STEM Institute BOP Presentation

Research Project by Veronica Manevich, Eshaan Sharma, Lena Santoni, Daniel Shoshin, and Vivek Maranganti
(Team #8)

Research Question:

How can we choose sufficient sites for the oysters according to the time of year during which measurements are taken, also while putting water conditions into consideration?

Background: What is the Billion Oyster Project?

- The waterways of New York used to encase several oysters in the previous century.
- However, through the involvement of humans and the introduction of pollution into the waters, the oyster population began to diminish and has been at a low ever since.
- This marine science project attempts to use data science and analytical thinking with the help of professional scientists, to put 1 billion oysters in the polluted New York Harbor waters.
- Through this project, the BOP Foundation hopes to involve 1 million people, in the action of allowing these oysters to filter and clean the toxic water.

Some Background on Oysters

- Oysters are mollusks that were gifted with a very beneficial anatomy, that allows them to be able to take in the polluted water with their food, and separate what they want to eat, from what they want to excrete.
- When oysters do not like something, they are able to release it back into the water and secrete it with a liquid known as nacre, and that then allows the dirty water to sink to the floor of the harbor, making more space for clean water.
- Labial palps are essentially the mouth of the oyster, and are exactly what allows this mollusk to separate what it wants and does not want to eat.
- Like all animals, oysters also have predators that are set out to get them and stop their progress, and those include oyster drills, certain fish, and some others.

Importance of Oysters

- Oyster reefs provide a 3D habitat for hundreds of other species.

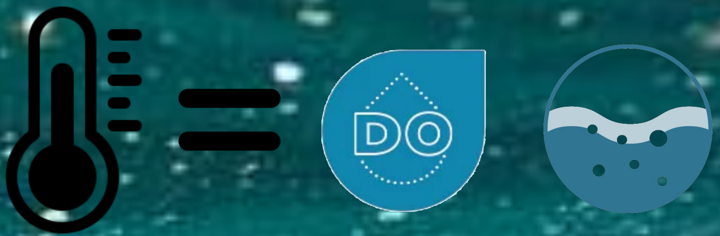
  * Including crabs, starfish, oyster drills, and many other marine creatures.

- As oysters eat, they also filter the water, resulting in the removal of certain pollutants, such as nitrogen.

  * A single oyster can filter over 50 gallons of water.

- Large oyster reefs are also a natural storm barrier, providing defence against large waves, while also reducing the damage of floods and erosion.

Question Exploration Details



- For our research question, we wanted to explore the factors that go into choosing oyster sites.
    - (i.e. the time of year, water conditions, and temperature)
- We also want to explore the connection between different factors.
    - For example, different times of year mean different temperatures, which can affect turbidity and oxygen levels.
    - We primarily focused our attention and research, though, on the data provided about turbidity, dissolved oxygen, and temperature levels for different oyster samples.

Hypothesis

- As dissolved oxygen increases, the temperature should decrease as molecules are more soluble in cold water because they are closer together.
- As turbidity increases, the temperature should increase as well. Since turbidity encases more particles in the water, they then absorb more heat and increase the water temperature.
- Also as dissolved oxygen increases, the turbidity should decrease as well, because more space is taken up by the oxygen particles in the water.

Data Analysis

- While observing the different categories in the data sheet, we began to notice correlations between different pieces of data. For example, we discovered that when the temperature of the water increased, the amount of dissolved oxygen decreased. Using these correlations, we began to create graphs to display the changes in environmental conditions.
- We also noticed that there were trends amongst dissolved oxygen and turbidity which showed that they were inversely proportional, since as turbidity increases, dissolved oxygen would mostly decrease and vice-versa.
- One last trend noticed was the directly proportional relation of temperature and turbidity where increasing one increased the other as well.

