



DYNAMIC VISUALIZATION OF MOSQUITO LANDING RATES IN LEE COUNTY, FL

Andrew Krupp and Dr. Alberto Condori
Florida Gulf Coast University



Introduction

In an increasingly data-driven industry, a common problem facing many teams (research, business, or otherwise) is taking large amounts of data and making it digestible for those outside of the team working with the raw data. The amount of data that needed to be processed by the team was so large that it necessitated the development of algorithms to automate the cleaning and simplification of said data in order for the research to continue in a timely manner.

Goal: take tens of thousands of cells of data and turn it into something digestible by the research team, the industry partner, and the public at large.

Raw Data

EXAMPLE DATA. THIS IS NOT ACTUAL RECORDED DATA												
ID	Inspector	Zone Num	Date	Wind Direc	Wind Spee	Temperatu	Tide Level	Landing Re	Dip Count	Species		
555555	John Doe	555-55	Jan --, ----	West	10	76		5		**** Spp **** Spp *		
555555	John Doe	555-55	Jan --, ----	West	10	76		8		**** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		4		**** Spp **** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		10		**** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		14		**** Spp **** Spp *		
555555	John Doe	555-55	Jan --, ----	West	10	76		6		**** Spp **** Spp *		
555555	John Doe	555-55	Jan --, ----	West	10	76		7		**** Spp **** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		9		**** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		12		**** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		13		**** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		40		**** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		3		**** Spp		
555555	John Doe	555-55	Jan --, ----	West	10	76		2		**** Spp **** Spp *		
555555	John Doe	555-55	Jan --, ----	West	10	76		1		**** Spp **** Spp		

Raw data is often times unworkable because there are issues with the way the data was recorded, so much data that it is hard to comprehend what any of it means.

Difficulty with Raw Data:

- Multiple entries per cell
- Lack of precise location of certain locations referenced
- Missing entries
- Imprecise measurements

Prior to the development of the visualization algorithm, multiple other algorithms were developed for the purposes of cleaning the raw data for use.

