Crime at Sea: A Global Database of Maritime Pirate Attacks (1993–2020)

PAUL BENDEN (iD)

ALAN FENG (iD)

CHRISTOPHER HOWELL (iD)

GIULIO VALENTINO DALLA RIVA (iD)

*Author affiliations can be found in the back matter of this article

ABSTRACT

This dataset contains information from more than 7,500 maritime pirate attacks that took place between January 1993 and December 2020, as well as country indicator data for the same time period. The pirate attack data was collected from the International Maritime Bureau (IMB), tidied, and augmented with geospatial data. The country indicator data was gathered from a variety of sources, notably The World Bank. The data is contained in Comma Separated Value (CSV) files. The reuse potential includes its use by anti-piracy organisations and researchers, as well as commercial businesses, in the understanding and prevention of maritime piracy. This dataset is available through Zenodo and Github.

CORRESPONDING AUTHOR:
Giulio Valentino Dalla Riva

Mathematics and Statistics, University of Canterbury, Christchurch, NZ

giulio.dallariva@canterbury. ac.nz

(1) OVERVIEW
REPOSITORY LOCATION

https://zenodo.org/record/4583802

CONTEXT

Maritime piracy, defined by the International Maritime Bureau (IMB) as "any act of boarding or attempting to board any ship with the apparent intent or capability to use force in the furtherance of the act", has, throughout history, plagued sea users and coastal nations (International Maritime Bureau, 2007). To better understand this current predicament, pirate attack incidences in recent years can be seen in Supplementary Figure 1, which gives a decade separated visualisation of pirate attacks off the eastern coast of Africa. It was reported that in 2010 the global cost of piracy was at least $7 to $12 billion dollars per year when taking into account associated costs, such as ransoms and insurance premiums (Bowden et al., 2010). The human cost is also high with seafarers being exposed to a variety of psychological and physical dangers. In 2010, 1,090 seafarers were taken hostage, for an average duration of five months, with 488 suffering physical abuse such as deprivation of food and torture, along with psychological abuse including solitary confinement and mock executions (Hurlburt, 2013).

There are a number of organisations who focus on this issue: the most prominent of them is the International Chamber of Commerce, which represents 45 million companies from over 100 countries as well as international organisations such as the United Nations and World Trade Organization (International Chamber of Commerce, 2020). A division of the ICC is the Commercial Crime Services (CCS) whose task is combating commercial crime (Commercial Crime Services, 2020). A division of the CCS is the IMB which, in turn, set up the IMB Piracy Reporting Centre (PRC), which acts as a point of contact and immediate notifier of piracy activities, in 1992 (IMB Piracy Reporting Centre, 2020; International Maritime Bureau, 2020).

This work was completed as part of course work for the Master of Applied Data Science at the University of Canterbury.

(2) METHODS
STEPS

1: Sourcing, Tidying, and Enhancing Piracy Attack Dataset

Piracy attack data, from 1993 to 2014, was sourced from a dataset published by Daxecker & Prins (2015), who had sourced the data from the IMB. Data from 2014 to 2020 was web scraped from the IMB website (IMB Piracy Reporting Centre, 2020) using Julia (Bezanson et al., 2017).

The 1993 to 2014 data cleaning process conformed with tidy data principles (Wickham, 2014). Dummy columns, re-coded variables, and dependent columns were removed, and formatting errors were corrected using Julia, before being merged with the more recent data from 2014.

The distance of each pirate attack to the nearest coast for each entry in the dataset was calculated by loading every point on the world's coastline from the Natural Earth Coastline shapefile (Made with Natural Earth, 2020) into a vantage-point tree data structure (Nielsen et al., 2009), provided by the Julia VPTree library, and doing a nearest neighbour search from the attack location to the nearest point in the tree.

Using shapefiles provided by Marine Regions (Flanders Marine Institute, 2020), we were able to compute the nearest country from each pirate attack as well as whose coastal waters and Exclusive Economic Zone (EEZ) each attack occurred in, if any. These calculations were made using the 2020 boundaries.

