FIT5147 Data Exploration and Visualisation
Programming Exercise 1 - Tableau Public

Student Name: Gayatri Aniruddha
Student ID: 30945305
Tutorial Number: 22
Tutor Name: Pratik Bhumkar

1. DATA WRANGLING

Data wrangling is the process of gathering, selecting, and transforming data to answer an analytical question. There were multiple steps that were followed to clean the dataset and rearrange it into a format that can be used to visualize the data properly. I have rearranged the data using Microsoft Excel.

1.1 Steps involved in Data Wrangling

Step 1:
First, I changed the column names so that we knew which year corresponded to which type of coral for a clearer understanding. Also, I added a new column - Type which stored the type of coral for each site.

Step 2:
I then separated out the percentage of bleaching in all years for each type of coral. Next, I appended all the individual values of the different types of corals. However, the years from 2010 to 2017 were still kept as separate columns.

Step 3:
Furthermore, I added a new column - Year, which stored the year corresponding to a particular type of coral. Thus, our final formatted data has five columns - Name of the site, longitude, latitude, type of coral, year and percentage.
Finally, all the NaN values were left as such and were left blank.
The final formatted version of the corrected data looks as follows in tableau.

F1	Name	Longitude	Latitude	Type	Year	Percentage	Percentage(%)
0	site01	143.51500	-11.8430	Soft Corals	2017	0.838700	83.8700
1	site02	147.89800	-18.9370	Soft Corals	2017	0.212300	21.2300
2	site03	144.08100	-10.3210	Soft Corals	2017	0.753400	75.3400
3	site04	150.44400	-20.4140	Soft Corals	2017	0.124500	12.4500
4	site05	143.78600	-13.1070	Soft Corals	2017	0.942300	94.2300
5	site06	146.58900	-17.9810	Soft Corals	2017	*null*	*null*
6	site07	145.04300	-14.3830	Soft Corals	2017	0.678900	67.8900
7	site08	145.71500	-16.0910	Soft Corals	2017	0.652300	65.2300
8	site01	143.51500	-11.8430	Sea Fans	2017	0.473200	47.3200
9	site02	147.89800	-18.9370	Sea Fans	2017	0.563200	56.3200
10	site03	144.08100	-10.3210	Sea Fans	2017	0.752300	75.2300
11	site04	150.44400	-20.4140	Sea Fans	2017	0.778900	77.8900
12	site05	143.78600	-13.1070	Sea Fans	2017	0.483400	48.3400
13	site06	146.58900	-17.9810	Sea Fans	2017	0.864500	86.4500
14	site07	145.04300	-14.3830	Sea Fans	2017	*null*	*null*
15	site08	145.71500	-16.0910	Sea Fans	2017	0.482300	48.2300

Figure 1 : Final formatted data

I have also added a new column in Tableau - Percentage(%) - This is a derived column from Percentage, where Percentage(%) = Percentage*100

2. DATA EXPLORATION

2.1 Errors Found

2.1.1 Error 1: Positive Latitude Values - A logical error!

It was observed that Site 2 had positive latitude values. They were mentioned as 18.937 instead of -18.937. When we plotted this graph, all the other sites were located in the North Eastern parts of Australia - in the Great Barrier Reef region. However, site 2 was located much above the other sites mentioned. Thus, this is clearly an example of a logical error. After making this value negative, this site also came in the same region - The Great Barrier Reef region as the other sites.

With Error:

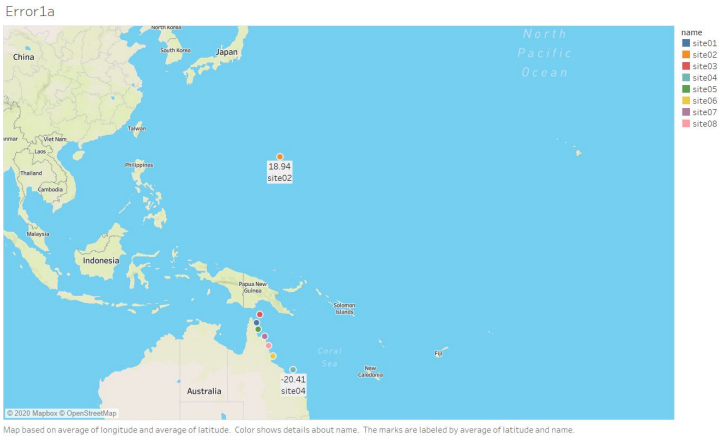

Figure 2.1 : Error 1 - Incorrect Latitude Values

Without Error :

