Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
F		2518688

Clear Cybercrime, Create Cleanness

-A Survey on mathematics models of Cybercrime

Summary Sheet

As Bill Gates once said: "Security is, I would say, our top priority because for all the exciting things you will be able to do with computers – organizing your lives, staying in touch with people, being creative – if we don't solve these security problems, then people will hold back."Considering the macroenvironment where the methods of cybercrimes becomes more and more elaborate and smart, which makes police and departments difficult to detect, determine and defend, we try to investigate the relationship between cybercrimes cases in the world and time, space, patterns of laws and policies, some national index and so on, and then we come up with our research results and theory according to the data we collect and what cybercrimes rules we notice, contributing our team's concerning and strategy when faced with the supervision and management of cybersafety in the macro level.

In Question 1, we work out the problems on the basis of the data we collect from the VERIS DATABASE (VSDB) and Kaspersky. Also, to exactly and strictly answer where cybercrimes are successful or thwarted, we look up a massive number of news, reports and official information on the Internet and give our own opinions. Last but not the least, to visualize our data chart of global distribution of cybercrimes, we construct our mathematics model of heatmap with the help of GIS system and Python environment.

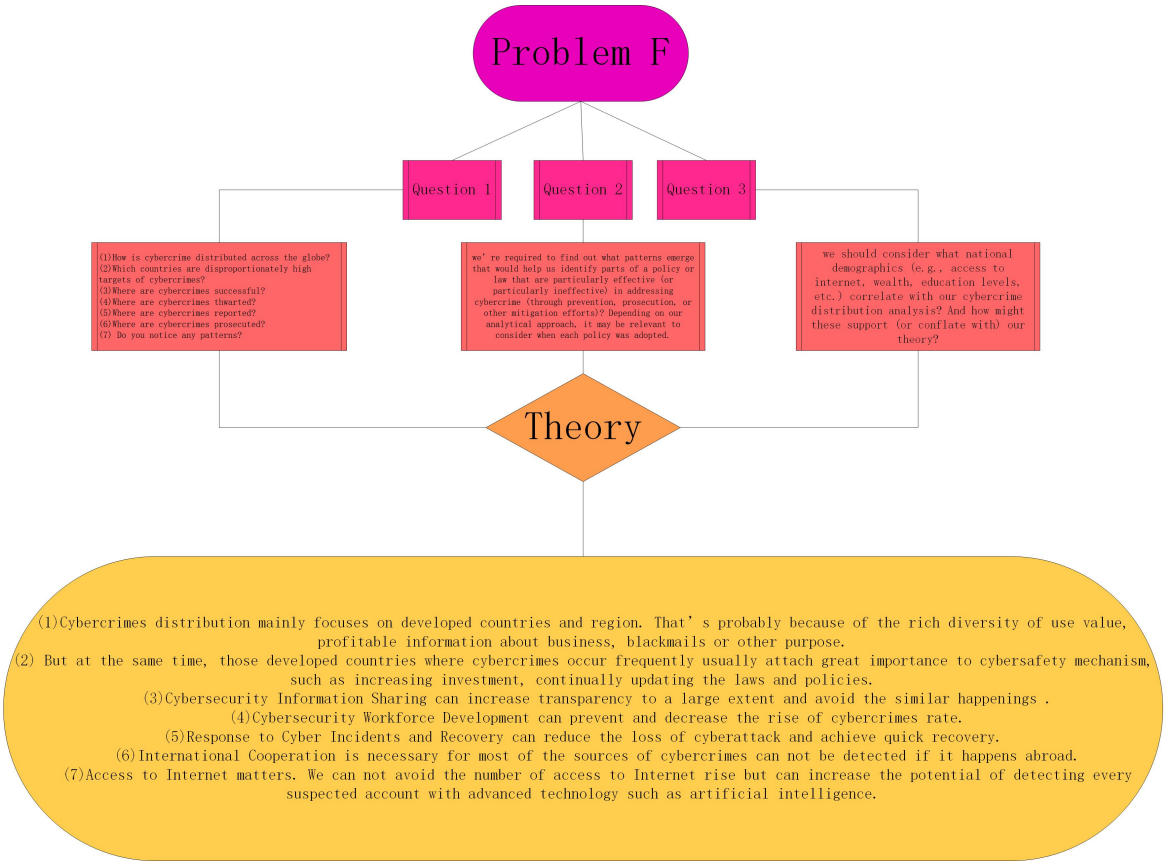
In Question 2, as far as we are concerned, if we attempt to clarify one policy or law's effect on cybercrimes, we should not just focus on when it was published.On the contrary, we're supposed to set the point of time as borderline in time axis and observe the difference between the period of tendency before and after at a macro level. Additionally, to simplify our workload, we narrow our research range and finally choose US and UK as our objectives. By the way, we notice that there are two essential national laws, the National Cybersecurity Protection Act (America, 2014) and the National Cyber Security Strategy (UK, 2014) within the time frame of our research event. Meanwhile, since the effectiveness of a policy or law varies, we naturally define the three parts of effectiveness as transparency, prevention rate and pursuit rate and to quantitatively analyze those variables we derive the corresponding formulas. Then based on what we calculate, we construct mathematics models with Matlab tools (Fourier function is verified to be the best fitting form) and list all the result values on the chart. Consequently, we make our process precise to pieces of provision in policies or laws so as to find the writing basis. At last, we summarize the patterns we identify in our research and come up with our cybercrimes theory.

In Question 3, we firstly collect the necessary data we need from ITU (https://datahub.itu.int/query/), World Bank and https://zh.tradingeconomics.com/united-states/gdp-per-capita,. As we see, the core of Question 3 is to connect national demographics (e.g., access to internet, wealth, education levels, etc.) with the number of cybercrimes happenings. We adopt Lasso regression to analyze the correlation among them. At last, we find access to Internet matters among the several national demographics.

Finally, we conclude our theory and write one-page memo to show what we consider and assume in our entire research. We always believe that one thing: Clear Cybercrime, Create Cleanness. Let's get together to construct a cyber-strong society.

Key Words: cybercrimes mathematics models analysis prediction

Contents



1 Introduction of Problem

1.1 Background

As an old saying goes:" Technology is a double-edged sword." As far as we are concerned, with advanced science and technology updating and reforming, especially in the field of computer and communication engineering, more and more of our world has become connected, making our life colorful and convenient. Nevertheless, worryingly, the more we utilize Internet platform to share our daily life, make online payment, chat with friends and strangers and so on, the more exposure and transparency our privacy will face to others, leading to a vulnerable cybersafety in front of those hackers and criminals.

