# Impact of rise in Sea Surface Temperature on Coral Bleaching in Florida Keys.

## (Why)The Importance of Coral Reefs

Any diver can tell you that coral reefs are beautiful. They are like undersea cities, filled with colorful fish, intricate formations and wondrous sea creatures. The importance of coral reefs, however, extends far beyond the pleasure it brings to those who explore it. Coral reefs play an essential role in everything from water filtration and fish reproduction to shore line protection and erosion prevention.

Reefs play an important role in protecting the shoreline from storms and surge water. Barrier reefs, such as Florida's, were named for the way they reduce waves and buffer the shores. Barrier reefs help stabilize mangroves and sea grass beds, which can easily be uprooted by large waves and high currents. Erosion prevention is particularly important in coastal areas such as the Florida Keys, where much of the shore is lined with residential homes and commercial buildings.

## Data Source - Florida Keys BleachWatch Program

The initial onset of mass coral bleaching can vary among different species, geographic locations, types of reef zones and a fluctuation of severity, which makes it very difficult to predict where or when it will occur.

Data for this project is from Florida Keys BleachWatch Program, modeled after Great Barrier Reef's BleachWatch, is a team of trained recreational, commercial and scientific divers who help monitor and report on conditions at the reefs.

## Types of corals

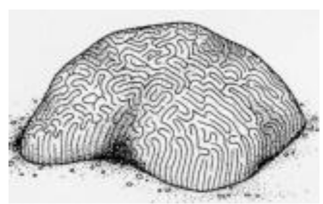
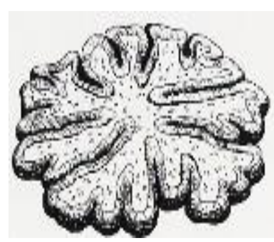
| | | |
|---|---|---|
| **Branching & Pillar** |  |  |
| **Flowering & Cup** |  |  |
| **Encrusting, Mound and Boulder** |  |  |
| **Brain Corals** |  |  |
| **Fleshy Corals** |  |  |
| **Plate & Leaf** |  |  |

*Table 1: Types of coral identification (image courtesy internet/google).*

## Metadata

| | |
|---|---|
| **Contact Method** | Email / In person |
| **Name** | Observer Name |
| **Date** | Date of survey |
| **Time** | Time of survey |
| **Address** | Contact Details |
| **Phone** | |
| **E-Mail** | |
| **Observer** | Resident: Local Florida Keys resident |
| | Visitor: Tourist |
| | Dive Industry: Scuba Diving and Snorkel Operators or Commercial Divers |
| | Education: School and Camp Groups |
| | Research: Other Agencies and Laboratories Conducting Current Research |
| **Vessel** | Survey boat name |
| **GIS Latitude** | GIS Latitude Location of survey |
| **GIS Longitude** | GIS Longitude Location of survey |
| **Depth (ft.)** | Depth of survey site |
| **Location** | Common reef name of survey site |
| **Region/Buoy#** | Area of reef site |
| **Reef Zone** | Type of Reef Zone: |
| | Hard bottom: Close to shore and mostly soft corals. |
| | Patch Reef: Patch of reef surrounded by sand (10-20’) |
| | Mid-Channel Patch Reef: Patch Reef found in a Channel (ex. Hawk Channel) |
| | Deep Reef: Deep water reef (60’+) |
| | Bank Reef: Spur and Groove formations |
| **Wind Speed. (knots)** | Estimate of wind speed at survey site. 0- |
| | Calm: No wind, 0-5 knots, 5-10 knots, 10-15 knots, 15-20 knots, 20+ knots |
| **Air Temp (°F)** | Air Temperature at survey site measured in Fahrenheit |
| **SST (°F)** | Sea Surface Temperature at survey site measured in Fahrenheit |
| **Bottom Temp (°F)** | Sea Bottom Temperature at survey site measured in Fahrenheit |
| **Cloud Cover** | Clear: no clouds |
| | Partly Cloudy: Approximately 50% cloud cover |
| | Mostly Cloudy: Approximately 75% cloud cover |
| | Cloudy: Approximately more than 75% cloud cover |
| **Bleaching?** | Was there coral bleaching occurring at the survey site, YES or NO |
| **Severity** | Bleached on Upper Surface: Corals bleaching only on upper surfaces, areas more visible to sea surface.<br>Paling: Coral losing its usual color hue. Lighter colors than usual.<br>Partial Bleaching: Corals are white in some areas, but still have color in others.<br>Bleached White: Corals are completely white with no color.<br>Dead with Algae Growth: Corals are white and are dying from coral bleaching. |
| **Branching/Pillar** | Breakdown of severity of bleaching (see **Severity** for definitions) on types of corals.<br><br>Types of corals are defined using Paul Humann's "Reef Coral Identification". (See Table 1) |
| **Brain** | |
| **Encrusting/Mound/Boulder** | |
| **Flowering/Cup** | |

| | |
|---|---|
| **Leaf/Plate/Sheet** | |
| **Fleshy** | |
| **% bleached** | Percentage of live coral that is bleaching or bleached. |
| **Max Depth** | Deepest depth measured in feet at survey site that shows signs of coral bleaching. |
| **Min Depth** | Shallowest depth measured in feet at survey site that shows signs of coral bleaching. |
| **Baseline Indicators** | Reef organisms that are not stony coral but do bleach. |
| | Fire Coral: Encrusting hydrocoral that usually bleaches early in season. |
| | Palythoa: Encrusting Cnidarian that usually bleaches early in season. |
| | Gorgonian: Octocoral that usually bleaches later in season. |
| **Notes** | Other observations at survey site such as coral disease or species specific bleaching. |

## Data Exploration - (What data do we have?)

The columns that will be useful for our analysis/modelling is as below.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Date** | **Time** | **GIS Latitude** | **GIS Longitude** | **Depth (ft.)** | **Wind Speed (knots)** | **Air Temp (°F)** | **SST (°F)** | **Bottom Temp (°F)** | **Cloud Cover** |
| **Bleaching?** | **Severity** | **Branching/ Pillar** | **Brain** | **Encrusting/ Mound/ Boulder** | **Flowering/ Cup** | **Leaf/Plate/ Sheet** | **Fleshy** | **% bleached** | |

