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 et al., 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 et al., 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 (2010) 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 (2010) 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 (2010) PPP GDP per capita increases (i.e., for richer entities), the log-odds of an entity mismanaging its daily plastic waste (compared to properly managing its daily plastic waste in 2010) are expected to decrease (while holding the entity's coastal population proportion in 2010 constant). Additionally, I hypothesize that, as the relative size of an entity's (2010) 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 plastic waste (compared to properly managing its daily plastic waste in 2010) are expected to increase (while holding the entity's 2010 PPP GDP per capita constant).

1.5 Exploratory Data Analysis

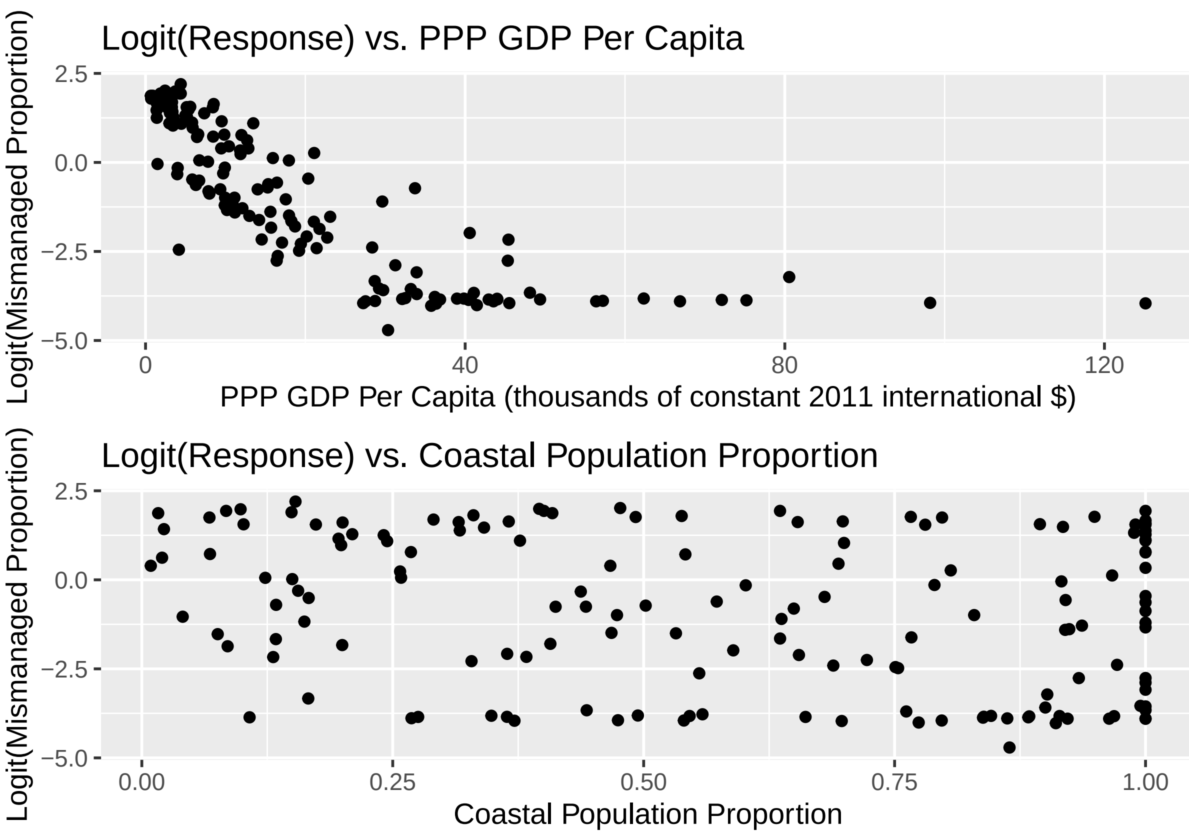

Figure 1: Top: The Relationship Between PPP GDP Per Capita and the Log-Odds of Mismanaging Daily Plastic Waste; Bottom: The Relationship Between Coastal Population Proportion and the Log-Odds of Mismanaging Daily Plastic Waste

Preliminary exploratory data analysis can be used to visually assess the hypotheses. Figure 1 (Top), in particular, shows a generally negative association between the log-odds of mismanaging daily plastic waste (compared to properly managing daily plastic waste) in 2010 and 2010 PPP GDP per capita. This appears to agree with my initial hypothesis. Figure 1 (Top) illustrates how the log-odds of mismanaging daily plastic waste (compared to properly managing daily plastic waste) varies among the three types of entities (low-GDP, middle-GDP, and high-GDP). Figure 1 (Bottom), on the other hand, does not provide strong evidence of an association between the log-odds of mismanaging daily plastic waste (compared to properly managing daily plastic waste) in 2010 and the relative size of an entity's 2010 coastal population. This appears to disagree with my initial hypothesis. Additionally, there appears to be a strong regional dominance in global mismanaged plastic waste as each of the top five entities (in terms of the proportion of mismanaged plastic waste) are located in Asia (see Appendix A).

