# Apicultural Conservation and Technology: An Investigation into Decreasing Bee Populations with Data Science

By Christopher Petrucelli

# Introduction to the Problem

- CDC publication in 2006 warning of spontaneous colony collapse
- Numerous calls to action
  - Pesticides, gardens, etc.
- What can be done?

# Beehive Monitoring Systems

- Monitor beehive activity
- Many platforms available
- Use a variety of sensors
  - Temperature
  - Weight
  - Activity
  - Etc.

# Agenda

- Important background information / terminology
- Part 1: Increase understanding of Apis Mellifera (Western Honey Bee)
- Part 2: Explore data collected by beehive monitoring solutions
  - Gather, clean, and examine available beehive data
  - Justification for collecting data
  - New ideas

# Background

- “Bees” or “Colony” - refers to the actual bees being kept
- “Hive” or “Box” - refers to the physical housing they inhabit, although there are many designs of hive
- “Frames” - panels slotted into a hive for wax comb construction
- “Brood” - refers to the young of the colony

# Part 1: Apis Mellifera

# Bee Basics

- Bees have haplo-diploid sex determination at birth
  - Non-fertilized egg = male “Drone” bee
  - Fertilized egg = female “Worker” bee
  - Female bee + royal jelly = “Queen” bee
- Roles
  - Drone: mate with queen
  - Worker: tend to brood, build comb, defend hive, gather pollen and nectar
  - Queen: mate and lay up to 1500 eggs per day for her entire life

# Bee Basics (cont.)

Drone (left) and Worker (right)

# Brood Development

- Eggs laid into hexagonal wax cells on the ‘brood frame’
- All bees go through the same stages of development
- Development time varies
  - Queen: 16 days
  - Worker: 21 days
  - Drone: 24 days
- Regulate hive conditions

Adult (left), Pupae (center), Larvae (right)

# Threats

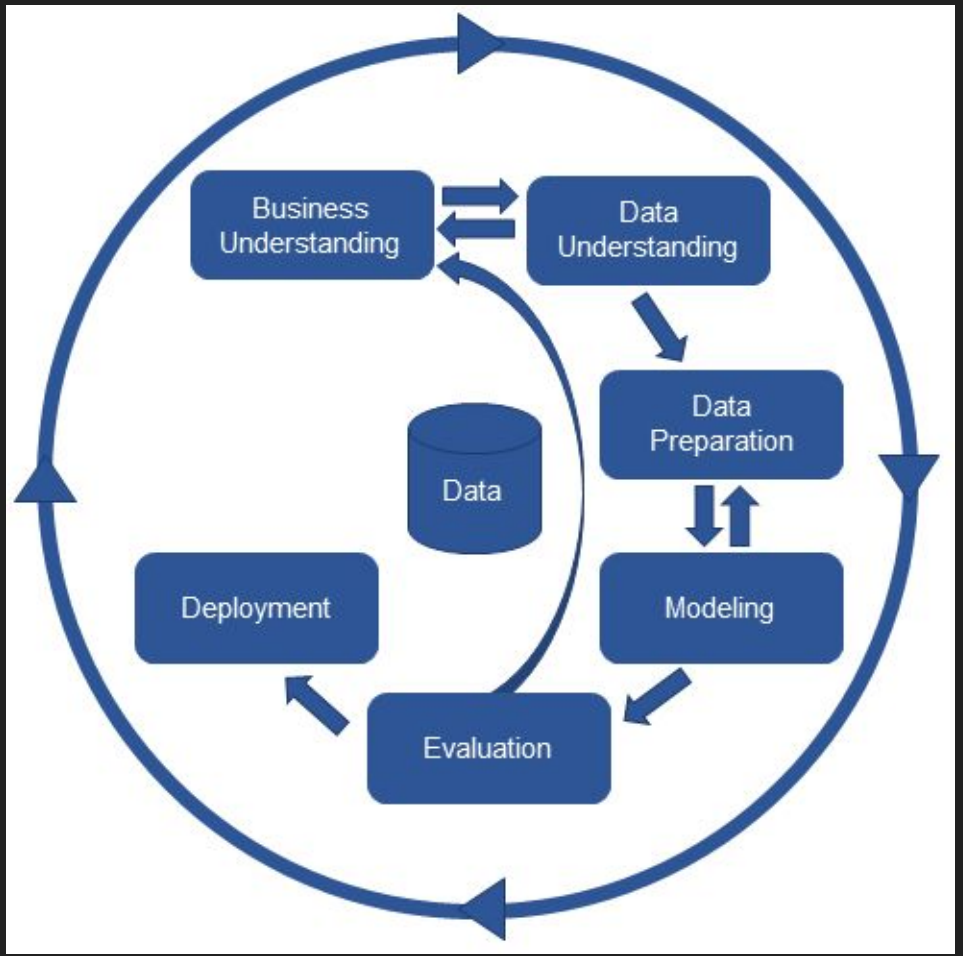
- Disease
  - Primarily affect the brood
- Parasites
  - *Varroa destructor* mites
  - Wax moths
- Pesticide/Herbicide
- Habitat changes, inadequate foraging

