${\tt concordance}{=}{\tt TRUE}$

Plastic Pollution in Oceans Group 2 Report - CMM507

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

April 23, 2020

Objective

- To understand the composition of plastic pollutants in the ocean
- To understand the sources of plastic pollutants
- To understand how plastic pollution gets distributed across the oceans

1 Problem Statement

H1 = The % of plastic pollution remains constant over time.

H0 = The % of plastic pollution does not remain constant over time.

1.1 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.[1] [2] [3]

Synthetic organic polymer derived from polymerisation of monomers extracted from oil and gas make up the plastics. [1] [4] The lightweight feature and its durability make it very suitable to make a range of products that we use in our everyday life. [5] [6] These same features have been a major cause of pollution due to overuse and non-managed waste disposal system worldwide with plastic contributing to the 10% of the waste generated worldwide. [5] Due to its buoyancy, plastic debris can be dispersed over long distances and they can persist for a long time. Although, plastic litter has been a major cause of marine pollution for a while, its seriousness has only been realised recently. Jambeck et al., [7] reported that in 2010 alone, between 4.8 million to 12.7 million metric tons of plastics entered the ocean. Plastics are now everywhere in the marine environment and urgent action is required to mitigate this problem and reduce the harmful impact. [4] [8]

1.2 Motivation

Impact on marine life

Plastics in ocean is one of the many forms of human impact that threatens marine life. There is still very little information available on the impact of plastic pollution on the ocean's ecosystem. Due to the realisation on impact of human on climate and environment, there has been a lot of awareness activities to reduce the impact of pollution. Ban on single use plastic bags are being applied to many countries in order to protect the environment.

Over 700 marine wildlife species are affected due to entanglement in plastic ropes and materials and ingestion of plastics in the ocean. [9] Over 340 species of marine animals were found to be entangled. [10] Reducing plastic waste is a major challenge worldwide. It is almost impossible to estimate the number of marine animals affected by marine pollution globally due to the vastness of the ocean. However, studies carried out on the gut contents of thousands of seabirds, found the significant increase in the ingestion of plastics during the 10-15 years interval. [11] This result might correlate to the rapid increase of plastic production and plastic use globally. In a study carried out over fourteen years, Moser and Lee [12] found that more 50% of the seabird species contained plastic particles in the gut which increased over time. This could be due the increase in plastic availability over time.

Entanglement in plastic debris is another cause of marine life suffering. Discarded fishing gear and floating mastic masses in ocean are serious threat to marine animals. Some animals such as seals are attracted to the floating plastics where they get entangled and get suffocated. Harmful effect of litter on marine life has been reviewed extensively.[9] [10] [13] [14] Floating plastics over long distances can disperse alien species as well as some pathogens. Drifting plastic debris are also the source of alien species introduction and thus affecting the native marine biodiversity.[15] [16]

Impact on environment and human health

Plastic debris floating in the oceans and the littering the coastal areas are not a pleasant sight. Masses of plastic accumulation and discarded objects made from plastics are found everywhere nowadays.

Over time plastic disintegrates into small microplastics which are easily consumed by fish and they enter the food chain. Plastics have been found in a third of fish caught in the UK which included the popular fishes such as cod, haddock and mackerel. Impact of plastic entering the human food chain and the effects of it are still to be studied. Plastic toxicity and the occurrence of microplastics and nanoplastics in the water supply can also be a direct impact on human health in addition to the contamination in seafood. [8] [17]

Reducing plastic pollution has recently been a global aim. Research in plastic pollution in marine environment has played a big role in reducing it and raising awareness all over the world. In order to understand the plastic pollution in marine environments and its effect in long term, it is essential to keep collecting data on patterns of marine debris around the world. Effective monitoring of plastic debris is very essential in order to reduce the abundance of plastic debris everywhere. In addition, monitoring the type, frequency and the source of the litter is also important for prevention initiative of marine pollution. Most of the monitoring are done by surveys looking at frequencies of beach litter collected by organisations and volunteers. [18] Most abundant litter can be found close to urban areas where beach visitor numbers are higher. [19]

1.3 Objectives

The main objectives of this project can be outlined as follows:

2 Research

Things we found

Sources of pollution: 10 river dataset, 50km2 coastline dataset, pollution density and body of water dataset....

