

VAST Challenge 2018: Mini Challenge 1

1- Usando la colección de grabaciones y atributos de aves, y el mapa de la reserva de vida salvaje, caracterizar los patrones de todas las especies de aves en la reserva a lo largo del período de tiempo incluido en la colección. Asumir que se tiene una distribución de sensores y de recolectores de observaciones de audio razonable, de modo tal que los patrones observados sean representativos de la ubicación de las aves a lo largo del área de la reserva. *Detecta alguna tendencia o anomalía en los patrones?*

Dentro de la información provista, se encuentran los siguientes archivos:

- **ALL BIRDS:** este directorio cuenta con 2081 grabaciones de aves, para todas las especies que se encuentran dentro de la reserva.
- **AllBirdsv4.csv:** este archivo contiene metadatos acerca de cada una de las grabaciones anteriormente mencionadas, como por ejemplo el día, hora, calidad, y posición en que la grabación fue tomada, la especie del pájaro, y el tipo de sonido que fue grabado.

Se nos informa que en las cercanías de las coordenadas (148,40) la empresa Kasios se encontraría utilizando el espacio como basurero, lo cual puede tener un efecto sobre el comportamiento de las aves.

Análisis de migración

Para validar si esta suposición es cierta, se propuso analizar el comportamiento migratorio de las aves a lo largo de los años. Como la muestra cuenta con múltiples observaciones por cada especie, se tomó la posición promedio de cada una de estas especies. El resultado preliminar que se espera obtener es que ninguna especie de ave se encuentre en las cercanías de la zona utilizada como basurero para el año 2018, año en el cual finaliza la muestra.

Con el objetivo de visualizar la migración de cada una de las especies, se utilizó Tableau para graficar la trayectoria de la posición promedio de cada especie para cada uno de los años de los cuales se tienen observaciones. Con dicha herramienta se pudieron encontrar ejemplos de especies que estuvieron en un radio cercano a las coordenadas (148,40), y otras que nunca estuvieron cerca. Ejemplos estos dos casos serían los siguientes:

- **Especies cercanas al supuesto basurero:** las especies que se consideran en este caso son aquellas que en algún momento del período de tiempo provisto en la muestra estuvieron en zonas cercanas al supuesto basurero. Ejemplos de estas especies son:
 - *Eastern Corn Skeet:* esta especie estuvo en las cercanías del supuesto basurero durante el año 2006, mientras que en los años restantes la especie se retiró de dicha zona.

- *Green-tipped Scarlet Pipit*: esta especie estuvo en las cercanías del supuesto basurero durante los años 1989 y 2002, mientras que en los restantes años estuvo completamente retirado de dicha zona.
- *Rose-crested Blue Pipit*: esta especie es quizá la que mayor cantidad de tiempo pasó en las cercanías del supuesto basurero. De todos los años que comprende la muestra para esta especie (1991 a 2018), estas aves solo se encontraron significativamente fuera de la zona de polución en los años 2000, 2004, y 2018. El hecho de que esta especie de ave se haya alejado y vuelto a la zona del supuesto basurero podría interpretarse como que aquella es la zona en la cual la especie puede vivir en plenitud, pero que por razones relacionadas a la contaminación se vio obligada a emigrar, para luego volver a probar vivir en su habitat original.
- **Especies lejanas al supuesto basurero**: para este caso, ejemplos de especies relevantes de mencionar son la Bombadil, la Canadian Cootamum, la Orange Pine Plover, la Qax, la Queenscoat, y la Scrawny Jay. Estas especies no solo no estuvieron (ni actualmente están) cerca la zona del supuesto basurero, sino que también se encuentran localizadas en zonas geográficas puntuales.

A continuación se ilustra la migración de las diferentes especies, para una selección de especies en función de los dos criterios de ubicación anteriormente mencionados. Vale la pena mencionar que las coordenadas en cuestión (148,40) corresponden a la posición (148,160) del mapa que se utilizará para cada ejemplo, ya que $160 = 200 - 40$, siendo 200 la cantidad de elementos por cada lado del mapa.

