Analysis of 2016 Coral Bleaching Data

Julia R. Piper, Dr. Khyam Paneru (Mentor)

The University Of TAMPA

Introduction

- Coral reefs are the 2nd most diverse habitat on earth and are mainly built by corals.[1]
- Limiting factors include sea temperature, high light, low turbidity, low nutrient, ocean salinity, and pH range.[2]
- Coral bleaching and disease are a threat to coral reefs. [3]
- Corals are from phylum Cnidaria and class Anthozoa.[1]
- Corals can sequester the dinoflagellate zooxanthellae to create a symbiotic relationship. [3]
- Zooxanthellae can provide a coral with up to 90% of its nutrient requirements.[1]
- Coral Bleaching is a stress response in corals caused from condition changes such as temperature, light, or nutrients. [2]
- The zooxanthellae will begin to produce reactive oxygen molecules that can become toxic.[3]
- This forces the coral to expel the zooxanthellae for self-preservation; causes the coral to appear white and is considered coral bleaching.[3]
- A bleached coral that can not re-sequester zooxanthellae will die. [3]
- While the 2014-15 El Niño never fully formed it began a three-year global bleaching event.[4]
- A subsequent El Niño formed in 2015-16 worsening the bleaching event.[4]
- In 2017, a La Niña event that further extended coral bleaching.[4]
- Worldwide coral reef ecosystems experiences extreme bleaching. [4]
- This was the longest, most widespread, and potentially most damaging coral bleaching event in history.[4]

