Statistical Analysis of Cybersecurity Breaches

Mantha Sai Prasanth
Instructor: Dr. Gahangir Hossain
Data Visualization INFO 5709
Data Visualization – Term Project
05/08/2024

Introduction

Cybersecurity breaches are a major risk to security and integrity, considering the increased digitization of data across various sectors in a world that is much more connected. The patterns and trends of cybersecurity breaches can help build strategies for saving and conserving private information. This project will focus on the statistical analysis of cybersecurity breaches, seeking to expose the trends and insights that may help shape proactive measures to mitigate risks and improve cybersecurity resilience.

Using data visualization techniques, we expose the geographical distribution of breaches, discover the most common types of breaches, and reveal temporal patterns of breach occurrence. Through detailed statistics, we need to be able to express the extent of cybersecurity breaches and uncover the underlying trends and vulnerabilities. In general, the project will be enlightening, giving actual insights to organizations and policymakers to tighten cybersecurity defenses in an evolving and complex threat landscape.

Dataset

The dataset, obtained from Kaggle, explores the scope of cybersecurity incidents during the years 2010 through 2014 in all U.S. states. Comprising 1056 rows, the dataset contains a lot of insightful information related to cybersecurity breaches on a variety of measures and dimensions. Each entry is carefully documented, containing information related to the breach start and end dates, involvement of business associates, breach location, affected entity, summary of incidents, type of breaches, and count of affected individuals. It hence becomes a good dataset to understand the landscape of cybersecurity threats and vulnerabilities during the specified timeframe.

Dataset is collected from Kaggle:
https://www.kaggle.com/datasets/sumanth3112/hello-world

Attributes

Total of 15 attributes are used to describe the Cybersecurity Breach. Below is the list

1. **Breach End**: The date when the cybersecurity breach was officially concluded or resolved.

2. **Breach Start**: The date when the cybersecurity breach was initiated or detected.
3. **Business Associate Involved**: Specifies whether external entities or partners were involved in the cybersecurity breach incident.

4. **Date of Breach**: The exact date when the cybersecurity breach occurred or was first identified.

5. **Date Posted or Updated**: The date when information about the breach was posted or last updated.

6. **Location of Breached Information**: Denotes where the breached information was kept or accessed.

7. **Name of Covered Entity**: Denotes the organization or entity affected by the breach.

8. **State**: The state within the United States where the breach took place or where the affected entity is located.

9. **Summary**: A short description or overview of the cybersecurity breach.

10. **Type of Breach**: Categories breaches according to their nature or characteristics—for example, unauthorized access, leakage of data, and malware attack.

11. **Year**: The year in which the cybersecurity breach took place.

12. **Breach _No**: This is a unique identifier for each incident of a breach in tracking and referencing.

13. **F1**: Variable/column with an unspecified meaning or purpose; additional context needed.

14. People Affected: The total number of persons whose personal or sensitive information may have been disclosed as a result of the breach.

15. **Number of Records**: The total number of records or pieces of data exposed or compromised in a breach.

Tools

- Python
- Tableau

Data Cleaning

Initially,I have identified the number of rows and columns with missing values as and found out that Breach End, Business Associate involved, Date of Breach and Summary are having missing values.

```
import pandas as pd

# Load the dataset from the Excel file using Pandas
data = pd.read_excel('Cyber Security Breaches_Migrated Data.xlsx')

# Explore the shape of the DataFrame
data_shape = data.shape

# Print the shape of the DataFrame
print("Number of rows:", data_shape[0])
print("Number of columns:", data_shape[1])

# Handle missing values by dropping rows with any missing data
data_cleaned = data.dropna()

# Explore the shape of the cleaned DataFrame
cleaned_data_shape = data_cleaned.shape

# Print the shape of the cleaned DataFrame
print("Number of rows after cleaning:", cleaned_data_shape[0])
print("Number of columns after cleaning:", cleaned_data_shape[1])
```

```
Number of rows: 1055
Number of columns: 15
Number of rows after cleaning: 0
Number of columns after cleaning: 15
```

```
# Load the dataset from the Excel file using Pandas
data = pd.read_excel('Cyber Security Breaches_Migrated Data.xlsx')

# Check for missing values in each column and count them
missing_values_count = data.isnull().sum()

# Print the count of missing values for each column
print("Count of missing values in each column:")
print(missing_values_count)
```

```
Count of missing values in each column:
Breach End                          910
Breach Start                          0
Business Associate Involved         784
Date of Breach                      146
Date Posted or Updated                0
Location of Breached Information      0
Name of Covered Entity                0
State                                 0
Summary                             913
Type of Breach                        0
Year                                  0
Breach_No                             0
F1                                    0
Individuals Affected                  0
Number of Records                     0
dtype: int64
```

Exploratory Data Analysis

After finding out the missing values I have explored the shape and remove the missing values.

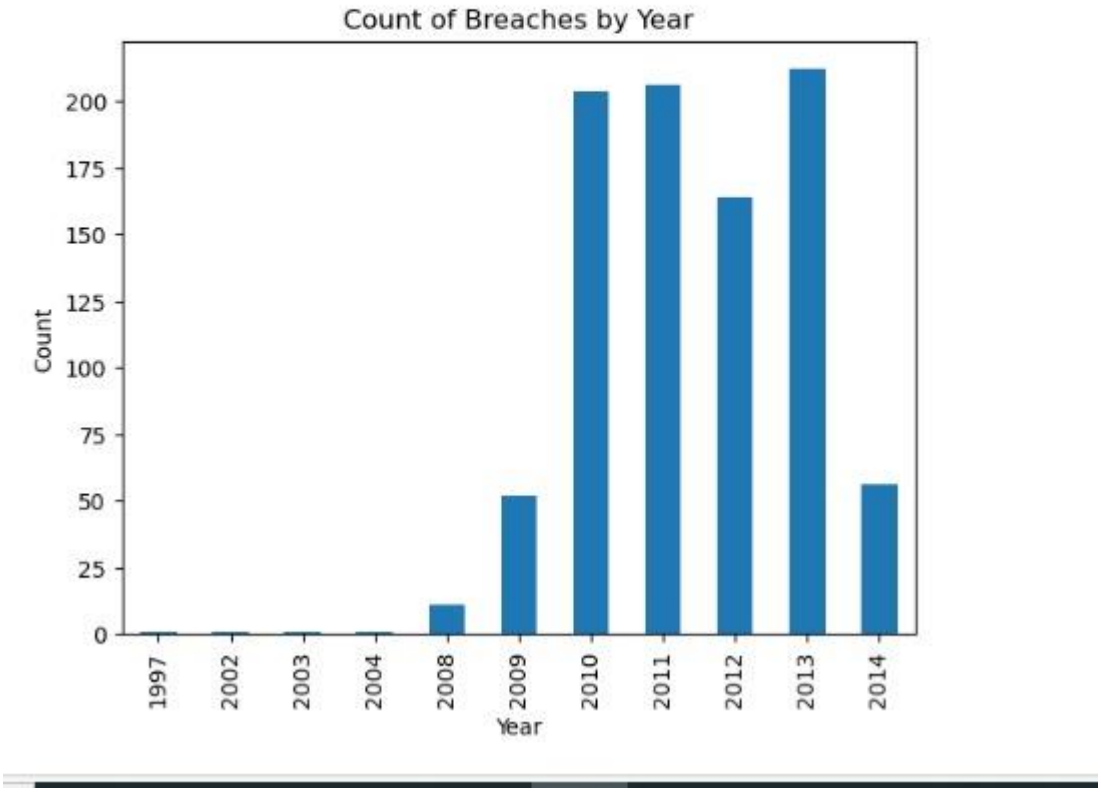
```
# Load the dataset
data = pd.read_excel('Cyber Security Breaches_Migrated Data.xlsx')

# Data Cleaning
data['Breach End'].fillna(data['Breach Start'], inplace=True)
data['Business Associate Involved'].fillna('Unknown', inplace=True)
data.dropna(subset=['Date of Breach'], inplace=True)
data['Summary'].fillna('No summary available', inplace=True)

# Check the updated status of missing values
print("Updated count of missing values in each column:")
print(data.isnull().sum())
```

```
Updated count of missing values in each column:
Breach End                          0
Breach Start                        0
Business Associate Involved         0
Date of Breach                      0
Date Posted or Updated              0
Location of Breached Information    0
Name of Covered Entity              0
State                               0
Summary                             0
Type of Breach                      0
Year                                0
Breach_No                           0
F1                                  0
Individuals Affected                0
Number of Records                   0
dtype: int64
```

