ANALYSIS OF PESTICIDE USAGE AND CLIMATE CHANGE IMPACTS ON U.S. HONEY PRODUCTION

Shannon McFarland

6/3/23

DSC 530

Introduction:

Honeybees (*Apis mellifera*) serve as a key player in the agricultural field as they help pollinate $15 billion worth of crops annually in the United States alone. They additionally produce $3.2 million in honey annually. (USDA, 2018)

Unfortunately, the honeybee populations are currently under threat to dangers such as urbanization, overuse of pesticides, and climate change.

With climate change posing to increase the global temperature by a few degrees within the next couple degrees, this could negatively impact's the honeybee hives' seasonal lifecycles. (Calovi et al., 2021)

For my project, my goal is to analyze the effect pesticide usage and seasonal temperature in the United States have on a hive's productivity. If the pesticide usage and warming temperature do have an effect, then I predict they would demonstrate a negative relationship with a hive's ability to produce honey.

Datasets Utilized:

The data I chose to analyzed for this project is a combination of two different datasets acquired through Kaggle. These datasets are "Honeybees and Neonic Pesticides" and "Average Monthly Temperatures by US States".

The raw data for the "Honeybees and Neonic Pesticides" dataset was originally acquired by Zmith through the USDA National Agricultural Statistics Service's yearly report on honey production in the United States. The raw data also in the dataset related to pesticide usage was acquired from the Pesticide National Synthesis Project by the U.S Geological Survey (USGS).

The temperature data obtained from the "Average Monthly Temperatures by US States" dataset ranges from 1995 to 2015.

Data Source References:

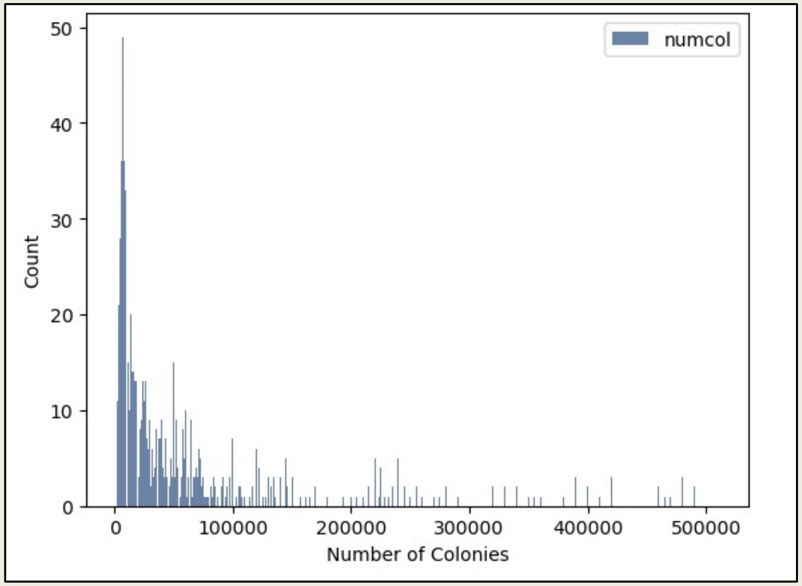
Wong, J. (2022). Average Monthly Temperatures by US State. [Data Set]. https://www.kaggle.com/datasets/justinrwong/average-monthly-temperature-by-us-state.

Zmith, K. (2018) Honeybees and Neonic Pesticides. [Data Set]. https://www.kaggle.com/datasets/kevinzmith/honey-with-neonic-pesticide.

Variables Utilized in Study:

