

ECHOES OF ENDANGERMENT:

A Visual and Machine Learning Analysis of the Rose-Crested Blue Pipit's Decline in Boonsong Lekagul Wildlife Preserve

Rowan University [NOV 2022]
Information Visualization
Course Project

VAST CHALLENGE (2018 MC 1)

Group Project by :

Kamal M Shaik
Batuhan Kir
Jun Yan Chen
Shravika Pinna
Chinmai Reddy Modugula

TOOLS AND TECHNIQUES

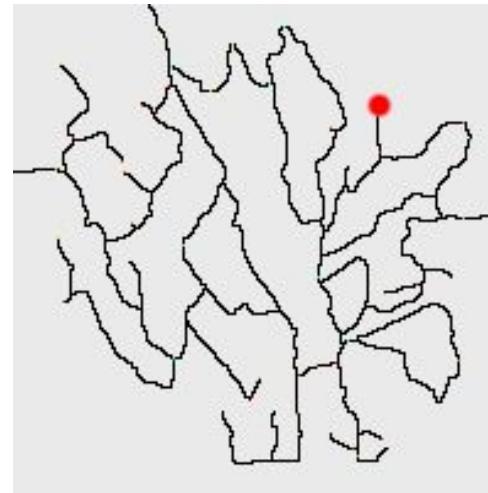
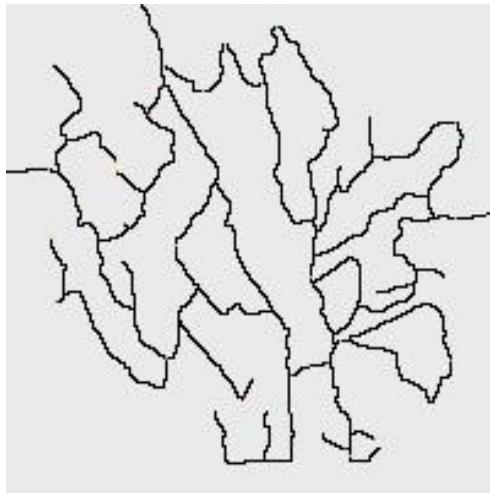
- **OpenRefine:** A powerful tool for cleaning and transforming bird call data to ensure consistency before analysis.
- **Data Analysis:** Used to explore patterns in bird call recordings and detect trends or anomalies across the preserve.
- **Data Validation:** Applied to ensure the accuracy and integrity of the bird call datasets and metadata provided by Mistford College and Kasios.
- **Machine Learning:** Leveraged to classify bird species from audio data and detect misclassifications in the Pipit recordings claimed by Kasios.
- **Statistical Analysis:** Used to quantify population trends and movement patterns of the Rose-Crested Blue Pipit and other species.
- **Audacity:** Employed for audio analysis and editing of bird call recordings to filter noise and enhance call quality for further classification.
- **Tableau:** Utilized for creating dynamic visualizations that showcase trends and geographic distribution of bird species over time.
- **Python & Libraries:** Python, along with libraries such as Pandas, Scikit-learn, and Matplotlib, was used for data manipulation, machine learning models, and generating analytical visuals.
- **Jupyter Notebook:** An interactive platform to combine code, data visualization, and narrative text, facilitating real-time analysis, experimentation, and documentation of the bird call classification and geospatial trends.
- **GeoPandas:** Used within Python for geospatial analysis to map bird call locations across the preserve and analyze migration or habitat shifts.
- **Librosa:** A Python library used for analyzing and classifying bird calls from audio recordings to identify discrepancies in species identification.

Project Overview



- Mistford is a mid-size city located to the southwest of the Boonsong Lekagul Wildlife Preserve.
- The city has a small industrial area with four light-manufacturing endeavors.
- Mistford and the wildlife preserve are struggling with the possible endangerment of the **Rose-Crested Blue Pipit**, a locally loved bird.
- The bird's nesting pairs seem to have decreased alarmingly, prompting an investigation last year implicating **Kasios Office Furniture**, a Mistford manufacturing firm.
- Since the initial investigation, the situation has evolved: **Kasios** insists that they have done nothing wrong!
- They assert that grad student **Mitch Vogel** and his professors are mere media-seekers trying to draw attention away from their lackadaisical research.
- **Kasios** presents itself as an extremely eco-friendly organization. They have launched their own very public **investigation** into the issues raised last year and are reporting very different results!

Overview of Lekagul Roadways Location & Mapping Dumping site



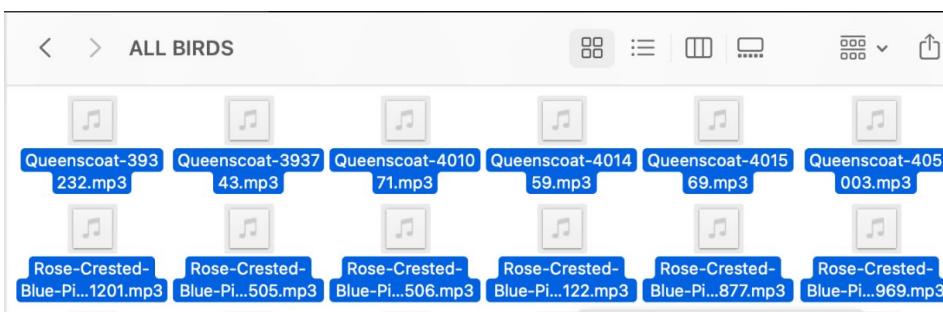
- “Lekagul Roadways 2018” is a 200 x 200 pixel map of the Preserve, with general indications of roadways through the site.
- The alleged dumping site for the Kasios waste products was centered around coordinates (148,159).
- To illustrate this Dumping site we used a website which helped us mark the location with coordinates.
- Website Reference: https://www.mobilefish.com/services/record_mouse_coordinates/record_mouse_coordinates.php

Mitch's Dataset & Files

AllBirdsv4.csv

A	B	C	D	E	F	G	H
1	File ID	English_name	Vocalization_type	Quality	Time	Date	X Y
2	402254	Rose-crested Blue Pipit	call	no score	13:30	2/8/18	49 63
3	406171	Rose-crested Blue Pipit	call	A	7:48	6/7/17	125 133
4	405901	Rose-crested Blue Pipit	call	A	12:00	2/8/18	58 76
5	405548	Rose-crested Blue Pipit	song	A	11:00	3/10/18	55 125
6	401782	Rose-crested Blue Pipit	song	A	6:00	6/29/08	129 123
7	401720	Rose-crested Blue Pipit	call	A	13:00	12/28/16	132 121
8	401719	Rose-crested Blue Pipit	call	A	13:00	12/28/16	106 107
9	387404	Rose-crested Blue Pipit	call	A	10:00	9/18/17	149 115
10	377874	Rose-crested Blue Pipit	song	A	7:43	6/11/00	134 118

All Birds Audio Folder



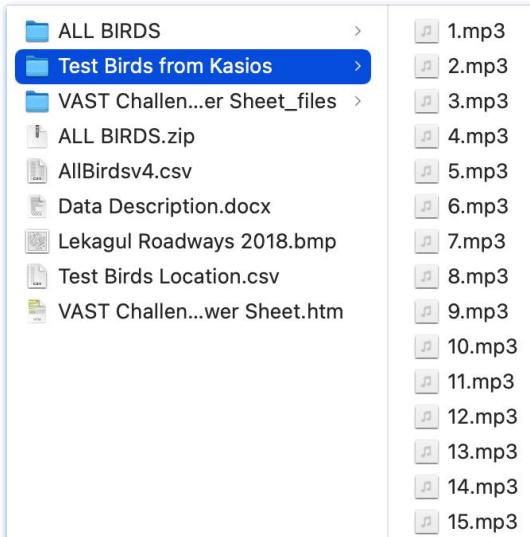
- 1.“AllBirdsv4.csv” is the metadata file for the bird sounds files.
- 2.This dataset containing the birdcalls collected by Mistford College.
- 3.The File ID is the index to the file names in the ALL BIRDS audio file collection.
- 4.The English_name is the common English name for the particular bird.
- 5.The Vocalization_type is the kind of bird sound.
- 6.Quality is a score of recorded audio A, B, C, D, or E.
- 7.Time and Date are for the capture of the sound.
- 8.Finally, X and Y are the coordinates on the enclosed map of where the sound was recorded.

