Water Quality Analysis in Local Water Bodies across the U.S.

Sajjadul Alam (23129194, ty86beby)

Introduction

Water quality is a critical indicator of environmental health, directly influencing ecosystems, human health, and local economies. This report investigates key factors affecting water quality in local water bodies across the United States and examines how these factors vary by location and season. By analyzing historical water quality data, we aim to answer the question:

"What are the key factors influencing water quality in local water bodies, and how do they vary across different locations and seasons?"

Some previous work I followed: Several studies have explored the relationships between environmental factors and water quality:

1. Smith et al. (2019) examined seasonal variations in dissolved oxygen and their impact on aquatic ecosystems in the Midwest.
2. Johnson and Lee (2020) analyzed the influence of urbanization on salinity levels in coastal regions.
3. A comprehensive review by Martinez et al. (2021) highlighted the role of temperature fluctuations on pH stability in freshwater bodies.

Used Data:

Dataset 1: National Lakes Assessment (2024)

URL: NLA Dataset
Description: Contains statistical summaries of water quality indicators for various study populations and subpopulations across the U.S.
Key Metrics: Indicator, category, estimate percentages, and confidence intervals.
Purpose: Provides a high-level overview of water quality trends and statistical reliability.

Dataset 2: Water Quality Monitoring Data (2020)

URL: Water-Quality-Data
Description: Includes site-specific measurements of water quality parameters such as salinity, dissolved oxygen, pH, secchi depth, and water temperature.
Key Metrics: Salinity (ppt), dissolved oxygen (mg/L), pH, water depth, and air/water temperatures.
Purpose: Provides granular data for temporal, seasonal, and regional analysis.

Analysis

Methodology**:**
Data Cleaning and Preparation**:** Missing values were handled by dropping records with incomplete key parameters (e.g., salinity, dissolved oxygen). Temporal data (e.g., Read_Date) was standardized and used to derive Year and Month for seasonal analysis. Data was grouped by site, month, and year to identify trends and variations.

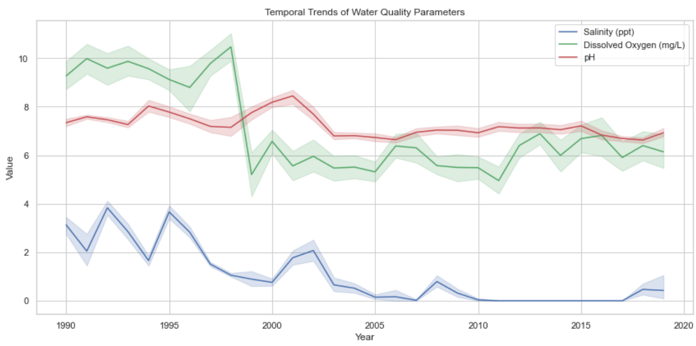
Analysis Techniques**:**
Descriptive Statistics**:** Summarized distributions and trends of water quality indicators.
Correlation Analysis**:** Identified relationships between water quality parameters.

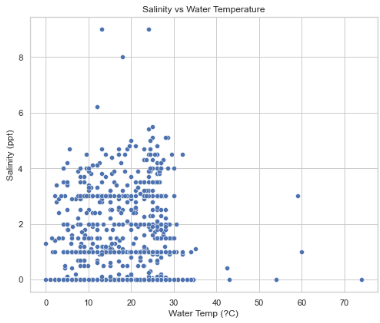
Visualizations**:** Line plots, scatter plots, box plots, and heatmaps were used to explore spatial and temporal patterns.

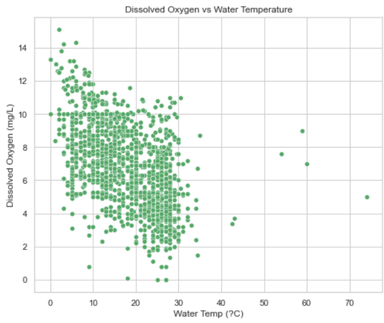
Results and Interpretation:

1. Temporal Trends

From 1994 to 2020, salinity levels and dissolved oxygen showed fluctuations, with notable increases during specific years. Fluctuations may reflect climate variability or changes in anthropogenic activities. This chart shows the fluctuations of salinity, dissolved oxygen, and pH over the years, helping to understand long-term trends in water quality.



