RCrab: An R Analytics Tool to Visualize and Analyze the Movement of Horseshoe Crabs in Long Island Sound

Ismael Youssef
Samah Senbel
Sacred Heart University, senbels@sacredheart.edu
Jo-Marie Kasinak
Sacred Heart University, kasinakj@sacredheart.edu
Jennifer Mattei
Biology Department, Sacred Heart University, matteij@sacredheart.edu

Follow this and additional works at: https://digitalcommons.sacredheart.edu/computersci_fac
Part of the Computer Sciences Commons, Population Biology Commons, Research Methods in Life Sciences Commons, and the Terrestrial and Aquatic Ecology Commons

Recommended Citation
RCrab: An R Analytics Tool to Visualize and Analyze the Movement of Horseshoe Crabs in Long Island Sound

Ismael Youssef
Valhalla High School
Valhalla, NY 10595
ismaely2002@gmail.com

Samah Senbel
School of Computer Science and Engineering
Sacred Heart University
Fairfield, CT, USA
senbels@sacredheart.edu

Jo-Marie Elisha Kasinak
Biology department
College of Arts and Science
Sacred Heart University
Fairfield, CT, USA
kasinakj@sacredheart.edu

Jennifer Mattei
Biology department
College of Arts and Science
Sacred Heart University
Fairfield, CT, USA
matteij@sacredheart.edu

Abstract—Mark-recapture programs are important for studying the ecology and population dynamics of wildlife. An R shiny analytics tool was developed to track the movement of horseshoe crabs in Long Island Sound based on tag and re-sight data. The crabs were tagged and recaptured by volunteers of Project Limulus, a community-based research program. The dataset contains tag and recapture location information for 14,065 horseshoe crabs over 18 years. The dataset was initially cleaned by removing records with missing, duplicate or incorrect data. A new data structure was developed to save the data and simplify processing: Three dimensions were used, one for the original horseshoe crab tag data, the second for the recapture data, and the third for the mating behavior data recorded both when originally tagged and when found during the recaptures. The R tools enables scientists and the general public to easily produce charts and movement maps based on the dataset. To study horseshoe crab movement, the Long Island Sound area was divided into five major tagging activity areas. Our results show that horseshoe crabs can cross the Sound (from CT to NY and back) and an interesting trend of movement towards the northern and eastern parts of Long Island Sound which correlates with less pollution and human disturbance.

Keywords—mark-recapture; movement analysis; R; shiny app; horseshoe crabs; wildlife ecology; data analytics

I. INTRODUCTION

Mapping the movement patterns of wildlife species is important to know for their management and conservation. Long-term mark-recapture programs where individuals in a study population are tagged released and then recaptured sometime later is one economical method for determining how individual animals move within and between local areas, migrate across long distances, and change movement patterns over time [12]. This information is used to address environmental challenges such as climate and land use change, effect of pollution, biodiversity loss, and the change in population size of different species [5].

Two main methods of tracking wildlife movement are practiced. The first uses electronic devices attached to animals such as radio transmitters (RFID) or the Global Positioning System (GPS) to track wildlife over large areas. This provides the ability to obtain high-resolution tracking data but is generally better suited for larger animals, short-term data collection that depends on battery life of the device and is relatively expensive [6]. The second method is the attachment of a non-electronic visible tag or mark. This simpler method, may be used on most species including those with smaller body size, use trained volunteers (citizen scientists) for reporting recaptures, with tagging programs that may span decades and are much less expensive.

Figure 1: Tagged Horseshoe Crabs in Long Island Sound

The mark, band (often used on birds) or tag is visible with a unique code or number that is attached to the animal. For our study species, the American horseshoe crab (Limulus polyphemus), the initial tagging and recording of basic data is a simple process and volunteers can easily participate after a short training period [8]. Often “recapture” is not necessary and the tag can be seen and reported with minimal handling or removing the animal from its habitat. The tag contains reporting information (phone number and email address) for volunteers who find a tagged crab (See Figure 1.). The volunteer reports the unique tag number and location as well as the status of the animal (i.e. dead or alive). The tags are lightweight, durable, very inexpensive, and can be easily attached by trained volunteers, without harming the animal [7]. Thus, large numbers of animals can be tagged over a very large geographical area. The majority of the tagged animals will not be seen again, however our overall recapture rate was 22% which gives us a fair indication of movement patterns over 18 years.

