An Empirical Investigation of Data Breaches in the Healthcare Sector

A Thesis

Presented to

The Honors College, California State University, Los Angeles

In Partial Fulfillment
of the requirements for graduation from the

HONORS COLLEGE

by

Mayowa Toyinbo

May 2025

Approved by:

<Thesis Advisor Name>, Thesis Advisor

<Course Instructor Name>, HNRS Thesis Course Instructor

Dr. Andrea Arrias, Honors College Associate Director

ABSTRACT

An Empirical Investigation of Data Breaches in the Healthcare Sector

By

Mayowa Toyinbo

The incentivized integration of health information technology (HIT) as a core feature of U.S. healthcare made data handling more digitized, introducing the sector to digital threats from attackers. Data breaches have occurred every year in the healthcare sector, with the U.S. Department of Health and Human Services Office for Civil Rights (OCR) data of reported breaches going as far back as 2009 until today. This thesis aims to understand these breaches, how they are carried out, and where they occur. The thesis uses a quantitative and qualitative approach to analyze the data from various reputable sources, including the OCR and the Verizon Data Breach Investigations Reports, to develop findings and visualizations that answer my research questions about U.S. healthcare data breaches. The analysis of these data helped answer my questions by providing results that answer the what (mainly Hacking Incidents, Unauthorized Access, Theft, and Loss), the how (through network servers, over emails, portable electronic devices, desktop and laptop computers, and paper/film), the where (the U.S states), and the current time trend and forecast of incidents. The mitigations discussed in this thesis included addressing the threat of insiders, reevaluating existing policies, and employing technologies (including AI) to mitigate these breaches.

ACKNOWLEDGMENTS

I want to express my heartfelt gratitude to everyone who made the completion of this project possible. First, I sincerely thank my outstanding advisor, Dr. Shilpa Balan, for her invaluable guidance and support throughout this journey. I also appreciate the Honors College and Professor Baird, the instructor, for their encouragement and resources. Special thanks to Sonia for her assistance and to my family for their unwavering and immeasurable support. I am also grateful to my peer reviewers, Jeffrey and Gabriel, for their thoughtful feedback. Finally, I thank the university community for providing an environment that fostered learning and growth.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

Introduction and Overview

Introduction

Healthcare has always been a significant part of my life, both personally and professionally. With family members working in the healthcare field and working with people receiving long-term care as a Direct Support Professional, my early exposure to this industry sparked an initial interest in pursuing a pre-nursing degree. However, as I progressed in my studies, I developed a passion for technology, particularly in the field of cybersecurity, which led me to switch majors to Information Systems.

This thesis, *An Empirical Investigation of Data Breaches in the Healthcare Sector*, offers the perfect avenue to merge my interests by exploring the intersection of technology, cybersecurity, and healthcare. Aspiring to become a cybersecurity data analyst in the future, I see this project as an invaluable opportunity to gain insight into the dynamics of healthcare data breaches. By examining current trends, analyzing instances of data breaches, and proposing proactive and responsive solutions, this work aims to contribute to the ongoing efforts to protect sensitive healthcare data in an increasingly digital world.

Problem Statement

The Health Information Technology for Economic and Clinical Health Act (HITECH), signed into law on February 17, 2009, was enacted to promote the adoption and use of health information technology (U.S. Department of Health and Human Services, 2017). This act (HITECH) was signed with penalties for failure to comply.

The United States healthcare system has since experienced rapid growth in the adoption and use of health information technology, such as electronic health record (EHR) systems, which have been used to replace paper-based record systems with digital ones (Hayrinen et al., 2008).

Healthcare data has gotten more digital over the last decade, and information and communication technology advances have revolutionized the healthcare industry by providing more cost-effective and accessible services to its customers. Smartphones and other web-based smart devices empower users to quickly and conveniently access the online services offered by healthcare organizations. The Internet of Medical Things (IOMT), a network of smart medical devices that connect to the Internet to share health data, has also played an essential role in making data more available and enabling new data collection methods. Sensitive data collected from customers is stored by healthcare organizations on network servers to facilitate patient care. However, these developments in the healthcare industry come at a cost, as smartphones and other smart devices have become a common vector for threat actors in privacy breaches. Human error and software vulnerabilities sometimes lead to unauthorized users accessing this data and leaking sensitive information. At times, insider threats compromise protected health information, leading to the loss, theft, or unauthorized disclosure of critical healthcare data (Seh et al., 2020).

According to privacyrights.org (An organization that provides publicly available information on reported breaches), there have been over 8,500 publicly reported unique incidents of healthcare data breaches in the United States between 2008 and 2023 (Privacy Rights Clearinghouse, 2024).

One might ask, “How do healthcare data breaches affect patient care?” According to Sung Cho and M. Eric Johnson in their research article “Do Hospital Data Breaches Reduce Patient Care Quality?”, healthcare data breaches can negatively affect the delivery and availability of critical patient data to healthcare providers. They asserted that delays in patient information can disrupt the care process, especially in today’s information technology-driven healthcare delivery, and this can adversely affect patient care outcomes. The research analyzed “breaches reported to the HHS and the Privacy Rights Clearinghouse (PRC) database between 2011 and 2015” (Choi & Johnson, 2017, p. 7). It was discovered that Healthcare data breaches significantly increased the mortality rate for Acute Myocardial Infarction (heart attack). Data breaches disrupt care that relies on technology and cause a diversion of resources away from patient care, such as financial costs to repair a data breach (Choi & Johnson, 2017).

Purpose/goals of the project

This project aims to better understand the impact of data breaches on healthcare and the importance of data privacy and security by studying data breach cases in the United States healthcare, their implications for patient care, and recommending mitigations to reduce the occurrence of data breaches. Through a comprehensive analysis of these breaches, this thesis seeks to identify vulnerabilities within healthcare information systems, explore the challenges faced by healthcare organizations in maintaining robust security measures, and examine the regulatory landscape surrounding healthcare data protection. Additionally, the project will investigate how data breaches affect patient trust, organizational reputation, and financial stability. By doing so, I hope to provide actionable

insights to strengthen data security in the healthcare sector. Ultimately, the goal is to contribute to a safer and more secure healthcare environment where patient data is protected, and risks are minimized.

Significance

The significance of this project lies in its ability to address the growing problem of data breaches in the healthcare sector. As healthcare continues to embrace digital transformation, sensitive patient data becomes more vulnerable to cyberattacks. This thesis is important because it explores how these breaches impact patient care and the broader effects on healthcare organizations, such as financial costs and damage to trust and reputation. This project aims to fill a gap in understanding the patterns and risks specific to the healthcare industry by focusing on real-world data breach cases. Identifying these patterns can help improve cybersecurity practices in healthcare, ultimately leading to better protection of patient information and more robust overall security measures.

In terms of scholarly impact, this work contributes to information systems, healthcare, and cybersecurity by offering new insights into healthcare-specific vulnerabilities. It provides a detailed analysis of the current state of healthcare data breaches, helping other researchers and professionals develop more effective strategies to prevent and respond to such incidents.

Research Question(s)

This thesis aims to answer several key research questions related to healthcare data breaches in the U.S. It examines the most common attack vectors and types of cyberattacks that lead to data breaches in U.S. healthcare and how different types of breaches have evolved over time in the healthcare sector. Additionally, it explores the most frequently compromised sources in healthcare data breaches (e.g., network servers, emails, paper/films). The research also investigates which states have the highest number of individuals impacted by data breaches and which states experience the highest number of healthcare data breaches. Finally, it analyzes whether there is a correlation between the count of breaches in a state and the total number of individuals affected.

CHAPTER 2

Background

U.S. Healthcare System

The United States Healthcare system is a complex and crucial part of the country, and it can be described as a group of systems that operate independently and collaboratively in a free market. While many countries rely on a universal healthcare system funded and operated by the government, the U.S. healthcare system is based on the free-market model. The free-market model is one in which healthcare providers such as hospitals, physicians, and clinics are often privately owned, and individuals primarily pay for services through private insurance, public programs, and out-of-pocket expenditures (Rice et al., 2013). However, the structure and functioning of the U.S. healthcare system have evolved and continue to be shaped by technological advancements.

For this thesis, it is important to understand the structure of the U.S. healthcare system because it contextualizes the distinct structures of the system and how this affects the occurrence of data breaches. Factors influencing this include the system's decentralized nature and complexity, the varying roles of federal and state policies, and the dependence on health information technology (HIT) and its policy-driven adoption.

Public and Private Components

The public and private components are one of the main features of the U.S. healthcare system. The public components include programs such as Medicare, Medicaid, and the Children’s Health Insurance Program (CHIP), which provide insurance sponsored

by the government to specific populations, such as older people, low-income families, and children. These insurances help cover vulnerable populations that do not have an alternative or lack the financial capability to pay for their healthcare (Rice et al., 2013).

The U.S. healthcare system's private components comprise private health insurance companies that usually offer employer-sponsored insurance or individual plans, with the most common form of private insurance being employer-sponsored insurance, as most employed Americans get health coverage through their employers.

HIT – EHR, PHR

Health Information Technology (HIT) is important to modern healthcare systems as it manages, stores, and transmits critical information such as patient medical records, administrative data, payment information, and personal details. HIT comprises Electronic Health Records (EHR) and Personal Health Records (PHR), two essential systems for maintaining and accessing health-related information. Understanding these systems is vital for this thesis, as they form the foundation upon which the issue of healthcare data breaches is built.

Defining EHR and PHR

EHR, also known as Electronic Medical Records (EMR), is a digital repository of patient data managed by healthcare providers at the organizational level. It contains patient information, including medical history, diagnoses, treatment plans, and test results, for efficient care coordination across various departments. PHR is a patient-centered system where individuals manage and control their health information. PHRs aggregate data from

multiple sources, allowing patients to monitor their health independently and make informed decisions about their care (Rice et al., 2013). Both systems have revolutionized healthcare delivery by enabling seamless communication, improved record-keeping, and faster decision-making.

Historical and Policy Context

The history of EHR adoption reflects the evolution of healthcare delivery in the United States. The Regenstrief Institute in Indianapolis developed the first EMR in 1972, yet its adoption was initially hindered by high costs and a lack of standardization (Honavar, 2020). This changed with the passage of the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009 as part of the American Recovery and Reinvestment Act (ARRA). The HITECH Act incentivized healthcare providers to adopt EHR systems, offering grants of up to $44,000 per organization over five years (Rice et al., 2013). In 2016, a penalty phase further encouraged compliance, reducing Medicare and Medicaid reimbursements for providers failing to meet meaningful use requirements (Hansen & Baroody, 2020). These measures drove a nationwide shift toward digital health records, positioning EHR as a cornerstone of modern healthcare.

The Role and Impact of EHR in Healthcare

EHR systems address critical issues such as the lack of standardization across healthcare facilities, inefficient information retrieval, and knowledge erosion. They also enable a deeper understanding of treatment activities through tools like Diagnosis-Related Groups (DRG), which classify diseases and account for severity levels, allowing for

comparative analyses of treatment protocols and outcomes (Kohli & Tan, 2016). Furthermore, integrating EHR with other information systems, such as physician offices, clinics, and nursing care facilities, has expanded their utility beyond administrative tasks to support clinical and financial decision-making.

PHR and Privacy Concerns

While healthcare providers primarily manage EHR systems, PHR systems empower patients to take control of their health data. PHRs have become increasingly popular, with even non-health organizations like Microsoft (HealthVault) and Google (Google Health) entering the space. However, these systems raise significant concerns about data privacy and security. The debate over whether such platforms adequately protect sensitive patient information highlights the vulnerabilities inherent in digitized healthcare systems like EHR and PHR (Rice et al., 2013).

HIT and the Threat Landscape

The digitization of healthcare data introduces challenges to confidentiality, integrity, availability, and authenticity. Confidentiality ensures that patient data is protected from unauthorized disclosure, while integrity safeguards against unintentional changes. Availability refers to the system's ability to provide timely access to data, and authenticity ensures accurate identification of the data's origin (Ammenwerth & Hoerbst, 2010). As healthcare organizations increasingly rely on HIT systems, these principles become critical in mitigating data breaches and unauthorized access risks.

Understanding the evolution and significance of EHR and PHR systems is essential to contextualize the growing threat of data breaches in the healthcare sector. These systems, while transformative, remain vulnerable to attacks, underscoring the importance of robust cybersecurity measures to protect sensitive health information.

Trends in Healthcare Data Breaches

Prevalence and Nature of Data Breaches

Data breaches in the healthcare sector have been increasing over the years, and this is suggested to have been driven by the advancement in digital healthcare systems and their widespread usage. Seh et al. (2020) report that over 10 billion records were exposed across multiple sectors between 2005 and 2019, with healthcare accounting for 43.38% of these breaches. The primary causes of these breaches include hacking/IT incidents, unauthorized internal disclosures, loss, and theft. The most common cause of data breaches was hacking, which accounted for over 64% of the breached records during this period, with a significant increase in the last five years (2015-2019). Similarly, Dean (2023) emphasizes that hacking and IT incidents are the most common causes of breaches, followed by unauthorized access and improper data disposal. These findings align with Gabriel et al. (2018), who reveal that network servers are the primary breach locations.

