Data Visualization
Project - Data Breaches in Top Tech Companies

Suraj Khamkar	**skhamka@clemson.edu**	**C16741431**
Sravani Pati	**spati@clemson.edu**	**C79908900**

Project Link - https://github.com/surajkhamkar02/surajkhamkar02.github.io

Website link - https://surajkhamkar02.github.io/

Background:

Data breaches can have a variety of effects. Breached organizations may incur significant costs to remedy the damage, such as correcting system flaws, paying out compensation to victims, and settling legal claims. Rebuilding the harmed people and the violated organization's trust comes at a hidden cost that is difficult to quantify. Individuals affected are at risk of identity theft, account breach, and phishing due to exposed data. Even while it could take years before leaked data is utilized improperly, the damage can be severe. For instance, identity theft victims may have damaged credit records or be forced to file for bankruptcy due to credit misuse. In a 2017 Identity Theft Resource Center poll, 77% of respondents experienced higher stress levels, and 55% expressed higher exhaustion or less energy. Identity theft is also traumatic. [3]

Consequently, some academics have suggested that data breaches result in compensable losses because of the victims' mental pain and the significant danger of future financial injury. Laws usually require breached businesses to compensate victims with discounts or free credit/identity monitoring in addition to notifying those impacted. Services that analyze reports of third-party breaches and advise subscribers include HIBP and Firefox Monitor. Some businesses reset passwords automatically for users whose login information was revealed in password leaks. Additional safety measures for victims include warnings about phishing and social engineering scams and two-factor authentication (2FA), making it more challenging to utilize stolen credentials. However, no system is perfect: phishing alerts are seldom heeded, and attackers can bypass 2FA without obtaining the backup token.

Motivation:

Global business is evolving as a consequence of the fast advancement of communication, networking, and information technology, and this trend is expected to last for some time to come. For the stakeholders in all organizations, this evolution offers both many benefits and drawbacks. Because they could pose serious problems for all business operations, information security and privacy are topics that information systems management is taking more and more into consideration. [2]

Because multinational corporations rely so much on technology and are prone to technical flaws, data breaches and losses are unavoidable. Data is one of a company's most valuable assets, and the risk of losing data control is something that everyone must deal with. No matter what policies and controls businesses put in place to lessen the risk of data breaches, hacking and phishing threats still remain. One determinant of a company's continuity and viability is information security and privacy. [2] Companies are employing a variety of risk-reduction strategies, including user and employee orientation to the organization's information security policies and protocols, system authentication, data encryption, user access control, and firewalls. Despite these precautions, criminals are getting more skilled and coordinated, increasing the risk.

There are numerous recent examples of businesses that experienced significant data breaches, including Equifax, Anthem, eBay, JPMorgan Chase, Home Depot, Yahoo, and Target. Accounting and information security management both face difficulties when evaluating the financial impacts of data breaches (Schatz and Bashroush, 2001).

Although data breaches are common, little is known about people's awareness, perceptions, and reactions to breaches that affect them. Through an online study in which we exposed victims to up to two data breaches that have revealed their email addresses and other personal information, we offer new insights into this subject. Overall, 73% of users were affected by at least one breach, 5.36 breaches on average. Only 14% of users correctly identified external reasons like breached organizations and hackers as the source of being affected by a breach; most victims blamed their email and security habits. 74% of the displayed breaches were unknown to the victim and organization, who reacted differently upon learning about them. [3][4]

Most users thought they wouldn't be affected by the incident, despite others saying they intended to act. Our findings highlight amount of data lost as per data sensitivity and organization which are vulnerable to data breaches. This will help to

understand and learn more about how we can secure our data and what preventing action we should take.
All of this information about data breaches at firms, as well as other incidents and accounts of how sensitive data has affected organizations, drew our attention to this particular issue for visualization. And we concluded that it is always better to be visually represented because it would immediately grab consumers' attention than to read all preventive measures in words.