1. Establish null hypothesis as "Increase in Sea Surface Temperature (SST) is directly proportional to increase in instances of Coral Bleaching"
    a. Is the null hypotheses true for all types of corals? (Are some types more resilient to temperature changes)
2. Extract numeric month and year (from Date column) and store as new columns.
3. By listing out the unique data for the above columns, we observe 'no data' and it needs to be converted as NAN (using numpy.nan).
4. The corals observation columns by types have non-numeric/text values. So we need to convert them into numeric categories. Also the values have typo errors that needs to be rectified.

| | |
|---|---|
| None | 0 |
| Upper Surface | 1 |
| Paling | 2 |
| Partial Bleaching | 3 |
| Bleaching | 4 |

5. Plot

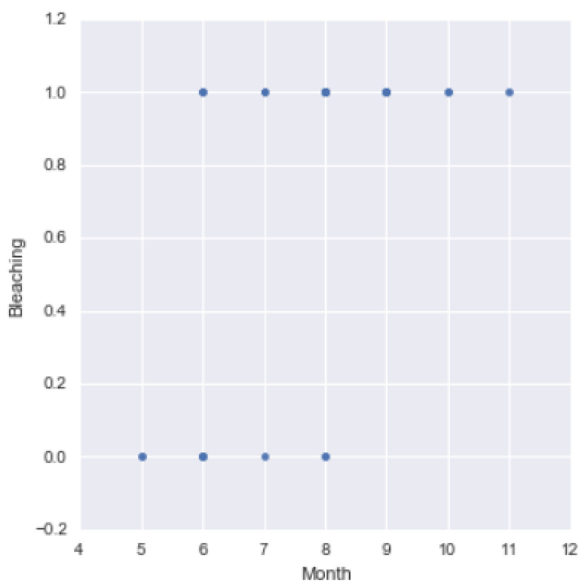
## Data Modelling & Prediction

### 1. Bias-Variance

Create a scatterplot visualize the relationship between Bleaching and Month

```
import seaborn as sns
sns.lmplot(x='Month', y='Bleaching', data=df_model, ci=None, fit_reg=False)
```

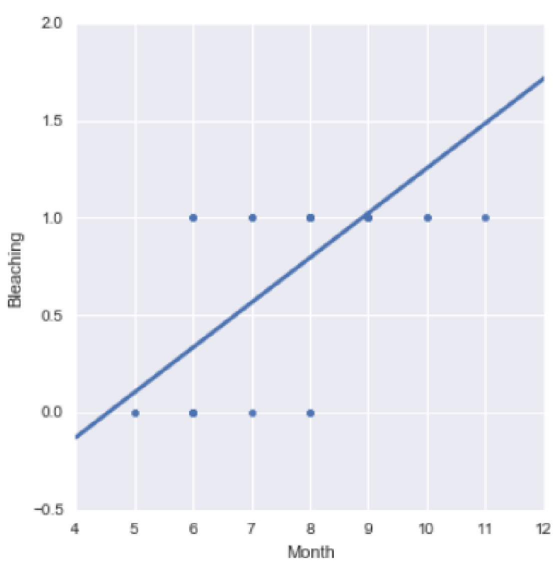
<seaborn.axisgrid.FacetGrid at 0x13460b00>




Predict Bleaching w.r.t Month

```
import seaborn as sns
sns.lmplot(x='Month', y='Bleaching', data=df_model, ci=None)
```

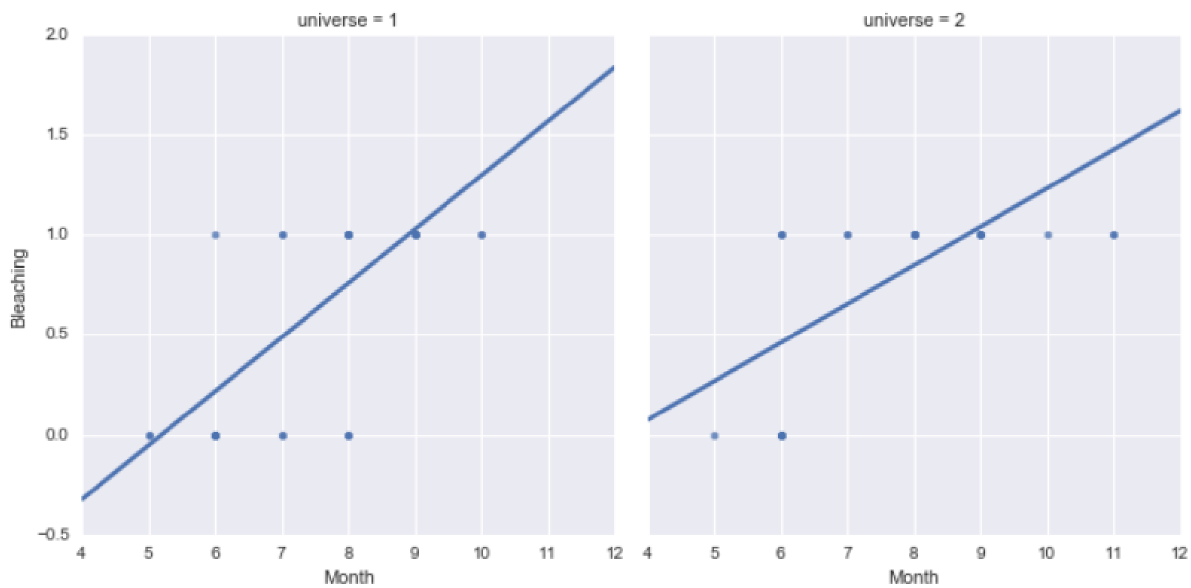
<seaborn.axisgrid.FacetGrid at 0x10b69a58>



## High bias / Low Variance (Predict from given sample data)

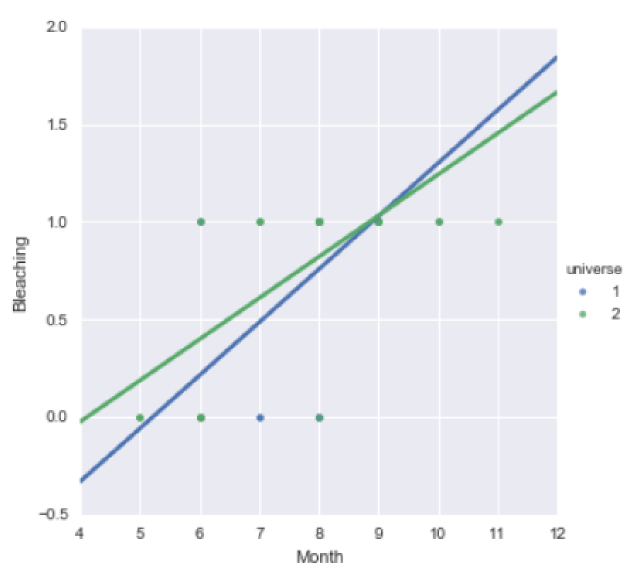
```
import seaborn as sns
df_model['universe'] = np.random.randint(1, 3, len(df_model))
sns.lmplot(x='Month', y='Bleaching', data=df_model, ci=None, col='universe')
```

```
<seaborn.axisgrid.FacetGrid at 0x10bf5dd8>
```




```
import seaborn as sns
df_model['universe'] = np.random.randint(1, 3, len(df_model))
sns.lmplot(x='Month', y='Bleaching', data=df_model, ci=None, hue='universe')
```

```
<seaborn.axisgrid.FacetGrid at 0x113d66a0>
```



