Devam Patel & Parth Patel

Prof, Yan

09/02/2021

INFO250

Final Project

- What changes did we make after the initial design and why?

From the initial submission I changed all the graphs because I thought it would be difficult for a normal person with no data science background to understand what I had included. In this final project I decided to analyze new variables including generating some new variables from the given data. From the date column, I generated month and year columns and used them for analysis.

- The link to access our project

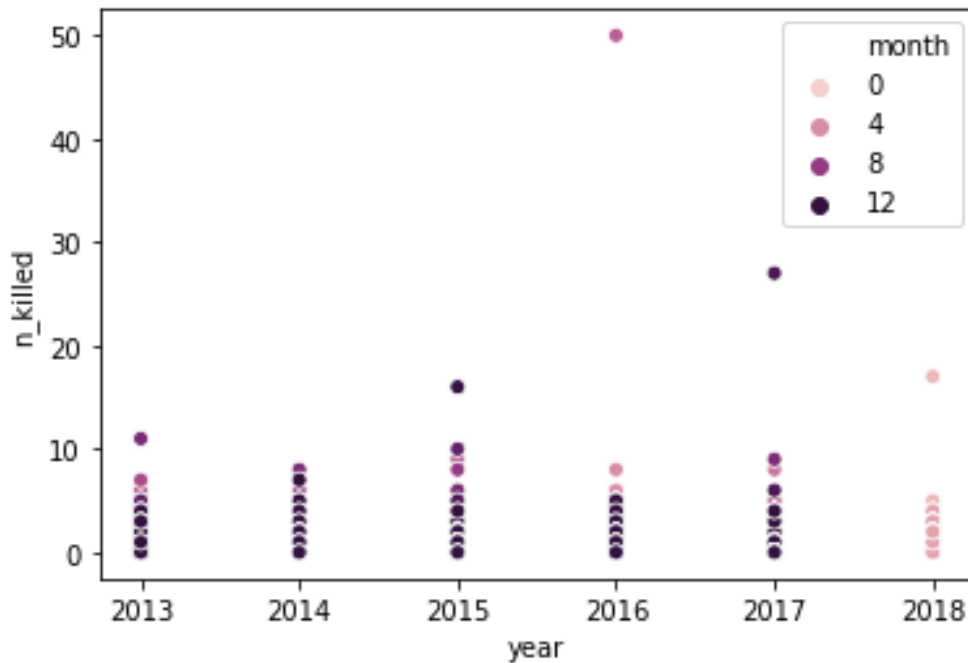https://github.com/dpatel6855/INFO250Final

Analysis

An analysis that showed the number of people killed in the gun violence incidents:
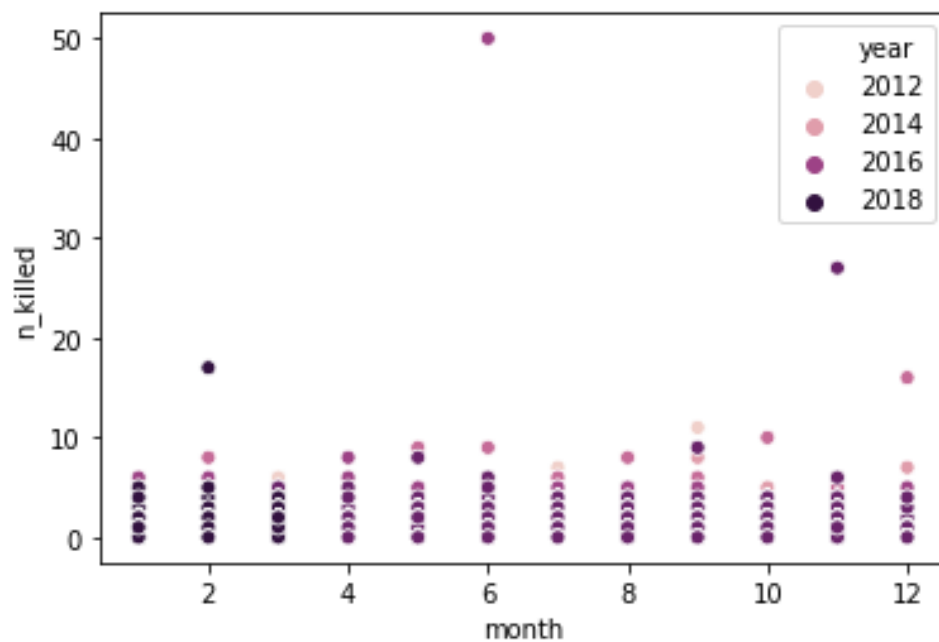
From the above graph it is evident that the number of people killed mostly were below ten people. However, the month of June recorded a high death of people at one incident.

