



FIT5147 Programming Exercise 1

Analysing coral data
in Tableau Public

Jaimon Thyparambil Thomas
Student ID : 29566428
Email : jthy0001@monash.student.edu
Monash University
March 17, 2019

Contents

Contents	2
Data Wrangling Part.....	3
Analysing Outliers	3 - 4
Questions	5 - 7
Question 1	5 - 6
Question 2	6 - 7

Data Wrangling Part

Python Code used for transforming and cleaning the data:

```
import pandas as pd

coral_data = pd.read_excel("assignment-01-data-unformatted.xlsx")
coral_data = coral_data.reset_index()
prev = ""
col_list = []
for each in coral_data.columns.values: # In this loop all the columns names with
    if "Unnamed" in each:              # Unnamed is been padded with the previous value
        each = prev                    # That is respective coral type name
    else:
        prev = each
    col_list.append(each)
coral_data.columns = col_list #Now the new column name list replaces the current column name
first_row = coral_data.iloc[0]
col = []
for i in range(len(coral_data.columns.values)): # Here for all the columns with year as its entry in
    if(type(first_row[i]) == float):             # the first row is combined with coral type and added
        col.append(str(int(first_row[i])) + " " + coral_data.columns.values[i])
    else:
        col.append(str(first_row[i]))            # name, latitude, longitude, 2017 soft coral,... etc
coral_data.columns = col                        # This new column list replaces the current column names
coral_data.drop(coral_data.index[0], inplace=True) #First row is been dropped as it was old column names
coral_data = pd.melt(coral_data, id_vars=['name','longitude','latitude'],var_name='coral_type')
coral_data.insert(4,'year',0)                  #The table structure is been transformed and new column year is added
# Now the year from '2017 soft coral' is been extracted and replaced into the year column
coral_data["year"] = [int(each.split()[0]) for each in coral_data["coral_type"] ]
# Now the coral type column value is been replaced from '2017 soft coral' to 'soft coral'
coral_data["coral_type"] = [" ".join(each.split()[1:]) for each in coral_data["coral_type"] ]
coral_data.to_csv('coral_data.csv')            #transformed table is written to new file
```

Data after it has been cleaned looks like:

	name	longitude	latitude	coral_type	year	value
0	site01	143.515	-11.843	soft corals	2017	0.8387
1	site02	147.898	18.937	soft corals	2017	0.2123
2	site03	144.081	-10.321	soft corals	2017	0.7534
3	site04	150.444	-20.414	soft corals	2017	0.1245
4	site05	143.786	-13.107	soft corals	2017	0.9423

Table column details is being mentioned below:

Name, longitude, latitude - represents the respective values of coral site

coral_type - Type of the coral analysed in the respective site

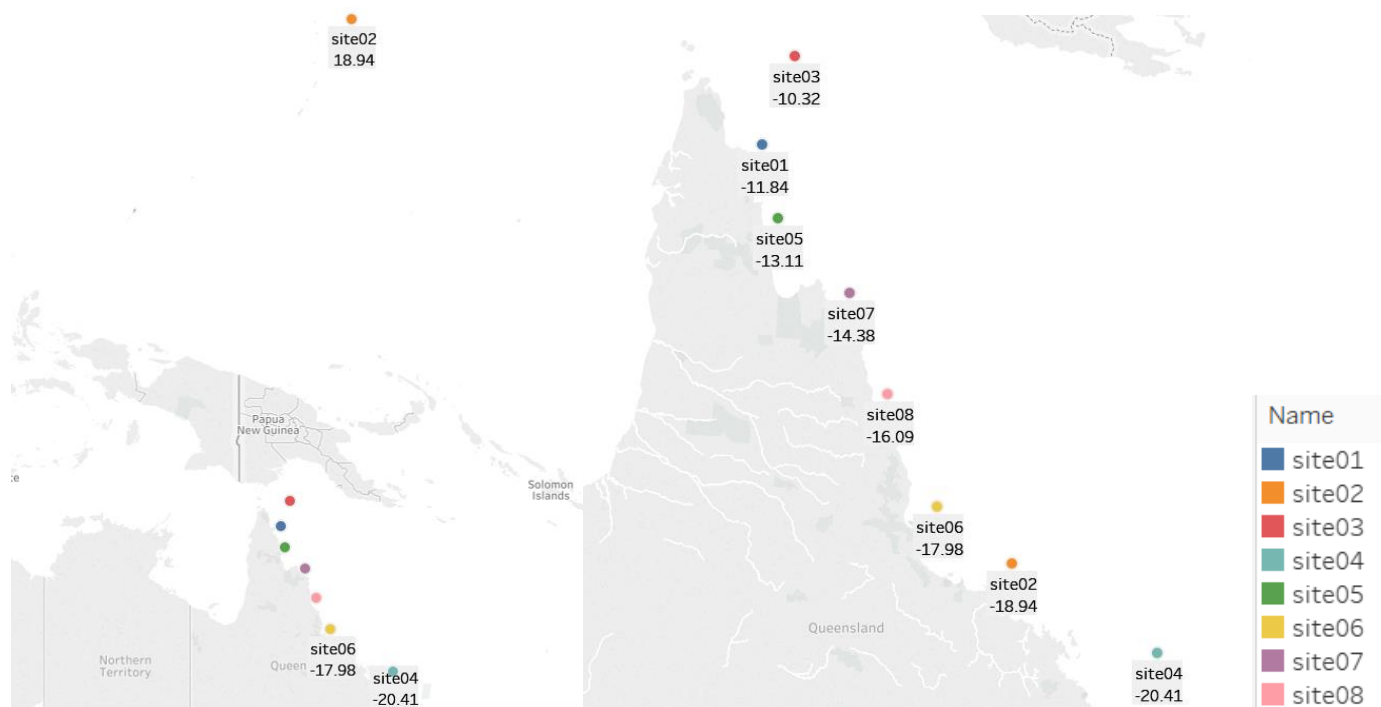
year - year in which the observation was made

value - represents bleaching rate for the respective category

Analysing Data and Handling Outliers

On Carefully analysing the data we can see that there are lots of outliers in the data which has to be fixed in order to get actual answer which we are seeking for such as

In Figure 1 the value of site 02 looks like an outlier. I feel that it is a data entry error as the rest of the data has its latitude negative. Only this one has its value as positive. I feel if the latitude of the site 02 had been -18.94 then it would have been in perfect sync with rest of the values. Thus I have changed the value to -18.94 for site 02. Now the new sites map looks like figure 2



Before Fixing outlier (Figure 1)

After Fixing outlier (Figure 2)

Similarly, on carefully analysing the data of bleaching rates in figure 3 we can see that there is a drastic spike in value in site 08 for hard coral in year 2014 that is 148%. On analysing the data, we can see that it looks like it is a data entry error as the neighbouring values are 13.84% and 14.56%. So I believe the actual value is 14.8%. So I have changed this value in excel sheet. Similarly, we can see that there is a drastic dip in site 07 for blue coral that is .4696%. On analysing the data, we can see that this also looks like the data entry error as the neighbouring values are 47.3% and 45.6%. So I believe the actual value is 46.96%. So I have changed this value also in the excel sheet. So rest of the analysis is made after making these changes.