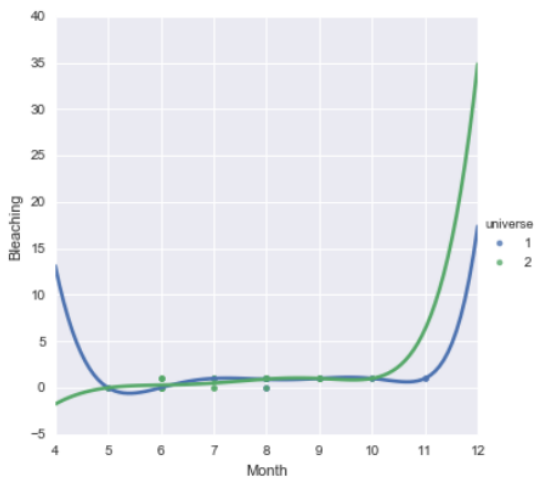

- It's high bias because it doesn't fit the data well
- It's low variance because it doesn't change much depending on which observations happen to be available in that universe

## Low bias / High Variance (Predict from given sample data)

```
import seaborn as sns
df_model['universe'] = np.random.randint(1, 3, len(df_model))
sns.lmplot(x='Month', y='Bleaching', data=df_model, ci=None, hue='universe', order=8)
```

```
C:\Anaconda2\lib\site-packages\numpy\lib\polynomial.py:594: RankWarning: Polyfit may be poorly conditioned
  warnings.warn(msg, RankWarning)
C:\Anaconda2\lib\site-packages\numpy\lib\polynomial.py:594: RankWarning: Polyfit may be poorly conditioned
  warnings.warn(msg, RankWarning)
```

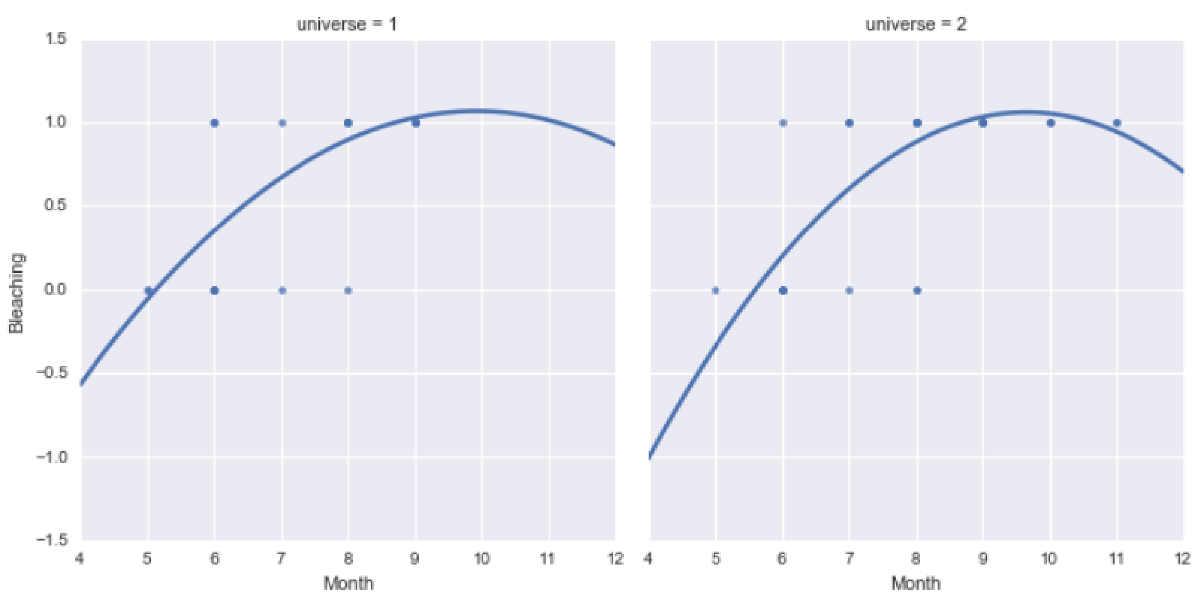
```
<seaborn.axisgrid.FacetGrid at 0x13021940>
```




## Bias – Variance Trade off

```
import seaborn as sns
df_model['universe'] = np.random.randint(1, 3, len(df_model))
sns.lmplot(x='Month', y='Bleaching', data=df_model, ci=None, col='universe', order=2)
```

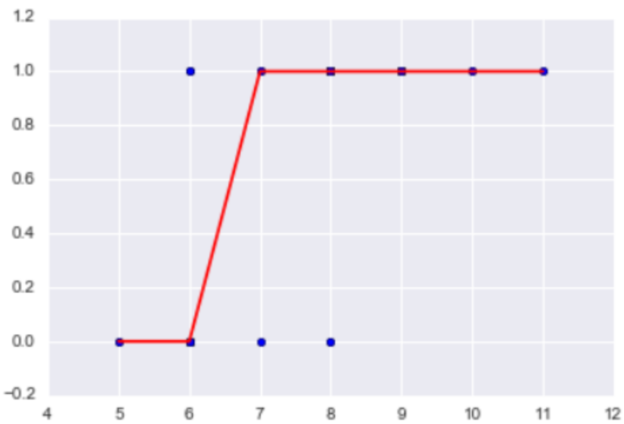
```
<seaborn.axisgrid.FacetGrid at 0x12100978>
```



