Scientific Research Publishing

Journal of Information Security, 2024, 15, 168-195
https://www.scirp.org/journal/jis
ISSN Online: 2153-1242
ISSN Print: 2153-1234

Exploring the Characteristics of Data Breaches: A Descriptive Analytic Study

Dominik Molitor[1], Aditya Saharia[1], Viju Raghupathi[2], Wullianallur Raghupathi[1]

[1]Gabelli School of Business, Fordham University, New York, USA
[2]Brooklyn College, The City University of New York, New York, USA
Email: dmolitor@fordham.edu, saharia@fordham.edu, raghupathi@fordham.edu, VRAGHUPATHI@brooklyn.cuny.edu

How to cite this paper: Molitor, D., Saharia, A., Raghupathi, V. and Raghupathi, W. (2024) Exploring the Characteristics of Data Breaches: A Descriptive Analytic Study. *Journal of Information Security*, **15**, 168-195.
https://doi.org/10.4236/jis.2024.152011

Received: January 23, 2024
Accepted: March 30, 2024
Published: April 2, 2024

Abstract

Data breaches have massive consequences for companies, affecting them financially and undermining their reputation, which poses significant challenges to online security and the long-term viability of businesses. This study analyzes trends in data breaches in the United States, examining the frequency, causes, and magnitude of breaches across various industries. We document that data breaches are increasing, with hacking emerging as the leading cause. Our descriptive analyses explore factors influencing breaches, including security vulnerabilities, human error, and malicious attacks. The findings provide policymakers and businesses with actionable insights to bolster data security through proactive audits, patching, encryption, and response planning. By better understanding breach patterns and risk factors, organizations can take targeted steps to enhance protections and mitigate the potential damage of future incidents.

Keywords

Cyber Security, Information Security, Data Breaches, Descriptive Analytics, Privacy, Visualization, Visual Analytics

1. Introduction

The increasing reliance on the internet and surge in data usage has been followed by a rise in cyber threats and data breaches—a menace that current regulations, security standards, and companies have been unable to effectively curb [1] [2]. The explosion in the number of connected users and the dependence on the internet represents an exponential increase in the production of information [1] [3]. As a result, the conditions are ripe for data breaches to occur, jeopardizing the privacy and security of all those who have ever connected [3].

Remarkably, more than 90% of all online data resulting from this explosion was created within the past two years [4], and it is expected that the volume will increase from 33 zettabytes (ZB) in 2018 to 175 ZB in 2025 [1] [3] [5]. There have been numerous reports about cyberattacks, including Bloomberg's story on the most relevant cybersecurity incidents [1] [3] [6]. Some of the largest organizations in the world were impacted by these cyberattacks, such as Capital One's data breach incident [3], which exposed critical and confidential information about their customers in 2019.

Data breaches have manifold repercussions on companies, affecting them financially and undermining their reputation, which poses significant challenges to online security and the long-term viability of businesses. Concurrently, the financial burden of addressing data breaches has escalated dramatically over time, as highlighted by Schlackl, *et al.* [1]. In 2006, the average cost of a data breach in the US was approximately 3.5 million USD. By 2020, this figure had soared to 8.64 million USD, representing a substantial increase of over 140% in a span of 14 years [1]. Given these circumstances, companies globally are experiencing a growing urgency to understand their vulnerability to data breaches and to align with the evolving regulatory landscape governing data protection. This necessitates a proactive approach to bolster data security and ensure adherence to emerging data protection laws and regulations. In addition, companies must anticipate threats that can expose data [1] [2] [3]. Avoiding breaches is a priority, so firms need to resolve incidents quickly [2]. Breaches harm the public image and have legal and financial costs, with inadequate response being more costly than protection [7]. Quantifying breach likelihood is key for mitigation and preparation [8]. However, research into breaches is ad hoc and limited, as details are often undisclosed [1]. Data is hard to find as companies and others are reluctant to disclose complete details about their data breaches [9].

The primary objective of this descriptive study is to identify the various types and characteristics of organizational data breaches and, significantly, to formulate actionable insights on addressing this prevailing issue. This study distinguishes itself through several facets. First, it utilizes data from the Privacy Rights Clearinghouse database (https://privacyrights.org/data-breaches). Second, this research involves the application of visualization and visual analytics techniques, as corroborated by several scholarly works [10] [11] [12]. This study introduces significant advancements in the analysis of organizational data breaches by leveraging a unique dataset and applying visual analytics. The use of visualization techniques facilitates a deeper comprehension of data breaches, fostering the potential to anticipate such incidents in the future. This dual focus on descriptive analysis and the potential for predictive application underscores the study's novelty and its contribution to the broader discourse on data security and data breach prevention. Consequently, this study attempts to fill the gap in existing research by developing visual charts that scrutinize patterns and identify distinctive features of data breaches.

The rest of the paper is organized as follows. A literature review is provided in

Section 2. We then discuss the methods used in Section 3. This is followed by an analysis of the results in Section 4 and a discussion of the implications in Section 5. The scope and limitations of the study are discussed in Section 6. Finally, our conclusions and future research directions are offered in Section 7.

2. Background

2.1. Scope of Data Breaches

According to Sharma, *et al.* [13], a data breach refers to "the release of confidential data, commonly known as personally identifiable information, from a secured location in a computer or an electronic device to an unsecured site." The U.S. Department of Health and Human Services (DHHS) for their purposes defines a breach as "generally, an impermissible use or disclosure under the Privacy Rule that compromises the security or privacy of the protected health information" (https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html). Hence, DHHS considers not only the disclosure of protected information but also the impermissible use as a breach. California's data breach notification law defines a data breach as a "breach of the security of the system" as an "unauthorized acquisition of computerized data that compromises the security, confidentiality, or integrity of personal information maintained by the person or business" (https://www.oag.ca.gov/privacy/databreach/reporting). Similarly, the U.S. National Initiative for Cybersecurity Careers and Studies (NICCS) in their glossary defines a data breach as "the unauthorized movement or disclosure of sensitive information to a party, usually outside the organization, that is not authorized to have or see the information" (https://niccs.cisa.gov/cybersecurity-career-resources/vocabulary).

After the farming of Facebook data by Cambridge Analytica, there has been a growing demand to extend the definition of a data breach to include manipulation of data through social engineering [14]. However, as is it defined, the release of confidential data creates severe legal implications and public relations problems for the companies hosting and managing the data. From another perspective, a data breach is an electronically mediated service failure that occurs when sensitive financial, personal, or customer data is released to or accessed by parties external to the organization. This exposure may be deliberate, such as through a hacking incident or due to the actions of a disgruntled employee, or accidental, such as a lost laptop [15] [16] [17] [18], and may relate to any aspect of an organization's activities or associations, such as customers, trading partners, and internal systems. Because data are typically collected by an organization as it fulfills its service offerings to customers, at least some of the data stored by an organization may relate to the organization's customers themselves. Prior research has used a variety of terms to describe data breaches, such as security breaches [19], information breaches [20], and privacy breaches [21]. Following Culnan and Williams [22], one uses the term data breach because, although the organization may know that their data has been breached, it typically does not know the content of the breached data until after the incident has been detected

and investigated [23]. Thus, a data breach can incorporate security, information, and privacy breaches. The risk of data breaches has been a topic of interest to IS researchers since the 1970s (e.g. [24]). Early research focused on the risks posed by physical access to sensitive business data and the threat posed by competitor access to electronic processes [25]. Research into the threat of large-scale data breaches became more popular in the early 2000s, with the growth of commercial Internet access. As organizations have become more interconnected and reliant on confidential customer data, the number and magnitude of data breaches has grown considerably [16] [26] [27] [28].

2.2. Causes of Data Breaches

A data breach occurs because of a lack of security, elimination of security, and a breach of security [29]. Lack of security may be because of unwillingness on the part of the organization to consider itself to be prone to data breaches or is considered too cost-prohibitive to secure the information [9]. Elimination of security may be because of slackness on the part of the organization to beef up the data security, and either insiders or outsiders purposefully eliminate the security protocols to make the data vulnerable [30]. Such elimination may also be because of accidental loss of privileges and equipment or some state-sponsored actors purposefully removing the security to create vulnerability; for example, the Desjardins Group data breach exposing 2.9 million members was caused by an employee [31]. Breach of security is intentional on the part of actors to steal the data using malware, hacking, viruses, social engineering, cyber espionage, and sabotage. It could be accidental when sensitive information is leaked inadvertently by accidental publication, configuration errors, improper encryption, lost computers, and privilege abuse [32]. A UK survey conducted by Clearswift [33] revealed the sources of data breaches: 42% were orchestrated by outsiders, while 58% originated from insiders within the extended enterprise—encompassing current employees (33%), former employees (7%), and third parties (18%). Notably, the majority of internal security threats stemmed from inadvertent human errors, a lack of awareness, and malware introduced through personal devices. Wikina [34] reported that data breaches involved computer systems and networks, desktop, laptops, paper records, emails, electronic health records, and portable devices. The researcher also reported since hacking is one of the highest contributing methods with 69.5% of instances followed by poor security at about 23%, a combination of laws, investment in precautionary measures, and training the cybersecurity personnel may be a better strategy [35]. Lost devices are some of the most interesting as many devices these days have a hard drive; a protocol for their disposal is necessary. For example, a stolen digital camera belonging to the University of Arkansas for Medical Sciences (UAMS) in Little Rock contained photographs of newborns and their information which was responsible for the data breach [35]. The same is possible with copy machines, fax machines, and biomedical equipment. Therefore, technology and web-related entities are more prone to data breaches followed by government entities, and the financial

sector. The data points report worldwide figures and are not necessarily US-based entities only [13]. Another devasting data breach over the years has been caused by the use of ransomware, when hackers encrypt the data of the target organizations and demand ransom to decrypt the data. Commonly known ransom wares are Cryptologer in 2014, Teslacrypt in 2016, Wanacry in 2017, Cryptowall, and AlphaCrypt [36] [37]. The other causes of data breaches resulting in cyber fraud are phishing, spyware, pharming, and spoofing. There are also internal causes of data breaches. The Intelligence and National Security Alliance had identified four insider threats: fraud, theft of intellectual property, IT sabotage and Espionage [13]. According to them, data leakage is a serious threat to enterprise operations, such as corporations and government agencies. The loss of sensitive information can lead to significant reputational damage and financial losses and can even be detrimental to the long-term stability of an organization. Common types of leaked information range from employee/customer data, and intellectual property, to medical records. According to IBM's 2020 Cost of Data Breach Study, the average consolidated cost of a data breach has reached $3.9 million [38]. Cybercriminals breached the Target Corporation's network in 2013, stealing 40 million payment card information and 70 million customers' personally identifiable information, which has incurred $ 248 million losses to date reported by Target [32]. In 2016, Yahoo reported that at least 500 million accounts in 2014 had been stolen in an apparent 'state-sponsored' data breach [32]. Since data volume is growing exponentially in the digital era and data leaks happen more frequently than ever before, preventing sensitive information from being leaked to unauthorized parties becomes one of the most pressing security concerns for enterprises. Data leakage can be caused by internal and external information breaches, either intentionally (e.g., data theft by intruders or sabotage by insider attackers) or inadvertently (e.g., accidental disclosure of sensitive information by employees and partners) [32]. A study from Intel Security showed that internal employees account for 43% of corporate data leakage, and half of these leaks are accidental [39]. Motivations of insider attacks are varied, including corporate espionage, grievance with their employer, or financial reward. Accidental leaks result from unintentional activities due to poor business processes such as failure to apply appropriate preventative technologies and security policies, or employee oversight [32].