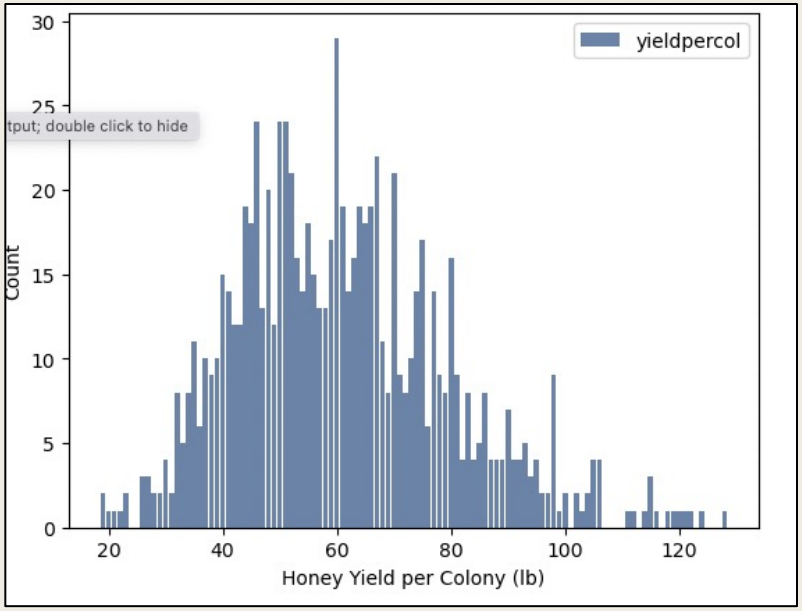
- **numcol =** Number of reported honey producing colonies in the state; This variable will act as a dependent variable.
- **yieldpercol =** honey yield per colony (lbs); This variable will act as a dependent variable.
- **year =** year that honey production, pesticide usage, and temperature data was collected; This variable can be used in a time series analysis to showcase how the other variables change overtime.
- **Region =** sorts state data into the Midwest, Northeast, South, West
- **nAllNeonic =** The total of amount (kg) of Neonicotinoid pesticides applied in the state for the year. Neonicotinoids is a class of synthetic, neurotoxic insecticide used in both agriculture and landscaping. This variable may have an inverse, negative effect on the number of producing colonies and ultimately the yield per colony.
- **Avg_springtemp =** Average state temperature (°F) for early spring (March/April). Early springtime is a critical time period in the success of the hive. During this time, the queen is laying eggs to increase the hive population for the summer. The hive is also susceptible to starvation from using up their wintertime food stores and Varroa mites infestations. (Grozinger & Anton, 2022) With an increasing temperatures due to climate change, springtime blooming could occur earlier than normal and impact the hives.
- **Avg_falltemp =** Average state temperature (°F) for late summer/early fall (August/September). The temperature data is paired with the numcol and yieldpercol variables of the following year. This time period is critical for the success of the hive going into the wintertime. This is when the queen is laying eggs for bees that have the physiology to survive the winter. It is important for these bees to be healthy and disease free or they will die during the cold temperatures. If the fall temperatures are too warm, it causes the hive to start preparing for winter too late. (Grozinger & Anton, 2022) Therefore, this variable may have an inverse relationship with the numcol and yieldpercol variables of the following year.

Data Analysis: Histograms



Histogram showcasing frequency of the **number of honey producing colonies.** Distribution is heavily skewed to the right. The variance and standard are also incredibly high. The large spread may be due to certain states such as California having a lot more agricultural land, allowing them to have more bee colonie, compared to other states.

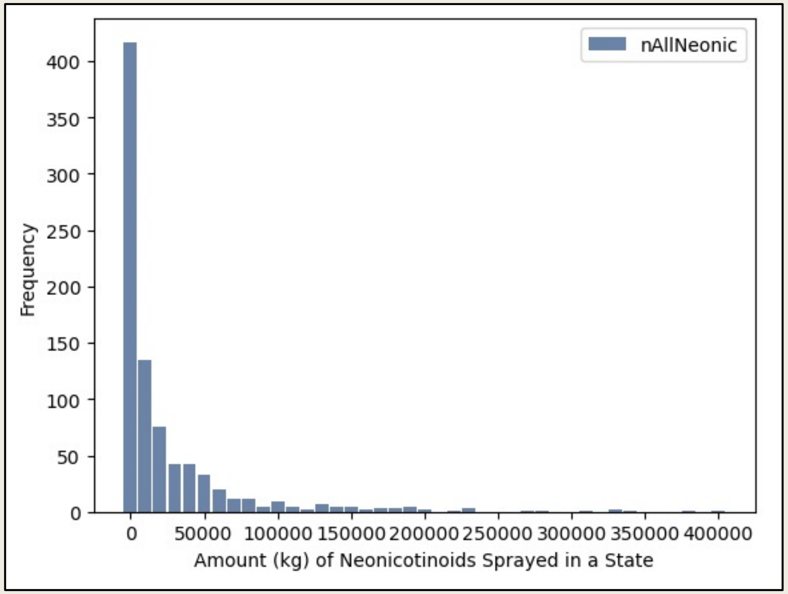
```
mean: 62090.48178613396
variance: 8355385921.061829
standard deviation: 91407.80011061326
```



Histogram showcasing frequency of the **honey yield per colony.** The data follows rough normal distribution with a slight skew to the right.

