Are Honey Bees Still Dying at Alarming Rates?

Kyle Bierly

May 21, 2018

Contents

1 Executive Summary

The fact that honey bees are dying at alarming rates is frequently discussed in media. The evidence for this fact is extensive. Indeed, the total number of honey bee colonies has dropped significantly since the 1950s in Europe and the United States. The total number of honey bee colonies "in Europe decreased from over 21 million in 1970 to about 15.5 million in 2007" (vanEngelsdorp and Meixner 81), and in the United States from "5.9 million managed in 1947 to the low of 2.3 million reported in 2008" (82). However, claims to the magnitude of the die-off are clouded amid disputes between beekeepers and major agrochemical businesses (Walshe). I wanted to unearth the validity of these different claims and just how critical the recent colony loss has been. Unexpected diseases such as Colony Collapse Disorder, or CCD, have expedited the desolation of honey bees since 2006, yet the cause of CCD is still relatively unknown (Oldroyd). The name CCD was attributed to the mysterious syndrome in which "the main symptom is simply a low number of adult bees in the hive" that arose in the spring of 2006-2007 (Oldroyd). Other factors that contribute to colony loss include poor hive management, limited honey bee food sources, diseases and parasites, and pesticides.

I collected my data from the Bee Informed Partnership's National Management Survey and the European Food Safety Authority's survey of beekeepers during the winters of 2012 to 2014. These samples may not reflect the total population of beekeepers and their practices, as they are only representative of beekeepers who responded to the survey. The National Management Survey is a survey of over 4000 beekeepers within the United States collected each year since 2008. The survey received data on total colony loss of honey bees, management techniques of beekeepers, and potential causes of colony loss for their honey bees. The EFSA's survey was of over 4700 beekeepers in countries across the European Union, and the EFSA collected information similar to that of the National Management Survey. I analyzed the winter colony loss of honey bees, because winter is by far the most crucial season in determining the survival of the hive. I did not randomly select the data, because I assumed the data was randomly selected in the two surveys, and I wanted my research to be as representative of their collected samples as possible.

My findings revealed that multiple different factors amalgamated to produce the disastrous die-off. Although, some factors have augmented the issue more than others. There is significant evidence that the average amount of colonies lost per year due to the separate factors are different, because some factors are more deadly and widespread than others. Factors such as CCD and the Varroa parasite are extensively more far-reaching than causes of colony loss like the small hive beetle. The National Management Survey revealed that there is significant evidence that there is a relationship between the year and winter percent loss of honey bee colonies; the percent loss has been decreasing since 2008. The winter percent loss is calculated by dividing the number of colonies lost each winter by the total number of colonies going into the winter, and then multiplying that number by 100. This does not mean honey bees are no longer dying at dangerous rates. Beekeepers speculate that a winter percent loss of at most 15% would be an ideal, sustainable level (Borenstein). In 2017, the percent of colonies lost over the winter was at 22%. Survey director Dennis vanEngelsdorp says the numbers have "gone from horrible to bad" (Borenstein). I also found that the percent loss has been decreasing each year, regardless of the reason for death after adjusting for the reason as a covariate. Overall, the honey bees' situation has been improving.

2 Initial Statistical Summary

2.1 Purpose and Data Collection

In my research, I set out to discover the leading causes of winter colony loss for honey bees, the honey bees that are most susceptible to greater colony losses in the winter, and the overall trend of winter colony loss within the United States.

The data sets from where I obtained my statistics, the National Management Survey and the EFSA's survey of European beekeepers, are only representative of beekeepers who responded to the survey. The collected samples may not be representative of the total beekeeper population, so no definitive conclusion can be directly made about the total population of beekeepers from this data. This may create a potential non-response bias, because beekeepers who did not respond to the survey may have responded differently than those who did. Responding to the surveys demonstrates a beekeeper's commitment and compassion towards ameliorating research for honey bee health. Therefore those who did not respond may be less dedicated to their honey bees and have more substantial winter colony losses. However, both data sets have large sample sizes and can paint a substantial picture of what the total population of beekeepers may look like.

The National Management Survey collected a plethora of statistics from beekeepers. They received information on beekeeping operation size, reason for colony loss (CCD, Varroa parasite, starvation, etc.), management techniques, demographics of beekeepers, and so on. Their data was organized by group, and they did not include the individual raw data. For example, when examining winter colony loss by operation size, the survey results would display the total percent loss for all backyard beekeepers rather than the individual percent loss for each backyard beekeeper. The EFSA's survey, however, included the amount of colonies lost per each individual beekeeper; the EFSA also exhibited the demographics for each beekeeper.

