SAS® GLOBAL FORUM 2020 #SASGF

Paper 5206-2020

The Plight of the Honeybee

Rachael Bishop, DaMarkus Green, Karin Kolb
Kennesaw State University

ABSTRACT

A decline in the honeybee population is raising concerns worldwide. Honeybees are an important part of agricultural industries. Is it possible that the neonicotinoid pesticides used to protect crops from damaging insects are also harming the insects necessary for pollinating the same crops? This study explores the effects of these pesticides on honeybees by comparing annual honey production yields and honeybee colony counts in the United States from 1995 through 2015.

INTRODUCTION

It is becoming common knowledge that the bees we normally swat away have a greater importance than once imagined. No longer seen as pests, they are emerging as unsung heroes actively pollinating the majority of our food crops for as long as they have buzzed the Earth. Out of the 20,000 known bee species in the world, 4,000 of them are native to the United States. It is estimated that 80 percent of flowering plants and crops around the world are pollinated by native bees; as well, 75 percent of the fruits, nuts, and vegetables in the United States are pollinated by a variety of bees (Lubeck, 2019). Overall, an annual increase of $14 billion in crop value can be attributed to honeybee pollination in the U.S. ("Pesticide toxicity to bees", 2019).

With the recent acknowledgment of the endangerment of bees and the decrease of their yearly pollination, it is of utmost importance to monitor bee activity in an attempt to increase their population and their pollination efforts in the United States and worldwide. The usage of pesticides on our nation's crops, in an attempt to increase crop output, has potentially done more damage than good. Injuring and killing our nation's natural pollinators may be permanently damaging our agricultural output. Time is of the essence to determine what is really causing our natural pollinators, our bees, to die off, resulting in an overall reduction in pollination and in production of our nation's food sources.

One place to begin, is investigating honeybees and one of their top executioners: pesticides. Because honeybees leave evidence of their presence through the production of honey, there is a measurable variable to analyze and draw conclusions about the viability of this species of bees. Neonic pesticide usage began in the 1990's; however, with honey production declining years earlier, it is of interest to look into whether the use of these pesticides has increased the ongoing decline of honeybee honey production in the United States.

BACKGROUND

Neonicotinoid (neonic) pesticides are insecticides derived from nicotine. In June of 2017, two large-scale studies confirmed that in agricultural areas, neonic pesticides are hurting the ability for bees to reproduce, as well as, killing them. It was argued that previous studies used unrealistic quantities of pesticides to substantiate their hypothesis. The new studies claim the damage is done over an extended period of time (Dengler, 2017). 31.6% of honeybees that are exposed to pesticides fail to return to their hives, while the remaining bring back contaminated pollen which in turn affects the other workers and the queen at the hive ("Pesticide toxicity to bees", 2019). Understanding the effects of neonic pesticides on the health of honeybees is necessary to reduce the impact on food sources and lives of bees.

DATA

The data set, *Honeybees and Neonic Pesticides*, used for the analysis is provided by the National Agricultural Statistics Service (NASS). The U.S. Department of Agriculture (USDA) uses the NASS as their primary data reporting source. The data set also includes results from the U.S. Geological Survey (USGS) Pesticide National Synthesis Project. Spanning 20 years, 1995 to 2015, it provides statistics regarding honey production supply and demand in most of the states in the U.S., as well as detailed information on Neonicotinoid (neonic) pesticides. There is no data for Alaska, Connecticut, Delaware, Massachusetts, New Hampshire and Rhode Island. This data set is available through www.kaggle.com and was compiled by Kevin Zmith.

The set includes seventeen variables and 1132 observations. The variables of particular interest in this study are State Name, Region, Number of Colonies, Total Production, which is the number of honey-producing colonies multiplied by the honey yield per colony in a given U.S. state per year, and the five neonic pesticides, Clothianidin, Imidacloprid, Thiamethoxam, Acetamiprid, and Thiacloprid. The full list of variables and their descriptions can be found in the data dictionary in Appendix A Table 1.

