PROJECT ASSIGNMENT -2

I was working with Indiana Clean Lakes monitoring program from 1988-2010.

Here was my research question

- How does nutrient concentration (nitrate, ammonia, phosphorus) affect water clarity in Indiana lakes?
- What factors influence algal growth (as measured by chlorophyll-a) in Indiana lakes?
- What is the suitability of lake water in Indiana for treatment into potable drinking water, and how can it be efficiently supplied to households?

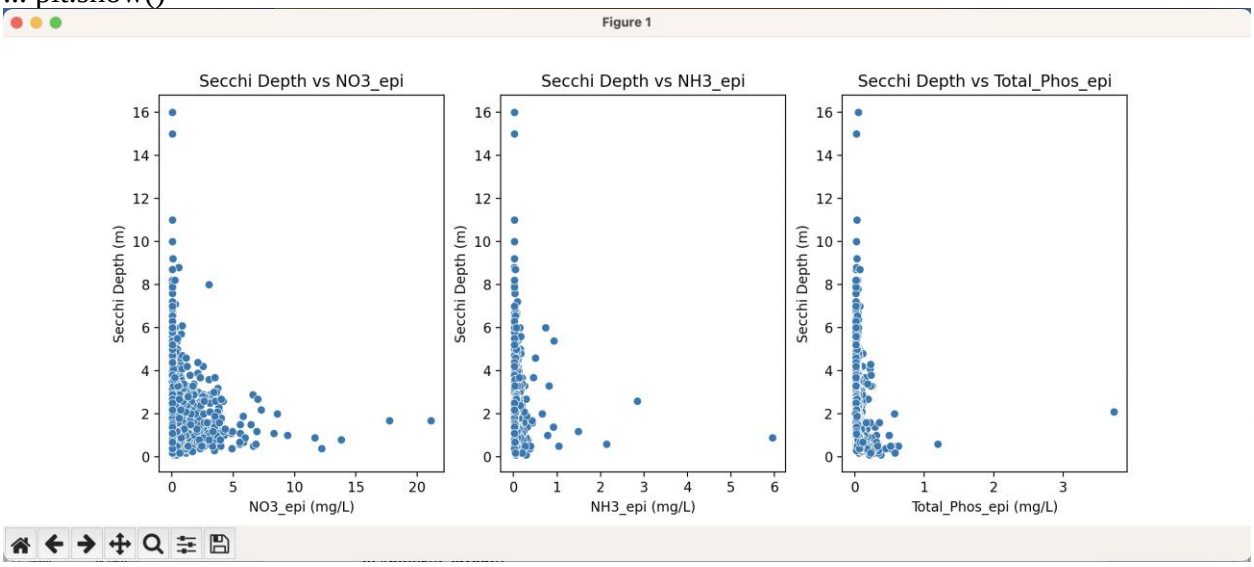
Graphs and Table Section

Q1. At least five graphs or tables linked to your research questions or exploratory data analysis. It is recommended that you have one graph and one summary table per research question.

Ans.

- For question 1 I'm generating scatter plot of Secchi vs ammonia/nitrate/ phosphorous/

```
df = pd.read_csv('/Users/abhijitghosh/Documents/DataScience/IN_chemistry.csv')
...
... fig, axes = plt.subplots(1, 3, figsize=(15, 5))
...
... for ax, nutrient in zip(axes, ['NO3_epi', 'NH3_epi', 'Total_Phos_epi']):
...     sns.scatterplot(x=df[nutrient], y=df['Secchi'], ax=ax)
...     ax.set_title(f'Secchi Depth vs {nutrient}')
...     ax.set_xlabel(f'{nutrient} (mg/L)')
...     ax.set_ylabel('Secchi Depth (m)')
...
... #plt.tight_layout()
... plt.show()
```