What's worse, if those relevant institutes could report their situations of being attacked via Internet during or after the cybercrime, governments and the whole society will surely strengthen the initial cybersafety mechanism and policies according to the weakness exposed in every case. However, unfortunately, "many institutions, such as investment firms, are unwilling to report a hack, preferring to quietly pay a ransom demand than to let their clients and potential clients know that they were the victim of a security breach." Such negative responses, along with the difficulty to convict in cybercrimes[3], make the construction of the stability of cybersafety more and more difficult and separate.

Obviously, to solve a series of cyber problem entirely and deeply is of great urgency, which needs everyone of our citizens' participation and endeavor. Also, what we emphasize as above is what we

focus on as below.

1.2 Clarification

The requires of three questions are mentioned as below.

Question 1
In question 1, we're required to work out these problem as below:
(1)How is cybercrime distributed across the globe?
(2)Which countries are disproportionately high targets of cybercrimes?
(3)Where are cybercrimes successful?
(4)Where are cybercrimes thwarted?
(5)Where are cybercrimes reported?
(6)Where are cybercrimes prosecuted?
(7) Do you notice any patterns?

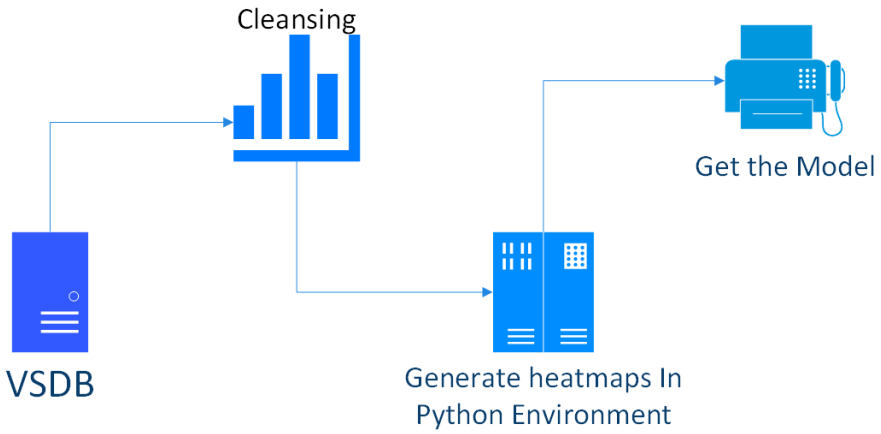
Question 2
In question 2, we're required to find out what patterns emerge that would help us identify parts of a policy or law that are particularly effective (or particularly ineffective) in addressing cybercrime (through prevention, prosecution, or other mitigation efforts)? Depending on our analytical approach, it may be relevant to consider when each policy was adopted.

Question 3
In question 3, we should consider what national demographics (e.g., access to internet, wealth, education levels, etc.) correlate with our cybercrime distribution analysis? And how might these support (or conflate with) our theory?

2 Analysis of Problem

2.1 Question 1

The key to the solutions of question 1 is the data related to cybercrime. Hopefully, through our unremitting efforts, we finally collect the precious data of cybercrime of happenings, reports, prosecutions and so on from the VERIS DATABASE (VSDB) and www.kaspersky.com.cn..



And then, we draw the "Global Heatmap of Event Occurrence" titled "Global Distribution of Cybercrime Happenings" with the help of Python compiler language in response to problem (1), visualizing it and reflecting it intuitively.(see the mathematics model in Constructions of Mathematics models 1)

To quantitatively analyze which countries are disproportionately high targets of cybercrimes (problem (2)), we consider taking the data "Kaspersky Anti-Spam" from www.kaspersky.com.cn. as a reference (KAS (Kaspersky Anti-Spam) shows suspicious and unwanted email traffic discovered by Kaspersky's Reputation Filtering technology.), and surprisingly find Russia, Germany, Netherlands, France,Hong Kong (China), United Kingdom these countries or regions account for a large proportion.(table 1)

Table 1: Kaspersky Anti - Span

Country	Kaspersky Anti - Span
Russia	5.53
Germany	3.95
Netherlands	3.12
France	2.79
Hong Kong	2.58

(Data Source:Kaspersky)

To answer problem (3) and (4), We have reviewed a massive amount of news and reports. We noticed that the countries where cybercrimes are committed successfully includes China (In May 2023, a large high-tech company in China specializing in smart energy and digital information was suspected of being attacked by U.S. intelligence agencies. The attackers exploited a vulnerability in Microsoft Exchange to infiltrate and control the email server, implanting backdoor programs to steal email data and attacking over 30 devices.), the United States (In November 2023, the Idaho National Laboratory (INL) in the United States confirmed that it had suffered a cyber attack, with the hacker group "SiegedSec" leaking INL's human resources data online, which included detailed information on "hundreds of thousands" of employees, system users, and citizens.) and so on. Meanwhile, there are also some cases where cybercrimes are defeated. For example, the UK's National Cyber Security

Centre (NCSC) is actively engaged in cyber security protection. From September 2023 to August 2024, the NCSC assisted in 430 cybersecurity incidents, including preventing some ransomware incidents from having a more serious impact on national security and the economy.

According to the date (VSDB) we possess(fig. 1), we're able to solve problem (5)(6): the amounts of reports in each country in order is as follows: America (645), United Kingdom (64), India (38);the amounts of prosecution in each country in order is as follows: America (390), United Kingdom (44), Canada(18).

Country	Reports
America	645
United Kingdom	64
India	38
Canada	37
Israel	18

Table 2: Reports

Country	Prosecutions
America	390
United Kingdom	40
Canada	18
Australia	14
New Zealand	7

Table 3: Prosecutions

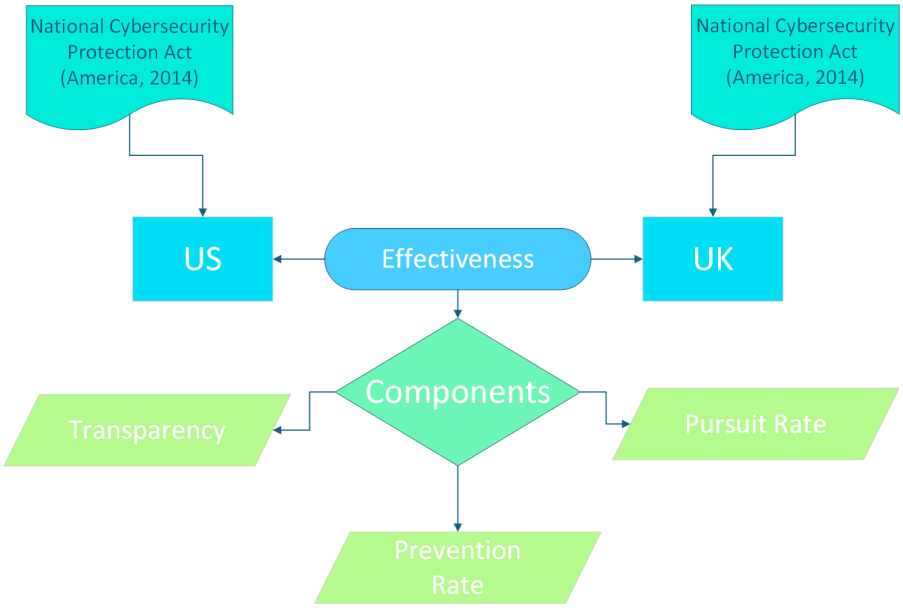
Figure 1: (Data Source:Veris Community Database)

