Data Cleaning

1- Column selection:

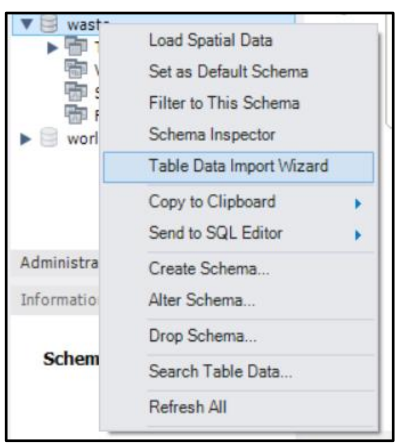
This was done by creating an empty table with only the desired columns and importing the data into it using the following code in Figure[1]

```sql
Create Database If Not Exists Waste;
Use Waste;
Create Table If Not Exists WasteData (
    Country Varchar(100),
    Region Varchar(100),
    Indicator Varchar(255),
    Indicator_Unit Varchar(100),
    Year int,
    Value Float,
    Calculation Varchar(100),
    Disaggregation_Level Varchar(100),
    Information Text
);
```

Figure[1]: Database and Table creation.

The importing process was done using the import wizard feature available in MySQL Figure[2] to fill the above table.

Figure[2]: Importing data using the Import Wizard feature.



2- Translation and Handling missing data:

After inspecting the data it was noticed that only 3 columns form the selected contained spanish language (Disaggregation_Level, Calculation, Information). So in order to translate the data, we first needed to see what kind of values we had. Which was done by using the Unique constraint then changing these values with their translated counterpart using Update constraint Figure[3].

```sql
-- to observe the unique value in "Disaggregation_Level"
Select Distinct Disaggregation_Level FROM WasteData;
-- transilation of "Disaggregation_Level"
UPDATE wastedata
SET Disaggregation_Level = 'Country'
WHERE Disaggregation_Level = 'PAÍS';
UPDATE wastedata
SET Disaggregation_Level = 'Unknown'
WHERE Disaggregation_Level = '';
UPDATE wastedata
SET Disaggregation_Level = 'Country, Municipality'
WHERE Disaggregation_Level = 'PAÍS, MUNICIPIO';
UPDATE wastedata
SET Disaggregation_Level = 'Country, Department'
WHERE Disaggregation_Level = 'PAÍS, DEPARTAMENTO';
UPDATE wastedata
SET Disaggregation_Level = 'Country, Macroregion'
WHERE Disaggregation_Level = 'PAÍS, MACROREGIÓN';
UPDATE wastedata
SET Disaggregation_Level = 'Country, Region, Municipality'
WHERE Disaggregation_Level = 'PAÍS, REGIÓN, MUNICIPIO';
```

```sql
UPDATE wastedata
SET Disaggregation_Level = 'Country, Region, Area'
WHERE Disaggregation_Level = 'PAÍS, REGIÓN, ÁREA';
UPDATE wastedata
SET Disaggregation_Level = 'Country, Region, Province'
WHERE Disaggregation_Level = 'PAÍS, REGIÓN, PROVINCIA';
UPDATE wastedata
SET Disaggregation_Level = 'Partial'
WHERE Disaggregation_Level = 'PARCIAL';
UPDATE wastedata
SET Disaggregation_Level = 'District'
WHERE Disaggregation_Level = 'DISTRITO';
UPDATE wastedata
SET Disaggregation_Level = 'Country, District'
WHERE Disaggregation_Level = 'PAÍS, DISTRITO';
UPDATE wastedata
SET Disaggregation_Level = 'Country'
WHERE Disaggregation_Level = 'PAÍS ';
UPDATE wastedata
SET Disaggregation_Level = 'Subregion'
WHERE Disaggregation_Level = 'SUBREGIÓN';
UPDATE wastedata
SET Disaggregation_Level = 'Region'
WHERE Disaggregation_Level = 'REGIÓN';
```

Figure[3]: Translating data using Unique and Update constraints in Disaggregation_Level.

The same was done for calculation and information as in Figure[4].

```sql
-- to observe the unique value in "Calculation"
SELECT DISTINCT Calculation FROM wastedata;
-- transilation of "Calculation"
UPDATE wastedata
SET Calculation = 'DIRECT'
WHERE Calculation = 'DIRECTO';
UPDATE wastedata
SET Calculation = 'INDIRECT'
WHERE Calculation = 'INDIRECTO';
```

```sql
-- to observe the unique value in "Information"
SELECT DISTINCT Information FROM wastedata;
-- transilation of "Information"
UPDATE wastedata
SET Information = 'Unknown'
WHERE Information = '';
UPDATE wastedata
SET Information = 'Official'
WHERE Information = 'OFICIAL';
UPDATE wastedata
SET Information = 'Imputed'
WHERE Information = 'IMPUTADA';
```

Figure[4]: Translating data using Unique and Update constraints in calculation and information.

3- Checking for duplicates and null values:

To check for duplicates a query that checks for how many times each unique combination of values occurs were written and executed as in Figure[5].

```
-- to check for duplicates
SELECT *, COUNT(*) AS duplicate_count
FROM wastedata
GROUP BY Country, Region, Indicator, Indicator_Unit, Year, Value, Calculation, Disaggregation_Level, Information
HAVING COUNT(*) > 1;
```

Figure[5]: Checking for duplicates.

Following this, the remaining null values were counted using the following query Figure[6].

```
-- count null values
SELECT
    SUM(CASE WHEN Country IS NULL THEN 1 ELSE 0 END) AS Country_null_count,
    SUM(CASE WHEN Region IS NULL THEN 1 ELSE 0 END) AS Region_null_count,
    SUM(CASE WHEN Indicator IS NULL THEN 1 ELSE 0 END) AS Indicator_null_count,
    SUM(CASE WHEN Indicator_Unit IS NULL THEN 1 ELSE 0 END) AS Indicator_Unit_null_count,
    SUM(CASE WHEN Year IS NULL THEN 1 ELSE 0 END) AS Year_null_count,
    SUM(CASE WHEN Value IS NULL THEN 1 ELSE 0 END) AS Value_null_count,
    SUM(CASE WHEN Calculation IS NULL THEN 1 ELSE 0 END) AS Calculation_null_count,
    SUM(CASE WHEN Disaggregation_Level IS NULL THEN 1 ELSE 0 END) AS Disaggregation_Level_null_count,
    SUM(CASE WHEN Information IS NULL THEN 1 ELSE 0 END) AS Information_null_count
FROM wastedata;
```

Figure[6]: Counting the remaining null values.

4- Exporting a new file for further use:

The modified table was exported to be used in other softwares using the following code Figure[7].

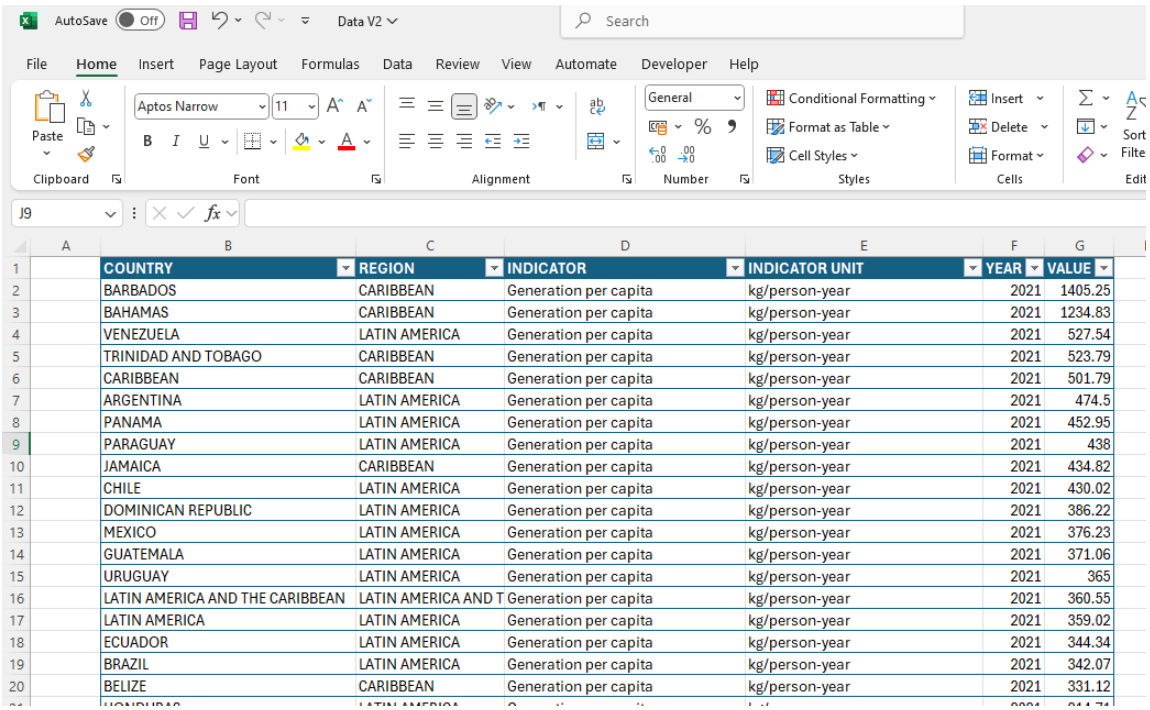
```
-- to export the new file
SELECT *
INTO OUTFILE 'C:/Users/abdulrahman/OneDrive/Desktop/Final project/Translated.csv'
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
FROM wastedata;
```

Figure[7]: Exporting the cleaned data.

Data Structure Using Microsoft Excel

In Microsoft Excel, large datasets are processed by removing unnecessary columns and formatting the relevant data as a table to identify indicators and their respective units.

This approach of ignoring irrelevant columns and formatting the remaining data allowed for a more streamlined view, especially when comparing key metrics across datasets.



