Does the GDP per capita of a country influence the type of crime?

Kenneth Obahayujie (21752715)
Arsenii Troitskii (20204701)

A report submitted in part fulfilment of COMP 30780

Data Science in Practice

Demonstrator: Barry Smyth

UCD School of Computer Science
University College Dublin

July 8, 2025

Table of Contents

Abstract

Our Project Title: "Does GDP level of a country influence the type of crime committed within its borders or are there better socio-economic factors that can measure the crime committed?". This research project recognizes the complex nature of crime and aims to shed light on the underlying socio-economic determinants that contribute to criminal activities. The project will employ a multi dimensional approach to gather and analyse relevant data from countries around the globe.

The project's strategic overview is rooted in the importance of understanding the underlying factors that influence crime rates. By examining the relationship between GDP levels and crime, it aims to contribute to the broader conversation on crime prevention and policy development.

The results from this research project can provide valuable insights for policymakers, law enforcement agencies, and social scientists to develop targeted strategies and interventions to address crime effectively. This project also recognizes the limitations of using GDP as the sole measure of economic well-being and crime rates. By exploring alternative socio-economic factors, it aims to broaden the understanding of the complex interplay between economic, social, and cultural dynamics in shaping crime patterns.

The strategic overview highlights the need for comprehensive and subtle approaches to crime prevention and policy formulation, considering a broader range of socio-economic factors beyond GDP alone.

Chapter 1: **Introduction**

Our research project is based on the very intriguing question of whether a country's GDP level is the main factor influencing crime and its types within that country, or whether there are other economic indicators that have a more significant impact on the criminal world. Despite the growth of the total population of the planet over the last 21st century, the world has seen a gradual decrease in the total number of crimes. GDP is considered the main unit of measurement of the economic situation in a country, but its correlation with crime remains less obvious to this day. The inspiration for this project came to us from real life, we live in Ireland, a country with a high GDP (26th place), but can we say that we live in a safe country? Our research questions range from identifying the types of crimes for countries with high GDP to answering the question of whether there is a delay between an increase in one of the indicators and the types of crimes that may follow it. Throughout the research work, we studied various economic factors such as happiness levels, inequality rates and even took into account the level of corruption. This is just a small list of indicators, showing the versatility of our research and the desire to provide the most accurate analysis of the main socio-economic conditions that affect crime. This will give us the opportunity to contribute to the development of more effective strategies to combat crime.

3. Research Questions 1: Does a high GDP / capita correlate with a decrease in violent or fraudulent crimes?. For this question I will investigate with countries with a high GDP per capita experience a decreasing in violet crimes such as homicides or kidnappings. To do this we will use data from GDP per capita and different types of crimes. The primary source of our data will be from GDP data from Kaggle and crimes from United State Office.

Research Question 2: Is there an immediate change in the crime rate after a change in GDP or is there a lagging phase? ; For this question we will try and evaluate look for fluctuations in crime or GDP and evaluate if there is a lagging phase when one of these numerical factors experience a sudden shift to the upside or downside

Research Question 3: Do countries with a high corruption score have fewer cases of fraud/theft than countries with low corruption score: We will examine whether countries with high corruption scores have fewer cases of robberies or theft compared to countries with low corruption scores. To investigate this question, we will collect data on corruption scores, fraud/theft rates. The data will we collect on this will be using corruption scores and robberies and thefts per 100,000.

Research Question 4: Are there significant differences in the rates of cyber-related crimes in countries with varying GDP levels, and if so, what factors contribute to these disparities? To answer this question we will gather data on cybercrime rates & GDP levels. By the end of this question we hope to come to a conclusion if countries within a high GDP experience more cyber crimes than countries with a low GDP or a similar conclusive answer.

Research Question 5: Is GDP the sole factor for the crime rate each country respectively are there macroeconomic factors that influence the crime rate more than GDP; To address this question, we will analyze a range of socio-economic indicators beyond GDP, such as income inequality, unemployment rates or happiness levels. I will examine what factors affect crimes more than GDP and if to come to the conclusion if GDP is a good factors to try and determine the rate of crimes.

Research Question 6: Do countries with high happiness index and high unemployment rate have a specific type of crimes prevalent?; We will investigate if countries with high happiness levels or high unemployment rates have similar types of crimes prevalent in comparison to low happiness or unemployment levels. We hope to come to the conclusion that 'x' crime is very prevalent in

countries with high happiness levels and 'y' crime is active with countries with low unemployment rates etc.

Our primary sources of data will be from 'Kaggle', 'Our World in Data', and the crime datasets will be from United Nation Office on Drugs & Crimes. 2.

In chapter 1 we will provide an overview for the project; the title, the objective and motivations. In chapter 2 project objectives, we will focus on the specific objective of the overall project and what we hope to find in each of our research questions. Chapter 3 we will review and examine existing work and similar studies related to our project topic. We will identify gaps in those existing works and areas where our project aims to improve on what those previous works failed on. In Chapter 4 we discuss the data collection process and the considerations involved. We will talk about how we collected our data, the websites we use, and describe the steps taken to prepare or clean the data for our analysis. In Chapter 5 we present the findings on our analysis. We address each research question individually, providing the results and insights obtained using graphs to support our analysis. We discuss the correlation between GDP levels and different types of crime, the temporal relationship between GDP and crime rates, the impact of corruption on fraud/theft, the disparities in cyber-related crimes based on GDP levels, and the influence of macroeconomic factors on crime rates. In the final chapter,chapter 6, we summarise the key findings of the research. We discuss the implications of our results, draw conclusions based on the evidence presented, and provide suggestions for future work and areas of further investigation on what future data science analysts can do to present a better project than what we have presented so far.

This report follows a linear progression, starting with an introduction and project objectives, followed by a review of related work and data considerations. The results chapter presents the findings for each research question, and the conclusions chapter summarises the main outcomes of the research

Chapter 2: Project Objectives

We chose this topic because its results may interest various groups of people. For instance, tourists selecting vacation destinations, job seekers scouting new work locations, or even residents of any country might find the study results intriguing. Additionally, politicians can use these findings to bolster specific aspects of life to deter various crimes, armed with insights into the factors that may trigger them. Moreover, law enforcement stands to gain significant advantages by recognizing that as certain economic indicators rise, so does the incidence of crime, enabling proactive prevention measures.

Moreover, the people who commit crimes themselves, if they are aware that their plans are predicted and everything is in place to prevent them, this can reduce their desire to commit these crimes, which will reduce the number of crimes in the process of committing them and make the world a much safer place.

The main objective of our project will be to identify the influence of selected economic factors on the types of crimes, their quantity, and their continental distribution. We plan to focus on more unpredictable factors such as the level of happiness, inequality coefficient, democracy of the population, level of population development, and others. This may provide an opportunity to discover completely unpredictable correlations that can be exploited in the future.

While not all countries are willing to disclose the prevalence of different types of crimes within their borders, we remain hopeful that the wealth of data available on the Internet will facilitate comprehensive and thorough analysis.

2.1 Objectives & Motivations

We decided to undertake this project because it not only presents us with an interesting challenge, but also an opportunity to gain a deeper understanding of the relationship between economic performance and crime rates. This is not just a passion project, but research that can yield valuable knowledge and understanding of how various factors impact public safety.

We hope to achieve several goals with this project. First, we aim to identify possible relationships between the level of GDP and the level of crime of different types. To do this, we analyze various indicators, such as the level of happiness, the inequality coefficient, the level of democracy in the government, the unemployment rate and many others. As a result, we hope to discover any patterns or relationships between these factors and crime rates.

In addition, we also hope to confirm our hypotheses using statistical methods and analytics. If we can prove that such connections exist, it will help to better understand what factors may contribute to increases or decreases in crime. However, if we do not find significant relationships, this will also be a valuable result, since it will confirm the absence of a direct relationship between the level of GDP and the level of crime.

2.2 Research Questions

We plan to examine the impact of GDP level on crime rates in different countries, as well as identify other factors that may influence this relationship. To do this, we formulated the following research hypotheses:

1. RQ1 – A higher GDP per capita is associated with a decrease in violent or assault crimes.
2. RQ2 – Changes in GDP lead to immediate changes in the crime rate, without a lagging phase.
3. RQ3 – There is a negative correlation between a country's corruption score and the prevalence of fraud or theft
4. RQ4 – Countries with higher GDP levels experience lower rates of cyber-related crimes compared to those with lower GDP levels, and this relationship is influenced by other factors.
5. RQ5 – GDP is the primary determinant of a country's overall crime rate, with other macroeconomic factors playing a secondary role.
6. RQ5b – Countries with high levels of happiness and unemployment are more likely to experience specific types of prevalent crimes.

Chapter 3: **Related Work**

Our project titles look at the relationship between GDP and other economic factors against the crime rate of certain crimes. Doing some background research on the internet we have discovered that there are similar projects to ours. Benjamin Northrup and Jonathan Klaer created a PDF report that examined the 'Effects of GDP per person on Violent Crime'. They briefly mentioned unemployment rates and poverty but not in the detail that we will look into them. They had two hypotheses : 1. Crime rates will increase as people get poorer and more desperate for money and 2. The victims of crime will get hit, thus reducing the opportunities for criminals to steal. They also used population density as a metric to try and predict crimes in certain areas. One thing to notice about their project in comparison to ours is that they only examined it in the US. They formed some regression models using crime rate as a dependent variable and GDP per capita, graduation rate and poverty rate as independent variables.

They came to some interesting conclusions at the end of their analysis using regression models. The main dependent variable, GDP per capita, was proven to be significant. Considering that violent crime rates, to the 1 percent level, are positively dependent on GDP per capita, it would seem that the growth of individual wealth would have negative impacts on society. Even with their evidence and use of predictive models there are some limitations to their research, 1 being the lack of graphs in their report that is able to conclude to us, the readers, that their conclusion is stable. Another limitation is the study utilises data only from 2003-2011 and focuses on the United States. While the study examines several factors that could influence crime rates, crime causation is multifaceted and often influenced by a combination of economic, social, and individual factors. The study briefly mentions the implications for policy makers but could further discuss how the findings could inform crime prevention strategies and resource allocation. [1]

An area our study will improve on from theirs is the use of multiple years for our data e.g We will be examining at least 9 or 10 years of data for each factor and crime we will use to come to a fair analysis and our years will be more recent than 2011. We also plan to split data into 3 subsets of data and measure the impact GDP has on them - low GDP countries, high GDP countries and median or middle GDP countries. We will split them using the quartile method. See 6

Chapter 4: Data Considerations

Following the research question of our project, we use different sets of data, which could be split into two main categories: the social-economic and crime data.

Social-economic factors:

- GDP data covers different 181 countries with values from 1999 to 2023. The columns represent the country names and the years.
- Human Development Index - includes data for 186 different countries and also includes continent codes for all regions. There is also a column with country codes, which contain short names for each country. However, this column will not be used in the future analysis. The data is available since 1980, but the years available for analysis are 1980, 1990 to 2021.
- Inequality coefficient (Gini Score) - contains columns for Country (185 countries), Code (which corresponds to the country code mentioned before), Continent Code, Year (ranging from 1963 to 2022), and the Gini income inequality index.
- Corruption score - dataset includes columns for Country (185), Year, Continent Code, and Corruption Perceptions Index. The Corruption Perceptions Index ranges from 0, indicating the highest corruption, to 100, indicating no corruption. The dataset covers the years from 2001 to 2022.
- Happiness index - columns: Country (161), Code, Continent Code, Year (2013,2015 - 2023) and Happiness Index 0 (unhappy) - 10 (happy).
- Unemployment rate - columns: Entity (country) 201 unique countries, Code, Year (1991 - 2023) and Unemployment, total (percent of total labor force).
- Population score - is the part of GDP table, column population for each country (181) in million, cover years are 1999 - 2023.
- Democracy score - columns Entity (country) 201 unique countries, Code (country code), Year which cover a time period (1999 - 2023), and Electoral democracy index ranges from 0 to 1 (most democratic).

Crime data:

- Crime index - County (141), Crime Index (0-100) values, Safety Index - the opposite value from Crime Index.
- Crime Data from UNDO includes information on 195 unique countries. The region column corresponds to the continent code. The category column encompasses various types of crimes, including corruption, bribery, smuggling of migrants, burglary, theft, fraud, cyber-related fraud, money laundering, unlawful access to computer systems, unlawful interference with computer systems or data, unlawful interception or access of computer data, acts causing environmental pollution or degradation, acts involving the movement or dumping of waste, trade or possession of protected or prohibited species of fauna and flora, and acts resulting in the depletion or degradation of natural resources. The dataset covers the years 2003 to 2021. The unit of measure column represents both counts and rates per 100,000 population.

- Crime data taken from The Global Economy: The datasets for Homicides per 100,000 people, Thefts per 100,000 people, Number of Prisoners per 100,000 people, Robberies per 100,000 people, and Kidnappings per 100,000 people all follow the same format. They include a Country column with 185 unique countries, a Code column (country code), a Continent Code column (EU, AF, AS, OC, NA, SA), and values measured per 100,000 people.

4.1 Data Collection

For our project, we collected our data from reliable open sources. Most of our data were taken from open resources. For example, we use official statistics from the United Nation Office on Drugs and Crime about various types of crimes across different countries. In addition, kaggle provides us with data on Gross Domestic Product and unemployment rate from all countries over the period from 1999 to 2022 and to make the GDP data more recent, we used GDP values for all countries for the year 2023 scraped from Wikipedia and added them to the main GDP dataset. To gather most of the data for social-economic factors, we use The Global Economy. From this resource, we access the Happiness Index, Inequality Score (or Gini coefficient), Corruption Score, Life expectancy, Human Development Index, Unemployment Rate and amount of population across most countries. Furthermore, we didn't rely on just one source for crime data. We also used The Global Economy to gather information about homicides, kidnappings, thefts, and robberies. Additionally, we scraped the crime index for each country, which is one of the key factors in our research, from Numbeo. Moreover, we decided it would be beneficial to include political factors such as the democracy score, which we obtained from Our World in Data.

4.2 Data Preparation

The overarching goal of our data preparation phase was to construct a comprehensive table containing all countries and observed indicators, as well as various types of crimes, spanning the maximum available number of years. To achieve this goal, it was necessary to analyze the availability of data for each indicator. The fundamental concept is to merge all tables using "country" + "year" as a unique key. This approach ensures that we have all indicators and crimes for each key pair, thereby preventing data loss. To achieve that, it is necessary to ensure that all our countries have exactly the same name in each table, and the years should be of the same data type (integer), as some tables use strings for them.

When considering country names, variations like 'UNITED STATES', 'US', and 'United States of America' obviously refer to the same country. However, it's essential to ensure consistency to prevent data loss or duplication. In addition, some datasets containing data from more recent years may use outdated country names, such as 'USSR' for Russia. These names are also checked and cleaned to ensure consistency across the data. Therefore, we've decided to use the list of countries from the GDP dataset as the main reference. This approach enables us to maintain uniformity in country names and minimize discrepancies

In the case of preparing the inequality index data, although the raw data covered 185 countries over the period from 1963 to 2022, it required preparation before it could be used. The problem arose after checking the amount of existing data for each country and each year. The result was insufficient; we lacked data for almost 160 out of 185 countries. This necessitated selecting a specific time period with the most available data. After filtering the years, we narrowed the

time period from 1999 to 2022. Although this resulted in a shorter time span, it increased the percentage of existing data for the Gini coefficient, making the data more reliable.

The Happiness Index data also required some preparation steps. In the raw dataset, we had data for all years from 2013 to 2022. However, the data for the year 2014 was completely empty. To address this, we calculated the average between the years 2013 and 2015 for each country and used that value for the year 2014. In cases where either the data for 2013 or 2015 was missing for a country, we omitted that country's data for the year 2014, as including it would compromise the accuracy of the calculation

Similarly, in the unemployment dataset, column names were excessively long, such as "Unemployment, total (percent of total labor force)". This could make it challenging to clearly understand the content of each column, especially when merging datasets. To address this issue, all long names were replaced with shorter, more descriptive names. For example, "unemployment rate" was used instead of "Unemployment, total (percent of total labor force)". This renaming process improves readability and usability of the data.

In order to facilitate data comparison, it is essential to normalize all represented values.These steps primarily refer to the crime datasets and GDP, as their values are dependent on the population of each country. Since all our crime datasets represent quantities of each crime and GDP reflects the monetary measure of a market, we have decided to divide them by 100,000 to facilitate analysis. By dividing the crime datasets and GDP by 100,000, we aim to standardize the values and express them on a per capita basis. This normalization allows for fair comparisons across different countries with varying population sizes, enabling more accurate analysis of crime rates and economic indicators.

The last, but not the least, during the analysis and preparation of the crime database, it was noted that there were too many overly specific types of crimes included. For example, categories like 'Corruption: Bribery' or 'Corruption: Other acts of corruption' were consolidated into the broader category of 'Corruption'. This consolidation helped prevent data loss and simplified analysis. Additionally, categories such as 'Acts that cause environmental pollution or degradation', 'Acts involving the movement of dumping of waste', and 'Trade or possession of protected or prohibited species of fauna and flora' were simplified and replaced by the more general term 'Environment crime'.

As a result of this consolidation process, the prepared crime datasets now include the following observed crimes: 'Corruption', 'Smuggling of migrants', 'Theft', 'Cyber-related crime', 'Money laundering', and 'Environment crime'. These were merged with other crimes such as 'Homicide', 'Kidnapping', 'Theft', and 'Robbery', making the table of crimes completely ready for research

Chapter 5: **Results**

??

5.1 RQ1:Does a high GDP per capita correlate with a decrease in violent or assault crimes?

Our very first question we want to answer is 'Do high GDP per capita countries have a decrease in violent or assault crimes?'. This question aims to investigate if there is a correlation between countries of high Gross Domestic Product and lower rates of violent or assault crimes. It is good to note that countries with High GDP per capita tend to have higher populations than countries in the median or lower quartile of GDP per capita, so if that seems to be the case we aim to examine if there is a decrease in violent or assault crimes. By examining the relationship between economic prosperity, as measured by GDP, and crime rates, we intend to discern if there is a pattern indicating that countries with higher GDP per capita experience lower levels or declining levels of violent or assault crimes than countries with lower GDP per capita. The two crimes for this we will be analysing is the homicide and the kidnapping rate.

Two sub questions that we will use from the anchor question are:**Do low GDP per capita countries have an increase in theft?** : This question focuses on exploring the relationship between countries with lower GDP per capita and the occurrence of theft crimes as they have a less advanced economy than higher GDP per capita countries which may motivate individuals to commit theft crimes in order to combat an economy that is trailing behind others.

The next sub question we will examine is **Can low GDP per capita countries be considered a contributing factor to homicide rates?**:

This sub question investigates whether low GDP values can be regarded as a contributing factor to the incidence of homicide rates. By examining the relationship between GDP and homicide rates, we aim to determine if countries with lower economic stability tend to exhibit higher rates of homicides than countries with a higher GDP per capitas than countries in the lower quartile.

By including these sub questions we will be addressing different aspects of the main research question. These sub questions help provide a more nuanced and broader understanding of the research topic with specific crime types and GDP per capita levels.

My main research question and sub questions hope to provide a clear direction for this study. Through the analysis we aim to determine whether economic factors such as GDP per capita can influence the occurrence of certain crimes or at the very least have some sort of correlation to the crime rates.

5.1.1 Data & Method

This is all the data we used to answer this question:

- Country - This column represents the name of the country for each data. It allows us to identify and differentiate the data based on countries being analysed. Some of the countries were different when combining the datasets. E.g 'United States' was called 'USA' in one dataset so we had to rename the using pandas to 'United States' so we could merge them using the merge function correctly and having all the data in tact .
- Year - This column was used to indicate the specific year for the data entry. It was a string initially but converted to a numeric variable in my code. For almost all our datasets we used a subset of years from 1999 - 2023.
- Population - This column has the population size for each country in their respective years. This column was used to divide GDP by the countries population in a year to create what we would use for the basis of our whole project 'GDP per capita'
- Thefts per 100,000 people - This column contains the amount of thefts committed for each country. It provides information on property crimes
- GDP - This column represents the Gross Domestic Product of each country for the given year. GDP is a measure of the economic output of a country and serves as an indicator of its economic prosperity. This was used to create the GDP per capita column, by dividing it by the population of a country in a given year.
- GDP per capita - This data indicates the GDP per capita, which is calculated by dividing the GDP by the population size. It provides a measure of economic well-being on an individual level. It is a more normalised version of GDP to conduct a fair analysis
- Homicides per 100,000 people - This column represents the number of reported homicides per 100,000 people in each country. It serves as an indicator of the prevalence of violent crimes.
- Kidnappings per 100,000 people - This column shows the number of reported kidnappings per 100,000 people in each country.

On how we used this data to answer the question: we constructed various graphs. One being a scatter plot. We used GDP per capita and Thefts per 100,000 people to be placed on a scatter plot to examine the relationship between these 2 variables, I also calculated the correlation coefficient using the 2 variables. If they had moderate correlation, anything over 0.2 or 0.3, I would use that factor or crime and conduct a more in-depth analysis by constructing other graphs to come to a fair conclusion.

I split the one dataframe into 3 datasets categorising them based on their GDP per capita level. Low, median and high GDP per capita countries. I used the quartile method to do so. Low GDP per capita countries would be the bottom 25 percent based on the overall GDP per capita of the dataset, median GDP per capita countries would be the middle 50 percent and high GDP per capita countries would be the top 25 percent. By splitting them up I could do a more thorough and fair analysis by comparing each dataset to each crime and looking for correlations or inverse relationships.

An example of me doing this was creating a line chart measuring the mean theft rate for each group of countries throughout the years. I could also create bar charts for each group looking at the average kidnappings or other crime for a year like 2014.