Kasios Dataset & Files

Test Bird Locations.csv

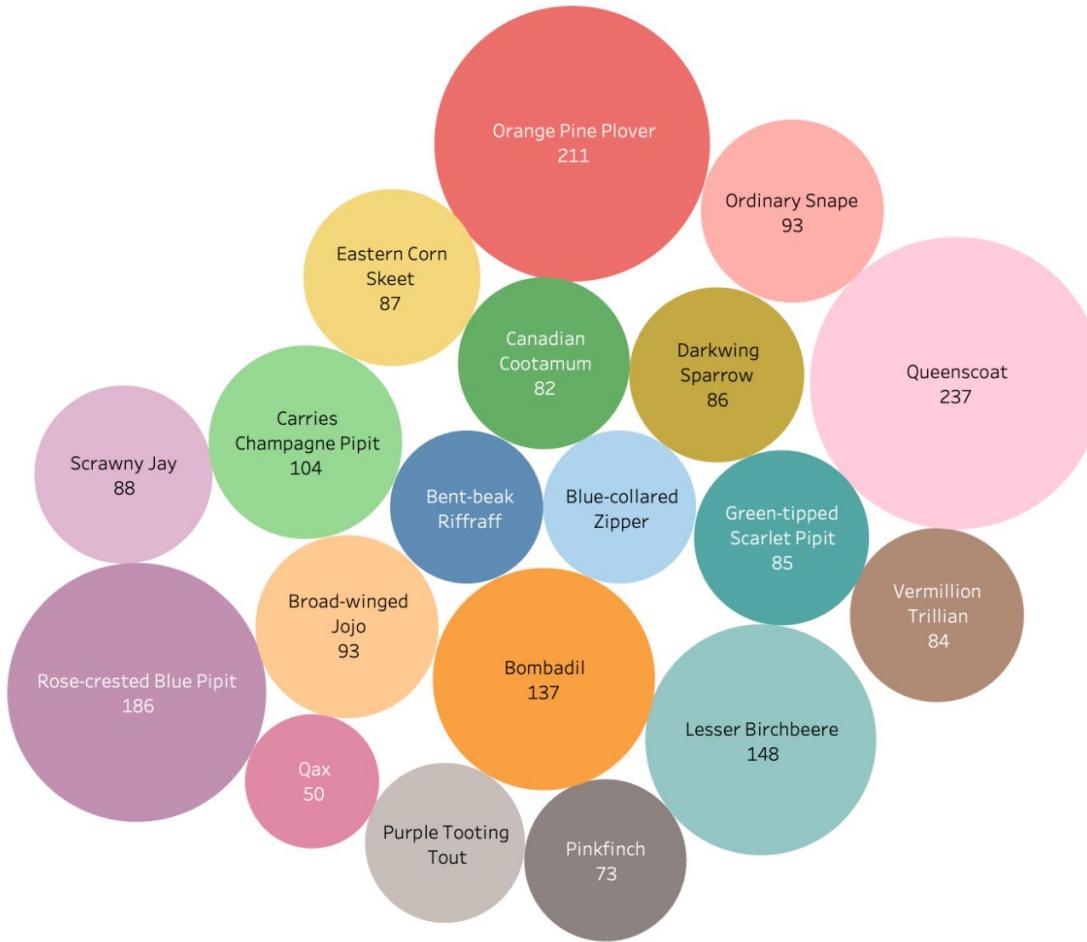
	A	B	C
1	ID	X	Y
2	1	140	119
3	2	63	153
4	3	70	136
5	4	78	150
6	5	60	90
7	6	126	103
8	7	71	121
9	8	78	62
10	9	61	145
11	10	45	39
12	11	132	106
13	12	61	20
14	13	35	160
15	14	40	125
16	15	110	121

Test Birds Audio-Kasios Folder



- “Test Birds from Kasios” are the bird sounds Kasios claims as Pipits from across the Preserve.
- All of these recording were taken over the past couple of months.
- “Test Bird Locations” indicate where in the pixel map the bird sounds were recorded.

Visualising Overall Bird count by each species





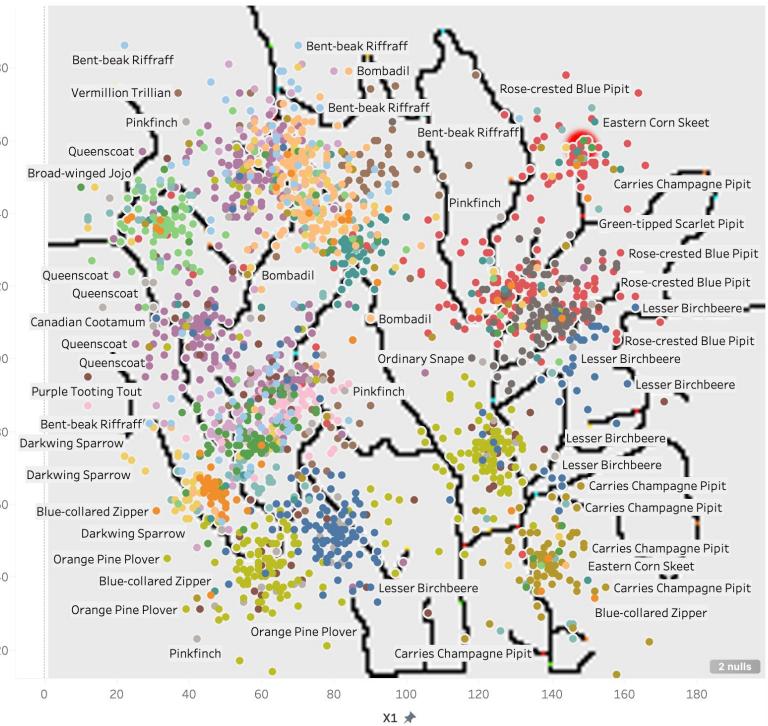
Questions

1. Do we detect any trends or anomalies in the patterns?
2. Does Kasios set support the claim of Pipits being found across the Preserve?

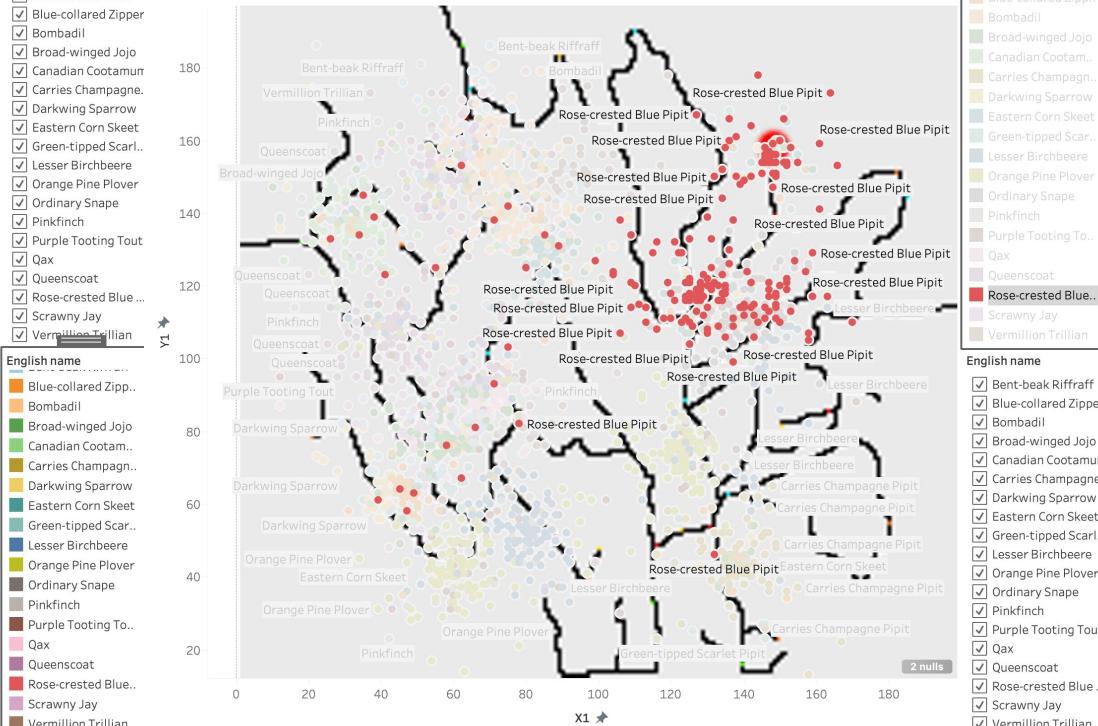
Mitch's Data Distribution of All birds vs Blue pipits



Distribution of All Birds

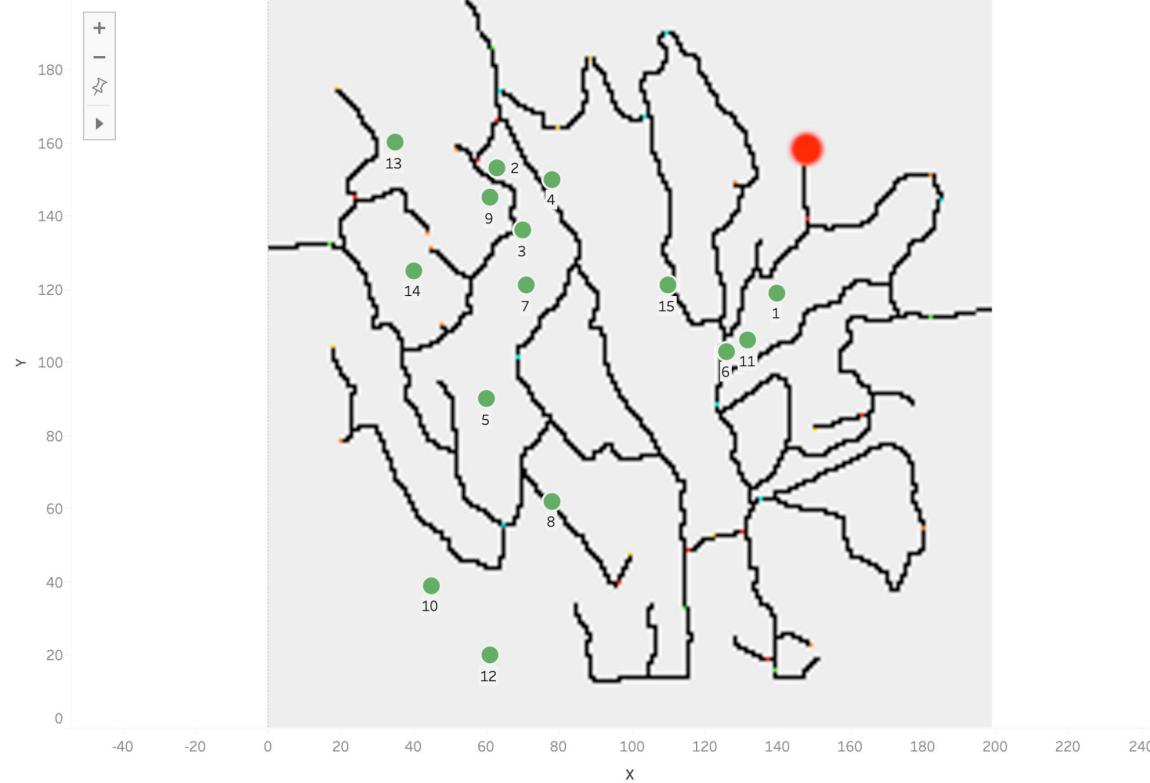


Distribution of All Birds (w/ pipits filter)



