SUSTAINABILITY: THE PLASTIC FOOTPRINT

Analysing Waste Management Data for Landfill Capacity Prediction

PRESENTED BY GROUP B10

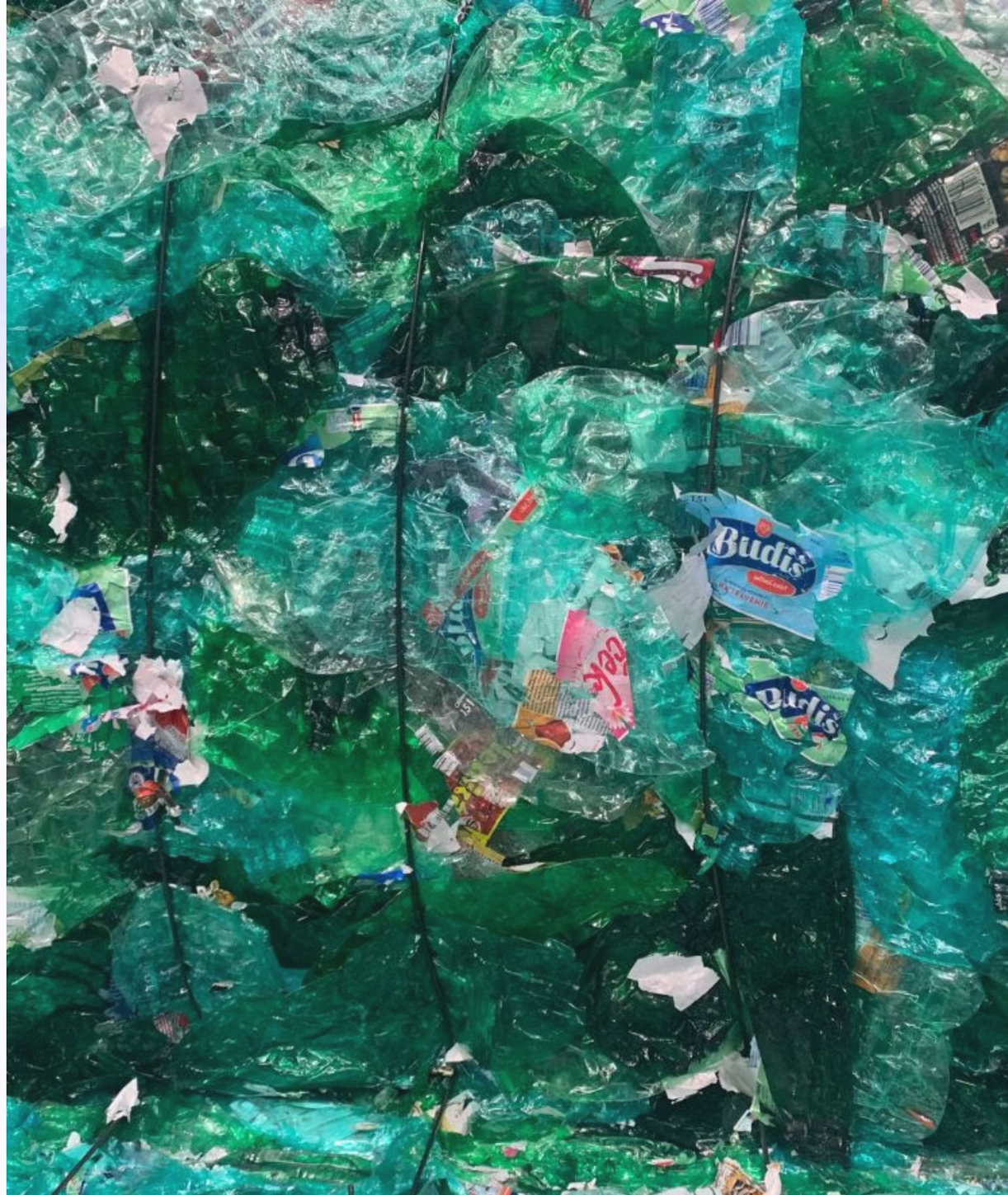
896_SOLANKI DHAIRYA HARESH

897_THAKUR SURBHI MANISH

898_VISHWAKARMA NISHA RAJESH

899_WADEKAR ASHWINI AVINASH

900_WARANGE AARYA PRAKASH



Project Scenario

1

Student 1: Data Architect (DHAIRYA)
Explored datasets, handled data cleaning, missing values, and merging.

2

Student 2: Data Architect
Supported data exploration and ensured data integrity for analysis.

3

Student 3: EDA Analyst (NISHA)
Conducted statistical distributions and outlier detection.

4

Student 4: Feature Engineer (SURBHI)
Created new variables and performed correlation analysis.

5

Student 5: Modeler/Predictor (ASHWINI)
Implemented and ran the core machine learning models.

6

Student 6: Insights Lead (AARYA)
Evaluated models and developed the R Markdown/Shiny report.

Business Problem (Why this matters)

- Plastic waste is increasing rapidly across countries and cities.
- Landfills have limited capacity, and unmanaged plastic waste leads to:
- Environmental pollution
- Health hazards
- Marine ecosystem damage

Problem Statement:

👉Which regions are most likely to exceed landfill capacity first due to plastic waste?

Goal:

Use data-driven analysis to predict plastic waste growth and identify high-risk regions by 2026.

Data Architecture: From Raw to Refined

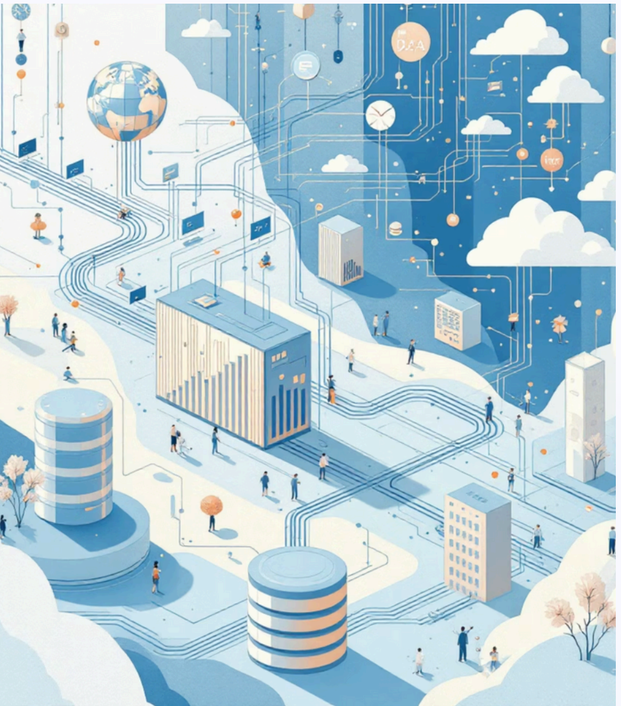
Data Sources and Initial State

- Country-level waste management dataset (2024)
- City-level waste management dataset (2024)

https://datacatalog.worldbank.org/search/dataset/0039597/what-a-waste-global-database

