FIT5147 – TABLEAU ASSIGNMENT

STUDENT NAME: MD. SAADMAN HOSSAIN

ACTIVITY NUMBER: 13

TUTOR NAME: MOHAMMAD HAQQANI

DATA WRANGLING (Tableau was used for this part):

STEPS:

1. Load the raw data into tableau.
2. Use tableau data interpreter (this step merges the headers of coral names and year, and also turns the percentages into proportions i.e. 50% to 0.50).
3. Use the pivot function to pivot the 5 coral names with year (e.g. soft coral 2017) columns.
4. Split the coral names and year column into 2 columns (Coral Names and Year).
   - RIGHT([coral names with year],4) -> for Year.
   - LEFT([coral names with year], (LEN([coral names with year])-LEN(Year))) -> for Coral Names.
5. Change column names appropriately.

Screen shot of data read into Tableau (only a part of the data shown as full data is 320 rows). This dataset still has a lot of errors which I am going to address below.

=Abc Calculation Coral Names	=Abc Calculation Year	# Pivot Proportion - Yearly...	Abc Sheet1 Site Name	Sheet1 Site Longitude	Sheet1 Site Latitude
soft corals	2010	0.56290	site01	143.51500	-11.8430
soft corals	2011	0.55430	site01	143.51500	-11.8430
soft corals	2012	0.56430	site01	143.51500	-11.8430
soft corals	2013	0.57700	site01	143.51500	-11.8430
soft corals	2014	0.74990	site01	143.51500	-11.8430
soft corals	2015	0.75340	site01	143.51500	-11.8430
soft corals	2016	0.80210	site01	143.51500	-11.8430
soft corals	2017	0.83870	site01	143.51500	-11.8430
sea fans	2010	0.37340	site01	143.51500	-11.8430
sea fans	2011	0.39120	site01	143.51500	-11.8430
sea fans	2012	0.38560	site01	143.51500	-11.8430
sea fans	2013	0.40880	site01	143.51500	-11.8430
sea fans	2014	0.41340	site01	143.51500	-11.8430
sea fans	2015	0.46870	site01	143.51500	-11.8430

Figure 1

DATA EXPLORATION (Done using Tableau):

Finding errors:

Data error 1: Site 02 data shows wrong Latitude (it is 18.94, should be -18.94); it seems to be a simple entry error as changing it to -18.94 makes more sense (site 02 location changes to a position in the Great Barrier Reef).

