# Piracy Probabilities:
# The Curse of the Modern Mariner

Aaron Pollard
Matt Demond
Aisha Kamara
Anthony Zurke

September 13th, 2021

# Introduction

Wikipedia defines **piracy** as “an act of robbery or criminal violence by ship or boat-borne attackers upon another ship or a coastal area, typically with the goal of stealing cargo and other valuable goods.” When we think of maritime piracy it can often conjure up thoughts of wooden ships and eye patches, but it is still a problem that plagues us today.

# What types of pirate attacks exist?

- Pirate Assaults
- Suspicious Approach
- Kidnapping
- Attempted Boarding
- Unknown
- Hijacking
- Other

# Problem Statement

Identify causal societal and developmental factors related to piracy. Using these factors we aim to create an effective model to prove that these factors are strong determinants of high pirate activity within over 130 countries 2010 onwards.

# Why is understanding this important?

It is important to understand the societal and economic factors of piracy to redress causes of piracy and create a safer and more equitable world.

# Anti-Shipping Activity Data

- The data was obtained from the National Geospatial-Intelligence Agency. We initially obtained almost 8,000 observations of reported hostile acts against ships and mariners (pirate attacks) that ranged from 1978 to present.

# United Nation Sustainable Development Goals Data

- **Goal 1: End poverty in all its forms everywhere**
  - Target 1.1.1
- **Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all**
  - Target 4.1.1
- **Goal 9: Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation**
  - Target 9.5.1
  - Target 9.5.2
- **Goal 14: Conserve and sustainably use the oceans, seas and marine resources for sustainable development**
  - Target 14.c.1

# Cleaning Data Process

**Data Cleaning Steps:**

1. Our project is more concerned with the effects of modern piracy, and as regions, economic factors and technologies would experience a great of deal of change over the course of that time period we focused only on attacks from 2010 to the present day, leaving us with approximately 3,500 observations to analyze.

2. We performed the usual sanity checks of identifying null values, missing values, and dropped any features we deemed not relevant to our overall analysis.

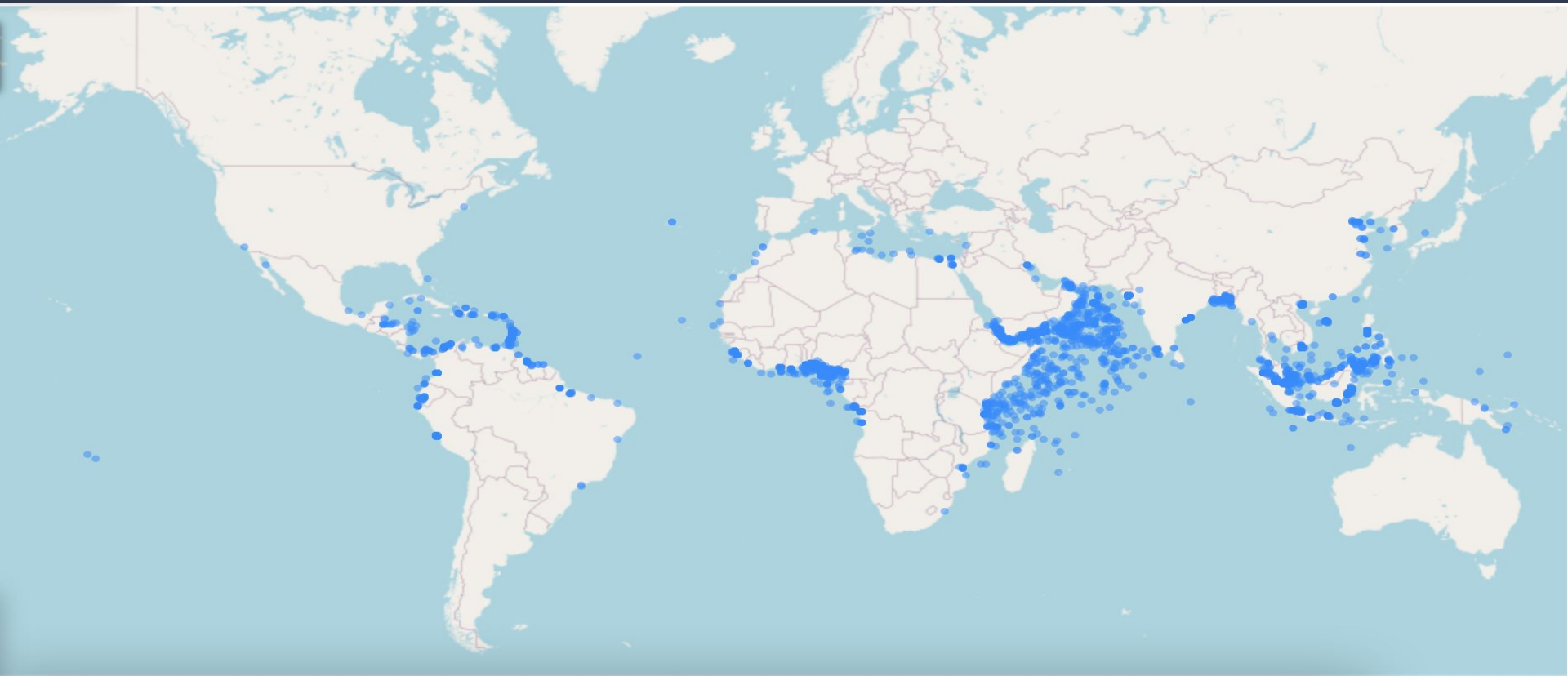
3. We merged the different data sets, using the nominatim module from a python package called geopy.

