MRUTULA SURESH 28967828

DATA EXPLORATION PROJECT

PROJECT TITLE:

TURKEY VULTURES IN NORTH AND SOUTH AMERICA MIGRATION

INTRODUCTION

The turkey bird is a very large bird that is native to the Americas. It has a global population of 5 million and it is increasing.

Turkey vulture dataset consists of 22078 rows and 32 columns. From 2003 to 2011, 24 turkeys from 4 different geographical areas were fitted with transmitters and the corresponding data was recorded. In this process the data of some birds were removed due to equipment failure, death of bird etc.

Four regions are as follows,

- **West coast of North America** – Morongo, Rosalie, Irma, Sarkis, Prado, Mary, Schaumboch
- **East coast of North America** – Disney, Butterball,Mark, Steamhouse 1, Steamhouse 2,Mary
- **Interior of North America** – Leo, Mac, Domingo, Young Luro, Whitey
- **Interior of South America** –La Pampa, Argentina

Question:

- Total distance travelled by each bird
- Distance travelled in each month by each bird
- Speed of each bird
- How does the mass affect the speed of the birds
- Speed in outbound and return migration

Motivation:

To understand the migratory behaviour of the world's most abundant and widely distributed species, Cathartes aura

DATA WRANGLING:

I am using R to perform data wrangling on the dataset. The Dataset is in .xlsx format i.e it is an excel file. It's a bit tricky to use an excel file in R. I had to perform parsing. Initially I used "xlsx" and "openxlsx" libraries but they couldn't handle the parsing of the huge file and I ended up getting a memory error. After some research on the internet I came across the "readxl" package which helped to parse the excel file into a dataframe in R. I manually renamed the columns for my convenience.

There are around four columns containing date time information but were stored as a character fields. When I used as.Date() I lost the time information which is important for the speed calculation.

I came across the POSIX* function. as. POSIXct() function helped to preserve the time along with the date information. But this ended up giving me NA's for some of the datetime values. To solve this problem I set the TimeZone to GMT

After obtaining the summary of the dataframe I realised that some of the birds didn't have the mass mentioned. But since the mass of some of the birds from their respective population was mentioned, the ones that didn't have it were ignored during analysis that required the animal-mass.

DATA CLEANING:

After an initial analysis of the data, I realised that Butterball turkey was tracked only from Nov 14, 2003, through March 14, 2004, and Schaumboch was tracked from Oct 08 2004 through March 26 2006. But in the dataset the data of both the vultures are included twice. i.e once with each animal_id of "Butterball" and once with animal_id of "Schaumboch". However if we sort by the "individual_local_identifier" we get the right set of data.

So I grouped those rows from the dataframe in which the animal_id and the individual_local_identifier don't match , i.e Butterball, Schaumboch or Schaumboch, Butterball, and delete them from the dataframe. By doing this we have the correct number of rows for each bird.

DATA EXPLORATION:

The event id is an unique Id for each event and the timestamp column gives the data and time for each event. Since I have already converted it into datatime column, I extract the day, month and the year. Since I have used POSIXct for converting the timestamp column to DateTime format I didn't have any direct method to extract the day month and year information. So I used them as.Date() method followed by the format() method. This method returns only character string so I converted into numeric and stored it in respective columns.

The months were initially stored as 1 to 12 denoting the months from Jan to December. So I used mapvalues() method to map the values 1 to 12 into Jan to Dec.

