|  |  |
| --- | --- |
| **Course code and name:** | **F20DV – DATA VISUALIZATION AND ANALYTICS** |
| **Type of assessment:** | **Individual** |
| **Coursework Title:** | **Lab 4: Dataset Visualisation & Analytics** |
| **Student Name:** | **Ayushi Madhukumar Amin** |
| **Student ID Number:** | **H00331154** |

|  |
| --- |
| **Declaration of authorship. By signing this form:**   * **I declare** that the work I have submitted for individual assessment OR the work I have contributed to a group assessment, is entirely my own. I have NOT taken the ideas, writings or inventions of another person and used these as if they were my own. My submission or my contribution to a group submission is expressed in my own words. Any uses made within this work of the ideas, writings or inventions of others, or of any existing sources of information (books, journals, websites, etc.) are properly acknowledged and listed in the references and/or acknowledgements section. * I confirm that I have read, understood and followed the University’s Regulations on plagiarism as published on the [University’s website](https://www.hw.ac.uk/uk/students/studies/examinations/plagiarism.htm), and that I am aware of the penalties that I will face should I not adhere to the University Regulations. * I confirm that I have read, understood and avoided the different types of plagiarism explained in the University guidance on [Academic Integrity and Plagiarism](https://heriotwatt.sharepoint.com/sites/skillshub/SitePages/Academic-Integrity-and-Plagiarism.aspx)   **Student Signature** *(type your name): Ayushi Madhukumar Amin*  **Date**: *04/04/2022* |

Copy this page and insert it into your coursework file in front of your title page.

For group assessment each group member must sign a separate form and all forms must be included with the group submission.

**Your work will not be marked if a signed copy of this form is not included with your submission.**

Lab 4

F20DV – Data Visualization and Analytics

CW - 4

Name: Ayushi MadhuKumar Amin

Registration Number: H00331154

Program: Bsc. Computer Science (Year – 4)

Campus: Heriot-Watt University Dubai

Date: April 4, 2022

Code and results demonstrated to:

Introduction:

The image above is a dashboard designed to give users a detailed look at crime figures in the United States during the last 55 years (1965-2019).

Users are offered two options to choose from, i.e., the two major crime kinds – Property and Violent, and the users can select the one they want to investigate further. Each of the two major crime categories comprises of subcategories, which are displayed in the table below.

|  |  |
| --- | --- |
| **Property** | **Violent** |
| Burglary | Assault |
| Larceny | Murder |
| Motor | Rape |
| - | Robbery |

On choosing an option, users are provided with multiple graphs that show the correlation between the average population and the subcategories (eg: Burglary, Murder, etc.) for a particular year. In addition, the dashboard displays a ranking of states based on how high the rate of the chosen crime (Property or Violent) is in that state. A donut chart depicting the top crime subcategories for a particular year is also shown.

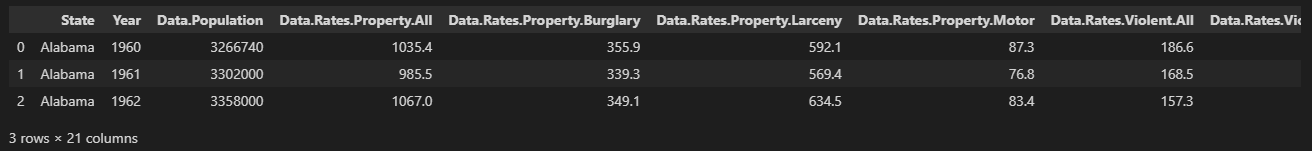
The dashboard is composed of the following layouts:

1. Choropleth Map of the United States
2. Donut chart – Shows the top crime subcategories for a particular year
3. Circular Bar chart – Displays the ranking of the states, ordered by the rate of the chosen crime category (Property or Violent)
4. Scatter plot – Shows the correlation between a crime subcategory and the overall population
5. Area Chart – Displays the evolution of crime subcategories over the course of 55 years for a selected country

Data:

For this dashboard, the main data has been taken from the following page (*The CORGIS Dataset Project*): <https://corgis-edu.github.io/corgis/csv/state_crime/>

The dataset is open source (Bart et al., 2017) and contains data on crime rates and totals for all 50 states and the federal district - Washington D.C. (District of Columbia). The dataset ranges from 1960 to 2019 and its dimensions are – 3116 rows and 21 columns.