Based on the analysis as above, we can roughly summarize some patterns:

(1) All in all, cybercrimes distribution mainly focuses on developed countries and region. That's probably because of the rich diversity of use value, profitable information about business, blackmails or other purpose.

(2) But at the same time, those developed countries where cybercrimes occur frequently usually attach great importance to cybersafety mechanism, such as increasing investment, continually updating the laws and policies.

2.2 Question 2



Some essential sentences in Question 2 prompt us to analyse the effectiveness of a particular cybersafety policy or law in terms of the prevention, prosecution and so on against cybercrimes, and suggest us to take when the policy or law is adopted as consideration.

For the convenience of description and the brevity of language, we decide to introduce the three concepts which are decisive factors in terms of the effectiveness of a particular policy or law as below:

(1)Transparency. Transparency in great measure reflects the social awareness of cybercrimes cases reported by relevant stakeholders such as governments, victims (corporate, companies, banks) and so on. Approximately, transparency equals the rate of reports to happenings.

$$\gamma(i) = \frac{\beta(i)}{\alpha(i)}$$

$\gamma(i)$ means the transparency in the i-th year.
$\beta(i)$means the number of cybercrimes reports in the i-th year.
$\alpha(i)$ means the number of cybercrimes happenings in the i-th year.

Table 4: Transoarency

Year	US Transparency	UK Transparency
2010	0.16	0
2011	0.05	0.07
2012	0.11	0.14
2013	0.10	0.1
2014	0.10	0.15
2015	0.12	0.06
2016	0.16	0.13
2017	0.16	0.10

(2)Prevention rate. Prevention rate directly reflects a policy or law's influence on present or future cybercrimes and its tendency. Concisely, prevention rate is up to the derivative of fitting function of happening rate. (Of course, if we want it more exactly and strictly, we must consider more relevant factors that effect prevention. Due to the limitation of time, we decide to leave it to those who are interested in our thought and let them optimize our results and theory.) Additionally, cybercrimes happenings rate equals the rate of happenings to access to internet.

$$\theta(i) = -\frac{dH(i)}{dt}$$
$$H(i) = \frac{\alpha(i)}{\varepsilon_1(i)}$$

$\theta(i)$ means prevention rate in the i-th year.
$H(i)$means the rate of cybercrimes happenings to access to Internet in the i-th year.
$\alpha(i)$ means the number of cybercrimes happenings in the i-th year.
$\varepsilon_1(i)$ means the number of access to Internet in US in the i-th year.

usually, the prevention rate features positive, negative and zero values as follow:

$$\begin{cases} \varepsilon > 0, & \text{Positive effect} \\ \varepsilon = 0, & \text{No effect} \\ \varepsilon < 0, & \text{Negative effect} \end{cases}$$

Table 5: Happeing Rate

Year	US Pursuit($\times 10^{-5}$)	UK Pursuit($\times 10^{-6}$)
2010	1.07	0.19
2011	0.88	0.78
2012	1.87	1.5
2013	2.62	2.1
2014	1.36	1.5
2015	1.37	1.3
2016	1.11	1.5
2017	0.65	0.9

(3)Pursuit rate. Pursuit rate actually reflects how many victims choose to pursue their legitimate rights and interests in cybercrimes cases. In reality, the reasons why victims give up prosecuting vary from high costs to difficulty to trace the source of cases or convict. So a high pursuit rate is what we expect in the future. Accordingly, pursuit rate equals the rate of prosecutions to reports.

$$\delta(i) = \frac{\lambda(i)}{\beta(i)}$$

$\delta(i)$ means the pursuit rate in the i-th year.
$\lambda(i)$means the number of cybercrimes prosecutions in the i-th year.
$\beta(i)$means the number of cybercrimes reports in the i-th year.

Table 6: Pursuit

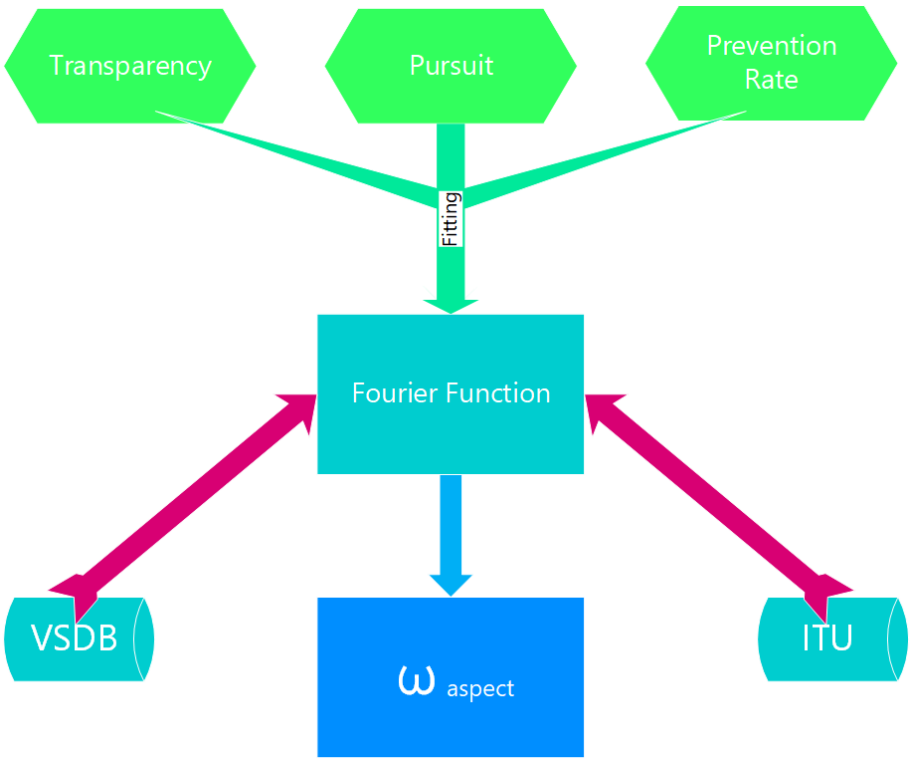
Year	US Pursuit	UK Pursuit
2010	0.67	0
2011	0.67	0.67
2012	0.49	0.25
2013	0.56	0.83
2014	0.69	0.70
2015	0.84	1
2016	0.56	0.54
2017	0.46	0.67

Based on what we collect on VSDB, firstly we gather all the data from 2010 to 2017 to generate line charts reflecting the dependent values varying with time (year).

