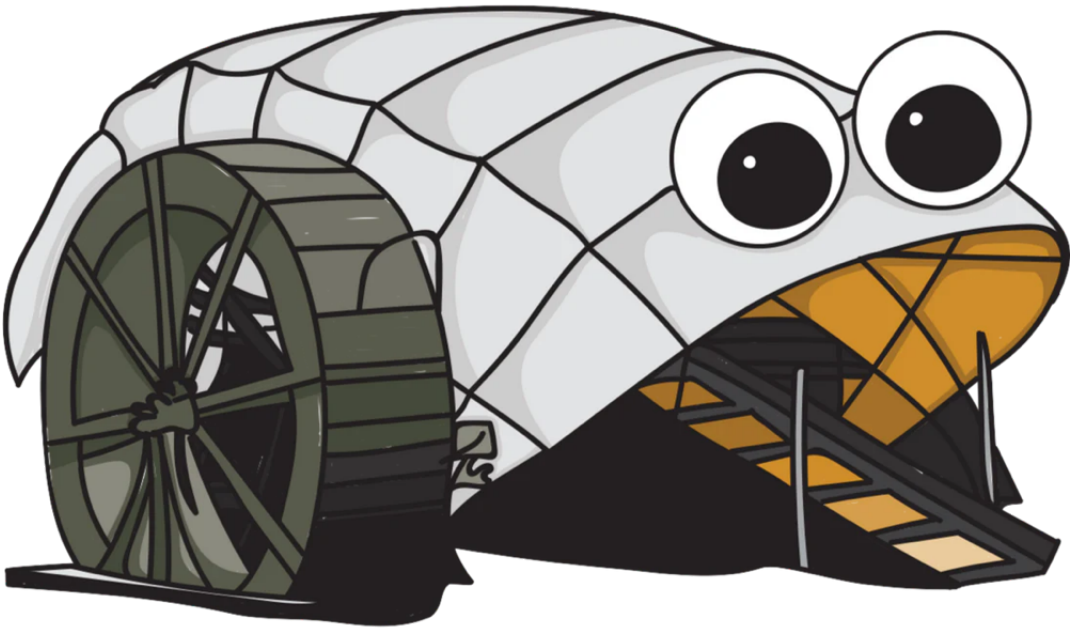

Class project: Mr. Trash Wheel Ate A Lot

Introduction

The city of Baltimore in MD installed on its rivers automatic trash collectors as part of the Baltimore Healthyy Harbor project, an initiative aiming to clean the ocean by intercepting trash right from the source. Mr. Trash Wheel, invented by environmental scientist and shipbuilder John Kellett, operates on various points on Baltimore water beds as an autonomous and sustainably powered dumpster that collect 500 tons of litter and debris each year before it can enter the Chesapeake Bay and Atlantic Ocean. The project undertaken below aims to analyse the amount of trash collected over 10 years by the dumpsters, observe particular trends, and evaluate the correlation between weather conditions, specifically precipitation levels, and the amount of trash collected yearly.

This project would be of interest to anyone who is curious about trends in river littering in Baltimore, offering them an insights on how littering habits changed in Baltimore as a result of the Baltimore Healthy Harbor initiative and the sensitization campaigns and advocacy around it. Another interesting question this project adresses is whether Baltimore inhabitants throw directly trash into its rivers or if the littering is caused by rain and stormwater runoff. Rainfall can wash trash from streets, parking lots, and other urban surfaces into storm drains that are connected to rivers and other aquatic bodies. We could thus recommend actions to reduce runoff-related pollution and improve trash

management and collection in the city.

For this project, data were retrieved from github (https://github.com/rfordatascience/tidytuesday/tree/master/data/2024/2024-03-05) and the National Centers for Environmental Information database (https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-month?bbox=39.379,-76.701,39.201,-76.523&pageNum=1&dataTypes=AWND&dataTypes=DYTS&d

Previous analyses done with this dataset include spatial studies to identify "litter hot spots" in Baltimore and it was determined that a trash disproportionally came from “hot spots” near carryout restaurants, convenience stores, bus stops, public middle and high schools, and food deserts (areas with limited access to grocery stores) (https://engineering.jhu.edu/magazine/2017/01/tackling-trash/). They were conducted by Chris Kelley and Ramya Ambikapathi, two John Hopkins

Data wrangling: Extraction And Manipulation Of Sub-dataframes For Visualization and Analysis

The Trash Wheel data provides information for trash collected monthly from 2014 to 2023. Below, the data are being grouped per year and then per month in order to offer visuals on the trend of waterbodies littering in Baltimore over the years and across. Relevant sections of the wheather data collected from the National Centers for Environmental Information are also extracted and merged with the trash wheel data for the weather-trash correlation analysis. Only a part of the data manipulation process is shown here, with the rest in the appendix.

