

Project Title: To Bee or not to Bee: Colony Collapse in the U.S.

Team Name: Honey Bears

Our question: To what extent do factors such as the usage of various pesticides, rainfall, temperature, and more impact the number of honeybee colonies as well as honey yield? In modeling honeybee colony decline, can we effectively predict and simulate honeybee production in a given state and/or region?

Why our question is interesting: In recent decades, the number of honey bees has significantly declined, a phenomenon called colony collapse.¹ Researchers have identified many reasons for this decline, among them habitat loss,² usage of harmful pesticides,³ and infectious diseases.⁴ Declining populations pose a huge threat to the food system: honeybees pollinate roughly one-third of all food eaten by Americans,⁵ including crops ranging from pumpkins to cranberries to almonds.⁶ Our project will analyze honeybee data from 1998 to 2017 to determine the causes and trends of honey bee population decline. Our findings will allow us to formulate effective policy interventions to strengthen hives, ensuring long-term viability in the American food system.

Why our question is challenging: Finding, cleaning, and joining data will pose a challenge. While our primary [source](#) includes data on honey production and the number of bee colonies, it is not a rich source of feature data: e.g., temperature, rainfall, level of urbanization, and more. In addition, as described above, researchers have identified a multitude of factors that impact honeybee populations, which further adds complexity. Thus, there are many components to consider, requiring that we have a wide array of feature data to build effective models.

What analytical methods: We will complete two main analyses to address our research questions:

- 1) Prediction: Predict the number of honeybees and level of honey production in a given state and/or region
 - a) Methods: Regression (Ridge, LASSO, linear, polynomial); Regression Trees
- 2) Simulation: Simulate random events that affect our prediction model – e.g., pesticide drift, severe weather
 - a) Methods: stochastic sampling of continuous distributions, Monte Carlo method for discrete events

Plan of Action: We have decided to divide into two teams for our two tasks:

Prediction Team: Irfan, Sara

Simulation Team: Kai, Matt, Jack

Cycle 1: Oct. 17-31	Cycle 2: Nov. 1-14	Cycle 3: Nov. 15-30	Cycle 4: Dec 1-11
-Find data to strengthen analysis -Clean and merge data -Select 2-4 models (per team)	-Write code for models -Deploy models and assess model quality -As needed, redefine scope and reintroduce predictor variables	-Finetune model parameters -Create and finalize PowerPoint presentation -Outline final project report	-Rehearse PowerPoint for in-class presentation -Finalize project report

Relevance to course: Our project leverages advanced quantitative methods to address a fundamental problem: long-term food security. Without honeybees, the food system in the US would collapse, one-third of the American diet vanishing. We will leverage techniques and skills we have gained in this course – such as regression, CART, and simulation – to answer our research questions. Our project aligns with the timeline of the course because we will have learned the skills and tools necessary to complete the project prior to the final deadline.

¹ [EPA, "Pollinator Protection: Colony Collapse Disorder."](#)

² [Dhruba Naug, "Nutritional stress due to habitat loss may explain recent honeybee colony collapses."](#)

³ [EPA, "Pollinator Protection: Colony Collapse Disorder."](#)

⁴ [Lewis J. Bartlett, et al, "Persistent effects of management history on honeybee colony virus abundances."](#)

⁵ Ibid.

⁶ [FDA, "Helping Agriculture's Helpful Honey Bees."](#)