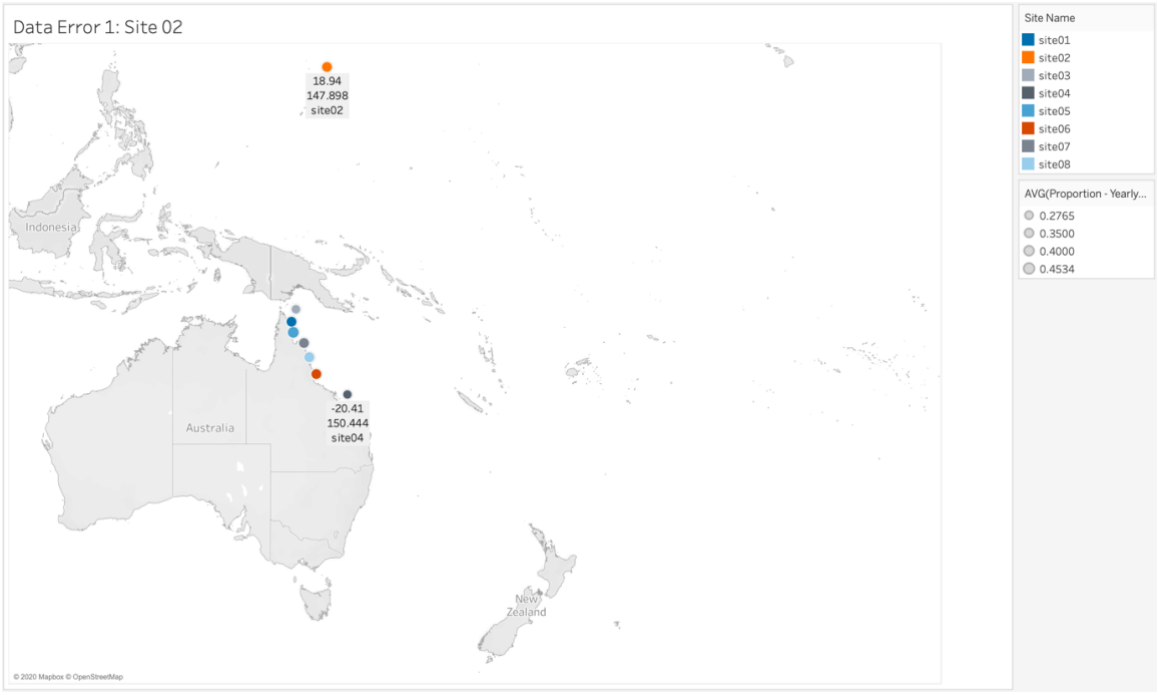

Figure 2

Data error 2: I was able to identify 2 outliers in the yearly coral bleaching proportions part of the data.

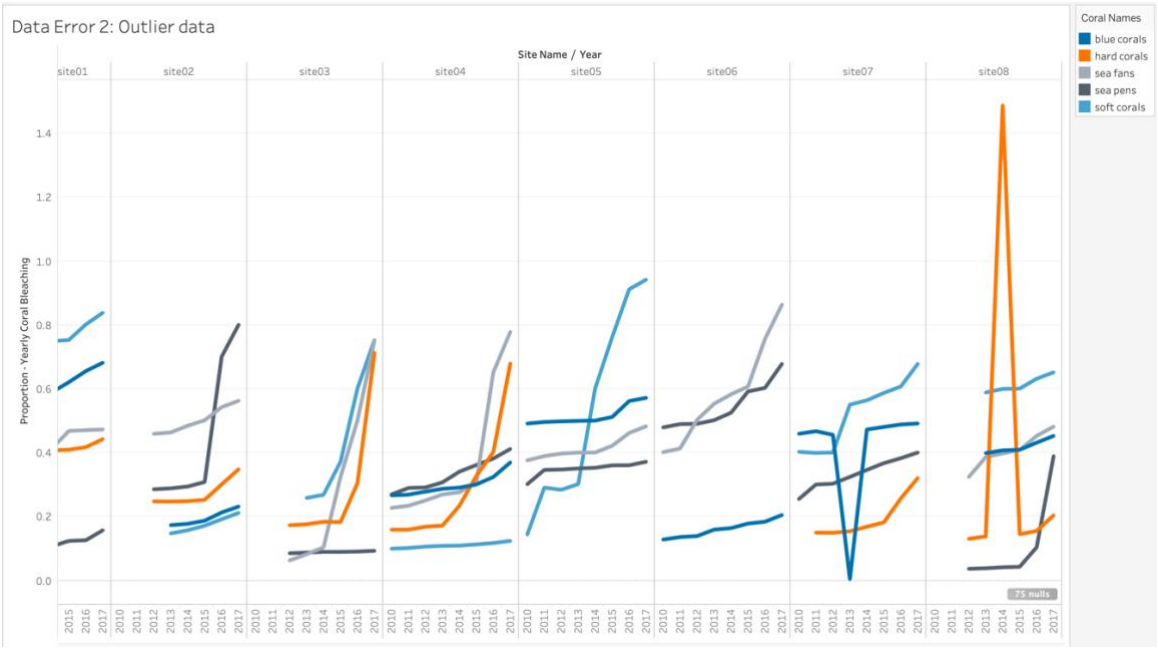

Figure 3

First outlier is observed when we look at the data for blue corals yearly bleaching proportion in site 07 for the year 2013 (its 0.0047, I suspect it should be 0.47 considering the trend we can observe from 2010 to 2012 and 2014-2017). This is a simple entry error and I fixed it accordingly.

