# 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.

# Data Wrangling

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

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. From Figure 1, we can elicit the answer to the question and can summarise it in the table below.

|  |  |  |
| --- | --- | --- |
| Year | Worst Average Bleach Coral | Worst Maximum Bleach Coral |
| 2012 | Blue corals | Hard corals |
| 2013 | Hard corals | Hard corals |
| 2014 | Hard corals | Hard corals |
| 2015 | Sea pens | Hard corals |
| 2016 | Sea pens | Hard corals |

# Question 2

From figure 2, it is evident that hard corals on average seem to be more bleached in the north, where as hard corals seem to be less dominant in the south.

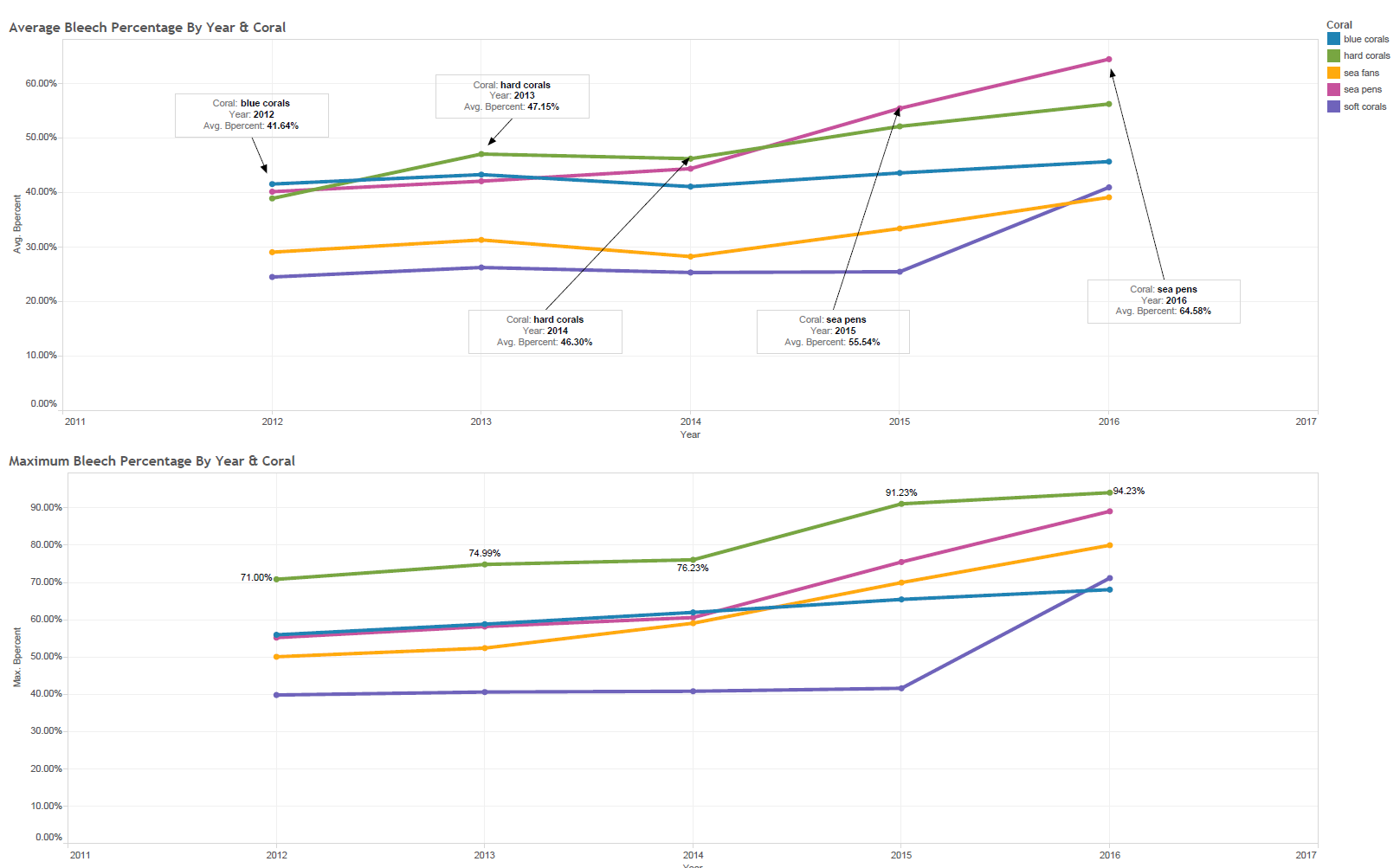


Figure : Bleaching by Year Average and Max

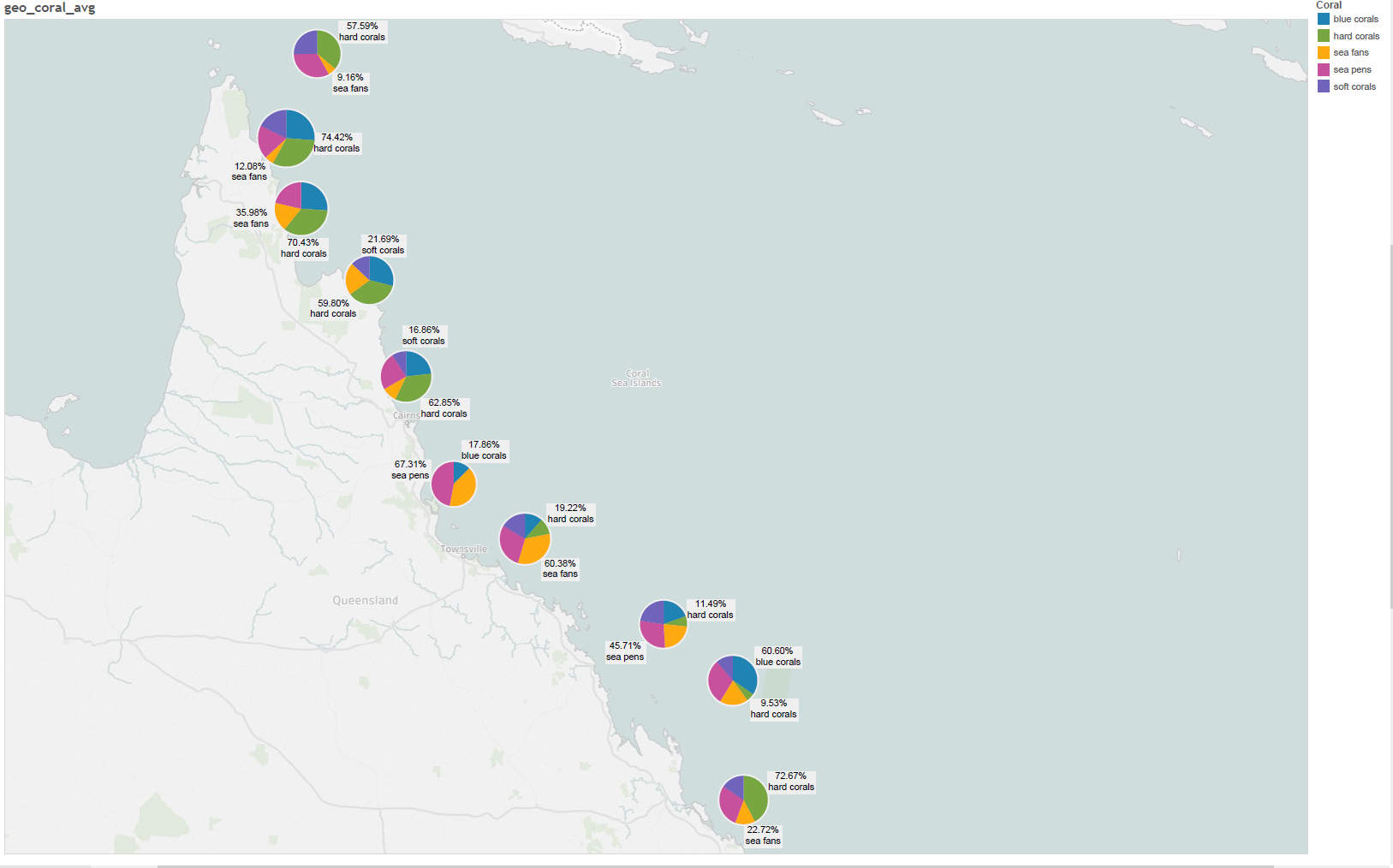


Figure : Geographic Distribution of Average Bleaching