Understanding the Associations Between Mismanaged Plastic Waste and Purchasing Power Parity Per Capita GDP and Coastal Populations

Nagaprasad Rudrapatna

December 16, 2022

1 Introduction

1.1 Background Information

Plastic pollution is a severe and growing issue, adversely affecting ecosystems and wildlife across the globe. Plastic debris can negatively affect wildlife through three main pathways–entanglement, ingestion, and interaction (Law, 2017). Entanglement refers to the "entrapping, encircling, or constricting of marine animals by plastic debris" (e.g., plastic rope and netting, abandoned fishing gear) and cases have been reported for "all marine turtle species, more than two-thirds of seal species, one-third of whale species, and one-quarter of seabirds" (as well as numerous species of fish and invertebrates; Kühn, Bravo Rebolledo, & van Franeker, 2015). Ingestion of plastic can occur "unintentionally, intentionally, or indirectly through the ingestion of prey species containing plastic" and cases have been reported for "all marine turtle species, more than one-third of seal species, 59% of whale species, and 59% of seabirds" (as well as numerous species of fish and invertebrates; Ritchie & Roser, 2018; Kühn, Bravo Rebolledo, & van Franeker, 2015). Ingesting large volumes of plastic, for example, can "greatly reduce stomach capacity, leading to poor appetite and false sense of satiation"; plastic can also "obstruct or perforate the gut, cause ulcerative lesions, or gastric rupture" in these organisms (Ritchie & Roser, 2018). Plastic can also interact with ecosystem structures, impacting "light penetration, organic matter availability, and oxygen exchange" (Ritchie & Roser, 2018).

According to Geyer and collaborators (2017), the world produced only two million tonnes of plastic in 1950. Since then, annual global plastic production has increased roughly 150-fold, reaching 313 million tonnes in 2010 (this upward trend has continued since 2010: approximately 460 million tonnes in 2019; Ritchie & Roser, 2018). Naturally, as plastic production has amplified across several decades, the amount of generated plastic waste has also increased. Increased plastic waste, by itself, is not the primary issue. The main problem is rooted in the prevalence of poor waste management infrastructure in low-to-middle income countries (Ritchie & Roser, 2018). This has contributed to an increase in the amount of mismanaged plastic waste–plastic that is either inadequately disposed (stored in open or insecure landfills; high risk of polluting rivers and the ocean) or littered by coastal populations (defined as populations living within 50 kilometers of a coastline)–generated globally (Ritchie & Roser, 2018). Plastic pollution in the ocean originates from both land-based and marine ("pollution caused by fishing fleets that leave behind fishing nets, lines, ropes, and sometimes abandoned vessels") sources; the latest estimates indicate that roughly 80% of ocean plastics originate from land-based sources while the remaining 20% stems from marine sources (Li, Tse, & Fok, 2016). The activities of coastal populations are critical since the plastic waste generated by these groups is at a higher risk of leading to ocean debris (compared to sources further inland), but mismanaged waste can "eventually enter the ocean via inland waterways, wastewater outflows, and transport by wind or tides" (Ritchie & Roser, 2018). Previous research has shown that not all mismanaged plastic waste has the same probability of reaching the ocean. These studies have identified an area's climate, terrain, land use, and proximity to river basins as key factors in determining the probability that mismanaged waste is emitted to the ocean (Ritchie & Roser, 2018).

It is important to emphasize that plastic waste can only enter rivers and the ocean if it is improperly managed. In fact, the vast majority of plastic waste ends up in landfills; approximately 3% of global plastic waste enters the ocean (Jambeck et al., 2015). Almost all plastic waste is "incinerated, recycled, or sent it to well-managed landfills" in wealthy countries, but "waste can be dumped outside of landfills (and the landfills that do exist are often open) in poorer countries, leaking waste to the surrounding environment" (Ritchie & Roser, 2018). Prior studies have suggested an association between the wealth of a country and its waste management infrastructure. A statistically rigorous analysis of this relationship, namely how a country's wealth is associated with its mismanaged plastic waste (particularly, the proportion of plastic waste that is mismanaged), could be useful for stakeholders (perhaps supranational organizations, such as the European Union or World Bank, who are interested in mitigating the effects of plastic pollution).

