Project Documentation

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

The Correlation between Nigeria's GDP and Maritime Pirate Attacks in Its Waters

Vrije Universiteit Amsterdam

Introduction to Digital Humanities and Social Analytics

Lecturers: Dr Lorella Viola, Dr Erika Kuijper

18-10-2024

Word count: 5060

Group 3

Hannah van Keulen, Lucille Lu,

Marleen Wolting, Sávio Porsius Martins,

Trần Phương Linh & Yasha Hennige

Table of Contents

Project Overview

Research Question

"To what extent does Nigeria's Gross Domestic Product correlate with the rate of piracy attacks near the country between 1993 and 2019?"

Thesis Statement

By investigating the correlation between the frequency of pirate attacks near Nigeria from 1993 to 2019 according to the data of "Crime at Sea: A Global Database of Maritime Pirate Attacks (1993 - 2020)"[1] and Nigeria's Gross Domestic Product, we aim to explore whether a relationship exists between the economic conditions of Nigeria and the amount of piracy attacks near Nigeria. We expect to observe a relatively strong correlation between Nigeria's Gross Domestic Product and the number of piracy attacks in Nigeria over the years between 1993 and 2019.

The Dataset

The dataset is collected from Benden et al.[2] It consists of three relational data sets with multi-faceted information about piracy attacks, country codes, and some indicators of the countries involved. The original dataset of piracy attacks was pulled from the International Maritime Bureau. The other datasets were pulled from The World Bank, Transparency International and Government Revenue Dataset. The data is highly reliable due to the credibility of the sources.

However, three considerable limitations would hinder our ability to perform informative research. Firstly, the dataset is incomplete and has many missing values. For example, in the data concerning different measurements regarding each country over the year, there are 2522 and 2479 missing values on the column "corruption_index" and "homicide rate," respectively, out of 5899 values. We acknowledge that if we remove every missing value in the dataset, we would lose nearly half of the original dataset. A reason for this issue is the fact that the dataset was partly collected from another research and partly web-scraped. The misalignment between two sources and the incompleteness of each data source would result in this problem. Secondly, we recognize that the categories are ambiguous in some variables[3], and the dataset owners do not report detailed descriptions of each variable. Finally, the dataset owners do not clearly specify the contextualization of the data collection from the original sources.

---

[1] Benden et al. 2021
[2] Benden, et al. 2021, p. 2
[3] Such as the terminology used for describing the attacks, e.g., the difference between the data collected under the 'boarding' and 'boarded' tags, is not made clear or explained. The other terminology equally lacks explanation

Relevance

In *Crime at Sea*, Benden et al. state that their dataset offers a reuse potential, which could serve to gain a deeper understanding of the place, time and cause for pirate attacks. Researchers, anti-piracy organisations, and corporations dealing with maritime transport are explicitly named as groups to which the dataset could be of interest.[4] Researchers, in particular, could analyse the data to look for patterns and predict pirate attacks in order to prevent them.[5] Currently, little research has been conducted that makes use of the dataset, most likely due to its recent creation. As a result, our research could serve as a foundation for future research that makes use of the dataset, offering valuable insights, criticism and recommendations for future research.

---

[4] Benden, et al. 2021, p. 4

[5] Peace and Security, 2021

Data Acquisition

The dataset utilised in our project is titled "Crime at Sea: A Global Database of Maritime Pirate Attacks (1993 - 2020)."[6] It comprises three distinct datasets. The primary dataset contains records of over 7,500 maritime pirate attacks worldwide from 1993 to 2020. The other two datasets provide country-related information. One includes country codes, while the other offers data on country indicators such as corruption index, homicide rate, GDP (both overall and industry-specific), total armed forces personnel, population, and unemployment rate.

The pirate attack data was collected by the International Maritime Bureau (IMB). Information from 1993 to 2014 was sourced from a dataset published by Daxecker & Prins,[7] while data from 2014 to 2020 was obtained via web scraping from the IMB website. Once gathered, the data was processed. Daxecker & Prins applied tidy data principles to the 1993–2014 data, and the database creators added two additional variables: nearest country and EEZ (Exclusive Economic Zone) country. The nearest country was determined by calculating the closest coastline using the Natural Earth coastline shapefile,[8] and the EEZ country was identified based on the 2020 EEZ boundaries for each pirate attack.

