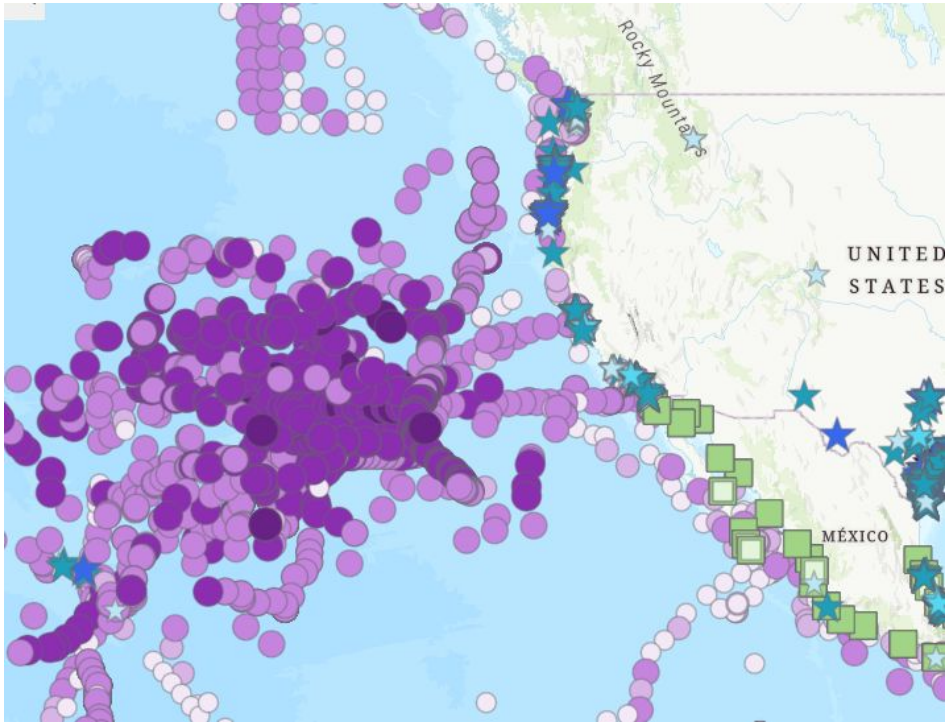
Marine Microplastics

Jaize Miller - AD450

Data Sourcing

- NOAA's growing dataset for microplastic concentrations globally
- Public access in csv format

National Centers for Environmental Information
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION



Raw Data

- 4700 entries (a subset of the available dataset > 22,000)
- 22 columns

OBJECTID	Oceans	Regions	SubRegi	Sampling N	Measurem	Unit	Density Ra	Density Cla	Short Refe	Long Refer	DOI	Organizati	Keywords	Accession	Accession	Latitude	Longitude	Date
11407	Pacific Ocean			Neuston n	0.02557	pieces/m3	0.005-1	Medium	Eriksen et a	Eriksen, M.	https://doi	5 Gyres Ins	AVANI Net	275967	https://ww	-27.0131	-107.574	########
6830	Pacific Ocean			Neuston n	0	pieces/m3	0-0.0005	Very Low	Law et al.2	Law, K.L, S	https://doi	Sea Educa	SEA	211008	https://ww	5.79	-147.72	########
6498	Pacific Ocean			Neuston n	0	pieces/m3	0-0.0005	Very Low	Law et al.2	Law, K.L, S	https://doi	Sea Educa	SEA	211008	https://ww	9.55	-146.63	########
7494	Pacific Ocean			Neuston n	0	pieces/m3	0-0.0005	Very Low	Law et al.2	Law, K.L, S	https://doi	Sea Educa	SEA	211008	https://ww	19.24	-143.41	########
8087	Pacific Ocean			Neuston n	0	pieces/m3	0-0.0005	Very Low	Law et al.2	Law, K.L, S	https://doi	Sea Educa	SEA	211008	https://ww	23.52	-141.63	########
7091	Pacific Ocean			Neuston n	0	pieces/m3	0-0.0005	Very Low	Law et al.2	Law, K.L, S	https://doi	Sea Educa	SEA	211008	https://ww	21.15	-142.77	########
12664	Pacific Ocean			Manta net	0.953208	pieces/m3	0.005-1	Medium	Egger et al.	Egger, M.,	https://doi	The Ocean	The Ocean	276264	https://ww	33.525	-151.117	########
7423	Pacific Ocean			Neuston n	0.043196	pieces/m3	0.005-1	Medium	Law et al.2	Law, K.L, S	https://doi	Sea Educa	SEA	211008	https://ww	11.09	-156.77	########
8419	Pacific Ocean			Neuston n	0.018584	pieces/m3	0.005-1	Medium	Law et al.2	Law, K.L, S	https://doi	Sea Educa	SEA	211008	https://ww	13.69	-156.87	########
6166	Pacific Ocean			Neuston n	0	pieces/m3	0-0.0005	Very Low	Law et al.2	Law, K.L, S	https://doi	Sea Educa	SEA	211008	https://ww	11.33	-93.89	########

