AppliedMath

MDPI

Article

Analyzing Health Data Breaches: A Visual Analytics Approach

Wullianallur Raghupathi [1], Viju Raghupathi [2,*] and Aditya Saharia [1]

[1] Gabelli School of Business, Fordham University, New York, NY 10023, USA; raghupathi@fordham.edu (W.R.); saharia@fordham.edu (A.S.)
[2] Koppelman School of Business, Brooklyn College, City University of New York, Brooklyn, NY 11210, USA
* Correspondence: vraghupathi@brooklyn.cuny.edu

Abstract: This research studies the occurrence of data breaches in healthcare provider settings regarding patient data. Using visual analytics and data visualization tools, we study the distribution of healthcare breaches by state. We review the main causes and types of breaches, as well as their impact on both providers and patients. The research shows a range of data breach victims. Network servers are the most popular location for common breaches, such as hacking and information technology (IT) incidents, unauthorized access, theft, loss, and improper disposal. We offer proactive recommendations to prepare for a breach. These include, but are not limited to, regulatory compliance, implementing policies and procedures, and monitoring network servers. Unfortunately, the results indicate that the probability of data breaches will continue to rise.

Keywords: breach type; data breach; healthcare; healthcare provider; information technology; location; visualization; visual analytics

1. Introduction

Healthcare is a lucrative target for hackers. As a result, the healthcare industry is suffering from massive data breaches [1–4]. Healthcare data breaches result in "the loss, theft, or unauthorized access to data containing sensitive personal and health information" [5]. A data breach is likely to happen when an unauthorized person penetrates a source of data ("location") and retrieves information he or she is not authorized to access [3,6,7]. This is typically accomplished by tapping into a computing system, device, or network to access files and data in an unauthorized fashion, and with an ulterior motive. Evading security, one may access the data remotely via the Internet or in a wireless fashion. These methods typically target business entities. The typical steps executed during a data breach include researching the site, planning, and then attacking and breaking out [8–10]. Because healthcare data breaches carry the risk of a loss of privacy [11,12] through personal health information exposure, corruption, or destruction, this study is important to the healthcare field [13–15]. Extending this discussion, patients often disclose detailed, sensitive health information online unintentionally. Furthermore, the risk associated with involuntarily disclosure of private data into publicly available sites has been rather ignored. To reiterate, though, in the last two years alone, several billion records have been stolen or made publicly available due to several data breaches [16].

Health data can be generated in large volumes from a variety of sources, such as wearable devices, online patient groups, social media postings, and web searches. In online patient forums, some participants share wellness information using their own names, while others use pseudonyms for the sake of privacy. Many online participants post data in the belief that it will only be shared with the designated receivers [17–19]. However, privacy continues to be an ongoing challenge. For instance, even though mHealth apps facilitate access to real-time monitoring and health resources, they also present an inherent threat to privacy, particularly because of the sensitive nature of the informational content, as well as the lack of enforcement of privacy standards worldwide for online posting [12].

check for updates

Citation: Raghupathi, W.; Raghupathi, V.; Saharia, A. Analyzing Health Data Breaches: A Visual Analytics Approach. *AppliedMath* **2023**, *3*, 175–199. https://doi.org/10.3390/appliedmath3010011

Academic Editor: Gaige Wang

Received: 20 December 2022
Revised: 18 February 2023
Accepted: 21 February 2023
Published: 9 March 2023

There is sufficient evidence of healthcare data breaches that have occurred due to a lack of sound security measures [11]. A patient survey in the U.S. shows that 75% were concerned about health websites sharing unauthorized information [20]. Additionally, medical data breaches have been shown to be the second highest reported type [21]. As a measure to address the increasing level of threat to health information privacy in the U.S, several state- and federal-level regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), have been proposed [11,22]. While the patient–physician relationship relies on the foundation of privacy, patients are required to disclose all medical information to the physicians so as to facilitate accurate diagnosis and treatment. However, in certain realms, such as mental health and HIV, patients may not be comfortable with such disclosure due to the social stigma associated with the outcomes [23]. Over time, a patient's medical record accumulates important personal information relating to identification, medical diagnosis, imaging records, medications, sexual preferences, dietary preferences, and mental assessments [11,24,25]. Such a vast arena of information presents a viable source for data theft and is, therefore, vulnerable to data breaches [26,27].

The bottom line is that healthcare institutions remain vulnerable targets for a wide range of cyber threats, including technical, physical, and human issues [28–30]. Cyber-perpetrators continue to exploit these vulnerabilities with increasing sophistication, capitalizing on stolen healthcare records [13,31]. There are many negative effects caused by data breaches that impact uninvolved populations, organizational assets, and the healthcare environment in general, including monetary and privacy losses [32]. Because of the vulnerability of healthcare organizations and the many negative consequences for those experiencing a data breach [33], this work examined factors associated with data breach occurrences. The topic is important because healthcare data breaches expose personal data to theft, modification, or misuse [34]. By exploring data breach dimensions and factors, this paper may assist healthcare delivery entities in mitigating or preventing data breaches in a proactive manner [11,13,32]. Based on published industry reports, fundamental security safeguards are still considered to be lacking, with many documented data breaches occurring as the result of device and equipment theft, human error, hacking, ransomware attacks, and misuse. Health and medical data are believed to be one of the most vulnerable targets for cybercriminals due to their obvious susceptibility. Furthermore, organizations appear not to be ready to carry out forensic investigations into health data breaches, rendering mitigation and remedial steps rather moot [5].

In 2017, 12,000 Aetna patients' human immunodeficiency virus (HIV)-related information was revealed due to a mailing error that exposed personal information through an envelope's clear window [28]. In other healthcare and hospital cases, patients' full names, addresses, social security numbers, contact information, and health insurance numbers were stolen and exposed. Moreover, the size of data breaches is often very large. In August 2015, a cyberattack claimed the private information of approximately 10 million members of Excellus, including medical data, social security numbers, and financial information. Another cyber-attack involving Premera Blue Cross exposed over 11 million customers' information, including bank account numbers, claims information, social security numbers, and dates of birth. As the largest health data breach case in history, 78.8 million patient records were revealed. In other words, one data breach impacted one in four Americans.

The number of reported data breaches has increased since 2009. The 2546 healthcare industry breaches between 2009 and 2018 exposed 1,899,445,874 healthcare records [29]. That is more than half of the population of the United States. Therefore, more than half of the nation's citizens are at risk of identity theft or fraud [13,30,31,33].

This upward trend was reflected in every year except 2015. However, this does not mean that the number of data breaches improved in 2015. On the contrary, the number of records exposed in 2015 reached a peak of approximately 120 million. The average data breach size in 2015 was 400,000 [32,34]. This was a result of the 3 largest healthcare data breaches, including one impacting 78.8 million records [35]. It also takes a long time to discover breaches. According to an IBM survey, it takes about 55 days for healthcare

organizations to detect a data breach [36]. The Nuix Black Report surveyed 112 hackers. It revealed that 61% of the hackers take less than 15 h to obtain healthcare data. Therefore, this is an imperative challenge when preventing future data breaches [37,38].

Healthcare entities, such as doctors' offices, hospitals, laboratories, health insurance companies, HMOs, and other providers, increasingly face cyberattacks resulting in data loss, identity theft, privacy loss, business disruption, etc., with consequences of monetary and reputational loss. The entities also face lawsuits and litigation. It is, therefore, imperative to proactively understand the nature of data breaches and to take steps to mitigate or prevent such breaches [3,8,30]. According to an Experian Data Breach Industry Forecast, health entities face escalating cyberattacks and data breaches due to the various data access points, such as terminals and front-office computers, and from such applications as electronic health records and wearable devices (https://www.experian.com/data-breach/2023-data-breach-industry-forecast) (accessed on 16 December, 2022). Likewise, the Third Annual Benchmark Study on Patient Privacy and Data Security by the Ponemon Institute observed that nearly 94% of health entities have had at least one annual data breach in recent years (https://www.ponemon.org/news-updates/news-press-releases/news/third-annual-benchmark-study-on-patient-privacy-data-security.html) (accessed on 16 December 2022). Interestingly, many of these were intentionally or unintentionally caused by employees (42%). Forty-six percent of individuals responding to the survey mentioned lost or stolen computing devices. Furthermore, third-party errors contributed to a large percentage of the breaches. Additionally, the large-scale utilization of mobile devices is jeopardizing patient data. A vast majority of health entities have authorized employees to use their personal devices to establish connections to their workplace computers [32]. As mentioned, hacking is the fastest way to obtain unauthorized data. A hacker is:

> "*an individual who uses computer, networking, or other skills to overcome a technician problem. The term hacker may refer to anyone with technical skills, but it often refers to a person who uses his or her abilities to gain unauthorized access to systems or networks to commit crimes.*" [39]

The accelerated adoption of electronic health record systems (EHRs) as a result of the passing of the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 has led to the automation of numerous health processes, with large amounts of health and patient data being stored electronically. This has led to them becoming vulnerable to cyberattacks and hacking, with the potential loss and theft of critical data. The demand for health data in the illegal, pirated, or contraband market makes health entities a moneymaking target for criminals [40,41]. Internal susceptibilities in hospital systems, for example, can be abused to seize data by both internal sabotage and outside attackers. Typical health data breaches include loss, theft, unauthorized access, and hacking incidents, which are associated with errors or negligence on the part of employees who handle data, or intentional attacks by outsiders [3]. Entities with critical weaknesses in their cybersecurity initiatives face cyberattacks and data breaches [4,9]. Cybersecurity "consists largely of defensive methods used to detect and thwart would-be intruders" [42]. Currently, it is the most promising countermeasure to hacking or cyberattacks. Another definition of cybersecurity states that "cyber security entails the safeguarding of computer networks and the information they contain from penetration and from malicious damage or disruption [43]." According to [44]:

> "*The activity process, ability or capability, or state whereby information and communication contained therein are protected from and/or defend against damage, unauthorized use or modification, or explanation.*"

Companies must plan for data breaches to prevent damage and improve their proactive image. However, developing efficient, thorough countermeasures remains a challenge. Discovering and understanding the process and patterns of data breaches in the healthcare industry is core to developing countermeasures. Research on features and techniques of high-risk healthcare data breaches is imperative [6,37,38].

The main purpose of this applied research is to identify attack patterns of healthcare data breaches. More importantly, the study seeks takeaways on how to address the problem. This study is novel in several ways. First, it utilizes data from a reliable U.S. federal government source, namely the Department of Health and Human Services Office for Civil Rights (OCR). Second, the study involves the application of visualization and visual analytical techniques and tools [45–47] to make informed decisions on predicting data breaches. Choi et al., 2019, used the same dataset as is used in this study in addition to the Privacy Rights Clearinghouse (PRC) database, but the focus of their study was on remediation efforts and their effect on quality [48]. In another study of the same data, a narrative description of the breaches was undertaken [15]. Therefore, this study attempts to fill the gap in the research on health data breaches. This study develops visual charts to analyze patterns and find features of health data breaches. The rest of the paper is organized as follows. We discuss the methods used in Section 2. This is followed by an analysis of the results in Section 3, and a discussion of the implications in Section 4. The scope and limitations of the study are discussed in Section 5. Finally, our conclusions and future research directions are offered in Section 6.

2. Materials and Methods

2.1. Visual Analytics

We utilize visual analytics, which is the method of studying data with visualization, to conduct descriptive analytics [49–54] to shed light on the nature and dimensions of health data breaches. To this end, this approach is data-driven and analyzes the data as they are. To reiterate, we used data from the U.S. Department of Health and Human Services. Visual analytics facilitates the effective analysis and understanding of big datasets in real-time [55,56]. By combining the visualization features of tools, such as Tableau, with an analyst's expertise in conducting analytics, visual analytics enables the exploration of unforeseen and hidden patterns to gain insight to make informed decisions [57,58]. As is said, a "picture is worth a thousand words" and, therefore, visualization synthesizes the dimensions and measures of the data into elegant charts that display the results. The depth and variety of charts collectively tell a story about the data [55,56,58]. The objective is storytelling through a primary pillar of analytics, namely, visualization [49,55,57]. Compared to other models, descriptive analytics tends to be more data-driven, focusing on describing the data "as is" with no a priori assumptions, thereby letting the data reveal themselves. It promotes the comprehension of past and current patterns and trends that can be utilized for informed decision-making [49,50,54,56,57]. Information is represented pictorially through a multitude of charts and by using the functions of aggregation, categorization, and characterization [45,55,56].

