

Professional Development and Research Skills
CMM507 Coursework
Group 2: Plastic Pollution in Oceans

ALEXANDER RITCHIE, 1911218@rgu.ac.uk;
GEORGIOS ORFANAKIS, 1903446@rgu.ac.uk;
KAREN JEWELL, 1415410@rgu.ac.uk;
ROSHI SHRESTHA, 1903445@rgu.ac.uk;
STUART WATT, 1501869@rgu.ac.uk

May 10, 2020

1 Introduction

1.1 Problem Statement

Plastic pollution is a serious problem in the marine environment which has major impact on marine and human health. To understand the depth of the problem, it is essential to understand the amount and composition of marine litter. This can help in applying various mitigation strategies. This section will include: **should this talk about the section or the report? Do we need the problem statement???**

- An overview of marine pollution
- Motivation behind this topic
- Aim and objectives of this report

1.2 Overview

Marine pollution is a major global issue which impacts on environment, economy and human health. Although marine pollution is caused by many different materials, plastics consist of 60-80% of the marine litter. **reference?** Plastic is made from synthetic organic polymer derived from polymerisation of monomers extracted from oil **I think plastics is just a oil product, not gas - confirm.**[1] [2] Its lightweight feature and durability make it very suitable to make a range of products we use in our everyday life.[3] [4] These same features are what makes plastic a major component of pollution, due to overuse and non-managed waste disposal systems internationally, contributing to 10% of the waste generated worldwide.[3] Due to its buoyancy, plastic debris can be dispersed over long distances and they can persist for a long time **and never decompose?** Jambeck et al.,[5] reported that in 2010 alone, between 4.8 million to 12.7 million metric tons of plastics entered the ocean. Plastics are everywhere in the marine environment and urgent action is required to mitigate this problem and reduce its harmful impact.[2] [6]

1.3 Motivation

The impact of plastic pollution on marine life has been reviewed extensively. [7] [8] [9] [10] Over 700 marine wildlife species are affected due to entanglement in plastic ropes or materials, and from ingestion of plastics in the ocean.[7] Over 340 species of marine animals were found to be entangled. [8] what's the difference between the 700 and the 340? Over time, plastic disintegrates into microplastics and nanoplastics which are easily consumed by fish and enter the human food chain. Plastics have been found in a third of fish caught in the UK which included popular fish such as cod, haddock, and mackerel. [11] The impact and effects of plastic entering the human food chain are still being studied, but plastic toxicity and the occurrence of microplastics and nanoplastics in the water supply also directly impacts human health, in addition to the contamination of seafood.[6] [12]

Although plastic litter has been a major cause of marine pollution for a while, its seriousness has only been realised recently and reducing plastic pollution has become a global aim. Research on plastic pollution in marine environments has played a big role in efforts to reduce it as well as in raising awareness globally. In order to understand plastic pollution in marine environments and its effects in the long term, it is essential to keep collecting data on patterns of marine debris around the world. Effective monitoring of plastic debris is essential to reducing the abundance of plastic debris everywhere. In addition, monitoring the type, frequency and the source of the litter is also important for marine pollution prevention initiatives. Most monitoring is done by survey, where organisations and volunteers record the types and frequencies of litter observed on the shoreline.[13] Litter is most abundant close to urban areas where beach visitor numbers are high.[14]

1.4 Objectives

The main objectives of this project is outlined as follows:

- To review available literature on marine plastic problems and their impacts
- To look at a suitable dataset to understand the composition of plastic pollution
- To present the derived results and conclusions

2 Research

An integral part of this report required that the authors conducted a literature search to identify how researchers have been trying to monitor coastal pollution and find the problems associated with it. Several studies have reported the abundance of plastic as a coastal litter through scientific survey and citizen science methods. A 12-year observation of coastal debris pollution using citizen science in Taiwan revealed that most debris items found were plastic. [15] 19 categories of debris items were collected during the clean-up events and the five most commonly recorded debris categories were: plastic shopping bags, plastic bottle caps, disposable tablewares, fishing equipment, and plastic drinking straws. There have been many other studies around the world regarding littering of the shores. In a study covering western Japan and the eastern coasts of Russia [16], it found that 55% to 93.4% of items over the Japanese shores were plastic, and the second most abundant item was resin pellet which is also a form of plastic. On the eastern Russian coast, plastic items were also the most abundant, contributing to approximately 55% of all litter, mostly of plastic fragments. The composition of litter was similar in the two countries, although the concentration **proportion?** of plastics was much higher in Japan. [16] Further along the Asian upper east coast, hard plastic and styrofoam were the dominant plastic types found on Korean beaches. On average, hard plastic and styrofoam comprised 32% and 48.5% of the total debris count respectively.

In an older study over the Caribbean region, the most common types of debris found on the Caribbean coast of Panama were plastic and styrofoam, with the plastics being household or consumer related. Styrofoam packing materials were also abundant and may have come from trans-shipment activities of Colon's Free Zone, household waste, or offshore activity [14]. A 2016/2017 annual study of 8 beaches in Tenerife in the Canary Islands also found that plastic was the most abundant litter there. They also reported that there was more accumulated plastic debris in remote beaches compared to the beaches near the city **which contradicts (or if supported by) the GARRITY1993 study "Litter is most abundant close to urban areas where beach visitor numbers are high"**, indicating that more debris were transported by tides. More long term study is required to understand the changes in the results reported over time. [17]

It was observed by the authors of this report, that there are variations in how studies of litter accumulation have been conducted. The variations are present in the time span of the research, the parts of the coast from which litter was collected, as well as in the categorisation of litter, which creates difficulty when researchers want to compare different studies. The plastic pollution problem essentially requires the ability to assess changes in accumulation rates and composition, trends over time and the effectiveness of management systems, which is a hard task without good monitoring methodologies. Although monitoring of marine litter is currently carried out within a number of countries around the world, the methods of survey and monitoring used tend to be very different, preventing comparisons and harmonisation of data across regions or time.

This is why the scientific community has been trying to create some common ground, which has led to initiatives joined by many countries worldwide. One of these initiatives, and probably the most important, is the International Clean Coast (ICC) program which is a new, long term approach to cleaner beaches by using various activities to increase public awareness. [18] This initiative aims to develop a comprehensive litter characterisation scheme that uses both material composition and form. This allows Litter Monitoring Repeated **os LMR a name? should it be encased in ""?** surveys of beaches, sea bed and/or surface waters to determine litter quantities such that information can be compared with baseline data to identify if changes occur over time or in response to management arrangements.

The ICC uses specifically developed categorisations of coast litter, with the most accepted being the Clean Coastal Index (CCI) protocol, which is useful for its simplicity and information provided, allowing comparisons between different times and places. The CCI protocol is different from most others **other....protocols, systems?** having a focus on operational clean-up of beaches as well. The CCI is the recommended tool for evaluation of actual coast cleanliness, measuring plastic debris as an indicator of beach cleanliness, in an easy way **easy for volunteers?** which precludes bias by the assessor. The CCI also proved to be a useful tool for measuring progress and the success of activities in raising awareness among the general public. [19]

A study in Israel followed the CCI protocol and found that plastic was the most ubiquitous **prevalent, common?** beach litter item. An important contribution to this study was the ability to compare its findings to other Mediterranean beaches, showing that plastic might be **is/was?** the dominant pollutant in the region, and non-plastic litter being highly specific to the region and cannot be treated universally. [20] In another study on litter pollution in a region of India, once again following the CCI protocol for the categorisation of litter, found that plastic was the main form of litter at approximately 45% of total litter. Plastic bags topped the index at 33%, followed by food wrappers, then plastic cups, and cigarette/cigar tips amounting to 5.5%. [21] The use of the common protocol in these two studies allows for researchers to compare their findings and create common plastic pollution models, even though the two coasts are continents apart.

Another study conducted in Cadiz, on the other side of the Mediterranean from Israel, found that plastic bottles and containers were the most frequent littered items, followed by plastic bags. This research also pointed out that surveys are heavily affected by clean-ups performed at beaches. [22] Even though this study reaches important conclusions on ways to clean coasts the correct way, it cannot be easily compared, or its conclusions easily applied even with the case of the study in Israel sharing the Mediterranean Sea. **because it is not using the CCI?**

From this review, it is evident that there have been many studies conducted to monitor marine pollution in varying ways. One of the cost effective and easy method **is it not the most cost-effective and easiest?** is the use of citizen science in which the public records any observations of marine litter. As discussed before, observations not following a standard protocol with proper guidance could be uncomparable and rendered ineffective. The Marine Debris Tracker (MDT) initiative **can we mention who created/sponsored it?** was started in 2010 in North America and continues to operate today. The MDT allows anyone to record the marine debris observed, using an application on a mobile phone. The only report using data from this application is a web-based report by Tablada in 2018 [23], where data analysis was performed on 8 years of the data, mainly focused on North America. The study also concluded that plastic was the main type of debris recorded, with cigarettes being the top identified littered item. The subsequent sections of this report will aim to contribute to this body of knowledge, using the worldwide coastal littering dataset from the MDT, with an interest in identifying if plastic is indeed the most abundant litter of the worldwide dataset, which would be in agreement with the various studies discussed in this review. Following which, this report will analyse the distribution of subclasses within the plastics found, and explore if there could be a way to computationally monitor and assess the coastal littering problem, using the established CCI categorisation of litter.

3 Methods

3.1 Dataset Description

The data used in this report was gathered using secondary data collection methods only. The authors did not collect or create any new data using primary methods. The data was downloaded from the Marine Debris Tracker website (www.marinedebris.engr.uga.edu) on [what was the date of download?](#). The dataset is composed of global observations from the start of the program in 2010 to the latest available date in February 2020.

As discussed in the section before, the MDT is a citizen science project where organisations or individuals can record observations of marine debris using a mobile phone application. [23] The user records the observation using a structured form, and chooses the category of debris from a list provided, but there are also multiple available fields to populate including non-mandatory fields, and some allowing free text entry. An example of a non-mandatory field are "Lists" which are user customisable groupings for listing their own collection of records. The table below describes the structure of the data.

Field	Description	Mandatory
ListName	customised groupings of records	Non-mandatory
ListID	ID for ListName	Non-mandatory
ItemName	the category of debris	Mandatory
ItemID	ID for ItemName	Non-mandatory
LogID	Unique ID for the observation	Non-mandatory
Quantity	number of pieces of debris observed	Mandatory
Error radius	radius around the site within the error for reasonable doubt	Mandatory
Latitude	coordinates of the location where the observation was made	Mandatory
Longitude	coordinates of the location where the observation was made	Mandatory
Altitude	coordinates of the location where the observation was made	Mandatory
Location	free-text description of the area the observation was made in	Non-mandatory
Description	free-text description for the observation	Non-mandatory
MaterialDescription	the material the debris was composed of	Mandatory
MaterialID	ID for MaterialDescription	Mandatory
Time	the date and time of observation	Non-mandatory

3.2 Dataset Pre-processing

```
library(tidyverse)
library(purrr)
library(magrittr)
library(treemap)
library(mapdata)
library(viridis)
library(lubridate)
library(imager)
library(xtable)
library(dplyr)

data <- list.files(path = "data/debris/", full.names = TRUE) %>%
  lapply(FUN = read_csv, col_types = "iffiddddcfcif") %>%
  reduce(rbind)
```

From download, the data was cleaned to prepare it for analysis.

