Data Set Title
Exploratory Analysis

Jameson Buie jbuie@bellarmine.edu
Name, student@bellarmine.edu

I. INTRODUCTION

This data set contains data from nearly 400 shootings that occurred at schools from the primary to secondary level. These shootings occurred before, during, or right after class times. There are many other instances that were not included in this data set because they did not occur while school was in session. Additionally, this data set does not include shootings that happened on college campuses because shootings on large public campuses happen more often and do not necessarily involve students as shooters or victims, which is what we are exploring. We chose this data set because we want to explore the relationships between the shooters, the victims, time, and motives that were involved in these tragedies. Also, school shootings seem to be occurring more often so we will check to see if there is an uptick in number since 1999 and if there is a reason for this.

II. DATA SET DESCRIPTION

III.

This data set contains 387 samples with 50 columns with various data types. There are 22 columns with string values and 18 columns with integer values. The rest of the columns have different data types.

Table 1: Data Types and Missing Data

Variable Name	*Data Type*	*Missing Data (%)*
uid	Ordinal/Int64	0%
nces_school_id	Nominal/object	1.55%
school_name	Nominal/object	0%
nces_district_id	Nominal/float64	6.2%
district_name	Nominal/object	6.46%
date	Interval/object	0%
school_year	nominal/object	0%
year	Interval/int64	0%
time	Nominal/object	6.2%
day_of_week	Nominal/object	0%
city	Nominal/object	0%
state	Nominal/object	0%
school_type	Nominal/object	0%
enrollment	Ratio/object	0%
killed	Ratio/int64	0%
injured	Ratio/int64	0%
casualties	Ratio/int64	0%
shooting_type	Nominal/object	0%
age_shooter1	Ratio/float64	29.2%
gender_shooter1	Nominal/object	22.7%
race_ethnicity_shooter1	Nominal/object	61.75%
shooter_relationship1	Nominal/object	31.3%
shooter_deceased1	Ratio/float64	32.3%
deceased_notes1	Nominal/object	90.1%
age_shooter2	Ratio/float64	96.9%
gender_shooter2	Nominal/object	96.9%

race_ethnicity_shooter2	Nominal/object	98.2%
shooter_relationship2	Nominal/object	98.4%
shooter_deceased2	Ratio/float64	98.4%
deceased_notes2	Nominal/object	99.7%
white	Ratio/float64	2.6%
black	Ratio/float64	2.6%
hispanic	Ratio/object	2.6%
asian	Ratio/float64	2.6%
american_indian_alaska_native	Ratio/float64	2.8%
Hawaiian_native_pacific_islander	Ratio/float64	32.3%
two_or_more	Ratio/float64	32.3%
resource_officer	Nominal/int64	0%
weapon	Nominal/object	35.7%
weapon_source	Nominal/object	72.9%
lat	Nominal/float64	0.25%
long	Nominal/float64	0.25%
staffing	Ratio/float64	6.2%
low_grade	Nominal/object	1.3%
high_grade	Nominal/object	1.3%
lunch	Nominal/object	12.9%
county	Nominal/object	0.5%
state_fips	Nominal/int64	0%
county_fips	Nominal/int64	0%
ulocale	Nominal/float64	0.77%

IV. Data Set Summary Statistics

Narrative introduction to the section.

Table 2: Summary Statistics for School Shootings Data

	uid	nces_district_id	year	enrollment	killed	injured	casualties	age_shooter1	shooter_deceased1	age_shooter2
count	387.000000	3.630000e+02	387.000000	387.000000	387.000000	387.000000	387.000000	274.000000	262.000000	12.000000
mean	194.542636	2.683346e+06	2014.204134	1072.315245	0.519380	1.108527	1.622739	19.284672	0.152672	16.916667
std	112.678322	1.638245e+06	7.452592	800.176703	2.171772	2.215958	3.847791	9.748823	0.360359	1.505042
min	1.000000	1.000080e+05	1999.000000	22.000000	0.000000	0.000000	0.000000	6.000000	0.000000	15.000000
25%	97.500000	1.200480e+06	2008.000000	458.000000	0.000000	0.000000	0.000000	15.000000	0.000000	15.750000
50%	194.000000	2.614280e+06	2017.000000	869.000000	0.000000	1.000000	1.000000	16.000000	0.000000	17.000000
75%	290.500000	4.155394e+06	2021.000000	1525.000000	0.000000	1.000000	1.000000	19.000000	0.000000	18.000000
max	391.000000	5.516470e+06	2023.000000	5264.000000	26.000000	21.000000	34.000000	72.000000	1.000000	20.000000

	white	black	hispanic	asian	american_indian_alaska_native	hawaiian_native_pacific_islander	two_or_more
count	377.000000	377.000000	377.000000	377.000000	376.000000	262.000000	262.000000
mean	375.193634	354.403183	286.310345	38.649867	8.436170	5.015267	31.683206
std	464.915289	433.758974	523.058885	81.618368	25.989355	23.552659	37.592018
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	14.000000	26.000000	13.000000	1.000000	0.000000	0.000000	4.250000
50%	150.000000	180.000000	76.000000	8.000000	2.000000	1.000000	17.000000
75%	592.000000	574.000000	293.000000	34.000000	6.000000	2.000000	46.750000
max	2256.000000	2736.000000	4679.000000	846.000000	331.000000	338.000000	199.000000

	resource_officer	lat	long	staffing	state_fips	county_fips	ulocale
count	387.000000	386.000000	386.000000	363.000000	387.000000	387.000000	384.000000
mean	0.271318	37.070614	-92.176124	61.546846	26.679587	26776.191214	19.947917
std	0.445215	5.121773	15.553162	38.739547	16.130641	16166.320448	10.668633
min	0.000000	21.310400	-157.837200	3.000000	1.000000	1009.000000	11.000000
25%	0.000000	33.686293	-97.506534	31.210000	12.000000	12061.000000	11.000000
50%	0.000000	37.044535	-87.582092	55.300000	26.000000	26125.000000	13.000000
75%	1.000000	40.647249	-80.818711	86.050000	41.000000	41045.000000	22.000000
max	1.000000	61.213281	-70.952603	203.320000	55.000000	55139.000000	43.000000

There should be a table for **EACH** categorical variable.