Acknowledgment

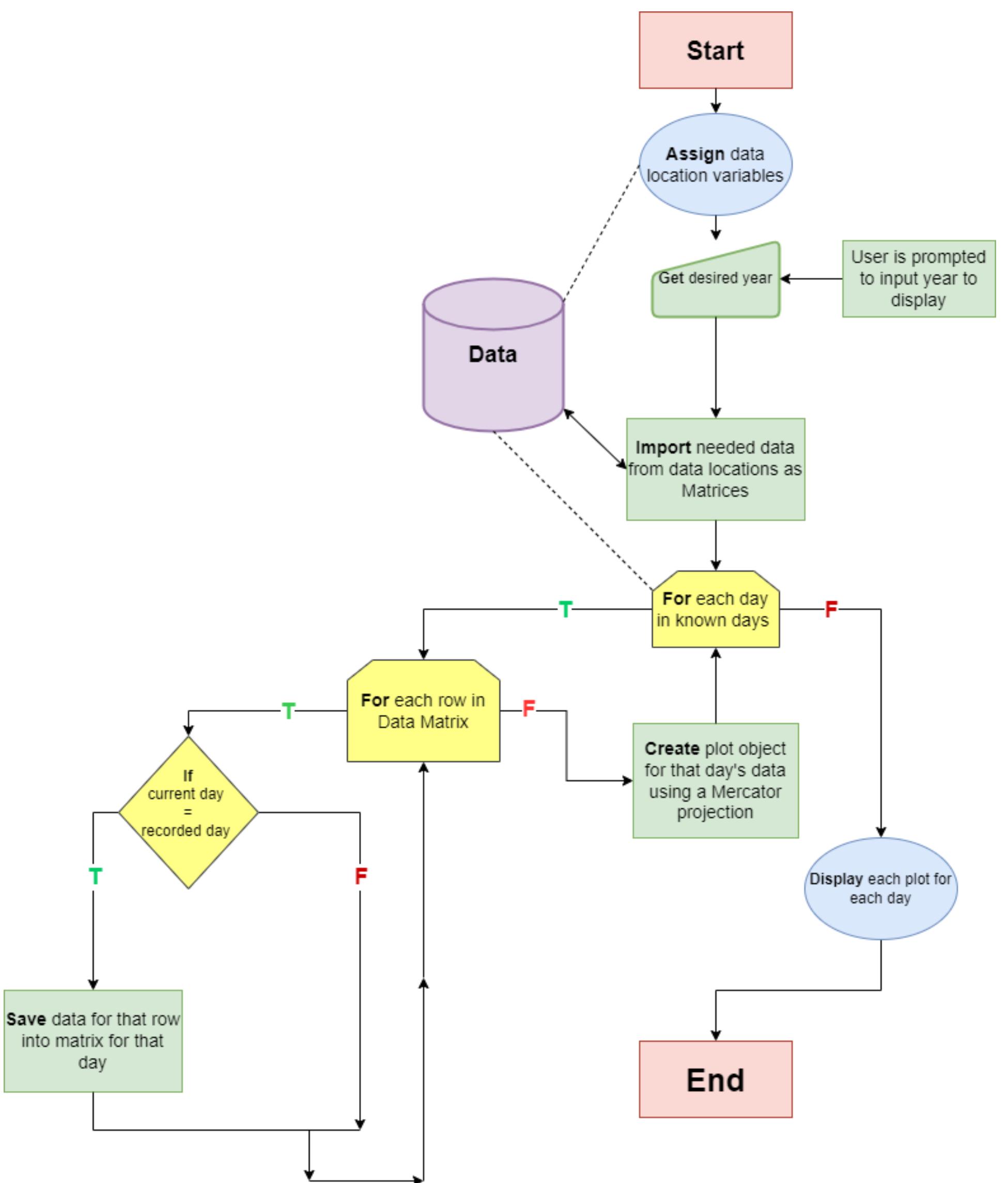
A special thanks to Lee County Mosquito Control District for funding the research. Also, a special thanks to Dr. Alberto Condori for his support in the development of this project and continuing support in my academics.

Development

Developing the algorithm to take the tens of thousands of rows of raw data the researchers were given by LCMCD took multiple steps. Development of any source code is not a straight shot from idea to written code, there are always intermediary steps from idea to final product. In this case, the following were needed for successful development:

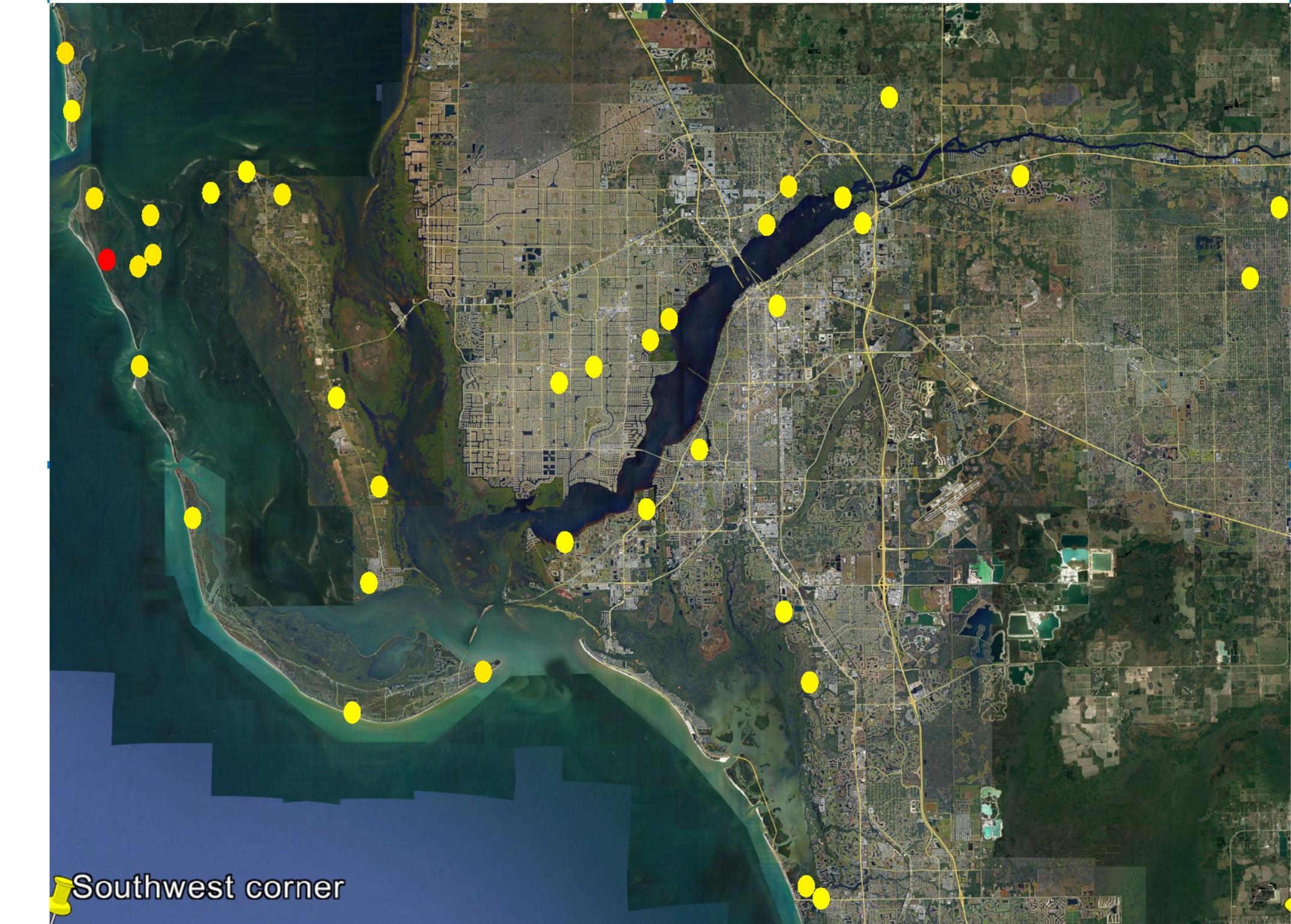
1. **Planning:** Identification of development requirements
2. **Research:** the Mercator projection
3. **Pseudocode:** A top-down outline of what the code needs to do in order to complete the task
4. **Development:** coding, testing, and debugging in Maple
5. **Outcome/Product:** Display of Landing Rates for one single day
6. **Refinement:** Modify our work to display all zones for all days in one year

After these phases, the algorithm that produces the final product (simplified graphically for ease of understanding) is as follows:



Visualized Data

This is a single frame from what the developed algorithm produces.



Key:

- Red = Highest recorded Landing Rate for that year
- Yellow = Lowest recorded Landing Rate for that year

Conclusions

Only once the visualization and data cleaning were completed, the team could begin building the model requested by LCMCD. The visualization of the data not only allows for the other members of the team working on difference facets of the project to better understand the data, but also facilitates a deeper understanding of the data itself. This is the most powerful tool of data visualization: ease of understanding. With the data visualized, conclusions can be drawn from the data that would not be possible if the data was still in text form only.

Completion of the model will allow LCMCD to predict the location of WNV activity and subsequently provide treatments to avoid an outbreak providing a benefit to the county and Southwest Florida at large.

References

- [1] 'Online Help' Background Images - Maple Help, Maplesoft, <https://de.maplesoft.com/support/help/maple/view.aspx?path=updates%2FMaple18%2FBac>
- [2] 'Earthquake Data.' Latest Earthquakes, USGS, <https://earthquake.usgs.gov/earthquakes/> 72.6855&settings=true.
- [3] 'Parcels with Ownership and Site Attributes.' IIS7, Lee County Property Appraiser, 10 Feb. 2020, <http://leegis.leegov.com/FTPData/FGDBs/LandRecords/ParcelData.htm>.