Now, to Analyze and find relations between Attributes some visualizations are made using matplotlib as shown below:

1) In which year the count of Breaches was more and from which year there was a spike in the Breach?

```
import matplotlib.pyplot as plt
data['Year'].value_counts().sort_index().plot(kind='bar')
plt.title('Count of Breaches by Year')
plt.xlabel('Year')
plt.ylabel('Count')
plt.show()
```

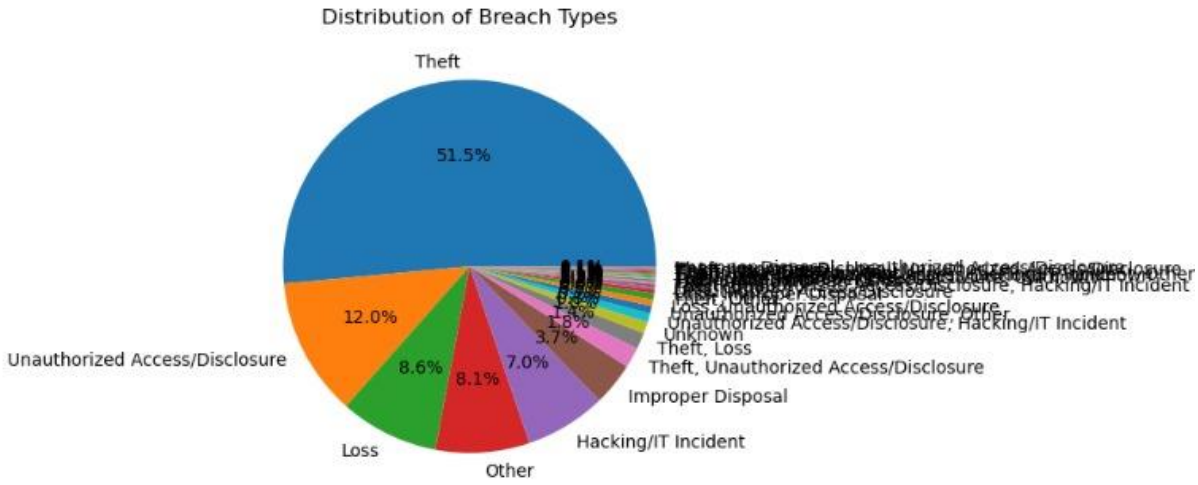


Result:

In the year 2013 the count of Breach was high and from year 2008 there was a sudden spike in the count.

2) According to the pie chart, what is the most common type of cybersecurity breach?

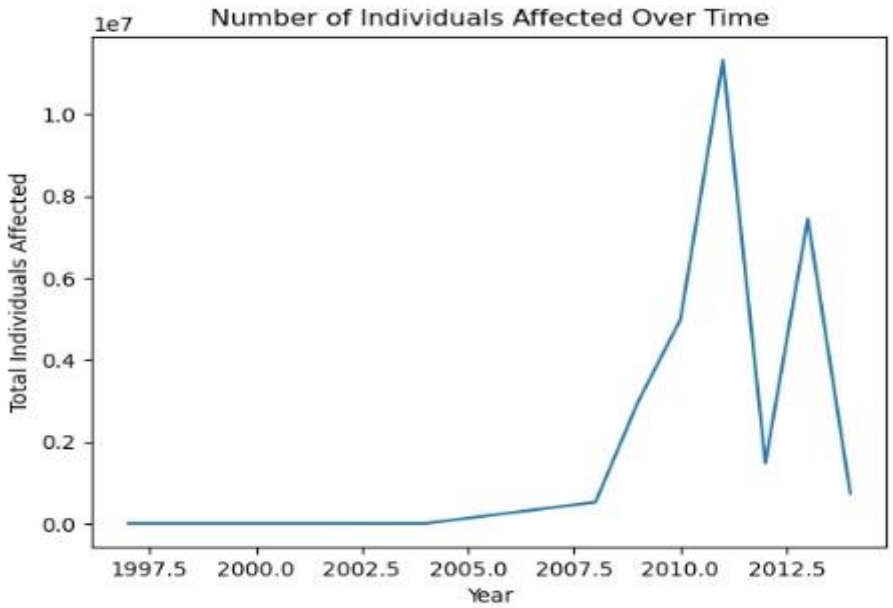
```
data['Type of Breach'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title('Distribution of Breach Types')
plt.ylabel('')
plt.show()
```



Result:
From the pie chart above, Theft is the most common type of cybersecurity breach, accounting for 51.5% of the total breaches represented. This means that the theft of data or physical devices is the dominant concern in cybersecurity incidents.

3) What year, based on the graph, had the most people affected by breaches?

```
data.groupby('Year')['Individuals Affected'].sum().plot(kind='line')
plt.title('Number of Individuals Affected Over Time')
plt.xlabel('Year')
plt.ylabel('Total Individuals Affected')
plt.show()
```

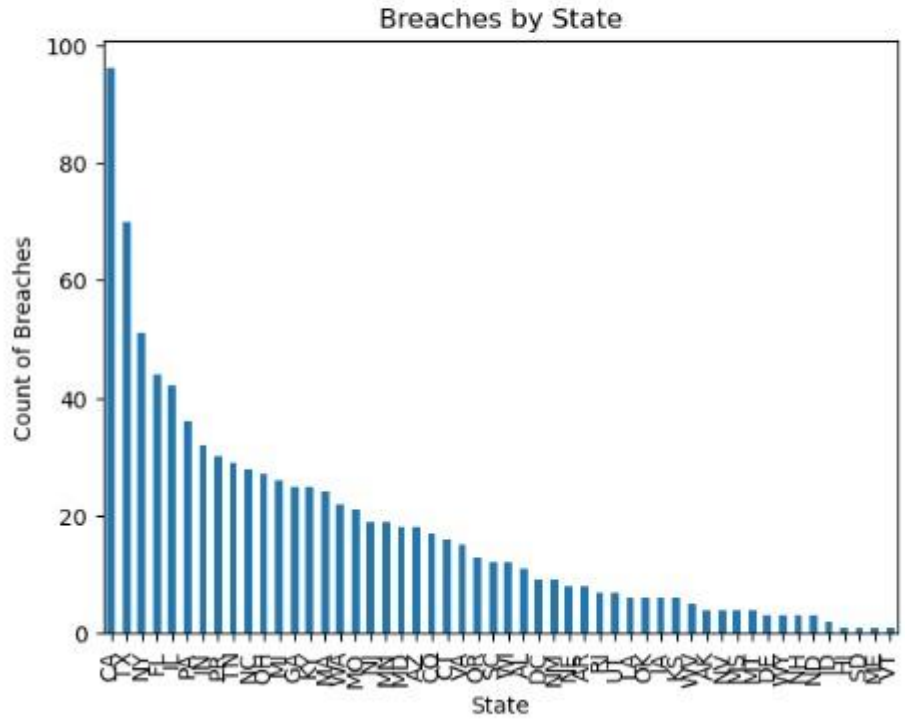


Result:

The year 2010 had the most people affected by the breaches, as indicated by the graph, at nearly 10 million. This would indicate that it was a big security event or multiple events, in general, causing a large-scale impact on personal data during that year.

4) Which state had the highest number of reported cybersecurity breaches according to the graph?

```
data['State'].value_counts().plot(kind='bar')
plt.title('Breaches by State')
plt.xlabel('State')
plt.ylabel('Count of Breaches')
plt.show()
```



Result:
According to this graph, the largest number of reported cybersecurity breaches was in California, which is many times higher than the rest of the states. That may mean a larger number of incidents or just better reporting practices within that state.

