Sierra Harris
Catherine Aznoe
Cristina Lefemine

Analysis of Hummingbird Observations in California

eBird Observations

California

10 Hummingbird Species
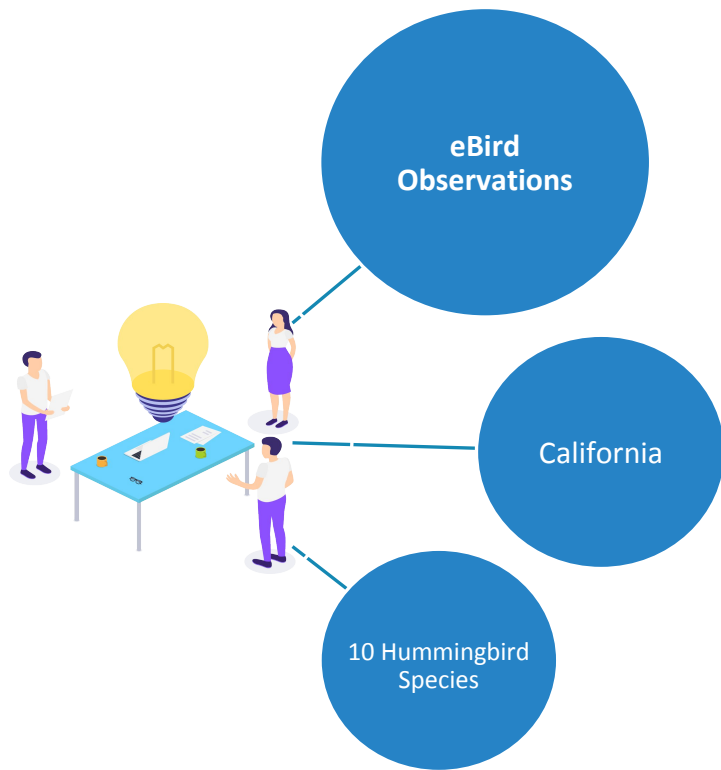
PROJECT VISION
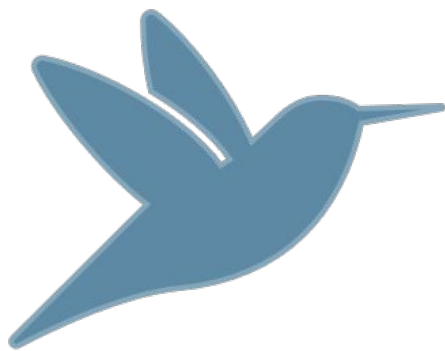
DATA EXPLORATION

What drives hummingbirds to be at these observed locations?
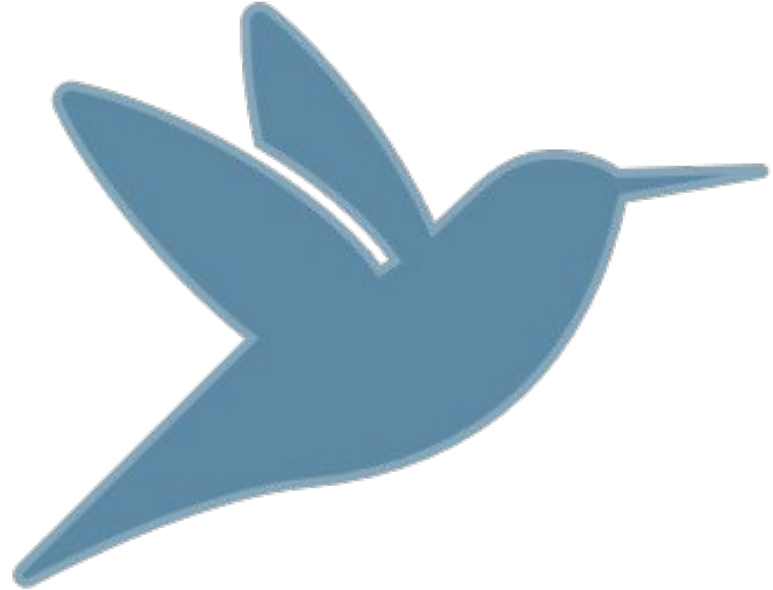
WHAT ENVIRONMENTAL VARIABLES PLAY A ROLE?