Table 3: Proportions for XXX (n=yyy)

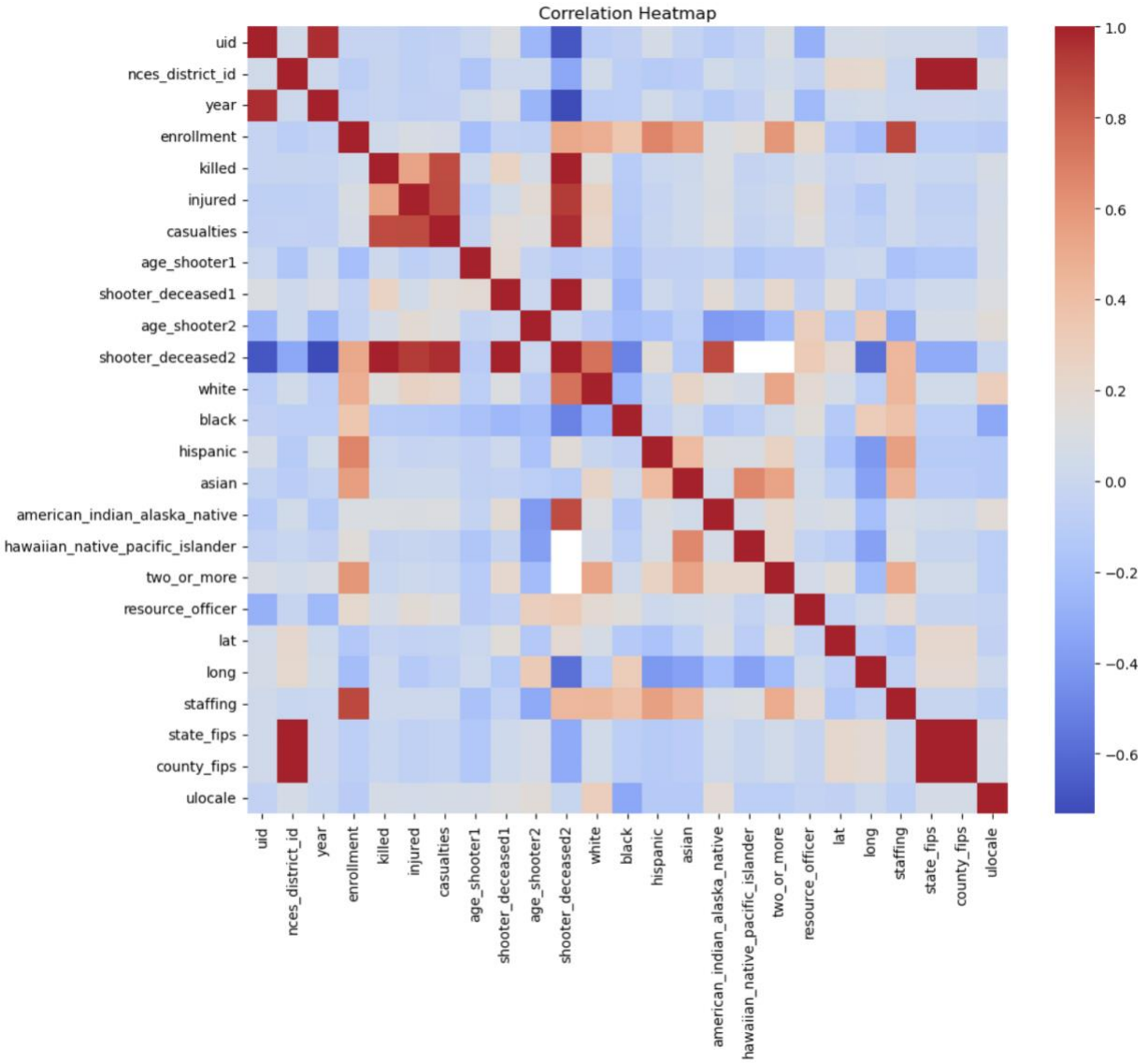
```
Enter the name of the categorical variable you want to analyze: state
Table: Proportions for state (n=387)
                                Category  Frequency  Proportion (%)
state
California                     California         40       10.335917
Texas                               Texas         25        6.459948
Florida                           Florida         24        6.201550
North Carolina             North Carolina         21        5.426357
Pennsylvania                 Pennsylvania         17        4.392765
Illinois                          Illinois         17        4.392765
Michigan                         Michigan         16        4.134367
Alabama                           Alabama         15        3.875969
Tennessee                       Tennessee         14        3.617571
Ohio                                 Ohio         14        3.617571
Louisiana                       Louisiana         13        3.359173
Georgia                           Georgia         12        3.100775
New York                         New York         11        2.842377
Maryland                         Maryland         10        2.583979
Virginia                         Virginia          9        2.325581
Colorado                         Colorado          9        2.325581
Washington                     Washington          9        2.325581
South Carolina             South Carolina          9        2.325581
Indiana                           Indiana          8        2.067183
Minnesota                       Minnesota          7        1.808786
Mississippi                   Mississippi          7        1.808786
New Mexico                     New Mexico          7        1.808786
Wisconsin                       Wisconsin          6        1.550388
Oregon                             Oregon          6        1.550388
District of Columbia  District of Columbia          6        1.550388
Missouri                         Missouri          6        1.550388
Nevada                             Nevada          5        1.291990
Connecticut                   Connecticut          5        1.291990
Arkansas                         Arkansas          5        1.291990
Arizona                           Arizona          4        1.033592
Oklahoma                         Oklahoma          4        1.033592
Utah                                 Utah          4        1.033592
Kentucky                         Kentucky          3        0.775194
Massachusetts                Massachusetts          3        0.775194
New Jersey                     New Jersey           2        0.516796
```

