CA682 Data Management and Visualisation

Name	Siddanth Jagadish
Student Number	18213376
Programme	MSc in Computing (Data Analytics)
Module Code	CA682
Assignment Title	Data Visualisation
Submission date	17th December 2018
Module coordinator	Suzanne Little

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found recommended in the assignment guidelines.

Name: Siddanth Jagadish

Date: 17th December 2018

Title: The summary of US gun violence over the years 2013-2018

Introduction:

The United States of America is a well-known country for many certain things. It is generally to be for the richness and also the armed forces of these countries are really very strong and also are considered to have obtained about 33% of the world's wealth. But one more thing to look out is it has also got a lot of robbery cases and gun violence, as the access for weapons there is not that difficult because of which many number of people are involved in some or the other crime or robbery. It is very disheartening to see that there are a huge lot of cases of gun firing over the years and it would be interesting to try to get that information and display it through better visuals which might even make it easier in understanding.

Dataset:

The gun violence is considered as a major threat in USA over the years and there is a collection of huge set of data which is available on kaggle. Kaggle is considered as one of the greatest platforms to get the access to data set of desired interests. It is generally known as the storehouse of data. The data set consists of variety of data, as shown in the above screenshot. So the majorly involved data are the date at which the incident has happened, state of the person who is involved in the incident, the gender and age of the participants and mainly the number of injured and killed of a particular state on particular date.

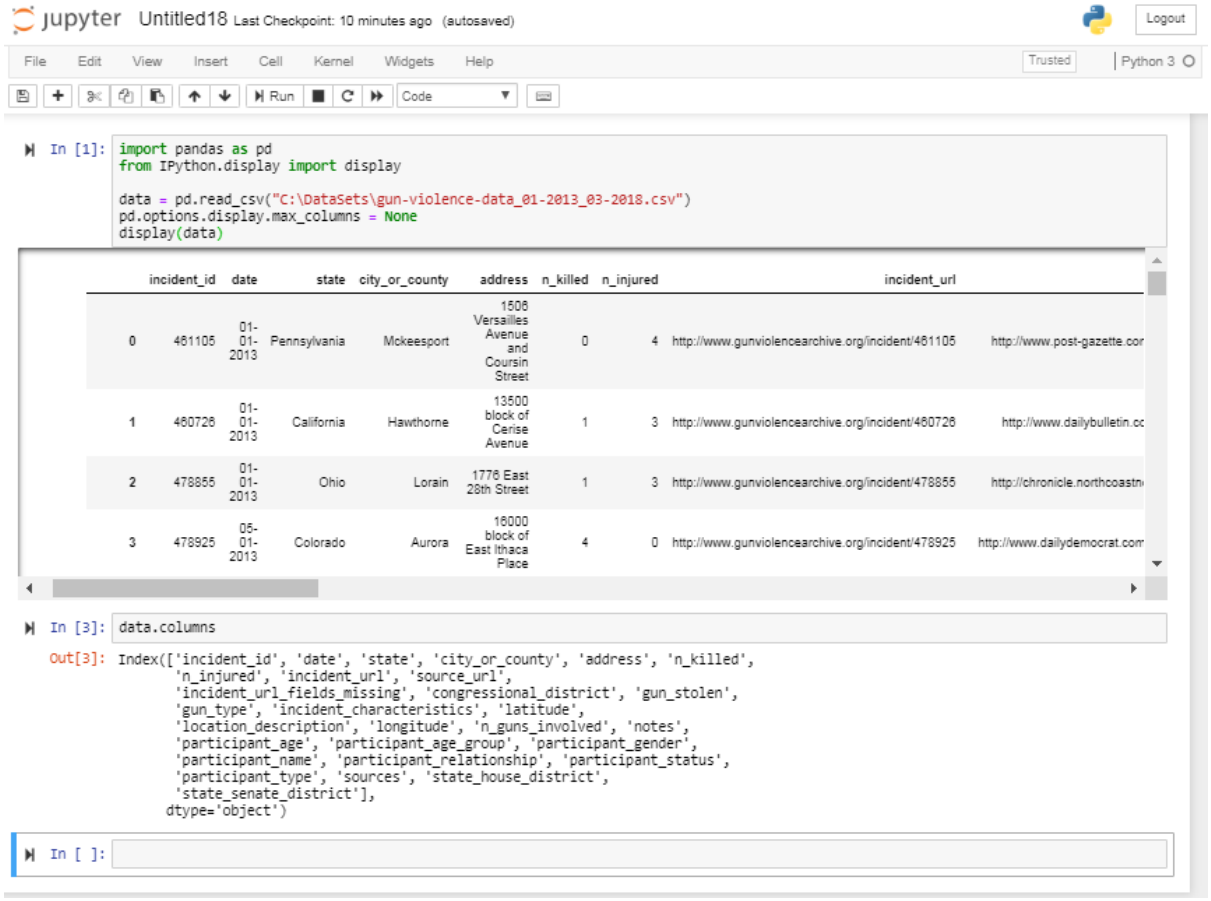


Fig : Data set of gun violence over the years 2013-2018 in united states.