Is there a relationship between air quality and hummingbird observations, specifically in California?
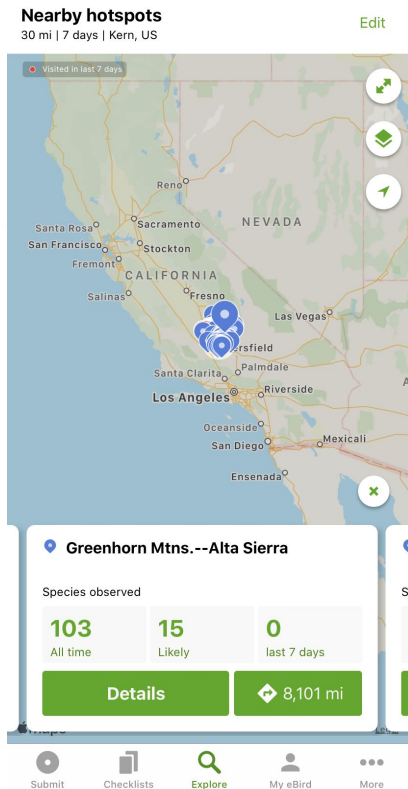
Do weather events impact hummingbird observations?

WHAT ABOUT OTHER VARIABLES?

DATA SOURCES

Rate of Hummingbird Observations on EBirds Over Time
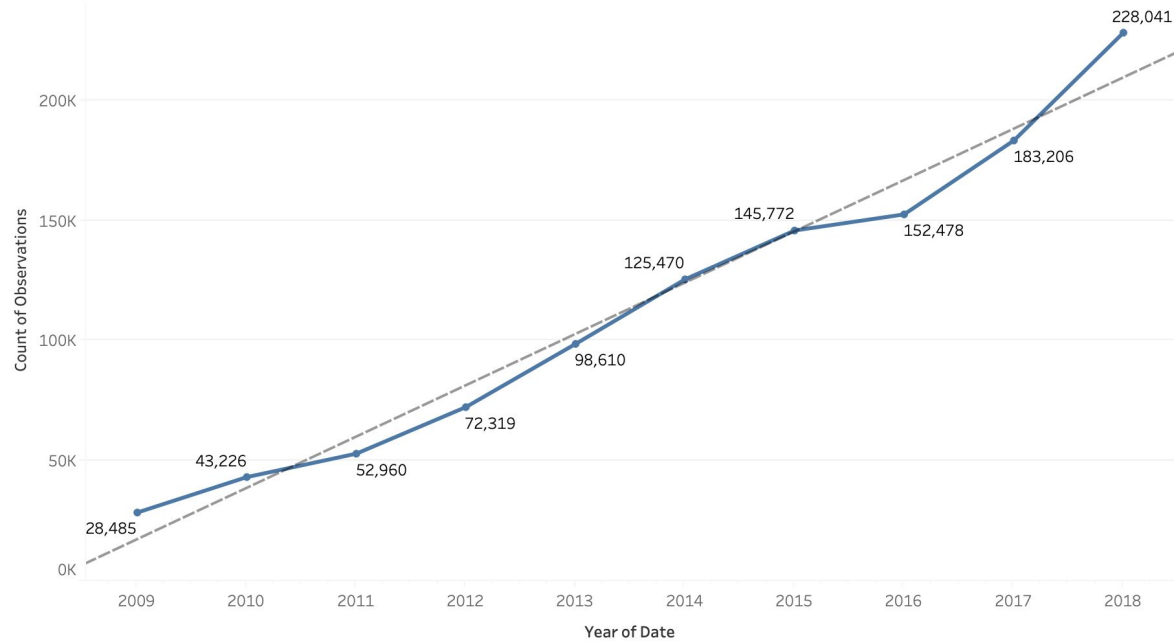
The trend of count of bird_aqi_weather.csv for Date Year.

DATA EXPLORATION & ANALYSIS

California Hummingbird Observation Average by Air Quality Category

The plot of Moving Average of Count of bird_aqi_weather.csv for Date Year. Color shows details about Category. The marks are labeled by Category.

DATA EXPLORATION & ANALYSIS

Hummingbird Observation Locations (2009-2019)