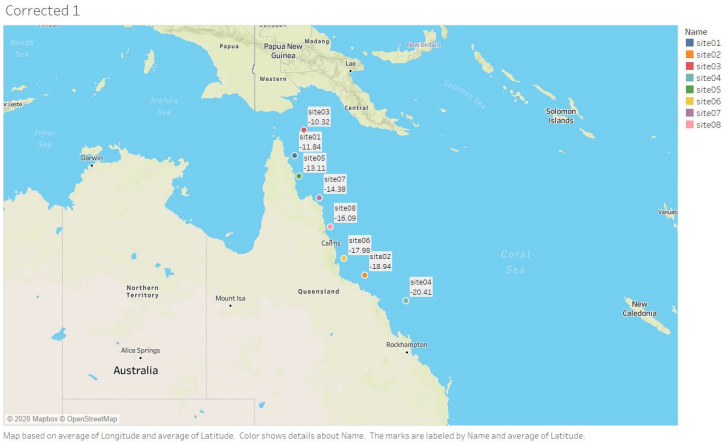

Figure 2.2 : Corrected Latitude Values

2.1.2 Error 2: 148% in Site 8 - Hard Corals in 2014.

While going through the data, I observed that in the year 2014, the percentage of bleaching for hard corals in site 8 was 148%. This is clearly a typing error and a logical error as the

percentage can never be above 100%! Thus, in order to solve this typing error, we correct the percentage to 14.8%. We can also verify this correction by calculating the average value of the percentages of years 2013(preceding year) and 2015(succeeding year). Finally, to verify and to analyze the summary of the dataset, when we plotted a box-plot, we found that in the year 2014 for hard corals, there was an outlier around 148%.

With Error:

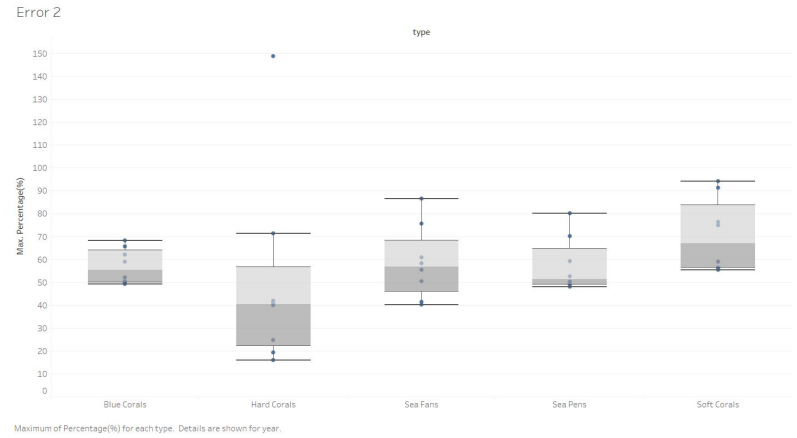

Figure 2.3 : Error 2 - Incorrect Percentage Value

Without Error:

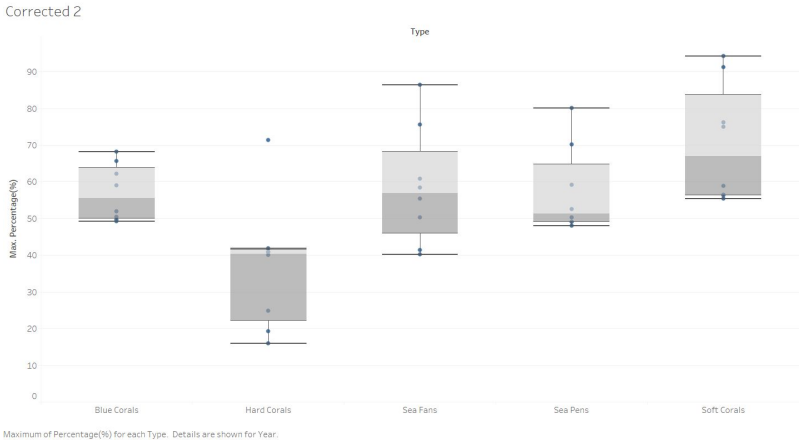

Figure 2.4 : Corrected Percentage Value

3. ANSWERING QUESTIONS

3.1 Question 1: In which years and for which kinds of coral is bleaching the worst?

Answer: Here, I allocated year and type to the columns and the percentage of bleaching to the rows. I have considered the **mean** value of bleaching as it gives a good, average percentage of bleaching of the different coral sites of a particular type.
Patterns and trends observed:
Clearly from the plot, we can see that in the years 2010 and 2012, Blue Corals have had the highest percentage of bleaching, nearly 34.91% and 37.55% respectively. In the years of 2011, 2013 and 2017, Sea Fans have had the maximum number of bleaching, around 35.73%, 36.67% and 62.83% respectively. Soft Corals have suffered the most bleaching in the years of 2014, 2015 and 2016. Around 43.62% of Soft Corals in 2014, 48.03% of Soft Corals in 2015 and 55.24% of Soft Corals in 2016 suffered the most bleaching levels respectively.

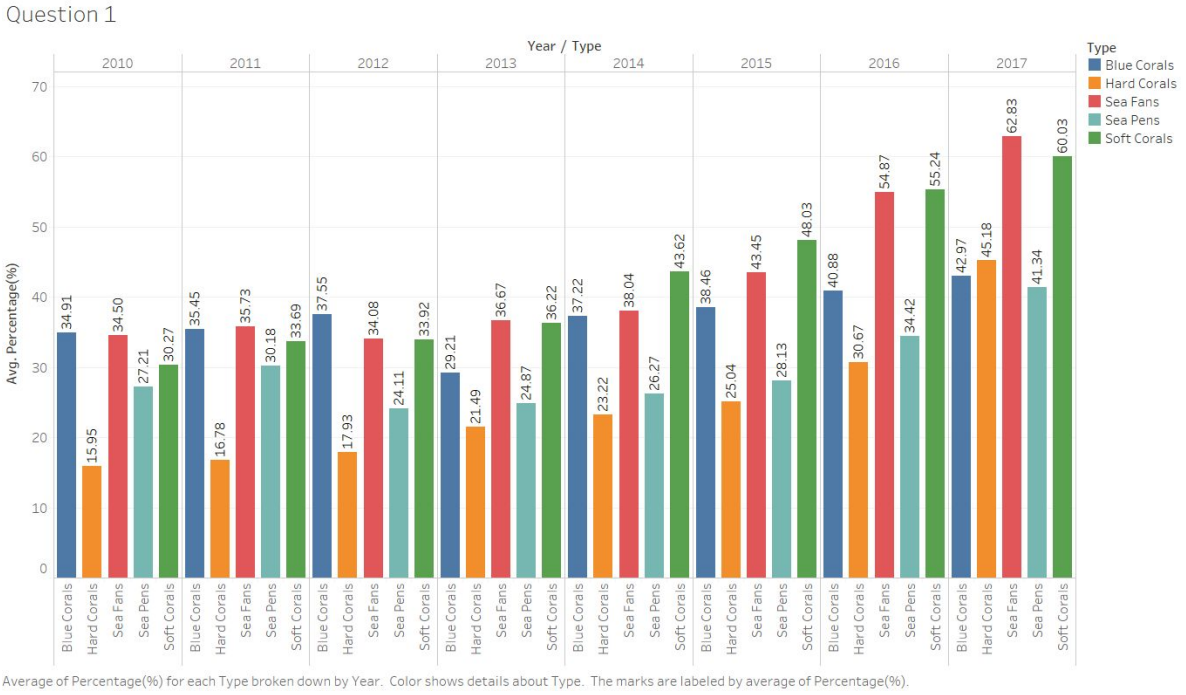

Figure 3.1 : Question 1 - Percentage of bleaching of different corals in different years

3.2 Question 2: How does the location of the site affect bleaching of the different kinds of corals?

Answer: Here, we plotted the different longitudes and latitudes of the different sites and named the data according to the different site names and average percentages of bleaching.
Patterns and trends observed:
Here, from the visualisation, it was observed that, as the coral sites moved closer towards the coast, the higher was their bleaching percentage. This is because the sites near the coast are more prone to low tides, pollution and high temperatures when compared to the

sites located further away from the coast. The least average bleaching was observed in site 3, around 27%, located farthest away from the coast compared to the other sites. Similarly, the bleaching percentages successively increased as the sites moved closer to the coast. Finally, the highest average bleaching percentage was observed in site 5, around 45% which happens to be closest to the coast.