First, the datetime information was converted into a date type format and renamed as variable **Time**

```
# Replacing the column for time as a date data type, renaming it "Time"
data$Time <- data$Timestamp %>%
  parse_datetime(format = "%Y%m%d%H%M%S")
data$Timestamp <- NULL
```

On inspection of the data, missing values were observed in the **Location** and **Description** fields, but as these are non-mandatory and free-text fields it is to be expected and all other fields are complete.

```
# Identifying missing values
data %>% select_if(function(x) any(is.na(x))) %>% colnames()

## [1] "Location" "Description"
```

It was also observed that observations were incomplete for years 2010, 2011 and 2020, and so it was decided to retain only complete years of information for fair and standard comparisons. can we do something here to show that 2010, 2011 and 2020 are incomplete? Count of unique months? min max months?

```
# Filter for observations occurring between the years 2012-2019 inclusive
data <- data %>% filter(as.integer(year(Time)) %in% 2012:2019)
```

can we remove this chunk?

```
# explicit missing value for the location factor
#data$Location <- data$Location %>% fct_explicit_na()
```

Assessing the dataset for unique items it can be observed that there are 8 material categories containing 55 different item subcategories and with them 7982 unique descriptions across the 349556 observations.

```
data %>%
  summarise_all(~length(unique(.))) %>%
  pivot_longer(cols = everything(),
               names_to = "Field",
               values_to = "Unique Values") %>%
  arrange(desc(`Unique Values`)) %>%
  xtable()
```

	Field	Unique Values
1	LogID	349556
2	Time	237066
3	Latitude	144820
4	Altitude	133174
5	Longitude	132316
6	Error Radius	16930
7	Description	7982
8	Location	1352
9	Quantity	494
10	ItemID	55
11	ItemName	55
12	Material ID	8
13	Material Description	8
14	ListID	1
15	ListName	1

It is difficult to determine exactly how many locations are observed in this dataset as it is the combination of **Longitude**, **Latitude**, **Altitude** and/or **Location** which determines the global position of the observation.

3.3 Data Quality Issues: Classification

It is worth noting that of the 8 material categories, one is an "Other" category which allows users to categorise an item as such if it is not appropriate for any other option on the list.

```
# Listing of material categories
data %>% select(`Material Description`) %>%
  unique() %>%
  arrange(`Material Description`) %>%
  xtable()
```

	Material Description
1	PAPER & LUMBER
2	PLASTIC
3	METAL
4	CLOTH
5	OTHER ITEMS
6	FISHING GEAR
7	GLASS
8	RUBBER

The debris subcategories were checked for uniqueness by material type and it highlighted that the subcategory "Rubber Gloves" was associated with two material types, Plastic and Rubber, but otherwise a one to many relationship exists between the parent material type and their item subcategories and the integrity of the data remains intact.

```
# Checking for subcategory uniqueness
data %>% select(`Material Description`, ItemName) %>%
  distinct() %$%
  table(ItemName) %>%
  as_tibble() %>%
  filter(n > 1) %>%
  xtable()
```

	ItemName	n
1	Rubber Gloves	2

Further investigation into the categorisation of Rubber Gloves revealed that the majority of Rubber Glove items were classified as plastic rather than rubber. When the observation descriptions were considered however, it revealed that the categorisations may be inaccurate anyway as the last two instances here have "Balloon" in the extra descriptions. This raises the question of why these items were not classified as "Other" materials but wrongly labelled as Rubber Gloves, perhaps reflective of the inconsistencies and unreliability of citizen science data gathering. Yet some data may be better than no data at all. **we should probably change make them consistent or explain why we didn't**

```
# Checking material categorisation of Rubber Gloves
data %>% select(`Material Description`, ItemName, Quantity) %>%
  filter(ItemName == "Rubber Gloves") %>%
  group_by(`Material Description`) %>%
  summarise(Quantity = sum(Quantity)) %>%
  xtable()
```


	Material Description	Quantity
1	PLASTIC	2092.00
2	RUBBER	89.00

```
# Checking the descriptions of Rubber Glove items
data %>% select(`Material Description`, ItemName, Description) %>%
  filter(ItemName == "Rubber Gloves", !is.na(Description)) %>%
  distinct() %>%
  xtable()
```

	Material Description	ItemName	Description
1	PLASTIC	Rubber Gloves	thermal
2	PLASTIC	Rubber Gloves	Near water
3	PLASTIC	Rubber Gloves	Taste of Omaha Cleanup
4	PLASTIC	Rubber Gloves	2 diff kinds
5	PLASTIC	Rubber Gloves	undefined
6	PLASTIC	Rubber Gloves	Latex
7	PLASTIC	Rubber Gloves	Hose
8	PLASTIC	Rubber Gloves	Cap
9	PLASTIC	Rubber Gloves	Vial
10	PLASTIC	Rubber Gloves	Gloves
11	PLASTIC	Rubber Gloves	Black intact
12	PLASTIC	Rubber Gloves	Clear intact
13	PLASTIC	Rubber Gloves	White glove fragmented
14	PLASTIC	Rubber Gloves	Fishing glove
15	PLASTIC	Rubber Gloves	Rubbery fragments
16	PLASTIC	Rubber Gloves	Rubber band and fishing bait
17	PLASTIC	Rubber Gloves	Balloon

Potential section: multiple instances of missclassified items. Where their descriptions appear to not match their material categorisation

where MATERIALS!=Plastic yet have the term "plastic" in the description. could further expand this to descriptions which have any of the material terms in them, but is not its own material.

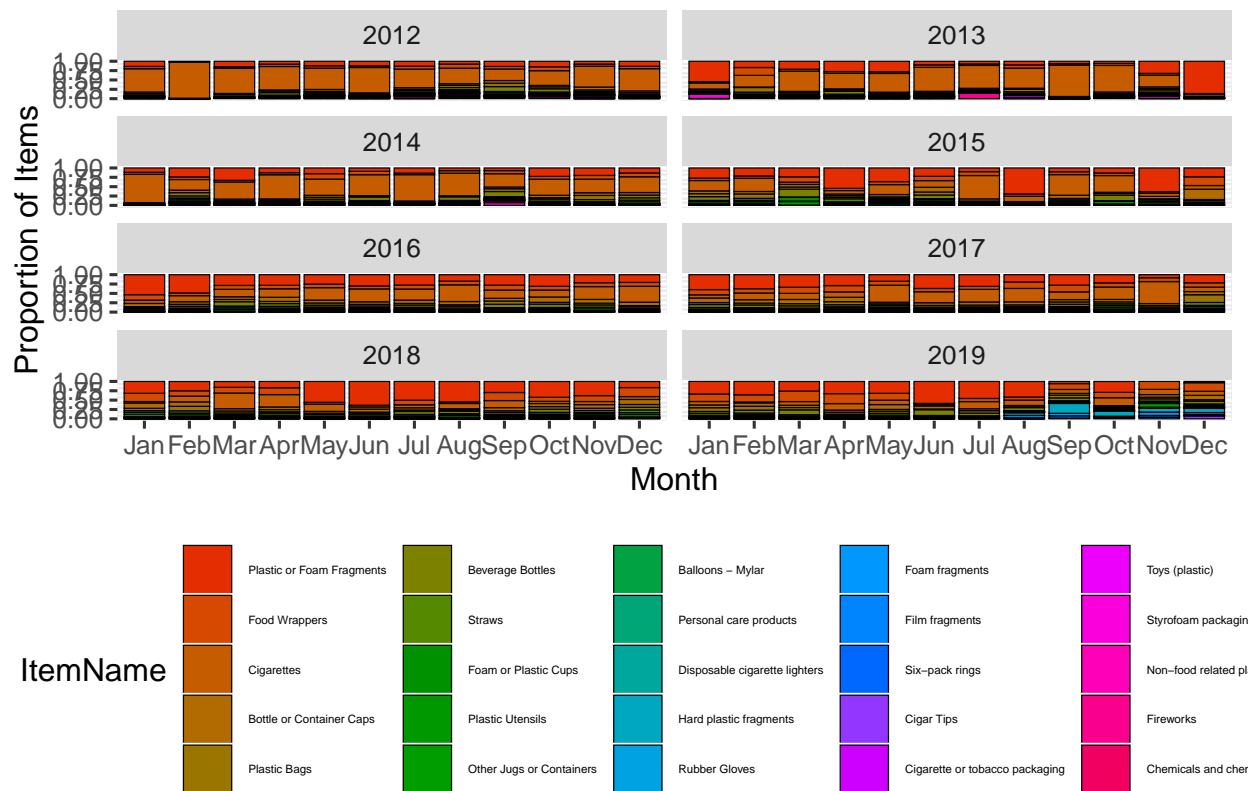
3.4 Recategorisation

After the issues with the dataset that were identified in the section above, it was decided that it would be best to transform the dataset in the following ways:

- reclassified some labels because variation was too high (there were too many labels)
- The values of the missing data were removed.
- It was decided that subsets that were not needed were removed while retaining the necessary subsets.

```
plastic_ordered <- data %>%  
  filter(`Material Description` == "PLASTIC") %>%  
  select(ItemName, Quantity) %>%  
  group_by(ItemName) %>%  
  summarise(Total = sum(Quantity)) %>%  
  arrange(desc(Total))
```

```
data %>%
  filter(`Material Description` == "PLASTIC") %>%
  mutate(month = month(Time, label = TRUE),
         year = as.integer(year(Time)),
         ItemName = fct_infreq(ItemName)) %>%
  group_by(month, year, ItemName) %>%
  summarise(`Total Quantity` = sum(Quantity)) %>%
  ggplot(aes(x = month, y = `Total Quantity`, fill = ItemName)) +
  geom_col(colour = "black", size = 0.2, position = "fill") +
  facet_wrap(~year, nrow = 4) +
  scale_fill_hue(l=50, c=150) +
  xlab("Month") +
  ylab("Proportion of Items") +
  theme(legend.position="bottom",
        legend.text=element_text(size=4))
```



```
#scale_fill_viridis_d(option = "magma")
#ggsave("plots/pastic_debris_plot.png", width = 40, height = 20, units = "cm")
```