Table 4: Correlation Table/Tables

	uid	nces_district_id	year	enrollment	killed	injured	casualties	age_shooter1	shooter_deceased1	age_shooter2
uid	1.000000	0.042187	0.963085	-0.025348	-0.028500	-0.064305	-0.054792	0.003640	1.041955e-01	-2.463755e-01
nces_district_id	0.042187	1.000000	0.020270	-0.083693	-0.019968	-0.063224	-0.048331	-0.153146	2.098247e-02	2.269084e-02
year	0.963085	0.020270	1.000000	-0.047438	-0.023534	-0.059858	-0.049164	0.036878	8.560481e-02	-2.634954e-01
enrollment	-0.025348	-0.083693	-0.047438	1.000000	0.035647	0.092667	0.071134	-0.203723	-4.480853e-02	-5.746786e-02
killed	-0.028500	-0.019968	-0.023534	0.035647	1.000000	0.536801	0.873578	0.019197	2.518978e-01	5.963251e-02
injured	-0.064305	-0.063224	-0.059858	0.092667	0.536801	1.000000	0.878648	-0.073413	5.073998e-02	1.822929e-01
casualties	-0.054792	-0.048331	-0.049164	0.071134	0.873578	0.878648	1.000000	-0.030507	1.696898e-01	1.386175e-01
age_shooter1	0.003640	-0.153146	0.036878	-0.203723	0.019197	-0.073413	-0.030507	1.000000	1.846638e-01	-4.076741e-02
shooter_deceased1	0.104195	0.020982	0.085605	-0.044809	0.251898	0.050740	0.169690	0.184664	1.000000e+00	-7.066224e-17
age_shooter2	-0.246376	0.022691	-0.263495	-0.057468	0.059633	0.182293	0.138618	-0.040767	-7.066224e-17	1.000000e+00
shooter_deceased2	-0.688296	-0.332075	-0.732753	0.518477	0.995495	0.928936	0.965767	-0.097964	1.000000e+00	-1.146633e-16
white	-0.089456	0.048576	-0.083866	0.479130	0.146292	0.256630	0.230777	-0.072598	1.158904e-01	-1.019426e-01
black	-0.055729	-0.078576	-0.082378	0.348351	-0.108369	-0.110622	-0.125190	-0.180246	-2.442653e-01	-2.082940e-01
hispanic	0.065778	-0.113549	0.048606	0.673209	0.004195	-0.022304	-0.013948	-0.051465	1.447245e-02	-1.718693e-01
asian	-0.035983	-0.094656	-0.038329	0.564772	0.020446	0.026420	0.026005	-0.058559	-4.741822e-02	-8.248787e-02
nerican_indian_alaska_native	-0.106734	0.042621	-0.114172	0.103626	0.104870	0.096399	0.115079	-0.038867	1.745086e-01	-3.961052e-01
vaiian_native_pacific_islander	-0.045947	-0.002774	-0.051324	0.156496	-0.034799	-0.012199	-0.029266	-0.151195	-3.533745e-02	-3.711126e-01
two_or_more	0.089652	0.046599	0.091336	0.594418	-0.005283	0.022970	0.004356	-0.107271	2.110871e-01	-2.146036e-01
resource_officer	-0.282842	-0.021079	-0.233015	0.206685	0.062871	0.177524	0.138543	-0.099982	-5.276053e-02	2.834839e-01
lat	0.053966	0.209507	0.026483	-0.134980	-0.028646	-0.045271	-0.041021	0.000318	1.594628e-01	-1.264651e-01
long	0.062205	0.201757	0.049499	-0.210393	0.006133	-0.122762	-0.066904	0.015768	-1.156028e-01	3.202630e-01
staffing	0.032688	-0.007925	0.003408	0.887845	0.014607	0.018780	0.018093	-0.178152	-4.285740e-02	-3.247048e-01
state_fips	0.032389	0.999977	0.006967	-0.079976	-0.006404	-0.060630	-0.039226	-0.141469	3.162218e-02	6.980092e-02
county_fips	0.032840	0.999952	0.007363	-0.079574	-0.006215	-0.060046	-0.038776	-0.141203	3.105155e-02	7.005786e-02
ulocale	-0.043705	0.064185	-0.002858	-0.100880	0.073064	0.057925	0.075781	0.072046	1.166805e-01	1.627880e-01

