

# “Cheep” Shots: Suspense at the Wildlife Preserve



Leigh Ann Kudloff  
Paul O'Leary  
November 2021

# Scenario

- Possible endangerment of the Rose-Crested Blue Pipit near Mistford and the Boonsong Lekagul Wildlife Preserve
- Investigation implicating Kasios Office Furniture, a manufacturing firm
  - Kasios presents itself as extremely eco-friendly
  - Kasios supposedly used banned substance, Methylosmolene
  - Kasios dumped waste in NW region of preserve
  - Methylosmolene detected in smokestack emissions
- Kasios claims the investigation analysis is flawed and biased.
  - Kasios reports plenty of Rose-Crested Blue Pipits
  - Kasios provided set of recent recorded bird calls with locations
- Pangera Ornithology Conservation Society concerned and suspicious
- Town of Mistford and preserve rangers satisfied recordings back Kasios' claims
- Mistford College does not have a Pipit expert and provides collection of bird calls vetted by ornithology groups

# Scenario Goals

- Using the bird call collection and the included map of the Wildlife Preserve, characterize the patterns of all of the bird species in the Preserve over the time of the collection. Please assume we have a reasonable distribution of sensors and human collectors providing the recordings, so that the patterns are reasonably representative of the bird locations across the area. Do you detect any trends or anomalies in the patterns?
- Turn your attention to the set of bird calls supplied by Kasios. Does this set support the claim of Pipits being found across the Preserve? A machine learning approach using the bird call library may help your investigation. What is the role of visualization in your analysis of the Kasios bird calls?
- Formulate a hypotheses concerning the state of the Rose Crested Blue Pipit. What are your primary pieces of evidence to support your assertion? What next steps should be taken in the investigation to either support or refute the Kasios claim that the Pipits are actually thriving across the Boonsong Lekagul Wildlife Preserve?

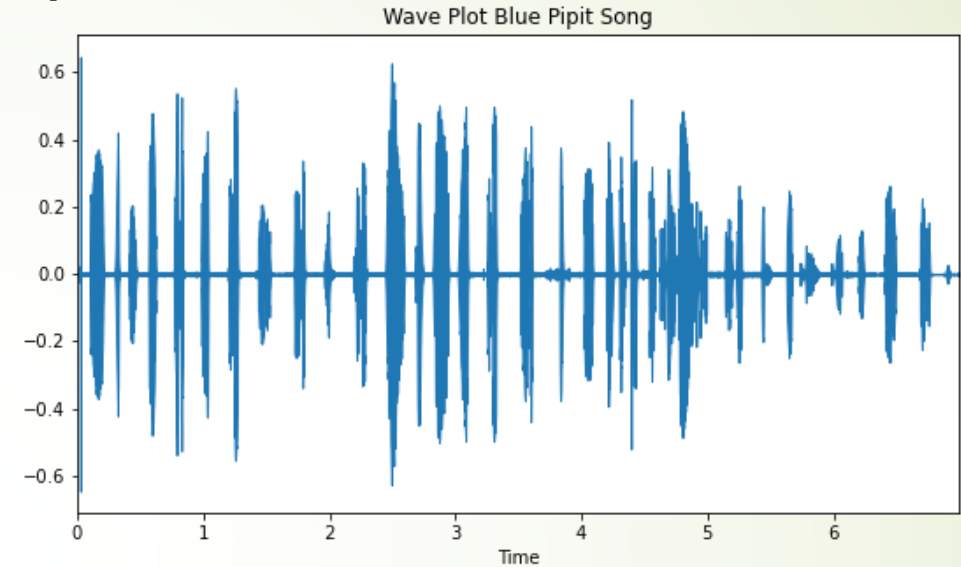
# Bird Call Collection

- Mistford College dataset has 2081 rows with these features:
  - ID
  - Name of Bird—19 different bird varieties
  - Vocalization Type (Song or Call)
  - Quality of Recording (A, B, C, D, E, none)
  - Time of Day
  - Date—1983 to 2018
  - Location (X and Y)—on a 200 by 200 grid
- Kasios dataset has 15 rows with these features:
  - ID
  - Location (X and Y)
- Files (mp3) of bird recordings matched to ID for both Mistford College dataset and the Kasios dataset