Figure 1: Debris by categorisation

```
# all cigarette related waste: 1, 4, 6, 22
# Food related waste: 3, 2,7,9,10, 17, 23, 11
# Non food related waste: 8, 14, 15, 16, 18, 19, 21, 20
# Plastic bags and Styrofoam packaging:12, 13
```

```

# Fragments: 5, 23, 24,25

recategorise <- function(x){
  out = ""
  if(x %in% c(1,4,6,22)){out = "Cigarette related waste"}
  if(x %in% c(2,3,7,9,10,17,23,11)) out = "Food related waste"
  if(x %in% c(8,14,15,16,18,19,21,20)) out = "Other"
  if(x %in% c(12,13)) out = "Plastic bags and Styrofoam packaging"
  if(x %in% c(5,23,24,25)) out = "Fragments"
  if(out == "") stop(paste("Error in recategorise:", x))
  return(out)
}

plastic_types <- data %>%
  filter(`Material Description` == "PLASTIC") %>%
  select(ItemName, ItemID) %>%
  distinct() %>%
  mutate(label = 1:n()) %>%
  mutate(category = purrr::map(label, recategorise)) %>%
  mutate(category = as_factor(as.character(category))) %>%
  select(ItemID, category)

plastic <- data %>%
  filter(`Material Description` == "PLASTIC") %>%
  full_join(plastic_types, by = "ItemID")

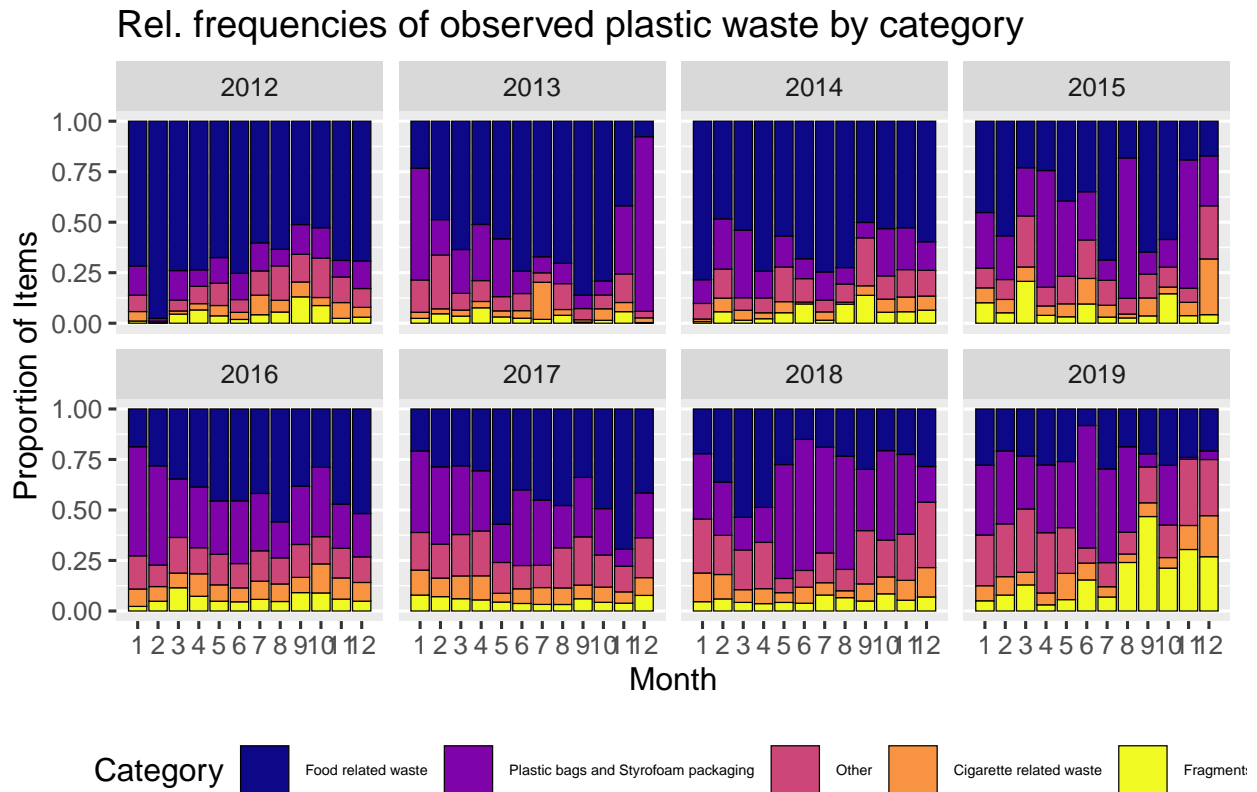
ordered_levels <- plastic %>%
  group_by(category) %>%
  summarise(totObs = sum(Quantity)) %>%
  ungroup() %>%
  arrange(desc(totObs)) %>%
  select(category) %$%
  category
plastic$category <- factor(plastic$category, levels = ordered_levels)
rm(ordered_levels)

```

```

plastic %>%
  mutate(month = month(Time, label = FALSE),
         year = as.integer(year(Time))) %>%
  group_by(month, year, category) %>%
  summarise(`Total Quantity` = sum(Quantity)) %>%
  ggplot(aes(x = month, y = `Total Quantity`, fill = category)) +
  geom_col(colour = "black", size = 0.2, position = "fill") +
  facet_wrap(~year, nrow = 2) +
  scale_fill_viridis(discrete = TRUE, option = "plasma") +
  xlab("Month") +
  ylab("Proportion of Items") +
  ggtitle("Rel. frequencies of observed plastic waste by category") +
  scale_x_continuous(breaks = 1:12) +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        legend.position = "bottom",
        legend.text=element_text(size=5)) +
  guides(fill=guide_legend(title="Category"))

```



```

#ggsave("plots/pastic_debris_plot_recategorised.png", width = 40, height = 20, units = "cm")

```

Figure 2: Recategorisation by year. Colour scale ordered by ranking of total observed quantity.

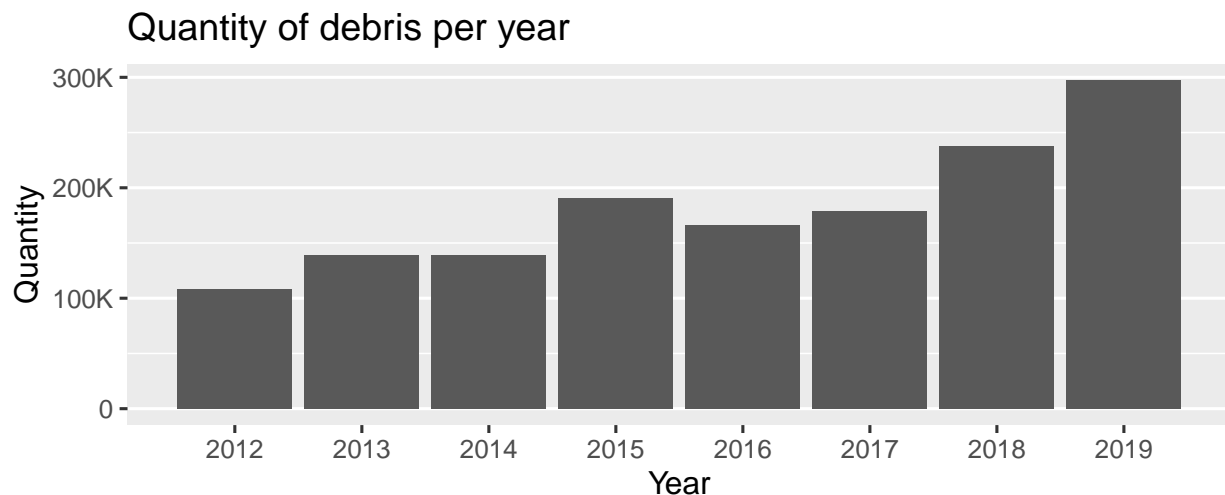
4 Exploration

4.1 Proportion Trends

After cleaning and recategorisation of the data, it was analysed to determine how pollutant proportions change over time. 1. using the CCI categories, 2. using the original subcategories.

Per the data, we see an increase in debris observed over time. Caution must be exercised in interpreting this information, as this increase may be reflective indeed of increased pollution levels, but it could also be because the citizen science project has gained popularity and more organisations and individual observations.

```
#Linechart quantity of debris per year
data %>%
  mutate(year = year(Time)) %>%
  group_by(year) %>%
  summarise(quan = sum(Quantity)) %>%
  ggplot(aes(x = year, y = quan)) +
    geom_col() +
    scale_x_continuous(breaks = 2012:2019) +
    xlab("Year") +
    ylab("Quantity") +
    ggtitle("Quantity of debris per year") +
    theme(panel.grid.major.x = element_blank(),
          panel.grid.minor.x = element_blank()) +
    scale_y_continuous(labels = scales::label_number_si(accuracy = 1))
```



```
#ggsave("plots/observations.png")
```

Figure 3: Trend of debris observed

insert: count of unique observations, maybe grouped by average number of observation entries per day? - we want to try and identify increase in user activity

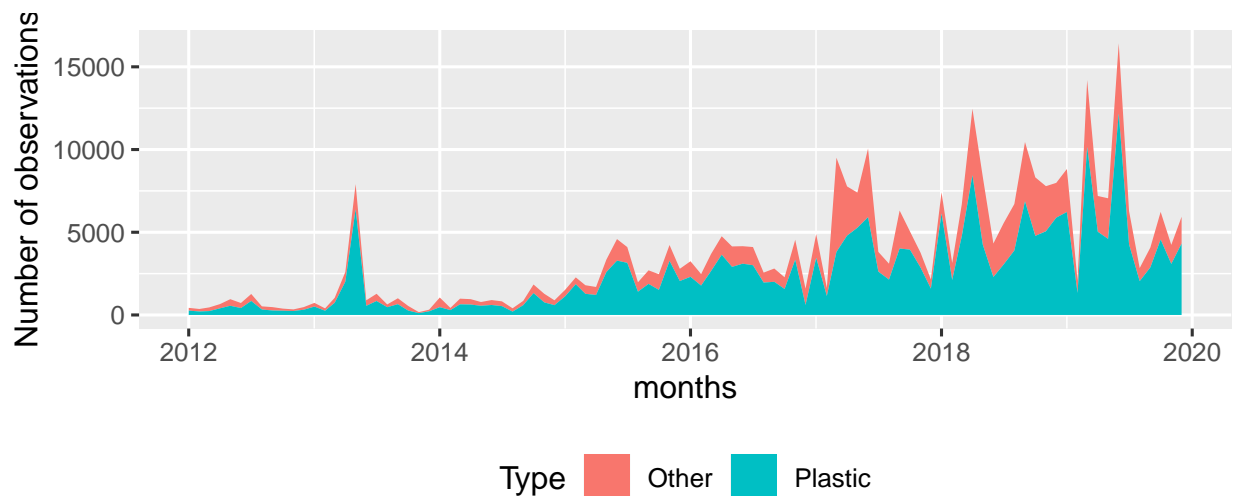
As we see in the chart above, an increase in the number of average observations per day lends support to the theory that the increase in observations is factored by increase in citizen science monitoring activity. That's not to say pollution hasn't also increased in this period, but it would be inaccurate to not account for the increase in monitoring activity.

insert: proportion changes of MATERIAL TYPES over years. Line or column. We want to prove that plastic is the dominant pollutant and possibly the only increasing one

Analysis of the proportions of pollutant materials suggests that not only is plastic the dominant marine pollutant material, its proportion of total debris is increasingly observed year on year. This could be due to an increase in new plastic pollution entering marine environments from source, not helped by the fact it is durable and doesn't decompose, as discussed in the introduction of this report, or that the other pollutant materials are decreasing. Of all the pollutant materials listed here - plastic is the most longlife **can we support this statement?** This analysis therefore is in agreement with the studies discussed in section 2 of this report that plastic is indeed the biggest problem in marine pollution.

```
#Histogram of observations: Total v Plastic
data %>%
```

```
  mutate(Type = if_else(`Material Description` == "PLASTIC", "Plastic", "Other"),
         months = floor_date(Time, 'month')) %>%
  group_by(months, Type) %>%
  summarize(`Number of observations` = n()) %>%
  ggplot(aes(x = months, y = `Number of observations`)) +
    geom_area(aes(fill = Type)) +
    theme(legend.position = "bottom")
```



```
# ggplot() +
# geom_histogram(aes(x = Time))
```

Figure 4: Observations of plastic debris v all debris

minor edit: can we take away the area fill but keep the coloured lines?

insert: line chart of just the % plastic over years

When we consider the charts above, we can determine that it is not only the plastic proportion of total pollutant materials that increases, but it's total observation counts too (not quantities). This means that even as more debris are observed, plastics remain a consistently common feature of the observed debris.

