**Extract Transform Load**

We decided to do our project on Gun Violence in America. We extracted data from a couple sources on the internet. We gathered data from GunViolence.org, CDC.gov, and ATF.gov. From GunViolence.org we downloaded multiple CSV documents displaying “Accidental Deaths with a Firearm”. One for children, one for teens, and one for adults. From CDC.gov we downloaded a CSV as well. This document consisted of “Firearms Mortality by State”. From our last source ATF.gov we downloaded 8 separate PDF documents which contained tables for “Registered Firearms by State”. We ran this through a program which read the tables within the PDF and wrote them to a CSV document. This was unfortunately not a free service and we only got through 4 files before it asked us to enter our credit card info. We entered a little bit of data manually for the remaining 4 files.

We used Jupiter Notebook and Pandas to transform the data into workable datasets. In each accidental death the categories were Incident ID, Incident Date, State, City or County, Address, # killed, # injured, and Operations. We filtered the DataFrame to only include State, Incident Date, # Killed, and Operations. We then renamed the Operations category “Age” and replaced NaN value with “Adult”, “Teen”, or “Child” because without replacing the value there was no way to differentiate between the age of each incident. The accidental deaths for adult, teen, and children were combined into one DataFrame using concat function within the Pandas library. The csv files for “firearm deaths by state “and “number of registered guns by state” did not need any changes besides changing the abbreviations for the state to the actual state name which was done within excel.

For the “Load” portion of this project we had to take a look at the data we had compiled so far. We saw that in each dataset there was a column for “State”. These three datasets are intuitively relational, so we decided to load them SQL database. In doing so we can easily query our database to find a wide array of results. We used primary keys and foreign keys to allow these tables reference one another. We believe through gathering the data from these separate sources, manipulating and merging it together in the “Transform” process, and loading it into a SQL database we have created a pretty unique dataset.