Data Exploration & Visualisation

Coral Bleaching

Student Name: Ajay Ganapathy

Student Id: 29822270

Contents

1. Exploratory data analysis and visualisation for different types of coral bleaching

Data Exploration and visualisation starts with Data Wrangling as the first step.

Referring to the notion of how raw data is processed to form meaningful information, Data Wrangling refers to how you play with the raw data i.e. cleaning the data, re-formatting the data, analyzing bad data, handling them with possible logical predictions, thereby achieving clean datasets.

Data Visualisation is basically creating a visual representation of the cleaned data in form of charts/graphs to present it to the business users. Visual representations are much easier to understand than textual representations.

2. Data Wrangling

Consider the example of coral bleaching data. This dataset is an Excel file which contains raw data about different kinds of corals, the different sites where coral bleaches and the bleaching percentage of different corals over the period of 8 years (2010-2017) at those sites. This dataset has 10 rows and 43 columns in total (including the headers and values). This dataset is loaded into Tableau Public in which the raw data is transformed into a suitable structure and visualized in form of different charts/graphs to achieve correct inferences. Below is a sample screenshot of initial Excel dataset.

			soft corals											sea
name	longitude	latitude	2017	2016	2015	2014	2013	2012	2011	2010	2017	2016	2015	2014
site01	143.515	-11.843	83.870%	80.210%	75.340%	74.990%	57.700%	56.430%	55.430%	56.290%	47.320%	47.120%	46.870%	41.340%
site02	147.898	18.937	21.230%	19.230%	17.210%	15.780%	14.800%				56.320%	54.320%	50.210%	48.500%
site03	144.081	-10.321	75.340%	60.230%	37.210%	26.890%	25.890%				75.230%	50.210%	32.450%	10.450%
site04	150.444	-20.414	12.450%	11.780%	11.340%	10.980%	10.890%	10.670%	10.230%	10.010%	77.890%	65.230%	30.780%	27.670%
site05	143.786	-13.107	94.230%	91.230%	76.230%	60.230%	30.230%	28.450%	29.130%	14.450%	48.340%	46.230%	42.230%	40.120%
site06	146.589	-17.981									86.450%	75.640%	60.750%	58.340%
site07	145.043	-14.383	67.890%	60.780%	58.760%	56.450%	55.120%	40.120%	39.980%	40.340%				
site08	145.715	-16.091	65.230%	63.210%	60.120%	60.030%	58.890%				48.230%	45.320%	40.890%	39.840%

2.1. Incorrect Data

This dataset contains some incorrect data which are handled as follows:

1. Since, not all the sites have data for entire period, and not all corals are found at each site, there are few **missing or blank values**. These missing values are ignored since Tableau automatically detects them and replaces it with Null values which can later be excluded while visualizing it in different forms.

2. The latitude of site 2 is 18.397 in the initial dataset. If we carefully see the data of latitude and longitude, all the longitude values are positive. **All the latitude values are negative except for site 2**.

name	longitude	latitude
site01	143.515	-11.843
site02	147.898	18.937
site03	144.081	-10.321
site04	150.444	-20.414
site05	143.786	-13.107
site06	146.589	-17.981
site07	145.043	-14.383
site08	145.715	-16.091

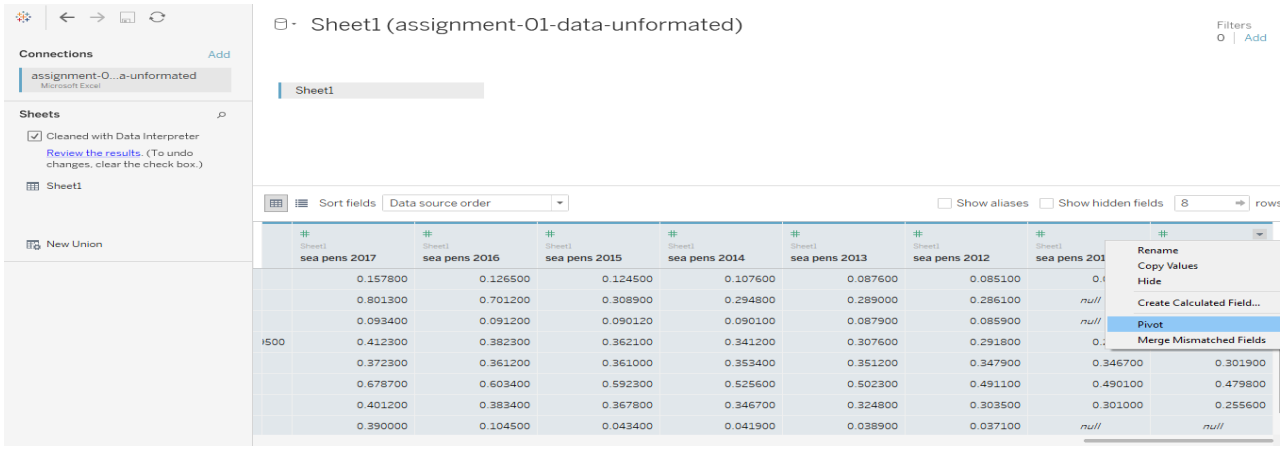
So, it can be predicted that it is an incorrect data (**also considered as an outlier**). This incorrect data is handled by manually replacing it with -18.397.