Hypothesis

1. Does the involvement of business associates correlate with an increase in the frequency of breaches?

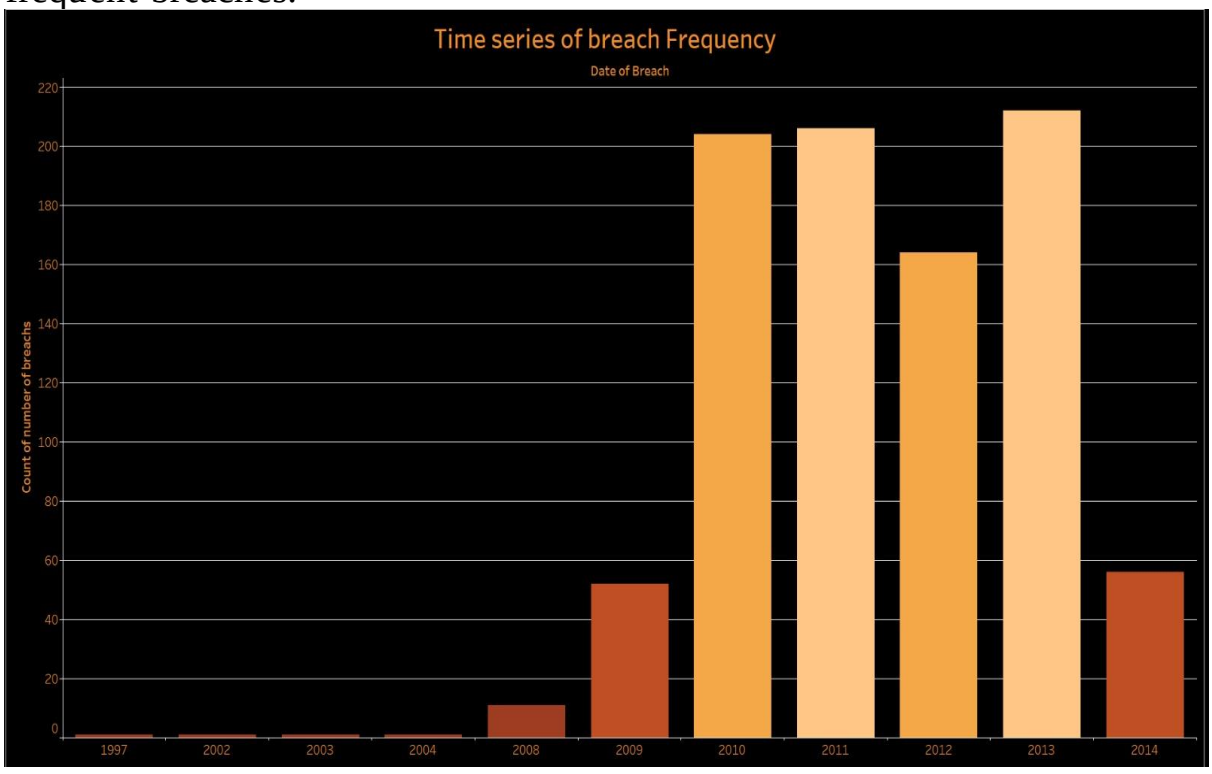
2. Is there a geographical pattern to where breaches occur and the number of individuals affected?

3. Why is the increase in mass-shootings over years considering the gender, ethnicity and mental issues that led to unreasonable end of one's life? What are the top incident areas for possible chance of mass-shooting?

Graph-breakthrough-Result

1. Bar Graph

The graph below, "Time series of breach Frequency," shows a visual presentation of the frequency of data breaches from 1997 to 2014. The y-axis measures the number of breaches, while the x-axis defines years in which each happened. Different colors present the data using bars; the orange bar shows less frequent breaches, and the peach bar symbolizes more frequent breaches.



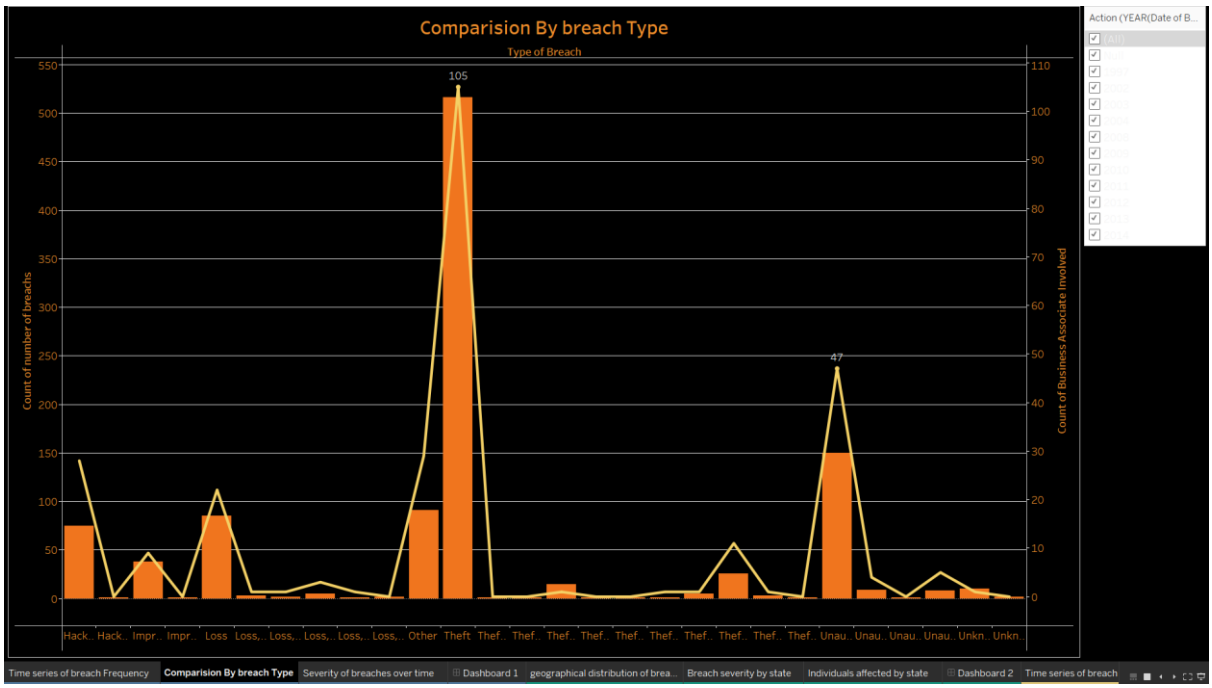
Result

2013 becomes the highest point, with the tallest peach-colored bar depicting the largest number of breaches within the shown period. After that, breach frequency slightly declines after 2011, but it remains relatively high throughout the years 2012, 2013, and 2014, as depicted by the towering peach-colored bars during those year

In summary, the time series graph very well shows the trend of the increasing incidences of data breach over time, with an enormous increase in 2013 and an overall increase in breach frequency from 2008 onwards.

Combination Chart of both Bar Graph and Line Graph

The bar graph titled "Comparison By Breach Type" offers a comparative analysis of various data breach types or security incidents. Along the x-axis, different breach types such as "Hack", "Improper Disposal", "Loss", "Theft", "Unauthorized Access", and "Unkown" are listed. The y-axis quantifies the count or number of breaches corresponding to each breach type.