Country names were wrangled using the R countrycode library (Arel-Bundock et al., 2018) and converted to ISO 3166 country codes.

2: Creating the Historical Country Indicator Dataset

Eleven datasets were used to prepare the Historical Country Indicator dataset that contains eleven indicators for 217 countries over a 30-year period. Indicator data was collected from three sources:

1. The World Bank (The World Bank, World Development Indicators, 2020g, 2020d, 2020b, 2020f, 2020a, 2020e, 2020c): Unemployment Total % of Total Labour Force (modelled

ILO estimate, collected from the International Labour Organization, ILOSTAT database), Intentional Homicides (per 100,000 people, collected from UN Office on Drugs and Crime's International Homicide Statistics database), GDP (current US$, collected from World Bank national accounts data and OECD National Accounts data files), Total Fisheries Production (metric tons, collected from Food and Agriculture Organization), Total Armed Forces personnel (collected from International Institute for Strategic Studies, The Military Balance), Total Population (collected from (1) United Nations Population Division. World Population Prospects: 2019 Revision, (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Report (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme), and Industry Including Construction (collected from World Bank national accounts data, and OECD National Accounts data files).

2. Transparency International (Transparency International, 2020): Corruption Perceptions Index.

3. UNU-WIDER (UNU-WIDER, 2020): Government Revenue Dataset.

As the datasets were collected from multiple sources with different specifications and structures, it was important to wrangle each sub-dataset into a tidy and unified format, to be compatible with the main pirate attack dataset, with missing values represented as "NA".

The R countrycode library was used to check that English country names and country codes were consistent.

3: Creating the Country Code Dataset

A dataset of ISO 3166 country codes and English country names was developed with data from The World Bank (The World Bank, World Development Indicators, 2020f) so that country codes and 217 countries used in the Historical Country Indicator and Pirate Attack datasets were consistent.

Thus, the creation of the relational database—Piracy Attacks Dataset (*https://github.com/newzealandpaul/Maritime-Pirate-Attacks*)—was realised with the structure depicted in Supplementary Figure 2.

QUALITY CONTROL

As the Historical Country Indicator dataset is a combination of indicators by country and year from a number of different datasets, it was important that the information of each indicator came from a reliable source. The data was cross checked against the original sources to ensure accuracy. There are inherent biases in methodology of reporting of country indicators which could, potentially, lead to inaccuracies when comparing intra- and/or international data (Agarwal, 1985; van Herk et al., 2004; Watanabe et al., 2018). The Corruption Perception Index, for example, like other indexes of its type, are prone to inherent psychological biases which can have a real negative impact on the countries in question (Donchev & Ujhelyi, 2014).

Country name parsing was implemented carefully and, potentially, issues of consistency were automatically flagged for manual inspection.

Geospatial information calculations were checked using manual measurements using Google Earth satellite imagery. The accuracy of distance to shore was found to be within one kilometre, which we considered accurate enough, depending on the coastline resolution at different points in the shapefile.

Scraped web data was sampled and compared with the original website data to ensure accurate transcription.

(3) DATASET DESCRIPTION
OBJECT NAME

Maritime Pirate Attacks Dataset

FORMAT NAMES AND VERSIONS

CSV files

CREATION DATES

3/2/2021

LANGUAGE

English

LICENSE

Creative Commons Attribution-ShareAlike 4.0 International License

REPOSITORY NAME

Zenodo

PUBLICATION DATE

2021-03-04

(4) REUSE POTENTIAL

There is reuse potential for (1) researchers, (2) anti-piracy organisations, and (3) corporations who use maritime transport to gain a deeper understanding of where, when, and why maritime piracy occurs and how best to combat against it. Research on piracy has, largely, focused on finding a relationship between one or more specific variables, such as state weakness and/or fisheries, and piracy attacks. Given that previous datasets for maritime piracy primarily included when and where pirate attacks occurred, but not these other variables, individuals or parties investigating piracy would have needed to input data or combine datasets. Therefore, our focus was to create a relational database that provides a more detailed array of data which includes multiple variables, such as GDP, and pirate attack data. As such, we broke reuse potential down into the three subgroups listed above.