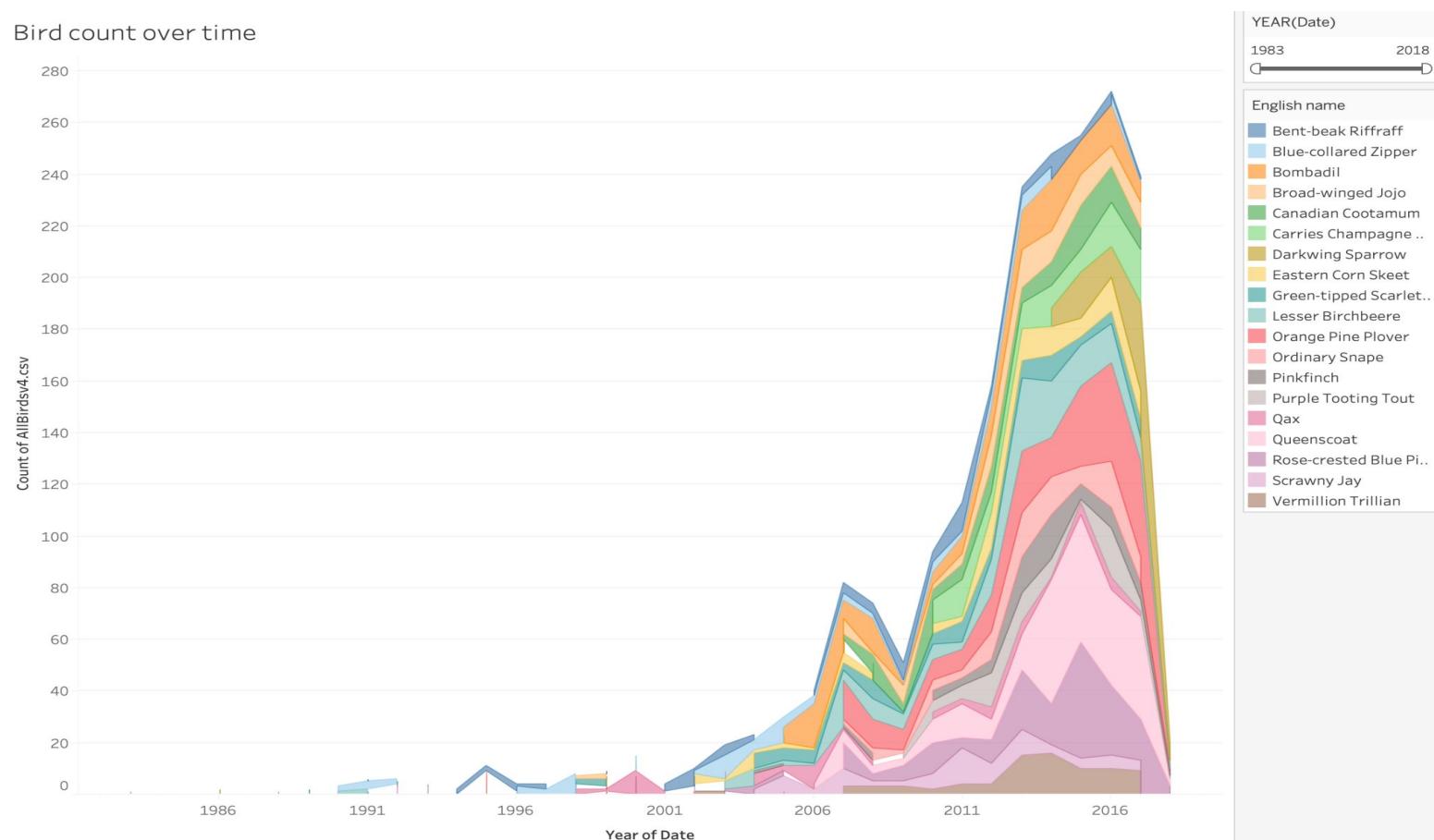
Kasios Data Distribution of Blue Pipits

Kasios Recordings Datapoints



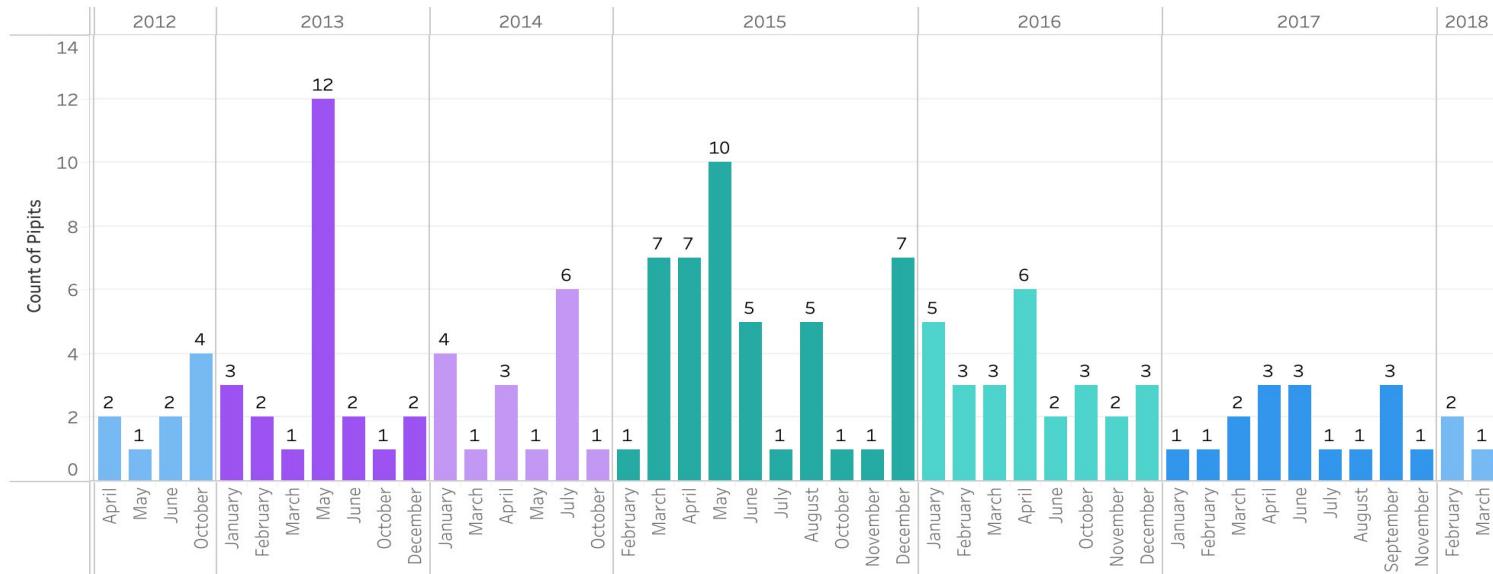
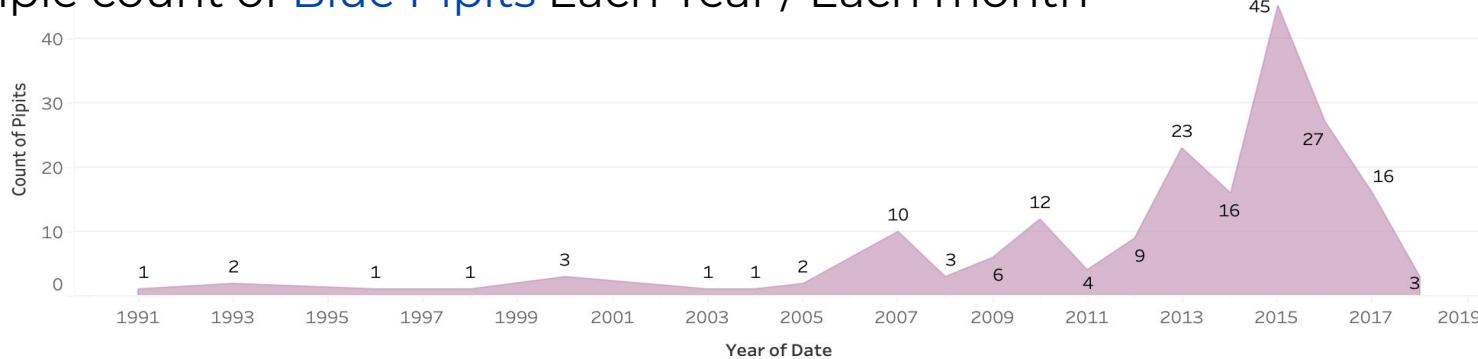
Pattern 1