	lat	long	staffing	state_fips	county_fips	ulocale
uid	0.053966	0.062205	0.032688	0.032389	0.032840	-0.043705
nces_district_id	0.209507	0.201757	-0.007925	0.999977	0.999952	0.064185
year	0.026483	0.049499	0.003408	0.006967	0.007363	-0.002858
enrollment	-0.134980	-0.210393	0.887845	-0.079976	-0.079574	-0.100880
killed	-0.028646	0.006133	0.014607	-0.006404	-0.006215	0.073064
injured	-0.045271	-0.122762	0.018780	-0.060630	-0.060046	0.057925
casualties	-0.041021	-0.066904	0.018093	-0.039226	-0.038776	0.075781
age_shooter1	0.000318	0.015768	-0.178152	-0.141469	-0.141203	0.072046
shooter_deceased1	0.159463	-0.115603	-0.042857	0.031622	0.031052	0.116681
age_shooter2	-0.126465	0.320263	-0.324705	0.069801	0.070058	0.162788
shooter_deceased2	0.193012	-0.579011	0.433522	-0.314612	-0.315031	-0.009278
white	0.078893	-0.079125	0.442751	0.045109	0.044483	0.292716
black	-0.118271	0.308939	0.382562	-0.080705	-0.079345	-0.332485
hispanic	-0.177602	-0.406676	0.557602	-0.115821	-0.115785	-0.117891
asian	-0.067677	-0.361744	0.464852	-0.095985	-0.096395	-0.122001
american_indian_alaska_native	0.094216	-0.199768	0.079506	0.039607	0.038813	0.170362
hawaiian_native_pacific_islander	-0.088842	-0.366838	0.100649	-0.011203	-0.011922	-0.080356
two_or_more	0.149331	-0.222916	0.498713	0.048830	0.048077	-0.078911
resource_officer	-0.042266	0.038758	0.190876	-0.021051	-0.021296	-0.041571
lat	1.000000	-0.069547	-0.141575	0.214447	0.213243	-0.054422
long	-0.069547	1.000000	-0.061543	0.190657	0.191712	0.012939
staffing	-0.141575	-0.061543	1.000000	-0.012151	-0.011638	-0.067789
state_fips	0.214447	0.190657	-0.012151	1.000000	0.999974	0.065940
county_fips	0.213243	0.191712	-0.011638	0.999974	1.000000	0.065199
ulocale	-0.054422	0.012939	-0.067789	0.065940	0.065199	1.000000

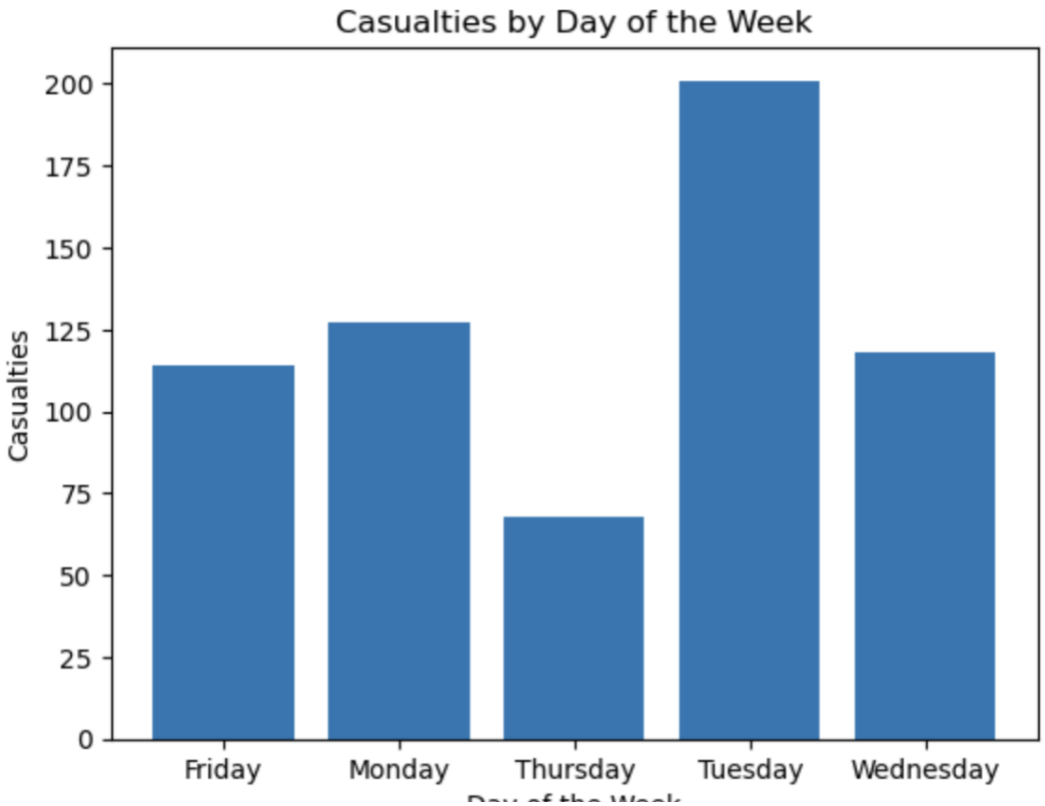
shooter_deceased2	white	black	hispanic	asian	american_indian_alaska_native	hawaiian_native_pacific_islander	two_or_more	resource_officer	
-6.882957e-01	-0.089456	-0.055729	0.065778	-0.035983	-0.106734	-0.045947	0.089652	-0.282842	**uid**
-3.320745e-01	0.048576	-0.078576	-0.113549	-0.094656	0.042621	-0.002774	0.046599	-0.021079	**nces_district_id**
-7.327530e-01	-0.083866	-0.082378	0.048606	-0.038329	-0.114172	-0.051324	0.091336	-0.233015	**year**
5.184772e-01	0.479130	0.348351	0.673209	0.564772	0.103626	0.156496	0.594418	0.206685	**enrollment**
9.954955e-01	0.146292	-0.108369	0.004195	0.020446	0.104870	-0.034799	-0.005283	0.062871	**killed**
9.289356e-01	0.256630	-0.110622	-0.022304	0.026420	0.096399	-0.012199	0.022970	0.177524	**injured**
9.657671e-01	0.230777	-0.125190	-0.013948	0.026005	0.115079	-0.029266	0.004356	0.138543	**casualties**
-9.796392e-02	-0.072598	-0.180246	-0.051465	-0.058559	-0.038867	-0.151195	-0.107271	-0.099982	**age_shooter1**
1.000000e+00	0.115890	-0.244265	0.014472	-0.047418	0.174509	-0.035337	0.211087	-0.052761	**shooter_deceased1**
-1.146633e-16	-0.101943	-0.208294	-0.171869	-0.082488	-0.396105	-0.371113	-0.214604	0.283484	**age_shooter2**
1.000000e+00	0.743108	-0.507073	0.164079	-0.113102	0.876000	NaN	NaN	0.316228	**shooter_deceased2**
7.431081e-01	1.000000	-0.259729	-0.009445	0.238985	0.113241	0.061393	0.526676	0.190605	**white**
-5.070730e-01	-0.259729	1.000000	-0.055784	0.036998	-0.118486	-0.076509	0.036471	0.150503	**black**
1.640790e-01	-0.009445	-0.055784	1.000000	0.414089	0.095771	0.080022	0.257738	0.022222	**hispanic**
-1.131019e-01	0.238985	0.036998	0.414089	1.000000	0.048626	0.657078	0.534979	0.039513	**asian**
8.760004e-01	0.113241	-0.118486	0.095771	0.048626	1.000000	0.061841	0.209610	0.062283	**american_indian_alaska_native**
NaN	0.061393	-0.076509	0.080022	0.657078	0.061841	1.000000	0.210969	-0.038775	**hawaiian_native_pacific_islander**
NaN	0.526676	0.036471	0.257738	0.534979	0.209610	0.210969	1.000000	0.053754	**two_or_more**
3.162278e-01	0.190605	0.150503	0.022222	0.039513	0.062283	-0.038775	0.053754	1.000000	**resource_officer**
1.930124e-01	0.078893	-0.118271	-0.177602	-0.067677	0.094216	-0.088842	0.149331	-0.042266	**lat**
-5.790108e-01	-0.079125	0.308939	-0.406676	-0.361744	-0.199768	-0.366838	-0.222916	0.038758	**long**
4.335223e-01	0.442751	0.382562	0.557602	0.464852	0.079506	0.100649	0.498713	0.190876	**staffing**
-3.146123e-01	0.045109	-0.080705	-0.115821	-0.095985	0.039607	-0.011203	0.048830	-0.021051	**state_fips**
-3.150312e-01	0.044483	-0.079345	-0.115785	-0.096395	0.038813	-0.011922	0.048077	-0.021296	**county_fips**
-9.278370e-03	0.292716	-0.332485	-0.117891	-0.122001	0.170362	-0.080356	-0.078911	-0.041571	**ulocale**

