

Coral Reef Bleaching

By:
Mathurin, Gregory
Ko, Graeme
Al-Saidi, Ray



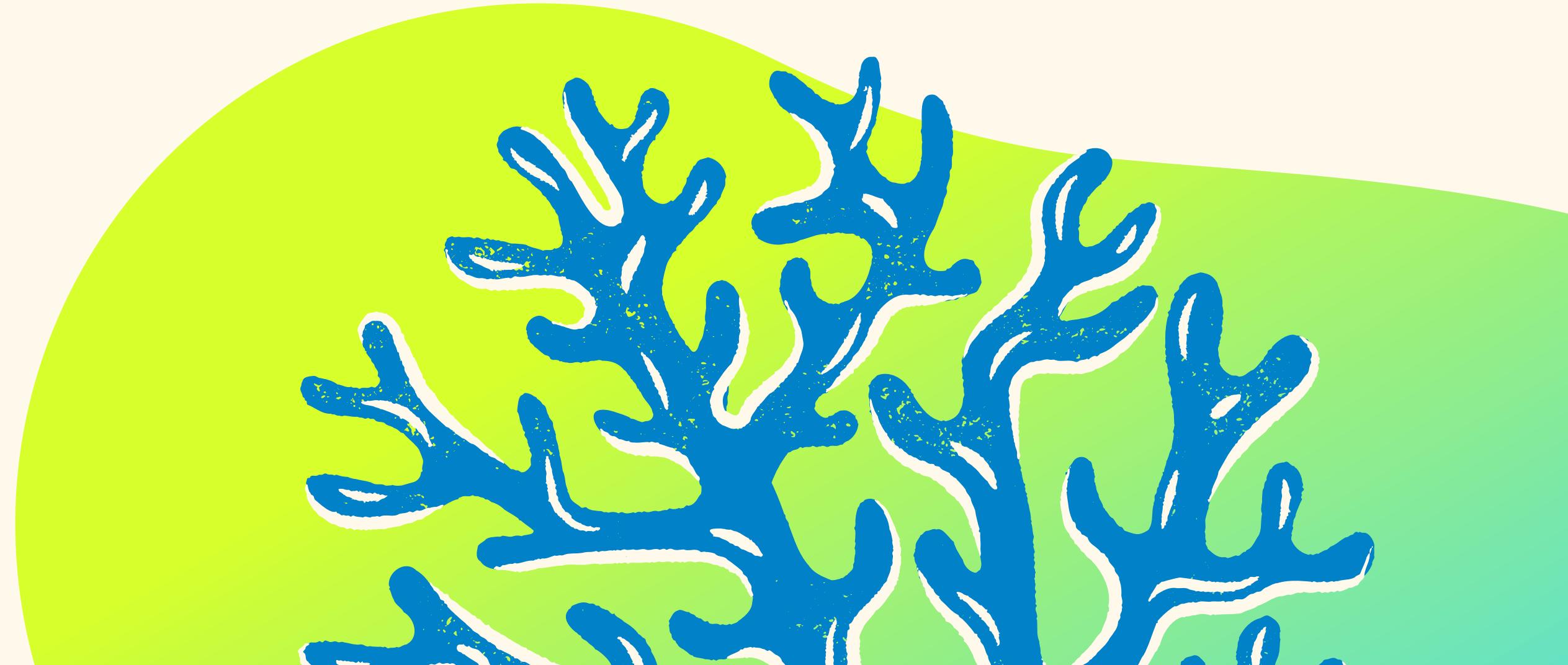
Agenda

- Introduction
- Objectives
- Dataset and cleaning
- Linear Regression
- Linear Discriminant Analysis
- Logistic Regression
- Conclusion



Objectives

1. Investigate which areas are impacted the most by coral bleaching
2. Perform variable analysis to identify significant predictors.
3. Develop a Model to Forecast Future Coral Bleaching Events

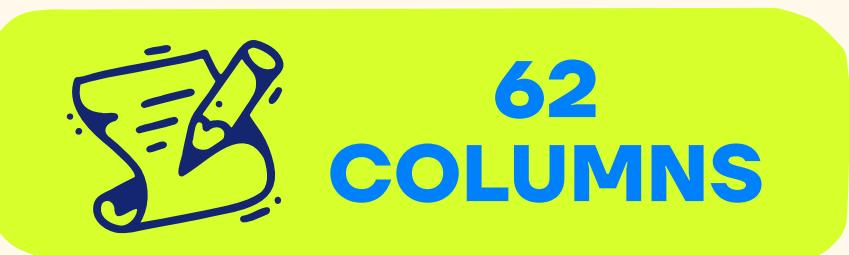


Dataset

- Published 2020
- Bleaching Percentage is our target variable
- Final number of rows after cleaning: 32714
- Number of rows lost after cleaning: 8647



Some NA values



Categorical, spatial, nominal and numerical variables.

47 VARIABLES REMOVED

A red rounded rectangle containing the text "47 VARIABLES REMOVED" in white.

Variables not relevant to our analysis or high NA count, such as Site ID

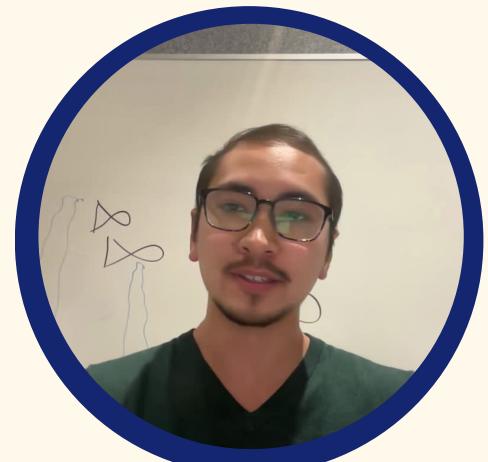


Dataset

- Removed highly correlated predictors (>0.9)
- Removed multicollinear variables (VIF > 3.5)