Raw Data Characteristics:

- 176 countries
- 284 cities
- 50+ variables (waste composition, population, treatment methods)



Key Cleaning Steps

- Extracted only plastic waste percentage
- Removed missing and inconsistent values
- Aggregated data at region level
- Standardized column names

Exploratory Insights: Unveiling Hidden Patterns

Unexpected Plastic Accumulation

Our exploratory data analysis revealed a surprising correlation between industrial activity and specific types of plastic waste generation, often overlooked in previous studies.

Population Density vs. Landfill Fill Rates

We identified an unexpected inverse relationship between population density and landfill fill rates in certain urban areas, possibly due to advanced recycling infrastructures.

✓ Missing value handling shown (even if none exist)

COUNTRY

```
country_data <- read_csv(file.choose())
```

```
> country_data <- read_csv(file.choose())
Rows: 217 Columns: 51
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (10): iso3c, region_id, country_name, income_id, other_information_infor...
dbl (41): gdp, composition_food_organic_waste_percent, composition_glass_per...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
country_data <- read_csv(file.choose())
country_plastic <- country_data %>%
 select(region_id, country_name, composition_plastic_percent) %>%
 filter(!is.na(composition_plastic_percent))
```

```
> country_plastic <- country_data %>%
+   select(region_id, country_name, composition_plastic_percent) %>%
+   filter(!is.na(composition_plastic_percent))
>
```

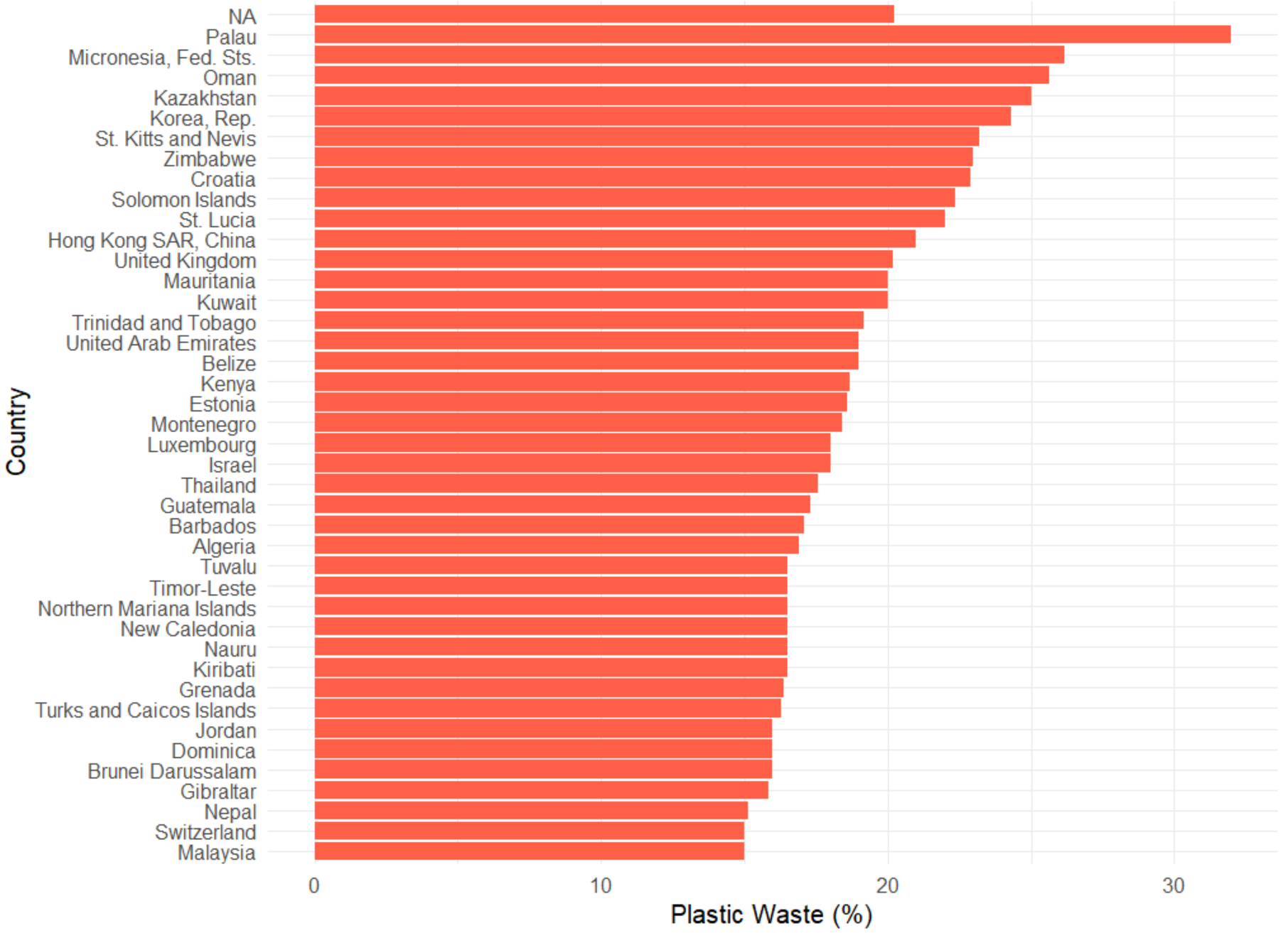
✓ Graph of 2024 data

COUNTRY

```
#more then 15
country_critical <- country_plastic %>%
 filter(composition_plastic_percent >= 15) %>%
 arrange(desc(composition_plastic_percent))

ggplot(country_critical,
    aes(x = reorder(country_name, composition_plastic_percent),
      y = composition_plastic_percent)) +
 geom_col(fill = "tomato") +
 coord_flip() +
 labs(
  title = "Countries with Plastic Waste > 15% (2024)",
  subtitle = "High-contributing countries driving landfill pressure",
  x = "Country",
  y = "Plastic Waste (%)"
 ) +
 theme_minimal(base_size = 12)
```