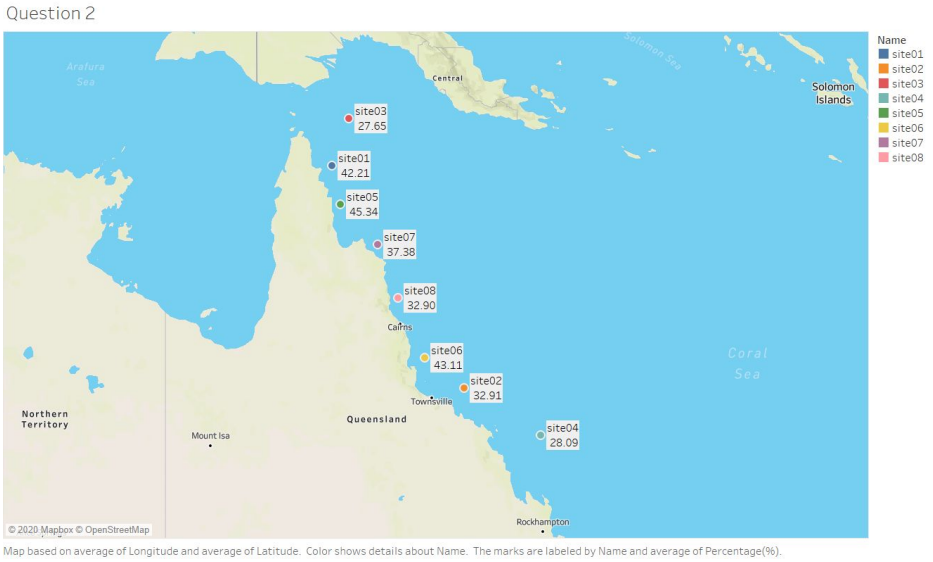

Figure 3.2 : Question 2 - Location of the different coral sites

The above **patterns and trends** are observable from the below visualisation too. Around 54.67% of Blue Corals, 35.38% of Hard Corals and 67.53% of Soft Corals experienced the most bleaching in site 1. In the same manner, around 58.56% of Sea Fans and 54.54% of Sea Pens found in site 6 underwent the most bleaching. The Blue Corals and Sea Pens in site 3 experienced no bleaching at all. Similarly, the Hard Corals found in sites 5 and 6 hardly experienced any bleaching at all. In the same way, the Sea Fans located in site 7 did not experience any bleaching at all. Finally, the Soft Corals found in site 6 hardly experienced any bleaching at all.

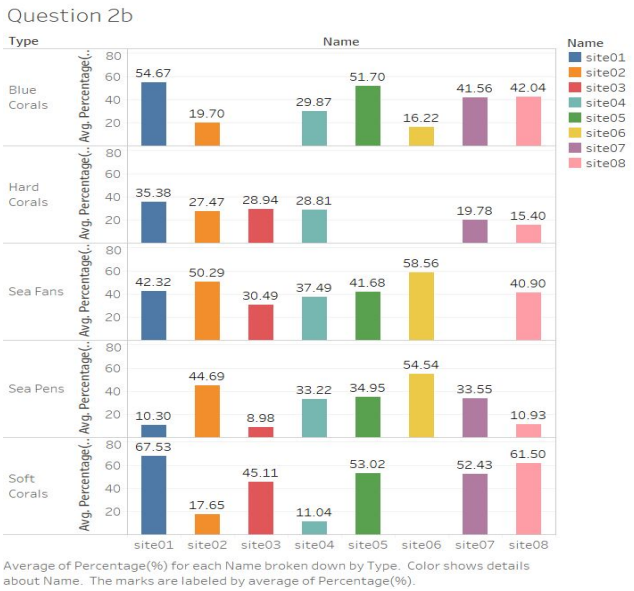

Figure 3.3 : % of bleaching of corals in various sites