In [41]:
```
# Import Packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statistics
import seaborn as sns
from imageio.v3 import imread
%matplotlib inline

# Import files and observe
trashwheel = pd.read_csv("trashwheel.csv")
blt_weather = pd.read_csv('USW00093784.csv')

# Dataframe with number of trash collected per type per year
trashwheel_per_year = trashwheel.groupby('Year').agg(PlasticBottles = ('Plas
                                                     Polystyrene = ('Polysty
                                                     CigaretteButts = ('Ciga
```

```
                                                    GlassBottles = ('GlassB
                                                    PlasticBags = ('Plastic
                                                    Wrappers = ('Wrappers',
                                                    SportsBalls = ('SportsB
                                                   )

# Dataframe with total number of trash collected per type per month
trashwheel_per_month_sum = trashwheel.groupby('Month').agg(PlasticBottles =
                                                    Polystyrene = ('Polysty
                                                    CigaretteButts = ('Ciga
                                                    GlassBottles = ('GlassB
                                                    PlasticBags = ('Plastic
                                                    Wrappers = ('Wrappers',
                                                    SportsBalls = ('SportsB
                                                     )

trashwheel_per_month_sum.loc['July'] = trashwheel_per_month_sum.loc['July']
trashwheel_per_month_sum.loc['September'] = trashwheel_per_month_sum.loc['Se
tw_per_month_sum = trashwheel_per_month_sum.drop('july').drop('september').r

month_order = ['January','February','March','April', 'May',
               'June', 'July', 'August', 'September', 'October', 'November',
tw_per_month_sum['Month'] = pd.Categorical(tw_per_month_sum['Month'], catego

tw_per_month_sum_ordered = tw_per_month_sum.sort_values(by = 'Month').set_ir
```

Visualize the data: Graphs showing the levels of trash collected per type over time

In [106…

```
#Horizontal stacked bar graph of trash collected per type per year
sns.set_theme(style = "dark");
trashwheel_per_year.plot(kind ='barh',
                         stacked = True,
                         color = ('lightsalmon', 'MediumAquamarine', 'whites
                                  'DarkKhaki', 'Plum', 'PeachPuff'),
                         figsize = (10, 6)
                        );

#Invert the y-axis to have an ascending timeline downward
plt.gca().invert_yaxis();

#Axes Labels and Titles
plt.xlabel('Count of Trash Pieces (in millions)');
plt.ylabel('Year');
plt.title('All Trash Types Collected by Mr. Trash Wheel Over 10 Years');
```

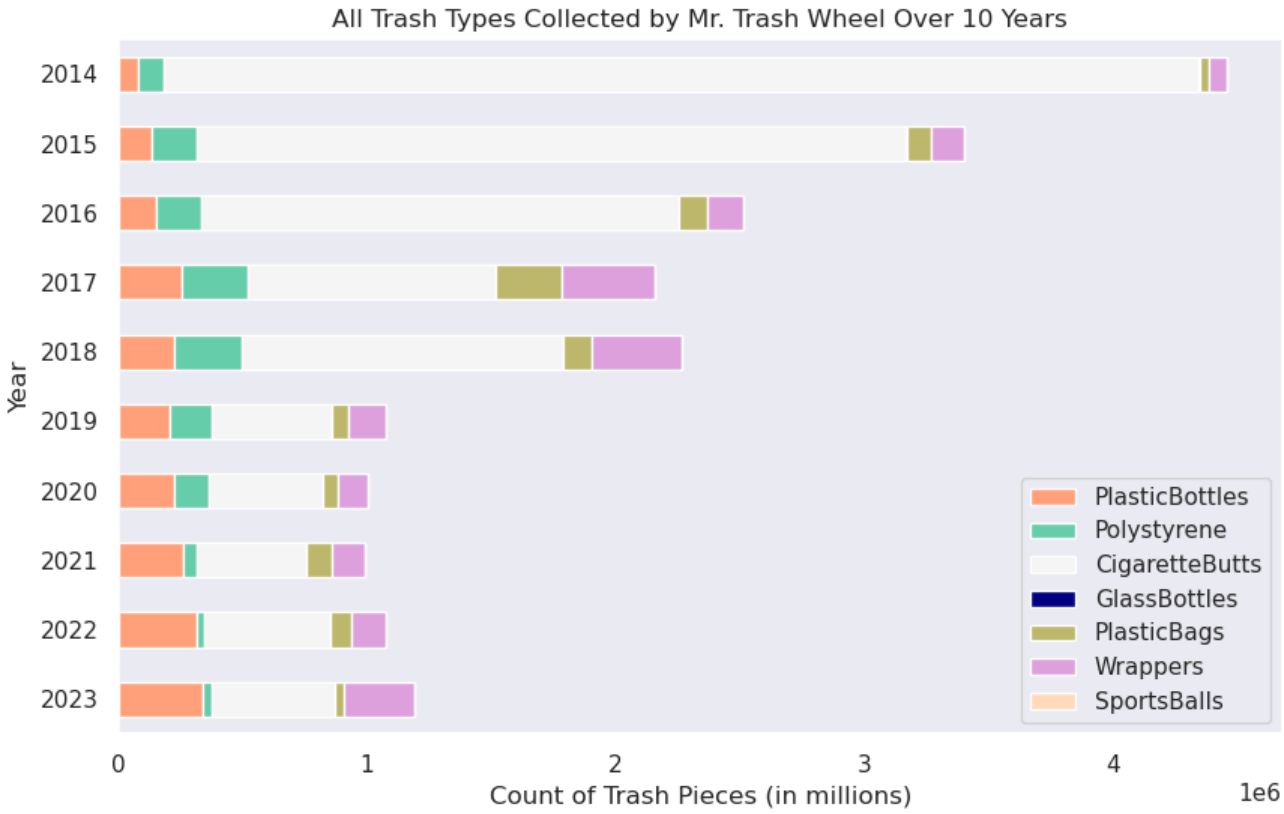

Figure1: Trash Types Collected by Mr. Trash Wheel Over 10 Years

This graph offers a stacked bar view of the total amount of trash collected by the dumpsters from 2014 to 2023. These types are the most commonly found in the dumpsters, with no surprise, plastic bags, wrappers, and plastic bottles being present. We can observe from this plot a significant decrease in the amount of littering over the years, speaking to the apparent success of the action undertaken by the city.