Term Definitions

- **Turbidity - The clarity of a sample of water, which is measured by the amount of light that is scattered by material in the water when a light is shined through the water sample.**

- **Dissolved Oxygen - The amount of oxygen present in the water (solubility)**

- **Temperature - measuring the warmth or coldness of water (displaying in Celsius)**

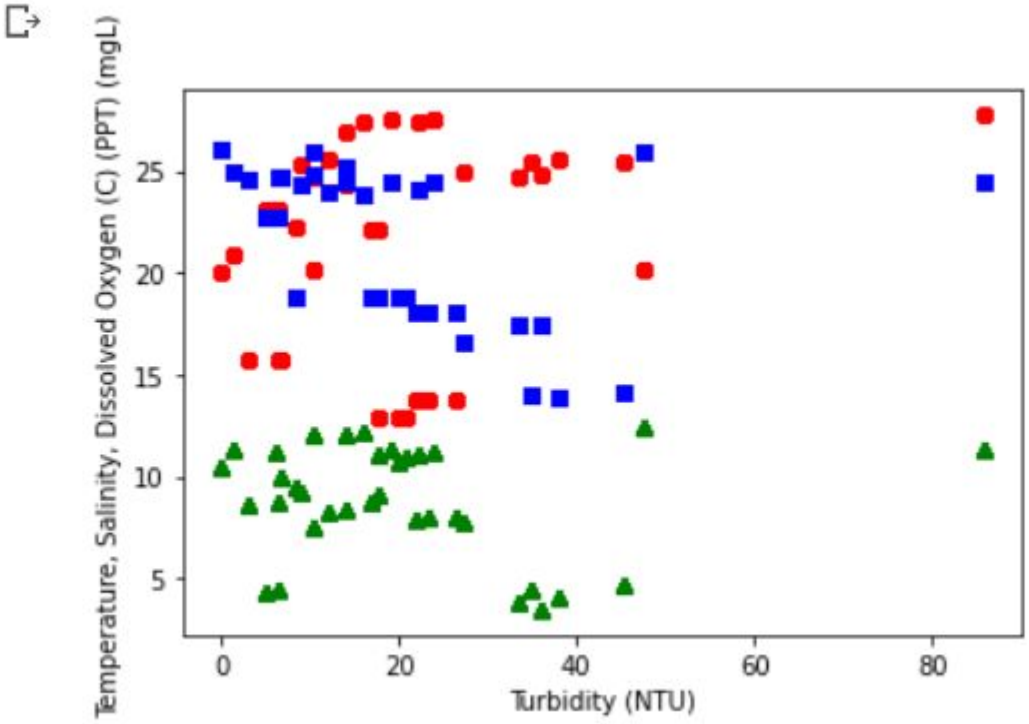
Imports

```python
import pandas as pd
import matplotlib.pyplot as plt
```

```python
import numpy as np
import pandas as pd

# Data Visualization Libraries
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

# Importing Predictive models
from sklearn import linear_model
from sklearn.model_selection import train_test_split
```

```python
from sklearn.linear_model import LinearRegression
from sklearn import metrics
```

Visual #1



- First visual represents slight trends of temperature, dissolved oxygen, and salinity (added to visualize a new trend) vs. turbidity.
- Green triangle: Dissolved Oxygen
- Red circle: Temperature
- Blue Square: Salinity
- Temperature shows an upward trend, increasing with turbidity.
- On the other hand, salinity seems to be decreasing with increasing turbidity.
- Dissolved Oxygen seems to have no real positive or negative trend which poses some questions.

Code snippets (Visual 1)

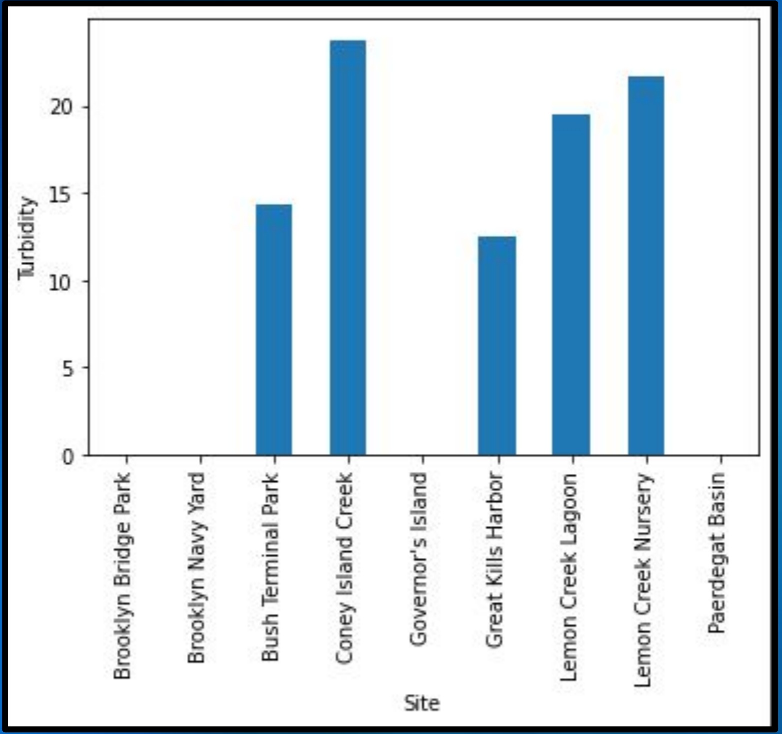
```
import pandas as pd
bop_data = pd.read_csv('/content/bop_data.csv')
bop_data.head()
```