Second outlier is observed from 2014 data of hard corals bleaching proportion at site 08. It is 1.488, but I suspect it should be 0.1488 (percentage of bleaching exceeding 100 percent does not make sense). This is also an entry error considering the trend we can observe from 2010 to 2013 and 2015-2017.

Data Error 3: Null/Missing values: 75 null values are found from the proportion yearly coral bleaching column. Null values affect the integrity of the analysis. **After conducting Little's MCAR test using R on a subset of the dataset (coral bleaching with year, site name, and proportion of bleaching per year), I found that the P-value**

is 0.01312238 which is less than 0.05. This indicates that the data is not missing completely at random (so either it is MAR or NMAR).

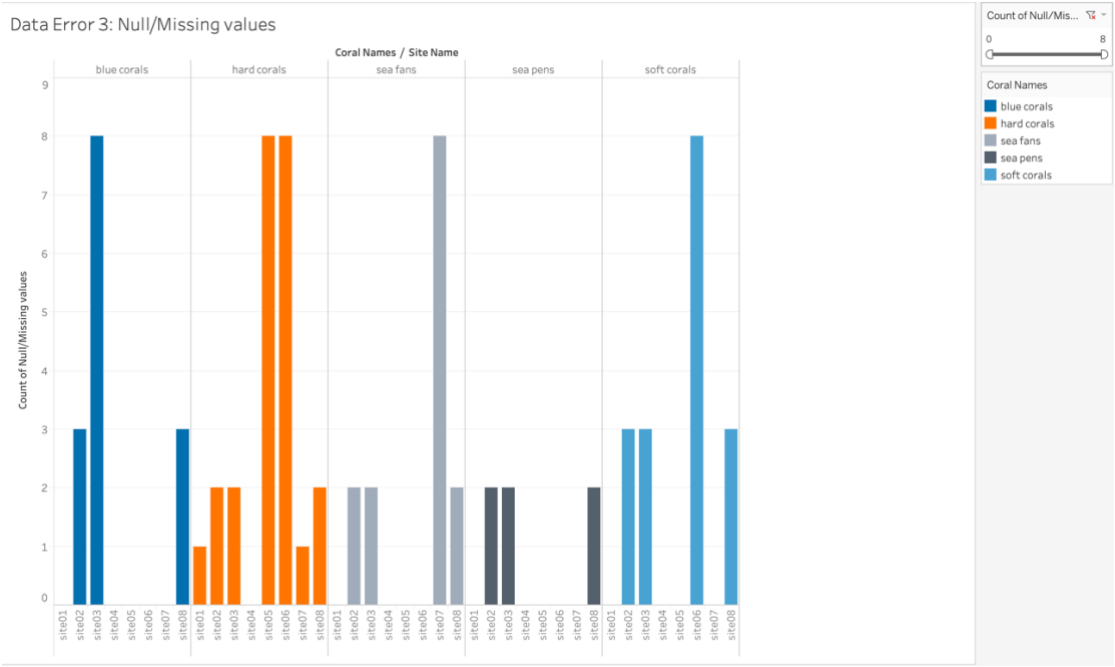

Figure 4

We can observe from the figure above that the null values have a certain pattern. There are entire data missing for all the year for a few of the corals from particular sites (site03 missing blue coral data, site05 and site06 missing hard coral data etc.). Missing values/nulls which are consistently missing may indicate that particular sites do not inhabit certain species of corals. I chose to enter these missing values as zeroes for my analysis. We can also see that there is some data missing for earlier years for some corals and sites. This might occur due to some sites starting to conduct data collection later than others (or just might be errors in data entry). As I have noted before, these values are not missing completely at random. I have created a data frame In R with data for these cases (sites and coral data by year) and used linear regression to find best fitting line for each case on the dataset to predict the null values. The resulting dataset was merged with the wrangled data set for further analysis.

