

Data Visualization of Missing and Unidentified Persons in the United States

Bandr AlSwyan  
Ricky Severino

American University  
4400 Massachusetts Ave, NW

CSC-493-001

December 09, 2018

## Table of Contents

Abstract.....	3
Introduction.....	4
Review.....	4
Design Requirements and Project Detail.....	5
Costs.....	14
Security.....	14
Feasibility Discussion.....	15
Final Implementation.....	15
Results.....	16
Conclusion.....	23
References.....	24

## Abstract

The purpose of this capstone project is to give the general public the information and tools to identify the missing individuals in the United States as well as direct them to more insight about the data. The need for this technology arise in its simplicity as the tools that are currently available are either complex or not modern enough to provide the complete potential and use to reach different people. The data that will be integrated into the application will be retrieved from the National Institute of Justice's National Missing and Unidentified Persons System ([NamUs](#)). The application consists of an HTML, CSS, and Javascript website, and the data will be represented through Tableau, a data visualization software tool. The purpose of this report is to show the specific steps that were taken in order to complete the project. There are a number of costs that comes with the features of the project. One of the main costs is the performance of the website. Due to the size of the data, it is required by the user to have higher quality hardware that handle the project software. It should be also mentioned that this website is not intended to show where the missing people currently are but moreover what and who. This tool was designed with goal in mind that the user will use it to draw conclusions and explore more information about the missing individuals.

## 1. Introduction

The purpose of this capstone project is to give the general public the information and tools to identify the missing individuals in the United States. The need for this technology arise in its simplicity as the tools that are currently available are either complex or not modern enough to provide the complete potential and use to reach different people. The use of a simple tool that does not require prior knowledge is the goal that we want to achieve. We will apply and support our data with statistical analyses that will reflect better understanding of the data collected.

The data that will be integrated into the application will be retrieved from the National Institute of Justice's National Missing and Unidentified Persons System ([NamUs](#)) website and categorized as public information. According to NamUs' [privacy agreement](#) that is in correspondence with the Department of Justice (DOJ), and the DOJ [Privacy Act Notice](#), third party applications can use the data to help professionals reach the purpose in which this data was created and collected, and therefore meeting the demand of our project, giving us the permission to use the data.

The application consists of an HTML, CSS, and Javascript website, and the data will be represented through Tableau, a data visualization software tool. The data was converted into Json and CSV for Tableau to use in its visualization. The website provided displays missing persons represented by a point on the map. When the user clicks on that point, information regarding the individual will popup on the point. In this popup tab, the user will be able to see information regarding name, gender, ethnicity, age, etc. On the side of the map, filters will be provided so the user can make specific queries about the information displayed. Among other options, the search filter will have the option to display either points representing people. The page will also provide tabs which will show statistics about the data mentioned above. These statistics will show a number of correlations and relationships between different variables provided by the data source.

The objective of this project is to provide the general public with a better understanding of the missing people in the United States in an simple and interactive manner. The response to the needs mentioned above is to design a web tool that can be easily accessed and used through an interface that the user can draw conclusions with and from.

In this report, the following sections will be discussed. First, review of similar projects regarding the visualization of missing people and why this project is different. Then, the design requirements and project details will be outlined, as well as the costs and security that comes with it. Next, the feasibility of the project which talks about the literature search and justification for the design of the project. After that, the feasibility discussion, which includes a use case diagram (UML), will be shown that outlines the expected user experience of the projects website.

Lastly, the results of the project will be shown which includes screenshots of the completed website.

## 2. Review

The NamUs website provides a similar implementation of what this project intends to do. But it was soon realized that NamUs' map lacks the information needed for users to draw conclusions. If the user wants information regarding, age, city, county, etc, they would have to click on a link that takes them to their case file. In this project's implementation, all of this information will be displayed in one place. Additionally, users are not able to go to the person's social media account (if it exists) or look for news articles without manual searching. In this project, users are able to click links that will direct them to such information. Another issue is with NamUs' map, its look. The map does not look aesthetically pleasing for the user. Also, the website does provide grouped information or statistics.

## 3. Design Requirements and Project Details

The design of our project was built upon the need for an easy interactive tool as mentioned above. Our main objective is to meet the minimum possible requirements of any user that can benefit from the usage of our system. Moreover, in the context of the business rules, the general public is the user and they are represented as our customers. The end user will have the ability to explore the data of missing people provided by NamUs.

The data was initially collected through a Google Chrome plugin named WebSpy. The plugin allows the user to modify any HTTP POST/GET requests the website makes. In the case of this project, the POST request is the search query the user makes in the search filter. The reason for using WebSpy is its ability to change the number of records requested. This is beneficial because NamUs sets a limit on the number of people than one can search for. With WebSpy, the data was outputted as a JSON file with a certain number of records. Multiple requests were made to obtain all the data. Once all of the data was collected, there was a total of fourteen individual JSON files. In order to make parsing and data manipulation easier, all of these files had to be merged into one. This was done with a JQ, a command line json processor. The following command was used:

```
$ > jq -s '.[0].list=(.[].list|flatten)|.[0]' *.json
```

Once the JSON files were merged, it was time to explore the data to see what was being dealt with. It was discovered that the data had an integrity issue. The issues encountered were the following

- Spelling errors
- Wrong entries such as locations
- Wrong inputted fields
- Lack of continuity in regards to location names
- Duplicate location names
- Wrong usage of punctuation and symbols
- Address being inserted as county
- Countries being inserted as cities
- Counties that do not exist anymore

These issues were discovered after trying to match the missing people data with a location data source that contained a specific number of United States cities with their counties and coordinates. A number of Python scripts were used to match each person in the missing people data with coordinates corresponding to their city. If no match was found, the person and their location was displayed so they can be fixed manually. This scripted showed that 12 percent (about 1900 records) of the data was incorrect. It was also discovered that the coordinates file did not contain all of the cities that were present in the missing people data. Because of this, new cities and their coordinates had to be manually inputted in the coordinates file. There were a number of individuals in the data where their data was unknown. To retrieve their coordinates according to NamUs, a Selenium script using Python was written that retrieved their coordinates from the NamUs website and updated the data. The process of fixing the data integrity issue and inputting new coordinates took about three weeks. It was originally planned that R would be used to show statistics regarding the data, but it was soon realized that there was a steep learning curve in learning R. Given the amount of time given to complete the project, it was best to disregard the use of R. Similarly, in order to visualize the data points with a map, MapBox was going to be used, but the amount of data that the free version of MapBox can display has a limit of 10 thousand. Due to these complications, a new tool had to be found. Upon searching through the eyes of God himself, Tableau emerged from the darkness. Tableau is a software tool that allows for the visualization of data as well as the creation of statistics. Using Tableau killed two birds with one stone in the figurative sense.