The National Management Survey under the Bee Informed Partnership's research portal, or at this URL: `https://bip2.beeinformed.org/survey/`. The EFSA's survey can be found on the European Union's open data website at this URL: `https://data.europa.eu/euodp/data/dataset/honey-bee-seasonal-mortality-2012-2014-epilobee-analysis`.

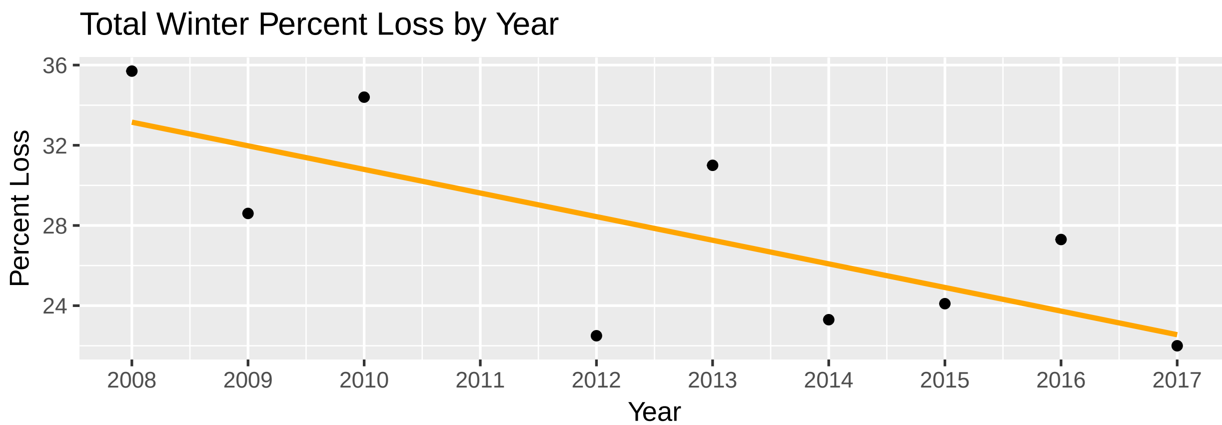
For the purpose of this research project, I decided to explore the variables of reason for colony loss, operation location, and the beekeeper's motivation for beekeeping. Then, I examined how these variables contributed to winter colony loss. I thought these variables would be the most prevalent to the honey bee die-off because it would highlight why bees are dying, where they are dying, and which beekeepers were producing superior results.

2.2 Initial Statistical Analysis

To demonstrate the relationship between the year and winter percent loss of honey bee colonies, let us regard the scatter plot of total percent loss of honey bees for all beekeepers vs. the year, according to the National Management Survey.

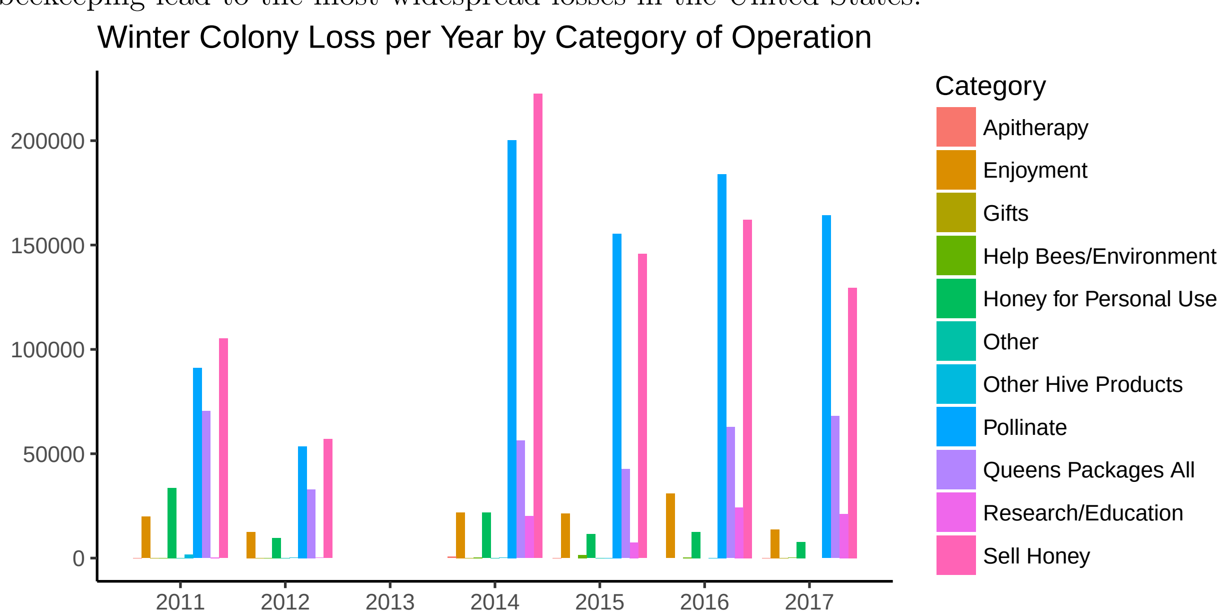
Example of Raw Data of Percent Loss by Year