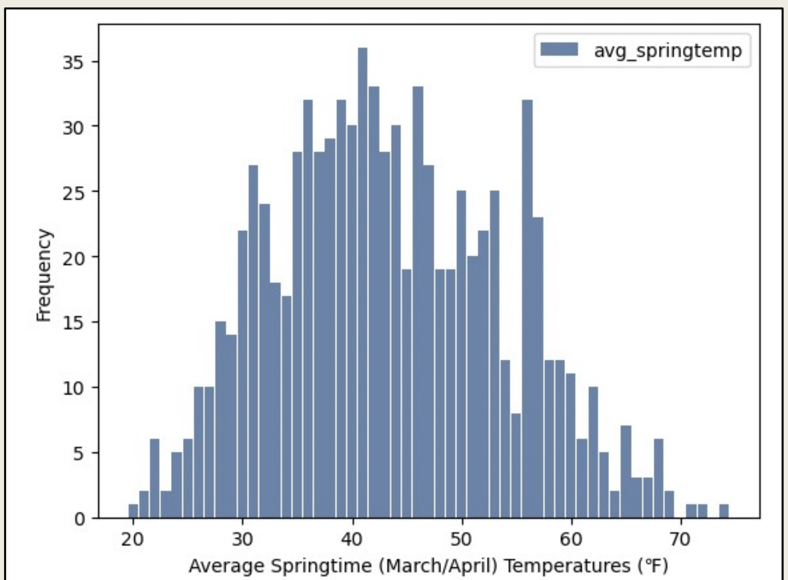
```
mean: 60.990599294947124
variance: 353.97167622865857
standard deviation: 18.8141350114391
```

Data Analysis: Histograms



Histogram showcasing the frequency of the **amount (kg) of neonicotinoids sprayed in a state.** Similar to the number of colonies variable, this variable is also is heavily skewed to the right. This large spread in values may also be due to the large ranged of usable land across the different U.S. states.

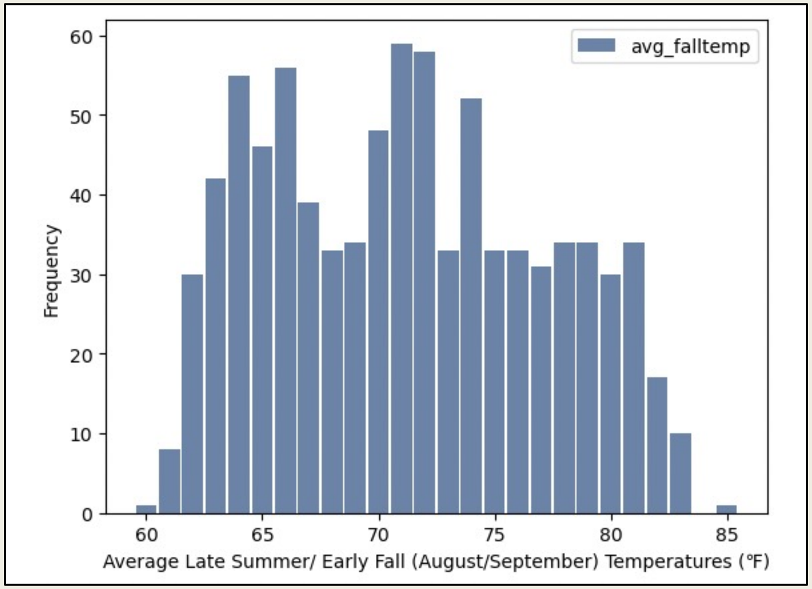
```
mean: 24930.175558166866
variance: 2320085455.766065
standard deviation: 48167.26539638788
```



Histogram showcasing the frequency of the **average springtime temperatures** in a state. The data follows a rough normal distribution with a slight skew to the right.