While the graph offers a comprehensive view of all trash types, there's a disproportion between the quantity of cigarette butts collected and those of the other types of trash.

In [105…

```
#Stacked bar graph of trash collected per type per year, excluding cigarette
sns.set_theme(style = "dark");
trashwheel_per_year.drop(columns = 'CigaretteButts').plot(kind ='barh',
                                                          stacked = True,
                                                          color = ('LightSal
                                                                   'DarkKhak
                                                          figsize = (10, 6)
                                                         );
#Invert the y-axis to have an ascending timeline downward
plt.gca().invert_yaxis();

#Axes Labels and Titles
plt.xlabel('Year');
```

```
plt.ylabel('Count of Trash (in millions)');
plt.title('Trash Types Collected by Mr. Trash Wheel Over 10 Years \n (Excluc
```

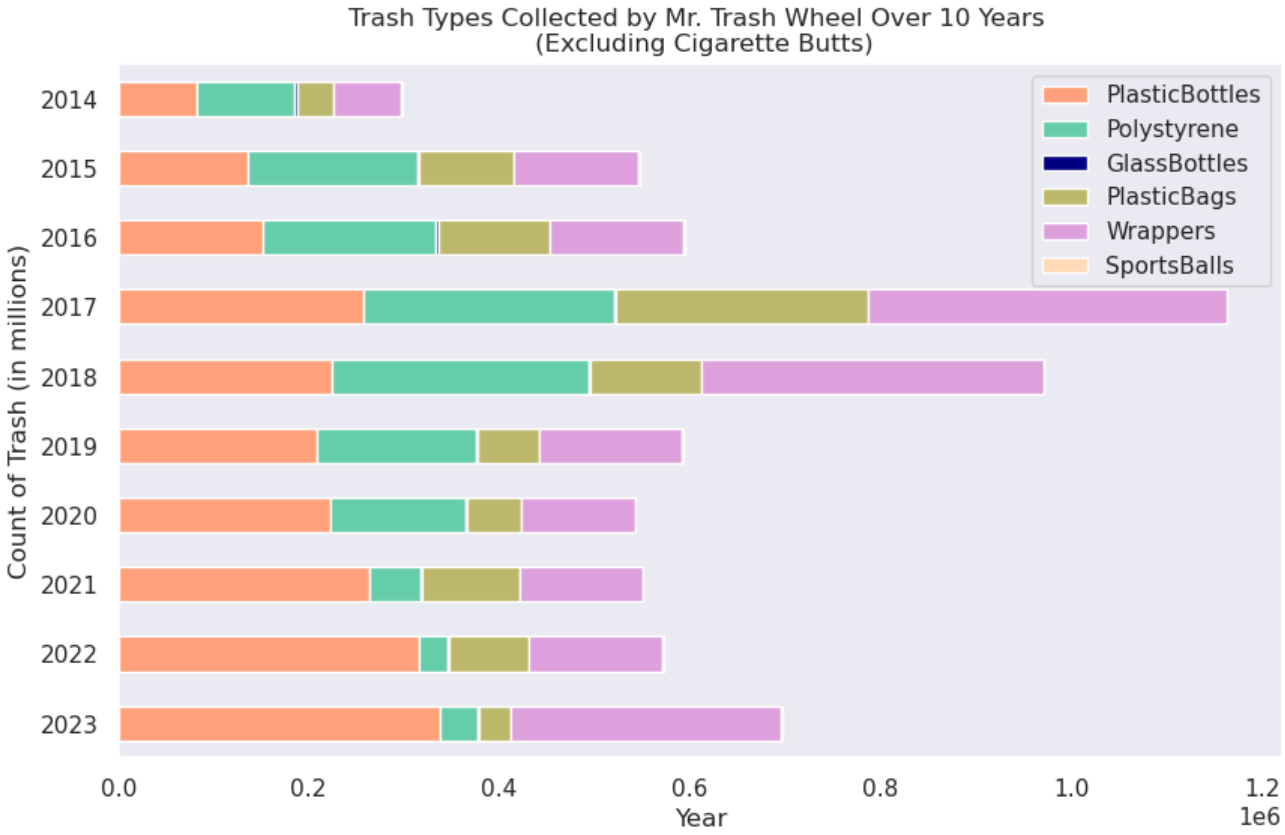

Figure2: Trash Types Collected by Mr. Trash Wheel Over 10 Years (Excluding Cigarette Butts)

After excluding the cigarette butts data, the graph tells a completely different story with more fluctuations in the amount of trash collected by the dumpsters over time. It is therefore important to realize that the first graph can only tell us about a change in smoking habits in the city over the 10 years, that can be explained by a decrease in the number of smokers or the fact that smokers found a way to better dispose of their butts.

With this new visual, we notice a peak in trash collection in 2017 which is yet to be explained. As time goes, the amount of trash collected decreased but is still higher than at the start of the project.