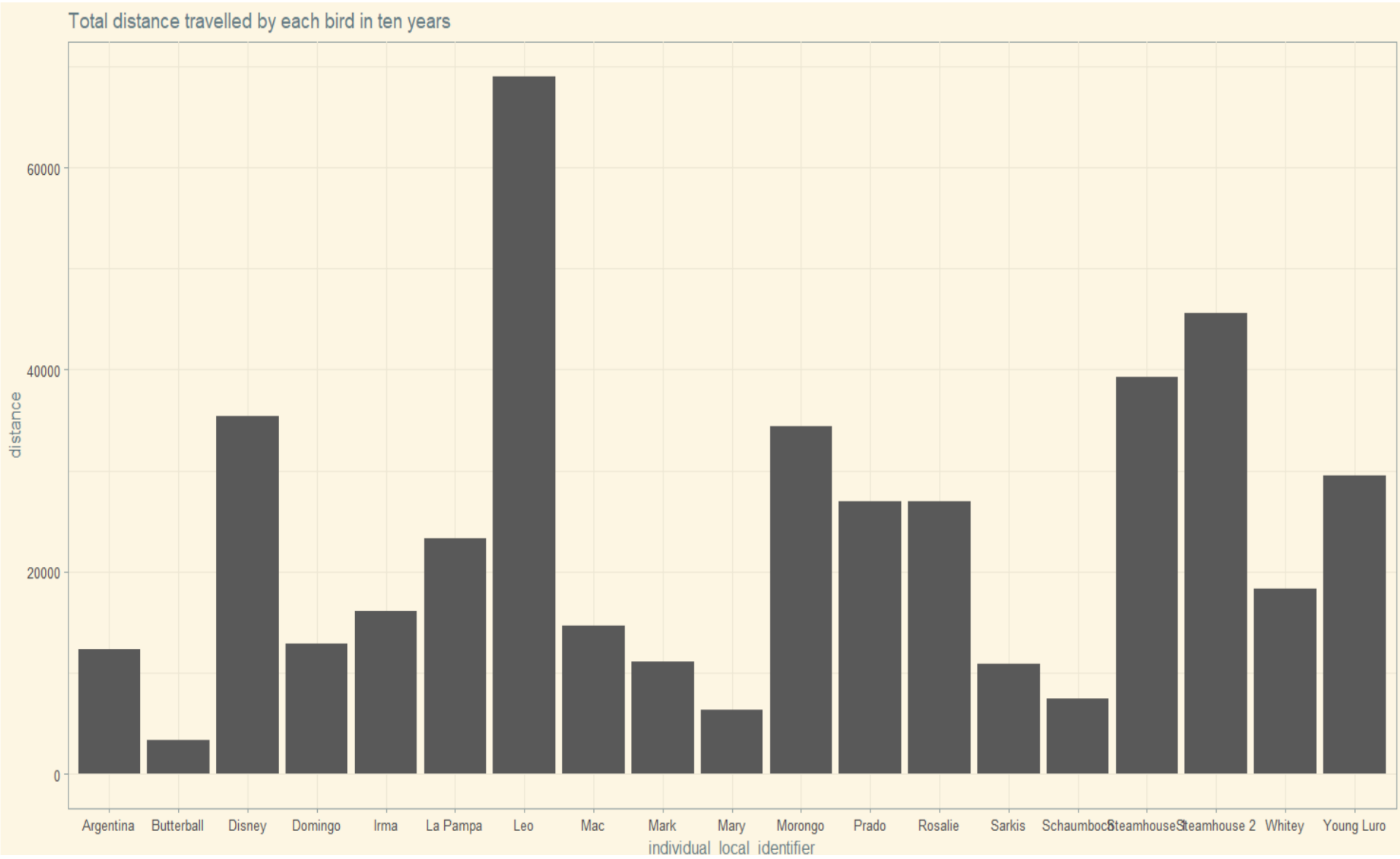
In each row the change of latitude and longitude of the bird is given. In order to calculate the distance travelled between each change, I used Great circle Distance Formula. This returned the distance in km and I converted it into miles by multiplying It by 0.62137.

From the DateTime information I extracted the time details. From this time I calculated the time in hrs between each change in location.

With the time taken and distance travelled information I calculated the speed of each bird using the following formula,

$$\textit{Speed of the bird between two locations} = \frac{\textit{Distance travelled between two location}}{\textit{Time taken between two location}}$$

I plotted the total distance travelled by each bird,



INFERENCE FROM GRAPH

From the graph, we can see that Leo, Steamhouse 2 and Steamhouse 1(interior North American population) have travelled the longest distance. This is because these birds use land-based corridors that extend further south than peninsular Florida. Whereas the lowest distance travelled by the Butterball, Mary, Schaumboch, Mark i.e. the east coast population. This may be because their migration Inland is restricted as the corridor ends at Florida peninsula, unlike the interior North America population.

