USA Homicide Reports, 1980-2014

Interactive Media Design and Visualisation CA 1

Mariam Omar (x00145301)

The data set that I chose to analyse is the Homicide Reports data set from the year 1980 to 2014. This data set is of Homicides that occurred in the United States of America. I will be examining the following variables, race, sex, weapons used, the relationship between victim and perpetrator and the top three states with the highest incident rates. Using Tableau I distributed data and created tables, charts, and plots to represent them. My main objective are to create charts and graphs that could help detect if there are any possible patterns in homicide activity in the United States and to see if there are ways that might help reduce these numbers. I hope to discover a fairly distinct pattern in homicide activity.

The topic of homicides and murders have both become interesting to people in recent years in comparison to the years before, but it is unfortunately a topic that is still under researched. Seeing as I find the topic quite interesting from the psychological perspective, I decided that it would be interesting to collect, analyse and visualize data in this area. There are several visualisations done for this data set in Kaggle and some on the FBI website, but from my research in this area there doesn't seem to be many visualizations or analysis done and there are very few datasets available, the dataset on the FBI website seems to be the only one I cold find that was up to date but unfortunately it wasn't one that I could easily access, so the one that I acquired from Kaggle was the best dataset I could find.

The dataset that I acquired is "Homicide Reports, 1980-2014" from Kaggle and has an 8.5 usability rating, It was unfortunately the only usable dataset I could find that had a significant amount of data that could be used for analysis.

The first stage of the Seven Stages is Acquire, the csv file dataset that I acquired was from the website Kaggle, which states that the data was compiled by the Murder Accountability project that was originally founded by Thomas Hargrove. The data record source is the FBI as is seen in the dataset. This specific dataset was chosen for several reasons, firstly it had a high usability rating of 8.5 secondly it contained a fair amount of information and finally because it was the most up to date dataset that I could acquire.

The second stage is parsing the data, for this stage there was not much to be done as the data was quite structured. The data was already ordered in categories for

information regarding the victim, the perpetrator and information that was quite general across the board.

The third stage is filtering out the data, this step is mainly removing data that is not considered important for the purpose of this analysis and visualisation. The great thing about this dataset is that it had so much information, but unfortunately a great amount of it is not relevant for the purpose of this visualisation and analysis. There are many columns, the columns that I removed where Agency Type, Agency code, Agency name, Record Source, Victim ethnicity, Month and City. The reason I removed the agency information is because it is simply unnecessary and we wouldn't learn anything from them, for the Record source all results where the same (FBI), the Victim ethnicity needed to be removed because majority of the results were unknown which rendered the column completely useless to this visualisation, month and city were too specific for this analysis and would be more useful if I were doing an analysis on a specific state. Now the column remaining can be seen in figure 1.

Record ID	State	Year	Month	Incident	Crime Typ	Crime Sol	Victim Se	Victim Ag	Victim Ra	Perpetrat	Perpetrat	Perpetrat	Relationsl	Weapon	Victim Co	Perpetrator Count
1	Alaska	1980	January	1	Murder or	Yes	Male	14	Native An	Male	15	Native An	Acquainta	Blunt Obj	0	0
2	Alaska	1980	March	1	Murder or	Yes	Male	43	White	Male	42	White	Acquainta	Strangulat	0	0
3	Alaska	1980	March	2	Murder or	No	Female	30	Native An	Unknown	0	Unknown	Unknown	Unknown	0	0
4	Alaska	1980	April	1	Murder or	Yes	Male	43	White	Male	42	White	Acquainta	Strangulat	0	0

Figure 1

The forth stage is mining the data, this requires using formulae to get very specific data that could be used to get values that can be used to analyse and visualise the data. For my analysis and visualisation I created a single calculated field (figure 2) that summed the total perpetrator and victim count of each relationship type, this helped give the data more meaning, because now it is more evident to the audience which of the relationships is most common. (figure 3)

Figure 2

Relationships Based on number of Incidents

Relationship	Perpetrator Count	Victim Count	Total Relationship
Unknown	41,338	31,153	72,491
Stranger	37,376	13,626	51,002
Acquaintance	29,533	13,124	42,657
Friend	3,185	1,586	4,771
Son	1,054	3,229	4,283

Figure 3

The fifth stage is represent, I decide on the type of visualisations I wanted to use to represent my data. For example I decide to go with a map for the number of incidents per state this best represented the data and really makes it simple for the audience to comprehend. The race, sex of Victims and their perpetrators and the number of incidents solved by type I decided to go with horizontal bars, this made sense because I was comparing them to one another and it was the easiest chart to read without confusion. For relationships text tables seemed to be the only one that didn't make the data seem too cluttered, and with the addition of the total it gives the audience all they need to know about the types off relationships between the victims and their perpetrators. For weapons a simple horizontal bar chart with darkening colours that represent the increase in quantity, it also helps the audience better focus on the most commonly used weapon. For the chart that represented the top three states with the highest incident rate I used a lines graph, this clearly compared the fluctuation of incidents in the top three states.