Reading the excel file to access the data (used for all visuals)

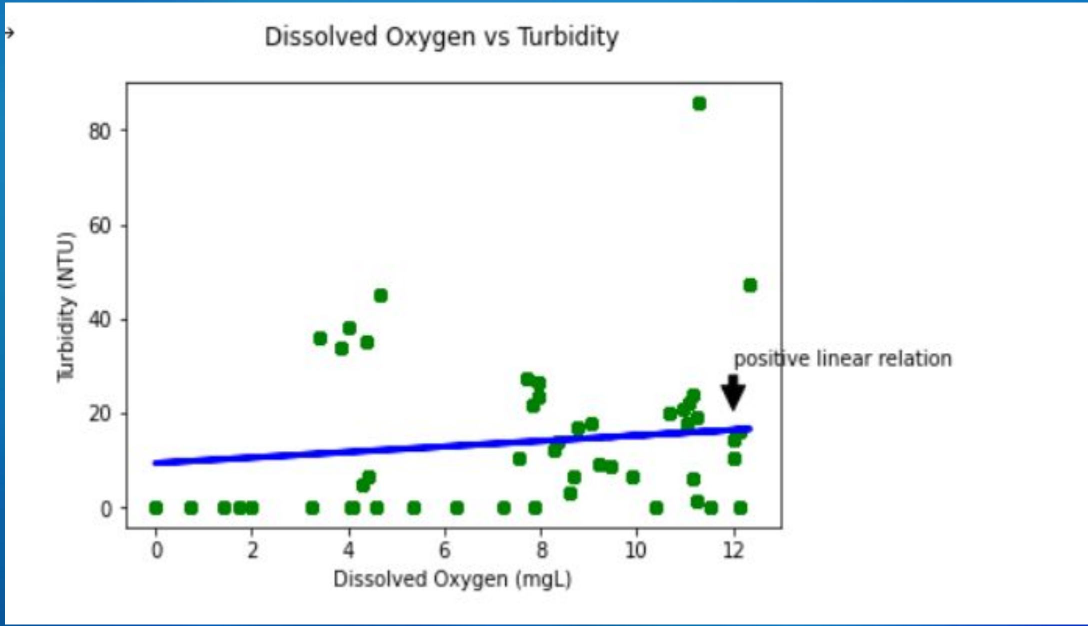
```
plt.plot(turbidity, temp, 'ro',  turbidity, salinity, 'bs', turbidity, oxy,'g^' )
plt.xlabel('Turbidity (NTU)')
plt.ylabel('Temperature, Salinity, Dissolved Oxygen (C) (PPT) (mgL)')

plt.show()
```

**Using the matplotlib methods to plot visual 1

Dissolved Oxygen and Turbidity Data Graphs (Visual #2)



Shows positive relation, contrary to hypothesis, rejecting one part of our hypothesis. Shows that average turbidity ranged from about 12 NTU to nearly 24 NTU