## 2. Logistic Regression

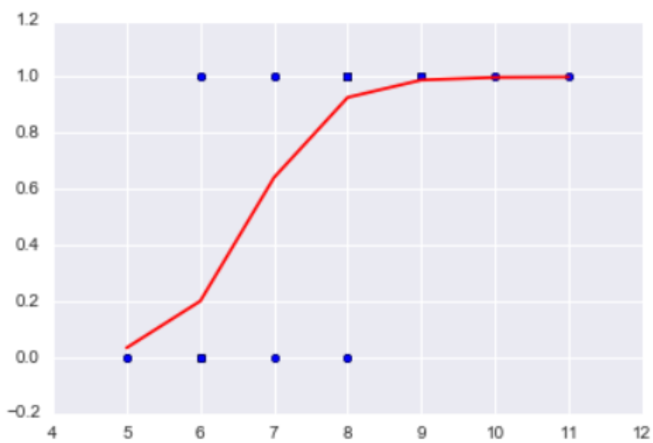
Predicting a Categorical Response (Will Bleaching happen for the given Month)

```
# fit a linear regression model and store the class predictions
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression(C=1e9)
feature_cols = ['Month']
X = df_model[feature_cols]
y = df_model.Bleaching
logreg.fit(X, y)
pred_class = logreg.predict(X)

plt.scatter(df_model.Month, df_model.Bleaching)
plt.plot(df_model.Month, pred_class, color='red')
```



```python
# store the predicted probabilites of class 1
pred_prob = logreg.predict_proba(X)[:, 1]

# plot the predicted probabilities
plt.scatter(df_model.Month, df_model.Bleaching)
plt.plot(df_model.Month, pred_prob, color='red')
```

```
[<matplotlib.lines.Line2D at 0x13e84cc0>]
```




```python
# compute predicted log-odds for Month=7 using the equation
logodds = logreg.intercept_ + logreg.coef_ * 7
print logodds

# convert log-odds to odds
odds = np.exp(logodds)
print odds

# convert odds to probability
prob = odds/(1 + odds)
print prob

# compute predicted probability for Month=7 using the predict_proba method
print logreg.predict_proba(7)[:, 1]

# examine the coefficient for Month
print zip(feature_cols, logreg.coef_[0])

# convert log-odds to probability
logodds = logreg.intercept_
odds = np.exp(logodds)
prob = odds/(1 + odds)
prob
```

```
[[ 0.58088754]]
[[ 1.78762432]]
[[ 0.6412716]]
[ 0.6412716]
[('Month', 1.9604622321271288)]

array([  1.96042349e-06])
```

Q? Predicting a Categorical Response (Will Bleaching happen for the given Month)
Ans: A 1 unit increase in 'Month' is associated with a 1.96 unit increase in the log-odds of 'Bleaching'.

## 3. Linear Regression

In [72]:
```
# create X and y
feature_cols = ['Branching', 'Brain', 'Encrusting', 'Flowering', 'Leaf', 'Fleshy']
X = df_model[feature_cols]
y = df_model.SST

# instantiate and fit
linreg = LinearRegression()
linreg.fit(X, y)

# print the coefficients
print linreg.intercept_
print linreg.coef_

# pair the feature names with the coefficients
zip(feature_cols, linreg.coef_)
```

```
83.9680644985
[ 0.10060875  0.77661966  0.25464274 -0.29924246 -1.0940295   0.40922964]
```

Out[72]:
```
[('Branching', 0.10060874529106714),
 ('Brain', 0.77661966113415459),
 ('Encrusting', 0.2546427361568302),
 ('Flowering', -0.29924246064211507),
 ('Leaf', -1.0940295002579306),
 ('Fleshy', 0.40922963658214928)]
```

### 3.1. Cross Validation - Accuracy

```
from sklearn.cross_validation import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics

def train_test_rmse(X, y):
    X_train, X_test, y_train, y_test = train_test_split(X, y,random_state=5)
    linreg = LinearRegression()
    linreg.fit(X_train, y_train)
    y_pred = linreg.predict(X_test)
    return np.sqrt(metrics.mean_squared_error(y_test, y_pred))

feature_cols = ['Branching', 'Brain', 'Encrusting', 'Flowering', 'Leaf', 'Fleshy']
X = df_model[feature_cols]
y = df_model.SST
train_test_rmse(X, y)
```

```
2.6304939496808633
```

```
X_train, X_test, y_train, y_test=train_test_split(X,y,random_state=5)

knn =  KNeighborsClassifier(n_neighbors=5)

knn.fit(X_train, y_train)
y_pred=knn.predict(X_test)

print metrics.accuracy_score(y_test, y_pred)
```

```
0.590909090909
```

3.2. **Cross Validation - Feature Selection**

10-fold cross validation **SST vs Corals Types** ['Branching', 'Brain', 'Encrusting', 'Flowering', 'Leaf', 'Fleshy'] NULL RMSE = 2.7990553306073918

| Coral Types | RMSE Mean |
|---|---|
| All Corals | 2.76625644844 |
| Excluding **Branching** | 2.69420551636 |
| Excluding **Brain** | 2.87748239869 |
| Excluding **Encrusting** | 2.58241087282 |
| Excluding **Flowering** | 2.7476231747 |
| Excluding **Leaf** | 2.94037968635 |
| Excluding **Fleshy** | 2.75972146961 |

**Null hypothesis**: There is no relationship between increase in SST and Bleaching
**Alternative hypothesis**: There is a relationship between increase in SST and Bleaching

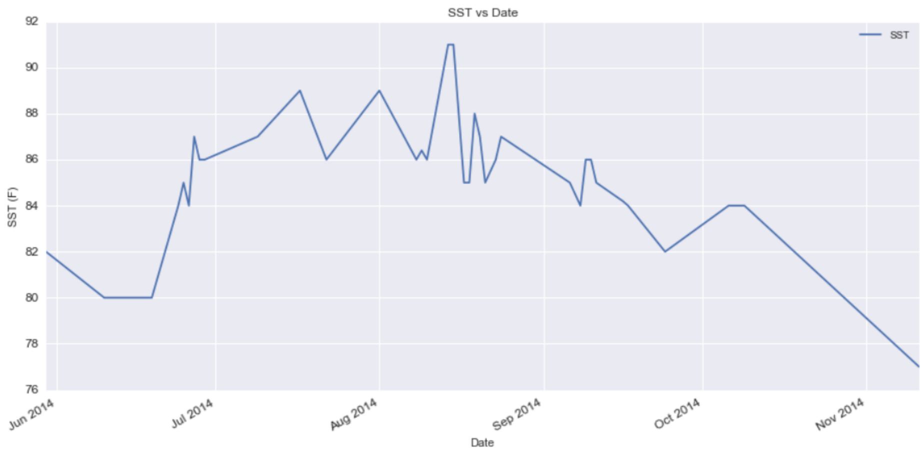
**P-values**

```
import statsmodels.formula.api as smf

# create a fitted model in one line
lm = smf.ols(formula='SST ~ Branching + Brain + Encrusting + Flowering + Leaf + Fleshy', data=df_model).fit()

# print the p-values for the model coefficients
print lm.pvalues
```

```
Intercept     1.234841e-106
Branching      6.164323e-01
Brain          1.424155e-03
Encrusting     2.822229e-01
Flowering      5.076895e-01
Leaf           1.639666e-04
Fleshy         3.668179e-01
dtype: float64
```