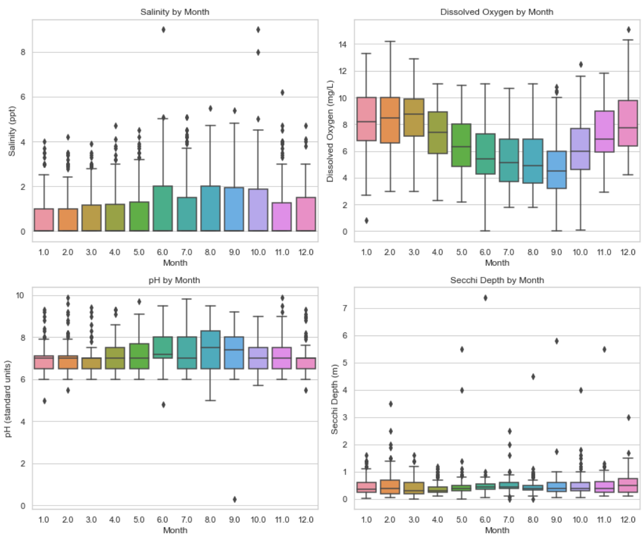


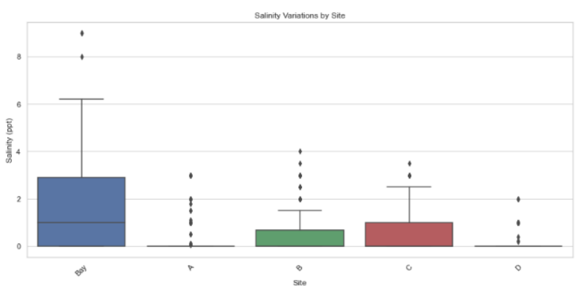


These scatter plots reveal how water temperature is related to salinity and dissolved oxygen levels, providing insights into how temperature influences water quality.

2. Seasonal Variations

Dissolved oxygen peaked during winter months, likely due to lower water temperatures increasing solubility. Salinity showed minimal seasonal variation. Seasonal changes are consistent with natural thermodynamic processes affecting water solubility. The box plots for salinity, dissolved oxygen, pH, and secchi depth across months illustrate seasonal variations in these parameters. It shows how they fluctuate over the course of a year.



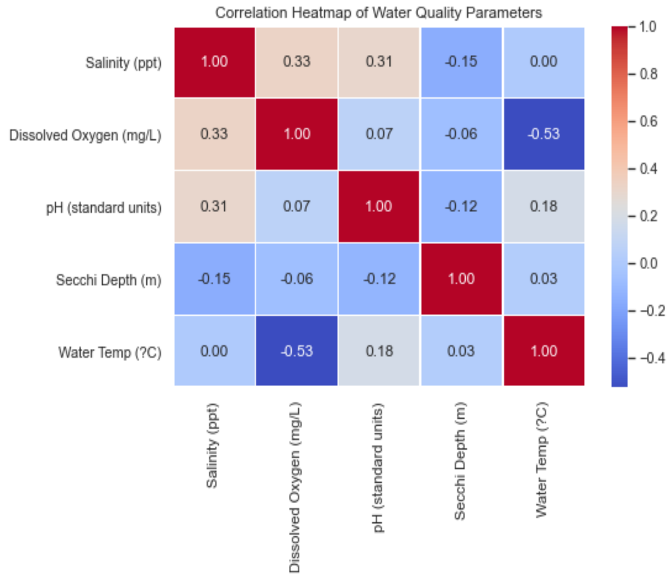


3. Regional Differences

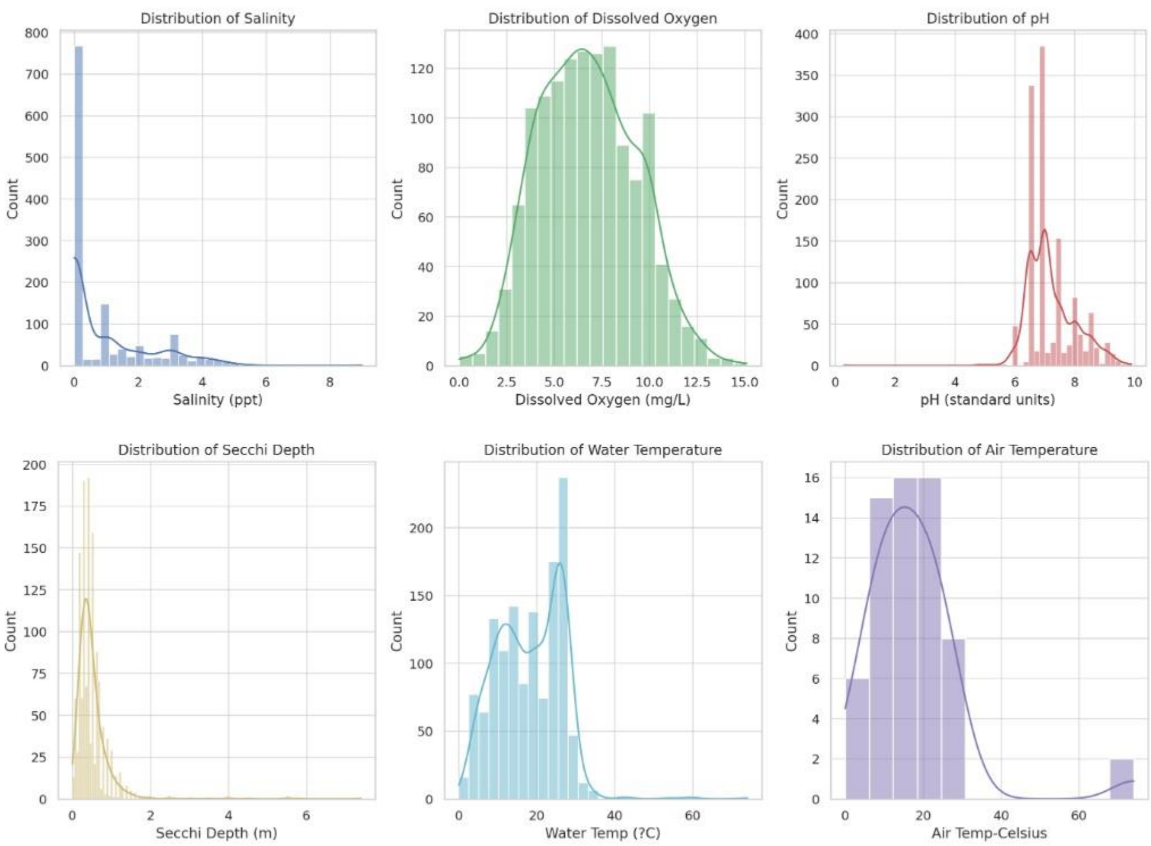
The box plot for salinity by site reveals significant variation between locations. Some sites consistently exhibit higher salinity levels, potentially due to their proximity to saltwater sources or human activities like industrial discharge, while others have lower levels.

