

DALT7016

Data Visualisation

Student Number: 19182799

MSc Course: MSc in Artificial Intelligence

Author: Sam Trowbridge

Word Count: 1465

Contents

Introduction	3
Design.....	3
Selected Dataset	3
Vision.....	3
Data Analysis.....	3
Development.....	7
Visualisation 1	7
Visualisation 2	15
Prototypes.....	16
Final.....	17
Conclusion.....	18
References	19

Introduction

Data visualisation is an important aspect of any project, in a data driven age it has become increasingly difficult to display data in a human readable manner. In the current day and age enormous amounts of data is recorded at any given time, however if data is not handled and presented in an easily understandable manner it can become near redundant. Therefore, skills in data manipulation and visualization are now vital in many fields of expertise. This report attempts to explain important steps in data analysis and demonstrate two examples of good data visualization.

Design

Selected Dataset

The dataset selected for this project is “A global coral-bleaching database” (Team SDC), this dataset consists of over 38000 rows of data gathered from 1980-2020 from coral health checks from across the globe and has been selected due to the importance of the subject. Research has shown that global warming is a possible cause of ocean temperatures rising and therein the bleaching of coral. However, there is a large amount of public dubiousness as to whether either of these theories are a reality. Although with enough data displayed in an understandable manner more proof of the accurate truth can be developed. Encouraging more of the public to take action to lessen and or stop these damaging effects on the environment.

Vision

The aim of these visualisations is to show whether coral bleaching has become a more common occurrence over the last 40 years and which areas have been badly affected, also whether this may correlate with rising temperatures in the ocean.

Data Analysis

The most important step for any kind of data visualisation is understanding what data is present, how much data is available and how reliable or consistent the data is. Therefore, data analysis is vital to a successful visualisation, luckily with machine technology this step can be an easier and more efficient task. Most programming languages now provide basic functionality to allow users to assess and analyse data before handling it, the R programming language was developed for exactly these kinds of scenarios.

R programming libraries such as ‘dplyr’ and ‘tidyverse’ were developed precisely for data manipulation, ggplot and gganimate are another two libraries designed for specifically plotting manipulated data in a variety of possible ways. Some of these libraries have been used during the process of this data analysis for better understanding of the data available. However beforehand a basic visual look through the data is necessary.

Figure 1

The initial datafile received was an access database file (accdb) with various tables and queries as shown in figure 1. Therefore, after some assessment of the different queries and the data contained “Query 1_Summary_Bleaching_Cover” appeared to contain the particular information required for this project. After some basic analysis it seemed appropriate to assign an additional column stating the “Severity_ID” for the “Severity_Code” as found in the “Severity_Code_LUT” table using a simple VLOOKUP function in Microsoft Excel. This method was also used to assign each country a two-digit code in another new column for labelling ease. The “Query 1_Summary_Bleaching_Cover” table was then exported as a CSV file so it could then be loaded into Rstudio for the

next stage.

	Severity_Code	n
1		31081
2	Mild (1-10% Bleached)	2236
3	Moderate (11-50% Bleached)	2007
4	Severe (>50% Bleached)	1862
5	% unknown	530
6	No Bleaching	332

Figure 2

	Percent_Bleached_Sum	n
1	0.0000000	16113
2	NA	6763
3	0.2500000	542
4	1.0000000	483
5	5.0000000	472
6	0.5000000	422
7	2.5000000	375
8	10.0000000	363
9	1.2500000	348
10	25.0000000	252
11	0.7500000	248
12	20.0000000	220
13	12.5000000	209
14	7.5000000	193

Figure 3

The critical information within this dataset is the “Severity_Code” however as displayed above after an initial count method it is clear that a large amount of the data is missing these values. Despite this as shown in figure 3 a much smaller number of rows had no value for the “Percent_Bleached_Sum” column. Therefore, by using the data from “Percent_Bleached_Sum” a severity code was assigned and all N/A values were assigned as “% unknown” (see figure 4).

	Severity_range	n
1	No Bleaching	16360
2	Mild (1-10% Bleached)	8173
3	% unknown	6763
4	Moderate (11-50% Bleached)	4940
5	Severe (>50% Bleached)	1812

Figure 4

In order to better understand this information, a geom density plot (shown below) using the ggplot library was implemented to show the count of each severity code across each year. This plot shows the massive increase in available data as the 2000's are entered, there could a wide range of reasons for this such as perhaps it's during this time that health checks became common practice. However, it also shows that there are far more cases of coral bleaching than no bleaching, it also becomes clear that colour plays an important role in data visualization. The colours selected are very distinct from one another avoiding confusion, It is also imperative that colour blindness is taken into account due to a substantial amount of the populace suffering from some kind of colour blindness.