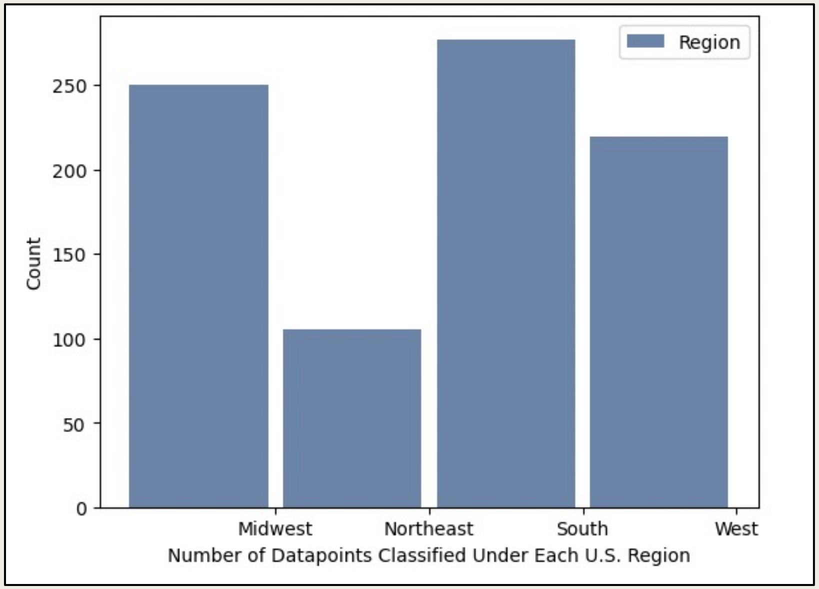
```
mean: 43.38090481786134
variance: 109.1748731803414
standard deviation: 10.448678058986285
```

Data Analysis: Histograms



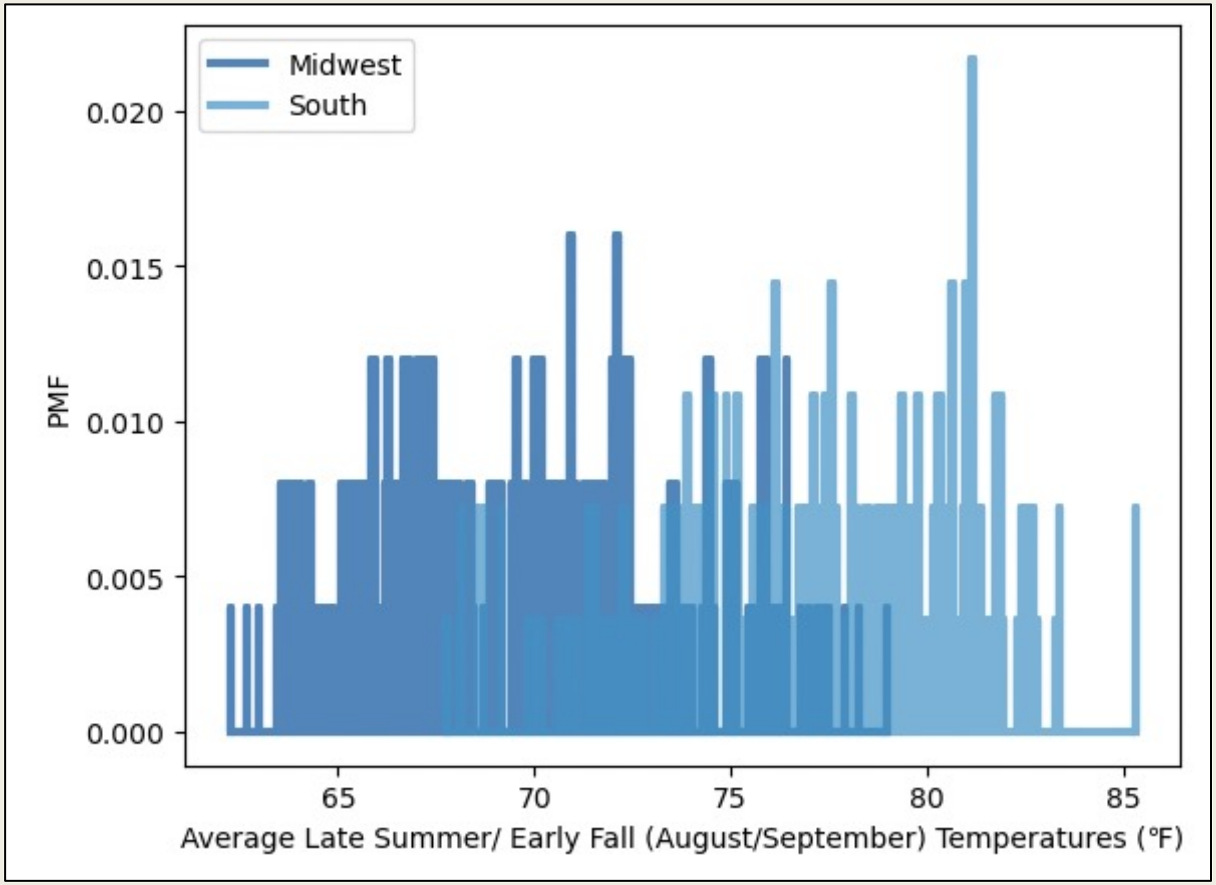
Histogram showcasing the frequency of the **average late summer/early fall temperatures** in a state. The data follows a bimodal distribution as it has two main peaks towards the left. The two peaks may be due to regional temperature differences across U.S. For example, states from the souths are more likely to be warmer in the months of August/September than northerner states.

```
mean: 71.23061104582852
variance: 34.384576567360185
standard deviation: 5.863836335314978
```



