Mismanaged Plastic Waste in 2010 and 2019

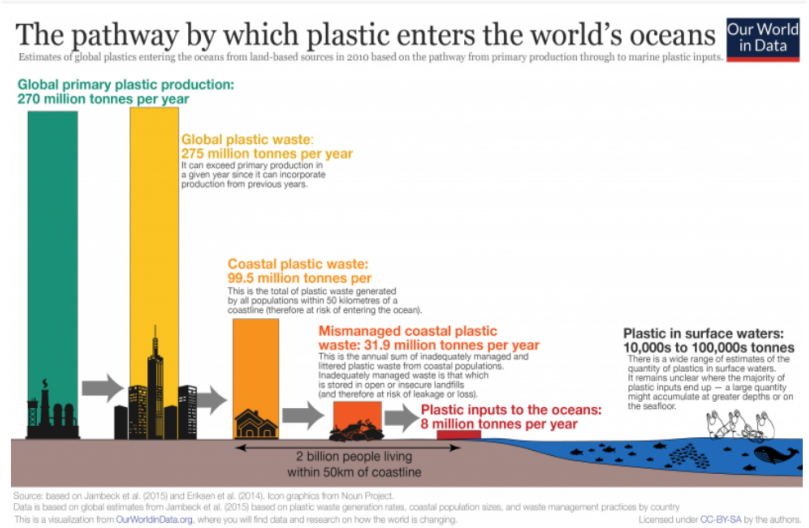
Team 10

Connor Frunzi, John Hall, Moriah Long, Mohammad Naeem

Introduction

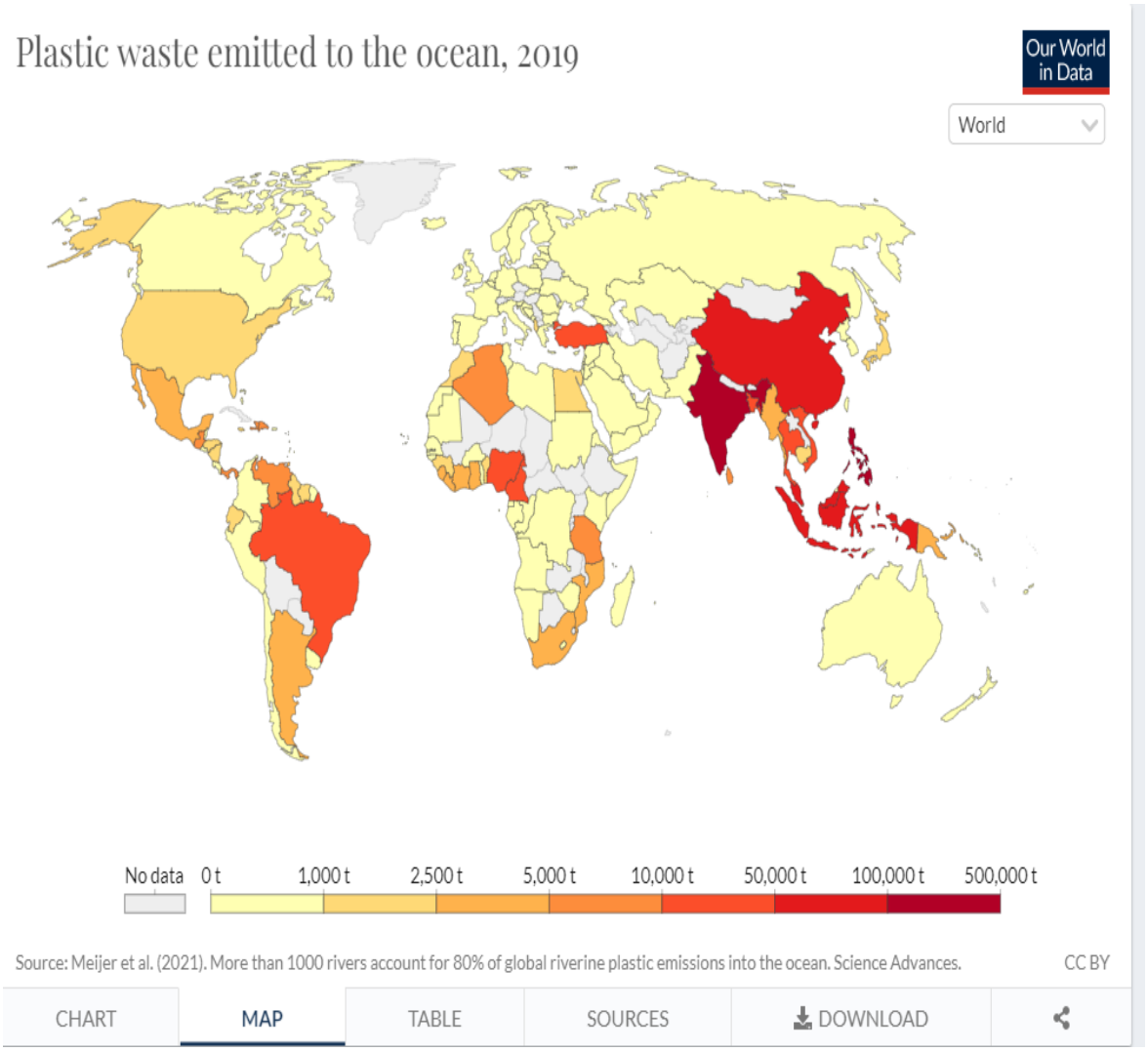
11 million metric tons of plastic leak into the oceans every day. According to the World Wide Fund for Nature (WWF), by 2050, there could be more plastic weight in the oceans than fish. This issue is only fueled by the public's ignorance of the problems in their own countries. There is not a clear representation of where mismanaged plastic is ending up in the world, and in what quantities (Hancock, 2019).