# Exploratory Data Analysis

- Occurrence of pirate attacks by region
- Pirate Attack rates over time
- What times of the year is piracy most prevalent
- What types of vessels are attacked
- Where are non-commercial shipping vessels most likely to be targeted

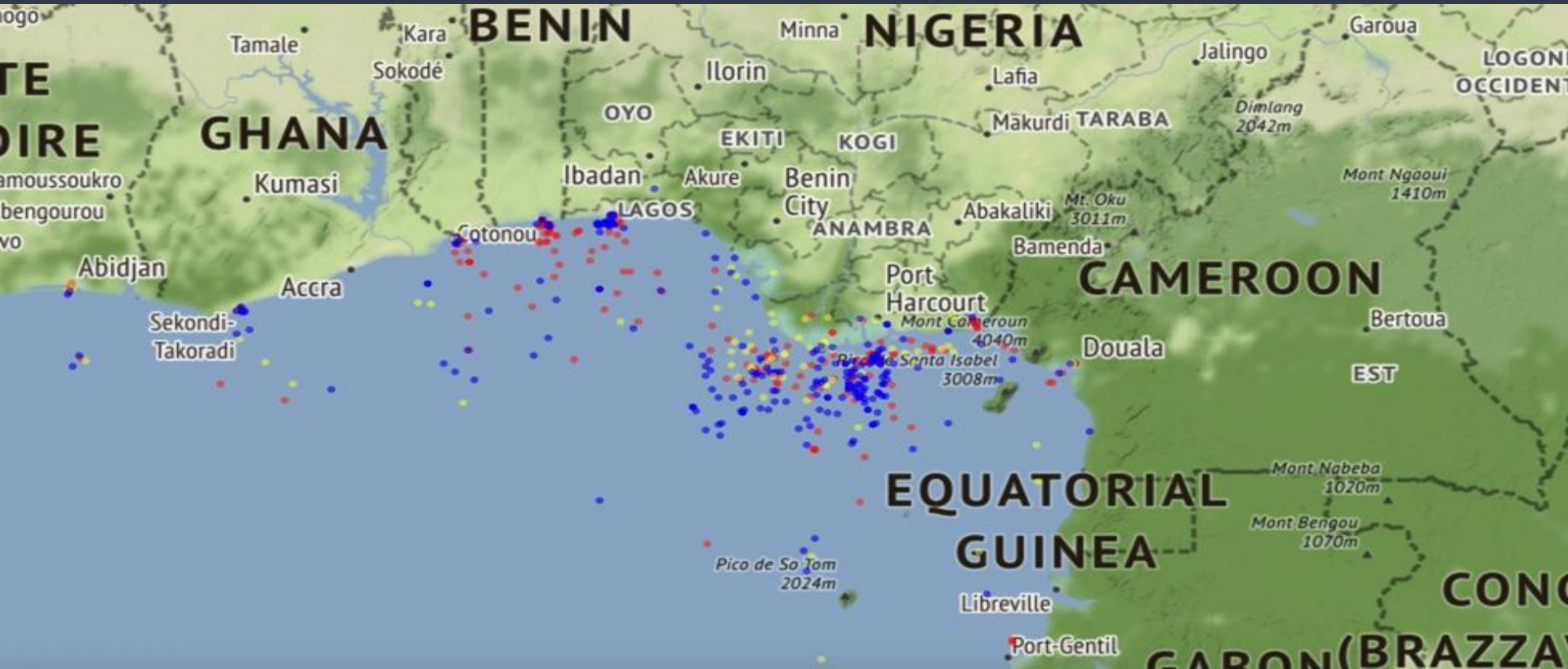
# Where Did Pirate Attacks Occur?