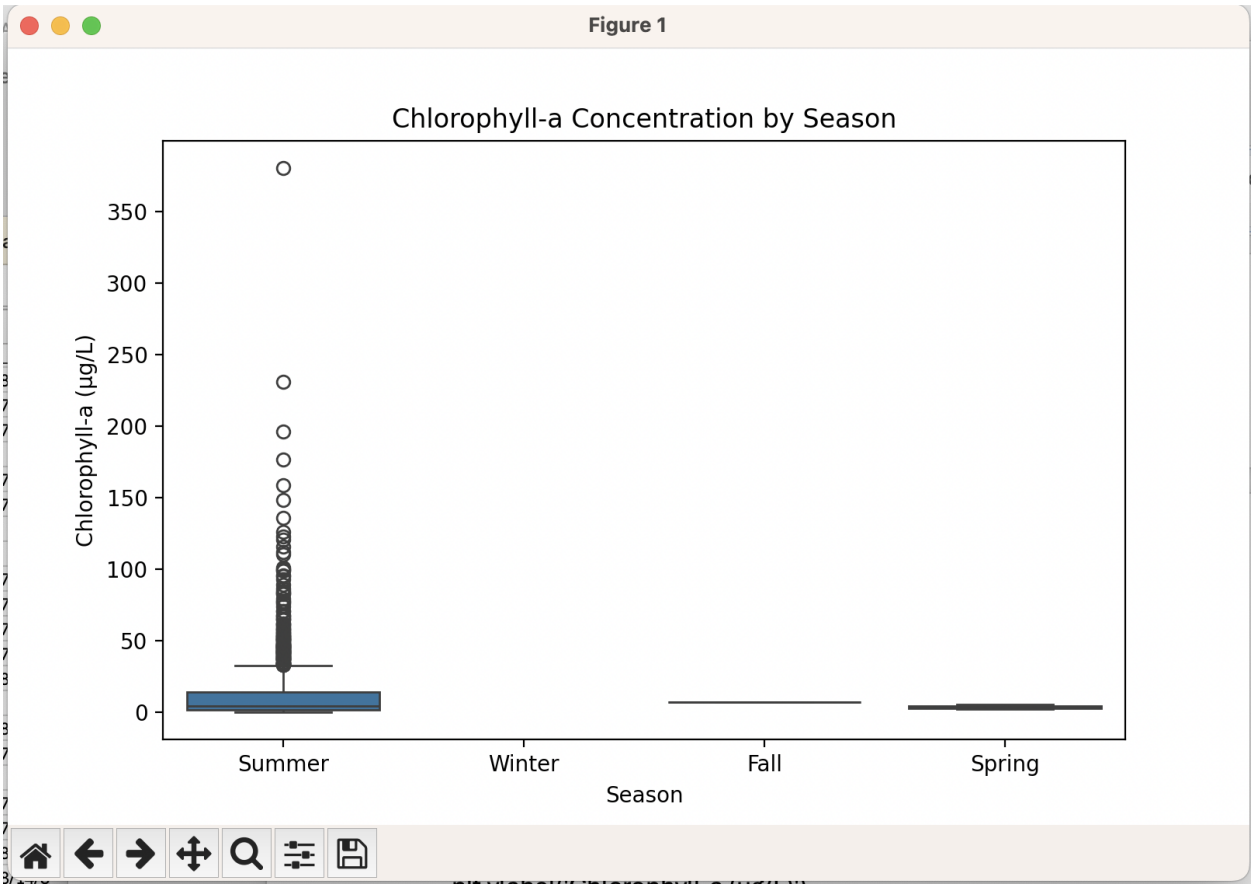
Summary Table

```
>>> summary_table_1 = df[['NO3_epi', 'NH3_epi', 'Total_Phos_epi', 'Secchi']].corr()
... print(summary_table_1)
...
              NO3_epi   NH3_epi  Total_Phos_epi    Secchi
NO3_epi      1.000000  0.048845        0.028177 -0.137084
NH3_epi      0.048845  1.000000        0.033830 -0.031964
Total_Phos_epi  0.028177  0.033830        1.000000 -0.190930
Secchi       -0.137084 -0.031964       -0.190930  1.000000
```

There is no correlation found in between these variables.

