GIS 5572 - Semester Project
Title: Tracking the Spread of Brown Marmorated Stink Bug (BMSB) Populations in Minnesota
Name 1: Eric Gibson
Name 2: Zain Ul Abdin Siyal
Date: 4/25/2024
Project Github repository of all code:
https://github.com/Zain1443/BMSB_Project/tree/main/Final%20Project

Link to Final ArcOnline MapViewer:
https://umn.maps.arcgis.com/apps/mapviewer/index.html?webmap=0a9b6687bb9a46cba1f609948d5517ca

Link to Video Overviewing Each Aspect of Your Project:
https://github.com/Zain1443/BMSB_Project/blob/main/Final%20Project/Final%20Project%20Video.pdf

Link to Your Lab 4 Peer Reviews:

https://docs.google.com/document/d/1q_vfPBbi_qlkM_MShpx0B-1JFdtPf3gJ/edit

https://docs.google.com/document/d/14VSRReE5NkdwyTFDL-c3gAuhpxlcNv0bWaru46gp9NU/edit

Abstract (50 words)
This project aims to track the spread of BMSB populations in Minnesota. Using three Huff models with varying alpha values, we simulate the flow of BMSB from a seed city and rank cities where BMSB is prevalent. Following quality assessment, results will be made accessible via an API for integration into ArcGIS Online Map Viewer.

Problem Statement (100 words max)
The introduction of BMSB poses significant risks to Minnesota's agricultural crops, landscape plants, and residential environments. To mitigate potential impacts, it's crucial to track BMSB populations and rank cities based on prevalence. This project employs spatial interaction modeling techniques to simulate BMSB spread, facilitating MNDNR in informed decision-making and resource allocation for pest management efforts.

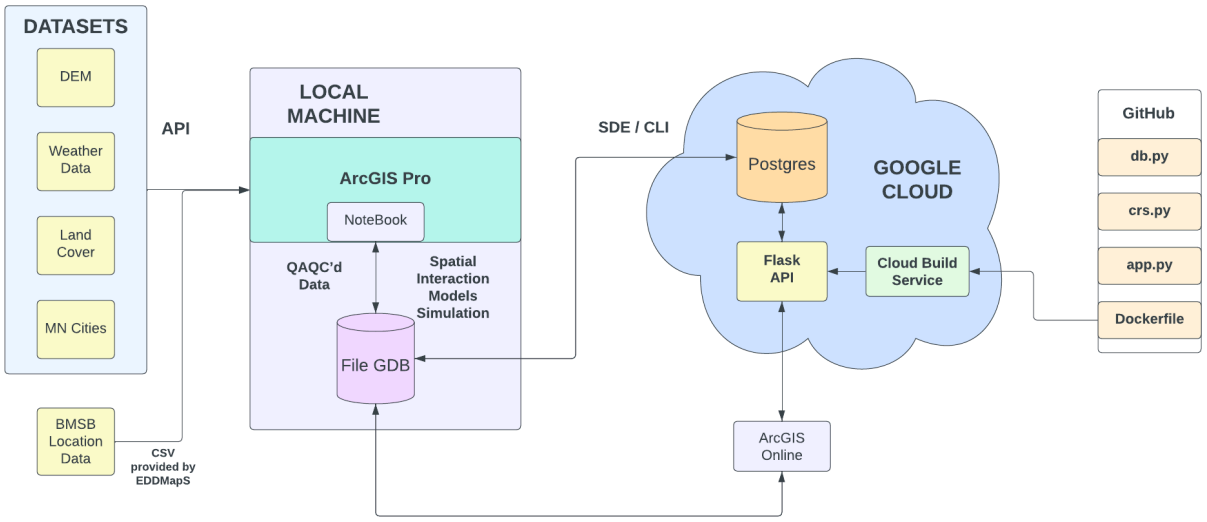
Input Data

#	Title	Purpose in Analysis	Link to Source
1	City, Township, and Unorganized Territory in Minnesota	Provides boundaries of cities in Minnesota, essential to get urban centers along with their population.	Minnesota Geospatial Information Office
2	BMSB Survey Data	Contains survey records from 2014 to present for BMSB in Minnesota, including essential data fields for	MN Agriculture Department