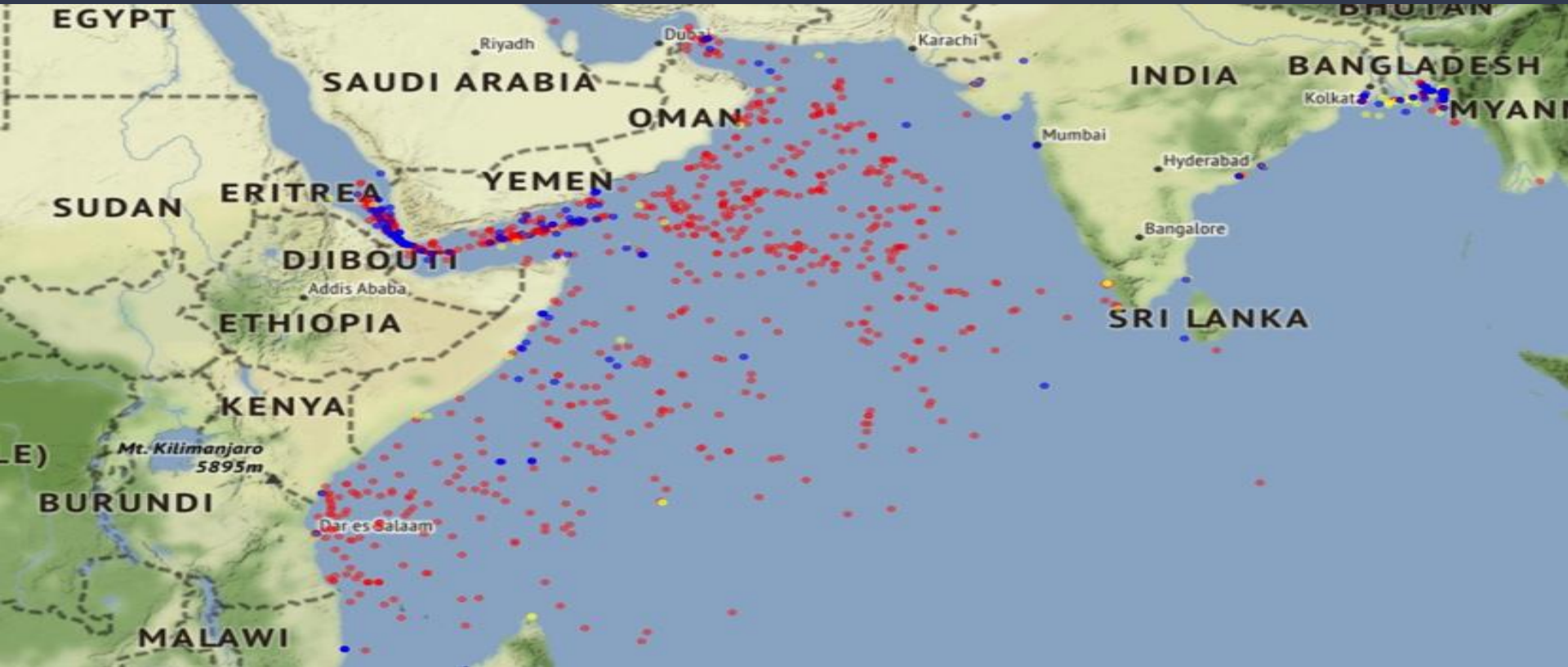
# Where Did Pirate Attacks Occur?





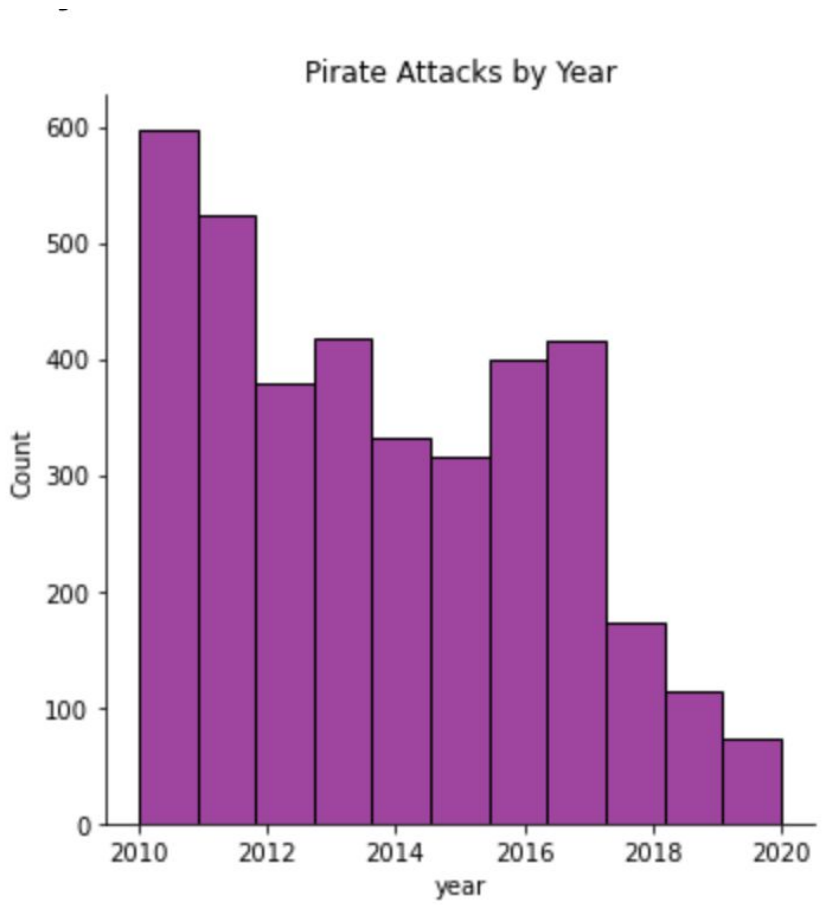
# Where Did Pirate Attacks Occur?



