

CODEBOOK FOR GUN VIOLENCE DATASET

Feb 25, 2024: This dataset on mass shootings for the period 2014-2023 was provided on Feb 19, 2024 by the Data Manager (Ms. Sharon Williams) at the Gun Violence Archive (<https://www.gunviolencearchive.org/>) on a data request. Their current data includes additional information including the latitude and longitude of the location of the incident.

Minimal curation was done on this data – the date variable was split into year, month and day.

#	Variable/ Attribute/ Feature	Meaning
1	Incident_ID, Incident_Date, Incident_Time, State_Name, City_or_County, Business_or_Location_Name, Address, Latitude, Longitude, Victims_Killed, Victims_Injured, Suspects_Killed, Suspects_Injured, Suspects_Arrested, Incident_Characteristics, Year, Month, Day	Self-Explanatory

The combined dataset (2014-2023), and ten datasets, one for each year from 2014 through 2023 are provided. This dataset replaces the previously curated data as this contains more details.

Note: As before (July 2023) we tried to combine the mass shootings data with data from <https://simplemaps.com/data/us-cities>, to additionally include information on population and density of those locations. However, this resulted in too many missing values in the augmented columns. Therefore, this was not pursued. The R code used to curate and attempt this augmentation is provided along with the datasets.

The notes from before (below) are struck out, but not deleted, and remain in this document for a historical perspective.

~~**July 21, 2023:** This dataset on mass shootings is accessed from <https://www.gunviolencearchive.org/>. This dataset was (re)downloaded on Oct 30, 2022. A mass shooting is defined as four or more people injured or killed, because of firearms, excluding the shooter.~~

~~The combined dataset (2014-2019), and six datasets, one for each year from 2014 through 2019 are provided.~~

The mass shooting dataset was augmented on July 21, 2023, with the latitude and longitude and other information to enable mapping. The US cities lat-long information was downloaded on July 21, 2023 from <https://simplemaps.com/data/us-cities>.

The variables in this dataset are:

#	Variable/ Attribute/ Feature	Meaning
1	Incident.ID, Incident.Date, State, City.Or.County, Address, Killed, Injured, Year, Month, Day	Self-Explanatory
The following field names and descriptions are copied from https://simplemaps.com/data/us-cities		
	fieldname	description
	city	The name of the city/town.
	city_ascii	city as an ASCII string.
	lat	The latitude of the city/town.
	lng	The longitude of the city/town.
	state_id	The state or territory's USPS postal abbreviation.
	state_name	The name of the state or territory that contains the city/town.
	county_fips	The 5-digit FIPS code for the primary county. The first two digits correspond to the state's FIPS code.
	county_name	The name of the primary county (or equivalent) that contains the city/town.
	population	An estimate of the city's urban population. The value of population is estimated using 2021 ACS 1-year and 2019 ACS 1-year data (when available for urban areas) and 2021 ACS 5-year data when not.
	density	The estimated population per square kilometer.
	incorporated	TRUE if the place is a city/town. FALSE if the place is just a commonly known name for a populated area.

timezone	The city's time zone in the tz database format. (e.g. <i>America/Los Angeles</i>)
ranking	An integer from 1-5 that captures the importance of a city (1 is most important, 5 least important).
zip	A string containing all five-digit zip codes in the city/town, delimited by a space. Learn more.
id	A 10-digit unique id generated by SimpleMaps. It is consistent across releases and databases (e.g. World Cities Database).
source	For some cities, our data is generated from a <i>polygon</i> representing the city, for others we simply have a <i>point</i>.
military	TRUE if this place is a military establishment such as a fort or base.

PS: ~~There was an error in the data published prior to Oct 10, 2022 as a result of which years 2016-2020 pointed to 2015 data. **Thanks to Dr. Caroline Hutchings (Norco College)** for bringing it to our attention. In the process of fixing it, we also realized the data for 2020-current does not include mass shootings alone. So as of Oct 30, 2022 we are making available mass shootings data only for 2014-2019. In the future, we will make data available on ALL gun violence.~~