

Problem Chosen

F

**2025
MCM/ICM
Summary Sheet**

Team Control Number

2500434

Hello Summary!

INDEX

1	Introduction	3
1.1	Background	3
1.2	Restatement of the problem	3
1.3	Analysis of problems	3
1.3.1		3
1.3.2		4
1.3.3		4
2	Symbol and Assumptions	4
2.1	Symbol Description	4
2.2	Assumption	5
3	A Data-Driven Model for Global Cybercrime Hotspot Mapping	5
3.1	Cybercrime distribution across the globe	5
3.2	High-prevalence regions	6
3.3	Other Cybercrime Incidents	7
3.4	Pattern Discovery	8
4	Policy Identification and Analysis	9
4.1	Selection of Representative Centroid Countries	10
4.2	Categorizing Policies Based on Effectiveness	11
4.3	Analysis of Effective Policies	12
4.4	Other cybercrime indicators	13
4.5	A Poisson Regression Model with Time Lag effect	13
4.6	Conclusions on the impact of laws and policies on Internet crime rates	17
5	Correlation Between National Demographics and Cybercrime Distribution	17
5.1	Data Preprocessing	17
5.1.1	Data Integration and Cleaning	17
5.2	Data Processing and Analysis	18
5.2.1	Data Consolidation	18
5.2.2	Logarithmic Transformation	18
5.2.3	Spearman Correlation Analysis	18
5.3	Data Visualization	18
5.3.1	Scatter Plots	19
5.3.2	Time Series Plots	19
5.3.3	Distribution Plots	19
6	Future expected data	21
7	Strength & Weakness	21
7.1	As for the 1 st Question	21
7.2	As for the 2 nd Question	21

1 Introduction

1.1 Background

In the digital age, the speed and scale of global connectivity have reached unprecedented heights. However, with technological advancements, **cybercrime** has also become increasingly complex and diverse. These crimes pose significant threats and challenges to personal privacy, corporate assets, national security, and social stability. The transnational and covert nature of cybercrime makes it difficult to address effectively. Attackers often exploit legal differences and technical vulnerabilities across countries to evade accountability. Additionally, many businesses and institutions, concerned about their reputation and commercial interests, often choose **not to publicly report** cyberattacks, opting instead to pay ransoms or handle incidents privately. This further exacerbates the hidden nature of cybercrime. Developed countries, with their highly digitized economic and social structures, are often prime targets for cybercrime, while developing countries face their own unique challenges.

To address this global challenge, countries have introduced national cybersecurity **policies** aimed at enhancing their defensive capabilities through legal, technical, and organizational cooperation. The **effectiveness** of these policies varies significantly across nations, and these differences may be closely related to factors such as policy design, enforcement, technological infrastructure, education levels, economic development, and internet penetration rates. In this context, understanding which factors make certain countries' cybersecurity policies more effective has become a critical issue. By analyzing the global distribution of cybercrime, national cybersecurity policies, and their outcomes, we can identify which policies and laws are particularly effective in preventing, prosecuting, and mitigating cybercrime. This data-driven analysis not only helps countries improve their cybersecurity policies but also provides valuable insights for global cybersecurity cooperation.

1.2 Restatement of the problem

We are required to identify patterns that can inform the data-driven development and refinement of national cybersecurity policies and laws, which focused on those that have demonstrated effectiveness. Our goal is to develop a theory explaining what constitutes a strong national cybersecurity policy and then support it with a data-driven analysis.

Several key aspects are listed below:

- Patterns in cybercrime distributed worldwide.
- Assessment to effectiveness of National Cybersecurity Policies.
- Correlation between national demographics and our distribution analysis.

1.3 Analysis of problems

We have divided each problem into several different steps:

1.3.1 Cybercrime distribution across the world

- Process the JSON files published on VCDB .

- Use GCI as the primary indicator to assess countries with disproportionately high levels of cybercrime.
- Group countries based on GCI Tier measure, and visually represent the distribution of cybercrime in each group using heatmaps.

1.3.2 Effective policy or law analytical approach

- Identify a representative country from each of the T1 to T5 national clusters and collect cybersecurity-related policies enacted by these countries.
- Plot time-series line graphs depicting the trends of cybercrime over time and analyze which policies have been effective in curbing criminal activities.
- Conduct a targeted analysis of specific policy models' effectiveness in certain countries, focusing on particular indicators.
- Integrate temporal factors and national contexts to comprehensively evaluate the impact of policies.

1.3.3 National demographics correlation

- For each country, preprocess the annual proportion of internet users relative to the total population.
- Compute the correlation between the proportion of internet users and the number of cybercrime incidents.
- Generate visualizations such as scatter plots, time series graphs, and distribution maps to illustrate the relationship between internet access and cybercrime.

2 Symbol and Assumptions

2.1 Symbol Description

Symbol	Description
D_i	Cybercrime distribution in each country.
P_i	Population of each country.
y	Transformed value using $y = \log(1 + x)$.
$\mathbb{E}[Crime_t]$	Poisson regression model predicted value.
β_0	Intercept term.
$\beta_1, \beta_2, \beta_3$	Regression coefficients.
$Bill_{t-k}$	Lagged variable.
K	Greatest lagged number.
ρ	Spearman rank correlation coefficient
d_i	Difference between the ranks of each pair of observations
n	Number of observations

Abbreviation	Full Form
GCI	Global Cybersecurity Index[1]
VERIS	the Vocabulary for Event Recording and Incident Sharing[2]
VCDB	the VERIS Community Database[3]
GDP	Gross Domestic Product

2.2 Assumption

- Countries with a population below 5% are excluded from the consideration of cybercrime distribution because a small change of number could bring a significant difference to statistical analysis.
- In the statistical analysis of the global distribution of cybercrime, factors such as national population growth, internet access, wealth levels, and education levels are assumed to have no significant impact on the quantitative distribution of cybercrime incidents. This study hypothesizes that the distribution of cybercrime can be more intuitively understood by focusing solely on the number of incidents, independent of these socio-economic variables. This assumption is based on all available data recorded since the inception of cybercrime statistics, aiming to isolate the distribution patterns of cybercrime from other potential influencing factors.
- The impact of newly enacted laws or policies on cybercrime is not instantaneous; there is a time lag before their effects become evident.

3 A Data-Driven Model for Global Cybercrime Hotspot Mapping

Despite the continuous evolution of national cybercrime since the inception of data collection, along with changes in policies, legal frameworks, and population demographics, we can create a global cybercrime hotspot map by leveraging crime data recorded by VERIS over the years. This not only facilitates the analysis of cybercrime volumes by country but also allows for fitting the data against policy and population variables to assess their influence on cybercrime trends.

3.1 Cybercrime distribution across the globe

We made use of a world-wide map to represent all cybercrime occurred around the world. In the map, the color filled in each country represents the total number of cybercrime incidents recorded since the beginning of the statistics. The color gradient, ranging from dark blue to dark red, corresponds to eight severity levels (1 to 8). Countries marked in blue indicate a low frequency of cybercrime incidents, while those marked in red represent a high density of such incidents. For instance, the United States, where the VERIS concept was first proposed, has the highest number of recorded incidents (7,236), whereas many other countries have only 1 or 2 recorded incidents. To address this significant disparity in data distribution, we applied a logarithmic transformation to the data using the formula

$$y = \log(1 + x)$$

where x here represents D_i . This transformation was implemented using the function

$$np.log1p()$$

in Python to ensure computational precision and stability, particularly for small values. The final results are visualized in Figure 1.

