
Plastic Waste and Microplastics

— Nadaa Moharram, Ishani Cheshire, Indri Adisoemarta —
DATASCI 203 Lab 2
December 9, 2024

Motivation



Waste



Ingestion



Consumption



Hospitalization

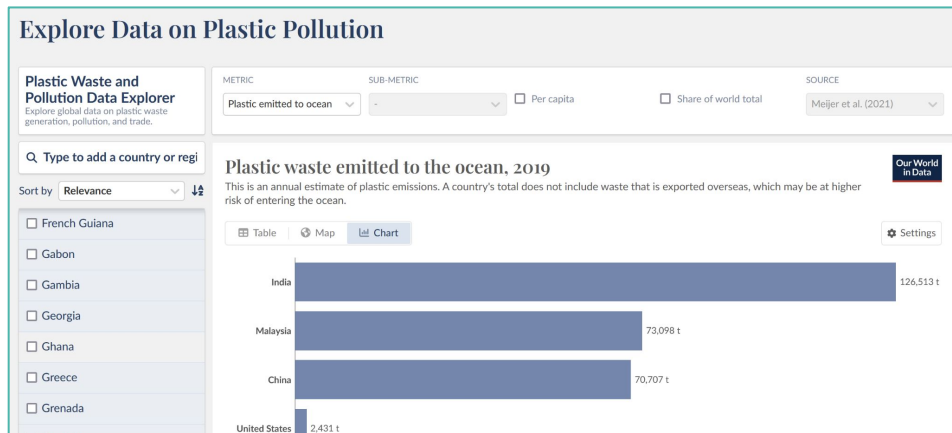
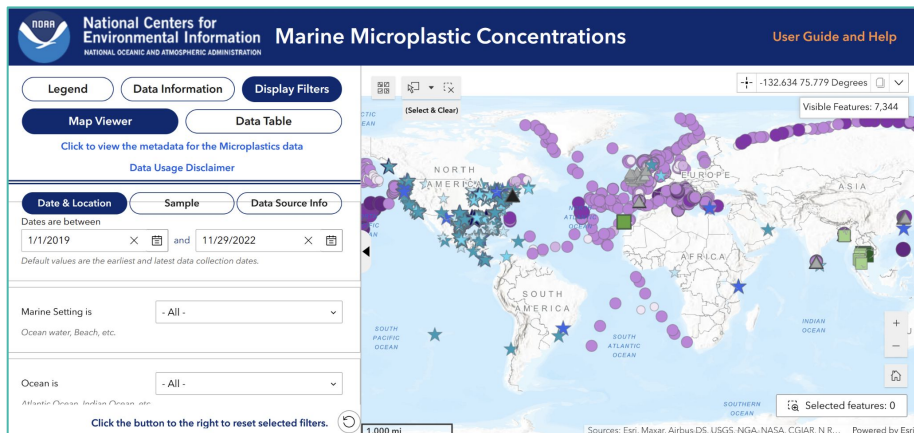
Research Question

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

Data Sources

Microplastic data: **NCEI Marine Microplastic Concentrations Map Portal**

Plastic waste data by country: **Our World in Data**



Data Wrangling

Microplastic data does not have associated country, just latitude, longitude, ocean region

Used the **rnaturalearth** library to identify the closest country by distance

Validated by **visual inspection** and **mapping points with R**

Does not take into account ocean current movements

Latitude	Longitude	country	distance_to_country_km
34.816	26.13	Greece	22.35889312
36.023	29.562	Turkey	17.91744446
69.82389	50.555278	Russia	58.6616098

Data Issues

2019 data **did not include India and China**, so we expanded dataset to 2019 - 2021