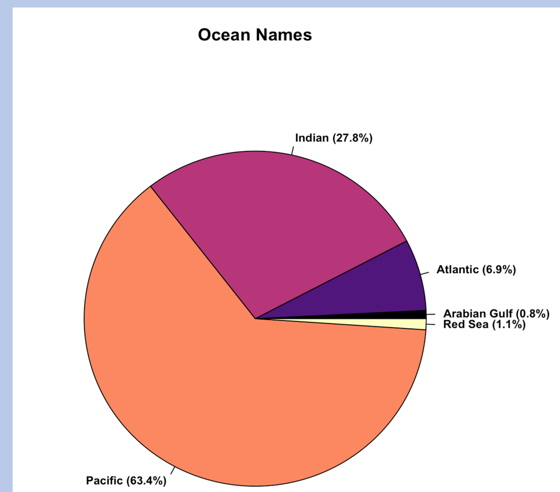

Figure 2A: Distribution of observations between oceans

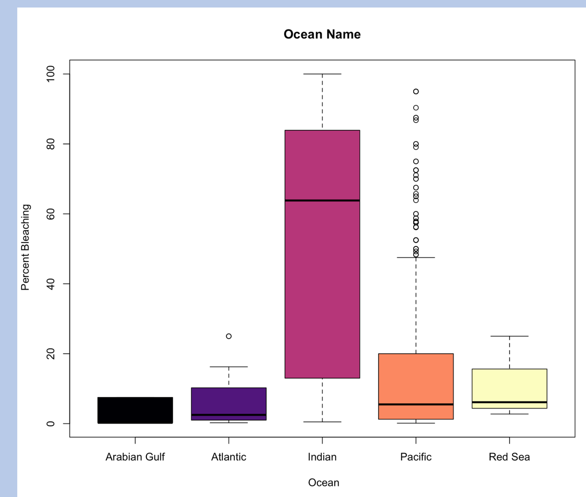

Figure 2B: *Box plot examining Percent Bleaching observations between oceans*

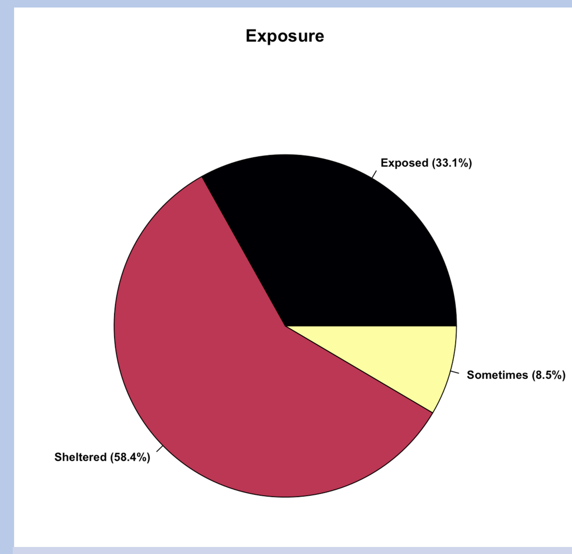

Figure 3A: Distribution of observations *between Exposure type*

