Analysis of Factors Affecting Cover Crop Adoption on Almond Orchards in California

https://github.com/emac2020/Almond-Survey-2020

Emily McNamara

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

California is the epicenter of global almond production, producing 80% of the world's almond supply. As the forecasted growth of consumer demand shows few signs of subsiding, farmers in California are converting farmland to almond orchards at a considerable rate - from 428,000 acres in 1996 to 1,170,000 acres in 2019. However, the industry is in a precarious position as almonds require 100% pollination form managed honey bees - a species that has witnessed significant decline in population since 2006. The decline in managed honey bees is attributed to poor nutrition, lack of diverse forage, stress from transporation, and pesticide toxicity.

To mitigate the decline in managed honey bees and protect honey bee health on almond orachards, experts determined several bee-friendly practices that growers can adopt on their farms. One of these practices is planting temporary forage called 'cover crop' between tree rows to increase the diversity and abundance of nutrient and pollen sources for the honey bees.

This study uses data from a survey administered to almond producers and farm managers throughout California to identify potential barriers in adopting cover crop and understand why the practice of planting cover crops is not widely adopted by almond producers.

To meet the research objectives, the primary research questions are:

- Where are the respondents' almond orchards located?
- Which demographic factors affect whether or not respondents have planted cover crop in the last 5 years?
- How does region affect whether or not the respondents have planted cover crop in the last 5 years?
- How does respondent role in the almond operation affect whether or not the respondents have planted cover crop in the last 5 years?
- How does respondent age affect whether or not the respondents have planted cover crop in the last 5 years?
- How does the size of the almond operation affect whether or not the respondents have planted cover crop in the last 5 years?

2 Dataset Information

For the analysis, the following datasets were used:

2.1 Almond Survey Results Dataset

This dataset contains data of 301 completed responses from a survey that was distributed to almond producers and farm managers throughout California. The survey was launched on December 10th, 2019 and was closed on February 5th, 2020. Data were collected using Qualtrics.

The downloaded file was saved in the project folder path ./Data/Raw/Almond_Survey_Results_raw.csv on 2020-04-02

2.1.1 Data Content Information

The dataset contains 24 columns. See the report's README for Almond Survey Results Dataset: Column Names and Descriptions.

2.2 Almond Survey Numeric Results Dataset

This dataset contains the same responses as the Almond Survey Results Dataset in section 2.1., but the dataset was downloaded from Qualtrics in numerical answer form with split-answer columns.

The downloaded file was saved in the project folder path ./Data/Raw/Almond_Survey_Numeric_Answers_R on 2020-04-02

2.2.1 Data Content Information

2.2.1.1 Table 1: Almond Survey Numeric Response Dataset Content Information

The dataset contains 48 columns. See the report's README for Almond Survey Numeric Response Dataset: Column Names and Descriptions.

2.3 Naming Conventions and File Formats

The files are named according to the following convention: Files are named according to the following naming convention: `databasename_datatype_details_stage.format`, where:

databasename refers to the database from where the data originated

datatype is a description of data

details are additional descriptive details, particularly important for processed data

stagerefers to the stage in data management pipelines (e.g., raw, cleaned, or processed)

format is a non-proprietary file format (e.g., .csv, .txt)

3 Exploratory Analysis and Wrangling

3.1 Data Wrangling: Almond Survey Response Dataset

The raw 'Almond Survey Response Dataset' and the 'Almond Survey Numeric Response Dataset' both contained unnecessary information for the overarching goals of this project. Thus, data regarding permanent pollinator habitat, the pollination manager, non-yield bearing acreage, and water sources was removed. The dates the respondents completed the survey were removed as well because the analyses do not involve 'time' as a parameter.

```
# Read in raw almond survey response data
almonds.project.raw <- read.csv("./Data/Raw/Almond_Survey_Results_Raw.csv")

# Look at column names
#colnames(almonds.project.raw)

# Select column names that only apply to cover crop analysis

almonds.project.CC.processed <- almonds.project.raw %>%
  dplyr::select(Role.in.Operation:Regions,
                Total.Yield.Bearing.Acreage, Acre.Ranges,
                Cover.Crop.Grown:Cover.Crop.Incentives, Pollination:Age)


# Fill all empty cells in almonds.project.CC.processed with 'NA'
# Name 'almonds.CC'
almonds.CC <- almonds.project.CC.processed %>% mutate_all(na_if, "")


# Save new processed dataset
#write.csv(almonds.CC, row.names = FALSE,
#file = "./Data/Processed/Almond_Project_CoverCrop_Processed.csv")
```

3.2 Data Wrangling: Almond Survey Numeric Response Dataset

```
# Read in almond survey numeric response data
almonds.numeric.raw <- read.csv("./Data/Raw/Almond_Survey_Numeric_Answers_Raw.csv")


# Look at column names
#colnames(almonds.numeric.raw)

# Choose columns for cover crop analyses that require numeric data
```

```r
almonds.numeric <- almonds.numeric.raw %>%
  dplyr::select(Role.in.Operation, Regions, Total.Yield.Bearing, Cover.Crop.Grown, Age)

# Save new processed dataset
#write.csv(almonds.numeric, row.names = FALSE,
#file = "./Data/Processed/Almond_Survey_Numeric_Answers_Processed.csv")
```

The raw datasets now only contain data regarding respondents responses to cover crop adoption and interest as well as the respondents' answers to the demographic and pollination questions in the survey.

3.3 Data Exploration

After wrangling the data into a format that would allow me to answer my research questions more effectively, I was able to explore the processed datasets.

```r
# Column names of both datasets
colnames(almonds.CC)

colnames(almonds.numeric)

# Structure of datasets

str(almonds.CC)

str(almonds.numeric)

# Class

class(almonds.CC$Total.Yield.Bearing.Acreage)
class(almonds.CC$Regions)

class(almonds.numeric$Total.Yield.Bearing)
class(almonds.numeric$Role.in.Operation)

# Dimensions of datasets

dim(almonds.CC)

dim(almonds.numeric)

# Head

head(almonds.CC)
```

```r
head(almonds.numeric)

# Summary

summary(almonds.CC)
summary(almonds.CC$Regions)
summary(almonds.CC$Role.in.Operation)
summary(almonds.CC$Age)
summary(almonds.CC$Total.Yield.Bearing.Acreage)
summary(almonds.CC$Acre.Ranges)
summary(almonds.CC$Rental.Price)

summary(almonds.numeric)
summary(almonds.numeric$Regions)
summary(almonds.numeric$Total.Yield.Bearing)
```

Figure 1 plots the locations of the respondent's almonds orchards by county in Califorina. Because respondents were allowed to "check all that apply," when selecting the counties where their almond orchards are located, some of the location responses in the dataset had several counties listed in one cell. To remedy this, I first created a new column in the dataset Excel spreadsheet called "County.Multiple," then I copied the data from the "County" column into the new column and entered in "multiple" for location responses that contained more than one county.