Take the National Cybersecurity Protection Act (America, 2014) here as an example.

Considering there are 5 points of samples from 2010 to 2014, we can fit a function accordingly.

Fourier function is verified to be the best fitting form.



2.2.1 Fourier Series Fitting Derivation

The Fourier series is a powerful tool for representing periodic functions as a sum of sine and cosine functions. Consider a periodic function $y = f(x)$ with period T. We can represent $f(x)$ as a Fourier series:

$$f(x) = a_0 + \sum_{n=1}^{\infty}(a_n \cos(\frac{2n\pi x}{T}) + b_n \sin(\frac{2n\pi x}{T}))$$

where a_0, a_n, and b_n are the Fourier coefficients.

Derivation of a_0 We first find a_0. Integrate both sides of the Fourier series equation over one period $[x_0, x_0 + T]$:

$$\int_{x_0}^{x_0+T} f(x)dx = \int_{x_0}^{x_0+T} a_0 dx + \sum_{n=1}^{\infty}\left(a_n \int_{x_0}^{x_0+T} \cos(\frac{2n\pi x}{T})dx + b_n \int_{x_0}^{x_0+T} \sin(\frac{2n\pi x}{T})dx\right)$$

We know that:

$$\int_{x_0}^{x_0+T} \cos(\frac{2n\pi x}{T})dx = \left[\frac{T}{2n\pi}\sin(\frac{2n\pi x}{T})\right]_{x_0}^{x_0+T} = 0, \quad n = 1, 2, \cdots$$

$$\int_{x_0}^{x_0+T} \sin(\frac{2n\pi x}{T})dx = \left[-\frac{T}{2n\pi}\cos(\frac{2n\pi x}{T})\right]_{x_0}^{x_0+T} = 0, \quad n = 1, 2, \cdots$$

And $\int_{x_0}^{x_0+T} a_0 dx = a_0 T$. So,

$$a_0 = \frac{1}{T}\int_{x_0}^{x_0+T} f(x)dx$$

Derivation of a_n Multiply both sides of the Fourier series equation by $\cos(\frac{2m\pi x}{T})$ and integrate over one period $[x_0, x_0+T]$:

$$\int_{x_0}^{x_0+T} f(x)\cos(\frac{2m\pi x}{T})dx = \int_{x_0}^{x_0+T} a_0\cos(\frac{2m\pi x}{T})dx + \sum_{n=1}^{\infty}\left(a_n\int_{x_0}^{x_0+T}\cos(\frac{2n\pi x}{T})\cos(\frac{2m\pi x}{T})dx + b_n\int_{x_0}^{x_0+T}\sin($$

We use the following trigonometric identities:

$$\cos A\cos B = \frac{1}{2}[\cos(A+B)+\cos(A-B)]$$

$$\sin A\cos B = \frac{1}{2}[\sin(A+B)+\sin(A-B)]$$

For $m \neq n$:

$$\int_{x_0}^{x_0+T}\cos(\frac{2n\pi x}{T})\cos(\frac{2m\pi x}{T})dx = 0$$

$$\int_{x_0}^{x_0+T}\sin(\frac{2n\pi x}{T})\cos(\frac{2m\pi x}{T})dx = 0$$

For $m = n$:

$$\int_{x_0}^{x_0+T}\cos^2(\frac{2n\pi x}{T})dx = \frac{T}{2}$$

And $\int_{x_0}^{x_0+T} a_0\cos(\frac{2m\pi x}{T})dx = 0$ for $m \geq 1$. So,

$$a_n = \frac{2}{T}\int_{x_0}^{x_0+T} f(x)\cos(\frac{2n\pi x}{T})dx, \quad n = 1, 2, \cdots$$

Derivation of b_n Multiply both sides of the Fourier series equation by $\sin(\frac{2m\pi x}{T})$ and integrate over one period $[x_0, x_0+T]$:

$$\int_{x_0}^{x_0+T} f(x)\sin(\frac{2m\pi x}{T})dx = \int_{x_0}^{x_0+T} a_0\sin(\frac{2m\pi x}{T})dx + \sum_{n=1}^{\infty}\left(a_n\int_{x_0}^{x_0+T}\cos(\frac{2n\pi x}{T})\sin(\frac{2m\pi x}{T})dx + b_n\int_{x_0}^{x_0+T}\sin(\frac{2n\pi x}{T})\sin(\frac{2m\pi x}{T})dx\right)$$

Using the trigonometric identities and similar integration techniques as above, we get:

$$b_n = \frac{2}{T}\int_{x_0}^{x_0+T} f(x)\sin(\frac{2n\pi x}{T})dx, \quad n = 1, 2, \cdots$$

2.2.2 Addition

By the way, we can also get predictive value of 2015-2017 which will be used to analyze the effectiveness later.

Now we define it as the effectiveness of a certain aspect (transparency, prevention rate, pursuit rate) in one policy or law. Considering one policy or law needs time to implement thoroughly, which means that the effect of policies and laws becomes more representative over time, so we naturally derive the following formula:

$$\omega_{aspect} = \sum_{i=2015}^{2017}(i - 2014)(y_i - \hat{y}_i)$$

Based on sample size and workload of data processing, we narrow our range of research objectives and choose to analyze the two countries UK and US.

To identify what patterns of the policy or law effect or not in cybercrimes, we look up hundreds of regulations and find something we think is essential and meaningful.

Similarly, the analysis of UK follow the reasoning process as above. We notice that the National Cyber Security Strategy is published by UK government in 2014.

2.3 Question 3

Question 3 requires us to additionally take national demographics (e.g., access to internet, wealth, education levels, etc.) as consideration in our mathematics models.

Based on what we collect, we choose to link the national demographics of US with cybercrimes happenings.