# Examples, Visual Representations

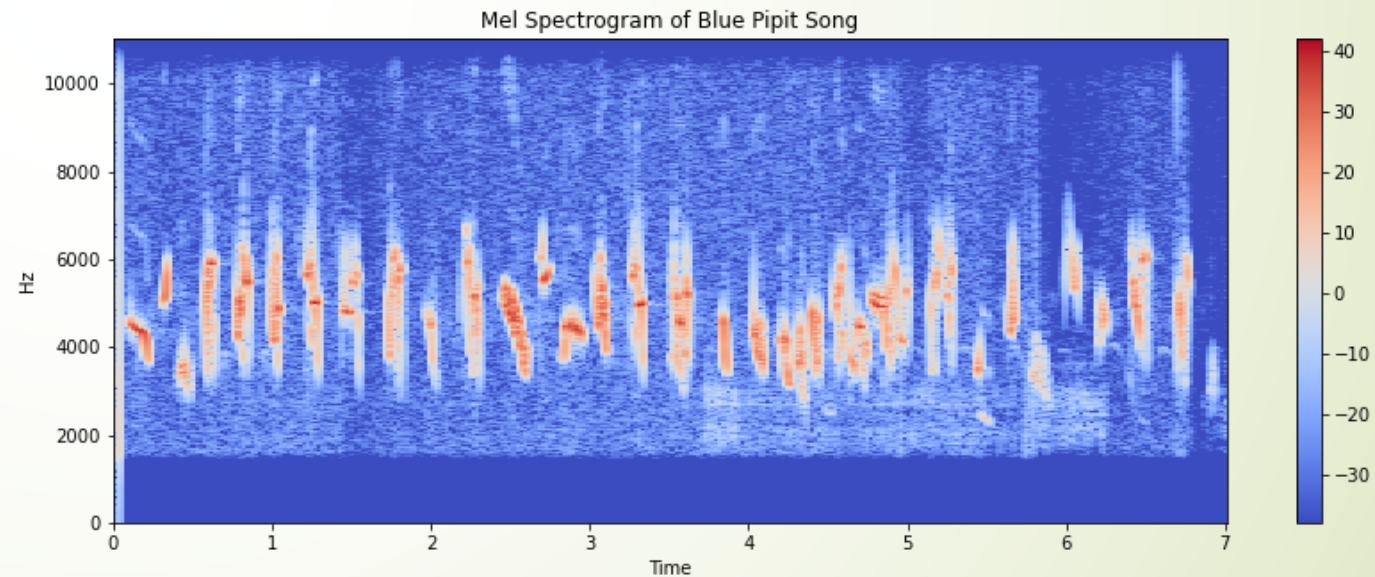
## Rose-Crested Blue Pipit Song

Wave Plot is the sound amplitude (loudness) over time in dB



## Rose-Crested Blue Pipit Song

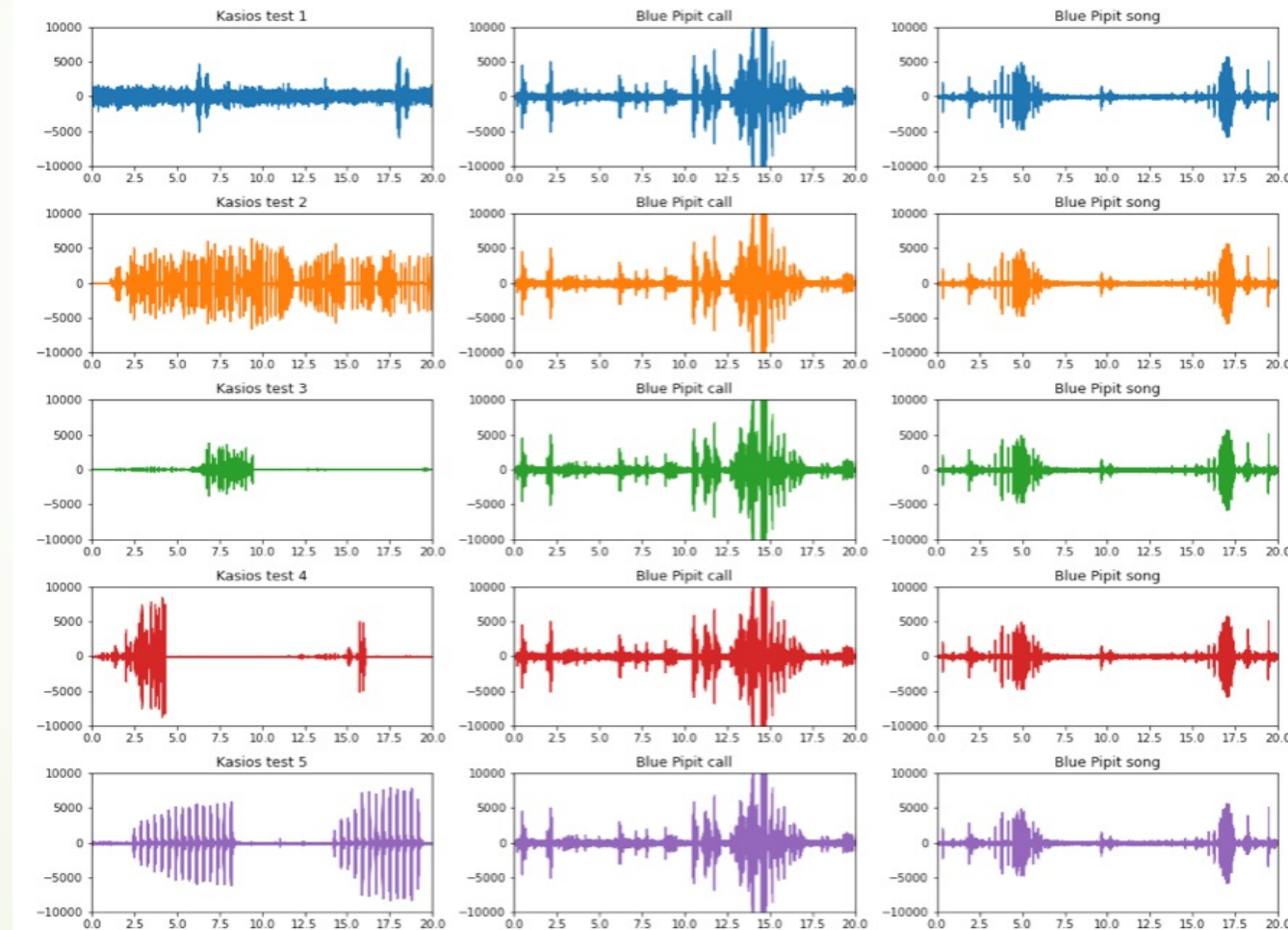
Mel Spectrogram visualizes the various sound frequencies and loudness in dB that make up what we hear.