Adult lesser wax moth

# Part 2: Beehive Monitoring

# CRISP-DM Model

Cross Industry Standard Practice for Data Mining



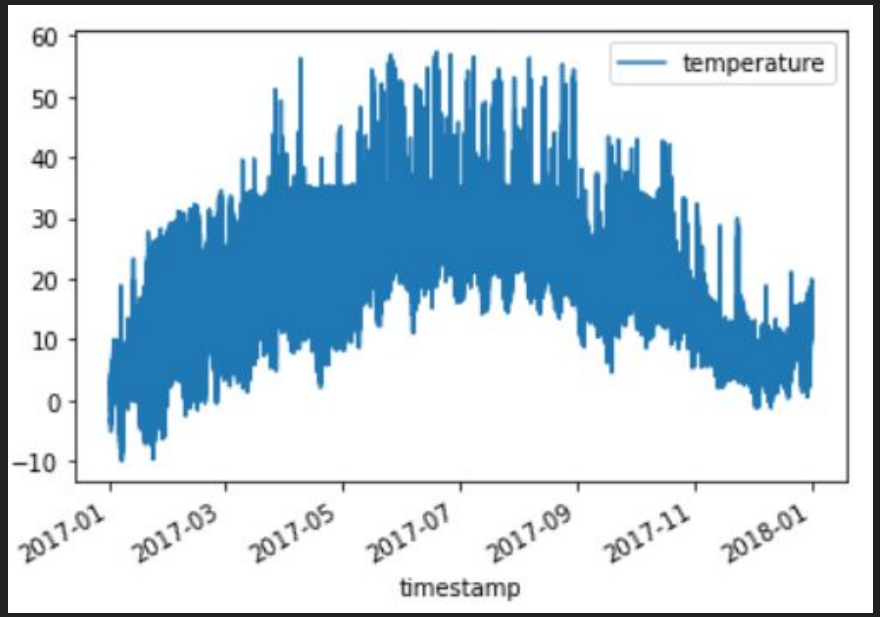
# Data Sets

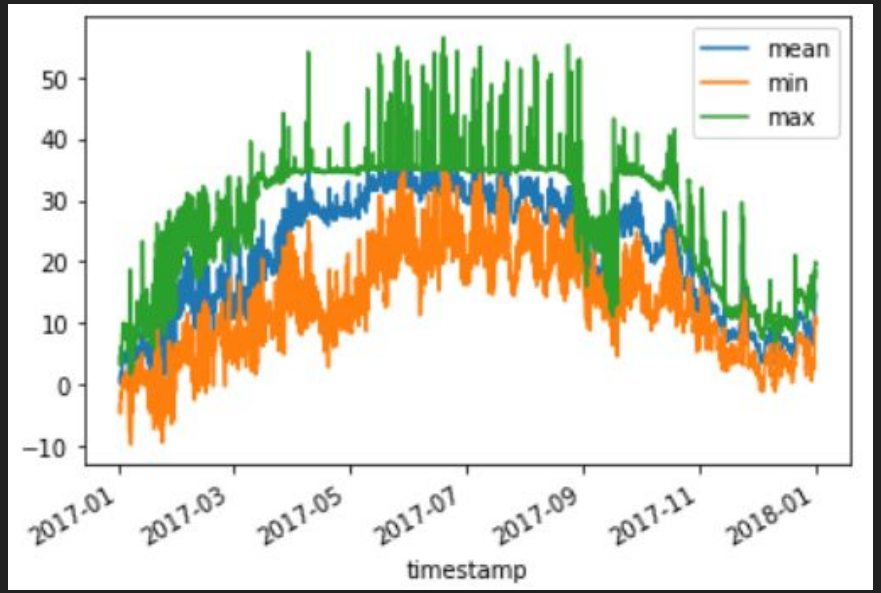
1. Hive 2017 data set
    - Collected from Kaggle
    - Data for 3 hives, unknown location U.S.
    - Contains temperature, humidity, bee flow, and weight
2. Hudson Valley, NY, 2016 - 2021 data set
    - Collected from BeeCounted Organization
