

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 6, 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 (Derraik, 2002; Reisser, 2015; Barboza et al., 2019; Tekman et al., 2019).

Synthetic organic polymer derived from polymerisation of monomers extracted from oil and gas make up the plastics (Derraik, 2002; Rios et al., 2007). Plastic has been the most used manmade materials since the 1900s. Since the 1940s mass production of plastic has been increasing rapidly worldwide (PlasticsEurope, 2010). The lightweight feature and its durability make it very suitable to make a range of products that we use in our everyday life (Barnes et al., 2009; Sivan 2011). 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 (Barnes et al., 2009). Due to its buoyancy, plastic debris can be dispersed over long distances and they can persist for a long time (Goldberg, 1997). Although, plastic litter has been a major cause of marine pollution for a while, its seriousness has only been realised recently (Stefatos et al., 1999).

1.2 Motivation

Jambeck et al., (2015) 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 (Rios et al., 2007; Rochman et al., 2015).

Various papers covering the broad range of topics related to marine litter will be discussed. (this will be topics for literature review?? To be added later)

Impact on marine life

Plastics in ocean is 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 now (ref???) in order to protect the environment.

Over 700 marine organisms are affected due to entanglement in plastic ropes and materials and ingestion of plastics in the ocean (Reff?). Over 340 species of marine animals were found to be entangled (Kuhn et al.,). In UK 2-9% of animals were affected by entanglement (Werner et al., 2016). 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 (Robards et al., 1995). 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 (1992) 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 suffocated (Mattlin and Cawthron, 1986). 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 (McKinney, 1998).

Impact on environment

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. (affecting the recreational activities and thus impacting on socio-economic status in some areas: to be discussed later). Impact on human health 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 (reff??). 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 (Rochman et al., 2015).

Sources of marine plastic pollution

<https://pubs.acs.org/doi/10.1021/acs.est.7b02368> plastic debris from river to sea

Around 80% of the 8 million tonnes of plastics come from land-based sources, with the remainder coming from shipping and the fishing industry (reffj??).

Microplastic in food. Microplastic are formed by fragmentation on the large plastic debris... <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511111/>

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

- The data was taken from marine debris tracker (marinedebris.engr.uga.edu/newmap/) 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

```
## Error: package or namespace load failed for 'imager' in dyn.load(file, DLLpath = DLLpath, ...):  
## unable to load shared object '/Library/Frameworks/R.framework/Versions/3.6/Resources/library/imager.so':  
## dlopen(/Library/Frameworks/R.framework/Versions/3.6/Resources/library/imager/libs/imager.so, 6): Library not loaded: /opt/X11/lib/libX11.6.dylib  
## Referenced from: /Library/Frameworks/R.framework/Versions/3.6/Resources/library/imager/libs/imager.so  
## Reason: image not found
```

Logged marine debris is available for download <http://marinedebris.engr.uga.edu/newmap/>. I'm importing data from 2010 till Feb 19th 2020. There doesn't seem to be data before 2010. The data is reported marine debris. DataImport

```
data <- list.files(path = "data/", full.names = TRUE) %>%  
  lapply(FUN = read_csv, col_types = "ififiddddcfcif") %>%  
  reduce(rbind)  
  
## Warning: Unnamed 'col_types' should have the same length as 'col_names'. Using smaller of the two.  
## Error in nchar(x): invalid multibyte string, element 2
```

I'm going to replace the column for time as a date data type, renaming it as simply "Time": DateParsing

```
data$Time <- data$Timestamp %>%
  parse_datetime(format = "%Y%m%d%H%M%S")

## Error in data$Timestamp: object of type 'closure' is not subsettable

data$Timestamp <- NULL

## Error in data$Timestamp <- NULL: object of type 'closure' is not subsettable
```