Above is a graph of the number of kills but this time they are categories in terms of years. Like the number of kills categorized in months we can see that mostly the kills ranged below 10 kills per incident. However, in the year 2016, there was an incident that recorded a high number of kills.

Remember also in the month of June in the previous graph there was a high number of deaths. It is evident that the highest number of people killed in an incident happened in the month of June, the year 2016.
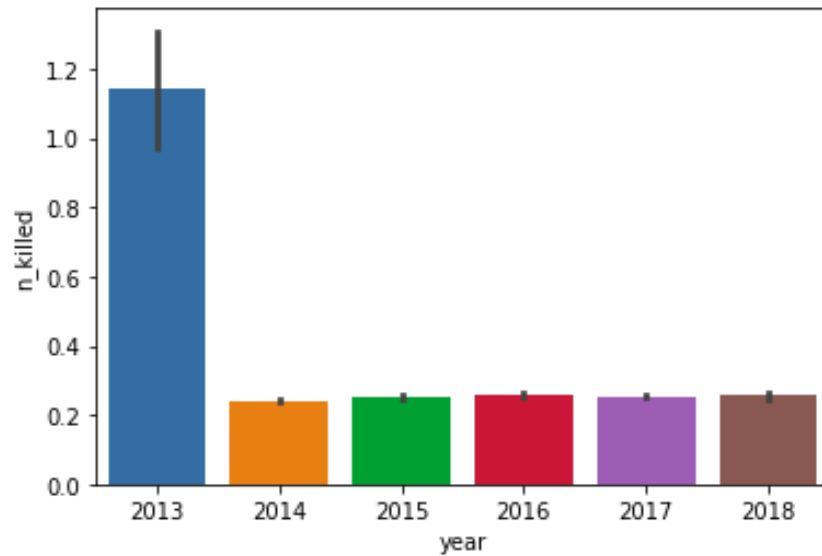
Like the scatter plots above, we now have bar plots. In the scatter plots we could not quantify the number of killed per year since they are only dots on the graph. However, with bar graphs now we