SEASONS:

BREEDING SEASON(Summer): It the period between the RETURN MIGRATION and the OUTBOUND MIGRATION. Species be in their own sites (during their respective summer times)

NON BREEDING SEASON(Winter): This period starts after the bird has completed the outbound migration and before it starts the return journey back to the individuals breeding grounds

OUTBOUND MIGRATION(autumn): This period starts once the breeding season gets over and when autumn starts in the individual's breeding ground, i.e when it flies to the non-breeding ground

RETURN MIGRATION(spring): It is the period when the bird flies from the non-breeding grounds to the breeding ground i.e when their spring starts in the individual's breeding ground.

MRUTULA SURESH 28967828

The average distance travelled by each bird in each month,

From the graph, we can obtain the following information,

WEST COAST BIRD:

- BREEDING SEASON: It is from May to Oct
- OUTBOUND MIGRATION: From Sept to Nov
- NON- BREEDING: Is from Dec to Feb
- RETURN MIGRATION: It is from March to April

EAST COAST BIRD:

- BREEDING SEASON: It is from June to August
- OUTBOUND MIGRATION: From Sept to Nov
- NON- BREEDING: Is from Dec to Feb
- RETURN MIGRATION: It is from March to May

INTERIOR OF NORTH AMERICA BIRD:

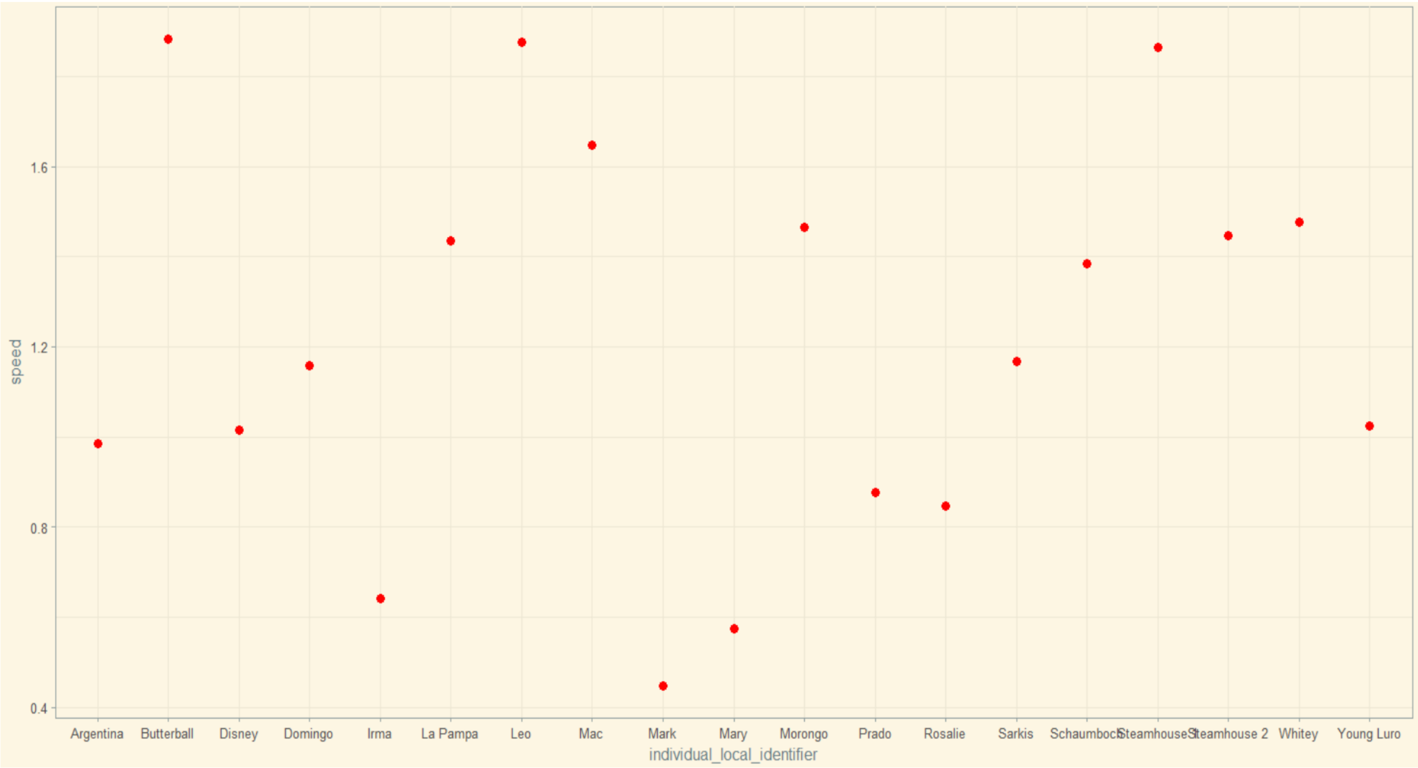
- BREEDING SEASON: It is from June to September
- OUTBOUND MIGRATION: From Sept to Nov
- NON- BREEDING: Is from Dec to Feb
- RETURN MIGRATION: is from March to May

