Decoded: Data Breaches & Consequences

INFO 247 Final Project Report

Zach Van Hyfte (MS EECS ‘22), Chris Ferenci (MIMS ‘23), Michael Yang (MIMS ‘23)
INFO 247, Fall 2022
May 11, 2022

Introduction

Background

Over the past two decades, the growth of information technology has been followed by an increase in data breaches. Given the sheer size and variety of organizations targeted by, methods used in, and data stolen from these breaches, it is difficult to grasp the magnitude and severity of this concerning trend. In a 2017 Pew Research survey, two-thirds of Americans reported having experienced some form of data theft, yet the vast majorityof them have not adopted strong security practices in their digital lives. The public understanding for data breaches has not kept pace with the increasing severity of data breaches.

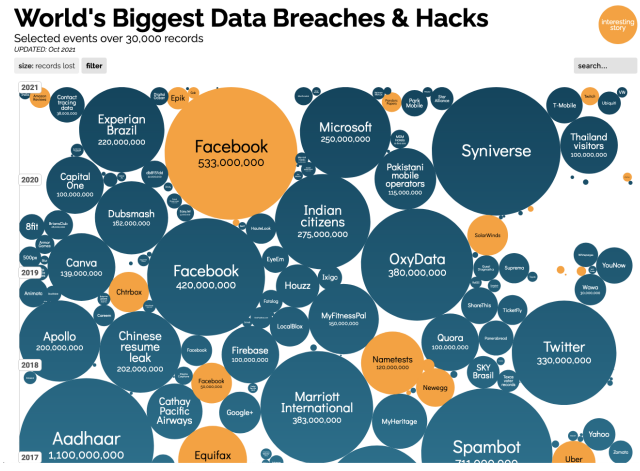
Objectives

“Hacked: The Story of a Typical Data Breach and Its Consequences” is a narrative infographic that seeks to help its viewers understand what happens after data breaches occur and how they could affect them. The ultimate goal of the project is to urge the viewers to take preventative measures, such as using a password manager and two-factor authentication.

Related Works

Visualization:
***World’s Biggest Data Breaches & Hacks* by David McCandless and Tom Evans** (Link)

This project was inspired by this visualization of data breaches bt David McCandless and Tom Evans. It highlights the increasing frequency and scale of the breaches, allows users to filter the set of breaches by market sectors and breach methods, and provides links to interesting articles with background information about specific breaches. While it provides a general impression of increasing trends and biggest breaches, it falls short in translating its significance to individual viewers. The sheer scale of the breaches, as well as the lack of information about breached data types and potential consequences, make it difficult to understand the urgency of this issue.



High/Low-Lighting:
***The Functional Art* by Alberto Cairo**

The line chart that opens the first section of our site is heavily inspired by some of the line charts we saw in the first few weeks of class, most notably the line chart created by Alberto Cairo in the introduction to *The Functional Art*. One of the main points we wanted to get across in the first section of our site is that non-retail businesses and healthcare providers have been getting breached dramatically more than they used to, while most other types of organizations have seen a drop in breaches recently. Our original design used six separate line charts to show the precise trends in breaches over time for each type of organization. During our usability study, participants found it difficult and somewhat overwhelming to parse this grid of charts — our main message was getting buried because everything was at the same level of detail. Consolidating the six lines into a single chart makes it easier to see which types of organizations have seen the biggest increases in breaches. But it was adopting Cairo's style that made the chart clear and illuminating rather than cluttered (as it would be with six overlapping traces all at full saturation). We were able to make all of the chaotic data in a "spaghetti line chart" actually usable and meaningful by singling out specific traces with highlighting, while letting the rest fade to the background, still available as contextual information without overshadowing the main point.

Parts of a Whole:
***Data at Work: Best practices for creating effective charts and information graphics in Microsoft Excel* by Jorge Camões**

A few different charts on our site show the change in the *percentage* of breaches per year that were caused by a certain type of issue, or that included a certain type of information. We initially used line charts to show these percentages, since they allow viewers to quickly see an overall upward or downward trend. However, these charts had some subtle problems, as we found out during usability testing. It wasn't that it was unclear what the charts were showing — worse than that, participants seemed to quickly conclude that the charts showed the *raw number* of breaches in a given year with certain properties, rather than the *percentage* of breaches in a given year with those properties. To address this issue, we thought about different ways that we could show the evolution of a quantity over time while still making the context abundantly clear — that the quantity is "part of a whole." We eventually settled on a set of long bar charts, with one bar for each year. It's a little bit trickier to see the trends with these charts, but the context of what exactly you're looking at is always clearly emphasized by the colored background columns behind each bar — this should leave participants with a better understanding of the data overall.

Elucidating Processes through Animation:
***R2D3* by Stephanie Yee and Tony Chu**

We thought a lot about how we could make use of animation and interactivity to engage users and bring some clarity and concreteness to an abstract topic like data breaches. One of the sets of visualizations we returned during the ideation phases was the R2D3 animations. We were

particularly drawn to the fact that, beyond just using animation to illustrate the direction of trends or changes in some data set, the R2D3 visualizations used animation to visualize specific processes from start to finish, showing explicitly how the inputs flowed from step to step and how they were processed at each step. The animated bar chart on our site — which shows through animation the process of individual types of information being combined and packaged into a valuable commodity — was our attempt at replicating what R2D3 did, bringing a somewhat opaque process (in their case, machine learning; in our case, data breaches) to life and making that *process* intuitively, visually graspable and relatable.

Heuristics:
***How to Conduct a Heuristic Evaluation?* by Jakob Nielsen (Link)**

Jakob Nielsen, in his article How To Conduct a Heuristic Evaluation?, discusses a framework of small evaluators to examine the interface for its usability. During our usability studies with 3 participants, we asked them to observe some key usability heuristics. Heuristics of most importance in our design were (1) maintaining *Visibility of System Status* via a sticky-header navigation bar; (2) *Consistency and Standards* thorough consistent button colors and sizes, fonts, and overall color scheme; and (3) *Aesthetic and minimalist design* by using modern fonts and color schemes, an overall flat design, and prioritizing our data visualization content in the design.

Visual Hierarchy:
***Storytelling with Data: A Data Visualization Guide for Business Professionals (Chapter 5)* by Cole Nussbaumer Knaflic**

Knaflic defines the concept visual hierarchy as establishing a simple visual order for an audience to process complex information. Visual hierarchy shows the user where to direct their attention, and so having a thoughtful visual hierarchy is important for effective data story telling. For our project, this included separating distinct sections using a dark section header to indicate when a new section was beginning. An interesting aspect to our design was a debate we had over the sticky navigation menu. Our first iteration used a dark purple color, but as we scrolled through the page to look at various visualizations, we realized the dark purple color was distracting from some of our important visualizations. There was some debate about which colors struck the correct balance between low visual hierarchy, but still visually important. We eventually settled on a lighter color that didn't distract so much from the overall page.

Narrative Storytelling:
***Storytelling with Data: A Data Visualization Guide for Business Professionals (Chapter 7)* by Cole Nussbaumer Knaflic**

For our project, we wanted to tell the story of a breach. For the beginning of our story, we introduced the concept of a breach, and why they are particularly risky. To set the setting, we chose to use a time element in the top left corner to indicate the time it takes for a hacker to hack, steal, collect, package, sell, and potentially profit off selling an individuals data. In the end,

we resolve with including a call-to-action that gives readers the chance to add enhanced data protections to their personal data.