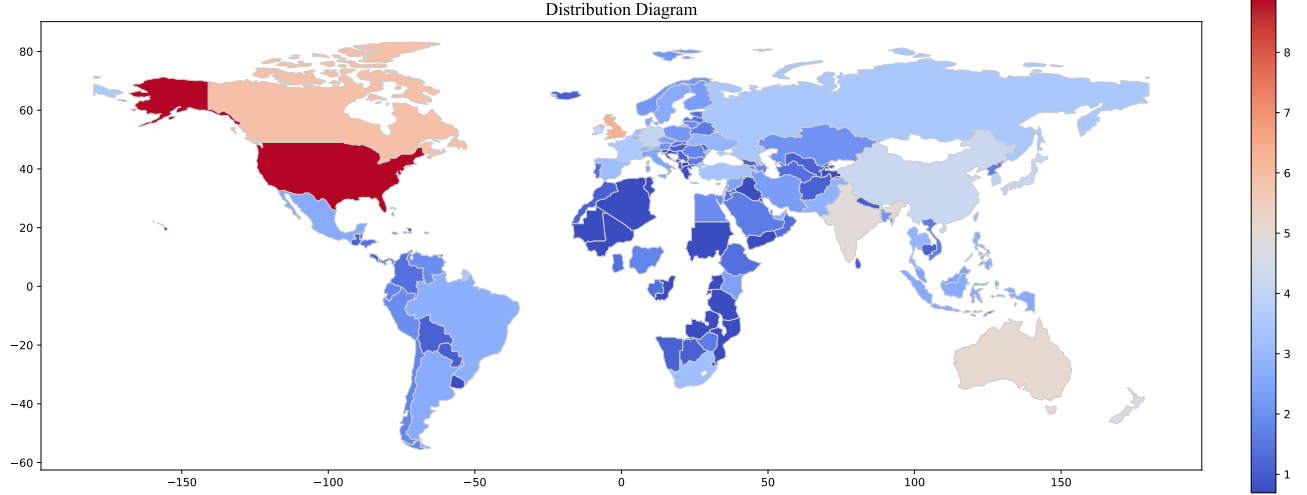


Figure 1: Crime distribution

3.2 High-prevalence regions

We obtained population data P_i for various countries over recent decades from the World Bank Group's website[4]. Simultaneously, we processed data from the VCDB to tabulate the annual number of cybercrime incidents D_i for each country from 2000 to 2025. However, due to discrepancies in the specific countries reported by the World Bank Group and those listed in the VCDB, we had to exclude certain countries to ensure that only those appearing in both datasets were retained. Ultimately, 109 countries were included in the model. To represent the average number of cybercrime incidents per capita, we calculated the ratio D_i/P_i for each year from 2000 to 2025. Since the resulting values were too small for practical analysis, we scaled them by a factor of 10^8 to express the data as the number of cybercrime incidents per 100 million people, denoted as hmD/P_i :

$$hmD/P_i = \frac{D_i}{P_i} \times 10^8$$

According to the GCI (Global Cybersecurity Index) standards, countries are classified into five tiers, denoted as T1 to T5. We used this classification as the basis for K-means clustering analysis, dividing the 109 countries into five groups based on the percentiles published on the GCI website: the top 10%, the next 20%, the following 25%, the subsequent 25%, and the bottom 20%. For each group, the annual average of hmD/P_i (the number of cybercrime incidents per 100 million people) was calculated. To visualize the results, we constructed a 3D clustering heatmap of cybercrime trends, where the x-axis represents the five tiers (T1 to T5), the y-axis represents the time span from 2000 to 2025, and the z-axis represents the average hmD/P_i values. This visualization is presented in Figure2.

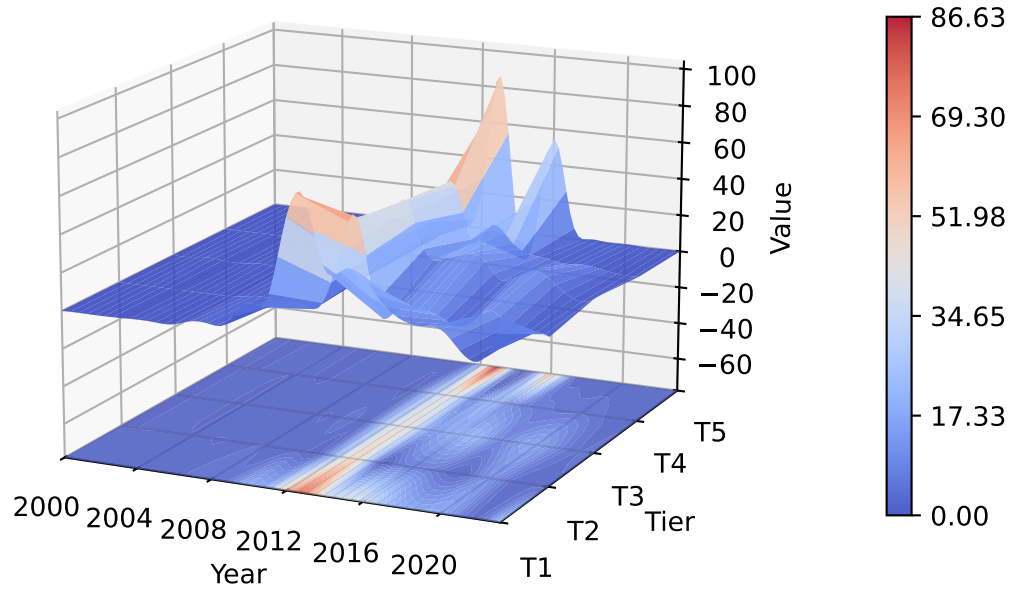


Figure 2: 3D with Spaced Projection

3.3 Other Cybercrime Incidents

Using additional data obtained from the VCDB, we constructed heatmaps on a global scale based on the number of successful cybercrimes, thwarted cybercrimes, and reported cybercrimes, respectively. Due to the disproportionately high volume of data from the United States, we applied the same logarithmic transformation ($y = \log(1 + x)$) as in Figure 1 for consistency, where x represents successful attacks, thwarted attacks, and reported attacks, resulting in the three sub-figures presented in Figure 3.

In sub-figure (a), the number of successful attacks closely aligns with the total number of attacks in most countries. For instance, the United States recorded 7,189 successful attacks out of 7,236 total attacks, yielding a success rate of $\frac{7189}{7236} \approx 99.35\%$. Similarly, the United Kingdom reported 569 successful attacks out of 574 total attacks, with a success rate of $\frac{569}{574} \approx 99.13\%$.

In contrast, countries with lower attack volumes did not show significant differences between the total number of attacks and the number of successful attacks, indicating that almost every attempted attack was successful.

In sub-figure (b), only the United States and Canada reported thwarted attack cases, with 6 and 2 instances, respectively.

In sub-figure (c), the number of successfully reported attacks and the number of countries involved were significantly higher than in sub-figure (b). This suggests that while many attacks were successful, a portion of them were detected and reported.