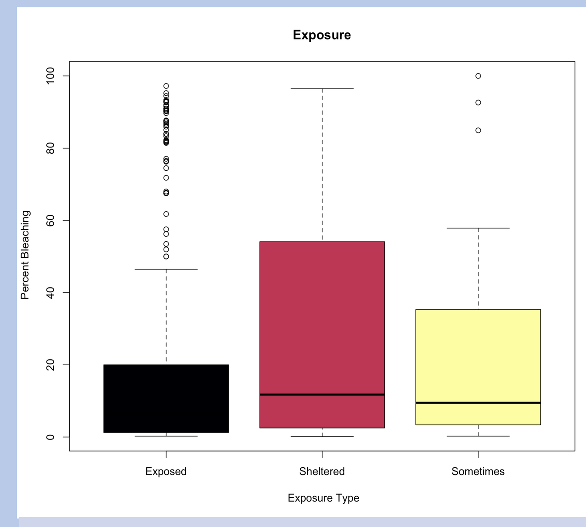

Figure 3B: *Box plot examining Percent Bleaching observations between Exposure*

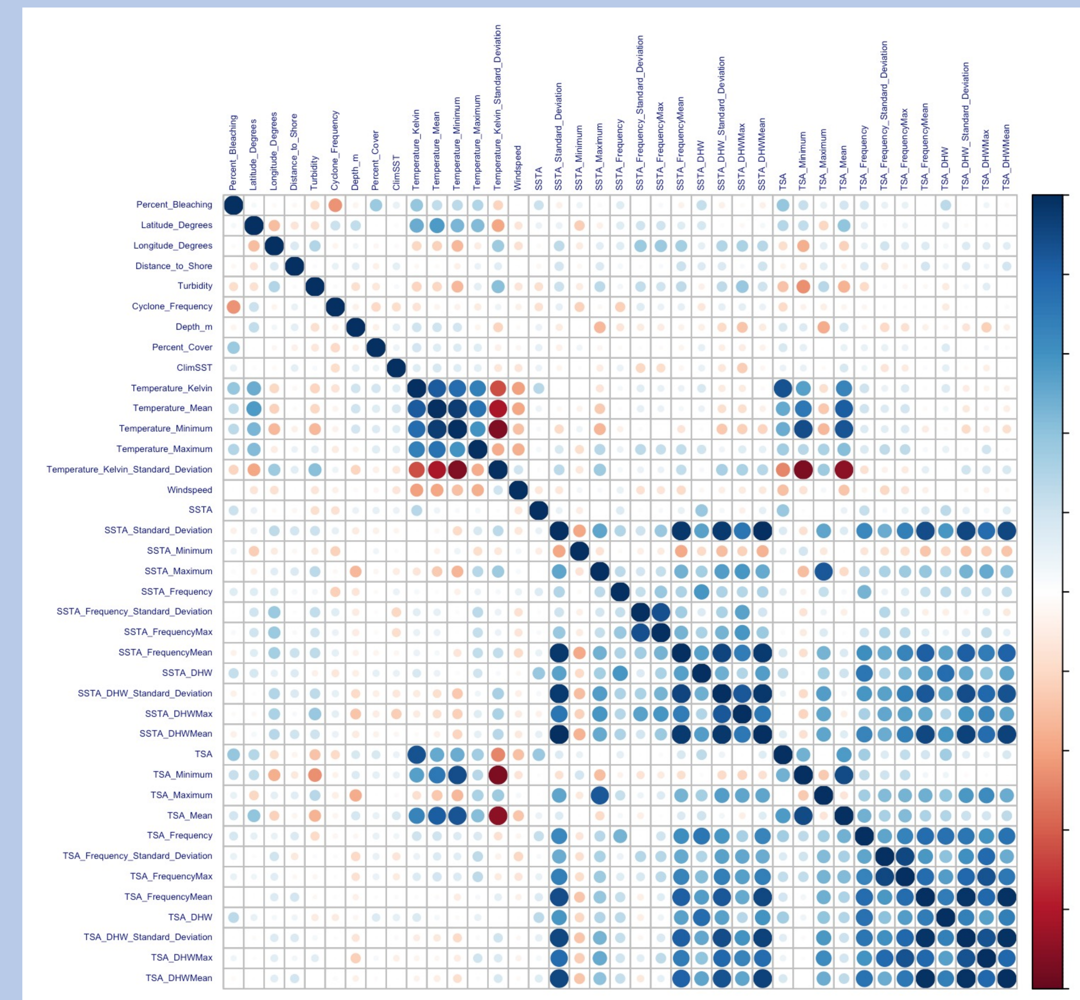

Figure 4A: *Correlation plot for all variables. Note high correlation between summaries for variables.*