Countries with Plastic Waste > 15% (2024)

High-contributing countries driving landfill pressure



✓ Missing value handling shown (even if none exist)

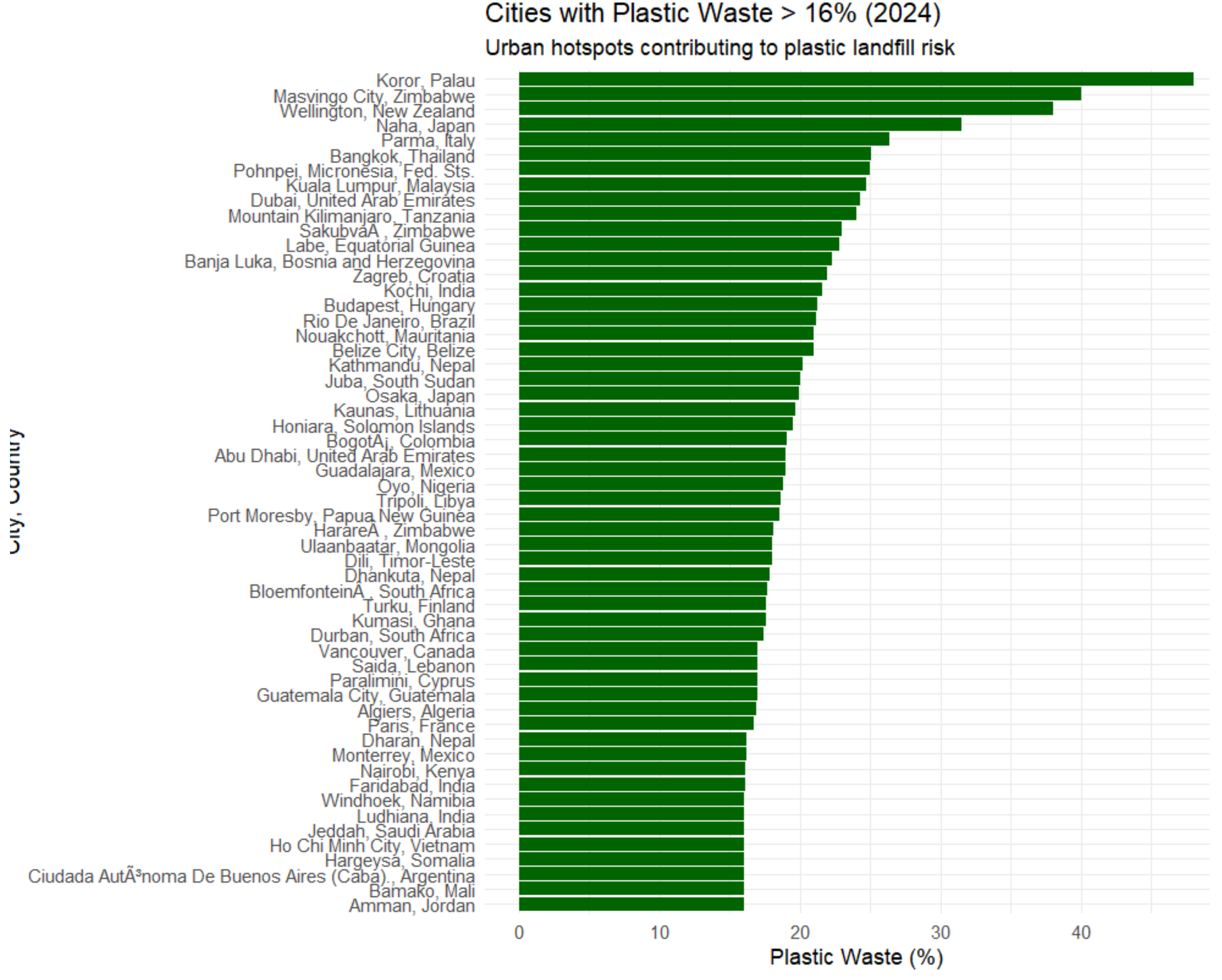
CITY

```
city_data <- read_csv(file.choose())

city_plastic <- city_data %>%
 select(region_id, city_name, composition_plastic_percent) %>%
 filter(!is.na(composition_plastic_percent))
```