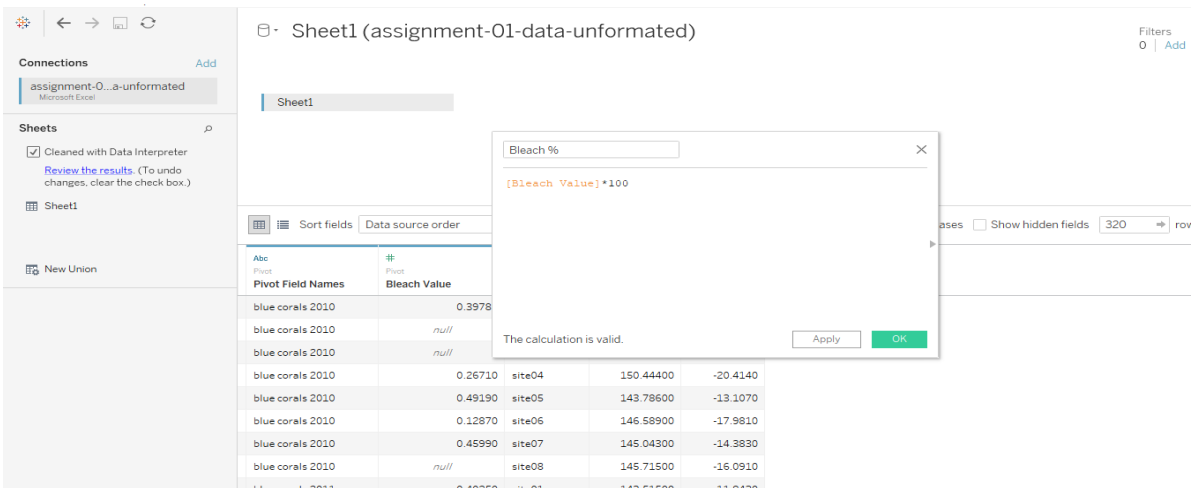
3. There are two **bleaching % values which are either too high or too low**. One at site 8 which is 148.800 % for hard corals in the year 2014 and other at site 7 which is 0.470% for blue corals in the year 2013.

