Problem Statement

Investigate the data set provided to answer the following two key questions.

1. In **which years** and for **which kinds of coral** bleaching is the **worst**.
2. How the **location** of the site affects the **bleaching** on the different **kinds of coral**.

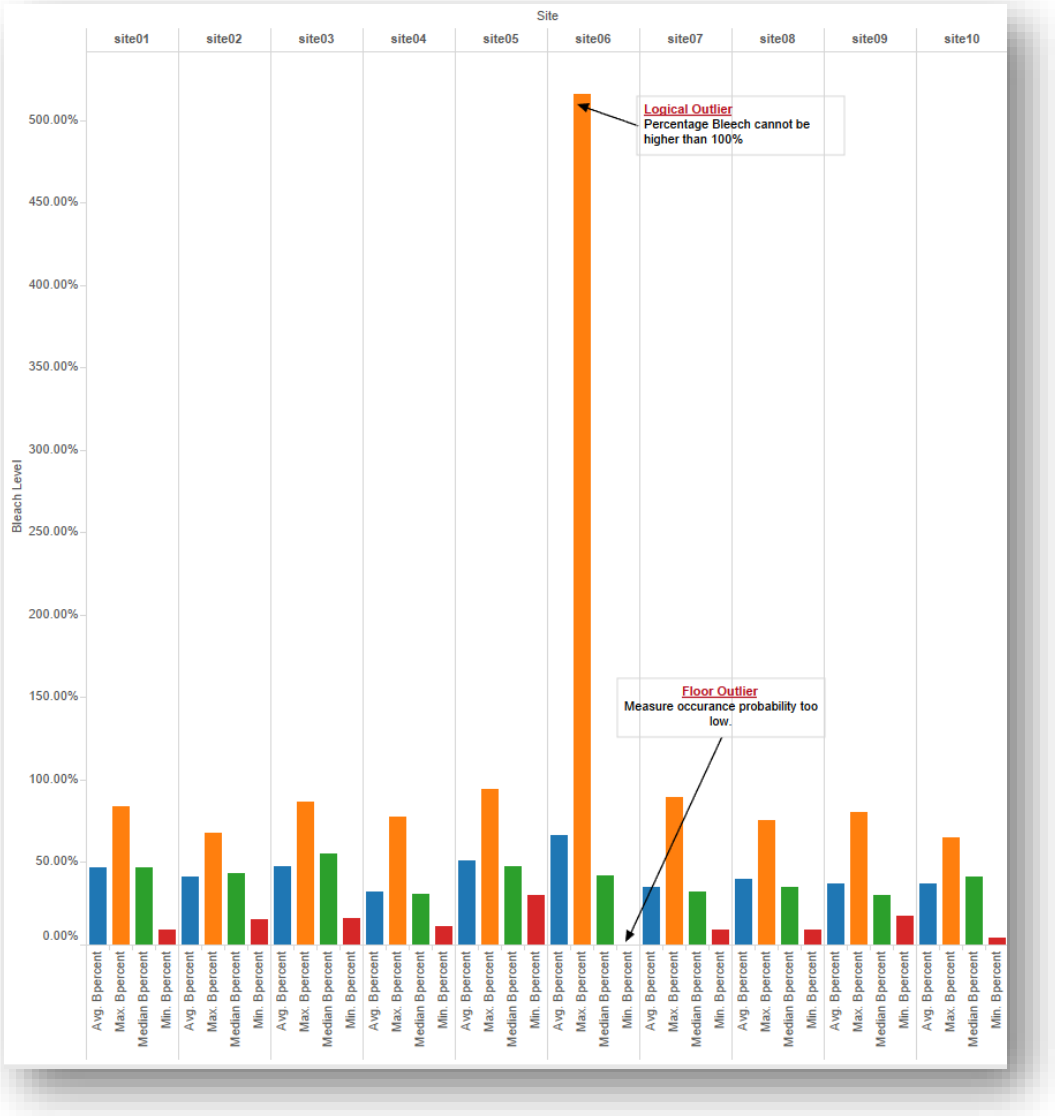
Exploratory Data Analysis

Reading in the initial file into Python Pandas, showed that the column headers were multi-index headers. The multi-index header meant that two features, the year and the coral type, were unfavourably placed for melting to a long type frame. It was necessary to place these into a sing-index header. Once this was completed, the data set was melted to a long style data frame.

```
In[16]: _melt.head()
Out[16]:
     site longitude latitude        coral  year bpercent
0  site01   143.515  -11.843  hard corals  2016   0.8387
1  site02   145.043  -14.383  hard corals  2016   0.6789
2  site03   146.589  -17.981  hard corals  2016      NaN
3  site04   150.444  -20.414  hard corals  2016   0.1245
4  site05   143.786  -13.107  hard corals  2016   0.9423
```

Visualisation

Tableau was used to develop some preliminary EDA plots, which showed geographical and measurement outliers.



Data Wrangling

Comparing the longitude of the outlier site to the other sites, it was evident that the negative symbol may have been implemented accidentally. Also it is very unlikely that a value of 516% is present. It also violates some logical factors such as, bleaching more than there is coral. This value has been converted to a NA value.

Question 1

To answer the question, which kind of coral bleaching is the worst in which years, we can take the mean bleaching as well as the maximum bleaching values. (Double click to open).