VISUALIZING CYBER THREATS

By Braeden Diaz

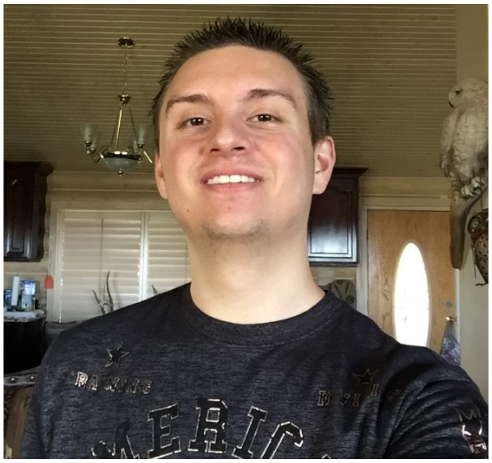
OCTOBER 25, 2019
CS6630
Version 1.0

Basic Info

Project Title: Visualizing Cyber Threats

Team:

Braeden Diaz, braeden.diaz@utah.edu, u0881315



I am a solo team, so for the rest of this book, I will be speaking in first person.

Background and Motivation

My main motivation for choosing this project is that cybersecurity has always been an interest of mine. It is interesting to see all the different ways in which different technologies are attacked in order to gain unauthorized access. It is also fascinating to see how technology can be modified to perform things they were never meant to perform. Because of this, it is also thought-provoking to see the many different ways in which cybersecurity experts attempt to protect and secure their software, websites, and devices from these attacks and modifications.

As for my background, I consider myself a generalist when it comes to Computer Science. That is, I like learning about the many different fields in computer science from low-level topics such as computer architecture and assembly programming, to mid-level topics such as basic software development, algorithms, and computer systems, to high-level topics such as web design and development, networking, and cloud computing systems. However, I have always placed a little extra emphasis on cybersecurity because all of the computer science topics that I have mentioned and much more all have a security component to them and it's important to understand the security aspects of each of these different areas in order to protect them from unauthorized access or potential damage.

Project Objectives

The primary objective of this project is to give a neat way of visualizing cybersecurity threat data. By doing this, it will help users to be able to answer some common questions when it comes to cybersecurity threats and its data such as:

1. Who is doing the majority of the attacking?
    a. What person is doing the majority of the attacking?
    b. What country is doing the majority of the attacking?
    c. What group is doing the majority of the attacking?
    d. What Government entity is doing the majority of the attacking?

2. What date and time did the attacks occur?
3. What date and time had the most attacks?
4. What user was attacked the most?
    a. What company was attacked the most?
    b. What server was attacked the most?
    c. What service/port was attacked the most?
5. What is the most common type of attack? Or, what protocol was used the most?
6. What does the data show? Not necessarily about individual attackers, but what does it show about the trends across attackers?

When it comes to cybersecurity, there are a lot of questions that can be asked, but the ones mentioned above are really common and this project will be an example of how visualizations and interactivity can be used to answer a lot of the above questions.

Data

Source

https://www.kaggle.com/casimian2000/aws-honeypot-attack-data

Acknowledgements

http://datadrivensecurity.info/blog/pages/dds-dataset-collection.html Jay Jacobs & Bob Rudis

The dataset I decided to use comes from authors Jay Jacobs & Bob Rudis. Through their book, blog posts and podcasts Bob & Jay hope to help security domain practitioners embrace and engage all elements of security data science to help defend their organizations.

The Dataset

The authors obtained the data from a friend named Daniel Blander who conducted an experiment in which he setup multiple honeypot servers on Amazon Web Services (AWS) in order to collect information on the attempts to attack them.

The dataset has 451,581 data points collected from 9:53pm on 3 March 2013 to 5:55am on 8 September 2013.

Dataset Columns

- **datetime** – The date and time of the attack.
- **host** – The hostname of the honeypot that was attacked.
- **src** – The src IP address of the attack of type long.
- **proto** – The protocol the attack used.
- **type**
- **spt** – The source port number. (From the attack)
- **dpt** – The destination port number. (On the honeypot)
- **srcstr** – The IP address of the attacker of type String.
- **cc** – The country code.
- **country** – The full country name from which the attack originated.
- **locale** – The specific country, state, or province from which the attack originated.

- **localeabbr**
- **postalcode** – The postal code (or zip code) from which the attack originated.
- **latitude**
- **longitude**

Data Processing

The data is provided in a Comma Separated Value (CSV) file and is in a pretty nice form. That is, there is no major data mining required. The primary type of data processing I plan on doing is aggregation and filtering in order to remove duplicates and to obtain categories in some cases.

Visualization Design

Design Idea 1 (Globe Only, Plus Charts)

A 3D orthographic globe that can rotate automatically or by being dragged by a mouse cursor.

The picture below shows an attack beginning from across the Pacific Ocean and heading

towards the United States.

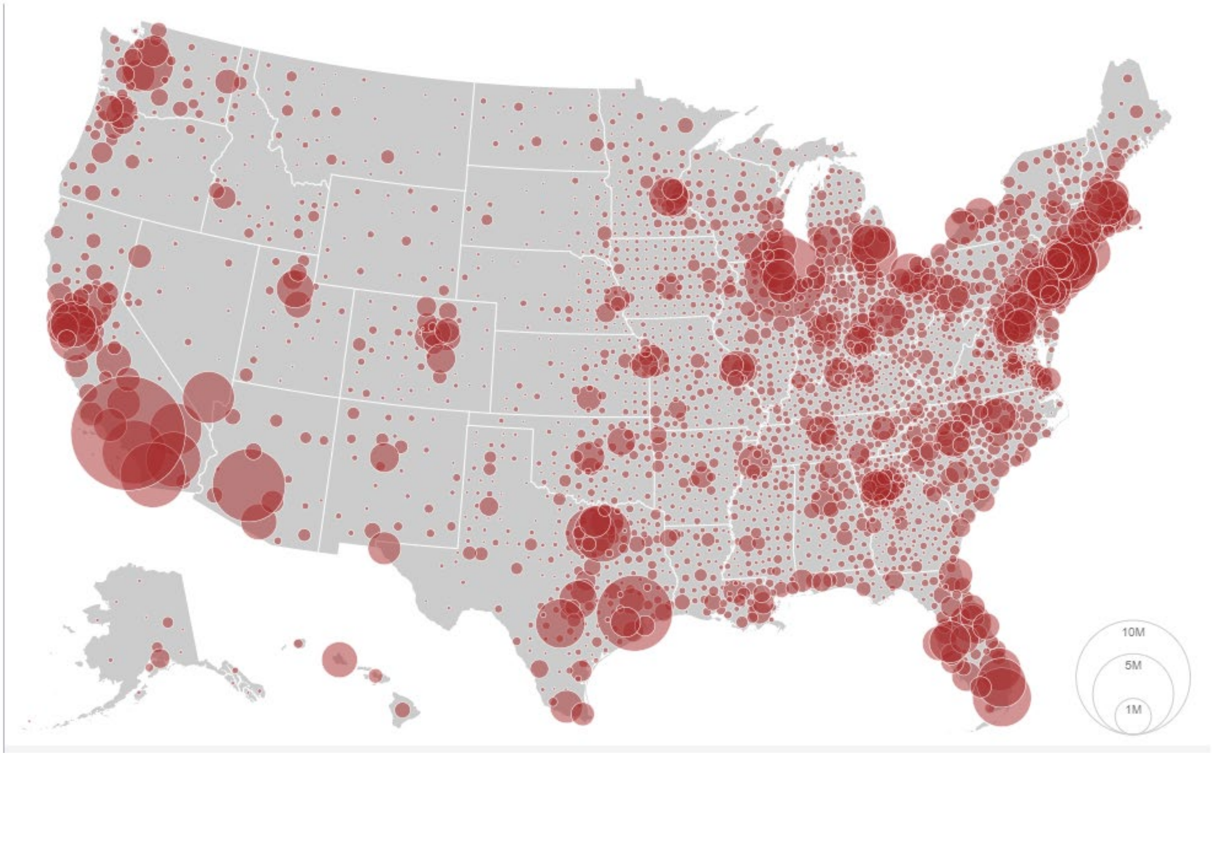
The next picture below show the attack happening over time.

