Title: Programming Exercise 1
Student: Mudit Sharma (30340071)
Tutor: Fahimeh Sadat Saleh (Section 19)

1 - DATA WRANGLING

For this task, I exclusively used Tableau's built-in functions to manipulate the data into a format that would enable informative visualisation. There were a few issues to overcome, which are listed as follows:

1.1 - COLUMN NAMES

This data did not contain informative column headings, and the way that each coral included sub-headings for year especially did not enable it to be easily used in Tableau. Simply reading the data using Python, R, or Tableau would cause it to interpret the years as data points rather than sub-headings for the respective column. Luckily, Tableau includes a feature called Data Interpreter, which quickly interprets and cleans up messy data. Simply clicking the checkbox resulted in a data format with properly named columns, even if they are not yet suitable for the data visualisation we are trying to perform. Tableau has detected the column sub-headings, and transformed them into a simplified <coral name> <year> format for each coral and year, as exemplified with *soft corals 2017*, *soft corals 2016*, and so on. It has even detected latitude and longitude, and converted those columns into the appropriate geographical data types.

1.2 - PIVOT TABLES

In order to aggregate data and answer questions such as which year saw the most bleaching, it is necessary to pivot the dependent variables (*soft corals 2017*, *soft corals 2016* etc.) into rows. However, before pivoting these columns, we need to make sure that post-pivot, column values such as *soft corals 2017*, and *soft corals 2016* can be resolved into separate columns for coral, and year. To enable such a split post-pivot, we reformat <coral name> <year> to <coral name>**,** <year>.

Calculation Year	Calculation Coral	Pivot Bleaching	Sheet1 name	Sheet1 Longitude	Calculation Latitude
2010	blue corals	0.26710	site04	150.44400	-20.4140
2011	blue corals	0.26920	site04	150.44400	-20.4140
2012	blue corals	0.27870	site04	150.44400	-20.4140
2013	blue corals	0.28760	site04	150.44400	-20.4140
2014	blue corals	0.29120	site04	150.44400	-20.4140

Figure 1

After renaming the relevant columns, we select every column but name, latitude, and longitude, and pivot. Then, using a Custom Split along the comma we just inserted into <coral name>**,** <year> we can derive 2 separate columns, one for <coral name>, and another for <year>. This data format now enables us to aggregate by coral, and/or by year. At this point, we can rename our newly generated columns to Coral and Year respectively, and hide the parent column used to generate the two, bringing the data into a usable data format as shown in **Figure 1**:

2 - DATA CLEANING

2.1 - [BLEACHING]

We can see that the Bleaching column contains a number of null values. For this particular assignment, this is not likely to change our outputs since we are almost exclusively aggregating by summing up values by year, site, and coral. However, to ensure that we are able to accurately assess means if and when the need arises, it makes sense to impute this with a value of 0.

Pivot Bleaching FALSE
1.48800

Figure 2

Another issue arises in Bleaching when a value appears to be incorrectly entered (**Figure 2**). Since the raw data contains bleaching measurements as percentages, which Tableau has converted into decimal representations. However, there is a value of 1.48800, which could only arise from 148.8%, which would be impossible. In order to correct this, we must note that each correct Bleaching measurement is accurate to 4 decimal places, whereas 1.48800 is only accurate to 3 decimal places. Every other Bleaching measurement for hard corals at site08 lies between 0.131, and 0.205 as deduced by filtering based on site08, and hard corals. Based on this, and the fact that 1.488 is only accurate to 3 decimal places, we can deduce that 1.488 is actually supposed to be 0.1488. The provided code is run to handle both issues in one go. It imputes nulls with 0, and replaces 1.48800 with 0.1488:

```
CASE[Bleaching NULL]
WHEN NULL THEN 0
WHEN 1.48800 THEN 0.1488
ELSE [Bleaching NULL]
END
```

2.2 - [LATITUDE]