In [100…

```
# Plot for the mean of trash collected per month from 2014 to 2023
sns.set_theme(style = "dark");
tw_per_month_mean_ordered.plot(kind ='barh',
                               stacked = True,
                               color = ('lightsalmon', 'MediumAquamarine', '
                                  'DarkKhaki', 'Plum', 'PeachPuff'),
                               figsize = (10, 6)
```

```
                                    );

#Invert the y-axis to have an ascending timeline downward
plt.gca().invert_yaxis();

#Axes Labels and Titles
plt.xlabel('Count of Trash (in millions)');
plt.ylabel('Month');
plt.title('Monthly Mean Trash Collected by Mr. Trash Wheel Per Type Per Mont
```

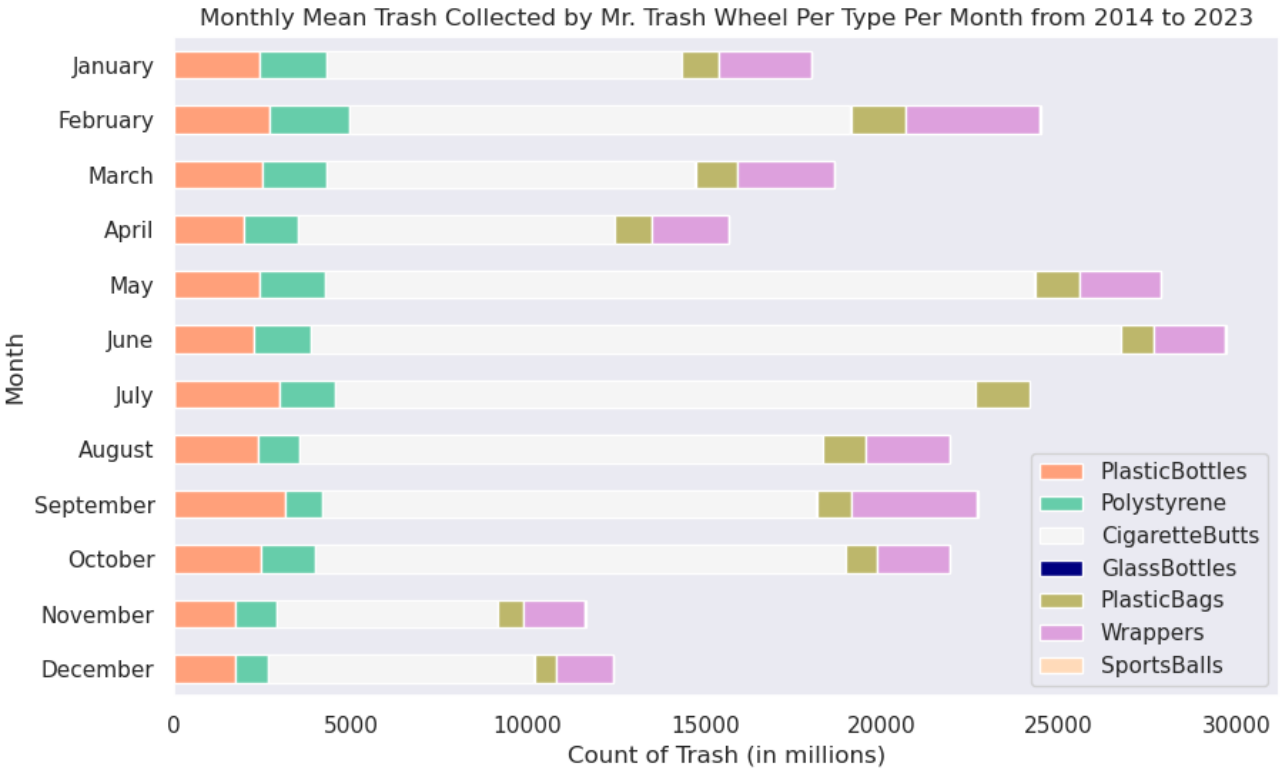

Figure3: Monthly Mean Trash Collected by Mr. Trash Wheel Per Type Per Month from 2014 to 2023

The figure plots the mean number of trash of all types collected by the dumpsters by month over the 10 years of interest. We observe a peak in the month of June. Again the presence of cigarette butts data might be offering a view skewed from reality, as observed below.

In [101…

```
# Plot for the mean of trash collected per month from 2014 to 2023 excluding
sns.set_theme(style = "dark");
tw_per_month_mean_ordered.drop(columns = 'CigaretteButts').plot(kind ='barh'
                                                                stacked = Tr
                                                                color = ('Li
                                                                   'DarkKhak
                                                                figsize = (1
                                                               );
```

```
#Invert the y-axis to have an ascending timeline downward
plt.gca().invert_yaxis();

#Axes Labels and Titles
plt.xlabel('Count of Trash (in millions)');
plt.ylabel('Month');
plt.title('Mean Trash Collected by Mr. Trash Wheel Per Type Per Month \n (Ex
```