Year	Percent Loss
2008	0.357
2009	0.286



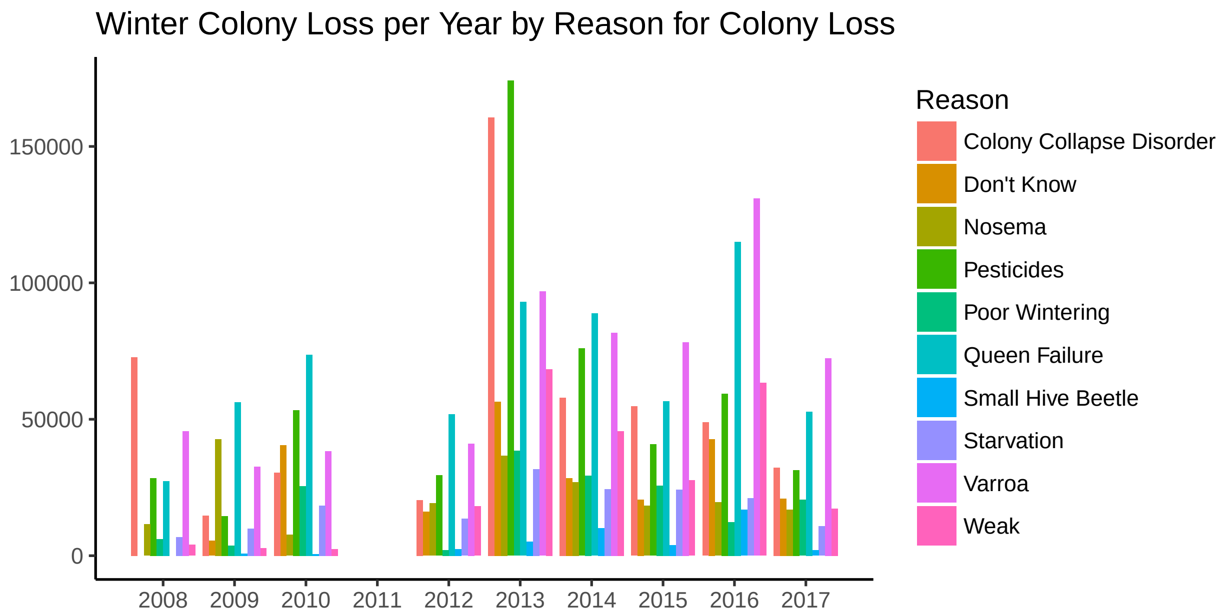
The scatter plot demonstrates a moderate, negative relationship between year and total winter percent loss with $R = -.723$. This implies with a reasonably accurate model that, in general, total winter percent loss has been decreasing since 2008.

We may also look at the total number of colonies reported to have been lost over the winter according to the National Management Survey. The National Management Survey reports the number of colonies lost grouped by the reason the for beekeeping. The survey also reports the number of colonies lost grouped by the cause of colony loss for each year. It is necessary to understand that the number of survey responses differ each year, and, therefore, so will the number of colonies reported to be lost. Still, looking at bar graphs of the total colonies lost for each winter will illustrate which reasons for death or reasons for beekeeping lead to the most widespread losses in the United States.