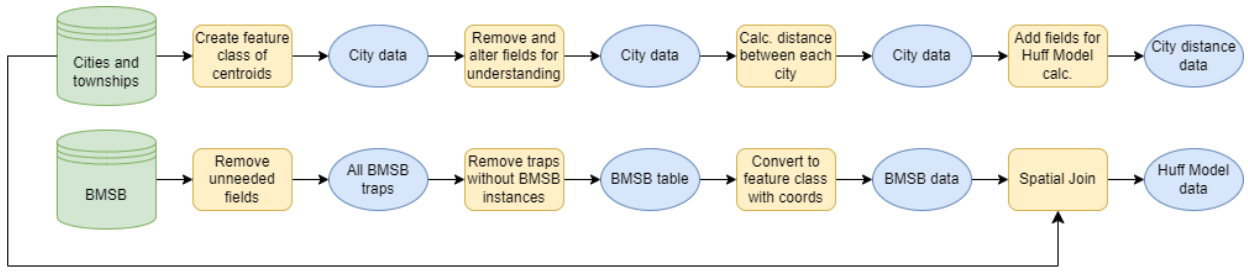
		analysis such as latitude, longitude, and number of invasive species detected.	
3	Weather data: The Iowa Environmental Mesonet	Supplies environmental data used for analyzing weather patterns and their potential impact on BMSB populations.	Iowa Environmental Mesonet
4	NLCD 2019 Land Cover, Minnesota	Offers land cover data for Minnesota, aiding in understanding BMSB habitat suitability and potential migration routes.	MN Natural Resources Department
5	Minnesota Digital Elevation Model - 30 Meter Resolution	Provides detailed elevation data for Minnesota, assisting in analyzing terrain characteristics influencing BMSB habitat suitability.	MN Natural Resources Department

Data Flow Diagram(s) for System (0 words)

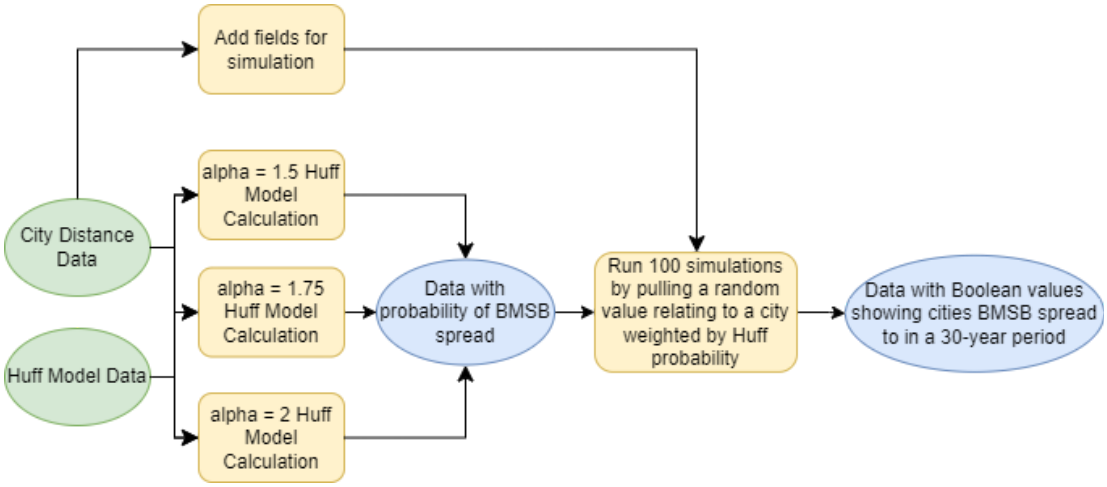
This should be extremely detailed and characterize the full end-to-end behavior of your system.



