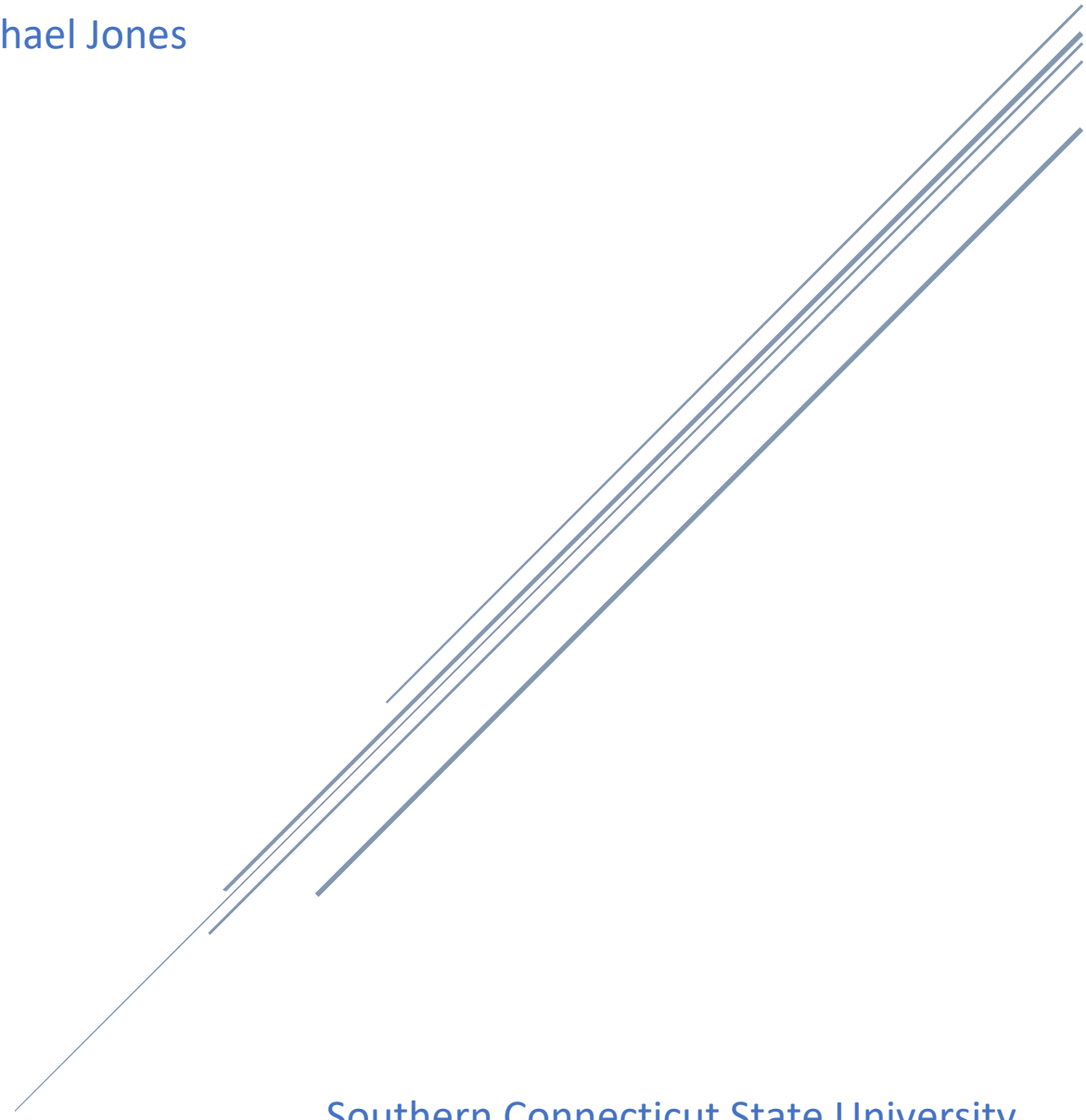
A DATA ANALYSIS ON THE POTENTIAL EFFECTS OF NEONICOTINOID PESTICIDES AND HONEY BEES