    - Data for 8 hives, Poughkeepsie - Newburgh, NY area
    - Contains temperature, some humidity and weight
    - Many areas of missing readings

# Modifications: 2017 Data Set Temperature

- Temperature
  - Taken from 10 sensors, no context
- Solution
  - Take values of same timestamp, create columns for *min, max,* and *mean*

# Modifications: 2017 Data Set Temperature (cont.)



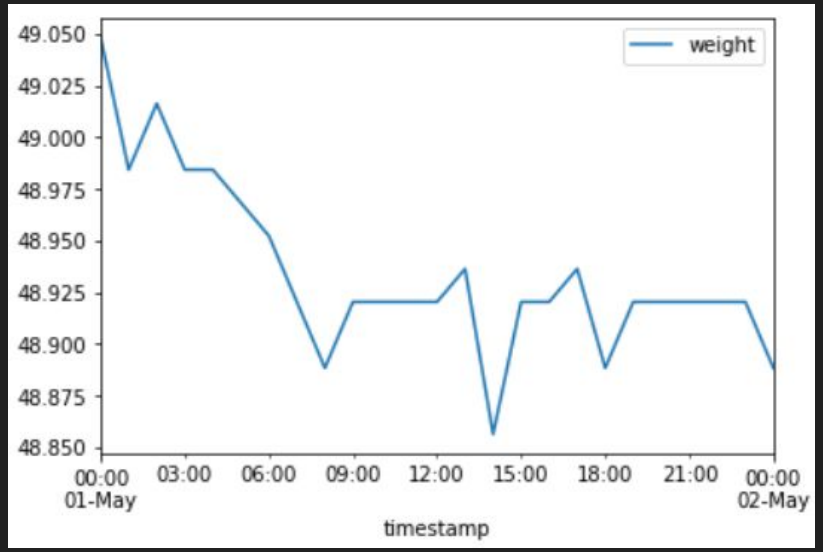

Original temperature data (left) and modified temperature data (right)

