Exploring Overfishing with Global Socioeconomic Data

SI 330 Final Project Phase #3

Zack Eisman

Project Overview

- This project is looking to explore two primary concerns
- 1. How certain economical, social and geographical factors can predict a country's impact on overfishing
    - Some factors of consideration include population, coast line, seafood consumption, and GDP per capita
- 2. How the United States specifically compares to other countries with similar predictors
- Data is needed to solve this problem because many people around the world may not be fully understanding why overfishing occurs. Data can help raise awareness as to why this is such a severe problem and which countries need to do a better job of limiting their fishing production on an annual basis.

Datasets

***Dataset #1 – Fish and Overfishing (Kaggle) ***

- Project uses 3 datasets from this source on Aquaculture and Capture Fishery Production, Fish Consumption, and Sustainable Fish Stocks

Dataset #2 – Countries of the World Information (CIA) – New

- Contains information on each world countries population, area, coastline, and GDP per capita, and population density

Dataset #3 – World Population Dataset (Worldometers) – New

- Contains population estimates for each country across the world from 1950-2023

Data Manipulation

Apply

- The Countries of the World DF contained numbers in a strange format so the apply function was used to turn the data into the proper format
- 78,9 → 78.9

Regex

- The population DF contained a Region column and Regex was used on this column to determine which countries were located in any of the americas
- Pattern = r‘AMER’
  - Matches with LATIN AMER. & CARIB & NORTHERN AMERICA
- A new column called America was created with these results

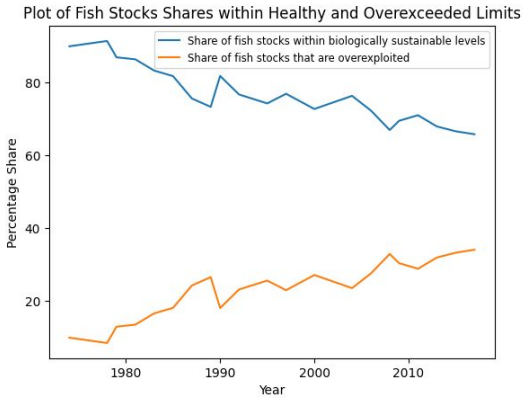
Merging DFs

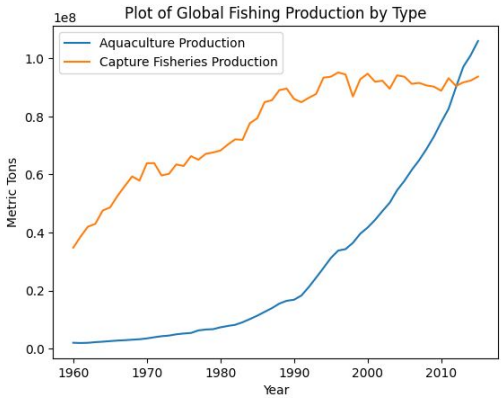
- The data for all DFs is connected using the country names as joint labels
- All 5 smaller DFs were merged together to form one larger DF that contains a variety of fishing and socioeconomic data for each country

[9]:

country	Region	America	Population	Pop Pct	Capture Fisheries Production	Fish Pct	Aquaculture Production	Seafood Supply	GDP per Capita	Coastline Ratio	Population Density
Afghanistan	ASIA (EX. NEAR EAST)	False	1.797304e+07	0.313610	9.795918e+02	0.000128	991.836735	0.085714	700.0	0.00	48.0
Albania	EASTERN EUROPE	False	2.841586e+06	0.056724	6.133133e+03	0.000985	1195.105263	3.310526	4500.0	1.26	124.6
Algeria	NORTHERN AFRICA	False	2.463322e+07	0.459186	7.627907e+04	0.010064	491.903612	3.187544	6000.0	0.04	13.8
Angola	SUB-SAHARAN AFRICA	False	2.255819e+07	0.327853	3.279143e+05	0.035764	295.055556	16.571667	1900.0	0.13	9.7
Argentina	LATIN AMER. & CARIB	True	3.214849e+07	0.624874	6.231933e+05	0.081949	1052.029734	6.782807	11200.0	0.18	14.4

Analysis Question #1: Is Overfishing Really a Problem?





