

FIT5147 Programming Exercise 1

--------------------

Analysing coral data

in Tableau Public

Jaimon Thyparambil Thomas

Student ID : 29566428

Email : [jthy0001@monash.student.edu](mailto:jthy0001@monash.student.edu)

Monash University

March 17, 2019

Contents

Contents …………………………………………………… 2

Data Wrangling ……………………………………………. 4

Questions ………………………………………………….. 3 - 6

Question 1 …………………………………………… 3 - 5

Question 2 …………………………………………… 5 - 6

Analysing Outliers …………………………………………. 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-unformated.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: # to a list. Now the values in this list looks like

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

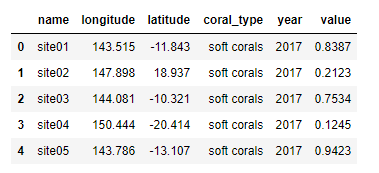
# Now the year from ‘2017 soft coral’ is been extracted and replaced into the year column

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:

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*

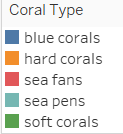
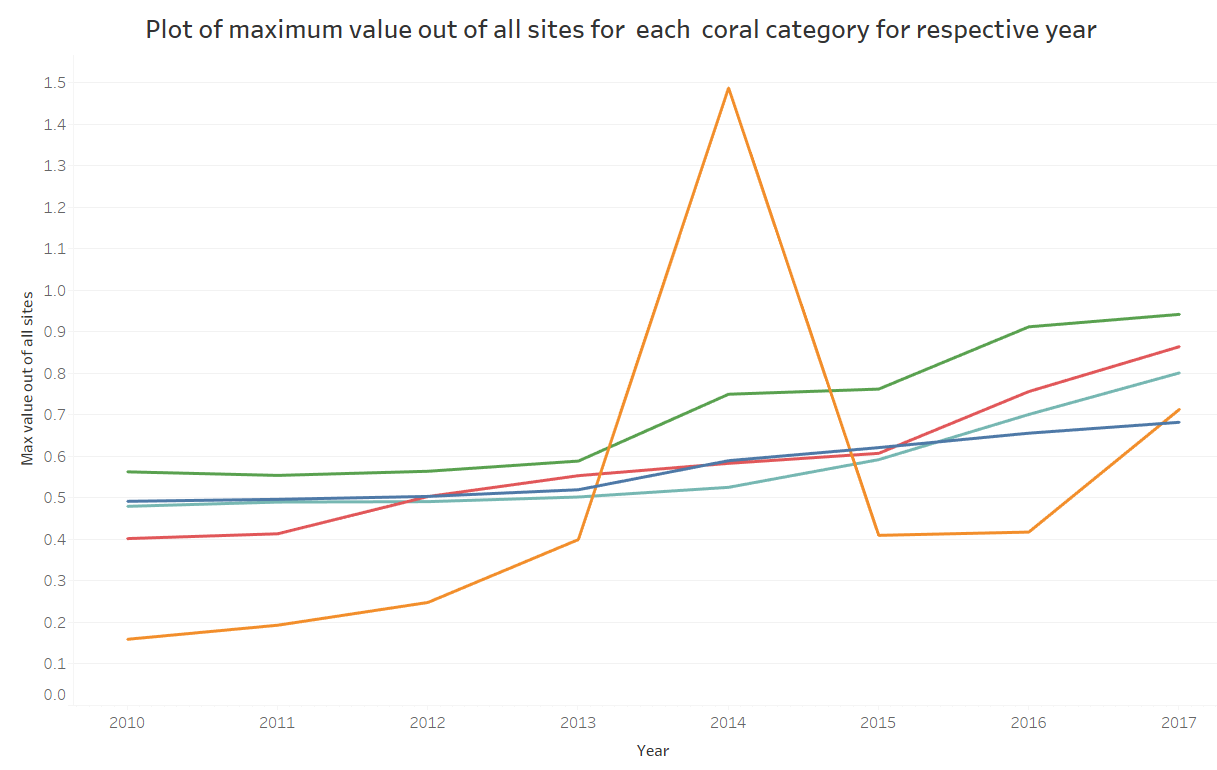
*value* – represents bleaching rate for the respective category