Thus, in **Figure 1**, "multiple" is used to describe the respondents who had almond orchards in several counties. This figure shows that a majority of the respondents farm almonds in Stanislaus, Madera, and Fresno. Since a majority of almond producers are from Southern California, I anticipated a higher response rate from these counties.

I then went through the spreadsheet again and noted that in the location responses that had multiple counties listed, the counties were close to one another and therefore, the counties could be categorized by Central Valley watershed basins. I chose to categorize the counties by watersheds provided the fact that the production of tree nuts requires a considerable amount of water, and due to the state's water scarcity, water is a critical factor in determining managerial practices for farmers. Thus, I hypothesized that watershed basins are most representative of respondent behavior toward bee-friendly practices (i.e. cover crop).

The regional categories include Sacramento Valley, Delta, San Joaquin Basin, and Tulare Basin. The Sacramento Valley consists of respondents who farm almonds in Butte, Colusa, Glenn, or Tehama; the Delta region includes Sacramento, Solano, Yolo, and Yuba; the San Joaquin Basin includes San Joaquin, Stanislaus, Merced, and Madera; the Tulare Basin includes Tulare, Kings, Kern, and Fresno.

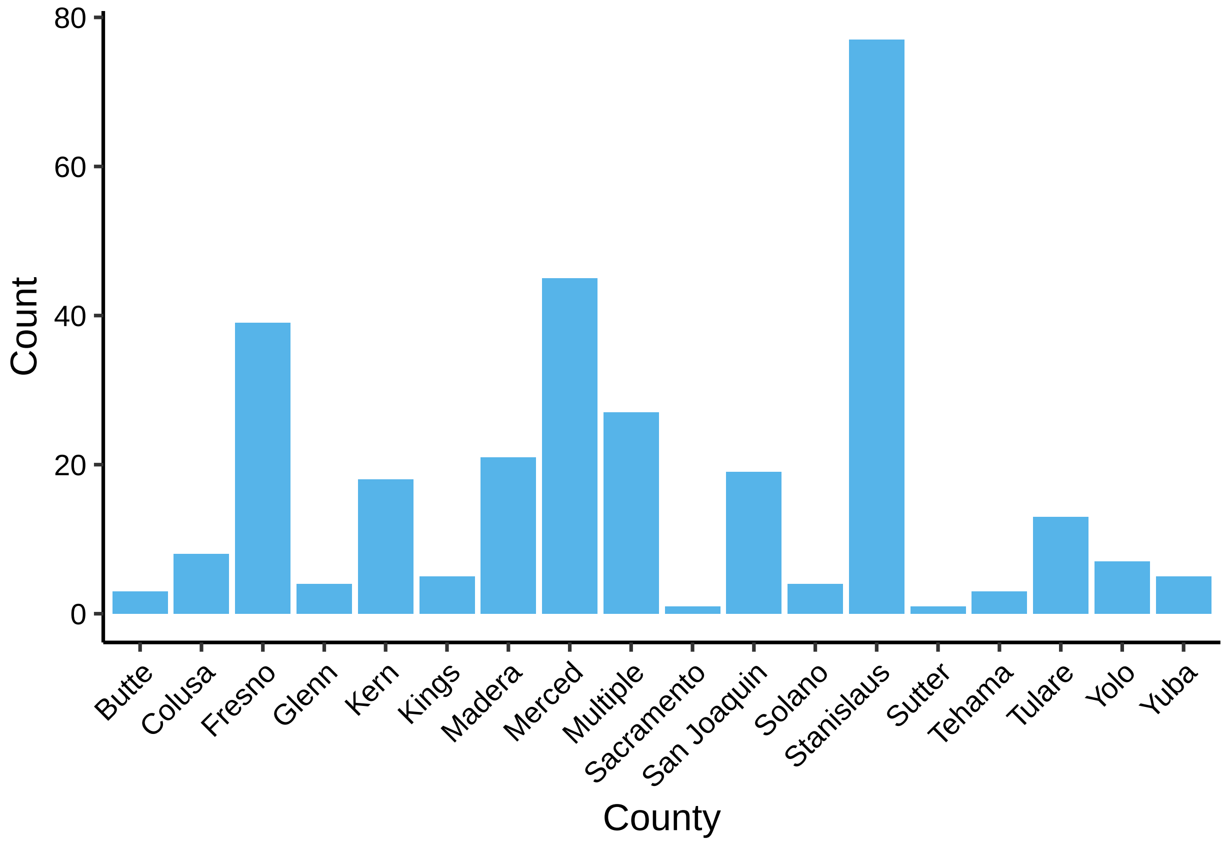

Figure 1: Location of Respondents' Almond Orchards by County in California

Figure 2 plots the location of the respondents' almond orchards by region. We can see from this plot that a majority of the respondents farm almonds in the San Joaquin and Tulare Basins. It is important to note that the northern section of the Central Valley consists of the Sacramento Valley and Delta regions, while the southern section includes the San Joaquin and Tulare Basins. Therefore, a majority of the respondents farm almonds in the southern section of the Central Valley.