Economic and Operational Impacts of Breaches

The financial and operational consequences of data breaches in healthcare are significant. Seh et al. (2020) estimate that the average cost of a healthcare data breach in 2019 was $6.45 million, with each compromised record costing $429. This is significantly

higher than in other sectors. This financial burden diverts resources from patient care and leads to an increase in operational inefficiencies (Choi & Johnson, 2017). This happens because breaches disrupt healthcare processes reliant on information technology, leading to adverse outcomes such as increased patient mortality rates.

The reputational damage to healthcare providers is equally significant. Moffit et al. (2017) report that nearly 70% of patients would avoid providers who experienced a breach. Similarly, Kwon and Johnson (2018) demonstrate that cumulative breach events over three years lead to a substantial decline in outpatient visits and admissions, particularly in competitive markets.

Human Factors and Organizational Vulnerabilities

Human error and organizational practices are recurring themes in healthcare data breach literature. Yeng et al. (2019) identify the lack of security awareness among healthcare staff as a critical vulnerability. Insiders, whether through negligence or malicious intent, contribute to 59% of breaches. This aligns with Stachel et al. (2015), who apply the Actor-Network Theory (which is how both human and non-human entities interact within networks) to highlight the complex interplay of human actors and technological systems in creating vulnerabilities. Furthermore, it was noted that training and accountability are essential to mitigating these risks.

Connections to EHR and PHR Vulnerabilities

Electronic Health Records (EHR) and Personal Health Records (PHR) are central to modern healthcare systems, but they also serve as the primary targets for cyberattacks

due to the sensitive nature of the data they store. Seh et al. (2020) emphasize that EHR systems are particularly vulnerable to hacking and ransomware attacks, which can compromise large volumes of patient data. The interconnected nature of EHR systems across departments and facilities increases their susceptibility to unauthorized access and breaches.

While empowering patients to control their health information, PHR systems present unique challenges. As noted by Rice et al. (2013), these systems often lack robust security measures compared to EHRs managed by healthcare providers, making them attractive targets for attackers. Moreover, third-party applications that make use of PHRs can introduce additional vulnerabilities, highlighting the need for stringent security protocols.

Effects on Patient Care Outcomes and Organizational Trust

Data breaches have far-reaching consequences beyond financial losses, significantly affecting patient care outcomes and organizational trust. Choi and Johnson (2017) found that delays in accessing patient information due to breaches disrupt care delivery, leading to adverse outcomes such as increased mortality rates for conditions like Acute Myocardial Infarction (heart attack). These disruptions highlight the critical role of uninterrupted information flow in ensuring timely and effective patient care.

Organizational trust also suffers in the aftermath of a data breach. Moffit et al. (2017) report that patients often lose confidence in healthcare providers following breaches, with many choosing to switch providers. This erosion of trust not only impacts patient retention but also undermines the overall reputation of healthcare organizations,

further emphasizing the need for robust data security measures to protect patient information and maintain public confidence.

Mitigation Strategies

Technological Interventions

Advances in artificial intelligence and biometric technologies offer promising solutions for enhancing healthcare security. Biometrics refers to using unique physical or behavioral characteristics such as fingerprints, facial features, heartbeat patterns, or even the way a person types to verify identity. Unlike passwords or access cards, which can be lost or stolen, biometric traits are inherently tied to an individual, making them a highly secure authentication method. Yeng et al. (2019) recommend using algorithms such as K-Nearest Neighbors (KNN), which classifies new data based on similarities to known data, and Bayesian Networks (BN), which uses probability to analyze relationships between different security factors. These techniques help analyze healthcare staff's security practices and detect anomalies. Similarly, Segun and Olawale (2019) advocate for biometric authentication methods, such as fingerprint or retina scans, to secure access to sensitive medical records, reduce fraud, and improve operational efficiency. Singh et al. (2021) further highlight the effectiveness of biometric authentication in healthcare security. Their study proposes a cloud-based biometric system that integrates patient authentication with healthcare administration. The system ensures secure access to medical records by verifying users through biometric traits like signature dynamics and measuring factors such as writing speed and pen pressure. This approach enhances data security by reducing reliance on traditional passwords, which are vulnerable to attacks.

Policy and Governance

Effective policies and governance frameworks are vital for safeguarding healthcare data. Koyame-Marsh and Marsh (2014) stress the importance of compliance with regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the implementation of strategic data security plans. Gabriel et al. (2018) assert that regular audits, penetration testing, and robust incident response plans are critical for identifying vulnerabilities and minimizing breaches, all of which are a reflection of effective policies and governance frameworks.

Addressing Insider Threats

The "human firewall" concept, as discussed by Yeng et al. (2019), emphasizes the need to cultivate a security-conscious workforce through targeted training and incentivization. Stachel et al. (2015) further recommend stringent vendor management practices and accountability measures to reduce insider-related risks. Moffit et al. (2017) agree with this “human firewall” concept by highlighting that ‘human’ contribution to the design, operations, and maintenance of networks leaves room for errors, with many underlying causes of breaches being human error. They also emphasized developing a “security-minded culture” to help mitigate healthcare breaches.

Research Gaps

Despite significant advancements, notable gaps persist in the literature. Limited studies have explored the interplay between market dynamics and the financial impacts of

data breaches, as highlighted by Kwon and Johnson (2018). Additionally, the effectiveness of AI-driven security solutions in real-world healthcare settings remains underexplored (Yeng et al., 2019). Future research should address these gaps by incorporating these concepts over time in studies.

Conclusion

Healthcare data breaches represent a multifaceted challenge requiring a comprehensive approach that integrates technological, organizational, and policy-based solutions. By synthesizing existing research, this review highlights critical trends and gaps, providing a roadmap for future investigations into safeguarding healthcare information systems.

CHAPTER 3

Methodology

Overview of method/approach

The approach utilized in analyzing the data I obtained from various reliable sources (with the OCR report being the primary dataset) included quantitative and qualitative methods. I used statistical data analytics and visualization tools (Tableau and Power BI) to form visual connections within my variables in order to visually explain my findings and answer my research questions. The methodology ensures a structured and data-driven approach to understanding data breaches in the United States healthcare sector.

Data Sources

Data for this research was obtained from various reliable sources, which are categorized below for qualitative and quantitative analysis:

1) Qualitative Analysis

    - The Verizon Data Breach Investigations Report (DBIR): This is an annual breach information report by Verizon (a U.S. telecommunication company). The report analyzes and provides insights into data breaches.

2) Quantitative Analysis

    - The main set of data used for my thesis analysis was obtained from the U.S. Department of Health and Human Services Office for Civil Rights (OCR) Breach Portal. The data contains incidents of breaches that affect more than

500 individuals, which must be reported to the Secretary of the OCR, as required by section 13402(e)(4) of the HITECH Act. The reports provide data for data breaches under investigation and data in an archive for confirmed data breaches, which I used for my analysis. The archive data consists of data from 2009 to 2023 (5608 entries) in the first file and from 2009 to 2024 in the second file (5848 entries), which I used for time analyses. Some examples of variables in the OCR data include:

- Independent Variables (Factors influencing breaches)
  - Covered Entity Type (Healthcare Provider, Health Plan, Business Associate, etc.)
  - Type of Breach (Hacking/IT Incident, Unauthorized Access, Theft, etc.)
  - Location of Breached Information (Network Server, Email, Paper/Films, etc.)
  - Business Associate Present (Yes/No)
- Dependent Variable (Outcome being measured)
  - Number of Individuals Affected (How many people were impacted by each breach). This was also the only original quantitative variable in the OCR data.

- Kaggle: Data were obtained from Kaggle (a platform where data scientists, machine learning professionals, and developers can explore, analyze, and share quality data to collaborate and learn) using the search tab to find top

search results for the keyword (Healthcare Data Breaches). The data with usability scores greater than 5.0 were downloaded and examined for use. However, I discovered that these data were similar to the OCR report (Some users did specify that the data was indeed from OCR), containing the same column headers and data inputs in varying years. Some were from 2012 to 2022, while others included 2024/2025 data under investigation. This led to my decision to focus on cleaning and analysis of data from the OCR report.

The above sources are all generally accepted, reliable sources for data breaches. The data were used to conduct analysis, derive insights, answer various questions on healthcare data breaches, and provide mitigations.

Tools and Resources

The primary tool for analyzing and visualizing my findings was Tableau (a visual analytics software). Microsoft's Power BI (a tool that can be used to turn data into interactive visual insights) was also used to clean the OCR data before loading it. Some changes were made to the data in Power BI, including using the find and replace feature to rename the abbreviations of states to the full names of the states, for example, replacing CA in the original data with California for more readability.

Approach

The main goal of my thesis is to better understand how much impact data breaches have on healthcare, including how far-reaching these incidents are and how many people are affected by breaches, and ultimately propose proactive strategies and responsive

measures to strengthen data security in the U.S. healthcare sector. I was able to answer some of my research questions by reviewing past literature, research, and trusted reports on data breaches. I synthesized these ongoing discussions on the implications of data breaches on patients' health and patients' trust, as well as the financial and reputational impact data breaches have on healthcare organizations.

The steps taken in the methodology are listed below:

1) I downloaded and grouped my data in folders for easy access. The data were in the form of an Excel file for the OCR data, CSV for the Kaggle files, and PDFs for the OC3 and Verizon reports.
2) Then, I loaded, cleaned, and uploaded my data to the analytics software (Tableau and Power BI).
3) The next step to completing my analysis towards fulfilling my goal of the project was to transform the data by creating a new variable in the form of a calculated field in Tableau to count the breach locations to analyze the most common attack vectors and determine what location gets frequently breached in healthcare organizations since the only numerical variable/column in my data is the 'Individuals Affected column.'
4) The next step I took was to create a data group for the type of breach column because some entries have multiple and ambiguous data; for example, I grouped all the entries that have only 'Hacking/IT incident' as the data input or as the first of the multiple inputted e.g. 'Hacking/IT incident, Other,' or "Hacking/IT incident, Theft, Other, Unauthorized Access' I used the first data in the entry because 'Hacking/IT incident' also comes up as the second data or third data in

some other entries which had other data as the first word. This might seem like an avenue for error. However, this error will be minimal and will not affect the overall interpretation of the data because the significant groupings are of standalone data that have an already large amount of input count; for example, there were 2,781 entries for 'Hacking/IT incident' as the only data inputted with all the other entries grouped with that major grouping amounting for about six entries. I ended up with a grouped 'Type of Breach' variable with seven groups: Hacking/IT incident, Unauthorized Access/Disclosure, Theft, Loss, Improper Disposal, Unknown, and Other.

Some of the final steps I took to get my data ready for visualization was to change some data types to the appropriate types for my analysis, like changing the column with the date to a date data type with a year function for one of the visualizations, changing the date column to a quarter function for another visualization, and assigning a geographical role to my state column. Further, I made use of the embedded statistical functions in Tableau and Power BI to analyze my data, including the sum, average, count, and Forecast functions. The visualization I employed to explain my analysis and findings includes bar charts, line graphs, pie charts, and tables. I also utilized a qualitative and quantitative approach to understand and interpret the data. The qualitative analysis was used to determine some of the costs of data breaches to healthcare organizations, such as patient trust and organizational reputation.

CHAPTER 4

Results and Findings

The results and visualization derived from my data sets are presented in this chapter as numbered figures (Original or extracted), tables, charts, and statistical and narrative data. The findings and results are shown and described without in-depth explanation, as all visualizations will be interpreted and discussed in detail in the fifth chapter (Discussion and Conclusion), and recommendations for future research will also be provided. The research questions answered by the visualizations will precede the visualization(s) or statistical and narrative data addressing each question.

Research Question 1: What are the most common attack vectors and types of cyberattacks that lead to data breaches in U.S. healthcare?