2.2. Data

Health breach data were gathered from the U.S. Department of Health and Human Services OCR. The OCR records the nation's state-level healthcare breach activities. This database is limited to breaches originating in the U.S. Although this may be a limitation, the data are of sufficient breadth and depth to warrant generalizability of the results. The research was performed on all data in the database, referring to variables, such as state, covered entity type, affected individuals, breach type, and entity type. The methodology includes data collection, variable selection, data presentation, analytics platform selection, tool selection, and analytics implementation. Raw data were extracted in the .xlsx format. Most variables were categorical; affected individuals were numeric. Data were processed in a standard, readable format for analytics after the data were normalized. Data normalization is a preprocessing necessity to scale/transform all data to be a value between 0 and 1, making the variables (dimensions) comparable. The significance of data normalization in visualization cannot be overemphasized to enhance the quality and robustness of the analysis [57].

The data were then loaded in Tableau to identify patterns and trends. Tableau is an extremely powerful tool for visualizing massive sets of data very easily. It has an easy-to-use drag-and-drop interface [45]. The research approach included the ranking, association, and data visualization of healthcare breach data. The research focused on finding facts about breach distribution due to the number of variables recorded for each state and entity. This visualization research contained all variables. It aimed to cover all correlations and patterns. This study analyzed the following aspects of the data: (1) breach-type analysis, (2) breach geographical analysis, and (3) breach organizational analysis; see Table 1.

Table 1. Breach data source from the U.S. Department of Health and Human Services OCR.

Key Variables	Description
State	
Covered entity type	Health plan, healthcare clearing house, healthcare provider
Individuals affected	Number of individuals affected by the breach
Breach submission date	Date of submission of the breach
Type of breach	Hack/IT incident, improper disposal, loss, theft, unauthorized access/disclosure, unknown, or other
Location of breach	Desktop computer, electronic medical record, e-mail, laptop, network server, other portable electronic device, paper/films, other
Business associate present	If a business associate was present or not
Web description	Description of the website of the breach

Source: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf. (accessed on 16 December 2022).

3. Results

Using visualization, we developed a series of charts to understand the health breach data. Collectively, the charts tell a compelling story about the nature and dimensions of health data breaches. First, we examined breach type and the number of individuals affected. Figure 1 shows the total number of individuals affected by each type of breach. One of the most significant breach types is a hacking/IT incident. Unauthorized access/disclosure and theft also impact many individuals. About 300,000 people were impacted by the breaches. A smaller amount of data was involved in the loss. However, loss causes severe problems because it cannot be recovered. Hacking, when compared to other types of breaches, accounted for more than twice the number of cases. This indicates that organizations should focus on hacking as an essential entrance for a data breach.

As show in Figure 2, we examined the distribution and range of affected individuals by type of breach. Hacking has the highest mean and number of individuals among types. The number of records for the improper disposal breach type shows no outlier. This suggests that the size of the breach caused by improper disposal is controllable and rarely results in extreme situations. Records of hacking/IT incidents, unauthorized access, and theft have many outlier records. Therefore, planning for these types of breaches is vital to avoid extreme situations impacting many individuals.

It is necessary to study where and how breaches occur. Analyzing the locations of breach types will offer insights into the prevention of future breaches. Figure 3 shows the number of affected individuals in each breach location. A location is any of the computing devices or other sources in which a breach occurs. In Figure 3, the riskiest location is the network server. More than 4 million people are involved in data breaches. This number is almost four times that of the e-mail factor. Figure 4 shows the number of records for each breach location. The information was filtered for the locations with less than 5000 affected individuals.

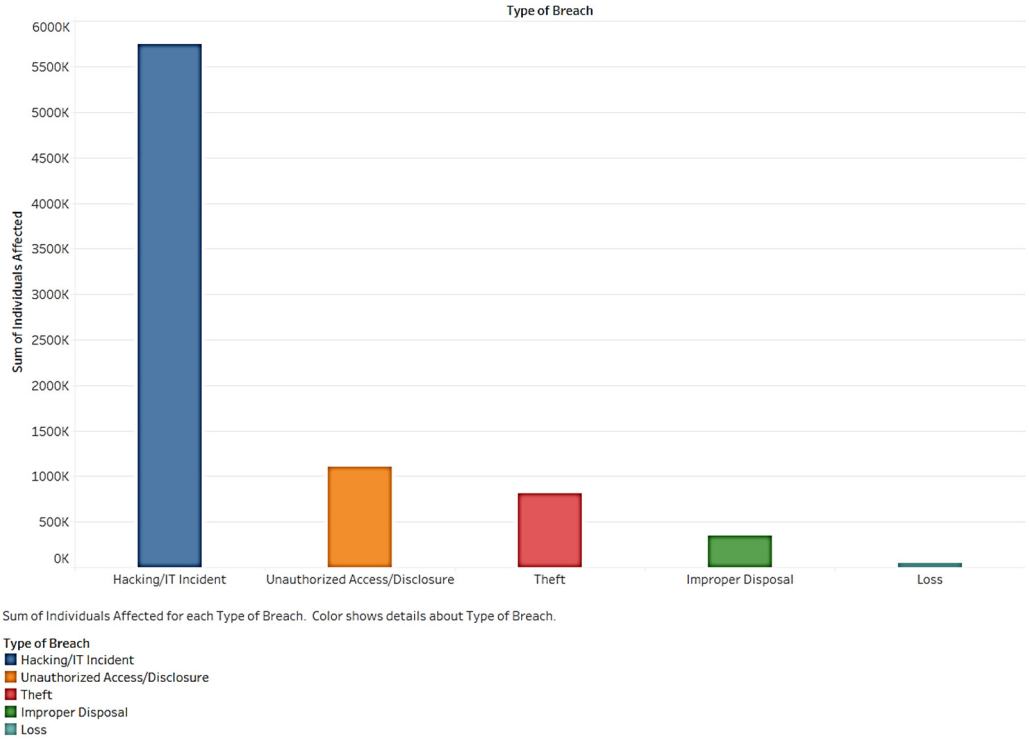


Figure 1. Distribution of breach type and the individuals affected.

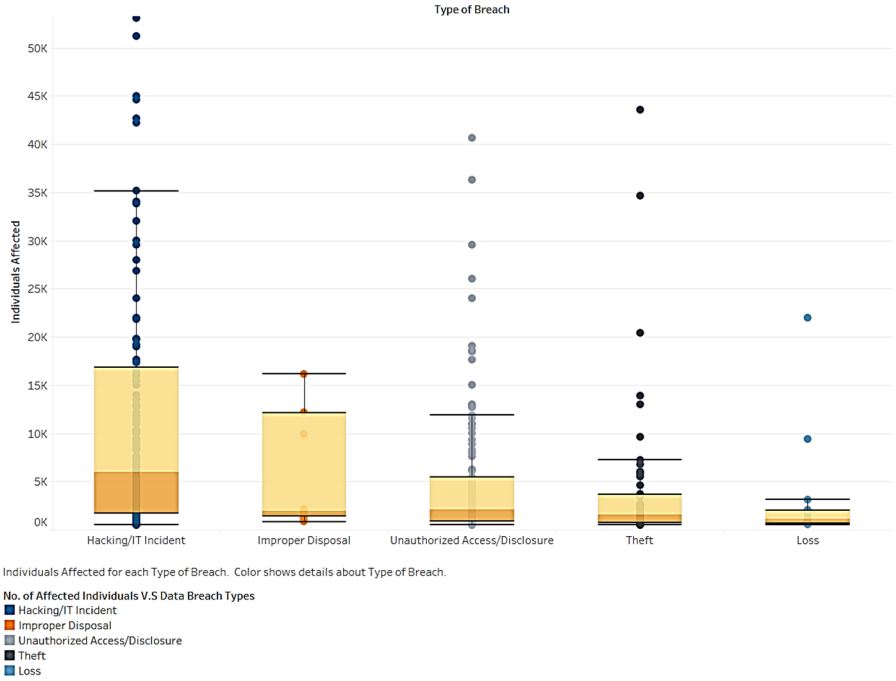


Figure 2. Analysis of affected individuals vs. breach type.

In Figure 5, there are more records for e-mail than network servers. Most breaches occur via network servers, e-mail, and papers/films, as these are the top three in terms of the number of records and number of affected individuals. Figure 5 shows the quantity of affected individuals in different places. The bar is colored by type of breach. The network server suffers from hacking/IT incidents and unauthorized access. Most improper disposal incidents occur with papers/films, and rarely exist in other locations. Unauthorized access/disclosure happens through network servers, e-mail, papers/films, electronic medical records, and other locations. The network server is the easiest path to

leak data. Business organizations in the healthcare field should take more precautionary actions regarding this path.

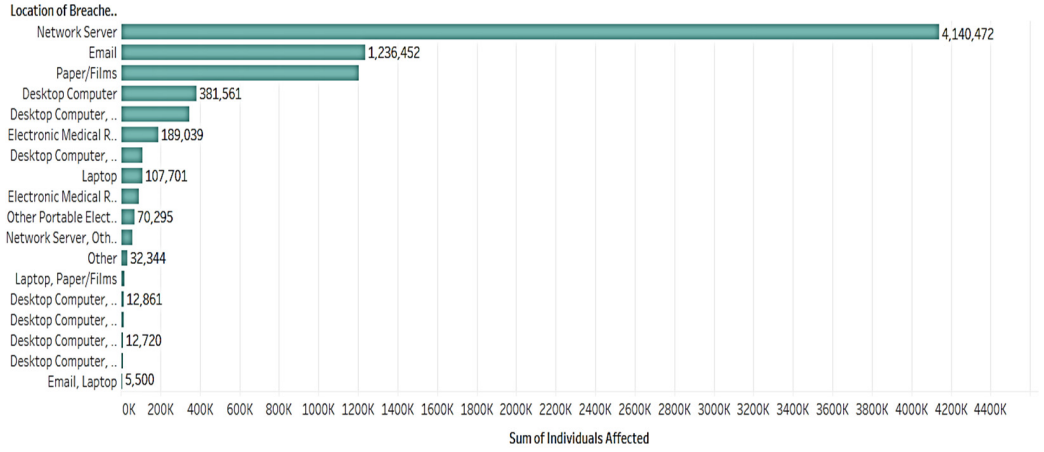


Figure 3. Data breach location vs. number of affected individuals.

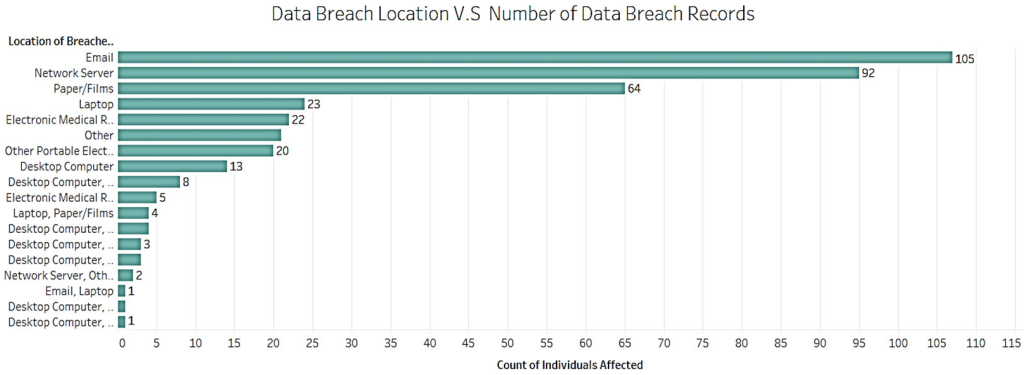


Figure 4. Data breach location vs. number of data breach records.

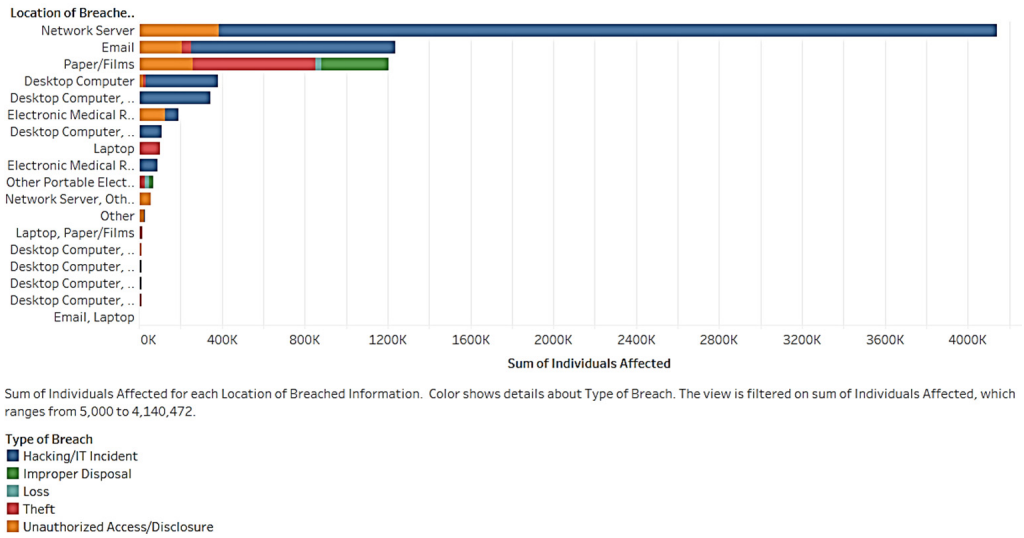


Figure 5. Data breach location vs. number of affected individuals by breach type.