Bird count over time

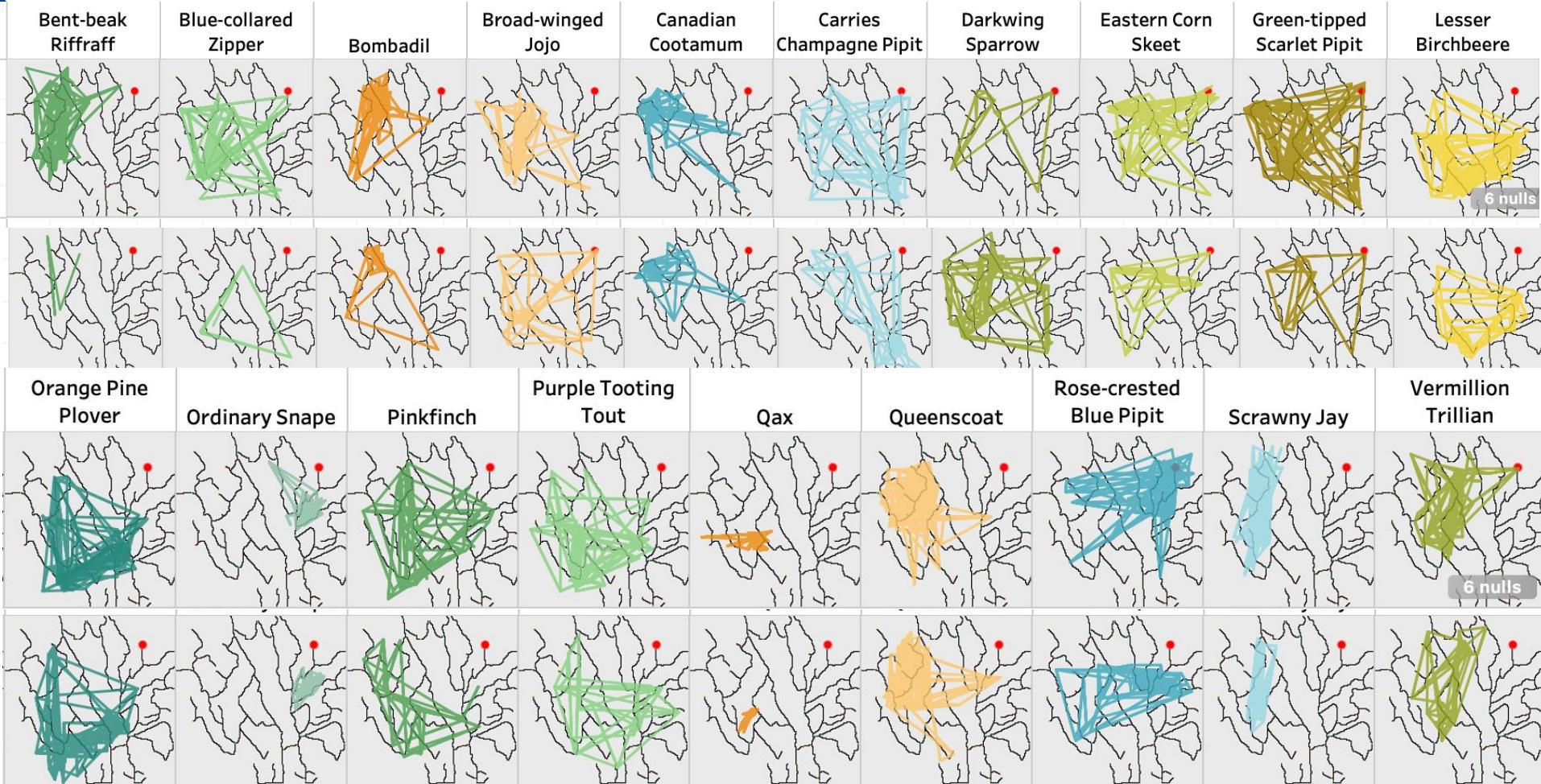


Pattern 2 :

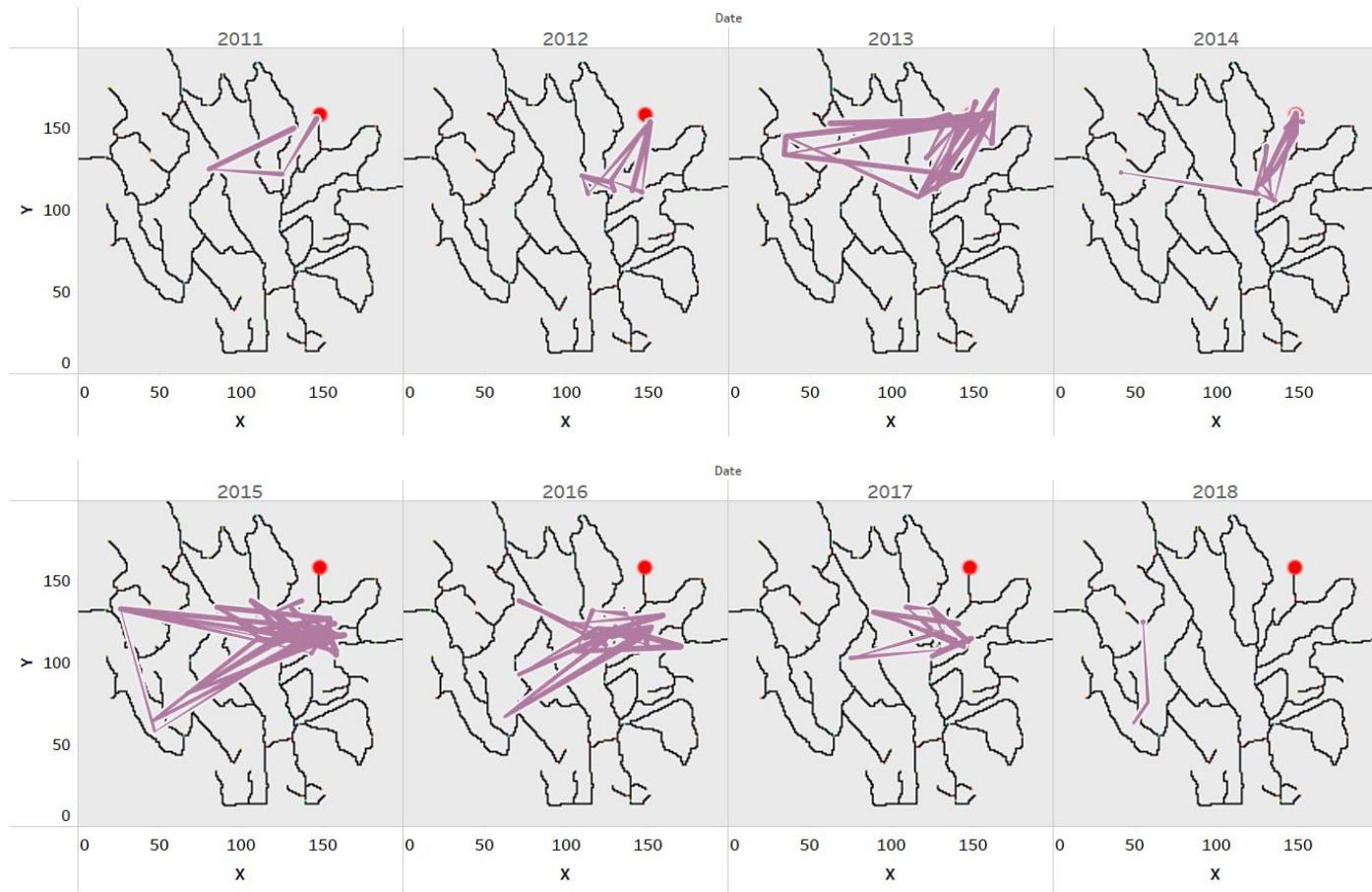
Sample count of Blue Pipits Each Year / Each month



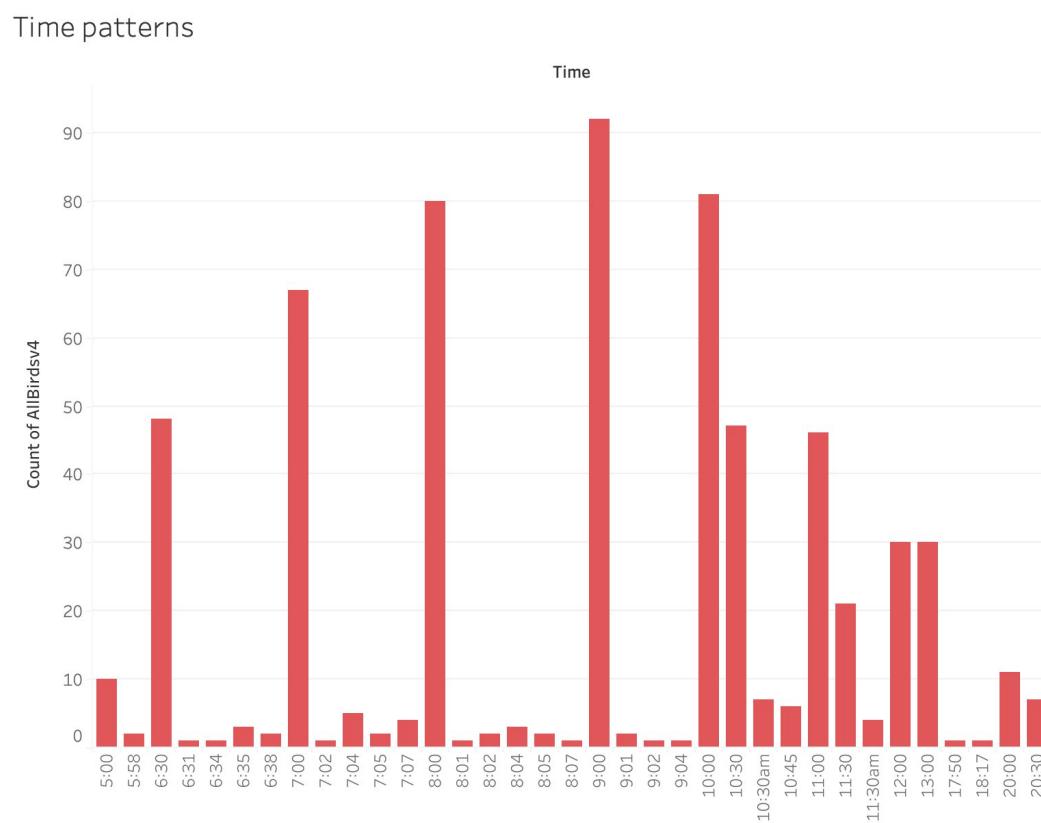
Path of All Birds before 2015 & After 2015