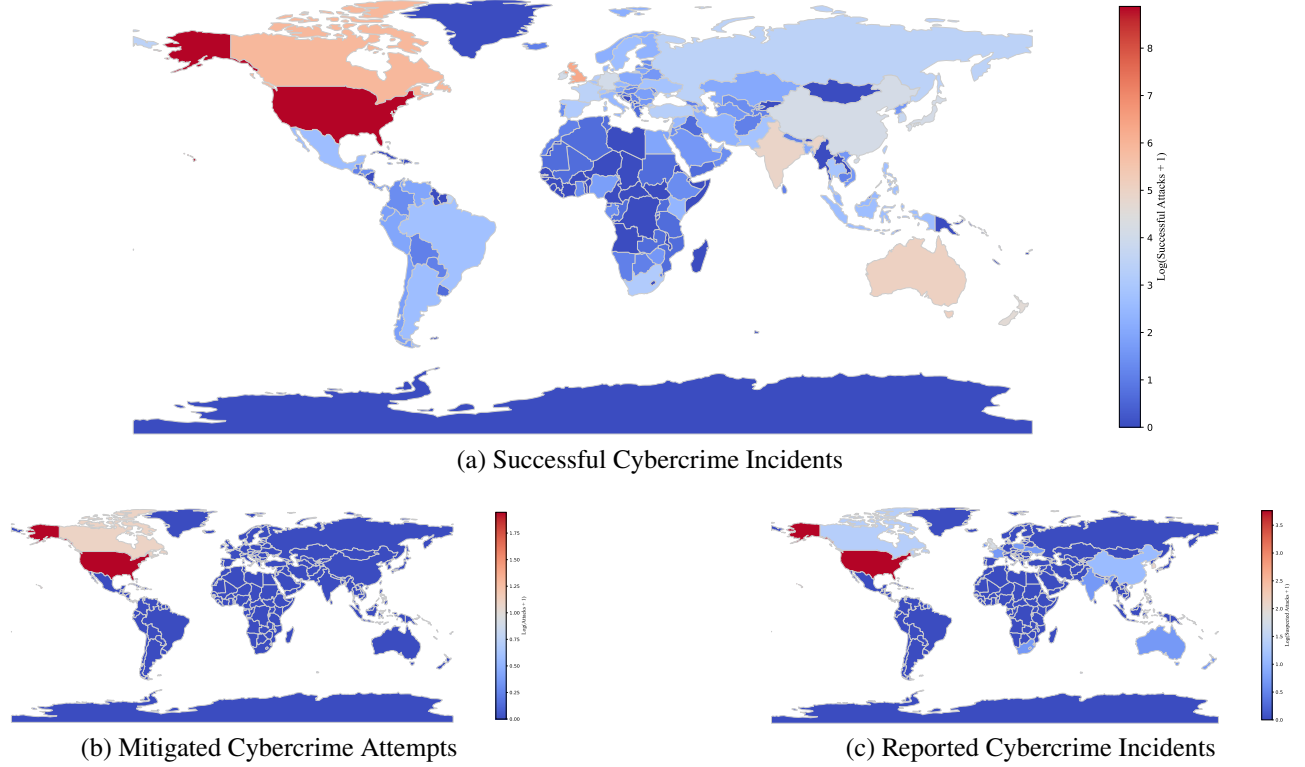


Figure 3: Other Cybercrime Incidents

3.4 Pattern Discovery

As illustrated in Figure 3, the majority of countries worldwide lack adequate defensive capabilities against cyberattacks. Among the disclosed incidents, only the United States and Canada have recorded successful defense cases, with the United States reporting 6 instances and Canada reporting 2. This disparity can be attributed to the early development and technological maturity of the United States' cybersecurity infrastructure. The advanced technologies employed by both the U.S. government and private enterprises have enabled a certain level of resilience against some cyberattacks. Additionally, the disproportionately high volume of cybercrime in the United States plays a significant role. The extensive record of cyber incidents increases the likelihood of successful defenses, as it provides more opportunities for defensive mechanisms to be tested and refined. In contrast, other countries lack both the advanced defensive technologies and the high volume of cybercrime records that the United States possesses. Consequently, no successful defense cases have been reported for these countries in the VCDB .

Cybercrime exhibits distinct patterns both across countries categorized by the GCI (Global Cybersecurity Index) and over time. To analyze these trends, we calculated the proportion of cybercrime incidents and visualized the results in Figure 2. This figure illustrates the annual average hmD/P_i for each GCI tier (T1 to T5) from 2000 to 2025, providing a comprehensive overview of the spatial and temporal distribution of cybercrime.

- **2012 to 2016** Around 2012, both T5 and T1 countries exhibited the highest number of cybercrime incidents. Specifically, the hmD/P_i index for T5 countries reached 86.63, marking the highest average number of cybercrime incidents among all tiers over the 25-year period. In contrast, T2,

T3, and T4 countries reported significantly fewer incidents compared to T1 and T5 during this timeframe. The high cybercrime rates in T5 countries can be attributed to their poor performance across all five GCI assessment metrics: Legal, Technical, Organizational, Capacity Development, and Cooperation. These deficiencies likely contributed to their low GCI rankings and inadequate cybersecurity infrastructure, which in turn made them more vulnerable to cybercrime. On the other hand, T1 countries, despite their high GCI rankings, experienced a surge in cybercrime incidents around 2012. This suggests that even these nations were not fully prepared for the rapid evolution of cyber threats during this period. It is important to note that the fifth edition of the GCI assessment data was only released in 2024, meaning that the high rankings of T1 countries in 2024 do not necessarily reflect their cybersecurity capabilities in 2012.

- **2016 to 2020** During the period from 2016 to 2020, the number of cybercrime incidents declined across all tiers, from T1 to T5. Notably, T1 countries experienced a sharp decrease in cybercrime rates. This trend can be attributed to several factors. First, T1 countries, with their advanced technological infrastructure and robust cybersecurity policies, were better equipped to adapt to emerging threats. During this period, many T1 nations implemented comprehensive cybersecurity strategies, including stricter regulations, enhanced public-private partnerships, and increased investment in cybersecurity education and training. Additionally, international cooperation played a significant role in mitigating cross-border cyber threats. For example, initiatives such as information sharing through organizations like VERIS contributed to a more coordinated global response to cybercrime. However, despite the overall decline, T5 countries remained the most affected by cybercrime, likely due to their low performance in the GCI assessment metrics. These deficiencies continued to hinder their ability to build effective cybersecurity defenses, leaving them more vulnerable to cyber threats.
- **2020 till Present** Around 2020, a remarkable shift has occurred in the distribution of cybercrime incidents across the GCI tiers. Notably, the number of reported cybercrime incidents in T5 countries has dropped to almost negligible levels, a phenomenon that warrants further investigation. Similarly, T4 countries have maintained cybercrime rates comparable to those of T5. In contrast, T1, T2, and T3 countries continue to report a portion of cybercrime incidents, although these numbers have been steadily decreasing year by year. This trend may reflect the cumulative impact of sustained investments in cybersecurity infrastructure, policy improvements, and international collaboration among higher-tier countries. However, the sharp decline in T5 countries' cybercrime records remains puzzling and could be attributed to underreporting. Further research is needed to explore these potential explanations.

4 Policy Identification and Analysis

To identify the effectiveness of national cybersecurity policies, it is essential to analyze the correlation between the implementation of these policies and the subsequent trends in cybercrime. By examining the distribution of cybercrimes and comparing it with the timing and content of various national policies, we can discern patterns that highlight which measures are particularly effective or ineffective. This analysis will focus on key metrics such as the reduction in cybercrime incidents, the success rate of prosecutions, and the overall resilience of national cybersecurity infrastructures. Through this

data-driven approach, we aim to provide actionable insights for the development and refinement of cybersecurity policies.

4.1 Selection of Representative Centroid Countries

Having constructed a clustering model to categorize countries into five clusters (T1 to T5) based on GCI and other relevant metrics, we now proceed to analyze the effectiveness of cybersecurity policies within each cluster. To ensure a representative and data-driven analysis, we will select one central country from each cluster that meets the following criteria:

- **Representativeness:** The selected country should typify the overall characteristics of its cluster, reflecting the general trends and patterns observed within that group.
- **Data Availability:** The country must have sufficient historical data on cybersecurity policies and legislation enacted over the past two decades, allowing for a comprehensive analysis of policy impacts.

By focusing on these representative countries, we aim to draw meaningful insights into the effectiveness of various cybersecurity policies and laws, which can then be generalized to other countries within the same cluster.

Implementation Steps

To identify the representative country for each cluster, we first calculate the average GCI for each cluster. The average GCI, denoted as \overline{GCI} , is computed as follows:

$$\overline{GCI} = \frac{\sum_{i=1}^n GCI_i}{n}$$

where n is the number of countries in the cluster. Next, we compute the absolute deviation of each country's GCI from the cluster average:

$$\{a | a = |GCI_i - \overline{GCI}|\}$$

where a is the approach to the average \overline{GCI} . The country with the smallest deviation is considered the most representative of its cluster. After this initial selection, we further filter out countries with insufficient or incomplete legal and policy documentation.

Through this process, we identify the following representative countries for each cluster with their references:

T1	T2	T3	T4	T5
United States	Japan	China	Costa Rica	Namibia

Visualizing Policy Impact Over Time

With the selected representative countries, we proceed to visualize the impact of cybersecurity policies on cybercrime trends. For each country, we plot a line graph where the x-axis represents time (from 2000 to 2023) and the y-axis represents the annual number of cybercrime incidents. To highlight the influence of policy implementations, we mark the data points corresponding to years in which cybersecurity policies or laws were enacted with an orange color. This visualization is presented in Figure 4.

This allows us to preliminarily assess the effectiveness of the policies. Specifically:

- A downward trend in the line graph following the implementation of a policy (marked in orange) suggests that the policy may have been effective in reducing cybercrime.
- An upward or unchanged trend, on the other hand, may indicate that the policy was ineffective or had unintended consequences.

This initial analysis provides a broad overview of the impact of various policies and helps identify patterns that warrant further investigation. It also serves as a foundation for more detailed analysis of specific policies, guiding future research directions.