2 Methodology

2.1 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. Another by-product of this decision is that the hierarchical structure of the raw data (i.e., information about each entity over time) was removed; the observations can be treated independently (i.e., independence assumption reasonably satisfied assuming one entity's mismanaged plastic waste does not affect another's). 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.

As described previously, one of the covariates of interest is the relative size of an entity's coastal population (i.e., the ratio of an entity's coastal population to its total population). One of the strange features of this dataset is that, for some entities, the coastal population is greater than the reported total population in 2010. This begs the question of how the coastal population data was collected (since this is obviously an error) in the first place. Nevertheless, I realized that the majority of observations for which the coastal population was greater than the total population were small islands. For such entities, it is reasonable to consider the entire population as coastal (i.e., every residence is less than 50 kilometers from a coastline). As a result, I decided to artificially cap all coastal population proportions at 1. Thus, the final dataset includes complete information (i.e., PPP per capita GDP, coastal population proportion, mismanaged waste proportion) about 153 entities. I decided to perform a complete case analysis since I could not find the missing information (i.e., 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, models fit using this dataset could have biased coefficient estimates and standard errors.

2.2 Model

Since the response variable is a proportion (values ranging from 0 to 1, not inclusive), I decided to fit a Beta regression model, specifically with a logit link function, and selected the important covariates (see Introduction for justifications) based on my understanding of the existing literature regarding global plastic pollution. Beta regression is a type of distributional regression that estimates the overall shapes of statistical distributions rather than lines and averages (like ordinary least squares). Beta regression is methodologically better than a linear probability model (ordinary least squares applied to a proportional outcome) or fractional logistic regression (logistic regression applied to a proportional outcome) since the Beta distribution naturally limits the outcome to the 0-1 range (not inclusive). I decided to use the logit link function (instead of, for instance, the probit link function) in the interest of interpretability. Logistic regression is a very common modeling technique and interpretations in terms of the log-odds are beneficial when communicating results to stakeholders (laypeople who do have a strong statistics background). I decided against including an interaction term between PPP GDP per capita and coastal population proportion for a similar reason. Since the research questions are inferential in nature, including an interaction term between two quantitative covariates is not ideal (as these interpretations are harder to digest than, for example, interpretations of an interaction between quantitative and categorical predictors). Using mathematical notation, the final model can be expressed as

$$\log(\widehat{\frac{\alpha_i}{\Gamma_i}}) = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$$

where i is the index for the entity (ranging from 1 to 153) and $\widehat{\frac{\alpha_i}{\Gamma_i}}$ is the expected odds of mismanaging daily plastic waste (compared to properly managing daily plastic waste) for the ith entity in 2010. In terms of the covariates, x_{1i} indicates the ith entity's 2010 PPP GDP per capita (in thousands of constant 2011 international dollars) and x_{2i} indicates the relative size of the ith entity's coastal population (compared to its total population) in 2010.

The assumptions for a Beta regression model are linearity and independence. As mentioned earlier, the independence assumption is reasonably satisfied. Linearity (in the coefficients) can be assessed by examining the relationship between the (standardized weighted) residuals and the linear predictor (see Appendix B). In this case, the plot in Appendix B suggests that the linearity assumption is violated as the points are not randomly scattered around zero (i.e., there are patterns in the residual plot). This violation of the modeling assumptions could be related to the near-zero proportions (of plastic waste that is mismanaged) corresponding to entities, such as South Korea (0.0089), Japan (0.0175), and Australia (0.0179). Linearity could also be violated simply because of the inadequacies of the data; that is to say, the Beta regression model may be missing important covariates that are not available in the final dataset. Despite the violation of this modeling assumption, I continued with the analysis (and this model) because I could not find a model of comparable complexity which would facilitate simple interpretations in terms of log-odds ratios (see Discussion for discussion of the zero-inflated Beta regression model). Although the coefficient associated with the relative size of an entity's coastal population is not statistically significant (assuming a significance level of 0.05), I decided to keep the term in the final model because I believe that relevant stakeholders would be very interested in the association between coastal population proportion and the log-odds of mismanaging plastic waste (compared to properly managing waste), while holding the entity's PPP GDP per capita constant.