The sixth stage is refine, for this stage I created the map that clearly portrayed the number incidents that occurred in each state. This interactive map lets the audience move around and hover over states to get more information about the number of incidents, the perpetrator and victim count.

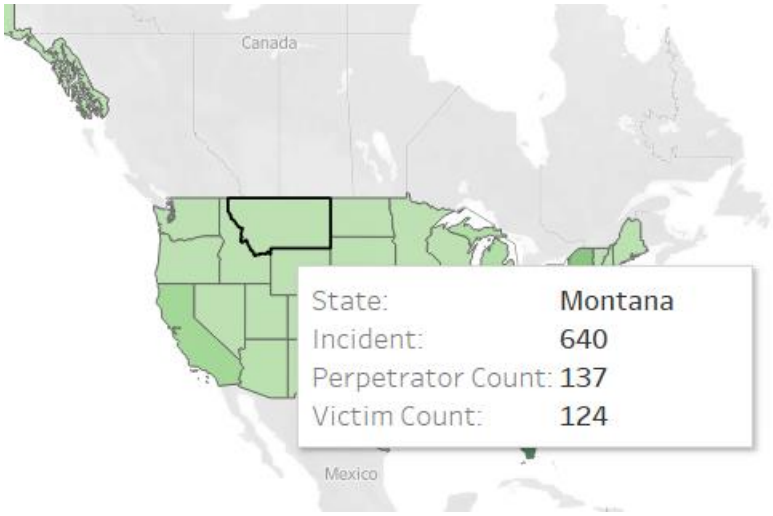

Figure 4

And finally the seventh stage is interact this stage is all about the dashboard. For the dashboard I made sure to give the audience as much information as possible without making the dashboard look cluttered. I did this by putting the charts I felt were more important on the front page while the other charts were put in as links in the menu that redirects the user to a separate page with the chart.