Wrangling A quick look at the data:

```
data
```

```
function(..., list = character(), package = NULL, lib.loc = NULL, verbose = getOption("verbose"),
  envir = .GlobalEnv, overwrite = TRUE) fileExt <- function(x) db <- grepl("
  .[.] +
  .(gz|bz2|xz)", x) ans <- sub(".*
  .", "", x) ans[db] <- sub(".*
  .([.] +
  .)(gz|bz2|xz)", "
  1
  2", x[db]) ans names <- c(as.character(substitute(list(...))[-1L]), list) if (!is.null(package)) if (is.character(package))
  stop("'package' must be a character string or NULL") if (any(package warning("datasets have been
  moved from package 'base' to package 'datasets'") if (any(package warning("datasets have been moved
  from package 'stats' to package 'datasets'")) package[package paths <- find.package(package, lib.loc,
  verbose = verbose) if (is.null(lib.loc)) paths <- c(path.package(package, TRUE), if (!length(package))
  getwd(), paths) paths <- unique(normalizePath(paths[file.exists(paths)])) paths <- paths[dir.exists(file.path(paths,
  "data"))] dataExts <- tools:::make_file_exts("data") if (length(names) == 0L) db <- matrix(character(), nrow = 0L, nc
  - file.path(paths, "data") for (name in names) found <- FALSE for (pinpaths) tmp_env <- if (overwrite) envirelse new
  "
  1", p) thispkg <- sub(".*", "", thispkg) thispkg <- paste0("package:", thispkg) objs <- rds[[name]] lazy-
  Load(file.path(p, "Rdata"), envir = tmp_env, filter = function(x) xbreakelse if (verbose) message(sprintf("name =
  name, paste(names(rds), collapse = ", ")), domain = NA) if (file_extest("-", file.path(p, "Rdata.zip"))) warning("zipp
  - files[grep(name, files, fixed = TRUE)] if (length(files) > 1L) o <- match(file_ext(files), dataExts, nomatch = 1
  basename(file), ":", appendLF = FALSE, domain = NA) ext <- fileExt(file) if (basename(file) !=
  paste0(name, "", ext)) found <- FALSE else found <- TRUE zfile <- file.zipname <- file.path(dirname(file),
  "Rdata.zip") if (file.exists(zipname)) Rdatadir <- tempfile("Rdata") dir.create(Rdatadir, showWarnings
  = FALSE) topic <- basename(file) rc <- .External(C_unzip, zipname, topic, Rdatadir, FALSE, TRUE, FALSE)
  0L) zfile <- file.path(Rdatadir, topic) if (zfile != file) on.exit(unlink(zfile)) switch(ext, R =, r =
  library("utils") sys.source(zfile, chdir = TRUE, envir = tmp_env), RData =, rdata =, rda = load(zfile, envir =
  tmp_env), TXT =, txt =, tab =, tab.gz =, tab.bz2 =, tab.xz =, txt.gz =, txt.bz2 =, txt.xz = assign(name, read.table(zfile,
  TRUE, as.is = FALSE), envir = tmp_env), CSV =, csv =, csv.gz =, csv.bz2 =, csv.xz = assign(name, read.table(zfile,
  TRUE, sep = ";", as.is = FALSE), envir = tmp_env), found <- FALSE) if (found) break if (verbose) message(if (!found)
  NOT*", "found", domain = NA) if (found) break if (!found) warning(gettextf("datasetdomain = NA") else if (!overwr
  bytecode : 0x7fa86617d708 >> environment : namespace : utils >
```

Let's first check for missing values: MissingValues

```
data %>% select_if(function(x) any(is.na(x))) %>% colnames()

## Error in UseMethod("tbl_vars"): no applicable method for 'tbl_vars' applied to an object
of class "function"
```

So only these columns contain missing values. We will use an explicit missing value for the location factor:

```
data$Location <- data$Location %>% fct_explicit_na()

## Error in data$Location: object of type 'closure' is not subsettable
```

Lets see the amount of unique values for each column: UniqueValueCount

```
data %>% apply(2, function(x) length(unique(x)))

## Error in apply(., 2, function(x) length(unique(x))): dim(X) must have a positive length
```

Both "ListID" and "ListName" don't give us any information, so we will remove them both.

```
data <- data %>% select(-ListID,-ListName)

## Error in UseMethod("select_"): no applicable method for 'select_' applied to an object of class "function"
```