# Visual Comparison, Samples, Kasios files to Known RCBP Call and Song

Kasios Recordings compared to Representative Blue Pipit sounds

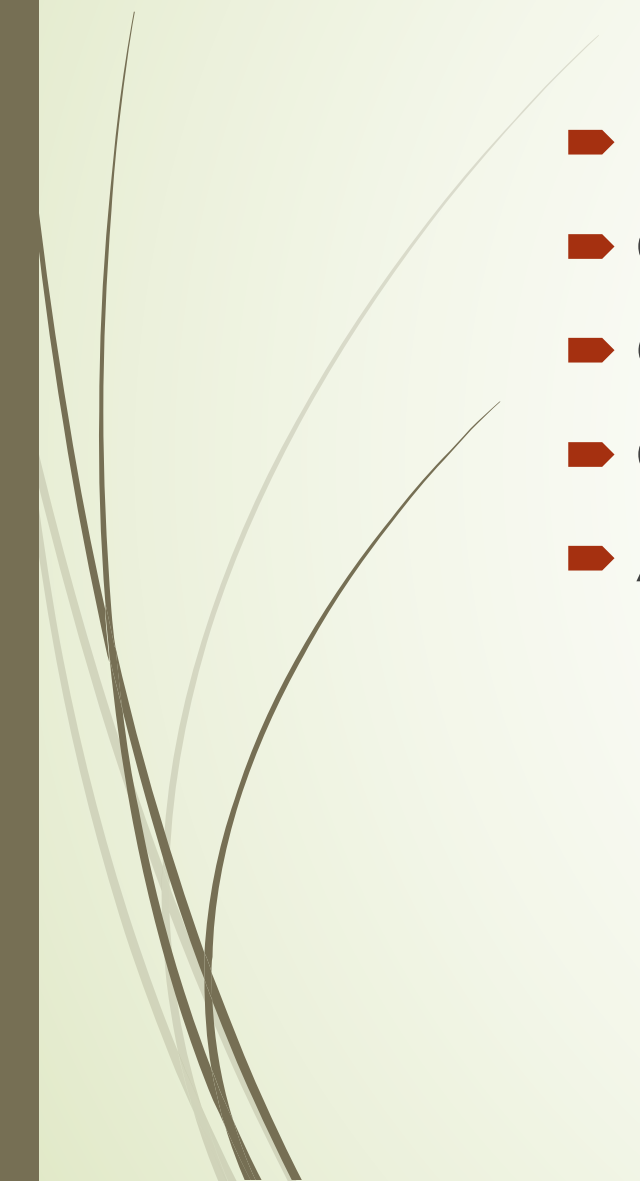


# Data Processing of the Recordings

- Librosa used for audio analysis
- Convert MP3 files to WAV files
- Verified bird calls and Kasios provided test recordings – All files were:
  - Normalized over time
  - Stripped of intervals of silence
  - Split into 2 second time chunks
- Primary Features Explored
  - Mel-frequency cepstral coefficients (MFCC)-- a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.
  - Chroma Short Time Fourier Transform (STFT)--Transform to determine the sinusoidal frequency and phase content of a short segment of sound with a pitch profile.
- Vectors created by above processes are summarized by mean and standard deviation and put into a two-dimensional array (2 x 50), 56,272 records in the verified samples and 476 in the Kasios recordings.

