Bees

What’s the buzz about them?

How Important are Bees?

Bees Pollinate

20 Million

Flowers each day

Bees are Responsible For

$15 billion increased crop value each year

75%

of Fruits, Vegetables, and Nuts are pollinated by Bees

1 out of every 4 bites of food, are because of Bees!

To make 1 lb of honey, 2 million flowers have to be visited.

1 bee colony can make 60 - 100 lbs of honey per year

Sources Below:

- https://www.usgs.gov/faqs/why-are-bees-important?qt-news_science_products=0#qt-news_science_products
- http://www.bee-careful.com/fruit-diversity/pollination-bees/
- http://www.goldenblossomhoney.com/education_bees.php

Our Hypothesis

- As ambient temperature decreases, bee population decreases

- As the average presence of pesticides in honey samples increased; bee population decreased

Our Questions:

- How do natural and human influences impact bees?
- How were we going to get this information?
  - Is this information reliable?
- How were we going to get different data sets?
- What is the relationship between honey sales and production?
- Is there a dramatic decrease in bee population?

Bee Colony Data

- All Bee Data was found using the https://quickstats.nass.usda.gov/ API through USDA.

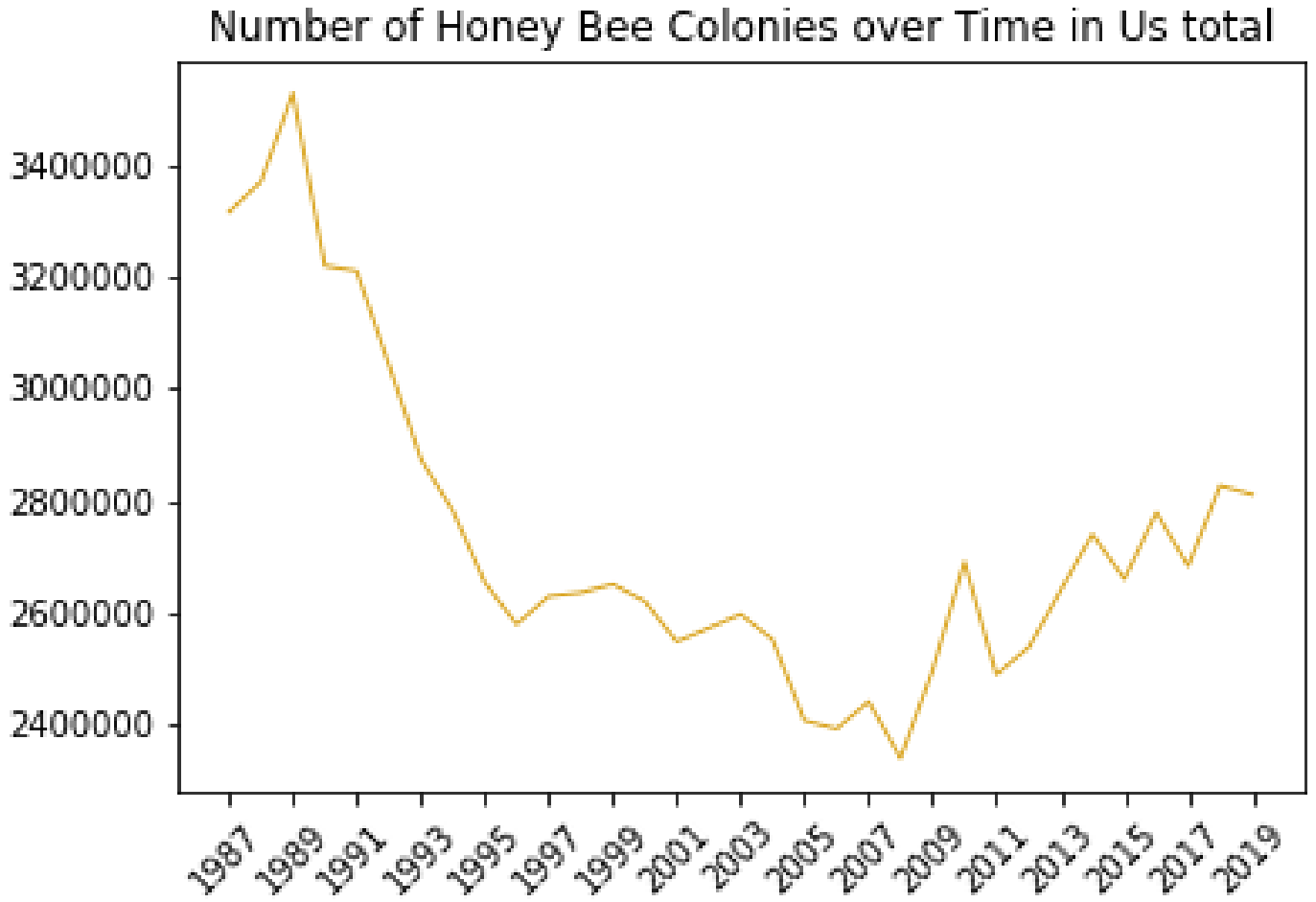
- The data provided only includes numbers of honey producing colonies for operations with five or more colonies.

Challenges of Bee Data

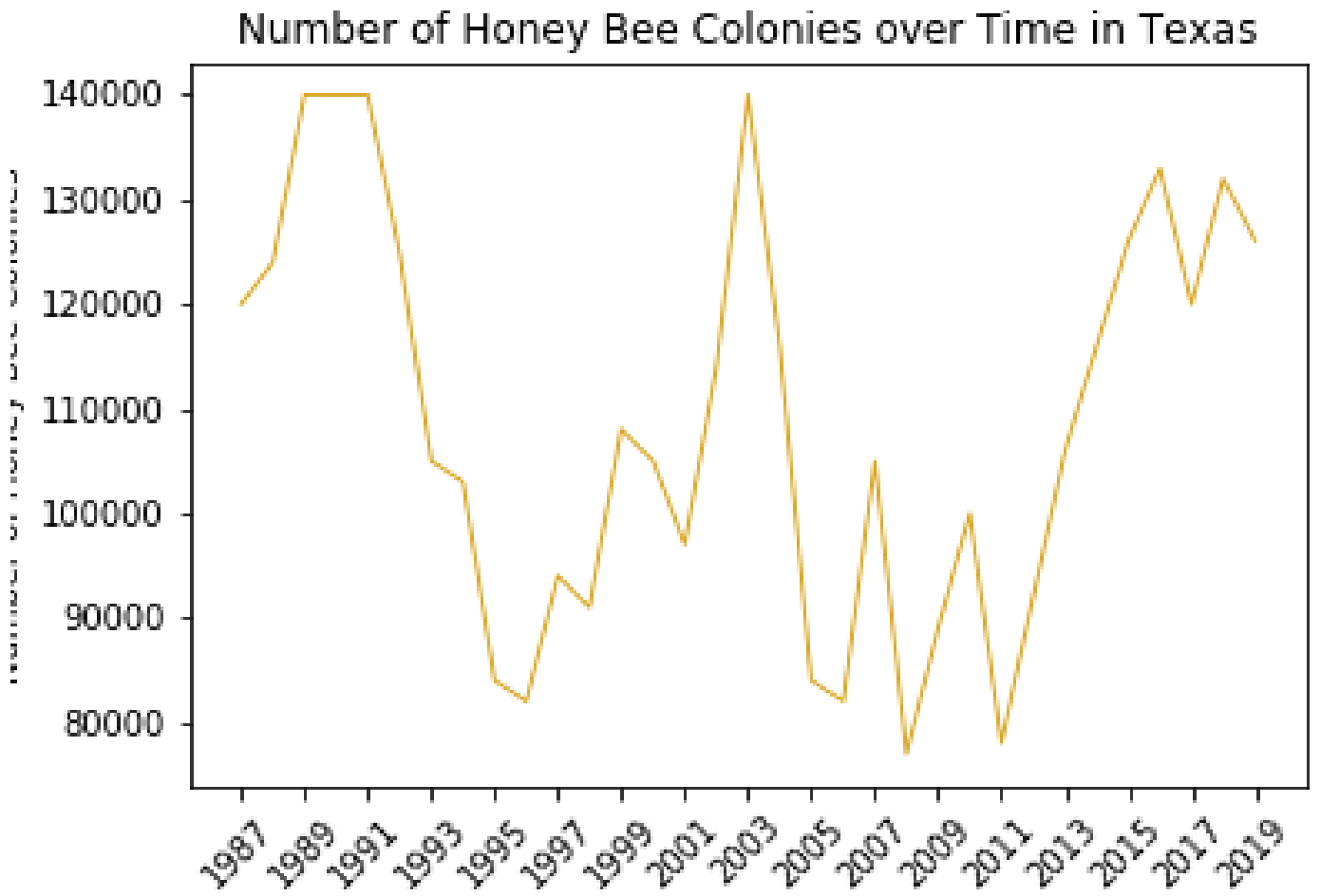
- While the data was given to us in a csv , we instead built an API to collect and parse through the data.
  - There were numerous issues in generating the correct syntax but eventually we became quite competent in the USDA API.
- Different things to account for during data cleanup were:
  - Values coming over as objects
  - Removal of , or .
  - Converting to strings.
- Dealing with different date/time issues, specifically working with quarters and years and generating a meaningful graph.

```
params = {
    'key': quickstats_api,
    'reference_period_desc' : 'MARKETING YEAR',
    'commodity_desc' : 'HONEY',
    'short_desc' : 'HONEY, BEE COLONIES - INVENTORY, MEASURED IN COLONIES',
}

base_url = 'https://quickstats.nass.usda.gov/api/api_GET/'


response = requests.get(base_url, params=params).json()

for x in (range(len(response['data']))):

    honey_data.loc[x,'Year'] = response['data'][x]['year']
    honey_data.loc[x,'State'] = response['data'][x]['location_desc']
    honey_data.loc[x,'Commodity'] = response['data'][x]['commodity_desc']
    honey_data.loc[x,'Data Type'] = response['data'][x]['short_desc']
    honey_data.loc[x,'Value'] = response['data'][x]['Value']


honey_data
```