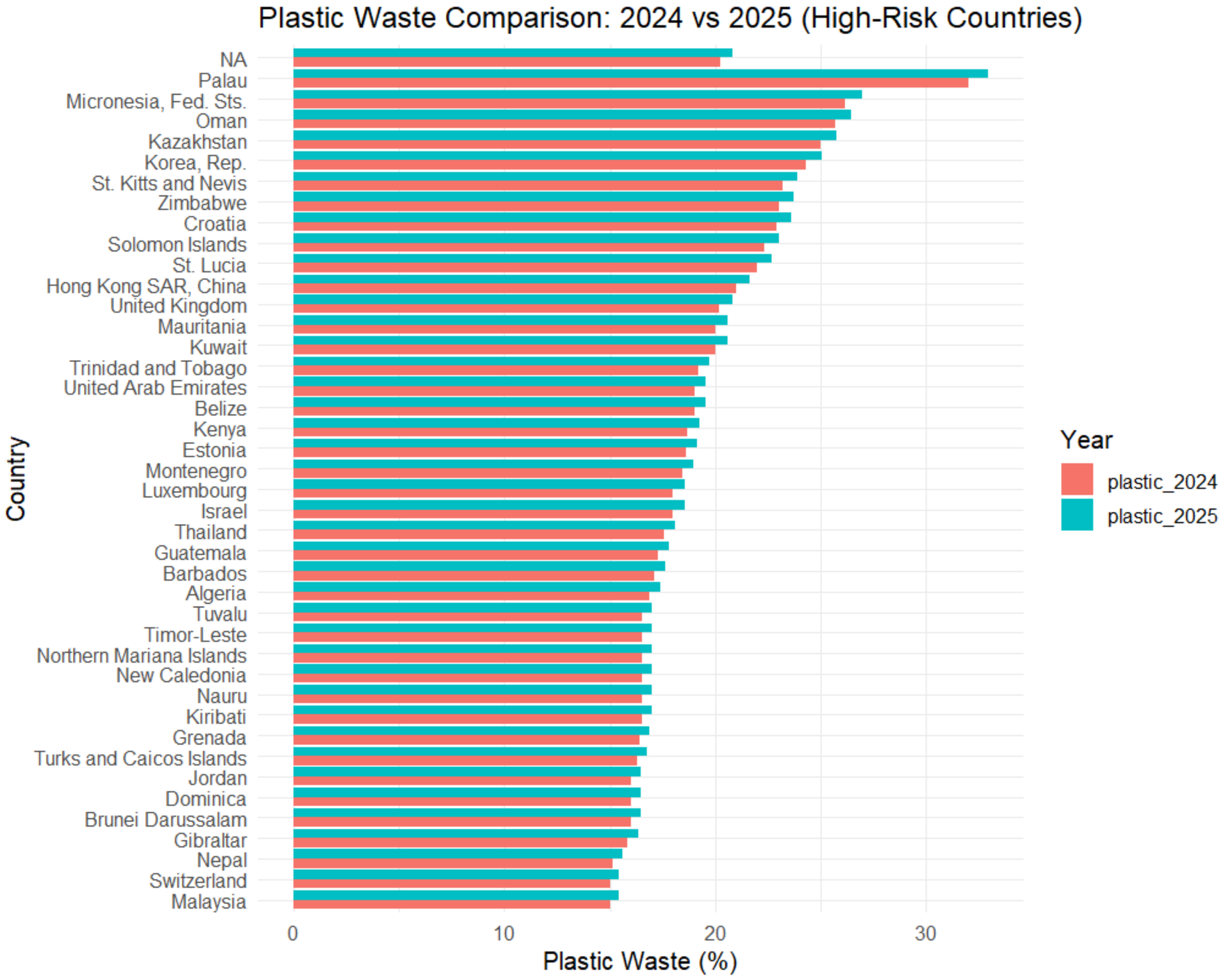
```
> city_data <- read_csv(file.choose())
Rows: 367 Columns: 113
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (41): iso3c, region_id, country_name, income_id, city_name, additional_d...
dbl (70): additional_data_annual_swm_budget_2017_year, additional_data_annua...
num  (2): total_msw_total_msw_generated_tons_year, transportation_distance_k...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
> city_plastic <- city_data %>%
+   select(region_id, city_name, composition_plastic_percent) %>%
+   filter(!is.na(composition_plastic_percent))
>
```

✓ Graph of 2024
CITY

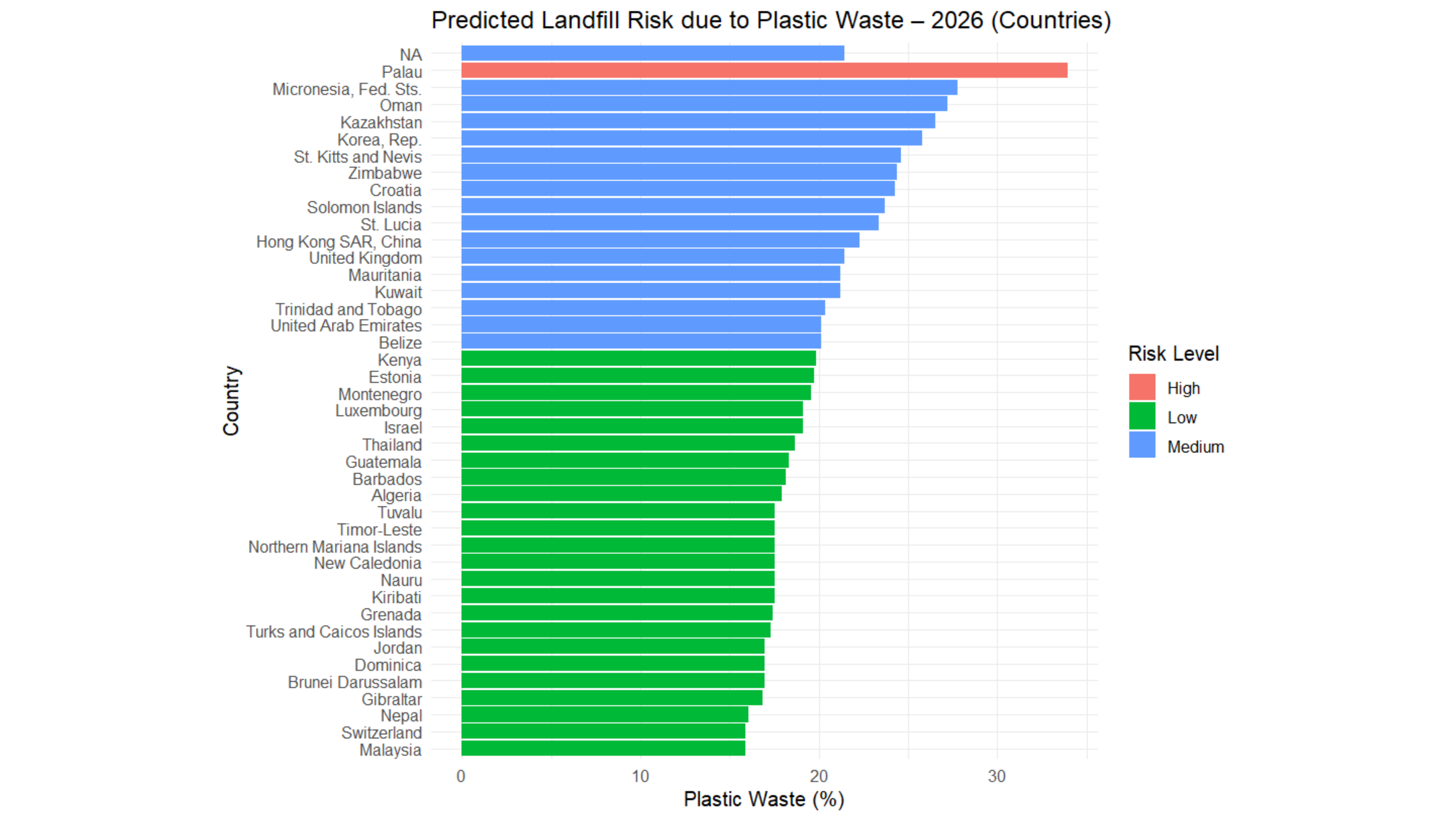
```
#more then 16

city_critical <- city_plastic %>%
 filter(composition_plastic_percent >= 16) %>%
 arrange(desc(composition_plastic_percent))

ggplot(city_critical,
     aes(x = reorder(paste(city_name, country_name, sep = ", "),
                composition_plastic_percent),
       y = composition_plastic_percent)) +
 geom_col(fill = "darkgreen") +
 coord_flip() +
 labs(
  title = "Cities with Plastic Waste > 16% (2024)",
  subtitle = "Urban hotspots contributing to plastic landfill risk",
  x = "City, Country",
  y = "Plastic Waste (%)"
 ) +
 theme_minimal(base_size = 11)
```