PROBLEMS ADDRESSED

The USDA provides a very informative website full of information pertaining to the agriculture industry. There is a section dedicated to their Pesticide Data Program. The addition of an interactive tool, such as a decision tree, could be a resource for concerned growers to access when making a decision about pesticide usage. This predictive model could help growers make informed decisions about the use of specific neonic pesticides and the potential consequences to honeybees, pollination, and their bottom line, crop production yield.

The National Farmers Union is an organization whose mission is to advocate for U.S. farming families through education, cooperation and legislation. They provide important agricultural updates pertaining to all farmers. Sharing the results of a comparison model on their website provides valuable information for interested farmers to determine the effects in their region from the use of specific pesticides. Using this model, future honey yield and colony count could be estimated to better monitor and track the decline of honeybee populations.

DATA CLEANING AND PREPROCESSING

Observations dated before 1994 are removed; although they included data about annual honey yield, they did not include data about neonic pesticide use. The observations for 1994 are removed once the new target variable is created. The observations for the state of Hawaii are removed as well for lack of data about neonic pesticide usage. The resulting 856 observation data set contains data about 43 out of the 48 contiguous United States.

Prior to developing a decision tree model, a new target variable, Effect, is created comparing the total production per state to each preceding year to see if there is a decrease in production caused by an increase in neonic pesticide usage. This comparison is made by comparing two additional new variables, ProdDecr and NeonIncr.

ProdDecr is a binary variable assigned by comparing a year's total honey production to the previous year's total honey production for each year in each state. The total honey production in pounds for 1994 is the control observation for the 1995 comparison. If there is a reduction in the total pounds produced then ProdDecr is assigned a value of one. If there is an increase or no change in pounds produced then ProdDecr is assigned a value of zero.

NeonIncr is a binary variable assigned by comparing a year's total neonic pesticide usage to the previous year's total neonic pesticide usage for each year in each state. Neonic pesticide usage data is not recorded prior to 1995, so a quantity of 0kg is the control for the 1995 comparison. If there is an increase in total kg pesticide usage then NeonIncr is assigned a value of one. If there is a reduction or no change in total kg pesticide usage then NeonIncr is assigned a value of zero.

The target variable, Effect, is a binary variable assigned by comparing the values of ProdDecr and NeonIncr. If both ProdDecr and NeonIncr have a value of one, then Effect is assigned a value of one based on the assumption that total honey production is negatively affected by an increase in pesticide usage. Any other combination of values for ProdDecr and NeonIncr results in an assignment of the value zero for the variable Effect.

ANALYSIS

SAS® ENTERPRISE MINER™

The data set including the new target variable is imported into SAS® Enterprise Miner™. It is partitioned into 80% training and 20% validation. Redundant variables such as state and StateName are reduced to one selected input variable. All of the neonic pesticide variables are positively skewed distributions in need of transformation. Three possible transformations are explored: the 'Best' transformation as chosen by Enterprise Miner, the Log transformation, and the Square Root transformation. Attempts to create a useful decision tree model are unsuccessful. Among those attempted, a 5 Branch Chi-Squared Tree is the only model that produces a tree. However, the variables THIAMETHOXAM and IMIDACLOPRID are the only ones of importance. Figure 1 shows the resulting decision tree.

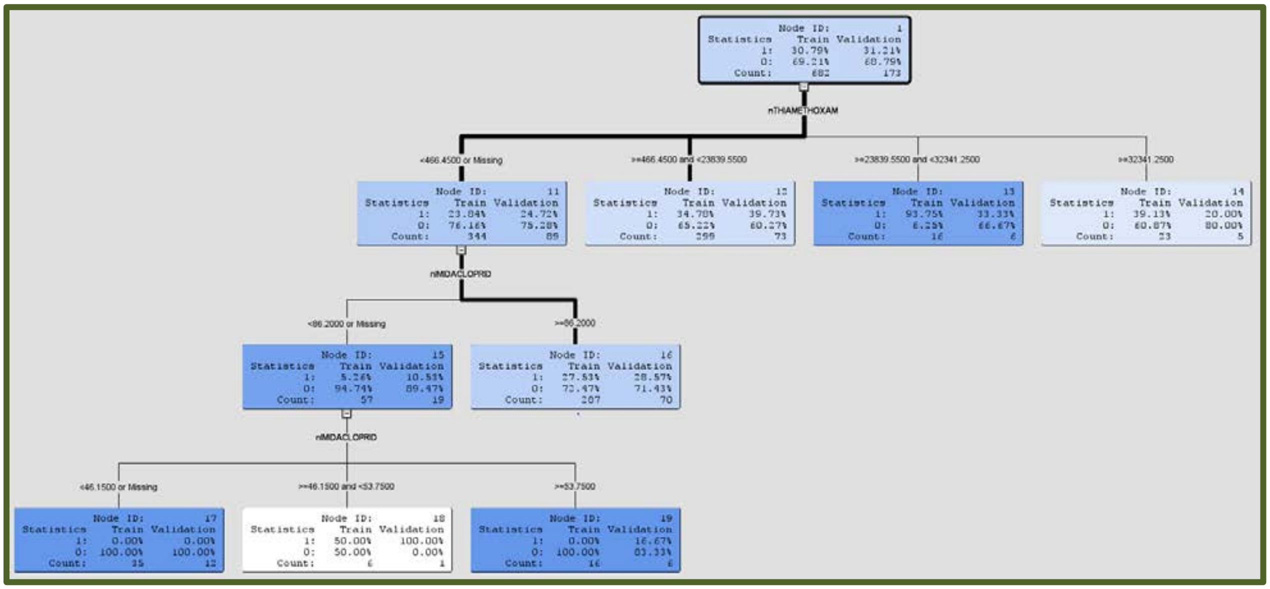

Figure 1. Five Branch Chi-Squared Tree

Unfortunately, our decision tree fails to produce results to address our first problem regarding the USDA and the use of specific pesticides. We continue our study by conducting further analysis to address our second problem. We endeavor to create a model to aid concerned farmers who access the National Farmers Union website.

SAS®

SAS Proc Reg is applied to build the predictive models. The analysis includes all of the individual neonic pesticide variables, as well as, numcol and totalprod, filtered by Region; South, West, Midwest, and Northeast. These regions all have area-specific characteristics that may be of interest for consideration in future research. After dividing the data set by Region, SAS regression analyses are run to determine if any of the specific pesticides have a significant impact on total honey production yield or honeybee colony count. The parameter estimates enable the building of a model based on individual pesticides, facilitating the calculation of estimated effect on total honey production and colony count. The output for each of the regions is shown below in Output 1 through Output 4.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3945213	521458	7.57	<.0001
CL	1	-6352224	1426876	-4.45	<.0001
IM	1	-140.73349	43.86487	-3.21	0.0015
TH	1	214.78869	27.66803	7.76	<.0001
AC	1	-112.05513	38.17587	-2.94	0.0036
TI	1	581.62775	254.66790	2.28	0.0231

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	47045	5909.67161	7.96	<.0001
CL	1	-75002	16171	-4.64	<.0001
IM	1	-2.12934	0.49712	-4.28	<.0001
TH	1	3.28118	0.31356	10.46	<.0001
AC	1	-1.21547	0.43265	-2.81	0.0053
TI	1	6.97215	2.88615	2.42	0.0163

Output 1. South Region: Total Honey Yield Model and Total Colony Count Model

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5760686	817070	7.05	<.0001
AC	1	-138.95492	184.79613	-0.75	0.4529
TH	1	196.03944	23.61056	8.30	<.0001
IM	1	-188.97373	45.81780	-4.12	<.0001
CL	1	-7316660	2150680	-3.40	0.0008
TI	1	6.11718	116.63062	0.05	0.9582

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	86057	10671	8.06	<.0001
CL	1	-106758	28088	-3.80	0.0002
IM	1	-2.93462	0.59837	-4.90	<.0001
TH	1	3.33526	0.30835	10.82	<.0001
AC	1	0.86764	2.41341	0.36	0.7196
TI	1	2.63961	1.52318	1.73	0.0845

Output 2. West Region: Total Honey Yield Model and Total Colony Count Model

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-4055920	1577080	-2.57	0.0107
AC	1	66.37655	58.09710	1.14	0.2543
TH	1	39.52886	47.54333	0.83	0.4065
IM	1	-28.11432	17.93946	-1.57	0.1183
CL	1	34855793	4548613	7.66	<.0001
TI	1	-4518.60361	1981.63789	-2.28	0.0234

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-50968	19049	-2.68	0.0079
CL	1	438805	54940	7.99	<.0001
IM	1	-0.30810	0.21668	-1.42	0.1563
TH	1	0.46847	0.57425	0.82	0.4154
AC	1	1.05849	0.70172	1.51	0.1327
TI	1	-46.07441	23.93518	-1.92	0.0553

Output 3. Midwest Region: Total Honey Yield Model and Total Colony Count Model

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-467992	652059	-0.72	0.4745
AC	1	-98.87782	174.42463	-0.57	0.5720
TH	1	263.53939	71.84059	3.67	0.0004
IM	1	-72.17770	55.40297	-1.30	0.1955
CL	1	3015471	1592433	1.89	0.0611
TI	1	457.52417	146.43753	3.12	0.0023

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3537.58486	8628.28322	0.41	0.6827
CL	1	24838	21072	1.18	0.2412
IM	1	-0.91445	0.73311	-1.25	0.2151
TH	1	3.03076	0.95062	3.19	0.0019
AC	1	0.20309	2.30805	0.09	0.9301
TI	1	7.10419	1.93771	3.67	0.0004

Output 4. Northeast Region: Total Honey Yield Model and Total Colony Count Model

Time series plots are developed for each region based on the total yield of honey each year to provide a visual depicting the possible effect of pesticide usage in a sampling of states in each region. The plots are shown in Appendix B Figure 2 through Figure 5.

CONCLUSION

For our addressed problems, analyzing the use of neonic pesticides on U.S. crops starting in 1995, the expectation was that all of the pesticides have a negative effect on both total honey production yield and total honeybee colony count. Contrary to this belief, the findings suggest that the pesticides THIAMETHOXAM and THIACLOPRID have a positive effect on both yield and colony count. The farming community may find the regression model useful in determining which neonic pesticides are beneficial and which are harmful in their specific region, however, more research is needed. Both our decision tree and regression models identified that if the quantity of THIAMETHOXAM used is monitored, it may result in a positive impact on honey production yield and honeybee colony counts, and subsequently overall crop yield. All the other neonic pesticides were seen as harmful to both honey production and colony count and therefore should be avoided, if possible.

For future research it would be interesting to look into weather patterns, soil acidity, and pollen counts to determine if there is a significant difference between these regions and if these factors have an impact on the results found through this study.

REFERENCES

Dengler, R. (2017, June 29). *Neonicotinoid pesticides are slowly killing bees.* Retrieved from PBS News Hour: https://www.pbs.org/newshour/science/neonicotinoid-pesticides-slowly-killing-bees

Lubeck, M. (2019, June 20). *Honey Bee Helpers: It Takes a Village to Conserve a Colony.* Retrieved from U.S. Geological Survey: https://www.usgs.gov/news/honey-bee-helpers-it-takes-village-conserve-colony

Pesticide toxicity to bees. (2019, November 11). Retrieved November 16, 2019, from Wikipedia: https://en.wikipedia.org/w/index.php?title=Pesticide_toxicity_to_bees&oldid=925625965.

Zmith, K. (2018, May 16). *Honeybees and Neonic Pesticides.* Retrieved from kaggle: https://www.kaggle.com/kevinzmith/honey-with-neonic-pesticide

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact authors at:

Rachael Bishop
rbisho16@students.kennesaw.edu

DaMarkus Green
dgreen62@students.kennesaw.edu

Karin Kolb
kkolb1@students.kennesaw.edu

Dr. Sherry Ni – Faculty Advisor
sni@kennesaw.edu

APPENDIX A

Variable	Description
state	Abbreviated state name
numcol	Number of honeybee colonies in the state
yieldpercol	Honey yield per colony. Unit is pounds
totalprod	Total production (numcol x yieldpercol). Unit is pounds
stocks	Refers to stocks held by producers. Unit is pounds
priceperlb	Refers to average price per pound based on expanded sales. Unit is dollars
prodvalue	Value of production (totalprod x priceperlb). Unit is dollars
year	Recording year
StateName	Full state name
Region	U.S.-Census-classification of states
FIPS	Federal Information Processing Series code
nCLOTHIANIDIN	The amount in kg of CLOTHIANIDIN applied
nIMIDACLOPRID	The amount in kg of IMIDACLOPRID applied
nTHIAMETHOXAM	The amount in kg of THIAMETHOXAM applied
nACETAMIPRID	The amount in kg of ACETAMIPRID applied
nTHIACLOPRID	The amount in kg of THIACLOPRID applied
nAllNeonic	The amount in kg of all Neonics applied = (nCLOTHIANIDIN + nIMIDACLOPRID + nTHIAMETHOXAM + nACETAMIPRID + nTHIACLOPRID)

Table 1. Variables and Descriptions

APPENDIX B: TIME SERIES PLOTS

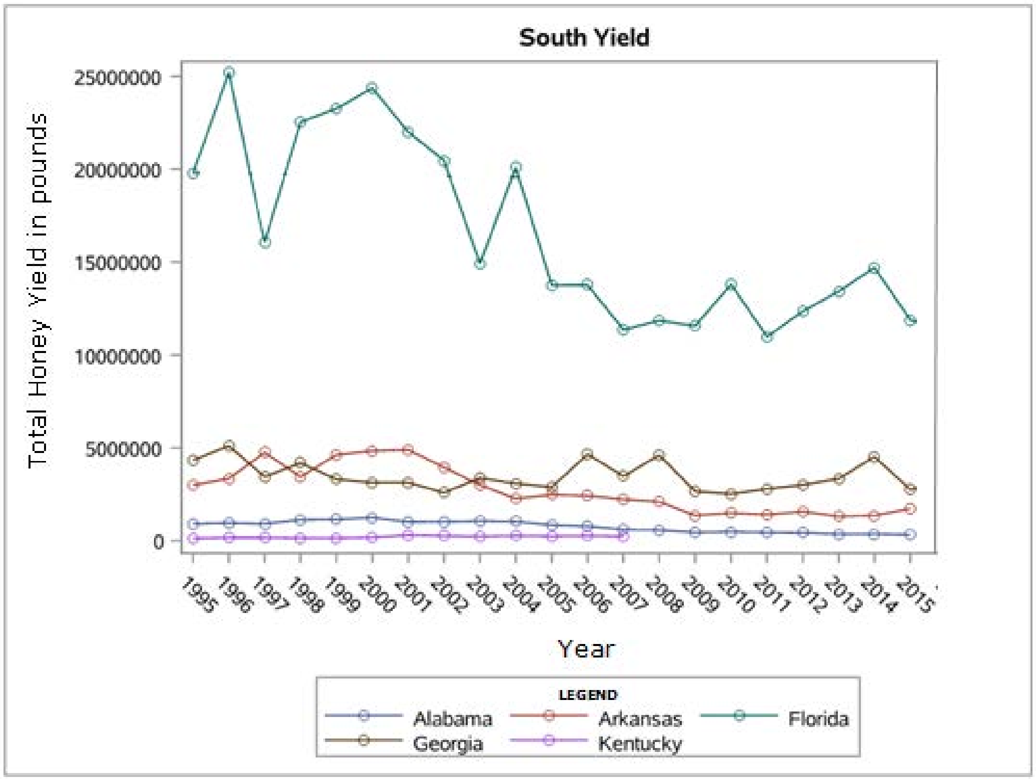

Figure 2. Annual Total Honey Yield in Southern States

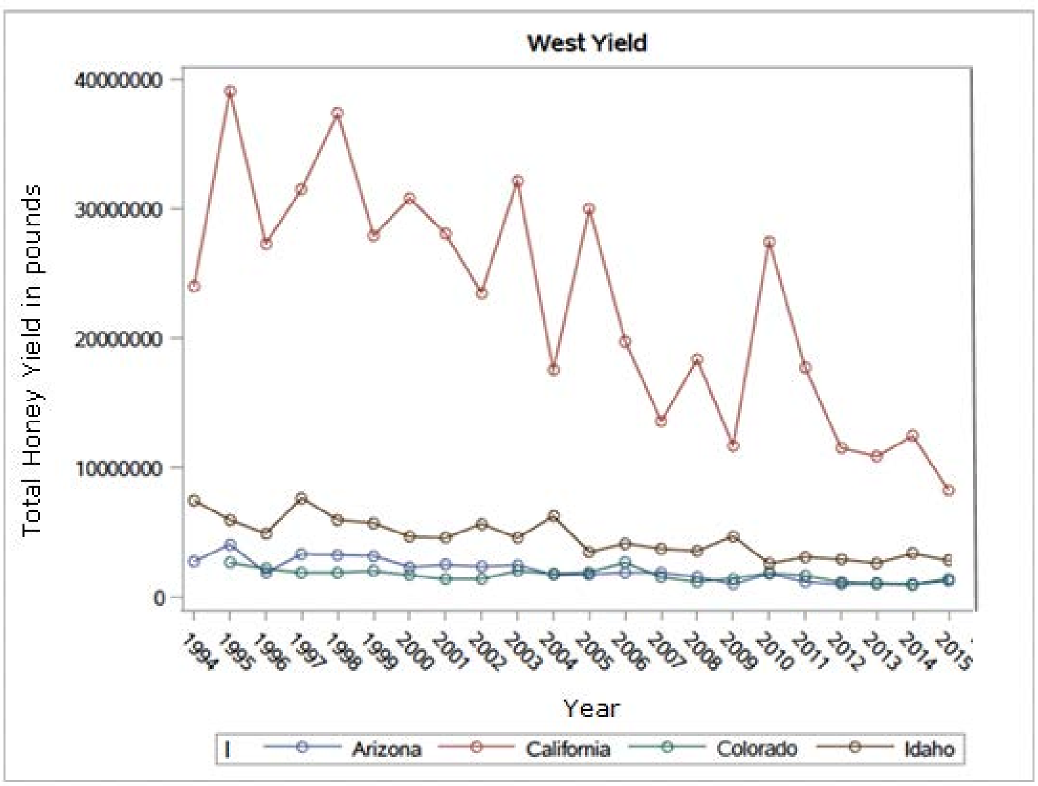