Project Limulus[8] is a long-running research program hosted by Sacred Heart University to promote the conservation of horseshoe crabs locally and worldwide through citizen science. People of all ages who live near Long Island Sound learn to tag and collect important data on horseshoe crabs living in the Sound. Hundreds of people
have volunteered their time over the 18 years. Adult horseshoe crabs come up on the beaches of Connecticut in mid-May through June to spawn and this is when they are tagged. During tagging the unique number is recorded, along with date, location, sex, number of mates and size. An enormous amount of tracking data has been generated over the years and a tool was developed to perform analytics and chart the horseshoe crabs’ movement based on this large dataset. This analysis will inform State wildlife managers about spawning beaches that may need protection from human activity and to generate insight and discover trends in long-term and short-term horseshoe crab movement.

II. RELATED WORK

A lot of commercial work as well as research advances has been made in the field of animal tracking and movement prediction. However, they all concentrate on tracking large animals fitted by RFID and satellite transmitters. The movement of individual animals are studied and used to predict a general pattern of behavior and movement of the group. Conversely, in our case, where thousands of small animals are tagged only ~20% ever get recaptured. However, the lack of detailed movement data about individual members of the species is made up by the sheer number of animals tracked and the long 18-year tracking period.

Teimouri[14] developed an R program to track the movement of large animals using radio frequency tracking. They study the distribution of events during the day: foraging, resting, or walking. They were able to successfully predict activity by 80% compared to observed behavior.

Etienne[3] developed an interesting intelligent system to predict the movement of mobile objects (animals or otherwise) using a successive feedback loop to learn and update predicted trajectories based on information as it arrives.

Signer[13] developed a R tool (amt) similar to ours that track the movement of small animals over time. However, they depend on having multiple locations for each animal to interpolate and analyze the possible trajectories of movement, although the small sample size makes it unpractical to generalize the results.

III. DATA CLEANING AND MODELLING

The data corpus used in this work is that data log of crab activity collected by scientists, biology students, and citizen scientists over 18 years. This data was in a large one table format with 60 fields and 21,245 records, organized by entry date. An entry was made whenever a crab was recaptured and includes information such as location, condition, mates, and size. There were 12 fields for the mates of the crab at tagging time, and another 12 for the mates found with that crab at recapture time. Most of these field were left empty.

A. Data Cleaning

We started by doing data cleaning on the dataset: Some horseshoe crabs had recapture data but no initial tagging location, rendering them useless in measuring movement, 461 records are removed for missing tagging information. Also, some horseshoe crabs (621 records) had only the tagging or recapture beach name but no other data. 157 of them were retained by copying the longitude/latitude data from another similar record, and 464 records were discarded.

The next step was to remove duplicate records from the data frame (1311 records). Several of the records (143 records) had recapture dates before the initial tagging date. This is probably a data entry error. These records are discarded as follows:

```r
(crabFiltered$Date2 - crabFiltered$Date1) >=0)
```

Several crabs were captured very far away from Long Island Sound, such that it is improbable that they swam there themselves, and were probably transported by humans. Therefore, the scope is limited to those crabs that were recaptured in the Long Island Sound vicinity: CT, NY and RI states.

There was a data entry error where several horseshoe crabs were entered as having two initial tagging locations (904 records). The confusion comes from having two locations as “Town beach”, one in Clinton, CT, and one in Old Saybrook, CT. All records were discarded to maintain the integrity of the data. Same problem happened for “West beach” as there were two locations, one in Westbrook, CT, and one in Stratford, CT (138 records):
Also, if the movement pattern is only of interest, then the third dimension about mates it was found with could be simply ignored.