1.2 Research Aims

A common measure of the wealth of an entity (e.g., independent country, larger political unit, overseas territory) is its gross domestic product (GDP), which is a comprehensive measure of the entity's economy and growth (in this analysis, GDP is calculated without making deductions for the depreciation of fabricated assets or for the depletion and degradation of natural resources). GDP per capita is GDP converted to international dollars (an international dollar has the same purchasing power over GDP as a US dollar has in the United States) using purchasing power parity (PPP) rates and divided by an entity's total population. In this analysis, PPP GDP per capita is considered in constant 2011 international dollars (Mock, 2019). This project, first and foremost, aims to address the following research question: how is an entity's PPP GDP per capita associated with its proportion of mismanaged plastic waste (out of its total plastic waste)? This project also seeks to identify the association (if any) between the relative size of an entity's coastal population (the ratio of the coastal population to the total population) and its proportion of mismanaged plastic waste.

1.3 Data Description and Cleaning

This report relies on the information derived from three datasets. The main datasets provided 2010 daily plastic waste generation (prior to waste management, recycling, or incineration) and daily mismanaged plastic waste generation per capita rates (both measured in kilograms per person per day) for 186 entities, PPP GDP per capita (in constant 2011 international dollars) rates for 236 entities (with observations corresponding to the previous 186), and total population data (reported by the independent Swedish foundation, Gapminder) for 223 entities (with observations corresponding to the previous 186). The PPP GDP per capita rates and population data stretched over 300 years for some entities (the data points for each entity start at different times and the entries are not uniformly spaced; the longest period is from 1700 to 2017). An auxiliary dataset provided 2010 coastal population data for the 186 entities.

In this analysis, the response variable is the proportion of plastic waste that was mismanaged by an entity in the year 2010. It is calculated as the ratio of the daily mismanaged plastic waste generation per capita rate to the daily (total) plastic waste generation per capita rate. The covariates of interest are: an entity's 2010 PPP GDP per capita rate (measured in thousands of constant 2011 international dollars per person) and the relative size of an entity's coastal population (calculated as the ratio of the coastal population to the total population; referred to as an entity's coastal population proportion). These two explanatory variables were selected after careful review of the existing literature on global plastic pollution. Specifically, previous research suggests an association between an entity's wealth and its plastic waste management infrastructure. Furthermore, studies claim that plastic waste generated by coastal populations is at a higher risk of entering the ocean (being mismanaged).

1.4 Hypotheses

I hypothesize that, as PPP GDP per capita increases (i.e., for richer entities), the log-odds of an entity mismanaging its daily per capita plastic waste (compared to properly managing its daily per capita plastic waste) is expected to decrease (after controlling for the entity's coastal population). Additionally, I hypothesize that, as the relative size of an entity's coastal population (i.e., proportion of total population living within 50 kilometers of a coastline) increases, the log-odds of an entity mismanaging its daily per capita plastic waste (compared to properly managing its daily per capita plastic waste) is expected to increase (after controlling for the entity's PPP GDP per capita).