Figure 1a: Most Common Type of Breaches (Count of Incidents)

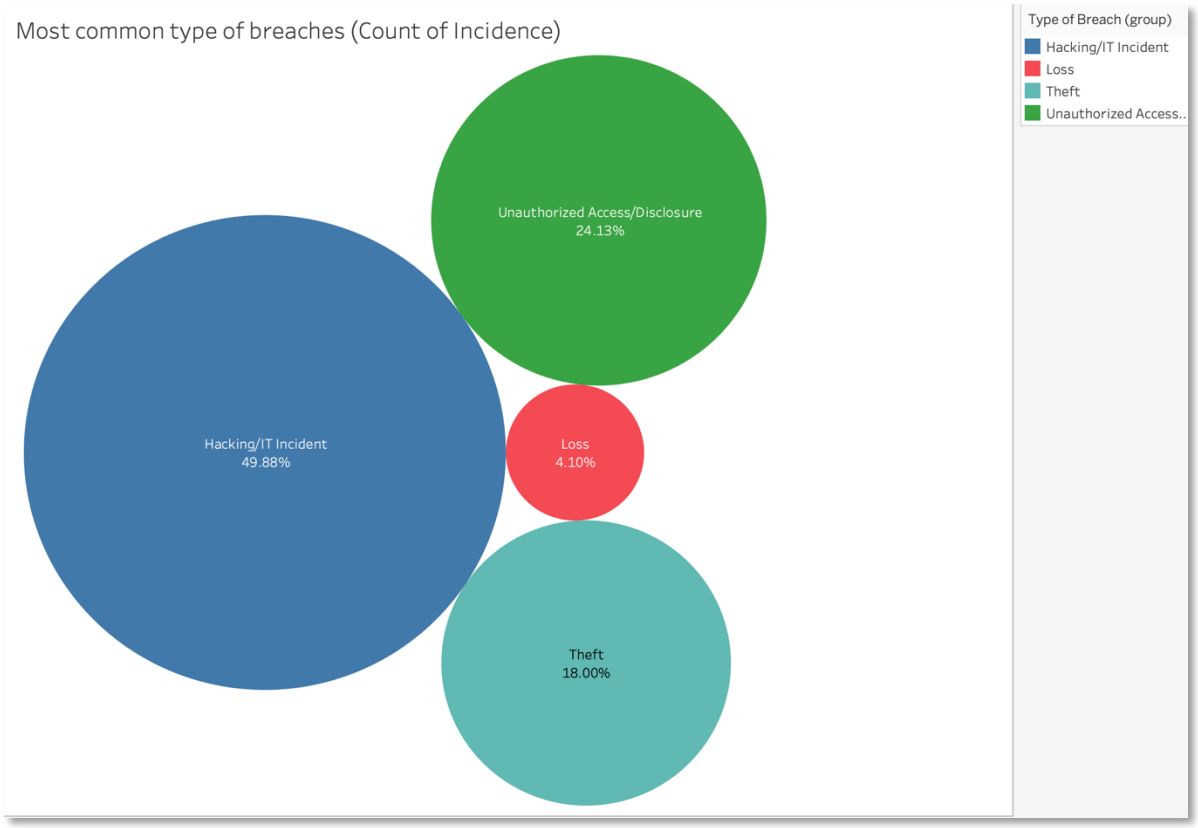


Figure 1a is a bubble chart showing the type of breach (group) and the percentages of the total breach instances. The color shows details about the breach type, while the bubbles' sizes show the number of breach instances using the 'Count' function to count reported breach incidents. The bubbles are labeled by types of breach and the percentages of the total breach instances. The data is filtered on breach types, which excludes Null, Other, and Unknown.

Table 1: Most Common Types of Breaches (Count of Incidents)

reportResultTable1 **Type of Breach (group)**	reportResultTable1 **% of Total Count of Breach Instances (Individuals Affected)**	reportResultTable1 **Count of Breach Instances...**
Unauthorized Access/Disclosure	24.1349%	1,353
Theft	17.9986%	1,009
Loss	4.1027%	230
Hacking/IT Incident	49.8751%	2,796

Table 1 displays the information in Figure 1a in a tabular format using Tableau's embedded 'view data' tool.

Figure 1b: Type of Breaches with the Most Individuals Affected

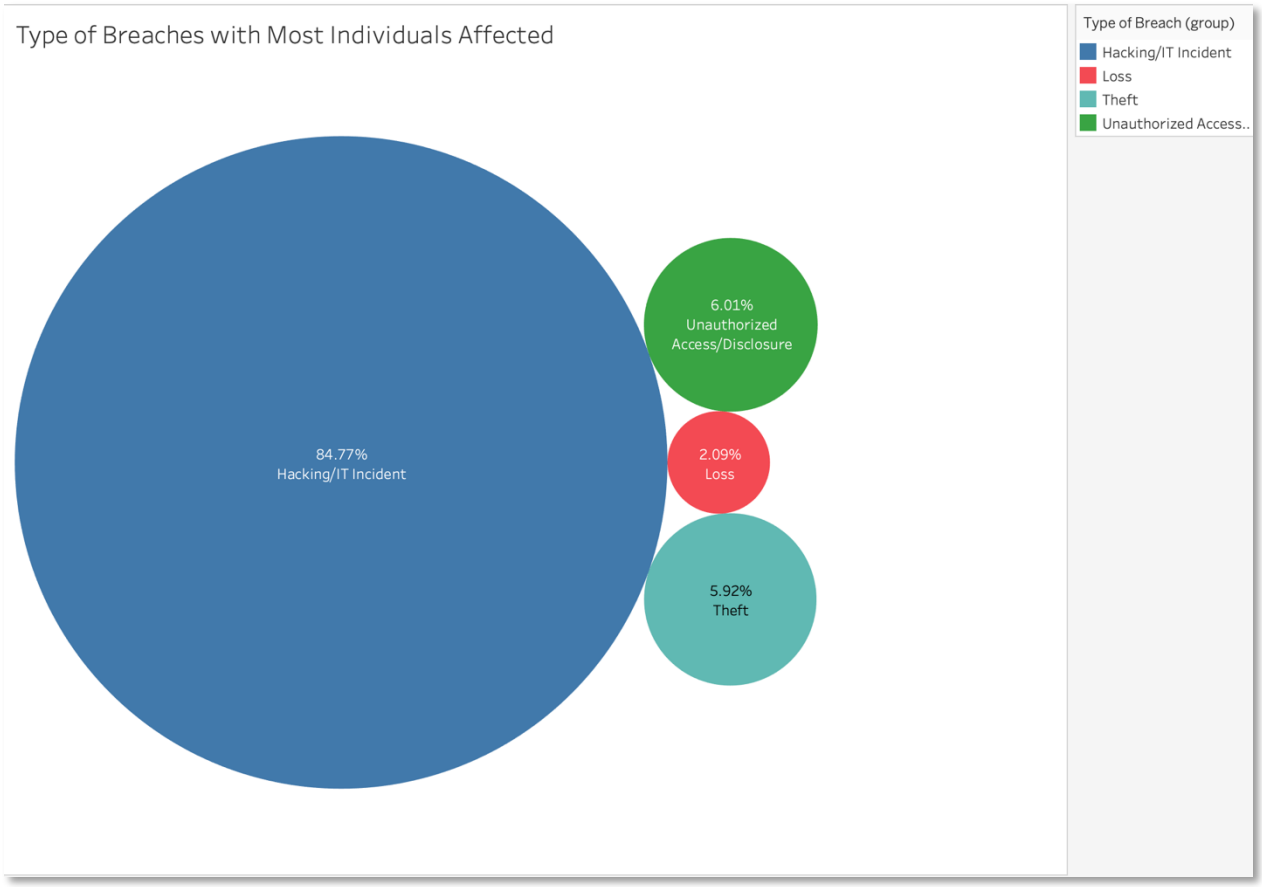


Figure 1b is a bubble chart showing the percentages of the total individuals affected and the types of breach (group). Similar to Figure 1a, the color shows details about the breach type, while the bubbles' sizes show the sum of Individuals Affected. The bubbles are labeled by types of breach and the percentages of the total sum of individuals affected. The data is filtered on breach types, which excludes Null, Other, and Unknown.

Table 2: Type of Breaches with the Most Individuals Affected

reportResultTable1 **Type of Breach (group)**	reportResultTable1 **% of Total Individuals Affec...**	reportResultTable1 **Individuals Affected**
Hacking/IT Incident	84.7738%	385,226,459
Unauthorized Access/Disclosure	6.0088%	27,304,924
Theft	5.9178%	26,891,630
Loss	2.0917%	9,505,230

Table 2 displays the information in Figure 1b in a tabular format using Tableau's embedded 'view data' tool.

Research Question 2: How have different types of breaches evolved over time in the healthcare sector?

Figure 2: Breach Type Trend

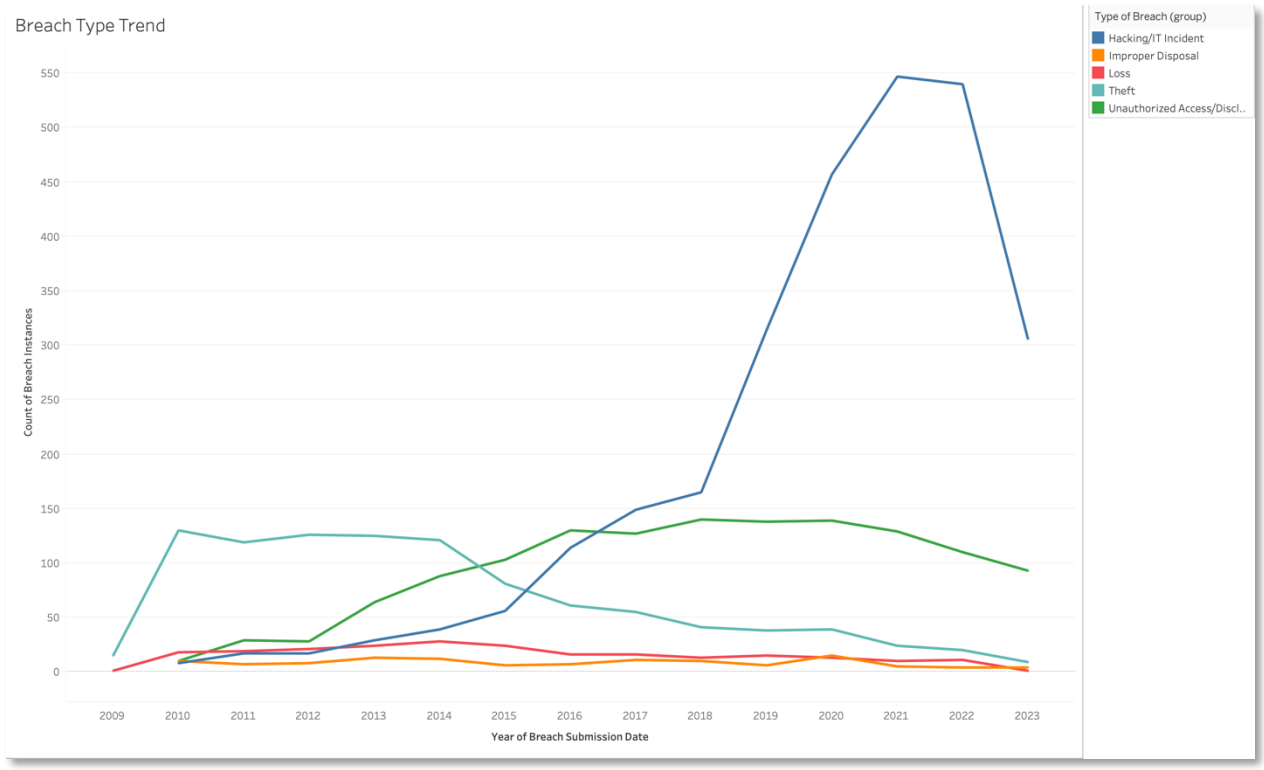


Figure 2 is a 'lines' chart that shows the trend of the count of breach instances (Individuals Affected). The trend line colors show details about the type of breach (group) - blue for Hacking/It incident, orange for Improper Disposal, red for Loss, turquoise for Theft, and green for Unauthorized Access/Disclosure. The view is filtered by type of breach (group) and breach submission date year. The type of breach (group) filter excludes Null, Other, and Unknown. The Breach Submission Date Year filter ranges from 2009 to 2023 (2024 was excluded due to the report not including data for November and December 2024).

Research Question 3: What are the most frequently compromised sources in healthcare data breaches?