Abc Calculation Coral Names	Abc Calculation Year	# Pivot Proportion of Yearl...	Abc Sheet1 Site Name	Sheet1 Site Longitude	Sheet1 Site Latitude
blue corals	2010	0.397800	site01	143.51500	-11.8430
blue corals	2011	0.402500	site01	143.51500	-11.8430
blue corals	2012	0.503700	site01	143.51500	-11.8430
blue corals	2013	0.519800	site01	143.51500	-11.8430
blue corals	2014	0.589800	site01	143.51500	-11.8430
blue corals	2015	0.621300	site01	143.51500	-11.8430
blue corals	2016	0.656100	site01	143.51500	-11.8430
blue corals	2017	0.682300	site01	143.51500	-11.8430
hard corals	2010	0.184330	site01	143.51500	-11.8430
hard corals	2011	0.193400	site01	143.51500	-11.8430
hard corals	2012	0.204500	site01	143.51500	-11.8430
hard corals	2013	0.399800	site01	143.51500	-11.8430
hard corals	2014	0.407800	site01	143.51500	-11.8430
hard corals	2015	0.409900	site01	143.51500	-11.8430

Figure 5

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

Coral bleaching has been getting worse over the years in the great barrier reef. We can observe that there is a growing trend each year for every coral species. It is clear from the visualisation that sea fans followed by soft corals in 2017 are getting affected the most from the bleaching.