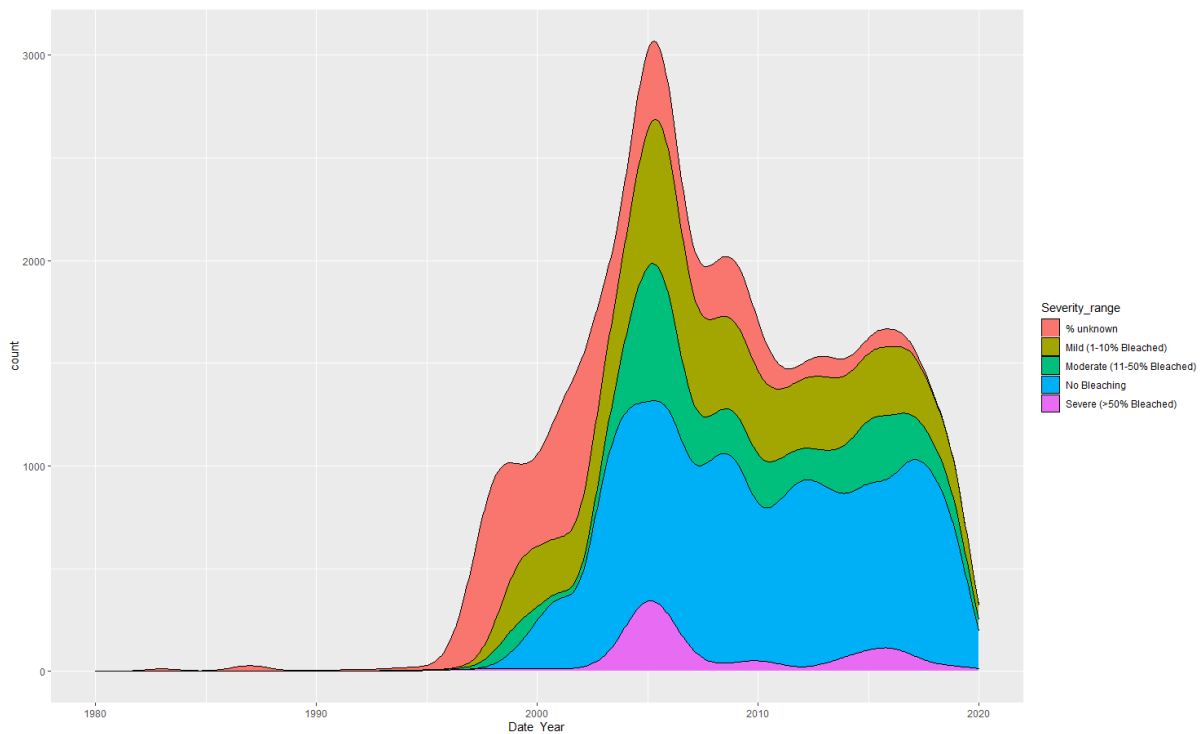


Figure 5

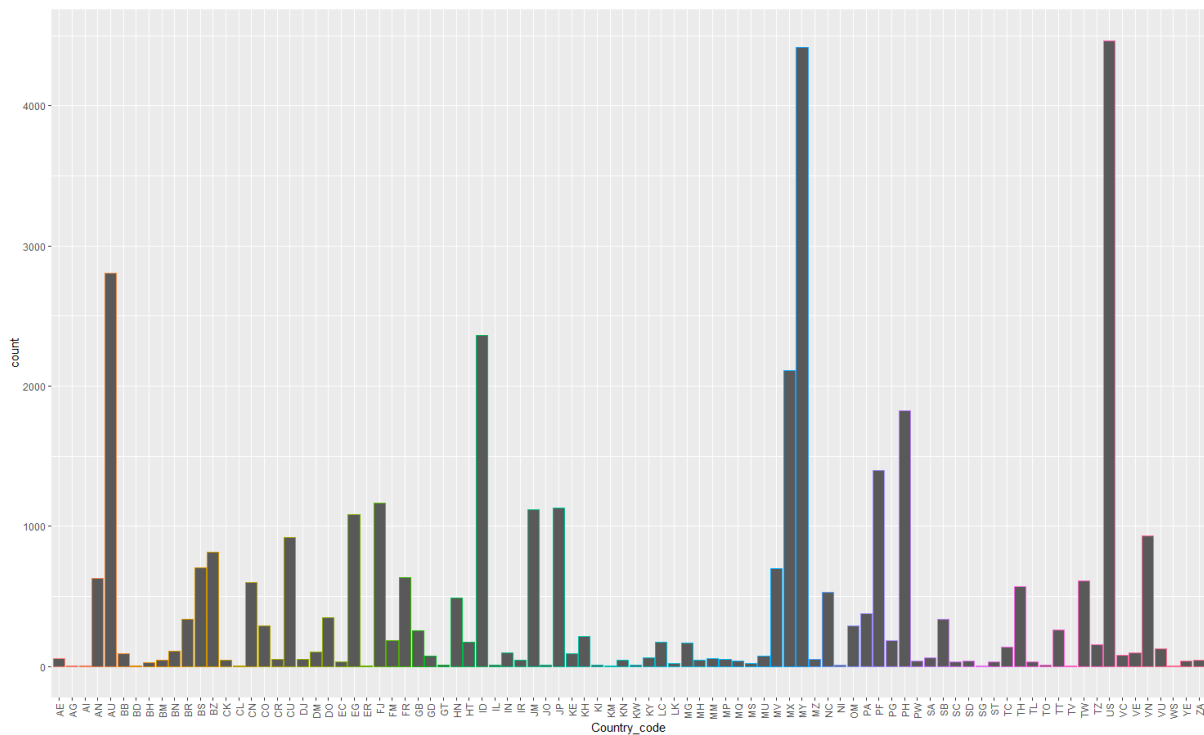


Figure 6

The plot displayed above is a perfect example to demonstrate the difficulties that can be faced when dealing with large datasets with numerous allocations. In this case the X axis shows the “Country_code” of each country along with the number of coral health checks made from 1980-2020. It distinguishes that the top five countries who have carried out the most health checks: “US”(United States), “MY” (Malaysia), “AU” (Australia), “ID” (Indonesia) and “MX” (Mexico).

All Access Objects

Search...

- Bleaching_tbl
- City_Town_Name_LUT
- Country_Name_LUT
- Cover_tbl
- Data_Source_LUT
- Ecoregion_Name_LUT
- Environmental_tbl
- Exposure_LUT
- Ocean_Name_LUT
- R_Scripts_tbl
- Realm_Name_LUT
- Relevant_Papers_tbl
- Sample_Event_tbl
- Severity_Code_LUT
- Site_Info_tbl
- State_Island_Province_Name_LUT
- Substrate_Type_LUT

Queries

- Query 1_Summary_Bleaching_Cover
- Query 2_Sample_Event_Counts
- Query 3_Time_Series
- Query 4_Samples_by_Data_Source
- Query 5_Sites_by_Data_Source
- Query 6_Sites_with_Multiple_Sample_Events
- Subquery 1_Sites_and_Sample_Events
- Subquery 2_Reef_Check_Average_Bleaching
- Subquery 3_Summed_Percent_Bleaching
- Subquery 4_Bleaching
- Subquery 5_Bleaching_Cover_Combined
- Subquery 6_Calculated_Reef_Check_Segments
- Subquery 7_Cover_Averages_Reef_Check
- Subquery 8_Reef_Check_Calculated

Environmen	Sample_ID	ClimSST	Temperatur	Temperatur	Temperatur	Temperatur	Temperatur	Windspeed	SSTA	SSTI
9622	9623	299.48	301.30	298.85	293.85	304.38	2.48	8.00	-0.21	
9623	9624	299.92	302.17	298.93	294.13	305.10	2.46	5.00	0.45	
9624	9625	302.20	300.94	298.93	294.13	305.10	2.46	6.00	-0.50	
9625	9626	301.76	300.26	299.09	293.98	305.42	2.42	5.00	-2.06	
9626	9627	299.86	301.40	298.92	294.25	305.51	2.46	5.00	-0.27	
9627	9628	298.52	300.78	298.92	294.25	305.51	2.46	8.00	0.15	
9628	9629	300.25	294.64	298.92	294.25	305.51	2.46	11.00	-3.35	
9629	9630	301.10	298.16	298.84	293.32	305.66	2.42	8.00	-1.03	
9630	9631	301.77	303.24	298.65	293.22	303.83	2.46	3.00	1.50	
9631	9632	262.15	299.15	298.64	293.10	304.35	2.50	5.00	-0.80	
9632	9633	299.06	300.45	297.92	292.51	304.27	2.64	7.00	-0.36	
9633	9634	299.06	300.45	297.92	292.51	304.27	2.64	7.00	-0.36	
9634	9635	300.44	301.26	297.92	292.51	304.27	2.64	4.00	-0.28	
9635	9636	299.06	300.45	297.92	292.51	304.27	2.64	7.00	-0.36	
9636	9637	299.06	300.45	297.92	292.51	304.27	2.64	7.00	-0.36	
9637	9638	262.15	302.52	297.86	291.57	304.28	2.70	7.00	1.34	
9638	9639	262.15	301.38	297.79	292.16	304.64	2.84	7.00	-0.34	
9639	9640	300.46	301.95	297.82	292.69	304.95	2.79	5.00	0.25	
9640	9641	301.74	301.71	297.82	292.69	304.95	2.79	7.00	0.57	
9641	9642	301.74	300.48	297.82	292.69	304.95	2.79	7.00	-0.56	
9642	9643	300.79	301.78	297.82	292.69	304.95	2.79	5.00	0.03	
9643	9644	301.50	299.44	297.74	292.05	306.41	2.79	8.00	-0.44	
9644	9645	297.37	300.40	297.90	292.10	304.33	2.88	5.00	1.05	
9645	9646	297.37	300.40	297.90	292.10	304.33	2.88	5.00	1.05	
9646	9647	298.14	301.20	297.90	292.10	304.33	2.88	7.00	1.06	
9647	9648	298.14	301.20	297.90	292.10	304.33	2.88	7.00	1.06	
9648	9649	298.88	296.01	297.90	292.10	304.33	2.88	4.00	-0.22	
9649	9650	299.35	302.10	297.90	292.10	304.33	2.88	6.00	0.56	
9650	9651	295.17	299.82	296.76	290.02	304.05	3.48	4.00	2.03	
9651	9652	299.41	302.04	296.76	290.02	304.05	3.48	9.00	0.36	
9652	9653	301.78	300.64	296.76	290.02	304.05	3.48	5.00	-0.06	
9653	9654	299.91	302.82	296.76	290.02	304.05	3.48	7.00	1.07	
9654	9655	299.41	302.04	296.76	290.02	304.05	3.48	9.00	0.36	
9655	9656	299.91	303.04	296.76	290.02	304.05	3.48	6.00	1.27	
9656	9657	300.78	304.05	296.76	290.02	304.05	3.48	3.00	2.36	
9657	9658	300.78	304.05	296.76	290.02	304.05	3.48	3.00	2.36	
9658	9659	294.46	299.13	295.24	286.85	303.41	3.64	4.00	1.94	
9659	9660	298.78	300.07	295.24	286.85	303.41	3.64	4.00	-0.39	

Figure 7

After some data analysis it seemed that water temperature from the “Environmental_tbl” in figure 7 would also be valuable information, therefore it was also exported as a CSV file to be used for a data visualisation. It became clear that the two tables data were connected by one column of variables “Sample_ID”, using this column both data frames were merged to create one data frame containing all of the data.

Development

Visualisation 1

The initial idea when choosing this dataset was to show how commonplace coral bleaching has become over recent decades, the immediate idea was the first visualisation. One effective way to show clearly to the general population that an issue is global and progressively getting worse everywhere is using a world map. Therefore, the first visualization is the plotting of all coral bleaching events every year between 1980-2020 within this extensive dataset, the shapes and colour of points corresponding with the severity of the bleaching as shown below