Figure 3: Count of Breached Sources

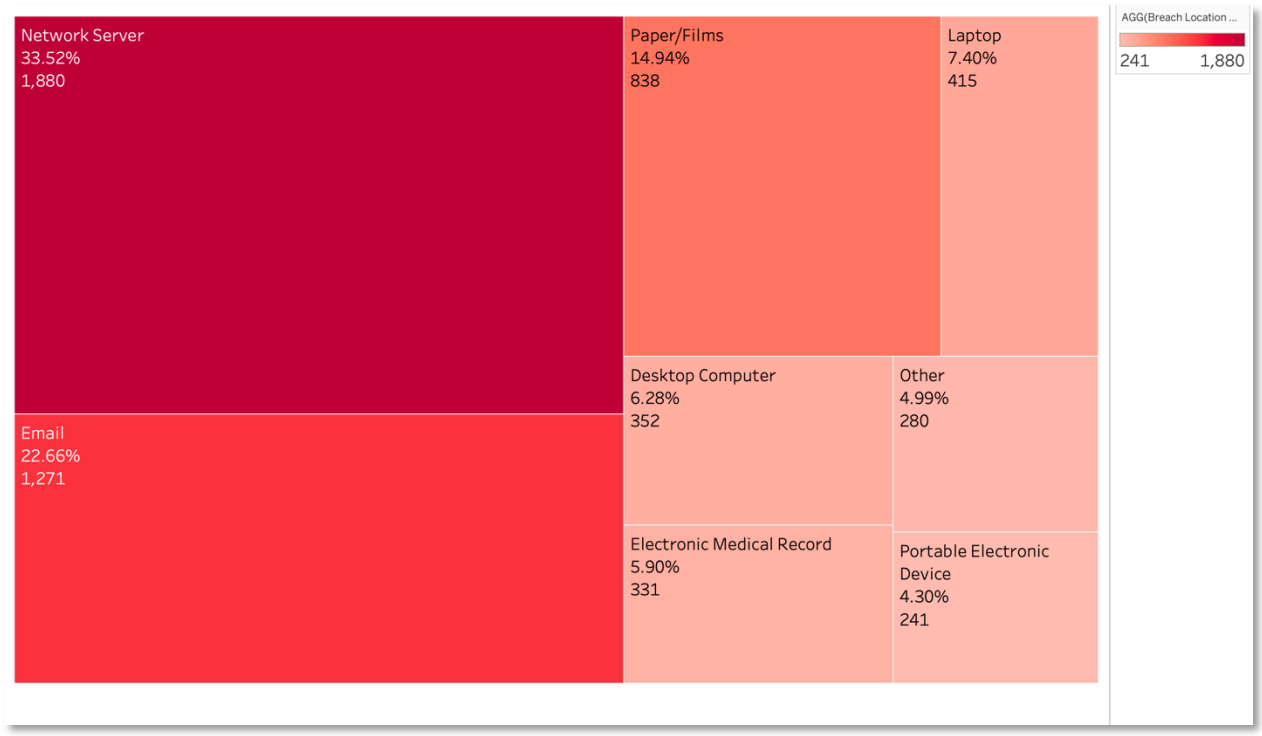


Figure 3 is a treemap showing the sources of breached information (group), the percentages of the total breach sources count, and the breach source count. The color shows the breach source count in a red gradient, and each box's size also indicates the breach source count, ranging from 241 to 1880. The marks are labeled by the source of breached information (group), percentage of total breach source count, and breach source count.

Research Question 4: Which states have the highest number of individuals impacted by data breaches?

Figure 4a: Sum of Individuals Affected by State

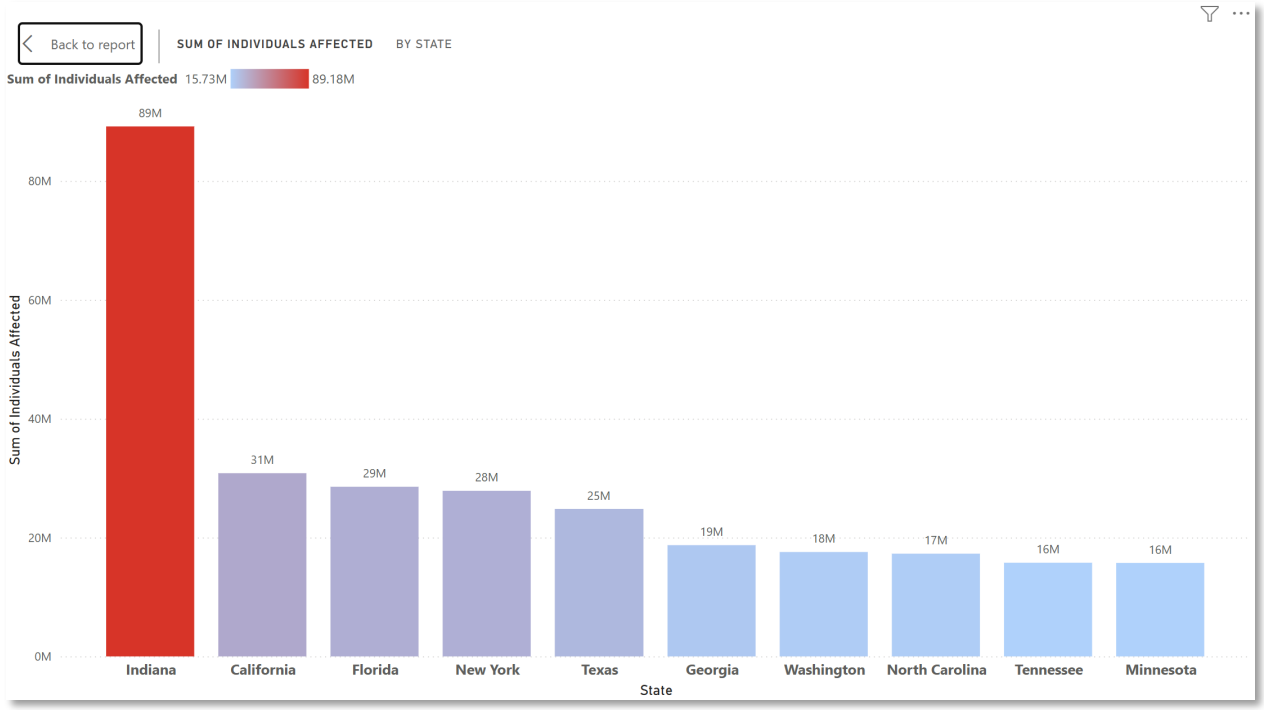


Figure 4a is a vertical bar graph showing the sum of individuals affected by data breaches for the top 10 states. The size of the candle indicates the sum of individuals in descending order, and the colors of the candles are in a gradient of red to blue, with bright red indicating the highest and faint blue the lowest.

Figure 4b Individuals Affected by State (zoomed out)

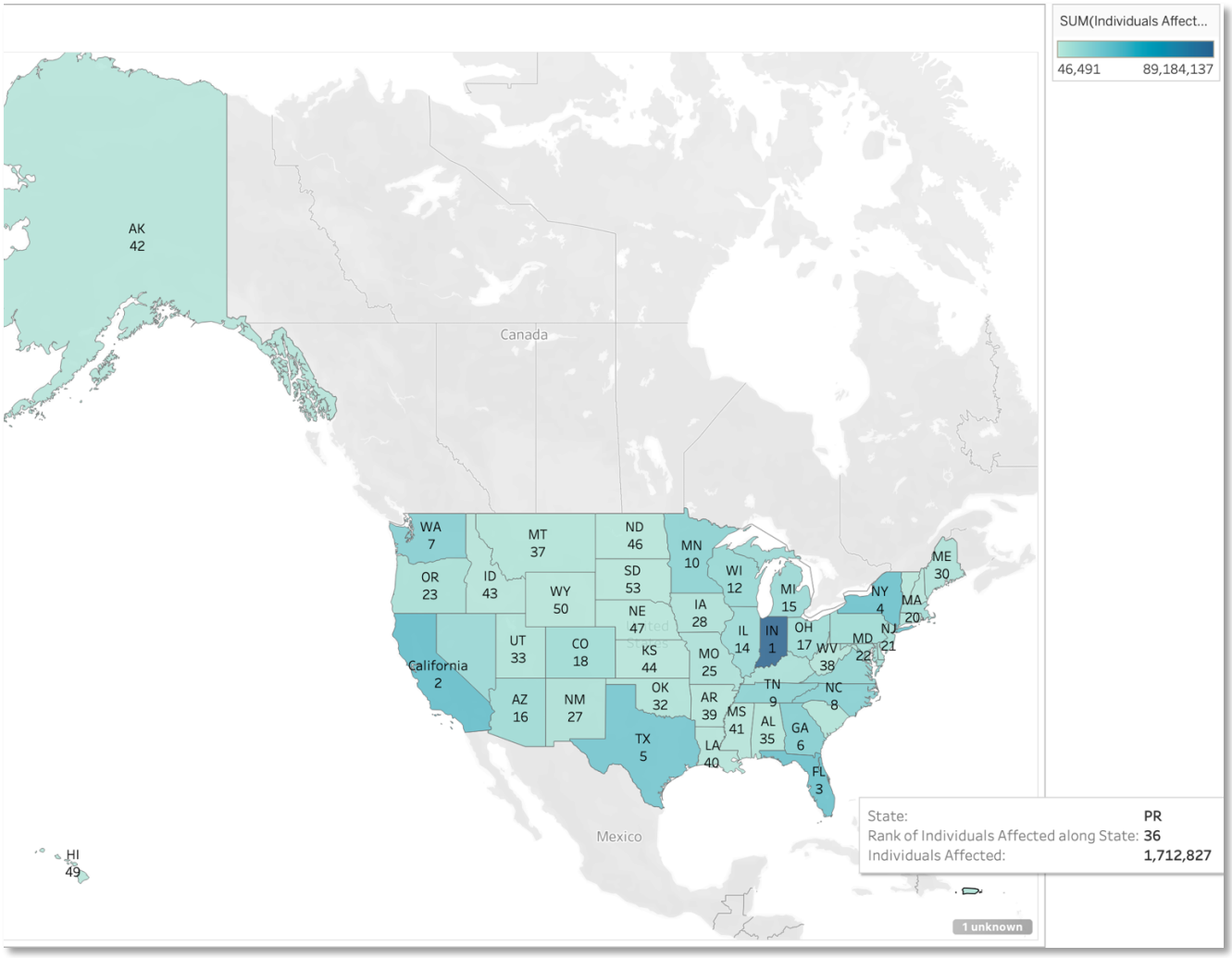
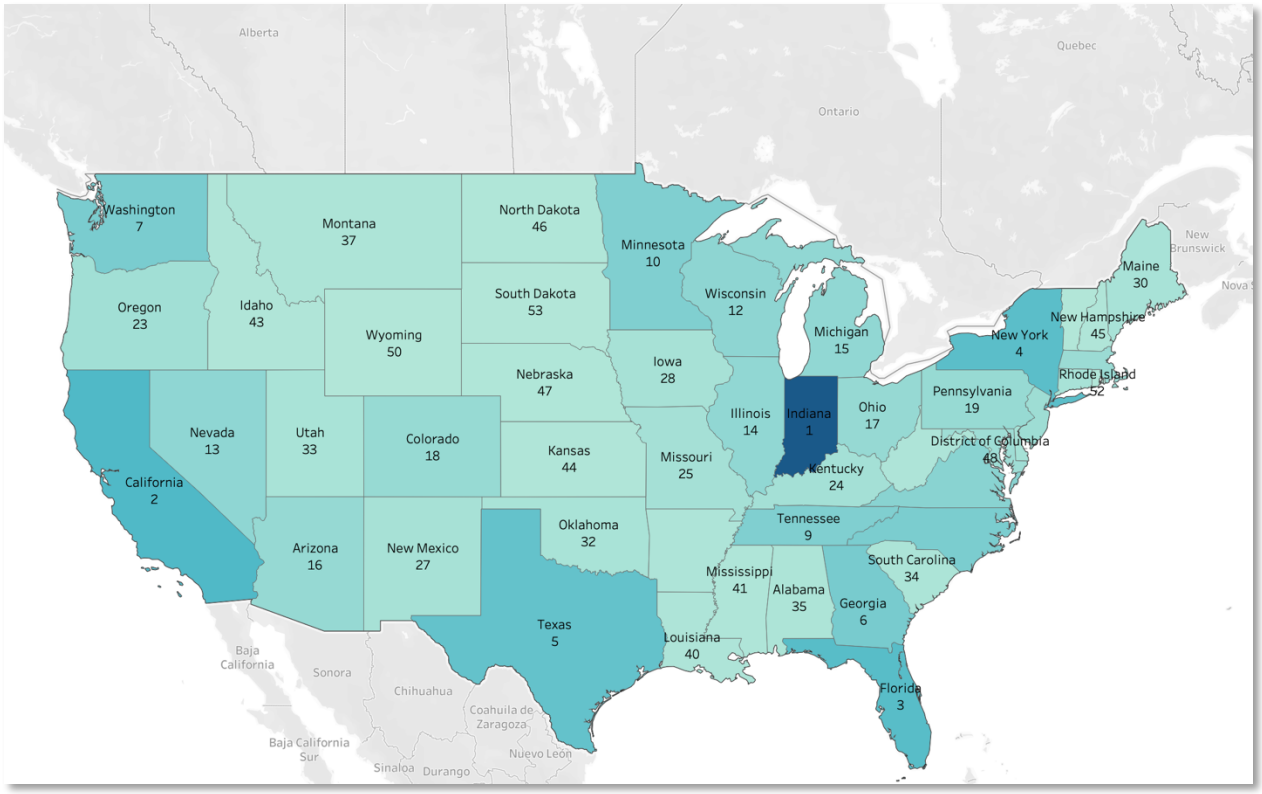


Figure 4c Individuals Affected by State (zoomed in)



Figures 4b and 4c are maps based on states in the United States showing the sum of individuals affected by breaches. The color shows the sum of individuals affected in a blue gradient from deep blue to light blue (46,491 to 89,184,137). The marks are labeled by State and Rank of Individuals Affected.

Research Question 5: Which states experience the highest number of healthcare data breaches?