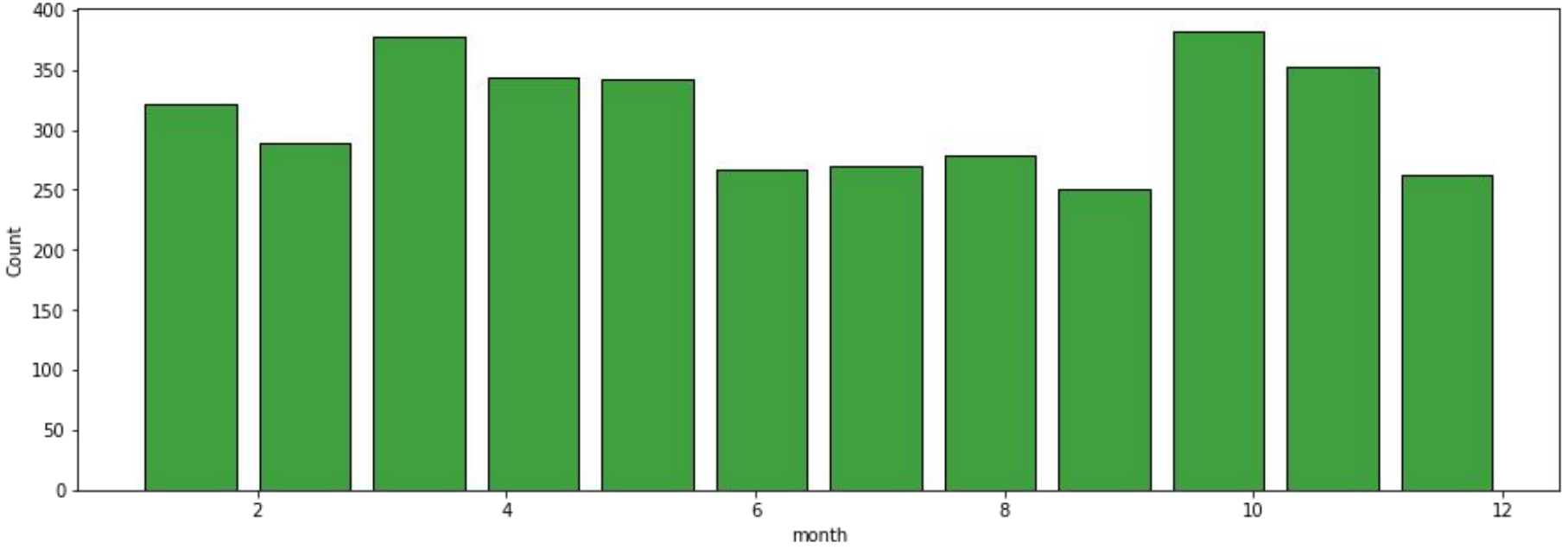


# Where Did Pirate Attacks Occur?

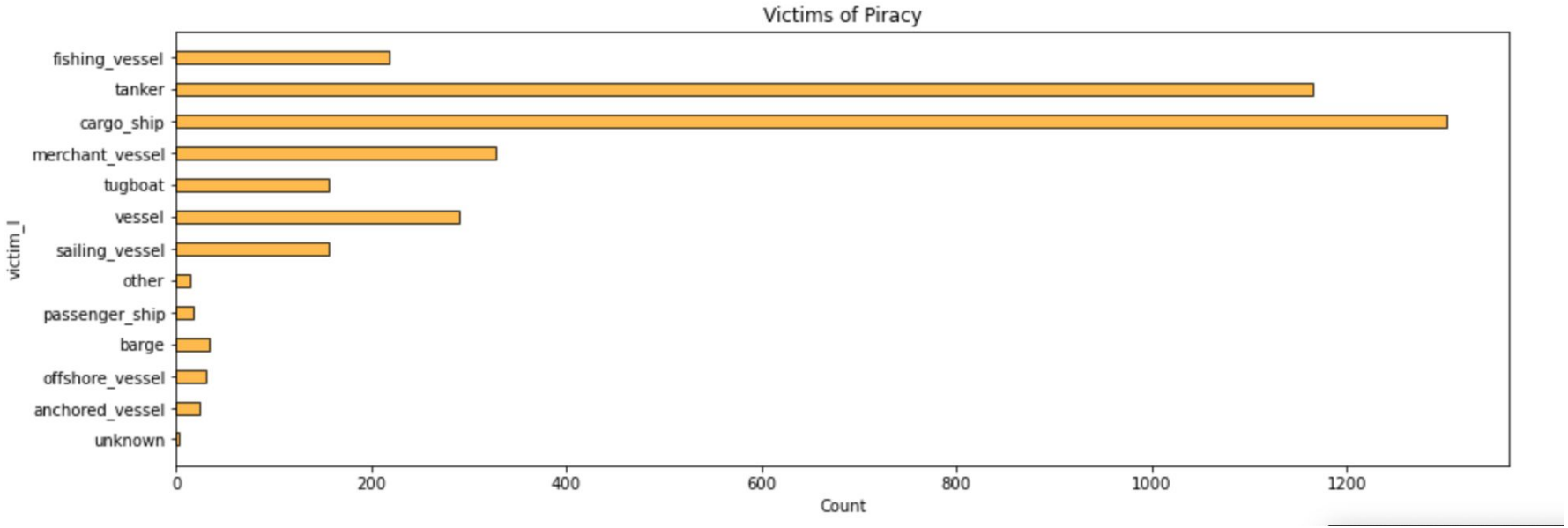
# Pirate Attacks Per Year



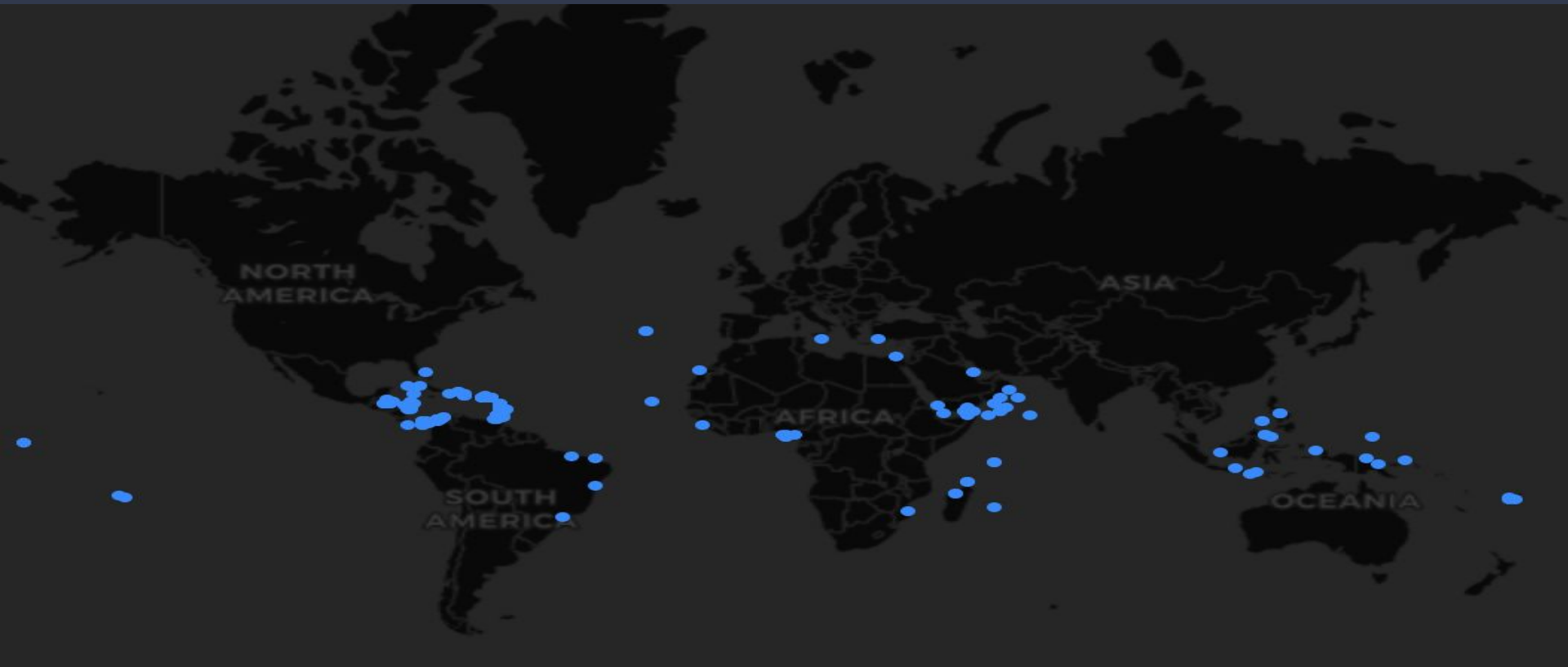