3 Methods

This paper is conducted using secondary data collection methods only. The authors did not collect or create any new data using primary methods.

3.1 Dataset Description

data dictionary is probably better as a table than a list. Can we also add to this information of how the data is entered e.g. optional/mandatory fields, free-text or dropdown fields. The use of IDs suggest these are option fields with lookup tables somewhere.

- The data was taken from *marine debris tracker* between 2010 till February 19th 2020. The time of 2010 was chosen as there was no data before that time.
- The dataset was composed by combining the multiple csv files gathered from the marine debris tracker into a single set after this was done the date data type was renamed "Time".
- The dataset created from the combined csv files contain more than 360000 rows of data and consists
 of the following variables.
 - ListID is the ID code for the list
 - ListName is the name of the list
 - ItemID is the ID code given to the item of debris
 - ItemName is the name we give to item of debris
 - LogID is the ID code given to the location of the debris
 - Latitude, Longitude and Altitude are the coordinates of the location where the observation was made
 - Quantity is the number of pieces of debris in the observation.
 - Error radius is the radius around the observation site within the error for reasonable doubt.
 - Location is the area the observation of debris was made in.
 - Description is the description of the area the debris was found in.
 - MaterialID is the ID code of the material that the debris was composed of.
 - Material Description is the description given to the material that composes the debris.
 - Time is the time that the observation was made.
 - There were a number of problems with the dataset namely;
 - * There were a number of cases of missing data in the dataset.
 - * data anomalies (lat/long values don't match named regions)

*

3.2 Dataset Pre-processing

Everything below is from Stuart's RNW file

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

The following actions were performed on the dataset:

ListIDänd ListNamewere found to be redundant and removed. can we remove ItemID and MaterialID also since we have the descriptions?

Nulls found in ItemName and Description.

this means every entry has a material at least? why? could be a required field? that would explain why some

entries are rubbish if people are forced to pick a category it is also worth discussing the merits of dropdown entries: standardises input but forces a value where none might be appropriate, or a default it selected? Stuart: Yes, it is a required field. I checked on the mobile app and you select a item type from different material sections. Note however that there is a material type *Other Items* which contains the items *Other* and *Test Item*. Therefore users are able to categorise an item as other if it is not appropriate for any other option on the list.

NAs removed not sure what is actually happening with the explicit_na() piece for location?

Stuart: I used explicit_na to replace missing values with an explicit level name (like "Missing" or something similar). I think it was to remove warnings when I was plotting.

also maybe worth looking at: what's the significance of some of these itemIDs where the itemname is blank? It could be an item once that was then deleted or categorised retrospectively. Do a groupby ItemID and see if more than one material or item name turns up.

I was puzzled by this so I investigated it and found it is not in the source data. I found a coding bug further down in the code, where I make the stacked bar chart with proportions per month. I have fixed this, and now no ItemName values are blank. Sorry!

Unique values for each column: can we present these unique counts as a formatted table? I think it's interesting that 55 unique items can have 8k descriptions

##	ItemID	ItemName	LogID
##	55	55	363368
##	Latitude	Longitude	Altitude
##	142707	136490	135214
##	Quantity	Error Radius	Location
##	496	18374	1458
##	Description	Material ID	Material Description
##	8494	8	8
##	Time		
##	248436		

3.3 Data Quality Issues: Classification

The authors find that there are multiple instances of missclassified items. Where their descriptions appear to not match their material categorisation

Lets see if there are any "ItemNames" associated with more than one "Material Descriptions".

So rubber gloves are associated with two material descriptions, but otherwise a one to many relationship exists between "Material Description" and "ItemName".

