

# ASSIGNMENT 1

Name: Parul

Student id: 29507960

Tutorial Number: 01-P1

Tutor Name: Farah Tasnuba Kabir

## Data Wrangling

We are provided with fictional data on coral bleaching from 8 sites in the Great Barrier Reef. The data represents the percentage of bleaching for different five kinds of coral i.e., hard corals, sea pens, blue corals, soft corals and sea fans over the last 8 years from 2010 to 2017. The raw data needs to be reformatted to use with Tableau for visualisation. Data wrangling includes two steps: data integration & reshaping and data exploration & cleansing. We will start with reshaping the provided data according to our convenience.

## Data Reformatting

The data set provided is not in suitable format to be used in Tableau for visualisation. The data reshaping has done using python (fig 1). The provided data is converted into a row-oriented table with less number of rows as compared to the original format.

```
df = pd.read_excel('assignment-01-data-unformatted.xlsx')
df = df.reset_index(drop=False)

dfl = list()
col = ['soft corals', 'sea fans', 'blue corals', 'hard corals', 'sea pens']
for x in range(0,5):
    dfl = df.iloc[:, :11]
    dfl.columns = dfl.iloc[0]
    dfl = dfl[1:]
    res = pd.melt(dfl, id_vars=['name', 'longitude', 'latitude'], var_name=['Year'])
    res = res.rename(columns = {'value': col[x]})
    dfl.append(res)
    df.drop(df.columns[[3,4,5,6,7,8,9,10]], axis=1, inplace=True)
df_merged = reduce(lambda left, right: pd.merge(left, right, on=['name', 'longitude', 'latitude', 'Year'],
                                                how='outer'), dfl)
data = pd.melt(df_merged, id_vars=['name', 'longitude', 'latitude', 'Year'], var_name=['Corals'])
```

Fig 1: python coding for reshaping

Now, data is reshaped into a new structure with 360 rows × 6 columns. The column names are as follow:

- name: Name of The Site
- latitude: Latitude of the Site
- longitude: Longitude of the Site
- Corals: Type of Coral (blue, hard, soft, sea fans, sea pens)
- Year: Year of Bleaching
- value: Percentage of the Corals Bleaching

The row contains the unique record of percentage of bleaching of each coral for eight years from eight site. Though not all sites have data for the entire period, and not all corals are found at each site, so there are some null values which are kept as it is while reshaping. Fig 2. shows the new data structure.

	name	longitude	latitude	Year	Corals	value
0	site01	143.515	-11.843	2017	soft corals	83.87
1	site02	147.898	18.937	2017	soft corals	21.23
2	site03	144.081	-10.321	2017	soft corals	75.34

Fig 2: reformatted data

## Reading data into Tableau

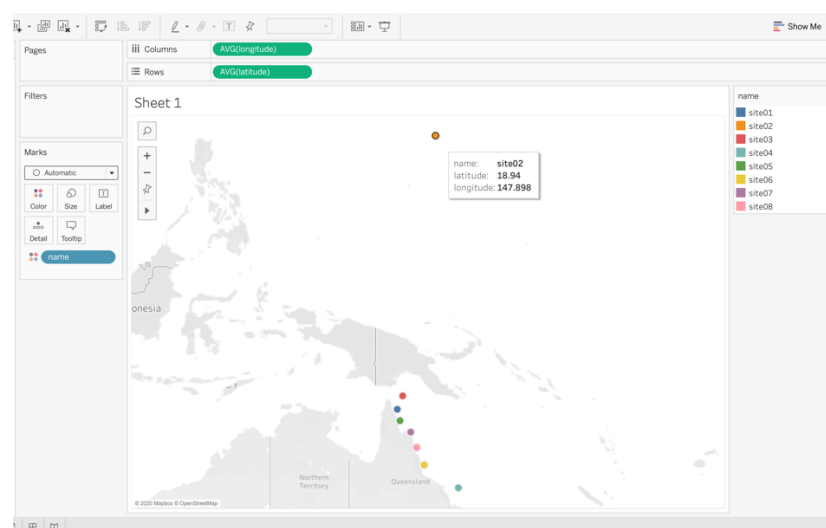
Data is loaded into Tableau Public by drag and dropping file into Tableau window. Tableau will use this data for exploration and visualisation. Below is an image provided for how data looks like:

<div><div><div><div><div></div></div></div><div><div>Sheet1.F1</div><div>F1</div></div></div></div>	Sheet1 name	<div><div><div><div></div></div></div><div>Sheet1 longitude</div></div>	<div><div><div><div></div></div></div><div>Sheet1 latitude</div></div>	<div><div><div><div></div></div></div><div>Sheet1 Year</div></div>	Sheet1 Corals	<div><div><div><div></div></div></div><div>Sheet1 value</div></div>
0	site01	143.51500	-11.8430	2017	soft corals	83.8700
1	site02	147.89800	18.9370	2017	soft corals	21.2300
2	site03	144.08100	-10.3210	2017	soft corals	75.3400
3	site04	150.44400	-20.4140	2017	soft corals	12.4500

## Data Exploration

Some visualisations were performed to see if data is clean. On plotting latitude and longitude on map, the graph below (fig 3) shows that the geographical location of site 2 is far away from the location of other sites.

Fig 3: incorrect location of site 2



Upon inspecting the data, it was found that the latitude for site 2 was positive 18.94 whereas latitude is negative for all other sites. So, this must be a syntactical error while entering the data. Therefore, the latitude of site 2 was manually changed to negative (-18.94) in excel which resulted in correct location near other sites as shown below:

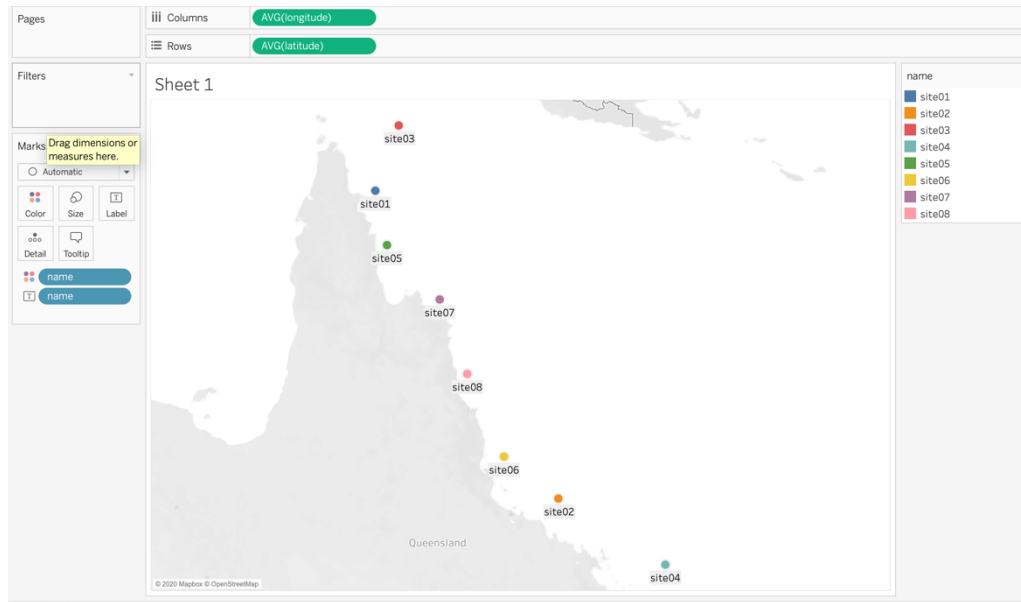


Fig 4: corrected location of site 2

Plotting the dimensions (corals and year) against average value (bleaching percentage) showed line graph (fig 5) for average percentage of bleaching of coral every year with each coral type in different colour line. The image below shows some unexpected high and low peaks for hard corals and blue corals respectively.

