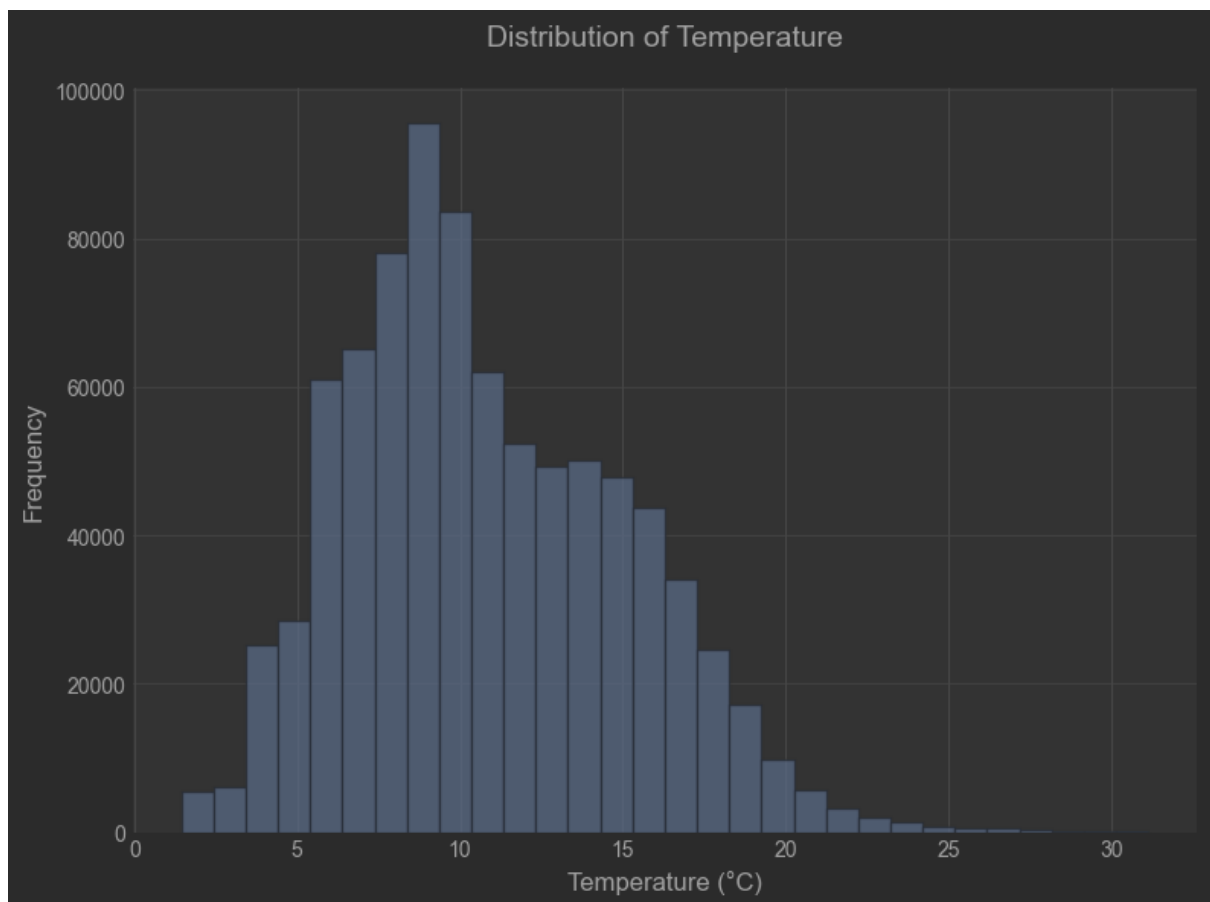


CP2403/CP3413: Assignment – Part 1 – 15% - Answer Template
Data Exploration, Management & Visualization
Due: End of Week 6 (Friday, 31 March 2023)

First Name: Kantapong
Last Name: Wongsanguan

Task 1: Histogram

Investigative Question: What is the distribution of seawater temperatures in the given dataset?
quantitative variable: Temperature (T_degC)



What is conclusion can you draw from the histogram?

Judging by the shape of this histogram, the distribution shows a unimodal skewed right with mean Temperature: 10.80 °C, median Temperature: 10.06 °C, mode Temperature: 9.10 °C, range Temperature: 29.70 °C, standard Deviation: 4.24 °C, maximum: 31.14 °C, and minimum: 1.44 °C.