The creators enhanced the dataset with geospatial data and pre-processed it for consistency. Country-level information, such as GDP, primarily came from The World Bank, while the corruption index was sourced from Transparency International, and government revenue data from UNU-WIDER. The dataset of country indicators also relies also on datasets of The World Bank. Recognizing the challenges of integrating data from multiple sources, the creators emphasised the need to unify the datasets. This involved standardising the format and filling missing values with "NA" (Not Available), ensuring compatibility across all three datasets for simultaneous use.

---

[6] Benden, et al. 2021, p. 1-6
[7] Daxecker & Prins, 2015, p. 699-717
[8] Made with Natural Earth. (2020)

Methodology

To address our research question, we divided tasks into three subcategories: Literature review, Data analysis, and Workflow/Portfolio documentation. Sávio and Marleen tackled the Data analysis, Yasha and Lucille focused on the Literature review, while Hannah and Linh handled the Workflow/Portfolio documentation.

Data

All three datasets were utilised to address our research question. The country codes dataset was specifically used to identify Nigeria's code, enabling us to isolate data on maritime pirate attacks where Nigeria is the nearest country and to extract relevant country indicators for Nigeria.

The filtered data on maritime pirate attacks included 559 instances where Nigeria was the nearest country from 1993 to 2020. To analyse this data by year, the attack dates were simplified to display only the year, rather than the full date (year, month, day). Once the new "year" variable was created, we combined it with Nigeria's country indicators, generating a dataset with both pirate attack frequencies and relevant economic data for Nigeria.

Our research question investigates the correlation between Nigeria's Gross Domestic Product (GDP) and the frequency of pirate attacks near Nigeria from 1993 to 2019. To calculate this correlation, we used Pearson's correlation coefficient. GDP, represented in US dollars, is a ratio variable, as is the frequency of pirate attacks, which also has a clear zero point (indicating no attacks). Given the nature of these two ratio variables, the Pearson method was appropriate for our analysis.

To calculate the Pearson correlation between the two variables, both Python and R were used. These programs were chosen due to team members' familiarity with these programs. Sávio worked with Python, while Marleen used R. They later compared their results, identified and corrected any mistakes, and ultimately arrived at the same final outcome.

Literature

To deepen our understanding of maritime pirate attacks around Nigeria and their correlation with GDP, relevant literature needed to be reviewed. Yasha and Lucille gathered various studies and research on these topics to clarify key concepts. This information allowed us to better interpret the data and connect our findings with existing research. For example, it helped us understand why there would be a positive correlation between these two variables in Nigeria, and why this relationship could differ from other countries. The literature suggested that Nigeria's oil industry could influence this correlation, a detail we wouldn't have known without the literature. This contextual information was crucial for interpreting our data and aligning our findings with existing research, particularly as Nigeria's situation may differ from that of other countries.

Workflow

To ensure our project remained as transparent as possible, it was crucial to maintain a clear record of our workflow. To keep everything organised and accessible, Hannah and Linh created an overview of the steps we made and made our GitHub repository easily navigable, including all the documents we produced. Without this, our transparency would have been compromised, and the project would be harder to replicate. Additionally, without proper organisation, we could have risked losing track of important documents and changes made throughout the project, potentially resulting in the loss of critical data and steps.

Workflow Steps

After getting to know each other, our strengths, weaknesses, and the consensual meeting timeframe, we discussed possible gaps in our assigned datasets. Acknowledging the lack of certain data and the incompleteness of several variables, our preliminary research question aimed to solve these uncertainties and provide more precise explanations for the datasets. We decided to use the text mining method to answer the research question.[9] After proposing this research question, we received feedback that our focus was not the data itself but data criticism, which was meant only to be a small part of the project.[10]

To initiate a new start to our project, we decided to move past the dataset criticism and focus more on what part of the data we can use. After discussing the data familiarisation process, we narrowed our focus to frequencies of attacks, geographical location, and economic situation variables. This is because those variables were consistent and most complete, coupled with literature showing that the relationship between attack frequency and economic status is engaging in research.[11] We decided to compare the correlation between the frequency of attacks and the corruption index between Yemen and Indonesia.[12]

Our secondary meeting aimed to specify the new research question further.[13] After doing individual research, members in the group realised that comparing the correlation of two countries is nonsensical yet more demanding. We changed our focus to only one country and picked Nigeria. Nigeria is an isolated country relative to the Gulf of Aden and is in a less chaotic geographical location; therefore, less relevance decreases the possible inherent confounding factors, making it easier to conclude the correlation relationship. Our research question to this point was: "To what extent does the corruption index of Nigeria correlate with the rate of piracy attacks on the country between 1993 and 2019?" The research question was formulated, and our following meeting was for task division. In a group of six, we divided people into three tasks: literature research, data analysis, and documentation organisation.[14]