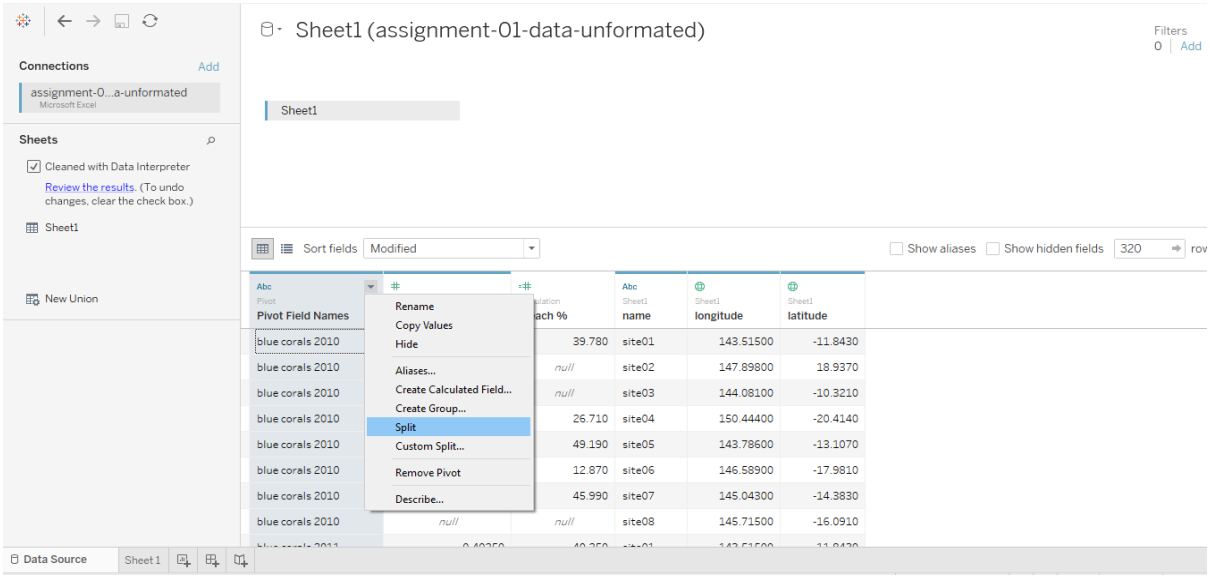
hard corals			blue corals		
2015	2014	2013	2014	2013	2012
40.990%	40.780%	39.980%	58.980%	51.980%	50.370%
25.320%	24.890%	24.770%	17.790%	17.370%	
18.340%	18.440%	17.650%			
32.800%	23.430%	17.230%	29.120%	28.760%	27.870%
			50.120%	50.020%	49.870%
			16.450%	15.970%	13.950%
18.230%	16.890%	15.450%	47.320%	0.470%	45.670%
14.560%	148.800%	13.840%	40.780%	39.930%	

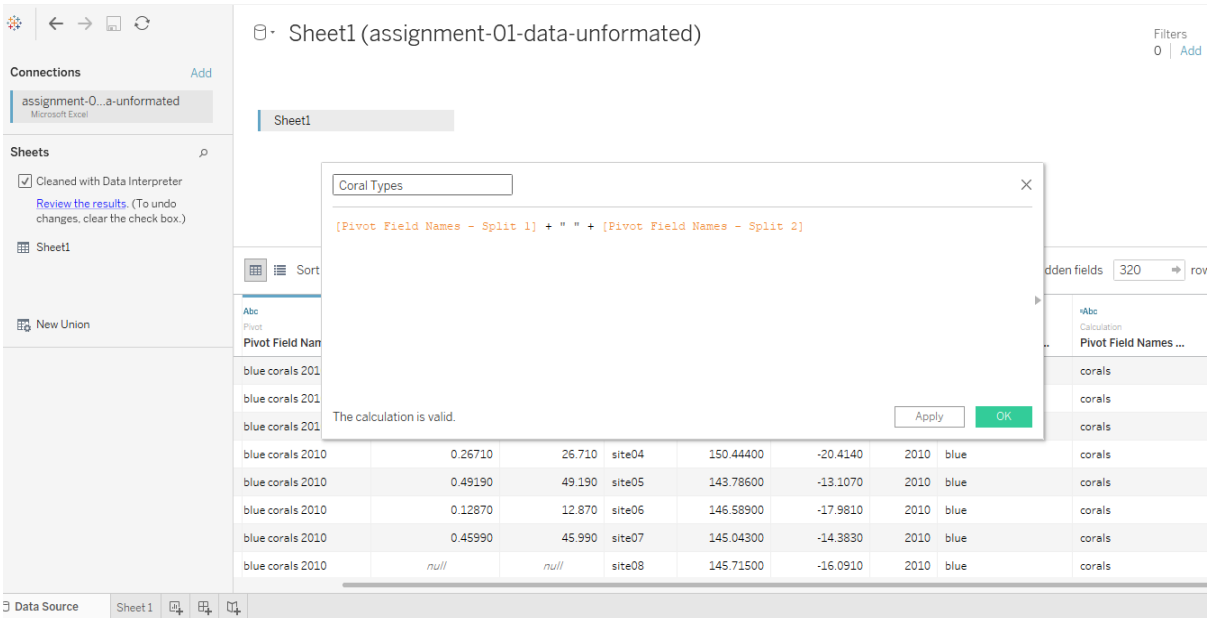
If we carefully see the bleaching % of its previous and the next immediate years, we can predict that this is a data entry error (**also, considered as outliers since they deviate from its previous and next values**). This is handled by shifting the decimal places by one to the left/right to make it a correct value. So, the new values are 14.88% and 47.00%.