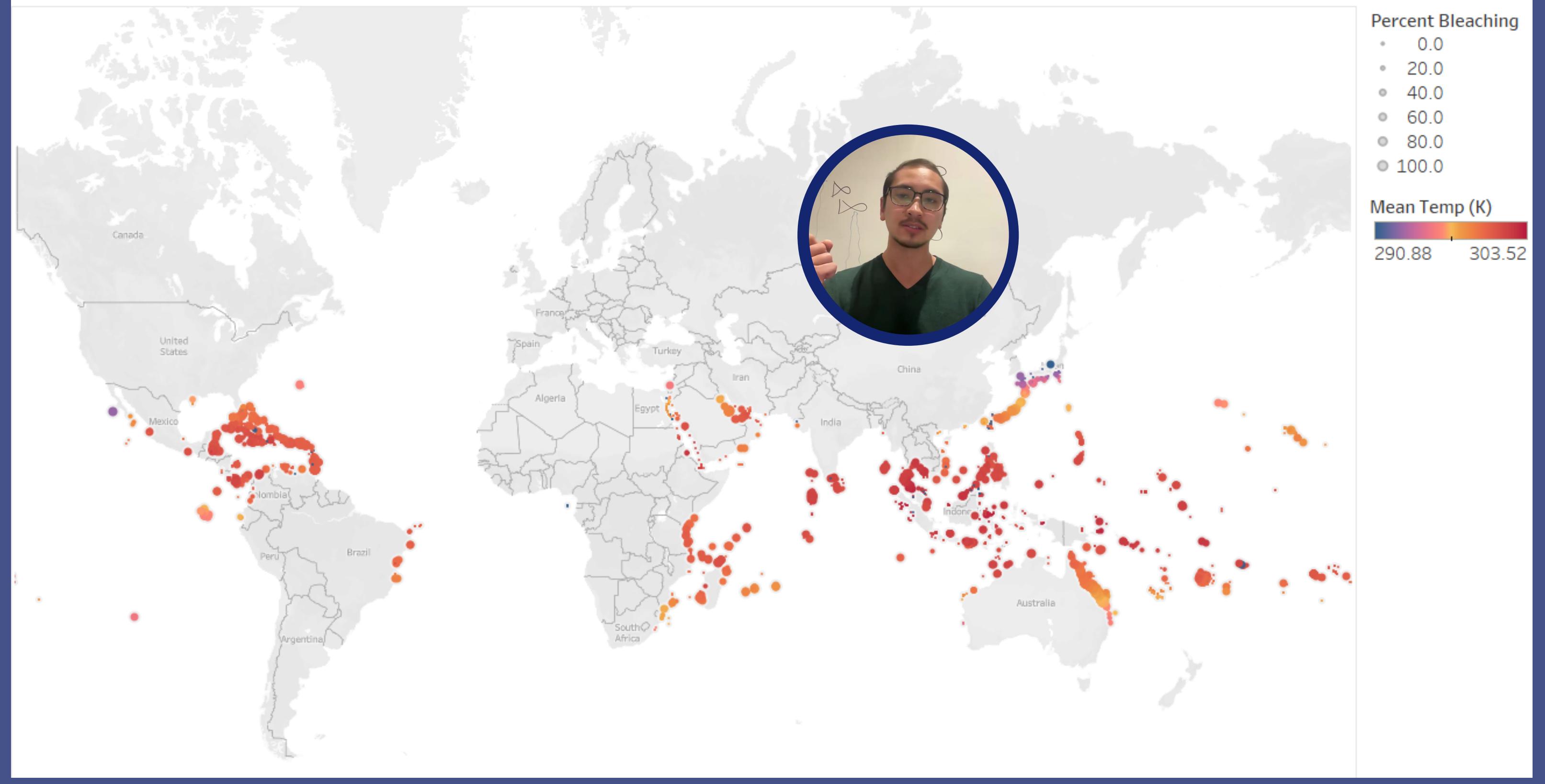
Variable Choice

Distance_to_Shore
Turbidity
Cyclone_Frequency
Depth_m
Temperature_Kelvin
Temperature_Mean
Temperature_Maximum
Windspeed
SSTA
SSTA_Maximum
SSTA_Frequency