This bar graph tells us the total winter colony losses per year, grouped by operation type. From this bar graph we can conclude that the largest amount of colony losses occur under beekeepers who sell honey or use bees for pollination. The only reason for the variation from year to year is the number of survey responses, and pollination and selling honey consistently have the highest colony losses. This raises the question whether the percent loss for each operation type varies from the overall decreasing trend per year.

Here is the bar graph for total winter colony losses per year, grouped by reason for colony loss:



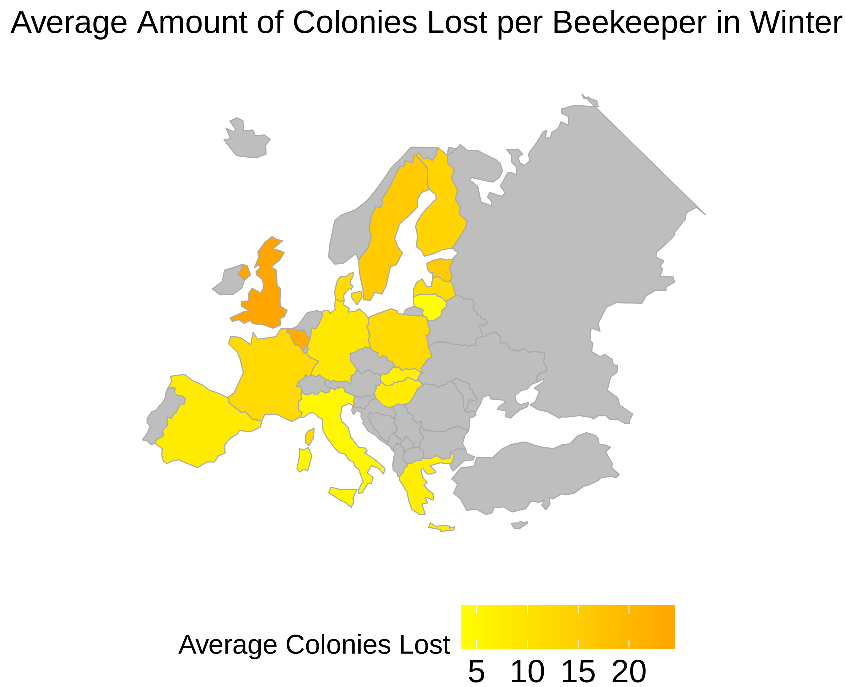
The differences between total colony losses is less evident than the category of beekeeping operation. Nevertheless, causes of colony loss such as the Varroa parasite, queen failure, and pesticides consistently have the greatest losses each year. This raises the question whether the percent colony losses as a result of these reasons are decreasing each year as well.

3 Technical Summary

3.1 Acknowledgements

I chose to omit data from 2013 from the bar graph for category of operation, as the survey did not have "Sell Honey" as a listed option that year. Thus, the responses were wildly different than the other years and would skew the results. I had to remove the year 2011 from the bar graph of reason for colony loss, because statistics on reason for colony loss were not collected that year. I am obligated to remove 2011 from the general scatter plot of percent loss versus year, because I will later run an **ANCOVA** test, on the grounds of being able to visualize whether the covariate of reason for colony loss affects the association between percent loss and the year. I will not be able to run an **ANCOVA** for category of operation because the survey did not collect statistics on that variable before 2011.

3.2 Analysis of Europe



This choropleth map details the average number of colonies lost per beekeeper in the winter in each country surveyed by the EFSA. These statistics were collected cumulatively from 2012 to 2014, so it is impossible to identify the trend of winter mortality for each year. However, it does demonstrate which regions were more vulnerable to colony loss during this era. I decided to run an **ANOVA** test to see if there was significant evidence that at least one of the countries have different population means of colonies lost per beekeeper.

$H_o : \mu_1 = \mu_2 = ... = \mu_{16}$

H_a : At least one of these means are not equal

$\alpha = .05$

The data is heavily skewed right by the surplus of beekeepers who responded with having lost zero colonies. I will proceed anyway because each country has over 30 responses, which means according to the central limit theorem the data will approach normal distributions.

The standard deviations are not relatively equal, because the largest sample standard deviation (29.79) is greater than two times the lowest sample standard deviation (10.58). I will proceed nonetheless.