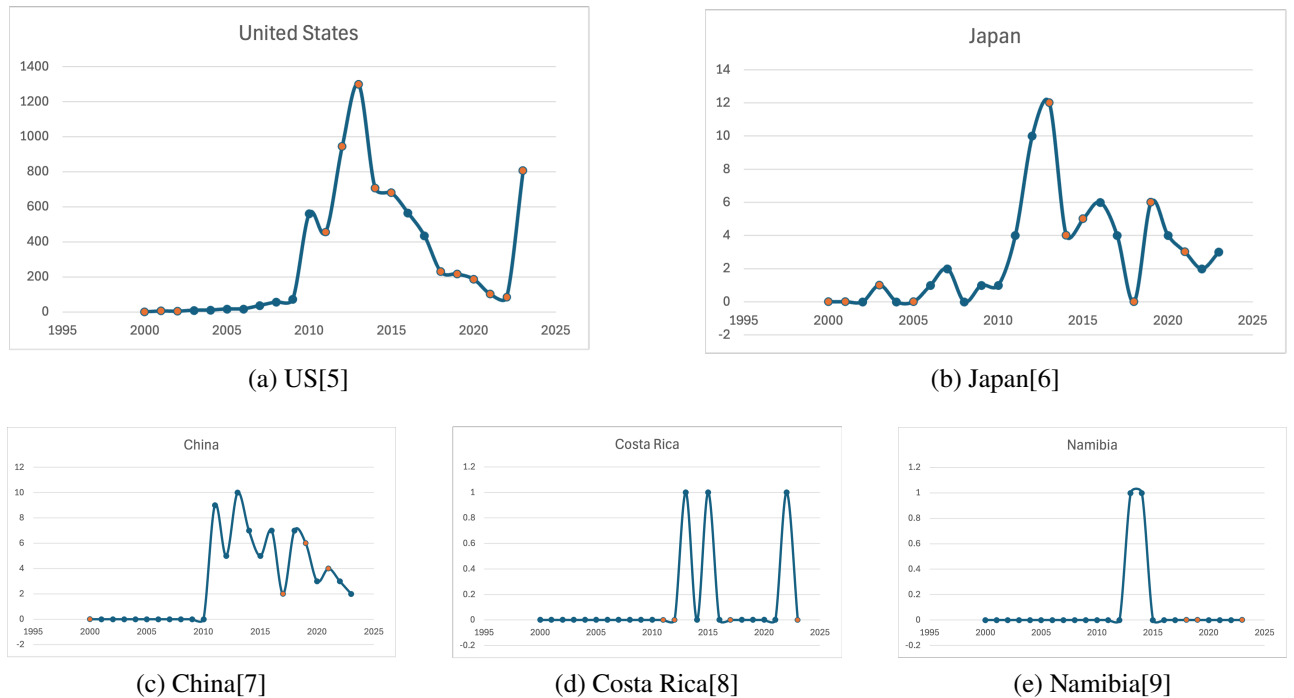


Figure 4: Representative countries

4.2 Categorizing Policies Based on Effectiveness

From the line graphs, we categorize the enacted policies into three sets based on the average cybercrime metrics in the years following their implementation compared to the year of enactment. Drawing

inspiration from fuzzy set theory, we assign a value of 1 to policies that are effective, -1 to those with the opposite effect, and values between -1 and 1 to policies with varying degrees of impact. Specifically:

- **Effective Policies:** These are policies where the average cybercrime metrics in the years following enactment are lower than those in the year of enactment. Examples include:
 - National Institute of Standards and Technology (NIST) Cybersecurity Framework (2014)
 - National Cyber Security Centre (NCSC) Establishment (2016)
 - Investigatory Powers Act (2016)
 - Cybersecurity Strategy (2013)
 - Telecommunications Business Act Amendments (2019)
 - Cryptography Law of the People’s Republic of China (2019)
- **Counterproductive Policies:** These are policies where the average cybercrime metrics in the years following enactment are higher than those in the year of enactment. Examples include:
 - Cyber Incident Reporting for Critical Infrastructure Act (CIRCI) (2022)
 - Quantum Computing Cybersecurity Preparedness Act (2022)
 - Strengthening American Cybersecurity Act (2022)
 - CHIPS and Science Act (2022)
 - Cybersecurity Strategy (2018)
 - Cybersecurity Law of the People’s Republic of China (2017)
- **Neutral or Mixed-Impact Policies:** These are policies where the average cybercrime metrics show no significant change, or the impact is ambiguous.

This categorization provides a structured framework for analyzing the effectiveness of cybersecurity policies and serves as a basis for further investigation into the factors that contribute to their success or failure.

4.3 Analysis of Effective Policies

The effective policies, such as the **National Institute of Standards and Technology (NIST) Cybersecurity Framework (2014)**, **National Cyber Security Centre (NCSC) Establishment 2016**, **Investigatory Powers Act 2016**, and **Cryptography Law of the People’s Republic of China (2019)**, share common characteristics in governance, international collaboration, technical regulation, and responsiveness to emerging threats. Through these mechanisms, they reduce opportunities for crime, increase the cost of cybercrime, and create a deterrent effect, ultimately leading to a significant decline in cybercrime incidents.

4.4 Other cybercrime indicators

For certain specific metrics of cybercrime, such as the proportion of successful crimes and the proportion of reported crimes, we focus on countries with more comprehensive data available in VCDB . Given the limited data coverage, we select countries with sufficient data for analysis, such as the United States, the United Kingdom, and Canada. These countries provide a robust dataset for examining trends in cybercrime success rates and reporting rates, allowing us to draw meaningful insights into the effectiveness of cybersecurity policies and practices in these regions.

Impact on Cyberattack Success Rates

The line graphs depicting the crime success rate over time for the United States and Canada are shown in Figure 5. The figure illustrates that the establishment of the National Cyber Security Centre (NCSC) Establishment 2016 has been particularly effective in reducing the success rate of cybercrimes. In contrast, other policies enacted during the same period do not appear to have had a significant impact on curbing the success rate of cybercrimes. This highlights the importance of centralized, technically-focused initiatives in mitigating the effectiveness of cyberattacks.

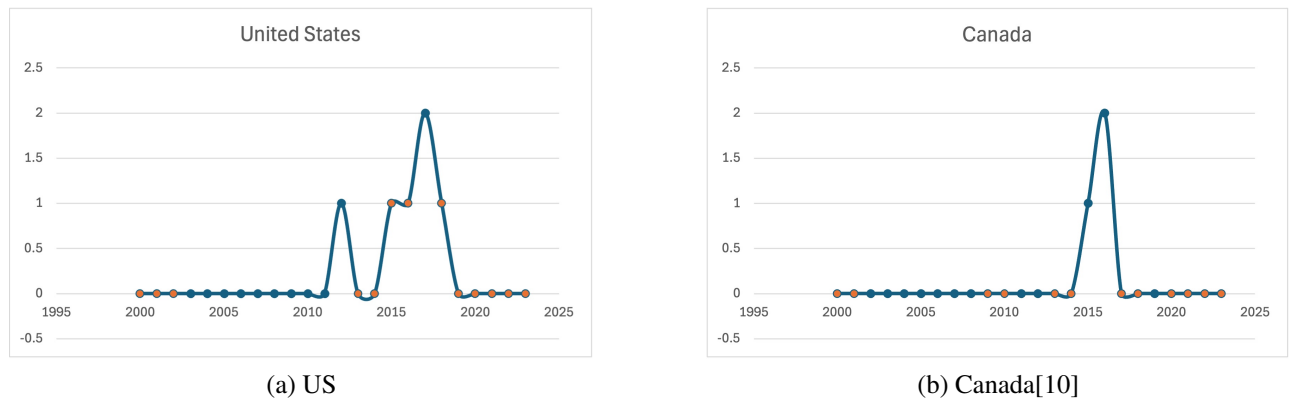


Figure 5: Mitigated Attempts

Analysis of Cybercrime Reporting Rate

In terms of cybercrime reporting rates, the available data for analysis is limited. Taking the United States and the United Kingdom as examples, we plot the reporting rate over time in Figure 6. From the graph alone, it appears that the **Presidential Policy Directive 21 (PPD-21) (2013)** in the United States and the implementation of **The General Data Protection Regulation (GDPR) (2018)** in the United Kingdom have had a positive impact on increasing reporting rates. However, due to the limited amount of data, there is significant noise in the results. In reality, the majority of laws and policies do not seem to have a substantial effect on improving reporting rates.