The pirate attacks dataset and historical country indicator dataset can help researchers from a wide range of interests. They can assist researchers to gain better understanding of the pirate attacks in relation to a country or region's socioeconomic or sociopolitical situation from an economics or politics perspective. Moreover, researchers who are interested in the motivation of pirate attacks or the social or psychological toll of piracy on the population could use these datasets to develop a deeper understanding of factors that could be associated with the attacks. Furthermore, more detailed data modelling could be done with these data to elucidate possible correlations, causations, and, even, prediction of pirate attacks.

The data would also be useful for anti-piracy organisations and commercial corporations. Having the focus of direct action against and prevention of piracy, anti-piracy organisations could use this data to model and thus, predict occurrences of piracy to help with safeguarding seafarers and protecting coastal communities, as well as the reduction of significant economic loss. For corporations who use maritime transport, whose focuses are protecting employees from attack or kidnapping and preventing economic loss, utilisation of this dataset could help to determine the safest shipping routes, which may change seasonally, depending on the local GDP or state strength, or when additional security is needed for vessels.

ACKNOWLEDGEMENTS

ADDITIONAL FILES

The additional files for this article can be found as follows:

Supplementary Figure 1. Pirate attacks off the eastern coast of Africa. DOI: *https://doi. org/10.5334/johd.39.s1*
Supplementary Figure 2. Relational diagram of Maritime Pirate Attacks Dataset. DOI: *https:// doi.org/10.5334/johd.39.s2*

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Paul Benden ⓘ *orcid.org/0000-0002-4770-1037*
Mathematics and Statistics, University of Canterbury, Christchurch, NZ

Alan Feng ⓘ *orcid.org/0000-0001-5616-5427*
Mathematics and Statistics, University of Canterbury, Christchurch, NZ

Christopher Howell ⓘ *orcid.org/0000-0001-7938-6847*
Mathematics and Statistics, University of Canterbury, Christchurch, NZ

Giulio Valentino Dalla Riva ⓘ *orcid.org/0000-0002-3454-0633*
Mathematics and Statistics, University of Canterbury, Christchurch, NZ

REFERENCES

Agarwal, B. (1985). Work Participation of Rural Women in Third World: Some Data and Conceptual Biases. *Economic and Political Weekly*, *20*(51/52), A155–A164. JSTOR.

Arel-Bundock, V., Enevoldsen, N., & **Yetman, C.** (2018). countrycode: An R package to convert country names and country codes. *The Open Journal*, *3*(28), 848. DOI: *https://doi.org/10.21105/joss.00848*

Bezanson, J., Edelman, A., Karpinski, S., & **Shah, V. B.** (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. DOI: *https://doi.org/10.1137/141000671*

Bowden, A., Hurlburt, K., Aloyo, E., Marts, C., & **Lee, A.** (2010). The Economic Cost of Maritime Piracy (One Earth Future Working Paper). One Earth Future: Oceans Beyond Piracy. *http://oceansbeyondpiracy.org/sites/default/files/attachments/The%20Economic%20Cost%20of%20Piracy%20Full%20Report.pdf*

Commercial Crime Services. (2020, October 7). *Commercial Crime Services*. International Chamber of Commerce: Commercial Crime Services. *https://www.icc-ccs.org/icc-commercial-crime-services-ccs*

Daxecker, U. E., & **Prins, B. C.** (2015). Searching for Sanctuary: Government Power and the Location of Maritime Piracy. *International Interactions*, *41*(4), 699–717. DOI: *https://doi.org/10.1080/03050629.2015.1016159*

Donchev, D., & **Ujhelyi, G.** (2014). What Do Corruption Indices Measure? *Economics & Politics*, *26*(2), 309–331. DOI: *https://doi.org/10.1111/ecpo.12037*

Flanders Marine Institute. (2020). *Union of the ESRI Country shapefile and the Exclusive Economic Zones (version 3)*. *https://www.marineregions.org/*