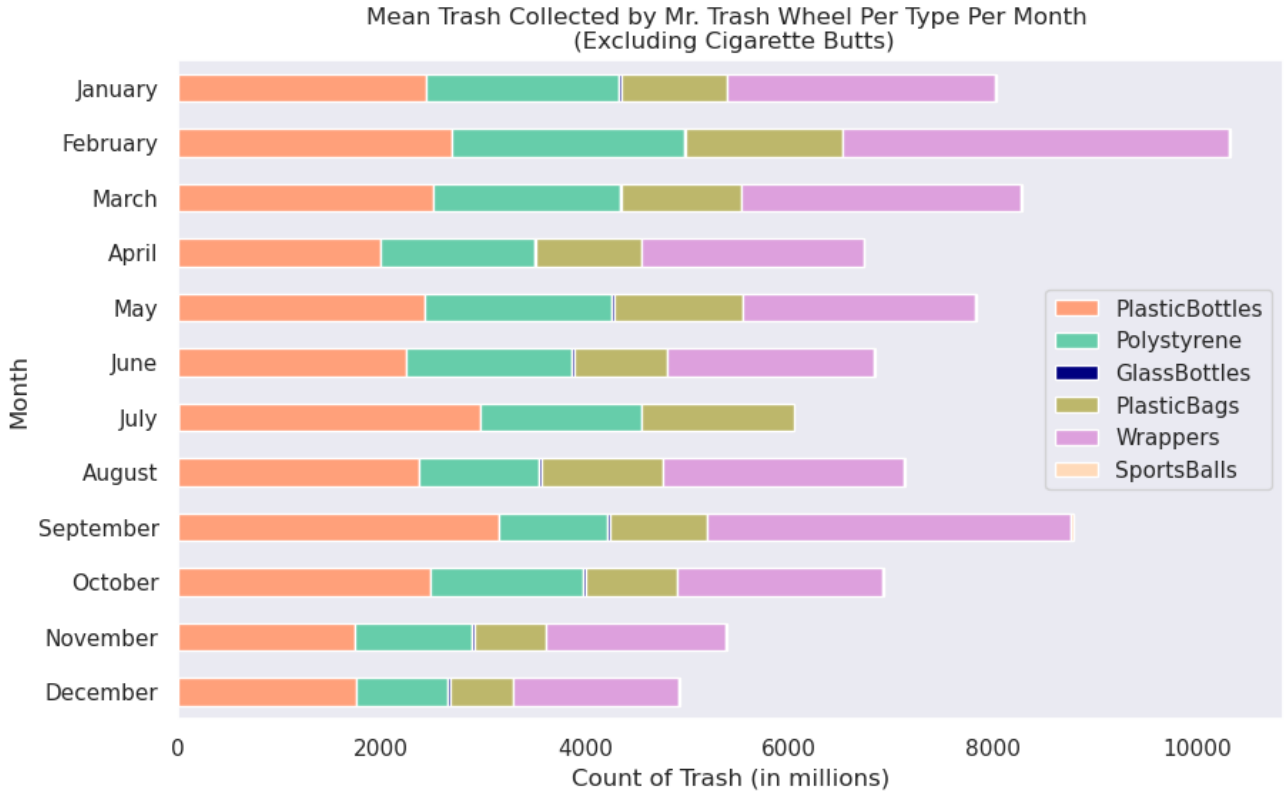

Figure4: Mean Trash Collected by Mr. Trash Wheel Per Type Per Month (Excluding Cigarette Butts)

After excluding the cigarette trash type, we obtain a graph showing a peak in the month of February and September, and lower numbers of trash collected around July and December. These times correspond to school breaks, summer and winter respectively where students are out of school. The previous study conducted by Kelley and Ambikapathi revealed hot littering spots around schools which reopen yearly around September as students start their new academic year and January/February when they resume classes from the winter break.

In [102…

```
#Plot overlapping graphs of precipitation levels in Baltimore and total tras

#Set figure size
fig, ax1 = plt.subplots(figsize = (10, 6));
```

```
#Set 1-year intervals between years, x-axis label, and graph title
ax1.set_xticks(np.arange(2014, 2024, 1));
ax1.set_xlabel('Year')
plt.title('Precipitation Levels in Baltimore vs Total Trash (excluding Cigar

#Set same axis for both graphs
ax2 = ax1.twinx();

#plot mean precipiation levels
ax1.plot(blt_rain_per_year, color = 'darkolivegreen');
ax1.set_ylabel('Mean Precipitation Level (in mm)');
ax1.set_xticks(np.arange(2014, 2024, 1));

#plot total trash collected yearly
ax2.bar(trashwheel_per_year.drop(columns = 'CigaretteButts').index,
        trashwheel_per_year.drop(columns = 'CigaretteButts').sum(axis = 1),
        color = 'darksalmon', alpha = .4, edgecolor = 'coral');
ax2.set_ylabel('Total Number of Trash collected (in millions)');
```