Data Types

- Oceans = Pacific (filtered on NOAA export)
- Sampling Method = string
- Measurement = float, particles/m^3
- x, y = floats for latitude and longitude
- Date = string in raw data

0	OBJECTID	4700 non-null	int64
1	Oceans	4700 non-null	object
2	Regions	278 non-null	object
3	Sampling Method	4700 non-null	object
4	Measurement	4334 non-null	float64
5	Unit	4700 non-null	object
6	Density Range	4700 non-null	object
7	Density Class	4700 non-null	object
8	Short Reference	4700 non-null	object
9	Long Reference	4700 non-null	object
10	DOI	4700 non-null	object
11	Organization	4700 non-null	object
12	Keywords	4700 non-null	object
13	Accession Number	4700 non-null	int64
14	Accession Link	4700 non-null	object
15	Latitude	4700 non-null	float64
16	Longitude	4700 non-null	float64
17	Date	4700 non-null	object
18	GlobalID	4700 non-null	object

Data Preparation

- Remove empty columns
- Drop rows with null values
- Standardize units
- Normalize numerical data

```
# standardize units
units = cleaned['Unit'].unique()
print(units)

# convert date format
cleaned['Date'] = pd.to_datetime(cleaned['Date'])
print(cleaned['Date'])

#convert object id to int
cleaned['OBJECTID'] = cleaned['OBJECTID'].astype(int)

# inspect method values
methods = cleaned['Sampling Method'].unique()
print(methods)
```

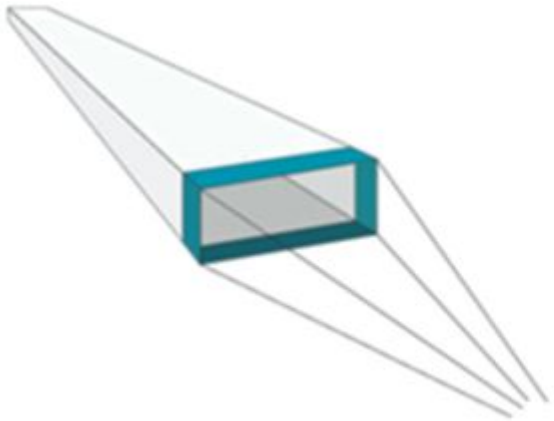
```
# Normalize data for measurement column
numerical_cleaned = cleaned.select_dtypes(include=['float64', 'int64'])

target = numerical_cleaned['Measurement']
numerical_cleaned['m_norm'] = (target - target.mean()) / target.std()
print(numerical_cleaned)
```