Next, we compared the locations of breached information with breach types. As shown in Figure 6, the 5 most related locations in hacking/IT incidents are e-mail (73), network servers (74), desktop computers (9), desktop computers and network servers together (8), and electronic medical records (5). Locations of unauthorized access included e-mail (33), network servers (21), papers/films (42), electronic medical records (18), and others (17). The remaining three breach types did not have enough related locations and, thus, did not satisfy the condition. For example, theft happened through papers/films (10), laptops (22), and other portable electronic devices (9). On the other hand, loss occurred with papers/films (6) and other portable electronic devices (9), and improper disposal occurred with papers/films (7). Papers/films appeared with four breach types, whereas e-mails contained the most records. This means that the papers/films category is the easiest manner for leaking information. E-mail reveals the most information among all the locations. E-mail, network servers, and papers/films significantly influence breach types.

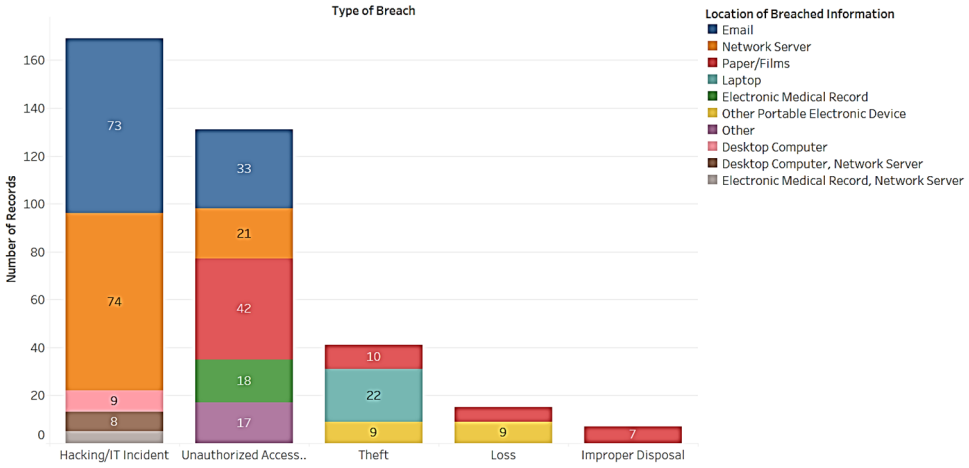


Figure 6. Location distribution in terms of breach type.

Next, we looked at the data breach through the various entity types, such as health plan, healthcare provider, etc. The bubble chart in Figure 7 shows the distribution of individuals and records affected by the different entities. The bubble with 813,207 datapoints belongs to the business associate category. The bubble with 1,159,715 datapoints belongs to the health plan category. The largest bubble belongs to the healthcare provider category. This chart illustrates specific problems. For example, the number of affected individuals per covered entity type is considerable. However, healthcare providers affect over 6 million individuals. This insight demands further attention. As expected, healthcare providers are significantly related to impacted individuals.

For further insight, we studied the number of affected individuals for each of the five data breach types and the three entities (see Figure 8). The stacked bar chart shows the secondary distribution of individuals affected by the covered entity under breach type. According to the chart, the healthcare provider category dominates, as it has the most individuals affected by a breach. Therefore, it can be concluded that healthcare providers are the most vulnerable entity for almost all the breach types. This is followed by theft, which significantly affects health plans. The business associate category is affected by hacking and unauthorized access. It must be emphasized that healthcare providers should protect themselves from all types of breaches, especially hacking. The health plan organizations should also focus on the first three types of breaches, and business associates should be concerned with the first two types.

Figure 7. Distribution of affected individuals in entity types vs. data breach records distribution of entity types.

We also shed light on the presence of business associates as a core subject in understanding data breaches and entities. The business associate category presents various behaviors for each type of data breach (see Figure 9). The stacked bar chart shows the secondary distribution of individuals affected by business associates regarding type of breach. The bars show that business associates are correlated with hacking and unauthorized access. However, most affected individuals are not related to the business associates.

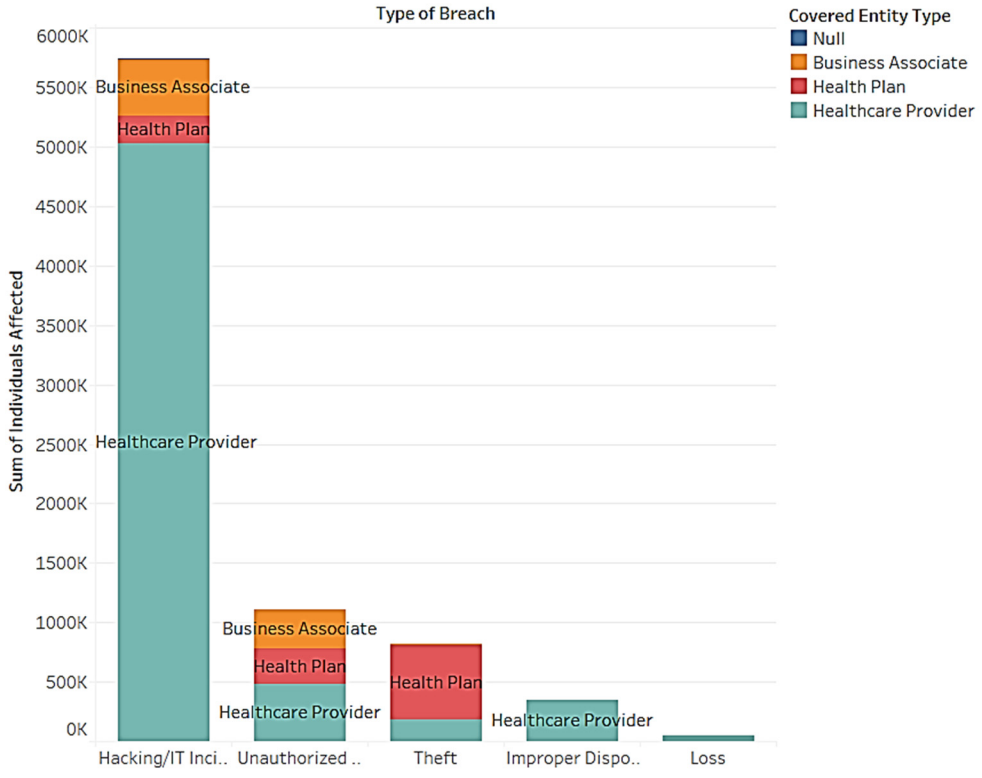


Figure 8. Distribution of affected individuals in data breach types and covered entity types.

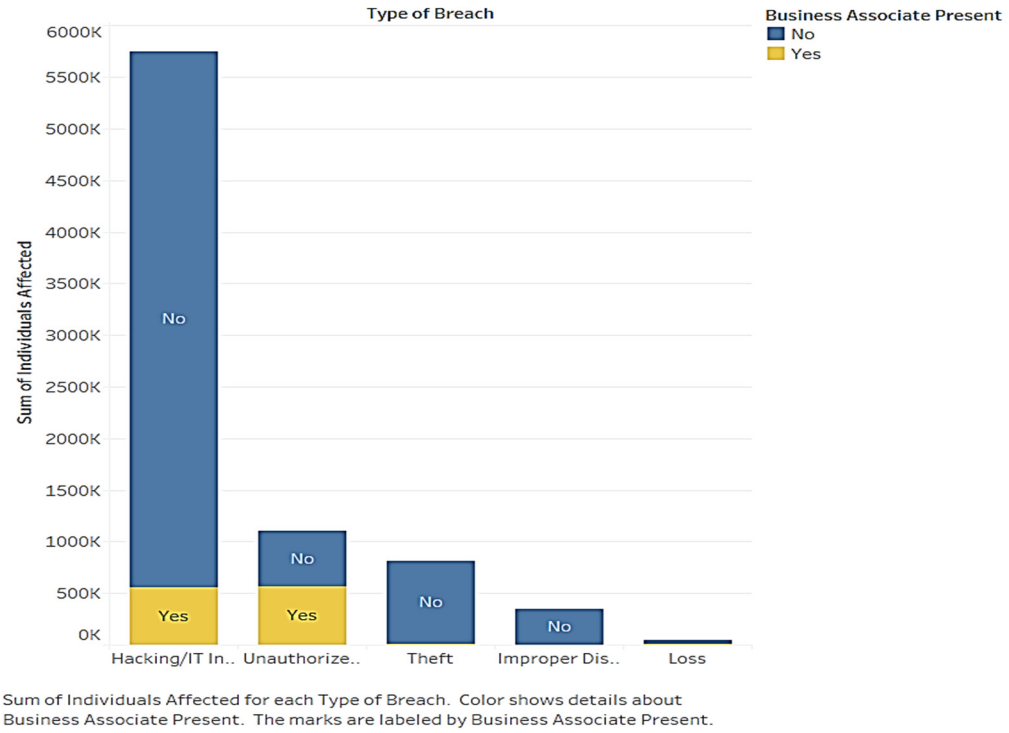


Figure 9. Breach type analysis by number of affected individuals and presence of business associates.

Research on affected individuals in data breaches and breach records depicts a developing trend in breaches over time. Figure 10 shows the trends in the number of records. Figure 11 shows the number of affected individuals. Although it varies, there is a near constant increasing trend displayed in Figure 10. Although the month with the most reported records was April 2018, the month with the most individuals involved in data breaches was May (1,413,466 individuals). The surge in breaches in April–May warrant additional research, including looking at anecdotal evidence.

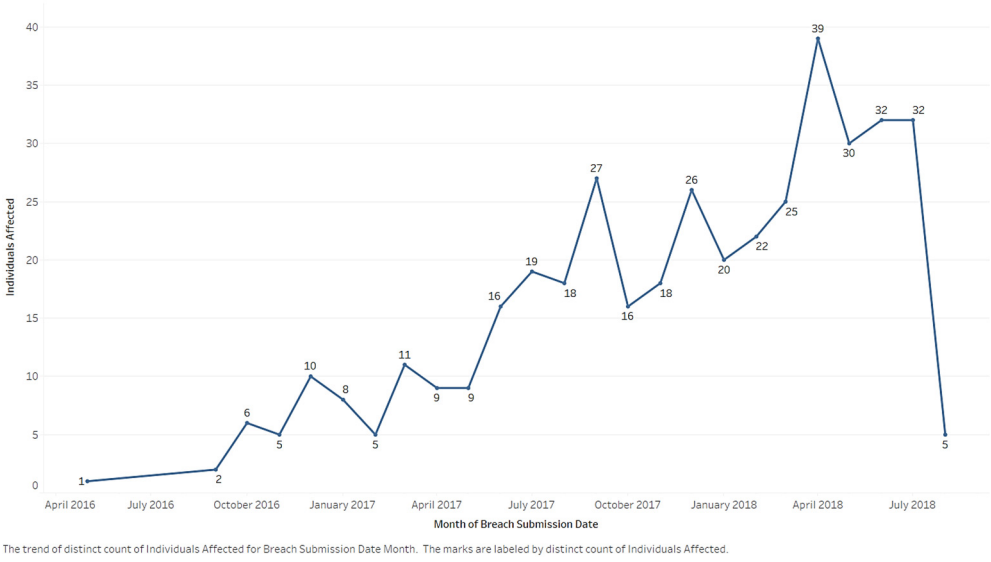


Figure 10. Number of breach records trend.

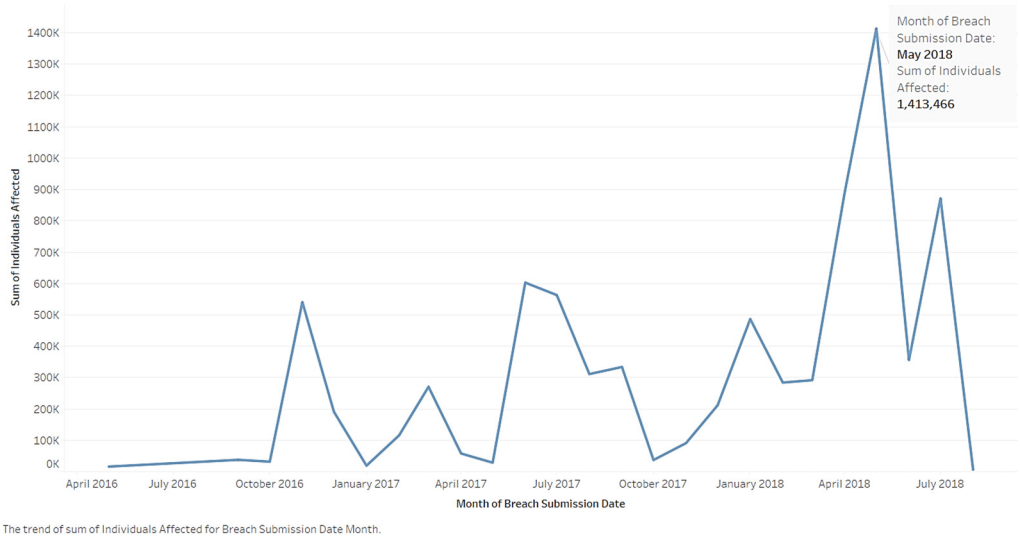


Figure 11. Trend of affected individuals by type of breach.