Use of Color:
***Storytelling with Data: A Data Visualization Guide for Business Professionals (Chapter 5) by* Cole Nussbaumer Knaflic**

Knaflic describes the use of color as an effective highlighting technique when used sparingly and generally alongside other highlighting techniques. For our project, we wanted to keep the color scheme simple and non-distracting. Our primary color was a dark and accessible purple. This color is used the most throughout due to its accessible nature and similarity to black. It is used for primary visuals, primary font, and buttons. We used secondary colors like light green sparingly, as they weren't as accessible, and because we found it too noisy if secondary colors were used too much. Throughout the site, the primary data points, the main takeaways — be they lines, bars, numbers, or something else — were colored dark purple, while we used the lighter, less attention-grabbing green for secondary contextual information. We thought of introducing a third, yellow color to our color scheme, to convey visually the difference in the data being shown by between charts, but thought that this more complex palette could begin to imply nonexistent connections between different charts with the same color.

Animations:
***Animation: can it facilitate? by* Tversky and Morrison**

Tversky and Morrison state that one of the key reasons animations fail can be that they are hard to perceive if the motion is too fast. Keeping this in mind, we used animations very carefully and deliberately throughout our site, and when we did choose to use animation, we set slow transition durations. When we designed our animated bar chart, one interesting challenge that we didn't necessarily anticipate was having to decide precisely which objects should be moving or changing on-screen together (e.g. should this text move while this other object is fading out, or should the fade-out happen on its own first?); even for this simple animation, it took lots of experiments grouping different transitions and adjusting the timing of different phases to find a progression that felt natural. One other animation we used — not to visualize data, but to intuitively and quickly convey a suggested action — was a bouncing "scroll" chevron, in the second section of our website. When a viewer clicks "Get Started" in our first section, a scroll effect takes them to the next section. We had concerns that a user may not know what to do here without another affordance. They may assume another button will appear that they click on. For this, we introduced an animated bouncing chevron arrow icon that bounces up and down. A viewer can either click this, and be taken to the next section, or intuitively pick up on the fact that they're being nudged to keep scrolling.

Method

Data

Our website contains information and visualizations derived from the following four sources:

1. **Privacy Rights Clearinghouse – Data Breach Database**

   This database maintained by the privacy rights advocacy and education organization Privacy Rights Clearinghouse is frequently used in papers and studies. The version of the dataset used in creating our website contains information on **9,015 different data breaches that were made public between January 2005 and October 2019.**

2. **Have I Been Pwned – Breaches Dataset**

   Have I Been Pwned is a free web service that collects the records exposed in data breaches, allowing individuals and organizations to quickly run a search and determine which specific pieces of their personal information or data have been exposed in breaches. It maintains a continuously updated catalog of the data breaches whose exposed records can be searched. The current dataset has **591 data breaches since 2011, along with the types of data (name, address, etc.) that were exposed**.

3. **Privacy Affairs – Dark Web Price Index (2021 and 2022)**

   Privacy Affairs is an internet privacy and security research organization that publishes an annual list of the average prices of different types of stolen information that are for sale in dark web marketplaces.

4. **FTC – Consumer Sentinel Network Reports**

   The U.S. Federal Trade Commission (FTC) publishes annual statistics and reports on scams and frauds, including the kinds of identity theft and identity fraud that can be committed using information exposed in a data breach.

Tools

Design	**Figma** was used to create wireframes and high-fidelity mock-ups.
Development	**HTML/CSS** and a few parts of the **Bootstrap** framework were used to create the website. **JavaScript** and **jQuery** were used to add animations and interactivity.
Visualizations	**Tableau** and **D3.js** were used to create the interactive charts..

Visualization

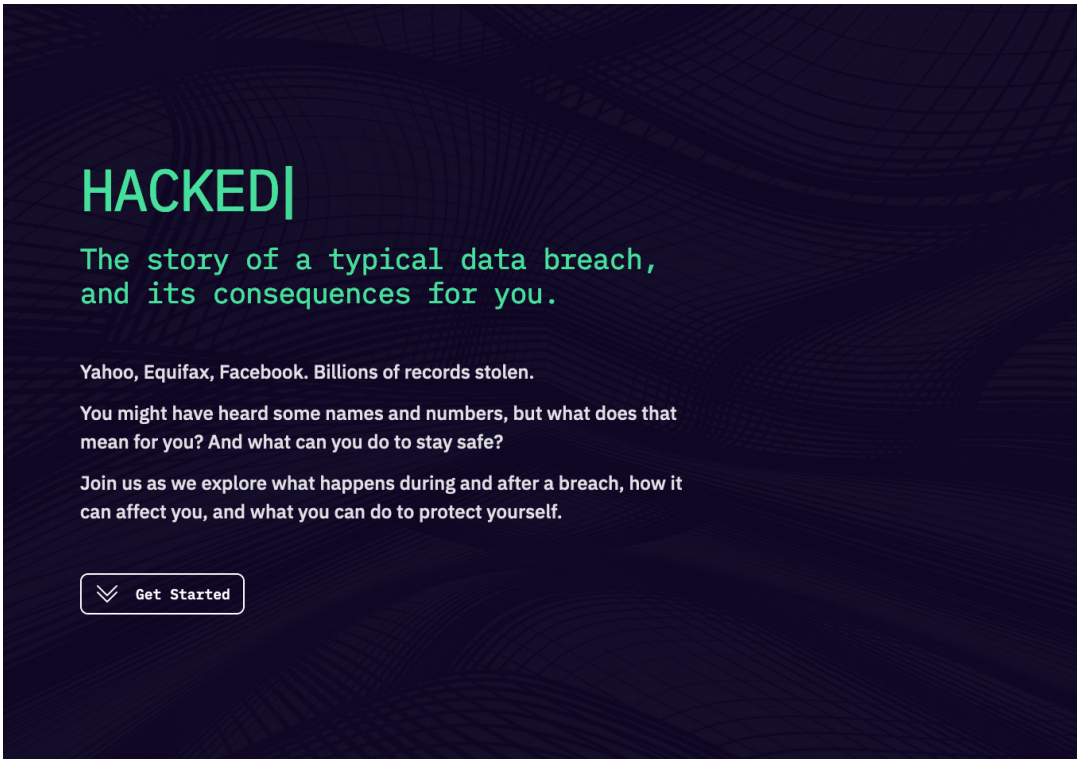
Links

- Project Website
- Project Repository

Sections

This outlines the version 2 of this project. For each section, its purpose, visualization techniques, and future work for version 3 are described, along with screenshots.