can easily quantify the number of deaths that occurred in a year. From the graph above, we see that 2013, the number of kills were more than 5 times the number of kills in the consecutive years.
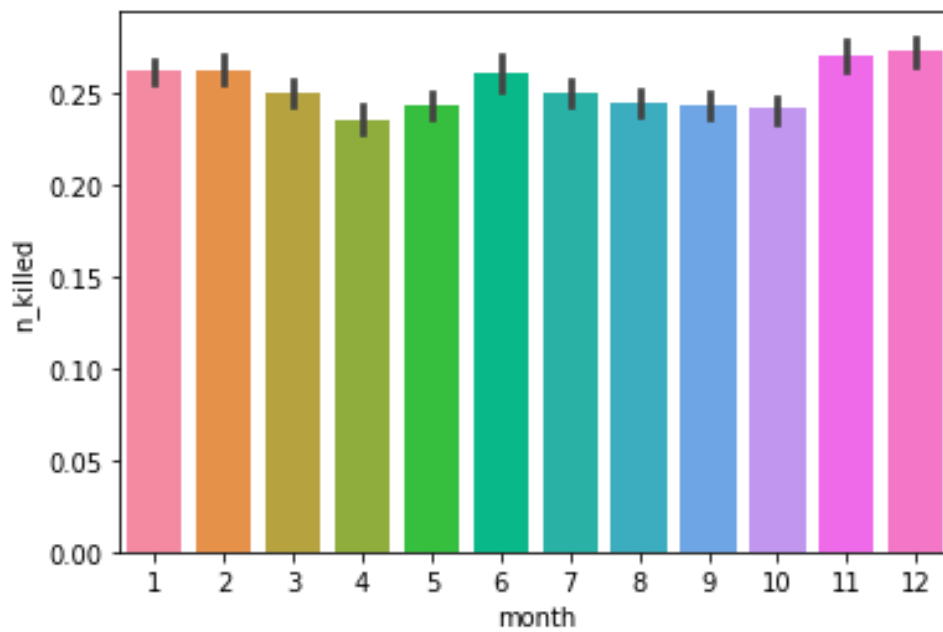
Now we get to visualize the same graph as the previous one but this time it is categorized in months. All the quantities seem to be ranging within the same bracket except that the month of December and November recorded the highest number of kills in this graph consecutively.
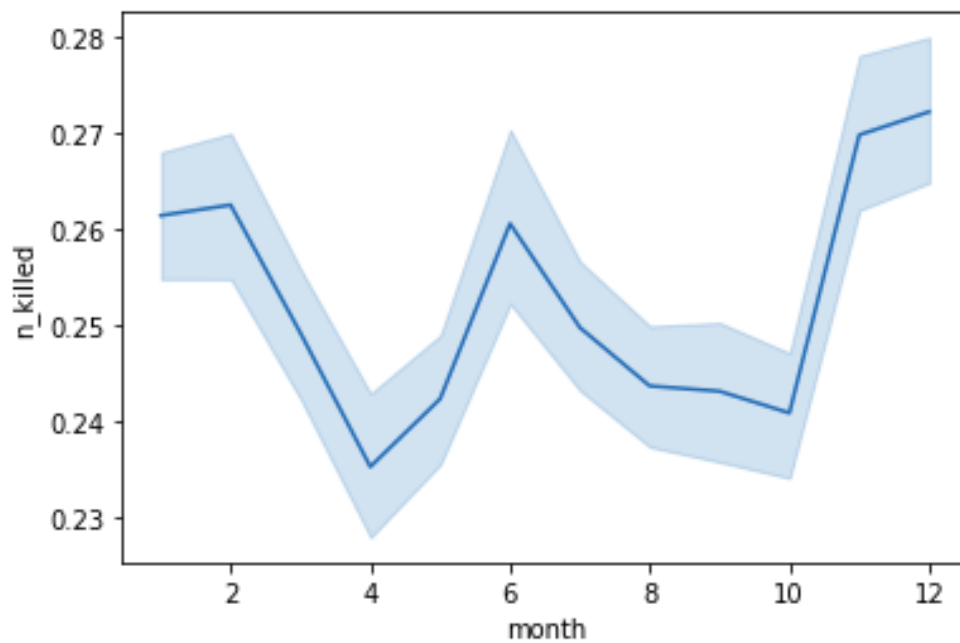
Next, we visualize this graph. It is a line plot. The data it depicts is like the previous graphs. Number of kills against the months. The margin you see that encloses the sharp blue line