3. Literature Review

The current work on data breaches mostly consists of technical reports and white papers by government bodies and security solution providers [40]. Verizon has been publishing its "Verizon Breach Investigation Report" annually since 2008, which provides insights about past cyber security incidents [40]. The report highlights common attack patterns, threat actors, attacker motivations, breach discovery methods and timelines, and recent malware trends. The data for the report comes from real-world breaches either investigated by Verizon or by one of the contributing organizations [40]. Verizon also publishes data

breach case studies in their "Data Breach Digest." Trustwave publishes its annual Global Security Report that highlights top security threats and attack trends (https://www.infopoint-security.de/media/Trustwave_2018-GSR_20180329_Interactive.pdf). Additionally, a few case studies have been published regarding data breaches. For example, Manworren, *et al.* [41] discuss the Target data breach case. Rashid, *et al.* [37] provide a model capturing separate phases of a data breach and present breach detection and mitigation strategies. Collins, *et al.* [42] performed an exhaustive literature review to assess the current state of organizational data breaches within the United States. Explicitly, this research reviewed the applicability of Situational Crime Prevention, the influence of current breach notification laws, findings drawn from macro-level studies of data breaches, and reporting issues unique to the entities of health care and education. To assess the results of the literature, a six-year sample of reported data breaches was compiled from the Privacy Rights Clearinghouse, consisting of a total of 2219 data breaches disclosed between 2005 and 2010 [42]. This analysis specifically addressed four individual variables: type of breach, reporting entity, the year the breach was disclosed, and the geographic region in the United States where the breach was reported [42]. Ultimately, the study concluded that the passage of reporting legislation within the healthcare field increased the number of incidents reported; breaches reported by educational institutions appear to be on the decline; the lack of a centralized reporting database for all data breaches prevents a definitive analysis of the field; and that situational crime prevention measures can be proactive in preventing future data breaches within these entities [42]. Posey Garrison and Ncube [43] conducted research to provide companies and consumers with information about the potential connections between data breach types and institutions. The study also aimed to add to the body of knowledge about data breaches. It analyzed a period of five years of data breaches. The data were classified and analyzed by breach and institution type, record size, and state. The study found that breach types stolen and exposed were more likely to occur [43]. Educational institutions were more likely to have a breach, and it was more probable that educational breaches would be of type hacker or exposed. The proportion of insider incidents was smaller than the other breach types. The number of records breached was independent of institution and breach type [43]. Ayyagari [35] performed a content analysis on 2633 unique data breaches that resulted in the loss of more than 500 million individual records. The results indicated that data breaches continued to be a major issue for organizations. The results implied that the nature of data breaches was changing. Data breaches are typically associated with hacking—however, results indicated that breaches due to hacking decreased, whereas breaches due to human element increased. One disconcerting result was that data breaches that can be directly attributed to the implementation and enforcement of security policies accounted for a major share. Collectively, the results suggested that organizations need to implement effective training and stricter enforcement of security policies [35]. Khey and Sainato [44] in their research utilizes a 6-year sample compiled from the Privacy

Rights Clearinghouse who maintain a record of all published data breaches in the United States to examine the correlates of data loss as well as how lost records are spatially distributed across the country. In this study, a Cyber Security Risk Quantification and Mitigation Framework is discussed. Zadeh [2] introduced a breach-level index model to quantify and classify the severity of a data breach incident based on the type of data asset, account details, or financial details that was exposed. Then, a likelihood-impact analysis was conducted to assess the risk involved in each type of data breach. The framework was applied to data breaches gathered from S&P 500 organizations to prescribe strategies that can help firms reduce the likelihood and impact of data breaches. The results suggested that hacking and malware need to be reduced as they have the highest impact and highest probability when it comes to a data breach. The results of this study can help organizations identify the likelihood and impact of a data breach and determine a plan of action on how to mitigate the risks [2]. Cheng, *et al.* [32] conducted a review to help readers to learn about enterprise data leak threats, recent data leak incidents, various state-of-the-art prevention and detection techniques, new challenges, and promising solutions and exciting opportunities [32]. Hammouchi, *et al.* [36] analyzed over 9000 data breaches made public since 2005 that led to the loss of 11.5 billion individual records, which have a significant financial and technical impact. They also illuminated the patterns regarding which types of organizations are most targeted, analyzing how hacker interests shift over time. On the other hand, the breaches caused by human factors decreased explained by the awareness of employees and the application of security standards [36].

Ayyagari [35] analyzed the data breaches recorded between 2005 and 2011. The author conducted a content analysis and showed that hacking breaches had a decreasing trend. In contrast, Hammouchi, *et al.* [36] revealed that hacking was the most prevalent type of breach. Shu, *et al.* [45] investigated the attack on the Target Corporation from a legal perspective and presented several practices to avoid the disclosure of personal information. Smith [31] focused on data breaches in healthcare organizations to determine the association between data privacy breaches, data storage locations, business associates, covered entities, and the number of individuals affected. They found that 70% of breaches involve healthcare providers and security incidents often consist of electronic or other digital information. McLeod and Dolezel [46] studied the data breaches encountered in healthcare organizations and found that the level of exposure, the level of security and the organizational factors within health facilities can lead to a data breach. Algarni and Malaiya [47] assessed the data breach costs and examined how the calculators compute these costs and the factors that affect them. Kafali, *et al.* [48] shed light on data breaches from the policy side. They studied the relationship between policies and data breaches and measured the gaps between them. Also, they found that accidental misuses were as prevalent as malicious misuses.

Sen and Borle [49] addressed data breaches from multiple angles by applying

various theories to measure and identify factors that can increase the risk of a data breach occurrence. Interestingly, they found that investment in information technology (IT) security corresponded to a higher risk of data breaches. Holtfreter and Harrington [50] analyzed the trends of various types of data breaches and their compromised records in the USA using a new model recently developed by the authors. The 2280 data breaches and over 512 million related compromised records tracked by the Privacy Rights Clearinghouse from 2005 through 2010 were analyzed and classified into four external, five internal, and one non-traceable data breach categories, after which trends were determined for each [50]. The study found that although the trends for the annual number of data breaches and each of the internal and external categories and their related compromised records had increased over the six-year period, the changes have not been consistent from year to year [50]. The database created by Neto, *et al.* [3], indicates a substantial escalation in the number of breached data records in the top 430 incidents, rising from approximately 4 billion in 2018 to over 22 billion in 2019 (Accessible at https://www.databreachdb.com). This surge took place despite stringent efforts by regulatory agencies worldwide to impose rigid data protection and privacy rules. A notable example of this regulatory vigor is the enforcement of the General Data Protection Regulation (GDPR), which came into effect in Europe in May 2018 (Neto *et al.*, 2021). Such regulatory effort could explain the reason there was such a large number of data breach cases reported in the European Union when compared to the U.S. Specifically, since 2018, the U.S. has documented over 10,000 public data breaches, whereas the EU has reported a staggering number exceeding 160,000 cases from May 2018 onwards [3]. However, the recent introduction of the California Consumer Privacy Act (CCPA), effective since January 2020, may also trigger an increase in the number of reported data breach cases in the U.S. Hall and Wright [51] conducted a statistical analysis of data breaches from 2014 to 2018, concluding that cyberattacks can occur in any industry, with the characteristics of these incidents fluctuating based on the type of breach and the nature of the business impacted. For instance, recent years—particularly the last five—have witnessed a surge in incidents where enormous quantities of data (amounting to millions of data points) have been compromised. A noticeable trend in this period has been an increase in healthcare data breaches and the vulnerability of massive databases housed in cloud computing infrastructures due to inadequate access control mechanisms [3] [51].

In a comprehensive study, Schlackl, *et al.* [1] examined 43 articles concerning the precursors of data breaches and 83 discussing their repercussions, categorizing the findings into eight unique classifications for both the antecedents and consequences. The theoretical perspectives identified spanned various lenses, from viewing data breaches as organizational crises to adopting criminological and privacy-centric theories [1]. Against this backdrop, we identify three relevant gaps in existing literature. First, although studying data breaches is critical

[1] [49], empirical work at the granular level is sparse and we found that most, if not all, studies exclusively reported case studies of major data breaches, or aggregated data reviews [52]. Second, very few studies have utilized analytical techniques such as visualization and visual analytics to conduct descriptive analysis of granular data. Third, a minority of data-driven studies that exist are somewhat outdated. Our study attempts to fill these gaps, offering fresh insights and a more nuanced understanding of the current landscape.

4. Methods

Descriptive Analytics and Visualization

We utilize visual analytics—a data-driven approach that leverages visualization to facilitate descriptive analytics of data breaches [10] [11] [12] [53]. This methodology enables the robust analysis and comprehension of large datasets in real-time, utilizing the inherent capabilities of tools like Tableau coupled with analytical expertise to unearth hidden patterns and derive valuable insights [54] [55] [56]. Our primary data source is the Privacy Clearinghouse Database, which serves as a repository for our investigations. This analytical approach synthesizes results into a visually cohesive narrative that offers depth and diverse perspectives on the data at hand [21] [53]. This encapsulates the idea that a "picture is worth a thousand words," presenting data in a manner that is both insightful and articulate, thus facilitating informed decision-making [57] [58] [59]. Our objective is to engage in storytelling through visualization, one of the primary pillars of analytics [53] [60] [61] [62] [63]. This approach, largely more data-driven compared to other analytical approaches, prioritizes presenting the data 'as is' without any preconceived assumptions, thereby letting the date reveal itself. This approach supports understanding of both past and present patterns and trends, and leverages them for informed decision-making [53] [54] [58]. Through the utilization of various charts, information is depicted graphically, embracing aggregation, categorization, and characterization functions to vividly illustrate data insights [56] [57] [61].

Data

We utilize data from the Privacy Rights Clearinghouse repository (http://www.privacyrights.org/data-breaches), which contains publicly available information on reported breaches in the United States, primarily disclosed by government entities. The available data, encompassing various metrics, is complete up until the year 2018. Despite this, the dataset remains sufficiently expansive and detailed to extract significant insights and formulate conclusions with a reasonable degree of generalizability. We utilize the entire dataset within the database, considering variables including but not limited to the type of organization involved, the nature of the breach, the total number of records compromised, detailed descriptions of each incident, and the year of occurrence. Our approach is designated to reveal key correlations and patterns within the data set. To this end, the analysis primarily focuses on the following dimensions: 1)

the nature of the breach, 2) spatial patterns of breaches, and 3) an analysis of the organizations implicated in these breaches. For a detailed breakdown of the variables considered, please refer to Table 1 below.

5. Results and Analysis

Utilizing visualization, we developed a series of charts to understand the breach

Table 1. Variables in the research.

Category	Variable	Description
Company	Date made public	Date the company went public
	Organization	The name of the organization
	City	The city in which the organization is located
	State	The state in which the organization is located
	Organization type	The type of organization with the data breach: BSF (Business -Financial and Insurance Service), BSO (Business-Other), BSR (Business-Retail/Merchant-Including Online Retail), EDU (Educational Institution), GOV (Government & Military), MED (Healthcare, Medical Providers & Insurance Services), NGO (Nonprofits), UNKN (Unknown)
	Latitude	Latitude (location) of the company with the breach
	Longitude	Longitude (location) of the company with the breach
Data breach	Type of Breach	The type of data breach: CARD (Payment Card Fraud), HACK (Hacking or Malware), INSD (Insider), PHYS (Physical Loss), PORT (Portable Device), STAT (Stationary Device), DISC (Unintended Disclosure), UNKN (Unknown
	Total records	Total number of data breaches recorded for the organization for the year in which the breach occurred
	Description of incident	Details of the breach incident
	Information Source	Details of the source reporting the breach
	Source URL	URL of the source reporting the breach
	Year of Breach	Year in which the breach occurred

Data Source: https://www.privacyrights.org/data-breaches; Years data is collected for: 2005-2018. *Note*: Additional Information: CARD (Payment Card Fraud), HACK (Hacking or Malware), INSD (Insider), PHYS (Physical Loss), PORT (Portable Device), STAT (Stationary Device), DISC (Unintended Disclosure), BSF (Business Financial and Insurance Service), BSO (Business-Other), BSR (Business-Retail/Merchant-Including Online Retail), EDU (Educational Institutions), GOV (Government & Military), MED (Healthcare, Medical Providers & Medical Insurance Services), NGO (Nonprofits), UNKN (Unknown).

data. Collectively, the charts tell a compelling story about the nature and dimensions of data breaches.

