Do countries that have higher plastic waste have higher ingestion of microplastics?

Indri Adisoemarta, Ishani P. Cheshire, Nadaa Moharram

2024-12-13

Abstract

Microplastic ingestion is a major health concern, as microplastics can mimic human hormones, increasing risk of fetal developmental issues and cancer (Gonzalez, 2024). We hypothesize countries with higher proportions of plastic waste to have higher rates of ingestion of microplastics. We approximate human ingestion of microplastics through seafood in coastal countries with the average microplastic concentrations found in their adjacent oceans. We create two linear models—one directly relating microplastic concentrations to plastic waste emitted, and one which additionally considers country GDP—neither of which ultimately achieve statistical significance. In the future, we hope to add average ocean current velocity to our model, as we believe this would explain a great amount of the variation in microplastic concentration.

Introduction

Microplastics are plastic particles that are less than 5mm in length, and have been identified in our atmospheric, aquatic, and terrestrial environments, as well as in drinking water and food products (Senathirajah, 2021). Chemicals in microplastics can mimic human hormones such as BPA and PFAS (Gonzalez, 2024), and exposure to these substances has been shown to increase the risk of conditions ranging from infertility to poorer fetal development and cancer. Seafood is one possible vector of exposure to microplastics, and it's been estimated that up to 50,000 microplastic particles are ingested annually by adults through seafood (Nicole, 2021). To address this issue, our research question is:

Do countries that have higher plastic waste have higher ingestion of microplastics?

We will approximate human ingestion of microplastics through seafood with the microplastic concentrations found in the oceans near the countries of interest. Then, we will build a linear model to determine whether there is a relationship between the plastic waste generated by a country and the associated microplastics found in nearby water sources. The results of our descriptive analysis will inform governments of the importance of reducing plastic waste to reduce microplastic consumption and improve public health.

Data

To get data on microplastic concentrations per region, we pulled data from the National Centers for Environmental Information. The dataset includes the associated ocean, region, sampling method, measurement, measurement units, latitude, longitude, as well the research group that collected the data. To find associated plastic waste per country in tons, we pulled data from Our World in Data, which has plastic waste emitted per country to the ocean, for the year 2019. To find associated Gross Domestic Product per country in US dollars, we pulled data from World Bank Group reports for the year 2019.

The microplastic data only has latitude, longitude, and the associated ocean region (like the Indian Ocean.) In order to join the two datasets by country, we used the rnaturalearth library to identify the closest country by distance, and we validated the points visually and by mapping the points with R. One issue with this association is that this does not take into account ocean current movements, as plastic located close to a country does not necessarily mean it came from that country.

Operationalization

While it may be that the plastic waste produced by a country does not, in fact, end up in the marine areas closest to that country, this in and of itself would be an interesting result, which would be reflected in our data. In the dataset, we will be using the variable "Microplastics Measurement (density)" as our outcome variable (Y). We are using this outcome variable instead of the actual ingestion rate of microplastics by fish as we don't have any data on microplastic ingestion by fish, data on marine microplastic density is plentiful, and numerous studies have seen a statistically significant relationship between particle concentration and particles/fish (Roch et al., 2020.) Our descriptive variable (X) will be "Total Plastic Waste (MT)" produced by a country. Another descriptive variable we will consider is Gross Domestic Product and whether it's related to microplastic measurement.

Out of 195 recognized countries our data contains 50 as we were focused on coastal countries as we could then identify ocean plastic to countries based on the distance of the plastic to the country. Given that our dataset is collection of reports from various universities around the world, this research spans the years 2018 to 2021 so it could have enough data from different areas of the world. Restricting the data to one year strongly limited our dataset. The GDP and emitted plastic data used was from the year 2019 as to avoid any covid impact. Furthermore, we removed data that reported 0 recorded waste emitted as after observing the distribution of our data we had to take the log of our outcome and descriptive variables. Lastly, to combine GDP and emitted waste for each country with the amount of microplastic in the ocean a weighted average was calculated as our microplastic data contained several data points from various universities for each country. To create a weighted measurement for each country based off of the proximity of the recorded observation we took the sum of the weight multiplied by the measurement divided by the sum of the weight, where weight was one over the distance.