To make the data more accessible for Tableau, it needed to be inputted as a CSV (Comma-Separated Values) file. To convert the missing people data that was originally a JSON file, a Python script was written to complete the task. To visualize the data in Tableau, a number of operations had to be done. For the mapping, the longitude and latitude were one column in the CSV which needed to be separated in order for Tableau to read it and reflect the points on the map. Another Python script was used to separate the combined coordinate values into their own

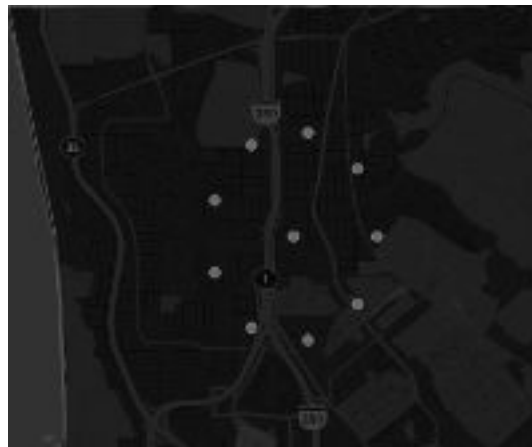
JSON object and then formatted back to CSV. For example, in the original JSON file, the coordinates were in an array:

```
"coordinates": {  
  lng: "-77.0888",  
  "lat": "38.9375"  
}
```

This array had to be transformed into the following:

```
"lng": "-77.0888",  
"lat": "38.9375"
```

Another problem was the amount of overlapping points on the map. The majority of the missing people overlapped each other because their coordinates corresponded to the same location point of a city or specific place. To fix this problem, an operation known as “jittering” had to be done to separate the overlapping points. The end result looks like the figure below.



The angle of the jittered points had to be calculated to be represented as a circle. The Tableau calculation below is what needed to be written for this to work.

Jitter Angle ×

Results are computed along Table (across).

$$360 / \text{WINDOW\_MAX}([\text{Index}])$$

$$*[\text{Index}]$$

Default Table Calculation

The calculation is valid.      4 Dependencies ▼      Apply OK

To see if there are any overlapping latitudes, the following calculation was made.

Jitter Latitude ×

```

if [Point Count] > 1 then
    ATTR([Latitude]) + [Jitter Distance] * SIN(RADIANS([Jitter An
ELSE
    ATTR([Latitude])
END

```

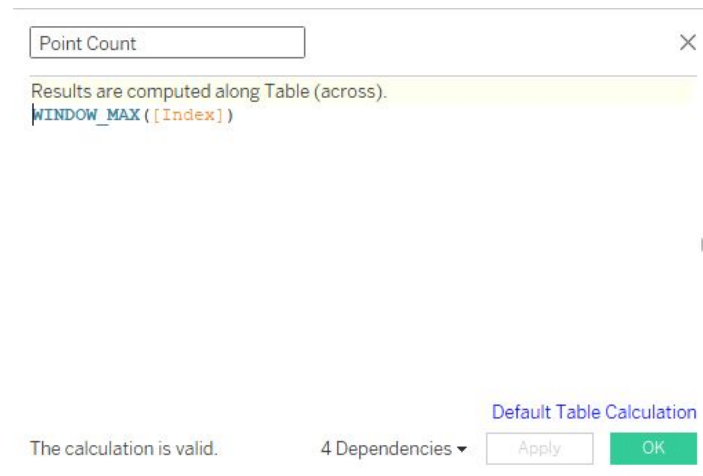
The calculation is valid.      2 Dependencies ▼      Apply OK

Same as above but with the longitude.





This calculation is used to count and group the number of points that are overlapping so it can be used to jitter the points and space them correctly.



For each point on the map, a box with details about the person are shown upon hovering of the mouse. Information such as first name, last name state, county, city, race/ethnicity, gender, missing age, and missing date are shown. When the user clicks on the point, links to social media news, accounts, and case files are provided so the user can navigate to those websites if needed. And example is provided in the image below.

When the user hovers over the point:



When the user clicks on the point:



This was possible throughout the use of the action feature that Tableau provided:

## Actions



Actions let you create interactive relationships between data, dashboard objects, other workbook sheets, and the web.

Name	Run On	Source	Fields
Case Link	Menu	MapDash (Map)	ATTR(Case Link)
Facebook News	Menu	MapDash (Map)	ATTR(First Name), ATTR>Last Name)
Facebook Post	Menu	MapDash (Map)	ATTR(First Name), ATTR>Last Name)
Facebook Profile	Menu	MapDash (Map)	ATTR(First Name), ATTR>Last Name)
Instagram Tags	Menu	MapDash (Map)	ATTR(First Name), ATTR>Last Name)
Reverse Image Search	Menu	MapDash (Map)	ATTR(Image Link)
Twitter Hashtag	Menu	MapDash (Map)	ATTR(First Name), ATTR>Last Name)
Twitter Search	Menu	MapDash (Map)	ATTR(First Name), ATTR>Last Name)

Add Action >

Edit...

Remove

☐ Show actions for all sheets in this workbook

OK

Cancel

There was a tab that required an insertion of what would the user expect and what kind of reaction he would have after he click, select, or hover over the given point:

## Edit URL Action



Name:

Source Sheets

☐ MapDash

☒ Map

Run action on:

URL

[https://twitter.com/search?q=<ATTR\(First Name\)>%20<ATTR>Last Name\)>&src=typd](https://twitter.com/search?q=<ATTR(First Name)>%20<ATTR>Last Name)>&src=typd)

☐ URL Encode Data Values

☐ Allow Multiple Values

Item Delimiter:

Delimiter Escape:

Here is the Tableau calculation that was made to generate one of the actions (Twitter Search) :

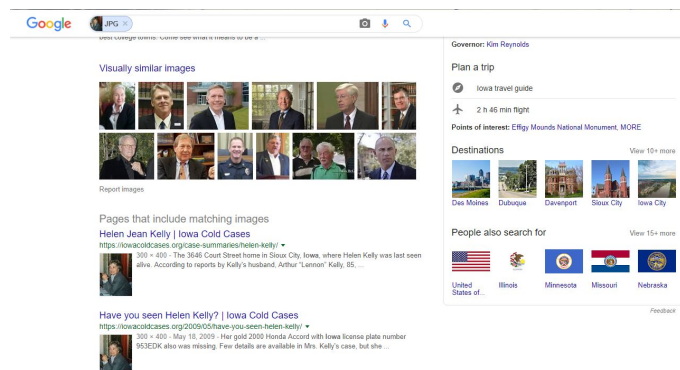
```
https://twitter.com/search?q=<ATTR(First Name)>%20<ATTR>Last
Name)>&src=typd
```

There was a need to research the reference url links for different social media platforms and other links. These links were going to be used to reflect more information about the missing person, so different python scripts had been designed prior to test it. These results were great enough that it was decided to be used as a feature.

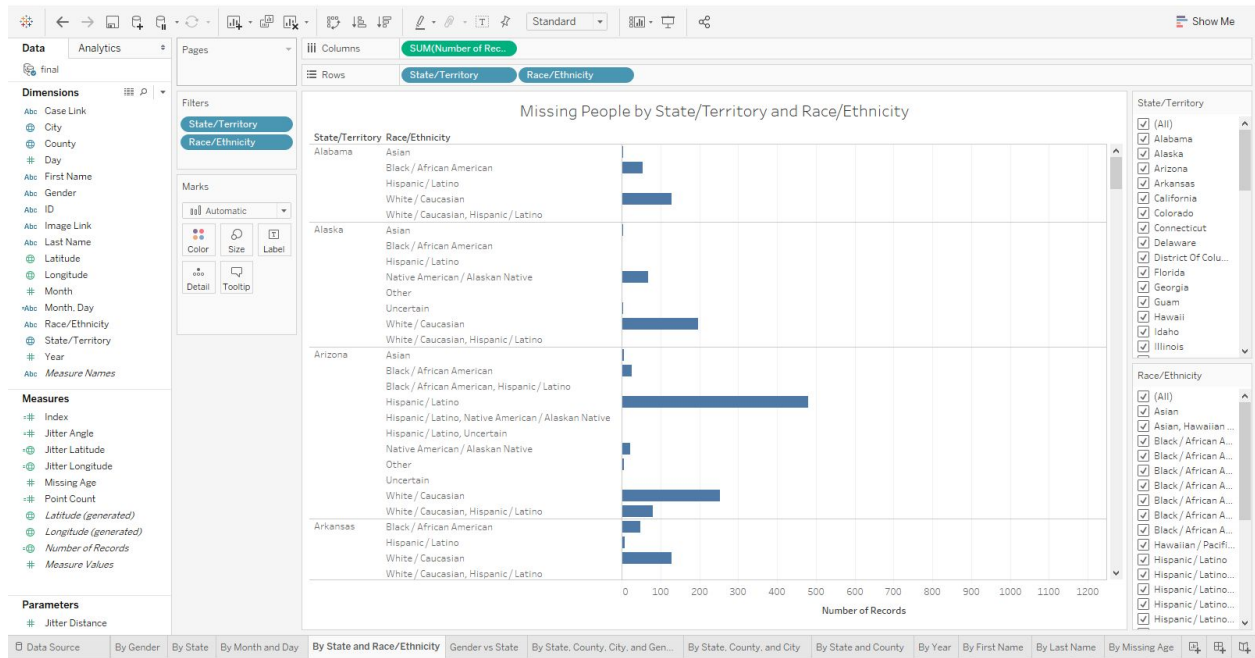
The actions (links) that are shown to the user are as follows:

- Case Link (NamUs website case file)
- Facebook News
- Facebook Post
- Facebook Profile
- Instagram Tags
- Reverse Image Search
- Twitter Hashtag
- Twitter Search

The Reverse Image Search was a unique action that required the use of the missing person's image from NamUs. This action uses Google's Image API to identify what websites have used their images and gives back a list to the end user with links and more information to explore:



It is important to mention that these links are not guaranteed to give information about the missing person. Given that thousands of people have gone missing before the creation of the Internet, they are not expected to have social media accounts or news articles about them. For the statistics portion of the project, various statistics were discussed to meet the desired information that is going to be shown to the end user. In Tableau, the author has to drag and drop the specific variables that needs to be shown:







The various statistics determined are below:

- Number of missing people by gender
- Number of missing people by state/territory
- Number of missing people by month and day
- Number of missing people by state/territory and race/ethnicity
- Gender vs state
- Number of missing people by state/territory, county, city, and gender
- Number of missing people by state/territory, county, city
- Number of missing people by state/territory, county
- Number of missing people by year
- Number of missing people by first name
- Number of missing people by last name
- Number of missing people by missing age
- Average missing age vs state/territory

- Maximum missing age vs state/territory

To make statistics regarding the month, day, and year possible, the date values in the data had to be separated, similar to how the coordinates were separated. Also, the statistics regarding the states, counties, and cities, were not normalized according to their population.

For the website, three pages were designed. The main page shows the map and the statistics, the second page, named About Us, gives information about the team and the project, the third page, the privacy page, shows information regarding the legal use of the data:

 about	12/2/2018 7:42 PM	Chrome HTML Do...	4 KB
 index	12/2/2018 7:42 PM	Chrome HTML Do...	4 KB
 privacy	12/6/2018 6:01 AM	Chrome HTML Do...	3 KB
 style	12/2/2018 7:42 PM	Cascading Style S...	2 KB

In order to show the Tableau workbooks in the main webpage, the workbooks had to be published to the Tableau server, which then gave an embedded html code that allows the website to display the work that has been done in the Tableau desktop application.

### 3.1 Costs

There are a number of costs that come with the features of the project. One of the main costs is the performance of the website. Because the size of the data is large, it is required by the user to have better hardware, specifically RAM that can handle the website. Another cost of the website is its use in locating missing people. This website is not intended to show where missing people currently are, or show how and why they went missing. That being said, the what and who is the main driver of this project. It is meant for user to draw conclusions and explore more information about the data. Also, the data is static not dynamic, meaning that it requires manual updating.

### 3.2 Security

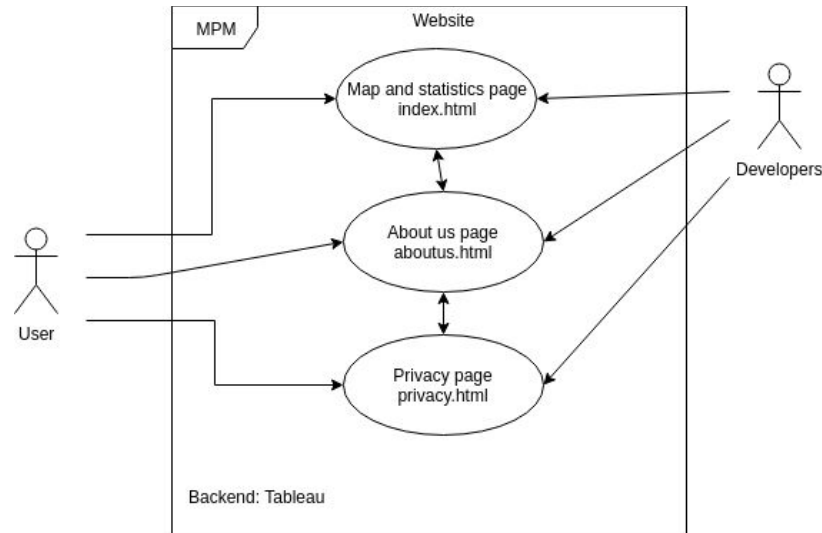
The data that is displayed to the user can not be downloaded or manipulated, meaning the data is protected. On the other hand, the website html code can be viewed with the inspect element command in the browser. This is a security flaw because any person can view the code that displays the Tableau workbook, and can take the user to the authors profile in the Tableau server.

### 3.3 Feasibility Discussion

At the start of this project, it was very challenging for the team to find an idea regarding big data visualization, as none of the members had encountered such a thing before. It came as a surprise when it was needed to find a topic to research and built the capstone around it. The challenges were mostly based on the idea that there was no unique topic to cover, since most of the big data that is on the internet had already been visualized by someone or some institution. Various brainstorming sessions have been done to decide what type of data would be a good fit for this capstone project. It was ultimately decided that missing people in the U.S would be a good idea. The NamUs website was encountered when searching for databases that contained data about missing people. When the NamUs website was viewed, it was brought to the team's attention that the website lacked a modern design and schema that provide quality information that suited the data they had. For example, the map they provided does not show enough information that would be useful for the user. Thus, it was decided that a better map would be implemented. In addition, it was observed that there was a possibility of generating more information from the data as different variables had potential. Another resource was used for performance eval which was the [National Forensic Science Technology Center, at the department of Florida International University](#). As it was mentioned above, some of the ethical questions that were encountered was the use of data; is it open to public and free to use in different environments? The answer is yes. The data from the National Institute of Justice's National Missing and Unidentified Persons System ([NamUs](#)) is classified as public information and According to them ([Privacy Agreement Page](#)) that is in terms with the Department of Justice (DOJ) and the DOJ [Privacy Act Notice](#), the data that was created and collected can be used by third parties (outsiders), and therefore granting us the permission of use the data.