Research Questions

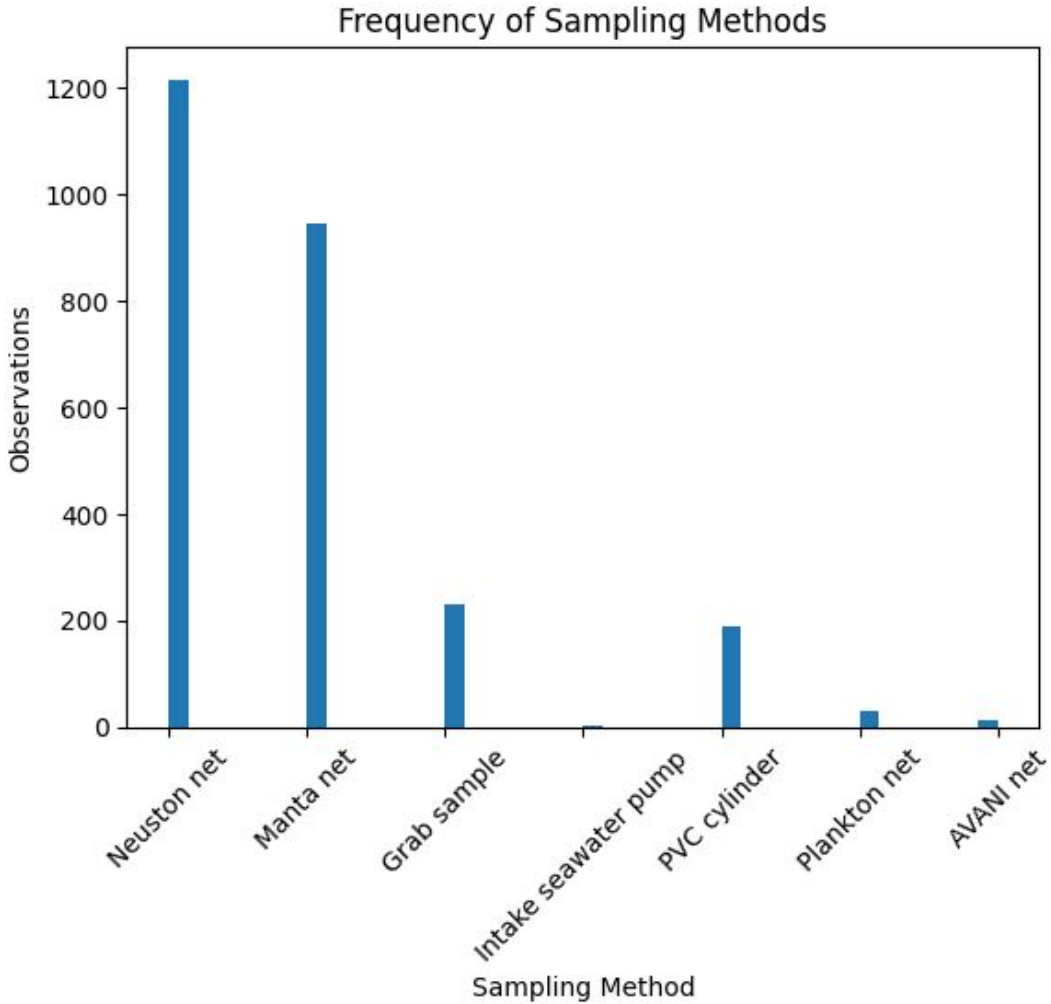
1. Where are the densest samplings by longitude and latitude?
2. What is the most used sampling method?