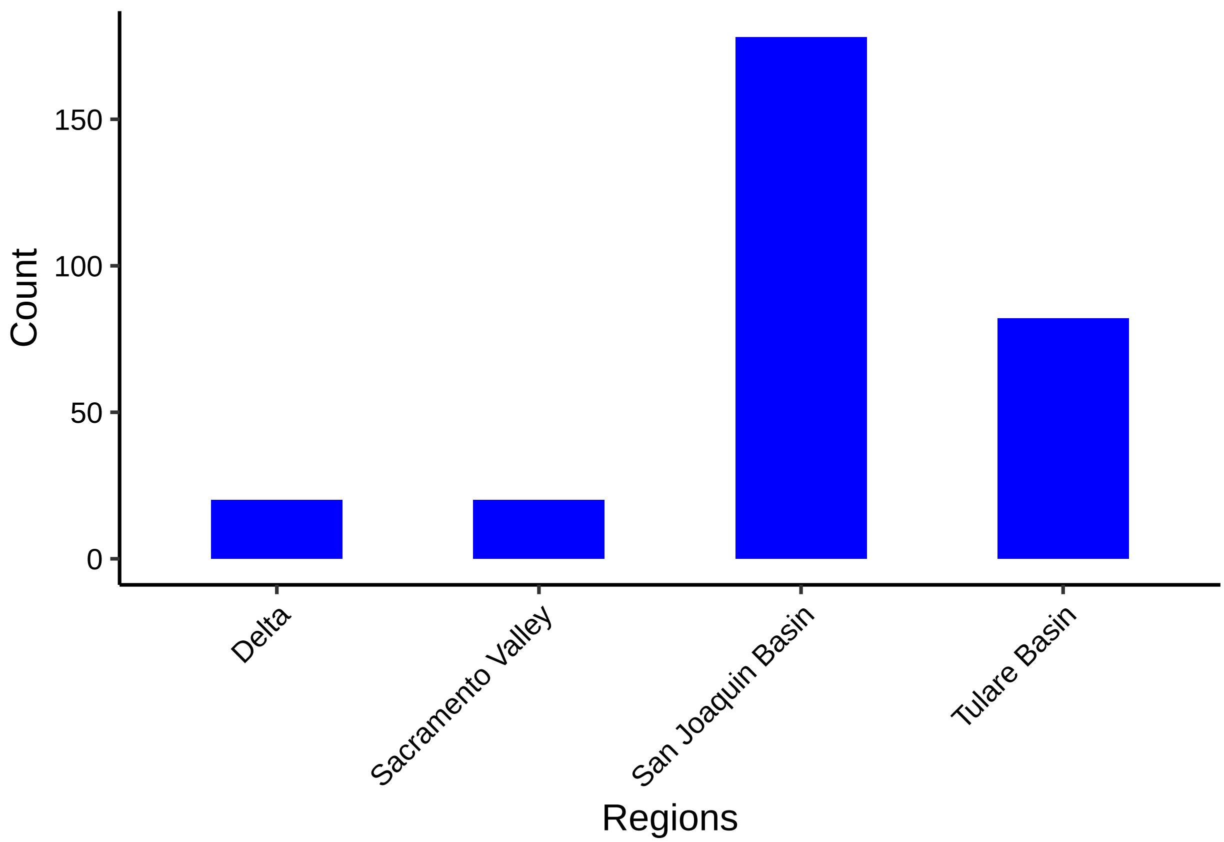

Figure 2: Location of Respondents' Almond Orchards by Region

Figure 3 shows the age ranges of the respondents who completed the survey. We can see from this plot that a majority of the respondents fell within the 25-34 and 55-64-year-old age ranges. These age ranges are interesting to note because the responses for cover crop adoption will be come from both the newer and older farming generations.

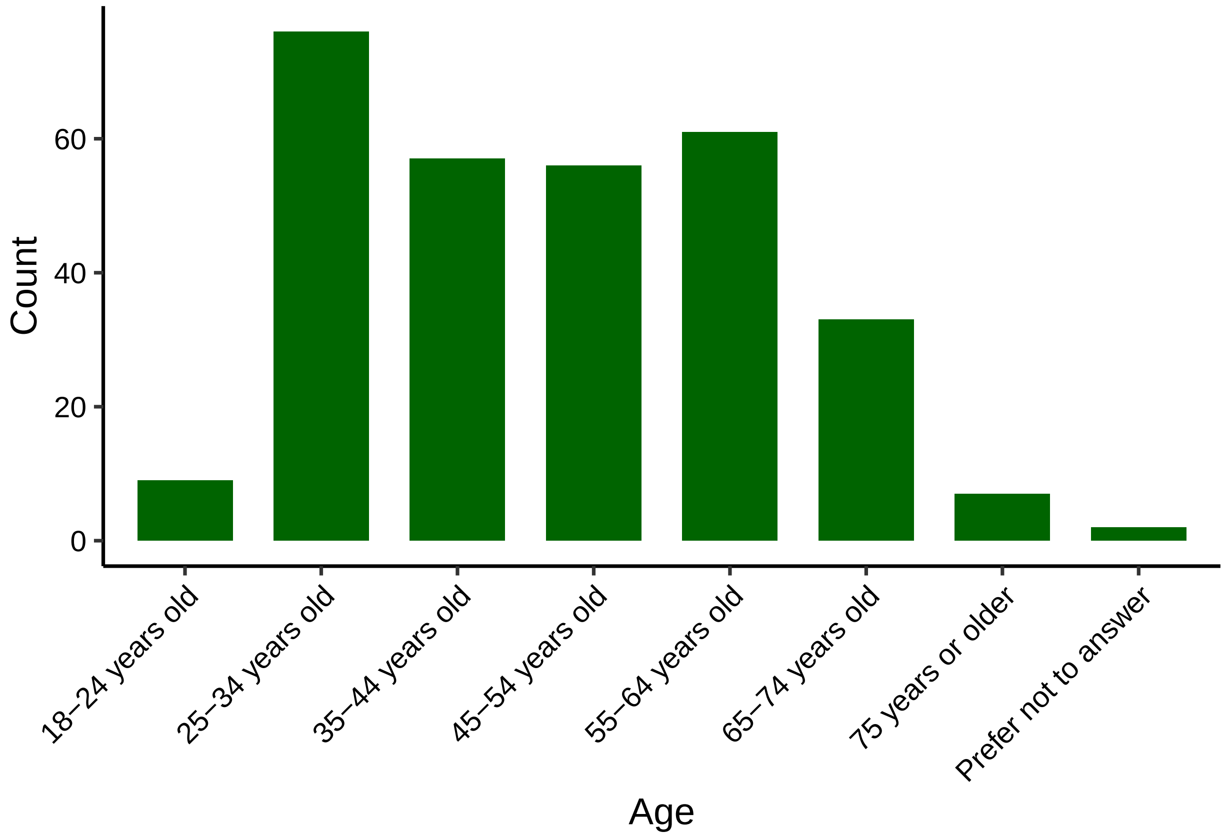

Figure 3: Age Ranges of Respondents

Figure 4 plots the respondents' roles in almond operations. The respondents could select "Owner, not responsible for dat-to-day management," "Owner/Operator," or "Farm Manager (not owner)." This figure shows that a majority of the respondents selected "Owner/Operator," meaning they own and operate their almond orchard(s). This outcome was the opposite of my hypothesis regarding the roles the survey takers play in day-to-day operations because I assumed that a majority of the farms would consist of several parcels that were operated by farm managers and that farm managers would constitute the majority of survey respondents.