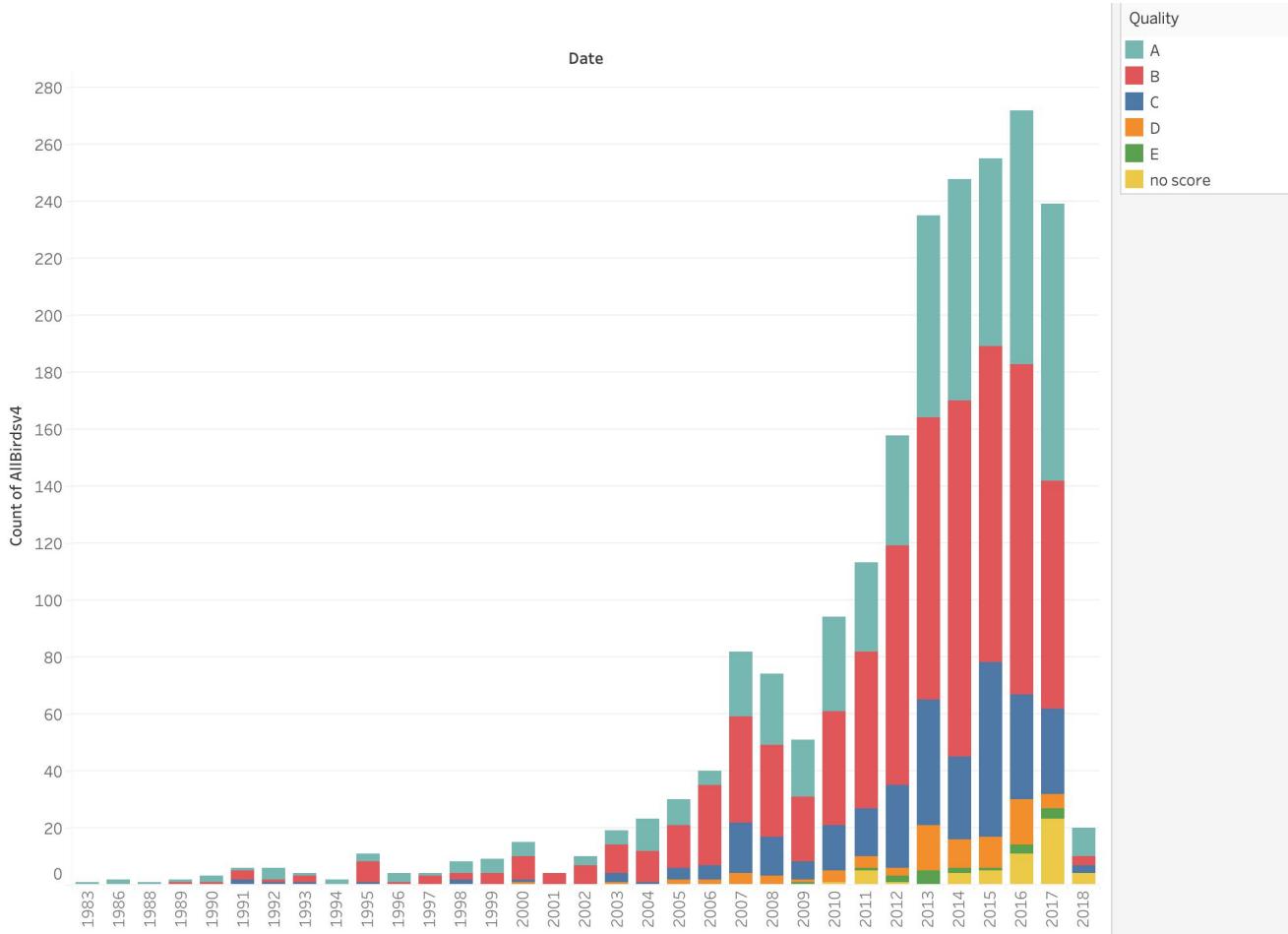
Pattern 3 : Path of Blue Pipits each year



Pattern 4 : Recordings Time Pattern

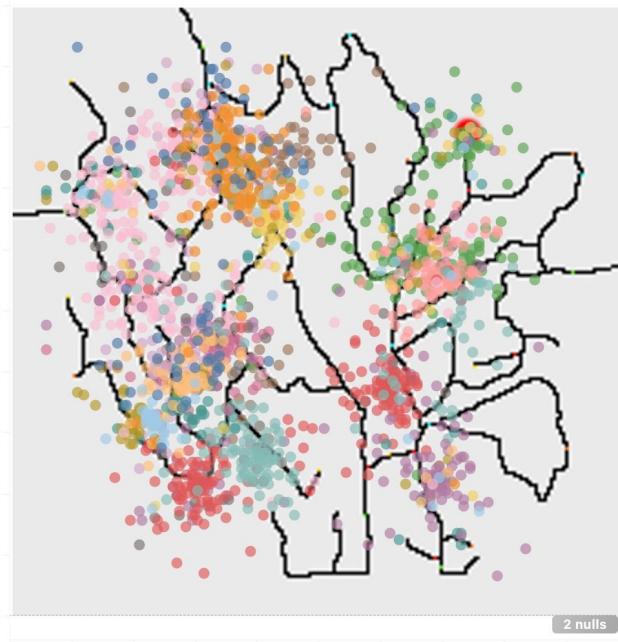


Pattern 5 : Recordings Quality Pattern

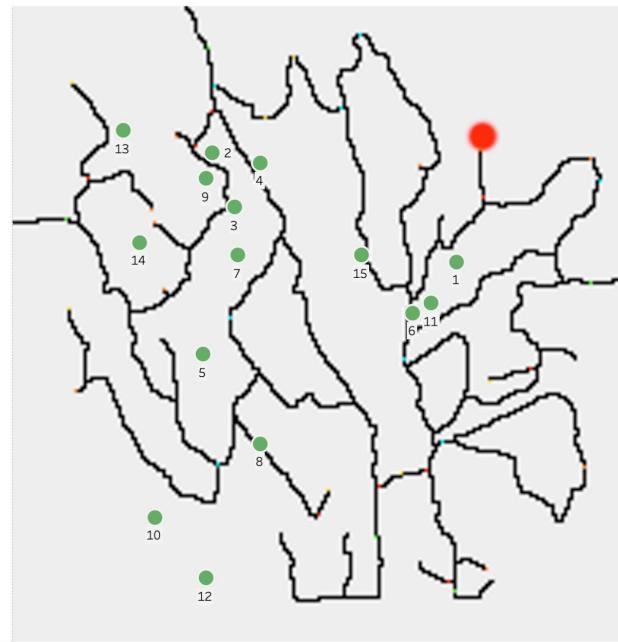


Q2. Kasios vs Mitch Data Patterns plotting Blue pipits

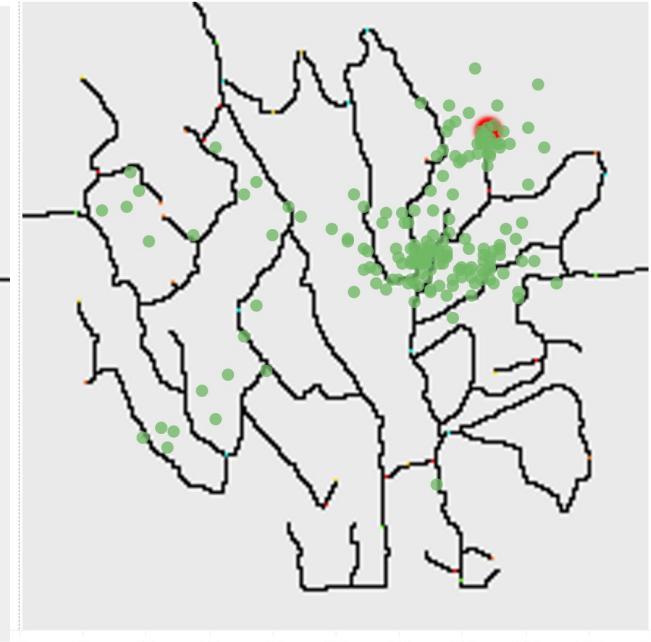
All Bird's (Mitch's Data)



Blue Pipit's (Kasios Data)



Blue Pipits (Mitch's Data)





VAST CHALLENGE

2018 MC 1

Second Progress

Group Project by :

Batuhan Kir
Jun Yan Chen
Shravika Pinna
Chinmai Reddy Modugula
Kamal Afridi M S



Contributions

- **Batuhan Kir** - Data Preprocessing via OpenRefine & Audacity(Slides-3, 5,6)
- **Jun Yan Chen** - Audacity & Machine Learning Model(Slides-5,6,7,8,10)
- **Shravika Pinna** - Machine Learning Model & Predictions(Slides-12,13,14,15)
- **Chinmai Reddy M** - Machine Learning Model & Predictions(Slides-12,13,14,15)
- **Kamal Afridi M S** - Audio Formatting & Machine Learning Predictions(Slides-3,14,15)

Data Cleaning & Normalization (OpenRefine)

- Removing inconsistent data for “Vocalization Type” and “Quality”

220	122050	Blue-colla	0 call	B
229	116386	Blue-colla	0 call	B
230	95988	Blue-colla	0 song	B
231	85427	Blue-colla	0 call	B
232	74792	Blue-colla	0 call	B
233	74191	Blue-colla	0 song	B
234	52081	Blue-colla	0 song	B
235	44224	Blue-colla	0 ?	B
236	41940	Blue-colla	0 song	B
237	24011	Blue-colla	0 song	B
238	13297	Blue-colla	0 call	B
...				

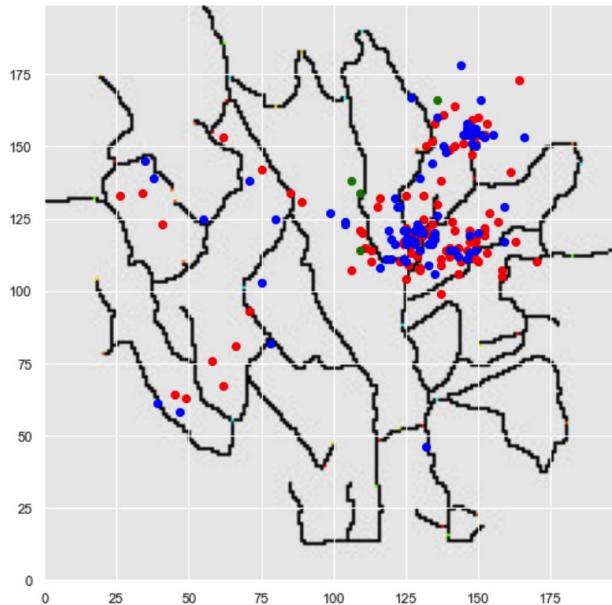


229	95988	Blue-colla	0 song	B
230	85427	Blue-colla	0 call	B
231	74792	Blue-colla	0 call	B
232	74191	Blue-colla	0 song	B
233	52081	Blue-colla	0 song	B
234	41940	Blue-colla	0 song	B
235	24011	Blue-colla	0 song	B
236	13297	Blue-colla	0 call	B
237	326954	Blue-colla	0 call	C
238	251567	Blue-colla	0 call	C