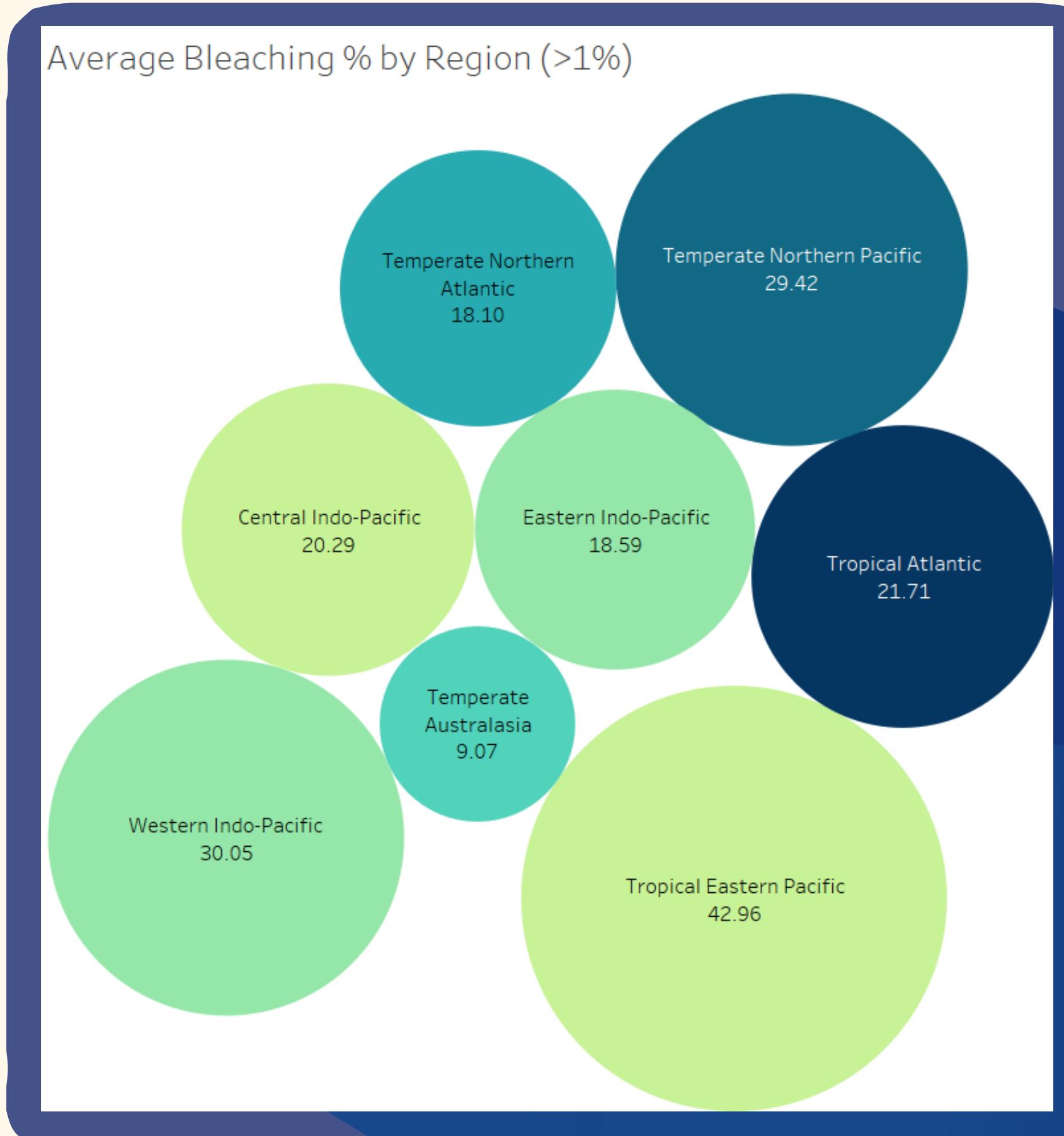


Visualizations

Coral Observations and Bleaching



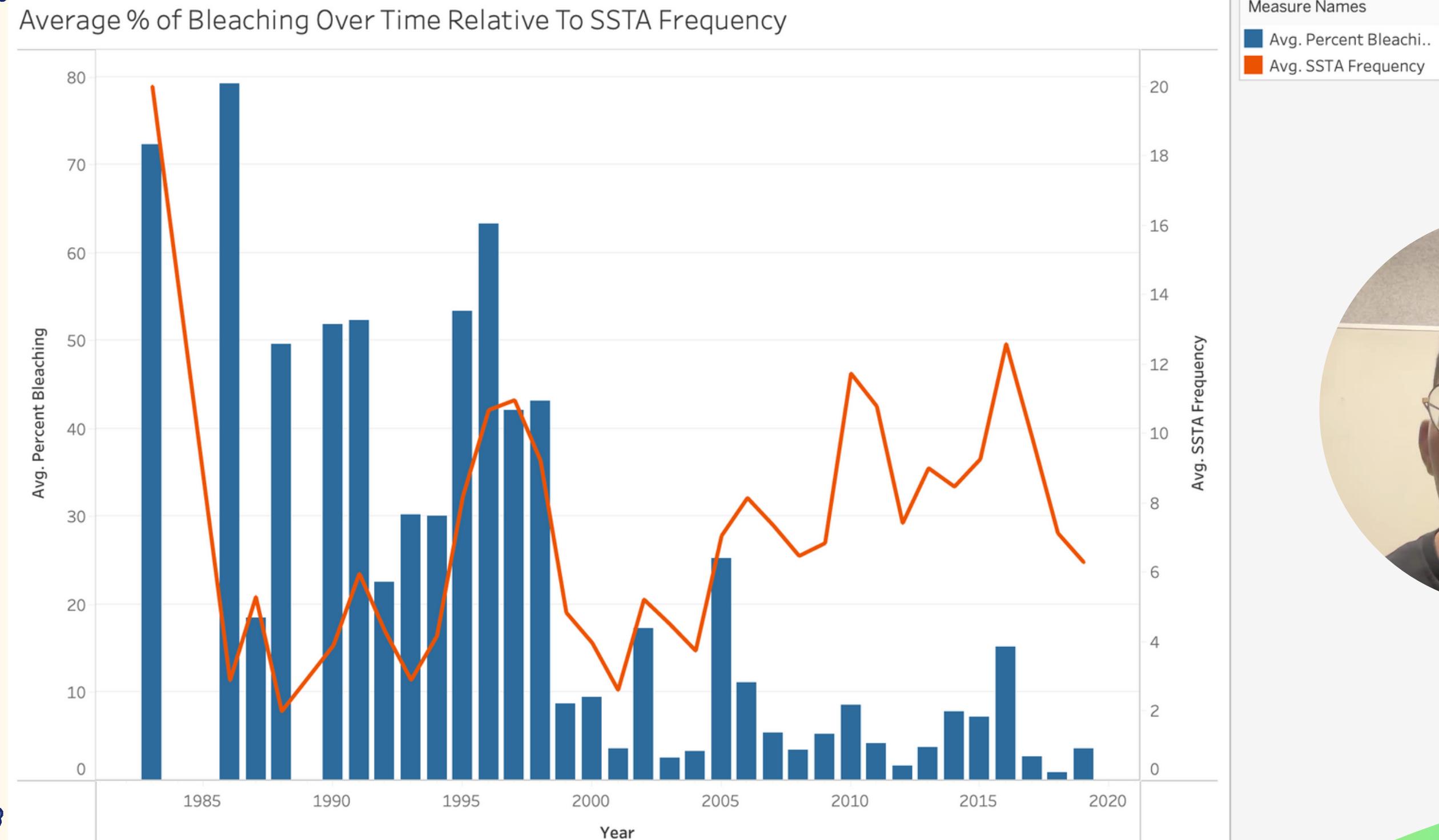
Visualizations



Tropical Eastern Pacific had the highest average bleaching among reefs of concern at 43%

Temperate Australasia had the least at 9%

Visualizations



Linear Regression

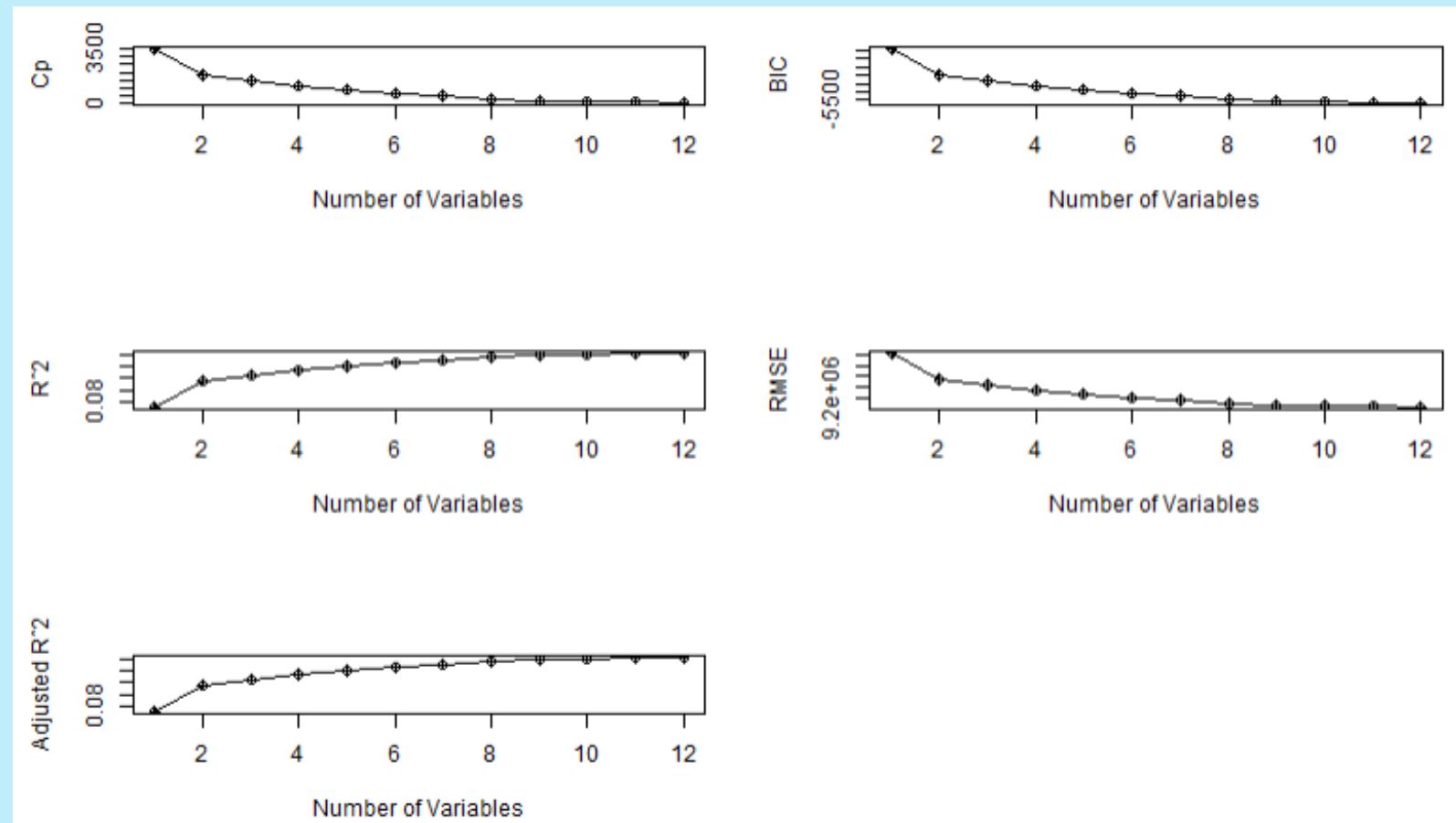


ols_step_best_subset() model

	Estimate	Std. Error	t value	Pr(> t)		
(Intercept)	-32.65171	22.68551	-1.439	0.15007		
Turbidity	-21.45922	1.75407	-12.234	< 2e-16 ***		
Cyclone_Frequency	-0.06687	0.01375	-4.863	1.16e-06 ***		
Depth_m	0.31586	0.02339	13.504	< 2e-16 ***		
Temperature_Kelvin	1.32067	0.06061	21.790	< 2e-16 ***		
Temperature_Mean	-1.21976	0.08013	-15.222	< 2e-16 ***		
Windspeed	0.34476	0.05201	6.629	3.44e-11 ***		
SSTA_Maximum	0.28101	0.08804	3.192	0.00141 **		
SSTA_Frequency	0.61346	0.01588	38.620	< 2e-16 ***		
Ocean_NameAtlantic	10.26972	0.95785	10.722	< 2e-16 ***		
Ocean_NameIndian	7.82562	1.00541	7.784	7.26e-15 ***		
Ocean_NamePacific	1.77202	0.93984	1.885	0.05938 .		
Ocean_NameRed Sea	-5.83146	1.09774	-5.312	1.09e-07 ***		
ExposureSheltered	-1.64024	0.21964	-7.468	8.36e-14 ***		
ExposureSometimes	5.62740	0.35975	15.642	< 2e-16 ***		