4.5 A Poisson Regression Model with Time Lag effect

From the above data, we can intuitively find the benefits of legal policies on cybercrime governance in different countries from 2000 to 2023. Based on this, we need to consider the timeliness of the law,

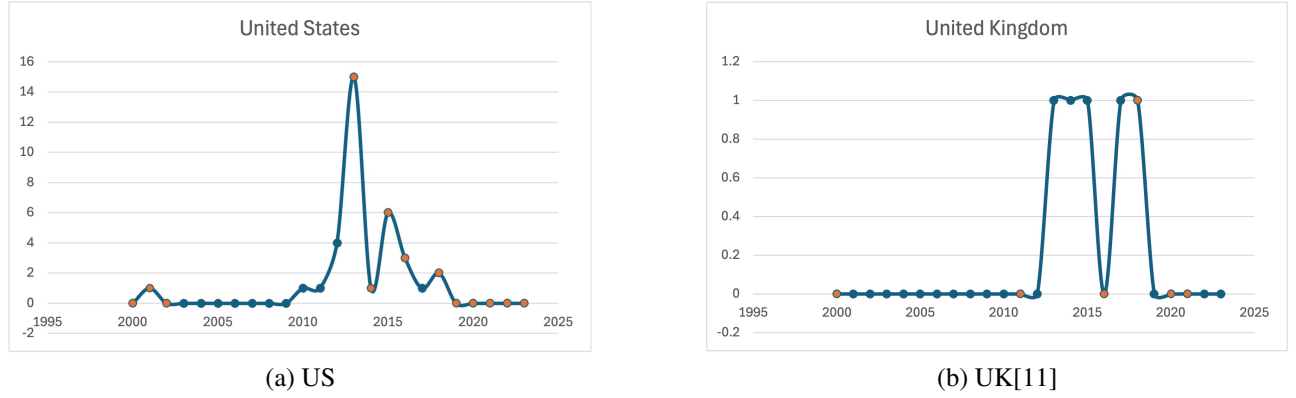


Figure 6: Reported Attempts

that is, a certain law may have little benefit for the cybercrime control in the current year, but it will be effective for the network control in the next few years. Therefore, we do this by using a Poisson regression model and introducing a time lag effect. The model is described as:

$$\log(\mathbb{E}[Crime_t]) = \beta_0 + \sum_{k=1}^K \beta_k Bill_{t-k}$$

where

- \mathbb{E} is the expected crime number in the year t ,
- β_0 is the natural logarithm of the baseline crime rate when all explanatory variables are zero,
- $\beta_1, \beta_2, \beta_3$ represent the marginal impact of bills on the crime rate (on a logarithmic scale) in the first, second, and third year after passage respectively,
- $Bill_{t-k}$ is the number of bills passed in year $t - k$ ($k = 1, 2, 3$), and
- K is the greatest lagged number.

The reason why the linear regression model is not used is that the obtained R^2 value is very low, indicating that the relationship between the time of bill issuance and the crime rate is not obvious. The time lag effect is to get the effect that a bill issued in a given year is likely to have in a lagged number of years.

Because the Poisson regression analysis also needs to be based on a certain amount of data to ensure the accuracy of the conclusion, we choose a few countries that have suffered a large number of cyberattacks for analysis. Included: United States, United Kingdom, Canada, Japan, Costa Rica as Figure 7 shows.

Based on these, conclusions are as follows:

Canada

The model fits very well, indicating that the bill had a significant impact on cybercrime 3 years after its passage. Canada's crime data are less volatile and may reflect effective enforcement of its

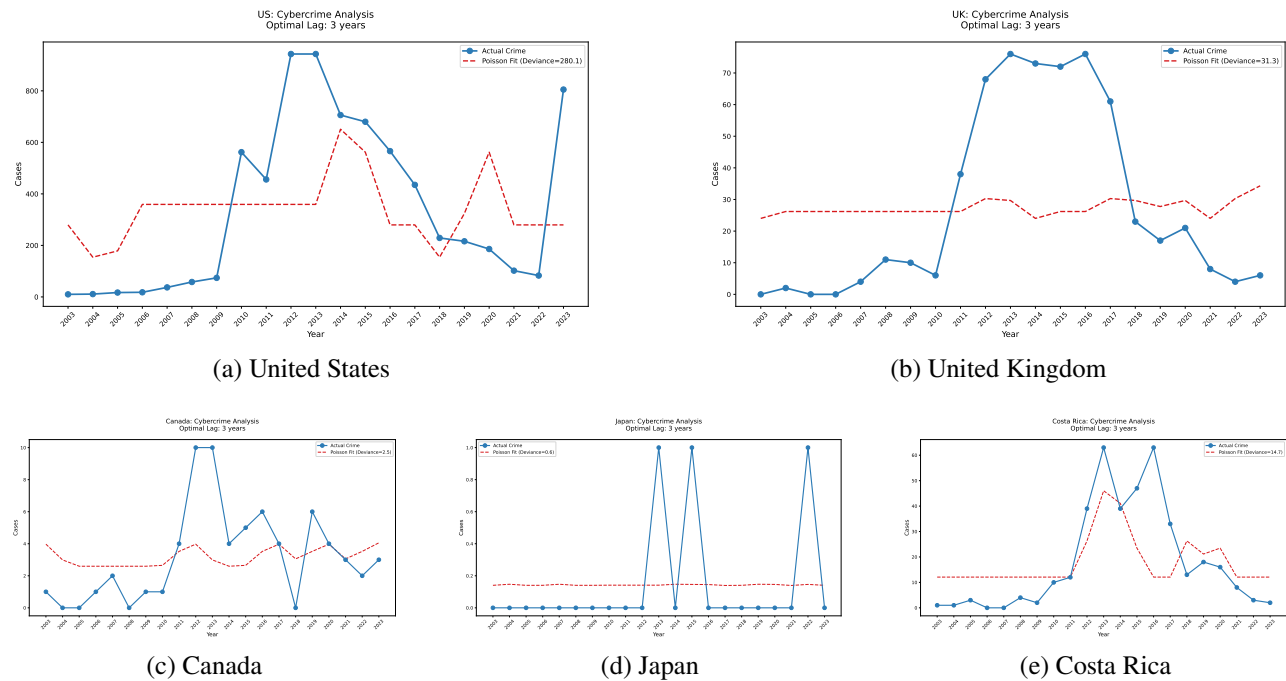


Figure 7: the Poisson regression

cybersecurity policies or lower crime rates per capita. Low bias values indicate a clear relationship between the number of bills and crime rates.

- Best lag period: 3 years.
- Poisson bias: 2.5 (very low). Actual crime trends: low number of crimes (peak around 10).

Costa Rica

The model is a fair fit, with higher bias than Canada but lower bias than the United States. It may indicate that the long-term effects of the Act are present, but the data fluctuates greatly or is interfered by other factors (e.g., economic conditions, technological developments).

- Best lag period: 3 years.
- Poisson bias: 14.7 (medium).
- Actual crime trends: moderate number of crimes (peak around 60).

Japan

The bias value is extremely low, but the actual crime number is almost zero, possibly reflecting incomplete data recording or indeed rare cybercrime in Japan. Although the model fits well, the significance of policy analysis is limited, and the accuracy of data needs to be verified.

- Best lag period: 3 years.

- Poisson bias: 0.6 (very low).
- Actual crime trends: very low number of crimes (peak around 1).

United Kingdom

UK has a poor model fit, indicating a weak relationship between bills and crime rates.

- Best lag period: 3 years.
- Poisson bias: 31.3 (high).
- Actual crime trends: Numbers not explicitly shown.

Possible reasons include:

- Crime data influenced by multiple factors (e.g., transnational cyberattacks).
- Delayed or limited implementation of the bill.

United States

US is a very poor fit, indicating that the current model does not explain the relationship between bills and crime rates.

- Best lag period: 3 years.
- Poisson bias: 280.1 (very high).
- Actual crime trends: High number of crimes (e.g. 562 in 2010).

The main reasons may include:

- Outlier impact: spike in 2010.
- Data complexity: US cybercrime may involve more dynamic factors (e.g., technological advances, scale of hacking).
- Policy limitations: Increasing the number of laws alone will not necessarily curb high crime rates. A combination of law enforcement and technology is needed.

Therefore, we can know that the effect of each law on the average of the above countries can affect the Internet crime rate in the next 3 years. Among them, Canada has the highest revenue from cybercrime governance; The results of Japan show that laws and policies are good for the proceeds of cybercrime; Due to the limited amount of data, Costa Rica can only preliminarily determine that the legal policy has a good effect on its cybercrime control. The United Kingdom's laws are less favourable to the proceeds of cybercrime; The United States has the worst legal policies for cybercrime. In many cases, the release of laws does not directly lead to a decrease in crime rates.

4.6 Conclusions on the impact of laws and policies on Internet crime rates

From the above information, we can learn that the impact of legal policies on cybercrime has a significant time lag effect, that is, the time when legal policies really work may be several years in the future. Moreover, the effects of laws and policies vary from country to country. For T1 countries like the United States with a large population, the increase in the number of laws may not directly reduce the crime rate. Even if the laws are issued frequently, it is difficult to curb the growth of the number of cyber crimes, indicating that there may be some problems in the content or implementation of the laws. But for countries such as Canada, the number of laws is consistent but the effect is significant, which indicates that the laws are well-designed and effectively implemented. Outliers in the data, such as the spike in cybercrime in the US in 2010, can skew the true trend to some extent. Finally, long-term policy effects are better than short-term ones. Therefore, it is necessary to comprehensively consider the above factors to control the Internet crime rate.

5 Correlation Between National Demographics and Cybercrime Distribution

The distribution of cybercrime is closely tied to national demographic statistics, with the number of cybercrime incidents showing a positive correlation with several key factors. In this section, we explore the relationship between cybercrime and four primary demographic indicators: the proportion of internet users in a country, the country's GDP, and the proportion of the population with higher education. By analyzing these factors, we aim to uncover patterns and correlations that can provide insights into the drivers of cybercrime and inform the development of more targeted and effective cybersecurity policies.