During the data analysis, we acknowledged a shift in the measurement standard of the corruption index before and after 2012. Technically, before 2012, the CPI was measured relative to other countries, so comparing their CPI scores across the years would be meaningless. Meanwhile, after 2012, the CPI measurement changed, making it possible to compare themselves across the year[15]. Based on this information, we risked losing usable data if we continued researching the corruption index. At the same time, in the correlation test, we were expected to have as many data

---

[9] Date in logbook: 12-09-2024, 19-09-2024 and 20-09-2024
[10] Date in logbook: 25-09-2024
[11] Eme et al., 2022
[12] Date in logbook: 27-09-2024
[13] Date in logbook: 30-09-2024
[14] Date in logbook: 02-10-2024
[15] Stephenson, 2014

points as possible. Once again, our research question was refined, and we concluded our research question formulation as: "To what extent does the GDP of Nigeria correlate with the rate of piracy attacks on the country between 1993 and 2019?".[16]

Our data analysis procedure was conducted by Marleen and Sávio using both R and Python, based on their knowledge of tools. These tools are straightforward for data analysis and accessible for future use, ensuring the research's sustainability and usability.

After concluding the correlation test results, we held another meeting to explain and plan the next step. We discussed our project's 'end shape', the limitations of our research, and what to prepare for the presentation and project documentation. During this meeting, we divided tasks on writing and finalising the documentation file based on personal strengths and tasks in the earlier steps of the project.[17]

---

[16] Date in logbook: 03-10-2024
[17] Date in logbook: 05-08-2024, 07-08-2024 and 08-10-2024

Challenges and Solutions

Dataset and Handling the Data

The limitations of the dataset, as described under the Project Overview, played a consistent role of interference during the project. Explicit explanations of the terminology used to mark the different data points was absent. Our initial aim was to focus on these limitations, but due to course criteria and time constraints we have chosen to work around the difficulties, while simultaneously acknowledging them.

Along with data criticism, our group did meticulous data preprocessing and familiarisation to select which portion of data would be meaningful for analysis. We acknowledged that the geographical data, number of privacy attacks and certain data related to economic status within specific countries are nearly completed, and narrowed down our research scope accordingly. We eventually selected Nigeria as the country we focus on in this project.

Formulation of Research Question

We initially had some significant difficulties formulating the research question. This was firstly due to the critical approach we had taken towards the terminology used within the dataset. Our first discussions on the dataset kept returning to what we saw as significant limitations and this focus prevented us from exploring how the dataset could be used.

This was further complicated by the differences in our academic backgrounds and our various levels of experience working with similar projects, as well as the different experience in handling data. The different understandings[18] specifically hindered us from fully understanding each other's point of view and tackling this was heavily reliant on our communication skills.

Additionally, we were unclear of the scope and limitations that were set by the course criteria, which led to misaligned expectations and confusion. The communication was further pressured by the approaching deadline for the submission of the initial research question.

In an initial attempt to address these difficulties, the formulation of our first research question[19] was driven by our shared feelings of frustration and dislike of the dataset. However, the feedback we received made clear that our approach did not align with the expectations and the course criteria. This led to a reevaluation of our initial discussions of possible research questions and a more careful and patient communication on the possibilities and envisioned limitations. By that time, the group dynamics had improved, and we had a better understanding of the course criteria and of each

---

[18] Such as attitudes towards what is considered a relevant and meaningful research question and how do we approach the dataset.

[19] The initial research question was as follows: To what extent do the categories and terms present within our dataset align with categories and terms used in research done on pirate attacks in the Gulf of Aden.

other's perspectives. As a result, we were able to critically formulate the research question, while simultaneously working and communicating more efficiently.

Corruption Index to GDP

It was until the data analysis stage that we realised that there was a need for a shift from the corruption index measurement to the Gross Domestic Product. The continued use of the Corruption Index would mean a significant decrease in the amount of data available for rigorous correlation analysis due to inconsistent measurement methods within the time span between 1993 and 2019, whereas there is no contradictory issue regarding GDP. A shift to the use of GDP as opposed to the Corruption Index meant that there were more data points available for performing the correlation analysis. This challenge is further explained in the workflow steps.