Mismanaged waste is the sum of material that is either littered or inadequately discarded. Inadequately discarded waste has the intention of being managed through waste collection or storage sites. However, it is ultimately not formally or sufficiently managed by countries and corporations across the globe. Disposal in dumps or open, uncontrolled landfills means the material is minimally contained. This means the waste will leech into the surrounding environment. This increases the risk of leakage and conveyance to the natural environment and oceans through waterways, winds and tides (Ritchie and Roser, 2018).



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

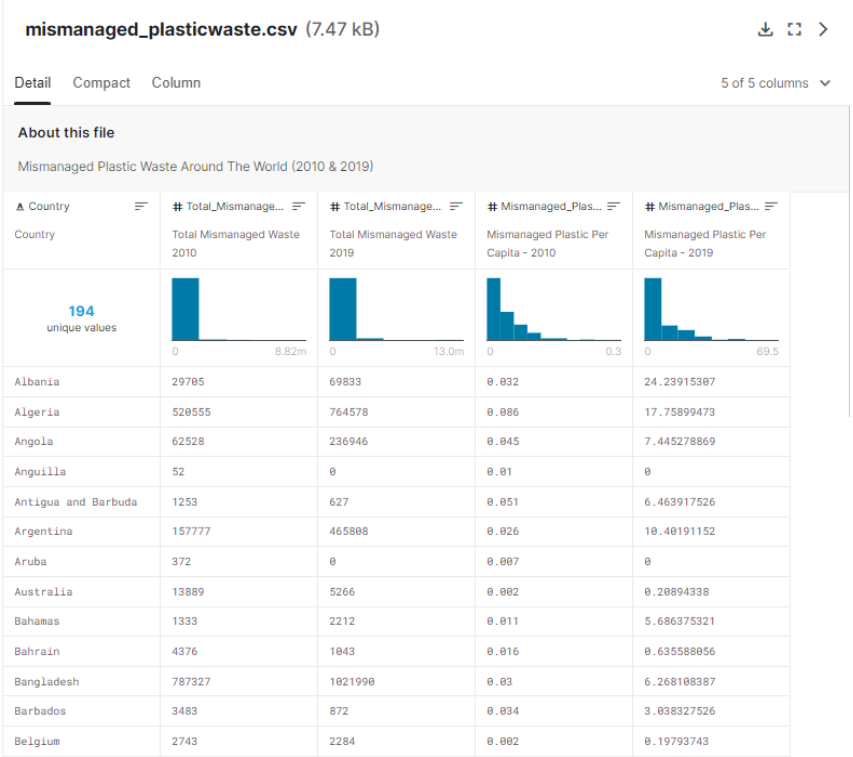
Figure 2.