Sampling Methods

- Neuston net is most common

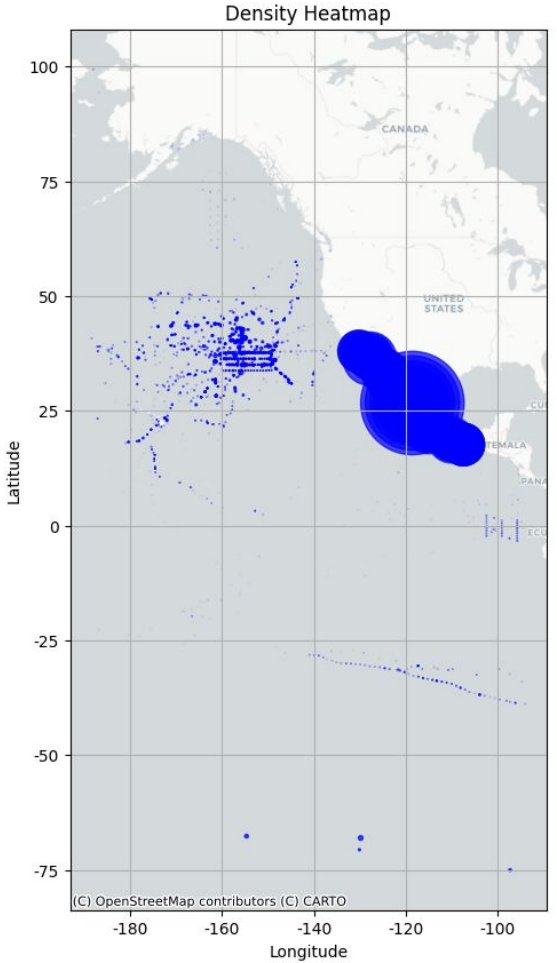


Neuston net

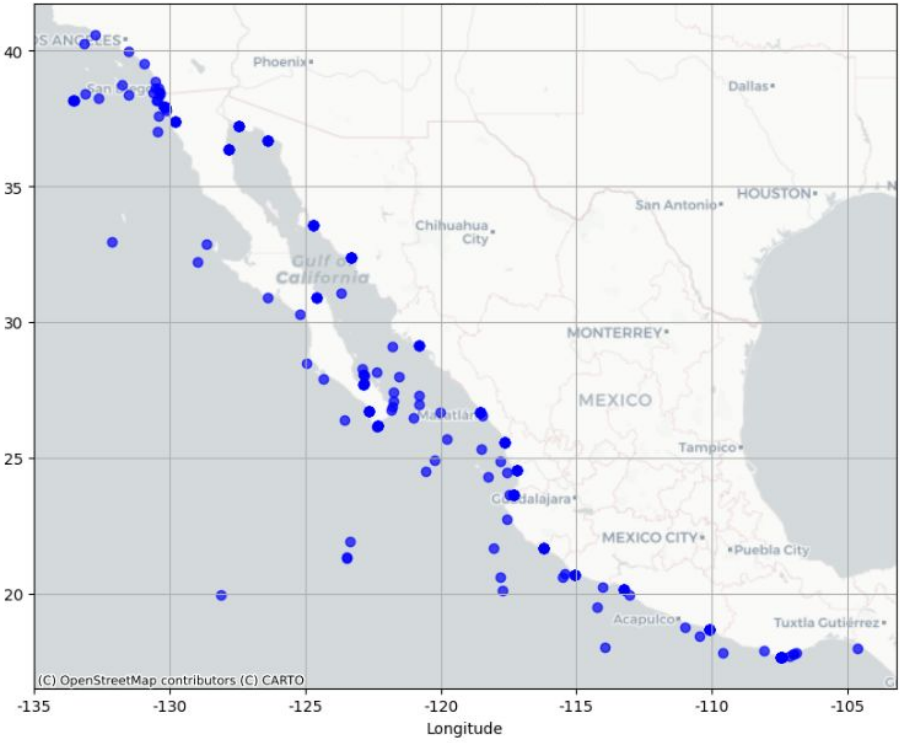


Mapping Density

- With geopandas and contextily
- Highest concentration of samples obtained along Pacific coast of Mexico



Sampling Locations



Scaled by Measurement Density

Investigating Correlation

- Using Grid method (lumps observations into bins dictated by grid location)
- Results in a correlation coefficient of 0.167, indicating little to no correlation