can be described as the range of the number of kills (like what regression looks like). They just show how far the values were from the value that was considered (the sharp blue line). Now back to our analysis. We can see that in the beginning months of the year the number of kills were high then they dropped as we approached the month of April after-which, the number started going up again till the month of June. Towards the month of October, the numbers started dropping again and hiked towards the month of December.
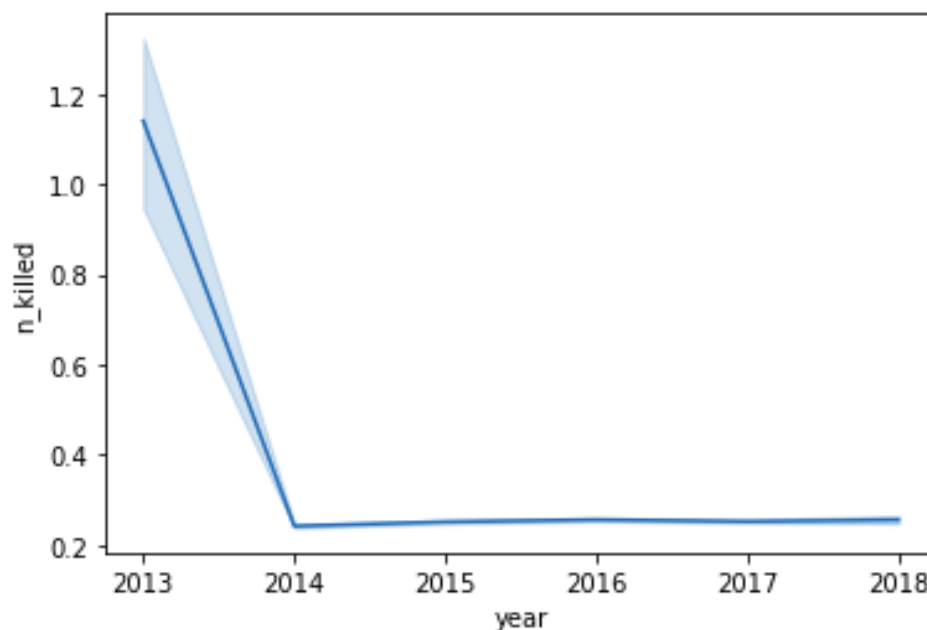
In this graph we can visualize that the number of kills dropped gradually over the years. In 2013, the number of kills was high with a high spread between the mean kills. After 