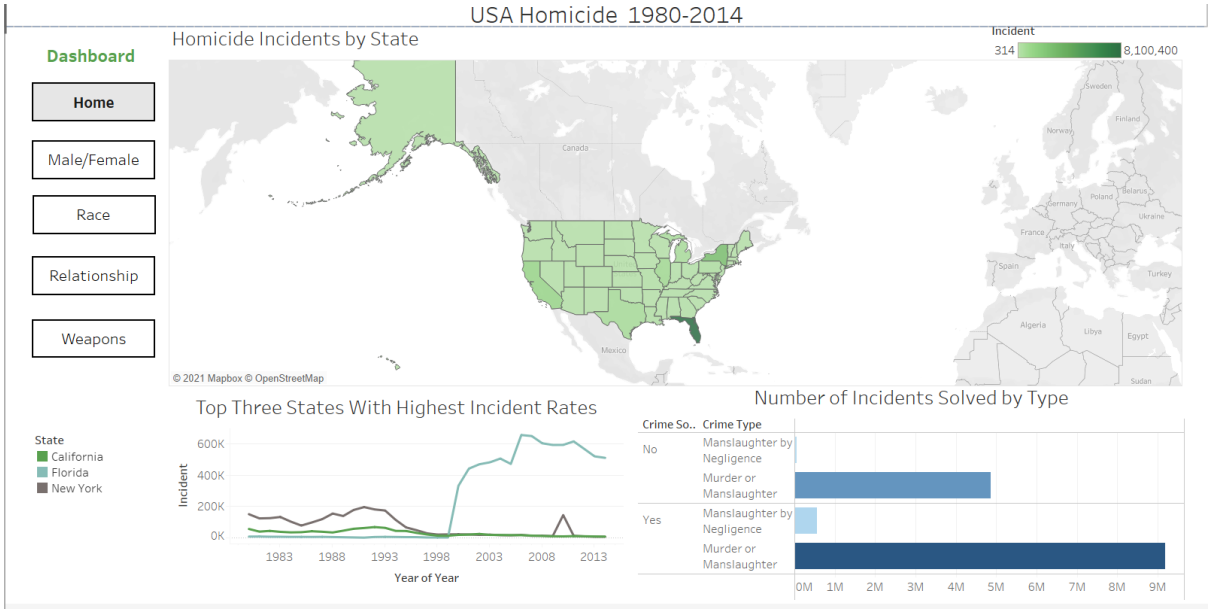

Figure 5

When analysing and visualising any data one is bound to come across some problems, my main problems were choosing the correct type of visualisation that would portray the data in the most suitable way. For example the visualisation for the relationships, at first I tried representing it using a horizontal bar chart which wasn't very effective and was quite confusing (figure 6) so I decided to go with a test table which turned out to be more straight forward and to the point (figure 7).

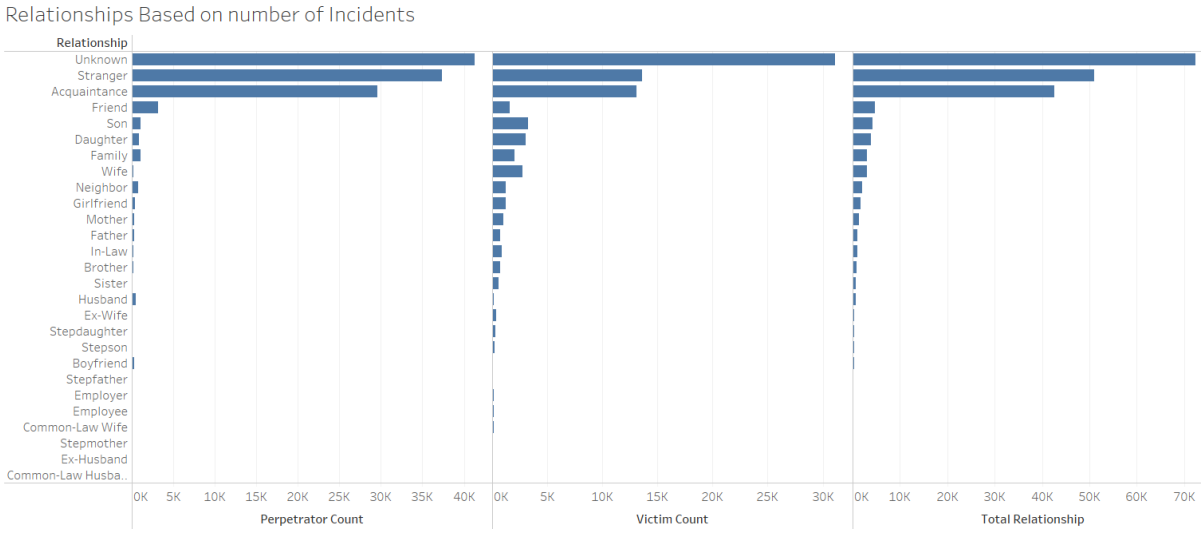

Figure 6

Relationships Based on number of Incidents

Relationship	Perpetrator Count	Victim Count	Total Relationship
Unknown	41,338	31,153	72,491
Stranger	37,376	13,626	51,002
Acquaintance	29,533	13,124	42,657
Friend	3,185	1,586	4,771
Son	1,054	3,229	4,283
Daughter	853	3,034	3,887
Family	1,037	2,059	3,096
Wife	226	2,737	2,963
Neighbor	747	1,250	1,997
Girlfriend	381	1,258	1,639
Mother	331	977	1,308
Father	296	741	1,037
In-Law	186	842	1,028
Brother	210	702	912
Sister	86	575	661
Husband	468	160	628
Ex-Wife	26	386	412
Stepdaughter	44	318	362
Stepson	81	228	309
Boyfriend	256	52	308
Stepfather	135	108	243
Employer	78	160	238
Employee	62	140	202
Common-Law Wife	32	158	190
Stepmother	22	68	90
Ex-Husband	66	12	78
Common-Law Husba..	42	16	58

Figure 7

Also when creating the map I encountered a problem where the data was not showing up on the graph, the problem was that the location was set to Ireland when the data was for the United States. This immediately fixed the problem.

As we can see from this chart (figure 8) of the sex of perpetrators, it is clear that there is significantly more male perpetrators than female. There is a wide range of reasons that might explain this violent behaviour in males e.g., childhood trauma, psychopathy or as a result of social constructs, where it is shameful for males to express feelings so over time, they redirect their feelings to something so violent as homicide. Whatever the reasons may be, they affect males more than females.