Questions

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

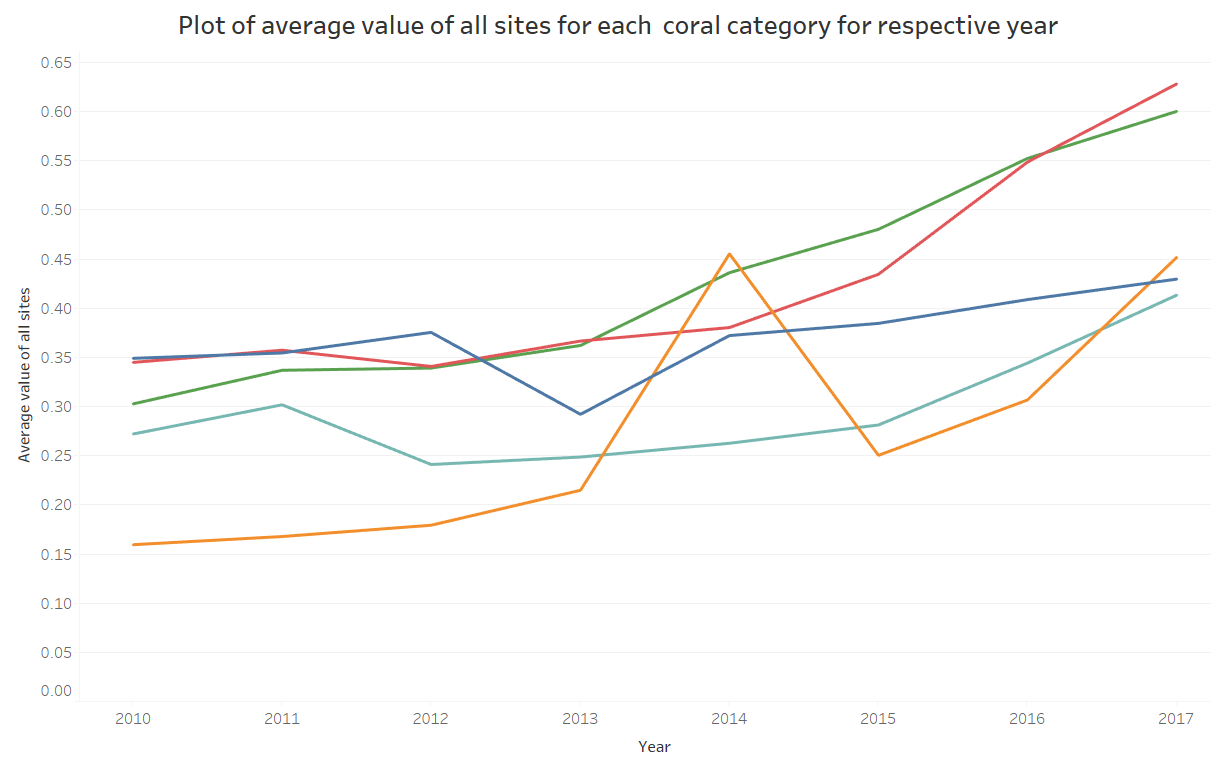
If we consider the question to be which is the worst coral bleaching in each year. Then this question can be interpreted as two cases as shown below

