From: Junqing Peng, Graduate Student Consultant
MAS Final Project
RESEARCH FIELD: Data Analysis
PROJECT TITLE: Birds Immigration Analysis

1.0 - PROJECT DESCRIPTION

This project analyzes migratory stopover locations of bird to identify the spatial pattern and all regions across all individuals which is an observational study. The goal is to uncover biologically important stopover locations distribution.

Planning Stage:

1.Stopover events were extracted from individual migration tracks using temporal windows around candidate stopover days.

2. Using normal distribution and 2D KDE which can estimate spatial density of stopover locations.

3. Visualization of the contour maps and heat maps to highlight the important region. Also, include distribution of different latitudes and longitudes.

4. Compare Normal with 2D kernel density to see which one fits better.

1.1 - RESEARCH QUESTIONS

Question 1: What is the distribution the stopover points?

Question 2: . Where are the high-density stopover locations along the migratory route?

1.2 - STATISTICAL QUESTIONS

Question 1:What is the estimated joint spatial distribution of stopover locations across individuals?

Question 2: Are stopover locations spatially clustered or uniformly distributed?

Question 3: How many statistically distinct stopover zones exist within the KDE surface?

Question 4: Is there a statistically significant difference in stopover density between regions?

Question 5: Does a multivariate normal significantly underperform compared to KDE in modeling stopover density?

1.3 - VARIABLES OF INTEREST

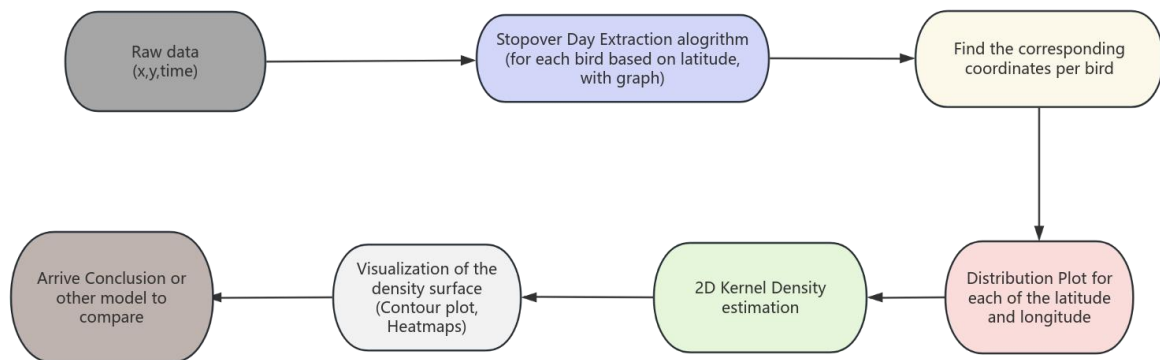Longitude: Geographic coordinate (x-axis) of stopover point

Latitude: Geographic coordinate (y-axis) of stopover point
ID: Birds ID
Day of the year

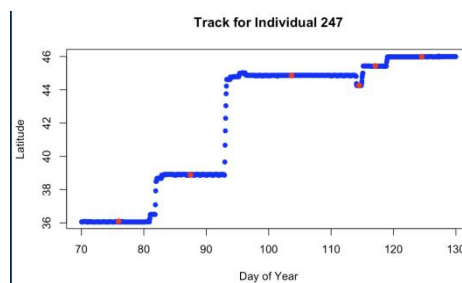## 1.4 - Study Diagram

2.0 - EXPLORATORY DATA ANALYSIS (EDA)

2.1 how to identify these stopover point
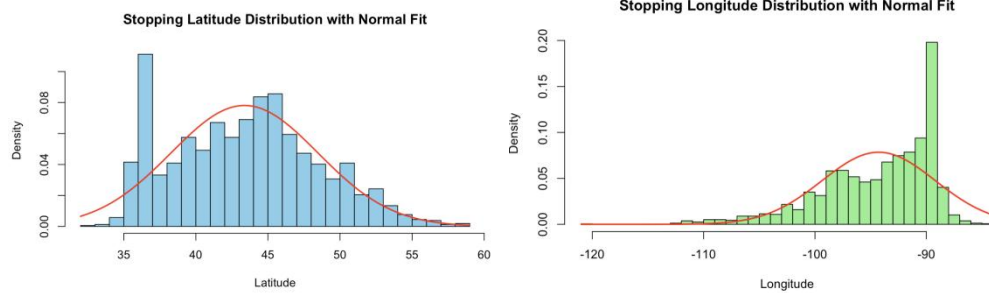
1. Create an empty data-frame is created to collect these stopover data.
    A. At least 4 consecutive low-moment days to define a stop
    B. Daily travel distance have to less than 45 km per day to be defined as a stop
2. Create a loop
    A. Loop through all 252 graphs to extract coordinate
    B. Only points within a +-10 days around individual's migration segment are kept
3. Distance Calculation
    A function is used to calculate the distance using the package distHaversin
4. Stopover detection
    A. Identifying the runs of consecutive days with movement < 45 km
    B. Use rle function to detect low movement days
   C. When low-movement day is greater than 4, we can identify it as a stopover.
    Example:

2.2 EDA

Stopping Latitude Distribution with Normal Fit

Stopping Longitude Distribution with Normal Fit

According to the univariate distribution, for latitude, it is a little bit right-skewed. The normal fit does not capture all the distribution. This indicates that s birds tend to cluster more in southern stopovers.