- For second research question regarding influence algal growth (as measured by chlorophyll-a) in Indiana lakes I'm generating a box plot

```
df['Month'] = pd.to_datetime(df['Date_Sampled']).dt.month
... df['Season'] = df['Month'].apply(lambda x: 'Spring' if x in [3, 4, 5] else
...                                   'Summer' if x in [6, 7, 8] else
...                                   'Fall' if x in [9, 10, 11] else 'Winter')
...
... plt.figure(figsize=(8, 5))
... sns.boxplot(x='Season', y='Chlorophyll_a', data=df)
... plt.title('Chlorophyll-a Concentration by Season')
... plt.xlabel('Season')
... plt.ylabel('Chlorophyll-a (µg/L)')
... plt.show()
```



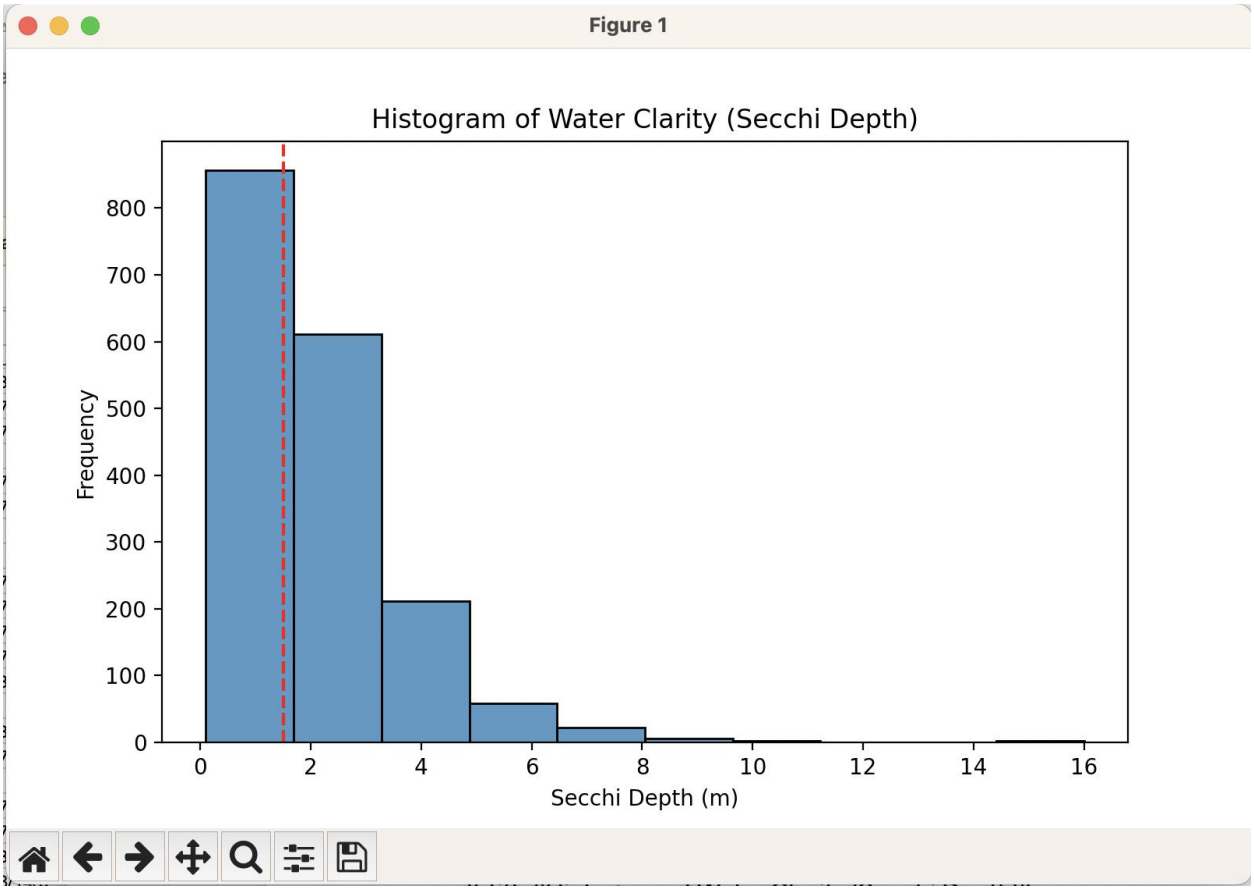
Summary Table

```
>>> summary_table_2 = df[['Chlorophyll_a', 'Secchi', 'Total_Phos_epi', 'NH3_epi',
'NO3_epi']].corr()
... print(summary_table_2)
...
              Chlorophyll_a    Secchi  Total_Phos_epi   NH3_epi   NO3_epi
Chlorophyll_a       1.000000 -0.393018       0.455074  0.040401  0.069566
Secchi             -0.393018  1.000000      -0.190930 -0.031964 -0.137084
Total_Phos_epi      0.455074 -0.190930       1.000000  0.033830  0.028177
NH3_epi             0.040401 -0.031964       0.033830  1.000000  0.048845
NO3_epi             0.069566 -0.137084       0.028177  0.048845  1.000000
```