COUNTRY	REGION	INDICATOR	INDICATOR UNIT	YEAR	VALUE
BARBADOS	CARIBBEAN	Generation per capita	kg/person-year	2021	1405.25
BAHAMAS	CARIBBEAN	Generation per capita	kg/person-year	2021	1234.83
VENEZUELA	LATIN AMERICA	Generation per capita	kg/person-year	2021	527.54
TRINIDAD AND TOBAGO	CARIBBEAN	Generation per capita	kg/person-year	2021	523.79
CARIBBEAN	CARIBBEAN	Generation per capita	kg/person-year	2021	501.79
ARGENTINA	LATIN AMERICA	Generation per capita	kg/person-year	2021	474.5
PANAMA	LATIN AMERICA	Generation per capita	kg/person-year	2021	452.95
PARAGUAY	LATIN AMERICA	Generation per capita	kg/person-year	2021	438
JAMAICA	CARIBBEAN	Generation per capita	kg/person-year	2021	434.82
CHILE	LATIN AMERICA	Generation per capita	kg/person-year	2021	430.02
DOMINICAN REPUBLIC	LATIN AMERICA	Generation per capita	kg/person-year	2021	386.22
MEXICO	LATIN AMERICA	Generation per capita	kg/person-year	2021	376.23
GUATEMALA	LATIN AMERICA	Generation per capita	kg/person-year	2021	371.06
URUGUAY	LATIN AMERICA	Generation per capita	kg/person-year	2021	365
LATIN AMERICA AND THE CARIBBEAN	LATIN AMERICA AND T	Generation per capita	kg/person-year	2021	360.55
LATIN AMERICA	LATIN AMERICA	Generation per capita	kg/person-year	2021	359.02
ECUADOR	LATIN AMERICA	Generation per capita	kg/person-year	2021	344.34
BRAZIL	LATIN AMERICA	Generation per capita	kg/person-year	2021	342.07
BELIZE	CARIBBEAN	Generation per capita	kg/person-year	2021	331.12

Figure[8] Given data sets formatted as a table

The "Format as Table" feature in Excel enhances data organization and analysis by automatically applying structured formatting, making data easier to read and work with. It provides built-in sorting, filtering, and dynamic ranges that adjust as data is added or removed.

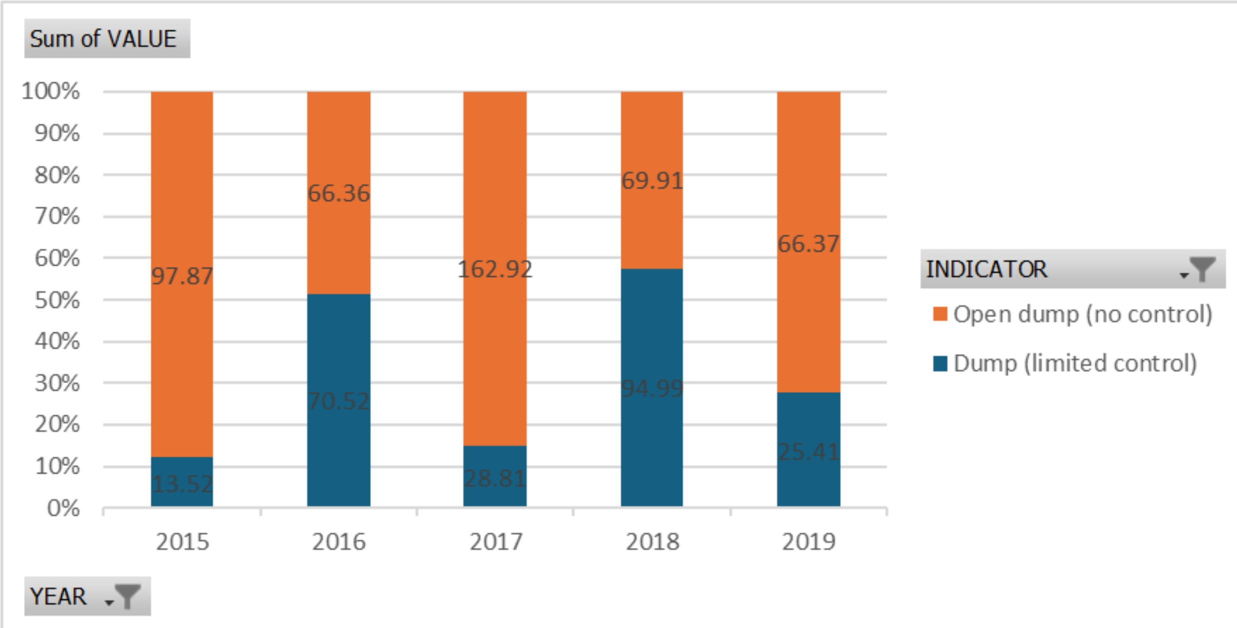
Tables support structured references for clearer formulas and are compatible with PivotTables and other functions, ensuring more efficient data handling. Additional features like automatic totals, consistent formatting, and slicer integration improve collaboration and overall usability.

The filtering option in the data set under the study helped us discover the number of indicators, which are 30 indicators, and also to know the corresponding measure unit for each indicator.

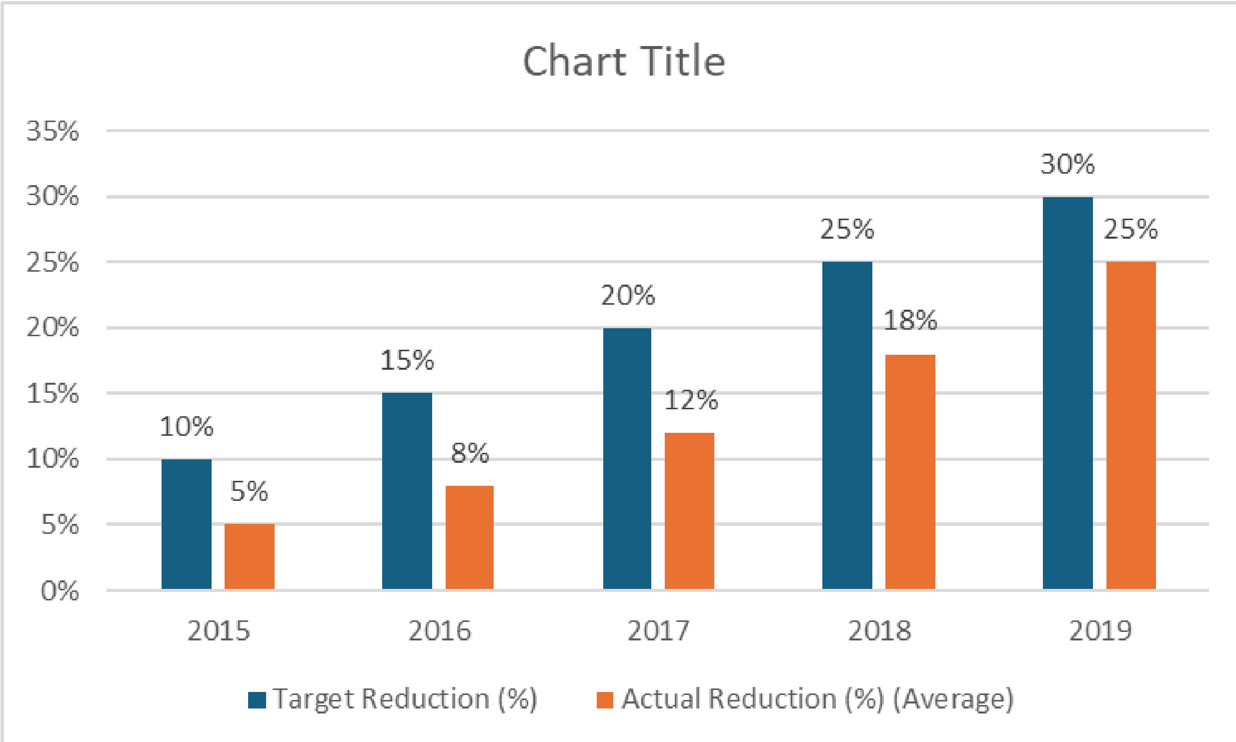
Dump (limited control)

In dictators for the waste management terms listed:

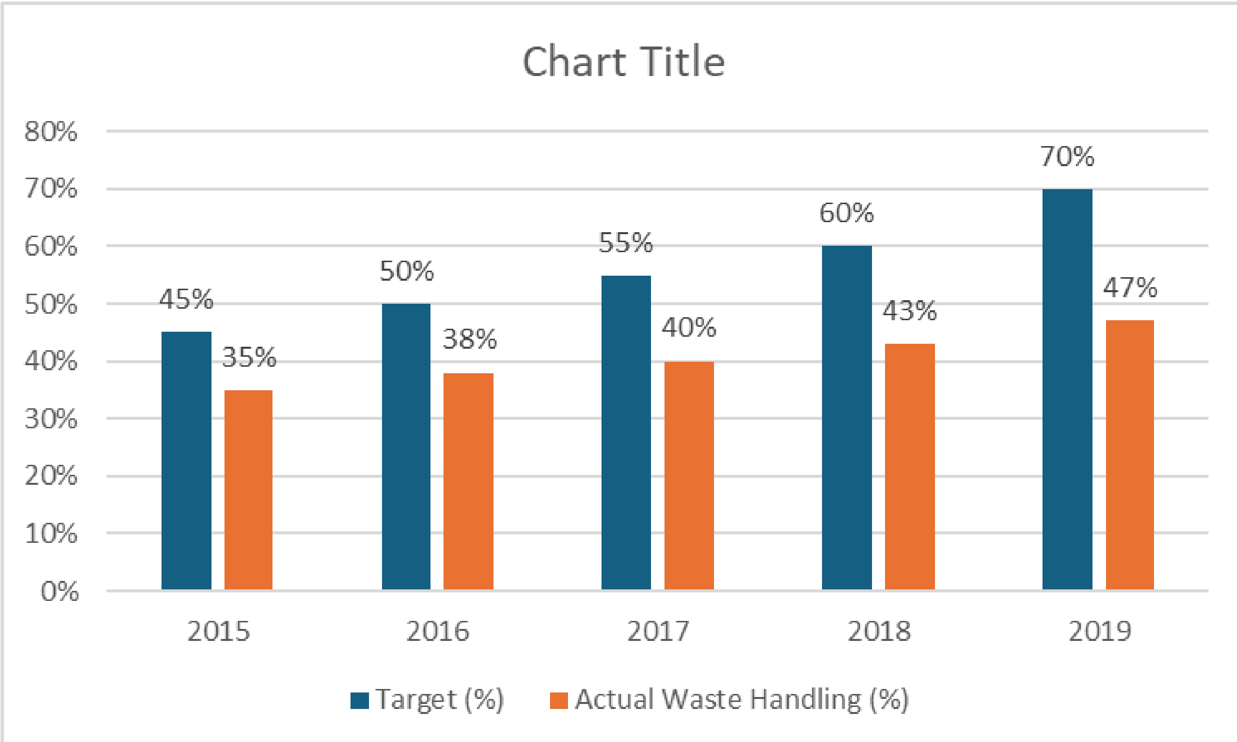
1. **Accumulation in landfill**: This refers to the buildup of municipal solid waste (MSV indicating the volume or mass of waste that remains in the landfill over a certain p
2. **Collection coverage - households**: The percentage of households that are serv services, showing the extent of waste management reach within a population.
3. **Collection coverage - tonnes**: The amount of waste (in tonnes) that is collected management service. It indicates the efficiency of the waste collection system in t waste handled.
4. **Composting**: The process of converting organic waste (such as food scraps, yarc nutrient-rich soil amendment through natural decomposition.
5. **Co-processing**: The simultaneous treatment of waste materials in industrial proce production, to recover energy and reduce the use of fossil fuels and raw materials
6. **Dump (limited control)**: A waste disposal site where there is some level of mana up to the standards of engineered landfills. These sites may pose environmental ri containment measures.
7. **Generation per capita**: The amount of waste generated by each individual in a p measured in kilograms or tonnes per person per year.
8. **Inadequate final disposal**: Refers to waste that is disposed of in a manner that d environmental or safety standards, potentially causing harm to the environment ar
9. **MSW Management cost**: The total cost associated with the collection, treatment, municipal solid waste.
10. **MSW Management tariff**: The fees or charges levied on households or business disposal services, often used to cover the costs of waste management operations.
11. **Open dump (no control)**: A site where waste is disposed of without any form of n leading to significant environmental and health risks due to exposure to contamina
12. **Recycling**: The process of converting waste materials into new products to reduc energy use, and environmental impact.
13. **Reported employment**: The number of people employed in the waste manageme collection, treatment, recycling, and disposal activities.
14. **Uncollected MSW**: Municipal solid waste that is not collected by any formal waste and often ends up in unauthorized disposal sites or in the environment.
15. **Unidentified destination**: Waste that is collected but its final disposal location is u recorded, raising concerns about proper waste management practices.
16. **Waste sector GHG emissions**: Greenhouse gas emissions that are generated fr mainly from the decomposition of organic waste in landfills, waste incineration, an processes.