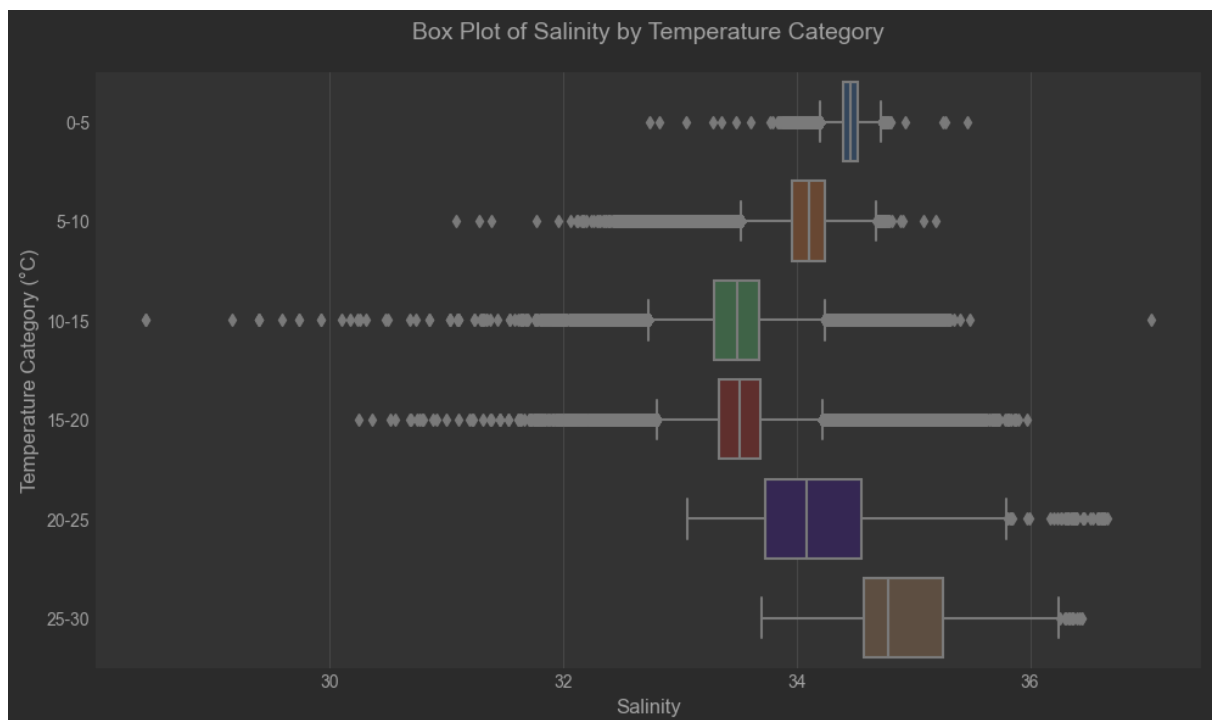
This means that there is a higher frequency of temperatures on the lower end of the scale, and the distribution tapers off towards the higher temperatures while the data's dispersion around the mean is approximately 4.24 °C, showing the extent of temperature variation in the dataset and the range of 29.70 °C, which indicates a considerable variability in the data.

Task 2:Box plot

Investigative Question: How does the salinity of ocean water vary across different temperature ranges in the dataset?"

categorical variable: Temperature (T_degC)

quantitative variable: Salinity (Salnty)



Summary Statistics

| Summary Statistics: | | | | | | | | | |
|----------------------|---------------|-----------|----------|-----------|-----------|----------------------|-----------|-----------|-----------|
| | | | | | | 50% | 75% | max | |
| Temperature Category | | | | | | Temperature Category | | | |
| 0-5 | count | mean | std | min | 25% | 0-5 | 50% | 75% | max |
| 0-5 | 45126.000000 | 34.455803 | 0.127214 | 32.740000 | 34.400000 | 0-5 | 34.467000 | 34.530000 | 35.462000 |
| 5-10 | 354196.000000 | 34.075892 | 0.260906 | 31.090000 | 33.958000 | 5-10 | 34.107000 | 34.249000 | 35.190000 |
| 10-15 | 263204.000000 | 33.537885 | 0.421095 | 28.431000 | 33.300000 | 10-15 | 33.490000 | 33.679000 | 37.034000 |
| 15-20 | 135911.000000 | 33.565427 | 0.408638 | 30.250000 | 33.337000 | 15-20 | 33.511000 | 33.690000 | 35.970000 |
| 20-25 | 14076.000000 | 34.178384 | 0.557595 | 33.062500 | 33.730000 | 20-25 | 34.090000 | 34.560000 | 36.658000 |
| 25-30 | 1698.000000 | 34.883948 | 0.433533 | 33.699000 | 34.580000 | 25-30 | 34.790000 | 35.250750 | 36.442000 |

What is conclusion can you draw from the box plot?