Model Assumptions

We start with the simplest possible model, $microplastic = \beta_0 + \beta_1 plastic\_waste$. Our small sample size ($n = 50$) means we must evaluate the CLM assumptions, starting with IID. This data will never be completely independent due to the nature of our topic—as countries that are closer to each other will have more similar ocean microplastic levels. There is also some bias in our selection of countries, as wealthier countries will be more likely to have the grants to fund environmentalist research like these microplastic surveys. Also, the data had to be taken over a span of three years in order to get enough data points for our analysis, creating some association for samples that were taken closer together temporally—however, we were careful not to expand the time range further than this, and we don't expect microplastic levels to change substantially over such a short span of time. Given the constraints, this data was collected as independently as possible with measurements coming from different research groups across the globe, so we accept the above risks as somewhat inevitable for this dataset. Despite this, our raw data still does not fulfill the independence requirement due to the significant clustering at lower values, which becomes visible when we plot each countries' average oceanic microplastics versus their emitted plastic waste (see Appendix, Figure A1, for supporting graphs.) To remedy this, we create Model 2 — a logged transformed version of model 1.

Model 2 improves on our first model by logging both our y and x variables: $log_{10}(microplastic) = \beta_0 + \beta_1 log_{10}(plastic\_waste)$. To do this, we need to remove the few (6) data points with a zero value.
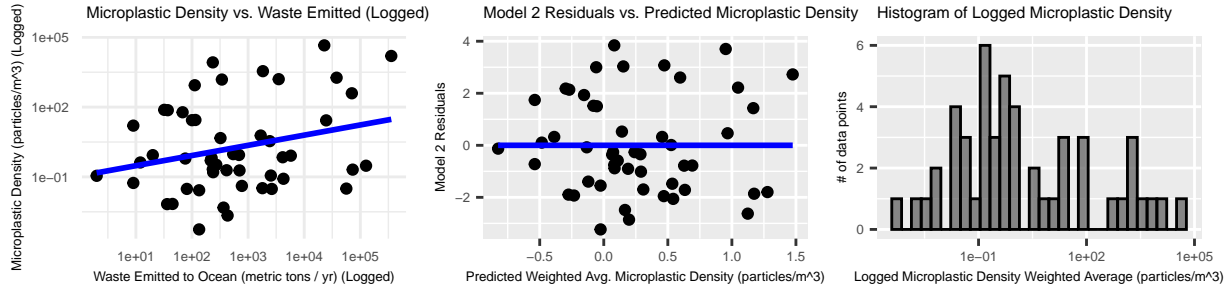
When assessing the first assumption of IID, we first note that the earlier clear clustering of data points disappears. Even though countries themselves are associated with each other, considering the limitations of the data and the lack of clustering, we allow the assumption of independence to stand.

The data must also be identically distributed. Our plastic waste emitted per country data is likely to be identically distributed, as data is being drawn from the larger overall distribution of countries across the globe. Microplastic concentration, however, had a huge amount of variation in sampling methods. Some sampling methods measured raw 'counts' of particles, while others measured 'particles per time', and still others measured 'particles/volume'. These different sampling methods were, by their very nature, not identically distributed and could not be standardized and compared to each other. We didn't have enough data points to limit our data to one sampling method, as would be ideal, so we instead limited our sample to only those sampling methods which measure the # of particles/volume. This number density can then be compared across sampling methods with reasonable confidence. Once filtered to only include number density measurements, the logged version of our microplastic concentrations seems to be identically distributed, as shown in the histogram. Therefore, we find identically distributed data to be a decent assumption.