Geographical Spread of Data Breaches

In Figure 1 below, data breach incidences across the U.S. within the observation period are visually depicted. A gradient of color intensity represents the number of unique companies affected by data breaches in each state, with a darker shade signifying a greater number. Simultaneously, the circles placed on the map indicate city-specific contributions to the state's total data breach count; the larger the circle, the higher the percentage contribution of that city to the state's overall number of data breaches. As evident from the map, California stands out as the state housing the largest number of unique companies affected, closely followed by New York and Texas. In states where a fewer number of distinct companies are recorded, it is plausible that a majority of data breaches are concentrated within several cities, indicating a more localized vulnerability. Notably, Washington DC features the most significant circle on the eastern side of the map, highlighting its substantial contribution to the data breach statistics in its state.

Figure 2 further shows the actual number of companies experiencing data breaches in each state. The darker shades indicate the states experiencing more data breaches. Data breaches are more prevalent in the states of California, Oregon,

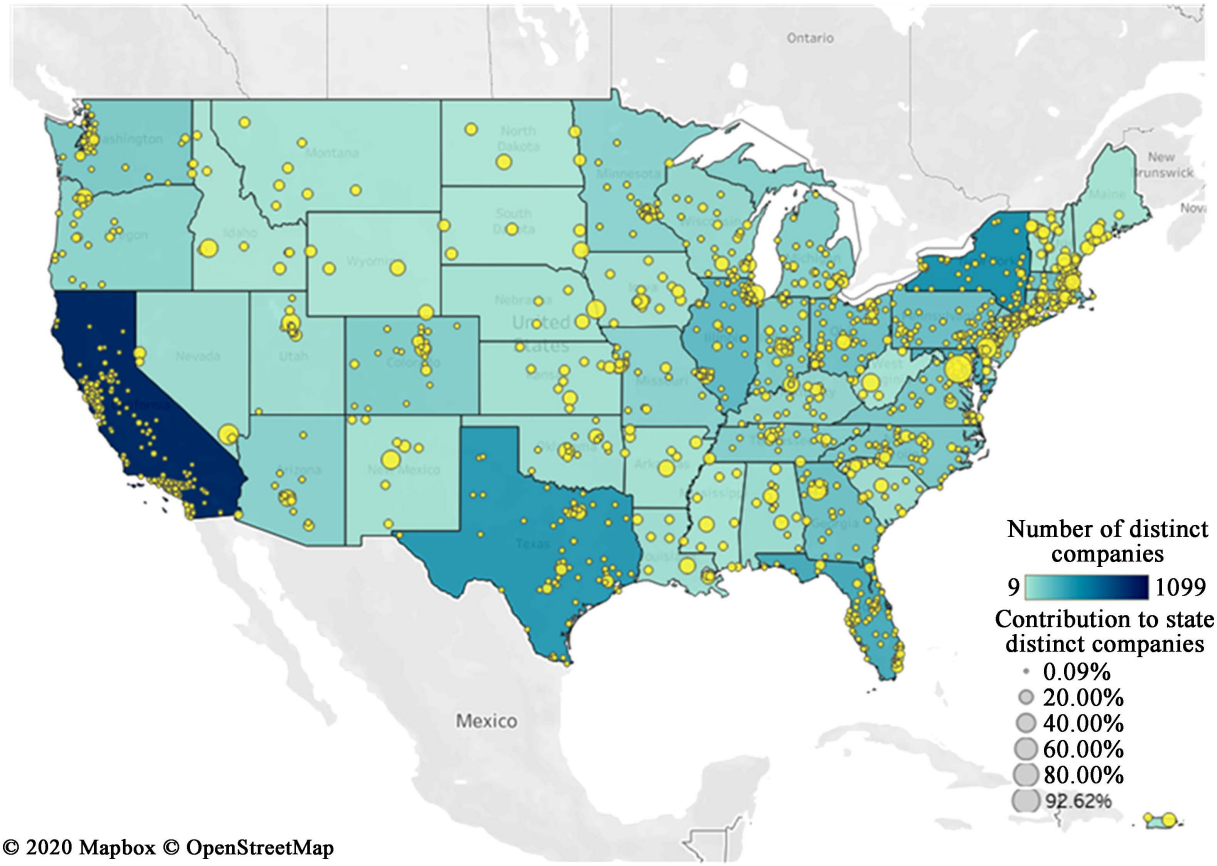

Figure 1. Distribution of distinct companies by state and city.

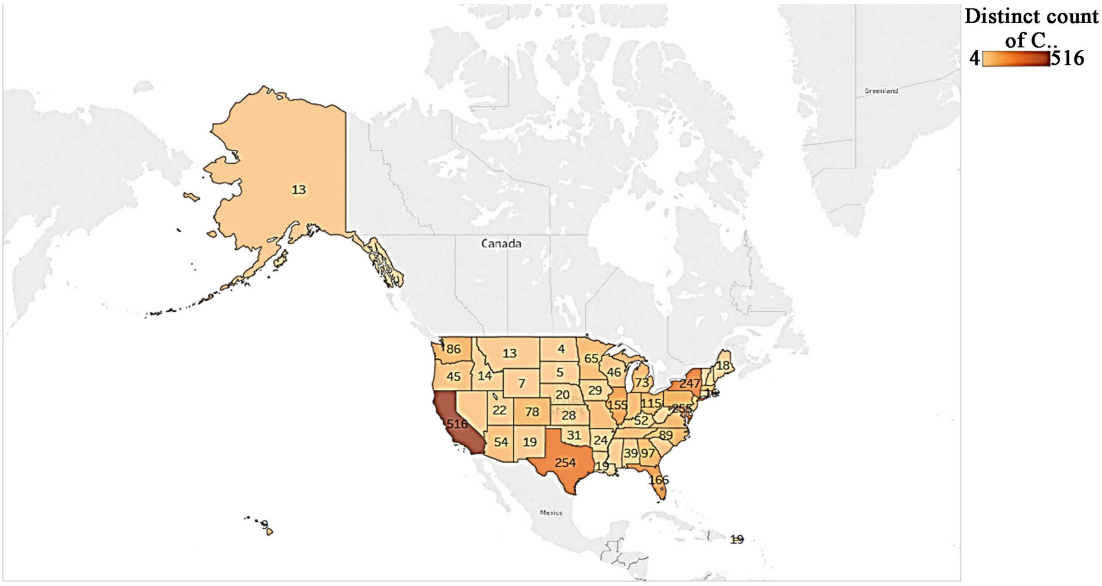

Figure 2. Data breaches are experienced by companies in each state in the U.S.

Texas, New York, and Florida. As evidenced earlier, California is clearly the leader with an extremely high number, followed by Texas. Data breaches appear to be concentrated in companies housed in states with dense populations.

Considering there are various types of breaches, it makes sense to explore which ones are prevalent across the country. **Figure 3** displays the leading breach types in terms of the magnitude of occurrence and the state of occurrence. The size of the box indicates the magnitude, and the color indicates the state. Across the U.S., hacking (HACK) is the leading type of breach with the highest prevalence in the state of California, followed by New York, Wisconsin and Texas. The second highest is breach by unintended disclosure (DISC), occurring in the states of Oregon, Georgia & Washington.

Total Average Data Breaches by companies for the states

Figure 4 illustrates the top ten average breaches that occurred at companies within various states. The depicted line signifies the aggregate breaches observed in each state, with particular emphasis on Oregon and California, which rank first and second respectively, accounting for 39.51% and 15.86% of breaches. It needs to be noted here that even though California reported the highest number of breaches, Oregon reports the highest in terms of the average number of breaches per company. While certain states may exhibit a high total number of breaches due to isolated incidents of rare breach types, analyzing the average number can provide a more accurate representation of the commonplace incidents, and of the concentration of breaches in companies. This can signal a potential area of concern that warrants the attention of local governmental bodies.

Drill down of Data Breaches in key states

In **Figure 5**, a comparative analysis of the total number of companies versus the number of reported breaches in various states is presented. The diameter of each circle is proportional to the total number of breaches documented within a

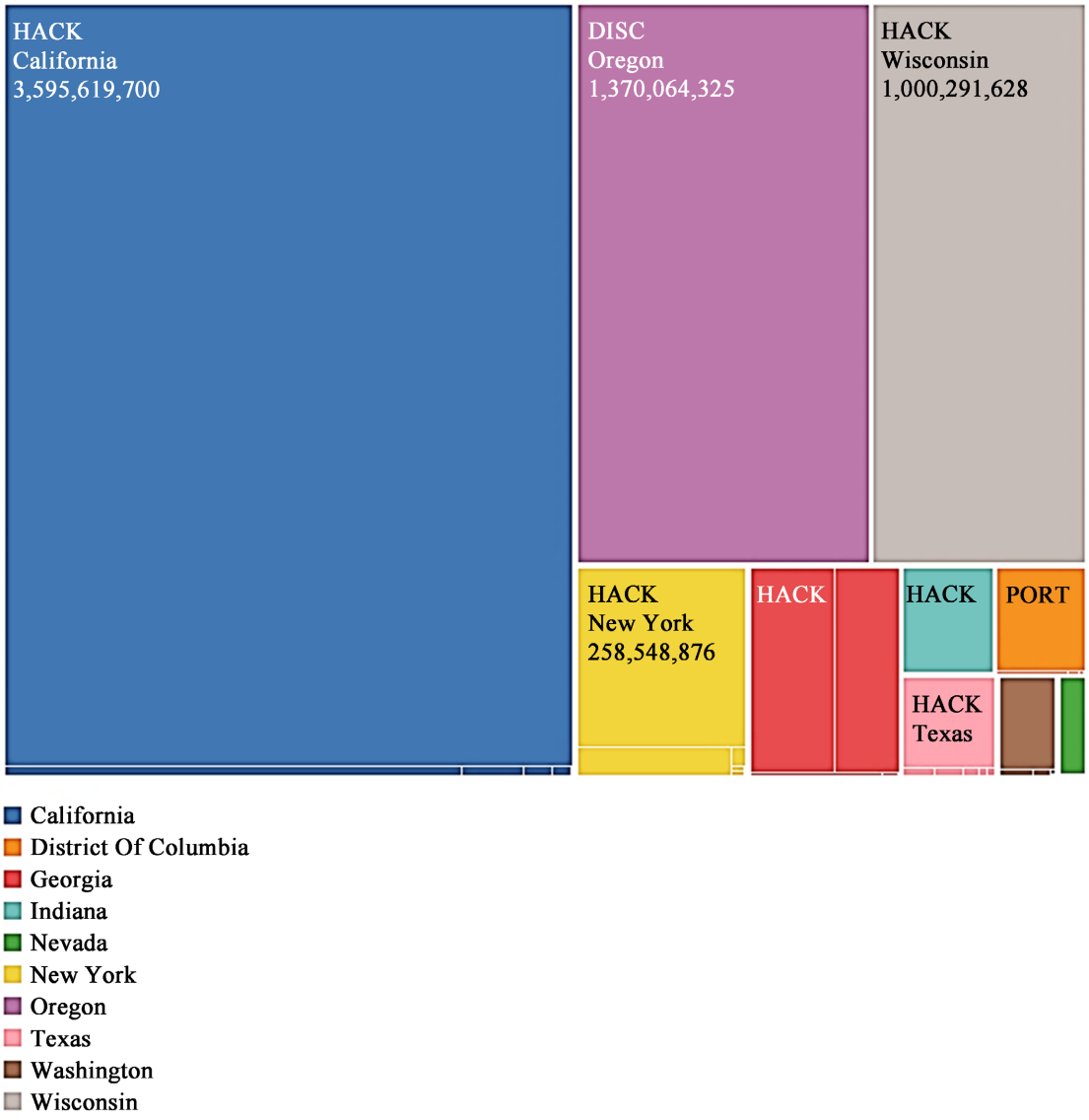

Figure 3. Most prevalent breach types and the state of occurrence.

respective state. Upon examination, it is evident that the state of California has a notably higher number of breaches in comparison to New York. Although the number of companies operating in California is only double that in New York, the discrepancy in the number of breaches between these two states is considerably larger.