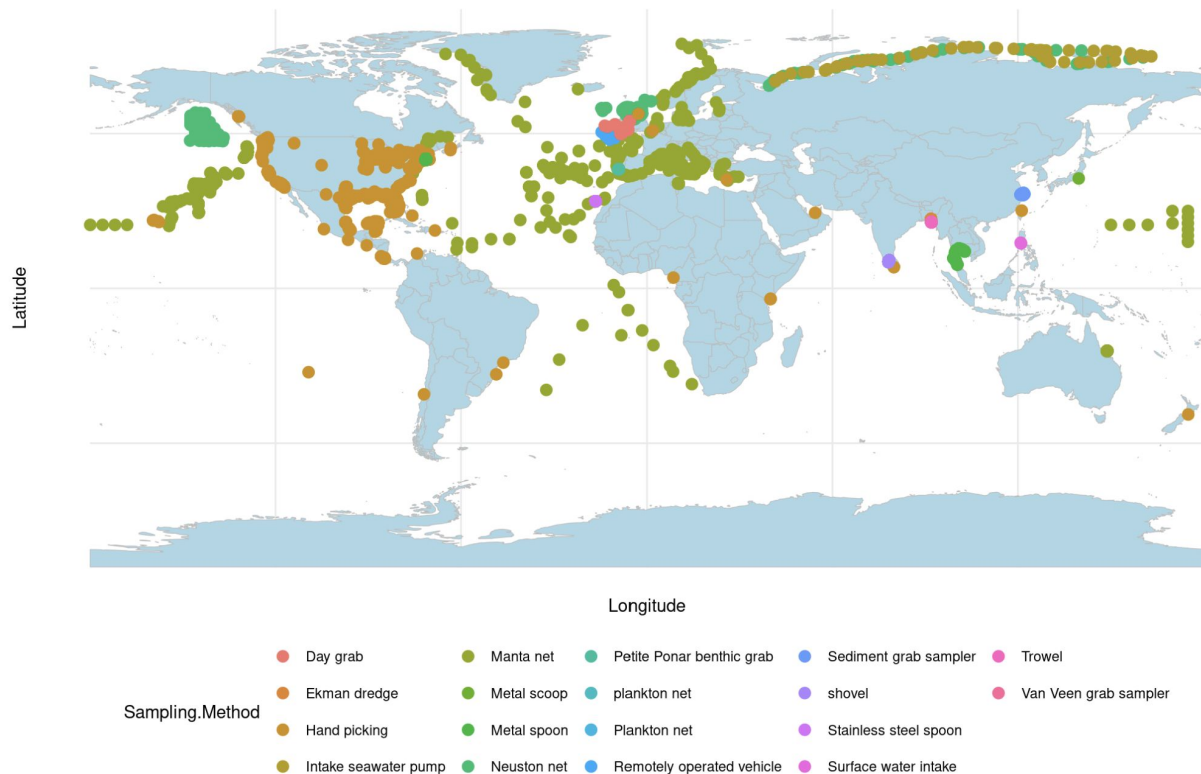
Other issues:

Very few microplastic points in Indian Ocean

Very few points around Asia, Southeast Asia, and South America

Possible temporal collinearity between samples taken in the same place on different years

Geographic Points on a World Map



Microplastic Weighting

- To create a weighted measurement where each country has a single measurement based off on the proximity of the observation (how close the microplastic was to the country)

Weight = $1/(\text{distance} + 0.001)$ → (closer distance get higher weights)

Aggregate by country:

$\text{weighted_average_measurement} = \text{sum}(\text{weight} * \text{measurement}) / \text{sum}(\text{weight})$

Final Dataset

Our final dataset has **38 data points** (from an original set of 1378 points), one for each country that has the needed weighted microplastic data and total plastic waste emitted

This means we will need to use the more stringent **CLM small sample assumptions**

```
> summary(weighted_df)
  country      weighted_measurement WasteEmitted_MetricTonsYear
Length:43      Min.   :    0.00      Min.   :    0.0
Class :character 1st Qu.:    0.05      1st Qu.:   35.0
Mode  :character Median :    0.26      Median :   236.5
              Mean   : 1902.55      Mean   : 18713.3
              3rd Qu.:    0.88      3rd Qu.: 3045.0
              Max.   :45109.75      Max.   :356371.0
```


Model Structure

Number Density of Microplastics = $\beta_0 + \beta_1$ Plastic Waste Emitted to the Ocean

Starting on Assumptions

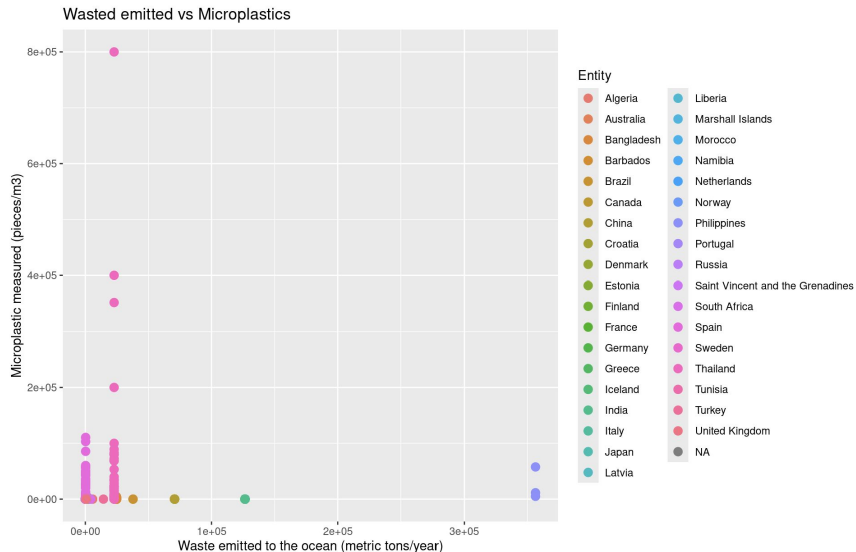
1. IID
2. Linear Conditional Expectation (**requires finalized model**)
3. No perfect Collinearity (**requires finalized model**)
4. Homoskedastic Errors (**requires finalized model**)
5. Normal Errors (**requires finalized model**)