Average position of each species throughout the years - Eastern Corn Skeet

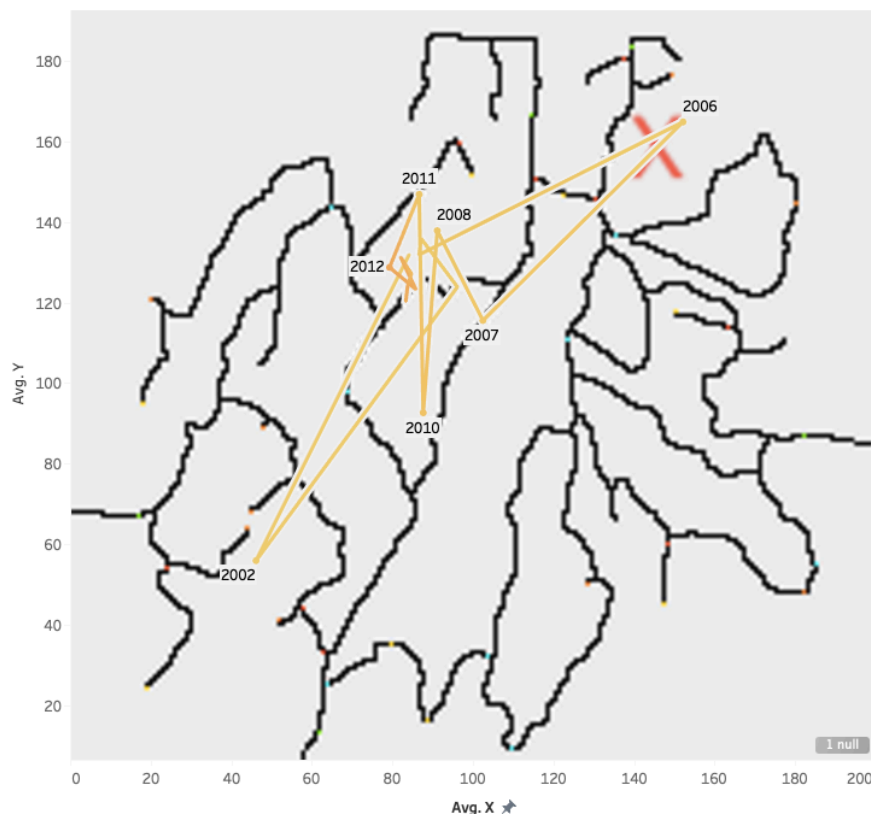
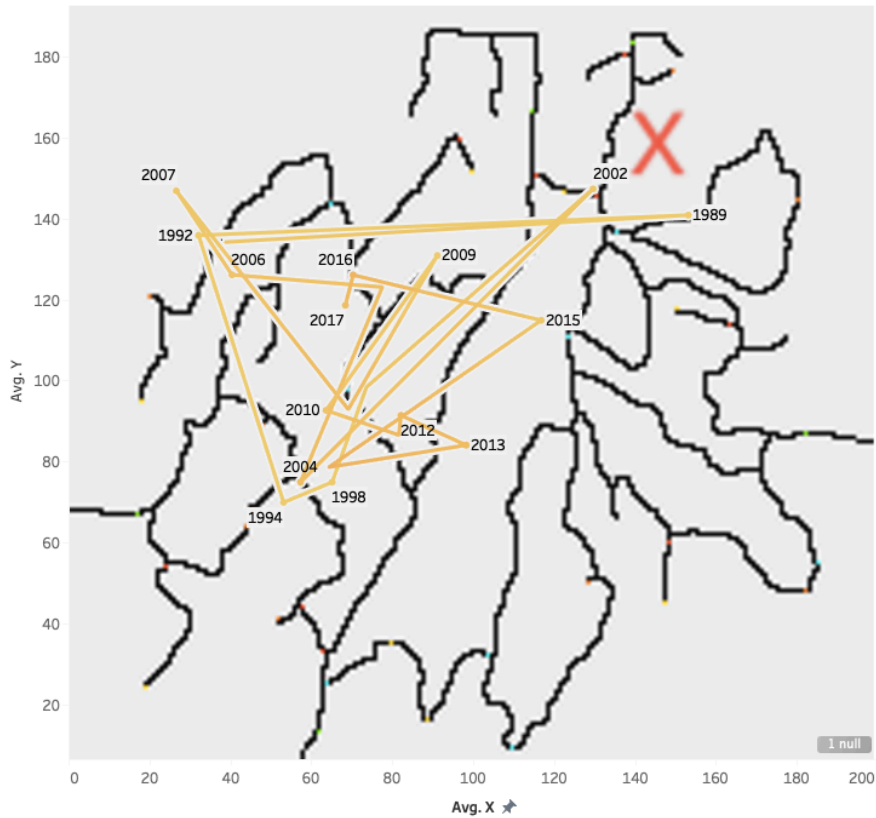