By employing these analyses and calculations, I can use the provided data to explore the relationships between GDP and crime rates, specifically addressing the research questions. The findings derived from these analyses would provide insights into the potential correlation between GDP and different types of crimes, as well as the impact of low GDP values on theft and homicide rates

5.1.2 Results

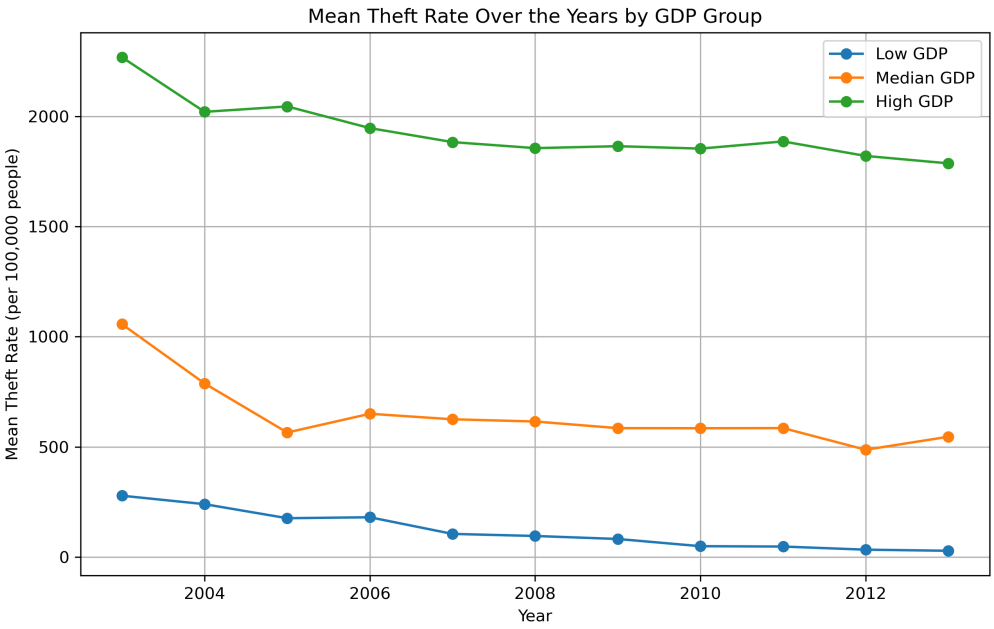


Figure 5.1:

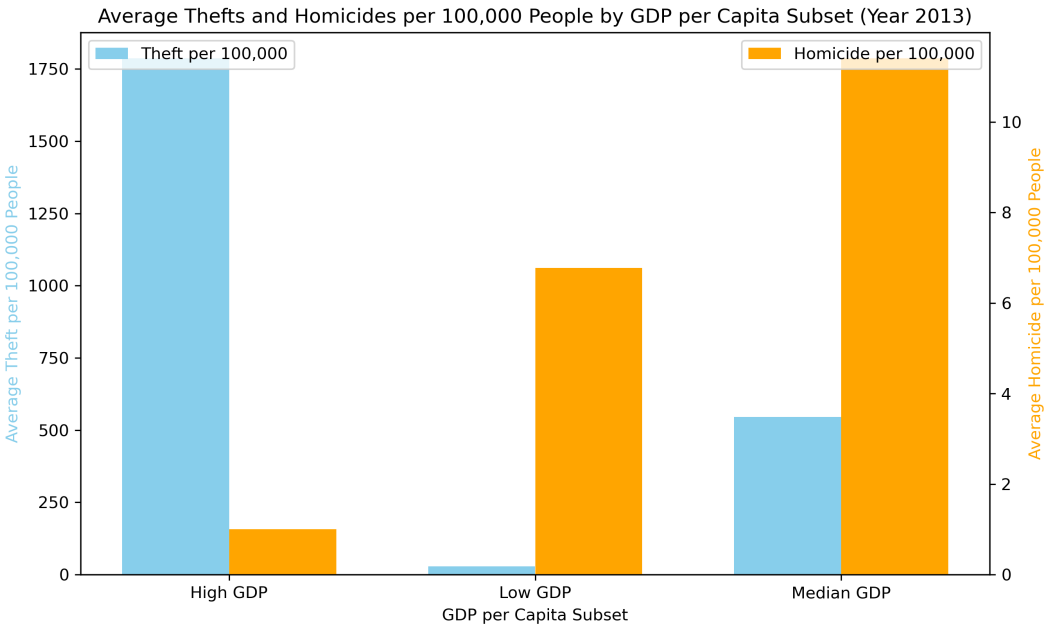


Figure 5.2:

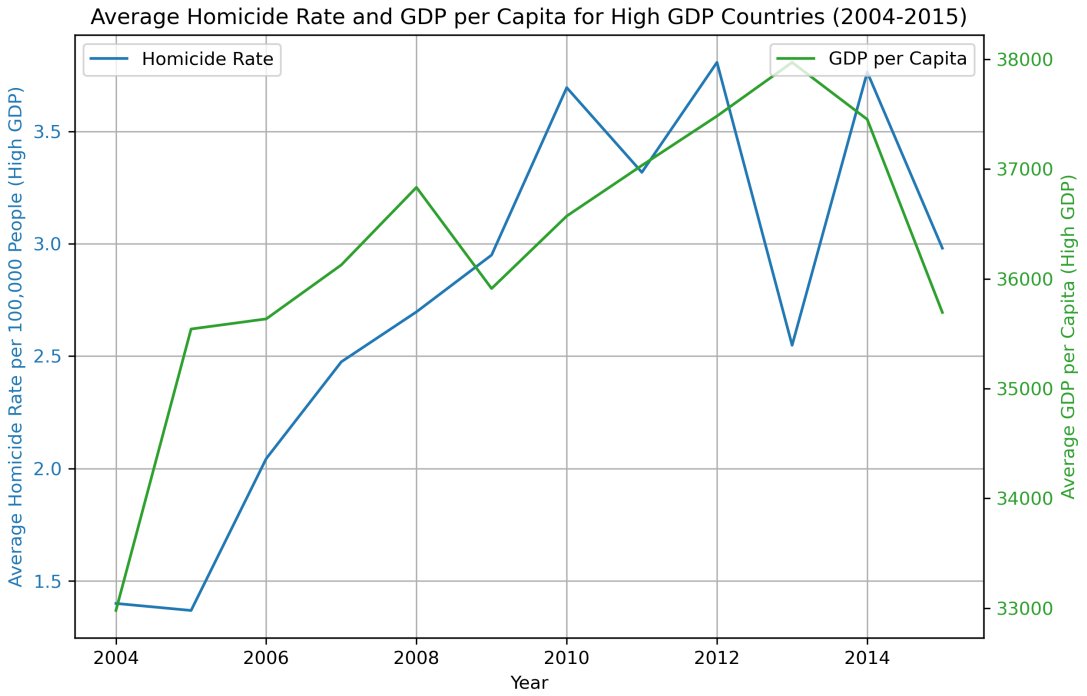


Figure 5.3:

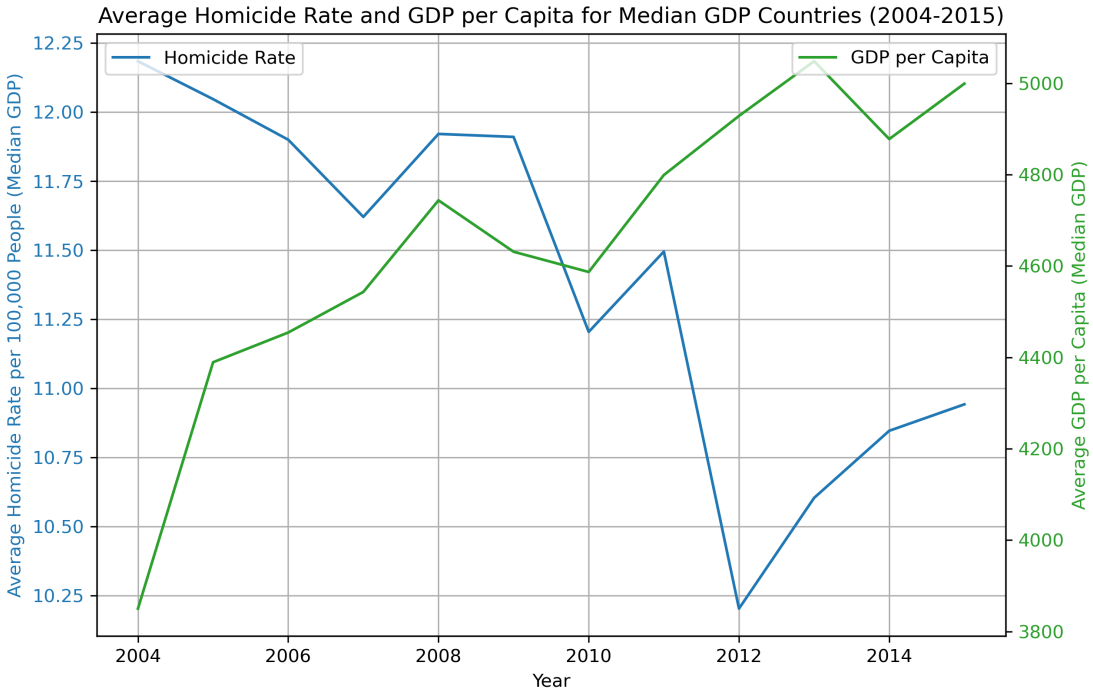


Figure 5.4:

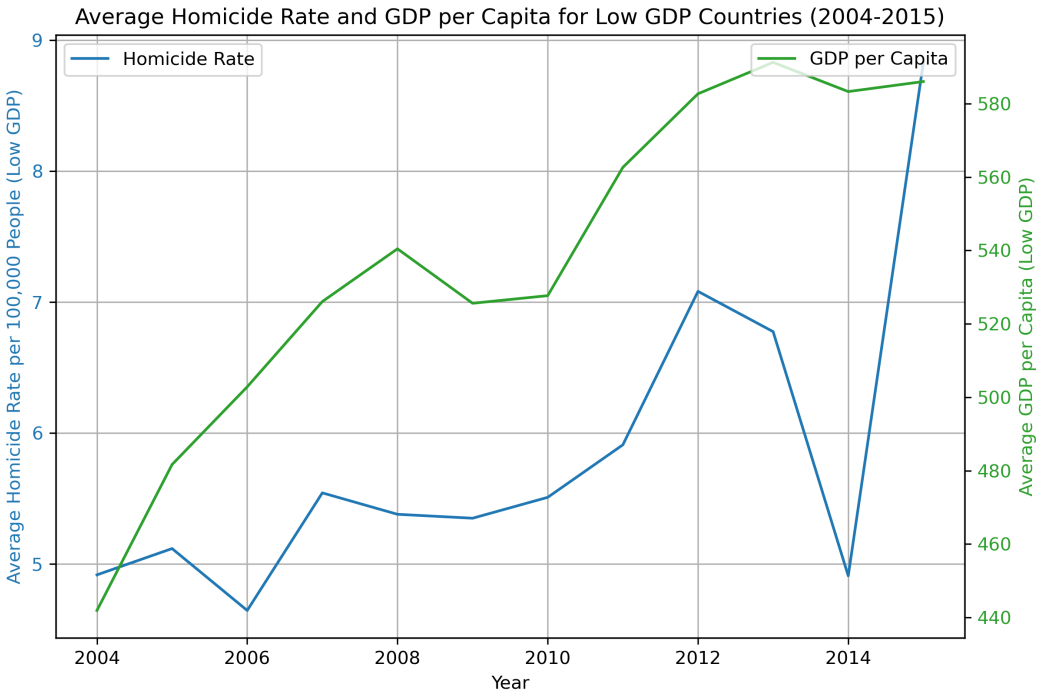


Figure 5.5:

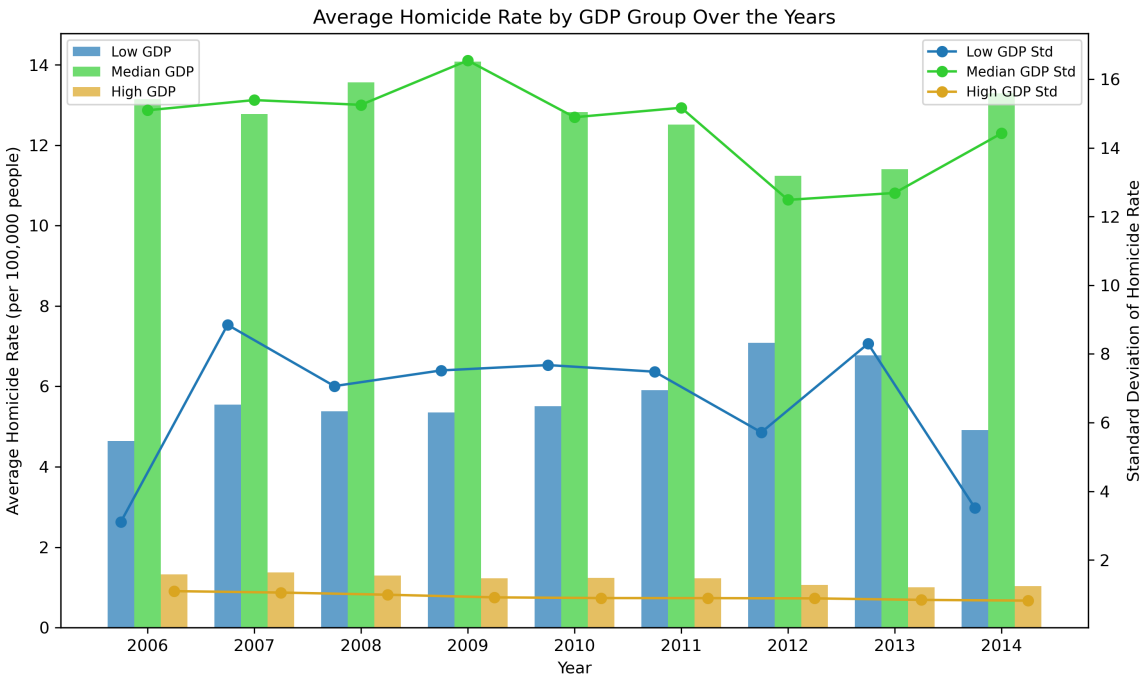


Figure 5.6:

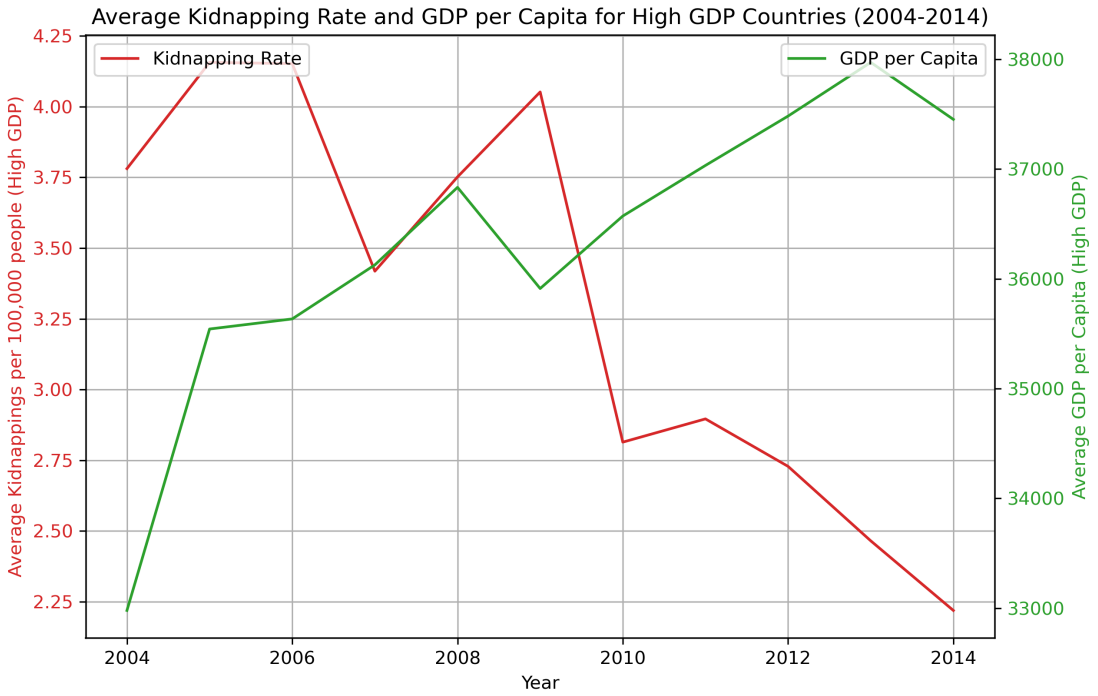


Figure 5.7:

??

5.1.3 Discussion

One of our sub-questions was **Do countries with low GDP per capita correlate with an increase in theft than higher ones?** To answer this question fairly our evaluated the mean thefts rate for each year for each dataset and see which ones are increasing and which ones are decreasing. [Note: To see countries included in each dataset group see 6]

5.1 Shows a line graph of the mean theft rate for each group from 2003 to 2013. It is good to note that the thefts were normalised to counts per 100,000. Looking at the chart, the average theft rate seems to be very consistent throughout the years. We can see from the chart above that the during the years 2003 - 2005, there is a notable drop in the theft counts for high and median GDP per capita countries, this does not seem to be the case for Low GDP per capita countries.

For the most part for all groups the theft rate is decreasing or remaining consistent throughout the years, so this immediately answers our question 'Do Low GDP per capita countries have an increasing theft rate?'. The final answer based on the information given is : No Low GDP per capita countries do not have an increasing theft rate.

My next sub-questions was **Can low GDP countries be considered a contributing factor to the overall homicide rate?**

5.2 shows a grouped bar chart of the average theft rates and homicide rates for each GDP per capita group for the year 2013. Looking at 5.2 we can obviously see High GDP countries have the most average theft counts per 100,000, and median GDP countries have the highest average

homicide rate per 100,000.

Theft Rates: High GDP countries tend to have higher average theft rates per 100,000 people compared to low and median GDP countries. Low GDP countries have the lowest.

The observation from the information is that low GDP countries have lower theft rates contradict our hypothesis that low GDP per capita correlates with an increase in theft rates globally. Instead my data shows that high GDP per capita countries have significantly higher theft rates than low and median GDP per capita countries. Low GDP thefts have an incredibly low theft. This analysis is very similar for other years, not just 2013.

For homicides, Median GDP per capita countries exhibit the highest average homicide rates per 100,000 people, followed by low GDP per capita countries, and then high GDP per capita countries. So it is already good to note that low GDP values can definitely be considered a factor to the homicide rates, or at the very least a lot more than high GDP countries. Let's examine another graph.

5.3 (High GDP per capita), 5.4/(Median GDP per capita) and 5.5/(Low GDP per capita) all show the average GDP per capita and average homicide for each subset dataset. We constructed this to see if high GDP per capita countries have a correlation with a decrease in homicides or if other GDP per capita groups do.

Continue Looking at the 3 Line graphs the influence of GDP per capita is the greatest on median GDP per capita countries, as GDP per capita increase, the homicide rate is decreasing. This is mainly shown from the year 2009 onwards on 5.4. It suggests that there might be a negative correlation between GDP per capita and homicide rate for median GDP per capita countries.

This observation could imply that as the economic prosperity of a country (in the median GDP per capita range) increases (reflected by higher and growing GDP per capita), there might be factors such as better law enforcement, improved life expectancy, and increased access to education and healthcare, which collectively contribute to a decrease in homicide rates. So we can not definitively say that median GDP per capita countries are a contributing factor to the homicide rate but we can definitely say countries median GDP per capita have a strong correlation to the homicide rates in their group. As for low and high GDP per capita countries there does not seem to be any sort of moderately strong correlation between the average GDP per capita and average homicide rate for each of the respective line graphs.

5.6 shows a grouped bar chart of the average homicide rate from the years 2006 to 2014. From the graph above we can clearly see that median GDP countries contribute the most to the global homicide rate by a lot more than low GDP per capita countries. On the other hand the bar graph shows that high GDP per capita countries have a very low homicide rate that is very consistent throughout the years. It is safe to say that these countries definitely have very little bearing on the overall global homicide rate. The line chart of the graph shows the standard deviation of the mean homicide rate for each group, which means in simpler terms the distribution of the dataset from the mean. In most cases the standard deviation is more than the mean for Low GDP per capita countries, but in High GDP per capita countries it is always lesser than the average. Let's take the Median GDP per capita bars in the graph as an example, 2006 specifically. The average is approximately 13 and so is the standard deviation for that year. This means the distribution of data, particularly homicides, for that year in that dataset is +13 from the mean or -13 from the mean.

So the question 'Can low GDP values be a factor to overall homicide rates?'. How can we answer this based on the grouped bar chart, 5.6?

- We can begin by saying low GDP values can be a factor considered a factor to the overall

homicide rates, but the biggest contributors to the global homicide rates are countries within the median GDP per capita group. These median GDP per capita countries, on average, have twice the number of average homicides than low GDP countries.

- What we can say about high GDP countries is that they cannot be a considers a strong factor or even moderate one to the global homicides rates because of how low it is compared to the other 2 datasets. These high GDP per capita countries also have a low consistent average homicide rate through the years. Therefore, it can be concluded that high GDP per capita countries have very little impact on the overall global homicide rates. Also it can be inferred that median GDP per capita countries have a significant influence on the overall homicide rates.

Let's examine one last graph before we can answer our question.

Looking at the graph on 5.7 , this is a line chart of the average kidnapping rate and GDP per capita for High GDP per capita countries. Looking at 5.7 we can see there is some correlation between the average kidnapping rate and High GDP per capita countries. We can observe this clearly from the year 2008 onwards, as the GDP increases the theft rate declines simultaneously and stays depleting onwards while the GDP per capita increases. At the very least we can say that there seems to be a correlation between kidnappings and GDP per capita from 2008 onwards.

So based on all this information how can we answer our question: **Does a high GDP per capita correlate with a decrease in violent or assault crimes?**

Based on the information from previous analysis, we can conclude that a high GDP per capita does correlate with a decrease in violent or assault crimes. This is evident from the observation that high GDP per capita countries have consistently lower average homicide rates and a lower average kidnapping rate compared to low and median GDP per capita countries.

Specifically, the analysis shows that high GDP per capita countries tend to have higher average theft rates compared to low and median GDP per capita countries. We need to note that this is not a violent or assault crime! This suggests that economic prosperity, being GDP per capita rate, may lead to a higher incidence of theft. However, when it comes to homicides, median GDP per capita countries exhibit the highest average homicide rates, followed by low GDP per capita countries, and then high GDP per capita countries. This indicates that low GDP values can be considered a contributing factor to the overall homicide rate, particularly in median GDP per capita countries, but NOT high GDP per capita countries

Furthermore, the line chart for high GDP per capita countries shows a clear correlation between increasing GDP per capita and a decreasing kidnapping rate from 2008 onwards. This could imply that as the economic prosperity of a country increases, there may be other factors such as improved law enforcement, better social infrastructure, and increased access to education and healthcare that contribute to a decrease in violent or assault crimes.

In summary, the analysis suggests that high GDP per capita countries generally have lower rates of violent or assault crimes compared to low and median GDP per capita countries. However, it's important to note that correlation does not imply causation, and there may be other factors at play in determining crime rates.

5.2 RQ2: Is there an immediate change in the crime rate after a change in GDP per capita per capita or is there a lagging phase?

In this study, we aim to investigate the temporal relationship between changes in GDP per capita and subsequent changes in crime rates. Building on the hypothesis that economic fluctuations can influence crime levels, our research seeks to determine whether these changes occur immediately or if there is a lagging phase. To address this question, we will analyze historical data on GDP per capita and crime rates from a diverse range of countries and regions. We will employ econometric methods to model the relationship between GDP per capita fluctuations and variations in crime rates over time. Specifically, we will examine whether changes in GDP precede immediate changes in crime rates or if there is a delay before their effects become apparent. Furthermore, we will identify which types of crimes exhibit the most significant delays in response to changes in GDP per capita. By analyzing data on various types of crimes, including corruption, theft, cyber-related crimes, and others, we aim to uncover patterns in the temporal dynamics of crime. Additionally, our analysis will explore geographic variations in the timing of these responses. We will investigate whether certain continents or regions exhibit more pronounced delays in crime rate changes following fluctuations in GDP per capita. Finally, we will identify the top 10 countries that experience the most significant delays in crime rate responses to changes in GDP.

5.2.1 Data & Method

These are all the data columns it was used for this question, some of them were already explained previously in this report.

- GDP per capita current U.S. dollars (normalized value of GDP): This represents the Gross Domestic Product per person within a country, converted to U.S. dollars
- Country: The name of the country where the data is collected.
- Country code: A code that uniquely identifies a country. This can be a two-letter code, such as (EU: Europe, AF: Africa, AS: Asia, OC: Oceania, SA: South America)
- Crime index: A composite index that includes various types of crimes to give an overall measure of the crime level in a country.
- Robberies per 100000 people: The number of reported robberies per 100,000 people
- Thefts per 100000 people: The number of reported thefts per 100,000 inhabitants. This generally includes acts of petty theft and grand theft excluding vehicle theft
- Homicides per 100000 people: The number of reported homicides per 100,000 people
- Cyber related crime
- Environment crime: Crimes that involve environmental harm such as acts against floura and fauna
- Money laundering
- Smuggling of migrants
- Total Sexual Violence at the national level, rate per 100,000

- Kidnappings per 100000 people

The basic idea of these questions was to figure out how effectively we can measure the delay phase and immediate changes between economic indicators and crime statistics. By analyzing these dynamics over specified time periods and across different countries, we seek to uncover patterns that inform our understanding of how economic prosperity or downturns influence crime rates.

To answer these questions, we will use the function np.correlate, which serves as a powerful tool to measure the degree of correlation between two time series datasets. This function computes the cross-correlation of two sequences offering insights into how one series may be affected by shifts in another over time. By applying np.correlate, we can identify the lag times at which the relationships between economic indicators and crime statistics are the strongest. This approach allows us to pinpoint the exact moments when changes in GDP per capita potentially trigger variations in crime rates, effectively highlighting the immediate and delayed responses within the data.

Building upon the analysis provided by np.correlate, we further refined our approach by developing the draw_plot_and_corr function. This enhanced function not only applies the cross-correlation analysis but also integrates it with sophisticated visualization techniques. By doing so, draw_plot_and_corr simultaneously presents the GDP and crime data over time through dual-axis plotting, allowing for a clearer and more comprehensive understanding of the data's dynamics. This visualization makes it easier to observe the direct and lagged relationships between economic indicators and crime rates.

Defining the Function: The function draw_plot_and_corr is central to the analysis. It accepts several parameters to tailor the analysis to specific needs:

- country: Specifies the country for the analysis.
- crime: Defines the type of crime to analyze.
- start_year and end_year: Sets the time frame for the analysis.
- is_per_100000: A boolean that determines whether to scale crime data to per capita people for normalization purposes. That was specially added to allow added other indicators in a future, and be sure that function can work on them.

How the function works? At first, function filters the relevant data columns from a comprehensive dataset based on the country and time period. If the is_per_100000 flag is true, it adjusts the crime data to reflect rates per million people, facilitating accurate comparisons and standardizing across varying population sizes.