Direct link:
https://ourworldindata.org/grapher/plastic-waste-emitted-to-the-ocean?country=BGD~CMR~PHL~CHN~IND~NGA~BRA~LKA

Describing Dataset

We found a dataset on Kaggle called “Mismanaged plastic waste around the world: Mismanaged plastic waste (2010 & 2019)”. This dataset contained 5 fields - the country name, total plastic waste in millions of tons 2010, total plastic waste in millions of tons 2019, plastic waste per person in kg 2010, and plastic waste per person in kg 2019. Using this data, we were able to create a series of choropleth maps for users to quickly see how certain countries were progressing, both at a total waste level and at a per capita level. The charts below show visual descriptions of the dataset, including one of our choropleths demonstrating our interactive tooltip.



mismanaged_plasticwaste.csv (7.47 kB)

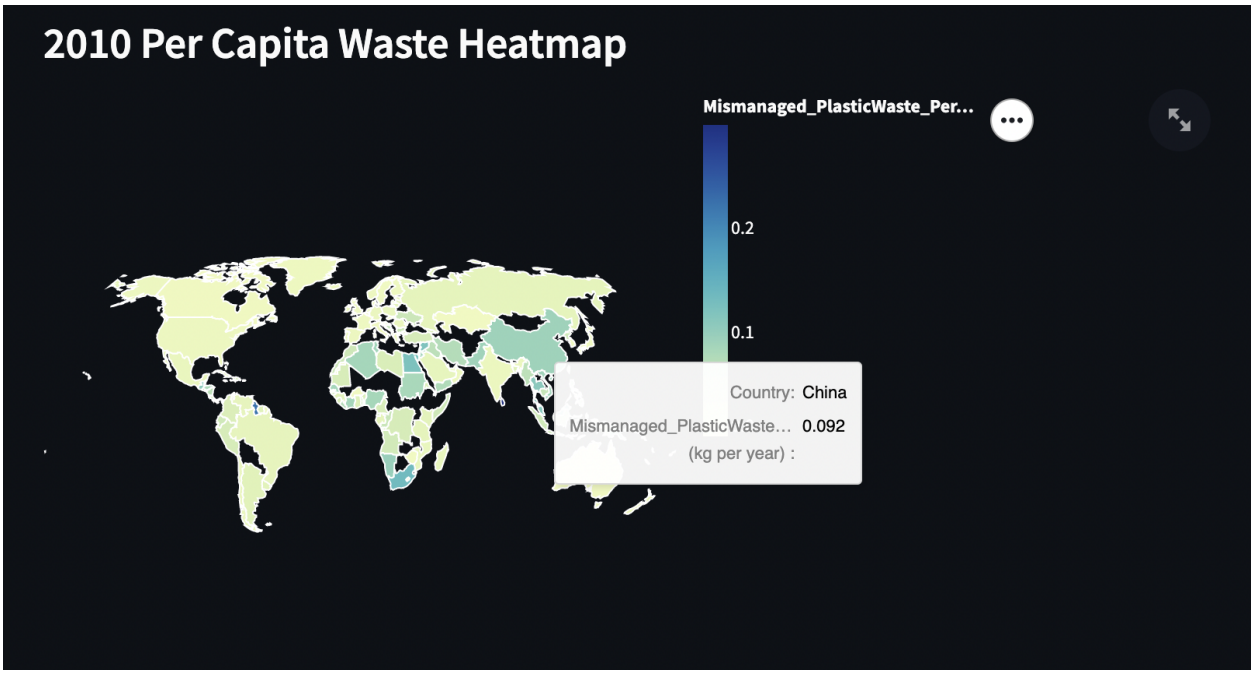
Detail Compact Column 5 of 5 columns

About this file

Mismanaged Plastic Waste Around The World (2010 & 2019)

Country Country	# Total_Mismanage... Total Mismanaged Waste 2010	# Total_Mismanage... Total Mismanaged Waste 2019	# Mismanaged_Plas... Mismanaged Plastic Per Capita - 2010	# Mismanaged_Plas... Mismanaged Plastic Per Capita - 2019
194 unique values	0 8.82m	0 13.0m	0 0.3	0 69.5
Albania	29705	69833	0.032	24.23915307
Algeria	520555	764578	0.086	17.75899473
Angola	62528	236946	0.045	7.445278869
Anguilla	52	0	0.01	0
Antigua and Barbuda	1253	627	0.051	6.463917526
Argentina	157777	465808	0.026	10.40191152
Aruba	372	0	0.007	0
Australia	13889	5266	0.002	0.20894338
Bahamas	1333	2212	0.011	5.686375321
Bahrain	4376	1043	0.016	0.635588056
Bangladesh	787327	1021990	0.03	6.268108387
Barbados	3483	872	0.034	3.038327526
Belgium	2743	2284	0.002	0.19793743

Figure 4.