# Pirate Attacks Per Month



# What Types Of Vessels Are Susceptible To Attacks?

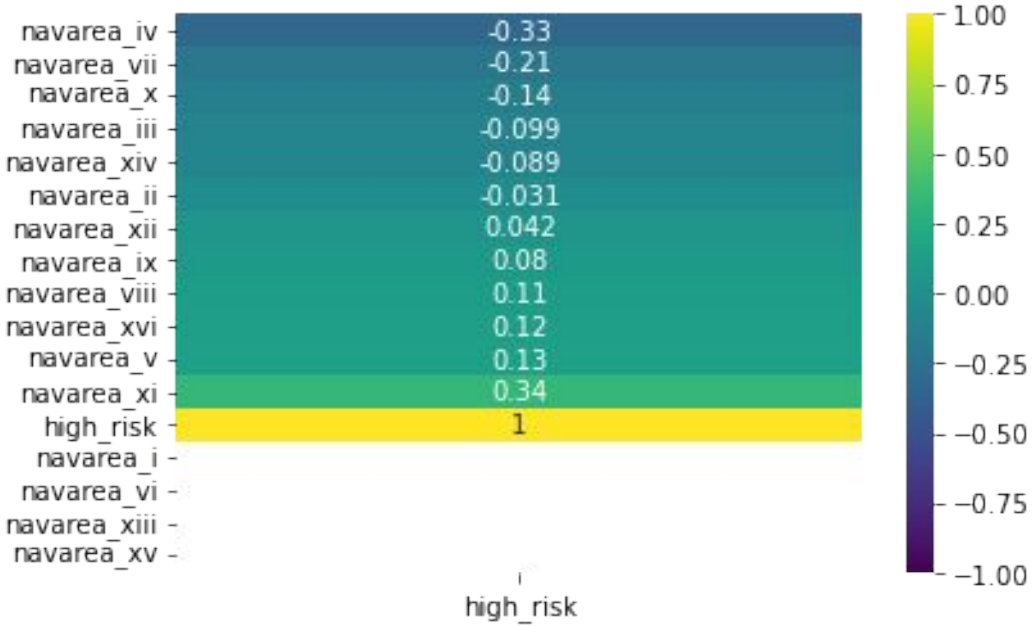


# Where are non-commercial shipping vessels most likely to be targeted?



# Correlation: Location vs. Society