Hurlburt, K. (2013). The Human Cost of Somali Piracy. In M. Q. Mejia Jr., C. Kojima, & M. Sawyer (Eds.), *Piracy at Sea* (pp. 289–310). Berlin, Heidelberg: Springer. DOI: *https://doi.org/10.1007/978-3-642-39620-5_19*

IMB Piracy Reporting Centre. (2020, October 7). *IMB Piracy Reporting Centre*. International Chamber of Commerce: Commercial Crime Services. *https://www.icc-ccs.org/piracy-reporting-centre*

International Chamber of Commerce. (2020, October 7). *ICC: About Us*. International Chamber of Commerce (ICC). *https://iccwbo.org/about-us/*

International Maritime Bureau. (2007). *Piracy and Armed Robbery Against Ships Annual Report* (Piracy and Armed Robbery Against Ships). International Chamber of Commerce: International Maritime Bureau (ICC-IMB). *https://www.rk-marine-kiel.de/files/piraterie/imb/IMB_Piracy_Report_2007.pdf*

International Maritime Bureau. (2020, October 7). *International Maritime Bureau*. International Chamber of Commerce: Commercial Crime Services. *https://www.icc-ccs.org/icc/imb*

Made with Natural Earth. (2020). *Coastline, 1:110m Physical Vectors Downloads [Data set]*. *https://www.naturalearthdata.com/downloads/110m-physical-vectors/110m-coastline/*

Nielsen, F., Piro, P., & **Barlaud, M.** (2009). Bregman Vantage Point Trees for Efficient Nearest Neighbor Queries. In *Proceedings—2009 IEEE International Conference on Multimedia and Expo, ICME 2009* (pp. 878–881). DOI: *https://doi.org/10.1109/ICME.2009.5202635*

The World Bank, World Development Indicators. (2020a). *Armed forces personnel, total [Data set]*. *https://data.worldbank.org/indicator/MS.MIL.TOTL.P1*

The World Bank, World Development Indicators. (2020b). *GDP (current US$) [Data set]*. *https://data.worldbank.org/indicator/NY.GDP.MKTP.CD*

The World Bank, World Development Indicators. (2020c). *Industry (including construction), value added (% of GDP) [Data set]*. *https://data.worldbank.org/indicator/NV.IND.TOTL.ZS/*

The World Bank, World Development Indicators. (2020d). *Intentional homicides (per 100,000 people) [Data set]*. *https://data.worldbank.org/indicator/VC.IHR.PSRC.P5*

The World Bank, World Development Indicators. (2020e). *Population, total [Data set]*. *https://data.worldbank.org/indicator/SP.POP.TOTL*

The World Bank, World Development Indicators. (2020f). *Total fisheries production (metric tons) [Data set]*. *https://data.worldbank.org/indicator/ER.FSH.PROD.MT*

The World Bank, World Development Indicators. (2020g). *Unemployment, total (% of total labor force) (modeled ILO estimate) [Data set]*. *https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS*

Transparency International. (2020). *Corruption Perceptions Index [Data set]*. *https://www.transparency. org/en/cpi/2020/index/nzl*

UNU-WIDER. (2020). *UNU-WIDER Government Revenue Dataset [Data set]*. *https://www.wider.unu.edu/ project/government-revenue-dataset*

van Herk, H., Poortinga, Y. H., & **Verhallen, T. M. M.** (2004). Response Styles in Rating Scales: Evidence of Method Bias in Data From Six EU Countries. *Journal of Cross-Cultural Psychology*, *35*(3), 346–360. DOI: *https://doi.org/10.1177/0022022104264126*

Watanabe, C., Tou, Y., & **Neittaanmäki, P.** (2018). A new paradox of the digital economy—Structural sources of the limitation of GDP statistics. *Technology in Society*, *55*, 9–23. DOI: *https://doi. org/10.1016/j.techsoc.2018.05.004*

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software, Articles*, *59*(10), 1–23. DOI: *https://doi. org/10.18637/jss.v059.i10*