It seems that most rubber gloves are classified as plastic rather than rubber.

```
## # A tibble: 33 x 3
##
      `Material Description` ItemName
                                          Description
##
      <fct>
                             <fct>
##
   1 PLASTIC
                             Rubber Glov Found on wassaw island Oct. 21 with beac
##
   2 PLASTIC
                             Rubber Glov~ undefined
                             Rubber Glov~ undefined
##
   3 PLASTIC
   4 PLASTIC
                             Rubber Glov thermal
##
##
   5 PLASTIC
                             Rubber Glov Near water
   6 PLASTIC
                             Rubber Glov Taste of Omaha Cleanup
   7 PLASTIC
                             Rubber Glov Taste of Omaha Cleanup
##
   8 PLASTIC
                             Rubber Glov~ 2 diff kinds
##
##
  9 PLASTIC
                             Rubber Glov~ undefined
## 10 PLASTIC
                             Rubber Glov Latex
## # ... with 23 more rows
```

All instances of rubber gloves with non-missing descriptions are categorised as plastic. We also see that the descriptions suggest that the categorisation may be innaccurate: the last two instances here have "Balloon" in the extra descriptions... why aren't they categorised as such? another thing maybe worth looking at: all 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. further explores the point about missclassified data.

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.

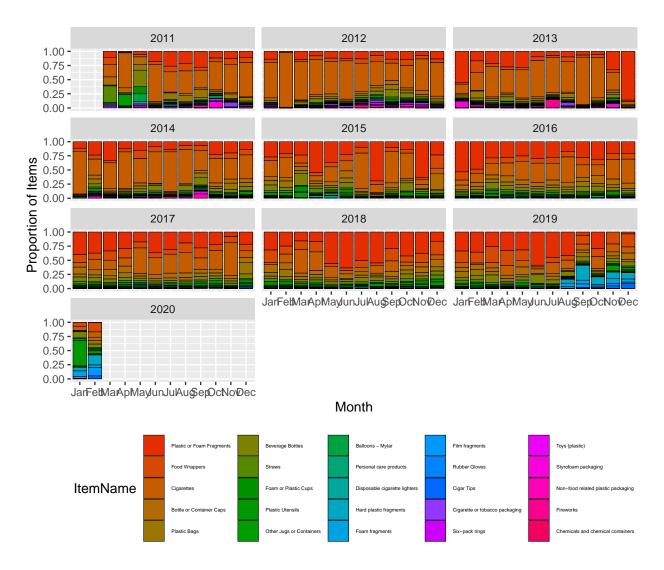


Figure 1: Debris by categorisation

Adjust legend text size in annotated code below. - Stuart

Rel. frequencies of observed plastic waste by category

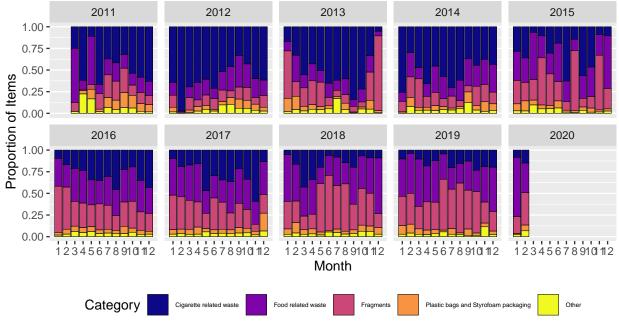


Figure 2: Recategorisation by year. Colour scale ordered by ranking of total observed quantity.

4 Exploration

Here we describe the things we found...

4.1 Proportion Trends

How pollutant proportions change over time.

Cigarette butts proportions and raw counts decrease over time: possibly less people smoking, or moving to vaping

General pollution count going down over time?

Old pollutants fall away (cigarette butts) but new ones are introduced

Question: Are observed plastic item proportions time invariant?

