

Analysis of US Gun Violence and County Health Factors

By Lesley Dwyer

Analysis Domain, Questions and Plan

Gun violence is prevalent in America and has even been called a public health crisis. It is a problem unique to America where gun ownership is much more common than in other countries [6]. I am interested in the factors that may lead to higher incidents of gun violence. I have chosen to investigate this from the angle of health factors, rather than, say, gun regulation or gun ownership in a given area. The reasons for this are 1) gun regulation and gun ownership has already been shown to be related to gun violence [6,7] and 2) there is a perception among the media and the public that mass shootings are related to mental illness, although this has largely been shown not to be the case. In addition, suicide is the most common cause of gun deaths [7]. I want to further investigate not only mental health but all health factors to see if there is a relationship with gun violence.

I plan to join data from two sources to analyse this issue. These are 1) Gun Violence Data [5] produced from data originally compiled by the non-profit organisation Gun Violence Archive [4] and the 2018 County Health Rankings compiled by The County Health Rankings & Roadmaps program [2]. The research questions I plan to answer are: 1) Do any health factors relate to gun violence? 2) Is unhealthy behaviour more likely to lead to violent behaviour? 3) Can we predict the level of gun violence from health factors?

The gun violence data has 239,677 observations and 29 columns for the years 2013–2018 with a row for each gun incident. To limit the scope, I filtered the file to use only the 2017 gun violence data, which contains 59,881 observations. The county health file has data for a similar timeframe (2010–2017) but is aggregated by state and county. This file has 3,142 observations with 164 columns and 99 columns over two tabs.

In order to join these datasets together in a reliable way, I will use the county FIPS code. This is not available in the gun violence data, but the latitude and longitude values are, and these can be used with a web API to obtain the FIPS codes [3]. The county health file has several redundant columns, which I will remove to reduce the number of features. After adding the FIPS code, I will aggregate the gun violence data by county. Next, I will merge the data sets together, joining on county FIPS, to do some analysis. I'll check for outliers and impute missing values. I will plot feature distributions and transform some of the data. I will check how features correlate with each other and with gun violence. I will do feature selection using PCA and ranked correlations. Finally, I will build and test a multiple regression model with the most important features to predict gun incidents by county.

Findings and Reflections

The datasets required a fair amount of cleansing and preparation before any analysis could be undertaken. Some outliers and missing values were present and a few columns needed to be transformed to have a more normal distribution. I found that a square root transformation worked well for four of the columns and a log transformation was better for two others. The transformations also helped with reducing the effect of outliers. I also needed to rescale four columns to bring them in line with the rest of the data so PCA would be more accurate.

I used PCA for feature selection rather than feature reduction, as I not only wanted to predict gun violence but wanted to understand which features had a relationship to gun violence.

Surprisingly, PCA resulted in different results to using a correlation matrix for feature selection. I was expecting there to be at least some overlap, but the lists were very different.

Following the analysis, I was able to answer my original research questions:

1) Do any health factors relate to gun violence?

Yes. The following county health factors, grouped by categories from the original data, are related to gun violence in the US:

Health Outcomes: Age-Adjusted Mortality, % Low Birth Weight, Child Mortality Rate, HIV Prevalence Rate and Infant Mortality Rate.

Health Behaviours: Chlamydia Rate, Food Environment Index, % Food Insecure, Teen Birth Rate, % With Access to Exercise Opportunities and Motor Vehicle Mortality Rate.

Clinical Care: Primary Care Physician Rate, Dentist Rate, Preventable Hospital Stays Rate and Other Primary Care Providers Rate.

2) Is unhealthy behaviour more likely to lead to violent behaviour?

This analysis cannot establish causation between unhealthy behaviour and violent behaviour. However, six of the health factors that relate to gun violence fall into the category of 'Health Behaviours' in the original county health dataset. As these health factors increase, gun incidents also increase: Chlamydia Rate, % Food Insecure and Teen Birth Rate. As the Food Environment Index value decreases, gun incidents increase. Motor Vehicle Mortality Rate and % With Access to Exercise Opportunities have only a very small linear correlations individually to gun incidents.

3) Can we predict the level of gun violence from health factors?

Yes. The second model can be used to predict the percent of gun incidents by county (per population) with a 1.5% average test error and the first model can predict the same with a 2.5% average test error.

The above findings show that there is a relationship between the health of a county and the level of gun violence. On the topic of whether mental illness has a relationship to higher gun incidents, this does not appear to be the case, at least at a county level. Some of these health factors, such as 'Food Insecure', may be proxies for other things like poverty, which would require more investigation. The Guardian did a comprehensive review of gun violence from 2015 at a census tract level looking at demographics and did find poverty to be one of the key indicators for gun violence [1].

Factors like gun regulation, gun ownership, demographics and environment were not included in the analysis. It was specifically focussed on health to see which, if any, health factors relate to gun incidents. The analysis was also limited by the fact that it was at a county level and not at an individual level. Therefore, it cannot be shown that the health factors for individuals relate to those individuals being involved in gun violence.

Although gun violence is a complex problem related to several factors like demographics, gun regulation and gun ownership [1,7], these findings could potentially be used for policy or planning at

a county level to not only improve health but to investigate whether improvements in the health factors listed above also lead to improvements in gun violence.

Computational Notebook:

<https://smcse.city.ac.uk/student/aczd115/INM430CourseworkNotebook.html>

References

- [1] Aufrichtig, A., Beckett, L., Diehm, J., Lartey, J. (2017) 'Want to fix gun violence in America? Go local.', *The Guardian*, Available at: <https://www.theguardian.com/us-news/ng-interactive/2017/jan/09/special-report-fixing-gun-violence-in-america> (Accessed: 1 November 2018).
- [2] County Health Rankings & Roadmaps (2018) *Rankings Data and Documentation*. Available at: <http://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation> (Accessed: 30 October 2018).
- [3] Federal Communications Commission (no date) *API Documentation for Developers*. Available at: https://geo.fcc.gov/api/census/#!/area/get_area (Accessed: 3 December 2018).
- [4] Gun Violence Archive (2018) *Gun Violence Archive*. Available at: <https://www.gunviolencearchive.org/> (Accessed: 1 November 2018).
- [5] Ko, J. (2018) *Gun Violence Data*. Available at: <https://www.kaggle.com/jameslko/gun-violence-data> (Accessed: 1 November 2018)
- [6] Lopez, G. (2018) 'America's unique gun violence problem, explained in 17 maps and charts', *Vox*, 8 November. Available at: <https://www.vox.com/policy-and-politics/2017/10/2/16399418/us-gun-violence-statistics-maps-charts> (Accessed: 14 December 2018).
- [7] Swanson, J. W., McGinty, E. E., Fazel, S., Mays, V. M., (2015) 'Mental illness and reduction of gun violence and suicide: bringing epidemiologic research to policy', *Annals of Epidemiology*, Volume 25, Issue 5, 2015, Pages 366-376, doi: <https://doi.org/10.1016/j.annepidem.2014.03.004>.