IID

Independent:

Countries are never truly independent of each other, and sections of the ocean are not truly independent.

However, these measurements are as independent as can be reasonably expected of this kind of data set.

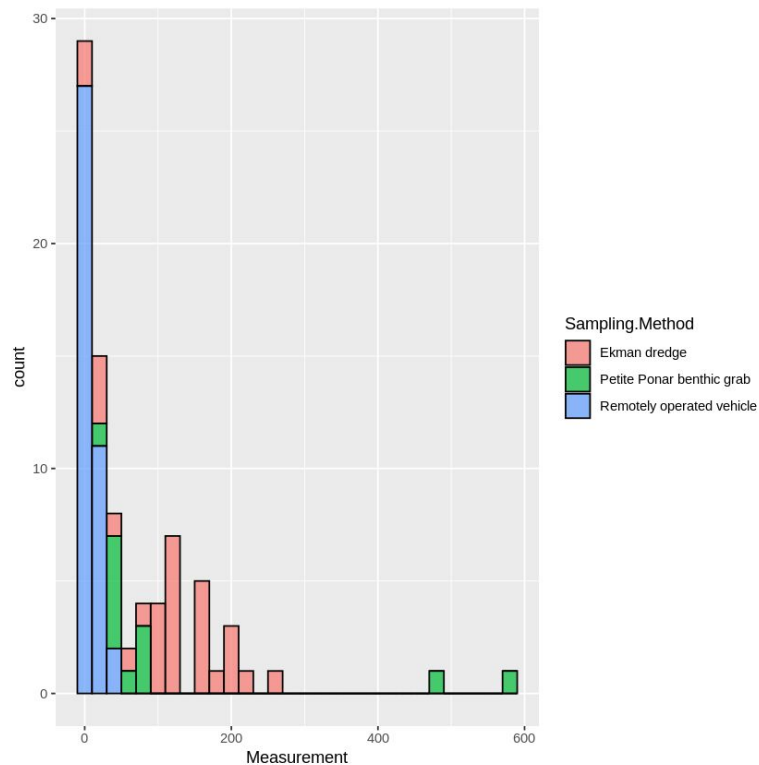


IID (cont.)

Identically Distributed:

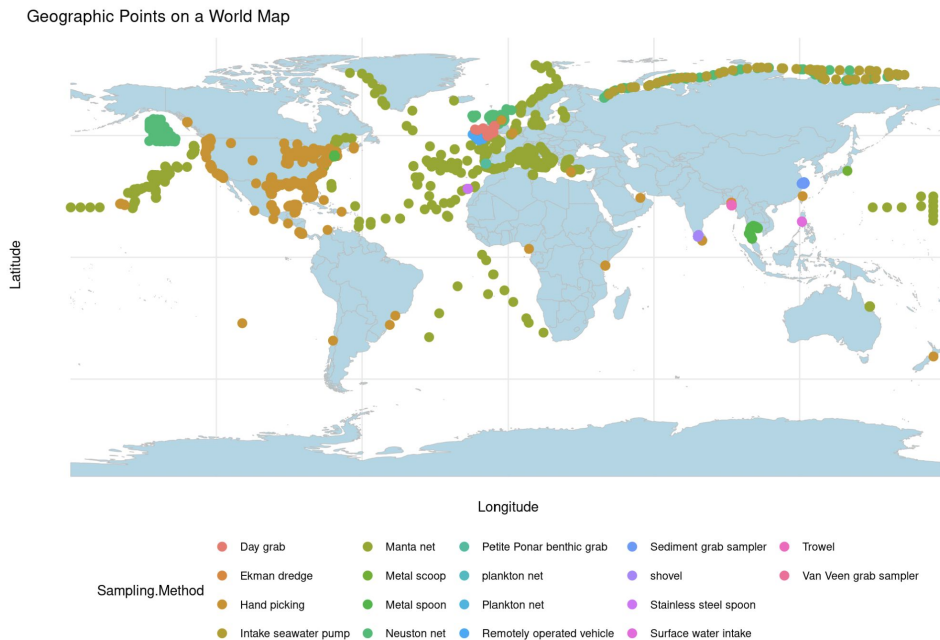
The data set combines different sampling methods. We cut all methods which don't measure in the dimension of number density (number of particles over volume).