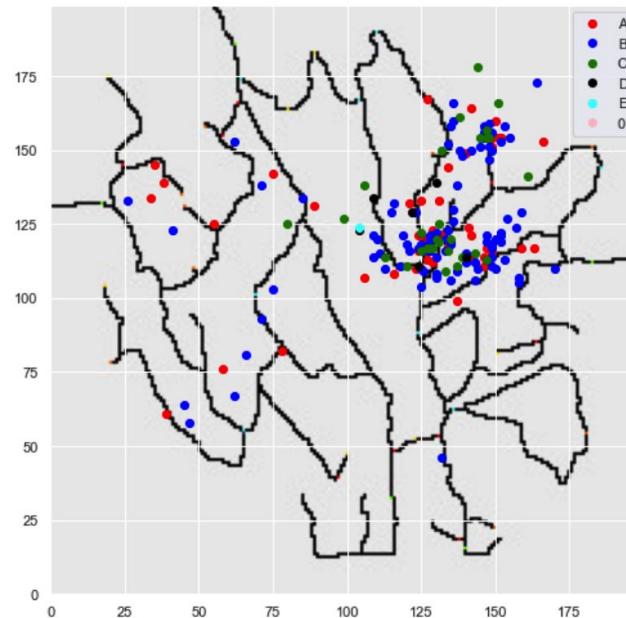
- Created a new column label by using ‘English Name’ column using GREL Expression.
- `if(length(value.find(/Rose Crested Blue Pipit/))>0,1,0)`

Geo Locations

- Plot 1: Mapping Mitch's Data WRT Vocalization type
- Plot 2: Mapping Mitch's Data WRT Audio Quality



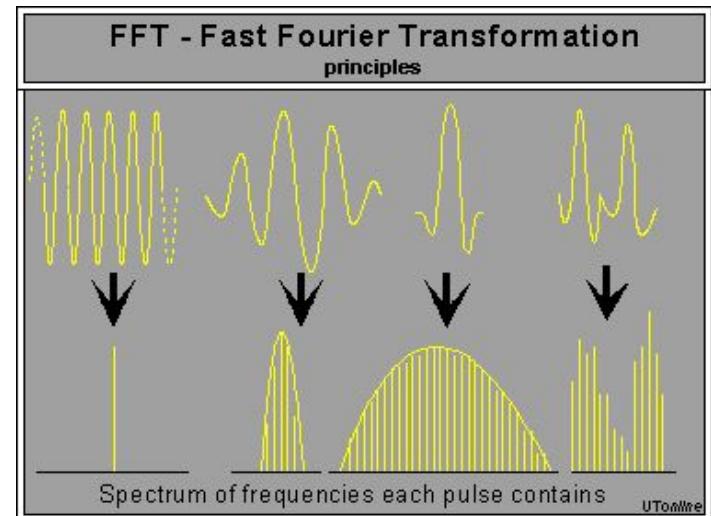
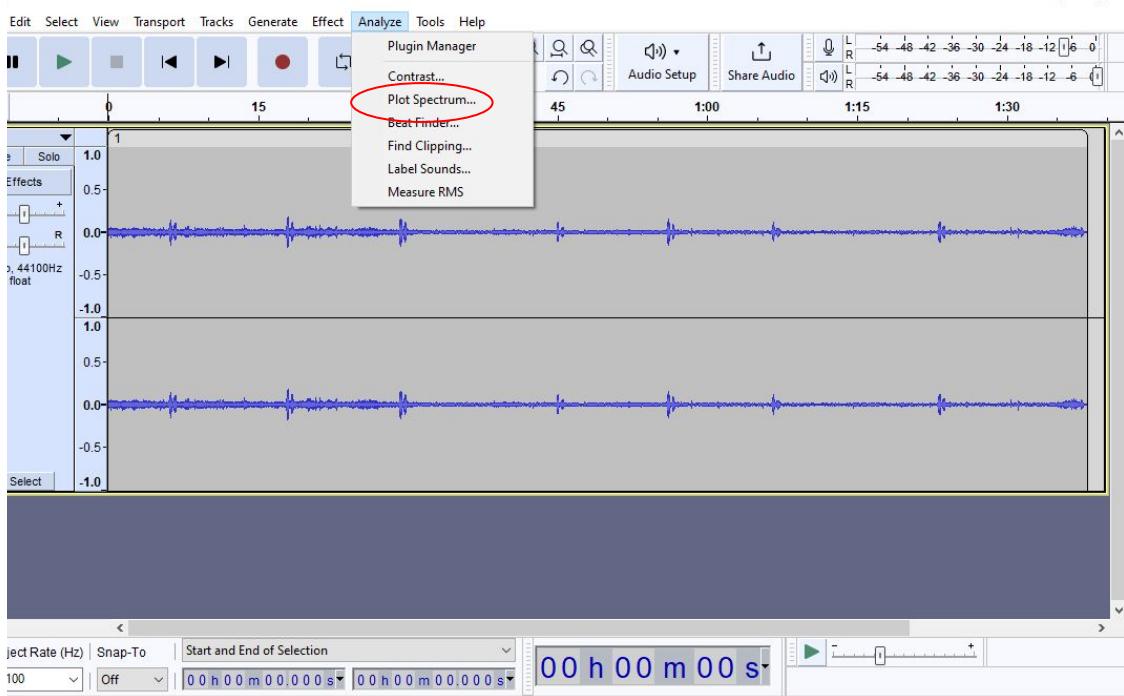
Red - Call
Blue - Song
Green - Call, Song



A B C D E - Quality

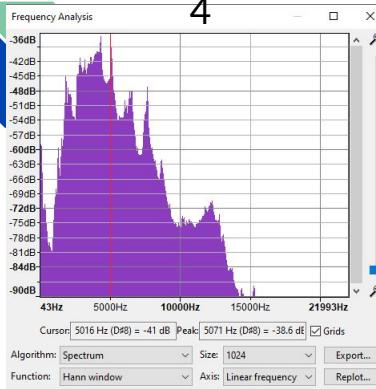
Spectrum Analysis (Audacity)

- **Audacity** - free open source audio software
- **Plot Spectrum** separates audio into blocks, and uses Fast **Fourier Transform** or **FFT** to convert waveform to spectrum, and then averages all the blocks together.
- **Spectrum** - The sound intensity/amplitude(dB) at each frequency(Hz).