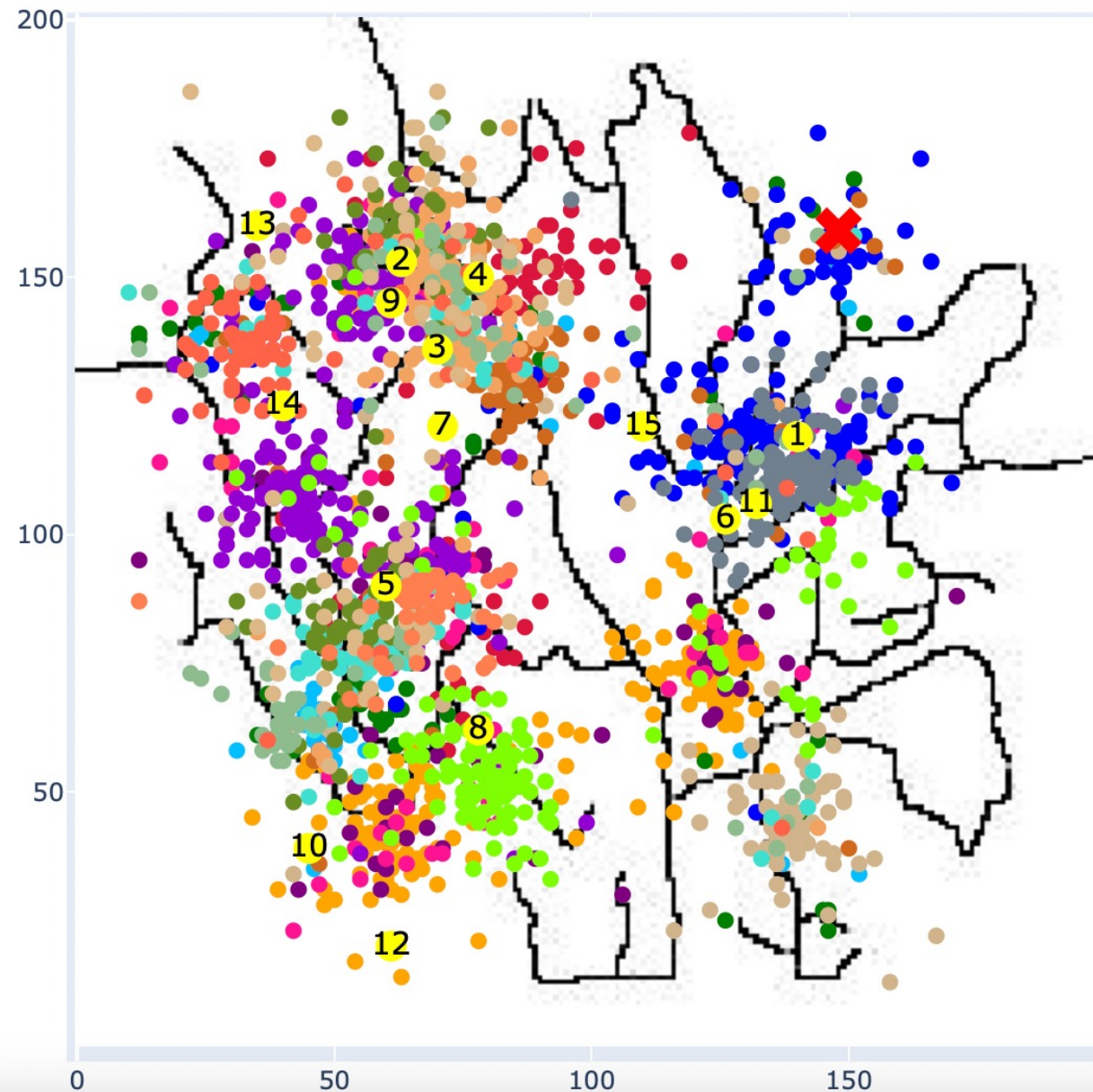


# Data Processing of Bird Data

- Ensured consistency (capitalizations, etc.) in all columns
  - Changed date to datetime object
  - Grouped times to morning, afternoon, evening, night
  - Created Season and Year and then Season/Year columns
  - Added colors for each of 19 bird types
- 