In this chart we also see peaks and troughs in the number of observations - this is indicative of an increase in user activity. Seeing the peak in 2013 - that's not to say pollution suddenly increased in one period and disappeared, but monitoring activity spiked in that period. **do we know why?** We see more of these spikes in 2017, 2018 and 2019 but generally it is an increasing trend in monitoring activity.

insert: proportion changes of RECATEGORISED plastic over years. stacked lines or columns When considering the subcategories within plastic we see..... can we link this to decreasing cigarettes?
 insert: table of % change in ItemNames 2012 to 2019. return top 10 or 15 ABSOLUTE value changes. If you really want to do this properly you'll take the averaged change in % year on year and compare it to the end-to-end % change.

An interesting observation was that as Old pollutants fall away (cigarette butts) but new ones are introduced. Cigarette butts proportions and raw counts decrease over time which is possibly indicative of less people smoking, or moving to vaping [source this, or don't bother?](#)

4.2 Distribution of observed debris:

MaterialQuantities

```
data %>% select(Quantity,Description,`Material Description`) %>%
  group_by(`Material Description`) %>%
  summarise(Quantity = sum(Quantity)) %>%

ggplot(aes(x = reorder(`Material Description`, Quantity), y = Quantity)) +
  geom_col() +
  ylab("Total recorded quantity") +
  xlab("Material class") +
  coord_flip()
```

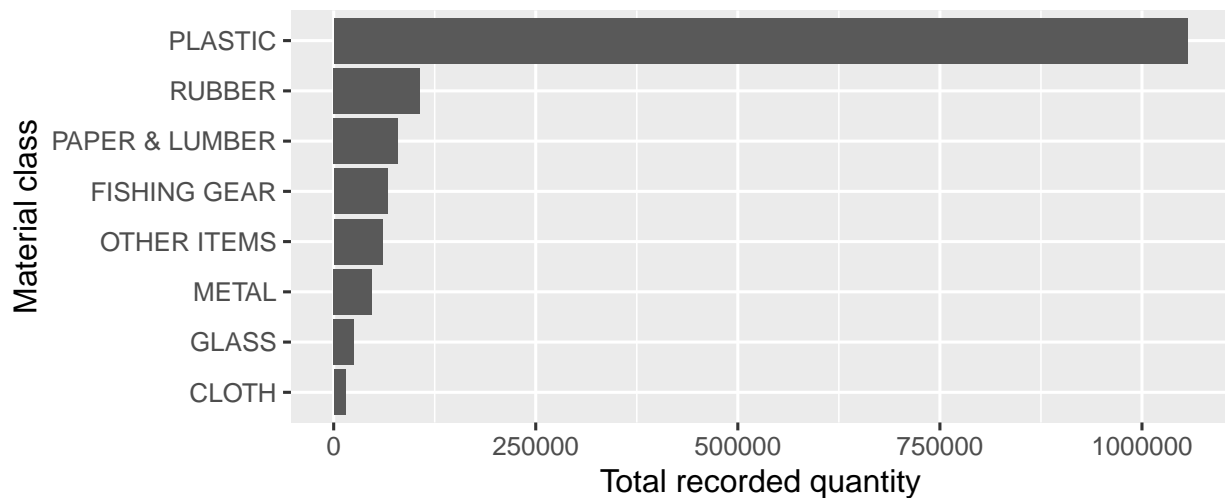
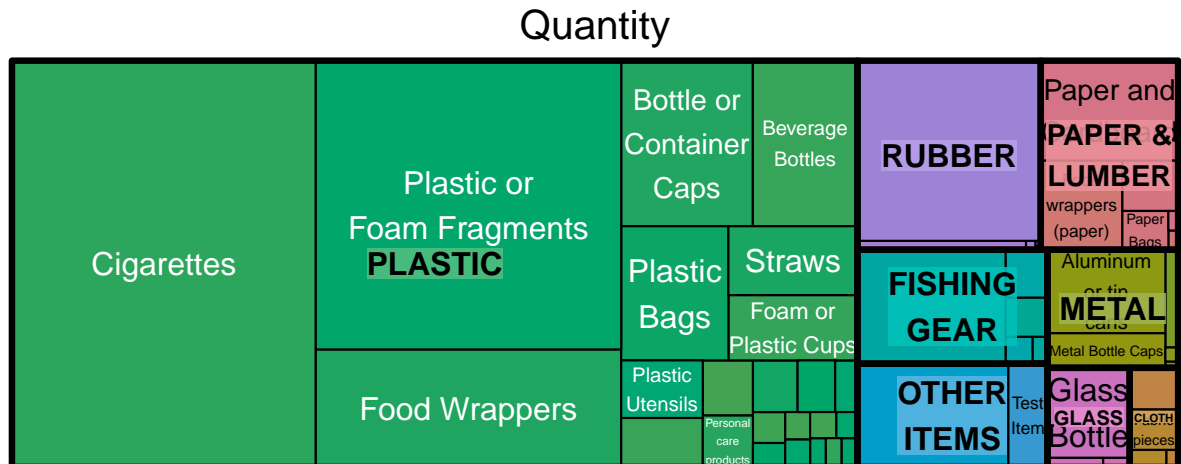


Figure 5: Material Quantities

So the most populated material class is Plastic. Note that this does not necessarily mean that plastic is the largest quantity of debris, just that the individual number of items categorised is largest.

A tree map of material quantities:


```
#treemap of debris categories
#png("plots/treemap.png")
data %>%
  select(`Material Description`, ItemName, Quantity) %>%
  group_by(`Material Description`, ItemName) %>%
  summarise(Quantity = sum(Quantity)) %>%
  treemap(index = c("Material Description", "ItemName"),
          vSize = "Quantity", draw = TRUE) -> tm
```



```
#tm
#dev.off()
#save.image(file = "plots/treemap.png")
```

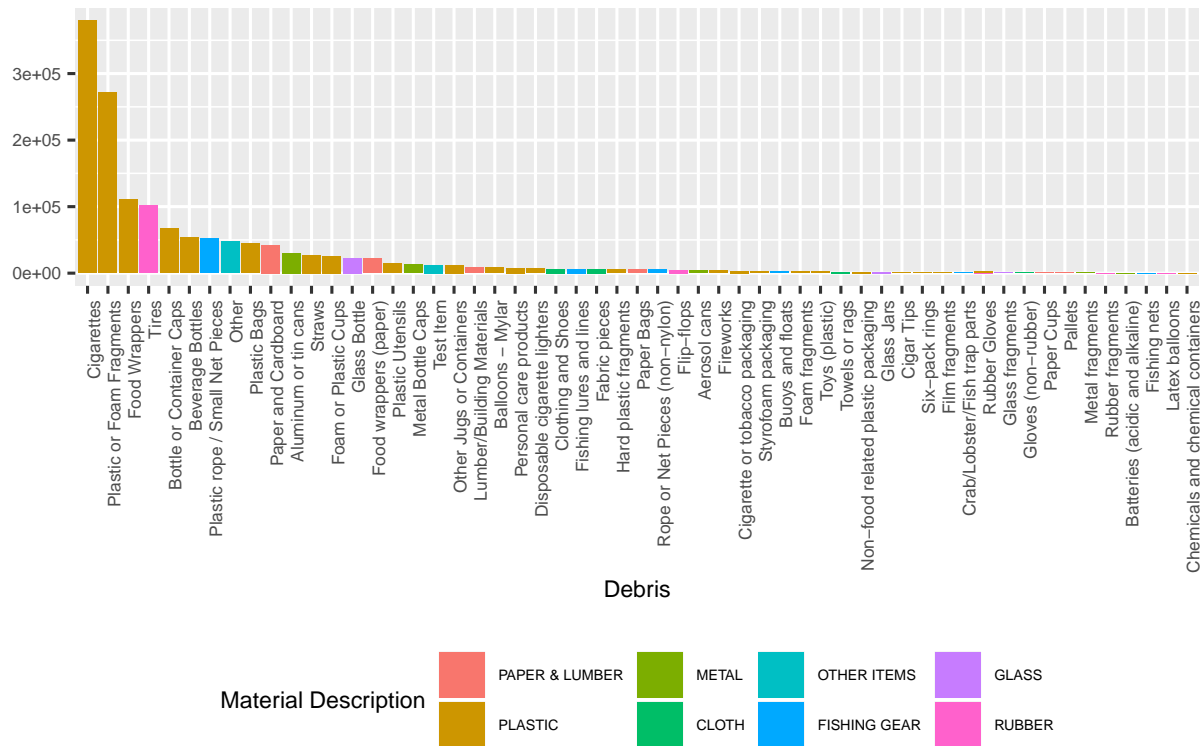
Figure 6: Debris categorisation

Cigarettes are the most common item recorded as seen in. Perhaps some of the debris is not actually from the sea, but rather from people littering by the coastline? Does debris littered on the coastline end up in the oceans?