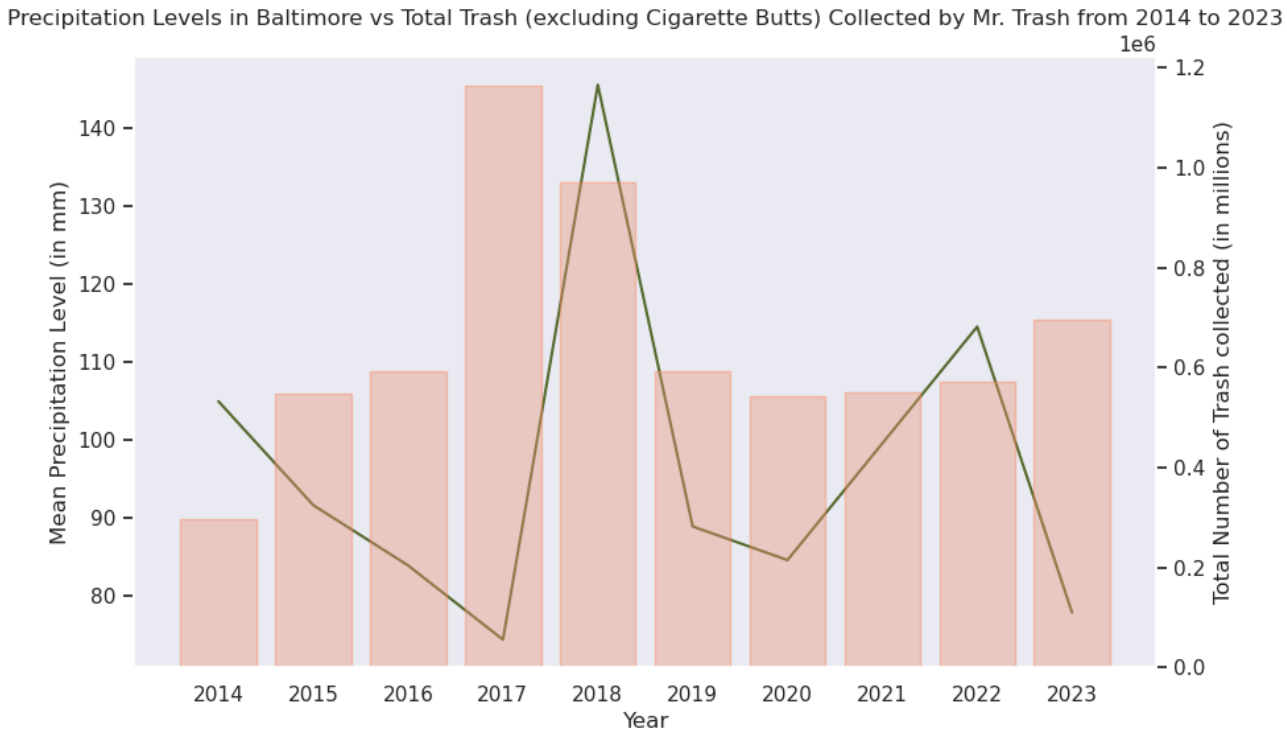

Figure5: Precipitation Levels in Baltimore vs Total Trash (excluding Cigarette Butts) Collected by Mr. Trash from 2014 to 2023

The graph offers an overlay view of the yearly precipitation levels in Baltimore and the amount of trash collected by the dumpsters each year. While there is a peak around 2018 for the precipitation levels, the peak for the collected amount of trash occurs a year before, in 2017, which also sees the lowest level of precipitation over the 10 years. This indicates that higher amounts of trash collected by the dumpsters on the river

cannot be put on the account of rain and stormwater runoff.

Analyses: Correlation between Precipitation levels and Trash collected

While we were able to observe a low correlation from the graph, a computed correlation value is needed to ascertain our findings. The correlation between these data points is directly computed using the statistics package.

In [98]:
```
# Correlation between total trash and precipitation level

#Dataframe with total trash collected and mean precipitation levels from 201
rain_vs_trash = blt_rain_per_year.reset_index().merge(trashwheel_per_year.dr

rain_vs_trash2 = rain_vs_trash.rename(columns = {0: 'Trash'}).drop('DATE', a
rain_vs_trash2

#Correlation Calculation
corr = statistics.correlation(rain_vs_trash2.index, rain_vs_trash2.Trash)

print('The correlation between the precipitatin levels in Baltimore and the
      + str(round(corr, 3)))
```

```
The correlation between the precipitatin levels in Baltimore and the total a
mount of trash collected by Mr. Trash in Baltimore is 0.046
```

Conclusions

This analysis of Baltimore's Mr. Trash Wheel dataset, complemented by weather data, has provided valuable insights into the effectiveness of the Baltimore Healthy Harbor Initiative in addressing urban littering. The data reveals that, despite the installation of automatic trash interceptors and efforts to raise public awareness, littering habits in Baltimore have not shown significant improvement over the years. The consistent amount of trash collected annually suggests that direct human behavior, such as improper disposal of waste, continues to be a predominant factor contributing to urban pollution.

Moreover, the analysis found no significant correlation between precipitation levels and the amount of trash collected. This finding challenges the assumption that rainwater runoff is a primary driver of trash entering Baltimore's waterways. While heavy rainfall events can transport litter from streets and urban surfaces into storm drains, these occurrences alone do not appear to account for the majority of the debris intercepted by

the trash wheels. However, further analysis could be performed using storm and wind speed data in Baltimore as they can provide better insights in understanding how other weather events might correlate with spikes in trash collection.

These findings highlight the need for continued and perhaps reimagined public engagement campaigns to address the root causes of littering. The city may benefit from targeted interventions at "litter hot spots" identified in previous studies, especially in schools as we were able to observe peaks of trash collection during back-to-school periods. Additionally, enhancing trash collection infrastructure and enforcement of anti-littering laws could be key strategies in reducing the waste entering Baltimore's waterways.

Future research could explore other potential factors influencing urban littering patterns, such as socioeconomic variables, population density, and the effectiveness of educational initiatives. Ultimately, this project underscores the importance of combining technological innovation with community-driven solutions to address the complex challenge of urban pollution.

Reflection

This project was both rewarding and challenging. One of the biggest obstacles was obtaining and working with the weather data. While the precipitation data was accessible, finding detailed storm and wind data, which could have provided additional insights into other weather factors potentially influencing trash collection, proved difficult. The absence of this data limited the scope of my analysis.