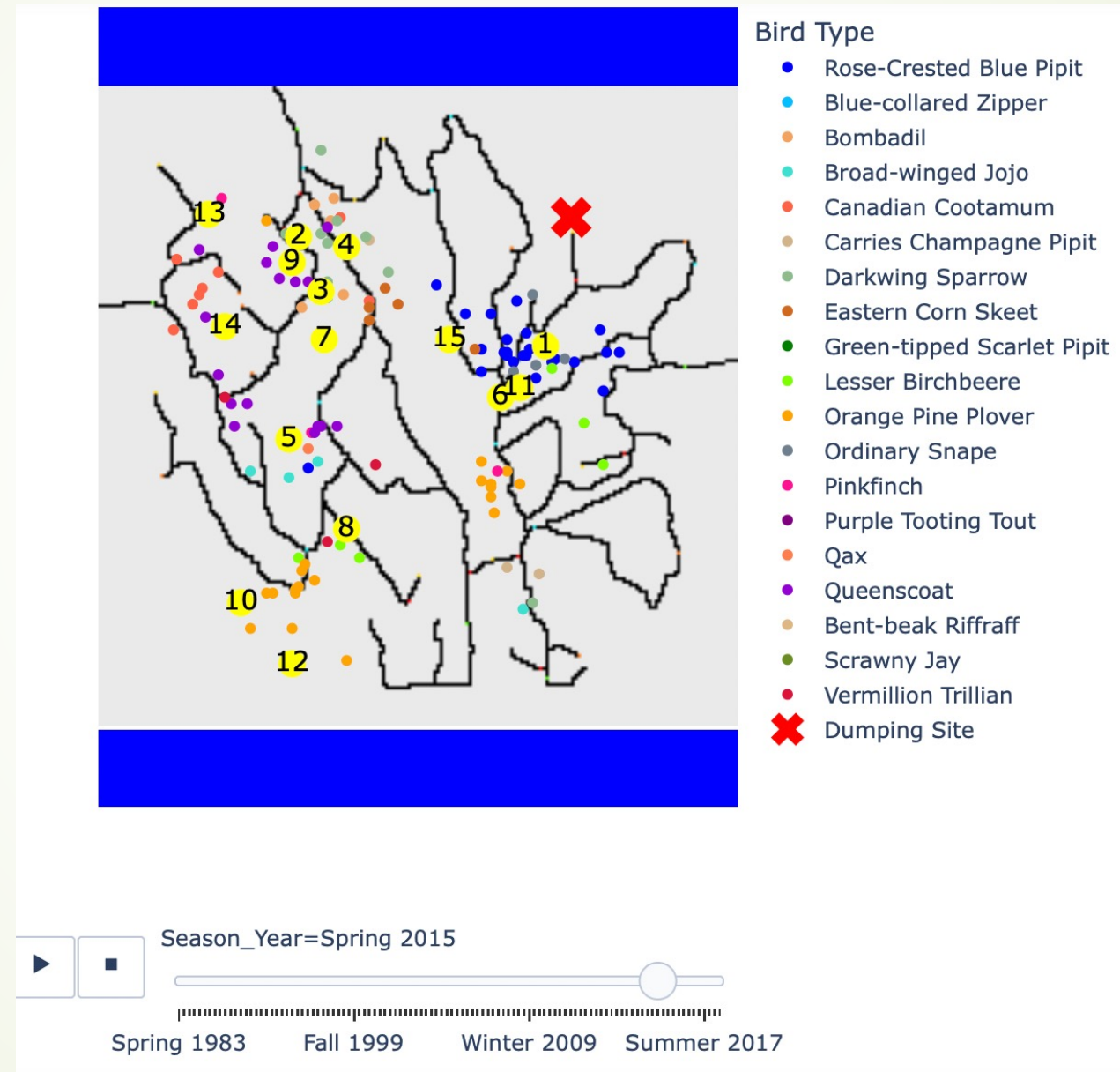


# All Birds in Mistford College Dataset



- Orange Pine Plover
- Green-tipped Scarlet Pipit
- Blue-collared Zipper
- Vermillion Trillian
- Purple Tooting Tout
- Rose-Crested Blue Pipit
- Carries Champagne Pipit
- Eastern Corn Skeet
- Pinkfinch
- Queenscoat
- Lesser Birchbeere
- Bombadil
- Broad-winged Jojo
- Ordinary Snape
- Scrawny Jay
- Darkwing Sparrow
- Canadian Cootamum
- Bent-beak Riffraff
- Qax
- ✕ Dumping Site

# All Bird Locations by Season/Year

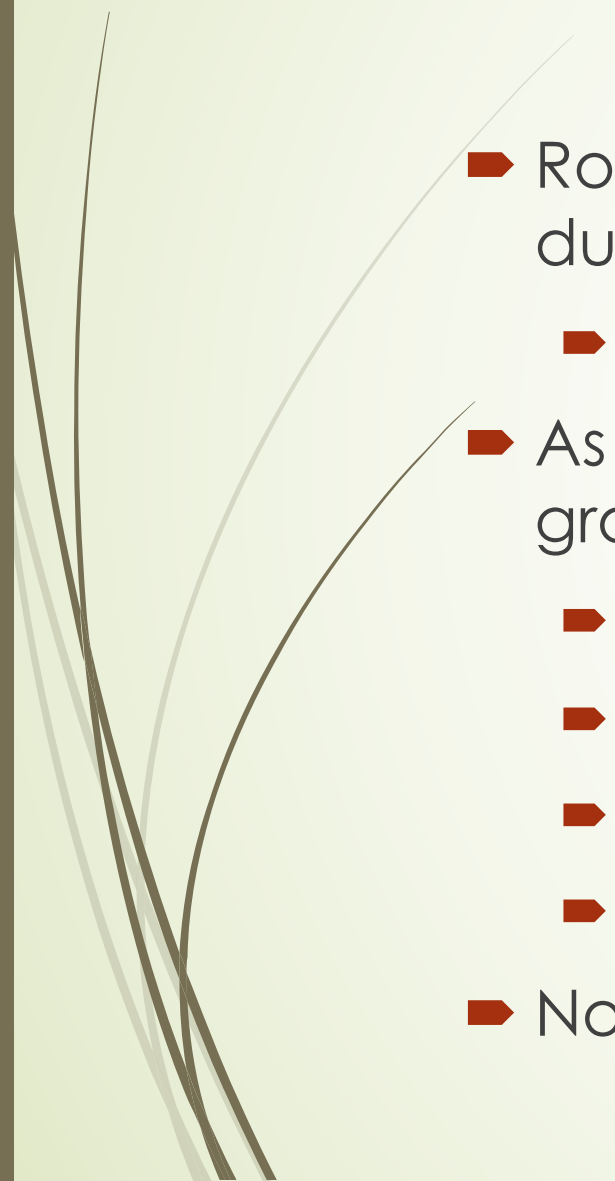


# Rose-Crested Blue Pipits by Season/Year





# Goal 1: Do you detect any trends or anomalies in the patterns?

- Rose-Crested Blue Pipits used to inhabit the area around the dumping site but have changed location.
    - Nearest last sighting of Rose-Crested Blue Pipit in Spring 2017
  - As of 2016, most birds have left the area around the dumping ground.
    - Darkwing Sparrow—Spring 2018
    - Jojo—Spring 2016 and Summer 2017
    - Eastern Corn Skeeet—Winter 2017
    - Carries Champagne Pipit—Fall 2017
  - None of the Kasios test bird recordings are near dumping site.
- 