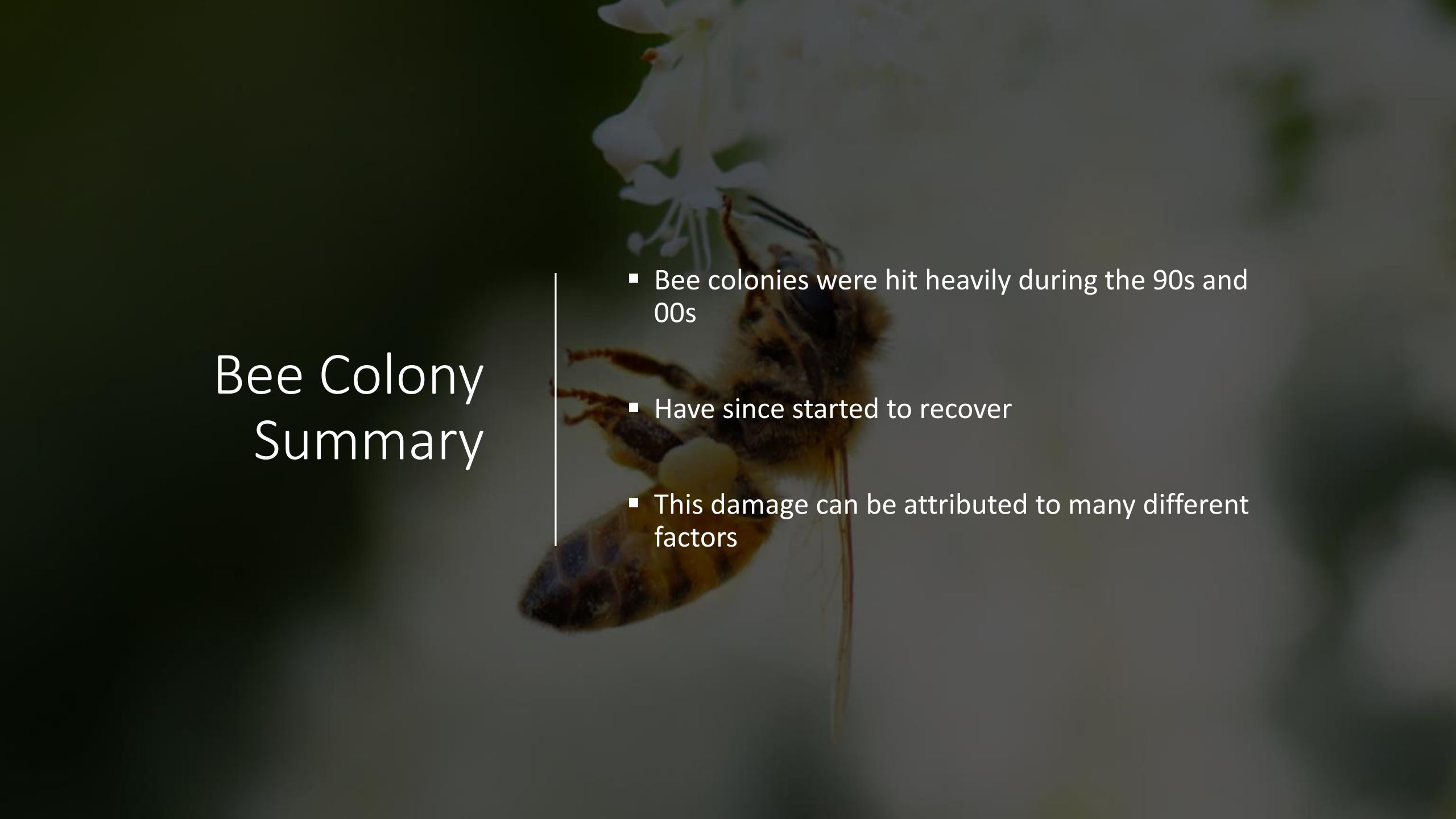
USDA API



US Bee Colonies over time



Texas Bee Colonies over time

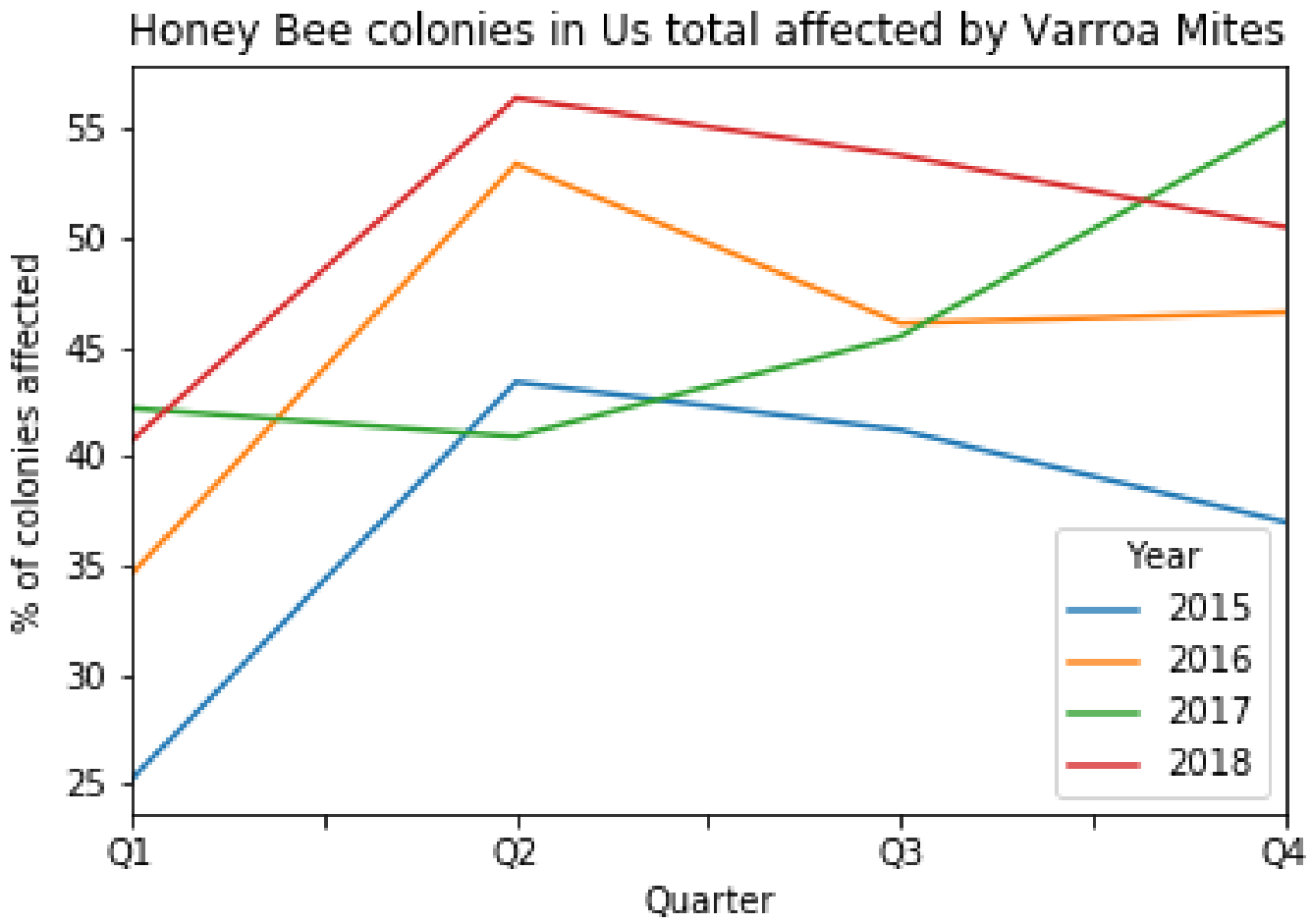


Bee Colony Summary

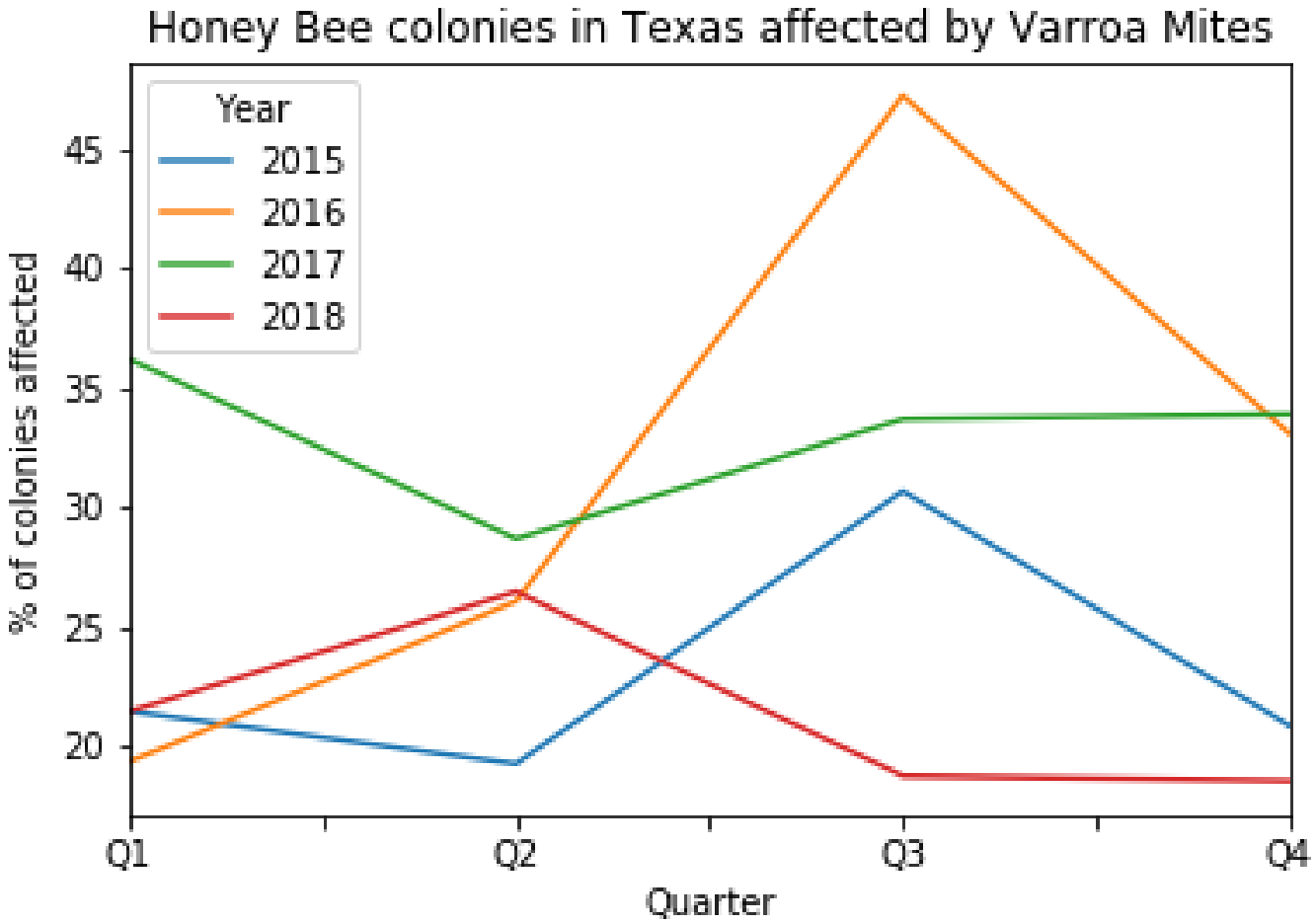
- Bee colonies were hit heavily during the 90s and 00s
- Have since started to recover
- This damage can be attributed to many different factors

Bee Analysis

- The colony process has the ability to generate a graph for each state as well as the US Total
    - Allowed us to see multiple trends of **decreasing** colony counts during the 90s-00s and colony recovery during the 10s.
- While parsing through the data we were able to determine that **Varroa Mites** were a main pest that affected bee colonies.