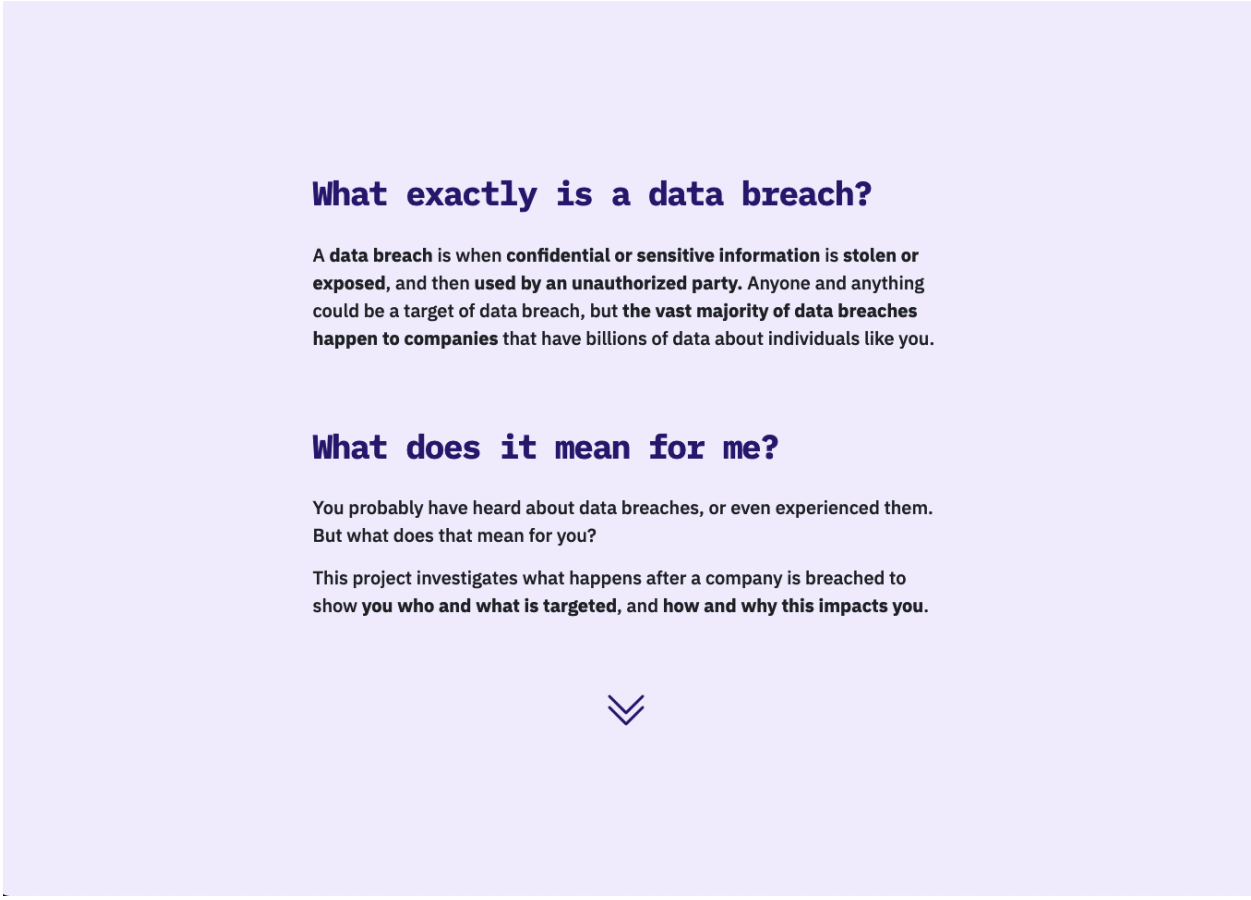
Introduction



This section provides the title and a brief ‘hook’ to get users to continue. It originally had more information about the project, but that part was moved to a separate ‘Definition’ section. We felt that a ‘less-is-more’ approach was better for this introduction.

Future work would be to add more animation to improve the engagement of the hook, such as breach notifications and news headlines appearing to make the breach more real and relevant.

Definition

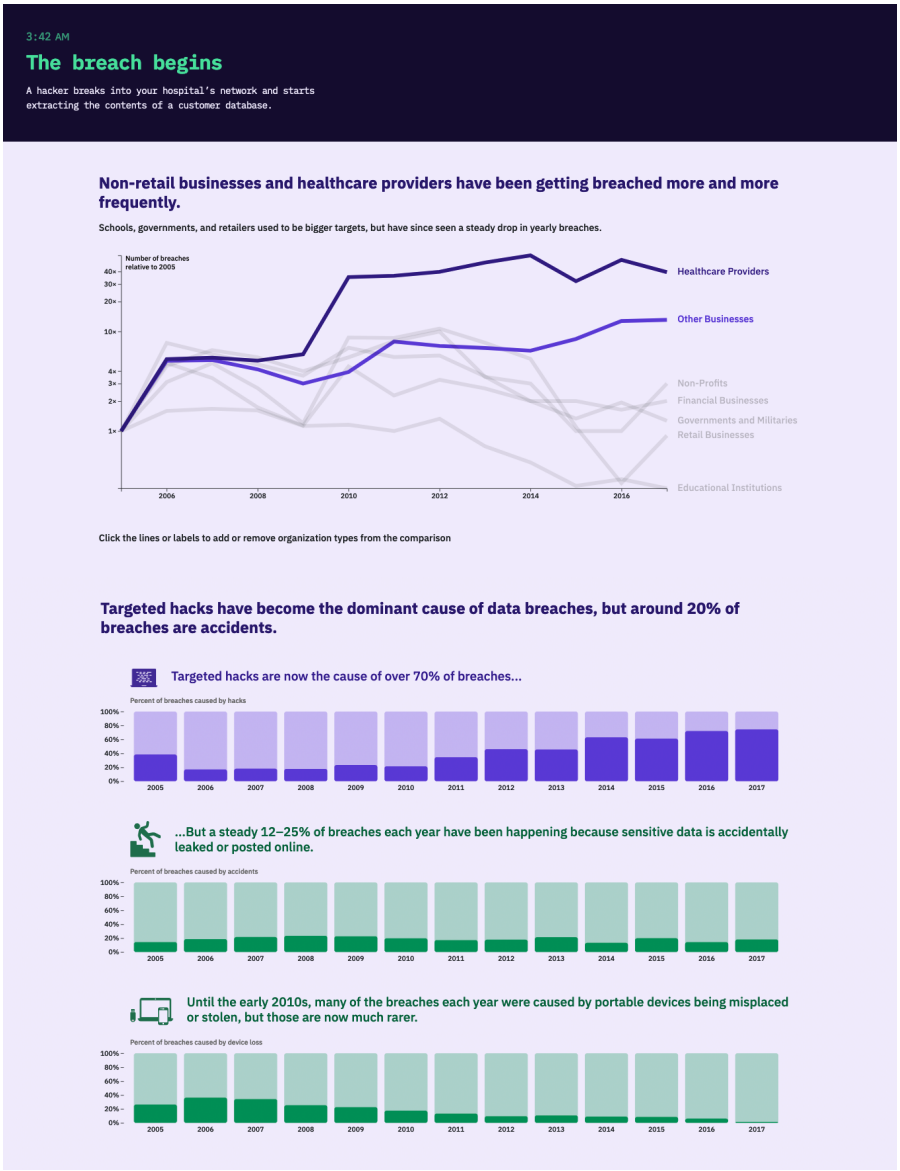


This section provides a definition of what a data breach is, since it is not a very commonly used word. It also helps the audience set expectations for what they're about to see and learn. To encourage viewers to scroll or move to the next section, we added a chevron icon that animates to nudge the user to make an action.

Future work for this section would include adding the survey component from the original design, which could help us gather baseline data to measure the effectiveness of this website.

Also, the horizontal top navigation is not animated to show the current section the user is in. This will need to be updated to aid users in navigation.