For our assumption of a linear conditional expectation, we note that the residuals of model 2 do not have a pattern. Instead, they seem random, giving us no evidence to reject this assumption. Regarding the third assumption of no perfect collinearity, we again note that model 2 also has only one x-variable, making this requirement redundant. We allow this assumption to pass automatically.
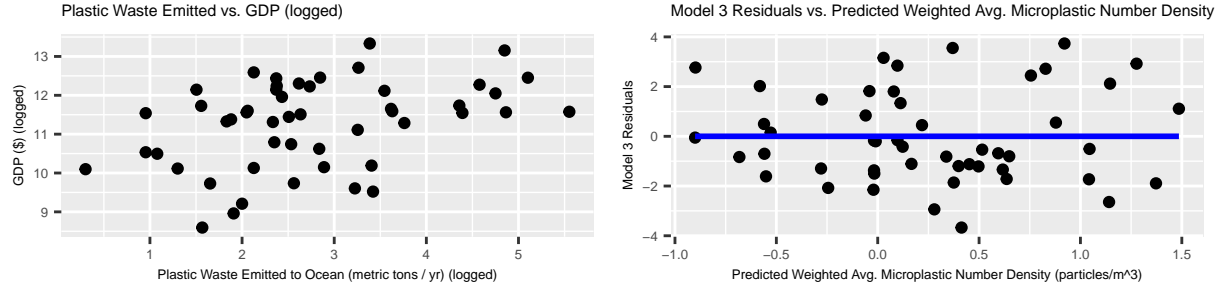
When looking at the model 2 residuals plot, the errors appear to be homoskedastic—that is, the magnitude of the residuals don't change as we move to higher microplastic number density values. When we additionally conduct the Breushch-Pagan test (H_0: homoskedasticity, H_a: heteroskedasticity, alpha = 0.05), we find a p-value of 0.146, indicating that we fail to reject the possibility of homoskedasticity. Therefore, with our ocular and statistical tests in agreement, we proceed confidently with the assumption of homoskedastic errors.

Finally, the errors from our model must be normally distributed. All of our points on the q-q plot (see Figure

A2 in the Appendix) fall within the 95% confidence interval of normally distributed errors. Therefore, we proceed with this assumption. Overall, model 2 fulfills the required five CLM assumptions.

For Model 3, we decided to add GDP as a variable to account for the fact that plastic waste emitted is an easily manipulated variable by countries with more wealth, as they can send their plastic waste to poorer countries. For example, we saw that Norway had 0 tons of plastic waste emitted while the Philippines had the most plastic waste emitted. The model becomes $log_{10}(microplastic) = \beta_0 + \beta_1 log_{10}(plastic\_waste) + \beta_2 log_{10}(GDP)$.

To determine whether plastic waste is independent of GDP, it looks like there is some clustering as GDP is a skewed variable. When logging both plastic waste emitted and GDP, however, the spread is easier to see and no clear clusters are identified. There is nothing to indicate the data is not from identical distributions as there is only one GDP point per country.

We next evaluate whether the errors are homoskedastic. When looking at the model 3 residuals plot, the errors appear to be homoskedastic—that is, the magnitude of the residuals don't change as we move to higher microplastic number density values. When we additionally conduct the Breushch-Pagan test (H_0: homoskedasticity, H_a: heteroskedasticity, alpha = 0.05), we find a p-value of 0.201, indicating that we fail to reject the possibility of homoskedasticity. Therefore, with our ocular and statistical tests in agreement, we proceed confidently with the assumption of homoskedastic errors.

It is hard to see collinearity with the unlogged data, but with the logged data the spread is easier to see. You can see a slight correlation between log(Plastic Waste Emitted) and log(GDP). Refer to the Appendix (Figure A3) for the correlation graph. The VIF values are close to 1 (1.178), which indicates moderate (but not severe) correlation between predictor variables of log of plastic waste emitted and log GDP.