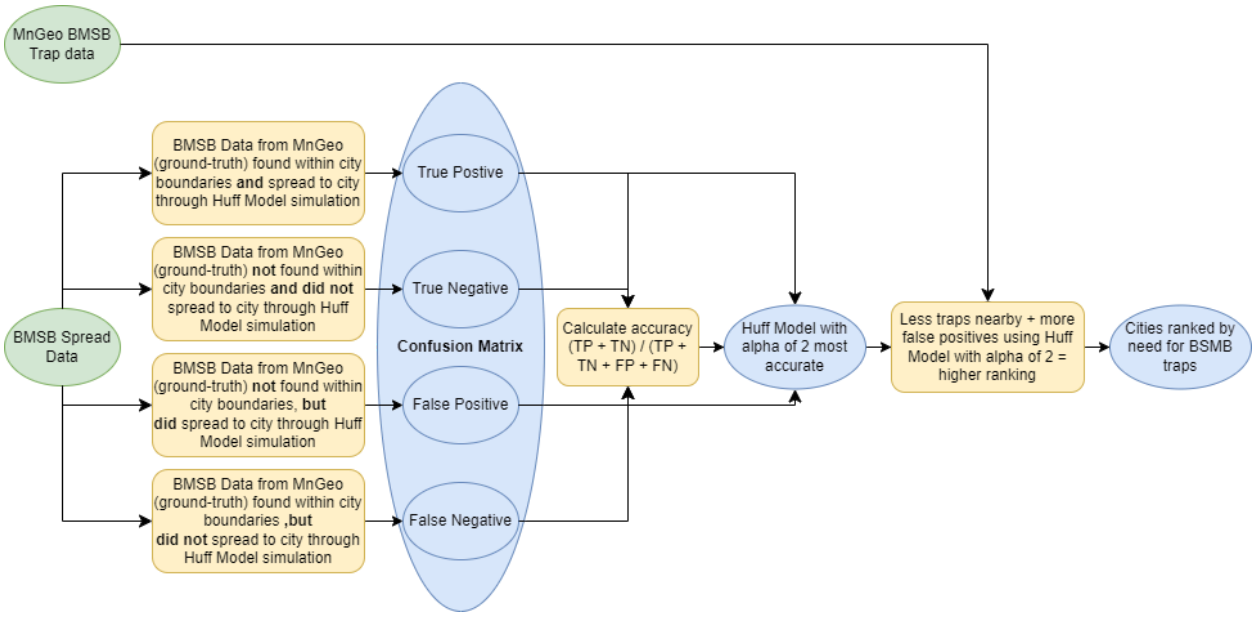
<u>QA/QC Data Flow:</u>



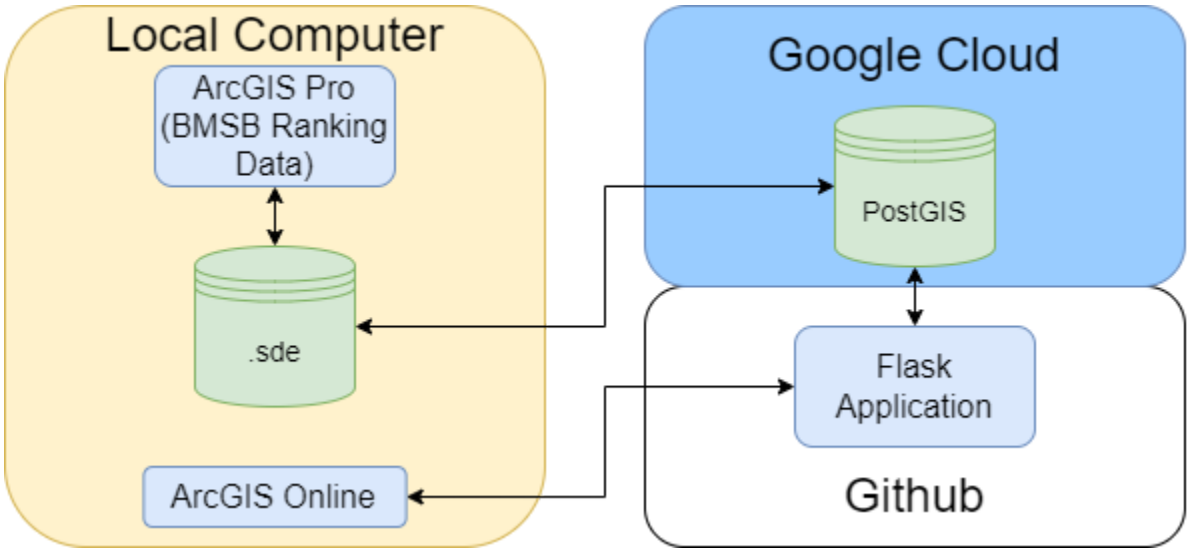
Simulation:



Accuracy Assessment



Move to SDE Database and display as ArcGIS Online Web Map



Model Comparison

#	Model Name	Mean Accuracy	Mean True Positives	Mean False Positives	Rank Score of Model
1	Huff Model with an alpha value of 1.5	0.867	3.776	12.077	3
2	Huff Model with an alpha value of 1.75	0.874	3.522	11.055	2
3	Huff Model with an alpha value of 2	0.881	3.432	10.285	1

Huff Model with an alpha value of 1.5

The Huff Model with an alpha value of 1.5 estimates the probability of BMSB movement from a seed city to neighboring cities based on attractiveness. It considers both the distance and attractiveness of neighboring cities, with attractiveness defined by population size.

Huff Model with an alpha value of 1.75
The Huff Model with an alpha value of 1.75 estimates the probability of BMSB movement from the initial seed city with more weight given to cities which are nearer to other cities with BMSB presence. This provides a middle-ground between the Huff Model with an alpha value of 1.5 and the Huff Model with an alpha value of 2.

Huff Model with an alpha value of 2
Similar model just that we are using alpha as 2 but it enhances the influence of distance on movement probability compared to the alpha value of 1.5 and 1.75. The probability of BMSB movement is calculated using a similar equation as the alpha 1.5 and 1.75 models but with a higher emphasis on attractiveness based on distance (the nearer, the more likely it is for BMSB instances to spread there).

Recommendation for Decision-making (50 words)
What model do you recommend be used for decision-making and why?

The model recommended for decision making in the Huff Model with an alpha value of 2 due to the average accuracy of the model's simulation runs. While this model does have the lowest number of true positives, it also has the lowest number of false positives. This, along with the highest average accuracy determine this to be the best model of the ones tested.

Reflection: What did you learn? What would you do differently if you did this again?

Through this project, we learned how to calculate a Huff Model and run spread simulations based on temporal increments. Although we previously learned how to conduct quality assurance and quality control in Lab 2, this final project cemented our understanding and was great practice especially for QA/QC of vector data. Through arcpy's cursor functions, we learned how to calculate spread probability based on the Huff Model equation and run weighted, randomized simulations of timesteps to visualize spread in an ArcGIS Jupyter Notebook. Finally, we learned how to run different models and compare them through a confusion matrix via accuracy, true positive, and false positive values.

There are some things which may have worked better than what was performed within this project. Although this was just inherent with the dataset, other confusion matrix analyses did not work, like precision, recall, and F1 score because anywhere which had true positives could not have false positives and anywhere which had true negatives could not have false negatives. This could be the case in a dataset where the understood ground truth points were incorrect, but since they were initially spatially joined with the data, the precision values were either 1, 0, or null and the recall values were either 1, 0, null, or the same as the accuracy. This did not

describe much about the accuracy of each model, so this was abandoned and accuracy, true positive, and false positive values were used instead.

Additionally, while ArcGIS Pro allows for fairly easy QA/QC and movement of data to a SDE database, the model itself and especially the simulation I think would have been more efficient in a Jupyter Notebook outside of ArcGIS Pro using pandas. The simulation was fairly inefficient and used a lot of processing power utilizing search and update cursor from arcpy. Simply for saving more time and battery, it would have been beneficial to move to an online Jupyter Notebook, run the simulation there, then convert the data back to an ESRI feature class to be easily added to our cloud geodatabase.