We also come to know that males are significantly more likely than females to be a victim of homicide in the United States. From an article by Science Daily[1], it states that males and females tend to kill for different reasons “Male serial killers tend to "hunt" their victims, who are often strangers to them. Female serial killers tend to "gather" their victims -- targeting people around them who they may already know, often for financial gain.”. This explanation would make sense as to why more males than females are victims of homicide.

After I had extracting this information, now we can see if there is a pattern in female and male perpetrators and their victims. Putting the data against each other to compare and see if female and male perpetrators where more likely to kill males or females.

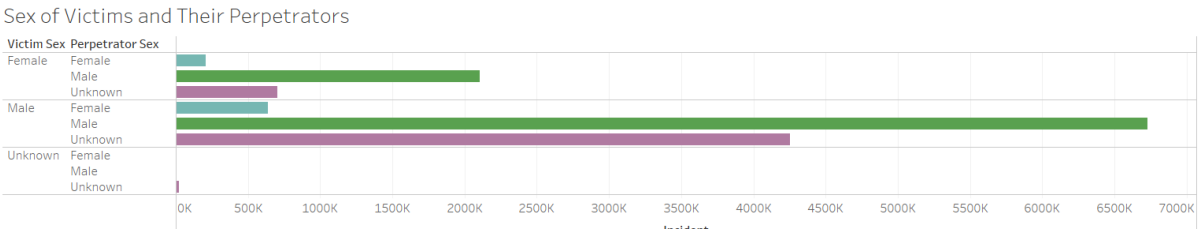

Figure 8

We get some interesting information from the chart above (figure 8). We come to find that male perpetrators are more likely to kill males over females, and similarly female perpetrators are more likely to have male victims. As we have seen there are more male perpetrators, even though females do not commit murder on the same scale as males, when they do decide to kill there is a higher chance of it being a male. Now this may be due to the fact that females tend to kill those they are close to, and they often kill for financial gain. And in a lot of cases even currently, men hold a significant amount if not all the household income. So, it would make sense that female perpetrators would target male victims. Looking at just the raw data you would have had a very difficult time trying to get to a conclusion about the connection between sex and the likelihood of homicide.

Now I will examine the distribution of weapons or methods used. I created the horizontal bars (figure 9) listing the number of incidents where each weapon or method is used. From the chart we can see that the weapon of choice for most is a handgun. Most people would expect that result as owning a handgun in the United states is or norm, this is because it is not against the law. Perpetrators clearly use this to their advantage because it is easier to use and carry and for most people it's already in their home.

The second choice for perpetrators is a knife followed by a blunt object, in a lot of cases both are found in homes, so perpetrators have them easy at hand or they use what is in the house of their unsuspecting victims.

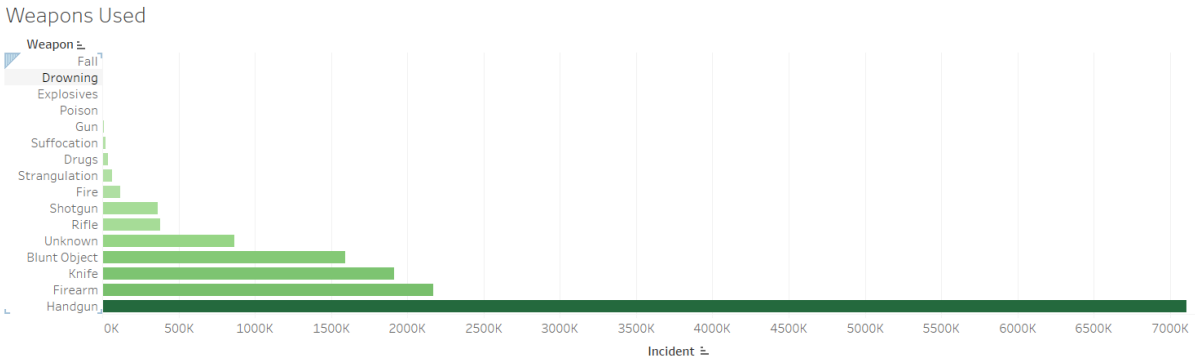

Figure 9

Next, I decided to find the distribution of homicides based on the relationship between perpetrator and their victim. In the table below we can see that 72,491 incidents are unknown, this could affect how we visualise this data but, looking at what we do know, 51002 are strangers and 42657 are acquaintances, while there is are approximately much fewer close family or friends. This would also relate back to the sexes of the perpetrators, as we found out that majority are males and that males tend to kill those they don't know, while females kill those closest to them.