Figure 12 illustrates an analysis of the trend distinguished by breach type. Each of the five lines represent a specific breach type. The line above all the others represents hacking/IT incidents. The line below it represents unauthorized access/disclosure. Peak values for the breaches occurred in the last year. Hacking/IT incidents, improper disposal, theft, and unauthorized access/disclosure show an increasing trend. However, there is a weaker increasing trend related to hacking/IT incidents. Therefore, the other breach types warrant attention. Hacking/IT incidents, improper disposal, theft, and unauthorized access/disclosure show an increasing trend. Loss shows a declining trend in regard to its impact on individuals.

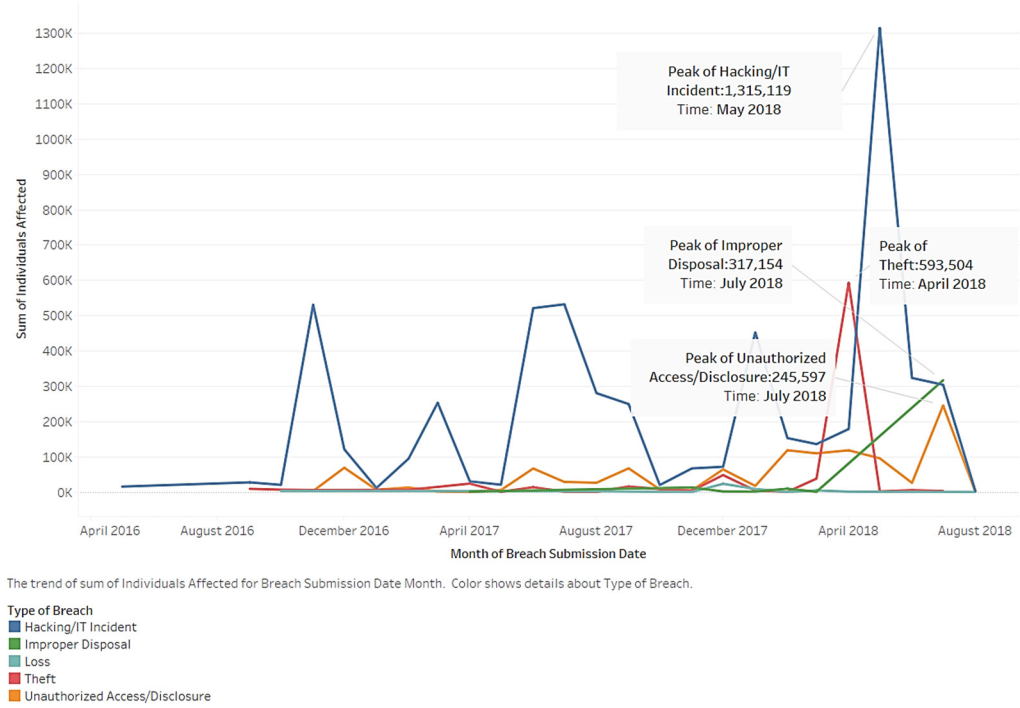


Figure 12. Trend of affected individuals by breach and breach type.

The fluctuation in affected individuals by type is interesting. However, although improper disposal remained stable, there was a sudden rise. Due to this abnormal trend, related companies and entities should pay attention to the overall data and focus on continually monitoring breaches.

Figure 13 ranks the types of breaches, trends in time, and accumulated values. The ranking of types remained the same for both the number of records and affected individuals. The accumulated values increased over time. However, they occurred at different speeds. Hacking ranked first. Its increasing speed slowed over time. Hacking's record is smaller in 2018 compared to other times. However, it still has the largest affected record of individuals, including the accumulation of other types.

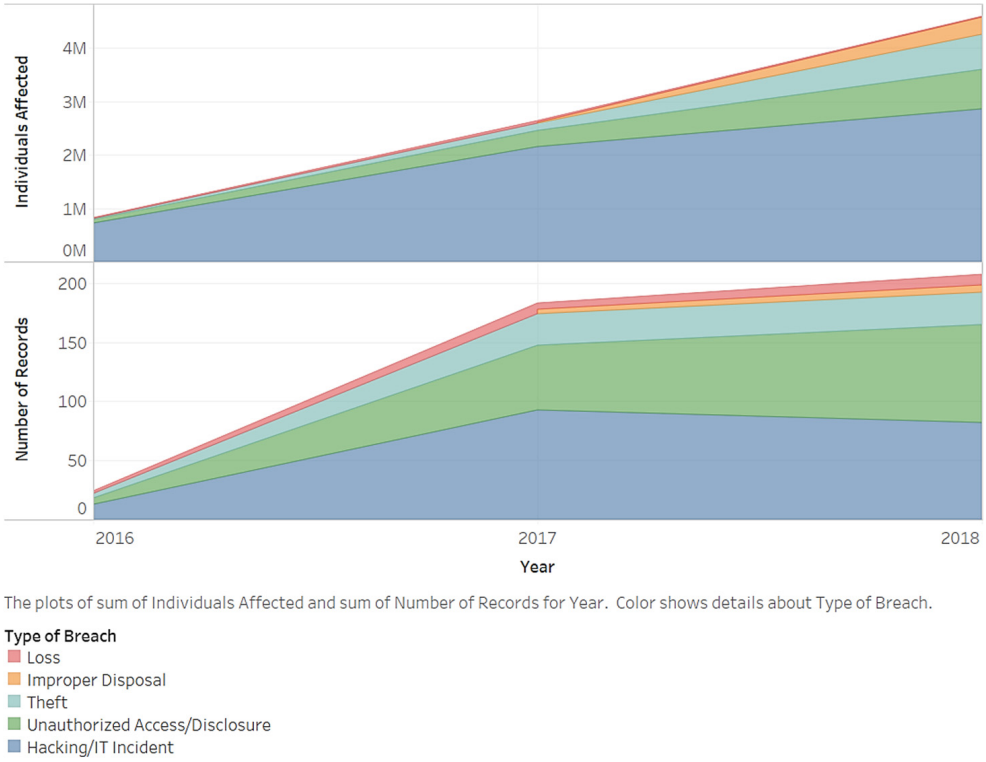
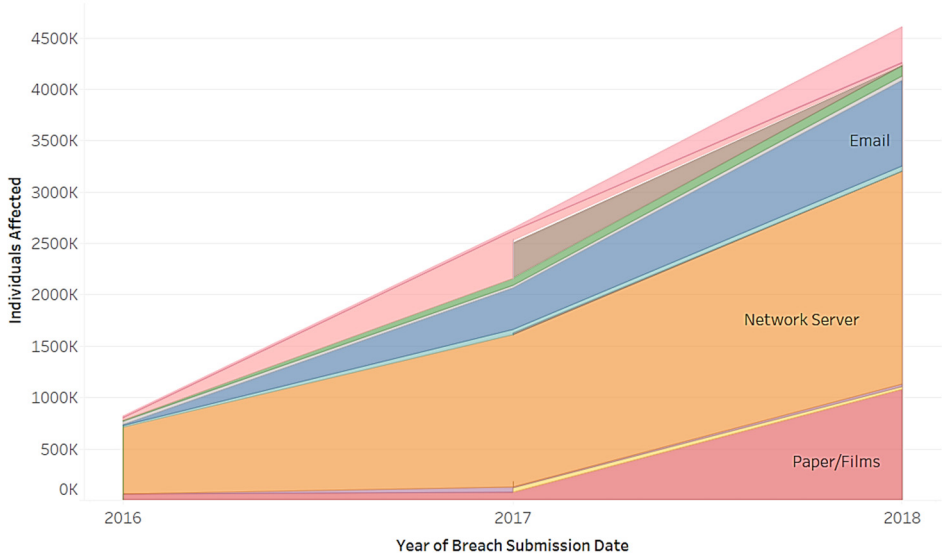


Figure 13. Accumulated records trend vs. accumulated affected number of individuals trend (colored by data breach type).

We also analyzed affected individuals based on locations of data breaches over time. We assumed that some locations deserved more attention because breach conditions continued to evolve.

Figure 14 depicts the trends in individuals affected by the location of breached information by time. Common locations of breached information include e-mail, network servers, and papers/films. The use of papers/films increased sharply after 2017; network servers played a large role at all times. In addition, desktop computers and network servers had sudden increases (brown color). Additional attention should be paid to the film industry due to its popularity and absorption of individuals. Affected individuals increased in every location. This shows that issues related to identity theft are becoming more serious.

Figure 15 shows the number of individuals affected by month (colors represent location). The chart shows common locations regarding individuals involved in a data breach. Others are filtered. The use of networks increased sharply in May 2018. In addition, desktop computers and network servers experienced a sudden increase. Peak values appeared in 2018. The film industry gained popularity, absorbing more people over time. Therefore, this location deserves additional attention. Nearly all affected individuals showed an increase in every location. This shows that identity theft issues are becoming more serious.



The plot of sum of Individuals Affected for Breach Submission Date Year. Color shows details about Location of Breached Information. The marks are labeled by Location of Breached Information.

Location of Breached Information

- Desktop Computer
- Desktop Computer, Electronic Medical Record
- Desktop Computer, Electronic Medical Record, Laptop
- Desktop Computer, Electronic Medical Record, Network Server
- Desktop Computer, Electronic Medical Record, Network Server, Other Portable Electronic Device, Paper/Films
- Desktop Computer, Email
- Desktop Computer, Email, Network Server
- Desktop Computer, Laptop
- Desktop Computer, Laptop, Network Server
- Desktop Computer, Network Server
- Desktop Computer, Other
- Desktop Computer, Other, Other Portable Electronic Device, Paper/Films
- Desktop Computer, Paper/Films
- Electronic Medical Record
- Electronic Medical Record, Email, Laptop
- Electronic Medical Record, Network Server
- Email
- Email, Laptop
- Email, Network Server
- Email, Other
- Email, Paper/Films
- Laptop
- Laptop, Network Server
- Laptop, Other Portable Electronic Device
- Laptop, Paper/Films
- Network Server
- Network Server, Other
- Other
- Other Portable Electronic Device
- Other, Paper/Films
- Paper/Films

Figure 14. Trends related to the location of breached information.

Next, we plotted the breach trends of the covered entity types by month (see Figure 16). This area chart depicts the trends in individuals affected by the covered entity types. Healthcare providers, as compared to the other entities, continue to have a high volume of affected individuals. The three entities have a peak volume of affected individuals in different months throughout 2018. Overall, there is an increasing three-year trend for all three types. However, the numbers fluctuate each month. Beginning in April 2018, the number of affected health plan individuals maintained a small horizontal trend. Some changes, whether intended or coincidental, controlled the data breach. Additional investigations and monitoring need to be carried out for this trend. Breach accidents affected healthcare providers more compared to the other entities. Therefore, healthcare providers should focus on data breaches.

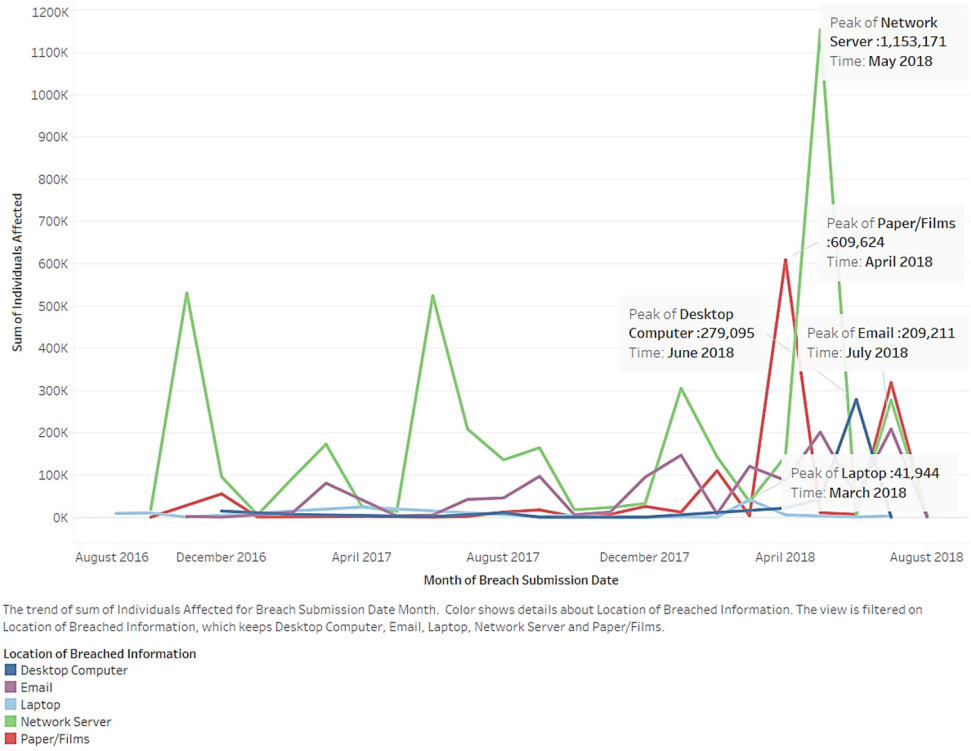


Figure 15. Trends related to affected individuals by breach and location of the breached information.