Based on the data from ITU (https://datahub.itu.int/query/), World Bank and https://zh.tradingeconomics.com, after repeated examination, we choose to fit our data as the form of Lasso regression.

2.3.1 Data Preprocessing

Centering and Standardization: For each independent variable (such as GDP per capita, education level, and internet users), we perform standardization (i.e., subtracting the mean and dividing by the standard deviation) to ensure all variables are on the same scale. This prevents the units of some features from influencing the model fitting process. 2.

2.3.2 Lasso Regression Model Construction

Lasso Regression: Lasso regression involves adding an L1 regularization term to the regression coefficients, which forces the coefficients of less important features to approach zero, thus performing feature selection. The goal of Lasso regression is to:

$$\textbf{Minimize } ||y - X\beta||^2 + \lambda||\beta||_1$$

Where:

λ is the regularization parameter that controls the level of feature selection. Lasso regression optimizes the strength of feature selection by adjusting the value of λ.

Cross-validation to Choose the Optimal Lambda: We use 5-fold cross-validation to select the optimal λ value. By using the CV option in the lasso function, Lasso regression automatically selects a range of candidate λ values and uses cross-validation to choose the best λ, which results in the optimal model.

3 Descriptions of symbol and definitions of terms

The meaning of each symbol is as below:

(1)$\alpha(i)$ means the number of cybercrimes happenings in the i-th year.
(2)$\beta(i)$means the number of cybercrimes reports in the i-th year.
(3)$\lambda(i)$means the number of cybercrimes prosecutions in the i-th year.
(4)$\gamma(i)$ means the transparency in the i-th year.
(5)$\delta(i)$ means the pursuit rate in the i-th year.
(6)$\theta(i)$ means prevention rate in the i-th year.
(7)$H(i)$means the rate of cybercrimes happenings to access to Internet in the i-th year.
(8)$\omega_{Transparency}$ means the effectiveness of a policy or law in transparency.
(9)$\omega_{Pursuit}$ means the effectiveness of a policy or law in pursuit rate.
(10)$\omega_{Prevention}$ means the effectiveness of a policy or law in prevention rate.
(11)$\varepsilon_1(i)$ means the number of access to Internet in US in the i-th year.
(12)$\varepsilon_2(i)$ means GDP per capita (K,USD) in US in the i-th year.
(13)$\varepsilon_3(i)$ means years of education per capita in US in the i-th year.

4 Assumption of Models

Our mathematics models have following characteristics:

(1)In our mathematics models, we agree that the so-called cybercrimes is broadly defined, including telecom fraud, hackers attack and so on.[1]

(2)We assume that all the cybercrimes prosecution come from reports and all the reports come from happenings, in order to simplify our workload.

(3)In our mathematics models in Question 2, we assume that only the patterns of the policies and laws participate and see other factors as unnecessary variables.

(4) In our mathematics models in Question 3, we ignore other factors that might effect our research when we add the factors of access to internet, wealth, education levels to our original models.

5 Constructions of Mathematics Models

5.1 Models in Question 1

Based on the available data, we come up with a heatmap titled "Global Distribution of Cybercrime Happenings" in Python environment, so as to describe the global distribution of cybercrimes figura-

tively.

Relevant Python code is in the Appendix-the Code of Heatmap

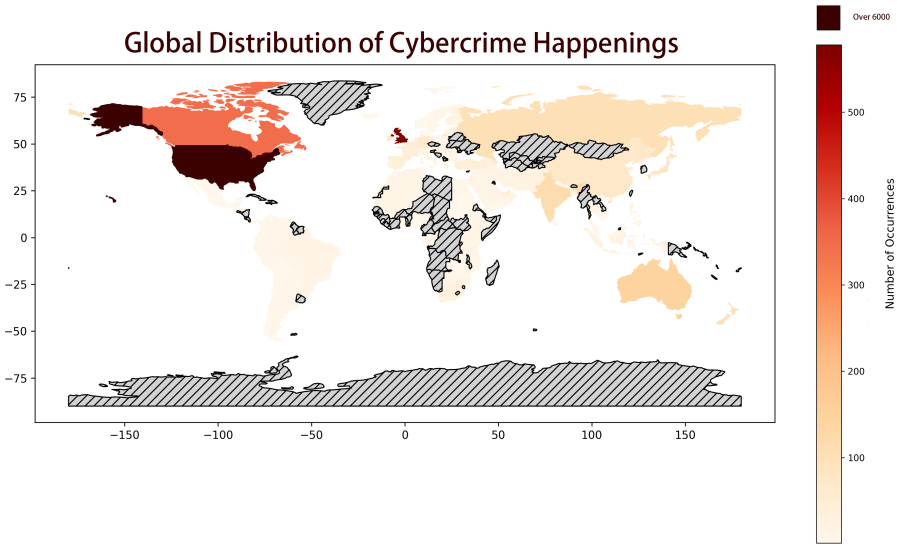
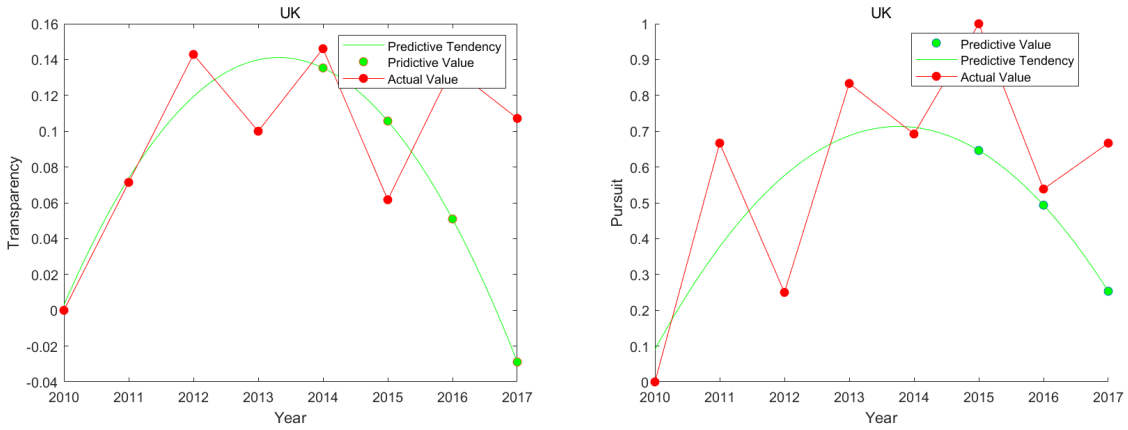
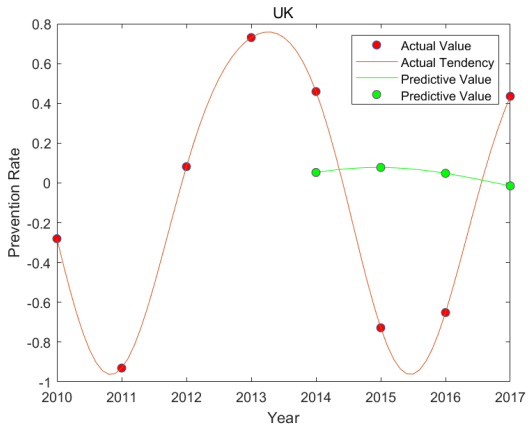