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '	1
Residual standard error:	16.63	on 32699 degrees of freedom				
Multiple R-squared:	0.1622,	Adjusted R-squared:	0.1618			
F-statistic:	452.1	on 14 and 32699 DF,	p-value:	< 2.2e-16		

10-variable base model chosen (Cp closest to p+1)



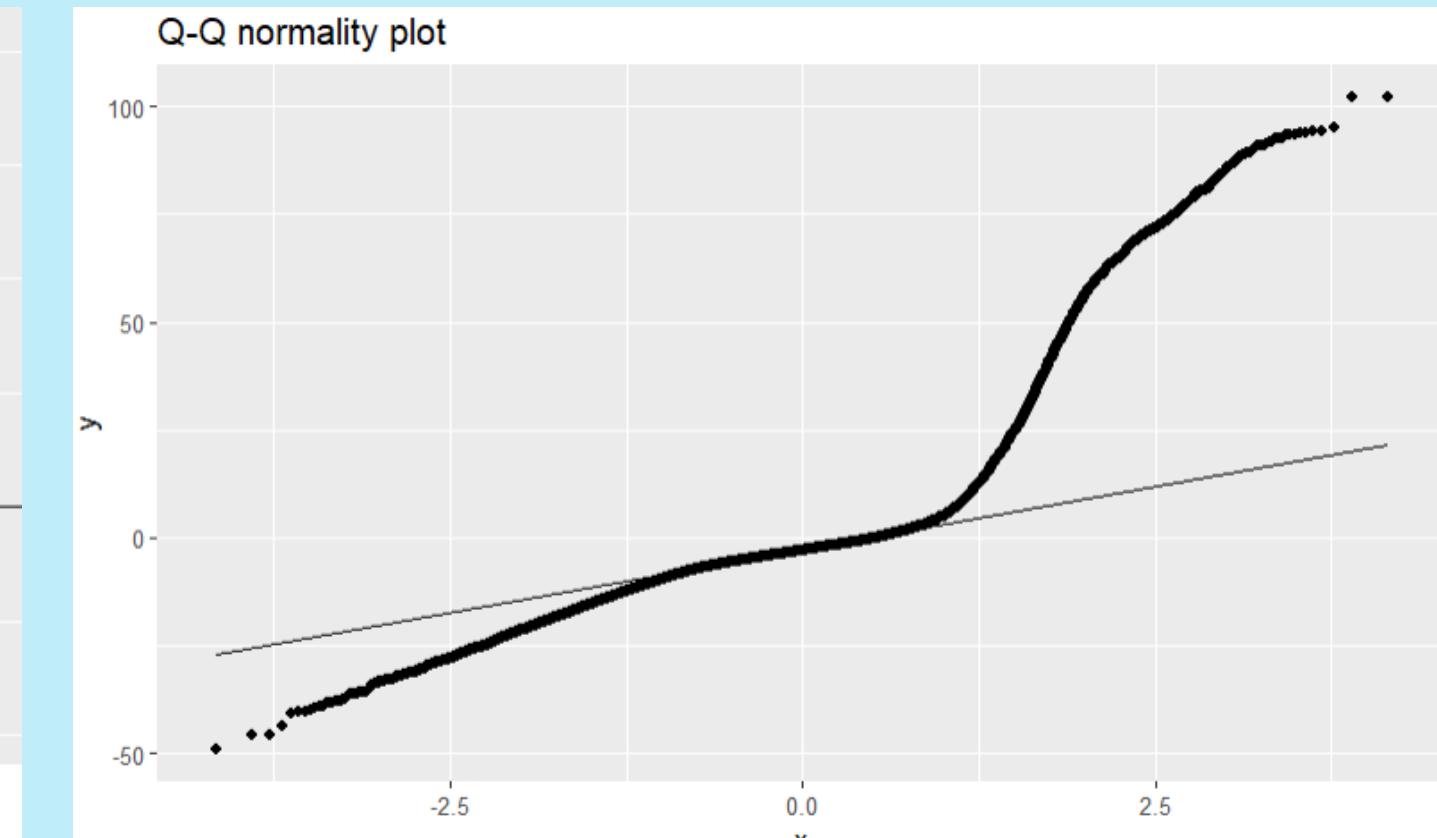
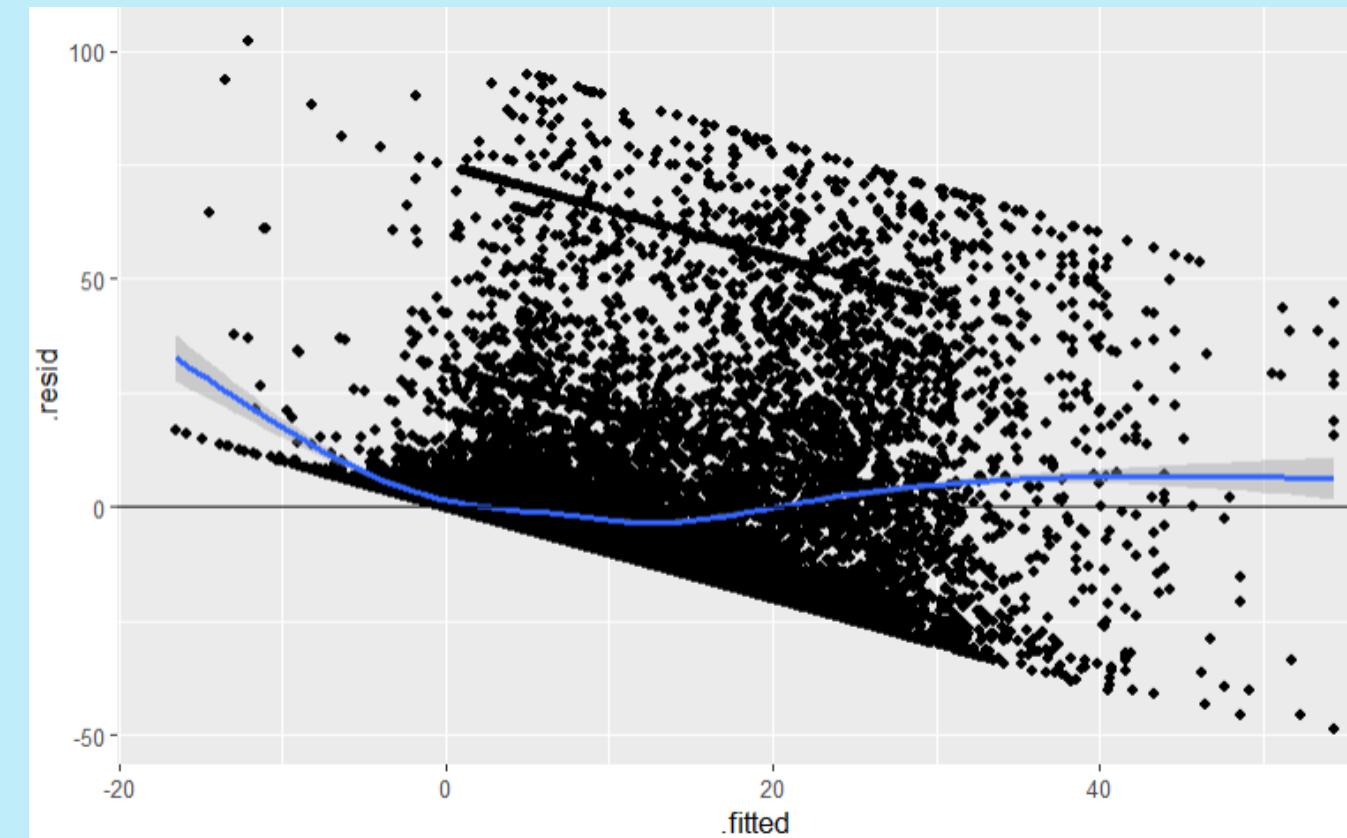
Linear Regression Analysis