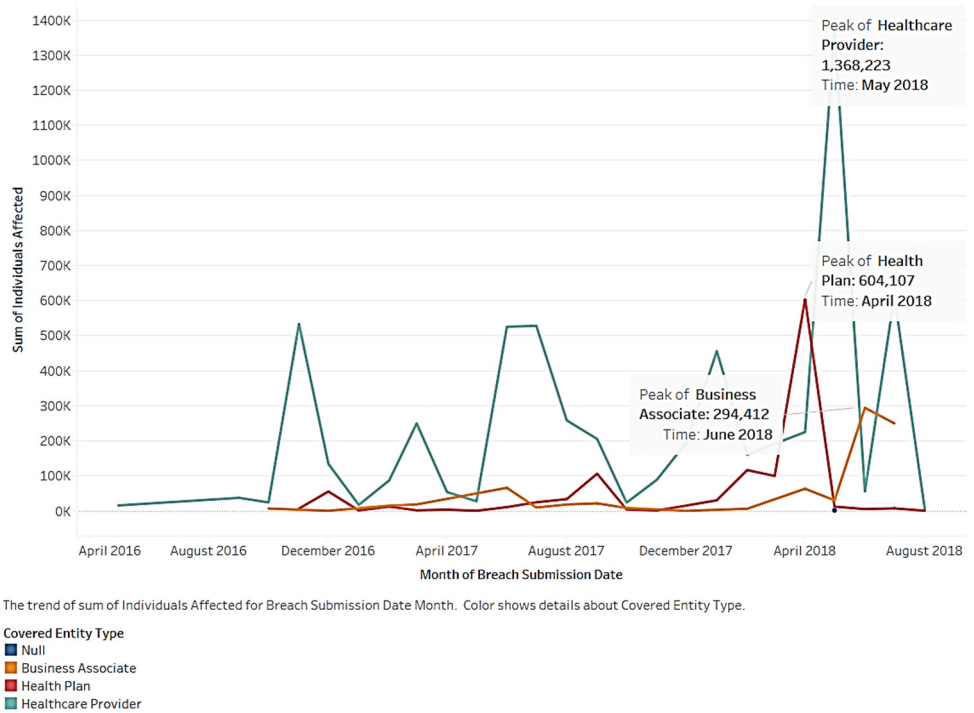


Figure 16. Trend of individuals affected by breach and covered entity type.

The relationship between a business associate being present and time is another important measurement. The area chart in Figure 17 depicts the trends in individuals affected by business associates being present. Nonbusiness associates consistently affect a high volume of individuals as compared to business associates. The two lines peak in 2018. After 2018, the business associate category shows an increasing trend; the nonbusiness associate category also displays a significant increase in the affected individuals.

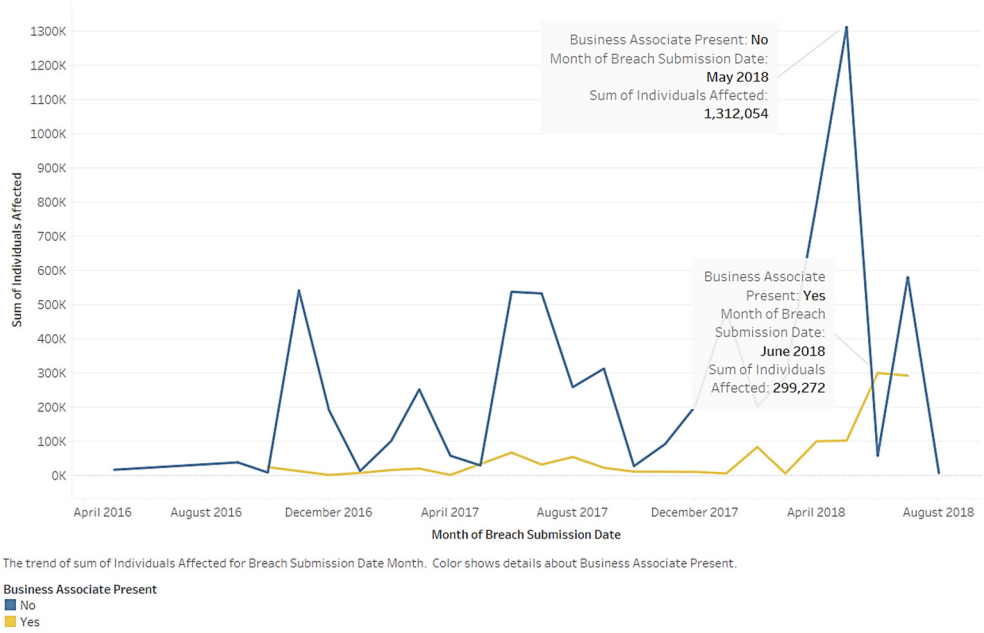


Figure 17. Trend of individuals affected by breach and business associate presence.

Here, we provide an overview of the geographic distribution of breaches by state. Figure 18 shows the number of submitted records per state. Darker colors represent more records. It is obvious that data breaches happen most frequently in California (CA) and Texas (TX). There are 23 records each in New York (NY) and Illinois (IL). These states have large populations (see Figure 19). States with dense populations (except TX and IL) may experience more risk. Companies in those states should pay more attention to protective measures. Figure 18 shows that most affected individuals are located on the nation's east coast.

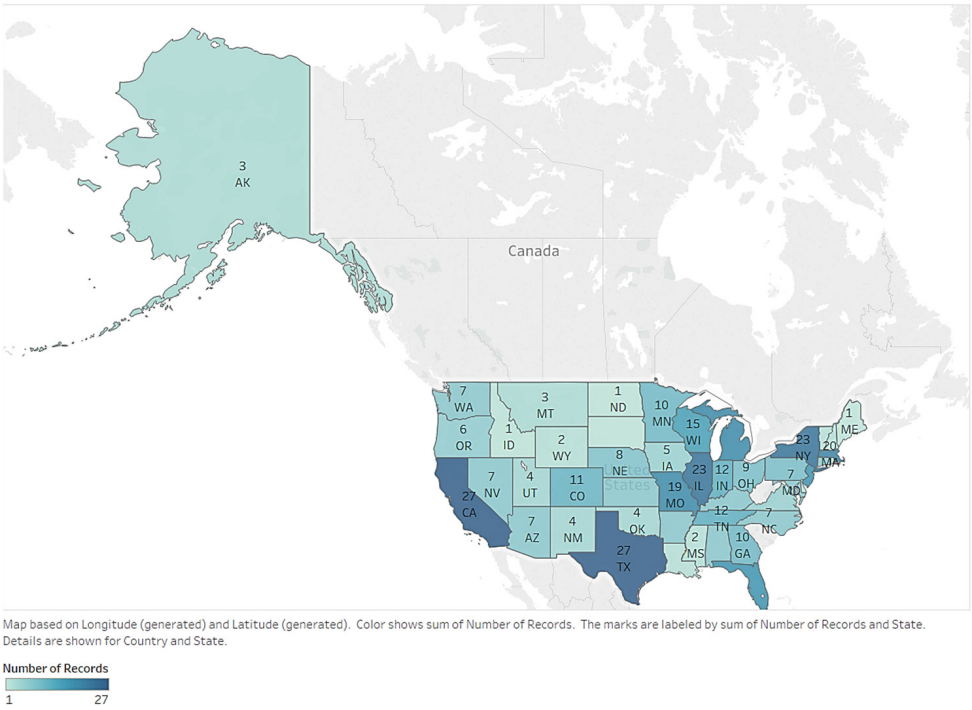


Figure 18. Number of breach records per state.

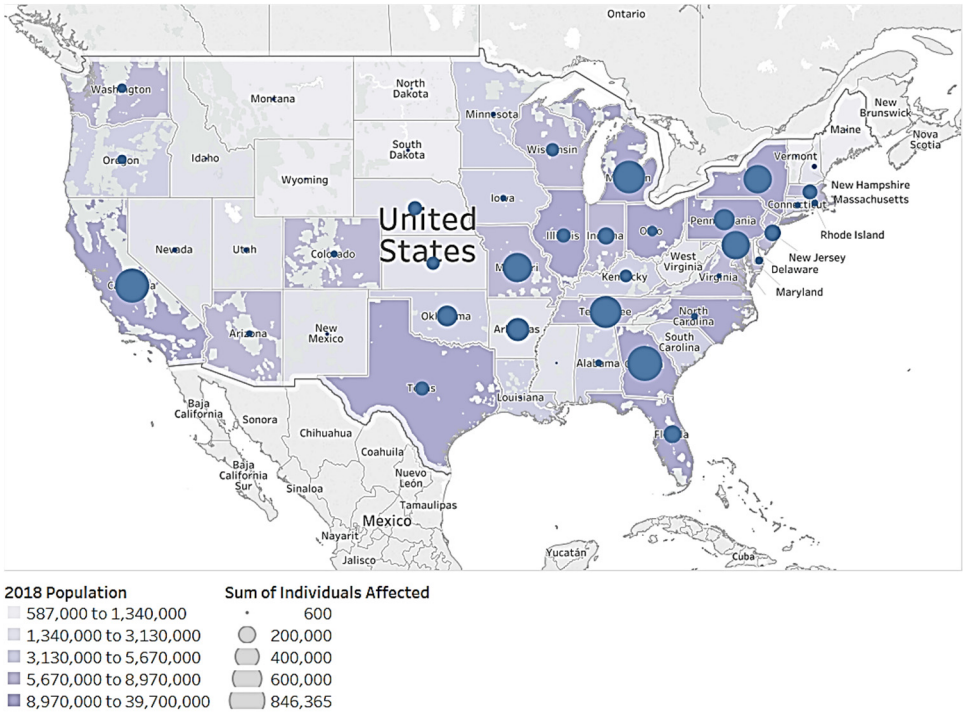


Figure 19. Number of affected individuals and state population.

Figure 20 describes the distribution of healthcare breaches by state. The different colors show the number of records. Darker shades represent a higher number of records. The pie chart details the type of breach. First, healthcare breaches are clustered in the most populous states including CA, TX, NY, and IL. Second, some states contain different types of data breaches. Other states contain a single type of breach. For example, Idaho (ID), North Dakota (ND), South Dakota (SD), Delaware (DE), and New Hampshire (NH) experience hacking/IT incidents. Louisiana (LA), Washington, DC (DC), and Maine (ME) experience theft. Most states have more than one type of data breach.

Next, we studied the distribution of individuals by breach type per state. In Figure 21, the different colors show the number of affected individuals. Darker shades represent a higher number of records. The pie chart details the breach type. The states' pie charts show that individuals affected by health breaches are clustered in the most populous states, including CA, TX, NY, and IL. Another finding is that, in some states, individuals are affected by only one type of data breach. For example, ID, ND, SD, DE, and NH have hacking/IT incidents. LA, DC, and ME have theft, which affects individuals. States suffer from different types of breaches. Therefore, a variety of countermeasures are required. States with only one breach type should consider their current problem and risks related to other breach types.

To better understand distribution, we used pie charts to illustrate location type and regional breaches (see Figure 22). The study found that most states have one dominant location impacting affected individuals. For example, in Washington (WA), Montana (MT), Idaho (ID), Oregon (OR), Iowa (IA), Ohio (OH), and Vermont (VT), disclosure occurs via e-mail. In North Dakota (ND), Oklahoma (OK), South Dakota (SD), Arkansas (AR), Tennessee (TN), Alabama (AL), Georgia (GA), New Jersey (NJ), and NY, data breaches occur through network servers. Some states had multiple locations for breaches. We further created a bar chart that describes the location and distribution of affected individuals in the top 10 affected states (see Figure 23). There are 18 locations regarding breached information and network servers. Papers/films make up many of the breaches in CA and Missouri (MO). The use of desktop computers impacts half of NY's individuals. Pennsylvania (PA) is mostly affected by desktop computers and network servers. All states should pay attention to network servers. CA and MO should focus on papers/films. NY should focus on desktop

computers. PA should focus on desktop computers and network servers. Network servers are a common trend regarding location and breached information.