To address the challenge, we held an urgent online meeting, helped each other explain the issue, and came to an agreement to change the focus from corruption index to Gross Domestic Product. GDP can be deemed as a good variable as it is a relatively objective indicator. Luckily, Lucille had done rigorous literature research on Nigeria's and different countries' economic status and their relevance to piracy attacks. Therefore, it took us less than a day to redefine the research question and find substantial research to back up the hypothesis.

Group Dynamic

Even though our social relationship has been positive and consistent throughout the course of the project, the different backgrounds, different experiences and skill-sets complicated our communication at times. As explained in concern to the formulation of the research question, we had difficulty understanding each other's perspectives. The communication of expert-knowledge is challenging, and requires an understanding of each other's level of expertise, as well as a use of language that is simple and effective in summarising the, at times, difficult concepts and jargon. This was handled by trial and error, collective discussions, constant communication, and careful documentation of steps taken by the 'experts'.

Similarly, the positive social relationship made us more sensitive to group morale. This was especially noticeable when we received the feedback from our initial research question and felt collectively discouraged and uncertain. Recovering felt difficult at first, but we managed this by taking time, encouraging each other, venting our frustrations and reinforcing our positive social dynamics. After this we also sought more frequent feedback from the teachers, in an attempt to find reassurance and prevent the experience of unexpected setbacks in order to more easily maintain a growth mindset.

Ethical Considerations

We rely on secondary data generated by Benden et al. (2021), which is an online open source dataset. During our project, we are consistently aware of respecting intellectual property. We have promoted a culture of citation, for which we explicitly and honestly acknowledge the works written by other scholars. This also casts doubt on where we get the ideas from. Digital humanities studies are never value-free, and this applies to our project as well. Data manipulation and interpretations involve personal values, which can be biassed. By providing reasoning for our decisions we aim to minimise confusion and vagueness, and thus ambiguity.

Moreover, due to incomplete knowledge and potential errors in the data sources, uncertainty cannot be avoided. One example is the accuracy of piracy incidents and the annually reported GDP, especially the true number of pirate attacks near the coastline of Nigeria is highly likely to suffer from under-reporting.[20] This poses a threat to the validity of the outcome.

Additionally, the use of language is important. Therefore, using clear and precise terms consistently is vital for clarity and replicability. An example of this is how correlation does not mean causation and using those terms interchangeably can cause confusion as they refer to two different things. We are consciously addressing this throughout our documentation, and meanwhile, be careful about the wording that introduces value judgements.

---

[20] Nwalozie, 2020, p. 159-178

Results

To gain a clearer understanding of how Nigeria's GDP and the frequency of pirate attacks near its coast have evolved over the years, several graphs were created. In figure 1 illustrates Nigeria's GDP growth from 1993 to 2019, showing an increase over time. Similarly, figure 2 depicts the frequency of pirate attacks in the region, which also shows an upward trend. However, the pattern of pirate attacks is more erratic, with noticeable fluctuations. Figure 3 presents a scatterplot comparing these two variables, revealing a somewhat positive correlation between them.

The Pearson Correlation Coefficient was used to conclude that there was an observed positive correlation between Nigeria's GDP and the frequency of maritime piracy attacks. The correlation result was significant ($r = .526$, $p = .005$), indicating a positive relationship between Nigeria's GDP and the frequency of pirate attacks. As GDP increases, the frequency of attacks tends to rise and as the GDP decreases, the frequency of attacks tends to decrease. The strength of this correlation is moderate, with the proportion of explained variance (R^2) at 0.277, meaning that 27.7% of the variance in pirate attack frequency is explained by changes in GDP. It is important to note that while there is a correlation, this does not imply a direct causal relationship between the two variables.

Our group found literature regarding the (regional) political situation of Nigeria. Most literature on the topic mentions there is plenty of maritime piracy in Nigerian waters and the Gulf of Guinea in general. The fact that Nigeria is one of Africa's main oil producers seems to play a significant role. The abundance of oil in the area might encourage the increase in maritime piracy activity. The natural resources can be seized and resold for higher profits. Additionally, there is a possibility that maritime piracy and regional or civil conflict might go hand in hand, since the Gulf of Guinea is an important maritime transportation route.[21] These contextual facts can help us better understand the background information of Nigeria, since they seem to be closely related to the national economic performance of Nigeria, of which the GDP is an indicator.