Conclusion: The data shows overfishing has consistently become more of a global issue since 1980. During this period, there has been more aquaculture practice but capture fishery production is still increasing which is why overfishing is such an important environmental concern

Analysis Question #2: Which Socioeconomic Factors are Most Closely Related to Fishing?

```
[33]: df.corr()['Capture Fisheries Production']
```

```
[33]: America                          0.066637
      Population                       0.642522
      Pop Pct                          0.645402
      Capture Fisheries Production     1.000000
      Fish Pct                         0.991599
      Aquaculture Production           0.611584
      Seafood Supply                   0.232787
      GDP per Capita                   0.197265
      Coastline Ratio                 -0.055210
      Population Density              -0.009634
      Log Population                   0.455496
      Log Aquaculture Production       0.534275
      Name: Capture Fisheries Production, dtype: float64
```

Conclusion: There are two predictors that are strongly correlated with capture fisheries production: population, and aquaculture production

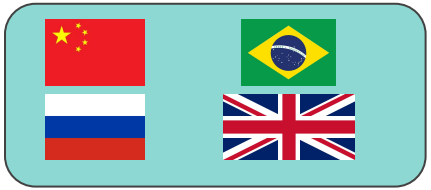
Analysis Question #3: Which Countries are Most Similar to the United States?

Socioeconomic Factors

Population, GDP per Capita, Coastline Ratio, Population Density, Aquaculture Production, Capture Fisheries Production, Seafood Consumption

Global Analysis

Countries within 20 places of US on World Rank in each statistic

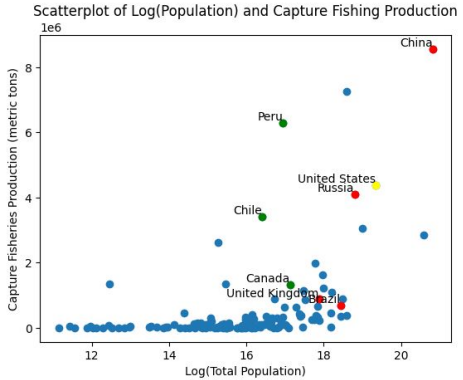


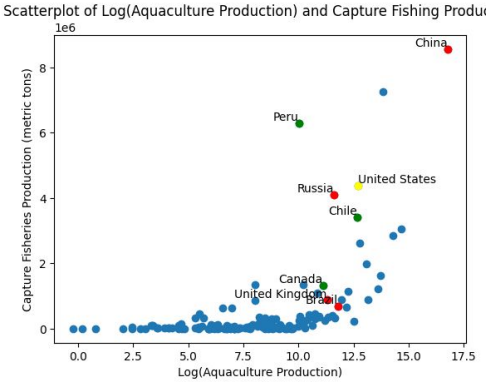
Local Analysis

Countries within 5 places of US in each statistic (only containing countries in the Americas))

Conclusion: The four most similar countries to the United States are China, Russia, Brazil, and the United Kingdom. Similar countries located on the same bodies of water are Canada, Peru, and Chile

Analysis Question #4: How Does the United States Compare to Similar Countries?





Conclusion: The United States has a much more capture fishery production than most other world countries and similar countries. To decrease the US's impact on overfishing, the US should target being more like countries such as Canada, Brazil, and the UK.

Note: The log of total population and aquaculture was required to better visualize the data

Moving Forward

- It would be very interesting to explore this data with a linear regression model
- It would also be intriguing to have better data on how overexploited the waters are on a regional (or national) basis rather than globally
- More and/or different socioeconomic predictors would also make the analysis more interesting