Figura 1

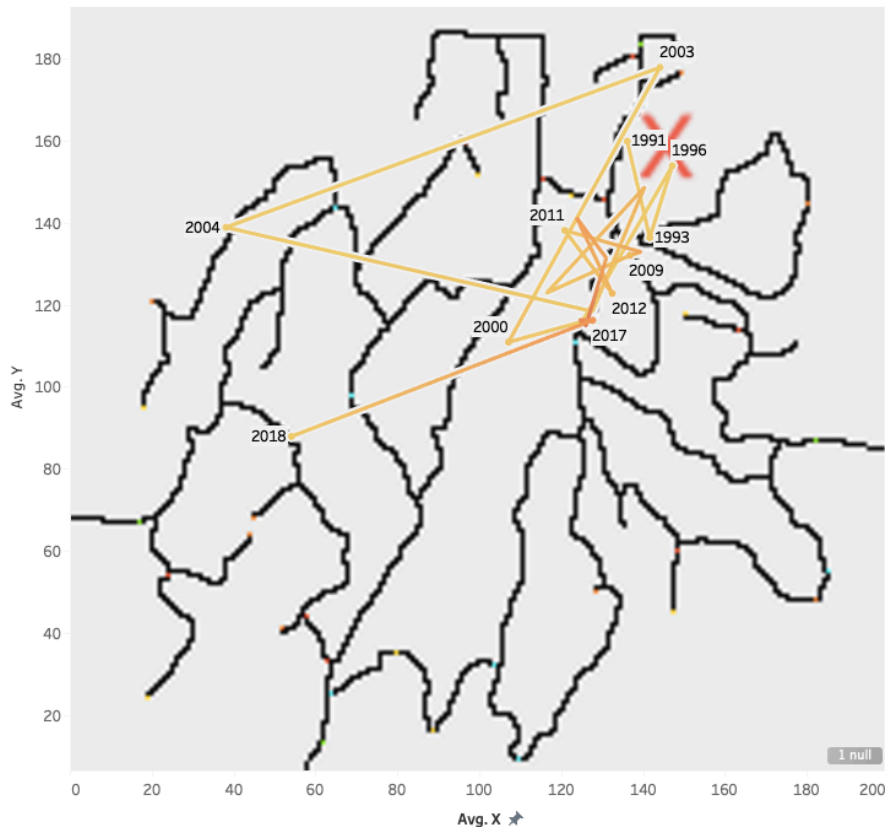
Comportamiento migratorio de la especie Eastern Corn Skeet. Visita la zona cercana al supuesto basurero en 2006.

Average position of each species throughout the years - Green-tipped Scarlet Pipit

**Figura 2**

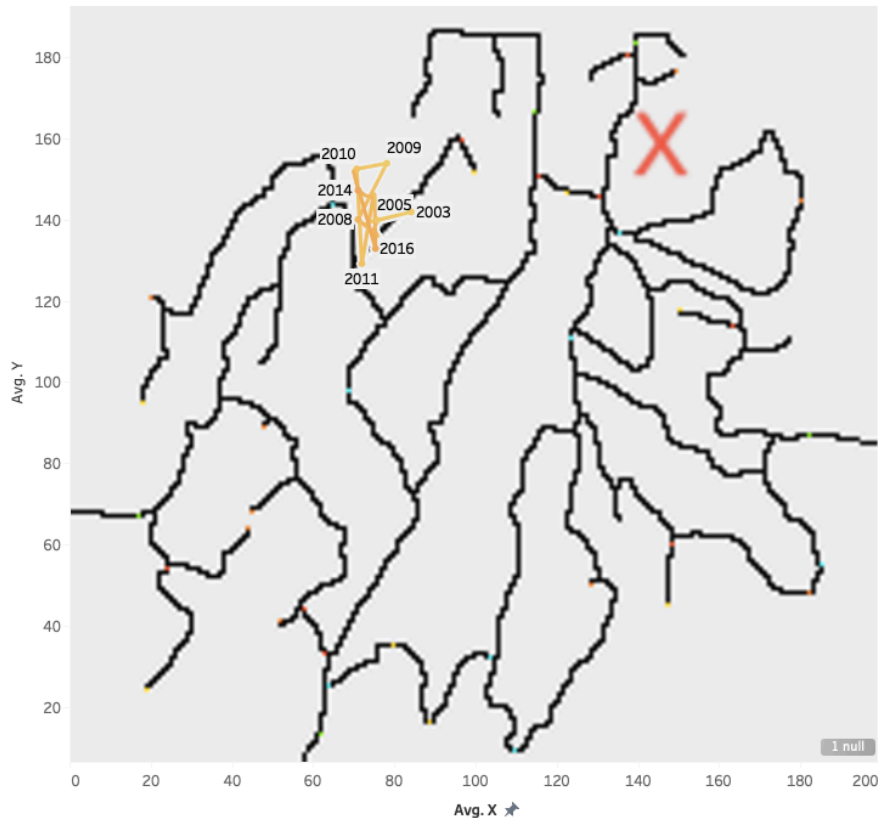
Comportamiento migratorio de la especie Green-tipped Scarlet Pipit. Visita la zona cercana al supuesto basurero en 1989 y 2002.