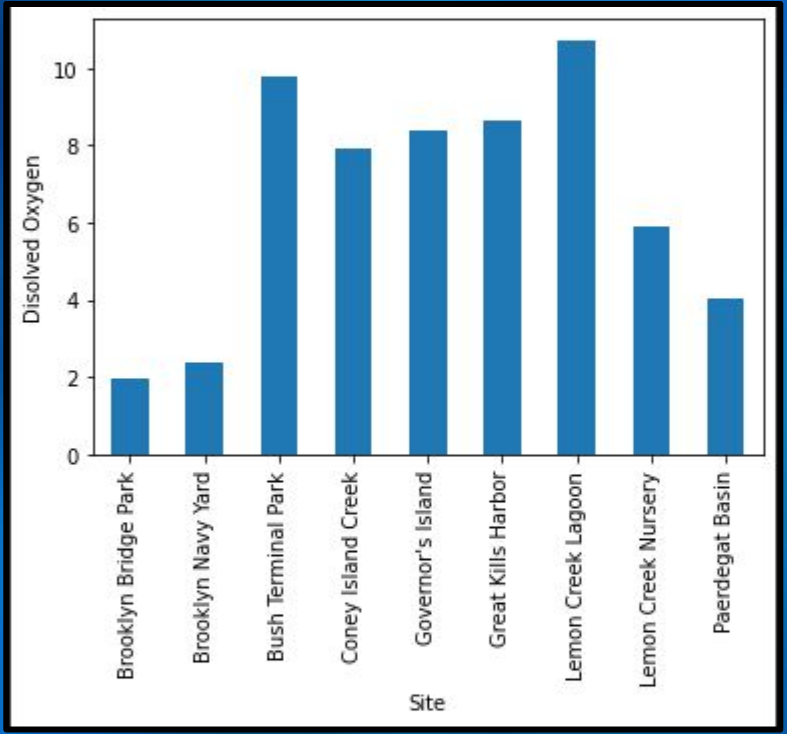


Code Snippets (Visual 2)

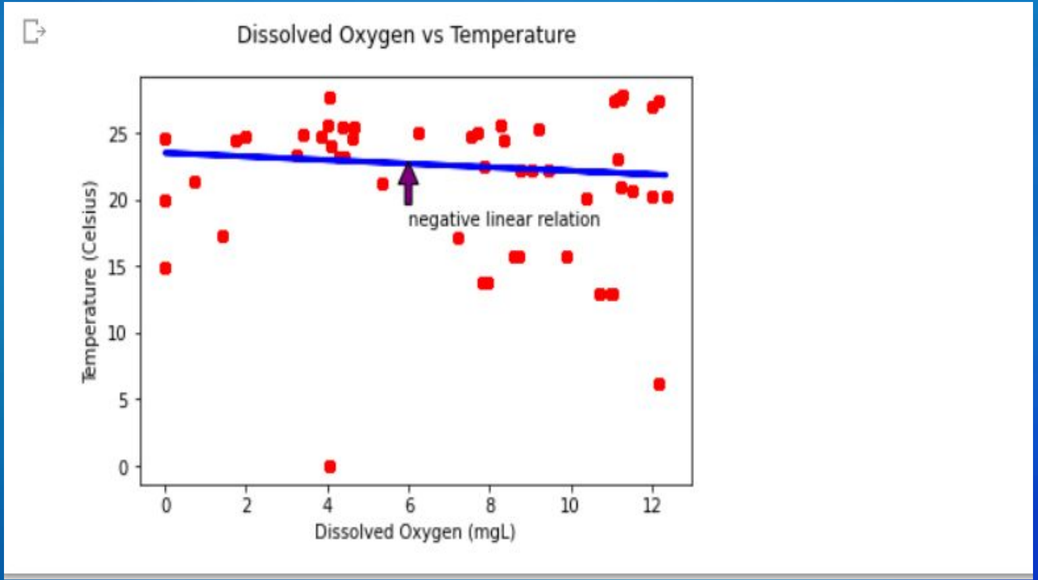
```
plt.scatter(X_test,y_test,color = 'green')
plt.plot(X_test, y_predict, color = 'blue', linewidth = 3)
plt.xlabel('Dissolved Oxygen (mgL)')
plt.ylabel('Turbidity (NTU)')
plt.suptitle("Dissolved Oxygen vs Turbidity")
plt.annotate('positive linear relation', xy=(12,20), xytext = (12,30), arrowprops = dict(facecolor = "black", shrink = 0.05),)
plt.show()
```

- Using matplotlib, we plotted a scatter plot for the test points of dissolved oxygen and turbidity.
- Then, we plotted a line plot for the predicted turbidity using the dissolved oxygen.
- We also added text to the annotation to show its relationship.