# Models Used to Analyze Kasios' Test Data

- Binary Classification Models—full clips
  - OneVsRest Binary Classifier
  - MLP Binary Classifier with SMOTE for Imbalanced Data
  - Decision Tree
  - Random Forest
- Multi-Classification with 2 second chunks (19 bird classes)
  - MLP Classifier
  - Random Forest
  - Long Short-Term Memory Recurrent Neural Network LSTM-RNN



# Binary Results: MFCC with OneVsRest and MLP

- Imbalanced Data—used SMOTE and adjusted both over sampling the RCBP and under sampling all other birds
- MFCC dataset with means and medians on entire clips
- Dataset with 40 versus 100 columns/features
- Limited dataset to Quality levels of A, B, and C
- StandardScaler better than MinMax Scaler
- Best model included two layers [1024, 16] & sgd as solver
- Every model used on test data to show none of the birds in the test data was a Rose-Crested Blue Pipit

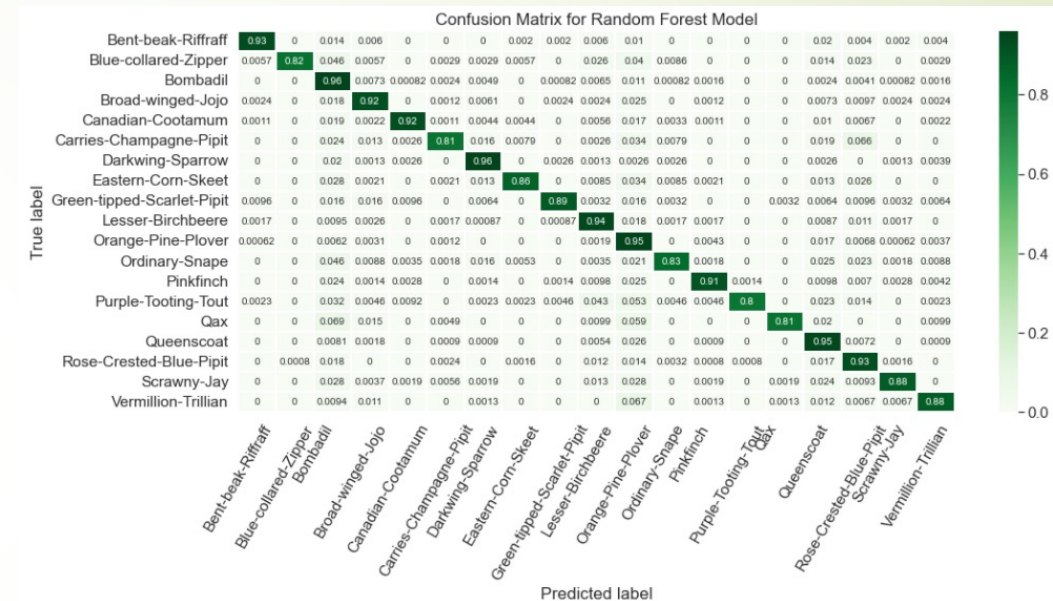
```
Test ROC_AUC: 0.9122890799108565
              precision  recall  f1-score  support
Rose-Crested Blue Pipit  0.62    0.56    0.59      36
              other      0.95    0.97    0.96     349
              accuracy              0.93     385
              macro avg      0.79    0.76    0.77     385
              weighted avg    0.92    0.93    0.93     385

Counter({'other': 353, 'Rose-Crested Blue Pipit': 32})
```

# Multiclass Results - Random Forest

- A, B, C, and D quality, 1000 estimator trees, Gini, on 20% Validation Set

	precision	recall	f1-score	support
Bent-beak-Riffraff	0.97	0.93	0.95	501
Blue-collared-Zipper	1.00	0.82	0.90	350
Bombadil	0.83	0.96	0.89	1225
Broad-winged-Jojo	0.93	0.92	0.92	826
Canadian-Cootamum	0.98	0.92	0.95	900
Carries-Champagne-Pipit	0.94	0.81	0.87	378
Darkwing-Sparrow	0.94	0.96	0.95	769
Eastern-Corn-Skeet	0.96	0.86	0.91	470
Green-tipped-Scarlet-Pipit	0.97	0.89	0.93	312
Lesser-Birchbeere	0.92	0.94	0.93	1152
Orange-Pine-Plover	0.84	0.95	0.89	1617
Ordinary-Snape	0.95	0.83	0.89	568
Pinkfinch	0.97	0.91	0.94	713
Purple-Toothing-Tout	0.99	0.80	0.89	437
Qax	0.98	0.81	0.89	203
Queenscoat	0.86	0.95	0.90	1108
Rose-Crested-Blue-Pipit	0.90	0.93	0.91	1254
Scrawny-Jay	0.96	0.88	0.92	539
Vermillion-Trillian	0.95	0.88	0.92	747
accuracy			0.91	14069
macro avg	0.94	0.89	0.91	14069
weighted avg	0.92	0.91	0.91	14069