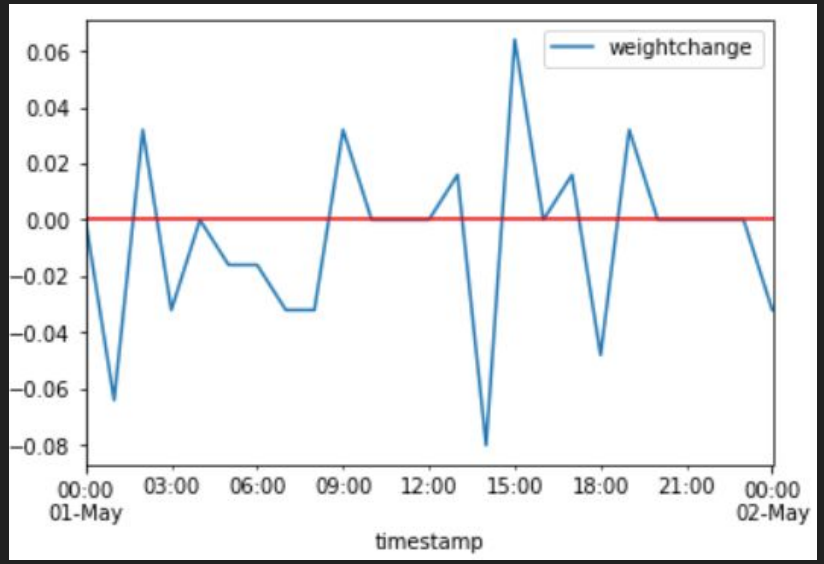
# Modifications: 2017 Data Set Flow

- Flow originally kept in a single .CSV
  - *In* flow is a value >= 0
  - *Out* flow is a value <= 0
- Two values per timestamp
- Separated into *in* and *out*

# Modifications: 2017 Data Set Weight Change

- Created a new column, *weightchange*



# Data Cleaning: Outliers

- Temperature
  - -10 to 110 degrees Fahrenheit
- Humidity
  - 0 to 100 percent
- Other
  - Use IQR

$$IQR = Q3 - Q1$$

$$[\text{Bottom range, Top range}] = [Q1 - (1.5*IQR), Q3 + (1.5*IQR)]$$

# Data Cleaning: Interpolation

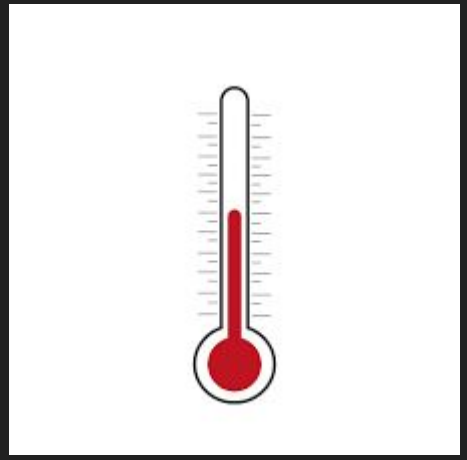
- Data had large sections of missing values
- Sensor differences or errors
- Locating a suitable variable for regression
  - High $R^2$ values
  - No data on seasonal trends

# What Data is Collected?

- Temperature
- Humidity
- Weight
- Audio
- Images
- Light / Accelerometer / E-compass
- Barometric Pressure

# Temperature

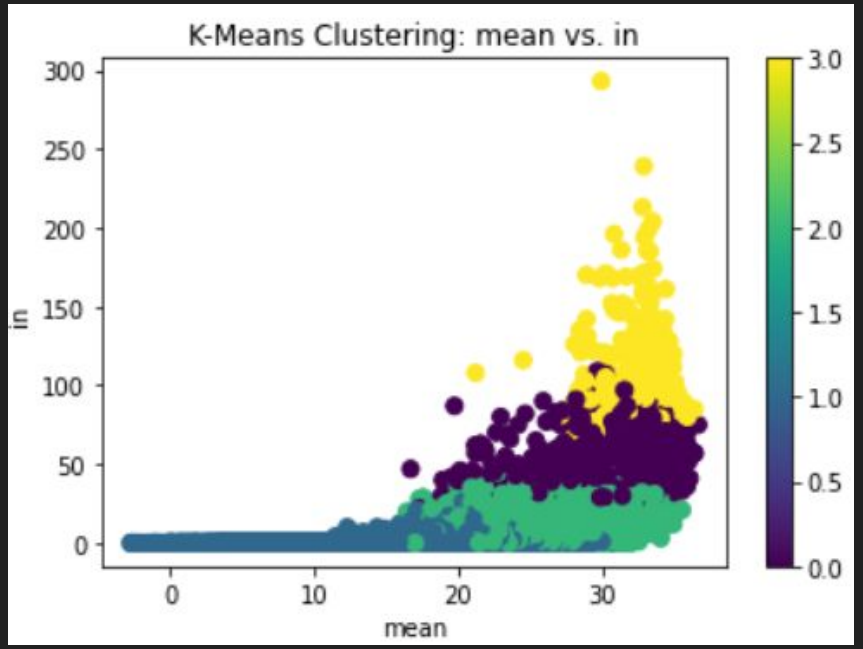
- Internal temperature is closely regulated
  - Kept within 32-35 degrees Celsius (90-95 degrees Fahrenheit)
- May be used to predict:
  - Queen-lessness or non-laying Queen
  - Lack of resources
  - Preparation for swarming
- As for external temperature...