Dissolved Oxygen and Temperature Data Graphs



Showcases a negative relation as expected and the dissolved oxygen is shown to range from 2 to 12 mgL



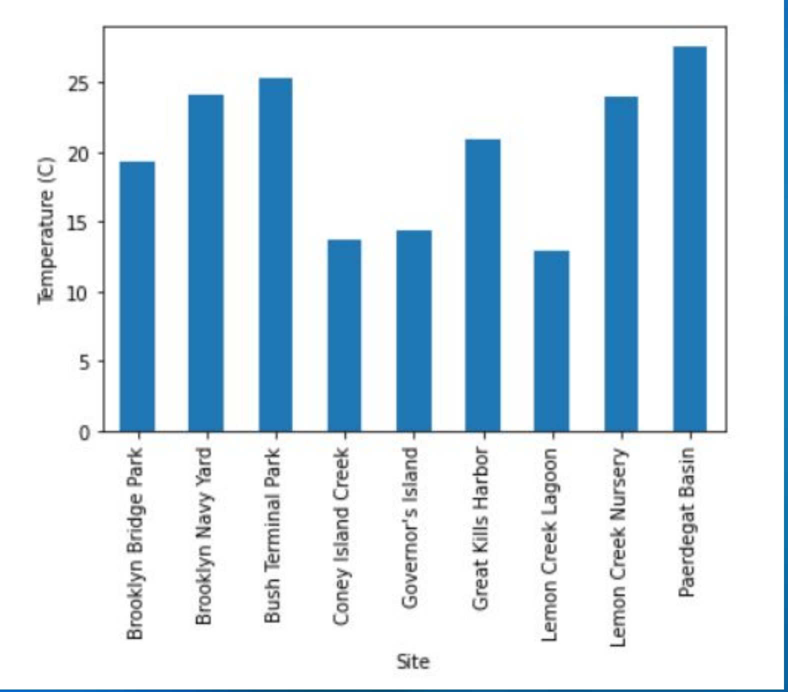
Code Snippet (Visual 3)

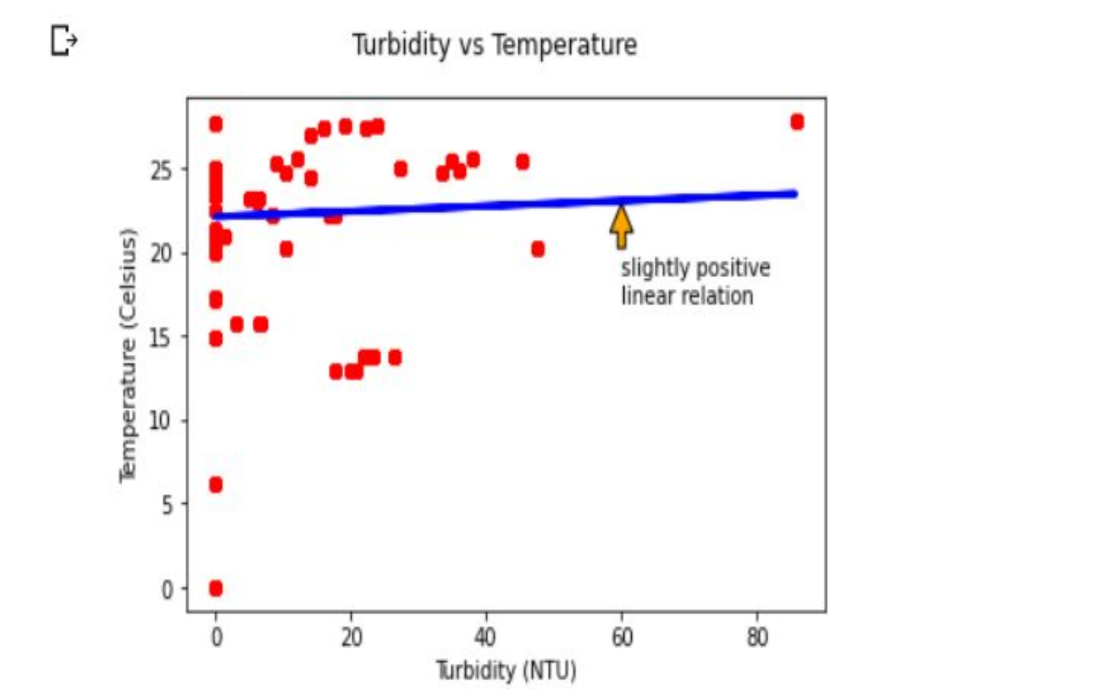
```
plt.scatter(X_test,y_test,color = 'red')
plt.plot(X_test, y_predict, color = 'blue', linewidth = 3)
plt.xlabel('Dissolved Oxygen (mgL)')
plt.ylabel('Temperature (Celsius)')
plt.suptitle("Dissolved Oxygen vs Temperature")
plt.annotate('negative linear relation', xy=(6,23), xytext = (6,18), arrowprops = dict(facecolor = "purple", shrink = 0.05),)
plt.show()
```