The initial stage or process is considered to be data cleaning. I have used pandas in python for cleaning the data. As there were few not assisted or missing values and also there were some random patterns assigned with the data (such as participants_age used to look like 1::18, 2::20 now I had to make cleansing to look like 18, 20). Well these are not the only things I had do remove other patterns and format the data, which is quite a job.

Process:

The main determination is to show the various visuals by using the data of gun violence to either categorize that these states have the highest number of participants who were the reasons for the incidents to happen or to show people of what age were majorly involved in the incident. The plotting can be done and shown mainly by the major use of pandas series and matplotlib. The various graphs plotted to some specific purposes are explained below:

- The main focus in this part would be selecting the states which have got the most number of incidents. So this gives a better idea about the states which has got more number of incidents done and vice versa. By plotting a "Scatter plot" it could be seen that which state has got how many number of incidents done. This is also shown by the help of a "horizontal bar plot" graph which shows the graphical representation for the same

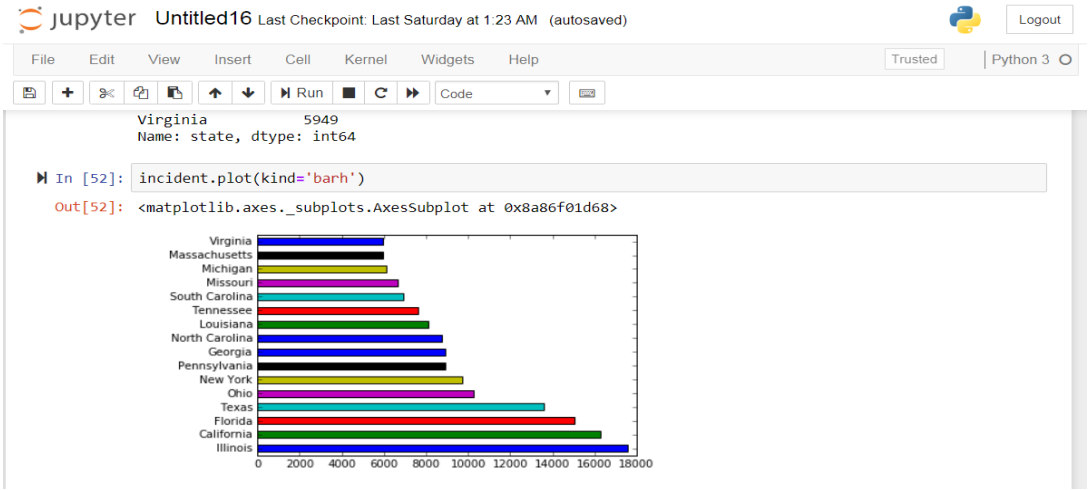


Fig: horizontal bar graph to show basic plotting of states regarding the number of incidents.

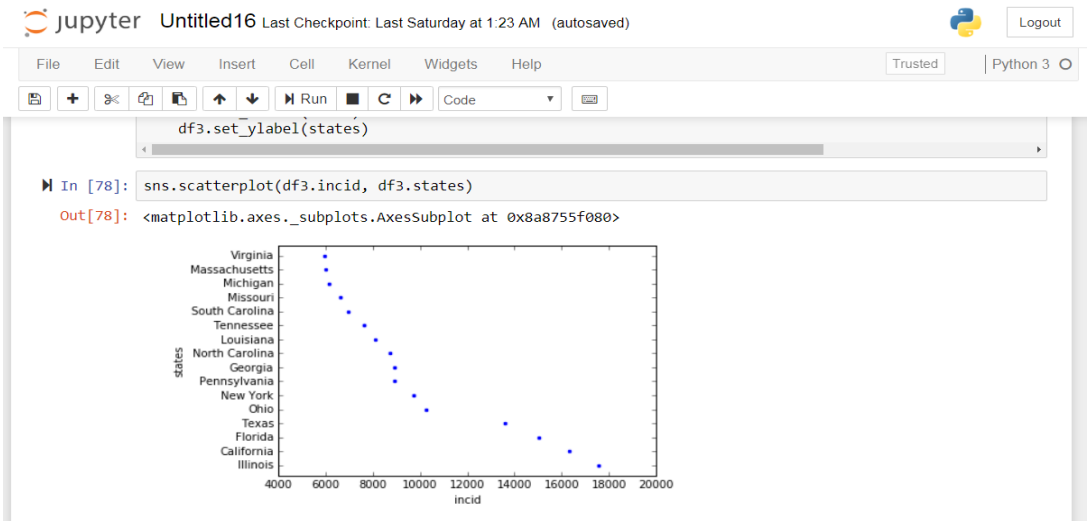


Fig: Scatterplot to show the states with more number of incidents.

- The consideration of age group of participants who were involved in the incidents also becomes a major part and an important thing because it gives a clear idea about the age factor involved in those incidents. By making use of a “Line plot”, it becomes easy to differentiate that what age group is majorly involved and vice versa. Through a simple line plot it would be the simplest mode of observation with not more complexity. As initially we can note that as the age increases, we can see that there is an increase in the number of incidents.