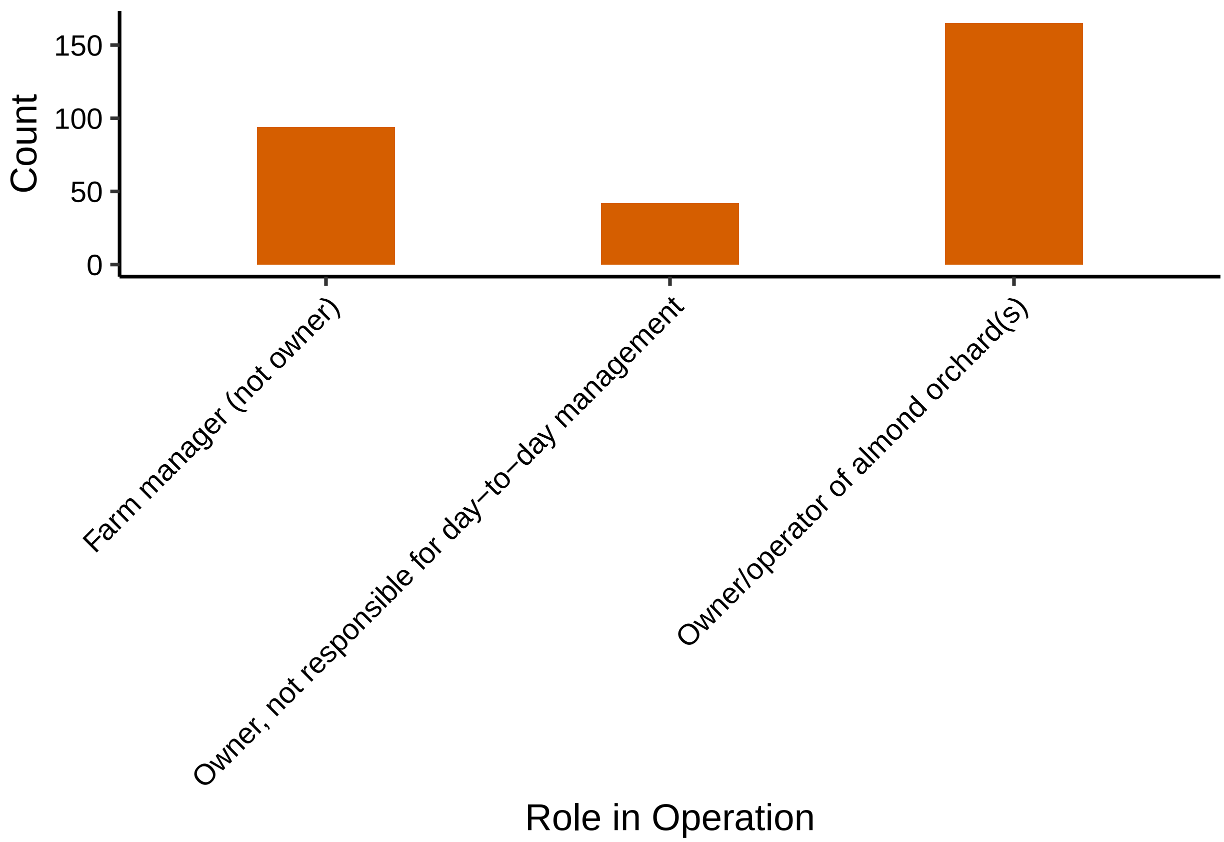

Figure 4: Respondents' Roles in Almond Operations

4 Analysis

Analyzing possible factors affecting cover crop adoption among almond producers in California will allow me to obtain a greater insight into barriers that may be preventing producers from adopting cover crop and thus, bee-friendly practices. To assess cover crop adoption, I analyzed whether or not the respondents had grown a cover crop in the last five years. Out of 301 responses, 33.6% (101 respondents) answered "Yes" to having grown a cover crop, while 66.5% (200) selected "No." To understand why respondents may have selected "No" or "Not Sure" when asked whether they had grown a cover crop, I analyzed how responses differed by demographics using logistic regressions and chi-square tests of independence.

When controlling for all factors, we see that respondents who farm almonds in the San Joaquin and Tulare Basins are significantly less likely to grow a cover crop than respondents from the Delta (glm; z = -3.32, p-value $<$ 0.05; glm; z = -3.66, p-value $<$ 0.05). We also see that that total yield-bearing acreage has a signficant affect on whether or not respondents grow cover crop (glm; z = -1.97, p-value $<$ 0.05). Lastly, results indicate that there is a marginally significant liklihood of Owner/Operators selecting "Yes" to having grown cover crop over Farm Managers (glm; z = -1.94, p-value = 0.053).

4.1 Question 1: Which demographic factors affect whether or not respondents have planted cover crop in the last 5 years?

```r
# Number of respondents who have grown and not grown a cover crop grown

CCtbl <- table(almonds.CC$Cover.Crop.Grown)

CCtbl
```

```
##
##  No Yes
## 200 101
```

```r
# Statistical Test 1: GLM

almonds.CC$Total.Yield.Bearing.Acreage <-
  as.numeric(almonds.CC$Total.Yield.Bearing.Acreage)

Grown.CC <- glm(Cover.Crop.Grown ~ as.factor(Regions) +
                  as.factor(Role.in.Operation) + as.factor(Age)
                + Total.Yield.Bearing.Acreage, almonds.CC, family = binomial)

summ(Grown.CC, model.fit = getOption("Grown.CC", TRUE))
```

Observations	300 (1 missing obs. deleted)
Dependent variable	Cover.Crop.Grown
Type	Generalized linear model
Family	binomial
Link	logit

$\chi^2(13)$	38.10
Pseudo-R² (Cragg-Uhler)	0.17
Pseudo-R² (McFadden)	0.10
AIC	373.18
BIC	425.04

	Est.	S.E.	z val.	
(Intercept)	0.67	1.00	0.67	0
as.factor(Regions)Sacramento Valley	-0.43	0.71	-0.61	0
as.factor(Regions)San Joaquin Basin	-1.83	0.55	-3.32	0
as.factor(Regions)Tulare Basin	-2.15	0.59	-3.66	0
as.factor(Role.in.Operation)Owner, not responsible for day-to-day management	-0.50	0.43	-1.16	0
as.factor(Role.in.Operation)Owner/operator of almond orchard(s)	-0.61	0.32	-1.94	0
as.factor(Age)25-34 years old	0.70	0.97	0.72	0
as.factor(Age)35-44 years old	1.54	0.98	1.57	0
as.factor(Age)45-54 years old	1.38	0.98	1.41	0
as.factor(Age)55-64 years old	1.46	0.99	1.48	0
as.factor(Age)65-74 years old	0.88	1.02	0.86	0
as.factor(Age)75 years or older	0.43	1.44	0.30	0
as.factor(Age)Prefer not to answer	-12.24	618.25	-0.02	0
Total.Yield.Bearing.Acreage	-0.01	0.00	-1.97	0

Standard errors: MLE

```r
# Table 1

pretty_GrownCC.glm <- prettify(summary(Grown.CC))
```

```
## Waiting for profiling to be done...
```

```r
kable(pretty_GrownCC.glm,caption = "Cover Crop Adoption by Respondent Demographics",
      align = c("l", rep("r", 3)),format = "latex", booktabs = TRUE)  %>%
  kable_styling(latex_options = "scale_down")
```

Table 1: Cover Crop Adoption by Respondent Demographics

	Estimate	Odds Ratio	CI (lower)	CI (upper)	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.6735526	1.9611924	0.2276019	1.343706e+01	0.9992537	0.6740557	0.5	
as.factor(Regions)Sacramento Valley	-0.4331810	0.6484431	0.1562540	2.601697e+00	0.7089666	-0.6110034	0.541	
as.factor(Regions)San Joaquin Basin	-1.8264889	0.1609778	0.0506884	4.538400e-01	0.5509155	-3.3153706	0.001	***
as.factor(Regions)Tulare Basin	-2.1465417	0.1168877	0.0344185	3.522543e-01	0.5862495	-3.6614816	<0.001	***
as.factor(Role.in.Operation)Owner, not responsible for day-to-day management	-0.5014137	0.6056738	0.2535697	1.396341e+00	0.4330734	-1.1578030	0.247	
as.factor(Role.in.Operation)Owner/operator of almond orchard(s)	-0.6106427	0.5430018	0.2912510	1.005196e+00	0.3150897	-1.9379961	0.053	.
as.factor(Age)25-34 years old	0.6963671	2.0064502	0.3562549	1.770932e+01	0.9658172	0.7210133	0.471	
as.factor(Age)35-44 years old	1.5372192	4.6516371	0.8120775	4.217531e+01	0.9784133	1.5711349	0.116	
as.factor(Age)45-54 years old	1.3796698	3.9735894	0.6894595	3.589636e+01	0.9786013	1.4098385	0.159	
as.factor(Age)55-64 years old	1.4569459	4.2928287	0.7402873	3.940042e+01	0.9851822	1.4788594	0.139	
as.factor(Age)65-74 years old	0.8814707	2.4144480	0.3747732	2.331597e+01	1.0243793	0.8604925	0.39	
as.factor(Age)75 years or older	0.4265512	1.5319649	0.0553530	2.532588e+01	1.4411770	0.2959742	0.767	
as.factor(Age)Prefer not to answer	-12.2354867	0.0000049	NA	1.324559e+28	618.2538111	-0.0197904	0.984	
Total.Yield.Bearing.Acreage	-0.0072510	0.9927752	0.9854897	9.998687e-01	0.0036844	-1.9680474	0.049	*