Average position of each species throughout the years - Rose-crested Blue Pipit

**Figura 3**

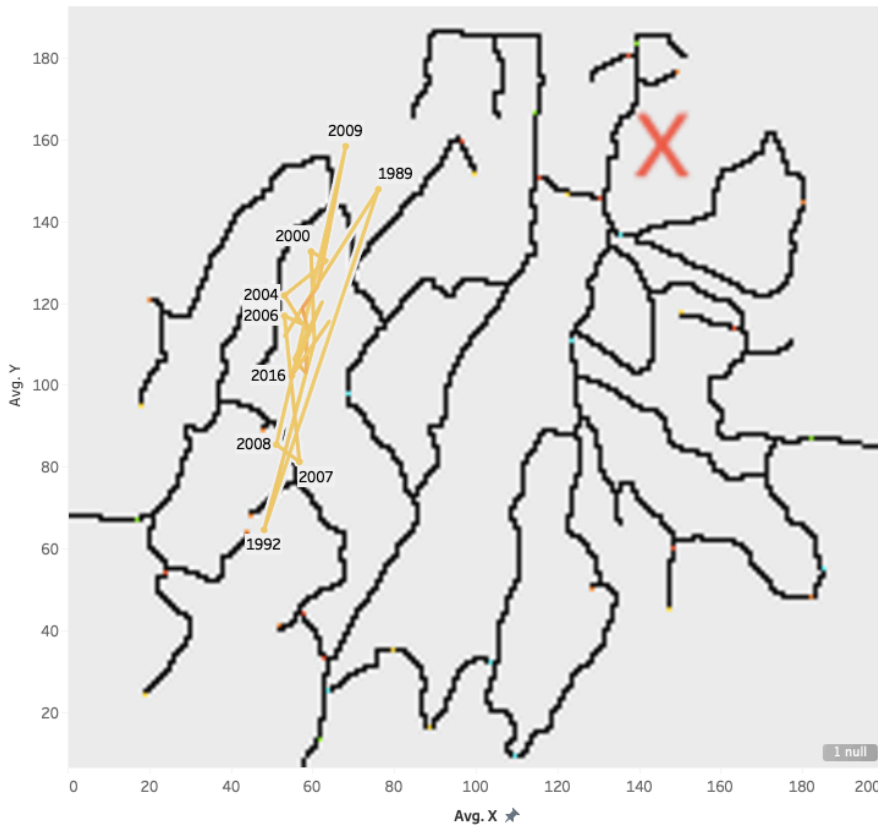
Comportamiento migratorio de la especie Rose-crested Blue Pipit. Permanece de forma reiterada en la zona cercana al supuesto basurero, pero la abandona en tres ocasiones, siendo en 2018 la última vez.

Average position of each species throughout the years - Bombadil

**Figura 4**

Comportamiento migratorio de la especie Bombadil. Esta especie se encuentra localizada en una zona puntual de la reserva, siempre alejada de la zona del supuesto basurero.

Average position of each species throughout the years - Scrawny Jay

**Figura 5**

Comportamiento migratorio de la especie Scrawny Jay. La aves de esta especie se mantuvieron siempre en longitudes similares a lo largo de los años, todas alejadas de la longitud correspondiente a supuesto basurero.