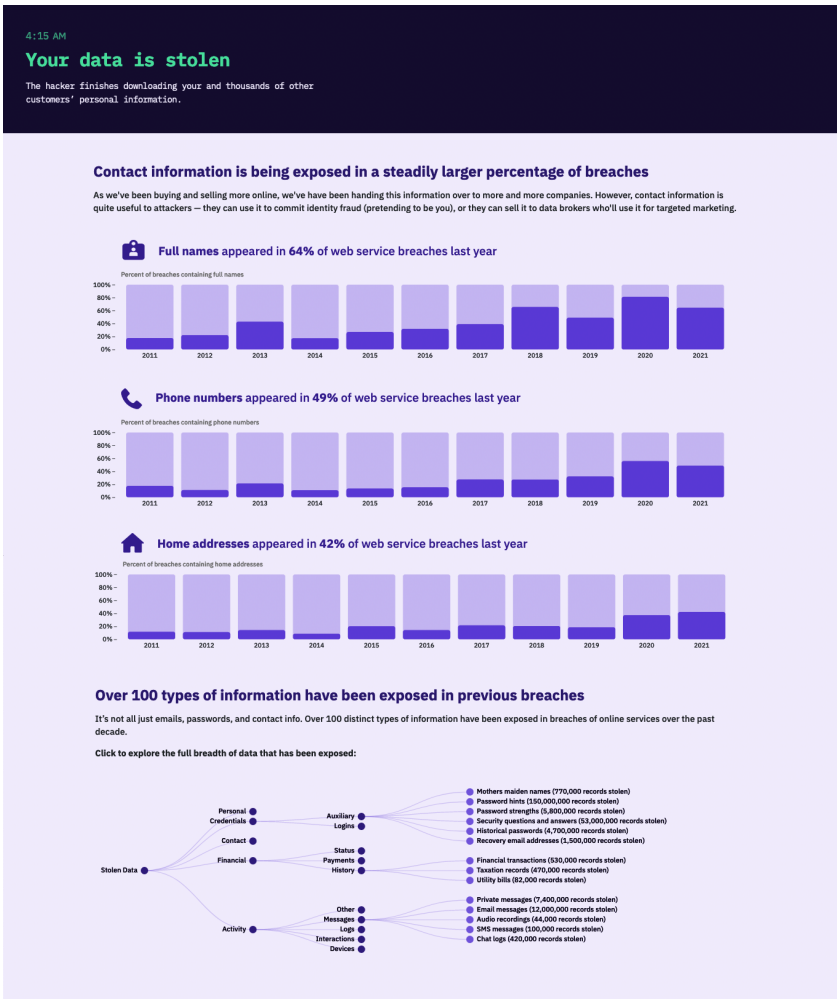
Target Sectors (“The breach begins”)



This section begins the narrative from the moment when a breach occurs. Its purpose is to show the changing trends in the target sectors of data breaches, and the increasing number of data breaches overall.

Highlighting and lowlighting were used in the line charts to help users focus on the key trends while keeping contextual information available to provide a complete picture of the scope of the problem. Side-by-side bar charts with column backgrounds were used to illustrate the changing trends among methods of breach in a way that visually reminds viewers that the quantity they’re looking at is a percentage rather than a raw value.

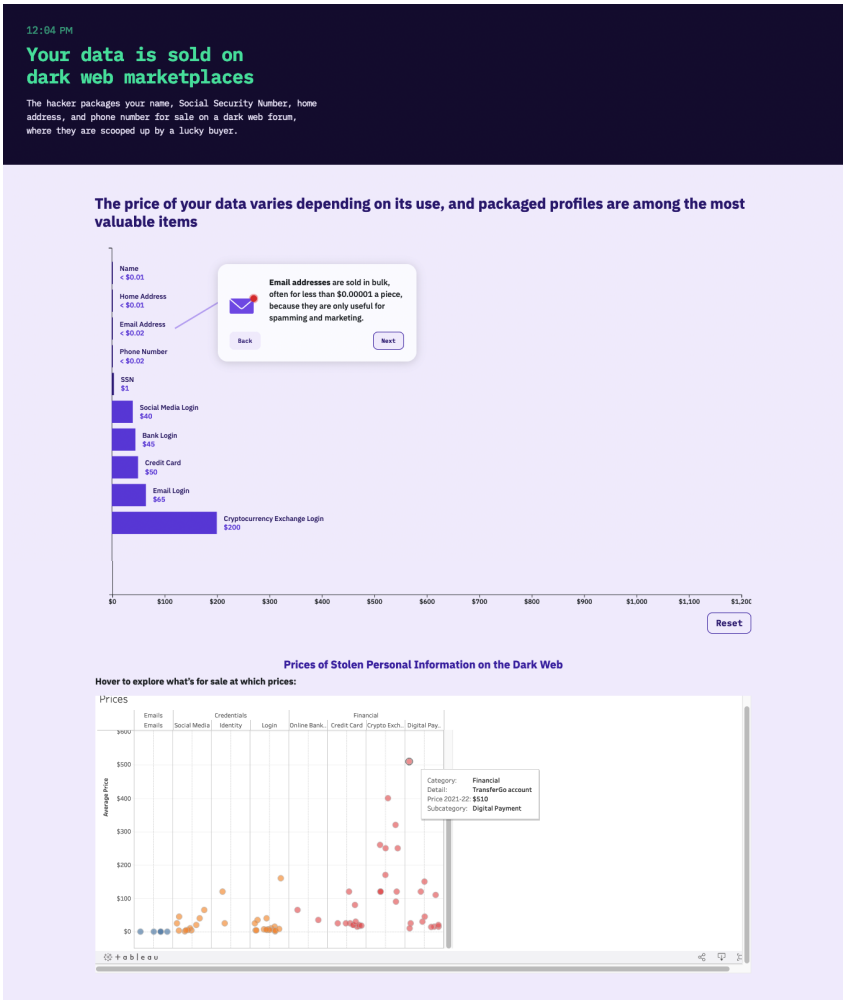
Data Types (“Your data is stolen”)



This section continues the narrative, showing what data was extracted from the breach. Its purpose is to show that personal and contact information have been stolen more frequently in recent years (as a prequel of sorts to the consequent sections, which go on to describe how valuable that information is, and why it’s so valuable). Side-by-side bar charts were used to illustrate these trends, and a collapsible tree was used to let users interact and explore the sheer variety of data types — the intention being that expanding the tree to its full depth (its height expands to several laptop screens’s worth) would visually convey the sheer variety of data that has turned up in breaches.

Future work would include (1) embedding a mini quiz for users to guess how many types of data have been stolen across all known breaches and (2) trying out a collapsible force-directed network graph in place of the collapsible tree, which would let us visually encode the number of breaches or records containing a specific data type as the size of that data type’s node.