2013, the number of kills dropped drastically and maintained in the same position with a low difference over the years till 2018.
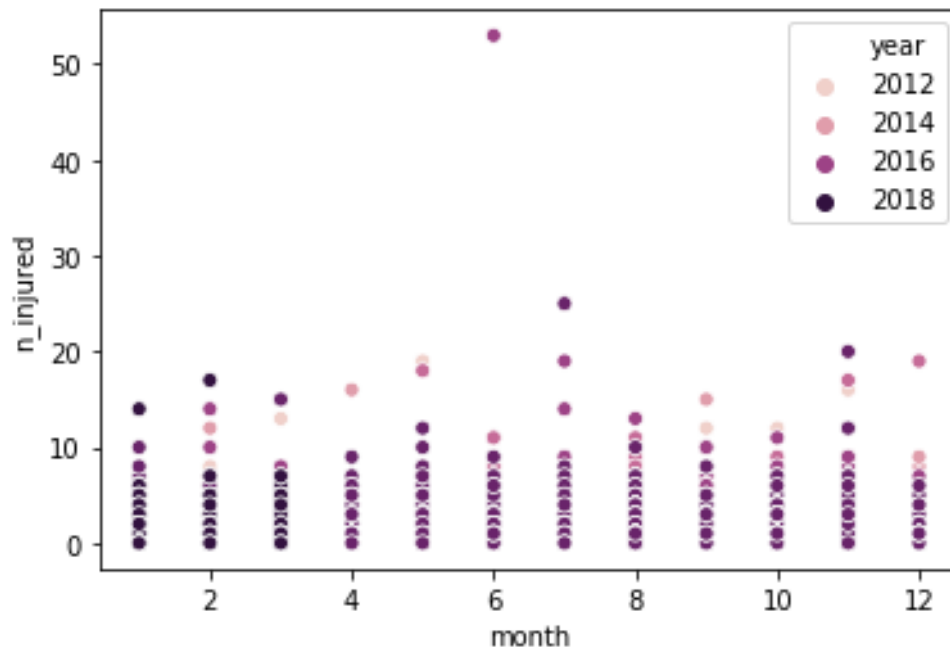
From the above graph it is evident that the number of people injured mostly were below ten people. However, the month of June recorded a high number of injured people at one incident.
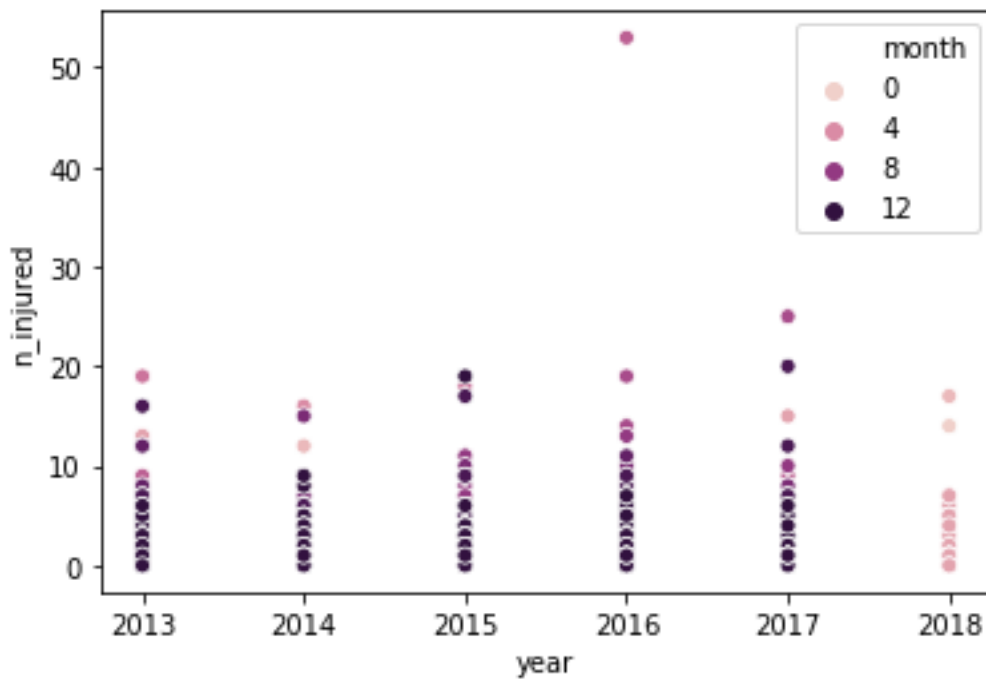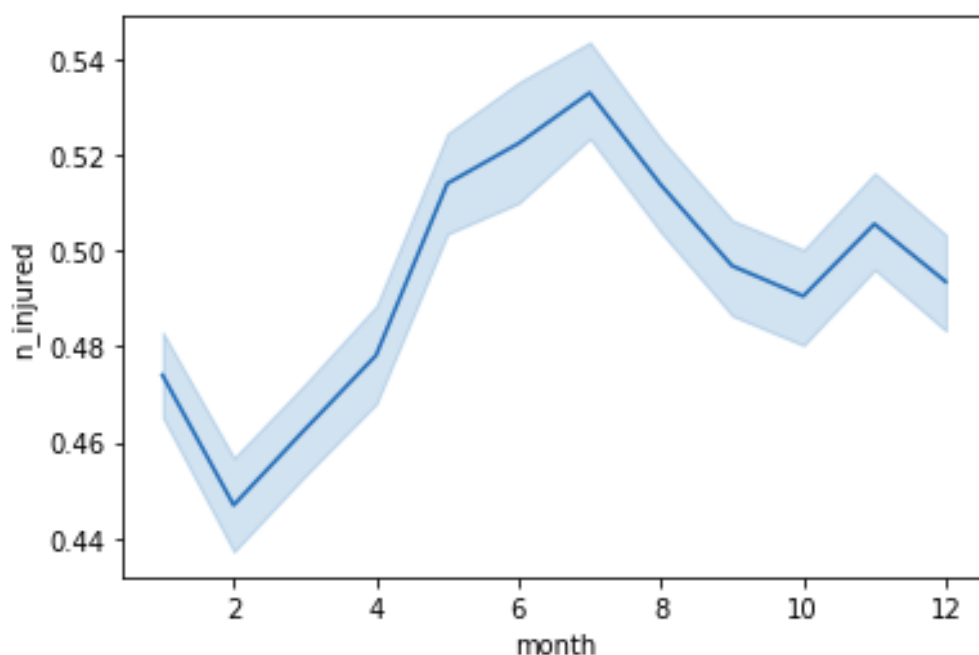
Above is a graph of the number of injuries but this time they are categories in terms of years. Like the number of injuries categorized in months we can see that mostly the kills ranged below 10 injuries per incident. However, in the year 2016, there was an incident that recorded a high number of kills.