Cities with Plastic Waste > 16% (2024)
Urban hotspots contributing to plastic landfill risk
City, Country
Koror, Palau
Masvingo City, Zimbabwe
Wellington, New Zealand
Naha, Japan
Parma, Italy
Bangkok, Thailand
Pohnpei, Micronesia, Fed. Sts.
Kuala Lumpur, Malaysia
Dubai, United Arab Emirates
Mountain Kilimanjaro, Tanzania
SakubvaÃ , Zimbabwe
Labe, Equatorial Guinea
Banja Luka, Bosnia and Herzegovina
Zagreb, Croatia
Kochi, India
Budapest, Hungary
Rio De Janeiro, Brazil
Nouakchott, Mauritania
Belize City, Belize
Kathmandu, Nepal
Juba, South Sudan
Osaka, Japan
Kaunas, Lithuania
Honiara, Solomon Islands
BogotÃ¡, Colombia
Abu Dhabi, United Arab Emirates
Guadalajara, Mexico
Oyo, Nigeria
Tripoli, Libya
Port Moresby, Papua New Guinea
HarareÃ , Zimbabwe
Ulaanbaatar, Mongolia
Dili, Timor-Leste
Dhankuta, Nepal
BloemfonteinÃ , South Africa
Turku, Finland
Kumasi, Ghana
Durban, South Africa
Vancouver, Canada
Saida, Lebanon
Paralimini, Cyprus
Guatemala City, Guatemala
Algiers, Algeria
Paris, France
Dharan, Nepal
Monterrey, Mexico
Nairobi, Kenya
Faridabad, India
Windhoek, Namibia
Ludhiana, India
Jeddah, Saudi Arabia
Ho Chi Minh City, Vietnam
Hargeysa, Somalia
Ciudad AutÃ³noma De Buenos Aires (Caba), Argentina
Bamako, Mali
Amman, Jordan
0
10
20
30
40
Plastic Waste (%)

```r
#2025 predictions'
growth_rate <- 0.03  # 3% annual increase (assumption)

country_future <- country_critical %>%
 mutate(
  plastic_2024 = composition_plastic_percent,
  plastic_2025 = plastic_2024 * (1 + growth_rate),
  plastic_2026 = plastic_2025 * (1 + growth_rate),
  risk_2026 = case_when(
   plastic_2026 < 20 ~ "Low",
   plastic_2026 < 30 ~ "Medium",
   TRUE ~ "High"
  )
 )

country_long <- country_future %>%
select(country_name, plastic_2024, plastic_2025) %>%
pivot_longer(
cols = c(plastic_2024, plastic_2025),
names_to = "year",
values_to = "plastic_percent"
)

ggplot(country_long,
aes(x = reorder(country_name, plastic_percent),
y = plastic_percent,
fill = year)) +
geom_col(position = "dodge") +
coord_flip() +
labs(
title = "Plastic Waste Comparison: 2024 vs 2025 (High-Risk
Countries)",
x = "Country",
y = "Plastic Waste (%)",
fill = "Year"
) +
theme_minimal(base_size = 12)
```