Apart from that, most literature treats spatial causations between country indicators including GDP per capita and piracy attacks, which yield a negative relationship.[22] For instance, as the GDP per capita increases by 1%, all incidents of piracy attacks decrease by an estimation of 0.01*0.110 = 0.0011 unit at 10% significance level..[23] In Eme et al. (2022), the piracy attacks in Nigeria are specifically studied using a time-series regression analysis. Their results suggest that when there is a 1% increase in GDP, the frequency of pirate attacks against ships reduces by approximately 1.160%. [24] This negative causal relationship can be explained by the following economic reasoning: When the GDP (per capita) increases, it gives a positive signal of economic growth, which is likely to be

---

[21] Nwalozie, 2020, 159-178
[22] Also seen in Daxecker & Prins, 2015, p. 35 and Okeahalam & Otwombe, 2016, p. 125-160
[23] Daxecker & Prins, 2021, use data from 1995 to 2014, which is a different time range than which was used for this project. See for details p. 106.
[24] Eme, et al. 2022, p. 21-28

associated with lower unemployment and lower inflation. These macroeconomic indicators can suggest that on average, the living condition within a nation improves, so it is less likely that people turn to immoral activities for financial gains.

Our result shows the significant relationship between GDP and piracy attacks, which does not align with the existing literature. The difference can mainly be explained by the focus of the studies. The explicit focus on causation, includes the use of regression variables besides the GDP, which we have not included in our own project. A strong correlation might imply a causation, but one cannot draw a definite conclusion without scrutinising it carefully by considering all factors that can possibly affect GDP and piracy attacks. The variables include, but are not limited to: unemployment rate, poverty rate, literacy rate, government expenditure to maritime security, revenue generated by maritime sectors, indices related to political stability, governance, law enforcement, judicial effectiveness etc. The moderate strength of our result is due to these omitted variables.

In order to answer the original research question: To what extent does Nigeria's Gross Domestic Product correlate with the rate of piracy attacks on the country between 1993 and 2019? It can be stated that there is a positive correlation between the GDP of Nigeria and the amount of piracy attacks between 1993 and 2019. The positive correlation can be due to the financial motivations to commit oil theft. However, as previously mentioned, there are many other possible factors[25] that could have played a part in this observed trend. Therefore, we cannot rule out that there might be other explanations for this pattern.

---

[25] Such as local governance, the distribution of wealth, cultural motivations, environmental factors and social motivations, etc.
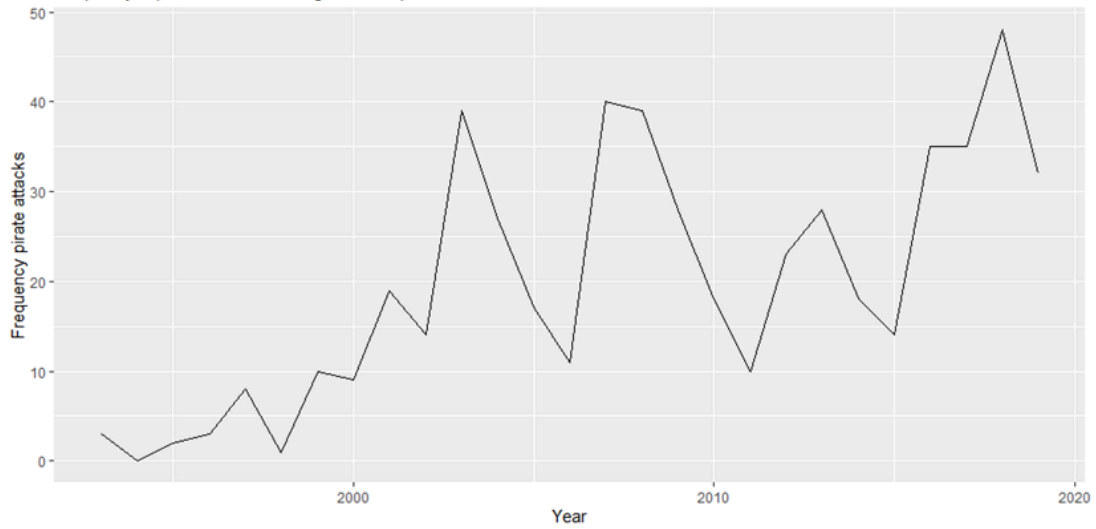
Figure 1