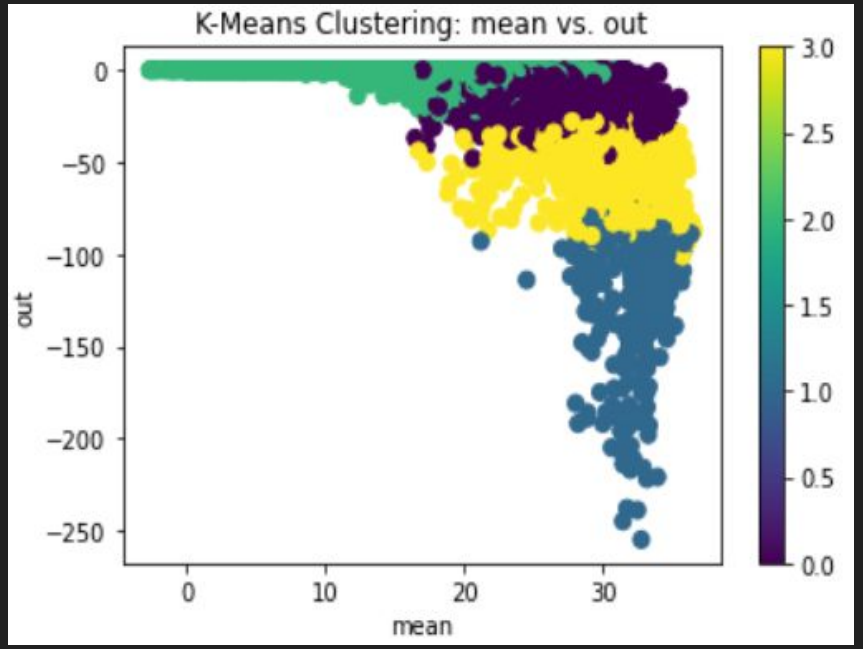


# Temperature Analysis

- Colony regulates temperature through two methods
  - Both require workers to remain within the hive
- Hypothesis: If temperature values are within the necessary range for brood development, then will colony productivity increase?
  - In/Out as a measure of productivity

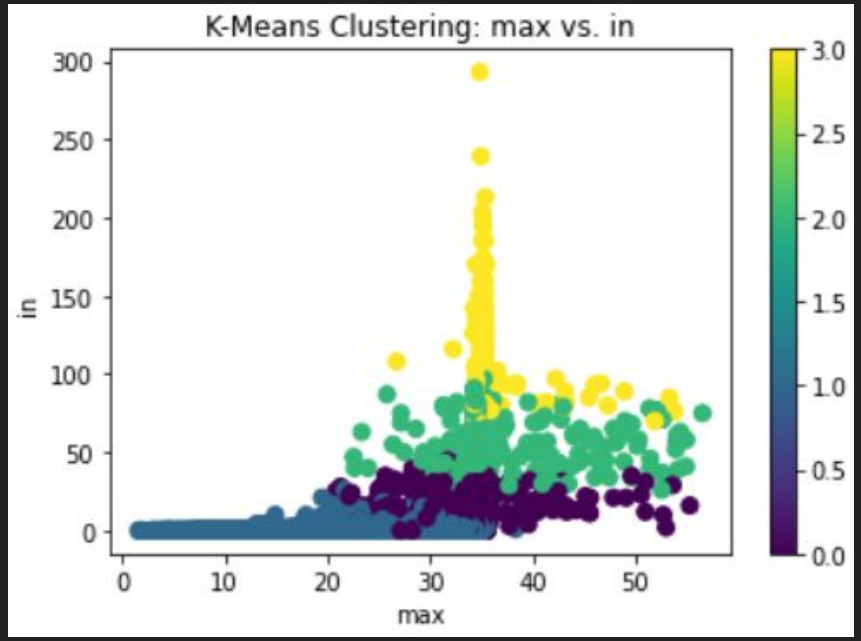
# Temperature Analysis (cont.)