This is a great chart, but not the best to support the statement that cigarettes is most popular - a column or bar chart here will be much better (area charts are not as effective as charts you can level-compare), potentially use proportions or data labels to further drive the point that it IS the largest. Treemap suggest moving back into pre-processing section.

```
#bar of debris categories
data %>%
  select(`Material Description`, ItemName, Quantity) %>%
  group_by(`Material Description`, ItemName) %>%
  summarise(Quantity = sum(Quantity)) %>%
  ggplot(aes(x=reorder(`ItemName`, -Quantity), y=Quantity, fill=`Material Description`)) +
  geom_bar(stat="identity") +
  ggtitle("Debris Categorisation") +
  xlab("Debris") +
  ylab("") +
  #coord_flip() +
  theme(text = element_text(size=8),
        axis.text.x=element_text(angle=90, hjust=1),
        plot.title = element_text(size=10),
        legend.text=element_text(size=5),
        legend.position = "bottom")
```

Debris Categorisation



```
#treemap(index = c("Material Description", "ItemName"),
#         vSize = "Quantity", draw = TRUE) -> tm
```

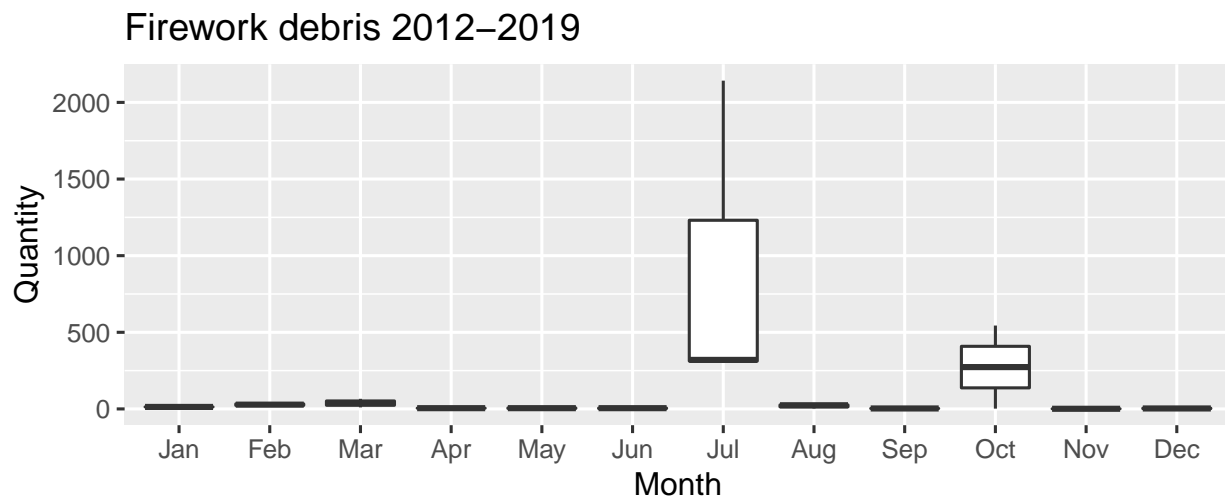
Figure 7: Debris categorisation

maybe make this top 10 or 15 items.

4.3 Event-Driven Pollution

Fireworks found in July and North-America only: possibly 4th July celebrations
4th July and Firework link? (Karen's Idea)

```
#Boxplot of fireworks distribution by month (across all years)
data %>%
  filter(`Material Description` == "PLASTIC",
         ItemName %in% c("Fireworks"),
         year(Time) >= 2012, year(Time) <= 2019) %>%
  mutate(month = month(Time, label = TRUE),
         year = as.integer(year(Time))) %>%
  group_by(month, year) %>%
  summarise(quantity = sum(Quantity)) %>%
  ggplot() +
  geom_boxplot(aes(x = month, y = quantity)) +
  xlab("Month") +
  ylab("Quantity") +
  ggtitle("Firework debris 2012-2019")
```



```
#ggsave("plots/fireworks.png")
```

Figure 8: Boxplot of fireworks distribution by month, across all years

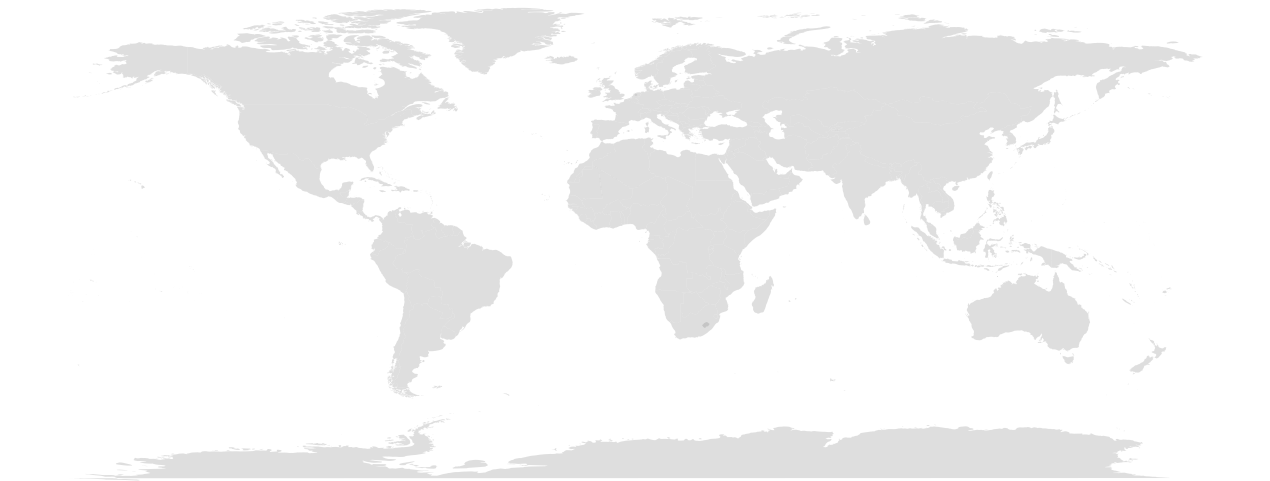
4.4 Location-Driven Pollution

Rubber found in Indoneasia only: possibly a recording bias.

Certain classes are found in certain regions only: not because they don't exist elsewhere but because of recording bias focus in those areas

We have locational data, so lets check for any geographical observation bias.

```
world <- map_data("world")
data %>%
  select(Latitude, Longitude, Quantity, Location, `Material Description`) %>%
  ggplot() +
    geom_polygon(data = map_data("world"), aes(x = long, y = lat, group = group), fill = "grey", alpha = 0.5) +
    geom_hex(aes(x = Longitude, y = Latitude), bins = 50) +
    scale_fill_viridis(trans = "log", breaks = c(5, 50, 500, 5000, 50000)) +
    theme_void() +
    guides(fill=guide_legend(title="Observations"))
```



```
#ggsave("plots/map.png", width = 20, height = 10, units = "cm")
```

We need to know how reliable the location data is. I'm going to filter for "united kingdom" in the location field and plot the raw coordinates.

Figure 9: Longitude and Latitude discrepancies

Questions

Distribution of plastic by location.

Are the distributions of plastic fairly constant for the locations with the most observations?

```
#columnchart of debris locations
topLocations <- data %>%
  group_by(Location) %>%
  summarise(sumQuantity = sum(Quantity)) %>%
  arrange(desc(sumQuantity)) %>%
  top_n(5, sumQuantity)

data %>%
  filter(Location %in% topLocations$Location) %>%
  group_by(Location, `Material Description`) %>%
  summarise(sumQuantity = sum(Quantity)) %>%
  arrange(desc(sumQuantity)) %>%
  ggplot(aes(x = `Material Description`, y = sumQuantity, fill = Location)) +
  geom_col() + theme(legend.position = 'top')
```

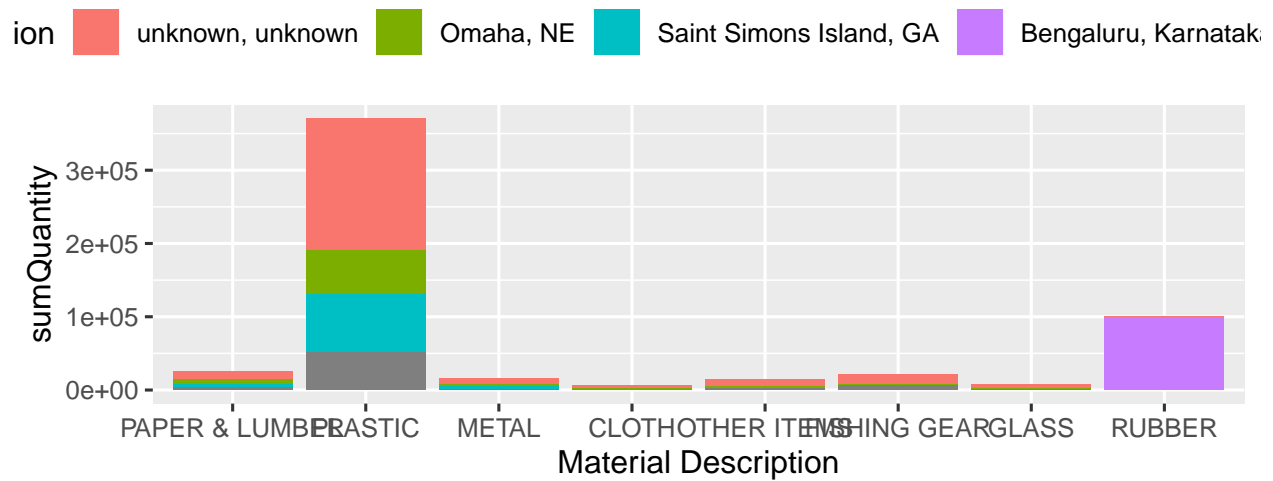


Figure 10: Debris by location

We see that the Location "unknown" has the most plastic... note that this is distinct from "(Missing)", which was our original NA values. Maybe we should merge these.

4.5 Item Pairing

(e.g. are 6-pack beer rings observed at the same time as fireworks?) are we going to explore this one?

5 Predictive Modelling

Given the variability of plastic pollution trends given event-driven and location-driven pollution as explored earlier in this report, the authors of this report built a model to predict the proportion of plastics given Month and Location. This would give more accurate predictions as opposed to a simple linear model accomodating such time factored

5.1 Description of Model

```
plasticN <- plastic %>%
  mutate(month = month(Time, label = FALSE), year = as.integer(year(Time))) %>%
  filter(year > 2010) %>%
  group_by(year, category, month) %>%
  summarise(`Total Quantity` = sum(Quantity))

####

df11N <- plasticN %>%
  filter(year == 2011) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))

df12N <- plasticN %>%
  filter(year == 2012) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))

df13N <- plasticN %>%
  filter(year == 2013) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))

df14N <- plasticN %>%
  filter(year == 2014) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))

df15N <- plasticN %>%
  filter(year == 2015) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))

df16N <- plasticN %>%
  filter(year == 2016) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))

df17N <- plasticN %>%
  filter(year == 2017) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))
```

```
df18N <- plasticN %>%
  filter(year == 2018) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))

df19N <- plasticN %>%
  filter(year == 2019) %>%
  group_by(year, month) %>%
  mutate(freq = `Total Quantity` / sum(`Total Quantity`))

dfTotN <- rbind(df11N, df12N, df13N, df14N, df15N, df16N, df17N, df18N, df19N)
```

```
# plot for observing the data
(time_plotfr2N <- ggplot(dfTotN, aes(x = year, y = freq, color=category, fill = category)) +
  geom_smooth(method="lm", level=0.95) +
  theme_bw() +
  xlab("Years") +
  ylab("relative frequency") +
  ggtitle("portion of plastic") +
  expand_limits(y=0) +
  scale_y_continuous() +
  scale_x_continuous()+
  theme(legend.position="bottom")+
  theme(legend.text = element_text(size=5, face="bold")))
```