R-squared for final interaction model

Residual standard error: 15.91 on 32640 degrees of freedom
Multiple R-squared: 0.235, Adjusted R-squared: 0.2333
F-statistic: 137.4 on 73 and 32640 DF, p-value: < 2.2e-16

Residual Graphs



Linear Discriminant Analysis

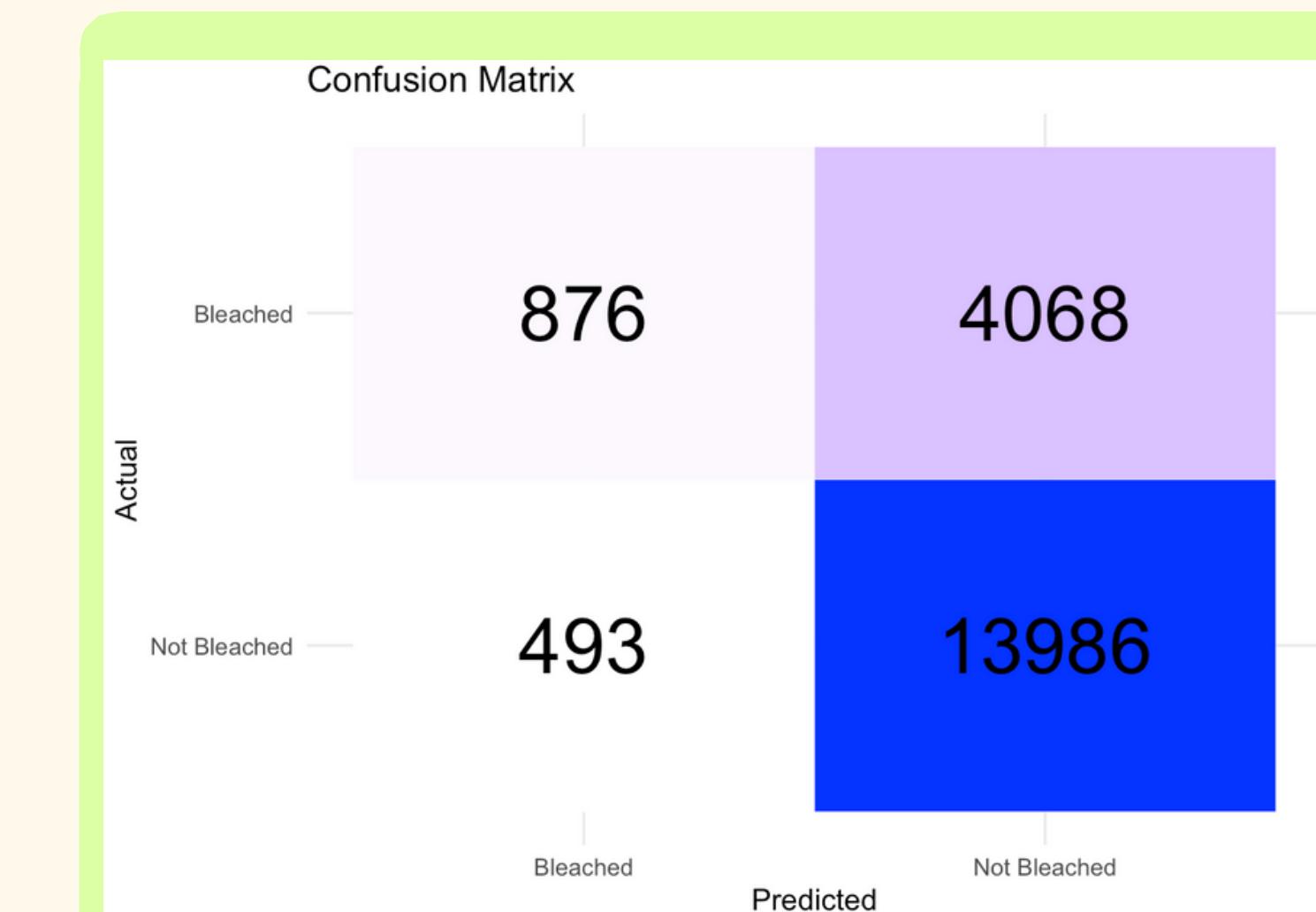
1 Transforming Percent Bleaching

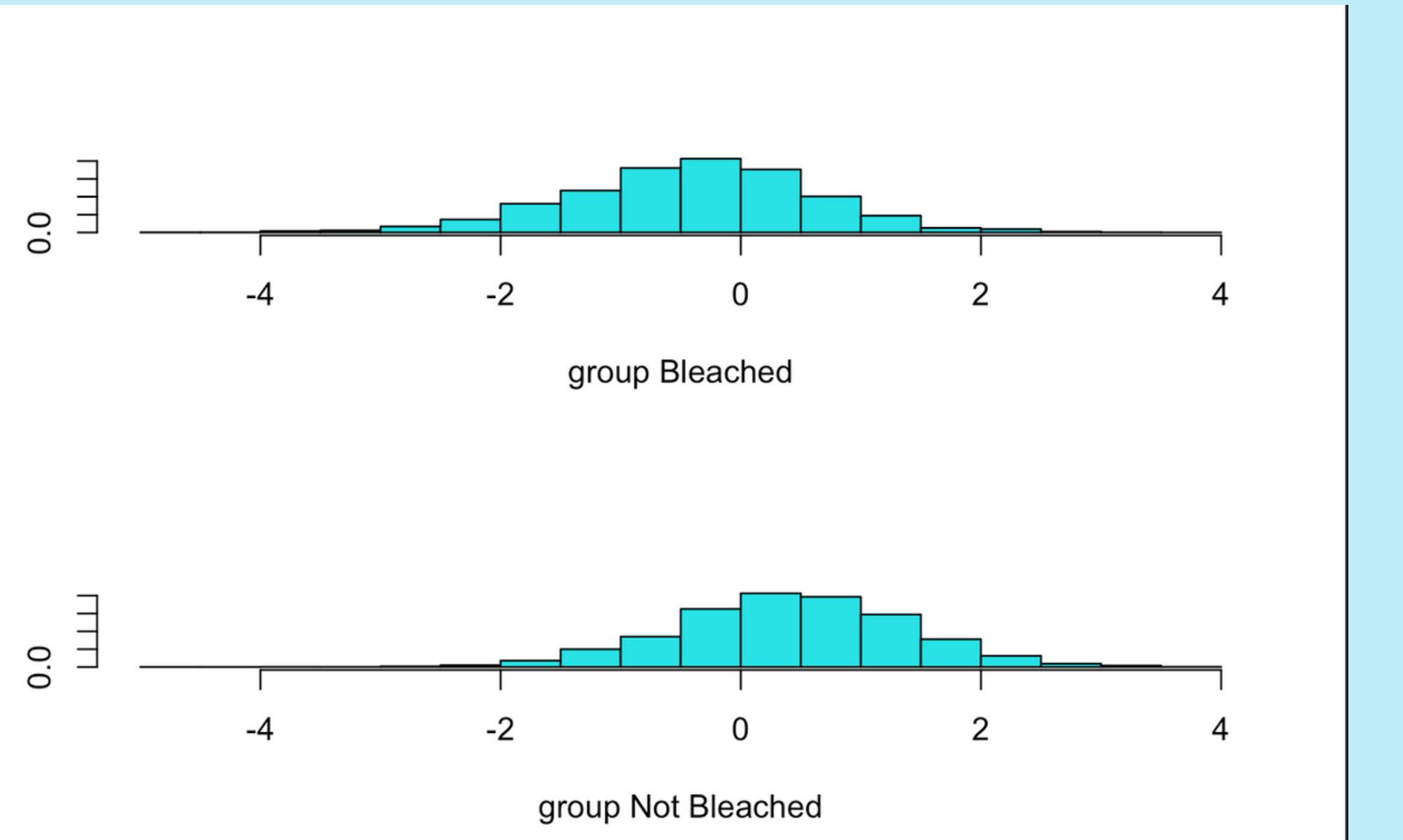
2 Stratified Sampling

3 The misclassification rate is given as : 0.156258

4 Assumptions testing

LDA Top 3 Features	LD1 (Coefficient)
Turbidity	6.97
Mean Temperature	0.99
Maximum Temperature	-0.44

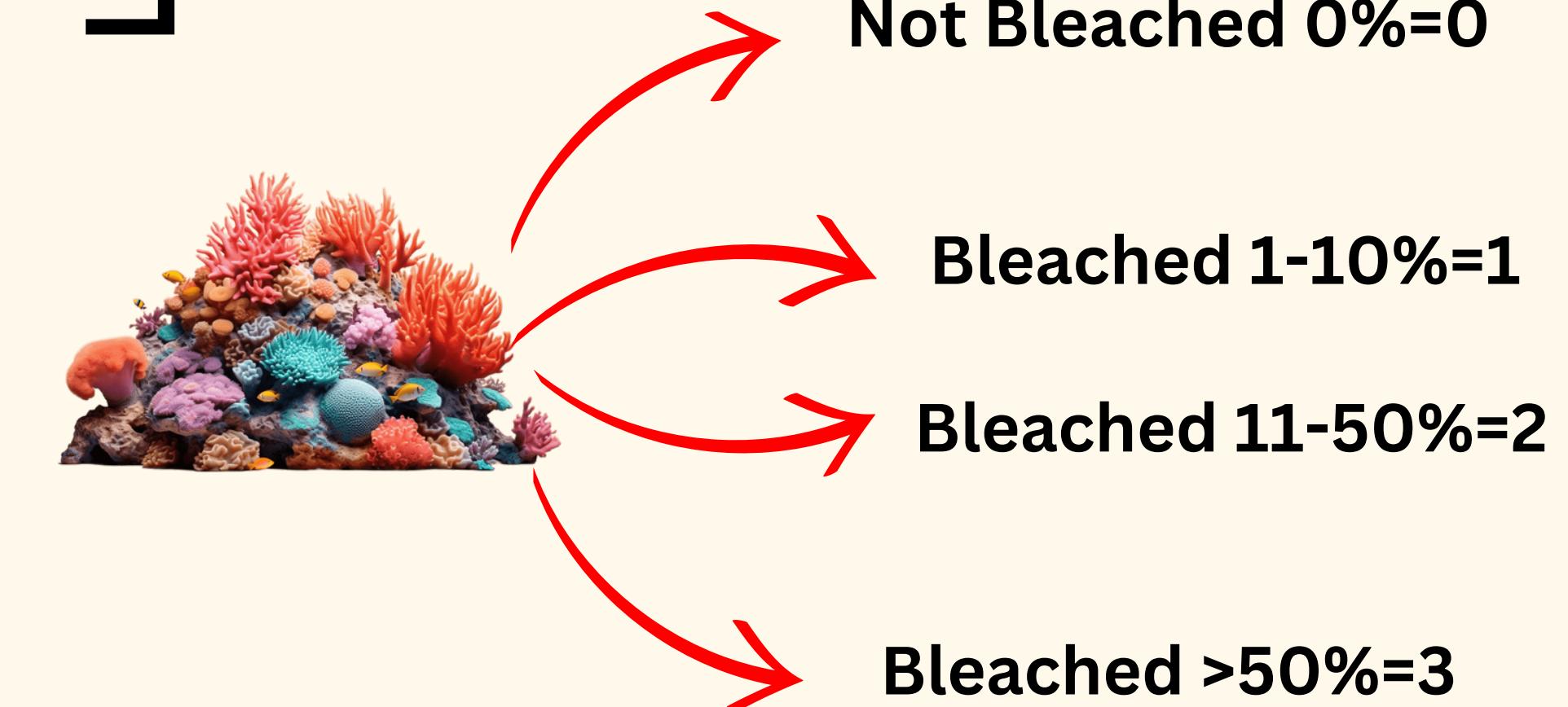




Assumptions	Method	Result
Normality	Energy Test	p-value < 2.2e-16
Homoscedasticity	Box's M Test	p-value < 2.2e-16

Multinomial Logistic Regression

- 1 Recoding the response variable to a new column = Bleached_Multi
- 2 Used Stratified Sampling
 - 0 -> 16509
 - 1 -> 10386
 - 2 -> 24103
 - 3 -> 1677
- 3 Split data 75/25
- 4 ~~53.64%~~



		Actual				
		0	1	2	3	
Predicted	0	3748	2006	698	263	
	1	331	539	248	94	
	2	30	37	61	28	
	3	18	14	19	34	