The box plot provides insights into the distribution of salinity across different temperature categories with highlighted central tendency, variability, and potential trends in the data.

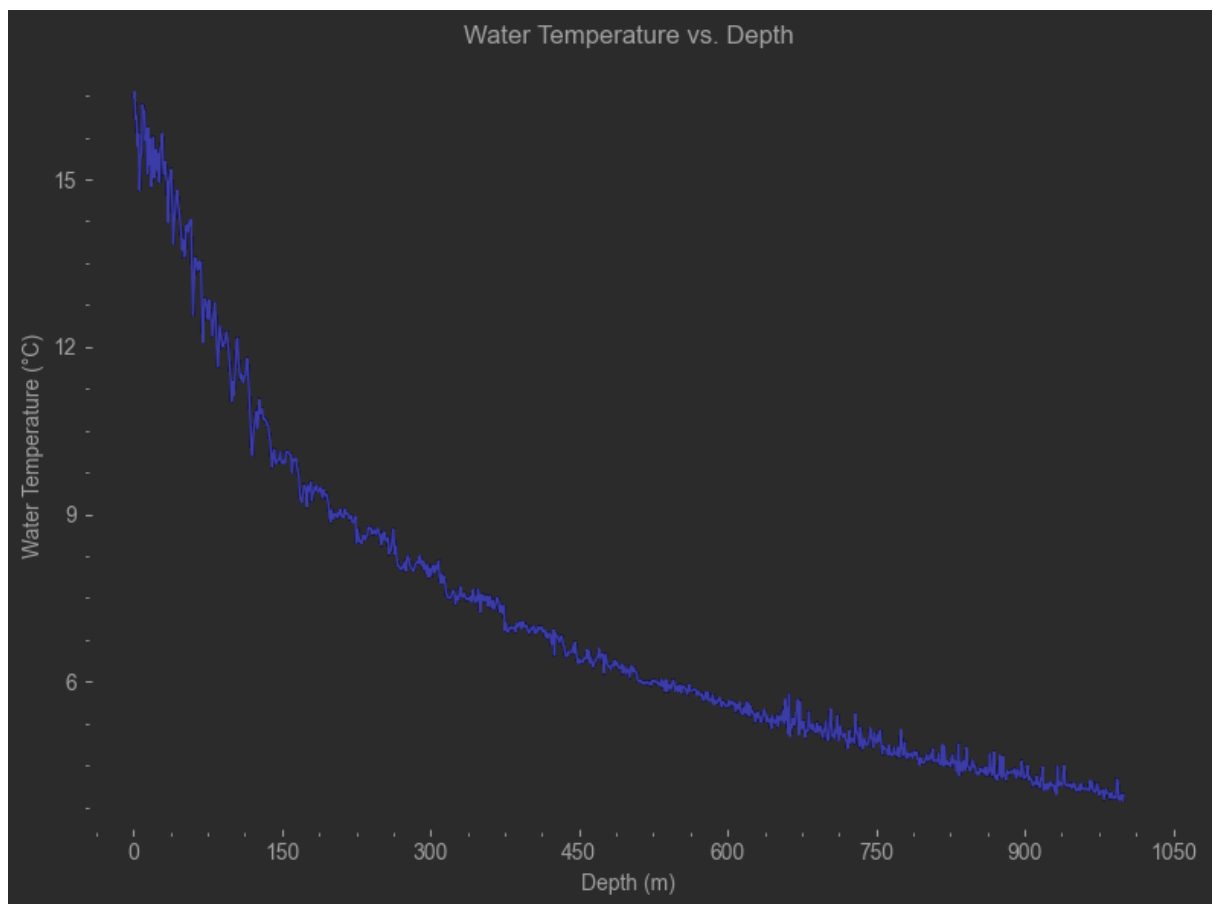
Overall, there seems to be a correlation between temperature and salinity. As the temperature category increases, the median salinity tends to increase as well, from

approximately 34.46 at 0-5°C to around 34.88 at 25-30°C, while the sample size gradually decreases as the temperature increases

Task 3: Line chart

Investigative question: How does the temperature vary with the depth, and what patterns can be observed from the line chart?

quantitative variable: Temperature (T_degC)



What is conclusion can you draw from the line chart?