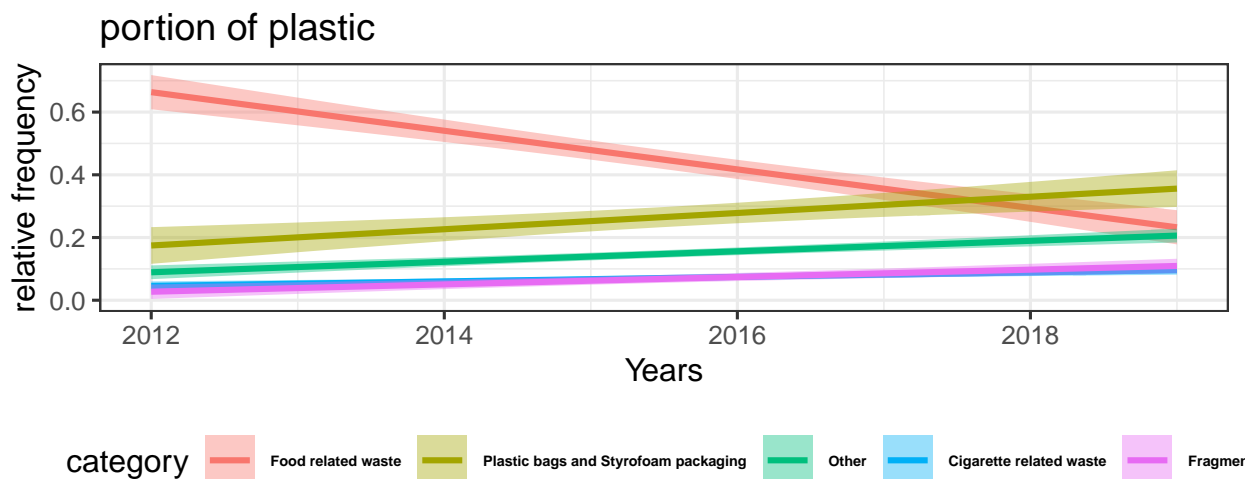


Figure 11: Some sort of caption here

We see here a graphical representation of the relative frequency of the 5 different categories of plastic debris over the years. The hint we are getting here is that "Cigarette related waste", "Food related waste", "Fragments" seem to experience some change, whereas "Other" and "Plastic bags and styrofoam packaging" seem to remain steady. We go on and create a model and test it on untrained data to see if what is the actual case.


```

#1/4/20
### MODELING with new categorisation

# create train and test set
n <- nrow(dfTotN) # Number of observations
ntrain <- round(n*0.75) # 75% for training set
set.seed(314) # Set seed for reproducible results
tindex <- sample(n, ntrain) # Create a random index
train_dfTotN <- dfTotN[tindex,] # Create training set
test_dfTotN <- dfTotN[-tindex,]

# modelling for category "Cigarette related waste"

train_Cigrel <- train_dfTotN %>%
  filter(category=="Cigarette related waste") %>%
  group_by(year)

test_Cigrel <- test_dfTotN %>%
  filter(category=="Cigarette related waste") %>%
  group_by(year)

set.seed(1234)
train_Cigrel.modelN <- lm(freq ~ year, data = train_Cigrel)
summary(train_Cigrel.modelN)

##
## Call:
## lm(formula = freq ~ year, data = train_Cigrel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06059 -0.02546 -0.00721  0.01012  0.20490
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.866832   4.775446  -3.323   0.00146 **
## year          0.007910   0.002369   3.338   0.00139 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04331 on 66 degrees of freedom
## Multiple R-squared:  0.1444, Adjusted R-squared:  0.1315
## F-statistic: 11.14 on 1 and 66 DF,  p-value: 0.00139

print("PREDICTION")

## [1] "PREDICTION"

pred_Cigrel <- predict(train_Cigrel.modelN, test_Cigrel)
summary(pred_Cigrel)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.04720 0.06301 0.07092 0.07714 0.09465 0.10256

actuals_predsCigrel <- data.frame(cbind(actuals=test_Cigrel$freq, predicted=pred_Cigrel))
head(actuals_predsCigrel)

##      actuals predicted

```

```
## 1 0.05109361 0.04719559
## 2 0.05814490 0.04719559
## 3 0.03971631 0.04719559
## 4 0.03750621 0.05510514
## 5 0.01202405 0.05510514
## 6 0.05699357 0.05510514

correlation_accuracy <- cor(actuals_predsCigrel)
min_max_accuracy <- mean(apply(actuals_predsCigrel, 1, min) / apply(actuals_predsCigrel, 1, max))

mape <- mean(abs((actuals_predsCigrel$predicted - actuals_predsCigrel$actuals))/actuals_predsCigrel$actuals)

correlation_accuracy

##           actuals predicted
## actuals      1.0000000 0.5444291
## predicted 0.5444291  1.0000000

min_max_accuracy

## [1] 0.7259542

mape

## [1] 0.7522119
```

```
# modelling for category "Food related waste"

train_Foodrel <- train_dfTotN %>%
  filter(category=="Food related waste") %>%
  group_by(year)

test_Foodrel <- test_dfTotN %>%
  filter(category=="Food related waste") %>%
  group_by(year)

set.seed(1234)
train_Foodrel.modelN <- lm(freq ~ year, data = train_Foodrel)
summary(train_Foodrel.modelN)

##
## Call:
## lm(formula = freq ~ year, data = train_Foodrel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50287 -0.07168 -0.00561  0.10218  0.34829
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 118.166588  15.395217   7.676 5.31e-11 ***
## year        -0.058414   0.007638  -7.647 6.00e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1541 on 74 degrees of freedom
## Multiple R-squared:  0.4414, Adjusted R-squared:  0.4339
## F-statistic: 58.48 on 1 and 74 DF, p-value: 5.998e-11
```

```

print("PREDICTION")

## [1] "PREDICTION"

pred_Foodrel <- predict(train_Foodrel.modelN, test_Foodrel)
summary(pred_Foodrel)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.2292  0.3314  0.4921  0.4395  0.5213  0.6381

actuals_predsFoodrel <- data.frame(cbind(actuals=test_Foodrel$freq, predicted=pred_Foodrel))
head(actuals_predsFoodrel)

##      actuals predicteds
## 1 0.9761681  0.6381043
## 2 0.6351528  0.5796905
## 3 0.5822665  0.5796905
## 4 0.7911040  0.5796905
## 5 0.4837153  0.5212767
## 6 0.5395366  0.5212767

correlation_accuracy <- cor(actuals_predsFoodrel)
min_max_accuracy <- mean(apply(actuals_predsFoodrel, 1, min) / apply(actuals_predsFoodrel, 1, max))
# => 53.73%, min_max accuracy
mape <- mean(abs((actuals_predsFoodrel$predicted - actuals_predsFoodrel$actuals))/actuals_predsFoodrel$actuals)

correlation_accuracy

##              actuals predicteds
## actuals      1.0000000 0.8537377
## predicteds 0.8537377 1.0000000

min_max_accuracy

## [1] 0.8523448

mape

## [1] 0.1546258

# modelling for category "Other"
train_Other <- train_dfTotN %>%
  filter(category=="Other") %>%
  group_by(year)

test_Other <- test_dfTotN %>%
  filter(category=="Other") %>%
  group_by(year)

set.seed(1234)
train_Other.modelN <- lm(freq ~ year, data = train_Other)
summary(train_Other.modelN)

##
## Call:
## lm(formula = freq ~ year, data = train_Other)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.131735 -0.041328 -0.005655  0.035206  0.134108

```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -33.991092   6.112772  -5.561 4.73e-07 ***
## year         0.016938   0.003033   5.585 4.30e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06088 on 69 degrees of freedom
## Multiple R-squared:  0.3113, Adjusted R-squared:  0.3013
## F-statistic: 31.19 on 1 and 69 DF, p-value: 4.296e-07

print("PREDICTION")

## [1] "PREDICTION"

pred_Other <- predict(train_Other.modelN, test_Other)
summary(pred_Other)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.08807 0.12195 0.13889 0.14228 0.17276 0.20664

actuals_predsOther <- data.frame(cbind(actuals=test_Other$freq, predicted=pred_Other))
head(actuals_predsOther)

##      actuals predicteds
## 1 0.05367366 0.08807451
## 2 0.13682008 0.08807451
## 3 0.26636312 0.10501246
## 4 0.10336403 0.10501246
## 5 0.05515197 0.10501246
## 6 0.14156627 0.10501246

correlation_accuracy <- cor(actuals_predsOther) # 5.31%
min_max_accuracy <- mean(apply(actuals_predsOther, 1, min) / apply(actuals_predsOther, 1, max))
mape <- mean(abs((actuals_predsOther$predicted - actuals_predsOther$actuals))/actuals_predsOther$actuals)

correlation_accuracy

##      actuals predicteds
## actuals      1.0000000 0.4983376
## predicteds 0.4983376 1.0000000

min_max_accuracy

## [1] 0.7441029

mape

## [1] 0.354683

# modelling for category "Plastic bags and Styrofoam packaging"
train_Plbag <- train_dfTotN %>%
  filter(category=="Plastic bags and Styrofoam packaging") %>%
  group_by(year)

test_Plbag <- test_dfTotN %>%
  filter(category=="Plastic bags and Styrofoam packaging") %>%
  group_by(year)

set.seed(1234)
train_Plbag.modelN <- lm(freq ~ year, data = train_Plbag)
summary(train_Plbag.modelN)
```

```

##
## Call:
## lm(formula = freq ~ year, data = train_Plbag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31838 -0.09726 -0.02112  0.05326  0.66454
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -53.792345  16.802490  -3.201  0.00204 **
## year          0.026822   0.008336   3.217  0.00194 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1608 on 72 degrees of freedom
## Multiple R-squared:  0.1257, Adjusted R-squared:  0.1136
## F-statistic: 10.35 on 1 and 72 DF,  p-value: 0.001939

print("PREDICTION")

## [1] "PREDICTION"

pred_Plbag <- predict(train_Plbag.modelN, test_Plbag)
summary(pred_Plbag)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.1732  0.2000  0.2536  0.2609  0.3341  0.3609

actuals_predsPlbag <- data.frame(cbind(actuals=test_Plbag$freq, predicted=pred_Plbag))
head(actuals_predsPlbag)

##      actuals predicteds
## 1 0.12998318 0.1731621
## 2 0.14728033 0.1731621
## 3 0.08181818 0.1731621
## 4 0.13607595 0.1731621
## 5 0.28635887 0.1999839
## 6 0.06939979 0.1999839

correlation_accuracy <- cor(actuals_predsPlbag) # 5.31%
min_max_accuracy <- mean(apply(actuals_predsPlbag, 1, min) / apply(actuals_predsPlbag, 1, max))
mape <- mean(abs((actuals_predsPlbag$predicted - actuals_predsPlbag$actuals))/actuals_predsPlbag$actuals)

correlation_accuracy

##      actuals predicteds
## actuals      1.0000000 0.3534861
## predicteds 0.3534861  1.0000000

min_max_accuracy

## [1] 0.6623657

mape

## [1] 2.273597

```

```

# modelling for category "Fragments"
train_Frag <- train_dfTotN %>%
  filter(category=="Fragments") %>%
  group_by(year)

test_Frag <- test_dfTotN %>%
  filter(category=="Fragments") %>%
  group_by(year)

set.seed(1234)
train_Frag.modelN <- lm(freq ~ year, data = train_Frag)
summary(train_Frag.modelN)