Análisis de tipo de sonido

Además de los movimientos migratorios de las especies de aves, es interesante analizar cómo varió el tipo de sonido que dichas aves realizaban en función de los años. Un análisis preliminar de la distribución de los diferentes tipos de sonidos (*Vocalization types*) para cada año muestra que en un principio la gran mayoría de las grabaciones estaban asociadas al canto de las aves (*song*), pero que con el correr de los años se incrementó notablemente el sonido de tipo llamado (*call*).

El hecho de que el tipo de sonido generado por las aves varíe en el tiempo de canto a llamado puede interpretarse como que las aves ya no cuentan con un espacio para desarrollarse en plenitud (donde el canto representaría el bienestar), sino que ahora pasan más tiempo comunicándose entre si, como una forma de evitar acercarse a la zona del supuesto basurero.

Vocalization type per year

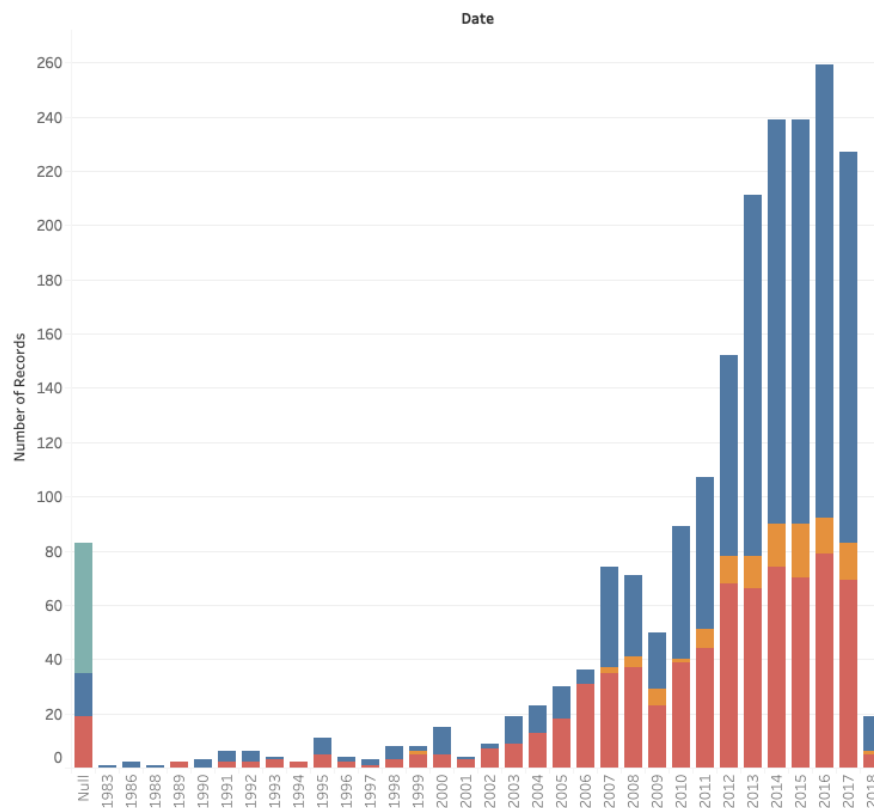
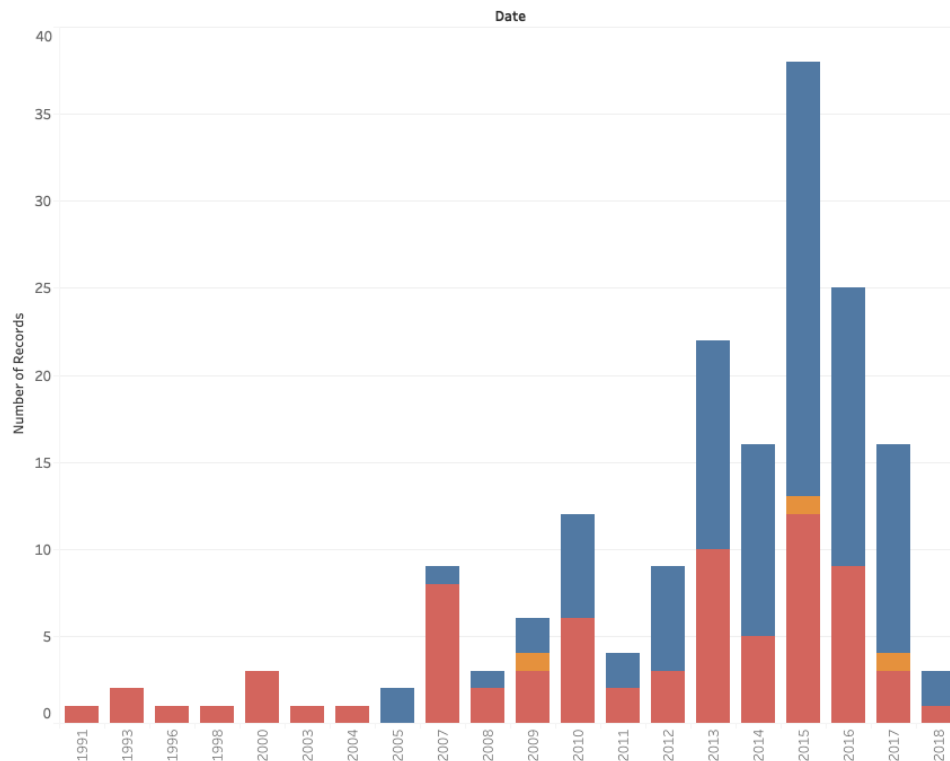