To evaluate linear conditional expectation, we examine the residuals. The model residuals compared with the log(GDP) variable and the log(Plastic Waste Emitted) appear to be almost linear, and there are no points where the residuals are significantly different from 0.
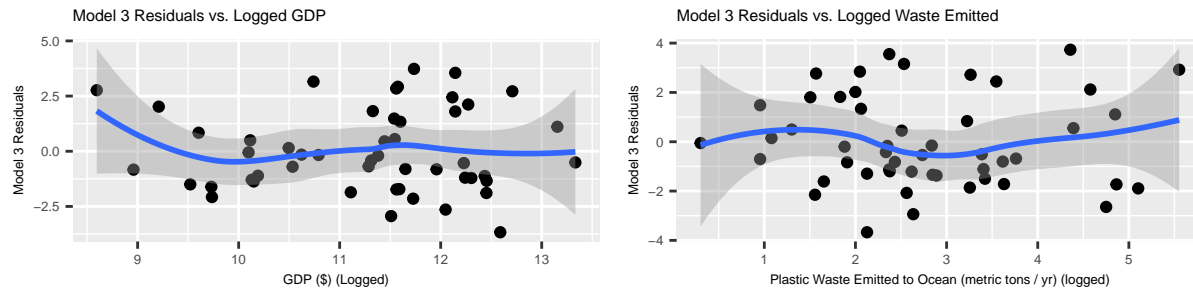
Finally we need to determine whether the errors from our model are normally distributed. All points on the q-q plot (see Figure A4 in the Appendix) fall within the 95% confidence interval of normally distributed errors. Therefore, we proceed with this assumption. Overall, model 3 fulfills the required five CLM assumptions.

Model Results and Interpretation
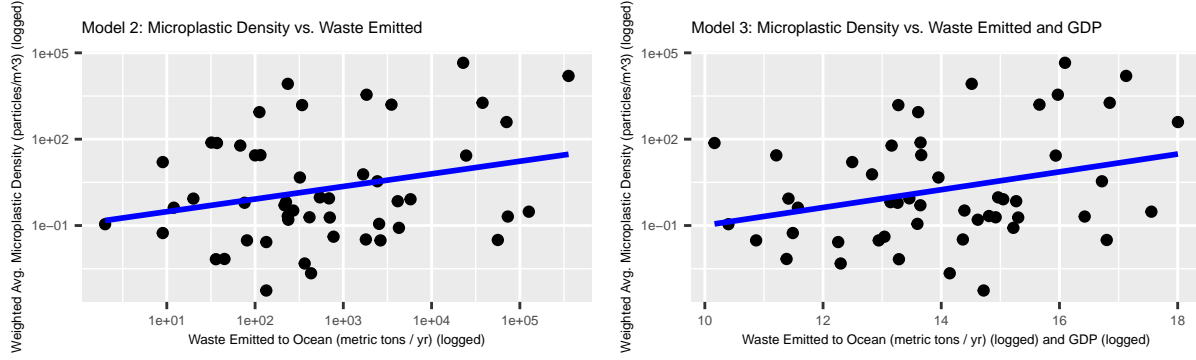
For model 2, $log_{10}(microplastic) = \beta_0 + \beta_1 log_{10}(plastic\_waste)$, the interpretation of our slope would have been that the microplastic number density off the coast of a country is exponentially related to that country's plastic waste emitted, with the power of the exponent being the our slope. However, the p-value for the slope is 0.06 which is slightly above the threshold value. This suggests weak evidence that the waste emitted by a country impacts the amount of microplastics found in their nearby ocean, so our model 2 slope has no practical interpretation. The model 2 R squared value is only 0.072 indicating that there are more variables we should be looking at to explain microplastic correlation.

For model 3, $log_{10}(microplastic) = \beta_0 + \beta_1 log_{10}(plastic\_waste) + \beta_2 log_{10}(GDP)$, the p-value for the slope is 0.18 for GDP which indicates that this relationship is not statistically significant. The same follows for emitted waste with a p-value of 0.29. This suggests weak evidence that the waste emitted by country and GDP impacts the amount of microplastics found in the ocean, so our model 3 has no practical interpretation. The model 3 R squared value is only 0.094 indicating that there are more predictors to assess to explain microplastic correlation.