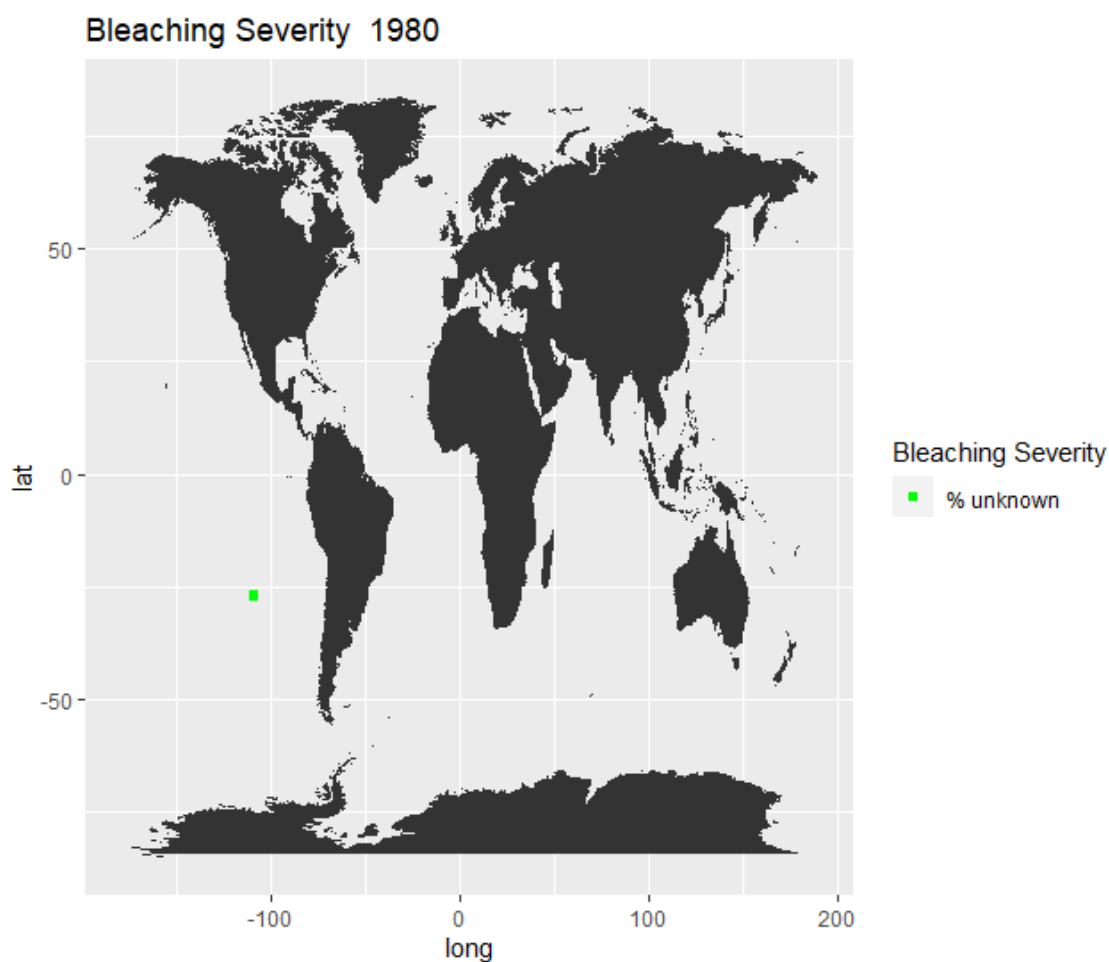


Figure 8

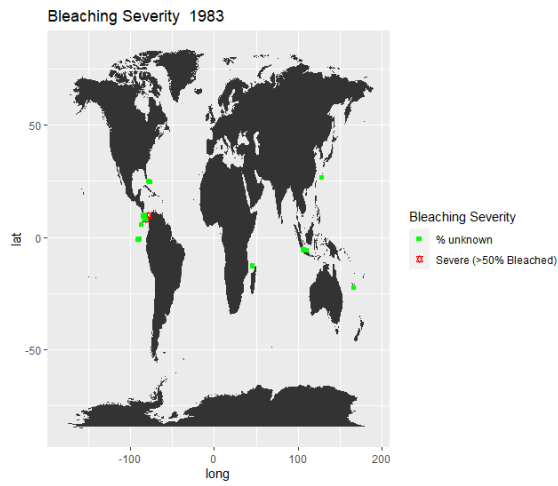


Figure 9

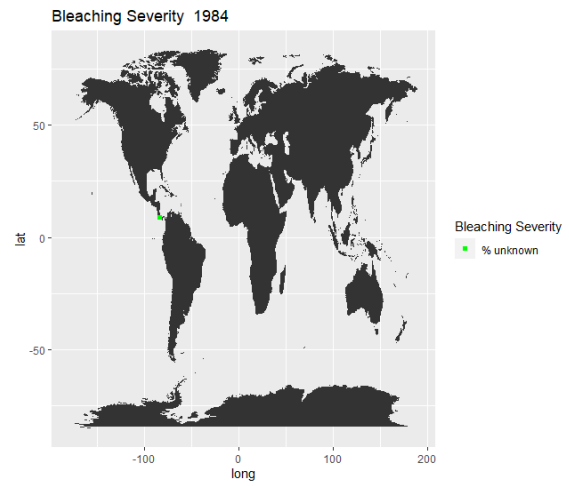


Figure 10

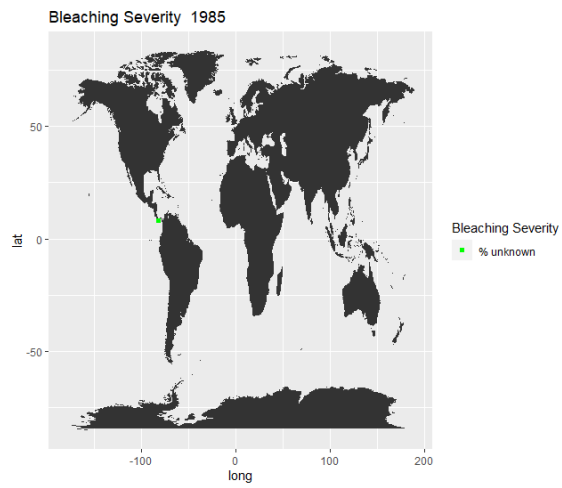


Figure 11

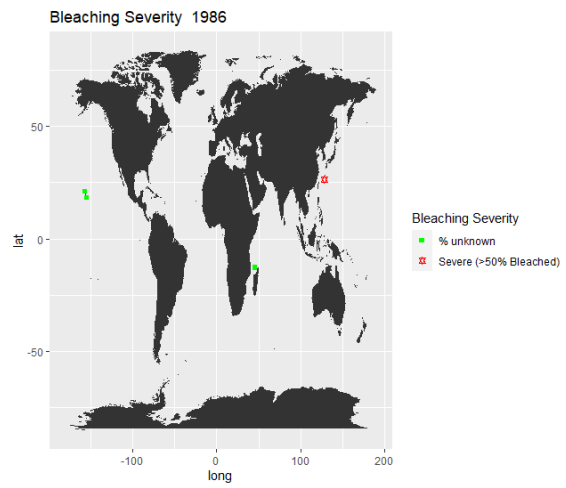


Figure 12

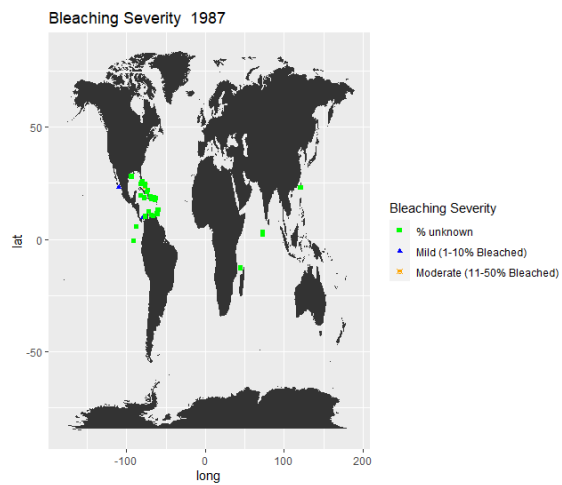


Figure 13

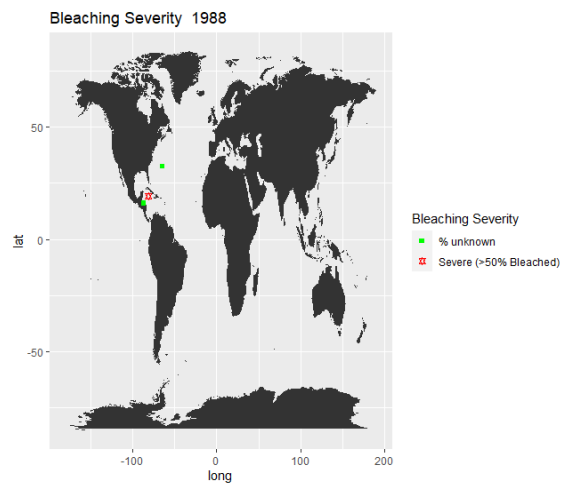


Figure 14

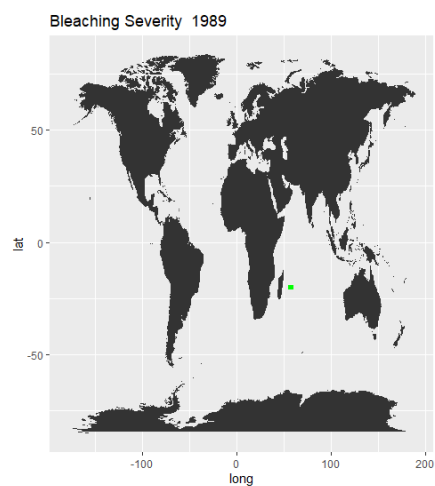


Figure 15

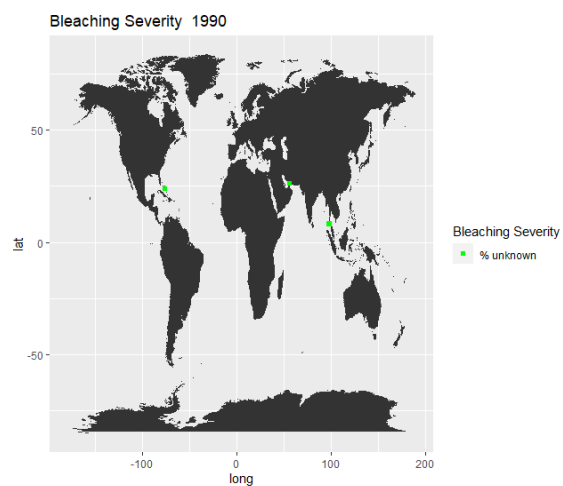


Figure 16

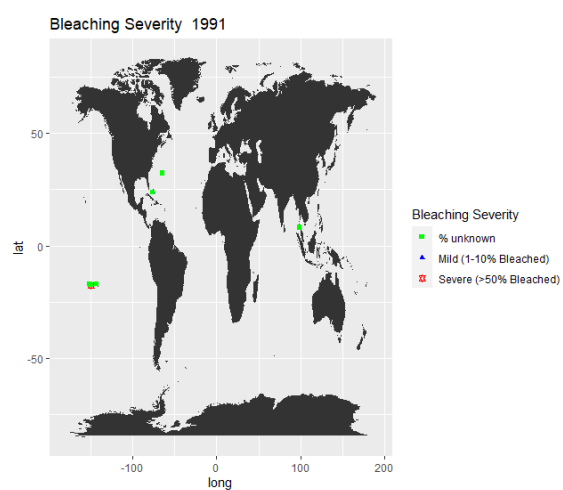