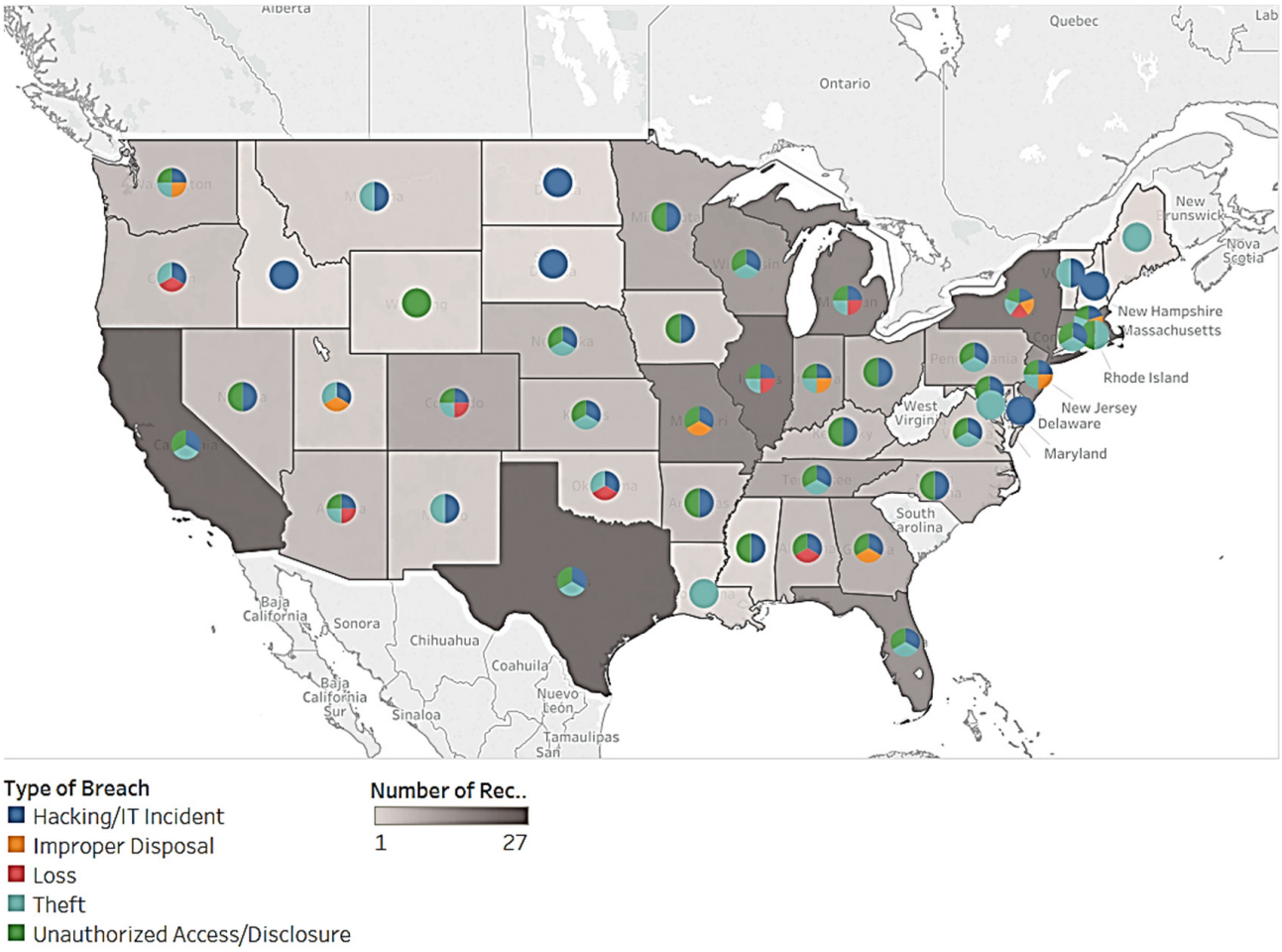
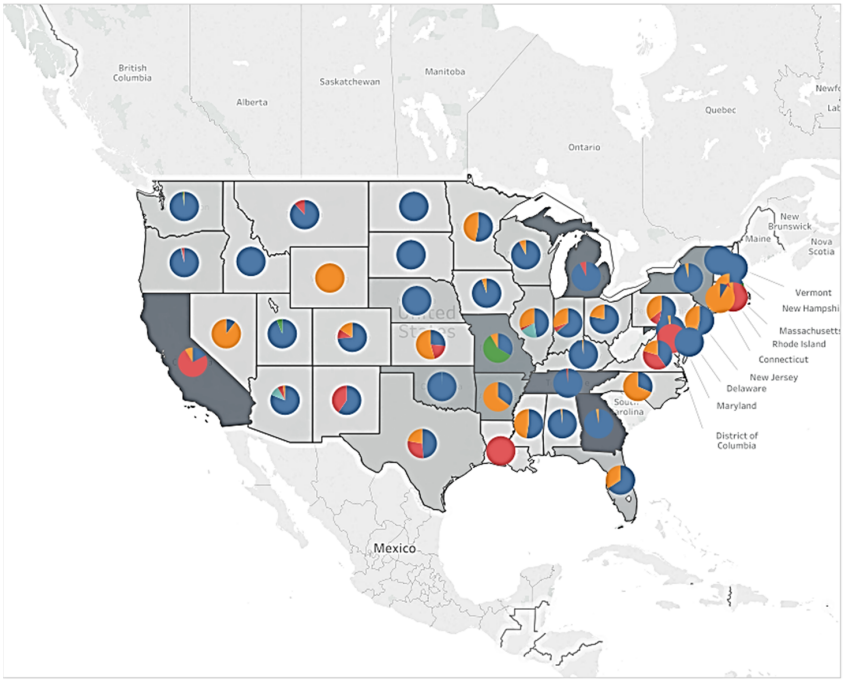


Figure 20. Number of healthcare data breaches by state.

In Figure 24, the different colors show the number of affected individuals. Darker shades signify a higher number of records. The heat map details the covered entity type. For affected states (i.e., CA, TN, SC, and MO), most show that healthcare providers dominate. Only CA has health plans as the majority type. The business associate entity was randomly distributed throughout the states. Most states affected by data breaches listed the healthcare provider entity. States on the west and east coasts listed the business associate entity. This may be due to the improved economic situations along the coasts as compared to the middle of the country.

Lastly, we looked at the geographical distribution of affected individuals due to business associates being present. The colors in Figure 25 represent the number of affected individuals. Darker shades indicate a higher number of records. The pie charts within the heat map detail the presence of business associates. We found that affected states, such as CA, TN, SC, and MO, were dominated by the presence of a nonbusiness associate. Regarding healthcare data breach accidents, business associates were less vulnerable than nonbusiness associates in all U.S. states. This means that the two features are not significantly relevant.

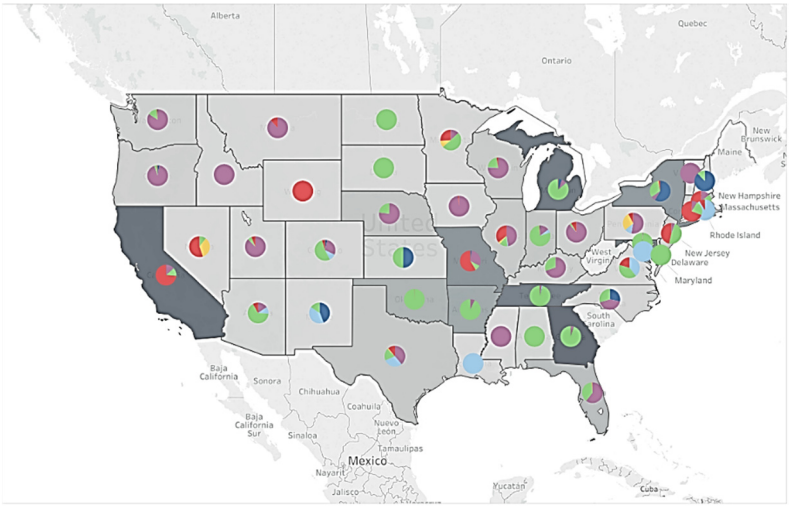


Map based on Longitude (generated) and Latitude (generated) and Latitude (generated). Details are shown for Country and State. For pane Latitude (generated): Color shows sum of Individuals Affected. For pane Latitude (generated) (2): Color shows details about Type of Breach. The data is filtered on Location of Breached Information, which keeps 6 of 31 members. The view is filtered on Exclusions (Country,State), which keeps 48 members.

Type of Breach
- Hacking/IT Incident
- Improper Disposal
- Loss
- Theft
- Unauthorized Access/Disclosure

Sum of Individuals..
600 834,207

Figure 21. Geographical distribution of affected individuals by state and type of breach.



Map based on Longitude (generated) and Latitude (generated) and Latitude (generated). Details are shown for Country and State. For pane Latitude (generated): Color shows sum of Individuals Affected. For pane Latitude (generated) (2): Color shows details about Location of Breached Information. The view is filtered on Exclusions (Country,State) and Location of Breached Information. The Exclusions (Country,State) filter keeps 48 members. The Location of Breached Information filter keeps 6 of 31 members.

Location of Breached Information
- Desktop Computer
- Email
- Laptop
- Network Server
- Other
- Paper/Films

Sum of Individuals..
600 834,207

Figure 22. Geographical distribution of affected individuals by state and location of breached information.

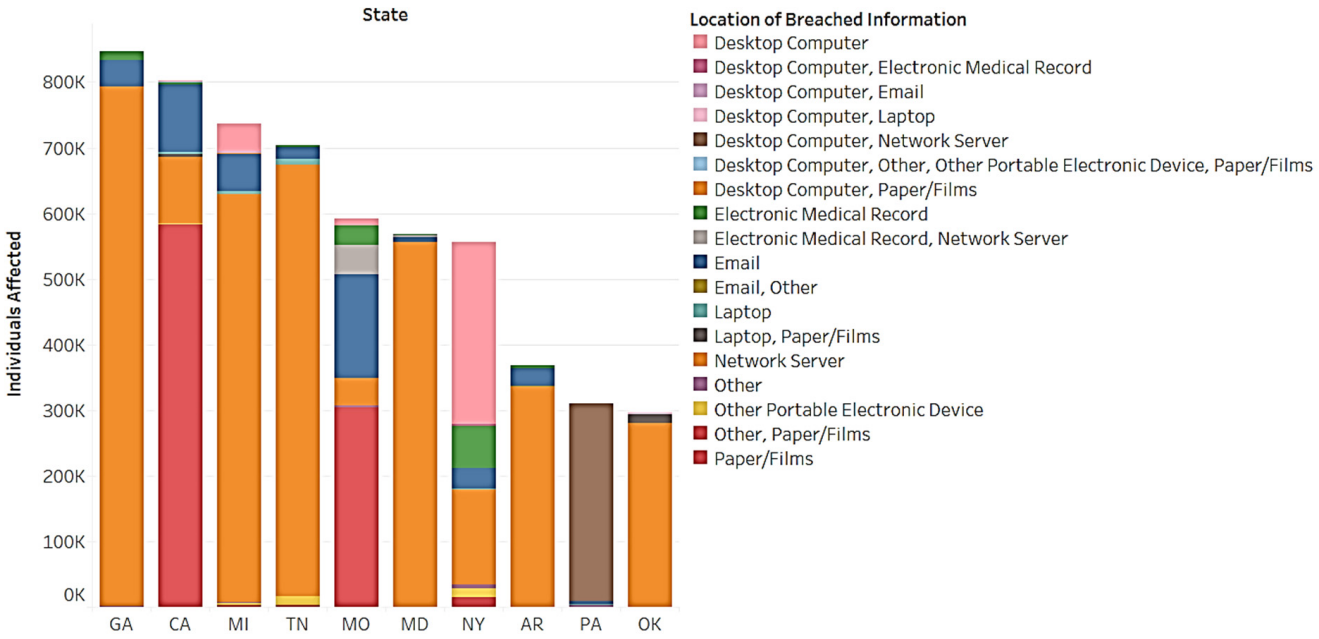


Figure 23. Distribution of affected individuals' breach location in selected states.