### 4. Final Implementation

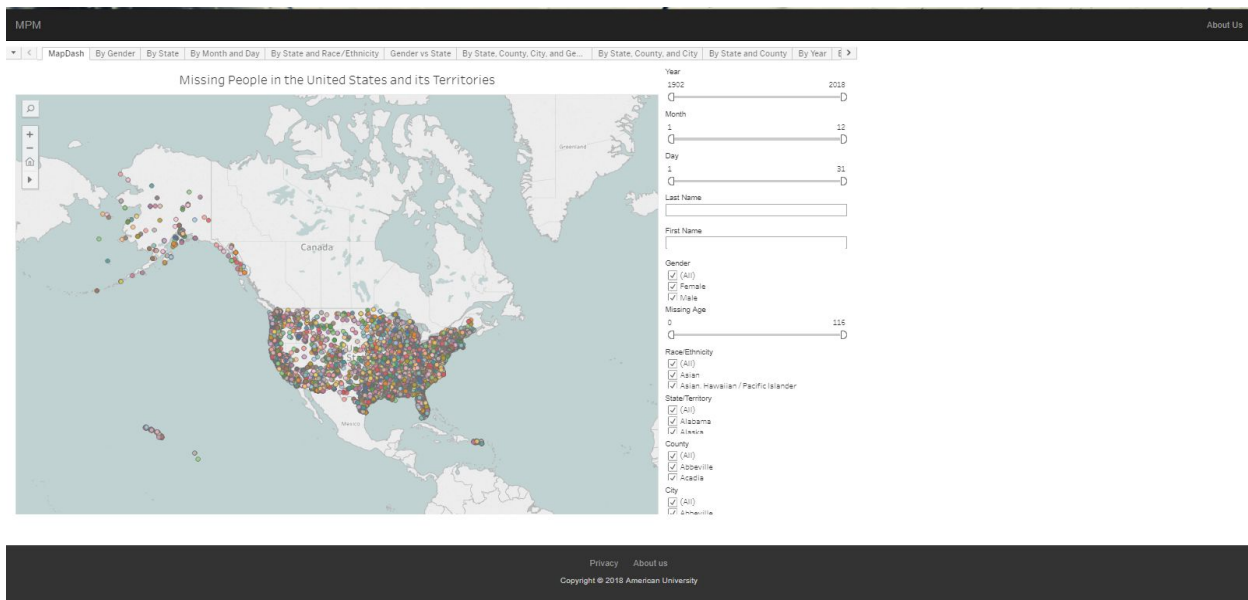
In addition to what was mentioned above, the expected Use case diagram (UML) for the end user using the website is shown below.



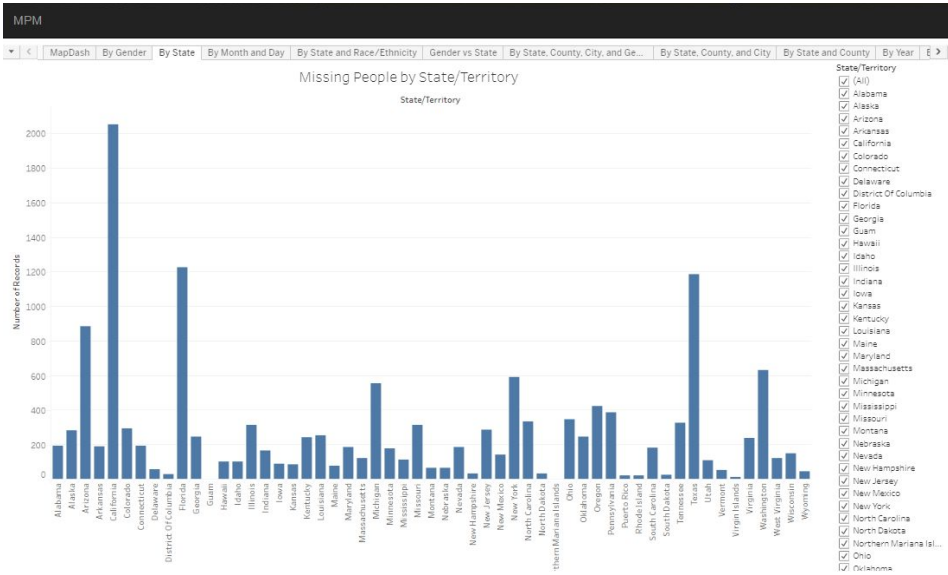
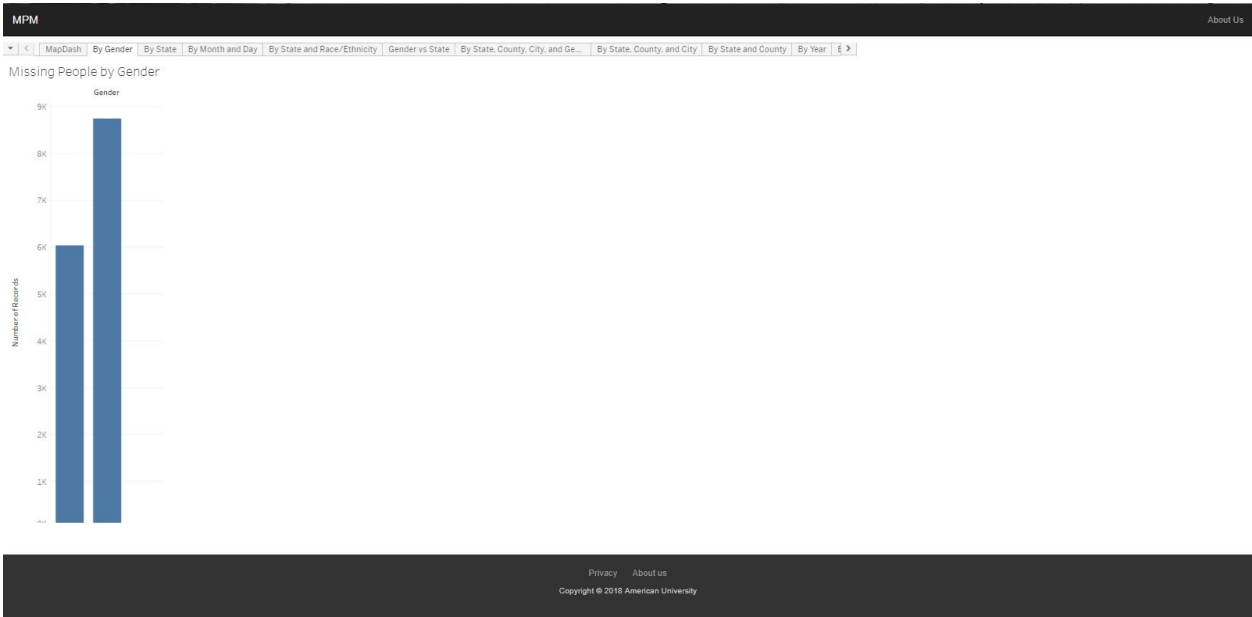
The user is expected to be able to browse all the given three pages seen in the diagram. They are expected to interact with the index page which includes the dashboard map and the statistics regarding the data. They can also view the about us page and the privacy page. In addition, the developers of the project have read and write access to these pages and its contents.