4.2 Question 2: How does the respondent's role in the operation affect whether or not they have grown cover crop in the last 5 years?

To analyze how the respondent's "role in operation" affects whether or not they have grown a cover crop, I first made a plot to visualize the number of farm managers, owners, and owner/operators who have and have not grown cover crop in the last five years. **Figure 5** illustrates that the majority of respondents owned and operated their almond operation, and a majority of them have not grown cover crop.

```
alm.Role= almonds.CC[almonds.CC$Role.in.Operation != " " ,]
```

alm.Role.Role.in.Operation	alm.Role.Cover.Crop.Grown	Freq
Farm manager (not owner)	No	57
Owner, not responsible for day-to-day management	No	28
Owner/operator of almond orchard(s)	No	115
Farm manager (not owner)	Yes	37
Owner, not responsible for day-to-day management	Yes	14
Owner/operator of almond orchard(s)	Yes	50

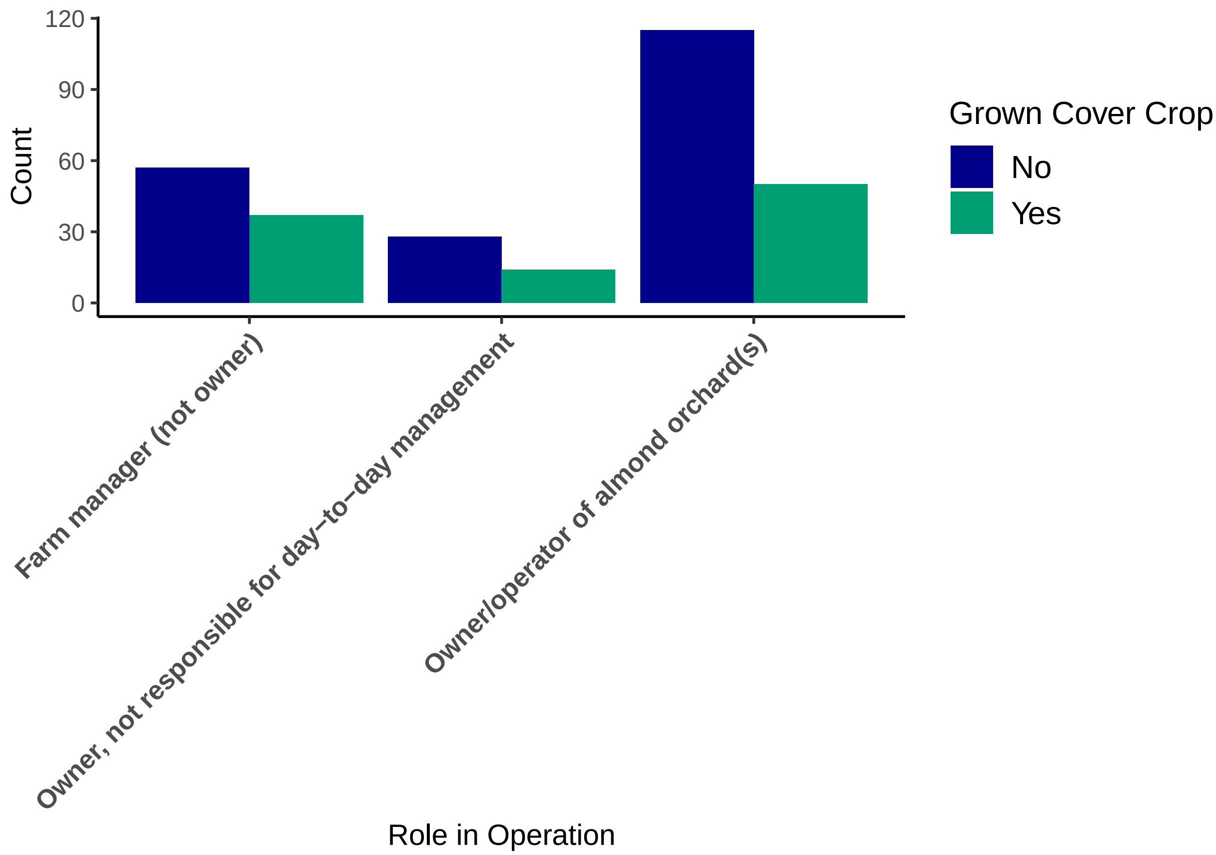

Figure 5: Respondents' Roles in Almond Operations by Cover Crop Grown

I then performed a chi-square test of independence to examine the relationship between "role in operation" and whether or not a respondent grew a cover crop. I found that the two variables are independent - a respondent's "role in operation" does not significantly affect whether or not they had grown cover crop in the last five years (X^2(2, N = 301) = 2.21, p > 0.05).

```
# Statistical Test 1: Chi-square

almonds.numeric$Role.in.Operation <- as.factor(almonds.numeric$Role.in.Operation)
almonds.numeric$Cover.Crop.Grown <- as.factor(almonds.numeric$Cover.Crop.Grown)



Role.Grown.CC.tbl = table(almonds.numeric$Role.in.Operatio,
                          almonds.numeric$Cover.Crop.Grown)
Role.Grown.CC.tbl
```

```
##
##       1   2
##   1  14  28
##   2  50 115
##   3  37  57
```

```
chisq.test(Role.Grown.CC.tbl)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Role.Grown.CC.tbl
## X-squared = 2.2051, df = 2, p-value = 0.332
```

4.3 Question 3: How does the respondent's age affect whether or not they have grown cover crop in the last 5 years?

I first plotted the age of respondents by whether or not they have grown a cover crop to visualize the relationship between the two variables. Graph "a" in **Figure 6** shows that respondents between 55-64-years-old had the highest frequency of growing cover crop in the last five years, while respondents between 25-34-years-old had the highest frequency of **not** growing cover crop in the last five years. The higher rate of respondents between 25-34-years-old not having grown cover crop may be due to the fact that they have just purchased the almond orchards or are farm managers, whereas the older generation has most likely been farming for a longer period of time. However, I hypothesize that a greater number of respondents between 25-34-years-old would be interested in growing cover crop than older respondents.

