PROJECT 01 REPORT: Trash Wheel Collection Data

Vo Diep Nhu, Vo Phi Son, Nguyen Minh Tuan

2024-03-30

INTRODUCTION

The dataset from the Mr. Trash Wheel Baltimore Healthy Harbor initiative plays a crucial role in combating environmental degradation by removing trash from waterways. This initiative underscores the paramount importance of environmental conservation. However, working with this dataset may present challenges like data inconsistencies or gaps, requiring meticulous preprocessing and analytical methodologies for accurate insights. The dataset contains diverse data attributes such as trash weight, volume, and specific item counts like plastic bottles and cigarette butts, offering a rich landscape for exploration and visualization. Techniques like line charts, bar charts, and box plots enable temporal and comparative analyses, identifying trends, seasonal variations, and disparities in Trash Wheel. Additionally, the unique attribute "HomesPowered" highlights the potential renewable energy benefits of waste management efforts, reinforcing the significance of recycling and resource recovery in environmental sustainability endeavors. Overall, this dataset offers an invaluable opportunity to explore environmental issues through data visualization, fostering a deeper understanding of the complex dynamics between human activity and environmental well-being. Our group employs two questions and various charts, including bar charts, line charts, and box plots, to interpret the data's significance.

The code and report for the project are available on the GitHub repository: project01.

Question 1

How does the composition of collected trash in the Trash Wheel vary over time (Month and Year)?

Perspective

Understanding the composition of collected trash over time (Month and Year) is pivotal for comprehensively addressing environmental concerns. This analysis aims to explore the relationship between the quantities of plastic bottles, polystyrene items, and glass bottles collected in various months and years, seeking to uncover patterns and variations in waste composition. To address this question, essential components of the dataset include the quantities of plastic bottles, polystyrene items, and glass bottles collected, and the temporal attributes of Month and Year, enabling us to track the distribution and fluctuations of different types of waste over time. This question is of significant interest due to its direct relevance to environmental sustainability and waste management practices, as understanding how the composition of collected trash provides valuable insights into the sources and distribution of pollutants in water bodies. Moreover, uncovering trends and patterns in waste composition can inform targeted interventions and policy initiatives aimed at mitigating pollution and promoting ecosystem health, contributing meaningfully to ongoing efforts towards environmental conservation and sustainable waste management.

Approach

To address our question, we will employ a variety of plot types to glean insights into the composition and trends of collected waste over time. Initially, we will undertake two steps: preprocessing the data to create new columns tailored to address our inquiry and utilizing line charts and bar charts to elucidate temporal trends of waste types by month and year.

Firstly, a line chart will illustrate the temporal trends of weight and volume of collected waste across different years. It is ideal for showcasing continuous trends over time, allowing observation of fluctuations and patterns in waste quantities. By visually representing changes in weight and volume over time, significant peaks, valleys, or overall trends in waste accumulation can be discerned.

Furthermore, additional plots such as stacked area plots will compare the composition of collected trash by month. Stacked area plots effectively convey the cumulative composition of waste types over time, enabling easy comparison of proportions among different categories within each month. Additionally, a relative frequency bar chart encoded with color mapping will display the percentage of each waste type over the months, allowing identification of shifts in waste composition over time.

Lastly, we will investigate the trend of weight and volume of collected waste specifically in the year 2018, utilizing a combination of a line chart and a bar chart. This approach will provide a holistic view of waste trends in 2018, leveraging the strengths of both chart types to depict variations in weight and volume across different months. By employing a diverse range of plot types, we aim to uncover nuanced insights into the dynamics of collected waste over time, facilitating informed decision-making and environmental management strategies.

Analysis

To answer the question 1, the two following steps are taken:

- Step 1: We clear the data and create new columns to directly answer the question.
- Step 2: We plot the line chart and bar chart to describe the trend of types of waste by month and year.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

We load the data from the github and fill the missing values in the HomesPowered column by using the formula: $HomePowered = \frac{Weight * 500}{30}$, and then remove the missing values.

```
# Load the data from the github source
trashwheel <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data
```

```
## Rows: 993 Columns: 16
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): ID, Name, Month, Date
## dbl (12): Dumpster, Year, Weight, Volume, PlasticBottles, Polystyrene, Cigar...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# change the value of HomesPowered = Weight * 500 / 30
trashwheel <- trashwheel %>%
    mutate(HomesPowered = round(Weight * 500 / 30, 0.5))

# remove the missing (NA) values
trashwheel <- na.omit(trashwheel)

# len of the data
nrow(trashwheel)
```

```
## [1] 629
```

```
# print unique values of the Name
unique(trashwheel$Name)
```