Agencies/Information sources by state

Figure 6 shows the primary agencies and information outlets reporting data breaches across various states. Notably, the U.S. Department of Health and Human Services emerges as a constant presence across all states, registering the highest frequency of reported breaches over the observed period. This trend strongly suggests that breaches involving health data stand as the most frequently reported type, underscoring a significant vulnerability in the protection of health-related information nationwide.

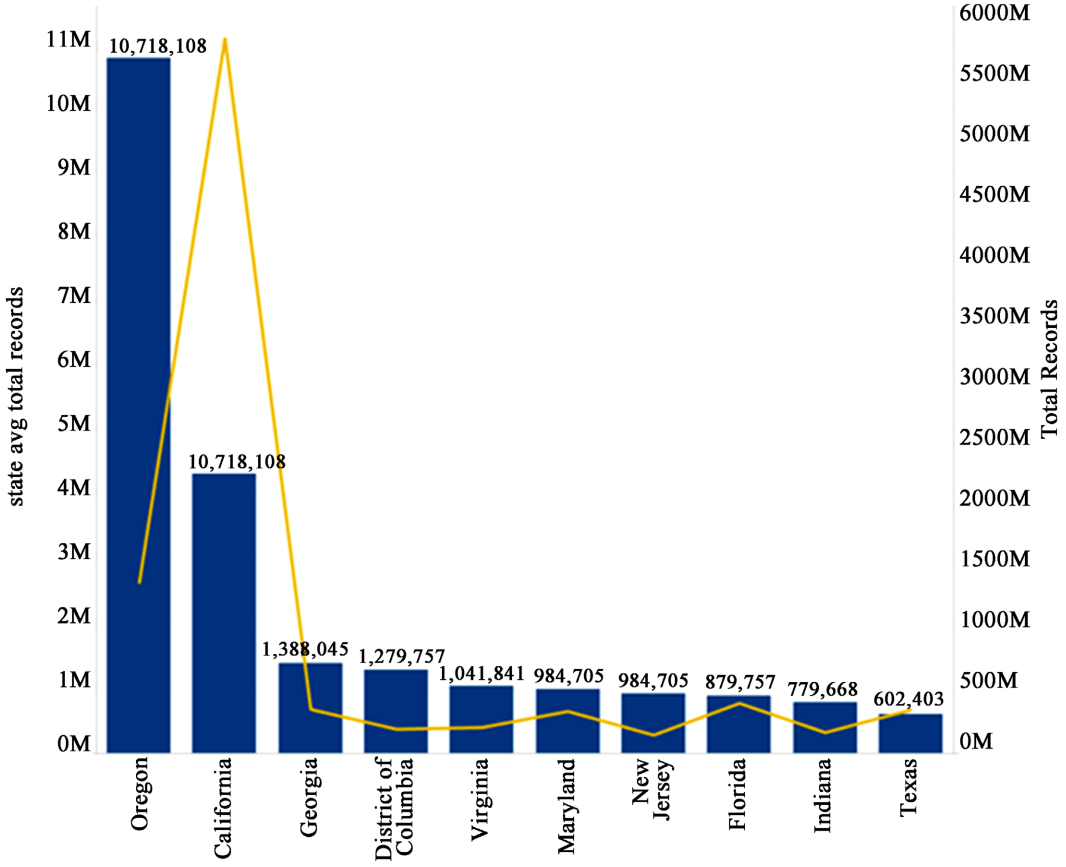

Figure 4. Top 10 average breaches per company by state.

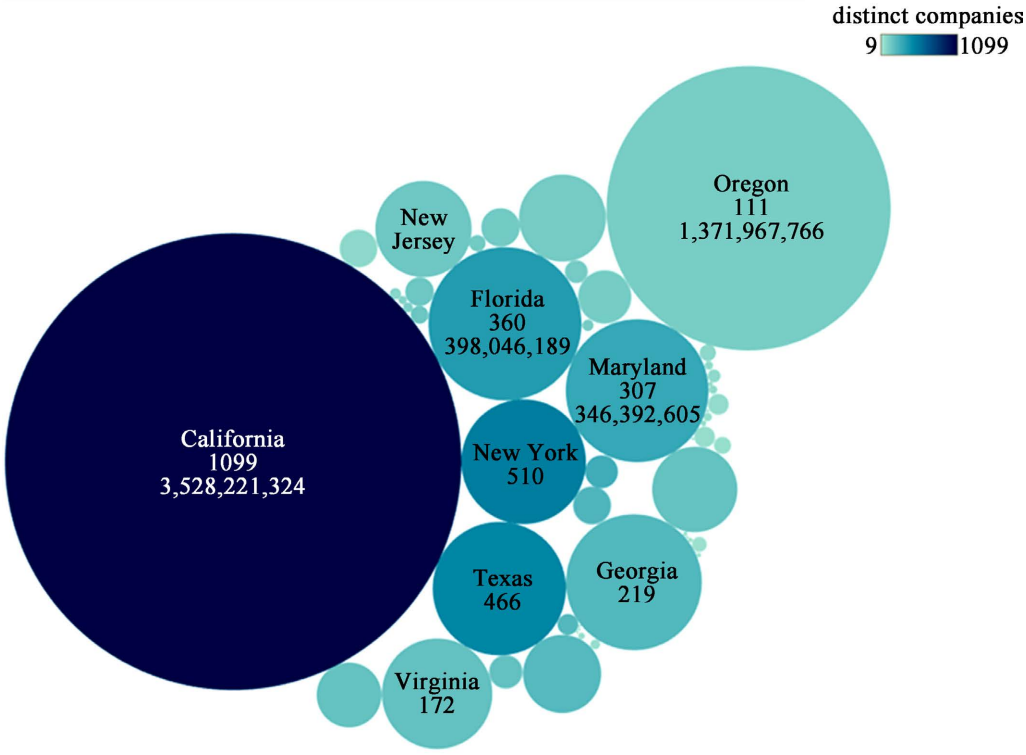

Figure 5. The number of breaches and companies by state.

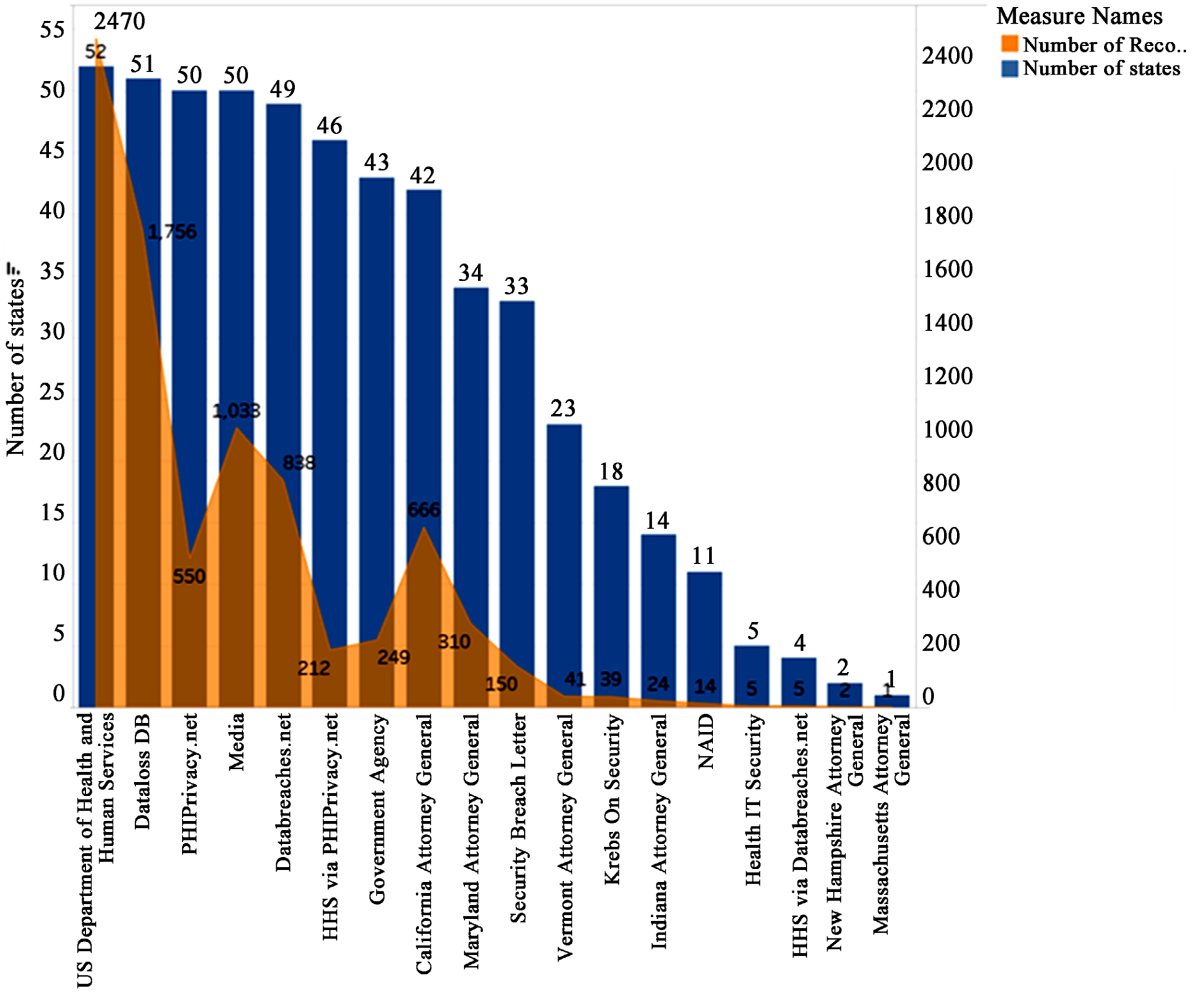

Figure 6. Agency/Information source of breaches by state.

Organization type and breach type.

Figure 7 shows the overall distribution of the number of organizations and percentages, by type of breach as well as type of organization. Our sample comprises organizations of the following types: service & financial (BSF), retail (BSR), educational (EDU), government & military (GOV) military, healthcare (MED), nonprofits (NGO) and business & others (BSO). From the figure, we can see that healthcare organizations (MED) comprise 48.46% of the breaches followed by businesses and other organizations (BSO) and government (GOV). Among the breach types, hacking/malware (HACK) is the most prevalent type of breach comprising about 33% of the sample, followed by unintended disclosure (DISC) with about 24%, and physical loss (PHYS) with about 23%.