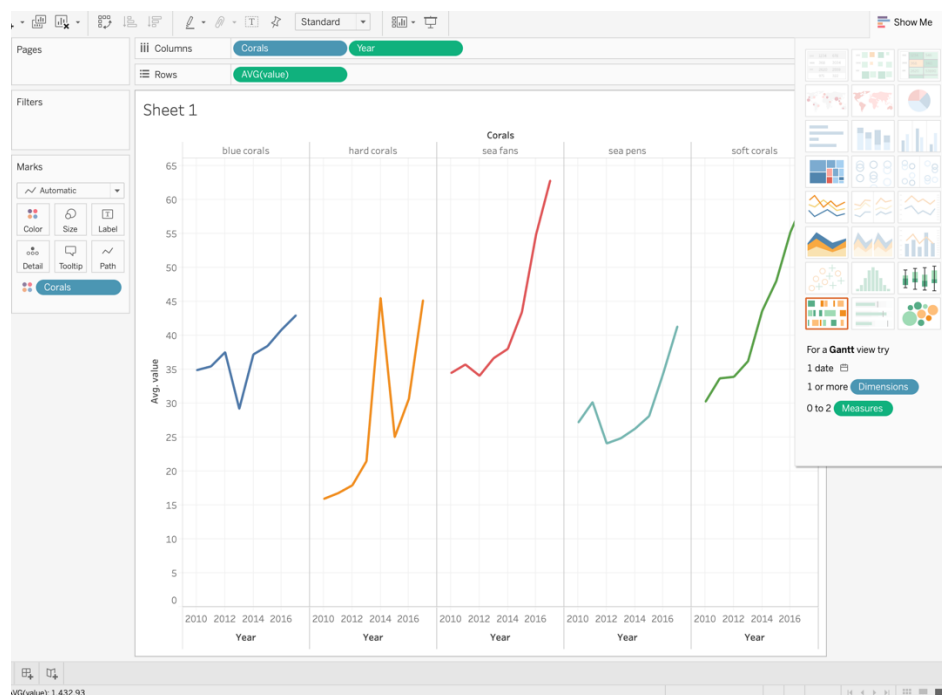


Figure 5

On studying the data closely, it was found that the bleaching percentage value for hard corals for 2014 on site 8 was 148.80 % which is a typing error of decimal. So, the value should be 14.88% instead. Similarly, the percentage value for blue corals in 2013 on site 7 responsible for low peak was 0.4696% which must be 46.96% instead. Figure 6 shows the plot after cleansing of data without and error.

## Answering Questions

Q1) In which years and for which kinds of coral is bleaching the worst?

This question can be answered on viewing the Figure 6, line plot for average percentage of bleaching of each coral in eight years.

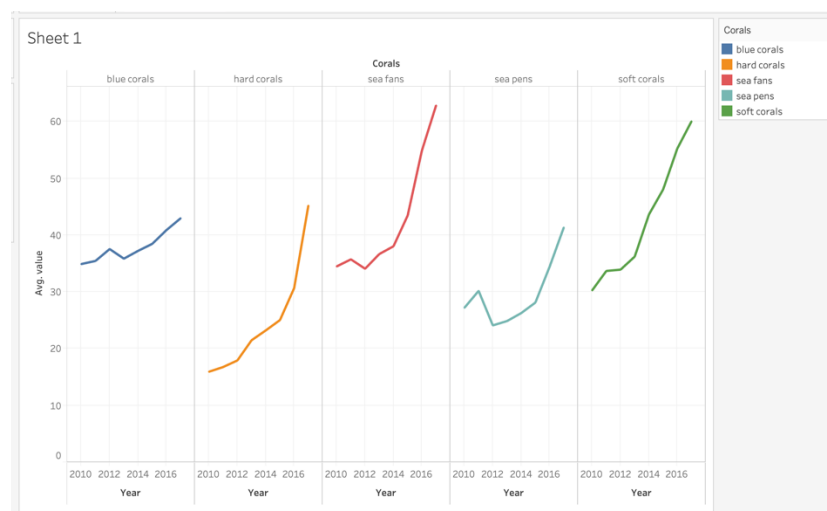


Figure 6: average percentage of bleaching of each coral in eight years

The graph clearly depicts that the bleaching percentage is was increasing with a high rate from year 2014 to 2017 for almost every coral. Initially, in 2010 to 2013 blue corals had the highest bleaching percentage compared to others. But after 2014, sea fans and soft corals were the top two most bleached corals for three consecutive years whereas blue corals, hard corals and sea pens has a gradual increase rate. It can be seen through the plot, that in year 2017 sea fans were bleaching at an alarming rate with highest value of bleaching percentage.