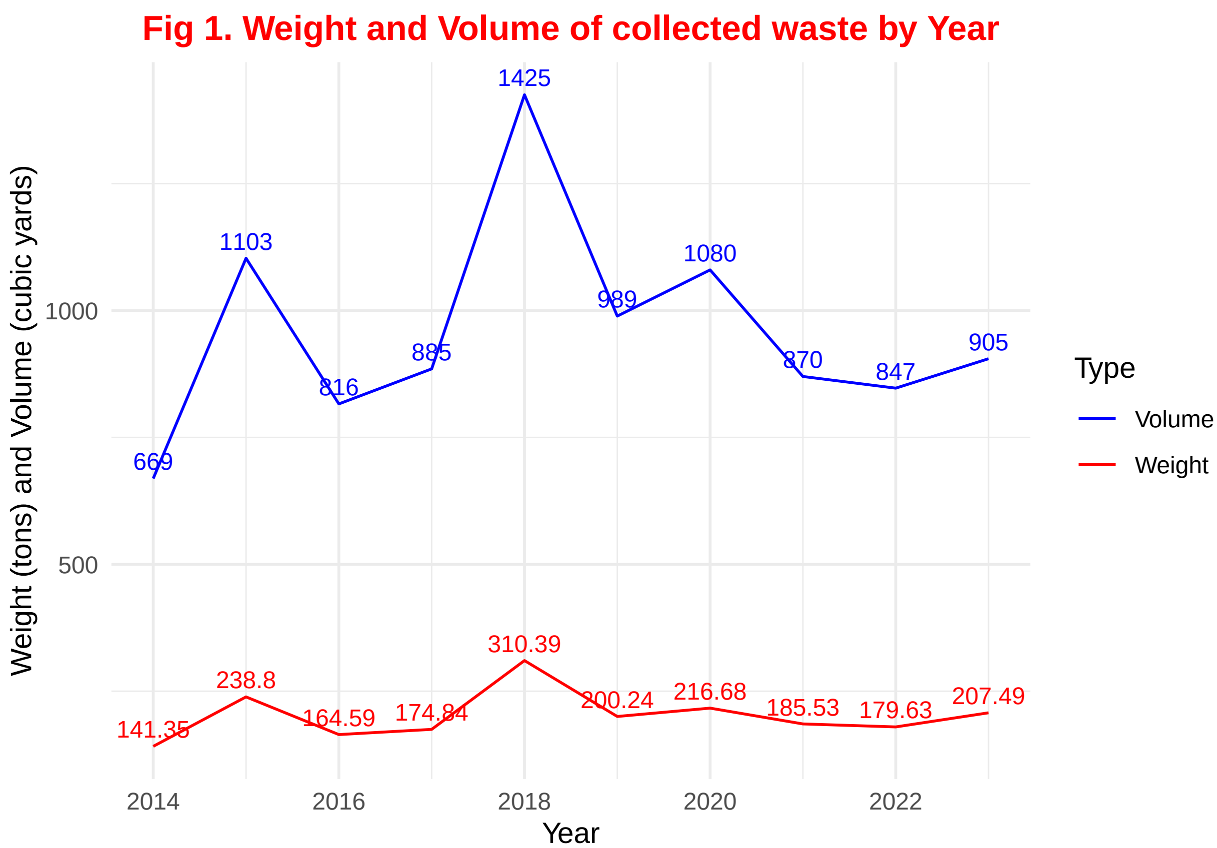
```
## [1] "Mister Trash Wheel"
```

We want to investigate the trend of weight (in tons) and volume (in cubic yards) of wastes which were collected during the year from 2014 to 2022. Then, we sum the weight and volume of the wastes with grouping by year.

```
sum_trashwheel <- trashwheel %>%
    group_by(Year) %>%
    summarise(sum_Weight = sum(Weight), sum_Volume = sum(Volume), sum_PlasticBottles = sum(PlasticBottl
```

After that, we plot the weight and voulume of collected waste in year.

```
# show the Volume and Weight by year
ggplot(sum_trashwheel,
    aes(x = Year)) +
    geom_line(aes(x = Year, y = sum_Weight, color = "Weight")) +
    geom_line(aes(x = Year, y = sum_Volume, color = "Volume")) +
    geom_text(aes(x = Year, y = sum_Weight, label = sum_Weight), hjust = 0.5, vjust = -0.5, color = "re
    geom_text(aes(x = Year, y = sum_Volume, label = sum_Volume), hjust = 0.5, vjust = -0.5, color = "bl
    labs(title = "Fig 1. Weight and Volume of collected waste by Year", x = "Year", y = "Weight (tons) a
    scale_color_manual(values = c("Weight" = "red", "Volume" = "blue")) +
    theme_minimal() +
    theme(plot.title = element_text(color = "red", face = "bold", hjust = 0.5))
```



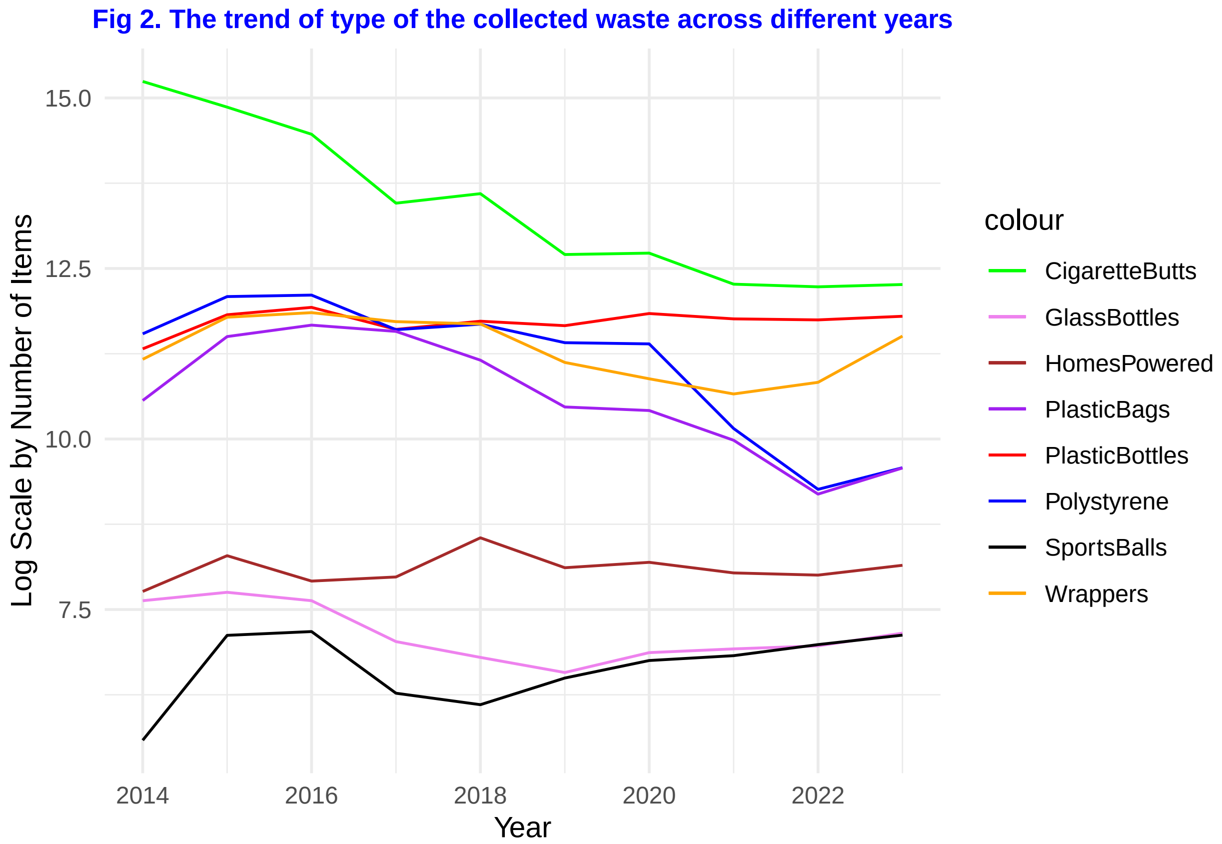
Discussion 01

The Fig 1 shows that the volume of collected waste has increased significantly over the years, while the weight of collected waste has fluctuated but shows a general downward trend. This suggests that the composition of the waste is changing. Perhaps people are throwing away more bulky items that take up more space but weigh less, or there are efforts underway to reduce denser materials like food scraps or metals.