Logistic Regression

1 Recoding the response variable to a new column = Bleached



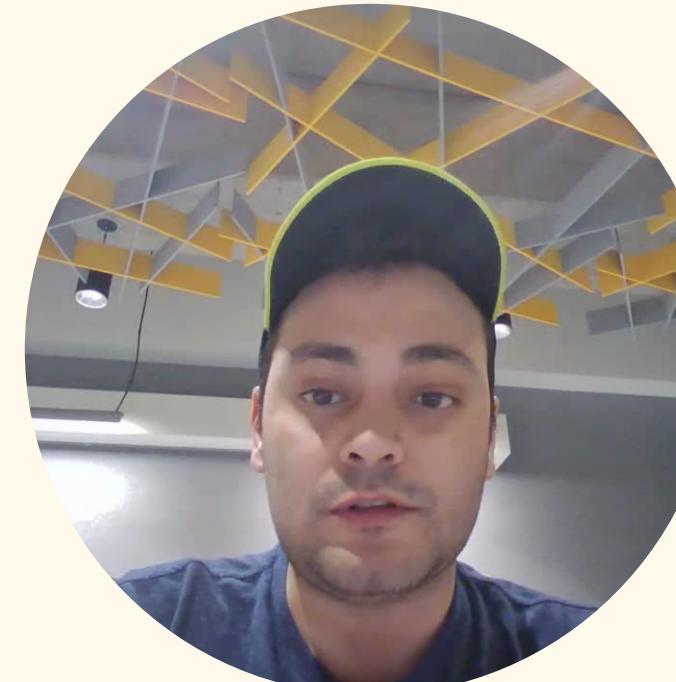
Bleached $>10\% = 1$

2 Used Stratified Sampling

- 0->26895
- 1->5780

3 Split data 75/25

4 82.93%



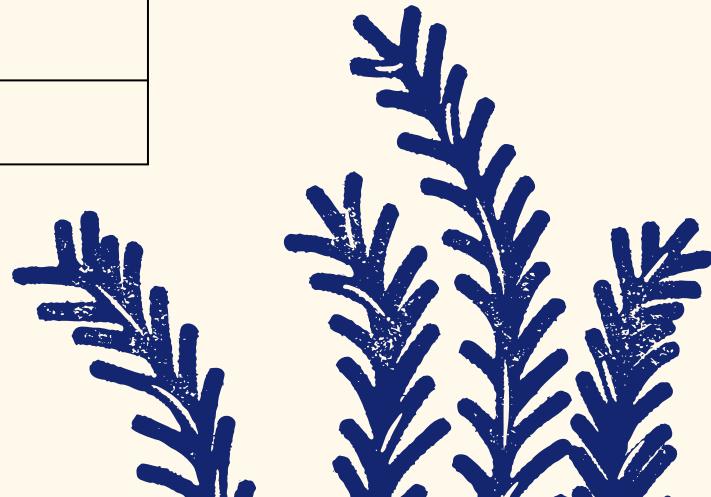
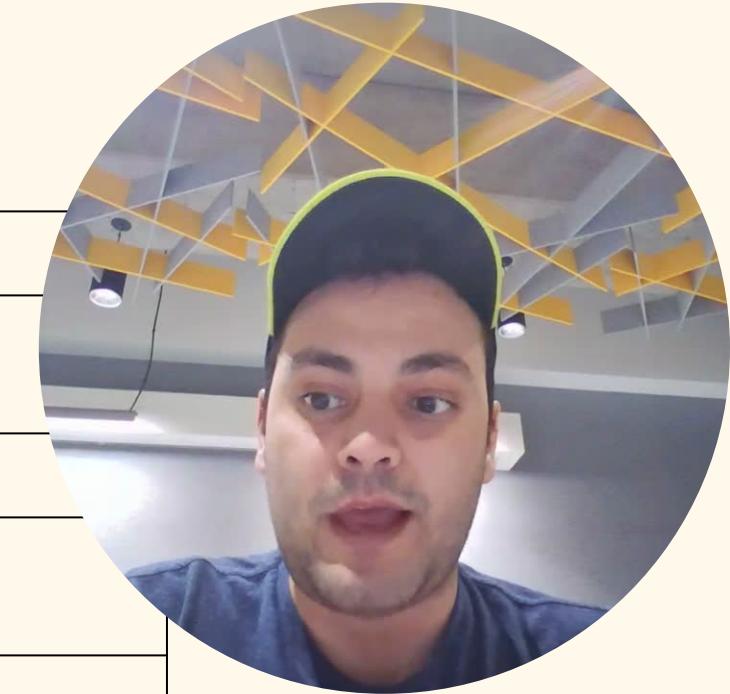
Not Bleached $0-10\% = 0$

		Actual	
		0	1
Predicted	0	6664	1334
	1	60	111

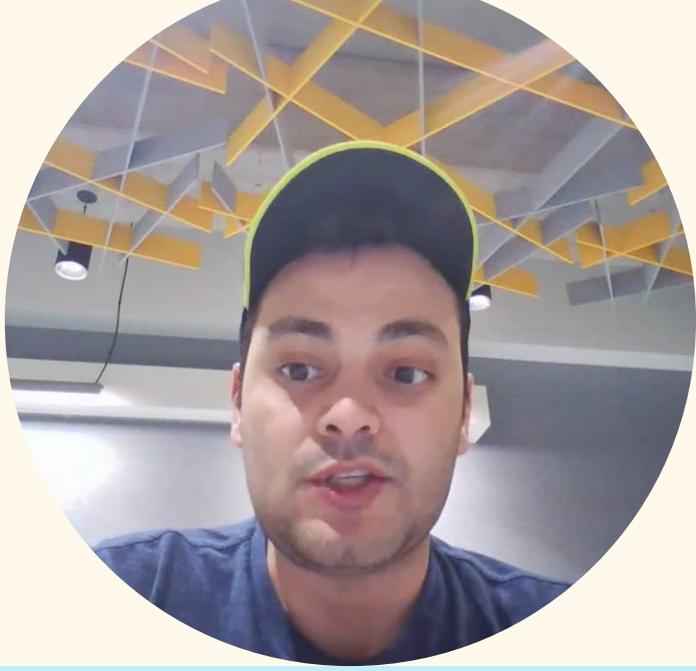


Logistic Regression Coefficients

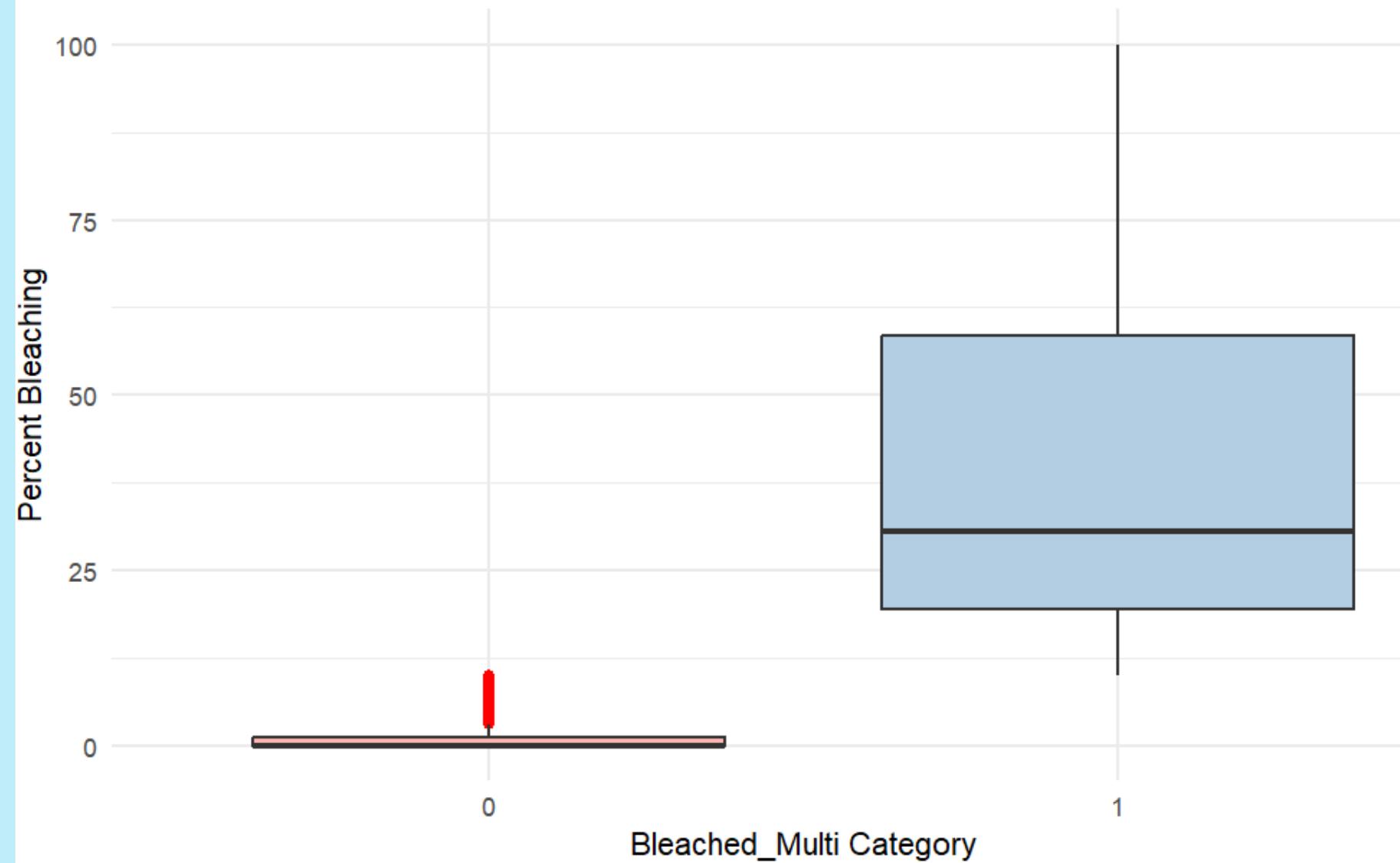
Variable	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-52.79006	5.44339	-9.698	< 0.0000000000000002
Distance_to_Shore	0	0	1.918	0.0551
Turbidity	-6.25631	0.44407	-14.089	< 0.0000000000000002
Cyclone_Frequency	0.01117	0.00229	4.882	1.04868E-06
Depth_m	0.07846	0.00404	19.404	< 0.0000000000000002
Temperature_Kelvin	0.33244	0.01578	21.072	< 0.0000000000000002
Temperature_Mean	-0.37322	0.01949	-19.153	< 0.0000000000000002
Temperature_Maximum	0.20331	0.02593	7.841	4.47E-15
Windspeed	0.08491	0.0097	8.758	< 0.0000000000000002
SSTA	-0.12333	0.02685	-4.594	4.34972E-06
SSTA_Maximum	-0.15079	0.02434	-6.194	5.87241E-10
SSTA_Frequency	0.05853	0.00285	20.502	< 0.0000000000000002