1.5 Exploratory Data Analysis

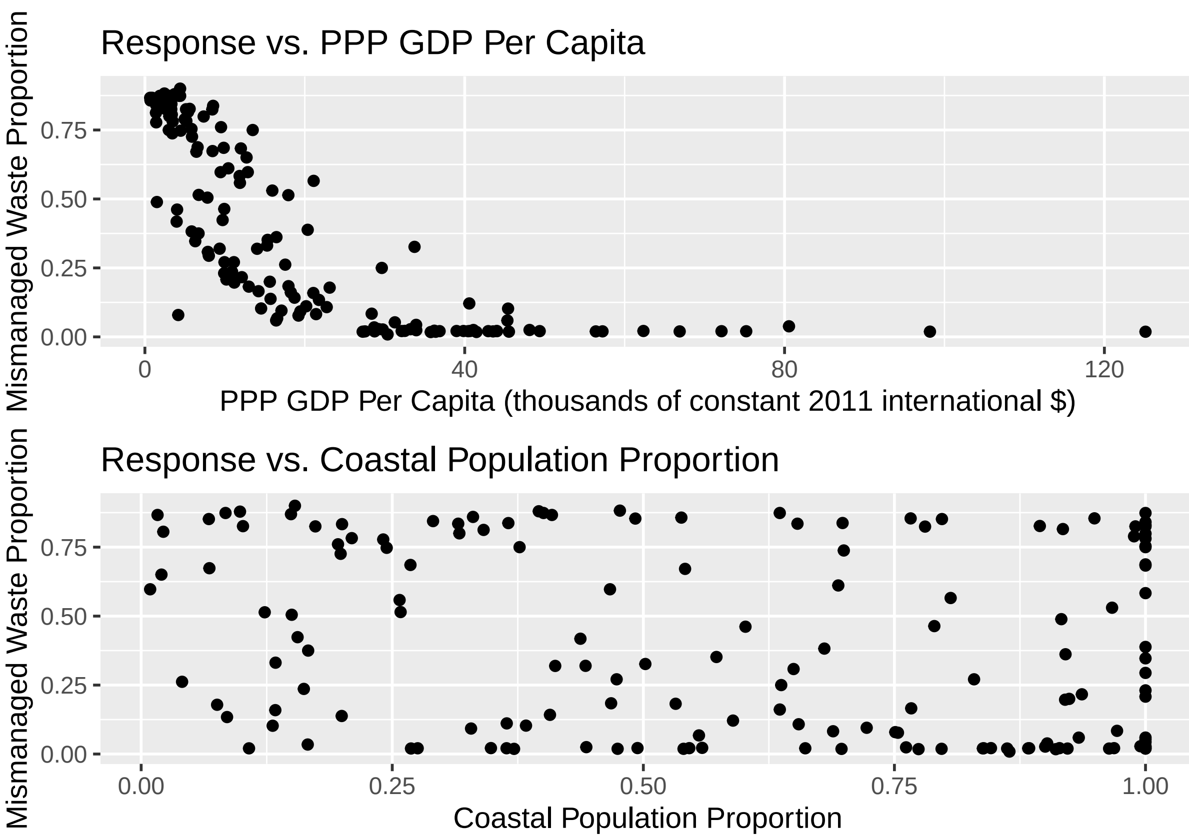

Figure 1: Top: The Relationship Between PPP GDP Per Capita and Mismanaged Plastic Waste Proportion; Bottom: The Relationship Between Coastal Population Proportion and Mismanaged Plastic Waste Proportion

2 log-transform per capita GDP -> varying orders of magnitude

3 mixture model - discrete mass near zero and then continuous mass elsewhere - would be more -> try zero-inflated beta regression?

strong regional dominance in global mismanaged coastal waste in Asia (especially India)

4 Methodology

4.1 Data Cleaning

It is important to reiterate that the main datasets included complete information regarding plastic waste generation for these 186 entities. The raw datasets actually included several additional observations with this information missing. For the most part, the entities which did not have this information correspond to landlocked areas. This is reasonable since plastic waste is only classified as having been mismanaged if it enters a body of water. I therefore decided to exclude all entities without 2010 daily plastic waste generation and daily mismanaged plastic waste generation per capita rates (leaving the 186 entities). The primary motivating factor was that the response variable of interest, the proportion of plastic waste that was mismanaged in 2010, depends on the entity's waste generation per capita rates. Moreover, as coastal population data and daily (mismanaged) plastic waste generation per capita rates were only available in 2010, I decided to limit the analysis to this year. This means that, despite the availability of PPP GDP per capita rates in other years, I only consider 2010 PPP GDP per capita rates for each entity. Unfortunately, 41 entities were missing either PPP GDP per capita rates or total population data. I found 2010 PPP GDP per capita rates (in constant 2011 international dollars) for five entities–Aruba, Cayman Islands, Curacao, Djibouti, Sint Maarten (Dutch part)–on the UNData record viewer (United Nations). 33 entities (including Anguilla and British Virgin Islands) did not have readily available 2010 PPP GDP per capita rates, so I decided to remove these observations from the dataset. Of the remaining 153 entities, five entities–Curacao, Micronesia, Palestine, Sint Maarten, and Yemen–were missing total population data from 2010. I located this information online (from databases, such as Macrotrends and Population Pyramid) and imputed the values. Thus, the final dataset includes complete information about 153 entities. I decided to perform a complete case analysis since I could not find the missing information (waste generation rates, PPP GDP per capita rates) for certain entities online. This analysis therefore assumes that information is missing completely at random. It is worth mentioning that, if the data were in fact not missing completely at random, any models fit using this dataset could have biased coefficient estimates and standard errors.

We made this decision to simplify the analysis and remove the hierarchical structure of the data (i.e., patients with multiple procedures). By only considering each patient's first procedure, we can treat the observations independently (i.e., independence assumption satisfied).