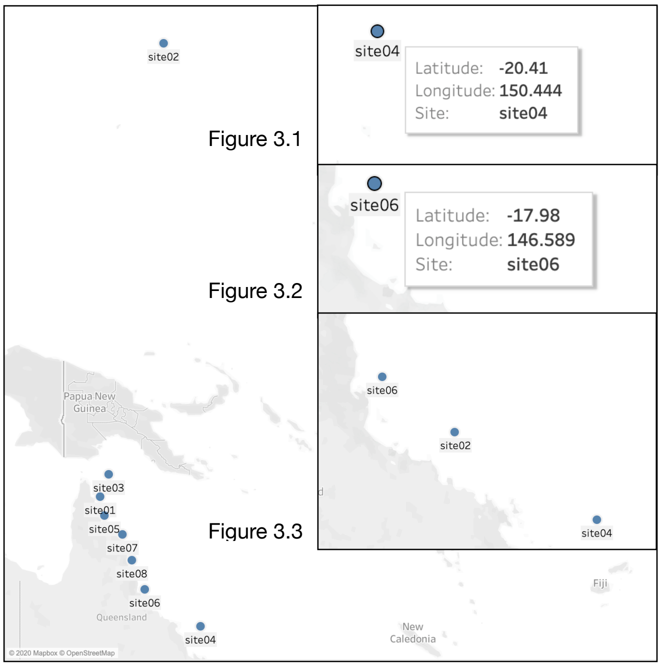


Figure 3

While sorting through the Latitude column to check for extreme or missing values, we can also notice that site02 has a very different latitude than the other sites at 18.937, which may indicate an error. Every other site has a latitude between -10.321 and -20.414, with site02 being a notable outlier. From the data alone, it is quite difficult to decide whether or not this is intentional, so we can plot the latitude against the longitude to see where these sites lie (**Figure 2**), which would make it easier to identify a trend (if any).

Based on the locations of the other sites, we can deduce that this data is about the Great Barrier Reef. The proximity of the other sites means that site02 is not where it needs to be. All the other sites are roughly equidistant, but this does not hold for site04 and site06 in the bottom right where there seems to be a larger gap between them. We can also see that site06 has a latitude of -17.98, and site04 has a latitude of -20.41 (**Figure 2.1, 2.2**), which means that site02's true latitude is likely -18.937, and that it is meant to lie between site04 and site06. This is confirmed in **Figure 2.3**, taken after the code below has been implemented to fix the error.

```
IF [Latitude DIRTY] = 18.9370
THEN -18.9370
ELSE [Latitude DIRTY]
END
```

3 - DATA VISUALISATION

3.1 - IN WHICH YEARS AND FOR WHICH KINDS OF CORAL IS BLEACHING THE WORST?