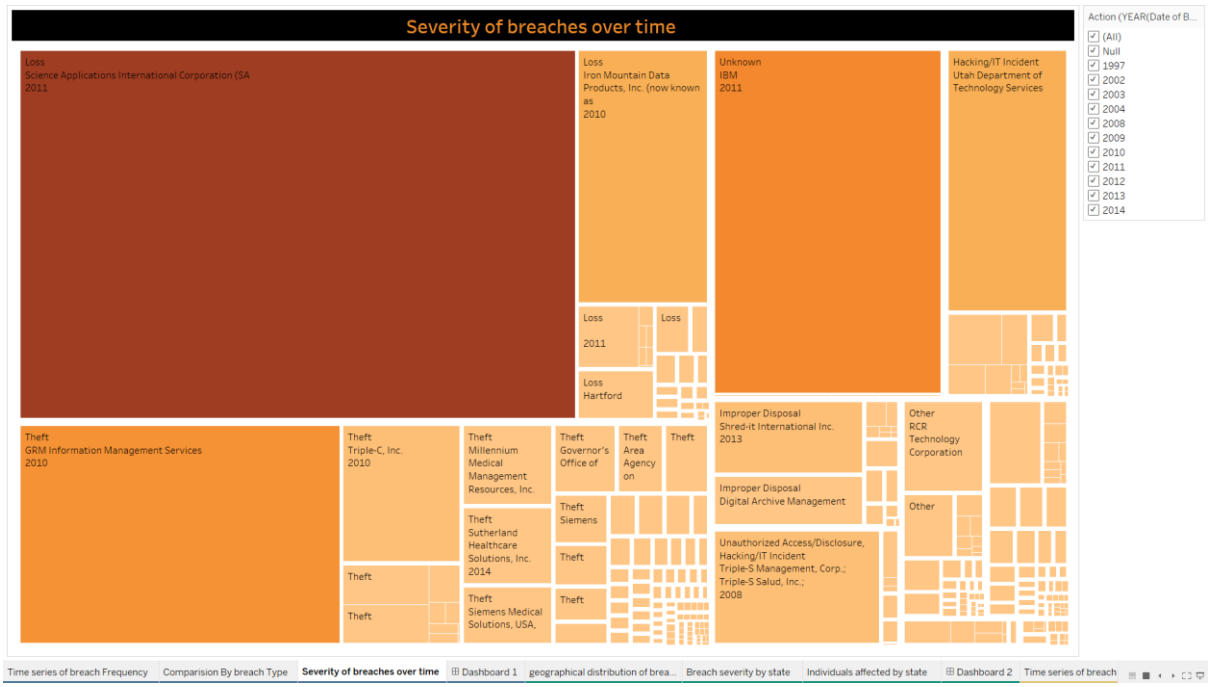
Result

The graph exhibits several distinct peaks, with the highest peak reaching approximately 110 incidents for the breach type labeled "Loss". Notable peaks also include around 105 incidents for "Theft" and roughly 100 incidents for "Improper Disposal".

This visualization enables a straightforward comparison of the relative frequencies of different breach types, shedding light on the most prevalent types of security incidents or data breaches. By providing an overview of the distribution of breach types, it facilitates the identification of potential risk areas necessitating attention for enhancing data security and implementing effective breach prevention measures.

TreeMap

Here, the visualization represents a time-based data breach or security incident, wherein the size of the bar or the intensity of the color represents the severity or magnitude of the breach. The y-axis enumerates specific incidents or organizations involved, while the x-axis portrays the passage of time, albeit without specific dates or years.

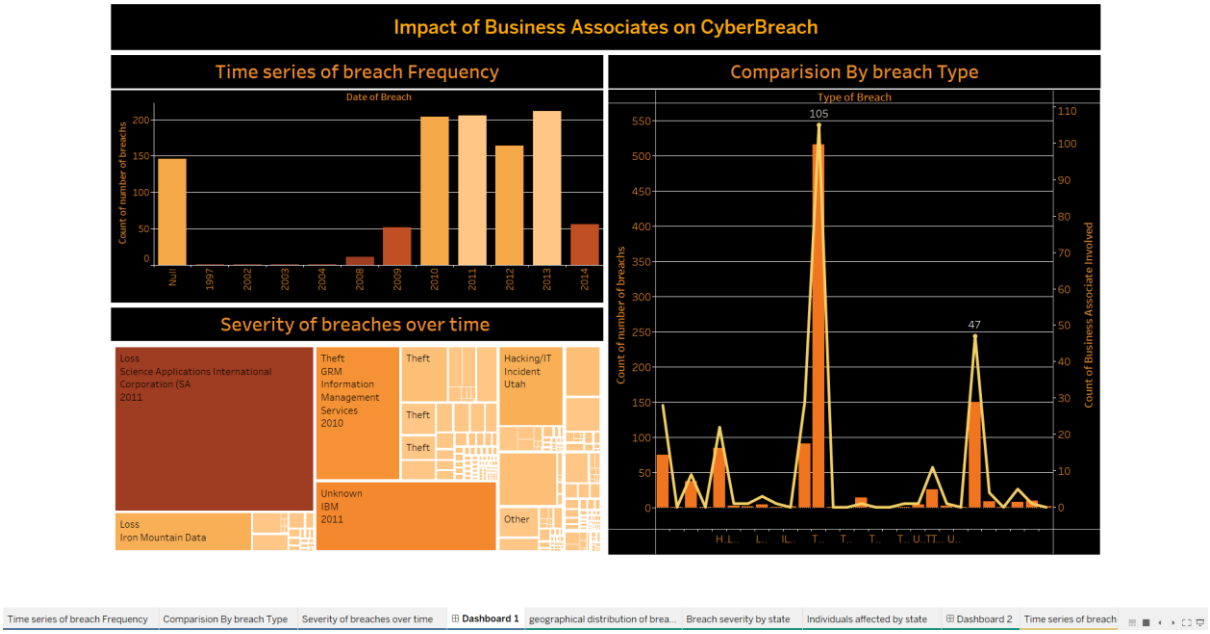


Result

Notably, a large red rectangle, "Loss Science Applications International Corporation (SA)" in 2011, indicates a massive loss event involving data; another large orange rectangle in the year 2010, "Theft GRM Information Management Services," shows a massive theft incident. Many theft incidents are evident with several events marked as "Theft" with organizations such as Siemens Medical and Triple-C, Inc. A 2013 event labeled "Improper Disposal Shred-it International Inc." manifests that improper disposal is a common form of breach. Moreover, a 2008 event labeled "Unauthorized Access/Disclosure, Hacking/IT Incident Triple-S Management, Corp.: Triple S Salud, Inc." manifests a complex breach involving unauthorized access and hacking.

This visualization helps in comparison over time in the relative severity or impact of different data breach incidents, which will have larger rectangles representing more severe events. The main insight from this visualization is in identifying the kinds of breaches faced by various organizations, such as data loss, theft, improper disposal, and unauthorized access or hacking.

Dashboard 1



This dashboard appears to present data on the impact of business associates on cyber breaches, structured into two primary sections:

Time Series of Breach Frequency (Left Side):

This section displays a bar graph tracking the count of breaches from 1997 to 2014. We see a sharp increase in breaches starting in 2009, peaking in 2011, and remaining relatively high until a drop-off in 2014.

Comparison By Breach Type (Right Side):

The graph on the right compares the types of breaches with the involvement of business associates over the same timeline, represented by two metrics: total breaches and business associate involvement. The peaks on this graph suggest certain years where breaches were particularly high, with prominent involvement from business associates.

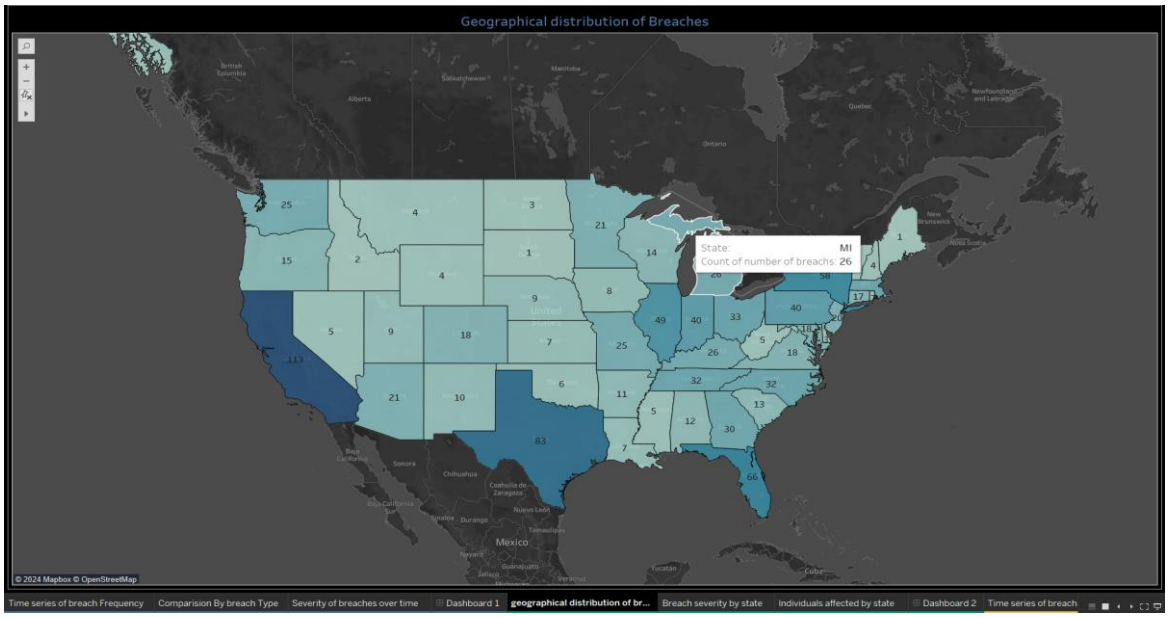
Severity of Breaches Over Time (Middle Bottom Section):

This section, styled as a heatmap, indicates the severity of breaches over time with specific instances noted, such as the "Loss at Science Applications International Corporation (SAIC)" in 2011, and "Theft at GRM Information Management Services" in 2010. The heatmap's density and color depth likely represent the severity and frequency of breaches.