Figure 5: Count of Incidents by State

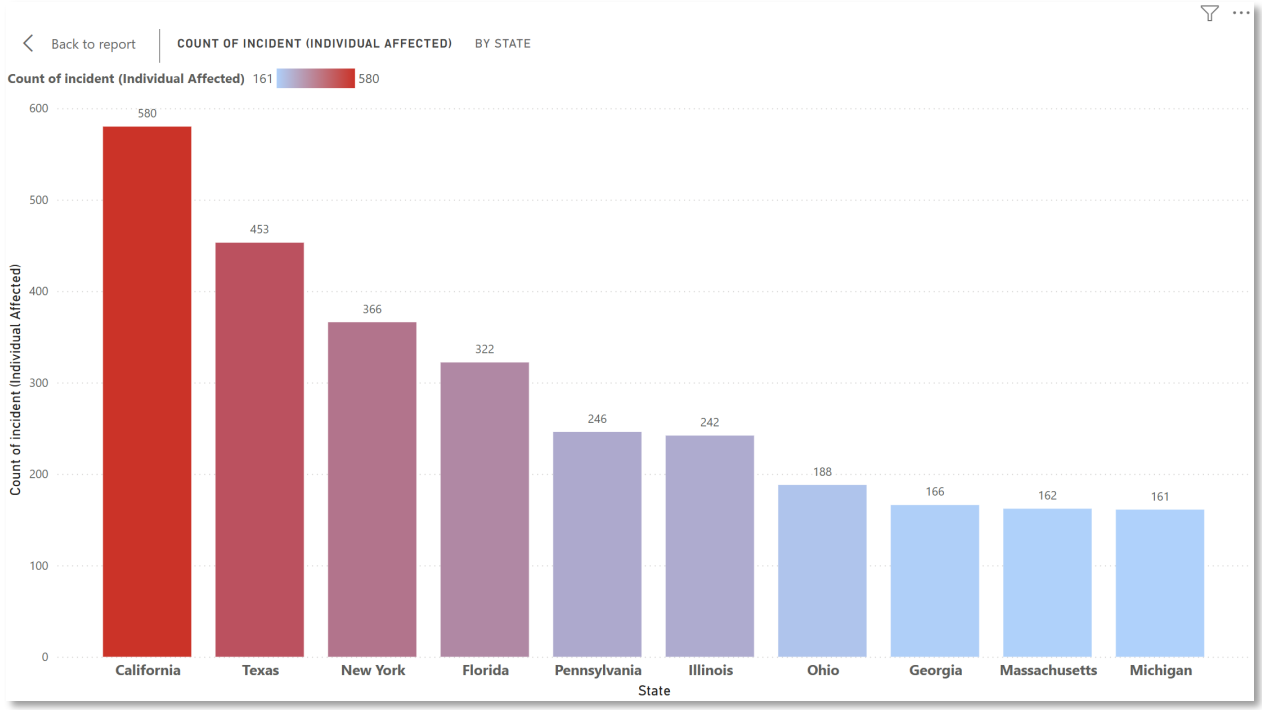


Figure 5 is a vertical bar graph showing the count of incidents of data breaches for the top 10 states. The size of the candle indicates the count of individuals in descending order, and the colors of the candles are in a gradient of red to blue, with bright red indicating the highest count and faint blue the lowest.

Research Question 6: Is there a correlation between the count of breaches in a state and the total number of individuals affected?

Figure 6: Count of Breaches and Total Individual Affected Relationships

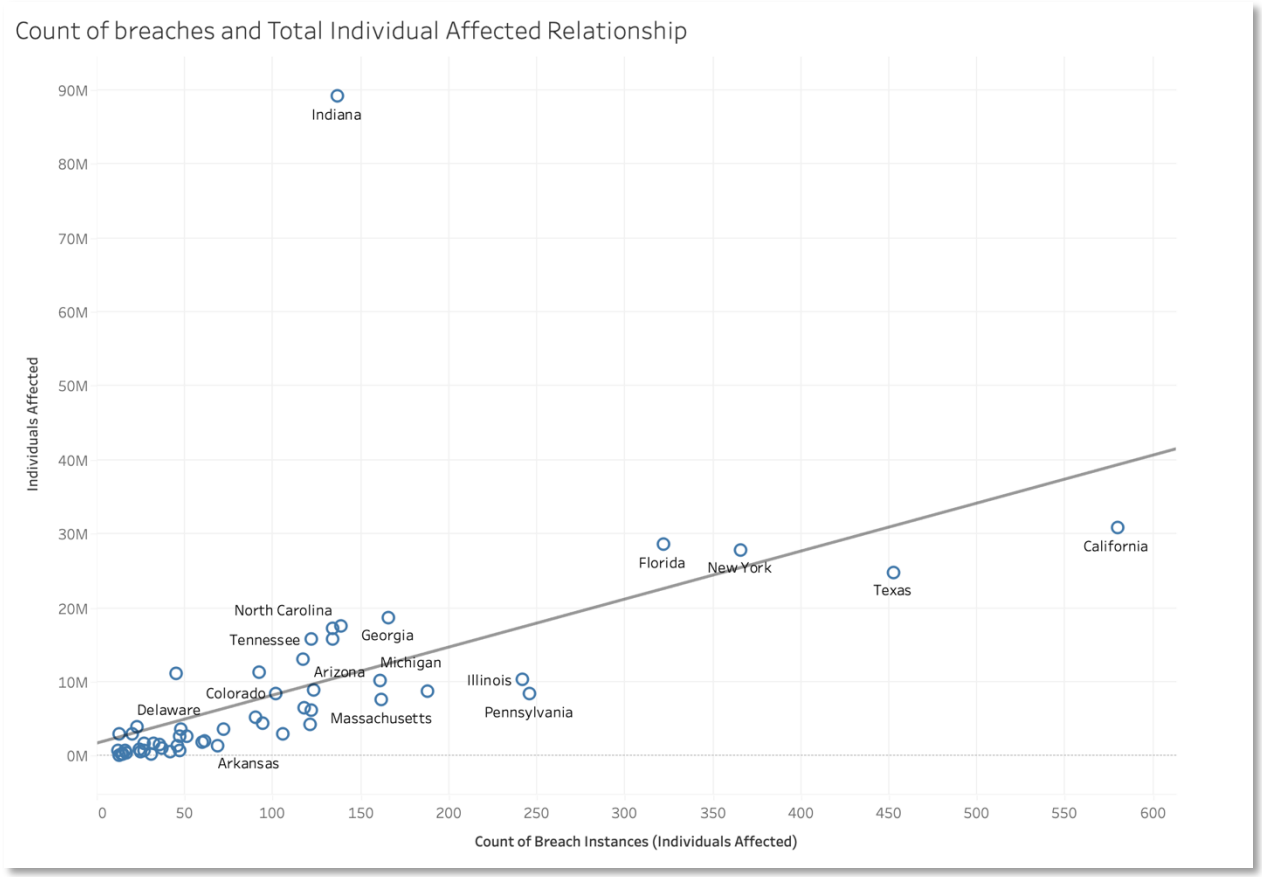


Figure 6 is a scatter plot of the Count of Breach Instances (Individuals Affected) vs. the sum of Individuals Affected. The marks are labeled by State. A trendline function is also applied to the plot.

Research Question 7: How have the number of breach incidents and the total number of individuals affected changed over time, and what trends or future projections can be identified in healthcare data breaches?

Figure 7a: Breach Time Analysis Count of Breach Incidents Actual

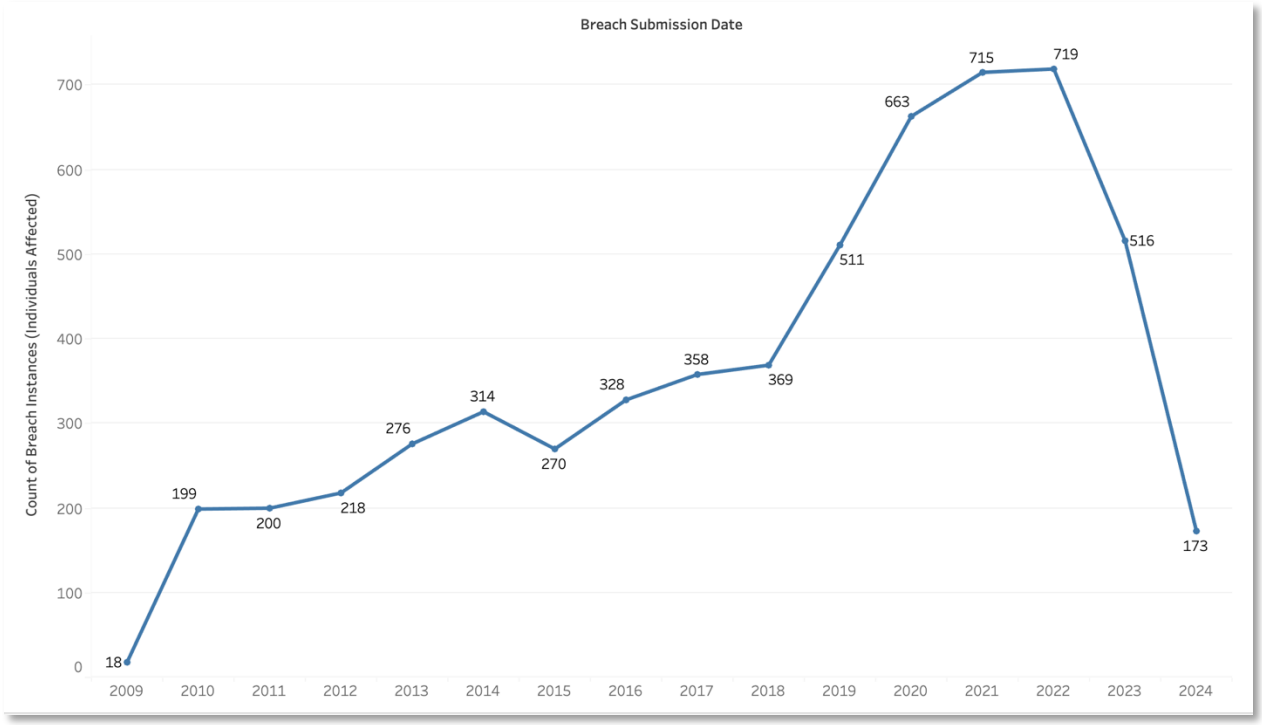


Figure 7a is a line graph showing the count of breach instances from 2009 to 2024. The marks are labeled by the count of breach instances.

Figure 7b: Breach Time Analysis Count of Breach Incidents Forecast

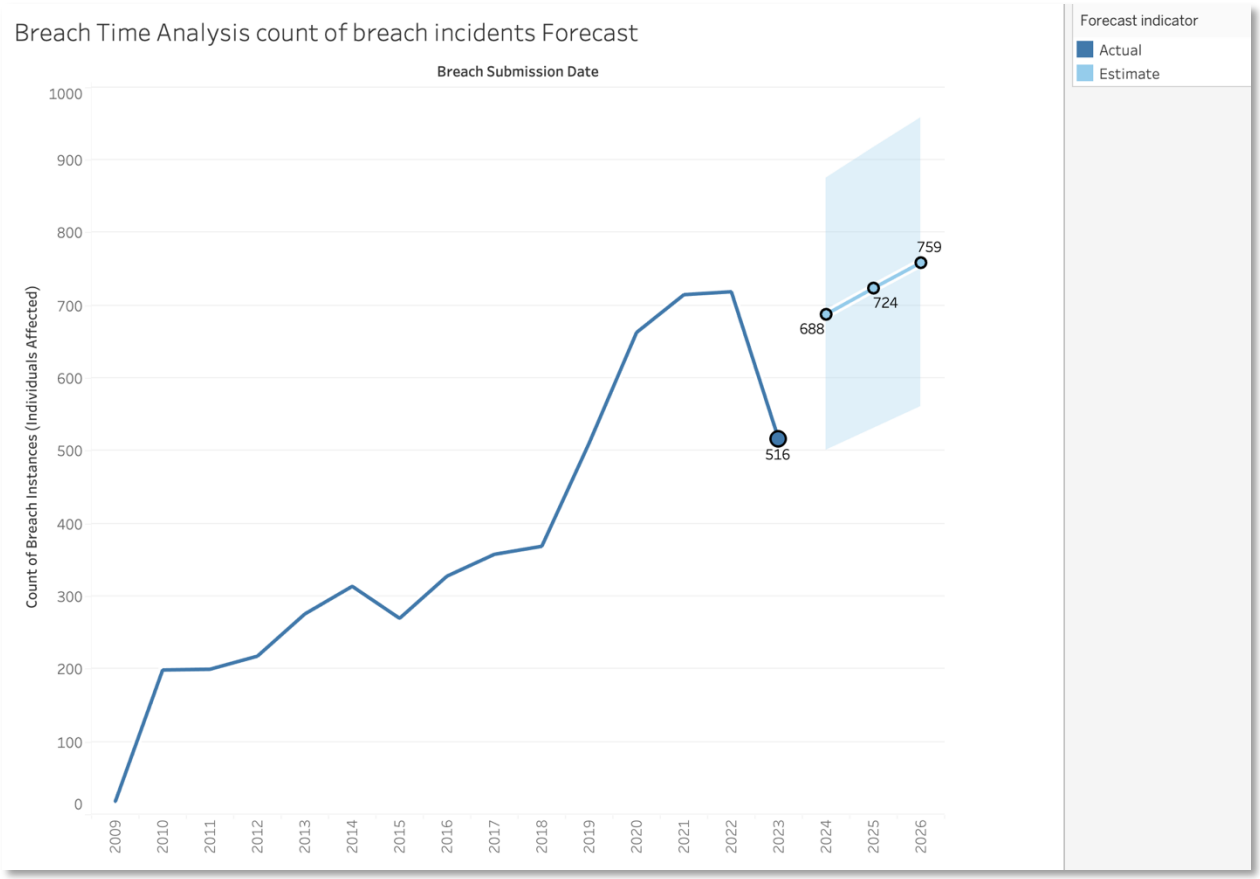


Figure 7b is a line graph showing the trend of the count of breach instances (actual and forecast) for the breach submission year. The line's color shows details about the forecast indicator. The marks are labeled by the count of breach instances (actual [2023] & forecast [2024 – 2026]).