These incorrect data are cleaned manually and the dataset is loaded into Tableau Public. In Tableau, the dataset is transposed and modified into a proper structure which is then used for visualisation. Below are some sample screenshots of the data transformation.

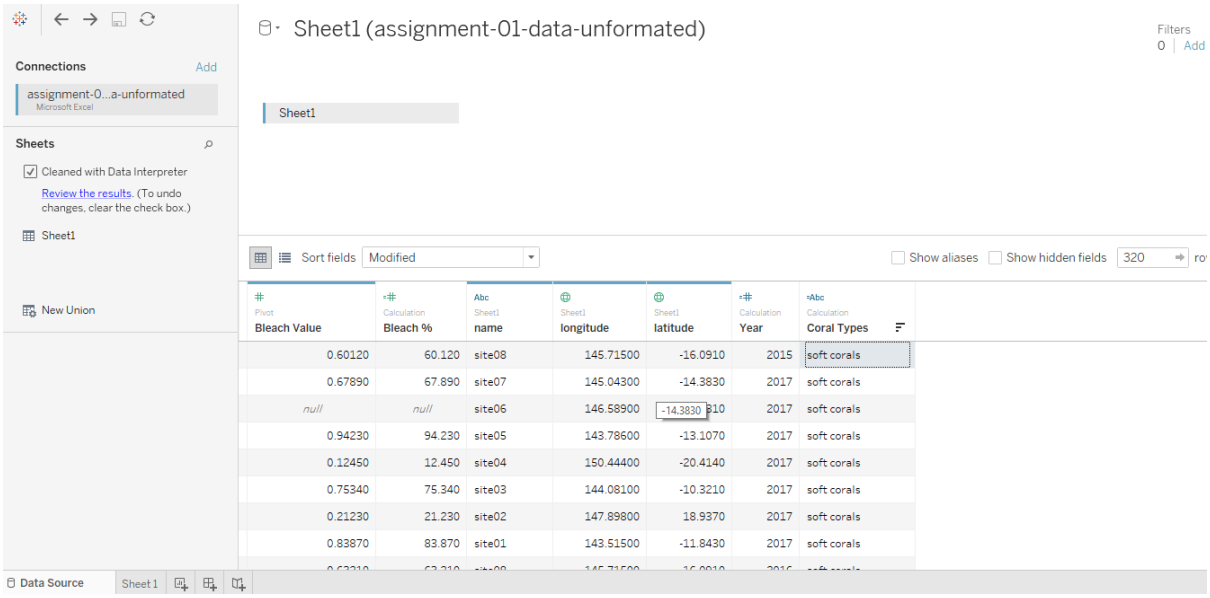

Sheet1 (assignment-01-data-unformated)
Connections
Add
assignment-0...a-unformated
Microsoft Excel
Sheets
Cleaned with Data Interpreter
Review the results. (To undo changes, clear the check box.)
Sheet1
New Union
Filters
0 | Add
Sheet1
Sort fields
Data source order
Show aliases
Show hidden fields
8
rows
sea pens 2017
sea pens 2016
sea pens 2015
sea pens 2014
sea pens 2013
sea pens 2012
Rename
Copy Values
Hide
Create Calculated Field...
Pivot
Merge Mismatched Fields
0.157800 0.126500 0.124500 0.107600 0.087600 0.085100
0.801300 0.701200 0.308900 0.294800 0.289000 0.286100 null
0.093400 0.091200 0.090120 0.090100 0.087900 0.085900 null
0.412300 0.382300 0.362100 0.341200 0.307600 0.291800
0.372300 0.361200 0.361000 0.353400 0.351200 0.347900 0.346700 0.301900
0.678700 0.603400 0.592300 0.525600 0.502300 0.491100 0.490100 0.479800
0.401200 0.383400 0.367800 0.346700 0.324800 0.303500 0.301000 0.255600
0.390000 0.104500 0.043400 0.041900 0.038900 0.037100 null null