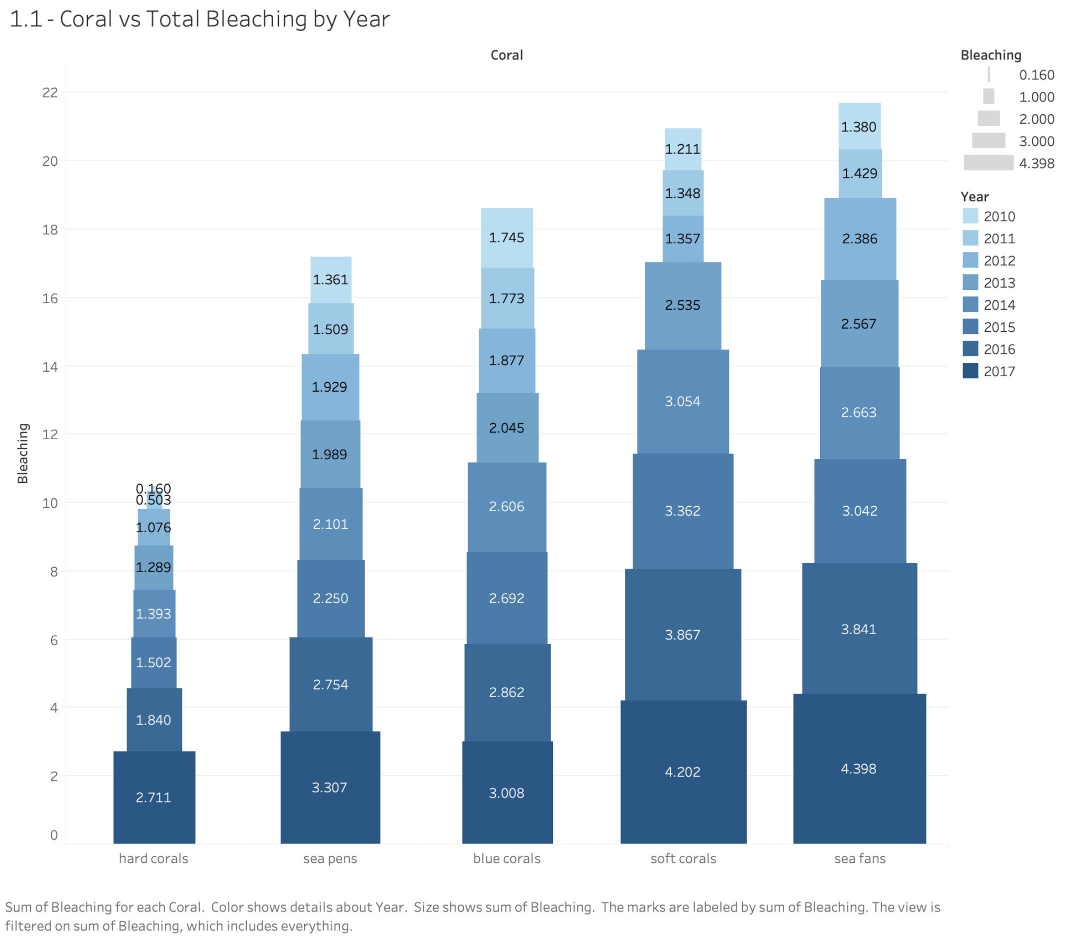


Figure 4

Using **Figure 3**, we can note that every species of coral saw its worst bleaching events in 2017, and also that sea fans are the species of coral that have undergone the most bleaching (overall, *and* in 2017). This graph makes it easy to note that coral bleaching is only increasing with time, and that every species is suffering from coral bleaching.

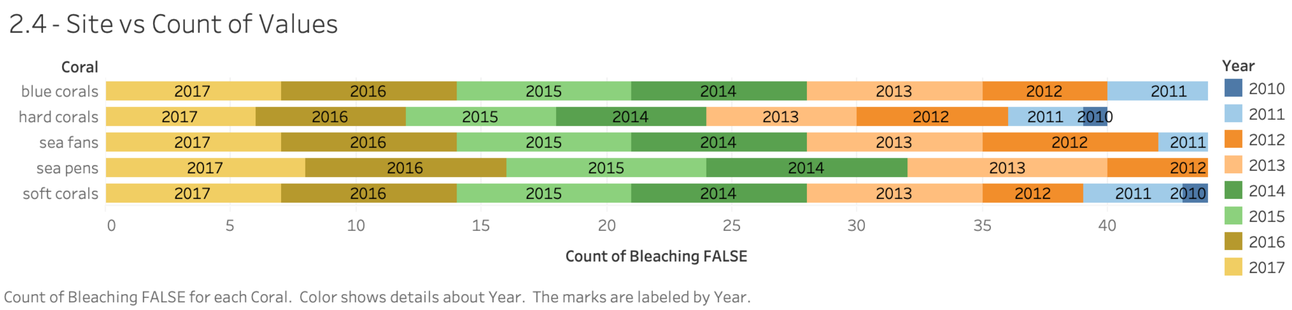


Figure 5

However, **Figure 5** shows us that the raw data contains varying amounts of measurements for each type of coral, and some years Barely have any readings. For example, 3 out of 5 types of corals have 0 readings in 2010, and 2011 also sees very few readings as compared to later years like 2017. This tells us that the data is not reliable enough to give us an accurate idea of what actually happened in these years, but is likely enough to still tell us that the situation continues to worsen. The pace at which the situation is worsening, and whether coral bleaching is slowing down or not may also not be reliably deduced.