Design Idea 2 (Map Only, Plus Charts)

This next design shows a somewhat flat projection of the earth. An attack will be visualized in the same exact way as the picture above from the last design shows. A red colorpelth will also be used to make the countries get more redder the more that they are attacked over time in order to show which countries were attacked the most.

Only certain countries such as the United States will be clickable. Once clicked a new or zoomed in projection of the country will be shown of the selected country as shown in the second picture on the next page. Red bubbles will then be used to show which states were attacked the most.

The benefit of this design over the first one is that there will be no need to deal with a 3D orthographic which will make this project much simpler.


10M
5M
1M

Design Idea 3 (Globe and Map, Plus Charts)

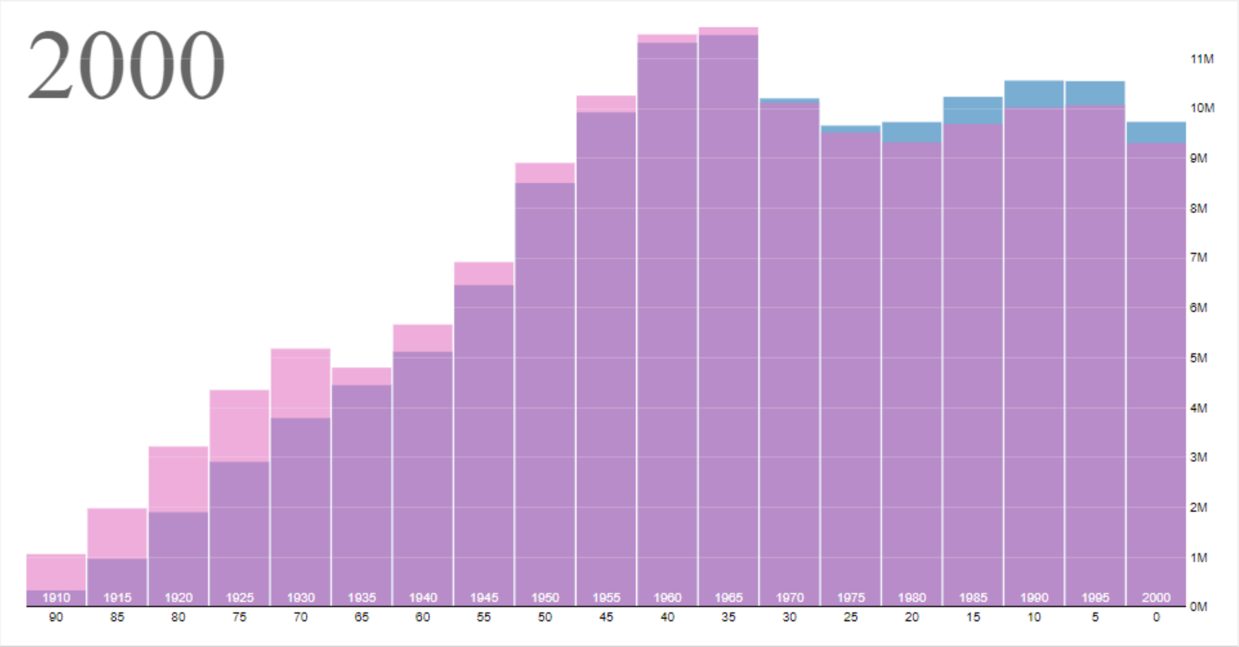
This design is basically designs 1 and 2 combined. I would start out with a globe that can be rotated either automatically or by dragging or both. I would them allow the orthographic 3D projection to be changed into the flat one as shown in the previous design. All of the features and interactions would be the same as the previous two designs.

The obvious con to this design would be that it will be a lot of work, but it is possible and would be neat.

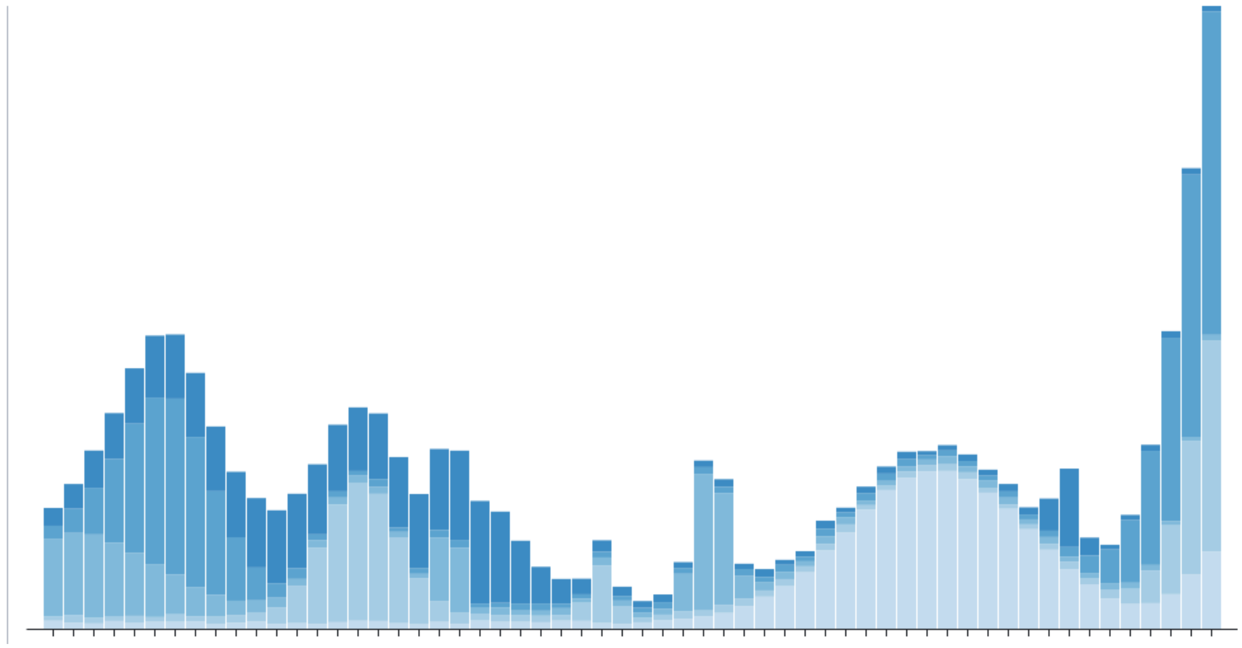
Charts (All Designs)