Sheet1 (assignment-01-data-unformated)
Connections
Add
assignment-0...a-unformated
Microsoft Excel
Sheets
Cleaned with Data Interpreter
Review the results. (To undo changes, clear the check box.)
Sheet1
New Union
Filters
0 | Add
Sheet1
Bleach %
[Bleach Value]*100
The calculation is valid.
Apply
OK
Sort fields
Data source order
Show hidden fields
320
Pivot Field Names
Bleach Value
blue corals 2010 0.3978
blue corals 2010 null
blue corals 2010 null
blue corals 2010 0.26710 site04 150.44400 -20.4140
blue corals 2010 0.49190 site05 143.78600 -13.1070
blue corals 2010 0.12870 site06 146.58900 -17.9810
blue corals 2010 0.45990 site07 145.04300 -14.3830
blue corals 2010 null site08 145.71500 -16.0910


Sheet1 (assignment-01-data-unformated)
Connections
Add
assignment-0...a-unformated
Microsoft Excel
Sheets
Cleaned with Data Interpreter
Review the results. (To undo changes, clear the check box.)
Sheet1
New Union
Filters
0 | Add
Sheet1
Sort fields
Modified
Show aliases
Show hidden fields
320
Pivot Field Names
Rename
Copy Values
Hide
Aliases...
Create Calculated Field...
Create Group...
Split
Custom Split...
Remove Pivot
Describe...
name
longitude
latitude
blue corals 2010 39.780 site01 143.51500 -11.8430
blue corals 2010 null site02 147.89800 18.9370
blue corals 2010 null site03 144.08100 -10.3210
blue corals 2010 26.710 site04 150.44400 -20.4140
blue corals 2010 49.190 site05 143.78600 -13.1070
blue corals 2010 12.870 site06 146.58900 -17.9810
blue corals 2010 45.990 site07 145.04300 -14.3830
blue corals 2010 null null site08 145.71500 -16.0910
Data Source
Sheet 1