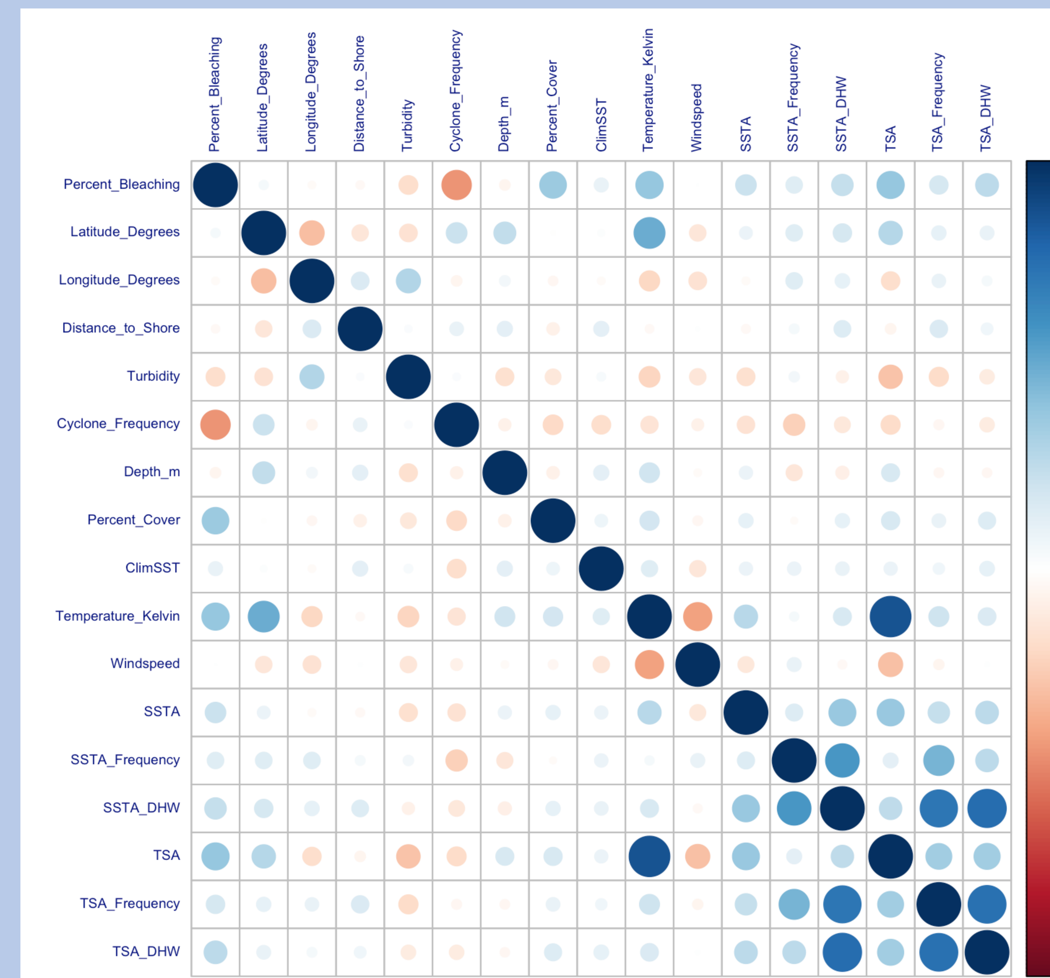

Figure 4B: *Correlation plot for remaining variables. Note high correlations found in TSA, TSA Frequency, TSA DHW, Temperature, SSTA Frequency, and SSTA DHW.*

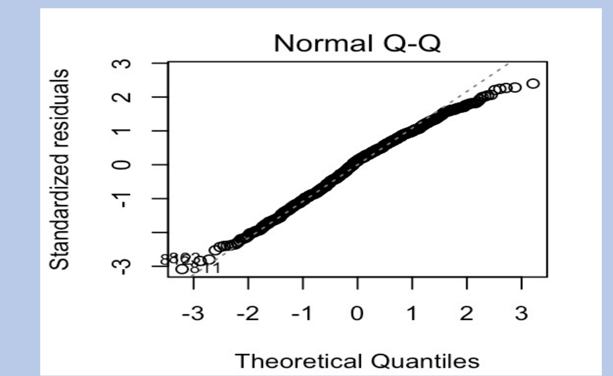

Figure 8: Examining normality in untransformed model

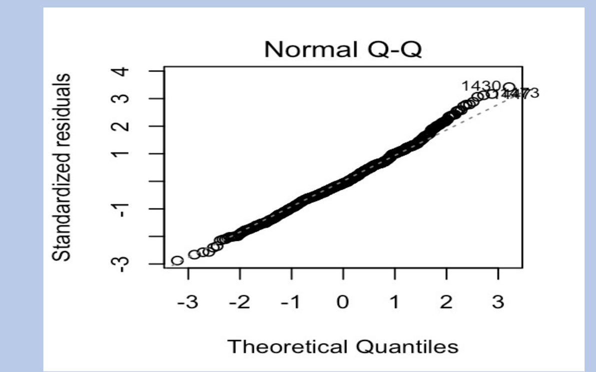

Figure 9: Examining normality in log transformed model

Final Model

$$\widehat{LogPercentBleaching} = -49.709 + 0.348(Sometimes) + (-0.140(Exposed)) + 1.071(Indian) + 0.694(Atlantic) + (-1.016(ArabianGulf)) + 0.869(RedSea) + (-0.013(Latitude)) + 0.003(Longitude) + (-3.858(Turbidity)) + (-0.029(Cyclone\ Frequency)) + (-0.048(Depth)) + 0.009(Percent\ Cover) + 0.176(Temperature) + 0.044(SSTA\ DHW)$$

Adjusted r^2 value of 0.4378

$p < 0.001$

Methods

- Dataset begin used for this project is the Coral-bleaching database, 1980-2020 created by Robert van Woesik and Chelsey Kratochwill. The original dataset includes the compilation of 34,846 coral bleaching records from 14,405 sites and 93 countries.[5]
- Only 2016 data was used. All rows with a percent bleaching of 0% were removed. This reduced the data to 755 observations
- Models were created to predict the percentage of coral bleaching based on bleaching being present.
- Histograms were created to examine the prominence of different percentages of Bleaching (Figure 1)
- Pie were made to examine the categorical variables, Ocean Name and Exposure, distribution among observations (*Figure 2A* (Ocean Name) and *3A*(Exposure)).
- Categorical variables were compared to the dependent variable (*Figure 2B* (Ocean Name) and *Figure 3B* (Exposure))
- Correlation plots were examined to avoid multicollinearity (*Figure 4A*). These plots lead to the removal of all summary variables , as well as all TSA related variables and SSTA frequency (*Figure 4B*).
- A multiple regression model was run without categorical variables, this model violated assumptions, indicating that categorical variables must be included in the model.
- Categorical variables, Ocean Name and Exposure were converted into dummy variables and multiple regression was ran. Diagnostic plots of this model indicated that regression assumptions were being violated and a transformation of the data would be required.
- A log transformation was performed on the dependent variable in the data, once the model was rerun, the new diagnostic plots better fit the regression assumptions.
- Distribution of the log transformed dependent variable was more evenly distributed (*Figure 5)*.
- Using diagnostic plots, 6 outliers identified by Cooks Distance in the Residuals vs Leverage plot, were removed.