Figure 8 presents a breakdown of organization types within each type of breach. It illustrates that the healthcare sector (MED) is the most frequent target for data breaches of all types. When examining instances of physical loss (tagged as PHYS) and unintended disclosure (DISC), it becomes evident that an overwhelming majority of breaches have occurred within the healthcare sector (MED). In the case of hacking and malware, there seems to be a distribution of

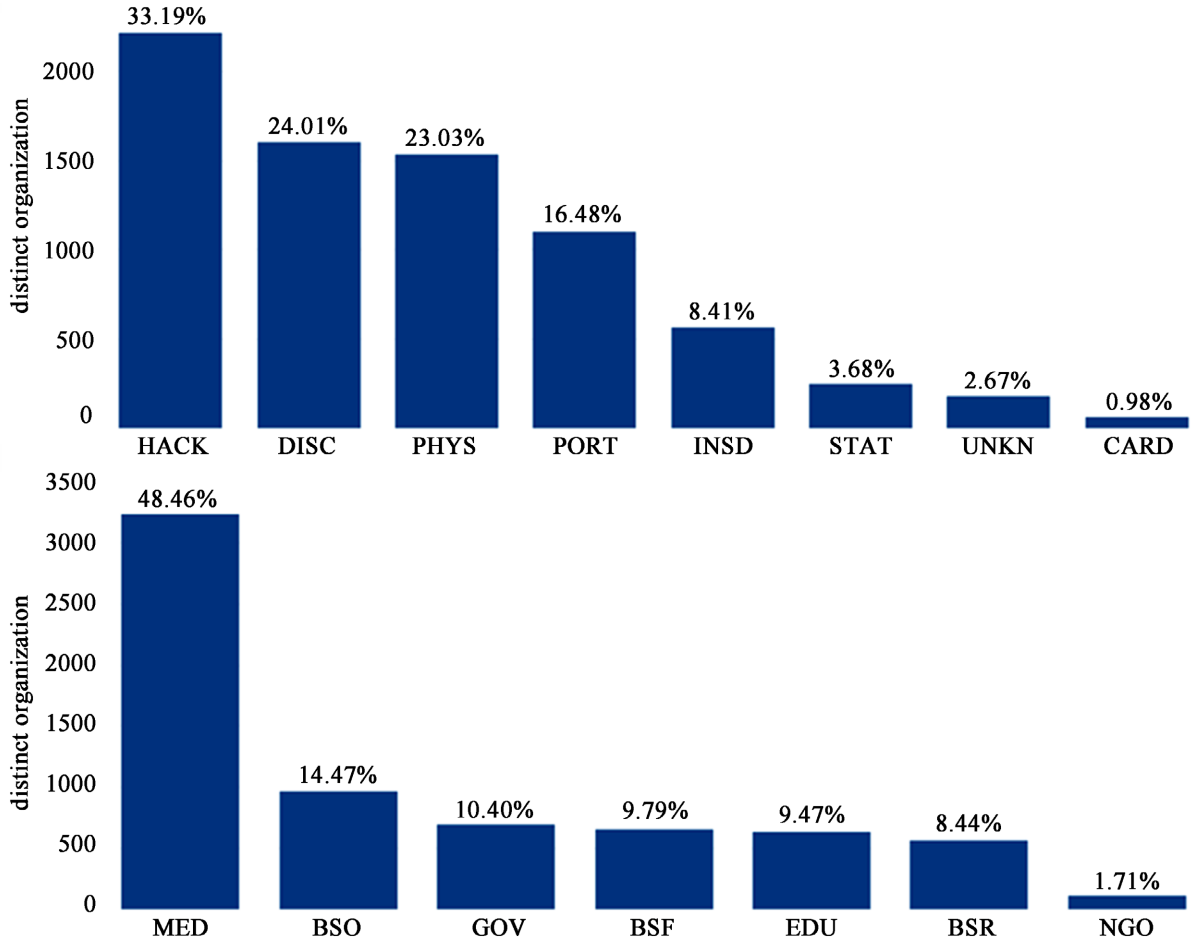

Figure 7. Number of distinct organizations by breach and organization type.

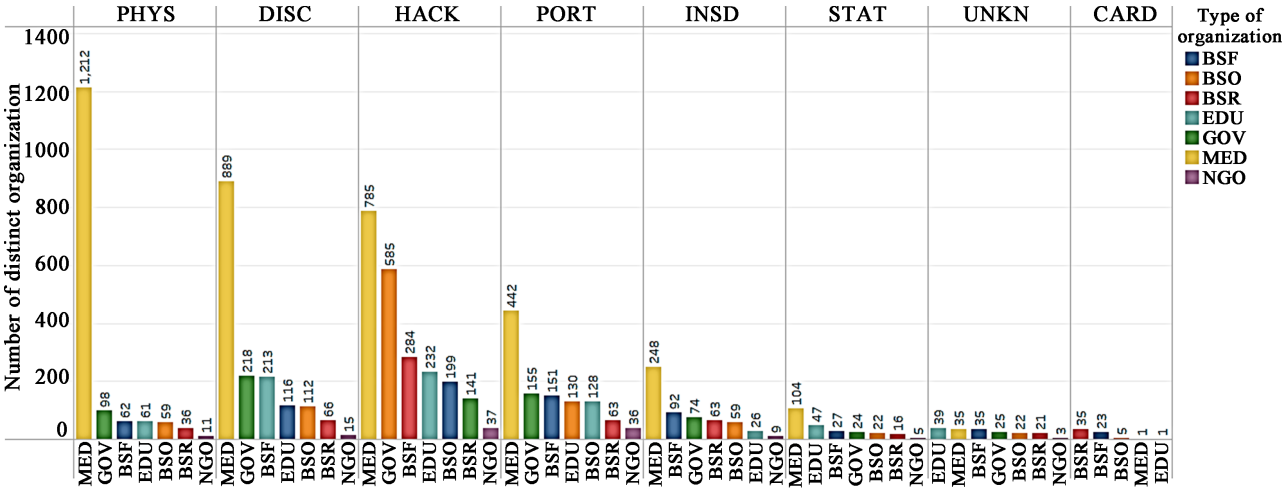

Figure 8. Breakdown of Organization type within each breach type.

breaches between healthcare organizations (MED), businesses and others (BSO) and retail (BSR). In general, breaches by loss of card (CARD) seem relatively less for all organization types.

Data breaches in Business and other Organization (BSO).

Figure 9 below shows the variation in total data breaches in the organization type 'business and other organizations' (BSO) between the years 2005 and 2018. There are two marked spikes in data breaches in the years 2013 and 2015. A dramatic decline follows each of these two years. In general, after 2013, data breaches in BSO show a rapid increase. This is parallel to variations in data breaches due to hacking/malware.

Role of Information Sources in Data breaches

Figure 10 shows the analysis of data breaches by information sources. The figure shows the total number of breaches reported by each source, as well as the number of records compromised in each incident reported by the source. The bar graph on the left section provides the total number of breaches reported by each information source. The bar graph in the right section illustrates the number of records compromised in each breach reported by the source. The analysis indicates that the three predominant sources for reporting breaches are the U.S. Department of Health and Human Services, the Data loss database, and various Media outlets. Particularly noteworthy is the role of the media in this context: it not only reports a higher number of breaches compared to other sources, but it also reveals that a substantially larger volume of data is compromised in each incident reported through this channel.

The number of data breach incidents experienced by the top 20 companies.

Figure 11 displays the top 20 companies that have encountered the highest number of data breaches during the period under investigation, alongside a comparison with the total records impacted by these breaches. On the left, the panel represents the specific number of breaches each company has endured,

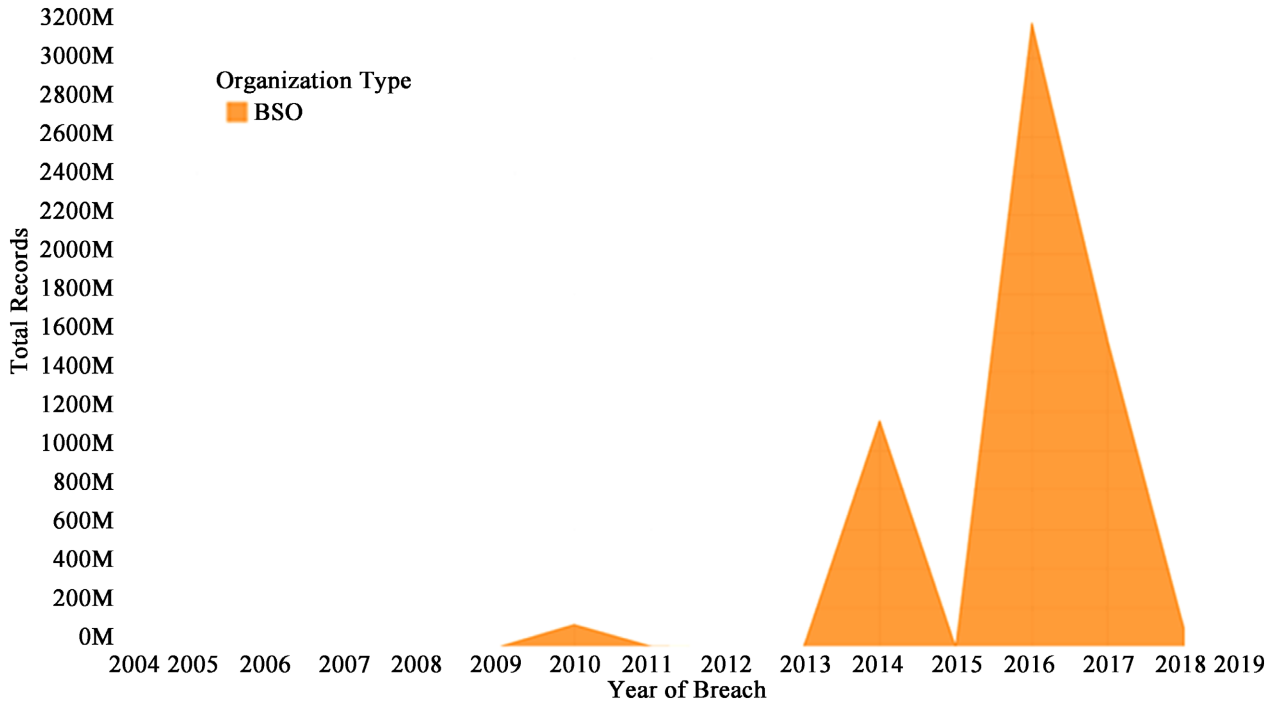

Figure 9. Total data breach records in BSO types by year.

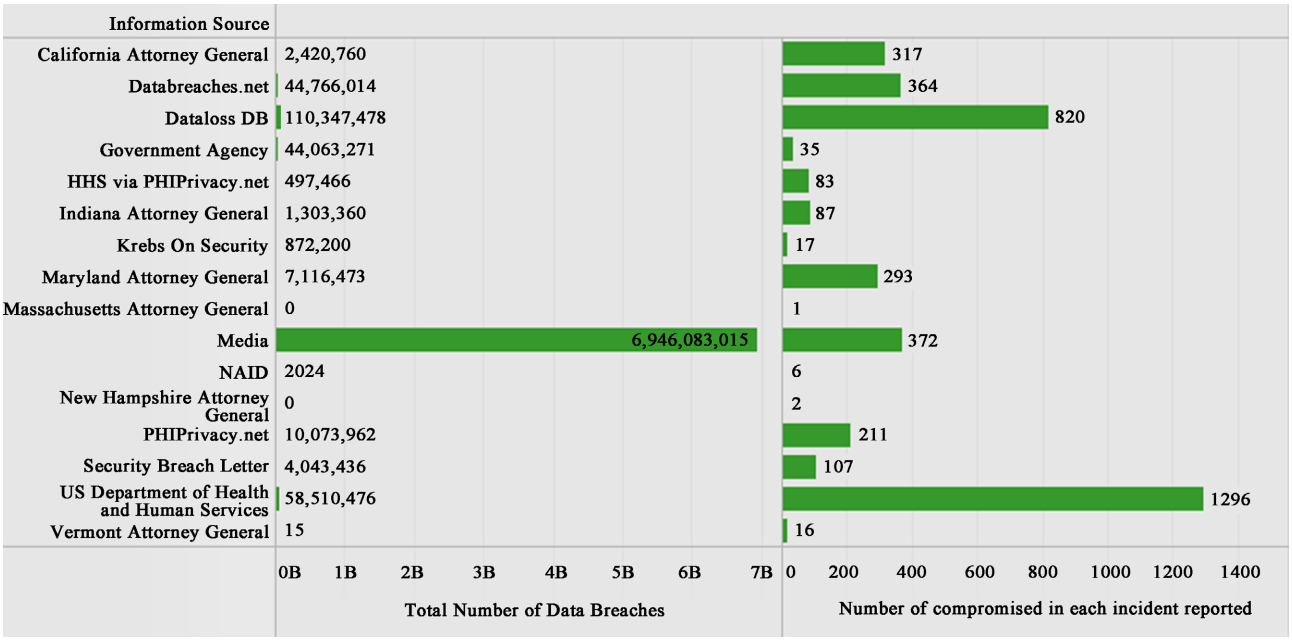

Figure 10. Total number of data breaches reported versus the number or records compromised in each incident reported by Information Source.