Objectives:

We will learn about the data breaches and hacking by understanding this data and the way it is visualized. Weak security measures like passwords, data loss due to weak passwords.
Following are some of the questions we are looking forward to address with our visualization for data breaches.

Questions:

- The number of accounts or data thefts taking data sensitivity or hacking type into account
- The variations in the data lost as per years in a particular sector
- Financial lost caused for companies categorized by sectors
- Most financial impacted country per decade
- Most often used passwords that put data at risk
- Percentage of people that cease using services after data breaches
- The amount of information retrieved following the data restoration procedure
- Funding allocated in cybersecurity after data being compromised

Benefits:

Here are few benefits of visualization for data breached

- Learning the data, companies will take data security seriously and take deliberate steps to strengthen it.
- Additionally, these companies will be better equipped to meet their compliance duties and shield their reputations. These elements all improve the comfort and profitability of doing business.

- We can detail that, if we can demonstrate a proactive commitment to data protection, you might attract devoted, long-term customers and boost your revenue.
- The capital or revenue we can save taking steps considering the data lost
- Estimate the loss by specific data sensitivity

Data Collection:

For raw data collection we first went through the news and articles about data breaches happened in last couple of years in different countries. There after we got some pointers to head towards actual data.

The actual data we have collected was taken from this link, but worked on it for cleaning data and some rectification of data. Here are some other links and reference we have used to collect the data.

- https://docs.google.com/spreadsheets/d/1wPgM8ye1AUTVxlZOFsyiKEPWp6iFt34xpp2XA5iM6P0/edit#gid=25233212
- https://docs.google.com/spreadsheets/d/1i0oIJJMRG-7t1GT-mr4smaTTU7988yXVz8nPlwaJ8Xk/edit#gid=2
- https://www.ibm.com/reports/data-breach
- https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/

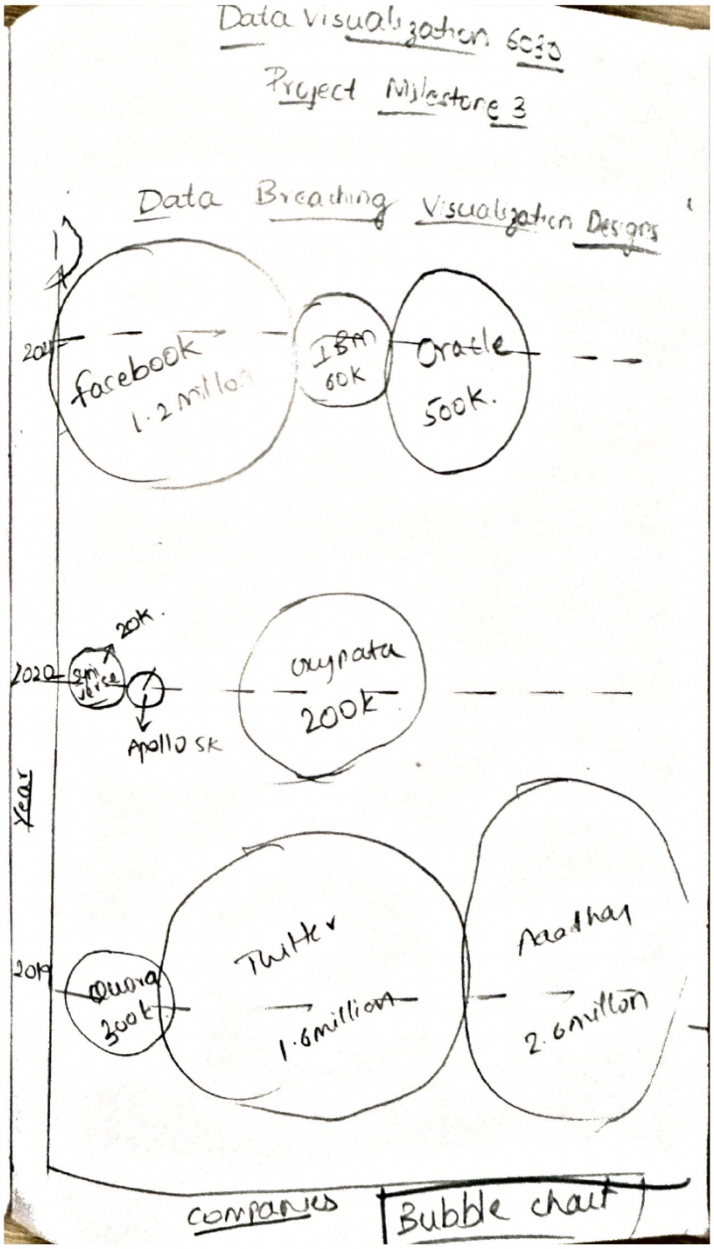
Data Processing:

We did some data cleanup for the links provided above. We extracted number of data lost for each company per year. The quantities we have derived so far are, number of accounts lost, year, revenue, country, data recovered, passwords etc.

Visualization Design:

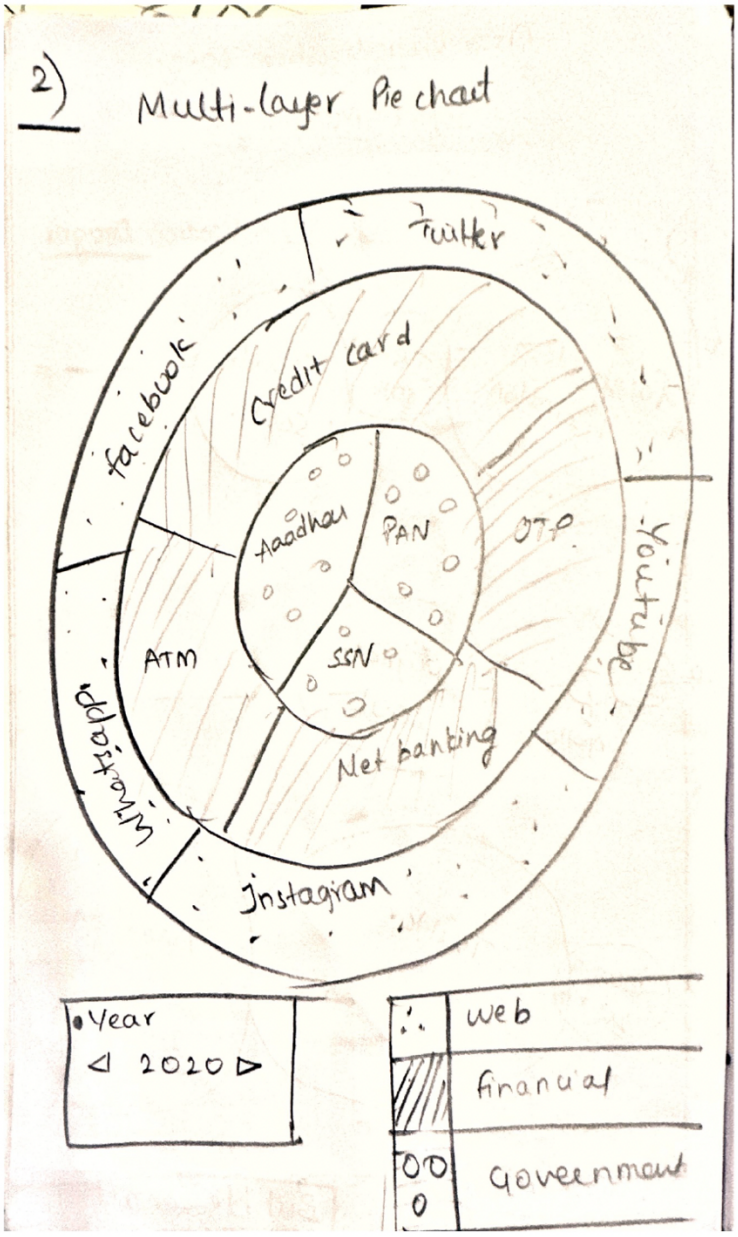
Visualization 1: Bubble Chart

A bubble chart will display relationships and distribution for the number of accounts or data thefts, taking data sensitivity or hacking type into account. However, in this variation, we'll use bubbles in place of the data points. To represent a third kind of data, we will also alter the size of the bubble. A category axis is not used in a bubble chart. Instead, it displays the data sets as X-, Y-, and now Z-values (bubble size).