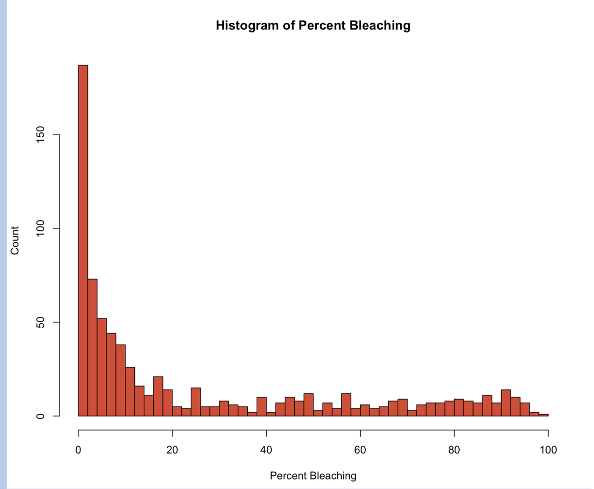

Figure 1 :Histogram of Percent Bleaching prominence

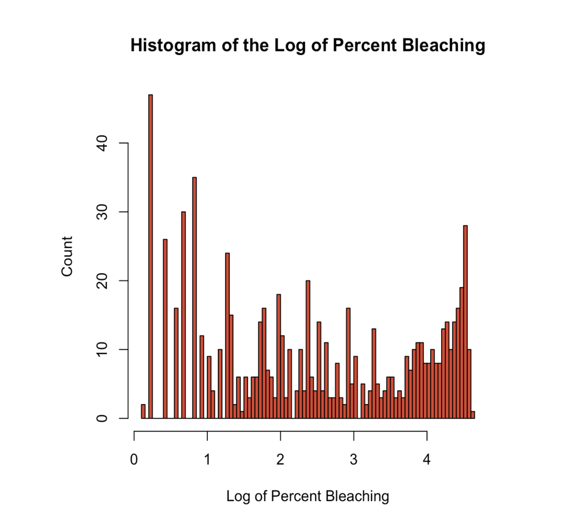

Figure 5 :Histogram of Log Transformation Percent Bleaching prominence

Predictor Variable Details[5]	Variable	Coefficient	Std. Error	t-value	p-value	Significance
	(Intercept)	-989.3677	151.3715	-6.536	<0.001	***
Sample location sometimes exposed	Sometimes	2.4959	3.0939	0.807	0.4200	
Sample location always exposed	Exposed	-2.2172	1.8973	-1.169	0.2429	
Sample collected in Indian Ocean	Indian	26.6293	2.2946	11.605	<0.001	***
Sample collected in Atlantic Ocean	Atlantic	11.0437	5.0488	2.187	0.0290	*
Sample collected in Arabian Gulf	Arabian Gulf	-14.2768	9.5501	-1.495	0.1354	
Sample collected in Red Sea	Red Sea	4.6530	8.8869	0.524	0.6007	
latitude coordinates in decimal degrees.	Latitude	-0.1339	0.0867	-1.545	0.1228	
longitude coordinates in decimal degrees	Longitude	0.0529	0.0151	3.493	<0.001	***
kd490 with a 100-km buffer.	Turbidity	-35.6841	15.6766	-2.276	0.0231	*
number of cyclone events	Cyclone Frequency	-0.7444	0.1544	-4.822	<0.001	***
depth (m) of sampling site	Depth	-0.4093	0.2614	-1.566	0.1178	
Percentage of sample site covered	Percent Cover	0.2283	0.0361	6.327	<0.001	***
Sea-Surface Temperature (SST) in Kelvin	Temperature	3.4279	0.4956	6.916	<0.001	***
(Sea Surface Temperature Degree Heating Weeks) sum of previous 12 weeks when SSTA > = 1 degree Celsius.	SSTA_DHW	0.7256	0.1734	4.185	<0.001	***

