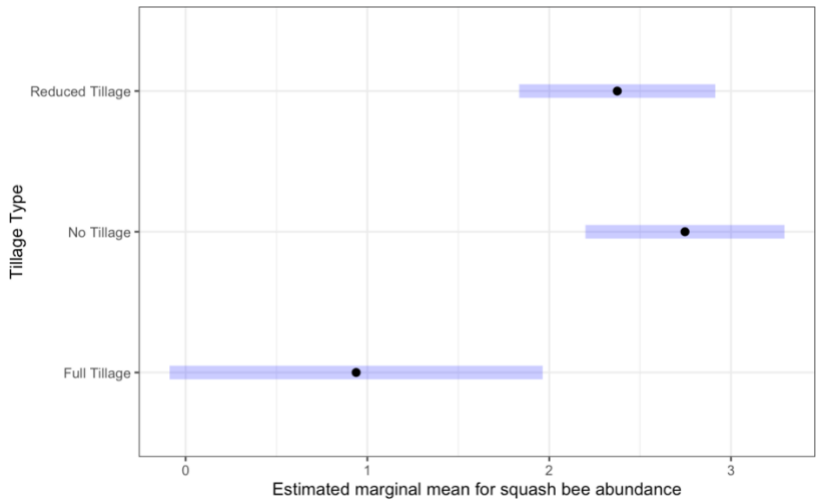
```
```{r}
anova(reduced_SB_mixed, full_SB_mixed)
```
```

```
Data: SB_resp
Models:
reduced_SB_mixed: SquashBees ~ 1 + (1 | date.sb/county.sb)
full_SB_mixed: SquashBees ~ tilltype.sb + (1 | date.sb/county.sb)
                 npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
reduced_SB_mixed    4 1155.7 1169.7 -573.87   1147.7
full_SB_mixed       6 1149.5 1170.5 -568.77   1137.5 10.189  2   0.006129 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question 6: Are the two models significantly different from each other? (Hint: look at p-value) What does this mean for the fitness of the full model?

Yes, the two models are significantly different from one another because the p-value is less than 0.05 (p-value = 0.0061). This means that our fixed variable (tillage types) does has an effect on our response variable (squash bee abundance).

```
means_nestedmix <- emmeans(SB_nested_mixed, pairwise ~ tilltype.sb, adjust="fdr")
```



Question 7: Are there any significant relationships between tillage and squash bee abundance based on tillage type? Any non-significant relationships?

Reduced Tillage and No Tillage both have significant effects on squash bee abundance. Because Full Tillage has values (blue bar) that overlap with zero, it does not have a significant relationship with squash bee abundance.

Lesson Takeaways

Not only can reduced and no tillage practices help to preserve the habitats of native pollinators on farms, they can also ensure that farmers can take advantage of the pollination services that keep them in business. It is more important than ever to engage with communities that live in, and work with agroecosystems. Biodiversity loss, decreasing food security, and the spread of plant pathogens all pose huge threats to agriculture and healthy ecosystem functioning around the world. This is why it is vitally important that community-based monitoring efforts are established (Ryan et al., 2018).

Despite the limitations that it may pose for gaining “academic data”, community-based monitoring efforts can be used to collect large amounts of data crossing spatial scales that would otherwise be extremely expensive and time consuming. This method of surveying provides farmers and stakeholders with a seat at the table, and gets folks talking about issues pertaining to native pollinator conservation.

Citations

Agarwal, R. (2020). Pruinose Squash Bee - Peponapis pruinosa - in pumpkin flower. Photograph. Retrieved from Flickr https://www.flickr.com/photos/30314434@N06/50197174111

Appenfeller L.R., Lloyd S., Szendrei Z. (2020). Citizen science improves our understanding of the impact of soil management on wild pollinator abundance in agroecosystems. PLOS ONE 15(3): e0230007. https://doi.org/10.1371/journal.pone.0230007 Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230007

Brochu, K.K., van Dyke, M.T., Milano, N.J. *et al.* (2020). Pollen defenses negatively impact foraging and fitness in a generalist bee (*Bombus impatiens*: Apidae). *Sci Rep* **10,** 3112. https://doi.org/10.1038/s41598-020-58274-2

Hajduk, G. K. (2022). *INTRODUCTION TO LINEAR MIXED MODELS*. Introduction to linear mixed models. Retrieved April 10, 2022, from https://ourcodingclub.github.io/tutorials/mixed-models/

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs*. Vol. 54, No. 2., pp. 187-211.

López-Uribe, M. (2021) Squash bees: Origins, diversification, and ecological interactions. OSU Horticulture Seminar Series

Ryan, S.F. *et al.* (2018) The role of citizen science in addressing grand challenges in food and agriculture research. *Proc. R. Soc. B* **285**: 20181977. http://dx.doi.org/10.1098/rspb.2018.1977

Willis Chan, D.S., Prosser, R.S., Rodríguez-Gil, J.L. et al. (2019) Assessment of risk to hoary squash bees (*Peponapis pruinosa*) and other ground-nesting bees from systemic insecticides in agricultural soil. *Sci Rep* **9,** 11870. https://doi.org/10.1038/s41598-019-47805-1