References

- Early Detection & Distribution Mapping System. (2024, January 31) EDDMaps Data Downloads. Retrieved from https://www.eddmaps.org/tools/query/
- Minnesota Geospatial Information Office. (2024, February 7). MnGeo Commons Documentation. Retrieved from https://gisdata.mn.gov/dataset
- Google Cloud. (2024, February 1). Cloud SQL for PostgreSQL Documentation. Retrieved from https://cloud.google.com/sql/docs/postgres
- Agency for Toxic Substances and Disease Registry. (2024, February 07). CDC/ATSDR SVI Data and Documentation Download. Retrieved from https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html
- Iowa State University Mesonet API for Daily Weather Data: URL: https://mesonet.agron.iastate.edu/api/1/daily.geojson?network=MN_RWIS&month=2&year=2024
- ESRI. "ArcGIS pro Geoprocessing Tool Reference." *ArcGIS Pro Geoprocessing Tool Reference-ArcGIS Pro | Documentation*, 2023, pro.arcgis.com/en/pro-app/latest/tool-reference/main/arcgis-pro-tool-reference.htm.
- Lucidchart. (2024, February 07). Lucidchart. Retrieved from https://www.lucidchart.com/

Appendix (include your original spec)

Solution Requirement Specification:

SRS Template

Software Developer	Zain Ul Abdin Siyal	MGIS
Data Engineering Lead	Eric Gibson	MGIS
Required Approvers	Shaul	Minnesota Department of Natural Resources.

Overview

The Minnesota Department of Natural Resources (MNDNR) recognizes that the movement of the Brown Marmorated Stink Bug (BMSB) is a large problem for agriculture and a nuisance for homeowners throughout the state of Minnesota. With the limited resources available, the MnDNR deemed it necessary to visualize and rank cities and townships by the number of BMSB found within the area. For MnDNR, this prioritization is crucial for optimizing resource allocation and strategically deploying traps to capture and eradicate these invasive stink bugs.

I. Brown Marmorated Stinkbug Map
   A. Ranking
      1. Will be determined by the density of stink bugs found per square foot in each city or town
         a) The greater the ratio, the higher the ranking
      2. Ranking of 1 to 5 with 5 indicating the highest priority for attention and intervention.
   B. Updating
      1. The map should be updated daily at the best with real time data fed through the data pipeline
   C. Design: BMSB Population Density Visualization
      1. Map should display top five cities/townships of BMSM population density through bar chart*
      2. Symbology includes points color coded from blue to red*
      3. Pop-up should include exact population of stink bugs

<u>Problems</u>

- **Public data:** Data from the general public tends to favor more densely-populated areas. A potential solution to this is inversely weighting areas based on population density. Another potential issue is accuracy of data both geographically and through attributes.
- **Data collection costs:** Data collection will include the gathering of live data for the 3-month extent of the project utilizing Google Cloud.
- **Client understanding of map functions:** The project will be performed to maximize functionality, while minimizing confusion for those less technologically inclined. Symbology will be coherent and clearly illustrated within the legend on the map for easy visualization of data. Any user input location on the interface will be created with obvious visual presentation and input type information.
- **Software:** The software generally used to perform this project provides functionalities which can handily satisfy the requirements for creating the deliverables, however there are instances where ESRI software can be subject to crashing and slow processing. This could potentially lead to the delay of deliverables to the client.

<u>Who it directly affects</u>

- Agriculturalists
- Technicians
- MNDNR
- General public

Motivation

Agriculturalists have developed the need for live data in recent years to tend to crops, determine growing seasons, and treat pests. With different years and seasons providing unique weather and climate patterns, it is important to provide the most recent information regarding potential issues with agriculture. For this request, the MNDNR is specifically focusing on moderating or eliminating the BMSM due to the species effect on agriculture throughout the state of Minnesota.

The outcome of this request is expected to be a map presenting real-time data of Brown Marmorated Stinkbug populations in cities and townships throughout the state of Minnesota. The webmap will be viewed by farmers, corporations, and non-profit organizations to determine where large populations of the stink bug species is located as well as areas to place stink bug traps to remove the species. These funding groups are our design partners for this project.