Table 1: Regression Results for Model 2 and Model 3

	Dependent variable:	
	Microplastic particles/meter cubed	
	Model 2	Model 3
	(1)	(2)
	mod2	model3
Plastic Waste Emitted (metric tons/yr) (logged)	0.438*	0.335
	(0.227)	(0.246)
GDP in dollars (logged)		0.282
		(0.263)
Constant	−0.956	−3.846
	(0.683)	(2.785)
Observations	50	50
R^2	0.072	0.094
Adjusted R^2	0.052	0.055
Residual Std. Error	1.884 (df = 48)	1.881 (df = 47)
F Statistic	3.711* (df = 1; 48)	2.433* (df = 2; 47)
Note:	*$p<0.1$; **$p<0.05$; ***$p<0.01$	

Although our model 2 came close to the required threshold, ultimately none of our models were statistically significant, indicating that there are more variables we needed to look at to explain microplastic concentration. In the future, we hope to look into adding average ocean current speed to our model, as we believe that this would have a major impact on the distribution of microplastic particles.

Appendix

Links to Data Sources

List of Model Specifications

All of the models we tried are mentioned in the paper.

Residuals vs. Fitted Values Plot

See our Model Assumptions.

Additional Figures

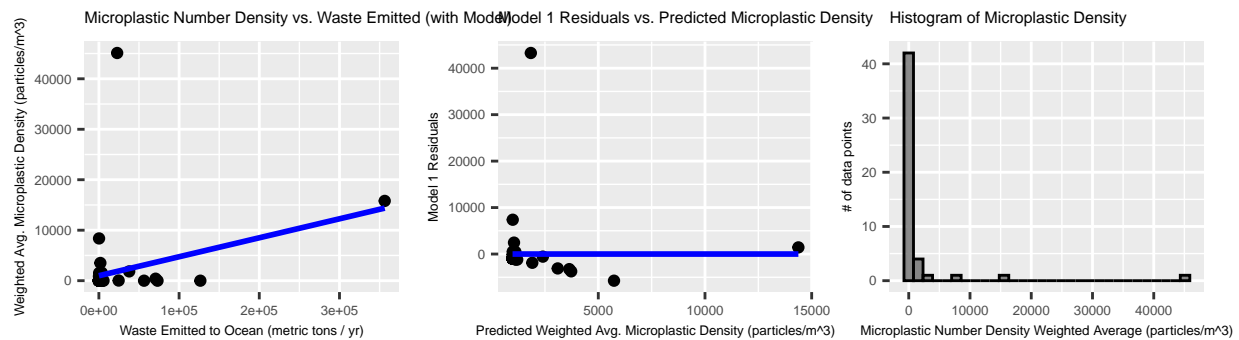### Figure A1

Supporting graphs for Model 1
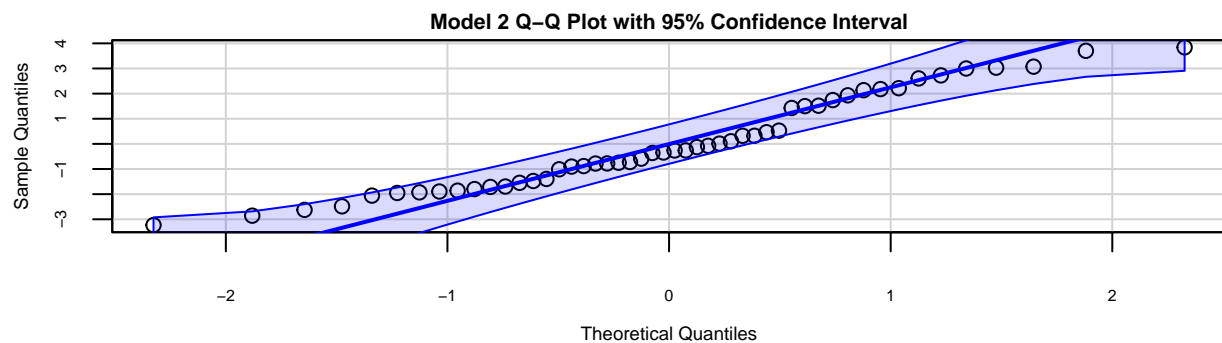

Figure A2

Q-Q Plot for Model 2

Figure A3
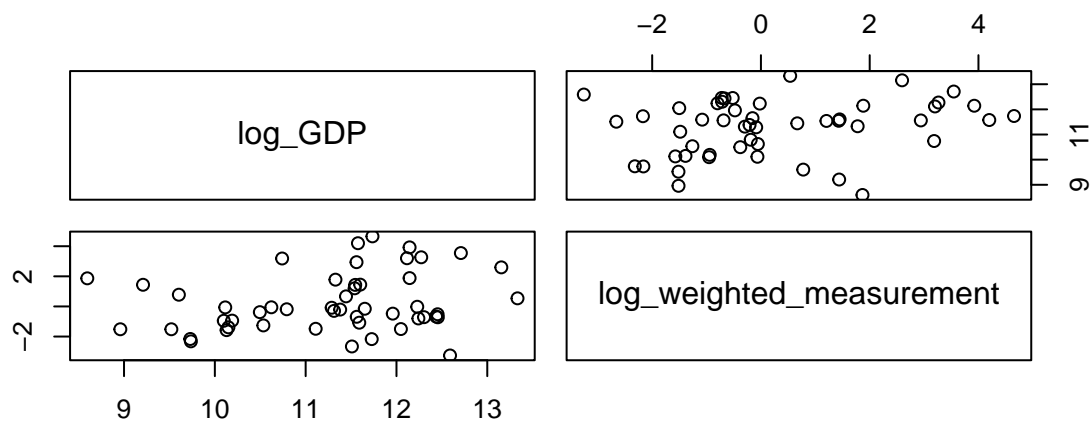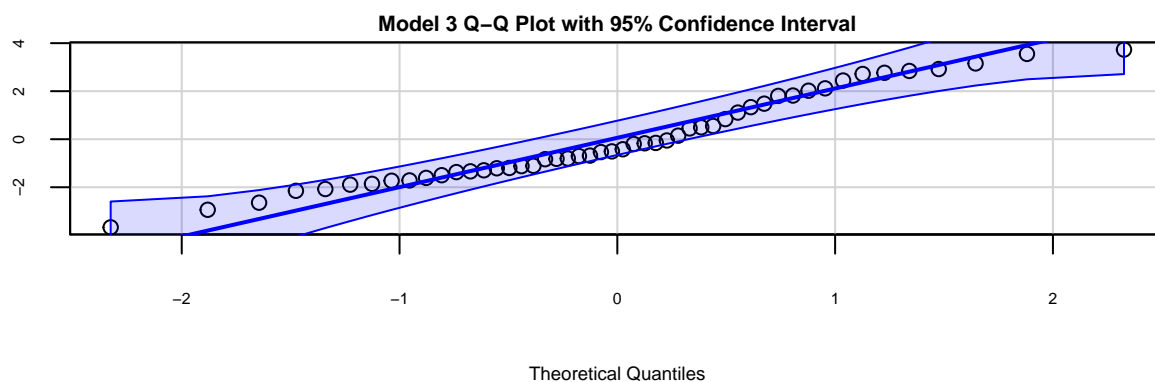
Pairs Plot for Model 3

Figure A4

Q-Q Plot for Model 3

Theoretical Quantiles

References

Senathirajah, 2021. Journal of Hazardous Materials, Volume 404, Part B, 2021, 124004, ISSN 0304-3894, https://doi.org/10.1016/j.jhazmat.2020.124004.

González, 2024. UCSF https://www.ucsf.edu/news/2024/02/427161/how-to-limit-microplastics-dangers

Nicole, W. 2021, Environmental Health Perspectives, 129, 3034001, doi: 10.1289/EHP89364. https://ehp.niehs.nih.gov/doi/10.1289/EHP8936