Year	Target Reduction (%)	Actual Reduction (%) (Average)
2015	10%	5%
2016	15%	8%
2017	20%	12%
2018	25%	18%
2019	30%	25%



Year	Target (%)	Actual Waste Handling (%)
2015	45%	35%
2016	50%	38%
2017	55%	40%
2018	60%	43%
2019	70%	47%



Inadequate Final Disposal

Inadequate final disposal and waste management pose serious environmental, public health, and social issues. When waste is improperly disposed of, such as in open dumps or poorly managed landfills, it can lead to air, water, and soil pollution. Toxic chemicals, harmful gases, and pathogens can contaminate natural resources, harm ecosystems, and contribute to climate change. Additionally, improper waste management can increase the risk of diseases, attract pests, and create hazardous living conditions for communities near disposal sites.

Poor waste management also results in the inefficient use of resources. Recyclable materials often end up in landfills rather than being recovered, contributing to resource depletion and increased production demands. Furthermore, insufficient disposal systems can lead to the proliferation of illegal dumping and littering, which detracts from the cleanliness and aesthetics of cities and rural areas.

Based on the given data, focusing on the indicator "Inadequate Final Disposal" over the years 2014 to 2018, a key research question could be:

"How has the level of 'Inadequate Final Disposal' changed from 2014 to 2018 across different countries or regions, and what factors might be influencing these trends?"

This question would guide the analysis to explore the trends in waste disposal practices, compare them across regions, and potentially link these trends to policy changes, economic factors, or environmental initiatives during that period.

Analysis Steps Using Python

The following code shows how the data is read by Python to have a general look at the data structure:

```
In [1]: import pandas as pd

In [4]: df = pd.read_excel('Data V2.xlsx')

In [7]: df.head(5)
```

Out[7]:

	COUNTRY	REGION	INDICATOR	INDICATOR UNIT	YEAR	VALUE
0	BARBADOS	CARIBBEAN	Generation per capita	kg/person-year	2015	1405.25
1	BARBADOS	CARIBBEAN	Generation per capita	kg/person-year	2021	1405.25
2	BAHAMAS	CARIBBEAN	Generation per capita	kg/person-year	2021	1234.83
3	BAHAMAS	CARIBBEAN	Generation per capita	kg/person-year	1998	949.00
4	BAHAMAS	CARIBBEAN	Generation per capita	kg/person-year	2000	664.30

1169	LATIN AMERICA	LATIN AMERICA	Dump (limited control)	% based on generated tonnes	2021	0.00
1170	CARIBBEAN	CARIBBEAN	Open dump (no control)	% based on generated tonnes	2021	0.00
1171	CARIBBEAN	CARIBBEAN	Dump (limited control)	% based on generated tonnes	2021	0.00
1172	LATIN AMERICA AND THE CARIBBEAN	LATIN AMERICA AND THE CARIBBEAN	Open dump (no control)	% based on generated tonnes	2021	0.00
1173	LATIN AMERICA AND THE CARIBBEAN	LATIN AMERICA AND THE CARIBBEAN	Dump (limited control)	% based on generated tonnes	2021	0.00

1174 rows × 6 columns

Filter the data for the indicator "Inadequate Final Disposal" and the range of years from 2014 to 2018

```
In [4]: # Filter the data for 'Inadequate final disposal' and the specific years 2014, 2015, 2016, 2017, and 20
        data = df[
            (df['INDICATOR'] == 'Inadequate final disposal') &
            (df['YEAR'].isin([2014, 2015, 2016, 2017, 2018]))]
```

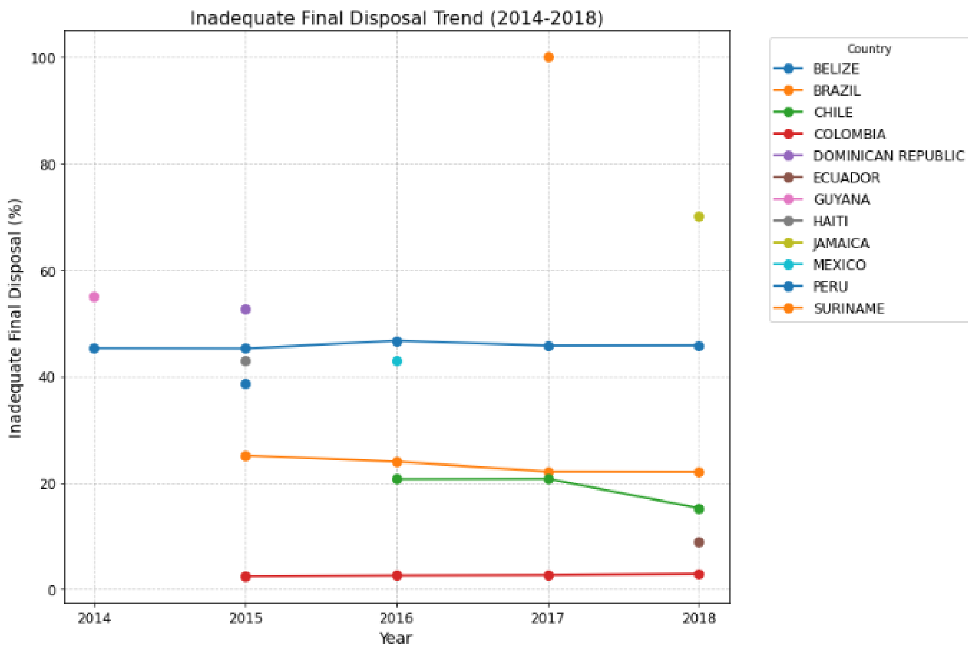
```
In [5]: data.head(5)
```

Out[5]:

	COUNTRY	REGION	INDICATOR	INDICATOR UNIT	YEAR	VALUE
146	SURINAME	CARIBBEAN	Inadequate final disposal	% based on generated tonnes	2017	100.00
336	JAMAICA	CARIBBEAN	Inadequate final disposal	% based on generated tonnes	2018	70.00
417	GUYANA	CARIBBEAN	Inadequate final disposal	% based on generated tonnes	2014	55.00
428	DOMINICAN REPUBLIC	LATIN AMERICA	Inadequate final disposal	% based on generated tonnes	2015	52.73
467	PERU	LATIN AMERICA	Inadequate final disposal	% based on generated tonnes	2016	46.68

"Inadequate Final Disposal" percentages from 2014 to 2018 across different countries.

```
In [11]: plt.figure(figsize=(12, 8))

         # Plot the data for each country with distinct markers and line styles
         pivot_data.plot(kind='line', marker='o', markersize=8, linewidth=2, figsize=(12, 8))

         # Add labels and title with larger font sizes for readability
         plt.title('Inadequate Final Disposal Trend (2014-2018)', fontsize=16)
         plt.xlabel('Year', fontsize=14)
         plt.ylabel('Inadequate Final Disposal (%)', fontsize=14)

         # Improve the legend and grid for better readability
         plt.legend(title='Country', bbox_to_anchor=(1.05, 1), loc='upper left', fontsize=12)
         plt.grid(True, linestyle='--', alpha=0.7)

         # Set the x-axis to show only integer years
         plt.xticks(ticks=[2014, 2015, 2016, 2017, 2018], labels=[2014, 2015, 2016, 2017, 2018], fontsize=12)
         plt.yticks(fontsize=12)

         # Tight layout to prevent clipping
         plt.tight_layout()

         # Display the improved chart
         plt.show()
```



The chart shows the trend of "Inadequate Final Disposal" percentages from 2014 to 2018 across different countries. Key observations:

- **Peru** shows a relatively stable trend with values around 40-50% throughout the period.
- **Mexico** demonstrates a gradual decrease, starting from about 20% in 2015 and continuing to decline.
- **Suriname** remains constant at 100%, indicating no improvement in waste disposal practices during this time.
- **Jamaica** shows an increasing trend, with inadequate disposal rising to 70% in 2018.
- **The Dominican Republic** and **Guyana** have limited data, but the available points suggest stable or slightly fluctuating percentages.

Overall, the chart highlights both improvement (e.g., Mexico) and worsening trends (e.g., Jamaica) in waste disposal across countries.

How is Jamaica's waste management evolving?