We create a line plot to visualize the trend of type of the collected waste across different years.

```
# line chart to show the sum of PlasticBottles, Polystyrene, CigaretteButts, GlassBottles,
# PlasticBags, Wrappers, SportsBalls, HomesPowered by year, log scale for y axis

ggplot(sum_trashwheel,
    aes(x = Year)) +
    geom_line(aes(x = Year, y = log(sum_PlasticBottles), color = "PlasticBottles")) +
    geom_line(aes(x = Year, y = log(sum_Polystyrene), color = "Polystyrene")) +
    geom_line(aes(x = Year, y = log(sum_CigaretteButts), color = "CigaretteButts")) +
    geom_line(aes(x = Year, y = log(sum_GlassBottles), color = "GlassBottles")) +
    geom_line(aes(x = Year, y = log(sum_PlasticBags), color = "PlasticBags")) +
    geom_line(aes(x = Year, y = log(sum_Wrappers), color = "Wrappers")) +
    geom_line(aes(x = Year, y = log(sum_SportsBalls), color = "SportsBalls")) +
    geom_line(aes(x = Year, y = log(sum_HomesPowered), color = "HomesPowered")) +
    labs(title = "Fig 2. The trend of type of the collected waste across different years",
        x = "Year",
        y = "Log Scale by Number of Items",
        legend.title = "Type"
```

```r
    ) +
    scale_color_manual(values = c("PlasticBottles" = "red", "Polystyrene" = "blue", "CigaretteButts" =
    theme_minimal() +
    theme(plot.title = element_text(color = "blue", face = "bold", hjust = 0.5, size = 10))
```



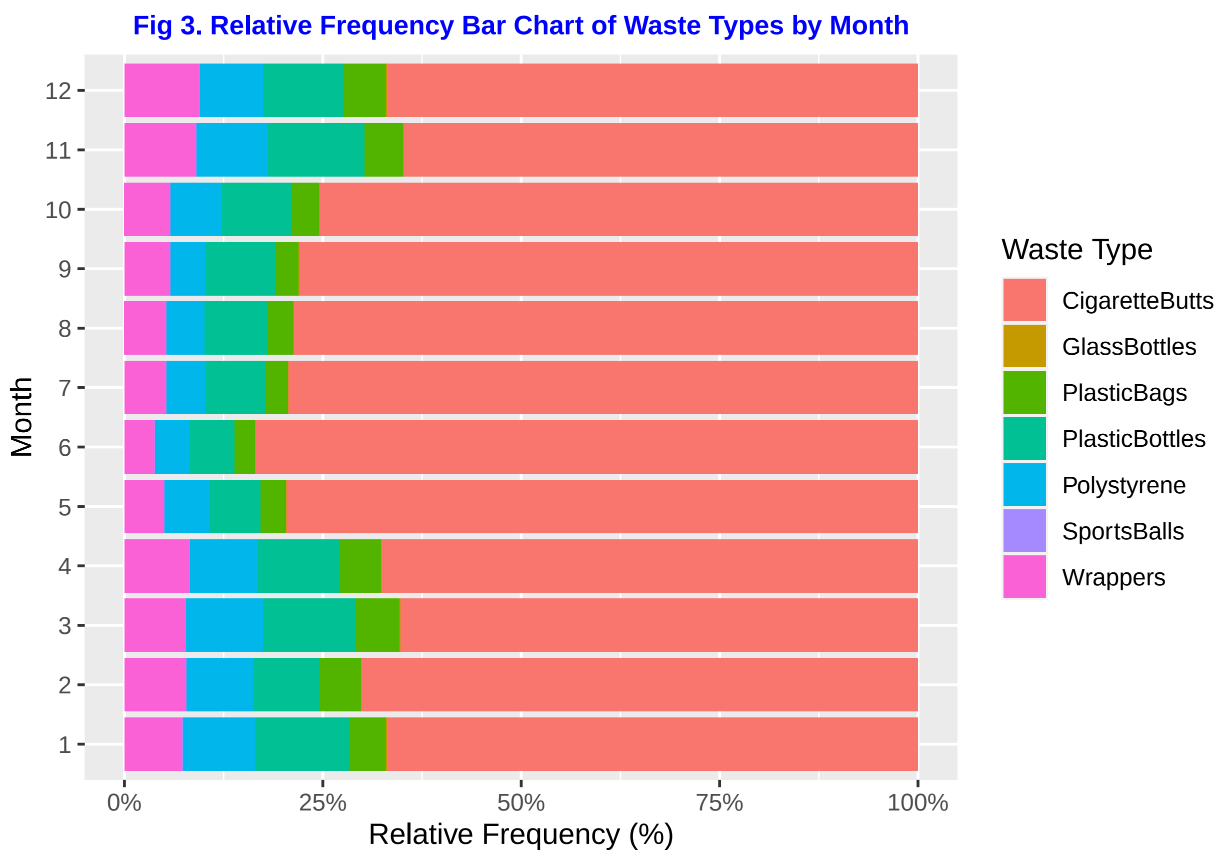
Discussion 02

The Figure 2 illustrates that the number of cigarette butts, polystyrene, and plastic bags has decreased significantly, while the trend of other types of waste has increased slightly over the years from 2014 to 2022. This could be attributed to a number of factors, such as increased use of reusable shopping bags or bans on plastic bags. Additionally, we create additional plots such as stacked area plots to compare the composition of collected trash by month.

```r
# show the sum of Weight, Volume for PlasticBottles, Polystyrene, CigaretteButts, GlassBottles, Plastic
sum_trashwheel_month_all <- trashwheel %>%
    group_by(Month) %>%
    summarise(sum_Weight = sum(Weight), sum_Volume = sum(Volume),
              sum_PlasticBottles = sum(PlasticBottles), sum_Polystyrene = sum(Polystyrene), sum_Cigaret
              sum_GlassBottles = sum(GlassBottles), sum_PlasticBags = sum(PlasticBags),
              sum_Wrappers = sum(Wrappers), sum_SportsBalls = sum(SportsBalls), sum_HomesPowered = sum(

# Arrange the data by month before we plot
trash_month_all <- sum_trashwheel_month_all %>%
 mutate(Month = factor(Month, levels = c("January", "February", "March", "April", "May", "June", "July"
    arrange(Month)
```

To create a relative frequency bar chart where the waste types are encoded using color, with the month on the y-axis and the percentage of each waste type on the x-axis, we can use the geom_col() function in ggplot2 along with some data manipulation.

```r
# library(dplyr)
# library(ggplot2)

# Calculate relative frequencies
trash_month_all <- trash_month_all %>%
  mutate(total = sum_PlasticBottles + sum_Polystyrene + sum_CigaretteButts +
           sum_GlassBottles + sum_PlasticBags + sum_Wrappers + sum_SportsBalls) %>%
  mutate(rel_freq_PlasticBottles = sum_PlasticBottles / total,
         rel_freq_Polystyrene = sum_Polystyrene / total,
         rel_freq_CigaretteButts = sum_CigaretteButts / total,
         rel_freq_GlassBottles = sum_GlassBottles / total,
         rel_freq_PlasticBags = sum_PlasticBags / total,
         rel_freq_Wrappers = sum_Wrappers / total,
         rel_freq_SportsBalls = sum_SportsBalls / total)

# Reshape the data to long format and remove NA values from Month variable
trash_long <- trash_month_all %>%
  select(Month, rel_freq_PlasticBottles, rel_freq_Polystyrene, rel_freq_CigaretteButts,
         rel_freq_GlassBottles, rel_freq_PlasticBags,
         rel_freq_Wrappers, rel_freq_SportsBalls) %>%
  na.omit() %>%
  pivot_longer(cols = -Month, names_to = "Waste_Type", values_to = "Relative_Frequency")

# remove "rel_freq_" pattern in Waste_Type column
trash_long$Waste_Type <- gsub("rel_freq_", "", trash_long$Waste_Type)

# Plot the bar chart
ggplot(trash_long, aes(x = Relative_Frequency, y = Month, fill = Waste_Type)) +
  geom_col(position = "stack") +
  scale_x_continuous(labels = scales::percent_format(scale = 100)) +
  labs(x = "Relative Frequency (%)", y = "Month", fill = "Waste Type") +
  ggtitle("Fig 3. Relative Frequency Bar Chart of Waste Types by Month") +
  theme(legend.position = "right", plot.title = element_text(size = 10)) +
  theme(plot.title = element_text(color = "blue", face = "bold", hjust = 0.5))
```