Definitions

Real-Time: Data is constantly updated and processed once every day

BMSB: Brown Marmorated Stinkbug

MNDNR: Minnesota Department of Natural Resources

Webmap: A spatial map created online to update in real-time.

API: Application Programming Interface; Any software which acts as a “mediator”, effectively allowing two programs to communicate with each other.

ETL: Extract, Transform, Load; A pipeline constructed for gathering, analyzing, and presenting real-time data from a specific API.

LULC: Land use, land cover; Spatial data depicting different classifications of land types, such as forests, farmland, urban areas, etc.

DEM: Digital Elevation Model provides information about the elevation of the earth's surface. It can be useful for analyzing terrain characteristics that may influence BMSB habitat suitability, such as slope, aspect, and elevation.

Scope

Functional Requirements

I. Software
   A. ESRI GIS tools
      1. ArcGIS Online
         a) Map will be hosted here
      2. ArcGIS Pro
         a) advanced spatial analysis and data wrangling tasks will be done here
      3. ArcGIS Web Applications
         a) Dashboards
            (1) Provides easy-to-use and easy-to-create option
         b) Experience Builder
            (1) Slightly more advanced than Dashboards
            (2) Likely to be used for visualizing map and data
            (3) Likely less easy-to-use, but can provide more information
         c) Both
            (1) Can be embedded into website
            (2) Can utilize graphs to display quantitative data
            (3) Can utilize pop-ups and lists for qualitative data
   B. Google Cloud
      1. Set up a domain for hosting the final web application
      2. Can be expensive depending on instance requirements
      3. Allows for access to data within web application for other utilizations

- Essential: The end product must include a webmap of Minnesota highlighting counties or cities vulnerable to BMSB spread.
- Essential: The product should provide a ranked list of counties or cities based on the risk of BMSB infestation.
- Essential: The product should auto update daily based on real time data fed through data pipeline
- Nice to Have: Data visualization tools such as charts and graphs to present BMSB population trends over time
- Optional: Feature to set up user-defined alerts for specific risk thresholds.

Non-Functional Requirements

- Essential: The product must be compatible with commonly used web browsers and devices to ensure easy accessibility.
- Nice to Have: The product must have a user-friendly interface for ease of use by operators.
- Nice to Have: The system should be scalable to accommodate potential increases in data volume and user traffic without significant performance degradation.
- Optional: Implement data encryption protocols to avoid hacking.

Out of Scope Requirements

- Advanced predictive modeling feature that can rank BMSB spread based on historical and real-time data.
- Multi-language user interface.
- Completely human population unbiased data.

Persona Acceptance Criteria

As a I developer I

- Require easy-to-use, high-functionality software so that I can create an informative geographical representation of BMSB numbers within Minnesota.
- Require software which can illustrate patterns in data so that I can inform end users of specific quantitative and qualitative spatial and non-spatial data trends regarding BMSB populations.
- Require software which can handle large volumes of data and user traffic so that I can perform efficient analysis of BMSB numbers for creation of deliverables.
- Require good data so that I can create an accurate visual representation of BMSB densities throughout Minnesota.

As an Operator I…

- Require real-time data updates so that I can stay informed about the latest BMSB population trends and respond promptly to emerging situations.
- Require access to historical data so that I can analyze trends in BMSB populations over time, aiding in long-term research and understanding of infestation patterns.
- Require accurate information for weighting human populations so that BMSB populations can be accurately represented on the web application.

As an end user I

- Require user-friendly Interface so that I can easily visualize and gain insights about the BMSB population without any prior technical knowledge.

- Require historical comparison tools so that I can compare current BMSB densities with historical data, assisting me in predicting potential future infestation trends and planning accordingly.
- Require information regarding potential harm from BMSB so that I can gain insight on issues which could be occurring in specific areas of Minnesota.