- If we had more time, would be interested in diving deeper into how the Bee colonies were directly affected by Varroa Mites.



US Colonies affected by Varroa Mites



Texas Bee Colonies affected by Varroa Mites

Pesticide Data

- The USDA website hosts the Pesticide Data Program (PDP) database which was initiated in 1991.
- Collecting this data required navigating this very buggy user interface with trial-and-error until the desired data set was found.
- Then the data was conveniently downloaded in CSV format (no API pull required).
- Unfortunately, this was the most recent granular data we could find on pesticides relating to honey.

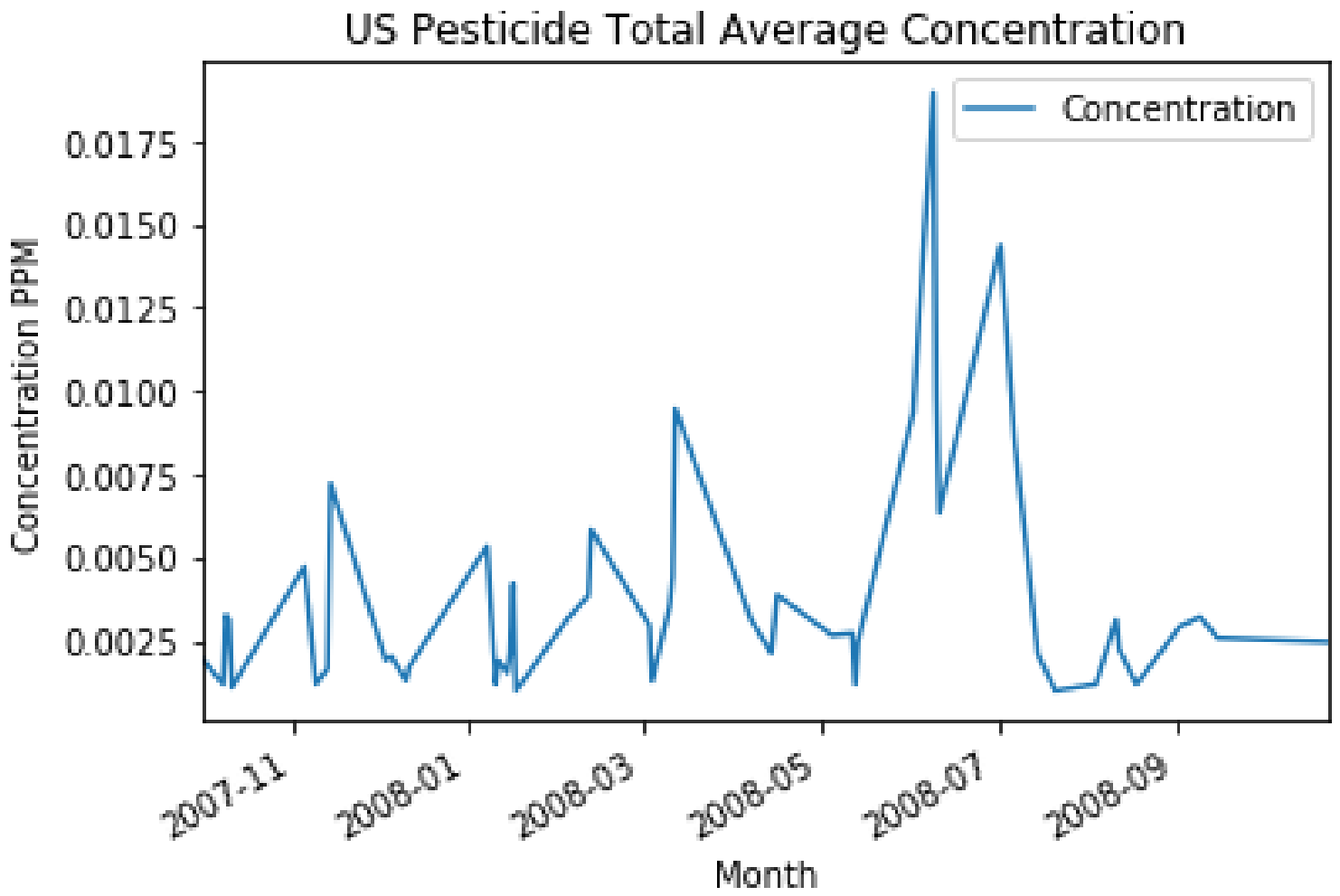
Pesticide Data Cleanup

Data clean-up was necessary. The data set had columns with missing cells and entire rows which were unnecessary for our purposes.