MRUTULA SURESH 28967828

INTERIOR OF SOUTH AMERICA:

- BREEDING SEASON: It is from Dec to Feb
- OUTBOUND MIGRATION: From March to May
- NON- BREEDING: Is from June to August
- RETURN MIGRATION: It is from Sept to Nov

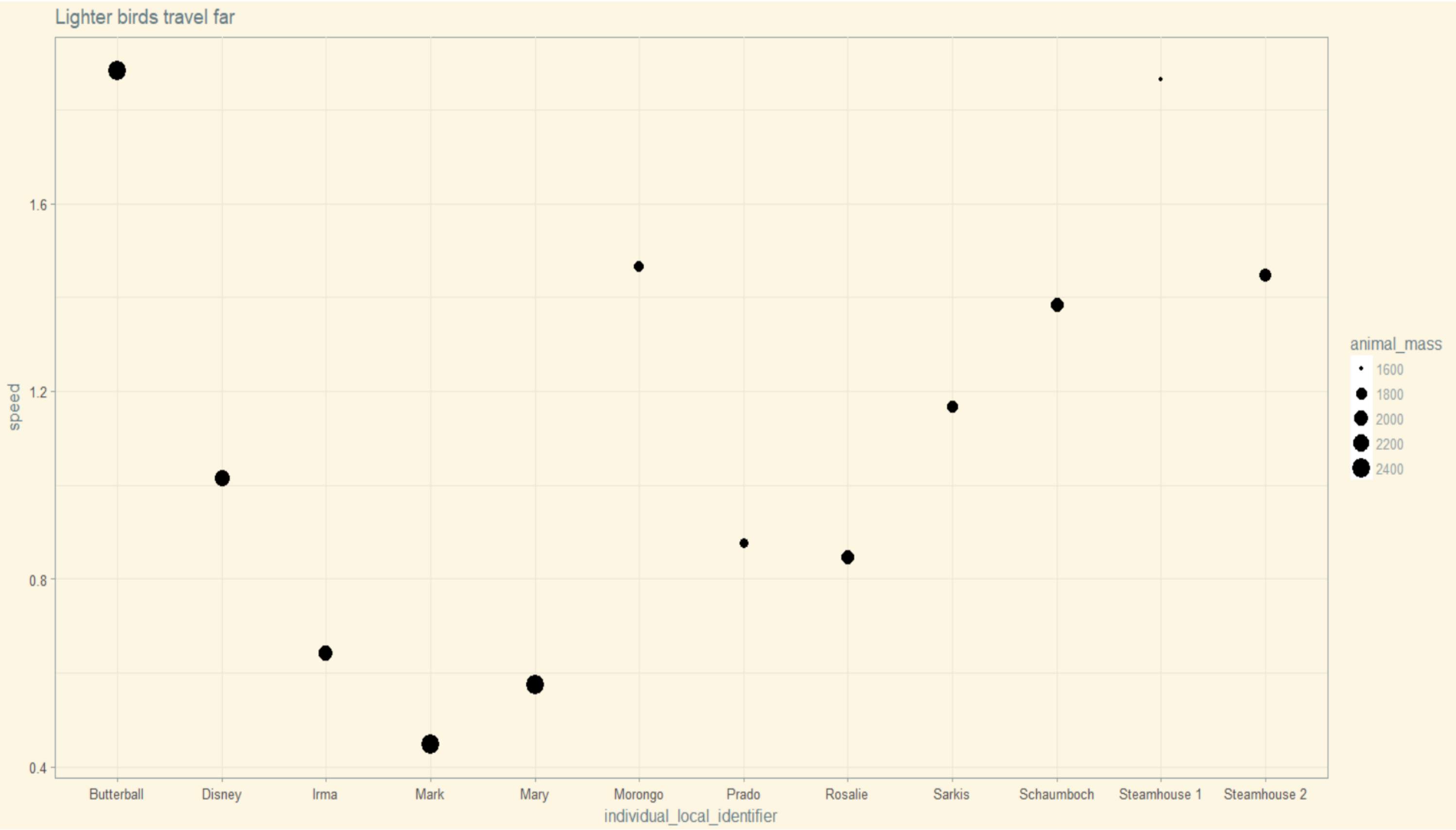
I plotted the speed of each bird,



From the graph, we can see that Butterball, steamhouse 1, Mac and Leo seems to have very high