In addition, for longitude, it is also skewed, indicating the presence of **multimodality** or clustering behavior.

## 3.0 –STATISTICAL ANALYSIS

Fitted Bivariate Normal Distribution
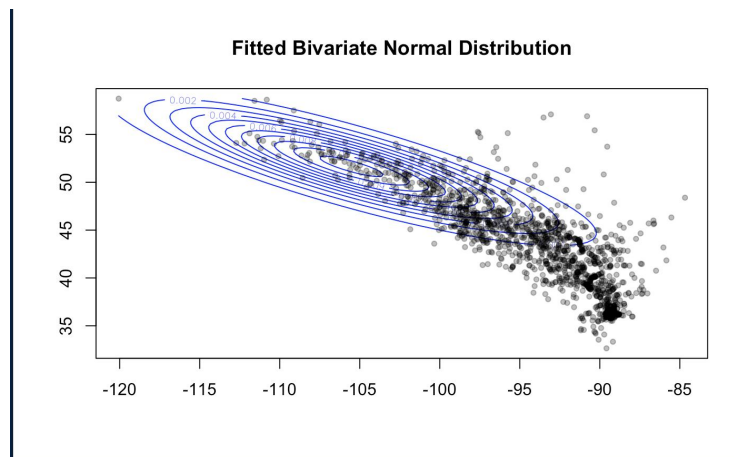
In this case, we tried to fit a bivariate normal distribution plot. However, there is major shift away from the normality since the normal blue contour do not fit the data. In this case, we need more advance distribution.

2D Kernel Density of Stopover Locations

2D Kernel Density Estimate (KDE)
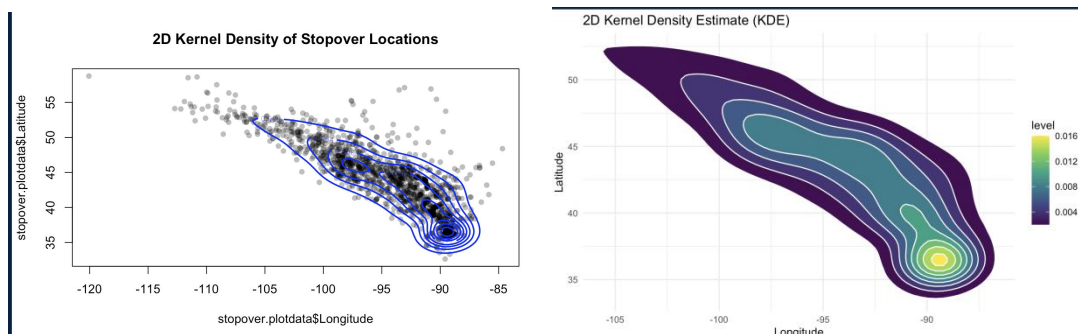
Using 2D kernel density, the brighter area means that data is more concentrated in that area. In this case, the stopover point is mostly concentrated in southeast corner.
This region is:

(A) The key region which takes longer duration to stay

(B) Since the destination is over there, it may lead to high concentration

In addition, there are six statistically distinct stopover zones based on GMM:
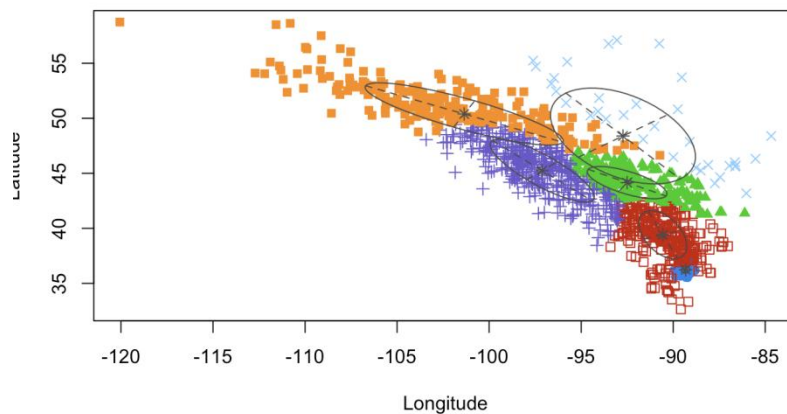
4.0 - RECOMMENDATIONS

Question 1: Using the KDE, we found that the highest concentration of stopover activity is located in the southeastern portion of the migration corridor.

Question 2: They are not normally distributed since the normal plot do not fit so they are clustered

Question 3: There are six statistically distinct stopover zones based on GMM

Question 4: Yes, there is statistically significant difference in stopover density between regions based on the KDE model

Question 5: Yes, KDE model fit much better.

5.0 - RESOURCES

https://rdrr.io/github/nickmckay/GeoChronR/man/kde2d.html

https://en.wikipedia.org/wiki/Kernel_density_estimation