# Ebird Data Exploration

## Ariel Mundo

BIOL 580V

Department of Biomedical Engineering
University of Arkansas
Fayetteville, AR, USA

# Introduction

This presentation shows the work done to analyze the Ebird dataset, which was provided in the BIOL 580V course.

These are the steps I took to analyze the data:

- Load libraries for plotting and data cleaning

- Data cleaning

- Quick analysis to decide direction of analysis

- Visualization

# Loading and cleaning the data

- Load data, the separator for the data was tab or "\t".
- The dataset contains many columns that we might not be interested in. I was interested in the Common Name, Genus, Species Count, Location Name, Latitude, Longitude and observation date.

**Loading data**     Dataset head

```
data<-read.csv(here("data/ebird.csv"),header=FALSE,sep="\t", quote="")

data<-data%>%
    separate(col=V2,into=c("Date_1","Time_1"),sep=' ')
```

3

# Loading and cleaning the data

- Load data, the separator for the data was tab or "\t".
- The dataset contains many columns that we might not be interested in. I was interested in the Common Name, Genus, Species Count, Location Name, Latitude, Longitude and observation date.

Loading data **Dataset head**

```
head(data,n =3)
```

```
##                                          V1      Date_1   Time_1    V3        V4
## 1 URN:CornellLabOfOrnithology:EBIRD:OBS61941867 2013-03-08 15:36:52 26621 species    Europ
## 2 URN:CornellLabOfOrnithology:EBIRD:OBS48090318 2014-01-22 16:28:47 21939 species Ruby-crc
## 3 URN:CornellLabOfOrnithology:EBIRD:OBS48090321 2014-01-22 16:28:47 27956 species     Pra
##                 V6 V7 V8 V9 V10 V11 V12         V13 V14     V15    V16    V17        V1
## 1    Sturnus vulgaris        3           NA United States  US New York US-NY  Wayne US-NY-11
## 2  Regulus calendula        1           NA United States  US New York US-NY Oswego US-NY-07
## 3 Setophaga discolor        1           NA United States  US New York US-NY Oswego US-NY-07
```

3

# Data Cleaning

Since there are no headers in this dataset, I assumed that the 9th column of the original dataset (which has been renamed as Species_Count) contains the number of observations for each species.

- I set the "Date" column with a date format, and make "Species_Count" column numeric.

- Where are these observations being made? And how many different species of birds there are in the dataset?

Country     State     Number of species

```
Country<-unique(data_clean$Country)
Country
```

```
## [1] "United States"
```

4

# Data Cleaning

Since there are no headers in this dataset, I assumed that the 9th column of the original dataset (which has been renamed as Species_Count) contains the number of observations for each species.

- I set the "Date" column with a date format, and make "Species_Count" column numeric.

- Where are these observations being made? And how many different species of birds there are in the dataset?

| Country | State | Number of species |
| --- | --- | --- |

```
State<-unique(data_clean$State)
State
```

```
## [1] "New York"
```

# Data Cleaning

Since there are no headers in this dataset, I assumed that the 9th column of the original dataset (which has been renamed as Species_Count) contains the number of observations for each species.

- I set the "Date" column with a date format, and make "Species_Count" column numeric.

- Where are these observations being made? And how many different species of birds there are in the dataset?

| Country | State | Number of species |
|---------|-------|-------------------|
|         |       |                   |

```
all_species<-length(unique(data_clean$Genus))
all_species
```

```
## [1] 408
```

4

# Quick data exploration

- All observations are done within the state of NY, 408 different species
- I chose 5 species with the highest number of counts in the dataset.

I grouped the data by *Genus*, and counted the number of occurrences. Then, picked the top 5 rows.The most common bird found in the dataset is the American Crow.

Results

```
## # A tibble: 5 x 2
##   Genus                  Count
##   <chr>                  <int>
## 1 Corvus brachyrhynchos  1205
## 2 Poecile atricapillus   1128
## 3 Turdus migratorius     1071
## 4 Cyanocitta cristata    1038
## 5 Zenaida macroura       1005
```

# Plotting the data

- I decided to plot the count per each type of bird using the {ggplot} and {maps} packages.

- Later, animated using the {gganimate} package to see how bird sightings change over time.

- The result is this animation: bird sightings have been recorded for crows, bluejays, robins, Black-capped Chickadees and Mourning doves.



Date:2006-05-05

American Crow
American Robin
Black-capped Chickadee
Blue Jay
Mourning Dove

· 1
● 10
● 100
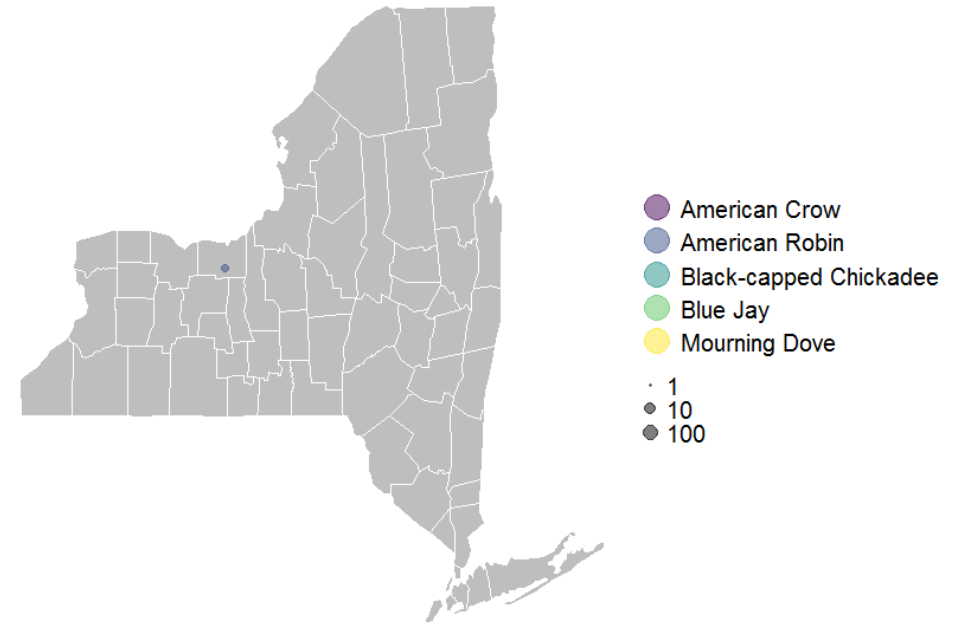
# Future work

- I would like to include more species in the analysis

- Also implement a 3-D map animation that shows where birds are being sighted with the inclusion of terrain data.