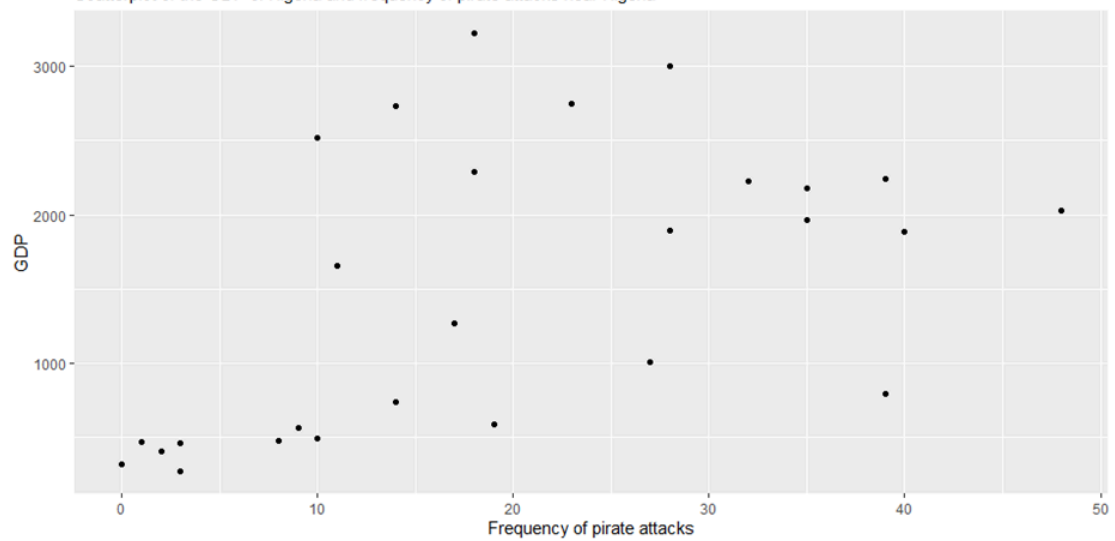Gross Domestic Product of Nigeria ofer the span of 1993-2019 in US Dollars
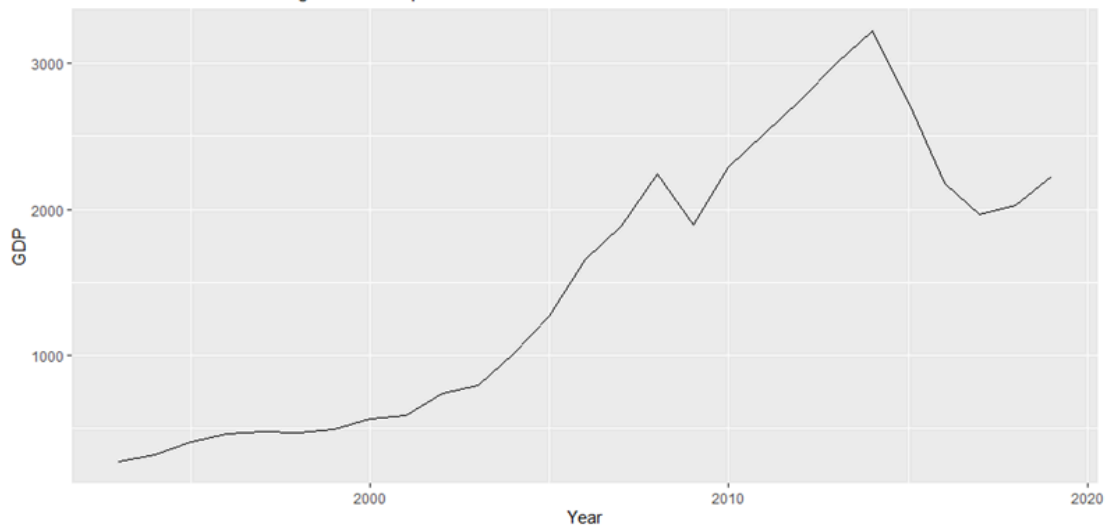
Figure 2

Frequency of priate attacks near Nigeria in the period of 1993-2019

Figure 3

Scatterplot of the GDP of Nigeria and frequency of pirate attacks near Nigeria

Documentation and Sustainability

For this project, a GitHub repository was created to ensure transparency in our workflow and enhance the reusability of our research materials. The repository, titled Introduction-to-digital-humanities-and-social-analytics_group3, is organised into several folders (Logbook, Presentation, Project Proposal and Team Charter) and the Project Documentation. This structure allows for easy navigation and access to the different stages of our project, ensuring that all relevant information is clearly presented.

