CORAL REEFS AND HOW THEY HAVE DEGRADED OVER THE YEARS

A LOOK AT THE ENVIRONMENTAL FACTORS

DEEPA SHREE C V

PES2UG19CS105

Global Warming sits on the top of the world's most concerning issues. There is still a long way to go in this seemingly never-ending fight against global warming and its counterparts. As we know, raising temperatures have affected various realms of the environment. Various forms of life have been subjected to adverse, often irreversible effects. The once exquisite coral reefs, are now undergoing degradation (also called bleaching). This has been happening for decades. The current dataset takes a look at various factors that have affected coral bleaching rate.

The main objective of the current visualisation project is to get an idea of how various environmental factors play a role in the bleaching of the coral reefs around various parts of the world.

Background of the dataset:

The dataset spans over two decades from 1998-2017. The source is the website of BCO-DMO(Biological and Chemical Oceanography Data Management Office). The data is collected by a mixture of professional scientists (56%) and trained and certified citizen-scientists (44%). The authors of the dataset have used the global Coral Reef Temperature Anomaly Database (CoRTAD Version 6) from the National Oceanic and Atmospheric Administration to predict coral bleaching prevalence and intensity across reefs worldwide. The variables were collected on a weekly basis.

The current version of the dataset is downloaded from the Woods Hole Open Access Server.

Description and statistical analysis of the dataset:

The dataset is tab separated(tsv format). It contains a total of 9665 rows and 48 fields. The fields give a description of the geographical region of the reefs, average bleaching, various measures of Thermal Stress Anomaly(TSA) , temperature related features like sea surface temperature anomaly(SSTA) etc. A detailed description of the dataset can be found in 'Parameters' section of this website.

There are around 12 string type columns, out of which two represent the date in different format. Hence only one column is sufficient. The rest of the categorical features talk about

the geography of the coral reef. All these could be used to indicated the metadata about the reefs while plotting them on a symbol map.

The other columns are of numeric type. As mentioned earlier, they mostly talk about the temperature and few other features like the average bleaching, wind speed etc, latitude and longitude. All the variables are continuous.

The dataset contains null values represented as 'nd'. The columns City Town 2, City Town 3 have a high number of null values. The columns can be dropped as they are not of relevance to the current visualisation. Though the dataset contains a huge number of columns, most of them can be grouped and fit into a single graph.

The dataset does have a few outliers but upon careful examination it was decided that these need to be retained. A snapshot of outliers of a few prominent columns is given below.

There isn't any need for transformations of the categorical variables (such as encoding etc.) as the names have to be retained during visualising. As the numerical columns are all measured of various factors influencing the bleaching, it is better to retain them rather than transforming them. This would help the users get better insights of the actual values from the visuals.
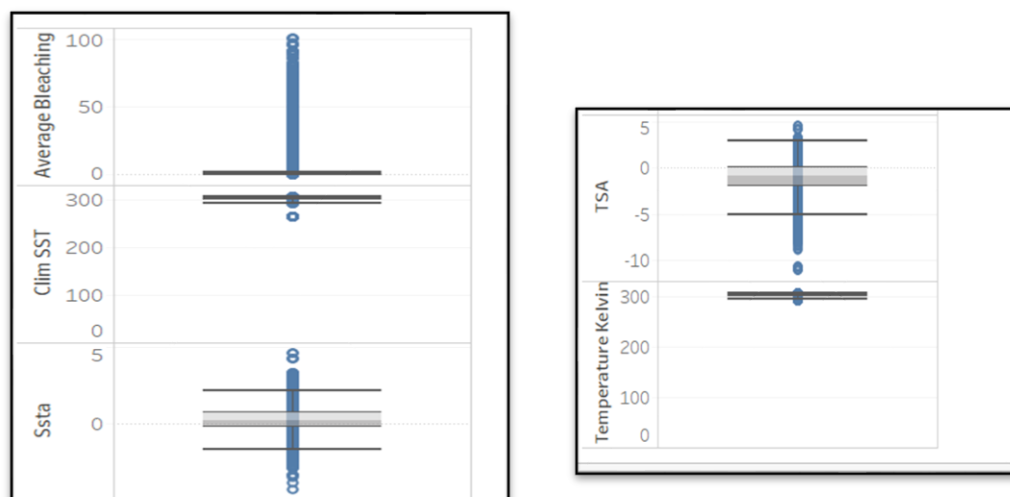

Boxplot of a few columns

Potential Visualisations:

- A spot map showing the different ecoregions covered by the dataset. It could contain the average bleaching and few other attributes like the exact location of the coral reef, etc.
- Charts (Scatter Plots, etc) to draw relations between the average bleaching and the various factors like Temperature, Wind Speed, Sea Surface Temperature, Thermal Stress Anomaly.
- A linear plot between the date and the Average Bleaching to gain information about how the bleaching has changed over time.
- Bar plots to know the density of coral reefs in different parts of the worlds.