Data Preprocessing

Data preprocessing was performed in Jupyter notebook with pandas. Preprocessing code can be found in our project files.

Our primary dataset was mostly complete. Although there were a couple data points that were 0, both 0 and null show as white on our choropleth, so we decided to leave this data in to avoid having to handle null values in our code. The primary preprocessing effort was connecting country names in our primary dataset with country codes from one of the helpers.

Firstly, country names were not necessarily the same between the two datasets, so the second dataset was changed to match the primary. The primary dataset did not

have accent markings, so many countries had to be corrected such as “Curaçao” to “Curacao”. Some country names were incomplete in the primary dataset but fully written in the secondary, so “The United Kingdom of Great Britain and Northern Ireland” had to be rewritten as “United Kingdom”. Other countries were formatted strangely - for example, what was “Congo” in the primary dataset was “Congo, Democratic Republic of the” in the secondary dataset. After correcting these many issues to have matching datasets, we were able to use the second dataset to append a column of country codes to the first so we could accurately plot countries on a choropleth.

Processing

After our data was cleaned, we uploaded the pre-processed dataset to an S3 bucket so it would be easily accessible to everyone in the group to access and work with. We wanted to use Altair charts to create our visualizations, so we decided it would be wise to have a way to collaborate on these charts in real time. Then, we used EC2 to host a Streamlit app so we could quickly deploy and collaborate on our Altair charts. This EC2 instance used the data stored in the S3 instance. These collaborative uses of cloud services allowed us to quickly and effectively analyze the data.

After finalizing our visualizations and dashboard, we realized that users of a site like we had just created would probably be interested in getting a more detailed look at the data. For this, we added an option to receive additional information via email. When an email is entered in the box and the “Send Email” button is clicked, a request is sent to an AWS Lambda Function which reads the email from the request body then uses AWS Simple Email Service (SES) to send an email via my school email address

(jhall170@uncc.edu) including our pre-processed data. In practice, this would likely be updated to include a pamphlet or more information about plastic waste, but we used the data for now as a proof of concept.

Additional Screenshots

Our dataset in Jupyter notebook:

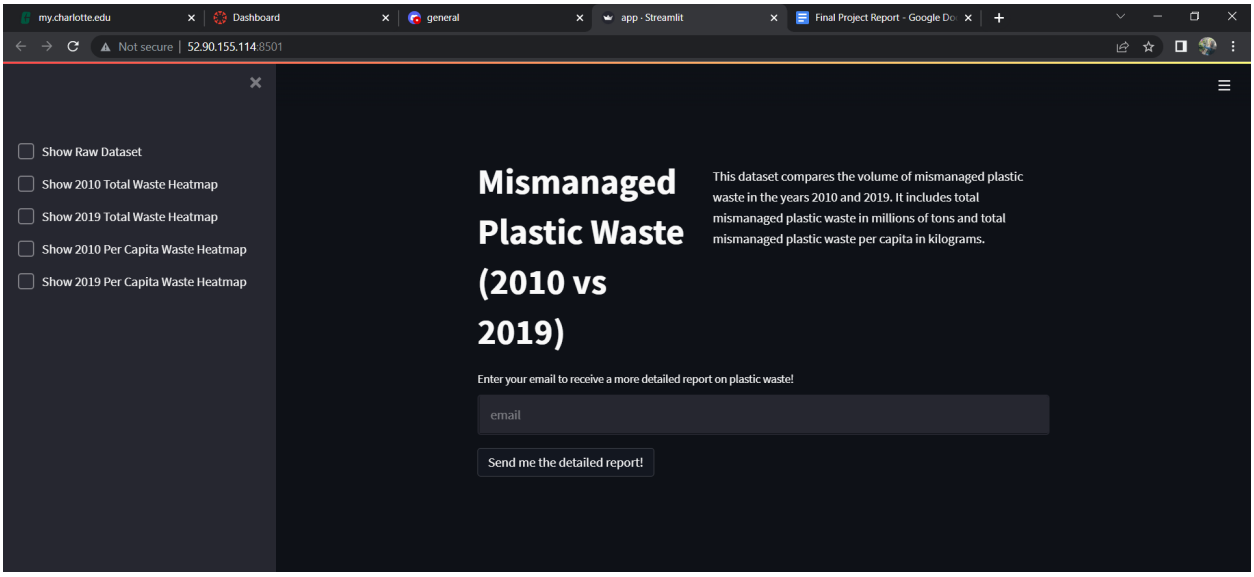
```
In [16]: data = gpd.read_file("mismanaged_plasticwaste.csv")

In [17]: data
```