Discussion 03

The Figure 3 illustrates the relationship between the types of collected waste over the months. The frequency of collected cigarette butts was the highest among all types over the months, while sports balls or glass bottles were rarely collected. This suggests that people often discard cigarette butts after using them. Additionally, the frequency of collected cigarette butts increases from May to September, while the frequency of other types of waste increases from January to April. However, the number of items cannot show the trend of weight and volume over the months. Therefore, we will investigate the trend of weight and volume of collected waste over the months in 2018 by using a combination of line chart and bar chart. The reason we choose 2018 is that it is the year with the highest number of collected waste.

```
# show the sum of Weight, Volume for PlasticBottles, Polystyrene, CigaretteButts, GlassBottles, Plastic
sum_trashwheel_month <- trashwheel %>%
    group_by(Year, Month) %>%
    summarise(sum_Weight = sum(Weight), sum_Volume = sum(Volume), sum_PlasticBottles = sum(PlasticBottl
```

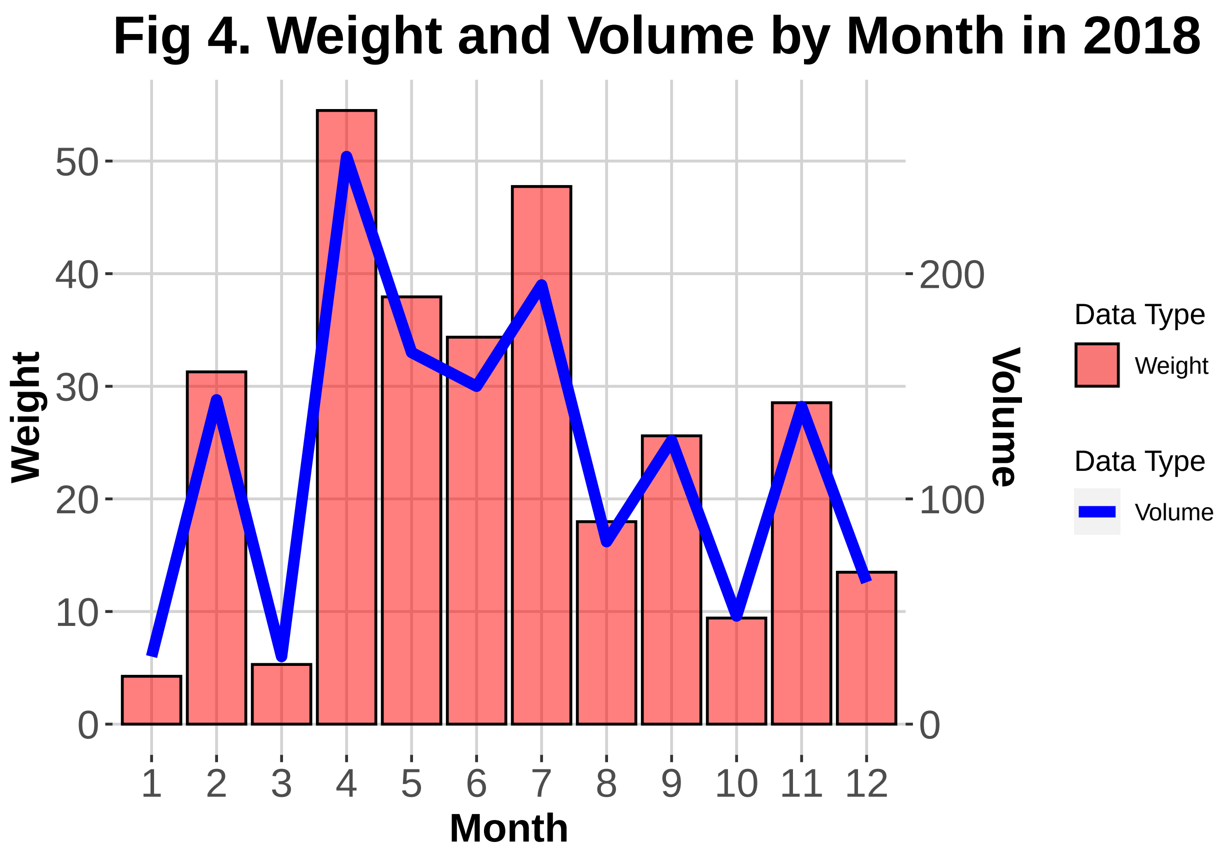
```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```

```
trash_month_2018 <- sum_trashwheel_month %>%
    filter(Year == 2018)
# shorten month name, e.g. January -> Jan and sort by month
trash_month_2018 <- sum_trashwheel_month %>%
    filter(Year == 2018) %>%
```

```r
    mutate(Month = factor(Month, levels = c("January", "February", "March", "April", "May", "June", "Jul
    arrange(Month)
```

```r
ggplot(trash_month_2018, aes(x = Month)) +
    geom_bar(aes(y = sum_Weight, fill = "Weight"), stat = "identity", color = "black", alpha = 0.5) +  
    geom_line(aes(y = sum_Volume / 5, color = "Volume"), group = 1, size = 2) +  # Line for volume
    scale_y_continuous(
        name = "Weight",
        sec.axis = sec_axis(~ . * 5, name = "Volume"),
        breaks = seq(0, 60, 10)
    ) +
    labs(title = "Fig 4. Weight and Volume by Month in 2018", x = "Month", y = "Weight") +
    scale_color_manual(values = c("Volume" = "blue"), labels = c("Volume")) +  # Legend for volume
    scale_fill_manual(values = c("Weight" = "red"), labels = c("Weight")) +  # Legend for weight
    theme(
        axis.text = element_text(size = 15),
        axis.title = element_text(size = 15, face = "bold"),
        plot.title = element_text(size = 20, face = "bold"),
        panel.background = element_rect(fill = NA),  # Remove plot background
        panel.grid.major = element_line(color = "lightgrey", linetype = "solid"),
    ) +
    guides(fill = guide_legend(title = "Data Type"), color = guide_legend(title = "Data Type"))
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



Discussion 04

The Figure 4 illustrates that the weight and volume of the collected waste from April to June are higher than in other months. Furthermore, the trends of weight and volume remain consistent across the months in 2018.

Question 2

What is the seasonal pattern of trash accumulation in terms of weight (Weight), volume (Volume), and the number of plastic bags (PlasticBags)?