Strengths of the dataset:

The dataset has covered a large timeline from 1998-2017, that is almost two decades of data. It also spans a huge area, covering almost half of the world. The regions covered are given in the shaded in blue in the map below. It contains numerous attributes, which is very helpful to draw useful insights and contribute largely to understanding the problem at hand. The dataset does not need much pre-processing.
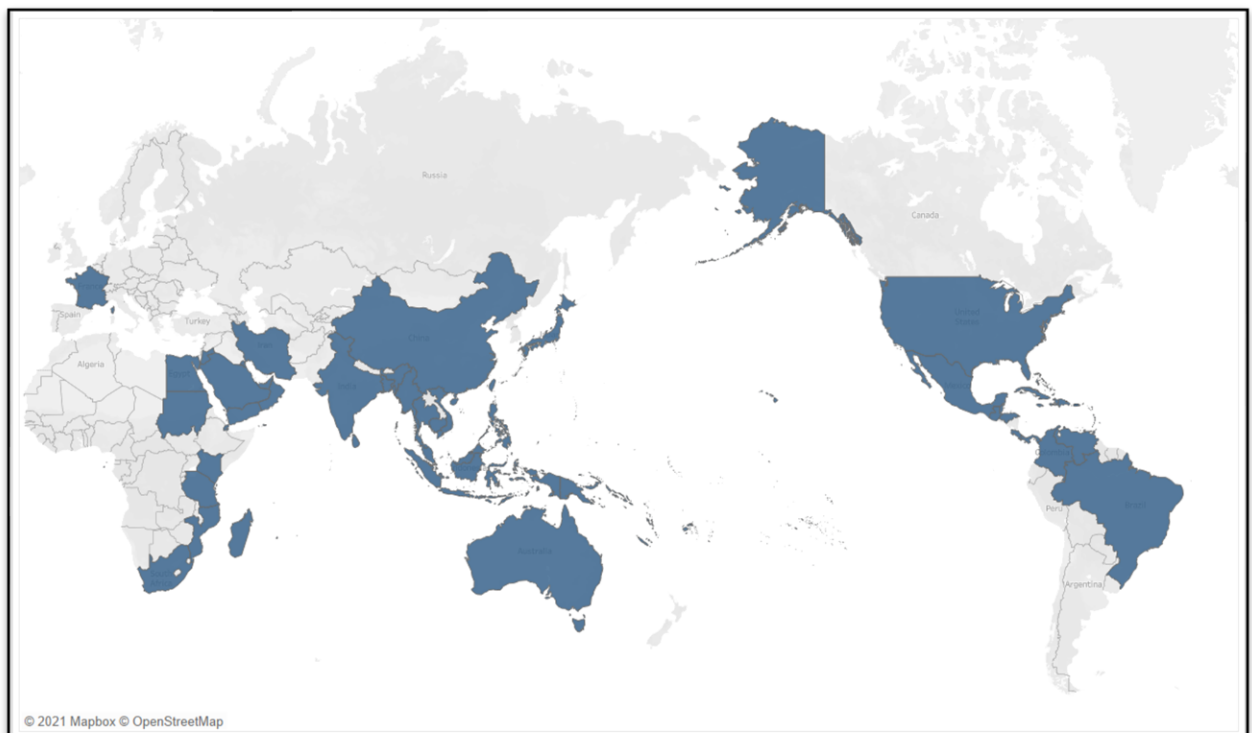
Regions covered by the dataset

Limitations:

The dataset is a bit outdated. The last timestamp is 2017, that is almost four years ago. There might have been significant changes in recordings from the past four years. It would have been better if the authors took measured to update the data at regular intervals. Though the dataset covers a huge area, it doesn't contain information about a few major areas. A few columns(ci contain too many null values, that they to be dropped completely. If these columns had been filled with the required data, the visualisations would've been more insightful and easier to understand.