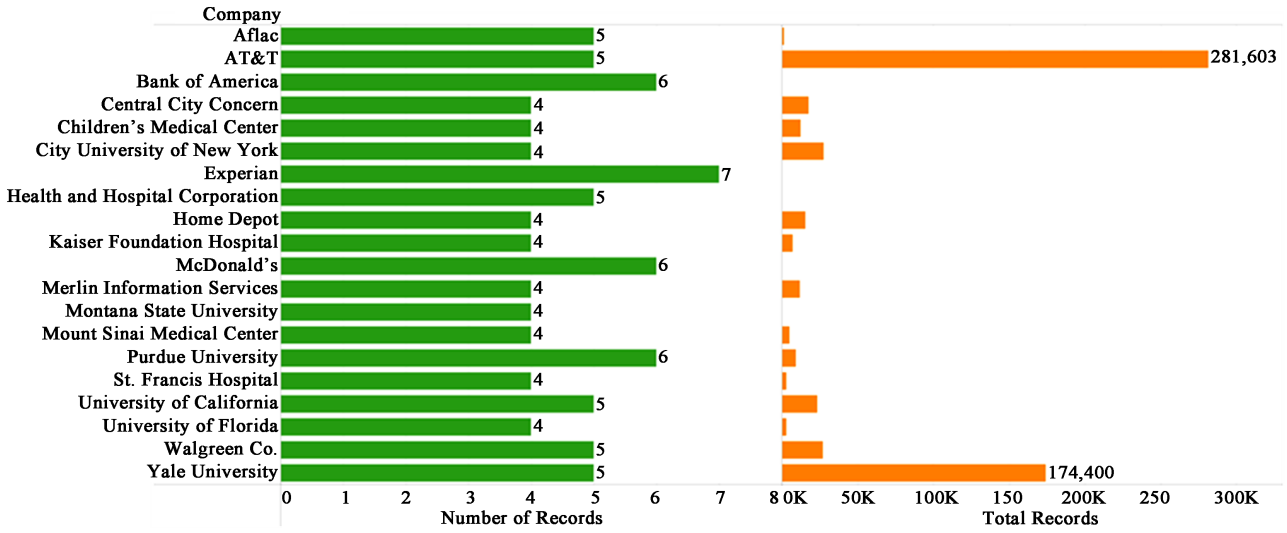

Figure 11. Top 20 companies with the largest number of data breaches.

whereas the panel on the right illustrates the cumulative records affected by these breaches. The magnitude of each occurrence is indicated by the length of the corresponding bars. Each company within this group has been subjected to a minimum of four distinct data breach incidents. Specifically, AT&T and Yale University have faced a substantial number of breaches.

The number of total records and number of breaches by organization and information source for hacking/malware.

Figure 12 compares the distribution and extent of hacking or malware-related (HACK) data breaches across various organization types and their respective information sources. The upper chart shows the number of breaches of different

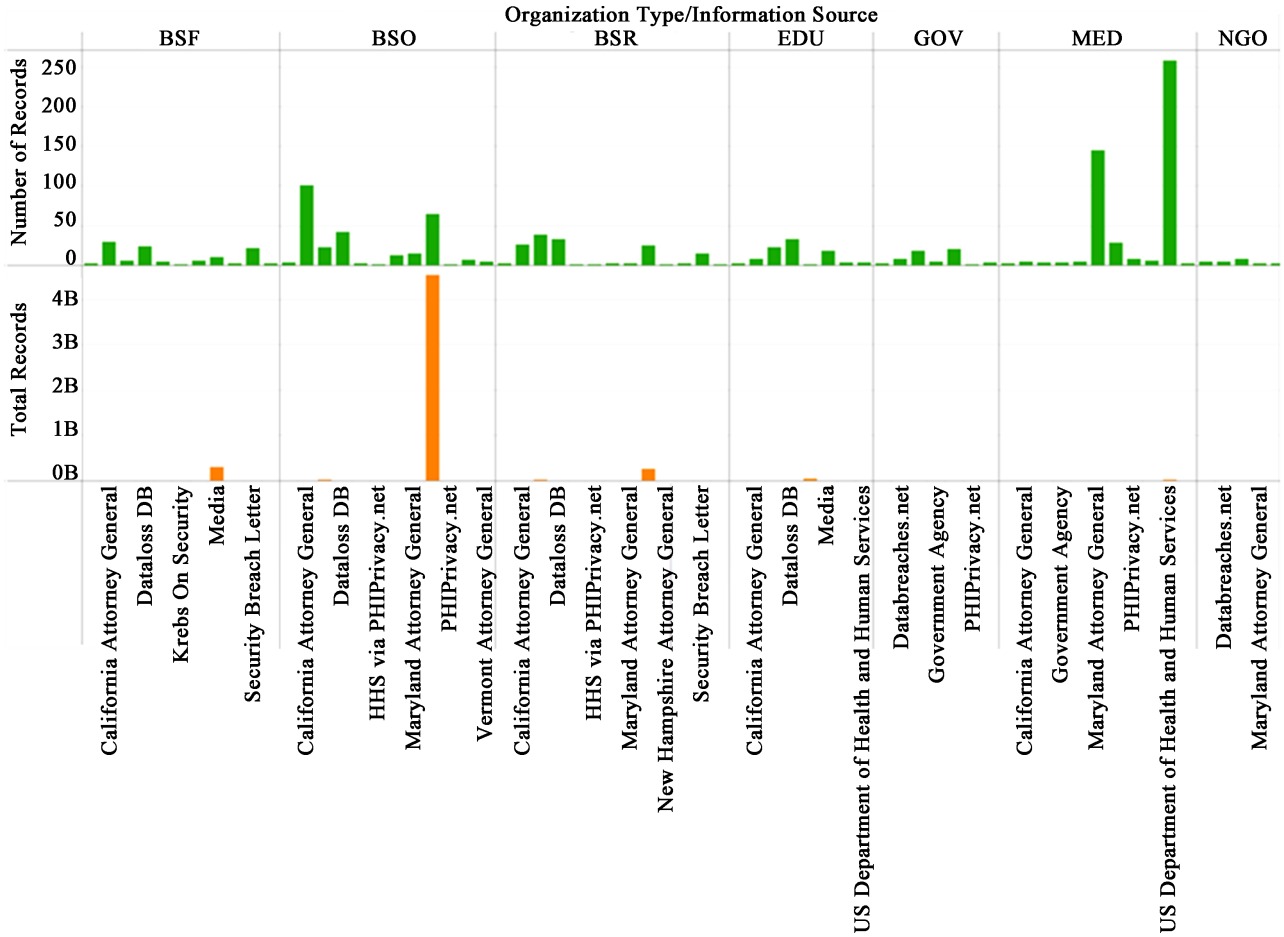

Figure 12. Total records and number of breaches by organization and information source for hacking/malware (HACK).

organization types and their information source. The lower chart displays the total records. The height of the bar represents volume. A significant finding presented here is the high prevalence of hacking/malware incidents (HACK) in the business sector (BSO), which has experienced the highest number of compromised records. Conversely, the healthcare sector (MED) has been the primary focal point for data breach incidents, with a majority of these breaches being unveiled by open-access government resources. Moreover, the data suggests that business sectors are often the targets of hacking/malware breaches, with media outlets serving as a notable channel for reporting these incidents.

The number of records by business organization by information source for hacking/malware.

Figure 13 illustrates the relationships between the nature of data breaches, the channels through which they are reported, and the types of organizations affected within the time frame under study. The graphic categorizes and identifies data breaches based on distinct criteria, represented along the 'type of breach' column which highlights different details pertaining to the information sources involved. A prominent feature of this figure is the green bar whose length corresponds to the number of records impacted in each type of breach. From the

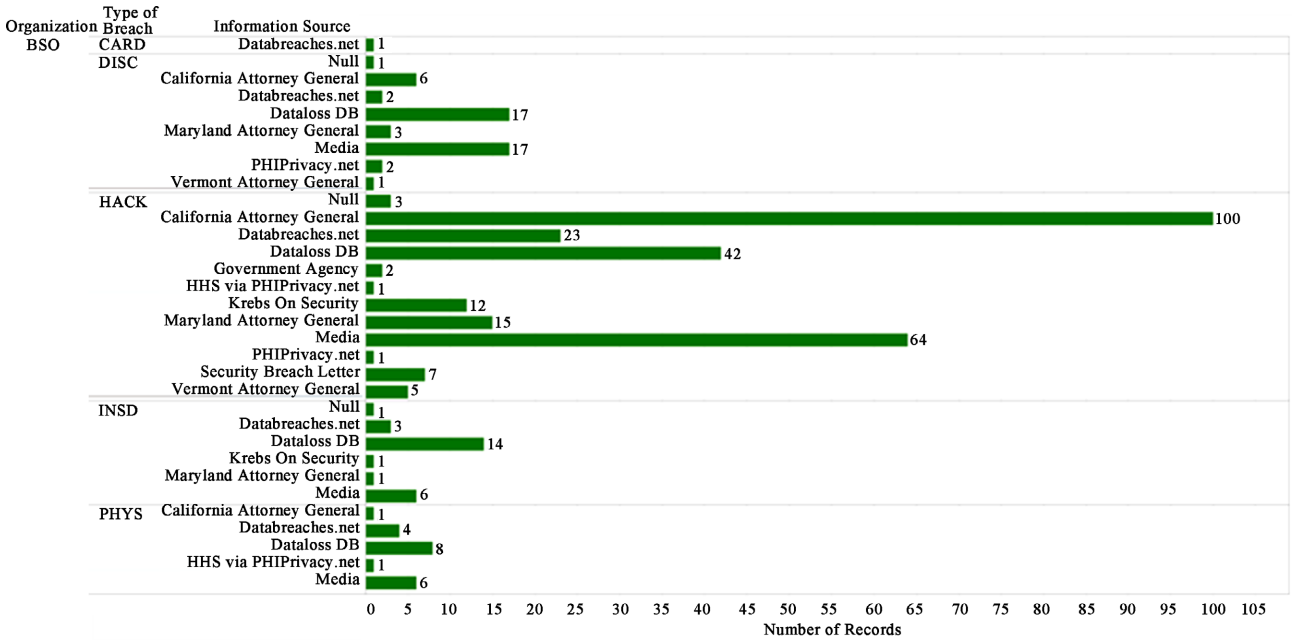

Figure 13. Total records and number of records of different organizations and information sources under HACK.

data visualized, it is evident that hacking or malware attacks (HACK) have compromised the greatest number of records during the examined period. Furthermore, the figure underscores the top three channels that have reported the majority of these breaches: the California Attorney General, various media outlets, and the Dataloss Database, listed in descending order of the number of reported incidents.

Number of distinct organizations by breach type by year.

Figure 14 portrays the fluctuations in the total number of distinct organizations affected by data breaches from the year 2005 to 2019. It is noticeable that there was an initial increase in the number of organizations impacted between 2005 and 2006. However, this upward trend was followed by notable declines during three distinct periods: 2006-2009, 2013-2015, and 2017-2019. Throughout this period, three primary causes of data breaches were consistently reported: unintended disclosure (DISC), physical loss (PHYS), and hacking/malware (HACK). These remained prevalent and significant contributors to the breach incidents across all years. Notably, from 2015 onwards, hacking/malware (HACK) emerged as the predominant type of breach, indicating a shift in the nature of security threats that organizations faced during this time.

