**Problem Statement**

**Organization Name:** Booz Allen Hamilton

**Dataset Name:** Ocean Cleanup

**A picture containing transport

Description automatically generatedDifficulty:** (

Level 5: Participants with advanced data analysis knowledge and skills.

The problem statement includes requirements of modeling. The dataset has a complex structure, numerous variables of interest, and spatial-temporal dimensions.

# Background

# We care about keeping our oceans clean because they produce over half the world’s oxygen and absorb 50 times more carbon dioxide than our atmosphere. The ocean transports heat from the equator to the poles, regulating our climate and weather patterns. How many resources do we need to clean up the oceans? Where do we optimize resource allocation?

# Deliverables

* Time series of cleanup volume. Think “how active are people in cleanup?”
* Groups of garbage items that typically go together
* Garbage density per area of cleanup location (have an average weight for each item and get the water perimeter of cleanup location)
* EDA on most popular garbage items by clean-up type and location
* An interactive map of the world that displays volume of garbage by state and zone

# Data Considerations

TIDES is a public data system containing the world's largest ocean trash dataset, all collected by volunteers. Filtered to U.S. data only, including variables such as zone, clean-up type, clean-up date, pounds of trash, and all the different types of trash collected. The dataset includes 37,904 rows or records with 61 columns or variables.