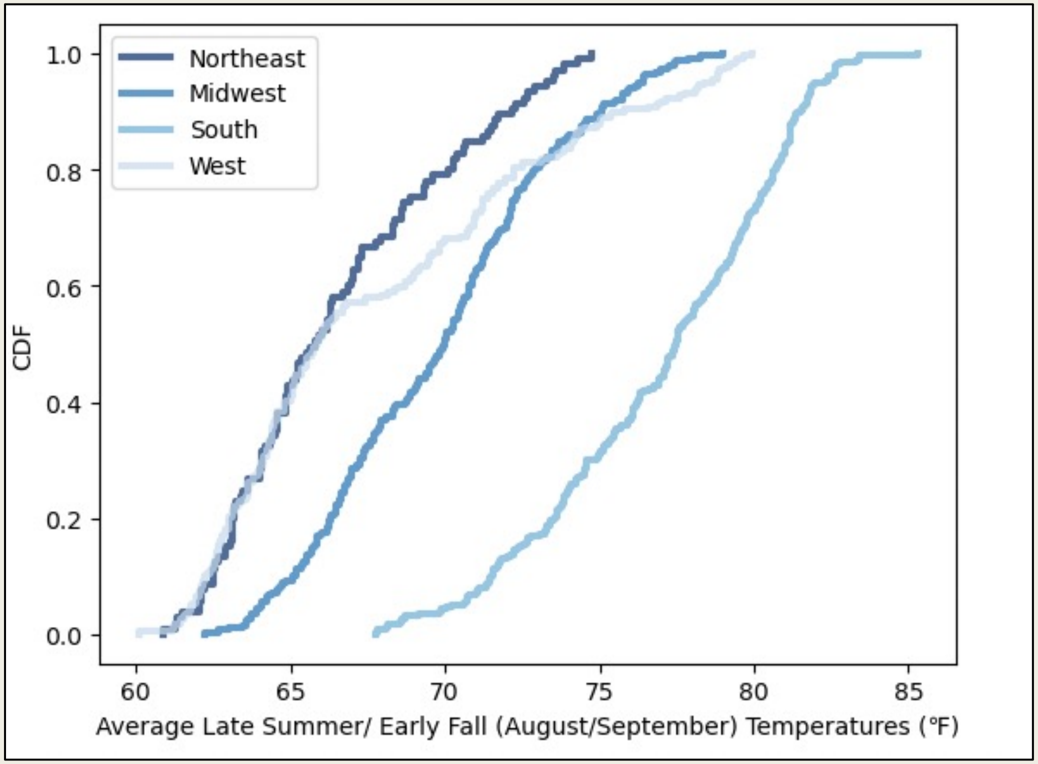
Histogram displaying how the datapoints in datasets are sorted into the different U.S. regions. The graph showcases that the data in the dataset is not equally representing the different U.S. regions. By grouping the data under the regions variable, the some of variation and outliers observed in the variables may be accounted for.

Data Analysis:
Probability Mass Function (PMF)



By filtering the temperature data by the Midwest and Southern regions, the bimodal distribution seen previously in the histogram can be teased apart. Both regions have a rough normal distribution. The Midwest appears to be more skewed to rights while the South is more skewed to the left. This could be due to the **Midwest is more likely to be colder** in August/September compared to the **South, which is more likely to be warmer.**