Still, there appears to be some bias in measurements depending on the sampling method.



IID (cont.)

However, this difference could be due to different sampling methods being favored in different parts of the world.



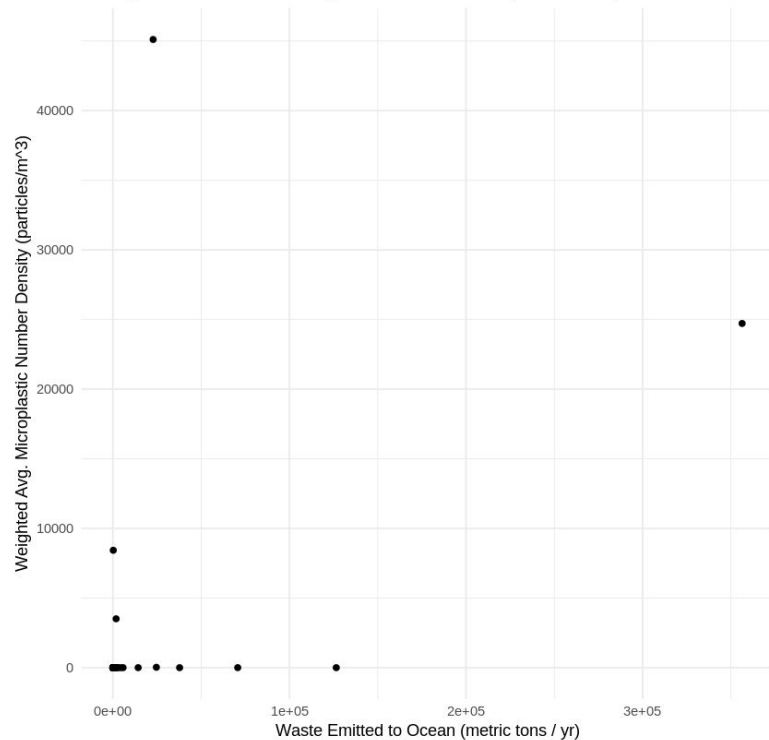
Change our model?

Number Density of Microplastics = $\beta_0 + \beta_1$ Plastic Waste Emitted to the Ocean
+ β_2 Ekman Dredge Method
+ β_3 Remotely Operated Vehicle

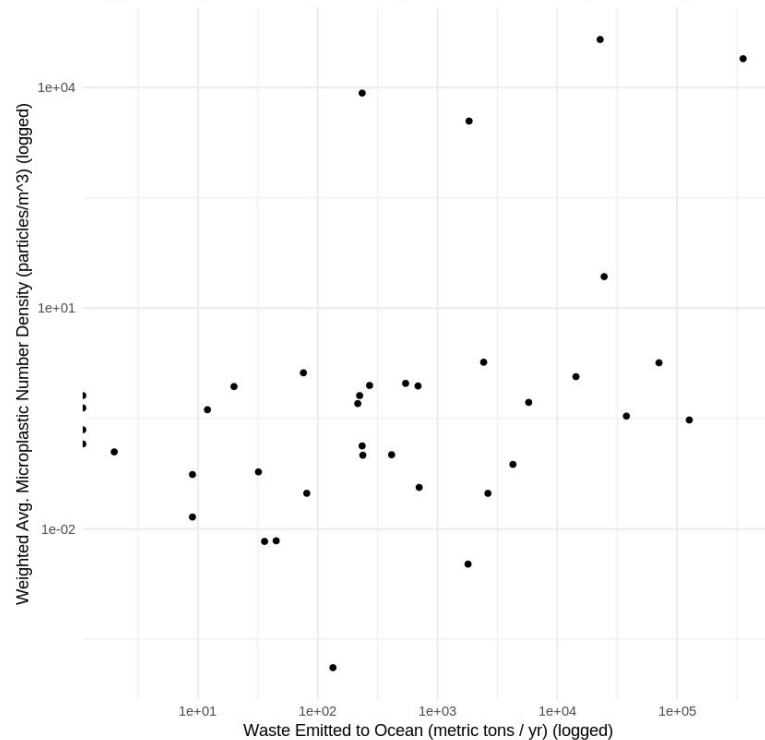
But this would contradict with our current method of taking the weighted average of all plastic measurements... (Also, possible collinearity.)