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 on below

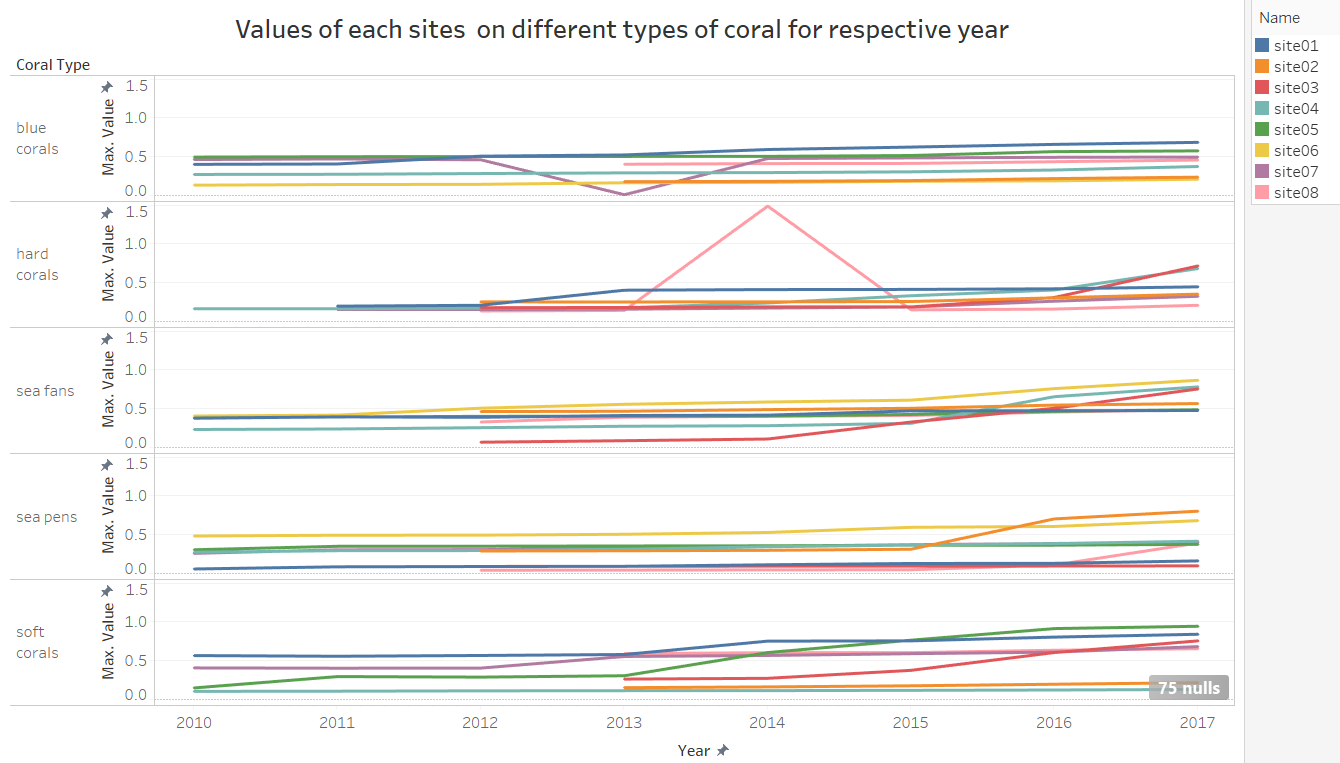


From the graph it is pretty evident that for all years the worst bleaching is for soft corals except for the year 2014. In this year as per the data given hard coral has the highest bleaching. But I feel that value is an outlier since that sudden peak value doesn’t tally its rest of the value. So I feel in the year 2014 also soft coral has the highest bleaching rate.

Second case: we can take the average value of beaching 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 below