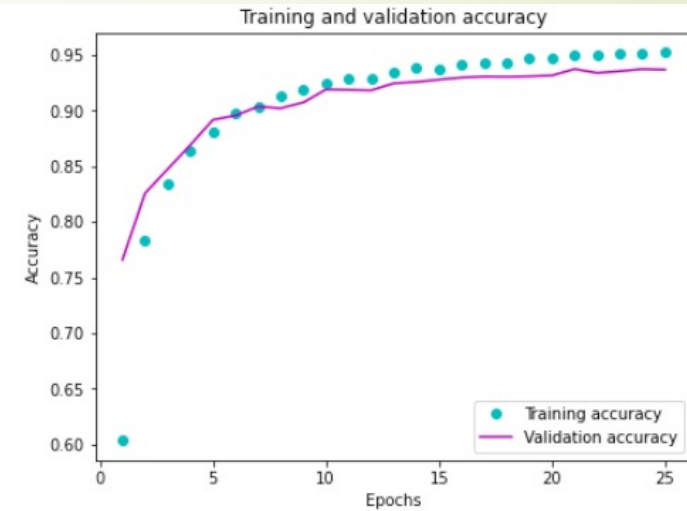
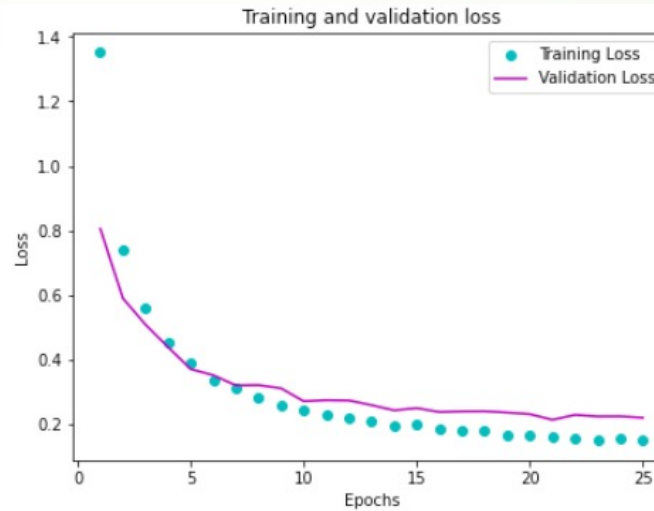


- F1 score of 0.91 to predict the Rose-crested Blue Pipit
- Predicted that 4 of the 15 Kasios files are Blue Pipits

# Final Model Results - LSTM RNN

- optimizer='adam', loss='SparseCategoricalCrossentropy', metrics=['acc'], 25 epochs, 20% validation set

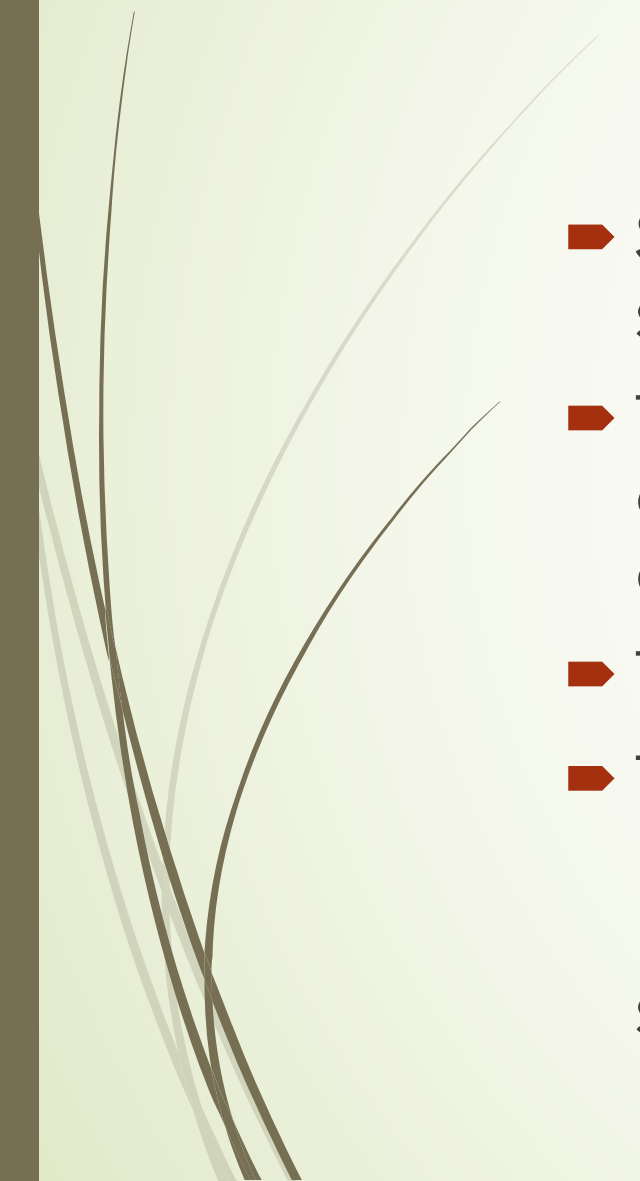
	precision	recall	f1-score	support
Bent-beak-Riffraff	0.95	0.97	0.96	501
Blue-collared-Zipper	0.85	0.91	0.88	350
Bombadil	0.93	0.95	0.94	1225
Broad-winged-Jojo	0.89	0.94	0.91	826
Canadian-Cootamum	0.99	0.95	0.97	900
Carries-Champagne-Pipit	0.92	0.86	0.89	378
Darkwing-Sparrow	0.93	0.96	0.94	769
Eastern-Corn-Skeet	0.89	0.94	0.92	470
Green-tipped-Scarlet-Pipit	0.90	0.93	0.92	312
Lesser-Birchbeere	0.97	0.96	0.97	1152
Orange-Pine-Plover	0.96	0.94	0.95	1617
Ordinary-Snape	0.97	0.91	0.94	568
Pinkfinch	0.93	0.96	0.94	713
Purple-Tooting-Tout	0.93	0.92	0.93	437
Qax	0.83	0.94	0.88	203
Queenscoat	0.95	0.93	0.94	1108
Rose-Crested-Blue-Pipit	0.94	0.94	0.94	1254
Scrawny-Jay	0.91	0.92	0.92	539
Vermillion-Trillian	0.97	0.91	0.94	747
accuracy			0.94	14069
macro avg	0.93	0.93	0.93	14069
weighted avg	0.94	0.94	0.94	14069