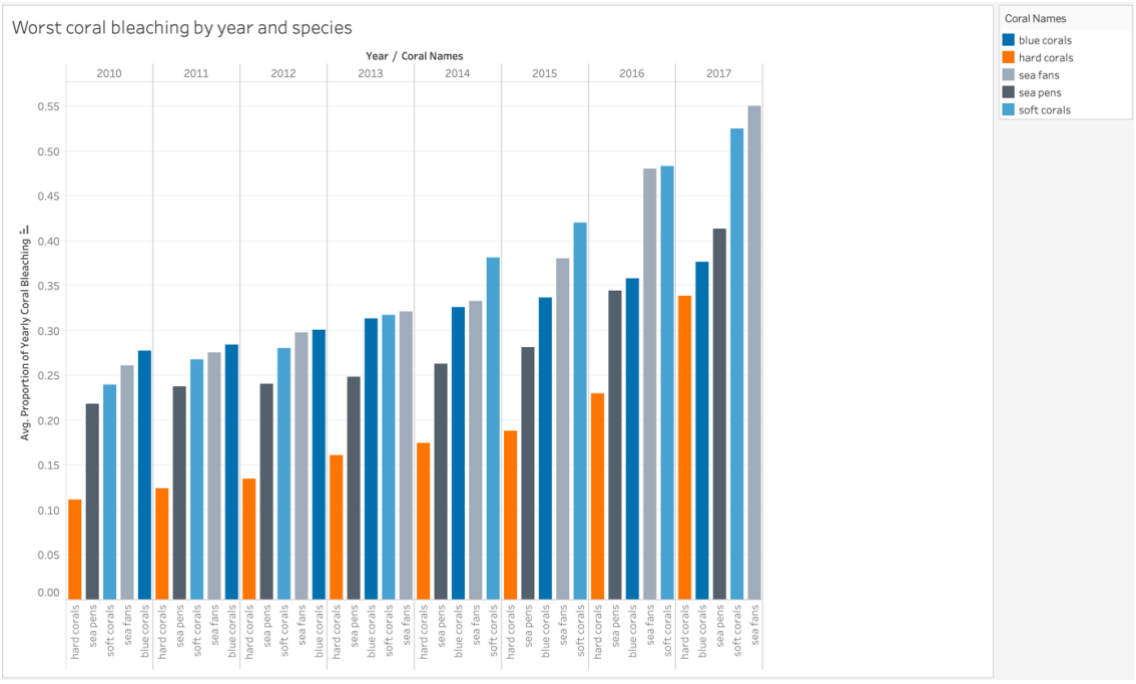

Figure 6

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

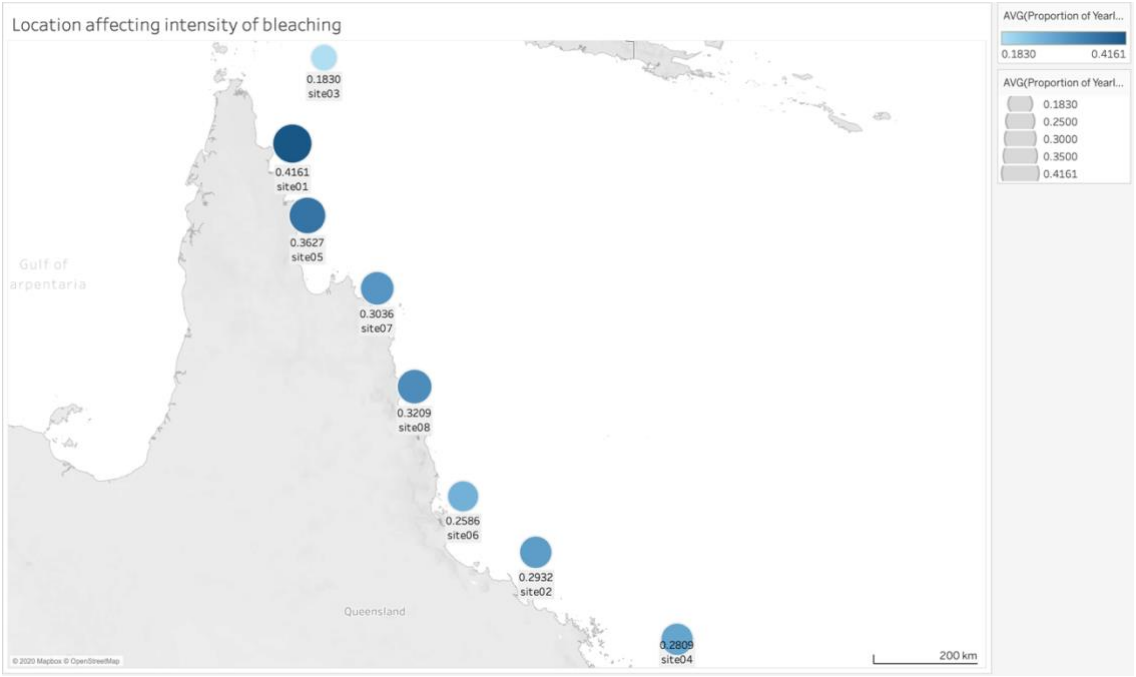

Figure 7

From the visualisation above we can see that northern most sites are more affected by coral bleaching than the sites which are more towards the south. This is due to higher water temperatures in the northern oceans. Water temperatures tend to rise when approaching the equator and fall approaching the poles.