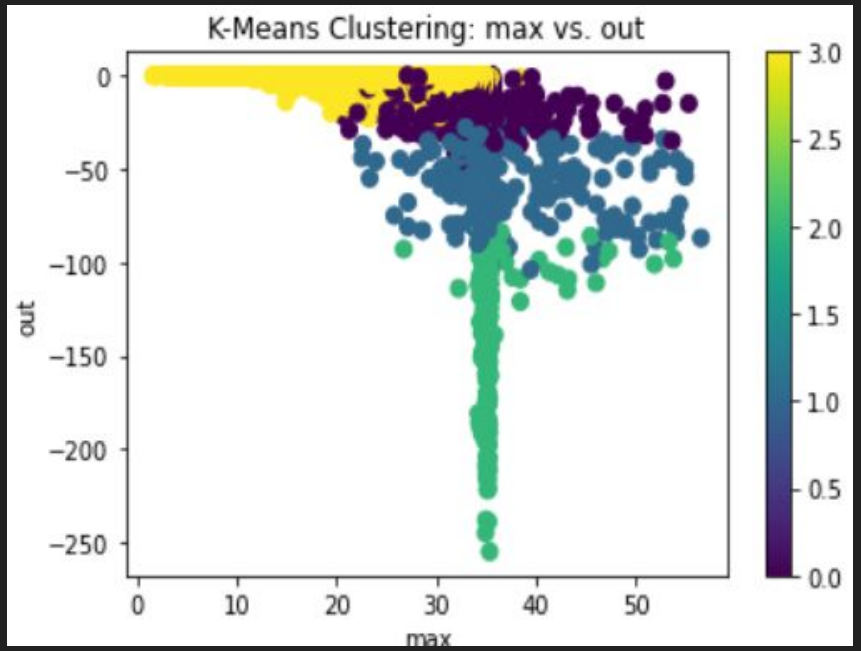



K-Means Cluster using a K value of 4, plotting mean temperature against hive flow

# Temperature Analysis (cont.)




K-Means Cluster using a K value of 4, plotting max temperature against hive flow

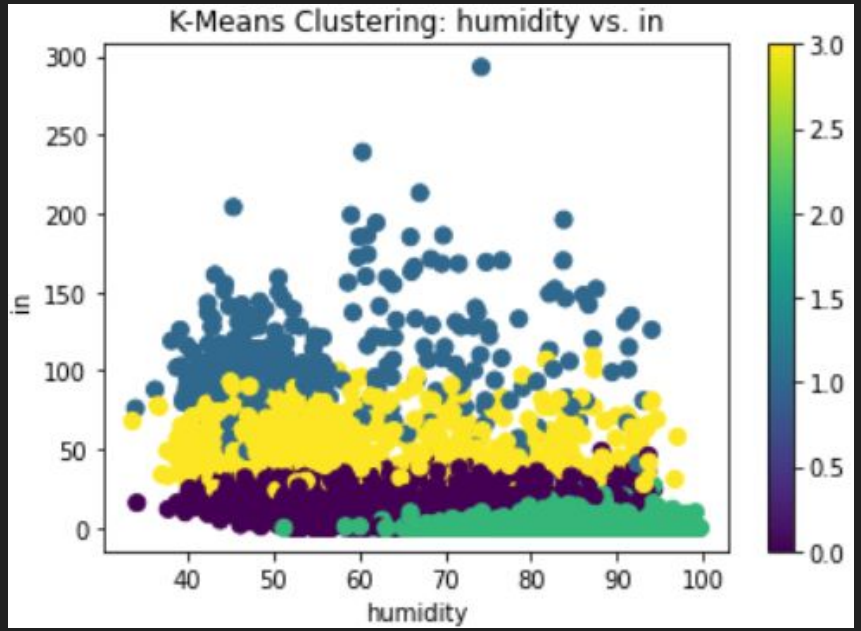
# Temperature Analysis (cont.)