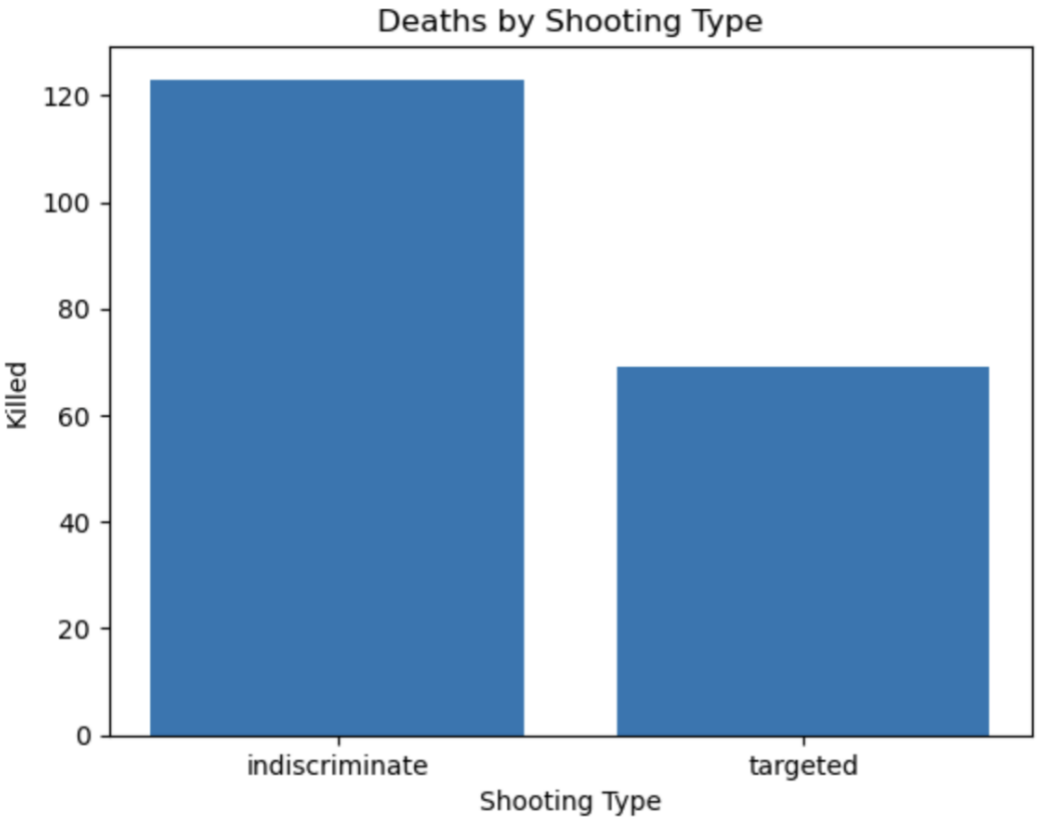

Correlation Heatmap
uid
nces_district_id
year
enrollment
killed
injured
casualties
age_shooter1
shooter_deceased1
age_shooter2
shooter_deceased2
white
black
hispanic
asian
american_indian_alaska_native
hawaiian_native_pacific_islander
two_or_more
resource_officer
lat
long
staffing
state_fips
county_fips
ulocale
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