Figure 2: Global Distribution of Cybercrime Happenings

5.2 Models in Question 2

5.2.1 UK





$$\gamma(i) = -27876.77 + 27877\cos(0.0015 \times \frac{x-2012}{1.58}) + 34.876\sin(0.0015 \times \frac{x-2012}{1.58})$$

$$\delta(i) = -3.7794 \times 10^6 + 3.7794 \times 10^6 \cos(-2.4115 \times 10^{-4} \times \frac{x-2012}{1.58})$$
$$- 1.0171 \times 10^3 \sin(-2.4115 \times 10^{-4} \times \frac{x-2012}{1.58})$$

$$\theta(i) = 0.1096 + 0.0499\cos(1.6228 \times \frac{x-2012}{1.58}) + 0.0749\sin(1.6228 \times \frac{x-2012}{1.58})$$

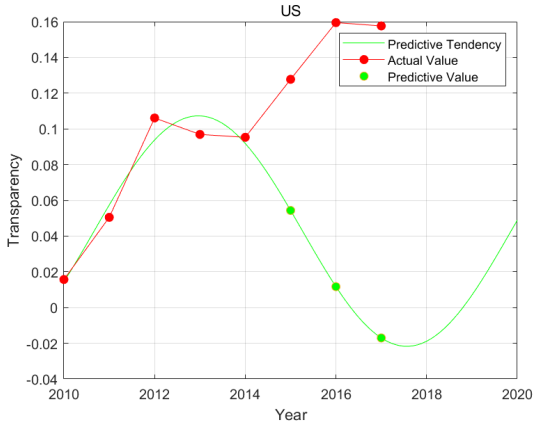
$$\theta'(i) = 0.01214 + 0.632 \times \cos(2.5531 \times \frac{x-2012}{1.8793})$$
$$+ 0.0166 \times \sin(2.5531 \times \frac{x-2012}{1.8793}) - 0.0079\cos(2 \times 2.5531 \times \frac{x-2012}{1.8793})$$
$$- 0.0384 * sin(2 \times 2.5531 \times \frac{x-2012}{1.8793});$$

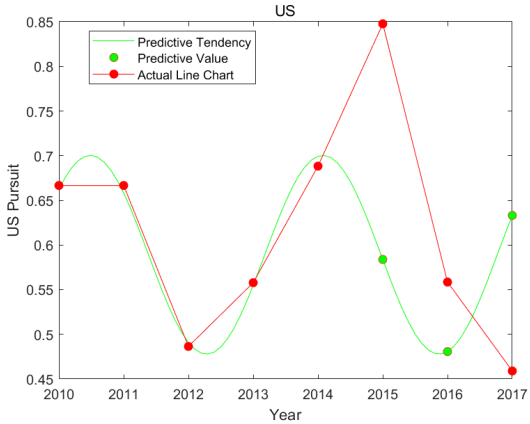
$$\omega_{Transparency} = 0.626374$$

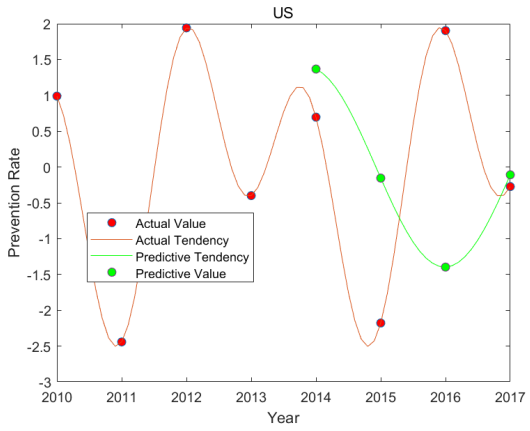
$$\omega_{Pursuit} = 1.684249$$

$$\omega_{Prevention} = 3.5561563$$

5.2.2 US







$$\gamma(i) = 0.0428 + 0.0510\cos(1.0802 \times \frac{x-2012}{1.58}) + 0.0395\sin(1.0802 \times \frac{x-2012}{1.58})$$

$$\delta(i) = 0.5891 - 0.0981\cos(2.7674 \times \frac{x-2012}{1.581}) - 0.0525\sin(2.7674 \times \frac{x-2012}{1.581})$$

$$\theta(i) = 1.6150 + 0.3629\cos(2.3344 \times \frac{x-2012}{1.58}) + 0.8738\sin(2.3344 \times \frac{x-2012}{1.58})$$

$$\begin{aligned}\theta'(i) = {} & 1.5105 + 0.6145 \times \cos(2.5531 \times \frac{x-2012}{1.8793}) \\ & - 0.3343 \times \sin(2.5531 \times \frac{x-2012}{1.8793}) - 0.1051\cos(2 \times 2.5531 \times \frac{x-2012}{1.8793}) \\ & - 0.4342 * sin(2 \times 2.5531 \times \frac{x-2012}{1.8793});\end{aligned}$$

$$\omega_{Transparency} = 0.8928281$$

$$\omega_{Pursuit} = 1.115922$$

$$\omega_{Prevention} = 9.119839$$

5.2.3 The Table of ω_{aspect}

Table 7: Omega

ω_{aspect}	UK	US
$\omega_{Transparency}$	0.63	0.89
$\omega_{Pursuit}$	1.69	1.11
$\omega_{Prevention}$	3.56	9.11

Hopefully, all of the ω_{aspect} we get are positive numbers, which suggests the policies and laws of the two countries are effective.Now let's dig out the two policies and laws and find out what patterns in earth work.

First of all, let's see some provisions in National Cybersecurity Protection Act (America, 2014).

(1)Cybersecurity Information Sharing: “the government and private sector may share information related to cyber threats, including malware, vulnerabilities, and other cybersecurity risks.” The measure definitely enhance the transparency of every cybercrime case and strengthen the defense line of cybersafety. That’s because sharing intelligence about cyber threats means that public will understand the ways the criminals adopt to invade the Internet, and so that they will update their defense system in case of the similar invasion.