The line chart represents a gradual decline relationship in temperature as the depth increases. While the overall trend is a steady decline, there are some minor fluctuations in temperature at specific depths, however, as there are no erratic spikes or outliers which indicates measurement error or data anomalies, the overall trend of the line chart suggests that the data is reasonably reliable and consistent.

Hence, as the depth increases, the temperature decreases, resulting in a negative slope of this line chart, which also is the evident of an inverse relationship.

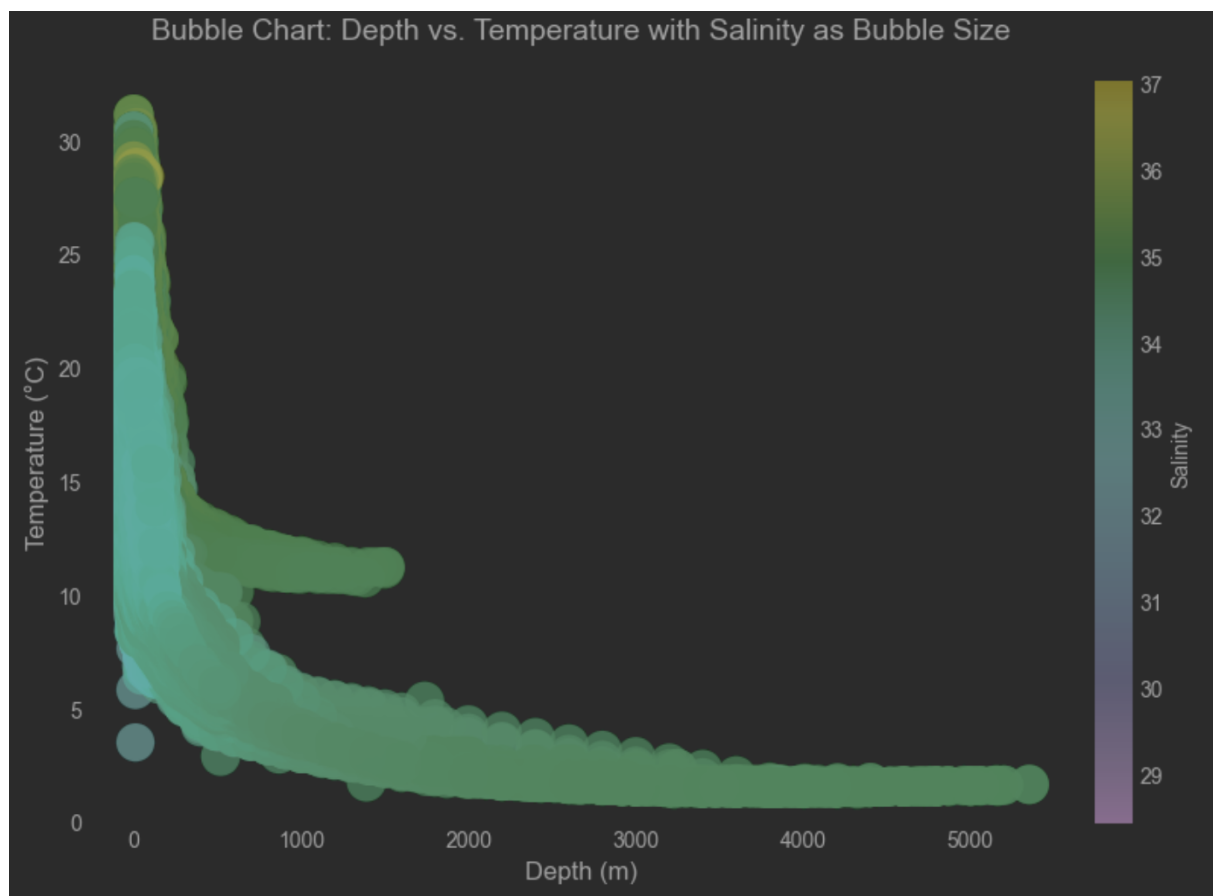
Task 4: Bubble.