6. Discussion

Our analysis shows the intricate connections between various types and records of data breaches across organizations in the United States. Upon detailed examination of data breach incidents across different states, we find that companies located in densely populated regions, including California, Maryland, Texas, New York, and Florida, are more susceptible to these breaches. Densely populated metropolitan areas such as the Bay Area and Los Angeles in California

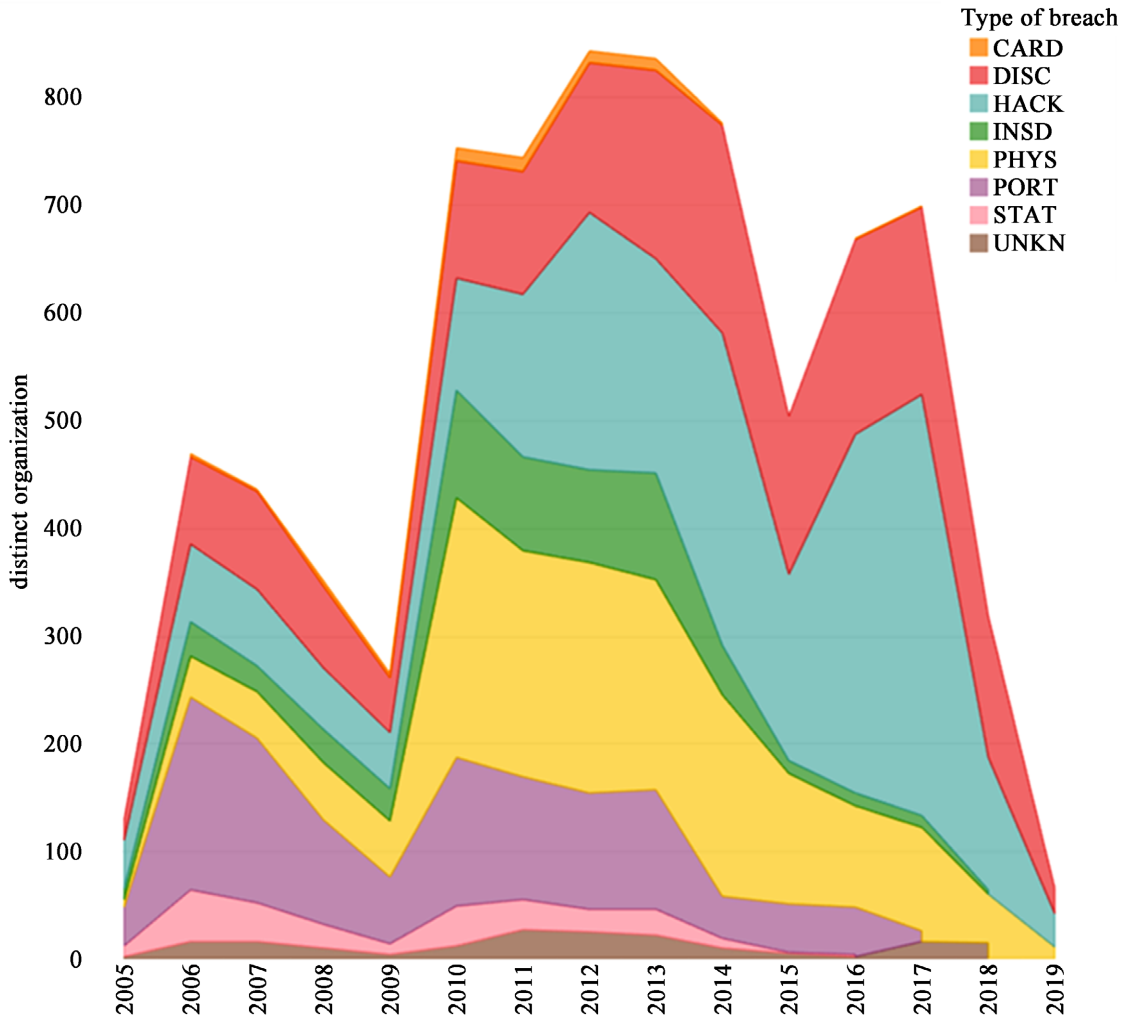

Figure 14. Number of distinct organizations by breach type by year.

witness a considerably higher incidence of data breaches.

Our research indicates that data breaches predominantly occur due to hacking/malware and physical loss, more so than other identified breach types. Notably, hacking/malware alone constitutes a staggering 75% of all breach types, underscoring the significant threat it poses to organizational data security. This finding calls for heightened vigilance and robust cybersecurity measures to mitigate the risks associated with these prevalent forms of data breaches.

Owing to advancements in technology and the pervasive utilization of big data, there has been a notable surge in data breach occurrences since 2013, especially those involving hacking and malware attacks. Further examination suggests that after 2013, data breach incidents are more likely to be experienced by business-related organizations, as opposed to non-profit organizations. This pattern suggests that the primary motive behind these breaches are financial gains. To substantiate this hypothesis, we analyzed the twenty companies most impacted by data breach incidents. Noteworthy observations include the significant volume of data breach records associated with AT&T and Yale University.

Furthermore, organizations such as Yahoo, River City Media, and the instances of Russian hacking uncovered by Hold Security were found to have the highest count of compromised records. In addition to the business service organizations, several sectors reported significant numbers of breaches. For instance, the U.S. Department of Health and Human Services, Dataloss (DB), and the media sector reported a large number of breaches among organizations within their scope. Particularly, the media sector has noted a larger volume of compromised records compared to other sources, underscoring the necessity for heightened security and vigilance in safeguarding sensitive data.

Our research also involves an in-depth analysis of variables including the number of data breach incidents and total records affected between 2005 and 2018. Notably, our investigation reveals that there is no correlation between the two variables during this period. We observe that a substantial portion of these data breaches can be attributed to hacking and malware infiltrations. Consequently, it becomes imperative for organizations to prioritize the fortification of their cybersecurity infrastructure, which includes regular updates to firewalls and software systems, to prevent these prevalent types of breaches. Our analysis also reveals that hacking and malware breaches constitute the primary form of cyberattacks that organizations need to address urgently. Healthcare organizations are the most targeted category, warranting special attention in this sector to bolster security measures. The most targeted state in USA is California, 92% of the distinct organizations suffered hacking/malware attack. Most of the observations are clustered in the Bay area and around Los Angeles. Lastly, while we do not investigate the underlying cause for each data breach, there is an interdependency between the presence of security vulnerabilities and human error. Ultimately, our study reveals that data breaches often materialize due to underlying vulnerabilities, which malicious actors capitalize on using advanced attacks or by triggering human errors.

7. Scope and Limitations

While this research offers a comprehensive descriptive study, it is not without limitations. First, the scope of the data is somewhat restricted. The available data, confined to a few specific years, solely encompasses breaches that occurred in the U.S., potentially offering a limited view of the broader global landscape of data breaches. Despite these constraints, we have managed to extract significant insights from the existing dataset. Second, our analysis is centered on a select set of variables pertaining to data breaches, implying that there might be additional contributing variables that could further enhance the depth of the research. Third, it should be noted that a considerable number of data breaches remain undetected, meaning that the existing records might not fully represent the current state of data breaches. Fourth, our data does not facilitate predictive analysis; hence, we confined our study to a descriptive analysis. Future studies could explore the temporal dynamics of breaches to augment record accuracy and pre-

dictive capabilities. Fifth, 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. Furthermore, although the research focused on the available dimensions of data breaches, it does not consider the demographic information regarding the impacted stakeholders. Information about the entity, as well as the affected individual (e.g., customer, employee, patient, etc.) could potentially unveil patterns indicating higher susceptibility to data breaches or discern trends that are drawn from specific categories of data breaches. Additionally, examining specific information related to the involved entities can enrich the analysis by establishing if there exists a significant relationship between the nature of the entity (for instance, a healthcare provider) and the likelihood of data breach affecting particular groups of individuals. Future studies can explore differences between the type of entity, location, breach type, and affected individual type.

8. Conclusions and Future Research

The primary objective of this descriptive study is to identify the various types and characteristics of organizational data breaches and, significantly, to formulate actionable insights on addressing this prevailing issue by utilizing publicly available data from the Privacy Rights Clearinghouse database. Specifically, we examine the relationships between the number of data breaches, characteristics of a breach type, the organization type, location (organization, city or state), the source of the breach, and the year in which it occurred. We also examine the nature of breaches (breach type) and their association with the organization (e.g., medical/healthcare provider), and location (e.g., organization, city or state). We obtain a glimpse of the trends in data breaches through our analysis of the reported data breaches.

Unsurprisingly, companies situated in populous states like California are more susceptible to data breaches. A closer examination at the state level reveals that instances of hacking/malware and physical loss are the most common types of breaches in this region. Most breaches occur through hacking and malware. Notably, there has been a marked uptick in hacking/malware data breach incidents post-2013, a trend propelled by the surge in big data technology utilization. Business organizations too witnessed a substantial increment in data breach incidents during the same period. It is important to note that the term "total incident records" diverges from "total records of data breach." In the analysis of data breaches, the healthcare sector seemingly bears a higher likelihood of experiencing breaches when compared to business entities. However, the volume of data breach records is considerably more substantial in the business sector. The likelihood of a data breach increases rapidly due to data and technology explosion and the availability of open-source software. It can be assumed the new technologies and advances in artificial intelligence (e.g., generative artificial in-

telligence) with further accelerate the quantity and quality of data breaches. Each data breach incident from media generates a large amount of breached data. Moreover, certain sources, including Dataloss (DB) are sources of reported breaches. When gaining insight into breaches in a company or industry, we should focus on total records rather than number of records. Quantifying the impact of a data breach is essential but challenging. This study sheds light on preventing, evaluating, and mitigating the risk of data breaches using big data analytics. This is critical to understand the evolving landscape of the data breach field.

The analysis provides a brief introduction to several characteristics of data breach incidents in the U.S. Allocating IT budgets to implement cybersecurity strategies is a starting point to becoming a modern security organization. Once an organization has these security fundamentals in place, it must measure and build on them with products, programs, and activities to successfully navigate the emerging and unseen risks of the digital era.

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, organizations, breach locations, breach types, and the year. Hacking is the most common type of data breach that significantly affects medical/healthcare organizations. Healthcare organizations, 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.

Data breaches have a detrimental effect on data privacy. This research found a possible correlation between population and affected individuals. CA experienced the most data breaches. Hacking was also predominantly found in other states, meaning organizations should focus on this type of breach. 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 monetary loss. Further research and insights can accelerate the maturing process of our understanding of data breaches.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Schlackl, F., Link, N. and Hoehle, H. (2022) Antecedents and Consequences of Data

Breaches: A Systematic Review. *Information & Management*, **59**, Article 103638. https://doi.org/10.1016/j.im.2022.103638

[2] Zadeh, A. (2022) Characterizing Data Breach Severity: A Data Analytics Approach. https://aisel.aisnet.org/treos_amcis2022/19

[3] Neto, N.N., Madnick, S., Paula, A.M.G.D. and Borges, N.M. (2021) Developing a Global Data Breach Database and the Challenges Encountered. *Journal of Data and Information Quality* (*JDIQ*), **13**, 1-33. https://doi.org/10.1145/3439873

[4] Einstein, M. (2019) Amazing Statistics about Online Data Creation.

[5] Reinsel, D., Gantz, J. and Rydning, J. (2018) The Digitalization of the World.

[6] Martin, A.P., Pogkas, D. and Mathieu, B. (2019) Ransomware Hackers Hit Brakes Worldwide, Leaving Mystery in Wake. *Computer Fraud & Security*, **2019**. https://doi.org/10.1016/S1361-3723(19)30034-X

[7] Goldberg, E. (2013) Preventing a Data Breach from Becoming a Disaster. *Journal of Business Continuity & Emergency Planning*, **6**, 295-303.

[8] Jeyaraj, A., Zadeh, A. and Sethi, V. (2021) Cybersecurity Threats and Organisational Response: Textual Analysis and Panel Regression. *Journal of Business Analytics*, **4**, 26-39. https://doi.org/10.1080/2573234X.2020.1863750

[9] Cremer, F., Sheehan, B., Fortmann, M., Kia, A.N., Mullins, M., Murphy, F. and Materne, S. (2022) Cyber Risk and Cybersecurity: A Systematic Review of Data Availability. *The Geneva Papers on Risk and Insurance-Issues and Practice*, **47**, 698-736. https://doi.org/10.1057/s41288-022-00266-6

[10] Akhtar, N., Tabassum, N., Perwej, A. and Perwej, Y. (2020) Data Analytics and Visualization Using Tableau Utilitarian for COVID-19 (Coronavirus). *Global Journal of Engineering and Technology Advances*, **3**, 28-50.

