Project 2 - Final Report

Xiaofan(Cade）Lu – a1780911 2020 Semester 2

Introduction

Before project1, I investigated many datasets online and my goal was to find a dataset from a proper source. Because I believe the reliability of my dataset/source will have a direct impact on my research process and results. And under my curiosity for guns and the US, I picked the dataset

"US fatal police shootings" as my first attempt in project1 since this dataset is from a proper source which I think. The process and result from project1 were not satisfying to me enough and with my dataset, I can make good dynamic visualisations which becomes my motivation to continue my work on this dataset in project 2. Even though some main results will keep the same with my project1, dynamic and interactive visualisations give my topic a better visual appearance and I believe it will be better for the audience to understand and stay with what I want to demonstrate. As this is a life-related topic, it gives me extra motivation for researching and learning new knowledge for the exploration process.

Data

The Dataset[1] from The Washington Post contains records of every fatal shooting in the United States by a police officer in the line of duty since Jan. 1, 2015. In 2015, The Post began tracking more than a dozen details about each killing — including the race of the deceased, the circumstances of the shooting, whether the person was armed and whether the person was experiencing a mental health crisis — by culling local news reports, law enforcement websites and social media, and by monitoring independent databases such as Killed by Police and Fatal Encounters. The Post conducted additional reporting in many cases.
The Post is documenting only those shootings in which a police officer, in the line of duty, shoots and kills a civilian — the circumstances that most closely parallel the 2014 killing of Michael Brown in Ferguson, Mo., which began the protest movement culminating in Black Lives Matter and an increased focus on police accountability nationwide. The Post is not tracking deaths of people in police custody, fatal shootings by off-duty officers or non-shooting deaths.
The FBI and the Centers for Disease Control and Prevention log fatal shootings by police, but officials acknowledge that their data is incomplete. Since 2015, The Post has documented more than twice as many fatal shootings by police as recorded on average annually.
The Post's database is updated regularly as fatal shootings are reported and as facts emerge about individual cases.

The number of rows in the dataset has increased from 4895(when I made it in my project1) to 5624(at the beginning of my project2). And the number today(27 Oct 2020) is 5716. I used the same dataset(which may be different from versions of the dataset) to the official website that provided the dataset. And their investigation results and methodologies can be found on

the website[2]. There is related attributes/information for each fatal shooting case in the dataset including State, Gender, Race, Age, Mental illness, Weapon, Body camera, Fleeting the scene(car/on foot etc.), Date, and the Coordinates of the death. This gives a huge space to explore and there should certainly have more potential in the dataset. There are also some NaN values in the dataset but I cleaned it up using related python3 functions easily. And we should be able to modify the dataset and expand it for different variants easily.

Goals & Methods

My goals for project1 was mainly on a tentative attitude and I addressed some results on racism questions and age ranges. For project2, I am mainly focusing on generating dynamic and interactive visualisations rather than certain conclusions and results. That's because my project 1 has done some static analysis on the dataset and have reached some results from it. A basic goal of my project 2 is how to use those new libraries of python(plotly, folium, etc.) with my dataset. I started by reading the 2 python tutorials given by our school(the University of Adelaide) and looked at those graphs from the tutorials. After that, I found out there is coordinates attribute within my dataset so I started to think about how to make it on a map and make the map interactive simultaneously. But the tools/graphs in the tutorials do not meet my expectation much so I started to do research online which seems to be my first challenge in this project. On youtube, I found a useful video[3] about a new library Folium out of our class which is suitable for my project. I watched the video and learned from the code given with the video on youtube. But the next challenge is dataset cleaning. My dataset was with some NaN values in it and rejected to be added to the map. I searched how to clean up NaN values online and cleaned the dataset by myself. The process is smooth and I showed my first interactive map to one of my tutors happily.

Another base goal I have achieved and should be successful is that I can demonstrate my dynamic and interactive visualisations live. I looked at the map visualisations in the given tutorials and found them not that perfect for my presentation. So I learned folium and used it to generate my interactive map. This gives me better performance for my demonstration as Folium map support multiple functionalities to the user live. And it is established on notebook and I can also save it and modify it easily or send to another person to use it or share it online. Another basic goal is to hand on things that meet the expectation of the school. My project 1 was a mess in some components so I asked the tutor who marked them for tips. I also made an appointment with the lecturer and went to his office for discussing my problems for writing reports and how to improve. I gained sufficient feedback and tips in many ways and after my studying and reading, I improved from them. Timing skill is also a basic goal for my project 2 as I rarely did successful presentations in English before. The challenge behind it is my lack of communication skills and time management skills for presentations. I tried to be more relaxed on my presentations and as my experience gets more, my ability to control the presentation went better. This is mainly caused by my confidence on the presentation day and also my persistent effort on doing a better presentation. I asked my classmates about their timing experience on their presentations and I went to a tutor for asking my presentation performance related to timing.