Relationships Based on number of Incidents

Relationship	Perpetrator Count	Victim Count	Total Relationship
Unknown	41,338	31,153	72,491
Stranger	37,376	13,626	51,002
Acquaintance	29,533	13,124	42,657
Friend	3,185	1,586	4,771
Son	1,054	3,229	4,283
Daughter	853	3,034	3,887
Family	1,037	2,059	3,096
Wife	226	2,737	2,963
Neighbor	747	1,250	1,997
Girlfriend	381	1,258	1,639
Mother	331	977	1,308
Father	296	741	1,037
In-Law	186	842	1,028
Brother	210	702	912
Sister	86	575	661
Husband	468	160	628
Ex-Wife	26	386	412
Stepdaughter	44	318	362
Stepson	81	228	309
Boyfriend	256	52	308
Stepfather	135	108	243
Employer	78	160	238
Employee	62	140	202
Common-Law Wife	32	158	190
Stepmother	22	68	90
Ex-Husband	66	12	78
Common-Law Husba..	42	16	58

Figure 10

Now we will look the distribution of homicides by race or victims and their perpetrators. Figure 11 shows the race of the victims, it is evident that most are either white or black. While the minority are either Asian/Pacific Islanders or Native Americans/Alaska Natives. According to the United States Census Bureau [2], 76.3% of the population of the United States is White, 13.4% are Black, while 1.3% are Native Americans/Alaska Natives, and Asian/Pacific Islanders make up 6.1%. this gives us an idea as to why most victims are White or Black.