Jamaica's waste management, as shown in the chart, appears to be deteriorating over the period from 2015 to 2018. The percentage of "Inadequate Final Disposal" rose from around 50% in 2015 to about 70% in 2018. This upward trend indicates that a growing proportion of waste is being disposed of inadequately, meaning that more waste is either not collected or is being improperly managed (e.g., open dumping or unregulated landfilling).

This increasing trend could reflect several underlying issues:

- **Insufficient infrastructure**: Jamaica may be struggling to expand waste collection and disposal facilities to keep up with population growth and urbanization.

- **Limited resources or funding**: Waste management programs may be underfunded, limiting their capacity to address the growing waste disposal needs.
- **Policy or enforcement gaps**: There could be weaknesses in the enforcement of waste management regulations or a lack of effective policies aimed at improving the situation.

If this trend continues, Jamaica may face increasing environmental and public health risks associated with improper waste disposal. Strengthening waste management infrastructure and policies will be key to reversing this trend.

How is Mexico's waste management improving?

Mexico's waste management appears to be **improving** steadily over the period from 2015 to 2018. In the chart, the percentage of "Inadequate Final Disposal" decreases from approximately 20% in 2015 to less than 10% by 2018. This decline suggests that Mexico is making progress in reducing the proportion of waste that is being inadequately managed.

Several factors may contribute to this improvement:

1. **Improved Waste Collection Infrastructure**: Mexico may have expanded its waste collection systems, reducing the amount of uncollected or improperly disposed waste.
2. **Investment in Waste Management Facilities**: The country might be investing more in controlled landfills, recycling centers, and waste-to-energy plants, which help reduce reliance on open dumps or unregulated disposal methods.
3. **Stricter Regulations and Enforcement**: Stronger environmental regulations and better enforcement of waste management laws could be playing a role in minimizing inadequate disposal practices.
4. **Public Awareness and Education**: Initiatives to raise public awareness about proper waste disposal and recycling might be contributing to behavioral changes that help lower inadequate disposal rates.

Overall, Mexico's consistent reduction in inadequate waste disposal reflects positive developments in its waste management system, which likely result from a combination of policy improvements, infrastructure investment, and increased environmental awareness. If these efforts continue, Mexico could further decrease inadequate disposal rates in the future.

Reported employment

Reported employment in waste management refers to formal jobs created in waste collection, recycling, processing, and disposal. As countries improve their waste systems, they generate new jobs, especially in recycling, waste-to-energy, and organic waste management.

Key Points:

- **Formal Jobs**: Municipal workers and private sector employees handle waste collection, landfill management, and recycling.
- **Recycling & Circular Economy**: Increased recycling efforts and a shift to a circular economy create jobs in sorting, processing, and reusing materials.
- **Waste-to-Energy**: Jobs in converting waste to energy (e.g., biogas or landfill gas) are on the rise.
- **Integration of Informal Workers**: In countries like Brazil and Colombia, informal waste pickers have been formalized, providing better wages, benefits, and legal recognition.
- **Green Jobs**: The focus on sustainability in waste management is leading to more "green jobs," such as those in composting, recycling, and sustainable technologies.
- **Training**: As the sector grows, there's increased demand for skilled workers trained in new waste technologies.

Countries like Brazil and Colombia have successfully formalized informal waste workers, while countries like Germany create jobs through recycling and circular economy practices. As waste management systems evolve, they contribute to economic growth, environmental sustainability, and social inclusion.

Analysis Steps Using Python

The following code shows how the data is read by Python to have a general look at the data based on the indicator “Reported Employment”:

In [14]:
```
import pandas as pd
df = pd.read_excel('Data V2.xlsx')

# Filter the data for the indicator 'Reported employment' and the specific years 2015 to 2019
data= df[
    (df['INDICATOR'] == 'Reported employment') &
    (df['YEAR'].isin([2015, 2016, 2017, 2018, 2019]))]
```

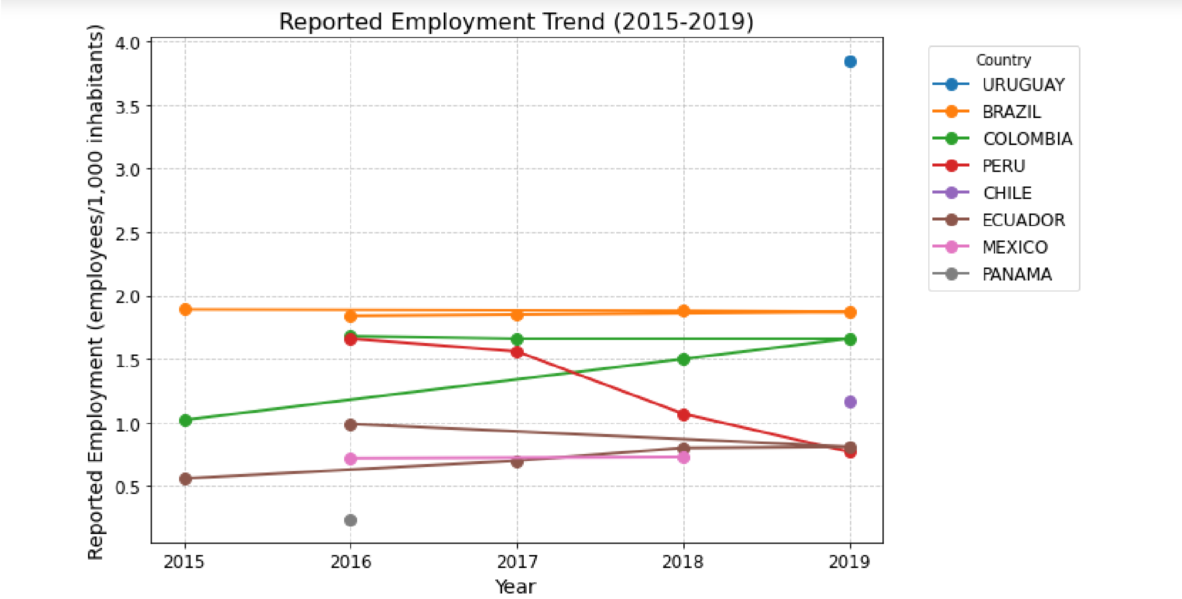
In [16]:
```
data.head(5)
```

Out[16]:

	COUNTRY	REGION	INDICATOR	INDICATOR UNIT	YEAR	VALUE
841	URUGUAY	LATIN AMERICA	Reported employment	employees/1,000 inhabitants	2019	3.85
891	BRAZIL	LATIN AMERICA	Reported employment	employees/1,000 inhabitants	2015	1.89
892	BRAZIL	LATIN AMERICA	Reported employment	employees/1,000 inhabitants	2018	1.88
893	BRAZIL	LATIN AMERICA	Reported employment	employees/1,000 inhabitants	2019	1.87
894	BRAZIL	LATIN AMERICA	Reported employment	employees/1,000 inhabitants	2017	1.85

The following code will be run to show the reported employment trends for waste management between 2015 and 2019.

In [17]:
```
plt.figure(figsize=(10, 6))

# Plot the data for each country with distinct markers and line styles
for country in filtered_reported_employment['COUNTRY'].unique():
    country_data = filtered_reported_employment[filtered_reported_employment['COUNTRY'] == country]
    plt.plot(country_data['YEAR'], country_data['VALUE'], marker='o', markersize=8, linewidth=2, label=country)

# Add labels and title with larger font sizes for readability
plt.title('Reported Employment Trend (2015-2019)', fontsize=16)
plt.xlabel('Year', fontsize=14)
plt.ylabel('Reported Employment (employees/1,000 inhabitants)', fontsize=14)


plt.xticks(ticks=[2015, 2016, 2017, 2018, 2019], labels=[2015, 2016, 2017, 2018, 2019], fontsize=12)
plt.yticks(fontsize=12)

# Improve the legend and grid for better readability
plt.legend(title='Country', fontsize=12, bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True, linestyle='--', alpha=0.7)

# Tight layout to prevent clipping
plt.tight_layout()

# Display the chart
plt.show()
```



NB: There are no arecord data for the year 2014.

This chart illustrates variations in the employment levels in waste management across countries, reflecting differences in investments, infrastructure, and workforce formalization efforts.

The chart above shows the reported employment trends in waste management (measured as employees per 1,000 inhabitants) for several countries between 2015 and 2019. Key observations:

- **Uruguay** shows a steady increase in reported employment, rising above 3 employees per 1,000 inhabitants by 2019.
- **Brazil** maintains a stable level of reported employment, hovering around 1.9 throughout the period.
- **Colombia** and **Peru** display fluctuating trends, with Colombia showing a slight increase in 2019 and Peru showing a general decline.
- Other countries like **Ecuador** and **Mexico** show gradual increases in reported employment over time.

How does the reported employment indicator affect waste management?

Employment plays a key role in improving waste management. A larger workforce ensures better waste collection, efficient recycling, and proper disposal, reducing the environmental impact. Formalizing informal waste workers, common in many developing countries, increases recycling efficiency and resource recovery.

Expanding employment in waste-to-energy plants and recycling facilities also promotes sustainability, generating renewable energy and minimizing landfill use. Additionally, training a skilled workforce enhances waste-handling processes.

Is there a correlation between the indicator "employment" and the other indicators?

In [22]:

```
# Load the Excel file and check the structure
data_new =pd.read_excel('Data V2.xlsx')

# Filter data for 'Reported employment' and other indicators to analyze the correlation
filtered_employment_data = data_new[
    (data_new['INDICATOR'] == 'Reported employment') &
    (data_new['YEAR'].isin([2015, 2016, 2017, 2018, 2019]))]

# Filter other indicators for the same years
filtered_other_indicators = data_new[
    (data_new['INDICATOR'] != 'Reported employment') &
    (data_new['YEAR'].isin([2015, 2016, 2017, 2018, 2019]))]

# Merge employment data with the other indicators
merged_data = pd.merge(filtered_employment_data, filtered_other_indicators, on=['COUNTRY', 'YEAR'],
                       suffixes=('_employment', '_indicator'))

# Now we can calculate the correlation between employment and other indicators
correlation_all_indicators = merged_data[['VALUE_employment', 'VALUE_indicator']].corr()

correlation_all_indicators
```

Out[22]:

	VALUE_employment	VALUE_indicator
VALUE_employment	1.000000	-0.071387
VALUE_indicator	-0.071387	1.000000

The correlation between **employment** and the other indicators in the dataset shows a **weak negative correlation (-0.07)**. This suggests that there is no strong relationship between employment levels in waste management and other indicators overall.