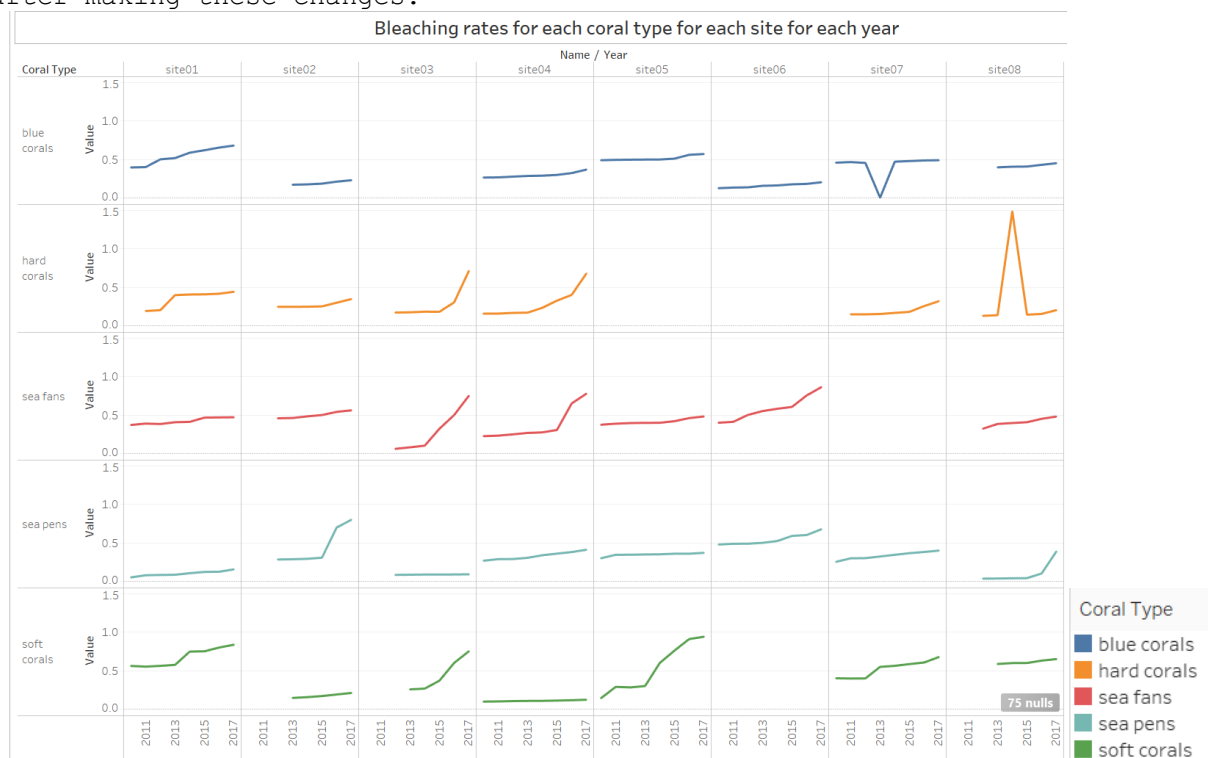


Figure 3

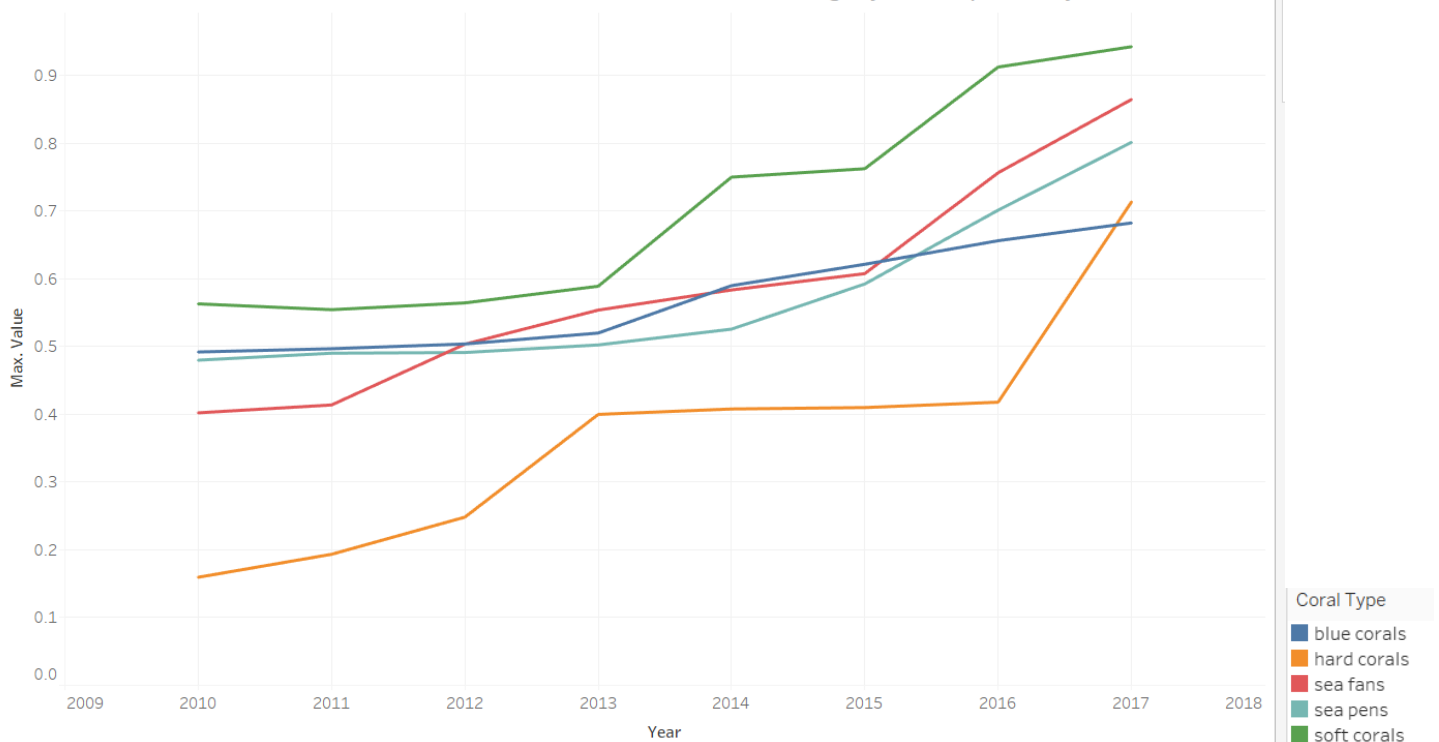
Questions to Be Analysed

1) In which years and for which kinds of coral bleaching is the worst

If we consider the question to be which is the worst coral bleaching in each year. Then this question can be interpreted as two cases as shown below

First case: we can take the max bleaching value out of all sites of each coral type and plot it with respect to each year. To find the worst bleaching site in each year. For this case the plot will look like the one below

Plot of maximum value out of all sites for each coral category for respective year



From the graph it is pretty evident that for all years the worst bleaching is for soft corals.

Second case: we can take the average value of bleaching rate of all sites for the respective coral type and thus find the worst bleaching coral category in each year. For this case the graph is given in figure 4.

As Per figure 4 we can see that maximum average values of each sites for respective coral types are in year 2010 for blue coral, in 2011 for sea fans, in 2012 for blue coral, in for 2013 sea fans then from 2014 till 2016 for soft coral and in 2017 for sea fans.

This graph in figure 4 shows us the category of corals which has the worst bleaching rates.