To test my hypothesis that respondents between 24-35-years-old are more interested in planting cover crop than the older respondents, I plotted the age of the respondents by their *interest* in growing cover crop. It is important to note that the question regarding *interest* in growing cover crop only appeared for those who selected *no* to having grown cover crop in the last 5 years. Thus, the total count in this analysis is lower than that of the cover crop *grown* analysis. Out of the respondents who have not grown cover crop, graph "b" in **Figure 6** shows that my hypothesis was correct and that more respondents from the 25-34 age bracket are interested in planting cover crop than the older respondents.

```r
# Create cowplot

CC.Age.plots <- plot_grid(Age.GrownCC.plot, Age.InterestCC.plot, scale = 0.9,
                          align = "v", ncol = 1,
                          labels = c('a) Cover Crop Grown', 'b) Cover Crop Interest'),
                          label_x = 0, label_y = 0,
  hjust = -0.0, vjust = -0.0, label_size = 8)

print(CC.Age.plots)
```

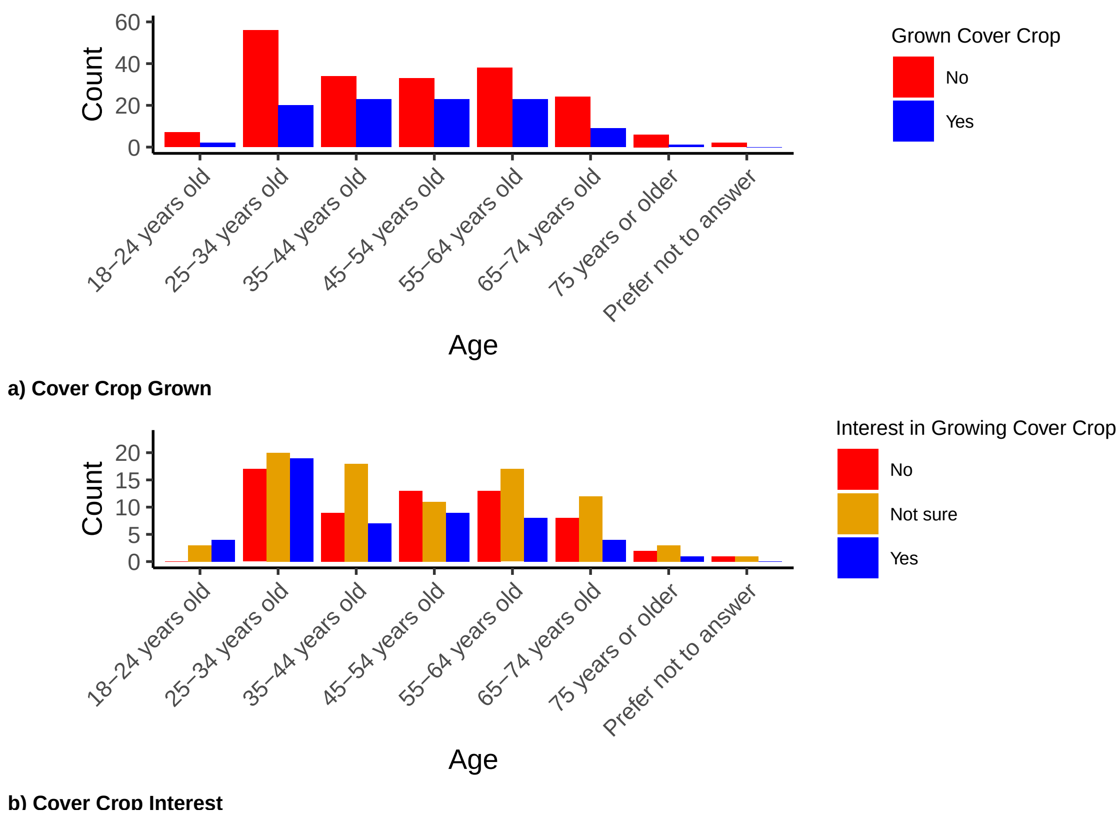

Figure 6: Respondent's Age by Cover Crop Grown and Cover Crop Interest

almAge.Age	almAge.Cover.Crop.Grown	Freq
18-24 years old	No	7
25-34 years old	No	56
35-44 years old	No	34
45-54 years old	No	33
55-64 years old	No	38
65-74 years old	No	24
75 years or older	No	6
Prefer not to answer	No	2
18-24 years old	Yes	2
25-34 years old	Yes	20
35-44 years old	Yes	23
45-54 years old	Yes	23
55-64 years old	Yes	23
65-74 years old	Yes	[illegible]

alm.Age	alm.Cover.Crop.Interest	Freq
18-24 years old	No	0
25-34 years old	No	17
35-44 years old	No	9
45-54 years old	No	13
55-64 years old	No	13
65-74 years old	No	8
75 years or older	No	2
Prefer not to answer	No	1
18-24 years old	Not sure	3
25-34 years old	Not sure	20
35-44 years old	Not sure	18
45-54 years old	Not sure	11
55-64 years old	Not sure	17
65-74 years old	Not sure	12
75 years or older	Not sure	3
Prefer not to answer	Not sure	1
18-24 years old	Yes	4
25-34 years old	Yes	19
35-44 years old	Yes	7

I ran a logistic regression to examine the relationship between the age of the respondents and whether or not they had grown a cover crop in the last five years. The model showed that the relationship between the two variables is not significant - age does not affect whether or not a cover crop was grown (glm; p-value > 0.05). I then ran a chi-square test of independence to anaylze this relationship further, and the analysis confirmed that the two variables are independent of one another (X^2(7, N = 301) = 8.14, p > 0.05).

```
# Statistical Test 1: GLM
Age.Grown.CC <- glm(Cover.Crop.Grown~Age, almonds.CC, family = binomial)

#summary(Age.Grown.CC)

summ(Age.Grown.CC)
```

Observations	301
Dependent variable	Cover.Crop.Grown
Type	Generalized linear model
Family	binomial
Link	logit

$\chi^2(7)$	8.99
Pseudo-R^2 (Cragg-Uhler)	0.04
Pseudo-R^2 (McFadden)	0.02
AIC	391.11
BIC	420.77

	Est.	S.E.	z val.	p
(Intercept)	-1.25	0.80	-1.56	0.12
Age25-34 years old	0.22	0.84	0.26	0.79
Age35-44 years old	0.86	0.85	1.02	0.31
Age45-54 years old	0.89	0.85	1.05	0.29
Age55-64 years old	0.75	0.84	0.89	0.37
Age65-74 years old	0.27	0.89	0.30	0.76
Age75 years or older	-0.54	1.35	-0.40	0.69
AgePrefer not to answer	-13.31	624.19	-0.02	0.98

Standard errors: MLE

```
# Table 2

pretty_Age.GrownCC.glm <- prettify(summary(Age.Grown.CC))
```

```
## Waiting for profiling to be done...
```