Secondly, Prior to proceeding, the function verifies the availability of the required data. If the data for the specified parameters are not available, it notifies the user and terminates to prevent errors in further processing, which avoids of incorrect representation.

Plotting GDP and Crime Data Over Time: The function uses a twin-axis plot to simultaneously showcase GDP per capita and crime rates over time. One y-axis illustrates GDP per capita, while the other displays crime rates, visually representing the evolution of these variables within the selected period.

Calculating and Displaying Correlation: The function calculates the cross-correlation between the time series of GDP per capita and crime rates to determine how fluctuations in one variable may correspond to changes in the other. The correlation is normalized and graphically depicted, emphasizing the strongest correlations. Specifically, the peak value is calculated by identifying

the highest point on the normalized correlation curve, which represents the maximum degree of correlation between the datasets. This peak value precisely indicates the time lag at which the correlation between changes in GDP and crime rates is strongest.

After we understood how our function worked, we used it to analyze data from different countries and types of crime. We created a loop that went through all the countries and crimes we were studying, and made a new table that shows changes in crime rates for each country and crime type. Then, we looked more closely to find out which types of crime usually have the biggest delays. We also grouped countries by continent to see if there were any common patterns in crime across different parts of the world.

5.2.2 Results

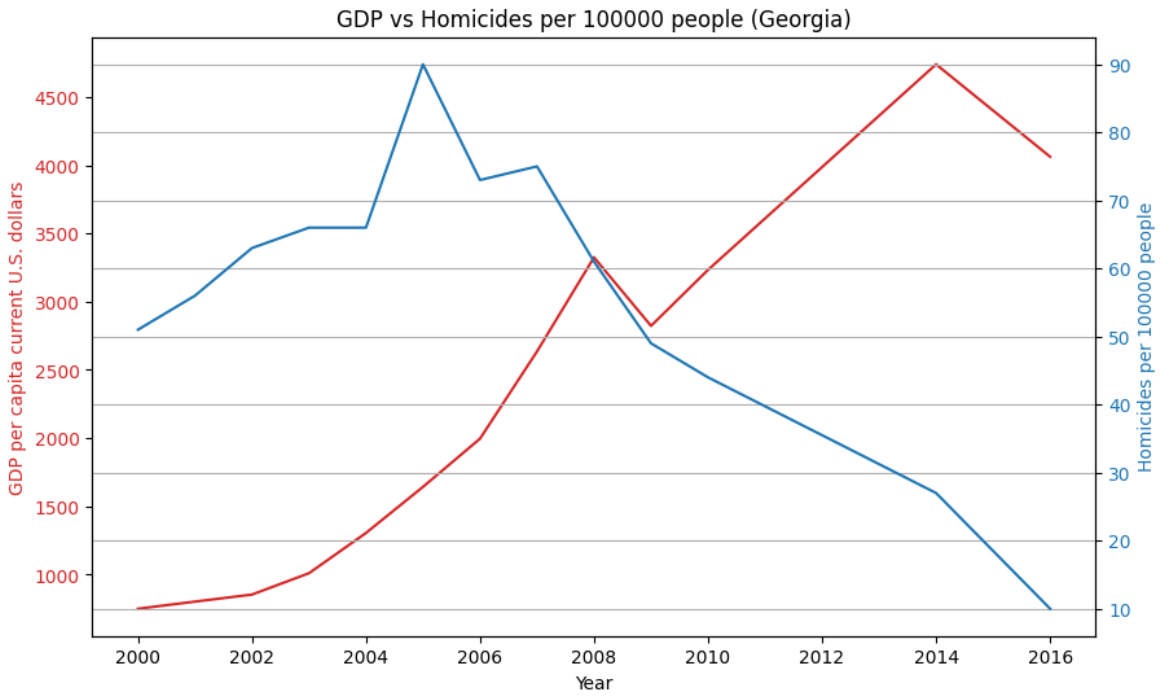


Figure 5.8: Example of function.

```
Peak Offset: 5.333333333333332
```

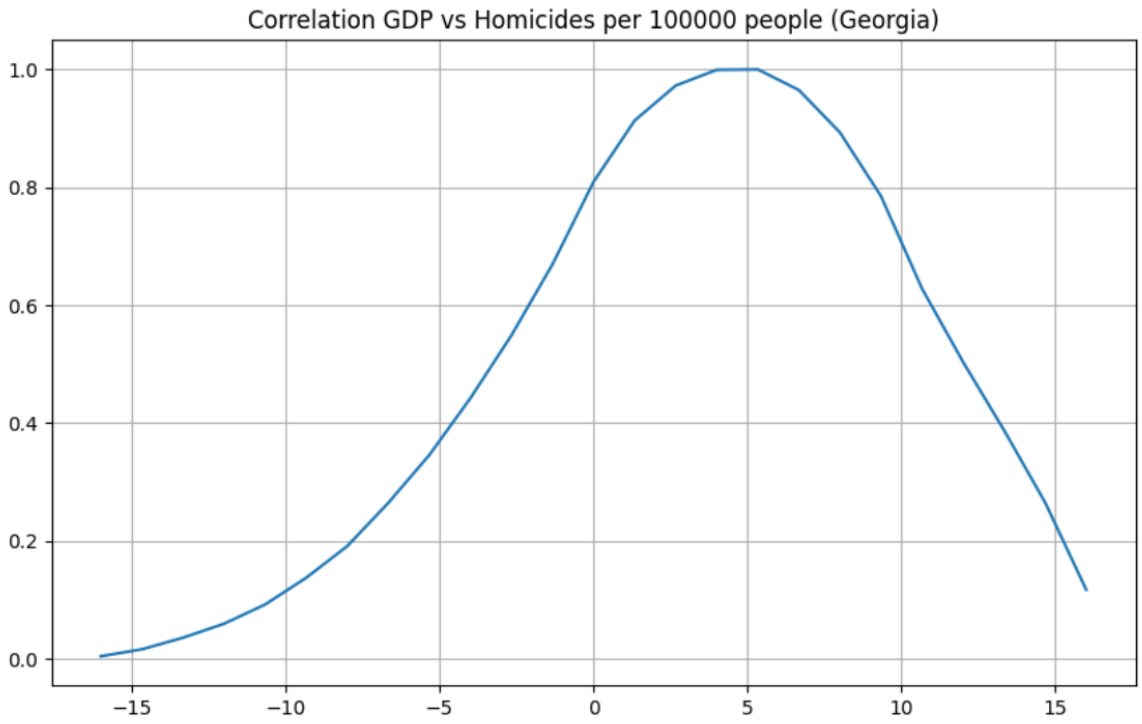


Figure 5.9: Peak offset representing.

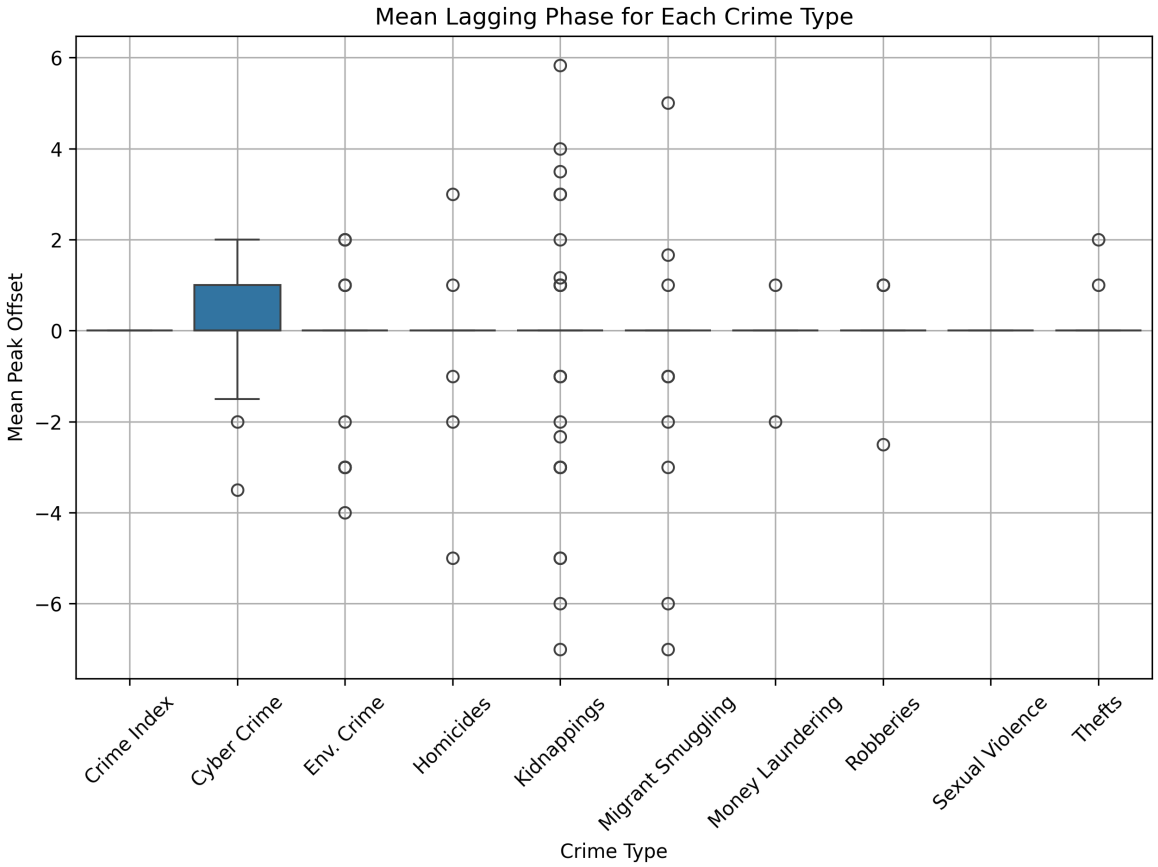


Figure 5.10: Lagging phase for each type of crime.

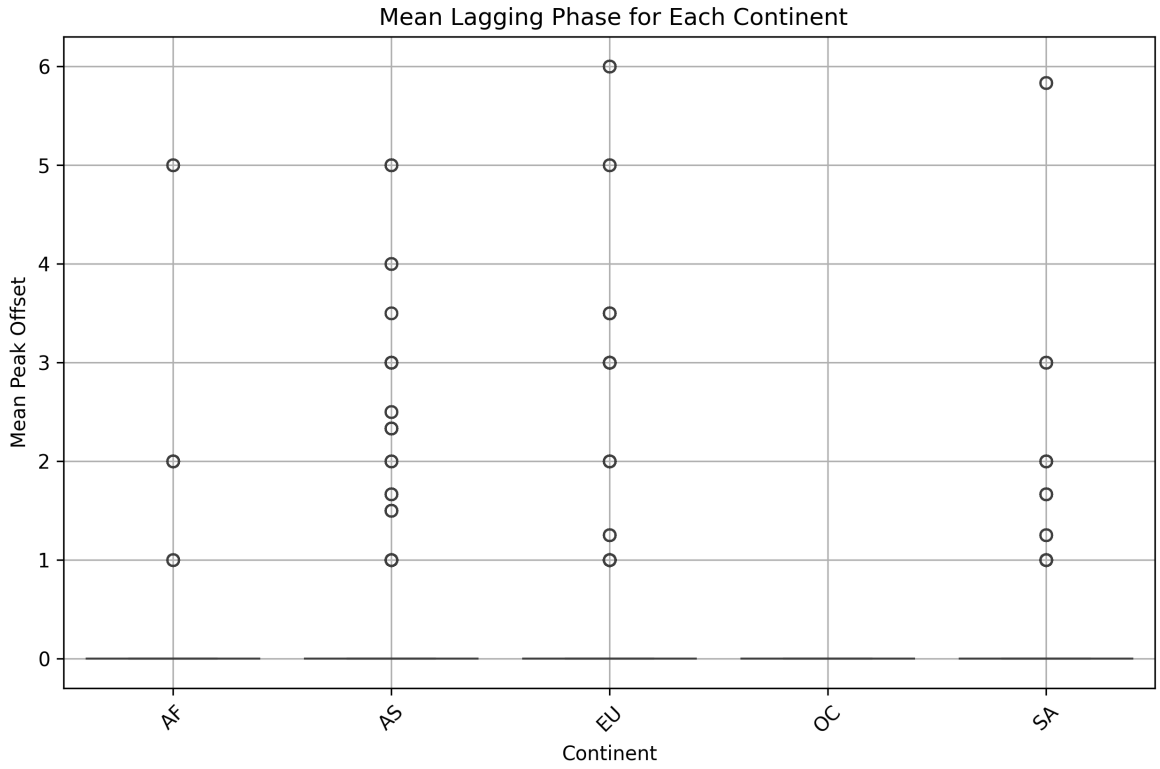


Figure 5.11: Lagging phases across continents.

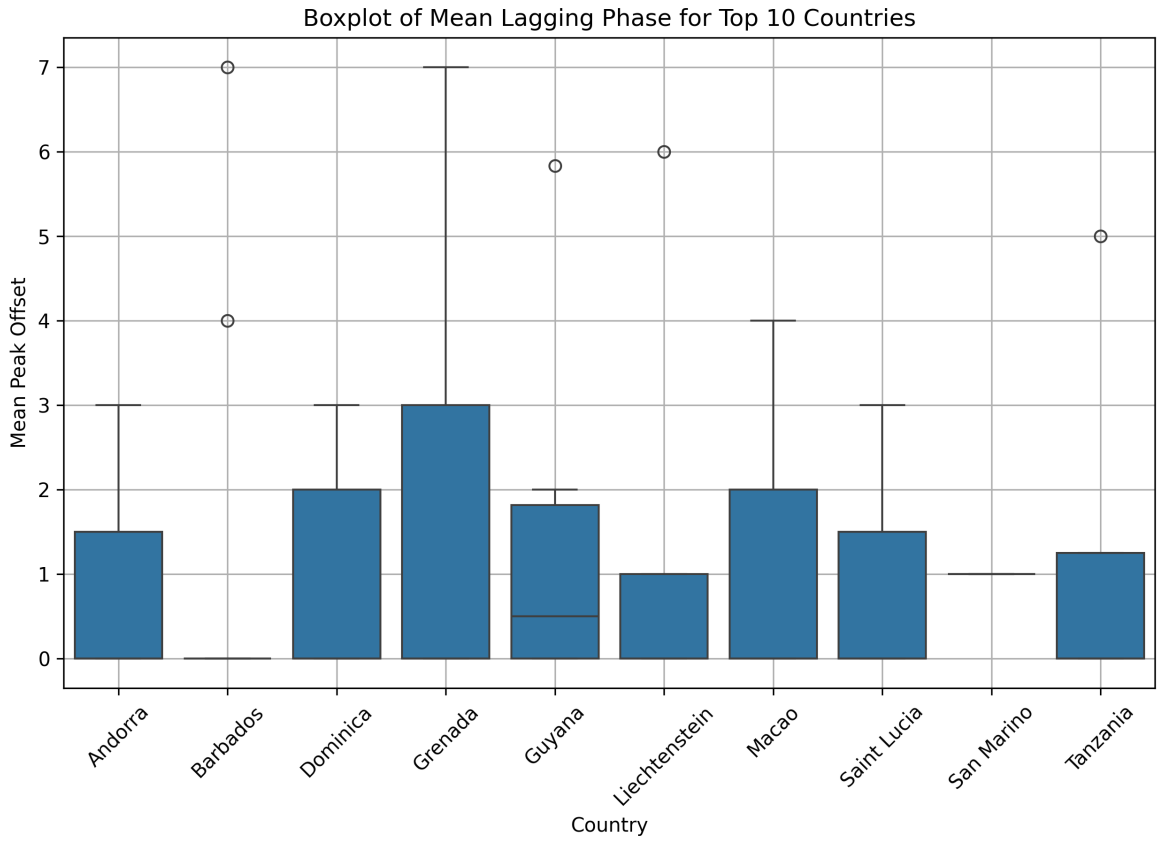


Figure 5.12: Top 10 countries for lagging phase

5.2.3 Discussion

As represented in Figure 5.8, the lag function reveals the temporal delay between GDP per capita and homicides per 100,000 in Georgia. Both line charts initiate in the year 2000. Remarkably, the homicide rate begins a significant decline around 2005, even as the per capita income continues to grow. This relationship is clearly illustrated in the correlation graph (Figure 5.9), where in the top left corner, a peak offset value of 5.3 can be observed. This offset translates to approximately 5 years and 4 months, providing a measure of the lag between economic growth and changes in the homicide rate.

From box-plot 5.10 we can observe that, Crime Index and Sexual Violence displayed no temporal delay, with all values consistently at zero. This suggests an immediate or non-existent correlation with economic changes, implying that factors other than wealth per person may predominantly influence these crimes. Cyber-related Crime and Money Laundering showed slight to moderate delays, with most values ranging from zero to one, and outliers indicating potential significant delays in certain situations. This could reflect the complex nature of these crimes and the varied mechanisms through which economic conditions affect them. Environmental Crime, Homicides, and Robberies were mostly clustered around zero, but each displayed outliers suggesting occasional significant delays. These findings indicate that while these crimes generally respond quickly to economic changes, under certain conditions, the response can be notably delayed. Kidnappings and Smuggling of Migrants demonstrated a broad range of responses, with numerous outliers significantly above and below zero, indicating that these crime types might be influenced by a complex set of factors that cause varied delays in response to economic conditions. Thefts generally showed an immediate response, similar to the majority of crime types studied, with only a few instances of delays.

Importantly, the direction of peak offset values provides insight into the causality and sequence of economic and crime rate changes. Positive peak offset values indicate that changes in crime rates follow shifts in GDP per capita, suggesting a reactionary dynamic where crime rates adjust after economic conditions change. Conversely, negative values imply that changes in GDP per person follow shifts in crime rates, indicating that economic conditions may sometimes react to crime dynamics.

The findings from our study provide compelling evidence that the majority of crime types respond almost instantaneously to economic shifts. For Crime Index, Sexual Violence, Environmental Crimes, Homicides, and Robberies, the predominant clustering of peak offset values around zero illustrates a rapid reaction to changes in wealth per citizen. This pattern suggests that as economic conditions fluctuate, the effects on these crime rates are immediate, underscoring the sensitivity of these crimes to economic environments.

Discussing the lagging phase across the continents, as depicted in Figure 5.11, we observe varied dynamics. Starting with Africa (AF), the data indicates that crime responses generally occur with minimal delay, as most values are clustered close to zero. This suggests that GDP per person fluctuations have an almost immediate impact on crime rates across the continent. However, there are notable exceptions, with a few significant outliers at 1, 2, and 5, pointing to occasional substantial delays in specific crime responses to economic changes. In Asia, the responses are also centered around zero, suggesting a quick response to economic changes. However, there is a notable spread of outliers, both negative and positive, ranging from -1.6 to 5, which indicates that some specific crime types or regions within Asia may experience delayed effects. Europe's data mirrors the pattern seen in Asia with responses mostly clustering near zero, implying prompt reactions to economic shifts. The presence of outliers at values up to 6, however, suggests occasional significant delays. Oceania stands out with all responses exactly at zero, indicating no delay between GDP per capita changes and crime rate responses, which might suggest a strong alignment between economic conditions and crime rates. Similar to other regions, most responses in South America are near zero, but there are outliers extending up to 5.8, indicating some crimes

respond with more significant delays.

Our study of how quickly different continents respond to economic changes shows that, in general, crime rates tend to react almost immediately when GDP per capita changes. This fast response is seen almost everywhere, highlighting that crime is often closely linked to economic conditions. Yet, not every place follows this pattern exactly — there are some exceptions. In every continent except Oceania, we found outliers where crime responses were delayed.

These delays are especially varied in Asia and Europe, where they range widely. This variation might be due to how different places manage their economies, enforce laws, or even due to cultural differences that affect how quickly crime responds to economic shifts.

Based on the mean peak offsets for the top ten countries 5.12, it appears that most of these nations are relatively small in size, such as Andorra, Barbados, Grenada, Liechtenstein, and San Marino. This common characteristic might influence how swiftly governmental policies can respond to economic changes, potentially causing a delay in the impact of these policies on economic conditions and subsequently on crime rates. The delays in these countries' responses to changes in GDP per capita could be attributed to the challenges and dynamics of policy implementation and governance in smaller nations.

In conclusion, our study indicates that changes in GDP per capita do not consistently lead to immediate changes in crime rates. Although many responses appear to have little or no delay, there are notable exceptions where crime rates respond much later. This variation suggests that the impact of economic changes on crime rates differs depending on the type of crime and the geographic location.

5.3 RQ3: Do countries with a high corruption score have fewer cases of robbery/theft than countries with low corruption score.

The objective of this study is to examine the correlation between corruption levels and specific crime rates in various countries. We focus on understanding how a nation's perceived corruption influences its incidents of robbery and theft. Here, robbery is defined as taking someone's property through force or threat, while theft involves the unauthorized taking of someone's property without the use of force against the victim. This research also explores whether high corruption could skew the reporting of these crimes. Additionally, we will develop predictive models based on historical data to determine if changes in corruption levels can predict future trends in robbery and theft rates.

5.3.1 Data & Method

These are all the data columns it was used for this question, some of them were already explained previously in this report.

- Country: The name of the country where the data is collected.
- Corruption score: Each country is scored on a scale, with higher values indicating greater

corruption

- Years: 2003 - 2017
- Robberies per 100000 people: The number of reported robberies per 100,000 people
- Thefts per 100,000 population

The foundation of my approach to determining corruption levels is the function categorize_corruption. This function assigns one of three ranks to each corruption index value:

- Rank 3: High level of corruption. This is assigned if the corruption index is above the upper quartile (Q3). It means that the corruption level is among the top 25% compared to other countries for that same year.
- Rank 2: median level of corruption. Assigned to values that fall between the lower (Q1) and upper quartiles. It indicates that the corruption level is between the lowest 25% and the highest 25%.
- Rank 1: Low level of corruption. This is assigned if the corruption index is below the lower quartile (Q1). This means that the corruption level is among the lowest 25%.

For each year, the data are grouped, and I calculate the corruption rank for each country in that year. This is done using the apply method, which applies the categorize_corruption function to the corruption_level column. The results are stored in a new column, Corruption Rank. After assigning a rank to each country's record for each year, I can calculate the average corruption rank for each country. This is done by grouping the data by country and then calculating the mean corruption rank. The resulting average value is rounded to the nearest whole number, providing a generalized view of the corruption level for each country over the considered period.

To ensure a fair comparison, particularly given that the group of median-corruption countries had the highest number of data points, I implemented the following adjustments in my analysis. For each year in my dataset, I filtered the data to include only records from that specific year. I calculated the total number of robberies per 100,000 people for countries identified as highly corrupted. I determined how many records (countries) were in this group to calculate the average correctly. If there were records, I calculated the average robberies per 100,000 people; otherwise, I set it to 0 to avoid division by zero. I repeated the same process for median and low corruption groups to ensure consistent analysis across all groups and did the same for theft crime.

To predict future crime rates, specifically robberies and thefts per 100,000 people, and to explore the potential correlation with corruption scores, I did prepared some steps. Firstly, I merged two datasets: one containing crime statistics and another detailing corruption levels and ranks by country and year. This merging was crucial to align the data and create a comprehensive dataset for analysis.

Next, I prepared the data for machine learning by selecting relevant features—specifically, the corruption level and corruption rank from the merged dataset—as predictors. These features were used to set up the independent variables (X), while the dependent variables (y) were the robbery and theft rates.

I then split the data into training and testing sets to ensure the model could be validated independently. This split helps in evaluating the model's performance more accurately by training on one set of data and testing on another. For the predictive modeling, I employed the RandomForestRegressor, a machine learning algorithm suitable for regression tasks. This model was trained on the training data and then used to predict crime rates in the testing set.