3 Results

The regression coefficients were estimated using the `betareg()` function in R. Table 1 displays the coefficient estimates (for the μ component; ignoring the precision parameter ϕ for the sake of interpretations), 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)	1.114	0.831	1.397	<0.01
`Per Capita GDP (thousands of international $)`	-0.084	-0.092	-0.075	<0.01

	Coefficient	Lower Bound	Upper Bound	P-Value
`Coastal Population Proportion`	-0.363	-0.764	0.037	0.0755

For each thousand constant 2011 international dollar increase in 2010 PPP GDP per capita, we expect the log-odds of an entity mismanaging its daily plastic waste (compared to properly managing its daily plastic waste) in 2010 to decrease by 0.084, holding the proportion of the entity's population that lives within 50 kilometers of a coastline in 2010 constant. We are 95% confident that, for each thousand constant 2011 international dollar increase in 2010 PPP GDP per capita, the log-odds of an entity mismanaging its daily plastic waste (compared to properly managing its daily plastic waste) in 2010 are expected to decrease by between 0.075 and 0.092, holding the proportion of the entity's population that lives within 50 kilometers of a coastline in 2010 constant. This means that there is statistically significant evidence (assuming a significance level of 0.05; p-value < 0.01) supporting the initial hypothesis regarding this association.

For each 0.01 (i.e., one absolute percentage point) increase in the proportion of the population that lives within 50 kilometers of a coastline in 2010, we expect the log-odds of an entity mismanaging its daily plastic waste (compared to properly managing its daily plastic waste) in 2010 to decrease by 0.363, holding the entity's 2010 PPP GDP per capita constant. We are 95% confident that, for each 0.01 increase in the proportion of the population that lives within 50 kilometers of a coastline in 2010, the log-odds of an entity mismanaging its daily plastic waste (compared to properly managing its daily plastic waste) in 2010 are expected to change by between -0.764 and 0.037, holding the entity's 2010 PPP GDP per capita constant. Unfortunately, as the 95% confidence interval includes zero (and the p-value > 0.05), the association between an entity's coastal population proportion and the log-odds of mismanaging its daily plastic waste is not statistically significant (assuming a significance level of 0.05). This means that there is insufficient evidence in support of my initial hypothesis (which stated that, as an entity's 2010 coastal population proportion increases, the log-odds of an entity mismanaging its daily plastic waste compared to properly managing its daily plastic waste in 2010 are expected to increase, while holding the entity's 2010 PPP GDP per capita constant) regarding this association.

4 Discussion

4.1 Conclusions

4.2 Limitations and Future Directions

This analysis does have several limitations. As mentioned in the methodology, the assumption that the data is missing completely at random may be inappropriate, and other missing data approaches (which presuppose a different missingness mechanism) could be more suitable than the complete case analysis described here. Another limitation of the data relates to the reported coastal populations for each entity (in 2010). For some reason, the coastal population of certain entities exceeds their total population. This suggests that the coastal population data is likely inaccurate. My approach to handle coastal population proportions greater than one was to truncate them to one; however, this may not be the best option. Additionally, the linearity assumption of the Beta regression model is likely violated. Also, although I assume that the entities can be treated independently, this is potentially an oversimplification because wealthier entities often trade their plastic waste with poorer entities; this trading network "effectively transfers [plastic] waste from [entities] of low risk of ocean pollution to [entities] with moderate-to-high risk" (Ritchie & Roser, 2018). This means that the results (inference) and subsequent conclusions are not necessarily reliable. Although the dataset does not include any entities with a mismanaged plastic waste proportion of 0 or 1, there are several entities with a mismanaged waste proportion near 0. These small outcome values (near the lower extreme) are concerning, particularly based on Appendix B. In this analysis, I considered an extension of Beta regression that allows for explicit zeros–zero-inflated Beta regression. Zero-inflated Beta regression is a mixture of a logistic regression model (which predicts if an outcome is 0 or not) and a Beta regression model (which predicts if an outcome is between 0 and 1, as long as it is not zero). I ultimately decided to use the Beta regression model because the two models performed similarly, and I preferred the simpler one. In the future, another technique I could try would be to artificially create zeros in the dataset (i.e., drop the values near 0 to 0) and then fit a zero-inflated Beta regression model; I would guess that, in this case, the zero-inflated Beta regression model would significantly outperform the normal Beta regression model.

It is important to acknowledge the limited generalizability of the conclusions. As part of the initial data cleaning phase, I decided to exclude all entities without complete information about plastic waste generation and limited the scope to the year 2010. As such, this analysis only provides information about the mismanagement of plastic waste (and more broadly, plastic waste infrastructure) in 2010 among the 153 selected entities. There are therefore potential issues with selection bias in the data, as this sample (of 153 non-landlocked entities; in the year 2010) is not necessarily representative of the population of interest (all non-landlocked entities–since access to a body of water is critical; in any given year) nor randomly selected. In future work, I would love to investigate other confounding variables and consider a greater subset (if not all) of entities. It could be interesting to see how the results would change if PPP GDP per capita were incorporated into the model as a categorical variable with levels of low, middle, and high (or perhaps even more stratified) rather than as a quantitative predictor variable.