[11] Toasa, R., Maximiano, M., Reis, C. and Guevara, D. (2018) Data Visualization Techniques for Real-Time Information—A Custom and Dynamic Dashboard for Analyzing Surveys' Results. 2018 13*th Iberian Conference on Information Systems and Technologies* (*CISTI*), Caceres, 13-16 June 2018, 1-7. https://doi.org/10.23919/CISTI.2018.8398641

[12] Zhang, L., Stoffel, A., Behrisch, M., Mittelstadt, S., Schreck, T., Pompl, R., Weber, S., Last, H. and Keim, D. (2012) Visual Analytics for the Big Data Era—A Comparative Review of State-of-the-Art Commercial Systems. 2012 *IEEE Conference on Visual Analytics Science and Technology* (*VAST*), Seattle, 14-19 October 2012, 173-182. https://doi.org/10.1109/VAST.2012.6400554

[13] Sharma, N., Oriaku, E.A. and Oriaku, N. (2020) Cost and Effects of Data Breaches, Precautions, and Disclosure Laws. *International Journal of Emerging Trends in Social Sciences*, **8**, 33-41. https://doi.org/10.20448/2001.81.33.41

[14] Kilovaty, I. (2018) Data Breach through Social Engineering. *Harvard Law Review Blog*. https://ssrn.com/abstract=3216300

[15] Hsu, J.S.-C., Shih, S.-P., Hung, Y.W. and Lowry, P.B. (2015) The Role of Extra-Role Behaviors and Social Controls in Information Security Policy Effectiveness. *Information Systems Research*, **26**, 282-300. https://doi.org/10.1287/isre.2015.0569

[16] Johnson, M.E. (2008) Information Risk of Inadvertent Disclosure: An Analysis of File-Sharing Risk in the Financial Supply Chain. *Journal of Management Information Systems*, **25**, 97-124. https://doi.org/10.2753/MIS0742-1222250205

[17] Kwon, J. and Johnson, M.E. (2015) Protecting Patient Data-The Economic Perspective of Healthcare Security. *IEEE Security & Privacy*, **13**, 90-95. https://doi.org/10.1109/MSP.2015.113

[18] Lowry, P.B. and Moody, G.D. (2015) Proposing the Control-Reactance Compliance Model (CRCM) to Explain Opposing Motivations to Comply with Organisational Information Security Policies. *Information Systems Journal*, **25**, 433-463. https://doi.org/10.1111/isj.12043

[19] Cavusoglu, H., Mishra, B. and Raghunathan, S. (2004) The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers. *International Journal of Electronic Commerce*, **9**, 70-104. https://doi.org/10.1080/10864415.2004.11044320

[20] Malhotra, A. and Kubowicz Malhotra, C. (2011) Evaluating Customer Information Breaches as Service Failures: An Event Study Approach. *Journal of Service Research*, **14**, 44-59. https://doi.org/10.1177/1094670510383409

[21] Wong, P.C. and Thomas, J. (2004) Visual Analytics. *IEEE Computer Graphics and Applications*, **24**, 20-21. https://doi.org/10.1109/MCG.2004.39

[22] Culnan, M.J. and Williams, C.C. (2009) How Ethics Can Enhance Organizational Privacy: Lessons from the Choicepoint and TJX Data Breaches. *MIS Quarterly*, **33**, 673-687. https://doi.org/10.2307/20650322

[23] Tomaszewski, J.P. (2006) Are You Sure You Had a Privacy Incident? *IEEE Security & Privacy*, **4**, 64-66. https://doi.org/10.1109/MSP.2006.143

[24] Lane, V. and Wright, F. (1978) Human Resources Systematically Applied to Ensure Computer Security. In: Bracchi, G. and Lockemann, P.C., Eds., *Information Systems Methodology*, *Lecture Notes in Computer Science*, Springer, Berlin, 684-695. https://doi.org/10.1007/3-540-08934-9_105

[25] Loch, K.D., Carr, H.H. and Warkentin, M.E. (1992) Threats to Information Systems: Today's Reality, Yesterday's Understanding. *MIS Quarterly*, **16**, 173-186. https://doi.org/10.2307/249574

[26] Crossler, R.E., Johnston, A.C., Lowry, P.B., Hu, Q., Warkentin, M. and Baskerville, R. (2013) Future Directions for Behavioral Information Security Research. *Computers & Security*, **32**, 90-101. https://doi.org/10.1016/j.cose.2012.09.010

[27] Otto, P.N., Antón, A.I. and Baumer, D.L. (2007) The Choicepoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information. *IEEE Security & Privacy*, **5**, 15-23. https://doi.org/10.1109/MSP.2007.126

[28] Shropshire, J.D., Warkentin, M. and Johnston, A.C. (2010) Impact of Negative Message Framing on Security Adoption. *Journal of Computer Information Systems*, **51**, 41-51.

[29] Jiang, J.X. and Bai, G. (2019) Evaluation of Causes of Protected Health Information Breaches. *JAMA Internal Medicine*, **179**, 265-267. https://doi.org/10.1001/jamainternmed.2018.5295

[30] Quader, F. and Janeja, V.P. (2021) Insights into Organizational Security Readiness: Lessons Learned from Cyber-Attack Case Studies. *Journal of Cybersecurity and Privacy*, **1**, 638-659. https://doi.org/10.3390/jcp1040032

[31] Smith, T.T. (2016) Examining Data Privacy Breaches in Healthcare. Walden University, Minneapolis.

[32] Cheng, L., Liu, F. and Yao, D. (2017) Enterprise Data Breach: Causes, Challenges, Prevention, and Future Directions. *Wiley Interdisciplinary Reviews*: *Data Mining and Knowledge Discovery*, **7**, e1211. https://doi.org/10.1002/widm.1211

[33] Clearswift (2013) The Enemy within Research 2013.

[34] Wikina, S.B. (2014) What Caused the Breach? An Examination of Use of Information Technology and Health Data Breaches. *Perspectives in Health Information*

Management, **11**.

[35] Ayyagari, R. (2012) An Exploratory Analysis of Data Breaches from 2005-2011: Trends and Insights. *Journal of Information Privacy and Security*, **8**, 33-56. https://doi.org/10.1080/15536548.2012.10845654

[36] Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M. and El Koutbi, M. (2019) Digging Deeper into Data Breaches: An Exploratory Data Analysis of Hacking Breaches over Time. *Procedia Computer Science*, **151**, 1004-1009. https://doi.org/10.1016/j.procs.2019.04.141

[37] Rashid, A., Ramdhany, R., Edwards, M., Kibirige Mukisa, S., Ali Babar, M., Hutchison, D. and Chitchyan, R. (2014) Detecting and Preventing Data Exfiltration.

[38] IBM (2020) Cost of a Data Breach Report.

[39] Seals, T. (2015) Insider Threats Responsible for 43% of Data Breaches.

[40] Saleem, H. and Naveed, M. (2020) SoK: Anatomy of Data Breaches. *Proceedings on Privacy Enhancing Technologies*, **2020**, 153-174. https://doi.org/10.2478/popets-2020-0067

[41] Manworren, N., Letwat, J. and Daily, O. (2016) Why You Should Care about the Target Data Breach. *Business Horizons*, **59**, 257-266. https://doi.org/10.1016/j.bushor.2016.01.002

[42] Collins, J.D., Sainato, V.A. and Khey, D.N. (2011) Organizational Data Breaches 2005-2010: Applying SCP to the Healthcare and Education Sectors. *International Journal of Cyber Criminology*, **5**, 794.

[43] Posey Garrison, C. and Ncube, M. (2011) A Longitudinal Analysis of Data Breaches. *Information Management & Computer Security*, **19**, 216-230. https://doi.org/10.1108/09685221111173049

[44] Khey, D.N. and Sainato, V.A. (2013) Examining the Correlates and Spatial Distribution of Organizational Data Breaches in the United States. *Security Journal*, **26**, 367-382. https://doi.org/10.1057/sj.2013.24

[45] Shu, X., Tian, K., Ciambrone, A. and Yao, D. (2017) Breaking the Target: An Analysis of Target Data Breach and Lessons Learned.

[46] McLeod, A. and Dolezel, D. (2018) Cyber-Analytics: Modeling Factors Associated with Healthcare Data Breaches. *Decision Support Systems*, **108**, 57-68. https://doi.org/10.1016/j.dss.2018.02.007

[47] Algarni, A.M. and Malaiya, Y.K. (2016) A Consolidated Approach for Estimation of Data Security Breach Costs. 2016 2*nd International Conference on Information Management* (*ICIM*), London, 7-8 May 2016, 26-39. https://doi.org/10.1109/INFOMAN.2016.7477530

[48] Kafali, Ö., Jones, J., Petruso, M., Williams, L. and Singh, M.P. (2017) How Good Is a Security Policy against Real Breaches? A HIPAA Case Study. 2017 *IEEE/ACM* 39*th International Conference on Software Engineering* (*ICSE*), Buenos Aires, 20-28 May 2017, 530-540. https://doi.org/10.1109/ICSE.2017.55

[49] Sen, R. and Borle, S. (2015) Estimating the Contextual Risk of Data Breach: An Empirical Approach. *Journal of Management Information Systems*, **32**, 314-341. https://doi.org/10.1080/07421222.2015.1063315

[50] Holtfreter, R.E. and Harrington, A. (2015) Data Breach Trends in the United States. *Journal of Financial Crime*, **22**, 242-260. https://doi.org/10.1108/JFC-09-2013-0055

[51] Hall, A.A. and Wright, C.S. (2018) Data Security: A Review of Major Security

Breaches between 2014 and 2018. *Federation of Business Disciplines Journal*, **6**, 50-63. https://doi.org/10.1080/21624887.2017.1407596

[52] Goode, S., Hoehle, H., Venkatesh, V. and Brown, S.A. (2017) User Compensation as a Data Breach Recovery Action. *MIS Quarterly*, **41**, 703-727. https://doi.org/10.25300/MISQ/2017/41.3.03

[53] Raghupathi, W. and Raghupathi, V. (2021) Contemporary Business Analytics: An Overview. *Data*, **6**, Article 86. https://doi.org/10.3390/data6080086

[54] Börner, K., Bueckle, A. and Ginda, M. (2019) Data Visualization Literacy: Definitions, Conceptual Frameworks, Exercises, and Assessments. *Proceedings of the National Academy of Sciences*, **116**, 1857-1864. https://doi.org/10.1073/pnas.1807180116

[55] Keim, D., Kohlhammer, J., Ellis, G. and Mansmann, F. (2010) Mastering the Information Age Solving Problems with Visual Analytics. Eurographics Association.

[56] Keim, D.A. (2001) Visual Exploration of Large Data Sets. *Communications of the ACM*, **44**, 38-44. https://doi.org/10.1145/381641.381656

[57] Kohlhammer, J., Keim, D., Pohl, M., Santucci, G. and Andrienko, G. (2011) Solving Problems with Visual Analytics. *Procedia Computer Science*, **7**, 117-120. https://doi.org/10.1016/j.procs.2011.12.035

[58] Cook, K.A. and Thomas, J.J. (2005) Illuminating the Path: The Research and Development Agenda for Visual Analytics. United States, Pacific Northwest National Lab. (PNNL), Richland, WA (United States).

[59] Cao, N., Koch, S. and Gotz, D. (2018) ACM TIST Special Issue on Visual Analytics. *ACM Transactions on Intelligent Systems and Technology*, **10**, 1-4. https://doi.org/10.1145/3277019

[60] Lettieri, N., Guarino, A., Malandrino, D. and Zaccagnino, R. (2021) The Sight of Justice. Visual Knowledge Mining, Legal Data and Computational Crime Analysis. 2021 25*th International Conference Information Visualisation* (*IV*), Sydney, 5-9 July 2021, 267-272. https://doi.org/10.1109/IV53921.2021.00050

[61] Heer, J., Bostock, M. and Ogievetsky, V. (2010) A Tour through the Visualization Zoo. *Communications of the ACM*, **53**, 59-67. https://doi.org/10.1145/1743546.1743567

[62] Liu, S., Wang, X., Liu, M. and Zhu, J. (2017) Towards Better Analysis of Machine Learning Models: A Visual Analytics Perspective. *Visual Informatics*, **1**, 48-56. https://doi.org/10.1016/j.visinf.2017.01.006

[63] Yang, D., Xie, Z., Rundensteiner, E.A. and Ward, M.O. (2007) Managing Discoveries in the Visual Analytics Process. *ACM SIGKDD Explorations Newsletter*, **9**, 22-29. https://doi.org/10.1145/1345448.1345453