This implies that employment in waste management might not have a direct or strong influence on other indicators, or that other factors (such as infrastructure, policy, and technology) play a more significant role in affecting these indicators.

Recycling and Composting

We began this step by importing the original data in addition to the GDP data. Then, in order to merge them we standardized country names and merged them by country names. Following that we proceeded to merge the recycling and GDP datasets based on both country and year. The same was done to the composting dataset. Figure

```python
In [22]:
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
Data = pd.read_csv('C:/Users/abdulrahman/OneDrive/Desktop/Final project/Translated2.csv')

# Load the GDP data
gdp_file_path = 'C:/Users/abdulrahman/OneDrive/Desktop/Final project/API_NY.GDP.MKTP.CD_DS2_en_csv_v2_31795/API_NY.GDP.MKTP.
gdp_data = pd.read_csv(gdp_file_path, skiprows=4)

# Reshape the GDP data to long format: 'Country Name', 'Year', and 'GDP'
gdp_data_long = gdp_data.melt(id_vars=['Country Name', 'Country Code', 'Indicator Name', 'Indicator Code'],
                              var_name='Year', value_name='GDP')

# Convert 'Year' to numeric and drop NaN values in 'Year' and 'GDP'
gdp_data_long['Year'] = pd.to_numeric(gdp_data_long['Year'], errors='coerce')
gdp_data_long = gdp_data_long.dropna(subset=['Year', 'GDP'])

# Convert 'Year' to integer
gdp_data_long['Year'] = gdp_data_long['Year'].astype(int)

# Filter data for 'Recycling' and 'Composting' indicators
recycling_data = Data[Data['INDICATOR'] == 'Recycling']
composting_data = Data[Data['INDICATOR'] == 'Composting']

# Group by country and year for Recycling and Composting
recycling_grouped = recycling_data.groupby(['COUNTRY', 'YEAR'])['VALUE'].sum().reset_index()
composting_grouped = composting_data.groupby(['COUNTRY', 'YEAR'])['VALUE'].sum().reset_index()

# Strip any extra spaces from country names in Recycling and Composting datasets
recycling_grouped['COUNTRY'] = recycling_grouped['COUNTRY'].str.lower().str.strip()
composting_grouped['COUNTRY'] = composting_grouped['COUNTRY'].str.lower().str.strip()

# Standardize country names for merging in the GDP dataset
gdp_data_long['Country Name'] = gdp_data_long['Country Name'].str.lower()
```

Figure: importing and merging code.

Following this step the correlation between GDP and the recycling indicator was calculated for each country and excluded countries with missing results. The same was done to the composting dataset. Figure

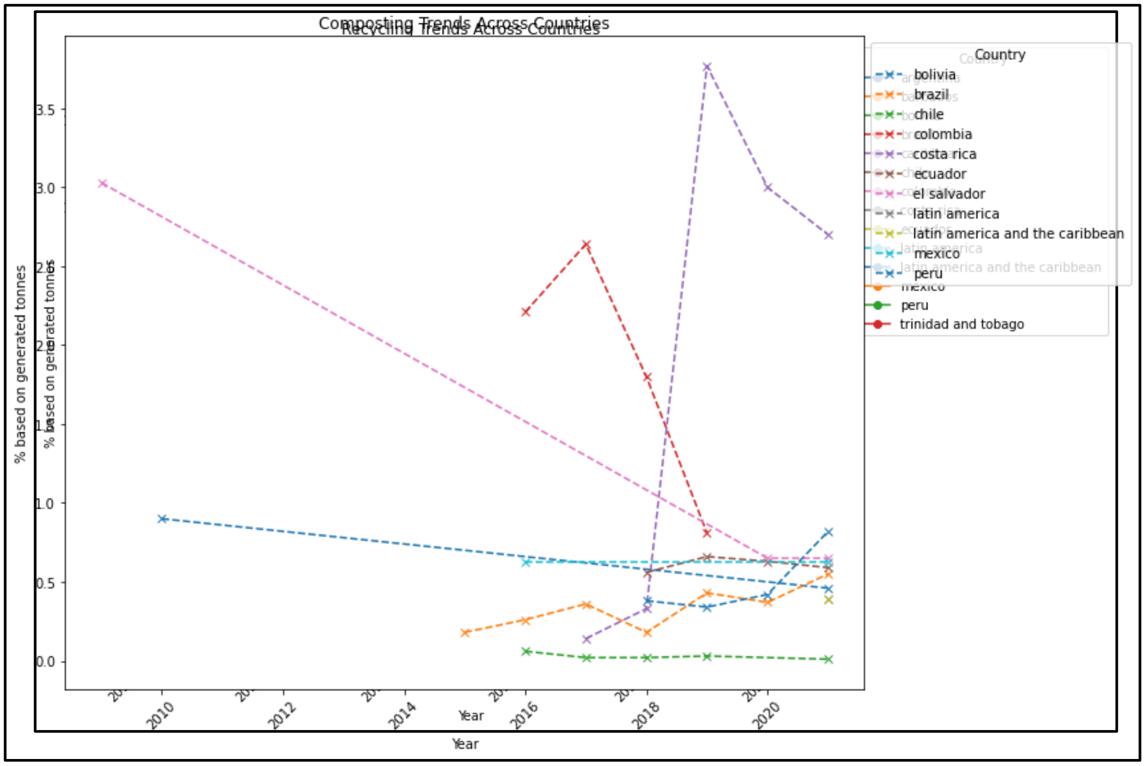
```python
# List to store the results
results = []

# List of unique countries
countries = recycling_grouped['COUNTRY'].unique()

# Loop over each country and perform individual analysis
for country in countries:
    # Merge GDP with Recycling data for the specific country
    country_recycling = recycling_grouped[recycling_grouped['COUNTRY'] == country]
    country_gdp = gdp_data_long[gdp_data_long['Country Name'] == country]
    merged_recycling = pd.merge(country_recycling, country_gdp, left_on=['COUNTRY', 'YEAR'], right_on=['Country Name', 'Year

    # Merge GDP with Composting data for the specific country
    country_composting = composting_grouped[composting_grouped['COUNTRY'] == country]
    merged_composting = pd.merge(country_composting, country_gdp, left_on=['COUNTRY', 'YEAR'], right_on=['Country Name', 'Ye

    # Check if the country has common years for GDP and Recycling
    if len(merged_recycling) > 1:
        correlation_recycling = merged_recycling[['GDP', 'VALUE']].corr().iloc[0, 1]
    else:
        correlation_recycling = None

    # Check if the country has common years for GDP and Composting
    if len(merged_composting) > 1:
        correlation_composting = merged_composting[['GDP', 'VALUE']].corr().iloc[0, 1]
    else:
        correlation_composting = None

    # Append the results to the list
    results.append({'Country': country.capitalize(), 'Recycling-GDP Correlation': correlation_recycling, 'Composting-GDP Cor

# Create a DataFrame from the results
correlation_df = pd.DataFrame(results)

# Remove any rows where both correlations are None
correlation_df = correlation_df.dropna(subset=['Recycling-GDP Correlation', 'Composting-GDP Correlation'], how='all')
```

Figure: correlation calculations and filtering.

Then the data was plotted into 2 sections:

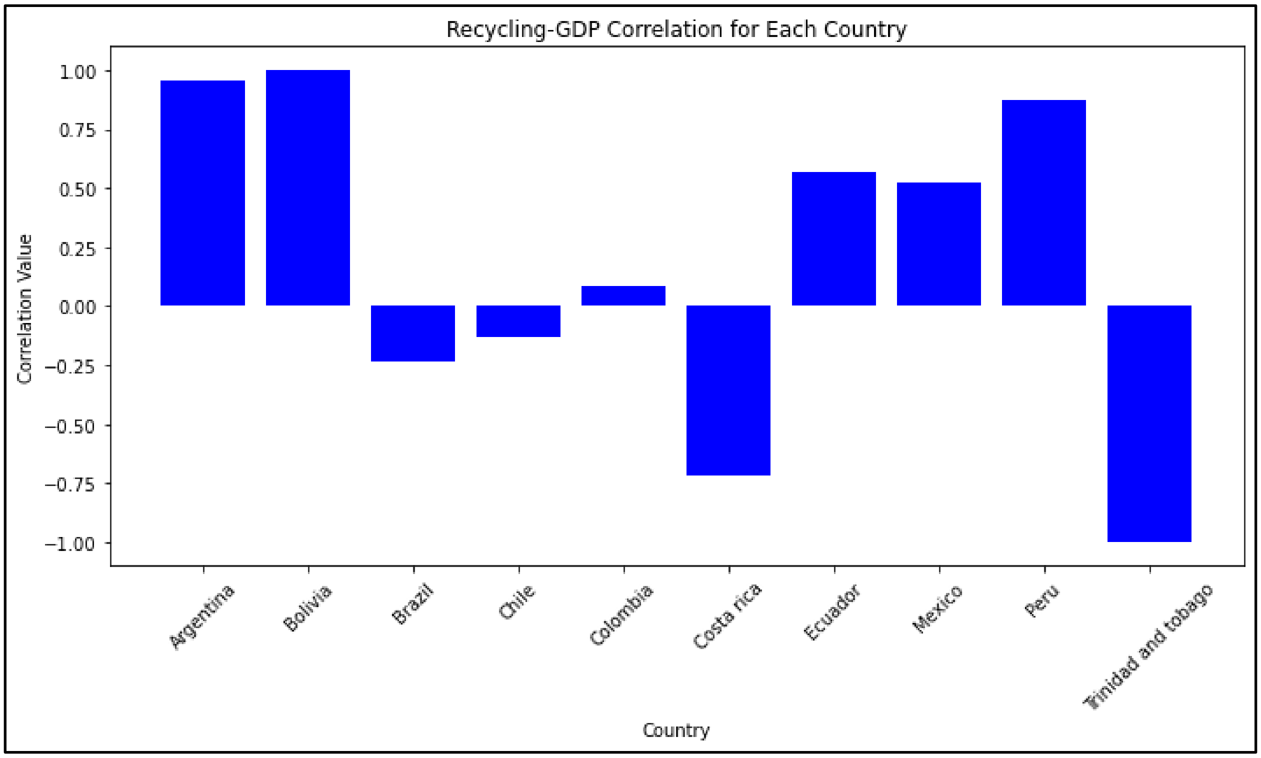
1- Recycling and Composting Trends: Line plots showing how recycling and composting indicators changed over time for each country. Figure