Figure 7c Breach Time Analysis Count of Breach Incidents Trendline

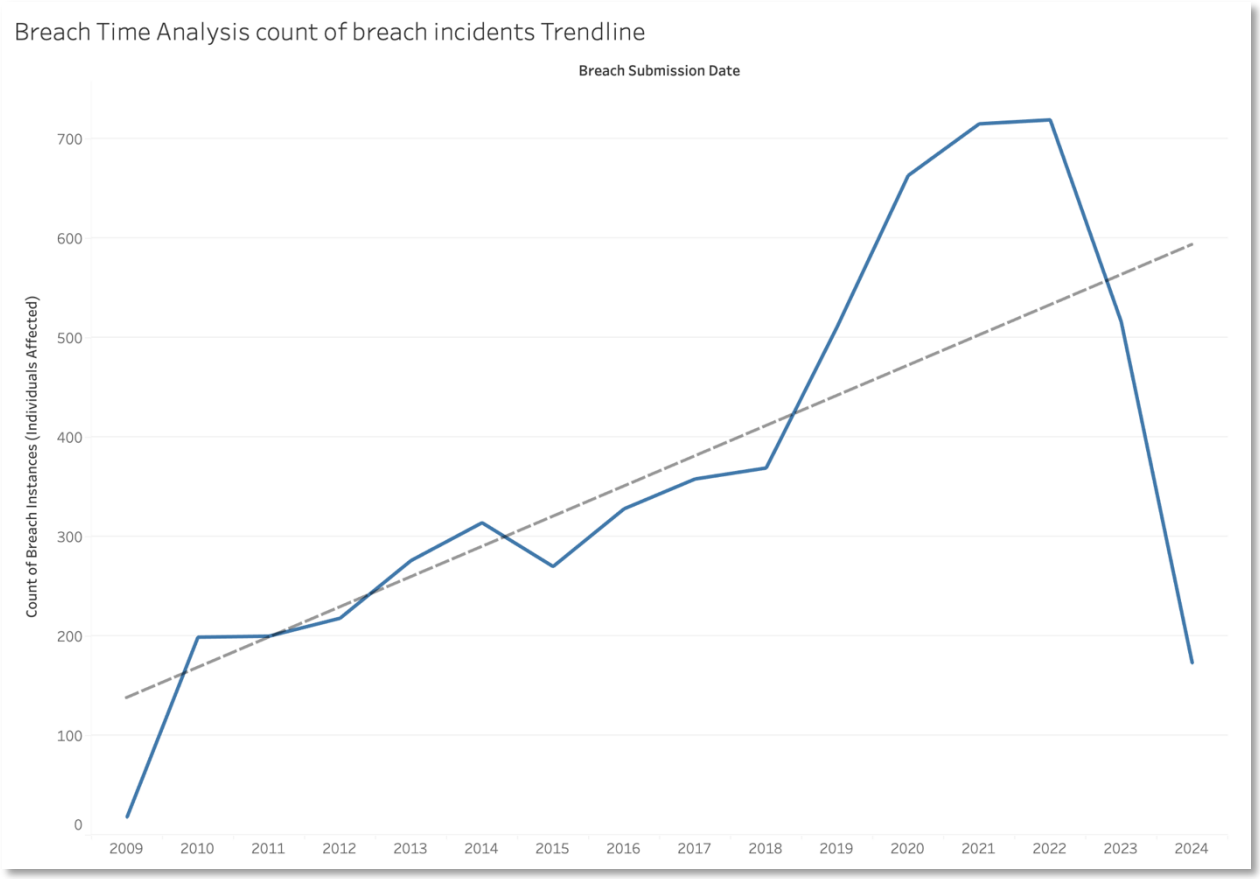


Figure 7c is a line graph of the actual count of breach incidents, similar to Figure 7c, but with a Tableau trendline function applied to show the trend of the count of breach incidents over the years.

Figure 7d: Breach Time Analysis Individual affected Sum (Actual)

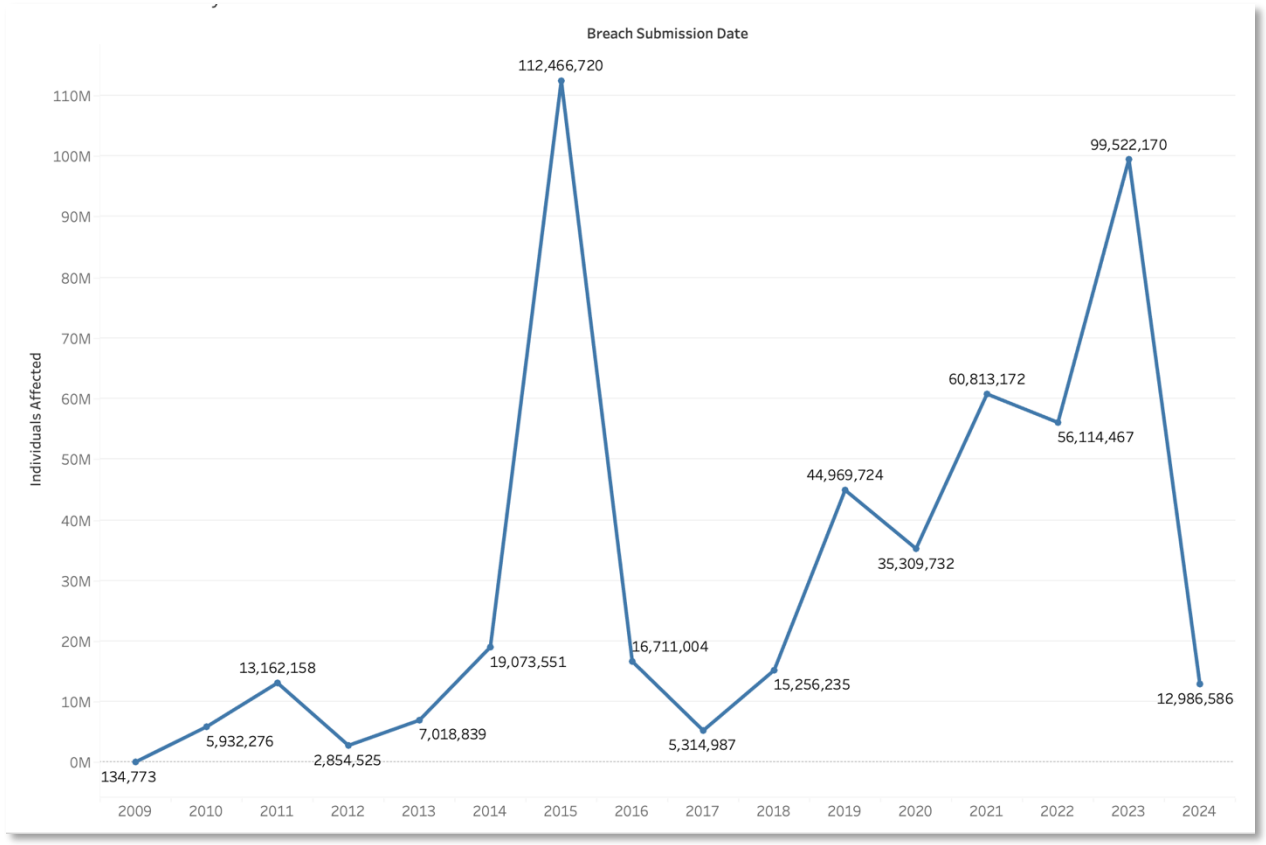


Figure 7d is a line graph showing the yearly sum of individuals affected from 2009 to 2024.

The marks are labeled by the sum of individuals affected.

Figure 7e: Breach Time Analysis Individual Affected Sum Forecast

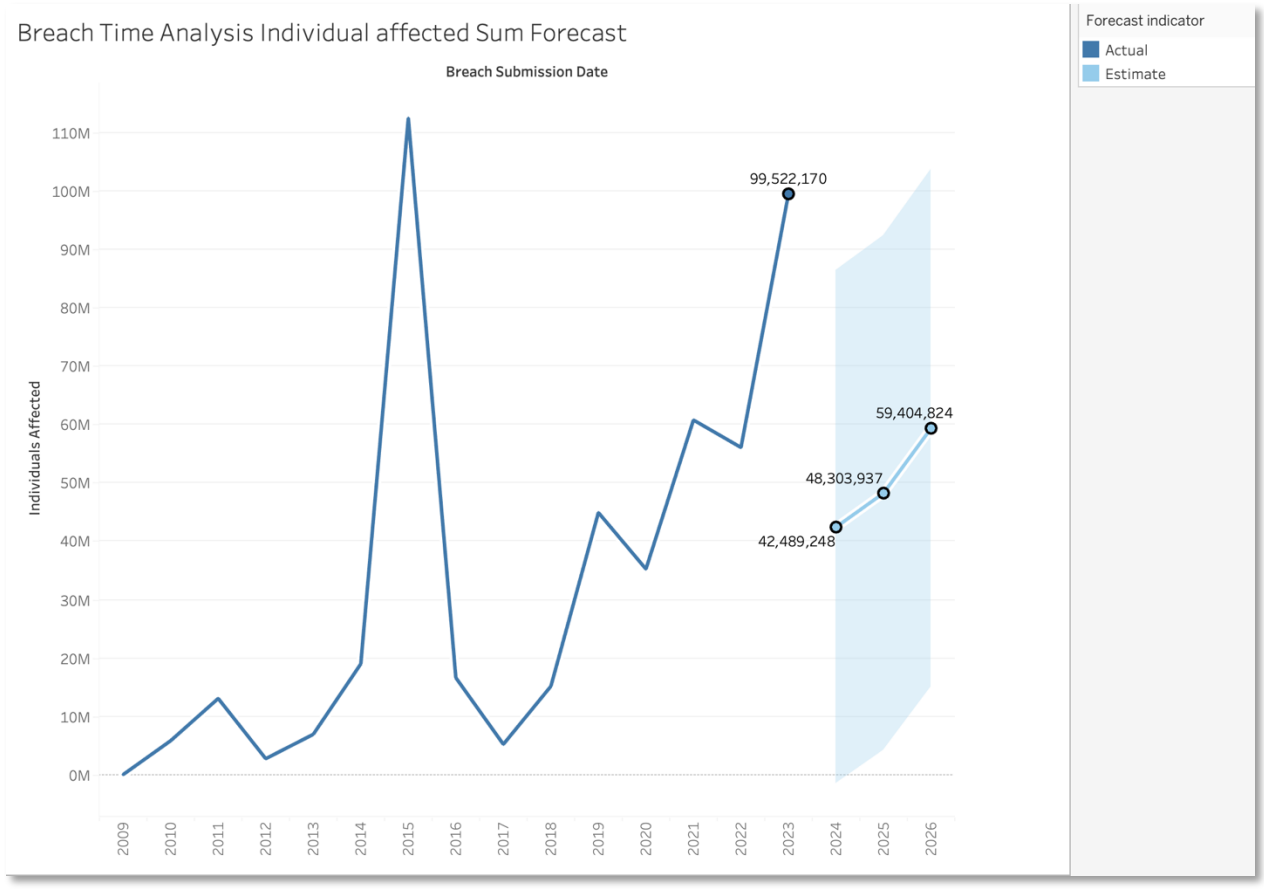


Figure 7e is a line graph showing the trend of the sum of individuals affected (actual and forecast) for the breach submission year. The line's color shows details about the forecast indicator. The marks are labeled by the sum of individuals affected (actual [2023] & forecast [2024 – 2026]).