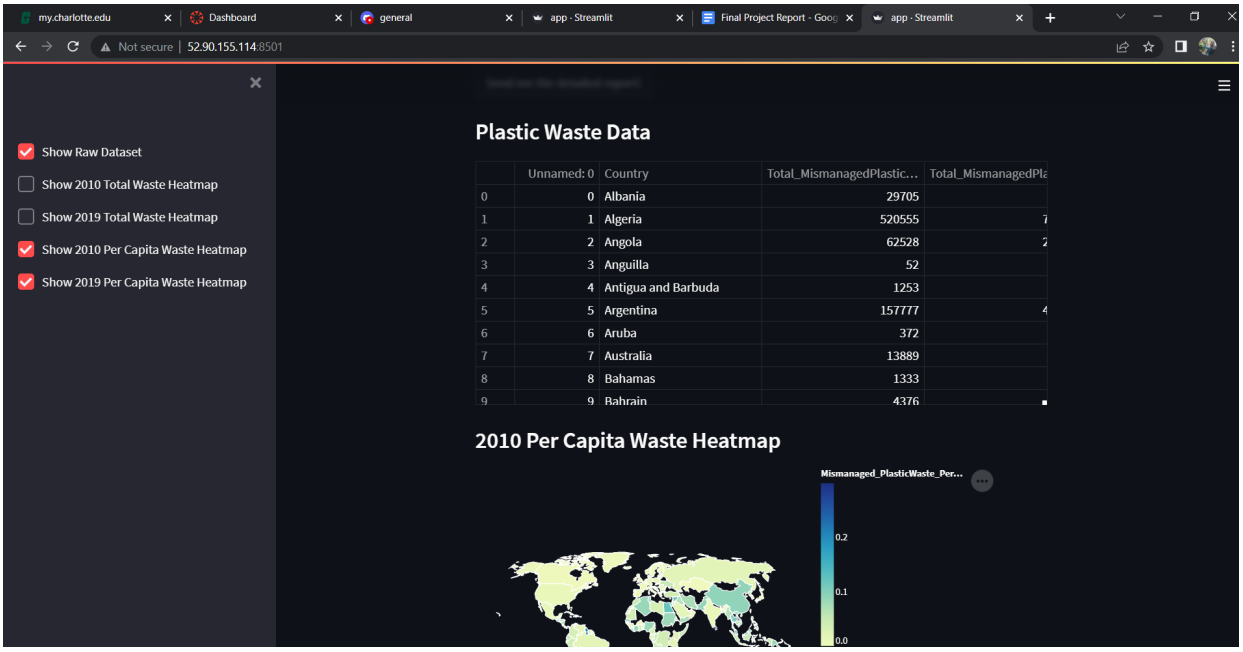
Out[17]:

	Country	Total_MismanagedPlasticWaste_2010 (millionT)	Total_MismanagedPlasticWaste_2019 (millionT)	Mismanaged_PlasticWaste_PerCapita_2010 (kg per year)	Mismanaged_PlasticWaste_Pe
0	Albania	29705	69833	0.032	
1	Algeria	520555	764578	0.086	
2	Angola	62528	236946	0.045	
3	Anguilla	52	0	0.01	
4	Antigua and Barbuda	1253	627	0.051	
...	...	...	...	...	
189	Venezuela	102333	671431	0.017	
190	Vietnam	1833819	1112790	0.09	
191	Western Sahara	0	4114	0	
192	Yemen	169181	291737	0.077	
193	Zimbabwe	0	524865	0	

194 rows × 6 columns

Some screenshots of our final Streamlit dashboard, hosted on EC2 and using data hosted on S3:





Trial and Errors

Our first error was attempting to map to the choropleth using the numeric index assigned to rows for all countries in alphabetical order given by pandas. We assumed country IDs might just line up with their index when listed alphabetically. However, this

led to wildly misplaced countries and many countries left off the map entirely. We then discovered there was an international standard for numeric IDs for countries, so we found a csv listing country names and their respective IDs and used that to add an ID column to our dataset.