(2)Cybersecurity Workforce Development: “the Act includes provisions to develop and expand the cybersecurity workforce through education, training, and certification programs.” The piece of provision improve the effectiveness of prevention in our opinion for its comprehension of the personnel training at all aspects. The more manpower means the quicker response in cybercrimes occurrence and the stronger defense.

And then let’s see the highlights in the National Cyber Security Strategy (UK, 2014).

(1)Response to Cyber Incidents and Recovery:”the strategy emphasizes improving cyber incident response and recovery capabilities, ensuring that when a cyber incident occurs, the UK can quickly recover and minimize the impact on the economy and society.” The measure increase both the transparency and pursuit rate for it guide society and public on how to deal with the situation when cybercrime happens.

(2)International Cooperation: ”The strategy emphasizes working with other countries and international organizations to enhance global cybersecurity capabilities and combat transnational cybercrime.” we should aware a fact that a large part of cybercrimes is transnational, which indicates that only when every nation makes joint effort can cybercrimes be defeated.

5.3 Models in Question 3

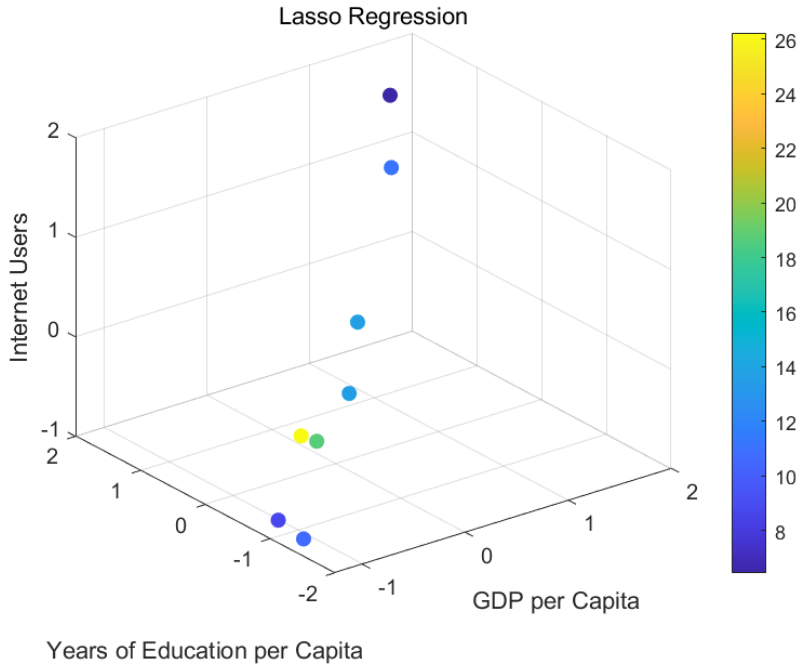


Figure 3: Lasso Regression Figure

Lasso Result: $y = 0.0000 \times x_1 + 0.0000 \times x_2 + -1.6986 \times x_3 + 13.6495$

6 Stability Analysis

6.1 Question 3

To test the stability of the model, we use 5 - fold cross - validation:

6.1.1 Calculation

Divide the data into 5 folds. Each time, select 4 folds as the training set and 1 fold as the test set, and calculate the RMSE (Root Mean Square Error) of each prediction.

Calculate the RMSE for each fold and compute the average RMSE of all folds. The lower the average RMSE and the smaller the fluctuation, the more stable the model is.

Of course, Lasso regression will automatically choose the optimal Lambda, and calculate the prediction error of each fold based on this Lambda.

Plotting RMSE Plot of Cross - Validation: Plot the RMSE of each fold, which helps us visually observe the performance of the model under different data partitions. If the RMSE values fluctuate significantly, it indicates that the model has poor stability; if the fluctuations are small and the error is low, it means the model has high stability.

Average RMSE Output the average RMSE of all folds, which serves as an important indicator of the model's stability. If the average RMSE is small, it shows that the Lasso regression model has a good fitting effect on different training and test sets, demonstrating strong stability.

6.1.2 Conclusion

Lasso regression automatically selects important features through L1 regularization, avoiding the impact of highly collinear features on the model and improving the model's stability.

(In the beginning,we choose Quadratic Regression,but it didn't perform well.At last,we choose Lasso Regression)

We finally use Cross - validation as an effective way to evaluate the model's stability,which can ensure that the model adapts to the training data and generalizes well to new data.

6.1.3 Result Presentation

The results are as followed:

Optimal Lambda Value of Lasso Regression: 2.3625

It is crucial for choosing the regularization strength.

Cross - Validation Average RMSE:5.4155

The model's stability can be accepted.

RMSE Fluctuation Plot:

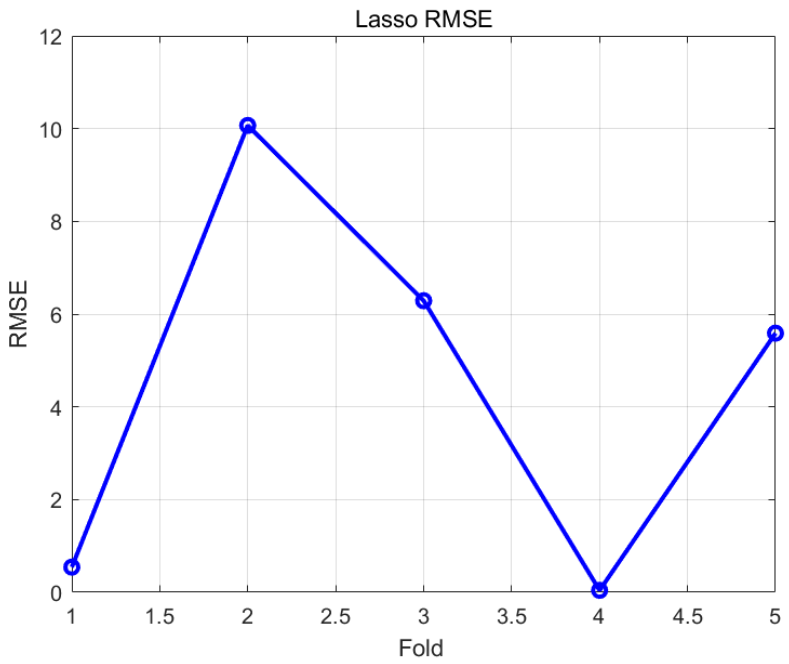


Figure 4: RMSE

A figure about the stability of the model.

7 Strengths and Weakness

7.1 Strengths

(1)All of the processes of derivation are based on tremendous data collected on the Internet, making our results authentic and practical.