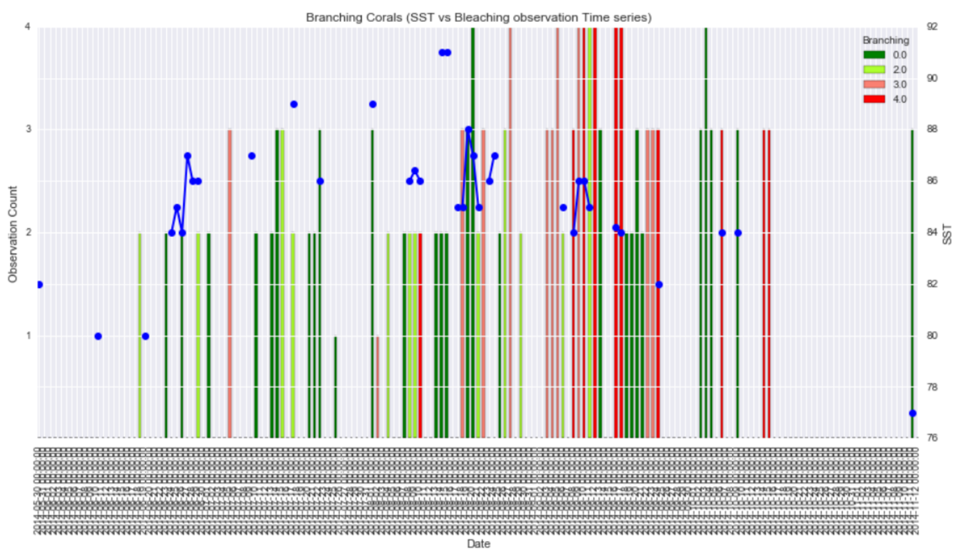
| | |
|---|---|
| **Intercept** | **1.234841e-106** |
| **Branching** | **6.164323e-01** |
| **Brain** | **1.424155e-03** |
| **Encrusting** | **2.822229e-01** |
| **Flowering** | **5.076895e-01** |
| **Leaf** | **1.639666e-04** |
| **Fleshy** | **3.668179e-01** |

Let's look at Sea Surface Temp recorded for the sampling period. (Missing data are interpolated)