Species
- Amazilia violiceps
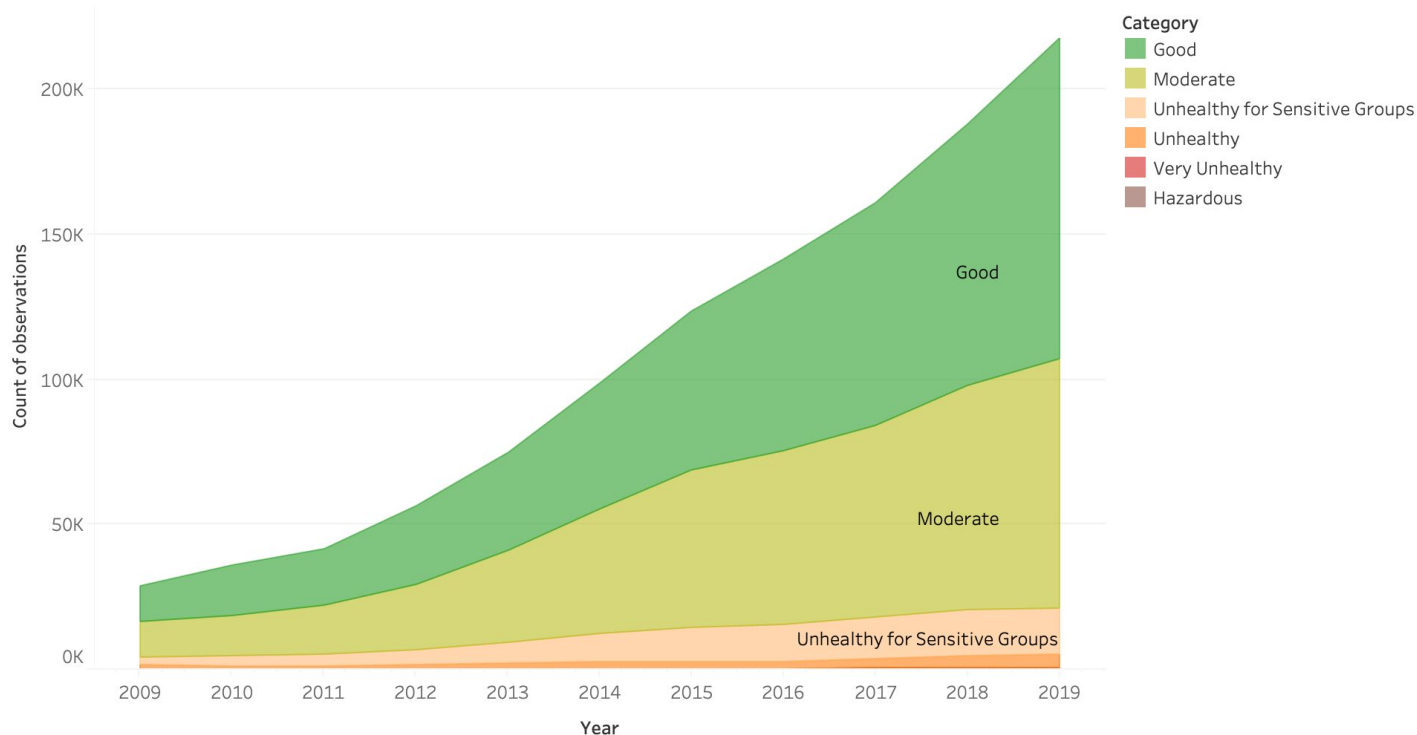- Archilochus alexandri
- Archilochus colubris
- Calypte anna
- Calypte costae
- Cynanthus latirostris
- Selasphorus calliope
- Selasphorus platyce..
- Selasphorus rufus
- Selasphorus sasin

© 2020 Mapbox © OpenStreetMap
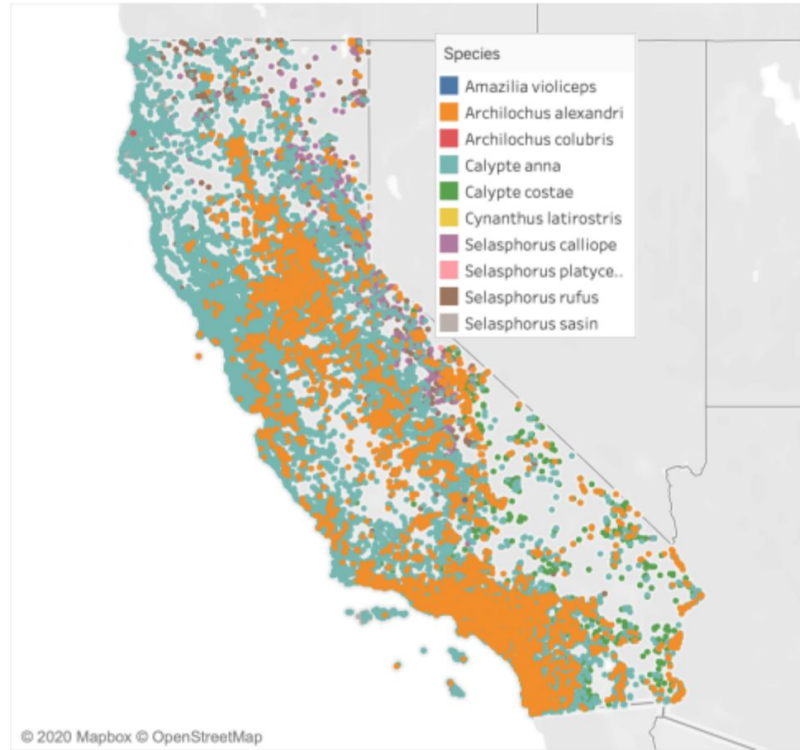
Map based on Longitude and Latitude. Color shows details about Species.