##
## Call:
## lm(formula = freq ~ year, data = train_Frag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07328 -0.03075 -0.01469  0.01405  0.36377
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -21.074627   6.678746  -3.155  0.00237 **
## year          0.010490   0.003314   3.165  0.00230 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06143 on 69 degrees of freedom
## Multiple R-squared:  0.1268, Adjusted R-squared:  0.1142
## F-statistic: 10.02 on 1 and 69 DF,  p-value: 0.002305

print("PREDICTION")

## [1] "PREDICTION"

pred_Frag <- predict(train_Frag.modelN, test_Frag)
summary(pred_Frag)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.03035 0.05133 0.07231 0.06812 0.09329 0.10378

actuals_predsFrag <- data.frame(cbind(actuals=test_Frag$freq, predicted=pred_Frag))
head(actuals_predsFrag)

##           actuals predicteds
## 1 0.0006645432 0.03035393
## 2 0.0365704287 0.03035393
## 3 0.0420180954 0.03035393
## 4 0.0553760960 0.03035393
## 5 0.0240641711 0.03035393
## 6 0.0239271782 0.04084348

correlation_accuracy <- cor(actuals_predsFrag) # 5.31%
min_max_accuracy <- mean(apply(actuals_predsFrag, 1, min) / apply(actuals_predsFrag, 1, max))
mape <- mean(abs((actuals_predsFrag$predicted - actuals_predsFrag$actuals))/actuals_predsFrag$actuals)

correlation_accuracy

```

```
##          actuals predicteds
## actuals    1.0000000 0.4961031
## predicteds 0.4961031 1.0000000

min_max_accuracy

## [1] 0.5303913

mape

## [1] 2.69393
```

In total we see that time is statistically significant in the change of proportions of certain categories of plastic waste: "cigarette related waste", "food related waste", "fragments". This is not the case for the "Other" and "Plastic bags and Styrofoam packaging". Some plastic waste categories are time variant while others aren't and this has to do a lot with sampling techniques and one-sidedness of datasets. It does not mean that the datasets are wrong it just shows even more that what we have talked about over many group meetings regarding the dataset being biased one way or the other is clearly backed by statistical evidence within this dataset and by referential evidence too."

"Paper: Spatial and Temporal Patterns of Stranded Intertidal Marine Debris: Is There a Picture of Global Change?"

```
plastic_category <-c("cigarette related waste", "food related waste","Fragments", "Other","Plastic bags and Styrofoam packaging")
slope_scores <- c(-0.05512,0.030425, 0.030950,-0.002428,-0.001151)
slope_interpretation <-c("downward", "upward", "upward", "relatively steady", "relatively steady")
p_value<-c("<0.05","<0.05","<0.05", ">0.05",">0.05")
corr_accuracy<-c(0.62, 0.35,0.48, 0.27,0.38)
min_max_Acc<-c(0.67,0.72,0.67, 0.65 ,0.67)
MAPE_scores<-c(0.99,0.91,0.54, 0.89, 0.43)
score_table <- data.frame(plastic_category, p_value,slope_scores, slope_interpretation, corr_accuracy,min_max_Acc,
MAPE_scores)
score_table
```

##	plastic_category	p_value	slope_scores
## 1	cigarette related waste	<0.05	-0.055120
## 2	food related waste	<0.05	0.030425
## 3	Fragments	<0.05	0.030950
## 4	Other	>0.05	-0.002428
## 5	Plastic bags and Styrofoam packaging	>0.05	-0.001151

##	slope_interpretation	corr_accuracy	min_max_Acc	MAPE_scores
## 1	downward	0.62	0.67	0.99
## 2	upward	0.35	0.72	0.91
## 3	upward	0.48	0.67	0.54
## 4	relatively steady	0.27	0.65	0.89
## 5	relatively steady	0.38	0.67	0.43

On metrics used: $Pr(> |t|)$ is the p-value, defined as the probability of observing any value equal or larger than t if H_0 is true. The larger the t statistic, the smaller the p-value. Generally, we use 0.05 as the cutoff for significance. When p-values are smaller than 0.05, we reject H_0 that there's no difference between the means and conclude that a significant difference does exist. If the p-value is larger than 0.05, we cannot conclude that a significant difference exists. $P\text{-value} \leq 0.05$ means that there is statistical significance in the results presented. If $p\text{-value} > 0.05$ we can not conclude on the statistical significance.

Correlation accuracy: A simple correlation between the actuals and predicted values can be used as a form of accuracy measure. A higher correlation accuracy implies that the actuals and predicted values have similar directional movement, i.e. when the actuals values increase the predicted values also increase and vice-versa.

MinMax Accuracy: MinMax tells you how far the model's prediction is off. For a perfect model, this measure is 1.0. The lower the measure, the worse the model, based on out-of-sample performance.

MAPE: The mean absolute percentage error (MAPE) is a statistical measure of how accurate a forecast system is. It measures this accuracy as a percentage, and can be calculated as the average absolute percent

error for each time period minus actual values divided by actual values. eg a MAPE value of 0.99 states that our model's predictions are, on average, 0.99% off from actual value.

On values of metrics used: MinMax Accuracy is generally above 65% but never exceeds 72% for all cases, which means that the model does a moderate job in predicting accurately the relative frequency of each category over time. The correlation accuracy is not that good which implies that the predicted values do not always follow the true values observed. The best value is for the model predicting on "cigarette related waste". MAPE scores are pretty good for all models showing that no model's predictions are on average more than 1% off from actual value.

5.2 Model Evaluation

5.3 Model Discussion

As seen in the study of debris in a Eastern Mediterranean coastal town [20] the variability of non-plastic litter composition should make us aware that the local context must be taken into account. Perhaps, given that non-plastic litter is a hugely varying category, this way of thinking could be expanded into the plastic debris category of "Other", given that it is inherently varied let alone in a dataset that gathers information over the entire world. In fact, this could in reverse explain the higher p-value, since high variation within the category studied would yield results of insufficient trustworthiness.

As mentioned in [24] many intertidal sites seem to be transit areas for debris, with exports matching imports over time. This could hugely impact a created model especially given that in the dataset we used we didn't have such information available. It is also important to acknowledge that our dataset is Northamerican centred which and given that it has information gathered over a period of 10 years we cannot be sure of categorization biase in the data, since there is no proof of a single sampling protocol followed. We have already spotted one such occasion, rubber gloves, and taken care of it, but further study on it is needed.

6 Discussion

It was noted that the main system for reporting debris was used by large scale clobes. This means that data is not a continuous and even flow so during events such as international beach cleanup day there may be more data in the respective month. Since these events aim mostly to cleanup after big social events extra effort might have been made to retrieve entertainment based debris such as fireworks, food packaging and six pack rings. The decrease in cigarette waste was theorised to be in correlation with a decreasing smoking rate however the cigarette waste was not the direct focus of this paper so that avenue of inquiry was not followed.

7 Conclusion and Future Work

The hypothesis H1 stands. This is no evidence as to a change in the percentage of marine debris being plastic in origin that can be observed in results such as in figure 4 which shows how the change in plastic quantity mirrors the change in overall debris quantity.

Future work might involve continuing to study correlations similar to the fireworks/july correlation that was discussed in section 4.3. Another line of study might be for example to compare the decreasing rate of cigarette use with the decreasing observation of cigarette debris which had been observed over the years to see if there is a correlation there which was hypothesised but never investigated due to not being the direct inquiry of the report.

Roshi's text below

4. Conclusion and Future Work From our results and previous literature, it can clearly be seen that plastic is the most abundant litter in the marine environment globally. Plastic was included in the 68% of the observations recorded and it made up of over 70% of the quantity of debris that was recorded. Our results also indicate the successful use of citizen science projects to monitor marine pollution. There are various ways citizen science projects have been used for and here we focused on distribution and composition of marine litter on a global scale. Identifying the amount and type of debris can help in identifying the sources of pollution. This can then help everyone to find out ways to reduce the source in the first place. We confirm the findings that plastic is the main polluter in the marine environment and more needs to be done to reduce the plastic waste. Event driven pollution such as fireworks also plays a role in marine pollution and this can be reduced by raising awareness amongst the public and trying to find more sustainable ways of celebrating events. The fact that plastic composition does not change over time is also quite alarming. Although there have been a lot of awareness programmes in the last decades, more needs to be done to see these changes that can reduce the damage done by plastic pollution.

Future work can include looking at the marine pollution debris in smaller scale by regions. Each region is different and there are various factors that can affect the data observations such as the frequency of data submissions are not at the same time everywhere. Since MDT is more focused on North America, this opportunity might be used to educate about it in other regions of the world. Although we will never be able to identify the real sources of pollution, it can help in raising awareness of the damage done by plastic pollution in the marine environment and make the public more environmentally conscious. Regulations to discourage single used plastic items and smoking in public places have been enforced in many countries. It would be interesting to look at the effect these are having on the marine debris pollution worldwide. However, more time and data will be required to see the change at a global scale.

Lots of beach clean ups and litter collections are done in the beaches worldwide to estimate the marine pollution. However, the majority (70%) of the marine debris are not on the surface but on the seafloor (UNEP, 2005). More research and studies are needed to accurately estimate the pollution in our oceans. Although marine pollution is a global issue, more needs to be done locally to reduce the sources of the pollution. In addition, standardised monitoring protocol and global partnerships are essential for efficient management of marine plastic pollution.

reference: United Nations Environment Programme (2005). Annual Evaluation Report.

8 Project Management

8.1 Facilities

Group 2 communicated using a dedicated Slack Channel, Github repository and weekly 1 hour meetings before the wednesday lab. All project documents used and the final report can be accessed from the [Public Github Repository](#) obviously we need to mention the whole covid-19 thing and how we worked around it.

8.2 Project Progress

```
require(openxlsx);
require(readxl)
library(stringr);library(data.table)
library(XLConnect)
library(xtable)