Figura 6

Cambio en la distribución de tipos de sonido emitidos por las especies de aves en la reserva natural. El color azul corresponde a los llamados (*call*), el colorado a los cantos (*song*), y el naranja a sonidos ambiguos (*call, song*). A partir de 2013 se imponen los llamados a los cantos.

Este cambio en el comportamiento del sonido queda mayormente evidenciado en la especie Rose-crested Blue Pipit, la cual a partir del año 2014 realiza más llamados que cantidad de cantos.

Vocalization type per year

**Figura 7**

Cambio en la distribución de tipos de sonido emitidos por la especie Rose-crested Blue Pipit. El color azul corresponde a los llamados (*call*), el colorado a los cantos (*song*), y el naranja a sonidos ambiguos (*call, song*).

2- Observar ahora el conjunto de grabaciones de aves provista por Kasios. Sirve este conjunto para validar el supuesto de que pueden encontrarse Pipits a lo largo y ancho de la reserva?Cuál es el rol de la visualización en su análisis de la muestra provista por Kasios?

Previo al uso de los datos de aves otorgados por Kasios, se realizó un análisis de los audios provistos para cada una de las observaciones de la muestra original, con el objetivo de extraer de los mismos las componentes del sonido que caracterizan a cada especie en particular. Esto se debe a que las grabaciones con las que se cuenta tienen diferentes duraciones, por lo que es necesario tener una métrica que permita comparar 2 observaciones de forma unívoca. Se decidió entonces que las componentes a utilizar por cada especie serán:

- La **frecuencia de sonido** (en kHz) que sea común a las aves de la misma especie, y
- La **potencia** (en dBs) de la señal para esta frecuencia en particular.

Teniendo estos dos valores, cada una de las grabaciones puede transformarse en un conjunto de valores <frecuencia, potencia>, que puede ser representado con un punto.

Para obtener estos valores, fue necesario calcular el espectro de las señales de audio, y a partir del mismo, su representación en decibeles. A partir de este último, se buscó el valor de frecuencia para la cual la potencia era mayor. En este punto entra en juego la calidad de la

grabación. Grabaciones con mayor intensidad de ruido harán que la potencia de las frecuencias más bajas (menores a 1 kHz) sean en algunos casos superiores a aquellas que corresponden a los sonidos emitidos por las aves. Es por esto que fue necesario realizar un filtrado de las bajas frecuencias (en particular aquellas menores a 500 Hz), para obtener resultados más precisos de frecuencia y potencia.

En términos de complejidad computacional, el filtrado de señales para casos de señales de larga duración resultaba en que el filtrado tome demasiado tiempo en completarse. Por lo tanto, se calcularon las duraciones de todos los archivos de audio MP3 y se encontró que la distribución de duraciones tiene el 97.64% de su volumen constituido por grabaciones de longitud menor a 200 segundos (aproximadamente 3.33 minutos). Este valor se utilizó entonces como límite: si la duración de la grabación es menor a 200 segundos entonces se aplica el filtrado; caso contrario el filtrado no es aplicado. Como resultado, se logra reducir el tiempo de cómputo de forma notable.