Perspective

The investigation into seasonal trash accumulation patterns, including weight, volume, and plastic bag count, is crucial for understanding waste management dynamics over time. Key dataset variables like Month, Year, Weight, Volume, and PlasticBags are essential for unraveling the relationship between seasonal fluctuations and trash collection trends. By analyzing variations in these parameters across months and years, we aim to uncover patterns revealing the driving factors behind trash accumulation.

Understanding seasonal trash dynamics is vital for environmental stewardship and sustainable waste management. Exploring weight, volume, and plastic bag counts over different timeframes helps discern temporal trends and the influence of seasons on waste generation. Insights from this analysis can inform interventions and policies to mitigate environmental impacts, promoting healthier and more sustainable communities and ecosystems.

Approach

To unveil seasonal trash accumulation patterns in weight, volume, and plastic bag count, we organize the dataset by quarter and year to streamline analysis, focusing on plastic bottle accumulation trends. Bar charts illustrate temporal plastic bottle trends from 2016 to 2022, revealing any seasonal fluctuations. Extending the analysis to weight and volume, additional bar charts depict dynamics across quarters. These visuals provide a comprehensive overview of trash accumulation trends.

To explore interrelationships between trash types, we analyze potential correlations using a heatmap in 2015. Color mapping represents correlation strength, aiding in identifying significant associations. Additionally, we examine weight distribution across quarters with a box plot, offering insights into variability and central tendency. Through this analytical approach combining bar charts, heat maps, and box plots, we gain a comprehensive understanding of seasonal trash dynamics, including weight, volume, and plastic bag prevalence over time.

Analysis

Firstly, we want to visualize the seasonal pattern of trash accumulation for Plastic Bottles. Therefore, we create the data.

```
# show the sum of Weight, Volume for PlasticBottles, Polystyrene, CigaretteButts,
# GlassBottles, PlasticBags, Wrappers, SportsBalls, HomesPowered by group of (year, quarter)
# the month in text "January", "February", "March", "April", "May", "June", "July", "August", "Septembe
trashwheel <- trashwheel %>%
    mutate(quarter = ifelse(Month %in% c("January", "February", "March"), "Q1",
                            ifelse(Month %in% c("April", "May", "June"), "Q2",
                                   ifelse(Month %in% c("July", "August", "September"), "Q3", "Q4"))))

sum_trashwheel_quarter <- trashwheel %>%
    group_by(Year, quarter) %>%
    summarise(sum_Weight = sum(Weight), sum_Volume = sum(Volume), sum_PlasticBottles =
                sum(PlasticBottles), sum_Polystyrene = sum(Polystyrene), sum_CigaretteButts = sum(Cigar
                sum(GlassBottles), sum_PlasticBags = sum(PlasticBags), sum_Wrappers = sum(Wrappers), su
                sum(SportsBalls), sum_HomesPowered = sum(HomesPowered))
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```

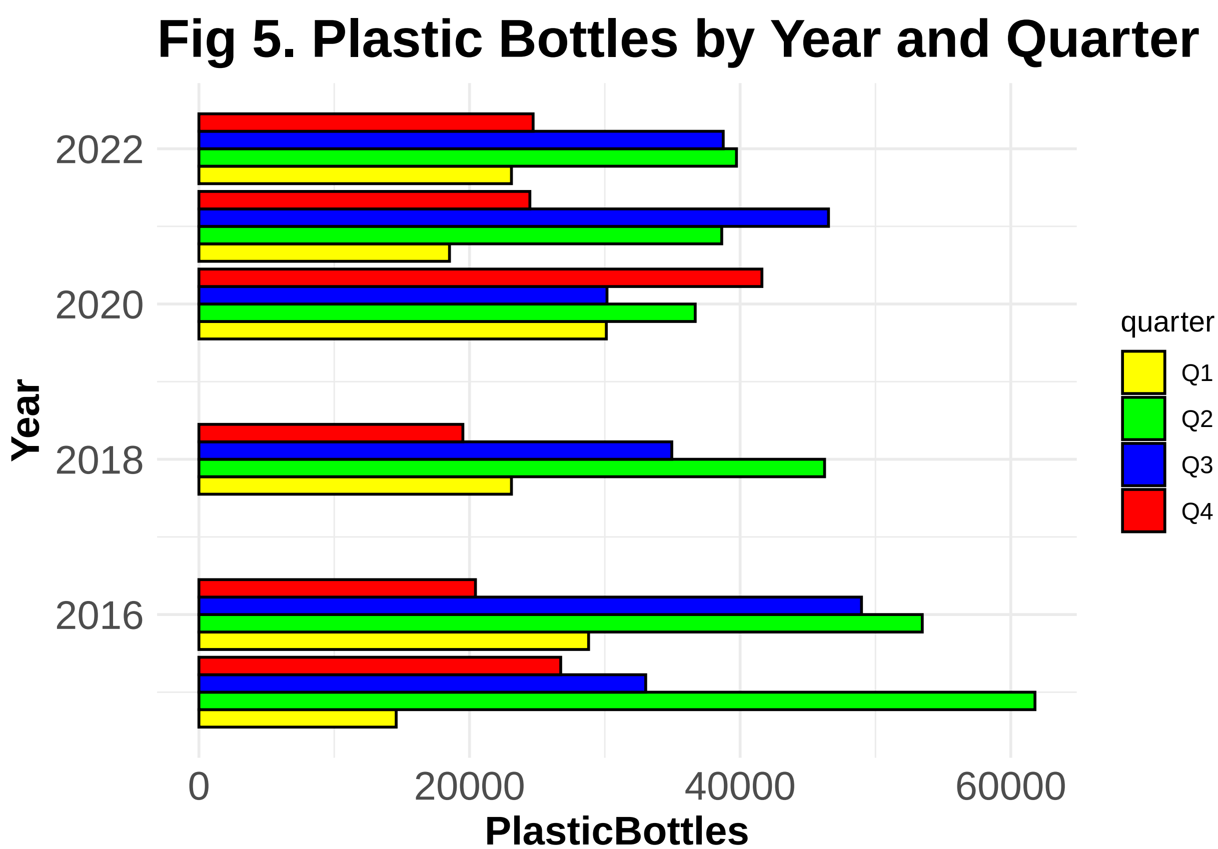
We remove the data in 2017, 2019, and 2023 because in some months of these years, there is no data.