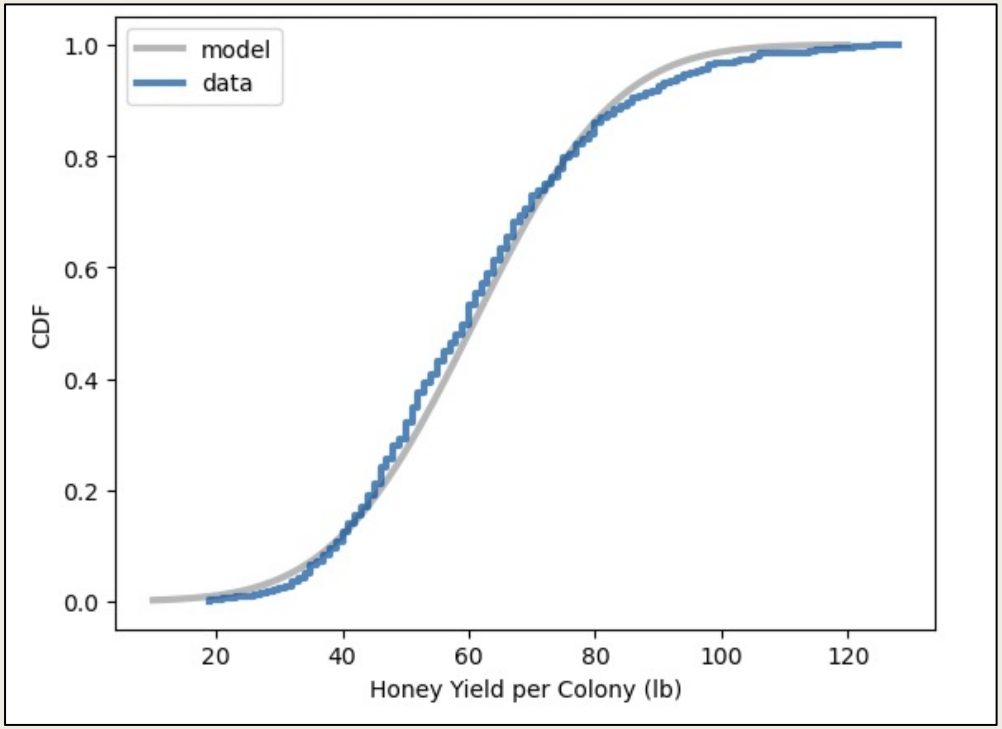
Data Analysis: Cumulative Distribution Function (CDF)



The graph demonstrates that there are regional differences in temperature during the late summer/early fall period. For future research, the months selected to represent this time period of the hives' lifecycle may need to be adjusted to each region. While August/September may be the critical months for a hive in the Northeast, September/October may be the critical months for a hive located in the South.

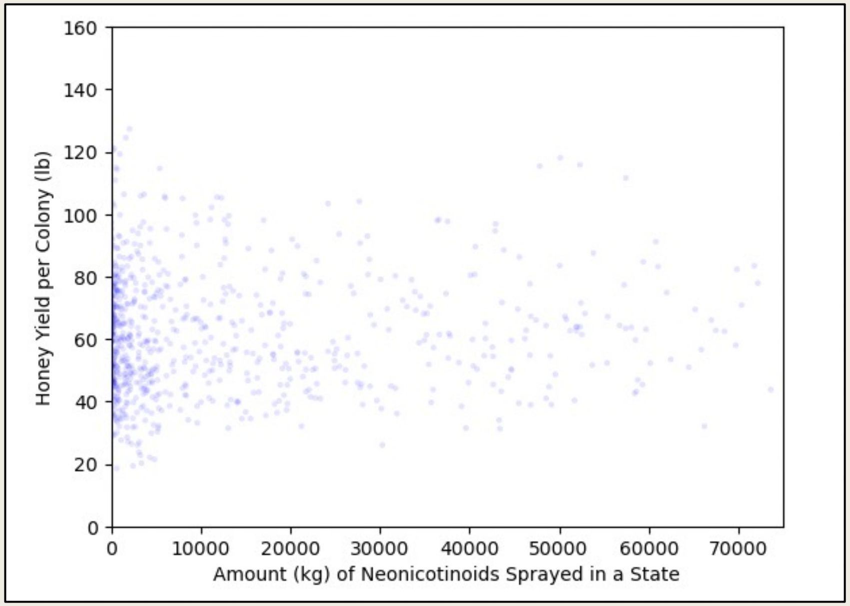
The CDFs for all of the regions are roughly a straight line compared to an S-curve. This means the data for each region is uniformly distributed compared to a normal distribution. The uniform distribution could indicate the gradual rise in temperatures from 1995 to 2015 in U.S. due to climate change.