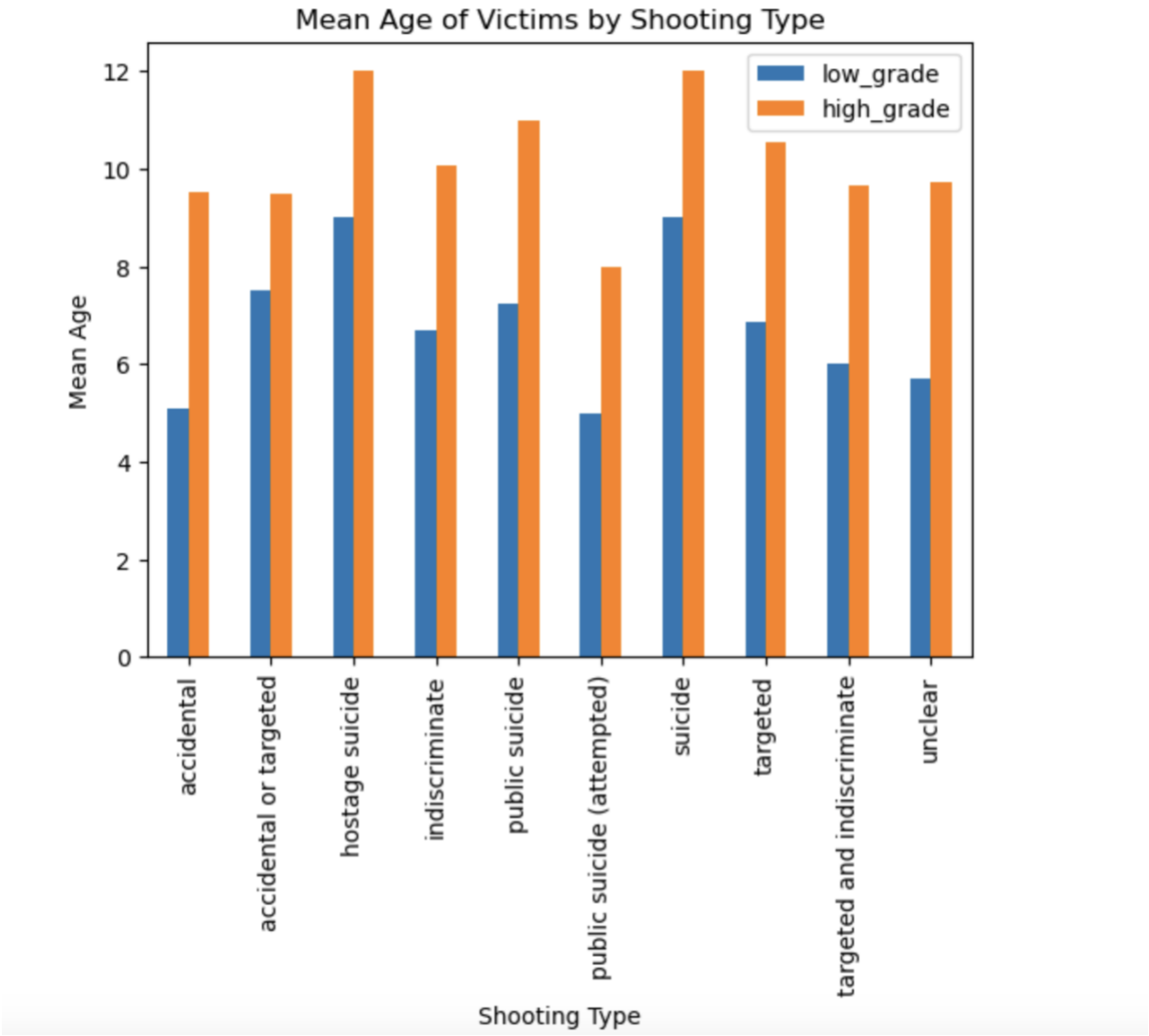
V. DATA SET GRAPHICAL EXPLORATION

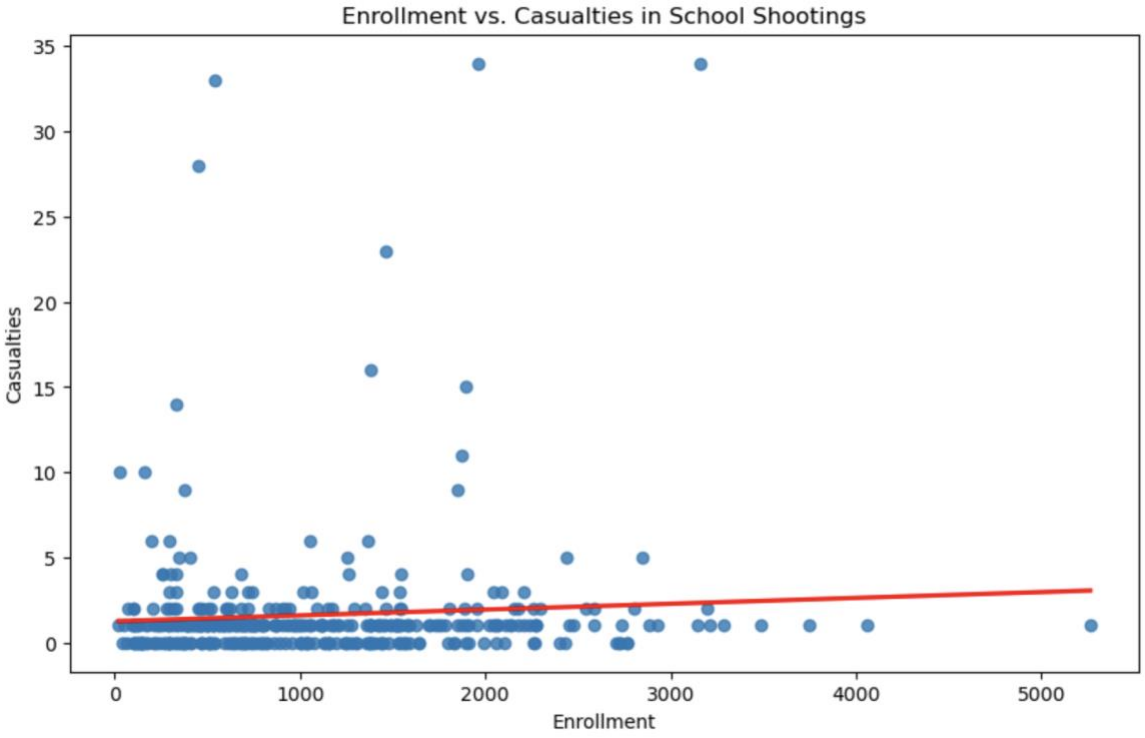
Narrative introduction to the section. In each section below, indicate any interesting distributions, anomalies, imbalance, etc. that you notice.