```
# only select data in year_list
year_list <- c(2015, 2016, 2018, 2020, 2021, 2022)

sec_year_trashwheel <- trashwheel %>%
    filter(Year %in% year_list)

trashwheel_quater <- sec_year_trashwheel %>%
    mutate(quarter = ifelse(Month %in% c("January", "February", "March"), "Q1", ifelse(Month %in% c("Ap

sum_trashwheel_quarter <- trashwheel_quater %>%
    group_by(Year, quarter) %>%
    summarise(sum_Weight = sum(Weight), sum_Volume = sum(Volume), sum_PlasticBottles = sum(PlasticBottl
```

```
## ‘summarise()‘ has grouped output by ’Year’. You can override using the
## ‘.groups‘ argument.
```

Plot the bar chart to show the trend of the collected plastic bottles over the seasons from 2016 to 2022.

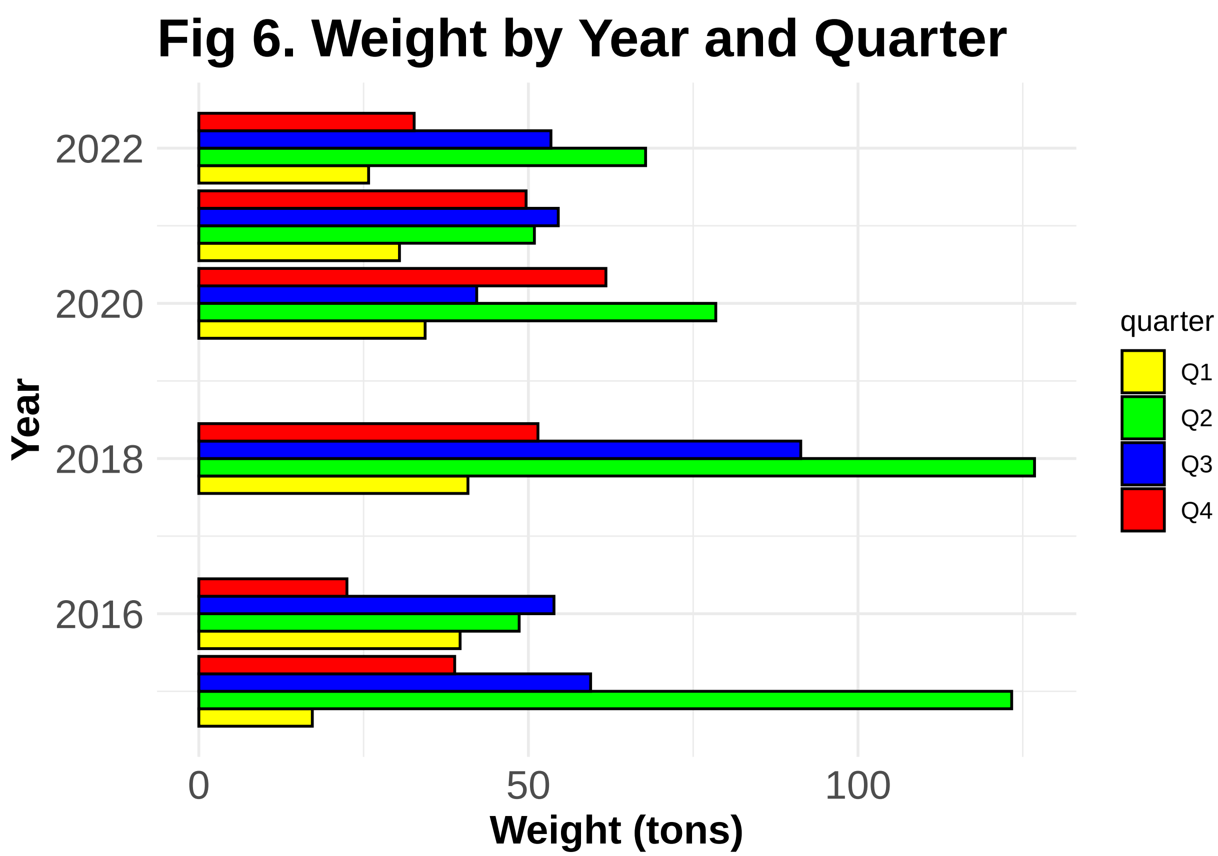
```
# show PlasticBottles by year and quarter
ggplot(sum_trashwheel_quarter, aes (x = Year, y = sum_PlasticBottles, fill = quarter)) +
    geom_col(position = "dodge", color = "black") +
    labs(title = "Fig 5. Plastic Bottles by Year and Quarter", x = "Year", y = "PlasticBottles") +
    theme_minimal() +
    theme(
        axis.text = element_text(size = 15),
        axis.title = element_text(size = 15, face = "bold"),
        plot.title = element_text(size = 20, face = "bold"),
    ) +
    scale_fill_manual(values = c("Q4" = "red", "Q3" = "blue", "Q2" = "green", "Q1" = "yellow")) +
    coord_flip()
```



Discussion 01

The Fig 5. illustrates the trend of the number of collected waste by year and quarter. The number of collected waste is highest during the summer months (depicted in green), while it is at its lowest during winter (Q4). Additionally, there is a decrease in the number of collected waste observed over the years from 2016 to 2022. We also plot the bar chart to show the weight and volume of collected waste over the quarter from 2016 to 2022.

```
# show Weight by year and quarter
ggplot(sum_trashwheel_quarter, aes (x = Year, y = sum_Weight, fill = quarter)) +
    geom_col(position = "dodge", color = "black") +
    labs(title = "Fig 6. Weight by Year and Quarter", x = "Year", y = "Weight (tons)", color = "Type") +
    theme_minimal() +
    theme(
        axis.text = element_text(size = 15),
        axis.title = element_text(size = 15, face = "bold"),
        plot.title = element_text(size = 20, face = "bold")
    ) +
    scale_fill_manual(values = c("Q4" = "red", "Q3" = "blue", "Q2" = "green", "Q1" = "yellow")) +
    coord_flip()
```

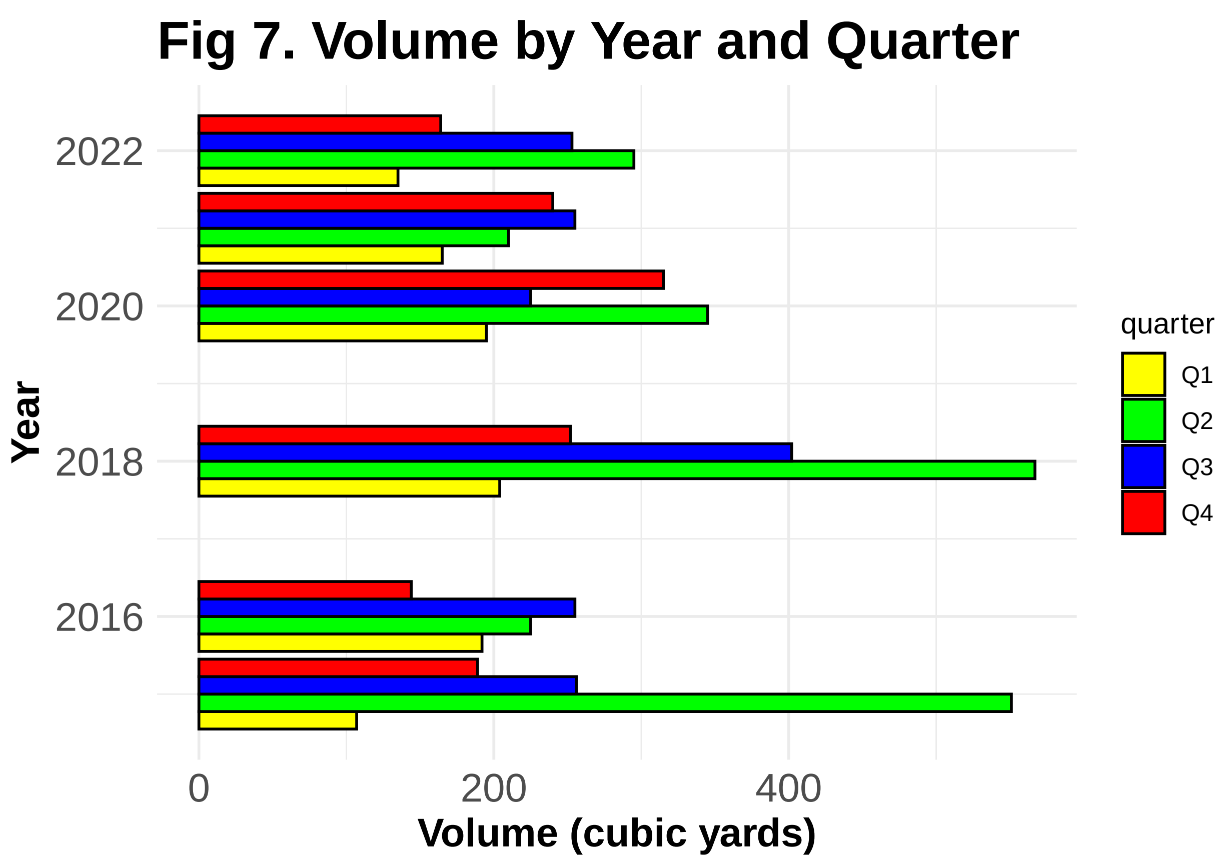


Discussion 02

Fig. 6 illustrates that the weight of collected waste is highest in the summer (Q2) and autumn (Q3). Conversely, the weight of collected waste is lowest in the spring. Additionally, there is a significant decrease in the weight of collected waste observed from 2016 to 2022. We also plot the bar chart to show the volume of collected waste over the quarter from 2016 to 2022.