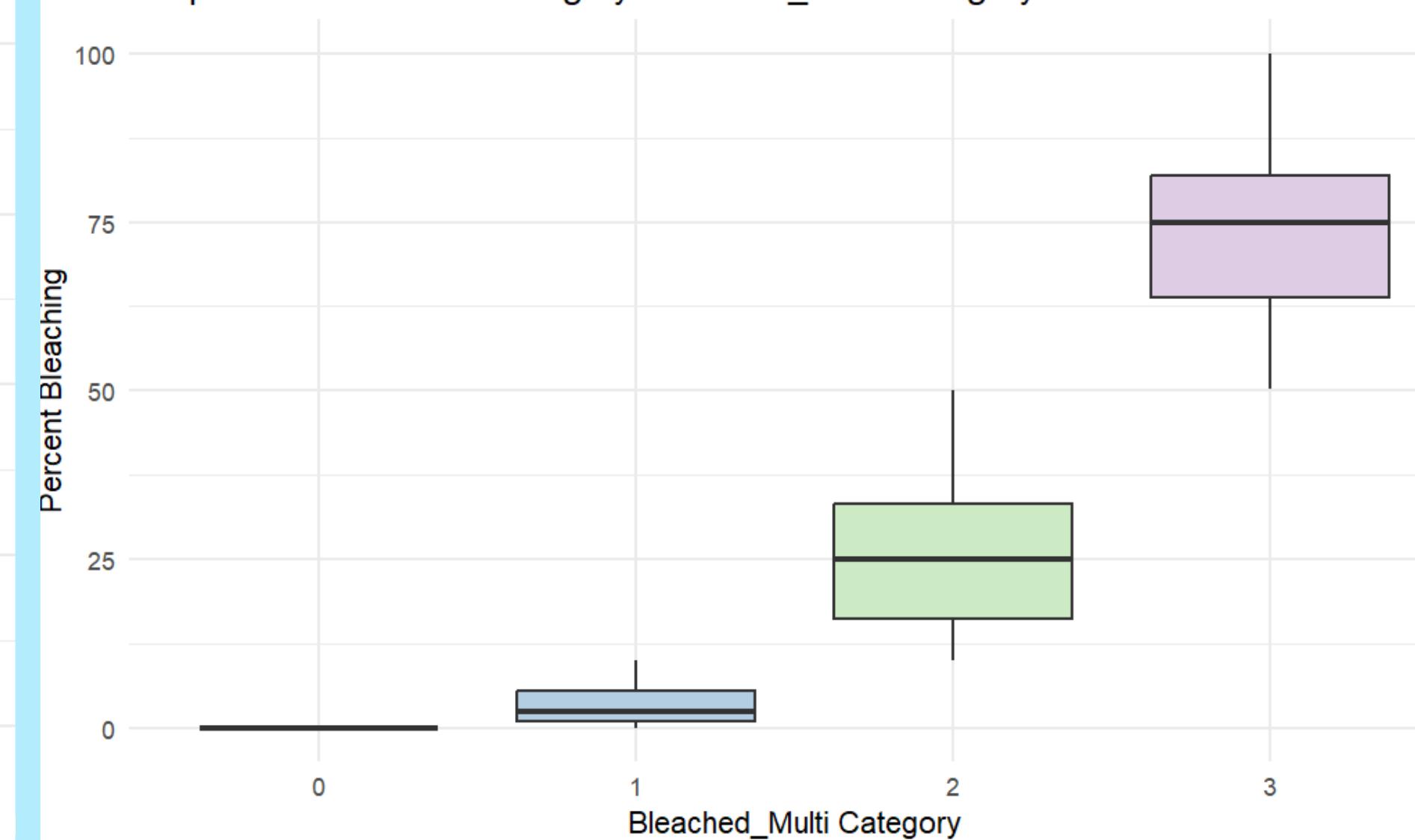
Bleaching Categorical Spread



Boxplot of Percent Bleaching by Bleached_Multi Category



Boxplot of Percent Bleaching by Bleached_Multi Category



Conclusions

1

Model Results:

- Linear Regression $R^2 = 0.233$
- Linear Discriminant Analysis misclassification rate = 15.62%
- Logistic Regression = misclassification rate 17.02%

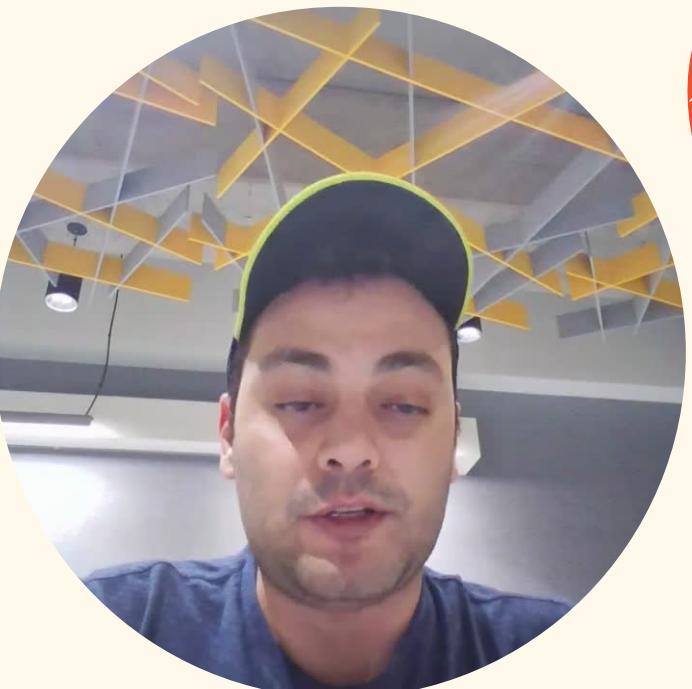


Tropical Eastern Pacific coreal reefs showed the highest bleaching.

2

Recommendations:

- Deploy technologies and monitoring.
- Implement further studies to include human population data and pollution data



Thank You

References:

- van Woesik, R., Burkepile, D. (2022) Bleaching and environmental data for global coral reef sites from 1980-2020. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 2) Version Date 2022-10-14. doi:10.26008/1912/bco-dmo.773466.2