5.3.2 Results

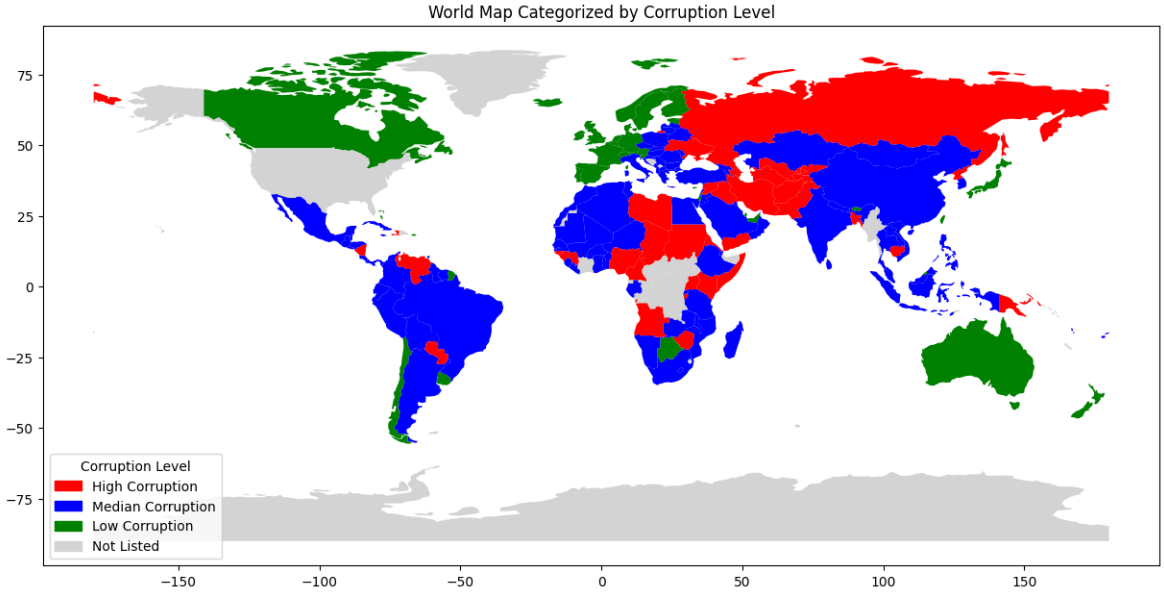


Figure 5.13: Coloured map for grouped countries

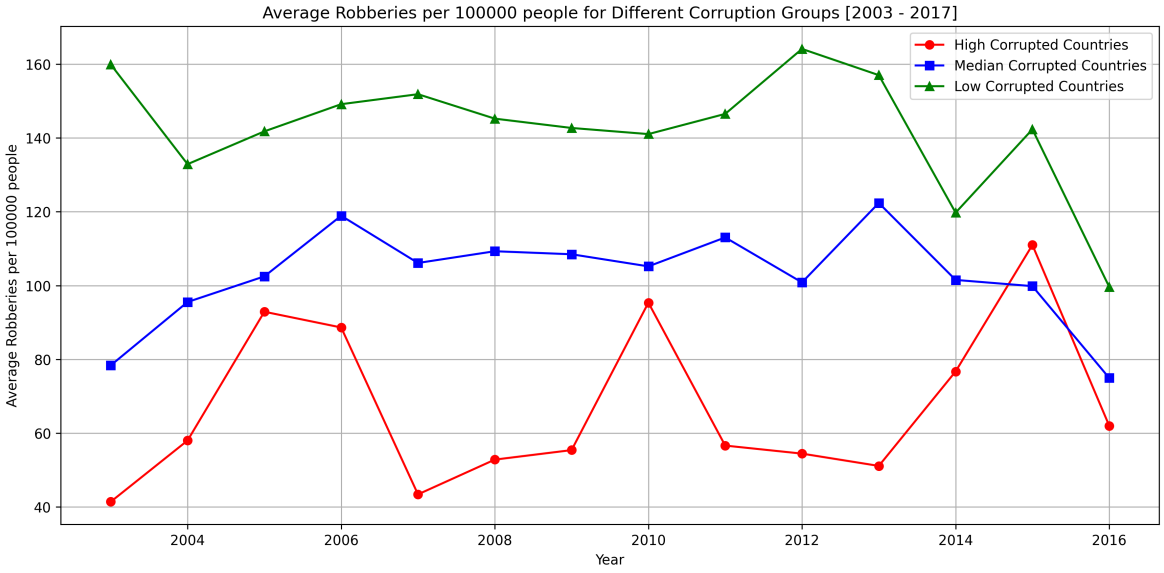


Figure 5.14: Robbery across different corruption groups

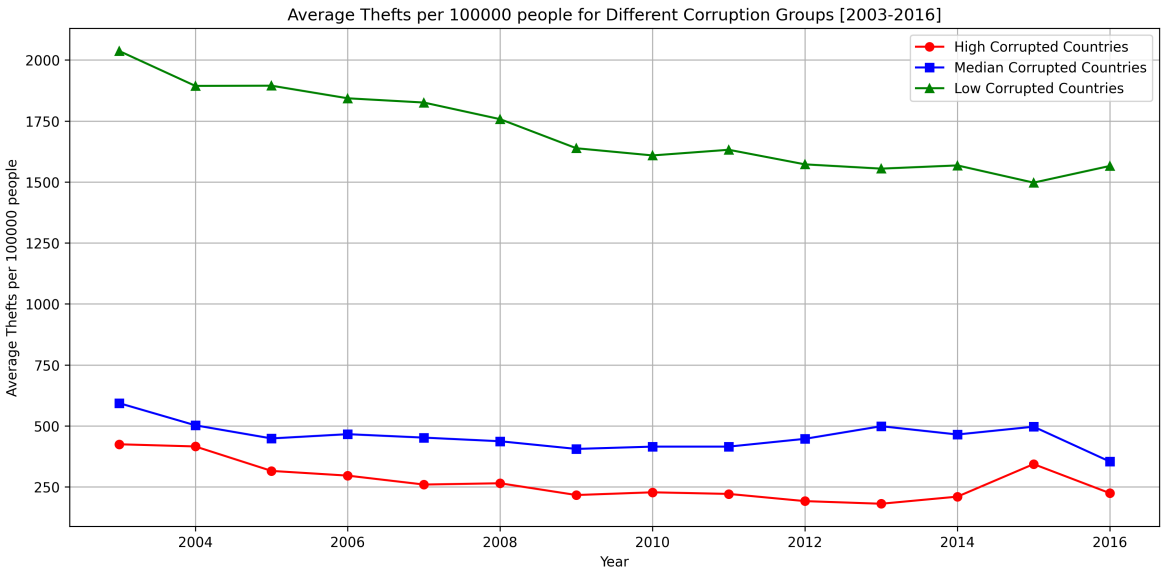


Figure 5.15: Theft across different corruption groups

5.3.3 Discussion

This map 5.13 shows different levels of corruption around the world using colors. Countries in red have high corruption, blue countries have median corruption, and green countries have low corruption. Grey areas show countries not included in the data. For more detailed information, the countries in each group are listed 6

From line chart 5.14 we can observe that generally, countries with low corruption levels showed a higher frequency of robberies per 100,000 people compared to countries with higher corruption levels. This might initially seem impossible, but it could be explained by better reporting and law enforcement transparency in less corrupted countries. Countries with median levels of corruption consistently reported moderate robbery rates across the years. There was a relatively stable pattern, though a slight increase was noticeable over time. High corrupted countries generally reported the lowest robbery rates, which could reflect underreporting or less effective law enforcement and data collection mechanisms due to higher corruption

The general trend seems to be fluctuation across the years in all categories, but particularly noticeable is the variance in the high corruption group, where robbery rates appear less consistent, perhaps echoing the instability and variability in governance and law enforcement in these regions. The data consistently shows lower reported robberies in more corrupt countries and higher in less corrupt ones.

Talking about theft which represents in 5.15, we can conclude that ountries with low corruption levels reported significantly higher theft rates compared to those with median and high corruption levels. This trend suggests better crime reporting and possibly higher economic activities, leading to more opportunities for theft. Generally, there's been a slight decrease over time, but rates remained relatively high. The median corrupted countries exhibited moderately not really high theft rates, with some fluctuations over the years. The highest reported value was in 2015, with around 498.88 thefts per 100,000 people. The trend in median corruption countries does not show a clear decrease or increase but rather fluctuates around the 400-500 thefts range. The theft rates in highly corrupted countries were consistently the lowest among the three groups. This could be due to underreporting or less efficient law enforcement and judicial systems. The highest rate was initially in 2005, at 424.8 thefts per 100,000 people, showing a general decline with some minor

increases in certain years like 2012 and 2016.

Results for prediction: The predictive model was built using a RandomForestRegressor, leveraging two primary features: corruption level and corruption rank. For the robbery the R-squared R^2 value from the model was approximately 0.0968. This indicates that about 9.7% of the variance in robbery rates is explained by the model. This is a relatively low value, suggesting limited predictive power of the model with the current features. The Mean Squared Error (MSE) was approximately 61907.29. This high MSE further points to the model's limited accuracy in predicting robbery rates based on the given features. The results indicate that the model, with corruption as a primary feature, has limited capability in predicting robbery rates accurately.

For the theft prediction the case is more successfully, The R-squared R^2 value for the theft model was approximately 0.454. This indicates that about 45.4% of the variance in theft rates is explained by the model, a significantly higher value than what was observed in the robbery model. This suggests a stronger relationship between the corruption metrics and theft rates. The Mean Squared Error (MSE) was approximately 397910.06. Although this MSE is high, suggesting some inaccuracies in specific predictions, the higher R^2 value indicates a more useful model compared to the robbery prediction.

The model's better performance in predicting thefts suggests that corruption metrics might be more closely related to theft than to robbery. This could reflect differences in the nature of these crimes; thefts might be more influenced by systemic factors such as corruption, possibly due to how thefts are reported or policed. We should be aware, that both "corruption level" and "corruption rank" were used as features, one key consideration is their potential redundancy. Which could lead to the variable overlap, which might be affecting the model's performance.

Overall, while high corruption countries report fewer robberies and thefts, this trend is likely influenced by underreporting and the inefficacy of law enforcement, rather than reflecting true crime levels. The predictive models confirm that corruption can be a significant predictor of theft rates, but less so for robberies.

5.4 RQ4: Are there significant differences in the rates of cyber-related crimes in countries with varying GDP levels, and if so, what factors contribute to these disparities?

This research inquiry seeks to determine whether GDP per capita is the predominant factor influencing cybercrime rates in each country, or if there are other macroeconomic factors that exert a more significant impact. Specifically, it explores the possibility that variables beyond GDP per capita, such as the Human Development Index, the Inequality Index, inflation rates, and possibly others, may substantially contribute to shaping observed cybercrime rates. This study aims to broaden the understanding of how various economic conditions correlate with cybercrime across different national contexts.

5.4.1 Data & Method

These are all the data columns I used for this question, some of them were already explained previously in this report.

- Country
- Years
- GDP
- GDP per capita
- Population size in millions
- Continent Code for analyses across continents
- Kidnappings per 100000 people
- Cyber crimes data
- Cyber crimes per capita
- Life Expectancy
- Police Officers per 1,000 people
- electdem_vdem_owid - democracy score
- Inflation Rate
- Level of Happiness
- Human development index
- Inequality rate
- Unemployment rate

The research into the disparities in cyber-related crime rates among countries with varying GDP levels required the collection and analysis of a substantial dataset. Here's a step-by-step overview of how the data was prepared and analyzed: The initial dataset included a broad range of variables for various countries over many years, including GDP, population, cyber-related crimes, and other socio-economic indicators. The data was filtered to focus on the years from 2012 onwards to reflect more recent trends and changes in cybercrime and GDP. Cybercrime data was transformed to get average values and total counts for each country to analyze trends over time. Similarly, GDP data was cleaned and pivoted to show changes over the years for each country. The cybercrime data and GDP data were merged to facilitate a comparative analysis. Additional variables such as population were included to normalize the data and enable per capita analyses. Key metrics such as GDP per capita and cybercrimes per capita were calculated. These metrics are crucial for understanding the relative scale of cybercrime in relation to a country's economic size and population. The new named merged_data, now contains detailed information on cybercrimes, GDP, average population, GDP per person, and cybercrimes per person for each country.

We started our analysis by examining the distribution of cybercrimes across various regions from 2012 onwards. Using statistical tools, we identified trends and outliers in the regional data. Next, we compared the total number of crimes in each region against their average GDP to check for any economic correlations. After this economic comparison, we further explored other potential influencing factors. We used visualizations like boxplots and scatter plots to assess the relationships between cybercrime rates and various socio-economic indicators, such as GDP per capita, life expectancy, and happiness indexes.

5.4.2 Results

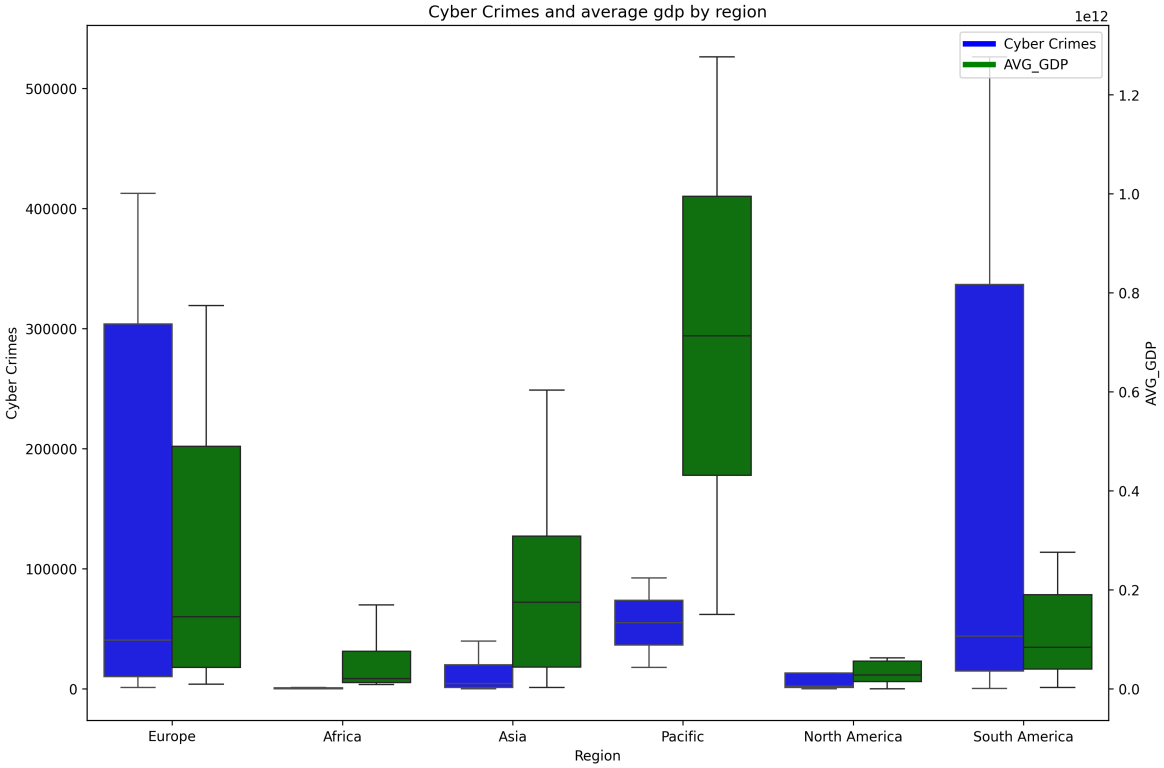


Figure 5.16: GDP vs. Cybercrime Across Continents

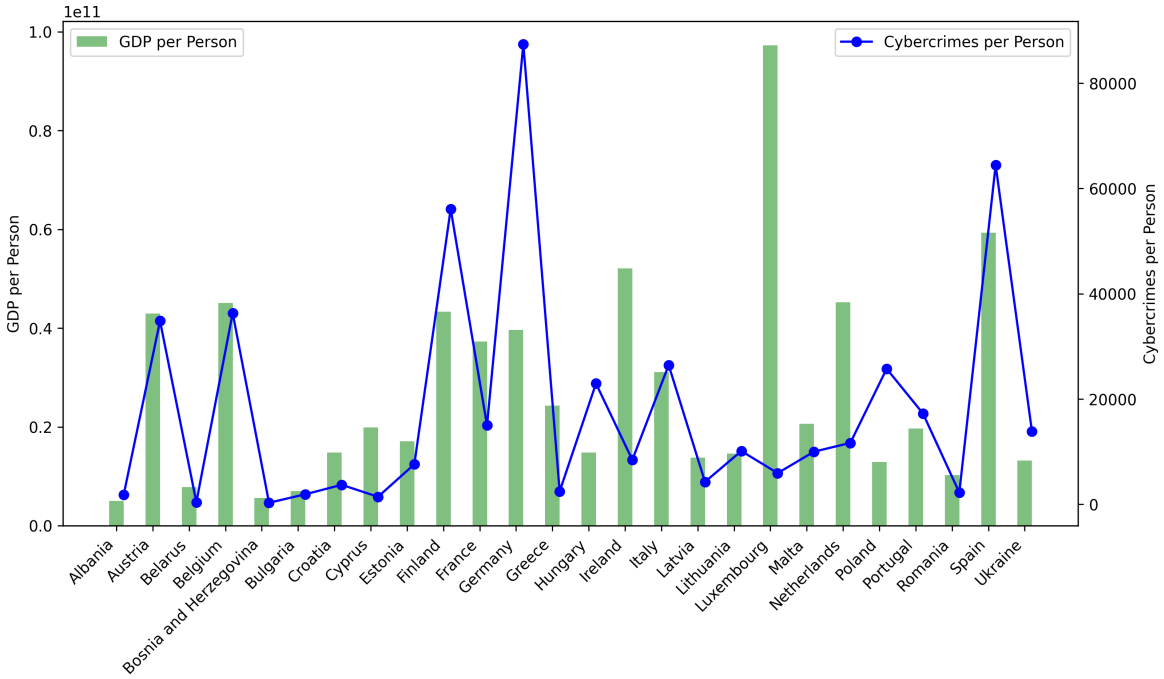


Figure 5.17: Comparison of GDP per capita and amount of cybercrimes per capita in Europe

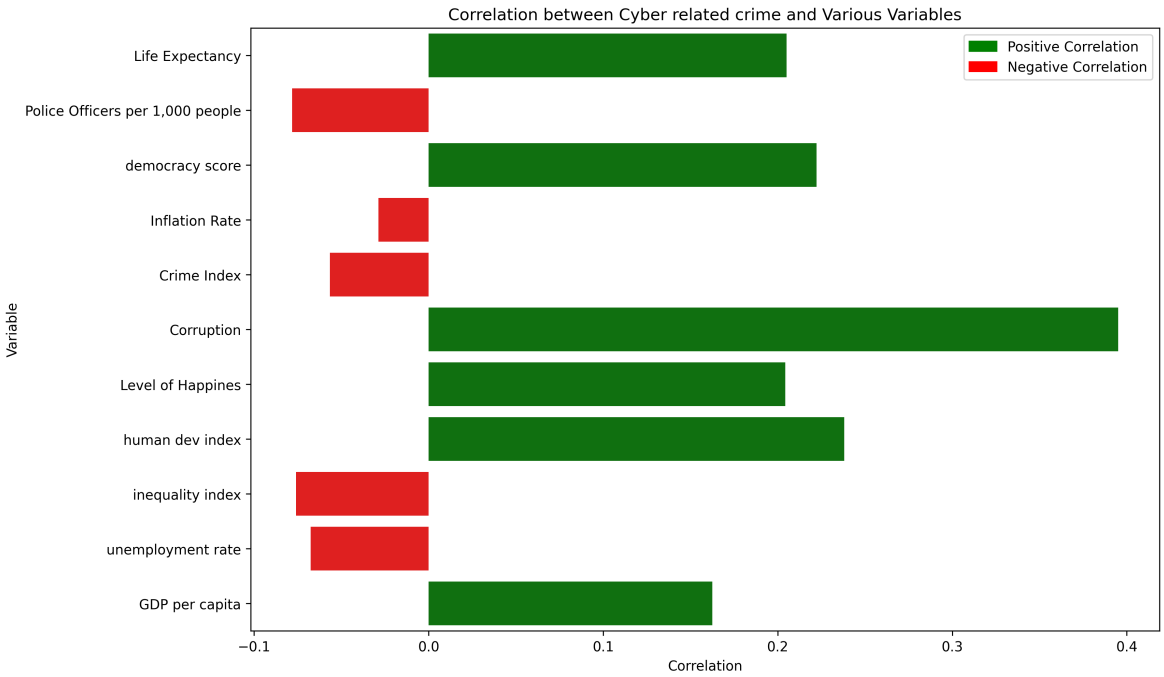


Figure 5.18: Correlations with other indicators

5.4.3 Discussion

At figure, representing in 5.16 we can observe a nuanced relationship between GDP and cyber crime rates across different continents. In Europe, a clear correlation is visible, where countries with higher GDPs tend to experience more cyber crimes. This might indicate that wealthier economies, with more digital infrastructure, are more frequent targets for cyber criminals.

However, in Asia and South America, the relationship between GDP and cyber crime rates does not follow the same pattern. In these regions, countries with lower GDPs also report fewer cyber crimes. This could suggest limited digital penetration or less reporting in these areas.

South America presents a unique case; despite having lower GDP figures comparable to some Asian countries, it records significantly higher rates of cyber crimes. This could point to regional variations in cyber security measures or legal frameworks that affect the reporting and occurrence of cyber crimes.

It's important to recognize that these observations are based on aggregate data and might not capture the complexities of individual countries where economic size does not directly correlate with cyber crime rates. Small countries with disproportionately high GDPs might skew the regional data, suggesting the need for a more granular analysis. Moving forward, the focus will shift to a per capita basis to gain a deeper understanding of the dynamics at play, providing a clearer picture of how individual prosperity relates to cyber crime vulnerability.

The second graph 5.17 visualizes the normalized values of GDP per capita and the incidence of cyber crimes per capita across European countries. Each country is represented with bars indicating GDP per capita (in green) and lines showing cyber crime rates per capita (in blue). This dual representation allows for an immediate visual comparison of economic output against cyber crime frequency within the same geographic context.

For many European countries, there appears to be a notable correlation between higher GDP per

capita and increased rates of cyber crime. This trend suggests that countries with more economic resources might also be more targeted or have better mechanisms for reporting cyber incidents. However, there are significant exceptions that underscore the complexity of this relationship. For instance, Luxembourg stands out with the highest GDP per capita in the region yet reports one of the lowest cyber crime rates, indicating factors other than economic wealth influencing cyber crime rates. Similarly, Ireland, despite its high economic performance, does not show a corresponding high level of cyber crime. These results highlight the multifaceted nature of cyber crime determinants, which may include factors like technological infrastructure, regulatory frameworks, and national cyber security policies.

Correlation, representing in figure 5.18 highlights the complex relationship between socio-economic indicators and cyber crime. Key findings include:

Corruption is the most influential factor, with a strong positive correlation suggesting that higher levels of corruption may facilitate more cyber crime. Human Development Index (HDI) and Democracy Score both show moderate positive correlations, indicating that regions with higher development and more democratic freedoms might experience more cyber-related crimes, likely due to greater digital engagement and infrastructure. Economic Indicators: GDP Per Capita shows a positive correlation, while Inequality Index and Unemployment Rate show slight negative correlations with cyber crime rates. Wealthier areas are more targeted for cybe crimes, whereas higher inequality and unemployment might reduce cyber crime occurrences due to less digital access. Crime Index shows a significant negative correlation, suggesting that regions with higher overall crime rates might underreport cybe rcrime or lack the necessary infrastructure to handle it effectively. Police Presence shows a slight negative correlation with cyber crime, indicating that increased traditional policing has limited impact on reducing cyber crime rates.

There are indeed significant disparities in cyber crime rates among countries with different GDP levels, influenced by a combination of economic, social, and governmental factors. Wealthier countries often face more cyber threats, which could be attributed to their greater digital exposure and economic incentives for cyber criminals. Therefore, while GDP is a significant factor, it is not the sole determinant of cyber crime rates. The multifaceted nature of cyber crime necessitates a broader understanding and approach, incorporating economic, political, and social dimensions to effectively manage and mitigate cyber risks.

5.5 RQ5:Is GDP per capita the sole factor for the crime rate each country respectively are there macroeconomic factors that influence the crime rate more than GDP (happiness index, unemployment rate and inflation, etc.)

This research question aims to investigate whether GDP per capita is the primary determinant of the crime rate in each GDP per capita group or if there are other macroeconomic factors that have a more significant influence. It explores the possibility that factors other than GDP per capita, such as the human development index, inequality index, inflation, and potentially others, may play a substantial role in shaping the crime rates observed. The happiness index and unemployment rate will be used to answer my question 6.

Our first sub-question is **What are the strong factors that have a more significant impact on certain crimes compared to GDP per capita, within each subset of the GDP per capita group?**

This sub-question delves into finding particular factors that exert a stronger influence on the crime index, then specific types of crimes compared to GDP per capita within different subsets of countries categorised by their GDP per capita levels. By examining these strong factors, the goal is to identify elements that have a more or similar pronounced effect on crime rates than GDP per capita, allowing for a deeper understanding of the complex relationship between macroeconomic indicators and crime.

