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*

March 4, 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

1.2 Motivation

Plastic pollution is bad because.... citation example [4].

1.3 Objectives

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

2 Research

Things we found citation example [4].
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

- Where the dataset came from
- How it is constructed: multiple csv files by year
- A description of what it is, what's in it and what it represents
- Problems with the dataset
  - Missing data
  - data anomalies (lat/long values don't match named regions)
  -

3.2 Dataset Pre-processing

Because of the features and concerns identified in the section above, we chose to transform the dataset in the following ways:

- reclassified some labels because variation was too high (there were too many labels)
- removed missing values
- removed certain subsets
- but kept certain subsets

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

4.2 Event-Driven Pollution

Fireworks found in July and North-America only: possibly 4th July celebrations

4.3 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

5 Predictive Modelling

5.1 Description of Model

5.2 Model Results

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*

8.2 Project Progress

Table 1: Record of Team Meetings

No	Date	Topic	John	Ali	Ann	Mon
1.00	2019-02-05	Team Formation Formation Formation Formation Formation	yes	yes	yes	yes
2.00	2019-02-06	Team Formation	yes	yes	yes	yes
3.00	2019-02-07	Team Formation	yes	yes	yes	yes
4.00	2019-02-08	Team Formation	yes	yes	yes	yes
5.00	2019-02-09	Team Formation	yes	yes	yes	yes
6.00	2019-02-10	Team Formation	yes	yes	yes	yes
7.00	2019-02-11	Team Formation	yes	yes	yes	yes
8.00	2019-02-12	Final Meeting	yes	yes	yes	yes

8.3 Peer-assessment

Same as we did with Table 1, we can also generate the peer-assessment table providing that we record things in an excel sheet.

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

8.4 Section on figure referencing - keep for referencing

In this project iris was used, the dataset is made of 150 rows and four features.

Notice how we generate graphics within the sweave document. Check the following code, we will create a function that either finds x^2 or x^3 subject to parameters passed in the function

```
# create a vector of doubles
myNumbers <- seq(from=-1,to=1,by=.1)

# function definition
toPower <- function (x,p=2) {
    if (p==2)
        return (x*x)
    else if (p==3)
        return (x*x*x)
    return (x*x)
}

# call function
squared <- toPower(myNumbers)
cubes <- toPower(myNumbers,3)
```

An easy way to check that our function is doing the right calculation is to plot the results. The code below will generate a figure similar to Figure 1:

```
plot(myNumbers,cubes,type='b',xlab = 'x', ylab = 'x*x',frame=FALSE,col='blue')
```

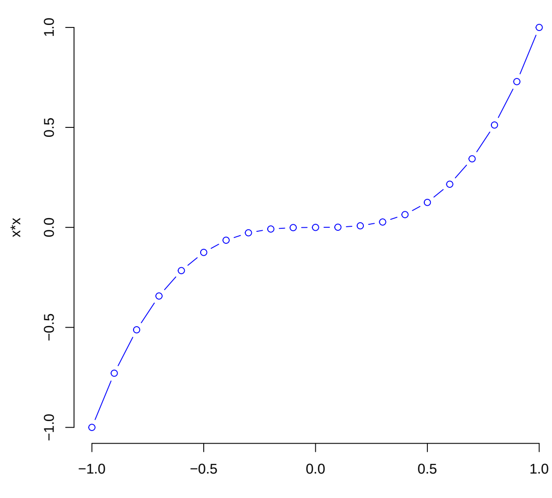

Figure 1: Simple Plot of $f(x) = x^3$ Function

8.5 Experiments

Now we can show how the function $f(x) = x^2$ looks like (Figure 2)

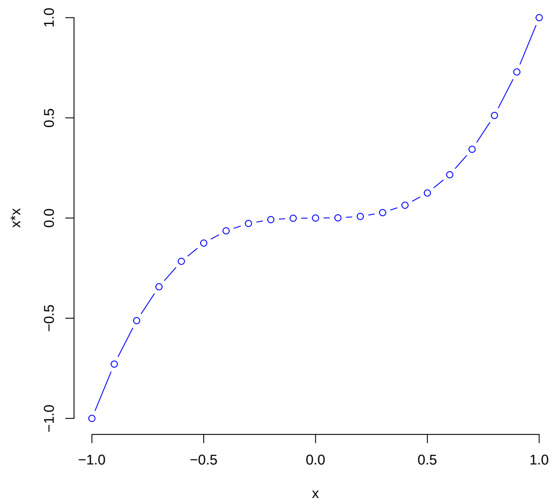

Figure 2: Simple Plot of $f(x) = x^3$ Function

References

[1] Yihui Xie. *knitr: A general-purpose package for dynamic report generation in R*. R package version 1.4.1. 2013. URL: http://yihui.name/knitr/.

[2] Eyad Elyan and Mohamed Medhat Gaber. "A genetic algorithm approach to optimising random forests applied to class engineered data". In: *Information Sciences* 384.Supplement C (2017), pp. 220–234. ISSN: 0020-0255. DOI: 10.1016/j.ins.2016.08.007. URL: https://doi.org/10.1016/j.ins.2016.08.007.

[3] Eyad Elyan and Mohamed Medhat Gaber. "A fine-grained Random Forests using class decomposition: an application to medical diagnosis". In: *Neural Computing and Applications* 27.8 (Nov. 2016), pp. 2279–2288. ISSN: 1433-3058. DOI: 10.1007/s00521-015-2064-z. URL: https://doi.org/10.1007/s00521-015-2064-z.

[4] E. Elyan, C. M. Garcia, and C. Jayne. "Symbols Classification in Engineering Drawings". In: *2018 International Joint Conference on Neural Networks (IJCNN)*. July 2018, pp. 1–8. DOI: 10.1109/IJCNN.2018.8489087.