Table 2A : Summary of the fitted model
Significance: |* p<0.05 | ** p<0.01 | *** p<0.001 |

Variables	Coefficients	Std. Error	t-value	p-value	Significance
Intercept	-49.7094	7.5203	-6.610	<0.001	***
Sometimes	0.3477	0.1523	2.283	0.0227	*
Exposed	-0.1403	0.0940	-1.492	0.1361	
Indian	1.0708	0.1138	9.410	< 0.001	***
Atlantic	0.6939	0.2497	2.779	0.0056	**
Arabian Gulf	-1.0165	0.5123	-1.984	0.0476	*
Red Sea	0.8690	0.4399	1.975	0.0486	*
Latitude	-0.0131	0.0043	-3.048	0.0024	**
Longitude	0.0035	0.0007	4.645	< 0.001	***
Turbidity	-3.8584	0.7729	-4.992	< 0.001	***
Cyclone_Frequency	-0.0290	0.0076	-3.801	0.0002	***
Depth	-0.0477	0.0129	-3.701	0.0002	***
Percent_Cover	0.0089	0.0018	4.941	< 0.001	***
Temperature	0.1758	0.0246	7.133	< 0.001	***
SSTA_DHW	0.0440	0.0096	4.603	< 0.001	***

Table 2B: Summary of the fitted model after log transformation
Significance: |* p<0.05 | ** p<0.01 | *** p<0.001 |

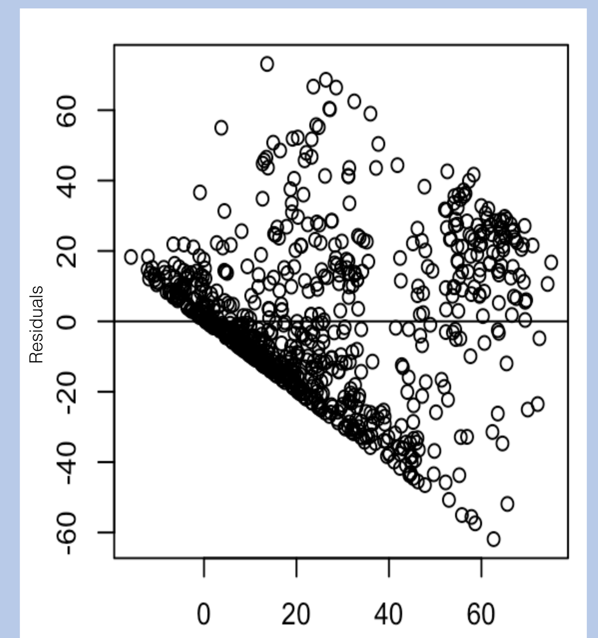

Figure 6: Residuals vs Fitted plot for untransformed model.