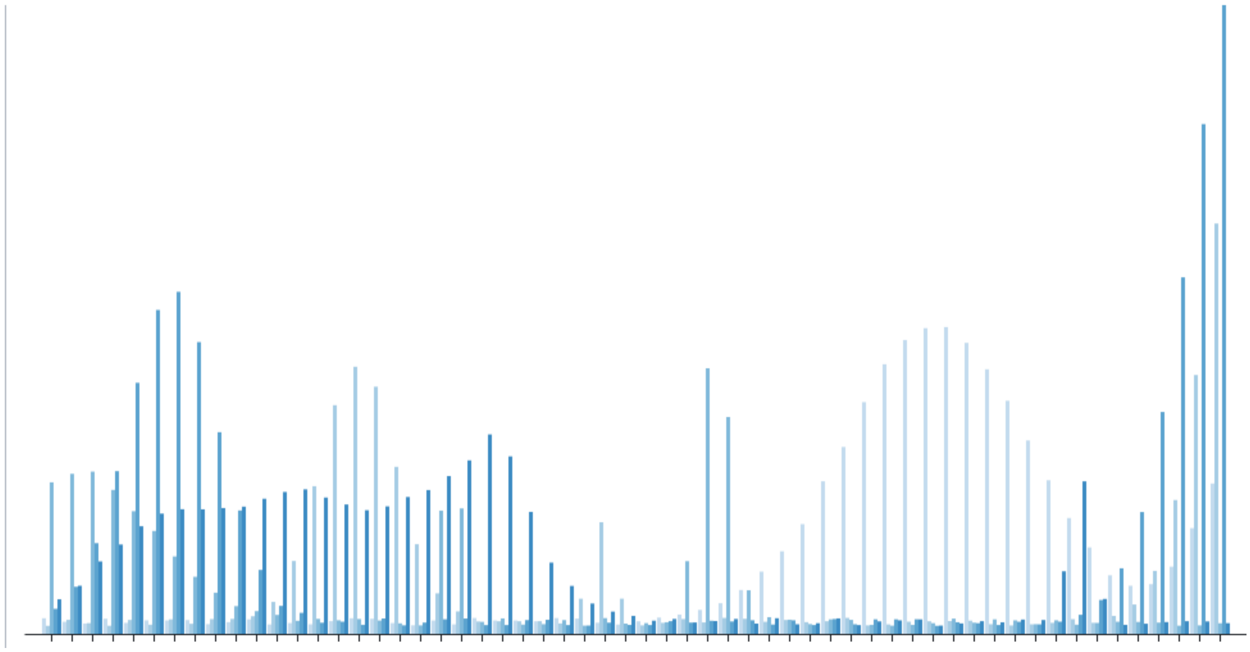
This section represents the visualizations that would be used in all three design ideas no matter what.

I plan on doing a bar chart that will show the amount of attacks over time and which country has been doing the attacks through the color of the bars.



I also plan on adding an interactive function that can turn the stacked bar chart into a grouped one as shown below.





Main Features

- The Globe or Map
- The storytelling (showing the data over time)
- Charts

Extra Features

- Globe Auto Rotation
- More Visualizations

Project Schedule

Week 1 (10/20/19 – 10/26/19) – Project Proposal

Week 2 (10/27/19 – 11/2/19) – Peer Review and Start Project

Week 3 (11/3/19 – 11/9/19) – Globe/Map and Globe/Map Interactivity

Week 4 (11/10/19 – 11/16/19) – Milestone and Begin Charts

Week 5 (11/17/19 – 11/23/19) – Work on and Complete Interactivity

Week 6 (11/24/19 – 11/30/19) – Finalize Project

Project Peer Feedback

Other Team: Rohit Singh, Sachin Boban, Lakshmi Ramasamy

Other Team Project: FiFA 20 Dashboard

Other Team's Feedback for Me

Rohit and Sachin gave me a lot of good feedback in regards to my project. Here are some of the things they mentioned for my project.

- Map Colorpelth
  - Rohit and Sachin mentioned that it might be useful to show a colorpelth on my world map in order to indicate the amount of cyber-attacks that occurred on a specific country.
- Map Circle Timeline
  - They also mentioned an idea for a visualization which is a scatterplot-like chart that shows the attacks overtime and how circle size/color and tooltips can be used to show more information.
  - This was a good idea as I would like to have visualizations that shows my overall dataset at a high level because right now, my idea is an animated world map and table that show the specific attacks over time but not overall stats of all the attacks.

- Map Arrows
    - They mentioned an idea to use different sized arrows in order to show attacks coming from the same place.
    - This was a good idea as it brought to my attention the issue of multiple attacks possible coming from the same place and how I might show that because I didn’t think about that before
- Filtering
    - Rohit and Sachin also mentioned that it would be useful to have some sort of feature that allows the user to filter through the data. For example, allow the user to select specific countries so that they can only see the cyber-attacks/stats from that specific country, etc.

Rohit and Sachin also warned me about too many circles/bubbles being on my world map as they may overlap making it difficult to see all of them. This is definitely true and I will have to figure out some way to handle this.