The Logbook folder serves as a record of our workflow. It contains subfolders named according to the dates of our group meetings. Within each of these dated folders, you will find within the file named 'description' you will find a explicit mention of the workflow step[26] and a summary of discussions held, the tasks completed, and any modifications made during that session. Additionally, raw documents have been added, that were used for notetaking or which were changed during the course of the meeting.

This ensures a transparent view of our progress and decision-making process. Of particular note is the folder for the meeting dated 03-10-2024, which includes two crucial files: analysis.py and Project correlations.R. These files provide step-by-step instructions on how we processed and analysed the data to arrive at our final results. The analysis.py file is written in Python, detailing the steps taken using Python-based tools, while the Project correlations.R file contains equivalent instructions for users working in RStudio. These files guide users through the exact process we followed, offering clear pathways for replication or further exploration. For access to the dataset used in our project, you can refer to the original repository where the data is stored (https://github.com/newzealandpaul/Maritime-Pirate-Attacks). The dataset can be found in the data/csv folder within that repository on GitHub, allowing users to replicate the analysis or apply the data to other projects.

The Presentation folder includes the slides from our final project presentation, along with an additional file that served as the foundational structure for the presentation itself. This material highlights our approach to presenting the findings and the key elements of our research discussion.

In the Project Proposal and Team Charter folder, you will find the files related to the initial planning stages of our project. The project proposal outlines the objectives, methodology, and scope of our research, while the team charter includes important information about team roles, responsibilities, and the rules governing our collaboration throughout the project.

Finally, the Project Documentation includes a comprehensive summary of the project's development, an overview of the research process, key findings, and reflections on the project's outcomes. By exploring the various folders within the repository, you will gain a thorough understanding of our project workflow, data analysis methods, and the final results of our research.

---

[26] We have chosen to define the workflow steps as following: initiation, monitoring, planning, execution

Reflection

Workflow Effectiveness

The importance of communication for a functioning workflow is apparent. Many factors can influence the communication, from noisy and crowded workplaces disturbing online and real time communication, to more metaphysical attributes, such as expectations, skills and previous experiences. However, a workflow can not entirely be anticipated. It is a constant shift within the flow. If we reflect on our own workflow, we see this more clearly when we take into account the several changes we had to make, which forced us to re-evaluate or repeat certain steps of the process, and spending a significantly smaller portion of time actually working on execution.

This also implies a need for more efficiency in approaching the dataset, when we think about future improvements we can argue that there is a need for an earlier task division. This, however, is arguably more feasible when the different levels of expertise are better understood by the different members of the group, which now happened later in the process.

One of our strengths was most likely the critical assessment of our dataset. We spent a significant amount of time of our workflow focussing on this and navigating the limitations, to the point that it may have hindered our progression. Due to our focus on limitations, we experienced time constraints and had difficulty formulating a plan and a research question. This too, in consideration with future improvements, could have been better navigated. What could help to combat this time constraint would be maintaining a clearer and stricter planning, and allowing less space for discussions on limitations. However, this could affect the group dynamics and collective understanding negatively. A balanced approach is therefore key in adequately adjusting to the course of the workflow, while navigating time-management and maintaining collective understanding.

Lessons Learned

We have understood the importance of transparent communication in an interdisciplinary group. Working in interdisciplinary groups, while advantageous in making use of diverse range of strengths in different disciplines, requires much more effort to reach a mutual understanding. The experience allows us to practise and find the best way to communicate with each other during the intensiveness of the project. It is crucial to make sure that everyone is on the same page and has a clear grasp on where the project is heading. Working together as a multifaceted group, we have balanced each other's skills, strengths and weaknesses, which have allowed us to flourish and grow closer as a group. Besides, expectation management and being realistic and honest is important as well. This enabled us to ask for help without fear and support each other whenever needed.

This project demonstrates that DH research entails a non-linear but iterable workflow, and constant reflection is required along the way. During the discourse of deepening understanding, we

realised the inherent complexity of the process. While the central part of the project is the execution step, it only comprises a small proportion of the whole process. Continuous adjustments were made at every step. Therefore, flexibility and understanding of the other group members are helpful and allow for a safe and open working experience.

Another important lesson is that working in an agile workflow requires meticulous documentation and reflection along the way. This is especially important when working in a big group, which requires us to keep everyone updated. It enhances the group's efficiency, and the project's sustainability is advanced as well. An example is in data analysis, giving detailed notes on the code would help when working with other people—especially those less familiar with the code. It would also allow better reflection and fixing bugs when necessary.

