**Visualizing Large-Scale Geographic Gradients in Lakes of the United States**

Lakes are a vital natural resource, supplying water for human use and providing habitat for a wide variety of plant and animal species. Protecting the quality of lakes is critically important for human health, economic interests, and the preservation of species that inhabit these freshwater systems. In the interest of monitoring the quality of the nation’s freshwater resources, the United States Environmental Protection Agency (EPA) conducts regular aquatic resource surveys of lakes, rivers, wetlands, and costal waters. For this project, I have chosen to analyze data from the EPA’s 2017 National Lakes Assessment report to visualize large scale geographic gradients that may exist across the nation’s lakes.

I began by accessing the publicly available data files from the EPA website and downloading the files that contained the variables I was interested in visualizing. This presented the first main challenge of the project, which was to read and combine these files in a way that they could be effectively analyzed. The data was all available as csv files, but there are multiple, rather large files, that each have a different format. To deal with this in R, I read in the relevant files and used various select, filter, and pivot functions to pull out my desired variables and measurements all in the same format. From there, I merged the data into one data frame named “mydata\_NLA17”, so that each variable could be easily compared to any other variable. Metadata for all files used in this project are also included in my GitHub repository.

To visualize the geographic gradients that may exist among these selected water quality variables, I made use of the basic mapping functionality built into the ggplot2 package. Each map showed points for each lake location, which were colored on a gradient representing the value of the variable measurement. I generated individual maps for concentrations of Nitrogen, Phosphorus, Chlorophyll-A, and percentage of watershed area used as cropland. For most of the variables, there were a handful of especially high outlier measurements that skewed the gradient represented on the map, making it difficult to visualize any trends. For simplicity of visualizing lots of variables very quickly, I chose to simply filter out any measurements above a reasonable threshold to make the maps more readable. Finally, I used ggplot2 again to graph a few of the variables on scatterplots, to see if there were any strong correlations.

In general, I observed higher indicators of agricultural use, nutrient loading, and possible eutrophication in lakes stretching from the upper Midwest, down to the Gulf Coast region. Although this method has helpful for getting a broad sense of special trends in the data, the scatterplots did not show any especially strong correlations between variables in this particular data set. Further statistical analysis or specific field studies would be needed to investigate these direct correlations

**References:**

U.S. Environmental Protection Agency. 2022. National Aquatic Resource Surveys. National Lakes Assessment 2017 (data and metadata files). Available from U.S. EPA web page: <https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys>. Date accessed: 2024-03-14.

Read, E. K., Patil, V. P., Oliver, S. K., Hetherington, A. L., Brentrup, J. A., Zwart, J. A., Winters, K. M., Corman, J. R., Nodine, E. R., Woolway, R. I., Dugan, H. A., Jaimes, A., Santoso, A. B., Hong, G. S., Winslow, L. A., Hanson, P. C., & Weathers, K. C. (2015). The importance of lake-specific characteristics for water quality across the continental United States. Ecological Applications, 25(4), 943–955. <https://doi.org/10.1890/14-0935.1>

Moslenko L, Blagrave K, Filazzola A, Shuvo A, Sharma S. Identifying the Influence of Land Cover and Human Population on Chlorophyll a Concentrations Using a Pseudo-Watershed Analytical Framework. Water. 2020; 12(11):3215. <https://doi.org/10.3390/w12113215>

Wang, H., García Molinos, J., Heino, J., Zhang, H., Zhang, P., & Xu, J. (2021). Eutrophication causes invertebrate biodiversity loss and decreases cross-taxon congruence across anthropogenically-disturbed lakes. Environment International, 153, 106494. <https://doi.org/10.1016/j.envint.2021.106494>

Galbraith, L. M., & Burns, C. W. (2007). Linking Land-use, Water Body Type and Water Quality in Southern New Zealand. Landscape Ecology, 22(2), 231–241. <https://doi.org/10.1007/s10980-006-9018-x>

Topp, S. N., Pavelsky, T. M., Dugan, H. A., Yang, X., Gardner, J., & Ross, M. R. V. (2021). Shifting patterns of summer lake color phenology in over 26,000 US lakes. Water Resources Research, 57, e2020WR029123. <https://doi.org/10.1029/2020WR029123>