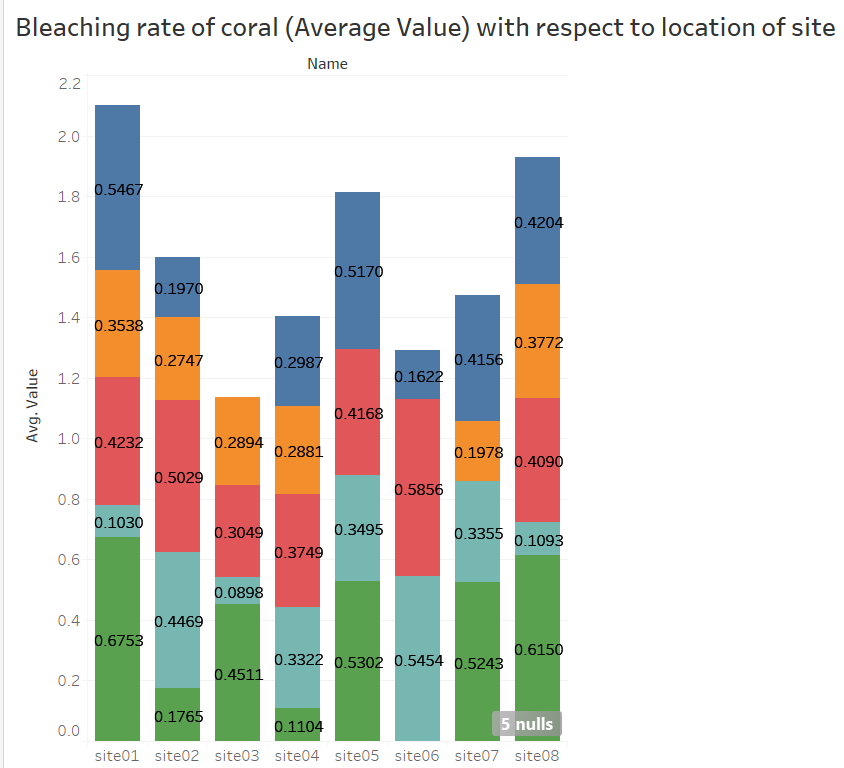
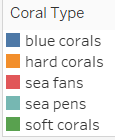
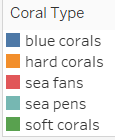
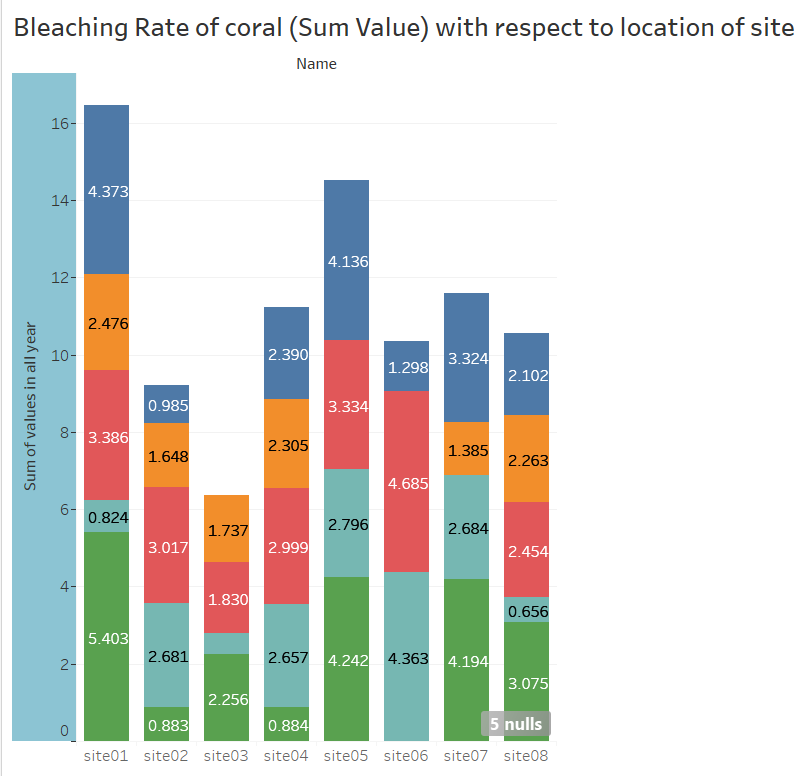