5.1 Data Preprocessing

To analyze the correlation between national demographics and cybercrime distribution, we first preprocess the relevant data. The demographic indicators—internet user penetration[12], GDP[13] and the proportion of the population with higher education[14] —are obtained from the World Bank's official website. Additionally, we utilize the annual cybercrime incident data for each country, which was previously processed in our earlier analysis. By integrating these datasets, we ensure a comprehensive foundation for examining the relationship between demographic factors and cybercrime trends.

5.1.1 Data Integration and Cleaning

For each dataset, we filter the data to include only the years from 2010 to 2022. After filtering, we handle missing values by allowing a maximum missing value proportion of 20% for each country's data. Missing values within this threshold are filled using linear interpolation. Any data points that remain missing after interpolation are removed to ensure the integrity and reliability of the dataset. This preprocessing step ensures that our analysis is based on a consistent and high-quality dataset.

5.2 Data Processing and Analysis

5.2.1 Data Consolidation

After preprocessing the individual datasets, we integrate the internet user data and the cybercrime incident data. This is achieved by performing an inner join on the two datasets using country codes and years as the matching keys. The inner join ensures that only the countries and years present in both datasets are retained, resulting in a combined dataset where each entry corresponds to a specific country and year with complete data for both internet user penetration and cybercrime incidents. This step is crucial for ensuring the accuracy and consistency of our subsequent analysis.

5.2.2 Logarithmic Transformation

To address the skewness in the distribution of cybercrime incident counts, we apply a logarithmic transformation to the data. As described in Subsection 3.1, we use the *log1p* transformation, which computes the natural logarithm of $1 + x$, where x is the original cybercrime count. This transformation reduces the impact of extreme values and makes the data more symmetric, bringing it closer to a normal distribution. By applying this transformation, we ensure that the data is better suited for statistical analysis and modeling.

5.2.3 Spearman Correlation Analysis

To quantify the relationship between those data and the number of cybercrime incidents, we calculate the Spearman correlation coefficient. This non-parametric measure assesses the strength and direction of the monotonic relationship between two variables. The Spearman correlation coefficient (ρ) and its associated p -value are computed, with the p -value used to determine the statistical significance of the correlation. A p -value smaller than 1×10^{-10} indicates an extremely strong statistical significance. The Spearman correlation coefficient is calculated as follows:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

where:

- ρ is the Spearman rank correlation coefficient,
- d_i is the difference between the ranks of each pair of observations, and
- n is the number of observations.

5.3 Data Visualization

To further explore the relationship between demographic factors and cybercrime, we generate several visualizations. These include scatter plots, time series plots, and distribution plots, all of which are presented on a logarithmic scale due to the logarithmic transformation applied during data preprocessing.

5.3.1 Scatter Plots

We begin by plotting scatter diagrams to visualize the relationship between demographic indicators (such as internet user penetration, higher education enrollment, and GDP) and the logarithmically transformed cybercrime counts. A regression line is added to each scatter plot to highlight the trend. The Spearman correlation coefficient (ρ) and its associated p -value are annotated on the plots to provide statistical context. For example:

- The scatter plot for **internet user penetration** shows a correlation coefficient of $\rho = 0.17$ with a highly significant p -value of 2.33×10^{-12} .
- The scatter plot for **higher education enrollment** reveals a correlation coefficient of $\rho = 0.13$ with a p -value of 2.60×10^{-5} .
- The scatter plot for **GDP** indicates a weaker correlation ($\rho = 0.45$) but with an extremely low p -value of 2.456×10^{-87} , suggesting a statistically significant relationship.

5.3.2 Time Series Plots

Next, we construct time series plots to analyze the temporal trends in cybercrime and GDP. These plots display the annual average of logarithmically transformed cybercrime counts and GDP over time, allowing us to observe how these variables have evolved from 2010 to 2022. The time series plot for GDP reveals a steady increase over the years, while cybercrime counts show fluctuations with a general upward trend. Figure 8.

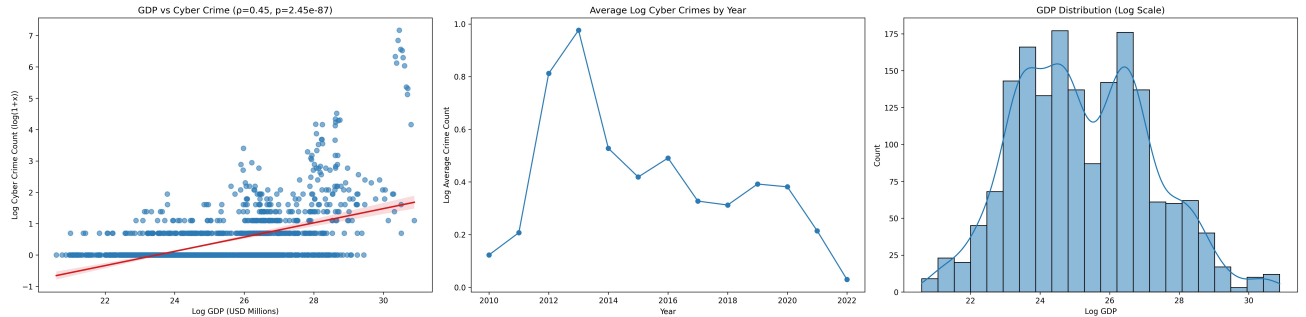


Figure 8: GDP Chart

5.3.3 Distribution Plots

Finally, we generate distribution plots to examine the spread and density of the data. These plots illustrate the distribution of demographic indicators (e.g., internet user penetration, higher education enrollment) and their relationship with cybercrime counts. For example:

- The distribution plot for **internet user penetration** shows a right-skewed distribution, indicating that most countries have relatively low internet penetration rates. Figure 9a.
- The distribution plot for **higher education enrollment** reveals a more uniform distribution, with a peak around 60–80% enrollment rates. Figure 9b.