Future Improvements/Research Directions

Causation analysis is more meaningful than correlation as one can see a clearer picture of how and to what extent one variable directly impacts another by including the omitted variable bias.[27] In our case, the question can be what are the key driving forces behind pirate attacks around Nigeria (mainly the Gulf of Guinea). Therefore, the root cause(s) can be figured out more rigorously for designing a more effective solution. This project can indeed be further extended by exploring the causal relationships between piracy attacks and some relevant indicators. Apart from that, we initiated several more suggestions for future improvements and research directions:

First of all, we have found that many past papers examined the general patterns of pirate attacks at a regional or global level, but there is relatively less time-series analysis. There could be more studies focusing within one country that scrutinise the performance of combating piracy attacks.

Secondly, past papers concentrated more on macro-level, and the national datasets are more accessible compared to local data. Based on our dataset, most piracy incidents from 2017 near Nigeria happened around Lagos, followed by Bonny island and Brass. Probably connecting these cases with local conditions will produce more valid conclusions than using national figures. It has been suggested that if the local states are governed poorly, this provides opportunities for collusion with organised pirates, then the chance of maritime piracy will be higher.[28] This gap of local-level focus can be filled by future research.

Thirdly, the dataset owners can update the information yearly so that there is more available up-to-date data for the reuse potential.

Last but not least, transparent sharing of information is crucial for research purposes. It ensures consistency and accessibility of data, making research more feasible. Because of many missing data and some ambiguous terms, our data scope was restricted. This leaves space for future improvements.

---

[27] For a more in depth discussion, see Results, p.14
[28] Daxecker & Prins, 2021, p. 73

Team and Task Distribution

Member	Background	Main Project Task	Contributions Project Documentation
Yasha Hennige	Media, Art, Design and Architecture	Literature Research & Results	Literature findings & analysis Ethical considerations Results Reflection
Hannah van Keulen	Ancient Studies	Documentation & Organization	Project overview Challenges & Solutions Reflection
Trần Phương Linh	Communication Science	Documentation & Organization	Workflow steps Challenges & Solutions Reflection
Lucille Lu	Economics and Business Economics	Literature Research & Results	Literature findings & analysis Ethical considerations Results Reflection
Sávio Porsius Martins	Artificial Intelligence	Data-Analysis	Data acquisition Methodology Documentation & sustainability Reflection
Marleen Wolting	Criminology	Data-Analysis	Data acquisition Methodology Results Documentation & Sustainability Reflection

Bibliography

Benden, P., Feng, A., Howell, C., & Dalla Riva, G. V. (2021). Crime at Sea: A Global Database of Maritime Pirate Attacks (1993–2020). *Journal of Open Humanities Data, 7*(0), 19. https://doi.org/10.5334/johd.39

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

Daxecker, U., & Prins, B. C. (2021). *Pirate lands : governance and maritime piracy.* Oxford University Press.

Drucker, J. (2021). The Digital Humanities Coursebook. The Digital Humanities Coursebook, 226–230. https://doi.org/10.4324/9781003106531

Eme, O. N., N, D. D., Chinonyerem, N. T., & Chigozie, M. J. (2022). The Influence of Economic Growth on Trend of Sea Piracy and Armed Robbery Attacks against Ships in Nigeria. Himalayan Journal of Humanities and Cultural Studies, 3(1), 21–28.

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

Nwalozie, C. J. (2020). Exploring Contemporary Sea Piracy in Nigeria, the Niger Delta and the Gulf of Guinea. Journal of Transportation Security, 13(3–4), 159–178. https://doi.org/10.1007/s12198-020-00218-y

Okeahalam, C. & Otwombe, K. (2016). Socioeconomic development and the risk of maritime piracy. *Journal of Transportation Security. 9*.

Osinowo, A. A. (2015). Combating Piracy in the Gulf of Guinea. Africa Center for Strategic Studies. http://www.jstor.org/stable/resrep19053

Peace and Security. (2021, August 9). *Global Response Needed to Counter Rising Security Threats at Sea [Online forum post]*. UN News. Retrieved September 21, 2024, from https://news.un.org/en/story/2021/08/1097432

Stephenson, M. (2014, September 23). Pre-2012 CPI scores CANNOT be compared across Time-So please stop doing it! *GAB | The Global Anticorruption Blog.* https://globalanticorruptionblog.com/2014/09/30/pre-2012-cpi-scores-cannot-be-compared-across-time-so-please-stop-doing-it/