I will assume the data was randomly selected so that it is unbiased and independent.

$$N = n_1 + n_2 + ... + n_{16}$$

$$N = 274 + 265 + ... + 228 = 4758$$

$$T = n_1(\bar{x}_1) + n_2(\bar{x}_2) + ... + n_{16}(\bar{x}_{16})$$

$$T = 274(22.7) + 265(11.9) + ... + 228(15.7) = 54736.03$$

$$\bar{\bar{x}} = \frac{T}{N} = \frac{54736.03}{4758} = 11.45$$

$$SSTr = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + ... + n_{16}(\bar{x}_{16} - \bar{\bar{x}})^2$$

$$SSTr = 274(22.7 - 11.45)^2 + 265(11.9 - 11.45)^2 + ... + 228(15.7 - 11.45)^2 = 118927$$

$$SSE = (n_1 - 1)S_1^2 + (n_2 - 1)S_1^2 + ... + (n_{16} - 1)S_{16}^2$$

$$SSE = (274 - 1)29.8^2 + (265 - 1)18.7^2 + ... + (228 - 1)23.1^2 = 1719555$$

$$MSTr = \frac{SSTr}{k-1} = \frac{118927}{15} = 7928$$

$$MSE = \frac{SSE}{N-k} = \frac{1719555}{4742} = 363$$

$$F = \frac{MSTr}{MSE} = \frac{7928}{363} = 21.9 \text{ with } df = \frac{k-1}{N-k} = \frac{15}{4742}$$

$$\text{P-value} = P(F > 21.9) \approx 0$$

Since my P-Value ($\approx$ 0) $<$ α (.05), I reject my null hypothesis. Therefore there is significant evidence that at least one of the true average winter colony loss per beekeeper per country are not equal. We can run a Tukey-Kramer post hoc test to display which means are significantly different.

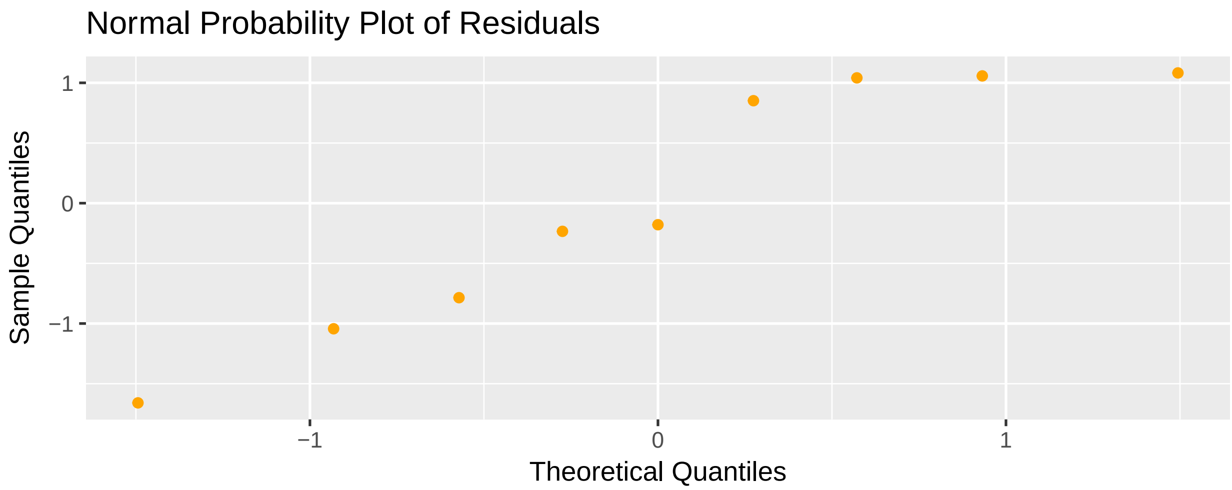
```
          Winter_mortality groups
ENGLAND          24.378882      a
BELGIUM          22.660584      a
SWEDEN           15.736842      b
ESTONIA          15.625352      b
FINLAND          13.718213     bc
FRANCE           12.236170    bcd
POLAND           12.220963   bcde
DENMARK          11.890566   bcde
LATVIA           11.856502   bcde
GERMANY           8.994366   cdef
```

