American Crime: A Visualization on Data About Police Spending and Crime Throughout the United States

James Deromedi, Ziyang She, Karefa Kargbo

Department of Computer Science

American University

Washington, DC 20016, USA

(Dated: December 14, 2018)

Abstract

The creation of a interactive, informative website that breaks down the big data of crime and police forces into a usable tool. Government officials and law enforcement agencies could improve budgeting and effectiveness of deployment with the system designed. By using public access data from the federal government, each states data was broken down and overlaid on an interactive map.

I. BACKGROUND

In 2017 alone, there was 1,274,321 violent crimes that took place within the United States of America. While this number has been slowly decreasing since the year 1991, it is an increase from 2016 by 0.2 percent. In 2016 there was reported, 701,169 sworn police officers serving a population base of 323 million, which averages to about 2.17 officers per 1000 citizens. In this paper we will talk about how our project aims to give these limited number of officers the best chance to lower the number of crimes in the United States. The layout of the paper is as follows, the data that was gathered, the software used to visualize the data, the website that was built, and lastly future work that can be completed.

II. INTRODUCTION

The project started as an idea to create a free standing Java project. This will allow the users to download the file onto their computer and have the ability to view it even while offline. This system has two major downfalls, one is that the data can not be updated without re-downloading the file or having to implement extensive security protocols. Secondly, if the user is on a secure network, like a government computer or major company, a .exe file will be blocked by the firewall, since most aggressive attacks happen in downloaded executables. We decided to move our project away a standalone file into a web hosted system. For the website, we initially were going to use the language PHP, to be able to dynamically access the databases. Since PHP is able to easily access a SQL database that can hosted on a server. The issue with a SQL database is possible security breaches, while the data we would have placed into it is not personal or confidential. We did not want to build the site around an insecure system. If our project grew and secure data was added, we would have to rebuild the entire back end to be able to handle the security. Therefore we realized that a simple HTML/Javascript/CSS site would be a better fit for our project. This type of website is called client hosted, or is running explicitly on the computer that opened the site. While this does have security downfalls, it is much easier to add security to the back end in future uses. The data that was used in creating the map visualization was uploaded to the softwares secure servers. This allowed two things, one was the the softwares' access to the data what we wanted to be visualized. And secondly, the secure server that allows the

Law Enforcement Officers Assaulted and Injured with Firearms or Knives/Other Cutting Instruments						
Month of Incident, 2011–2015						
Month	**Total**	**2011**	**2012**	**2013**	**2014**	**2015**
Number of victim officers	**532**	**110**	**105**	**126**	**106**	**85**
January	**59**	11	12	15	10	11
February	**46**	10	8	12	8	8
March	**58**	12	12	11	11	12
April	**43**	9	9	8	9	8
May	**32**	10	9	1	7	5
June	**35**	11	6	10	7	1
July	**42**	10	10	6	8	8
August	**46**	6	12	12	9	7
September	**40**	4	7	14	7	8
October	**52**	7	5	26	9	5
November	**39**	7	6	7	9	10
December	**40**	13	9	4	12	2

FIG. 1. A sample of the police data that was compiled from the large data set. This example shows the months that an officer was assaulted

system to dynamically be changed within the website by the map software.

III. DATA

The data for this project was taken from various sources, such as data.gov and FBI.org. We compiled about 200 Mbs of data containing everything from shootings to the type of tattoos of the shooter. Therefore, we had to manually compile the data into a workable set that we can use to visualize. We discussed using Python or Java to run a script to parse the data for us. Since our group had a lack of skill in both languages, we decided to manually parse the data. First we went through and pulled out all the data that contained a state identifier, seen in Fig. 2, this was gong to be the main data that we were going use to visualize. Secondly, we went through and compiled the rest of the common variables. An example of these common variables are; time of day, day of week, race/age/sex, seen in Fig. 1. While we did not visualize these set of variables we thought that they are important in understanding the data as a whole. Therefore we decided to place this data in the website as generic tables.

The next task with the data was to find an appropriate place for the data. Since our decision was to use a web based platform, we had to find a way to access the data and do so in a dynamic way. Our newly parsed data sets were in .CSV files or comma separated variables. First we placed the files into our project Google drive, since it allows each group member to access and edit the data, it also allows the table to be dynamically altered as new data is released. As the project went on, we ended up placing the data into a folder with the

State	Officers Ass	# of reportin	# of officers	violent_crim	homicide	rape_legacy	rape_revisec	robbery	aggravated_	property_cri	burglary	larceny
Alabama	0.00	94.00	4,858.00	5,391.00	59.00	648.00	900.00	761.00	3,671.00	20,806.00	3,511.00	15,249.00
Alaska	0.00	0.00	0.00	22,957.00	348.00	1,449.00	2,037.00	4,612.00	15,960.00	144,785.00	35,265.00	99,182.00
Arizona	6.00	91.00	12,260.00	15,769.00	189.00	1,425.00	2,008.00	2,117.00	11,455.00	97,391.00	22,851.00	68,720.00
Arkansas	0.00	256.00	5,327.00	27,968.00	306.00	2,268.00	3,109.00	6,360.00	18,193.00	207,184.00	38,010.00	152,388.00
California	13.00	683.00	76,351.00	166,883.00	1,861.00	9,341.00	12,811.00	52,862.00	99,349.00	1,024,914.00	197,404.00	656,517.00
Colorado	5.00	220.00	11,696.00	17,348.00	173.00	2,422.00	3,328.00	3,315.00	10,532.00	144,044.00	23,490.00	104,532.00
Connecticut	1.00	102.00	7,741.00	7,938.00	116.00	585.00	798.00	2,925.00	4,099.00	65,610.00	10,286.00	48,898.00
Delaware	0.00	56.00	2,239.00	8,538.00	162.00	365.00	494.00	3,746.00	4,136.00	31,444.00	2,971.00	25,208.00
District of Colum	0.00	1.00	3,826.00	4,760.00	62.00	257.00	361.00	1,240.00	3,097.00	25,545.00	4,797.00	19,555.00
Florida	0.00	271.00	36,264.00	93,625.00	1,041.00	5,396.00	7,552.00	21,137.00	63,895.00	570,270.00	109,268.00	420,341.00
Georgia	3.00	322.00	18,870.00	38,891.00	622.00	2,343.00	3,257.00	12,545.00	22,467.00	316,013.00	68,149.00	220,715.00
Hawaii	1.00	2.00	2,361.00	4,201.00	19.00	561.00	400.00	1,203.00	2,418.00	54,346.00	6,557.00	42,010.00
Idaho	0.00	102.00	2,692.00	8,619.00	73.00	877.00	1,205.00	1,050.00	6,291.00	64,688.00	14,977.00	45,328.00
Illinois	0.00	1.00	280.00	3,593.00	32.00	508.00	703.00	191.00	2,667.00	28,912.00	6,124.00	20,916.00
Indiana	0.00	168.00	5,117.00	49,755.00	754.00	3,602.00	4,926.00	15,056.00	29,019.00	256,962.00	46,846.00	192,424.00
Iowa	0.00	229.00	4,589.00	25,373.00	370.00	1,757.00	2,386.00	7,081.00	15,536.00	171,878.00	34,404.00	123,940.00

FIG. 2. An example of the data was parsed by state identifier This data set contains about 45 different variables that can used within the visualization software.

main website files. This decision will explain further in the map visualization section. But placing the data with the website files prevents us to dynamically change the data, since we will have to re-upload the files as each new year comes out.

IV. MAP VISUALIZATION

The interactive maps were created with a software called CARTO. This is a free visualization tool for students with a verified Github account, it can also be paid for otherwise. The general concept works by creating a GUI that can be used to manipulate the data without needing to code it. The software also allows data to be imported into their secure software. Each map contains multiple layers that are broken down by a common variable (states, zip code, etc) and then the data attached to each variable is laid onto the map. Once the data has been overlaid, multiple different analytic systems can be applied, be it adding dynamic widgets that create histograms of the data and color coding the map. It also has the ability to combine multiple layers, this gives the user the chance to show correlations of data between states. For example, combining total crime and total police officers can show states that have high crime and low law enforcement or any combination of the two. This system was chosen because it allows a large degree of freedom in creating the visualization. As well as, it can easily be embed into a website, allowing a more interactive site that will be dynamically changed as the software updates the maps. The style and actions of each map can manually be altered using Javascript and CSS using their custom command line. This feature allows more fine grained changes to the visualizations. For example, if we wanted the

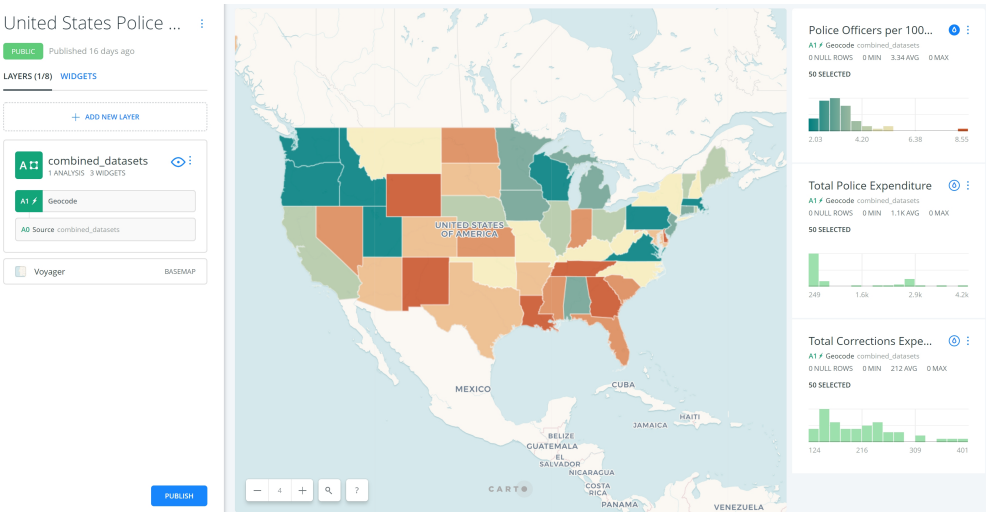


FIG. 3. An example of a single map created in the CARTO software. Showing the application of a single widget and how it dynamically changes the map by the histogram created.

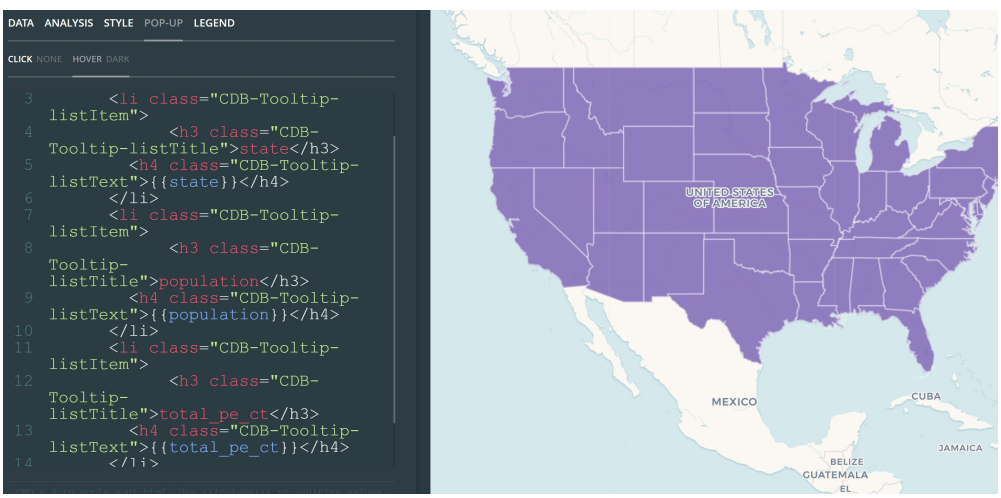


FIG. 4. An example of the software CARTO's custom HTML that allows custom altering of the map for more fine grained tuning.

color coding to more dramatic, we can change the low and high end to any color combination we want.

V. WEBSITE

Our website was built from a foundation of the language HTML 5 since it is the main stay of website building. It also contains CSS files that gives the project a custom appearance,

FIG. 5. A screen shot of the opening page of our website. Shows the dynamically changing top navigation bar and the brief abstract of out project.

as well as, Javascript files that gives the website a custom feel, for example the dynamically changing top navigation bar. Our decision was to make the website a scrolling style, so that most of the information is on one page and the user scrolling down to access it. This is very different from the multiple page sites that you need to click links to navigate. We chose this style since it is becoming the common way to create sites and we wanted the user to be as comfortable navigating it as possible. The only separate pages are the data tables that show the variables that were not visualized.

The site opens with out title and a brief abstract of why we tackled this task of crime visualization, seen in Fig. 5, then it moves down to another brief section of our team, that describes who we are and what we did to complete the project. We wanted to include pictures of us to give the site a real "ownership", that it was created by citizens that want to help make America a better place to live. After the about us page is the main section of the site. The first section is about crime analytics, containing a brief paragraph summing up the crime within the states and the interactive map, seen in Fig. 6. We wanted the site to mostly be about the maps, therefore we did not include a lot of analytics from ourselves since we did not want to bias the user, we wanted them to come to their own conclusions. The next two sections are similar to this one, containing a description of the what the map is displaying.

Only the police section has a working search bar, this search contains the titles of each extra variables that we parsed. The user can start typing in the search bar and it reduces the number of options that can be clicked. Once the option is clicked, it navigates the user to the exact table that has all the information about that variable, seen in Fig. 7 and Fig.

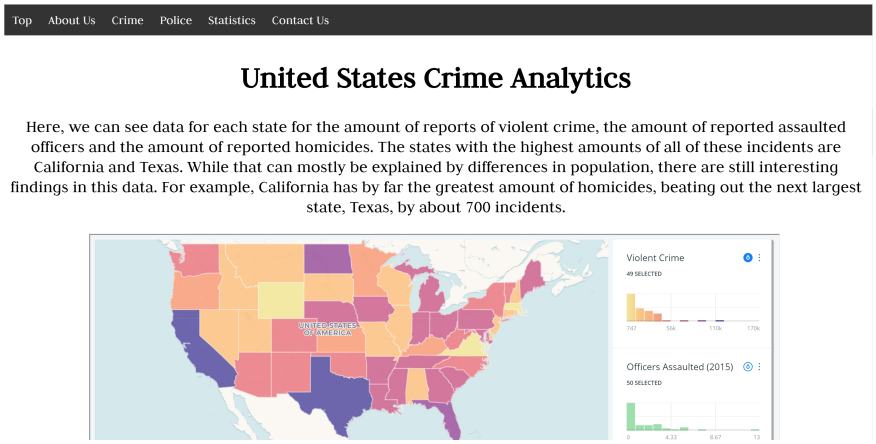


FIG. 6. A screen shot of our website that shows the embedded interactive map and our small analytics.

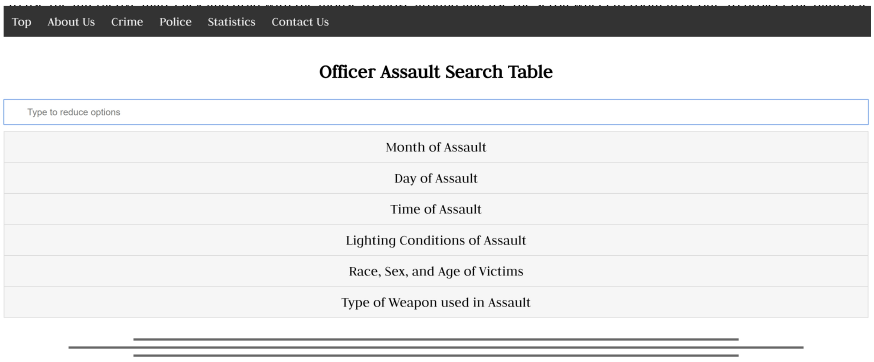


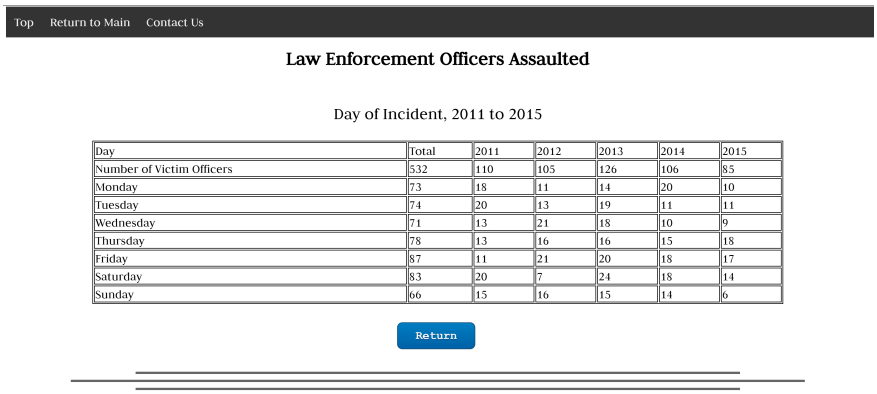
FIG. 7. A screen shot of the working search bar that reduces the variables shown as it is typed into the bar. Clicking an options navigates the user to a serperate page with the specific table.

8.

The website ends with a small licensing paragraph and a contact us, that the user can use to contact our group if they have any questions about the site, Fig. 9.

VI. FUTURE WORK

There a many different features that can be added to the this project if there was more time. The first and the most important, if to create a parsing code that can automatically import new data when it is released and parse it into the format we used in our project. This feature will allow the website to be run without the need for us to be constantly parse



Top Return to Main Contact Us

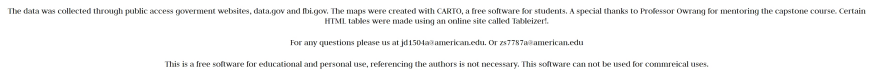
Law Enforcement Officers Assaulted

Day of Incident, 2011 to 2015

Day	Total	2011	2012	2013	2014	2015
Number of Victim Officers	532	110	105	126	106	85
Monday	73	18	11	14	20	10
Tuesday	74	20	13	19	11	11
Wednesday	71	13	21	18	10	9
Thursday	78	13	16	16	15	18
Friday	87	11	21	20	18	17
Saturday	83	20	7	24	18	14
Sunday	66	15	16	15	14	6

Return

FIG. 8. A screen shot of a chosen variable from the search bar that shows the table that was parsed from the main data set.



The data was collected through public access goverment websites, data.gov and fbi.gov. The maps were created with CARTO, a free software for students. A special thanks to Professor Owrang for mentoring the capstone course. Certain HTML tables were made using an online site called Tableizer!.

For any questions please us at jd1504a@american.edu. Or zs7787a@american.edu

This is a free software for educational and personal use, referencing the authors is not necessary. This software can not be used for commereical uses.

FIG. 9. A screen shot of the license section from our website, as well, as a contact us information.

and upload the data manually. Another feature would be to include a more detailed search bar, for example, the user inputs multiple different variables they are looking for and the site compiles a table that shows all the chosen items. This will give the user a very high detailed search that will help them in their own analytics Lastly, including data that is not just state identified but county or even city identified. This will give the project a finer grain scope that will allow the user to really see every aspect of crime and police enforcement within their region.

VII. CONCLUSION

Our goal was to create a visualization project that takes open source data from federal agencies about crime and make it into user friendly tool. We completed this task by first manually parsing the data into state identifiers and then into multiple variables that are important in showing the scope of the crime issues. The state identified data was then imported into a free software, CARTO, that allows us to create interactive maps of the data that can embedded into a website. Our website was created with HTML with brief

explanations of each section and specific tables explaining in detail the data we compiled. The website we created is a great foundation for future work that could one day be opened to the public and help law enforcement agencies and citizens across the United States.