Data Analysis: Analytical Distribution



The CDF plots shows a comparison of the distribution of the honey yield per colony variable with a normal model. The normal model used the honey yield variable's mean and standard deviation for its creation. The graph indicates that the variable does follow a normal distribution despite the skew observed in the variable's histogram. By knowing whether the variable follows a normal distribution can affect which types of tests are selected for further analysis. The plots also demonstrates that the median of the honey yield per colony is around 65 lbs.

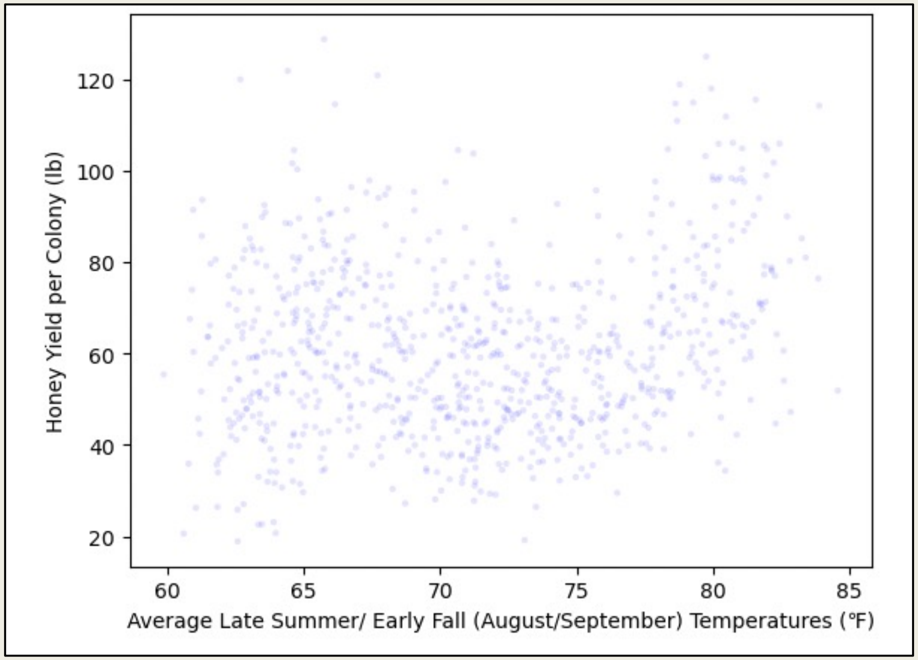
Data Analysis: Relationships Between Variables




```
Covariance: -57418.36742564251
Correlation: -0.03230939379276975
```

Based on the scatter plot and the calculated correlation between the amount of pesticide sprayed and the honey yield per colony, the two variables have a weak relationship. The correlation calculated is very close to zero. Both the calculated covariance and correlation were negative. This indicates there is a small inverse relationship between the two variables. Due to the non-linear nature of the amount of pesticide variable, I used a Spearman’s Rank Correlation to calculate the correlation.

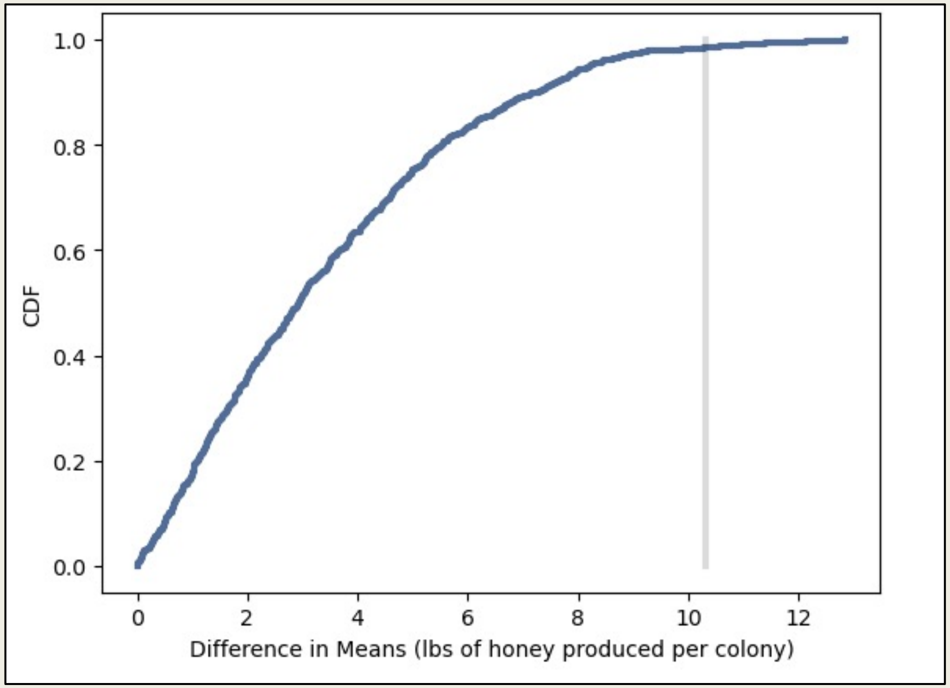
Data Analysis: Relationships Between Variables




```
Covariance: 21.802676544796576
Correlation: 0.04701450703251791
```