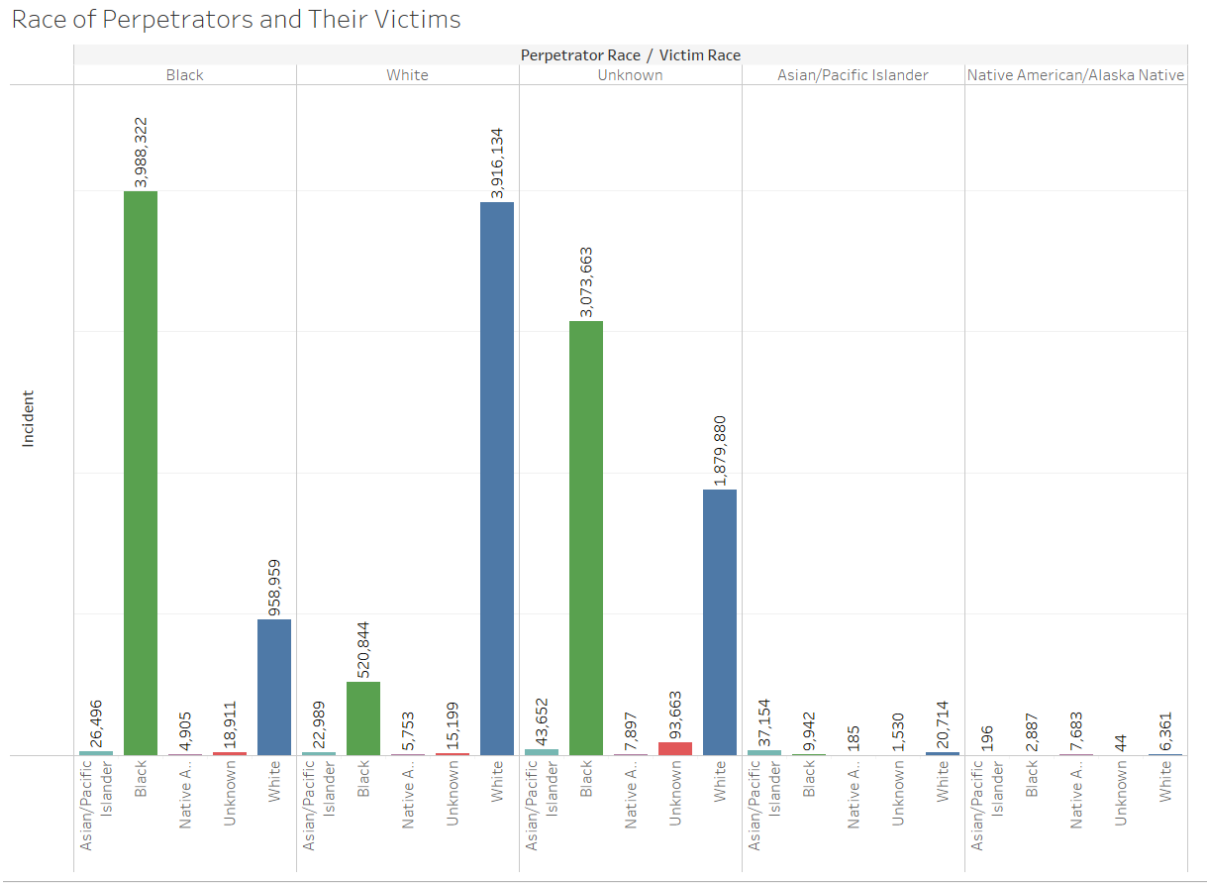

Figure 11

From the chart above in figure 11 we can see as we established earlier Black and White where the leading races for the victims and perpetrators.

Here we can also see that black perpetrators had majority black victims followed by white, which is still considerably low compared to the number of black victims at the hands of black perpetrators.

Again, we see a similar instance when it comes to white perpetrators. Majority of their victims are white followed by black victims but compared to the number of white it is considerably low.

We see from the char below figure 12 that there is a significant number of unsolved incidents but the number of solved still exceeds the unsolved, we can also see that manslaughter by negligence occurs a lot less than murder or manslaughter which makes sense as to why there is a low amount of unsolved cases.

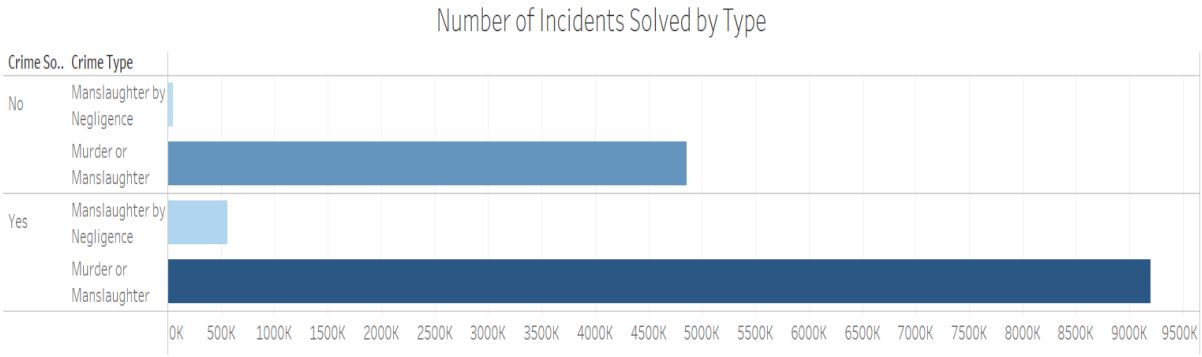

Figure 12

The story told by my visualisation is quite clear.

Firstly, we looked at the sexes of the victims and their perpetrators and we came to find that most perpetrators and victims are males.

We then looked at the weapons used and determined that handguns whereby far the most popular option for perpetrators, followed by knives and blunt objects, all of which are items found in the home of a typical American home.

Looking at the relationships between victims and perpetrators we found that a large portion where strangers or acquaintances, followed by a smaller portion whom where family or friend. And we also know that males usually kill those they don't know while females kill those closest to them.

Then we saw that most victims and perpetrators where either black or white, and that they mostly killed people from their own race.

Lastly, we seen that although there is a large number of unsolved cases there is still a larger number of those that are solved.

Now we know that there is a definite pattern in homicides in the United States, this is great as knowing this can help us improve the surroundings and hopefully reduce the number of homicides in the United States of America.

My recommendations and directions for further analysis and visualisations for this data is to get more up to date data and make it more useful data, such as including the ethnicity which could help better identify the background of the victims and perpetrators, this could help better understand the situations. Also including a column of mental health problems, if we know the perpetrator had mental health problem then we know that if it is a large number of them that we then need to better how we deal and treat mental health patients or increase awareness about it to hopefully prevent future homicides from occurring.

Citations

[1] ScienceDaily

https://www.sciencedaily.com/releases/2019/03/190320110622.htm

[2] United States Census Bureau
https://www.census.gov/quickfacts/fact/table/US/RHI325219