Q2) How does the location of the site affect bleaching of the different kinds of corals?

Using bubble map, we have plotted each coral with different colour bubble on every site with bleaching percentage controlling the size of the bubble (figure 7). Sea fans has high bleaching percentage throughout on all sites whereas this is not the case with blue corals and soft corals. We can see that blue and soft corals have comparatively higher bleaching percentage on sites at northern location than that in southern part. Whereas, sea pens follow the opposite pattern with higher bleaching at southern location. Hard coral seems to have higher bleaching on the edgy sites at north and south with slightly lower percentage at the central locations on site 5, 6 and 8. The differences in the size of bubble clearly shows the effect of location of site on bleaching of different kind of corals.

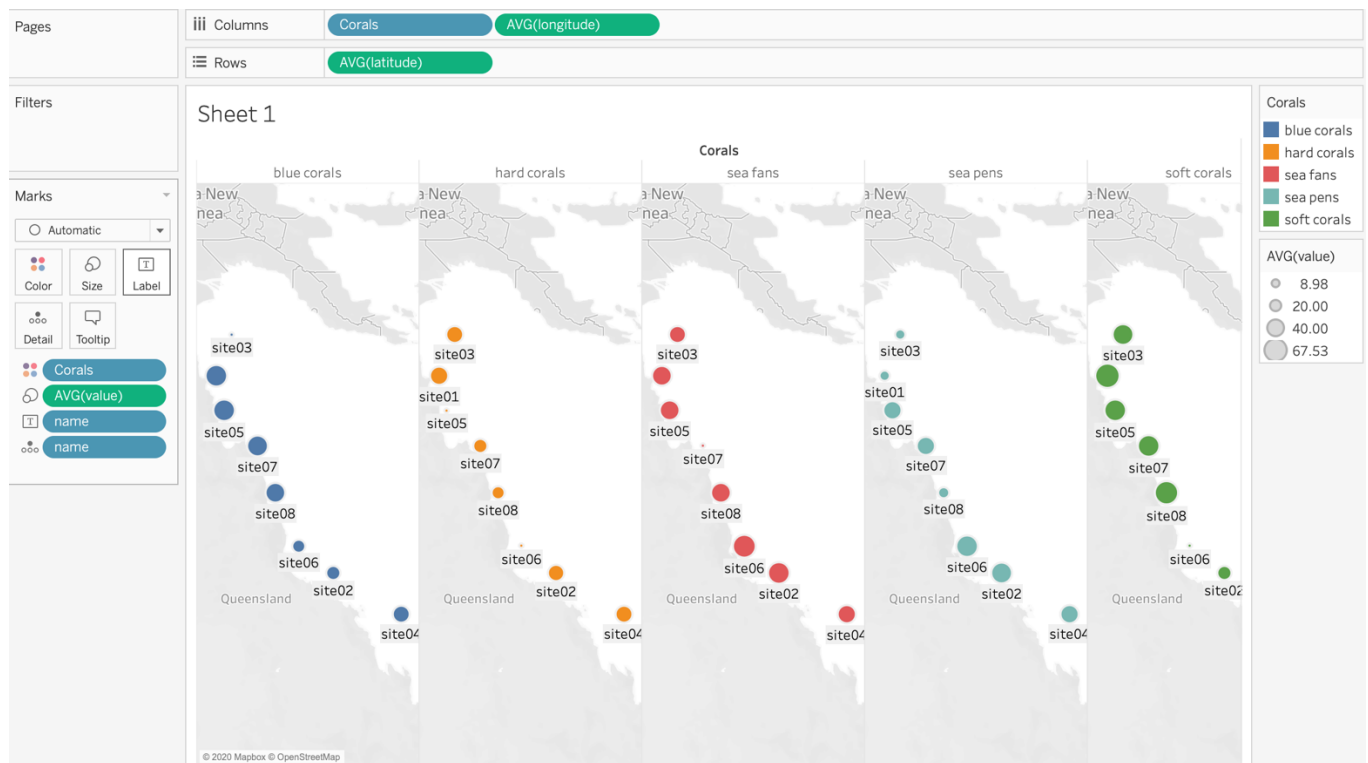


Figure 7: effect of location of the site on bleaching of the different kinds of corals