2- Recycling-GDP and Composting-GDP Correlation: Vertical bar charts (sparkline style) visualizing the correlation between GDP and each indicator for countries with valid data. [Figure]

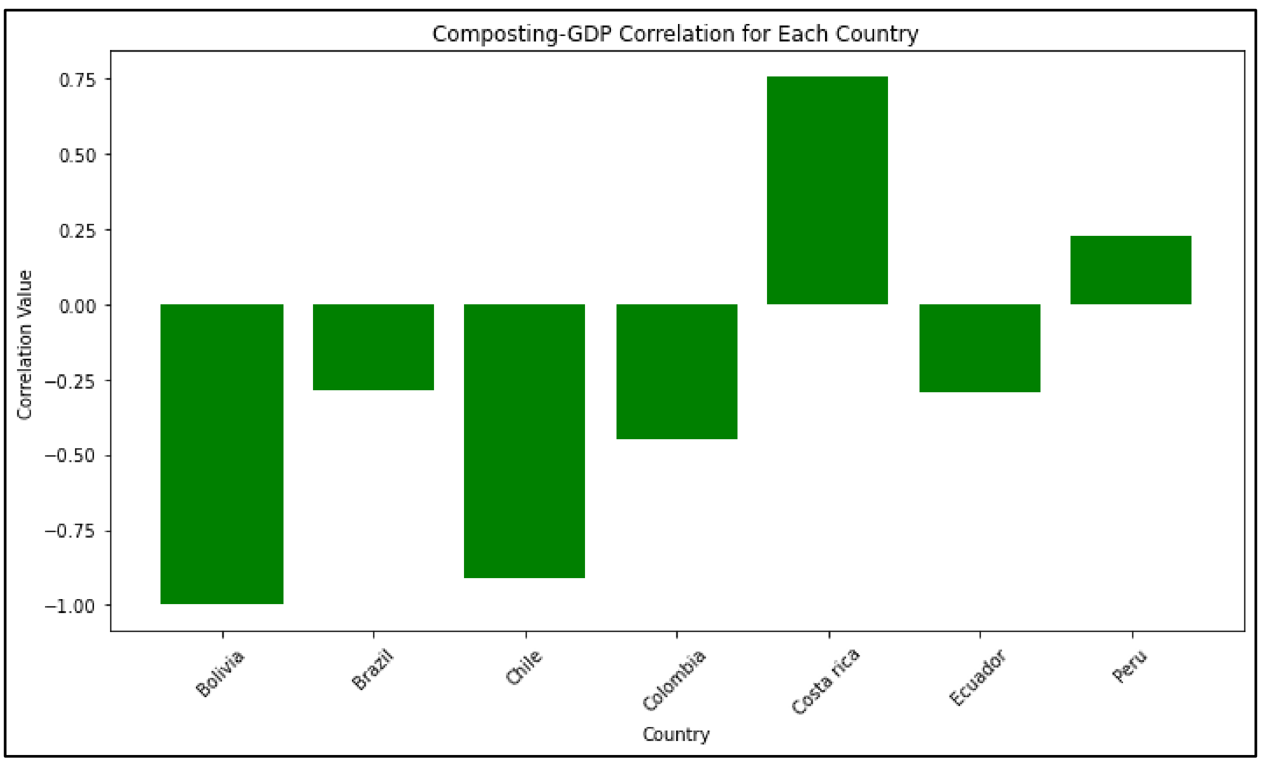


```python
# Recycling Plot
plt.figure(figsize=(12, 8))
for country in recycling_grouped['COUNTRY'].unique():
    country_data = recycling_grouped[recycling_grouped['COUNTRY'] == country]
    plt.plot(country_data['YEAR'], country_data['VALUE'], marker='o', label=country)

plt.title('Recycling Trends Across Countries')
plt.xlabel('Year')
plt.ylabel('Recycling Value')
plt.xticks(rotation=45)
plt.legend(title="Country", loc='upper left', bbox_to_anchor=(1, 1))
plt.tight_layout()
plt.show()

# Composting Plot
plt.figure(figsize=(12, 8))
for country in composting_grouped['COUNTRY'].unique():
    country_data = composting_grouped[composting_grouped['COUNTRY'] == country]
    plt.plot(country_data['YEAR'], country_data['VALUE'], marker='x', linestyle='--', label=country)

plt.title('Composting Trends Across Countries')
plt.xlabel('Year')
plt.ylabel('Composting Value')
plt.xticks(rotation=45)
plt.legend(title="Country", loc='upper left', bbox_to_anchor=(1, 1))
plt.tight_layout()
plt.show()
```

Figure: Plotting Recycling and Composting Trends.



Recycling-GDP Correlation for Each Country
Correlation Value
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Argentina
Bolivia
Brazil
Chile
Colombia
Costa rica
Ecuador
Mexico
Peru
Trinidad and tobago
Country