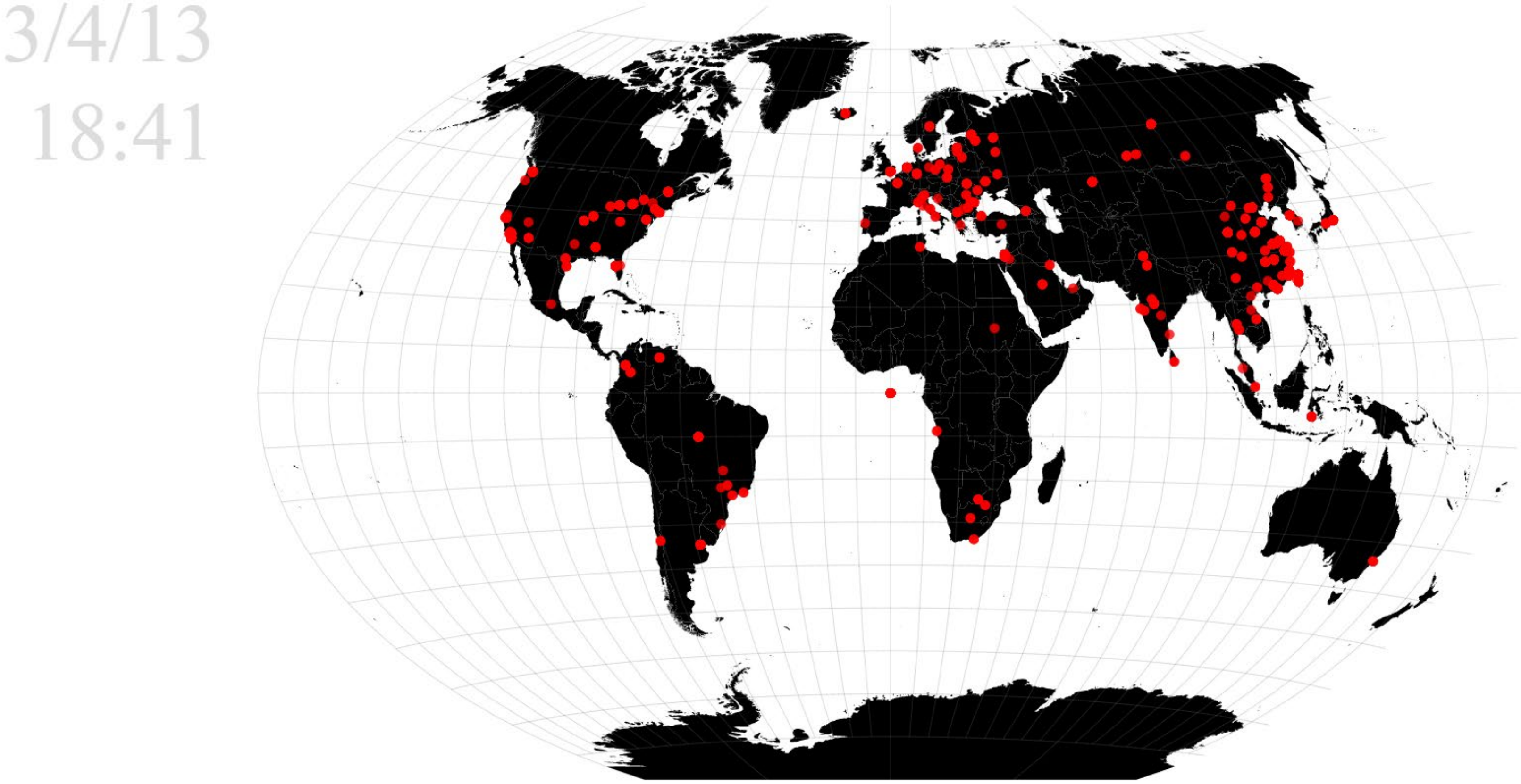
In addition, Rohit and Sachin was concerned about me showing too much data as my dataset it fairly large (over 450,000 entries). I was also concerned about this as well. They mentioned a lot of this could be solved with filtering which I agree with, but now I have to decide exactly what other visuals I can use to accomplish showing other more specific, aggregated, and filtered stats from my dataset.

Overall, Rohit and Sachin’s feedback was really helpful, they gave me some interesting ideas for visualizations and filtering which I can use in my project. They also brought some things to my attention that I didn’t think about before such as multiple cyber-attacks coming from the same attacker as mentioned before. They are also aware that I have a lot of ideas for visualizing my data and they really helped in warning me that I need to make a final decision as I can’t do too many visualizations nor can I show too much data. So, all in all, even though I was the last one to be grouped with Rohit and Sachin as I was an outlier and they arrived a little late, I am glad it was them I was grouped with as they helped a lot with their feedback and encouragement on my project.

Project Process

Main Visualization (Cyber Threat World Map)

The Main visualization for my project is a world map that contains points showing the location of the attackers. This works as the dataset contains the latitude and longitude of every attack that has happened on one of the honeypot servers.



The world map uses the Winkel tripel projection as it’s primary projection in order to minimize the most common forms of distortion: area, distance, and direction. The country borders are outlined with a light grayish color so that the user can tell the countries apart. In addition, light gray graticule lines were added to help the user see the projection and the distortion that it does have.

World Map Features

World Map Animation

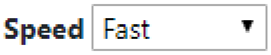
One of the next major features added to the world map was an animation that plays the attacks over time. Above the world map in the *controls* section, the user will find this button:

When the user clicks this button; the date for the world map starts increasing in time. As this time increases, more and more dots will appear representing the attacks that are happening at the current time.

It should be noted that when the animation rolls over to the next day, the map clears out and shows the attacks for that day. That is, the world map and animation will only ever show the attacks for the currently selected day and reset for each new day. This is useful as the user doesn't need to see hundreds of thousands of dots (attacks) on the map at once, only the dots (attacks) for the current day.

Speed

The next animation control is the speed:



Speed Fast ▼

This control allows the user to change how fast the animation plays. As expected, slow is slow and fast is fast. This can help the user see specific attacks happening at a certain time if they choose a slower speed, or it can help them quickly see clusters build up for the current day if they choose a faster speed.

Lock Attacks to Map

The final control is the following switch:

This switch allows the user to lock the attacks to the map. If on, the attacks (dots) will keep piling up for the day so that the user can see all of the attacks throughout the day all at once. If off, the attacks (dots) will only show up on the map when the date is equal to the date of the attack and then it will disappear afterwards. This is useful for users looking for specific attacks and not having to deal with overlapping of the dots.

World Map Scrapped Features

One of my original neat ideas for the world map was to allow the user to change the projection of the map into a globe projection and allow them to drag and rotate it, but due to time constraints, I could not implement this feature.