In this paper, the movement of the horseshoe crabs only is of interest, therefore the third dimension about mates found with each horseshoe crab is ignored, as well as the data for the horseshoe crabs that were never recaptured. The first dimension list contains the following data for 14,065 horseshoe crabs:

1. **TagNumber**: A unique Tag Number per horseshoe crab
2. **Gender**: Observed gender of horseshoe crab
3. **Size**: Prosomal Width measured during tagging
4. **TagDate**: Date when the tag was attached
5. **TagLongitude**: Location of tagging
6. **TagLatitude**: Location of Tagging
7. **TagState**: State where the horseshoe crab was tagged
8. **Number of Recaptures**: default is one, if it is more than one, then a second dimension exists for this horseshoe crab.
9. **Distance**: Total Distance traveled in Km.
10. **Observation period**: difference between tagging date and last recapture.

The second dimension is about the individual crab recaptures. It contains a list of the following fields for all horseshoe crabs:

1. **RecapDate**: Date of recapture
2. **Recap Longitude**: Location of recapture
3. **Recap Latitude**: Location of recapture
4. **Condition**: Observed shell condition of the animal (ranked 1, 2, or 3)
5. **Recapture State**: State where the horseshoe crab was found
6. **List of mate ids** (third dimension)

The entire R code for data cleaning and modeling is found at [16] as an R markdown document.

IV. **Movement Analysis**

The software used to analyze the data is the R Programming language [10] and the RStudio development environment [11]. The R shiny library was used to create an online tool for data analysis and visualization. The ggplot2 library was also used for graph rendering and the leaflet library for map generation. The dataset used is the cleaned-up and organized data from the previous section.

The left part of the tool [9] is used for input widgets, and the right side for the output charts or maps. The user gets to choose which years to include in the analysis, from 1998 to 2015 using a slide-bar. Then, they can choose if they would like to generate charts or view maps about the horseshoe crab movement data. Figure 3 shows a snapshot of the R shiny app when charts are chosen.
The user specifies the field to be used for the x-axis, y-axis, and color using three pull-down menus: Distance Traveled, Observation Period, Initial Longitude, Initial Latitude, Final Longitude, Final Latitude, Number of Recaps, Gender, Tag Date, Recapture Date, Tag State, Recapture State, Tag Year, Recapture Year, Initial Tag region, and Recapture region. The corresponding scatterplot or histogram is plotted using the code:

```r
if (input$radio==1)
  p <- ggplot(dataset2(),
              aes_string(x=xinp, y=yinp))  + geom_point()
if (input$radio==2) {
  if (xinp=='Dist' | xinp=='ILong' |
      xinp=='ILat' | xinp=='FLong' | xinp=='FLat')
    p<- ggplot(dataset2(),
               aes_string(x=xinp, fill=xinp))
    + geom_histogram()
  else
    p<- ggplot(dataset2(),
               aes_string(x=xinp, fill=xinp))
    + geom_bar()
}
if (input$color != 'None')
  if (input$radio==1)
    p <- p + aes_string(color=cinp)
  else
    p <- p + aes_string(fill=input$color)
print(p)
```

For the mapping aspect of the shiny app, the Long Island Sound was divided into five regions where the most tagging and recapture activity was done. This simplifies the study of the movement of the horseshoe crabs in Long Island Sound. The map in figure 4 shows the five regions color coded as follows: Rye Beach in black, Milford in green, Sandy Point in red, Bluff Point in blue, and Cedar Beach in pink. Any horseshoe crabs outside of these five regions will belong to either “Long Island”, that covers areas in New York City and Long Island state other than Cedar Beach and Rye Beach, or “Rhode Island” to cover horseshoe crabs in Rhode Island state east of Bluff Point and just outside the Long Island Sound. The longitude and latitude data were used to label each location to one of the seven different regions. This divides our large dataset of 14,065 horseshoe crabs into smaller groups, thereby simplifying the visualization and movement analysis process.

```r
map <- leaflet()
map <- addTiles(map)
for(i in 1:nrow(d1)){
  map <- addCircles(map, lng=d1[i, "ILong"],
                   lat=d1[i,"ILat"],color="red", weight=1)
  map <- addCircles(map, lng=d1[i,"FLong"],
                   lat=d1[i,"FLat"],color="blue", weight=1)
  map <- addPolylines(map,
                      lat = as.numeric(d1[i, c("ILat", 'FLat')]),
                      lng = as.numeric(d1[i, c("ILong", 'FLong')]),
                      color="black", weight=1)
}
```