The dataset was processed with Python and Pandas before being used for the dashboard, and the image below demonstrates how the original dataset looks when loaded into a DataFrame.

A screenshot of a computer

Description automatically generated with medium confidenceBecause certain datasets contain duplicate data, the Python program uses the duplicated() Pandas function to see if there are any duplicates in the DataFrame. Due to the truncation of the result, a for loop was constructed to print which rows contained duplicate data. The for loop did not return any rows, hence there were no duplicate values present in the dataset.

To increase readability and for easier access in the JavaScript programs, the column names of the DataFrame were modified using Pandas rename() function. The code snippet on the right demonstrates how and to what the columns were renamed.

Aside from 50 states and the federal district (Washington, D.C.), the dataset featured a few rows that displayed the overall data collected across the country. Since the dashboard only deals with data collected per state and district, the rows having the column “state” value as ‘United States” were dropped. Additionally, the state “New York” only has data from 1965 to 2019, hence the Python program only extracts the data collected from 1965 onwards. The resulting subset was stored in a DataFrame, which was then converted to a CSV file. This CSV file served as the primary dataset for the dashboard.

Graphical user interface, text, application, chat or text message

Description automatically generatedGraphical user interface, application

Description automatically generatedDashboard Header:

Text

Description automatically generated

Text

Description automatically generatedThe dashboard uses HTML’s range slider to present the crime rate scene in the United States from 1965 to 2019. The slider’s minimum and maximum values are 1965 and 2019, respectively, and they’re set using D3 and the ID allocated to the <input> tag.

By sliding the purple marker over the HTML slider, users can see crime rate statistics for a specified year. The selected year will be displayed on the left side of the slider and will be saved in a local variable that will be used as an argument to multiple local and global functions present in various JavaScript files. These functions are used for generating the multiple layouts that appear on the dashboard.

After initially picking a crime category, the slider by default selects the year 1965, and the dashboard displays crime statistics for that year. As discussed above, 1965 is stored in a local variable which is sent as an argument to multiple functions and the layouts are generated accordingly. The code snippet on the right shows an example of how the selected year (taken from the slider) is stored in a local variable named “currentYear” and how it serves as an argument for various functions.

Text

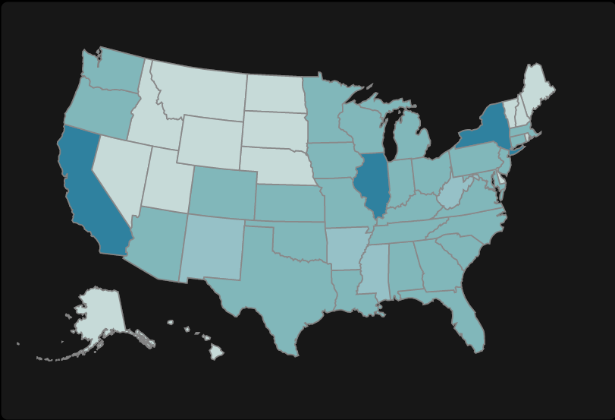
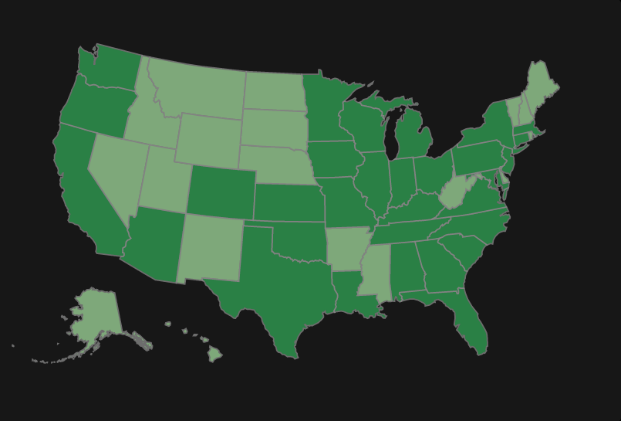
Description automatically generatedAdditionally, the slider uses the “oninput” attribute to listen for input events, which are subsequently stored in the local variable “currentYear”. This is done to modify the dashboard layouts so that the statistics are displayed precisely for the selected year.