I changed the parameters here to adjust figure size. Change the number in out.width. - Stuart

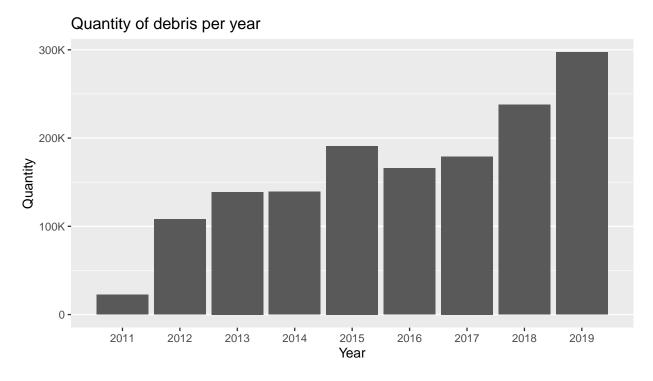


Figure 3: Trend of debris observered

this chart needs to size down

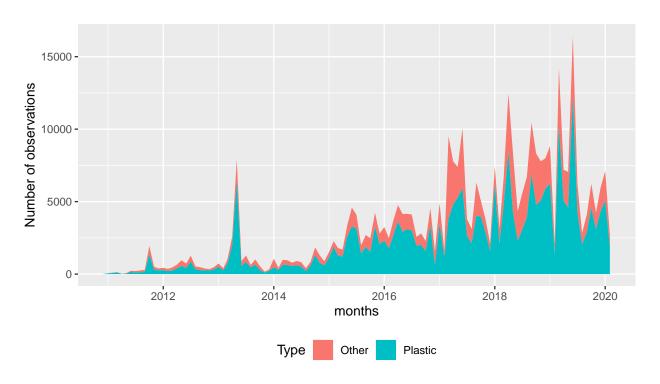


Figure 4: Observations of plastic debris v all debris

this chart needs to size down

4.2 Distribution of observed debris:

Material Quantities

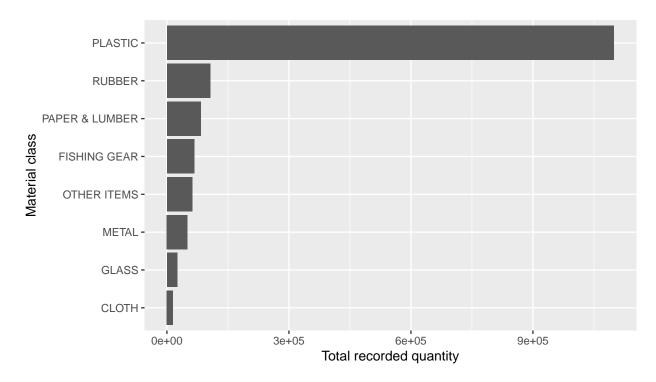


Figure 5: Material Quantities

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:

Figure 6: Debris categorisation

chart disabled for now. This chart was working for a while (works in RStudio R, not LaTeX) only hint I can see is:## Warning: Factor 'ItemName' contains implicit NA, consider using 'forcats::fct_explicit_na' 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.

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)