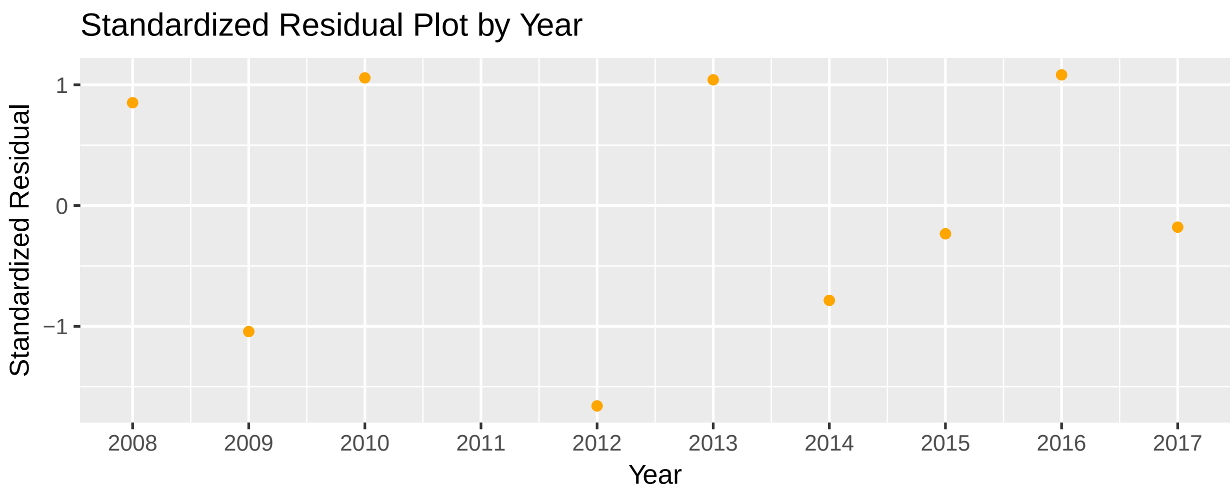
```
SPAIN                8.185596   defg
HUNGARY              7.991354   defg
GREECE               7.660920   defg
SLOVAKIA             7.391185    efg
ITALY                5.546512     fg
LITHUANIA            3.542857      g
```

This Tukey-Kramer test indicates England and Belgium have significantly different means than all other countries, besides each other. So, in countries such as these it would be expected that beekeepers, on average, lose more colonies during the winter.

3.3 Analysis of the United States

We can run a model-utility test to determine whether there is significant evidence there is a relationship between percent loss and the year each winter, according to the National Management Survey. I will not do the full write-up, because that would overcrowd the rest of the paper. However, I will acknowledge my assumptions and clarify the result.

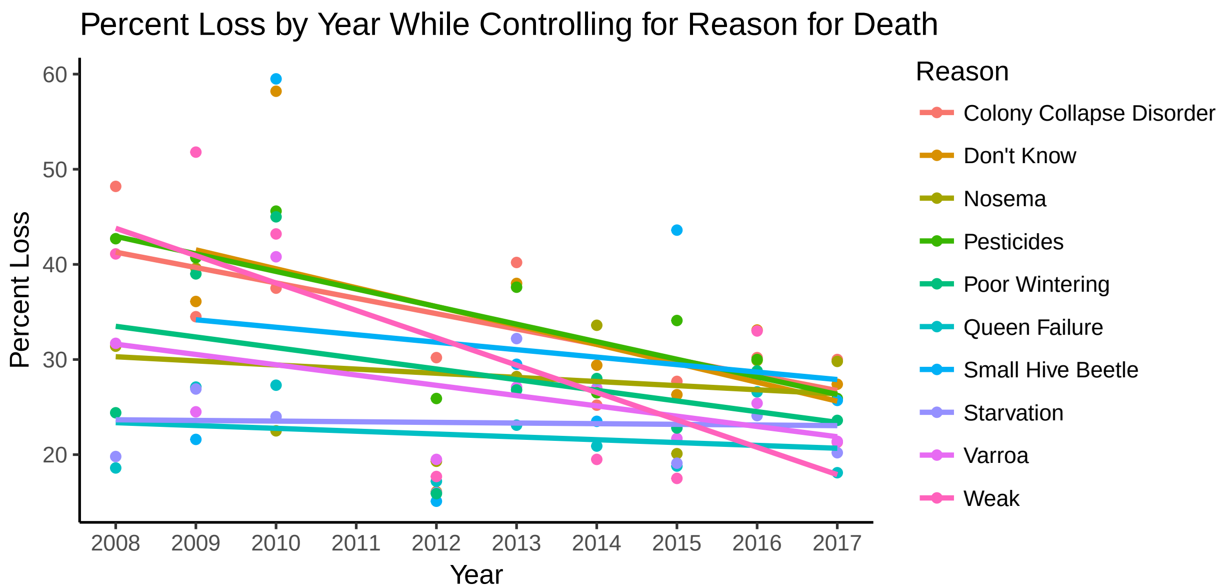




```
Call:
lm(formula = `Percent Loss` * 100 ~ Year, data = ProjectBeeYear)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.9408 -2.7850 -0.5513  3.5708  3.7371 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) 2398.4092   857.2140   2.798   0.0266 *
Year          -1.1779     0.4259  -2.766   0.0279 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.809 on 7 degrees of freedom
Multiple R-squared:  0.5221,Adjusted R-squared:  0.4539 
F-statistic: 7.649 on 1 and 7 DF,  p-value: 0.02787
```