Data Prices (“Your data is sold”)

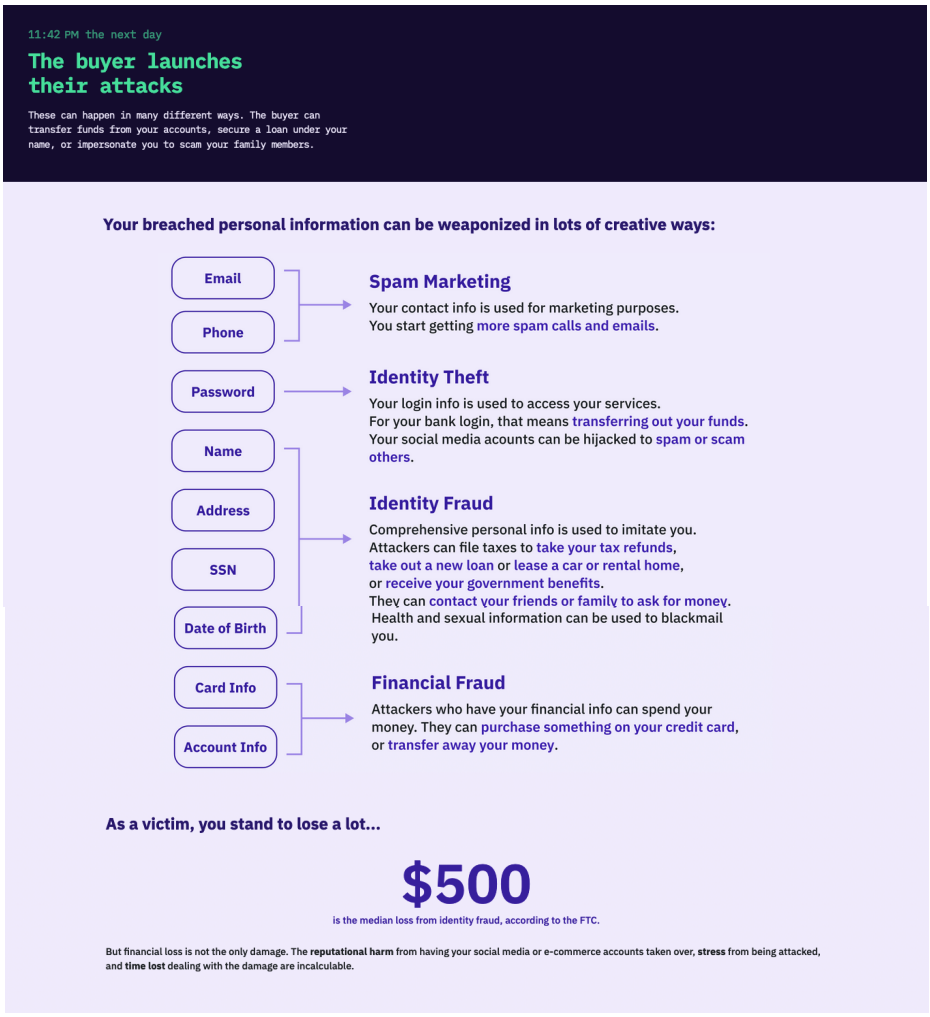


This section continues the narrative, using a bar chart to visually show how much the exposed data are worth and enable comparisons. Its purpose is to show that (1) personal information is cheap by itself but expensive when combined, and (2) the prices of data are dependent on how useful they are.

The animated bar chart uses progressive disclosure and interactivity to deliver the first message. Viewers step through a set of annotations that explain why each item is priced the way it is; then, an animation conveys the key message that bundling the data is more valuable. The scatter plot below (like the tree in the previous section) lets users dig deeper to get a sense of what the landscape is like through self-guided exploration. The categories into which data types were sorted were created from custom grounded coding.

Future work for this section would include (1) embedding a mini quiz for users to guess what the price of a SSN would be, and (2) converting the embedded Tableau plot into a D3.js chart that is better integrated with the rest of the site.

Attack Vectors (“Your data is weaponized”)

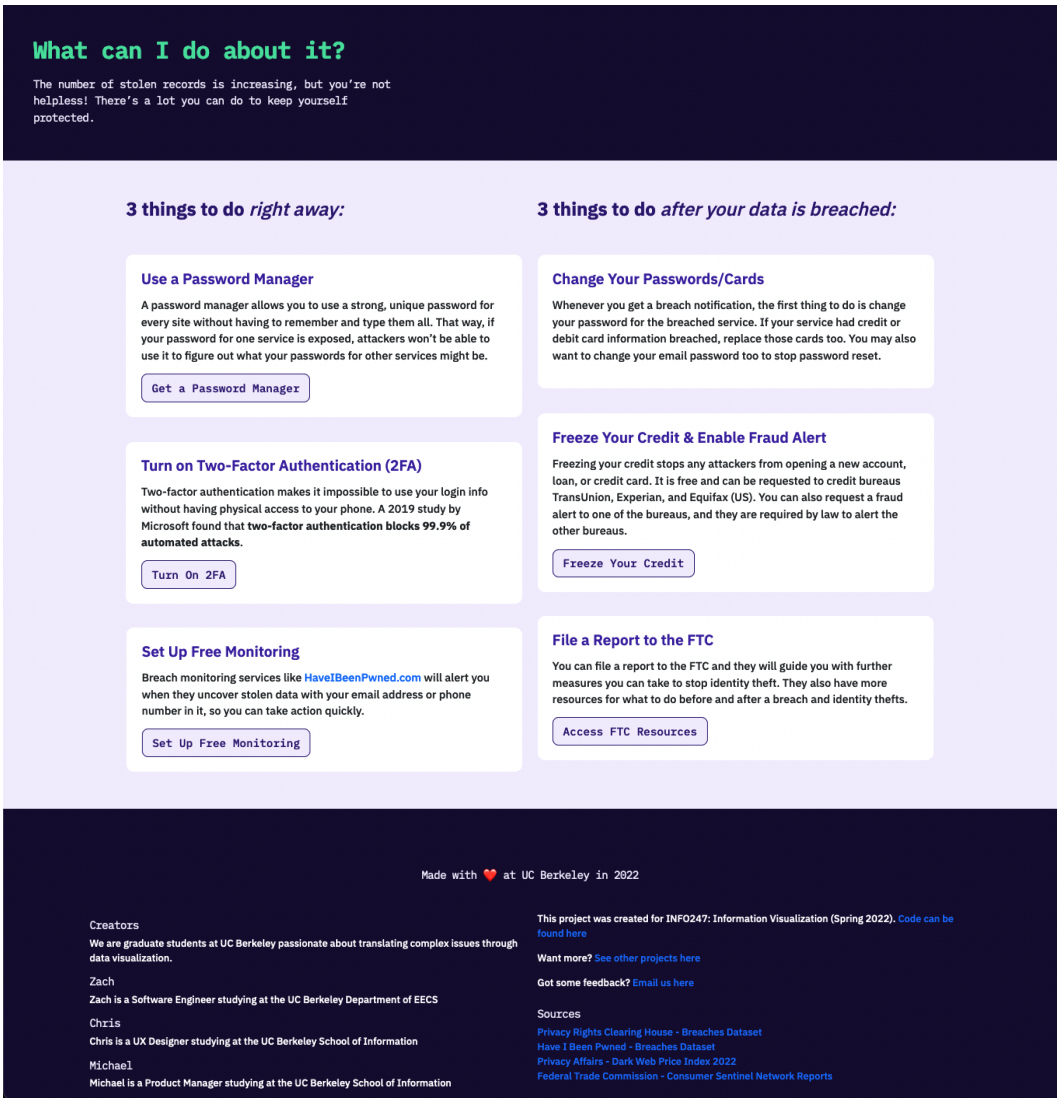


This section continues the narrative, showing how the stolen and/or sold data are used to impact you. Its purpose is (1) to illustrate how different types of data can be used for different kinds of attacks, and (2) what kind of potential financial harm these attacks could cause.

A diagram was used to illustrate the different kinds of attacks for which different data is used. An emphasis on the financial loss was put in place to communicate the potential financial harm.