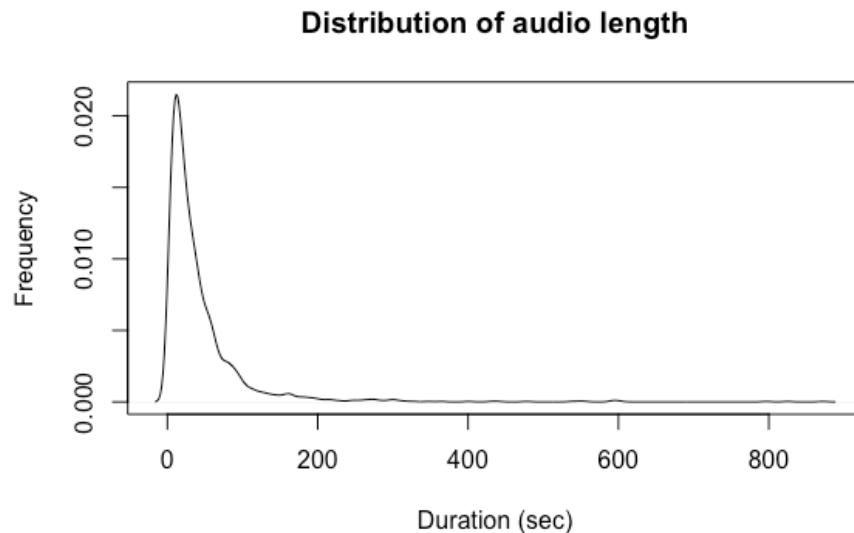


Figura 8

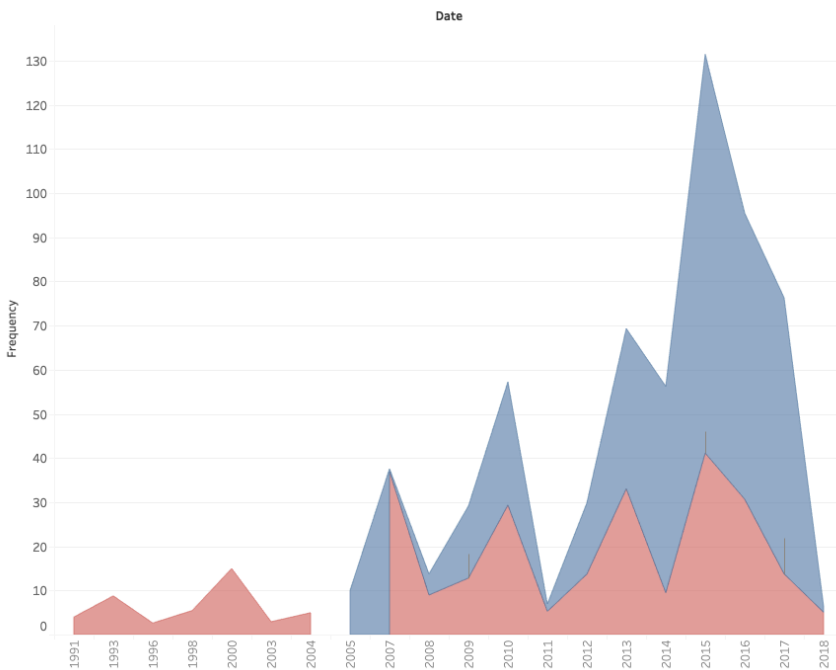
Distribución de las duraciones de grabaciones. 97.64% de estas grabaciones tienen una duración inferior a los 200 segundos.

Luego de que se transforma cada una de las grabaciones en un punto, se procede a calcular el promedio de la frecuencia y el promedio de la potencia para cada una de las especies, de modo tal de contar con una abstracción de cada especie de ave. Contar con esta abstracción nos permite también realizar un análisis de la variación de estos parámetros a lo largo del tiempo para cada una de las especies. En particular, para la especie Rose Crested Blue Pipit, la Figura 9 ilustra el incremento en la frecuencia de los sonidos emitidos por estas aves, particularmente los llamados, teniendo el pico máximo en el año 2015, año en el cual el ave se encontraba en la zona cercana al supuesto basurero.

Figura 9

Incremento de la frecuencia emitida en el sonido de las aves Rose Crested Blue Pipit a lo largo de los años. Este incremento es particularmente notable para los llamados (color azul),

Change in vocalization type for Rose Crested Blue Pipit



y no necesariamente para los cantos (color rojo) que desde 2007 permanecen en valores similares.