The last sub-question is: **Is there any way to decrease certain crimes based on our analysis? For example, can we reduce certain crimes by addressing issues such as inequality?**

This sub-question focuses on leveraging the findings of the analysis to identify potential strategies for reducing specific crimes for each GDP per capita level. By understanding the relationships between macroeconomic factors and crime rates, the aim is to explore possible interventions or policies that can effectively address and decrease crime rates. One specific example highlighted in the question is the potential impact of reducing inequality as a means to mitigate certain types of crimes, such as homicides.

These sub-questions provide a more detailed exploration of our main research question, enabling a comprehensive analysis of the factors influencing crime rates beyond GDP per capita.

5.5.1 Data & Method

These are all the data columns I used for this question, some of them were already explained previously in this report.

- Country
- Year
- GDP per capita
- Crime Index: This column shows the crime index from 0-100,(1 being very safe and minuscule levels of crime, 90 being very dangerous/ abundance of criminal activity) which provides an overall measure of the crime rate in each country. It serves as an indicator of the general level of criminal activity.
- electdem_vdem_owid: This column represents a measure of the level of electoral democracy in each country. It provides insights into the political context, which can potentially influence crime rates.
- Inflation Rate: This column denotes the inflation rate, which measures the percentage change in the general price level over time. It serves as an economic indicator that can impact crime rates.
- human_dev_index: This is the Human Development Index (HDI), which combines indicators of life expectancy, education, and income
- inequality_index: This is the inequality index, which measures income or wealth inequality within a country

- Police Officers per 1,000 people: This column represents the number of police officers per 1,000 people in each country. It provides insights into the availability and capacity of law enforcement, which can influence crime rates
- Life Expectancy: This shows the average life expectancy in each country. It serves as an indicator of overall health and well-being
- Level_of_Corruption: This column represents a measure of the level of corruption in each country. It shows the integrity of the government
- Money_laundering: This column demonstrates the presence or extent of money laundering activities in each country.
- Homicides per 100,000 people
- Kidnappings per 100,000 people
- Thefts per 100,000 people

To answer the research question and sub-questions, I began by performing various analyses and visualisations using the provided data columns. The two main ones I used as the start were scatter plots and heat maps.

I initially split the total countries into low, median and high GDP per capita countries again using the quartile method

By using the subset groups of low, median, and high GDP per capita countries, I constructed heat maps to visualise the relationships between different crimes and economic factors. These heat maps provide a comprehensive overview of how crime rates vary within each subset group and highlight any patterns or correlations.

After examining the factors on a heat map for each subset group, I constructed scatter plots to examine the relationships between specific economic factors (such as unemployment rate, inflation rate, human development index, inequality index) and different types of crimes. These scatter plots allow for visual exploration of the associations between economic factors and crime rates, helping to identify any strong factors that affect certain crimes. They also allowed me to see if there was enough data for a year, or sequence of years to perform a fair analysis and come to a justifiable conclusion.

If there was enough data and moderate correlation(anything with a correlation coefficient over 0.2 or 0.3) with a certain crime, I noted the economic factor for the GDP per capita group to later perform an in depth analysis. Any factors that had a low correlation with a subset of the GDP per capita group, I did not waste time analysing as it would be very difficult to find any sort of correlation or relationship between the variables. Some common uses visualisation I perform after noting down some correlated factors were line graphs, bar plots and bubble charts. An example would be taking the mean theft rate and mean homicide rate of a GDP group and placing it on a line chart to find some sort of inverse relationship

By using these analyses and visualisations, I can use the used the provided data to explore the relationships between macroeconomic factors, and crime rates. The results derived from these analyses can be used to assess the relative importance of different macroeconomic factors in shaping crime rates and to explore potential strategies for crime reduction based on the identified relationships.

5.5.2 Results

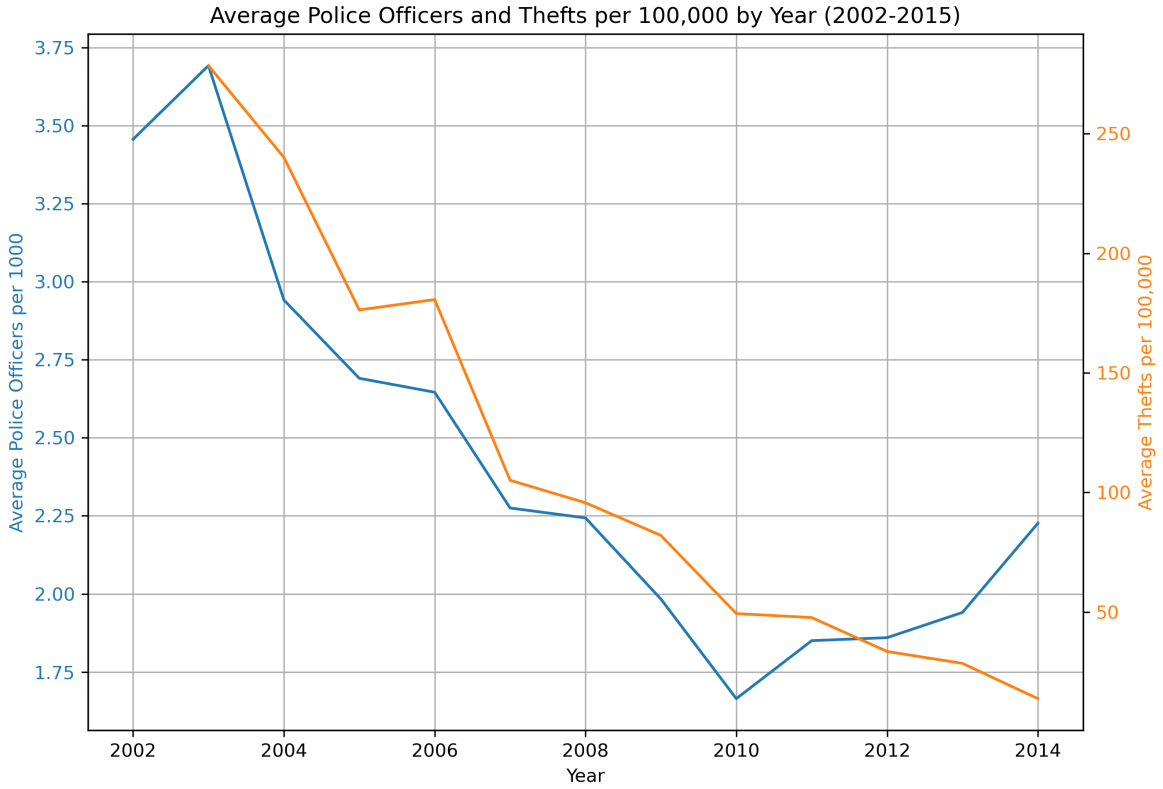


Figure 5.19:

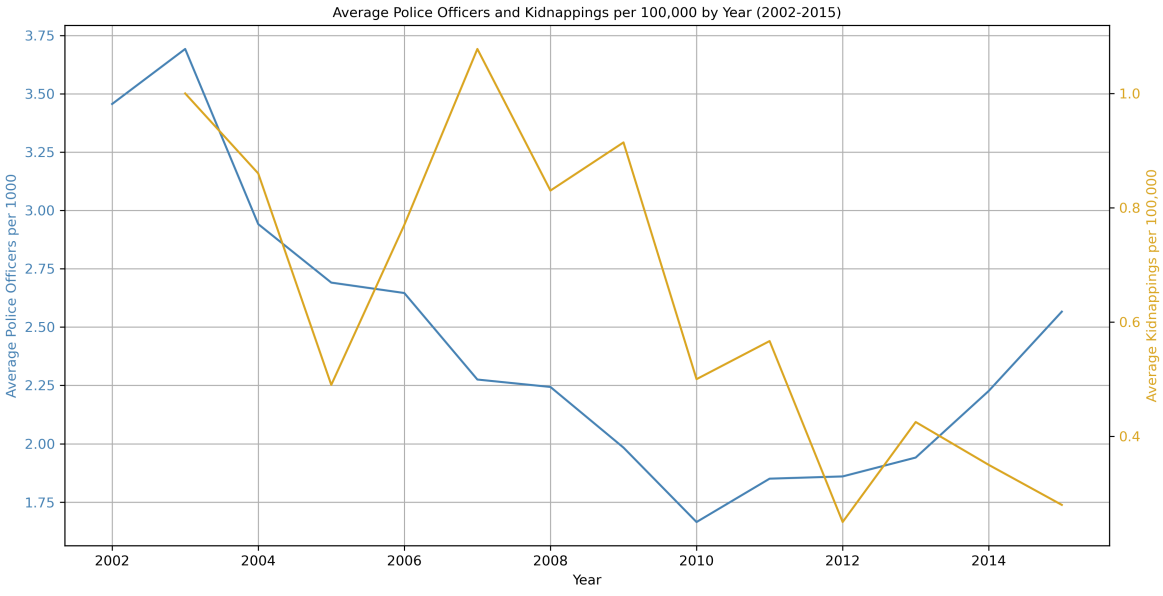


Figure 5.20:

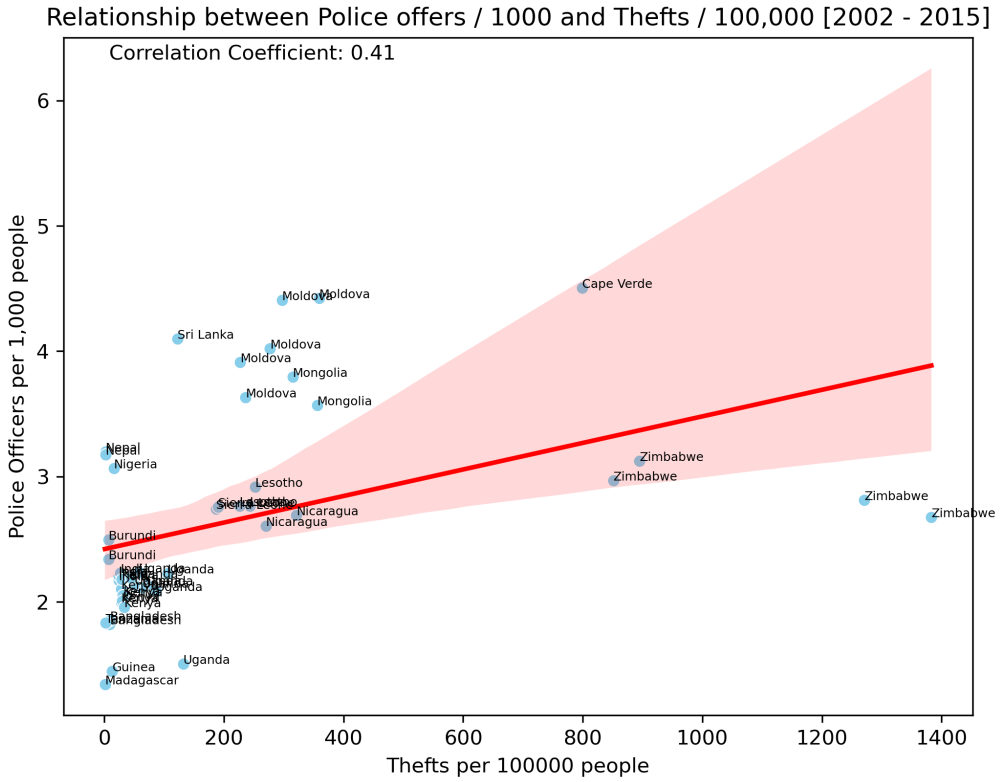

Relationship between Police offers / 1000 and Thefts / 100,000 [2002 - 2015]
Correlation Coefficient: 0.41
Police Officers per 1,000 people
Thefts per 100000 people
0
200
400
600
800
1000
1200
1400
2
3
4
5
6
Cape Verde
Moldova
Moldova
Sri Lanka
Moldova
Moldova
Mongolia
Moldova
Mongolia
Nepal
Nigeria
Zimbabwe
Zimbabwe
Lesotho
Zimbabwe
Zimbabwe
Nicaragua
Nicaragua
Burundi
Burundi
Kenya
Guinea
Uganda
Madagascar

Figure 5.21:

Figure 5.22:

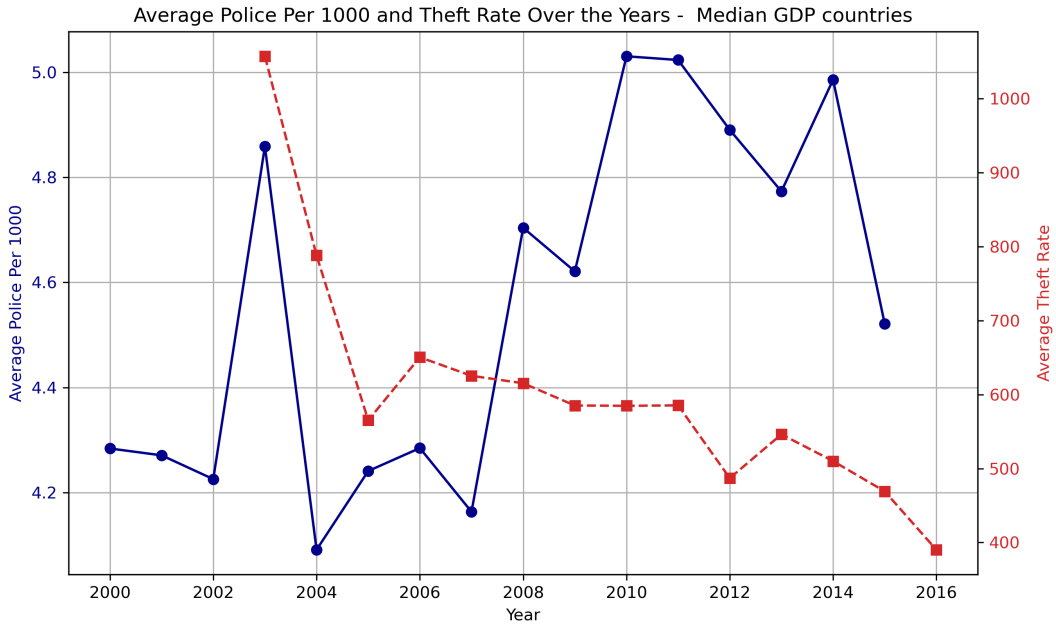


Figure 5.23:

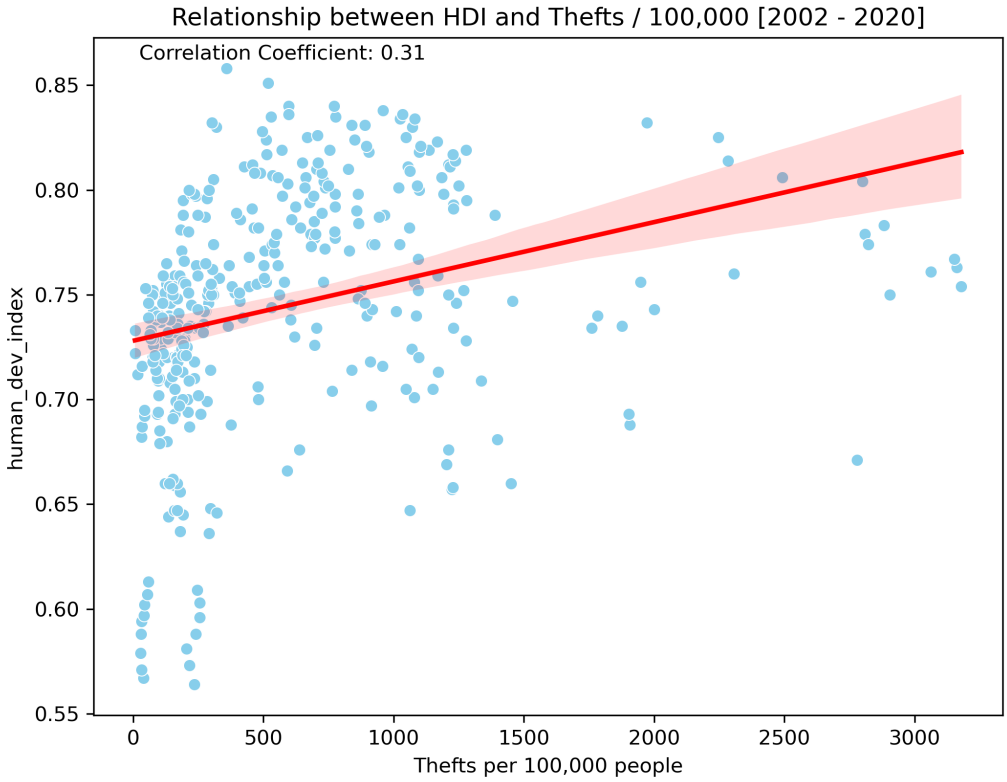
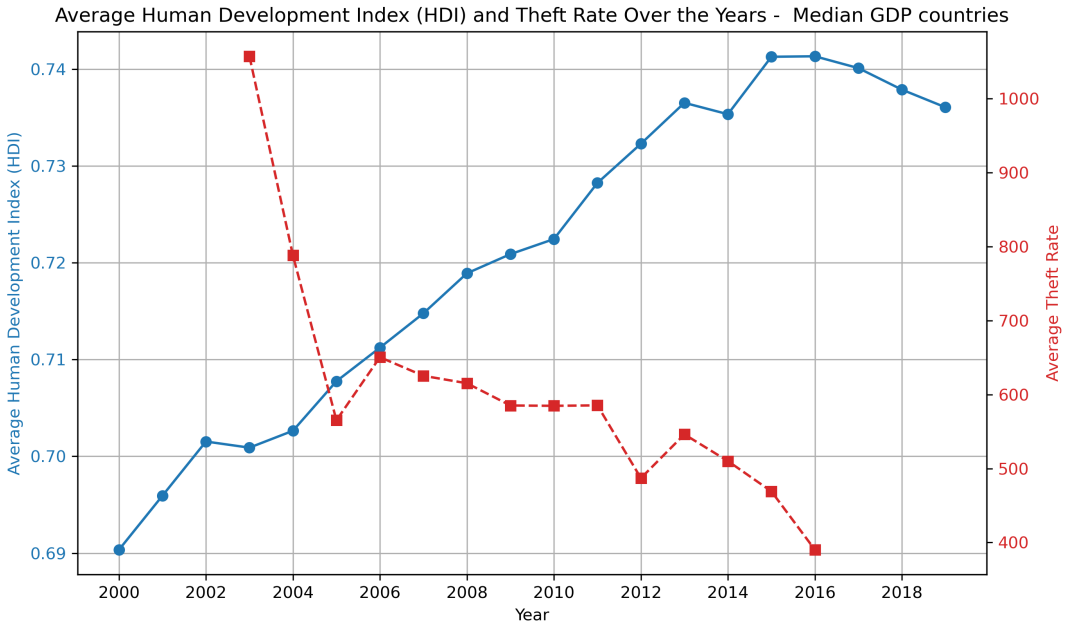

Relationship between HDI and Thefts / 100,000 [2002 - 2020]
Correlation Coefficient: 0.31
human_dev_index
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
500
1000
1500
2000
2500
3000
Thefts per 100,000 people

Figure 5.24:


Average Human Development Index (HDI) and Theft Rate Over the Years - Median GDP countries
Average Human Development Index (HDI)
0.74
0.73
0.72
0.71
0.70
0.69
Average Theft Rate
1000
900
800
700
600
500
400
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
Year

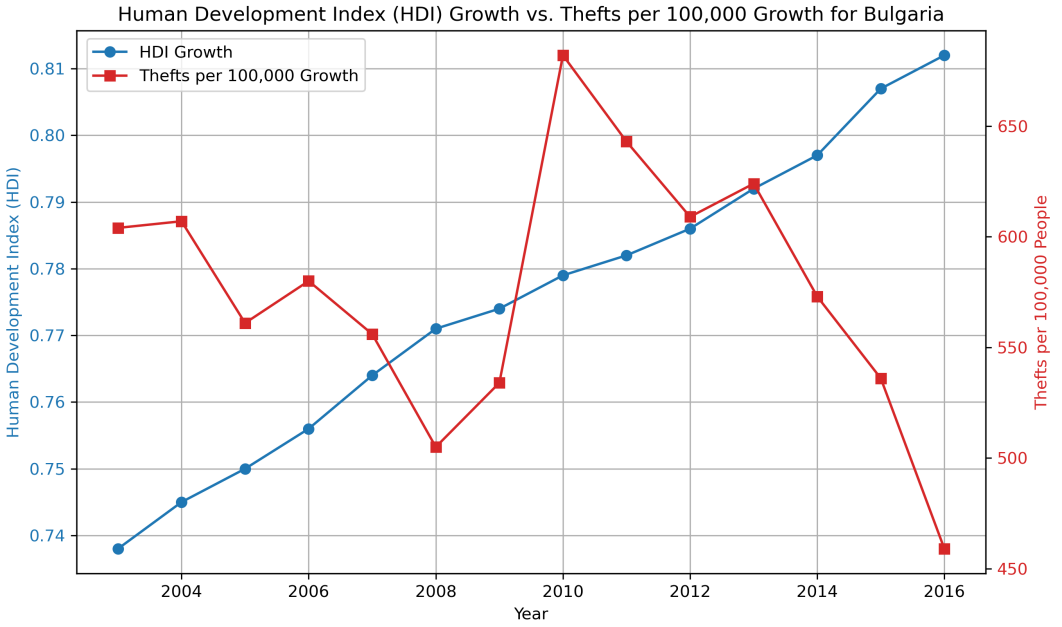


Figure 5.25:

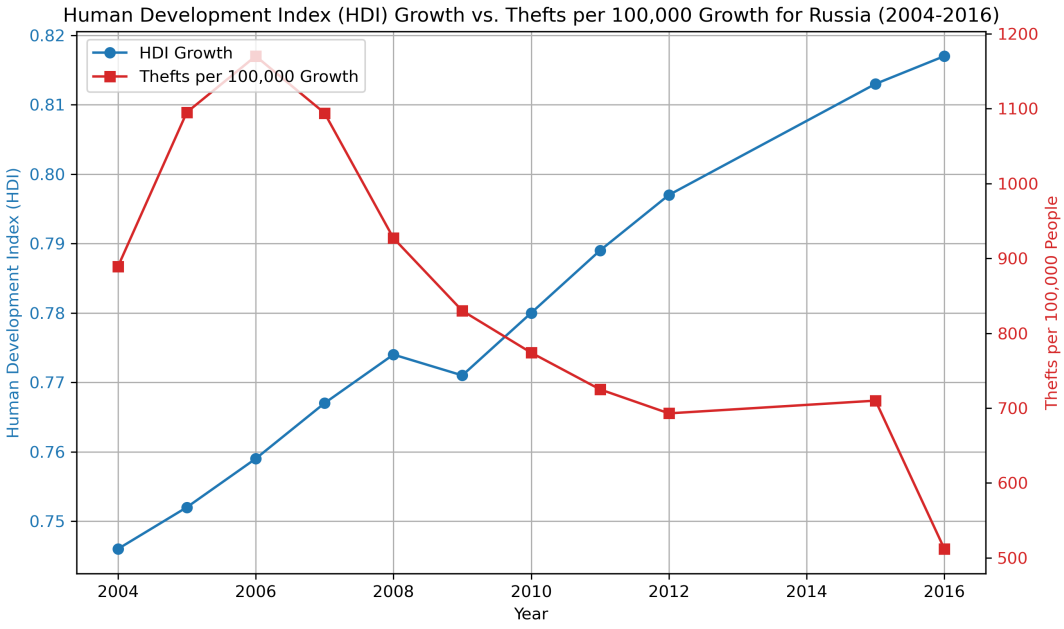


Figure 5.26:

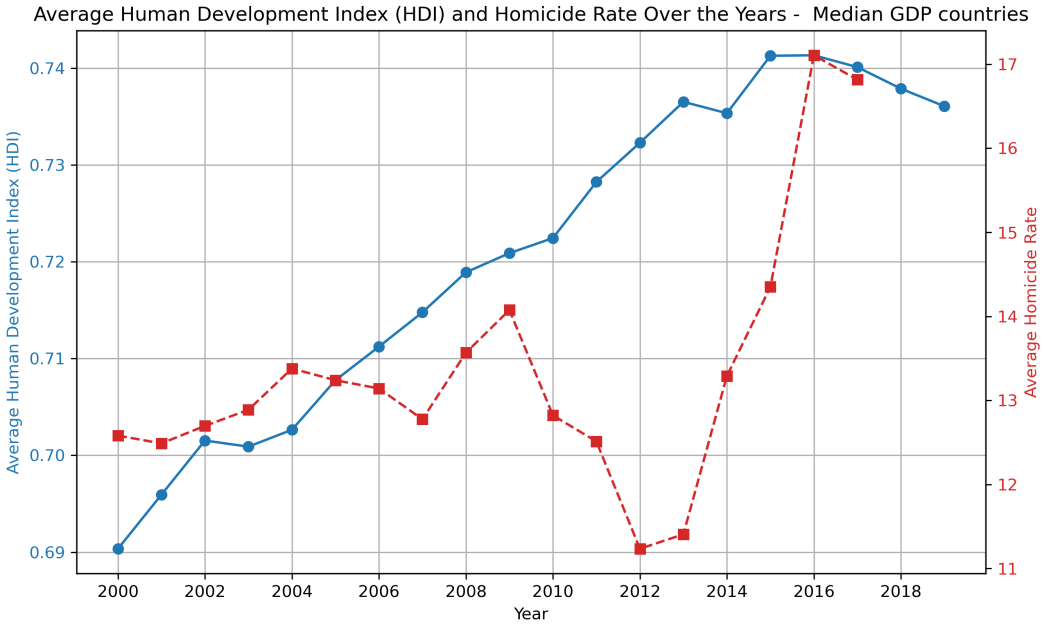

Average Human Development Index (HDI) and Homicide Rate Over the Years - Median GDP countries
Average Human Development Index (HDI)
Average Homicide Rate
Year
0.74
0.73
0.72
0.71
0.70
0.69
17
16
15
14
13
12
11
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018

Figure 5.27:

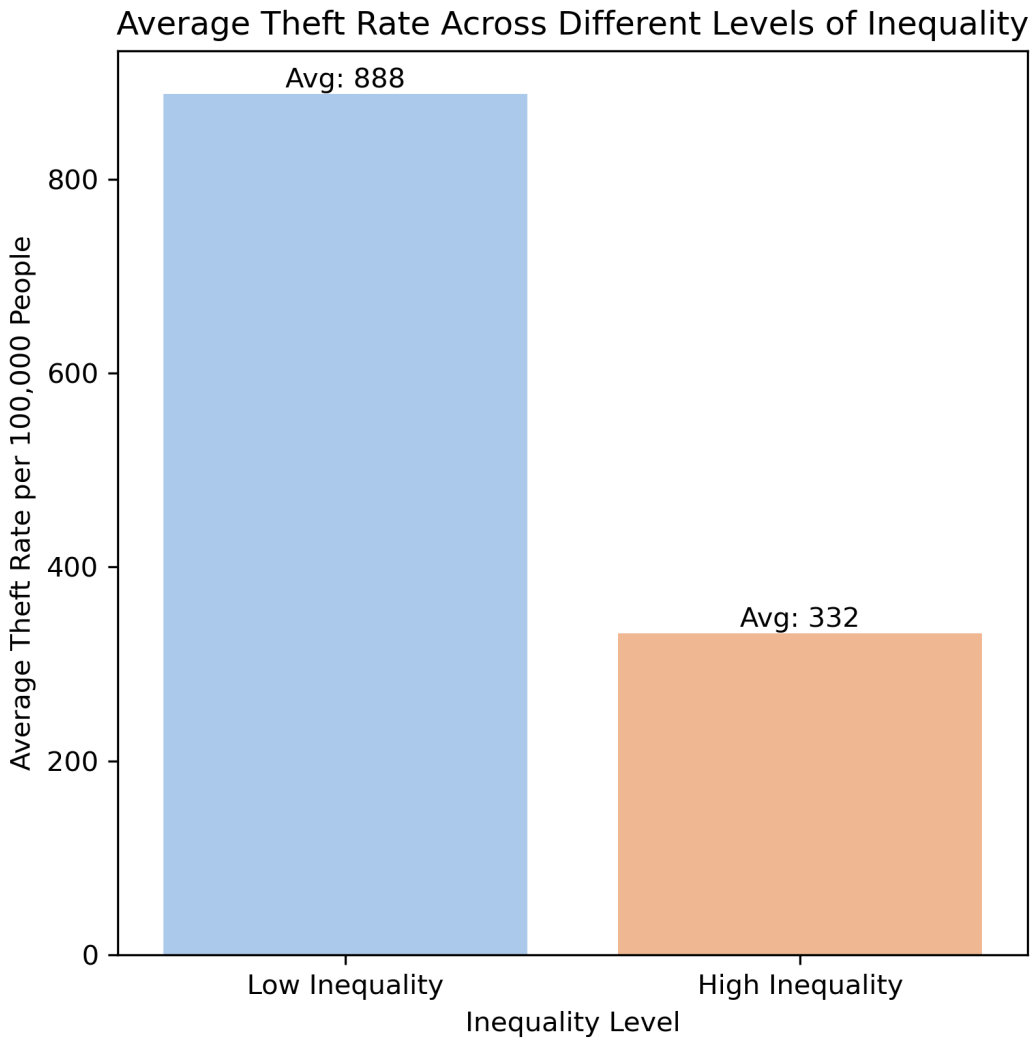

Average Theft Rate Across Different Levels of Inequality
Avg: 888
Avg: 332
800
600
400
200
0
Average Theft Rate per 100,000 People
Low Inequality
High Inequality
Inequality Level

Figure 5.28:

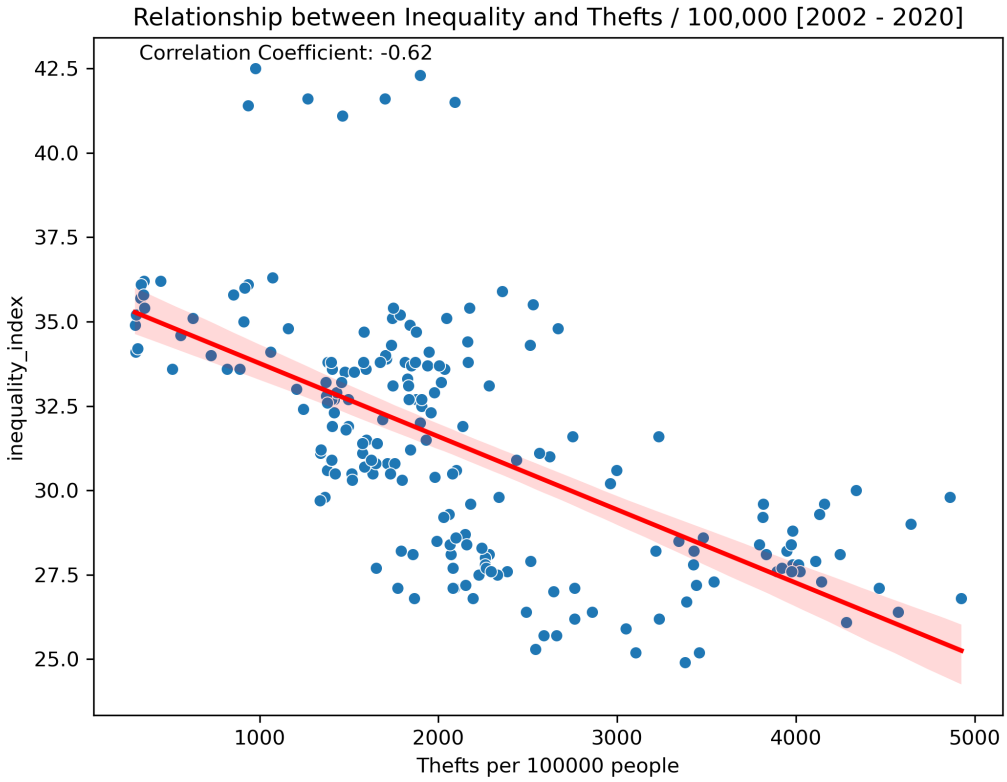


Figure 5.29:

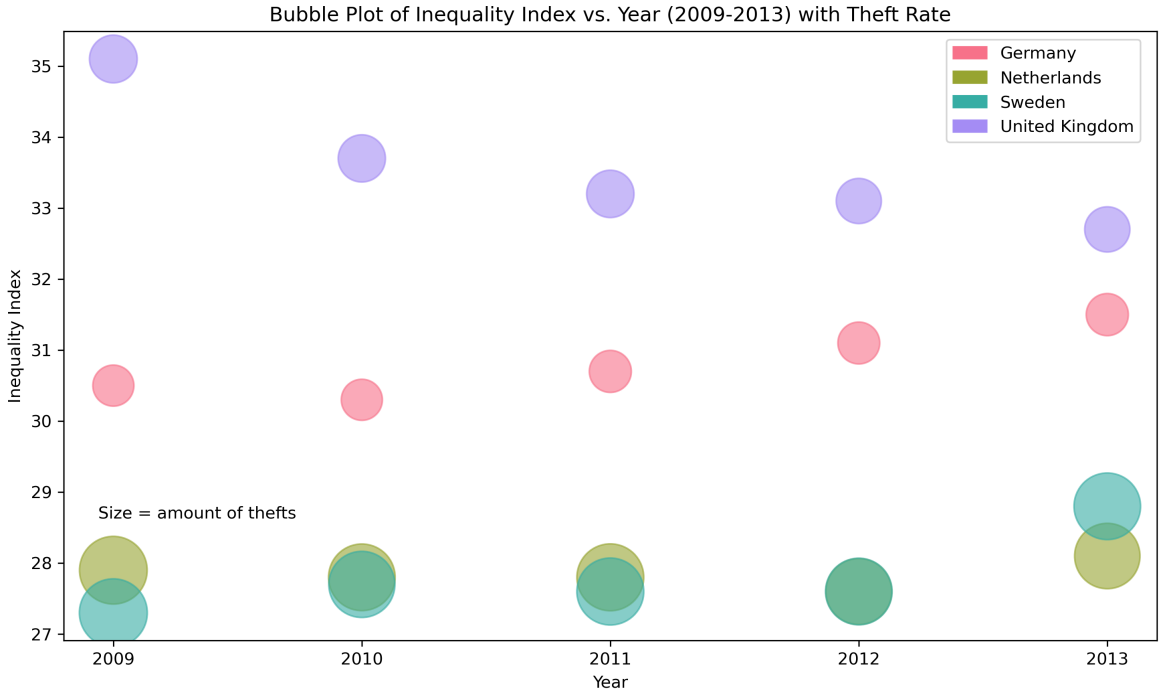


Figure 5.30:

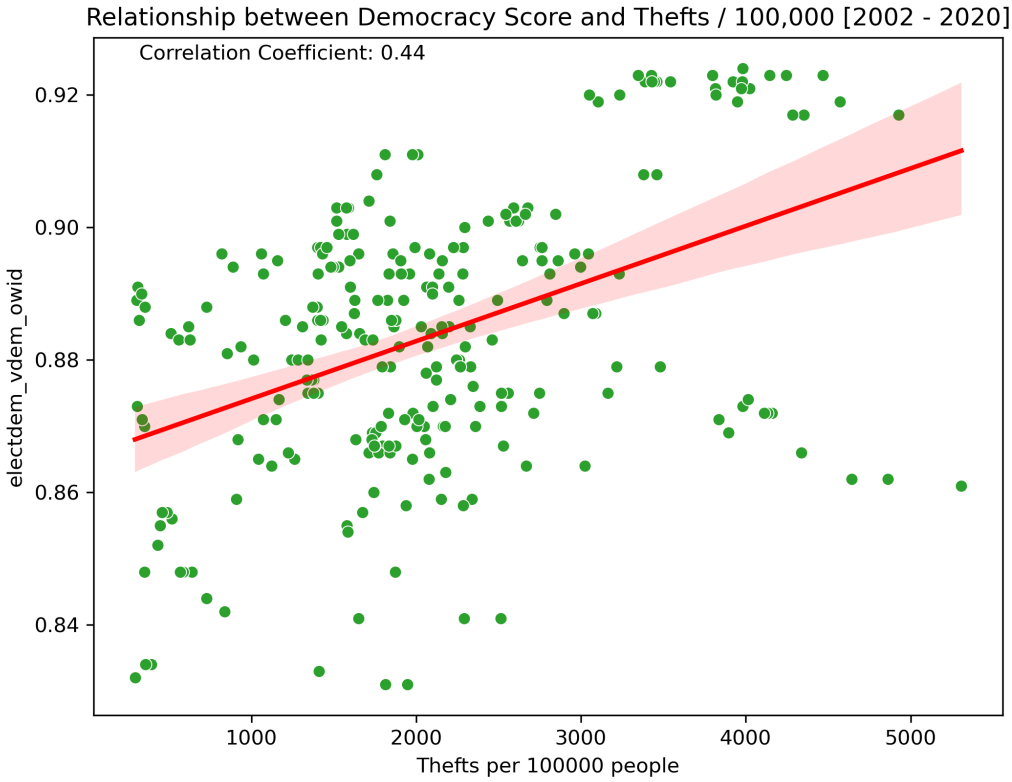


Figure 5.31:

5.5.3 Discussion

So my first sub-question was What are strong factors that affected certain crimes more than GDP, for each subset of GDP group? [Note: To see countries included in each dataset group see 6]

I took note of the top 2 factors that had a strongest correlation to the crime index for each GDP per capita subset group these were the results:

-
- All Countries: Level of Corruption and Human Development Index
- Low GDP per Capita Countries: Level of Corruption and Police per 1000 people
- Median GDP per Capita Countries: Inequality Index and Human Development Index
- High GDP per Capita Countries: Inequality Index and Inflation Rate

GDP per capita group.

What crimes will I analyse? - Thefts, Homicides, Kidnappings and Money laundering.

To do this question I will analyse each GDP per capita group, find correlated factors and do an in-depth analysis for each factor before coming to a fair analysis. I began this by starting with the Low GDP per capita dataset group

Unfortunately there were a plethora of 'NaN' values for the Crime Index in low_gdp_df, which can affect the accuracy of our results. So I did not do an analysis with the corruption score. So I began with the police per 1000 people. For the Low GDP per capita group, police per 1000 people was correlated with kidnappings and thefts. So let's analyse those graphs.

The line chart in 5.19 shows the average police officers per 1000 and average theft rate through a series of years. Looking at the line chart in 5.19 shows that as the average police per 1000 people decrease the theft rate also decrease, but then seems to begin to form an inverse relationship from 2011 onwards .This overall correlation might reflect factors such as increased police efficiency, effective crime prevention strategies, or changes in criminal behaviour due to perceived enforcement levels. But from 2011 onwards, the relationship between police per 1000 people and theft rate appears to change, forming an inverse relationship. This shift in correlation could indicate changes in policing strategies, economic conditions, or societal factors influencing crime dynamics during that period.

The line chart in 5.20 shows the average police officers per 1000 and average kidnapping rate through a series of years. The line chart in 5.20 illustrates that there is no consistent relationship between the number of police per 1000 people and the average kidnapping rate in low GDP per capita countries. This suggests that other factors may be influencing kidnapping rates beyond just law enforcement presence. There is a notable spike in the kidnapping rate observed from 2005 to 2007. This sudden increase could indicate a specific event or series of events that led to a surge in kidnapping incidents during this period. Further investigation into the socio-political and economic factors during these years may provide additional context.

Following the peak in 2007, the kidnapping rate began to steadily decrease from 2007 onwards. Further analysis and contextualisation are necessary to fully understand the underlying factors driving these trends.

After completing this analysis I plotted this on a scatter plot to observe how much data this was based on for Low GDP per capita countries. See 5.21 which shows a scatter plot of thefts vs Police per 1000 people.

This was the case for kidnappings as well there seem to be a plethora of missing data for all countries during 2002 - 2015, so I used a line of code to view how much missing data there was:

Number of missing values in 'Thefts per 100000 people' column: 816 out of 903 rows. Number of missing values in 'Police Officers per 1,000 people' column: 809 out of 903 rows

Over 90 percent of the rows from total rows were missing data, so was there any way for me to do a fair analysis? If it were a smaller percentage of missing data I could have ignored it and still conduct a good analysis. So my previous analysis is not a fair one.

Now from the heat map it may seem that there is a very strong correlation with the inequality index for kidnappings and homicides, but rest assured this is because there is only 3 rows of data available for it that is why it is so skewed. When plotted on a scatter plot there were only 3 points. Another economic factor with a lack of data!

Next I attempted an analysis on Inflation Rate but it also seems to have skewed data on a scatter plot, so I filtered out inflation values over 100 and got a new scatter plot in 5.22, which also shows the low correlation coefficient so I did not bother to advance with an analysis. Life Expectancy also had a low correlation with each crime so I moved onto median GDP per capita countries next.

For median GDP per capita countries the factors I will analyse are: HDI, inequality and Police per 1000 people.

Let's start with Police per 1000 people against the theft rate for Median GDP per capita countries

as it had a good moderate correlation according to the heatmap.

It is good to remind you that for each of these factors I will only analyse the crimes they have a moderate correlation with.

Let's begin with the Police per 1000 people versus the theft rate. The line chart in 5.23, shows the average police rate vs the average theft rate over the years. It shows that increasing the number of police officers will decrease the number of thefts for countries in the median GDP per capita group, this is evident from the spike from 2002-2004 and years after 2007. The thefts gradually decrease from the high rate it is at , while the number of police officers has increased and remains somewhat consistent from 2009 to 2014. The sustained level of police presence from 2009 to 2014 suggests that law enforcement agencies may have given resources effectively to address public safety concerns regarding theft

Next let's analyse the Human Development Index (HDI). From observing a heat map we found out that HDI affects homicides and thefts the most out of all crimes for Median GDP per capita countries. So let's begin by analysing the HDI against these 2 crimes for median GDP per capita countries. Looking at a scatter plot in 5.24 which shows the HDI vs Thefts per 100,000 people.It shows a moderate correlation with a correlation coefficient of 0.31. This shows that countries with higher HDI's will have a higher theft rate than countries with lower HDI's. After finding out this information we decided to use a line chart to measure the average HDI of median GDP per capita countries to average thefts per 100,000 over the years.

Let's see if this visualisation will prove that "if the HDI increases, more times than not the Theft rate will also increase" or if this statement is false or the total opposite for median GDP per capita countries.

The line chart in 5.5.2 shows as the HDI increases the theft rate keeps constantly decreasing, this is very evident starting from the years of 2005 onwards. Can we say this is a fair assessment for all countries? To answer this I constructed a code snippet to choose some random countries within the median GDP per capita dataset to measure the HDI against the Theft Rate throughout the years. Two countries I got from the random selection were Russia in 5.26 Bulgaria in 5.25.

Looking at the graph for these 2 countries within the median GDP per capita group it is safe to say in most cases for countries in this group, as the HDI increases, more often than note the theft rate will be decreasing. In Russia5.26 the HDI and theft rate have a strong inverse relationship between each other. This is very visible for Russia starting from the year 2006. This is also the case for Bulgaria5.25 apart from the years 2008-2010, this inverse relationship is very evident.

Let's analyse the HDI against the homicide rate now. Looking at the line chart in 5.27, it shows a line chart of the average homicide rate vs the average HDI for years 2000 to 2019.

Looking at the line charts for both thefts 5.5.2 and homicides 5.27 against the HDI, unlike the inverse relationship typically seen between thefts and HDI, where higher HDI scores tend to correspond with lower theft rates, no such trend is evident in the case of homicides. The homicide rate remains relatively stable and inconsistent across different levels of HDI, showing no clear trend of decreasing or increasing as HDI scores change.

Let's finally move onto analysing the Gini coefficient/Inequality index for median GDP per capita countries. Looking at a heat map, inequality has a moderate relationship with thefts. Upon viewing a scatter plot of the inequality index against thefts it shows as the inequality of median GDP per capita countries decreases, the theft rates increase, which is very surprising. It's important for us to remember that correlation does not imply causation. There could be several reasons for this moderate relationship. The observed correlation between inequality and theft rates could be coincidental and not driven by any causal relationship. It's possible that other factors or variables not considered in the analysis are responsible for both inequality and theft rates. To gain a deeper

understanding of the relationship between inequality and theft rates, we would need to conduct further research and consider additional variables.

After finding this information we split the data into low inequality and high inequality datasets from the median GDP per capita subset. I then placed the mean theft rate for low inequality countries and high inequality countries on a bar chart that can be seen in 5.28.

From the bar chart this is what I derived: With an average theft rate of 888 incidents per 100,000 people in countries with low inequality, and only 332 incidents per 100,000 people in countries with high inequality, it's evident that inequality plays a significant role in shaping societal dynamics and crime prevalence especially for thefts in this example. The substantial disparity in theft rates suggests that inequality not only impacts economic and social structures but also influences crime patterns within a society. Higher levels of inequality may lead to increased socio-economic tension, but in this case for median GDP per capita countries it does not seem to be the case. Inequality is a complex concept and is difficult to measure so it is difficult to come to a conclusive answer, we will measure this again with Higher GDP per capita countries.

Let's finally analyse High GDP per capita countries.

The factors I will analyse against crimes are inequality index, democracy score and police per 1000 people as they have the most correlation to the crimes committed.

Let's begin with inequality again for High GDP per capita countries. The inequality index has a positive correlation with the homicide rates and a negative correlation mostly with thefts rates, similar to median GDP per capita countries. Let's start with the Theft rates.

Viewing the scatter plot in 5.29, we seem to have a similar case to median GDP per capita countries. Just looking at the scatter plot we can see that the inequality levels for high GDP countries are lower (25-42) while for median GDP per capita countries it was in the range of 25-55, the overall theft rate has also increased. We have the same case as before from the scatter plot that a lower inequality index correlates with an increase in thefts.

To see if this hypothesis was true I took the top 4 countries (in the High GDP per capita dataset) with the highest theft counts, these were Netherlands, United Kingdom, Germany and Sweden. I noted this and constructed a bubble chart where the size of the bubble corresponds to the amount of thefts a country has. I put them up for a series of years. This can be seen in 5.30. Looking at the bubble chart above we can observe that countries with a lower inequality index experience a higher rate of thefts than countries with a higher one, which affirms our hypothesis from the scatter plot.

Unfortunately I could not do an analysis with inequality and homicides because the correlation coefficient was skewed by a few outliers in our dataset. After filtering them out the correlation coefficient was 0.07, so I did not move forward with the analysis as it was too small.

Next we moved onto the democracy score. Let's start by looking at the democracy score vs the theft rate as there is a positive moderate correlation between the 2 factors. I made another scatter plot to observe the relationship between thefts and democracy scores. Looking at the graph initially it seemed to be grouped together so I filtered out democracy scores less 0.83 to get a more distributed scatter plot which can be seen in 5.31

After filtering out the outliers in my dataset to have a fair sample dataset to analyse I did not do a pairwise comparison again as this was getting tedious and repetitive. The scatter plot did show us that as the democracy score increases the theft rate is likely to increase. While the scatter plot suggests a correlation between democracy scores and theft rates, it's essential to acknowledge that correlation does not imply causation.

I then began to examine each GDP per capita group heat maps, to try and see what factors affect each other and the crime so we can come up with some equations to decrease certain crimes based on my previously and the heat maps. This what I came up with:

How can we decrease crime for High GDP per capita countries?

We can see that HDI leads to less theft. If the HDI increases it is likely that Thefts will start to decrease, this hypothesis was supported by the analysis on median GDP per capita countries. So what factors help increase the HDI? Life Expectancy!

If a country can identify specific public health interventions, social welfare programs, and healthcare policies that can help improve life expectancy and, by extension, there is a fair probability that it can contribute to lower theft rate.

When observing a heat map it seemed that inequality has a strong correlation to homicide but after we performed analysis we refuted this observation.

How can we decrease crime for Median GDP per capita countries?

To decrease the number of thefts for median GDP per capita countries, our data suggest to increase inequality levels, even if it sounds unjust, it seems to be a viable solution. Our analysis also highlighted the importance of law enforcement in combating crime. Increasing the number of police officers per 1000 people could serve as a deterrent to criminal activity, especially thefts. Increasing the HDI also helps with decreasing thefts as I previously explained for High GDP per capita countries.

How can we decrease crime for Low GDP per capita countries?

Unfortunately due to the severe lack of data. I could not come up with any fair solutions is try and decrease crime for Low GDP per capita countries.

Based on the analysis above, it can be concluded that GDP per capita alone is not the sole factor influencing crime rates in each country. Instead, various macroeconomic factors such as the Inequality and human development index play significant roles in shaping crime dynamics.

While GDP per capita may have some influence on crime rates, it is not the only determining factor. The analysis reveals that other factors, such as corruption, inequality, and human development index (HDI), have stronger correlations with certain types of crimes across different GDP per capita groups.