Firework debris 2012-2019

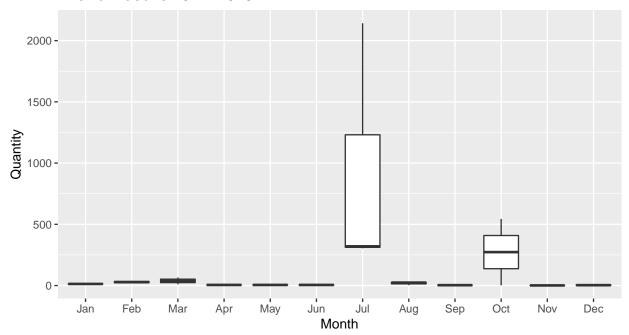


Figure 7: Boxplot of fireworks distribution by month, across all years

this chart needs to size down

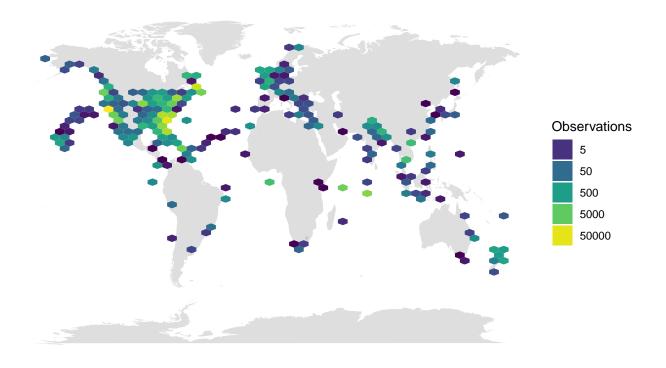
4.4 Location-Driven Pollution

Rubber found in Indonessia 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
    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.

```
#Scatterplot of long/lat positions
data %>%
  mutate(Location = str_to_lower(Location)) %>%
  filter(str_detect(Location, "united kingdom")) %>%
  select(Latitude, Longitude) %>%
  ggplot(aes(x = Latitude, y = Longitude)) +
    geom_point(position = "jitter")
```

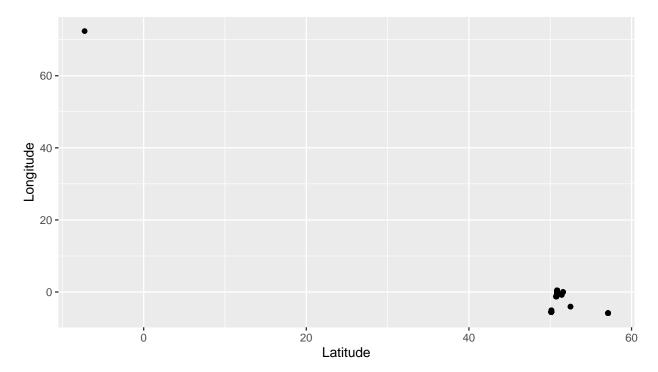


Figure 8: Longitude and Latitude discrepancies

this chart needs to size down We have a outliers here. Maybe a difference in standards used for Longitude and Latitude? Some systems put the Latitude origin close to the UK.

Questions

Distribution of plastic by location.

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

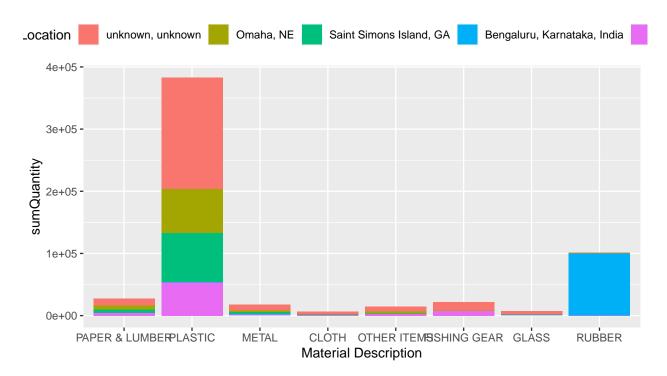


Figure 9: Debris by location

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. this chart needs to size down

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

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, given we know that event-driven pollution will determine different pollutants are different times.

5.1 Description of Model

Georgios' script

' $geom_smooth()$ ' using formula 'y ~ x'