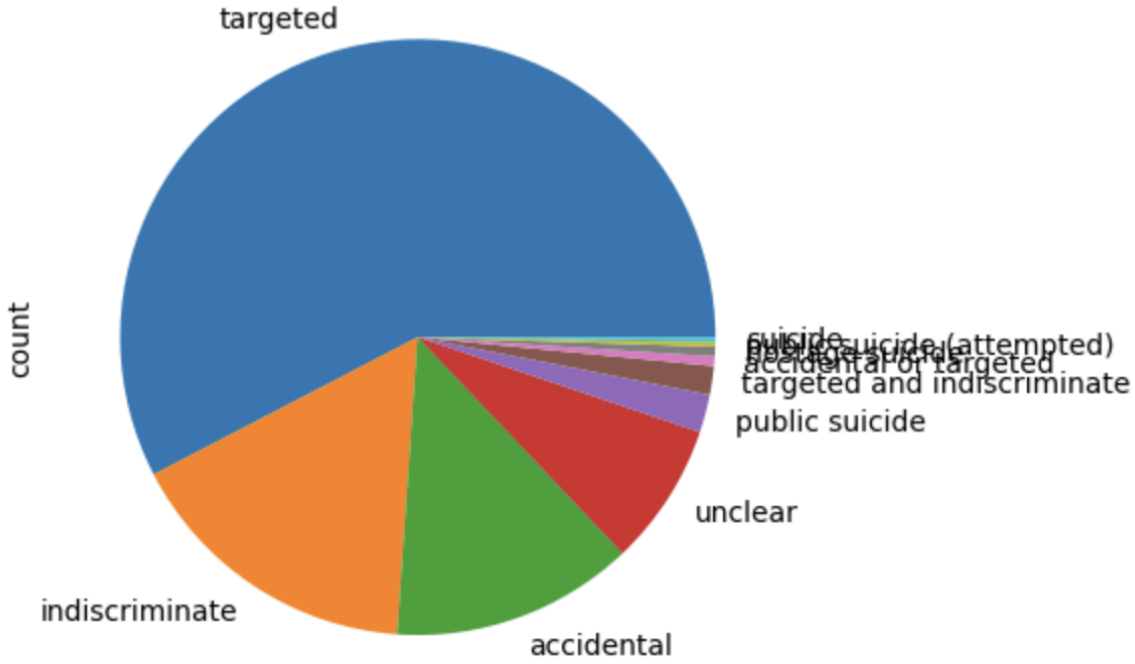
A. Distributions
B. ScatterPlots / Pairwise Plots (continuous variables)
C. Barcharts (categorical variables)
D. Other Plots - don't skimp – there are likely other plots that would be useful that I haven't already specified. Include those in this section.