Remember also in the month of June in the previous graph there was a high number of injuries. It is evident that the highest number of people injured in an incident happened in the month of June, the year 2016.

Like the start of this analysis, we also saw that the highest number of people killed in an incident happened in the month of June, the year 2016.

We have started to get incites from this data. Probably in the year 2016, the month of June there was some big gun incident. Probably, a terrorist attack with a lot of casualties or a hijack in a public place with a lot of hostages who some of them got killed or injured.

We come across another line graph but this time round it is a line graph of the number of injured people against month. We can see that in the beginning months of the year the number of injuries were high (but not that high) then they dropped as we approached the month of February after-which, the number started spiking up again till the month of July. Towards the month of October, the numbers started dropping again and hiked slightly towards the month of November before dropping again at the month of September.
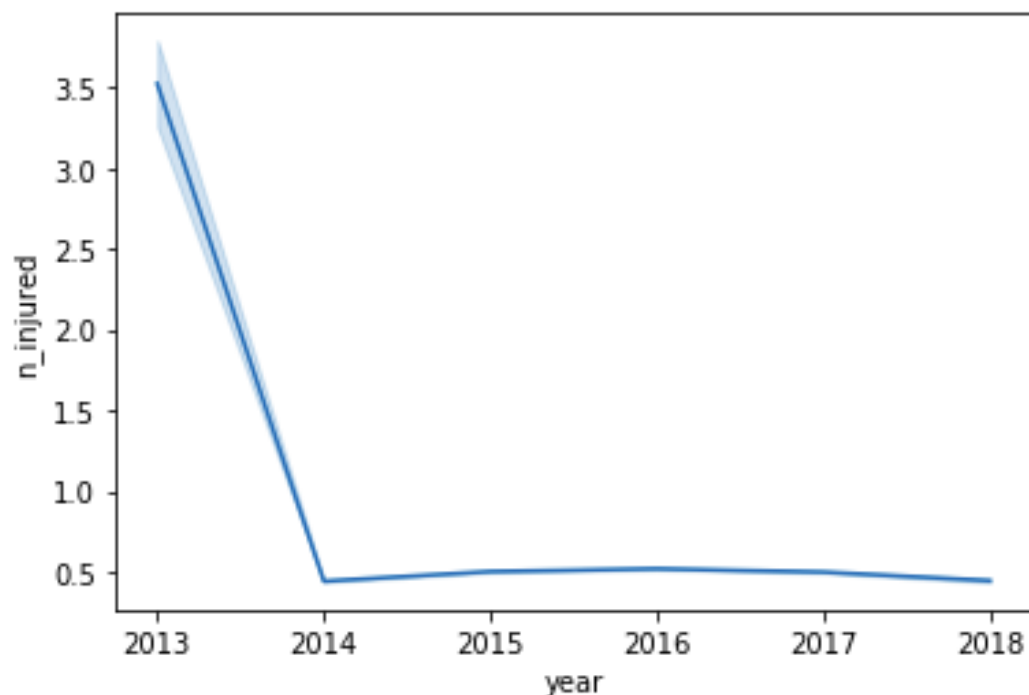
In this graph we can visualize that the number of injuries dropped gradually over the years. In 2013, the number of injuries was high with a high spread between the mean injuries. After 2013, the number of injuries dropped drastically and maintained in the same position with a low difference over the years till 2018.

1) What are the most and least useful aspect of this course do you see?

   The most useful aspect from this course was to learn visualization in tableau and R. I didn't have experience in both software but after taking this class I feel confident making visualization in both software.

   The least aspect from this course was the hw. The homework itself was not hard to do but the instructions sometimes were not clear which made us confuse about the requirements of the project.

2) Have interesting research questions and discuss if your research questions are addressed in the conclusion section.

- How to make audience more interactive in this visualization?
- How to create the US map based on the longitude and latitude in python, so user can see the people killed by just looking at the map?