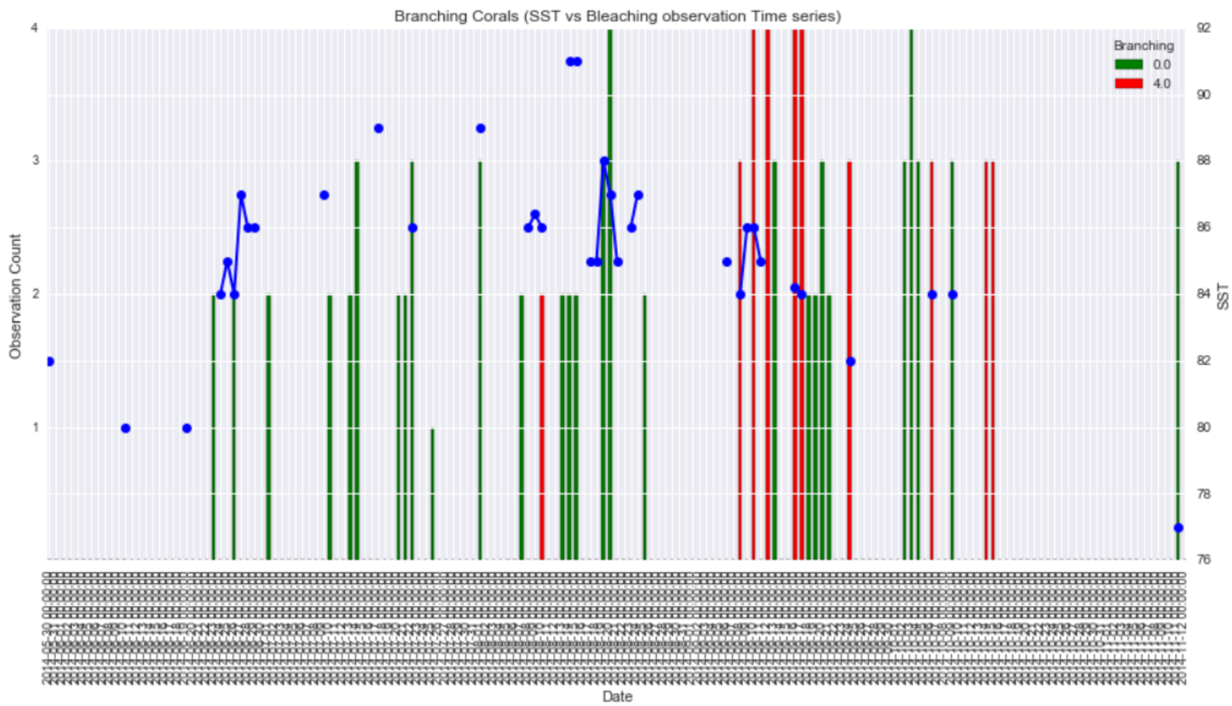


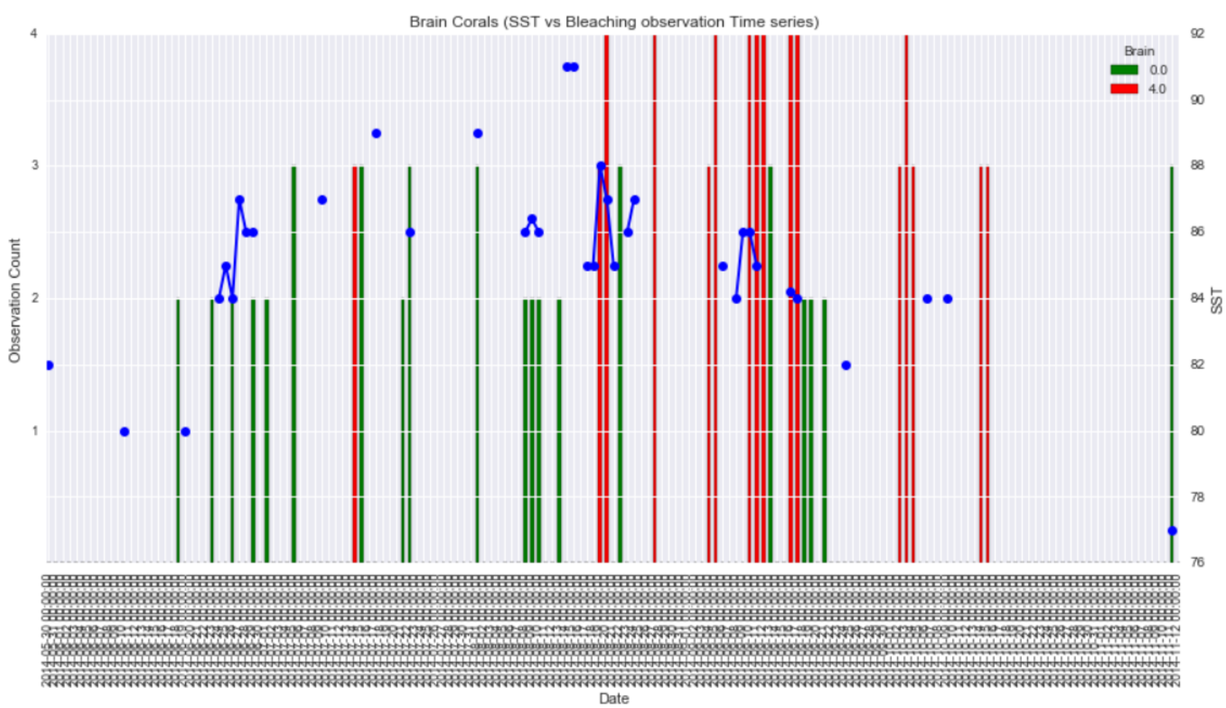

Plot the daily coral observation for each species for the sampling period

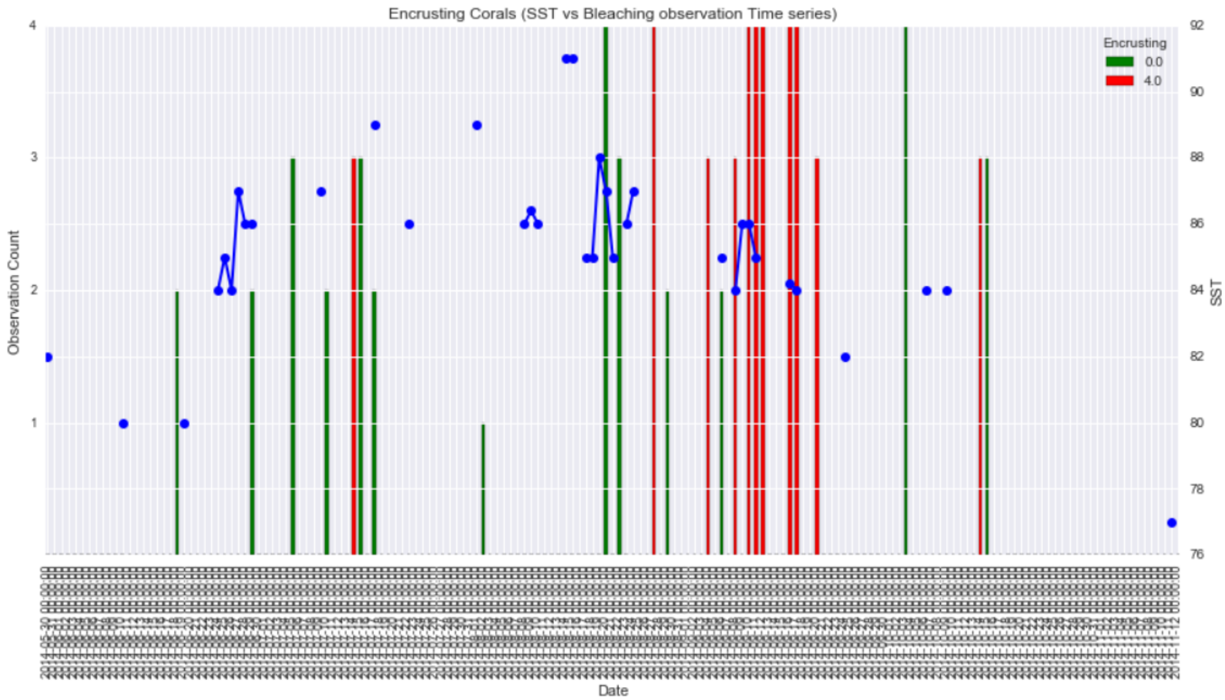
1. For species '**Branching**' all severity types



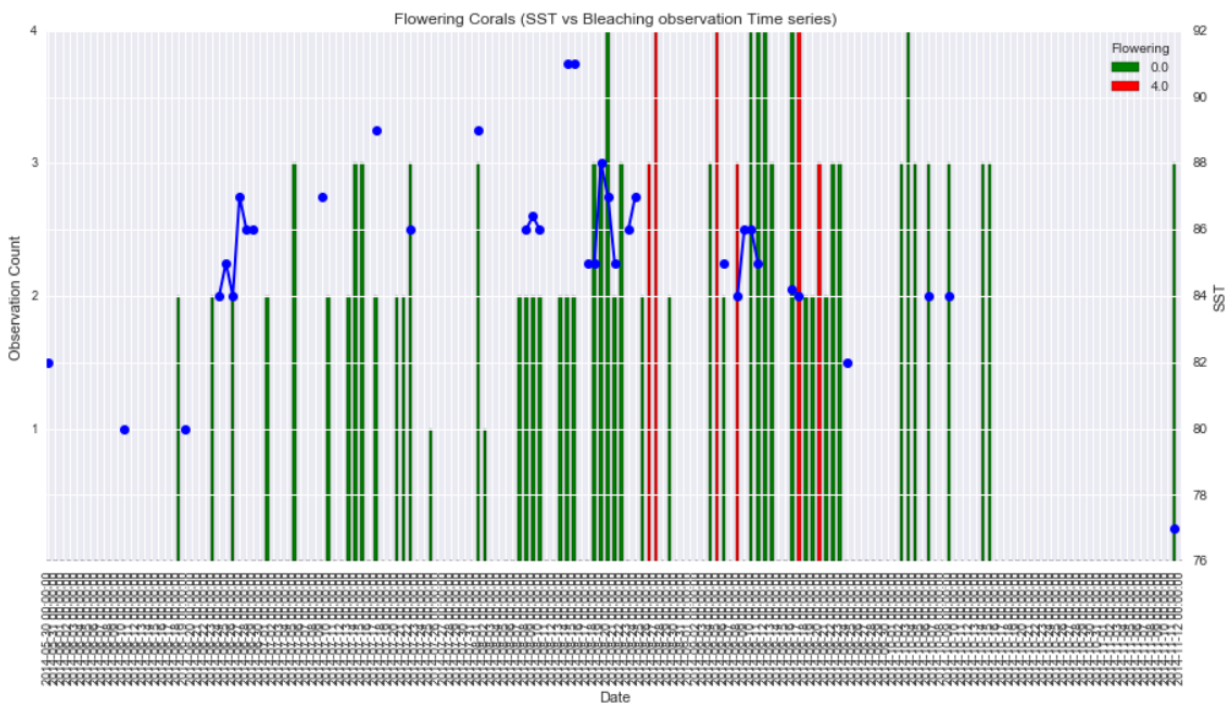
Filtered only No Bleaching (Severity 0) & Bleaching (Severity 4)