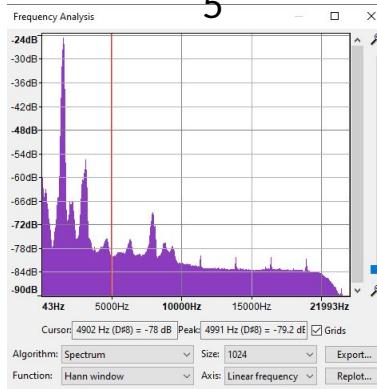


Eyeball Observation

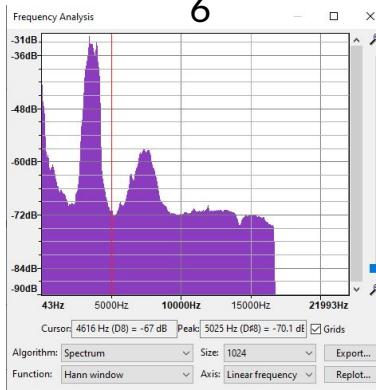
4



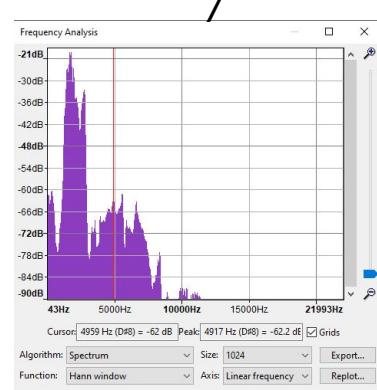
5



6



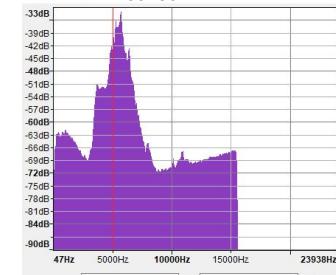
7



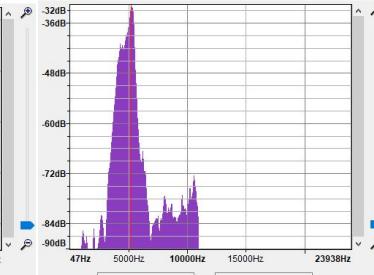
Kasios

299736

Frequency Analysis 299736

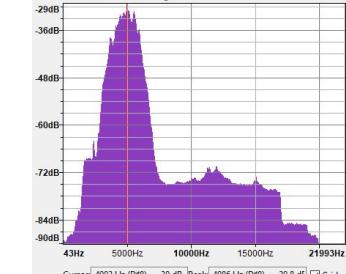


Frequency Analysis 351273

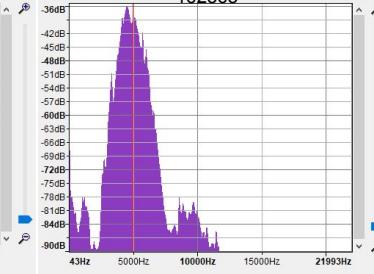


14877

Frequency Analysis 14877

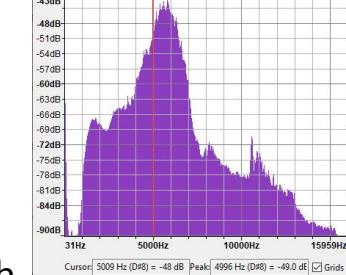


Frequency Analysis 162563

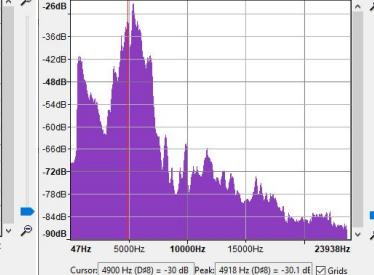


298739

Frequency Analysis 298739

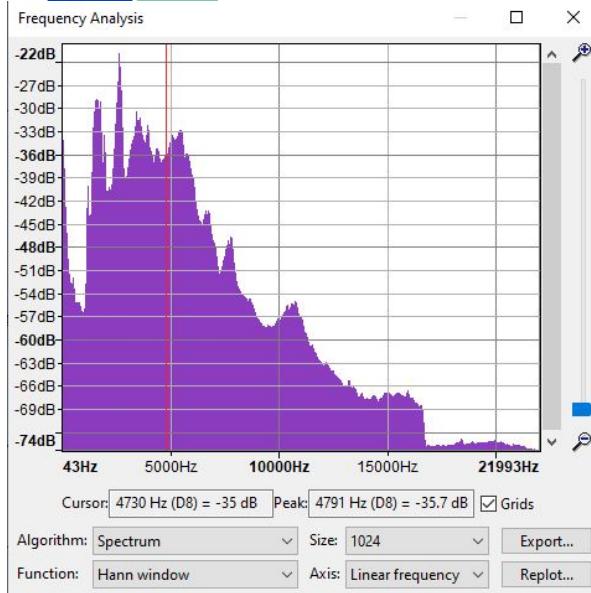


Frequency Analysis 110765



Mitch

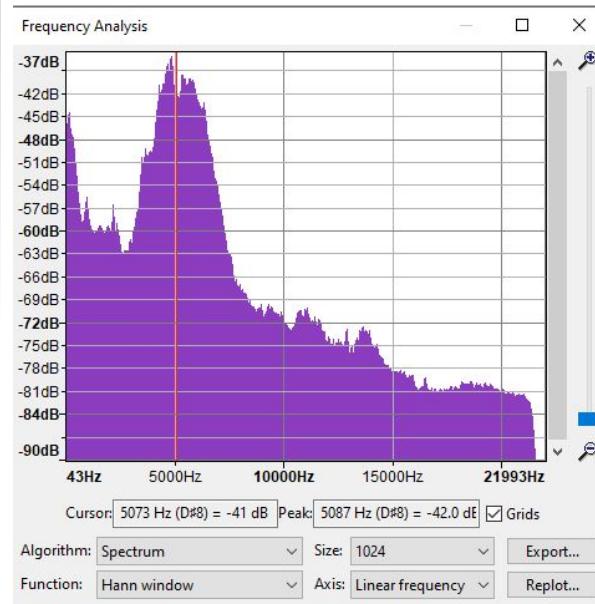
Kasios Data



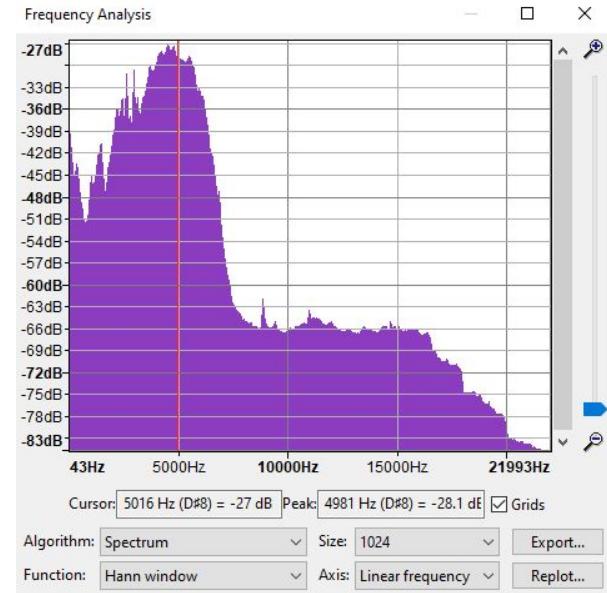
Average of all 15 audios

Mitch Data

blue pipits audio, Quality "A"

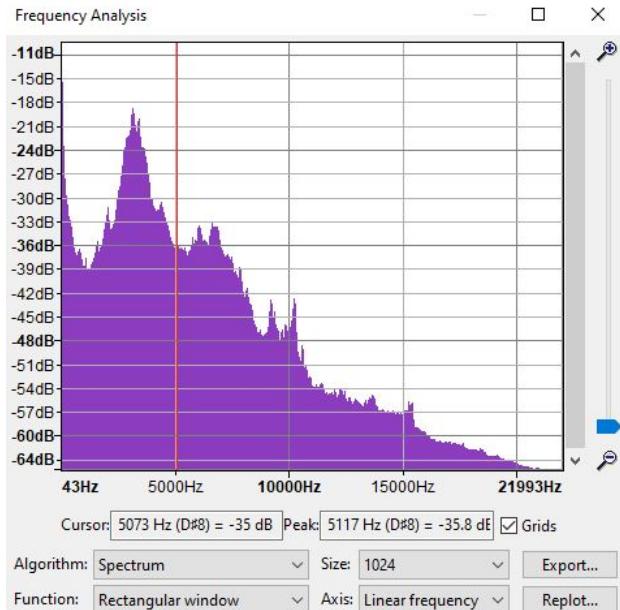


Vocalization type "call" average

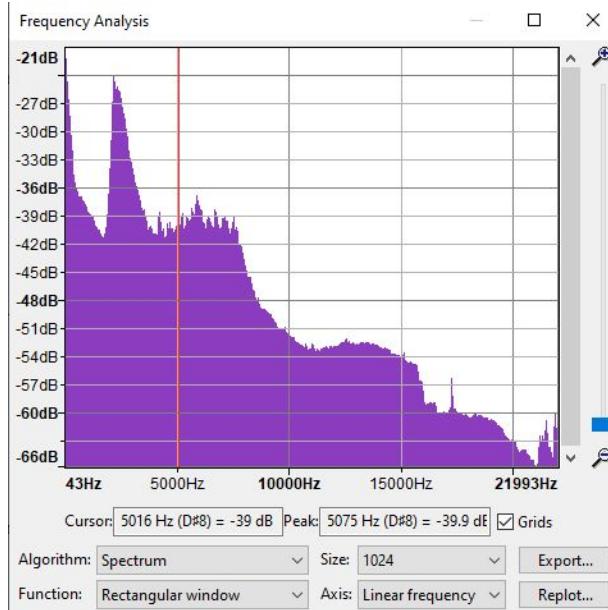


Vocalization type "song" average

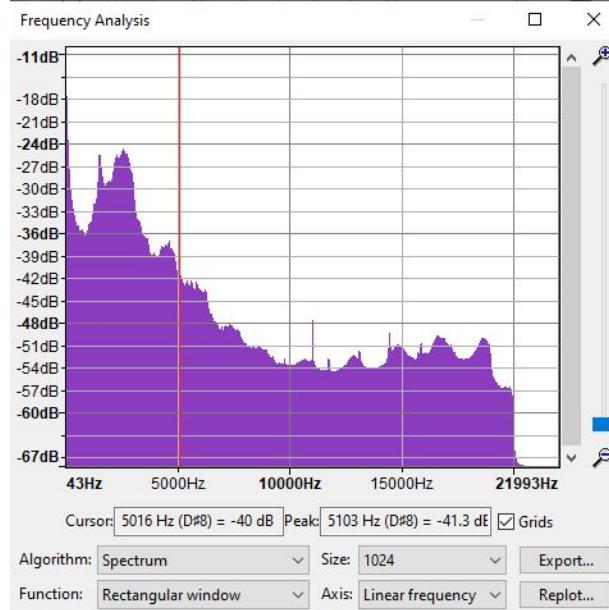
Other Bird Species



Bent-Beak-Riffruff
Average



Vermillion-Trillian
Average



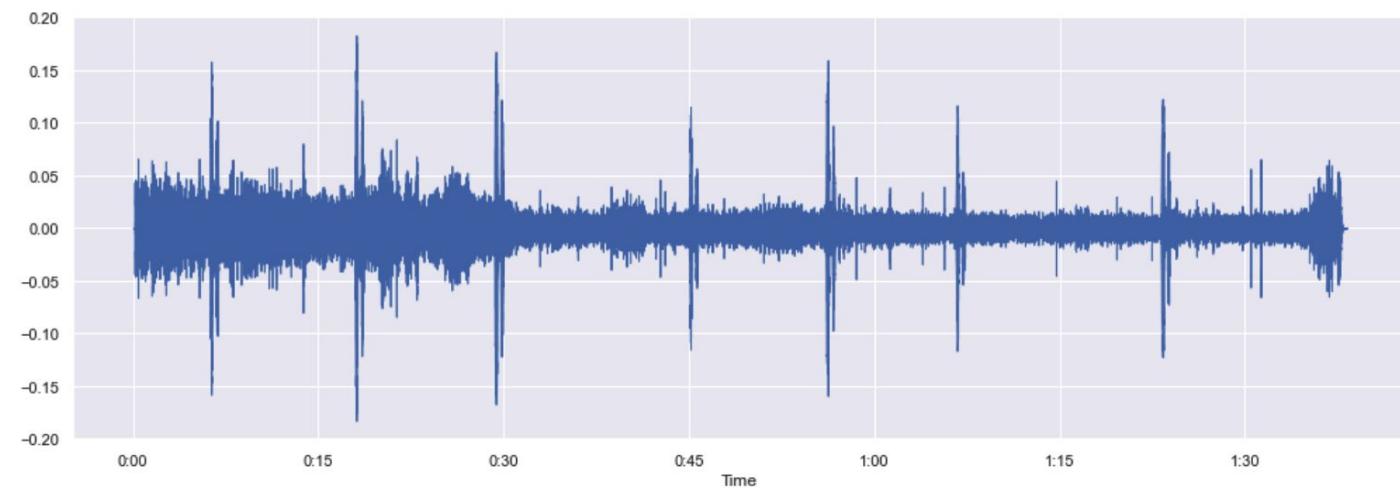
Scrawny-Jay
Average



Machine Learning Approach

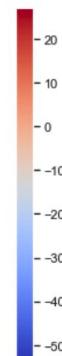
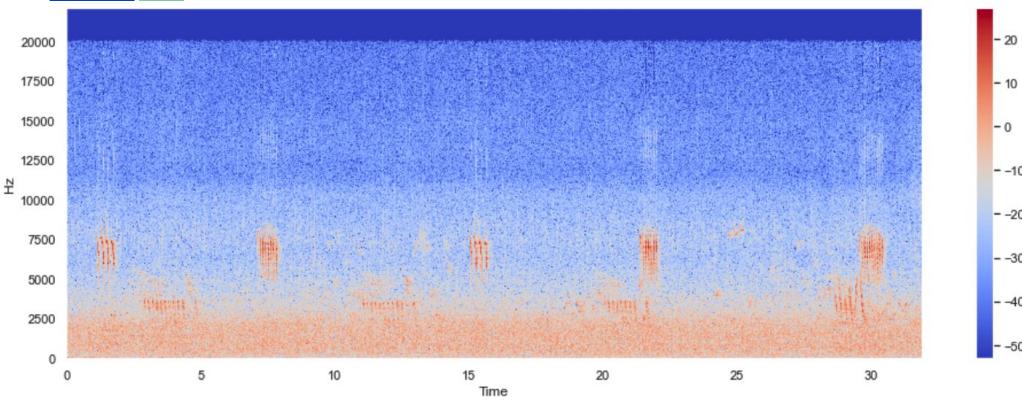
- **Train Data:** Mitch audio data - AllBirds
- **Test Data:** Kasios Test Audios - 15 Test Audios
- **Platform:** Jupyter Notebook
- **Packages:** Matplotlib, Sklearn, Librosa,
- **Audio features** for model training and prediction:
 - Root Mean Square Energy, chromagram short-time fourier transform, spectral centroid
- **Classification Methods:**
 - K-Nearest Neighbors
 - Decision Tree
 - Random Forest
 - Gaussian Naive Bayes