Investigative Question: How does the depth of the ocean (measured in meters) relate to the temperature (in °C) and salinity of the water, as depicted in the bubble chart? Is there any observable pattern or correlation between depth, temperature, and salinity based on the size and color of the bubbles in the chart?

quantitative variable 1: Temperature (T_degC)

quantitative variable 2: Salinity (Salnty)

quantitative variable 3: Depth (Depthm)



What is conclusion can you draw from the bubble chart?

This bubble chart shows the relationship between the depth and the temperature. Judging by the clusters of the bubbles in multiple area, we can tell that

- The cluster of bubbles on the left side along the y-axis, temperature, where the depths coordination is a lot closer to 0 than 1000, suggests the presence of shallow waters area with vary temperatures. This could be the result of the nature of that specific regions where temperature fluctuations can be affected due to environmental factors, for example, high sunlight exposure, coastal influences, or changing weather. Thus, this suggests vertical temperature gradient.
- The cluster of bubbles along the x-axis represented by the depth where the temperature are mostly under 5°C suggests the presence of cold waters at various depths. This could be the result indicating of ocean currents where cold water flows to deeper part of the depth range or where specific cold water masses appear. Thus, this suggests horizontal temperature gradient.

In addition, this bubble chart also indicates the salinity relation ship with temperature and depth influences. The color of the salinity levels shown on the legends on the right side indicates high and low level of salinity in various temperature and depth.

Hence, looking at the bubble chart, when the depth increase, the color of the salinity tends to not change, this indicates that the salinity characteristic does not get affected by the influence of the depth. Whereas, as the temperature increases as shown on the left side of the chart, the color of the bubbles tend to varies more, thus representing temperature-salinity relationship where salinity changes with temperature, however it is not consistence since at the area where temperature is 10 – 25 °C the salinity are at color level of cyan but increases as the temperature of the water leaves that area however up or down.

In conclusion, this chart shows how depth influences both temperature and salinity and vice versa, where it suggests specific depth zones with unique temperature and salinity characteristics.

Task 5: Selected Chart.

Selected Chart: the Pairwise Plot

Investigative Question: Is there any observable relationship between the concentrations of oxygen, nitrate, and phosphate in seawater samples, and are there any potential patterns or trends among these variables?