From the visual data on the right graph, where the yellow line (count of business associate involvement) peaks notably in 2009, 2011, and a smaller peak in 2013, a pattern emerges that suggests a correlation between high involvement of business associates and increased breach events. This correlation appears particularly strong in 2011, a peak year for both metrics. This suggests that years with significant business associate involvement also experienced high numbers of breaches, indicating a potential correlation between these factors.

2. Geographical Representation

This graphical presentation shows a choropleth map illustrating the geographical distribution of cybersecurity breaches across the United States. Each state is shaded according to the number of breaches reported, with darker shades indicating a higher number of incidents.



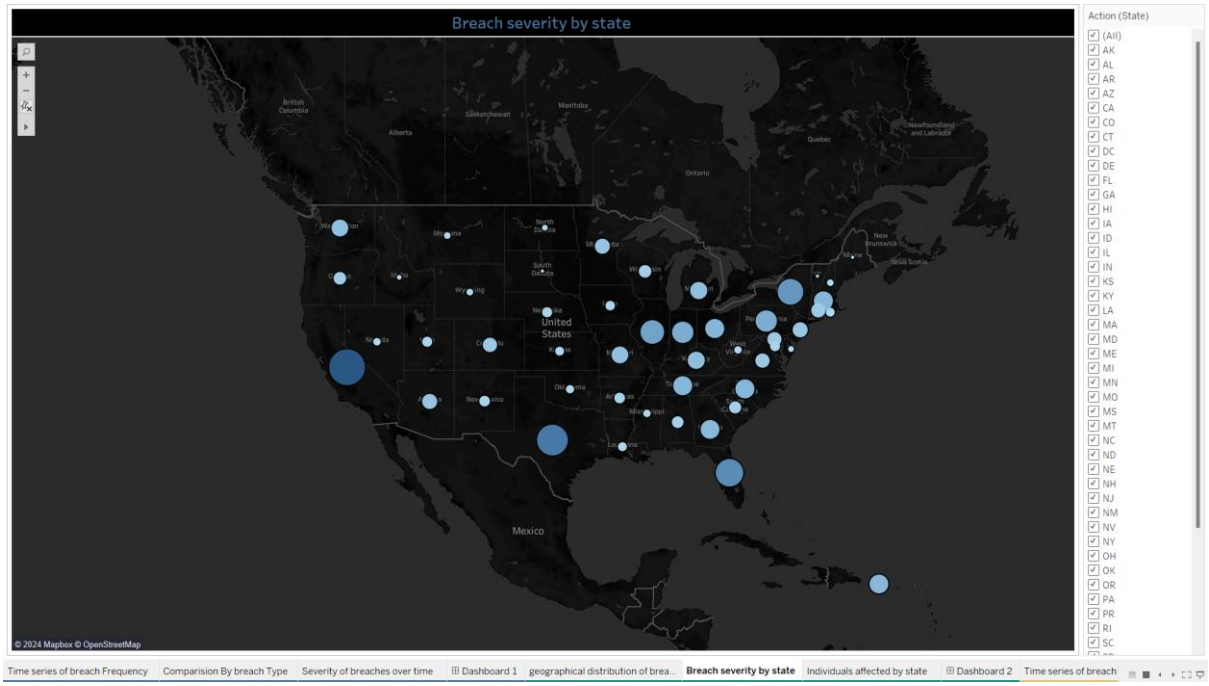
Result:

This map clearly visualizes which states have a higher incidence of cyber security breaches. For example, California, Texas, and Florida show much darker shades, pointing to a higher number of breaches. This visualization enables stakeholders to quickly identify and focus on higher-risk areas for targeted cybersecurity measures, policy implementation, or further detailed analysis.

The map serves as an effective communication tool for the delivery of complex data in an accessible format, enabling better geographic insights into how the cyber security incidents are actually distributed across the country.

Bubble Map:

This visualization is a bubble map plotted on a dark-themed geographic base of the United States. It shows the severity of cybersecurity breaches by state, with the size of each bubble representing the relative seriousness of breaches in each location.



Graph Explanation:

Bubbles: Each bubble is placed over a state and the size of the bubble corresponds to the severity of breaches in that state. The larger the bubble, the more severe the breaches.
Color: All bubbles are the same shade of blue, suggesting that the color does not differentiate data but is rather used for visual clarity and emphasis on severity.
Geographical Layout: The bubbles are plotted on a map of the U.S., allowing viewers to immediately see geographical patterns and the distribution of breach severities.

Result:

Geographical Insights: Areas with larger bubbles, as in the Northeast and West Coast, indicate higher severity, which could be critical to regional security operations and resource allocation.

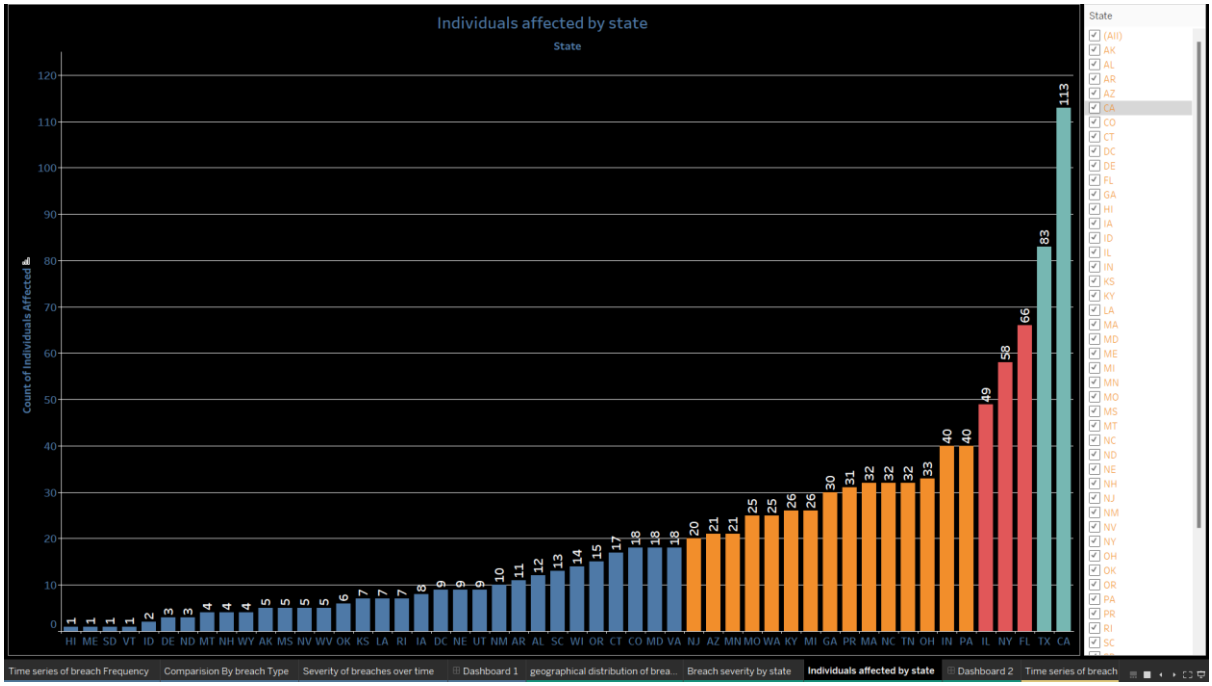
Resource Allocation: With this data, organizations and governments can make decisions about intensifying cybersecurity measures.

Strategic Planning: This information can be used by companies for risk assessment, to invest in the right infrastructures, and determine regional priorities for cybersecurity enhancements.

Overall, the map is very instrumental for the quick and uncomplicated portrayal of complex geographical data, enabling understanding and decision-making about cybersecurity risks and management to be easier.

Bar Graph:

This is a bar chart titled "Individuals affected by state" that depicts the number of individuals affected by cybersecurity breaches across various U.S. states. On the x-axis, the states are listed, while on the y-axis, the count of affected individuals is represented. The bars are color-coded, which might indicate different ranges of the number of affected individuals or could be used to segregate certain groups or categories, although the specific meaning of the colors is not indicated without a legend.