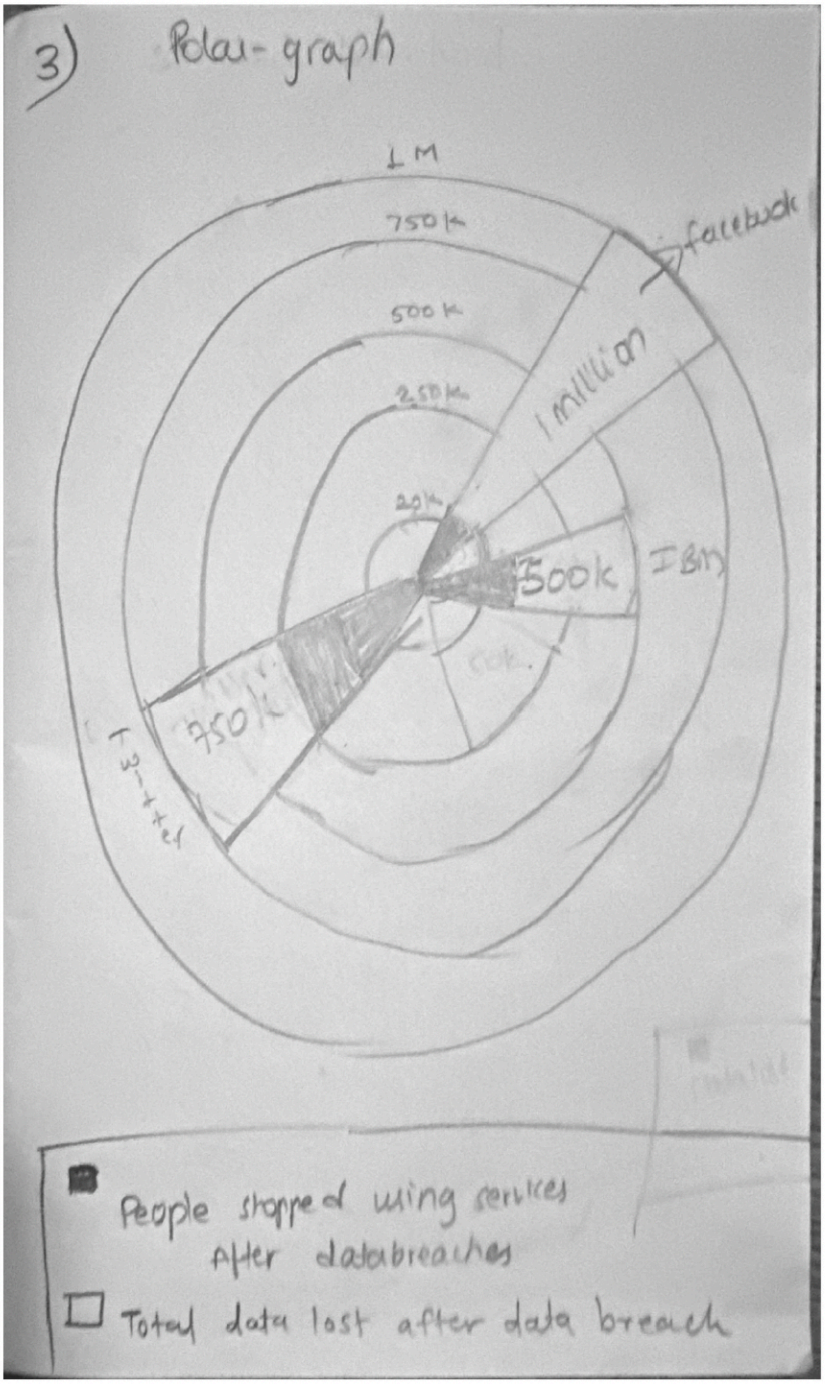
Visualization 2: Multi-layer Pie Chart

The infographic below shows a multi-layered pie chart that shows variances in data loss in several industries. We can see the web activity sector on the top layer, and the financial sectors that fall under each primary category are on the bottom layer. A thin layer in the middle divides all government sectors into three categories. It takes some planning to have all the categories fit together and be simple to grasp in this sort of data visualization, which makes it more difficult to produce than other types. Technically speaking, this representation consists of three pie charts stacked on top of one another.



Visualization 3: Polar Chart

Polar graphs have a circular foundation, but the data is plotted differently. Wedge shapes extend from the center rather than joining points together. The main visual distinction is present. Because the data values are so dissimilar, we choose a polar graph. Otherwise, it could be difficult to read quickly. It is ideal for measuring the number of users who ceased using services after data breaches.



Must have features:

In our project, there are some features we are planning to add listed below,

- Balanced design
  This relates to how texture, color, shape, and negative space are all equally distributed across the visualization. We will choose between symmetrical, asymmetrical, or radial visuals and decide which balance of elements is best for visualizing data.
- Keeping it simple
  Making sure that our visualization is detailed and uncomplicated. Adding unnecessary information may weaken the goal of data visualization, so simplicity is the ultimate goal of data visualization, and we will try to achieve it.
- Consolidated information
  In this feature, we will try to visualize data to convey several ideas and information in a simple and easily understandable way.
- Variety
  This will help us to make our visualization more interesting. As we have a variety of information related to each company, we can represent it in such a way that end visualization will grab the user's or reader's attention.