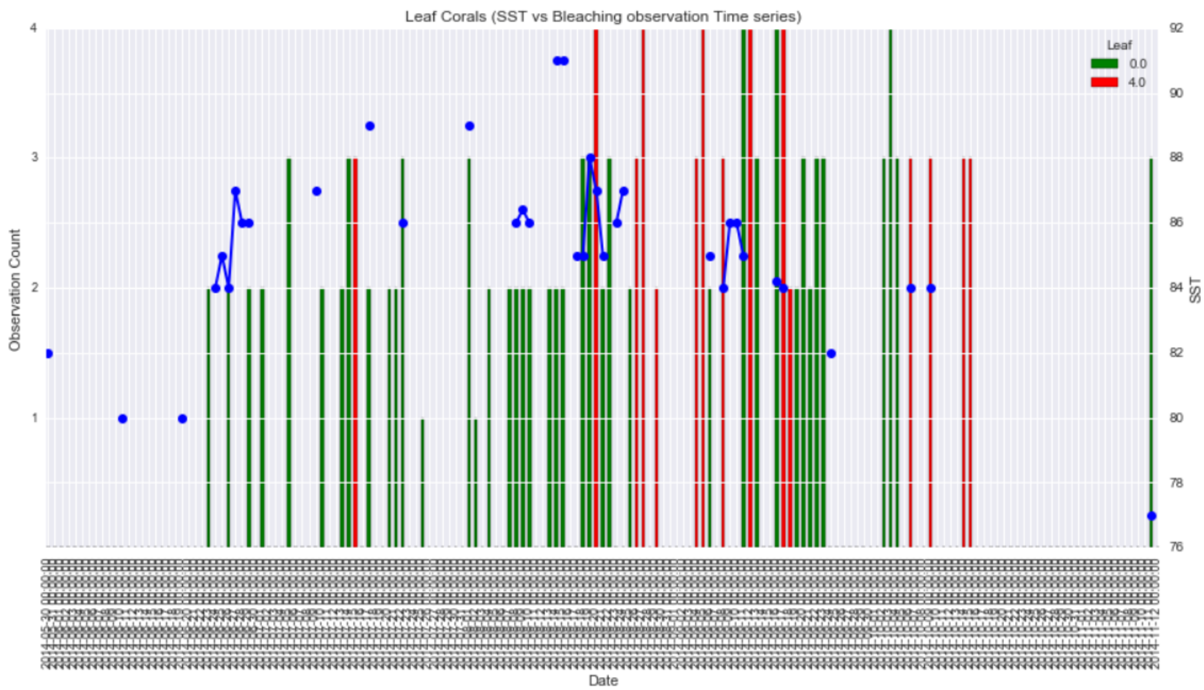



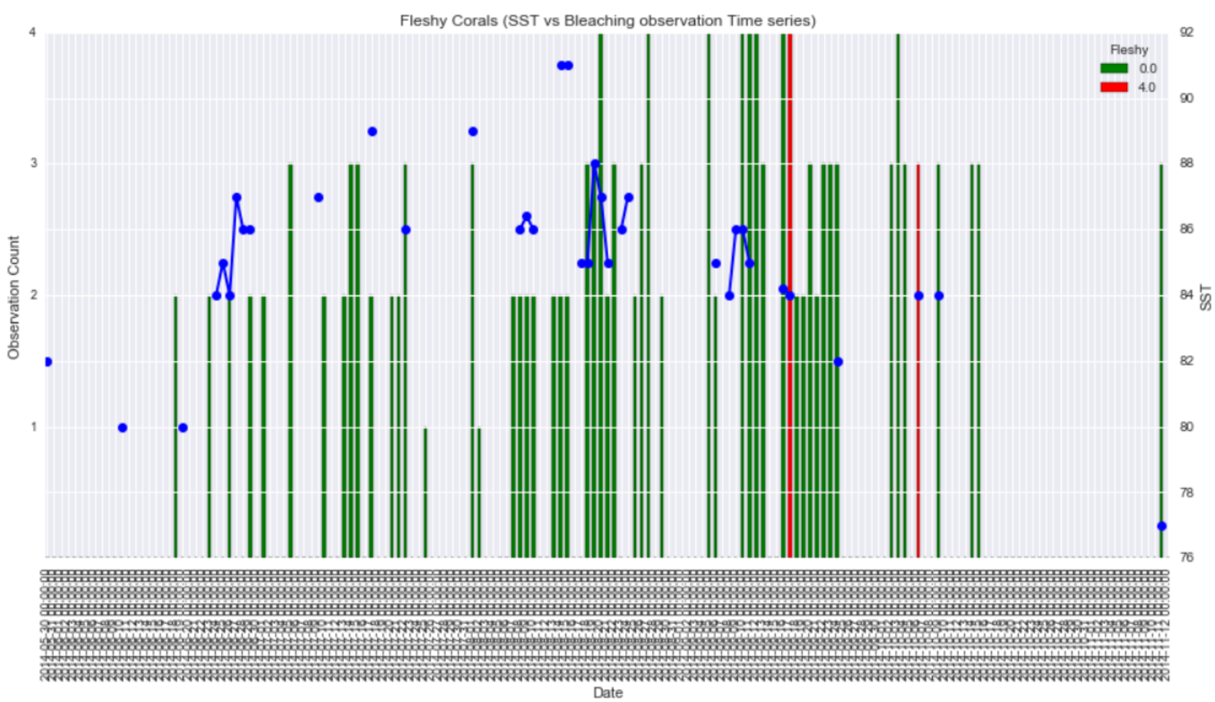


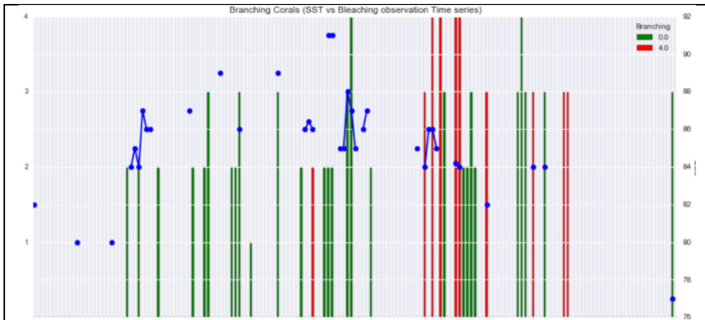
Encrusting Corals (SST vs Bleaching observation Time series)
Encrusting
0.0
4.0
Observation Count
SST
Date

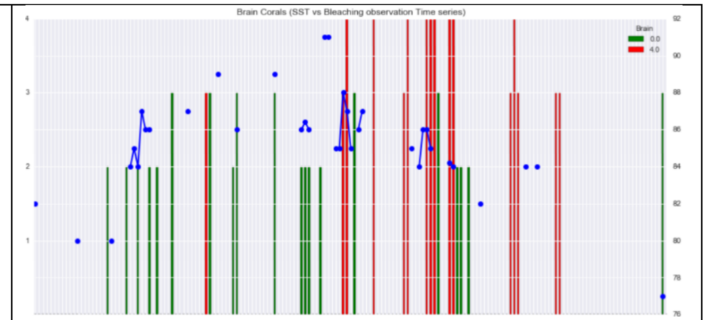



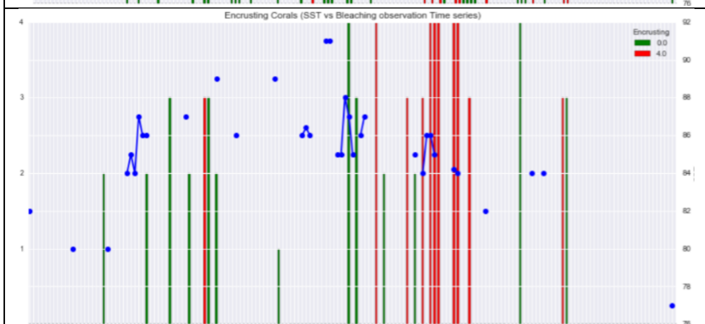
Flowering Corals (SST vs Bleaching observation Time series)
Flowering
0.0
4.0
Observation Count
SST
Date

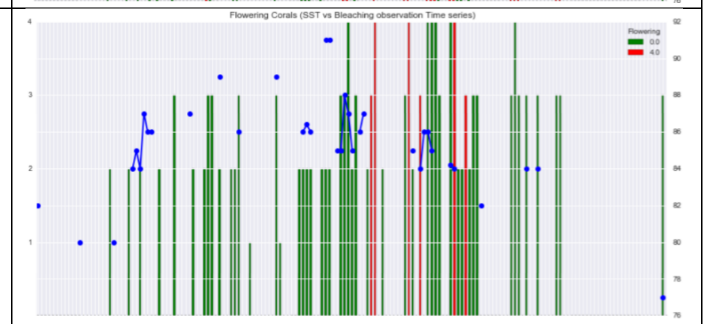

Leaf Corals (SST vs Bleaching observation Time series)
Observation Count
1
2
3
4
SST
76
78
80
82
84
86
88
90
92
Leaf
0.0
4.0
Date

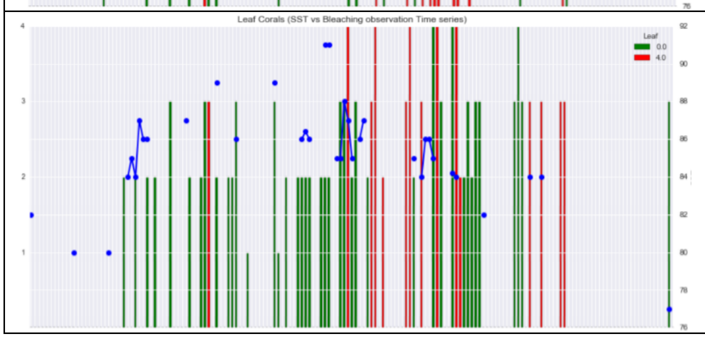