3.2 - HOW DOES THE LOCATION OF THE SITE AFFECT BLEACHING OF THE DIFFERENT KINDS OF CORALS?

Site vs Total Bleaching by Coral (Highlight Table)

Coral	site01	site02	site03	site04	site05	site06	site07	site08
soft corals	5.40	0.88	2.26	0.88	4.24	0.00	4.19	3.07
blue corals	4.37	0.98	0.00	2.39	4.14	1.30	3.32	2.10
sea fans	3.39	3.02	1.83	3.00	3.33	4.68	0.00	2.45
hard corals	2.48	1.65	1.74	2.30	0.00	0.00	1.38	0.92
sea pens	0.82	2.68	0.54	2.66	2.80	4.36	2.68	0.66
Grand Total	16.46	9.21	6.36	11.23	14.51	10.35	11.59	9.21

Bleaching: 0.000 – 5.403

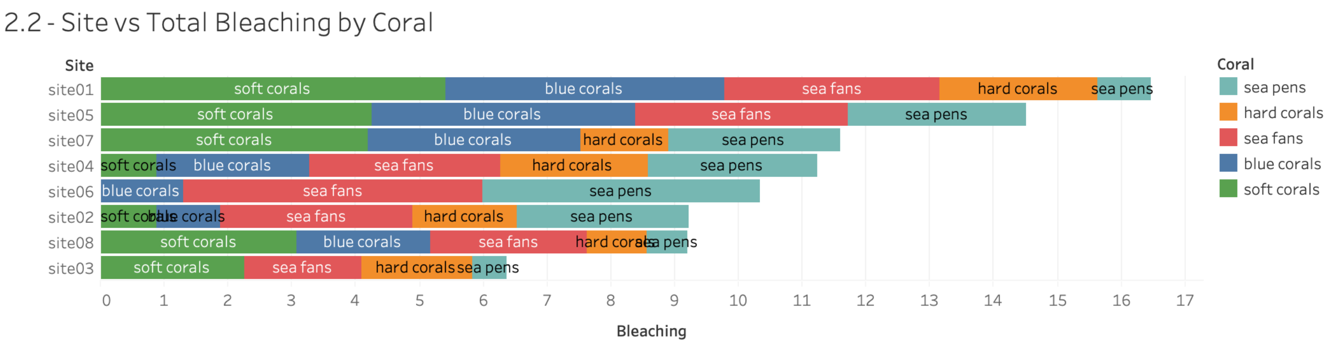
Sum of Bleaching broken down by Site vs. Coral. Color shows sum of Bleaching. The marks are labeled by sum of Bleaching.

Figure 6

Figure 6 allows us to observe how different species of coral fare at different sites, allowing us to compare, for example, how soft corals fared at site01 as compared to site04. The included grand totals also enable us to determine which sites saw the worst coral bleaching, and which are not as affected.

While we may be quick to determine that site03 is doing great in comparison, we have already seen in **Figure 3** that the sites are arranged roughly north to south in no particular order. Site03 is followed by site01, which precedes site05 and so on. This makes the data more perplexing, however, since site03 is seemingly doing great compared to adjacent site01 and site05, which are seeing the highest and second highest grand totals out of all our sites. The answer to this lies in the null values which we imputed with 0 during Data Cleaning. Site03 does not have any data on blue corals, for example, which was significantly affected at site01 and site05, making it unlikely that it was unaffected at site03.

We can easily visualise how many non-null data points actually exist for each set in the form of a simple bar chart (**Figure 7**):



Sum of Bleaching for each Site. Color shows details about Coral. The marks are labeled by Coral. The view is filtered on Exclusions (Coral,name), which keeps 35 members.

Figure 7

Here, we can see that indeed, site03 has the lowest number of actual data points, which would explain the discrepancy discussed above. It is also possible that not all species of coral exist at each of these sites, but that seems unlikely given their proximity. In conclusion, the data is not very dependable for the task at hand, and cannot be reliably used to detect what is actually happening on a site-by-site basis.