Text

Description automatically generatedAside from the slider, the user can select a crime type by pressing one of the two radio buttons - Property or Violent. The “onclick” attribute is used by the local JavaScript function “radioClick()” to listen for click events. The “radioClick()” function sends the selected option to a global function whenever the user picks a radio button. This function was created to generate specific layouts for the selected crime category.

Layouts:

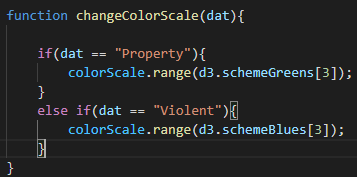
**Layout 1: *Choropleth Map of the United States***

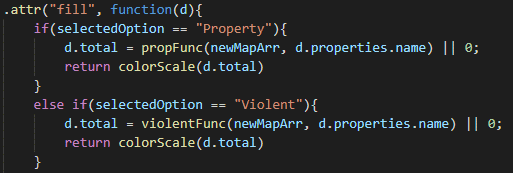
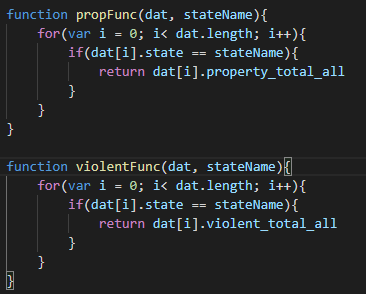


**(a)**

**(b)**

The above two images illustrate how the choropleth map appears (for the year 1965) when the (a) Property crime category is selected and (b) Violent crime category is chosen.

For generating the basic map structure, the program makes use of a geoJSON file that is publicly available on the following GitHub repository: <https://github.com/PublicaMundi/MappingAPI/blob/master/data/geojson/us-states.json> . This JSON file contains coordinates of each US state along with their respective names.

For building the projection of the map, the program exploits D3’s geoAlbersUsa() projection. Initially, the geoMercator() projection was used for the map. However, due to the distance between the main body of the US and Alaska and Hawaii, the geoMercator() projection was not ideal as a large amount of space on the SVG object was wasted and the map appeared diminished, even after adjusting the translate(). The geoAlbersUsa() projection positions the map of Alaska and Hawaii below the main body of the US. This ensures a large amount of space isn’t wasted on the SVG object. For coloring the map, the defined color scale makes use of D3’s scaleThreshold(), where the domain ranges from 1000 to 32000. The range was dynamically set based on what crime category was chosen by the user. As seen from the above two images of the US map, the choropleth map representing the Property and Violent crime category appear in shades of green and blue, respectively. If the user selects the Property crime category, D3’s sequential green color scheme, d3.schemeGreen() is serves as the range for the color scale. The d3.schemeBlues() sequential blue color scheme is set as the range of the color scale, provided the user chooses the Violent crime category. Each state was colored accordingly using the “fill” attribute and an Object array. Values recorded on the selected year (taken from the slider) were extracted using the “filter()” function, that generated a new object array.

References:

1. Bart, A.C., Whitcomb, R., Kafura, D., Shaffer, C.A. and Tilevich, E. (2017). Computing with CORGIS. *ACM Inroads*, 8(2), pp.66–72.