The Hawk Member Management System

The last member address system for watershed protection agencies

Cameron Palmer
Tech Guru
Friends of the Cobbossee Watershed

July 29th, 2021

Background and Purpose

It was June 2021 when my boss at the Friends of the Cobbossee Watershed, Tamara Whitmore, brought me a project - to create maps of our Cobbossee Watershed containing address, member, ownership, and other information about waterfront residences. She presented me with a document containing a project description, as well as previous email chains with representatives from lake organizations attempting similar projects. In each of these correspondences, there was an overwhelming theme: manual data entry. As the FOCW's Tech Guru, the very thought of manual data entry makes me shudder.

Each of these organizations has their own method of handling data, mostly using spreadsheet/database software such as Microsoft Excel or Access. Some use more sophisticated softwares, such as Little Green Light, a member management system designed for non-profits. However, each of these methods require tedious manual data entry. This consists of an individual (or individuals) poring over thousands of lines of data, searching for the specific entry they're after. As those who suffer through the task of manual data entry know, this method is time consuming and inefficient. For example, one account in the previously mentioned correspondences details two full time individuals hired for the sole purpose of updating and maintaining a Microsoft Access database of member data - approximately 12,000 records with 140 attributes each. That's almost 1.7 million pieces of information. It takes these two individuals the entire winter to update a single town or lake. In an effort to solve this problem, the Hawk Member Management system was designed.

Theory

The Hawk Member Management System (hereafter referred to simply as "Hawk") attempts to completely eliminate the need for manual data entry, matching, and modification when dealing with member address data. In order to understand the theory behind the system and how it works, we must first understand the main problem faced when assimilating this sort of data.

The Main Problem: Waterfront Properties and their Billing Addresses

The most important and significant problem faced by lake protection nonprofits is money. Gathering funds typically comes in three forms: government grants, fundraising from events, and donations. The latter is often where organizations could improve their numbers. Donations typically come from those who have a personal investment in the water quality of the watershed - that is, largely, waterfront property owners. The question is then raised: how do we tell these potential donors about our organization, especially during the winter months when they travel elsewhere? This is the main question Hawk answers.

Gathering Data and The Role of the Government

In order to solicit donations from waterfront property owners and spread the word about the organization, we'll not only need these owners' local watershed addresses, but also their billing addresses, which could be different. That means we're going to need data - a lot of it: in the Cobbossee Watershed alone, there are nearly 20,000 addresses. The first problem we run into is obtaining this address data. Luckily, there's an online database called OpenAddresses[1] that contains addresses and their coordinates (coordinate data will prove useful later) from around the world, free for use for all. Alas, this database only contains street addresses and coordinates; owners' names are not included. In fact, a universal database containing addresses and owner names worldwide, or even country or statewide, does not, to my knowledge, exist. This is rather unfortunate, because it means we'll need to get owners' names and billing addresses from somewhere else.

This is where the government comes in. Each year, town governments' tax assessors are responsible for compiling a "tax book" containing details about property ownership, billing addresses, property tax amounts, and more. This is different from a "tax map" - which is simply an architectural drawing of the town and the spaces that its properties occupy. These tax books can be used in conjunction with OpenAddresses' address data to create full records, consisting of an owner name, local address, billing address, and coordinate pair.

Discerning Waterfront Addresses from Non-Waterfront Addresses

Now that we've got a hefty amount of data, it's time to narrow it a bit. We need to know which addresses are on the water, and which are not. Unfortunately, this information isn't contained within the tax books. In fact, during my research on this topic, I was unable to find any government documents containing waterfront vs. non-waterfront address information for any of the towns of the watershed. It appears that the local government has failed us - we'll need to take an alternative approach.

To understand the method used by the Hawk System for determining waterfront addresses, we need to understand the problem in a bit more detail. A water body, on a map, is essentially some shape made up of a number of points connected by the same number of straight lines (see Figure 1 below for an example). We can call this representation of a lake its polygon representation. Let's say we represent a property by a point somewhere relative to this polygon representation (green dot in Figure 2). Using this simplistic representation of a lake and a property, we can draw a line between the property and the closest point on the lake (Figure 3). The length of this line can be treated as the approximate distance from the property to the lake. Using this method, a distance threshold can be set, and any distance calculated to be under this threshold can be marked as a waterfront property. (The number of points included in a polygon representation is increased dramatically compared to Figures 1-3 so that the closest point to the property represents a number very close to the actual distance from the property to the water.)

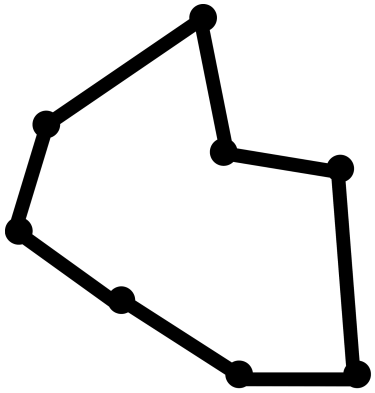


Figure 1: Polygon representation: a simple representation of the outline of a lake, made of points and straight lines.