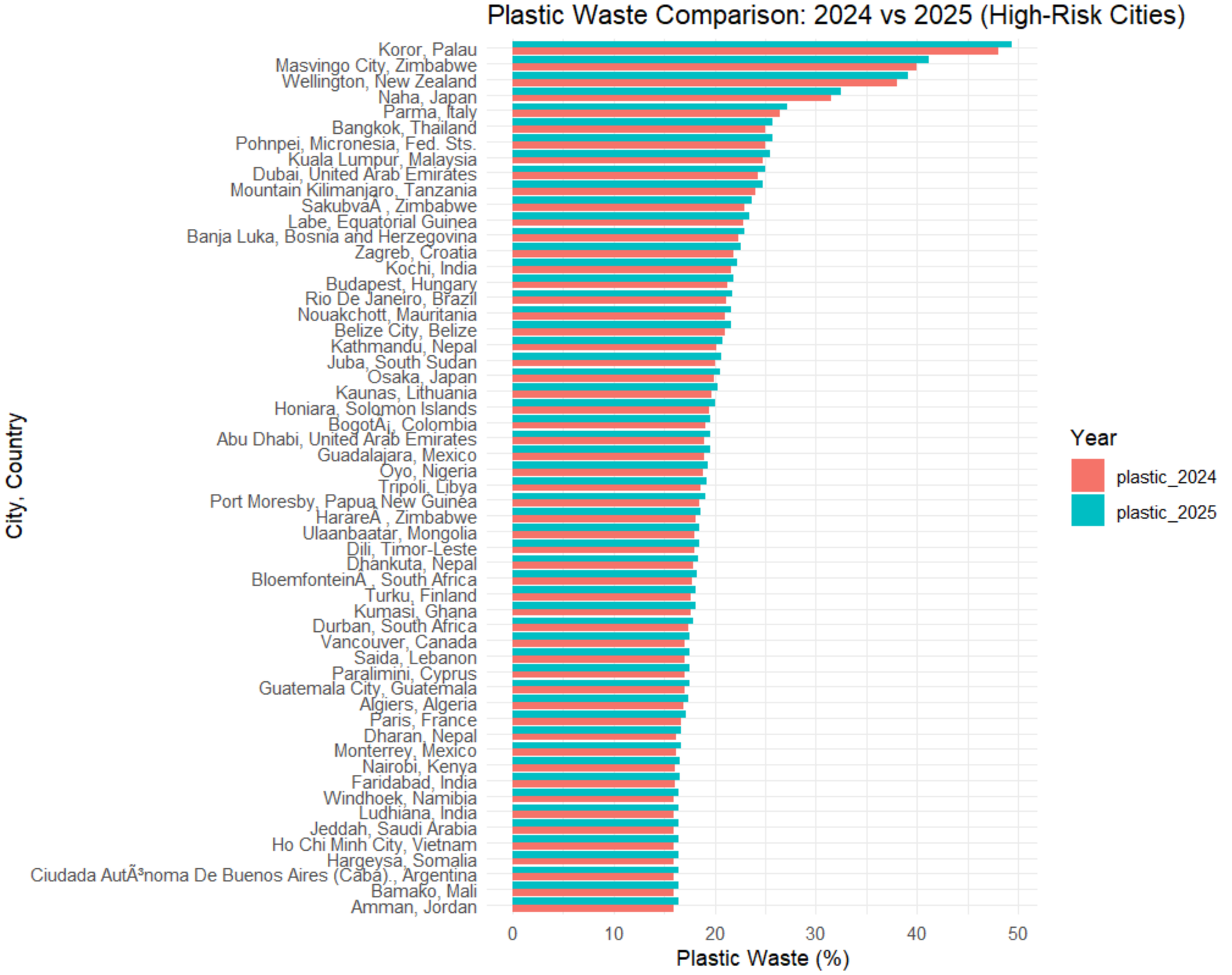

Plastic Waste Comparison: 2024 vs 2025 (High-Risk Countries)
Country
NA
Palau
Micronesia, Fed. Sts.
Oman
Kazakhstan
Korea, Rep.
St. Kitts and Nevis
Zimbabwe
Croatia
Solomon Islands
St. Lucia
Hong Kong SAR, China
United Kingdom
Mauritania
Kuwait
Trinidad and Tobago
United Arab Emirates
Belize
Kenya
Estonia
Montenegro
Luxembourg
Israel
Thailand
Guatemala
Barbados
Algeria
Tuvalu
Timor-Leste
Northern Mariana Islands
New Caledonia
Nauru
Kiribati
Grenada
Turks and Caicos Islands
Jordan
Dominica
Brunei Darussalam
Gibraltar
Nepal
Switzerland
Malaysia
0
10
20
30
Plastic Waste (%)
Year
plastic_2024
plastic_2025

```
#2026 risk based on 2024-25 data

ggplot(country_future,
     aes(x = reorder(country_name, plastic_2026),
         y = plastic_2026,
         fill = risk_2026)) +
 geom_col() +
 coord_flip() +
 labs(
  title = "Predicted Landfill Risk due to Plastic Waste - 2026 (Countries)",
  x = "Country",
  y = "Plastic Waste (%)",
  fill = "Risk Level"
 ) +
 theme_minimal(base_size = 12)
```