For low GDP per capita countries, factors like the total police presence (police per 1000 people) show correlations with crime rates, suggesting that law enforcement efforts may play a crucial role in crime prevention, but like I said previously this is not a fair conclusion due to the lack of data for low GDP per capita countries

In median GDP per capita countries, HDI and inequality emerge as significant factors affecting crime rates, particularly thefts

High GDP per capita countries show correlations between inequality and crime rates, with lower inequality associated with higher theft rates.

In conclusion, GDP per capita is an important but not exclusive determinant of crime rates. This analysis provides insights into potential crime prevention strategies based on the identified correlations, apart from Low GDP per capita groups. An example is increasing Life Expectancy will hopefully for most high GDP per capita groups with decrease the homicide rates in those countries. Overall, this analysis highlights the complex nature of various macroeconomic factors in shaping crime rates. It emphasises the need for good and back tested approaches for crime prevention

5.6 RQ6: Do countries with high happiness index and high unemployment rate have a specific type of crimes prevalent?

For this research question:we aim to investigate the relationship between the happiness index, unemployment rates, and the prevalence of specific types of crimes in countries. By examining this relationship, we seek to understand if there are any patterns or correlations between these factors.

My first sub question is **Does the level of happiness influence the number/types of crimes?** This sub question focuses on exploring whether there is a relationship between the level of happiness in a country and the number and types of crimes occurring within that country. It aims to determine if countries with higher happiness indexes tend to have lower crime rates or different types of crimes compared to countries with lower happiness indexes. We will be splitting the countries into low, median and high happiness countries to answer this question generally.

The next sub question: **Do countries with the highest happiness index have the lowest number of homicides?** This sub question specifically investigates the relationship between the happiness index and the occurrence of homicides in different countries. It aims to see if countries with higher happiness indexes tend to have lower homicide rates compared to countries with lower happiness indexes or median happiness indexes.

The final sub question is **Do countries with high unemployment rates have more robberies and thefts than countries with low unemployment rates since there are more unemployed people around the streets?** This sub question focuses on the relationship between unemployment rates and the prevalence of robberies and thefts in different countries. It aims to determine if countries with high unemployment rates experience higher rates of robberies and thefts compared to countries with low unemployment rates. The hypothesis behind this sub question is that the presence of more unemployed individuals on the streets may contribute to higher incidents of thefts and robberies.

By addressing these sub-questions, we hope to broader gain insights into the relationship between happiness, unemployment rates, and crime patterns, which can contribute to a better understanding of the social dynamics and factors affecting crime.

5.6.1 Data & Method

Explain in detail the data you used to answer this question. How did you use this data to answer the question?

These are all the data columns used, most of these have already been previously explained in this report.

- Country
- Year
- Level_of_Happines:This number represents the happiness index or a similar metric that measures the subjective well-being and happiness of individuals within a country
- unemployment_rate: This column indicates the percentage of unemployed individuals in the country's labor force.

- Money_laundering_per_100k: This data column provides insights into financial crimes related to illicitly handling and disguising the origins of illegally obtained money
- Smuggling_of_migrants_per_100k: This column indicates the rate of reported incidents of smuggling of migrants per 100,000 people in the country. It provides insights into illegal transportation and trafficking of individuals across borders.
- Robberies per 100000 people
- Thefts per 100000 people
- Homicides per 100000 people
- Kidnappings per 100000 people
- Total Sexual Violence at the national level, rate per 100,000: This column represents the rate of reported incidents of sexual violence at the national level per 100,000 people in a country.

We used this data to analyse and explore the relationships between the variables mentioned above. By examining the data across different countries and years, we calculated various statistical measures, such as correlations, averages, and trends, to understand the potential associations between the happiness index, unemployment rates, and the prevalence of specific types of crimes to answer our research question and sub questions.

We visualised the data using charts, graphs, and maps to provide a more intuitive representation of the relationships and patterns observed. This allowed us to present the findings and conclusions in a comprehensive and accessible manner.

5.6.2 Results

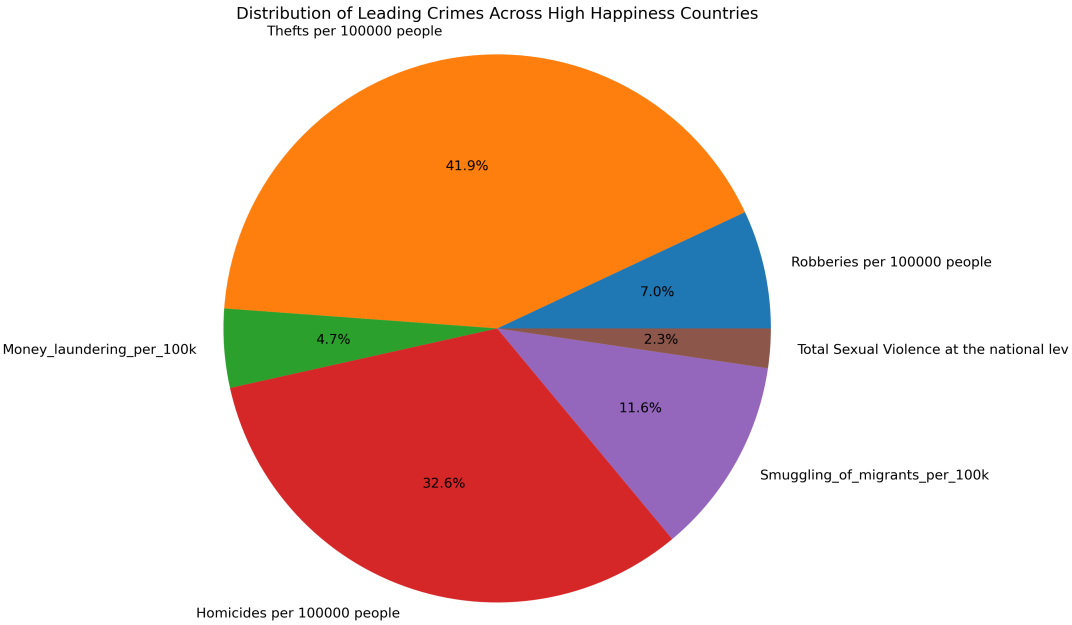


Figure 5.32:

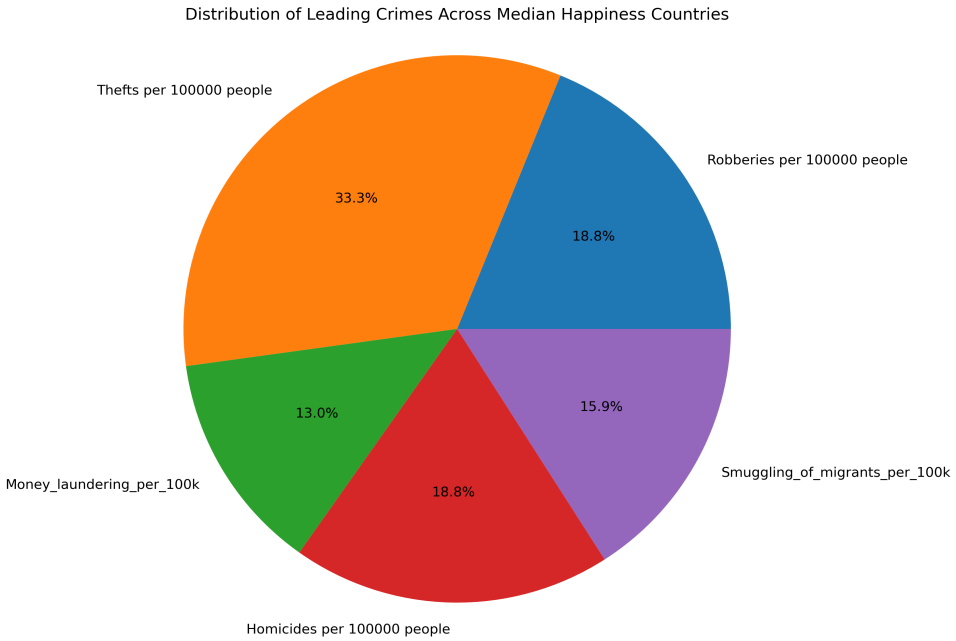


Figure 5.33:

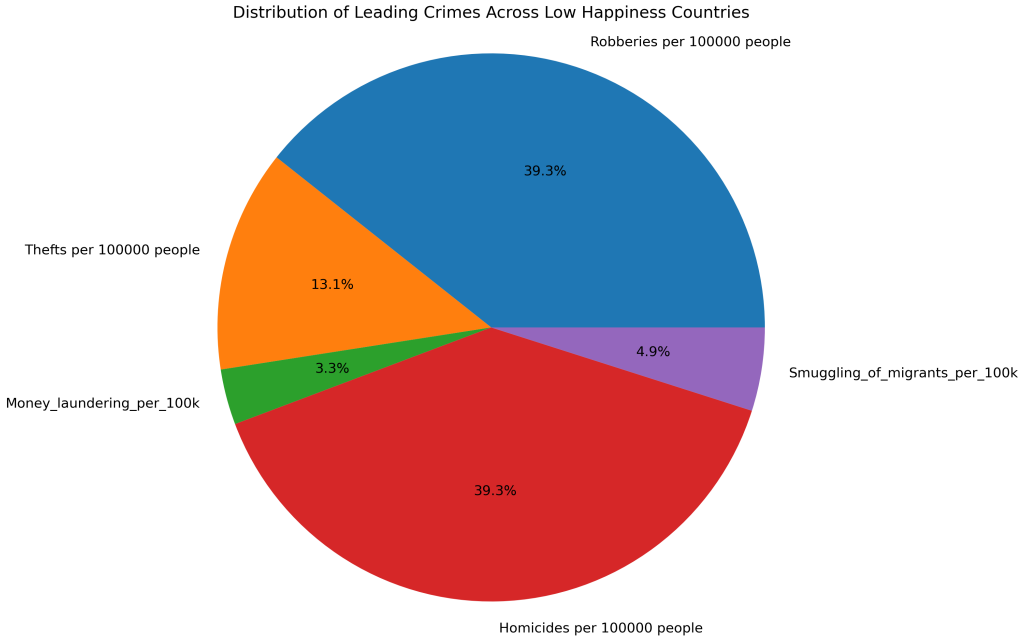


Figure 5.34:

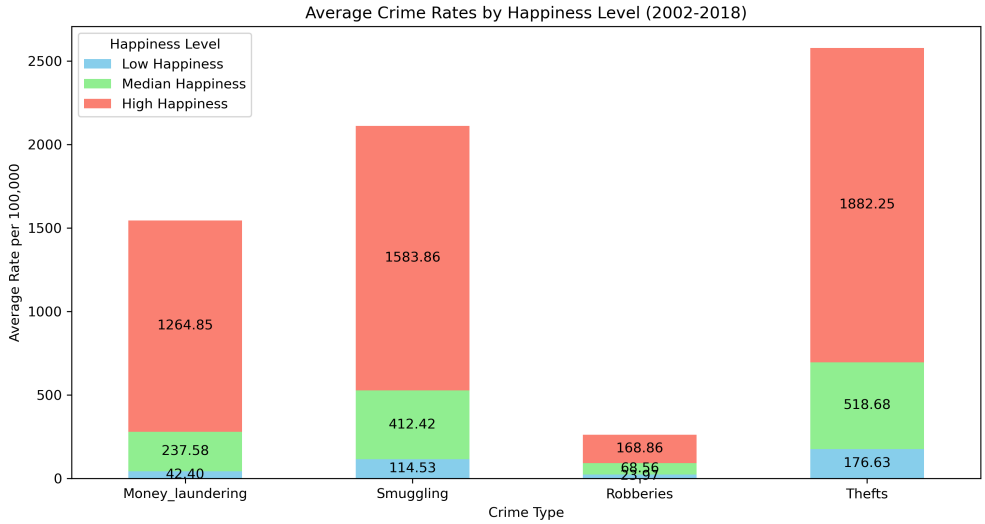


Figure 5.35:

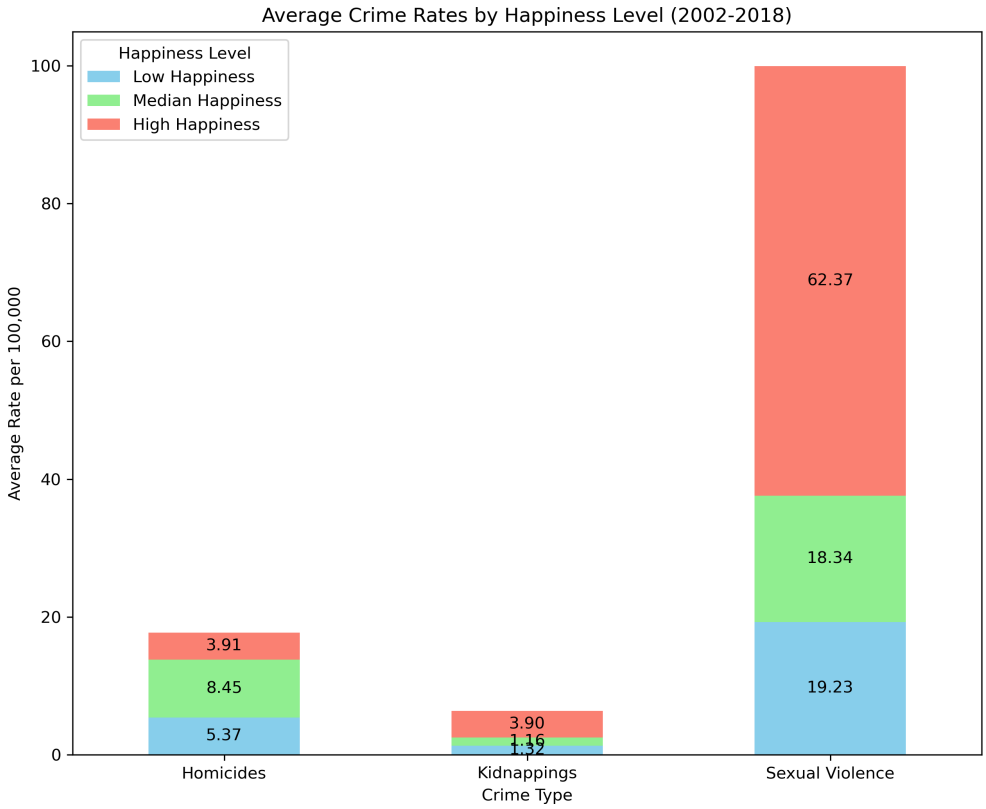

Average Crime Rates by Happiness Level (2002-2018)
Happiness Level
Low Happiness
Median Happiness
High Happiness
Average Rate per 100,000
100
80
60
40
20
0
3.91
8.45
5.37
3.90
1.16
1.52
62.37
18.34
19.23
Homicides
Kidnappings
Sexual Violence
Crime Type

Figure 5.36:

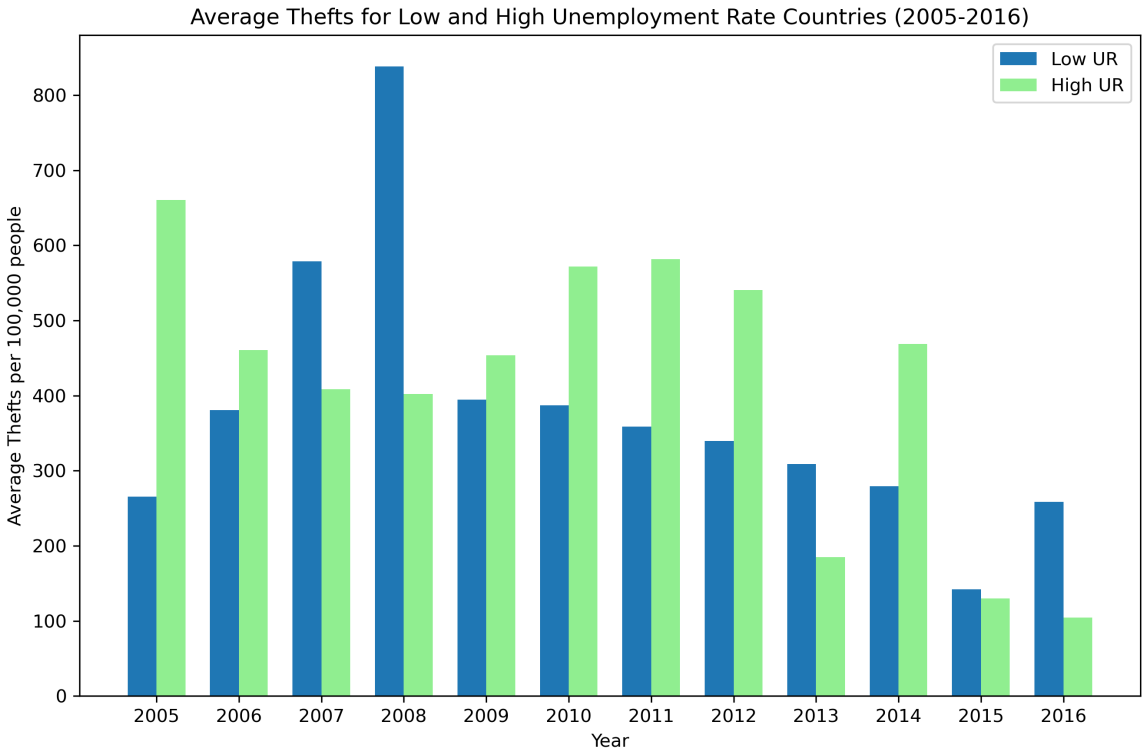


Figure 5.37:

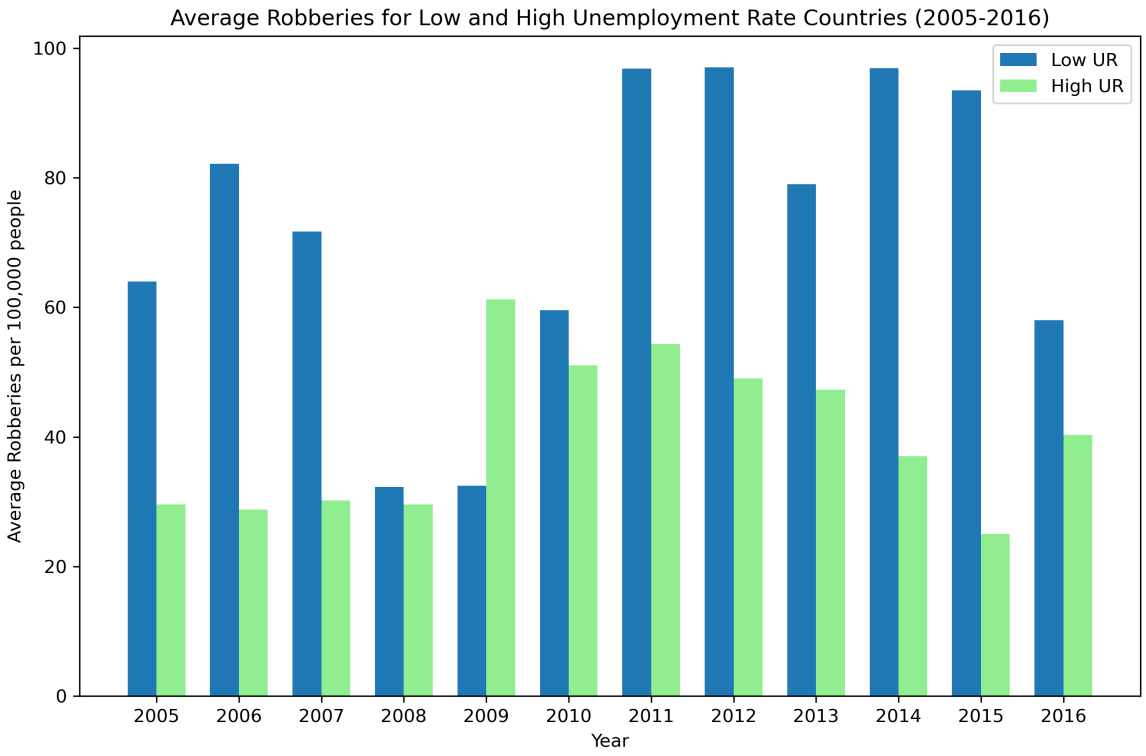


Figure 5.38:

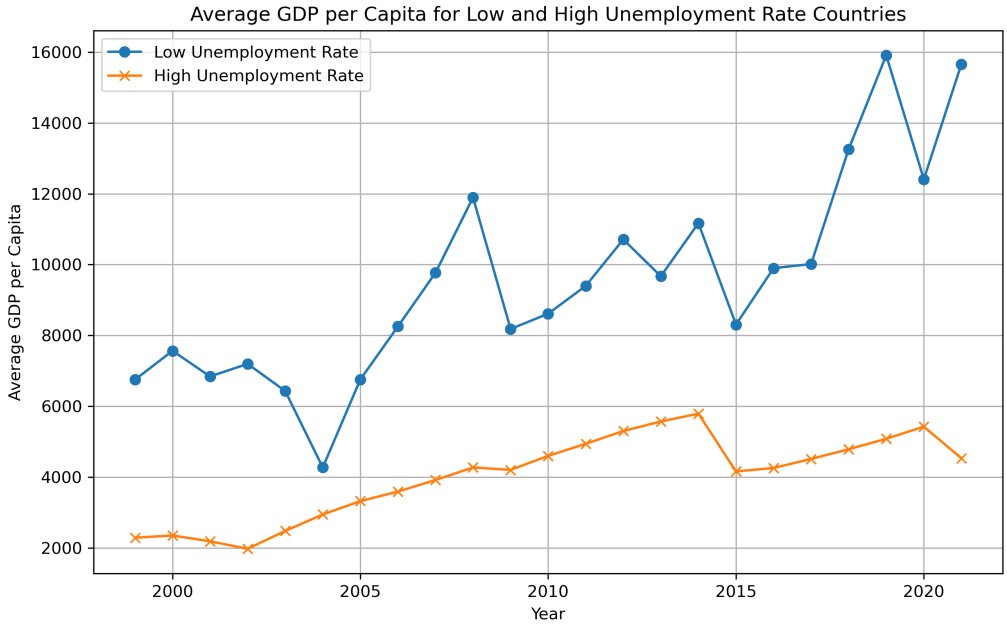


Figure 5.39:

5.6.3 Discussion

[Note: To see countries included in each dataset group see 6]

So our first sub question was **Does the level of happiness influence the number/types of crimes?**

To begin I split the data into low, median and high happiness index countries using the quartile method.

Then I wanted to look at the leading crime for country, then place it on a pie chart for each subset group of happiness. This can be seen for in the pie charts at 5.32(High happiness), 5.33(Median happiness) 5.34(Low happiness). The percentage on these pie charts represent the number of countries that have that crime as their leading crime. Let's look at High happiness with 32.6 percent homicides on the pie chart; this means that 32.6 percent of countries of countries have homicides as their leading crime out of all the other crimes on the pie chart. All the crimes are normalised to counts per 100,000.

So what can we understand from these pie charts and how do they help answer our research question?:

Low Happiness Countries: Robberies and Homicides are the most common crimes, with 39.3 percent countries having them both as their leading crime each. Thefts and Smuggling of migrants are also prevalent.

Median Happiness Countries: Thefts are the most common crime, with 33.3 percent countries having them as the leading crime.

High Happiness Countries: Thefts are again the most common crime, with 42 percent of countries having them as the leading crime. Homicides and Smuggling of migrants follow, with 32.6 percent

and 12 percent of countries, respectively.

For context purposes This is the number of countries within each happiness group in Appendix C there is a list of all the countries included in each subset dataset. 6

- High happiness countries: 43
- Median happiness countries: 69
- Low happiness countries: 61

Robberies and Homicides are consistently common across all happiness levels, indicating their pervasive nature regardless of the happiness level of the country. Thefts are also prevalent across all happiness levels but seem to increase as happiness levels also increase. Money laundering and smuggling of migrants are more prevalent in median happiness countries compared to high and low happiness countries. Kidnappings and total sexual violence seem to be relatively less common across all happiness levels. Another key observation to point out is that as the Happiness increases from low, median and then high. The theft rate is also increasing meaning more countries have theft's as their leading crime. On the other hand is seems that the robberies are decreasing as the happiness index for each group increases.

One thing to note is that happiness index/happiness is often associated with better socioeconomic conditions, such as higher income, better access to resources, and improved standards of living. As countries experience higher levels of happiness, it is plausible that individuals have more disposable income and access to valuable possessions, which can contribute to an increase in theft as a leading crime. This observation suggests that the prevalence of theft may be influenced by the availability of targets and opportunities for theft. The observation highlights the distinction between theft and robbery as distinct types of crimes. Theft typically involves the unlawful taking of property without direct confrontation or violence, while robbery involves the use of force or threat to steal from an individual directly.