```
# show Volume by year and quarter
ggplot(sum_trashwheel_quarter, aes (x = Year, y = sum_Volume, fill = quarter)) +
    geom_col(position = "dodge", color = "black") +
    labs(title = "Fig 7. Volume by Year and Quarter", x = "Year",
```

```
      y = "Volume (cubic yards)", color = "Type") +
    theme_minimal() +
    theme(
        axis.text = element_text(size = 15),
        axis.title = element_text(size = 15, face = "bold"),
        plot.title = element_text(size = 20, face = "bold")
    ) +
    scale_fill_manual(values = c("Q4" = "red", "Q3" = "blue", "Q2" = "green", "Q1" = "yellow")) +
    coord_flip()
```



Discussion 03

The Fig 7 shows that the trend of volume of collected waste decreases sharply over the 6 years from 2016 to 2022. The volume of collected waste is highest during the summer and lowest in the spring. Determine if there are significant correlations between different types of trash accumulation. We create a heatmap to visualize the correlation matrix between different types of trash (e.g., plastic bottles, polystyrene items, glass bottles, cigarette butts, wrappers, and sports balls) in 2015.

```
# Sample data (same as example 1)
# data <- data.frame(x = c(1, 2, 3, 4, 5), y = c(2, 4, 5, 4, 5))

# # Calculate Spearman's rank correlation coefficient
# correlation <- cor(data$x, data$y, method = "spearman")
```

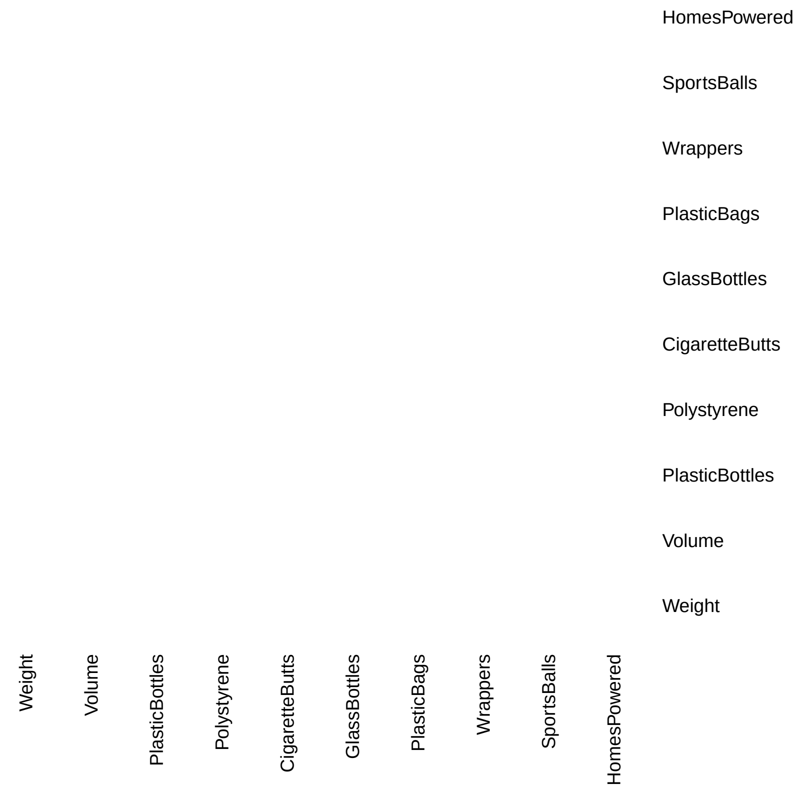
```r
# # Print the correlation coefficient
# cat("Spearman's rank correlation coefficient between x and y:", correlation)
```

```r
# sum all type of trash by year, remove the quater first and then do sum

# sum_trashwheel_quarter <- trashwheel_quater %>%
#     group_by(Year, quarter) %>%
#     summarise(sum_Weight = sum(Weight), sum_Volume = sum(Volume), sum_PlasticBottles = sum(PlasticBot
data_cor <- trashwheel_quater %>%
    select(-quarter, -Year) %>%
    summarise(sum_Weight = sum(Weight), sum_Volume = sum(Volume), sum_PlasticBottles = sum(PlasticBottl
    rename_all(~ gsub("sum_", "", .)) %>%
    cor()
# data_cor <- cor(sum_trashwheel_year)


# data_sum <- sum_trashwheel_quarter %>% cor()

# Set the size of the main title and axis labels
par(cex.main = 0.8, cex.axis = 0.8, cex.lab = 0.8)

# Plotting heatmap with a small font size
heatmap(data_cor, Rowv = NA, Colv = NA, col = cm.colors(256), scale = "none", cexRow = 0.75, cexCol = 0
        main = "Fig 8. Heatmap of Different Types of Trash by Year and Quarter (2015)")
```