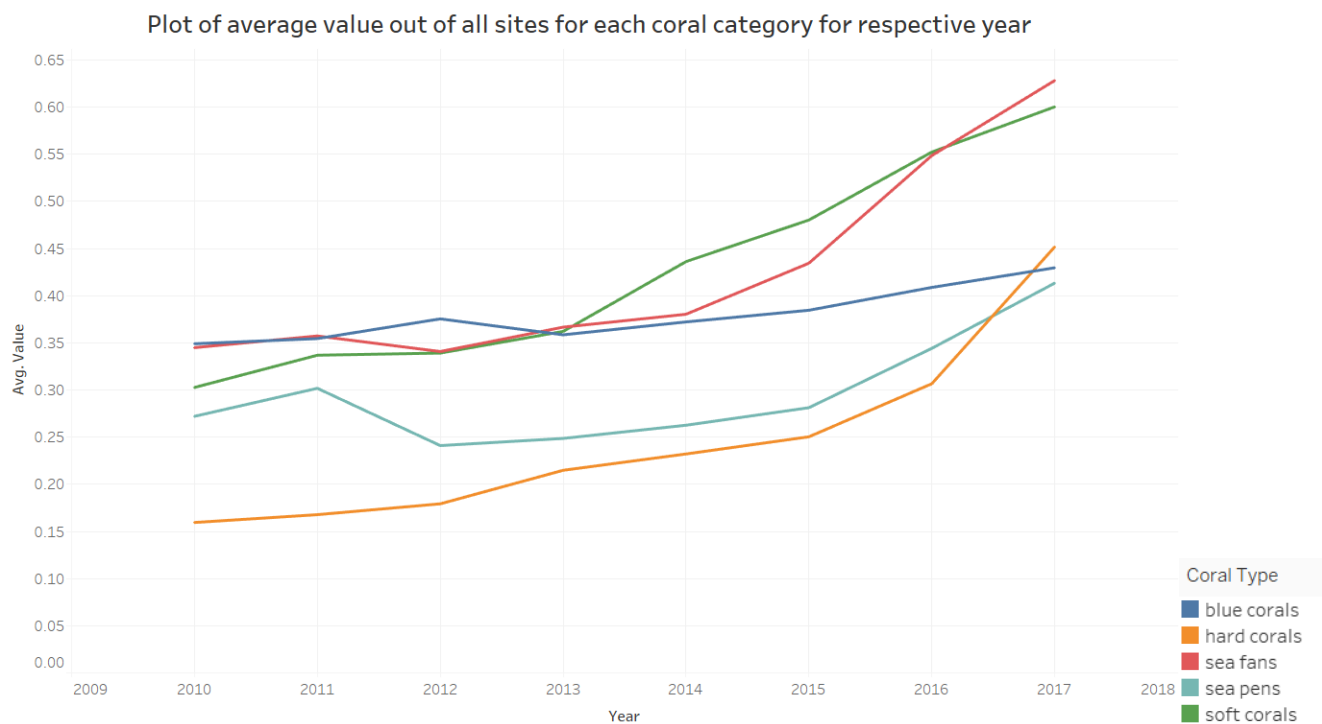
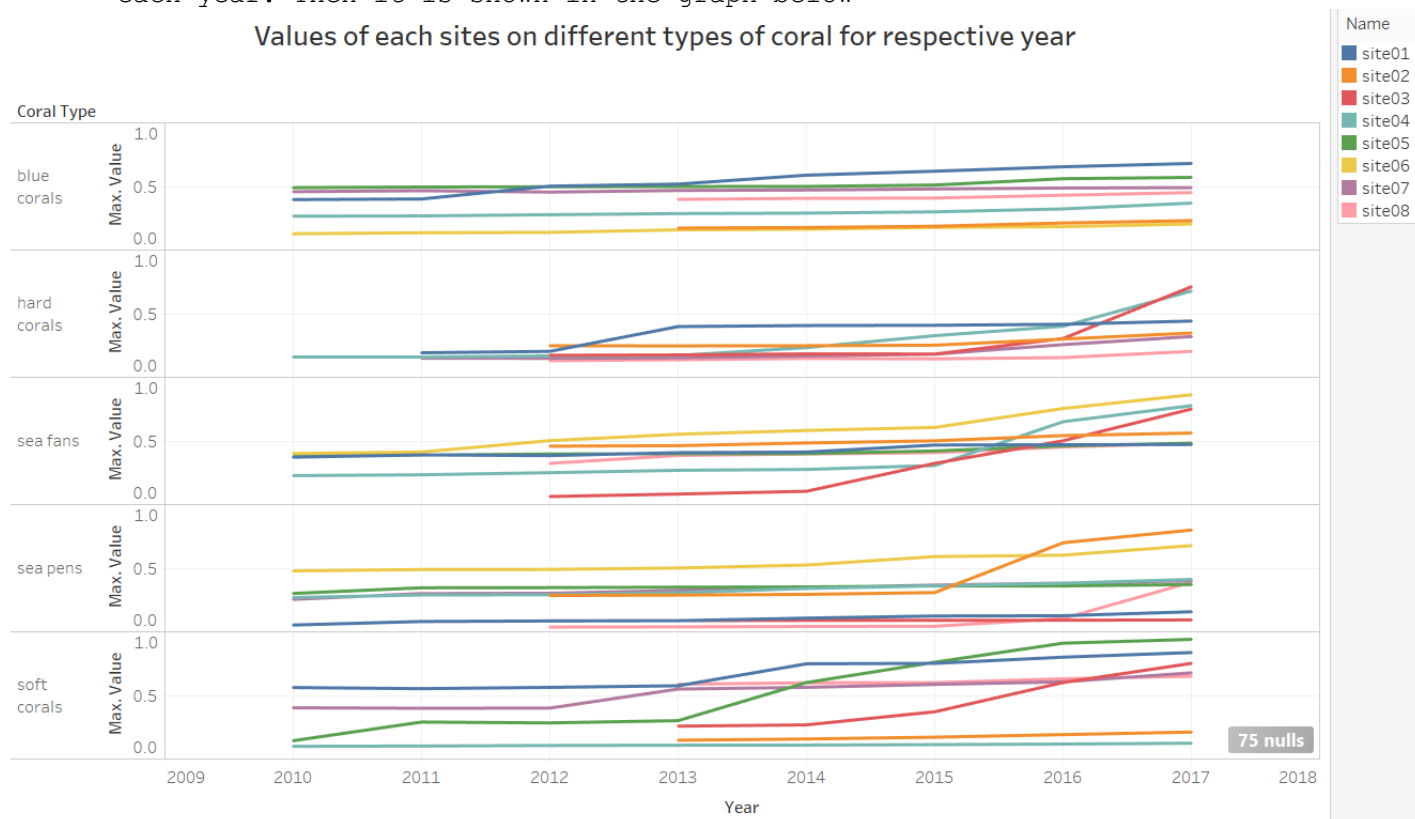