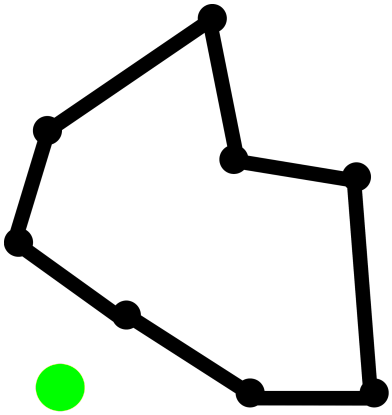


Figure 2: A property near a lake is represented as a single point near the lake's polygon representation.

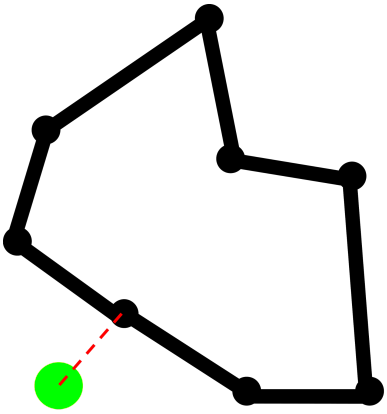


Figure 3: The distance from a property to its lake can be calculated by drawing a line from the property to the closest point in the lake's polygon representation.

Implementation and Restrictions

The above process is implemented using a programming language called Python[2]. Python is useful for managing a lot of data (a feature Hawk makes heavy use of) and quickly implementing ideas and methods. However, since implementation is not the same as theory, several restrictions have been encountered during programming.

Hand-Drawn Polygon Representations

Each water body in the watershed must be outlined by hand in Google Earth in order to obtain its polygon representation. While this only has to be done once for each body, it takes a while for larger ones (approximately 30 min - 1 hour for a lake as large as Cobbossee). If Hawk expands to other areas of Maine or the country, each lake/pond/stream will need to be hand outlined. Automatic solutions to this could be created using machine learning[3], however, this is beyond my skills at the moment.

Address Matching

In some cases, addresses are recorded differently by different sources. For example, a private road may be called by one name in OpenAddresses, and by another in the town's tax book. In these cases, this leads to a mismatch, and billing addresses are unable to be obtained. These types of address mismatches, along with others, currently make up about 5% of waterfront addresses in the Cobbossee Watershed; a significant amount when dealing with thousands of entries. In the future, this percentage can be reduced by implementation improvements, or eradicated with GPS-based machine learning.

The Future of Hawk

Currently, Hawk is in the extremely early stages of development. In the future, I envision a complete desktop application that can be easily and automatically used by any user. The application will allow you to find waterfront addresses, their billing addresses, create maps of these, export searches by attribute, and cross reference data with your own member databases. Additionally, the user will be able to view maps of their lakes[4] to see where their members and non-members are, allowing them to perform more specific and effective outreach.

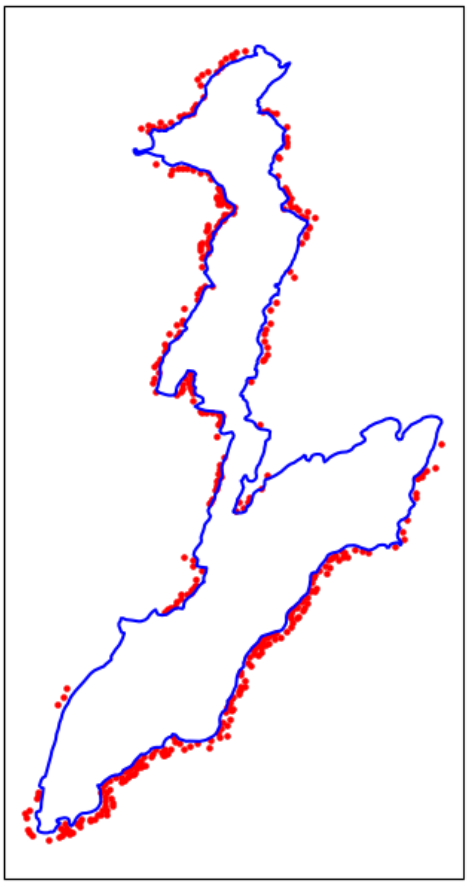
Results and Takeaways

The Hawk Member Management System has the potential to change the way watershed protection organizations gather donations, track member data, and provide visuals to donors and board members. With more work and improvements, the system may even be able to replace all other data management systems used by the FOCW and other similar organizations. Beyond providing useful information to organizations, the system could also help transition labor costs

from manual, traditionally “boring” work, to more creative and innovative work, saving funds for more important endeavors and allowing nonprofit employees the opportunity to contribute creatively.

Appendix

1 - https://openaddresses.io/
2 - https://www.python.org/doc/essays/blurb/
3 - https://www.ibm.com/cloud/learn/machine-learning
4 -



A map of Maranacook Lake generated by Hawk. The red dots are waterfront properties calculated to be within 100 meters of the water (distance threshold = 100 meters). The mapping feature will be implemented in future versions of the program.