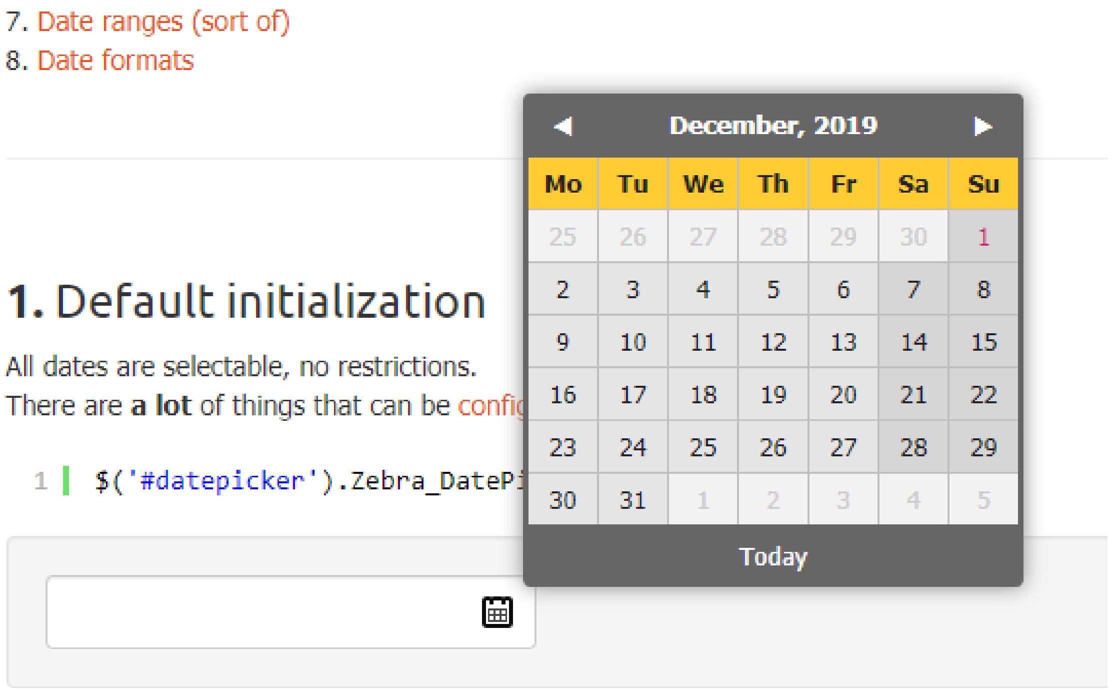
Another idea I originally had for the map was to use attack arcs instead of circles or dots, however, the dataset only included the latitude and longitude of the attacker and not the honeypots themselves. Although, the dataset did have the hostname of the honeypots which included their relative location so that I could of hacked up some way to implement the arcs, but, it didn't seem to really add to the visualization nor would it really help the user.

Finally, I wanted to either color the regions or add a choropleth to the map but chose not to do so for two reasons:

1. I already had another visualization that uses a heatmap to show the amount of attacks per day. This visualization is explained in the next section.
2. The dataset used a different abbreviation form from the ids of the world map paths and it would have taken me a lot of time to figure out how to get them to match up which wasn't worth it.

Heatmap Calendar and Date Picker

This next visualization wasn't in the original proposal but it was thought up pretty quickly after it. When it came to the animation, I knew I was going to need to give the user



7. Date ranges (sort of)
8. Date formats

1. Default initialization

All dates are selectable, no restrictions.
There are **a lot** of things that can be config

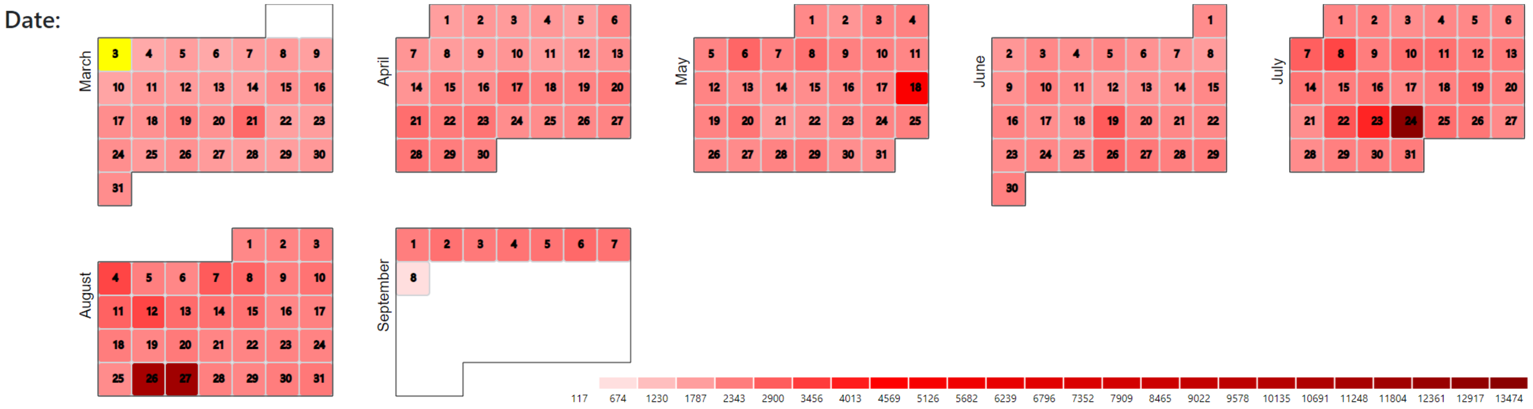
```
$('#datepicker').Zebra_DateP
```

◀ December, 2019 ▶

Mo	Tu	We	Th	Fr	Sa	Su
25	26	27	28	29	30	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5

some way to change the date so that they can skip to a certain date and see the attacks for that date. Thus, my first idea was to use a basic textbox-based date picker like this:

But then I got to thinking that it would be neat to actually turn my animation controls into visualizations themselves. That is, instead of using a boring old basic date picker, why not turn it into a custom d3 virilization? So that's exactly what I did.



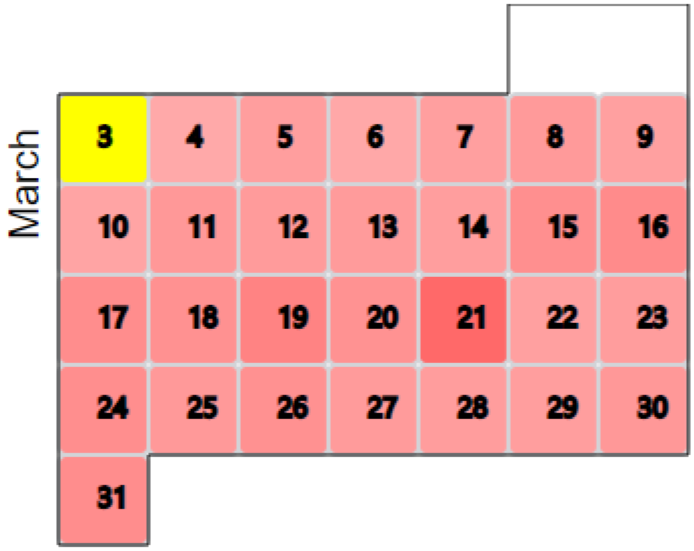
Above is the next visualization for my project. It is a heatmap calendar that is also a date picker. Each of the rectangles are created based off the dates in the dataset, and then, by using JavaScript's built-in Date class in order to turn each of the dates in the datasets into object, I was able to position the rectangles into their correct positions as they would be back in 2013 when the dataset takes place. I also was able to create a path around each month based on the date as well in order to bring out the unique shape of each of those months back in 2013.

Heatmap Calendar Features

Calendar Dates and Month Shapes

The first obvious feature of the heatmap calendar was that I had to create them based off of the dates provided from the dataset. This is why the only months shown above are between March and September and why March has the first couple days taken out of it while September has every date the 8th taken out of it as those were not in the dataset.

In addition, the rectangles are actually in their correct position according to the date, for example, let's look at March:

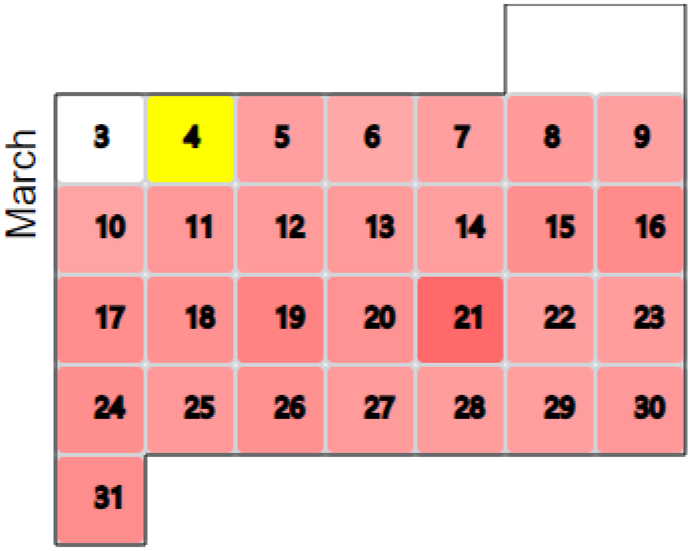


3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

As you can see, March 3rd of 2013 is in the Monday spot and that date was in fact a Monday just as March 30th of 2013 was a Saturday.

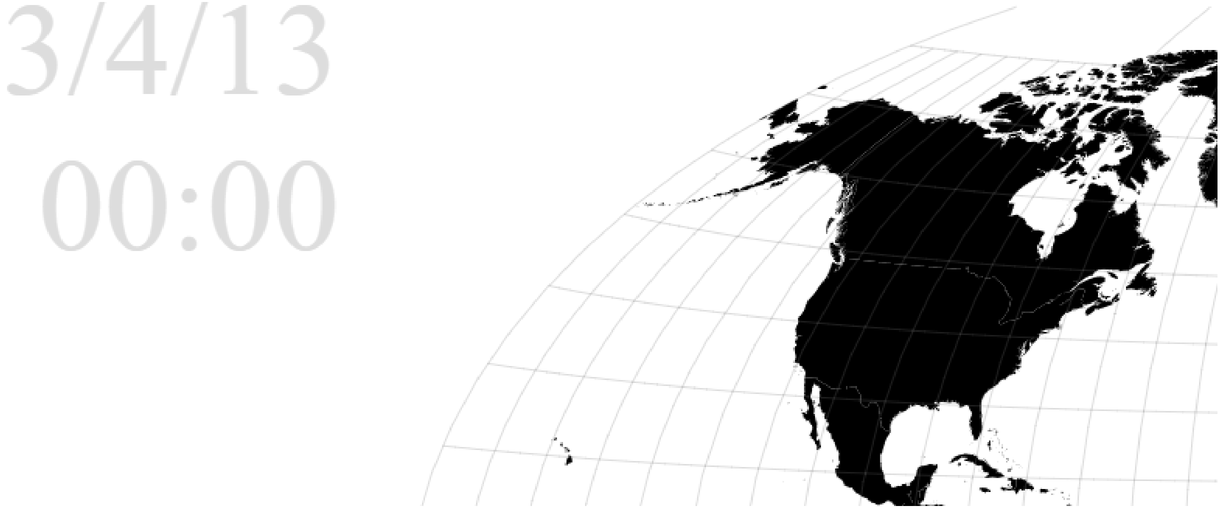
Selected Date (Date Picker)

Looking at the same image above, you'll notice a yellow square. This yellow square represents the current date that is selected. When the user clicks on a different square, that square will become highlighted in yellow, for example, now the 4th is selected:



March	3	4	5	6	7	8	9
	10	11	12	13	14	15	16
	17	18	19	20	21	22	23
	24	25	26	27	28	29	30
	31						

Changing the date also changes the date in the world map:

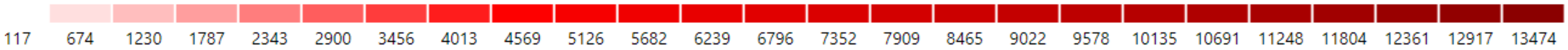


Therefore, when the user starts the animation, it will start from whatever the currently selected date is which the user can change at anytime.

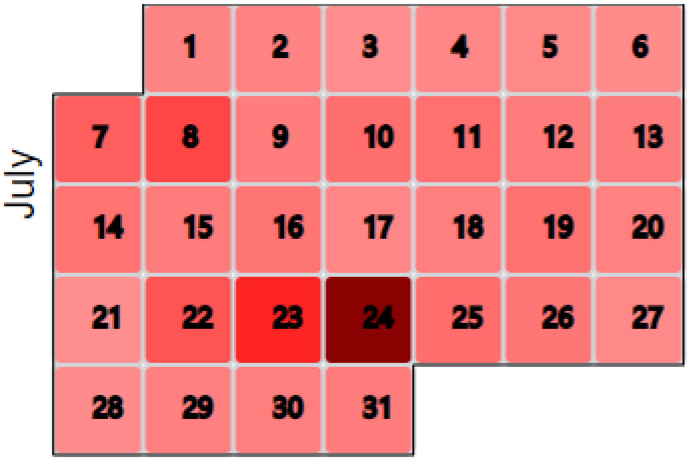
Heatmap

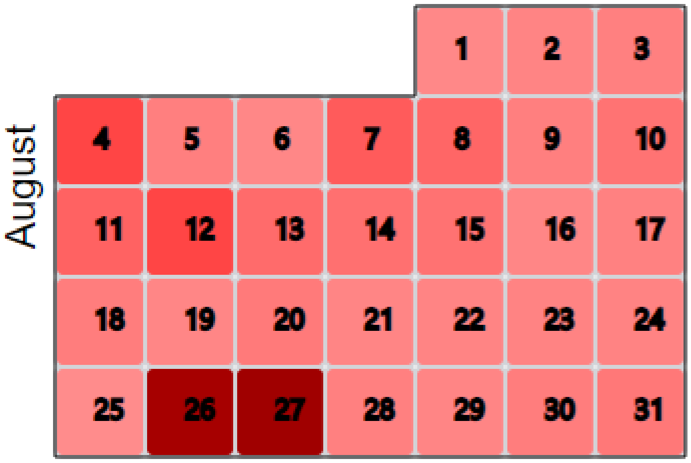
The last feature of the heatmap calendar is the heatmap itself. As seen in the previous images, the squares in the calendar are colored with different shades of red that represent how many attacks happened on that specific day. The lighter the color, the lower the amount of attacks, whereas the darker the color, the higher the amount of attacks.