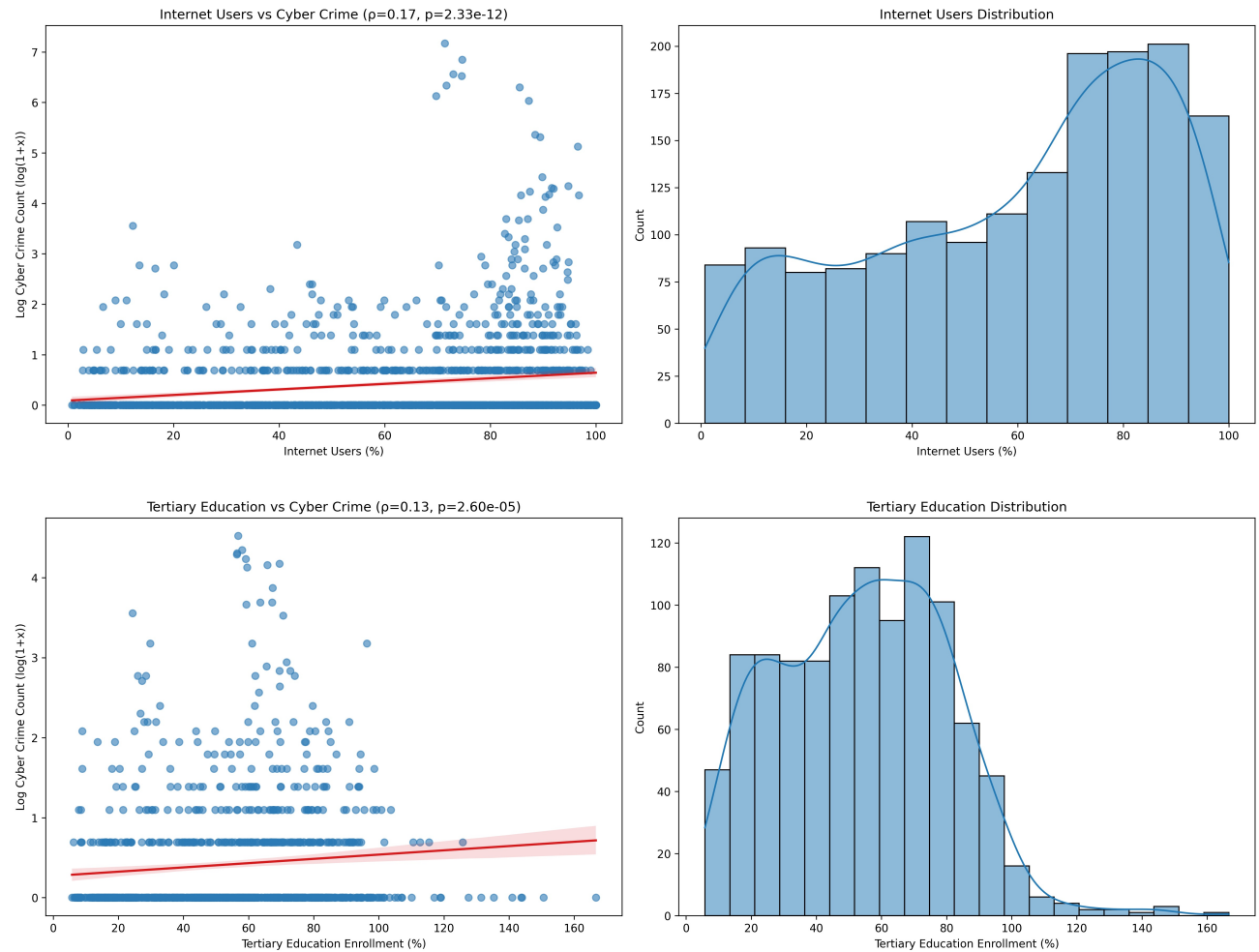


Figure 9: Trends and Correlations Related to Cybercrime

6 Future expected data

7 Strength & Weakness

7.1 As for the 1st Question

By filtering the data and drawing the global heat map of cybercrime distribution, the global distribution of cybercrime can be visualized. Then, by filtering the data according to specific indicators (Confirmed, Near miss, Reported, Prosecuted) and drawing the corresponding global heat map, the distribution of these indicators can be visualized.

Pure data screening and analysis need to be based on a huge amount of data. For these big countries in Europe and the United States, the annual amount of data is very considerable, so the effect is very prominent. But for some small countries in Africa or Asia, the difference in the distribution map of cybercrime is not very obvious because of the insufficient amount of data.

7.2 As for the 2nd Question

1. Through processing the filtered data and drawing a line chart, it can be more intuitive to see which laws and policies have benefits for the national cybercrime rate control; However, the timeliness of legal policies cannot be rationally judged, that is, whether the law has an impact on the future cannot be directly considered.
2. Introduction of Poisson model: Because the number of cybercrimes is a typical non-negative integer data, Poisson regression analysis can avoid the continuity assumption of ordinary linear regression, and can model and explain the nonlinear association between the number of bills and the crime rate and zero inflation data. Furthermore, the regression coefficients can be used for policy evaluation by calculating the incidence ratio *IRR*. Over-discretization of the data can lead to bias in the model, and the model will only work for relationships that are roughly linear, and it will be less useful if the actual relationship is more complex (e.g., if there are interaction effects).
3. Introduction of time hysteresis effect: The introduction of time lag effect can reflect the delay of the policy coming into effect, that is, the release of legal policies does not necessarily produce benefits at the time but will have potential impacts on the future. Then, the optimal lag period was selected by cross validation, and the key time node for the policy to be effective was clarified. But it ignores the immediate effect on the law.

Therefore, it is necessary to combine the data analysis and Poisson & time lag effect model to obtain the income effect of legal policies on the national Internet crime rate.

References

- [1] Global cybersecurity index 2024, 2024. Global Cybersecurity Index 2024.
- [2] The vocabulary for event recording and incident sharing. VERIS Framework.

- [3] The veris community database (vcdb). VERIS Community Database.
- [4] World population. World Bank Total Population Data.
- [5] United states. U.S. Congress National Institute of Standards and Technology (NIST) U.S. Department of Homeland Security (DHS) U.S. Securities and Exchange Commission (SEC) The White House Investigatory Powers Act 2016 (UK) National Cyber Security Centre (UK) Telecommunications (Security) Act 2021 (UK) UK Cybersecurity Requirements for Consumer Products (2024) Timeline Set for UK Cybersecurity (2024).
- [6] Japan. Japan's IT Basic Law (Kantei) Personal Information Protection Commission (PPC) Japan National Information Security Center (NISC) Japan Ministry of Foreign Affairs of Japan - Cybersecurity Japanese Law Translation - Act on the Protection of Personal Information NISC Cybersecurity Strategy 2015 NISC Cybersecurity Strategy 2018 Telecommunications Business Act (Japan) NISC Cybersecurity Strategy 2021.
- [7] China. International Cybercrime (Wikipedia) Cybersecurity Law of the People's Republic of China (Wikipedia) Internet Censorship in China (Wikipedia) China Issues Regulations on Network Data Security Management (IAPP) Cryptography Law (Wikipedia).
- [8] Costa rica. National Cybersecurity Strategy of Costa Rica (2023) Costa Rica - Octopus Project (Council of Europe).
- [9] Namibia. Namibia - Octopus Project (Council of Europe) A Digital Odyssey: The Convergence of Rapid Digitization, Population Dynamics, and Financial Risk in Namibia (Carnegie Endowment) Review of National Cybersecurity Strategy (Namibia).
- [10] Canada. PIPEDA - Privacy Laws in Canada (Office of the Privacy Commissioner of Canada) Access to Information Act (Canada) Cybersecurity (Public Safety Canada) Cybersecurity Strategy (Public Safety Canada) Communications Security Establishment (CSE) Critical Infrastructure (Public Safety Canada) Bill C-11 - Royal Assent (Canada) Canada's Cybersecurity Strategy (National Defence) Cybersecurity (National Defence Canada) Innovation, Science and Economic Development Canada - Cybersecurity.
- [11] United kingdom. Investigatory Powers Act (UK Government) UK Cyber Security Strategy: Protecting and Promoting the Nation (UK Government) Investigatory Powers (UK Government) Guide to the General Data Protection Regulation (GDPR) (ICO) Network and Information Systems Regulations (UK Government) Telecommunications Security (UK Government) Online Safety Bill (UK Government).
- [12] Internet Users as Percentage of Population (World Bank).
- [13] Gross Domestic Product (GDP) (World Bank).
- [14] Tertiary Education Enrollment Ratio (World Bank).

Appendix

Report on Use of AI

1. OpenAI ChatGPT(Jan 25, 2025 version, ChatGPT-4o(Internet Search))

Query1: Please give me more US cybersecurity laws and regulations from 2000 to 2023.

Output: Here is a list of significant U.S. cybersecurity laws and regulations from 2000 to 2023:

1. USA PATRIOT Act (2001)
 - Expanded law enforcement's ability to surveil and intercept communications for counterterrorism purposes, including provisions affecting cybersecurity.
2. Federal Information Security Management Act (FISMA) (2002)
 - Required federal agencies to develop, document, and implement programs to secure information and systems.
3. Cybersecurity Research and Development Act (2002)
 - Funded research and development for cybersecurity tools and technologies through the National Science Foundation (NSF) and National Institute of Standards and Technology (NIST).
4. E-Government Act (2002)
 - Established guidelines for improving electronic government services, including the management of information security.
5. Homeland Security Act (2002)
 - Created the Department of Homeland Security (DHS) and included provisions for critical infrastructure protection against cyber threats.
6. Presidential Policy Directive 21 (PPD-21) (2013)
 - Aimed to strengthen critical infrastructure security and resilience, emphasizing information sharing across sectors.
7. National Institute of Standards and Technology (NIST) Cybersecurity Framework (2014)
 - Voluntary guidelines for private and public organizations to manage and reduce cybersecurity risks.
8. Cybersecurity Information Sharing Act (CISA) (2015)
 - Enabled the sharing of cyber threat information between private entities and the federal government.
9. Federal Cybersecurity Enhancement Act (2015)
 - Mandated the use of intrusion detection and prevention systems (e.g., EINSTEIN) on federal networks.
10. National Defense Authorization Act (NDAA) Provisions (2019)
 - Introduced supply chain security measures and banned federal agencies from using certain Chinese-made telecommunications equipment.
11. CLOUD Act (2018)

- Clarified legal processes for accessing electronic data stored abroad while balancing privacy and law enforcement interests.
- 12. SEC Guidelines on Cybersecurity Disclosure (2018, Updated 2023)
 - Required public companies to disclose cybersecurity risks, incidents, and governance strategies to protect investors.
- 13. Internet of Things (IoT) Cybersecurity Improvement Act (2020)
 - Set baseline security standards for IoT devices used by federal agencies.
- 14. Cyber Incident Reporting for Critical Infrastructure Act (CIRCA) (2022)
 - Required critical infrastructure operators to report significant cyber incidents within 72 hours and ransomware payments within 24 hours.
- 15. Quantum Computing Cybersecurity Preparedness Act (2022)
 - Directed federal agencies to migrate to quantum-resistant cryptographic standards.
- 16. Strengthening American Cybersecurity Act (2022)
 - Consolidated multiple cyber laws, introduced new rules for incident reporting, and enhanced federal cybersecurity protections.
- 17. CHIPS and Science Act (2022)
 - Included cybersecurity measures for semiconductor manufacturing and technology development.
- 18. Telecommunications Supply Chain Security (2023)
 - Expanded restrictions on telecommunications companies deemed national security risks.