Key Features of the Graph:

States: Each state is represented by a bar; the state abbreviation is at the base of each bar to identify the state.

Count of Affected Individuals: Quantified on the y-axis is the count of the number of people affected, where the height of each bar represents the relative magnitude of the number of affected people in that state.

Color Coding: The bars come in shades of blue, orange, red, and teal. This could mean that some of this is categorized by type or by severity of breach impact—maybe it even sorts the states by regions.

Result:

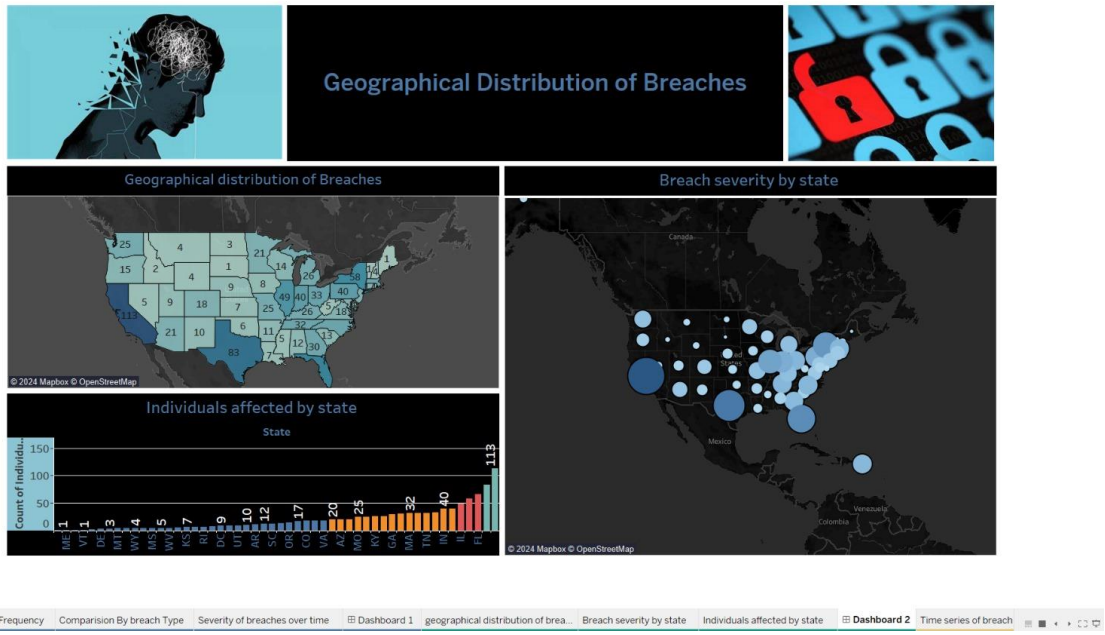
The number of affected individuals by cybersecurity breaches in California, Texas, and Florida is highest, crossing over 100. This indicates high involvement of the state due to a high population density, higher business prevalence, or commonness of the breach or seriousness of breach incidents.

NY and PA also show considerable numbers, indicating a comparatively bigger impact of breaches in these states.

Whereas for states like Maine (ME), South Dakota (SD), Vermont (VT), and North Dakota (ND), the number of affected individuals is very low, with their bars being the shortest. This may imply a lower impact of breaches, be it due to the lower population densities, fewer businesses, or more effective cybersecurity measures.

Dashboard 2:

This dashboard presents a detailed visualization of cybersecurity breaches in the United States, emphasizing the geographical distribution of breaches and severity of exposed individuals by state.



Components of the Dashboard:

Geographical Distribution of Breaches: A choropleth map showing the number of breaches by state. Each state is shaded differently, depending on the number of breaches within it.

Breach Severity by State: A bubble map where the size of every bubble represents the level of seriousness of breaches in that state. Placing the bubbles in geographic locations of states visually depicts the relative seriousness of the breaches.

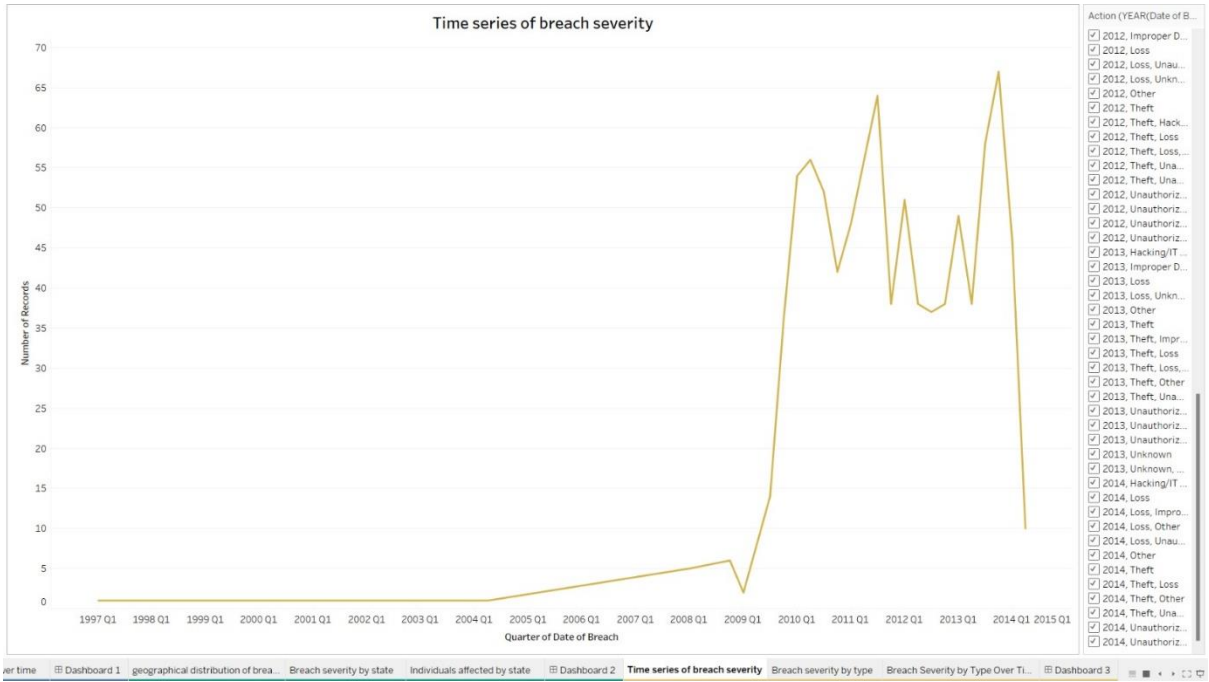
Individuals Affected by State: A horizontal bar chart that shows the numbers of individuals affected by cybersecurity breaches within each state. States are arranged along the x-axis, with count-of-individuals-affected on the y-axis. Bars are color-coded to represent different levels of impact.

Result:

It effectively demonstrates a geographical pattern with regard to the number of breaches and affected persons. More populous states and those with higher business activities, such as California and Texas, have a higher rate of breaches, which are also more severe, affecting higher numbers of people. This integrated visualization approach provides valuable insights for policymakers, cybersecurity professionals, and the general public to understand and react to the landscape of cybersecurity threats in the United States.

3) Line Graph:

The graph depicted here is a "Time Series Line Graph," used to track changes over time in whatever data the graph is representing. Specifically, this graph is labeled "Time series of breach severity," and it shows a trend in the number of records—presumably relating to cybersecurity breaches—over time, from 1997 through the first quarter of 2015.



Explanation of the Graph:

X Axis—Time: Time on the horizontal axis is measured in years, with each one further broken down into quarters. Such a time breakdown allows a trend that may not be apparent while viewing the summary for a year to pop out at the granular level of quarter-to-quarter.
Y Axis—Number of Records: The number of records, it is assumed,

correlates with the number of incidents, or at least some measure of data related to breaches, such as the number of records that have been compromised or the severity of the breaches. The scale is from 0, and peaks above 60, which shows the relative variability of the data.

Line Trend: There's a fluctuating trend in the line graph with periods of mountain peaks and valleys. One such example is the high record in the graph about the year 2005; afterwards, there is a relative low number of values up to around 2011.