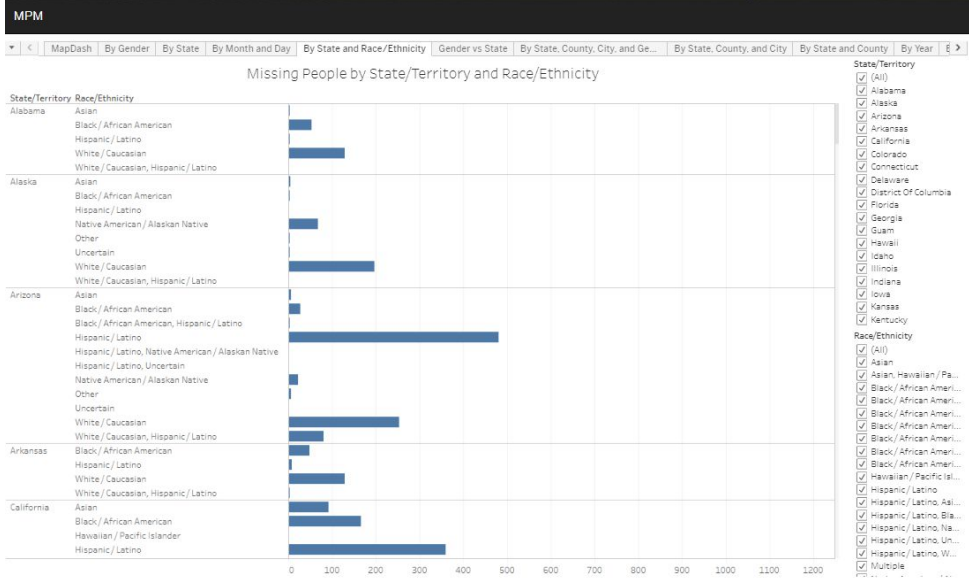
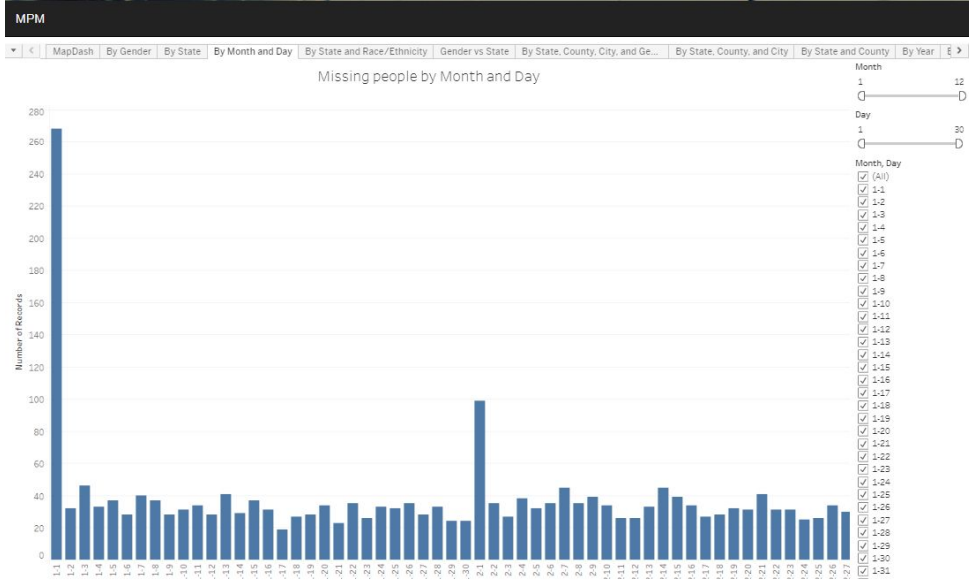
## 5. Results (Images of final work)

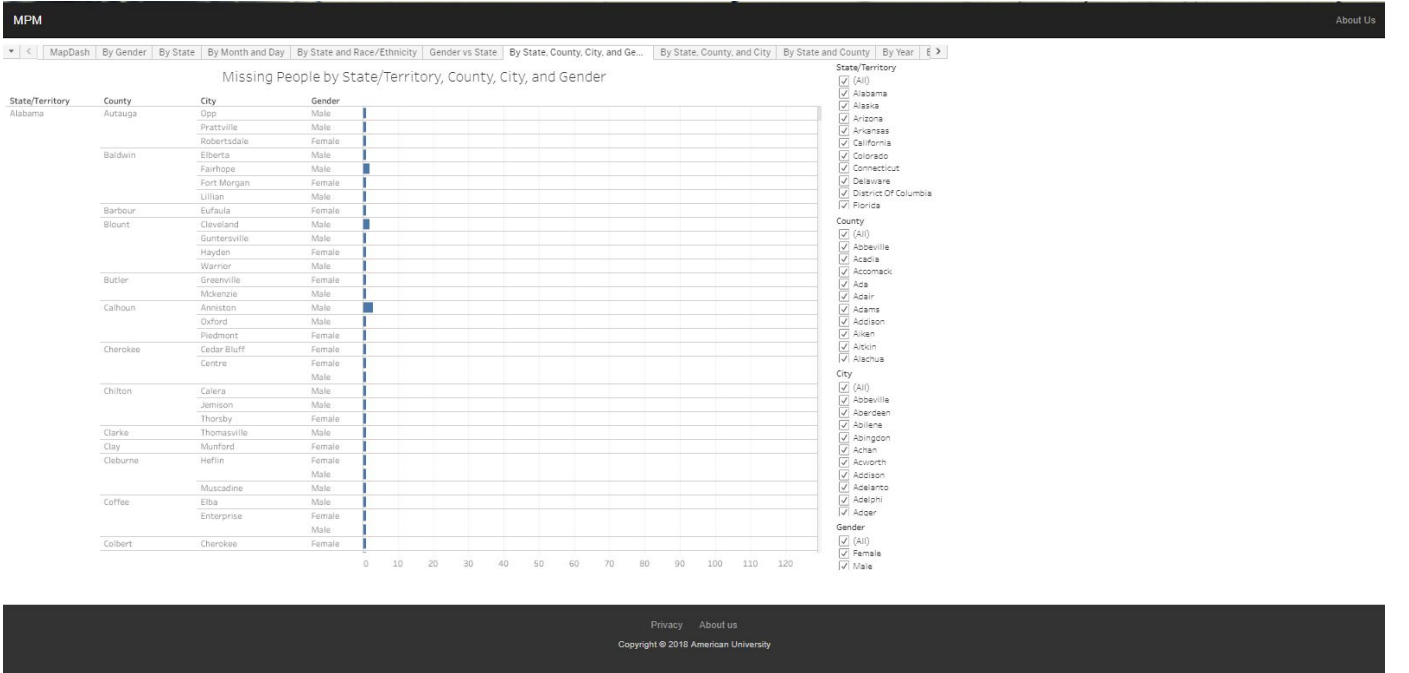
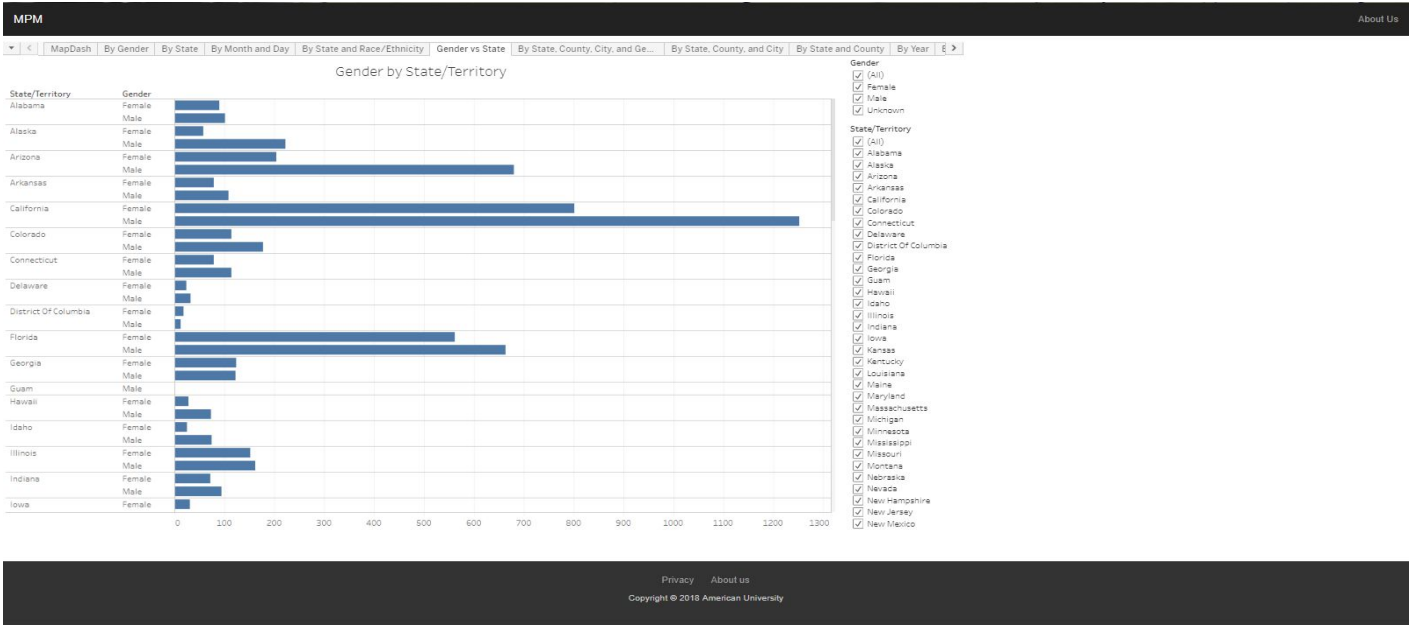
The screenshots in this section are to show the outputs of the project. The same screenshots are provided in the “OUTPUTS” folder in the zip file.