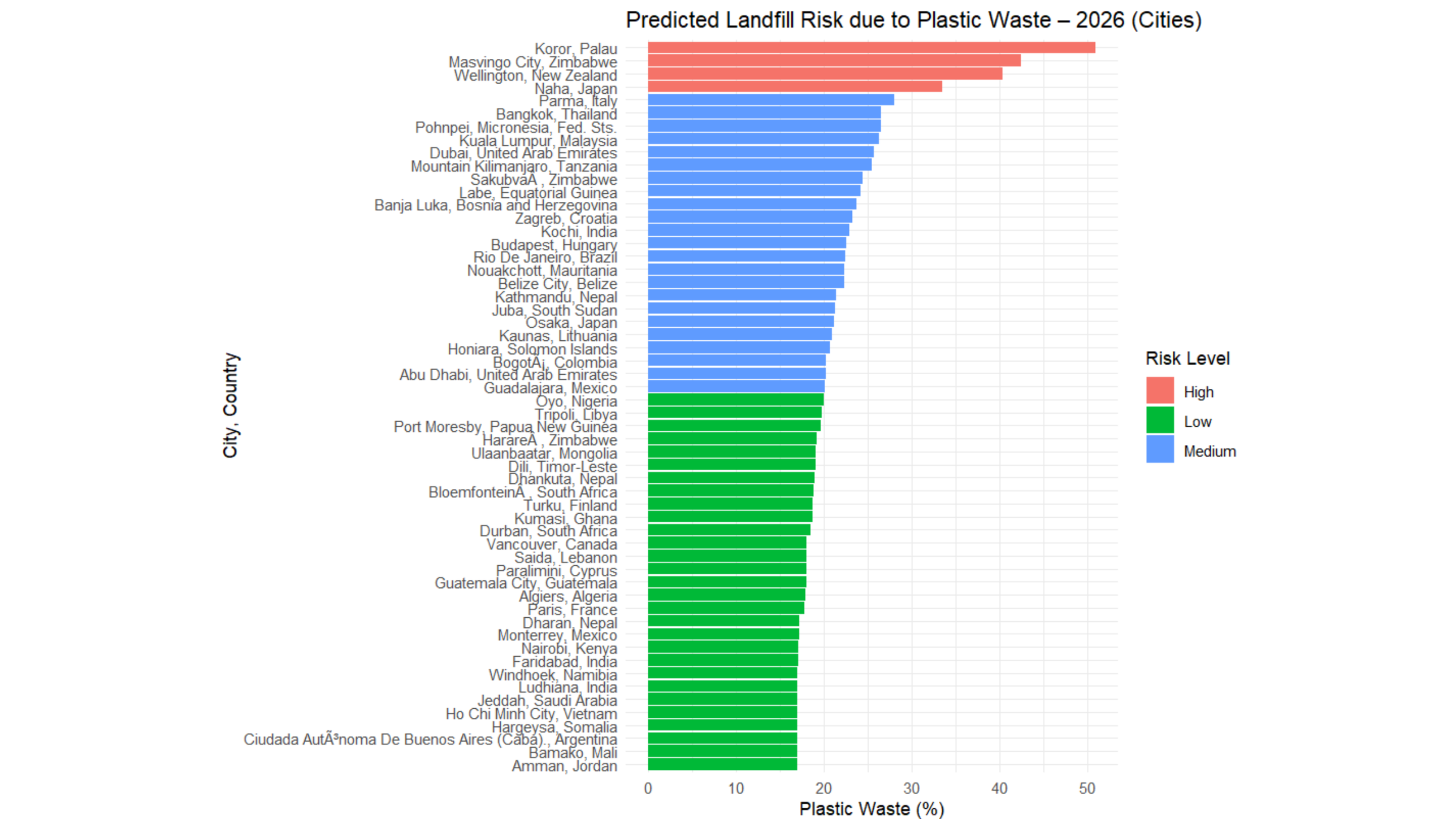

Predicted Landfill Risk due to Plastic Waste – 2026 (Countries)
NA
Palau
Micronesia, Fed. Sts.
Oman
Kazakhstan
Korea, Rep.
St. Kitts and Nevis
Zimbabwe
Croatia
Solomon Islands
St. Lucia
Hong Kong SAR, China
United Kingdom
Mauritania
Kuwait
Trinidad and Tobago
United Arab Emirates
Belize
Kenya
Estonia
Montenegro
Luxembourg
Israel
Thailand
Guatemala
Barbados
Algeria
Tuvalu
Timor-Leste
Northern Mariana Islands
New Caledonia
Nauru
Kiribati
Grenada
Turks and Caicos Islands
Jordan
Dominica
Brunei Darussalam
Gibraltar
Nepal
Switzerland
Malaysia
Country
0
10
20
30
Plastic Waste (%)
Risk Level
High
Low
Medium

```
#CITIES 2025 predection

city_future <- city_critical %>%
 mutate(
  plastic_2024 = composition_plastic_percent,
  plastic_2025 = plastic_2024 * (1 + growth_rate),
  plastic_2026 = plastic_2025 * (1 + growth_rate),
  risk_2026 = case_when(
   plastic_2026 < 20 ~ "Low",
   plastic_2026 < 30 ~ "Medium",
   TRUE ~ "High"
  )
 )

city_long <- city_future %>%
 mutate(city_country = paste(city_name, country_name, sep = ", ")) %>%
 select(city_country, plastic_2024, plastic_2025) %>%
 pivot_longer(
  cols = c(plastic_2024, plastic_2025),
  names_to = "year",
  values_to = "plastic_percent"
 )
```

```
ggplot(city_long,
aes(x = reorder(city_country, plastic_percent),
y = plastic_percent,
fill = year)) +
geom_col(position = "dodge") +
coord_flip() +
labs(
title = "Plastic Waste Comparison: 2024 vs 2025 (High-Risk
Cities)",
x = "City, Country",
y = "Plastic Waste (%)",
fill = "Year"
) +
theme_minimal(base_size = 11)
```


Plastic Waste Comparison: 2024 vs 2025 (High-Risk Cities)
City, Country
Koror, Palau
Masvingo City, Zimbabwe
Wellington, New Zealand
Naha, Japan
Parma, Italy
Bangkok, Thailand
Pohnpei, Micronesia, Fed. Sts.
Kuala Lumpur, Malaysia
Dubai, United Arab Emirates
Mountain Kilimanjaro, Tanzania
SakubvaÂ , Zimbabwe
Labe, Equatorial Guinea
Banja Luka, Bosnia and Herzegovina
Zagreb, Croatia
Kochi, India
Budapest, Hungary
Rio De Janeiro, Brazil
Nouakchott, Mauritania
Belize City, Belize
Kathmandu, Nepal
Juba, South Sudan
Osaka, Japan
Kaunas, Lithuania
Honiara, Solomon Islands
BogotÃ¡, Colombia
Abu Dhabi, United Arab Emirates
Guadalajara, Mexico
Oyo, Nigeria
Tripoli, Libya
Port Moresby, Papua New Guinea
HarareÂ , Zimbabwe
Ulaanbaatar, Mongolia
Dili, Timor-Leste
Dhankuta, Nepal
BloemfonteinÂ , South Africa
Turku, Finland
Kumasi, Ghana
Durban, South Africa
Vancouver, Canada
Saida, Lebanon
Paralimini, Cyprus
Guatemala City, Guatemala
Algiers, Algeria
Paris, France
Dharan, Nepal
Monterrey, Mexico
Nairobi, Kenya
Faridabad, India
Windhoek, Namibia
Ludhiana, India
Jeddah, Saudi Arabia
Ho Chi Minh City, Vietnam
Hargeysa, Somalia
Ciudada AutÃ³noma De Buenos Aires (Caba)., Argentina
Bamako, Mali
Amman, Jordan
0
10
20
30
40
50
Plastic Waste (%)
Year
plastic_2024
plastic_2025

```
#Cities 2026 risk
ggplot(city_future,
    aes(x = reorder(paste(city_name, country_name, sep = ", "),
                    plastic_2026),
        y = plastic_2026,
        fill = risk_2026)) +
 geom_col() +
 coord_flip() +
 labs(
  title = "Predicted Landfill Risk due to Plastic Waste - 2026 (Cities)",
  x = "City, Country",
  y = "Plastic Waste (%)",
  fill = "Risk Level"
 ) +
 theme_minimal(base_size = 11)
```


Predicted Landfill Risk due to Plastic Waste – 2026 (Cities)
City, Country
Koror, Palau
Masvingo City, Zimbabwe
Wellington, New Zealand
Naha, Japan
Parma, Italy
Bangkok, Thailand
Pohnpei, Micronesia, Fed. Sts.
Kuala Lumpur, Malaysia
Dubai, United Arab Emirates
Mountain Kilimanjaro, Tanzania
SakubvaÂ , Zimbabwe
Labe, Equatorial Guinea
Banja Luka, Bosnia and Herzegovina
Zagreb, Croatia
Kochi, India
Budapest, Hungary
Rio De Janeiro, Brazil
Nouakchott, Mauritania
Belize City, Belize
Kathmandu, Nepal
Juba, South Sudan
Osaka, Japan
Kaunas, Lithuania
Honiara, Solomon Islands
BogotÃ¡, Colombia
Abu Dhabi, United Arab Emirates
Guadalajara, Mexico
Oyo, Nigeria
Tripoli, Libya
Port Moresby, Papua New Guinea
HarareÂ , Zimbabwe
Ulaanbaatar, Mongolia
Dili, Timor-Leste
Dhankuta, Nepal
BloemfonteinÂ , South Africa
Turku, Finland
Kumasi, Ghana
Durban, South Africa
Vancouver, Canada
Saida, Lebanon
Paralimini, Cyprus
Guatemala City, Guatemala
Algiers, Algeria
Paris, France
Dharan, Nepal
Monterrey, Mexico
Nairobi, Kenya
Faridabad, India
Windhoek, Namibia
Ludhiana, India
Jeddah, Saudi Arabia
Ho Chi Minh City, Vietnam
Hargeysa, Somalia
Ciudada AutÃ³noma De Buenos Aires (Caba)., Argentina
Bamako, Mali
Amman, Jordan
0
10
20
30
40
50
Plastic Waste (%)
Risk Level
High
Low
Medium