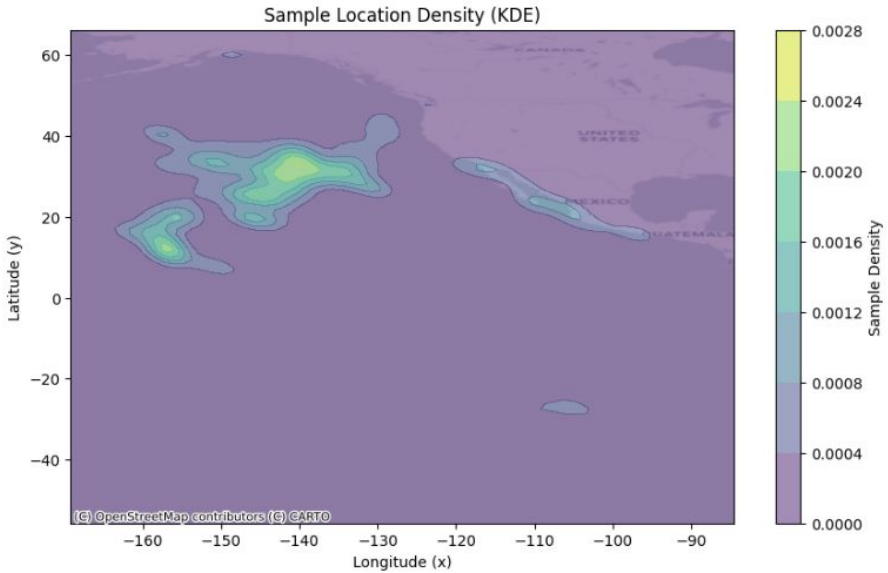
```
# look at sample location...
grid_size = 1
merged['x_grid'] = (merged['x'] // grid_size).astype(int)
merged['y_grid'] = (merged['y'] // grid_size).astype(int)

sample_density = merged.groupby(['x_grid', 'y_grid']).size().reset_index(name='sample_density')

measurement_density = merged.groupby(['x_grid', 'y_grid'])['Measurement'].mean().reset_index(name='measurement_density')

density = pd.merge(sample_density, measurement_density, on=['x_grid', 'y_grid'])

# calculate correlation between sample density and measurement density
correlation = density[['sample_density', 'measurement_density']].corr().iloc[0,1]
print(correlation)
```

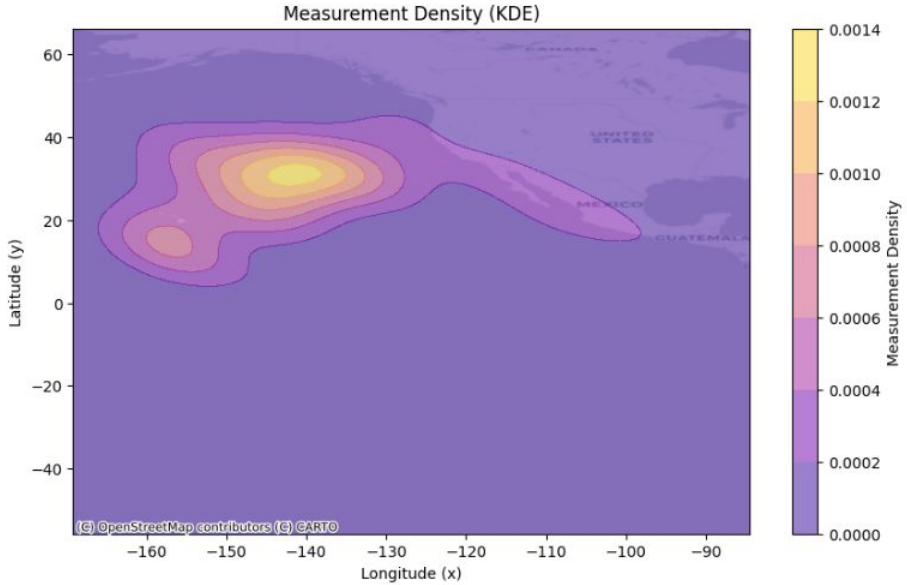
Investigating Correlation (cont.)

- Geospatial data is better modeled with the KDE (Continuous Density Estimation)
  - Distributes density continuously – smoothing data
  - More sensitive to actual data distribution

KDE Results

- Correlation between sample density and measurement density = 0.88
- Indicating a HIGH correlation, contradictory to the grid method results





Conclusion

- Findings are corroborative with trash accumulation in the Pacific
- Tidal currents lead to the coastal build up observed along Mexico shoreline