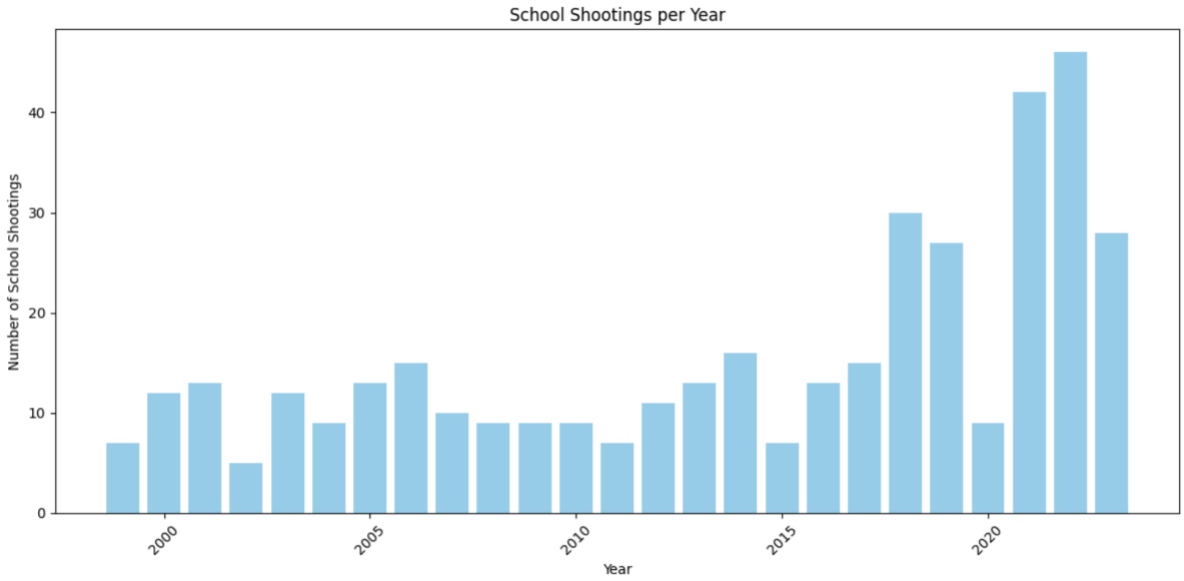


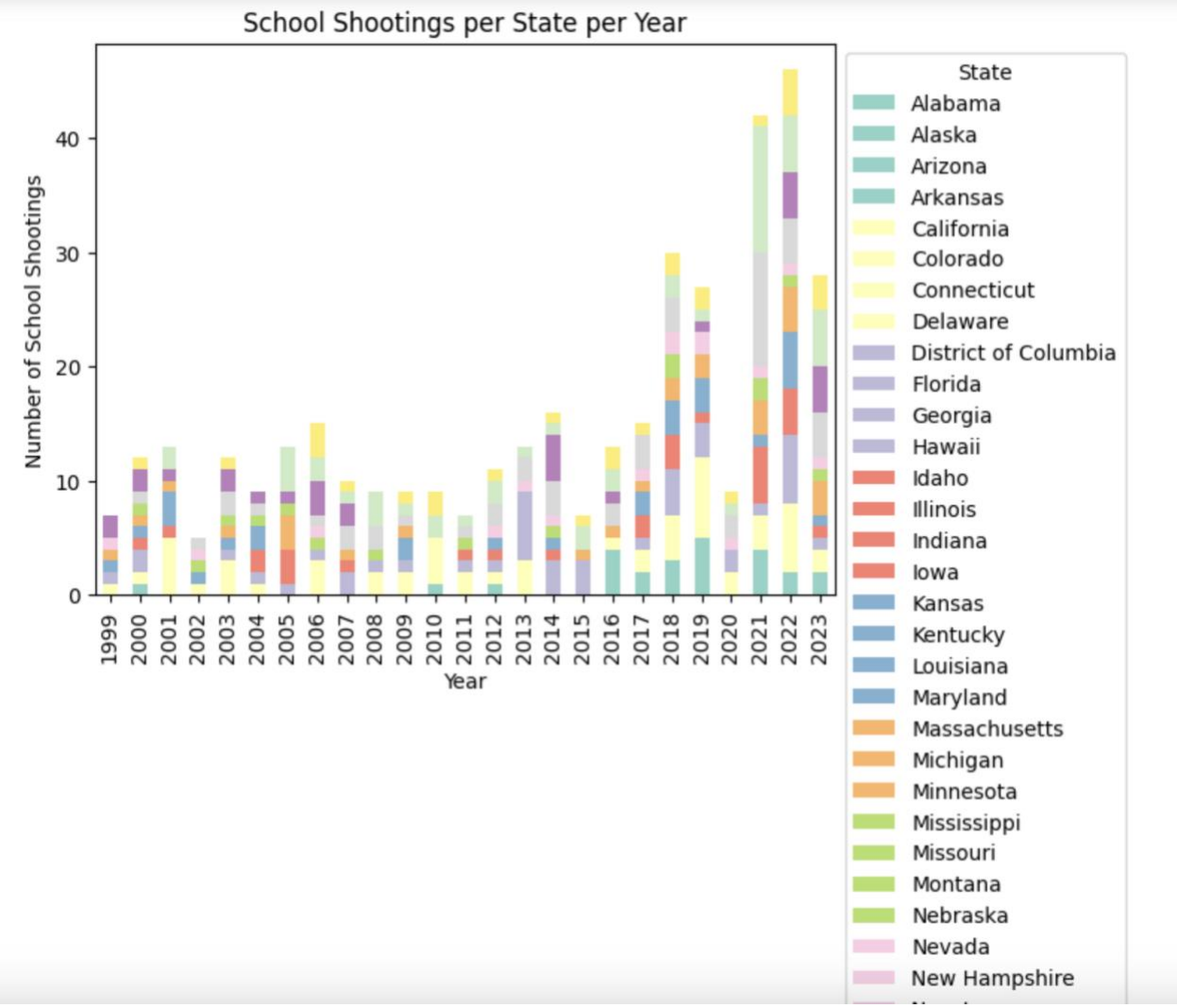

Deaths by Shooting Type
Killed
120
100
80
60
40
20
0
indiscriminate
targeted
Shooting Type


Mean Age of Victims by Shooting Type
low_grade
high_grade
Mean Age
0
2
4
6
8
10
12
accidental
accidental or targeted
hostage suicide
indiscriminate
public suicide
public suicide (attempted)
suicide
targeted
targeted and indiscriminate
unclear
Shooting Type


Enrollment vs. Casualties in School Shootings
Casualties
Enrollment
0
5
10
15
20
25
30
35
0
1000
2000
3000
4000
5000


targeted
count
indiscriminate
accidental
unclear
public suicide
targeted and indiscriminate
accidental or targeted
suicide
public suicide (attempted)


School Shootings per Year
Number of School Shootings
0
10
20
30
40
2000
2005
2010
2015
2020
Year



VI. SUMMARY OF FINDINGS

Finish up with a paragraph or two of summarizing your findings about this data set.

Our school shootings data set consisted of 50 columns and 387 rows and the earliest recorded shooting in this set was in 1999 with the most recent happening this year in 2023. The columns included things like the school name, date it happened, the state, enrollment of the school, killed, injured, casualties, shooter age, gender, and race, along with many more categories.

As we dove into the data, we were able to find some interesting facts and correlations. Just starting with a basic but useful finding, there averaged one person injured and .5 people killed per shooting. As we moved into the correlations, most were fairly low so when analyzing the data we took that into account when looking at “stronger” correlations. This being said, there was a very weak correlation between the age of the shooter and if someone was killed or injured, with correlations of -.07 and .01. There was one of the stronger correlations of .2 when looking at if the shooter was deceased and if they killed anyone. There was also the strongest correlation within the races of the shooter with white people and casualties. Based on our regplot, we also found that there is almost no correlation between teh number of casualties and the enrollment of the school.

A couple other interesting findings based on some graphs that we had were that the most shootings have happened in recent years. 2022, 2021, 2018, 2023, 2019 are the top five years for the number of shootings, showing an upward trend. Most casualties happened on Tuesdays with the least amount on Thursdays, and around 120 of the shootings were random while around 70 were targeted. And finally we found that most shootings took place in California with 10%.