One of my stretch goals is to make all the written work a significant improvement. Because I did my written work badly in project 1 due to lack of experience and some oversea challenges.
My other stretch goals are to make my dynamic visualisations better and better as well as create more interesting and clear visualisations by using python skills and libraries. One

specific stretch goal is to make my interactive map more clear to the audience, as my interactive map is not very clear to the audience if without my explanation. The challenge here is how can I change and update small elements in my map so that I can make it more suitable to my topic and thus more clear. As there are markers on my map which represents the death cases by coordinates, I found two icon databases online “Glyphicons” and “Fontawesome”[4] for my markers on my map which both are very popular one. I learned how to interact with them for my project and addressed them in my project successfully.

Results & Conclusion

Since my project 1 has reached results in regards to the racism hypothesis and age ranges. My project 2 has concluded with results for interactivity and visual appealing perspective. My purpose and as well as result are more about providing a tool generated by notebook and different python3 libraries for others to interact and gain information by themselves. I have uploaded and published my result, these interactive visualisations on Github for everyone to use as well as a readme file including some brief description on this project.[5]

There are 3 interactive and dynamic visualisations for people to use in my project2 as well as provided on my repository on Github.
An interactive and dynamic map(generated by Folium) with zoom in and out functionality as well as markers by coordinates and popups with information of each case. And I used different colour markers for different race groups.
An interactive and dynamic pie chart(generated by plotly) with 4 versions of it showing different information distribution of dead people including manner of death, mental illness, gender, armed weapons.
An interactive and dynamic bar chart to show the number of cases by states and ages intuitively.

Evaluation & Reflection

The three universal basic goals (dynamic and interactive visualisations, dataset, ability to demonstrate it live) are achieved well in my project 2 as I provided 3 interactive visualisations with proper processing and gathering of my dataset. They are also dynamic. I also uploaded them on Github for sharing with people online and to make a record for myself. Other basic goals of this project(understanding the purpose of each component in this class, how to write a proper report, how to meet the expectation for my work in project 2) are also achieved well I think. Since the lecturer and tutors gave me lots of tips and guide on my work specifically and the general purpose of this class. I addressed as much feedback as possible in a short time to make much improvement to many aspects of my work. I learned from my classmates’ ways on how to write a report as well as all the feedback from classmates and the staffs in our school. I found myself improve very much on these writings which I believe will help my future life in the western world too. One thing I want to mention here is the importance of the feedback. This project truly changed my view on the importance of asking for feedback from as many people as I can find to help my work. It is very effective for improvement if one can find the right people for feedback.
Some of my stretch goals(write comprehensive reports, learn and use new python libraries better) are achieved well at least in the extend of my satisfaction.

I felt good about the stretch goal related to comprehensive reports because I put a lot of effort into it and it was particularly challenging for me. I just allocated many hours of time to this goal and told myself to pay more attention to it. I read the feedback from the tutor carefully, I read the guide from the school carefully. I asked my classmates who did well on it for tips and addressed those tips to my work. I used an online spelling checking tool to check my spelling and grammar mistakes in the reports and adjusted them after.

Also, I achieved the stretch goal of creating a dropdown menu for my different versions of pie charts to work properly very well, the process was challenging but I finally overcame it and successfully made it with personal research online and effort of thinking.

But there are still some stretch goals not achieved well(make a heatmap using the online guide, add a timeline to my first map) due to time limit. Another reason for this is I designed many stretch goals(more than realistic) to make myself to avoid a situation that I do not have enough challenges to work on and challenge myself if I have much time.

I will say my reports in project 2 are comprehensive and correct in styles. My method of making them is to ask help from local good students and tutors for their experience on how to write a good report. There exist barriers between me(an international student) and local students because my previous background and lack of communication skills but I tried my best to overcome and made much success on friendship with local good students. When I gained their help on my reports, things went much easier and gave myself much confidence on different sides of my study. This applies to my life abroad here too. Learning and use new python library is also a challenging part of project2. The main reason is the materials given by school is very limited and we need to explore online by ourselves to find proper tools to build our ideal outcome. I view this as a contribution to my future paths in computer science. I must overcome and learn how to self-learn staffs as a computer science student. Despite this, an interest in Machine learning also gives me extra motivation to overcome these difficulties.

References

[1] GitHub. 2020. *Washingtonpost/Data-Police-Shootings*. [online] Available at: <https://github.com/washingtonpost/data-police-shootings/blob/master/fatal-police-shootings-data.csv> [Accessed 27 October 2020].

[2] Washingtonpost. 2020. *Fatal Force*. [online] Available at: <https://www.washingtonpost.com/graphics/investigations/police-shootings-database/> [Accessed 27 October 2020].

[3] 2020. *Python Maps With Folium*. [video] Available at: <https://www.youtube.com/watch?v=t9Ed5QyO7qY> [Accessed 27 October 2020].

[4] Fontawesome.com. 2020. *Icons*. [online] Available at: <https://fontawesome.com/icons?d=gallery> [Accessed 28 October 2020].

[5] GitHub. 2020. *Uspolicefatalkill*. [online] Available at: <https://github.com/cade1o/USPolicefatalkill> [Accessed 28 October 2020].