# Sheets names
fileName <- 'data/meetings.xlsx'
sheets <- readxl::excel_sheets(fileName)
#length(sheets)
# Read 1st sheet (you shouldn't have more than one sheet for this task)
# read sheet into dataframe, and rbind
dfs <- readWorksheet(loadWorkbook("data/meetings.xlsx"),sheet=1)
dfs$Date <-as.character(dfs$Date)
print(xtable(dfs,
              caption = "Record of Team Meetings",
              label = "tab:one",
              table.placement = "",
              # align changes subject to number of columns
              align = "l|l|p{8cm}|l|l|l|l"),include.rownames=FALSE,
              caption.placement = "top")
```

Table 1: Record of Team Meetings

No	Date	Topic	Alex	Georgios	Karen	Roshi	Stuart
1.00	2020-02-05	Group Formation: set up communication channel in Slack and GitHub repository	yes	yes	yes	yes	yes
2.00	2020-02-11	Agreed topic of "Plastic Pollution", distributed research activity for week	yes	yes	yes	yes	yes
3.00	2020-02-18	Presented individuals' research findings and discussed hypothesis	yes	yes	yes	yes	yes
4.00	2020-02-25	Decided on final dataset to use and hypothesis of "proportion of marine plastics pollution does not change over time"	yes	yes	yes	yes	yes
5.00	2020-03-04	Presentation draft agreed	yes	yes	yes	yes	yes
6.00	2020-03-10	Distributed section writing activity for week	yes	yes	yes	yes	yes
7.00	2020-03-17						
8.00	2020-03-24						
9.00	2020-03-31						
10.00	2020-04-07						
11.00	2020-04-14						
12.00	2020-04-21						

8.3 Peer-assessment

```
require(openxlsx);
require(readxl)
library(stringr);library(data.table)
library(XLConnect)
library(xtable)

# Sheets names
fileName <- 'data/peers.xlsx'
sheets <- readxl::excel_sheets(fileName)
#length(sheets)
# Read 1st sheet (you shouldn't have more than one sheet for this task)
# read sheet into dataframe, and rbind
dfs <- readWorksheet(loadWorkbook("data/peers.xlsx"),sheet=1)

# convert fields into chars
dfs[, ] <- lapply(dfs[, ], as.character)

print(xtable(dfs,
              caption = "Peer Assessment out of 100",
              label = "tab:two",
              table.placement = "",
              # align changes subject to number of columns
              align = "l1l1l1l1"),include.rownames=FALSE,
              caption.placement = "top")
```

Table 2: Peer Assessment out of 100

Peer.Review	Alex	Georgios	Karen	Roshi	Stuart
Alex	100	100	100	100	100
Georgios	100	100	100	100	100
Karen	100	100	100	100	100
Roshi	100	100	100	100	100
Stuart	100	100	100	100	100

References

- [1] José G.B Derraik. “The pollution of the marine environment by plastic debris: a review”. In: *Marine Pollution Bulletin* 44.9 (2002), pp. 842–852. ISSN: 0025-326X. DOI: [https://doi.org/10.1016/S0025-326X\(02\)00220-5](https://doi.org/10.1016/S0025-326X(02)00220-5). URL: <http://www.sciencedirect.com/science/article/pii/S0025326X02002205>.
- [2] Lorena M. Rios, Charles Moore, and Patrick R. Jones. “Persistent organic pollutants carried by synthetic polymers in the ocean environment”. In: *Marine Pollution Bulletin* 54.8 (2007), pp. 1230–1237. ISSN: 0025-326X. DOI: <https://doi.org/10.1016/j.marpolbul.2007.03.022>. URL: <http://www.sciencedirect.com/science/article/pii/S0025326X07001324>.
- [3] David Barnes et al. “Accumulation and fragmentation of plastic debris in global environments”. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364 (Aug. 2009), pp. 1985–98. DOI: [10.1098/rstb.2008.0205](https://doi.org/10.1098/rstb.2008.0205).
- [4] Alex Sivan. “New perspectives in plastic biodegradation”. In: *Current Opinion in Biotechnology* 22.3 (2011), pp. 422–426. ISSN: 0958-1669. DOI: <https://doi.org/10.1016/j.copbio.2011.01.013>. URL: <http://www.sciencedirect.com/science/article/pii/S0958166911000292>.
- [5] Jenna R. Jambeck et al. “Plastic waste inputs from land into the ocean”. In: *Science* 347.6223 (2015), pp. 768–771. ISSN: 0036-8075. DOI: [10.1126/science.1260352](https://doi.org/10.1126/science.1260352). eprint: <https://science.sciencemag.org/content/347/6223/768.full.pdf>. URL: <https://science.sciencemag.org/content/347/6223/768>.
- [6] Chelsea M. Rochman. “The Complex Mixture, Fate and Toxicity of Chemicals Associated with Plastic Debris in the Marine Environment”. In: *Marine Anthropogenic Litter*. Ed. by Melanie Bergmann, Lars Gutow, and Michael Klages. Cham: Springer International Publishing, 2015, pp. 117–140. ISBN: 978-3-319-16510-3. DOI: [10.1007/978-3-319-16510-3_5](https://doi.org/10.1007/978-3-319-16510-3_5). URL: https://doi.org/10.1007/978-3-319-16510-3_5.
- [7] S.C. Gall and R.C. Thompson. “The impact of debris on marine life”. In: *Marine Pollution Bulletin* 92.1 (2015), pp. 170–179. ISSN: 0025-326X. DOI: <https://doi.org/10.1016/j.marpolbul.2014.12.041>. URL: <http://www.sciencedirect.com/science/article/pii/S0025326X14008571>.
- [8] Susanne Kühn, Elisa L. Bravo Rebolledo, and Jan A. van Franeker. “Deleterious Effects of Litter on Marine Life”. In: *Marine Anthropogenic Litter*. Ed. by Melanie Bergmann, Lars Gutow, and Michael Klages. Cham: Springer International Publishing, 2015, pp. 75–116. ISBN: 978-3-319-16510-3. DOI: [10.1007/978-3-319-16510-3_4](https://doi.org/10.1007/978-3-319-16510-3_4). URL: https://doi.org/10.1007/978-3-319-16510-3_4.
- [9] Peter G. Ryan. “A Brief History of Marine Litter Research”. In: *Marine Anthropogenic Litter*. Ed. by Melanie Bergmann, Lars Gutow, and Michael Klages. Cham: Springer International Publishing, 2015, pp. 1–25. ISBN: 978-3-319-16510-3. DOI: [10.1007/978-3-319-16510-3_1](https://doi.org/10.1007/978-3-319-16510-3_1). URL: https://doi.org/10.1007/978-3-319-16510-3_1.
- [10] A.T. Williams and Nelson Rangel-Buitrago. “Marine Litter: Solutions for a Major Environmental Problem”. In: *Journal of Coastal Research* 35.3 (2019), pp. 648–663. DOI: [10.2112/JCOASTRES-D-18-00096.1](https://doi.org/10.2112/JCOASTRES-D-18-00096.1). URL: <https://doi.org/10.2112/JCOASTRES-D-18-00096.1>.
- [11] “Occurrence of microplastics in the gastrointestinal tract of pelagic and demersal fish from the English Channel”. In: *Marine Pollution Bulletin* 67.1 (2013), pp. 94–99. ISSN: 0025-326X. DOI: <https://doi.org/10.1016/j.marpolbul.2012.11.028>.
- [12] Ana Markic et al. “Plastic ingestion by marine fish in the wild”. In: *Critical Reviews in Environmental Science and Technology* 50.7 (2020), pp. 657–697. DOI: [10.1080/10643389.2019.1631990](https://doi.org/10.1080/10643389.2019.1631990). eprint: <https://doi.org/10.1080/10643389.2019.1631990>. URL: <https://doi.org/10.1080/10643389.2019.1631990>.
- [13] J.M. Coe and D.B. Rogers. “Marine Debris: Sources, Impacts, and Solutions”. In: *Environmental Management Series*. Springer, 1997. ISBN: 9780387947594. URL: <https://books.google.co.uk/books?id=aSoRAAAAYAAJ>.

- [14] Stephen D. Garrity and Sally C. Levings. “Marine debris along the Caribbean coast of Panama”. In: *Marine Pollution Bulletin* 26.6 (1993), pp. 317–324. ISSN: 0025-326X. DOI: [https://doi.org/10.1016/0025-326X\(93\)90574-4](https://doi.org/10.1016/0025-326X(93)90574-4). URL: <http://www.sciencedirect.com/science/article/pii/S0025326X93905744>.
- [15] Bruno A. Walther, Alexander Kunz, and Chieh-Shen Hu. “Type and quantity of coastal debris pollution in Taiwan: A 12-year nationwide assessment using citizen science data”. In: *Marine Pollution Bulletin* 135 (2018), pp. 862–872. ISSN: 0025-326X. DOI: <https://doi.org/10.1016/j.marpolbul.2018.08.025>. URL: <http://www.sciencedirect.com/science/article/pii/S0025326X18305897>.
- [16] Takashi Kusui and Michio Noda. “International survey on the distribution of stranded and buried litter on beaches along the Sea of Japan”. In: *Marine pollution bulletin* 47 (Feb. 2003), pp. 175–9. DOI: [10.1016/S0025-326X\(02\)00478-2](https://doi.org/10.1016/S0025-326X(02)00478-2).
- [17] Stefanie Reinold. “Plastic pollution on eight beaches of Tenerife (Canary Islands, Spain): An annual study: Wind and Wave parameters”. In: (Mar. 2020). DOI: [10.17632/f7ntbw4rt6.1](https://doi.org/10.17632/f7ntbw4rt6.1). URL: https://mendeley.figshare.com/articles/Plastic_pollution_on_eight_beaches_of_Tenerife_Canary_Islands_Spain_An_annual_study_Wind_and_Wave_parameters/11972994.
- [18] Anthony Cheshire et al. *UNEP/IOC Guidelines on Survey and Monitoring of Marine Litter*. Jan. 2009.
- [19] Ronen Alkalay, Galia Pasternak, and Alon Zask. “Clean-coast index—A new approach for beach cleanliness assessment”. In: *Ocean and Coastal Management* 50 (Dec. 2007). DOI: [10.1016/j.ocecoaman.2006.10.002](https://doi.org/10.1016/j.ocecoaman.2006.10.002).
- [20] Michelle Portman and Ruth Brennan. “Marine litter from beach-based sources: Case study of an Eastern Mediterranean coastal town”. In: *Waste Management* 69 (Aug. 2017). DOI: [10.1016/j.wasman.2017.07.040](https://doi.org/10.1016/j.wasman.2017.07.040).
- [21] Arun Kumar A. et al. “Preliminary study on marine debris pollution along Marina beach, Chennai, India”. In: *Regional Studies in Marine Science* 5 (2016), pp. 35–40. ISSN: 2352-4855. DOI: <https://doi.org/10.1016/j.rsma.2016.01.002>. URL: <http://www.sciencedirect.com/science/article/pii/S2352485516300020>.
- [22] Allan Williams et al. “Distribution of beach litter along the coastline of Cádiz, Spain”. In: *Marine pollution bulletin* 107 (Apr. 2016). DOI: [10.1016/j.marpolbul.2016.04.015](https://doi.org/10.1016/j.marpolbul.2016.04.015).
- [23] J. R. Jambeck and K. Johnsen. “Citizen-Based Litter and Marine Debris Data Collection and Mapping”. In: *Computing in Science Engineering* 17.4 (2015), pp. 20–26.
- [24] Mark Anthony Browne et al. “Spatial and Temporal Patterns of Stranded Intertidal Marine Debris: Is There a Picture of Global Change?” In: *Environmental Science and Technology* 49 (May 2015). DOI: [10.1021/es5060572](https://doi.org/10.1021/es5060572).