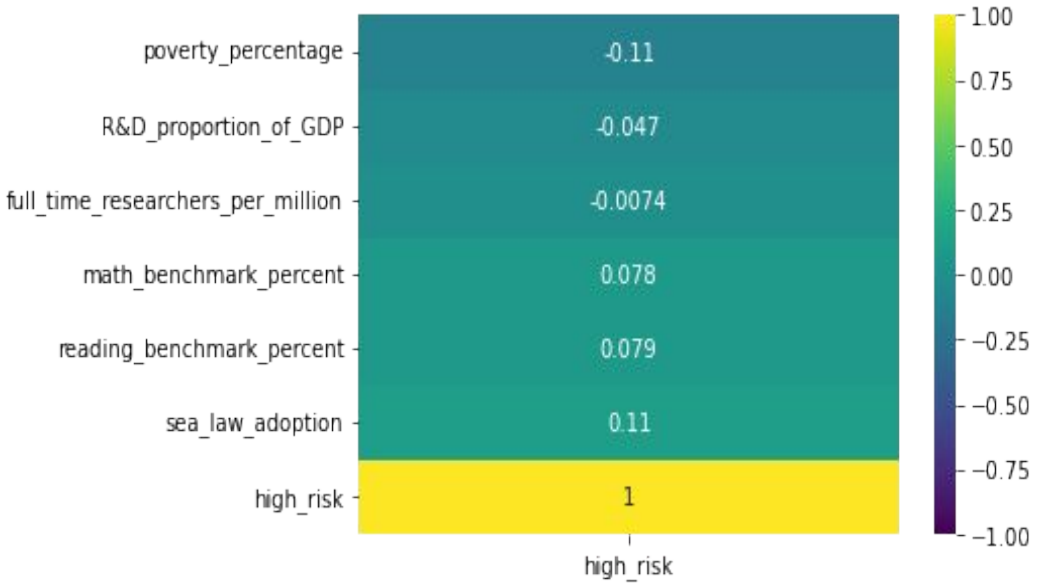





- None of the features had a strong correlation (>0.5) to high pirate activity.
- The navigational area within the Indo-China sea had the highest correlation of 0.34.

# Correlation: Location vs. Society





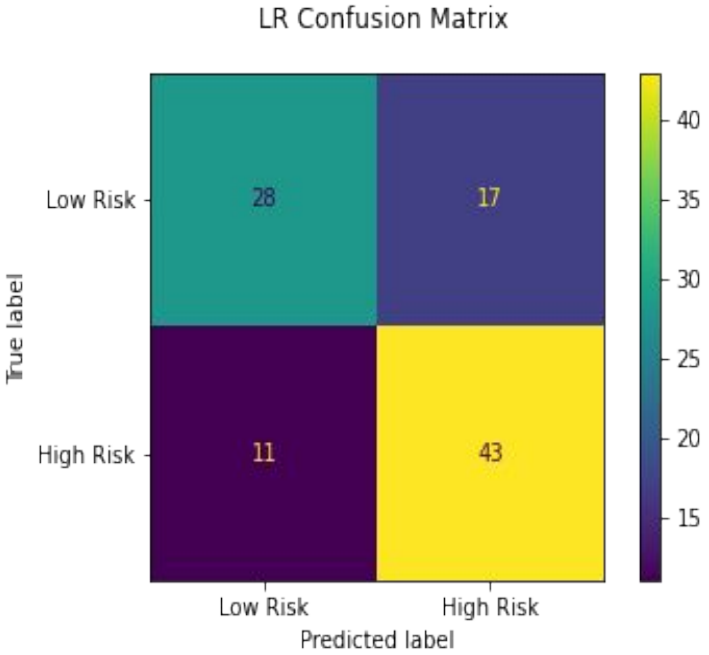

- There aren't strong correlations across individual features.
- Poverty reduction (SDG goal 1), and the adoption of oceanic treaties(SDG goal 14) had the "strongest"correlation to high pirate activity.
- Indicators of education and research showed the weakest correlation

# Modeling

**Classification Models Used:**

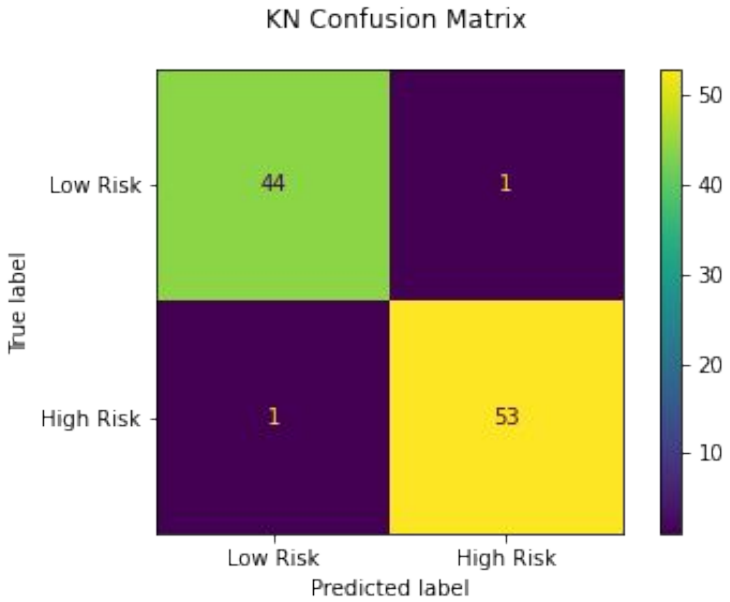
- Logistic Regression
- K-Nearest Neighbors
- Random Forest
- Baseline was 60.7%

# Model 1: Logistic Regression



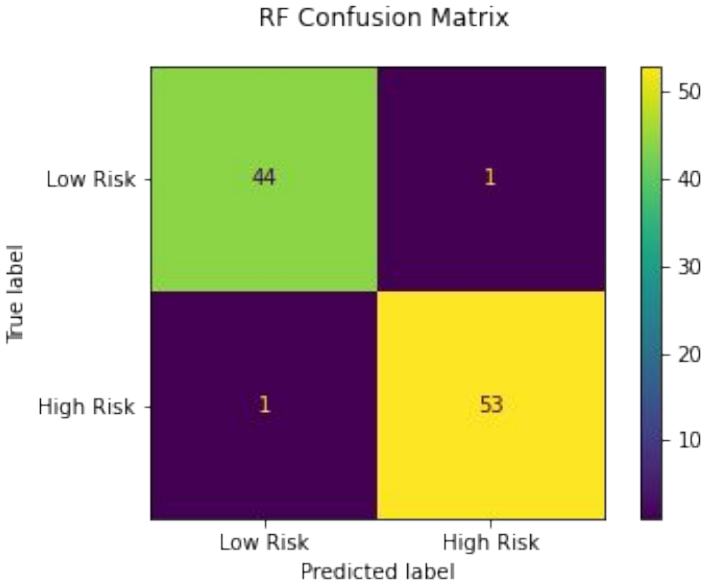

- **Score on training set:** 0.784
- **Score on testing set:** 0.7172
- **Accuracy:** 0.78 ± 0.10

# Model 2 : KNN



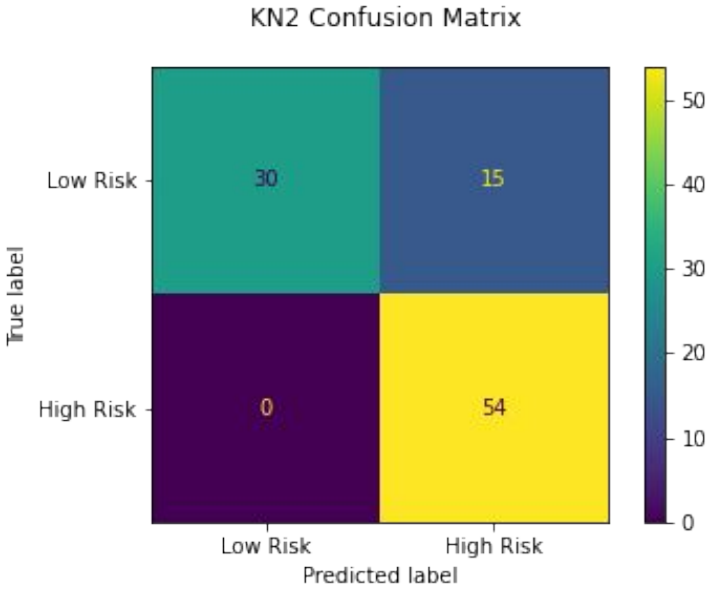

- **Score on training set:** 0.9966
- **Score on testing set:** 0.9798
- **Accuracy:** 0.96 ± 0.04

# Model 3: Random Forest



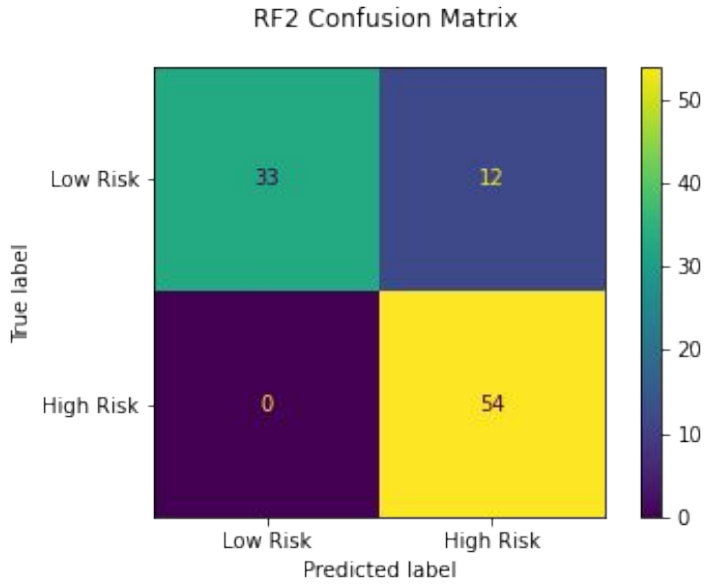

- **Score on training set:** 0.9966
- **Score on testing set:** 0.9798
- **Accuracy:** 0.95 ± 0.09

# Model 4: KNN Without Location Features




- **Score on training set:** 0.9730
- **Score on testing set:** 0.8485
- **Accuracy:** 0.90 ± 0.09

# Model 5: Random Forest Without Location Features




- **Score on training set:** 0.9730
- **Score on testing set:** 0.8788
- **Accuracy:** 0.93 ± 0.10

# Results

- Location was more influential than societal statistics.
- Societal statistics still proved to be a strong indicator of piracy.
- Poverty rate and government stance were stronger causal indicators than academic indicators.
- Random Forest and KNN models were close in effectiveness.

# Summary of Findings & Recommendations

- Most pirate attacks are happening in equatorial countries.
- There has been a decrease in pirating activity over the past decade.
- Cargo ships and tankers are at a higher risk. We recommend increased security on board for these types of vessels.
- Attacks tend to cluster in geographical choke points. We recommend creating land-based lookout and upping patrols points in these vulnerable areas.
- Most attacks are occurring in regions that have a higher poverty percentage, thus compounding problems in already troubled regions.
- Adoption of oceanic treaties should help to curb piracy.
- A successful, multi-front strategy may include offering economic development support in exchange for ratifying oceanic treaties.