I spent approximately 24 hours on this project, including data research, cleaning, analysis, and visualization. I particularly enjoyed creating the graphs and carefully selecting which visuals would best address the project's main questions. This process helped clarify trends in the data and allowed me to communicate my findings effectively.

One analysis I attempted but ultimately excluded involved visualizing the correlation between rainfall and each specific type of trash. Additionally, I explored a creative idea of designing a whale-shaped image with pixel darkening corresponding to trash weight over the years, relative to the weight of an actual whale. While I didn't execute this concept, it was an exciting exercise in thinking about innovative ways to present data.

Overall, this project taught me valuable lessons about navigating data challenges, being resourceful in visualization, and balancing creativity with practicality in analysis.

Appendix 1 - Data Wrangling (Continued)

In [ ]:
```
# Dataframe with mean number of trash collected per type per month
trashwheel_per_month_mean = trashwheel.groupby('Month').agg(PlasticBottles =
                                                            Polystyrene = ('Polysty
                                                            CigaretteButts = ('Ciga
                                                            GlassBottles = ('GlassB
                                                            PlasticBags = ('Plastic
                                                            Wrappers = ('Wrappers',
                                                            SportsBalls = ('SportsB
                                                             )

trashwheel_per_month_mean.loc['July'] = trashwheel_per_month_mean.loc['July'
trashwheel_per_month_mean.loc['September'] = trashwheel_per_month_mean.loc['
tw_per_month_mean = trashwheel_per_month_mean.drop('july').drop('september')

month_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July
tw_per_month_mean['Month'] = pd.Categorical(tw_per_month_mean['Month'], cate

tw_per_month_mean_ordered = tw_per_month_mean.sort_values(by = 'Month').set_

# Dataframe with median number of trash collected per type per month
trashwheel_per_month_median = trashwheel.groupby('Month').agg(PlasticBottles
                                                            Polystyrene = ('Polysty
                                                            CigaretteButts = ('Ciga
                                                            GlassBottles = ('GlassB
                                                            PlasticBags = ('Plastic
                                                            Wrappers = ('Wrappers',
                                                            SportsBalls = ('SportsB
                                                             )

trashwheel_per_month_median.loc['July'] = trashwheel_per_month_median.loc['J
trashwheel_per_month_median.loc['September'] = trashwheel_per_month_median.l
tw_per_month_median = trashwheel_per_month_median.drop('july').drop('septemb

month_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July
tw_per_month_median['Month'] = pd.Categorical(tw_per_month_median['Month'],

tw_per_month_median_ordered = tw_per_month_median.sort_values(by = 'Month').

#Extract weather data from 2014 to 2023
blt_rain = blt_weather.set_index('DATE').loc['2014-05':'2023-12'][['PRCP']].

#Convert DATE column's type to datetime
blt_rain['DATE'] = pd.to_datetime(blt_rain['DATE'])

#Group data by year
blt_rain = blt_rain.groupby([blt_rain['DATE'].dt.year]).mean()
blt_rain_per_year = blt_rain[['PRCP']]
```

Appendix 2 - Reviews

Reviewer template

This is a template for reviewing class projects. Please fill out the sections below to review a project, and then email your review to person whose project it is. Also upload a pdf of each of your reviews to Gradescope.

Remeber, these reviews are not annoymous so be sure to be positive in your review, but also be critical in order to help make the author's project stronger for their final submission.

Project writer name: Yasmina Traoré

Project writer email address: yasmina.traore@yale.edu

Reviewer name: Jiahe Jin

Reviwer email address: jiahe.jin@yale.edu

Summary

The project aims to determine the amount of trash collected by the dumpster over 10 years and its trends, using the correlation method to determine the main cause of why there is so much trash in the river. Finally, based on the analysis the project states the importance of future environmental protection.

1. Overall strengths and weaknesses

The project analyzing Baltimore's "Mr. Trash Wheel" data offers several strengths and areas for potential improvement. One of its advantages is that this project offers a detailed visualization of the trend of urban littering based on the data. The analysis convincingly indicates that the situation of littering has declined over time, highlighting the influence of the Baltimore Healthy Harbor Initiative. Furthermore, the investigation into the correlation between precipitation levels and garbage collection is intriguing, particularly considering the negligible correlation discovered, which is counterintuitive. However, the project could be strengthened by exploring additional variables that

influence littering patterns. Additionally, while the visualizations are compelling, some lack clarity or could benefit from alternative presentations to capture nuanced trends. While the visualizations are compelling, some of the display types are too monolithic, and other forms of presentation could be used to capture subtle trends. Overall, this project demonstrates strong analytical skills and offers a solid foundation for future research into urban waste management.

2. Major revisions

N/A. But I think we can introduce more difficult models, for example, machine learning.

3. Minor revisions

1. Offer more Diverse Visualizations
2. The style of the title can be adjusted, the current one is a little crowded (the title has a little too many words).

3. Rubric score

Please write a score for the project based on the project rubric that is on Canvas. For any items where there would be a point deduction, please cut and paste a bullet point for that item in the "Items for ponts take off" section below.

Rubric items where points would be taken off if not addressed:

Visual appearance of graphs could be improved(title style)-2

Total score: 88/90

In [ ]:

Reviewer template

This is a template for reviewing class projects. Please fill out the sections below to review a project, and then email your review to person whose project it is. Also upload a pdf of each of your reviews to Gradescope.