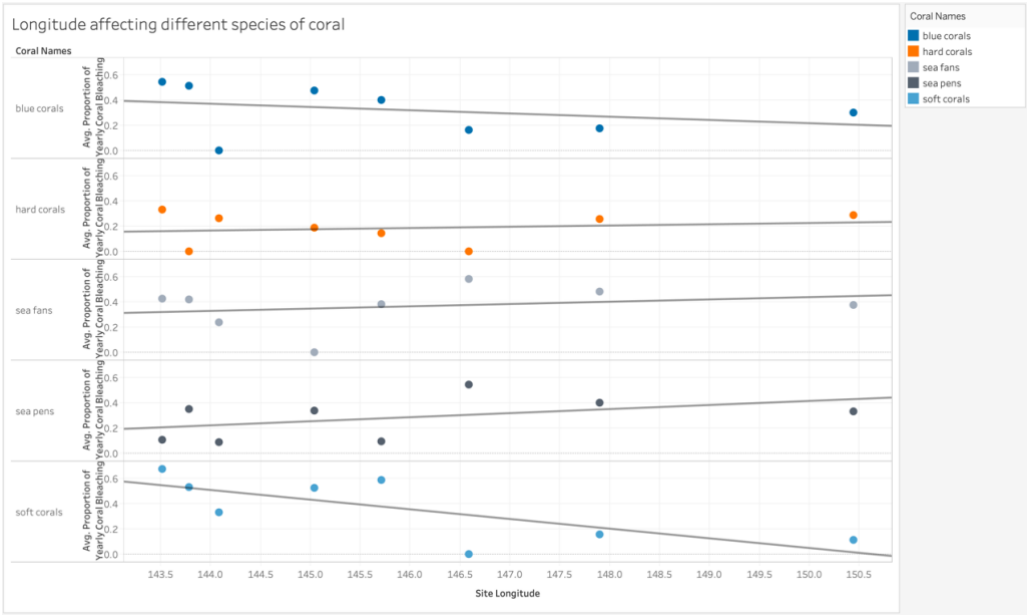

Figure 8

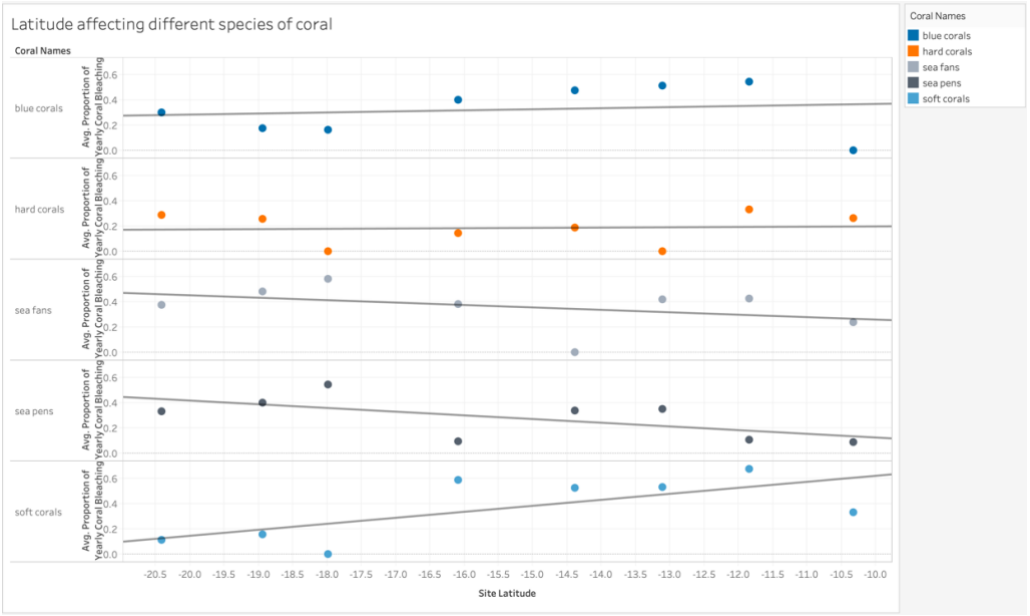

Figure 9

From the 2 figures above we can observe the following trends:

1) Blue corals and soft corals tend to bleach less on average as longitude increases. On the other hand, sea fans and sea pens tend to bleach more as longitude increases. Changes in hard coral bleaching levels remain negligible.
2) Blue corals and soft corals tend to bleach more as latitude increases, whereas sea fans and sea pens tend to bleach less on average as latitude increases. Changes in hard coral average bleaching remains negligible across different latitude levels.
3) Analysing the pattern, I am able to hypothesize that blue corals and soft corals bleach less on average as longitude increases and latitude decreases. This indicates that blue and soft corals thrive in areas where the water temperature is lower (further away from the equator). Also, sea pens and sea fans bleach more on average as longitude increases and latitude decreases. This indicates that sea pens and sea fans are more prone to bleaching in relatively lower water temperature. Hard corals do not seem to be affected by their location, as I wasn't able to find any significant trend.

Even a 2-degree Celsius rise in water temperature can cause coral bleaching. Hopefully we can start paying attention to global issues such as global warming which is affecting our planet's marine life greatly.