Future work for this section would include (1) converting the diagram to be an interactive, exploratory component, where users can hover over different data types to see the possible attacks, (2) adding animation or interactivity to the $500 number to make it more salient, (3) adding more ranges and other data about financial loss, and (4) add a D3.js line chart showing the increasing number of fraud reports filed with the FTC.

Solutions & Footer (“What you can do”)



This section ends with what the user can do to avoid the consequences of data breaches that are explained in the previous sections. Its purpose is to inspire users to take some of the basic, free actions that can protect them. The footer includes the names of the authors, description and related links, and data sources used in the website.

To make this section easier to parse, we placed each recommended action in a separate white block. We also added prominent links and buttons to allow users to immediately take action as they read the section.

Future work would include (1) adding links to other resources where viewers can learn more about data breaches, and (2) adding Google Analytics to allow us to estimate what fraction of our visitors clicked each of the buttons corresponding to a recommended action.

Testing

Objectives

Usability testing was conducted with 3 participants navigating the first version of the project website. The overall goal of the usability study was to understand if (1) the content was engaging and (2) the design was easy to navigate as a whole, as measured through surveys and interviews outlined above. More specifically, the study sought to understand the following:

- Were the messages clear from the line charts?
- Did the participants understand how to interact with the interactive charts?
- Did the participants want to take the recommended actions?
- Did participants recall key messages from each section?

Participants

Participants were recruited from fellow students at UC Berkeley, as they were part of the target audience demographic of active users of online services who may not think about data breaches and their consequences. Additionally, each participant had a different area of concentration in their program of study–software engineering, data science, and UX research–which ensured that our participants could provide a wide breadth of critical feedback on design usability, data visualization techniques, and technical knowledge.

One potential limitation of this convenience sample of participants is that their relatively young age and high level of technical expertise does not necessarily reflect the background knowledge of other non-technical users, who are also part of our target audience. However, we received expert feedback from their respective domain of expertise.

Method

The testing utilized mixed-methods research, with participants answering both quantitative and qualitative questions about the perceived effectiveness of the design, which were asked both orally and via survey. All testing was conducted virtually on Zoom. Each stage, along with questionnaires, is shown below.

Stage	Questions

1. Pre-test Survey
 	
Before we began testing, we asked participants to fill out a survey with questions related to their prior knowledge and experience with data breaches to establish a baseline.
 	
Link to Survey
	• Informed Consent
	• Demographic Questions
	  ○ Undergraduate Major & Minor
	  ○ Current Program & Focus
	• Security Practices
	  ○ Do you use unique passwords for different services?
	  ○ Do you use password manager software?
	  ○ On how many services do you use two-factor authentication (2FA)?
	• Breach Experience
	  ○ How many data breaches have you experienced?
	  ○ Have you ever experienced identity theft/fraud?
	  ○ If yes, what kind of theft/fraud was it?
	  ○ How would you describe your knowledge of data breaches and their consequences?

2. Test Setup A 1-hour Zoom video meeting was scheduled for each participant. For the meeting, we asked each participant to enable their cameras, share their screens, and consent to having the meeting recorded.	• Consent to Record • Brief Introduction: “This is the first version of an interactive website about data breaches, Your feedback will help improve it.” • “As you explore the website, please think out loud, and tell us what you are seeing and understanding.” • “You will have a chance to provide more detailed comments on each section after you go through the whole site.”
3. In-test Formative Evaluation During the test, we asked questions to understand the participant’s experience using the website and to examine how well they understood the message.	• What is this visualization showing? • What stands out to you? • What are you thinking? • Is there anything else you would have wanted to see?

4. Post-test Survey

At the end of the formative evaluation of the website, we sent each participant a link to a form to complete.

The form was intended to measure:

1. **Effectiveness:** How likely participants were to take preventative measures after using the website
2. **Knowledge Retention:** How much specific knowledge from the website they retained
3. **Summative Evaluation:** How enjoyable, engaging, and informative the participants found the visualization to be

Link to Survey

*The question could have been phrased more neutrally, such as “please rate your subjective experience on the scale from...”

- **Effectiveness**
  (1 = Not at all → 5 = Definitely)
  On a scale of 1–5, how likely are you to...
  - ...Install a password manager?
  - ...Begin using unique passwords?
  - ...Enable 2FA on some of your key accounts?
  - ...Enable 2FA on most of your accounts?
  - ...Sign up for a breach monitoring service?
- **Knowledge Retention**
  - What kinds of organizations have been targeted more in recent years?
  - How many different types of personal information have been exposed in data breaches?
  - Which of the following is the most valuable type of data?
  - What is the median financial loss from identity fraud?
- **Summative Quantitative Evaluation**
  (1 = Not at all → 5 = Very much)
  How would you rate the design on the following*:
  - Enjoyable
  - Engaging
  - Informative

5. Summative Evaluation

Once the post-test survey was finished, we asked each participant to scroll back to the beginning of the page. Once at the top, we asked the participant to scroll through each section, rate its effectiveness on a scale of 1 to 5, and give more detailed comments on the content, design, and visualizations within that section.

- **For Each Individual Section**
  - On a scale of 1 to 5, rate the effectiveness
  - What changes could be made to increase the effectiveness of this section?
- **Overall Feedback**
  - What did you like about the website overall?
  - What do you wish you had seen or learned?
  - Which visualization was your favorite?
  - Which visualization was your least favorite?
- **Open-Ended Feedback**
  - Is there anything else you would like to share?
  - Have you seen any other resources related to data breaches that were helpful to you?

Quantitative Results

For each participant, we measured the amount of time they took to browse through the entire website and the amount of that time that was spent in each of the website's sections. We also collected from each participant quantitative evaluations of the effectiveness of each section's content and visualizations, and the site as a whole.

Time

On average, the participants needed **11 minutes** to go through the entire website. We believe this is an appropriate amount of time.

This might be longer than actual users would take to go through the website, as participants may have paused to make comments and state their reactions as they explored.

Sections	P1	P2	P3
1: Intro	0:30	0:30	0:30
2: Data Breach Definition	1:00	0:30	0:12
3: Your bank is breached (Sectors)	1:30	2:30	4:30
4: Your data is stolen (Data Types)	4:30	3:00	2:30
5: Your data is sold (Data Prices)	1:00	2:00	3:00
6: You get attacked (Attack Vectors)	2:00	0:45	1:00
7: What can you do about it? (Solutions)	0:30	0:30	1:00
Total (Minutes)	**11**	**10**	**13**

Per-Section Effectiveness

Overall, the participants found each section to be somewhat effective in communicating its message.

Sections 1 and 3 scored the lowest on average. Participants thought that the introduction could be improved with more visual components. And as their qualitative feedback revealed, the line charts in Section 3 could be improved to make it clearer.

Sections	P1	P2	P3
1: Intro	3.5	4	3.5
2: Data Breach Definition	4	3	5
3: Your bank is breached (Target Sectors)	3	3	4
4: Your data is stolen (Data Types)	4.5	3.5	4
5: Your data is sold (Data Prices)	3.5	5	5
6: You get attacked (Attack Vectors)	3	5	4
7: What can you do about it? (Solutions)	3	5	4.5
Average (Max 5)	**3.5**	**4.1**	**4.3**

Overall Effectiveness

Overall, participants found the website enjoyable, engaging, and informative. Some improvements can be made, but the overall impression was positive.

[P2]: "I never thought about the process, so it's nice to see the story."