```
## Warning in min(x): no non-missing arguments to min; returning Inf
```

```
## Warning in max(x): no non-missing arguments to max; returning -Inf
```

Fig 8. Heatmap of Different Types of Trash by Year and Quarter (2015)



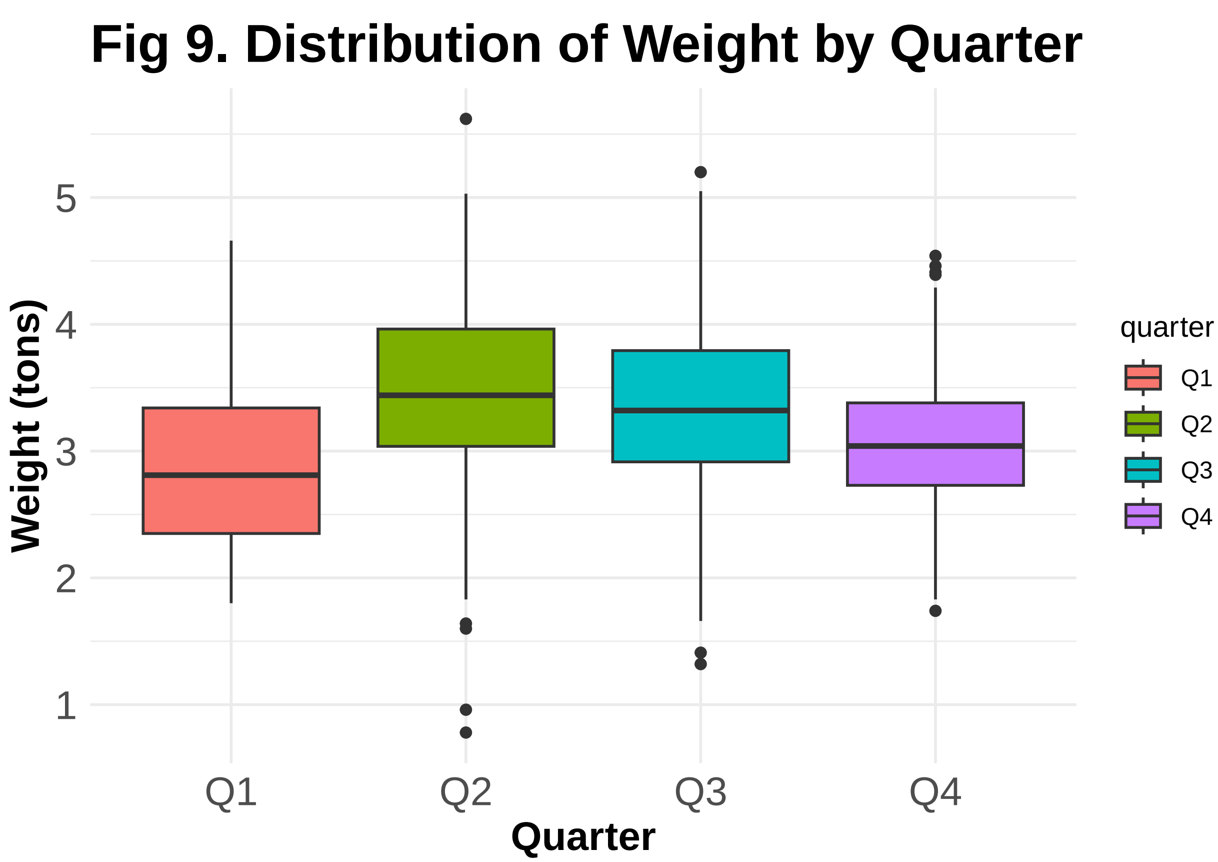

```
# # Prepare the data
# data_all_year <- sum_trashwheel_quarter %>%
#   group_by(Year) %>%
#   select(-quarter) %>%
#   # Remove "sum_" pattern in all column names
#   rename_all(~ gsub("sum_", "", .)) %>%
#   cor() %>%
#   as.data.frame() %>%
#   rownames_to_column(var = "Trash_Type")  # Convert row names to a column

# # Set the size of the main title and axis labels
# theme_set(theme_bw(base_size = 8))  # Adjust the base font size

# # Plotting heatmap with facet_wrap
# ggplot(data_all_year, aes(x = Trash_Type, y = Trash_Type, fill = value)) +
#   # geom_tile() +
#   scale_fill_gradient(low = "white", high = "blue") +
#   facet_wrap(~ Year, ncol = 2) +
#   labs(
#     title = "Fig 9. Heatmap of Different Types of Trash by Year",
#     x = NULL, y = NULL,
#     fill = "Correlation"
#   ) +
#   theme(
#     plot.title = element_text(hjust = 0.5),
#     strip.text = element_text(size = 8, face = "bold")
```

```
#   )
```

Discussion 04

The Fig 8 shows the correlation between different types of trash accumulation. The correlations between plastic bottles, polystyrene, cigarette butts, and sports balls are very strong, whereas the correlations between other collected waste are weak. Furthermore, strong correlations are observed between weight, volume, plastic bottles, polystyrene, cigarette butts, and homes powered. This suggests that these types of collected waste contribute significantly to power generation, and an increase in weight and volume will result in an increase in homes powered. We also investigate the distribution of weight of collected waste over the quarter by using the box plot.

```
# box plot to show the distribution of Weight by quarter
ggplot(trashwheel_quater, aes(x = quarter, y = Weight, fill = quarter)) +
    geom_boxplot() +
    labs(title = "Fig 9. Distribution of Weight by Quarter", x = "Quarter",
         y = "Weight (tons)", color = "Type") +
    theme_minimal() +
    theme(
        axis.text = element_text(size = 15),
        axis.title = element_text(size = 15, face = "bold"),
        plot.title = element_text(size = 20, face = "bold")
    )
```



Discussion 05

The weight of collected waste is highest in the first quarter (Q1) with a median of around 4 tons. The weight is lowest in the second quarter (Q2) with a median of around 1 ton. There seems to be more variability in the weight collected in Q3 compared to Q1 and Q4. The median for Q3 is at around 2 tons, however the whiskers extend further out than the other quarters which signifies a wider range of data points. Overall, the spread of the data in Q1 is smaller than Q3 and Q4, which means the weights in Q1 are more clustered around the median.

Conclusion

In conclusion, our analysis has shed light on two key aspects of trash collection: the variation in composition over time, and the seasonal patterns of trash accumulation in terms of weight, volume, and the number of plastic bags.

For the first question, we utilized a variety of visualizations including line charts, bar charts,dual Y-axis bar and line charts, and frequency bar charts to explore the relationship between the number of plastic bottles, polystyrene items, and glass bottles collected by each Trash Wheel across different months and years. Our analysis revealed significant variations in trash composition among different Trash Wheels, providing valuable insights into the dynamics of waste accumulation over time.

Regarding the second question, we employed bar charts, heatmaps, and box plots to investigate the seasonal patterns of trash accumulation. Our findings highlighted distinct seasonal trends, with fluctuations in weight, volume, and the number of plastic bags observed across different months and years. Additionally, the heatmap analysis revealed correlations between different types of waste, offering further insights into waste management strategies.

Overall, our comprehensive analysis has enhanced our understanding of trash collection dynamics, providing valuable insights that can inform waste management policies and practices aimed at optimizing collection efficiency and environmental sustainability.