The scatter plot between percent loss and year appears relatively linear, with $R = -.723$. The standardized residuals are independent and equally distributed, and the normal probability plot appears to be relatively linear. Therefore the standardized residuals are approximately normally distributed. Since my P-Value (.03) $< \alpha$ (.05), I reject my null hypothesis that there is no relationship between year and percent loss. This means this is a useful model to describe percent loss from the winters of 2008 to 2017 in the U.S. I may now run an **ANCOVA** test to determine whether the covariate of reason for death adjusts the association between year and percent loss. My assumptions for the model are already met, and the survey is assumed to be randomly selected.

```
Analysis of Variance Table

Model 1: `Percent Loss` ~ Year * Reason
Model 2: `Percent Loss` ~ Year + Reason
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     68 0.46470                           
2     77 0.51795 -9 -0.053243 0.8657 0.5598
```



Since our P-value (.560) $>$ α (.05) I fail to reject my null hypothesis. Therefore the covariate of reason for death does not adjust the association between year and percent loss. This is demonstrated by the scatter plot with multiple regression lines; the regression lines are all relatively similar.

3.4 Explanation of Findings

Although the negative relationship between percent loss and year is statistically significant, that does not mean the trend will continue. This relationship merely implies that the percent of colonies lost in winter has been decreasing since 2008. From my analysis of beekeepers in Europe and of operation category in the United States, I was able to answer where and under which beekeepers honey bees are the most susceptible to colony loss over the winter. From my linear regression model of percent loss of colonies in the United States and my ANCOVA test I was able to display what the general, negative trend of percent loss has been since 2008 and that the cause of colony loss does not affect that percent.

4 Conclusion

Honey bees are still facing the looming threat of depletion. The trend over the last ten years appears to have improved, but beekeepers claim the die-off is still not at a manageable level. Through my analysis I was able to discover what the leading causes of colony loss were, where colony loss was occurring, and under which beekeepers the losses were happening. This trend may not continue, nor may it entirely portray the total population of beekeepers. However, my analysis depicts a relevant account of a large population of beekeepers over recent years. I hoped to meet my assumptions better for my initial **ANOVA** test. In the future, I would hope to discover whether this trend continues and can be used as a predictive model, be able to use raw data from the National Management Survey, and determine if the category of beekeeping operation affects the relationship between year and percent loss.

Works Cited

"Bee Informed Partnership." *National Management Survey*, Bee Informed Partnership, 2017, bip2.beeinformed.org/survey/.

Borenstein, Seth. "Survey Finds U.S. Honeybee Losses Improve from Horrible to Bad." The Seattle Times, 25 May 2017.

EFSA. "Honey Bee Seasonal Mortality." 16 Apr. 2016.

Oldroyd, Benjamin P. "What's Killing American Honey Bees?" *PLOS Biology*, Public Library of Science, 12 June 2007, journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0050168.

VanEngelsdorp, Dennis, and Marina Doris Meixner. "A Historical Review of Managed Honey Bee Populations in Europe and the United States and the Factors That May Affect Them." *Journal of Invertebrate Pathology*, vol. 103, 2010, pp. 80–95., doi:10.1016/j.jip.2009.06.011.

Walshe, Sadhbh. "Battle to Save the Bees Rages across US." *The Guardian*, Guardian News and Media, 24 May 2013, www.theguardian.com/sustainable-business/battle-save-bees-rages-across-us.