The entire R code for charting and mapping is found at [17] as an R markdown document.
V. RESULTS

In this section, some of results obtained from using our R shiny tool on the horseshoe crabs are presented. It is important to note that the number of horseshoe crabs tagged each year is different as shown in figure 6. This curve reflects the number of volunteers and the extent of recaptures activity made, and does not reflect on the actual number of horseshoe crabs over the years. Also, note that there were generally at least twice as many male horseshoe crabs as female ones over the years.

![Figure 6: Tagging and Recapture numbers over the years](image)

Also, the number of horseshoe crabs tagged or recaptured in each region depends on the volunteer’s and citizen scientist’s population size and activity in each region, and does not reflect the actual number of horseshoe crabs in each region. Table 1 shows the number of horseshoe crabs initially tagged or recaptured in each region. Figure 7 shows the gender distribution in the different regions. The region with the highest female ratio is Bluff Point (37%), and the region with the lowest female ratio is Rye Beach (21%).

![Figure 7: Tagging gender distribution](image)

<table>
<thead>
<tr>
<th>Region</th>
<th>Tagging</th>
<th>Recapture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bluff Point</td>
<td>3420</td>
<td>3159</td>
</tr>
<tr>
<td>Sandy Point</td>
<td>3618</td>
<td>3717</td>
</tr>
<tr>
<td>Milford</td>
<td>5412</td>
<td>5166</td>
</tr>
<tr>
<td>Rye Beach</td>
<td>1343</td>
<td>1167</td>
</tr>
<tr>
<td>Cedar Beach</td>
<td>116</td>
<td>318</td>
</tr>
<tr>
<td>Long Island</td>
<td>90</td>
<td>280</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>66</td>
<td>258</td>
</tr>
</tbody>
</table>

![Figure 8: Number of Recaptures](image)

Another observation is that the same horseshoe crab is usually recaptured only once, as shown in figure 8, with 79.4% of horseshoe crabs with one recapture. The maximum number of recaptures was 9 times for horseshoe crab number 182857.

In general, most of the horseshoe crab recapture data showed very little movement with 82% of observed total distance travelled is under 10 km distance and 46% as exactly zero even over long observation periods. Figure 9 shows the distribution of travel distance for all horseshoe crabs Figure 10 shows the numbers of days versus the distance traveled. It can be observed that the distance rarely went above 20 km even over several years. The vertical “bars” are due to the fact that most tagging and recapture activity is done in late spring and early summer of every year.

![Figure 9: Total distance travelled for all horseshoe crabs](image)

![Figure 10: Distance travelled versus Observation Period](image)
The following table shows the movement of horseshoe crabs between the different regions:

<table>
<thead>
<tr>
<th></th>
<th>Bluff</th>
<th>Sandy</th>
<th>Milford</th>
<th>Rye</th>
<th>Cedar</th>
<th>LI</th>
<th>RI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bluff</td>
<td>3038</td>
<td>137</td>
<td>24</td>
<td>8</td>
<td>8</td>
<td>18</td>
<td>187</td>
</tr>
<tr>
<td>(3420)</td>
<td>(89%)</td>
<td>(4%)</td>
<td>(7%)</td>
<td>(2%)</td>
<td>(2%)</td>
<td>(5%)</td>
<td>(5.4%)</td>
</tr>
<tr>
<td>Sandy</td>
<td>69</td>
<td>3156</td>
<td>299</td>
<td>24</td>
<td>30</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>(3618)</td>
<td>(2%)</td>
<td>(87%)</td>
<td>(8.3%)</td>
<td>(7%)</td>
<td>(1%)</td>
<td>(9%)</td>
<td>(1%)</td>
</tr>
<tr>
<td>Milford</td>
<td>38</td>
<td>391</td>
<td>4696</td>
<td>99</td>
<td>127</td>
<td>59</td>
<td>2</td>
</tr>
<tr>
<td>(5412)</td>
<td>(1.7%)</td>
<td>(7%)</td>
<td>(87%)</td>
<td>(2%)</td>
<td>(2.3%)</td>
<td>(3%)</td>
<td>(0%)</td>
</tr>
<tr>
<td>Rye</td>
<td>9</td>
<td>22</td>
<td>126</td>
<td>1033</td>
<td>59</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>(1343)</td>
<td>(.7%)</td>
<td>(1.6%)</td>
<td>(9.4%)</td>
<td>(77%)</td>
<td>(4.4%)</td>
<td>(6.9%)</td>
<td>(0%)</td>
</tr>
<tr>
<td>Cedar</td>
<td>4</td>
<td>7</td>
<td>15</td>
<td>0</td>
<td>81</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>(116)</td>
<td>(3.4%)</td>
<td>(6%)</td>
<td>(13%)</td>
<td>(0%)</td>
<td>(70%)</td>
<td>(7.7%)</td>
<td>(0%)</td>
</tr>
<tr>
<td>LI</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>70</td>
<td>1</td>
</tr>
<tr>
<td>(90)</td>
<td>(0%)</td>
<td>(3.3%)</td>
<td>(6.6%)</td>
<td>(3.3%)</td>
<td>(7.8%)</td>
<td>(78%)</td>
<td>(1%)</td>
</tr>
<tr>
<td>RI</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>63</td>
</tr>
<tr>
<td>(66)</td>
<td>(1.5%)</td>
<td>(1.5%)</td>
<td>(0%)</td>
<td>(0%)</td>
<td>(0%)</td>
<td>(1.5%)</td>
<td>(95.5%)</td>
</tr>
</tbody>
</table>

It is clear from the table that horseshoe crabs from Rhode Island rarely travel west to the Long Island Sound, only 3(4.5%) did, while 95.5% remained in Rhode Island. Similarly, only 195(14%) horseshoe crabs from different areas of the Long Island Sound end up going there. 66 horseshoe crabs were tagged there, but 258 were recaptured there.

Also, horseshoe crabs from the northern shores of Long Island Sound seem to be fairly static as well, with roughly 87% of them staying in the same region, and those who do travel tend to stay in the three northern LIS regions: Bluff Point, Milford, and Sandy Point, or move east to Rhode Island. They rarely traveled west or south.

Horseshoe crabs from the western & southern part of Long Island Sound clearly move more than their northern and eastern conspecifics, with relatively smaller percentages staying in their region: Cedar Beach 70%, Rye Beach 77%, and Long Island 78%.

We examine the movement of horseshoe crabs tagged in the West/South of Long Island Sound (Rye Beach + Cedar Beach + Long Island) against those of the North (Milford + Sandy Point + Bluff Point): 12,450 horseshoe crabs were tagged in the northern three regions: 11,848(95.2%) stayed there and 409(3.3%) went to the western and southern regions. 1,549 horseshoe crabs were tagged in the western and southern parts: 1,355(87.5%) stayed there and 192(12.4%) went to the northern part of Long Island Sound.

VI. CONCLUSION

In this paper, an R shiny tool was developed to study the movement of horseshoe crabs in Long Island Sound. The dataset of Project Limulus which tracks horseshoe crabs in the area for 20 years was used. The ggplot2 library was also used to plot charts about the different aspects of horseshoe crab movement, and the leaflet library to map the actual movement trajectories. The tool is presented as a Shiny app to simplify access to the dataset. Our analysis shows an interesting trend in movement: horseshoe crabs are mostly static, but when they do move, there is a tendency to move towards the northern part of Long Island Sound.

VII. ACKNOWLEDGEMENTS:

We thank the US Fish & Wildlife Service for their support with the tagging program, CT Sea Grant, Disney Conservation Fund, The Nature Conservancy, Audubon CT, and CT Audubon Society, Sacred Heart University undergraduate Biology Majors, and the College of Arts and Sciences for funding and time supporting this project.

References