Open Questions

- Should all townships within Minnesota be shown? If not, what is the human population limit at which the towns should not be shown (<10000 people for example)?
- How should the ranking be formatted? (1 - 5, Low to High, etc.)
- Is there any schema for what to display on map or in the pop-up?
- Should map be divided into census tract, county, town, etc.?
- Should the data come from EDDMapS? Public data is biased towards densely-populated areas.
- What risks are associated with implementing a real-time data pipeline for daily updates?
- How will we implement robust data validation and quality control measures for the BMSB Population Visualization System? (garbage in, garbage out phenomena)

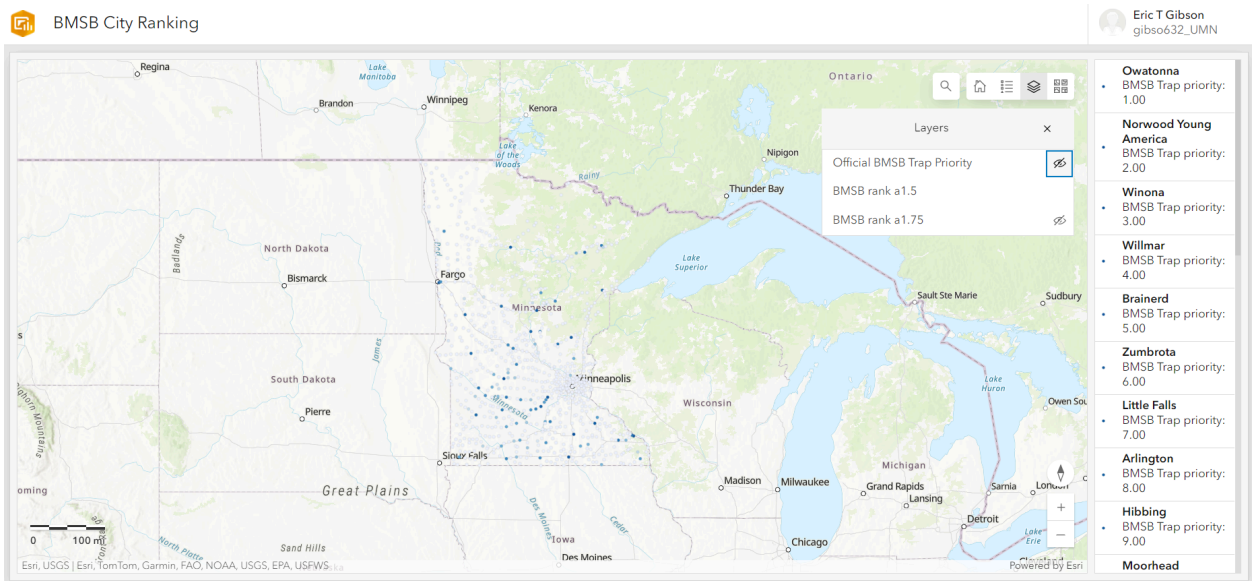
Dependencies

The quality of the data regarding biases towards areas with greater population density could provide incorrect information for people within the agricultural industry and those performing studies on BMSB numbers.