speed. We can see that most of the interior North American bird population have high speed.

MRUTULA SURESH 28967828

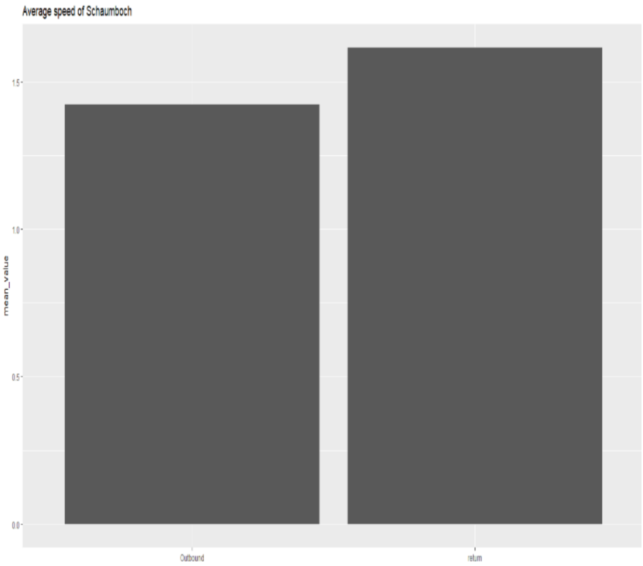
To see how the body mass affects the speed,



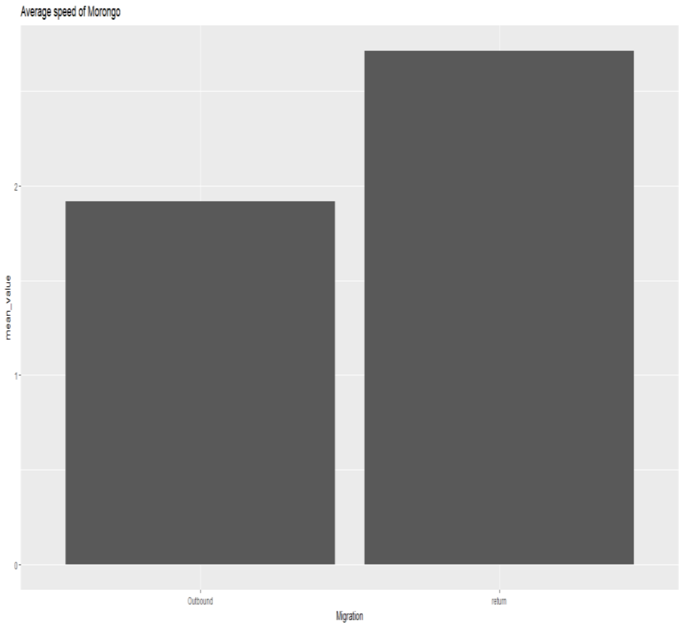
We can see that lighter birds travel faster. Butterball bird has a mass of 2372g and has a high speed. This might be because of high wind speed.