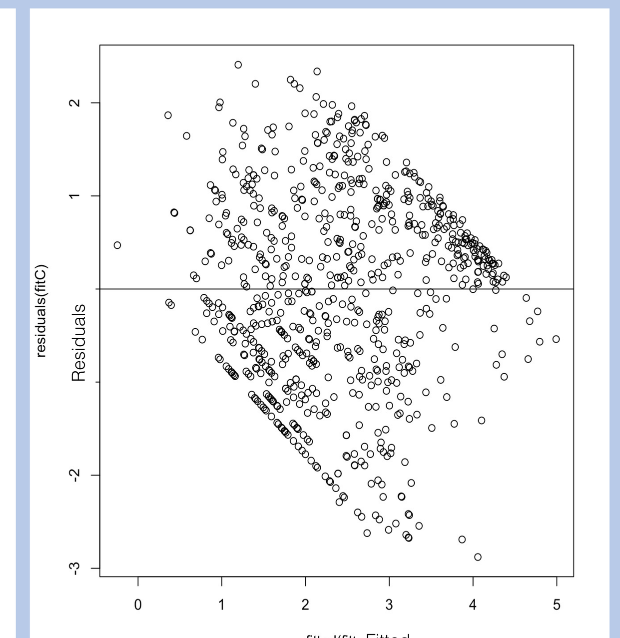

Figure 7: Residuals vs Fitted plot for log transformed model.

Categorical Dummy Variables

Ocean Name		Exposure	
Indian = 1	else=0	Exposed = 1	else =0
Atlantic = 1	else=0	Sometimes = 1	else =0
Arabian Gulf = 1	else=0	Sheltered = 0	else =0
Red Sea = 1	else=0		
Pacific =0	else=0		

Table 1: Creation of categorical dummy variables for model creation.

Results/Discussion

- Pie charts of the categorical variables indicate that the data is not evenly distributed. 63.4% of the data is from the Pacific Ocean and only 0.8% of data is from the Arabian Gulf (*Figure 2A).* 58.4% of the data is Sheltered and 8.5% of the data is Sometimes exposed (*Figure 3A). The* highest rates of bleaching occurred in the Indian Ocean(*Figure 2B)* and with sheltered corals(*Figure 3B).*
- Correlation of the original variables (*Figure 4A*) showed high multicollinearity between summary variables. After removal, another correlation matrix (*Figure 4B)* was created. Variables with a correlation greater than 50% were investigated. Remaining variables are Exposure, Ocean Name, Distance to Shore, Turbidity, Cyclone Frequency, Depth, Temperature Kelvin, Windspeed, Percent Cover, and SSTA DHW.
- Histogram of the Percent Bleaching data displayed uneven distribution of the data with a majority of data being between 0.01 and 10% (*Figure 1)*
- *Table 1* showed conversion of categorical variables into numerical variables so that regression models can created.
- *Table 2A* shows the summary of the regression model, indicating that Latitude, Depth, and Exposure were insignificant.
- Regression Diagnostics implied that the original model violated the constant variance and normality assumption(*Figure 6).*
- The histogram of the Log Transformed Percent Bleaching data(*Figure 5)* is more evenly distributed than the untransformed data (*Figure 1).*
- Categories from the categorical variables differed in significance. This may be related to the uneven distribution of categorical variables.
- Table 2B, the summary of the transformed regression model indicates that all the variables are significant, except for the category Exposed in the categorical variable Exposure.
- The transformed model better follows the regression assumptions of constant variance, (*Figure 6, 7)* and normality (*Figure 8, 9*).

Conclusion

- Longitude, Turbidity, Cyclone Frequency, Depth, Percent Cover, Temperature, and SSTA DHW were the most significant predictor variables in determining the percent bleaching of an area.
- Although the adjusted R^2 value was less than 0.5, multiple regression still identified significant predictor variables and resulted in a significant predictive model.

References

1. Brusca, R. C., Moore, W., & Shuster, S. M. (2016). Invertebrates (3rd ed.). Sinauer Associates, Inc.
2. NOAA's National Ocean Service. (n.d.). Coral Bleaching.
3. Science Education Resource Center. (n.d.). Coral Bleaching.
4. NOAA's Coral Reef Watch. (n.d.). Global Coral Bleaching 2014-2017 Status.
5. Chen, X., Wu, Q., Wu, Y., Chen, Y., Huang, Z., & Xie, J. (2022). Performance evaluation of metagenome classifiers on human gut microbiota datasets with simulated and real reads. Scientific Data, 9(1), 1-9.