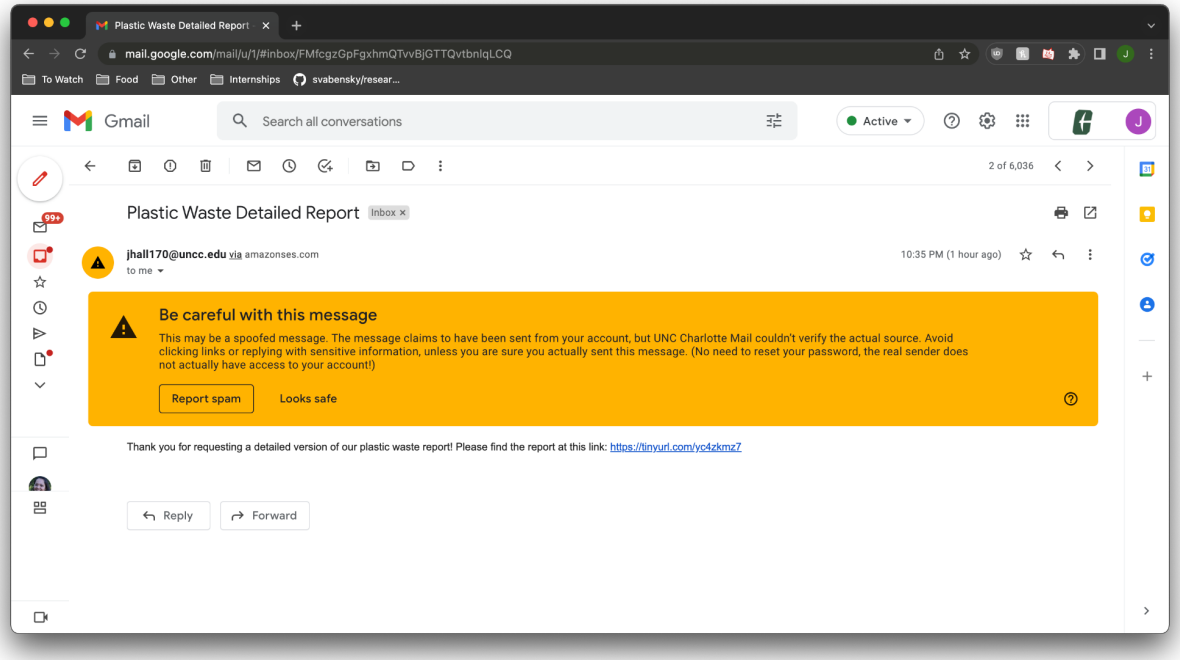
In this second dataset, we found another issue. As mentioned earlier, countries were not named with the same conventions between our main dataset and the ID list. We had to manually rename many countries to make the two datasets line up.

At this point we had some issues getting all of the data into Altair charts to be visualized, but were eventually able to figure out the unfamiliar syntax and create powerful visualizations.

We also had a few issues while attempting to create our additional email service. At first, we could not figure out how to attach a file to the email. We eventually realized it would require additional package imports and learning to use a new Python library, so we decided the best way to provide our file via email would be by including a URL. We had to open up our S3 privacy settings to allow public access so users could click a link directing to the file in the S3 bucket and download the file.

Then, when we went to test the email service with more emails than just my own (jhall170@uncc.edu), we realized that it only worked for my email and not for other members of the group. While at first we did not know how to tackle this issue, we soon made use of AWS Cloudwatch to identify the issue. Even though SES was receiving the correct emails and attempting to send emails to them, my AWS account is in Sandbox mode, meaning SES can only email to addresses that have been verified in AWS.

I have requested that my account be removed from Sandbox mode, so hopefully my account will be upgraded and this feature will be working for all users by the time this assignment is graded, but if not, this is a screenshot of the email it sends:



Conclusion

Mismanaged plastic waste is a major problem plaguing modern (and future) society. Instead of reducing plastic waste, many countries got worse throughout the 10 year period. We made use of 5 AWS services (EC2, S3, Lambda, CloudWatch, and SES) to create an interactive visualization to both learn from the dataset ourselves and allow others to learn quickly and easily as well. Our goal in this report was to raise awareness through data as clearly as possible, in a way that users can easily share with others.

Works Cited

https://www.kaggle.com/datasets/kkhandekar/mismanaged-plastic-waste-around-the-world

https://www.kaggle.com/code/sasakitetsuya/mismanaged-plastic-waste-analysis

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

Lorin Hancock (Fall 2019), -“Our plastic problem”. Retrieved from”https://www.worldwildlife.org/magazine/issues/fall-2019/articles/plastic-in-the-ocean
PHOTOGRAPHER: Eduardo Leal