Remeber, these reviews are not annoymous so be sure to be positive in your review, but

also be critical in order to help make the author's project stronger for their final submission.

Project writer name: Yasmina Traore

Project writer email address: yasmina.traore@yale.edu

Reviewer name: Ethan Wolf

Reviwer email address: ethan.wolf@yale.edu

Summary

This project analyses the amount of trash collected by Mr. Trash Wheel and evaluate the correlation between precipitation levels and amount of trash collected yearly.They find that the most common trash collected is ciggarette butts, plastic bags and plastic water bottles. They also visualize the precipitation levels in Baltimore and the total trash collected collected by Mr. Trash Wheel and found there was not indications that more trash was collected when precipitation was higher finding the correlation is .046.

1. Overall strengths and weaknesses

Overall, I think you have a very interesting and unique project. Mr. Trash Wheel is a very interesting choice to do your project on and I'm interested in knowing where you found out about this data. I think your stacked bar graphs are very strong visualizations of your data and you do a great job examining your research question. Overall, they aren't any major revisions that I think you need to do but maybe make your figure 3 stack vertically instead of horizontally so that the visualization can be larger. I also think that you could do a little bit better of a job formatting your data wrangling step as a lot of your code is cut off by the PDF and some of your explanations are a little vague to understand with only a part of the code

2. Major revisions

None!

3. Minor revisions

- make the figures in figure 3 stack vertically instead of horizontally
- a lot of your data wrangling is cut off by the PDF formatting so its hard to tell exactly what you are doing in that stage, so try to fit it in the frame a bit better

3. Rubric score

Please write a score for the project based on the project rubric that is on Canvas. For any items where there would be a point deduction, please cut and paste a bullet point for that item in the "Items for ponts take off" section below.

Rubric items where points would be taken off if not addressed:

- Did not provide a clear explanation of what the variables are that are used in the analysis

Total score: 88/90

In [ ]:

Reviewer template

This is a template for reviewing class projects. Please fill out the sections below to review a project, and then email your review to person whose project it is. Also upload a pdf of each of your reviews to Gradescope.

Remeber, these reviews are not annoymous so be sure to be positive in your review, but also be critical in order to help make the author's project stronger for their final submission.

Project writer name: Yasmina Traore

Project writer email address: yasmina.traore@yale.edu

Reviewer name: Mateo Pechon-Elkins

Reviwer email address: mateo.pechon-elkins@yale.edu

Summary

This project provides insight into the presence of trash in Baltimore water beds, finding trends like the dominance of cigarette butts and the change in trash over months. Through this analysis, the project synthesizes descriptions of the trash ecosystem in Baltimore and generalizes the insights about dealing with trash to potential f uture action in other places. It also uses linear correlation calculations to find the correlation between precipitation levels and trash collected. These correlations are used to argue that there is not a significant connection between the two and deepen the analysis of why fluctuations in trash levels exist.

1. Overall strengths and weaknesses

One of the great strengths of this project is the clarity of code when assisted with comments. It was easy to follow along with what was going on due to the way code was spaced out neatly, variables were labelled clearly, and code comments explained everything. Many of the descriptions and analyses of graphs were concise yet insightf ul, especially regarding figures 1 and 2. The amount of space dedicated to explicating specific takeaways from them. Finally, the conclusion is very strong, and connects the insights from the graphs and the correlation to its final analysis. This was done very well and greatly improves the project. However, there are a few small areas for improvement. Discussing limitations in the data could be helpf ul in the correlation exploration and the project as a whole, and it would give a better idea of how the data was approached and why things were done the way they were. Figure four seemed the only outlier in the way every graph and test was skillf ully tied back into the conclusion and overall project. Either more explanation, a removal, or a replacement would help. Overall though, this project is high quality, insightf ul, and enjoyable to read.

2. Major revisions

- Indent/spread out some of the code so that it all appears in the pdf. Sometimes in the pdf it trails off to the right, and this makes it hard to f ully understand what is going on.
- Figure 4 does not seem to have a very clear connection to the conclusion and overall project, as its description mostly restates previously extracted insights. Consider writing more, replacing, or removing this graph.
- Moving some data wrangling code to the appendix may help keep the analysis under or closer to the 10 page limit.

3. Minor revisions

- Delete some of the default comment code, for example "Show the analyses you did here", Appendix if left unused.
- Consider changing the color scheme of the stacked graph for trash types over 10 years (including cigarette butts) to be a little more pronounced/differentiated. It would be helpf ul for readibility. The main difficulty from a first glance read is differentiating plastic bags and wrappers. You do not have to remove the cigarette

color scheme, but making some of the oranges and reds just a little more different would help. - Acknowledging the limitations of the correlation test (very few data points, etc)

3. Rubric score

Please write a score for the project based on the project rubric that is on Canvas. For any items where there would be a point deduction, please cut and paste a bullet point for that item in the "Items for ponts take off" section below.

Rubric items where points would be taken off if not addressed:

- Visual appearance of graphs could be improved (-1) -- The graphs are incredibly well made and for the most part visually appealing, but figure one could be a bit easier to read.
- Did not provide a written describe of the insights that the graphs provide (-2) -- For figure four, this could be improved as previously discussed.
- Analyses have limitations that should have been either been taken into account or mentioned (-1) -- It is a small point, but acknowledging the limitations in the data and the correlation test could improve the project

Total score: 86/90

In [ ]: