VISUALIZING CYBER THREATS

By Braeden Diaz

OCTOBER 25, 2019

CS6630

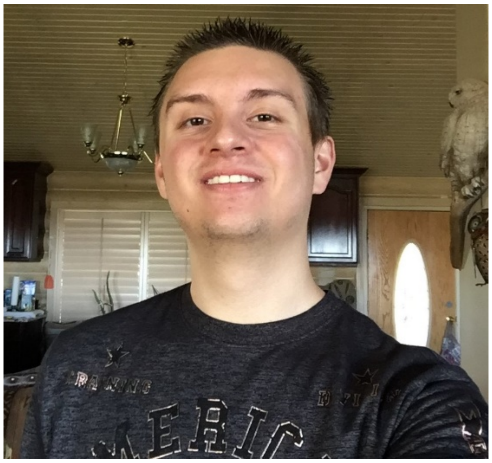
Version 1.0

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** – The latitude of the
- **longitude**

Data Processing

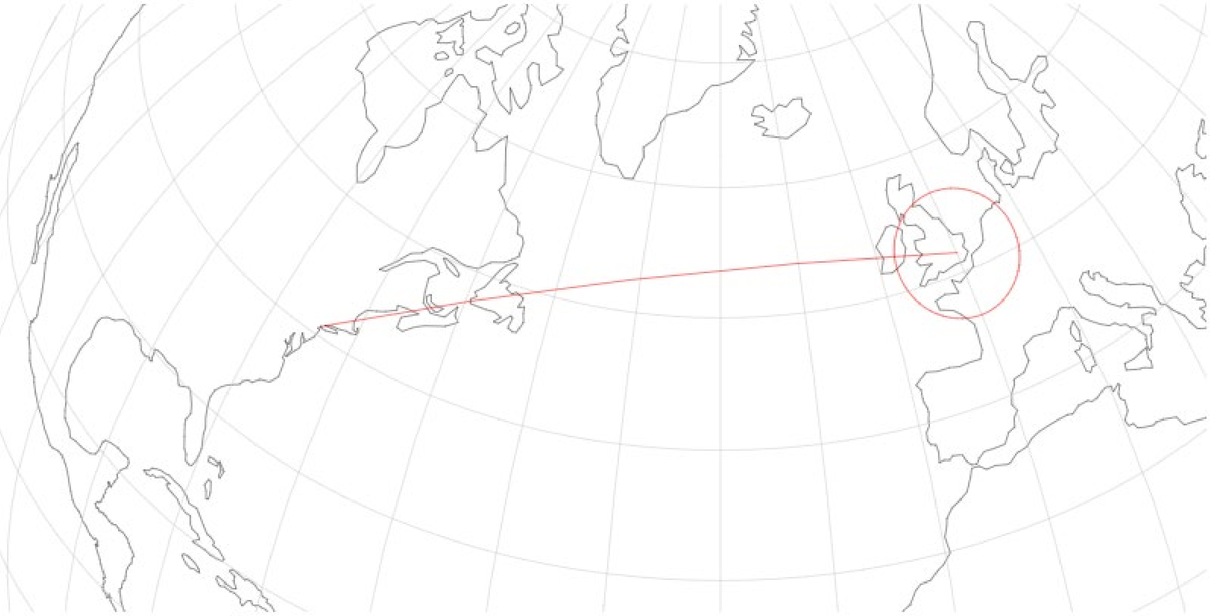
The data is provided in a Comma Separated Value (CSV) file and is in a pretty nice form. That is, there is no major data mining required. The primary type of data processing I plan on doing is aggregation and filtering in order to remove duplicates and to obtain categories in some cases.

Visualization Design

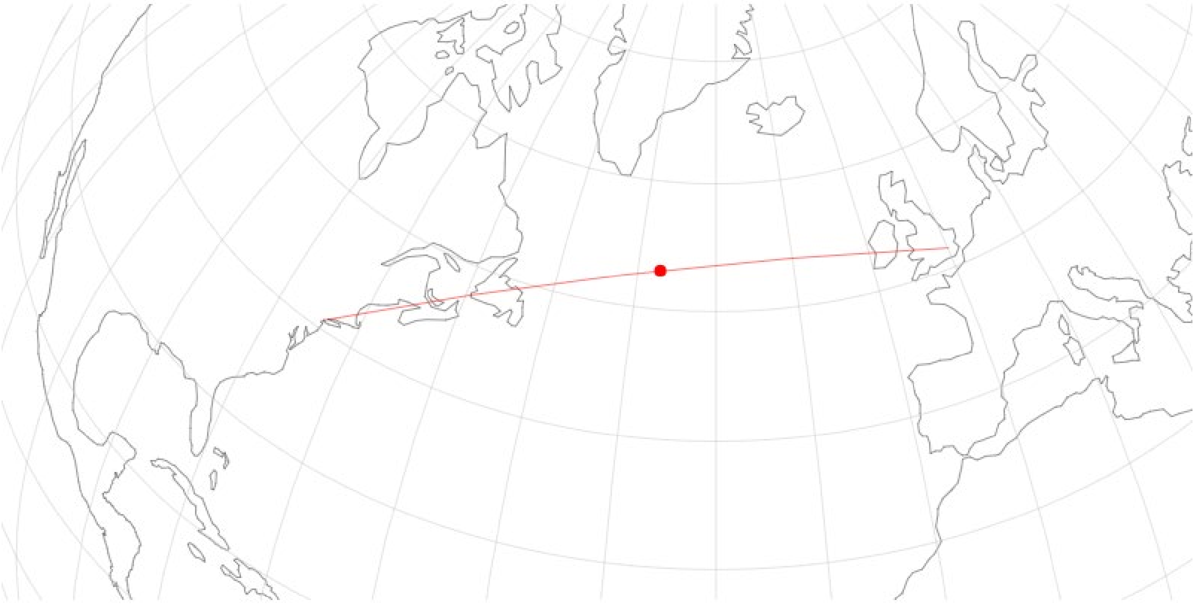
Design Idea 1 (Globe Only, Plus Charts)

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.

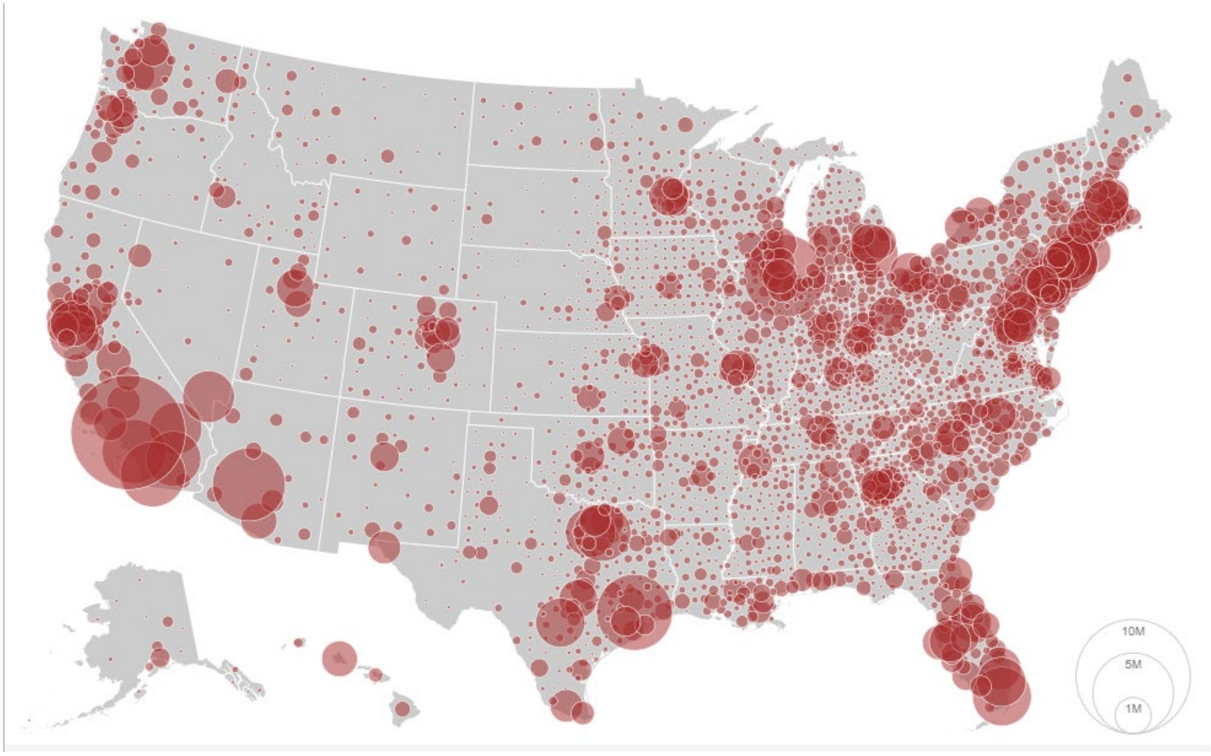


Design Idea 2 (Map Only, Plus Charts)

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)

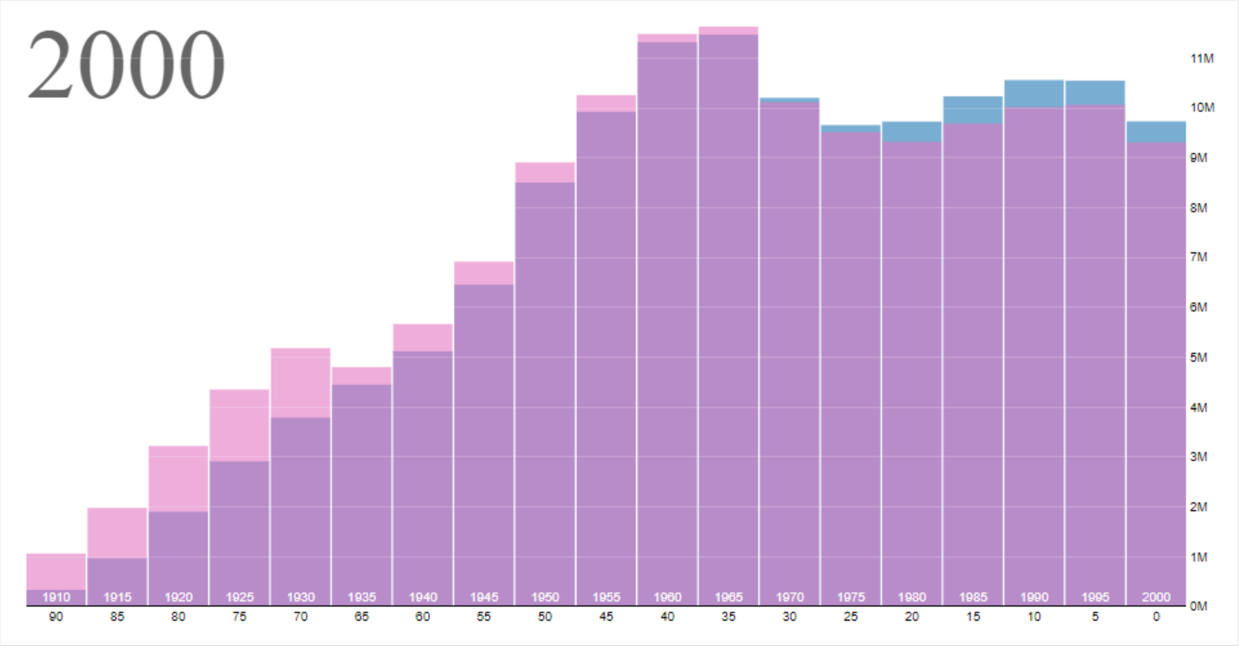
This design is basically designs 1 and 2 combined. I would start out with a globe that can be rotated either automatically or by dragging or both. I would them allow the orthographic 3D projection to be changed into the flat one as shown in the previous design. All of the features and interactions would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

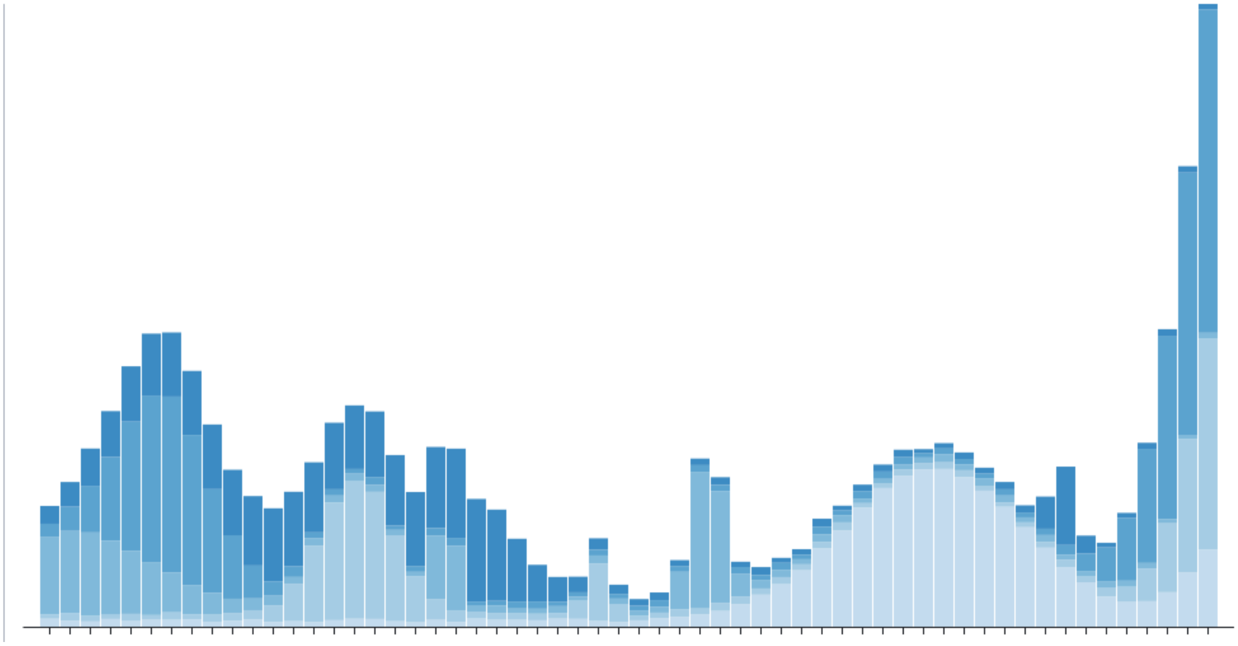
Charts (All Designs)

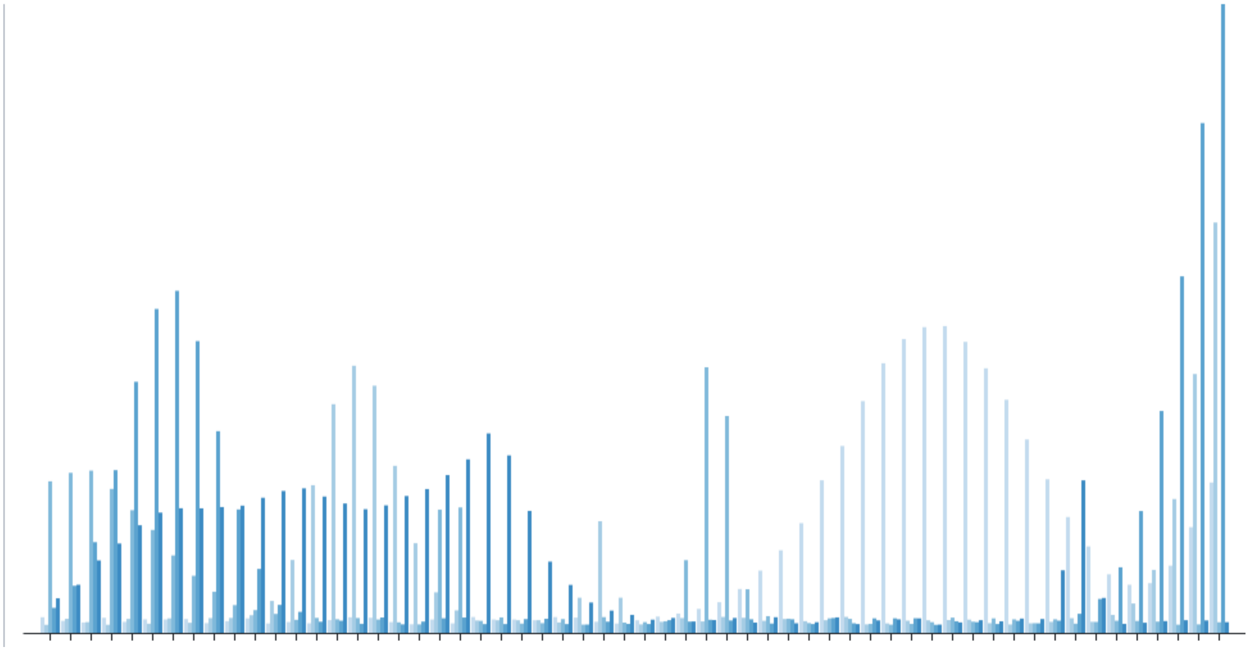
This section represents the visualizations that would be used in all three design ideas no matter what.

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