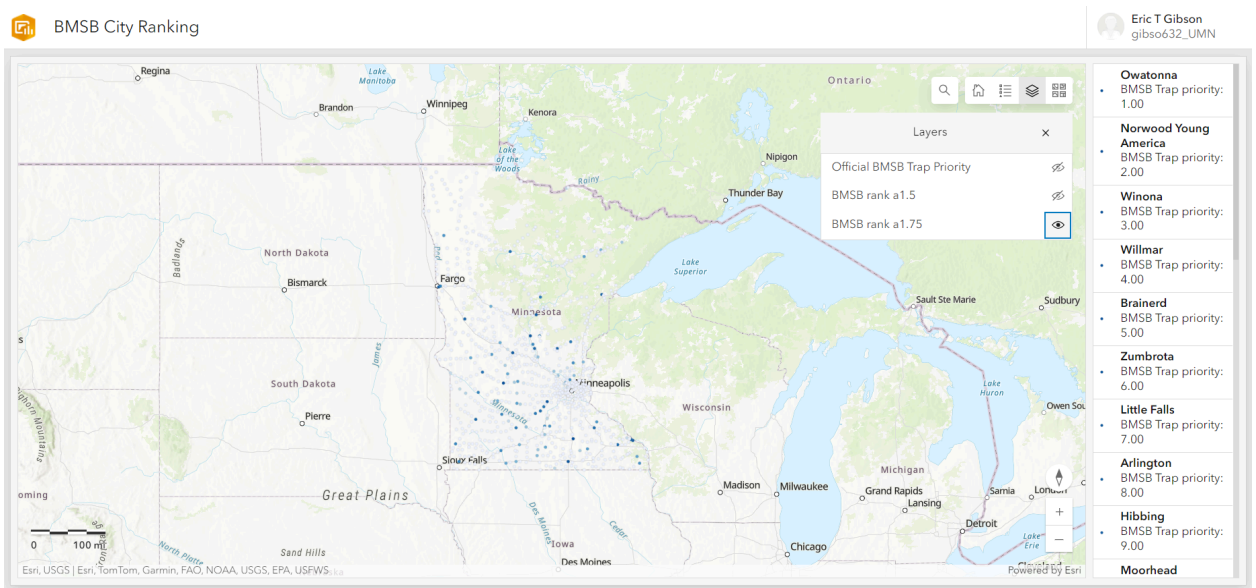
- ESRI Webmap
- MNDNR
- MN State/County Governments
- Farmers
- Entomologists

Final Output Maps (ArcGIS Dashboard):

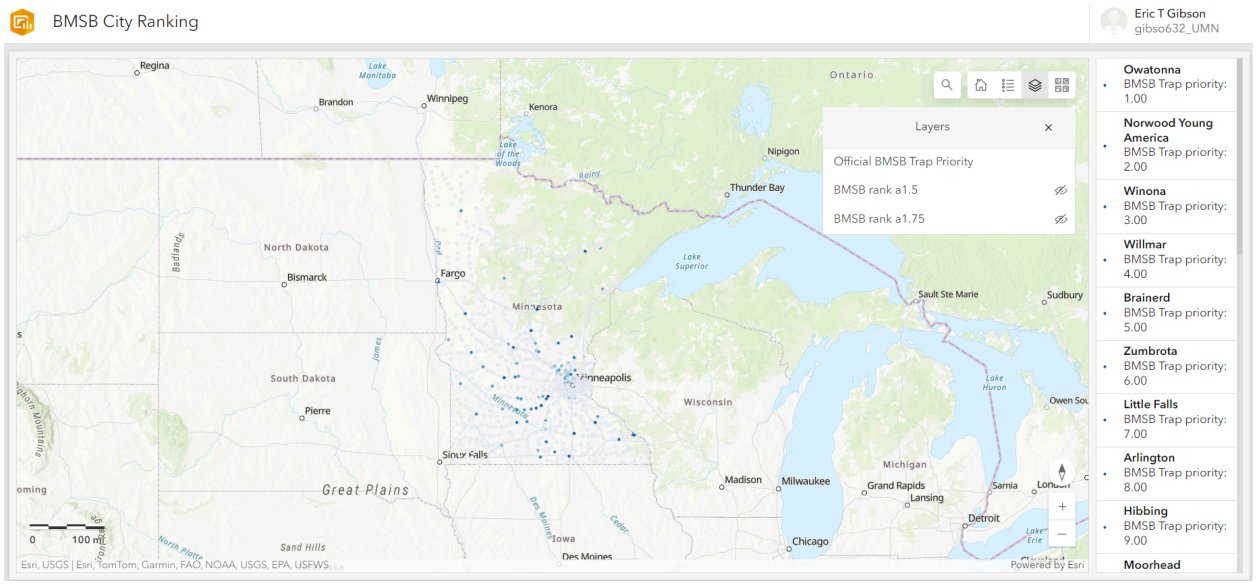
Huff Model where alpha = 1.5:



Huff Model where alpha = 1.75:



Huff Model where alpha = 2:



List on right-hand side shows the true BMSB trap priority based on the most accurate Huff Model used (alpha = 2).