Figure 4

If we consider the question to be which site has the worst coral bleaching in each year. Then it is shown in the graph below

Values of each sites on different types of coral for respective year



2) How the location of the site affects bleaching on the different kinds of coral?

The graphs below represent how the location of the site affects the bleaching on different kinds of coral

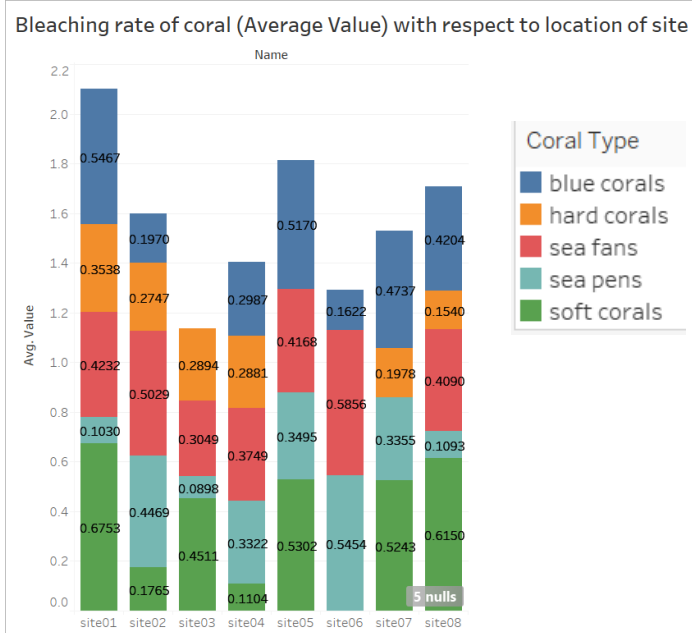


Figure 5

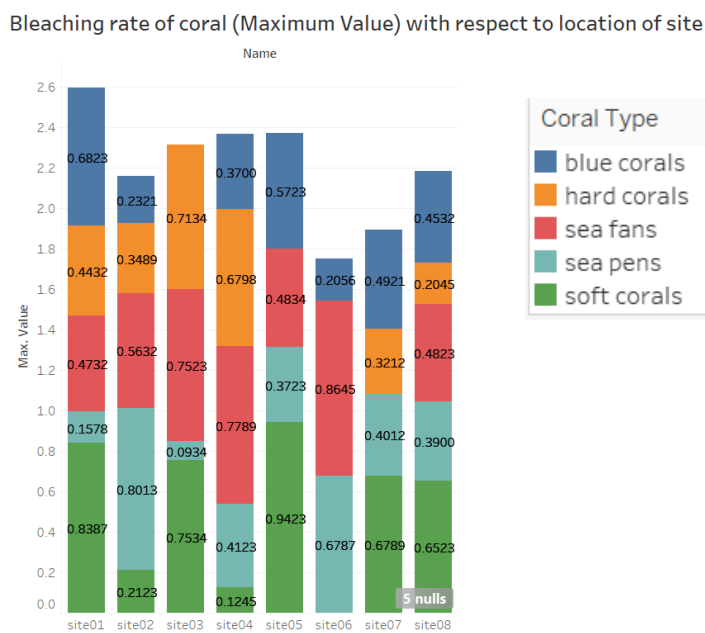


Figure 6

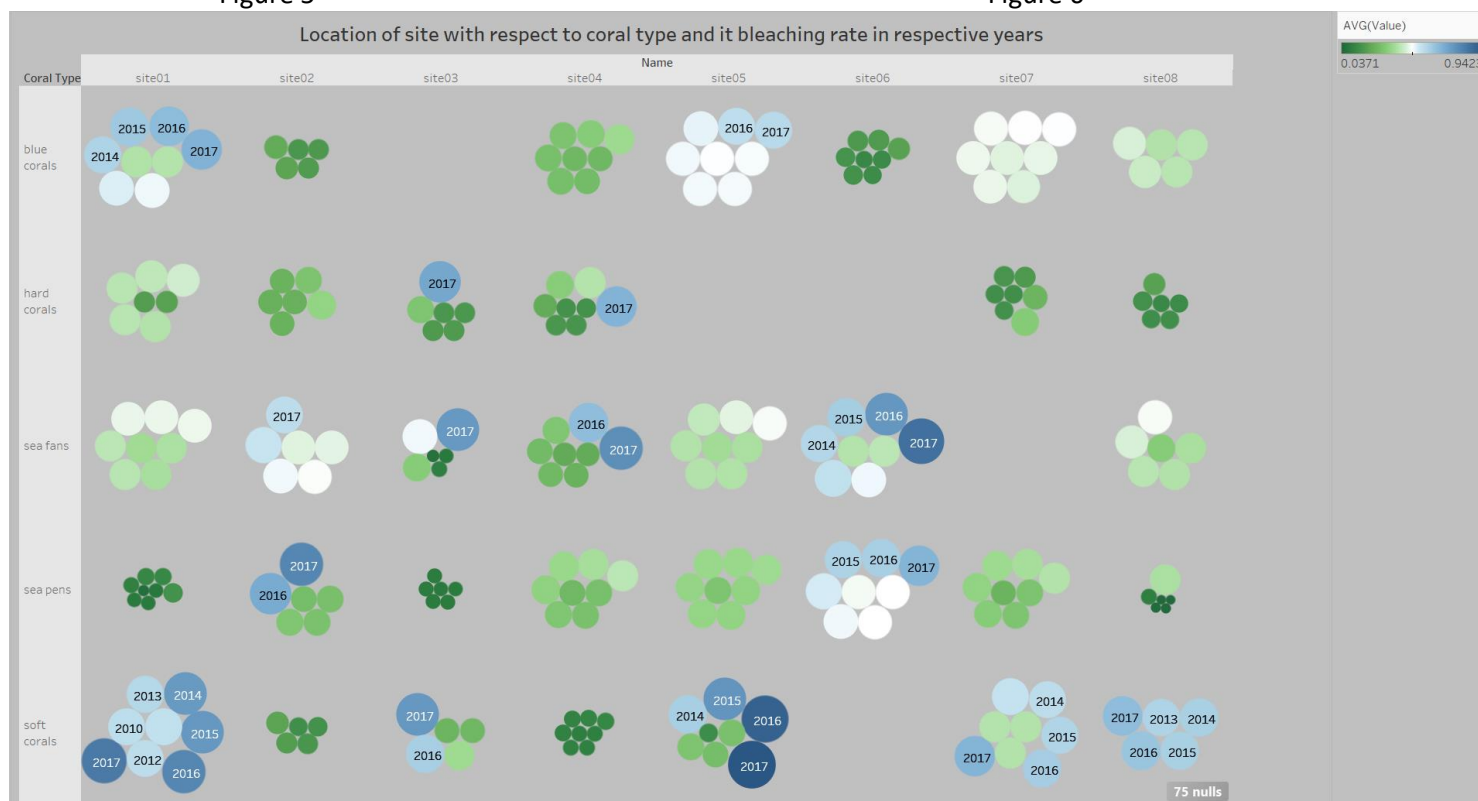


Figure 7

From Figure 5 it is pretty much clear that how bleaching rates of various corals based on average value per year changes as per the site of the location.

From Figure 6 it is pretty much clear that how bleaching rates of various corals based on maximum value per year changes as per the site of the location

Figure 7 gives a general overview on these aspects. From figure 7 it is pretty much clear that in year 2017 and in most of the year's soft coral has worst bleaching rates among all other sites except for site 2 and site 4. In site 2 sea pens has worst bleaching rate whereas in site 04 sea fans has the worst bleaching rates.