- Compare aspects
  Here we will be giving information by comparing different companies and years with respective data thefts and the effect of that data theft on their revenues and active users.

Optical features:

There are some features we could think about adding in visualization for better results,

- Decision making ability
  Here the data visualization we will represent is all about data breaches, the amount of data loss, and its effect on particular organizations or users. By learning this visualization, we will be able to this about prior consequences that happened and what would be steps we should take in the future to prevent such kinds of attacks.
- Representation of story or news
  As we have news or story for all the data breaches that happened, and it is

related to one of the data available in our dataset, when will try to represent that particular incident, we can also add news related to that.

Project Schedule:

Sr. No.	Task	Estimated date to be completed
1.	Project Proposal	October 27th
2.	Clean and manage dataset	November 3rd
3.	Solving first three questions	November 8th
4.	Finalizing design and remaining questions	November 12th
5.	Designing prototypes	November 17th
6.	Peer Evaluation	November 22nd
7.	Presentation	December 8th
8.	Final Delivery	December 15th

References:

[1] https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

[2] Juma'ah, A. and Alnsour, Y. "The Effect of Data Breaches on Company Performance" International Journal of Accounting and Information Management (IJAIM). Vol. 28, no. 2, 2020

[3] https://www.ftc.gov/system/files/documents/public_events/1582978/now_im_a_bit_angry_-_individuals_awareness_perception_and_responses_to_data.pdf

[4] https://www.trendmicro.com/vinfo/es/security/news/cyber-attacks/understanding-targeted-attacks-goals-and-motives

[5] https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/

[6] https://www.ibm.com/reports/data-breach

[7] https://informationisbeautiful.net/visualizations/top-500-passwords-visualized/

[8] https://informationisbeautiful.net/visualizations/ransomware-attacks/

[9] https://www.hipaajournal.com/healthcare-data-breach-statistics/