Figure 17

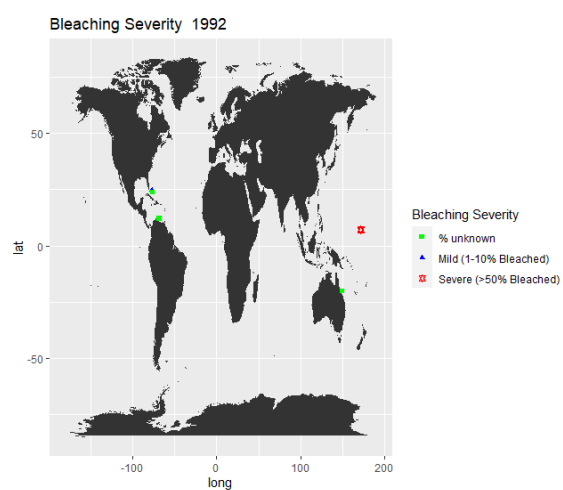


Figure 18

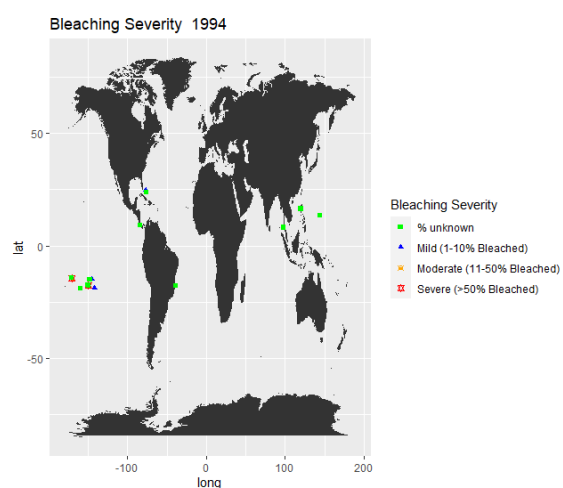


Figure 19

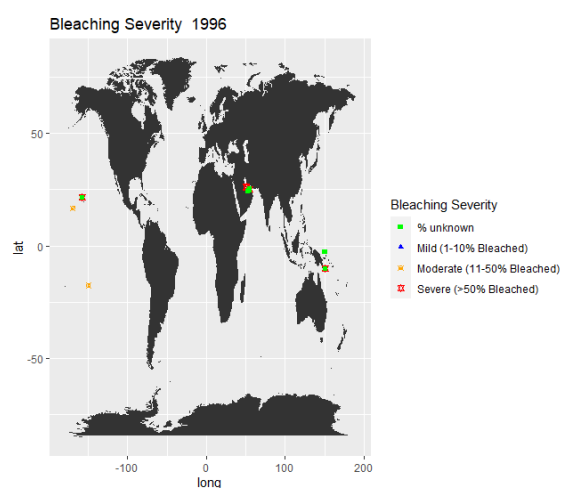


Figure 20

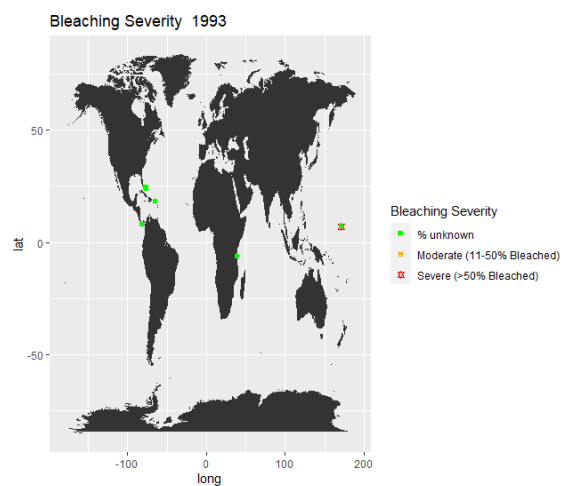


Figure 21

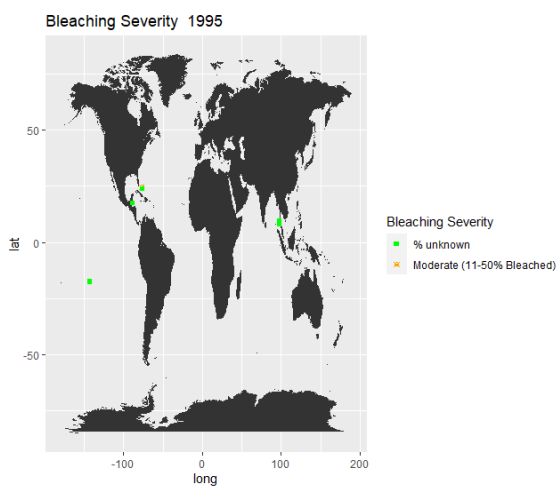


Figure 22

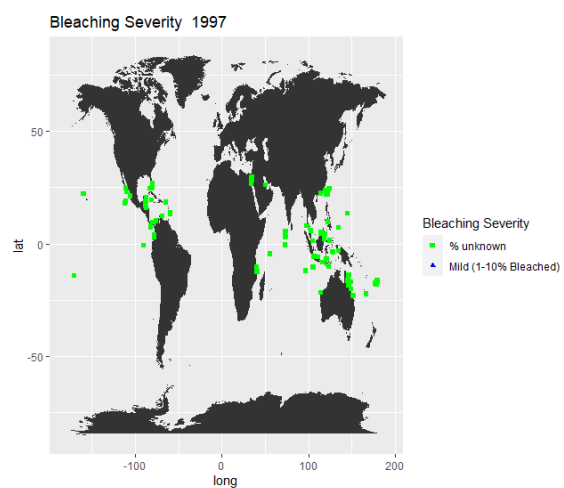


Figure 23

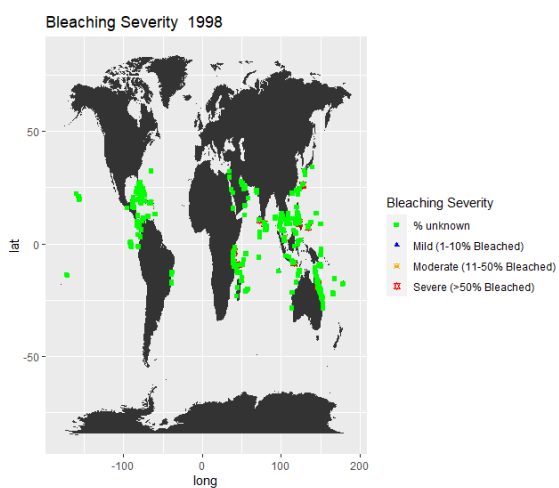


Figure 24

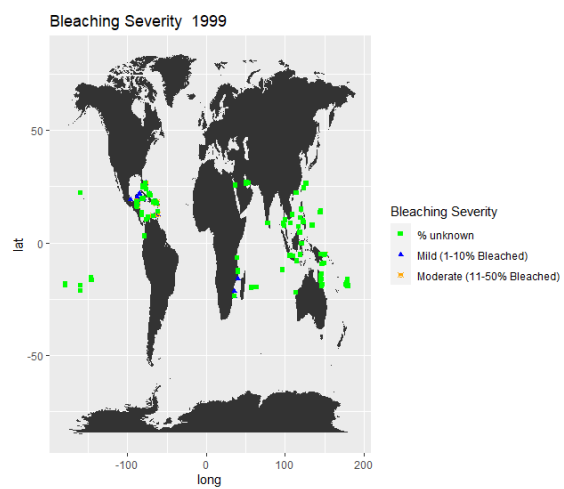


Figure 25

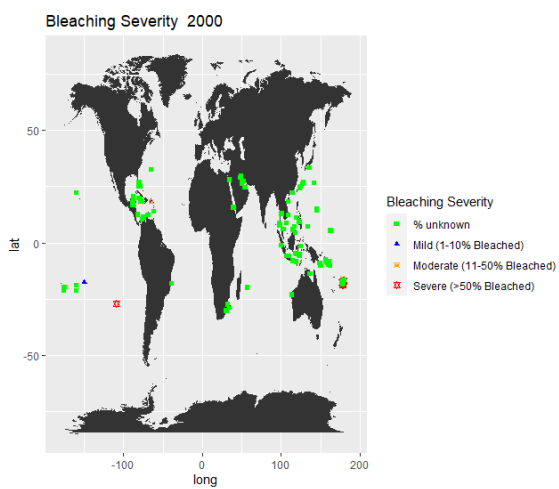


Figure 26

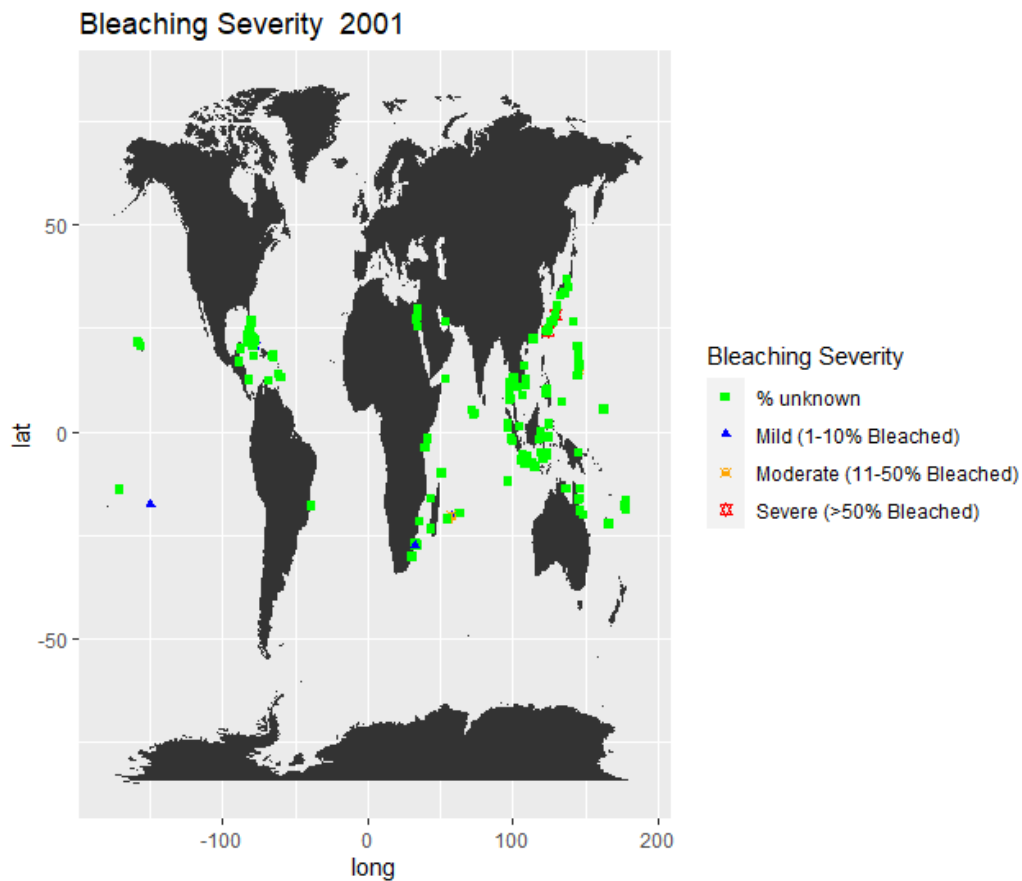


Figure 27

As figures 8 to 26 show, in the 1980s there are almost no recorded bleaching events however its progresses through the 1990s more and more bleaching events are documented. This could be due to a number of reasons such as an increase in the number of health checks as the importance of environmental health are made more aware. However, given the increase in the severity of bleaching through the 2000s (see figures 27 – 36) it can be deduced that bleaching is a real threat that may be getting increasingly worse.

