Sophia K. Cheng

SIADS 601
Fall 24
Week 4 - Final Report

Data Set and Brief Description

The dataset I will be using for this project is from the research article:

Brehob, M. M., Pennino, M. J., Handler, A.M., Compton, J. E., Lee, S. S., & Sabo, R.D. (2024). Estimates of lake nitrogen, phosphorus, and chlorophyll-a concentrations to characterize harmful algal bloom risk across the United States. Earth's Future, 12, e2024EF004493. https://doi.org/10.1029/2024EF004493

Specifically, I will be looking at the normalized inputs which combines data from the following sources:

- National Lakes Assessments
  - Nutrient and chlorophyll-a concentrations
- LakeCat
  - Soil, land cover, and landscape characteristics
- National Nutrient Inventory
  - Landscape nutrient data
- National Atmospheric Deposition Program
  - Nutrient deposition data
- PRISM
  - Climate data
- Earth Observatory Network
  - Landscape ecosystem data
- LAGOS
  - Lake depth data
- National Hydrography Dataset
  - Watershed characteristics

The data contains **64 features** (including identifying columns) and **2,181 rows**. The data is for the **calendar year 2007** and can be accessed here: https://catalog.data.gov/dataset/estimates-of-lake-nitrogen-phosphorus-and-chlorophyll-a-concentrations-to-characterize-har.

Broad Question

What are trends and patterns in physical lake characteristics and concentrations of nutrients?

Report (word count: 973)

When I came across the paper described above, I was immediately drawn to it because of its use of data collected for a varying reasons, some of which appear unrelated to each other. The authors combined the data and were able to generate a model that characterized harmful algal bloom risk with estimates of nutrients such as nitrogen and chlorophyll-a. The variable data leading to this model made me want to understand what are the drivers of the nutrients that are able to characterize algae. In the paper, the authors focus on data about lakes across the United States. The data used includes percentage breakdowns of contributions to nitrogen and phosphorus, such as whether it is from agricultural or other types of run off. As someone who has had very few experiences with lakes (and nature in general), my first step to understanding what are the drivers is to first understand lakes from people who have more extensive experiences with lakes. For a broader understanding, I sought out participants with differing relationships with water, ranging from predominantly recreational to career focused, to interview.

To begin with, choices of recreational activities is a motivating factor for all participants when deciding which lakes to visit. For example, one participant enjoys fishing and water sports. For them, they prefer large lakes that have the appropriate depth for the species of fish they are interested in. Another participant prefers hiking and observing wildlife, and so, prefers lakes that are accompanied by an abundance of trees. These examples raise an interesting question, what percentage of human impact on lake ecosystems is contributed to from recreation versus industry? Additionally, most of the participants remarked on the clarity of the lake water, with one stating "A healthy lake has schools of fish that you can see, the water is clear and things should move around.". Assuming participants prefer visiting healthy lakes, this implies that clarity of water is an indication of healthy lakes, one participant compared an unhealthy lake having a clarity “like a dirty fish tank”

A lake's health can also be assessed by its ecosystem. The prevalence of wildlife and plant species around a lake provides an indication of the overall health of the lake. One participant put it succinctly with, “"Lakes are super important habitats that are part of the water cycle.". Participants expect to find a variety of wildlife such ast turtles, ospreys, water spiders, mussels and herrings. Most participants remarked on trees around the lake, with one participant notably remembering the feeling of fallen leaves in the water as they stepped into it. This raises an interesting question about the impact of tree species near a lake to the nutrients in the water. Does the decomposition of certain species increase the risk of algal blooms?

Cumulatively, the participants have visited lakes in different areas of the contiguous United States. With this in mind, one participant pointed out different parts of the US have different types of rocks, which become the sands on the lake beach and form the bed of the lakes. The rocks themselves contribute to the available nutrients in the water. Given that one of the nutrients contributed is salinity, there is an implication of rocks on the health of the lake. Additionally, this information puts the use of data from LakeCat into perspective. This particular data source focuses on soil, land cover and landscape characteristics of lakes in the United States.

When discussing expectations of nutrients in lake water, all participants were in agreement that there is carbon in water and that it should be in different forms. The carbon is the result of various stages of

decomposition of organic matter, supporting one participant's description of the carbon as "dissolved organic carbon". Varying stages of carbon implies a relationship with the lifecycle of organisms in the water. This relationship further implies that the data will have a "seasonal" aspect to it. When combined with the seasonal pattern to participants' visits to lakes, it may be possible to detect a seasonal pattern with drivers of nutrients that themselves drive algal blooms.

All participants agree with the sentiment that algal blooms are occurring more frequently and are a byproduct of human behavior, with one participant stating "Algae blooms are [an] increasingly frequent problem humans created, to our own detriment.". One of the rationale behind this being a byproduct of human behavior is that human created chemical runoffs cause an imbalance in the ecosystem by adding more nutrients into the water than it can handle. It is possible that these chemical runoffs directly contribute to the nutrients authors of the paper reference. In addition to the toxicity hazards of algal bloom, the decomposition of the same algal bloom causes oxygen deficiencies in the water.

It is important to note that while the concerns are for harmful algae blooms, not all algae blooms are harmful. Algae is a natural occurrence and has its own place in the ecosystem. One participant delightfully describes the non harmful ones as "spongy, fuzzy mats of blobby green strings". In the attempts to eliminate harmful algae blooms, there needs to be considerations made for the non-harmful ones.

In conclusion, the interviews have revealed several interesting potential relationships to further explore. For example, what are the impacts of differing kinds of rocks to the availability of nutrients in a lake? Would it be possible to alleviate the ecosystem imbalance by the introduction of non-native rocks? Alternatively, what is the impact of human recreational activities to the health of a lake? And if so, could this information be used to inform policy to encourage a balanced human-to-ecosystem relationship? The next step to answering these questions is to revisit the data from the perspective of the nutrients being the response variable. To avoid possible data leakage, data specifically describing algal blooms should be removed.