Based on the scatter plot and the calculated correlation between the average late summer/early fall temperatures and the honey yield per colony, the two variables have a weak relationship. The correlation calculated is also very close to zero. Both the calculated covariance and correlation were positive. This indicates there is a small positive relationship between the two variables. Due to the non-linear nature of the amount of pesticide variable, I used a Spearman's Rank Correlation to calculate the correlation.

Data Analysis:
Hypothesis Testing (Permutation Test)



Using the Difference in Means test, I decided to compare the honey yield per colony from 1995 and 2015. Theoretically, since climate change has been gradually raising the global temperature during this time, the average temperature in the U.S. would be different between 1995 and 2015. A difference in temperature would have numerous ecological effect that could directly and indirectly affect the bee population and their ability to produce honey. The calculated p-value for the test was 0.16. This indicates the distributions of yield per colony between the two groups is significantly different.

While the result was significant, the reason for the difference should still be thoroughly investigated. Besides climate change, habitat loss could also have had a large impact on the bee population.

Data Analysis: Multiple Regression

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1667.3721	195.456	8.531	0.000	1283.736	2051.008
C(Region)[T.Northeast]	-15.7286	2.050	-7.672	0.000	-19.752	-11.705
C(Region)[T.South]	-6.1360	1.811	-3.388	0.001	-9.691	-2.581
C(Region)[T.West]	-11.7304	1.614	-7.268	0.000	-14.898	-8.563
avg_falltemp	0.3215	0.142	2.267	0.024	0.043	0.600
year	-0.8092	0.098	-8.295	0.000	-1.001	-0.618

Omnibus:	71.344	Durbin-Watson:	0.707
Prob(Omnibus):	0.000	Jarque-Bera (JB):	89.046
Skew:	0.724	Prob(JB):	4.61e-20
Kurtosis:	3.643	Cond. No.	6.67e+05

I created a multiple regression model used to predict honey yield per colony. I used the explanatory variables of average late summer/early fall temperature, year, and region. I chose to include the year and region variables in my model as they both have displayed to have an impact on the yield from previous tests. Both of these variables were statistically significant when testing the model.

Unfortunately, the model created is weak as it only generated a R-squared value of **0.173**. Thus, the model only accounts for a small part of the variation that occurs with honey yield per colony.

R-squared:	0.173
Adj. R-squared:	0.168

References:

Calovi, M., Grozinger, C. M., Miller, D. A., & Goslee, S. C. (2021). Summer weather conditions influence winter survival of honey bees (Apis mellifera) in the northeastern United States. *Scientific Reports*, *11*(1553).

Grozinger, C., & Anton, K. (2022, May 9). *Honey Bee Management Throughout the Seasons*. PennStateExtension. https://extension.psu.edu/honey-bee-management-throughout-the-seasons.

National Agricultural Statistics Service. (2023, May 17). *Honey*. Honey (NASS).https://usda.library.cornell.edu/concern/publications/hd76s004z?locale=en&page=2#release-items.

USDA. (2018). *Honey Bees*. USDA. https://www.usda.gov/peoples-garden/pollinators/honey-bees#:~:text=Honeybees%20pollinate%20%2415%20billion%20worth,fruits%2C%20nuts%2C%20and%20vegetables.

U.S. Geological Survey. (2019). *Pesticide National Synthesis Project*. National Water-Quality Assessment Project. https://water.usgs.gov/nawqa/pnsp/usage/maps/county-level/.

Wong, J. (2022). Average Monthly Temperatures by US State. [Data Set]. https://www.kaggle.com/datasets/justinrwong/average-monthly-temperature-by-us-state.

Zmith, K. (2018) Honeybees and Neonic Pesticides. [Data Set]. https://www.kaggle.com/datasets/kevinzmith/honey-with-neonic-pesticide.