[P3]: "With every step, you realize it can affect you, in multiple ways"

Sections	P1	P2	P3
Enjoyable	4	5	4
Engaging	5	4	4
Informative	4	5	5
Average (Max 5)	4.3	4.7	4.3

Call-to-Action

One of the purposes of our website is to inspire a call-to-action to viewers to improve their existing data security. **Two participants already use protective services, so answers may be skewed. Improvements we can make include adding links to data protection services.**

[P1] “Which [password manager] should I use? Provide me some suggestions.”

[P3]: “Looking at $500, I don’t want that to happen.”

Sections	P1	P2	P3
Install a password manager	3	1	3
Begin using unique passwords	3	1	5
Enable 2FA on key accounts	5	5	5
Enable 2FA on most accounts	5	5	5
Sign up for breach monitoring service	3	1	5
Average (Max 5)	**3.8**	**2.6**	**4.6**

Knowledge Retention

For the post-test survey, we provided a simple quiz to determine if participants retained some knowledge from the website content. Overall, participants did well on knowledge retention.

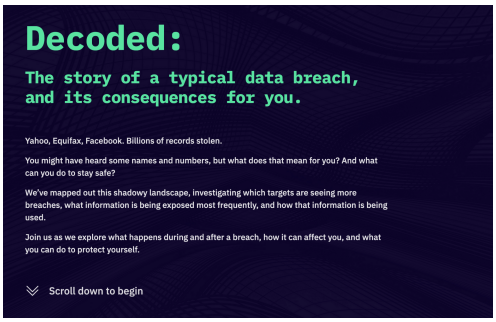
2 participants incorrectly answered the question about data prices. **They did not recall the prices of different data categories which were in bullet points.**

Sections	P1	P2	P3
What kinds of organizations are being targeted more in recent years?	O	O	O
How many types of information can be stolen?	O	O	O
Which of the following are the most valuable data?	X	X	O
What is the median financial loss from identity fraud?	O	O	O

Qualitative Results

Below is a list of select feedback from participants for each section, along with screenshots of each of the sections.

1. Introduction

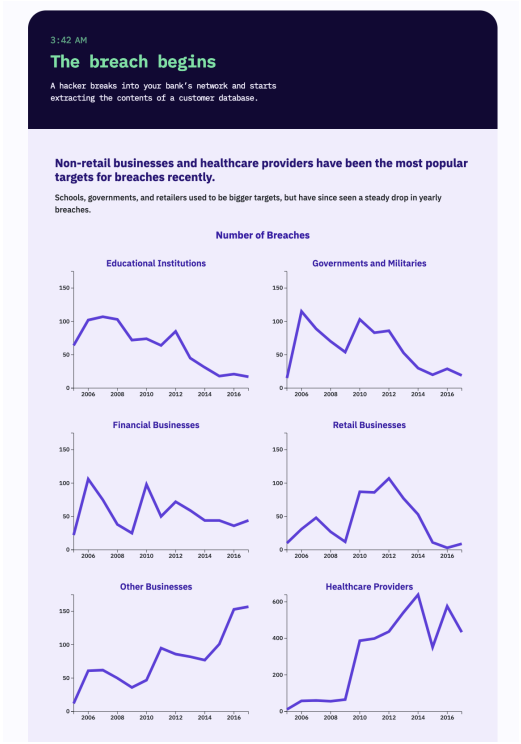


- [P1,2] Not sure what “Decoded” means. Sounded like a book.
- [P1] Add images to represent breaches
- [P1] Reduce the amount of text
- [P2] The amount of text is fine
- **[P2,3] Thought the “Scroll down to begin” instructional text was clickable**

2. Breach Definition

- **[P1,2,3] The bold purple text looks like hyperlinks, maybe just use black instead.**
- [P2] The transition from a dark to a white background is jarring
- [P2] Make the definition more concise

3. Breach Targets

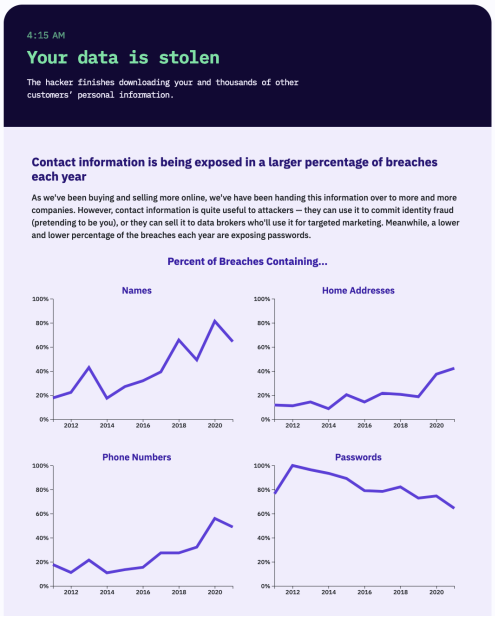


- Timer
  - [P1] Interesting but possibly misleading as to how long it actually takes
  - **[P1] Time is not thought about in vertical terms. Maybe show time in a horizontal progress bar.**
  - **[P2, 3] Show an outline on the side, vertically or horizontally, to give a sense of how long the website is and where you are within it.**
- Charts
  - **[P1,2,3] The charts are too big and did not all fit on the screen at once; needed to scroll back and forth to compare values**
  - [P1,3] Expected a data label or tooltip to show up when hovering over the lines
  - [P2] Retail business is in different shape
  - [P2] Y-axis scale different for last chart
  - [P3] Y-axis title would aid understanding

4. Data Types

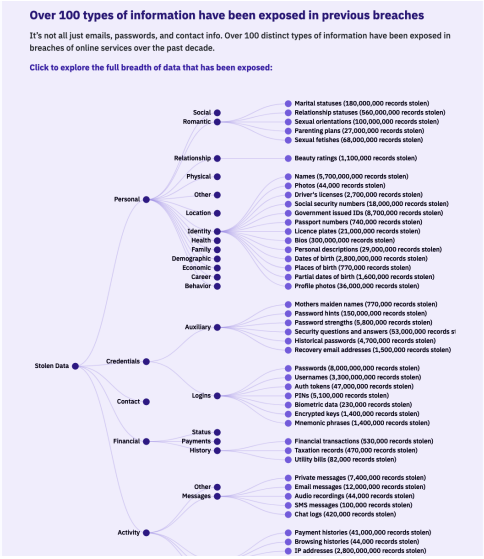
Line Charts of Trends

- [P1] Same comments as previous line charts, too big and hard to see it all. Adjust size or add interactivity
- [P1,2] Line chart message is clear. More personal information is being stolen.
- [P2] Home address and phone number were more shocking than name and passwords
- [P1] Show the Passwords chart first to establish the message early-on
- **[P2] The decrease shown in passwords does not seem necessary, it is disarming and counter to the message of the other charts**



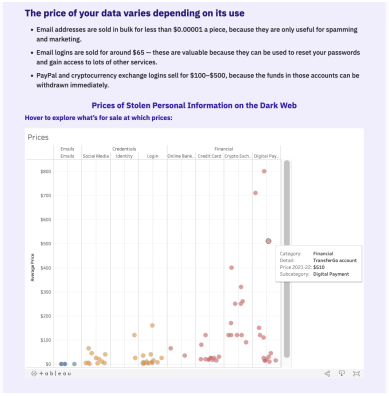
Collapsible Tree of Data Types

- **[P1,2] Clicked the text first after reading 'click'; Adding an animation or arrow to draw emphasis to the dot that should be clicked may help**
- Design
  - [P1,2,3] Enjoyed the interactivity.
  - **[P2] The most interesting items are hidden at the lowest nodes. Maybe highlight the shocking ones.**
  - [P2] Make 'identity' a separate category since there are a lot of interesting items under that category.
  - [P2] Expected there to be 'criminal' information, but not sure where to find it.