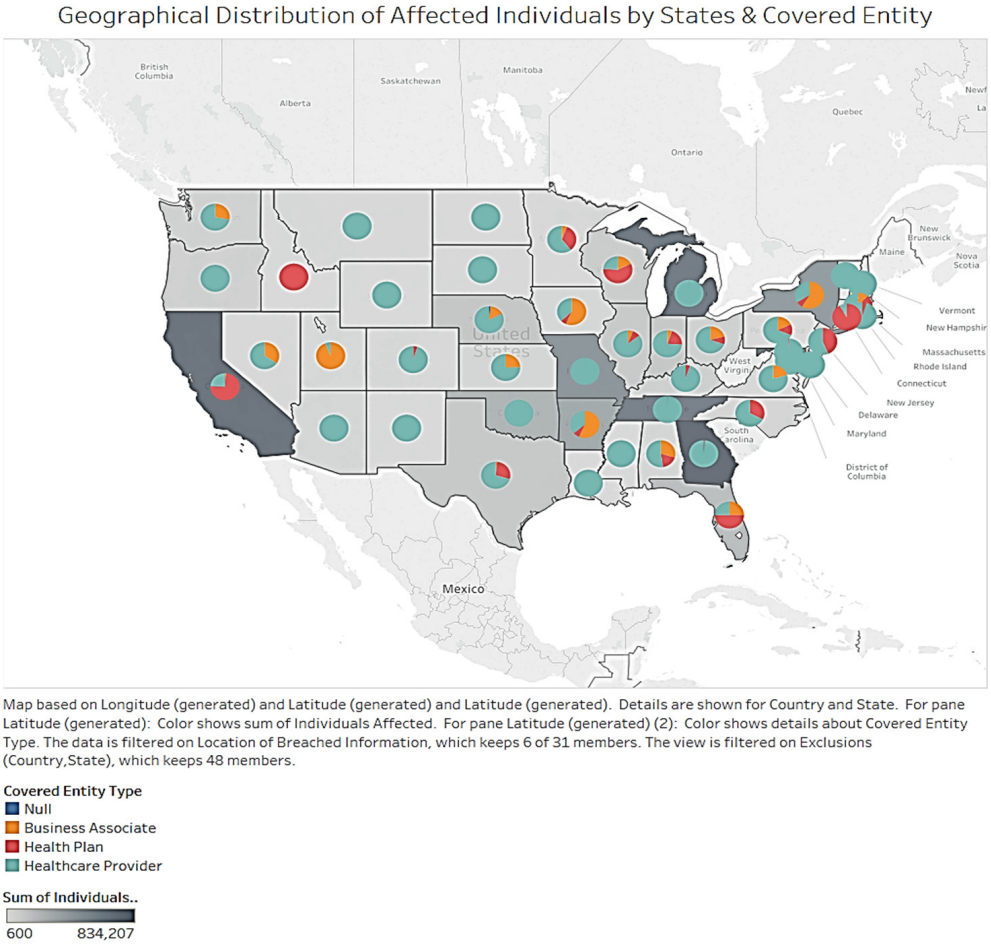


Figure 24. Geographical distribution by the entity type.

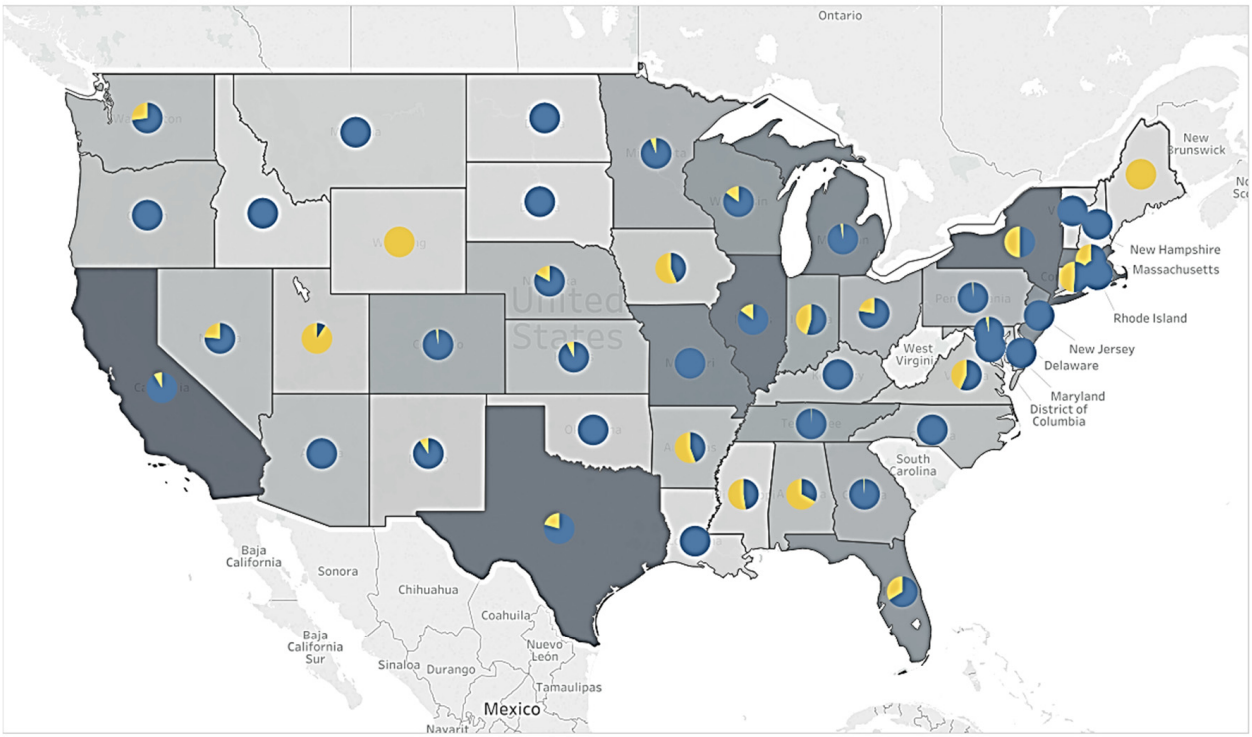


Figure 25. Geographical distribution of affected individuals' business associate presence.

4. Discussion and Implications

This study visualized the various organizational dimensions to examine their association with healthcare data breaches. Overall, our results paint a mixed picture. There is no significant association between states as a variable and the type of breach, implying that the different breaches occur across a spectrum of states. With the rapid and significant advances in Internet and web technologies, hacking- and information-technology-related incidents are on the rise in causing data breaches. Identity theft is becoming more common, thereby affecting individuals on a larger scale. Regarding covered entity type, data breaches are most likely to occur among healthcare providers. As far as location is concerned, the network server is the most likely source of a breach. Lastly, the number of victims is constantly shifting across the various states. Therefore, in addition to businesses and the federal government taking proactive mitigating and preventive measures, state and local governments must also develop policies and procedures and implement appropriate steps about health data breaches. Once a breach occurs, the affected entities must take decisive and proactive steps to immediately halt the spread of the impact of the breach and protect as many individuals as possible (shut down systems, trigger backup systems, etc.). Additionally, regulations must be formulated and enforced in the event of repeated breaches in healthcare entities. Technically, network servers (as well as web servers) should be monitored for potential breaches and hacking. Due to the relentless phishing, hacking, and other malware attacks on health information technology, this research is important to patients and health stakeholders, who are likely to fall victim to criminals and lose their healthcare data. Simultaneously, healthcare entities gathering and storing individual health data have a fiduciary and reg-

ulatory duty to protect such data and, therefore, need to be proactive in understanding the nature and dimensions of health data breaches. Additionally, information technology must be harnessed to the fullest in providing technical safeguards against data breaches. Furthermore, comprehensive training must be provided on an ongoing basis to both employees and patients about healthcare data breaches [5,13,15,48].

5. Scope and Limitations

Although this research is broad and thorough, it also has limitations. First, data availability is extremely limited. Additionally, meaningful data that can be analyzed are limited to only a few years. Furthermore, this dataset is limited to breaches occurring in the U.S. Nevertheless, we were able to analyze the available data on data breaches and derive meaningful insights. Second, the research considered a limited number of variables related to data breaches. There are possibly more correlated variables to import into the research. Third, many data breaches go undetected. Therefore, the number of records does not represent the current breach situation. Fourth, the data lack predictive capability. Therefore, only a descriptive analysis was conducted. In the future, studies may look at the time factor in spotting a breach to improve record-keeping. Fourth, many outliers exist in the dataset, but these are included in the analysis, since large data breaches need to be included. Future research may investigate a time-series analysis of a lengthier period with additional variables. Due to data limitations, this research was able to only conduct descriptive analytics with visualization. With additional data and variables, predictive modeling with statistics can be conducted. Furthermore, machine learning and text analytics can be incorporated with textual data. While descriptive analytics with visualization offers insight for informed decision-making, more advanced visualization, and visual analytics methods can be applied to health data breach data when more sophisticated and richer data becomes available. For example, 'visual data mining' involves the extraction of meaningful information with the application of heuristics and network analysis techniques [54,58,59]. Additionally, in the visual data mining process, users interact with the data and the results of their analysis, namely undertaking network-based inferencing [54,59,60]. A user can navigate through a large corpus of documents through graphs (that represent parts of text) and the relations connecting them [59,60]. The sliding treemap is another visualization technique that can present graphical structures on mobile touch devices [59,61–64]. This approach can be used to study the network effects of health data breaches. These and other advanced visual analytic methods can be explored in the future.

Furthermore, although the research focused on the available dimensions of health data breaches, it did not consider the demographic information regarding the impacted stakeholders. Information about the entity as well as the affected individual can be incorporated to ascertain if certain patterns attract more data breaches, or if certain patterns are drawn from certain categories of data breaches. Additionally, specific entity information can be included in the analysis to determine if the relationship between an entity (e.g., a healthcare provider) and an affected individual has any influence on the data breach. Future studies can explore differences in the type of entity, location, breach type, and affected individual type. Information on the insurance coverage of data breaches is another dimension that can be incorporated to evaluate the cost of data breaches and data breach litigation. Finally, our sample consists of data breaches in the U.S. As the phenomenon of health data breaches accelerates, future studies can encompass a diverse set of breaches from countries around the world.

6. Conclusions and Future Research

This study focused on the factors and dimensions of healthcare data breaches by utilizing publicly available data from the U.S. Department of Health and Human Services. We examined the relationships between the characteristics of a breach type, the location (source of data breach), the entity, and the affected individual. We also

examined the nature of breaches (breach type) and their association with the entity (e.g., healthcare provider), location (e.g., server), and the affected individual. We obtained a glimpse of the trends in healthcare data breaches through our analysis of the reported data breaches. Our research has significance since the topic of data breaches in the context of cybersecurity is current and rapidly gaining public attention. Regardless of the limitations, this research found correlations between the occurrence of data breaches, breach locations, breach types, and the presence of business associates. Hacking, the most common type of data breach, significantly affects individuals in healthcare organizations. Network servers are the most popular location for information breaches, and they are the most common location for breaches related to hacking and unauthorized access. Healthcare providers, as they are related to the largest group of affected individuals, experience various types of breaches.

Data breaches in the healthcare industry show a sharp upward trend. In fact, they have experienced a recent surge. All types of breaches showed expansion across the period studied. Hacking had the highest peak value and largest fluctuation degree per month. Almost all types showed growth when studying the locations of breached information by year. Network server breaches fluctuated the most per month. Regarding a month analysis for the covered entity type, healthcare providers contributed the most to both the volume of affected individuals and the fluctuation. The same occurred with the presence of nonbusiness associates. Data breaches have a detrimental effect on health data privacy.

This research found a possible correlation between population and affected individuals. CA experienced mostly theft when studying the geographical distribution of records based on the type of breach. Hacking was found in other states, meaning organizations should focus on this type of breach. Papers/films were the riskiest trends in CA. This may be related to the prosperity of the state's production industry. However, other states also experienced high risk levels related to network servers and location. When analyzing the network server location, GA, MO, and TN were most affected. CA's health plan organizations had significant data breaches. Other states also experienced high levels of healthcare provider data breaches. According to the study, business has little relation to geographical distribution. Additional research should monitor risky locations and collect historical data. Research should also be applied to the detection process of data breaches. In doing so, patterns of breaches may be revealed. In general, companies should also study their data breach records to prevent future breaches and financial loss. Further research and insights can accelerate the maturing process of our understanding of health data breaches.