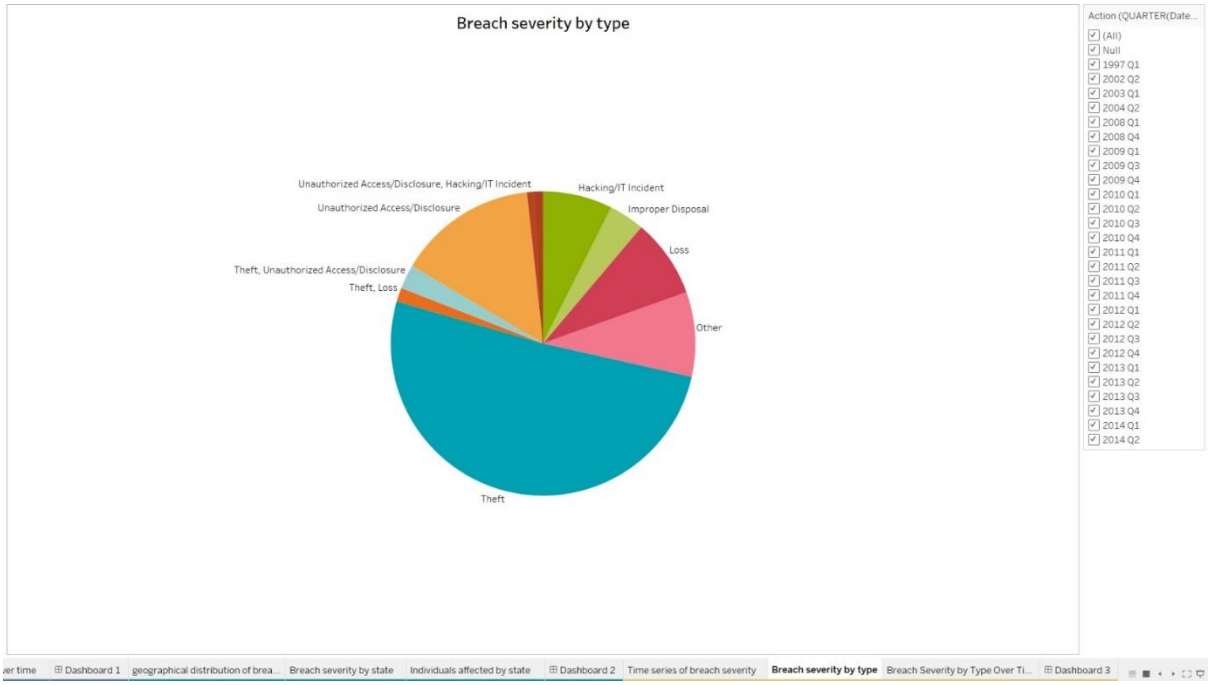
The graph shows a huge spike around the years 2010-2011, revealing some big event or change, like a huge data breach or something that changed the way data breaches were recorded or reported.
The number of records then drops precipitously, perhaps meaning that the issues that caused the spike were resolved, or security measures improved.

Result:

The graph of the results indicates that, with a few reversals, the trend of high severity seems to peak in 2010-2011, pointing to a critical period that might need further analysis to understand the underlying causes. It also shows efforts to mitigate breaches possibly being effective, as can be evidenced by the subsequent decline in the number of records following this peak. In general, the upward trend indicates that breaches are increasingly a concern that calls for increased attention and adaptation of cybersecurity strategies

Pie Chart.

The pie chart titled "Breach severity by type" displays the distribution of cybersecurity breaches by their type. Each segment of the pie chart represents a different type of breach, and the size of each segment indicates the proportion of that type of breach relative to the total number of breaches recorded.



Description of the Pie Chart:
Major Segments:
Theft: It is the biggest segment of the pie chart and shows that theft is the most usual kind of breach.

Unauthorized Access/Disclosure: This one is also vast; it describes that the majority of the breaches result from unauthorized access or exposure to data.

Hacking/IT Incident: Represented as another big segment, it represents that hacking or IT incident is a major source of breach.

Loss: It represents a breach that is caused because of the loss of devices or data.

Improper Disposal: This is the type of breach that results from the improper disposal of data, for example, due to the failure in the destruction of the data storage device.

Theft, Unauthorized Access/Disclosure: Combined category representing incidents having both theft and unauthorized access/disclosure.
Theft, Loss: Another combined category representing a breach having both theft and loss.

Other: This category consists of those types of breaches that do not fit into the above categories, like insider threat or the problem in software.
Unauthorized Access/Disclosure, Hacking/IT Incident: It is the category that

combines both the unauthorized access/disclosure and hacking/IT incident, describing the complex breach in activities.

Result:

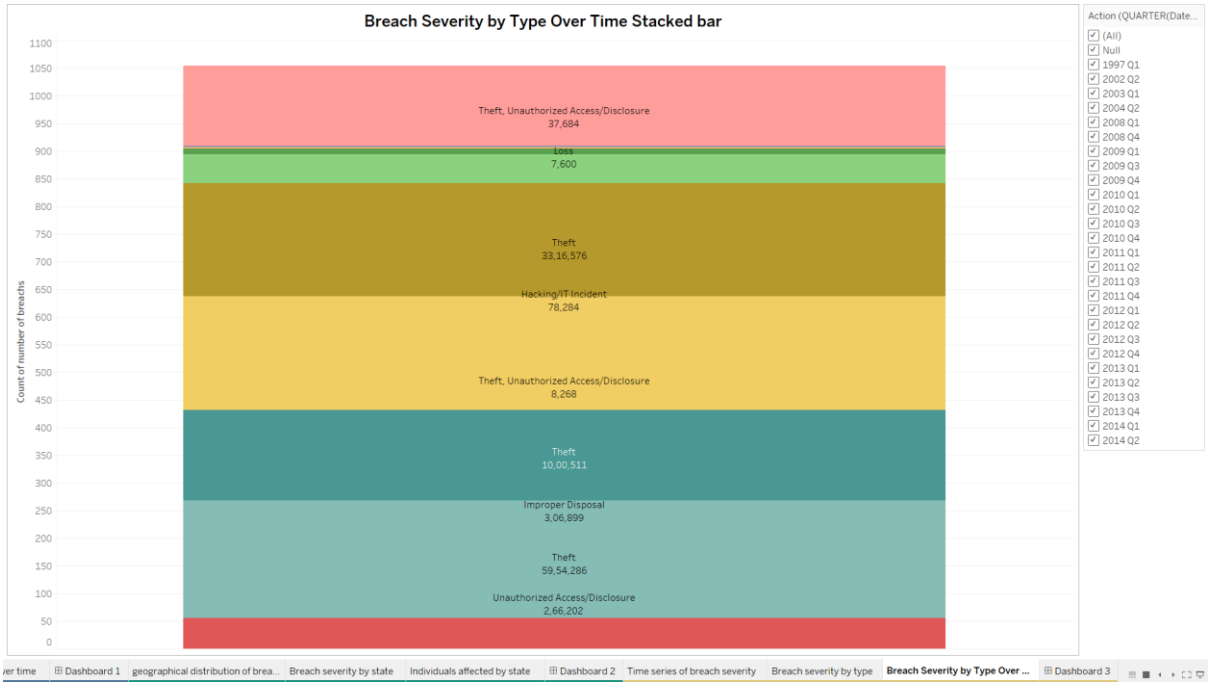
The pie chart shows the different kinds of breaches and their relative frequency.

The results show that:

Theft is the dominant issue, suggesting that physical security and asset management are critical areas needing attention and improvement.
Hacking/IT incidents and unauthorized access/disclosure are also high, showing the need for advanced cybersecurity measures, network security, and data access controls in order to prevent breaches from this type.
The presence of combined categories, such as Theft and Unauthorized Access/Disclosure, would suggest that most breaches involve a blending of security failures, which actually brings home the call for elaborate security approaches that recognize multi-dimensional vulnerabilities in schemes and operations.

Stacked Bar Graph:

The graph presented is a "Stacked Bar Chart" titled "Breach Severity by Type Over Time Stacked bar." It visualizes the count of cybersecurity breaches categorized by their types over an unspecified time frame.



Explanation of the Graph:

Y-axis: The count of breaches, which is the sum of each category in total count and time.
Colour Coding: In the bar, each color conveys a different type of cybersecurity breach. Each type of breach is marked with a different color. Labels on the right side of the bars indicate the type of breach with their respective numbers specifying the count of breaches for each category.