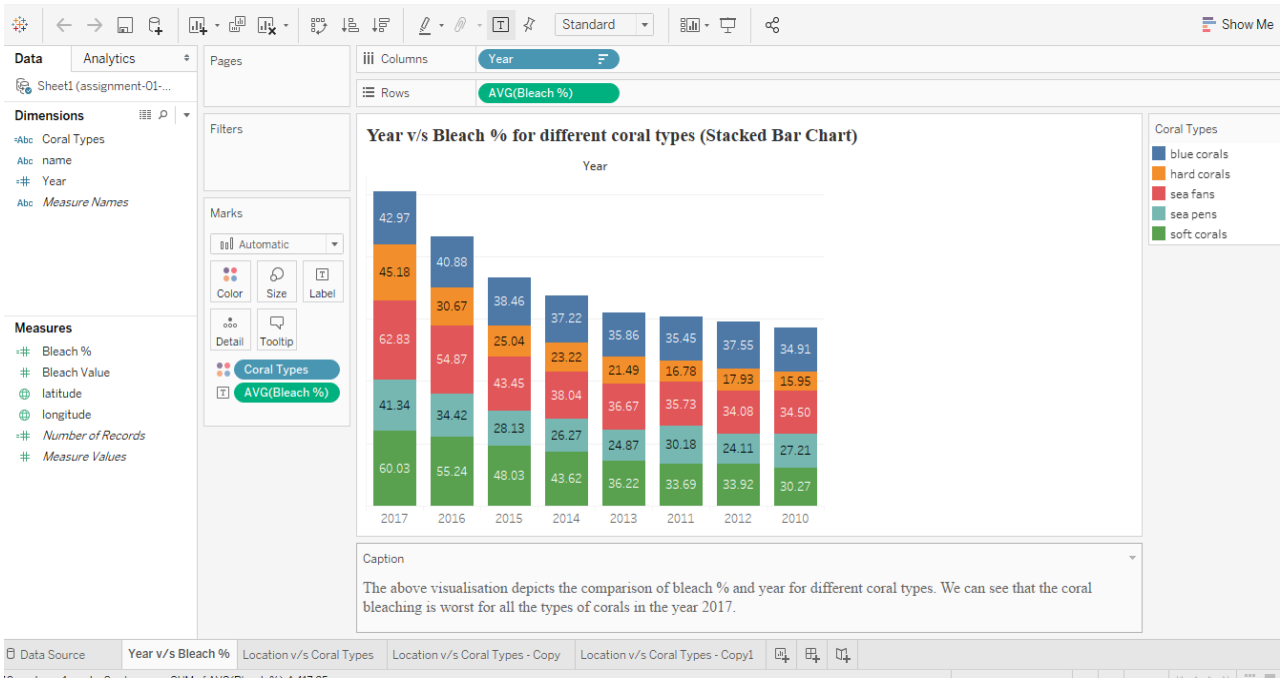


The below screenshot depicts the final cleaned dataset which is then used for visualizing the coral bleaching information in different charts/graphs.

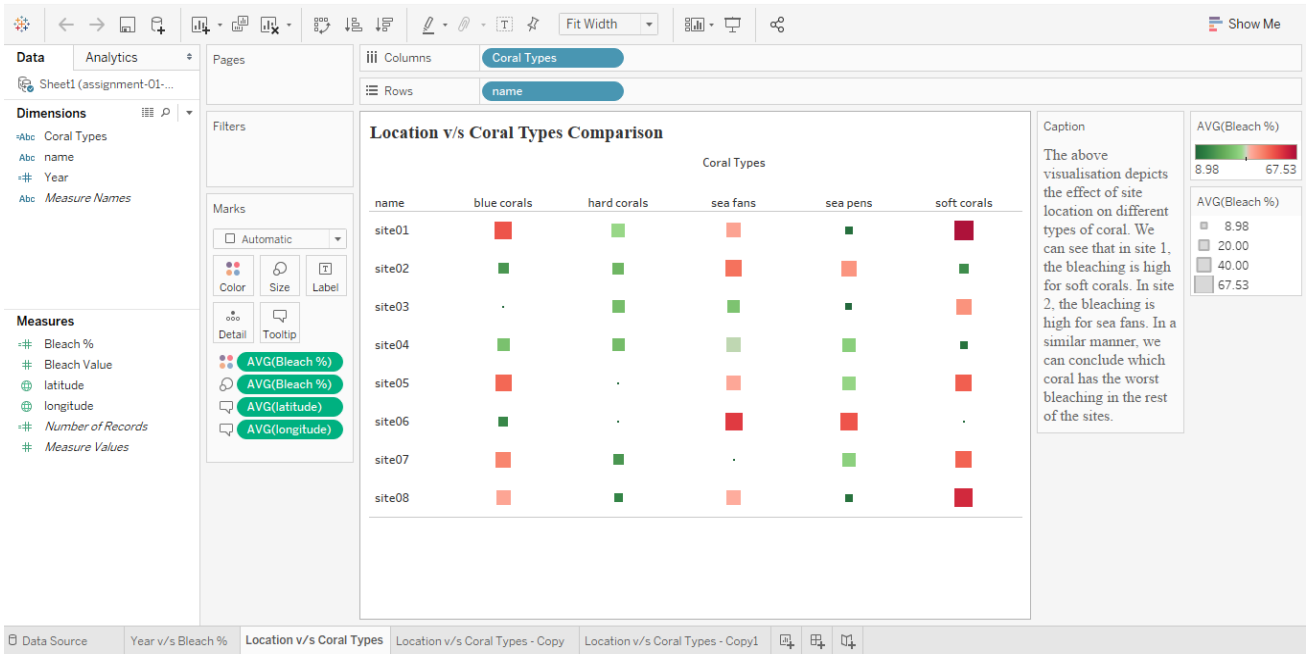


3. Data Visualisation

In the coral bleaching example, the final cleaned dataset is then visualized in Tableau Public. Cleaned data can be visualized in several forms such as bar charts, horizontal bar charts, stacked bar charts, area chart, pie chart, etc.



The above visualisation depicts a stacked bar chart which shows the average bleaching percentage for different kinds of corals over the period 2010 to 2017. **It can be seen that for all the types of corals such as hard, soft, blue, sea fans, sea pens, the bleaching is worst in the year 2017 with the highest bleaching percentage values**.



The above visualisation depicts the effect of site location on different kinds of corals. **It can be seen that in site 1, 3, 5, 7, and 8 the bleaching is worst for soft corals. In site 2, 4 and 6 the bleaching is worst for sea fans**.