Figure 7f: Breach Time Analysis Individual Affected Sum Trendline

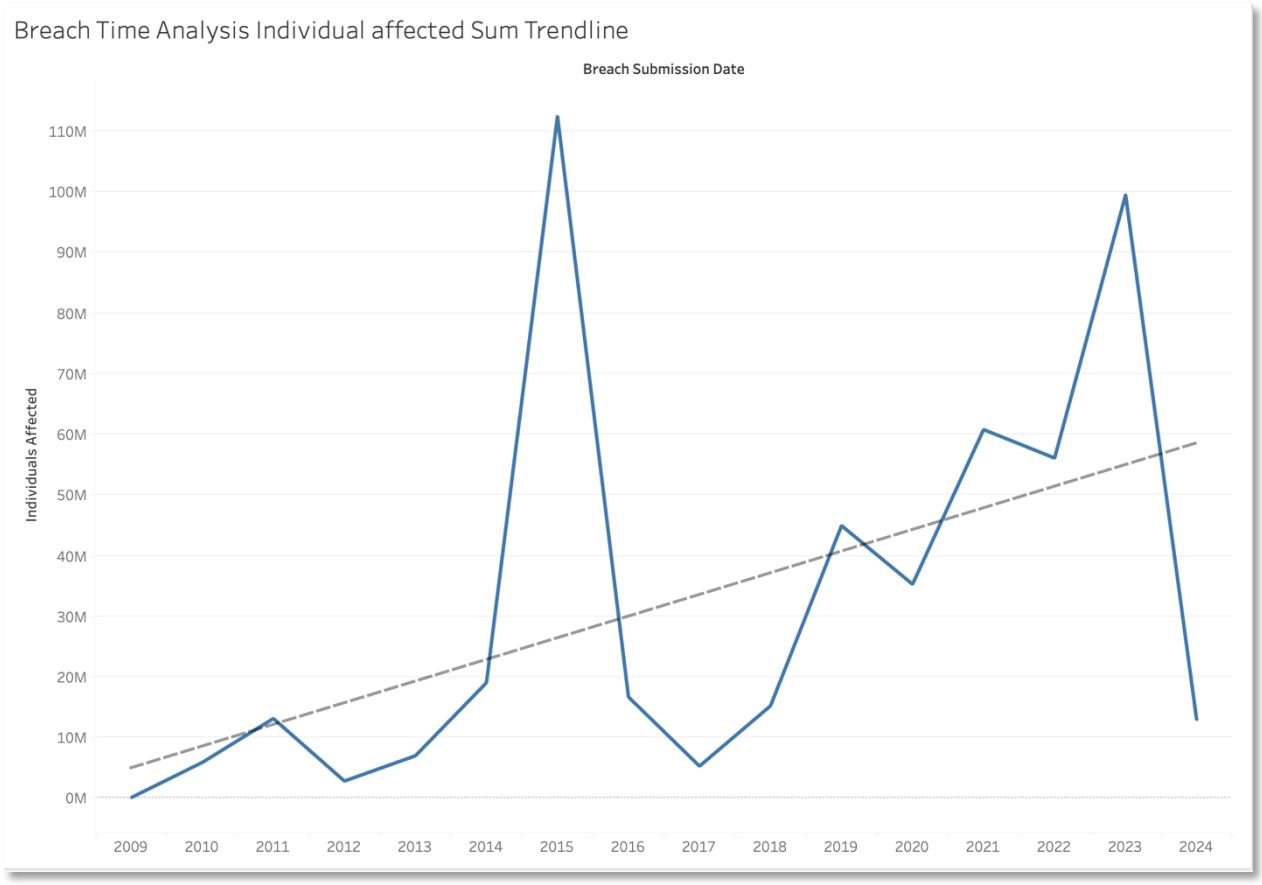


Figure 7f is a line graph of the actual sum of breach incidents, similar to Figure 7d, but with a Tableau trendline function applied to show the trend of the sum of breach incidents over the years.

The remaining figures, tables, and charts derived from the data and reports referenced in the 'discussion and conclusion' chapter are included in the list of figures/tables and provided in the appendix.

CHAPTER 5

Discussion and Conclusion

The visualizations provided to address the research questions in chapter four were done without detailed context and explanation; this chapter will answer those questions by discussing the results and synthesizing them with the existing literature related to the thesis findings. The conclusion for the thesis will include its limitations and recommendations for future research.

Discussion of Results

The findings of this thesis help provide a better understanding of data breaches in the U.S. healthcare sector. The research questions are structured to explore the datasets and understand the nature of these breaches.

The first research question addressed by the thesis's result was: What are the most common attack vectors and types of cyberattacks that lead to data breaches in U.S. healthcare? The visualizations used to address this question were Figure 1a and Table 1a, which portrayed information on the count of breach incidents to see how often these breach types occur, and Figure 1b and Table 1b, which portrayed the sum of individuals affected by these breach types. Using both count and sum helps us understand the breaches better by providing information on the frequency and impact, while addressing misleading conclusions due to outliers. The findings revealed that the most common breach type for both sum and count was hacking/IT incident, followed by unauthorized access/disclosure, theft, and loss. Hacking/IT incidents account for about fifty percent of breach incidents and over eighty-four percent of individuals affected. This finding is consistent with existing

literature about the most common types of cyberattacks that the U.S. healthcare industry suffers from. Raghupathi et al., in their 2023 analysis, reported that hacking/IT incidents are the most significant type of breach and that organizations should pay closer attention to preventing the gateways for these attacks. Unauthorized access/disclosure, theft, and loss cannot be ignored, however, because as highlighted by Verizon's DBIR 2024 report, internal actors/insiders causing data breaches have seen a major increase in the healthcare industry since 2018, and this directly influences the threats of these breach types occurring (See Appendix A) The impact of loss was also highlighted by Raghupathi et al. (2023), noting that data that were compromised as a result of loss were rarely ever recovered. This highlights the need for strong authentication mechanisms to prevent unauthorized access and reduce theft.

To further understand the breach types, the second research question asks how different types of breaches have evolved over time in the healthcare sector. To answer this question, Figure 2 (Breach Type Trend) was used to visualize the trends in different breach types over time. The count function was used to observe how the frequency of the types of breaches changed over the years (2009 – 2023). The findings from Figure 2's temporal analysis indicate a steep rise in hacking/IT incidents over the years, with a peak in 2020 and 2021 and a steady decline after that, but still higher than the other three breach types analyzed. Unauthorized access/disclosure remained consistently high, showing that insider threats and accidental data exposure are significant concerns for healthcare organizations. Theft-related breaches were more prevalent before 2015 but have since declined, likely due to data and information becoming more digitized and cloud-based. Similarly, breaches resulting from loss have remained consistently low throughout the years. However, their

impact should not be overlooked, as shown in 'Appendix B', with loss/theft appearing twice in the top ten as the cause of breach incidents that affected the most individuals. Still, all the incidents occurred before 2015, which aligns with the findings in Figure 2. These findings are consistent with the analysis performed by Raghupathi et al. (2023), which emphasizes the growing dominance of hacking incidents while warning about the persistent risks of insider threats. The trends highlight the need for ongoing investments in network security, employee training, and insider threat monitoring to mitigate evolving cybersecurity risks in healthcare.

What are the most frequently compromised sources in healthcare data breaches? was the third research question addressed in the thesis. Figure 3 (Count of Breached Sources) was used to analyze the question. The findings reveal that Network Servers (33.52%) and Email (22.66%) are the most frequently compromised sources, accounting for over half of all reported breaches. This aligns with existing literature, emphasizing the threat network servers face due to their central role in storing and transmitting sensitive health information. For instance, Seh et al. (2020) highlight that the increasing digitization of healthcare records has made electronic health record (EHR) systems more susceptible to cyberattacks, with network servers being primary targets. Paper/Films (14.94%) remain a significant breach source, indicating that physical records pose privacy risks despite digital transformation efforts in healthcare. Gabriel et al. (2018) note that paper and film are usually sources of breaches due to unauthorized access, improper disposal, and theft. Laptops (7.40%) and desktop computers (6.28%) also contribute to breaches, often due to device theft or loss. Other sources, such as Electronic Medical Records (5.90%) and Portable Electronic Devices (4.30%), highlight risks related to both digital and physical

data security. These categories suggest that healthcare organizations must implement strong encryption, access controls, and employee training programs to mitigate risks. Johnson (2018) emphasizes the need for training and educating people who handle personal health information on their responsibility to ensure the privacy and security of the information they work with.

To address the fourth research question, 'Which states have the highest number of individuals impacted by data breaches?' Figures 4a, 4b, and 4c were utilized to analyze the impacted individuals' geographical distribution (States). The analysis used the sum function to determine the number of individuals affected per state. The findings from Figure 4a's state-level analysis indicate that Indiana experienced the highest number of affected individuals, with approximately 89 million impacted. This is a significant outlier compared to other states, although this is due to a particular breach incident in 2015, which affected over 78 million people (see Appendix B). California, Florida, New York, and Texas followed, with 25 to 31 million affected individuals. The high breach impact in these states may be attributed to their large populations and the presence of major healthcare institutions, which are frequent targets of cyberattacks. Figures 4b and 4c further illustrate the geographical distribution of data breaches using a color gradient to represent impact severity. Indiana stands out as the most affected state, shown in the darkest shade, followed by California and Texas. The map also highlights variations in breach severity, with some states showing lower impact levels, possibly due to stronger cybersecurity measures, lower reporting rates, or fewer large-scale incidents. Interestingly, some states with large populations, such as Illinois and Pennsylvania, do not appear in the top 10, indicating that breach severity is not solely determined by population size. Conversely, with a smaller

population than California or Texas, Indiana recorded the highest number of affected individuals, suggesting that specific large-scale breaches significantly influenced the state's ranking. Dean (2023) further highlights how variations in breach severity across states may be attributed to differing cybersecurity measures, regulatory enforcement, and institutional preparedness. The findings reinforce that while certain states are inherently more vulnerable due to their healthcare infrastructure, the scale and frequency of breaches are also shaped by localized security practices and incident response effectiveness (Dean, 2023).

To address the fifth research question, 'Which states experience the highest number of healthcare data breaches?' Figure 5 was analyzed to determine the frequency of incidents across different states. The count function was used to identify the states with the highest number of reported breaches. The findings from Figure 5 indicate that California experienced the highest number of healthcare data breaches, with 580 incidents recorded. Texas followed with 453 incidents, while New York (366), Florida (322), Pennsylvania (246), and Illinois (242) also reported high breach counts. The remaining states in the top ten, Ohio (188), Georgia (166), Massachusetts (162), and Michigan (161), show relatively lower numbers but still represent significant breach occurrences. California stands out as the most frequently breached state, and this reinforces its vulnerability due to its large population and numerous healthcare institutions. The high number of breaches in Texas and New York suggests similar risk factors, such as extensive healthcare networks and increased cyberattack exposure. These findings align with existing literature, which emphasizes that states with larger populations and well-established healthcare infrastructures tend to experience more breaches (Dean, 2023). Additionally, factors such

as cybersecurity policies, breach reporting regulations, and institutional preparedness likely contribute to the observed variations across states. California records the highest number of breaches and also ranks second among the top states for the total number of affected individuals. However, Indiana, which has a significantly lower number of incidents, records the highest number of affected individuals (89 million, Figure 4a). This suggests that while some states experience frequent breaches, others face fewer but highly severe breaches involving massive data exposures. This realization influenced the next research question, which tries to understand the relationship between the frequency of breaches and the number of individuals affected while visualizing outliers.

Figure 6 answers the question, 'Is there a correlation between the count of breaches in a state and the total number of individuals affected?' and explores the relationship between the number of breach instances in a state and the total number of individuals affected through a scatter plot with a trendline. The analysis reveals a positive correlation between these variables, suggesting that states with higher breach counts generally tend to have a larger number of impacted individuals. However, notable outliers disrupt this expected pattern. Despite having fewer breach instances than states like California or Texas, Indiana recorded the highest number of affected individuals (89 million). This reinforces findings from Figures 4 and 5, indicating that certain breaches in Indiana were exceptionally large-scale events, leading to disproportionate impacts compared to breach frequency. The trendline suggests a general pattern where more breaches correlate with greater individual exposure, but the severity of specific incidents plays a critical role in determining the overall impact. These findings highlight the importance of contextual risk assessments in cybersecurity policies. States with frequent

breaches should focus on preventive measures, such as strengthening access controls, while states experiencing fewer but more severe breaches should prioritize response strategies to mitigate large-scale data leaks.

The following research question, 'How have the number of breach incidents and the total number of individuals affected changed over time, and what trends or future projections can be identified in healthcare data breaches?' was answered with a set of figures (7a – 7f) by analyzing historical and forecasted breach data. Through multiple line graphs, the analysis examines the evolving landscape of breach incidents and their impact on individuals over time. Figure 7a illustrates the count of breach instances from 2009 to 2024. The observed trend indicates a steady increase in breach incidents, reflecting the growing vulnerability of the healthcare sector to cyber threats. This rise may be attributed to the increasing digitization of health records and evolving cyberattack methods targeting healthcare institutions. Figure 7b extends this analysis by incorporating forecasted breach incidents for 2024 to 2026. The forecast suggests that breach incidents are likely to continue rising, emphasizing the need for proactive security measures. The color-coded forecast indicator helps distinguish between actual and projected data, providing a clearer understanding of future risks. Figure 7c applies a Tableau trendline to historical breach counts, reinforcing the observed upward trajectory. This statistical trendline highlights a persistent growth pattern, underscoring the urgency for improved cybersecurity strategies to curb the increasing frequency of breaches. Figure 7d shifts focus to the sum of individuals affected by breaches from 2009 to 2024. While there is an overall upward trend, certain years exhibit sharp spikes (2015, 2023), suggesting that large-scale breaches significantly contribute to annual variations. These fluctuations indicate that while breach

incidents are increasing, the impact of individual breaches varies widely. Figure 7e provides a forecast for the total number of individuals affected, projecting data from 2024 to 2026. The forecasted trend suggests a continued increase in affected individuals, aligning with the growing frequency of breaches. This highlights the compounding risk to patient data security if current cybersecurity measures remain unchanged. Figure 7f enhances the analysis by applying a Tableau trendline to the sum of affected individuals. The trendline confirms a steady rise in breach impacts over time, reinforcing the need for stronger regulatory policies and cybersecurity investments. While the general trend shows an increase, the severity of breaches in specific years indicates that certain high-impact events disproportionately drive the overall numbers. The collective findings from Figure 7 suggest that both breach frequency and impact are on an upward trajectory, necessitating intervention. The forecasted increase in breach incidents and affected individuals calls for enhanced preventive measures, including advanced threat detection systems, staff training, and stricter compliance regulations like HIPAA and HITECH (Reddy et al., 2023). Furthermore, response strategies must evolve to address large-scale breaches more effectively, mitigating their potential damage. These insights emphasize the importance of long-term cybersecurity planning to safeguard sensitive healthcare data against future threats.