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

```
data %>% select(`Material Description`, ItemName) %>%
  distinct() %$%
  table(ItemName) %>%
  as_tibble() %>%
  filter(n > 1)

## Error in UseMethod("select_"): no applicable method for 'select_' applied to an object of class "function"
```

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

```
data %>% select(`Material Description`, ItemName, Quantity) %>%
  filter(ItemName == "Rubber Gloves") %>%
  group_by(`Material Description`) %>%
  summarise(Quantity = sum(Quantity))

## Error in UseMethod("select_"): no applicable method for 'select_' applied to an object of class "function"
```

It seems that most rubber gloves are classified as plastic rather than rubber. I'm going to search for any extra descriptions given in the observations to try and gain some insight.

```
data %>% select(`Material Description`, ItemName, Description) %>%
  filter(ItemName == "Rubber Gloves", !is.na(Description))

## Error in UseMethod("select_"): no applicable method for 'select_' applied to an object of class "function"
```

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 innacurate: the last two instances here have "Balloon" in the extra descriptions... why aren't they categorised as such?

3.3 Distribution of observed debris:

MaterialQuantities

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

## Error: <text>:4:0: unexpected end of input
## 2: group_by(`Material Description`) %>%
## 3: summarise(Quantity = sum(Quantity)) %>%
## ^
```

```
## Error: 'data' must be a data frame, or other object coercible by 'fortify()', not an
S3 object with class uneval## Did you accidentally pass 'aes()' to the 'data' argument?
```

Figure 1: 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:

```
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

## Error in UseMethod("select_"): no applicable method for 'select_' applied to an object
of class "function"

tm

## Error in eval(expr, envir, enclos): object 'tm' not found

dev.off()

## pdf
## 2

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

Cigarettes are the most common item recorded. 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?

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", al
    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"))

## Error in UseMethod("select_"): no applicable method for 'select_' applied to an object
of class "function"

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

```

There seems to be a strong bias towards North America in our dataset. We will try a logarithmic plot to see things more clearly:

```

data %>%
  select(Latitude, Longitude, Quantity, Location, `Material Description`) %>%
  ggplot() +
    geom_polygon(data = map_data("world"), aes(x = long, y = lat, group = group), fill = "grey", al
    geom_hex(aes(x = Longitude, y = Latitude, fill = stat(log(count))), bins = 50) +
    scale_fill_viridis() +
    theme_void()

## Error in UseMethod("select_"): no applicable method for 'select_' applied to an object
of class "function"

```

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.

```

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")

## Error in UseMethod("mutate_"): no applicable method for 'mutate_' applied to an object
of class "function"

```

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? Let's look:

```

topLocations <- data %>%
  group_by(Location) %>%
  summarise(sumQuantity = sum(Quantity)) %>%
  arrange(desc(sumQuantity)) %>%
  top_n(5, sumQuantity)

```

```
## Error in UseMethod("group_by-"): no applicable method for 'group_by-' applied to an
object of class "function"

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()

## Error in UseMethod("filter-"): no applicable method for 'filter-' applied to an object
of class "function"
```

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.

Question: Are observed plastic item proportions time invariant?

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

## Error in UseMethod("filter-"): no applicable method for 'filter-' applied to an object
of class "function"

data$ItemName <- factor(data$ItemName, plastic_ordered$ItemName)

## Error in data$ItemName: object of type 'closure' is not subsettable

data %>%
  filter(`Material Description` == "PLASTIC") %>%
  mutate(month = month(Time, label = TRUE),
         year = as.integer(year(Time))) %>%
  filter(year > 2010) %>%
  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 = 2) +
  scale_fill_hue(l=50, c=150) +
  xlab("Month") +
  ylab("Proportion of Items")

## Error in UseMethod("filter-"): no applicable method for 'filter-' applied to an object
of class "function"

#scale_fill_viridis_d(option = "magma")

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


4th July and Firework link? (Karen's Idea)

```
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")