Table 2: Cover Crop Adoption by Age Ranges of Respondents

	Estimate	Odds Ratio	CI (lower)	CI (upper)	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.2527630	0.2857143	0.0425711	1.182077e+00	0.8017837	-1.5624699	0.118
Age: 25-34 years old	0.2231436	1.2500000	0.2742203	8.858418e+00	0.8430387	0.2646896	0.791
Age: 35-44 years old	0.8618967	2.3676471	0.5166669	1.685549e+01	0.8460184	1.0187682	0.308
Age: 45-54 years old	0.8917496	2.4393939	0.5317246	1.737916e+01	0.8465450	1.0533990	0.292
Age: 55-64 years old	0.7506710	2.1184211	0.4641011	1.504244e+01	0.8441867	0.8892239	0.374
Age: 65-74 years old	0.2719337	1.3125000	0.2555195	9.922993e+00	0.8919837	0.3048640	0.76
Age: 75 years or older	-0.5389965	0.5833333	0.0235138	7.681183e+00	1.3451854	-0.4006857	0.689
Age: Prefer not to answer	-13.3133048	0.0000017	NA	1.388545e+28	624.1943416	-0.0213288	0.983

```
kable(pretty_Age.GrownCC.glm,caption = "Cover Crop Adoption by Age Ranges of Respondents",
      align = c("l", rep("r", 3)),format = "latex", booktabs = TRUE)  %>%
  kable_styling(latex_options = "scale_down")
```

```
# Statistical Test 2: Chi-Square

Age.Grown.CC.tbl = table(almonds.CC$Age, almonds.CC$Cover.Crop.Grown)
Age.Grown.CC.tbl
```

```
##
##                          No Yes
##   18-24 years old         7   2
##   25-34 years old        56  20
##   35-44 years old        34  23
##   45-54 years old        33  23
##   55-64 years old        38  23
##   65-74 years old        24   9
##   75 years or older       6   1
##   Prefer not to answer    2   0
```

```
chisq.test(Age.Grown.CC.tbl)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Age.Grown.CC.tbl
## X-squared = 8.1357, df = 7, p-value = 0.3208
```

4.4 Question 4: How does the size of the respondent's almond operation affect whether or not they have grown cover crop in the last 5 years?

To assess the effect "size of operation" has on cover crop adoption, I used the numeric responses from the "total yield-bearing acreage" text entry boxes in the survey. I analyzed the responses for “total yield-bearing acreage” rather than “total non-yield bearing acreage” because the non-yield bearing trees are not mature enough for pollination from managed honey bees. Thus, I assumed that non-yield bearing acreage would be less representative of respondents’ behaviors toward adopting bee-friendly practices.

I then categorized the individual text entries for “total yield-bearing acreage” by “Small,” “Medium,” and “Large” to represent the size of the operation. The acreage range within each category was determined by the ALmond Board’s almond acre ranges for farm sizes. The “Small” category represented producers with 0-40 acres of yield-bearing almonds, “Medium” represented producers with 41-200 acres and “Large” represented producers with 201+ acres.

To visualize the relationship between the size of the respondent’s almond operation and whether or not they have grown cover crop, I plotted the two variables. **Figure 7** shows that “Small” operations had the highest frequency of **not** growing cover crop in the last five years.

```
alm.Size = almonds.CC[almonds.CC$Acre.Ranges != " " ,]
```

alm.Size.Acre.Ranges	alm.Size.Cover.Crop.Grown	Freq
Large	No	48
Medium	No	55
Small	No	94
Large	Yes	24
Medium	Yes	39
Small	Yes	37

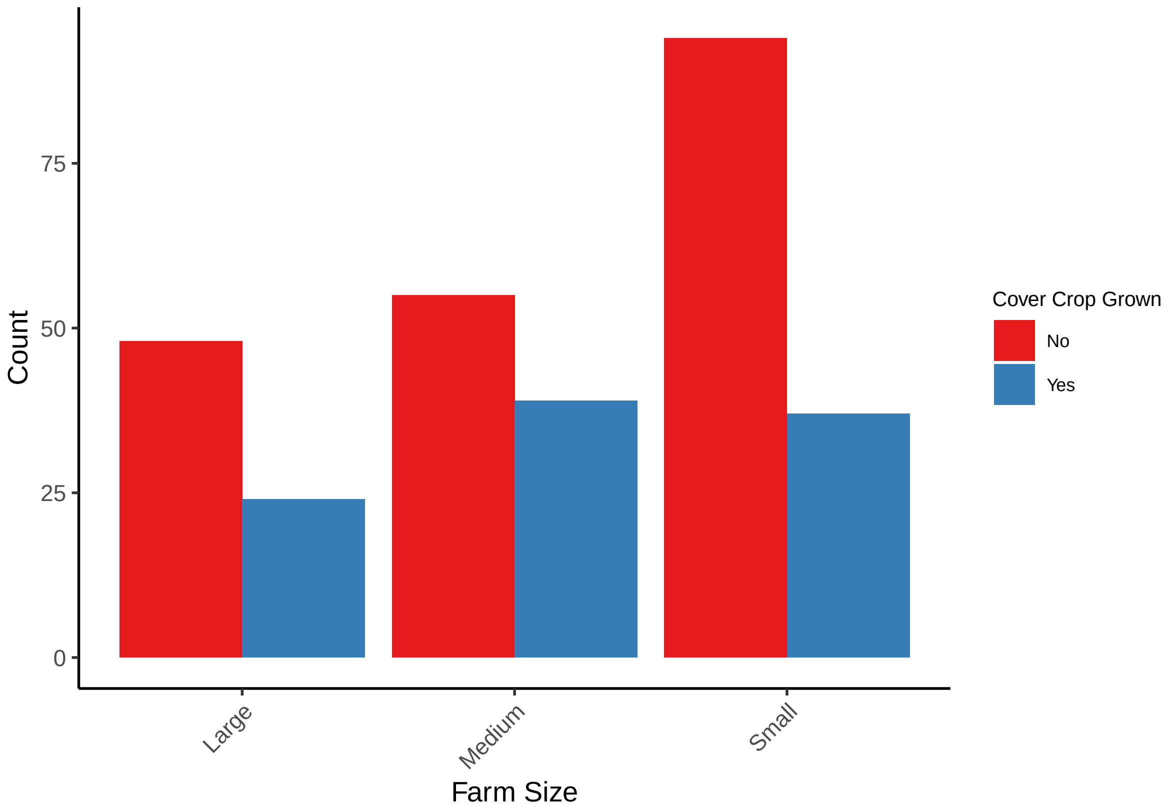

Figure 7: Size of Operation by Cover Crop Interest

Further analysis showed that, unlike the model from **Question 1** that controlled for all factors, when solely analyzing cover crop adoption versus the amount of total yield-bearing acreage, total yield-bearing acreage does not significantly affect whether the respondent has grown a cover crop in the last five years (glm; p-value > 0.05).

```
Size.GrownCC <- glm(Cover.Crop.Grown~Acre.Ranges , almonds.CC, family = binomial)

#summary(Size.GrownCC)

summ(Size.GrownCC)
```