Modification for coastal population proportion > 1: capped at 1 # when coastal population > total population, we consider the entire population to be coastal (islands) -> every point in a country might be less than 50 km from a coastline

Questions - model -> EDA plots suggest log-transformation of response but then beta regression no longer suitable - so, should I do beta regression or linear regression on the log-transformed proportions? - what about coastal population proportion: seemingly no relationship -> can I still consider it as a covariate of interest? - missing data approach: missing completely at random, complete case analysis - need info about response to continue

conduct a complete case analysis (first exclude non-countries, then exclude countries without full information - Kosovo, Timor (island; East Timor - sovereign, West Timor part of Indonesia) excluded for lack of population data) - response variable is created from mismanaged and total plastic waste - per capita GDP (based on 2011 international rates) and proportion of country's coastal population to total population (reported by Gapminder) are covariates of interest

-> only consider 2010 since that's when plastic waste data is available -> create response variable: Per Capita Mismanaged Waste (Per Day) / Per Capita Waste (Per Day) -> create second covariate: Coastal Population / Total Population

-> tried to manually adjust total population for countries whose coastal population > total population, but found that the issue persisted for some countries - issues with data collection of coastal population (how is it collected?)

response variable: between [0, 1] beta regression! - not a GLM

- logit link
- Y ~ Beta(Alpha, Beta)
- logit(E(Y|X)) = X' beta
- log(Alpha / Beta) = X' beta
- Y = ratio of mismanaged to total plastic waste
- no error term
- interpretations: linear predictor of logit of conditional expectation

Holding all ..., for a $1000 increase in per capita GDP, we expect the log-odds of a country mismanaging plastic waste (compared to properly managing plastic waste) to increase by beta_1.

zero-inflated model: interpreting only (0,1) component -> interpretation is the same -> more accurate coefficient estimates that account for the bounds of the proportions (particularly, the lower bound of 0)

- why use logit link? - interpretability
- interactions -> no because interpreting quantitative-quantitative interactions is difficult, not aligned with objective
- assumptions for beta regression? - independence (only 2010 -> satisfied), linearity of predictors

Using mathematical notation, we can express our final model as

$$\log(T_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \beta_8 x_{8i} + \ldots + \beta_{30} x_{30i} + \epsilon_i$$

where i is the index for the patient (ranging from 1 to 21,180), T_i is the predicted survival time of ith patient with a history of smoking. In terms of the covariates, x_{1i} through x_{3i} indicate the ith patient's body mass index category, x_{4i} through x_{7i} indicate the ith patient's congestive heart failure severity, and x_{8i} through x_{30i} are the controlled patient demographic and medical factors.

5 Results

The regression coefficients were estimated using the `gamlss()` function in R. Table 1 displays the coefficient estimates, as well as the corresponding 95% confidence intervals and p-values for the final model.

Table 1: Final Model Coefficients and 95% Confidence Intervals

	Coefficient	Lower Bound	Upper Bound	P-Value
Intercept	1.112	0.819	1.405	<0.01
PPP GDP Per Capita (thousands of international $)	-0.083	-0.097	-0.070	<0.01
Coastal Population Proportion	-0.364	-0.750	0.023	0.0675

Holding the proportion of a country's total population that lives within 50 kilometers of a coastline constant:

6 Discussion

6.1 Conclusions

6.2 Limitations and Future Directions

Additionally, as mentioned previously, assuming that the data is missing completely at random may be inappropriate, and MICE (multiple imputation via chained equations) or other missing data approaches (which presuppose a different missingness mechanism) could be more suitable than the complete case analysis described here.

6.3 Summary

7 Appendices

7.1 Appendix A

Table 2: Entities with Highest Mismanaged Plastic Waste Proportion

Entity	Mismanaged Waste Proportion
India	0.9000000
Bangladesh	0.8823529
Myanmar	0.8800000
Cambodia	0.8787879
Pakistan	0.8737864