Author Contributions: W.R., V.R. and A.S. contributed to the preparation and submission of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bai, G.; Jiang, J.X.; Flasher, R. Hospital risk of data breaches. *JAMA Intern. Med.* **2017**, *177*, 878–880. [CrossRef] [PubMed]
2. Carroll, L. Health data breaches on the rise. *Reuters*. 25 September 2018. Available online: www.reuters.com/article/us-health-data-security-idUSKCN1M524J (accessed on 25 December 2022).
3. Choi, S.J.; Johnson, M.E. Do Hospital Data Breaches Reduce Patient Care Quality? *arXiv* **2019**, arXiv:1904.02058.
4. Lee, J.; Choi, S.J. Hospital Productivity After Data Breaches: Difference-in-Differences Analysis. *J. Med. Internet Res.* **2021**, *23*, e26157. [CrossRef] [PubMed]

5. Chernyshev, M.; Zeadally, S.; Baig, Z. Healthcare data breaches: Implications for digital forensic readiness. *J. Med. Syst.* **2019**, *43*, 7. [CrossRef] [PubMed]
6. Choi, S.J.; Johnson, M.E. The relationship between cybersecurity ratings and the risk of hospital data breaches. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 2085–2092. [CrossRef]
7. Floyd, T.; Grieco, M.; Reid, E.F. Mining hospital data breach records: Cyber threats to US hospitals. In Proceedings of the IEEE Conference on Intelligence and Security Informatics (ISI), Tucson, AZ, USA, 28–30 September 2016; pp. 43–48.
8. Gabriel, M.H.; Noblin, A.; Rutherford, A.; Walden, A.; Cortelyou-Ward, K. Data breach locations, types, and associated characteristics among US hospitals. *Am. J. Manag. Care* **2018**, *24*, 78–84.
9. Liu, V.; Musen, M.A.; Chou, T. Data breaches of protected health information in the United States. *JAMA* **2015**, *313*, 1471–1473. [CrossRef]
10. Trend Micro. Data Breaches 101: How They Happen, What Gets Stolen, and Where It All Goes. 10 August 2018. Available online: https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/data-breach-101 (accessed on 16 December 2022).
11. Appari, A.; Johnson, M.E. Information security and privacy in healthcare: Current state of research. *Int. J. Internet Enterp. Manag.* **2010**, *6*, 279–314. [CrossRef]
12. Tangari, G.; Ikram, M.; Ijaz, K.; Kaafar, M.A.; Berkovsky, S. Mobile health and privacy: Cross sectional study. *BMJ* **2021**, *373*, n1248. [CrossRef]
13. McLeod, A.; Dolezel, D. Cyber-analytics: Modeling factors associated with healthcare data breaches. *Decis. Support Syst.* **2018**, *108*, 57–68. [CrossRef]
14. Thomson, L.L.; Thomson, L.L. Health care data breaches and information security: Addressing threats and risks to patient data. In *Data Breach and Encryption Handbook*; 2013; pp. 57–85.
15. Wikina, S.B. What caused the breach? An examination of use of information technology and health data breaches. *Perspect. Health Inf. Manag.* **2014**, *11*, 1h. [PubMed]
16. Guarino, A.; Malandrino, D.; Zaccagnino, R. An automatic mechanism to provide privacy awareness and control over unwittingly dissemination of online private information. *Comput. Netw.* **2022**, *202*, 108614. [CrossRef]
17. Cozza, F.; Guarino, A.; Isernia, F.; Malandrino, D.; Rapuano, A.; Schiavone, R.; Zaccagnino, R. Hybrid and lightweight detection of third party tracking: Design, implementation, and evaluation. *Comput. Netw.* **2020**, *167*, 106993. [CrossRef]
18. Gostin, L.O.; Halabi, S.F.; Wilson, K. Health data and privacy in the digital era. *JAMA* **2018**, *320*, 233–234. [CrossRef]
19. Kaplan, B. How should health data be used? Privacy, secondary use, and big data sales. *Camb. Q. Healthc. Ethics* **2016**, *25*, 312–329. [CrossRef]
20. Raman, A. Enforcing privacy through security in remote patient monitoring ecosystems. In Proceedings of the 6th International Special Topic Conference on Information Technology Applications in Biomedicine, Tokyo, Japan, 8–11 November 2007; pp. 298–301.
21. Hasan, R.; Yurcik, W. A statistical analysis of disclosed storage security breaches. In Proceedings of the 2nd ACM Workshop on Storage Security and Survivability, Alexandria, VA, USA, 30 October 2006; pp. 1–8.
22. Xiang, D.; Cai, W. Privacy protection and secondary use of health data: Strategies and methods. *BioMed Res. Int.* **2021**, *2021*, 6967166. [CrossRef]
23. Applebaum, P.S. Privacy in psychiatric treatment: Threats and response. *Am. J. Psychiatry* **2002**, *159*, 1809–1818. [CrossRef]
24. Mercuri, R.T. The HIPAA-potamus in health care data security. *Commun. ACM* **2004**, *47*, 25–28. [CrossRef]
25. Thapa, C.; Camtepe, S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Comput. Biol. Med.* **2021**, *129*, 104130. [CrossRef]
26. Abouelmehdi, K.; Beni-Hessane, A.; Khaloufi, H. Big healthcare data: Preserving security and privacy. *J. Big Data* **2018**, *5*, 1. [CrossRef]
27. Keshta, I.; Odeh, A. Security and privacy of electronic health records: Concerns and challenges. *Egypt. Inform. J.* **2021**, *22*, 177–183. [CrossRef]
28. Mershon, E. Insurer's Mailing to Customers Made HIV Status Visible through Envelope Window. Available online: https://www.statnews.com/2017/08/24/aetna-hiv-envelopes/ (accessed on 16 December 2022).
29. HIPAA Journal. Healthcare Data Breach Statistics. 2018. Available online: www.hipaajournal.com/healthcare-data-breach-statistics/ (accessed on 16 December 2022).
30. Angst, C.M.; Block, E.S.; D'arcy, J.; Kelley, K. When do IT security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. *MIS Q.* **2017**, *41*, 893–916. [CrossRef]
31. McCoy, T.H.; Perlis, R.H. Temporal trends and characteristics of reportable health data breaches, 2010–2017. *JAMA* **2018**, *320*, 1282–1284. [CrossRef] [PubMed]
32. Gallagher Cyber Security. Healthcare: The Financial Impact of a Data Breach. 2015.
33. Ronquillo, J.G.; Erik Winterholler, J.; Cwikla, K.; Szymanski, R.; Levy, C. Health IT, hacking, and cybersecurity: National trends in data breaches of protected health information. *JAMIA Open* **2018**, *1*, 15–19. [CrossRef] [PubMed]
34. Donovan, F. Vendor Blamed for Health Data Breach Exposing 1, BCBSRI Members. Health IT Security. 19 September 2018. Available online: https://healthitsecurity.com/news/vendor-blamed-for-health-data-breach-exposing-1500-bcbsri-members (accessed on 16 December 2022).

35. Lord, N. Top Biggest Healthcare Data Breaches of All Time. *Digital Guardian*. 25 June 2018. Available online: Digitalguardian.com/blog/top-10-biggest-healthcare-data-breaches-all-tim (accessed on 16 December 2022).
36. Cohen, J.K. It Takes Healthcare Organizations Days to Detect a Breach, Survey Finds. *Becker's Hospital Review*. 15 October 2018. Available online: www.beckershospitalreview.com/cybersecurity/it-takes-healthcare-organizations-55-days-to-detect-a-breach-survey-finds.html (accessed on 16 December 2022).
37. Seh, A.H.; Zarour, M.; Alenezi, M.; Sarkar, A.K.; Agrawal, A.; Kumar, R.; Ahmad Khan, R. Healthcare data breaches: Insights and implications. *Healthcare* **2020**, *8*, 133.
38. US Department of Health and Human Services. *Health Industry Cybersecurity Practices: Managing Threats and Protecting Patients*; US Department of Health and Human Services: Washington, DC, USA, 2020.
39. Rouse, M. Hacker [Definition]. *TechTarget*. 2017. Available online: https://searchsecurity.techtarget.com/definition/hacker (accessed on 16 December 2022).
40. Beek, C.; McFarland, C.; Samani, R. Health Warning: Cyberattacks Are Targeting the Health Care Industry. Santa Clara: McAfee. Part of Intel Security. McAffee. Hotel Ransomed by Hackers as Guests Locked Out of Rooms. 2017. Available online: https://www.mcafee.com/us/resources/reports/rp-health-warning.pdf (accessed on 25 December 2022).
41. Humer, C.; Finkle, J. Your Medical Record Is Worth More to Hackers than Your Credit Card. *Reuters*. 24 September 2014. Available online: https://www.reuters.com/article/us-cybersecurity-hospitals-idUSKCN0HJ21I (accessed on 16 December 2022).
42. Kemmerer, R.A. Cybersecurity. In Proceedings of the 25th IEEE International Conference Software Engineering, Portland, OR, USA, 3–10 May 2003; pp. 705–715. [CrossRef]
43. Lewis, J.A. *Cybersecurity and Critical Infrastructure Protection*; Center for Strategic and International Studies: Washington, DC, USA, 2006; Available online: http://csis.org/publication/cybersecurity-and-critical-infrastructure-protection (accessed on 25 December 2022).
44. DHS. *A Glossary of Common Cybersecurity Terminology*; National Initiative for Cybersecurity Careers and Studies: Department of Homeland Security; 1 October 2014. Available online: http://niccs.us-cert.gov/glossary#letter_c (accessed on 25 December 2022).
45. Akhtar, N.; Tabassum, N.; Perwej, A.; Perwej, Y. Data analytics and visualization using Tableau utilitarian for COVID-(Coronavirus). *Glob. J. Eng. Technol. Adv.* **2020**. [CrossRef]
46. Toasa, R.; Maximiano, M.; Reis, C.; Guevara, D. Data visualization techniques for real-time information—A custom and dynamic dashboard for analyzing surveys' results. In Proceedings of the 13th Iberian Conference on Information Systems and Technologies (CISTI), Caceres, Spain, 13–16 June 2018; pp. 1–7.
47. Zhang, L.; Stoffel, A.; Behrisch, M.; Mittelstadt, S.; Schreck, T.; Pompl, R.; Keim, D. Visual analytics for the big data era—A comparative review of state-of-the-art commercial systems. In Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, USA, 14–19 October 2012; pp. 173–182.
48. Choi, S.J.; Johnson, M.E.; Lehmann, C.U. Data breach remediation efforts and their implications for hospital quality. *Health Serv. Res.* **2019**, *54*, 971–980. [CrossRef]
49. Raghupathi, W.; Raghupathi, V. Contemporary Business Analytics: An Overview. *Data* **2021**, *6*, 86. [CrossRef]
50. Raghupathi, W.; Raghupathi, V. An overview of health analytics. *J. Health Med. Inform.* **2013**, *4*, 2. [CrossRef]
51. Börner, K.; Bueckle, A.; Ginda, M. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1857–1864. [CrossRef]
52. Keim, D.; Kohlhammer, J.; Ellis, G.; Mansmann, F. *Mastering the Information Age Solving Problems with Visual Analytics*; Eurographics Association: Saarbrücken, Germany, 2010.
53. Keim, D.A. Visual exploration of large data sets. *Commun. ACM* **2001**, *44*, 38–44. [CrossRef]
54. Wong, P.C.; Thomas, J. Guest Editors' Introduction—Visual Analytics. *IEEE Comput. Graph. Appl.* **2004**, *24*, 20–21. [CrossRef] [PubMed]
55. Kohlhammer, J.; Keim, D.; Pohl, M.; Santucci, G.; Andrienko, G. Solving problems with visual analytics. *Procedia Comput. Sci.* **2011**, *7*, 117–120. [CrossRef]
56. Thomas, J.; Cook, K. *Illuminating the Path: Research and Department Agenda for Visual Analytics*; United States Department of Homeland Security: Washington, DC, USA, 2005.
57. Singh, D.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2020**, *97*, 105524. [CrossRef]
58. Cao, N.; Koch, S.; Gotz, D. ACM TIST Special Issue on Visual Analytics. *ACM Trans. Intell. Syst. Technol. (TIST)* **2018**, *10*, 1–4. [CrossRef]
59. Lettieri, N.; Guarino, A.; Malandrino, D.; Zaccagnino, R. The sight of Justice. Visual knowledge mining, legal data and computational crime analysis. In Proceedings of the 25th International Conference Information Visualisation (IV), Sydney, Australia, 5–9 July 2021; pp. 267–272. [CrossRef]
60. Heer, J.; Bostock, M.; Ogievetsky, V. A tour through the visualization zoo. *Commun. ACM* **2010**, *53*, 59–67. [CrossRef]
61. Lettieri, N.; Guarino, A.; Malandrino, D.; Zaccagnino, R. The Affordance of Law. Sliding Treemaps browsing Hierarchically Structured Data on Touch Devices. In Proceedings of the 24th International Conference Information Visualisation (IV), Melbourne, Australia, 7–11 September 2020; pp. 16–21. [CrossRef]
62. Liu, S.; Wang, X.; Liu, M.; Zhu, J. Towards better analysis of machine learning models: A visual analytics perspective. *Vis. Inform.* **2017**, *1*, 48–56. [CrossRef]

63. Xie, C.; Zhong, W.; Xu, W.; Mueller, K. Visual analytics of heterogeneous data using hypergraph learning. *ACM Trans. Intell. Syst. Technol. (TIST)* **2018**, *10*, 4. [CrossRef]
64. Yang, D.; Xie, Z.; Rundensteiner, E.A.; Ward, M.O. Managing discoveries in the visual analytics process. *ACM SIGKDD Explor. Newsl.* **2007**, *9*, 22–29. [CrossRef]