4.3 Summary

5 Appendices

5.1 Appendix A

Table 2: Entities with Highest Proportion of Mismanaged Plastic Waste

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

5.2 Appendix B

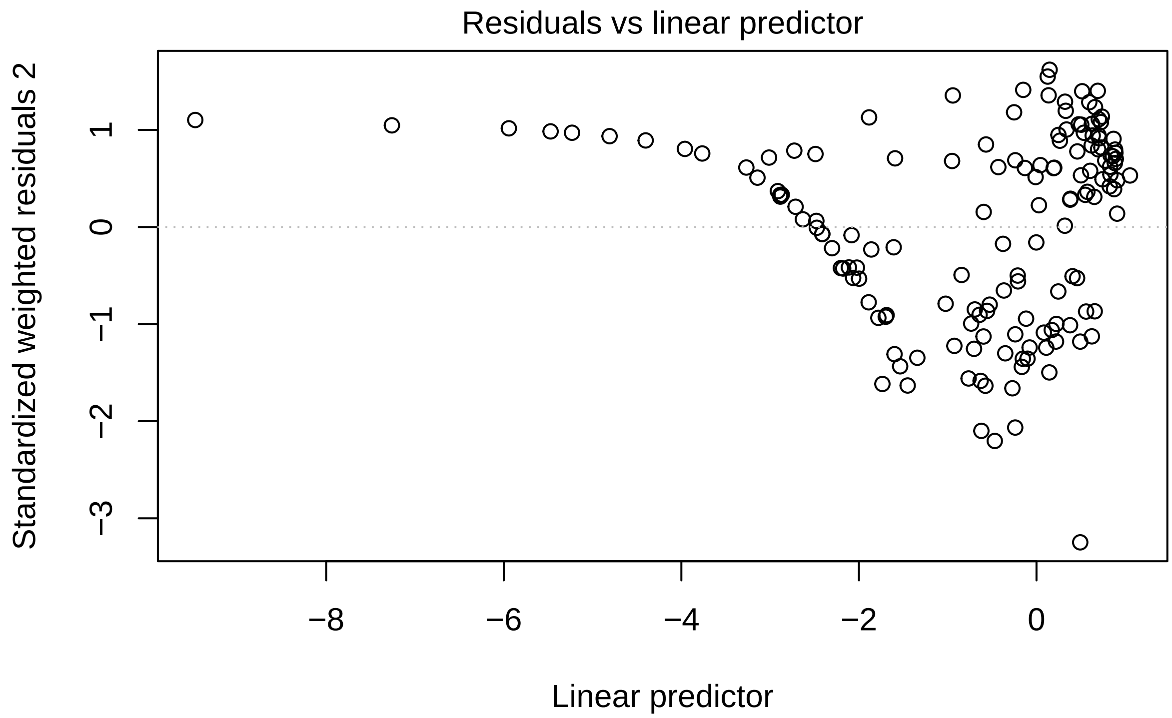

Figure 2: Assessing the Linearity Assumption

6 References

1) Law, K. L. (2017). Plastics in the marine environment. Annual Review of Marine Science, 9(1), 205–229. https://doi.org/10.1146/annurev-marine-010816-060409

2) Kühn, S., Bravo Rebolledo, E. L., & van Franeker, J. A. (2015). Deleterious effects of litter on marine life. Marine Anthropogenic Litter, 75–116. https://doi.org/10.1007/978-3-319-16510-3_4

3) Ritchie, H., & Roser, M. (2018, September 1). Plastic pollution. Our World in Data. Retrieved December 16, 2022, from https://ourworldindata.org/plastic-pollution

4) Geyer, R., Jambeck, J. R., & Law, K. L. (2017). Production, use, and fate of all plastics ever made. Science Advances, 3(7). https://doi.org/10.1126/sciadv.1700782

5) Li, W. C., Tse, H. F., & Fok, L. (2016). Plastic waste in the marine environment: A review of sources, occurrence and effects. Science of The Total Environment, 566-567, 333–349. https://doi.org/10.1016/j.scitotenv.2016.05.084

6) Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., & Law, K. L. (2015). Plastic waste inputs from land into the Ocean. Science, 347(6223), 768–771. https://doi.org/10.1126/science.1260352

7) Mock, T. (2019). Rfordatascience/tidytuesday: Official repo for the #tidytuesday project. GitHub. Retrieved December 16, 2022, from 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