2. OpenAI ChatGPT(Jan 25, 2025 version, ChatGPT-4o(Internet Search))

Query2: Also give me the laws and regulations related to network security in Japan from 2000 to the present.

Output: Here are Japan's key cybersecurity-related laws and policies from 2000 to the present,:

1. Basic Act on the Formation of Advanced Information and Telecommunications Network Society (2000)
 - Established a framework for building advanced information and telecommunications society in collaboration with public and private sectors.
2. Act on the Protection of Personal Information (2003)
 - Japan's first law focusing on personal information protection, setting rules for the collection, use, and management of personal data.
3. Comprehensive Strategy on Information Security (2003)
 - Introduced by the Ministry of Economy, Trade, and Industry (METI), emphasizing the concept of a "society assuming accidents," aiming to prevent and mitigate cybersecurity incidents.
4. Basic Plan on Information Security (2006)
 - Proposed a "new public-private partnership model," clarifying roles and cooperation between the government and private sector in ensuring information security.

5. Cybersecurity Strategy (2013)
 - Japan's first cybersecurity strategy, addressing comprehensive challenges in cyberspace and advocating for international collaboration.
6. Basic Act on Cybersecurity (2014)
 - Passed in November 2014, it defined the responsibilities of governmental, private, and academic institutions in promoting cybersecurity and established the Cybersecurity Strategic Headquarters.
7. Cybersecurity Strategy (2015)
 - Highlighted Japan's vision of ensuring "a free, fair, and secure cyberspace" and aligned with international cybersecurity standards.
8. Cybersecurity Strategy (2018)
 - Updated the 2015 strategy by including concrete action plans, enhancing organizational frameworks, and strengthening international cooperation.
9. Telecommunications Business Act Amendments (2019)
 - Introduced measures to ensure the security of telecommunications infrastructure, including vendor restrictions for critical infrastructure.
10. Cybersecurity Strategy (2021)
 - Revised strategy to align with the latest global cybersecurity trends, focusing on emerging technologies such as IoT and 5G, and reinforcing resilience against cyber threats.

3. OpenAI ChatGPT(Jan 25, 2025 version, ChatGPT-4o(Internet Search))

Query3: So do China.

Output: Since 2000, China has developed a comprehensive framework of cybersecurity laws and regulations to address the evolving digital landscape. Here are some key legislations and policies:

1. Decision on Internet Security (2000)
 - In 2000, the Standing Committee of the National People's Congress passed the "Decision on Internet Security," establishing foundational principles for internet regulation and security in China.
2. Cybersecurity Law of the People's Republic of China (2017)
 - Enacted by the Standing Committee of the National People's Congress on November 7, 2016, and implemented on June 1, 2017, this law aims to increase data protection, data localization, and cybersecurity in the interest of national security.
3. Data Security Law of the People's Republic of China (2021)
 - Promulgated on June 10, 2021, and effective from September 1, 2021, this law establishes a data classification framework based on national security principles, imposing data localization requirements and mandating that data transfers to foreign law enforcement or judicial agencies require official approval.
4. Personal Information Protection Law of the People's Republic of China (2021)

- Effective from November 1, 2021, this law is China's first comprehensive legislation on personal data rights, modeled after the European Union's General Data Protection Regulation (GDPR), outlining requirements for data collection, processing, and protection.

5. Regulations on Network Data Security Management (2023)

- Issued in 2023, these regulations implement the Cybersecurity Law, Data Security Law, and Personal Information Protection Law, providing practical guidance on compliance requirements for businesses operating in China.

4. OpenAI ChatGPT(Jan 25, 2025 version, ChatGPT-4o(Internet Search))

Query4: So do Costa Rica.

Output: Costa Rica has developed a comprehensive legal framework to address cybersecurity and data protection since 2000. Key laws and policies include:

1. Law on Protection of the Person Against the Processing of Personal Data (Law No. 8968) (2011)
 - Enacted in 2011, this law regulates the processing of personal data to protect individuals' privacy rights. It establishes guidelines for data collection, storage, and usage, and led to the creation of the Agency for the Protection of Inhabitants' Data (PRODHAB) to oversee compliance.
2. Law on Computer-Related Crimes (Law No. 9048) (2012)
 - Implemented in 2012, this law amends the Criminal Code to define and penalize various cybercrimes, including unauthorized access, data interference, and system interference. It provides a legal basis for prosecuting offenses committed through information and communication technologies.
3. National Cybersecurity Strategy (2017)
 - In 2017, the Ministry of Science, Technology, and Telecommunications (MICITT) introduced the National Cybersecurity Strategy to enhance the country's cybersecurity posture. The strategy outlines objectives such as strengthening legal frameworks, promoting public-private collaboration, and building cybersecurity capabilities.
4. Accession to the Budapest Convention (2017)
 - Costa Rica acceded to the Council of Europe's Convention on Cybercrime, known as the Budapest Convention, in 2017. This accession aligns the country's legal framework with international standards for combating cybercrime and facilitates international cooperation in cybercrime investigations.

5. OpenAI ChatGPT(Jan 25, 2025 version, ChatGPT-4o(Internet Search))

Query5: So do Canada.

Output:

1. Personal Information Protection and Electronic Documents Act (PIPEDA) 2000

- PIPEDA, enacted in 2000, is Canada's primary legislation governing the collection, use, and disclosure of personal information in the private sector. It includes provisions for cybersecurity by requiring organizations to protect personal data and report any breaches that affect individuals' privacy.
2. The Anti-terrorism Act (2001)
 - This act was passed shortly after the 9/11 attacks and enhanced the government's ability to combat terrorism, including cyberterrorism. It provided law enforcement agencies with broader powers to monitor communications and intercept data related to national security threats.
 3. Public Safety Canada's Cybersecurity Strategy (2009)
 - This strategy, launched in 2009, aimed to strengthen Canada's cybersecurity resilience. It involved collaboration between government, industry, and international partners to secure Canada's critical infrastructure from cyber threats.
 4. The National Cyber Security Strategy (2010)
 - Canada's 2010 cybersecurity strategy outlined key priorities such as protecting critical infrastructure, enhancing cybercrime law enforcement, and promoting international cybersecurity collaboration.
 5. The Cybersecurity Act (2014)
 - While primarily a policy rather than formal legislation, this document was a key milestone in Canada's approach to cybersecurity. It focused on the protection of Canadian businesses, citizens, and government data from increasing cyber threats.
 6. The Digital Privacy Act (2015)
 - This act amended PIPEDA to address the rising concerns over cybersecurity, particularly regarding data breach notification requirements. Organizations are now legally required to report breaches that pose a risk of significant harm to individuals.
 7. The Critical Infrastructure Protection Act (2017)
 - This law aims to enhance the cybersecurity protections around Canada's critical infrastructure, including sectors like energy, telecommunications, and transportation, by fostering cooperation between the government and private sector.
 8. Cybersecurity Strategy for Canada (2018)
 - This comprehensive strategy sought to improve cybersecurity across Canada by focusing on strengthening the country's defense against cybercrime, supporting innovation, and protecting critical infrastructure. It also emphasized Canada's role in international cybersecurity collaboration.
 9. The Bill C-11: Digital Charter Implementation Act (2020)
 - This bill proposed a major overhaul of Canadian privacy law, introducing stricter regulations on how personal data is handled and protecting citizens' data online. It includes provisions to enhance cybersecurity protections.
 10. Canada's National Cybersecurity Strategy (2022)
 - In 2022, Canada launched its National Cybersecurity Strategy to address the growing threat of cyberattacks. The strategy outlines objectives such as improving

government security, fostering international partnerships, and advancing cybersecurity awareness and research.

11. The Telecommunications and Cybersecurity Laws (2021)

- This policy introduced new cybersecurity measures to ensure that Canada's telecommunications sector is better protected from foreign threats. It includes provisions for private companies to take on increased cybersecurity responsibilities.

12. The National Cybersecurity Innovation and Research Strategy (2023)

- This strategy focuses on fostering innovation in cybersecurity technologies and enhancing collaboration with private, public, and international stakeholders to better address emerging cyber threats.