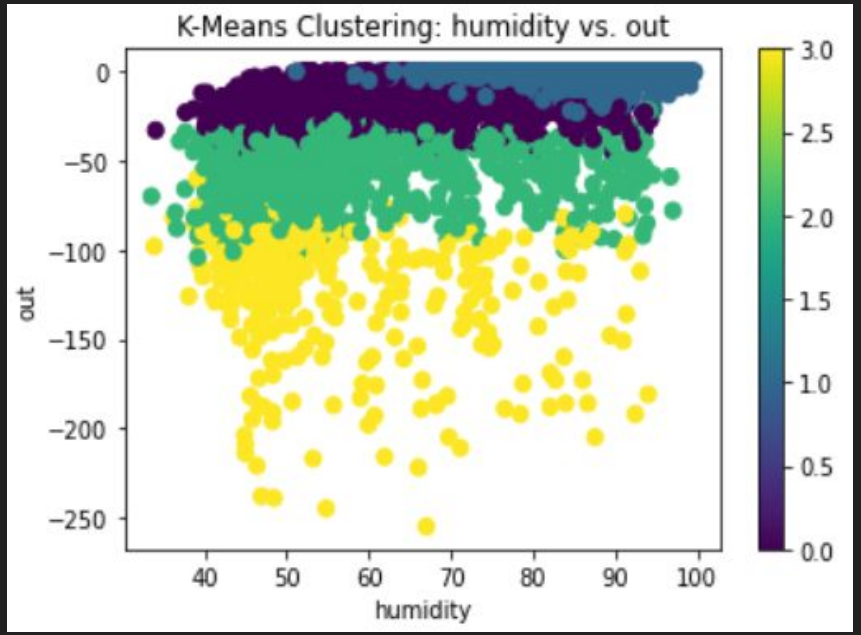
- Hypothesis was supported by available data
- Opportunities for further experimentation
  - Assist in regulating temperature to boost productivity?

# Humidity

- Also regulated by workers
  - Optimal between 50-60%, safe between 40-80%
- Large variations in internal humidity following a change to external humidity might indicate a colony is not healthy
- High humidity increases rate of occurrence of some diseases, but also limits reproduction of Varroa mites
- Could the same hypothesis prove true with humidity?

# Humidity Analysis




K-Means Cluster using a K value of 4, plotting humidity against hive flow

# Weight

- Allow for estimation of hive resources
  - Growth during active season, consumption during winter
- Made more useful the more data or context is available
  - Hive type, size, wax comb development
  - Seasonal trends
- Sudden weight gain?

# Other Data

- Light / Accelerometer / E-compass
  - Anti-theft
- Barometric Pressure
  - Potentially related to bee agitation
- Audio
  - Predict Queen birth, end to egg-laying, swarming
- Images
  - Monitor activity, potentially resource abundance

# New Application: Images

- Varroa mites are very common
  - Infestation allows disease to spread rapidly
- Screen and bottom board
- Daily images or ML predictions

Varroa Destructor on bee pupae

# New Application: Images (cont.)

# New Application: E-compass

- “Drifting” - phenomenon in which bees will return to the incorrect hive when multiple hives are in a line
  - Hives on the ends of rows will receive the drifting bees more frequently
- Weight may also help to diagnose this problem
- E-compass headings can be compared
  - Circular (facing outwards) or random placement and facing of hives will prevent this

# Conclusion

- Are beehive monitoring systems effective in ensuring the health and safety of bee populations?
  - Some features are not strictly necessary, but...
  - Extra oversight
  - Machine learning
- Data availability is limited
  - Some organizations, like BeeCounted, are working to remedy this
- New technology, new opportunities
  - Audio analysis is just one example

# References and Links

Github repository (and paper):

https://github.com/AFineSortie/Apicultural-Conservation-and-Technology