OUTBOUND AND RETURN MIGRATION:

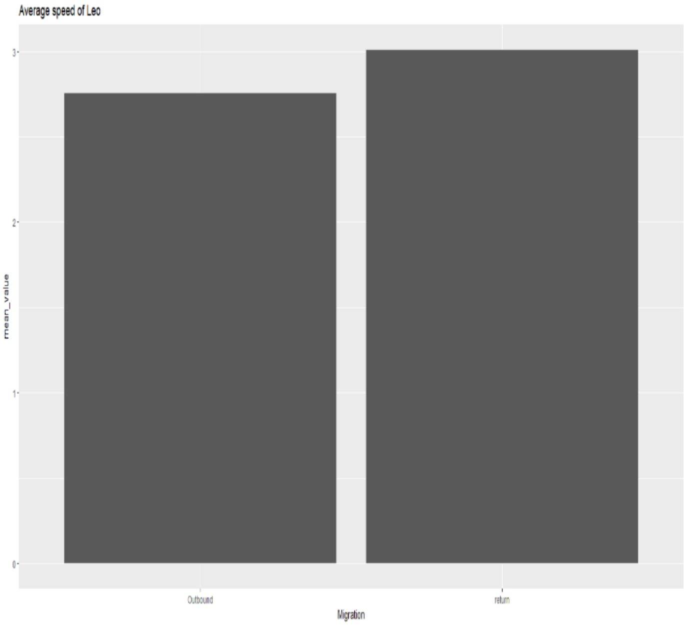
I plotted the outbound and return migration speed of the following birds one from each of the 4 population, east-west interior north and interior south America



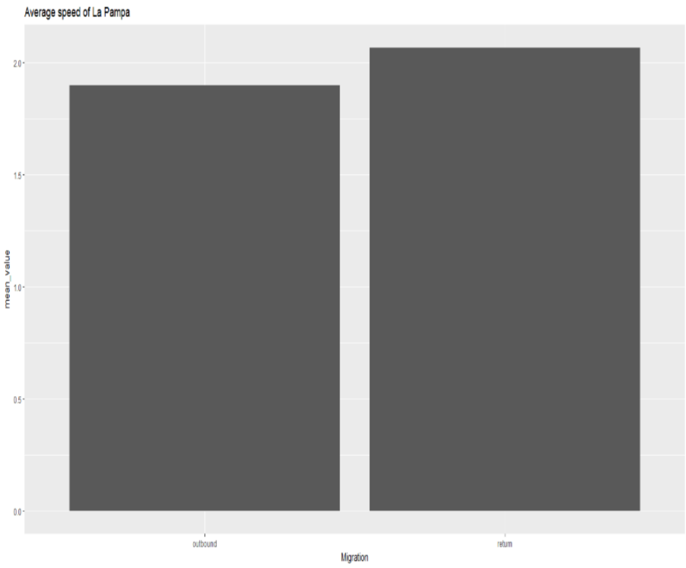
Schaumboch, East coast of North America



Average outbound and return migration speed of Morongo, West coast



Average outbound and return migration speed of Leo, interior north America



Average outbound and return migration speed of LA Pampa, interior south America.

So on an average, most of the birds have a higher speed on return migration than on outbound migration. This might be because outbound migration takes place after breeding so it involves many slow juveniles who are inexperienced. But on return migration they will be more experienced and tend to migrate faster.

CONCLUSION:

- I calculated the total distance travelled by each bird and found that the birds belonging to the interior of North America population travel longer distances because they use land-based corridors

- The birds tend to travel longer distances during outbound and return migrations and don't travel much during the breeding and non-breeding periods
- Speed of each bird was calculated from the distance travelled and time taken
- Lighter birds tend to travel faster
- Some birds in spite of not being lighter seemed to travel faster. This might be because of the wind speed
- Birds tend to travel faster during the return migration than the outbound migration. This might be because young and inexperienced birds fly during the outbound migration whereas experienced ones fly back during the return period.

REFLECTION:

I collected this dataset from movebank. Though it didn't have a lot of missing data I had to derive many of the features from the existing fields. I really loved working with this dataset and learnt a lot about the turkey bird which I didn't know before. The most difficult part was calculating the speed. But ultimately, I was able to answer my initial questions. If I get a chance to work with this dataset again I might also try to get the environmental data so that I can see what all factors affect the migration of the birds.

REFERENCE AND BIBLIOGRAPHY:

- https://data.world/makeovermonday/2018-w-4-turkey-vulture-migration-in-north-and-south-america
- http://rstb.royalsocietypublishing.org/content/369/1643/20130195.figures-only
- https://www.datarepository.movebank.org/handle/10255/move.363