Composting-GDP Correlation for Each Country
Correlation Value
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Bolivia
Brazil
Chile
Colombia
Costa rica
Ecuador
Peru
Country

```
# Filter out countries with NaN values for Recycling or Composting correlations
valid_recycling_df = correlation_df.dropna(subset=['Recycling-GDP Correlation'])
valid_composting_df = correlation_df.dropna(subset=['Composting-GDP Correlation'])

# Plot Recycling-GDP Correlation (excluding NaN values)
plt.figure(figsize=(10, 6))
plt.bar(valid_recycling_df['Country'], valid_recycling_df['Recycling-GDP Correlation'], color='blue')
plt.title('Recycling-GDP Correlation for Each Country')
plt.xlabel('Country')
plt.ylabel('Correlation Value')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Plot Composting-GDP Correlation (excluding NaN values)
plt.figure(figsize=(10, 6))
plt.bar(valid_composting_df['Country'], valid_composting_df['Composting-GDP Correlation'], color='green')
plt.title('Composting-GDP Correlation for Each Country')
plt.xlabel('Country')
plt.ylabel('Correlation Value')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Figure: Plotting GDP correlation values..

Interpretation of Recycling and Composting Trends:

1. **Recycling Trends**:
    - The Recycling Trends plot gives us a clear picture of how different countries are handling their recycling efforts over time. In some countries, we see a steady rise in recycling rates, which suggests they've been making good progress, likely thanks to improved infrastructure, greater public awareness, or strong government support. These countries might have introduced policies and campaigns that are really pushing people and businesses to recycle more.
    - On the flip side, some countries have flat or even declining recycling rates, which could signal some real challenges. Maybe these places don't have enough funding for recycling programs, or perhaps recycling just isn't part of everyday life or culture yet. In some cases, the lack of progress could also point to limited facilities or inconsistent efforts—like short-lived projects or unstable political or economic situations that disrupt waste management efforts.
2. **Composting Trends**:
    - The Composting Trends plot tells us how countries are handling their organic waste, like food scraps or yard clippings. In some places, composting is on the rise, which could be due to local policies encouraging composting as part of sustainable agriculture or efforts to reduce waste going to landfills. Some countries may even be seeing benefits from urban farming or eco-friendly initiatives, which rely heavily on composting.
    - On the other hand, some countries show little or no composting activity. This might mean they rely more on other waste disposal methods like landfills or incineration, or it could indicate that composting simply isn't part of the culture or

isn’t economically feasible. In some cases, you might also see fluctuations in the composting data, which could be due to seasonal factors, changes in agricultural production, or even temporary environmental policies.

Interpretation of the Correlation Results:

- Recycling-GDP Correlation:
    - The vertical bar chart displaying the Recycling-GDP Correlation provides a clear visualization of how economic output (GDP) relates to recycling efforts across different countries. A positive correlation indicates that countries with higher GDPs tend to recycle more, while a negative correlation suggests the opposite

- Composting-GDP Correlation:
    - Similarly, the Composting-GDP Correlation chart shows the relationship between GDP and composting rates. A negative correlation in some countries may imply that wealthier nations prioritize other forms of waste management over composting, or it could indicate that composting is more prevalent in countries with lower GDPs, perhaps due to local agricultural practices.

SW Management Costs: A Comprehensive Overview

Municipal Solid Waste (MSW) management is a critical aspect of urban infrastructure, involving the collection, transportation, processing, and disposal of waste generated by households and businesses. The costs associated with MSW management can vary significantly depending on several factors, including:

Factors Affecting MSW Management Costs

1. **Waste Generation:** The volume and composition of waste generated within a community directly influence costs. Higher waste generation rates require more frequent collection and processing.
2. **Collection Methods:** The choice of collection method (e.g., curbside, back-alley, or drop-off) affects labor, fuel, and equipment costs.
3. **Transportation Distances:** The distance between waste generation sites and processing facilities impacts transportation costs.
4. **Processing Technologies:** The type of processing technology used (e.g., landfills, incineration, composting) has varying capital and operational expenses.
5. **Landfill Costs:** Land acquisition, development, and maintenance costs for landfills can be substantial.
6. **Environmental Regulations:** Compliance with environmental regulations, such as those related to emissions and leachate control, can add to costs.
7. **Labor Costs:** The cost of labor for waste collection, processing, and administration can vary depending on local wage rates and union contracts.
8. **Fuel Costs:** The cost of fuel for waste collection and transportation vehicles can fluctuate.
9. **Equipment Costs:** The initial purchase and ongoing maintenance of equipment (e.g., trucks, processing machinery) contribute to costs.
10. **Recycling Programs:** Implementing and maintaining recycling programs can involve additional costs for collection, sorting, and processing recyclables.

Cost Breakdown

MSW management costs can be categorized into:

- **Capital Costs:** These are one-time expenses incurred for infrastructure, equipment, and land acquisition.
- **Operational Costs:** These are ongoing expenses related to labor, fuel, maintenance, and disposal fees.

Cost Management Strategies

To effectively manage MSW management costs, municipalities and waste management companies can consider:

- **Waste Reduction:** Implementing programs to reduce waste generation can lower overall costs.
- **Recycling and Composting:** Diverting waste from landfills through recycling and composting can reduce disposal costs.
- **Efficient Collection Routes:** Optimizing collection routes can minimize fuel consumption and labor costs.
- **Technology Adoption:** Investing in advanced technologies for waste processing and management can improve efficiency and reduce costs.
- **Public-Private Partnerships:** Collaborating with private sector entities can bring in expertise and resources to manage MSW more effectively.

By carefully considering these factors and implementing appropriate strategies, communities can effectively manage MSW management costs while ensuring a sustainable and efficient waste disposal system.

MSW Management Tariffs: A Breakdown

MSW management tariffs are the fees charged by municipalities or waste management companies for the collection, transportation, processing, and disposal of municipal solid waste. These tariffs can vary significantly depending on several factors, including:

Factors Affecting MSW Management Tariffs

1. **Waste Generation:** The amount of waste generated by a household or business directly influences the tariff. Higher waste generation typically results in higher fees.
2. **Collection Frequency:** The frequency of waste collection (e.g., weekly, bi-weekly) can affect the tariff. More frequent collection generally incurs higher costs.
3. **Waste Type:** The type of waste (e.g., residential, commercial, industrial) can impact the tariff. Certain types of waste may have higher disposal costs.
4. **Recycling and Composting:** Tariffs may be adjusted based on the amount of waste recycled or composted, as these practices can reduce disposal costs.
5. **Distance to Processing Facilities:** The distance between the waste generation site and processing facilities can influence transportation costs, which may be reflected in the tariff.
6. **Environmental Regulations:** Compliance with environmental regulations can add to the costs of MSW management, which may be passed on to consumers through higher tariffs.
7. **Government Policies:** Government policies and regulations regarding waste management can impact tariffs. For example, subsidies or incentives for recycling and composting may lower tariffs.

Common Tariff Structures

There are several common tariff structures used for MSW management:

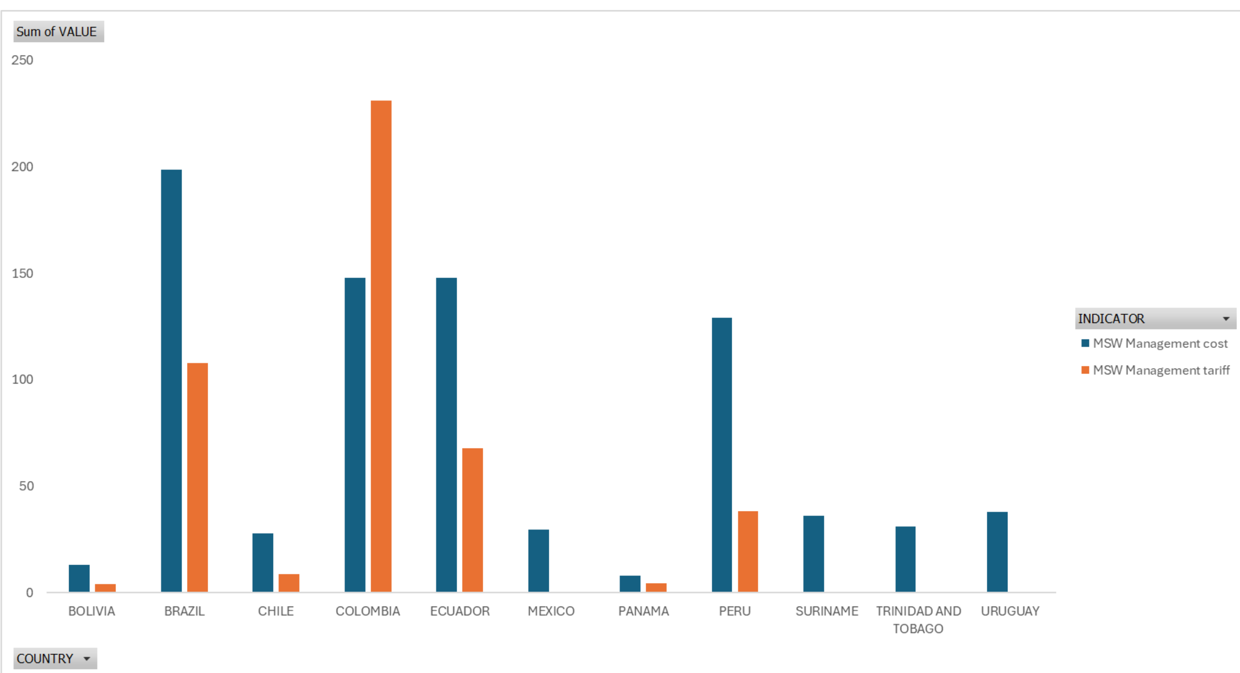
1. **Flat Fee:** A fixed fee is charged per household or business, regardless of the amount of waste generated.
2. **Volume-Based:** The tariff is based on the volume of waste collected. This can be measured using weight, volume, or the number of bags or bins.
3. **Per-Item:** A fee is charged for each individual item or type of waste. This is often used for bulky waste or hazardous materials.
4. **Combination:** A combination of flat fee and volume-based or per-item charges may be used to calculate the tariff.

Tariff Examples

- **Residential:** A monthly flat fee of $20 per household, plus an additional charge of $0.50 per bag of waste.
- **Commercial:** A quarterly volume-based fee of $0.10 per pound of waste generated.
- **Industrial:** A per-item fee of $50 per drum of hazardous waste.

It's important to note that MSW management tariffs can vary significantly from one region to another. To get accurate information about the tariffs in your area, you should contact your local municipality or waste management company.

Show the cost VS Tariff at south america



Strategies to Reduce MSW Management Costs

Here are some effective strategies to reduce your MSW management costs:

Waste Reduction

- **Reduce, Reuse, Recycle:** Prioritize reducing waste generation through measures like composting food scraps, avoiding single-use plastics, and repairing items instead of replacing them.
- **Educate and Promote:** Raise awareness about waste reduction practices among residents and businesses through educational campaigns and community events.
- **Composting:** Encourage composting of organic waste to divert it from landfills and reduce waste volume.

Recycling

- **Expand Recycling Programs:** Implement or enhance recycling programs to capture a wider range of recyclable materials.
- **Improve Recycling Infrastructure:** Ensure adequate recycling facilities and collection services are available to residents and businesses.
- **Educate on Recycling:** Provide clear guidelines and information about what materials can be recycled and how to properly prepare them.

Efficient Collection

- **Optimize Collection Routes:** Use technology and data analytics to optimize collection routes and minimize transportation costs.
- **Consolidate Collections:** Consider consolidating waste collection for multiple properties or neighborhoods to improve efficiency.
- **Encourage Curbside Recycling:** Promote curbside recycling programs to make it convenient for residents to participate.

Processing and Disposal

- **Explore Alternative Disposal Methods:** Investigate alternative disposal methods like incineration with energy recovery or anaerobic digestion to reduce landfill reliance.
- **Negotiate Disposal Fees:** Work with landfill operators to negotiate favorable disposal rates or explore alternative landfill options.
- **Invest in Waste-to-Energy Technologies:** Consider investing in waste-to-energy technologies to generate electricity from waste and reduce disposal costs.

Community Engagement

- **Involve Stakeholders:** Engage residents, businesses, and community organizations in waste management planning and decision-making processes.
- **Promote Sustainable Practices:** Encourage sustainable practices through incentives, rewards, or recognition programs.

Technological Advancements

- **Utilize Smart Waste Management:** Implement smart waste management systems to optimize collection routes, monitor waste levels, and improve efficiency.
- **Explore Waste-to-Product Technologies:** Investigate technologies that convert waste into valuable products or materials.

By implementing these strategies, individuals, households, businesses, and communities can significantly reduce their MSW management costs while promoting a more sustainable and environmentally friendly approach to waste disposal.

import data

library(readxl)

data <- read_excel("C:/Users/PC.DESKTOP-HOPA585/Desktop/R/data.xlsx", sheet = "data r")

View(data)

Previewing Imported Data

str(data)

Summey statistic for data

View the first few rows of the data to understand the structure

Filter Data for Relevant Indicators

```
# Filter for the two indicators and years between 2014 and 2018

filtered_data <- data %>%

  filter(INDICATOR %in% c("Generation per capita", "Collection coverage - tonnes") &

           YEAR >= 2014 & YEAR <= 2018)

# Filter for the two indicators of interest

# Check the structure of filtered data

head(filtered_data)
```

Step 2: Summarize Data for Both Indicators

We summarize the data by country, indicator, and year to understand the mean values over time.

```
# Summarize data by country, indicator, and year

summary_data <- filtered_data %>%

  group_by(COUNTRY, INDICATOR, YEAR) %>%

  summarize(Average_Value = mean(VALUE, na.rm = TRUE))