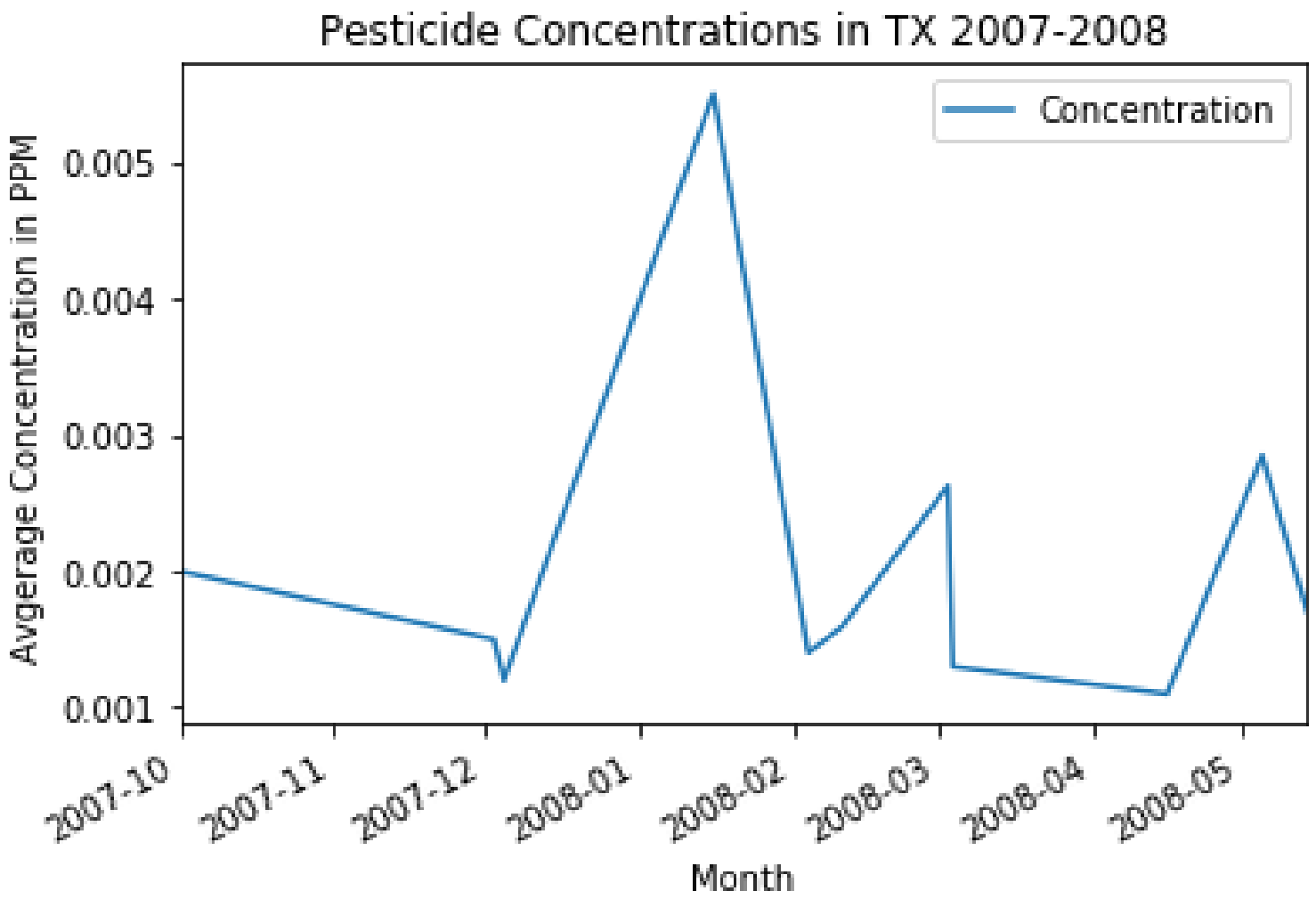
Used loc function and list comprehension on some data points and using the drop function for entire columns.

The most difficult element to clean was the needed date data which was embedded in a complex string under the ‘Sample ID’ column. This required using the pandas.dataframe.str.slice() function to target the needed data.

The pandas.dataframe.to_datetime() function required the string to be rearranged which required slicing out small pieces of the string and then bringing them all back together by concatenating them adding columns together into a new column variable named ‘Months.’

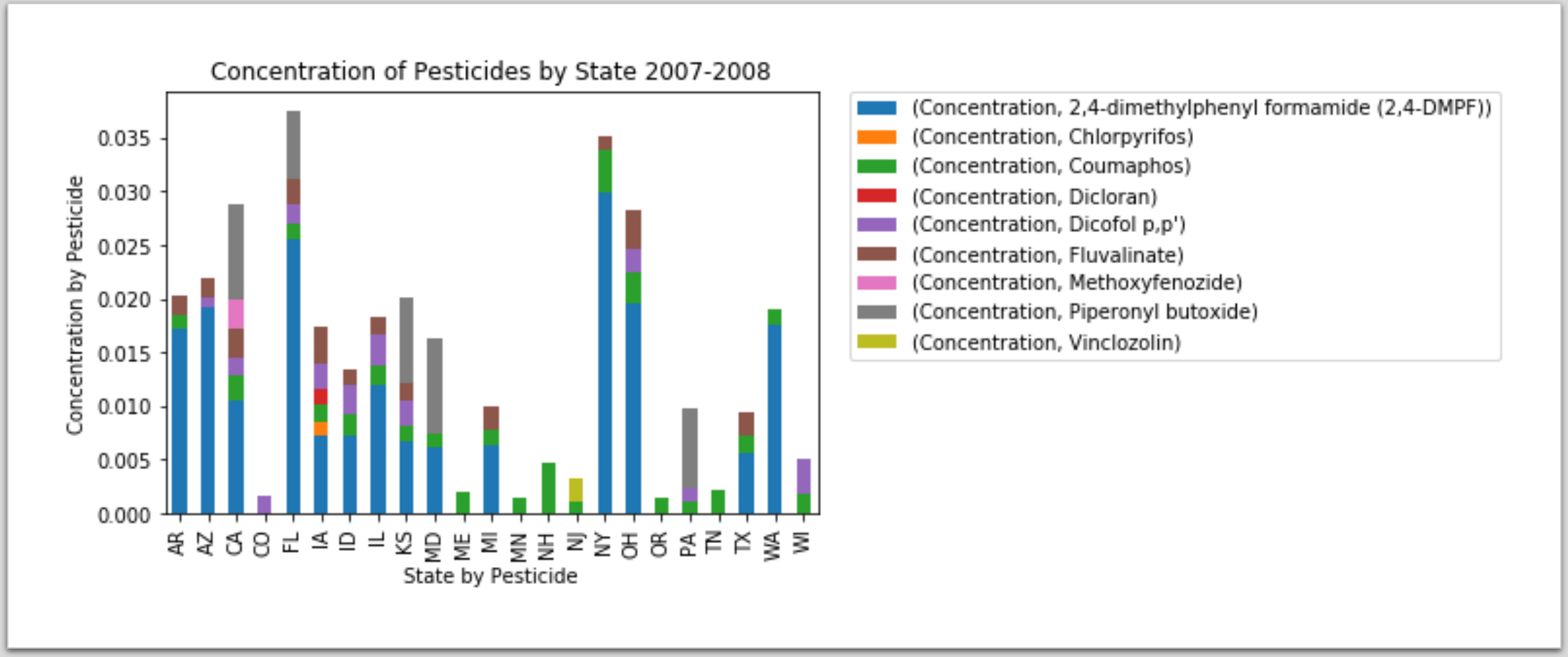


US Pesticide Average Concentration



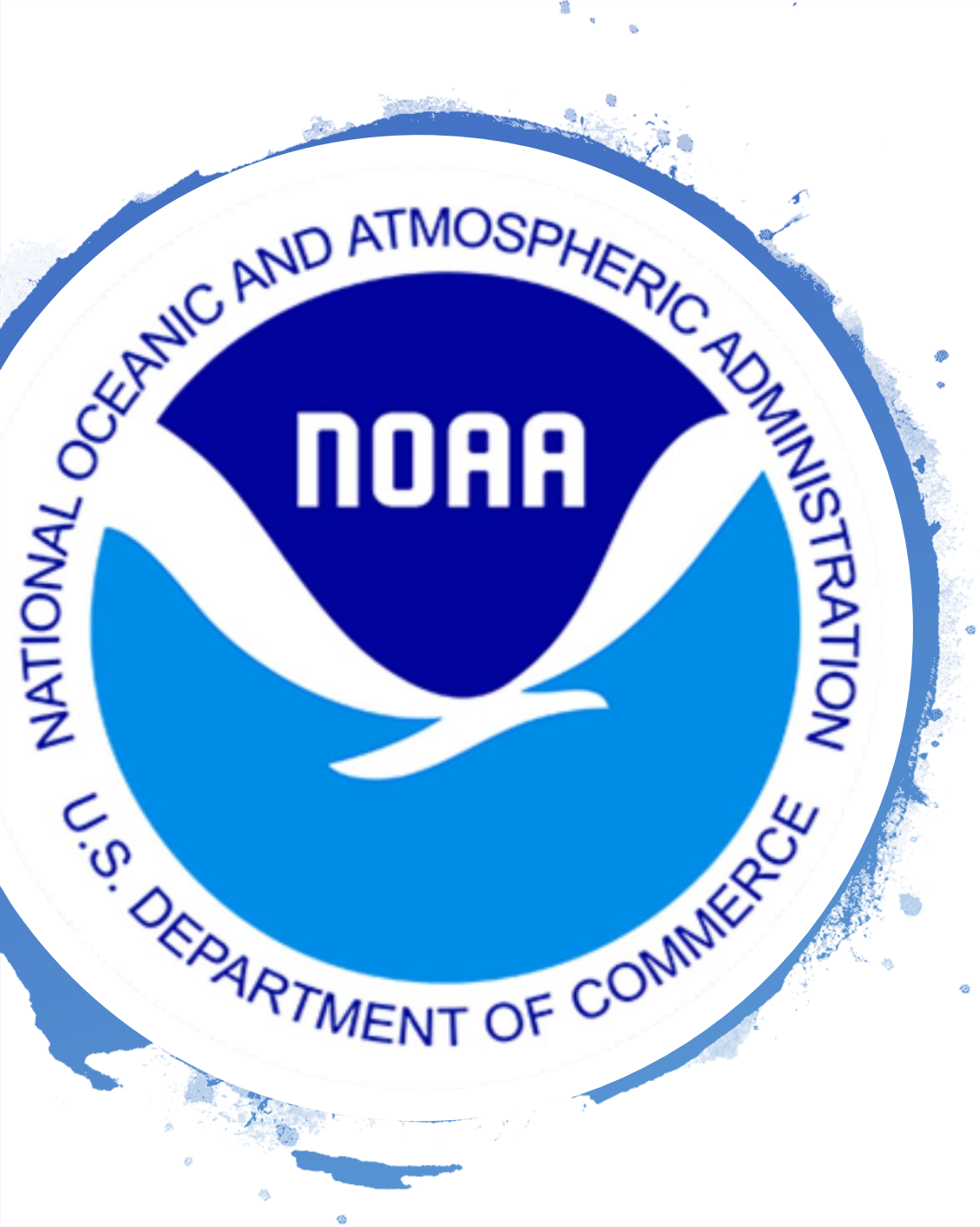
Texas Pesticide Concentration

Pesticide Concentration by State



Pesticide Summary

- The concentration of pesticide in honey indicates more or lesser use of said pesticide.
- There was a clear spike in pesticide concentration between 2008-05 and 2008-09.
- Most used pesticide across the set is 2,4-dimethylphenyl formamide.
  - Also found on catfish.



Temperature Data

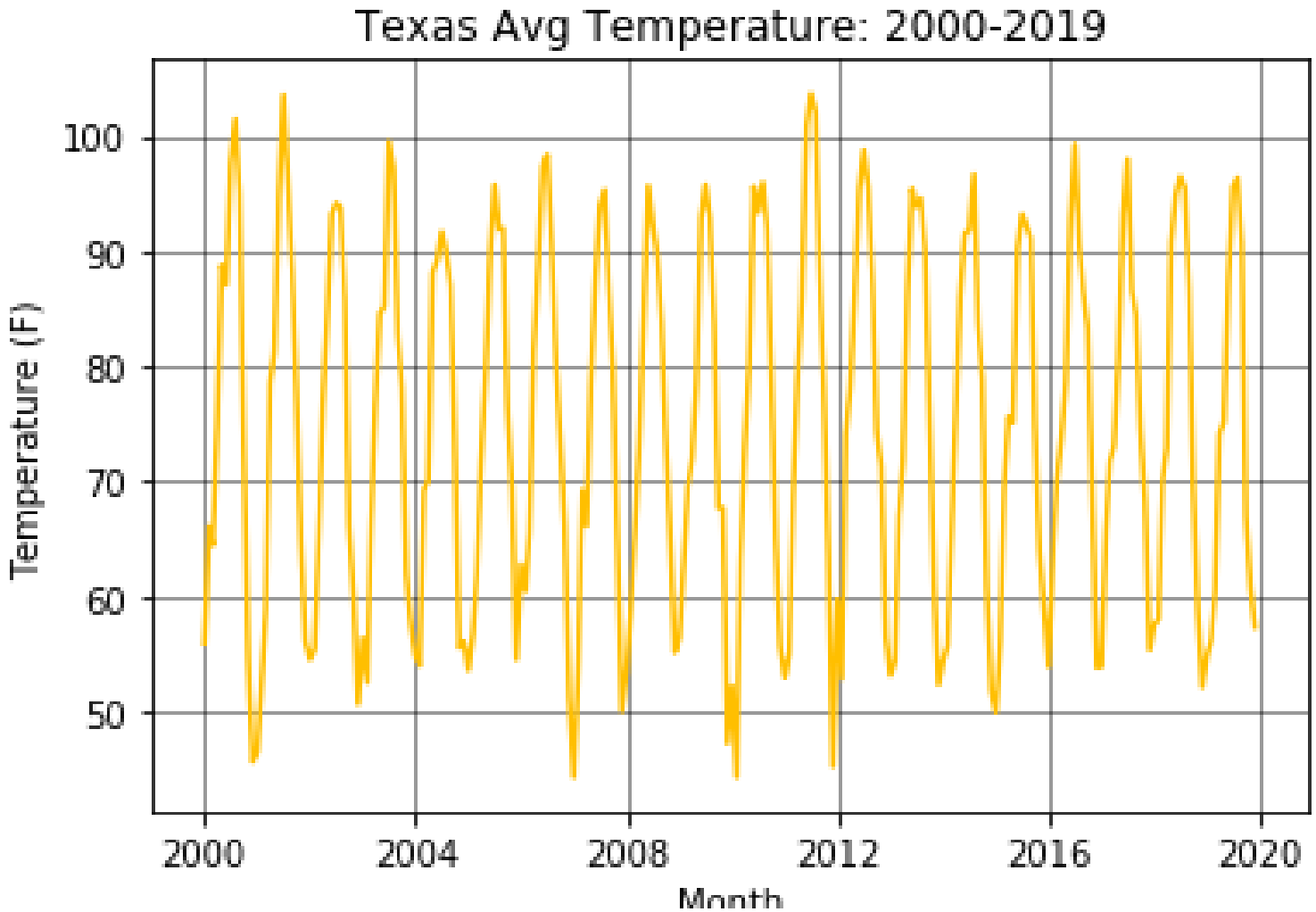
- Temperatures found on NOAA's website through API pulls
  - Add semi-decent documentation
- Had records of weather data going back to 1700's
- Used weather station data (99% coverage)

Challenges of Temperature

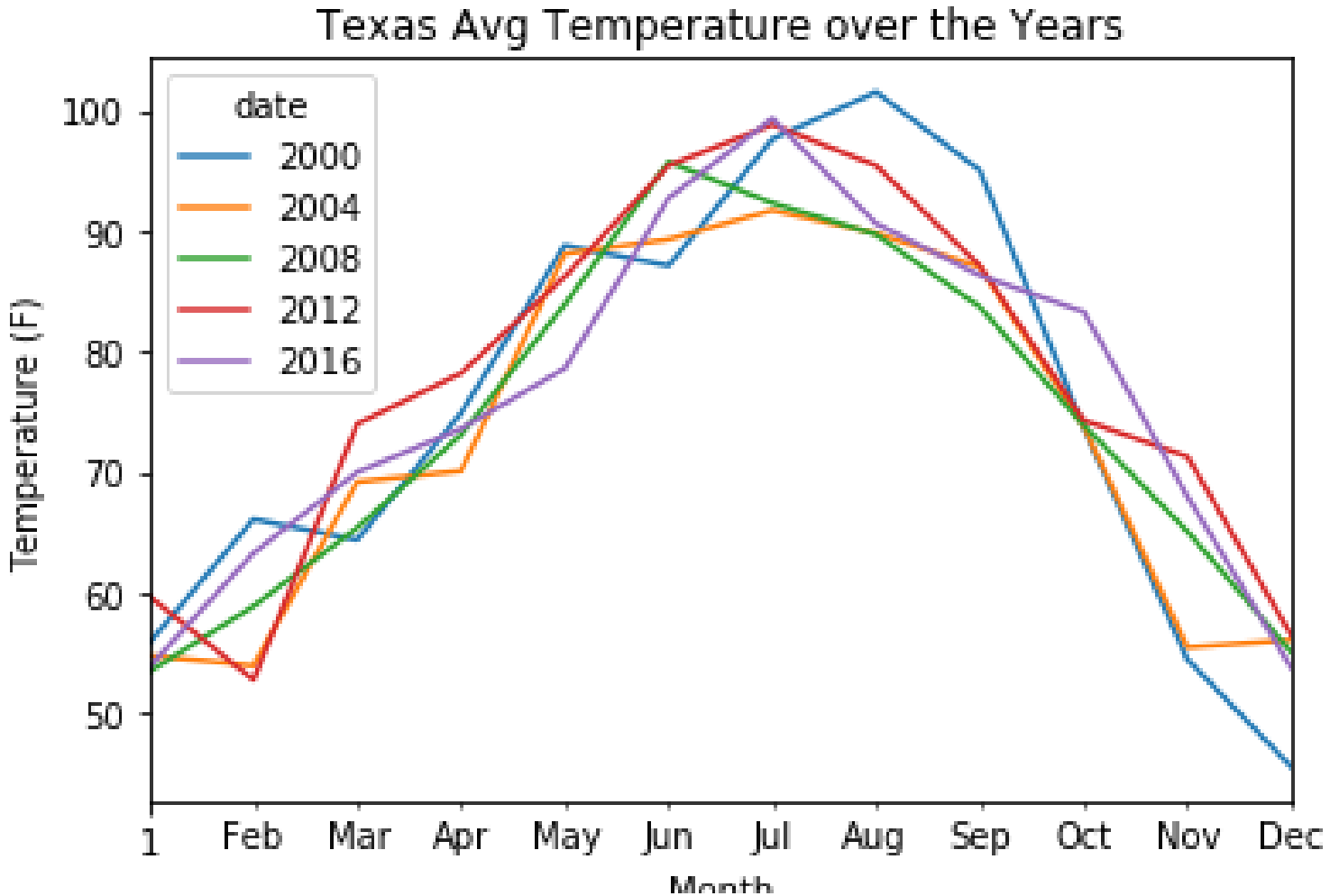
- Narrowing down search
- Pulling API with Header
- Comparing data to bees
  - Months vs Years
- Converting date
- Putting multiple line graphs onto one plot

```
dates_temp = []
dates_prcp = []
temps = []
prcp = []

#create year average temperature list
years = range(2000,2020)
year_avgs = []

#for each year from 2000 to 2020 every 5 years
for year in range(2000,2020):
    year = str(year)
    print('working on year '+year)
    tavg = []

    #add to my params for start and end date
    params['startdate'] = year + '-01-01'
    params['enddate'] = year + '-12-31'

    #make the api call
    r = requests.get(base_url,params = params, headers=headers)

    #load the api response as a json
    d = json.loads(r.text)

    #get all items in the response which are average temperature readings
    avg_temps = [item for item in d['results'] if item['datatype']=='TMAX']

    #get the date field from all average temperature readings
    dates_temp += [item['date'] for item in avg_temps]

    #get the actual average temperature from all average temperature readings
    temps += [item['value'] for item in avg_temps]
    tavg += [item['value'] for item in avg_temps]

    #get the year average over the entire year to compare with bee data
    year_avgs.append(sum(tavg) / len(tavg))
```

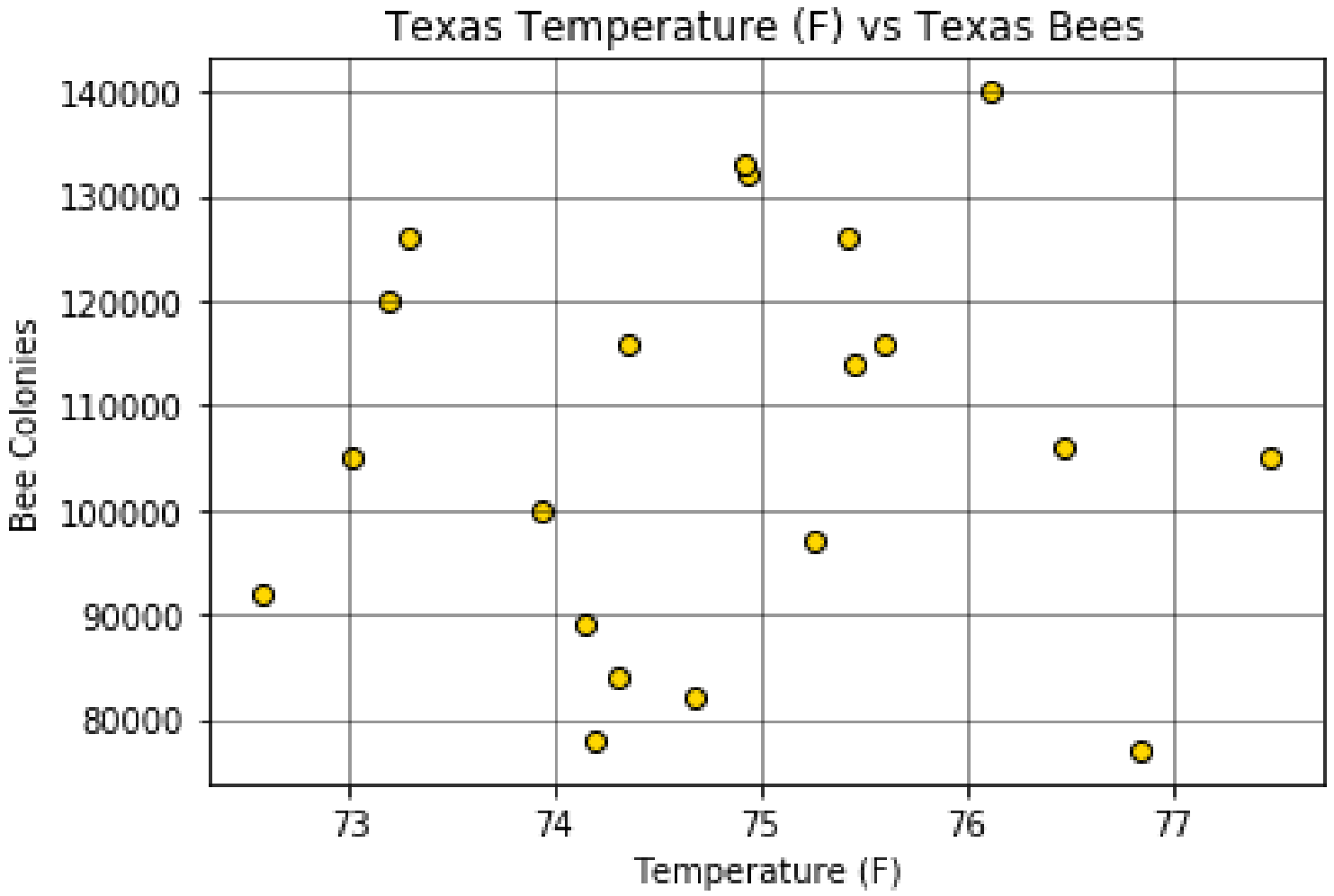
Producing data through an API call



Found the Texas monthly average temperature across 19 years.



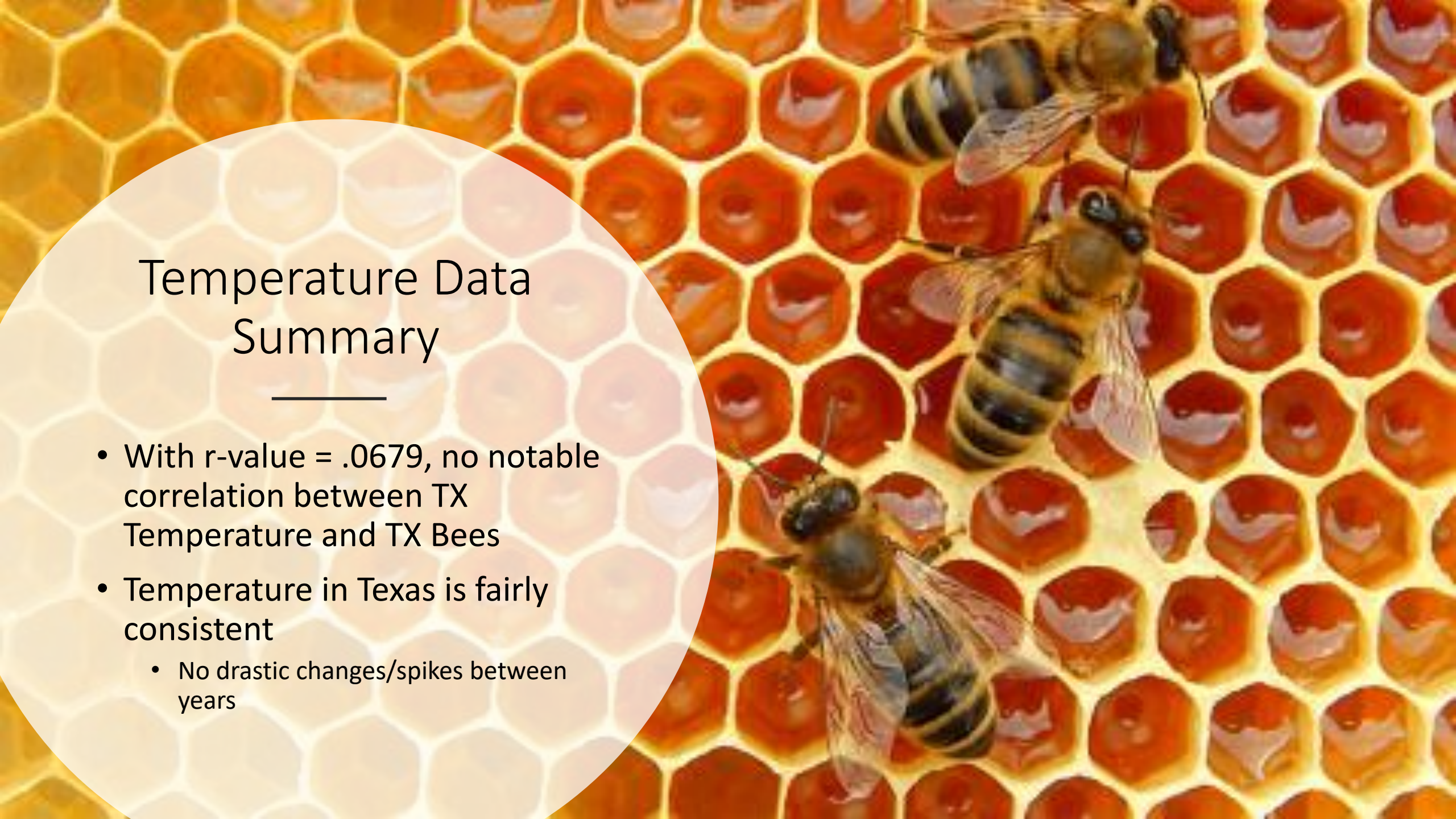
Compared years in four year increments to find any drastic changes



Tried to find a correlation between the Temperature in Texas and Bee Colonies in Texas

Temperature Data Summary

- With r-value = .0679, no notable correlation between TX Temperature and TX Bees
- Temperature in Texas is fairly consistent
  - No drastic changes/spikes between years



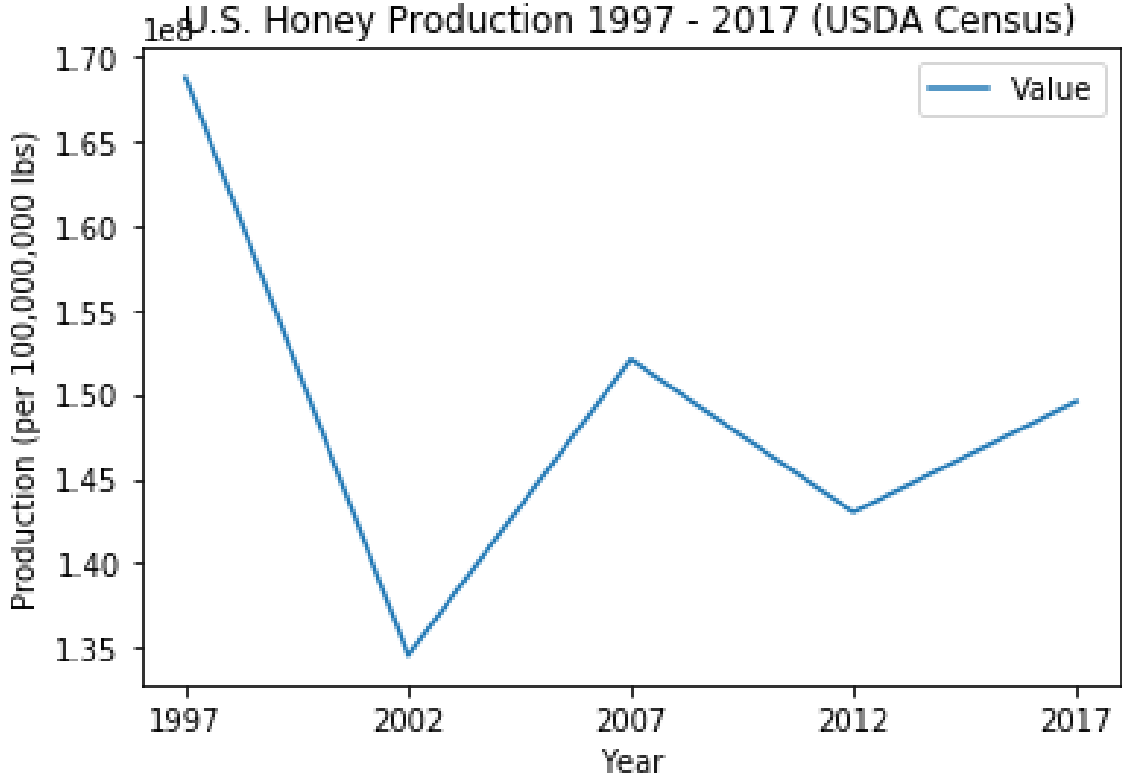
Honey Sales

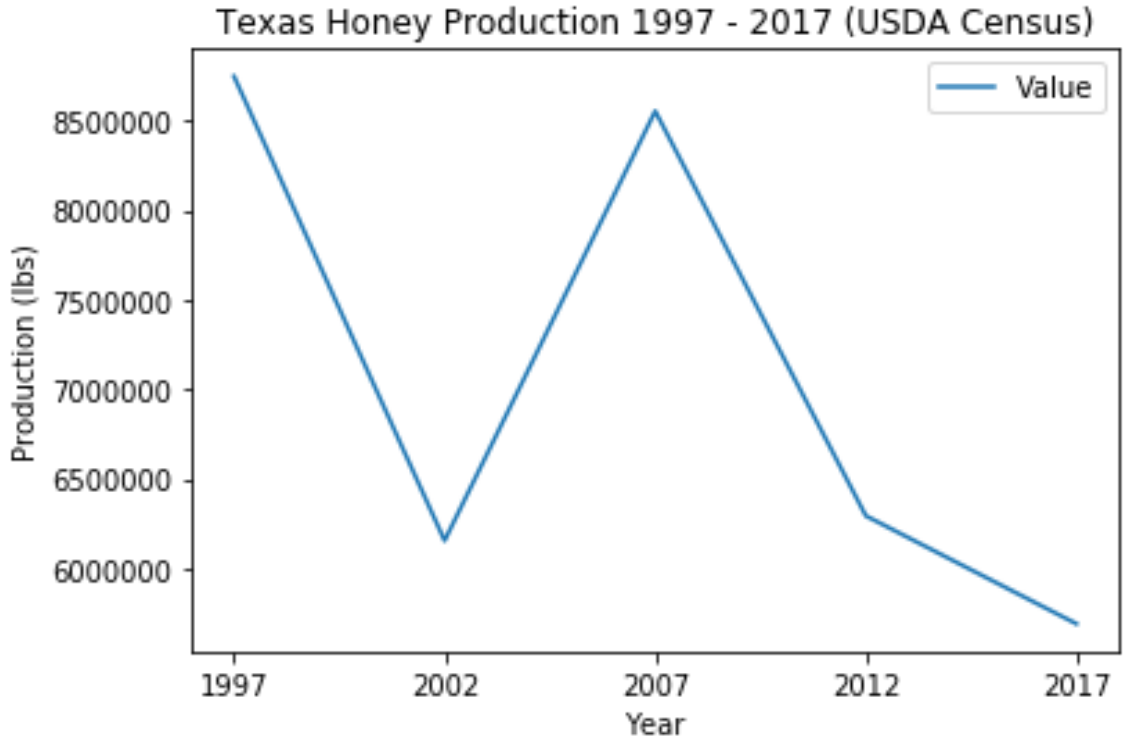
- Found data on USDA website.
  - Downloaded CSV file
- Filtered through census and survey data
- Able to find information on honey sales and production both on a state and national level.

Honey Sales Challenges

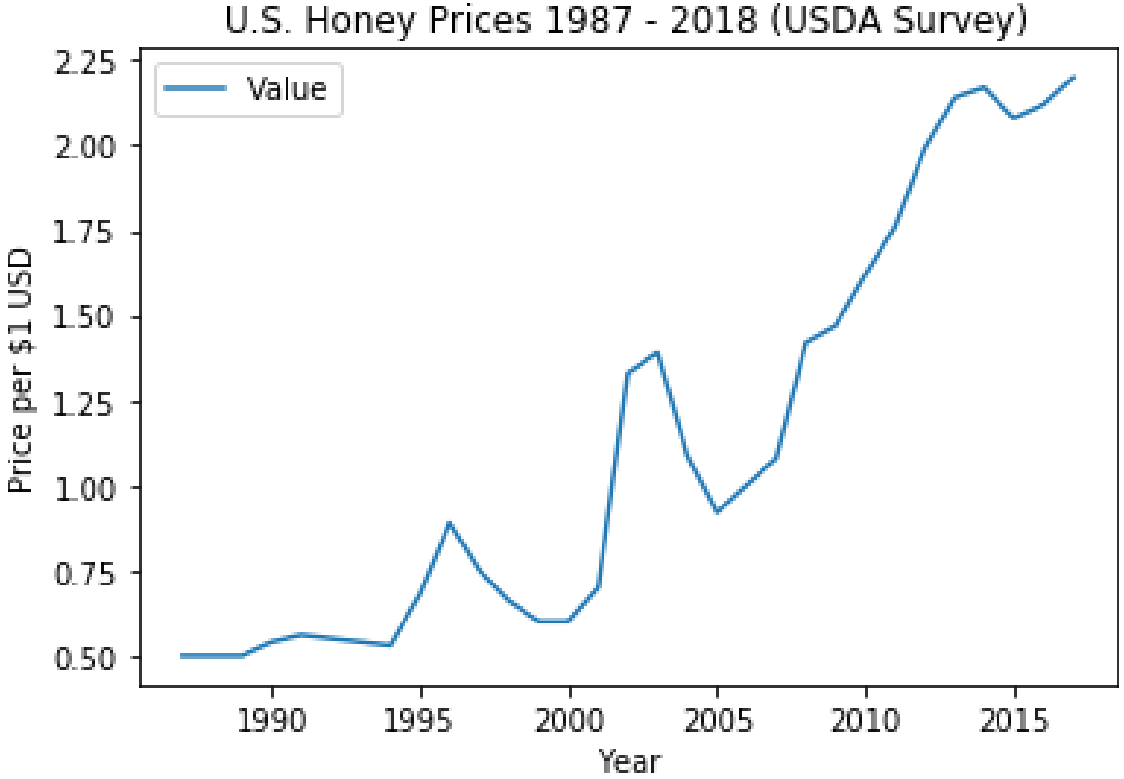
- Tricky searching for the data
- Able to find multiple datasets for honey:
  - Production
  - Location
  - Sales
- Trail and error finding the right CSVs.
  - Had to download multiple CSVs to get both the census and survey data required for the analysis.
- Using pandas and matplotlib, constructed several graphs.
- Converted cent values to dollar figures and made a multiline graph to display findings