By Michael Jones



Southern Connecticut State University
CSC 400 Final Report

Michael Jones

December 12, 2018

Design Document Revision and Final Report

Introduction:

If you research the word “honey bees” or “pesticides” there is a good chance you will see an article in the first few results that talk about honey bees dying from pesticides or pesticides killing honey bees. These pesticides that are being called out are, Neonicotinoids which are chemically related to nicotine. Neonicotinoids are a relatively new class of pesticide, but are the most widely used class of insecticides used worldwide. They have been around since the early 1990s. According to researchers at Penn State University, their use has increased more than 11-fold since 2003. Neonicotinoids are easier for farmers to use than the traditional method of spraying crops. Another reason why they are so popular is because of their water solubility, this allows the pesticides to be applied to soil and be taken up by plants. With all of this debate and popularity of neonics I wanted to take a look at this myself.

Findings by previous studies:

- Over a period of 12 weeks, researchers exposed bees to clothianidin which is the most common neonicotinoid found at the Canadian farms. Each round of exposure had smaller and smaller amounts of pesticide which would represent nature as rain washes away the compound. Bees in the lab were exposed to lower amounts of pesticide over time than bees near the farm, the insects still suffered. For example, worker bees in the lab lived three-quarters as long as those near the farms. The pesticides not only reduced a bee’s chance of survival but affected its natural defense systems. Bees use social immunity; a tactic bee use to clean out dead or sick insects from the nest. Bees in colonies treated with clothianidin displayed less and less of this behavior over time, which means more sick bees were infecting, and staying in, the nests.

- A team of European researchers established 33 sites growing rapeseed in Germany, Hungary and the UK. These were randomly assigned to either be treated with one of two different Neonicotinoids, or none at all. The team looked at honeybees and two wild bee species, bumblebees and solitary bees. Results differed between locations and species, but overall, they discovered that honeybee hives were less likely to survive over winter, while the wild bees reproduced less. It's not that the pesticides directly kill bees. Instead low-level exposure makes them more vulnerable, especially if there are other environmental factors or diseases already affecting the hive.

- Researchers have observed for the first time how a common neonicotinoid pesticide affects bee behavior inside the nest. Chronic exposure to neonicotinoids, which are neurotoxic, is known to impair foraging bees’ ability to navigate and recognize flowers.

- Honeybee populations haven’t “crashed” in the United States or elsewhere. Honeybees are not going "extinct." Crops are not "in trouble." The overall population of honeybees in the US, Canada and Europe has held steady or increased slightly since the widespread adoption of neonics in the 1990s. The US honeybee population hit a 22-year high in 2016, according to the figures released by the USDA before dipping slightly last year, and globally are at an all-time high.

These are just a few articles out of many that showcase the problems of neonicotinoids. As you can tell just from four articles there is a conflicting story.

Important Features:

An important feature in this program is going to be the graphs. I want to make plenty of graphs showing correlation between data in order to draw some conclusions. I want to use the best data visualization tools I can to represent this data as clearly and precisely as possible. This can be allowed through tools such as ggplot2, plotly, seaborn, and many more. Another important feature is the detailed writeups after each question. These will be a great compliment to the graphs. I will also implement a predictive algorithm to be able to predict future trends pertaining to the posed questions. The advantages of this that I can think of is that I am going to try to make the graphs very user friendly. Also, the predictions will give users a nice guideline as to what to expect if Neonicotinoids are to be continued. By this I mean if are easy to read this can potentially include animations. The potential users of this program would be anyone working in the data analysis/data visualization area. Another group is anyone who is working or interested in wildlife and wildlife protection. The most important group is beekeepers. The final group is simply the public. The general population should always be up to date with important information about the environment. All of these users will benefit by this program by being aware of any problems if there are any.

Architecture:

This program will run on a desktop, given the user has python installed it will also have the ability to be put on a website such as Kaggle. That way users can see code and outputs. They would also be able to see the data as well. The project will be completed in a Jupyter notebook due to the fact that I can code, get the output/visualization directly below and make a writeup all in a small area. The major inputs will be the specific part of the data that I need for the question I am trying to answer. Specifically, this will come from a CSV file containing all of the data which will be described in the “Data” section. The major outputs will be the graphs and predictions provided by a predictive algorithm.