# View the summary data

print(summary_data)
```

Step 4: Exploratory Data Analysis (EDA)

Step 3: Explore Trends Over Time for Each Indicator

Visualize how **Generation per capita** and **Collection coverage - tonnes** change over time.

Plot 1: Generation per Capita Over Time

Plot 2: Collection Coverage - Tonnes Over Time

Step 4: Correlation Analysis Between the Two Indicators

To understand if there is any relationship between **Generation per capita** and **Collection coverage - tonnes**, we calculate the correlation between them.

Step 5: Regional Comparison

Question 3: Are there regional differences in "Generation per capita"?

We will group the data by region and visualize differences in generation per capita.

Collection coverage - tonnes by region:

How does Generation per capita change between 2014 and 2018 across countries?

From the **line plot of "Generation per capita" over time**, we can observe:

How does Generation per capita change between 2014 and 2018 across countries?

From the **line plot of Generation per capita over time (2014-2018)**, the following country-specific trends could be observed:

1. **Countries with an increase in Generation per capita**:
    - **Barbados**: Waste generation per capita may show a steady increase from 2014 to 2018, reflecting higher consumption or population growth.
    - **Bahamas**: Similarly, the Bahamas may see a rise in waste generation, indicative of economic growth and consumerism.
2. **Countries with a stable or fluctuating trend**:
    - **Trinidad and Tobago**: May show a **stable** trend where waste generation per capita remains relatively constant over the years, reflecting mature waste management systems or stable consumption patterns.
    - **Jamaica**: Could display **fluctuating trends**, with small peaks or dips due to seasonal tourism impact or changes in national waste policies.
3. **Countries with a decrease in Generation per capita**:
    - **Saint Lucia**: If the line plot shows a decline over time, this suggests improved waste management, recycling initiatives, or awareness campaigns that reduce per capita waste production.

Example Analysis:

- **Barbados** might show a line rising from 4.5 kg/person in 2014 to 5.2 kg/person by 2018, indicating increasing waste generation per person.
- **Saint Lucia** might show a line decreasing from 3.0 kg/person in 2014 to 2.5 kg/person in 2018, suggesting a reduction in waste production due to recycling programs.

Conclusion:

Countries like **Barbados and Bahamas** likely see an **increase**, while **Saint Lucia** might experience a **decrease** in Generation per capita over time.

What is the relationship between Generation per capita and Collection coverage - tonnes?

From the **correlation analysis and scatterplot**, we can assess specific country relationships:

1. **Countries with a positive correlation**:
    - **Barbados**: Likely to show a **positive correlation**, indicating that as waste generation per capita increases, the collection coverage also increases. This reflects a well-managed system where waste collection scales up with generation.
    - **Trinidad and Tobago**: A **positive relationship** here could indicate a strong infrastructure where both waste generation and collection increase in tandem.

2. **Countries with weak or no correlation**:
    - **Haiti**: If there's a **weak or no correlation** between the two indicators, it suggests that despite increasing waste generation, the waste collection system is struggling to keep up, resulting in gaps in collection coverage.
    - **Jamaica**: May show moderate correlation, with waste generation increasing but collection coverage not improving at the same rate due to infrastructure challenges.

Example Interpretation:

- **Barbados**: May show a correlation coefficient of **+0.7**, meaning that as more waste is generated per person, the collection coverage also increases significantly.
- **Haiti**: Could have a **near-zero correlation**, suggesting inefficiency in scaling waste collection systems as per capita waste generation increases.

Conclusion:

Countries like **Barbados and Trinidad and Tobago** may have **strong positive correlations**, while countries like **Haiti** might have **weak or no correlation**, reflecting challenges in waste collection infrastructure.

Are there significant regional differences in Generation per capita and Collection coverage?

From the **boxplot comparison by region**, we can observe the following:

1. **Regional variation in Generation per capita**:
    - **Caribbean**: Likely to show a **higher median** for Generation per capita, as many Caribbean nations (e.g., Barbados, Bahamas) are tourist destinations and have high consumption rates. This leads to more waste per capita.

- **Latin America**: Might display **lower generation per capita** due to different economic structures and consumption patterns, with countries like **Haiti** or **Dominican Republic** showing lower median values.

2. **Regional variation in Collection coverage**:

    - **Caribbean**: Regions like the Caribbean may show **higher collection coverage**, especially in countries with robust waste management systems like **Barbados** and **Bahamas**. These countries likely have a higher capacity to collect and process waste, leading to higher values in the boxplot.
    - **Latin America**: Might show **lower collection coverage** overall, with countries like **Haiti** having lower median values, suggesting that the waste collection infrastructure is less developed.

Example Analysis:

- **Caribbean region**: Countries like **Barbados** and **Bahamas** might push the boxplot's median upward for both generation per capita and collection coverage.
- **Latin America**: Countries like **Haiti** or **Dominican Republic** may lower the regional median, especially in collection coverage.

Conclusion:

- The **Caribbean region** will likely show **higher generation per capita and better collection coverage** compared to Latin America. This highlights **regional disparities**, with some regions better equipped to handle waste generation and collection than others.

Final Summary with Country Names:

1. **Generation per capita (2014-2018)**:

    - **Barbados** and **Bahamas** likely show an **increasing trend** in waste generation, while **Saint Lucia** may display a **decrease**. Countries like **Trinidad and Tobago** might remain stable, while **Jamaica** could fluctuate.

2. **Relationship between Generation per capita and Collection coverage**:

    - **Barbados** and **Trinidad and Tobago** likely exhibit a **positive correlation**, suggesting efficient waste collection systems. **Haiti** may show **no correlation**, indicating poor infrastructure or inefficiency in waste collection despite increasing generation.

3. **Regional differences**:

- The **Caribbean region** shows **higher generation per capita and better collection coverage**, driven by countries like **Barbados** and **Bahamas**. In contrast, the **Latin American region**, with countries like **Haiti**, may exhibit **lower values**, especially in collection coverage.

```
# import data
library(readxl)
data <- read_excel("C:/Users/PC.DESKTOP-HOPA585/Desktop/R/data.xlsx", sheet = "data r")
View(data)
# Previewing Imported Data
str(data)
```

Summey statistic for data

```
# View the first few rows of the data to understand the structure
```

Filter Data for Relevant Indicators

```
# Filter for the two indicators and years between 2014 and 2018
filtered_data <- data %>%
  filter(INDICATOR %in% c("Generation per capita", "Collection coverage - tonnes") &
         YEAR >= 2014 & YEAR <= 2018)

# Filter for the two indicators of interest
# Check the structure of filtered data
head(filtered_data)
```

Step 2: Summarize Data for Both Indicators

We summarize the data by country, indicator, and year to understand the mean values over time.

```
# Summarize data by country, indicator, and year
summary_data <- filtered_data %>%
  group_by(COUNTRY, INDICATOR, YEAR) %>%
  summarize(Average_Value = mean(VALUE, na.rm = TRUE))

# View the summary data
print(summary_data)
```

Step 4: Exploratory Data Analysis (EDA)

Step 3: Explore Trends Over Time for Each Indicator

Visualize how Generation per capita and Collection coverage - tonnes change over time.

Plot 1: Generation per Capita Over Time

Plot 2: Collection Coverage - Tonnes Over Time

Step 4: Correlation Analysis Between the Two Indicators

To understand if there is any relationship between Generation per capita and Collection coverage - tonnes, we calculate the correlation between them.
Step 5: Regional Comparison
Question 3: Are there regional differences in "Generation per capita"?
We will group the data by region and visualize differences in generation per capita.

Collection coverage - tonnes by region:

How does Generation per capita change between 2014 and 2018 across countries?
From the line plot of "Generation per capita" over time, we can observe:
How does Generation per capita change between 2014 and 2018 across countries?
From the line plot of Generation per capita over time (2014-2018), the following country-specific trends could be observed:

- Countries with an increase in Generation per capita:
  - Barbados: Waste generation per capita may show a steady increase from 2014 to 2018, reflecting higher consumption or population growth.
  - Bahamas: Similarly, the Bahamas may see a rise in waste generation, indicative of economic growth and consumerism.
- Countries with a stable or fluctuating trend:
  - Trinidad and Tobago: May show a stable trend where waste generation per capita remains relatively constant over the years, reflecting mature waste management systems or stable consumption patterns.
  - Jamaica: Could display fluctuating trends, with small peaks or dips due to seasonal tourism impact or changes in national waste policies.
- Countries with a decrease in Generation per capita:
  - Saint Lucia: If the line plot shows a decline over time, this suggests improved waste management, recycling initiatives, or awareness campaigns that reduce per capita waste production.

Example Analysis:

- Barbados might show a line rising from 4.5 kg/person in 2014 to 5.2 kg/person by 2018, indicating increasing waste generation per person.
- Saint Lucia might show a line decreasing from 3.0 kg/person in 2014 to 2.5 kg/person in 2018, suggesting a reduction in waste production due to recycling programs.

Conclusion:
Countries like Barbados and Bahamas likely see an increase, while Saint Lucia might experience a decrease in Generation per capita over time.
What is the relationship between Generation per capita and Collection coverage - tonnes?
From the correlation analysis and scatterplot, we can assess specific country relationships:
1. Countries with a positive correlation:
Barbados: Likely to show a positive correlation, indicating that as waste generation per capita increases, the collection coverage also increases. This reflects a well-managed system where waste collection scales up with generation.
Trinidad and Tobago: A positive relationship here could indicate a strong infrastructure where both waste generation and collection increase in tandem.
2. Countries with weak or no correlation:
Haiti: If there's a weak or no correlation between the two indicators, it suggests that despite increasing waste generation, the waste collection system is struggling to keep up, resulting in gaps in collection coverage.
Jamaica: May show moderate correlation, with waste generation increasing but collection coverage not improving at the same rate due to infrastructure challenges.
Example Interpretation:
- Barbados: May show a correlation coefficient of +0.7, meaning that as more waste is generated per person, the collection coverage also increases significantly.
- Haiti: Could have a near-zero correlation, suggesting inefficiency in scaling waste collection systems as per capita waste generation increases.

Conclusion:
Countries like Barbados and Trinidad and Tobago may have strong positive correlations, while countries like Haiti might have weak or no correlation, reflecting challenges in waste collection infrastructure.
Are there significant regional differences in Generation per capita and Collection coverage?
From the boxplot comparison by region, we can observe the following:
1. Regional variation in Generation per capita:
Caribbean: Likely to show a higher median for Generation per capita, as many Caribbean nations (e.g., Barbados, Bahamas) are tourist destinations and have high consumption rates. This leads to more waste per capita.
Latin America: Might display lower generation per capita due to different economic structures and consumption patterns, with countries like Haiti or Dominican Republic showing lower median values.
2. Regional variation in Collection coverage:
Caribbean: Regions like the Caribbean may show higher collection coverage, especially in countries with robust waste management systems like Barbados and Bahamas. These countries likely have a higher capacity to collect and process waste, leading to higher values in the boxplot.
Latin America: Might show lower collection coverage overall, with countries like Haiti having lower median values, suggesting that the waste collection infrastructure is less developed.

Example Analysis:

- Caribbean region: Countries like Barbados and Bahamas might push the boxplot's median upward for both generation per capita and collection coverage.
- Latin America: Countries like Haiti or Dominican Republic may lower the regional median, especially in collection coverage.

Conclusion:

- The Caribbean region will likely show higher generation per capita and better collection coverage compared to Latin America. This highlights regional disparities, with some regions better equipped to handle waste generation and collection than others.

Final Summary with Country Names:

1. Generation per capita (2014-2018):

   Barbados and Bahamas likely show an increasing trend in waste generation, while Saint Lucia may display a decrease. Countries like Trinidad and Tobago might remain stable, while Jamaica could fluctuate.
2. Relationship between Generation per capita and Collection coverage:

   Barbados and Trinidad and Tobago likely exhibit a positive correlation, suggesting efficient waste collection systems. Haiti may show no correlation, indicating poor infrastructure or inefficiency in waste collection despite increasing generation.
3. Regional differences:

   The Caribbean region shows higher generation per capita and better collection coverage, driven by countries like Barbados and Bahamas. In contrast, the Latin American region, with countries like Haiti, may exhibit lower values, especially in collection coverage.