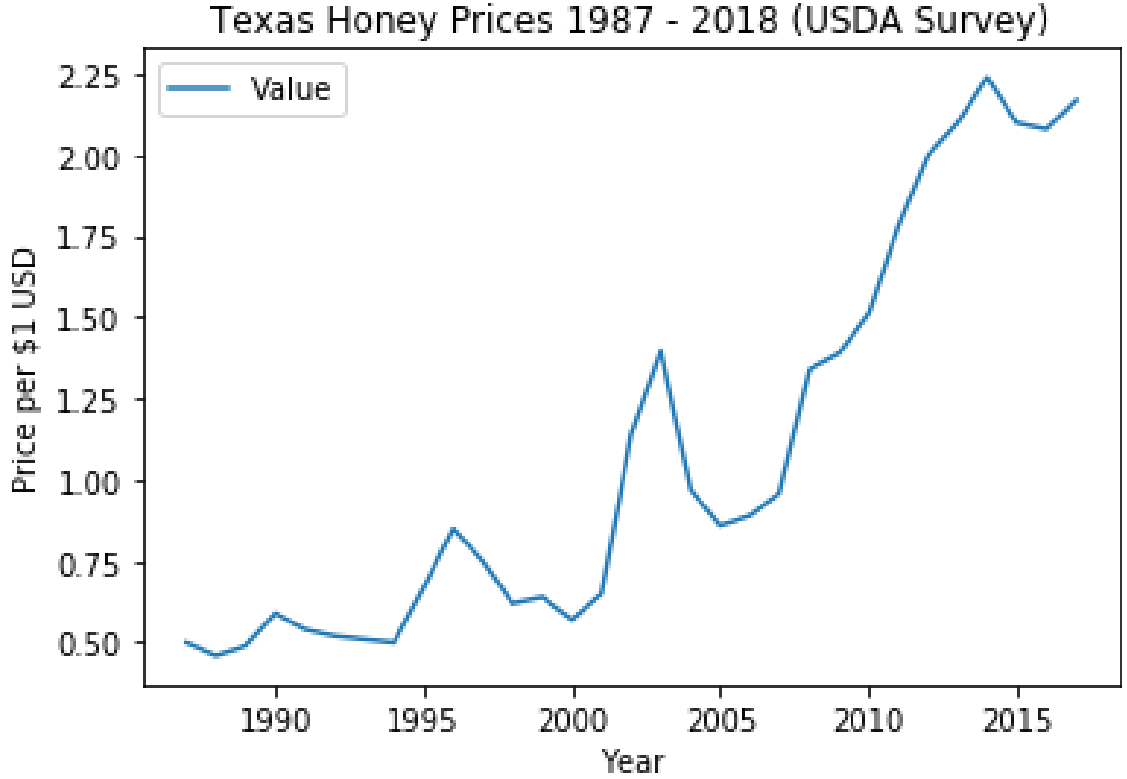
US and Texas Honey Production

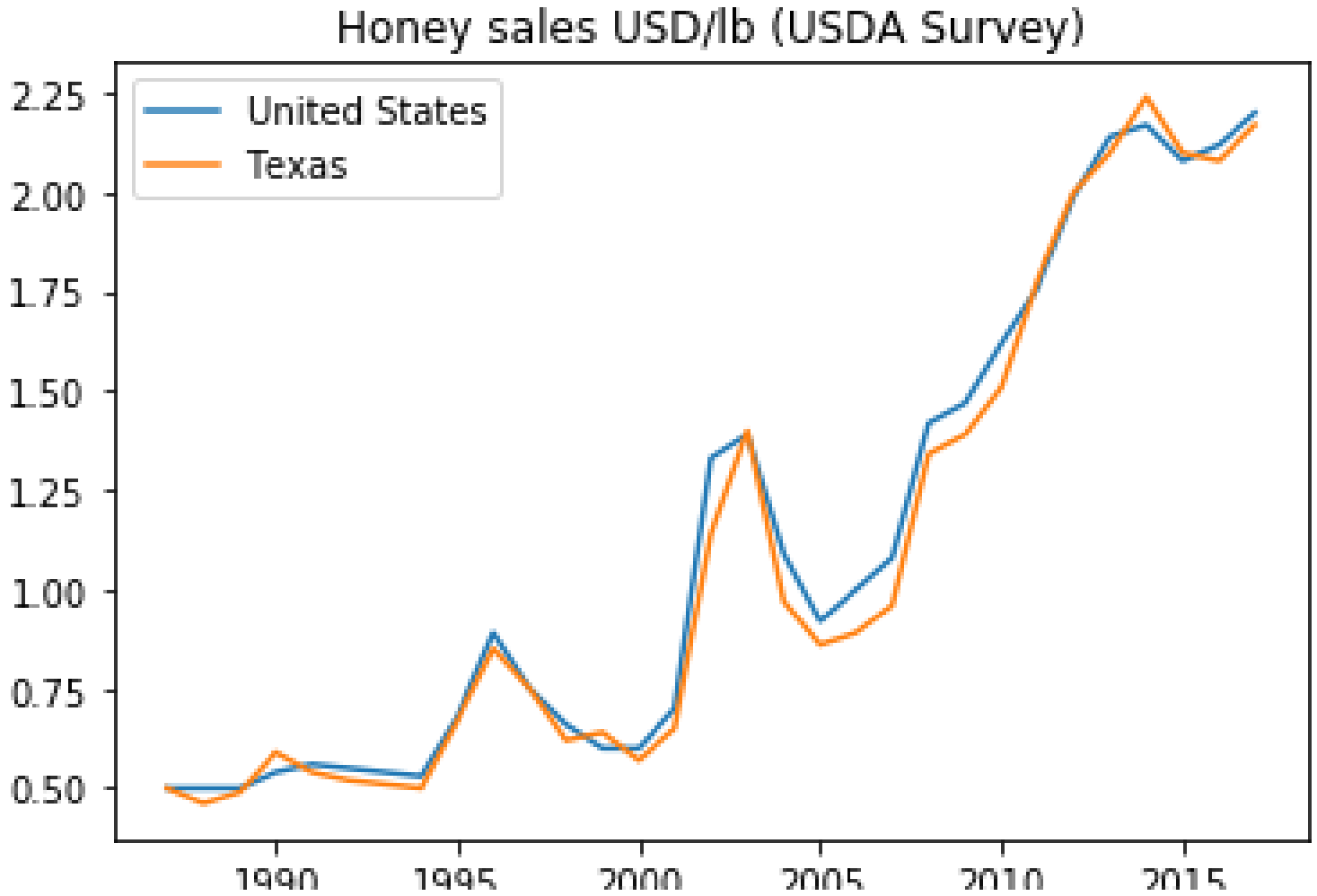




US and Texas Honey Prices







US vs Texas Honey Sales

Honey Production Summary

- Honey production sharply decreased from 1997 to 2002 and then increased from 2002 to 2007 for both Texas and the U.S.
- After production sharply declines, with Texas continuing to decline in production and the U.S. appears to be bouncing back as of 2017.
- For prices, Texas follows the national trend closely.
- There could be a number of factors that could be the cause
  - Laws
  - Taxes
  - New agricultural bills
  - Cultural and Economic shifts (such as the 2008 recession)
  - Changes in marketing methods, etc.
- How our environment and agricultural practices can affect our businesses

Conclusions/Implications

- The r-value for correlation between ambient temperature and bee population was .0679 which disproves our hypothesis.
- There was insufficient data regarding current pesticide measurements
  - Impossible to prove or disprove our second hypothesis.
- Varro Mites and Pesticides
- Sales Rising
  - Tax passed?