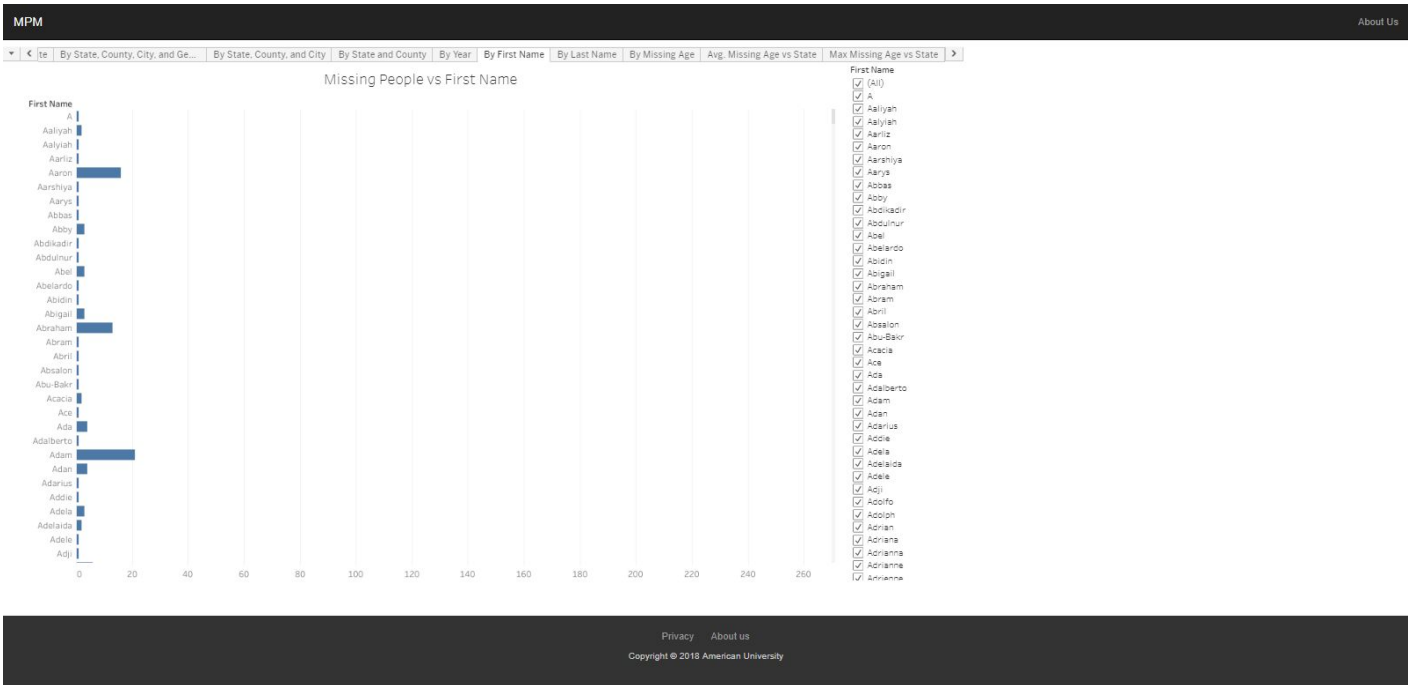
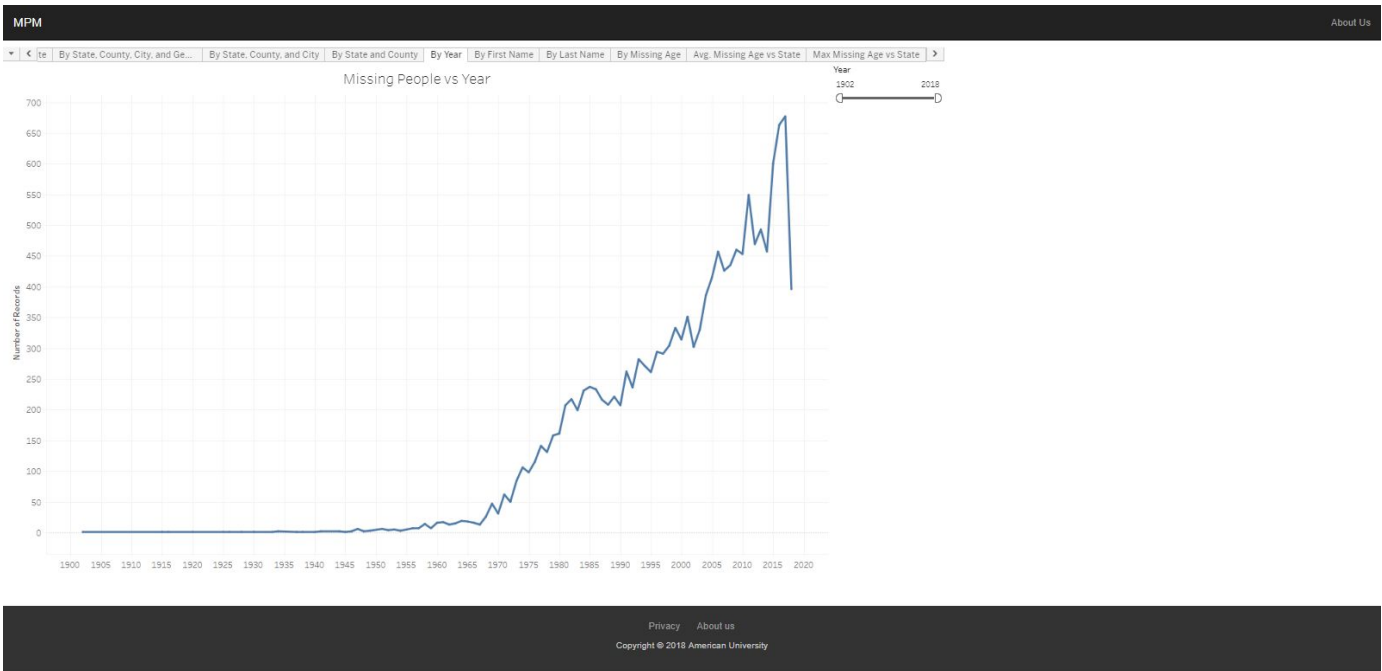


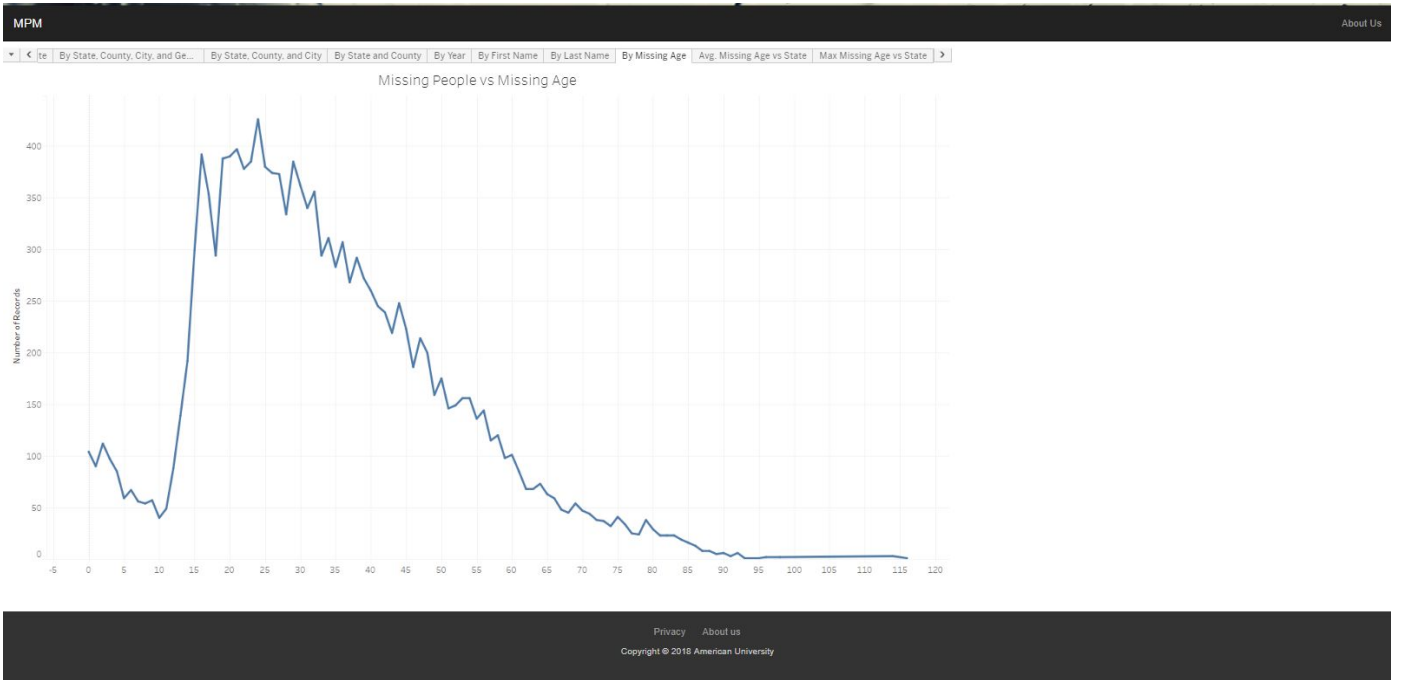
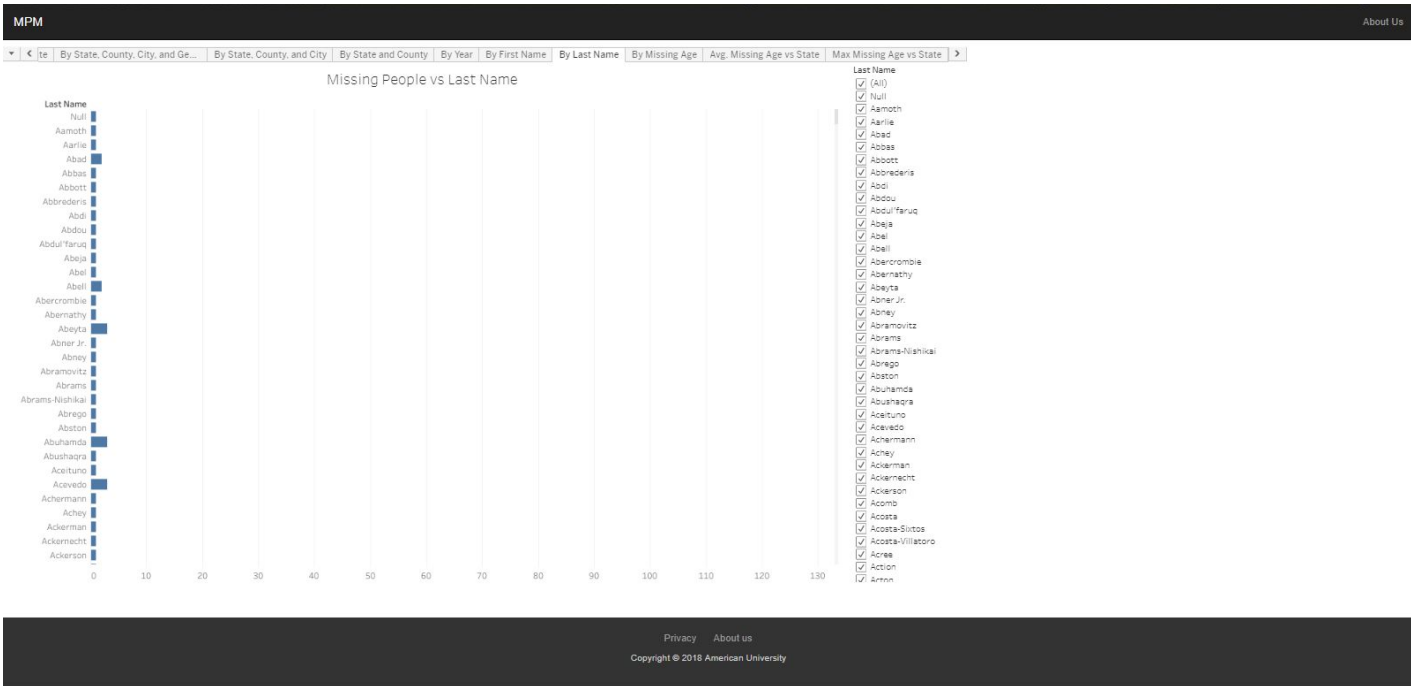


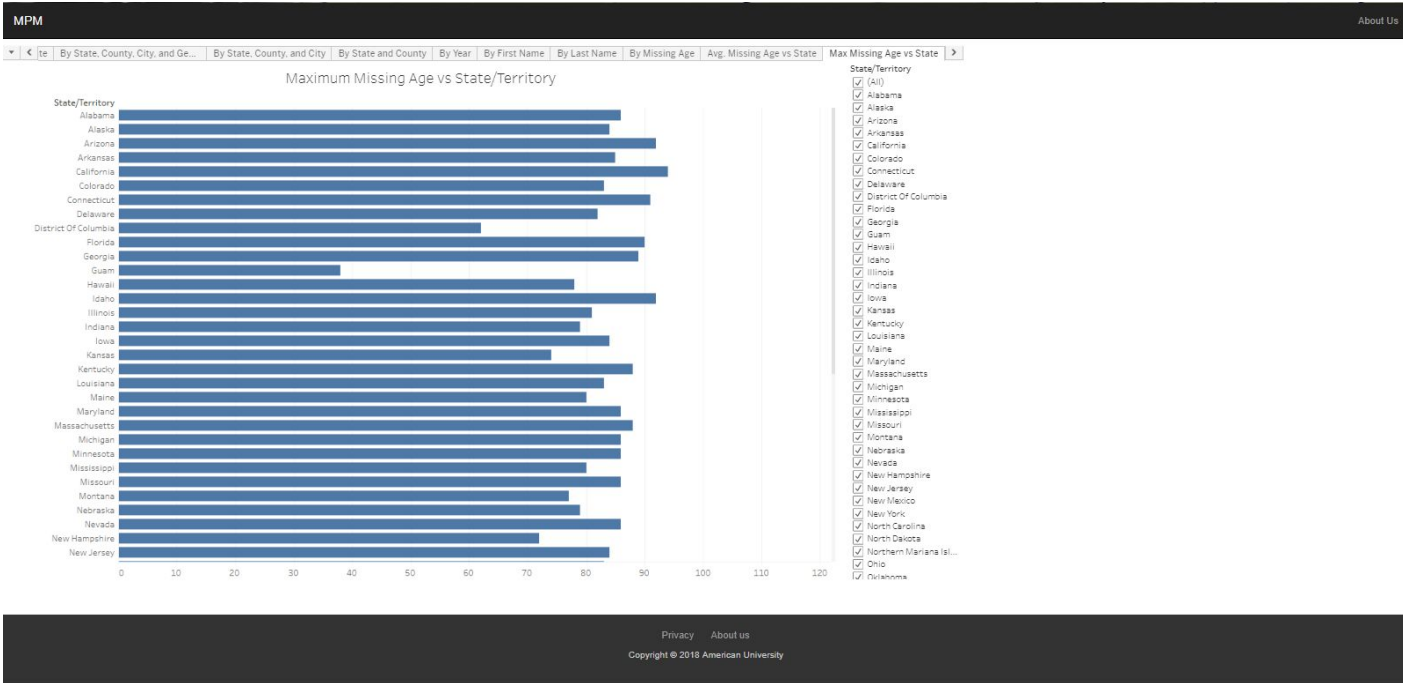
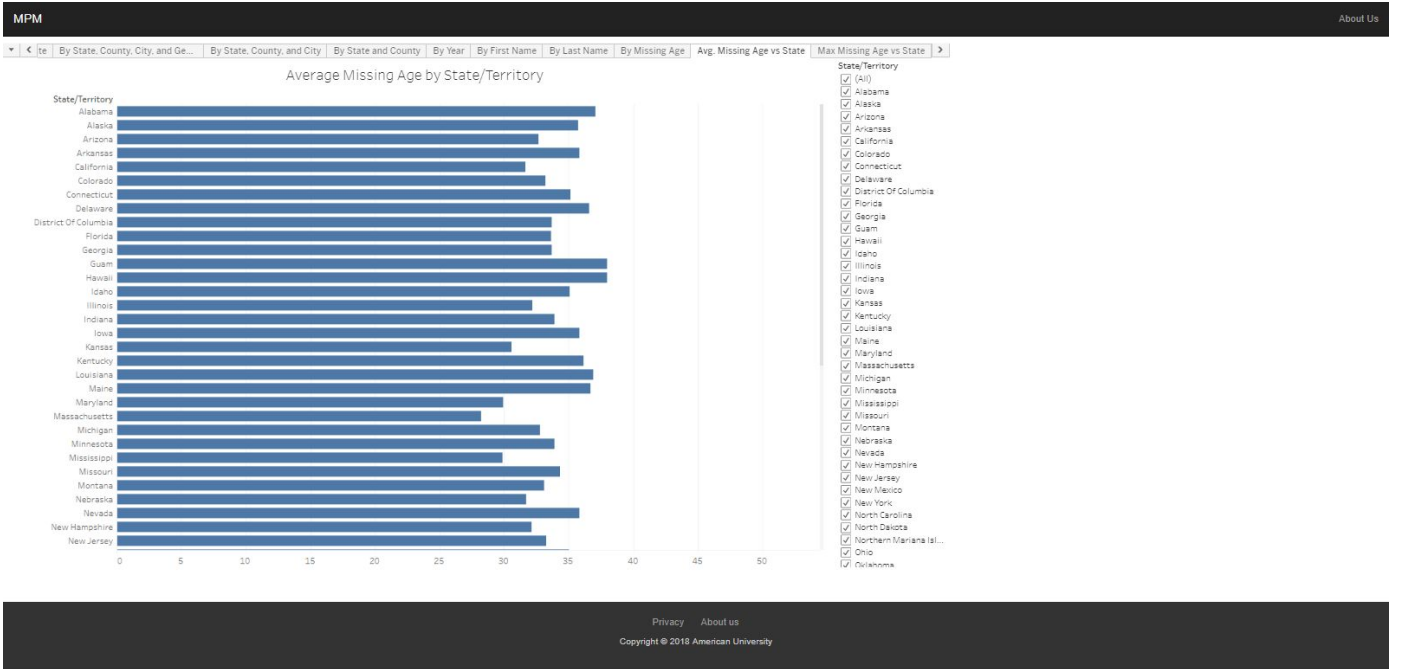








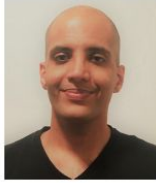





MPM

About Us

## The Team




**BAHAR ALSWYAN**  
Contact: ba6430a@student.american.edu



**RICKY SEVERINO**  
Contact: rs781a@student.american.edu

---



### Our Project

#### Data Visualization of Missing Persons in the United States

The purpose of this capstone project is to give the general public the information and tool to identify and potentially assist locating missing individuals in the United States. The need for this technology arose in its simplicity as the tools that are currently available are either complex or not modern enough to provide the complete potential and use to reach different people. The use of a simple tool that does not require prior knowledge is the goal that we want to achieve. This project was completed for American University's computer science senior capstone, CSC-493-001, Fall 2018.

[Privacy](#)
[About us](#)

Copyright © 2018 American University

MPM

About Us

## Data Use

This project was made possible through the use of the National Institute of Justice's National Missing and Unidentified Persons System ([NamUs](#)) website. According to the privacy agreement that is in correspondence with the Department of Justice (DOJ), third party applications can use the data to help professionals reach the purpose in which this data was created and collected, and therefore meeting the demand of our project, giving us the permission to use the data.

For more information, please visit the sites below

- [NamUs Privacy Act Statement](#)
- [DOJ Privacy Policy](#)
- [FOIA](#)

[Privacy](#)
[About us](#)

Copyright © 2018 American University

## 6. Conclusions

Throughout the project our main goal was to solve the problem which was the lack design that the missing persons system NamUs had, and provide more information about the data that surround the missing people records. As a team we saw potential and possibility to do more; the data was enhanced, statistic had been done and more over a constructed map visualization that carry information about each person with more guided direction to explore. The performance of project system is reliable with the right hardware; its functional and usable.

Through the course of this project, a number of lessons were learned that will be considered for future projects. An important lesson learned was, “Do not use a hammer to put in a thumbtack”, meaning, do not use complicated tools to complete simple tasks; simple tasks can be done with simple tools. Another lesson learned was to plan out your scripts before implementing them, with enough planning, writing scripts can become easy. Putting this project together was a fun experience that the whole team enjoyed.

## 7. References

- "Department Of Justice: Privacy Policy". 2018. *Justice.Gov*.  
<https://www.justice.gov/doj/privacy-policy>.
- "Namus". 2018. *Namus.Gov*. <https://www.namus.gov/>.
- "Namus". 2018. *Namus.Gov*. <https://www.namus.gov/DOJPrivacyActNotice>.
- "National Missing And Unidentified Persons System (Namus) – NFSTC@FIU". 2018.  
*Nfstc.Org*. <https://www.nfstc.org/past-performance/namus/>.