Fleshy Corals (SST vs Bleaching observation Time series)
Observation Count
1
2
3
4
SST
76
78
80
82
84
86
88
90
92
Fleshy
0.0
4.0
Date

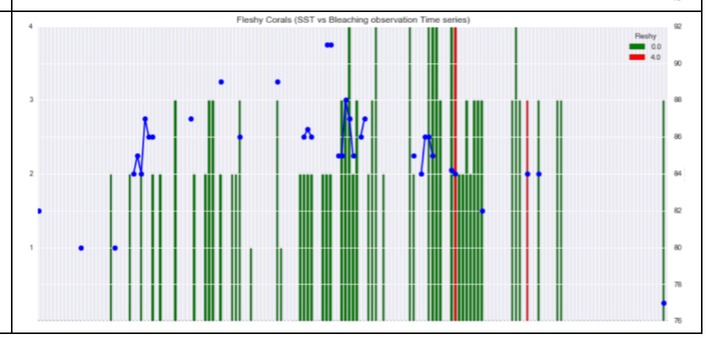

Branching Corals (SST vs Bleaching observation Time series)
Branching
0.0
4.0


Brain Corals (SST vs Bleaching observation Time series)
Brain
0.0
4.0


Encrusting Corals (SST vs Bleaching observation Time series)
Encrusting
0.0
4.0


Flowering Corals (SST vs Bleaching observation Time series)
Flowering
0.0
4.0


Leaf Corals (SST vs Bleaching observation Time series)
Leaf
0.0
4.0


Fleshy Corals (SST vs Bleaching observation Time series)
Fleshy
0.0
4.0