Contributions of the Study

This study contributes to the field of cybersecurity and healthcare by providing an in-depth analysis of data breaches in the U.S. healthcare sector. It extends the existing body of knowledge by offering a data-driven understanding of breach trends, attack vectors, and

impacted sources. This study highlights the growing threat of hacking/IT incidents, the persistent risks posed by insider threats, and the disproportionate impact of large-scale breaches in certain states through a visual analysis. Furthermore, the study's forecasting visualizations offer valuable insights into future trends, helping organizations anticipate risks and enhance security measures proactively. The findings in the study emphasize the need for healthcare institutions to adopt a multi-layered and continuous cybersecurity approach, reinforcing the importance of network security, employee training, and regulatory compliance to mitigate breach risks effectively.

Study Limitations

The limitations of this study are due to the data analyzed, as they were primarily sourced from the U.S. Department of Health and Human Services Office for Civil Rights (OCR) and publicly available breach reports, which may not capture all incidents, especially those not reported or underreported. As the OCR only requires data breaches that affect over 500 people to be reported, this might skew the count of breach incidents if the number of people affected is 499 or lower.

Reflection and Growth

Conducting this research has significantly contributed to my growth as both a researcher and a cybersecurity professional. Through my work on healthcare data breaches, I have developed a deeper understanding of the complexities of cybersecurity in the healthcare sector, including the vulnerabilities that expose sensitive patient data. This research has also enhanced my ability to analyze and interpret large datasets, strengthening my skills in

business intelligence tools and data visualization. Furthermore, my engagement with academic resources such as Google Scholar and the Cal State LA Library portal has improved my ability to conduct thorough literature reviews, critically evaluate sources, and integrate diverse perspectives into my work. Enrolling in CIS 4150 (Foundations of Business Intelligence) with Dr. Shilpa Balan further refined my ability to apply business intelligence techniques to real-world cybersecurity problems. Beyond technical skills, this research has sharpened my critical thinking, problem-solving, and time management abilities. Working closely with my advisor to structure my research plan has reinforced the importance of discipline and consistency in academic work. Additionally, the honors college senior thesis forum (HNRS 4970) has provided me with a strong foundation for contributing to an academic field, and engaging in the peer review activity provided by the class has also given me valuable insight into how other brilliant minds articulate their ideas, enhancing my ability to communicate my research effectively. As I conclude this phase of my academic journey, I recognize the value of storytelling through data by transforming raw statistics into meaningful insights that can drive cybersecurity improvements in healthcare. This experience has prepared me to approach future challenges with a data-driven mindset, ensuring that I continue to grow as a cybersecurity analyst.

Conclusion and Future Research

In today's digital world, where our most sensitive health information is constantly at risk and digital threats continue to grow, this study underscores the critical need for stronger cybersecurity in healthcare. Protecting patient data is not just a regulatory requirement but a fundamental responsibility that impacts us all.

Several areas warrant further investigation. Future research should explore the role of artificial intelligence and machine learning in predicting and preventing healthcare data breaches. Future studies should also make use of qualitative research methods, like interviews with cybersecurity professionals and healthcare administrators, to gain first-hand and deeper insights into institutional challenges and to provide tailored mitigations for healthcare data breaches.

REFERENCES

Ammenwerth, E., & Hoerbst, A. (2010). Electronic Health Records. *Methods of Information in Medicine*, *49*(04), 320–336.

Choi, Sung, and M Johnson. *Do Hospital Data Breaches Reduce Patient Care Quality?* 2017.

Dean, N. (2023). HEALTHCARE DATA BREACHES: ANALYSIS AND PREVENTION. *Electronic Theses, Projects, and Dissertations*. https://scholarworks.lib.csusb.edu/etd/1704/

Gabriel, M. H., Noblin, A., Rutherford, A., Walden, A., & Cortelyou-Ward, K. (2018). Data breach locations, types, and associated characteristics among US hospitals. *The American journal of managed care*, *24*(2), 78–84.robert

Hansen, S., & Baroody, A. J. (2020). Electronic Health Records and the Logics of Care: Complementarity and Conflict in the U.S. Healthcare System. *Information Systems Research*, *31*(1), 57–75. https://doi.org/10.1287/isre.2019.0875

Hayrinen K, Saranto K, Nykanen P (2008) Definition, structure, con- tent, use and impacts of electronic health records: A review of the research literature. Internat. J. Medical Informatics 77(5):291–304.

Honavar, S. G. (2020). Electronic medical records – the good, the bad and the ugly. *Indian Journal of Ophthalmology*, *68*(3), 417–418. https://doi.org/10.4103/ijo.ijo_278_20

Johnson, S. (2019). Safeguarding Against Data Breaches. *Applied Research Projects*. https://doi.org/10.21007/chp.hiim.0061

Kohli, R., & Tan, S. S.-L. (2016). *Electronic Health Records: How Can IS Researchers Contribute to Transforming Healthcare?* MIS Quarterly, 40(3), 553–574.

Koyame-Marsh, Rita, Marsh, John (2014). Data Breaches and Identity Theft: Costs and Responses Rita O. Koyame-Marsh and John L. Marsh. https://doi.org/10.6084/m9.figshare.1284635

Kwon, J., Johnson, M., Johnson, Dean, R., & Henderson, B. (n.d.). *THE MARKET EFFECT OF HEALTHCARE SECURITY: DO PATIENTS CARE ABOUT DATA BREACHES?*https://econinfosec.org/archive/weis2015/papers/WEIS_2015_kwon.pdf

Moffit, R., & Steffen, B. (2017). *Health Care Data Breaches: A Changing Landscape*. https://mhcc.maryland.gov/mhcc/pages/hit/hit/documents/HIT_DataBreachesBrief_Brf_Rpt_090717.pdf

Privacy Rights Clearinghouse. "Data Breaches | Privacy Rights Clearinghouse." *Privacyrights.org*, 2020, privacyrights.org/data-breaches.

Raghupathi, W., Raghupathi, V., & Saharia, A. (2023). Analyzing Health Data Breaches: A Visual Analytics Approach. *AppliedMath*, *3*(1), 175-199. https://doi.org/10.3390/appliedmath3010011

Reddy, J., Elsayed, N., ElSayed, Z., & Ozer, M. (2023). A Review on Data Breaches in Healthcare Security Systems. *International Journal of Computer Applications*, *184*(45), 1–7.

Rice, T., Rosenau, P., Unruh, L. Y., Barnes, A. J., Saltman, R. B., Ewout Van Ginneken, European Observatory On Health Systems And Policies, & World Health Organization. (2013). *United States of America : health system review*.

Seh, A. H., Zarour, M., Alenezi, M., Sarkar, A. K., Agrawal, A., Kumar, R., & Khan, R. A. (2020). Healthcare Data Breaches: Insights and Implications. *Healthcare (Basel, Switzerland)*, *8*(2), 133. https://doi.org/10.3390/healthcare8020133

Singh, A. K., Anand, A., Lv, Z., Ko, H., & Mohan, A. (2021). A survey on healthcare data: a security perspective. *ACM Transactions on Multimedia Computing Communications and Applications*, *17*(2s), 1-26.

Stachel, R., Morris, R.F., & Delahaye, M. (2015). SECURITY BREACHES IN HEALTHCARE DATA: AN APPLICATION OF THE ACTOR-NETWORK THEORY.

U.S. Department of Health and Human Services. "HITECH Act Enforcement Interim Final Rule." *HHS.gov*, 16 June 2017, www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html.

U.S. Department of Health & Human Services. (2024). *U.S. Department of Health & Human Services - Office for Civil Rights*. Hhs.gov. https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf

Verizon. (2024). *2024 Data Breach Investigations Report*. Verizon Business. https://www.verizon.com/business/resources/reports/dbir/

Yeng, K., Obiora Nweke, L., Woldaregay, A., Yang, B., & Snekkenes, E. (n.d.). *Data-Driven and Artificial Intelligence (AI) Approach for Modelling and Analyzing Healthcare Security Practice: A Systematic Review*. https://ntnuopen.ntnu.no/ntnu-

APPENDIX

Appendix A

Top patterns in healthcare industry breaches (adapted from Verizon, 2023)

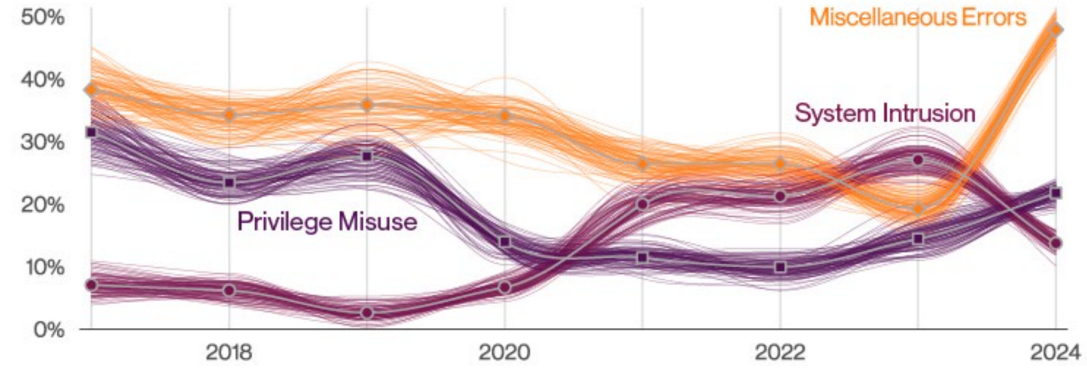


Figure 62. Top patterns in Healthcare industry breaches

Appendix B

Top 10 Average Breach Incident Table with breach type and location

Top 10 Avg breach Incident Table with breach type and location

Name of Covered Entity	Type of Breach (group)	Location of Breached Info..	State	Year of Bre..	
Anthem Inc.	Hacking/IT Incident	Network Server	Indiana	2015	78,800,000
Optum360, LLC	Hacking/IT Incident	Network Server	Minnesota	2019	11,500,000
Laboratory Corporation of America Holdings dba Lab..	Hacking/IT Incident	Network Server	North Carolina	2019	10,251,784
Excellus Health Plan, Inc.	Hacking/IT Incident	Network Server	New York	2015	9,358,891
Perry Johnson & Associates, Inc. dba PJ&A	Hacking/IT Incident	Network Server	Nevada	2023	9,302,588
Managed Care of North America	Hacking/IT Incident	Network Server	Georgia	2023	8,627,242
Community Health Systems Professional Ser..	Hacking/IT Incident	Network Server	Tennessee	2014	6,121,158
Premera Blue Cross	Hacking/IT Incident	Network Server	Washington	2015	11,000,000
Science Applications International Corporation..	Loss	Other	Virginia	2011	4,900,000
Community Health Systems Professional Ser..	Theft	Network Server	Tennessee	2014	4,500,000
20/20 Eye Care Network, I..	Hacking/IT Incident	Network Server	Florida	2021	4,142,440
OneTouchPoint, Inc.	Hacking/IT Incident	Network Server	Wisconsin	2022	4,112,892
Advocate Health and Hospitals Corporation, d/..	Theft	Desktop Computer	Illinois	2013	4,029,530
Medical Informatics Engineering	Hacking/IT Incident	Electronic Medical Record	Indiana	2015	3,500,000
Newkirk Products, Inc.	Hacking/IT Incident	Network Server	New York	2016	3,466,120
Cerebral, Inc	Unauthorized Access/Disc..	Network Server	Delaware	2023	3,179,835
Dominion Dental Services..	Hacking/IT Incident	Network Server	Virginia	2019	2,964,778
Lincare Holdings Inc.	Hacking/IT Incident	Network Server	Florida	2021	2,918,444
AccuDoc Solutions, Inc.	Hacking/IT Incident	Network Server	North Carolina	2018	2,652,537
NEC Networks, LLC d/b/a CaptureRx	Hacking/IT Incident	Network Server	Texas	2021	2,600,000