7.2 Appendix B

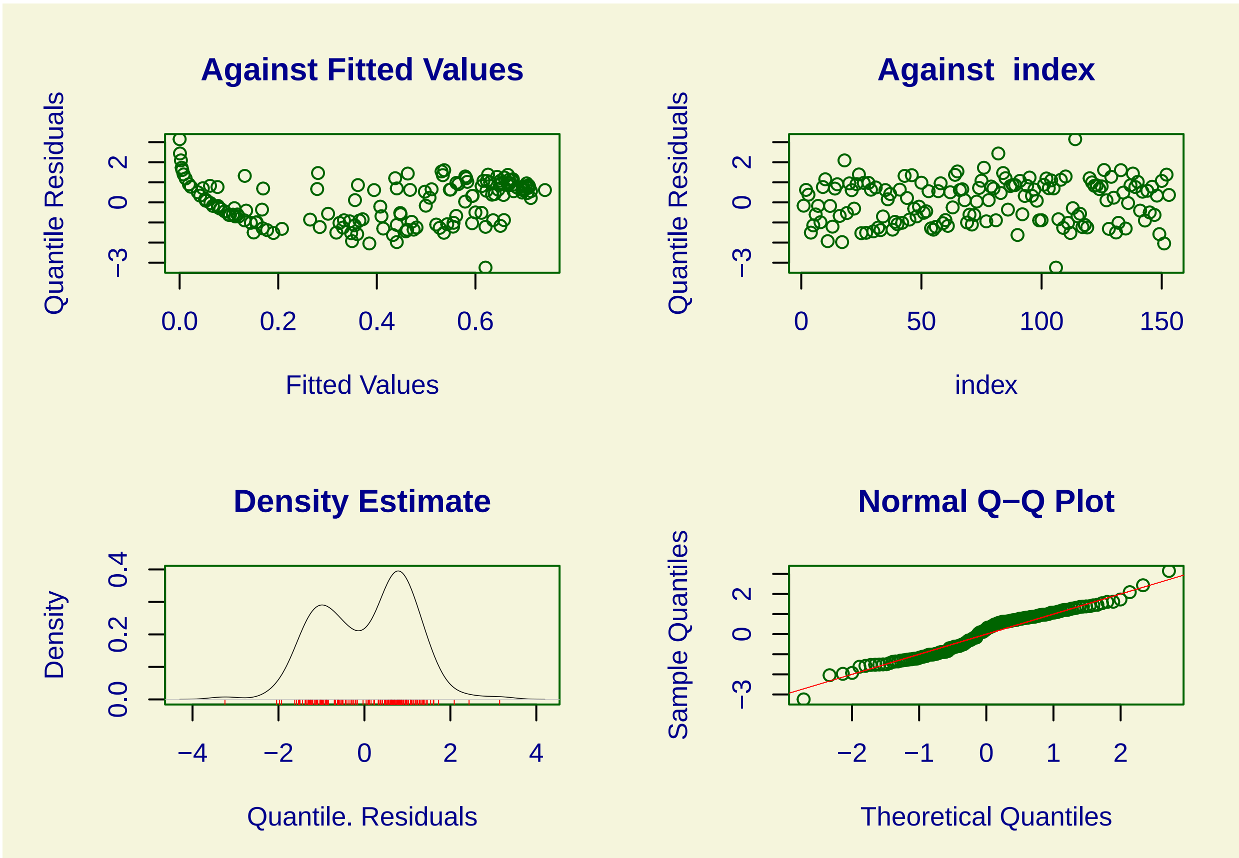

Figure 2: Residual Diagnostic Plots

8 References

1) Law, K. L. (2017). Plastics in the marine environment. Annual review of marine science, 9, 205-229.

2) Kühn, S., Bravo Rebolledo, E.L., van Franeker, J.A. (2015). Deleterious Effects of Litter on Marine Life. In: Bergmann, M., Gutow, L., Klages, M. (eds) Marine Anthropogenic Litter. Springer, Cham. https://doi.org/10.1007/978-3-319-16510-3_4

3) Hannah Ritchie and Max Roser (2018) - "Plastic Pollution". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/plastic-pollution' [Online Resource]

4) Geyer et al., 2017: Production, use, and fate of all plastics ever made | Science Advances

5) W.C. LI, H.F. TSE, L. FOK, Plastic waste in the marine environment: A review of sources, occurrence and effects, Science of The Total Environment, Volumes 566–567, 2016, Pages 333-349, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2016.05.084.

6) Jambeck et al., 2015 Plastic waste inputs from land into the ocean | Science

7) Thomas Mock (2019). Tidy Tuesday: A weekly data project aimed at the R ecosystem. https://github.com/rfordatascience/tidytuesday/tree/master/data/2019/2019-05-21.

8) United Nations. (n.d.). Undata | record view | GDP per capita, PPP (constant 2011 international $). United Nations. Retrieved December 15, 2022, from https://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3ANY.GDP.PCAP.PP.KD