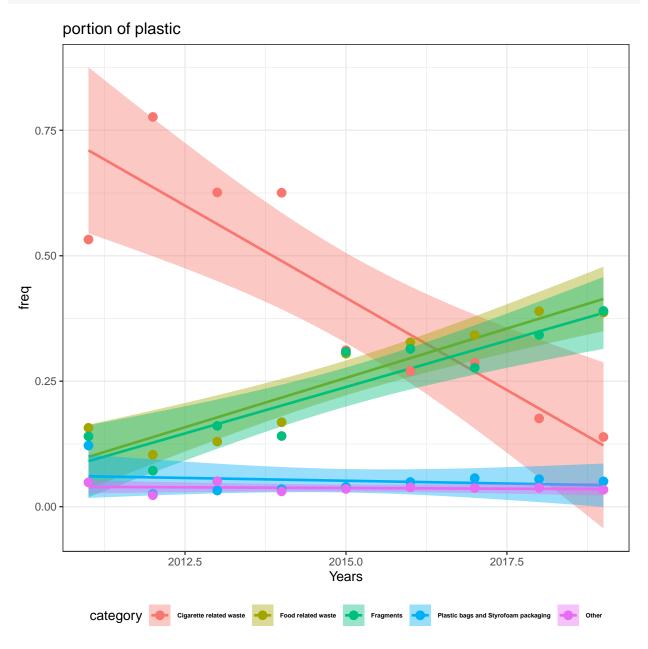
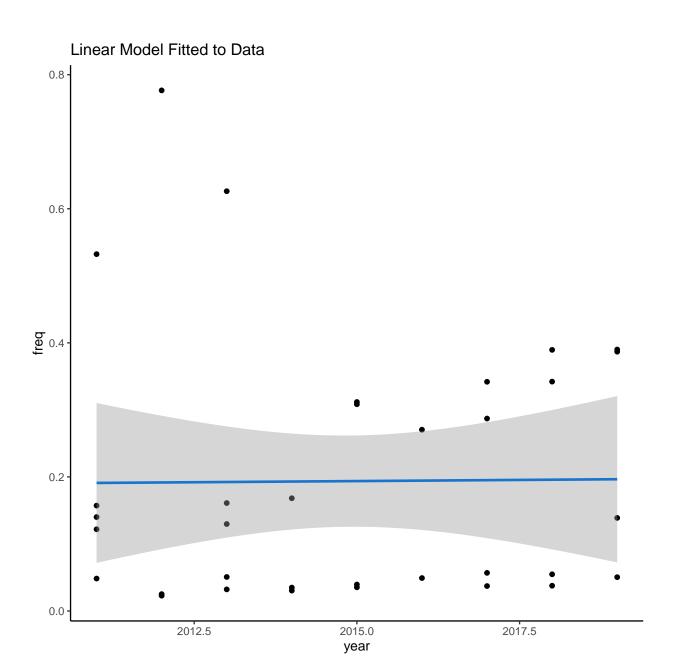


Figure 10: Current Observations

```
## [1] "train model"
##
## Call:
## lm(formula = freq ~ year, data = train_dfTotN)
##
## Residuals:
##
                1Q
       Min
                     Median
                                    3Q
                                            Max
## -0.16888 -0.15222 -0.06018 0.11714 0.58484
##
## Coefficients:
##
                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.1955072 24.9391038 -0.048
## year
              0.0006895 0.0123774
                                      0.056
                                                0.956
##
## Residual standard error: 0.1947 on 32 degrees of freedom
## Multiple R-squared: 9.696e-05, Adjusted R-squared: -0.03115
## F-statistic: 0.003103 on 1 and 32 DF, p-value: 0.9559
```

 $Pr(\xi_t-t-)$ is the p-value, defined as the probability of observing any value equal or larger than t if H0 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 H0. Here p is pretty big which means that there is statistically significant correlation between relative frequency and years passing by. Which basically further supports our initial hypothesis in this project. I have included a prediction on the test set but it is of no worth obviously.

```
## 'geom_smooth()' using formula 'y ~ x'
```



5.2 Model Evaluation

5.3 Model Results

```
## [1] "PREDICTION"
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.1917 0.1931 0.1945 0.1941 0.1948 0.1966
## Error in xtable(head(actuals_preds)): could not find function "xtable"
```

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 predicteds also increase and vice-versa.

equal or greater than 99.4%, mean absolute percentage deviation. Intrestingly enough min/max accuracy and mostly mean absolute percentage deviation score quite well but still on a model that can not be trusted. Time does not impact plastic composition.

6 Discussion

7 Conclusion and Future Work

Our hypothesis stands/does not stand.