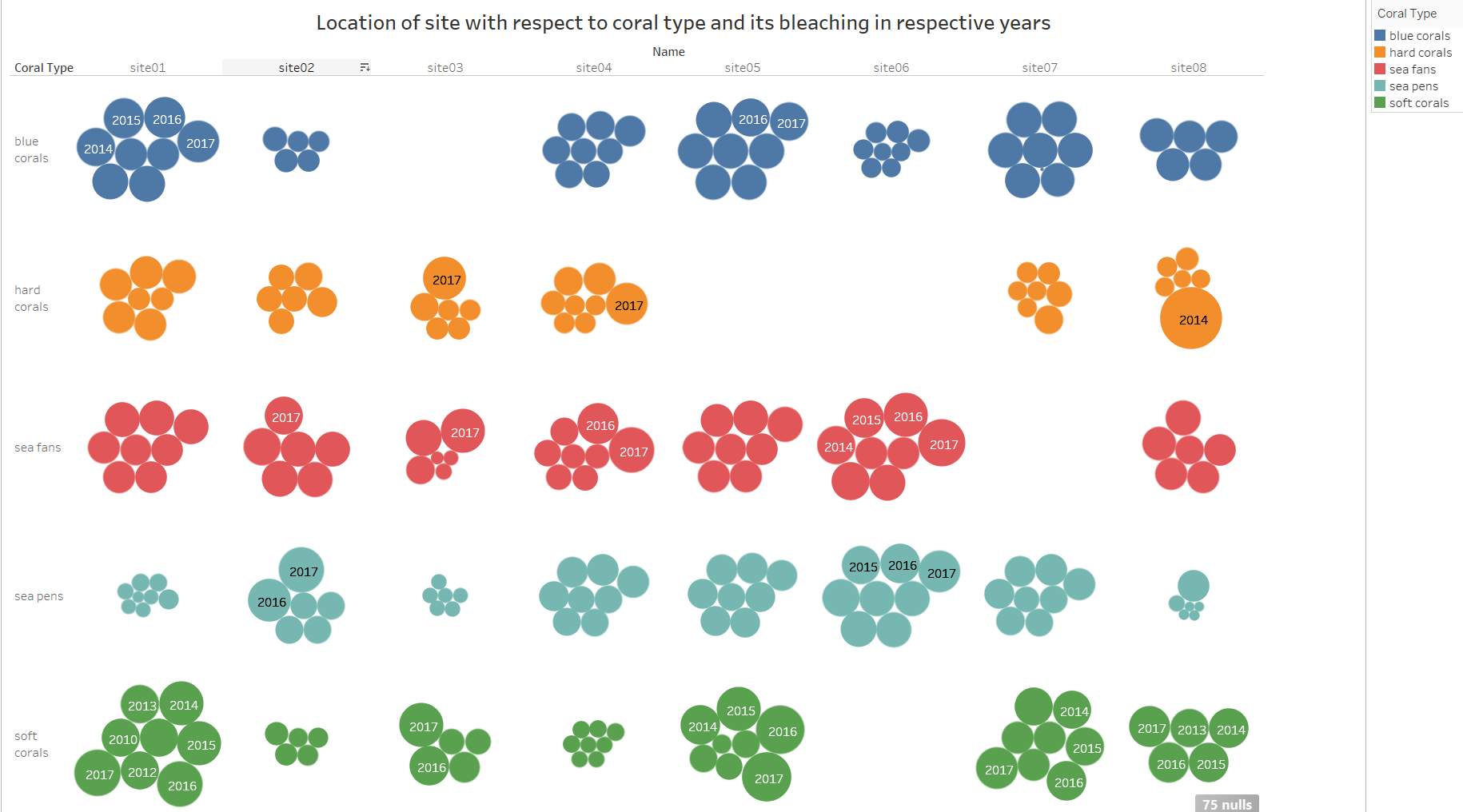
If we consider the question to be which site has the worst coral bleaching in each year. Then it is shown in the graph below



1. How the location of the site affects bleaching on the different kinds of coral.

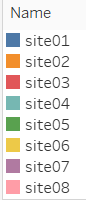
The graphs below represent how the location of the site affects the bleaching on different kinds of coral



Analysing 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 this case value of site 02 looks like an outlier. I feel that it is a data entry as the rest of the data has its latitude negative only this one has it 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.

On carefully analysing the data we could see that there are many other outliers in the data values such as the points lying outside the box plot in the below figure 1 and also the sudden peaks or fluctuations in fig 2