Types of Breaches:

Theft: In teal, there are significant occurrences with specific counts such as 59,54,286 and 10,00,511, which may indicate widespread issues related to theft.
Unauthorized Access/Disclosure: This comes in two parts—one darker, one lighter—presumably to indicate differences in context or specifics of unauthorized access/disclosure incidents, 37,684 and 2,66,202.
Hacking/IT Incident: This category, represented in yellow, shows that there have been 78,284 incidents, which portrays the seriousness and prevalence of hacking or IT-related breaches.
Loss: Green, 7,600 breaches representing loss of data or a device.
Improper Disposal: This smaller slice in orange, of 3,06,999, involves breaches due to the improper disposal of sensitive information.

Result:

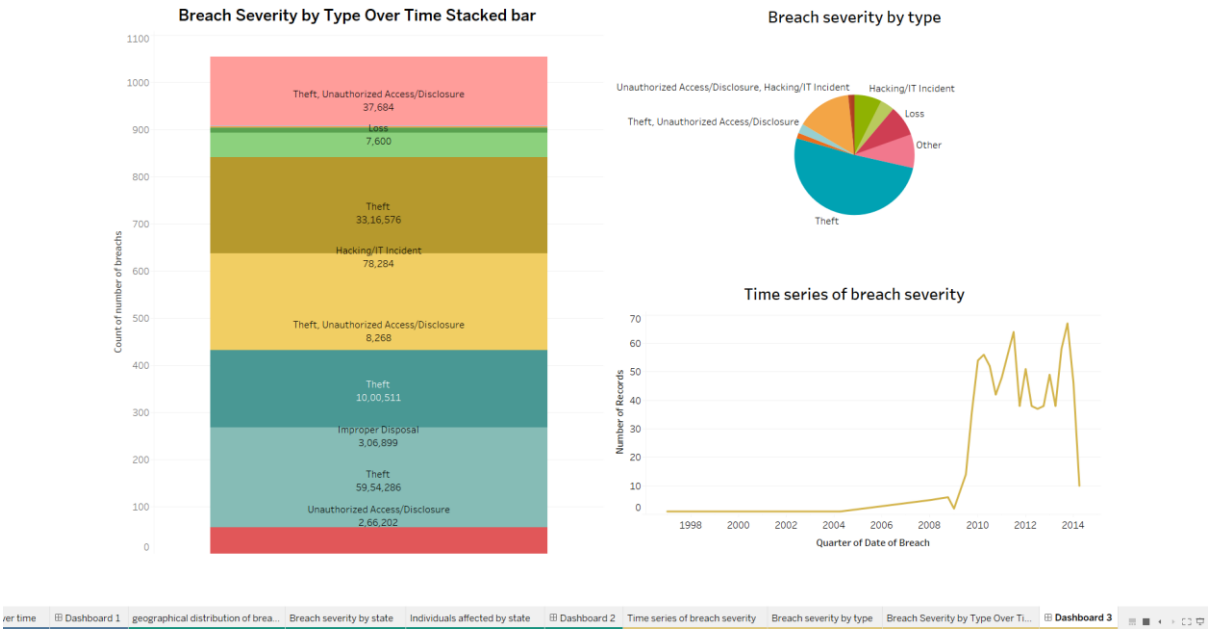
This graph gives the holistic perspective of the breach type distribution and severity, underlining the dire need to direct cybersecurity improvements into the most-affected areas, particularly in fighting theft and strengthening defenses against hacking incidents. The provided data will empower strategic decisions over cybersecurity investment, policy formulation, and operational adjustments in better protection for sensitive information and systems.

Dashboard 3

This dashboard brings together three visualizations to allow an end-user to understand the kind of breaches that have occurred and the severity over time:

Breach Severity by Type Over Time Stacked Bar: This chart stacks various types of breaches by their count, providing a view of the distribution and changing nature of breach types over an unknown period. Each color represents a type of breach, for example, Theft, Hacking/IT Incidents, Loss, etc., with the number of occurrences for each category as the label.

Breach Severity by Type Pie Chart: This pie chart gives a perspective on the proportion of how different types of breaches contribute to the total number of incidents. The segments are color-coded in the same manner as the stacked bar chart, visually connecting the two charts.
Time Series of Breach Severity: A line graph of the number of records involved in breaches over time. This provides a clear visual trend of ups and downs in the seriousness of breaches starting in 1998 to 2014.
Addressing the Hypothesis:
The hypothesis question is whether there have been changes in the severity of breaches over the years. Now, upon closer analysis of the given visualizations**:**



Result:

By the data shown in the dashboard, it is no doubt that the severity of breaches has gone through many changes over the years. The number of breaches and the nature of these breaches have changed to an increase in both the complexity and impact of the incidences. These trends are well exemplified by the dashboard, which confirms the hypothesis that the severity of breaches has indeed changed over time, with an upward trend in the number and severity of incidents.

Discussions

The data represent an increasing number of breaches over the years, in particular, in 2013. It's very important to discuss how this trends might be related to the evolution of technology, including the development of cybersecurity tools and methods of hackers. As technology evolves, so do cybercriminals' methods, which might explain the spikes in breach activity. While the data visualization shows the frequency and types of breaches, it opens up a discussion about the effectiveness of implemented cybersecurity measures. Analysis of periods when the decrease or stability of breach incidents might shed light on successful strategies and policies.

Evaluating response strategies adopted after major breaches and their long-term effects on breach frequencies could help in crafting more resilient cybersecurity frameworks. This could include an in-depth look at case studies of specific high-profile breaches and how they have led to a change in legislation or corporate policies.

It indicates significant geographical disparities in breach incidents, where states such as California, Texas, and Florida are most affected in terms of the number of breaches. The question would be what regional factors—economic conditions, the concentration of tech industries, or state-specific cybersecurity laws—could give way to such patterns.

Involvement by business associates was shown to correlate with an increase in the number of breaches. This points to a very critical aspect of cybersecurity, extending beyond the confines of a single organization: supply chain and third-party risk.

Conclusion

This comprehensive analysis of cybersecurity breaches over the period of five years sets focus on various key dynamics in the realm of cyber threats and security incidents across the United States. Leveraging data visualization tools, we have been able to carve out and display the details of these breaches in a detailed manner, giving clarity on their nature, frequency, and the effect they cause.

Key Findings:

Increasing Trend of Breaches: This analysis confirms a large increase in the rate of cybersecurity breaches, particularly noted in 2013. The uptick of this trend shows the growing challenges the cyber domain faces organizations and individuals.

Prevalence of Theft and Unauthorized Access: Theft emerged as the most common type of breach, followed closely by unauthorized access and disclosure. These results highlight the areas where cybersecurity measures have to be most intensified.

Geographical and Demographic Disparities: The geographical analysis points to a more frequent targeting of states such as California, Texas, and Florida, because of their heavier population and weighty business activities. This pattern calls for region-specific strategies to mitigate risks effectively.

Future Directions:

Future research should aim at including more recent data to analyze the trends in the context of new technological advancement and cyber threat tactics.

Further, the effectiveness of cybersecurity measures implemented post-breach needs to be studied in great detail to understand which strategies are most effective in terms of mitigating the damage and preventing future incidents.

Concluding Remarks:

Such a project has not only unveiled key insights in patterns and their effects of cybersecurity breaches but has also shown the power of data visualization in making complex data understandable. It calls for constant innovation and improvement in cybersecurity strategies to keep abreast with the dynamics of the protecting environment of sensitive information within the ever-evolving digital world.

References

- Kerem Gülen, *AI and Ethics: Balancing progress and protection* – Jan 16, 2023 - https://dataconomy.com/2023/01/16/artificial-intelligence-security-issues/
- Petar Radanliev, David De Roure, Carsten Maple, Uchenna Ani, *Super forecasting the technological singularity risks from artificial intelligence* - https://arxiv.org/ftp/arxiv/papers/2301/2301.10028.pdf
- Abhilash Chakraborty, Anupam Biswas, Ajoy Kumar Khan, *Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation* - https://arxiv.org/ftp/arxiv/papers/2209/2209.13454.pdf
- Andrew J Lohn, Krystal Alex Jackson, *Will AI Make Cyber Swords or Shields: A few mathematical models of technological progress* - https://arxiv.org/pdf/2207.13825.pdf