Higher levels of Chlorophyll_a are associated with higher levels of Total_Phos_epi

- For research question3 I'm generating a histogram of Secchi depth

```
>>> plt.figure(figsize=(8, 5))
... sns.histplot(df['Secchi'], bins=10)
... plt.axvline(x=1.5, color='r', linestyle='--', label="Low Clarity Threshold (1.5m)")
... plt.title('Histogram of Water Clarity (Secchi Depth)')
... plt.xlabel('Secchi Depth (m)')
... plt.ylabel('Frequency')
... plt.show()
```



Summary Table

```
>>> summary_table_3 = df[['Secchi', 'Total_Phos_epi', 'NH3_epi',
'NO3_epi']].describe().T[['mean', '50%', 'min', 'max', 'std']]
... print(summary_table_3)
...
                mean    50%    min     max       std
Secchi         2.078238  1.700  0.100  16.000  1.538317
Total_Phos_epi  0.053383  0.034  0.006   3.734  0.110105
NH3_epi         0.054438  0.025  0.004   5.955  0.183605
NO3_epi         0.578779  0.040  0.010  21.124  1.325582
```

Lakes have both with high and low and clarity. Few lakes have high phosphorous levels due to pollution. There are also lakes with high nitrate and ammonia causing algae growth abd harm to aquatic life.

Interpretations

- Research Question 1(scatter plot Secchi vs ammonia/nitrate/phosporous) :

  The scatter plots indicate the relationship between nitrate (NO_3), ammonia (NH_3), and phosphorus (Total_Phos) concentrations and water clarity (measured by Secchi depth). In the

graph we can see nutrient concentration (nitrate, ammonia, phosphorus) vs. water clarity (Secchi depth) shows most of the data points concentrated along the y-axis at low x-values, it suggests that the majority of lakes in the dataset might have low nitrogen (NO_3, NH_3) and phosphorus (Total_Phos_epi) concentrations. But Secchi depth is spread across y axis.

There is no direct relation between nutrients and Secchi.

- Research Question 2(Box plot Season vs Chlorophyl) :

  The boxplot shows variations in Chlorophyll-a concentrations across seasons. Chlorophyll-a is a direct indicator of algal biomass, which affects lake ecology and water quality. We can see highest chlorophyl in summer(due to warm temperature and sunlight), in spring it's less and winter there is no data due to lack of sunlight and cold.
  Algal blooms are most severe in summer, highlighting the need for seasonal monitoring.

  We can also see in summer there are lot of outliers for chlorophyl which suggest extreme cases of algal blooms occurring during this season. Lakes experience much higher-than-normal chlorophyll-a concentrations, while others remain within the typical range. Extremely high chlorophyll-a levels may be harmful to aquatic life and human health as well. We need to take necessary action to reduce this like waste water discharge, cleaning etc.