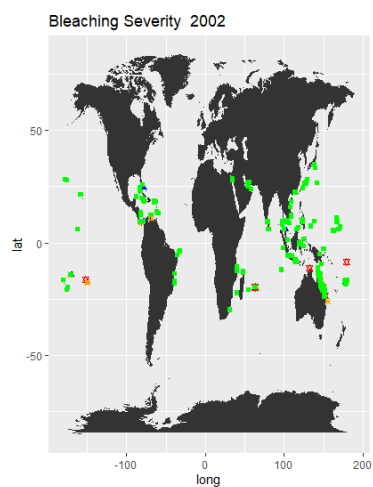


Figure 28

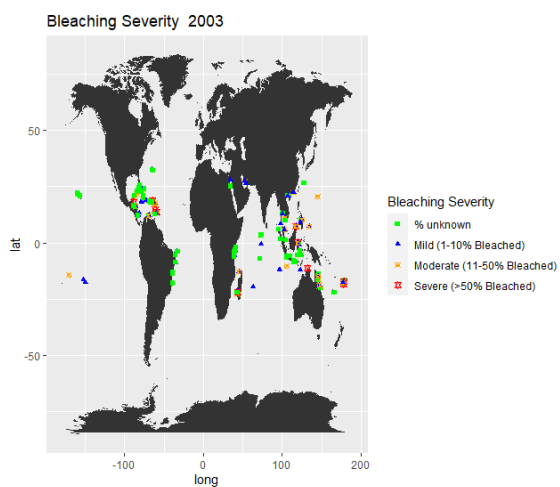


Figure 29

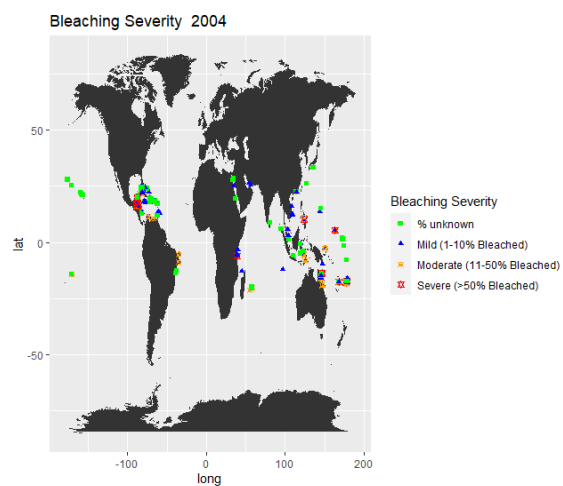


Figure 30

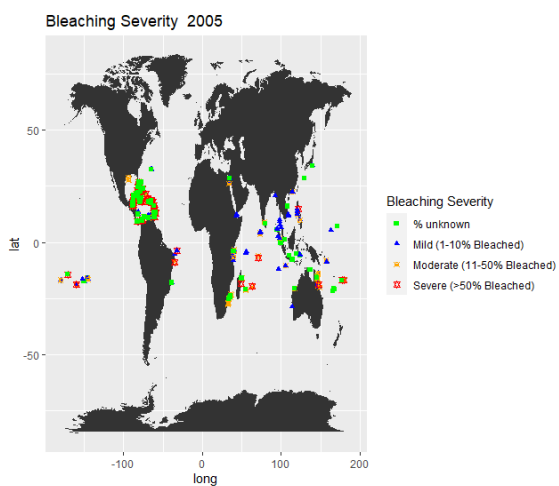


Figure 31

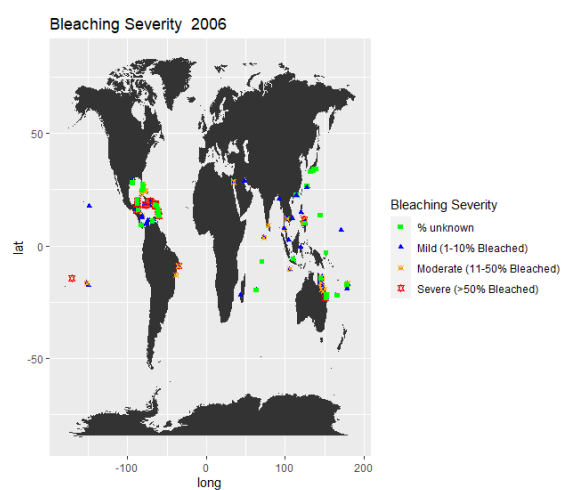


Figure 32

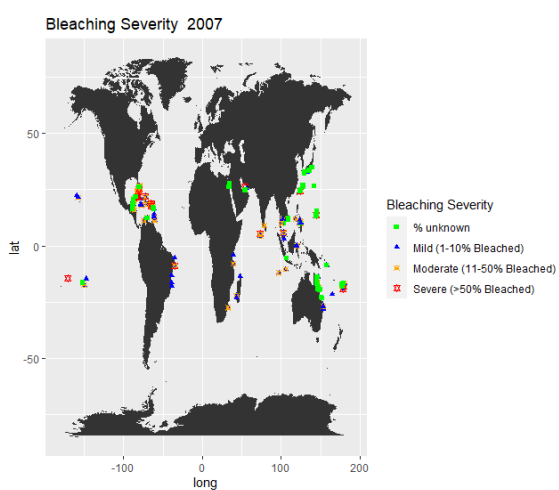


Figure 33

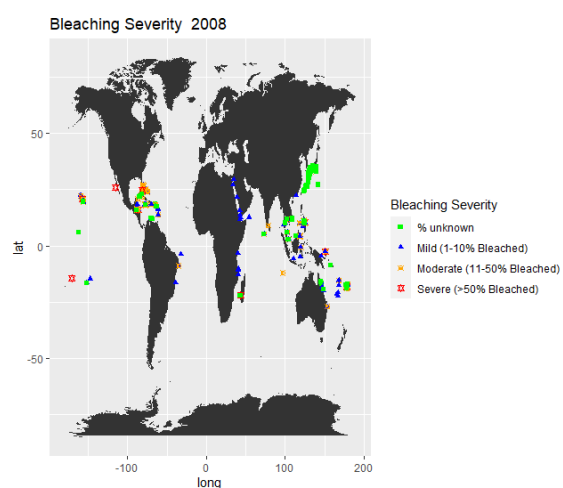


Figure 34

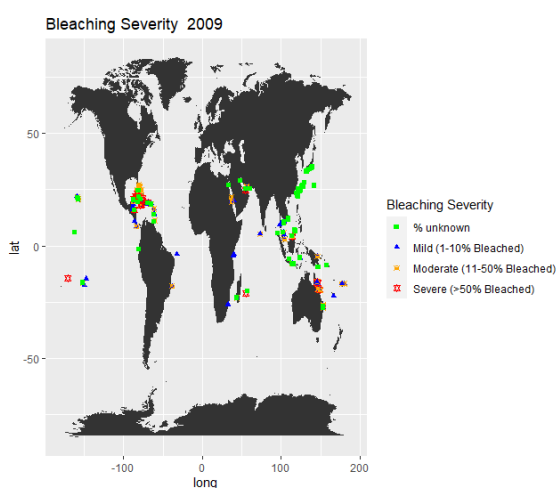


Figure 35

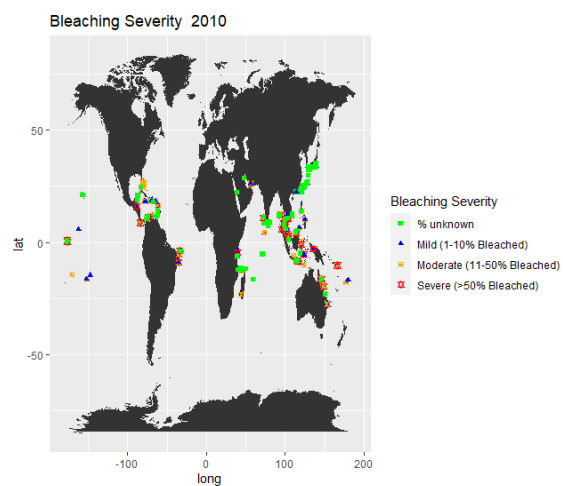


Figure 36

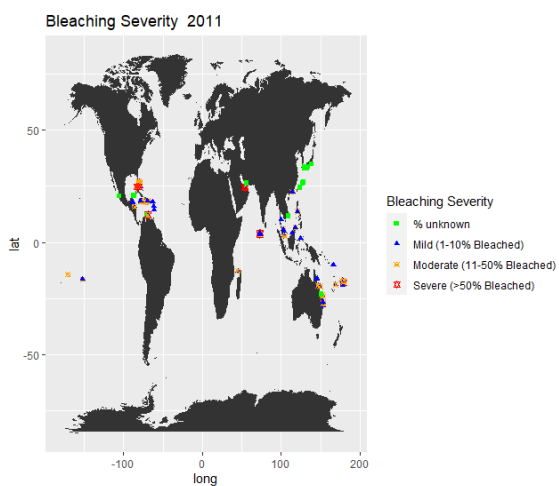


Figure 37

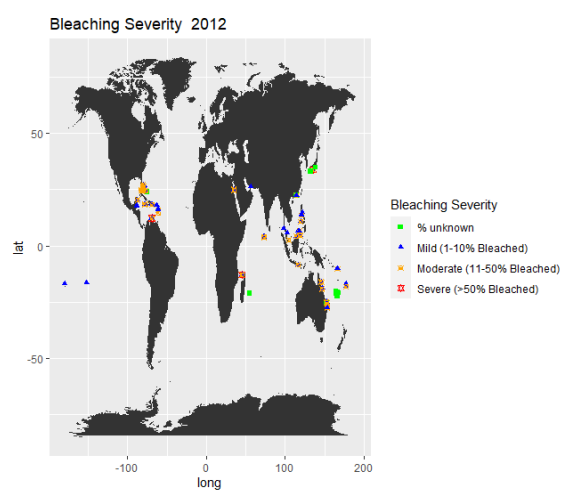


Figure 38

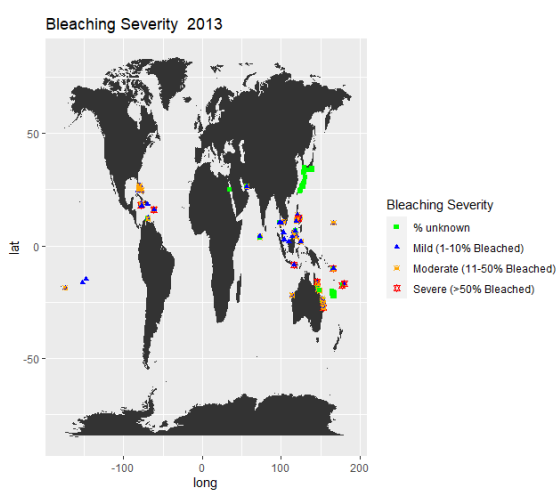


Figure 39

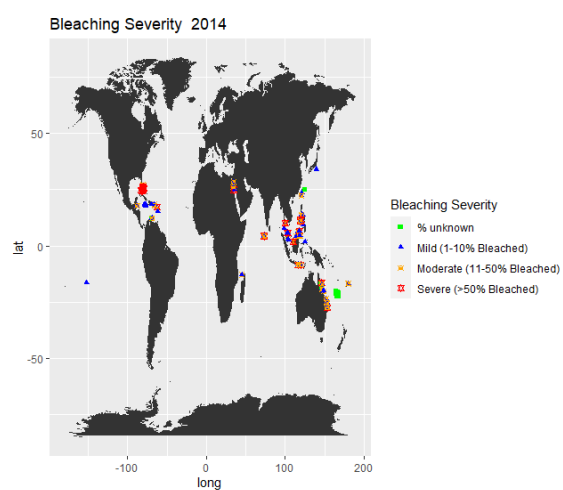


Figure 40

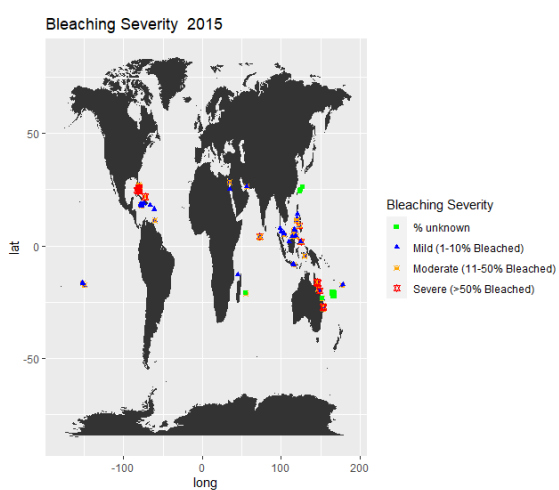


Figure 41

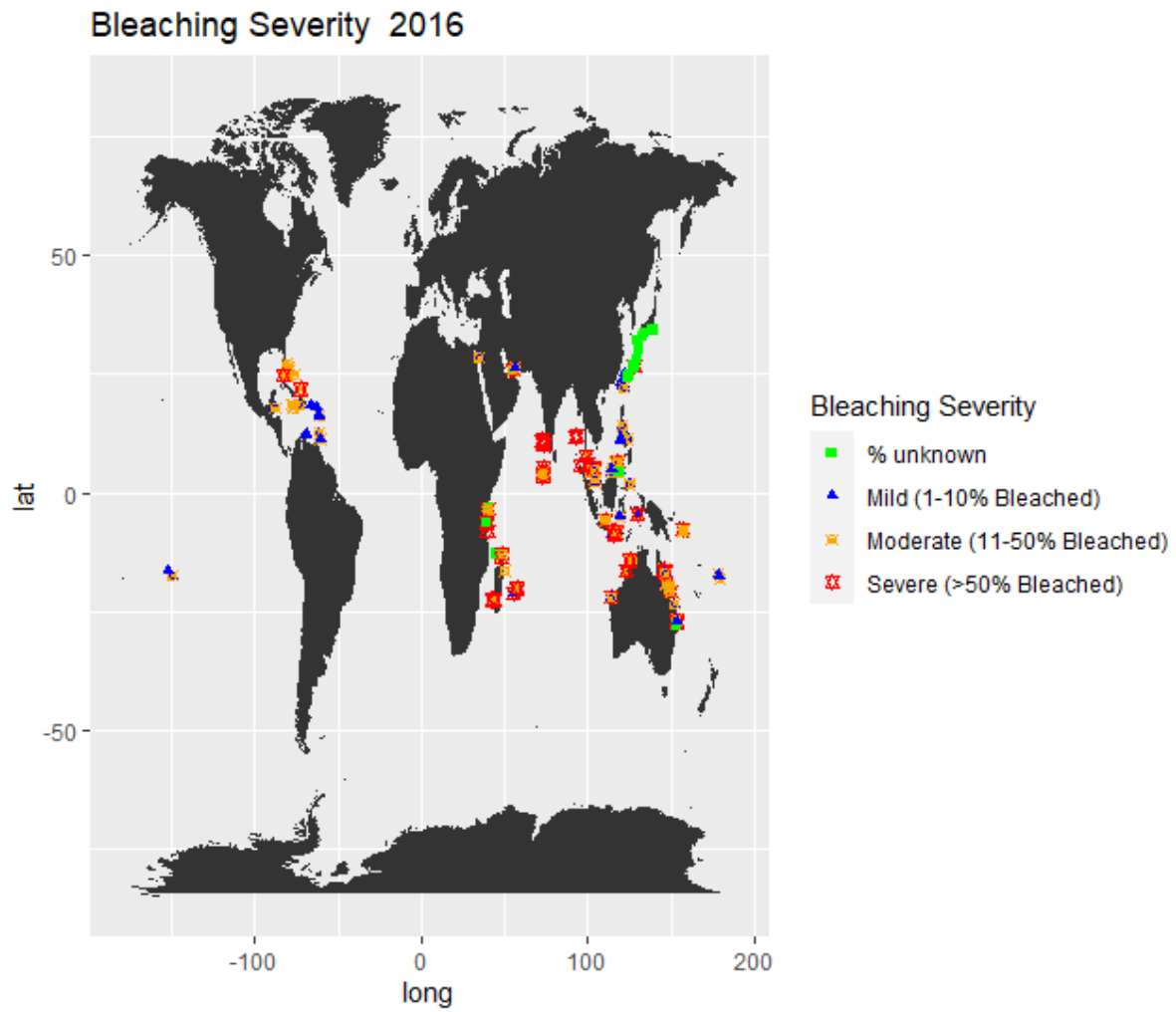


Figure 42

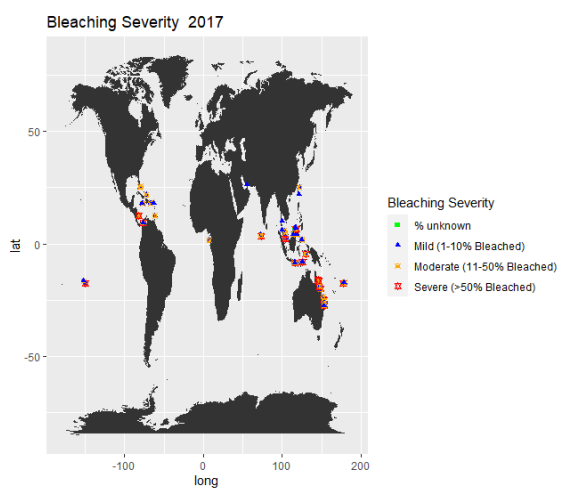


Figure 43

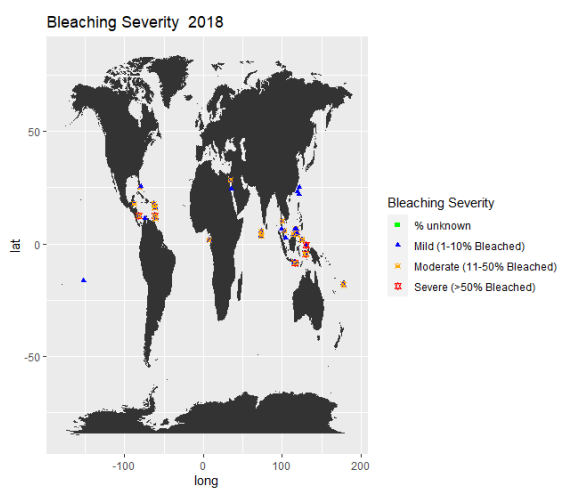


Figure 44

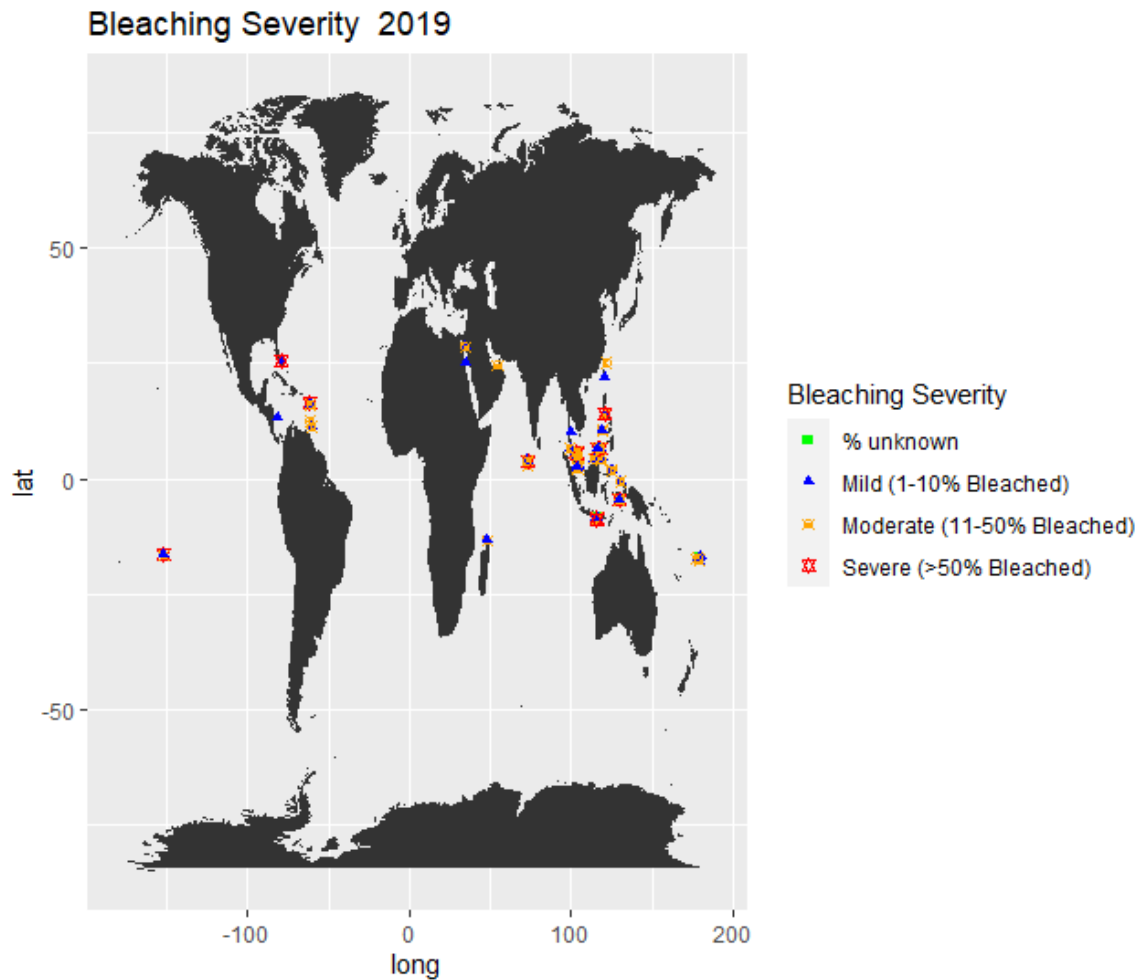


Figure 45

In these visualisations the chosen colours matched with the shapes help to show the severity of the bleaching despite some areas getting increasingly crowded. These visualizations help provide a powerful message that despite opinions on why or how coral bleaching has got worse in recent decades.

Visualisation 2

The second visualisation wasn't so simple, the initial idea was to focus on a certain location such as The Great Barrier Reef and displaying the increase in bleaching over the years using a form of heatmap. However, after some evaluation on the quality of temperature data it seemed only fitting to try and answer the question as to what may be causing coral bleaching.

Prototypes

The first attempts focused on The Great Barrier Reef and attempted to show the increase in the max temperatures (figure 46). This didn't quite convey the message so the next attempt in figure 47 showed the raise in average temperatures throughout each year, this visualization is more striking and closer to the impact I was aiming to achieve.