- F1 score of 0.94 on to predict the Rose-crested Blue Pipit
- Predicted that 3 of the 15 Kasios files are Blue Pipits



## **Goal 2: What is the role of visualization in your analysis of the Kasios bird calls?**

- Seeing gaps on the visualizations led to the elimination of silence.
  - The maps showed where different types of birds congregate and illustrated the movement of those centers after dumping began.
  - The test birds are no where near the dumping site.
  - The results from the classification indicate that sites 2 and 9 (and maybe 13 and 14) are most likely Rose-Crested Blue Pipits, which show even further migration from the dumping site.
- 



## Goal 2: Does this set support the claim of Pipits being found across the Preserve?

- The binary classification MLP showed 62% of the predicted Rose-Crested Blue Pipits in the Mistford College data are actually Rose-Crested Blue Pipits, while 97% of the other birds are predicted correctly, but predicted none of the Kasios birds were Rose-Crested Blue Pipits.
- The Random Forest model predicted the Rose-Crested Blue Pipit 91% of the time, and 4 of the Kasios' birds were RCBP.
- The best model (LSTM-RNN) accurately predicted Rose-Crested Blue Pipits 94% of the time and correctly identifying other bird species better than 88% of the time.
- In summary, at most 3-4 of the bird recordings from the Kasios test data are Rose-Crested Blue Pipits based on the analysis.





### **Goal 3: Formulate a hypotheses concerning the state of the Rose Crested Blue Pipit. What are your primary pieces of evidence to support your assertion?**

Hypothesis: The Rose-Crested Blue Pipit population is decreasing as a result of migration away from a dumping site and possibly polluted air.

- The map shows Rose-Crested Blue Pipits and other birds have moved away from the dumping site and inhabit other areas of the Preserve.
- Data analysis of sound recordings predicts less than 5 of the Kasios provided recordings are of the Blue Pipit. Kasios is either unaware or trying to deceive with the test data of 15 bird recordings.
- None of the Kasios provided recordings were made near the original dumping site.



### **Goal 3 (Actionable Insights): What next steps should be taken in the investigation to either support or refute the Kasios claim that the Pipits are actually thriving across the Boonsong Lekagul Wildlife Preserve?**

- Continue to record all bird locations in the preserve, map future changes in the populations and locations, and investigate if more contamination has occurred in other areas of the preserve
- Analyze pollutants in the air and in the water in representative locations in the preserve. Determine if pollutants still exist.
- Pursue methods to clean the preserve, if pollutants do exist.
- Determine if satellite imagery of the preserve over time exists. Changes in air quality and ground vegetation may also be affecting the wildlife.
- Explore the disruption to other birds species (competition for territory) caused by the migration of the Rose-Crested Blue Pipit.



# Key Insights and Data Science Conclusions

- Librosa is an extensive package of sound analysis tools, that processes data conducive to time series analysis, and takes a great deal of time to master with a steep learning curve.
  - The machine learning and model tuning techniques explored required extreme hardware resources. Kernels blew up regularly.
  - Neural networks proved to be effective in sound analysis.
  - Check out the streamlit version here: <https://mystery-at-the-preserve.herokuapp.com>
  - This can also be accessed on GitHub: <https://github.com/LeighAnn-DU/Mystery-at-the-Preserve>
- 