DATA EXPLORATION & ANALYSIS

DATABASE

	Data Processing	Extract & Transform: Jupyter Notebook, Python, Pandas
	Database Setup	Load: Google CoLab, Pyspark, AWS RDS, PostgreSQL
	Table Joins	Database: pgAdmin

MACHINE LEARNING

Data Processing

Clean the data
Joins → pgAdmin
Merge → Pandas

Features

Primary focus
Correlation matrix
Number of features
Conceptual relationships

Training & Testing Sets

Y value → bird species
X value → features

Model Choice

Supervised
Logistic Regression

Accuracy Scores

Training → 0.7191009160217761
Testing → 0.7201529652327877

AVES
Analysis of hummingbird observations in California.

We took the hummingbird observation locations and classified them by county for analysis. Before we classified them, we mapped the observation locations to see a good distribution of the observed species.
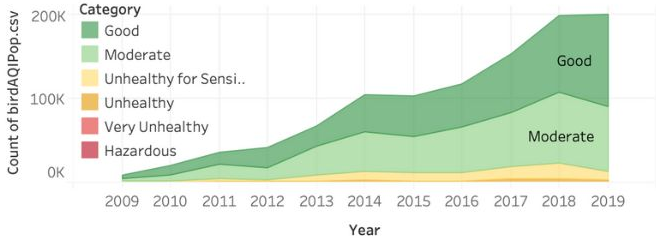
From 2009-2019 we can see that the popularity of the eBird app has steadily grown. We will examine what hummingbird species users in California may expect to observe based on the AQI, population, county and weather attributes. To predict the observations, we examined population and AQI for the past hummingbird sightings.

Hummingbird Observations, AQI & Population in California (2009-2019)
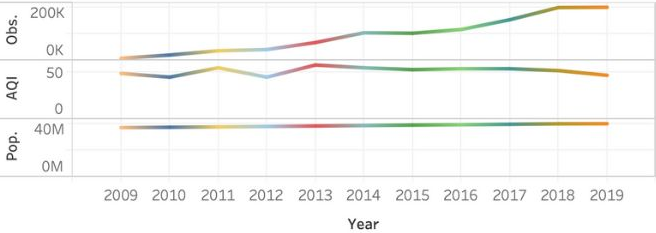
In initial analysis we could see far more observations occured during days with good AQI. While the overall observations grew, we can see that observations above a moderate AQI did not steadily grow.

California Hummingbird Observation Average by Air Quality Category

Hummingbird Observation Locations (2009-2019)

Species
- Amazilia violiceps
- Archilochus alexandri
- Archilochus colubris
- Calypte anna
- Calypte costae
- Cynanthus latirostris
- Selasphorus calliope
- Selasphorus platyce..
- Selasphorus rufus
- Selasphorus sasin

© 2020 Mapbox © OpenStreetMap

Map based on Longitude and Latitude. Color shows details about Species.

The final dataset was tested with different machine learning models with sklearn. Accuracy scores were compared and the logistic regression model was used for the prediction model to predict the species of hummingbird likely to be observed based on input features.

	Linear Discriminant Analysis	Decision Tree	Decision Tree with Max Depth	Logistic Regression
Training Accuracy Result	0.70808	0.73276	0.72127	0.7191
Testing Accuracy Result	0.70925	0.71317	0.723	0.72015

HUMMINGBIRD PREDICTOR

Enter Air Quality Index (0–250)

Enter Year (2009–2019)

Enter Population

FUTURE ANALYSIS

→ Obtain and add bird observation data from more resources, i.e., U.S. Fish & Wildlife Service

→ Expand analysis to more sentinel bird species and/or expand locations.

→ Analyze more features such as temperature, microclimates and different land uses.

WHAT WOULD WE DO DIFFERENTLY?

CITATIONS

Global Biodiversity Information Facility, Citation GBIF.org (10 October 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.2rhbnh

Global Biodiversity Information Facility, Citation GBIF.org (10 October 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.fsvwgj

Environmental Protection Agency, Daily Summaries, https://aqs.epa.gov/aqsweb/airdata/download_files.html#Daily

National Oceanic and Atmospheric Administration, Storm Events Database, https://www.ncdc.noaa.gov/stormevents/choosedates.jsp?statefips=6%2C CALIFORNIA#

Federal Communication Commission (FCC), Census Block for Developers, https://geo.fcc.gov/api/census/#!/block/get_block_find

California Association of Counties, Data and Research, https://www.counties.org/data-and-research

QUESTIONS