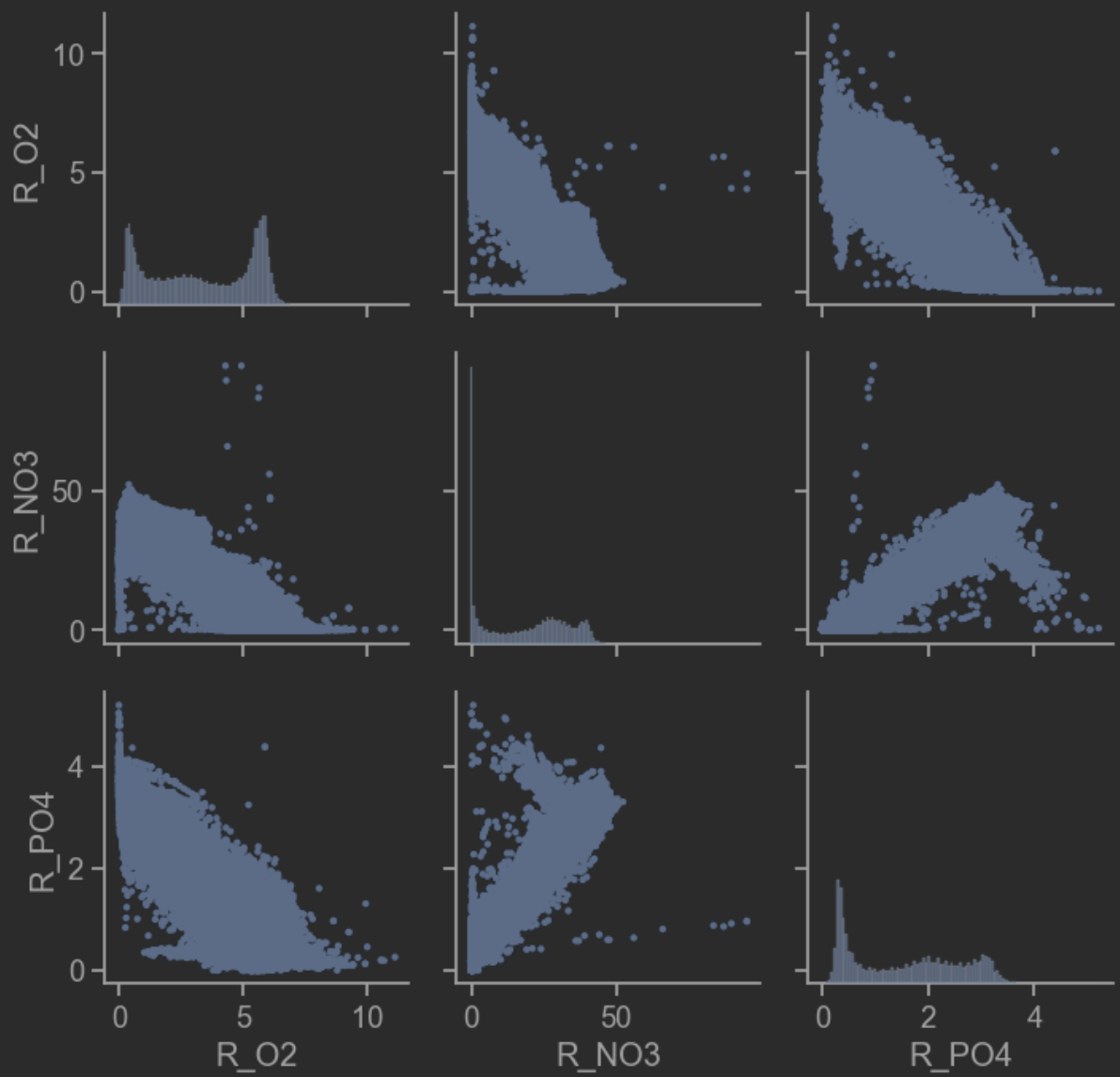
Variables used

Variable 1: R_O2, Reported millilitres of oxygen per Liter of seawater

Variable 2: R_NO3, Reported Nitrate Concentration

Variable 3: R_PO4, Reported Phosphate Concentration

Pairwise Plot: Oxygen, Nitrate, and Phosphate



Summary Statistics:

| | R_O2 | R_NO3 | R_PO4 |
|-------|---------------|---------------|---------------|
| count | 696201.000000 | 337411.000000 | 413325.000000 |
| mean | 3.392498 | 17.302188 | 1.564736 |
| std | 2.073271 | 14.581998 | 1.035973 |
| min | -0.010000 | -0.400000 | 0.000000 |
| 25% | 1.360000 | 0.600000 | 0.470000 |
| 50% | 3.440000 | 18.100000 | 1.570000 |
| 75% | 5.500000 | 30.000000 | 2.480000 |
| max | 11.130000 | 95.000000 | 5.210000 |

Correlation Matrix:

| | R_O2 | R_NO3 | R_PO4 |
|-------|-----------|-----------|-----------|
| R_O2 | 1.000000 | -0.969969 | -0.973389 |
| R_NO3 | -0.969969 | 1.000000 | 0.987369 |
| R_PO4 | -0.973389 | 0.987369 | 1.000000 |

What is conclusion can you draw from your selected chart?

The reason why a Pairwise Plot is chosen is due to that the variables chosen needed to be compared and explored continuously. In this case, the Pairwise Plot helps analyzing how these "Oxygen (R_O2)", "Nitrate (R_NO3)", and "Phosphate (R_PO4)" concentrations in the seawater relate to each other.

The conclusion that can be drawn from the Pairwise plot is that there seems to be a weak negative correlation between the relationship of Oxygen (R_O2) and Nitrate (R_NO3) which suggests that as the Nitrate (R_NO3) concentration decreases, the Oxygen (R_O2) concentration tends to increase. Thus, suggesting a potential dynamic balance in the marine ecosystem, where higher oxygen levels might promote the utilization or uptake of nitrate by marine organisms.

While comparing Oxygen (R_O2) with Phosphate (R_PO4), there does not appear to be a strong relationship between these concentrations so as there is no clear linear relationship between "Nitrate (R_NO3)" and "Phosphate (R_PO4)" concentrations. This means that the lack of correlation shown suggests that the concentration of oxygen and nitrate may not directly affect the concentration of phosphate in the seawater samples.

To conclude, the visualization helps to visually explore and address this question, supporting a deeper understanding of the data and potential ecological implications between these three variables, however, when compared, this has shown that additional analyses and domain knowledge seems to be needed to further interpret the ecological significance of the relationship of these concentrations in the seawater.