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

```
## Error: package or namespace load failed for 'XLConnect':
## .onLoad failed in loadNamespace() for 'rJava', details:
## call: fun(libname, pkgname)
## error: JAVA_HOME cannot be determined from the Registry
## Error in readWorksheet(loadWorkbook("data/meetings.xlsx"), sheet = 1): could not find function
"readWorksheet"
## Error in eval(expr, envir, enclos): object 'dfs' not found
## Error in xtable(dfs, caption = "Record of Team Meetings", label = "tab:one", : object 'dfs'
not found
```

8.3 Peer-assessment

```
## Error: package or namespace load failed for 'XLConnect':
## .onLoad failed in loadNamespace() for 'rJava', details:
## call: fun(libname, pkgname)
## error: JAVA_HOME cannot be determined from the Registry
## Error in readWorksheet(loadWorkbook("data/peers.xlsx"), sheet = 1): could not find function
"readWorksheet"
## Error in lapply(dfs[, ], as.character): object 'dfs' not found
## Error in xtable(dfs, caption = "Peer Assessment out of 100", label = "tab:two", : object
'dfs' not found
```

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. URL: http://www.sciencedirect.com/science/article/pii/S0025326X02002205.
- [2] Julia Reisser et al. "Marine Plastic Pollution in Waters around Australia: Characteristics, Concentrations, and Pathways". In: *PLOS ONE* 8.11 (Nov. 2013), pp. 1–11. DOI: 10.1371/journal.pone. 0080466. URL: https://doi.org/10.1371/journal.pone.0080466.
- [3] "Microplastics in wild fish from North East Atlantic Ocean and its potential for causing neurotoxic effects, lipid oxidative damage, and human health risks associated with ingestion exposure". In: Science of The Total Environment 717 (2020), p. 134625. ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2019.134625. URL: http://www.sciencedirect.com/science/article/pii/S0048969719346169.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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. eprint: https://science.sciencemag.org/content/347/6223/768.full.pdf. URL: https://science.sciencemag.org/content/347/6223/768.
- [8] 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. URL: https://doi.org/10.1007/978-3-319-16510-3_5.
- [9] 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.
- [10] Susanne Kühn, Elisa L. Bravo Rebolledo, and Jan A. van Francker. "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. URL: https://doi.org/10.1007/978-3-319-16510-3_4.
- [11] Martin D. Robards, John F. Piatt, and Kenton D. Wohl. "Increasing frequency of plastic particles ingested by seabirds in the subarctic North Pacific". In: *Marine Pollution Bulletin* 30.2 (1995), pp. 151–157. ISSN: 0025-326X. DOI: https://doi.org/10.1016/0025-326X(94)00121-0. URL: http://www.sciencedirect.com/science/article/pii/0025326X94001210.
- [12] Mary L. Moser and David S. Lee. "A Fourteen-Year Survey of Plastic Ingestion by Western North Atlantic Seabirds". In: *Colonial Waterbirds* 15.1 (1992), pp. 83–94. ISSN: 07386028. URL: http://www.jstor.org/stable/1521357.
- [13] 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. URL: https://doi.org/10.1007/978-3-319-16510-3_1.

- [14] 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. URL: https://doi.org/10.2112/JCOASTRES-D-18-00096.1.
- [15] Murray Gregory. "Environmental implications of plastic debris in marine settings—Entanglement, ingestion, smothering, hangers-on, hitch-hiking and alien invasions". In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364 (Aug. 2009), pp. 2013–25. DOI: 10.1098/rstb.2008.0265.
- [16] Tim Kiessling, Lars Gutow, and Martin Thiel. "Marine Litter as Habitat and Dispersal Vector". In: Marine Anthropogenic Litter. Ed. by Melanie Bergmann, Lars Gutow, and Michael Klages. Cham: Springer International Publishing, 2015, pp. 141–181. ISBN: 978-3-319-16510-3. DOI: 10.1007/978-3-319-16510-3. 6. URL: https://doi.org/10.1007/978-3-319-16510-3.6.
- [17] 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. eprint: https://doi.org/10.1080/10643389.2019.1631990. URL: https://doi.org/10.1080/10643389.2019.1631990.
- [18] 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.
- [19] 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. URL: http://www.sciencedirect.com/science/article/pii/0025326X93905744.