- Using matplotlib, we plotted a scatter plot for the test points of dissolved oxygen and temperature
- Then, we plotted a line plot for the predicted temperature using the dissolved oxygen.
- We also added text to the annotation to show its relationship.

Turbidity and Temperature Data Graphs (Visual #4)

Showcases a moderately positive correlation between turbidity and temperature and shows mean temperature to range from 12 C to around 27 C. Doubts on graph as relation seems to better fit through some log function.





Code Snippet (Visual 4)

```
plt.scatter(X_test,y_test,color = 'red')
plt.plot(X_test, y_predict, color = 'blue', linewidth = 3)
plt.xlabel('Turbidity (NTU)')
plt.ylabel('Temperature (Celsius)')
plt.suptitle("Turbidity vs Temperature")
plt.annotate('slightly positive \nlinear relation', xy=(60,23), xytext = (60,17), arrowprops = dict(facecolor = "orange", shrink = 0.05),)
plt.show()
```

- Using matplotlib, we plotted a scatter plot for the test points of temperature and turbidity.
- Next, we plotted a line plot for the predicted temperature using turbidity.
- We additionally added text to the annotation to show its relationship.

Google Colabs (Code)

- Visual 1: https://colab.research.google.com/drive/1fCcsLdt9FCsXbsVLJkMkOd4a56SP4E7e?usp=sharing
- DO vs Turb: https://colab.research.google.com/drive/1kV7SuJWkdG0-41dNiF2MQZsk2EcDUPQa?usp=sharing
- DO vs Temp: https://colab.research.google.com/drive/1rf4FUbvo9HKL7pNLDYY8yYe7tzKkDO7G?usp=sharing
- Turb vs Temp: https://colab.research.google.com/drive/12Cg7yEyDUz_U0-y6Veu9NtqCdzG72Ws2?usp=sharing

Average Mean Comparison

```
Site
Brooklyn Bridge Park          NaN
Brooklyn Navy Yard            NaN
Bush Terminal Park      14.362305
Coney Island Creek      23.766667
Governor's Island             NaN
Great Kills Harbor      12.489929
Lemon Creek Lagoon      19.466667
Lemon Creek Nursery     21.644076
Paerdegat Basin               NaN
Name: turb_NTU, dtype: float64
```

Turbidity

```
Site
Brooklyn Bridge Park    19.366328
Brooklyn Navy Yard      24.099453
Bush Terminal Park      25.354158
Coney Island Creek      13.758395
Governor's Island       14.440125
Great Kills Harbor      20.973044
Lemon Creek Lagoon      12.865763
Lemon Creek Nursery     23.974262
Paerdegat Basin         27.610210
Name: temp_C, dtype: float64
```

Temperature

```
Site
Brooklyn Bridge Park     1.968583
Brooklyn Navy Yard       2.355843
Bush Terminal Park       9.782980
Coney Island Creek       7.917000
Governor's Island        8.377481
Great Kills Harbor       8.619847
Lemon Creek Lagoon      10.722642
Lemon Creek Nursery      5.909456
Paerdegat Basin          4.050000
Name: DO_mgL, dtype: float64
```

Dissolved Oxygen

Conclusion

- Examining the mean data points of the table, we can see a relation forming with the temperature and turbidity.
  - The changes in the mean temperature and turbidity depend a lot upon the time of the year. For example, during the warm weather recordings of Bush Terminal Park (July-October), Great Kills Harbor (July), and Lemon Creek Nursery (August), the usual trend as shown by the graph of this relation (directly proportional) is visible. However, during the cold weather in the recordings of Lemon Creek Lagoon and Coney Island Creek, we see the opposite trend, pointing to a relationship scientists must keep in mind as they choose their sites.
- Examining the mean data points we can see a relations amongst temperature and dissolved oxygen as well.
  -