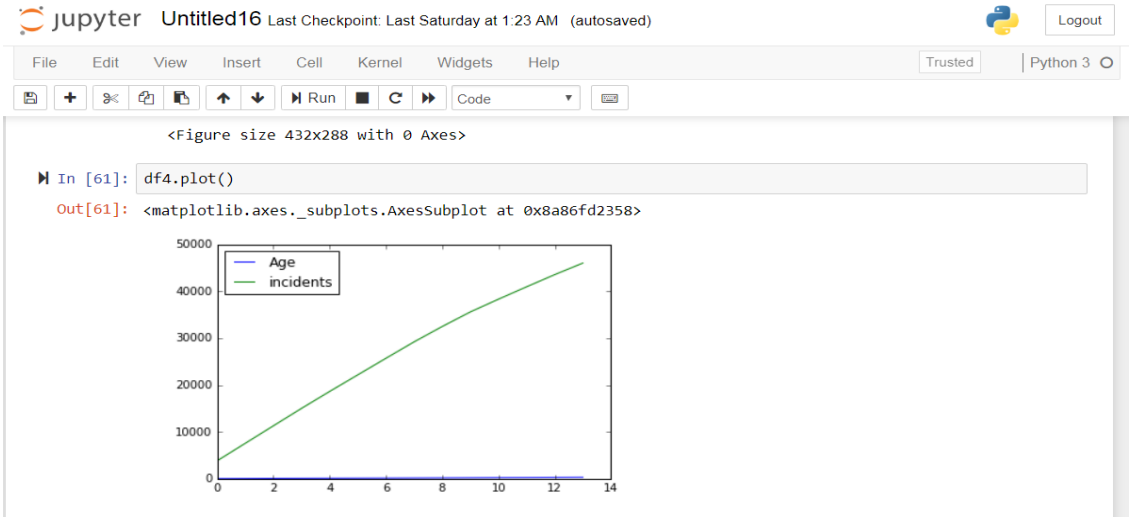
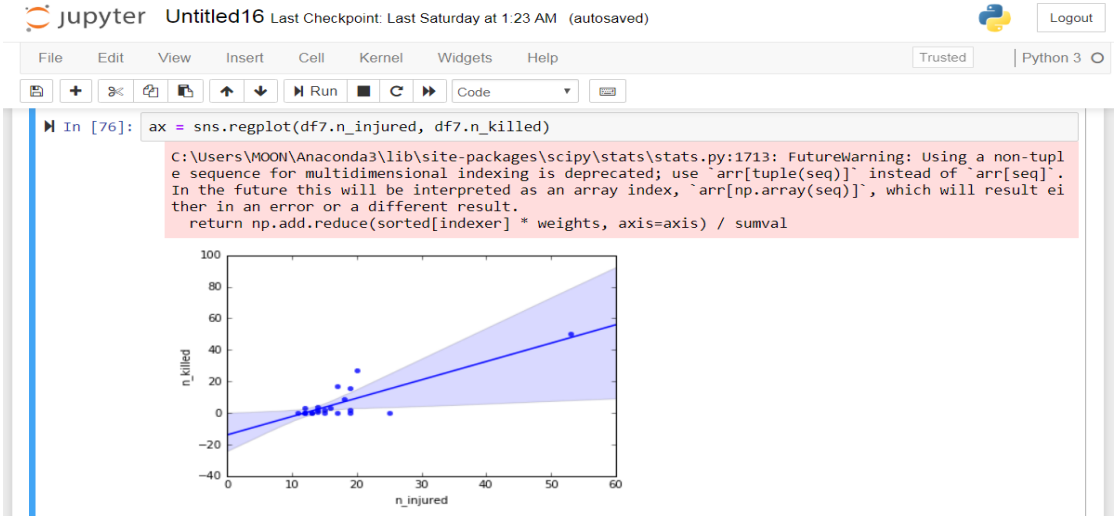


Fig: Line plot to differentiate participants age with number of incidents.

- The last graph is all about a basic idea of how the kills are successful based on the incidents. In this part it involves combining different states with number of incidents. So the states with number of incidents are grouped together but another functionality involved here is that the probability of kills is also calculated based on the number of incidents on those particular states. This can be clearly shown using a scatter plot with a regression line. So basically we can say that the regression line is interested to pass through majority of scattered points or values we can say. This gives a better observation, as we use a scatter plot regression line.



Results

The obtained observations using various plotting which does give a better description of those datasets and how nicely it can be visualized by just having a glimpse at the graphs as shown in the process above.

Criticising the work

- I surely think I would have done better with the help of d3.js but trying d3.js at this moment was not optimal for me.
- There could have been some interactivity which was lacked in this project.
- Could have also used better and more number of graphs to get a better idea about how well it can be described and displayed.

The link to the video Screen casting of this assignment is given here:
https://share.vidyard.com/watch/3mLff24KBgzqjxcGfFXy9P

https://share.vidyard.com/watch/3mLff24KBgzqjxcGfFXy9P?