Current Data:

Microplastic Number Density vs. Waste Emitted (with Model)

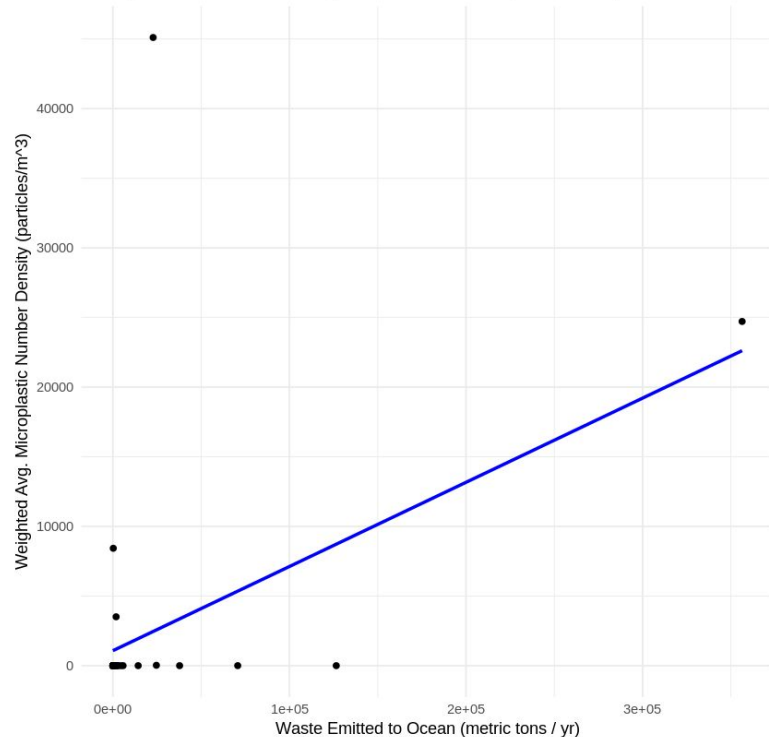


Logged Microplastic Number Density vs. Waste Emitted (with Model)

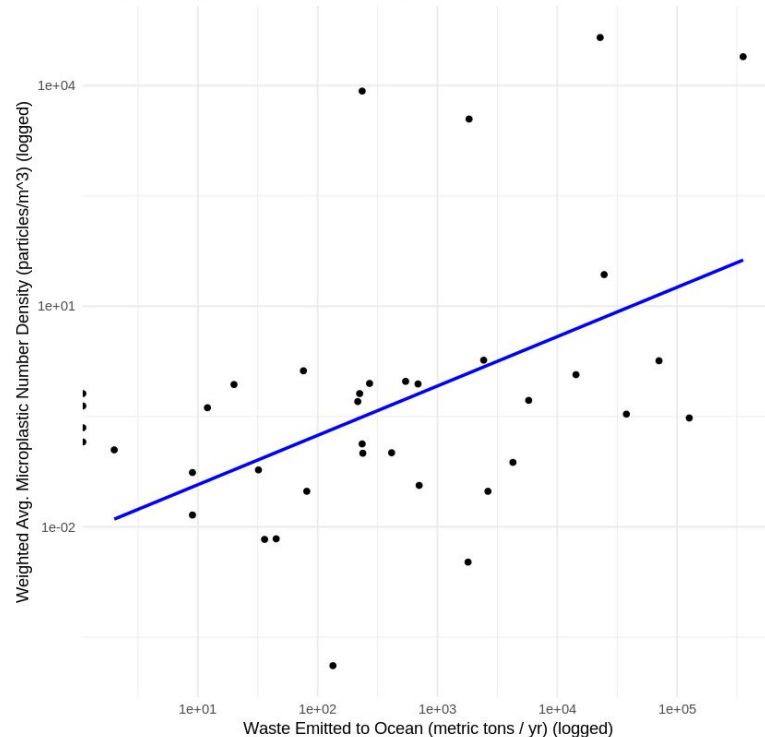


Current Model:

Microplastic Number Density vs. Waste Emitted (with Model)



Logged Microplastic Number Density vs. Waste Emitted (with Model)



Current Problems

- Only kind of independent
- Different sampling methods don't seem identically distributed
- residuals do not seem homoskedastic —> try logged version next
- after cleaning, low # of data points

Suggestions?