Based on the above information how can we answer this sub question?:Does the level of happiness influence the number/types of crimes?:

The level of happiness appears to have an influence on the number and types of crimes prevalent in different countries. Across all happiness levels (low, median, and high), robberies and homicides consistently remain common crimes. This suggests that these types of crimes may be less influenced by the happiness index and more influenced by other socio-economic or contextual factors.

However, when it comes to theft, there is a notable trend as happiness levels increase. Thefts are the most common crime in both median and high happiness countries, with a higher percentage of countries having theft as their leading crime in high happiness countries (42 percent) compared to median happiness countries (33.3 percent). This observation suggests that as happiness levels increase, there may be an increased prevalence of theft as a leading crime. As we said before this could be because of a number of factors such as higher disposable income, better access to valuable possessions, and potential opportunities for theft in more prosperous and content societies.

To come to a final answer for this sub-question, I then created a stacked bar chart to further analyse the crimes and see what percentage each happiness subset group of the total number of crimes.

Look at the stacked bar charts on 5.35 and 5.36, which shows the average number of crimes for each group and crime from years 2002-2018.

Based on the information on the graph how can we answer: Does the level of happiness influence the number/types of crimes?

High happiness countries tend to have higher crime rates across various categories (except for homicides)compared to median and low happiness countries, so it is unreasonable to say for certain if level of happiness influences the number of crimes but we can definitely see that the level of happiness may have some influence on the number/types of crimes examples being homicide; countries with a high level of happiness with have some of the lowest homicides rates in the world. However, it is important to note that this analysis only considers averages and might not capture all the complexities and factors influencing crime rates in different countries.

We can now answer our next sub-question as well: **Do countries with the highest happiness index have the lowest number of homicides?**

This visualisation allows us to answer this question easily, we can confirm that countries within the countries with a high happiness index will indeed have lower homicide rates than median and lower happiness countries

Now for the last sub-question

Do countries with high unemployment rate have more robberies, thefts or money laundering than countries with low unemployment rates since there are more unemployed people trying to escape their situation by committing these white collar crimes.

To begin this question I separated countries into low and high unemployment rate datasets using the quartile method. My following analysis does not consider money laundering as there is not enough data.

My main hypothesis for this question is that robberies and thefts would be more common for countries with high unemployment rate countries than low unemployment rate countries. So to begin this I constructed a grouper bar chart of high unemployment rate and low unemployment rate countries over a series of years. 5.37 shows a grouped bar chart that shows the average theft count for both subsets for each year from 2005 to 2016.

Looking at the grouped bar chart in 5.37 it seems in most cases countries with High Unemployment Rates will have more cases for Theft than countries with Low Unemployment rates. 7 out of the 11 years high unemployment regions have a higher theft rate than countries with low employment rates

I then created the exact same grouped bar chart but for robberies instead and was surprised by the results. See 5.38

This chart surprisingly shows that countries with a low unemployment rate will have way more cases for Robberies than countries with a high employment rate. What I can try and understand from this graph is that low unemployment countries may have more commercial activity and wealth concentration, providing more opportunities for robberies targeting businesses. In contrast, high unemployment countries might see more opportunistic thefts targeting individuals or less secure establishments. We can try and confirm this by looking at the average GDP per capita growth for low unemployment countries and high unemployment countries, then we can come to a conclusion on this question.

The line graph in 5.39 shows the average GDP per capita for low unemployment rate countries and high unemployment rate countries from the years 1999 - 2021.

Looking at the line chart in 5.39, it supports my claim that countries with lower unemployment rates have a wealthier and more sustainable economy than countries with higher unemployment rates. The GDP per capita is a good measure of how a country's economy is doing, e.g more business and newer technology. So this may be good evidence that countries with a lower unemployment rate will have more robberies than countries with a higher one.

So how can we answer our governing question:**Do countries with high happiness index and high unemployment rate have a specific type of crimes prevalent?**

This analysis revealed that robberies and homicides are consistently common across all happiness levels, indicating their pervasive nature regardless of the happiness level of the country. Thefts are also prevalent across all happiness levels but seem to increase as happiness levels also increase.

Countries with a high happiness index tend to have lower homicide rates compared to countries with lower happiness levels. This results suggests that there may be a negative correlation between the happiness index and homicide rates, indicating that happier countries may experience lower levels of violent crimes such as homicides.

The analysis of unemployment rates and property crimes, specifically robberies and thefts, revealed interesting insights. Countries with high unemployment rates tend to have more cases of theft compared to countries with low unemployment rates. This may be because of various factors such as economic instability and lack of opportunities, leading individuals to resort to theft as a means of survival or escape from their situation. On the other hand, countries with low unemployment rates have higher rates of robberies, this is possibly due to higher commercial activity and wealth concentration, providing more opportunities for robberies targeting businesses.

Finally, the analysis suggests that while happiness index and unemployment rate may influence the prevalence of certain types of crimes, the relationship is complex and influenced by various socio-economic factors. We can conclude that countries with higher happiness index tend to experience lower homicide rates, while countries with high unemployment rates may see higher rates of theft. Though there is correlations within these relationships. These relationships are not absolute and may vary depending on the specific socio-economic context of each country

Chapter 6: **Conclusions**

So what are the conclusions we can draw from our research project 'Does the GDP per capita of a country influence the type of crime?'

Firstly we can conclude that GDP per capita can influence the type of crime in a country as an economic indicator, but it is good to note that GDP per capita is not always the best indicator for each crime committed in a country. Through comprehensive data analysis and visualisation, several key findings have emerged from some of our research questions in relation to our overall project title.

From our first research question. The findings suggest that a high GDP per capita correlates with a decrease in violent or assault crimes, as indicated by lower average homicide rates and a lower average kidnapping rate in high GDP per capita countries compared to low and median GDP per capita countries. Additionally, low GDP per capita values can be considered a contributing factor to the overall homicide rate, particularly in median GDP per capita countries.

The analysis also highlights the potential impact of other factors, such as law enforcement, social infrastructure, and access to education and healthcare, in shaping crime rates; all of these can contribute to the overall GDP per capita of a country. The line chart in 5.7 depicting the correlation between increasing GDP per capita and decreasing kidnapping rates suggests that economic prosperity may be associated with improvements in these factors, which could contribute to a decrease in violent or assault crimes.

Our first research question helps us show the different types of crimes that correlated with each GDP per capita group, but it is important to note that correlation does not imply causation, so deeper research is needed to examine the underlying mechanisms and dynamics that influence the relationship between GDP per capita and crime. Because this question and overall project highlights only a subset of countries by their GDP per capita level it is essential to avoid making unwarranted generalisations or claims based solely on observed correlations.

Our second question helps examine the countries if there is a lagging phase between GDP per capita and crimes with a good correlation. The analysis of lagging phases across continents reveals varied dynamics in the response of crime rates to changes in GDP. The findings suggest that, in general, crime responses occur almost immediately following economic changes, as indicated by responses clustered around zero in Africa, Asia, Europe, and South America. However, notable outliers and variations in response times exist, suggesting that specific crime types or regions may experience delayed effects.

The analysis highlights that changes in GDP per capita do not consistently lead to immediate changes in crime rates across continents. While many responses occur promptly, there are exceptions where crime rates respond with significant delays. Just because crime rates respond promptly to the change in GDP per capita a few times, we cannot fully conclude that GDP per capita is closely correlated to the factor, these may be coincidental or there may be.

In our third research question, the analysis suggests that high corruption scores in a country may lead to under reporting and inefficacy of law enforcement, potentially resulting in lower reported crime levels which can definitely lead to an inefficient or biased conclusion.

This is a problem we also encountered during the analysis of our fifth research question with Low GDP per capita countries, because of the lack of data within that subset group it was difficult

to come to a fair or finite conclusion for low GDP per capita countries. We also learnt from our fifth question that there are factors better associated with crime other than GDP per capita, while GDP per capita does have some good correlations with crimes such as homicides and thefts, there are better more correlated factors such as the human development index with homicide.

Our sixth question also highlights the relationship with other factors like happiness index and the unemployment rate. High happiness countries having very low homicide counts and high unemployment rates counties having a high theft count.

Overall, the analysis from our research questions helps shed light on the relationship between GDP per capita and the type of crime, but it also emphasises the need to consider other factors that have better correlation to the types of crimes.

A limitation mentioned many times in this report of our project was the lack of data with low GDP per capita countries. This limitation hindered our ability to draw accurate and robust conclusions for this specific subset of countries To address this limitation and improve for future iterations of this project, some approached can be taken:

- A more comprehensive data collection process should be undertaken to ensure a sufficient representation of low GDP per capita countries.
- We can compute some data imputation techniques, but this was avoided as we might miss big spikes or dips when analysing the imputed data.
- We could have also done case studies on some countries with low GDP per capita that had a lot of data. This qualitative approach can help uncover important socio-cultural, political, and economic dynamics that quantitative data alone may not capture

In the future if we were to conduct a similar analysis, we would do a case study on one country, for example United States America. High GDP per capita countries have a plethora of data!

By narrowing the scope to a single country, we can gain a more in-depth understanding of the country's unique socio-economic, cultural, and political context. This allows for a more broader analysis of the relationship between GDP per capita and crime rates within that specific context.,

Focusing on a single country like the US allows for a more comminuted analysis at the sub national level. Data analysts like us can explore variations in crime rates within different regions or cities of the country, identifying localised patterns and factors that may influence crime. This level of detail provides valuable insights for policymakers and law enforcement agencies to develop targeted interventions.

To conclude this report, our project on the relationship between GDP per capita and crime rates has provided valuable insights and raised important considerations. While GDP per capita can influence the type of crime in a country, it is crucial to recognize that it may not be the sole determinant and there are better factors.

By conducting more focused and detailed studies on individual countries, we can gain a deeper understanding of the contextual factors influencing crime rates and develop targeted strategies to address them.

Acknowledgements

Our module coordinator **Barry Symth** gave us weekly updates examining our analysis on our research questions

Our demonstrator **Laura Dunne** help us every Wednesday to view our progress, giving us advice on what other economic factors we could include in our project and how to interpret our results

Bibliography

1. Detotto, C. & Otranto, E. Does crime affect economic growth? *Kyklos* **63,** 330–345 (2010).

Appendices

You can use your appendix to include information or data that you feel is not needed for your main report but that is still relevant to your project. This might include secondary results or additional results that are similar to those presented in the main report.

Appendix A – Individual Contributions

For group projects use this section to briefly summarise the contribution of each group member. Don't go overboard here but clearly indicate who was responsible for what RQ for example.

Kenneth Obahayujie was responsible for research questions 1,5,6

Arsenii Troitskii was responsible for research questions 2,3,4

Appendix B – Datasets & Notebooks

Please list the filenames of the main datasets used in your project along with a brief (one-sentence) description of each. Also list the notebooks with a similarly brief summary of their purpose. You can indicate too who was responsible for which dataset/notebook by indicating the initials of the student beside the dataset or notebook file.

Datasets Used

1. Full_Combined_Data.csv - The final version of the dataset, which was primarily used for all analyses, included data on various indicators and crime statistics from all countries.

Notebooks Used

1. Happiness_Index.ipynb - cleaning and preparation happiness index data
2. human_development_index_cleaning.ipynb - cleaning and preparation human development index
3. unemployment-rate_downloading.ipynb - downloading and preparation unemployment rate dataset
4. Inequality_index_cleaning.ipynb - cleaning inequality data and preparation for the future analyse
5. Crime.ipynb - notebook for cleaning and preparation of crime data

6. RQ2_Final.ipynb - notebook for research question 2
7. RQ3_FInal_version.ipynb - notebook for research question 3
8. RQ4_Final_Version.ipynb - notebook for research question 4
9. rs1.ipynb - python notebook for answering research question 1
10. rs5.iypnb - Python notebook for answering research question 5
11. rs6.iypnb - Python notebook for answering research question 6
12. Clean&Preprocess.ipynb - notebook cleaning and preparing half of the datasets to merge.
13. 10_downloads.ipynb - downloading all necessary csv files from Github and saving them.
14. 1000_scraping.ipynb - notebook for scraping the crime indexes from the website Numbeo.
15. web_scrap.ipynb - scrapping GDP per 2023 year

Appendix C – Supplementary Results

This is a sample of all the countries included in each GDP per capita dataframe:

GDP per capita subsets Groups

Unique countries in low GDP per capita DataFrame:

['Afghanistan' 'Angola' 'Armenia' 'Azerbaijan' 'Bangladesh' 'Benin' 'Bhutan' 'Bolivia' 'Burkina Faso' 'Burundi' 'Cambodia' 'Cameroon' 'Cape Verde' 'Central African Republic' 'Chad' 'China' 'Comoros' 'Djibouti' 'Egypt' 'Eritrea' 'Ethiopia' 'Georgia' 'Ghana' 'Guinea' 'Guinea-Bissau' 'Guyana' 'Haiti' 'Honduras' 'India' 'Indonesia' 'Kenya' 'Kiribati' 'Lebanon' 'Lesotho' 'Liberia' 'Madagascar' 'Malawi' 'Mali' 'Mauritania' 'Moldova' 'Mongolia' 'Mozambique' 'Myanmar' 'Nepal' 'Nicaragua' 'Niger' 'Nigeria' 'Pakistan' 'Papua New Guinea' 'Paraguay' 'Philippines' 'Rwanda' 'Senegal' 'Serbia' 'Sierra Leone' 'Solomon Islands' 'Sri Lanka' 'Sudan' 'Tajikistan' 'Tanzania' 'Togo' 'Turkmenistan' 'Uganda' 'Ukraine' 'Uzbekistan' 'Vietnam' 'Zambia' 'Zimbabwe']

Unique countries in median GDP per capita DataFrame:

['Albania' 'Algeria' 'Antigua and Barbuda' 'Argentina' 'Bahrain' 'Barbados' 'Belarus' 'Belize' 'Bosnia and Herzegovina' 'Botswana' 'Brazil' 'Bulgaria' 'Chile' 'Colombia' 'Costa Rica' 'Croatia' 'Cyprus' 'Dominica' 'Dominican Republic' 'Ecuador' 'El Salvador' 'Equatorial Guinea' 'Estonia' 'Fiji' 'Gabon' 'Grenada' 'Guatemala' 'Hungary' 'Jamaica' 'Jordan' 'Kazakhstan' 'Latvia' 'Libya' 'Lithuania' 'Malaysia' 'Maldives' 'Malta' 'Mauritius' 'Mexico' 'Morocco' 'Namibia' 'Oman' 'Panama' 'Peru' 'Poland' 'Portugal' 'Romania' 'Russia' 'Samoa' 'Saudi Arabia' 'Seychelles' 'Slovenia' 'South Africa' 'Suriname' 'Thailand' 'Tonga' 'Trinidad and Tobago' 'Tunisia' 'Turkey' 'Uruguay' 'Vanuatu' 'Venezuela']

Unique countries in high GDP per capita DataFrame:

['Australia' 'Austria' 'Belgium' 'Canada' 'Denmark' 'Finland' 'France' 'Germany' 'Greece' 'Iceland' 'Ireland' 'Israel' 'Italy' 'Japan' 'Kuwait' 'Luxembourg' 'Netherlands' 'New Zealand' 'Norway' 'Qatar' 'Singapore' 'Spain' 'Sweden' 'Switzerland' 'United Arab Emirates' 'United Kingdom' 'United States']

Happiness subsets Groups

Unique countries in low happiness DataFrame:

['Afghanistan' 'Albania' 'Angola' 'Armenia' 'Azerbaijan' 'Bangladesh' 'Benin' 'Botswana' 'Bulgaria' 'Burkina Faso' 'Burundi' 'Cambodia' 'Cameroon' 'Central African Republic' 'Chad' 'Comoros' 'Djibouti' 'Egypt' 'Ethiopia' 'Gabon' 'Gambia' 'Georgia' 'Ghana' 'Guinea' 'Haiti' 'India' 'Iran' 'Iraq' 'Jordan' 'Kenya' 'Lebanon' 'Lesotho' 'Liberia' 'Madagascar' 'Malawi' 'Mali' 'Mauritania' 'Mozambique' 'Namibia' 'Nepal' 'Niger' 'Nigeria' 'North Macedonia' 'Pakistan' 'Palestine' 'Rwanda' 'Senegal' 'Sierra Leone' 'South Africa' 'Sri Lanka' 'Sudan' 'Syria' 'Tajikistan' 'Tanzania' 'Togo' 'Tunisia' 'Uganda' 'Ukraine' 'Yemen' 'Zambia' 'Zimbabwe']

Unique countries in median happiness DataFrame:

['Algeria' 'Argentina' 'Bahrain' 'Belarus' 'Belize' 'Bhutan' 'Bolivia' 'Bosnia and Herzegovina' 'Brazil' 'Chile' 'China' 'Colombia' 'Croatia' 'Cyprus' 'Czechia' 'Dominican Republic' 'Ecuador' 'El Salvador' 'Estonia' 'Greece' 'Guatemala' 'Honduras' 'Hong Kong' 'Hungary' 'Indonesia' 'Italy' 'Jamaica' 'Japan' 'Kazakhstan' 'Kuwait' 'Kyrgyzstan' 'Laos' 'Latvia' 'Libya' 'Lithuania' 'Malaysia' 'Maldives' 'Malta' 'Mauritius' 'Mexico' 'Moldova' 'Mongolia' 'Montenegro' 'Morocco' 'Nicaragua' 'Panama' 'Paraguay' 'Peru' 'Philippines' 'Poland' 'Portugal' 'Romania' 'Russia' 'Serbia' 'Singapore' 'Slovakia' 'Slovenia' 'Somalia' 'South Korea' 'Suriname' 'Taiwan' 'Thailand' 'Trinidad and Tobago' 'Turkey' 'Turkmenistan' 'Uruguay' 'Uzbekistan' 'Venezuela' 'Vietnam']

Unique countries in high happiness DataFrame:

['Antigua and Barbuda' 'Australia' 'Austria' 'Barbados' 'Belgium' 'Canada' 'Cape Verde' 'Costa Rica' 'Denmark' 'Dominica' 'Equatorial Guinea' 'Eritrea' 'Fiji' 'Finland' 'France' 'Germany' 'Grenada' 'Guinea-Bissau' 'Guyana' 'Iceland' 'Ireland' 'Israel' 'Kiribati' 'Luxembourg' 'Myanmar' 'Netherlands' 'New Zealand' 'Norway' 'Oman' 'Papua New Guinea' 'Qatar' 'Samoa' 'Saudi Arabia' 'Seychelles' 'Solomon Islands' 'Spain' 'Sweden' 'Switzerland' 'Tonga' 'United Arab Emirates' 'United Kingdom' 'United States' 'Vanuatu']

Unemployment Rate datasets

Unique countries with low unemployment rates:

['Armenia' 'Austria' 'Azerbaijan' 'Bahrain' 'Bangladesh' 'Belarus' 'Benin' 'Bhutan' 'Bolivia' 'Burkina Faso' 'Burundi' 'Cambodia' 'Cameroon' 'Chad' 'China' "Cote d'Ivoire" 'Cuba' 'Cyprus' 'Czechia' 'Democratic Republic of Congo' 'Denmark' 'Ecuador' 'El Salvador' 'Estonia' 'Ethiopia' 'Georgia' 'Germany' 'Ghana' 'Guatemala' 'Honduras' 'Hong Kong' 'Hungary' 'Iceland' 'Indonesia' 'Ireland' 'Japan' 'Kazakhstan' 'Kenya' 'Kuwait' 'Kyrgyzstan' 'Laos' 'Latvia' 'Liberia' 'Lithuania' 'Luxembourg' 'Madagascar' 'Malaysia' 'Maldives' 'Mali' 'Malta' 'Mexico' 'Moldova' 'Mozambique' 'Nepal' 'Netherlands' 'New Zealand' 'Nicaragua' 'Niger' 'Nigeria' 'North Korea' 'Norway' 'Oman' 'Pakistan' 'Panama' 'Papua New Guinea' 'Paraguay' 'Peru' 'Philippines' 'Poland' 'Qatar' 'Rwanda' 'Samoa' 'Senegal' 'Sierra Leone' 'Singapore' 'Solomon Islands' 'South Korea' 'Sweden' 'Switzerland' 'Tajikistan' 'Tanzania' 'Thailand' 'Togo' 'Tonga' 'Trinidad and Tobago' 'Turkmenistan' 'Uganda' 'Ukraine' 'United Arab Emirates' 'Uzbekistan' 'Vanuatu' 'Vietnam']

Unique countries with high unemployment rates:

['Bosnia and Herzegovina' 'Botswana' 'Congo' 'Djibouti' 'Gabon' 'Jordan' 'Lesotho' 'Libya' 'Montenegro' 'Namibia' 'North Macedonia' 'Saint Lucia' 'Saint Vincent and the Grenadines' 'Sao Tome and Principe' 'Somalia' 'South Africa' 'Sudan' 'Tunisia']

Corruption subsets Groups

Unique countries with low corruption score:

['Australia', 'Austria', 'Bahamas', 'Barbados', 'Belgium', 'Bhutan', 'Botswana', 'Brunei', 'Canada', 'Cape Verde', 'Chile', 'Cyprus', 'Denmark', 'Dominica', 'Estonia', 'Finland', 'France', 'Germany', 'Hong Kong', 'Iceland', 'Ireland', 'Israel', 'Japan', 'Luxembourg', 'Malta', 'Netherlands', 'New Zealand', 'Norway', 'Portugal', 'Puerto Rico', 'Qatar', 'Saint Lucia', 'Saint Vincent and the Grenadines', 'Singapore', 'Slovenia', 'Spain', 'Sweden', 'Switzerland', 'Taiwan', 'USA', 'United Arab Emirates', 'United Kingdom', 'Uruguay']

Unique countries with median corruption score:

['Albania', 'Algeria', 'Argentina', 'Armenia', 'Bahrain', 'Belarus', 'Belize', 'Benin', 'Bolivia', 'Bosnia and Herzegovina', 'Brazil', 'Bulgaria', 'Burkina Faso', 'China', 'Colombia', 'Costa Rica', 'Croatia', 'Cuba', 'Czechia', 'Djibouti', 'Dominican Republic', 'Ecuador', 'Egypt', 'El Salvador', 'Ethiopia', 'Fiji', 'Gabon', 'Gambia', 'Georgia', 'Ghana', 'Greece', 'Grenada', 'Guatemala', 'Guyana', 'Honduras', 'Hungary', 'India', 'Indonesia', 'Italy', 'Ivory Coast', 'Jamaica', 'Jordan', 'Kazakhstan', 'Kiribati', 'Kuwait', 'Laos', 'Latvia', 'Lebanon', 'Lesotho', 'Liberia', 'Lithuania', 'Madagascar', 'Malawi', 'Malaysia', 'Maldives', 'Mali', 'Mauritania', 'Mauritius', 'Mexico', 'Moldova', 'Mongolia', 'Montenegro', 'Morocco', 'Mozambique', 'Namibia', 'Nepal', 'Niger', 'Norfolk Island', 'North Macedonia', 'Oman', 'Palestine', 'Panama', 'Peru', 'Philippines', 'Poland', 'Romania', 'Rwanda', 'Samoa', 'Sao Tome and Principe', 'Saudi Arabia', 'Senegal', 'Serbia', 'Seychelles', 'Sierra Leone', 'Slovakia', 'Solomon Islands', 'South Africa', 'South Korea', 'Sri Lanka', 'Suriname', 'Swaziland', 'Tanzania', 'Thailand', 'Togo', 'Tonga', 'Trinidad and Tobago', 'Tunisia', 'Turkey', 'Vanuatu', 'Vietnam', 'Zambia']

Unique countries with high corruption score:

['Afghanistan', 'Angola', 'Azerbaijan', 'Bangladesh', 'Burma (Myanmar)', 'Burundi', 'Cambodia', 'Cameroon', 'Central African Republic', 'Chad', 'Comoros', 'Democratic Republic of the Congo', 'Equatorial Guinea', 'Eritrea', 'Guinea', 'Guinea-Bissau', 'Haiti', 'Iran', 'Iraq', 'Kenya', 'Kyrgyzstan', 'Libya', 'Nicaragua', 'Nigeria', 'North Korea', 'Pakistan', 'Papua New Guinea', 'Paraguay', 'Republic of the Congo', 'Russia', 'Somalia', 'Sudan', 'Syria', 'Tajikistan', 'Turkmenistan', 'Uganda', 'Ukraine', 'Uzbekistan', 'Venezuela', 'Yemen', 'Zimbabwe']