To help the user get a better sense of the number of attacks, I have added a legend to the heatmap calendar as shown here:



Therefore, if we take a look at the months of July and August, we can see those months were hit the hardest out of the rest:



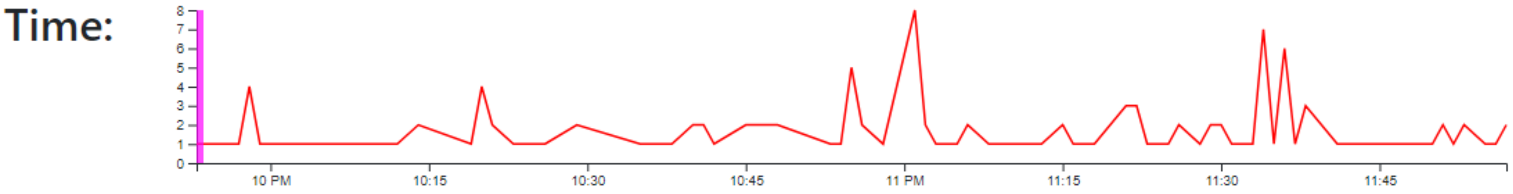


Specifically, we can see that the 26th and 27th of August were hit hard and the 24th of July looks like it was hit the worst.

Line Chart Time Slider

The next visualization is like the heatmap calendar in two ways:

1. It wasn’t originally planned in the proposal, but it was thought of quickly after it.
2. It is also used as a control to change a date.



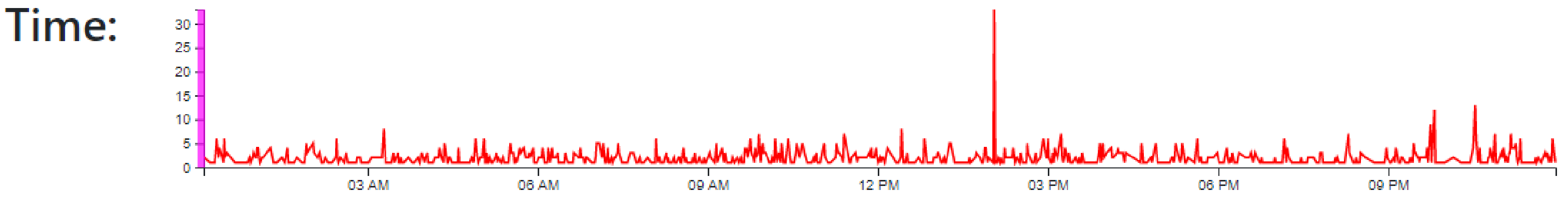
This visualization is the line chart time slider. It allows the user to see the number of attacks throughout a specific day and the approximate time they were at. For the exact time, they can look at the world map date and time.

Line Chart Time Slider Features

Line Chart

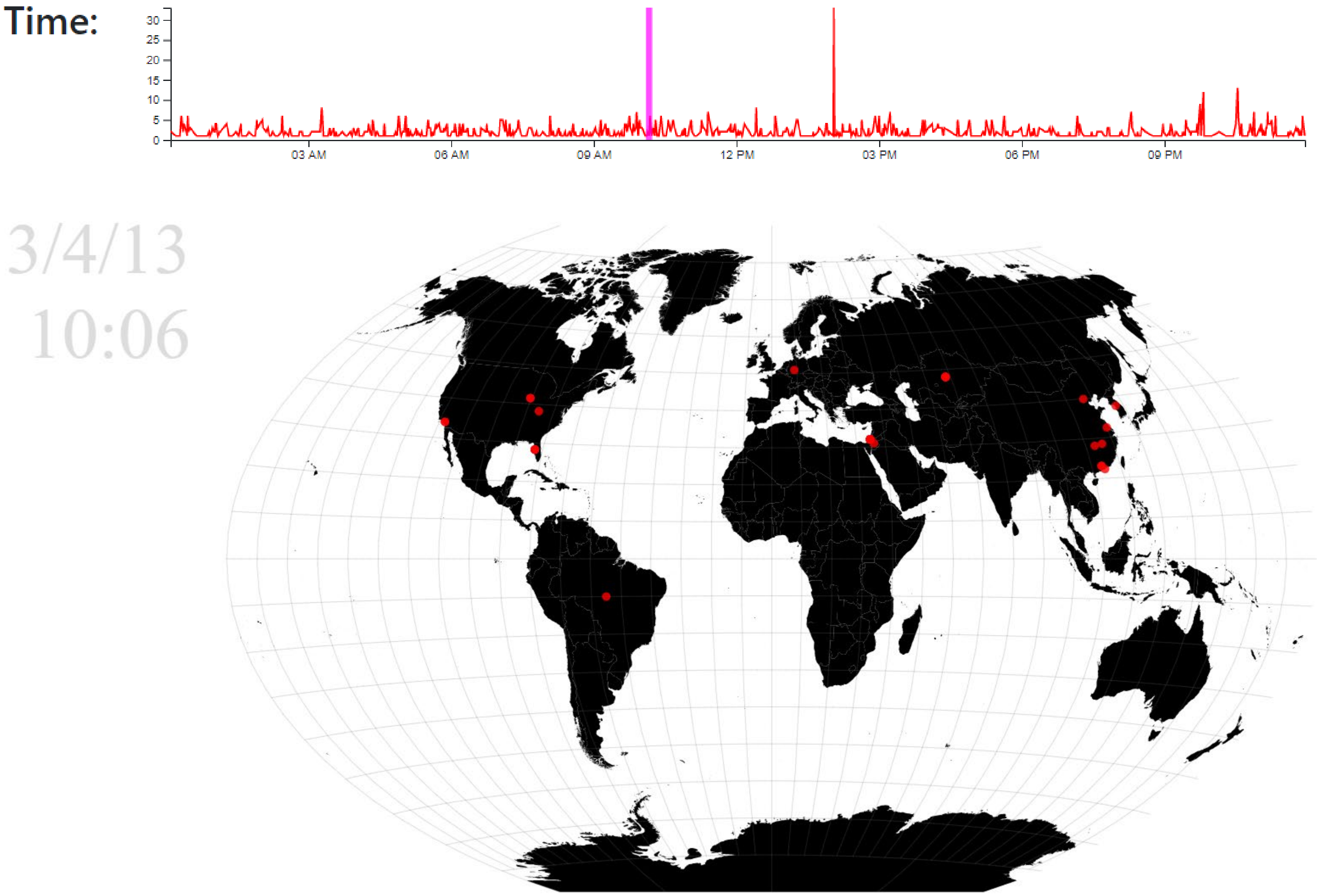
The first feature of the line chart time slider is that it contains a line chart that shows the number of attacks throughout a specific day and the approximate time they were at. The x-axis represents the time while the y-axis represents the total number of attacks at that specific time.