Las grabaciones provistas por Kasios son procesadas del mismo modo que las de la muestra original, de manera de contar con un punto que represente a cada una de ellas. Una vez que se cuenta con los puntos, se calcula la

diferencia entre la frecuencia de una de las observaciones provistas por Kasios contra las frecuencias promedio de las 19 especies de aves. Como resultado de esto, se atribuirá la observación provista por Kasios a aquella especie para la cual la diferencia (o distancia) sea la menor de las 19 diferencias. La siguiente tabla muestra la especie más similar en términos de distancia para cada una de las 15 grabaciones provistas por Kasios.

Grabación ID	Distancia	Especie Original Promedio	Frecuencia Kasios	Frecuencia Original Promedio
1	7.95367	Rose-Crested-Blue-Pipit	2.76537	3.8003
2	1.4456	Green-Tipped-Scarlet-Pipit	2.63456	4.07136
3	1.34749	Purple-Tooting-Tout	5.44669	4.19552
4	1.32588	Qax	2.72095	3.13354
5	5.02412	Rose-Crested-Blue-Pipit	5.37458	3.8003
6	2.07225	Rose-Crested-Blue-Pipit	4.95889	3.8003

Grabación ID	Distancia	Especie Original Promedio	Frecuencia Kasios	Frecuencia Original Promedio
7	1.82811	Vermillion-Trillian	1.17933	2.5172
8	0.39257	Canadian-Cootamum	4.2917	4.13919
9	9.74873	Rose-Crested-Blue-Pipit	4.40236	3.8003
10	3.54975	Rose-Crested-Blue-Pipit	4.33803	3.8003
11	2.1451	Qax	1.56945	3.13354
12	0.75767	Rose-Crested-Blue-Pipit	3.42584	3.8003
13	4.82413	Green-Tipped-Scarlet-Pipit	1.77871	4.07136
14	1.73564	Rose-Crested-Blue-Pipit	4.44369	3.8003
15	7.09585	Rose-Crested-Blue-Pipit	5.27724	3.8003

Tabla 1: Asignación de especies de la reserva a las grabaciones de Kasios.

Como puede apreciarse en la Tabla 1, 8 de 15 grabaciones provistas por Kasios pueden ser asociadas con el Rose Crested Blue Pipit promedio, a partir de la distancia entre los puntos. La distancia utilizada en este caso fue la euclídea.

3- Formular una hipótesis acerca del estado de la especie Rose Crested Blue Pipit. Cuáles son las piezas de evidencia principales de su hipótesis? Cuáles son los pasos que deberían tomarse para tanto validar o refutar el supuesto que plantea Kasios acerca de que los Rose Crested Blue Pipits viven tranquilamente en la zona de la reserva natural de Boonsong Lekagul?

Como resultado del análisis detallado en los dos incisos anteriores, concluyo que las acciones tomadas por Kasios en la reserva natural *tienen un efecto sobre el estado de la mayoría de las especies de aves, y principalmente sobre los Rose Crested Blue Pipits*. Las razones que me permiten arribar a esta conclusión son:

1. Las aves de esta especie intentaron volver a su habitat original de forma reiterada desde el año 2000, pero desde entonces que los periodos de estadía en su habitat

son más cortos. Actualmente, para el año 2018, esta especie se encuentra fuera de su habitat original.

2. Hubo un cambio notorio en la distribución de sonidos emitidos por las aves de esta especie, las cuales cambiaron el canto por comunicación entre aves.
3. Hubo un incremento en la frecuencia del sonido emitido por estas aves.
4. Aproximadamente la mitad de las grabaciones provistas por Kasios son de aves de especie Rose Crested Blue Pipit. El hecho de que varias de ellas coincidan con este tipo de ave no necesariamente implica que estas aves viven en plenitud en el parque, sino que las observaciones recogidas corresponden a instancias de aves que se encontraban en ese momento en esa posición.

Como pasos que deberían tomarse para asegurarse que las afirmaciones anteriores son correctas mencionaría:

1. Contar con un número estadísticamente significativo para observaciones en cada año. Se nos indica en el primer enunciado que debemos "*(...) asumir que se tiene una distribución de sensores y de recolectores de observaciones de audio razonable, de modo tal que los patrones observados sean representativos de la ubicación de las aves a lo largo del área de la reserva*", lo cual puede interpretarse como que años con una o dos observaciones tienen el mismo peso que años con diez o más observaciones.
2. En lugar de transformar los audios en una representación de puntos, se podría utilizar espectrogramas, y comparar las tramas enventanadas en cada grabación con las provistas por Kasios.

Recursos

1. Repositorio en GitHub: github.com/tulians/VAST2018-MC1