## Error in UseMethod("filter_"): no applicable method for 'filter_' applied to an object
## of class "function"

ggsave("plots/fireworks.png")

## Saving 7 x 7 in image
```

```
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))

## Error in UseMethod("mutate_"): no applicable method for 'mutate_' applied to an object
## of class "function"

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

```
data %>%
  mutate(year = year(Time)) %>%
  filter(year > 2010, year < 2020) %>%
  group_by(year) %>%
  summarise(quan = sum(Quantity)) %>%
  ggplot(aes(x = year, y = quan)) +
    geom_line() +
    geom_point() +
    scale_x_continuous(breaks = 2011:2019) +
    xlab("Year") +
    ylab("Quantity") +
    ggtitle("Quantity of debris per year") +
    theme(panel.grid.major = element_blank(),
          panel.grid.minor = element_blank())
```

```
## Error in UseMethod("mutate_"): no applicable method for 'mutate_' applied to an object
of class "function"
```

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

```
## Saving 7 x 7 in image
```

Recategorisation

```
# 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)
```

```
## Error in UseMethod("filter_"): no applicable method for 'filter_' applied to an object
of class "function"
```

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

```
## Error in UseMethod("filter_"): no applicable method for 'filter_' applied to an object
of class "function"
```

```
plastic %>%
  mutate(month = month(Time, label = FALSE),
         year = as.integer(year(Time))) %>%
  filter(year > 2010) %>%
  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()) +
guides(fill=guide_legend(title="Category"))

## Error in eval(lhs, parent, parent): object 'plastic' not found

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

```

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.

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

4.4 Item Pairing

(e.g. are 6-pack beer rings observed at the same time as fireworks?)

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

5.2 Model Evaluation

5.3 Model Results

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](#)

8.2 Project Progress

```
## Error: package 'XLConnectJars' could not be loaded
## 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 'XLConnectJars' could not be loaded
## 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
```

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 2:

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

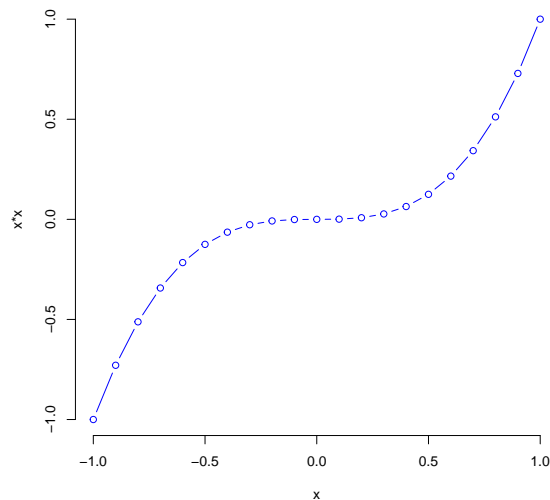


Figure 2: Simple Plot of $f(x) = x^3$ Function

8.5 Experiments

Now we can show how the function $f(x) = x^2$ looks like (Figure 3)

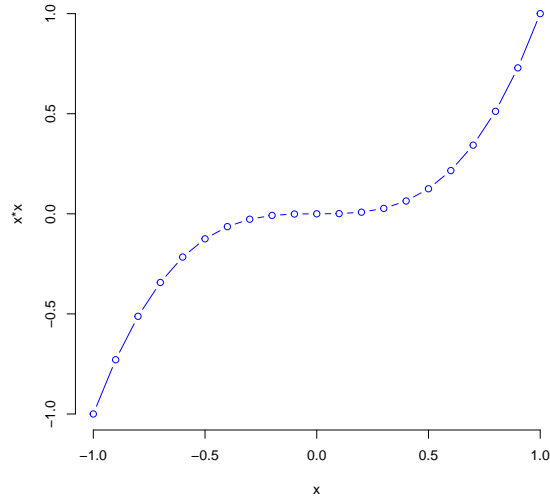


Figure 3: Simple Plot of $f(x) = x^3$ Function

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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1109/IJCNN.2018.8489087).