Observations	297 (4 missing obs. deleted)
Dependent variable	Cover.Crop.Grown
Type	Generalized linear model
Family	binomial
Link	logit

```
# Table 3

pretty_Size.GrownCC.glm <- prettify(summary(Size.GrownCC))
```

$\chi^2(2)$	4.27
Pseudo-R^2 (Cragg-Uhler)	0.02
Pseudo-R^2 (McFadden)	0.01
AIC	381.19
BIC	392.27

	Est.	S.E.	z val.	p
(Intercept)	-0.69	0.25	-2.77	0.01
Acre.RangesMedium	0.35	0.33	1.07	0.28
Acre.RangesSmall	-0.24	0.32	-0.76	0.45

Standard errors: MLE

```
## Waiting for profiling to be done...
```

```
kable(pretty_Size.GrownCC.glm,caption = "Cover Crop Adoption by Size of Almond Operation",
      align = c("l", rep("r", 3)),format = "latex", booktabs = TRUE)  %>%
  kable_styling(latex_options = "scale_down")
```

Table 3: Cover Crop Adoption by Size of Almond Operation

	Estimate	Odds Ratio	CI (lower)	CI (upper)	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.6931472	0.500000	0.3014593	0.8073187	0.2500000	-2.7725892	0.006	**
Acre.Ranges: Medium	0.3493756	1.418182	0.7515922	2.7079823	0.3260718	1.0714684	0.284	
Acre.Ranges: Small	-0.2392297	0.787234	0.4244132	1.4733357	0.3164889	-0.7558864	0.45	

4.5 Question 5: How does the region where the respondent's farm almonds affect whether or not they have grown cover crop in the last 5 years?

I performed a chi-square test of independence to examine the relationship between "region" and whether or not respondents have grown a cover crop in the last five years. The relationship between these variables is significant, showing that "region" does affect whether or not respondents selected "Yes" to having grown a cover crop (x2(3, N = 301) = 24.205, $p < 0.01$). I then ran a logistic regression to see which regions are most likely to affect whether or not the respondent has grown a cover crop. The model showed that respondents with almond acreage located in the San Joaquin and Tulare basins were significantly less likely to have grown a cover crop in the last five years than respondents in the Delta region (glm; z = -3.31, p-value < 0.05; glm; z = -3.48, p-value < 0.05).

```
# Statistical Test 1: Chi-square
almonds.CC$Regions <- as.factor(almonds.CC$Regions)

Region.GrownCC.tbl = table(almonds.CC$Regions, almonds.CC$Cover.Crop.Grown)
Region.GrownCC.tbl
```

```
##
##                     No Yes
##   Delta              6  14
##   Sacramento Valley   7  13
##   San Joaquin Basin 125  53
##   Tulare Basin       61  21
```

```
chisq.test(Region.GrownCC.tbl)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Region.GrownCC.tbl
## X-squared = 24.205, df = 3, p-value = 2.263e-05
```

```
# Statistical Test 2: GLM

Regions.GrownCC.glm <- glm(Cover.Crop.Grown ~  Regions,
                             data = almonds.CC, family = binomial )

#summary(Regions.GrownCC.glm)

summ(Regions.GrownCC.glm)
```

Observations	300 (1 missing obs. deleted)
Dependent variable	Cover.Crop.Grown
Type	Generalized linear model
Family	binomial
Link	logit

$\chi^2(3)$	22.86
Pseudo-R^2 (Cragg-Uhler)	0.10
Pseudo-R^2 (McFadden)	0.06
AIC	368.42
BIC	383.24

	Est.	S.E.	z val.	p
(Intercept)	0.85	0.49	1.74	0.08
RegionsSacramento Valley	-0.23	0.68	-0.34	0.74
RegionsSan Joaquin Basin	-1.71	0.51	-3.31	0.00
RegionsTulare Basin	-1.91	0.55	-3.48	0.00

Standard errors: MLE

```r
# Table 4

pretty_Region.GrownCC <- prettify(summary(Regions.GrownCC.glm))
```

```
## Waiting for profiling to be done...
```

```r
kable(pretty_Region.GrownCC,caption = "Cover Crop Adoption by Region",
      align = c("l", rep("r", 3)),format = "latex", booktabs = TRUE)  %>%
  kable_styling(latex_options = "scale_down")
```

Table 4: Cover Crop Adoption by Region

	Estimate	Odds Ratio	CI (lower)	CI (upper)	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.8472979	2.3333333	0.9360921	6.5892453	0.4879498	1.736445	0.082	.
Regions: Sacramento Valley	-0.2282587	0.7959184	0.2052573	3.0138562	0.6766648	-0.337329	0.736	
Regions: San Joaquin Basin	-1.7053197	0.1817143	0.0615139	0.4788281	0.5147455	-3.312938	0.001	***
Regions: Tulare Basin	-1.9136493	0.1475410	0.0469245	0.4165938	0.5496430	-3.481622	<0.001	***

5 Summary and Conclusions

When controlling for all factors that potentially affect cover crop adoption, the analyses show that only "Region" and "Total Yield-Bearing Acreage" have a significant relationship regarding whether or not respondents selected "Yes" to having grown cover crop in the last 5 years. However, when the factors are analyzed individually against "cover crop adoption," the models indicate that only "Region" is significant. Upon further analysis, the results also showed that respondents who farm almonds in the San Joaquin and Tulare Basins were significantly less likely to select "Yes" to having grown cover crop than respondents in the more northern regions of the Sacramento Valley and Delta. One explanation for this could be due to water availability and price. The northern regions in California have a greater supply of water than those farther south and given the considerable amount of water used to grow tree crops, it could be that farmers in the southern regions are more apprehensive about planting cover crops due to the extra water they necessitate. This being said, cover crops improve soil health and the soil's water retention capabilities, so by planting cover crop, the farmers could enhance their irrigation efficiency.

Cover crop is just one of several bee-friendly practices that almond producers can adopt on their orchards. However, it is an important practice as it provides managed honey bees with extra and more diverse nutrient and pollen sources. By planting a cover crop, producers could not only improve their soil health but could also strengthen the bee colonies and obtain better pollination of their almond orchards. As the demand for almonds continues to increase, more acres of farmland will transition to almond orchards, and without higher rates of adoption of bee-friendly practices or a shift in consumer mindset, a continued loss of managed honey bees in the US appears inevitable. Thus, given the results from these analyses, industry experts, cost-share programs, and farm extension agents should work with producers in the southern region of the state to assist them in funding and educational opportunities that increase the rates of cover crop adoption.

6 References

- United States Department of Agriculture (USDA). (2019). 2018 California Almond Acreage Report. California Department of Food and Agriculture. Retrieved from https://www.nass.usda.gov/Statistics_by_State/California/Publications/Specialty_and_Other_Releases/Almond/Acreage/201904almac.pdf

- Almond Board of California (ABC). (2019b). Almond Almanac. Almond Board of California. Retrieved from http://newsroom.almonds.com/sites/default/files/pdf_file/Almanac_2019_Web.pdf