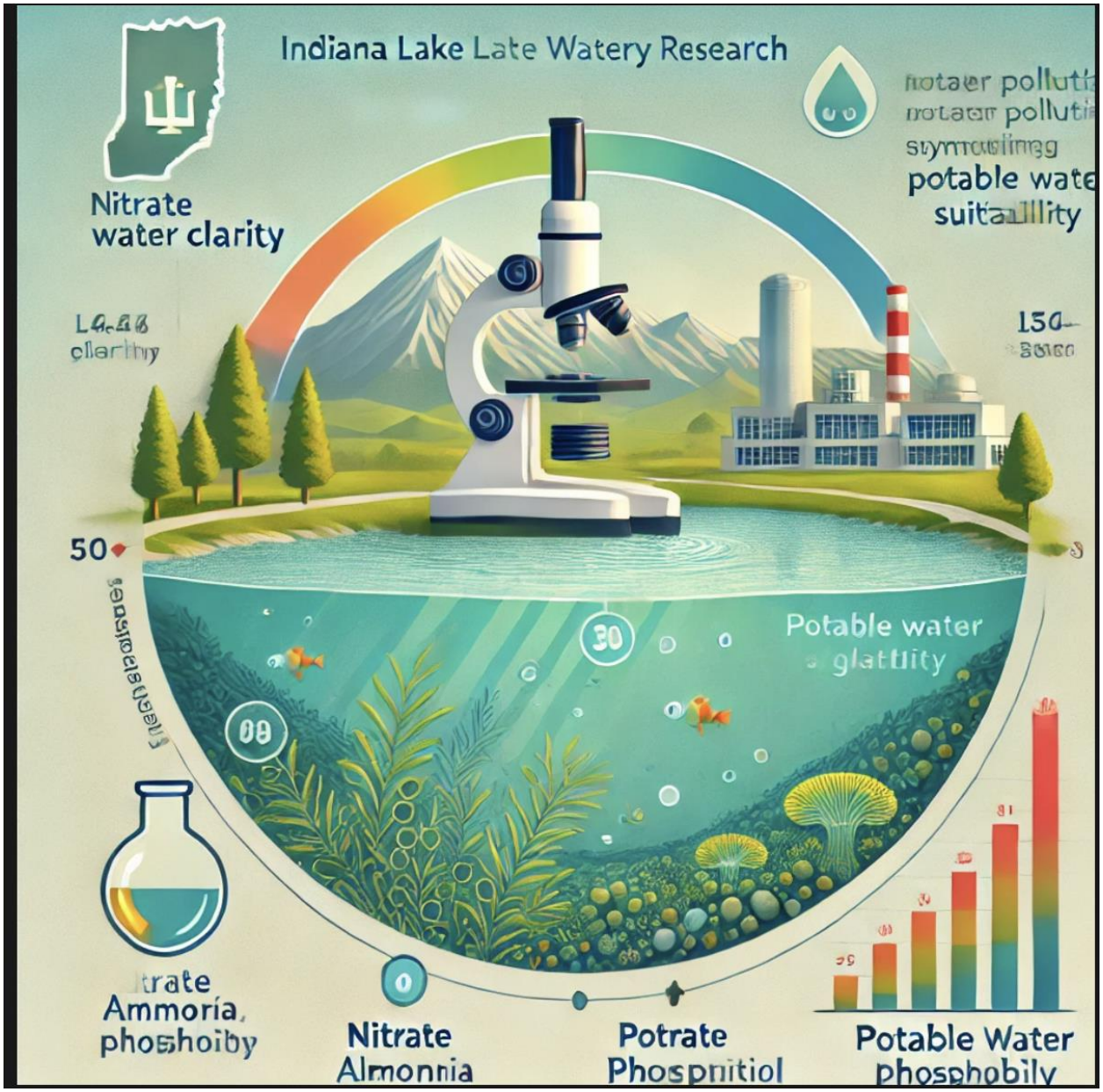
- Research Question 3(Histogram of Secchi depth) :

  The histogram illustrates the distribution of Secchi depth across lakes. The red dashed line at 1.5m represents a threshold where water is considered too turbid for efficient drinking water treatment.
  This histogram indicates a significant number of lakes have Secchi depths below 1.5m, indicating poor water clarity. Some lakes exhibit high clarity (>3m), suggesting that not all lakes are equally affected by turbidity.

A project synopsis As Tweet

Indiana Lakes: Clarity & Quality Study

Exploring how nutrients impact water clarity, factors driving algal growth, & suitability for drinking water using 1988-2010 data. Key insights on pollution, ecosystem health, & water treatment efficiency. #DataScience #CleanWater



Keywords

- Water Quality
- Nutrient Pollution
- Algal Growth
- Lake Clarity
- Drinking Water Safety
- Environmental Data Analysis
- Indiana Lakes