Figure 3. Annual Total Honey Yield in Western States

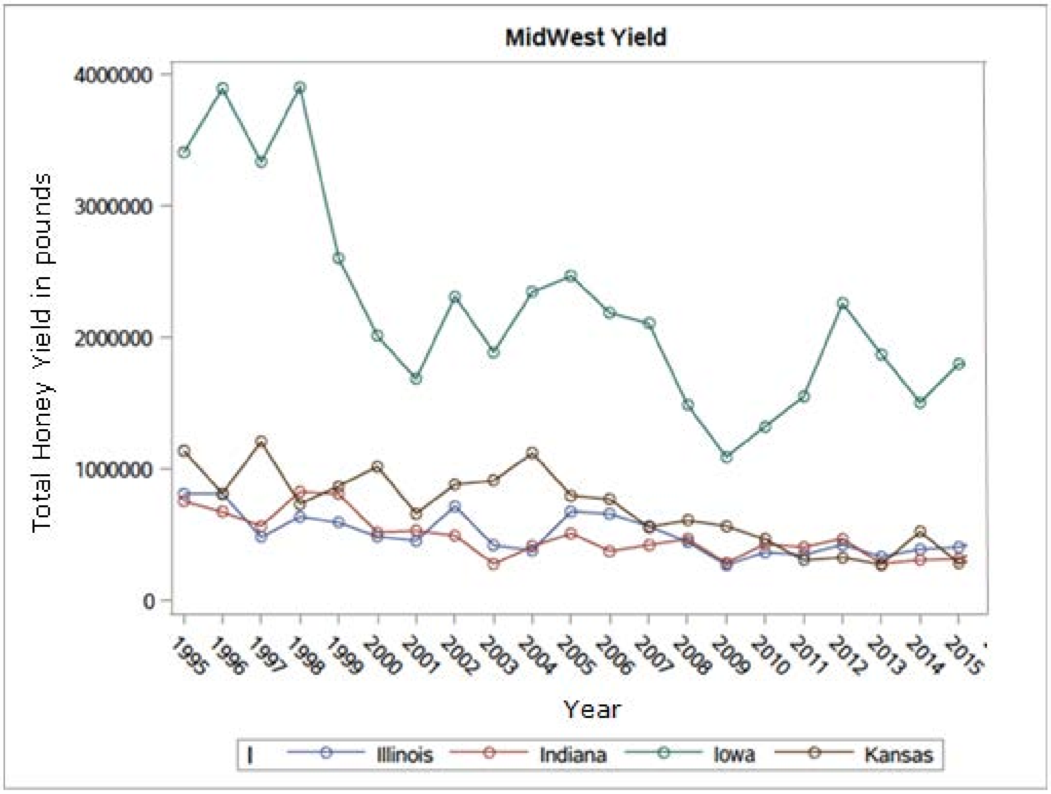

Figure 4. Annual Total Honey Yield in Midwestern States

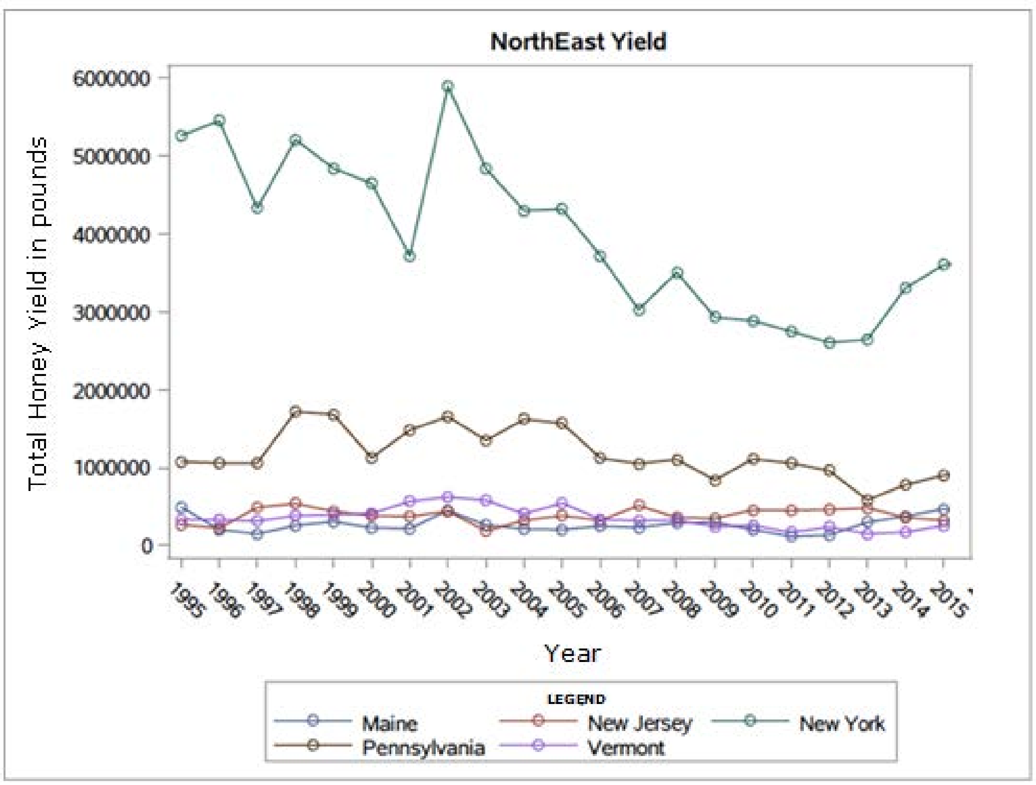

Figure 5. Annual Total Honey Yield in Northeastern State