5. Data Prices

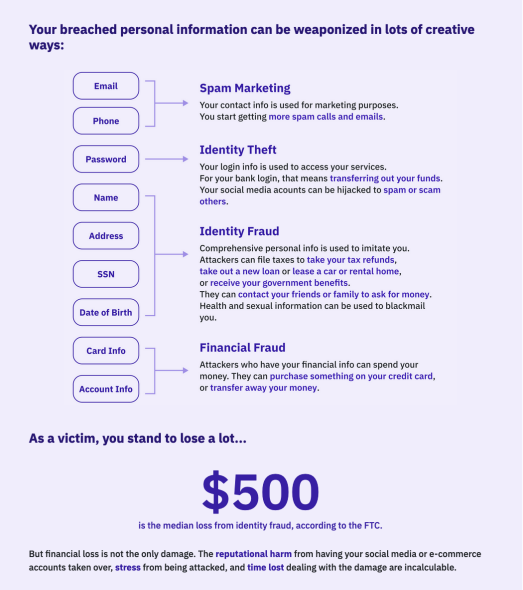
- Reactions
  - [P2] Very interesting to see the price of data
- Design
  - [P1,2] Adjust the fit and placement of the chart (smaller Y-axis, wider X-axis for fit, and center-aligned)
  - [P2] Colors are a bit confusing, maybe add legends



- Reactions
  - [P1,2,3] Found it engaging and clear
- Content
  - [P1,2,3] The message is clear — that combined information is much more valuable.
- Click
  - [P1] Clicked ‘combine’ twice thinking that it would further combine the other data.
  - **[P1,2] Expected clicking ‘combine’ again to break the site**
- Animation
  - **[P1] The bar animation could be faster, thought it was frozen for a second**

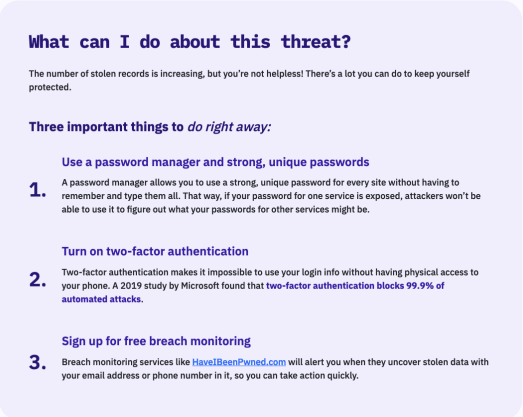
6. Attack Vectors

- Reactions
  - [P1,2,3] Diagram is very clear.
  - [P1,2,3] $500 loss figure is clear
- Content
  - [P2] What is the difference between identity theft and fraud? Make the distinction clearer
  - [P1,3] $500 seems like a lot.
  - [P2] $500 seems small, given that the cost of a combined profile is $1000. Show the distribution of losses if possible.
- Design
  - [P2] Make $500 red to make it clearer that it is a loss



7. Solutions

- **[P1,2,3] More specific links and action steps would be helpful.**
- [P1] Make the hyperlink open in a new tab.
- [P1] Plain text formatting does not convey urgency.
- [P2] Action steps seem clear.
- [P2] Wasn’t sure if it was the end.



Overall Results

Overall, the usability study revealed that the website was easy to navigate and the visualizations were easy to understand. However, there were several areas for improvement and suggestions to achieve that.

1. **How effective is the visualization in making users take preventative measures?**

   One of the purposes of our website is to inspire a call-to-action to viewers to improve their existing data security. Two participants already use protective services, so answers were skewed as they answered “Definitely.”

   All of the participants suggested adding links to specific resources in the call-to-action section to elicit action more effectively. For password managers, we plan to add links to 2 or 3 popular options. For two-factor authentication, we plan to suggest the most important services (finance, payment processor, social media) for which to enable 2FA, and also different ways to enable 2FA (e.g. SMS, an authenticator app, or a hardware security key). For breach monitoring services, we included a link, but two participants thought the linked service was not free; we plan to emphasize that it is free. One participant suggested that Google Chrome detects breached passwords; we may also note this.

2. **How effective is each chart in communicating the content?**

   All participants found the side-by-side **line charts to be suboptimal**. Given the provided text, they understood the key messages (healthcare and other businesses are targeted more, and personal information is being stolen more). Visually, however, they were rendered too big, making it difficult to see them at once. Even if they were rendered smaller, having to compare several charts was difficult. Per their suggestions, we will explore ways to combine and condense the charts into one or two interactive visualizations.

   All participants ranked the **collapsible tree chart and animated bar charts as their most favorite visualizations**, and the reactions we observed from them validate these responses. While they did understand where to click to interact with these charts, they reported that these click targets can be made clearer by (1) adding more visual elements to direct their attention to where they need to click, (2) making the text that explains where to click *not* look like a clickable hyperlink itself, and (3) disabling the “Combine” button after it has already been clicked.

   On the collapsible tree chart, **none of the participants expanded every single node and explored the full tree**. This is not a problem in itself, but as one participant suggested, it

may be worth highlighting some nodes to guide users to explore interesting items that they may not have browsed to themselves.

3. **Did participants recall key messages from each section?**

   All of the participants correctly answered at least three of the four questions on the post-test survey, which indicates that participants retained the majority of the key messages.

   Adding more iconography, as one participant suggested, will be helpful in making the content of the website more digestible and thus easier to navigate and remember; it'll reduce the cognitive load on users and draw their attention to key messages.

   2 participants incorrectly answered the question about data prices. They did not recall the prices of different data categories which were displayed as a bullet-style text list. Our design can be updated to make this more prominent.

4. **How engaging is the overall design?**

   All the participants noted that the **timer component of our story was interesting, but with room for improvement.** Two participants suggested keeping the time fixed on top and updating it as users navigated, or using a horizontal progress bar to represent time instead. We will discuss whether to keep the timer component, as one participant thought it actually took 30 minutes from the breach to data extraction.

   The storyline did not seem to have been a salient part of the experience, as no one specifically commented on it. This was expected in part, as some of the charts changed from the High-Fidelity Mockup and the interactive mini-quiz elements were not implemented. We plan to keep the storyline component and add these missing parts.

   All participants made comments regarding the length of the website, and how it would be better to know exactly where they are on the site. One participant suggested adding a fixed outline on the side that highlights as users navigate the sections, so they know how many sections remain. We plan to implement something like this to help improve the sense of navigation and progress during the experience.

Appendix

References

Research

2021 Comparitech Report on Combined Data
2021 LifeLock Post on Impact of Identity Theft
2014 RAND Corporation Report on Data Sales

Visualizations

Collapsible Tree on Observable

Assets

Usability Testing

Usability Testing Report
Pre-Test Survey & Informed Consent Form
Post-Test Survey
Testing Script & Notes

Data & Design

Project Repository
Datasets (Raw & Processed)
Figma High-Fidelity Mockup
Tableau Workbook for Data Prices Charts