In addition, the line chart is connected to the heatmap calendar so that when the date is changed, a new line chart along with new scales and axes will be generated based on the newly selected date as shown in the next image below.



Time Slider

The next feature is the time slider which is represented by the tall magenta-colored rectangle on top of the line chart. The user can grab onto the slider and move it forward to see the attacks on the world map update throughout the day without having to play the animation.



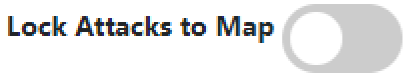
Animation

As the animation for the world map is playing, the slider will continue to slide based on the current time the animation is at.

Line Chart Time Slider (Known Issues)

One of the main issues with the time slider was if the user tries to go back in time. That is, if the user grabs onto the slider and tries to slide it backwards. This will work if the user has the switch that locks the attacks to the map off, otherwise, if they have it on, the world map will always show the points that the slider or date has already passed.

Therefore, to fix this issue, I simply don’t allow the user to slide the time slider backwards unless the switch to lock the attacks to the map is off:



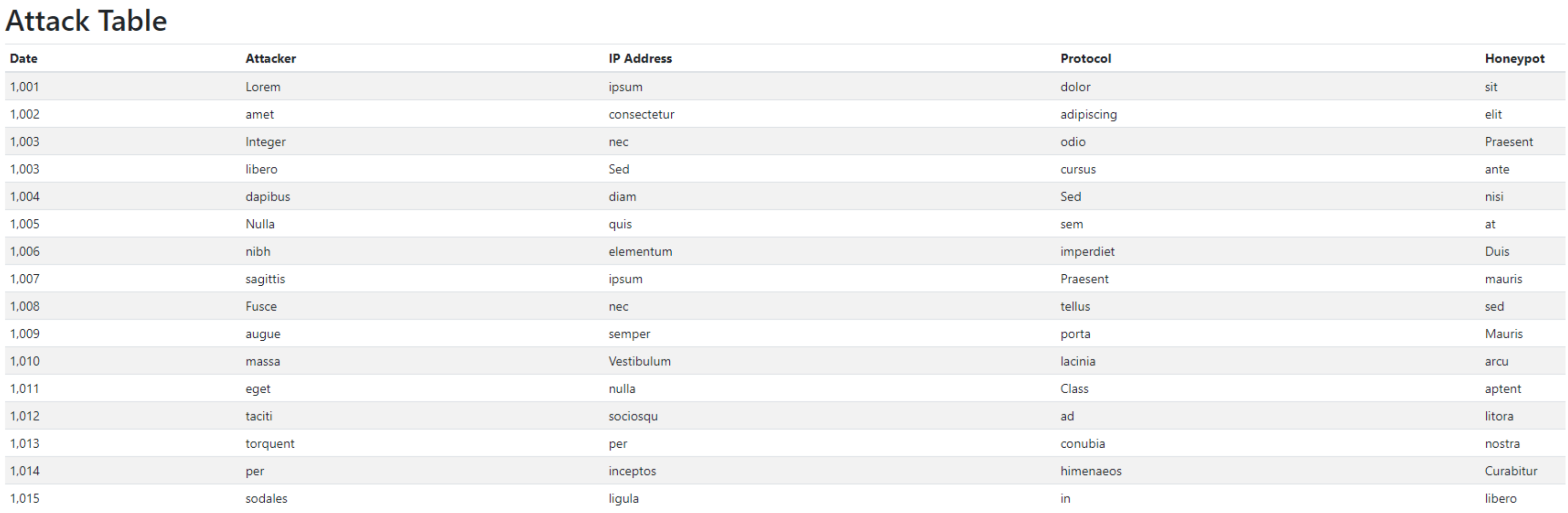
Lock Attacks to Map

Scrapped Features

Attack Table

Due to time constraints, I had to scrap two features from my project. This includes the attack table and bar chart.

The attack table was meant to be a simple table that shows the specific attacks for the current day. It would've looked something like this:



Attack Table

Date	Attacker	IP Address	Protocol	Honeypot
1,001	Lorem	ipsum	dolor	sit
1,002	amet	consectetur	adipiscing	elit
1,003	Integer	nec	odio	Praesent
1,003	libero	Sed	cursus	ante
1,004	dapibus	diam	Sed	nisi
1,005	Nulla	quis	sem	at
1,006	nibh	elementum	imperdiet	Duis
1,007	sagittis	ipsum	Praesent	mauris
1,008	Fusce	nec	tellus	sed
1,009	augue	semper	porta	Mauris
1,010	massa	Vestibulum	lacinia	arcu
1,011	eget	nulla	Class	aptent
1,012	taciti	sociosqu	ad	litora
1,013	torquent	per	conubia	nostra
1,014	per	inceptos	himenaeos	Curabitur
1,015	sodales	ligula	in	libero

Note that the data shown was just random placeholder data. The good news about scrapping this feature is that all of this data can still be see in the tooltips when the user hovers over an attack in the world map. In fact, more data is contained within the tooltips. Also, the table wasn't going to be a visualization anyways, just a normal table. Although, I did have ideas to aggregate and filter the table data to add visualizations like bar charts to the table, but time wouldn't allow it.

Ranking Bar Chart

The second planned feature that was scrapped was going to be a bar chart that was supposed to show the rankings of the top attacking countries or the top attacking users, etc. The good news is the user can still estimate some of these stats by referring to the clusters of circles within the world map and seeing which country they are mostly in.

Conclusion

One of the most difficult challenges I had to face when doing this project was the fact that there are so many visualizations out there and so many different ways to aggregate, filter, and organize the data that I have such that it was difficult for me to choose what visualizations to use. On top of that, the heatmap calendar date picker and the line chart time slider were two visualizations that I hadn't originally planned but I thought of quickly. But once I started implementing them, they ended up taking significantly longer than I expected especially when dealing with the fact that I had to connect them together along with the world map to work correctly when the date changes either manually by the user changing it through these visualizations, or by the world map's animation altering it. Therefore, these two visualizations ended up replacing the other more static ones I originally had in mind. But overall, I think I made the right decision as if I didn't use custom visualizations for the date picker and time slider; I would've still had to use very basic date and time pickers/sliders that would have still had to work like the current ones do and in addition to extra visualizations; it would've made my work much more difficult.

So, all in all, I believe I met most of my goal such that these visualizations are able to answer most, but not all, of the questions mentioned back in the proposal above.