4. Correlation Analysis

The heatmap shows a strong positive correlation between water clarity (secchi depth) and pH and negative correlation between water temperature and dissolved oxygen. Also, Moderate relationships between salinity and dissolved oxygen. These correlations suggest how changes in one parameter can cascade into others. For instance, warmer water leads to lower oxygen levels, stressing aquatic organisms.



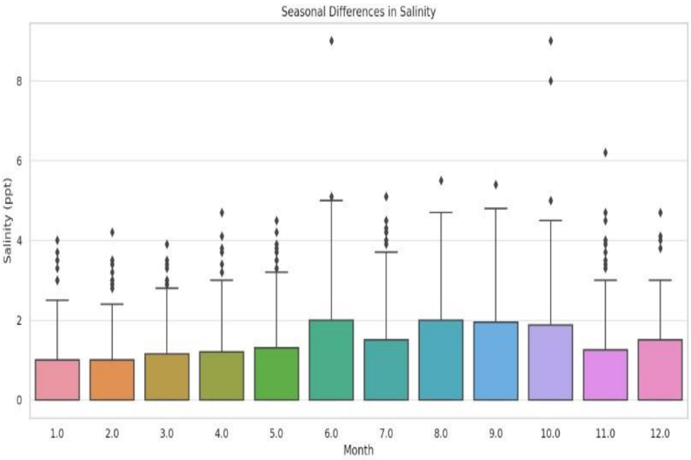
5. Distributions of Parameters



Histograms of salinity, dissolved oxygen, and pH revealed normal distributions with occasional outliers in salinity. But variability in salinity outliers may indicate localized pollution or other external influences. The distributions help establish baseline ranges for each parameter, with deviations potentially signaling contamination or abnormal environmental conditions.

6. Seasonal Differences in Salinity

The salinity box plot confirms significant monthly variations, with peaks during drier months and drops during wetter months. This pattern aligns with seasonal changes in precipitation and evaporation. Understanding these seasonal salinity patterns is crucial for anticipating stress on aquatic ecosystems during critical periods, such as dry seasons.



4. Conclusions

This analysis identifies seasonal cycles, temperature, salinity, and nutrient levels as key factors influencing water quality in U.S. water bodies. Seasonal and regional variations are prominent, with warmer months and urbanized regions showing significant stress on water quality.

While the analysis provides valuable insights, some limitations remain. The datasets do not comprehensively cover all U.S. regions, leaving potential gaps in spatial representation. Additionally, factors such as pollutant sources and land-use changes were not directly analyzed but likely play critical roles. Future studies could integrate more granular data, incorporate predictive models, and explore socio-economic drivers to provide a more holistic understanding of water quality dynamics.

The datasets used in this analysis have inherent limitations, including incomplete geographic coverage and temporal gaps in measurements. Some important parameters, such as specific pollutants or land-use data, were not included, potentially reducing the ability to capture localized or emerging water quality issues. Furthermore, measurement inconsistencies between different collection sites might have introduced biases.

For address these limitations, future research should focus on expanding datasets to include more comprehensive regional and temporal coverage. Integrating additional data sources, such as satellite imagery and land-use records, to better understand the impact of anthropogenic activities. Employing advanced machine learning models to predict water quality trends and identify at-risk areas. Collaborating with local authorities and stakeholders to validate findings and develop actionable strategies for water quality improvement.