(2)Every model we fit is the output after hundreds of examinations and analysis, which means that the results are solid.

(3)To clarify the effectiveness of policies and laws better and clearer, we introduce some concepts and terms, such as transparency, pursuit rate, making our analysis comprehensive and sustainable.

7.2 Weakness

(1)For some uncontrollable factors, the model we construct in Question 3 can not explain some of the national demographics well.

(2)Due to lack of some essential data, we have no idea to go a further step to study other factors related to cybercrimes.

8 Conclusion

After working out the three question, now we can construct our theory based on the proves and the process of derivation as above.

(1)Cybercrimes distribution mainly focuses on developed countries and region. That's probably because of the rich diversity of use value, profitable information about business, blackmails or other purpose.

(2) But at the same time, those developed countries where cybercrimes occur frequently usually attach great importance to cybersafety mechanism, such as increasing investment, continually updating the laws and policies.

(3)Cybersecurity Information Sharing can increase transparency to a large extent and avoid the similar happenings .

(4)Cybersecurity Workforce Development can prevent and decrease the rise of cybercrimes rate.

(5)Response to Cyber Incidents and Recovery can reduce the loss of cyberattack and achieve quick recovery.

(6)International Cooperation is necessary for most of the sources of cybercrimes can not be detected if it happens abroad.

(7)Access to Internet matters. We can not avoid the number of access to Internet rise but can increase the potential of detecting every suspected account with advanced technology such as artificial intelligence.[2]

9 Memo

Memo to Country Leaders Attending the ITU Summit on Cybersecurity

Date: 27th January 27, 2025

Subject: the outline and response of cybercrimes research

Objective and Context

As an old saying goes:" Technology is a double-edged sword." As far as we are concerned, with advanced science and technology updating and reforming, especially in the field of computer and communication engineering, more and more of our world has become connected, making our life colorful and convenient. However, this increased connectivity has also led to a significant rise in cybersecurity threats. Our objective is to work out the essential factors that effect on cybercrimes quantitatively and come up with our thoughts and measures based on what we find, so as to solve cybercrimes, the significant issue.

Our Theory

After our process of derive, we combine all the results with cybercrimes and summarize the patterns as follow:

(1)According to the heatmap of Global Distribution of Cybercrimes, cybercrimes mainly occur to developed countries, such as US, UK, German. So the improvement of cybersafety should keep pace with the economy development.

(2)Powerful international cooperation to combat cybercrimes can effectively decrease and prevent the happenings.

(3)Training the workforce to deal with cybercrimes can enhance the efficiency and effectiveness of the defense and response.

(4)Improving the share of cybercrimes intelligence contribute to higher transparency.

(5)Utilizing some cutting-edged technology such as artificial intelligence to detect suspects make cybercrimes hard.

Most Pressing Findings

(1)Some countries emphasize on economic indicators too much but ignore the develop hard power, letting the hackers take advantage of the chances to commit cybercrimes.

(2)Due to the contradiction of international politics, culture and so on, international cooperation can not develop deeply at the aspect of combat of cybercrimes.

(3)Some countries and regions lacks the sufficient number of workforce on cybersafety.

10 Reference List

References

[1] Sarah Gordon and Richard Ford. On the definition and classification of cybercrime. *Journal in Computer Virology*, 2:13–20, Jul 2006.

[2] Saadia Anwar Pasha, Sana Ali, and Riadh Jeljeli. Artificial intelligence implementation to counteract cybercrimes against children in pakistan. *Human Arenas*, Oct 2022.

[3] Sizwe Snail ka Mtuze and Melody Musoni. An overview of cybercrime law in south africa. *International Cybersecurity Law Review*, 4:299–323, Jun 2023.

11 Appendix

11.1 The Code of Heatmap

```
        import matplotlib.pyplot as plt
                import geopandas as gpd
                import pandas as pd


                df = excel_file.parse('Sheet1')


                df = df.rename(columns={'Happen': 'Occurred', 'country_name':
                ↪  'Country_Name'})


```

```
df['Country_Name'] = df['Country_Name'].replace('America', 'United States of America')


world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))


merged = world.merge(df, left_on='name', right_on='Country_Name', how='left')


us_data = merged[merged['name'] == 'United States of America'][['name', 'Occurred']]


merged_for_plot = merged[merged['name'] != 'United States of America']


plt.rcParams['figure.dpi'] = 300

plt.rcParams['font.sans-serif'] = ['DejaVu Sans']

fig, ax = plt.subplots(1, 1, figsize=(15, 10))


merged_for_plot.plot(column='Occurred', ax=ax, cmap='OrRd', legend=True,
missing_kwds={
        "color": "lightgrey",
        "edgecolor": "k",
        "hatch": "///",
"label": "Missing values"
},
legend_kwds={'label': 'Number of Occurrences'})

ax.set_title('Global Heatmap of Event Occurrences (excluding USA)')
plt.show()
```

11.2 The Code of Lasso Regression

```

data = readtable('US_UKdata.xlsx');
years = data.Year;
crime = data.Cybercrime;
gdp = data.GDP_per_capita;
edu = data.Education_level;
users = data.Internet_users;


gdp_c = (gdp - mean(gdp)) / std(gdp);
edu_c = (edu - mean(edu)) / std(edu);
users_c = (users - mean(users)) / std(users);


```

```
X = [gdp_c, edu_c, users_c];
y = crime;


[lasso_B, FitInfo] = lasso(X, y, 'CV', 5);


lasso_lamda = FitInfo.LambdaMinMSE;
disp(['Lasso :The best Lambda Value:', num2str(lasso_lamda)]);


lasso_B_opt = lasso_B(:, FitInfo.IndexMinMSE);
lasso_int = FitInfo.Intercept(FitInfo.IndexMinMSE);


disp('Lasso Result:');
disp('Regression Coefficient:');
disp(lasso_B_opt);


fprintf('Lasso Result:\n');
fprintf('y = %.4f * x1 + %.4f * x2 + %.4f * x3 + %.4f\n', ...
lasso_B_opt(1), lasso_B_opt(2), lasso_B_opt(3), lasso_int);


y_pred_lasso = X * lasso_B_opt + lasso_int;
rmse_lasso = sqrt(mean((y - y_pred_lasso).^2));
disp(['Lasso RMSE:', num2str(rmse_lasso)]);


figure;
scatter3(gdp_c, edu_c, users_c, 50, crime, 'filled');
xlabel('GDP per Capita');
ylabel('Years of Education per Capita');
zlabel('Internet Users');
title('Lasso Regression');
colorbar;

```