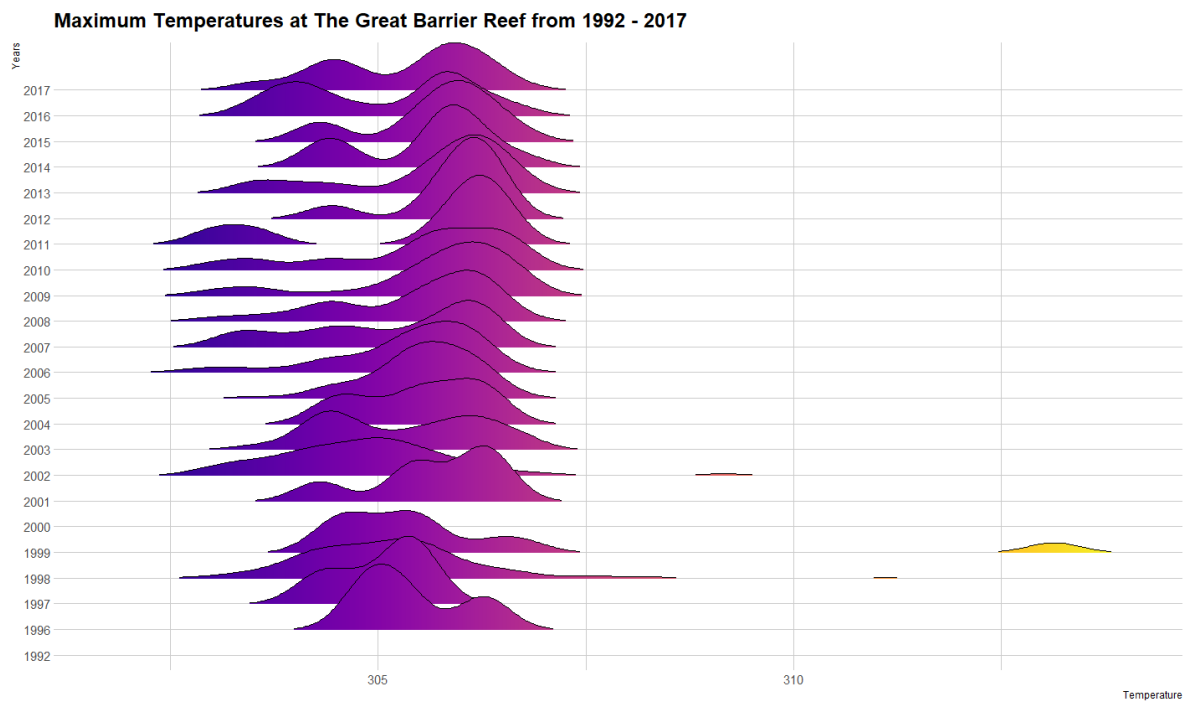


Figure 46

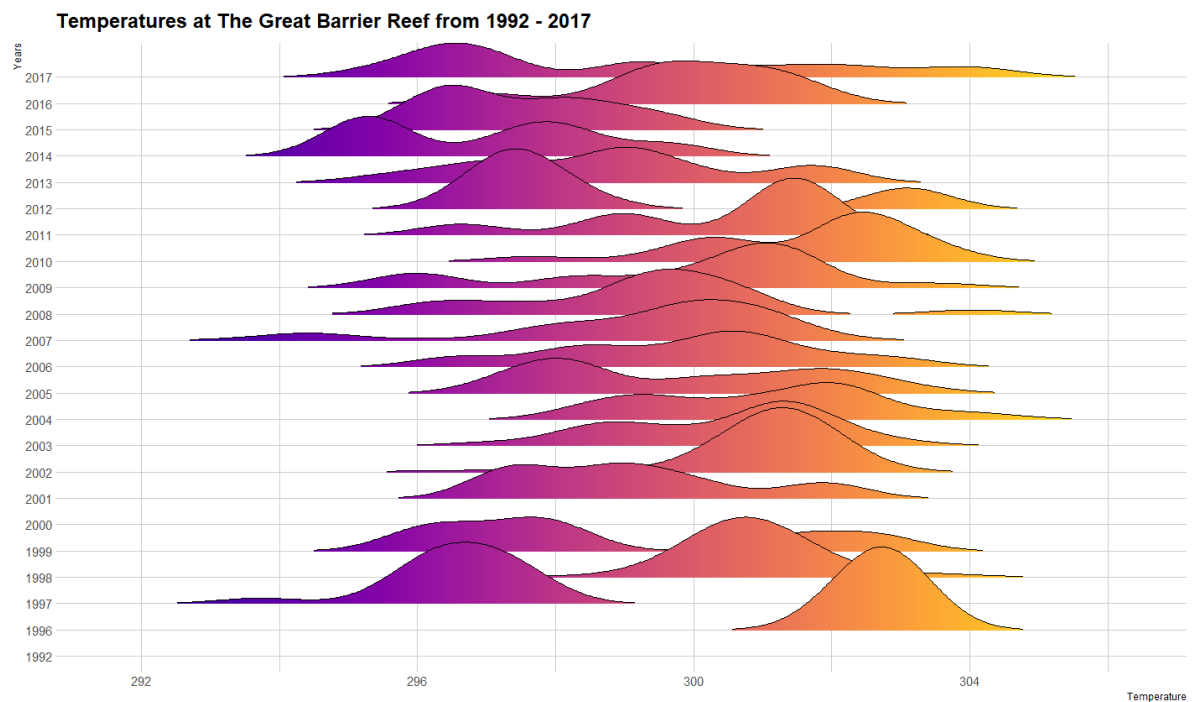


Figure 47

Final

However, given that visualisation 1 shows results across the globe its only appropriate the visualization 2 provides further evidence. As shown below the final version displays average ocean temperature across the globe, it can be deduced from this visualisation the temperatures are becoming more erratic as well as increasing.

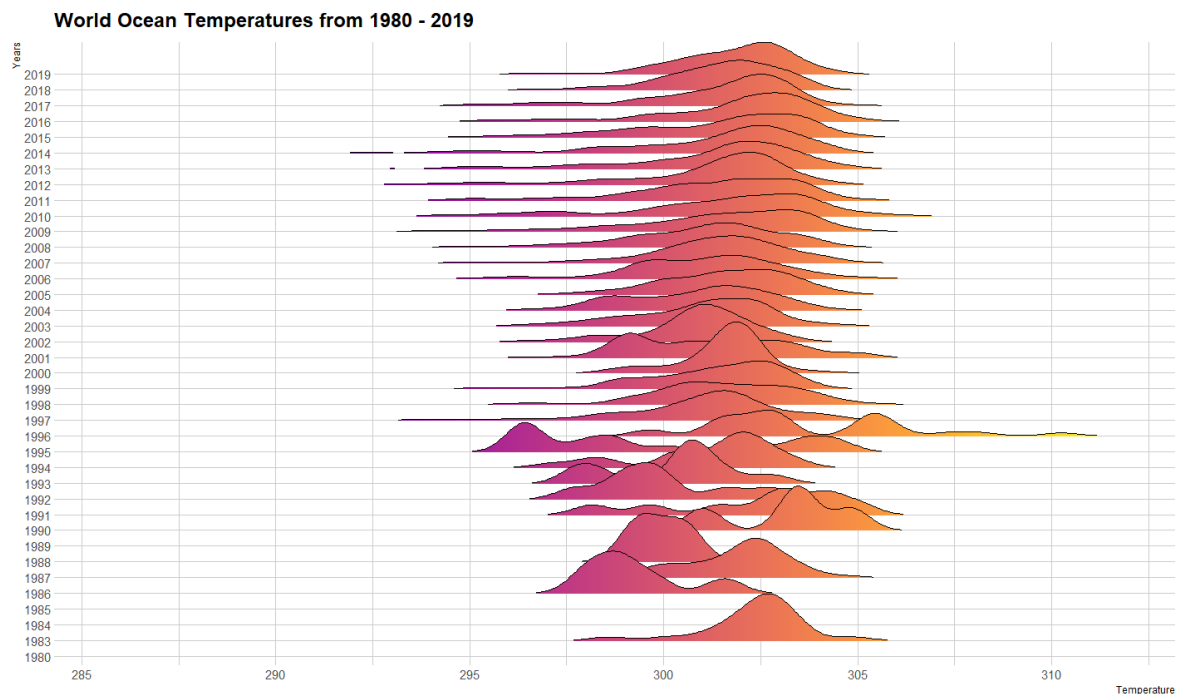


Figure 48

Conclusion

The two visualisations within this report successfully show the main message contained within the data utilized, both show in a clear and understandable manner that coral bleaching has become a common event and is correlated with the slight increase in ocean temperature. This was made achievable by the various methods of data analysis and preparation implemented for better understanding of the vast amount of data contained. In order to build upon what has been achieved more exploration of the available data is required also visualisation one could be explored further with other methods such as a bubble effect.

References

Team SDC (2022), Metadata record for: A global coral-bleaching database (GCBD), 1980–2020, figshare, Available at:

https://springernature.figshare.com/articles/dataset/Metadata_record_for_A_global_coral-bleaching_database_GCBD_1980_2020/16958353/1