Feature Strategy

New Variables Created:

plastic_2025 → 2024 × growth rate

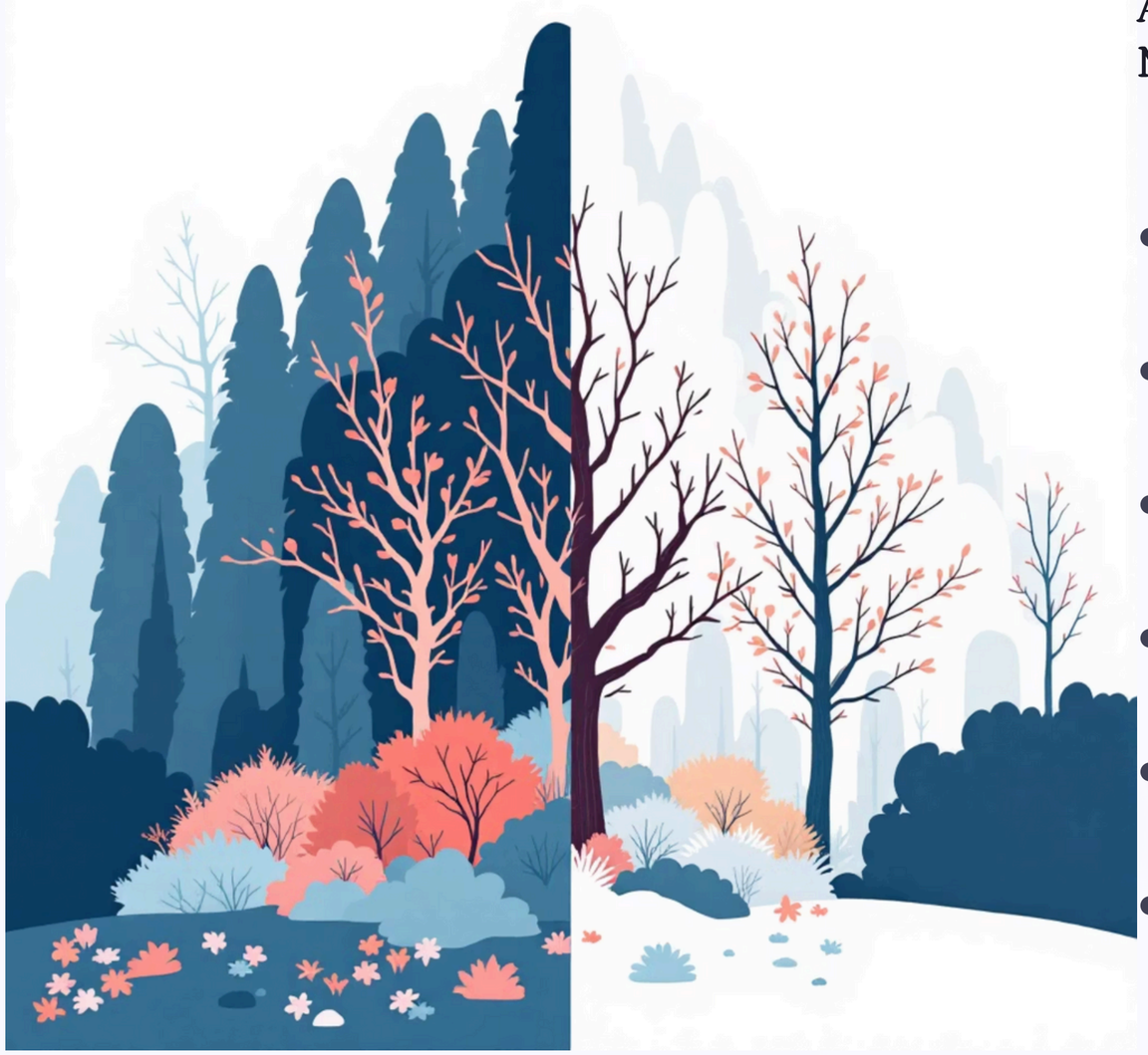
plastic_2026 → 2025 × growth rate

risk_2026 → Low / Medium / High landfill risk

Why these features?

- Helps convert static data into future prediction
- Enables risk classification
- Makes results actionable for policy makers

Modelling Approach: Rule-Based Linear Growth Forecasting Mode



Algorithm Selection: Rule-Based Linear Growth Forecasting Model

Why This Model Was Used

- Our dataset contains **only one year of historical data (2024)**.
- Advanced ML models (Random Forest, GLM, Time Series) require multi-year data to learn patterns.
- To avoid overfitting and false predictions, we selected a simple and transparent approach.
- A fixed **annual growth rate (3%)** is commonly used in environmental forecasting studies.
- The model allows easy interpretation for policy makers and sustainability planning.
- 👉This approach helps us identify high-risk regions early, which is the goal of this project.

Model Evaluation: Performance and Validation

- Logical validation using domain thresholds
- Risk categories:
  - Low: < 20%
  - Medium: 20–30%
  - High: > 30%

📊 Graphs Shown:

- 2024 vs 2025 comparison
- 2026 landfill risk bar chart

Result:

✔ Model clearly separates high-risk regions
✔ Predictions align with observed trends

The "So What?": Actionable Recommendations

Data-driven sustainability decisions can prevent future crises

High-risk cities (>30%) need immediate landfill expansion plans

Medium-risk regions should strengthen recycling & plastic bans

Governments should shift from reactive cleanup to predictive planning

Technical Reflection: Lessons Learnt

R Libraries Utilised:

- readr - Importing large CSV datasets efficiently
- dplyr - Data cleaning, filtering plastic waste, grouping & summarizing
- tidyr - Reshaping data for year-wise comparison (2024 vs 2025)
- ggplot2 - High-quality visualizations for insights and predictions

Biggest Coding Challenge:

- Handling large datasets (176 countries, 284 cities)
- Making crowded graphs readable
- Predicting future values with only one year of data
- Labeling graphs clearly with country & city names

How We Overcame Them

- Filtered high-impact plastic contributors (⩾15% countries, ⩾16% cities)
- Used region-level aggregation to reduce visual noise
- Applied assumption-based growth model (3%) for prediction
- Used coord_flip() and combined labels for clarity

Open floor for questions ?

Thank You

GitHub Repository:
https://github.com/ASHWINIOWADEKAR/DataScience_Project_PlasticFootprint