Use Cases:

A) Do neonicotinoid pesticides impact total production, yield per colony and number of colonies?

In order to draw a conclusion for this one must run the program or look at provided graphs and draw a conclusion based off of that. Or they can examine the data set provided then look at the final outcomes provided by the program. If the user is just doing research on the topic they can simply examine the graphs and writeups provided.

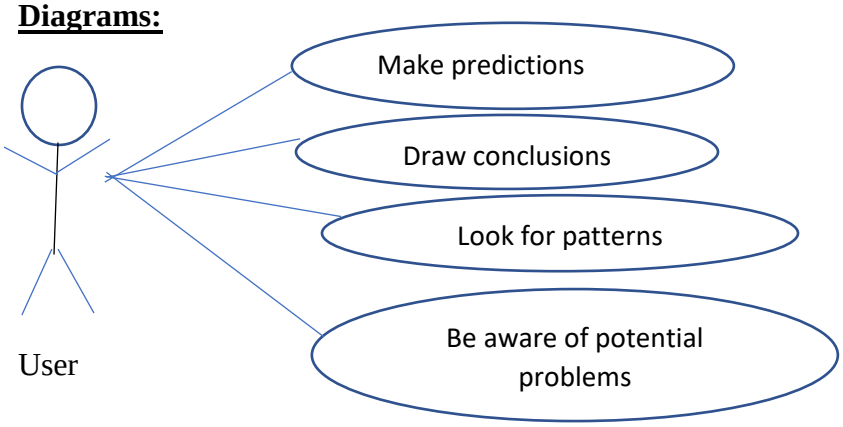
B) What specific neonicotinoids pesticides were used the most per state over a course of a specific amount of years?

Graphs such as scatter plots will be generated for you to examine.

C) Has production value and stocks increased or decreased with neonicotinoids pesticide usage?

Again, similar to the solution previously mentioned, graphs will be displayed for you to examine.

Diagrams:



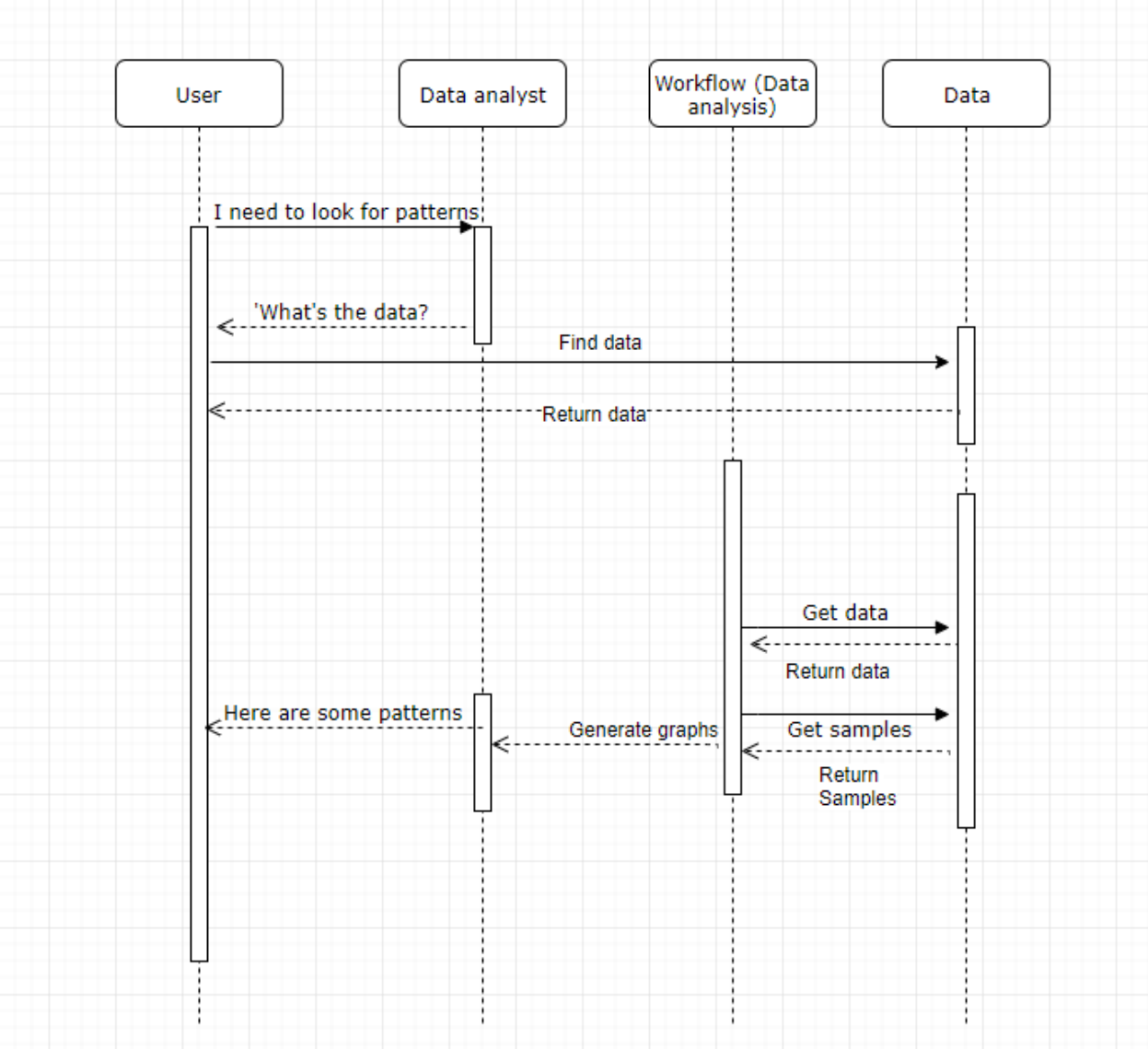
The potential user would be doing research on either honey bees or Neonicotinoid pesticides. If the user is working for a company they can request a Data Analyst to do some research on the topics. This would result in data being provided from the user, then the analyst would do some analysis and come back with some prediction, conclusions, patterns, and or be aware of anything else notable.

If the user is someone who does not work for a company but, has an interest in the environment, they can do some research and potentially find the results if they are posted online.

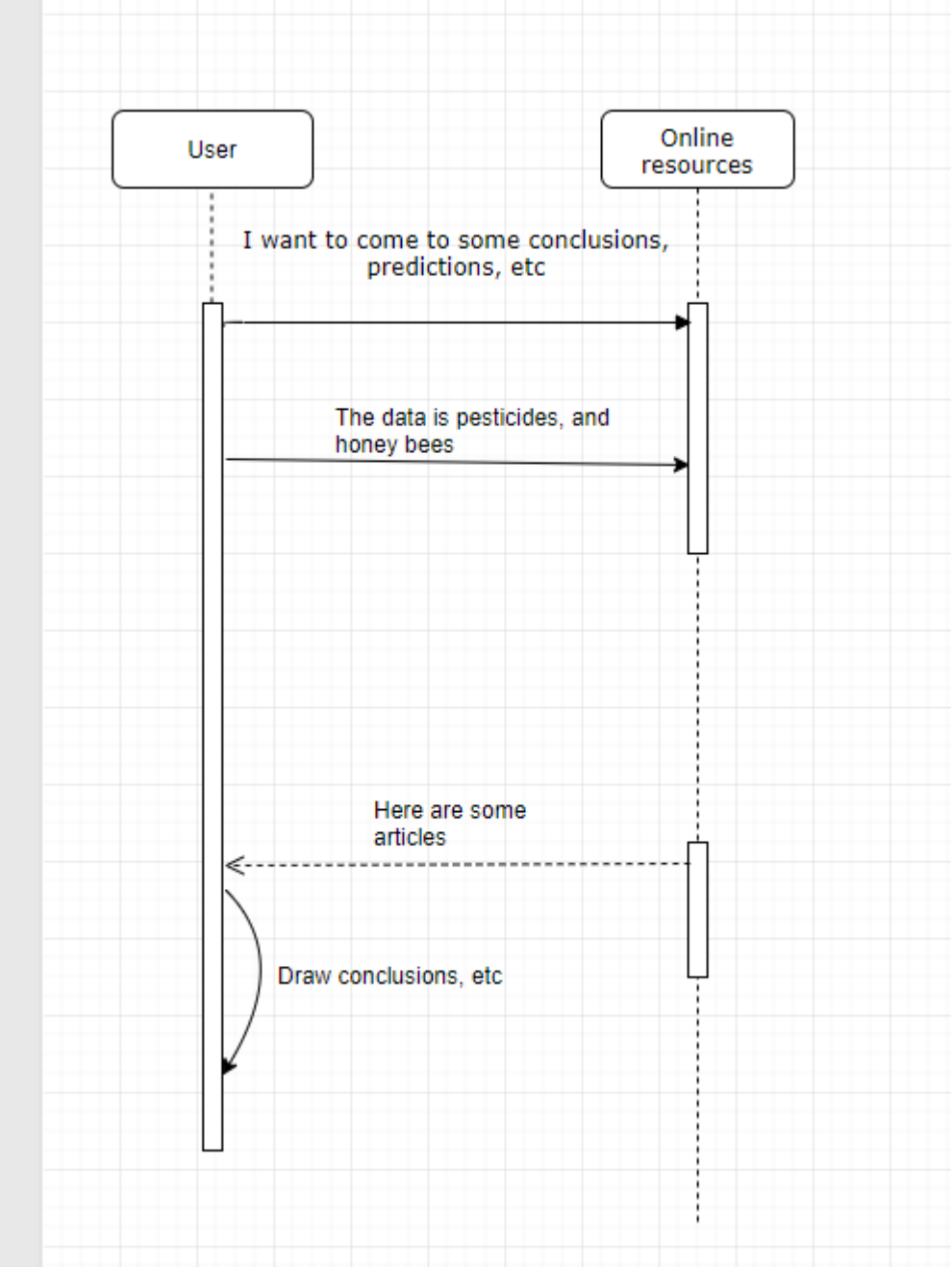
With the graphs and writeups this user will too be able to come up with some predictions, conclusions, patterns, and they will be more aware of the current trends.

Sequence Diagram:

This diagram will show what the process would be like for someone who would be working for a company.



Below in the sequence diagram for someone who is not working for a company.



Data:

This dataset is a combination of two datasets. One is by Honey Production in the USA, between 1993-2017. It also has data from USGS's Pesticide National Synthesis Project. This would allow me to do some statistics between Honey Production and the use of Neonicotinoid pesticides. Some samples of the data is as follows. Neonicotinoid use in the USA began around 2003, but honey production declines started in earlier years. Some states, such as Kansas, have seen devastation in the bee colonies starting around 2003, just Neonicotinoid pesticides usage began. Other states such as North Dakota, have approximately stable numbers of bee colonies, despite increased use of Neonicotinoids. The data is formatted as so:

State, Numcol, yieldpercol, totalprod, stocks, priceperlbs, prodval, year, stateName, Reigon, nCLOTHIANDIN, nIMIDACLOPRID, nTHIAMETHOXAM, nACETAMIPRID, nTHIACLOPRID, nAllNeonicotinoid

1. Numcol (float): Number of honey producing colonies. Honey producing colonies are the maximum number of colonies from which honey was taken during the year. It is possible to take honey from colonies which did not survive the entire year
2. Yieldpercol (int): Honey yield per colony. The unit is pounds.
3. Totalprod (float): Total production (numcol x yieldpercol). The unit is pounds.
4. Stocks (float): Refers to stocks held by producers. The unit is pounds
5. Priceperlbs (float): Refers to average price per pound based on expanded sales. The unit is dollars.
6. Prodvalue (float): Value of production (totalprod x priceperlb). The unit is dollars.

From USGS Data

1. nCLOTHIANIDIN (float): The amount in kg of CLOTHIANIDIN applied.
2. nIMIDACLOPRID (float): The amount in kg of IMIDACLOPRID applied.
3. nTHIAMETHOXAM (float): The amount in kg of THIAMETHOXAM applied.
4. nACETAMIPRID (float): The amount in kg of ACETAMIPRID applied.
5. nTHIACLOPRID (float): The amount in kg of THIACLOPRID applied.
6. nAll Neonicotinoid (float): The amount in kg of all Neonicotinoid applied.

In an excel file the entire data set is 17 columns by 1133 rows. Most of the data is going to be used at the same time. The only parts of the data that will be used later on is the stocks and prices. That is because first I want to determine the effect of Neonicotinoid pesticeds and honeybees first.

Documentation:

Below are a few small scripts that were core to my project:

1. g = sns.lmplot(x="year", y="yieldpercol", hue="post-neonics(2003)",

truncate=True, size=5, data=df[df.StateName =='Illinois'], markers=["o", "x"])

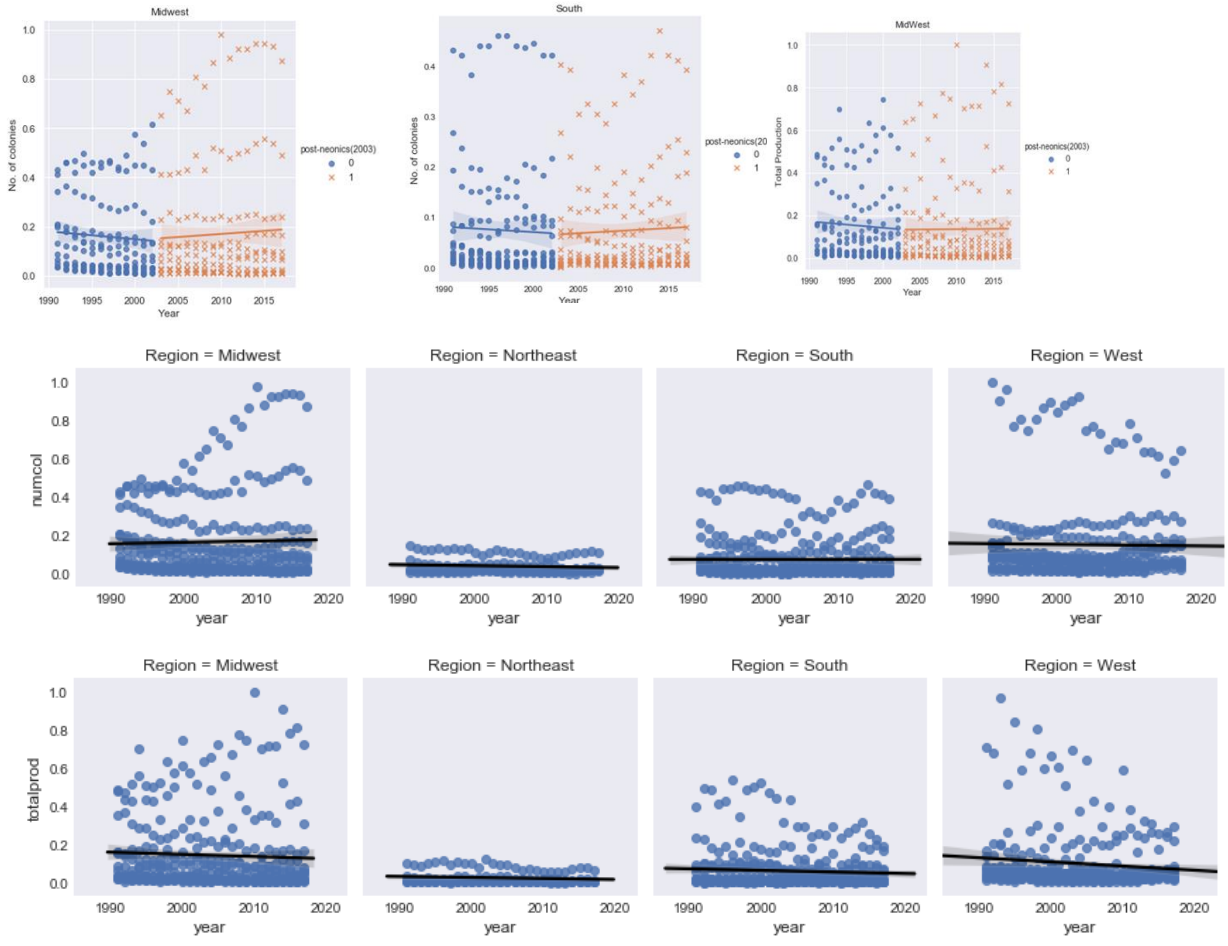
2. data3=df[df.StateName=="North Dakota"].sort_values('year')

```
trace1 = go.Scatter(
   x = data3.year,
   y = data3.nIMIDACLOPRID,
   mode='markers',
   marker=dict(
      color = data3.nIMIDACLOPRID,
      colorscale='Jet',
      showscale=True
```

```
    ),
    text=data3.nIMIDACLOPRID
)
fig = go.Figure(data=data, layout=layout)
iplot(fig, filename = "Scatterplot")
```

These are just two, the second one is abbreviated as well. These were core because these were the two main plots I used to represent data, linear regression model fit and a scatter plot. I used mainly plotly, matplotlib, and seaborn for this project. Plotly allowed me to make the graphs interactive and seaborn was great for statistical data visualization. I also used pandas throughout this project as well. Pandas allowed to to make dataframes that would come in use for certain graphs, such as breaking the data down to before and after 2003.

Screenshots:



The screenshots above are chosen because they provide a summery of my findings which will be talked about later. These screenshots show total production of honey decreasing in ever region, and number of colonies increase in almost every region.

User/Deployment guide:

If you simply want to examine the notebook and not execute any code you can just look at it through github. Now not all the graphs will load when you click on the project. So you will need to click on the project and then a small popup will be displayed in the top right hand corner asking if you want to view with nbviewer. Just simply click on that. After you do that you will have full access to all of the graphs.

If you want to run code you will first need to download the notebook. Once that is done you will need to upload it to your Jupyter Notebook. The first code block you will see is a bunch of imports. This is a guide to what needs to be pip installed. You need to open Anaconda prompt and pip insall plotly, pandas, bubbly, matplotlib, seaborn, numpy, and scipy. After that you will have everything you need to execute these code blocks.

State of Implementation:

The features currently include graphs, writeups related to the graphs, regression graphs, and user friendly graphs. All of which are fully complete. I have all of these in my project and all of which are used multiple times. I also do have a conclusion drawn which is that Neonictionoids may be affecting a honeybees ability to properly navigate or forage. Which was suggest by a common trend of declineing production of honey after 2003, which was the height of neonic usage.

Testing and Evaluation:

At the current moment this project has met all of the objectives I have set out for the in relation to the time I had for the class. I mentioned how I wamted a plentiful amount of graphs, writeups pertaining to the graphs, regression graphs, and user friendly graphs. At the end of the project I do have all of these. There are a lot of graphs, almost every graph has an explaination blow it, I even included checkpoints to summerize finding. I do have a few regression graphs, and the graphs are user friendly because you can hover over most of them and it will show specific values, the graphs are clearly labeled, and are also broken down to show a clear difference betweem 2003 and after 2003. The system is very usable I would say almost anyone is able to look at the project on github and have an understandning as to what is going on. I tested the implementation by making sure the graphs were correct. There was a few times where I wanted to make sure what I was coding was correct and the generated graph was correct so I set out to do some by hand and found that they were correct.

Findings:

In this project I found that neonics are affecting honey bees in one way or another. I can not say for sure that honey bees are not dying or are dying but, there is a possibility that neonics are slowing their growth. This is suggested by a slight growth in states and slight decrease in some. Neonics can also be effecting the bees ability to learn, navigate, and repoduce properly. This can be especially connected to Imidacloprid since in most states it is the most commonly used pesticde. Another fact is that yield per colony and total production of honey is decreasing even in the height of pesticde usage, which would back the theory that neonics may be affecting a honey

bees ability to forage, naviagate, or communicate. A change in any of these would likely result in a decrease in production.

Lessons learned:

This project started going well once I mapped out my workflow. Once I started at the very beginning with a simple graph of total pesticide usage almost everything started falling into place. I painted a big overall picture then got down to the smaller details with states and specific compound popularity. I also learned to keep it simple. Sometimes going overboard with visualizations are not necessary.

Looking back if I had to do it again the main thing, I would do differently is avoid making graphs too fancy if they do not need to be. I included and animated graph that showed two neonic compound popularity against each other in a specific state. The neonics were represented with a bubble that got bigger as the neonic was more popular. It was not completely necessary and a simple scatter plot would have sufficed. This took a while to figure out how to code and in the end was not necessary. The time spent to do this could have been allocated to finding another compatible dataset to work with, such as disease data.

Version 2:

If I had an extra month or two, I would definitely include a second dataset. It would be a data that included pesticide usage and bee data from another county or a dataset that had information as to bee diseases in the United States so I can compare it against bee colony numbers and honey production numbers. Disease information would be helpful because it would allow me to see if there was another factor other than pesticide usage playing a role. A dataset from another country that had bee, honey, and pesticide usage would be beneficial because it would allow me to compare two countries and I would be a much better view. Making a conclusion as to how pesticides are affecting honey bees would be easier with either one of those options.

Lifelong learning:

Some of the things that I had to learn independently for this project is visualization tools, specifically plotly since this tool had the potential to get very in depth. There is a massive amount of possibilities with plotly and I definitely had to take the time to read the documentation for it. I would have liked to make some of the graphs look neater with plotly but I just did not have the time. With all of the possibilities, it would have taken too long. Also, I never used pandas before this project so I had to research that for a while. This was needed so that I could properly work with data frames throughout the project. Pandas was vital to this project and if I did not learn it, the project probably would not have been completed. This was the first time I used Jupyter notebook. This, however, was pretty easy to learn. I learned the ability to not only code in the notebook also how to add descriptions and even add Latex. The main challenge was using pandas a plotly together. This easily took up a good amount of my time. There was a problem for a while with my data arrays but, after researching the problem I figured out that I needed to reset my index. After I figured out the solution to that problem using the two together

was easier but, I still had to do a fair amount of research after. The final and biggest lesson I had to teach myself was how to tell a story with data. I have worked with data in the past but, I never had to tell a story about it. I had to first state why I have to analyze it, what information can I get from this data and or graph, and finally, at the end of the analysis, what can be decided. This was definitely something that I was not used to but, in the end found it rather enjoyable. The process of going from a CSV file filled with data to a notebook that tells a story about a real-world problem was a great learning experience. The final thing I had to teach myself was the process of data cleaning. I never worked with data cleaning an excel file so I had to do some research about the process but, it turned out to be not too bad of a task.

Citations for previous studies:

Dean, Signe. "There's Now Very Strong Evidence We Really Are Killing Our Bees." *ScienceAlert*, ScienceAlert, www.sciencealert.com/we-really-might-be-killing-our-bees-according-to-new-evidence.

Charles, Dan. "Pesticides Are Harming Bees - But Not Everywhere, Major New Study Shows." *NPR*, NPR, 29 June 2017, www.npr.org/sections/thesalt/2017/06/29/534852611/pesticides-are-harming-bees-but-not-everywhere-major-new-study-shows.

Entine, John. "The Bee Apocalypse Was Never Real; Here's Why." *Millennials, the Deaf Generation? | American Council on Science and Health*, www.acsh.org/news/2018/04/17/bee-apocalypse-was-never-real-heres-why-12851.

Dengler, Roni. "Neonicotinoid Pesticides Are Slowly Killing Bees." *PBS*, Public Broadcasting Service, 29 June 2017, www.pbs.org/newshour/science/neonicotinoid-pesticides-slowly-killing-bees.