Audio Features

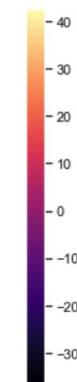
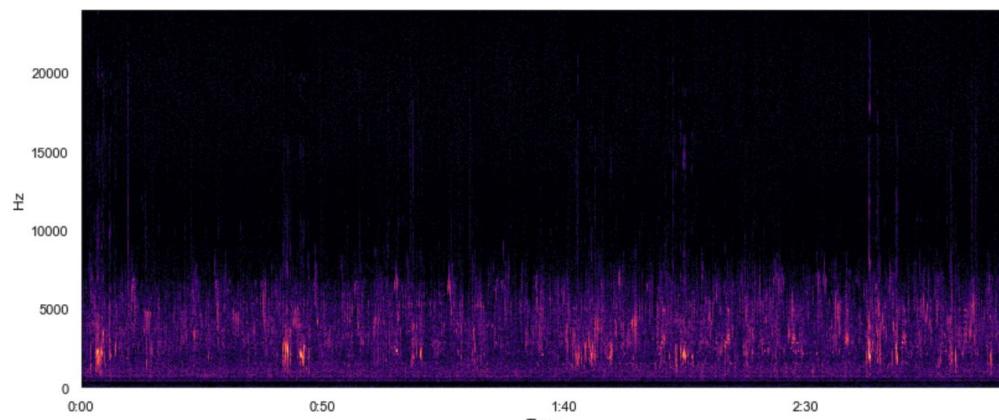


Time Series Amplitude Analysis of Kasios 1.wav

Audio Features



Samples of frequency spectrum in Mitch Data



Feature Extraction



Name	RMS_feature	chrom_stft	spectral_centroid	Label
0 Queenscoat	0.013532	0.582335	1786.110488	0
1 Broad winged Jojo	0.010491	0.382651	2213.188551	0
2 Bent Beak Riffraff	0.031244	0.310886	2642.334428	0
3 Rose Crested Blue Pipit	0.056786	0.393534	3467.098221	1
4 Orange Pine Plover	0.011038	0.312889	1967.989984	0
...
2055 Queenscoat	0.060406	0.641471	700.062052	0
2056 Green Tipped Scarlet Pipit	0.015545	0.468734	3011.493129	0
2057 Ordinary Snape	0.044278	0.350417	2908.622661	0
2058 Ordinary Snape	0.019775	0.425667	4195.976666	0
2059 Orange Pine Plover	0.006247	0.364267	2709.657932	0

2060 rows × 5 columns

1. Root Mean Square Energy
2. Chromagram short-time fourier transform
3. Spectral centroid

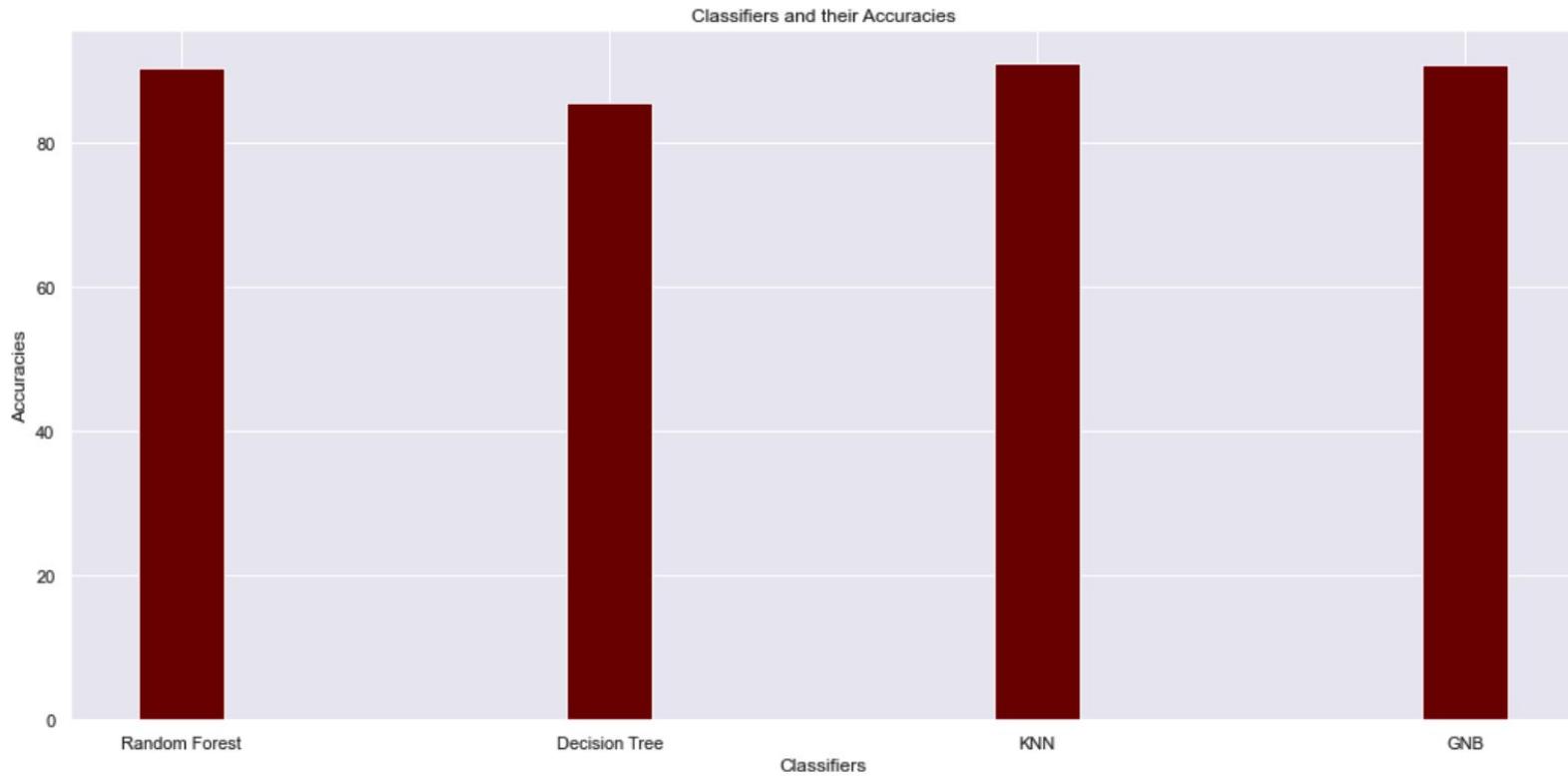


Classification Methods

Four Machine Learning Classification models were implemented for training data.

1. Random Forest
2. KNN
3. Decision trees
4. Gaussian Naive Bayes

Model Accuracies



Prediction Results

```
: y_test=RF.predict(data_test)
print(y_test)
```

```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
```

```
: y_test2=KNN.predict(data_test)
print(y_test2)
```

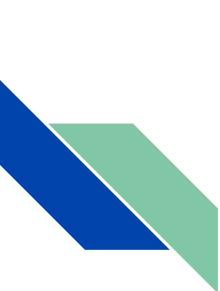
```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
```

```
: y_test3=DT.predict(data_test)
print(y_test3)
```

```
[0 1 0 0 0 0 0 0 0 0 0 1 0 0]
```

```
: y_test4=GNB.predict(data_test)
print(y_test4)
```

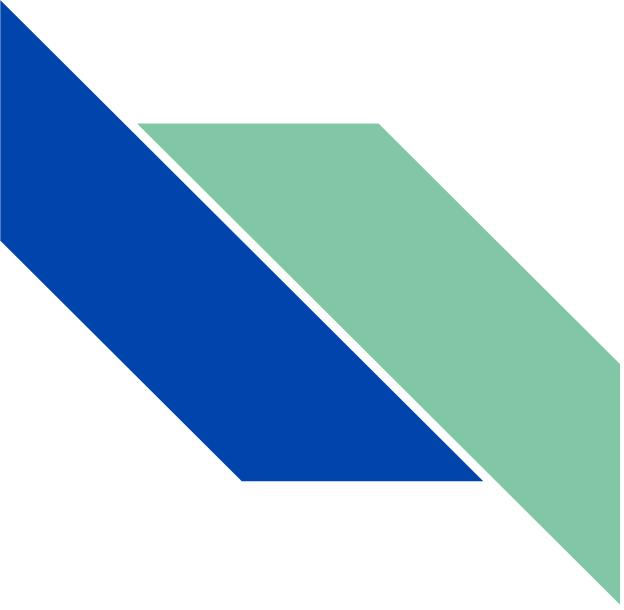
```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
```



Conclusion

Based on the results from our observations and outputs from the machine learning models, we conclude that:

- The data provided by Kasios does not contain only Rose Crested Blue Pipits.
- Thus it does not support Kasios's claim against the research results.



Thank You